Survey Methods for Transport Planning

Anthony J. Richardson
Elizabeth S. Ampt
Arnim H. Meyburg
# TABLE of CONTENTS

1. **Introduction**
   1.1 THE ROLE OF SURVEYS
      1.1.1 Types of Surveys
      1.1.2 Survey Purposes
   1.2 THE TRANSPORT SURVEY PROCESS
   1.3 TRADE-OFFS IN TRANSPORT SURVEY DESIGN
   1.4 STRUCTURE OF THE BOOK

2. **Preliminary Planning of Surveys**
   2.1 OVERALL STUDY OBJECTIVES
   2.2 SURVEY OBJECTIVES
   2.3 REVIEW OF EXISTING INFORMATION
   2.4 FORMATION OF HYPOTHESES
   2.5 DEFINITION OF TERMS
   2.6 DETERMINATION OF SURVEY RESOURCES
      2.6.1 Costs of Surveys
      2.6.2 Time Estimates for Surveys
      2.6.3 Personnel Requirements
      2.6.4 Space Requirements
   2.7 SURVEY CONTENT

3. **Selection of Survey Method**
   3.1 TIME FRAME FOR SURVEY
      3.1.1 Longitudinal Surveys
   3.2 TYPES OF DATA COLLECTION TECHNIQUE
      3.2.1 Documentary Searches
      3.2.2 Observational Surveys
      3.2.3 Household Self-Completion Surveys
      3.2.4 Telephone Surveys
      3.2.5 Intercept Surveys
      3.2.6 Household Personal Interview Surveys
      3.2.7 Group Surveys
      3.2.8 In-Depth Interviews
   3.3 SUMMARY OF SURVEY METHOD SELECTION

4. **Sampling Procedures**
   4.1 TARGET POPULATION DEFINITION
   4.2 SAMPLING UNITS
   4.3 SAMPLING FRAME
   4.4 SAMPLING METHODS
      4.4.1 Simple Random Sampling
      4.4.2 Stratified Random Sampling
      4.4.3 Variable Fraction Stratified Random Sampling
      4.4.4 Multi-Stage Sampling
      4.4.5 Cluster Sampling
      4.4.6 Systematic Sampling
      4.4.7 Non-Random Sampling Methods
   4.5 SAMPLING ERROR AND SAMPLING BIAS
   4.6 SAMPLE SIZE CALCULATIONS
      4.6.1 Sample Sizes for Population Parameter Estimates
      4.6.2 Sample Sizes for Hypothesis Testing
# Table of Contents

4.7  **VARIANCE ESTIMATION TECHNIQUES**  126  
   4.7.1  Variability in Simple Random Samples  126  
   4.7.2  Design Effects  127  
   4.7.3  Replicate Sampling  133  
4.8  **DRAWING THE SAMPLE**  142  

5.  **Survey Instrument Design**  147  
   5.1  **MORE TRADE-OFFS IN TRANSPORT SURVEY DESIGN**  147  
   5.2  **SCOPE OF THIS CHAPTER**  150  
   5.3  **QUESTIONNAIRE CONTENT**  151  
      5.3.1  Length of the Questionnaire  151  
      5.3.2  Relevance of the Questions  153  
      5.3.3  Reasonableness of the Questions  154  
      5.3.4  The Context of Questions about Trips  154  
      5.3.5  Questionnaire Design to Maximise Trip Recording  155  
   5.4  **PHYSICAL DESIGN OF SURVEY FORMS**  159  
   5.5  **QUESTION TYPES**  166  
      5.5.1  Classification Questions  166  
      5.5.2  Factual Questions  167  
      5.5.3  Opinion and Attitude Questions  168  
      5.5.4  Stated Response Questions  183  
   5.6  **QUESTION FORMAT**  187  
      5.6.1  Open Questions  187  
      5.6.2  Field-Coded Questions  189  
      5.6.3  Closed Questions  191  
   5.7  **QUESTION WORDING**  194  
      5.7.1  Use Simple Vocabulary  194  
      5.7.2  Use Words Appropriate to the Audience  195  
      5.7.3  Length of Questions  196  
      5.7.4  Clarify the Context of Questions  196  
      5.7.5  Avoid Ambiguous Questions  196  
      5.7.6  Avoid Double-Barrelled Questions  197  
      5.7.7  Avoid Vague Words About Frequency  197  
      5.7.8  Avoid Loaded Questions  197  
      5.7.9  The Case of Leading Questions  198  
      5.7.10  Avoid Double Negatives  202  
      5.7.11  Stressful Questions  202  
      5.7.12  Avoid Grossly Hypothetical Questions  202  
      5.7.13  Allow for the Effect of Response Styles  203  
      5.7.14  Care with Periodicity Questions  203  
      5.7.15  Use of an Activity Framework  204  
      5.7.16  Flow of the Question  204  
      5.7.17  Always Use a Pilot Test  204  
   5.8  **QUESTION ORDERING**  205  
   5.9  **QUESTION INSTRUCTIONS**  207  
      5.9.1  Self-Completion Surveys  207  
      5.9.2  Personal Interview Surveys  209  

6.  **Pilot Surveys**  213  
   6.1  **WHY A PILOT SURVEY?**  213  
      6.1.1  A Test of ALL Aspects of Survey Design  214  
      6.1.2  A Need for Several Tests  215  
   6.2  **USES OF THE PILOT SURVEY**  216
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2.1 Adequacy of the Sampling Frame</td>
<td>216</td>
</tr>
<tr>
<td>6.2.2 Variability of Parameters Within the Survey Population</td>
<td>216</td>
</tr>
<tr>
<td>6.2.3 Non-Response Rate</td>
<td>217</td>
</tr>
<tr>
<td>6.2.4 Method of Data Collection</td>
<td>217</td>
</tr>
<tr>
<td>6.2.5 “Skirmishing” of Question Wording</td>
<td>218</td>
</tr>
<tr>
<td>6.2.6 Layout of the Questionnaire</td>
<td>218</td>
</tr>
<tr>
<td>6.2.7 Adequacy of the Questionnaire in General</td>
<td>218</td>
</tr>
<tr>
<td>6.2.8 Efficiency of Interviewer and Administrator Training</td>
<td>220</td>
</tr>
<tr>
<td>6.2.9 Data Entry, Editing and Analysis Procedures</td>
<td>220</td>
</tr>
<tr>
<td>6.2.10 Cost and Duration of Survey</td>
<td>221</td>
</tr>
<tr>
<td>6.2.11 Efficiency of Survey Organisation</td>
<td>221</td>
</tr>
<tr>
<td>6.3 SIZE OF THE PILOT SURVEY</td>
<td>222</td>
</tr>
<tr>
<td>7. Administration of the Survey</td>
<td>223</td>
</tr>
<tr>
<td>7.1 GENERIC PROCEDURES FOR SURVEY ADMINISTRATION</td>
<td>223</td>
</tr>
<tr>
<td>7.1.1 Recruitment of Survey Staff</td>
<td>224</td>
</tr>
<tr>
<td>7.1.2 Survey Staff Training</td>
<td>228</td>
</tr>
<tr>
<td>7.1.3 Pre-Publicity</td>
<td>232</td>
</tr>
<tr>
<td>7.1.4 Survey Execution and Monitoring</td>
<td>234</td>
</tr>
<tr>
<td>7.1.5 Follow-up and Validation Procedures</td>
<td>237</td>
</tr>
<tr>
<td>7.1.6 Confidentiality</td>
<td>237</td>
</tr>
<tr>
<td>7.1.7 Response Rates</td>
<td>238</td>
</tr>
<tr>
<td>7.2 ADMINISTRATION OF SELF-COMPLETION SURVEYS</td>
<td>239</td>
</tr>
<tr>
<td>7.2.1 The Use of a Reminder Regime</td>
<td>239</td>
</tr>
<tr>
<td>7.2.2 Validation Methods</td>
<td>240</td>
</tr>
<tr>
<td>7.2.3 Sponsorship of Survey</td>
<td>242</td>
</tr>
<tr>
<td>7.2.4 Consideration of Respondents</td>
<td>243</td>
</tr>
<tr>
<td>7.2.5 Use of Incentives</td>
<td>243</td>
</tr>
<tr>
<td>7.2.6 Covering Letter</td>
<td>243</td>
</tr>
<tr>
<td>7.2.7 Use of Comments Section</td>
<td>243</td>
</tr>
<tr>
<td>7.2.8 Provision of a Phone-in Service</td>
<td>244</td>
</tr>
<tr>
<td>7.2.9 Preparation of Questionnaires for Mailing</td>
<td>244</td>
</tr>
<tr>
<td>7.2.10 Type of Postage Used</td>
<td>244</td>
</tr>
<tr>
<td>7.2.11 Response Rates</td>
<td>245</td>
</tr>
<tr>
<td>7.3 ADMINISTRATION OF PERSONAL INTERVIEW SURVEYS</td>
<td>246</td>
</tr>
<tr>
<td>7.3.1 Use of a Robust Interview Regime</td>
<td>246</td>
</tr>
<tr>
<td>7.3.2 Training of Interviewers</td>
<td>247</td>
</tr>
<tr>
<td>7.3.3 Checking of Interviews</td>
<td>247</td>
</tr>
<tr>
<td>7.3.4 Satisfactory Response Rate</td>
<td>249</td>
</tr>
<tr>
<td>7.3.5 Correct Asking of Questions</td>
<td>249</td>
</tr>
<tr>
<td>7.3.6 Number of Interviewer Call-Backs</td>
<td>250</td>
</tr>
<tr>
<td>7.4 ADMINISTRATION OF TELEPHONE INTERVIEW SURVEYS</td>
<td>251</td>
</tr>
<tr>
<td>7.4.1 Sampling</td>
<td>251</td>
</tr>
<tr>
<td>7.4.2 Dealing with Non-Response</td>
<td>252</td>
</tr>
<tr>
<td>7.4.3 Improving Data Quality</td>
<td>253</td>
</tr>
<tr>
<td>7.4.4 Other Aspects of Administration</td>
<td>254</td>
</tr>
<tr>
<td>7.5 ADMINISTRATION OF INTERCEPT SURVEYS</td>
<td>255</td>
</tr>
<tr>
<td>7.5.1 Sampling</td>
<td>255</td>
</tr>
<tr>
<td>7.5.2 Training of Staff</td>
<td>257</td>
</tr>
<tr>
<td>7.5.3 Other Administrative Details</td>
<td>258</td>
</tr>
<tr>
<td>7.6 ADMINISTRATION OF IN-DEPTH INTERVIEW SURVEYS</td>
<td>258</td>
</tr>
<tr>
<td>7.6.1 Organisation of &quot;Pre-Interviews&quot;</td>
<td>258</td>
</tr>
<tr>
<td>7.6.2 Preparation for the Main Interview</td>
<td>259</td>
</tr>
</tbody>
</table>
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>The Transport Survey Process</td>
<td>5</td>
</tr>
<tr>
<td>1.2</td>
<td>Trade-Offs in the Transport Survey Process</td>
<td>9</td>
</tr>
<tr>
<td>1.3</td>
<td>Effects of Instrument Quality and Sample Size on Uncertainty</td>
<td>11</td>
</tr>
<tr>
<td>1.4</td>
<td>Effects of Instrument Quality and Sample Size on Cost</td>
<td>11</td>
</tr>
<tr>
<td>1.5</td>
<td>Combined Effects of Instrument Quality and Sample Size</td>
<td>12</td>
</tr>
<tr>
<td>1.6</td>
<td>Optimal Combinations of Instrument Quality and Sample Size</td>
<td>13</td>
</tr>
<tr>
<td>2.1</td>
<td>The Transport Planning Process</td>
<td>18</td>
</tr>
<tr>
<td>2.2</td>
<td>The Dimensions of System Boundaries</td>
<td>21</td>
</tr>
<tr>
<td>2.3</td>
<td>The Systems Modelling Process</td>
<td>22</td>
</tr>
<tr>
<td>3.1</td>
<td>Trade-offs in Selection of the Survey Method</td>
<td>34</td>
</tr>
<tr>
<td>3.2</td>
<td>A Completed HATS Display Board</td>
<td>67</td>
</tr>
<tr>
<td>3.3</td>
<td>Decision Trees in the Situational Approach Survey</td>
<td>69</td>
</tr>
<tr>
<td>3.4</td>
<td>The Roles of Interactive Interviews</td>
<td>71</td>
</tr>
<tr>
<td>4.1</td>
<td>A Population of 100 Sampling Units</td>
<td>82</td>
</tr>
<tr>
<td>4.2</td>
<td>A Population of 100 Sampling Units with Identifiers</td>
<td>82</td>
</tr>
<tr>
<td>4.3</td>
<td>A Simple Random Sample of 10 Sampling Units</td>
<td>83</td>
</tr>
<tr>
<td>4.4</td>
<td>A Simple Random Sample from a Stratified Population</td>
<td>84</td>
</tr>
<tr>
<td>4.5</td>
<td>A Stratified Random Sample from a Stratified Population</td>
<td>85</td>
</tr>
<tr>
<td>4.6</td>
<td>Examples of Respondent Selection Grids</td>
<td>91</td>
</tr>
<tr>
<td>4.7</td>
<td>A Systematic Sample from a Stratified Population</td>
<td>94</td>
</tr>
<tr>
<td>4.8</td>
<td>The Distinction between Accuracy and Precision</td>
<td>97</td>
</tr>
<tr>
<td>4.9</td>
<td>The Confusion between Accuracy and Precision</td>
<td>98</td>
</tr>
<tr>
<td>4.10</td>
<td>Distribution of the Parameter in the Population</td>
<td>104</td>
</tr>
<tr>
<td>4.11</td>
<td>Distribution of the Means of Independent Samples</td>
<td>105</td>
</tr>
<tr>
<td>4.12</td>
<td>Confidence Limit Estimator for Continuous Variables</td>
<td>112</td>
</tr>
<tr>
<td>4.13</td>
<td>Confidence Limit Estimator for Discrete Variables</td>
<td>113</td>
</tr>
<tr>
<td>4.14</td>
<td>Input Screen for Sample Size Design Parameters</td>
<td>114</td>
</tr>
<tr>
<td>4.15</td>
<td>Input Screen for Expected Values of Variables in Population</td>
<td>115</td>
</tr>
<tr>
<td>4.16</td>
<td>Output Screen for Estimated Values of Variables in Sample</td>
<td>115</td>
</tr>
<tr>
<td>4.17</td>
<td>Probability of a Type I Error</td>
<td>118</td>
</tr>
<tr>
<td>4.18</td>
<td>Probability of a Type II Error</td>
<td>119</td>
</tr>
<tr>
<td>4.19</td>
<td>Variation in $\beta$ as a function of $\beta$</td>
<td>125</td>
</tr>
<tr>
<td>5.1</td>
<td>Questionnaire Imprecision (Variability) and Inaccuracy (Bias)</td>
<td>149</td>
</tr>
<tr>
<td>5.2</td>
<td>Recall-Only Method of Trip Reporting</td>
<td>156</td>
</tr>
<tr>
<td>5.3</td>
<td>Maximising Trip Reporting in a Self-Completion Form</td>
<td>157</td>
</tr>
<tr>
<td>5.4</td>
<td>Verbal Activity Recall Framework in Personal Interviews</td>
<td>157</td>
</tr>
<tr>
<td>5.5</td>
<td>Example of an Activity Diary</td>
<td>158</td>
</tr>
<tr>
<td>5.6</td>
<td>Sample Page of Personal Interview Survey Form</td>
<td>161</td>
</tr>
<tr>
<td>5.7</td>
<td>Example of Self-Completion Questionnaire Survey</td>
<td>164</td>
</tr>
<tr>
<td>5.8</td>
<td>Classification Question, with Branching</td>
<td>167</td>
</tr>
<tr>
<td>5.9</td>
<td>Distribution of Response to a Single Stimulus</td>
<td>172</td>
</tr>
<tr>
<td>5.10</td>
<td>Distribution of Response to Two Different Stimuli</td>
<td>172</td>
</tr>
<tr>
<td>5.11</td>
<td>Part of a Paired Comparison Question</td>
<td>173</td>
</tr>
<tr>
<td>5.12</td>
<td>Similarities Ranking Question</td>
<td>175</td>
</tr>
<tr>
<td>5.13</td>
<td>Category Scale Question</td>
<td>176</td>
</tr>
<tr>
<td>5.14</td>
<td>Likert Scale Question</td>
<td>177</td>
</tr>
<tr>
<td>5.15</td>
<td>Semantic Differential Scale Question</td>
<td>178</td>
</tr>
</tbody>
</table>
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.16</td>
<td>Comparing Category Scale and Magnitude Estimation Ratings</td>
<td>180</td>
</tr>
<tr>
<td>5.17</td>
<td>An Open Question in a Personal Interview</td>
<td>187</td>
</tr>
<tr>
<td>5.18</td>
<td>Field-Coded Questions (D&amp;E) in a Personal Interview</td>
<td>190</td>
</tr>
<tr>
<td>5.19</td>
<td>A Closed Question in a Personal Interview</td>
<td>191</td>
</tr>
<tr>
<td>5.20</td>
<td>An Open Alternative in a Closed Question</td>
<td>192</td>
</tr>
<tr>
<td>5.21</td>
<td>A Show Card for a Closed Question</td>
<td>193</td>
</tr>
<tr>
<td>5.22</td>
<td>The Quintamensional Plan of Question Design</td>
<td>194</td>
</tr>
<tr>
<td>5.23</td>
<td>Poor Placement of Tick-Box Labels</td>
<td>197</td>
</tr>
<tr>
<td>5.24</td>
<td>An Introduction to a Set of Questions</td>
<td>205</td>
</tr>
<tr>
<td>5.25</td>
<td>The Use of a Completed Sample Questionnaire</td>
<td>208</td>
</tr>
<tr>
<td>7.1</td>
<td>Graphical Representation Used in Validation Interviews</td>
<td>241</td>
</tr>
<tr>
<td>7.2</td>
<td>Typical HATS Interview Procedure</td>
<td>259</td>
</tr>
<tr>
<td>8.1(a)</td>
<td>Map used in Self-Coding of Location Coordinates (front)</td>
<td>277</td>
</tr>
<tr>
<td>8.1(b)</td>
<td>Map used in Self-Coding of Location Coordinates (back)</td>
<td>278</td>
</tr>
<tr>
<td>8.2</td>
<td>General Procedure for Geocoding of Locations</td>
<td>279</td>
</tr>
<tr>
<td>8.3</td>
<td>The Location of Full Street Addresses using MapInfo</td>
<td>283</td>
</tr>
<tr>
<td>8.4</td>
<td>The Location of Cross Street Addresses</td>
<td>284</td>
</tr>
<tr>
<td>8.5</td>
<td>The Occurrence of Multiple Cross Street Locations</td>
<td>285</td>
</tr>
<tr>
<td>8.6</td>
<td>Cross Streets with Multiple Bordering CCDs</td>
<td>286</td>
</tr>
<tr>
<td>8.7</td>
<td>The Allocation of a Cross Street Address to a CCD</td>
<td>286</td>
</tr>
<tr>
<td>8.8</td>
<td>Sample Questionnaire to be Coded</td>
<td>293</td>
</tr>
<tr>
<td>8.9</td>
<td>A Simple Spreadsheet Layout</td>
<td>294</td>
</tr>
<tr>
<td>8.10</td>
<td>A Replicate Spreadsheet Layout</td>
<td>294</td>
</tr>
<tr>
<td>8.11</td>
<td>A Relational Database for Travel Survey Data.</td>
<td>295</td>
</tr>
<tr>
<td>8.12</td>
<td>A Relational Database Created in FoxBASE.</td>
<td>297</td>
</tr>
<tr>
<td>8.13</td>
<td>Selection of a Trip Record in the Trip File</td>
<td>297</td>
</tr>
<tr>
<td>8.14</td>
<td>Selection of Corresponding Person Record in the Person File</td>
<td>297</td>
</tr>
<tr>
<td>8.15</td>
<td>Selection of the Corresponding Household Record.</td>
<td>298</td>
</tr>
<tr>
<td>8.16</td>
<td>A Flatfile Created for Statistical Analysis</td>
<td>299</td>
</tr>
<tr>
<td>9.1</td>
<td>The Effect of Reminders on Questionnaire Response Rates</td>
<td>325</td>
</tr>
<tr>
<td>9.2</td>
<td>The Speed of Response by Response Type</td>
<td>326</td>
</tr>
<tr>
<td>9.3</td>
<td>The Effect of Household Size on Level and Speed of Response</td>
<td>327</td>
</tr>
<tr>
<td>9.4</td>
<td>The Effect of Age on Level and Speed of Response</td>
<td>328</td>
</tr>
<tr>
<td>9.5</td>
<td>The Effect of Employment and Sex on Speed of Response</td>
<td>328</td>
</tr>
<tr>
<td>9.6</td>
<td>Travel Characteristics as a Function of Response Speed</td>
<td>331</td>
</tr>
<tr>
<td>9.7</td>
<td>Average Stop Rate as a Function of Response Wave</td>
<td>332</td>
</tr>
<tr>
<td>9.8</td>
<td>Stop Rate as a Function of Cumulative Response</td>
<td>333</td>
</tr>
<tr>
<td>10.1</td>
<td>A Flat File Created for Statistical Analysis</td>
<td>340</td>
</tr>
<tr>
<td>10.2</td>
<td>Data Desk's Representation of Variables as Icons</td>
<td>340</td>
</tr>
<tr>
<td>10.3</td>
<td>The Contents of Data Desk Variable Icons</td>
<td>341</td>
</tr>
<tr>
<td>10.4</td>
<td>A Histogram of the Number of Trips per Person per Day</td>
<td>342</td>
</tr>
<tr>
<td>10.5</td>
<td>Summary Statistics of the Number of Trips per Person per Day</td>
<td>342</td>
</tr>
<tr>
<td>10.6</td>
<td>A Boxplot of the Number of Trips per Person per Day</td>
<td>343</td>
</tr>
<tr>
<td>10.7</td>
<td>A Scatterplot of Trip Rate vs Household Vehicles</td>
<td>344</td>
</tr>
<tr>
<td>10.8</td>
<td>Splitting the Data into Groups Based on Vehicle Ownership</td>
<td>344</td>
</tr>
<tr>
<td>10.9</td>
<td>Boxplots of Trips per Day for the Vehicle Ownership Groups</td>
<td>345</td>
</tr>
<tr>
<td>10.10</td>
<td>Scatterplot of Person trips vs Household Vehicles</td>
<td>345</td>
</tr>
<tr>
<td>10.11</td>
<td>A Simple, Bivariate Linear Relationship</td>
<td>349</td>
</tr>
<tr>
<td>10.12</td>
<td>A Simple, Bivariate, Monotonic Non-linear Relationship</td>
<td>349</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>10.13</td>
<td>Specification Error and Model Complexity</td>
<td>351</td>
</tr>
<tr>
<td>10.14</td>
<td>Measurement Error and Model Complexity</td>
<td>352</td>
</tr>
<tr>
<td>10.15</td>
<td>Relationship between Total Error and Model Complexity</td>
<td>353</td>
</tr>
<tr>
<td>10.16</td>
<td>The Effect of Bad Data on Total Model Error</td>
<td>353</td>
</tr>
<tr>
<td>10.17</td>
<td>Scatter Diagram of a Positive Linear Relationship</td>
<td>357</td>
</tr>
<tr>
<td>10.18</td>
<td>Scatter Diagram of a Negative Linear Relationship</td>
<td>357</td>
</tr>
<tr>
<td>10.19</td>
<td>Scatter Diagram showing No Relationship</td>
<td>358</td>
</tr>
<tr>
<td>10.20</td>
<td>Scatter Diagram showing Non-Linear Relationship</td>
<td>358</td>
</tr>
<tr>
<td>10.21</td>
<td>Geometric Representation of the Regression Hypothesis</td>
<td>361</td>
</tr>
<tr>
<td>10.22</td>
<td>Scatter Diagram of n Data Points</td>
<td>362</td>
</tr>
<tr>
<td>10.23</td>
<td>Prediction if Error Is Assumed to Be in Xi Measures</td>
<td>363</td>
</tr>
<tr>
<td>10.24</td>
<td>Selection of Linear Regression Option in Data Desk</td>
<td>367</td>
</tr>
<tr>
<td>10.25</td>
<td>Results of Linear Regression Analysis in Data Desk</td>
<td>367</td>
</tr>
<tr>
<td>10.26</td>
<td>Selecting a Residuals Scatterplot in Data Desk</td>
<td>368</td>
</tr>
<tr>
<td>10.27</td>
<td>Results of Residuals Scatterplot in Data Desk</td>
<td>368</td>
</tr>
<tr>
<td>10.28</td>
<td>Results of the Ungrouped Linear Regression</td>
<td>369</td>
</tr>
<tr>
<td>10.29</td>
<td>Distribution of the Residuals</td>
<td>370</td>
</tr>
<tr>
<td>10.30</td>
<td>Raw Attribute Scores for Four Car Models</td>
<td>378</td>
</tr>
<tr>
<td>10.31</td>
<td>Factor Scores for Four Car Models</td>
<td>379</td>
</tr>
<tr>
<td>10.32</td>
<td>Frequency Distribution of Car versus Transit Users</td>
<td>380</td>
</tr>
<tr>
<td>10.33</td>
<td>Definition of Discrimination between Two Subpopulations</td>
<td>382</td>
</tr>
<tr>
<td>10.34</td>
<td>Logit and Probit Curves for Equations 10.65 and 10.66</td>
<td>394</td>
</tr>
<tr>
<td>11.1</td>
<td>Comparison of 3-D and 2-D Pie-Charts</td>
<td>397</td>
</tr>
<tr>
<td>11.2</td>
<td>Comparison of Non-Zero and Zero Origin Charts</td>
<td>398</td>
</tr>
<tr>
<td>11.3</td>
<td>Representing One-Dimensional Quantities with 2-D Objects</td>
<td>399</td>
</tr>
<tr>
<td>11.4</td>
<td>Comparison of Low Data-ink and High Data-ink Ratio Graphs</td>
<td>400</td>
</tr>
<tr>
<td>11.5</td>
<td>Poor Choices of Shading Patterns</td>
<td>400</td>
</tr>
<tr>
<td>11.6</td>
<td>Superfluous Gridlines and 3-D Effects</td>
<td>401</td>
</tr>
</tbody>
</table>
List of Tables

LIST of TABLES

Table 2.1  Sources of Existing Transport Data  
Table 3.1  Comparison of Successive Sample Surveys and Panel Studies  
Table 3.2  Uses of Each Survey Method  
Table 4.1  Possible End-States of Hypothesis Testing  
Table 4.2  Results of the Stratified Sample Airport Upgrading Survey  
Table 4.3  Results of the Cluster Sample Airport Upgrading Survey  
Table 4.4  Results of the Replicated Sample Airport Upgrading Survey  
Table 4.5  Sample Data for Half-Sample Replication Example  
Table 4.6  Sample Data for Balanced Half-Sample Replication  
Table 5.1  Summary of Scale Types  
Table 5.2  A Simple Stated Response Experimental Design  
Table 6.1  Pilot Studies for the 1981 Sydney Travel Survey  
Table 7.1  Response Behaviour of Non-Response Validation Households  
Table 9.1  Responses by Age and Sex  
Table 9.2  Trip-Rates in Responses by Age and Sex  
Table 9.3  Population Breakdown by Age and Sex  
Table 9.4  Response Rates by Age and Sex  
Table 9.5  Population Expansion Factors by Age and Sex  
Table 9.6  Total Trips in Population by Age and Sex  
Table 9.7  Marginal Population Totals by Age and Sex  
Table 9.8  Expanded Population Totals after First Iteration  
Table 9.9  Expanded Population Totals after Second Iteration  
Table 9.10  Estimated Population Expansion Factors by Age and Sex  
Table 9.11  Total Trips in Population by Age and Sex  
Table 9.12  Incomplete Information for Various Question Types  
Table 9.13  Incomplete Information for Various Types of Respondent  
Table 9.14  Non-Reported Trips for Various Types of Respondent  
Table 9.15  Trip Characteristics of Non-Reported Trips  
Table 9.16  Increases in Mobility after Allowing for Non-Reported Trips  
Table 9.17  Non-Reporting Correction Factors for Expected Added Stops  
Table 9.18  Non-Reporting Correction Factors for Unexpected Added Stops  
Table 9.19  Non-Reported Stop Weights (phone connected)  
Table 9.20  Non-Reported Stop Weights (phone not connected)  
Table 9.21  Reasons for Non-Response in Self-Administered Surveys  
Table 9.22  Household Characteristics of SEQHTS Respondents by Wave  
Table 9.23  Personal Characteristics of SEQHTS Respondents by Wave  
Table 10.1  Input Data and Results for Weighted Regression  
Table 10.2  Attributes Used to Rate Various Car Models  
Table 10.3  Factor Groupings for Attributes of Table 10.2  
Table 10.4  Results of Discriminant Analysis on Chicago Data  
Table 10.5  Additional Chicago Commuters to be Classified  
Table 10.6  Observations on Carbon Monoxide Level (parts per million)  
Table 11.1  Characteristics of the Friendly Data Graphic
1. Introduction

1.1 THE ROLE OF SURVEYS

All persons involved in transport and land-use planning will at some stage be involved with data collection. Even if not directly concerned with the design and conduct of surveys, they will certainly wish to use data at some time, and at that stage they will realise what should have been done in the design and administration phases of the survey. Each individual's introduction to survey data may have widely differing emotional connotations. For some, it may be a horrific experience as they try to grapple with data which has been collected by someone else, only to find that the documentation associated with that dataset is incomplete, misleading or simply non-existent. For others, who face the prospect of collecting a new set of data, it may be a challenging professional experience with ample opportunity for initiative and the exploration of new frontiers of knowledge in survey methodology. This book is intended to guide the latter group, and console the former.
Chapter 1

1.1.1 Types of Surveys

Surveys are of particular relevance to transport and land-use planning in several specific areas. Land-use surveys are an integral component of transport planning information requirements. This is due to the fact that travel is a so-called "derived demand". That is, travel, in itself, has no inherent value; it is useful only insofar as it facilitates participation in other activities. Thus with respect to passenger transport, travel enables individuals to participate in an activity at a location some distance away from their present location. If the activity did not exist, or if it could be undertaken at the individual's present location, then there would be no need for travel to occur; that is, if there were no land-use activities, there would be no travel. The spatial location and intensity of land-use activities is measured by land-use surveys.

The amount of travel which takes place between land-uses will depend on the quality and quantity of the transport system which connects the land-uses, and surveys of the transport system inventory play a major role in specifying the location and characteristics of the available transport system. This system, which includes both public and private modes of transport, may be described in terms of three basic components: the right-of-way, terminals, and vehicles. The right-of-way includes the roads, tracks and paths which may be used by different types of vehicles and which can be described in terms of length, direction of movement, capacity, and design standards. The terminals include public transport stations, modal interchanges and parking stations and can be described in terms of location, throughput capacity, and holding capacity. The vehicles include both public and private vehicles and may be described in terms of total number, passenger (or goods) capacity, comfort level, and various operating characteristics.

The combination of land-use activity and a transport system invariably results in trip-making, and to measure the type and extent of trip-making it is necessary to conduct travel pattern surveys by one means or another. Such travel patterns may be described in terms of who is going where, with whom, at what time, by which mode and route, and for what purpose. The measurement of such travel patterns is perhaps the unique part of transport survey methods, but determining the most effective way of obtaining the above information has often received little attention.

One effect of trip-making is to change the way in which the transport system operates. To establish whether the transport system is coping adequately with the demands being placed on it, it is therefore necessary to conduct transport system performance surveys. Such surveys seek to measure performance
Introduction

characteristics, such as travel times, travel time variability, passenger waiting times, vehicle occupancies, and system safety.

Each of the above types of surveys has the common characteristic that it attempts to measure system characteristics as they now exist. However, a major task of the transport planner is to attempt to predict changes which will occur in the system. In particular, the planner is often required to predict changes in travel demand as a result of changes in the physical system or changes in the operating characteristics of that system. In attempting to predict the demands which will be placed on a transport system, it is well recognised that various groups in the population will react differently to changes in the transport system. To identify these groups, it is necessary to incorporate demographic and socio-economic surveys within the overall transport survey framework. It is also well recognised that individuals react not to actual changes in the transport system but to perceived changes in that system. For that reason, perception and attitude surveys often form a vital component of transport planning surveys. Data from these surveys are often used as inputs to travel demand models. In many cases, a transport survey will fulfil more than one of the above roles.

While all types of surveys in transport are referred to, and while the principles of questionnaire design can apply to any survey type, the main focus of this book is on travel pattern surveys, that is surveys which ask people how, where or why they travel. Furthermore, while we cover those surveys which we have called intercept surveys (i.e. on board public transport vehicles, at the road side, and at activity centres such as shops and airports) a great deal of attention is placed on household surveys - since the methodology used here can frequently be applied to all other survey types.

1.1.2 Survey Purposes

Regardless of the subject matter to be covered within a survey, transport surveys may serve several purposes, either alone or in combination. First, they may merely attempt to describe existing conditions at a given time in order to ascribe an order of magnitude to various transport phenomena. Secondly, they may seek to establish causal explanations of conditions at a given time so that greater understanding of transport system behaviour may be obtained. Thirdly, it may be desired that after analysis of the survey results, predictive models will be derived so as to forecast future transport conditions or to predict the effects of system changes. Fourthly, rather than predict the effects of system changes, it is often more appropriate, or convenient, to measure the effects of system changes. In this case, before-and-after surveys may be used to assess the effects of these system changes. Fifthly, an extension of the before-and-after concept (where surveys are generally conducted at two points in time) is the establishment of a
regular series of monitoring surveys whereby changes in transport system characteristics or behaviour may be established over a long period.

In addition to the above-mentioned purposes, surveys may also play two further roles which, while perhaps technically undesirable, should be clearly recognised. Surveys, for example, may often be used as "report padding" to fill an otherwise empty document and to lend weight to the conclusions or opinions contained in it. Alternatively, the conduct of a survey may be a convenient method of putting off a decision. The use of surveys in this way, as an aid to procrastination, may often be embedded within the more general gambit of calling for a Committee of Inquiry when decisions are not wanted (yet). A clear recognition of the purpose of a survey, in terms of any one of the above seven categories, can greatly aid in the initial selection of the survey technique, the design of the survey method and the later interpretation of results.

This is not to imply, however, that the latter two survey purposes should be accepted unquestioningly by the professional transport surveyor. If the purpose of the report appears to be report-padding or procrastination, the client should first be questioned more insistently on the objectives of the survey. If no professionally acceptable objectives can be specified, then you, as the survey designer, may face a difficult decision. You can either refuse to carry out the survey or else agree to design the survey (perhaps under protest). The grounds for refusing to carry out the survey are not simply academic piety; rather it is for the highly pragmatic reason that if too many such surveys are carried out, the whole area of transport surveys may be brought into ill-repute in the public mind. If, however, the conduct of the survey is inevitable, either by you or by someone else, then you should attempt to make the most of the situation and try to incorporate as much experimentation into the conduct of the survey as possible. As will be seen later, there are many ideas in survey design and administration which require empirical validation. If the survey can be used to test several survey design hypotheses, then that survey will have gained a legitimate (research) objective.

1.2 THE TRANSPORT SURVEY PROCESS

The conduct of a survey is not an informal procedure. Rather, it should follow a series of logical, interconnected steps which progress toward the final end-product of the survey. The stages in a typical sample survey are shown in Figure 1.1, and the issues to be addressed within each of these stages are listed on the next pages.
Figure 1.1 The Transport Survey Process
(a) Preliminary Planning
   (i) Overall Study Objectives
   (ii) Specific Survey Objectives
   (iii) Review of Existing Information
   (iv) Formation of Hypotheses
   (v) Definition of Terms
   (vi) Determination of Survey Resources
   (vii) Specification of Survey Content

(b) Selection of Survey Method
   (i) Selection of Survey Time Frame
   (ii) Selection of Survey Technique
   (iii) Consideration of Survey Errors

(c) Sample Design
   (i) Definition of Target Population
   (ii) Sampling Units
   (iii) Sampling Frame
   (iv) Sampling Method
   (v) Sampling Error and Sampling Bias
   (vi) Sample Size and Composition
   (vii) Estimation of Parameter Variances
   (viii) Conduct of Sampling

(d) Survey Instrument Design
   (i) Types of Survey Instrument
   (ii) Question Content
   (iii) Trip Recording Techniques
   (iv) Physical Nature of Forms
   (v) Question Types
   (vi) Question Format
   (vii) Question Wording
   (viii) Question Ordering
   (ix) Question Instructions

(e) Pilot Survey(s)
   (i) Adequacy of Sampling Frame
   (ii) Variability within Survey Population
   (iii) Estimation of Non-Response Rate
   (iv) Size of the Pilot Survey
   (v) Suitability of Survey Method
   (vi) Adequacy of Questionnaire (schedule)
   (vii) Efficiency of Interviewer Training
Introduction

(viii) Suitability of Coding, Data Entry, and Editing Procedures
(ix) Suitability of Analysis Procedures
(x) Cost and Duration of Surveys
(xi) Efficiency of Organisation

(f) Administration of the Survey
   (i) Procedures for Survey Administration of: Self-Completion, Personal Interview, Telephone, Intercept and In-depth Interview Surveys
   (ii) Survey Execution and Monitoring
   (iii) Quality Control
   (iv) The Use of the Computer in Transport Surveys

(g) Data Processing
   (i) Selection of Coding Method
   (ii) Preparation of Code Format
   (iii) Development of Data Entry Programs
   (iv) Coder and Data Entry Training
   (v) Coding Administration

(h) Data Editing
   (i) Editing of Field Sheets
   (ii) Verification of Data Entry
   (iii) Development of Editing Computer Programs
   (iv) Consistency and Range Checks

(i) Data Correction and Expansion
   (i) Editing Check Corrections
   (ii) Secondary Data Comparisons
   (iii) Corrections for Internal Biases

(j) Data Analysis and Management
   (i) Exploratory Data Analysis
   (ii) Model Building
   (iii) Interpretation of Results
   (iv) Database Management
   (v) Provision of Data Support Services

(k) Presentation of Results
   (i) Verbal Presentations
   (ii) Visual Presentations
   (iii) Preparation of Reports
   (iv) Publication of Results
Chapter 1

(i) Tidying-Up
   (i) Documentation of Survey Method
   (ii) Storage and Archival of Data
   (iii) Completion of Administrative Duties

In the survey process outlined above, there are three types of linkages between activities: forward, feedback and backward linkages. The forward linkages are relatively obvious, e.g. the questionnaire design cannot begin until the survey method has been selected. The feedback linkages indicate that two or more activities must be performed sequentially in a closed loop, e.g. having performed the pilot survey, it may be necessary to redesign the questionnaire and then pilot test the new questionnaire. Backward linkages indicate that information must be transferred from an activity which occurs later in the process to one which occurs early in the process. For example, the design of the questionnaire will be affected by the coding procedure to be used later, while the coding procedure will depend on the type of analysis to be performed on the data. While such backward linkages may not be highly visible, it is important that consideration be given to them so that decisions made early in the process will not proscribe options for later data analysis.

1.3 TRADE-OFFS IN TRANSPORT SURVEY DESIGN

As will become evident at several points throughout this book, the authors believe that the essence of good survey design is being able to make trade-offs between the competing demands of good design practice in several areas (such as sample design, survey instrument design, conduct of surveys, and data weighting and expansion) so as to arrive at the most cost effective, high quality survey which meets the needs of the client within budget constraints. The overall nature of these trade-offs is shown in Figure 1.2.

The underlying nature of this trade-off process is the so-called "Architect's Triangle", in which quantity and quality are traded-off against cost. A trade-off occurs because it is impossible to control all three of the major elements in Figure 1.2.; at best, only two of the three can be controlled by the survey designer. Thus given a fixed budget, as is normally the case, the decision to obtain data of a specified quality will automatically control the quantity of data which can be collected. Alternatively, within a fixed budget, specification of the quantity of data to be collected will immediately dictate the quality of data which can be collected. That is, we can collect a greater quantity of low quality data or we can collect a limited amount of higher quality data for a given budget. Generally, the latter course of action is to be preferred.
The quality of data to be collected is a function of the survey method selected and the quality of the sample (insofar as the sample is free of bias). The quality of data obtained from any survey method will, in turn, be a function of the quality of the survey instrument design (i.e. does it collect information on the variables of interest in an unbiased way) and the quality control procedures put in place for the implementation of that survey method (i.e. what follow-up procedures will be used to verify the quality of the data collected). The quality of the sample will depend on the ability of the sampling frame to truly represent the population, and the extent to which the sample selection procedures result in a random selection from the sampling frame.

The quantity of data collected will be a function of the number of respondents in the final dataset and the amount of information obtained from each respondent. This, in itself, presents a trade-off situation because any attempt to collect more information from each respondent (beyond a threshold level of information) may result in less respondents responding. The total number of respondents will
Chapter 1

obviously depend on the size of the sample drawn from the population and the response rate obtained from that sample. The amount of information obtained from each respondent will depend on the number of questions asked in the survey, as well as the depth of the questions asked. Thus some surveys can be effective with only a large number of highly directed questions, while others need to explore a few topics in depth. The extent of this trade-off is therefore a specific design decision on the part of the survey designer. The trade-off will also be partly determined by the respondents themselves. As the length of the survey increases, the response rate will generally decrease (the rate of decrease will depend on such factors as the interest level of the survey topic to the respondents, and the overall quality of the survey instrument design). There will therefore be a point at which an increase in the number of questions asked will result in the collection of less data in total, because of the more than proportional decrease in the response rate. The survey designer should therefore be cognisant of this interaction when making the explicit trade-off between number of respondents and information obtained per respondent.

The nature of these trade-offs can be illustrated by a simple example related to the quality of the survey instrument and the size of the sample. Both of these components of survey design have an effect on the quality of data obtained from a survey, in that each can affect how well the reported information matches the true situation in the population. As will be described later in this book, a larger sample can reduce the uncertainty involved in assuming that the sample results represents the results which would have been obtained from the population. Similarly, a good survey instrument (e.g. a good questionnaire) can reduce the uncertainty that the answers obtained from respondents represent the real conditions about which we intended to obtain information. These relationships between instrument quality, sample size and the uncertainty involved in the sample results are shown in Figure 1.3.

It can be seen that uncertainty can be reduced either by designing a better survey instrument or by selecting a larger sample, or both. However, improvements in the survey instrument or the sample do not come without some cost. In each case, there will be fixed costs involved (either in the design of a very poor quality instrument, or the selection of a very small sample), plus marginal costs for increases in the sample size and improvements in the survey instrument, as shown in Figure 1.4.
Introduction

Figure 1.3 Effects of Instrument Quality and Sample Size on Uncertainty

Figure 1.4 Effects of Instrument Quality and Sample Size on Cost

Figure 1.4 postulates that the marginal costs associated with improvements in instrument design and increases in sample size are different, but linear. In reality, the marginal costs will probably not be linear, but this does not detract from the present argument.

Since uncertainty is affected by both the instrument design and the sample size, this means that each combination of instrument design and sample size will possess a specific degree of uncertainty in the final results. This can be displayed in the form of uncertainty isoquants (lines of equal uncertainty) for combinations of instrument design and sample size, as shown in Figure 1.5. It can be seen that as either the instrument design is improved or the sample size is increased, the level of uncertainty with the sample results decreases. Importantly, different combinations of instrument design and sample size can produce the same level of uncertainty, as shown by the combinations of the two factors lying along a single
Chapter 1

isoquant. This implies that the two factors can be traded off against each other to achieve the same end results in terms of uncertainty of the sample results (note that this argument does not extend to the case of reducing bias in the results, where a poor survey instrument or a poor sample is consistently giving the wrong result. In such a situation, taking a bigger sample and using a biased survey instrument will simply give us the wrong answers with more certainty).

Figure 1.5 Combined Effects of Instrument Quality and Sample Size

Figure 1.5 also shows that different combinations of survey instrument quality and sample size will result in different total costs, as shown by the cost isoquants in the right half of Figure 1.5. As either the instrument design is improved or the sample size is increased, the cost of the survey increases. Importantly, different combinations of instrument design and sample size can produce the same survey cost, as shown by the combinations of the two factors lying along a single isoquant. This implies that the two factors can be traded off against each other to achieve the same end results in terms of the cost of the survey. One of these isoquants represents the budget for the survey and, even though combinations of instrument design and sample size above this isoquant would result in decreased uncertainty in the results, such combinations are not feasible within a fixed budget. The essence of effective survey design is to find the right combination of survey instrument quality and sample size which minimises the uncertainty of the results within the available budget.

The search for the most cost-effective survey design can be depicted by the diagrams shown in Figure 1.6, where the cost isoquant line from Figure 1.5 corresponding to the budget is overlaid on the uncertainty isoquant diagram of Figure 1.5. The point where the cost isoquant is tangential to the uncertainty isoquant (as indicated by the small circle) is the combination of instrument quality
and sample size which minimises the level of uncertainty for a given level of expenditure. In the case of the diagram on the left of Figure 1.6, where the cost isoquant reflects a lower marginal cost for improving instrument quality (see Figure 1.4), the optimal situation is one where more attention is paid to improving instrument quality than increasing the sample size.

![Diagram showing optimal combinations of instrument quality and sample size.](image)

**Figure 1.6 Optimal Combinations of Instrument Quality and Sample Size**

However, if the marginal cost of increasing sample size was lower than the marginal cost of improving instrument quality, the cost isoquant line would be steeper and this would result in a different optimal design as shown in the diagram on the right of Figure 1.6. Thus the "best" survey design is dependent on the quality and the relative costs of instrument quality and sample size.

The foregoing argument has been made to illustrate the nature of the trade-offs inherent in survey design, and to demonstrate that concentrating on either the instrument design or the sample design, to the virtual exclusion of the other, will most likely be a pointless strategy in trying to find the most cost-effective solution for the client's needs.

However, at this point in time in the evolution of the field of survey methodology, the above argument must remain qualitative rather than quantitative. This is because the areas of sample design and instrument design have developed in different directions and at different rates. Sample design is highly quantified and estimates of uncertainty can be calculated as a function of sample size. However, the area of instrument design has largely developed through the experience of individual survey designers and the dissemination of a limited amount of research findings. At this point in time, we do not know (quantitatively) how improvements in survey instrument design will affect the
Chapter 1

uncertainty in the results obtained from respondents, nor has much been documented about the costs involved in improving survey instrument designs.

One of the objectives of this book is to promote the adoption of a balanced approach to total survey design, and to advance the state of knowledge about the role of survey instrument design in the conduct of transport and activity surveys.

1.4 STRUCTURE OF THE BOOK

The structure of this book follows from the survey process outlined in Figure 1.1. Thus Chapter 2 describes some of the tasks to be addressed in the preliminary planning of a survey and, in particular, describes some steps which are often overlooked but which are essential to a survey if it is to function smoothly in later stages of the process. Chapter 3 describes the types of survey methods commonly used in transport and land-use planning and the circumstances where each method might be used most effectively. Advantages and disadvantages of each method are discussed and, in particular, the types of biases inherent in each method are described. Chapter 4 outlines some procedures which can be used in selecting a sample and highlights the important, but distinctively different, concepts of sampling error and sampling bias. Chapter 4 also follows on by describing the techniques used to determine minimum sample size requirements on the basis of achieving an acceptable sampling error.

Chapter 5 is concerned with the principles involved in the design of survey instruments (e.g. questionnaires and interview schedules). The various options available for obtaining detailed records of trip and activity patterns are also described in this chapter. The techniques involved in recording trip patterns are of particular importance and serve to distinguish transport survey methodology from the more general social survey methodology. The trip recording techniques are, however, more generally applicable to the recording of many forms of activity patterns.

The conduct of pilot surveys is the focus of attention in Chapter 6. Pilot surveys are often ignored, or undervalued, by survey designers; this chapter highlights the reasons for conducting pilot surveys and emphasises the value of such pilot surveys and pre-tests. Chapter 7 describes the design of administrative procedures for various survey types, and highlights the importance of a pre-planned approach to follow-up procedures to enable later corrections for non-response and other sources of survey bias.

Once the data has been collected via the methods outlined in Chapters 2 through 7, attention turns to working with the data which has been collected. It should be realised that, at this stage, not much more can be done to improve the quality of
the data; if the design has not accounted for the various aspects of bias and error before now, then little can be done subsequently. In Chapter 8, the areas of data processing known as coding, editing and data entry are covered. These tasks, which are often tedious, are essential prerequisites to the analysis of the data and should not be skipped over in one's hurry to get to the analysis phase of the survey process. The quality of the analysis is only as good as the quality of the input data; so ensure that the data set is "clean" (i.e. free from error and bias) before embarking on the analysis.

An important component of obtaining clean data is the use of various data correction and expansion techniques which attempt to make the sample data more nearly representative of the population which it is trying to represent. These techniques are described in detail in Chapter 9, where attention is focussed on the use of a variety of data weighting techniques to account for socio-demographic differences in the sample and the population, as well as for the effects of non-response and non-reported information.

Chapter 10 outlines the wide variety of data analysis techniques which can be used on the clean data set. Attention is focussed on the use of widely available microcomputer spreadsheets and database programs, which facilitate the entry, editing, analysis and presentation of data in one integrated package, and on the use of commercially available microcomputer statistical packages. This chapter outlines the concepts of "exploratory data analysis" which can be used to get a good feel for the nature of the data collected. There is also a relatively brief overview of a number of multivariate analysis techniques which can be used in the "model-building" phase of data analysis. Aspects of on-going database management are also covered in this chapter.

Chapter 11 concludes the book by providing some guidelines for the preparation of survey reports and the presentation of survey results in a variety of ways. It also highlights the importance of adequate survey method documentation and tidying-up of the many details associated with a survey project.

A special feature of this book is the Survey Design Checklist located in the yellow-paged Appendix at the rear of the book. This Checklist consists of a series of questions which record the major features of a survey, and which serve as a reminder of the decisions which must be made in the design and conduct of the survey. Each question is linked back to the content of the book by the provision of references to relevant sections in the book. The Checklist pages can be photocopied and used as a design aid for any survey. The completed Checklist also serves as a component of the documentation of the survey. Finally, when this book is used as a textbook, the Survey Design Checklist serves as a useful means of revision for the student of transport survey methodology.
Chapter 1

Whilst attempting to provide a comprehensive framework for the conduct of transport and land-use activity surveys, it will be obvious that this book does not attempt to cover all aspects of sample survey design in equal detail. In particular, it does not cover the detailed statistical design associated with various types of complex sampling strategies; numerous textbooks have been written on the details of sample design (e.g. Cochran, 1977; Kish, 1965; Yates, 1971) and the reader is referred to these for further details. Similarly, although an overview of exploratory and multivariate data analysis is given, relatively little is provided about the details of the many techniques which can be used in the analysis phase, since the range of such techniques is quite wide and the selection of the analysis technique will vary between individual surveys. For a comprehensive description of many multivariate data analysis methods, the interested reader is referred to Stopher and Meyburg (1979).

The reader should also be aware that this book concentrates on the application of social survey techniques to the recording of personal travel and activity behaviour. While many of the same principles apply, this book does not concentrate on traffic surveys (see Taylor and Young (1988) for an exposition of traffic survey methods) or commercial vehicle surveys.

Finally, as the reader either now knows, or will soon find out, the real learning experience in survey design comes not from reading books but from conducting surveys in the field. Only then will some of the seemingly trivial points made in this book assume their full importance; while this book provides a framework for the conduct of transport surveys, it is no substitute for experience.
2. Preliminary Planning of Surveys

At the preliminary planning stage of the survey process, a number of basic issues must be addressed before a decision is taken to proceed with the design and conduct of a sample survey. It is assumed throughout this book that it is always a sample survey being considered, because in transport it is usually very difficult to survey an entire population; it is also usually unnecessary to sample the entire population, as will be shown in Chapter 4.

2.1 OVERALL STUDY OBJECTIVES

It should be realised from the outset that the collection of data is one part of a more comprehensive transport planning process. Many contemporary authors have suggested various ways in which the urban transport planning process should operate (e.g. Hutchinson, 1974; Stopher and Meyburg, 1975; Dickey et al., 1975; Morlok, 1978; Ortúzar and Willumsen, 1994). By adapting the critical components of each of these versions of the planning process, a general transport planning systems process can be postulated as shown in Figure 2.1. It can be seen that data collection is but one of the tasks leading towards the evaluation, selection and implementation of a particular transport strategy. For example, in planning for the introduction of a toll road system, collecting data about who currently uses the route would be an important component of the process leading to a decision about whether a toll should be introduced, how much it should cost, and which route it should take.
The starting point in the planning process, however, (if indeed a starting point can be defined in such a continuous process) is the definition of the problem (or problems) under consideration. It is probably the most important single component of the planning process. Very often, the careful definition of the problem will greatly assist in suggesting possible solutions. Indeed, the explicit enunciation of the problem may well be a crucial step in the solution of the problem itself, and may obviate the need for surveys and subsequent data analysis.

As an aid to understanding the significance of problem definition, consider the following definition of a problem: "A problem for an individual or a group of individuals is the difference between the desired state for a given situation at a given time and the actual state. This difference cannot be eliminated immediately (if ever)." (Dickey et al., 1975). Four factors concerning this definition are of significance. Firstly, the problem is the concern of a limited population and is not necessarily of general concern to the population at large. This fact is of some importance when attempting to define the "goals of the community"; there is no one set of goals, they are different for everybody. This disparity is often a source of considerable misunderstanding, since those people (such as planners) not directly affected by the problem may find it difficult to comprehend the nature and/or magnitude of the problem.
Preliminary Planning

Secondly, the identification of the desired state may in itself present a problem, in that it may be very difficult to determine just what that desired state is, or should be, for a particular situation. For example, the identification of a desirable level of air pollution may present problems because the desirable state may not be represented by a zero pollution level. There are finite levels of air pollution which are easily tolerated by human, and other living, organisms. However, increasing air pollution levels above a threshold brings with it substantial problems. The determination of these threshold levels, as a possible representation of a desired state, may involve considerable difficulty and controversy.

Thirdly, the fact that the problem may never be solved makes transport problems somewhat different to other mathematical or physical science problems which may have been previously encountered. This insolubility is due to the complex interweaving of various transport problems. Thus, the best solution for one problem may well create problems in other areas. For example, reducing the air pollution emissions from internal combustion engines may result in increased fuel consumption. Alternatively, solving a problem for one group in the community may simply create a problem for another group of individuals, e.g. shifting heavy trucks off one road and onto another road.

Fourthly, it is important to talk about problems "at a given time". This is because, firstly, the nature of transport problems changes over time and, secondly, even if a transport problem could be solved, the solution to that problem would create a rise in the level of expectation of the population in question such that a gap would once again appear between the desired and actual state. This dilemma can be discouraging for the planner since it implies that most problems can never be solved "once and for all"; the planner can never satisfy the community.

Nonetheless, the importance of problem definition is extremely high, as highlighted by Armstrong (1978) who states that "a good job of identifying the problems will reduce the likelihood of Type III Errors (Type III Errors are good solutions to the wrong problems)."

Since a problem has been defined as a discrepancy between the desired and actual states of the system, it is obviously necessary to ascertain these two states before attempting the definition of the problem. The desired state of the system may be reflected in the values and goals of the community. Stopher and Meyburg (1975) made the useful distinction between values, goals, objectives and criteria. Each of these elements represents a restatement, and refinement in more operational terms, of the preceding element. To attempt problem definition, both the values and goals must have been defined. Thus, for example, if the value was defined as being "Environmental Amenity", one possible goal might be defined as "the containment of traffic noise levels in residential areas".
Chapter 2

To determine whether a problem existed with respect to this goal, it would be necessary to ascertain whether, in fact, traffic noise levels were presently considered to be satisfactory in residential areas. To do this, it would be necessary to resort to some form of data collection to determine the present state of the system. At this stage of the planning process, the data collection might be relatively informal. For example, letters of complaint to the authorities about traffic noise levels in residential areas might suffice as an indication of the present state of the system with respect to this parameter.

If a discrepancy exists between the desired and actual states, then a problem may be deemed to exist with respect to this aspect of the system. If no discrepancy exists, then other aspects of the system should be examined to determine whether problems exist in those areas.

Having defined the problem, or at least expressed one's perception of the problem, it is then possible to determine the system boundaries within which solutions to the problem will be sought. The definition of the system boundaries is performed on the basis of an intuitive assessment of the likely impacts of the problem and may be subject to later revision. As shown in Figure 2.2, these system boundaries may be defined in three dimensions; spatial, social and temporal.

The spatial dimension defines the system boundaries with respect to the geographical area to be considered in an evaluation study or a sample survey. With respect to transport, a useful trichotomy is between site, corridor and network studies. Alternatively, the spatial dimensions may be defined in accordance with geopolitical boundaries pertaining to local government, state government and national government boundaries. The social dimension in Figure 2.2 relates to the social groups who are to be included within the system boundary of the study. These groups may be defined in terms of their relationship to the transport system. For example, as a result of any change to the transport system there may be system users who benefit from the change, users who lose as a result of the change, and non-users who may be either adversely or beneficially affected by the change. The introduction of a high occupancy vehicle (or transit) lane readily brings to mind examples of all three of these. Alternatively, the social groups may be defined with respect to a wide range of socio-economic variables where, for example, the objective of the study might be to examine the distributional impacts of the current system or changes to that system. The temporal dimension in Figure 2.2 relates to the time horizon over which a solution is to be sought. Thus problems, or impacts, may be considered with respect to their short-term, medium-term, or long-term implications. Obviously, the interaction between solutions identified for each of these time-frames needs to be identified and assessed for compatibility.
Given an explicit statement of the problem, and the boundaries within which the problem is perceived to exist, it is then possible to refine the goals into more specific, operational measures, termed *objectives* and *criteria*, which will later be used as a basis for comparison in the evaluation phase of the planning process.

Continuing the traffic noise example quoted earlier, the objective might be "acceptable traffic noise levels on residential streets in inner urban areas", whilst the final criterion might be "a maximum noise level of 68dB(A) L_{10} (18 hour) at a distance of 1 metre from the building facade for residential buildings along Minor Street in Hushville". At this stage, there now exists a specific numerical desired state to work towards in seeking a solution. Obviously, other criteria could equally well be specified for this aspect of the transport system, as well as for other aspects. The degree to which each criterion is satisfied will depend, among other things, on the importance placed on each criterion.

The allocation of *resources* to the solution of the problem is a critical step since it affects other key components of the planning process. This allocation, in terms of financial and manpower commitments, will determine the amount and type of data which can be collected, the type, number and scale of feasible alternative solutions which can be considered, the degree of sophistication needed, and if possible, the models which will be used in the analysis.

A key feature of the transport planning process is the use of *models* to describe the operation of the system. Conceptually, it would be possible to investigate the operation of the system by actually implementing an alternative in the field and...
then observing its effects on the surrounding ecological, economic and social environment. However, this approach has a number of economic, financial, political, social and engineering drawbacks and, apart from a limited number of trial schemes, this method is seldom used. Hence some form of system model, or simplified abstraction of the real world, must be relied upon to generate predictions of consequences. A framework for consideration of the different types of system model is shown in Figure 2.3. There are three different types of model; supply models, which describe how physical systems work; demand models; which describe how users of the system make their decisions; and impact models, which describe the impacts of usage of the system in terms of the economic, environmental and social consequences.

![Figure 2.3 The Systems Modelling Process](image)

The formulation of the system models is governed by the resources available and the objectives of the analysis. In many cases, the "model" is no more than the experience of the particular planner involved in the analysis. In other situations, the model is a complex mathematical model which takes account of the many
systems interactions. In all cases, the model simply makes predictions of the likely consequences of the alternatives to be analysed.

The generation of the *alternatives* by which the problem might be solved is possibly the most challenging part of the process from a professional point of view, in that it requires considerable creativity on the part of the planner to generate alternatives which will meet the desired criteria within the constraints imposed on the problem solution. While much of the transport planning process is concerned with the application of logical and reasoned thought processes, the generation of alternatives should concentrate on illogical and unreasonable thought processes. All transport planners should become familiar with the writings on lateral thinking by Edward deBono (e.g. deBono, 1967, 1972, 1988). In his works, deBono stresses that no new ideas can come from logical thinking; all new ideas come from illogical and somewhat random thoughts. With this in mind, deBono invented a number of thinking strategies such as provocation, random words and “po” statements (po stands for "provocation operation"). deBono describes how such strategies can result in significantly new ideas which can overcome problems which could not have been solved through the application of conventional ideas. However, deBono also stresses that, once discovered, all of these new ideas must be able to be explained and justified in terms of conventional logical thought processes; in essence, all great new ideas are obvious in hindsight. Combining logical and illogical thinking is one of the great challenges in transport planning.

The range of alternatives which might be considered is quite wide and may include one of more of the following:

- do nothing
- change transport technology
- construct new facilities
- change methods of operation
- substitute some other activity for transport
- change the regulations or legislation
- change pricing policies
- change public attitudes
- tackle the urban problems which cause transport problems

The prediction of the *consequences* of various alternatives may necessitate a revision of the system boundaries if it appears that there are likely to be substantial impacts outside of the existing system boundaries. This may then involve a revision of the objectives and criteria and a change in the models to be used to predict an expanded set of consequences.

The comparison of the predicted consequences with the stated criteria is then performed in the *evaluation* process. If no alternatives are deemed to be acceptable
Chapter 2

as a result of the evaluation, then a search should be made for new alternatives which may be acceptable in accordance with the stated criteria. If, after an intensive search, there appear to be no acceptable alternatives then it may be necessary to perform a re-examination of the goals and objectives to determine whether they are unattainable and whether it may be possible to lower the standards of the criteria without serious consequences.

If one or more of the alternatives are finally deemed acceptable, then a selection of the best alternative is made on the basis of the stated criteria. This alternative is then the subject of implementation in the field, depending on any constraints which may be imposed on implementation by parties outside of the transport planning process. Examples of such external constraints include political and electoral considerations.

The final phase of the planning process is the monitoring of the performance of the selected alternative under actual field conditions. This monitoring process will give rise to data on the actual operation of the alternative. This data on operation and consequences may provide a basis for recalibration, or reformulation, of the system models to enable better predictions to be made of future consequences. This monitoring may also suggest changes which should be made to the selected alternatives to improve operations. The changes can then be modelled and evaluated to predict new operating conditions. Finally, monitoring should be performed to ascertain any changes in goals and objectives which may affect the selection of alternatives over time. The inclusion of this monitoring step is essential and highlights the fact that planning is a continual process which does not stop with the generation of a set of plans for implementation. These plans must be continually revised in accordance with changes in community goals, changing economic conditions and developing technology.

Only after the overall study objectives have been specified can the question of how to design and conduct the survey be considered. In particular, the objectives of the study will determine the content of the survey and the analytical techniques required to address the stated problem. The definition of the system boundaries will assist in specifying the survey population and other components relevant to the design of the sample. It should also be recognised and accepted that the explicit statement of the problem may well obviate the need for a survey, either because the statement of the problem provides the solution to the problem or because it becomes clear that a sample survey may not assist in devising solutions to the problem.

2.2 SURVEY OBJECTIVES

Given the objectives of the overall study, and noting the various survey purposes as outlined in Chapter 1, it is necessary at this stage to define in quantitative and
Preliminary Planning

qualitative terms the objectives of the survey. Basically, one must ask just what questions are to be answered by the survey and how the information obtained from such a survey can be used to assist in the overall planning process. Unless these questions can be answered clearly at this stage, it is highly likely that problems will occur at several later stages in the survey process.

It may well be useful, however, to distinguish between two different types of survey; project planning surveys which are surveys addressing a specific issue (e.g. what kind of people use public transport?) and research study surveys where data is being collected for more general use with multiple purposes rather than a specific application in mind. At the outset, the questions to be answered in a research survey will be much less clearly defined than in a project planning survey. This is understandable because, by the very nature of research, if one knew all the questions to be asked at the outset, then the study would no longer be a research study. However for project planning surveys, there would be much less uncertainty about the concepts and questions involved; all that is missing is the empirical validation of these concepts. In the discussion to follow, it will be assumed that project planning surveys are the central topic and hence it should be possible to clearly define the survey objectives.

An example of survey objectives for a large-scale travel survey may be drawn from the 1981 Sydney Travel Survey, which was a personal interview survey of 20,000 households carried out in Sydney, Australia between July and December 1981 (Ampt, 1981). The stated objectives of this survey were:

(a) To provide a description of current travel behaviour, so that requests for information from various transport and planning authorities could be accommodated; and

(b) To collect data for specific policy analysis purposes, and for short-term forecasting using disaggregate behavioural models of transport choice.

In order to achieve these major stated objectives, the following informal objectives were also formulated (Ampt and West, 1985):

(c) To improve on aspects of the methodology deemed to be problematic in the 1971 Sydney Area Transportation Study (SATS) survey and in other documented surveys;

(d) To ensure that the data reflected important travel-related factors such as flexible working hours; and

(e) To include information on behavioural (travel and activity) alternatives other than those reported on the sample travel day.
Chapter 2

2.3 REVIEW OF EXISTING INFORMATION

Before embarking on the collection of a new set of data, it is prudent to ascertain just how much is already known on the subject in question. Whilst this would appear to be a commonsense step, it is often overlooked, with the result that many surveys either obtain data which simply duplicates that which is already available, or else obtain data which is not completely appropriate to the problem at hand. The review of existing information can provide assistance in two major ways.

First, it may be possible to unearth existing relevant data sources. If one is extremely fortunate, it may be possible to find data which are exactly those required, and hence the need for a further survey is completely obviated. In many cases, however, even though appropriate data is found it may still be necessary to re-analyse the data in the light of the problem at hand. Wigan (1985) gives an excellent review of some of the opportunities and pitfalls involved in such secondary use of transport survey data.

Often not all the required data is found, and hence the need for a further survey is only partially removed. Whilst not removing the need for further data collection, other data sources may be of great assistance in the design of that survey. For example, estimates of population variance for relevant parameters from a previous survey may assist in sample size calculations, while knowledge of the composition of the population (e.g. what percentage of people own cars) may assist in stratified sample design. Finally, data may be found in a different form but one which nevertheless will be useful as a cross-check on the accuracy of the data which is subsequently collected. Many sources of existing data are available for transport planning purposes and a comprehensive list of such data sources would be extremely long - and would vary from country to country. An example of such data sources, compiled for Australian conditions by Bowyer (1980), is shown in Table 2.1. Axhausen (1994) has also prepared a very comprehensive list of major travel surveys which have been carried out throughout the world since the 1950s.

The second way in which an information review might provide assistance is in the revelation of methodological procedures which may be appropriate in the survey. The art of survey design is based on the acquisition of knowledge through experience, and there is no substitute to learning from one's mistakes in the field. Unfortunately, many survey designers appear to take this generalisation to extremes, with the result that their surveys contain the same basic mistakes that have been made in hundreds of previous surveys. Whilst there is no substitute for experience, there are many pitfalls which have been reasonably well documented in the literature. A cursory review of this literature will help avoid some of the common faults in survey design. There is a problem in
transport surveys in that the literature is widely dispersed. Often the survey method is only briefly described as an adjunct to the main topic of the paper which is usually the analysis of the survey data to assist in policy formulation. Nonetheless, there is a growing body of literature which will assist in suggesting survey techniques which are most appropriate for particular problems. For example, the 1981 Sydney Travel Survey adopted the use of a verbal activity recall framework on the basis of work done by Jones and colleagues in the late 70s (reported in Jones et al., 1983) and in response to dissatisfaction with other trip recall techniques used in the 1971 SATS survey.

Table 2.1 Sources of Existing Transport Data

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>VALUE</th>
<th>LIMITATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Bureau of Statistics: Population Census</td>
<td>Provide a broad outline of population groups and characteristics</td>
<td>Limited transport-related items. Individual records not available.</td>
</tr>
<tr>
<td>Urban Transport Studies; Home Interview Surveys</td>
<td>Detailed person and household travel data</td>
<td>Little data on non-travel activities nor on likely attitudes to future transport options.</td>
</tr>
<tr>
<td>Family Expenditure Surveys</td>
<td>Household expenditure behaviour.</td>
<td>Small nationwide sample; not yet suitable for longitudinal studies.</td>
</tr>
<tr>
<td>Community Services; Patient and attendee records at hospitals, welfare centres, schools, play groups.</td>
<td>Identifies &quot;disadvantaged&quot; groups; design and operation of small paratransit services.</td>
<td>Does not cover &quot;disadvantaged&quot; persons who are not on the records. Some records not easily accessed.</td>
</tr>
<tr>
<td>Employers and Unions; Employee and Membership data</td>
<td>Basis for identifying potential users of employer-based transport schemes.</td>
<td>Possible problems with privacy and availability.</td>
</tr>
<tr>
<td>Operator Data; Transit and Taxi Systems.</td>
<td>Identifying and monitoring users of these services</td>
<td>May not reflect characteristics and attitudes of non-users.</td>
</tr>
<tr>
<td>Political &quot;Network&quot;; Petitions by individuals or groups to government.</td>
<td>Enables local &quot;grass roots&quot; needs to surface.</td>
<td>Danger of demands snowballing or wrong solution being fitted to an expressed need.</td>
</tr>
<tr>
<td>Monitoring Schemes; Tracking Patronage and Productivity.</td>
<td>Can aid in &quot;fine-tuning&quot; a particular scheme and provides research data for other schemes.</td>
<td>Requires commitment to monitor over adequate time period, and a statement of success criterion.</td>
</tr>
<tr>
<td>Universities; Institutes; Special Purpose Studies.</td>
<td>Exploratory &quot;research value&quot; insights into population attitudes and activity/travel behaviour.</td>
<td>Location specific.</td>
</tr>
<tr>
<td>Government Departments and Authorities (e.g. Electric Utility Companies, Motor Vehicle Registrations)</td>
<td>Survey sample frames</td>
<td>Access to files</td>
</tr>
</tbody>
</table>

Source: Bowyer (1980)

In addition to identifying survey techniques and practical pitfalls, a preliminary literature review should also be conducted with respect to the subject of the survey. This may enable the survey objectives to be specified more precisely, in the light of both previous survey results and also of theoretical considerations with respect to the subject matter.
Chapter 2

2.4 FORMATION OF HYPOTHESES

If it appears that new data must be collected in order to satisfy the survey objectives, then it is necessary that the correct type of data be collected. Valuable guidance can be obtained in this respect by indulging in a little "theorising" or "kite-flying" as to the possible hypotheses which one may wish to test once the data have been collected. Such theorising is far from being an academic exercise; it is in fact fundamental to the methods of scientific enquiry on which much of transport systems analysis is based. As noted by Stopher and Meyburg (1979), the scientific approach consists of four major phases: hypothesis formation, observation, testing and refinement of hypotheses. Survey methods are the basis of the observation phase, but this must be preceded by hypothesis formation.

The rationale for hypothesis formation is relatively simple; without it, we have no idea of what we want to measure and to what level of accuracy such measurement is required. However, given an hypothesis about how the component of the transport system under investigation works, it is possible to clarify several aspects of the survey process. First, the dependent and independent variables in the hypothesised relationship are established. For example, the dependent variable may be trips per person per day while the independent variables might be personal income, age, sex and car ownership. Second, the types of analysis techniques which may be appropriate to derive the hypothesised relationship is determined. Third, the unit of analysis may be determined by the theoretical construct, e.g. is choice of a mode of travel a function of an individual or of the household to which the individual belongs? Fourth, the requirements of the specified analysis techniques will give an indication of the degree of accuracy needed in the survey data. Bates (1979) provides a good example of this stage of the survey process with respect to the construction of disaggregate models of transport choice.

It should be noted that not all hypotheses which are proposed before a survey is conducted turn out to be correct hypotheses. Therefore, it is prudent to establish a range of plausible hypotheses and then determine the survey requirements such that each of the hypotheses can be adequately tested.

2.5 DEFINITION OF TERMS

At the end of the hypothesis formation stage, the survey designer is usually left with a number of concepts and terms whose general meanings are understood, but which are not totally clear with respect to details. It is therefore important, at this early stage of survey design, to specify clearly the definition of terms to be used in the survey so that everyone concerned with the design and conduct of the survey is clearly aware of the interpretations to be used. The problem of communicating these definitions to survey respondents in a clear and
unambiguous fashion is a very different issue which will be covered later when dealing with survey instrument design.

In transport surveys, some common terms which need specific definition include:

(a) Household
In the past, for many of the cases encountered in transport surveys, the definition of a household was relatively straightforward where the household consisted of the standard nuclear family, i.e. husband, wife and dependent children. However, current estimates show that such a nuclear family is now in the minority and is further decreasing. Single parent families, extended families, and non-related people living in a family situation make the definition of a household more difficult. There do exist some standard definitions of households (related to how people in a dwelling utilise such items as the cooking facilities) although for the purpose of some transport surveys these definitions may not be appropriate. For example, if the survey objectives pertain to the way in which household members interact in choosing to take part in different types of activities outside the home and the methods of transport they use to reach them, non-related household members will generally behave very differently than those who are related.

(b) Trip
Probably the most difficult concept to describe in transport surveys is that of a "trip". There are many possible definitions, and different definitions may be appropriate under different conditions. Axhausen (1994) describes many of these in some detail. The definition of a trip may be related to a change in land-use activity at the origin and destination, it may be influenced by the mode used for the trip, by the length of the trip, and by the length of participation in activities at each end of the trip. Most trips are defined in terms of one-way movements, but some reported trip rates are given in terms of round-trips (sometimes called journeys or sojourns).

The most common differentiation between trips, however, is between unlinked trips (sometimes called stages or legs or stops) and linked trips. Unlinked trips are usually defined as all movements on a public street, meaning that walking to the bus stop, travelling on the bus, and walking from the bus to the destination would be three separate unlinked trips or stages. This type of definition is important in many cases; for example, when researching exposure to the risk of accident since walk trips are equally vulnerable to accidents as are other types of travel. In general, walking is being seen as a more important component of travel than in the past, so that this definition is becoming increasingly used. Linked
trips are usually defined based on activities. Each time the activity changes, a new trip is said to occur. In the above example, there would be only one linked trip to the destination.

Since it is clear that the definition of a trip is vital to the survey results, it is critical that the survey designer(s) agree on a definition at this stage so that questions can be designed with a single definition of "trip" in mind.

2.6 DETERMINATION OF SURVEY RESOURCES

In most cases, the available resources dictate the design of a survey, rather than vice versa. Whilst it may be theoretically desirable to first design the survey to achieve pre-specified aims, and then to work out the resources required, it is the case in most real-life surveys that the survey must be designed to fit within certain pre-specified resource constraints. The resources needed for the conduct of any survey can be defined in terms of three items: money, time and manpower.

2.6.1 Costs of Surveys

The cost of a survey can be divided into five major areas of expenditure:

(a) Salaries
   (i) Professional staff
   (ii) Administrative staff
   (iii) Computer data entry staff
   (iv) Field staff
   (v) Specialised consultants

(b) Travel Costs and Living Expenses for Field Staff

(c) Services
   (i) Printing of questionnaires
   (ii) Vehicle operating costs
   (iii) Data entry costs
   (iv) Publication costs
   (v) Postage costs
   (vi) Telephone and fax costs

(d) Equipment and Supplies
   (i) Computers

(e) Other Costs
   (i) Overheads
   (ii) Publicity
   (iii) Transport of material
   (iv) Rental of office space
   (v) Incentives and give-aways
Preliminary Planning

Because of inevitable cost over-runs, a good rule-of-thumb found in many surveys is to ensure that there is about 150% of the estimated cost actually available for conduct and analysis of the survey, if necessary. As will be discussed in Chapter 6, the pilot survey is a good time to check the estimated costs of the main survey.

2.6.2 Time Estimates for Surveys

The time needed to complete the survey should be estimated by ensuring that adequate time is available for each of the stages of the survey described in Chapter 1. A bar chart, or critical path diagram, showing overlapping and sequential activities is a useful aid in estimating the survey time requirements. However, so many things can go wrong in a survey that it is virtually impossible to state fixed rules for making time estimates under a wide range of circumstances. Therefore, it is again useful to ensure that there is at least 150% of the estimated time available, since several stages of the survey process can consume far more time than originally envisaged, e.g. editing of data.

The two components which are most often overlooked in terms of time allowance are the pilot surveys and time for analysis of the data. Both are critical to the success of the survey, and more importantly, to the quality and usefulness of the data being collected. Their omission can be to the detriment of an otherwise potentially robust data set.

2.6.3 Personnel Requirements

In estimating personnel requirements, five different categories of staff may be required: professional (for design and analysis), administrative, computer support, field staff, and specialist consultants. Depending on the size of the survey and the way in which it is to be conducted, some of the categories defined above may either be overlapped or eliminated. Thus, for small surveys, the professional and administrative tasks may be handled by one person (or group) while the field and editing tasks may be performed by the one group of people. Alternatively, for large surveys it has been common practice for authorities to contract the entire survey out to survey consultants, and to maintain liaison through a small professional or administrative staff group. Irrespective of the manner in which the survey is to be conducted, it is essential that adequate staff be available at all five levels; otherwise delays and/or inefficiencies will inevitably occur. It should also be remembered that not all tasks can necessarily be performed by the same person. The designer of a questionnaire may not necessarily be the best type of person to administer the questionnaire in the field; similarly, a person who likes the uncertainties of the field survey situation may not be suited for the repetitious task of entering data into the computer.
2.6.4 Space Requirements

One of the components of survey resources which is often overlooked is the requirement of space for all types of surveys. The storage of questionnaires, the seating of administrative staff and the location of numerous data enterers (with their computers) can take up considerable office and storage space and needs to be considered at an early stage of planning. For example, a self-completion travel survey of the kind described in Section 7.2 with a gross sample of 20,000 households needs storage for about 10 tonnes of paper - that is about 40 cubic metres, or a double garage (floor to ceiling) worth of space!

2.7 SURVEY CONTENT

Determining the content of a survey is an important yet sometimes difficult task. The degree of difficulty is increased greatly if the objectives of the survey have not yet been clearly defined. In such a case, where it is not known how the data is to be analysed, the determination of survey content is often left to the intuition of the survey designer; many topics are included in the survey content "just in case" they may turn out to be useful at a later date. Large-scale household travel surveys have often suffered from this "shot-gun" approach to survey content determination.

On the other hand, if the survey objectives are clearly defined, then the task of determining survey content becomes relatively straightforward; if an item of information is required for the intended analysis, then it should be included in the survey content. Even proceeding in this relatively systematic fashion, however, it is possible that the survey content list can be quite large. The task of refining the list to arrive at those questions actually asked on the questionnaire will be discussed later in the chapter on survey instrument design, when the trade-offs between desired information and questionnaire length are considered. At the moment, then, the survey content list is like a "wish-list" of desired information which one would like in order to achieve the objectives of the survey and the overall study.

It is important to realise at this early stage that the survey content list should include items which may not be directly related to the survey topic under investigation, but which will be necessary when weighting the data to represent the population from which the sample is drawn. Many surveys have floundered at the data weighting stage because the variables in the secondary data source, to which the weighting or expansion factors are to be keyed, have not been included in the questionnaire in an appropriate form.
3. Selection of Survey Method

The task of selecting the appropriate survey method is crucial to the efficiency of the overall survey effort. The choice of survey method will usually be the result of a compromise between the objectives of the survey and the resources available for the survey. This compromise, or trade-off, can be neatly illustrated as shown in Figure 3.1.

A trade-off occurs because it is impossible to control all three of the major elements in Figure 3.1; at best, only two of the three can be controlled by the survey designer. Thus given a fixed budget, as is normally the case, the selection of the survey method, with an associated degree of quality control, will automatically control the quantity of data which can be collected. Alternatively, within a fixed budget, specification of the quantity of data to be collected will immediately dictate the quality of data which can be collected. That is, we can collect lots of low quality data or we can collect a limited amount of higher quality data for a given budget. Generally, the latter course of action is to be preferred.

In determining the total quantity of data to be collected, a further trade-off is present between the sample size from which data is collected and the amount of data collected from each respondent in the sample (i.e. the number of questions in the questionnaire or interview). Within a limited budget for coding, editing and analysis, it will be necessary to trade-off the number of questions against the sample size; which one takes precedent will depend on the purposes of the
survey and the length of the survey content list. Thus some surveys can be effective with only a limited number of highly directed questions, while others need to explore topics in depth. The extent of this trade-off is, therefore, a specific design decision on the part of the survey designer. The trade-off will also be partly determined by the respondents themselves. As the length of the survey increases, the response rate will generally decrease (the rate of decrease will depend on such factors as the interest level of the survey topic to the respondents, and the overall quality of the survey instrument design). There will, therefore, be a point at which an increase in the number of questions asked will result in the collection of less data in total, because of the more than proportional decrease in the response rate. The survey designer should therefore be cognisant of this interaction when making the explicit trade-off between sample size and survey length.

![Figure 3.1 Trade-offs in Selection of the Survey Method](image)

In making the above trade-offs, it is important to remember that the survey resources are defined in three dimensions: time, money and personnel. The above trade-offs should be considered separately with respect to each of the three resources. Thus, a survey method and sample size combination which is feasible with respect to the money available may not be feasible with respect to the time which is available for the survey, or vice versa.

### 3.1 TIME FRAME FOR SURVEY

One basic decision which must be made for all surveys is the choice of time frame for the data collection effort. It is possible to collect data from a **cross-sectional survey** or a **time-series survey**. A cross-sectional survey involves the collection of data at one point in time over a large number of identifiable groups which constitute a fraction of the total population. On the other hand, a time-series
survey involves the collection of data using the same survey method, usually from a smaller number of groups in the population, at a number of successive points in time (a before-and-after survey is a limited form of time-series survey).

Some authors (e.g. Hensher, 1985) make the useful distinction between time-series and longitudinal data, whereby time-series data commonly defines time itself as the independent variable in the data set, whereas longitudinal data defines the independent variables to be those factors which are causally related to the dependent variables. Both types of data are collected over several periods of time, but the emphasis in the analysis is different. Time-series data is typified by quarterly data bases used by economists studying macro-economic relationships, or annual traffic counts used by traffic engineers to estimate traffic growth rates. The major emphasis in transport planning data, however, is on longitudinal data which can assist in increasing understanding of the underlying causes of travel behaviour.

Possibly the best known, and certainly longest running, longitudinal survey in transport is The Dutch National Mobility Panel survey which has been carried out by the Netherlands Ministry of Transport and Public Works since 1984 (van Wissen and Meurs, 1989). It was designed as a general-purpose travel survey, with the primary objective to analyse public transport policies. A large volume of research in many aspects of travel has in fact been drawn from the panel data. A more recent general purpose longitudinal study was begun in the Puget Sound Region of the United States in 1989 (Murakami and Watterson, 1992). Other examples of specific studies are a car ownership study in Sydney, Australia (Hensher, 1986), a survey of car and bus use in the U.K. (Goodwin, 1989) and of telecommuting in California (Goulias et al., 1992).

A further distinction is often made between longitudinal surveys and longitudinal data. Longitudinal data need not be generated from time-series surveys only; it is possible to use retrospective reports and prospectively focussed intentions surveys to obtain quasi-longitudinal data at a cost which is much closer to that of a conventional cross-sectional survey. However, given the problems with the recall ability of respondents and the propensity of respondents to "tidy up the past" to be more in accord with known outcomes (the cognitive dissonance effect, Festinger, 1957), it is likely that the use of retrospective reports creates deliberate or unconscious distortion of information supplied by the respondent (Wall and Williams, 1970). While retrospective reports may have a function in establishing respondents perceptions of prior conditions, they should not be used as surrogates for objective data which could have been, but were not, collected at those prior times. The use of surveys which ask people to say what they will do in the future (often under changed conditions) are discussed in Chapter 5 under stated preference surveys. The use of these surveys without
using retrospective data, however, is not generally considered to provide longitudinal data.

In many cases, it is implicitly assumed that longitudinal data and cross-sectional data will yield the same behavioural relationships. For example, if one wanted to predict future household trip generation rates at a time when car ownership is at a generally higher level than at present, a standard method would be to conduct a cross-sectional survey and, from this data, derive trip generation rates for households with different levels of car ownership. It would then be assumed that the relationship between car ownership and trip generation, which was obtained across different households in the population at one point in time, will be valid when used for one household over a period of time. Thus, over time, as a household moves to a higher car ownership level, it would be assumed that they will take on the trip generation characteristics of those households who currently have that higher car ownership level. While in many cases this assumption may be a reasonable approximation, it should be realised that it is an unproven assumption. In some cases, the assumption may well be grossly incorrect because the dynamic changes which occur as a result of a change in one variable may not be considered.

If survey data is to be used to develop predictive models of travel and activity behaviour, it therefore appears that longitudinal data, which can account for dynamic changes in behaviour, is the preferred type of data to collect (see Hensher (1982, 1985) for further arguments to this effect).

### 3.1.1 Longitudinal Surveys

Within longitudinal surveys, there are two distinct variations: successive samples and panel studies.

- **The successive samples** survey involves the selection of a separate sample from the same population at each stage in the time-series. Thus, while each sample represents the same population or market segment within a total population, the individuals within each sample are different from survey to survey. This is often called a repeated cross-section survey.

- A **panel study**, on the other hand, selects a sample at the start of the time-series and retains the same individuals within the sample for the entire series of surveys.

Longitudinal surveys exhibit a number of advantages and disadvantages compared to cross-sectional surveys. Each type of longitudinal survey also has relative advantages compared to each other. A discussion of some of the advantages and disadvantages of longitudinal panel studies will therefore
Selection of Survey Method

highlight some of the features of each longitudinal survey method, and of longitudinal surveys in general.

The advantages of a longitudinal panel study include:

(a) Whereas the successive sample survey provides information only on the net change in the characteristics of the sample, the panel study survey also provides information on the gross changes. Table 3.1 illustrates these features. For example, a successive samples survey may indicate that usage of a bus service within the population of interest has decreased from 40% to 35%. A panel study will provide the extra information that, as well as there being a net decrease of 5%, this change is the result of 25% of the total population ceasing to use the bus while a different 20% of the sample started to use the bus. This extra information is useful to the operator of the bus service in two ways. First, while there has been an overall trend of declining patronage, this trend is not entirely one-way. It would appear that there is scope for increased usage of the bus within a certain segment of the population; the operator should identify this group, establish the reasons for their increased usage of the bus, and then seek out ways in which this potential can be maximised. Second, while overall there has only been a modest decline in patronage, there has been a dramatic loss of patronage from the existing users (62.5% [250/400] of the original users stopped using the bus between the two surveys). Once again, the operator should identify those who discontinued using the bus, identify the reasons for their decisions, and adopt strategies to arrest the loss of existing users.

Table 3.1 Comparison of Successive Sample Surveys and Panel Studies

<table>
<thead>
<tr>
<th></th>
<th>1993</th>
<th>1994</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Successive Survey</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bus use</td>
<td>40%</td>
<td>35%</td>
</tr>
<tr>
<td><strong>Panel Survey</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• 1993 only</td>
<td>250</td>
<td>-</td>
</tr>
<tr>
<td>• 1993+1994</td>
<td>150</td>
<td>150</td>
</tr>
<tr>
<td>• 1994 only</td>
<td>-</td>
<td>200</td>
</tr>
<tr>
<td>i.e. Total users</td>
<td>400</td>
<td>350</td>
</tr>
<tr>
<td>i.e. Bus use</td>
<td>40%</td>
<td>35%</td>
</tr>
</tbody>
</table>

*Assume total population surveyed = 1000
(b) By tracing the behaviour of a group of individuals over a long period of time, one obtains an idea of the dynamics of change. That is, not only is it possible to observe static equilibrium states at different points in time, but it is also possible to trace the way in which individual members of the sample arrived at the equilibrium states (if, indeed, they are equilibrium states). In this way we can identify the conditions which triggered a response and the way in which the panel study member adapted over time to changes in conditions or to the responses that they actually made. For example, what pressures or opportunities caused a household to buy a second car, how did they adapt to the pressures before they bought the second car, and how has their behaviour and travel patterns changed after purchase of the car. Tracing panel study members in this way enables a much richer and more valid understanding of their responses to change. Trying to understand such dynamic behaviour by means of a cross-sectional survey, is like Newton trying to estimate the gravitational constant by looking at a photograph of an apple falling from a tree (if he had a camera available in those days!).

(c) By adding a few new questions to each round of a panel study survey, it is possible to accumulate more information about each respondent than would be possible in either a cross-sectional or a successive samples survey. A good discussion of updating a panel survey questionnaire can be found in Goulias et al. (1992). These new questions should pertain only to those respondent characteristics which are expected to remain relatively stable over time, such that it does not matter at which interview they are asked. The advantage of this procedure is that more information can be obtained without drastically increasing the size of any one interview.

(d) By interviewing the same people on a number of occasions, a greater degree of rapport and cooperation can be built up between the interviewer and the respondent. This may allow questions (of a slightly sensitive nature) to be asked on later occasions where they could not successfully be asked in an initial interview.

(e) A great advantage of panel studies is that the between-sample variance is eliminated from the sample design. Thus, whereas for successive samples surveys it is assumed that the samples at each stage of the time-series are (statistically) the same, for panel studies it is known that they are the same. This means that, all other things being equal, smaller samples can be chosen for panel studies than for successive samples surveys.
Selection of Survey Method

While panel studies do possess some significant theoretical advantages, as noted above, they also are fraught with a number of potentially serious survey methodology problems as follows:

(a) It is often difficult to find respondents who are willing and able to be interviewed on repeated occasions. More importantly, given that such people can be found, it is highly probable that, unless special care is taken in sample selection, they will represent a particular segment of the population. Significantly, for transport surveys, these people may well be less mobile in terms of both locational and travel patterns, i.e. they stay in one residence for a long time, and they are more often at home with nothing to do (except answer interview questions!). The selection of people in the initial sample may therefore be biased toward people who will agree to be interviewed on repeated occasions. Such self-selection poses special problems in sample design; although the socio-demographic composition of the sample may match that observed in secondary sources, there may well be an internal invisible substitution of willing panel members for unwilling panel members. Many of the issues of panel survey sampling design are discussed in van de Pol (1987).

(b) Even if an unbiased sample can be initially chosen for the panel study, then sample mortality is a major problem as the panel study progresses. People die, move out of an area, change jobs or socio-economic characteristics, or move to another stage in the life cycle. While such changes do not automatically rule panel members out of continued participation in a panel study, these effects (and others) can tend to rapidly diminish the number of people who remain eligible for membership of the panel study population. Not only is sample mortality a problem in terms of reduced sample size, but more importantly it can result in a biased sample remaining in the study if, for example, it is the more mobile members of the population who leave the panel. This attrition bias can cause serious problems in the interpretation of trends observed over the life of the panel. While it is extremely dubious to simply replace panel members who drop out of the study (Moser and Kalton, 1979), it is possible to adopt a hybrid design (such as a rotating panel design) which combines the features of panel studies and successive samples surveys (Curtin, 1981).

(c) Repeated interviewing can have a conditioning effect on the respondent whereby an initial attitude of indifference towards the subject of the survey can change to one of interest and greater involvement in the subject of the survey. This may, in some circumstances, result in changes in respondent behaviour (which may be measured in later surveys). More likely, however, it will result in changes in respondent attitudes (which
may again be the subject of measurement). If these effects are not allowed for in the initial design, then serious biases can occur in the results. On the other hand, one can use these effects to advantage as an instrument of attitude formation as is done, for example, in the Delphi survey method (Dalkey and Helmer, 1963).

(d) Repeated interviewing can also have the reverse effect whereby respondents tend to lose interest in what might have initially been an interesting subject to them. The same can also happen to interviewers. In such circumstances, both respondent and interviewer may, in later interviews, simply go through a stereotyped routine where both the interviewer and the respondent know the answer which will be given before the question is asked. Additional interviews of this type may result in very little additional information.

(e) In an effort to reduce some of the negative effects of repeated interviewing with the same interviewer, it is possible to change interviewers in each round of the study. However, unless careful experimental design is used in this procedure, the change in interviewer can produce equally undesirable effects (Jones et al., 1986). This is because changes in response attributed to the respondent may really be due to changes in the respondent-interviewer interaction over the different rounds of the survey.

(f) If a longitudinal survey is being used to develop an index for monitoring of a system (e.g. surveys of customer satisfaction, or monitoring trip rates over time) then it is essential that the questions used to obtain information for construction of the index should remain stable over time. This desire for stable questions, however, can pose a problem where it is realised after the early rounds of the longitudinal survey that the questions being asked may not be the most appropriate questions to ask. In this case, one is forced to make a trade-off between the stability and the appropriateness of the questions. This problem occurs, for example, in the selection of questions on transport for inclusion in the national Census of Population. The use of a pilot survey is, therefore, all the more critical in this case since it can largely eliminate the problem if designed carefully.

(g) Where a longitudinal survey is being used to assess changes in system performance as a result of change in the system’s physical characteristics (as, for example, in before-and-after surveys), it is often difficult to ensure that no other changes occur, either in the system or in the external environment, which would have an effect on the system performance measures. The longer the period that the longitudinal survey covers, the more difficult will be this problem.
Selection of Survey Method

(h) While panel surveys have the advantage that they eliminate between-sample variance, they have the problem that one must take extra precautions to ensure that the initial sample is indeed representative of the population under consideration, since one must stay with this sample for the entire series of surveys (van de Pol, 1987).

(i) As with all longitudinal surveys, panel studies have the considerable disadvantage that, because of their very nature, one must be prepared to wait for some time before the results will become available. Obviously one can allay this problem to some extent by releasing results from each wave of the panel study as if they were simple cross-sectional surveys. Because of the delay in obtaining the final results, however, it is sometimes difficult to convince prospective clients of the need for a panel study. This is particularly the case where the client is involved in the political arena, and where the length of the panel study may be longer than the term for which the politician is elected. Because of the difficulty in securing funding, it is often necessary to make a panel study part of an omnibus survey, which serves the different objectives of several different clients, in order to obtain sufficient funding to make the survey viable. In situations where results are needed relatively quickly (as is most often the case), longitudinal surveys are relatively impractical unless suitable data happen to have been collected over several time periods in the past. The corollary of this is that for longitudinal surveys, one needs a considerable degree of foresight to predict what the issues requiring data will be when one completes the surveys at some future date. One then only needs to convince clients that they should start funding the collection of that longitudinal data now!

In general then, it can be seen that while longitudinal data is often the most theoretically appropriate type of data, it is also beset by many practical (and some theoretical) difficulties which militate against its use. There is, however, a body of literature which will help the prospective user of this methodology, some part of which has already been mentioned. Choosing the sample and dealing with the idiosyncrasies of a changing population has been addressed by van de Pol (1987). Jones et al. (1986) have set out a series of guidelines on motivating interviewers and respondents in longitudinal research. If personal interviews are to be used, Morton-Williams (1993) has many useful comments. For a public transport users panel, Golob and Golob (1989) have some practical advice. And finally, Meurs et al. (1989) have some interesting comments on that vexing question of measurement error in panel surveys.

Despite the recent interest in the longitudinal method, cross-sectional surveys remain the most common form of survey used in transport research.
Chapter 3

3.2 TYPES OF DATA COLLECTION TECHNIQUE

Having decided on the time frame for the survey, the next decision is to decide on the type of data collection technique to be employed either once in a cross-sectional survey or repeatedly in a longitudinal survey. It should be noted that if a longitudinal survey is to be undertaken, then it is imperative that the same data collection technique be employed in each wave of the longitudinal survey. As will be shown later, each data collection technique has its own peculiar biases. By using different techniques in each wave of a longitudinal survey, it may be that any differences in results which are observed from wave to wave are due to the different data collection techniques used, rather than to any real differences in the subject of the survey.

Essentially, there are eight different data collection techniques which may be employed:

(a) Documentary searches
(b) Observational surveys
(c) Household self-completion surveys
(d) Telephone surveys
(e) Intercept surveys
(f) Household personal interview surveys
(g) Group surveys
(h) In-depth surveys

These survey methods vary in complexity, in the types of information which can feasibly be collected, and in the level of interaction between the survey designer and the respondents in the survey.

In documentary searches, the subjects of the survey are inanimate objects (documents) and there is no response required of these objects. In observational surveys, the subjects of the survey may be either inanimate (e.g. roadside features in an inventory survey) or animate (e.g. pedestrians in a pedestrian flow survey), but no specific response is required from these objects; they are merely expected to behave in their normal manner while they are being observed. As noted in Chapter 1, these two types of survey techniques are not the main focus of this book, although they are discussed briefly in the following pages to put them in context. In self-completion surveys, the subjects of the survey are definitely animate, and they are required to respond and participate in the survey. However, the contact between the respondent and the surveyor is second-hand, and is made only via a written questionnaire. With a telephone survey, while there is contact with an interviewer, it is not face-to-face. Intercept surveys are those surveys conducted outside the home and while the respondent is in the process of using a mode of transport (e.g. on a train) or of participating in activity (e.g. shopping).
Selection of Survey Method

They can be either personal interview or self-completion. Even with self-completion methods, however, the questionnaires are usually distributed by survey staff, meaning that there is some contact with the research group. With the personal interview survey, contact is more direct and involves face-to-face contact between the respondent and the interviewer. Under these circumstances, there is a far greater possibility for interaction (both beneficial and harmful to the purposes of the survey) between respondent and interviewer. While this interaction may allow more complex data to be collected, it also allows for a greater degree of bias to enter into the survey results. This interaction is compounded in the group survey which takes place with a group of people (often outside the home). In the in-depth survey, however, interaction is not just a by-product of the survey, but rather it is a design feature. In most cases, it is expected that there will be considerable interaction between the members of the group being surveyed, and that this interaction will yield much richer data than could be collected in one-on-one interviews with individual members of the household or other group.

3.2.1 Documentary Searches

A documentary search is simply a search of existing published and unpublished documents and databases in an attempt to uncover the type of information which is required in the survey. As noted in Chapter 2.3, a documentary search may either obviate the need for further data collection or else it may simply assist in the design of further data collection efforts. The use of documentary searches to generate items of information for secondary analysis is best illustrated by a number of documents which are regularly published containing collations of transport statistics. In the U.S. context, such documents would include the National Transportation Statistics Annual Report (USDOT, 1986), the Transportation Energy Data Book (Holcomb and Kosky, 1984), the National Urban Mass Transportation Statistics (Section 15 Data) (USDOT, 1982), the Transit Fact Book (APTA, 1985), and the Aircraft Operating Cost and Performance Report (USDOT, 1985). In Australia, Transport Indicators (BTCE, 1994) and in the UK (UK Department of Transport, 1993a; Central Statistical Office, 1994) are further examples.

These documents, which are only a sample of the wide range of documents covering various modes and geographic locations, provide much basic data which can at least provide order-of-magnitude estimates of many important parameters. In assembling data from such documentary searches, however, care must be taken to ensure that the conditions surrounding the initial collection of the data are known and allowed for. Thus, definitions and initial data collection procedures should be assembled as well as the results obtained from the initial surveys. In addition, it is often useful in interpretation of the assembled data, if the background of the originating organisation and the purpose of the initial data
collection are also known. In this way, any potential biases in the data can more readily be recognised. Wigan (1985) and Thoresen (1983) provide some interesting comments on experiences with documentary searches in an Australian context, which should be of general relevance in other geographic circumstances.

Documentary searches may also be useful when attempting to obtain transport system inventory data. Thus, timetables and route maps for public transport systems may provide useful information for the construction of a public transport network model. Similarly, street maps of various scales will often provide enough information for the construction of a network model for a road system. Aerial photographs may be useful in establishing a sampling frame of houses in a study area for use in a household interview survey (although it should be noted that the aerial photographs will not be able to determine whether multiple households exist within a single residential building).

In some cases, the documents themselves become the unit of analysis in a survey. For example, in a study of transport legislation, the Acts, Bills and Regulations become the units of the survey population. Similarly, in a study of editorial opinion on transport matters (Olsson and Kemp, 1976), the unit of analysis was a documentary source; the newspapers containing the editorials. In other cases, the documents become the subject of the survey (rather than the units of analysis) in, for example, the study by Cohen et al. (1981) where they examined which transport journals were most frequently read by the transport profession (at least the United States transport profession).

### 3.2.2 Observational Surveys

Observational surveys, while being relatively infrequent in social science surveys, are commonplace in transport and, more particularly, traffic surveys. Two basic types of observational survey are the direct and the indirect observational survey. Examples of direct observational surveys include:

(a) Transport inventory surveys (e.g. using techniques such as video recording (Fahey and Chuck, 1992), digital imaging (Dickinson and Waterfall, 1984; Wigan and Cullinan, 1984), and instrumented vehicles (e.g. Bonsall and Parry, 1991));

(b) Traffic count surveys of different types (e.g. link counts, intersection counts, cordon counts, screen-line counts, transit route counts, boarding and alighting counts, etc.);

(c) System performance surveys (such as travel time surveys, intersection delay surveys (Richardson and Graham, 1982), and public transport performance surveys (Attanucci et al., 1981));
Selection of Survey Method

(d) Traveller tracing, either to check accuracy of reported trip characteristics (e.g. Hensher and McLeod, 1977), to obtain direct information on route choice characteristics (Wright, 1980), or to observe the speed and acceleration characteristics of vehicles in traffic systems (Akcelik, 1983);

(e) Vehicle classification surveys, in which vehicle types are identified by various means such as manual visual recognition, profiles from inductance loops, and pattern recognition from single tube axle counters (Bayley, 1983). In addition to identifying vehicle types, a greater emphasis is being placed on obtaining the weights of heavy vehicles by means of a wide variety of weighing-in-motion (WIM) devices.

A critical feature of survey types (b) through (e) is that it should be recognised that the system performance measures obtained on any one day are but single points from the distributions of the system performance measures. For example, travel time will vary from day to day for any given trip, as will public transport passenger waiting time which is dependent on the coordination of arrivals and departures of different modes of transport. In designing surveys of system performance, allowance must be made for the variability inherent in each of the parameters under observation.

Examples of indirect observational surveys include:

(a) Wear patterns (caused by vehicles or pedestrians) which may indicate predominant traffic flows;

(b) Accident debris or skid marks to indicate hazardous sites in a road network; and

(c) Fuel sales, and other economic indicators, to estimate total activity in various transport sectors.

In general, indirect observational techniques are used less than direct observational techniques in transport surveys. This contrasts with the social science survey field where a large group of indirect observational survey techniques has been developed under the general title of "unobtrusive" survey techniques (Webb et al., 1966).

One of the major uses of observational surveys in transport planning has been to obtain data which has been used to check on the validity of results obtained from personal interview or self-completion surveys. Thus screen-line counts have been used to check on origin-destination predictions from a household survey. One should, however, realise certain factors when making such a comparison. First, the observational survey should account for variability in the system performance measures, as mentioned earlier. Second, it must be realised that
many household surveys are conducted over a reasonably long period of time (e.g. six months). In such circumstances, it is somewhat difficult to decide just which traffic counts should be used for comparison. Third, it should be ensured that the same definitions are used in each type of survey (e.g. personal interview surveys would generally predict person-trips, whereas cordon counts would observe vehicle-trips). Finally, biases in either or both of the survey methods may need to be accounted for before the results of the surveys are compared (e.g. the under-reporting of particular types of trips in personal interview surveys).

3.2.3 Household Self-Completion Surveys

Self-completion questionnaire surveys are one of the most widely used forms of survey technique in transport. Self-completion surveys are defined to be those which the respondent completes without the assistance of an interviewer. With this type of survey, respondents are required to perform the three tasks on their own. These three tasks are to read and understand the question, to mentally formulate an answer, and to transcribe this answer onto the questionnaire form itself. In personal interview surveys, the respondent is wholly responsible only for the second of these three tasks; the interviewer is primarily responsible for the first and third.

Several types of basic survey format can be described, depending on the methods used for collection and distribution of the questionnaire forms. These variations include:

(a) Mail-out/mail-back surveys

This is the most basic form of the self-completion survey and the one that is most often employed. The questionnaire is mailed to the respondents and they are asked to return it by mail after they have answered the questions. The return postage is generally paid by the survey designer, although the precise method by which this postage is paid can have an effect on response rate as will be described in Chapter 7.

(b) Delivered to respondent/mailed-back

Where it is suspected that response rates may be particularly low, or where it is believed that respondents may need some personal instructions on how to complete the survey form, it may be desirable to personally deliver the questionnaire form to the respondent. In this way the purpose of the survey can be explained, any questions about the survey can be answered immediately, instructions can be given, and questions answered, about how to fill out the survey form. Such personal contact will generally increase the response and will also result in a higher quality of response to the questions.
Selection of Survey Method

(c) Delivered to respondent/collection from respondent

In addition to delivering the questionnaire form, it is also possible to collect the form from the respondent at some later date. This further increases the response rate by putting some pressure on the respondent to complete the survey before the collector returns, and also enables the collector to resolve any specific problems which the respondent encountered while filling out the questionnaire form. Naturally, however, the increased response obtained in methods (b) and (c) can only be obtained at considerable extra expense for the personal delivery and collection of the questionnaire forms. However, where a high response rate is essential, as in a National Census, then this method may be the most cost-effective way of obtaining these responses. This method is frequently used when "long-term" diaries (e.g. 7 day diaries) are distributed (e.g. in the National Travel Survey in the U.K. (UK Department of Transport, 1993b).

Self-completion surveys have both advantages and disadvantages. The primary advantages of a self-completion survey include:

(a) Self-completion surveys are generally much less expensive than a comparable personal interview survey. Hitlin et al. (1987) estimate that the cost of a telephone interview survey was approximately three and one half times more expensive per completed response than a hand delivered and collected self-completion survey. A personal interview survey is even more expensive than a telephone interview survey (Ampt, 1989). Care needs to be taken, however, that this cost reduction is not a false economy. Self-completion surveys almost always have lower response rates than personal interview surveys, and hence the cost per returned questionnaire is much higher than the cost per distributed questionnaire. More important, the value of a returned questionnaire is a function of the overall response rate because of the effect of non-response bias. Thus, the value-for-money from self-completion surveys may be lower than it initially appears unless care is taken to minimise, or otherwise account for, non-response effects.

(b) A wide geographic coverage is possible in the sample, because postal charges do not generally vary as a function of the distance involved. It is usually just as expensive to send a questionnaire 1 kilometre in the mail as it is to send it 1000 kilometres. There is therefore no incentive to limit the sample to a local area (as there would be in a personal interview survey). The only restriction which needs to be considered is the time required for the mail to reach the more remote areas of the study area; this may be a critical factor in the design of a schedule of reminder letters.
Chapter 3

(c) Because there is no interviewer involved in the survey, interviewer effects are eliminated as a possible source of response bias. The type of interviewer effects which can occur will be discussed in Section 3.2.6.

(d) The respondent has ample time to consider the questions before providing an answer, and hence it is possible to obtain considered responses to questions in a self-completion survey. It is also possible for a respondent to consult documents, if necessary, in order to provide factual information as an answer to a question (e.g. consult a log book to provide information about vehicle operating costs).

(e) The respondent can choose the time and place in which to complete the questionnaire (except for questionnaires which must be completed quickly, such as on-board a vehicle). Because of this, there is less incentive for respondents to hurry through the questionnaire in order to resume the activity they were engaged in when they received the questionnaire. This generally results in a more complete answering of the questions. The danger with allowing the respondents to defer answering the questionnaire is that they will completely forget about answering it.

While the above advantages are attractive features of a self-completion survey in some circumstances, the self-completion questionnaire survey is not without some considerable difficulties, such as:

(a) The most consistent criticism of self-completion surveys has been the high level of non-response. Reported response rates of between twenty and fifty percent are not uncommon, and this allows ample opportunity for serious biases (see Chapter 9) to enter the data (Brög and Meyburg, 1981). However, by incorporating rigorous design and validation measures, more recent self-completion travel surveys in Australia have achieved response rates of up to 75% (Richardson and Ampt, 1993a). The most reliable way to address the response rate is to incorporate a series of reminder letters and/or questionnaires into the survey design - a methodology which raises the cost of the survey method to some extent.

(b) The layout and wording of the questionnaire MUST be extremely clear and simple. Definitions must be clear and easily understood by the population being studied (not just to the survey designer), because there is no interviewer on hand to clarify the intent of the questions. The amount of time and effort which must be put into questionnaire design is therefore considerable.

(c) With a self-completion survey, it is difficult to ensure that the correct person fills out the questionnaire form. Even though the questionnaire
may state quite specifically who should fill out the form, there is no guarantee that this occurs and there is no simple way of checking whether this is the case. In travel surveys, it is important that each person relates their own experiences to avoid the problems of under-reporting associated with proxy data (Ampt, 1981). It is therefore vital to incorporate validation measures which give some information on proxy reporting in the self-completion survey design.

(d) Responses from self-completion surveys tend to be skewed towards the more literate sectors of the population which tends to travel in a different way than the remainder of people. This means that rigorous follow-up procedures for non-respondents are required to ensure robust data.

(e) In general, only simple, one stage questions can be asked, since questions which require complex filtering, which depends on answers given previously, usually need the skill of an interviewer.

(f) In self-completion surveys the answers on the questionnaire form must be accepted as final; there is often no chance to probe further to clarify ambiguous or unclear answers. For this reason it is important to plan and test the inclusion of ways of following-up respondents. One way to do this is to ask for their phone number and phone if clarification is needed, and another is to follow-up a sample of people with a personal interview (Section 7.2.2).

(g) Spontaneous answers cannot be obtained in a self-completion survey, so opinions given may not in fact be the respondent's own opinion at that time, but may be the result of discussion with others at a later time. For this reason, self-completion surveys are generally not the best instruments for conducting attitudinal surveys.

(h) A self-completion questionnaire offers no opportunity for the collection of supplementary information on such things as the respondent's residential environment or the respondent's attitude towards the survey.

(i) Answers to questions cannot be treated as independent since the respondent has the opportunity to scan the entire list of questions before answering any of them. For this reason, obvious cross-checks which are transparent to the respondent cannot be used very effectively.

In summary, self-completion questionnaires are characterised by the ease with which they can cover a wide geographical area, and by their moderate cost. They need to have substantial effort invested in their physical design and appearance. Finally, the desirability of using reminder letters or questionnaires means that the
Chapter 3

entire survey period needs to be considerably longer than that for personal interviews.

3.2.4 Telephone Surveys

The telephone survey has been used for many years in market research surveys outside the area of transport (Blankenship, 1977; Frey, 1983, Groves, et al., 1988). In fact, in the United States it has become the most commonly used mode of conducting structured market research interviews (Frankel, 1989). It has been used to a much lesser extent in transport research, but there are a number of transport studies which report its use in one form or another (Stopher and Sheskin, 1982b; Hitlin et al., 1987; Stopher, 1992; Ampt, 1994).

The growth of telephone interviewing in the 1970s and early 1980s led to the setting up of centralised telephone interviewing installations for many surveys, a development which revolutionised telephone interviewing. Dedicated telephone interviewing facilities allow for stricter control and closer supervision of interviewer behaviour than is possible with from-home telephone surveys or with personal interviews (Morton-Williams, 1993).

The telephone survey method has a number of advantages which include:

(a) The telephone survey offers the possibility of wide geographic coverage - particularly in a given urban area where rates for phone calls frequently do not vary with distance. This is in significant contrast to the personal interview survey where cost factors usually mean that samples are clustered. And when compared with self-completion methods, although telephone rates are not as geographically uniform as postage rates, it is almost as easy and inexpensive to contact a remote site in a telephone survey as it is to contact an accessible site.

(b) Because telephone interviews are usually performed from a central location, it is possible to have much better supervision of the interviewers in order to maintain a higher level of quality control on the completed interviews.

(c) By centralising the interview facility, it is possible to use Computer Assisted Telephone Interviews (CATI). In this method, the interviewer reads the questions from the computer screen and then types the responses directly into the computer as they are received over the phone. This allows rapid checking of responses for consistency, and eliminates the task of transcribing the answers onto paper and then later entering these answers into the computer.
Selection of Survey Method

(d) Telephone surveys are generally cheaper than personal interview surveys because of the reduction in labour needed to conduct the survey and the absence of field and travel costs associated with having interviewers in the field. The cost advantage is more noticeable in the United States than in other countries since in this country special telephone lines are available at concessionary rates, and where random sampling for face-to-face interviewing is relatively complex and expensive. In Britain, however, the situation is reversed; long distance telephone calls are relatively expensive and random sampling for telephone surveys is more problematic than for person interview surveys (Morton-Williams, 1993, p. 156).

(e) Use of the telephone is ideal for validating and clarifying queries about responses people have made in self-completion surveys (Richardson and Ampt, 1994). Respondents are phoned after data enterers have identified "suspected" errors in their replies. This method achieves an almost 100% response rate (since people being phoned have already responded once to the survey) and has the advantage of significantly increasing the validity of results, with minimal additional cost.

(f) In conducting surveys in today's multilingual societies, dealing with people who do not speak the language of the interviewer occurs fairly often. Telephone surveys offer an effective method of dealing with these occurrences by having a central core of multilingual interviewers who can be called to the phone as required to complete an interview when language problems arise. This is much more efficient than trying to have a large number of multilingual interviewers who must perform interviews in the field (where their special talents are most often not needed).

(g) Because of the speed and low cost of contacting households, it is sometimes possible to use a telephone survey to identify rare populations (i.e. groups in the population who constitute a very small fraction of that population). By using an appropriate set of filter questions at the beginning of the interview, the interviewer can quickly identify whether the household belongs to the rare population or not.

The telephone survey method, however, has some potentially serious disadvantages, including:

(a) There is a limit on the length of survey which can be successfully completed over the phone. While some individuals will be willing to spend as much as 20 minutes or more being interviewed by telephone, the overall response rate drops rapidly after about 10 to 15 minutes (Stopher, 1985a).
(b) The number of people in a household with whom it is possible to carry out the interview is almost always limited to one. It is rarely possible to encourage more than one person to respond. As will be discussed in Section 7.4, proxy interviewing (one person reporting on behalf of another) is not recommended at all in travel surveys. This, combined with the limited length-tolerance for phone interviews, is a significant reason for not using telephone interviews for household travel surveys where all household members should be interviewed.

(c) With an increasing amount of direct marketing being done by means of the telephone (some of which is disguised as a sample survey), it is becoming more and more difficult for sample survey researchers to establish their credibility at the beginning of an interview. The general public is becoming rightfully wary of anyone who introduces themselves as conducting a survey. It is imperative that the survey designer identify a means whereby the credibility of the surveys can be established before the telephone call is made. This can be by means of an introductory letter (although this can be difficult since the respondent’s address will often not be known in advance if, for example, random digit dialling is used to select the sample) or by means of an introductory telephone call in which an appointment is made to call back at a more convenient time for the respondent. This latter method also has the possibility of reducing the intrusiveness of a telephone interview survey.

(d) Because of the fact that only those households with phones can be included in a telephone survey, there is an obvious potential for sample bias to occur. The extent of this bias will depend on the extent of phone ownership in the population under consideration, and the extent to which those households with phones display differing characteristics to those households without phones or with unlisted numbers. Certainly it has been shown that non-phone owners and people with unlisted phone numbers have different socio-demographic than those with phones (non-phone owners are younger, single person households, less affluent and more likely to be unemployed, low car owning or of ethnic background (Morton-Williams, 1993), and unlisted households are generally the reverse).

More important, however, is the variation among non-phone owners with respect to the variables under consideration (e.g. travel patterns). It has been demonstrated (Richardson, 1986) that such differences can occur, but that they can be corrected for by means of standard socio-demographic weighting techniques (to be described in Chapter 8). Nonetheless, it is essential to recognise the potential for such sample bias
Selection of Survey Method

and to plan the survey design such that these corrections can be made at a later stage.

(e) The use of the telephone interview usually means that telephone books have been used to select the sample. In addition to the problem of non-phone owning households being excluded from the sample, there are other problems such as out-datedness associated with their use. These are elaborated in Section 4.3.

(f) Unlike other forms of survey, there is no chance of follow up for non-respondents in a telephone survey. If a respondent refuses to participate in the survey, then little or no background information on this respondent can be obtained. This makes it extremely difficult to correct for response bias at a later stage.

(g) Because of the nature of a telephone survey, no visual aids can be employed in such a survey. All communication to and from the respondent must be by means of the spoken word, and this results in severe limitations on how questions can be asked and answered. For example, it is not possible to use show cards to overcome the "order bias" which occurs when a list of possible responses is read to the respondent. Some possibilities for overcoming this problem include sending visual aid materials to the respondent before the telephone interview takes place. Like sending an introductory letter, however, this requires that an address is known for each respondent. A novel technique which does away with voice communication is the use of digital touch-phone signals which can be used to enable the respondent to reply to a pre-supplied list of questions.

While telephone surveys are seen by some as an area with significant potential in the collection of transport survey data, we would recommend that they should be used with caution, especially for data which is not factual and straightforward.

Telephone surveys for gathering data on travel and activities can be summed up as follows. They are generally not useful:

• for surveys in which all persons in the household are to be interviewed,

• any collection of travel data where the household has not been contacted by letter or in person.

They are general useful:
Chapter 3

- to follow up queries arising from self-completion or personal interview surveys, and
- for commercial or business-based surveys.

For a good discussion of the advantages and pitfalls of telephone interviewing, see Groves et al., (1988).

3.2.5 Intercept Surveys

Intercept surveys are those surveys which take place at a site which is not in a household - where people are intercepted in the course of carrying out an activity of some type. They include surveys on-board public transport vehicles, at cordon points on roads, and at other activity sites such as shopping centres, work places or transport nodes such as airports. The surveys which are carried out at these places can have more or less interaction between surveyor and respondents, depending on the objectives of the survey and the location of the intercept. All intercept surveys, however, involve personal contact with respondents in one form or another - either to distribute questionnaire forms or to actually ask a series of questions.

The major types of intercept surveys are:

(a) On-board vehicle distribution/mail-back

In many cases, it is desired to conduct a survey of a particular group of transport system users, e.g. public transport patrons. To attempt to find these people by means of a general household survey would be almost impossible, because they represent such a small percentage of the total population. A more efficient method is to limit the population to include only those people, and to use a survey method which will identify only members of that population. On-board vehicle surveys are an effective means of conducting such surveys. In on-board surveys, surveyors ride on-board the vehicle and distribute questionnaire forms to those passengers on the vehicle. The passengers may then be required to fill out the questionnaire forms at their convenience, and then return them via the mail. A comprehensive description of such surveys is given in Stopher (1985). They have the advantage of being moderately cheap, but the disadvantage of generating low response rates, since it is not possible to encourage or remind people to respond in any way.

(b) On-board vehicle distribution/on-board vehicle collection

As an alternative to method (a), it may be possible to collect the completed questionnaire forms before the respondents leave the vehicle.
For some modes, such as in-flight surveys, this poses no particular problems since there will generally be ample time for the passenger to complete the survey before the end of the trip. They may even welcome the survey as something constructive to do to pass the time on the flight. A very successful way of distribution/collection on planes is to distribute questionnaires to all persons on entry and to have surveyors at the destination airport to collect them from disembarking passengers (Ampt Applied Research, 1988).

It should be noted that these types of survey on public transport vehicles may pose considerable practical difficulties. For example, not everyone is seated in the vehicle, not everyone has a writing instrument, and not everyone has sufficient time to complete the questionnaire form before the end of their trip. Nonetheless, such surveys are often used, and it is then a matter of careful design and administration to ensure that the survey can be successfully completed under the particular conditions expected to be encountered. Pilot surveys are always critical to the successful execution of this type of survey.

(c) On-board distribution/collection plus mail-back

In some studies, hybrid on-board surveys, which combine elements of both methods (a) and (b), have been used successfully (Sheskin and Stopher, 1982a; Hensher, 1991). The method involves using a two-part questionnaire form. The first part is the more usual postcard style, clearly marked for filling out and return on the bus. The second part is a more lengthy form to be taken away by the bus traveller, filled out later in the day, and mailed back. This method allows for considerably more information to be obtained than can be acquired from the standard on-board distribution/mail-back method.

The robustness of the results of the intercept surveys which are carried out on-board a vehicle and described here in (a) - (c) depends very heavily on the voracity of the sampling method used to select the vehicles, routes and people which are included in the sample survey. For a discussion of issues related to sampling for on-board surveys, see the discussion in Section 4.4.4.

(d) Roadside distribution/mail-back surveys

Where the mode of transport under consideration is the private car, then the method of distribution to pin-point those users is often the roadside questionnaire survey. In this survey method, questionnaire forms are distributed to motorists as they pass by a particular point, or set of points, on the road. To enable the questionnaire forms to be distributed, it is
necessary that the motorist be stationary at that point. This can be achieved in one of two ways; either the motorist can be stopped at a natural feature of the roadway (such as a traffic signal or a toll-booth), or else the motorist can be deliberately stopped by the survey team (preferably with the assistance of local police officers). After motorists are stopped, they are given a questionnaire form and a brief explanation of the purpose of the survey. Respondents are then asked to fill out the questionnaire form at their convenience and return it by mail. They have many similarities with the on-board vehicle distribution/mail-back surveys in (a) above and are mostly differentiated by the longer introduction time given to the surveyor. Descriptions of these types of surveys may be found in Richardson and Young (1981) and Richardson et al. (1980).

The small amount of research done to validate surveys such as these suggests that there can be some significant biases using this method - for example an analysis of the 1961 London Travel Survey (LCC 1963) compared results of intercept interviews of passengers using London rail termini with results from mail-back surveys handed out at the same sites, and concluded that the mail-back survey over-represented regular commuters to central London. Similarly, Sammer and Fallast (1985) reported on a mail-back survey near Salzburg which showed that local (Austrian) residents were over-represented relative to foreign (German) residents. In both cases its was hypothesised that the over-represented groups had a greater stake in the journeys they were making because they participated in them more frequently, and therefore had a greater incentive to respond.

(e) Intercept interviews

Sometimes intercept surveys involve personal interviews with the drivers of vehicles or travellers as they are stopped at the intercept point. In these cases, the respondents are stopped by an interviewer who asks them a series of questions - most usually about origin, destination and times of travel, with some details on socio-demographic status. The presence of an interviewer generally ensures a much higher response rate than for methods which involve mailing back a postcard.

Some recent work in Shipley in England presented an opportunity to test the difference between road-side or on-vehicle distribution/mail-back and intercept interviews (Bonsall and McKimm, 1993). It was found that response rates varied from 60% (at a morning peak site where a prize draw was on offer) to 33% (at an all-day site with no prizes) with responses being as high as 62% for the journey to work and as low as 29% for the journey to employer’s business. In addition,
people who made the trip more frequently tended to respond better as did males. Although roadside and on-vehicle distribution surveys have frequently been used to define trip matrices, trip frequency distributions and trip purpose, there is clearly a fairly high risk in doing this without carrying out verification surveys such as intercept interviews where response rates are much higher.

(f) Activity centre distribution/mail-back or interview

This type of survey is similar to the above intercept methods in that the distribution of the questionnaire or administration of the interview is designed to capture the population of interest at a natural point of congregation (e.g. shopping centres, airports). If the activity centre is of a type that repeat visits by the population could be expected on a regular basis (such as work or school), then a mail-back could be wholly or partly replaced by having respondents deposit the completed forms at some strategic locations at the activity centre.

The biggest challenge in activity centre surveys is choosing a sample which is representative of all people visiting the centre. This is a particularly vexing issue in public areas of airports or shopping centres where there is no "funnel-mechanism" to allow distribution of questionnaires or interview of all people or of a random sample of people. The usual sampling approach to surveys in these activity centres is to use an uncontrolled quota sample (Section 4.4.7), i.e. to survey a certain number of people of given types (e.g. certain age groups, gender, destination) without knowing the proportion of these types which exist in the population (of people who visit the activity centre). These type of surveys (using uncontrolled quota sampling) have three major problems:

• there is no knowledge of the population from which the sample is drawn;
• there is no rigid sampling rule applied in the selection of respondents; and
• there is no information collected on those people approached who do not respond.

If each of these problems are addressed, surveys at activity centres can be as reliable as other intercept surveys.

A characteristic which is common to all intercept surveys is the inability (or limited ability) of the researcher to follow-up non-respondents. As already noted in the discussion on self-completion surveys, and as will be seen in that on personal interview surveys, one of the key factors which ensures a high degree of reliability of the data is the ability to gain information on non-respondents, and thereby to weight the data according. Consideration should be given to this
limitation of the methods at the time the survey is being designed. In many cases it may be possible to set up a small control survey-within-the-survey where either a follow-up procedure is implemented (e.g. using car registration numbers at roadsides, or ticket numbers on planes) or where a small part of the survey is conducted with a close to 100% response rate (using personal interview with multiple interviewers).

3.2.6 Household Personal Interview Surveys

A personal interview survey is defined as one in which an interviewer is present to record the responses provided by the respondent in answer to a series of questions posed by the interviewer. For this reason, many of the intercept surveys discussed in Section 3.2.5 as well as the telephone surveys outlined in Section 3.2.4 could also readily fall under this category. Our discussion in this section is, however, limited to personal interviews which take place in the home. Personal interview surveys have long been associated with transport, with home interview surveys providing the major means of data collection for the transport studies of the 1960s, 70s and 80s.

A household personal interview survey may be chosen, in preference to a self-completion survey, for any of several reasons:

(a) In general, higher response rates may be obtained from personal interview surveys than from self-completion surveys. Response rates of the order of 75% to 85% are not uncommon. This tends to minimise the effects of non-response bias, although it does not completely eliminate this bias as will be discussed in Chapter 9.

(b) The personal interview survey allows for considerable flexibility in the type of information collected. Attitudes, opinions, open-ended verbal answers and other non-quantitative information are much more readily collected in a personal interview survey than in a questionnaire survey. Complex sequence guides or filters can be used if required, since interviewers (unlike respondents) are always given training prior to the commencement of the survey.

(c) The presence of an interviewer means that explanations can be given regarding the meaning of questions or the method in which answers are to be given. As will be explained later (Chapter 7.3.2), the interviewer must generally adhere to a fixed set of questions, some of which need to be asked verbatim, but explanations of the meaning of questions are generally permitted so long as they do not influence the answer by the respondent. In a travel survey, this is particularly important in relaying information to respondents about the level of detail required for reporting trip and activity behaviour.
(d) Personal interview travel surveys can be carried out over a much shorter time period than self-completion surveys which need up to 6 weeks elapsed time to incorporate sufficient reminder notices into the survey procedure (see Section 7.2.1).

(e) Since many surveys can be quite long, an interviewer can be effective in maintaining respondent interest and in ensuring that the full set of questions is completed.

(f) By noting the interest of the respondent in the survey and the way in which the questions (especially attitudinal questions) are completed, the interviewer can make a valuable assessment of the validity of the recorded answers.

(g) The interview situation is valuable where it is desired to obtain spontaneous answers from a particular individual. Thus interview surveys are particularly suited, perhaps even essential, for attitude surveys.

While being particularly effective in several aspects of transport data collection, personal interview surveys are not without their own distinct disadvantages, including:

(a) Personal interview surveys are relatively expensive. Typically they would be three to ten times more expensive per returned questionnaire than a self-completion survey (this of course depends on the quality of the self-completion survey). The high cost of personal interview surveys is primarily due to the high labour content of interview surveys. A comparison of costs between persona interview and self-completion appears in Ampt and Richardson (1994).

(b) In order to reduce travel expenses and interviewer lost-time, many household-based personal interview surveys make use of clustering of households or survey sites on a geographic basis. This causes the "effective sample size" to be reduced with consequent reductions in the accuracy of estimates from the data.

(c) The interview situation is basically a human interaction between an interviewer and a respondent. Such interactions are rarely, if ever, completely neutral. The resulting interaction (often termed interviewer bias) may affect each participant (and the data which is collected) in various ways including:

(i) The personal characteristics of the interviewer (e.g. age, sex, nationality, general appearance) may influence the answers because of the impression made by the interviewer on the respondent (and vice versa).
Chapter 3

(ii) Respondents are being asked to disrupt their normal routine in order to answer the interview questions. If such a disruption is inconvenient, this may distort the answers given by respondents (if, for example, they wish to return to their normal routine as soon as possible).

(iii) An interviewer with strong opinions may subconsciously communicate these opinions to the respondent by the way in which the questions are asked or the answers are received. In some circumstances, some respondents may agree, or disagree, with statements depending on their perception of the temperament of the interviewer.

(iv) Answers to questions early in the interview allow the interviewer to build up a picture of the respondent’s general behaviour and attitudes. The interviewer may then interpret later answers (especially vague answers) to be consistent with this picture even though the respondent may have intended to give a contradictory or apparently inconsistent answer. Alternatively, the interviewer may construct a picture of the type of person being interviewed (based, for example, on socio-economic characteristics of the respondent) and then interpret vague answers to fit within an idea of the expected response from that type of person.

(d) Personal interview surveys are not suited for situations where questions require a considered response, or where factual information is required which is not immediately available. The time delay involved in obtaining such responses is either a waste of the interviewer’s time, or else the respondent feels embarrassed at making the interviewer wait for the answer.

In summary, personal interview surveys are best for attitude surveys, for surveys where the concepts are complex or where there is a complex series of sequencing required. They are more expensive than their self-completion counterparts and have to be designed thoroughly to minimise interviewer bias, but their high response rates and their ability to be carried out within a relatively short time interval make them ideal in cases where high quality data is required within a medium time frame.

3.2.7 Group Surveys

In the personal interview surveys described above, attention was focussed on the role of interviewer effects in the survey procedure. Interaction between individuals is inherent in many forms of survey procedure which attempt to record human behaviour, and in most surveys our intention is to minimise this
Selection of Survey Method

interaction in an attempt to produce results which are unaffected by the presence of an interviewer (Collins, 1978). The survey techniques outlined in this and the following section, however, take a substantially different view of these interaction effects. Instead of seeing interaction as being a totally negative phenomenon, they utilise the interaction between interviewer and respondent, and more particularly, the interaction between respondents, to enable the collection of a data base which is much richer in terms of its ability to explain the dynamics of travel and activity patterns.

A wide range of interactive and/or group survey techniques are available for use. Because of the fact that the output of these surveys is unlikely to be a hard set of statistics, but rather a better understanding of the problem at hand, such survey methods have often been referred to as "qualitative" survey methods (Gordon and Langmaid, 1988). Basically, there are two types of qualitative methodologies: the group discussion and the in-depth interview (Section 3.2.7). This section deals with group discussions which are also often known as focus groups.

The basic concept of group discussions is that a small number of people (usually between seven and nine) who are specially recruited according to a predetermined set of criteria, exchange experiences, attitudes and beliefs about a particular issue. Depending on the survey objectives, the criteria may be that the group be similar (for example, they may all be public transport users, or they may all live in a certain area) or dissimilar (for example to include professional drivers, regular drivers and those who do not have a driver's licence). The discussion is carried out under the guidance of a trained facilitator who:

- directs the flow of the discussion over areas that are important to the purposes of the survey;
- recognises important points and encourages the group to explore these and elaborate upon them;
- observes all non-verbal communication between respondents, or between respondents and the facilitator, or between respondents and the issue being discussed;
- creates an atmosphere that allows respondents to relax and lower some of their defences;
- synthesises the understanding gained with the problems and objectives of the survey; and
- tests out hypotheses generated by the information gained so far.
Chapter 3

Group discussions are usually tape-recorded to ensure that there is an accurate record of the interactions and to relieve the facilitator of the responsibility of taking notes during the session. They are sometimes also video-recorded to allow analysis of non-verbal, as well as verbal reactions after the discussion.

These discussions have several advantages which contribute to their usefulness in such situations, including:

- the group environment with 'everyone in the same boat' is less intimidating than a one-on-one in-depth interview;
- one respondent's experiences or feelings tend to trigger reactions from other respondents, whereby ideas which had lain dormant in the second respondent are now brought to the surface. It is, therefore, a good vehicle for creative expression from all respondents;
- the process highlights the differences between respondents (especially if respondents have deliberately been chosen with different backgrounds and experiences), thus making it possible to observe a range of attitudes and behaviours in a relatively short time;
- groups can be observed (by people other than the facilitator), thus making it particularly useful for professional staff of a transport agency who can experience respondents vocabulary, attitudes and reactions first-hand;
- spontaneity of response is encouraged in a group setting, yielding insights which may not be available from a one-on-one interview; and
- by careful selection of members of the group, the social and cultural influences on attitudes and behaviour are highlighted.

Group discussions are not, however, without their disadvantages, including:

- group processes may inhibit the frank exchange of attitudes and beliefs, especially from minority members of the group, and may lead to unrealistic and excessive recounting of behaviour;
- the group may react negatively to the facilitator, the subject matter or the discussion environment, and may freeze up;
- the strong personality of one respondent may overawe the other respondents who either withdraw or simply agree; and
Selection of Survey Method

• the group may lose perspective on the real issue, by getting too close to the problem and by discussing something in great depth which, in reality, may be more of an instinctive reaction.

For the above and other reasons, group discussions by themselves are not the most appropriate survey technique under the following circumstances:

• when statistically significant data is needed;
• where the subject matter is of a personal nature or where it involves personal finances;
• where social norms strongly predominate, creating extra pressure for conformity within the group;
• where a detailed understanding of a process is involved, and where it is unrealistic to expect all in the group to have the required level of understanding;
• where personal opinions on the subject matter are highly varied, and where the group becomes too heterogeneous to obtain useful information which could be generalised outside of the group; and
• where difficulties are experienced in recruiting the target sample (e.g. members of minority groups with specific transport system usage patterns).

Within the overall structure of group discussions, there are a number of variations. Mini-group discussions consist of 4-6 respondents instead of the usual 8-9, and are useful when focussed information is required quickly (e.g. immediate reactions to a new public transport marketing campaign), and where it may be difficult forming a larger discussion group. While normal group discussions may last from 1-2 hours, extended discussions may last from 3-4 hours (usually with some form of break). Extended discussions are particularly useful when you wish to use a variety of methods which involve the respondents in actually performing some task, which will provide you with more information about their real attitudes and beliefs. This is particularly the case where the respondents need more time to become comfortable with each other, or where there may be some time required to explain the tasks to them and allow them some time to familiarise themselves with the tasks. Reconvened group discussions, as the name implies, involve the group meeting on more than one occasion. These discussions are particularly useful when the respondents are asked to engage in a particular activity between the two meeting times (e.g. when car drivers are asked to try various forms of public transport). This provides the
Chapter 3

researcher with the opportunity to obtain immediate reactions to such activities in a controlled environment.

Brainstorming is a particular form of group discussion in which the respondents are asked to suspend evaluation of any ideas proposed by others in the group (or by themselves) and to simply concentrate on thinking about innovative solutions to the problem proposed by the facilitator. While these sessions sometimes appear to be virtually out-of-control, they require a very disciplined approach from the facilitator to ensure that the respondents behave in this fashion. The facilitator needs to be specially trained in techniques to aid creativity such as various methods of lateral thinking, thinking by analogy, fantasy solutions, word associations and imagery. Role-playing is sometimes used to help direct respondents in this case. For example, a respondent who is a professional driver may be asked to take on the role of a person who always walks and be asked to comment on a proposal to reduce the pedestrian access to favour cars.

In summary, group discussions are very useful for exploring issues for which the researcher has no clear-cut expectations about what will be found, but where there may be some hypotheses which needs to be tested in a fairly informal setting. They are, therefore, ideal for pre-pilot testing of questionnaires, and for exploring the jargon used by the population on a particular survey topic. They are not appropriate for providing statistically significant data.

3.2.8 In-Depth Interviews

The rise in popularity of in-depth interviews in transport studies follows directly from viewing travel as a "derived demand"; that is, we almost always travel in order to carry out an activity outside the home - not for the sake of travel itself. Because of this, a full understanding of travel behaviour can only be achieved by viewing travel as but one of a series of activities which are carried out in time and space (Jones, et al., 1983; Carpenter and Jones, 1983).

In-depth interviews are defined as those which are orientated to penetrating below the superficial question-and-answer format of structured or semi-structured personal interviews, which have attention paid to building rapport in order to facilitate the expression of sincere beliefs and attitudes. The interviews usually last for an hour or more, and are tape-recorded (in general, no written notes are taken in the interview to enable the interviewer to concentrate on what the respondent is saying).

3.2.8.1 Individual In-Depth Interviews

In-depth interviews usually take place with a group of people (most commonly all people in a household) but they can also take place with just one person. Particular benefits of the individual in-depth interview are:
Selection of Survey Method

- longitudinal information can be gathered on one respondent at a time (e.g. travel patterns or expenditure information);
- problems with dominance and intra-group rivalry are absent, enabling both majority and minority opinions to be expressed;
- personal material can be discussed, especially with an experienced interviewer (e.g. values which may be considered socially unacceptable, such as a love of V-8 engines).

3.2.8.2 Interactive Group Interviews

The most common form of in-depth interview in transport studies is the interactive group interview. These interviews usually take place in households where all members of the household take part in the interview. Jones (1985) provides an excellent comprehensive description of these techniques.

As noted by Jones, the interactive group surveys developed in transportation (of which Jones (1979a) and Brög and Erl (1980) are the best known) share three common features: interaction between participants, use of visual aids as a structuring device, and the development of gaming simulation techniques. With respect to interaction between participants, the interactive group survey is different, compared to a conventional structured personal interview travel survey, in that:

(a) interaction in the interview is exploited rather than suppressed;
(b) interviews are tape recorded, rather than recorded in written form during the interview, because the interviewer would not have enough time to conduct the interview and record the responses;
(c) interviewers are highly skilled and require a detailed understanding of the problem being studied;
(d) there is no formal questionnaire, although the interviewer may refer to a topic list during the interview;
(e) questions are framed during the course of the interview, in response to previous discussion; and
(f) probes and supplementary questions do not need to be phrased neutrally, as they would be in a structured interview; in some circumstances, the interviewer may act as a "devil's advocate" in order to elicit more detailed responses from the respondents.
Visual aids have an important role in a successful interactive survey methodology by introducing some element of structure into the discussion. In particular they may provide
- an aid to comprehension;
- a (passive) prompt;
- an aide memoir;
- a check on the feasibility of responses;
- a means of dealing with complex issues;
- a device for obtaining quantifiable data; and
- a means of contributing to a more relaxed interview environment.

As an example of the use of visual aids in an interactive survey, consider the interactive survey described by Jones (1979a, 1980) as the Household Activity-Travel Simulator (HATS). This technique employs the use of a display board, known as a HATS-board, as a central focus for the interactive survey (see Figure 3.2). Travel or activity diaries which have been completed by members of the household for a specified period before the interview are translated into a physical representation of a given day by means of coloured blocks along three parallel axes to indicate at-home activities, away-from-home activities and connecting travel activities. The spatial arrangement of these activities and trips is also shown on the HATS-board map. Household members are then given the chance to describe their activity-boards and to discuss possible interactions between the activity-boards of several members of the household. The HATS-board provides the focus for these discussions, by fulfilling the functions listed above.

The third element of interactive surveys is the use of a gaming simulation approach within the interview. The use of these "what-if" techniques enables the analyst to investigate the range of likely adaptations to a variety of policy options. For example, in the HATS survey, after the group members have finished discussing their current travel and activity patterns, they are asked to respond to various changes in policies which may affect either their activity or travel patterns. The responses are shown by rearrangement of the blocks on the HATS-board and/or rearrangement of the spatial patterns on the map. Problems which arise as a result of these responses immediately become visible on the activity-boards and map, in terms of gaps between blocks, overlapping blocks, no travel block linking at-home and away-from-home activities, and inconsistencies between activity-boards for various members of the household.
These inconsistencies result in individual adaptations being tested by one member of the household and then re-checked with other members’ adaptations. In Monopoly-game fashion, the members of the household engage in conversation and trade-offs to find feasible household adaptations to the policy changes. In the process, numerous infeasible adaptations are suggested and
Chapter 3

rejected. After a final adaptation is adopted by all members of the group, the HATS-board displays are coded onto new activity diaries and these, together with a tape recording of the discussion, form the basis for future analysis.

The HATS technique can be most effectively used to investigate the adaptation to forced changes (e.g. curtailing of a public transport route), but may also be used to investigate the effects of a permissive change (e.g. providing a new activity at an away-from-home location) which allows, but does not necessitate, a change to be made. In the later case, greater attention needs to be paid to the feasibility of any suggested adaptations.

The HATS technique, or variations of it, has been used in a variety of situations including the assessment of the attitudes of bus driver's towards a variety of patterns of shift work (Bottom and Jones, 1982), the reactions of households to various policy measures aimed at reducing transportation energy consumption (Phifer et al., 1980), the identification of transportation needs of disabled people (Faulkner, 1982), and possible reactions to congestion management strategies (Ampt and Jones, 1992).

A slightly different approach to the use of simulation in interactive surveys has been adopted by Brög and Erl (1980). In their methods, they adopt the concept of the "situational approach" which assumes that an individual's travel patterns are not simply a function of that individual's preferences, but also depend on the social and environmental situation in which that individual exists. Once again, they use the idea of an in-depth survey in which an interactive "what-if" simulation is employed, but they present the results of the simulation in terms of a hierarchical decision-tree which illustrates the effects of situational constraints on that individual. An example of such a decision-tree is given in Figure 3.3, which demonstrates the potential for usage of a new transit route in Melbourne, Australia (Ampt et al., 1985). It can be seen that although a large proportion of the population under study had the new mode objectively available to them (25% in Reservoir and 75% in Kingsbury), most of these potential users were subsequently eliminated from consideration by a variety of situational and perceptual filters. This method of presenting the results of an interactive group survey have been found to be highly effective in explaining the results of such a survey to a non-technical client. The method has been used in a range of applications, including studies of the potential of bicycle travel (Brög and Otto, 1981), long distance personal travel (Brög, 1982), and urban public transport (Brög and Erl, 1981).
Selection of Survey Method

Figure 3.3 Decision Trees in the Situational Approach Survey
(Source: Ampt, et al., 1985)
Chapter 3

While interactive survey techniques have been found to possess a number of distinct advantages when attempting to understand travel behaviour, they do have a number of limitations, including:

(a) The interactive survey method is expensive. The cost per interview may be three to four times more expensive than a structured personal interview survey;

(b) Interactive surveys must be carried out by experienced interviewers who have a good knowledge of the subject matter of the survey. Often it is the researcher or principal investigator who must conduct the interview. This is necessary because only a person in this position will have enough detailed understanding of the topic to be able to follow up avenues of thought which arise spontaneously in the interview and which appear to be most productive. Thus, as well as being costly, interactive surveys, because of their limited manpower base, can also be very time consuming;

(c) Because of their time and cost requirements it follows that, within a limited budget, only a relatively small number of interactive surveys may be completed for any one study. This may pose problems when attempting to make generalisations from such a small sample; and

(d) Interactive surveys do not yet provide data in a form that is amenable to the construction of detailed mathematical models of travel behaviour or traffic flows. While progress has been made in this direction (e.g. the CARLA simulation model developed at Oxford University (Clarke, 1984)), the end-product of interactive surveys is not the production of seemingly precise estimates of traffic flows or transit patronage. Rather, they seek to provide a better basic understanding of travel behaviour.

On the other hand, there are many advantages to these methods. Although there are some cases where interactive methods may form the main approach to studying a transport problem or issues, most of the application of this method lies in using it in the context of a more conventional, quantitative travel study. Jones (1985) suggests that there are three broad roles which interactive techniques can perform in this context, as shown in Figure 3.4.

The first of these is exploratory. Before the main study is begun, interactive measurement can be used as an aid to define transport problems and/or policy issues, and to help formulate the appropriate quantitative methodology. This can include content and wording of the questionnaires, the hypothesis formation or the provision of guidelines for appropriate modelling structure. This is probably the best developed and most frequently used role of the interactive interview.
The second role for interactive surveys is *investigative*. Once a conventional travel survey has been completed, the analysis may raise a number of issues which cannot be answered by the data in a statistical way. This occurs often when the survey concentrated on what was happening in the study area, not why it happened - as is frequently the case in travel surveys. Interactive interviews can be employed in these situations to exploring certain issues in greater depth with respondents, or as a means of communicating findings and their implications to decision makers.

Finally, interactive surveys may be used as *investigative* tools. Rather than being confined to a role in preparation or conclusion of a survey, they may actually be used as a component of the main study phase (e.g. Ampt and Jones, 1992), as one means of investigating the subject matter. This is often achieved by using in-depth interviews or group discussions alongside structured surveys as a complementary means of obtaining information about behaviour. In some cases the association may be even close, with interactive methods being used to inform model development through the use of gaming simulation to guide model inputs, structures and forecasts (e.g. Brög and Zumkeller, 1983).

### 3.3 SUMMARY OF SURVEY METHOD SELECTION

The final choice of survey method will depend upon a matching of the characteristics of the individual survey methods, as outlined above, with the objectives of the survey. This will be tempered by the resources available for the conduct of the survey. Stopher and Banister (1985) summarise the issues to be faced in the selection of a survey method and Table 3.2 builds on this to include the survey types discussed in this chapter.
Chapter 3

Table 3.2 Uses of Each Survey Method

<table>
<thead>
<tr>
<th>Survey Type</th>
<th>Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Factual Travel</td>
</tr>
<tr>
<td>Documentary searches</td>
<td>Yes</td>
</tr>
<tr>
<td>Observational surveys</td>
<td>Yes</td>
</tr>
<tr>
<td>Household self-completion</td>
<td>Yes</td>
</tr>
<tr>
<td>Telephone surveys</td>
<td>No</td>
</tr>
<tr>
<td>Household</td>
<td>No</td>
</tr>
<tr>
<td>Individual</td>
<td>Yes</td>
</tr>
<tr>
<td>Validation</td>
<td>Yes</td>
</tr>
<tr>
<td>Intercept surveys</td>
<td>Limited</td>
</tr>
<tr>
<td>Household personal interview</td>
<td>Yes</td>
</tr>
<tr>
<td>Group surveys</td>
<td>Limited</td>
</tr>
<tr>
<td>In-depth surveys</td>
<td>Limited</td>
</tr>
</tbody>
</table>

In line with the "Total Design" concept proposed by Dillman (1978), several authors have suggested the need for a dual or multiple survey mechanisms for the collection of data. In such a process, a battery of survey methods would be used to collect data, with each method being used in its most appropriate context. For example, as noted earlier, Sheskin and Stopher (1980) used a combined on-board questionnaire survey and a take-home questionnaire survey for gathering data on public transport users. Stopher (1985a) suggests the combination of personal interviews with mail-back questionnaire surveys, and telephone surveys with mail-out/mail-back surveys (Stopher, 1982). Shaw and Richardson (1987) report on the combination of an activity centre interview with a mail-back questionnaire survey. Brög and Erl (1980) point to the need to use a variety of methods within an overall study design and offer the following general guidelines:

(a) Structured questionnaires are appropriate for obtaining socio-demographic data at a person and household level;

(b) Diaries are the best means of obtaining activity and travel information (otherwise respondents tend to recall typical behaviour rather than provide an accurate record for the survey period);

(c) Land-use and transport system information is best obtained from official statistics or by carrying out an inventory;
Selection of Survey Method

(d) Face-to-face, in-depth unstructured interviews are appropriate for exploring perceptions and attitudes;

(e) Group discussions are a better means of obtaining information about household organisation, decision rules, etc. than direct questioning, since household members are often unaware of their own organisational structures; and

(f) Simulation techniques are a useful means of exploring response to change, as they make explicit the constraints, options and decision rules which contribute to the observed adaptation.

In general, after reading this chapter, we think it becomes clear that the selection of the survey method can only sensibly take place after a careful period of preliminary planning to ensure that the survey method chosen is best suited to measuring what is needed to fulfil the study and survey objectives. Furthermore, we hope it also becomes clear that the selection of a survey method is vital to ensuring that the highest quality data is collected in the most cost effective way.
The selection of a proper sample is an obvious prerequisite to a sample survey. A sample is defined to be a collection of units which is some part of a larger population and which is specially selected to represent the whole population. Four aspects of this definition are of particular importance: first, what are the units which comprise the sample; second, what is the population which the sample seeks to represent; third, how large should the sample be; and fourth, how is the sample to be selected?

4.1 TARGET POPULATION DEFINITION

The target population is the complete group about which one would like to collect information. The elements of this group may be people, households, vehicles, geographical areas or any other discrete units. The definition of the target population will, in many cases, follow directly from the objectives of the survey. Even so, there may be several problems to be resolved in the definition of the population.

To outline the types of questions which may need to be resolved in the definition of a survey population, consider a survey which had the objectives of determining the travel patterns of elderly, low-income residents living in a particular area of Melbourne, Australia (Richardson, 1980). The objective of the study for which the survey was undertaken was to identify the degree to which...
this group was transport-disadvantaged and to suggest means by which they might be provided with an improved level of transport service. To define the population for this survey, several issues had to be resolved. First, a definition of "elderly" had to be adopted. With no particular theoretical justification, it was decided to define elderly as being equivalent to being "of retirement age". At the time of the survey, this meant 60 years of age for women and 65 years of age for men. Second, "low-income" also had to be defined. Whilst many income and/or wealth indicators could theoretically have been used for this purpose, there appeared to be no easily identifiable level of income which would serve as a definition of "low-income". In fact, the practical definition of "low-income" used in this survey was not resolved until later in the survey process when the selection of a sampling frame was being made. Third, the geographical extent of the study area had to be defined. Since the study was initiated and supported by the Sherbrooke Shire Council, it was easily resolved that the study should concentrate on residents of the Sherbrooke Shire. However, given that, there was still some debate as to whether the survey should cover the entire Shire or whether it should concentrate on a number of relatively inaccessible areas within the Shire. Mainly because very little was known of elderly travel patterns in the area at the time, it was decided to survey the entire Shire to build up a general picture of elderly travel patterns. The survey population thus defined was low-income (yet to be defined) residents of retirement age living in the Shire of Sherbrooke.

4.2 SAMPLING UNITS

The survey population is composed of individual elements. Thus continuing with the example of the Shire of Sherbrooke survey, the elements of the population were the individual elderly residents. However, the selection of a sample from this population was based on the selection of sampling units from the population. Sampling units may or may not be the same as elements of the population; in many cases, they are aggregations of elements.

Thus in the Sherbrooke survey, it was decided to select elderly households rather than individual elderly residents. This decision was made for three reasons. First, it was thought that elderly travel patterns and needs were best addressed by considering the total travel of the elderly household. Thus, even though an individual may not travel greatly, the needs of that individual may be serviced if another member of that household was able to travel on their behalf, at least for needs like shopping and personal business. Second, it was considered that where an elderly household consisted of two elderly individuals it would be difficult to restrict the survey to just one of them without the other feeling left out of things. Third, it was considered that the productivity of the interviewers would be increased by having them interview all elderly residents in one household rather than just one selected individual in each household.
Given the definition of the sampling unit as being an elderly household, it was therefore necessary to define an elderly household. The adopted definition of an elderly household was one in which at least one elderly individual (as previously defined) resided on a permanent basis.

In more general situations, sampling units may typically include such entities as:

(a) Individuals
(b) Households
(c) Companies
(d) Geographic regions (zones, cities, states, nations)
(e) Vehicles
(f) Intersections or road links
(g) Other features of the transport network

4.3 SAMPLING FRAME

Having identified the desired survey population and selected a sampling unit, it is necessary to obtain a sampling frame from which to draw the sample. A sampling frame is a base list or reference which properly identifies every sampling unit in the survey population. Clearly, the sampling frame should contain all, or nearly all, the sampling units in the survey population.

Depending on the population and sampling units being used, some examples of sampling frames which could be used for various transport surveys include:

(a) Electoral rolls
(b) Block lists (lists of all dwellings on residential blocks)
(c) Lists by utility companies (e.g. electricity service connections)
(d) Telephone directories
(e) Mailing lists
(f) Local area maps
(g) Census lists (if available)
(h) Society membership lists
(i) Motor vehicle registrations
Each of these sampling frames, however, suffers from one or more of the deficiencies outlined below:

(a) Inaccuracy
All sampling frames will generally contain inaccuracies of one sort or another. Lists of individuals will contain mis-spelt names and incorrect addresses. Maps will often have incorrect boundary definitions, will not be true to scale and will have streets and other features which simply do not exist.

(b) Incompleteness
As well as having incorrect entries, sampling frames may simply not have some valid entries at all. For example, electoral rolls may not have the names of individuals who have recently moved into an area or of non-citizens of the country; temporary workers may not be included on a list of company employees; off-road recreational vehicles may not be on motor vehicle registration records; some telephone numbers are unlisted.

(c) Duplication
Entries may also be duplicated on some sampling frames. For example, telephone directories may list individuals and companies more than once under slightly different titles; an individual may appear on the electoral roll for two different areas soon after they have moved from one area to another; an individual may appear on a Local Government electoral roll more than once, if they own land in more than one ward of the Local Government Area.

(d) Inadequacy
A sampling frame is said to be inadequate if it simply does not provide a close correspondence with the desired survey population, but has been adopted for the sake of convenience.

(e) Out-of-date
Whilst a sampling frame may once have been adequate, accurate, complete and with no duplications, this situation may not last forever. Conditions change and, with these changes, sampling frames go out-of-date. Thus telephone directories commonly do not include people who have moved in the previous 18 months; electoral rolls are usually up-to-date only at election time; lists of all sorts are compiled for special purposes but are not kept current after the immediate need for the list has passed.

In many cases, the reason for the deficiencies listed above is that the list which one wishes to use as a sampling frame, has been compiled for a completely
Sampling Procedures

different reason. Thus electoral rolls only need to be current at the time of elections; telephone directories only list those who have telephones; maps are compiled for specific purposes to an acceptable degree of accuracy for that purpose. While lists from utility companies are probably the most up-to-date, they too can rapidly become out-of-date as people move and houses become vacant. Before adopting a list as a sampling frame, it is wise to ascertain the reasons why the list was initially compiled, the criteria for inclusion on the list and the methods used for up-dating of the list. Having said this, it should be noted that the acceptability of lists for sampling frames will depend entirely on the scope and purpose of the survey to be conducted. If the deficiencies in a list will not affect the results of the survey, then such a list may be entirely satisfactory as a sampling frame for that survey. The question of acceptability revolves entirely around the absence of an introduction of bias into the survey results.

To illustrate the restrictions imposed, and the opportunities offered, by the availability of certain sampling frames, consider the sampling frame used in the Sherbrooke Shire elderly travel survey (Richardson 1980). It will be remembered that the desired population was all low-income, elderly residents living in the Shire of Sherbrooke at the time of the survey (April 1979). The question, then, was where could one find a list of such elderly residents. Since elderly had been defined as "of retirement age" and since all retired people were eligible for old-age pensions, the obvious source was the Department of Social Security pension mailing-lists for the Shire of Sherbrooke. However, enquiries revealed that confidentiality requirements would deny accessibility to individual names and addresses on that list. Even if such names and addresses were available, however, the problem of defining "low-income" would still have remained. Another sampling frame was therefore sought.

Several possibilities were suggested but each had attendant deficiencies. Two proposals were given serious consideration. The first entailed the use of an informal network of contacts established through Senior Citizens Clubs in the area and through the Shire’s community services activities (such as Meals-on-Wheels). It was considered that by taking the names of those directly involved in these activities, and then asking them for the names of other elderly residents whom they know to live in the area, that a reasonable sampling frame would be established. This method, however, had a number of problems. First, the establishment of the network would require considerable time and effort; second, the completeness of the sampling frame would be unknown, but would probably be biased towards those who already found it possible to travel to the Senior Citizens Clubs in the area and hence were less likely to be transport-disadvantaged; third, the accuracy of the addresses supplied for other elderly residents was open to question.
Chapter 4

The second proposal entailed the use of a list compiled by the Shire Council detailing all those pensioners who had applied for a rebate of the land rates paid to Council in the previous year. Such rate rebates were means-tested and hence this provided a definition of "low-income" which had so far proved elusive. This sampling frame was, however, not without its own problems. First, the list contained all pensioners who had received rate rebates and not just old-age pensioners; it would therefore be necessary to separate the old-age pensioners at a later stage. Second, the list obtained was for the previous year and hence might have been slightly out-of-date. Third, and most importantly, because of its very nature the list contained names of only those pensioners who owned a house (or land) in the Shire. Thus low-income elderly residents who rented property would not be on the list. Since it might be expected that renters would be lower-income than owners then this omission could be of some significance. To the extent that renters comprised a portion of the total elderly resident population, the sampling frame would be in error. Despite this shortcoming, this sampling frame was used in the study since time and resources prevented the establishment of a more comprehensive sampling frame. Nonetheless, the results of the study had to be interpreted with reference to the sampling frame used.

In the example cited above and in many other surveys, especially of relatively rare populations, the establishment of a sampling frame can be a major problem. In many cases, the sampling frame available may largely define the population and the sampling method. As a result, the detailed survey planning must often await the identification of the available sampling frames. If no adequate sampling frames can be found, then it may be necessary to conduct a preliminary survey with a view to establishing a suitable sampling frame. Alternatively, the survey can be designed using a larger than required sampling frame and using filter questions at the beginning of each questionnaire (or interview) to eliminate non-relevant sampling units from the survey.

4.4 SAMPLING METHODS

The object of sampling is to obtain a small sample from an entire population such that the sample is representative of the entire population. It is therefore of some importance to ensure that the sample is drawn with care to ensure that it is indeed representative. The need for sampling is based on the realisation that in transport studies we are often dealing with very large populations. To attempt to survey all members of these populations would be impossible. For example, the 1981 Sydney Travel Survey took approximately 12 months to collect, code and edit data on the travel patterns of 20,000 households. To attempt a 100% sample survey on the same annual budget would take over 50 years! Sample surveys are also used because, not only it is often not possible to collect data on all members of a population but, it is also not necessary. As will be seen later, quite accurate
estimates of population characteristics can be obtained from relatively small samples.

The accuracy of sample parameter estimates, however, is totally dependent on the sampling being performed in an acceptable fashion. Almost always, the only acceptable sampling methods are based on some form of random sampling. Random sampling entails the selection of units from a population by chance methods such as flipping coins, rolling a die (not two dice), using tables of random numbers or through the use of pseudo-random numbers generated by recursive mathematical equations.

The essence of pure random sampling is that sampling of each unit is performed independently and that each unit in the population has an equal probability of being selected in the sample (at the start of sampling).

There are many types of sampling methods, each of which is based on the random sampling principle. The most frequently encountered methods are:

(a) Simple random sampling
(b) Stratified random sampling
(c) Variable fraction stratified random sampling
(d) Multi-stage sampling
(e) Cluster sampling
(f) Systematic sampling

In addition, there are a number of other sampling methods which, while used in transport surveys, are not based on random sampling and are therefore not highly recommended. These methods include quota sampling and expert sampling. This section will describe each of the above sampling methods, indicating their strengths and weaknesses.

4.4.1 Simple Random Sampling

Simple random sampling is the simplest of all random sampling methods and is the basis of all other random sampling techniques. In this method, each unit in the population is assigned an identification number and then these numbers are sampled at random to obtain the sample. For example, consider a population of 100 sampling units, as depicted by the 100 asterisks in Figure 4.1. The task is to select a random sample of 10 sampling units from this population.
The first step in simple random sampling is to name each of the sampling units. This is often done by assigning a unique identification number to each of the sampling units, even if they already have unique identifying names. Thus the population of 100 sampling units now appears as shown in Figure 4.2.
Sampling Procedures

Using random number selection methods (to be described later in section 4.8), a set of ten random numbers is now selected, and the sampling units corresponding to these numbers are included in the sample, as shown by the bold numbers in Figure 4.3.

```
<table>
<thead>
<tr>
<th></th>
<th></th>
<th>02</th>
<th></th>
<th></th>
<th>04</th>
<th>05</th>
<th>06</th>
<th>07</th>
<th>08</th>
<th>09</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>31</td>
<td>32</td>
<td>33</td>
<td>34</td>
<td>35</td>
<td>36</td>
<td>37</td>
<td>38</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>41</td>
<td>42</td>
<td>43</td>
<td>44</td>
<td>45</td>
<td>46</td>
<td>47</td>
<td>48</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>51</td>
<td>52</td>
<td>53</td>
<td>54</td>
<td>55</td>
<td>56</td>
<td>57</td>
<td>58</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>61</td>
<td>62</td>
<td>63</td>
<td>64</td>
<td>65</td>
<td>66</td>
<td>67</td>
<td>68</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>71</td>
<td>72</td>
<td>73</td>
<td>74</td>
<td>75</td>
<td>76</td>
<td>77</td>
<td>78</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>81</td>
<td>82</td>
<td>83</td>
<td>84</td>
<td>85</td>
<td>86</td>
<td>87</td>
<td>88</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>91</td>
<td>92</td>
<td>93</td>
<td>94</td>
<td>95</td>
<td>96</td>
<td>97</td>
<td>98</td>
<td>99</td>
<td></td>
</tr>
</tbody>
</table>
```

Figure 4.3 A Simple Random Sample of 10 Sampling Units

The question arises in this method as to whether sampling should be performed with replacement, or without replacement. The usual practice is to sample without replacement such that each unit can be included in the sample only once; this is particularly so when dealing with individuals or households which are being sampled for an interview survey. It is possible, however, to sample with replacement and simply include the results from those sampling units selected more than once as many times as they are selected. That is, if a household is selected twice, then you do not interview the household twice but merely include the results from this household twice in the data set. Notwithstanding the above, conventional practice in transport surveys is to sample without replacement.

A variation on the random sampling method which is sometimes used in transport surveys is to select a number and then let the sampling units themselves introduce the randomness. For example, in licence plate travel time surveys, the usual practice is to record arrival and departure times of all vehicles whose licence plate ends in a certain digit (for a 10% sample) or digits (for higher sampling rates). Since it would be expected that the occurrence and time of arrival of such vehicles would be random, the resulting estimates of traffic flow and travel time should be unbiased.
Chapter 4

Depending on the sample size, simple random sampling often gives highly variable results from repeated applications of the method. It would therefore be desirable if this variability could be reduced whilst still maintaining the characteristics of a random sample. In other situations, the cost of simple random sampling would be excessive. Several improvements on simple random sampling have therefore been developed.

4.4.2 Stratified Random Sampling

The sample of ten sampling units identified in Figure 4.3 may well be a good representation of the 100 sampling units in the population. However, if we have some prior information about the population, it may be clear that this is not the case. For example, assume that the sampling frame depicted in Figure 4.2 comes from a list of employees in a company and that, for some other reason, the employees are listed by gender such that the first 40 employees are female and the second 60 employees are male, as shown in Figure 4.4.

<table>
<thead>
<tr>
<th>00</th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
<th>06</th>
<th>07</th>
<th>08</th>
<th>09</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
</tr>
<tr>
<td>30</td>
<td>31</td>
<td>32</td>
<td>33</td>
<td>34</td>
<td>35</td>
<td>36</td>
<td>37</td>
<td>38</td>
<td>39</td>
</tr>
<tr>
<td>40</td>
<td>41</td>
<td>42</td>
<td>43</td>
<td>44</td>
<td>45</td>
<td>46</td>
<td>47</td>
<td>48</td>
<td>49</td>
</tr>
<tr>
<td>50</td>
<td>51</td>
<td>52</td>
<td>53</td>
<td>54</td>
<td>55</td>
<td>56</td>
<td>57</td>
<td>58</td>
<td>59</td>
</tr>
<tr>
<td>60</td>
<td>61</td>
<td>62</td>
<td>63</td>
<td>64</td>
<td>65</td>
<td>66</td>
<td>67</td>
<td>68</td>
<td>69</td>
</tr>
<tr>
<td>70</td>
<td>71</td>
<td>72</td>
<td>73</td>
<td>74</td>
<td>75</td>
<td>76</td>
<td>77</td>
<td>78</td>
<td>79</td>
</tr>
<tr>
<td>80</td>
<td>81</td>
<td>82</td>
<td>83</td>
<td>84</td>
<td>85</td>
<td>86</td>
<td>87</td>
<td>88</td>
<td>89</td>
</tr>
<tr>
<td>90</td>
<td>91</td>
<td>92</td>
<td>93</td>
<td>94</td>
<td>95</td>
<td>96</td>
<td>97</td>
<td>98</td>
<td>99</td>
</tr>
</tbody>
</table>

**Figure 4.4 A Simple Random Sample from a Stratified Population**

It is clear from Figure 4.4 that we have inadvertently over-sampled females (selecting 5 out of 40) and under-sampled males (selecting 5 out of 60). As a result, any inferences drawn from this sample will be biased towards the behaviour or attitudes of females because they are over-represented in the sample compared to their representation in the population.

To overcome this problem, stratified random sampling makes use of prior information to subdivide the population into strata of sampling units such that the units within each stratum are as homogeneous as possible with respect to the stratifying variable. Each stratum is then sampled at random using the same
Sampling Procedures

Sampling fraction for each stratum. When the same sampling fraction is used in each stratum, this method is sometimes called proportionate stratified sampling. The resulting sample will then have the correct proportion of each stratum within the whole population, and one source of error will have been eliminated. For example, if the males and females in Figure 4.4 were each sampled at a rate of 10% then the sample shown in Figure 4.5, where sampling unit 55 is substituted for sampling unit 33, would be a more representative sample than that shown in Figure 4.4.

<table>
<thead>
<tr>
<th></th>
<th>00</th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
<th>06</th>
<th>07</th>
<th>08</th>
<th>09</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td><strong>17</strong></td>
<td>18</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>31</td>
<td><strong>32</strong></td>
<td>33</td>
<td>34</td>
<td>35</td>
<td>36</td>
<td>37</td>
<td>38</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>41</td>
<td>42</td>
<td>43</td>
<td>44</td>
<td>45</td>
<td>46</td>
<td>47</td>
<td>48</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>51</td>
<td>52</td>
<td>53</td>
<td>54</td>
<td>55</td>
<td>56</td>
<td>57</td>
<td>58</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>61</td>
<td><strong>62</strong></td>
<td>63</td>
<td>64</td>
<td>65</td>
<td><strong>66</strong></td>
<td>67</td>
<td>68</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>71</td>
<td><strong>72</strong></td>
<td>73</td>
<td>74</td>
<td>75</td>
<td>76</td>
<td>77</td>
<td>78</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>81</td>
<td>82</td>
<td>83</td>
<td>84</td>
<td>85</td>
<td>86</td>
<td>87</td>
<td>88</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>91</td>
<td>92</td>
<td>93</td>
<td>94</td>
<td>95</td>
<td>96</td>
<td>97</td>
<td>98</td>
<td><strong>99</strong></td>
<td></td>
</tr>
</tbody>
</table>

Female

Male

Figure 4.5 A Stratified Random Sample from a Stratified Population

While the stratified sample in Figure 4.5 is obviously more representative than that in Figure 4.4 (in that males and females are represented in their correct proportions), the question remains as to whether the stratified sample is still a random sample of the population. This question can be answered by reference to the two criteria for random sampling noted earlier in this chapter. That is, a sample is random if:

- each unit is sampled independently; and
- each unit in the population has an equal probability of being selected in the sample (at the start of the sampling process).

With respect to stratified random sampling, it is clear that the second condition obviously holds, in that all males and females each have the same chance of selection (i.e. 10%) at the start of the process. With respect to the first condition, within each strata each unit is selected independently because simple random sampling is being employed within each strata. Therefore, given that each strata
Chapter 4

is sampled independently of the other, the addition of two independent random samples will produce a third random sample.

To use stratified sampling, it is necessary that some prior information about the population is available before sampling takes place. The prior information should also relate to the variables which are to be measured in the survey. For example, if one were attempting to measure trip generation rates in a survey, then stratification on the basis of car ownership would be more useful than stratification on the basis of the day of the week on which the respondent was born (assuming both data sets were available prior to sampling). Whilst the latter stratification would ensure that we got the correct number of people born on each day in our sample, we would not expect this to improve our estimate of trip generation rates. On the other hand, by having the correct average car ownership in our sample, rather than (by chance) too high or too low an estimate of car ownership, we would expect a better estimate of trip generation rate.

Therefore stratified sampling requires that we have some prior information about each unit in our population which is relevant to the objectives of the survey.

Whilst stratified sampling is useful, in general, to ensure that the correct proportions of each stratum are obtained in the sample, it becomes doubly important when there are some relatively small sub-groups within the population. With simple random sampling, it would be possible to completely miss out on sampling members of small sub-groups. Stratified random sampling at least ensures that some members of these rare population sub-groups are sampled (assuming these sub-groups were used as strata and that the product of the sub-group population size and the sampling rate produces a number greater than one).

A final advantage claimed for stratified sampling is that it allows different survey methods to be used for each of the strata. An example given by Stopher and Meyburg (1979), concerning stratification on the basis of access distance to a rail station, suggests that while the strata with shorter access distances may be surveyed by a postal questionnaire, the stratum with the longest access distance should be surveyed by personal interview on the basis that they are less likely to be transit-oriented travellers. Whilst such a variation in survey method is possible, care should be taken when comparing, or combining, the results for the different strata because of the different biases built into each of the survey methods.

A variation on stratified sampling is the use of multiple stratifications. Thus, instead of stratifying with respect to only one variable, the stratification can be performed with respect to several variables thus creating an n-dimensional matrix of stratification cells. In selecting the number of dimensions for
Sampling Procedures

stratification and the number of strata within each dimension, attention should be paid to the total number of stratification cells produced. Since the number of cells increases geometrically with the number of dimensions or strata, it is possible to produce a large number of cells inadvertently. In such a case, the average number of units in the sample within each cell could be quite small (perhaps fractional). Under these conditions, the necessary round-off errors in drawing a sample could defeat the purpose of stratification, unless carefully controlled.

The method by which stratification is conducted will depend to a large extent on the structure of the sampling frame to be used. In some sampling frames, the stratification may have already been performed in the compilation of the sampling frame list. For example, students at a University may already be categorised by Faculty. In such a case, an unrestricted random sample is conducted separately within each of the stratified lists. In other sampling frames, the list ordering may be completely random but it may be known how many sampling units belong in each stratum and, therefore, how many in the sample should come from each stratum. In this case, a random sample may be drawn from the entire list and, upon selection, each unit is placed in its correct stratum. When the required quota for each stratum has been sampled, further selections for that stratum are rejected and another selection is made.

Finally, it should be noted that the concept of stratification can also be used after the data have been collected by means of a simple random sample survey. In such a case, the survey results can be adjusted so that each stratum is represented in the correct proportion. Such weighted "expansion" (see Section 9.1) is frequently performed when the required stratification information does not become available until after the survey has been performed (e.g. when travel surveys are performed in Census years). However, it should be noted that such a procedure is strictly valid only when there is a sufficiently large sample size within each of the strata to enable reasonable confidence to be held in each of the strata results.

4.4.3 Variable Fraction Stratified Random Sampling

The above discussion of stratified random sampling has implicitly assumed that within each stratum, the same sampling fraction will be used. Whilst this may often be the case, an added advantage of stratified sampling is that it allows different sampling fractions to be used in each stratum. Such variable fraction sampling may be desirable in three distinct situations.

First, as will be shown later, the accuracy of results obtained from a sample depends on the absolute size of the sample, not on the fraction of the population included in the sample. In some populations, stratification on the basis of a particular parameter may result in a highly variable number of sampling units in each stratum. If a constant sampling fraction were used for each stratum, one
would obtain highly variable sample sizes in each of the strata, and hence highly variable degrees of accuracy in each stratum, all other things being equal. To obtain equal accuracy in each of the strata, it would be necessary to use different sampling fractions in each stratum so that approximately equal sample sizes were obtained for each stratum.

The second factor which affects the accuracy of a parameter estimate obtained from a sample is the variability of that parameter within the population; higher variability parameters require higher sample sizes for a specified degree of accuracy. If the sampling units within different strata exhibit different degrees of variability with respect to a parameter of interest, then it would be necessary to use high sampling fractions for strata with high variability if equal degrees of accuracy were to be obtained for each stratum.

The third reason for choosing variable fraction sampling is more pragmatic than theoretical. It may be that the costs of sampling and/or collecting the data may vary across the different strata. In such a case, a trade-off is necessary between the anticipated costs of reduced accuracy and the known costs of obtaining the data. It may therefore be desirable to reduce the sampling fraction in those strata where data is more expensive to obtain.

Therefore, while the sample shown in Figure 4.5 is more representative of the population, the sample shown in Figure 4.4 may be more efficient in reducing the total sampling error in the sample of ten. The entire question of tailoring strata sampling fractions in accordance with the above principles is related to the idea of sampling optimisation, whereby the efficiency of the data collection procedure is maximised with respect to value-for-money.

For the moment, however, two drawbacks with variable sampling fraction methods should be noted. First, the method may require far greater prior information about the population, including the size of each strata, the variability of the specified parameter within each strata, and the cost of data collection within each strata. Second, because sampling units in each strata no longer have the same chance of selection (because some strata are being deliberately oversampled), one of the two basic conditions for random sampling has now been violated. Therefore, the raw data obtained is no longer a random sample representation of the entire population. It will be necessary to assign weightings to each of the strata samples during analysis to generate population estimates which are truly representative of the population. The added complications during design and analysis of the survey generally mean that variable fraction sampling is considered only for relatively large surveys where there is the opportunity for the extra costs involved to be recouped by large savings due to increased survey efficiency.
4.4.4 Multi-Stage Sampling

In simple random sampling, the first stage in the process is to enumerate (give names or numbers to) the entire population. While this may be feasible for small populations, it is clearly more difficult with larger populations. For example, identifying every individual in a large city or a nation is clearly a non-trivial task. In such circumstances, another variation of random sampling is called for. Multi-stage sampling is a random sampling technique which is based on the process of selecting a sample in two or more successive, contingent stages. Consider, for example, a multi-stage survey of travel patterns for an entire nation. Within an Australian context, the process may proceed in five stages as follows:

(a) First-stage: divide nation into states and sample from total population of states.

(b) Second-stage: divide selected states into Local Government Areas and sample from these Local Government Areas within each selected state.

(c) Third-stage: divide selected Local Government Areas into Census Collectors' Districts and sample Census Collectors' Districts.

(d) Fourth-stage: divide selected Census Collectors' Districts into households and sample households.

(e) Fifth-stage: divide selected households into individuals and sample individuals.

At the end of this process we have a random sample of individuals from the nation (i.e. every individual had an equal chance of being selected at the start of the process) provided that appropriate sampling procedures are used at each of the stages. Thus at the first three stages, it would be necessary to sample states, Local Government Areas and Census Collectors' Districts by means of a selection procedure with probabilities proportional to size (PPS) if all individuals are to have an equal probability of selection. Thus, larger population states would have a higher probability of selection at the initial stage. The PPS sampling procedure can be easily applied to the initial three stages because the population within each state, Local Government Area and Census Collectors' District would generally be known in advance of sampling (from other sources such as National Census statistics).

At the fourth stage, however, a problem arises because, without detailed knowledge of the size of each household in the selected Census Collectors' Districts, it would not be possible to use PPS sampling at this stage. Such detailed knowledge about individual household structure would generally be unavailable.
Without this information, it can easily be seen that if one individual is to be selected from each selected household then an individual in a small household has a higher probability of selection than an individual in a large household. To correct for this it may be necessary to place households in strata in the field by means of filter questions at the start of the interview. The number of households in each stratum would be directly proportional to the household size. When each stratum is filled, no further questions would be asked of that household. Thus if \( x \) interviews were conducted in single person households, then \( 2x \) interviews should be conducted in two-person households etc., such that each individual has an equal chance of selection, irrespective of household size. Alternatively, households could be selected randomly as if they were all of equal size and then adjustments could be made to the survey results by means of appropriate weighting factors to reflect the distribution of household sizes found in the population.

The fifth stage also requires care in sampling. The interview should not be conducted with whoever opens the door or with anyone who is simply willing to be interviewed. Rather, if individuals are the unit of investigation, random sampling should be performed across all members of the household who are members of the population under investigation (perhaps, for example, there is an age-limit on members of the population). This random sampling may be formalised by printing, on the interview form, instructions to the interviewer for selection of a household member depending on the size of the household. These sampling instructions would be varied randomly, or systematically, from form to form to ensure the desired distribution of household members was obtained in the sample (see Kish, 1965). Examples of such selection grids are provided by Stopher (1985a) for selection of adults aged 18 or over, and are reproduced in Figure 4.6. The interviewer uses each grid on alternating occasions (odd and even grids) and then, depending on the answers to filter questions about the number of adults aged 18 or over and the number of males aged 18 or over, asks to speak with a specified member of the household as indicated in the appropriate selection grid.
Sampling Procedures

ODD

<table>
<thead>
<tr>
<th>Males 18 +</th>
<th>Adults 18 and Over</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>THE WOMAN</td>
</tr>
<tr>
<td>1</td>
<td>THE MAN</td>
</tr>
<tr>
<td>2</td>
<td>THE OLDER MAN</td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Even

<table>
<thead>
<tr>
<th>Males 18 +</th>
<th>Adults 18 and Over</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>THE WOMAN</td>
</tr>
<tr>
<td>1</td>
<td>THE MAN</td>
</tr>
<tr>
<td>2</td>
<td>THE OLDER MAN</td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.6 Examples of Respondent Selection Grids
(Source: Stopher, 1985a)

Whilst multi-stage sampling may appear to be somewhat complicated from the above description, its major advantage over simple random sampling lies, in fact, in its convenience and economy, especially for surveys of large populations. Thus, in multi-stage sampling, it is not necessary to enumerate all the sampling units in the population. At each stage, the only sampling units which need to be listed are those which belong to the higher level sampling units selected in the previous stage. Thus the expensive and time-consuming compilation of a complete sampling frame list is avoided.

The disadvantage of multi-stage sampling is that the level of accuracy of parameter estimates for a given sample size tends to be less than if a simple random sample had been collected (see Section 4.7 for more details). However, this reduction in accuracy needs to be traded off against the reduction in costs. In many cases, an increase in sample size for multi-stage samples can be paid for by the savings accrued in not having to prepare a full sampling frame.

It should also be noted that at each stage in the multi-stage sampling process, different sampling methods can be applied. Thus stratified sampling and variable fraction sampling can be applied to meet certain objectives. For example, if
Chapter 4

certain states must be represented in the final sample (perhaps for political or other reasons outside the scope of the survey) then these states can be segregated into strata by themselves and sampled with certainty. The higher probability of selection of the state must, however, be compensated for by lower probabilities of selection at later stages in the process, such that individuals in all states have an equal probability of selection. This latter criteria is, in fact, the guiding light of multi-stage sampling; virtually anything is allowable at each stage provided that individuals (or whatever the last-stage sampling element happens to be) have an equal chance of selection after all stages have been completed.

Multi-stage sampling can also be used in the design of on-board transit surveys. In such surveys, the sampling unit is the transit passenger, but it would be impractical to have a sampling frame based on the names of all transit passengers. Rather, the sample of transit passengers can be drawn in a four stage process, where each stage takes account of a different dimension in the transit passenger population (see Stopher, 1985b; Fielding, 1985):

Stage 1: Geographically-stratified sampling of routes

Stage 2: Sampling of vehicles from the selected routes

Stage 3: Time-stratified sampling of runs on selected vehicles

Stage 4: Surveying of all passengers on selected runs

4.4.5 Cluster Sampling

Cluster sampling is a variation of multi-stage sampling. In this method, the total population is first divided into clusters of sampling units, usually on a geographic basis. These clusters are then sampled randomly and the units within the cluster are either selected in total or else sampled at a very high rate. Like multi-stage sampling, cluster sampling can be much more economical than simple random sampling both in drawing the sample and in conducting the survey. For example, interviewers' travel costs can be reduced substantially by the use of cluster sampling. Also, if interviews are being conducted in a small number of relatively well-defined areas, it is easier to maintain a higher degree of quality control on the conduct of the interviews.

Like multi-stage sampling, the main problem with cluster sampling is that sampling error will tend to be increased for any given sample size compared to simple random sampling. This is even more of a problem for cluster sampling than for multi-stage sampling. The effect of clustered sampling on sampling error will depend on the degree of similarity between the units in the cluster and those in the total population (see Section 4.7). At one extreme, if all units in a cluster
were identical to each other, but totally different from units outside the cluster, then each cluster could equally well be described by just one unit within each cluster. The other units in the cluster would add no new information. In such a case, the effective sample size would be the number of clusters, not the number of sampling units in all the clusters. At the other extreme, if sampling units within a cluster showed equal dissimilarity to other units in that cluster and to units in the total population, then cluster sampling would result in the same distribution of sampling units as would simple random sampling. In this case, the effective sample size for cluster sampling would be equal to the number of sampling units in all the clusters. Most sampling situations fall between these extremes and hence the effective sample size will be somewhere between the number of clusters and the total number of sampling units in the clusters. For any one survey, the effective sample size will depend on the definition of the clusters and the parameter to be estimated. The homogeneity of different parameters within different clusters can vary substantially. The art of cluster definition is to find economical clusters which maintain heterogeneity in the parameters to be estimated.

4.4.6 Systematic Sampling

When random sampling is being performed in conjunction with a sampling frame list, it is frequently more convenient to use a technique called systematic sampling rather than rely on the use of random numbers to draw a sample. Systematic sampling is a method of selecting units from a list through the application of a selection interval, I, such that every Ith unit on the list, following a random start, is included in the sample. The selection interval is simply derived as the inverse of the desired sampling fraction. For example, Figure 4.7 shows a systematic sample drawn from our population of 100 sampling units, where 04 has been randomly drawn as the starting number.

Systematic sampling is much simpler than truly random selection and, once the sampling frame list has been prepared, it can be carried out by inexperienced clerical staff. The major task in systematic sampling lies in the preparation of an appropriate sampling frame list. An appropriately structured list can result in a systematic sampling procedure which automatically performs stratification with respect to a number of variables. For example, it can be seen in Figure 4.7 that the systematic sample has also resulted in a stratified sample because of the way in which the sample frame list was constructed in a stratified fashion.
Whilst the ordering of the sampling frame list can be used beneficially as described above, it can also be a source of problems which cause the systematic sample to deviate substantially from a truly random sample. Three aspects of list structure are of particular concern in this respect. First, and perhaps most importantly, care should be taken to ensure that the lists do not exhibit a periodicity with respect to the parameter being measured, especially where the periodicity corresponds to the selection interval, I. As an extreme example, consider a household survey to determine traffic noise nuisance effects in an inner urban area. With repetitious grid-street layouts, which predominate in such areas, and the existence of uniform house block sizes, it is likely that the number of houses between intersections will be constant. If the selection interval happens to be equal to this number, then the sampled houses will all be in the same position with respect to intersections (e.g. perhaps all corner blocks). The effect of this on the resultant measurements of perceived traffic noise could be quite serious. Whilst this is an extreme example, there are many instances where periodicity does occur in lists. Ways of overcoming periodicity include choosing a selection interval which is neither a multiple nor fraction of the periodicity, or else dividing the list into several segments and then choosing a new starting point at the start of each segment.

The second problem with list ordering is that, when using systematic sampling, only certain combinations of sampling units can be sampled. Specifically, only \( I \) combinations of sampling units can be sampled. Whilst this does not prevent each unit from having an equal probability of selection, it does violate the other condition of random sampling, namely that each unit be chosen independently. Where the ordering of the list is completely random in itself, this will not be a
Sampling Procedures

problem. However, few lists are completely random. An alphabetical list, such as an electoral roll, has people of the same surname (often husband and wife) next to each other on the list. If one person is included in the sample then the next person cannot be in the sample. Thus husbands and wives will rarely be both selected by systematic sampling. Similarly, where systematic sampling is based on house order in a street, neighbours in adjacent households cannot both be included in the sample. Thus to the extent to which there is serial correlation between neighbours on a list with respect to the parameter to be estimated, the resultant sample will not be truly random. This non-randomness will have a consequent effect on the sampling error for the sample. In general, as the similarity between neighbours on the list increases, the sampling error estimated from the systematic sample will be a greater overestimate of the sampling error which would have been achieved from an equal size random sample.

The third problem with systematic sampling, especially when the selection interval is large, occurs when there is a trend in the parameter being measured within the list. The estimated average value of the parameter will then depend on the point at which sampling commences within the list. The degree to which this is a problem will depend on how much the parameter varies between the first and last units within each selection interval.

Despite the potential problems with systematic sampling, it is a very useful and simple sampling method which can be used whenever sampling frame lists are available. Care should simply be taken to ensure that whatever ordering is present in the list is used to advantage rather than disadvantage.

4.4.7 Non-Random Sampling Methods

In addition to the above methods, which are based to a greater or lesser extent on random probabilistic sampling, there are a number of other sampling methods which are not based on random sampling. The two principal methods are quota sampling and expert sampling.

Quota sampling, as the name suggests, is based on the interviewer obtaining responses from a specified number of respondents. The quota may be stratified into various groups, within each of which a quota of responses must be obtained. This method, for example, is often used when interviewing passengers disembarking from aircraft or other transport modes and for many types of street interviews where passers-by are stopped and asked questions. The major problem with quota sampling is not that quotas are used for each sub-group (after all, this is the basis of stratified sampling), but that the interviewer is doing the sampling in the field and this sampling procedure may be far from random, unless strictly controlled. Left to themselves, interviewers will generally pick respondents from whom they feel they will most readily obtain a response. Thus passers-by who appear more willing to cooperate, are not in a hurry, and are of a
social class comparable to the interviewer will more likely be interviewed. In a household survey, households which are closer to the interviewer's residence (and hence require less travel to reach), households whose members are more often at home, and households without barking dogs are more likely to be interviewed. Such preferential selection can often cause gross biases in the parameters to be estimated in the survey.

Expert sampling, on the other hand, takes the task of sampling away from the interviewer and places it in the hands of an "expert" in the field of study being addressed by the survey. The validity of the sample chosen then relies squarely on the judgement of the expert. While such expert sampling may well be appropriate in the development of hypotheses and in exploratory studies, it does not provide a basis for the reliable estimation of parameter values since it has been repeatedly shown that people, no matter how expert they are in a particular field of study, are not particularly skilled at deliberately selecting random samples. A more appropriate role for the expert in sample surveys is in the definition of the survey population and strata within this population, leaving the task of selecting sampling units from these strata to the aforementioned random sampling methods.

4.5 SAMPLING ERROR AND SAMPLING BIAS

Despite all our best intentions in sample design, the parameter estimates made from sample survey data will always be just that: estimates. There are two distinct types of error which occur in survey sampling and which, combined, contribute to measurement error in sampled data.

The first of these errors is termed sampling error, and is the error which arises simply because we are dealing with a sample and not with the total population. No matter how well designed our sample is, sampling error will always be present due to chance occurrences. However, sampling error should not affect the expected values of parameter averages; it merely affects the variability around these averages and determines the confidence which one can place in the average values. Sampling error is primarily a function of the sample size and the inherent variability of the parameter under investigation. More will be said about sampling error when techniques for the determination of sample size are discussed.

The second type of error in data measurement is termed sampling bias. It is a completely different concept from sampling error and arises because of mistakes made in choosing the sampling frame, the sampling technique, or in many other aspects of the sample survey. Sampling bias is different from sampling error in two major respects. First, whilst sampling error only affects the variability around the estimated parameter average, sampling bias affects the value of the
average itself and hence is a more severe distortion of the sample survey results. Second, while sampling error can never be eliminated and can only be minimised by increasing the sample size, sampling bias can be virtually eliminated by careful attention to various aspects of sample survey design. Small sampling error results in precise estimates while small sampling bias results in accurate estimates.

The difference between these two sources of error is sometimes confused, with attention being paid to reducing sampling error while relatively little attention is paid to minimising sampling bias. In an attempt to underscore the difference between the two concepts, consider an analogy with rifle marksmanship as illustrated by the targets shown in Figure 4.8.

**Figure 4.8 The Distinction between Accuracy and Precision**

These targets illustrate four essentially different ways in which rifle shooters may hit the target. The top left target shows a marksman who consistently hits the bullseye. The bottom left shows one who centres his shots around the bullseye but also tends to spray his shots; he seems to be able to aim at the right point but tends to suffer from slight movement of the rifle at the last moment so that his shots are not consistent. The top right target shows the results of a marksman who consistently misses the bullseye; he holds the rifle rock-steady but unfortunately he is aiming at the wrong point on the target, maybe because the telescopic sights on the rifle are out of adjustment. The bottom right shows a shooter who appears to be aiming at the wrong point, but because he also suffers from nervous jitters he sometimes hits the bullseye even though he is not aiming at it. These four situations may be categorised in terms of the precision and the
accuracy of the shots; precise shooters always hit the same spot, while accurate shooters aim at the right point on the target.

It is fairly clear which of the four shooters would be regarded as the best; the top left shooter shoots with both accuracy and precision in that he consistently hits the bullseye. It is also probably safe to say that the bottom left shooter is the second best in that he is at least on target (on average). However, it is not quite so clear which of the remaining two are the worst. Is it better to be consistently off-target, or inconsistently off-target (where at least you have some chance of hitting the bullseye)? This judgement of the quality of marksmanship is made more difficult when the bullseyes are removed to leave only the holes left by the rifle shots, as shown in Figure 4.9. In this case, it is difficult to say whether the top left or the top right group of shots came from the better marksman. Indeed, one may argue that both groups are equally good. In the absence of any knowledge about where the marksmen were aiming, one is more readily swayed by the precision of the shots in judging the quality of the shooter. Indeed, the top right group of shots is now vying for the best group of shots, whereas in Figure 4.8 it was vying for being the worst group of shots.

![Figure 4.9 The Confusion between Accuracy and Precision](image)

The above description of the marksman can be applied, by analogy, to the design and use of sample surveys. A precise survey is one which displays repeatability; that is, if administered on repeated occasions under similar conditions it will yield the same answers (irrespective of whether the answers are right or wrong). On the other hand, an accurate survey is one which displays validity, in that the
survey is aimed at a correct sample of the correct target population. The precision of a sample survey can be increased by increasing the sample size so as to reduce the possibility of unobserved members of the population having, by pure chance, characteristics which are different to those observed. The accuracy of a sample survey can be increased by ensuring that, first, the sampling frame does not systematically eliminate some members of the population and, second, that the sample is obtained from the sampling frame in a truly random fashion.

Much attention is often paid to reducing sampling error (i.e. increasing precision) by means of elaborate sampling designs and large sample sizes. Relatively little attention, however, is generally paid to increasing accuracy by means of reducing sampling bias to ensure that the questions are being asked of the right people. We are often guilty of "Type-III Errors", described by Armstrong (1979) as "good solutions to the wrong problems". By simply increasing sample sizes, and not paying attention to the quality of the sample, we can always ensure that we will be able to spend enough money to get precisely wrong answers! Indeed, by analogy with Figure 4.9, when we do not know much about the true population we are trying to survey, then we assume that a precise answer is better than an imprecise answer, irrespective of whether it is accurate or not.

In an attempt to improve the accuracy of sample surveys, we therefore need to be more aware of the likely sources of sampling bias (the issue of increasing accuracy by improving survey instrument validity will be discussed in Chapter 5). Some common sources of sampling bias include:

(a) Deviations from the principles of random sampling including: the deliberate selection of a "representative" sample which results in too many observations at the extremes of the population distribution; the deliberate selection of an "average" sample which results in too many observations near the average value and not enough at the extremes; the initial selection of a random sample from which the investigator discards some values because they are not considered to be random.

(b) Use of a sampling frame whose characteristics are correlated with properties of the subject of the survey, e.g. using a telephone interview survey to obtain car ownership rates. Richardson (1985) shows the effect of sampling bias in telephone surveys, and suggests means of correcting for this bias.

(c) Substitution sampling, where the interviewer in the field changes the specified sample because of difficulties experienced in obtaining the originally selected sample (e.g. barking dog in front garden, lack of time to reach next specified household, non-response from selected household).
(d) Failure to cover the selected sample. This can result in bias if those sampling units left unsurveyed are atypical of the total sample. For example, in a household survey of transit usage, where the interviewer uses transit to reach the survey area, lack of time may result in those households closest to the transit line being surveyed whilst those farther away are left unsurveyed.

(e) Pressures placed on interviewers by the method of payment adopted. If payment is by interview completed, then there is an incentive to the interviewer to complete as many interviews as possible in the shortest possible time. This increases pressures for substitution sampling and also encourages the interviewer to complete each interview as quickly as possible. For travel surveys, this will result in fewer trips being reported by respondents during the interview. If payment is by the hour, then the reverse incentives are present. While more desirable than hurried interviews, the expenditure of large amounts of time on each completed interview inevitably means that interviewer productivity must fall.

(f) Falsification of data by interviewer when the interview has not even been conducted. Where such falsification is caused by difficulties in contacting the respondent, then sampling bias may be introduced.

(g) Non-response effects. The bias introduced by non-response varies with the type of survey method used. Thus for mail-back questionnaires, non-response is generally an indication of a low level of interest in the subject of the survey by the non-respondent. For self-completion travel surveys, this often means that non-respondents travel less than respondents, at least for the purpose and/or mode which is the subject of the survey. Empirical verification of this trend can be found in Brög and Meyburg (1981) and Richardson and Ampt (1994). On the other hand, for personal interview surveys, non-response and in particular non-contact is generally more of a problem for those respondents who are more mobile, and hence less often at home to be contacted by the interviewer (see Brög and Meyburg, 1982). Thus for self-completion mail-back questionnaire surveys, non-response will bias travel estimates upwards whereas for personal interview surveys, non-response will bias travel estimates downwards.

Whilst all the above sources of bias are potentially serious, there are a number of safeguards against the introduction of sampling bias including:

(a) Use a random sample selection process and adopt, in full, the sample generated by the process.
Sampling Procedures

(b) Design the survey procedure and field administration such that there is no opportunity or need for interviewers to perform "in-field" sampling.

(c) Perform random call-backs on some respondents who have been interviewed to check on the accuracy of the data obtained and the adherence of the interviewers to the specified random sample.

(d) Perform cross-checks with other secondary sources of data to check on the representativeness of the respondents.

(e) Make every attempt to increase response rates by means of reminder letters for self-completion, mail-back surveys and repeated call-backs for personal interview surveys.

(f) Attempt to gain as much information about the characteristics of the entire sample (i.e. in terms of Figure 4.9, try to know the target at which you are shooting) by identifying the characteristics of non-respondents so that adjustments can be made to the survey results to account for the degree of non-response (see Brög and Meyburg, 1980; Meyburg and Brög, 1981).

One final point with respect to sampling bias is that it will vary with the type of survey method being used and with the parameters which the sample survey seeks to estimate. Only careful consideration of the individual circumstances will determine whether significant bias is likely to exist in the survey results. In all cases, however, it is only possible to correct for sampling bias if sufficient effort has been made to gather information about the entire sample and the entire population which the sample purports to represent.

4.6 SAMPLE SIZE CALCULATIONS

Of all the questions concerned with sample design, the one most frequently addressed is that of required sample size. As mentioned earlier, one way of reducing sampling error is to increase sample size; the question remains, however, as to how much one should increase sample size in order to obtain an acceptable degree of sampling error. This section will attempt to provide some guidance on this matter, particularly for the case of simple random sampling. Estimation of required sample sizes for other sampling methods rapidly becomes more complex (see, for example, Kish 1965) and will not be covered in detail in these notes.

In discussing required sample sizes for simple random samples, it is emphasised that guidance only can be given. Much to the chagrin of many investigators, no firm rules can be given for sample size calculations for use in all circumstances. Whilst the calculations are based on precise statistical formulae, several inputs to
Chapter 4

the formulae are relatively uncertain and subjective and must be provided by the investigator after careful consideration of the problem at hand. Importantly, it is often difficult for the survey designer to convey the nature of sample size calculations to clients, who are most often ignorant of the statistical concepts involved. This chapter will attempt to provide some assistance in conveying these concepts to clients with little or no statistical background.

The essence of sample size calculations is one of trade-offs. Too large a sample means that the survey will be too costly for the stated objectives and the associated degree of precision required. Too small a sample will mean that results will be subject to a large degree of variability and this may mean that decisions cannot reliably be based on the survey results. In such a situation, the entire survey effort may have been wasted. Somewhere between these two extremes there exists a sample size which is most cost-effective for the stated survey objectives.

In the context of survey objectives, it is useful to distinguish between two broad purposes for which survey data may be collected:

(a) The main purpose is often to estimate certain population parameters, e.g. average person trip rates, car ownership, mode split, etc. In such cases, a sample statistic is used to estimate the required population parameter. However, because all sample statistics are subject to sampling error, it is also necessary to include an estimate of the precision which can be attached to the sample statistic. This level of precision will be affected, \textit{inter alia}, by sample size.

(b) A second purpose of a survey (or surveys) may be to test a statistical hypothesis concerning some of the population parameters, e.g. are there significant differences in trip rates in different areas, or has mode use risen following introduction of a new transport service? To test such hypotheses, it is necessary to compare two sample statistics (each being an estimate of a population parameter under different conditions), each of which has a degree of sampling error associated with it. The tests are performed using statistical significance tests where the power of the test is a function of the sample size of the survey(s).

Whilst the use of sample survey data to fulfil each of these objectives requires different statistical techniques, they are linked by a common usage of the concept of standard error. This concept will now be described, initially with reference to the former objective of sample surveys - that of obtaining population parameter estimates.
4.6.1 Sample Sizes for Population Parameter Estimates

The determination of required sample size for the estimation of population parameters depends, as will be shown later, on three principal factors:

(a) The variability, over the population, in the parameters to be measured;
(b) The degree of precision required for each of the parameter estimates;
(c) The population size.

Of these three factors, the first two are by far the most important. This may at first seem surprising because, to many, it seems intuitive that larger samples will be required from larger populations to maintain accuracy of parameter estimates. This intuitive feeling is often summarised by statements which imply that sample sizes are always expressed in percentage form e.g. “a 10% sample should be big enough”. However, as shall be seen later, except for very small populations, the population size does not significantly affect the required sample size: it is the absolute sample size which is important.

This finding is so important that it is worth repeating. Except in surveys of very small populations, it is the number of observations in the sample, rather than the sample size as a percentage of the population, which determines the precision of the sample estimates. A sample of 200 people from a population of 10 million is just as precise as a sample of 200 people from a population of ten thousand.

4.6.1.1 Sample Sizes for Continuous Variables

Before proceeding too far in the determination of required sample sizes, it will be useful if we review one statistical theory which is at the very heart of sample size estimation. This theorem is called the Central Limit Theorem. This theorem states that estimates of the mean of a sample tend to become normally distributed as the sample size n increases. This normality of sample means applies irrespective of the distribution of the population from which the samples are drawn provided that the sample size is of reasonable size (n!>!30). For small sample sizes, the theorem still applies provided that the original population distribution is approximately bell-shaped.

This theorem often causes confusion but it is so basic to sampling theory that it must be understood before any progress can be made in understanding sample size determination. So let’s restate it. Assume that we have, for example, a continuous variable (x) whose variability among sampling units in the population may be described by the distribution shown in Fig.4.10. Such a variable may be, for example, the income of people in our population. The distribution may be of any form (for example, negatively skewed as shown). Assume that the population is of size N and the population distribution has some true mean value [μ] and a true standard deviation [σ].
Chapter 4

If we were to now draw a sample of size $n$ from this population, we could calculate the mean income for that sample as $m_1$ and the standard deviation for that sample as $S_1$. We could then draw a second sample of size $n$ from the total population and calculate $m_2$ and $S_2$. This could be repeated for a third sample to obtain $m_3$ and $S_3$, a fourth sample to get $m_4$ and $S_4$ etc. Having drawn $x$ samples, we could then construct a frequency distribution of the values $m_1, m_2, m_3, \ldots, m_x$. The Central Limit Theorem states that this distribution, as shown in Fig. 4.11, is normally distributed with mean $m$ (which is an unbiased estimate of the population mean $\mu$).

![Figure 4.10 Distribution of the Parameter in the Population](image-url)
The standard deviation of this distribution of sample means, which is referred to as the standard error of the mean (s.e.(m)), is given by:

\[ s.e.(m) = \sqrt{\frac{N-n}{N}} \frac{\sigma^2}{n} \]  

(4.1)

The above discussion has been based on taking repeated samples from a population. Generally, however, this is not possible and therefore it is necessary to make some estimates based on a single sample of size n. In such a situation our best estimate of \( \mu \) is given by \( m_1 \) and similarly the best estimate of \( \sigma \) is given by \( S_1 \) (hereafter referred to as S). Therefore on the basis of a single sample, we can estimate what the standard error of the mean would have been, if repeated samples had been drawn, as:

\[ s.e.(m) = \sqrt{\frac{N-n}{N}} \frac{S^2}{n} \]  

(4.2)

As noted earlier, the standard error is a function of three variables; the variability of the parameter in the population (represented by the standard deviation \( \sigma \)), the sample size (n) and the population size (N). However for large populations and small sample sizes (which is often the case in transport surveys), the finite population correction factor \((N-n)/N \) is very close to unity. In such situations, the equation for standard error of the mean may be reduced to the more familiar form of:
s.e.(m) = \sqrt{\frac{S^2}{n'}} = \frac{S}{\sqrt{n'}} \tag{4.3}

This equation highlights a most important aspect of sample size determination. That is, as sample size increases, the standard error of the mean will decrease but only in proportional to the square root of the sample size. Thus, quadrupling the sample size will only halve the standard error of the mean. Increasing sample size is therefore a clear case of diminishing marginal returns with respect to decreases in standard error of the mean.

Reference to the properties of the normal distribution, dictated by the Central Limit Theorem, also enables an estimate to be made of the accuracy of the sample mean \(m\) as a reflection of the true population mean \(\mu\). Such estimates are calculated using the concept of confidence limits associated with the normal distribution. Thus, some 95% of all sample means (from samples of size \(n\)) would lie within two standard errors on either side of the true mean, so that there is a probability of only about one in twenty that the deviation between a sample mean and the true mean will exceed a value greater than twice the standard error.

Given the foregoing discussion, the required sample size can be estimated by solving for \(n\) in equation (4.2). This is most easily done in two stages by first solving for \(n\) in equation (4.3) such that:

\[
n' = \frac{S^2}{(\text{s.e.}(m))^2} \tag{4.4}\n\]

and then correcting for the finite population effect, if necessary, such that:

\[
n = \frac{n'}{1 + \frac{1}{(n'/N)}} \tag{4.5}\n\]

Whilst the above procedure for the determination of sample size looks relatively straightforward and objective, there are two major problems in the application of the method; the estimation of the population standard deviation \(\sigma\) and the selection of an acceptable standard error of the mean \(\text{s.e.}(m)\). The problem with the estimation of the standard deviation is that this is one of the statistics which will be calculated after the survey has been conducted, and yet we are required to estimate it before we conduct the survey in order to calculate the sample size. It is therefore necessary to derive an estimate of the standard deviation from other sources. Three major sources suggest themselves:

(a) Previous surveys of the same, or a similar, population may provide an estimate of the standard deviation of the parameter in question. Due
allowance should be made for any differences in the sampling method used in the previous and the current survey.

(b) There may be some theoretical foundations on which to base an estimate of the standard deviation. This technique was used, for example, in the Australian National Travel Survey (Aplin and Flaherty, 1976).

(c) Where little previous information exists about the population, it may be necessary to conduct a pilot survey to obtain information needed to design the main survey. A problem with this method, however, is that often time and money resources do not permit the conduct of large enough pilot survey to enable serviceable estimates of the standard deviation to be obtained. In such circumstances, the standard deviation estimates may be more misleading than informative.

Sometimes the estimated sample size can be adjusted during the course of the main survey to overcome any uncertainty in the initial estimate of the standard deviation. Thus using the initial standard deviation estimate, a sample of minimum size could be collected. The standard deviation in this sample could then be computed and compared with the initial estimate. If the standard deviation is larger than estimated, thus indicating that a larger sample should be collected, then a supplementary sample could be collected to augment the initial sample. Whilst this two-step procedure sounds attractive in being able to lessen the demands of accurate estimation of standard deviation, it is only feasible in certain circumstances. Thus the conduct of the survey must be spread over a reasonable time period so that coding, editing and analysis of an initial sample can be completed in time for the supplementary sample data collection to follow on reasonably soon after the collection of the initial sample. Where strict time limitations are placed on the conduct of a survey, supplementary samples may not be feasible.

The second problem in the estimation of sample sizes using the above equations is the specification of an acceptable standard error of the mean. This task basically expresses how confident we wish to be about using the sample mean as an estimate of the true population mean. The specification of a standard error is rarely performed per se; rather it is usual to specify confidence limits of a specified size around the mean and at a certain level of confidence. For example, you may specify that you wish, with C% confidence, to obtain a sample mean which is within a specified range, either relative or absolute, of the population mean. Specified in this way, two judgements must be made in order to calculate the acceptable standard error.

First, a level of confidence must be chosen for the confidence limits. Basically, the level of confidence expresses how frequently the client is prepared to be wrong in accepting the sample mean as a measure of the true mean. For example, if a
Chapter 4

95% level of confidence is used, then implicitly it is being stated that the client is prepared to be wrong on 5% of occasions. If such a risk is deemed to be unacceptably large, then higher confidence limits (such as 99%) can be used. Higher confidence limits will, however, require larger sample sizes.

The specification of levels of confidence is a difficult task for the client and the survey designer. Not understanding the subtleties of sampling, most clients are unwilling to accept that the survey, for which they are paying, will come up with anything but the correct answer. On the other hand, the survey designer should know that nothing but a full population survey will produce results that are absolutely correct (assuming that everything else about the survey is acceptable). The task of the survey designer, therefore, is to get some indication from the client of what they think is an acceptable level of confidence in the results. By convention, 95% levels of confidence are often assumed for sample surveys in transport. This means that if repeated sample surveys were to be conducted on this topic with a sample of this size, then 5% of the estimates of the mean would lie outside the range of the population mean, plus or minus two standard errors.

Second, it is necessary to specify the confidence limits around the mean, either in absolute or relative terms. If relative measures are used (i.e. the confidence limit is a proportion of the mean) then this requires that an estimate of the mean be available so that an absolute measure of the confidence limit can be calculated. If the parameter being estimated is of some importance then smaller confidence limits can be specified but again this will result in higher sample sizes being necessary. The size of the confidence limits will depend on the use to which the results of the survey are to be put.

The important point to note about acceptable standard errors is that the specification of both the confidence limits and the level of confidence is relatively subjective. More important parameters can be assigned smaller confidence limits and/or higher levels of confidence. Each of these actions will result in a smaller acceptable standard error and thus a higher required sample size. The decision, however, lies in the hands of the sample designer in liaison with the client; is accuracy of a parameter estimate sufficiently important to warrant the higher costs involved in a larger sample size?

To illustrate the points outlined above, consider, as an example, a survey of a population of 1000 households in which an estimate of average household income is required such that there is a 95% probability that the sampling error will be no more than 5% of the sample mean. From a pilot survey of this population, it has been found that with a sample size of 30 the mean income was $24,000 and the standard deviation was $5,000.

The acceptable standard error can be calculated from the specified confidence limits and level of confidence. From a table of unit normal distribution values (see
Sampling Procedures

Appendix A), a 95% level of confidence corresponds to a value of 1.96 times the standard error. That is, there is a 95% probability that the error of the mean estimate will be no more than 1.96 times the standard error. However, in our case we want this error to be no more than 5% of our estimated mean. Using the pilot survey mean value as an initial estimate, the confidence limit will therefore be equal to $1,200 (= 0.05 \times 24,000)$. The acceptable standard error is then given by:

\[
\text{s.e.}(\bar{X}) = \frac{\text{confidence limit}}{z} = \frac{1200}{1.96} = 612
\] (4.6)

The sample size, for an infinite population, is then given by:

\[
n' = \frac{s^2}{\text{s.e.}(m)^2} = \frac{5000^2}{612^2} = 67
\] (4.7)

Applying the finite population correction factor, the final sample size is given by:

\[
n = \frac{n'}{1 + \frac{n'}{N}} = \frac{67}{1 + \frac{67}{1000}} = 63
\] (4.8)

Having collected data from these 63 households, it may be found that the estimated sample mean income has fallen to $23,200 while the sample standard deviation has increased to $6,000. In such a case, a new estimate of the required sample size may be given by:

\[
n' = \frac{6000^2}{592^2} = 103
\] (4.9)

and hence

\[
n = \frac{103}{1 + \frac{103}{1000}} = 93
\] (4.10)

If it is convenient, and if the extra expense is deemed worthwhile, then an extra 30 households should be sampled and surveyed. It should be noted that without the extra households in the survey, the sampling error of the mean, at a 95% confidence level, is equal to 6.2% of the sample mean. The question must be asked as to whether the expense of the extra surveys is warranted by the reduction in sampling error from 6.2% to 5.0% of the mean.

4.6.1.2 Sample Sizes for Discrete Variables

The preceding discussion on sample size estimation has concentrated on the collection of data on continuous variables. In many cases, however, we may wish...
to collect data on discrete variables which are characterised by their presence or absence in a household. An example of a discrete variable would be car ownership, where a household either owns a car or it does not. In such a case, the Central Limit Theorem still applies but this time it is applied to the proportion of a sample possessing a certain characteristic e.g. owning a car. The standard error for estimation of a proportion $p$ is given by:

$$s.e.(p) = \sqrt{\frac{N-n}{n} \cdot \frac{p(1-p)}{n}}$$

(4.11)

To illustrate the use of this equation, consider our previous example. In the pilot survey, it may have been found that 20% of the households did not own a car. Assume that we wished to estimate the percentage of households not owning a car such that, at a 95% level of confidence, the sampling error was not greater than 5 percentage points. Note that when estimating percentages it is important to define the acceptable sampling error clearly; in our example, we have specified 5 percentage points. (i.e. 20% +/- 5%), and not 5 percent of the mean percentage (i.e. 20% +/- 1%). Also note that if relative error is specified, it should be clear as to whether the relative error refers to the percentage possessing a characteristic or the percentage not possessing a characteristic. Thus if relative errors are specified as 5% of the mean, does it mean 5% of the percentage not possessing a car (i.e. 5% of 20%) or 5% of the percentage possessing a car (i.e. 5% of 80%). The acceptable standard error may be calculated as:

$$s.e.(p) = \frac{5}{1.96} = 2.55 \text{ percentage points}$$

(4.12)

The required sample size may be obtained in a two-step calculation by:

$$n' = \frac{p(1-p)}{s.e.(p)^2}$$

$$= \frac{20 \times 80}{2.55^2}$$

$$= 246$$

(4.13)

Applying the finite population correction factor, the final required sample size is given by:

$$n = \frac{n'}{1 + \frac{n}{N}}$$

$$= \frac{246}{1 + \frac{246}{1000}}$$

$$= 197$$

(4.14)
Sampling Procedures

Note that in this case the finite population correction factor has a much larger effect on the required sample size because the sample size is a substantial proportion of the population.

The two calculations described above highlight a problem with sample size calculations for real-life surveys. Very few surveys seek to estimate just one parameter in the population; the normal situation is that a survey seeks to estimate a large number of parameters for a population. However, as shown above, carrying out sample size calculations separately for each parameter may result in widely varying estimates of required sample size. Thus, in our example, an initial sample size of 63 was required for income but a sample of 197 was required for (non) car-ownership. Obviously a fail-safe procedure would be to select the largest calculated sample size across all the parameters. In this way, all parameters would be estimated (at least) as precisely as desired. However, this procedure may also be very inefficient if, for example, only one parameter requires a large sample size due to its own inherently high variability across the population. The more usual procedure in selecting an overall sampling rate therefore involves a degree of compromise across the parameters. In this way, some parameters will be obtained more precisely than desired, while other parameters will be estimated with less precision than desired. If possible, the more important parameters should have their precision requirements fulfilled in preference to those of less important parameters. An overall average sample size, weighted by parameter importance, should therefore influence the final selection of a required sample size.

A final problem to be noted in the above sample size calculations is that often a survey will be seeking to obtain parameter estimates not just for the entire sample, but for many different sub-samples within the total sample. In such situations, it is necessary to know what those sub-samples will be before the survey is designed so that adequate sample sizes can be obtained for each of the sub-samples to be analysed later.

4.6.1.3 Explaining Sample Sizes to Clients

Because of the complexities and subtleties involved in sample design, it is often difficult for a survey designer to explain the concepts to clients and to obtain useful information from the client to assist in the sample design. This input of information from the client is essential because, after all, it is the client who has to live with the results after the survey is completed. Therefore, every attempt should be made to tailor the survey and the sample to the needs of the client. To assist in explaining sample size calculations to clients, and in obtaining the required input from clients, the design aids outlined in this section have proven useful.
Chapter 4

The basis of these design aids is a set of two spreadsheet tables, shown in Figures 4.12 and 4.13. Figure 4.12 shows how the confidence limits around a continuous variable change as the sample size is changed, given estimates of the population mean and standard deviation, the size of the population and the level of confidence required by the client. Specification of the level of confidence automatically calculates the value of $z$ used in equation 4.6. The outputs of Figure 4.12 are the upper and lower confidence limits which could be expected from any specified sample size.

**Continuous Variable Confidence Limit Calculator**

<table>
<thead>
<tr>
<th>Population Mean</th>
<th>3.60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population S.D.</td>
<td>2.50</td>
</tr>
<tr>
<td>Population Size</td>
<td>10000</td>
</tr>
<tr>
<td>Level of Confidence</td>
<td>95%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>s.e.(m)</th>
<th>$z\times$s.e.(m)</th>
<th>Lower Limit</th>
<th>Upper Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.35</td>
<td>0.69</td>
<td>2.91</td>
<td>4.29</td>
</tr>
<tr>
<td>100</td>
<td>0.25</td>
<td>0.49</td>
<td>3.11</td>
<td>4.09</td>
</tr>
<tr>
<td>150</td>
<td>0.20</td>
<td>0.40</td>
<td>3.20</td>
<td>4.00</td>
</tr>
<tr>
<td>200</td>
<td>0.18</td>
<td>0.34</td>
<td>3.26</td>
<td>3.94</td>
</tr>
<tr>
<td>250</td>
<td>0.16</td>
<td>0.31</td>
<td>3.29</td>
<td>3.91</td>
</tr>
<tr>
<td>300</td>
<td>0.14</td>
<td>0.28</td>
<td>3.32</td>
<td>3.88</td>
</tr>
<tr>
<td>350</td>
<td>0.13</td>
<td>0.26</td>
<td>3.34</td>
<td>3.86</td>
</tr>
<tr>
<td>400</td>
<td>0.12</td>
<td>0.24</td>
<td>3.36</td>
<td>3.84</td>
</tr>
<tr>
<td>450</td>
<td>0.12</td>
<td>0.23</td>
<td>3.37</td>
<td>3.83</td>
</tr>
<tr>
<td>500</td>
<td>0.11</td>
<td>0.21</td>
<td>3.39</td>
<td>3.81</td>
</tr>
<tr>
<td>5000</td>
<td>0.03</td>
<td>0.05</td>
<td>3.55</td>
<td>3.65</td>
</tr>
</tbody>
</table>

**Figure 4.12  Confidence Limit Estimator for Continuous Variables**

For example, if the expected trip rate per person per day is 3.60 (with a standard deviation of 2.50), then with a sample of 300 persons in any one strata, we would expect that the mean trip rate would fall between 3.32 and 3.88 in 95% of samples of this size.

If the client believed that this range was too great, then they could experiment either with different sample sizes or with changing the level of confidence. The bottom line in Figure 4.12 is provided to allow the specification of any desired sample size, which may be outside the range of those provided in the table. Figure 4.12 may be re-used for any continuous variable in the survey by simply changing the population mean and standard deviation.
Sampling Procedures

A similar set of calculations is carried out in Figure 4.13 for discrete variables. In this case, however, what needs to be specified is the expected proportion in the population possessing a certain feature. For example, if the expected proportion of trips by bus is 20%, then with a sample of 300 trips in any one strata, we would expect that the mean proportion of trips by bus would lie between 16% and 24% in 95% of samples of this size.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>s.e.(m)</th>
<th>z*s.e.(m)</th>
<th>Lower Limit</th>
<th>Upper Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.06</td>
<td>0.09</td>
<td>0.11</td>
<td>0.29</td>
</tr>
<tr>
<td>100</td>
<td>0.04</td>
<td>0.07</td>
<td>0.13</td>
<td>0.27</td>
</tr>
<tr>
<td>150</td>
<td>0.03</td>
<td>0.05</td>
<td>0.15</td>
<td>0.25</td>
</tr>
<tr>
<td>200</td>
<td>0.03</td>
<td>0.05</td>
<td>0.15</td>
<td>0.25</td>
</tr>
<tr>
<td>250</td>
<td>0.03</td>
<td>0.04</td>
<td>0.16</td>
<td>0.24</td>
</tr>
<tr>
<td>300</td>
<td>0.02</td>
<td>0.04</td>
<td>0.16</td>
<td>0.24</td>
</tr>
<tr>
<td>350</td>
<td>0.02</td>
<td>0.03</td>
<td>0.17</td>
<td>0.23</td>
</tr>
<tr>
<td>400</td>
<td>0.02</td>
<td>0.03</td>
<td>0.17</td>
<td>0.23</td>
</tr>
<tr>
<td>450</td>
<td>0.02</td>
<td>0.03</td>
<td>0.17</td>
<td>0.23</td>
</tr>
<tr>
<td>500</td>
<td>0.02</td>
<td>0.03</td>
<td>0.17</td>
<td>0.23</td>
</tr>
<tr>
<td>2000</td>
<td>0.01</td>
<td>0.01</td>
<td>0.19</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Figure 4.13  Confidence Limit Estimator for Discrete Variables

Figures 4.12 and 4.13 show the effect of changing sample sizes on the confidence limits which could be expected for one variable at a time. However, as noted earlier, one of the problems in sample design for a real survey is that sample sizes must be calculated for many variables across many strata. The effects of varying sample size on the precision obtained for all variables can be summarised for the client as shown in Figures 4.14 through 4.16. These tables are constructed within a standard spreadsheet program, and are designed to be used interactively with a client to give them a feel for the implications of using various sized samples and varying number of strata (e.g. geographic regions). In Figure 4.14, the client can specify the number of strata and the sample size, the population size (in terms of number of households in the study area) and the required level of confidence (the latter item may be selected on the advice of the survey designer).
Chapter 4

Sample Size Design Parameters

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Strata =</td>
<td>12</td>
</tr>
<tr>
<td>Households in Sample per Strata</td>
<td>200</td>
</tr>
<tr>
<td>Persons in Sample per Strata</td>
<td>600</td>
</tr>
<tr>
<td>(assuming 3 persons per household)</td>
<td></td>
</tr>
<tr>
<td>Trips in Sample per Strata</td>
<td>2400</td>
</tr>
<tr>
<td>(assuming 4 trips per person)</td>
<td></td>
</tr>
<tr>
<td>Population Size =</td>
<td>60000</td>
</tr>
<tr>
<td>Level of Confidence =</td>
<td>95%</td>
</tr>
<tr>
<td>Total Households in Sample</td>
<td>2400</td>
</tr>
<tr>
<td>Total Survey Cost =</td>
<td>$144,000</td>
</tr>
<tr>
<td>(assuming $60 per responding household)</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.14 Input Screen for Sample Size Design Parameters

The survey designer can also input a unit price per responding household in Figure 4.14 to give the client an indication of the cost implications of their sample design decisions. A second input screen, shown in Figure 4.15, requires the client, or the survey designer, to specify the expected values of key variables in the population together with the expected variability of continuous variables.

The spreadsheet program calculates the standard error for each variable and, using the value of $z$ corresponding to the stated level of confidence, then calculates the upper and lower confidence limits for each key variable. These limits are then stated in simple English as shown in Figure 4.16. If one or more of these ranges are not acceptable to the client, because they think that the precision is not adequate for their purposes, then they can go back to the first input screen in Figure 4.14, change the sample size and observe the effects on the precision shown in Figure 4.16. They can also experiment with the precision obtained with different levels of stratification, by either increasing the number of strata and decreasing the sample size in each strata, or by decreasing the number of strata and increasing the sample size in each. In this way, the client and the survey designer can interactively experiment with different sample designs and observe the effects on the precision of sample estimates and the cost of the survey, thereby experiencing the nature of the trade-offs in sample design.
### Expected Population Values for Key Variables

<table>
<thead>
<tr>
<th>Household Variables</th>
<th>Proportion</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persons per Household</td>
<td></td>
<td>3.00</td>
<td>1.50</td>
</tr>
<tr>
<td>Vehicles per Household</td>
<td></td>
<td>1.50</td>
<td>0.80</td>
</tr>
<tr>
<td>Households without Vehicles</td>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trips per Household</td>
<td></td>
<td>12.00</td>
<td>5.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Person Variables</th>
<th>Proportion</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trips per Person</td>
<td></td>
<td>4.00</td>
<td>2.50</td>
</tr>
<tr>
<td>% Male</td>
<td>0.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Unemployed</td>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Personal Income ($K)</td>
<td></td>
<td>28.00</td>
<td>8.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trip Variables</th>
<th>Proportion</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Trip Length (minutes)</td>
<td></td>
<td>20.00</td>
<td>12.00</td>
</tr>
<tr>
<td>% Trips by Bus</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Trips to School</td>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average People per Vehicle</td>
<td></td>
<td>1.40</td>
<td>0.20</td>
</tr>
</tbody>
</table>

---

**Figure 4.15** Input Screen for Expected Values of Variables in Population

### Precision of Sample Estimates

**Household Variables**
- The estimated value of *Persons per Household* lies between 2.79 and 3.21
- The estimated value of *Vehicles per Household* lies between 1.39 and 1.61
- The estimated value of *Households without Vehicles* lies between 0.06 and 0.14
- The estimated value of *Trips per Household* lies between 11.31 and 12.69

**Person Variables**
- The estimated value of *Trips per Person* lies between 3.80 and 4.20
- The estimated value of *% Male* lies between 0.46 and 0.54
- The estimated value of *% Unemployed* lies between 0.08 and 0.12
- The estimated value of *Average Personal Income ($K)* lies between 27.36 and 28.64

**Trip Variables**
- The estimated value of *Average Trip Length (minutes)* lies between 19.52 and 20.48
- The estimated value of *% Trips by Bus* lies between 0.04 and 0.06
- The estimated value of *% Trips to School* lies between 0.09 and 0.11
- The estimated value of *Average People per Vehicle* lies between 1.39 and 1.41

---

**Figure 4.16** Output Screen for Estimated Values of Variables in Sample

### 4.6.2 Sample Sizes for Hypothesis Testing

The second purpose of a survey (or surveys) may be to test a statistical hypothesis concerning some of the population parameters e.g. are there significant differences in trip rates in different areas, or has use of a specific mode risen following introduction of a new transport service? To test such hypotheses,
it is necessary to compare two sample statistics (each being an estimate of a population parameter under different conditions), each of which has a degree of sampling error associated with it. The tests are performed using statistical significance tests where the power of the test is a function of the sample size of the survey(s).

In using sample survey data to test hypotheses about population behaviour, it is first necessary to ensure that the hypothesis to be tested is correctly specified. While hypotheses are often described as having been rejected or accepted, it should be realised that the rejection of a hypothesis is to conclude that it is false, while the acceptance of a hypothesis merely implies that we have insufficient evidence to believe otherwise. Because of this, the investigator should always state the hypothesis in the form of whatever it is hoped will be rejected. Thus if it is hoped to prove that car ownership is higher in one area than in another, the hypothesis should be that car ownership is equal in the two areas, and then we try to reject the hypothesis (statistically).

An hypothesis that is formulated with the hope of rejecting it is called the null hypothesis and is denoted by \( H_0 \). The rejection of \( H_0 \) leads to the "acceptance" of an alternative hypothesis denoted by \( H_1 \). Thus in the case of testing for differences in two average values, the null hypothesis could be specified as:

\[
H_0: \mu = \mu_0
\]  
(4.15)

The alternative hypothesis could be specified in a number of different ways depending on the purpose of the comparison. Possible alternative hypotheses might include:

\[
H_1: \mu > \mu_0
\]  
(4.16)

\[
H_1: \mu < \mu_0
\]  
(4.17)

\[
H_1: \mu \neq \mu_0
\]  
(4.18)

Note that only one of these alternative hypotheses can be used in any particular test. The first two alternative hypotheses would constitute a one-tailed test, while the third alternative hypotheses would constitute a two-tailed test.

Since the data to be used in the testing of these hypotheses is to be collected using a sample survey, we would be reluctant to base our decision on a strict deterministic interpretation of the stated hypotheses. For example, if we wished to test the null hypothesis that \( \mu = \mu_0 \), then we would base our decision on testing whether \( \mu \) fell within a critical region \( (D) \) around \( \mu_0 \). Thus, if \( \mu_0 \pm D/\sigma \), then we would, in practice, not reject the hypothesis that \( \mu = \mu_0 \). Similarly if \( \mu < \mu_0 - D \), then we would not reject the hypothesis that \( \mu = \mu_0 \) in favour of the alternative hypothesis that \( \mu > \mu_0 \). The definition of the critical region, \( D \), is...
Sampling Procedures

somewhat arbitrary and merely serves to give a workable rule for the rejection of hypotheses. Obviously a smaller critical region will make rejection of the null hypothesis easier.

In testing hypotheses, there are four possible end-states of the hypothesis testing procedure. Two of these states signify that a correct decision has been made while the other two indicate that an error has been made. The four end-states may be depicted as shown in Table 4.1.

Table 4.1 Possible End-States of Hypothesis Testing

<table>
<thead>
<tr>
<th>DECISION</th>
<th>TRUE STATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accept H₀</td>
<td>Correct</td>
</tr>
<tr>
<td>Reject H₀</td>
<td>Type I Error</td>
</tr>
<tr>
<td></td>
<td>Type II Error</td>
</tr>
</tbody>
</table>

Thus if the true state is described by the null hypothesis H₀ and we accept H₀ as being a description of the true state, then we have made no error. Similarly, we will be equally correct if we reject H₀ when the true state is actually described by the alternative hypothesis H₁.

A Type I Error will have been committed if we reject the null hypothesis when it is in fact true. For example, we assume that two groups of households have the same car ownership when, in fact, one group has more cars than the other. A Type II Error will have been committed if we accept the null hypothesis when it is in fact false. For example, we assume from two sample surveys conducted at different points in time that trip rates have risen over time, when in fact it was simply chance variations in the two samples which gave the appearance of an increase in trip rate.

Obviously in testing hypotheses we would be interested in trying to minimise the chances of making either a Type I or Type II error. Which one we would be most interested in avoiding will depend on the relative costs associated with each type of error. The degree to which we wish to avoid each type of error is expressed in terms of the maximum probability which we will accept for making each type of error. The acceptable probability of committing a Type I error is called the level of significance of the test and is denoted by \( \alpha \). The acceptable probability of committing a Type II error is denoted by \( \beta \). The value \( 1-\beta \) is often called the power of the test. The power of the test cannot be calculated unless a specific alternative hypothesis is stated in the form of an equality, such as \( H₁: m₁ = m₀ + \Delta \).

The inter-relationships between the variables described above can be more clearly seen by reference to Figures 4.17 and 4.18. Both figures depict the
distribution of sample means which would be obtained from repeated sampling from a population. In Figure 4.17, it can be seen that the null hypothesis expressed in equation (4.15) is in fact correct, i.e. \( \mu = \mu_0 \). If this was known with certainty, we would definitely accept the null hypothesis. However, because \( \mu \) is estimated from a sample survey, we would not know that the true mean is \( \mu_0 \). Rather, all we know is our sample estimate of the mean (denoted by a value of \( m \)). We also know, from the Central Limit Theorem, that the sample estimates of \( \mu \) will be normally distributed around \( \mu_0 \) (for large enough sample sizes), as shown in Figure 4.17. In such circumstances, a certain proportion (\( \alpha \)) of our estimates will be greater than \( \mu_0 + D \), and hence according to our decision rule, involving rejection of values outside the range of the critical region, we will reject \( H_0 : \mu = \mu_0 \) even though \( H_0 \) is, in fact, correct. We will therefore have committed a Type I Error (on a percentage of occasions).

![Figure 4.17 Probability of a Type I Error](image)

Consider, now, the situation depicted in Figure 4.18 which shows the situation when the specific alternative hypothesis is, in fact, correct, i.e. \( H_1 : \mu = \mu_0 + D \). Once again, however, our sample estimates of \( \mu \) will be distributed normally around \( \mu_0 + D \). Given the same critical region \( \alpha \) (around \( \mu_0 \)), we can see that on \( \alpha \) percent of occasions the sample estimate of \( \mu \) will be less than \( \mu_0 + D \) and hence we will be led to accept the null hypothesis when it is in fact false. Thus, a Type II error will have been committed.
It should be obvious from Figures 4.17 and 4.18 that we can change the values of \( a \) and \( b \) by changing the critical region \( D \). Thus increasing \( D \) will reduce \( a \) but will at the same time increase \( b \). Thus, we can make less Type II errors at the expense of making more Type II errors. Alternatively, decreasing \( D \) will increase \( a \) but reduce \( b \). Importantly, no adjustment of \( D \) is possible which will simultaneously reduce both \( a \) and \( b \). All one can do by changing \( D \) (i.e. changing the decision rule by which we accept or reject the null hypothesis) is to trade-off the frequency of Type I and Type II errors.

One can reduce \( b \) without affecting \( a \), by increasing \( D \). Thus, if one wishes to test for a larger difference between the null hypothesis and the specific alternative hypothesis, then one can reduce the probability of making a Type II error. However, specifying a large value of \( D \) makes the test procedure of little value, since it is then incapable of detecting small differences in the null and alternative hypotheses. All that the test then says is that if a large difference in a parameter is detected then we can be fairly sure that such a difference is real.

The only way in which \( a \) and \( b \) can be simultaneously reduced is by increasing the sample size \( n \). This is because, from equation (4.3), increasing the sample size will reduce the standard error (i.e. the standard deviation of the distributions shown in Figures 4.17 and 4.18). Reducing the standard error of each distribution will obviously reduce the area under the curve lying in the tails of the distribution to the right of the critical region line \((\mu = \mu_0 + D)\) for \( a \) and to the left of the critical region line for \( b \). The question then remains as to how much the sample size must be increased in order to reduce both \( a \) and \( b \) to acceptable levels.
Chapter 4

In testing a statistical hypothesis, the significance level ($\alpha$) is normally controlled by the investigator, while $\beta$ is controlled by using an appropriate sample size. Given an estimate of the standard error, the critical region ($\square$) will automatically be determined as a consequence of meeting the requirements of attaining a fixed level of $\alpha$. For the one-tailed test where:

\[ H_0: \square = \square_0 \]
\[ H_1: \square = \square_0 + \square \]  \hspace{1cm} (4.19)

or
\[ H_1: \square = \square_0 - \square \]

it can be shown (Walpole and Myers, 1978) that the required sample size to meet the requirements of $\alpha$ and $\beta$ is given by:

\[ n = \frac{(z_\alpha + |z_\beta|)^2 \square^2}{\square_0^2} \]  \hspace{1cm} (4.20)

where $z_\alpha$, $z_\beta$ = critical values from standard normal distribution tables at levels of significance $\alpha$ and $\beta$ (see Appendix A).

$\square^2$ = population parameter variance.

$\square$ = desirable detectable difference in hypotheses.

In the case of a two-tailed test of the form:

\[ H_0: \square = \square_0 \]
\[ H_1: \square = \square_0 \pm \square \]  \hspace{1cm} (4.21)

then the minimum sample size is given by:

\[ n = \frac{(z_{\alpha/2} + |z_{\beta/2}|)^2 \square^2}{\square_0^2} \]  \hspace{1cm} (4.22)

The discussion so far has concentrated on testing the results obtained from one sample survey ($\square$) against a known benchmark ($\square_0$). In many cases, however, it is useful to compare the results obtained from two sample surveys (such as before-and-after studies), or to compare the results obtained from two groups within the same sample survey (e.g. car ownership rates for two different socio-economic groups in a population). In such cases, it is necessary to determine the sample size required for each of the surveys in order to obtain pre-determined values of $\alpha$ and $\beta$. Assuming that the sample size for each survey, or for each sub-population within a single survey, is the same, then for a one-tailed test the required sample size is:
Sampling Procedures

\[ n = \frac{(z_1^2 + z_2^2)(1^2 + 1^2)}{\bar{s}^2} \]  \hspace{1cm} (4.23)

and for two-tailed test, the required sample size is:

\[ n = \frac{(z_1^2/2 + z_2^2)(1^2 + 1^2)}{\bar{s}^2} \]  \hspace{1cm} (4.24)

where \( \bar{s}^2 \) = population parameter variances for each of the samples.

Typically, in before-after surveys where one is attempting to detect an improvement (or degradation) following a change in the transport system, one is concerned only with a one-tailed test. Also since the same population is being surveyed on both occasions, it is usually safe to assume that \( \bar{s}_1^2 = \bar{s}_2^2 = \bar{s}^2 \).

Equation (4.23) then reduces to:

\[ n = \frac{2(z_1^2 + z_2^2)(\bar{s}^2)}{\bar{s}^2} \]  \hspace{1cm} (4.25)

Three implications of this equation, while obvious on reflection, are worth noting. First, the larger the values of \( z_1 \) and/or \( z_2 \), the larger the required sample size. Large values of \( z \) will occur if \( \bar{s} \) and/or \( \bar{s}_2 \) are small. Thus the smaller the probabilities that one wishes to accept of making either Type I or Type II errors, the larger the required sample size. For example, if one wishes to reduce the values of both \( \bar{s}_1 \) and \( \bar{s}_2 \) from 10% to 5%, then, all other things being equal, the required sample size will increase by a factor of 1.64. Second, as noted when estimating sample sizes for parameter estimation, the required sample size is directly proportional to the population variance of the parameter to be tested. Third, the sample size is inversely proportional to the square of the difference which one would like to detect in the parameter in the before-after situation. Halving the desired detectable difference will result in a fourfold increase in the required sample size.

To illustrate the use of the above equations, consider that a study wished to examine household public transport trip generation rates in an area before and after the introduction of a new public transport system. Because it would be unrealistic to expect a sample survey to detect a change of any size (no matter how small), it is necessary to have an idea of the expected change (\( \bar{s} \)) and to test whether this change has been realised. Assume that the initial weekly trip rate by public transport in the area is estimated to be 8 trips per household, with a variance of 4 trips per household, and that we wish to test whether this rate has increased by 25%, as expected. For no good reason, other than convention, let us...
assume that $\alpha = \beta = 0.05$. The minimum sample size for each survey would then be given by:

$$n = \frac{2(z_{\alpha/2} + z_{\beta})^2}{d^2}$$

$$n = \frac{2(1.645 + 1.645)^2}{0.05^2} = 87 \quad (4.26)$$

Further examples of sample size calculations for hypothesis testing may be found in Skelton (1982).

As an example of a slightly different situation in which sample surveys might be used to test hypotheses, consider the following situation. The suburb of Fairfield has recently had a new free minibus service established to serve residents of that suburb. The residents of nearby Glossodia, on hearing of this, petition their transport authority for establishment of a similar service in their area. They are informed that Fairfield received the service only because of the relatively high number of households in that suburb which do not own a car. The Glossodia residents reply that their suburb has even more non-car-owning households than Fairfield, and therefore they too deserve a free minibus service. They offer to back up this claim by conducting a survey of households in Glossodia and Fairfield to show the difference in car-ownership. The transport authority agrees to provide a free bus service for Glossodia if it can be shown, by an unbiased sample survey, that Glossodia has more non-car-owning households than Fairfield. However, to ensure that the claim is fully justified, the transport authority insists that the difference in the proportion of non-car-owning households be at least 5 percentage points at the 5% significance level. The residents agree to this, perhaps after some bartering, and proceed about the task of designing the survey (obviously with professional help).

The first thing which must be done is to define the hypotheses to be tested. Since the residents wish to prove that non-car-ownership is higher in Glossodia than in Fairfield, they choose as their null hypothesis the idea that the non-car-ownership is equal in both areas. That is:

$$H_0: \ p_1 = p_2 \quad (4.27)$$

where $p_1 = \text{proportion of households in Glossodia with no car}$

$$p_2 = \text{proportion of households in Fairfield with no car}$$

The alternative hypothesis, which they would like to prove, can be stated as:
Sampling Procedures

\[ H_1 : p_1 > p_2 \]  \hspace{1cm} (4.28)

However, to enable calculation of the power \((1-\beta)\) of the test, it is necessary to state the alternative in a specific manner such as:

\[ H_1 : p_1 = p_2 + \delta \]  \hspace{1cm} (4.29)

where \(\delta\) = expected difference in car-ownership.

The second step is to agree on the significance and power of the test to be performed. To the residents, this means agreeing on how certain they would like to be when drawing conclusions from the sample survey data. Initially, the residents may say that they want to be 100\% certain of making the right decision. However it is then explained to them, by their consultants, that to be 100\% certain they would need to survey all households in both areas. Assuming that each area is quite large, the cost of such a survey would be quite prohibitive - they would be able to buy several minibuses of their own for less than the cost of the survey! It is explained to them that if they are willing to accept being wrong with a certain small probability, then the cost of the survey could be reduced quite substantially. They agree that this seems reasonable and therefore set about determining how often they would be prepared to be wrong.

They must set two probabilities: one each for Type I and Type II errors. To them, a Type I error means accepting the conclusion that they do have more non-car-owning households than Fairfield when it is, in fact, not the case. A Type II error means accepting that Glossodia does not have more non-car-owning households than Fairfield, and hence they will not get the free bus service, even though in reality they do have more non-car-owning households and should have got the free bus. Quite naturally the Glossodia residents are not very concerned about Type I errors; from their point of view they would like them to happen as often as possible. However, the transport authority has wisely constrained them in this regard by stipulating a 5\% level of significance [i.e. \(\beta = 0.05\)]. In fact, the transport authority has constrained them more than they probably realise, because as well as stipulating the probability of Type I errors they have also stipulated the size of the critical region for the test (i.e. \(\delta = 0.05\)). By doing this, the transport authority has effectively predetermined the sample size needed for the surveys in each area, as will now be shown.

During their discussion with the residents, the authority let it be known that the proportion of non-car-owning households in Fairfield was 20\%. Given this, the variance in this proportion within Fairfield would be given by:

\[ \delta^2 = p_2(1 - p_2) \]

\[ = 0.20(1-0.20) \]
Since there is expected to be no drastic difference in car-ownership between the two areas it would be reasonably safe to assume the same variance for Glossodia.

If Type I errors are limited to 5%, the critical region is set at 0.05 and the variance of the difference in car ownership is estimated to be 0.32 \((=\sigma^2)\) then, as seen by reference to Figure 4.17, the sample size must be fully determined by:

\[
n = \frac{2\sigma^2}{(\text{s.e.}(p_1 - p_2))^2}
\]

\[
= \frac{2\sigma^2 \cdot 1.645^2}{0.05^2}
\]

\[
= 346
\]

Since surveys of this size would need to be carried out in both areas, a total of 692 households must be surveyed. At a cost per completed survey of (say) $40 this represents a total cost of $27,680. This appears to be quite a high outlay in order to get a free mini-bus and hence the residents begin to question whether it is worth doing the survey at all. To help in making this decision, it is useful to calculate the probability of the residents making a Type II error, i.e. accepting that Glossodia does not have more non-car-owning households than Fairfield, even though in reality they do and should have got the free bus.

To perform this calculation, it is necessary for the residents to specify a value of the expected difference in car-ownership between the two areas so that a specific alternative hypothesis can be formulated as per equation (4.29). Obviously if the difference is confidently expected to be quite large then it may still be worthwhile spending the money on the survey since the results are highly likely to be favourable to Glossodia residents. On the other hand, if they do not expect the difference to be much more than the 5% stipulated by the transport authority then the chances of achieving a favourable result are diminished. The chances of not being able to prove that Glossodia deserves the free bus, when in fact it does, can be expressed in terms of \(\beta\). The variation in \(\beta\) as a function of the expected difference in the proportion of non-car-owning households is shown in Figure 4.19.
As the expected difference in the proportion of non-car-owners in each area becomes greater, the probability of an unfavourable result for the Glossodia residents decreases. Assume that after consultation, and perhaps some reference to existing data sources, the residents decide that the most likely difference in the proportion of non-car-owning households in Fairfield and Glossodia is 0.07. i.e. given that the proportion of households without cars in Fairfield is 0.20, the expected value for Glossodia is 0.27. In this case, with a sample size of 346 households in each area, there is a probability of 25% (from Figure 4.19, \( \beta = 0.25 \)) that the survey will show a difference less than 0.05 even though the true difference may well be 0.07. Thus there is a 25% chance that the free bus will not be provided when in fact it should be provided. The question remains as to whether this probability is too high compared to the cost of the survey. Using simple expected utility theory, the expected gain from conducting the survey will be given by:

\[
\text{Expected Gain} = (\text{probability of getting bus})(\text{value of bus}) - \text{survey cost}
\]

\[
= (0.75)(40,000 \text{ (say)}) - 27,680
\]

\[
= 2,320
\]

\[(4.32)\]

In such a situation, it is just worthwhile for the residents to go ahead with their survey; on the balance of probabilities they should come out ahead, provided that their assumptions along the way have been accurate.
The foregoing example has shown the application of decision analysis to sample survey design (in this case to the decision of whether to proceed with a survey or not). It has also shown very clearly that the probabilities of Type I and Type II errors cannot be adequately specified without a consideration of the likely costs involved in committing each type of error. For example, if the value of the bus had been $60,000 in the previous example, then the value of $ could have been set as high as 0.65 and the decision to conduct the survey would still have been correct. While fraught with numerous practical difficulties, the concepts of decision analysis can be most useful in bringing a more rational basis to sample survey design (see Ramsay and Openshaw (1980), and Jessop and Gilbert (1981) for further discussion of this matter).

4.7 VARIANCE ESTIMATION TECHNIQUES

In sample surveys, the analyst is usually interested not only in measuring the mean of a variable for the sample in question, but in estimating the value of this mean for the total population from which the sample was drawn. If the sample in question is a random sample from the population, then the population mean may be taken to be the same as the sample mean. If, however, a new sample was drawn from the population (using the same methods), one would not expect the two sample means to be the same. Therefore, in stating that the population mean is the same as the sample mean, one is really stating that the population mean is best represented by the sample mean. However, repeated sampling will result in a distribution of estimates of the mean, and the standard deviation of this distribution is known as the standard error of the estimate of the mean.

4.7.1 Variability in Simple Random Samples

For simple random samples, the standard error of the mean may be calculated easily by the following equation:

\[ \text{s.e.}(m) = \sqrt{\frac{S^2}{n}} = \frac{S}{\sqrt{n!}} \]  

(4.33)

where \( \text{s.e.}(m) \) = standard error (of the mean)

\( S \) = standard deviation of the variable in the population

\( n \) = sample size

Thus larger samples produce lower standard errors (i.e. more precise estimates of the mean), while variables which are less variable in themselves can be estimated more precisely.
In real-life surveys, however, there are often considerable deviations from the ideas of simple random sampling. Such complex surveys often include design refinements such as stratification, multi-stage sampling and the use of clusters as sampling units. Whilst it is theoretically possible to extend equation (4.33) to account for these complexities, such extensions often become cumbersome, if not mathematically intractable. It is therefore desirable to use other methods to estimate the degree of sampling error in a sample estimate of the mean.

4.7.2 Design Effects

A particularly simple, but often used, method of estimating sampling error for complex survey designs is the use of "design effects". A design effect (Kish, 1965) is a factor which relates the variance in a sample design to the variance which would have been obtained if the sample had been chosen by simple random sampling. The design effect can be greater than or less than one, depending on whether the sample design tends to increase or decrease the sampling error. An example of a sample design which decreases sampling error is stratified sampling; designs which increase sampling error include multi-stage sampling and cluster sampling.

The following calculations, based on an example described by Moser and Kalton (1979), show the effect of stratification on the sampling error obtained from a survey and also shows the calculation of the design effect for this sampling plan. Consider a sample survey of one person from a household in different municipalities within a city. The objective of the survey was to determine the proportion of households favouring the upgrading of an airport. Since it was known beforehand that households in different areas of the county would have different views on the upgrading (e.g. those near the airport would prefer no upgrading, while those living away from the airport in more affluent suburbs would prefer the upgrading), the sample was chosen such that each municipality within the city was represented in the sample in proportion to the number of households within each municipality. The results of the surveys are shown in Table 4.2, together with the proportion in each municipality favouring the upgrade (p_j), and a product term (n_ip_i(1-p_i)) required for the calculation of the standard error for a stratified sample.
Table 4.2 Results of the Stratified Sample Airport Upgrading Survey

<table>
<thead>
<tr>
<th>Municipality</th>
<th>Households in Sample ( n_i )</th>
<th>Number Favouring Upgrade ( p_i )</th>
<th>( n_i p_i(1-p_i) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dryden</td>
<td>95</td>
<td>86</td>
<td>7.170</td>
</tr>
<tr>
<td>Freeville</td>
<td>43</td>
<td>22</td>
<td>10.736</td>
</tr>
<tr>
<td>Lansing</td>
<td>25</td>
<td>18</td>
<td>5.040</td>
</tr>
<tr>
<td>Montrose</td>
<td>39</td>
<td>31</td>
<td>6.355</td>
</tr>
<tr>
<td>Waverley</td>
<td>32</td>
<td>20</td>
<td>7.500</td>
</tr>
<tr>
<td>Belgrave</td>
<td>66</td>
<td>33</td>
<td>16.500</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>300</strong></td>
<td><strong>210</strong></td>
<td><strong>54.301</strong></td>
</tr>
</tbody>
</table>

The proportion in the city in favour of the airport can be obtained by calculating a weighted average of the proportions in each of the strata (municipalities). Since the same sampling rate was adopted in each strata, this average can be obtained as the proportion of the total number favouring the upgrade divided by the total sample size, i.e., 70% (210/300) of city households can be seen to be in favour of the upgrade. If these results had been obtained by a simple random sampling plan, then the standard error of the estimate of this proportion (s.e.(p_{srs})) could be calculated by the use of equation 4.11. Thus, assuming that the sample size is a small proportion of the total households in the city:

\[
s.e.(p_{srs}) = \sqrt{\frac{0.700! \cdot 0.300!}{300}} = 0.0265 \tag{4.34}
\]

However, it is known that the estimate of standard error for a proportionate stratified random sample is given by:

\[
s.e.(p_{prop}) = \sqrt{\frac{\sum n_i p_i(1-p_i)}{n^2}} \tag{4.35}
\]

\[
= \sqrt{\frac{54.301}{300^2}} = 0.0246
\]

Comparison of the above standard errors shows the reduction gained by the use of stratified random sampling. This gain can be expressed in terms of the design effect by calculating the ratio of the sampling error variances:

\[
\text{Design Effect} = \frac{(0.0246)^2}{(0.0265)^2} = 0.862 \tag{4.36}
\]

This design effect of 0.862 indicates that the use of stratified sampling brings about a reduction of 14% in the variance of the estimate of the mean, compared to what would have been obtained using simple random sampling. Put in more
practical terms, it means that a simple random sample of 348 households (\(=300/0.862\)) would be needed to give the same level of accuracy as provided by this stratified sample of 300 households. The design effect can therefore be interpreted as a measure of sample size savings by the use of the more complex sample design (i.e. we can use a sample which is only 86.2% of the simple random sample if we are willing to use a stratified sample design). We can then trade-off this reduction in sample size against the possible increase in cost per sampling unit to determine which is the most efficient way of obtaining a sample of specified precision.

While the above example has used the estimation of a proportion as the basic rationale for the survey, the same arguments apply when one is attempting to estimate the mean of a continuous variable. In this case, the standard error of the sample mean based on a small proportionate stratified random sample of size \(n\) is:

\[
s.e.(\bar{\text{prop}}) = \sqrt{\frac{\sum i n_i^2}{n^2}}
\]

(4.37)

Given this demonstration of the calculation of a design effect for a stratified random sample, the question arises as to whether there is a generally applicable value of the design effect which could be used for all samples of this type. Unfortunately, the answer to this question is in the negative. While it is generally true that stratification will reduce the standard error of estimates, the extent to which it reduces it will vary from case to case. As noted in Section 4.4.2, there is no gain in precision if the stratification of the sample is based on a variable which has no connection with the parameters under investigation (e.g. estimating trip rates for people born on different days of the week). The gain provided by stratified sampling can be expressed as the difference in variances of the estimate, as given by:

\[
\text{Var}(\bar{\text{srs}}) - \text{Var}(\bar{\text{prop}}) \approx \sqrt{\frac{\sum i n_i(\bar{\|i\|} - \bar{\|\|})^2}{n^2}}
\]

(4.38)

where \(\bar{\|i\|}\) is the population mean for the \(i\)th stratum and \(\bar{\|\|}\) is the overall population mean. This expression shows that if all the strata have the same mean, then there is no gain from stratification (i.e. the design effect is equal to one). As the difference in the means within each stratum increases, then stratification will have a greater effect on reductions in variance (i.e. the design effect will become smaller and smaller). The size of the design effect will therefore depend on the extent to which the stratification variables are able to make significant differences in the strata means for the parameters in question. An estimate of the size of the
Chapter 4

design effect can be made before the survey if you are able to estimate the likely means within each strata. Such information may come from the sources identified in Section 4.6.1.

The use of complex sample designs does not always reduce the standard error of the estimates. Whenever cluster sampling (including multi-stage sampling) is employed, then the design effect will generally be greater than one. As an illustration, consider once again the data presented in Table 4.2, and examine the results obtained from the municipality of Belgrave. The number of households in this municipality was 660, of whom 10% were chosen for the survey. In the discussion so far, it has been assumed that these households were chosen in a simple random sampling manner. However, consider for the following discussion that they were chosen as clusters, rather than individually. There may have been many practical reasons for this, but the usual one is to reduce field costs when performing the survey. For example, the regular grid street pattern within Belgrave may have conveniently divided the households consistently into groups of six along one side of a street between adjacent intersections. To reduce travel costs and time between interviews at households, a decision may have been made to select 11 of these clusters and then perform the interview with all heads of households in the selected clusters (for the moment we are assuming a 100% response rate to this survey). Thus we are selecting 10% of the clusters rather than 10% of the individual households to obtain our sample of 66 households. Assume that the results obtained from the 11 clusters in Belgrave are as shown in Table 4.3, together with some values required for later calculations of standard errors.

The proportion favouring the airport upgrading in the entire Belgrave sample is obtained by dividing the number favouring it by the total number of households in the sample. To estimate the standard error of this estimate, it is useful to think of the survey not as a random sample of households but as a random sample of clusters (i.e. change the sampling unit from household to cluster as described in Section 4.2), where the parameter being measured is the average proportion of households within clusters favouring the airport upgrading. Since it is a simple random sample of clusters, each with a value of $p_i$, then the standard error of the mean of $p_i$ can be obtained by use of a modification of equation 4.2, such that:
Table 4.3 Results of the Cluster Sample Airport Upgrading Survey

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Households in Cluster, n_i</th>
<th>Number Favouring Upgrade</th>
<th>p_i</th>
<th>(p_i-p)^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>3</td>
<td>0.500</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>5</td>
<td>0.833</td>
<td>0.111</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>2</td>
<td>0.333</td>
<td>0.028</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>3</td>
<td>0.500</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>4</td>
<td>0.667</td>
<td>0.028</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>1</td>
<td>0.167</td>
<td>0.028</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>4</td>
<td>0.667</td>
<td>0.028</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>2</td>
<td>0.333</td>
<td>0.028</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>6</td>
<td>1.000</td>
<td>0.250</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>1</td>
<td>0.167</td>
<td>0.111</td>
</tr>
<tr>
<td>11</td>
<td>6</td>
<td>2</td>
<td>0.333</td>
<td>0.028</td>
</tr>
<tr>
<td>TOTAL</td>
<td>66</td>
<td>33</td>
<td>0.723</td>
<td></td>
</tr>
</tbody>
</table>

\[
s.e.(p_i,\text{clust}) = \sqrt{\frac{M-m}{M}!} \frac{S_c^2}{m!} \tag{4.39}
\]

where we have drawn a sample of m clusters from a population of M clusters, and where \(S_c^2\) is the standard deviation of the cluster values \(p_i\). This standard deviation may be estimated from the sample results as:

\[
S_c^2 = \frac{1}{m-1} \sum_{i=1}^{m} (p_i-p)^2 \tag{4.40}
\]

In this case, \(m=11\), \(M=110\), and \(S_c^2 = 0.723/10 = 0.072\). Hence,

\[
s.e.(p_i,\text{clust}) = \sqrt{\frac{100}{110}!} \frac{0.072}{11!} \tag{4.41}
\]

\[
= \sqrt{0.0059} = 0.077
\]

If the results for Belgrave had been based on a simple random sample of households, then the standard error would have been given by:

\[
s.e.(p_i,\text{srs}) = \sqrt{\frac{N-n_i}{N}!} \frac{p(1-p)}{n-1} \tag{4.42}
\]

\[
= \sqrt{\frac{594}{660}!} \frac{0.5(1-0.5)}{65} \tag{4.42}
\]

\[
= 0.059
\]

The design effect for this cluster sample of households within Belgrave is therefore given as:
Design Effect = \frac{(0.077)^2}{(0.059)^2} = 1.70 \hspace{1cm} (4.43)

Such a value for the design effect in the range from 1 to 2 is typical for a cluster sample, although by no means universal. Just as the design effect for a stratified random sample depends on the degree to which the selected strata are homogeneous, so too the design effect for cluster samples depends on the degree of homogeneity within the cluster. It is more than likely that members of a cluster are more like each other than they are like others outside the cluster. In our airport upgrading example, it is likely that members of households living on the same side of the street on the same block may well have discussed the issue of the airport upgrading, or for many other reasons may have come to a similar conclusion based on where they live. The degree to which the members of a cluster are similar is measured by the intra-cluster correlation. This correlation measure can be integrated into our considerations of design effect by noting that for a cluster sample with C members per cluster:

\[ \text{Var}_{\text{clust}} = \text{Var}_{\text{srs}} [1 + (C-1)\rho] \hspace{1cm} (4.44) \]

Since the design effect is given as the ratio of the cluster sample variance to the simple random sample variance, then:

\[ \text{Design Effect} = [1 + (C-1)\rho] \hspace{1cm} (4.45) \]

Several points arise from this equation which are worth noting:

(a) if \( C = 1 \), that is if each cluster contains only one member, then the design effect is equal to one since there is essentially no clustering taking place, and the cluster sample degenerates to a simple random sample;

(b) if \( \rho = 0 \), then there is no intra-cluster correlation, in that the members are just as similar to members of the population at large as they are to members of their own cluster. In such a case, each cluster can be seen as just a random mini-sample of the population at large, and the summation of these clusters will simply be the same as a simple random sample, as shown by the calculated design effect being equal to one; and

(c) so long as the value of \( \rho \) is positive (and this is almost always the case), the design effect will be greater than one. The greater the similarity between members of a cluster, the higher will be the value of \( \rho \) and hence the higher will be design effect.
Sampling Procedures

In the airport upgrading example, the design effect was calculated from the data as being equal to 1.70. Given a cluster size (C) of 6, this results in a value of \( d \) of +0.14. This is a typical value of intra-cluster correlation, although it can sometimes be far greater.

The differences in design effect for stratified sampling and cluster sampling are worth noting. Both come about because of homogeneity within the strata or clusters, but the effect is in different directions. Thus greater homogeneity within strata results in a greater reduction in the design effect for stratified sampling, but greater homogeneity within clusters results in a greater increase in the design effect for cluster sampling. The difference lies in the fact that whereas we wish to include all the strata in a stratified sample, we wish to deliberately exclude some of the clusters in the cluster sample.

The concept of cluster sampling can be further extended, and the effects on sampling error mitigated to some extent, by employing the concept of multi-stage sampling in which a sample of sub-clusters is selected from within the originally selected clusters. In this way, the design effect can be reduced by spreading the final sample over a greater number of clusters. It is also possible to trade-off the effects of clustering and stratifying by the use of a stratified multi-stage sample, although usually this still results in a design effect greater than one since the clustering effect on sampling error is generally greater than the effect of stratification. Further details on the calculation of design effects may be found in Moser and Kalton (1979).

4.7.3 Replicate Sampling

The use of design effects gives only a rough approximation to the expected standard error and would be used mainly when trying to estimate sample sizes before the data is collected. Once the data has been collected, however, one can employ the methods of replication to directly calculate standard errors. Drawing on the idea expressed earlier, whereby the standard error is simply the standard deviation of the distribution of estimates of the mean from repeated sampling, it is possible to draw a number of independent samples from the population thus enabling the standard error to be estimated directly. When two separate samples are drawn, this method is sometimes called a paired selection design (Kish, 1965). When more than two samples are drawn, this is often referred to as replicated or interpenetrating sampling (Deming, 1960). The idea of replication has been used in transport surveys in the 1974 Survey of Seat Belt Utilization conducted for Transport Canada, in a study reported by Ochoa and Ramsey (1986), and in travel surveys conducted in Australia by the Transport Research Centre (Richardson and Ampt, 1994; Richardson and Cuddon, 1994).

With replicated sampling, sample estimates of the mean (or for that matter any other parameter) can be calculated for each sub-sample, and the variation of
these estimates is an indication of the sampling error which would have been obtained from the total sample. The calculated sampling error includes the effects of all elements of the sample design, and hence replication techniques can be used for sample designs of any degree of complexity.

The number of replications to be used is open to question. Choosing a few large sub-samples means that the variation within each sub-sample is well estimated, but the variation between sub-samples is based on only a few observations. On the other hand, choosing a large number of small sub-samples means that the variance within the sub-samples is based on only a small sample size. Somewhere a compromise needs to be struck, using a moderate number of moderately sized sub-samples. One guideline that is often adopted is that originally proposed by Deming (1960), whereby he recommends the use of ten sub-samples. This method has the added advantage that it is easy to choose the sub-samples by sub-dividing the entire sample based on the last digit of the case identification number, i.e. all households with identification numbers ending in "1" belong in sub-sample 1, all those with numbers ending in "2" belong in sub-sample 2 and so on.

If a total sample is selected by choosing \( r \) independent samples of equal size, and if \( z_i \) is the estimate of a parameter for the \( i \)th sample, then the overall estimate of this parameter is given by:

\[
z = \frac{1}{r} \sum z_i \quad (4.46)
\]

and the standard error of \( z \) is estimated by:

\[
s.e.(z) = \sqrt{\frac{s_z^2}{r}} \quad (4.47)
\]

\[
= \sqrt{\frac{\sum (z_i - \bar{z})^2}{r(r-1)}} \quad (4.48)
\]

where \( s_z \) is an estimate of the standard deviation of the sub-sample estimates of the parameter in question.

Suppose, for example, in the airport upgrading example introduced in the previous section that the total sample of 300 households was collected by means of 10 separate and random sub-samples with 30 households in each sample. The results from the 10 sub-samples is shown in Table 4.4.
Table 4.4  Results of the Replicated Sample Airport Upgrading Survey

<table>
<thead>
<tr>
<th>Replication</th>
<th>Households in Replicate</th>
<th>Number Favouring Upgrade</th>
<th>$z_i$</th>
<th>$(z_i - z)^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>18</td>
<td>0.600</td>
<td>0.0100</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>21</td>
<td>0.700</td>
<td>0.0000</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>23</td>
<td>0.767</td>
<td>0.0044</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>18</td>
<td>0.600</td>
<td>0.0100</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>25</td>
<td>0.833</td>
<td>0.0178</td>
</tr>
<tr>
<td>6</td>
<td>30</td>
<td>19</td>
<td>0.633</td>
<td>0.0044</td>
</tr>
<tr>
<td>7</td>
<td>30</td>
<td>24</td>
<td>0.800</td>
<td>0.0100</td>
</tr>
<tr>
<td>8</td>
<td>30</td>
<td>19</td>
<td>0.633</td>
<td>0.0044</td>
</tr>
<tr>
<td>9</td>
<td>30</td>
<td>22</td>
<td>0.733</td>
<td>0.0011</td>
</tr>
<tr>
<td>10</td>
<td>30</td>
<td>21</td>
<td>0.700</td>
<td>0.0000</td>
</tr>
<tr>
<td>TOTAL</td>
<td>300</td>
<td>210</td>
<td></td>
<td>0.0622</td>
</tr>
</tbody>
</table>

Given that the standard deviation of the $z_i$ values is calculated as 0.083, then the standard error of the mean may be estimated from equation (4.47) as:

$$s.e.(z) = \sqrt{\frac{0.083^2}{10}}$$

$$= 0.0263$$

The standard error of the mean may also be estimated from equation (4.48) as:

$$s.e.(z) = \sqrt{\frac{0.0622}{10(10-1)}}$$

$$= 0.0263$$

This value of 0.0263 compares favourably with the value of 0.0265 estimated in equation (4.34) for the case of simple random sampling. Such agreement is not, however, necessary and simply reflects the fact that in this particular case the assumptions underlying the analytical calculation of equation (4.34) appear to have been approximated in this particular data set.

One obvious problem with this approach, however, is that by selecting a small number of independent samples, the estimate of the standard error could be, in itself, subject to considerable sampling error. This could be overcome to some extent by drawing a larger number of independent samples. However if the size of the sub-samples were to remain at 30, this would result in an excessively expensive total sample. Conversely, if the total sample size was to remain constant at 300, then the size of the sub-samples would decrease and this may create problems within each sub-sample (e.g. with respect to the number of observations within stratification cells). In order to overcome the problem of sufficient replicates of sufficient size, several concepts of half-sample replication have been proposed as described below.
4.7.3.1 Half-Sample Replication

When working with a stratified sample, one can employ the idea of half-sample replications to generate a large number of "pseudo-replicates". The basic idea in such a scheme is to draw two independent samples within each stratum. By combining the half-samples within each stratum with the half-samples in the other stratum, in all possible ways, it is possible to generate a large number of different sub-samples. This concept is also termed "inter-penetrating samples" by Deming (1960). For example, consider a stratified sampling procedure where two independent selections are made in each stratum, and where each stratum is assigned a weight $W$ when estimating the population parameters from the strata values. Let the population and sample characteristics be denoted as follows (for the reader not interested in the following explanation, you may skip ahead to the conclusion following equation (4.60) without loss of meaning):

Table 4.5 Sample Data for Half-Sample Replication Example

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Weight</th>
<th>Population mean</th>
<th>Population variance</th>
<th>Sample observations</th>
<th>Sample mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$W_1$</td>
<td>$Y_1$</td>
<td>$S_1^2$</td>
<td>$y_{11}, y_{12}$</td>
<td>$Y_1$</td>
</tr>
<tr>
<td>2</td>
<td>$W_2$</td>
<td>$Y_2$</td>
<td>$S_2^2$</td>
<td>$y_{21}, y_{22}$</td>
<td>$Y_2$</td>
</tr>
<tr>
<td>...</td>
<td>....</td>
<td>.....</td>
<td>.....</td>
<td>........</td>
<td>....</td>
</tr>
<tr>
<td>...</td>
<td>....</td>
<td>.....</td>
<td>.....</td>
<td>........</td>
<td>....</td>
</tr>
<tr>
<td>$h$</td>
<td>$W_h$</td>
<td>$Y_h$</td>
<td>$S_h^2$</td>
<td>$y_{h1}, y_{h2}$</td>
<td>$Y_h$</td>
</tr>
<tr>
<td>...</td>
<td>....</td>
<td>.....</td>
<td>.....</td>
<td>........</td>
<td>....</td>
</tr>
<tr>
<td>$L$</td>
<td>$W_L$</td>
<td>$Y_L$</td>
<td>$S_L^2$</td>
<td>$y_{L1}, y_{L2}$</td>
<td>$Y_L$</td>
</tr>
</tbody>
</table>

An unbiased estimate of the population mean $Y$, ($y_{st}$), is:

$$y_{st} = \frac{1}{L} \sum_{h=1}^{L} (W_h \cdot y_h) \quad (4.51)$$

The ordinary sample estimate $v(y_{st})$ of the variance of the mean $V(y_{st})$ is given by:

$$v(y_{st}) = \frac{1}{2} \frac{1}{L} \sum_{h=1}^{L} (W_h^2 \cdot S_h^2) \quad (4.52)$$

$$= \frac{1}{4} \frac{1}{L} \sum_{h=1}^{L} (W_h^2 \cdot d_h^2) \quad (4.53)$$

where $d_h = (y_{h1} - y_{h2})$.  

136
Sampling Procedures

Under these circumstances, a half-sample replicate is obtained by choosing one of \(y_{11}\) and \(y_{12}\), one of \(y_{21}\) and \(y_{22}\), \ldots, and one of \(y_{L1}\) and \(y_{L2}\). The half-sample estimate of the population mean, \((y_{hs})\), is:

\[
y_{hs} = \frac{1}{L} \sum_{i=1}^{L} (W_h y_{hi})
\]

(4.54)

where \(i\) is either 1 or 2 for each \(h\). There are \(2^L\) possible half-samples, and it is easy to see that the average of all half-sample estimates is equal to \(y_{st}\). That is, for a randomly selected half-sample:

\[
E(y_{hs} \mid y_{11}, y_{12}, \ldots, y_{L1}, y_{L2}) = y_{st}
\]

(4.55)

where \(E(y_{hs} \mid y_{11}, y_{12}, \ldots, y_{L1}, y_{L2})\) is the expected value of \(y_{hs}\) given a sample randomly drawn from the population consisting of \(y_{11}, y_{12}, \ldots, y_{L1}, y_{L2}\).

If one considers the deviation of the mean determined by a particular half sample, for example \(y_{hs,1} = \frac{1}{L} \sum_{h=1}^{L} (W_h y_{h1})\), from the overall sample mean, the result is obtained that:

\[
(y_{hs,1} - y_{st}) = \frac{1}{L} \sum_{h=1}^{L} (W_h y_{h1}) - \frac{1}{2} \frac{1}{L} \sum_{h=1}^{L} (W_h(y_{h1} + y_{h2}))
\]

\[
= \frac{1}{2} \frac{1}{L} \sum_{h=1}^{L} (W_h(y_{h1} - y_{h2}))
\]

\[
= \frac{1}{2} \frac{1}{L} \sum_{h=1}^{L} (W_h d_h)
\]

(4.56)

In general, these deviations are of the form:

\[
(y_{hs} - y_{st}) = \frac{1}{2} (\pm W_1 d_1 \pm W_2 d_2 \pm \ldots \pm W_L d_L)
\]

(4.57)

where the deviation for a particular half sample is determined by making an appropriate choice of a plus or minus sign for each stratum. In the example given above, each sign was taken as plus indicating that the first of the two stratum values was used to form the half-sample. The squared deviation from the overall sample mean is, therefore, of the general form:
where the plus or minus signs in the cross-product summation are determined by the particular half-sample that is used (plus signs correspond to using the first value in the stratum, minus signs indicate that the second value was used).

If the squared deviations of a half-sample estimate from the overall sample mean are summed over all possible half-samples, then it is possible to demonstrate that the cross-product terms appearing in the separate squared deviations cancel one another. Thus, for a randomly selected half-sample:

$$E[(y_{hs} - y_{st})^2 | d_1, d_2, \ldots, d_L] = \frac{1}{4} \sum_{i=1}^{L} (W_i^2 d_i^2) = v(y_{st})$$

Since $v(y_{st})$ is known to be an unbiased estimate of the true variance $V(y_{st})$, if one takes expected values over repeated selections of the entire sample, we have the result that:

$$E(y_{hs} - y_{st})^2 = \frac{1}{2} \sum_{i=1}^{L} (W_i^2 S_i^2) = V(y_{st})$$

Thus although we know that the $2^L$ sub-samples are not independent, because they contain many common elements, it is possible to eliminate the effects of these covariances by taking the average variance across all of the possible half-sample combinations. This, however, poses a potentially serious computational problem. The number of possible sub-samples is known to be $2^L$, where $L$ is the number of strata. In a typical sample survey project, the number of stratification cells could easily be between 20 and 40. In this case, the total number of pseudo-replicates would be between $2^{20}$ and $2^{40}$ (i.e. between about 1 million and 1 trillion). Clearly, the task of calculating the variance estimates for several million pseudo-replicates would be excessive. What is needed is a method by which we can obtain most of the benefits of pseudo-replication without the excessive computational costs involved. Two methods have been proposed for this purpose.

The first method is to choose a random sample of pseudo-replicates from the available population of $2^L$. Clearly, taking a larger sample of pseudo-replicates will more nearly eliminate the effects of pseudo-replicate covariance. Techniques have been derived to determine appropriate sample sizes under these conditions. A more elegant technique, however, employs the concept of balanced half-sample replicates.
Sampling Procedures

4.7.3.2 Balanced Half-Sample Replication

As noted above, taking different samples of pseudoreplicates will introduce variability into the estimates of half-sample variance because of the covariance between the half-samples. These covariances are represented by the cross-product terms involving $d_i d_k$ in equation (4.58). These cross-product terms cancel one another over the entire set of $2^L$ half-samples, or when one uses an "infinite" number of half-sample replications. The question then arises as to whether one can choose a relatively small subset of half-samples for which these terms will also disappear. If this can be done, then the corresponding half-sample estimates of variance will contain all the information available from the total sample.

A simple example will show that it is possible to select a subset of half-sample replications that will have the desired property. Consider a three-strata situation with observations $(y_{11}, y_{12})$, $(y_{21}, y_{22})$ and $(y_{31}, y_{32})$. There are $(2^3 - 1) = 7$ possible half-sample replicates. Now consider the following subset of four replicates:

<table>
<thead>
<tr>
<th>Replicate</th>
<th>Stratum 1</th>
<th>Stratum 2</th>
<th>Stratum 3</th>
<th>Deviation from Mean $(y_{h,i} - y_{st})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$y_{11}$</td>
<td>$y_{21}$</td>
<td>$y_{31}$</td>
<td>$(1/2)(+W_{1d1} +W_{2d2} +W_{3d3})$</td>
</tr>
<tr>
<td>2</td>
<td>$y_{11}$</td>
<td>$y_{21}$</td>
<td>$y_{32}$</td>
<td>$(1/2)(+W_{1d1} +W_{2d2} -W_{3d3})$</td>
</tr>
<tr>
<td>3</td>
<td>$y_{12}$</td>
<td>$y_{22}$</td>
<td>$y_{31}$</td>
<td>$(1/2)(-W_{1d1} -W_{2d2} +W_{3d3})$</td>
</tr>
<tr>
<td>4</td>
<td>$y_{12}$</td>
<td>$y_{21}$</td>
<td>$y_{32}$</td>
<td>$(1/2)(-W_{1d1} +W_{2d2} -W_{3d3})$</td>
</tr>
</tbody>
</table>

The signs of the separate terms in the deviations are determined by the definition of $d_i = (y_{h1} - y_{h2})$. It is, of course, immaterial how the two observations within a stratum are numbered originally. Once the numbering is set, however, as in the first replicate, it is maintained in determining the remaining replicates. If these deviations are squared, the first part of each expression is $W_1^2d_1^2/4 + W_2^2d_2^2/4 + W_3^2d_3^2/4$, which is the desired estimate of variance. The second part of each expression contains the cross-product terms, and it can easily be checked that all these cross-product terms cancel when the squared deviations are added over the four replicates. This follows from the fact that the columns of the matrix of signs in the deviations are orthogonal to one another. Thus this set of balanced half-samples can be identified as:

```
+  +  +
+  -  -
-  -  +
-  +  -
```

where a plus sign indicates $y_{h1}$, while a minus sign denotes $y_{h2}$. Notice that this particular set of replicates has the property that each of the two elements in a
stratum appears in half the samples. Thus, the mean of the replicates is an unbiased estimate of the mean of the population, and because of the nature of the "cross-product balance" the variance estimate is also unbiased and unaffected by the correlations inherent in the composition of the individual half-samples.

If one wishes to obtain a set of half-samples that will have this feature of "cross-product balance", for any fixed number of strata, then it becomes necessary to have a method of generating matrices of + and - signs whose columns are orthogonal to one another. A method is described by Plackett and Burman (1943-46, p.323) for obtaining $k \times k$ orthogonal matrices, where $k$ is a multiple of 4. Suppose, for example, that we have 5, 6, 7 or 8 strata. The Plackett-Burman method produces the following $8 \times 8$ matrix, which is the smallest that can be used for these cases because of the multiple-of-4 restriction. The rows identify a half-sample, while the columns refer to strata.

\[
+ - - + - + + - \\
+ + - - + - + - \\
+ + + - - + - - \\
- + + + - - + - \\
+ - + + + - - - \\
- + - + + + - - \\
- - + - + + + - \\
- - - - - - - - 
\]

Any set of 5 columns for the 5 strata case (or in general, n columns for the n strata case) defines a set of eight half-sample replicates which will have the property of "cross-product balance". If it is necessary to use the eighth column, the resulting set of half-samples will not have each element appearing an equal number of times. This will not destroy the variance estimating characteristics of the set of half-samples, but it does mean that the average of the eight half-sample means will not necessarily be equal to the overall sample mean. When the number of strata is a multiple of four, it may then be wise to use the next highest multiple of four as the number of half-samples.

Since orthogonal matrices of plus and minus ones can be obtained whenever the order of the matrix is a multiple of four, it is always possible to find a set of half-sample replicates having cross-product balance. It follows that the number of half-samples required will be at most four more than the number of strata. Consider, for example, if the survey design were to be based on 21 stratification cells. In such a case, it would be necessary to use 24 half-sample pseudoreplicates. A possible set of balanced half-samples for this particular case is shown below. This design was obtained by using the first 21 columns of the construction given in the Plackett-Burman paper. Any two columns of this design are orthogonal, and each element appears in 12 of the 24 replicates. The entire pattern is determined by the first column. The 2nd column is obtained from the 1st by
Sampling Procedures

moving each sign down one position and placing the 23rd sign at the top of the second column. This rotation is applied repeatedly to obtain the remaining columns. The 24th position is always '-' and is not involved in the rotation.

<table>
<thead>
<tr>
<th>Half Sample</th>
<th>Stratum</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>22</td>
<td></td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>23</td>
<td></td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>24</td>
<td></td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

The use of replication, random half-sample replication, or balanced half-sample replication provides a means whereby the variability in a sample estimate can be obtained from samples of any degree of complexity. It should be noted, however, that this estimate of variability can be obtained only after the sample has been drawn and the data has been collected. Therefore, replication methods cannot be used to assist in the determination of the sample size. However, replication methods can be used on existing samples of various designs to enable the calculation of design effects which can then be used in sample design as outlined in the previous section.
There are many other ways of estimating variance directly, including various forms of "jack-knifing" techniques (Brillinger, 1966; Efron, 1981, 1983). The essence of jack-knifing is to remove one observation from the sample and recalculate the mean of the resultant sample. By choosing which observation to remove in a systematic manner, akin to the orthogonal matrix technique described for replication, the mean and variance of the jack-knife samples can be made to provide unbiased estimates of the population parameters.

4.8 DRAWING THE SAMPLE

The final stage in the sampling process is the actual drawing of a sample from the sampling frame. In some cases (for example, with systematic sampling), the procedure is very simple and can be easily automated, either in the office or in the field (although care should be taken to ensure that the sampling procedure is adhered to in the field). In most situations, however, the sample should be drawn by reference to a random process.

Ideally, the process should be truly random i.e. independent events with each outcome being equally probable. However, the only true random processes are strictly physical events such as tossing a coin or rolling a die. In most situations, such random processes are too time consuming to be useful in sample selection. We must, therefore, resort to some form of "pseudo-random" process which can quickly and easily generate a set of random numbers for use in sampling.

Two forms of such random number generation processes are commonly used; look-up tables and recursive mathematical equations. The use of tables of random numbers is widespread and for this purpose several publications have compiled tables of random numbers. The most well-known of these is "Rand's One Million Random Numbers" (The Rand Corporation, 1955) although there are several other compilations (Owen, 1962; Kendall and Smith, 1939). In addition, most statistics textbooks contain a reproduction of part of one of the more complete compilations. A table of random numbers is included as Appendix B in this book.

In using random number tables to select a sample, the first step is to number all sampling units in the sampling frame. The order of this numbering is immaterial, and should be done for maximum convenience. The next step is to pick a starting point in the table of random numbers to be used (anywhere will do) and then systematically work through the table until the required number of random numbers has been selected. In sampling without replacement, the final sample of random numbers must contain no replications. In using tables of random numbers, any selection procedure is acceptable so long as it is systematic. For example, numbers may be read down or across the page, from left to right or right to left. In addition, numbers may be truncated in any way to obtain
Numbers in the desired range, e.g. if the tabulated numbers are in five digit groups, as in Appendix B, and the user wants random numbers between 0 and 99 then either the first two or the last two (or any other two) digits of the five digit group may be used. The numbers may also be modified systematically to obtain numbers in the desired range, provided that no bias is introduced. For example, if random numbers in the range 0 to 39 are required, the user could sample numbers between 0 and 99, discarding numbers greater than 39. This procedure, while not introducing bias, is, however, wasteful of random numbers. A more efficient procedure is to subtract 40 from those numbers in the range 40 to 79 and then use the result as a valid random number in the range 0 to 39. Those numbers from 80 to 99 would still need to be discarded because they do not have the same range as the required numbers and hence, if included (by subtracting 80), would bias the selection towards numbers in the range 0 to 19.

Sometimes, tables of random numbers are not readily available for use in the field. In such situations, it is useful to be aware of other, more readily available sources of random numbers. One source which is almost universally available is a telephone directory. Whilst not recommended for large scale surveys, a telephone directory does provide a useful source of random numbers if used with due care. Thus, for example, random numbers could be selected by taking digits in the right-hand columns of the telephone number. For most telephone directories a maximum of four digits should be selected from any one telephone number (otherwise the biasing effect of a limited number of telephone exchange codes can become significant). Within these restrictions, the principle is again to choose numbers systematically. For example, choose a page and column at random, and then read down the list of numbers in that column.

The major difference in using a telephone directory to obtain random numbers, as opposed to using a specific table of random numbers, is that whereas the table of random numbers has been checked for randomness before publication, the numbers obtained from the telephone directory carry no such guarantee. It is therefore necessary to ensure that the numbers obtained are indeed (approximately) random, by means of a series of simple checks. As a minimum, three tests should be conducted: first, plot the frequency distribution of the listed numbers and perform a goodness-of-fit test e.g. Kolmogorov-Smirnov test; second, calculate the mean and standard deviation of these numbers; and third, perform a runs test to check whether the listed numbers are ordered non-randomly. The use of these tests is described in Fishman (1973) and Knuth (1969).

Whilst the above methods of using "look-up" tables are convenient when one is drawing a relatively small sample, they become rather cumbersome to use when one is drawing a large sample, or drawing repeated samples, especially when using a computer. The storage space required to store tables of random numbers can be quite large, especially when large independent repeated samples are
required. For that reason, use is often made of truly "pseudo-random" numbers which are generated by a recursive mathematical equation. The use of an equation to generate random numbers may at first appear to be in direct contradiction of the concept of random numbers because each pseudo-random number so generated is completely determined by its predecessor and, consequently, all numbers are determined by the initial "seed" number. While this is true, the critical point is that the random numbers so generated can pass the statistical tests for uniformity and independence required of truly random numbers. They are therefore indistinguishable from real random numbers.

While there exist a number of different random number generator algorithms (see Knuth, 1969) the most common type is the linear congruential method. In this method, numbers are generated by an equation of the form:

\[ x_i = (ax_{i-1} + c) \mod m \quad (4.61) \]

where \( x_0 = \) an initially specified seed number

Expanding equation (4.61) such that the modulo notation \((\mod m)\) is removed, we obtain:

\[ x_i = ax_{i-1} + c - \left\lfloor \frac{ax_{i-1} + c}{m} \right\rfloor \cdot m \quad (4.62) \]

where \([\ ]\) = integer portion of value inside brackets.

To convert the integer number obtained from equation (4.62) to a random number within a smaller specified range \((A,B)\), the following transformation is applied:

\[ x_i(A,B) = \frac{x_i}{m} \cdot (B-A) + A \quad (4.63) \]

The critical factor in the use of such random number generators is to choose appropriate values of \(a\), \(c\) and \(m\). The selection of these values will depend on the computer being used to perform the calculations. Generally, each computer will have a random number generator routine with appropriate values of \(a\), \(c\) and \(m\) which have been found to give satisfactory statistical results on that computer. One problem often noted with pseudo-random number generators is that, because they are deterministically calculated, then once a number is repeated an entire sequence of repetitions must follow. Obviously no more than \(m\) different random numbers can be generated, although the period \(p\) can be substantially less than \(m\) with inappropriate choice of \(a\) and \(c\). Specific rules apply for maximising \(p\) (see Fishman, 1973). However, this problem of periodicity is more of a concern when using pseudo-random numbers in discrete event simulation modelling than it is in survey sampling procedures. Only when the sample size is
very large might periodicity become a problem (and even then it can be avoided by suitable choice of $a$, $c$ and $m$).

Following selection, and checking, of a set of random numbers by one of the above methods, it then remains to select those units on the sampling frame with the corresponding number and then include them on the sample list for use in the survey.
5. Survey Instrument Design

5.1 MORE TRADE-OFFS IN TRANSPORT SURVEY DESIGN

The survey process described in this book has different levels of precision at each of its many stages. On the one hand, sampling theory, which was described in the previous chapter, has been developed to a relatively high degree of sophistication such that, with due care, sampling error can be reduced to pre-specified acceptable levels of precision.

The same cannot be said, however, about the design of the survey instrument. The challenge of this stage of the survey process is dealing with the fact that it is much more of an art form than a science. Indeed, the design of survey forms fits very neatly the definition of an art form as "an established form of composition (e.g. a novel, sonata, sonnet, etc.)" (Concise Oxford Dictionary). Here we have an art form which, like a sonata, has rules which must be followed to ensure the best possible results, but which is free to adapt to the constraints imposed by the overriding objectives of the study and to the skills of the survey designer.

Nonetheless very similar types of errors which are discussed in Chapter 4 can apply when talking about survey instrument design. Our designs of survey instruments can result in errors of variability (since this make the results non-repeateable, it is rather like sampling error) and bias. In other words, our instrument design can result in:
data which gives unacceptable levels of variability when measuring the respondent’s real behaviour, or real attitudes (we will call this instrument uncertainty); and/or

• data which is not at all what we are trying to measure (we call this instrument bias).

The first of these errors - instrument uncertainty - occurs, for example, when poor survey instrument design leads a group of respondents who actually all behaved in the same way to report a variety of different behaviours. This mostly happens when bad design means that respondents understand the same question in different ways. For example, the designer may want respondents to report linked trips, but the poor (or untested) design means that some people report unlinked/staged trips (too many trips in this case), and some people report journeys (too few trips).

The second type of error which is a result of poor survey instrument design - instrument bias - occurs when the questions actually lead to the wrong answers. An example of this would be a questionnaire design which led to people reporting their travel of more than, say, 10 minutes to places which they went regularly (like to school and work) very accurately, but often forgetting all about small trips to irregular activities (like the corner shop). These examples of instrument bias (sometimes called systematic bias) are serious because, without careful testing, they often go completely unnoticed. In our example, the results may show that very few people make short trips within their local areas, with the result that there may seem to be no need to improve pedestrian or bicycle facilities; an improvement which might have encouraged the use of these facilities - possibly giving economic, environmental and even social benefits to local communities.

Before proceeding with a detailed discussion of survey instrument design, it may be useful to review briefly the marksmanship analogy from Chapter 4 in the current context (Figure 5.1). Again, the first of the two sources of error - instrument variability - is analogous to imprecision. This occurs when the survey instrument asks the same question and gets a range of answers even when people have actually behaved in the same way.

On the other hand, if people are answering in the same (but incorrect) way (always telling us about longer, more regular trips), the resultant instrument bias means that we may be recording very precise (repeatable) data, but unless we carry out careful testing, we are not sure if the data is actually accurate!
Figure 5.1 Questionnaire Imprecision (Variability) and Inaccuracy (Bias)

Unless the measurements being made are valid, it does not really matter how precise those measurements are.

Relative to sampling error (uncertainty) and sampling bias (Chapter 4), very little research has been carried out on instrument uncertainty and bias. One of the main reasons for this is that instrument errors are not often obvious - either to the survey designer or to the analyst. While there are no fixed rules for designing a survey instrument, there are sound principles which have been derived from some controlled experiments as well as from recorded experience. Application of these principles, in conjunction with a good deal of common sense pertaining to the problem at hand, should result in a design which actually measures what the investigator set out to measure.

There are probably two important reasons why the art of survey instrument design has not progressed as far as that of sample design. When we design survey instruments we are often reluctant:

- to test the instruments; and
- to admit that someone else may understand the questions we design in a different way than we do ourselves.

It is possibly the latter point that makes us so reluctant to execute the former. This chapter is designed to make you more willing to test survey instruments by
Chapter 5

giving you an insight into the rewards of better understanding human behaviour (travel behaviour, in particular, in our case), and hence being able to be confident about the survey results you get - not just dismissing anomalies as "survey problems", but accepting them as being extremely likely to represent actual travel behaviour and attitudes.

5.2 SCOPE OF THIS CHAPTER

Throughout this chapter we will refer to "questionnaires". More correctly we should continue referring to "survey instruments", since the form of the questionnaire will vary depending on the survey method being used. Thus self-completion surveys, telephone interviews, household interviews, and interactive group interviews will all require some means by which the data is to be recorded. However, the details of these data recording techniques will vary considerably.

Despite the many different types of questionnaires, survey instrument design techniques can essentially be divided into:

- those which design for completion by the respondent and
- those which design for completion by a trained person (interviewer).

The major difference between these two methods is that, in the former, the respondents read and answer the questions by themselves whereas in interview surveys, the respondent is assisted by an interviewer who, by and large, reads the questions and records the answers. Obviously, there is a much greater onus on clear questionnaire design when an interviewer is not present. This is because, while it is possible to train a small number of interviewers to deal with considerable complexities in survey design, it is clearly not possible to expect the same of a large number of unknown respondents. However, since interviewer bias (all error associated with the use/presence of an interviewer) can be exacerbated by poor questionnaire design and layout, making it difficult for interviewers to conduct a survey without error, even personal interview questionnaires need to be designed with great care.

As with other components of the survey process, it is instructive to examine questionnaire design within the context of a number of specific factors which are important for the overall task. With respect to questionnaire design, the principal factors which will be addressed are:
(a) Questionnaire content;
(b) Physical nature of forms;
(c) Question types;
(d) Question format;
(e) Question wording;
(f) Question ordering;
(g) Question instructions.

The aim of this chapter is to present a number of general principles, and some specific recommendations, with respect to each of these key factors.

5.3 QUESTIONNAIRE CONTENT

This section deals with the issue of deciding exactly what information needs to be collected by a transport survey instrument. The discussion centres around the quantity of data being collected and therefore focuses on the length of the survey instrument. In addition, since gathering information on travel behaviour is often an integral part of a transport survey, particular emphasis is given to ways in which this type of data (about trips) is best collected.

In deciding on what information to include in the questionnaire, there are three basic guidelines:

- the data must be relevant to the purpose of the survey;
- the data must be reliable, i.e. the same results would be gained if the survey was replicated - this minimises instrument uncertainty;
- the data must accurately represent what is being examined - this minimises instrument bias.

5.3.1 Length of the Questionnaire

Already in the preliminary planning stage (Section 2.7), we should have constructed a wish-list of survey content. The task at this stage is to whittle down this list to those items which are particularly relevant. Specifically, the survey designer should now derive an explicit rationale for each item in the survey, covering not only why the information is needed but also how the data obtained is to be analysed. As noted in Chapter 1, this requires a backward linkage from the coding and analysis phases of the survey process. Oppenheim (1992) suggests that one way of testing the adequacy of the design (including content of the questionnaire) is to run through the natural sequence of survey stages in reverse.
order. For example, if we expect at the end of the survey to be able to show whether men make more trips to shopping than women, we would need some cross tabulations which related gender to trip rate. At this point we would have to draw up some dummy tables, showing the relevant variables cross-tabulated with certain sub-groups, for example, gender. In order to generate these tables we must have asked questions about trip making as well as about shopping (at the level of detail required in the cross-tabulations) and we must also know the sex of each of our respondents.

Having identified the need for various items of data, the final selection of items will be the result of a trade-off between:

- the expressed needs (i.e. the wish-list of questions to be covered);
- the survey resources (in terms of money, personnel and so on); and,
- the effect of survey length on response rates and validity of responses.

While there are always exceptions to the rule, it is generally agreed that, once above a critical-mass survey length, longer questionnaires and/or interviews result in poorer response rates, completion rates and/or response validity. Thus fewer people respond to longer questionnaires and, if they do respond, they may not complete all the questions. Even if they do complete the questions, the validity of responses to questions at the end of the questionnaire may be dubious.

An interesting way of asking more questions than would usually be possible in an on-board-survey is described by Sheskin and Stopher (1982b). Here they used the duel survey mechanism described in Section 3.2.5, which collected some information via self-completion on-board a bus, and gave respondents a further questionnaire to take home and mail back. This resulted in a significant increase in response rates, and hence decrease in sample bias.

The issue of trading off the wish-list vs. survey resources vs. effect of length presents an interesting idea for the way in which an experimental design may be introduced into a pilot study. For example, it would be possible to test the effect (in terms of resources and response rates) of having, say, 50 questions vs. 20 questions. This could be done by testing two questionnaires in parallel - one with the 50 "dream" questions and one with the 20 "absolutely essential" questions. Analysis of the results would then indicate any variation in response rates between the two designs, and, together with a comparison of the costs of each method it would be possible to make an informed decision on which method to use.
Survey Instrument Design

In economic terms, what one is faced with is an elasticity of response with respect to questionnaire length. The decision to be made is whether increased questionnaire length will provide more or less useful information overall (compare this problem with that of a transit operator who wishes to know whether increasing or decreasing fares will increase revenue). While better questionnaire design will improve response (for any length questionnaire) there is always a limit to the amount of information which can be sought in any questionnaire.

Given that the desired amount of information will often be greater than the possible amount of information which can be collected (based either on a response rate or resources argument), it is necessary to assign priorities to the items of information. Thus the items should be ranked in order of importance to the study in question (realising that some items of information may be complementary). The final selection should then be made in accordance with the available resources and the diminishing marginal value of extra items of information.

5.3.2 Relevance of the Questions

In making this final selection of which questions to include in the survey, two vital factors should be considered. First, in dealing with human populations the information sought should not only be relevant to the study purposes but should appear to be relevant to the respondent. Special care (and pilot testing) needs to be used because, to us as survey designers, it is always perfectly obvious why a question is included. After all, we have done the preliminary planning, set our objectives and even worked backwards to ensure that the questions we ask are the right ones for the analysis. Hence we know, for example, that the "usual activity" of a respondent (such as whether the respondent works, studies, is retired, etc.) is an important variable influencing travel behaviour. However, if the respondent sees (or hears) a questionnaire on travel behaviour beginning with a question about what they usually do (without explanation), they can certainly be excused for querying the relevance of this question to a travel survey. (A simple way of dealing with this particular case is shown in Figure 5.16.)

If questions are not perceived to be relevant then a number of adverse effects are possible. At the very least, the respondent may be annoyed at having to answer "irrelevant" questions. This may lead, in a personal interview survey, to a diminution in rapport between respondent and interviewer, and hence have a lasting adverse effect on the interviewer. In later surveys, the interviewer may omit or rephrase the question (perhaps thereby changing the intent of the question) in order to avoid annoying the respondent. More seriously, "irrelevant" questions (especially when their purpose is not adequately explained) can create a
mistrust of the stated survey purpose and this may well lead to spurious and inaccurate answers from a wary respondent. In the long term, such mistrust can only be to the detriment of all sample survey efforts as respondents become reluctant to respond to any such surveys. It is therefore wise to restrict questions to those which are, and can be explained to be, relevant to the immediate survey purposes.

Of course, this reinforces the importance of ensuring that all survey staff who are likely to have any dealings with respondents (primarily interviewers and those people answering phone queries), know a great deal about the purpose of the survey. This is discussed further in Chapter 7.

5.3.3 Reasonableness of the Questions

Another factor to consider in defining the questionnaire content is whether it is reasonable to expect respondents to be able to answer questions we ask of them. This difficulty applies equally to questions of opinion, knowledge and fact. It should, however, be considered that, in general, respondents find questions of fact "easier" than questions of opinion - since one is stating what has already occurred, while the other involves consideration of, or speculation about, an issue (which may or may not have been done by the respondent). Sections 5.6 and 5.7 discuss this in more detail.

It is unwise to assume that respondents will voluntarily admit ignorance. On the contrary, they (like many of us!) will generally attempt to give some answer even if they are ill-informed or have never thought about the subject matter before. If first impressions or perceptions are all that the investigator is interested in, then this may pose no serious problem. However, if the answer is to be interpreted as being a considered response or a correct statement, then respondents should be given ample opportunity to admit their ignorance without fear of recrimination.

5.3.4 The Context of Questions about Trips

Many (perhaps most) surveys which gather information for transport planning have data on people's trips as an "absolute essential" component of their wish-list for questionnaire content. For this reason it is important to understand the context in which people can be asked about trip-making.

Day-to-day travel behaviour can be gathered in two contexts:

1. Travel-only context - people can be asked to report details about their travel only. This will almost always give information on the purpose of travel, which is usually an activity (see Preface), but no further information on activities is obtained.
Survey Instrument Design

(2) *Activity context* - here people are asked to report all activities in which they take part - both at home and away from home. Travel will be a natural part of these activities, and in this context is often seen as an activity in itself. Research has indicated that this context not only gives data on activities, but also results in a much more accurate recording of travel (Clarke, Dix and Jones, 1981).

At the beginning of designing a survey instrument in which travel data is to be collected, it is essential to determine which of these approaches will be adopted. Each requires substantially different question types (Section 5.5) and each can be done using different survey methods (Chapter 7). A further discussion of these details appears in the relevant sections of this book.

5.3.5 Questionnaire Design to Maximise Trip Recording

Whichever of the above contexts are used in a travel survey, it is necessary to realise that travel outside the home can be recorded either by asking respondents to recall what happened at a past time (recall technique), or by announcing to respondents in advance that they will have to report travel about a future time (announce-in-advance technique). Both of these methods have variations as well as advantages and disadvantages.

The recall technique, by and large, generates the greatest error in reporting of actual travel. This is particularly the case when respondents are asked simply to report travel-only for a period of time in the past (Figure 5.2).

To test these problems for yourself, try remembering in detail where/when/how you travelled two days ago, or even yesterday! Early household travel surveys used this technique almost exclusively. There are two ways to assist with getting better data from the simple recall technique. First, respondents could be assisted with prompts such as "Where did you go next?". Another improvement would be to ask respondents to think in terms of activities and not just trips. In either case, however, there remain severe disadvantages in accurate reporting of trip data due to forgetting of travel.
The announce-in-advance technique has been shown to improve trip reporting considerably. The methodology involves contacting the respondent in one way or another prior to the "Travel Day" (the day about which trip reporting should occur). This means that respondents are more likely to either record travel as it is made or at least to be alert to the fact that they will have to report their travel, and thereby pay more attention to details on the Travel Day. The methods do not totally exclude the possibility that respondents will sometimes fail to record their travel, meaning that recall is their last resort, but it is generally limited significantly by the survey method used.

This method gives respondent the opportunity to actively remember their travel patterns and even to take notes on what they do. The latter option has led survey researchers to accompany the announce-in-advance method with a diary of some type to facilitate the note-taking behaviour. In personal interview surveys, a very brief diary (memory jogger) has been left at the pre-travel day visit (Ampt, 1981), and in self-completion surveys, although the questionnaires themselves are a kind of diary, some researchers (notably Stopher, 1992) have sent an additional memory jogger to stimulate even better travel recording. Both of these methods are usually accompanied by further assistance to ensure that travel and activities are not forgotten. The self-completion method can sometimes includes specific reminders such as "Did you go anywhere else after this?" (Figure 5.3) which helps to reduce the chance that major trips will be forgotten.
The personal interview method, on the other hand, can use prompts such as "And what did you do next?" (even while the respondent is referring to a completed memory jogger) (Figure 5.4). This method is appropriate only for the personal interview method, since while asking respondents to think in the framework of activities ("The next thing I did was bake a cake.") the activities are not recorded - only the next trip ("..and I found I was out of flour, so I walked to the shop"). Just the walk to the shop would be recorded.
Possibly the most thorough method of collecting comprehensive activity and travel data uses the announce-in-advance method, and gives respondents activity diaries (Figure 5.5) to carry with them throughout the day. As well as improving the recall of trips, activity diaries can also provide basic information on activity patterns which can be used in assessing potential trade-offs between travel and activities. The advantage of diary techniques is that they rely very little on the recall of past events (provided the dairies are filled in regularly). On the other hand, they require a greater degree of cooperation from respondents and there is also the possibility that the mere fact that a diary has to be filled in will affect the phenomena that are being measured. Certainly, in activity diary surveys, there must be an entry each day entitled "Filled in activity diary!"

In this method, which is sometimes called the "verbal activity recall framework" (Ampt, 1981), relatively minor trips which involve returning immediately to the same location (e.g. going to a local shop, going to lunch from work) are less likely to be forgotten as a coherent activity pattern is established.

![Activity Diary Example](image)

**Figure 5.5  Example of an Activity Diary**
Source: Jones et al., (1983)

Notwithstanding possible problems with the mechanics of diary completion, the concept of placing trips within the total context of a daily travel or activity pattern is one way of effectively minimising the problem of respondents not being able to answer questions accurately through lapses of memory.
5.4 PHYSICAL DESIGN OF SURVEY FORMS

An often overlooked aspect of questionnaire surveys is the physical nature of the form on which the data is to be recorded. Careful attention to this matter, however, can often lead to more efficient job performance by respondent, interviewer and data enterer. In addition, in self-completion surveys, an attractive, professional appearance of the survey form will always lead to higher response rates. In fact, for self-completion surveys, the overall appearance of the survey form is of vital importance since it is the only point of contact with the respondent. While the cost of producing a high quality survey form is obviously higher than for a single photocopied sheet of questions, it is money well spent, especially if the interest of the respondent needs to be aroused.

While more attention needs to be given to the physical nature of forms in self-completion surveys, it can never be disregarded altogether - even in personal interview surveys. The following guidelines will help in the design of forms for personal interview surveys and self completion questionnaires; there are some special notes for intercept surveys.

For personal interview surveys, attention needs to be given to the following points:

(a) It is first of all necessary to determine whether the respondent or the interviewer will actually fill out the form. If the respondent is to do the writing, then the form will need to be designed more as a self-completion questionnaire, since the interviewer is merely there to give assistance in the interpretation of questions. Normally, however, the interviewer reads the questions and records the answers on the form. It is in this situation where the following comments are more applicable.

(b) Generally, the form should require a minimum amount of writing. If the interviewer is required to do a lot of writing when recording responses, then it is quite likely that the attention of the respondent will be lost while they wait for the next question to be asked.

(c) The interviewer should be provided with a separate list of instructions for each question so that they may guide the interview and provide interpretations of questions to respondents. Generally, these detailed instructions do not need to be included on every copy of the interview form since they should be contained in the Interviewers’ Manual (Section 7.3). However, brief reminders of instructions and specific prompts can be included at appropriate points within the form.

(d) The interviewers should be trained to give an introduction which explains the purpose and background of the survey. Usually ample training will
mean that it is not necessary to write this on the questionnaire form, although notes may be useful as reminders of points to cover.

(e) The form should contain a detailed flow chart of the sequence of questions, by means of arrows or Go to instructions, especially when there are many filtering and branching questions.

(f) Each form should be identified by a unique identification number to enable records to be kept of the status of that unit in the sample, e.g. has interview been completed, refused or not yet attempted.

(g) It is extremely useful if each interview form is enclosed in a cover sheet for office and interviewer use only. It is used for administrative control and for recording selected data which the interviewer can obtain by observation. The cover sheet normally serves a range of functions such as:

(i) helping the interviewer to locate the sample household.
(ii) checking the accuracy of information about sample households.
(iii) maintaining a precise record of what happens at each sample household. Times and dates of successful and unsuccessful contacts should be recorded.
(iv) recording supplementary data, such as type of household and quality of surrounding environmental conditions.
(v) providing a space for recording response reports.
(vi) recording appointment times

The cover sheet is ideally made of heavier quality paper than the rest of the questionnaire.

(h) All questions should be numbered consecutively throughout the form with no omissions or repetitions. This applies even when the interview is broken up into several discrete sections since consecutive numbering facilitates easy and non-ambiguous cross-referencing and branching control.

(i) Different type faces or fonts should be used for different elements of the interview form to facilitate easier administration of the survey by the interviewer (Figure 5.6). Three elements, in particular, should be segregated:

(i) Instructions to the interviewer;
(ii) QUESTIONS TO BE READ VERBATIM TO THE RESPONDENT;
(iii) Coding categories for the recording of responses.
For self-completion questionnaire surveys (with no interviewer), the survey form takes on extreme importance and should be the subject of extra attention in design, wording, and layout. Some general guidelines include:

(a) The overall layout should be clear, concise and, in general, should lead respondents to the next question. In this respect, arrows from branching questions can be a useful device, provided that there are not too many of them; too many arrows tend to confuse rather than assist.

(b) A minimal amount of writing should be required. Questions should require a "tick the box" reply if at all possible (Figure 5.7). Many people rarely need to write in their day-to-day lives, so that it can be quite threatening to have to complete a survey form. In addition, many people

---

**Figure 5.6  Sample Page of Personal Interview Survey Form**
Source: Ampt (1992b)
with a less than perfect grasp of a language who are quit happy to *speak* it, find it embarrassing to write down their mistakes for perpetuity (as they perceive it).

(c) A short, non-technical summary of the aims of the survey should be included to increase respondent interest in the survey.

(d) Include general instructions on how to fill out the form at the start of the questions. If an example completed form is provided, refer to it in the introductory comments.

(e) Any detailed instructions for particular questions should be attached to the questions to which they refer.

(f) Assurances of confidentiality should generally be stated in the introduction, but should not be over-emphasised. Too much emphasis, especially when the survey covers material of a not very confidential nature, may have the counterproductive effect of arousing people's suspicions.

(g) The questionnaire should be endorsed and signed by someone with authority so as to lend credence to the survey effort. Just who this person should be varies with each survey (See Section 7.1). This is usually done on an accompanying letter.

(h) There would always be a phone number for respondents to ring about questions of clarification or legitimacy of the survey. The phone needs to be manned on evenings and weekends as well as weekdays, since that is the time respondents will usually want to ask questions.

(i) The form should be as small as possible, consistent with clarity, legibility, and sufficient space for recording of answers. For mail-back questionnaires, attention should be given to the approved maximum sizes for various postage rates. To maximise efficient use of space, consideration should be given to printing the introduction and instructions on a portion of the survey form which can be torn off and not returned through the mail. Going one step further, the questions can also be printed on the tear-off portion so that all that is returned in the post is a postcard containing the answers. This maximises the ratio of information received to return postage charges paid.

(j) Using two sides of a sheet of paper is better than using two sheets of paper to ensure that the questionnaire does not look longer than it actually is. However, if questions are printed on both sides then care should be taken in the selection of paper stock to ensure that it is
Survey Instrument Design

relatively opaque. For single sheet surveys, the use of good quality card stock is recommended.

(k) The survey form should look professional and be printed in clear, easily readable type face. Professional artwork design is recommended for all questionnaires. The use of desktop publishing and object-oriented graphics programs has greatly facilitated the production of professional-looking survey forms in recent years. For major surveys, the use of multi-colour printing has been found to be a cost-effective way of increasing response rate. An example of the type of survey form which is relatively easy to produce these days is shown in Figure 5.7. This trip form was part of a package of survey forms produced by the authors for a self-completion questionnaire travel survey, and were loosely based on the KONTIV survey form design described by Brög and colleagues (Brög et al., 1983; Brög et al., 1985). Such forms can be produced using object-oriented graphics programs which are becoming more and more common. They can then be colour printed directly from the final camera-ready copy. The use of readily-available graphics programs in this way greatly facilitates the preparation of draft versions of the questionnaire for committee discussion, and enables changes to be easily made right up until the day before the questionnaire forms go to the printers.

Bear in mind that even the advertising you receive in your mail box looks very professional! Who will complete a less than professional looking survey form?

(l) In summary, self-completion questionnaire survey forms should be designed to encourage every respondent to reply - whether they are used to filling out forms or not.
Figure 5.7  Example of Self-Completion Questionnaire Survey
Source: Richardson and Ampt (1993a)

For *intercept surveys* (e.g. public transport patronage studies), the following points need to be given special consideration, since often the questionnaires will be handed out or filled in on a moving vehicle:

(a) The form should be ergonomically designed. This means that ample space should be left for writing (if this is necessary) and data should not
Survey Instrument Design

need to be recorded towards the edges of the sheets (which is difficult with no firm support for the survey form).

(b) The forms should be weather-resistant if necessary (for example if they are being done at the roadside) and, if rain-affected should still be able to be filled in with pen or pencil. Note that instructions may need to be given to interviewers about the types of pen or pencil which are permissible. (Some tips! Felt-tipped pens smudge in the rain, ballpoint pens do not work at temperatures below 3°C, and pencils need lots of sharpening!)

(c) The forms should be of convenient size and format such that, for example, pages are easy to turn, and that the forms can be used by both right-handers and left-handers.

(d) The cardinal rule is for the investigator to test the survey form under actual, or simulated, field conditions before adopting the form for final use. The form should be tested under various conditions and especially when the survey workload is at its highest (e.g. with very full buses).

(e) The rule about the minimal amount of writing required is particularly important in intercept surveys. Questions should require a "tick the box" reply if at all possible. This is particularly important for on-board public transport surveys where vehicle movement can make writing extremely difficult. Another small but important point with on-board questionnaires is to provide the respondent with a pen or pencil to fill in the survey form. The pen could then be given to the respondent as a "freebie".

(f) Although response rates are best when intercept surveys are collected immediately, always provide a mail-back option for people who are genuinely in a hurry, have forgotten their reading glasses, or who simply cannot fill in the form while holding a brief case in one hand, and a child in the other.

5.5 QUESTION TYPES

In constructing a questionnaire there are three basic types of questions which may be asked: classification, factual and opinion questions.

5.5.1 Classification Questions

Classification questions are those questions which need to be asked in order to obtain a basic description, or classification, of the respondent. Such questions usually relate to socio-economic and demographic characteristics of the
respondent, e.g. age, sex, income. Responses to classification questions are typically used to form sub-groups within the sample for later analysis or cross-checking with secondary data sources.

Classification questions can often be used to perform a screening function whereby members of certain sub-populations are either included or excluded from our final population. For example, questions relating to car ownership could eliminate car-owners from a population for a study of the transport disadvantaged. Such screening questions would be asked at the very beginning of the interview (perhaps when contact has first been made with the respondent at the door). The use of screening questions in this way can be a useful method of locating members of rare populations, when no sampling frame list for this population is available. Thus a totally random sample is first selected, and then only those passing the screening test are retained in the final sample.

Classification questions can also be used as branching questions within the interview, so that respondents are only asked questions which are relevant to them, e.g. non-car-owners are identified so that they are not asked detailed questions about car ownership (see Figure 5.8).

Since many classification questions may be seen by respondents as being somewhat tangential to the main purpose of the survey, it may be wise for the interviewer to explain why such details are needed (generally along the lines of comparing results for different groups in the population). Also because some respondents may find the classification questions intrusive, it is often a good plan to have categories decided upon before the interview (on the basis of prior knowledge) and have the categories printed on cards. The interviewer can then simply present the card to the respondent and say "which of these groups applies to you?"

Classification data is very important when the dataset is likely to be used for secondary analysis - so that the new investigator can identify sub-populations within the original dataset. In these circumstances, it makes very good sense to use classification categories which are standard across many surveys, e.g. asking respondents the income categories used in the Census questions, since it is very likely that Census data will be used as secondary data at a later stage.
5.5.2 Factual Questions

Factual questions deal with the respondent’s experiences and knowledge and are the types of questions which are most suited to self-completion questionnaire surveys. Two points are of particular importance when shaping factual questions. First, it is necessary to ensure that the definitions of words and phrases used in the question (e.g. trip) are completely clear to the respondent so that the "facts" given by them are indeed the facts which we seek. The accuracy of the information received can only be as great as the clarity with which the question is asked. Second, one should ensure that it is reasonable for the respondent to be able to provide the facts we require.

The effect of memory lapses with respect to trip recall was discussed earlier as an example of this problem. One should also be careful in this regard when asking respondents to reply on behalf of other members of the family (i.e. to provide "proxy" responses). All that can be expected is the respondent’s perception of the "facts" as they apply to the other family members.

One useful point to realise about factual questions in interview surveys is that probing can be used by the interviewer without great fear of biasing the responses it is even possible to rephrase the question if necessary. The successful use of such strategies depends largely on the amount of training given to interviewers. If they are made fully aware of the scope and intention of such factual questions, then they can appraise the quality of the response and judge whether further clarification through probing is required.

At the outset, it is necessary to consider that reporting travel is essentially reporting factual information - i.e. what actually happened rather than opinions or attitudes. Since the respondent actually takes part in their own travel, the inability to answer factual questions accurately in this case could not be due to ignorance per se. If the questions are asked about past travel behaviour or trip making, it is much more likely to be due to the fallibility of human memory.
Past studies have consistently shown that, in surveys where respondents are required to recall the trips made at some time in the past, they do forget trips they have made (Clarke, Dix and Jones, 1981; Meyburg and Brög, 1981; Stopher and Sheskin, 1982). This leads to the phenomenon of under-reporting of trips, meaning simply that the number of trips reported is less than the number which actually occurred.

5.5.3 Opinion and Attitude Questions

In contrast to factual questions, opinion and attitude questions seek to obtain the opinions and attitudes, rather than the knowledge, of respondents. Because of this, they are much more sensitive to the wording and type of probing used. Questions of this type should be specifically identified on personal interview questionnaires as a reminder to the interviewer that they, like all other questions, must be asked verbatim. The difference between opinions and attitudes is somewhat difficult to define and there is, in fact, a large body of literature which addresses the topic (e.g. Pratkanis et al., 1989). For ease of understanding, some social psychologists make a rough distinction between different levels of a person's philosophy, calling the most superficial level "opinions", the next one "attitudes", a deeper level "values" or "basic attitudes", and a still deeper level, "personality" (Oppenheim, 1992). There are certainly relationships and patterns of connections between all of these layers. Usually the most important thing for the survey designer is to discover the way in which any of these levels is likely to affect behaviour.

Moser and Kalton (1979) suggest that opinion questions merely seek to determine whether a respondent agrees or disagrees with a given opinion statement. The Gallup Poll is an example of a collection of opinion questions. Attitude questions, on the other hand, often form a battery of coordinated opinion questions which attempt, through specific psychological theories, to form an assessment of the respondent's overall attitude towards a particular subject. The techniques of attitude measurement are highly developed and cannot be covered in detail here (see Golob (1973) and Tischer (1981) for detailed descriptions of attitude measurement). However, a few general comments on the types and use of opinion or attitude questions are warranted.

Unidimensional vs. Multidimensional Attitude Data

The collection of attitude data, particularly with respect to attitudes towards transport options, can proceed under one of two assumptions. The first states that each attribute of an alternative can be separately assessed in terms of its relative importance and its degree of satisfaction. These unidimensional attitude ratings can then be combined in some manner to obtain an overall attitude rating for the alternative.
Survey Instrument Design

The second assumption follows in the Gestaltist tradition and states that an alternative can only be assessed in its totality. Thus attitude ratings can only be obtained for the alternative and not for the individual attributes. Information integration theory (Anderson, 1974) is one example of the use of multidimensional attitude data.

The vast majority of attitude data used in transport choice analysis, and elsewhere, is of the unidimensional type. The majority of discussion in this section will therefore concentrate on unidimensional attitude data, although some examples of multidimensional attitude measurement techniques will be given in this section, and will be explored later when discussing stated preference survey methods (Section 5.3.4.).

Profile vs Similarities Data

A further general distinction which must be made in relation to attitude data is between profile and similarities data. Profile data characterises one or more alternatives in terms of one or more attributes. Thus, for example, several alternatives may be rated in terms of the satisfaction gained from several attributes.

Similarities data, on the other hand, attempts to show how alike two alternatives are seen to be, either in general or with respect to a specific attribute. Both profile and similarities data may be collected on either a unidimensional or multidimensional basis.

Types of Measurement Scales

Before we discuss several specific examples of attitude measurement techniques, consider the types of scales which may be used. There are essentially four different types of scales which may be constructed:

- Nominal Scales
- Ordinal Scales
- Interval Scales
- Ratio Scales

Nominal scales serve simply to categorise people and objects into groups. The codes attached to each group do not imply any ordering. Examples of such scales include categorisations by sex and occupation. The only information obtained from a nominal scale is that objects with the same rating belong to the same category.

Ordinal scales serve not only in a classification manner but also impart order to objects rated on such a scale. If, for example, three objects (A, B and C) were rated on an ordinal scale, then the only information available on these objects,
with respect to the characteristic in question, would be their order. Hence it could be stated, for example, that A > B > C but nothing would be known about the relative size of the difference between A and B and between B and C. An example of an ordinal scale is the Mor's Hardness Scale which simply states that materials with a higher number on the scale are harder than (i.e. will scratch) materials with lower numbers on the scale. Another example is the ranking procedure for destination zones as used in the intervening opportunities model of trip distribution. All that is known is that one zone is more attractive than another; how much more attractive is not known.

One consequence of using an ordinal scale is that it is impossible to perform any mathematical operations on the scale numbers (except ordering). For this reason, they are of limited usefulness as inputs to mathematical models of travel choice.

**Interval** scales impart both order and position to objects rated on such a scale. That is, they not only rank objects but they also give meaning to the distance between object ratings on the scale. Thus, relative values on the scale possess some meaning. However, there is no meaning attached to absolute values on the scale. This is because interval scales possess no fixed zero point; rather the selection of a zero point is completely arbitrary. Examples of interval scales in common usage are the Fahrenheit and Celsius temperature scales (with a zero point determined by the freezing point of water), and the Gregorian calendar of years (with zero point determined by the birth of Christ).

As a result of the absence of a fixed zero point, it is not possible to multiply or divide with interval scale numbers although it is permissible to add and subtract such numbers.

**Ratio** scales impart order and length to an object rating and also have a determinate zero point which enables all four mathematical operations to be performed with ratio scale numbers. Thus the distance between two ratings has a meaning as does the ratio of two ratings. Examples of ratio scales are the Kelvin scale of temperature, the Decibel scale of loudness and numerous other physical scales such as length, weight and duration. A summary of the properties of the four types of scale are presented in Table 5.1.
Survey Instrument Design

Table 5.1 Summary of Scale Types

<table>
<thead>
<tr>
<th>Scale Type</th>
<th>Central Tendency</th>
<th>Variability Measure</th>
<th>Individual Position</th>
<th>Permissible Uses</th>
<th>Permissible Transformations</th>
</tr>
</thead>
<tbody>
<tr>
<td>RATIO</td>
<td>Geometric Mean</td>
<td>Coefficient of Variation</td>
<td>Absolute Score</td>
<td>Find Ratios Between</td>
<td>Multiplication and Division</td>
</tr>
<tr>
<td>INTERVAL</td>
<td>Arithmetic Mean</td>
<td>Variance, Standard Deviation</td>
<td>Relative Score</td>
<td>Find Differences Between</td>
<td>Addition and Subtraction</td>
</tr>
<tr>
<td>ORDINAL</td>
<td>Median</td>
<td>Range</td>
<td>Rank Percentile</td>
<td>Establish Rank Order</td>
<td>Any that Preserve Order</td>
</tr>
<tr>
<td>NOMINAL</td>
<td>Mode</td>
<td>Number of Categories Belonging to Category</td>
<td>Identify and Classify</td>
<td>Substitution within Category</td>
<td></td>
</tr>
</tbody>
</table>

Note: The higher level scales subsume all the features of the lower level scales.

**Attitudinal Measurement Techniques**

Given these general definitions of scale types, consider some specific examples of attitudinal scaling techniques which are of use in transport choice analysis. The techniques discussed in this section are all unidimensional scaling techniques. Multidimensional techniques are discussed in the next section.

**Paired Comparisons**

Assume that there exists an ordering to a set of objects. One way of determining such an ordering would be by comparing all objects two at a time and noting the higher order object on each occasion. After carrying out all possible comparisons, the highest order object should have been selected every time it appeared in a comparison. The second-highest order object should have been selected every time it appeared except for when it was compared with the highest order object. This finding can be extended to all lower order objects until the lowest order object which is never selected in any comparison. The number of times an object is selected will constitute an ordinal scale since it imparts order but not separation distance to the objects on that scale (i.e. all objects will be equally spaced on this scale). The concept of a paired comparison test has been extended to produce an interval scale by Thurstone (1959) based on his famous Law of Comparative Judgement (Thurstone, 1927). This law states that a stimulus - whether physical or otherwise - gives rise to a perceptual response within an individual which, for various random reasons, varies from presentation to presentation of the same stimulus as shown in Figure 5.9.
Figure 5.9  Distribution of Response to a Single Stimulus on repeated Occasions

If the individual is presented with a different stimulus this too will result in a distribution of responses within the individual. If the two stimuli levels are close enough and the response variances are large enough then the two distributions will overlap as shown in Figure 5.10.

Figure 5.10  Distribution of Response to Two Different Stimuli

If now the individual is required to compare the two stimuli and make a selection of the higher order (e.g. the biggest, the best), then the probability of arranging
the two stimuli in the correct order will be a function of the distance between them (and hence the degree of overlap of the two distributions).

Hence by allowing for an error in the perception of the stimuli levels, the number of times an object is selected will now constitute an interval scale since the number of incorrect selections will be directly related to the distance between the mean values of the perceived stimulus distributions.

An example of a paired-comparison task is shown in Figure 5.11 for the estimation of system attribute importances. Paired-comparison tasks have been used by Golob et al. (1972) and Gustafson and Navin (1973).

In each of the following questions, please select the feature (A or B) which you would most prefer to be included in a public transport service for your journey to work.

1. A. Guarantee of obtaining a seat for the entire journey.
   B. Low waiting times at stations.

2. A. Low door-to-door travel time.
   B. Low fare.

Figure 5.11 Part of a Paired Comparison Question

Two major problems exist with the paired-comparisons technique. Firstly, the levels of adjacent stimuli must be close enough such that there exists some overlap between the two distributions in Figure 5.10. If there is no overlap, then no estimate of the distance between them can be inferred from the results of the comparisons and the resultant scale reverts to an ordinal scale. Secondly, the number of comparisons needed to compare all objects in all possible ways is given by \( n(n-1)/2 \). Thus when the number of objects is large, the number of comparisons becomes prohibitive. It is possible to compare only some pairings (Gullikson, 1956; Bock and Jones, 1968) but this makes the analysis more complicated and less accurate.

Rank Ordering

In a rank ordering task, an individual is asked to rank a set of alternatives with respect to some attribute. For example, modes of transport may be ranked in order of the level of comfort associated with each. For one individual, such a ranking produces an ordinal scale of comfort measurement. However, by again
employing Thurstone’s Law of Comparative Judgement, and by having the individual perform repeated rankings, an interval scale may again be generated. However, because all alternatives are presented for comparison at the one time in the rank order task, whereas they were presented only two at a time in the paired comparisons task, it is likely that there will be less inconsistency in judgements in the rank order task for any one individual (i.e. the variance in the distributions will be reduced). For this reason, it is more difficult to produce an interval scale with individual rank order data. It is therefore more usual to obtain interval scales from rank order data by means of utilising the results obtained from a number of different individuals with the assumption that the individuals come from a homogeneous population.

The obvious advantage of the rank ordering method is that the task is not as onerous as the paired-comparisons task and that the task can be more easily expanded to take account of new alternatives for comparison. However, like paired-comparisons, the alternatives must be sufficiently close with respect to the attribute in question such that the distributions overlap to some extent. The rank order method can be used to obtain both profile and similarities data. An example of the rank-order task for similarities data is given by Nicolaidis (1977) and is shown in Figure 5.12. The data obtained from this method was used by Nicolaidis in a multidimensional scaling analysis technique known as INDSCAL (Carroll and Chang, 1970).

Category Scales

The allocation of objects to categories can produce scales of a nominal, ordinal or interval nature depending on the way in which the categories are defined. At the nominal scale level, objects may be allocated to categories on the basis of unorderable classifications (e.g. sex). At the ordinal scale level, such categories may indeed possess an order such as income categories or statements of preference for an object. Interval scales may be derived by ensuring that the descriptions of each category accord with the numerical value associated with each category. To construct an interval category scale it is necessary to know the relationships between standard words and phrases in terms of the numerical interpretation of such labels. A semantic atlas showing dimensional intensity loadings for several hundred English words has been constructed for this purpose (Jenkins et al., 1958).
For each of the modes of travel shown below, please specify which of the remaining modes you think are most similar with respect to the comfort experienced whilst using each mode. Indicate the most similar mode by putting its identification letter in the first set of brackets under the name of each mode of transport. Put the identification letter of the second most similar in the second set of brackets and so on for all the other modes.

The identification letters for the modes are:

A. Automobile
B. Bus
C. Bicycle
D. Motorcycle
E. Taxi
F. Walking
G. Hitch-hiking

<table>
<thead>
<tr>
<th></th>
<th>Automobile</th>
<th>Walking</th>
<th>Taxi</th>
<th>Bicycle</th>
<th>Bus</th>
<th>Hitch-hiking</th>
<th>Motorcycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 [ ]</td>
<td>1 [ ]</td>
<td>1 [ ]</td>
<td>1 [ ]</td>
<td>1 [ ]</td>
<td>1 [ ]</td>
<td>1 [ ]</td>
<td>1 [ ]</td>
</tr>
<tr>
<td>2 [ ]</td>
<td>2 [ ]</td>
<td>2 [ ]</td>
<td>2 [ ]</td>
<td>2 [ ]</td>
<td>2 [ ]</td>
<td>2 [ ]</td>
<td>2 [ ]</td>
</tr>
<tr>
<td>3 [ ]</td>
<td>3 [ ]</td>
<td>3 [ ]</td>
<td>3 [ ]</td>
<td>3 [ ]</td>
<td>3 [ ]</td>
<td>3 [ ]</td>
<td>3 [ ]</td>
</tr>
<tr>
<td>4 [ ]</td>
<td>4 [ ]</td>
<td>4 [ ]</td>
<td>4 [ ]</td>
<td>4 [ ]</td>
<td>4 [ ]</td>
<td>4 [ ]</td>
<td>4 [ ]</td>
</tr>
<tr>
<td>5 [ ]</td>
<td>5 [ ]</td>
<td>5 [ ]</td>
<td>5 [ ]</td>
<td>5 [ ]</td>
<td>5 [ ]</td>
<td>5 [ ]</td>
<td>5 [ ]</td>
</tr>
<tr>
<td>6 [ ]</td>
<td>6 [ ]</td>
<td>6 [ ]</td>
<td>6 [ ]</td>
<td>6 [ ]</td>
<td>6 [ ]</td>
<td>6 [ ]</td>
<td>6 [ ]</td>
</tr>
</tbody>
</table>

Figure 5.12 Similarities Ranking Question

An example of such a category scale is shown in Figure 5.13. All the respondent has to do is to mark the importance of each attribute in the appropriate box. An analysis by Miller (1956) suggests that the most appropriate number of categories for such a scale, based on the limits of human discrimination, is seven plus or minus two. There should always be an odd number of categories to allow for a neutral rating by the respondent, if so desired.

There exists some debate as to whether scales of this type are truly interval or merely ordinal. Some researchers (e.g. Anderson, 1972) claim that the scale values are interval and use the numerical labels of the categories at face value. Others (e.g. Bock and Jones, 1968) state that they are really only ordinal but, by employing the Law of Comparative Judgement, they then proceed to convert the scale values into an interval scale. The category scale method has been applied to transport by Paine, et al. (1969) and Brown (1977).
Chapter 5

In considering the use of public transport for the journey to work, show the importance of the following features by circling one of the numbers 1 to 7 in accordance with the degree of importance you associate with each feature.

<table>
<thead>
<tr>
<th>FEATURE</th>
<th>IMPORTANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guarantee of getting a seat for the entire trip</td>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>Low waiting times at stations</td>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>Low door-to-door travel time</td>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>Low fare</td>
<td>1 2 3 4 5 6 7</td>
</tr>
</tbody>
</table>

Figure 5.13  Category Scale Question

Likert Scales

The Likert scale consists of a number of attitudinal statements of different polarities and degrees of extremity. The respondent rates each statement along a five-point dimension denoted by: strongly disagree, disagree, uncertain, agree and strongly agree. This type of scale is shown in Figure 5.14. The final score with respect to the attitude in question is given by a combination of the ratings on each statement and the extremity of view expressed by each statement.

The Likert scale is similar in concept to two other techniques which seek responses to statements of varying polarity and extremity. These are the Thurstone scale and the Guttman scale (Guttman, 1950; Fishbein, 1967). Although the question format and analysis techniques are somewhat different, the three scales have been shown to give very similar results.
An important aspect of the use of Likert, and other, scales is the need to pre-test extensively with a panel of respondents to establish the extremity of the view expressed by each statement (Oppenheim, 1992). This is done by writing a set of attitude statements, and then ascertaining the opinions expressed towards these statements by a panel of representative respondents. Once the polarity and extremity of each statement is established, a balanced pool of questions is then selected for use in the final survey.

Semantic Differential Scales

One of the most widely used of the attitude scales is the semantic differential scale developed by Osgood et al., (1957). The scale consists of a seven-point scale which is labelled at each end with bipolar adjectives describing the quality in question e.g. good/bad, satisfactory/unsatisfactory, expensive/cheap. The respondent rates the attribute in question by placing an X on the scale at a position which is indicative of the strength of the response. A sample semantic differential scale is shown in Figure 5.15. The semantic differential scale can be seen to be closely related to a category scale. The principal difference lies in the use of end-anchors...
only with the semantic differential whereas intermediate labels are used for category scales.

In considering the use of public transport for the journey to work, show the importance of the following features by marking with an X on the scale-line to correspond with the degree of importance you associate with each feature.

**Figure 5.15  Semantic Differential Scale Question**

Examples of transport studies which have used semantic differential scales include Golob (1970), Nicolaidis (1977), Ackoff (1965), Sherret (1971), Hartgen and Tanner (1971) and Golob, Dobson and Sheth (1973). Such scales can be used equally well to obtain profile or similarities data.

**Ratio Scales**

Although there are many methods for the generation of ratio scales (see, for example, Guilford, 1954; Carterette and Friedman, 1974; Torgerson, 1958), three methods are of particular importance in transport choice analysis.

(i) Fractionation

This scaling procedure involves presenting the respondent with an assumed rating for one object and then requiring the respondent to select another object whose rating is a particular fraction of the initial object’s rating. Thus, for example, a respondent may be told the accessibility of one location to the city centre and then asked to nominate other locations whose accessibility is a fraction
(e.g. a half) of the original location. The scale of accessibility so derived would be a true ratio scale of perceived accessibility.

(ii) Multiple judgement

Closely related to judgements of fractions are judgements of multiples. The respondent is given an initial rating for an object and then asked to select another object whose rating is a given multiple of the initial rating.

Although the methods of fractionation and multiple judgements have been considered by some (Torgerson, 1958; Hanes, 1949) to be similar, it is not altogether clear that they represent the same psychological process. For example, fractionation methods may be considered as an interpolation between the nominated initial rating and a fixed zero point, thus generating a true ratio scale. On the other hand, multiple judgement methods may be considered as an interpolation between the nominated rating and the highest rating of that type of object that the respondent has ever experienced. In this case the zero point is arbitrary, and hence the ratings do not necessarily generate a ratio scale.

(iii) Magnitude Estimation

In the magnitude estimation method, the respondent is asked to assign scale values to a series of objects in accordance with the subjective impressions they elicit. No arbitrary reference point is specified at the start; the respondent chooses both the reference point and the subsequent ratios to this reference point. The origins of this method are attributable to Richardson (1929) but the modern development of this theory has been most strongly associated with Stevens (1956, 1967).

As an example of this method, a respondent may be asked to assign a number (any number) to represent the accessibility of a location to the city centre. The presentation of other locations to the respondent should then elicit responses which are in direct proportion to the accessibility ratio between the first and the present locations.

Considerable debate exists as to whether category scales or ratio scales are the most appropriate measurement techniques. Stevens (1974), for example, states that "For the purposes of serious perceptual measurement, category methods should be shunned. The deliberate and ill-conceived imposition of a limited set of response categories forces the subject into portioning. At that point, the hope for a ratio scale must fail". On the other hand, Anderson (1976) is equally as forceful when he states that "It seems appropriate, therefore, to conclude that the rating method can yield true interval scales and that the method of magnitude estimation is biased and invalid". To be sure, there exists a non-linear relationship between category rating results and ratio scale results as demonstrated by
Stevens and Galanter (1957), and acknowledged by Anderson (1976). The form of such a relationship is shown in Figure 5.16, and demonstrates the concave relationship usually found between category and ratio scale results.

![Magnitude Estimation Rating of Loudness (sones)](image)

**Figure 5.16 Comparing Category Scale and Magnitude Estimation Ratings**  
(Source: Stevens and Galanter, 1957)

Although there is obviously a difference between the results there is no clear indication of which is correct. The problem is confused even further when it is realised that the number system which is used is also subject to considerable doubt as to which type of scale it represents. Is it an interval or ratio scale, and is it perceived as a linear scale? (Jones, 1974). It has been empirically demonstrated, for example, that the number series (from 1 to 10, at least) is subjectively perceived as a power function with an exponent of 0.49 (Rule, 1971). That these basic questions about scaling remain unresolved after a century of psychological research should be some comfort for transport planners attempting to come to grips with the area of attitudinal measurement.

**Constant Sum Allocation**

A final unidimensional ratio scaling technique which is of relevance to transport choice is the method of constant sum allocation (Comrey, 1950). The original technique involves the division of 100 points between pairs of objects in such a way that the assigned values indicate the relative amounts of some characteristic which each object possesses. The method used by Comrey is based on a paired comparison format. However, considerable economies in effort in data collection
can be effected by considering all objects at one time and dividing the 100 points between all the objects.

This constant sum allocation method has been used in transport choice analysis to measure attribute importances (Hensher, 1972) and behavioural intent (Hartgen and Keck, 1974). It appears that this relatively straightforward method could be used more often in the measurement of attitudes in terms of a ratio scale.

**Attitudinal Questions in Pilot Surveys**

One of the most important aspects of the use of attitude or opinion questions - and one which is almost invariably overlooked in travel-related research - is the pilot test. Its importance is documented again and again in the social science literature (e.g. Oppenheim, 1992). After studying the literature on the subject, these pilot tests need to take the form of in-depth interviews, the essential purpose of which is two-fold:

1) to explore the origins, complexities and effects of the attitude areas in question in order to decide more precisely what is to be measured, and;

2) to get vivid expression of these attitudes from the respondents in a form that could be used in the statements on an attitude scale.

Suppose we were trying to construct a scale to measure people's attitudes to improving the frequency of buses in their area. We may well find in exploratory interviews that almost everyone is in favour of this measure. A simple "for or against" scale on improving frequency would, therefore, show little differentiation. It can often happen that the pilot test actually causes a change in the aim of the scale (e.g. levels of frequency rather than to increase or not to increase), and possibly of the aim of the investigation.

Next, we may propose to build a scale dealing with "relevance of bus frequency to mode choice" - dealing with the extent to which considerations of bus frequency enter into people's minds when they choose a mode of transport. At first, it would seem that this attitude is directly related to people's knowledge about bus frequencies; and it would seem relatively easy to differentiate between those with more or less knowledge about the bus service on the basis of a few well-chosen factual knowledge questions. Further exploration may show, however, that many people with little correct knowledge about bus frequencies nevertheless are very interested in them and are influenced by friends'/neighbours'/newspapers' claims about them. We begin to find various links between a person's attitudes to bus frequencies and their attitudes to other aspects of their life; for example, community/neighbourhood awareness may influence the perception of bus frequencies, while concern about house prices...
may also influence whether people want increased bus frequencies. And we could continue to find a variety of other linkages outside the narrow realm of transport services! "This is a part of social psychology where clinical acumen, intuition and a capacity for listening with the third ear are invaluable and where chairbound, preconceived frameworks may constitute a real hindrance." (Oppenheim, 1992).

After doing perhaps thirty or forty preliminary interviews, it is then necessary to decide what it is we wish to measure. Only then will we be in a position to draw up a conceptual sketch of the clusters of attitudes in question with the likely linkages.

The Asking of Attitudinal Questions

It is essential when asking opinion questions in personal interview surveys that interviewer effects be reduced to a minimum. Thus, in addition to asking the questions verbatim, it is important to realise that even seemingly innocuous comments may have considerable impact on the way in which opinion questions are answered. All answers must be accepted by the interviewer as perfectly natural and no signs of positive or negative reaction should be displayed. These reactions may either be verbal (such as saying "I see" with an inflection after a response is given) or visual (such as raising the eyebrows).

Apart from possible interviewer effects, attempts to obtain opinions may meet with other problems. First, respondents may simply have no opinion on a particular topic - they may not have thought about it before or else, if they have, then the topic may be so unimportant to them that they have not bothered to form an opinion. Second, respondents may have conflicting opinions on a topic. They may see both good and bad sides to the topic and may not be able to give a definite opinion without a large number of qualifying statements. Third, although an opinion may be elicited from a respondent, the intensity of opinion may vary considerably between respondents. Thus some respondents may feel very strongly about a topic (either positively or negatively) while others may feel less strongly but in the same direction. It is particularly important that there is room for respondents to record all these types of responses either on the questionnaire form or to the interviewer.

Because opinion questions generally require a spontaneous and individual response from a respondent, opinion questions are generally more suited to personal interview surveys than to self-completion questionnaires, despite the above reservations about interviewer effects.

The preceding discussion of attitudinal measurement techniques has been necessarily brief. A vast array of literature awaits those who wish to delve more deeply into the subject. In terms of general reading on attitudinal measurement
Survey Instrument Design


5.5.4 Stated Response Questions

Two types of multidimensional scaling technique are of particular relevance to transport choice analysis. The first involves the rating of an alternative, overall, by means of one of the techniques mentioned above - that is, the application of a unidimensional scaling technique to a multidimensional object. This method is frequently used to ascertain how the unidimensional ratings of the individual attributes might be combined into an overall rating of the alternative.

The second method is known by various titles such as conjoint measurement (Luce and Tukey, 1964; Krantz and Tversky, 1971), information integration (Anderson, 1971, 1974), functional measurement (Anderson, 1970; Levin, 1979a; Meyer, Levin and Louviere, 1978), and, in recent years, stated preference or stated response (Pearmain et al., 1991; Hensher, 1994). The principle feature of each of these methods is that they seek the respondent’s reaction to a series of hypothetical combinations of attribute levels. The set of questions is determined on the basis of an experimental design which seeks to present a balanced set of situations to the respondent.

Stated response methods are particularly useful in two contexts:

- when a substantially new alternative is being introduced and there is little or no historical evidence of how people might respond to this new alternative
- when the investigator is trying to determine the separate effects of two variables on consumers’ choices, but where these two variables are highly correlated in practice.

Because of the manner in which the set of questions has been determined by an experimental design, the investigator has control over the combinations of attributes to which the respondent will respond. This is particularly important in the second context listed above, because it enables the investigator to isolate the individual effects of the various attributes.

The design of the choice situations to be presented to the respondents is an important component of the overall design of stated response surveys. Pearmain et al. (1991) offer a simple example of such design by considering a situation
Chapter 5

involving three attributes for a public transport service: fare, travel time and frequency. If each attribute has only two levels (viz, high-low, slow-fast, frequent-infrequent), then there are eight different combinations of these options as shown in Table 5.2.

Table 5.2  A Simple Stated Response Experimental Design

<table>
<thead>
<tr>
<th>Option</th>
<th>Fare</th>
<th>Travel Time</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Low</td>
<td>Fast</td>
<td>Infrequent</td>
</tr>
<tr>
<td>2</td>
<td>Low</td>
<td>Fast</td>
<td>Frequent</td>
</tr>
<tr>
<td>3</td>
<td>Low</td>
<td>Slow</td>
<td>Infrequent</td>
</tr>
<tr>
<td>4</td>
<td>Low</td>
<td>Slow</td>
<td>Frequent</td>
</tr>
<tr>
<td>5</td>
<td>High</td>
<td>Fast</td>
<td>Infrequent</td>
</tr>
<tr>
<td>6</td>
<td>High</td>
<td>Fast</td>
<td>Frequent</td>
</tr>
<tr>
<td>7</td>
<td>High</td>
<td>Slow</td>
<td>Infrequent</td>
</tr>
<tr>
<td>8</td>
<td>High</td>
<td>Slow</td>
<td>Frequent</td>
</tr>
</tbody>
</table>

The respondent could then be asked to rank these options in order of preference, and from the combined responses of a sample of respondents, the relative importance attached to fares, travel times and frequency could be determined. Importantly, because of the orthogonal nature of the experimental design (ie. each variable is independent of all other variables in the set of options presented to the respondent), the importances attached to each attribute are true reflections of the separate effects of each attribute.

As with unidimensional scales, the respondent may be asked to perform different tasks with the presented information. For example, they could be asked to:

- rank the alternatives in order of preference
- assign a rating to each alternative to reflect their degree of preference
- select the single alternative which they prefer the most
- select choices in a paired comparison manner from a series of two-way choice situations

Each of these methods has their own strengths and weaknesses, both from the point of view of the respondent and the analyst.

One of the problems with stated response methods is that the set of options shown in Table 5.2 is extremely limited. For example, it is likely that more than two levels of each of the attributes would need to be tested, and perhaps more
Survey Instrument Design

than three attributes will need to be evaluated. However, as the number of attributes and attribute levels increases, so too does the number of possible combinations of the attribute levels. For example, if we wish to test three levels of three attributes, that would result in 27 combinations; three levels of four attributes would require 81 combinations. Clearly, it is impossible to expect respondents to be able to consider this many different situations. Kroes and Sheldon (1988) suggest that a maximum of 9 to 16 options is acceptable, with most current designs now adopting the lower end of this range. With a maximum of 9 options for the respondent to consider, this severely limits the number of attributes that can be considered. For example, the following options are available with 9 or less options:

- with two attribute levels  \(2^2 = 4\) options  2 attributes with 2 levels  \(2^3 = 9\) options  3 attributes with 2 levels
- with three attribute levels  \(3^2 = 9\) options  2 attributes with 3 levels
- with mixed levels  \(2^1 \times 3^1 = 6\) options  1 attribute with 2 levels  1 attribute with 3 levels  \(2^1 \times 4^1 = 8\) options  1 attribute with 2 levels  1 attribute with 4 levels

To overcome this limitation, and yet be able to consider more attributes and/or more attribute levels, it is necessary to adopt one of the following strategies (Pearmain et al., 1991):

- use a "fractional factorial" design, whereby combinations of attributes which do not have significant interactions are omitted from the design. A significant interaction is said to exist when the combined effect of two attributes if significantly different from the combination of the independent individual effects of these two attributes.

- remove those options that will "dominate" of "be dominated" by all other options in the choice set. For example, in Table 5.x1, option 7 is dominated by all other options, while option 2 dominates all others. These options could be removed from the choice set, on the assumption that all "rational" respondents would always put option 2 first and option 7 last in any ranking, rating or comparison process.

- separate the options into "blocks", so that the full choice set is completed by groups of respondents, but with each group responding to a different sub-set of options. Each group responds to
Chapter 5

a full-factorial design within each sub-set of options, and it is
assumed that the responses from the different sub-groups will be
sufficiently homogeneous that they can be combined to provide the
full picture.

- present a series of questions to each respondent, offering different
  sets of attributes, but with at least one attribute common to all to
  enable comparisons to be made. Often the common attribute will be
time or cost to enable all other attributes to be measured against
easily understood dimensions.

- define attributes in terms of differences between alternatives (e.g.
  travel time difference between car and train). In this way, two
attributes are reduced to one attribute in the experimental design.
However, they may still be presented as separate attributes to the
respondent on the questionnaire.

Adoption of one, or more, of the above strategies will allow more information to
be obtained from stated response questionnaires while keeping the task
relatively manageable for the respondent.

The major weakness of stated response methods, however, is that they seek the
reactions of respondents to hypothetical situations and there is no guarantee that
respondents would actually behave in this way in practice. This is particularly the
case if the respondent does not fully understand the nature of the alternatives
being presented to them. There is thus a high premium on high-quality
questionnaire design and testing to ensure that respondents fully understand the
questions being put to them. Unfortunately, this does not appear to be the case at
the present time. While a lot of attention has been placed on refining the nature
of the experimental designs, and on increasing the sophistication of the analysis
techniques to be employed after the data has been collected, relatively little
attention has been paid to improving the quality of the questions being put to the
respondents. With few exceptions (e.g. Bradley and Daly, 1994), relatively little
attention has been focussed on testing for methodological deficiencies in the
survey techniques used to obtain stated response data. There are numerous
examples of stated response questionnaires in which the questions being asked of
the respondent are almost unintelligible (even to a trained professional). Future
work in this area must pay much greater attention to the quality of the survey
instrument.

5.6 QUESTION FORMAT

The format of a question describes the way in which the question is asked and,
more importantly, the way in which the answer is recorded. The choice of
Survey Instrument Design

question format is closely related to the choice of data processing procedures to be used later in the survey process. Three basic types of question format are available:

(a) open questions
(b) field-coded questions
(c) closed questions

5.6.1 Open Questions

Open questions are answered by the respondent in their own words which are then recorded verbatim (as much as possible) and coded at a later date (Figure 5.17). Open questions can be used in personal interview surveys, where the interviewer does the recording, or in self-completion questionnaire surveys where the respondent does the recording by means of a written answer.

<p>| WHAT IS YOUR COMPANY POLICY FOR BUSINESS TRAVEL? |</p>
<table>
<thead>
<tr>
<th>(i.e. specific modes, costs - upper limits, etc.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>---</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>---</td>
</tr>
</tbody>
</table>

Figure 5.17 An Open Question in a Personal Interview
Source: Ampt Applied Research and Transport Studies Unit (1989)

The open question format has a number of distinct advantages:

(a) This format generally improves the relationship with respondents because they feel that their own personal views are of some importance. In this respect, open questions are often good introductory questions in interviews to make the respondent feel at ease, and to stimulate interest in the survey topic. They are also useful concluding questions at the end of any survey to enable the respondent to "let off steam" about the survey topic.

(b) There is more opportunity for probing by the interviewer to bring out further points in discussion. Specific probes may be provided to each interviewer to guide discussion.

(c) Open questions can be of great help to the investigator in developing ideas and hypotheses which may later be tested by more structured questions.

(d) Open questions are most useful in pilot surveys where the range of responses to a question may be initially unknown (see Section 5.5.3 for a
discussion on this aspect of attitude questions). The pilot survey open question responses can be used to define an expected range of responses to make it possible to develop a satisfactory closed question format for the main survey.

(e) The use of an open question, immediately after a closed question, can be useful for interpreting the respondents’ perception or understanding of the closed question. This technique is particularly useful in pilot studies where wording is being tested.

(f) The quotations obtained from open question responses may be of direct value when writing the survey report, to provide a degree of human interest and to emphasise dramatically the meaning and significance of otherwise sterile statistical results.

Open questions are not, however, without their disadvantages. In particular, it has been found that open questions can have the following problems:

(a) By forcing respondents to express their feelings in their own words, they can be threatening.

(b) They can give answers which are vague and not related to the subject matter of the survey.

(c) They can also give quite valid answers, the range of which is so wide as to be completely unmanageable from a coding viewpoint.

(d) When used in personal interview surveys, they are generally subject to interviewer effects in two main ways. First, the amount of information obtained from an open question will depend, to a large extent, on the amount of probing done by the interviewer. Differences in response may therefore be a function of the interviewer rather than of the respondent. Second, although the answers to open questions are intended to be recorded *verbatim*, this is generally impossible in an interview situation. Interviewers are therefore somewhat selective in recording responses, recording those aspects which *they* feel are most important. In this way, interviewer bias will affect the results.

(e) The response to open questions can also be subject to respondent bias. Thus, in personal interviews, a larger number of comments will be received from the more loquacious members of the population irrespective of whether they have more information to contribute or not. Similarly, in self-completion questionnaires, a greater number of comments will be received from the more literate members of the population. Since both these characteristics are related to the social class
and level of education of the respondent this could introduce a bias into the interpretation of the results of open questions.

(f) For personal interview surveys, open questions generally mean that a double-processing of responses is necessary - once in the field when the response is recorded verbatim and once in the office where the responses are assigned a numerical code.

(g) If too many open questions are used in an interview, it can be very tiring, and/or annoying, for both interviewer and respondent. It can be particularly annoying if the same subject matter is covered later in the interview in closed question format. The respondent may rightly wonder why they are being asked a question to which they have already given a complete answer.

The general recommendation for open questions in personal interviews and in self-completion questionnaires is therefore to use them sparingly. Be aware that although they will often provide a richer source of information, they also carry with them some significant practical disadvantages.

5.6.2 Field-Coded Questions

In an attempt to avoid some of the disadvantages of open ended question in personal interview surveys, use is often made of field-coded questions, in which the interviewer asks what is apparently an open question but then, instead of recording the answer verbatim, classifies the response into one of several predetermined categories (Figure 5.18). The categories available are, however, known only to the interviewer.
The principal advantage of field-coded questions is that they eliminate double-processing of the responses by having the interviewer do the coding. However, this is also the source of their prime disadvantage since considerable undetectable interviewer bias may exist in the interpretation of the responses and the selection of the code to match the response. As in all situations, the training of the interviewer is crucial in these cases. The Interviewers' Manual (Section 5.9) becomes essential here since it should contain all definitions needed (e.g. a hairdresser might be coded as "Personal Business/Services" and "Education" may exclude hobby courses). It is essential that the interviewer not only knows what to do, but also understands the reasons for asking each question, so that their coding is done with comprehension and not simply according to pre-specified rules.

5.6.3 Closed Questions

In an effort to eliminate interviewer bias altogether, most questionnaires make greatest use of closed questions, which may be defined as those questions where
the respondent is presented with a list of possible responses and is then asked to fit themselves into the appropriate category in response to the question (Figure 5.19).

![Figure 5.19 A Closed Question in a Personal Interview](image)

Source: Ampt and Waters (1993)

A number of features of closed questions deserve specific mention:

(a) Closed questions are most useful for factual questions, although opinion questions can be asked by means of psychometric scaling techniques (see Tischer, 1981).

(b) The use of pre-selected categories is very valuable in helping to define the meaning and scope of the question for the respondent.

(c) On the other hand, the scope of the responses must not be unnecessarily restricted by the range of the response categories offered. The response categories should be:

(i) exhaustive (Note that by presenting some, but not all, categories you run the risk of suggesting to the respondent that their responses should fit into those options offered);

(ii) mutually exclusive

(iii) unambiguous.

(d) To avoid forcing people into categories which are not really appropriate (which may bias the results and will certainly annoy the respondent), an open alternative should be offered, where appropriate (e.g. "don't know", "not applicable", "other (please write in)") as shown in Figure 5.20.
Figure 5.20  An Open Alternative in a Closed Question
Source: Richardson and Ampt (1993a)

(e) If one does offer an open alternative category, as described above, it must be realised that this may attract a higher than acceptable response, since it offers respondents a chance to reply without really thinking about the question. In some circumstances it may be more desirable to force respondents to come down on one side of the fence or the other.

(f) In the search for exhaustiveness of categories, one must be careful to avoid creating confusion by offering too many categories. In opinion questions, a limit of seven response categories is often taken as a general rule-of-thumb (Miller, 1956).

(g) When there are numerous possible responses in a personal interview, it is preferable to show respondents a card on which the categories are printed rather than have the interviewer read out the possible responses (see Figure 5.21). Apart from the possibility of having the interviewer show preference for one response by the mere tone of their voice, it has been demonstrated that respondents show preferences for alternatives at the start and end of the list when they are read out (because of short-term memory effects). The use of show cards also allows the interviewer to obtain a variation in the ordering of the alternatives by a random selection of the card to be shown.
The use of cards with variable ordering can eliminate the possibility of a "donkey vote" biasing the results.

(h) If respondents are required to answer "Yes/No" questions or provide "Agree/Disagree" responses, it should be realised that it has been shown that certain individuals are prone to respond in a positive way regardless of the question. Low interest in the survey topic results in a larger tendency to respond positively.

(i) As a general rule, it is necessary to pre-test closed questions very carefully to insure that the proposed categories are understood, relevant, and more-or-less comprehensive.

---

**Survey Instrument Design**

<table>
<thead>
<tr>
<th>Per Week</th>
<th>Per Year</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>No income</td>
<td>No income</td>
<td>1</td>
</tr>
<tr>
<td>$1 - $58</td>
<td>$1 - $3,000</td>
<td>2</td>
</tr>
<tr>
<td>$59 - $96</td>
<td>$3,001 - $5,000</td>
<td>3</td>
</tr>
<tr>
<td>$97 - $154</td>
<td>$5,001 - $8,000</td>
<td>4</td>
</tr>
<tr>
<td>$155 - $230</td>
<td>$8,001 - $12,000</td>
<td>5</td>
</tr>
<tr>
<td>$231 - $308</td>
<td>$12,001 - $16,000</td>
<td>6</td>
</tr>
<tr>
<td>$309 - $385</td>
<td>$16,001 - $20,000</td>
<td>7</td>
</tr>
<tr>
<td>$386 - $481</td>
<td>$20,001 - $25,000</td>
<td>8</td>
</tr>
<tr>
<td>$482 - $577</td>
<td>$25,001 - $30,000</td>
<td>9</td>
</tr>
<tr>
<td>$578 - $673</td>
<td>$30,001 - $35,000</td>
<td>A</td>
</tr>
<tr>
<td>$674 - $769</td>
<td>$35,001 - $40,000</td>
<td>B</td>
</tr>
<tr>
<td>$770 - $961</td>
<td>$40,001 - $50,000</td>
<td>C</td>
</tr>
<tr>
<td>$962 - $1,155</td>
<td>$50,001 - $60,000</td>
<td>D</td>
</tr>
<tr>
<td>$1,156 - $1,346</td>
<td>$60,001 - $70,000</td>
<td>E</td>
</tr>
<tr>
<td>$1,347 - $1,538</td>
<td>$70,001 - $80,000</td>
<td>F</td>
</tr>
<tr>
<td>Over $1,538</td>
<td>Over $80,000</td>
<td>G</td>
</tr>
</tbody>
</table>

**Figure 5.21 A Show Card for a Closed Question**  
Source: Ampt and Waters (1993)
Chapter 5

In personal interview surveys, a combination of open, closed and field-coded questions may be most effective in exploring a topic. For example, Gallup (1947) suggested a "quintamensional plan of question design" in which a series of five types of question are used to explore various components of an issue, as shown in Figure 5.22.

| Q.1 | Awareness of issue (Open or field-coded question) |
| Q.2 | General feelings (Open) |
| Q.3 | Answers on specific parts of issue (Open, field-coded or closed) |
| Q.4 | Reasons for views (Open) |
| Q.5 | Intensity of belief (Open or closed) |

Figure 5.22 The Quintamensional Plan of Question Design
Source: (Gallup, 1947)

In general, the balance between open and closed questions will depend on a number of situational factors. These include the objectives of the survey, the prior knowledge held by the investigator, the extent to which the respondent has thought about the subject matter, and the ease with which the material in question can be communicated to the respondent.

5.7 QUESTION WORDING

It is obvious that a question must be understood by the respondent before a suitable answer can be given. The meaning of a question is communicated primarily by the words used in the question. While there are numerous studies on the problems of specific wording (e.g. Oppenheim, 1992) the following points are the essential features of question wording which should be borne in mind in the design of a questionnaire.

5.7.1 Use Simple Vocabulary

Use simple and direct vocabulary with no jargon or slang. The language used should be suited to the subject and the population in question (note that in some surveys a language other than English may be most appropriate). Many words and phrases which you think are simple may not be to the general population or, more importantly, to your particular survey population. For example, transport planners frequently refer to "modes" of travel to denote different forms of transportation such as cars, buses, trains, walking, etc. The general public, however, does not usually refer to these means of transportation as travel.
"modes". It will be much more easily understood if you say "What means of transport did you use to get there?" or, even better, "How did you get there?" "Destination" is another word with similar problems. Try variations on "Where did you go?"

Make sure that words which you use have the same meaning to your respondents. In some cases, it may be necessary to define the meaning of certain words (e.g. trip) which will be used in the questionnaire. If you are only going to use a word or phrase once only, it is best to give a full description immediately. For example, if you are only going to use the word trip once, rather than to say "a trip - which means...", it is easier to say something like "every time you move along a public street...". Pilot testing with people from the population to be surveyed will assist with checking these words.

5.7.2 Use Words Appropriate to the Audience

While the language used should be simple, you should not give the appearance of "talking down" to your respondents. If this situation is perceived by respondents, you run the risk of alienating them for the entire survey. It should be realised, however, that in surveying a large general population, it is virtually impossible to abide by both these first two points on question wording if a single questionnaire is used for the entire population. To attempt to satisfy both points, it would be necessary either to use different question wording for different segments of the population (and assume that both wordings are measuring the same phenomena) or else provide for an interviewer who can adapt and interpret the question wording according to each respondent's circumstances (and assume that interviewer effects are minimal). The best solution is to use the day-to-day language you use when speaking to your friends (who are not colleagues) about travel and transport.

A particular problem is encountered when dealing with a multilingual population. Not all members of this population are equally fluent in the majority language, and for this reason it would be desirable to have different versions of the survey in the languages which are most likely to be encountered. However, this poses a potential problem in ensuring that each version of the survey is in fact measuring the same thing because of subtle differences which may creep into open-ended questions during the translation process. A recent survey in Singapore (Olzewski, et al., 1994) used the personal interview method described in Section 7.3 which meant that short diaries were left for respondents to complete before the personal interview. By having these diaries ("Travel Memos") available in Mandarin, Malay and Tamil as well as English, it was possible to conduct many interviews in English since the difficult task (writing) had been done in the native language.
5.7.3 Length of Questions

Questions which are long, complicated, or which involve multi-dimensional concepts are better split up into a number of simpler, shorter questions. On the other hand, with open questions it has been found that the length of the response given by the respondent is directly proportional to the length of the question asked by the interviewer. Short abrupt questions generally elicit short abrupt replies. The choice of question length will therefore depend on the objectives of an open question. For closed questions, where the length of the response is pre-specified, short simple questions are recommended.

5.7.4 Clarify the Context of Questions

The context and the scope of questions should be made clear by giving the respondent an adequate frame of reference in which to answer the question. For example, in a study of residential location, the question "Do you like living here?" would be relatively useless (unless asked purely as an introductory not-to-be-coded question). Unless the respondents are advised on which factors to consider in their answer (e.g. closeness to work, quality of neighbourhood) the range of possible answers would make the question extremely difficult to evaluate.

5.7.5 Avoid Ambiguous Questions

Obviously, every attempt should be made to avoid ambiguous questions. The trouble with this guideline is that often the survey designer does not realise that a question could be legitimately interpreted by a respondent in a completely different fashion to that which was intended. If ambiguities were obvious, they would be easy to eliminate. Generally, though, ambiguities are only detected by someone other than the question designer. This emphasises another important role for pre-tests and pilot surveys. To properly test for ambiguity it is particularly important that these tests are done with people of the type who will be surveyed. Asking colleagues is the least likely method of removing any ambiguities. Never carry out even the most simple survey without a pilot test for wording.

One source of ambiguity, which is related more to layout than wording of a question, concerns the placement of labels for tick-the-box closed questions. The labels should clearly refer to one and only one box. For this reason labels should not be placed between adjacent boxes on one horizontal line unless there is ample separation between label/box pairs. This confusion applies especially to categories in the middle of the list of categories, as shown in Figure 5.23.
Survey Instrument Design

How many vehicles were garaged at this household last night?

0 □ 1 □ 2 □ 3 □ 4 □ 5 □ >5 □

Figure 5.23 Poor Placement of Tick-Box Labels

5.7.6 Avoid Double-Barrelled Questions

"Double-barrelled" questions should be avoided. Answers to questions like "Do you like buses and trains?" are difficult to interpret by both the respondent and the analyst. More subtle, but equally incorrect, is the question, "How do you feel about public transport?" To some extent, the problem is ameliorated if respondents are asked to respond in an open way. If, however, they are given a scale, the answers to these questions can be considered basically worthless since it cannot be determined which form of public transport is being commented on.

5.7.7 Avoid Vague Words About Frequency

It is desirable to avoid vague words like "usual", "regular", "often", etc., unless there is a specific intention for the respondent to interpret the word in their own way as, for example, in the question "How do you usually travel to work?" The reason is because you will find that interpretations of this question are as many as the number of times the question is asked. "Regular", for example, can mean "each day", "each week" or each month" and would therefore give no indication of frequency., should that be the reason it was being asked.

Such a question could be used where it is impossible and/or unnecessary to obtain complete details of the journey to work or where it is expected that there is little variation in the method of travelling to work. Generally, however, less bias will be introduced into the response if the question is made more specific, e.g. "How did you travel to work today?"

5.7.8 Avoid Loaded Questions

In most situations, "loaded" questions should be avoided. The classic example of a loaded question is "Have you stopped beating your wife yet?" The essence of such loaded questions is that it presumes that the question is relevant to the respondent with respect to their current activities. To eliminate such loaded questions, use can be made of filter questions to determine whether subsequent questions are relevant to the respondent.
Chapter 5

In some situations, however, the loaded question has been used successfully to obtain answers to potentially embarrassing questions. A famous situation can be found in the work of Kinsey et al. (1948, 1953) in their study of sexual behaviour. Rather than asking respondents whether they engaged in certain sexual practices, they went straight into questions about frequency and detail. In this way, respondents were made to feel that such practices were perfectly normal behaviour and were able and willing to answer questions in great detail. In transport surveys, however, questions are seldom likely to be as socially delicate as these, and, therefore, the use of filter questions is generally advised.

5.7.9 The Case of Leading Questions

While loaded questions severely limit the range of possible replies, leading questions, which merely encourage the respondent to reply in a particular manner, are an equal, though perhaps more subtle, threat to the validity of responses. Leading questions can be phrased in several different ways, as described in the next six items, each of which should be recognised.

• The use of emotionally charged words can induce respondents to answer in a particular fashion, e.g. "Do you think that powerful left-wing trade unions should have more say in the control of public transport systems?" would probably bring a different response if the question were phrased as "Do you think employees should provide an input to the management of public transport systems?"

• The partial mention of possible answers will bias responses towards those answers, e.g. "Did you perform any activities on the way home from work last week (e.g. go shopping, play sport, etc.)?" will result in a higher recall of these activities and a lower reporting of other activities which have not been mentioned.

• The mention of a likely response framework can also lead to a leading question. For example, "Do you think congestion has increased in the last five years?" is more leading than "Do you think travel conditions changed in the last five years?"

• An appeal to the status quo is another example of a leading question. If a question implies or states that one of the alternatives to be chosen from a list is a representation of the present state of affairs, then there will be a tendency for respondents, in the general population, to choose this alternative because of the widespread tendency among the community to accept things as they are now in the social order. This biasing effect could also be used in reverse if the population in question is, for example, a group of radical university students.
Survey Instrument Design

- The use of qualifying phrases at the beginning or end of questions is an obvious way to construct a leading question, e.g. "Don't you think that..." or "...isn't it?" Such qualifiers are definitely taboo.

- The degree of leading involved will also depend on the context within which a question appears in a questionnaire. If respondents are asked to agree with one statement in a list of statements, they will more likely agree with the statement if the other statements are supportive of, rather than antagonistic towards, the statement in question.

- The context of a question can also be affected by the sequencing of previous questions. Such sequencing can result in obtaining predictable replies to questions by forcing respondents into a corner because of their previous answers. This technique is a favourite ploy of salesmen, as illustrated by the following abbreviated dialogue between an encyclopedia salesman and the parents of young children.

SALESMAN: *(after knocking on door of household)* Good evening! I am conducting a "survey" concerned with the state of education in today's schools. Do you have any children and are you interested in their education?

PARENTS: Why, yes, we do have two kids, and we are also concerned about what's happening in the schools today. Come inside.

S: *(After coming inside and getting settled)* What do you think are the major problems with the education today's children are receiving?

P: *(The parents are allowed to talk freely about the issue, giving them the feeling that they "own" this "interview". The salesman carefully notes any mention they make about reading, sources of information, reference material, etc.)*

S: I note that you mentioned several times that you thought reading was important and that kids spend too much time in front of the television these days.

P: *(The parents are given a further opportunity to commit themselves to the idea that reading is important and that kids should have an alternative to watching television all the time.)*

S: And who do you think should be responsible for providing the kids with this alternative to watching television at home?

P: *(The parents are able to say anything they wish, but almost invariably they will, at some stage, say that it is their responsibility as parents.)*

S: As parents, then, you feel that it is largely your responsibility to provide adequate educational material for your children to read
at home? What sources of educational reading do you think are most useful?
P: (Once again, the parents are able to say anything they wish, but again almost invariably they will, at some stage, say that encyclopedias are a valuable resource material.)
S: So, as responsible parents, then, you feel that encyclopedias are a valuable educational resource for your children to have available in their home?

At this stage, the salesman has committed the parents to such an extent by their previous answers that they can hardly turn around and say that they do not believe that they should consider buying the encyclopedias. By careful sequencing of questions, the salesman has turned their position from one of probable indifference, or even antagonism, towards the encyclopedias to one of support. While this example is somewhat extreme, the same process can be seen at work in more subtle ways in many examples of survey design.

The effect which question sequencing can have on the outcome of later questions is well demonstrated by the following extract from the popular television series "Yes, Prime Minister" (Lynn and Jay, 1989, pp. 106-107), in which Bernard is instructed by Sir Humphrey Appleby on the more subtle points of survey design:

"He was most interested in the party opinion poll, which I had seen as an insuperable obstacle to changing the Prime Minister's mind.

His solution was simple: have another opinion poll done, one that would show that the voters were against bringing back National Service.

I was somewhat naive in those days. I did not understand how the voters could be both for it and against it. Dear old Humphrey showed me how it's done.

The secret is that when the Man In The Street is approached by a nice attractive young lady with a clipboard he is asked a series of questions. Naturally the Man In The Street wants to make a good impression and doesn't want to make a fool of himself. So the market researcher asks questions designed to elicit consistent answers.

Humphrey demonstrated the system on me. 'Mr. Woolley, are you worried about the rise in crime among teenagers?'
'Yes,' I said.

'Do you think there is a lack of discipline and vigorous training in our Comprehensive Schools?'

'Yes.'

'Do you think young people welcome some structure and leadership in their lives?'

'Yes.'

'Do they respond to a challenge?'

'Yes.'

'Might you be in favour of reintroducing National Service?'

'Yes.'

Well, naturally I said yes. One could hardly have said anything else without looking inconsistent. Then what happens is that the Opinion Poll published only the last question and answer.

Of course, the reputable polls don't conduct themselves like that. But there weren't too many of those. Humphrey suggested that we commission a new survey, not for the Party but for the Ministry of Defence. We did so. He invented the questions there and then:

'Mr. Woolley, are you worried about the danger of war?

'Yes,' I said, quite honestly.

'Are you unhappy about the growth of armaments?'

'Yes.'

'Do you think there's a danger in giving young people guns and teaching them how to kill?'

'Yes.'

'Do you think it is wrong to force people to take up arms against their will?'

'Yes.'

Would you oppose the reintroduction of National Service?'
Chapter 5

I'd said 'Yes' before I'd even realised it, d'you see?

Humphrey was crowing with delight. "You see, Bernard," he said to me, "you're the perfect Balanced Sample."

5.7.10 Avoid Double Negatives

In the interests of clarity, double-negatives should normally be avoided even though, to the survey designer, the idea to be tested might most accurately be described in terms of a double-negative. An example of a double negative: "Do you not think that the reduction of trams is undesirable?"

5.7.11 Stressful Questions

Questions should generally be non-stressful and non-threatening. Respondents should not be forced into admitting anti-social or non-prestigious behaviour in order to answer a question truthfully.

5.7.12 Avoid Grossly Hypothetical Questions

Grossly hypothetical questions should be avoided, or, at least, the answers should be treated with some caution unless the answers are cross-referenced to other questions or information, or unless the respondent is faced with a clear and realistic trade-off situation when making a response. Thus the question "Would you like a more frequent bus service?" is useless because most people would obviously answer in the affirmative. However, the question "Would you prefer the frequency of buses to be doubled if the fare also increased by 50 percent?" could provide useful information on the trade-off between frequency and fares.

This issue is of particular importance as the use of stated preference surveys (Section 5.5.4) gain popularity in transportation planning (Hensher and Louviere, 1979; Bonsall, 1985; Kroes and Sheldon, 1988; Polak, 1994).

5.7.13 Allow for the Effect of Response Styles

The effect of response styles should be accounted for in the wording and layout of closed questions, especially opinion questions. A response style is a tendency to choose a certain response category regardless of the question content. Several response styles are of particular importance. First, there is the acquiescence response style where some respondents, in an agree/disagree question format, consistently choose "agree" even when the content of the question is reversed. To overcome this, one could specify an equal number of positively and negatively worded statements such that the acquiescence effect is counterbalanced. Alternatively, one could rephrase the question such that a specific response, rather than an agree/disagree response, is required.
A second style is governed by *social desirability* where a respondent always gives answers which are most favourable to self-esteem irrespective of the respondent's true opinions.

A third response style consists of *order or position biases* when answering multiple choice questions or using rating scales. Thus, for example, Payne (1951) has noted that respondents will be more inclined towards the middle in a list of numbers, towards the extremes in a list of ideas, and towards the second alternative in a list of two ideas. Similarly, some individuals will consistently mark on the left, right, or centre of horizontal rating scales. Order and position biases can be controlled by having a number of alternative forms on which the positions and orders are reversed or randomly distributed. The expense incurred in printing multiple versions of a questionnaire, however, can only be justified if the biases involved are severe, although the increasing use of computer-based surveys makes this option much more feasible.

### 5.7.14 Care with Periodicity Questions

When asking questions concerning the frequency of periodical behaviour (such as moving house, or performing holiday travel) there are a number of alternative question formats. One could ask, for example, "How many holiday trips do you usually make per year?", or "How many times did you go to work last week?" The definition of last week, for example, can be any of the following:

- Sunday to Saturday
- Monday to Sunday
- Monday to Friday
- Today - back seven days
- Yesterday - back seven days

Clearly these self-defined weeks can lead to significantly different response time-frames and hence results. Possibly the best solution is to say "How many times in the last seven days...?" in the 1981 Sydney Travel Survey, this was supplemented with a question which ran, "Did you work yesterday (say, Tuesday), did you work Monday? did you work Sunday? etc., i.e. working backwards through each of the preceding seven days?

Naturally, the level of effort used in this type of question relates to the survey objectives, but it is possible to gather very precise data if the questionnaire is designed carefully.

### 5.7.15 Use of an Activity Framework

As discussed in Section 5.3.5, the recording of trips will be done most accurately by respondents if the trip is placed within an activity framework. This means that
respondents are asked "what they did" rather than "where they travelled" to encourage them to put travel in context. Further, if a diary is used to assist in the recording of trips as they are made, the results can be improved even further.

5.7.16 Flow of the Question

An important test of questionnaire wording is to check whether the question reads well. Several features should be considered, such as insuring that the key idea in the question appears last. To avoid the respondent prejudging the intent of the question, all qualifiers, conditions and other less important material should be placed at the start of the question. Punctuation, in the form of commas, semi-colons, etc., should be kept to a minimum. The objective is to get a question which reads well, not one which is necessarily strictly correct grammatically. For example, "Who did you travel with?" should actually be "With whom did you travel?", but the former is likely to give the impression of a much more respondent-friendly questionnaire than the latter.

Key words in the question should be identified by means of underlining or the use of a different type-face. Finally, abbreviations should not be used no matter how much space they save or how well you think the respondent might know the abbreviation.

5.7.17 Always Use a Pilot Test

The acid test of questionnaire wording is the conduct of a pre-test or pilot survey. As we have mentioned before, this test should not be confined to your work associates, who probably think along the same lines as you do anyway, but should include people from the same population that are to be surveyed in the main survey.

5.8 QUESTION ORDERING

Given a set of well-worded questions, consideration should be given to the ordering of questions to ensure a smooth, successfully completed interview. The opening questions may be of two forms. First, any screening questions should be asked so that wasted time is kept to a minimum. If respondents are not eligible for membership of the population it is best to find this out quickly to avoid wasting both the respondent's and the interviewer's time. Having asked these questions, or if screening questions are not needed, the opening questions should be used to put the respondent at ease, to establish rapport and to get the respondent interested in the survey. Open questions are often very useful in this context to enable the respondent to quickly air their views on the subject. Care should be taken, however, to ensure that this open-ended discussion does not carry on for too long. Otherwise, the respondent may tire before reaching the
Survey Instrument Design

main part of the interview and may become annoyed at having to repeat some of the points of closed question form later in the interview.

The body of the questionnaire should be arranged in a logical sequence moving from point to point. Where there is an unavoidable break in the train of thought, the interviewer should warn the respondent and briefly explain what the next set of questions will be about, as shown in Figure 5.24.

![Figure 5.24 An Introduction to a Set of Questions](source: Ampt (1992b))

The end of the interview is a good place for two types of questions. First, any personal questions which you believe may occasionally meet with refusal if asked earlier in the questionnaire may be asked at the end. The reason for this is twofold. First, the respondent (and the interviewer in a personal interview) will have had the chance to gain rapport by this time and hence personal questions will have a higher chance of being answered. Second, even if the respondent refuses to answer these questions at this stage and halts the interview, at least the rest of the questions will have already been answered and hence not much is lost by concluding the interview at this stage.

A second type of question at the end of an interview is an open-ended question in which the respondent's general comments are sought. These comments may be irrelevant to the survey topic but at least they will make the respondent feel better. Alternatively, the comments may be very useful in highlighting points which were not thought of in the design of the questionnaire. Either way, it provides a smooth, pleasant way of finishing an interview and taking your leave.

Within the body of the questionnaire there are a number of points worth noting. It is important to be aware of the conditioning effect of previous questions and the way in which answers to earlier questions can lead the respondent to answer later questions in a particular manner. In exploring a particular issue, one can use
Chapter 5

a funnel sequence of questions which start out very broad and successively narrows the topic down to very specific questions. Alternatively, one can use an inverted funnel in which initially narrow and specific questions lead on to more and more general questions. The use of either type of sequence will depend on whether it is difficult to elicit answers on either the general or specific topic.

While it has been stated several times that filter or skip questions or sequence guides should be used to save time, effort and frustration, one must be careful how and where such filter questions are used. For example, if filter questions are placed before each set of questions throughout the interview and if a "NO" response to a filter question means that a set of questions is skipped, it is not beyond the wit of a reluctant respondent to figure out that a "NO" response to all these questions will result in a very short interview!

A better way to use filter questions is to mix the required responses between "YES" and "NO" so that the respondent is not sure how to skip a set of questions or, alternatively, to ask all filter questions at the start of the interview so that the respondent does not realise that the answers are being used to determine whether or not to ask certain sets of questions later in the interview.

A final point on question ordering is to realise that fatigue is likely to set in towards the end of a long interview or questionnaire and that this may tend to bias the answers to latter questions. It is therefore desirable to keep the more important questions away from the end of the interview, if possible. Better still, keep the questionnaire short!

5.9 QUESTION INSTRUCTIONS

The questionnaire or interview form is only one part of the documentation needed to conduct the survey. Also needed is a set of instructions explaining to respondents how to fill in the questionnaire, or a set of instructions explaining to interviewers how to conduct the interview.

5.9.1 Self-Completion Surveys

With self-completion surveys, instructions may be of two types: general instructions which apply to the entire questionnaire and specific instructions which apply to particular questions. General instructions should be placed at the beginning of the questionnaire while specific instructions should be placed directly in front of the question to which they refer. Often, for small questionnaires, it is possible to include an example questionnaire which has already been completed to provide some guidance on the methods to be used in completing the questionnaire, as shown in Figure 5.25.
Survey Instrument Design

Care should, however, be taken to ensure that the responses on the example questionnaire do not lead the respondent to give specific answers. Also, if an example questionnaire is provided it should be referred to in the introductory instructions for the questionnaire.

In general, instructions for self-completion questionnaires should be kept to a minimum. In some respects, good design and wording of the questionnaire should largely eliminate the need for extensive instructions. Besides, many respondents do not read long instructions but merely prefer to answer the questions as they see fit. Therefore, instructions should only be used to define key terms, to give directions for skipping questions and to explain questions which, after extensive pilot-testing, still appear to need instructions for unambiguous completion.

Another set of question instructions in a self-completion survey should be the Administrator's Manual - a manual very similar to the Interviewers’ Manual described in the next section. It contains information on the survey objectives, the reasons for all questions, data definitions, and so on, to ensure that when the administrative personnel are phoned by any respondents, they are as knowledgable as the researchers on all these aspects.
5.9.2 Personal Interview Surveys

With personal interview surveys, the need for instructions is much more broadly based. Not only are instructions needed to assist the interviewer in asking the questions, but the instructions should help in all facets of the conduct of the interview. Often these instructions are formulated in the form of an Interviewer Instructions Manual. One widely-quoted manual is that produced by the Survey Research Center at the University of Michigan (University of Michigan, 1976). An excellent summary is available in Warwick and Lininger (1975).
Survey Instrument Design

In addition to an Interviewers’ Manual which interviewers can refer to at any time, a training session is important. At this training session it is important that the survey designer instil in the interviewers not only the answers to a set of rules, but an understanding of the entire survey process. In particular, interviewers need to understand why the survey is taking place, how the sample was chosen, why, for whom is the work being done, and so on. Only with this level of background can interviewers feel confident and hence achieve the levels of response which are necessary for gaining robust data. This implies that the traditional one-two hour “briefing” is rarely enough to adequately provide interviewers with the proper level of information about the survey. It is usually better to think in terms of “interviewer training” - even for experienced interviewers it is likely that the topic and application will be new - and plan to set aside a whole day for the exercise. It is certain to pay off in terms of the quality of the data obtained - and usually in the loyalty of the interviewers who find this type of treatment a sign of respect of their abilities. For new interviewers using the personal interview survey described in Section 7.3, the training sessions need to be about 3 days in length, including practice sessions and actual field experience in the evenings.

In this training session and in the Interviewers' Manual, guidance should be provided to the interviewer on the following components of the interview task:

(a) Understanding the survey objectives. The interviewers should know and understand the survey objectives, since this helps clarify the need for many of the features of the survey. If, for example, the objective of the survey is to measure people’s exposure to the risk of accident, it becomes relatively easy to understand why one of the questions for each walk trip is, “How many streets did you cross?”

(b) Explaining the survey - it is essential that respondents know a bit about the background of the survey. The manual should provide a graduated series of explanations for the interviewer to use to satisfy varying degrees of curiosity. Knowledge of the survey helps give the interviewers a professional demeanour which is essential to gaining high response rates.

(c) Finding the sample - often the sample may not be easy to locate, e.g. finding households in high-density inner urban areas or, conversely, in low density semi-rural or rural settings. The interviewer should also be instructed on how many call-backs to make if the respondent cannot be contacted immediately.

(d) The method by which the sample was selected. This provides the interviewer with an understanding on what do if there are problems in
finding the sample and in explaining the way the sample was chosen to the respondents.

(e) Gaining entry to household - often the most difficult task is perceived to be gaining initial acceptance by the respondent. The Interviewers' Manual should provide guidance on factors such as the effects of the appearance of the interviewer, the timing of the initial contact, the opening introductions and the methods of conveying a sense of respectability.

(f) Handling problem situations - often the interviewer will have to deal with comments like "I'm too busy", "I'm not interested", "Who's behind this survey?", "Do I have to fill it out?", "What's the use of the survey?", or "Sorry, he's not at home". The manual should provide methods of dealing with these, and other, potentially embarrassing situations. The manual should also describe how to deal with outright refusals.

(g) Dealing with third parties - often when asking opinion questions it is desirable for the interviewer to be alone with the respondent. In family-situations, however, this may not always be possible. The manual should describe how to "get rid of" unwanted third parties.

(h) Asking the questions - instructions to interviewers for asking the questions may be included generally in the instructions manual and specifically on the questionnaire form. In the instructions manual the following guidelines are generally appropriate:

(i) Use the questionnaire carefully, making sure to ask the questions verbatim. Asking questions in a different way is the same as asking different questions;
(ii) Know the specific purpose of each question;
(iii) Know the extent of probing allowed on each question;
(iv) Follow the order of questions indicated on the questionnaire;
(v) Ask every (appropriate) question;
(vi) Do not suggest answers;
(vii) Do not leave any questions blank.

(i) Probing techniques - in an effort to obtain an adequate response to each question, the manual should describe some of the types of probes which are acceptable, and give examples of those which are not acceptable. The type of probe allowed for each question (if any) should be indicated either on the questionnaire or in the manual. Some types of acceptable probes are:
Survey Instrument Design

(i) The "pregnant pause" which encourages respondents to feel they should fill in the silence with an answer (not to be confused with the "embarrassed silence");
(ii) Overt encouragement;
(iii) Elaboration;
(iv) Clarification;
(v) Repetition.

All probes are acceptable only if they are neutral and do not lead the respondent to answer in a specified way.

(j) Recording the responses - particular attention needs to be paid to the methods of recording responses to open and field-coded questions. The following instructions provide some useful guidelines:
(i) Record all answers immediately;
(ii) Abbreviate words and sentences according to an agreed system;
(iii) Include all probes used;
(iv) Think about the writing device to be used. It is easier for interviewers (and data processors) to record responses in a colour different from the one in which the questionnaire is printed. If it is an outdoor survey, consider the weather (see Section 5.4).

(j) Concluding the interview - convenient and non-hasty ways of concluding the interview should be suggested in the instructions.
6. Pilot Surveys

This chapter is relatively short but very important! At this stage in the survey process, we have a questionnaire and a sample of respondents to whom we wish to administer the questionnaire. It would therefore seem appropriate to go ahead and do just that. However, as shown in the survey process diagram (Figure 1.1), it is wise, if not essential, to first perform a pilot survey before embarking on the main survey.

6.1 WHY A PILOT SURVEY?

We can hear you saying that you know of many survey designers who have "jumped in at the deep end" without using a pilot survey and survived. They may have survived, but the credibility of their data has undoubtedly been questioned. And it is unlikely that they have had reasonable answers to justify the queries. If a survey question was asked badly, or if a sample was chosen incorrectly, there often is no real answer. The survey was simply a waste of time.

Check with anyone who has designed a survey - ask them if they can be sure the data they collected was reliable (ask yourselves if you have designed a survey!). We can almost guarantee that those people who did not do a pilot survey will be able to mention many weaknesses. Indeed, those of us who carry out pilot
surveys as a matter of course still find small problems as we go into the main survey, so it is not conceivable that pilot-free surveys do not have problems.

Although pilot testing is one of the most important components of the survey procedure, it is also one of the most neglected. The usual reasons for not doing pilot surveys are said to be either lack of time or money (or both). However, not doing a pilot survey, or at least a series of pre-tests, almost always turns out to be a false economy. If the survey designer has been correct in all the assumptions they have made in the design of the sample and the questionnaire, then the pilot survey will not pick up any problems and, in many cases, the data obtained can be combined with the rest of the data obtained from the main survey. In such a case the pilot survey will have effectively cost nothing. If, on the other hand, the survey designer has been less than perfect and the pilot survey does detect some problems, then at least these problems can be rectified before the main survey is carried out. In this case, the pilot survey will have saved the cost of the errors occurring in the main survey. For the above reasons, a pilot survey is a useful fail-safe precaution to take before the conduct of the main survey.

Even survey techniques and questionnaires which have been used successfully in similar circumstances in other surveys by the same survey designer have effectively been subjected to extensive pilot testing and need to be tested again if they are carried out on anyone other than the original population. An example of this was a large-scale self-completion travel survey carried out by the Transport Research Centre in Brisbane, Australia (Richardson and Ampt 1993a). The survey was conducted in 1992 and in 1993 the same questionnaire was to be used in Melbourne. Although it was known that the changes were very minor (different fare structures for public transport, no toll system in Melbourne etc.) a pilot survey was carried out. Some significant differences were, in fact found, which needed moderately major modifications to the questionnaire and survey design. For example, the large ethnic population needed addressing, there were many fewer holiday homes in Melbourne (which affected the sample) and a much larger number of people describing themselves as "other pensioners" needed to be dealt with in the questionnaire design (Ampt 1993, 1994; Richardson and Ampt, 1993b).

6.1.1 A Test of ALL Aspects of Survey Design

Pilot surveys are most often associated with testing the adequacy of the questionnaire instrument itself. The association of pilot surveys with questionnaire design is, however, slightly misleading. As described comprehensively by Ampt and West (1985) and Oppenheim (1992), pilot surveys are designed to test all aspects of the survey process, not just the adequacy of questionnaire design. Informal trial-and-error testing of various components of
Pilot Surveys

the survey process is performed by pre-tests, whereas pilot surveys are a systematic dress rehearsal of the main survey.

In many cases, various options may be tested within a pilot survey, within the context of a controlled experimental design, to establish which option will prove to be most effective in the main survey. An excellent example of this technique in a transportation setting is provided by Sheskin and Stopher (1982b). The survey being tested was an intercept survey of people travelling on-board buses and was testing the new design of asking respondents to fill out some questions while travelling and to take another part of the questionnaire home to fill in during their own time. Two versions of the on-board form and three versions of the take-home form were devised. The versions were distributed in a systematic mix to consecutive bus travellers as they boarded, to assure that, as far as possible, the full range of survey instruments was distributed at each stop.

There were many interesting findings, one of which was that the longer version of the on-board form actually gained a better response than its shorter counterpart. Apparently the presence of some perceptual questions sparked respondents' interest in the longer form.

Not surprisingly, some rewording was suggested after careful scrutiny of the returns in the pilot. In addition, it was found that it was possible to remove one entire section of the survey questions (about a specific mode), since sufficient responses were actually being obtained elsewhere in the questionnaire. Thus, some very positive, and in the long run, cost-saving measures were learned from the extensive pilot study of the on-board and take-home forms. More importantly, a small in-house pre-test on secretarial staff had failed to uncover the full extent of the problem revealed in the pilot study. As the authors summarise, "Had a decision been made on the basis of the in-house pre-test to use the form, the expensive main survey might have failed to generate data of sufficient quality to support the modelling effort".

6.1.2 A Need for Several Tests

In large surveys it may well be advisable to conduct more than one pilot survey. For example, in connection with the 1981 Sydney Travel Survey, Ampt and West (1985) describe the conduct of four different pilot surveys as described in Table 6.1.

The four pilot surveys were spread over a considerable period of time (nearly 2 years). The information gathered very early in the first exploratory skirmish was collected at a point in time when the final form of the survey was far from clear, but nonetheless it had a considerable impact on shaping the final survey. For example, it was initially hoped that some of the more advanced methods of travel demand modelling, exemplified, for example, by the work of Jones et al. (1983)
Chapter 6

and Brög and Erl (1980), could have been applied at a large metropolitan area scale. It soon became clear, however, that such an ambition was not feasible at that scale, and the direction of the data collection and modelling efforts were adjusted accordingly very early in the process before too much time, money and intellectual effort had been invested in further development of the modelling processes.

Table 6.1  Pilot Studies for the 1981 Sydney Travel Survey  
Source: Ampt and West (1985)

<table>
<thead>
<tr>
<th>PILOT STUDY</th>
<th>TIME</th>
<th>TESTING OBJECTIVES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Exploratory</td>
<td>July 1979</td>
<td>To observe travel determinants and to test the feasibility of collecting behavioural data on a large scale.</td>
</tr>
<tr>
<td>Skirmish</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Skirmish</td>
<td>Aug. 1980</td>
<td>To test the effectiveness of questionnaire wording and design.</td>
</tr>
<tr>
<td>4. Dress Rehearsal</td>
<td>April 1981</td>
<td>Final test of all aspects including coding and all operational details.</td>
</tr>
</tbody>
</table>

6.2  USES OF THE PILOT SURVEY

Pilot surveys (including various types of pre-test) are used to give guidance on all of the following aspects of survey design:

6.2.1  Adequacy of the Sampling Frame

Is the chosen sampling frame complete, accurate, up-to-date and easy to use? The basic question to be answered is how well does the sampling frame correspond to the desired survey population. If the frame gives very poor correspondence or is grossly deficient in any other way, it is as well to know this before the main survey begins. If the frame is only slightly deficient it may be possible to remedy this and still use it in the main survey. For example, the pilot survey might suggest some screening questions which might be added to the start of the interview to refine the final target population.

6.2.2  Variability of Parameters Within the Survey Population

As described in Chapter 4.6, it is necessary to have an estimate of the population variability of the parameter to be estimated before the required sample size can be calculated. Pilot surveys can be useful in supplying estimates of variability for a number of parameters of interest. It is of dubious value, however, to use pilot
Pilot Surveys

studies as the sole source of information on parameter variability. Pilot studies are generally restricted to a relatively small sample and, hence, the reliability of an estimate of the variance (which is itself subject to a standard error) will not be very great. To obtain reliable estimates of parameter variance would require a large sample size, which is generally not feasible in a pilot survey. The pilot survey should therefore be relied on to provide supporting evidence only, except where no other sources of prior information are available.

6.2.3 Non-Response Rate

The probable number of refusals or non-contacts (households where no contact can be made after repeated visits in a personal interview survey) and instances of sample loss (e.g. vacant dwellings, non-existent phone numbers, etc.) can be estimated from a pilot survey, provided that the same methods are used in the pilot survey as will be used in the main survey. Alternatively, a number of different methods of reducing non-response may be tried in a pilot survey and the results then compared to determine the most effective method.

Causes of non-response should be noted, to assist in the possible redesign of the survey procedure. The magnitude of non-response can also assist in estimating the total number of sampling units to be chosen from the population (to obtain a specified number of completed questionnaires).

If possible, some indication of the characteristics of non-respondents should be obtained to determine whether non-response bias is likely to be a major problem in the main survey. This can be done by following up non-respondents to check whether their travel behaviour is significantly different than that of respondents. A methodology for doing this is described in Chapter 7.

6.2.4 Method of Data Collection

The ease or difficulty of collecting the required data may indicate whether an appropriate data collection method is being employed. For example, very low response rates and a general lack of understanding of a self-completion questionnaire may indicate that the wrong survey method is being used. Perhaps a personal interview survey is necessary to increase response rates and to provide for the assistance of an interviewer in asking questions and recording responses.

An interesting example of testing various survey methods to collect the same data (Ampt, 1992a) used 6 different data collection methods: personal interview, and self-completion methods which tested 1 and 2 day variations of "linked" and "un-linked" (stage) data collection methods (see Chapter 2.5 for a definition).
Chapter 6

6.2.5 "Skirmishing" of Question Wording

An important reason for carrying out a pilot survey is to test the wording of the questions in the survey. One way of doing this is to get respondents to complete the questionnaire (if it is self-completion) or to answer the questions in a personal interview, and then go over each of the questions with an interview and, using a completely open question, check the respondent’s understanding of the question. This process, which is often time-consuming in the short term, is an invaluable method of ensuring that respondents not only understand the questionnaire in the same way as the survey designer, but also that everyone is understanding the question in the same way. As we have mentioned several times, pilot surveys - but tests of wording in particular - need to be carried with the type of people who will complete the main survey.

A similar procedure which has been used with success in self-completion surveys is a type of "participant observation" survey. This involves giving the respondents the questionnaires in the envelope (in the way they would receive them in the mail) and observing all behaviour from emptying the contents of the envelope to filling in the forms. This has proved invaluable in understanding many of the problems which cannot otherwise be understood in trying to develop these survey forms. Small hesitations, and observing common "wrong moves" with the pen, all give clues about the user-friendliness of the design.

6.2.6 Layout of the Questionnaire

Apart from question wording of the questionnaire, layout is also of critical importance in both self-completion and personal interview surveys. For example, is the location of the tick-boxes relative to the respective answers causing a problem (see Section 5.7.5)? To discover this, it will often be necessary to actually ask respondents or interviewers what they meant, since the analyst will only see a ticked box and the respondent/interviewer will have had no trouble ticking a box according to their own perceptions.

Colour and contrast can be used to good effect on the questionnaire form, but when things are written too small, together with the use of the "wrong" colours, they tend to be overlooked by the person completing the form, and the form may need re-working after the pilot survey.

6.2.7 Adequacy of the Questionnaire in General

This is the reason most often voiced for the need for pilot surveys. Although the questions will probably already have been tested on colleagues and friends, a pilot test using the same kind of respondents (and interviewers) as will be encountered in the main survey is the only true test. The aspects of questionnaire design which need to be specifically checked in pilot surveys include:
(i) The ease of handling the questionnaire in the field - are instructions clear and are answers easy to record? Are the survey forms of a manageable size and shape?

(ii) Are definitions clear? Are there any consistent misinterpretations on the part of either the respondents or the interviewers?

(iii) Are questions clear and unambiguous? Are there signs that respondents, or interviewers, have misunderstood the intent of the question?

(iv) Are the letters of introduction signed by the right people - too prestigious, or not prestigious enough? Is the phone query line being used?

(v) Are special techniques such as attitude rating scales producing valid answers. Too much bunching of answers may indicate a leading question or badly chosen categories. Too many "Don't know" responses might indicate a vague question, one which is misunderstood by respondents or one which is simply pointless. Too many refusals to a question may indicate that it should be asked more delicately, or that the order of questions should be changed or that the question should be omitted.

(vi) Is the questionnaire too long? Too many unanswered questions or hurried answers towards the end of the questionnaire indicate that perhaps the questionnaire is too long for the amount of interest shown in the subject matter. Perhaps the questionnaire should be shortened or else attempts should be made to raise the level of the respondents' interest in the survey.

(vii) Can some open-questions be converted into closed-questions in the main survey? The responses obtained in the pilot survey can be used to define the closed-question categories.

(viii) Is the sequencing of questions clear? Are respondents being forced to answer a series of questions which do not pertain to them? Are more branching and skip questions needed? Is it clear what question should be answered next after a branching question?

(ix) Are the coding categories clear and adequate for closed and field-coded questions? Should some categories be removed, combined or added to the list? Is there a problem with the lack of an open alternative with any of the closed questions?
Chapter 6

One of the most important ways to test these aspects in a personal interview questionnaire is for the survey designer (usually you!) to actually carry out some interviews on the same population as will be sampled in the main survey. In our experience, there is no better way to understand the complexities of survey design!

6.2.8 Efficiency of Interviewer and Administrator Training

The pilot survey provides an ideal opportunity for testing the interviewer training procedures as well as for in-field training of interviewers. The performance of interviewers under real survey conditions may indicate areas in which retraining is necessary. Areas of poor performance can be detected by checking the collected questionnaires and by supervision of selected interviews by a field supervisor or the survey designer.

Training of administration staff is an important part of many surveys, particularly those using mail-back self-completion questionnaires. The day-to-day management of the survey office requires extremely well-organised and knowledgable people. Not only do they have to understand the daily tasks of mailing (even though this is usually computer-automated, there understanding is essential for problem-solving on bad days!), but they have to be extremely competent in dealing with respondent's phone enquiries, and often also with data entry (where this is part of the responsibility). This means that these staff need substantial training - very similar to interviewer training, in fact (Section 7.1.2). Failing to test this in the pilot survey can have significant negative effects - particularly in the first busy days of the main survey!

6.2.9 Data Entry, Editing and Analysis Procedures

The pilot survey should include these aspects of the survey design process because of the backward linkages of these parts of the process to earlier parts of the survey. The effectiveness of pre-coded questions, the ease of data entry, the completeness of coding information, the required logic and editing checks, and the relevance of the analysis output to the objectives of the survey are examples of the sort of items which should be checked in the pilot survey. There have been many occasions when a data item collected in a survey has not been in the exact form anticipated by the user of the data, and has been subsequently rejected from the analysis. A simple check in the pilot survey could readily have averted this problem.

Do not put off testing these parts of the survey process just because it seems that it is still a long time before you will have to analyse the real data - it will be too late to rectify problems in the structure of the data when it comes time to perform the real analysis.
6.2.10 Cost and Duration of Survey

The pilot survey can also provide estimates of the cost and time requirements per completed interview which will assist in the determination of the total survey resources required for the survey of the total sample. If this amount exceeds the budgeted amount, then a decision can be made at this stage as to whether to seek an enlarged budget, or to accept a reduced accuracy (through smaller sample size) or to abandon the main survey altogether.

In pilot studies for personal interview surveys, it may be possible to pay interviewers by the hour and the distance travelled, and then to calculate a per interview rate for the main survey. This not only simplifies administration procedures for the main survey, but also ensures that interviewers work efficiently during the main study.

One of the costs which is most often under-estimated in survey design is that of data-entry, validation, editing and analysis. Particularly in large-scale travel surveys, time and again we hear of stories of lack of funds for these key processes. Here again, the simple inclusion of these processes in the pilot - together with detailed costing of each stage - can help to avoid this problem.

In order to facilitate least error in the projections from the pilot survey costs to the main survey, it is necessary to plan the accounting process prior to the pilot survey. Even though the pilot survey may be relatively low cost, it is important to document all items and time spent to ensure that proper projections can be made.

6.2.11 Efficiency of Survey Organisation

The efficiency of office administration and field supervision can be checked by treating the pilot survey as a small-scale dress rehearsal of the main survey. Is the workload for interviewers and supervisors manageable in a personal interview survey? Are there any problems with travel to the interview sites, security while performing the interviews, problems with minority languages? For a self-completion mail survey, does the outbound and return mail system work as it should? How long does it take for mail to travel in each direction? What are the postage rates and how does the weight of your survey material (both outbound and inbound) compare with any thresholds in postal rates? Are the procedures for stuffing and addressing envelopes working satisfactorily and, importantly, will they work the same when you are dealing with a much larger volume of mail in the final survey? Many of these issues may seem trivial at first, but taking the time to test them in the pilot may mean that you will not have to spend a lot of time trying to work around their deficiencies in the main survey when you will have a million-and-one things to do already!
6.3 SIZE OF THE PILOT SURVEY

The size of a pilot survey is a trade-off between cost and efficiency. It cannot be as extensive as the main survey but nevertheless it should be large enough to yield significant results. This is especially important if the pilot survey is being used to compare alternative procedures in survey method, sample design or questionnaire design. In such a situation the pilot survey should be of sufficient size to ensure that if substantial differences between the methods do exist then they can be detected statistically.

Usually, however, the size of the pilot survey will be related to the size of the main survey. A reasonable rule of thumb is to expect to spend about five to ten percent of the total budget on the pilot survey. While this may seem, to some, to be a waste of resources, it generally turns out to be a wise investment. Remember that many of the tasks required for the pilot survey will be required to be done later in any event, so they may as well be done early in the process when you have a chance of correcting your mistakes. In addition, if there are no problems detected in the pilot, then it may be possible to combine the pilot survey data with the main survey data in a final data set.

Finally, with suitable cautionary notes, you can use the results of the pilot survey to present an interim report to the sponsors of the survey to let them know that the survey is in progress, and to give them a feeling for the type of results which they could expect to see from the main survey. This, in fact, can also be seen as simply a pilot survey of the final stage of the survey process, which is the presentation of results and the documentation of the survey.
7. Administration of the Survey

While pilot surveys form a vital component of survey design, as discussed in the previous chapter, several other aspects of survey design must be considered before a pilot survey can actually be administered. This chapter will describe particular aspects of the administration of various types of survey including the roles of recruitment and training of interviewers, pre-publicity, field supervision and follow-up validation procedures. It will then provide some more specific comments with respect to the five major types of survey described earlier in Chapter 3.2 (viz., self-completion, personal interview, intercept surveys, telephone surveys, and in-depth surveys). Other aspects of survey administration may be found described in texts such as Warwick and Lininger (1975), Parten (1965) Hoinville and Jowell (1978) and Pile (1991). A particularly good reference is the book on Total Design Principles by Dillman (1978). While this book was written in a pre-computer era, it highlights the myriad of details which must be attended to in the design and administration of surveys.

7.1 GENERIC PROCEDURES FOR SURVEY ADMINISTRATION

The recruitment, training, and supervision of survey personnel and the need for follow-up validation procedures are four highly interrelated tasks, since the amount of effort required on any one of these tasks is closely related to the amount of time spent on the other three. For example, a poor recruitment
Administration of the Survey

campaign will mean that more time and effort must be placed into training and field supervision at a later date. Because of the interdependent nature of these tasks they will be discussed together in general terms in the first section of this chapter. Some of the tasks relate more to one type of survey than another, although there are some common threads running through the tasks for all types of survey discussed in this book.

7.1.1 Recruitment of Survey Staff

There are three different sources of recruits for survey staff for transport surveys:

(a) Captive markets, e.g. yourself, members of your company, students, members of organisations.
(b) Market research firms.
(c) Respondents to advertisements.

Captive market recruits usually have a number of particular problems. First, the staff tend to be atypical of the general population. They may be all students, or all professional people or all members of an organisation with distinct characteristics. Second, if they have been chosen because the topic of the survey is of special relevance to them and if they are to employed as interviewers, they may be overzealous in their pursuit of answers thereby introducing a certain degree of interviewer bias into the survey. On the other hand, if they have been directed to do the interviewing or survey administration as an adjunct to their normal duties within an organisation, they may have little motivation to do a good job (especially as it may often include evening work) and thereby introduce a different type of bias. Third, some of the people chosen may not have suitable personalities or aptitudes for the task. This is problematic since the recruits may, for example, all be volunteers from a community organisation, and it is difficult to reject recruits without appearing to be ungrateful for the organisation's assistance. This usually means that considerable extra training is needed to overcome these weaknesses. In general the main advantage of using a captive market of recruits lies in the considerable cost advantage over other methods of recruitment.

The practice of using market research firms to provide interviewers can be undertaken in one of two ways. Either the firms can be approached to simply provide a group of interviewers or else the firm can be approached to conduct a much larger part of the survey. Either method has a number of advantages. First, it eliminates the need for the investigator to get involved in the actual recruitment process. Second, it is most likely that a market research firm would not need to go through a full recruitment program since it probably has a number of interviewers already "on the books".
Using market research firms does, however, have some disadvantages, especially if they are used for more than just the supply of interviewers. First, the cost of using a market research firm can be considerable. Second, it is possible that very little of the expertise gained in the conduct of the survey will stay within the sponsoring authority. Third, it is necessary to realise that the quality of market research firms varies. In particular, a market research firm which is skilled in general survey techniques may not necessarily possess the special skills and experience needed for travel surveys. For example, many market research interviewers do not understand the severe travel-data implications of choosing a neighbouring household to replace one in the sample where no-one was at home, or of not calling back to a household numerous times to ensure that not only the easily accessible respondents are reached (see Section 7.3.6). It is therefore advisable for the investigator to first determine which aspects of the survey can be handled “in-house” and then to bring in specialised consultants for the remaining tasks for which they possess special skills. It is wise to ascertain the methods used by the market research firm and their previous experience and then retain some control over survey procedures.

The third type of recruitment is the most complex, but if performed correctly, it is the most satisfactory method of recruiting staff for a particular survey. In this method, the staff are hired by the investigator for the survey at hand. The first step in this recruitment process is for the investigator to decide just what type of person is needed to do the job. In particular, the following questions need to be answered:

(a) What type of person can do the work?

(b) Does it require highly skilled people with a special knowledge of the subject matter of the survey, or will a comprehensive training program be provided?

(c) What facilities are required by the staff/interviewers, e.g. is it essential that they have a driver's licence and access to a car.

(d) What degree of availability is required of the staff/interviewers, e.g. will they be required to work during the day or at night or on weekends?

(e) Are any specific abilities required such as the ability to speak a foreign language?

Given that the investigator can adequately answer these questions, the next step is to draw up an application form which will obtain a profile of the person applying for an survey staff position, which can then be matched against the desired staff profile to assist in staff selection.

The application form should ideally include questions on the following topics:
(a) Name;
(b) Home address;
(c) Home phone number;
(d) Date of birth;
(e) Highest education level reached;
(f) Previous study or work experience which may be relevant to interviewing;
(g) Foreign languages spoken (if any);
(h) Car availability;
(i) Current employment and hours of work/study;
(j) Health condition;
(k) Availability for training;
(l) References;
(m) Reasons for applying;
(n) Source of information on interviewing job.

Items (a) to (g) on the above list give a general description of the applicant and their abilities which may be compared with the desired profile identified earlier. Items (h) to (k) provide information on possible constraints which the applicant might face in carrying out the interviewing task. These questions are included not only to provide information to the investigator, but also to make the applicant aware of the constraints which they may face. It is probable that full realisation of these constraints may dissuade many potential applicants from making an application, thereby saving the investigator time and money by not having to process applications which will, eventually, probably be rejected. In asking for this information, you need to be aware of any legal restrictions you may face locally with respect to Equal Employment Opportunities and other legislation. Item (l), references, is a useful piece of supportive information for the earlier claims. Item (m), reasons for applying for the job, should provide some indication of the general motivation of the applicant. It may also be a warning to watch for any excessive interviewer bias if the applicant expresses a keen interest in the subject of the survey. Finally, item (n) can provide useful data to assess the effectiveness of different types of advertising. This information may be useful for subsequent recruitment of more survey staff.

Following on from this last point, it is necessary at this stage to determine where advertisements should be placed to attract the appropriate type of person. The choice of media for the placement of advertisements essentially lies between:
(a) Newspapers - national, urban, and local;
(b) Specialised journals and newspapers;
(c) Electronic media - radio, television.

The choice of media will depend to a large extent on the scale of the survey, both in terms of the number of staff required and the geographic spread of the interview locations in cases where interviewers are needed. The use of electronic media will generally attract a lot of applicants, but generally they are of a lower quality and this will entail a lot of expense in screening of unsuitable applicants.

The type of newspaper advertisements used can significantly affect the type of applicants obtained. First, the choice of newspaper will result in a considerable variation in the characteristics of applicants, depending on the readership educational level, political orientation, etc. Second, the placement of the advertisement within the newspaper will also affect response. Different applicants will be attracted depending on whether it is placed in the classified section or in the body of the newspaper. If placed in the body of the newspaper, different responses will result if it is placed in the business news section as compared to near the shopping specials or in the sports section. The guiding rule in all this is to know, beforehand, the type of person wanted and then to place advertisements where that type of person is most likely to see them.

As an alternative to placing advertisements in the general media, it might be useful to approach directly several sources of possible applicants, such as the local unemployment office. This has been shown to be very successful in Australia and New Zealand. It has the advantage that people are often very keen to do the work, and there is rarely the problem of "untraining" habits which market researchers sometimes exhibit. Second, most universities have a Careers and Appointments Office where students register for part-time and vacation employment. These should only be considered if the survey work is confined to holiday periods, or the employment is to be part-time. Third, it might be useful to contact schools with a view to offering part-time employment to teachers during vacation periods. Applications received from each of these three sources should, however, be treated the same as applications received in response to media advertisements, i.e. they will still need to be screened since not all applicants are suitable.

The selection of suitable applicants from the applications received should proceed in three stages. First, all applicants would be required to complete the formal application form described earlier. The completion of this form serves three functions in the selection process. By requiring all applicants to fill in the form in their own handwriting one can quickly see whether their writing is sufficiently clear not to present problems to other people who may later have to
Administration of the Survey

read their writing (e.g. data enterers). Completion of the form can also test the applicant's ability to follow written instructions. Finally, the information on the form can give a general indication of the applicant's suitability and, more importantly, will indicate any constraints which would make the applicant completely unsuitable.

The second stage of selection may involve some formal testing procedures to screen out applicants with insufficient skills in a number of areas. The general types of skill required for all survey staff may be thought of as intelligence, word knowledge, clerical skills (an ordered mind), familiarity with the computer if the staff are to do data entry, and telephone skills if the use of the phone is important.

The third stage, for applicants who have passed the two initial stages, is a personal interview to determine whether applicants have the necessary personal qualities to be good survey staff. The qualities which one would look for in survey staff would include:

(a) The person must be able to understand the general purpose of the survey.

(b) If they are to be interviewers:

- they should be of presentable appearance to encourage acceptance into respondent's homes
- they should have a similar background to the potential respondents to encourage the development of empathy during the interview.

(c) If they are to employed as either interviewers or office staff dealing with phone enquiries:

- they must be enough of an extrovert to enable them to keep an interview moving and to elicit answers from reluctant respondents.
- they must be resilient enough to withstand the occasional refusal.

(d) If necessary, the person should be able to speak a foreign language which will likely be encountered in the study area.

(e) They must be of general honest character such that there is no great danger of fabrication of results.

At the end of these three stages, a number of applicants will be selected for training.

7.1.2 Survey Staff Training

Having recruited apparently suitable applicants, it is now necessary to train them in the skills of survey administration, interviewing, or data entry. Three factors need to be considered when discussing staff training:
(a) Why should we train applicants?
(b) On what should we train them?
(c) How should we train them?

7.1.2.1 Why train staff?

It seems obvious that people should be trained for interviewing, but what particular reasons can be offered for the expenditure required in the training program. There appear to be three basic reasons for interviewer training:

(a) Increased response rates - experience has shown an 8-10% difference in response rates between "trainee" and "trained" interviewers. Besides increasing the value of the data in terms of reduced sampling bias, higher response rates also have the pragmatic advantage of decreasing the probability of irate respondents complaining to supervisors.

(b) Increased likelihood of following instructions - by training people in the purpose of the survey, they are more likely to see the need for following instructions exactly. For example, if people are trained in the purpose of sampling, there is likely to be no substitution sampling in the field. Similarly, awareness of the biases which can arise from changes in question wording will induce people to ask the questions verbatim where necessary.

(c) Reduced error rate - training can reduce clerical errors in completing the interview form quite substantially. This will result in a consequential reduction in editing effort required. Trained interviewers have about half the error rate of trainee interviewers.

But what about training for administration staff and data enterers? Administration staff in a self-completion survey have several important roles to fill - each of which is carried out best if the staff have a comprehensive understanding of the total survey process.

(a) Mailing of questionnaires. Knowledge of sampling procedures allows survey staff to understand important objectives such as why questionnaires have to reach the household before the "travel day", and what to do with inadequate addresses.

(b) Receipt of questionnaires. Appropriately designating returned questionnaires or completed interviews as fully or partly responding, or as sample loss, can only be done with a good understanding of the survey objectives.

(c) Answering phone enquiries. Since every phone enquiry - even an irate person - can be considered to be a potential respondent,
response rates can be increased considerably be knowledgeable, friendly staff answering phones.

Many travel survey designers are opting to use "intelligent" data enterers as part of the quality control process. If properly trained, these people can not only enter data, but can highlight suspected inconsistencies in the data and initiate procedures to follow-up respondents.

7.1.2.2 Content of Training Sessions

Given that we see training to be essential, the question then becomes one of what do we train these people to do. While training is often only considered to apply to interviewers, administrators (particularly of self-completion surveys) and data entry personnel should also attend training sessions. There are six broad areas of training - some are suitable for all three types of people, and others are confined to interviewers.

(a) Survey objectives and design - the underlying reason for carrying out the survey, the objectives as set out by the investigator and/or client, and the reason for selecting the methodology which has been chosen. Interviewers, administrators and data entry people all need to know these in order to carry out their jobs efficiently and to make wise decisions throughout the survey period.

(b) Subject matter of survey - all three groups of people need to be aware of several factors concerning the subject matter of the survey. They need to know the background of the survey organisation, the subject matter technical details, the survey administration procedure, and how the data will be used. Training on the subject matter will enable the interviewer and administrator to answer any questions raised by respondents in the field. More importantly, good training will give the all these people confidence in, and commitment to, the aims of the survey. Such confidence is highly contagious to respondents and generally results in higher response rates and respondent co-operation.

(c) Questionnaire details - interviewers, administrators and data entry personnel should be familiarised with the details of each question on the questionnaire to be used.

(d) Technique - the technique and skills of personal interviewing is a critical area in need of training for interviewers. Definition of terms should be clarified, coding conventions for field-coded questions outlined, and the degree of probing and clarification allowed for each question should be specified.

The type of issues which should be covered are:
Administration of the Survey

(i) How to gain co-operation;
(ii) How to handle difficult cases;
(iii) How and when to clarify answers;
(iv) How and when to probe;
(v) Clerical recording techniques.

These issues should also be addressed in the Interviewers’ Manual (Section 5.9) and in a manual for administrators and data enterers. All these manuals need to be given to the client as vital articles of documentation at the end of the study.

(e) Sampling frame - all people involved in the survey need to know how the sample was selected and what sampling frame was used. This very often helps answer respondents questions about "why me?". For example, if dwellings were selected from a list of people connected to water, it is possible simply to say that it was not they who were selected, but their dwelling.

(f) Sampling - interviewers should be trained in the basic principles of sampling and, in particular, they should be told about the various sources of sampling bias which exist. They should be made to realise the importance of adhering to the selected sample. In this way there is a greater chance of them adhering to sampling instructions or else, if in a difficult position in the field, they will be able to make a decision of whether to contact their supervisor for assistance or whether they can change the selected sample without affecting the randomness of the sample.

(g) Administrative details - information on this aspect of the survey is necessary for interviewers, in particular, if the administration of the survey is to proceed smoothly. Factors which need to be covered include:

(i) Who is the immediate supervisor in case of field difficulties?
(ii) Who do they report to for collection and return of interview forms?
(iii) How are time sheets to be filled out?
(iv) How do they get paid?
(v) What times are they expected to work?
(vi) Where do they obtain their lists of households for interviewing?
7.1.2.3 Training Methods

The methods to be used in training are many. It has been found that the distribution of a Manual prior to the training session (e.g. Ampt 1994), with some Home Study Exercises including a tape recording of the way an interview is likely to run, can be invaluable in ensuring that participants learn the maximum during the training sessions. Lectures and tutorials with extensive use of audio-visual materials (especially video-tapes) can also be very useful.

For interviewers, however, the most important aspect of training is practice, which can be of two types; group practice in class and field practice under supervision of a more experienced interviewer. Group practice in class is a convenient method of gaining practice where three or four interviewers interview each other in turn. In this way, lessons are learned by being both an interviewer and a respondent. It works best where mock answers are arranged beforehand, although if this is not possible, it is possible to interviewers to interview each other. Make sure that when they are "playing respondent" they use real examples and not fictitious ones since this can get very confusing. These group practice sessions should include both the asking of questions and the recording of answers, and should be video-taped if possible.

Survey administrators primarily need to be trained on the underlying reasons for the survey and the design of the questionnaire. This gives them important background information for all their work. If they are to be supervisors of interviewers, however, they should always complete the interviewer training course, including carrying out some interviews. Similarly, if they are to supervise data entry personnel, they should both undertake that training course and carry out some data entry themselves.

During training, data entry staff need to be given plenty of chances to see (and correct) the errors they make, as this is the best way of ensuring that their skills are of high quality.

While the procedures finally adopted in recruitment and training will depend to a large extent on the scale of the survey and the amount of previous experience held by the survey staff, the above discussion will provide a framework for such a process. More complete details on recruitment and on training may be found in Warwick and Lingen (1975) and Moser and Kalton (1979).

7.1.3 Pre-Publicity

In an attempt to increase response rate, it is often considered useful to conduct some pre-publicity for the survey some time before the survey takes place. This publicity may be pitched at two levels; at the general population, and at the respondents in the sample.
Administration of the Survey

Publicity aimed at the general population is normally by way of the general media, particularly the newspapers. As with the recruitment of survey staff, the selection of newspapers for publicity, by means of a community interest story, should be guided by the location and characteristics of the survey population. As well as arousing interest in the survey, the newspaper clipping of the story can be a very strong focal point when attempting to gain initial acceptance at the household’s front door on the day of the interview. With media publicity, care should be taken not to oversell the survey since this may arouse negative feelings and suspicions about the survey’s purpose.

Another issue lies in the fact that if the survey is being run over a period longer than a few weeks, it is difficult to decide when the publicity should be run so as to affect all respondents equally. This is important, because response rates may be affected (either positively or negatively) by publicity, and clearly it is important to know and measure which response levels are “normal” and which are affected by publicity. For this reason, it is often argued that no publicity should be done for longer-term surveys.

The more important form of pre-publicity is direct contact with the respondents in the sample. The normal form of contact is by means of a letter of introduction explaining the survey purposes and advising that an interviewer will be calling in the near future or that a questionnaire will be sent. The letter should be fairly brief and should be sent out only a few days before the interviewer is to call. The purpose of this letter is threefold. First, it makes respondents aware that an interviewer will be calling or that a questionnaire will be arriving in the mail, and hence they will be expected. In a personal interview survey, the interviewer is therefore less likely to be mistaken for a door-to-door salesman when initial contact is made, and hence refusal rates should decrease. Secondly, it gives the respondent the opportunity to think about the general subject matter and this may improve the quality of response.

In some survey designs, the interviewer is required to contact the respondent directly by phone to arrange for a mutually convenient time at which to hold the interview. While phone contact is sometimes useful as a follow-up to an introductory letter, it should never be used alone. It is much easier for someone to refuse an interview over the phone than it is when you are standing at the door. Phone contact should generally only be used in special circumstances.

The combination of local newspaper articles, introductory letter and phone contact was used to good effect in a survey of elderly travel in Sherbrooke Shire, Victoria, Australia (Richardson, 1980). In this survey, described earlier in Chapter 4.3, it was felt that the elderly residents may have been somewhat reluctant to grant the interviews since many interviews were performed at night and most of the interviewers were male (a rarity in household interviews). Effective pre-
Administration of the Survey

publicity, including the provision of a phone number for respondents to find out more about the survey, resulted in no outright refusals in a sample of 72 households.

7.1.4 Survey Execution and Monitoring

While attention paid to recruitment and training of survey staff will reduce the problems encountered in the field, such problems will never be eliminated and the survey administration process needs to take this into account with a rigorous monitoring process.

7.1.4.1 Personal Interview Surveys

Some interviewers, while performing well in training, cannot cope with the demands of real interviewing, while others who initially perform well in the field tend to become careless after some time. For these reasons, it is generally necessary that the administration of any survey of reasonable magnitude be managed by a field supervisor. The main tasks of a field supervisor are to organise work groups, to distribute work assignments, to supervise in the field, to review completed work, and to act as a liaison with the survey office. In addition, the administrative tasks connected with the survey, including payment of interviewers, will need to be handled efficiently to ensure smooth survey operation.

For large personal interview surveys, it is generally advisable to break up the total number of interviewers into work groups of five to ten interviewers who work under one supervisor in the same geographic area. The formation of small groups enables personal interaction between interviewer and supervisor and allows the development of a sense of camaraderie within the group. This is important if the survey is to continue over a long time period or if interviews must be made under difficult conditions.

The distribution of work assignments should be carefully managed by the supervisor. The general idea is to assign work to interviewers in relatively small bundles with the requirement that all completed interviews must be returned to the supervisor before more work will be assigned. In this way, tight control is kept over the flow of work and interviewers are prevented from deferring those interviews which they perceive to be most difficult. It is also a good general rule to require interviewers to return completed interviews no more than one or two days after they were completed. This insures that interviewers carry out their editing duties on each questionnaire before they have forgotten the details of that interview. Prompt return of completed questionnaires also ensures a continuous flow of work to the data enterers.
Before passing on completed interview forms to the survey office, the field supervisor should check through the questionnaires in some detail. In this way, recording errors and other problems can quickly be drawn to the attention of the interviewer who can then make the appropriate correction. This procedure also serves to alert the interviewer to such problems before more interviews are conducted, thereby avoiding the perpetuation of systematic error or misunderstanding.

At least once during every survey, a supervisor should attend one interview with each field person to ensure that the correct procedures are being carried out. Naturally while notes can be taken during the interview, any comments to the interviewers would be given in private after the interview. Good supervisors are quick to spot interviewers who are not in the habit of reading questions verbatim, who usually probe too much, and so on, and can help them (and the data quality) by reiterating the principles of good data collection after the supervised interview.

7.1.4.2 Office Procedures

To maintain control over the administration of any type of survey, it is necessary for the survey office to keep administrative records detailing the current status of the work program. Three types of record are particularly important. They are best kept as part of a computer package or spreadsheet, but can also be kept manually:

(a) Sample control log - this is a listing of the total sample by address, with space allowed to record the current status of each sample element, such as: not yet issued, in field, questionnaire returned/completed (with date of interview and name of interviewer if appropriate), refusal, sample loss, etc.

(b) Questionnaire control log - a list of the status of each questionnaire - and of all interviewers and interviewers assigned to them in a personal interview survey. This information is useful in assessing the quality and quantity of the interviewer's work. This log is also designed to keep control of questionnaires when they are given to data entry staff and when they are returned for shelving or filing.

(c) Completed questionnaire log - a list of all completed questionnaires which have been returned to the survey office. Details of the date of return or each questionnaire, the interviewer's name, etc. should be kept in the office records.

Several computer program packages are available for survey management and control, including some which are suitable for use on micro-computers. As early as 1981, Neffendorf (1981) described the Survey Management Information System
Administration of the Survey

(SMIS). More recently Bethlehem and Hundepool (1992) gave examples of using a computer for all stages of the survey process including administration. Eash (1987) describes how the management of a small telephone-based home interview survey was enhanced in a number of ways by using a microcomputer to carry out the major administrative tasks. The tasks computer selected the sample, mailed the pre-contact letters, organised interviews, checked the status of the interviews, kept a log of successful interviews and performed some data analysis.

One aspect of survey administration which has been the subject of debate for many years is the relative merits of different forms of payment to interviewers. The two major methods are "piecework rates", where interviewers are paid per completed interview, and hourly rates. Those in favour of piecework rates argue that it provides a strong incentive to interviewers to make contact and to complete interviews, allows for greater ability to predict the total cost of the survey, and results in higher productivity in terms of number of completed interviews per dollar. It also has the advantage of being much simpler to determine the payments to be made to each interviewer.

Proponents of hourly rates, on the other hand, first point to the disadvantages of piecework rates. They argue that rather than provide an incentive to interviewers to make contacts, there is a definite incentive for interviewers not to make contacts unless there is a high chance of completing an interview. Thus difficult households tend to be avoided and call-backs are neglected since they are less likely to result in payment. This can result in substantial sampling bias. It is also claimed that once a contact has been made, there is pressure on the interviewer to finish that interview as soon as possible so that another contact can be made. In these circumstances the interviewer tends not to probe too deeply since probing can be time consuming. In travel surveys, this probably results in under-reporting of trips by respondents who cannot immediately remember their complete travel pattern. Piecework rates are therefore seen to contribute towards quantity of interviews rather than quality of interviews.

The use of hourly rates, however, also has its practical problems. Interviewers have to keep time-sheets and have to be relied upon to fill them out correctly and honestly. There is an incentive for interviewers to pad each interview, spending longer than is really necessary in each household. The productivity in terms of interviews per dollar must therefore fall.

Thus each method of payment has its own advantages and disadvantages with opinion being divided among the two methods. Generally, the hourly rates method appears to be more favoured because of its greater emphasis on quality of response. With an efficient survey management information system, it is possible to identify those interviewers who appear to be particularly slow and checks can be made to determine whether they are padding their time-sheets or
whether the responses they are getting are actually of higher than average quality.

One commonly used approach to payment methods is to use hourly rates for the pilot survey, when only experienced interviewers are used, to allow the "time and mileage data" from the pilot to be converted into piecework payments for the main survey. This ensures fairness of payment and has the second benefit of allowing more realistic calculation of the survey costs in advance.

7.1.5 Follow-up and Validation Procedures

The field component of many surveys ends when the first round of completed questionnaires arrives at the survey office for coding. It is considered that "we did our best" and that the analysis will just have to proceed with the data which have already been obtained. Often, there is a time deadline for the results to be delivered to the sponsor, and it is considered that it is simply not worthwhile to "chase" those in the sample who did not respond. However, it is becoming increasingly common for follow-ups to be seen as an integral part of the total process of conducting a survey, whether it be a personal interview or a self-completion survey and for the time and costs of these process to be calculated as part of the survey process. Various methods of follow-up and validation which are applicable to the different types of survey will be addressed in the following sections.

7.1.6 Confidentiality

Confidentiality of survey data is an extremely important aspect of the administration of all surveys. Since respondents are giving the information to the researcher in good faith, we should respect this in every way possible. There are several simple things which can be done to assist this:

- Mark the questionnaires with "In Confidence" or something similar.
- Train interviewers and all other staff to "not only be confidential, but be seen to be confidential". This involves things like not taking anyone else in the car in a personal interview, covering the questionnaires in the back seat so that passers-by cannot see them, and so on.
- All respect should be given to confidentiality in the office. Data enterers, coders and administrative staff should not discuss individual respondents in the office or at home.
- Questionnaire forms should be shredded at some time after the survey data has been analysed.
Several more substantive things may also be done (and, indeed are mandatory in some countries):

- Addresses may be removed from the computer files subsequent to coding these to zones or x-y co-ordinates.
- Specific locked rooms (for which only certain persons have permission to enter) may be designated within the survey office for all completed questionnaire forms.

### 7.1.7 Response Rates

Every survey has a response rate associated with it, and it is imperative that these are reported correctly. The response rate measures the extent to which the sample responds to the survey instrument. The objective is to obtain a high response rate, such that there is a greater probability that the set of respondents more closely represents the chosen sample, which should in turn represent the target population. Response rates, in principle, are calculated in the following way. From the gross sample size is subtracted those members of the sample from whom a response could not possibly be obtained. These forms of sample loss (i.e. invalid households, such as vacant or demolished dwellings, invalid phone numbers) do not affect the quality of the sample, and are sometimes said to be quality neutral. The number left after subtracting the sample loss from the gross sample size is the net sample size. The number of total responses is then taken as a percentage of this net sample size. The following example is useful:

<table>
<thead>
<tr>
<th>Gross sample size</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample loss</td>
<td>minus 5</td>
</tr>
<tr>
<td>Net sample size</td>
<td>95</td>
</tr>
<tr>
<td>Responses</td>
<td>64</td>
</tr>
<tr>
<td><strong>Response rate</strong></td>
<td>$64/95 = 67.4%$</td>
</tr>
</tbody>
</table>

While the above method of calculating response rates is simple in concept, it is not quite so simple when being used for a particular type of survey. Importantly, however, consistent methods of calculating response rate must be used across all types of survey, especially when comparisons are being made between survey types to decide which type of survey method to select.

### 7.2 ADMINISTRATION OF SELF-COMPLETION SURVEYS

While it is often thought that organising a personal interview survey is more complex than organising a mail-out/mail-back survey, self-completion
questionnaire surveys have very special administrative needs - particularly with regard to quality control. As noted in Chapter 2, a major problem with self-completion surveys is that very often the response rate is quite low, and therefore the opportunity for non-response bias to occur is quite high. The greatest contribution to high quality in self-completion surveys is therefore to raise the response rate as much as possible, while also collecting information about non-respondents in order to account for the residual effects of non-response.

Several factors are important determinants of the response rate for self-completion surveys and, as such, are important components of the administration of self-completion surveys. They are outlined in the following sections.

7.2.1 The Use of a Reminder Regime

Reminder letters are undoubtedly the most effective way of increasing the response rate. A general procedure for a mail-out/mail-back self-completion survey, including reminder letters, is based on the well-tested KONTIV method, originating in Germany (Brög, Fallast, et al., 1985). The method described here has been widely used (for example in the U.S. and Europe (Brög, Meyburg et al., 1985), and in Australia (Richardson and Ampt 1995)) for a household travel survey where all people in the household are asked to report travel for a specified travel day. The same principles could, however, be used for any self-completion survey.

(a) Initial contact. This stage is to introduce the respondents to the fact that they have been selected to participate in the survey and to legitimise it in some way. This is done with an introductory letter and informational brochure which is sent just over one week prior to the Travel Day allocated to the household (each household is asked to provide complete travel and activity data for one pre-specified Travel Day).

(b) First mailing. The first mailing includes the following items:

- A follow-up covering letter
- A household and person form
- 6 trip forms (to cover the maximum expected number of persons in the household).
- A trip form with a pre-printed completed example.
- A postage-paid return envelope.

This mailing is sent in an envelope with a postage stamp to make the letter seem more personal. The letters are sent so that they arrive 2 working days prior to the Travel Day.
Administration of the Survey

(c) **First reminder.** This takes the form of a post-card either to thank respondents who have already returned their forms or to remind respondents to return the questionnaire and to allocate them a new travel date (one week after the initial one) in case the forms have not yet been filled in.

(d) **Second reminder.** The second reminder is a letter sent in an ordinary business shaped envelope, again signed by the Survey Director. Once again, a new travel date is suggested for those people who have not yet filled in the forms.

(e) **Third reminder.** This reminder contains all the items sent in the first mailing with the addition of a cover letter from the Survey Director stressing the importance of cooperation by respondents in returning the forms. Again, a new travel date is proposed.

(f) **Fourth reminder.** For this (final) reminder a postcard is again used - but in a different colour. A new travel date is again proposed.

Notice that each element of the reminder regime is either a different shape (or colour). This is designed to discourage people from discarding (without reading) "yet another letter from these survey people".

As a rough method of estimating returns, it appears that about the same percentage of persons sent questionnaires respond to each mailing. For example, if 50% respond to the first mailing then 50% of the remainder will respond to the second mailing (i.e. 50% of 50% = 25%), while 50% of the remainder (i.e. 50% of 25% = 12.5%) will reply to the third mailing etc. Therefore while the use of reminder letters will increase the response rate, the initial response is critical in determining what the final response might be. It is therefore important to consider all factors which influence the initial response rate.

7.2.2 Validation Methods

In addition to the postal reminders, a number of other techniques can be used to improve response rates and the quality of the reported data (Ampt and Richardson, 1994; Ampt, 1993), particularly in self-completion surveys.

7.2.2.1 **Phone Interviews**

When the data from the returned forms is initially entered into the data base, queries or apparent mistakes are "tagged" by the data enterers. These are then followed up by phone interviewers who telephone these households in order to clarify any points of uncertainty. The phone numbers are provided by the respondents in response to a question on the survey form (in Australia, about 85% of respondents provide their phone numbers), and approximately 60% of all
responding households are phoned. During phone interviews, there is a check made of which person in the household completed each travel form in order to gain a measure of proxy reporting.

7.2.2.2 Validation Interviews

A sample of responding households can be selected for a personal interview to check on the quality and completeness of the data provided in the self-completion phase of the survey. Each household member is asked to go through the information provided for their travel day. A variety of techniques have been used for this interview.

One method is to carry out a full personal interview (using the original self-completed form as a memory jogger). In this way data on all travel is verified personally. Since respondents are also asked who filled in the original trip form, this is of particular value for measuring the effects of proxy reporting. The personal interview also means that the method is directly comparable with data obtained in the non-response interviews.

In many cases a graphical summary of the travel and out-of-home activities is used in these validation interviews. The representation is based on the time line concept (Jones 1977) with a line for each of home, travel and out-of-home activities (Figure 7.1).

![Graphical Representation Used in Validation Interviews](image)

**Figure 7.1** Graphical Representation Used in Validation Interviews

This was developed to assist interviewers and respondents to view the travel day at a glance. For example, in the above example it would be easy to check whether the respondent left work for lunch by simply asking "Did you stay in the same place between 9 am and 5 pm?".

The main purpose of these interviews was to obtain information on the under-reporting of trips in the self-completion phase of the survey and thereby to be able to calculate non-reporting weights (Section 9.3).
7.2.2.3 Non-Response Interview Surveys

Finally, a sample of non-responding households can be selected for a personal interview to check on the reasons for their non-response. As shown in Table 7.1, in about half the cases in the Brisbane study area in South East Queensland, the household agreed to complete a travel survey when contacted by the interviewer, and this information is used later in the calculation of non-response weights.

Table 7.1  Response Behaviour of Non-Response Validation Households

<table>
<thead>
<tr>
<th>Final Response Type</th>
<th>No. of Households</th>
<th>% of Households</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid Response</td>
<td>107</td>
<td>45%</td>
</tr>
<tr>
<td><strong>Sample Loss</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No such address</td>
<td>2</td>
<td>1%</td>
</tr>
<tr>
<td>Vacant</td>
<td>29</td>
<td>12%</td>
</tr>
<tr>
<td>Other sample loss</td>
<td>38</td>
<td>16%</td>
</tr>
<tr>
<td><strong>Other Loss</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-contact (after 5 visits)</td>
<td>16</td>
<td>7%</td>
</tr>
<tr>
<td>Refusal</td>
<td>46</td>
<td>19%</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>238</td>
<td>100%</td>
</tr>
</tbody>
</table>

(Source: Richardson and Ampt, 1993a)

These non-response interviews have proven to be especially valuable in identifying those households which contained stubborn non-respondents, those who were merely forgetful, and those households which did not actually exist (i.e. sample loss) - again important pieces of information for the weighting process.

Using the above survey design and quality control procedures, average response rates of 65%-75% have been achieved.

7.2.3  Sponsorship of Survey

Obtaining official sponsorship for the survey from a well-known and respected group or individuals is likely to markedly increase the response rate, and therefore needs to get particular attention in self-completion survey design. Non-controversial government authorities, research institutions, universities and public figures are useful sponsors whose name, or letterhead paper, can be used in the covering letter. As noted earlier, if there is any doubt about the impact of these sponsors, it is important to check the effect of these sponsors during a pilot survey.

An interesting example occurred in New Zealand where (at the time of the survey) the Ministry of Transport was not only a policy and planning body, but also the policing authority for all traffic offences. Although it seemed as if the prestige of the Ministry was a positive aspect of the survey design, people who received a letter in the mail with a potential offence notice, did not react...
positively to the survey - and the sponsoring authority was changed after the pilot survey!

7.2.4 Consideration of Respondents

In general, self-completion surveys have most success where the population under study is literate and concerned and/or interested with the subject under study. For surveys of the general public, there is evidence to suggest that non-response is highest among the lower socio-economic groups. This reinforces the need for special measures to be introduced to ensure participation by all groups if this is required by the objectives of the survey and to carry out follow-up surveys to give information on non-respondents.

7.2.5 Use of Incentives

A special case of the consideration of respondents is the use of incentives. Opinions vary as to whether incentives actually help or retard response rates. Intuitively, the use of a small payment or gift would seem appropriate. There is some evidence to suggest, however, (e.g. Chipman, et al., 1992 and Bonsall and McKimm, 1993) that this is not always the case. It is likely to be safest to use incentives such as special postage (see Section 7.2.10 below). We would argue that one of the best incentives is a survey design where the purpose and layout is easily understood and where it is easy to contact someone if questions need to be asked.

7.2.6 Covering Letter

Since there is no opportunity to personally introduce and explain the questionnaire to the respondent, the use of a covering letter is essential in all self-completion surveys to increase response rate and the understanding of the questions. The letter need not be overly personalised but should be clear, friendly and not officious. Handwritten notes urging reply in reminder covering letters have been found to be effective. In addition, it is sometimes useful to enclose a brochure explaining the survey in a more informal manner than is generally possible in a letter.

7.2.7 Use of Comments Section

The use of a comments section at the end of the questionnaire can often improve response rates by giving an opportunity to respondents to air their own views on the subject, independent of the formal questions which may have been asked in the main part of the questionnaire. These comments may or may not be coded and used in the analysis.
7.2.8 Provision of a Phone-in Service

One of the key elements of the administration of a self-completion survey is the provision of a phone-in service for respondents. Since there is often no other personal contact with survey investigators, it is imperative that this service is available (free, if possible) for as many hours of the day as practicable. Given that most people are away from home during the day, and are likely to be completing the forms at nights and on weekends, it is not at all practical to limit the hours of operation to work times. This may mean having someone on duty during non-work times, or switching the phone through to the private homes of the survey administrators. As noted earlier (Section 7.1), all people who answer the phone need to have comprehensive training in the survey objectives as well as the questionnaire content.

7.2.9 Preparation of Questionnaires for Mailing

For small self-completion surveys, this may be a fairly minor task, but for large-scale household surveys, with complicated reminder regimes (e.g. Section 7.2.1), the preparation of questionnaires for mailing can be a major operation. In this case, piloting of times taken as well as space used, and ergonomic methods of preparation become critical. As an example, for a survey of 20,000 households (gross) using the 6 stage method of reminders it will take about 100 person days to prepare the questionnaires for mailing - and this does not include all reminders, since in some cases it is necessary to know which household needs reminding before customised questionnaires can be prepared!

7.2.10 Type of Postage Used

The type of postage used on outward and return letters can have an effect on response rates. On outward letters, the use of first class mail, priority paid post and registered letter (in that order) can increase response rates by demonstrating the apparent importance of the survey to the respondent. For return letters, three types of postage in increasing order of effectiveness are commonly used:

(i) **No return postage paid.** This usually results in very low response rates and is not recommended. Those people who do reply are likely to be very much more committed to the survey than those who do not.

(ii) **Reply-paid post permit.** This method involves payment of postage on return of the letter to the investigator. A licence number must usually be obtained from the Post Office for this method. While reducing postage costs, because only returned letters are paid for, it does not necessarily increase response rate to an optimal level because the respondent feels under no obligation not to waste the investigator’s money. There is some evidence to suggest that it may
be an effective method when the size of the letter (and hence postage) is quite large. It seems that the temptation to remove stamps (method (iii) below is greater for these larger letters, making the reply-paid method equally effective (while somewhat cheaper).

(iii) **Stamped self-addressed return envelope.** This involves placing a normal stamp on the return envelope before the questionnaire is sent out. On receiving this envelope the respondent has three options:

1. Throw it away with the questionnaire;
2. Throw the questionnaire away but steam the stamp off the envelope; or
3. Return the completed questionnaire in the envelope.

Option 1 leaves the respondent with the feeling of having wasted a resource - a perfectly good stamp. Option 2 leaves some respondents with a guilt feeling of having gone to a lot of trouble to get a stamp. The only way to overcome both of these guilt feelings is to adopt option 3. While this method appears to obtain better response rates when the questionnaire is small, the cost of individually affixing stamps to each envelope needs to be compared with simply getting the reply-paid envelopes printed. This process is being monitored in detail in the current Melbourne survey (Richardson and Ampt, 1993b).

### 7.2.11 Response Rates

As stated before, the objective in postal questionnaires is to increase the response rate. Every effort should be made to achieve this objective. Some valuable references on various aspects of self-completion surveys and response rates are Kanuk and Berenson (1975), Galin (1975) and Richardson and Ampt (1994).

Despite our best attempts to increase survey response rates, there will always be a certain proportion of people who do not respond. If these non-respondents are similar to the respondents, then there is no great cause for alarm since the non-response bias should be minimal. However if, as is usually the case, the non-respondents are atypical of the general population, then steps should be taken to account for the biasing effect of this non-response. It should be emphasised that the similarity between respondents and non-respondents should be assessed not simply on the basis of socio-economic characteristics but on the basis of the survey parameters of particular interest, e.g. in a travel survey, on the basis of trip rates. In a transport context, useful methods of accounting for non-response bias have been proposed by Brög and Meyburg (1981) and Stopher and Sheskin (1982), and are discussed further in Chapter 9.
7.3 ADMINISTRATION OF PERSONAL INTERVIEW SURVEYS

A vital component of the administration of personal interview surveys is the maintenance of quality control by means of various procedures. There are six major items to check to ensure high quality data.

7.3.1 Use of a Robust Interview Regime

A personal interview survey to gain information for transport planning must be one which ensures a high response rate and which is as easy as possible for interviewer to carry out and for the respondent to take part in. The following is a summary of a methodology for a household travel survey which has had widespread use in several countries since its inception in 1981. It is described in detail elsewhere (Ampt 1981).

It uses a verbal activity recall framework (Chapter 5.3) and is not based on recall of travel, since respondents are notified in advance of the day about which they will be required to give information.

7.3.1.1 Pre-Contact Letter

This letter is sent from the client informing the respondent of the survey and legitimising it, in the same way as it is done for self-completion methods. To add to its authority, the name of the interviewer can also be written on each letter, a measure which has been shown to be particularly conducive to high response rates. This letter is sent so that it arrives several days prior to the first attempt to carry out a pre-contact interview (below).

7.3.1.2 Pre-Contact Interview

At this initial contact, the interviewer gains information on the household including the such things as structure type, the number of household members, the number and type of vehicles and bicycles in the household. Since the information is entirely factual, a very structured method of interviewing can be used.

In addition to asking the above information, interviewers also leave Memory Joggers for each person in the household. These Joggers are personalised, diary-like notepads on which respondents keep track of all travel at the level of detail described to them by the interviewer. Finally, appointments are made to speak with each household member over a certain age (often 9 years) personally at an appointment after the travel day/s. Data for children younger than 9 years is usually accepted by proxy.

This pre-contact interview is best attempted about 3 days before the Travel Day (the day about which the main survey is conducted) to ensure that up to 4 calls...
can be attempted. This is important because each household must have an opportunity of being found at home. (Section 7.3.6). Surveys in which only 1-2 calls pre-contact calls are attempted suffer serious under-reporting of trips because the people who are most easily found at home are clearly not travelling - at least at that time!

7.3.1.3 Main Interview

The final stage of the personal interview takes place when the interviewer returns to carry out interviews with each household member after the Travel Day. Again it is best to use a structured form if factual travel data is being asked. This means that the Memory Joggers are used by respondents only for the purpose their name implies and it is not necessary for the interviewer to collect them.

In order to eliminate sample bias due to households "choosing" their own travel day, each household is allocated a specific travel date which should not be varied.

7.3.2 Training of Interviewers

This has already been mentioned in Section 7.1 but cannot be stressed enough. The two-fold aim of the interviewer training sessions is to give the interviewers a thorough understanding of the survey objectives and processes, and to teach the specific skills of interviewing related to the survey in question.

7.3.3 Checking of Interviews

Checking whether all interviews have, in fact, been carried out requires some form of follow-up contact with the respondent/s. This can be performed in one of several ways.

First, a sample of respondents could be followed up with a telephone validation check. A few short questions can usually ascertain whether or not the interview has been carried out. It is usually useful to ask some questions of content in addition to whether or not the survey actually took place to ensure that the interview was done as required. The phone check can also be used to ensure that no proxy interviews were carried out, if this was an important objective of the survey (which is usually the case in travel surveys).

The phone validation method should never be used alone since it would be possible for the "astute" interviewer wanting to avoid detection to omit the phone number from any questionnaires, claiming perhaps that respondents did not want to give it. For this reason, a sample of respondents who do not have the phone can be sent postcards containing a small questionnaire. These postcards could ask similar questions to the phone interviews - whether they were interviewed, when it took place, how long it took and what were their
impressions of the interview. The main problem with postcards is that respondents may not return them. In particular, those respondents who were not interviewed may think that they have received them by mistake and therefore not bother with them. Ideally, a small sample of non-respondents to these postcards should then be visited. As an alternative method, all respondent/s to be validated could be sent postcards, although combination with the phone validation is method is usually the most efficient.

A third means of follow-up is for call-backs to be made to selected households by a field supervisor to check whether the interview has been made. Such a process is however quite a costly method.

Three points need to be made with respect to the use of follow-up contacts.

(a) In general, if interviewers are trained to understand the importance of their task and of the study in general, and are treated as professional members of the study team, there is usually very little difficulty with interviews not being carried out. Quality control measures should, of course, still be carried out, but interviewers should be informed of the results. Often, for example, words of praise are heard of the interviewer's performance, and in any case, general information about the quality control process helps to encourage a high standard of performance from team members.

(b) Furthermore, it is usually the case that the mere statement that random call-backs will be made is enough to dissuade most interviewers from attempting falsification. Whether the follow-up procedures are fully effective in detecting non-contacts may be of secondary importance.

(c) In some cases, validation can be limited to cases where it is suspected that an interviewer may be guilty of falsification. Such suspicion arise after a routine statistical comparison of each interviewer's recorded responses with those obtained from the rest of the interviewers. Just as most people cannot deliberately pick a random sample from a list of numbers, so most interviewers are unable to falsify data such that they produce the correct distribution with respect to each of the survey parameters. Even more difficult is the task of correctly assessing the degrees of multiple correlation between all the parameters. Thus in their attempt to create realistic false data, interviewers give themselves away. Routine statistical assessment of each interviewer's recorded responses should therefore be a standard component of survey quality control. Note that this procedure would be effective only if each interviewer has
enough completed interviews to produce statistically reliable comparisons.

While the comparisons described above are invaluable to assist both the survey administrators and the interviewers, relying on this method to identify problem interviewers is not ideal; a strict regime of quality control is always recommended.

7.3.4 Satisfactory Response Rate

In personal interviews, the response rate (that is, the ratio of completed interviews to assigned interviews, minus any sample loss) should rarely be less than 80%. If the response rate for any interviewer is consistently lower than that obtained by other interviewers in similar circumstances then this is an indication that the interviewer may either be in need of retraining in the art of making initial contact, or else is incapable of making such contacts and should be dismissed.

The same test can be applied to individual questions within the questionnaire. If too many "don't knows" or refusals are being encountered then it is a sign that either the question is being asked incorrectly, that not enough probing is being used, or that the interviewer's motivation or attitudes are inappropriate. For example, it has been shown that interviewers who, themselves, do not understand the importance of asking income questions have a higher refusal rate on income questions. On the other hand, those who understand its significance rarely encounter a refusal for this question. Once again, retraining or detailed explanations of individual questions might be the appropriate remedy for recurring problems.

7.3.5 Correct Asking of Questions

Poor asking of questions is perhaps the most difficult point to check and remedy by follow-up procedures. The most effective steps in preventing it occur prior to the actual execution of the survey.

(a) At the design, planning and pilot stage the testing should include a check of the "comfort" which interviewers have with the flow of the question. When this is well controlled, there is usually little difficulty in the field.

(b) During training, a great deal of stress placed on the importance of verbatim question asking, plus some hours of practice, also serve to minimise the problem.

Once again, an indication of problems during the survey itself may be gleaned from statistical analysis of the data or from examination of the completed questionnaires. The task of definitely identifying and remedying the situation is,
however, not straightforward. Several methods are available, each of which has distinct deficiencies. Re-interviewing of the respondent by a high-level interviewer/ supervisor may pick up gross errors. However, there will always be differences between any two interviewers (even if they are both of a high standard) and it must also be assumed that the respondent stays constant for both interviews. This assumption may not be true because of lapses of memory by the respondent and possible genuine changes of attitude.

The other methods of validation all involve observation of the interviewer while an interview is in progress. This interview may be a real one in the field or a test interview in the office, while the method of observation may be by a supervisor or by using a tape recorder. While these methods may pick up errors which the interviewer is genuinely unaware of, it is difficult to detect the interviewer who is just getting lazy or careless. The mere fact that the interview is being observed will generally result in these tendencies disappearing (for a while). However, generally, a series of well-announced field supervisions serves to keep the standards high.

7.3.6 Number of Interviewer Call-Backs

An important decision in the administration of personal interviews in the number of call-backs which interviewers are required to make before determining that the household or person is not able to be contacted - often termed "a non-contact".

Theoretically, given time, the interviewer to a household would eventually be able to reach all members of the household. This is generally not cost-efficient.

On the other hand, there is ample evidence to suggest that there is significant difference in travel behaviour between those people who are interviewed on the first visit and those interviewed at second and subsequent attempts (e.g. Brög and Ampt, 1983), and it is therefore advisable to make up to 3-4 attempts at different times of the day and week. Interviewers who understand the principles behind high response rates have been shown to be ingenious in finding elusive respondents (particularly shift-workers) at home.

7.4 Administration of Telephone Interview Surveys

If telephone interview surveys are being conducted, four specific areas of administration need to be considered. They are sampling, dealing with non-response, maximising data quality, and several operational aspects.

7.4.1 Sampling

As the use of telephone interviewing has grown in the last 25 years, telephone sampling methods have increased in diversity. Early telephone sample designs
used telephone directories as the sampling frame because they were readily available and they were thought to contain a “representative” selection of the telephone household population. Given that telephone interviewing is only dealing with those people who have phones, and that problems associated with using this sampling frame has been discussed in Section 3.2.4, the problem of non-phone ownership will not be considered here.

Hence, when the proportion of unlisted (ex-directory) phone numbers is small, the representativeness argument is fairly persuasive, particularly considering the trade-offs in convenience and cost over other methods. However, in many countries the frequency of unlisted numbers has increased to levels that raise concern about the accuracy based on directory samples. Three other approaches are therefore possible in most countries:

(a) **Random digit dialling.** This provides coverage of both listed and unlisted telephone households by generating telephone numbers at random from the frame of all possible telephone numbers. In some areas, local telephone agencies can assist in providing ranges and constraints to these numbers, which greatly increases the efficiency of the task (e.g. Stopher 1985a).

(b) **List-assisted designs.** These designs use information in telephone directories to generate telephone number samples that include both listed and unlisted telephone households.

(c) **Multiple frame sampling.** This method basically combines the directory and random digit dialling sampling frames into a single design.

An excellent review of probability sampling methods for telephone household surveys in the United States is given in Lepkowski (1988). There is, in fact, a whole body of literature dealing with sample designs, comparing efficiency, cost factors and quality of the samples, as well as the relationship of the sample design to non-response rate (e.g. Groves, et al., 1988) which is worth pursuing if telephone sampling is to be an important part of the transport survey process.

### 7.4.2 Dealing with Non-Response

As noted in Section 3.2.4, telephone survey non-response has become increasingly troublesome and threatens to eliminate the unique property of sample surveys - statistical inference to a known population. This section focuses on non-response in cold household telephone surveys. It will omit discussion of telephone surveys of non-household populations such as businesses which were briefly mentioned in Chapter 3 as generally having higher response rates, in any case.
7.4.2.1 Sample Loss

It is important at the outset to realise that phone calls to households result in sample loss - non-response which does not affect the quality of the sample. There are a number of reasons for this:

- a person may answer the phone and report that the number is a business,
- a message may report that the phone number does not exist, or has been changed,
- in random digit dialling in some countries, it is possible to gain a ringing tone from a non-valid number.

These cases can be treated as "sample loss" since they do not affect the quality of the sample being surveyed.

7.4.2.2 Factors Affecting Non-Response

Part of the design of a telephone survey involves understanding factors which may affect non-response. They are, of course, many and varied, and as we mentioned earlier, the important ones are those which affect the values of the parameters being measured (usually travel behaviour of some type). Included among these factors are some which relate to respondents ((a) and (b)) and some which relate to the characteristics of the interviewers ((c) to (e)).

(a) Education. There is some evidence to suggest that there is higher non-response rate among lower education groups (Cannell et al., 1987).

(b) Urban-rural differences. While there is evidence that there is a higher non-response in large urban areas than in other areas for personal interviews, this seems to be diminished in phone surveys where urban-rural differences are minimal.

(c) Interviewer voice quality may affect levels of cooperation. Overall, interviewers rated as speaking loudly, with standard local pronunciation, and perceived as sounding competent and confident, have lower refusal rates than those with the opposite patterns. With regard to intonation patterns, interviewers using a falling tone on key words early in the introduction have lower refusal rates than those using a rising tone (Oksenberg and Cannell, 1988).

(d) Interviewer gender and experience. In general females seem to gain higher response rates than males, and more experience is also associated with better response rates.
(e) **Interview length.** Surprisingly, there is very little evidence to support the strong intuitive belief that shorter interviews are always more successful. In the U.S., researchers generally use shorter interviews for telephone surveys than for personal interviews, although Swedish researchers encounter less resistance to long telephone interviews. In the U.K., there is evidence of a 5% lower response rate to a 40-minute telephone interview as compared to a 20-minute interview (Collins et al., 1988).

### 7.4.2.3 Methods of Dealing with Non-Response

Several methods can be used to deal with some problems of refusals in telephone interviews.

(a) **Timing of calls.** When phone calls interrupt specific events like meals they can lead to rapid refusals. For this reason, method (d) below is often a useful adjunct in the survey process.

(b) There is some evidence that the strategy of **making appointments** to defer the initial resistance to the phone call can reduce initial non-response.

(c) **Advance letters** (similar to the letters advocated for personal interview and self-completion surveys) sent to respondents to explain the survey have been shown to reduce non-response since it reduces the element of surprise (Clarke, et al., 1987).

(d) **Attempts by a second interviewer** to convert those who initially refuse to take part, and

(e) **The selection and training of interviewers** to take into account the types of differences mentioned above.

### 7.4.3 Improving Data Quality

There are several things which may affect data quality in phone interviews which are different from those found in other modes of data collection.

The first relates to the way in which people respond to sensitive or threatening questions. There are two lines of argument about this effect. On the one hand, it is argued that because respondents cannot see the facial cues which are evident in face-to-face personal interviews, they experience some anxiety that they might not be giving the right answers, and may be less likely to respond. However, on some topics, this very fact may encourage people to give this type of information more freely. Thus, for items or topics particularly susceptible to such effects, the telephone may hold certain advantages of personal interviews. Pilot testing will obviously help to understand each particular case better.
Administration of the Survey

Furthermore, it has been shown that there is a tendency to obtain truncated or shorter responses to open questions administered over the phone, than to the same questions asked in person. Care needs to be taken to consider this during the question design stage.

Finally, there is the issue of response scales where, in a personal interview, people would normally be presented with a show card. One way to deal with this is to read out the scale points, although another method is to ask the question in two stages. For example a five point "satisfaction-dissatisfaction" scale would first be presented with three response options (satisfied, dissatisfied and neither satisfied nor dissatisfied). "Satisfied would then be asked: "Is that very satisfied or just satisfied, and similarly for dissatisfied.

7.4.4 Other Aspects of Administration

One of the key components of telephone interviewing is the use of a computer-assisted telephone interviewing (CATI) system to assist interviewers and their supervisors in performing the basic data collection tasks of the telephone surveys. In typical applications the interviewer is seated at a computer terminal, wearing a telephone headset. As survey questions are displayed on the screen, the interviewer reads them to the respondent and enters responses into the keyboard. In most systems, question wording and sequencing between items is computer controlled based on prior entries, and answers can be checked for logic and range errors as soon as they are input - meaning that inconsistencies can be checked on the spot. For further information on these systems refer to Baker and Lefes (1988).

The questionnaire design using this system requires careful management. For example, it needs to cater for the need to go back to verify information and to correct interviewer and respondent error. Without extensive testing, this can cause significant problems.

Finally, many supervisors complain that with a CATI survey that uses a sample-management schedule, they lose the "feel" of the progress of the survey. Instead of piles of cards to sort, they must deal with more abstract queues and reports. It is important that these people work together with the programmers to be sure that their decisions are implemented correctly and have the desired effects. As system are becoming more sophisticated, there are also many cases where computer graphics can assist in the alleviation of this problem.

7.5 ADMINISTRATION OF INTERCEPT SURVEYS

The administration of intercept surveys - those which are carried out with people that have made a choice of a certain mode of transport - also requires special care if high quality data is to be obtained. As mentioned earlier, these surveys can
either be self-completion or personal interview. The discussion in this section is relevant to both of these types.

7.5.1 Sampling

Although the theory of sampling has been discussed extensively in Chapter 4, some of the specific issues relating to intercept surveys are worth reiterating at this point.

7.5.1.1 On-Board Surveys

For surveys on-board a mode of transport (such as a bus, train or plane), there are two specific issues. The first relates to the choice of the vehicle itself. Which buses/trains/planes are to be selected from the many which operate over a given time period if the intercept survey is to be representative of that time? One option is to survey on all vehicles. While this may be practical if the study period is brief, it is generally both more time consuming and more costly than necessary for most surveys. Unfortunately, the large quantity of data obtained is often assumed to make up for any biases caused by a lower than desired response. Furthermore it is also often assumed that the survey on each vehicle in the sample must cover the entire operating day, or as much of it as is important for the collection. This also usually constitutes an unnecessarily large sample size and survey cost and, in general, these methods are not supported by a thorough understanding of sampling issues and procedures.

An excellent method of selecting vehicles is described in Stopher (1985b). Two levels of understanding are necessary. The first is a clear recognition that the sampling unit is the passenger not the vehicle; and the second is to know and understand the implication of the operational scheme of the transport system being surveyed. Knowing these two characteristics of the population makes it possible to develop a system in which specific routes of a transit service are selected using a system of stratification of routes or runs. A multi-stage sampling technique is then used to select specific bus/train/tram runs or services. Using this method, the process of expansion to the entire system is not as straightforward as it would be if a simple random sample of vehicles had been selected, but knowing the way in which the sample is selected is not really very difficult. However, it can readily be done and is described in more detail in Kaiser Transit Group (1982).

7.5.1.2 Roadside Surveys

Another specific sampling problem occurs with surveys in which vehicles are stopped at the side of roads or at signalised intersections. This is much more difficult than the case of on-board surveys since the vehicles in question in this case do not behave according to a set timetable.
Theoretically either all vehicles (or a random sample of vehicles) should be interviewed, because stratification of any type is usually not possible while standing at the roadside. Since both of these can prove difficult it is important that three controls are always practiced.

(a) There should be no systematic omission of any specific vehicles. This means that if vehicles are being stopped at the roadside, all vehicles in a given period should be stopped, and then a group missed, rather than allowing the traffic control person to select what they may consider to be "friendly" vehicles.

(b) There should be classification counts done in parallel with the surveys to determine whether the surveyed vehicles are representative of the entire fleet of vehicles passing the point during the survey period.

(c) As in all surveys, any refusals should be monitored.

Furthermore, as part of the survey sampling design, it needs to be clear whether the sampling unit is the driver or all persons in the vehicle, and the survey method (including the instrument design) needs to take this into account. This depends to a large extent on the objectives of the survey. Route choice, for example, may be largely decided by the driver, but other decisions about the travel (e.g. destination and time of day) may be a decision of all the passengers, not just the driver.

7.5.1.3 At Activity Points

Intercept surveys which take place at activity points such as shopping centres or airports are probably the most difficult from the point of view of sampling. In almost all cases the sampling unit here is the individual, and while finding people to interview is rarely a problem, keeping control of the population from which this sample is selected is much more difficult.

Because of this difficulty there has been very little research done on ways to improve the sampling method, and many reports simple make do with an over-representation of the people who are easy to find (like meeters and greeters at an airport) and under-representation of those who are difficult (like business people). This is not serious when the total population of each sub-group is known, since the data can be expanded appropriately. However, these cases are not common.

A method which can sometimes be useful to gain this information is to find a point at the activity centre where all persons need to pass before entering or leaving. While it may not be possible to carry out the interviews at the level of detail required at these points, it is often possible to ask one or two screening
questions to gain information on the population of those people in the centre, thereby getting a base point for expansion.

Although these methods seem time-consuming, they need to be considered if the data collected at these centres is not to be disregarded for its lack of representativeness.

7.5.2 Training of Staff

The staff training for intercept surveys is as important to the success of the survey as for all other survey methods. It is, however, frequently relegated to a quick gathering of personnel a half an hour prior to commencement of the survey. This results in poor performance by the staff and frequent "no-shows" with all their attendant problems.

Training sessions are used, as in other surveys, to give the staff information on the background to the survey. A knowledge of sampling methods is also often very important. In surveys at public transport hubs, unreliability of timetables often means that the staff will have to take back-up measures under some circumstances and understanding how and why a sample is chosen can be crucial to their making the correct decision.

Furthermore, in cases where a self-completion questionnaire is being distributed, particularly where this is happening on a moving vehicle, it is important that the training sessions are used actually practice the distribution of forms. Get all the "passengers" to file past an interviewer at the speed people board a bus and check that the interviewer can keep up! The art of giving out forms (and pencils) to all people boarding a bus in peak hour - with the appropriate smile of encouragement - is not for the faint hearted, but can be easily mastered with a little practice!

7.5.3 Other Administrative Details

There are several other details which need to be considered in the administration of intercept surveys.

(a) The surveyors need to have writing implements that work under the weather conditions being experienced at the intercept site (felt-tipped pens smudge in the rain, and biros do not work on wet paper or in temperatures less than about 3ºC).

(b) If you are distributing pencils to respondents on a moving vehicle, consider square rather than round ones to avoid their rolling when dropped.
(c) If weather will be a problem, make sure there is protection for the surveyors.

(d) In all cases the surveyors will need somewhere to put their personal belongings during their work periods.

(e) Finally, do not forget to record the non-response. Without these details, expansion to the population will not be meaningful.

7.6 ADMINISTRATION OF IN-DEPTH INTERVIEW SURVEYS

In-depth interview surveys, while always incorporating personal interview surveys of one form or another, require special care from an administrative perspective. A comprehensive description of the conduct of these surveys, not exclusively in the field of transport, is given by Gordon and Langmaid (1988).

Figure 7.2 shows the various stages involved in a typical in-depth interview procedure and highlights the complexity of the process. Correspondingly, the administrative procedures need to include the following components.

7.6.1 Organisation of "Pre-Interviews"

This phase begins with the selection of the sample of people to be interviewed. Depending on the survey objectives, and bearing in mind that one of the key uses of in-depth interviews is exploratory (Chapter 3), the sample is often not random and consists of people with known travel characteristics (who are likely to be most affected by changes to the status quo).

The pre-interview phases can often include the gathering of information on current travel behaviour which requires the distribution of self-completion diaries, or the organisation of personal interviews to collect travel data if needed. Given the length of the main in-depth interview, it is generally advisable to keep the pre-interviews completely separate from the main survey.

7.6.2 Preparation for the Main Interview

This will involve preparation of any interview aids such as boards in the example or the preparation of any other aids which will be needed in the interview. If a simulation approach is used the data from the pre-interviews can, in some cases, already be set up ready for immediate use. Other in-depth methods may include the use of physical models to help understand certain changes which are likely to occur (e.g. an architectural model of a High Street before and after the introduction of clearways). All of these items need to be prepared in advance, and often in a new configuration for each interview. Another important part of the preparation for this interview is the availability of highly sensitive (yet portable) tape recorders.
### 7.6.3 Training

In comparison with structured personal interviews, the training of the interviewers for in-depth surveys needs to be much more extensive. Frequently a detailed understanding of the political context of the topic under exploration, as well as of the survey components is a necessary prerequisite for an in-depth interviewer. For this reason, and also given the relatively small sample sizes usually used in in-depth studies, it is common that the investigator is the only interviewer on a given project.

### 7.6.4 Preparation of Questionnaires/Data

After the interview, interviewers need to be given plenty of time to record any important details which emerged from the interview and which may not be
recorded on the tape. For example, if people's changes in behaviour are being recorded on board games it might be necessary to record any revisions which respondents in writing.

7.6.5 Time and Resources for Transcriptions

Both time and resources need to be set aside for the transcriptions of the tapes after the interview. As a rule of thumb, it takes between just over twice the actual interview time to make the transcription. If appropriate, this has been found to be effectively done by the interviewers themselves, since issues such as attributing voices to the correct people is more easily done by them than other people.

7.6.6 Report Writing

Report writing is a component of the in-depth interview which is critical to the quality of the survey. On the one hand, reports need to record sensitively the feelings of the qualitative information presented by respondents. On the other hand, there are many quantitative components of the method, and these should not be disregarded. For example, the changes in behaviour which may occur after the introduction of a new measure during the interview (e.g. reactions to a doubling of travel time) can often be quantified quite easily. The output of methods like the situational approach (Figure 3.3) give an excellent example of this type of quantification.

7.7 COMPUTERS AND THE CONDUCT OF TRAVEL SURVEYS

Whereas, as recently as 15 years ago, computers were not used at all extensively in the administration and conduct of surveys for transport planning, they are now used increasingly in all phases from survey design, management and data collection. An excellent description of the uses up until 1990 is given in Jones and Polak (1992). Computers are, of course, also used for data entry, but this is described in more detail in Chapter 8.

7.7.1 Survey Design and Management

The sampling frame for most surveys - particularly those of a larger scale - is almost always made available in computer-readable format. Whether the samples are from voting or utility lists, postal address lists or telephone records, it would be rare that they are not provided on disk or tape. One of the chief (and often frustrating) tasks of survey administrators is to ensure that the data is comprehensible.

Another important area for computer applications is survey management and control, which covers the allocation of work and monitoring of returned questionnaires. This process has been described in Section 7.1.
7.7.2 Data Collection

As Jones and Polak (1992) point out, many types of transport survey already make use of computers for rapid data capture and entry and their role is well-established. Among these are

(a) **Telephone interviews.** Computer assisted telephone interviewing (CATI) has already been discussed and is now the norm in many countries, although its application in transport is more limited than in other areas. The ability to combine the data collection and management functions is a specialty. Non-answering numbers can automatically be stored, quality control through listen-in procedures is very straightforward, sequencing can be automatic, and data entry can occur as the interview progresses.

(b) **On-board surveys.** In the U.K. and many other countries, simple on-board surveys to collect passenger profiles on bus and rail are often carried using light pens for data capture linked to a small portable computer. Hamilton (1984) gives a good example of using a computer-based bus passenger monitoring system in the north east of England.

(c) **Traffic surveys,** while not the specific subject of this book, often use hand-held computers. These can be used for such diverse surveys as traffic classifications or counts, turning movements, or parking surveys.

The use of computers in the home and for personal interviews in general, is a relatively recent phenomenon. In some fields, however, it has had remarkable results. For example, the Dutch Labour Force Survey has had a fully computerised household interviewing system since 1987 in which about 25,000 people are interviewed each month. Data is down-loaded each night via modems from the 400 laptop computers kept by the interviewers, and each month’s data is ready for use nine days after the end of each month (Keller and Metz, 1989).

The computer can be used in two main ways in travel surveys.

(a) It can be used as though it were an electronic questionnaire. In this type of survey the questions and interview style are kept the same as they would be in a conventional paper survey, with the computer screen being read by the interviewer rather than the questionnaires. This was the earliest approach and it was sometimes used to help design experiments in stated preference exercises (e.g. Ampt et al, 1987; Jones et al., 1989).

(b) Both the interviewer and the respondent can use the screen, enabling among other things, the use of graphics and video scenes. Examples
of this type of use can be found in Polak et al. (1989) and Polak and Axhausen (1990).

Both of methods can add quality to the data being collected because they eliminate any potential sequencing errors made by respondents or interviewers and because they also do away with the need for data entry, thereby saving cost and reducing one phase where errors can be introduced.

Another very specific use of computers in travel surveys for their specific ability to adapt the questions to people’s responses. For example, in stated preference surveys (Section 5.5.4) respondent’s answers to questions on facts about a specific trip (such as travel time, wait time, length or cost) can be used to develop hypothetical scenarios for future changes to the system which are based on people's current travel. For example, if people are currently travelling 15 minutes on a bus line and the operator is trying to check how people would respond to a more circuitous route (with more helpful drivers), the new travel times offered them would be based on the current 15 minutes (perhaps +10% or +20%). In the past, without computers, all respondents tended to get presented with the same absolute changes in travel time, meaning that they were often very unrealistic. Much more reliable data is possible with the aid of this facet of computer interviews.
8. Data Processing

Once the field survey has been conducted and the completed interviews or questionnaires begin to flow into the survey office, it is time to begin the often tedious task of editing, coding and analysing the results. Although the physical component of this task begins now, it should be realised that the planning phase should have been completed much earlier in the survey process. It is too late to start designing the coding and analysis procedures once the completed questionnaires begin to arrive. Rather, these tasks should have been largely finalised when the design of the questionnaire and the sample was being considered. Indeed, attention given to these tasks at those earlier stages will greatly facilitate the smooth completion of these tasks now.

The task of transforming completed questionnaires into useable results is composed of several discrete tasks including initial editing of questionnaires, coding, data entry, computer editing, data correction, analysis, interpretation of results and preparation of reports. In addition, to enable use of the data for secondary analysis at a later date, it will be necessary to arrange for satisfactory storage of the data. This chapter will concentrate on the coding and editing of data in preparation for analysis. Later chapters will examine the tasks involved in the latter stages of data analysis.
Chapter 8

8.1 INITIAL QUESTIONNAIRE EDITING

Before the completed questionnaire from a personal interview survey reaches the survey office, it should already have been subjected to two forms of editing; field editing and supervisor editing. Field editing is performed by the interviewer and is used to check on the completeness of responses, to expand on notes jotted on the questionnaire by the interviewer and to check for legibility. Such field editing should be done as soon as possible after completion of the interview, so that problems can be cleared up while the interview is still fresh in the interviewer's mind. In many cases, the data can be checked after the interviewer returns to their vehicle and before they drive to the next interview (assuming they are driving). In this way, if any information is found to be missing or unclear, the interviewer can return immediately to the household and obtain the missing data or clarify the uncertainty. To avoid introducing interviewer bias at this stage, it is essential that in supplying missing information the interviewer only enters responses which were given in the interview but which, in the heat of the moment, were not recorded. No attempt should be made to infer what the answers "should" have been.

Supervisor editing should be carried out as a quality control check, and to ensure that the completed questionnaire sent on to the survey office for coding is legible, complete and consistent. If the supervisor detects persistent errors by the same interviewer then it is possible for this error to be drawn to the attention of that interviewer before too many more interviews have been conducted. It may be necessary to request that the interviewer attend a retraining course if the error rate is high. On the other hand, if a similar error or inconsistency is detected over a number of interviewers, this may point to an inadequacy in the question wording or in the instructions for asking or recording answers to that question. This would call either for rewriting of the question and/or instructions, or for retraining of all interviewers with respect to this question.

In these days when most survey data analysis is performed by computer, it is usual to suspend editing activities at this stage after these two types of editing. More complete editing is performed after the data has been transferred to the computer, because the computer is far quicker and less error-prone in picking up errors and inconsistencies than an individual human editor could ever hope to be. For this reason, we now turn attention to the coding of the data.
8.2 CODING

Coding is the translation of data into labelled categories suitable for computer processing. In the overwhelming majority of cases, this means numerical labelling. It should be realised, however, that in many cases the labels used, whilst numerical, are purely nominal; they do not necessarily imply an underlying numerical relationship for the categories of the variable in question - for example it would be ridiculous to compute a mean value for marital status from the numerical codes assigned to each state. Similarly, the conventional coding of sex as male (1) and female (2) does not imply a value ordering of the sexes!

In many cases, however, the codes do possess more than a nominal property. Three other types of code are possible; ordinal, interval and ratio codes. Whilst nominal codes merely attach labels to data items, ordinal, interval and ratio codes carry more information about the data item.

In addition to naming the data items, ordinal codes rank the data items in some order and hence allow the mathematical operations of equality, inequality, greater than and less than to be applied to the codes. An example of an ordinal code in travel surveys is the use of a priority ranking of modes which can then be used in the "trip-linking" process. Thus, in addition to giving a numerical label to each mode, the codes also order the modes in the order of priority to be used in the linking process, e.g. train=1, bus=2, car=3, bicycle=4, walk=5.

Interval (or cardinal) codes go one step further in that, in addition to ranking the data items, they also impute measures of separation to the numerical differences in the code values. The mathematical operations of addition and subtraction are therefore possible with interval codes. An example of a interval code is the time of day attached to the start and end of trips. These time codes order the sequence of start and end times and can also be subtracted to obtain the travel time on the trip. Similarly, the year of birth of a respondent is a interval code, which when subtracted from the year in which the survey is performed gives the age of the respondent at the time of the survey.

The most general form of code is a ratio code which imputes meaning to the order, differences and ratios of code values. Thus all four mathematical operations (+, -, x, ÷) are possible on ratio code values. Many codes used in travel surveys are ratio codes since they represent the quantities of physical attributes, e.g. distances of trips, number of cars in household, age of respondents. Note that ratio codes are often obtained by the substruction of interval codes, e.g. obtaining age from the reported year of birth.

In devising a coding procedure it is therefore important to know how the codes will be used in analysis, so that an appropriate choice of code type can be made at
Chapter 8

this stage. This is particularly the case for scales used in attitudinal questions. If it is desired to carry out a range of mathematical operations on the answers to attitudinal questions, then it is necessary that these answers are in the form of ratio codes. This can be done by suitable selection of the type of attitudinal scale questions to be used in the survey.

8.2.1 Coding Method

In devising a coding procedure, it is important to first decide on the general method which will be used for coding and data entry. Several options are open to the survey designer, each of which will entail a different amount of effort at the coding stage:

(a) Mark-sense or optical-scan questionnaire forms
For questionnaires which consist entirely of closed or field-coded questions, it may be possible to use questionnaire forms on which all answers are recorded directly in pre-specified locations. Each question would have a limited number of categories and the recording of answers would be done by filling in the square or circle corresponding to that category for that question. The questionnaire forms would then be fed directly into an optical scanner attached to the computer whereupon the code categories selected would be transferred directly to a data file ready for editing and analysis. The use of this method is possible only for highly structured questionnaires and is dependent upon the availability of a scanner. Where applicable however, it does provide a very rapid turnaround between data collection and calculation of results.

(b) Self-coded questionnaires
The next step down in automation of the coding procedure is the use of self-coded questionnaires where the respondent, or interviewer, records answers directly into boxes printed on the questionnaire. These boxes are often identified by small numbers indicating the columns into which the data will be recorded onto "card images" within a data file (the idea of card-images is a hangover from the days of mainframe computers, where data was often punched onto 80-character computer cards, which were then grouped into decks of cards representing the entire dataset). The advantage of self-coded questionnaires is that data entry personnel can type information into the computer directly from the questionnaire, thus minimising the need for the use of coders (the distinction between coders and data entry personnel is also a hangover from mainframe days, where the computer cards were typed out by datapunch operators who took no part in the coding of the data from the questionnaires; these days, coding and data entry are often done by the same person). Once again, however, this method is only applicable for highly structured
questionnaires. In addition, the "computerised" appearance of the survey form may be confusing and/or intimidating to respondents when used on a self-completion questionnaire form. Obviously, for the mark-sense and self-coded questionnaire methods, the coding procedure must be completely specified before the questionnaire is printed because both rely on the use of pre-specified codes on the questionnaire form.

(c) Coding sheets
In the past, the most commonly used coding procedure relied on the use of specially designed coding sheets onto which data was manually transcribed from the questionnaire. A typical coding sheet consisted of 24 lines of data each containing 80 columns (to conform with the standard computer card). All data, from both open and closed questions, were transferred to these sheets by coders, and then given to data entry personnel for typing into the computer. Coding sheets were useful when the questionnaire contained a large number of open-questions, but were rather wasteful for closed questions where the task simply involved rewriting all the information from the questionnaire form onto the coding sheet. A trade-off was necessary between the extra time and cost for coders to perform this task against the reduced time and cost for data entry (which was much quicker and less error-prone than typing directly from the questionnaire). These days, however, very little coding and data entry is performed using coding sheets, and the process has largely been replaced by the use of interactive computer programs which perform both the coding and data entry tasks.

(d) Interactive computer programs
With the recent rapid advances in computer technology, it is now possible to combine the tasks of coding and data entry into one by having the coder enter data directly into the computer from the questionnaire form. Thus, instead of writing the data onto coding sheets, the coder types it directly into the computer, each coder being provided with a separate terminal or microcomputer for data entry. This procedure provides for much quicker coding and data entry and has a number of other advantages as will become apparent later. Since this procedure has become the dominant mode of data entry, especially with the widespread and increasing use of microcomputer spreadsheets and databases, particular attention will be focussed on it in the following discussion of coding, although the general principles are equally applicable to the other methods of coding and data entry.

(e) In-field lap-top and hand-held computers
Even more recently, the advent of powerful lap-top and hand-held computers has meant that a greater amount of data entry can be
performed directly in the field. This applies both to observational surveys (such as passenger counts and running time surveys on public transport vehicles) and to various types of interview surveys (such as intercept surveys and stated preference surveys, e.g. Jones and Polak (1992)). The in-field use of computers can drastically reduce the amount of time and effort involved in the coding and entry of data. With appropriate software, it can also increase the accuracy of recordings (especially recording of times in observational surveys) and can greatly reduce the number of errors in the data.

8.2.2 Code Format

Irrespective of the coding procedure adopted, the data that is entered into the computer must be stored in a pre-determined format, i.e. the user of the data must know how to get access to individual pieces of the data set in order to perform the various analyses. In specifying a coding format, a basic question which must be answered is whether data is to be recorded and stored in "fixed" or "free" format. The difference between the two methods is that whereas fixed format specifies exactly where each item of data is to lie within a long string of data (previously, an 80-column card image field, but now more likely to be a 256-character word), free format simply allows for the sequential entry of data separated by some form of delimiter (such as a comma, a space, a tab or a carriage return). Thus, for example, the sequence of numbers 121, 6, 37, 0, 9, 12 might appear in fixed format as:

121 6 37 0 912

where the number of columns reserved for each number is 3,2,4,2,2,2 respectively.

In free format the numbers look like:

121,6,37,0,9,12

When using fixed format it is necessary to specify that all numbers should be "right-justified" in their field, i.e. any blanks must appear to the left of a number. For example in the above example, the number 37 when written in its four-digit field must be written as ∆∆37 rather than ∆37∆ or 37∆∆ (where ∆ represents a blank space). This is because most computers do not distinguish between blanks and zeros when dealing with fixed-length records, and hence ∆37∆ would be interpreted as 370 and ∆∆37 as 3700.

The choice between the two formats will often depend on the computer language and/or programs to be used. Some commercially available statistical programs require that data be in a specified format. Given a choice, however, each format
Data Processing

has its own strengths and weaknesses. Free format is often easier to enter since there is no need to ensure that each data item is typed exactly in the correct columns. On the other hand, fixed format is more efficient to use in programming because it is possible to specify exactly where a data item will be on every questionnaire’s computer record. One can therefore access that data item directly, without having to read all the preceding data items as one would have to with free format. Fixed format is also preferable when using relational databases (such as household, person, and trip files) where frequent cross-referencing is required.

The relative advantages of each type of format can be used most effectively when interactive terminals or microcomputers are used for data entry. In these situations, data is entered by the coder/data enterer in free format and the computer can then assign the data item to a fixed format data file for later analysis. The free format data entry can be taken one step further with interactive terminals by having the computer ask for the next item of data required, by means of a "prompt message" or input field written on the terminal screen. The coder responds to this prompt by entering the required code and then hitting a carriage return. The computer then asks for the next item of data. This technique is particularly powerful when a questionnaire contains numerous filter questions. Depending on the coder’s response to a filter question, the computer can then decide which is the next appropriate question to ask. All questions (and answers) skipped can be supplied automatically with missing values by the computer. This avoids the tedious and error-prone task of typing long rows of blanks for fixed format computer data entry, as often occurs in complex questionnaires.

8.2.3 Coding Frames

A coding frame describes the way in which answers are to be allocated codes. The collection of coding frames for an entire questionnaire is often referred to as a code book. The preparation of a code book involves two distinct processes; the definition of coding categories to be used, and the methods to be used in allocating individual answers to these categories.

Coding frames are required for two different types of questions; closed and field-coded questions, where the codes have already been decided by the time it gets to the coder: and open-questions, where the coder must decide on the choice of code corresponding to each particular answer. The techniques for each type of question will be described below.

Before describing the specifics of coding-frame construction, three general points need to be made. First, for closed and field-coded questions, the choice of coding frame had to have been made at the time of questionnaire design so that
appropriate codes could either be printed on the questionnaire or supplied to the interviewer. Open-question coding frames, on the other hand, generally cannot be determined until after the completed questionnaires are received. Second, in selecting the code categories for both open and closed-questions it is wise to choose categories, wherever possible, which have already been used in other surveys, such as the Census of Population or other transport surveys, with which comparison might later be made. The Census is particularly useful for categories dealing with classification variables, such as occupation and dwelling type. Third, in addition to codes for questionnaire answers it is also necessary to have coding details describing the questionnaire itself. A unique identification number, date of interview (or date of questionnaire return by post), and interviewer’s name are all useful data for control of the analysis procedure.

8.2.3.1 Open-question Coding Frames

The definition of coding frames for open-questions can be a difficult process. For each question, the frame should consist of mutually exclusive and exhaustive categories into which responses may be classified. The process of developing a coding frame for an open-question is usually a mixture of theory and empirical observation. Thus, while the investigator may have some theoretical idea of what the possible codes should be, these can only be verified by reference to empirical data. Such data might come from previous studies, pilot surveys or from the main survey itself.

Using responses from the main survey to define the coding frame entails the selection of a sample of completed questionnaires and then systematic recording of the responses received to the open-question. From this list of responses (or content analysis), a coding frame is designed which satisfies the criteria of being exhaustive and mutually exclusive. Thus every recorded response in the sample should fit into one, and only one, category in the selected coding frame.

The design of coding frames for open-questions is as much an art as a science. A delicate balance is required between over-generalisation and semantic niceties. One problem arises when handling the ubiquitous category of "other". One recommendation is that any response that occurs in the initial sample, however infrequent, should be allocated a category code just in case it turns out to be more significant than expected in the rest of the complete sample. A point worth considering is that it is easy to collapse categories in later analysis but you cannot expand general categories to obtain data that were never coded. On the other hand, if there is a long list of categories with minimal differentiation, the use of such categories for coding is likely to be unreliable and subject to considerable coder bias. One further point to remember is that if, in the analysis, it is desirable to find out how few responses of a given kind are observed, then it is necessary to include such a category code even though it may rarely be used.
Data Processing

8.2.3.2 Closed-question Coding Frames

The coding of closed-questions (or the entry of codes derived for open-questions) is relatively straightforward although the consistent use of certain conventions and techniques can increase the efficiency of the coding process. Some factors which need to be borne in mind when designing a coding frame for closed-questions include:

(a) A decision needs to be made as to what the coding unit will be for each question. Will each question be coded separately or will certain combinations of questions be used to give an overall code with respect to responses to the set of questions? For example, each person in a household may be asked their relationship to the oldest person in the household, and the set of these answers may then be used to derive a variable which describes the overall structure of the household relationships (e.g. nuclear family, single-parent household etc.). As will be described later, it is generally better to keep the coding tasks as simple as possible. Therefore it is generally advisable to code individual questions, and then let the computer derive any combinations at a later date.

(b) Several different types of code are possible. These include the field code in which numbers representing actual amounts are coded, more or less as they were given by the respondent e.g. income recorded to the nearest dollar. If field codes are used, it is necessary that sufficient space be left in the coding frame to accommodate the largest number of digits expected in any answer. Bracket codes involve the assignment of a code to cover a range of numbers. Generally bracket codes would not be used unless the questionnaire itself had used categories to simplify the question for the respondent. The only reason why bracket codes would be used in preference to field codes is if it was essential that space be saved in the coding frame. Pattern codes involve the assignment of a single code to a pattern of responses to either a single question or to a series of questions e.g. patterns of licence holding/car ownership in a household. Pattern codes are often used to save space in the coding frame although the coding task is made more difficult because the coder has to decide on the pattern before selecting the code.

(c) Many computers do not distinguish between blanks and zeros. This means that zero codes with a substantive meaning (e.g. zero car ownership) should not be used in a coding frame where a blank represents something else (such as "not applicable"). This point reinforces the statement made in Chapter 7 that interviewers should not leave any questions blank. It is good practice to differentiate between "not applicable", "don't know", "refused to answer", and simple omission on
the part of the interviewer, because all of these items may be needed for logical edits and quality control checks. Many computer packages provide facilities for several kinds of missing data, and for omitting these missing values from statistical calculations as necessary.

(d) If provision is made for missing data, it is convenient to apply consistent coding conventions for recording such data. For example, a "-3" code could always represent "don't know", a "-2" code could represent "not applicable", a "-1" could represent "no answer provided when one was expected", and so forth. Another situation where a consistent convention is required is in the coding of YES/NO answers. A convention often used is to code YES as "1" and NO as "2". On no account should these answers be coded as "1" and "0" because of the problem with zeros and blanks mentioned earlier.

(e) There are several ways in which space may be saved in coding frames. Sometimes it is useful to make use of embedded codes whereby codes are contained within codes. For example, in response to a filter question a respondent may be asked a different follow-up question depending upon their initial reply. The responses to both the filter and follow-up question may be recorded by one code if, for example, the responses to the first follow-up question were coded as 1 to 4 whereas the responses to the second follow-up question were coded as 5 to 8. In this case the response to the filter question could be ascertained by checking whether the code was less than or equal to 4. Another use of embedded codes is in the construction of identification numbers for respondents where in addition to uniquely identifying the respondent the ID number can also provide some basic information about the respondent. A third use of embedded codes is where the data may want to be used at different levels of refinement. An example of this is in the use of occupation codes, such as the Australian Standard Classification of Occupations (ASCO) used by the Australian Bureau of Statistics, where a one-digit code gives a fairly general split such as:

1  Managers & Administrators
2  Professionals
3  Para-professionals
4  Tradespersons
5  etc., etc.

Within each of these codes, however, a second digit may be used to refine the occupation codes, such that, for example:

21 - Natural Scientists
22 - Building Professionals & Engineers
Third and fourth digits could be added to further stratify the classification of occupations. Each occupation would then be coded as a four-digit number, but in using the data the analyst could choose only that number of digits which is required to give a level of definition commensurate with the purpose of the analysis of the survey data.

(f) When coding responses to menu items, such as the selection of modes used for a given trip from a list of modes of transport, two options are available. First, one could use a listing code wherein each mode is allocated a column in the coding frame and YES/NO responses are coded for each mode. Thus, for example, if six options are listed then a possible coded response might look like 122112 indicating that the first, fourth and fifth modes were used by that respondent. The trouble with listing codes is that, while simple to use, they do take up a lot of computer storage space. An alternative method is the use of a geometric code whereby the same information can be packed into a smaller number of columns. In the above example, the nth mode would be given the code \(2^{n-1}\). Thus the first mode has the code \(2^0 = 1\), the second \(2^1 = 2\), the third \(2^2 = 4\), the fourth \(2^3 = 8\) etc. The combination of modes used would then be coded as the sum of the codes for each mode. In the example given above, the code would be \(2^0 + 2^3 + 2^4 = 25\). This two digit code, which represents one and only one possible combination of modes, contains the same information as the six digit listing code used earlier. The problem with geometric codes is that when used manually by coders there is a large risk of errors in coding. A better method, if coders are using computers, is to enter the data in listing code format and then let the computer pack it into geometric code. With geometric code, a two digit code is equivalent to a six digit listing code, a three digit code is equivalent to a nine digit listing code, while a four digit code is equivalent to a thirteen digit listing code.

(g) While geometric codes are good for representing combinations, they are not suitable for representing permutations (or sequences). In this case a listing code is more suitable, where each mode is assigned a specific code number and these numbers are then listed in the appropriate sequence. One possible alternative is to use a pattern code where each sequence of modes is assigned a unique code. In this case there is a slight reduction in the space required (e.g. a nine digit listing code is reduced to a 6 digit pattern code) but there is also a substantial increase in coding effort if performed manually.
(h) In selecting the number and definition of code categories, attention should be paid to achieving compatibility with other sources, to being exhaustive and mutually exclusive (if required), to minimising the number of columns used in the coding frame, and to achieving a compromise between accuracy and simplicity.

(i) One factor which should not be forgotten when defining code categories is to specify the units in which the answer is required. In transport surveys, the major decisions regarding units will be concerned with clock time (ask in a.m./p.m. format but code in 24 hour format), distance (km or miles), money (dollars or cents), time (minutes or hours) and income (per week or per annum).

8.2.3.3 Location Coding

One particular type of open-question coding frame which is of special significance to transport surveys is the coding of geographical locations e.g. home, work and destination locations. Very often, the coding of locations is the most time-consuming component of transport survey coding, especially when complete trip records for a survey period are involved. The information on the questionnaire generally consists of a street name and suburb - the coding procedure requires this to be transformed into a numerical code. Several approaches have been adopted to attempt this transformation.

The most widespread location coding method, in the past, has been the allocation of locations to zones. Thus, each street/suburb combination lies within a particular traffic zone. This zone is normally determined by the use of look-up-tables - a very time consuming process. The use of zones also has the disadvantage that while they may be aggregated to form villages, towns, cities and counties, they cannot be broken down into finer divisions and the zone boundaries cannot be changed.

In some circumstances, an alternative to the use of specially defined traffic zones is the use of postcodes as a proxy zonal system. The use of postcodes, whilst very convenient in surveys where minimal resources are available for coding, has a number of severe disadvantages for transport surveys. First, they possess all the disadvantages of inflexibility mentioned for zones. Second, the boundaries are often not compatible with other zoning systems, such as town and county boundaries, and hence comparison with other data sources is difficult. Third, the postcodes zones are defined for a completely different purpose to that which is required for transport survey analysis. There is therefore little regularity in zone size or shape, which causes problems for transport analysis. For urban travel surveys, postcodes are often too large for meaningful analysis of origin-destination patterns. The major advantage of using postcodes is that, in many cases, this information can be supplied directly by the respondent on the
questionnaire, and thus the tedious task of coding locations is largely avoided except in cases where the respondent does not know the location's postcode.

In an attempt to avoid some of the problems associated with rigid zonal definitions, a number of surveys (Richardson, et al., 1980; Young, et al., 1978) adopted the use of an x-y coordinate system based on a city street directory. Using this system, which was based on a grid element size of 400m x 400m, coders looked up the street name and suburb in the street directory index and recorded the map number and page coordinates. These two pieces of information were later converted to an overall x-y coordinate system, which covered the entire study area, by a computer program subroutine. The use of such grid coordinates allowed for the amalgamation of grid elements into any larger zoning system such as postcodes or Census districts. It also allowed considerable detail to be retained in the location coding and enables detailed "scattergrams" of location to be plotted using conventional computer printers. Such scattergrams were most useful for graphical presentation of location results, and were also found to be useful in the editing of location information. The major disadvantage with this system is the considerable time involved for coders in looking up the map number and page coordinates in the street directory index.

To alleviate some of the work involved in using the street directory system of coordinates, a self-completion survey of student travel patterns at Monash University in Melbourne, Australia (Jennings, et al., 1983) asked respondents to include the street directory location reference for their home location from the most popular of the commercially available street directories - the "Melways" (Melway, 1992). A total of 60% of the completed questionnaires did include this information, considerably reducing the coding effort required for this survey. As the use of such grid reference systems becomes more widespread in the commercial and retail world (to indicate locations of shops etc. to potential customers), the self-reporting of location codes may become a technique which is worthy of further investigation.

Self-reporting of location codes may also be attempted by asking respondents to indicate the x-y coordinates of trip origins and destinations on a specially prepared map which is supplied to all respondents. The success of this method depend on the size of the study area and the precision with which geographic locations are to be recorded. Given a specific study area, there needs to be a trade-off between the size of the map supplied to respondents and the amount of detail included on the map. Naturally, self-reporting methods assume that the respondents can read a map and that they can, and will, report their trip origins and destinations in terms of x–y coordinates. An example of the use of such a location coding system is shown by the map depicted in Figure 8.1(a) and (b), which was used by the authors in a county-wide travel survey in Upstate New
York. Over 80% of the population in this study was able to correctly identify the location of their homes and their destinations for all trips during a day.

The final method of location coding, which is certainly the most efficient for large transport surveys, is the direct geocoding of street names and suburbs (and other principal locations). This method involves the use of a computer program whereby the user enters an address, the computer checks through a list of such addresses (in much the same way as a coder would look through a street directory index), and then returns an x-y coordinate for that location.

The computer program should be able to match coordinates with incorrect addresses (caused by misspelling of street, or the use of adjacent suburb name) and should take account of house numbers, especially on long streets. The coordinate system adopted will depend on the data base used. With the increasing availability of computer technology, the geocoding system is likely to replace existing systems of location coding since it gives a more accurate result with lower coder effort required. The use of Geographic Information Systems (GIS) for the conversion of geographic information about home addresses and trip destinations into machine-readable format (geocodes) was one of the more innovative aspects of the South East Queensland Household Travel Survey (SEQHTS) survey (Richardson and Ampt, 1993a). In past travel surveys, destination locations have often been coded directly to rather aggregate traffic zones, at suburb level of detail, with the result that considerable information has been lost about the location of destinations. However, coding survey data to the level of the Census Collectors District (CCD) is extremely useful for the plotting of trip information, for more accurate calculation of distances between destinations, and allows greater flexibility for the design of more specific zoning systems (e.g. for the analysis of public transport corridors).

Figure 8.1(a) Map used in Self-Coding of Location Coordinates (front)

Figure 8.1(b) Map used in Self-Coding of Location Coordinates (back)

The general procedure adopted for the geocoding of locations in the SEQHTS project is shown in Figure 8.2.
Locational information was obtained from the travel data files in the form of reported addresses. These addresses may be from the sample frame database of residential addresses obtained from the South-East Queensland Electricity Board (SEQEB), in the case of the household file addresses, or from the respondents, in the case of the stop file destination locations. These addresses were transferred to an address file which contained only the address and an identifier which enabled the geocoded CCD location to be transferred back to the travel files at the end of the geocoding process. The locational information, especially from respondents, was of varying degrees of completeness and accuracy. Therefore, before attempting to geocode the address information, the addresses had to be corrected to put them in a format which was compatible with the GIS database of address coordinates. These corrected addresses were then geocoded by one of various methods of geocoding, as described below. The x-y coordinates of the addresses were then transferred back to the address file. By comparing these coordinates with the CCD boundary files, the CCD in which the address was located was obtained, and this CCD number was then transferred back to the travel data files.

The geocoding procedure used in the SEQHTS survey consisted of a series of geocoding methods applied in a hierarchical fashion to obtain a likely geocode for an address. The accuracy of the geocode is dependent on the geocoding method used. Therefore, the more reliable methods were attempted first.
Chapter 8

The degree of accuracy of the geocoding depends on two factors; the accuracy with which the respondent can supply the locational information, and the accuracy with which the GIS program (in this case, MapInfo®) can use that information to generate a set of coordinates. For example, a respondent might know that they went shopping at the Coles supermarket in Chermside. From their point of view, this is the most accurate description of their destination. However, whether MapInfo® can geocode this location correctly will depend on what information it has about the location of Coles supermarkets. If all Coles supermarkets are in the landmarks datafile, then this should provide a very accurate geocode. However, if they are not in the landmarks file, then the very accurate locational information provided by the respondent will be of little use, unless an alternative method of locating Coles supermarkets can be found. For example, it is possible to look up the Yellow Pages (either the paper version or the electronic version on CD-ROM database) and find that the Coles supermarket in Chermside is on the corner of Gympie and Webster Roads. This information, in that form, is still not very useful since MapInfo® needs a street name and number to find a geocode. However, as will be described later, it is possible to write a special module which finds geocodes based on the specification of cross-streets. Therefore, the accurate locational information supplied by the respondent can eventually be converted into an accurate geocode. On the other hand, the information that MapInfo® is most accurate in working with (i.e. full street name, number and suburb) is often not easily supplied by the respondent. For example, very few people would know the street number of the Coles supermarket in Chermside, even if they knew what street it was on. If they provided only the street name, then we would be forced to select a random position along the street within the suburb - providing a less accurate geocode than that provided by use of the shop name.

In the actual computer implementation of the geocoding methods, four program modules were developed for the SEQHTS project. These are:

- geocoding using MapInfo®;
- geocoding using a cross-street database;
- geocoding with the assistance of a street directory; and
- geocoding by sampling.

In addition, an interactive spelling checker program was developed to automate the correction of spelling errors/mismatches of street names and suburb names. Spellings were checked against a dictionary created from the electronic reference maps provided with MapInfo®.

The next few sections will discuss how addresses are prepared to make them suitable for geocoding and then details of the four geocoding program modules mentioned above will be provided.
Data Processing

Preparation of the address data

A crucial factor in geocoding is the success of matching the address information (i.e. street name and suburb name) provided by the respondents to that used in the electronic reference maps. Slight differences in spellings result in a mismatch and consequently a geocoding failure.

Steps were made to minimise spelling mismatches in the SEQHTS data by providing a pop-up dictionary of street names and suburb names in the data entry program for the travel data. The pop-up dictionary even went as far as displaying only those streets which belong to a specified suburb. However, as the dictionary was not really complete, some addresses were still required to be entered manually. Also, a few entries in the dictionary were discovered to be misspelt but nonetheless they were assumed correct for matching purposes.

The more common causes of spelling mismatches are variations in abbreviations such as Ter & Tce for Terrace, and Mnt & Mt for Mount, and reversals of combinations of names such as Wynnum West and West Wynnum.

Considering that there is so much address information to check for mismatches, a rudimentary interactive program was developed for the purpose. The program starts off by extracting the address records from the travel data files and saving them into an address file. This latter database saves the spelling changes, with the original address information provided by the respondents left unmodified in the former database. A way of relating the information in the address file to that in the travel data file must be maintained to be able to transfer the geocoding information in the address file to the travel data file. This was done via the unique household, person or stop identification numbers.

The interactive spelling checker program was implemented using FoxBASE+/Mac and has the basic features of a word processing spelling checker. It finds an item that is not in the dictionary and displays candidate dictionary items using the "soundex" function of FoxBASE+/Mac. Soundex is used to determine if two words are phonetically similar, that is, if they sound alike.

It was expected that only a few addresses would turn up as mismatches owing to the use of the dictionary pop-up during data entry. However in the case of suburb or locality names there were quite a number of mismatches. This is because the suburb boundary maps for several areas were not complete at the time of the study. Postcode boundary maps were provided, however, so that mismatches in suburb names were resolved by entering postcode numbers.

To speed-up the process in most of the geocoding methods, identification numbers are used instead of the actual names of streets and suburbs. A table of unique identification numbers for each street name and suburb name was
created along with postcodes for each suburb. The identification numbers and postcodes are attached to the address file after the spelling changes have been made.

**Geocoding full street addresses**

The initial task in this procedure is to extract a unique listing of full street address records from the address file. A full street address is one whose street number, street name, and suburb name are given.

Geocoding of full street addresses is done using MapInfo®. MapInfo® basically needs two inputs for geocoding: a street address (which consists of a street number and a street name); and a bounded area (known as a boundary file) such as a suburb or a postcode to refine the search. When provided with this information, MapInfo® finds the street segment in the given suburb containing the specified house number. This street segment is defined in terms of its start point and end point, which are located by means of latitudes and longitudes. In Figure 8.3, this is shown as (X1,Y1) for the start point and (X2,Y2) for the end point. The data file also contains the range of house numbers on either side of the street segment. Thus in going from the start point to the end point of the segment shown in Figure 8.3, house numbers N1 through N2 are on the right while house numbers N3 through N4 are on the left. MapInfo therefore knows what side of the street the specified address is on. For an address on the left side of the street, it then divides the left side of the street segment into \( 1^{st} + \frac{(N4-N3)}{2} \) equal lengths and finds the position along the centre line at which the specified address is located. For example, the house with number N5 is located at a position which is on the left and \( \frac{(N5-N3+1)}{2+N4-N3} \) of the way along the link from the starting point. MapInfo then proceeds to automatically position the geocoded location (the latitude and longitude) 10 metres off the centre line of the street on that side of the street, as shown in Figure 8.3. For example, if N3=1 and N4=7, then house number 5 would be located at coordinates (X,Y) five-eighths of the way along the link and 10 metres to the left of the centre line of the link.

![Figure 8.3 The Location of Full Street Addresses using MapInfo](image)

It is quite common that respondents give incorrect suburb information and so the address cannot be geocoded. This, however, is often circumvented by assuming that respondents are likely to give a suburb not far from the correct suburb. Respondents often upgrade their suburb to a nearby, more socially
Data Processing

distinguished, suburb. By using this assumption, success in geocoding can be improved by re-attempting to geocode using an increasingly larger boundary file.

Postcode boundaries are generally larger than suburb boundaries and so they are used in the geocoding process after the suburb boundary. Larger boundaries are further defined using the nearest eight suburb boundaries and the nearest eight postcode boundaries. The number "eight" is chosen with the idea that if a suburb boundary is roughly square, then there will be four adjacent suburbs on each side of the square and another four on its corners. The nearest eight suburb and postcode boundaries were determined by comparing distances between boundary centroids. This was done only once with the result saved in a database file for use by the appropriate geocoding methods.

It is expected that the probability of correctly geocoding an address diminishes as the boundary used becomes larger.

When geocoding a small file of full street addresses, all methods may be attempted in MapInfo before attaching the geocodes onto the address file. However, when the full street address file is large, it saved time if geocodes were attached to the address file after each method was applied and then the full street address file was compressed by removing geocoded records.

Once an x-y coordinate had been attached to an address, the CCD in which this coordinate was located was found by overlaying the boundaries of the CCDs on the geocoded coordinates. The region (CCD) in which the geocoded point was located was then transferred back to the travel data files as the most disaggregate description of the location of that address.

Geocoding cross-street addresses

As in the geocoding of full street addresses, a list of unique cross-street addresses supplied by respondents was extracted from the address file to avoid unnecessary repetitions in geocoding. A cross-street address consists of two street names and a boundary (e.g. suburb or postcode).

As mentioned earlier, MapInfo did not have the capability to geocode cross-streets, at least as a standard function. A program was therefore written to fill this gap using a fairly straightforward procedure. A database of cross-streets with their coordinates was set-up from the reference maps provided with MapInfo with each record having the following fields:

- street_one - id number of the first street
- street_two - id number of the second street
- x_coord - longitude of intersecting point
- y_coord - latitude of intersecting point
Geocoding a cross-street address was then a matter of searching this cross-street database to find a match between the first and second streets within the appropriate suburb boundary. The latitude and longitude obtained from a cross-street matching correspond to the centre of the intersection of the two streets, i.e. it lies on the intersection of the centre lines, as shown in Figure 8.4. Thus, if the respondent nominates Smith Street as the address, with the nearest intersection being Brown Street (or vice versa), then that location was initially given the coordinates on (X,Y).

![Figure 8.4 The Location of Cross Street Addresses](image)

The last four fields of the cross-street database listed above are necessary because multiple occurrences of a cross-street in various locations are possible. To be able to identify which cross-street is pertinent, the cross-street database has to have a boundary field that qualifies each record. Searching a cross-street in turn must also have boundary information as part of the input. Thus, Smith and Brown Streets may have intersections in several suburbs (because they are different Smith and Brown Streets). So long as the suburb information is provided along with the cross street names, the overlaying of the suburb boundary file will identify which is the correct location.

However, this only partially solves the problem of multiple cross street matches, as multiples may also exist within a single boundary. A good example is a "court" type street where it intersects another street twice, with both intersections likely to be in the same suburb or postcode boundary, as shown in Figure 8.5. Knowing the number of multiples allows for a randomised approach to selecting...
a pair of X and Y coordinates among the multiples. It should be clear that multiple occurrences of a cross-street which are in different boundaries should not be considered as multiples.

![Diagram of Court and Main Street](image)

**Figure 8.5 The Occurrence of Multiple Cross Street Locations**

The geocoding of cross-streets, as in geocoding of full addresses, is also done in successive stages with the next stage using a larger boundary than the previous. Once again, the probability of a correct geocode decreases as a larger boundary is used. For cross-streets, this is aggravated by the random process of selecting a cross-street from its set of multiples, if any.

A further problem with cross street locations is the difficulty of allocating this location to a specific CCD. Because CCD boundaries often run along the centre lines of streets, a cross street often coincides with the junction of several CCDs, as shown in Figure 8.6. Thus each quadrant of the cross street belongs to a different CCD. The question is which CCD should be used to represent the location of the cross street, and hence the location of the destination described in terms of the cross street. Unless further information is provided about the destination, any of the CCDs might be correct. If the destination is on Smith Street, then it could be on either side of the street on either side of Brown Street. Thus any of the CCDs might be the correct one.
Under these circumstances, the (X,Y) coordinate was moved slightly so that it lay within one of the surrounding CCDs. Since there were several thousand destinations in the SEQHTS survey where this problem occurred, it was clearly impractical to move these points manually. Therefore, an automated procedure was devised to accomplish this task. Around each cross street location where there were multiple possible CCD allocations, four candidate sets of coordinates were generated by adding \( \pm 0.00001 \) to both the longitude and latitude of the cross street coordinate. This had the effect of generating four points roughly at 45° to the cross street location, as shown by the small points in Figure 8.7. One of these points was then randomly selected and the CCD in which the selected point was located was then found.
By examining Figure 8.7, it is clear that the above procedure will fail when the cross streets themselves are at 45° to the lines of latitude, since the four new points will lie on the centre lines of the four arms of the cross street. In these cases, the above procedure was rerun, with the four candidate locations being generated at angles of 0°, 90°, 180°, and 270°. This resulted in the selected point generally falling in one of the neighbouring CCDs.

*Geocoding landmarks*

It was allowable in the survey for respondents to nominate a landmark as a destination address. Examples of landmarks include the name of a restaurant, a school, a bank, a government office, a shopping centre, a park, a beach, etc. To be effective as a valid address, a landmark has to be qualified to identify it uniquely from all others with a similar name. A bank, for example, needs to have the branch (usually a suburb) appended to its name.

The geocoding of landmarks was done by one of several means. Firstly, it was anticipated that many landmarks would be able to be geocoded by searching a database of landmark names with geocodes, supplied with MapInfo. Unfortunately, the original landmarks database supplied with MapInfo only contained railway stations as landmarks. Because of the limited scope of the original landmarks file, it was necessary to create a new landmarks file containing a wider variety of landmark types. This was done by compiling information from various sources such as telephone books and street directories on a variety of landmarks, such as:

- schools, pre-schools and childcare centres
- universities and colleges
- shopping centres
- food outlets
- sporting centres
- places of interest
- parks, ovals and reserves
- caravan parks
- hospitals
- ambulance stations
- police stations
- fire stations
- churches
- Guide and Scout halls
- Masonic Centres
- bus and airline offices
- ferry terminals
- post offices
Chapter 8

- public libraries
- council offices
- bays and beaches
- boat ramps
- theatres and cinemas
- hotels and motels
- commercial buildings
- racecourses
- TAB agencies
- golf courses
- bowling centres
- swimming pools

For each of these landmarks, an equivalent full street address or cross-street address was determined manually from these printed sources, and then the geocoding methods for full street addresses and cross-street addresses (described earlier) were used to generate the geocodes.

Not all landmarks were easily identified by a street address. Finding an address for a landmark posed a problem in cases where one was not available and/or the area covered by the landmark was large (e.g. beaches and parks). For such large areas, the area centroid may represent a more appropriate definition of the location to be used as a geocode. Centroids of areas could be marked and geocoded in a MapInfo map, but this process proved to be laborious. An alternative geocoding method was, therefore, developed.

The alternative method involved the development of a computer program that generates a geocode given a map reference from a street directory. An example of a map reference is "A 4 15" where "A" and "4" are row and column references respectively while "15" is a map number in the street directory. A map reference may also be specified as a fraction for a more precise specification as in "B.2 6.3 48A", where "B.2" refers to a point which is 20% of the way between row B and C, "6.3" refers to a point which is 30% of the way between columns 6 and 7, and "48A" refers to map 48A.

The task of assigning map references to landmarks was made less taxing by having somebody who was knowledgeable of the study area do the work. In addition, some data entry personnel entered map references as part of the destination address on a number of occasions.

The street directory map reference method was used extensively in the SEQHTS survey to geocode full street addresses and cross-street addresses that failed to obtain a geocode in their respective methods, primarily because the street networks in that area were missing from the MapInfo electronic reference maps.
**Data Processing**

This method worked well where the address could be positively located in the street directory maps. However, even for a full street address, the task of identifying the exact location using the street number was sometimes difficult, especially when the street directory maps did not show street numbers.

The accuracy of the geocodes obtained using this method, however, depends greatly on the accuracy of the street directory maps and the accuracy with which the maps had been digitised into the computer files. Accuracy may be verified by mapping, in MapInfo, a sample of geocodes obtained using this method for each street directory map. This is important to determine the real position of each geocoding method in the hierarchy of geocoding methods.

*Geocoding by sampling*

Addresses provided by respondents were not always complete. Some respondents intentionally omitted street numbers or just indicated their suburb or locality - probably for privacy reasons. The approach that was used to geocode these cases in the SEQHTS survey was to sample a point along the length of the street, if a street name was given, or to sample a point within a suburb, if a suburb was all that was available.

A long or winding street in a MapInfo map is divided into short segments, usually at street intersections and when it bends or changes direction. Sampling a point along a street therefore consists of gathering all the segments belonging to the given street within a given area (suburb or postcode), then randomly selecting which segment to use (segments may be assigned relative weights based on their lengths), and then sampling a point along the selected street segment. Sampling a point within an area (suburb or postcode) also followed this procedure, with the added step of firstly randomly selecting a street among the streets within the area.

In addition, the selection of the side of the street was also randomised, and the sampled point was then offset transversely from the street by about 10 metres. This was felt to be necessary as the lines defining the streets on a MapInfo map represent the centre lines of the streets and thus an adjustment had to be made to account for the street width. This adjustment was required because CCD boundaries also follow the centre lines of streets, and this method minimised the incidence of locations falling on the boundaries between adjacent CCDs. The offset of 10 metres is consistent with the way in which MapInfo geocodes full street addresses, as shown in Figure 8.3. This practice, however, resulted in some geocodes "spilling out" of boundary files or onto water areas when the street segment was near a river bank or beach. These occurrences were corrected manually, after visually examining a plot of the geocoded points.
Chapter 8

As in geocoding of full street addresses and cross-street addresses, progressively larger boundaries were used when the given street could not be found within the given suburb boundary.

The methods belonging to this last category of geocoding were all implemented outside of MapInfo using specially written program modules, but using the reference maps provided with MapInfo. Locations in other parts of Queensland, in other States, and overseas were not geocoded, but were assigned a pseudo-CCD code to assist in identifying their location.

The methods described above gave rise to a range of geocoding methods, which were recorded in the respective data files using the following codes:

<table>
<thead>
<tr>
<th>Code</th>
<th>Geocoding Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>&quot;full address, exact match on suburb&quot;</td>
</tr>
<tr>
<td>11</td>
<td>&quot;full address, exact match on postcode&quot;</td>
</tr>
<tr>
<td>12</td>
<td>&quot;full address, exact match on nearest 8 suburbs&quot;</td>
</tr>
<tr>
<td>13</td>
<td>&quot;full address, exact match on nearest 8 postcodes&quot;</td>
</tr>
<tr>
<td>14</td>
<td>&quot;interactive matching&quot;</td>
</tr>
<tr>
<td>20</td>
<td>&quot;cross-streets, exact match on suburb&quot;</td>
</tr>
<tr>
<td>21</td>
<td>&quot;cross-streets, multiple exact matches on suburb&quot;</td>
</tr>
<tr>
<td>22</td>
<td>&quot;cross-streets, exact match on postcode&quot;</td>
</tr>
<tr>
<td>23</td>
<td>&quot;cross-streets, multiple exact matches on postcode&quot;</td>
</tr>
<tr>
<td>24</td>
<td>&quot;cross-streets, exact match on nearest 8 suburbs&quot;</td>
</tr>
<tr>
<td>25</td>
<td>&quot;cross-streets, multiple exact matches on nearest 8 suburbs&quot;</td>
</tr>
<tr>
<td>26</td>
<td>&quot;cross-streets, exact match on nearest 8 postcodes&quot;</td>
</tr>
<tr>
<td>27</td>
<td>&quot;cross-streets, multiple exact matches on nearest 8 postcodes&quot;</td>
</tr>
<tr>
<td>30</td>
<td>&quot;landmark, exact match using MapInfo landmarks&quot;</td>
</tr>
<tr>
<td>31</td>
<td>&quot;landmark, with equivalent full address&quot;</td>
</tr>
<tr>
<td>32</td>
<td>&quot;landmark, with equivalent cross-streets&quot;</td>
</tr>
<tr>
<td>33</td>
<td>&quot;landmark, exact match using UBD Refidex landmarks&quot;</td>
</tr>
<tr>
<td>40</td>
<td>&quot;sampling along a street, within a suburb&quot;</td>
</tr>
<tr>
<td>41</td>
<td>&quot;sampling along a street, within a postcode&quot;</td>
</tr>
<tr>
<td>42</td>
<td>&quot;sampling along a street, within nearest 8 suburbs&quot;</td>
</tr>
<tr>
<td>43</td>
<td>&quot;sampling along a street, within nearest 8 postcodes&quot;</td>
</tr>
<tr>
<td>50</td>
<td>&quot;sampling of street, within suburb&quot;</td>
</tr>
<tr>
<td>51</td>
<td>&quot;sampling of street, within postcode&quot;</td>
</tr>
<tr>
<td>60</td>
<td>&quot;not geocoded, but pseudo-SLA coded&quot;</td>
</tr>
</tbody>
</table>

8.2.4 Coding Administration

The process of coding needs to be carefully supervised if high quality data transmission is to be obtained from questionnaire to computer. The process really consists of two phases, although in many surveys these will be performed simultaneously by the same person. The two phases are coding and data entry. In many surveys, these tasks will be deliberately separated with a team of coders deciding on and allocating codes to the questions, and a team of high-speed
Data Processing

typists entering data into the computer. The following comments apply more to
the coders than the typists.

A decision which needs to be made before coding begins is whether data entry is
to be independently verified. This verification is designed to detect errors which
are purely typing errors and which might otherwise be undetectable. Verification
is performed by having a second typist type in every form which has already
been typed in and then having the computer check the second typed entry
against the first typed entry. If any discrepancies are noted an error message is
printed so that a check can be made against the coding form or, if necessary, the
original questionnaire.

The advantage of verification is that it reduces typing errors to virtually zero. The
error rate for experienced typists is in the order of 2%. If verification is used, the
error rate can be reduced to 0.04%. The disadvantage of verification is that it
obviously doubles the cost of data entry. It should also be realised that many of
the typing errors will be picked up by other editing checks at a later stage. The
only errors which will not be detected are those which do not go out of range or
cause a logical error. The decision as to whether to verify will depend on how
serious these types of errors are seen to be to the overall survey analysis.

The recruitment and training of coders follows a similar procedure to that
outlined earlier for interviewers. The main difference is in the skills required. The
principal requirement for coders is that they possess good clerical skills, clear
handwriting and plenty of patience. Note that these skills are not the same as
those required for interviewers. Therefore good interviewers will not necessarily
be good coders, and vice versa. As with interviewers, however, the major
training for coders should consist of practice coding. A useful technique is to have
all coders code the same set of questionnaires which have been chosen, or
artificially constructed, to illustrate many of the different types of problems
which the coders are likely to encounter.

In production coding, a decision must be made as to the order in which questions
are to be coded from the questionnaire. With simple questionnaires it is easiest if
all questions are coded in the order in which they appear on the questionnaire,
letting the computer rearrange the items into logical units of information. For
more complex questionnaires with several difficult questions to code, it is
worthwhile employing some special coders whose job it is to code the more
difficult questions such as location coding and the coding of open-questions.
These coders may mark their codes directly on the questionnaire forms and leave
it to the general coders to transfer these codes to the coding sheet for typing.
The advantage of special coders is that the variability between coders is reduced
and the productivity is increased since the special coders quickly develop a feeling
for their task and are able to innovate short-cuts to help in the task. With the use
Chapter 8

of computer terminals for data entry, special coders can also enter their codes directly into the computer by accessing the respondent’s record by means of the unique identification number. They then leave it to the general coders to access the record at a later date to enter the remainder of the codes.

It is useful in production coding for the supervisor to perform some check-coding on a sample of the coded questionnaires. This quality control procedure enables errors to be detected in an individual coder’s work or enables detection of systematic errors by many coders which may indicate a deficiency in the coding frame or the coding instructions. Typically, the sampling rate for check-coding should be quite high at the beginning of the coding process, and should reduce to random spot-checks when most problems in coding appeared to be rectified. If check-coding did detect deficiencies in the coding frame (such as not enough categories for a particular question), then it is essential that the supervisor be the one who issues the amendments to the coding frame. Individual coders should not be allowed to add extra codes as they see fit; certainly they should be encouraged to suggest changes to the supervisor, but it is the supervisor’s job to issue changes.

Despite all the best intentions and preparations, coding is rarely error-free. Two principal sources of error are coder variability and reliability. Coder variability is a measure of the discrepancy in codes assigned to the same response by two different coders whereas coder reliability is a measure of the discrepancy in codes assigned by the same coder at two different points in time. In a study reported by Durbin and Stuart (1954) it was shown that although coder variability was greater than coder unreliability, both were a cause for some concern. Significantly, it was shown that both were more of a problem with more complex questions. The implication of this is that the coding task should be kept as simple as possible with all editing and data transformations being performed by the computer after data has been entered.

8.3 COMPUTER-BASED DATA ENTRY AND EDITING

8.3.1 Data Entry

The methods of data entry which are becoming increasingly popular are the use of commercially available spreadsheet and database programs (such as Microsoft Excel, Lotus 1-2-3, dBase, and FoxBASE/FoxPro). These generic programs provide most of the facilities for data entry, editing and basic analysis. The concepts underlying spreadsheets and database programs are, however, subtly different.

A spreadsheet, such as Excel, subdivides the computer screen (and beyond) into a series of intersecting rows and columns. Each cell so created can hold an item of
Data Processing

information, where that item may be text, a number, or an equation which operates on the contents of other cells. Each cell is uniquely identified by the name of the intersecting row and column. In using a spreadsheet for data entry from a questionnaire, two options are available: first, the columns can represent data items for each questionnaire, while the rows represent different respondents to the questionnaire. Second, the screen can be set up as a replicate of the actual questionnaire form, and the coder then simply copies the data from the actual form to the replicate on the screen.

An example of the use of spreadsheets is given below. Suppose that a sample questionnaire such as that shown in Figure 8.8 is to be coded. Using the first spreadsheet method, the screen might appear as shown in Figure 8.9, whereas the second method would result in a screen such as that shown in Figure 8.10. Note that it is possible to eliminate the row and column gridlines and the row and column headings from the screen display, so that the data entry form looks like a normal sheet of paper.

![HOUSEHOLD FORM]

How many of the following vehicles are in this household?
(Please include all vehicles usually kept here overnight)

- Cars & Station Wagons
- Vans & Pickups
- Motor Cycles
- Bicycles
- Others (please specify type) ..........................  

Figure 8.8 Sample Questionnaire to be Coded
The use of a database for data entry has some similar features in that the screen can be set up with either row and column entries, or with a replicate of the questionnaire. The main difference between a database and a spreadsheet is that a spreadsheet identifies a data item by its position within the spreadsheet (in terms of row and column numbers), while a database identifies a data item by means of a field name and a record number. Another difference lies in the way that data items are normally displayed; a spreadsheet displays data for several respondents in the form of a matrix, whereas a database often displays the contents of each field for one record at a time (although this is not necessary).
different types of data structure encountered in typical transport surveys: flat-file databases, and relational databases. A flat-file database is often used with relatively simple surveys where the data from each respondent to the survey is stored in one record within one data file. A relational database is used when there is a natural hierarchical nesting to the data obtained from one respondent, or group of respondents. For example, in many transport surveys, travel data are collected from households. There is a set of data describing the household itself, some data describing each member of the household, and some more data describing each trip made by each member of the household. While it would be possible to store the data describing the person and the household with the data for each trip, this would be very wasteful of storage space since the household and person data would be repeated many times. A more efficient method is to store the household data in one file, the person data in a second file, and the trip data in a third file, and then to establish "linkages" between the three files. These linkages or "relationships" form the basis of a "relational database". The concept of a relational database for travel survey data is depicted graphically in Figure 8.11.

![Figure 8.11 A Relational Database for Travel Survey Data.](image-url)
Chapter 8

The household file is always the shortest of these files while the trip file is always the longest (in terms of number of records). In Figure 8.11, the relationships between the files are shown by the connecting lines. Thus the information about the people in the first household is stored in the 3rd and 4th records of the person file, while the information about the people in the third household is stored in the 8th, 9th and 10th records of the person file. For these three people, the trips for the first person are stored in the 11th, 12th and 13th records of the trip file, the second of these people made no trips, while the trips for the third person are stored in the 14th, 15th, 16th and 17th records of the trip file. In order to establish these relationships, both the household and person files must contain a field which uniquely identifies that household or person in the appropriate data file. Each record in the trip file then contains a "pointer" which uniquely links each trip with one and only one person in the person file. In turn, each record in the person file contains a pointer back to the household file. By storing only a pointer back to the person and/or household in the trip file, rather than the full details of the person or household, substantial savings in storage space can be achieved.

As an example of a relational database for travel surveys, consider the data illustrated in Figure 8.12. Because of the structure of the data, and the size of the database, it makes sense to store it in a relational database, rather than a flatfile database. In FoxBASE, an individual database is set up in a work area as represented by the circles in Figure 8.12. Within each of the ten work areas, a different database can be opened. In the case shown in Figure 8.12, the household data file has been opened in work area A, the person file in work area I, and the trip file in work area G.

The databases in the work areas can be related as shown by the arrows connecting the work area circles. Thus the trip file is related to the person file (i.e. extra information describing each trip is contained in the person file), and the person file in related to the household file. By inference, therefore, the trip file is also related to the household file. To create these relationships, a pointer variable must be common between the two files being related. In this case, the trip file contains a variable identifying the person making these trips, while the person file contains a variable identifying the household to which the person belongs. In addition, the file to which the relationship arrow points must be "indexed" on the pointer variable. "Indexing" a file is similar to sorting the file on the basis of that variable; however, whereas sorting rearranges the records in the file, indexing creates another index file which stores the order of the records if they had been sorted on that variable. Indexing essentially provides a fast way of finding records based on the indexing variable. It can be seen that the person and household files have been indexed in Figure 8.12, by means of the "index finger" pointing to those work area circles.
One of the advantages of relating database files is that one can now identify to which person and household a particular trip belongs. For example, by clicking on the fourth row of the trip file (as shown in Figure 8.13 by the highlighted cell) and then clicking anywhere in the person file window, the person to whom that trip belongs will immediately be highlighted (as shown by the highlighted cell for person 3 in Figure 8.14).
Figure 8.15  Selection of the Corresponding Household Record.

By clicking anywhere in the household file window, the household to whom that trip and person belongs will also be immediately highlighted (as shown by the highlighted cell for household 8 in Figure 8.15).

The ability to quickly find related trips, persons and households is very useful when coding and editing the data. However, a relational database is not the ideal format for performing statistical analysis when you wish to test for relationships between variables in different files (e.g. trying to construct a modal choice model based on trip, personal and household characteristics). For these types of activities, we need to convert the relational database structure back into a flatfile database (probably for use with a different statistical analysis program, since database programs are not particularly well-suited for multivariate statistical analysis).

To make this conversion, it will be necessary to copy some of the variables from the trip, person and household file to another file (which is a flatfile). This is done simply by specifying the name of the new file, and by selecting the variables from the three files that you want to copy across to the new file. Because of the relational database structure, FoxBASE will know which person and household record to go to when you select a variable from either of those files. Figure 8.16 shows a flatfile which has been created for the purpose of statistical analysis (to be described in Chapter 10), and which has subsequently been assigned to a new work area in FoxBASE. Whilst this file contains as many records as there are persons (in this case, 904), it contains only eleven variables (including household and person identifiers) instead of the entire set of 38 variables for all three databases. It is therefore possible to work more quickly and with less memory requirements when analysing this data set.
8.3.2 Data Editing

Once the data have been coded and entered into the computer, the major task of editing can begin. The editing phase of the survey process is perhaps the most boring but it is also one of the most important tasks. Most survey designers would admit that more time and effort goes into the editing task than almost any of the other tasks; and such effort is worthwhile. It is useless to proceed straight into analysis hoping that the data are free from error; there will always be errors in the data as initially coded. This error arises from several sources; the respondent, the interviewer, the coder and the typist. The errors may be genuine errors in judgement or reporting, or may arise from problems in legibility. Some of these errors may be detectable during editing and some of them may be able to be corrected.

The main editing technique for the detection of errors is the simple process of tabulation and cross-tabulation. The construction of frequency distributions, the calculation of means, standard deviations and Z-statistics, and the plotting of the data by computer may also assist in detecting outliers in the data. The three major problems which may be detected during editing are permissible range errors, consistency checks and missing data.

Permissible range errors:

Typing and recording errors may result in obvious errors where the code value is outside the range of codes permissible for that response, e.g. a code of 4 for a YES/NO question. In many cases, this type of error can be corrected by referring to the original questionnaire or coding sheet where, very often, a mistake has been made in transcription because of poor legibility. In other cases, a
misunderstanding as to the units to be used in the response will cause the answer to be outside the allowable range.

Permissible range error checks are all within a single data file. Examples of range error checks performed in the 1992 SEQHTS survey (Richardson and Ampt, 1993a) include:

**Household Form**

- Household size cannot be 0 when number of visitors is 0 (error)
- Household size is usually not more than 10 (warning)
- Household size cannot be negative (error)
- Number of visitors is usually not more than 10 (warning)
- Number of visitors cannot be negative (error)
- Number of vehicles should equal sum of all vehicle types (error)
- Number of vehicles is usually not more than 10 (warning)

**Person Form**

- Number of person records cannot be more than household size (incl. visitors) on household form (error)
- Birth year should not be more than 92 (warning)

**Vehicle Form**

- Number of vehicle records cannot be more than number of registered vehicles on household form (error)

**Stop Form**

- Arrival time cannot be less than 0400 (error)
- Departure time cannot be less than arrival time for stop (error)
- Arrival time cannot be less than departure time of previous stop (error)

*Logic checks:*

Cross-tabulations will often reveal logical inconsistencies in the data, e.g. a household with three people having four driving licence holders. Often, checking with the original questionnaire will show that one of the responses has been transcribed incorrectly and can be easily corrected. If both responses have been coded as they appear on the questionnaire then it is obvious that an error has been made in recording one of the responses. To determine which one is in error it is necessary to check other responses for that respondent to see which of the two responses is most likely to have been correct. For example, closer checking of the questionnaire, in the above example, may reveal that only two of the
Data Processing

household members were of licence holding age and two cars were owned by the household. It would therefore appear reasonable to recode the number of licence holders as two. In other cases, a logical inconsistency while not being impossible, as in the above example, may be highly improbable e.g. a low-income two-person household owning five cars. Again the questionnaire should be checked to determine whether responses to other questions indicate that one of the responses is in error. In both cases, if no evidence can be found that one of the responses is in error, it is best to leave the responses as they stand.

Logic checks are cross-tabulation checks, sometimes within one file and sometimes across more than one file. Examples of logic checks performed in the SEQHTS survey (Richardson and Ampt, 1993a) include:

**Within Stop File**

- The last trip of the day should normally be to home. Check other destinations.
- Trips with home as the origin and destination should be checked. Usually indicates a missing stop record. The same applies to any other location (not purpose or place) when it appears as both the origin and destination (e.g. "my workplace" to "my workplace" without an intermediate destination).
- Check for combinations of destination place and purpose (see section 2.5.2 for a full description)
- If the mode is public transport, then the destination place should normally be a terminal for that mode (e.g. a bus trip to a bus stop).
- Trips with very high speeds (for the mode concerned) should be checked.
- All trips of more than 2 hours duration should be checked.
- Walk trips of more than 1 hour duration should be checked.

**Within Person File**

- Check year of birth against driver’s licence (should not be less than 17)
- Check year of birth against full-time employment (investigate if less than 15)
- Check year of birth against retired/old age pension (investigate if less than 55)
- Check year of birth against preschool/childcare (investigate if more than 6 years old)
- Check year of birth against primary school (should normally be between 4 and 13 years old)
Chapter 8

- Check year of birth against secondary school (should normally be between 12 and 20 years old)
- Check year of birth against university/college (should normally be greater than 16 years old)
- Compare entries in employment, studying and other activities fields. Should normally be at least one valid entry in one of the fields; if so, other fields should be coded as "not applicable"; if not, all three fields should be coded as "missing".

Within Vehicle File

- Check spelling of vehicle makes and models
- Switch make and model name, if necessary (e.g. Falcon Ford should become Ford Falcon)
- Check number of cylinders against make and model.

Between Admin and Household Files

- Those households on the Administration file with a response code corresponding to a valid response should appear on the Household file.
- Conversely, those households on the Administration file without a response code corresponding to a valid response should not appear on the Household file.

Between Household and Person Files

- The number of records in the Person file for a household should correspond to the number of residents and visitors specified on the Household file.

Between Household and Vehicle Files

- The number of registered vehicles on the Household file should agree with the number of vehicles for which information is supplied on the Vehicle file. If vehicle details are missing from the Vehicle file, missing values should be entered.

Between Person and Stop Files

- People without licences in the Person file should not appear as car drivers in the Stop file.
Between Vehicle and Stop Files

- The vehicle number specified for any particular Stop record should correspond to an actual vehicle record on the Vehicle file.

**Missing data:**

Editing checks may reveal two types of missing data; firstly where complete interviews are missing, and secondly where responses to particular items are missing. With missing interviews (i.e. non-response) it is advisable to apply differential weighting to the survey results, if something is known about the non-respondents, so that the sample results more closely represent the population results. Various methods of using adjustments for non-response are described by Kish (1965) for general surveys, and are covered in greater detail in the next Chapter.

A more common problem of missing data relates to responses missing from particular questions. In this case, three options are available:

(i) Ignore the missing values and report the distribution of responses in terms of the total number of responses received;

(ii) Report the percentage missing for each question so that results are based on the total number of completed questionnaires; or

(iii) Attempt to estimate a probable value of the missing datum, using information contained in the responses to other questions.

As an example of the treatment of missing data, consider responses to a question on income. This question typically has a higher-than-average number of missing data values. If these missing values were distributed evenly across the range of income values, no particular problems would arise in the calculation of average income. However it has been observed in many surveys that higher non-reporting of income occurs with higher income respondents (e.g. Ministry of Transport Victoria, 1981). Therefore estimates of the average income based only on the data obtained from respondents will systematically underestimate the average income. More importantly, the difference in income between different areas will appear to be less than it really is, if options (i) or (ii) are used for treating the missing data.

It is possible however to estimate what the missing value of income might be for any one respondent by utilising information from responses to other questions (e.g. Bhat, 1994; Loesis and Richardson, 1994). For example, it can be shown that income is correlated with other variables such as car ownership, employment status, occupation, gender and age. Each of these variables is less likely to be
Chapter 8

missing in the survey responses than income and therefore in most cases an estimate can be made of income. Whilst the resultant estimate is not intended to represent what the income really is, the use of such estimates can markedly improve the validity of parameter estimates. It is better to make such estimates than to perform calculations where there is a high proportion of missing data.
9. Weighting & Expansion of Data

This section addresses the issues that arise from the fact that a number of factors will have interfered with obtaining the exact information, both in quality and quantity, from the survey data in spite of the analyst's best efforts to chose the appropriate survey method, to develop the best instrument possible, and to administer and execute the survey meticulously. Why then do we suspect that we will not quite get the information we want, and why should any corrections, adjustments and weightings be necessary?

The answers to these questions fall into a number of categories. After the survey instrument was distributed, many of the analyst's conceptual, theoretical, and logical considerations were up against a test in the real world; namely the behavioural characteristics of the human beings from whom the survey information was to be obtained. And these human beings do not necessarily respond to our request in line with our wishes, expectations, and theories. Some of them were not able to respond to our request, others did not want to cooperate, others responded only partially, others again misunderstood some questions on the survey instrument.

Yet in spite of the less than perfect response that is likely to have occurred, the investigator still wants to, and has to, use the data to obtain information that is relevant for the survey population and not just for the subsample of people that responded "perfectly". It should be remembered here from the discussion of
Chapter 9

sampling theory in Chapter 4 that the original intent, based on this theory, was to develop population estimates on the basis of a carefully selected sample of that population. Unfortunately, in virtually all surveys, the population estimates have to be derived on the basis of a response of less than one hundred percent, in most instances from substantially less than this ideal target.

The purpose of this chapter then is to make the analyst aware of both the likely reasons for, and the consequences of, having to deal with only a subset of the desired sample. An awareness of these reasons and, particularly, of the effects of an imperfect response rate, can go a long way towards understanding the limitations of the survey results, the likely magnitude, direction, and implications of any biases resulting from them, and towards the developments of any adjustments and compensating measures that might be possible.

The research literature is full of examples where researchers have concentrated their efforts exclusively on the intellectually more challenging and satisfying exercise of developing sophisticated mathematical models without proper attention to the quality of the data that they use to validate these models. However, as will be explained in Chapter 10, there is a tradeoff between data quality and sophistication of the modelling process. Without a proper knowledge of the characteristics of the dataset used, it is almost impossible to draw proper conclusions about the quality of such models, since the source of the problem could lie either in the data base or the model itself.

The other area where the improper use of survey subsample information can lead to disastrous results is in the area of simple statistical information and conclusions derived from sample information. Given the multitude of surveys conducted every day, this area is probably the more serious one, since such statistical information is used daily at all levels of government and in the private sector for short-term and long-term decision-making, investments, and projections.

There are three major sources of systematic error (bias, distortion) in a typical sample survey dataset, namely:

(a) Non-response

(b) Non-reporting

(c) Inaccurate reporting

*Non-response* pertains to the situation where a household or individual did not provide a response at all, i.e. no survey form was filled out. *Non-reporting* refers to survey responses where the analyst is in receipt of a survey form on which certain questions have not been answered. *Inaccurate reporting* describes the cases
Weighting & Expansion of Data

where the analyst has determined that some of the responses provided on the survey instrument are objectively incorrect, inaccurate, or incomplete.

In order to compensate for these deficiencies in the typical sample survey data set, a number of "repair" strategies can be pursued:

(a) Editing.

This standard process of "repairing" the survey responses simply eliminates obvious omissions, errors, etc. that can be rectified by objective judgment on the part of the survey administrator and his/her staff. Editing obviously will not address the problem of non-response, and in most cases will not contribute to overcoming the non-reporting issue.

(b) Coding.

The coding process eliminates additional errors and omissions in addition to identifying inconsistencies among the answers given by the respondent. This process does not address the non-response problem (obviously, coding and editing can only take place if the questionnaire was returned).

(c) Weighting Factors.

Socio-economic and statistical adjustments to account for non-observed information.

9.1 POPULATION EXPANSION FACTORS

As stated several times already, the eventual purpose of a sample survey is to be able to draw conclusions about the characteristics and behaviour of the population from which the sample was drawn. If the sample has been selected according to the simple random sampling method described in Chapter 4, then theoretically the results of the sample survey can be expanded back up to the population by multiplying by the inverse of the sampling fraction. For example, if a sample of 100 people has been randomly selected from a population of 1000, and if it has been found that this sample makes a total of 287 trips per day then the total number of trips made by the population can be inferred to be 2870. However, while the concept of sample expansion is quite simple, the process is rarely as simple as described above, for the following reasons:

(a) Even with a simple random sample, there is no guarantee that the sample is truly representative of the population. Chance random errors will result in some groups within the population being over-represented,
while others are under-represented (remember the example of the males and the females in the simple random sample in Chapter 4). If the variable in question (e.g. the number of trips per day) varies systematically across these groups, then simple expansion of the sample results will not necessarily provide good population estimates.

(b) In many situations, we will have used a more complex sampling procedure, some of which (such as variable fraction stratified random sampling) will never produce a sample which is representative of the population, because we have deliberately under- and over-sampled the strata. To obtain population parameter estimates, we need to take explicit account of the manner in which the sample was drawn, and then work backwards to reconstruct the population estimates.

(c) Even if we have accounted for the manner in which the sample has been drawn from the population, and if a perfectly representative sample had been drawn, there is still no guarantee that what we obtain from respondents is what we expected to obtain. For example, not all people will respond to the survey; furthermore, this non-response is unlikely to be evenly distributed across the groups within the population. Thus the distribution of respondents, across various characteristics, is unlikely to be the same as the distribution of the sample across those parameters.

For the above reasons, it is usually necessary to explicitly account for the composition of the respondents before expanding the results to represent the population to which the respondents belong. This explicit recognition is performed by means of population expansion factors, which relate the composition of the respondent group to a known composition of the population. In order to calculate these expansion factors, however, it is necessary to have a secondary source of data describing the population in terms which can also be related to the sample. The most common source of secondary data is a national Census of Population, which provides complete information about the population with respect to key socio-economic variables. Provided that your survey asks these same questions of your respondents (in the same way and using the same response categories), then you can calculate population expansion factors to obtain population parameter estimates.

To give an example of the calculation and use of population expansion factors, consider a survey in which a sample of 1000 people over the age of 15, from a total population of 10000, are surveyed by postal questionnaire about their trip making behaviour. As part of the survey, assume that each individual is asked their age, their sex, and the number of trips they made on a specified day. Assume that the sample was randomly drawn from the population, and that the overall response rate was 40%. The number of responses in each age/sex
Weighting & Expansion of Data

category are shown in Table 9.1, and the average number of trips per day for each category is shown in Table 9.2. Based on this information we can calculate, by a weighted average, that the average trip-rate in the sample was 3.17 trips per day.

Table 9.1 Responses by Age and Sex

<table>
<thead>
<tr>
<th>Age---&gt;</th>
<th>15-24</th>
<th>25-34</th>
<th>35-44</th>
<th>45-54</th>
<th>55-64</th>
<th>65+</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>56</td>
<td>30</td>
<td>14</td>
<td>13</td>
<td>19</td>
<td>20</td>
<td>152</td>
</tr>
<tr>
<td>Female</td>
<td>83</td>
<td>65</td>
<td>28</td>
<td>20</td>
<td>21</td>
<td>31</td>
<td>248</td>
</tr>
<tr>
<td>TOTAL</td>
<td>139</td>
<td>95</td>
<td>42</td>
<td>33</td>
<td>40</td>
<td>51</td>
<td>400</td>
</tr>
</tbody>
</table>

Table 9.2 Trip-Rates in Responses by Age and Sex

<table>
<thead>
<tr>
<th>Age---&gt;</th>
<th>15-24</th>
<th>25-34</th>
<th>35-44</th>
<th>45-54</th>
<th>55-64</th>
<th>65+</th>
<th>AVE.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>2.14</td>
<td>3.10</td>
<td>4.79</td>
<td>5.85</td>
<td>4.74</td>
<td>3.25</td>
<td>3.36</td>
</tr>
<tr>
<td>Female</td>
<td>2.53</td>
<td>1.91</td>
<td>3.50</td>
<td>6.70</td>
<td>4.90</td>
<td>2.81</td>
<td>3.05</td>
</tr>
<tr>
<td>AVE.</td>
<td>2.37</td>
<td>2.28</td>
<td>3.93</td>
<td>6.36</td>
<td>4.83</td>
<td>2.98</td>
<td>3.17</td>
</tr>
</tbody>
</table>

While it is known that the overall response rate was 40% (because we received 400 replies to the 1000 questionnaires distributed), we do not know the response rates in the individual categories. To calculate these response rates, we need to know the total number in the population in each category. Suppose that we have a secondary data source which provides the number in the population in each of these categories, as shown in Table 9.3.

Knowing this information, we can now calculate the response rates in each category (assuming that the sample of 1000 was randomly distributed across these categories) by dividing the number of responses in each category by the number in the sample in each category, to obtain the response rates shown in Table 9.4.

Table 9.3 Population Breakdown by Age and Sex

<table>
<thead>
<tr>
<th>Age---&gt;</th>
<th>15-24</th>
<th>25-34</th>
<th>35-44</th>
<th>45-54</th>
<th>55-64</th>
<th>65+</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>2040</td>
<td>1130</td>
<td>560</td>
<td>430</td>
<td>380</td>
<td>390</td>
<td>4930</td>
</tr>
<tr>
<td>Female</td>
<td>1880</td>
<td>1090</td>
<td>570</td>
<td>470</td>
<td>450</td>
<td>610</td>
<td>5070</td>
</tr>
<tr>
<td>TOTAL</td>
<td>3920</td>
<td>2220</td>
<td>1130</td>
<td>900</td>
<td>830</td>
<td>1000</td>
<td>10000</td>
</tr>
</tbody>
</table>
The population expansion factors are now calculated as the ratio of the number in the population to the number of respondents in each category, as shown in Table 9.5.

### Table 9.5  Population Expansion Factors by Age and Sex

<table>
<thead>
<tr>
<th>Age--&gt;</th>
<th>15-24</th>
<th>25-34</th>
<th>35-44</th>
<th>45-54</th>
<th>55-64</th>
<th>65+</th>
<th>AVE.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>36.4</td>
<td>37.7</td>
<td>40.0</td>
<td>33.1</td>
<td>20.0</td>
<td>19.5</td>
<td>32.5</td>
</tr>
<tr>
<td>Female</td>
<td>22.7</td>
<td>16.8</td>
<td>20.4</td>
<td>23.5</td>
<td>21.4</td>
<td>19.7</td>
<td>20.4</td>
</tr>
</tbody>
</table>

Thus whereas the average expansion factor for the entire sample is 25.0, the individual category expansion factors range from 16.8 to 40.0. When these expansion factors are applied to the number of trips made by respondents in each category (obtained by multiplying the number of respondents by the trip rate in each category), the total number of trips made in the population is found as shown in Table 9.6, yielding an average trip rate in the population of 3.21 trips per day (the difference between the sample trip rate and population trip rate is not substantial in this case, but this would obviously depend on the data set being used).

### Table 9.6  Total Trips in Population by Age and Sex

<table>
<thead>
<tr>
<th>Age--&gt;</th>
<th>15-24</th>
<th>25-34</th>
<th>35-44</th>
<th>45-54</th>
<th>55-64</th>
<th>65+</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>4370</td>
<td>3500</td>
<td>2680</td>
<td>2520</td>
<td>1800</td>
<td>1270</td>
<td>16140</td>
</tr>
<tr>
<td>Female</td>
<td>4760</td>
<td>2080</td>
<td>2000</td>
<td>3150</td>
<td>2210</td>
<td>1710</td>
<td>15910</td>
</tr>
</tbody>
</table>

The above procedure would be used when the data in the secondary source is of a comparable level of detail to that obtained in the survey. However, this is often not the case, and frequently the secondary source data can only be obtained at a more aggregate level. While we would like to know the number in the population in each of the categories, often all we can get is the "marginals"; that is, the total number in each of the rows and columns. Thus in the above example, all we may be able to get is a breakdown by age and a separate breakdown by sex,
but not a breakdown by age and sex together. This secondary data may be represented as shown in Table 9.7.

### Table 9.7 Marginal Population Totals by Age and Sex

<table>
<thead>
<tr>
<th>Age-→</th>
<th>15-24</th>
<th>25-34</th>
<th>35-44</th>
<th>45-54</th>
<th>55-64</th>
<th>65+</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOTAL</td>
<td>3920</td>
<td>2220</td>
<td>1130</td>
<td>900</td>
<td>830</td>
<td>1000</td>
<td>10000</td>
</tr>
</tbody>
</table>

In such cases, it is still possible to calculate population expansion factors, but since we are working with less information, the reliability of these expansion factors will depend on how much extra information is contained in the body of Table 9.7. To calculate expansion factors under these conditions, we need to adopt an iterative procedure where first we obtain agreement with respect to one of the marginals. For example, we can expand the values in Table 9.1, by multiplying each value by the ratio of the marginal total in Table 9.7 to the marginal total in Table 9.1, such that the correct number of males and females are obtained in the expanded total, as shown in Table 9.8.

### Table 9.8 Expanded Population Totals after First Iteration

<table>
<thead>
<tr>
<th>Age-→</th>
<th>15-24</th>
<th>25-34</th>
<th>35-44</th>
<th>45-54</th>
<th>55-64</th>
<th>65+</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>1820</td>
<td>970</td>
<td>450</td>
<td>420</td>
<td>620</td>
<td>650</td>
<td>4930</td>
</tr>
<tr>
<td>Female</td>
<td>1700</td>
<td>1330</td>
<td>570</td>
<td>410</td>
<td>430</td>
<td>630</td>
<td>5070</td>
</tr>
<tr>
<td>TOTAL</td>
<td>3520</td>
<td>2300</td>
<td>1020</td>
<td>830</td>
<td>1050</td>
<td>1280</td>
<td>10000</td>
</tr>
</tbody>
</table>

At this point, while the total number of males and females is correct, the number in each age group is incorrect. It is therefore necessary to perform a second iteration by adjusting the values in the matrix such that the column totals agree with the number in each age group, as shown in Table 9.9.

### Table 9.9 Expanded Population Totals after Second Iteration

<table>
<thead>
<tr>
<th>Age-→</th>
<th>15-24</th>
<th>25-34</th>
<th>35-44</th>
<th>45-54</th>
<th>55-64</th>
<th>65+</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>2030</td>
<td>940</td>
<td>500</td>
<td>460</td>
<td>490</td>
<td>510</td>
<td>4930</td>
</tr>
<tr>
<td>Female</td>
<td>1890</td>
<td>1280</td>
<td>630</td>
<td>440</td>
<td>340</td>
<td>490</td>
<td>5070</td>
</tr>
<tr>
<td>TOTAL</td>
<td>3920</td>
<td>2220</td>
<td>1130</td>
<td>900</td>
<td>830</td>
<td>1000</td>
<td>10000</td>
</tr>
</tbody>
</table>

At this point, in this example, the age and sex totals are correct and the iterations can cease. In other situations, however, especially where there are a larger number of control variables on which the sample is being expanded, it may be
necessary to iterate several times before a stable condition is achieved. Heathcote (1983) describes the iteration process in some detail. However, even though stability has been achieved with respect to the marginal totals, there is no guarantee that the values within the matrix in fact agree with the real values in the population. For example, by comparing Tables 9.9 and 9.3, it can be seen that males between the ages of 25 and 34 are under-represented in our expanded population while females in this age group are correspondingly over-represented. This occurs because of the correlation between age and sex in the population (females tend to be older) which is not accounted for in the iterative process based on the marginal totals. At the end of the iterative process, population expansion factors may be calculated as the ratios of the estimated totals in the population to the number of respondents in each category, as shown in Table 9.10.

**Table 9.10 Estimated Population Expansion Factors by Age and Sex**

<table>
<thead>
<tr>
<th>Age--&gt;</th>
<th>15-24</th>
<th>25-34</th>
<th>35-44</th>
<th>45-54</th>
<th>55-64</th>
<th>65+</th>
<th>AVE.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>36.3</td>
<td>31.3</td>
<td>35.7</td>
<td>35.4</td>
<td>25.8</td>
<td>25.5</td>
<td>32.5</td>
</tr>
<tr>
<td>Female</td>
<td>22.8</td>
<td>19.7</td>
<td>22.5</td>
<td>22.0</td>
<td>16.2</td>
<td>15.8</td>
<td>20.4</td>
</tr>
<tr>
<td>AVE.</td>
<td>28.2</td>
<td>23.4</td>
<td>26.9</td>
<td>27.2</td>
<td>20.7</td>
<td>19.6</td>
<td>25.0</td>
</tr>
</tbody>
</table>

These estimated expansion factors may then be applied to the number of trips made by respondents in each category to find the total number of trips made in the population as shown in Table 9.11, yielding an average trip rate in the population of 3.18 trips per day. In this case, the population expansion factors have moved the total number of trips closer to the real number (as given in Table 9.6), but have not provided the correct answer because of the information which was missing from the marginal totals.

**Table 9.11 Total Trips in Population by Age and Sex**

<table>
<thead>
<tr>
<th>Age--&gt;</th>
<th>15-24</th>
<th>25-34</th>
<th>35-44</th>
<th>45-54</th>
<th>55-64</th>
<th>65+</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>4340</td>
<td>2910</td>
<td>2400</td>
<td>2690</td>
<td>2320</td>
<td>1660</td>
<td>16320</td>
</tr>
<tr>
<td>Female</td>
<td>4780</td>
<td>2440</td>
<td>2210</td>
<td>2950</td>
<td>1670</td>
<td>1380</td>
<td>15430</td>
</tr>
<tr>
<td>TOTAL</td>
<td>9120</td>
<td>5350</td>
<td>4610</td>
<td>5640</td>
<td>3990</td>
<td>3040</td>
<td>31750</td>
</tr>
</tbody>
</table>

In addition to the mathematical problems involved in using marginal totals for the estimation of expansion factors, there are a number of other practical issues which need to be resolved. First, one has to find a good source of secondary data which hopefully will provide the control variables in a cross-tabulated fashion (and not just in marginal total fashion). Second, the data in the secondary source should have been collected in a similar manner to the survey currently being conducted. In particular, the coding categories used should be similar between
Weighting & Expansion of Data

the two data sets. Common definitions of items such as occupation, employment status and housing type should be used (this may involve a compromise on your part if the secondary data, such as the Census, has already been collected). Third, there may be problems with the timeliness of the secondary data becoming available. For example, the Census typically takes about two years from the time of data collection before the first results are available. Even then, these results are generally very aggregate in nature and may not be suitable for the purposes of calculating population expansion factors.

As a general rule, the design of the procedures for expansion of the data should be performed very early in the design process, since the availability of secondary data may often affect the choice and wording of questions on the survey.

9.2 CORRECTIONS FOR NON-REPORTED DATA

Non-reporting refers to the incompleteness of information in questionnaires that were returned. This incompleteness can refer either to questions or parts of questions which were answered incorrectly or incompletely, or to information which was not supplied at all. In the context of travel surveys, this non-reporting phenomenon is of particular importance in the non-reporting of trips and trip characteristics since conclusions about trip volumes (by mode) and general trip making behaviour and characteristics are the focus of travel surveys.

A reason for non-reporting of trips and trip characteristics can be simple memory lapses, especially when the respondent is asked to recall trips made over a significant period in the past. But even in short-term recollection, trips are frequently forgotten or misrepresented. Another reason for non-reporting can lie in the conviction by the respondent that a trip was not "important", or it was too short, or it was performed on foot or by bicycle. Proper instructions about trip definitions and reporting requirements can reduce this source of non-reporting. In certain situations it is possible that a respondent is unwilling to disclose all trips because of an embarrassing trip purpose or destination. Very little, however, can be done to overcome this latter problem.

The problem of incomplete information has been studied within the context of the "KONTIV"("Kontinuierliche Erhebung des Verkehrsverhaltens - A Continuous Survey of Travel Behaviour) survey design in West Germany by Brög, Erl, Meyburg and Wermuth (1982) and Wermuth (1985a). The KONTIV design is a highly refined self-administered survey developed by Werner Brög (reported in Brög et al., 1983) with a format similar to that shown in Figure 5.2. As such, it would be expected that the problem of incomplete information would be at a minimum compared to other less well-designed surveys. Nonetheless, the patterns of incomplete information are useful diagnostic information for the design of other surveys. Table 9.12 presents data on the percentage of responses
Chapter 9

for various questions for which there was incomplete information. These results are presented in two ways; the raw percentage of incomplete information on the survey form and the percentage incomplete after the coder had made any possible corrections.

Table 9.12 Incomplete Information for Various Question Types

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>On Survey Form</th>
<th>After Corrections</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>3.0%</td>
<td>2.5%</td>
</tr>
<tr>
<td>Age</td>
<td>4.5%</td>
<td>4.5%</td>
</tr>
<tr>
<td>Marital Status</td>
<td>3.5%</td>
<td>3.0%</td>
</tr>
<tr>
<td>Education Level</td>
<td>9.4%</td>
<td>3.1%</td>
</tr>
<tr>
<td>Employment Status</td>
<td>7.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Occupation</td>
<td>9.5%</td>
<td>2.0%</td>
</tr>
<tr>
<td>Drivers Licence</td>
<td>9.5%</td>
<td>4.5%</td>
</tr>
<tr>
<td>Trip Details:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Destination Address</td>
<td>15.0%</td>
<td>11.0%</td>
</tr>
<tr>
<td>Trip Purpose</td>
<td>10.0%</td>
<td>6.0%</td>
</tr>
<tr>
<td>Travel Mode</td>
<td>54.7%</td>
<td>51.3%</td>
</tr>
<tr>
<td>Travel Time</td>
<td>20.3%</td>
<td>12.3%</td>
</tr>
</tbody>
</table>

It can be seen that, initially, the extent of incomplete information on demographic and trip questions is in the range of 5 to 10%, but after coding and office editing this can be reduced to less than 5%. With respect to incomplete trip details, the major type of omission was with respect to travel mode. Wermuth (1985a) also shows that the extent of incomplete information for trip details varies with the trip purpose and travel mode, as shown in Table 9.13. It can be seen that shopping trips and recreational trips are most likely to have incomplete information, both before and after coder corrections, while non-motorised trips are more likely to be incompletely specified.

Table 9.13 Incomplete Information for Various Question Types

<table>
<thead>
<tr>
<th>Trip Purpose:</th>
<th>On Survey Form</th>
<th>After Corrections</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work</td>
<td>21.8%</td>
<td>9.8%</td>
</tr>
<tr>
<td>School</td>
<td>37.0%</td>
<td>6.8%</td>
</tr>
<tr>
<td>Shopping</td>
<td>60.3%</td>
<td>31.8%</td>
</tr>
<tr>
<td>Other Discretionary trips</td>
<td>30.8%</td>
<td>9.6%</td>
</tr>
<tr>
<td>Recreation</td>
<td>40.4%</td>
<td>13.7%</td>
</tr>
<tr>
<td>Return Home Trips</td>
<td>8.9%</td>
<td>0.7%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Travel Mode:</th>
<th>On Survey Form</th>
<th>After Corrections</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Motorised</td>
<td>28.6%</td>
<td>10.8%</td>
</tr>
<tr>
<td>Motorised</td>
<td>24.2%</td>
<td>9.6%</td>
</tr>
<tr>
<td>Public Transport</td>
<td>30.3%</td>
<td>7.9%</td>
</tr>
</tbody>
</table>

It is also possible to relate the extent of incomplete information to the type of respondent supplying the information, as can be seen in Table 9.14. Thus, the
Weighting & Expansion of Data

Incomplete information increases as the respondent gets older, and tends to decrease as the level of education of the respondent increases.

Table 9.14 Incomplete Information for Various Types of Respondent

<table>
<thead>
<tr>
<th>Age (years):</th>
<th>On Survey Form</th>
<th>After Corrections</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-15</td>
<td>31.9%</td>
<td>5.5%</td>
</tr>
<tr>
<td>16-25</td>
<td>22.4%</td>
<td>4.5%</td>
</tr>
<tr>
<td>26-45</td>
<td>25.1%</td>
<td>11.4%</td>
</tr>
<tr>
<td>46-64</td>
<td>26.3%</td>
<td>11.0%</td>
</tr>
<tr>
<td>&gt;65</td>
<td>49.3%</td>
<td>22.3%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Education Level:</th>
<th>On Survey Form</th>
<th>After Corrections</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elementary School</td>
<td>31.5%</td>
<td>12.2%</td>
</tr>
<tr>
<td>High School</td>
<td>22.0%</td>
<td>10.1%</td>
</tr>
<tr>
<td>University</td>
<td>12.5%</td>
<td>4.6%</td>
</tr>
</tbody>
</table>

While the problem of incomplete information is an inconvenience, especially to the coder who has to try to supply the missing information, a more serious problem is the complete non-reporting of trips (this may be seen as an extreme case of incomplete information). Brög, et al., (1982) and Wermuth (1985a) have shown that the extent of non-reported trips can be related to personal characteristics of the respondent and to various characteristics of the missing trips.

Table 9.15 Non-Reported Trips for Various Types of Respondent

<table>
<thead>
<tr>
<th>Age (years):</th>
<th>% Non-Reported Trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-15</td>
<td>27.4%</td>
</tr>
<tr>
<td>16-25</td>
<td>7.4%</td>
</tr>
<tr>
<td>26-45</td>
<td>15.5%</td>
</tr>
<tr>
<td>46-64</td>
<td>16.3%</td>
</tr>
<tr>
<td>&gt;65</td>
<td>20.2%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Education Level:</th>
<th>% Non-Reported Trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elementary School</td>
<td>23.2%</td>
</tr>
<tr>
<td>High School</td>
<td>9.6%</td>
</tr>
<tr>
<td>University</td>
<td>17.8%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Licence Holding:</th>
<th>% Non-Reported Trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>Driver's Licence</td>
<td>11.6%</td>
</tr>
<tr>
<td>No Licence</td>
<td>21.0%</td>
</tr>
</tbody>
</table>

The extent of non-reporting of trips as a function of respondent characteristics is shown in Table 9.15. With the exception of young teenagers, who tend to not report many trips made by car (as a passenger), there is again a tendency for older people to have more non-reported trips. Whether this is a function of memory lapses or is a result of the types of trips they tend to make will be explored later. There appears to be no clear tendency for non-reporting of trips.
to be associated with any education level, but respondents without a driver's licence tend to make more unreported trips.

**Table 9.16 Trip Characteristics of Non-Reported Trips**

<table>
<thead>
<tr>
<th>Trip length (km):</th>
<th>% Non-Reported Trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 0.5</td>
<td>26.5%</td>
</tr>
<tr>
<td>0.5 - 1.0</td>
<td>23.5%</td>
</tr>
<tr>
<td>1.0 - 3.0</td>
<td>13.8%</td>
</tr>
<tr>
<td>3.0 - 5.0</td>
<td>9.5%</td>
</tr>
<tr>
<td>5.0 - 10.0</td>
<td>7.5%</td>
</tr>
<tr>
<td>10.0 - 20.0</td>
<td>7.5%</td>
</tr>
<tr>
<td>&gt;20.0</td>
<td>5.4%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Travel Mode:</th>
<th>% Non-Reported Trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moped, Motorcycle</td>
<td>25.0%</td>
</tr>
<tr>
<td>Walk</td>
<td>22.9%</td>
</tr>
<tr>
<td>Bicycle</td>
<td>14.4%</td>
</tr>
<tr>
<td>Car Passenger</td>
<td>12.3%</td>
</tr>
<tr>
<td>Car Driver</td>
<td>8.9%</td>
</tr>
<tr>
<td>Public Transport</td>
<td>6.7%</td>
</tr>
<tr>
<td>Train</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trip Purpose:</th>
<th>% Non-Reported Trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shopping</td>
<td>18.4%</td>
</tr>
<tr>
<td>Recreation</td>
<td>17.2%</td>
</tr>
<tr>
<td>Other Discretionary Trips</td>
<td>14.5%</td>
</tr>
<tr>
<td>School</td>
<td>8.1%</td>
</tr>
<tr>
<td>Work</td>
<td>5.8%</td>
</tr>
</tbody>
</table>

In addition, the extent of non-reporting of trips appears to be a function of the characteristics of the trips themselves. As shown in Table 9.16, non-reported trips tend to be shorter than average, tend to be by non-motorised means of transportation and also tend to be of a more discretionary nature. As a result of the characteristics of the non-reported trips, the increase in mobility after accounting for these trips varies depending on the measure of mobility used. Thus the proportion of mobiles increases least, the trip rate per mobile increases more, and the trip rate across all people increases most as shown in Table 9.17.

**Table 9.17 Increases in Mobility after Allowing for Non-Reported Trips**

<table>
<thead>
<tr>
<th>Measure of Mobility:</th>
<th>% Increase in Mobility</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Mobiles</td>
<td>4.8%</td>
</tr>
<tr>
<td>Trip Rate per Mobile</td>
<td>10.4%</td>
</tr>
<tr>
<td>Trip Rate per Person</td>
<td>14.2%</td>
</tr>
</tbody>
</table>

The results reported in this section have been confirmed by other studies (e.g. Clarke, Dix and Jones, 1981; Barnard, 1985) and lend credence to the need to at least be aware of, if not explicitly correct for, the effects of non-reported trips when presenting the findings of travel surveys.
Methodological research conducted as part of the South-East Queensland Household Travel Survey (SEQHTS) (Richardson and Ampt, 1993a) has offered further insights into the issue of non-reporting of trips and has suggested a way of correcting for this non-reporting in the expanded data.

In the SEQHTS survey, validation interviews were performed with a sample of the responding households (as described in Chapter 7). The information for the estimation of non-reporting correction factors was obtained by means of identifying all additions made to the stop data as a result of the validation interviews. These added stops were also classified as to whether they were expected or unexpected. Expected extra stops were those where, during data entry (prior to validation), it had been identified that it was likely that an extra stop should have been reported - e.g. a person went to a shop and did not return home. Unexpected stops were those which had not been identified in this way, but which respondents reported during the validation interview checking.

As a result of experience gained in previous pilot surveys, it was decided to examine the characteristics of these added stops in terms of their mode, their purpose, and whether they were the last stop of the day. As in those pilot surveys, it was found that the added stops differed from the originally-reported stops most significantly in terms of their purpose and position in the day. The non-reporting correction factors were calculated by dividing the sum of the original stops, plus the expected added stops, plus the unexpected added stops by the original stops, i.e.

\[
\text{Non-Reporting Correction Factor} = \frac{\text{original stops} + \text{expected added stops} + \text{unexpected added stops}}{\text{original stops}}
\]

The resultant non-reporting correction factors for expected and unexpected stops are shown in Tables 9.18 and 9.19.
### Table 9.18 Non-Reporting Correction Factors for Expected Added Stops

<table>
<thead>
<tr>
<th>Destination Purpose</th>
<th>Last Stop of Day?</th>
<th>NO</th>
<th>YES</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change Mode</td>
<td></td>
<td>1.015</td>
<td>1.000</td>
<td>1.015</td>
</tr>
<tr>
<td>Pick Someone Up</td>
<td></td>
<td>1.012</td>
<td>1.000</td>
<td>1.012</td>
</tr>
<tr>
<td>Drop Someone Off</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Accompany Someone</td>
<td></td>
<td>1.022</td>
<td>1.000</td>
<td>1.022</td>
</tr>
<tr>
<td>Buy Something</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td>1.058</td>
<td>1.000</td>
<td>1.058</td>
</tr>
<tr>
<td>Work-Related</td>
<td></td>
<td>1.004</td>
<td>1.000</td>
<td>1.004</td>
</tr>
<tr>
<td>Go Home</td>
<td></td>
<td>1.021</td>
<td>1.071</td>
<td>1.052</td>
</tr>
<tr>
<td>Any Other</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Personal Business</td>
<td></td>
<td>1.016</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Social/Recreational</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Social/Welfare</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Medical/Dental</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Childcare</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Park/Unpark</td>
<td></td>
<td>1.200</td>
<td>1.000</td>
<td>1.200</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>1.014</td>
<td>1.070</td>
<td>1.024</td>
</tr>
</tbody>
</table>

It can be seen from Table 9.18, that the major impact of the non-reported stop correction factors for expected additions will be on trips home at the end of the day, which are frequently forgotten but often easy to detect. From Table 9.19, it can be seen that the major impact of the non-reported stop correction factors for unexpected additions will be on "change-mode" stops made during the day and trips home at the end of the day. These trips are primarily by walk or public transport modes. The fact that stop purpose and mode are correlated means that the application of these non-reported stop correction factors based on stop purpose will also result in an (upward) adjustment for stops made by walk and public transport during the day.
The non-reported stop weights are then applied in the following fashion:

- any household/person/stop which was phoned or validation-interviewed does not need to have the *expected or unexpected* non-reported stop weights applied (because they would already have been found during the phone or validation interview),

- any household for which the data was judged to be perfect, and hence would not have been phoned, needed to have *unexpected* non-reported stop weights applied (because had they been interviewed, there was a chance that an unexpected stop might have been found); and

- any household which had *expected* errors but which was neither on the list to be validated, nor could it be phoned (because no number was given), would need to have both the *expected and unexpected* weights added.

The procedure, therefore, for application of the non-reported stop weights was:

- if the household had been phone-edited, or was a participant in the validation or non-response interviews, then no non-reported stop weights were applied (this means a value of 1.00 was adopted)

- if the household had not been edited at all, then if they stated that they did not have a phone or they did not say whether they had a phone (either way they definitely could not be phoned) then the expected and
unexpected non-reported stop weights were applied to all stops made by that household

- if the household had not been edited at all, and if they stated that they did have a phone and they provided the phone number, then all stops in made by members of that household would receive only the unexpected non-reported stop weights.

The final sets of non-reported stop weights for households with and without phones are shown in Tables 9.20 and 9.21.

Table 9.20 Non-Reported Stop Weights (phone connected)

<table>
<thead>
<tr>
<th>Destination Purpose</th>
<th>Last Stop of Day?</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NO</td>
<td>YES</td>
<td>Total</td>
</tr>
<tr>
<td>Change Mode</td>
<td>1.072</td>
<td>1.000</td>
<td>1.072</td>
</tr>
<tr>
<td>Pick Someone Up</td>
<td>1.040</td>
<td>1.000</td>
<td>1.040</td>
</tr>
<tr>
<td>Drop Someone Off</td>
<td>1.022</td>
<td>1.000</td>
<td>1.022</td>
</tr>
<tr>
<td>Accompany Someone</td>
<td>1.049</td>
<td>1.000</td>
<td>1.049</td>
</tr>
<tr>
<td>Buy Something</td>
<td>1.006</td>
<td>1.000</td>
<td>1.006</td>
</tr>
<tr>
<td>Education</td>
<td>1.043</td>
<td>1.000</td>
<td>1.043</td>
</tr>
<tr>
<td>Work-Related</td>
<td>1.013</td>
<td>1.000</td>
<td>1.013</td>
</tr>
<tr>
<td>Go Home</td>
<td>1.011</td>
<td>1.086</td>
<td>1.057</td>
</tr>
<tr>
<td>Any Other</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Personal Business</td>
<td>1.020</td>
<td>1.000</td>
<td>1.020</td>
</tr>
<tr>
<td>Social/Recreational</td>
<td>1.025</td>
<td>1.000</td>
<td>1.024</td>
</tr>
<tr>
<td>Social/Welfare</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Medical/Dental</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Childcare</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Park/Unpark</td>
<td>1.449</td>
<td>1.000</td>
<td>1.449</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1.030</td>
<td>1.084</td>
<td>1.040</td>
</tr>
</tbody>
</table>

As with the application of all correction weights, a major conceptual limitation must be acknowledged in the use of non-reporting correction factors. The reason for the application of the non-reporting weights is that some people did not tell us about some of the trips they made. By way of the validation interviews, we determine which are the most likely types of trips not to have been reported. We then multiply those trips of this type, which have been reported, by a correction factor to compensate for the missing trips. In this way, the total number of trips in the population should be more accurately estimated. However, from an individual person viewpoint, we are adding trips to those people who have already told us about their trips, and not adding them to the people who have not told us about all their trips (because multiplying zero by any number still leaves us with zero trips). Therefore, while the total number of trips should be more accurately estimated, the distribution of trips per person will be pushed further away from the real situation. Statistically, we have improving the
estimation of the mean number of trips per person, but artificially increased the variance of the number of trips per person. This occurs because of the use of multiplicative correction factors. To overcome this problem, we would need to develop additive correction factors which add the non-reported trips onto those people who have not told us about all their trips; this however is logically difficult to implement. Therefore, multiplicative correction factors must be used in the realisation that they improve estimates of the mean, but worsen estimates of the variance. However, since estimates of the mean are generally more important, it is better to use some form of multiplicative correction factor than to not use any at all.

### Table 9.21 Non-Reported Stop Weights (phone not connected)

<table>
<thead>
<tr>
<th>Destination Purpose</th>
<th>Last Stop of Day?</th>
<th>NO</th>
<th>YES</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change Mode</td>
<td></td>
<td>1.083</td>
<td>1.000</td>
<td>1.083</td>
</tr>
<tr>
<td>Pick Someone Up</td>
<td></td>
<td>1.049</td>
<td>1.000</td>
<td>1.049</td>
</tr>
<tr>
<td>Drop Someone Off</td>
<td></td>
<td>1.022</td>
<td>1.000</td>
<td>1.022</td>
</tr>
<tr>
<td>Accompany Someone</td>
<td></td>
<td>1.066</td>
<td>1.000</td>
<td>1.066</td>
</tr>
<tr>
<td>Buy Something</td>
<td></td>
<td>1.006</td>
<td>1.000</td>
<td>1.006</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td>1.087</td>
<td>1.000</td>
<td>1.087</td>
</tr>
<tr>
<td>Work-Related</td>
<td></td>
<td>1.016</td>
<td>1.000</td>
<td>1.016</td>
</tr>
<tr>
<td>Go Home</td>
<td></td>
<td>1.027</td>
<td>1.139</td>
<td>1.097</td>
</tr>
<tr>
<td>Any Other</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Personal Business</td>
<td></td>
<td>1.032</td>
<td>1.000</td>
<td>1.032</td>
</tr>
<tr>
<td>Social/Recreational</td>
<td></td>
<td>1.025</td>
<td>1.000</td>
<td>1.024</td>
</tr>
<tr>
<td>Social/Welfare</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Medical/Dental</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Childcare</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Park/Unpark</td>
<td></td>
<td>1.600</td>
<td>1.000</td>
<td>1.600</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>1.040</td>
<td>1.137</td>
<td>1.058</td>
</tr>
</tbody>
</table>

### 9.3 CORRECTIONS FOR NON-RESPONSE

Having corrected for non-reported trips from those people who respond to the survey, it is now necessary to turn attention to people in the sample who do not respond to the questionnaire at all. It is quite easy to think of a number of reasons why a non-response to a survey might occur. In this context it is important to recognise that we can only speak of a true or genuine non-response in a situation in which a response was indeed possible, e.g. the addressee simply did not want to respond or was out of town at the time of the survey. Quite a different situation exists where a response was not even possible, e.g. the addressee was deceased, the survey was sent to a non-existing address. In this case we have what is often called "sample loss". Wermuth (1985b) provides data indicating the reasons for non-response to two self-administered, mail-back...
questionnaire surveys conducted in West Germany in 1981. He calls sample loss “non-genuine non-response” to distinguish it from “genuine non-response”. Table 9.22 shows the results of these analyses of non-response.

Table 9.22 Reasons for Non-Response in Self-Administered Surveys

<table>
<thead>
<tr>
<th>Reasons:</th>
<th>Number of Households</th>
<th>Survey #1</th>
<th>Survey #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>GROSS SAMPLE</td>
<td></td>
<td>5039</td>
<td>7688</td>
</tr>
<tr>
<td>SAMPLE LOSS</td>
<td></td>
<td>370</td>
<td>603</td>
</tr>
<tr>
<td>addressee deceased</td>
<td></td>
<td>24</td>
<td>40</td>
</tr>
<tr>
<td>household moved</td>
<td></td>
<td>150</td>
<td>359</td>
</tr>
<tr>
<td>addressee unknown</td>
<td></td>
<td>172</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td></td>
<td>24</td>
<td>204</td>
</tr>
<tr>
<td>NET SAMPLE</td>
<td></td>
<td>4669</td>
<td>7085</td>
</tr>
<tr>
<td>Genuine non-response</td>
<td></td>
<td>1710</td>
<td>2677</td>
</tr>
<tr>
<td>Reasons (as far as known):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objective non-responses</td>
<td></td>
<td>147</td>
<td>279</td>
</tr>
<tr>
<td>- too old</td>
<td></td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>- ill</td>
<td></td>
<td>36</td>
<td>206</td>
</tr>
<tr>
<td>- out of town</td>
<td></td>
<td>40</td>
<td>73</td>
</tr>
<tr>
<td>Subjective non-responses</td>
<td></td>
<td>249</td>
<td>437</td>
</tr>
<tr>
<td>- non-acceptance of questionnaire</td>
<td></td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>- answer refused</td>
<td></td>
<td>183</td>
<td>247</td>
</tr>
<tr>
<td>- lack of time, other</td>
<td></td>
<td>9</td>
<td>190</td>
</tr>
<tr>
<td>Genuine non-responses (with known reasons)</td>
<td></td>
<td>396</td>
<td>716</td>
</tr>
<tr>
<td>Respondents</td>
<td></td>
<td>2959</td>
<td>4408</td>
</tr>
<tr>
<td>Household response rate</td>
<td></td>
<td>63.4%</td>
<td>62.2%</td>
</tr>
</tbody>
</table>

It is important to understand the way in which response rate is calculated. This has been shown in Section 7.1.7. From the gross sample size is subtracted those members of the sample from whom a response could not possibly be obtained. These forms of sample loss (i.e. invalid households, such as vacant or demolished dwellings) do not affect the quality of the sample, and are sometimes said to be quality neutral. The resultant number is the net sample size. The response rate is then calculated as the ratio of the number of respondents as a percentage of this net sample size.

While the exact composition of non-responses will vary with the survey and the population, it is clear that the sample loss is directly related to the quality of the sampling frame from which the potential respondents were sampled. The more dated and inadequate the sampling frame, the more likely that there will be an undesirably large number of sample losses.

The two basic concerns with respect to non-response that need to be stressed are the importance of recognising the existence of non-response and of the need to find ways of assessing its impact on the quality, representativeness and reliability of the information derived from the survey. The analyst has to answer satisfactorily the questions as to whether the results of the survey would have been the same even if a one hundred percent response rate had been achieved.
Weighting & Expansion of Data

This question translates into the recommendation that the analyst try to establish some information about the non-respondents that will permit judgment about whether the information that could have been obtained from the non-respondents would have been statistically different from that actually collected.

Ideally, it would be desirable to have available a series of adjustment factors that could be applied for different surveys and population groups in order to account for the information lost through non-response. Unfortunately, these adjustment factors can only be obtained through significant survey research efforts into the characteristics of "typical" non-respondents. As the reader can well imagine, follow-up surveys to investigate the reasons for non-response and to establish appropriate adjustment factors are costly and time-consuming. Since survey budgets generally tend to be very tight, it is virtually impossible to advance the state-of-the-art of adjustments for non-response through regular survey activities. Separately funded and carefully staffed research efforts are necessary to achieve significant and analytically sound advancements in this area. On the other hand, it has been shown through the limited research efforts that exist in this area (e.g., Brög and Meyburg 1980, 1981, 1982; Wermuth 1985b; Richardson and Ampt, 1993a, Richardson and Ampt, 1994) that an understanding of non-response effects can lead to significantly more accurate and representative survey results.

Moser and Kalton (1971) identify five general sources of non-response:

(a) No longer at available address ("movers")
(b) Physical (health-related) inability to respond
(c) Refusals
(d) Away from home during survey period
(e) Out at time of call

Several strategies have been proposed to compensate for people whose addresses have changed since the sampling frame was prepared. One approach is to substitute for the moved household the new household that has moved to that address (if it is the household address and not the specific residents that are the sampling unit). Another strategy could be to try to "pursue" the household to its new address and to obtain a response at that location (if the identity of the specific residents is important to maintain). A third strategy is to determine the number of households that have moved out of the survey area during the \( m \) months preceding the survey and to double the weight of an equal number of respondents who have moved into the area during that same time period. In this way, the movers-in are included in the sample on their own behalf and also in place of the movers-out (Gray, et al., 1950).
Only the last four reasons for non-response are of major interest to the analyst because the first reason could be considered as falling into the category of sample loss, i.e. they are out of the analyst’s control once the survey sample has been drawn. It is the segment of the non-respondents that legitimately belong in the sample that is of particular interest to the analyst because, under these conditions, carefully designed survey procedures can help reduce the problem. Very little, if anything, can be done about correcting for the non-response in the second category, neither in mail-back self-administered surveys nor in home interview surveys. However, we ought to keep in mind with respect to the other reasons that non-response is a relative term. It depends very much on the surveyor’s level of perseverance, quite aside from the quality of the overall survey design and administration. For example, in mail-back surveys, it would be very unwise to omit follow-up reminders and to be satisfied with whatever is returned in the first wave (i.e. after the questionnaire has first been distributed). The use of reminders can significantly increase the number of respondents, as shown in Chapter 7, and as demonstrated in Figure 9.1.

The results in Figure 9.1 are based on surveys conducted in West Germany (Wermuth, 1985b). It can be seen that in all three sets of survey data, the response rate increased significantly with the use of reminders. In the surveys carried out in three West German cities, a very extensive system of reminders was used, consisting of the following steps:

(a) First announcement of survey by postcard
(b) First mailing of questionnaires (two weeks later)
(c) First reminder (postcard, one week later)
(d) Second reminder (postcard, one week later)
(e) Second mailing of questionnaires (one week later)
(f) Third reminder (postcard, one week later)
(g) Third mailing of questionnaires (one week later)
(h) Fourth reminder (postcard, one week later)
(i) Fifth reminder (postcard, one week later)
In the survey covering nine cities, only steps (a), (b), (c), (d) and (e) were implemented, while in the Munich survey only steps (a), (b), (c) and (e) were implemented. Several points arise from consideration of Figure 9.1. First, if each of the surveys had omitted all reminders then a response rate of only 30 to 35% would have been obtained. This, coincidentally, is the response rate often quoted for self-administered surveys. The use of the reminders, however, increased the response rates to over 60% for all surveys. Secondly, it appears that only two mailings and reminders are needed. While further reminders do increase the response rate, they do so only marginally and are probably not very cost effective. Thirdly, the results are remarkably consistent over all of the surveys.

A similar result has been obtained in the SEQHTS survey (Richardson and Ampt, 1993a) with a variation on this program of reminders, as shown in Figure 9.2.
In addition to the responses being stimulated by the use of reminders, common sense alone tells us that the early respondents are likely to be different "from the rest of us" in that they might have a particular interest in the topic of the survey, or they might have plenty of time available to sit down to respond to a survey very promptly. It is conceivable, for example, that a disproportionate percentage of retired people are among the respondents of the "first wave". Wermuth (1985b) investigated the socio-economic status of respondents in the various response groups, and found the results which are summarised in Figures 9.3 through 9.5.
Figure 9.3 The Effect of Household Size on Level and Speed of Response
(Source: Wermuth, 1985b)
Figure 9.4 The Effect of Age on Level and Speed of Response

Figure 9.5 The Effect of Employment and Sex on Speed of Response
Weighting & Expansion of Data

In both the Munich survey and the "three cities" surveys, larger households were more likely to respond and to respond earlier, probably because of the increased chance of finding someone in the household willing to complete the survey. Older people are more likely to respond, probably because of their greater amounts of free time. Employed people are more likely to respond, probably because of their greater extent of trip making and hence the greater perceived relevance of the travel survey. There appears, however, to be no difference in response between males and females.

Similar results were found in the SEQHTS survey (Richardson and Ampt, 1993a) as shown in Tables 9.23 and 9.24.

Table 9.23 Household Characteristics of SEQHTS Respondents by Wave

<table>
<thead>
<tr>
<th>HOUSEHOLD SIZE</th>
<th>RESPONSE WAVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>61%</td>
</tr>
<tr>
<td>2</td>
<td>64%</td>
</tr>
<tr>
<td>3</td>
<td>59%</td>
</tr>
<tr>
<td>4</td>
<td>63%</td>
</tr>
<tr>
<td>5</td>
<td>56%</td>
</tr>
<tr>
<td>6</td>
<td>53%</td>
</tr>
<tr>
<td>7</td>
<td>54%</td>
</tr>
<tr>
<td>8</td>
<td>63%</td>
</tr>
</tbody>
</table>

Since household size, employment status, age and time availability are likely to have an impact on trip-making characteristics, it is reasonable to assume that the travel characteristics and data for non-respondents will be different from that for the respondents. Brög and Meyburg (1980, 1981, 1982) have demonstrated that trip-making characteristics do change substantially as additional response waves due to reminders are recorded. For example, for the nine cities survey, Figure 9.6 shows that the trip frequency and the proportion of mobile persons in the population (i.e. people who make at least one trip) both decrease as the time to respond increases. Thus mail-back questionnaire surveys which do not include follow-up reminders would tend to over-estimate trip-making because of the higher mobility of the early respondents. This should be contrasted with home interview or telephone interview surveys where the early respondents tend to be the "stay-at-homes" who generally have lower than average trip rates. Thus personal interview surveys without call-backs would tend to under-estimate trip rates.
Chapter 9

Table 9.24 Personal Characteristics of SEQHTS Respondents by Wave

<table>
<thead>
<tr>
<th>AGE GROUP</th>
<th>RESPONSE</th>
<th>WAVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 -&gt; 4</td>
<td>58%</td>
<td>19%</td>
</tr>
<tr>
<td>5 -&gt; 9</td>
<td>60%</td>
<td>18%</td>
</tr>
<tr>
<td>10 -&gt; 14</td>
<td>61%</td>
<td>19%</td>
</tr>
<tr>
<td>15 -&gt; 19</td>
<td>60%</td>
<td>21%</td>
</tr>
<tr>
<td>20 -&gt; 24</td>
<td>52%</td>
<td>22%</td>
</tr>
<tr>
<td>25 -&gt; 29</td>
<td>54%</td>
<td>21%</td>
</tr>
<tr>
<td>30 -&gt; 34</td>
<td>60%</td>
<td>19%</td>
</tr>
<tr>
<td>35 -&gt; 39</td>
<td>61%</td>
<td>19%</td>
</tr>
<tr>
<td>40 -&gt; 44</td>
<td>61%</td>
<td>18%</td>
</tr>
<tr>
<td>45 -&gt; 49</td>
<td>64%</td>
<td>17%</td>
</tr>
<tr>
<td>50 -&gt; 54</td>
<td>61%</td>
<td>21%</td>
</tr>
<tr>
<td>55 -&gt; 59</td>
<td>71%</td>
<td>17%</td>
</tr>
<tr>
<td>60 -&gt; 64</td>
<td>75%</td>
<td>14%</td>
</tr>
<tr>
<td>65 -&gt; 69</td>
<td>76%</td>
<td>11%</td>
</tr>
<tr>
<td>70 -&gt; 74</td>
<td>74%</td>
<td>14%</td>
</tr>
<tr>
<td>75+</td>
<td>70%</td>
<td>16%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SEX</th>
<th>RESPONSE</th>
<th>WAVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>60%</td>
<td>19%</td>
</tr>
<tr>
<td>Female</td>
<td>61%</td>
<td>18%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ACTIVITY STATUS</th>
<th>RESPONSE</th>
<th>WAVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full-Time Employment</td>
<td>58%</td>
<td>21%</td>
</tr>
<tr>
<td>Part-Time Employment</td>
<td>62%</td>
<td>18%</td>
</tr>
<tr>
<td>Primary School</td>
<td>60%</td>
<td>18%</td>
</tr>
<tr>
<td>Secondary School</td>
<td>60%</td>
<td>20%</td>
</tr>
<tr>
<td>Tertiary College</td>
<td>59%</td>
<td>24%</td>
</tr>
<tr>
<td>Not yet at School</td>
<td>59%</td>
<td>19%</td>
</tr>
<tr>
<td>Pre-School</td>
<td>58%</td>
<td>20%</td>
</tr>
<tr>
<td>Childcare</td>
<td>53%</td>
<td>22%</td>
</tr>
<tr>
<td>Keeping House</td>
<td>63%</td>
<td>17%</td>
</tr>
<tr>
<td>Currently Unemployed</td>
<td>59%</td>
<td>19%</td>
</tr>
<tr>
<td>Retired or Pensioner</td>
<td>73%</td>
<td>14%</td>
</tr>
<tr>
<td>Other Pensioner</td>
<td>67%</td>
<td>12%</td>
</tr>
<tr>
<td>Other</td>
<td>77%</td>
<td>17%</td>
</tr>
</tbody>
</table>

The observation that trip rate declines with increasing time to respond to the survey has been interpreted as meaning that the trip-making characteristics of late respondents are different to that of earlier respondents. However, before this interpretation can be accepted, we need to account for two other possible explanations. Firstly, it could be that late respondents simply belong to different socio-demographic groups to early respondents and that, while they do make fewer trips, they make no fewer trips than early respondents in the same socio-demographic group. It has been shown above that the socio-demographic characteristics of early and late respondents are indeed different, and therefore
socio-demographic expansion will tend to partially correct for the non-response problem.

Figure 9.6 Travel Characteristics as a Function of Response Speed

Second, while the observed (i.e. reported) trip rates are lower for late respondents, it may be that they don't make fewer trips but simply report fewer trips (i.e. they have a higher non-reporting rate than early respondents). For the above reasons, it is first necessary to correct reported trip rates in each response wave for socio-demographic and non-reporting differences as described in the previous two sections of this Chapter.

Ideally, it would then be desirable to have available a series of adjustment factors that could be applied for different surveys and population groups in order to account for the information lost through non-response. These adjustment factors can only be obtained through significant research efforts into the characteristics of “typical” non-respondents. This was attempted in the SEQHTS survey by the follow-up non-response surveys in that study, and the use if these surveys will be described later in this section.

In the SEQHTS survey, responses were classified into six groups according to the time taken for them to respond. Those who respond within 7 days (1 week) of their original travel day are classified as Wave 1 respondents. Those who respond within 2 weeks of their original travel day are classified as Wave 2 respondents, up to Wave 4 respondents. Those responding of their own volition after 4 weeks are classified as Wave 5 respondents. Those who respond as a result of the non-response interviews are classified as Wave 6 respondents. These waves of
respondents are often called "Brögian Response Waves", after Werner Brög who did much of the early research on the calculation of corrections for non-response effects in postal travel surveys.

Calculation of the average, and upper and lower percentiles, of the number of stops per person per day for respondents in each of the response waves gave rise to the curve shown in Figure 9.7.

![Figure 9.7 Average Stop Rate as a Function of Response Wave](image)

It can be seen from Figure 9.7 that the average number of stops per day decreases as the time taken to respond increases. Thus the respondents in the first wave have the highest stop rate and those in the last waves have the lowest stop rates. This trend is consistent with that found in previous work (Wermuth, 1985b).

A better picture of this trend can be obtained by considering the number of responding households in each of these waves, and the cumulative percentage of responses. Thus in the SEQHTS survey, 60% of the total respondents (weighted for demographic characteristics) responded within one week of the initial travel date, another 19% responded in the week after this and so on. By the end of week two, a cumulative total of 79% of the total respondents had responded. The 6th response wave consists of those respondents who were obtained from the non-response interviews. They represent more than themselves, however, since only a sample of non-responding households were included in the non-response interview sample. The non-response sample was found to consist of three sub-
Weighting & Expansion of Data

groups; those who agreed to respond, those who refused to respond to the survey, and those who did not respond for other (perhaps travel-related) reasons. The second group have been referred to as "stubborn" non-respondents; Wermuth (1985b) has noted that approximately 10% of the net sample are stubborn non-respondents. The SEQHTS survey found that approximately 11% of the net Brisbane sample fell into this category. Therefore, it would never be possible to get more than 89% response from the net sample, and this value has been used as the upper limit of cumulative response for response wave 6. The relationship between stop rate (stops/person/day) and cumulative percentage of net sample is shown in Figure 9.8.

![Figure 9.8 Stop Rate as a Function of Cumulative Response](image)

**Figure 9.8 Stop Rate as a Function of Cumulative Response**

This relationship (for the first five waves) is very similar to that obtained by Wermuth (1985a) in that the stop rate falls relatively uniformly after the second response wave. Using such a relationship, Brög and Meyburg (1982) have postulated that non-respondents are more likely to have trip-making characteristics like those who respond late to travel surveys than those who respond early to travel surveys. They have assumed a linear decrease in stop rate after the second response wave up till the last respondents to the mailed questionnaire. They then project forward to estimate the likely stop rate of the non-respondents.

In the case of Figure 9.8, a linear relationship is postulated as given by the dashed line overlaid on the response curve. This would give an estimate of approximately 2.65 stops/person/day for the non-respondents. As it happens, in
Chapter 9

the SEQHTS survey, there was the unusual situation of actually having an empirical measurement of the stops/person/day for the non-respondents from the non-response interviews. This actual value was 2.61 stops/person/day. This confirms the overall validity of the approach adopted by Brög and Meyburg (1982) and Wermuth (1985).

Given that these non-respondents have a lower stop rate than the respondents, then it is necessary to apply a correction factor to all observed stops to reduce the estimated population stop rate to account for the lower stop rate of the non-respondents. While it is possible that the reductions in stop rate apply non-uniformly to various types of stop, such differentiation has not yet been attempted; the non-response correction factor is applied equally to all stops. Later research should investigate variations in non-response correction factors by stop purpose, mode of travel etc.

The non-response correction factor is calculated by considering the three major groups in the net sample and the stop rates associated with each group. These three groups are the respondents, the non-respondents and the stubborn non-respondents. In the SEQHTS survey, these groups make up approximately 73%, 16% and 11% of the net sample. The stop rates associated with the first two groups can be found from the data for the waves of respondents and the wave of non-respondents. Thus the average stop rate of the respondents is 4.38 (for respondents in the first 5 waves), and the average stop rate of non-respondents is 2.61. The average stop rate for stubborn non-respondents is assumed to be 4.38 (the same as the respondents, on the assumption that their unwillingness to participate in the survey has nothing to do with their travel behaviour). Thus, the weighted average stop rate of the entire sample is 0.73*4.38 + 0.16*2.61 + 0.11*4.38 = 4.10. Since the average stop rate for the respondents would have been calculated as 4.38, a correction factor of 0.935 (=4.10/4.38) was applied to all stops reported by respondents in order to obtain the correct stop rate for the entire net sample. This weighting factor was applied to all records in the stop file.

To minimise the effect of possible non-response bias, it is therefore good survey practice to send out at least one combination reminder/thank you postcard, followed a week later by another reminder postcard with a new survey form, and, if funds and time permit, followed by a third postcard after another week. This procedure will generate several response waves and it will both reduce the non-response rate and increase the quality and representativeness of the survey results, and also provide information on those respondents to the later reminders who might otherwise have been non-respondents. This information can then be used to investigate any trends in travel characteristics as a function of response speed, which can then be used to infer the travel characteristics of those who remain as non-respondents.
Weighting & Expansion of Data

For personal interview surveys the refusal rate is largely a function of the skill and experience of the interviewer. Of course the subject matter of the survey also plays a significant role in the respondents’ willingness to answer questions on a specific topic. Conflicting results have been reported on the desirability of making interview appointments by prior telephone call or by postcard announcement. Sudman (1967) reported that the number of calls required to complete an interview was reduced from 2.3 to 1.7 calls per completed interview. On the other hand Brunner and Carroll (1967) found that prior telephone appointments had the undesirable effect of reducing the response rate. People will find it easier to refuse cooperation through the relative anonymity of a telephone contact than when confronted in a face-to-face situation at their home.

An interesting approach to dealing with the not-at-home problem in personal interview surveys was developed by Politz and Simmons (1949). The survey population is grouped into strata according to the probability of the interviewer finding people at home on the first call. All calls are assumed to be made during the same period of the day, e.g. in the evening. On the basis of the respondent’s answer to the question of how many of the previous five evenings he/she spent at home, the probability of each respondent being at home on any random evening could be calculated. In order to derive population estimates, the interview results for each stratum should be weighted with the reciprocal of the probability of being at home. Of course, a slight bias is introduced due to the fact that those people who are never at home will not be considered in this procedure. While the Politz-Simmons method makes one-call interviews a palatable survey procedure, it is not clear whether this re-weighting procedure is more efficient in terms of cost and quality of results than the more conventional approach of multiple calls until a successful contact and interview is made. Also, it is not clear whether people are necessarily willing to disclose to the interviewer their typical behaviour about presence or absence at home (because of fears about security).

Finally, it ought to be noted that an “unrepaired” non-response bias might be so serious that it might be worth considering selecting a smaller initial sample and placing all resources on a concentrated effort to obtain a higher response rate. Cochran (1977) and Deming (1953) have indeed taken this position. It is clear that ignoring the effect of non-response is a highly unprofessional and unscientific approach to survey sampling.
10. Data Analysis

Having coded and edited the data, we should now have a "clean" dataset ready for analysis. The analysis phase is often a welcome relief after the tedium of editing the data, in that it allows the investigator to show some creative flair.

Two types of analysis are possible. First, there may be a simple enumerative or exploratory analysis which seeks to explore the contents of a dataset and to describe the dataset in a number of ways (e.g. response rates, means and standard deviations of responses, frequency distributions and cross-classifications). Second, one may proceed to a more complex analysis which seeks to confirm statistical hypotheses and find causal relationships among the variables. This model-building phase is frequently the purpose of many transport-related surveys. In this Chapter, we shall first concentrate on simple exploratory analysis and then describe some of the basic methods of building causal relationships from the data. A more complete description of multivariate analysis and model building may be found in texts such as Stopher and Meyburg (1979).

For the analysis of sample data, two alternatives present themselves. You can write your own computer programs for the analysis of data or else use a commercially available computer package for analysis. Writing your own program is usually applicable for special survey methods where many calculations and transformations are required. However, for most sample
Chapter 10

surveys it is now possible to make use of one of the many statistical analysis programs commonly available. They save a lot of time, they present results in neat easily-readable format and they don't require a great deal of computer expertise (although knowledge of the underlying statistical techniques is essential). The one danger with using some statistical packages is that it is often tempting to run all sorts of statistical analyses just to see what eventuates. Such "fishing trips" should generally be avoided since, apart from being inconsistent with the objectives of scientific social surveys, they can turn out to be very expensive and can lead to simplistic and misleading conclusions. As mentioned much earlier, the type of analysis to be performed should be known before the survey is designed.

The choice of which package to use will often depend on the computer facilities available. In the past, much of the analysis of travel survey data was performed on mainframe or minicomputers (because of the needs for substantial data storage and rapid computation speed). Under these conditions, the SPSS (Statistical Package for the Social Sciences) package was probably the most widely used package. It is relatively easy to use (even though it is rather inefficient, computationally, in its operation). No computer programming expertise is required as it is only necessary to prepare a few commands describing the datset and then tell SPSS what analysis to perform. SPSS can be used both to edit the data and to carry out final statistical analysis of the data. Full details of the SPSS package can be found in Nie et al., (1975). Apart from SPSS, a number of other statistical packages are available for mainframe computers and minicomputers. The most well-known packages include SAS and IMSL.

In recent years, however, the trend has been away from the use of mainframe and minicomputers and towards the use of microcomputers and workstations (as exemplified by the IBM and Apple Macintosh range of microcomputers). The use of micros make available a vast array of analytical software previously not available, and enables small agencies to conduct and analyse their own surveys. While mainframes and minicomputers may still be desirable (because of computational speed) for very large survey databases, the availability, ease of use and graphics capabilities of modern microcomputers makes them the system of choice for most survey tasks. As noted in Chapter 5, the micro can also be used in other aspects of survey design such as the graphic design of survey forms.

In using microcomputers for the analysis of survey data, there are three major types of program which are useful, depending on the size of the survey and the type of analysis to be performed. For data entry and editing, spreadsheets and database manager programs are most useful, as described in Chapter 8. These programs can also be useful for some forms of data analysis which do not require the use of complex statistical procedures. When attempting to perform
Data Analysis

statistical analysis on the data, however, it is better to use a dedicated statistical package (many of which have some spreadsheet and database capabilities). There are essentially two types of statistical package which are useful in data analysis; "exploratory" statistical packages and "confirmatory" statistical packages. Exploratory statistical analysis is most useful when you are trying to understand the data and to get an intuitive feel for what the data are trying to tell you. Confirmatory statistical analysis is most useful when you are trying to test whether the data supports a preconceived hypothesis (which may have been obtained from the exploratory analysis, or which you may have had before conducting the survey). Confirmatory statistical analysis can take the form of statistical hypothesis testing or the construction of multivariate causal models. Both types of analysis will be discussed in this Chapter.

10.1 EXPLORATORY DATA ANALYSIS

The concept of exploratory data analysis is best exemplified by the works of Tukey (1977) who coined the term "exploratory data analysis" and who drew the analogy between data analysis and "detective work". To quote Tukey:

"A detective investigating a crime needs both tools and understanding. If he has no fingerprint powder, he will fail to find fingerprints on most surfaces. If he does not understand where the criminal is likely to have put his fingers, he will not look in the right places. Equally, the analyst of data needs both tools and understanding."

Tukey’s tools were largely derived for manual use; he proposed a number of innovative graphical techniques for representing data which could be done manually on relatively small datasets, using "graph paper and tracing paper when you can get it, backs of envelopes if necessary". Tukey was openly sceptical of these tasks being replaced by computerised methods, mainly because of the inaccessibility of computers to the average analyst (in the mid-1970’s). He could not have foreseen the emergence and widespread adoption of microcomputers in the 1980’s, although he did foresee one development in this area; "But there is hope. Paul Velleman has been putting together facilities adapted to the use of such techniques in research...". Ten years later, Velleman is the author of an innovative data analysis program for microcomputers which utilises many of Tukey’s concepts. This program, Data Desk, utilises the interactive graphics capabilities of the Macintosh to enable analysts to be "detectives" in true Tukey fashion.

Within this section, we shall describe several of Tukey’s methods of exploring data, and we shall use Data Desk to show how these ideas can be used in practice. In particular, we will perform analysis on the flat file data file described in Chapter 8. Remember that we had used FoxBASE to extract a subset of data from
the trip and person data files and combine them in a single flat file, as shown in Figure 10.1.

![Figure 10.1 A Flat File Created for Statistical Analysis](image)

To use these data for statistical analysis, it is necessary to transfer them to a statistical analysis program (in this case, Data Desk). Generally, each program stores the data in a form which is unique to that program. However, one can almost always store data from any of these programs in a text file, and one can almost always import data from a text file into any one of these programs. This, in fact, is how the transfer was made from FoxBASE to Data Desk. When imported into Data Desk and supplied with the names of the variables, the data appears in Data Desk as shown in Figure 10.2.

![Figure 10.2 Data Desk's Representation of Variables as Icons](image)
Each variable appears on-screen as a separate column of data. When these icons are opened (by double-clicking on them), the data values appear as shown in Figure 10.3 for the "trips" and "vehicles" variables.

Although the data values still appear in different windows, they are in fact part of the same larger data file. When a record is selected in one of the variable windows, the corresponding value in the other variable window is also selected, as shown in Figure 10.3. Thus the seventh person in the data file has made 7 trips and belongs to a household which owns 6 vehicles.

The variable icons may be re-ordered on the screen simply by dragging them with the mouse to the position required in the group of variable icons.

Many options are available for examination of the raw data, and only a few of them can be covered here. For a more complete coverage, see Tukey (1977), Velleman and Hoaglin (1981), and Velleman and Velleman (1988). The simplest way to examine data is to look at it, and to calculate some simple statistics. The most conventional way to look at the data is by means of a histogram as shown in Figure 10.4.
This histogram may also be summarised statistically by means of a number of summary statistics such as the mean, standard deviation, median, mode (50th!%ile) as shown in Figure 10.5.

Several of these statistics may also be summarised graphically by means of a boxplot (also known as a "box and whiskers") as shown in Figure 10.6. Examining the data in this way is very useful for detecting outliers in the data, and for determining whether these data are genuine or whether they are mistakes. By searching in the variable list for the outlier, one can then identify the person and household to which it belongs (this is the reason such identifiers are always attached to a dataset for statistical analysis), and then examine the questionnaire or interview schedule to determine the validity of the data point.
In this boxplot, the box represents the middle half of the data between the 25th and 75th percentiles. The line in the middle of the box is the median. The whiskers extend beyond the box to the data points which are not further away from the box than 1.5 times the depth of the box. Points between here and 3 times the depth are circled, while points further away are marked with a "starburst".

Figure 10.6  A Boxplot of the Number of Trips per Person per Day

Once this one-way analysis has been conducted for each variable, it is then useful to examine the variables two-at-a-time to see, first, if there are any logical inconsistencies that have not been discovered by the previous editing routines, and secondly to see whether any causal or correlative relationships appear to exist between the two variables. For example, it might be expected that there may be a relationship between the number of trips per person per day and the number of vehicles owned by that persons household. To test this, both variable icons can be selected on the screen, and then a menu item selected to create a scatterplot as shown in Figure 10.7.

At first glance, this plot may appear to suggest that there is not likely to be a relationship between the two variables. However, it should be remembered that there are 904 data values in the data file and only 94 points appearing on the scatterplot. Therefore, on average, each plotted point represents about 10 data values. It is also unlikely that the data values will be evenly distributed across all the visible data points. There may well be a relationship "lurking" beneath the surface which only further exploration will uncover. One way of trying to uncover this relationship is to place the data values into groups based on vehicle ownership and then see whether there is any systematic variation in trips per day between the groups. This can be done by selecting both the variable icons, and then selecting a menu item which will perform this grouping, as shown in Figure 10.8.
Having selected the option shown in Figure 10.8, Data Desk will now split the
data values for the first variable (no. of trips) into groups based on the second
variable (vehicles) and will create a new variable icon for each of these groups.
These data groups can then be analysed individually or in concert. For example, if
all the data groups are selected and the Boxplot menu item is then selected, a
combined set of boxplots for each of the groups will be produced as shown in
Figure 10.9. When presented in this fashion, it can be seen that there may be a
relationship between the two variables because of the way that the "boxes" and
the medians move upward as the number of vehicles increases.

![Figure 10.7 A Scatterplot of Trip Rate vs Household Vehicles](image)

![Figure 10.8 Splitting the Data into Groups Based on Vehicle Ownership](image)
Data Analysis

Figure 10.9  Boxplots of Trips per Day for the Vehicle Ownership Groups

This trend can be even more clearly established, if the mean trip rate within each of these groups is plotted against the number of vehicles as shown in Figure 10.10.

Figure 10.10  Scatterplot of Person trips vs Household Vehicles
Chapter 10

From Figure 10.10, it is quite clear that there is a relatively clear relationship between the average number of trips per person per day and the number of vehicles owned by that person’s household.

In going beyond relationships between two variables, it has traditionally been difficult to investigate such relationships visually. While it is at least theoretically possible to represent the relationship between three variables in three dimensional space, the difficulties of preparing graphs to represent this have been considerable. However, several microcomputer statistical programs, including Data Desk, now have the ability to not only graph the variables in three-dimensional space, but to rotate them at the request of the user, thus enabling you to truly look at the data from many different perspectives. While this feature provides an important tool for exploring data, it is, unfortunately, not possible to present it’s capabilities on the printed page.

10.2 CONFIRMATORY DATA ANALYSIS

The preceding section has dealt with the exploratory analysis of data to obtain an intuitive feel for the data and perhaps to develop some additional hypotheses. This exploratory analysis is usually followed by a more formal confirmatory data analysis, which firstly summarises the data in conventional statistical terms, and may then attempt to relate elements of the data to each other in terms of statistical and causal models. The intention in this section is to provide some of the background knowledge necessary for the development of statistical models, where the use of the term statistical implies the use of sample data to produce some form of mathematical relationship. This section will not provide complete details about the statistical model building procedures discussed; rather it will hope to provide sufficient material at an introductory level to enable you to decide whether such an analysis may be useful to you, and to interpret the output obtained from statistical packages which perform these analyses.

The first thing that should be emphasised is that statistical modelling techniques should not be regarded as a means of pulling some relationship out of thin air. This is a frequent and serious misuse of confirmatory statistical analysis techniques. Rather, the use to which confirmatory statistical modelling techniques should be put is that of attempting to determine the parameters of a theorised relationship (you should use exploratory data analysis techniques if you are just trying to get some ideas about the data).

In other words, a hypothesis (or hypotheses) must be advanced, expressing a relationship that might be expected between certain phenomena (variables). The statistical techniques described in the remaining sections of this chapter represent one means by which certain types of hypotheses may be tested. It is extremely important that the reader understands this idea of hypothesis testing, since it
underlies and explains much of the procedure of statistical model building and testing. This point has been outlined in Chapter 4 of this book, and will be reiterated in specific contexts of certain statistical methods in the succeeding sections of this chapter.

Before detailing the various statistical techniques, it is appropriate to review the reasons for statistical modelling and the general aims of model building. The work "model" is used in the purely statistical sense as invoked in many areas of engineering and applied science. A model may be defined as an abstraction of reality. Specifically, this means that a model is intended to be a simplification of reality, not a replica of reality. In this sense, one may distinguish between a model and a physical law. A physical law is reality, without approximation. Thus Newton's law of gravitation (Newton, 1713) is an exact phenomenological statement (if one accepts the precepts of Newtonian physics). It is the reality.

\[ F = \frac{GM_1M_2}{d^2} \]  

(10.1)

where \( F \) = the gravitational attraction between two bodies  
\( G \) = Newton's gravitational constant  
\( M_1, M_2 \) = the mass of bodies 1 and 2  
and \( d \) = the distance between the two bodies.

The relationship of equation 10.1 (Newton's law of gravitation) is not a model within this definition. It is accurate, precise, and complete statement of a relationship that always holds. To illustrate more clearly just what a model is, one might consider a relationship between the yield of wheat per acre \( (W) \), the rate of fertiliser application \( (F) \), the number of years since wheat was last grown on that land \( (Y) \), and the number of inches of rain since seeding \( (R) \). A simple linear relationship might be:

\[ W = a_0 + a_1F + a_2Y + a_3R \]  

(10.2)

Indeed, this equation may represent the hypothesis to be tested. The questions to be answered by the analyst are: can non-zero values be found for \( a_1, a_2, \) and \( a_3 \), and how does the right side of equation 10.2 predict the observed values of the left side? There is no known law that relates yields of wheat to these three other variables precisely and completely. However, it seems reasonable to expect that some relationship might exist between them. Herein lies the essence of a model. There may be many other variables that will affect the relationship, for example, the quality of the seed, the number of hours of sunshine, the amount of rain and sun at particular growth periods, and the type of soil. Equation 10.2 is, however, an abstraction of reality. It is incomplete, not very precise, and subject to error. But it provides insights into the relationship and may represent a useful relationship for predicting yields of wheat from limited available data.
Chapter 10

It should be clear from this illustration that a number of demands must be made of the process for testing the hypothesis represented by equation 10.2. First, it is necessary to be able to test the null hypothesis that the relationship adds nothing to our knowledge and understanding. This is equivalent to saying that values of \( W \) in equation 10.2 could be predicted by guesswork (i.e., a random process) as accurately as by use of the model. If this hypothesis can be rejected, it is appropriate to determine whether all the variables in the model are necessary for reasonable prediction of \( W \), whether some variables add no new information, and whether other variables are necessary to obtain good predictions. As will be seen, these concerns are raised for each statistical model-building process treated in the remainder of this chapter.

Sufficient has not yet been said about the basic properties of a model. It has been emphasised that a model is an abstraction of reality. It is also true that a model should be as accurate as possible, while yet retaining its simplicity. It may be used as a predictive or forecasting tool, or it may be used to gain understanding of a process. For each of these uses, the probably conflicting goals of accuracy and simplicity are desired. For the purpose of subsequent discussions, simplicity may be interpreted as implying parsimony of variables and the use of the least complex of functional forms, consistent with the hypothesis to be tested. In this later respect, a linear relationship (such as that shown in equation 10.2 and Figure 10.11) is the simplest. A monotonic, continuous non-linear function is next (as shown in Figure 10.12) and various types of discontinuous non-linear functions and continuous non-monotonic functions are next in their degree of complexity.
Figure 10.11 A Simple, Bivariate Linear Relationship

Figure 10.12 A Simple, Bivariate, Monotonic Non-linear Relationship
While retaining these properties of accuracy and simplicity, a model must also be useful for its purpose; that is, if predictive, it must be capable of predicting; if explanatory, it must provide explanation and understanding of the phenomenon being modelled. The model must also be economical to use. It should not need the use of data that are extremely difficult and costly to obtain, and it should be cheaper by far to use than real-world experimentation. A model must also be valid. Many interpretations can be given to validity, although principally a dichotomy of meaning may be made between descriptive and predictive models. If a model is intended to be descriptive only, validity may imply logic in the model structure and transferability to other geographic locations or other modelling situations (depending upon the discipline of study). Validity in a predictive model must also imply causality. In other words, the direction of "affect" must be correctly specified in the model. There are a number of other properties that a model should possess, but these are among the most important and serve adequately to set the stage for the description of the techniques themselves.

This discussion of models should also have pointed out another property possessed by any statistical model; error. Clearly, the preceding statements about accuracy, simplicity, and abstraction of reality all imply the existence of error. Furthermore, models in transport planning are generally built with the use of sample data. As discussed in Chapter 4, sample data possess sampling errors. In fact, there are three primary sources of error in statistical modelling (Alonso, 1968). These are specification error, measurement error, and calibration error. Specification error is generally unknowable and derives from the inclusion of unnecessary variables or the exclusion of necessary variables. While exhaustive testing may eliminate the former, it is never possible to determine the extent of the latter. Measurement error derives from the sampling error and from impreciseness and inaccuracy of the actual measurements made to generate the data. The measurement error that arises from sampling can be computed, as described in Chapter 4, but that arising from imprecision of actual measurement often cannot be determined, particularly when humans are the subjects of the data collection and model building. Calibration error occurs in the modelling process as a result of the use of data containing measurement errors and the building of a model with specification error.

Measurement error is a property of the data and cannot be ameliorated substantially in the modelling process, with the exception of error propagation (i.e., the means by which a model magnifies or diminishes errors in different variables). One of the goals of statistical model building is to minimise error. Specification error is minimised by careful selection of variables in the model through prior exploratory statistical analysis and visual examination of the data, together with a carefully reasoned hypothesis of structure and variable content.
Calibration error is minimised in most statistical techniques as the major principle of model building. The degree to which it can be minimised, however, depends upon a wide range of properties of the data and the model-building technique.

The trade-off between specification and measurement error provides an interesting perspective on the relative effort which should be spent on data collection and modelling in any transport study. Alonso (1968) has postulated that, although specification error cannot be measured exactly, it is likely that specification error will drop rapidly as complexity of a model increases as shown in Figure 10.13 (complexity is defined as being measured by the number of relevant explanatory variables included in the model).

However, each variable included in the model will have a degree of measurement error associated with it, so that the inclusion of more variables will mean that there is a greater total amount of measurement error in the model, as shown in Figure 10.14.

---

**Figure 10.13** Specification Error and Model Complexity
Since the total error in the model includes both specification and measurement error (ignoring calibration error for the moment), then the total error can be obtained by:

$$ E_{\text{total}} = \sqrt{e_{\text{spec}}^2 + e_{\text{meas}}^2} \quad (10.3) $$

This total error is shown in Figure 10.15, wherein it can be seen that, because of the countervailing effects of each source of error, there is in fact an optimum degree of model complexity in order to obtain the minimum total error in the model. The implications of Figure 10.15 are important enough to bear repeating. The best model is not necessarily the most complex model. This is due to the fact that error arises from both specification and measurement error. While a more complex model will reduce the specification error, it will also increase the measurement error. At some point, the inclusion of more variables into the model will increase the measurement error more than it will reduce the specification error.

This trade-off between specification error and measurement error can be further demonstrated by considering the use of a dataset which has a higher degree of measurement error (perhaps because we spent less time and effort on quality control in the data collection process). Under these conditions, the measurement error will be higher at all levels of model complexity, as will be the total error, as shown in Figure 10.16.
The really important feature of Figure 10.16, however, is that apart from simply being higher, the total error curve is minimised at a lower level of model complexity (i.e. the valley in this total error curve is shifted to the left). The implication of this is that if you have worse data, then you should also use simpler models. This finding runs counter to actual modelling practice in many cases, where modellers believe that they can overcome the effects of bad data by
Chapter 10

using more complex models. As shown in Figure 10.16, using more complex models with bad data simply increases the total error in the model.

The policy lesson that comes from consideration of the above arguments is that there must be a balance between the time and effort spent on data collection and data analysis (model building). You cannot compensate for poor quality data by doing better analysis (remember: GIGO, garbage-in/garbage-out). On the other hand, having collected high quality data, you should use more complex methods of analysis to obtain the most information out of this datset if you wish to minimise the total error.

To complete this discussion of errors, it is necessary to consider some specialised terminology. In statistical model building, there are generally two types of variables: dependent and independent. (As discussed later, this taxonomy of variables is not always adequate, particularly in the case of more complex and specialised model-building techniques.) A dependent variable is the variable to be explained or predicted by the model. It therefore depends upon the other variables. Dependency implies a strict one-way causality in almost all cases. A dependent variable is caused to change by changes in the independent variables. The reverse causality, however, must not occur. Changes in the dependent variable cannot cause changes in the independent variables.

The example of equation 10.2 may be used to illustrate this: W, the yield of wheat, is the dependent variable. Its value is changed by the other variables – fertiliser application, number of years since wheat was last grown, inches of rainfall. These latter three variables are independent variables. If the rate of fertiliser application is changed, the yield of wheat may be expected to change as a result. The yield of wheat, however, will not change the rate of application of fertiliser (in the year) or the amount of rainfall. The unidirectional causality is upheld by this model specification and the meanings of dependent and independent variables should be clear. It is important to note, however, that independence, as applied to variables, does not necessarily imply that the independent variables do not cause changes in each other or change together. This may be a required property in some instances and a preferred property in all instances, but it is not part of the definition of independence. Thus the rate of fertiliser application may depend to some extent on rainfall amounts or the period since wheat was last grown on the land, without violating the independence property of these variables.

To calibrate a model, observations are needed on a sample of values of both dependent and independent variables for the units of concern. The resulting calibrated model would then be used to explain or predict the dependent variable. These sample values will have errors associated with them that will affect the estimation of model parameters in calibration. Various mathematical
operations on the independent variables may serve to propagate (magnify) individual errors to a much larger error in the dependent variable, while other operations may reduce errors or at least leave them unchanged. A number of rules can be put forward to reduce error propagation in models (Stopher and Meyburg, 1975). These may be summarised as follows: the preferred mathematical operation is addition, followed by multiplication and division, while the least desirable operations are subtraction and raising variables to powers.

10.2.1 Bivariate regression

The simplest form of relationship that can be hypothesised is a linear one, as typified by equation 10.2. Furthermore, the simplest linear relationship is one involving one dependent and one independent variable. The statistical procedure for developing a linear relationship from a set of observations is known as linear regression. The simple case of one dependent and one independent variable is called bi-variate linear regression, while the use of multiple independent variables is called multivariate linear regression. The procedures for estimating linear relationships can be described most economically and understandably by considering the simple bi-variate case.

A relationship is hypothesised of the form:

$$Y_i = \beta_0 + \beta_1 X_i$$

(10.4)

where

- $Y_i = \text{ith value of the dependent variable}$
- $X_i = \text{associated ith value of the independent variable}$
- $\beta_0, \beta_1 = \text{true values of the linear regression parameters}$

The parameter values, $\beta_0$ and $\beta_1$, are those sought by the estimation procedure. The true values are those that can be obtained by estimating the model from the entire population of values of $Y$ and $X$. In other words, the connotation "true" refers to population values as opposed to sample values, but does not refer to the existence or non-existence of a linear relationship of the form of equation 10.4. Thus the true value of $\beta_1$ may be zero, indicating that there is no relationship between $X$ and $Y$. This means that equation 10.4 is the equation sought, but the equation that can be estimated is one based on sample estimates, such as:

$$Y_i = a_0 + a_1 X_i + \epsilon_i$$

(10.5)
Chapter 10

where \( Y_i, X_i \) are as defined before,

\[
\begin{align*}
  a_0, a_1 &= \text{sample estimates of } a_0 \text{ and } a_1 \\
  \epsilon_i &= \text{the error associated with the } i\text{th observation}
\end{align*}
\]

The addition of the error term, \( \epsilon_i \), is significant in several respects. First, it is a clear indication of the acceptance of sample error in the calibration process. This error term, however, does not account for measurement or specification errors. Secondly, it is significant that the error is represented as being additive in the model. This indicates that the error is seen as being effectively independent of the independent variable, \( X_i \), since it does not interact with it. In fact, the calibration procedure of linear regression expressly assumes that the independent variables are known without error for all observations and that the error resides only in the dependent variable, \( Y_i \). Thirdly, the value \( \epsilon_i \) is assumed to exist, but cannot be measured until the values of \( a_0 \) and \( a_1 \) are obtained. The existence of \( \epsilon_i \) is, however, the mathematical basis for obtaining estimates of \( a_0 \) and \( a_1 \).

Before attempting to build a linear relationship, such as that presented in equation 10.5, it is necessary to determine whether a linear relationship is appropriate to describe the data. The various measures, discussed in later sections of this chapter, do not indicate whether a linear relationship is the best one to describe a phenomenon. They indicate only how good a fit is obtained by the linear function.

The first step in the analytical process, as described in Chapter 10.1, is to conduct exploratory data analysis. With respect to linear regression analysis, this first step must always be to construct a scatter diagram. Having selected a variable to be the dependent variable, a plot should be constructed of each independent variable against the dependent variable. These plots will generally show fairly clearly how appropriate a linear relationship is, or whether any relationship may be expected.

Some typical, hypothetical scatter diagrams are shown in Figures 10.17 through 10.20. Figure 10.17 shows evidence of a positive linear relationship between \( Y \) and \( X \). The relationship is positive because increases in \( X \) give rise to increases in \( Y \). Figure 10.18 shows the reverse relationship, but still a linear one. Figure 10.19 shows a more-or-less horizontal band of values, indicting no relationship between \( Y \) and \( X \). Figure 10.20 indicates a probable non-linear relationship between \( Y \) and \( X \) for which a linear regression would be inappropriate. Having constructed a scatter diagram and finding prima facie evidence of a linear relationship, the analyst may then proceed to construct a regression relationship.
Figure 10.17 Scatter Diagram of a Positive Linear Relationship

Figure 10.18 Scatter Diagram of a Negative Linear Relationship
If either of the plots of Figures 10.19 or 10.20 were obtained, then no linear-regression procedure should be applied to the data.

Given equation 10.5, the problem is now to determine a solution method for $a_0$ and $a_1$. Clearly, any solution should seek to minimise the error of calibration
Data Analysis

(measurement and specification errors must be considered as given for an assumed model and dataset). By definition, the sum of the error terms for the population is zero,

$$\sum_{i=1}^{N} e_i = 0 \quad (10.6)$$

where $N = \text{population size}$.

This is so because the true model is defined as the one that would be obtained for the population (equation 10.4) in which $e_i$ is absent. Hence, both the sum and average of the error terms over the population must be zero. In turn, this implies that in any sample there must be a range of values of $e_i$ that are distributed across both negative and positive values.

These properties lead to the conclusion that minimising the sum of the errors for the sample data is not an effective procedure. The absolute sum of the errors (if the sample is indeed drawn correctly) must approach zero as sample size is increased. In fact, in any large sample (where large may be defined as being in excess of 100 observations), the deviation of the sum of the error values from zero will be very small. It may also be noted that a strict minimisation of the sum of the error values would lead to a search for values of $a_0$ and $a_1$ that would produce extremely large negative values of $e_i$ since minus infinity is the smallest number known. Such a procedure would lead to estimates of $a_0$ and $a_1$ that would be as far from the true values as possible. Indeed, the calibration would generate a value of plus infinity for $a_0$ and zero for $a_1$, thus generating all values of $e_i$ as minus infinity.

To avoid trivial or absurd calibration results, it would appear that the best procedure would be to minimise the square of the error terms. By squaring the terms, all values to be summed become positive. Hence, the absolute minimum value becomes zero, occurring only when all values of $e_i$ are also zero. Thus minimising the sample sum of squared error terms leads to a calibration that is consistent with the true population model. The least-squares approach, as this procedure is termed, also has another important property that is discussed in Chapter 10.2.5 -- that of providing maximum-likelihood estimates of the regression coefficients that are unbiased for large samples.

Suppose there are a set of $n$ observations of the values of $Y_i$ and $X_i$. The hypothesised relationship between $X_i$ and $Y_i$ is that of equation 10.5. This equation may be rewritten to express $e_i$ in terms of the observed values of $Y_i$ and $X_i$ and the unknowns, $a_0$ and $a_1$,

$$e_i = Y_i - a_0 - a_1 X_i \quad (10.7)$$
Chapter 10

The sum of the squares of the deviations, $S$, from the true line is:

$$S = \frac{1}{n} \sum_{i=1}^{n} (Y_i - a_0 - a_1 X_i)^2$$  \hspace{1cm} (10.8)

where $n$ = sample size.

The least squares method then states that the linear-regression line is that which minimises the sum of the squares of the deviations, that is, the true line is that for which $\sum_i (Y_i - a_0 - a_1 X_i)^2$ is a minimum. Obtaining estimated values of $a_0$ and $a_1$ is a mathematical operation of no great complexity.

To find the minimum value of $S$, partial derivatives of $S$ are taken with respect to each coefficient in the equation, as shown in equations 10.9 and 10.10.

$$\frac{\partial S}{\partial a_0} = -2\sum (Y_i - a_0 - a_1 X_i)$$  \hspace{1cm} (10.9)

$$\frac{\partial S}{\partial a_1} = -2\sum X_i (Y_i - a_0 - a_1 X_i)$$  \hspace{1cm} (10.10)

Minimisation of $S$ occurs when the partial derivatives of $S$ are zero and the second derivatives are positive. Hence, $a_0$ and $a_1$ may be found by setting equations 10.9 and 10.10 to zero.

$$\sum (Y_i - a_0 - a_1 X_i) = 0$$  \hspace{1cm} (10.11)

$$\sum X_i (Y_i - a_0 - a_1 X_i) = 0$$  \hspace{1cm} (10.12)

With the necessary mathematical manipulations, the solutions for $a_0$ and $a_1$ may be obtained.

$$a_1 = \frac{\sum X_i Y_i - \sum X_i Y^2}{\sum X_i^2}$$  \hspace{1cm} (10.13)

$$a_0 = Y - a_1 X$$  \hspace{1cm} (10.14)

It is often helpful to visualise the linear-regression procedure as a geometric problem. The hypothesised relationship of equation 10.5 is shown in Figure 10.21. From this, it can be seen that $a_0$ is the intercept on the Y-axis and $a_1$ is the slope of the line. By virtue of equation 10.14, the point $(X,Y)$ must be on the line, as shown in Figure 10.21.
The linear regression problem is shown in Figure 10.22 as the problem of finding the best straight line that fits the scatter of data points, representing the n observations of $X_i$ and $Y_i$. The assumption that the errors, $\epsilon_i$ are in the $Y_i$ values only and that the $X_i$'s are known without error indicates that the $\epsilon_i$'s are the distances in the $Y$-dimension between the observed point $(X_i,Y_i)$ and the vertical projection of that point on the regression line $(X_i,Y_i)$ as shown in Figure 10.22. Thus, the least-squares solution is the one that minimises the sum of the squares of these vertical displacements.
Figure 10.22 Scatter Diagram of \( n \) Data Points

One reason for assuming the errors to reside in the \( Y_i \) values and not in the \( X_i \) values relates to the use of the resulting model for prediction of \( Y_i \) values. Since the errors are not known, it is necessary to assume that any new values of \( X_i \) for which values of \( Y_i \) values are to be predicted, are known exactly. If the reverse assumption were made, that all of the error were in the \( X_i \)s, it would be necessary to know the error to predict \( Y_i \) values. This is shown in Figure 10.23. The value \( Y_i \) would be obtained if the error were assumed to be zero. The best estimate of \( Y_i \), however, requires a knowledge of the error \( \hat{e}_i \). As shown by the comparative positions of \( Y_i \) and \( \hat{Y}_i \), this knowledge of the error is critical. Clearly, the geometry of this shows that the independent variable must be assumed to be known without the error.

This discussion leads to further insights into the definition or meaning of dependent and independent variables. Clearly, the independent variables are those to be used for prediction, while the dependent variable is the one to be predicted. Hence, the direction of causality is imposed on the regression model.
A property of regression models of particular interest concerns the sum of the residuals or error terms for the regression model. The residuals for the calibration data are the differences between the regression estimate $Y_i$, and the observed value of the dependent variable, $Y_i$. This difference, as shown previously, is the error term, $\hat{\epsilon}_i$:

$$\hat{\epsilon}_i = Y_i - \bar{Y}$$  \hspace{1cm} (10.15)

The sum of the residuals is:

$$\sum_i \hat{\epsilon}_i = \sum_i (Y_i - \bar{Y})$$  \hspace{1cm} (10.16)

To determine this value, it is necessary to substitute for $Y_i$, using the regression equation:

$$Y_i = \bar{Y} - a_1X + a_1X_i$$  \hspace{1cm} (10.17)

Substituting equation 10.17 into equation 10.16 yields:

$$\sum_i \hat{\epsilon}_i = \sum_i (Y_i - \bar{Y} + a_1X_i - a_1X_i)$$  \hspace{1cm} (10.18)
Summing the terms on the right side of equation 10.18 yields:

$$\sum_{i} \Delta i = \sum_{i} Y_i - \bar{Y} n + a_1 \sum_{i} X_i$$ (10.19)

However, $\sum_{i} Y_i$ is $nY$ and $\sum_{i} X_i$ is $nX$. Hence, it is clear that the right side of equation 10.19 is zero.

$$\sum_{i} \Delta i = 0$$ (10.20)

This property is important primarily because it indicates that what might appear to be a potential test of a regression model is not a very good test. Specifically, re-substituting the values, $X_i$, used to calibrate the model for a test of predictive power is a very weak test. Clearly, each individual prediction, $y_i$, can be examined against the observed value of $y_i$. However, summation of the $y_i$ values will yield exactly the summation of the $y_i$ values. This will occur regardless of how good or how poor the model fit may be. Thus summed predictions from the calibration data provide no test of goodness-of-the-fit.

What, then, can be used to measure the goodness-of-fit of a regression equation? Remember that the basis of the linear regression procedure is to minimise the error sum of squares. Using concepts based on Analysis-of-Variance (ANOVA), Stopher and Meyburg (1979) show that it is possible to derive a measure of goodness-of-fit termed the coefficient of determination ($R^2$) which may be expressed as:

$$R^2 = \frac{\sum (Y_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$$ (10.21)

Values of $R^2$ must lie between 0 and 1. The closer they are to unity, the better is the fit, while the closer they are to zero, the poorer is the fit. The coefficient of determination also signifies the percentage of the variance in $Y$ which is "explained" by the regression model.

The square root of the coefficient of determination is called the correlation coefficient and is denoted as $R$ (or $r$). While the coefficient of determination must always be positive, the correlation coefficient has a sign which indicates the direction of the relationship between the observed values of $X_i$ and $Y_i$. A positive value of $R$ indicates that as $X_i$ increases, so too does $Y_i$, whereas a negative value of $R$ indicates that as $X_i$ increases, then $Y_i$ decreases.
While the correlation coefficient provides a measure of the goodness-of-fit, it is by itself insufficient to assess the worth of the regression model. A simple, if somewhat extreme, example suffices to demonstrate this. Suppose a dataset comprises just two observations from a much larger population of data points. The correlation coefficient for a regression estimated from this dataset will obviously be unity, since the regression line must pass through all (two) data points. Suppose now that extra data points from the population are included in the sample. Since it is unlikely that all the new data points will lie on the line between the two original data points, then the correlation coefficient must fall below unity. Thus the coefficients in the regression model and the measure of goodness-of-fit are both functions of the number of data points used to estimate the regression model. What is needed is some way of estimating the reliability of these estimates as the sample size used in estimation is changed.

To assess the statistical reliability of the entire linear regression relationship, it is necessary to determine the probability that the correlation coefficient could have been obtained from random un-correlated data. As noted earlier, any two random data points will always give a correlation coefficient of unity; the probability of random data giving high values of the correlation coefficient will decrease as the sample size increases. Again referring to the concepts of ANOVA, it is possible to show that the F-statistic can be used to test the probability that the correlation coefficient could have been obtained by chance, given the size of the sample used to estimate the regression coefficients. The calculated value of F is compared against the tabulated value of F for the required level of confidence of avoiding a Type-I error (a value of $F$ at the 5% level is often used), and for degrees of freedom of $F_1 = 1$ and $F_2 = n-2$ (where $n$ is the sample size). If the calculated value of F is greater than the tabulated value, then the overall regression can be accepted as being statistically significant (i.e. not likely to have been generated by random data).

To assess the reliability of the coefficients in the regression equation, use is made of the t-statistic. If independent random samples had been drawn from the population, then the resulting regression equation coefficients would have varied somewhat. If the data in the population is actually described by some underlying relationship between the independent and the dependent variables, then we would expect that essentially the same regression equation would be reproduced on these repeated samplings. However, we know that random fluctuations in the data will cause the coefficients to vary. What we would like to estimate is the extent to which these random fluctuations contribute to the values obtained for the coefficients. The t-statistic measures the relative magnitude of these fluctuations by taking the ratio of the average value of the coefficient to the standard error of the coefficient (i.e. the standard deviation of the coefficient if repeated sampling had been conducted). Fortunately, it is generally unnecessary to conduct repeated sampling, and the standard error of the estimate of the...
Chapter 10

coefficient can be obtained by ANOVA techniques from a single sample regression. The calculated value of the t-statistic can then be compared with tabulated values of t to test the significance of the coefficient. If the calculated value of t is greater than the tabulated value of t at the 100(1 - \(\alpha\)/2) percentage level with \((n-2)\) degrees of freedom, then the hypothesis that the true value of the coefficient is equal to zero (i.e. no correlation) can be rejected with a confidence of 100(1 - \(\alpha\))%. At a 95% level of confidence, and with a sample size of greater than about 50, a critical value of the t-statistic is equal to 2.00. The same t-test can be applied to the constant in the regression model, but the interpretation of the result is somewhat different. While a low value of t for a regression coefficient implies that the dependent variable is unlikely to be related to the independent variable (at least in a linear fashion), a low value of t for the constant implies that the regression (if one exists) is likely to pass through the origin.

To illustrate the potential application of linear regression, consider the finding in Chapter 10.1 that there appeared to be a linear relationship between the number of trips per person per day and the number of vehicles owned by that person's household (see Figure 10.10). This hypothesis can be tested by confirmatory analysis by seeing whether a statistically significant linear regression model can be fitted to the data.

Within Data Desk, this regression can be performed by clicking on the triangular arrow at the foot of the scatterplot (a so-called "HyperView" button), and then selecting the regression option as shown in Figure 10.24. The regression is then performed with the variable on the horizontal axis being assumed to be the independent variable and the variable on the vertical axis assumed to be the dependent variable. The results are then presented in tabular fashion in a window on top of the scatterplot, as in Figure 10.25.

In terms of the previous discussion, the following interpretation can be applied to the results of the regression. First, it appears from the coefficient of determination \((R^2)\) that a good regression may exist because of the high value of 80.5%. However, as described earlier, this high value may be misleading because of the small number of data points used (only 9 values of mean trip rate were used). To test this, we need to examine the value of the F-statistic. With 7 degrees of freedom, an F value of 28.9 would indicate a significant regression at almost the 0.1% level of confidence (where the critical value of F is 29.25). That is, we could accept this regression and only be proved wrong (i.e. that no such regression exists) 1 in 1000 times. We can also see that the constant and the coefficient are also significant at better than the 5% level of confidence (i.e. the values of t are both greater than 2.00). The t-statistic on the independent variable merely confirms what we have already seen from the F-statistic (because in bivariate linear regression, the values of F and t are directly related), while the t-statistic on the constant indicates that the regression does not pass through the
origin (i.e. people in zero-vehicle households still make some trips; in fact they make an average of 2.698 trips!).

![Trips vs Vehicles](image1)

**Figure 10.24** Selection of Linear Regression Option in Data Desk

![Regression of Trips per Person vs Vehicles per Household](image2)

**Figure 10.25** Results of Linear Regression Analysis in Data Desk

To further check the validity of this linear regression, we can plot the residuals from the regression against the predicted values of the dependent variable. The residuals are the differences between the predicted values of the dependent variable and the actual values. Obviously, if a perfect regression had been obtained (i.e. $R^2 = 1$), then all the residuals would have been equal to zero. However, in real regressions this never occurs and what we seek to check is
whether there is any relationship between the residuals and the predicted values. To plot the residuals scatterplot, click on the HyperView button and select that option as shown in Figure 10.26. This will produce a scatterplot as shown in Figure 10.27.

![Figure 10.26 Selecting a Residuals Scatterplot in Data Desk](image)

![Figure 10.27 Results of Residuals Scatterplot in Data Desk](image)

It can be seen that the residuals are fairly evenly distributed around the zero line with values of plus/minus 0.5 trips per person per day. Importantly, we can see that there is no relationship between the residuals and the predicted values and that the variance of the residuals does not vary with the size of the predicted values (this could be verified by performing a further regression on this scatterplot using the HyperViews button at the bottom of this scatterplot).
lack of a relationship is important because it signifies a lack of heteroscedasticity (i.e. non-uniform variances) which would invalidate the use of simple linear regression. If such a relationship is observed, then you would be advised to consult a regression textbook (e.g. Draper and Smith, 1968) to determine the seriousness of this effect.

Given the results described above, it would appear that a linear relationship does exist between the two variables. This is not surprising given the scatterplot of Figure 10.10, but it appears somewhat surprising given the scatterplot of Figure 10.7, where the ungrouped data points are plotted. However, if a true relationship exists, then it should exist at all levels of aggregation. Therefore, consider performing a linear regression on the 904 data points shown in Figure 10.7. The results of this regression are shown in Figure 10.28.

![Summary Table]

**Summary**

<table>
<thead>
<tr>
<th>Dependent variable is: Person Trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>904 total cases of which 7 are missing</td>
</tr>
<tr>
<td>$R^2 = 2.6%$ $R^2(\text{adjusted}) = 2.5%$</td>
</tr>
<tr>
<td>$s = 2.925$ with $897 - 2 = 895$ degrees of freedom</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>207.459</td>
<td>1</td>
<td>207</td>
<td>24.3</td>
</tr>
<tr>
<td>Residual</td>
<td>7655.56</td>
<td>895</td>
<td>8.55370</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>s.e. of Coeff</th>
<th>t-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>3.15395</td>
<td>0.1732</td>
<td>18.2</td>
</tr>
<tr>
<td>Household Yehi...</td>
<td>0.291337</td>
<td>0.0592</td>
<td>4.92</td>
</tr>
</tbody>
</table>

**Figure 10.28 Results of the Ungrouped Linear Regression**

It appears, from the low value of the coefficient of determination, that this regression is not as good as that obtained from the grouped data. That is, the regression is not explaining as much of the variance in the dependent variable. However, this should not be too surprising since we eliminated most of the variance from the dependent variable when we performed the aggregation. To check on the real significance of the regression, we need to consider the F-statistic. With 985 degrees of freedom, an F value of 24.3 would indicate a significant regression at well beyond the 0.1% level of confidence (where the critical value of F is 10.83). In addition, both the variable coefficient and the constant term are highly significant as indicated by the high values of the t-statistic. Thus the regression on the ungrouped data appears to be even stronger than the regression on the grouped data, even though the proportion of the variance explained is far less. This example clearly demonstrates that the reporting of the correlation coefficient is meaningless without also reporting the F-statistic.
A further test on the validity of the regression can be obtained by plotting the distribution of the residuals, as shown in Figure 10.29. If the assumptions underlying the regression are upheld, then the distribution of the residuals should be approximately normal. This would appear to be the case for the ungrouped data used in the above analysis, although the distribution is skewed slightly to the right.

![Distribution of the Residuals](image)

**Figure 10.29 Distribution of the Residuals**

One major difference between the two regressions is that the estimated coefficients and constants are different, with the regression on the ungrouped data having a higher intercept and a lower slope. This difference serves to illustrate a major misuse of regression on survey data. As described earlier, when looking at the data, it always appears that the grouped data will give a better regression model because of the reduction in scatter of the plotted points. For this reason, regressions are often erroneously performed on the grouped data. However, in the regression on the grouped data described above, it is implicitly assumed that each of the group means has equal weight in the determination of the regression line. However, in the calculation of the group means, a different number of data values has been used in the calculation. One would expect, intuitively, that those group means which have been calculated from the larger number of data values are more reliable and hence should be given more weight when estimating the regression equation. This can also be shown to be the case statistically, and gives rise to a method known as *weighted regression*.

The essence of weighted regression is to weight each data point by the inverse of the variance within that data point (i.e. give more reliable data points more weight). This can also be shown to be equivalent to weighting each data point by the number of data values contributing to each of the data points. Thus if there
Data Analysis

are \( m_i \) data values grouped into the \( i \)th group, then the independent and dependent variables for each group are entered into the regression \( m_i \) times. Many regression packages have the facility to automatically carry out this procedure simply by specifying the number in each group (Data Desk lacks this facility, but SYSTAT is one package which can be used in this way). The input data and results from a weighted regression using SYSTAT is shown in Table 10.1.

<table>
<thead>
<tr>
<th>Trip Rate</th>
<th>Vehicles</th>
<th>Group Size</th>
<th>Regression Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.95</td>
<td>0</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>3.42</td>
<td>1</td>
<td>267</td>
<td></td>
</tr>
<tr>
<td>4.01</td>
<td>2</td>
<td>267</td>
<td></td>
</tr>
<tr>
<td>4.39</td>
<td>3</td>
<td>126</td>
<td></td>
</tr>
<tr>
<td>3.85</td>
<td>4</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td>4.36</td>
<td>5</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>4.60</td>
<td>6</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>6.00</td>
<td>7</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>5.44</td>
<td>8</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>

\[ R^2 = 0.622 \]
\[ a_0 = 3.158 \]
\[ t_{\text{constant}} = 142.4 \]
\[ t_{\text{coeff.}} = 0.291 \]
\[ F = 1473.7 \]
\[ \text{d.o.f} = 895 \]

It can be seen that the regression coefficients are now in agreement with those of the ungrouped regression shown in Figure 10.28. However, the t-statistics are larger and the regression appears more significant than it really is (by means of the very high F-statistic) because an implicit assumption has been made that all of the data values within a group are identical, whereas in reality there is still a considerable degree of variability within each group as shown in Figure 10.7. Nonetheless, weighted regression at least overcomes the biasing effect on the estimated coefficients and hence should always be used with grouped data if the original data values are not available.

10.2.2 Multivariate regression

The last section provided a detailed treatment of bi-variate regression, that is, a linear relationship between only two variables. However, it is relatively unlikely that a useful model can be obtained from a bivariate relationship, because this would generally suggest a much too close relationship of variables to be plausible. For example, in the case of the ungrouped regression model of trips per person as a function of vehicles per household, only a relatively small portion of the variance in the observed trips per household was captured by the variable vehicles per household. It is clearly not very satisfactory to be able to account for such a small proportion of the variance in the model. Thus it appears extremely probable that real modelling efforts will require a more complex relationship to be established.

Additional complexity can involve two basic alternatives: a more complex relationship between the variables, that is, a nonlinear relationship; or the use of
more than one independent variable. The second alternative is both simple and the subject of this chapter. The extension of bivariate linear regression to multivariate linear regression is principally a matter of simple mathematical generalisations of the relationships developed in Chapter 10.2.1.

Before undertaking these generalisations, it is useful to consider the concepts and visualisation of the multivariate linear relationship. Bivariate regression involved fitting a straight line to data in a two dimensional space. Multivariate regression involves the fitting of an (n-1) dimensional surface to data in n dimensional space. It should be noted that the surface is a plane, since the relationship is linear. Illustrations of higher dimensionality problems cannot, of course, be provided. However, the extension is conceptually relatively simple.

The three-dimensional problem may be examined in another light. The relationship postulated is:

\[ Y_i = a_0 + a_1 X_{1i} + a_2 X_{2i} + \epsilon_i \] (10.22)

If one of the two independent variables, \(X_i\) or \(X_2\), were to be held constant, the relationship could be written:

\[ Y_i = a_0' + a_1 X_{1i} + \epsilon_i \] (10.23)

\((X_{2i} = \text{fixed value}).\)

Alternatively, equation 10.24 could be used:

\[ Y_i = a_0'' + a_2 X_{2i} + \epsilon_i \] (10.24)

\((X_{2i} = \text{fixed value}).\)

In these two equations, \(a_0'\) and \(a_0''\) represent adjusted constant terms that take account of the effect of the variable whose value is fixed. Thus in equation 10.23, assuming that the fixed value of \(X_{2i}\) is some arbitrary value \(g\) then \(a_0'\) is:

\[ a_0' = a_0 + a_2 g \] (10.25)

Similarly, if the fixed value of \(X_{1i}\) in equation 10.24 is 0, then \(a_0''\) is:

\[ a_0'' = a_0 + a_1 0 \] (10.26)

The relationships of equations 10.23 and 10.24 could be represented then as a family of straight lines in a two dimensional space. Note that the lines are parallel, since at each value of \(X_2\) that variable has no further effect on the value of \(Y_i\) as \(X_{1i}\) changes in value. In engineering terms, the lines represent a series of
Data Analysis

projections of the surface onto the two-dimensional surface formed by the Y and X₁ axes.

It is important at this point to reconsider the concepts of dependence and independence. As stated in Chapter 10.2.1, the basic concept of dependence and independence is one of causality. Thus changes in an independent variable cause changes in the dependent variable, while changes in the dependent variable do not cause changes in the independent variable. Thus, in the weather-forecasting example, humidity may not be an appropriate independent variable, since precipitation will cause changes in humidity (although it may not be strictly possible to say that the likelihood of precipitation will cause changes in humidity). In the context of multivariate regression, independence does not mean that the independent variables are unrelated to each other. Hence it is not a statement that the variables are independent of each other. However, if any two or more independent variables are strongly related to each other, problems will arise. First, one may consider that if there are two or more highly related independent variables, the use of them in a model creates a redundancy. Conceptually, such a redundancy is undesirable and may contribute to some problem in understanding how the phenomenon being modelled behaves in reality. Statistically (discussed later in the chapter), the presence of such redundancy will lead to incorrect estimates of the coefficients of all such interrelated variables. Hence it may be stated that high intercorrelations are conceptually and statistically undesirable, but the principle of independence by itself does not exclude such intercorrelations. Such variables should therefore be checked for and, if found, excluded from the proposed linear relationship.

Given the basic concepts of multivariate linear regression, attention can now be given to the estimation of the coefficients of the regression equation. In general, a model of the form of equation 10.27 may be postulated as representing the multivariate linear-regression equation.

\[
Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + ... + b_mX_{mi} + e_i
\]  

(10.27)

To calibrate such a model, the procedure is to minimise the sum of the squared error terms, \(e_i\), to find the values of the coefficients and the constant. Rearranging equation 10.27 to isolate \(b_i\) on one side produces:

\[
b_i = Y_i - b_0 - b_1X_{1i} - b_2X_{2i} - ... - b_mX_{mi}
\]  

(10.28)

The process is then completely analogous to that described for bivariate linear regression. Equation 10.28 is squared and summed over all observations, \(i\), an partial differentials taken with respect to each of the unknown values, \(a_k\). The result of this procedure is a set of normal equations of the following form:

\[
b(Y_i - b_0 - b_1X_{1i} - b_2X_{2i} - ... - b_mX_{mi}) = 0
\]
The set has \((m + 1)\) equations, all of which are linear in the unknown values, \(a_0\) through \(a_m\). Thus a unique solution may be found by solving this set of linear equations as was done before. As an example, consider the case in which \(m\) equals 2. The normal equations are:

\[
\begin{align*}
\sum (Y_i - a_0 - a_1 X_{1i} - a_2 X_{2i}) &= 0 \quad (10.30) \\
\sum X_{1i} (Y_i - a_0 - a_1 X_{1i} - a_2 X_{2i}) &= 0 \quad (10.31) \\
\sum X_{2i} (Y_i - a_0 - a_1 X_{1i} - a_2 X_{2i}) &= 0 \quad (10.32)
\end{align*}
\]

As before, the first normal equation provides an identity in terms of the means:

\[
Y - a_0 - a_1 X_1 - a_2 X_2 = 0 \quad (10.33)
\]

Rearranging this equation to yield a definition of the constant, \(a_0\), produces:

\[
a_0 = Y - a_1 X_1 - a_2 X_2 \quad (10.34)
\]

The remainder of the solution becomes algebraically tedious and is best solved by matrix algebra. Defining the vector of observations of \(Y_i\) as \(Y\), the matrix of observations of \(X_{1i}\) and \(X_{2i}\) as \(X\), the vector of coefficients as \(a\), and the vector of error terms, \(e_i\), as \(e\), the solution for the coefficients is:

\[
a = (X'X)^{-1} X'Y \quad (10.35)
\]

This is the general solution of the multivariate regression.

In Chapter 10.2.1, a number of properties of the bivariate linear-regression model were described and discussed. Without exception, the multivariate linear-regression model holds these properties.

Thus, the sum of the residuals is zero, regardless of the number of independent variables used. This follows from the first of the normal equations. The sum of the residuals is given by equation 10.36, which is a rearrangement of equation 10.27, summed over all observations, \(i\).

\[
\sum e_i = \sum (Y_i - a_0 - a_1 X_{1i} - a_2 X_{2i} - ... - a_m X_{mi}) \quad (10.36)
\]
However, the first normal equation is:
\[
\frac{\partial}{\partial \beta_0} (Y_i - \beta_0 - \beta_1X_{1i} - \beta_2X_{2i} - \cdots - \beta_mX_{mi}) = 0
\] (10.37)
Hence it follows that the sum of the errors (residuals) is zero.
\[
\frac{\partial}{\partial \beta_0} (Y_i - a_0 - a_1X_{1i} - a_2X_{2i} - \cdots - a_mX_{mi}) = 0
\] (10.38)
The implications of this property are again the same as those discussed for bivariate linear regression.

To test the significance of each individual variable in the regression equation, \( t \)-tests can again be conducted on each coefficient, as described for the bivariate regression, and the computation is exactly the same. The constant may be tested by the same procedure as before. The degrees of freedom of the \( t \) tests are all \( (n-2) \).

10.2.3 Factor analysis

Factor analysis is a procedure with roots in the basic principals of multivariate linear-regression analysis. One of the problems noted in Chapter 10.2.2, with respect to multivariate linear regression, is that multicollinearity of two or more independent variables causes mis-estimation of the coefficients of such variables. Frequently, exclusion of one of the multicollinear variables will result in a less than satisfactory relationship, while inclusion leads to counter-intuitive values for the coefficients. The analyst is thus placed on the horns of a dilemma. Factor analysis and, more particularly, principal components or principal factors (a special case of factor analysis) provide a means to resolve this dilemma.

In very broad terms, factor analysis is a method for reformulating a set of natural or observed independent variables into a new set (usually fewer in number, but necessarily not more in number) of independent variables, such that the latter set has certain desired properties specified by the analyst. Kendall (1965) has suggested a difference between factor analysis and principal-components analysis. He suggests that principal-components analysis is the search through data to try to find factors or components that "may reduce the dimensions of variation" and may "be given a possible meaning". On the other hand, he suggests that factor analysis starts with the hypothesis of a model and tests it against the data to see if it fits. This distinction seems useful and is pursued in the latter part of this section. (It is notable that other authors, e.g., Harman (1960) and Tintner (1952), do not appear to make use of this distinction.) This section concentrates on principal-components analysis, where among other things the latter set of variables has the property of zero correlations between all new variables. In all versions of factor analysis, the variables (factors) in the new set are formed as linear combinations of all variables in the original set.
Chapter 10

The method was originally developed in mathematical psychology as a means to reduce a large number of measures (probably collinear) obtained in psychological measurement to a small number of largely uncorrelated salient characteristics, capable of describing character traits or other concepts in psychology. For readers familiar with psychological measurement, factor analysis was originally developed as the principal means of constructing multidimensional scales (Harman, 1960). Although the method has been criticised by many psychologists, it still remains one of the most tractable and understandable methods of undertaking multidimensional scaling (Koppelman and Hauser, 1977).

Some examples may be useful to illustrate the purpose and procedure of factor analysis. Consider, first, a situation in marketing cars. A sample of people have been asked to rate some of the aspects of the finish and comfort of a number of car models. The ratings were made by allocating points on each attribute to each model, where the points may be selected from zero to 100. Fourteen attributes are used, as shown in Table 10.2.

Table 10.2 Attributes Used to Rate Various Car Models

<table>
<thead>
<tr>
<th>Colour</th>
<th>Heating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seat width</td>
<td>Interior noise</td>
</tr>
<tr>
<td>Seat support</td>
<td>Carpeting</td>
</tr>
<tr>
<td>Interior finish (panels, fascia)</td>
<td>Entry/exit space</td>
</tr>
<tr>
<td>Radio/tape player</td>
<td>Safety belts</td>
</tr>
<tr>
<td>Window controls</td>
<td>Seat adjustment</td>
</tr>
<tr>
<td>Air conditioning</td>
<td></td>
</tr>
</tbody>
</table>

After obtaining the scores of each person, one could obtain the means for each model on each attribute (see Figure 10.30). There is a great deal of useful information in Figure 10.30. However, it is very hard to absorb the information and even harder to gain an impression of which model is superior. Furthermore, many of the attributes may be related, either in the minds of the raters or in the survey design. For example, the efficiency and effectiveness of the air conditioning and heating are likely to be related, while these same items are probably quite unrelated to the exterior finish.

Applying factor analysis to these ratings will lead to greater insights into the images of the various models and will also show how the various attributes link together. The grouping of attributes is shown by the coefficients of the original variables in the composition of the factors. Those with large (negative or positive) coefficients make a major contribution to the factor, while those with small (near zero) coefficients make little contribution to the factor. Suppose a factor analysis of the ratings has yielded the groupings of attributes shown in Table 10.3, where the three factors account for more than 95% of the variance of
the original attribute ratings. Factor 1 might be termed exterior design, factor 2 interior comfort, and factor 3 environmental controls.

For each factor, an average score for each car model could be obtained and plotted, as shown in Figure 10.31. The information provided by Figure 10.31 is at
once much easier to comprehend, and has also shown how certain items are correlated to produce a major effect, that is, the three factors.

Table 10.3 Factor Groupings for Attributes of Table 10.2

<table>
<thead>
<tr>
<th>Factor</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Colour</td>
</tr>
<tr>
<td></td>
<td>Exterior finish</td>
</tr>
<tr>
<td></td>
<td>Entry/exit space</td>
</tr>
<tr>
<td>2</td>
<td>Space width</td>
</tr>
<tr>
<td></td>
<td>Seat support</td>
</tr>
<tr>
<td></td>
<td>Interior finish</td>
</tr>
<tr>
<td></td>
<td>Interior noise</td>
</tr>
<tr>
<td></td>
<td>Carpeting</td>
</tr>
<tr>
<td></td>
<td>Seat adjustment</td>
</tr>
<tr>
<td>3</td>
<td>Radio-tape player</td>
</tr>
<tr>
<td></td>
<td>Window controls</td>
</tr>
<tr>
<td></td>
<td>Air conditioning</td>
</tr>
<tr>
<td></td>
<td>Heating</td>
</tr>
<tr>
<td></td>
<td>Safety belts</td>
</tr>
</tbody>
</table>

It should be noted that factor analysis is not a modelling technique. Rather, it is a procedure for manipulating data prior to developing models. The results of factor analysis can be used directly in some forms of decision making or appraisal, but such use does not conform to the usual definitions of modelling.

10.2.4 Discriminate analysis

Discriminant analysis was originally developed for use in the biological sciences by Fisher (1936). The basic hypothesis of discriminant analysis runs as follows: it is assumed that a population is made up of two distinct subpopulations. It is further assumed that it is possible to find a linear function of certain measures or attributes of the population that will allow an observer to discriminate between
the two subpopulations. Originally, the technique was devised to assist biologists in identifying subspecies. In this context, suppose a researcher has two subspecies of a particular plant species, say a daffodil. In general appearance the two subspecies are so alike that one cannot with any certainty state which is which. However, the length, width, and thickness of the leaves and the maximum diameter of the bulb can be measured. It can be hypothesised that a linear combination of these measures can be devised which will allow the analyst to discriminate between the two subspecies with the least possible likelihood of error.

This technique appears to have many possible applications in transportation planning. An application frequently used pertains to the choice of transport mode, where the two populations are considered as being car users and transit users, and a linear combination of system and user characteristics is sought as a basis of discriminating between the two populations. Using this as an example, the frequency distributions of the two populations--car users and transit users--can be plotted against some function $z$. Now over the range $z_1$ to $z_2$ (see Figure 10.32), it is not certain whether an individual with a $z$ value in that range is an car or a transit user. Suppose it has to be decided whether each person in this total population is an car or a transit user. How can one proceed so as to make as few mistakes as possible? To put it in a slightly different way, how can one minimise the number of mis-classified people?

![Figure 10.32 Frequency Distribution of Car versus Transit Users](image)

Two important things should be noted here, namely, that each member has to be assigned to one or the other subpopulation, and that the decision as to how to divide the population has already been made, that is, into auto and transit users.
In other words, discriminant analysis is not designed as a procedure for seeking population groupings, like cluster analysis.

It should be clear from this that discriminant analysis has many potential uses beyond the original biological ones. Whenever it is desired to find relationships that would permit classifying human populations into groups, discriminant analysis may be an applicable method. Likewise, transport engineering often calls for classification of various physical items into groups having distinct properties. Whenever such grouping cannot be made on obvious grounds, discriminant analysis may be used to find relevant compound measures to permit appropriate classification.

The problem is one of determining the function $z$ that will best permit the discrimination between members of the two populations. Let $z$ be defined as a function of a number of factors $x_1, x_2, x_3, \ldots, x_k$, and the subpopulation be designated by subscript $i$ and the members of the subpopulations by subscript $j$. Then,

$$z_{ij} = a_1 x_{1ij} + a_2 x_{2ij} + \ldots + a_k x_{kij} \quad (10.39)$$

For convenience, equation 10.39 may be abbreviated to:

$$z_{ij} = \sum_{p=1}^{k} a_p x_{pij} \quad (10.40)$$

where $i = 1, 2$

and $j = 1, 2, \ldots, n$

and $a_p$ = the weighting coefficient of the $p$th factor, $x_{pij}$

The task is to determine the values of the set of weighting coefficients, $[a_p]$, such that one can best discriminate between the two subpopulations. To set about determining these coefficients, it is necessary to define what is meant by discrimination between two subpopulations.

Two alternative definitions could be postulated. Consider Figure 10.33. One may postulate that the rule sought is to state that all members of the population with a value of $z$ less than $z'$ are to be classified as being in subpopulation 1, while those greater than $z'$ are classified in subpopulation 2. Clearly, those members of population 2 who fall in the shaded area $M_2$, will be mis-classified by this rule, as will those of population 1 who fall in $M_1$. 
Figure 10.33 Definition of Discrimination between Two Subpopulations

Neyman and Pearson (1933) suggested that discrimination be treated as an attempt to minimise the total number of mis-classifications, that is, the sum of $M_1$ and $M_2$. The task is to state this mathematically so as to define the coefficients of the discriminant function, $z$.

Fisher (1936) suggested, alternatively, that discrimination could be defined as achieving the maximum separation of the two subpopulations. This is, of course, equivalent in effect to minimising mis-classifications, but it generates a different mathematical statement of the problem and is the derivation used here. Of course, with Fisher's definition of discrimination, some care is needed in selecting the measure of separation of the two subpopulations. If one were to measure this as the distance, $D_m$ between the two measures, $z_1$ and $z_2$, one could clearly increase $D$ simply by multiplying the $z$ function by some factor greater than one. This, however, would not increase the separation in any meaningful way. Therefore, Fisher proposed that separation be measured as the distance between the means, $D$, relative to the within-subpopulation variances. This would make it clear that scaling the discriminant function has no effect upon the separation of the two subpopulations. Finally, the simple distance between the two means may not be the best measure, since one may experience sign problems. (A priori, one may not know which subpopulation has the small $z$ values. Hence, by measuring from the second subpopulation mean, say, from Figure 10.13, the distance would be negative and the largest negative distance would be desired.) This can be overcome by maximising the square of the distance between the two populations, where this distance is also the between-population variance.
Chapter 10

Stopher and Meyburg (1979) describe the procedures involved in estimating a set of coefficients for use in the discriminant function (eqn. 10.58). The next step is to determine whether the discriminator is significant. There are three factors to consider in this determination. First, there may be a real difference between the populations, but they are so close together that a discriminator is not very effective. This is measured by the errors of misclassification which, though the minimum obtainable, may still be large. Second, there may be a real difference between the populations but the sample is not large enough to produce a very reliable discriminator; this is really a matter of setting confidence limits on the interpretation of the sample results as described in Chapter 4. Third, it may be that the parent populations are identical and that a discriminant function is illusory. For the purposes of the use of discriminant analysis in this context, the latter is unlikely to occur, since this technique is only suggested when separate populations can in fact be readily identified. However, the first two questions of significance are very relevant. Considering the first point, the populations may be interpreted as being too close together in two ways. Either the wrong factors were chosen or else not all the significant factors were used to build the discriminant function. Alternatively, the assumptions of rationality and consistency implicit in the modelling technique are so tenuous that no set of factors can discriminate effectively between the two populations. In either case, the significance of the observed discrimination can be measured in terms of the probability of the observed discrimination occurring at random. This can be done using a variance ratio (Kendall, 1965). Because the variance ratio involves measures of both the distance between the populations and the numbers in the sample, it can also be used as an indicator of the reliability of the discriminator due to sample size. One may also wish to test the significance of each factor in the discriminant function, and this can be done using the t-statistic described in Chapter 10.2.2.

As an example of the use of discriminant analysis, consider the following study taken from Lisco (1967) relating to travel to work in Chicago. For 159 people, information was obtained on the costs of their travel to work, the time it took, the mode of travel used, income, age and sex of the respondent, and information on the availability of a car for the trip to work. The problem was to see if a discriminant function could be used to separate transit and auto users. A total of 61 respondents used autos and 98 used transit.
Data Analysis

Car availability was entered as a dummy variable (i.e. a Yes/No response), while income, sex and age were entered as values. Time and cost were entered as differences between transit and auto. The results of the discriminant function analysis are shown in Table 10.4. For this model, the variance ratio was significant beyond the 99% confidence level, indicating that a significant discriminator appeared to exist. The total number of misclassifications was 37 out of the 159 respondents, where 11 car users were assigned to transit and 26 transit users were assigned to auto.

Table 10.4 Results of Discriminant Analysis on Chicago Data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>t-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Travel Time</td>
<td>0.0330</td>
<td>2.48</td>
</tr>
<tr>
<td>Travel Cost</td>
<td>0.0050</td>
<td>3.15</td>
</tr>
<tr>
<td>Car Availability 1</td>
<td>0.0631</td>
<td>0.36</td>
</tr>
<tr>
<td>Car Availability 2</td>
<td>-0.5039</td>
<td>2.98</td>
</tr>
<tr>
<td>Income</td>
<td>-0.8167</td>
<td>3.36</td>
</tr>
<tr>
<td>Sex</td>
<td>-0.2713</td>
<td>1.81</td>
</tr>
<tr>
<td>Age</td>
<td>-0.0203</td>
<td>2.39</td>
</tr>
<tr>
<td>Constant</td>
<td>1.1790</td>
<td>----</td>
</tr>
</tbody>
</table>

Examining the results of Table 10.4, it may be seen that all but two of the coefficients are significantly different from zero at the 95% confidence level (i.e. only two t-scores are less than 2.00). Car users are found to be associated with positive values of the discriminant function and transit users with negative values. It appears therefore that car users are associated with positive travel time and travel cost differences. For classification purposes, since these differences are expressed as transit minus auto times and costs, auto drivers are more likely, *ceteris paribus*, to be associated with shorter travel times and costs than for the transit trip. Similarly, transit users are likely, *ceteris paribus*, to be older and have higher incomes than car users and are more likely to be female (sex was entered as a 1, 2 variable, such that 2 indicated female). Transit users are also more likely to have more favourable time and costs by their chosen mode than by car.

The interpretation of this example illustrates the correct use of discriminant analysis as a classification tool. Thus, given data on a new individual, one could classify that person as a transit or auto user by evaluating their value of the discriminant function. Suppose, for example, that data are obtained on three additional people in the study area, as shown in Table 10.5 (note that age and income were coded as normalised variables, i.e. unit mean and standard deviation).


Chapter 10

Table 10.5 Additional Chicago Commuters to be Classified

<table>
<thead>
<tr>
<th>Person</th>
<th>Time Diff.</th>
<th>Cost Diff.</th>
<th>Car Avail.1</th>
<th>Car Avail.2</th>
<th>Income</th>
<th>Sex</th>
<th>Age</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0.50</td>
<td>1</td>
<td>0.5</td>
<td>0.340</td>
</tr>
<tr>
<td>2</td>
<td>-20</td>
<td>-10</td>
<td>0</td>
<td>0</td>
<td>0.95</td>
<td>2</td>
<td>3.2</td>
<td>-0.914</td>
</tr>
<tr>
<td>3</td>
<td>-15</td>
<td>25</td>
<td>1</td>
<td>0</td>
<td>0.76</td>
<td>2</td>
<td>2.2</td>
<td>-0.336</td>
</tr>
</tbody>
</table>

The value of the discriminant function, Z, for each individual is also shown in Table 10.5. The discriminant function was defined in Table 10.4 such that zero represents the boundary between the two populations, with negative values of Z indicating a transit user. Hence the discriminant function suggests that individual 1 is likely to be a car user, while individuals 2 and 3 are likely to be transit users. These classifications are intuitively satisfying. The time and cost differences for individual 1 show that for this person the car is quicker and cheaper. Hence the car is a logical choice and the classification is reasonable. Similar reasoning applies to individual 2. Individual 3 has a time advantage by transit and a cost disadvantage. While the cost disadvantage is larger, the relative sizes of the coefficients in Table 10.4 mean that the travel time advantage has far more effect on the classification. This is reinforced by the fact that this individual is female and older than average, and hence the transit classification is intuitively correct.

10.2.5 Maximum likelihood estimation

The regression techniques described in Chapters 10.2.1 and 10.2.2 have both used least-squares regression to obtain their estimates of the coefficients of the regression equation. However, the use of least-squares is based on assumptions of normality and independence, and these conditions may not always prevail. In such circumstances, it has been found that the use of Maximum Likelihood Estimation (MLE) techniques proves to be a valuable methodology. MLE can also be used in more general situations related to the likelihood of outcomes.

The concept of likelihood is derived from the ideas of probability and experimental outcomes (Ehrenfeld and Littauer, 1964). Consider an experiment to measure the level of carbon monoxide in an urban area. Because of local variations in concentration, dispersion, proximity of sources, and so forth, one will obtain a range of different values from these different points. Before or after undertaking the experiment, one may postulate a general level of carbon monoxide in the urban area. For example, the experiment may have generated the data of Table 10.6.

One may postulate that the average level of carbon monoxide in the area is 20 parts per million (ppm). Now one would like to know how likely it is that these values could have been obtained if the average urban level is 20 ppm. This can be determined by calculating the likelihood of the values of Table 10.6, given the assumption the average level is 20 ppm. One may also calculate the likelihood of
these values, given some other assumed average level, say, 25 ppm. As is seen later, additional assumptions must be made to estimate the value of the likelihood. For the moment, however, it is appropriate to consider further the meaning of the likelihood.

Table 10.6 Observations on Carbon Monoxide Level (parts per million)

<table>
<thead>
<tr>
<th>Location</th>
<th>Carbon Monoxide Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>26</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
</tr>
<tr>
<td>4</td>
<td>32</td>
</tr>
<tr>
<td>5</td>
<td>23</td>
</tr>
<tr>
<td>6</td>
<td>25</td>
</tr>
<tr>
<td>7</td>
<td>17</td>
</tr>
</tbody>
</table>

For any assumed average level of carbon monoxide, it is possible to define a probability that each of the values in Table 10.6 could have occurred. The likelihood is then defined as the joint probability under the assumed average value. Suppose the probability of measuring 26 ppm at location 1 under the assumed value of 20 ppm is denoted \( p(1/20) \), the value of 13 ppm at location 2 denoted \( p(2/20) \), and so forth. Then the likelihood of the values of Table 10.6 is:

\[
L(CO = 20) = p(1/20) \cdot p(2/20) \cdot p(3/20) \cdot \ldots \cdot p(7/20) 
\]  

(10.41)

Similarly, equation 10.42 defines the likelihood for the assumption of a mean level of 25 ppm.

\[
L(CO = 25) = p(1/25) \cdot p(2/25) \cdot p(3/25) \cdot \ldots \cdot p(7/25) 
\]  

(10.42)

One might wish now to consider which of the assumptions is more likely: an average level of 20 ppm or 25 ppm. The more likely value is clearly going to be the one that yields the higher value of the likelihood. This notion leads to the extension of the concept of likelihood to hypothesis testing.

To test a hypothesis, you may recall from Chapter 4.6.2 that one may calculate the likelihood under the hypothesis and under some alternative hypothesis. Clearly, if the likelihood under the first hypothesis is larger, one would be inclined to accept that hypothesis in preference to the alternative hypothesis. This notion leads to several very important considerations.

First, it is clearly of considerable importance to construct two hypotheses for such tests to determine which hypothesis is the more likely and hence the more acceptable. Second, the choice of the alternative hypothesis is as important as that of the principal hypothesis. It is clear that if, in the example, one chose the alternative hypothesis as 200 ppm against the principal hypothesis of 20 ppm, the principal hypothesis would be accepted. However, if one chose 21 or 19 ppm, it is
no longer clear that the principal hypothesis would be accepted. The test is now more stringent.

Third, it is not sufficient simply to determine which of the two likelihoods is larger. One may raise the question of how much larger one likelihood must be to consider that the difference could not have been caused by chance. This is, of course, the principle underlying most statistical hypothesis testing, much of which has been utilised in earlier portions of the book. It is worthwhile exploring these considerations in more detail here since they led to the notion of maximum-likelihood estimation and its accompanying statistical tests, which are the concern of subsequent sections in this chapter.

Before developing the notions of likelihood and hypothesis testing further, it is worthwhile to see how one may calculate likelihoods. Return to the data of Table 10.6. It was noted that additional assumptions must be made to compute the likelihood. Essentially, it is necessary to assume there is some underlying distribution of the values of carbon monoxide, from which the samples of Table 10.6 were drawn. Suppose the distribution is assumed to be normal with a standard deviation of 6 ppm. Then the probability of obtaining one measurement, \( y \), of carbon monoxide is:

\[
p(1,y) = \frac{e^{((y-\mu)^2/2\sigma^2)}}{\sqrt{2\pi}\sigma}\]

(10.43)

Assuming the value of \( y \) to be 20 ppm, \( \sigma \) to be 6 ppm, and the values of Table 10.5, the likelihood can be estimated from:

\[
L(y_1,y_2,y_3,...,y_7 | y = 20) = \frac{e^{((y_1-\mu)^2/2\sigma^2)}}{\sqrt{2\pi}\sigma} \cdot \frac{e^{((y_2-\mu)^2/2\sigma^2)}}{\sqrt{2\pi}\sigma} \cdot \ldots \cdot \frac{e^{((y_7-\mu)^2/2\sigma^2)}}{\sqrt{2\pi}\sigma}.
\]

(10.44)

By gathering the terms, this may be simplified to:

\[
L(y_1,y_2,y_3,...,y_7 | y = 20) = \frac{e^{\sum_{i=1}^{7}((y_i-\mu)^2/2\sigma^2)}}{[(2\pi\sigma)^7]}
\]

(10.45)

Using the values listed above, the value of equation 10.45 is found to be \( 1.2435 \times 10^{-10} \). In a similar manner, one can calculate the likelihood under the assumption that \( y \) is 25 ppm, rather than 20 ppm. This yields an estimate of \( 7.2322 \times 10^{-11} \) which is smaller than the previous estimate. So one may conclude that it is less
likely that the values of Table 10.6 would have been obtained if the mean value was 25 ppm than if it was 20 ppm. The unresolved question here is whether the difference in the two values is significant. It is also important to note that, in this example, if one rejects the hypothesis of a mean value of 25 ppm, one accepts (or does not reject) the hypothesis of 20 ppm. However, this hypothesis may still be wrong, simply because both hypotheses chosen were wrong.

Two further points are worth noting here. It has been shown that one additional assumption is needed to compute the likelihood, that is, the distribution of the phenomenon being measured. Second, the likelihood is shown to be a very small value. This follows since the likelihood is a joint probability obtained by multiplying together all the individual probabilities for each observation. Since probabilities must lie between zero and one, it follows that the likelihood will be very small, decreasing in size as the sample size increases.

As noted earlier, maximum likelihood methods can be used to estimate the values of coefficients in an equation. The use of the methods in this way is known as Maximum Likelihood Estimation (MLE). The basic concept underlying MLE is to find the set of coefficients for the independent variables in the equation which is most likely to have given rise to the observed values \( y_1, y_2, y_3, ..., y_n \) of the dependent variable. Thus we wish to find:

\[
\max L[y_1, y_2, y_3, ..., y_n] \mid \Omega \]

where \( \Omega \) defines a region of possible sets of coefficients.

To find the maximum value of the likelihood function, \( L \), the standard procedure would be to differentiate the likelihood function with respect to \( \Omega \) and then equate this to zero:

\[
\frac{dL[y_1, y_2, y_3, ..., y_n]}{d\Omega} \bigg| \Omega = 0
\]

Naturally, it is important to ascertain that the solution of this differential equation is a maximum and not a minimum or a point of inflection.

It is reasonable at this point to ask why MLE estimates of coefficients are useful or desirable. To understand this, it is necessary to consider the desired properties of parameter estimates. In general, estimates of parameters should have four properties: consistency, lack of bias, efficiency, and sufficiency. A consistent estimator has an accuracy which increases with the size of the sample used to compute the estimate. thus an infinite sample would yield the exact value of the parameter. An unbiased estimator has a sampling distribution with a mean of the true value of the estimator. In other words, if several samples are taken and estimates made of the parameter from each separate sample, the mean of the
estimates will be the true value, or at least not statistically significantly different from the true value, of the parameter if the estimates are all unbiased. An efficient estimator exhibits a small variance for the sampling distribution of the unbiased estimator. In fact, the efficient estimator has the smallest variance of any estimator. The idea of a sufficient estimator is a complex concept, but essentially it means that it uses all the available information from a sample for the estimation of the parameter. It has been shown that MLE estimates are consistent, efficient, and sufficient, but not always unbiased (Ehrenfeld and Littauer, 1964); however, the bias decreases rapidly with increasing sample size. In addition, it is known that the distribution of the MLE estimate approaches that of a normal distribution with a known variance as the sample size increases, thus permitting the use of significance tests on the coefficient estimates obtained.

To illustrate the application of MLE to the estimation of a set of coefficients, consider a linear regression as described earlier in Chapter 10.2.1. Suppose, as hypothesised, that the error terms \( e_i \) are independent and drawn from a normal distribution. The likelihood function for the \( y_i \)'s \((y_1, y_2, y_3, \ldots, y_n)\) is therefore:

\[
L[(y_1, y_2, y_3, \ldots, y_n) \mid \beta] = \left( \frac{1}{\sqrt{2\pi}} \right)^n \exp \left( -\frac{1}{2} \sum_i e_i^2 \right) \tag{10.48}
\]

This follows because \( e_i \) is assumed to have a zero mean. Taking logs of both sides, the log likelihood is:

\[
\ln L = n \ln \left( \frac{1}{\sqrt{2\pi}} \right) - \frac{1}{2} \sum_i e_i^2 \tag{10.49}
\]

For a bivariate regression between \( y \) and \( x \), \( e_i^2 \) is:

\[
\sum_i e_i^2 = (y_i - a_1 x_i - a_0)^2 \tag{10.50}
\]

Substituting eqn. 10.50 into eqn. 10.49 yields:

\[
\ln L = n \ln \left( \frac{1}{\sqrt{2\pi}} \right) - \frac{1}{2} \sum_i (y_i - a_1 x_i - a_0)^2 \tag{10.51}
\]

To obtain MLE estimates of \( a_0 \) and \( a_1 \), the log-likelihood is differentiated with respect to each parameter, and the differentials set to zero.

\[
\frac{\partial \ln L}{\partial a_0} = \frac{1}{2} \sum_i [-2(y_i - a_1 x_i - a_0)] = 0 \tag{10.52}
\]
\[ \frac{\partial \ln L}{\partial a_1} = \frac{1}{2}\sum (-x_i)(y_i - a_1 x_i - a_0) = 0 \] (10.53)

Assuming that \( \sum \) is not equal to zero, these can be solved to yield:

\[ \sum (y_i - a_1 x_i - a_0) = 0 \] (10.54)
\[ \sum x_i (y_i - a_1 x_i - a_0) = 0 \] (10.55)

These are the identical estimating equations to those derived on the basis of least squares (see eqns. 10.11 and 10.12). Hence provided that \( \sum \) is a random normal variable with mean zero and variance \( \sum^2 \), the least-squares estimators are also the MLE estimators. In the above example, the second derivatives should also have been found to ensure that a maximum and not a minimum likelihood had been reached. In this example, since the likelihood function was linear, it was relatively easy to compute an analytical derivative. However, MLE is not bound to simple linear functions. It can be used with highly non-linear functions (as will be seen in the next section) provided that an iterative search procedure is used to find the maximum. In such situations, it is particularly important to check the second partial derivatives to see if one is climbing towards a maximum, away from it, or has landed at a point of inflection.

**10.2.6 Logit analysis**

There are many situations in transport data analysis and modelling where a simple linear model is not appropriate. This is particularly the case when trying to construct a probability model of choice (e.g. mode choice, route choice etc.). In such cases, the dependent variable (i.e. probability) must be bound between zero and one, and this necessitates the use of a non-linear model structure. Two major types of model have been proposed for this purpose; the probit model and the logit model.

The logit model appears to have been put forward originally by Berkson (1944) as a simplified version, in essence, of the probit model (which describes the response to a stimulus in terms of a cumulative normal distribution). At the time of Berkson’s work, probit models could be calibrated only by the tedious graphical methods described at length by Finney (1965). No further justification of the model form was given at the time, it being noted that the logit and probit models produced very similar symmetrical ogives.

In recent years, however, there has been a upsurge of interest in the logit model, particularly in consumer economics and transport planning. In this new work, theoretical developments of the logit model have been put forward, showing the model to be derivable from certain micro-economic assumptions of behaviour and from specific distributional assumptions on an error term. Since these
derivations are well-documented, they are not repeated in detail here (see, for example, McFadden and Domencich, 1975; Stopher and Meyburg, 1975). Suffice it to say that the model is developed in a utility context, in which it is assumed that people maximise their own utility in making a choice from a set of possible purchases or courses of action, and the utility is expressed by:

$$U(X_j, S_i) = U'(X'_j, S_i) + \epsilon_{ji}$$  \hspace{1cm} (10.56)

where $U(X_j, S_i) =$ the utility of alternative $j$ to individual $i$ as a function of characteristics, $x_j$, of the alternative and characteristic, $s_i$, of the individual

$U'(X'_j, S_i) =$ the observable utility for alternative $j$ to individual $i$,

$\epsilon_{ji} =$ the unobservable utility (or error)

The error term, $\epsilon_{ji}$, is assumed to be identically and independently distributed as a Weibull distribution (Weibull, 1951). Based on this assumption, the general form of the multinominal logit model can be derived as:

$$p^i_j = \frac{\prod e[U'(X'_j, S_i)]}{\prod_{k} e[U'(X'_k, S_i)]}$$  \hspace{1cm} (10.57)

where $p^i_j =$ the probability that individual $i$ chooses alternative $j$

and $k =$ subscript denoting any alternative from the set $K$

So far, the derivation does not specify the functional form of the observable utility, $U'(X'_j, S_i)$. In fact, the analyst may choose any functional form that seems useful or applicable. In the remainder of the chapter, however, the simplest functional form will be used, that is, a linear form.

$$U'(X'_j, S_i) = \epsilon_0^j + \sum_s \epsilon_s x'_s + \sum_t \epsilon_t s_i$$  \hspace{1cm} (10.58)

Thus the observable utility is assumed to consist of an alternative-specific dummy variable (coefficient $\epsilon_0^j$), a set of alternative-specific characteristics with generic coefficients ($\epsilon_s x'_s$), and a set of individual specific characteristics with alternative-specific coefficients ($\epsilon_t s_i$) (see Stopher and Meyburg, 1975).

The model developed above is known as the multinominal logit (MNL) model. Much of the early work in these models used a simpler form of the multinominal logit model. The MNL model is a general model that holds for any number (finite) of alternatives, $K$. A simple form that is also useful for exploring many of
the properties of the model is the binary logit model, shown for choice of alternative 1 as:

\[
p_1^i = \frac{e^{U'(X_1'S_i)}}{1 \cdot e^{U'(X_1'S_i)} + \cdot e^{U'(X_2'S_i)}}
\] (10.59)

Similarly, the binary logit model for the choice of alternative 2 is:

\[
p_2^i = \frac{e^{U'(X_2'S_i)}}{1 \cdot e^{U'(X_1'S_i)} + \cdot e^{U'(X_2'S_i)}}
\] (10.60)

This model form is used in much of the discussion in the remainder of the chapter because of its simple form.

Before proceeding further on the statistical properties of the model, there are certain other properties that should be explored since they affect the conceptual value of the logit model. First, both the multinominal and binary forms of the model, (see equations 10.57 and 10.59) are symmetrical. That is, the probability of an alternative being chosen is directly proportional to a function of the utility of the alternative and is inversely proportional to a sum of functions of all the utilities. This latter term will be the same for all alternatives. Hence the logit model is symmetrical.

However, the model is frequently expressed in an asymmetrical form. Suppose, in the binary case, one were to divide the numerator and the denominator of equation 10.59 by \(e^{U'(X_2'S_i)}\). This would produce equation 10.61 which is still identical in fact to equation 10.59.

\[
p_1^i = \frac{e^{U'(X_1'S_i)} - e^{U'(X_2'S_i)}}{e^{U'(X_1'S_i)} - e^{U'(X_2'S_i)}}
\] (10.61)

A similar manipulation of equation 10.60 yields:

\[
p_2^i = \frac{1}{e^{U'(X_1'S_i)} - e^{U'(X_2'S_i)}}
\] (10.62)

While equations 10.61 and 10.62 are in fact identical to equations 10.59 and 10.60, respectively, the appearance of symmetry has been lost. Given the equivalence of the models, it should also be noted that one could equally well have chosen to divide numerator and denominator by \(e^{U'(X_1'S_i)}\). For the multinominal case, any alternative may be selected as the base (i.e., the one used to divide numerator and denominator) and a general form produced as equations 10.63 and 10.64 where \(t\) is the arbitrary base alternative.
Several important concepts emerge from this process of manipulation. First, it reveals that the logit model is structured around the concept that choice (or probability of occurrence of an event) is determined by the difference between the respective utilities. Thus the logit model is based on an implicit decision rule that is revealed by this manipulation. Second, given a linear utility function, it should now be clear why the coefficients of the \( S_i \) characteristics would have no effect on the probabilities, since all would cancel out the difference formulation. The third concept derives from the model form used for calibration. As discussed later in the chapter, the model form for calibration is always that of equations 10.63 and 10.64. From this, it can be seen that all the alternative-specific dummy variables will be in difference form, that is, \( a_{0k} - a_{0t} \). Hence there are \( K \) alternatives in the dataset.

The second property of interest for the logit model concerns the shape of the curve that it yields. For the binary case, this curve is shown in Figure 10.34, where the model is:

\[
p_j^i = \frac{e^{(U_j^i)}}{e^{(U_k^i)}} \quad \text{where} \quad e^{(U_j^i)} = \frac{e^{U_j^i}}{e^{(U_k^i)} + e^{(U_k^i)}}
\]
If a probit curve is drawn for equation (10.66), this is found to produce a very similar curve, as shown in Figure 10.34.

\[ p_j^i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u_j} e^{-1/2u^2} \, du \]  

(10.66)

In fact, calibrating probit and logit curves on the same data produces curves that are generally statistically indistinguishable from each other (Stopher and Lavender, 1972). It must be noted, however, that the assumptions and concepts behind these two models are significantly different and these differences may have far-reaching effects upon the performance and use of the models. A more complete description of the calibration of the logit model using MLE and the use and interpretation of this type of model is beyond the scope of this book. The interested reader is referred to texts such as Stopher and Meyburg (1979), Hensher and Johnson (1980), and Ben-Akiva and Lerman (1985).
11. Finalising the Survey

Once the data has been collected and the analysis has been completed, there is a natural tendency for you to think that the project is nearly finished, and that it is time to move on to something new. However, there are two major tasks still to be completed. First, the results of the survey must be communicated effectively to the sponsors and other interested parties. Obviously there is little point to conducting surveys if this task is not handled well. Second, there is a need for you to make sure that the data are stored in a manner which ensures easy retrieval by others. While you may not want to make further use of the data right now, it is most likely that either you, or more likely someone else, will want to use the data in the future; therefore it has to be easily accessible. This ease of accessibility covers both the physical media on which the data is stored (tapes, disks) plus documentation which will allow someone else to make sense of the contents of the data set.

This chapter will address the issues and methods involved in the presentation of the results of the survey, in the storage of the data, and in the documentation of the survey method.
11.1 PRESENTATION OF RESULTS

It is a well-worn phrase that data is not the same as information. Data must be analysed, interpreted and presented properly before it can be classified as information. The analysis of data and the interpretation of results have been covered in the previous chapter; this chapter describes some of the techniques by which the information content of data may be effectively presented. In doing so, it draws upon two of the most readable and interesting books on the topic of graphical presentation of quantitative data that it has been my pleasure to read (Tufte, 1983, 1990), and which should be on the bookshelf of anyone involved in data collection and analysis. Tufte (1983) describes "data graphics" as devices which "visually display measured quantities by means of the combined use of points, lines, a coordinate system, numbers, symbols, words, shading, and color". Although data graphics are a relatively recent invention (about 200 years old), the recent explosion in multi-media computer graphics capabilities has given many people the illusion that they can create attractive and meaningful graphics. If you are one of these people, then you should buy and read Tufte's books quickly.

As an underpinning to data graphics, Tufte (1983) lists several principles of graphical excellence which help to communicate complex ideas with clarity, precision and efficiency. These principles are:

- show the data
- induce the viewer to think about the substance rather than about methodology, graphic design, the technology of graphic production, or something else
- avoid distorting what the data have to say
- present many numbers in a small space
- make large data sets coherent
- encourage the eye to compare different pieces of data
- reveal the data at several levels of detail, from a broad overview to the fine structure
- serve a reasonably clear purpose: description, exploration, tabulation, or decoration
- be closely integrated with the statistical and verbal descriptions of a data set.
Tufte states that "graphics reveal data", and in this he is closely aligned with the ideas of Tukey (1977) and the role of graphics in exploratory data analysis; the first rule of any data analysis is "plot the data". Thus, good graphics can be useful both in analysis and presentation; indeed, good graphical presentations encourage the reader (viewer) to conduct exploratory analysis beyond what is reported by the original analyst.

11.1.1 Avoid Distorted Graphics

One of Tufte's primary rules is to avoid the use of graphics which distort the data they are trying to represent. There are three classic distortions which unfortunately appear far too often in practice; the three-dimensional (3-D) pie-chart, the use of non-zero origins, and the use of multi-dimensional objects to represent uni-dimensional data values.

The 3-D pie-chart is often used to add "glamour" to a presentation which would normally only use a 2-D pie-chart. A comparison of 3-D and 2-D pie-charts which use the same data is shown in Figure 11.1. It is clear from the 2-D pie-chart that the values at the top, the right and the bottom are all the same size (they are each 8% of the total). However, in the 3-D pie-chart, it looks as though the 8% segment at the front (bottom) is bigger than the 8% segment at the back (top) which is, in turn, much bigger than the 8% segment on the right. The perspective view used by the 3-D pie-chart has distorted the data in the pie-chart (which is exactly what perspective views are supposed to do!). Unfortunately, many analysts use 3-D pie-charts without realising the distortion involved, because they know the values that they put into the chart and they don't see the distortion with an unbiased set of eyes. The distortion is particularly severe when value labels are not printed next to the pie segments, as in Figure 11.1.

![Figure 11.1 Comparison of 3-D and 2-D Pie-Charts](image-url)
Chapter 11

The second major type of distortion is the use of a non-zero value as the origin of the vertical axis of charts. The non-zero origin distortion is used so frequently that one would think that it is an accepted part of graphical presentations. An example is shown in Figure 11.2. By having the vertical axis start at a non-zero value (in this case, 396), the trend in the data is much more pronounced in the chart on the left than when it is plotted at natural scale on the right. While close inspection will show that the data in the two charts are the same, that is not the impression given at first glance. When such charts are used in audio-visual presentations, where the viewer does not have much time to read the axis labels in detail, a very false impression can be conveyed. Unfortunately, the chart on the left is the one that was produced automatically by a popular spreadsheet package; the chart on the right was only obtained by manual intervention.

![Figure 11.2 Comparison of Non-Zero and Zero Origin Charts](image)

The third major type of distortion is the use of multi-dimensional objects, such as circles, cylinders and spheres to represent a one-dimensional value. Tufte (1983, pp 69-73) gives some lovely examples, but a simple example will suffice. In transport, one often tries to represent populations or trip destinations in a graphical format. A method that is often used is to use circles centred on the location, the size of which are proportional to the number of trips (or the population of an area). An example is shown in Figure 11.3.

In the diagram on the left, the diameter of the circles is proportional to the number of trips, whereas in the diagram on the right the area of the circles is proportional to the number of trips. In both cases, the biggest circle is five times the smallest circle (either in diameter or area). In neither case, however, does the biggest circle look five times bigger. When diameters are used, it looks much more than five times bigger, and when areas are used it does not look as though it is five times bigger. Neither representation is correct, however, because we are trying to use a two-dimensional object (a circle) to represent a one-dimensional quantity (the number of trips). This problem is becoming more widespread as
Finalising the Survey

spreadsheet and graphing packages now allow the user to substitute multi-dimensional graphical objects for simple shading in histograms and other graphs.

![Figure 11.3 Representing One-Dimensional Quantities with 2-D Objects](image)

11.1.2 Maximise the Data-Ink Ratio

Tufte (1983) invented the concept of the data-ink ratio as a means of measuring whether a graphic was concentrating on the data content or on decoration. He defined "data-ink" as the "non-erasable core of a graphic, the non-redundant ink arranged in response to variation in the numbers represented". He then defines the data-ink ratio as:

\[
\text{Data-ink Ratio} = \frac{\text{data-ink}}{\text{total ink!used!to!print!the!graphic}}
\]

Two principles which then follow are to:

- maximise the data-ink ratio, within reason
- erase non-data-ink, within reason

Awareness of the concept of the data-ink ratio leads to the generation of graphics which are elegant in their simplicity, and which eschew the use of decoration for decoration's sake. It leads one to question whether or not certain lines, shadings, borders and text need to be on a graphic or not. If they can be removed without removing meaning from the graphic, then remove them. Compare the high data-ink ratio graph on the right of Figure 11.4 with the low data-ink ratio graph on the left. Both convey the same amount of information, but the graph on the right does so with much less ink. Its appearance is clean and distinctive.
11.1.3 Minimise Chart Junk

One of the major causes of a low data-ink ratio is the proliferation of “chart junk” on graphs and charts. Chart junk, or graphical decoration, has proliferated since the widespread availability of computer graphics packages. Major sources of chart junk are shading patterns, gridlines, and superfluous decorations (what Tufte terms “ducks”). Poor choice of shadings can result in severe Moiré vibration effects, where the patterns appear to be continually moving. Some examples of poor shading are shown in Figure 11.5.

Figure 11.5 Poor Choices of Shading Patterns

Figure 11.6 shows a combination of superfluous gridlines and a 3-D "duck" in the figure on the left. The gridlines on the side, rear and floor of the three-dimensional space do little to improve the clarity of the graph. This is compounded by the use of a 3-D ribbon graph which adds nothing to the
presentation of the data. Firstly, it is difficult without close inspection to determine which of the ribbons is at the front of the graph (i.e. which data belongs to series S1 and which belongs to S2). Secondly, it looks like the data series move in sympathy with each other in that they both appear to dip in the middle of the graph. However, examination of the simpler graph on the right shows that at the fourth data point along the horizontal axis, the two data sets are moving in opposite directions (series 1 is dipping while series 2 is peaking). A more accurate and easier to understand picture of the data is obtained by use of the simpler graphic.

![Graph with ribbons and simpler graph](image)

**Figure 11.6 Superfluous Gridlines and 3-D Effects**

### 11.1.4 Shape and Orientation of Graphs

Tufte (1983) notes that graphics should tend towards the horizontal, greater in length than in height. If there is a cause-effect relationship being displayed, then the cause should be plotted on the horizontal axis and the effect on the vertical axis (unlike an economist's demand curve). Labels should read from left to right, rather than having words stacked vertically, or having the words run from top to bottom or vice versa. While some prefer the ratio of the horizontal width to the vertical height to be that of the "Golden Section" (viz. 1.618), which is such that the ratio of the height to the width is the same as the width to the sum of the height and width, Tufte merely suggests that the width should be about 50% greater than the height.

### 11.1.5 The Friendly Data Graphic

Tufte (1983) encapsulates many of his principles of graphic design in what he terms the "friendly data graphic"; that is, one that looks as though the designer had the viewer in mind at every point while constructing the graphic. The features of the "friendly data graphic" are summarised in Table 11.1 by way of contrast with the unfriendly data graphic.
Chapter 11

Table 11.1 Characteristics of the Friendly Data Graphic

<table>
<thead>
<tr>
<th>Friendly</th>
<th>Unfriendly</th>
</tr>
</thead>
<tbody>
<tr>
<td>words are spelled out, mysterious codes are avoided</td>
<td>abbreviations abound, requiring the reader to find the explanations in the text</td>
</tr>
<tr>
<td>words run from left to right</td>
<td>words run vertically, particularly along the Y-axis; words run in several directions</td>
</tr>
<tr>
<td>messages are attached to explain the data</td>
<td>graphic is overly cryptic, requiring repeated references to the text</td>
</tr>
<tr>
<td>elaborate shadings, cross-hatchings and colours are avoided; labels are placed directly on the graphic, removing the need for a legend</td>
<td>obscure codings and shadings require repeated cross-referencing between the legend and the graphic</td>
</tr>
<tr>
<td>graphic attracts the viewer, promotes curiosity</td>
<td>graphic is overpowering, filled with chartjunk</td>
</tr>
<tr>
<td>if colours are used, they are chosen with the colour blind in mind; blue can generally be distinguished by most colour-deficient people</td>
<td>design is insensitive to colour-deficient viewers; red and green are used for essential contrasts</td>
</tr>
<tr>
<td>type is clear, precise and modest, with minimum number of fonts used</td>
<td>type is all bold, large and in many fonts</td>
</tr>
<tr>
<td>type is upper and lower case, with serifs</td>
<td>type is all capitals, sans serif</td>
</tr>
</tbody>
</table>

11.2 DOCUMENTATION OF SURVEY

The preparation and distribution of reports is the investigator’s primary means of conveying the results and methodology of the survey to interested parties. In preparing the survey report, four principles must be kept in mind. First, it is up to the investigator to explain fully the purpose and scope of the survey. It may be completely obvious to the investigator why the survey was carried out. However, it is almost certain that it is not as obvious to the reader of the report who is generally not as deeply immersed in the subject as the investigator. Without this background, it will be difficult for the reader to place the remainder of the report in its proper context.

Second, the investigator must keep in mind the type of reader(s) for whom the report is being written, the extent of their knowledge, the type of problem that is likely to be of interest to them and the kind of language to which they are accustomed. In many cases, a question-and-answer format for the reports will assist readers to find the information they seek from the report.
Finalising the Survey

Third, and related to the above, it is the responsibility of the investigator to translate statistical technicalities into language which will be understood by the reader who is primarily interested in the substantive results of the survey. Masses of standard errors, significance levels and the like are of no help if they are not understood in the context of the subject matter of the survey.

Fourth, non-quantified verbal descriptions of results are very useful in helping to develop understanding of the survey results. To many readers, statistical tables are dull and difficult to comprehend. A certain amount of verbatim quotation (obtained directly from open-ended questions), as well as verbal summaries of tables, helps to make the report easier to digest.

As regards the format and content of survey reports, there are a number of different formats that can be adopted. The United Nations recommendations for the preparation of sample survey reports serve as a useful guide. They recommend three types of report - a preliminary report, a general report and a technical report. Full details of these reports are described in Moser and Kalton (1979) but the major features of each are summarised below.

The preliminary report is often required to make available data of current interest as soon as possible. It should contain a brief statement concerning the survey methods and the limitations of the data. As a minimum requirement, information should be given concerning the size of the sample, the method of selecting the sample and any discrepancies observed between the sample and external data sources.

The general report should contain a general description of the survey for the use of those who are primarily interested in the results of the survey rather than in the technical aspects of the sample design, survey execution and analysis. Nonetheless, there must be sufficient description of the survey methodology to ensure that the survey results are not taken out of context or misinterpreted. The general report should include information on the following aspects of the survey:

(a) Statement of purposes of survey;
(b) Description of the sample coverage;
(c) Method of collection of information;
(d) Repetition details (if a continuing survey);
(e) Numerical results;
(f) Date and duration of survey;
(g) Description of accuracy;
(h) Cost of survey;
(i) Assessment of survey success;
(j) Responsibility for survey;
(k) References.
Chapter 11

The technical report should be issued for surveys of particular importance and those using new techniques and procedures of special interest. The report should deal in detail with technical statistical aspects of the sampling design, survey execution and analysis. It should include information on the following points:

(a) Specification of the sampling frame;
(b) Design of the sample;
(c) Personnel and equipment;
(d) Statistical analyses and computational procedures;
(e) Accuracy of results including discussion of random sampling errors and sampling bias;
(f) Adequacy of sampling frame;
(g) Comparisons with other sources of information;
(h) Costing analysis;
(i) Efficiency of survey method;
(j) Observations on survey staff.

Although the above procedures provide a useful way in which to document the survey, the authors believe that the framework adopted in this book provides an equally effective means of documenting a survey. Since the survey should have been designed using the procedures outlined in the preceding chapters, then the following outline should provide a useful means of writing up the survey documentation.

Chapter 1 Preliminary Planning

- Administrative Details of the Survey
  - the name of the survey?
  - who sponsored the survey?
  - who designed the survey?
  - who collected the survey data?
  - who analysed the survey data?
  - was there an Advisory Committee or Panel?
  - dates and duration of the survey?

- Overall Study Objectives
  - what were the objectives of the project to which this survey contributed?
  - why was a survey needed?

- Specific Survey Objectives
  - what were the specific objectives of this survey?

- Review of Existing Information
Finalising the Survey

- what prior information was available?
- what secondary information was available for sample expansion?

• Formation of Hypotheses
  - what specific hypotheses, if any, were to be tested?

• Definition of Terms
  - what definitions are being used by the survey team for key items such as trip, household, mode, income etc. (as relevant to the specific survey)?

• Determination of Survey Resources
  - what time was available for completion of the survey?
  - how much money was available for the survey?
  - what people were available to work on the survey?

Chapter 2 Selection of Survey Method

• Selection of Survey Time Frame
  - was the survey cross-sectional or time-series (and why)?

• Selection of Survey Technique
  - what methods were considered for the survey technique?
  - what testing was performed on the different methods?
  - what method was finally selected (and why)?

Chapter 3 Sample Design

• Definition of Target Population
  - what was the population for the survey?
  - how was this population defined and identified?

• Sampling Units
  - what unit was used for sampling?

• Sampling Frame
  - what sampling frame was used?
  - where was the sampling frame obtained from?
  - how was the sampling frame obtained?
  - why was the sampling frame first compiled?
  - how did the sampling frame perform in term of:
    - accuracy
    - completeness
    - duplication
    - adequacy
    - up-to-dateness
Chapter 11

- **Sampling Method**
  - what sampling methods were considered?
  - what sampling method was finally chosen (and why)?
  - was the selected sample representative of the population?
    - if not, how will this be corrected later?
  - what was the specific sampling procedure (full details)?

- **Consideration of Sampling Bias**
  - what sources of sampling bias were considered?
  - how serious were these biases considered to be?
  - what steps were taken to overcome these sources of bias?

- **Sample Size and Composition**
  - what was the final sample size?
  - what stratifications were used in the sample design?
  - how was the sample size calculated?
    - what were the key variables considered?
    - what was the variability of these variables?
    - what confidence limits were used?
    - what levels of confidence were used?

- **Estimation of Parameter Variances**
  - how are parameter variances to be estimated in the data analysis?

- **Conduct of Sampling**
  - what procedure was used in selecting the sample?
  - was random sampling used at all stages of sampling?

Chapter 4 Survey Instrument Design

- **Question Content**
  - what are types of information being sought in the survey?

- **Trip Recording Techniques**
  - how are trips and activities being sought from respondents?

- **Physical Nature of Forms**
  - what is the physical nature of the survey forms?
    - what paper size and weight was used?
    - what colours and printing methods were used?

- **Question Types**
  - what classification questions were asked?
    - where did the classification categories come from?
  - what attitude questions were asked?
    - what testing was performed on the attitude scales?
Finalising the Survey

• Question Format
  - which questions were asked as open questions (and why)?
  - which questions were asked as closed questions (and why)?
    - where did the closed question categories come from?

• Question Wording
  - how has the question wording been tested for:
    - simple vocabulary
    - words appropriate to the audience
    - length of questions
    - ambiguous questions (get someone else to read them)
    - double-barrelled questions
    - vague words
    - loaded questions
    - leading questions
    - double negatives
    - stressful questions
    - grossly hypothetical questions
    - the effect of response styles
    - periodicity questions

• Question Ordering
  - what reasons are there for the question ordering?

• Question Instructions
  - what instructions were provided for respondents/interviewers?

Chapter 5 Pilot Survey(s)

• Description of Pilot Surveys
  - what pilot testing was performed?
  - if no pilot testing was done, why not?

• Size of the Pilot Survey

• Lessons from the Pilot Survey
  - how adequate was the sampling frame?
  - what was the variability within the survey population?
  - what response rate was achieved?
  - how suitable was the survey method?
  - how well did the questionnaire perform?
  - how effective was the interviewer training?
  - did the coding, data entry, editing and analysis procedures work satisfactorily?

• Cost and Duration of Pilot Surveys
Chapter 11

Chapter 6 Administration of the Survey

- Survey Procedures
  - Self-Completion Questionnaires
    - pre-contact procedures
    - mail-out procedures
    - response receipt procedures
    - phone enquiry procedures
    - postal reminder regime
    - telephone follow-ups
    - validation interviews
    - non-response interviews
  - Personal Interviews
    - pre-contact procedures
    - call-back procedures
    - maintenance of survey logs
    - interviewer payment methods
    - field supervisor tasks
    - work distribution procedures
  - Telephone Interviews
    - sampling procedures
    - dealing with non-response
    - use of CATI systems
  - Intercept Surveys
    - procedures for obtaining respondents
    - distribution of surveys
    - collection of surveys
  - In-depth Interview Surveys
    - pre-contact procedures
    - call-back procedures
    - maintenance of survey logs
    - recording methods
    - transcription methods
    - interpretation of responses

Chapter 7 Data Processing

- Selection of Coding Method
  - what physical method was used for data coding?

- Preparation of Code Format
  - what coding frame was used?
    (provide full coding frame in Appendix)
  - what location-coding method was used?
Finalising the Survey

- Development of Data Entry Programs
  - what special data entry programs were developed?
    (provide screen-shots of data entry screens in Appendix)

- Coder and Data Entry Training
  - what training was provided for coders and data enterers?
    (provide training manual in Appendix)

- Coding Administration
  - how was the coding administered?
  - what quality control procedures were implemented?
  - how were changes made to coding frames?

Chapter 8 Data Editing

- Initial Questionnaire Editing
  - what in-field checking was done by interviewer/supervisor?
  - what checking was done on receipt in survey office?

- Verification of Data Entry
  - was data entry verified for accuracy?

- Development of Editing Computer Programs
  - were special data editing programs developed?

- Consistency and Range Checks
  - what permissible range checks were applied?
    (provide full list of checks in Appendix)
  - what logic checks were applied?
    (provide full list of checks in Appendix)

- Missing Data
  - how was missing data coded?
  - were estimates made of missing values?

Chapter 9 Data Correction and Expansion

- Editing Check Corrections
  - what procedures were used for office edits?

- Secondary Data Comparisons
  - what secondary data was used for sample expansion?
  - what variables were used for expansion purposes?
  - was expansion based on cross-tabulations or marginal totals?
  - what were the final expansion factors?
  - how are they to be applied when using the data?
Chapter 11

- Corrections for Internal Biases
  - what recognition was there of non-reported data?
  - were non-reporting factors calculated?
  - if so, how are they to be applied to the data?
  - what recognition was there of non-response?
  - were non-response factors calculated?
  - if so, how are they to be applied to the data?

Chapter 10  Data Analysis and Management

- Exploratory Data Analysis
  - what EDA methods were used?

- Model Building
  - is the data to be used to build specific models?

- Interpretation of Results
  - are any limitations on the data clearly stated?
  - how is the sampling error expressed?

- Database Management
  - is the structure of the datafiles clearly described?
  - are the relationships between data files clear?

- Provision of Data Support Services
  - what support is available for users of the data?
  - is it clear where such support can be obtained?

Chapter 11  Presentation of Results

- Presentation of Results of Analysis
  - are the major descriptive results presented:
    - in a clear visual manner?
    - with accompanying written explanations?
    - with appropriate interpretations?
    - and with clear statement of any qualifications?

- Publication of Results
  - are the results of the survey or the survey methodology written up in concise form, and available in the general literature?
Chapter 12  Tidying-Up

- Storage and Archival of Data
  - where is the data stored?
  - who is the contact person?
  - are telephone, fax and e-mail numbers provided?
  - is this documentation stored electronically with the data?
  - has the data been lodged with any other archival service?

- Completion of Administrative Duties
  - have all survey staff been fully paid?
  - have all outstanding bills been paid?
  - what arrangements have been made for destroying original
    questionnaires?

These same headings are reproduced, using a question and answer format, in Appendix 3 in the form of a Survey Design Checklist document that can be copied and used to assist in the planning and execution of travel and activity surveys.

In addition to the preparation of reports by conventional means, consideration should also be given to the preparation of verbal, visual and electronic presentations describing the survey and its results. Recent advances in "desktop presentation" techniques on widely available microcomputers promise a much more effective means of communication than conventional written reports. Programs such as Hypercard on the Apple Macintosh and presentation software such as Microsoft Powerpoint enable the compilation of audio-visual presentations which can effectively and dramatically convey the results of the survey to decision-makers in a readily understandable and flexible fashion. Wigan (1988) has taken steps in this direction with a prototype system based on Australian transportation surveys; further work is currently being performed by the authors. It is expected that in the next few years, graphics-based microcomputers will be playing a major role in the design, analysis and presentation of results from most transportation surveys.
11.3 TIDYING-UP

The final task in the survey process is the often thankless task of tidying-up. The effective tidying-up of the data is the only way in which the data will be available for secondary analysis at a later date (Hyman 1972; Wigan 1985).

In discussing the requirements of data for secondary analysis, Wigan (1985) raises, among many other issues, the following points:

(a) The distribution of data for secondary analysis has been greatly improved by the establishment of the Inter-University Consortium for Political and Social Research (ICPSR) at the Institute for Social Research in Ann Arbor, Michigan. This institute gathers data sets and arranges for their storage, maintenance and distribution according to pre-set standards.

(b) The availability of data for secondary analysis is often restricted because of confidentiality requirements imposed on the data by the original sponsor of the survey. Key data items may be omitted from the data set after primary analysis, or the data may be aggregated in such a way that prevents meaningful secondary analysis.

(c) As mentioned earlier, considerable effort can be expended in the estimation of weighting factors which enable the sample data to more closely represent the population from which it was drawn. In many cases, these weighting factors are not stored with the original data, with the result that the person doing secondary analysis must repeat this time-consuming process.

(d) Because of the many adjustments which may be made to a data file during primary analysis, it is usual for several versions of a data file to exist at any one time. It is essential to know which version has been archived, and to ensure that the documentation which has been archived matches the version of the data file which has been archived.

(e) It is essential that the code books and other documentation be archived with the data tape, so that definitions of variables and the codes used are clearly defined.

The importance of tidying-up is only really appreciated when you attempt to use someone else's data and realise how much you wish they had documented their survey technique, arranged for storage of the data with enough information to permit easy access, and let someone know where the data was stored and who had continuing responsibility for its upkeep. The golden rule in tidying-up is

"Do unto others as you would have them do unto you".
Survey Methods for Transport Planning

List of References


Ampt ES (1981). Some Recent Advances in Large Scale Travel Surveys. Presented at the PTRC 9th Summer Annual Meeting, University of Warwick, UK.


Baughan CJ (1979). Public Attitudes to Alternative Sizes of Lorry - a Study in a Residential Area. Supplementary Report 509, Transport & Road Research Laboratory, Crowthorne UK.


Bhat C (1994). Estimation of Travel Demand Models with Grouped and Missing Income Data. Presented at the 7th International Conference on Travel Behaviour Research, Santiago, Chile.


Brög W (1982). 'Application of the Situational Approach to Depict a Model of Personal Long-Distance Travel'. Transportation Research Record, 890, 24-33.


Clark AC & Goldstucker C (1986). 'Mail-Out/Mail-Back Travel Survey in Houston Texas'. Transportation Research Record, 1097, 13-20.


Cooper BR & Layfield RE (1978). Nottingham Zones and Collar Study - Results of the "after" Surveys. Supplementary Report 365, Transport & Road Research Laboratory, Crowthorne UK.


Das M (1978). Collection and Dissemination of Transport Information in a Local Authority. Laboratory Report 815, Transport & Road Research Laboratory, Crowthorne UK.


Difiglio C & Disbrow JA (1981). 'Impact of Travel Survey Sampling Error on Travel Demand Forecasting'. Transportation Research Record, 815, 31-40.


Dixon CJ & Leach B (1979), Questionnaires and Interviews in Geographic Research. Concepts and Techniques in Modern Geography No 18, Study Group in Quantitative Methods Institute of British Geographers University of East Anglia, Norwich UK.


Dreisbach AF (1980). 'Small-scale, Ongoing Home-interview Survey in Pennsylvania'. Transportation Research Record, 779, 10-16.


Eash R (1987). 'Management of a Small Home Interview Travel Survey Using a Microcomputer'. Transportation Research Record , 1134, 49-56.


Journal of Marketing Research, 14, 3, 280-293.


Fulton PN (1984). 'Allocating Incomplete Place-of-Work Responses in the 1980 Census 
Urban Transportation Planning Package'. Transportation Research Record, 981, 15-21.

Supplementary Report 633, Transport & Road Research Laboratory, Crowthorne UK.

Galin D (1975). 'How Reliable are Mail-Surveys and Response Rates: A Literature Review'. 
Journal of Marketing Research, 12, 440-453.


Garber NJ & Bayat-Mokhtari F (1986). 'A Computerized Highway Link Classification 


Attributes of Transportation Alternatives'. Transportation Research Board, Special 
Report 149, 58-81.

Panel. Presented at the International Conference on Dynamic Travel Behaviour 
Analysis, Kyoto University, Japan.

Speed Ground Transportation Journal, 4, 103-116.

Golob TF (1973). 'The Development of Attitudinal Models of Travel Behavior'. Highway 
Research Board, Special Report 143, .

for a Public Transportation System'. Transportation Research, 6, 81-102.

Golob TF, Dobson R & Sheth JN (1973). Perceived Attribute Importance in Public and 
Private Transportation Proceedings. Presented at the 5th Annual Meeting of the 
American Institute for Decision Sciences, Washington, DC.

Golob TF, Horowitz AD & Wachs M (1979). Attitude-Behavior Relationships in Travel 
Demand Modeling. In DA Hensher & PR Stopher (Eds), Behavioural Travel Modelling. 
Croom Helm: London.


Hardman EJ & Walker MJ (1979). Anglo-Scottish Car Travel from Surveys on the A74 Road near Lockerbie. Laboratory Report 856, Transport & Road Research Laboratory, Crowthorne UK.


Jones PM (1980). 'Experience with Household Activity-Travel Simulator (HATS)'. Transportation Research Record , 765, 6-12.


Kroes E & Sheldon R (1988). Are there any Limits to the Amount Consumers are Prepared to Pay for Product Improvements?. Presented at the 15th PTRC Summer Annual Meeting, The University of Bath, UK.


Liss S (1986). 'Nationwide Personal Transportation Study: Experiences with Previous Surveys and Options for the Future'. Transportation Research Record, 1097, 31-33.


Lomax DE & Downes JD (1977). Patterns of Travel to School and Work in Reading in 1971. Laboratory Report 808, Transport & Road Research Laboratory, Crowthorne UK.


Mackie AM & Griffin LJ (1977). Before and After Study of the Environmental Effects of Tring Bypass. Laboratory Report 746, Transport & Road Research Laboratory, Crowthorne UK.


Miller GA (1956). 'The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information'. Psychological Review , 63, 2, 81-97.


Norris BB & Shunk GA (1986). 'Special-Purpose Travel Surveys'. Transportation Research Record 1097, 1-23.


Phifer SP (1982). 'Use of Sampling in Bus Line Data Collection'. Transportation Research Record, 877, 41-44.


Quenault SW (1979). Cycle Routes in Portsmouth III - Attitude Surveys. Laboratory Report 875, Transport & Road Research Laboratory, Crowthorne UK.


Richardson AJ & Ampt ES (1994). Non-Response Effects in Mail-Back Travel Surveys. Presented at the 7th International Conference on Travel Behaviour Research, Santiago, Chile.


Richardson AJ (1986). The Correction of Sample Bias in Telephone Interview Travel Surveys. Presented at the 65th Annual Meeting of the Transportation Research Board, Washington, DC.


Rigby JP (1977). An Analysis of Travel Patterns using the 1972/73 National Travel Survey. Laboratory Report 790, Transport & Road Research Laboratory, Crowthorne UK.


Slavik MM (1986). 'Errors in Origin-Destination Surveys done by Number-Plate Technique'. Transportation Research Record, 1050, 46-53.


Smith ME (1979). 'Design of Small-Sample Home-Interview Travel Surveys'. Transportation Research Record, 701, 29-35.


Stopher PR & Sheskin IM (1982). 'Toward Improved Collection of 24-Hour Travel Records'. Transportation Research Record , 891, 10-17.


Stopher PR (1992). 'Use of an Activity-Based Diary to Collect Household Travel Data'. Transportation , 19, 2, 159-176.


Tanner JC (1981). Methods of Forecasting Kilometres per Car. Laboratory Report 968, Transport & Road Research Laboratory, Crowthorne UK.


Appendix A

Survey Design Checklist

This document should be used as a checklist to ensure that all aspects of the survey design process have been addressed.

In most cases, a short answer will suffice to demonstrate that the particular point has, indeed, been addressed. In other cases, more detailed information may be provided, or attached, and this checklist may then serve as a useful shorthand form of survey documentation. It may also be used as a guideline for writing more comprehensive documentation for the survey.

Where appropriate, page references are given to a text on Travel Survey Methods (Richardson, Ampt and Meyburg (1995), Survey Methods for Transport Planning, Eucalyptus Press: Melbourne) to help you understand what is required to be answered in this Checklist.
**Section 1  Preliminary Planning**

**Administrative Details of the Survey**

<table>
<thead>
<tr>
<th>The name of the survey?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Who sponsored the survey?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Who designed the survey?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Who collected the survey data?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Who analysed the survey data?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

Was there an advisory committee?  YES ☐  NO ☐

If so, who was on the Committee?

<table>
<thead>
<tr>
<th>Dates and duration of the survey?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>
**Survey Design Checklist**

**Overall Study Objectives (p17)**

What were the objectives of the project to which this survey contributed?

<table>
<thead>
<tr>
<th>Why was a survey needed?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

**Specific Survey Objectives (p24)**

What were the specific objectives of this survey?

<table>
<thead>
<tr>
<th>What were the specific objectives of this survey?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>
## Survey Design Checklist

### Review of Existing Information (p26)

What prior information was available?

What secondary information was available for sample expansion?

### Formation of Hypotheses (p28)

What specific hypotheses, if any, were to be tested?
<table>
<thead>
<tr>
<th>Survey Design Checklist</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Definition of Terms (p28)</strong></td>
</tr>
<tr>
<td>What definitions are being used by the survey team for key items such as: trip, household, mode, income etc. (as relevant to the specific survey)?</td>
</tr>
</tbody>
</table>

| **Determination of Survey Resources (p30)** |
| What time was available for completion of the survey? |
| How much money was available for the survey? |
| What people were available to work on the survey? |
## Section 2 Selection of Survey Method

### Selection of Survey Time Frame (p34)

Was the survey cross-sectional or time-series (and why)?

- [ ] Cross-sectional
- [ ] Time Series

Why?

---

### Selection of Survey Technique (p42)

What methods were considered for the survey technique?

What testing was performed on the different methods?

What method was finally selected (and why)?
Section 3   Sample Design

Definition of Target Population (p75)

What was the population for the survey?

How was this population defined and identified?

Sampling Units (p76)

What unit was used for sampling?

Sampling Frame (p77)

What sampling frame was used?

Where was the sampling frame obtained from?

How was the sampling frame obtained?
## Survey Design Checklist

<table>
<thead>
<tr>
<th>Why was the sampling frame first compiled?</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>How did the sampling frame perform in term of:</th>
</tr>
</thead>
<tbody>
<tr>
<td>accuracy</td>
</tr>
<tr>
<td>completeness</td>
</tr>
<tr>
<td>duplication</td>
</tr>
<tr>
<td>adequacy</td>
</tr>
<tr>
<td>up-to-dateness</td>
</tr>
</tbody>
</table>

## Sampling Method (p80)

<table>
<thead>
<tr>
<th>What sampling methods were considered?</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>What sampling method was finally chosen (and why)?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey Design Checklist</td>
</tr>
<tr>
<td>------------------------</td>
</tr>
<tr>
<td>Was the selected sample representative of the population?</td>
</tr>
<tr>
<td>If not, how will this be corrected later?</td>
</tr>
<tr>
<td>What was the specific sampling procedure (full details)?</td>
</tr>
</tbody>
</table>
Consideration of Sampling Bias (p96)

What sources of sampling bias were considered?

How serious were these biases considered to be?

What steps were taken to overcome these sources of bias?
## Sample Size and Composition (p101)

<table>
<thead>
<tr>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>What was the final sample size?</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>What stratifications were used in the sample design?</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>How was the sample size calculated?</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>what were the key variables considered?</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>what was the variability of these variables?</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>what confidence limits were used?</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>what levels of confidence were used?</td>
</tr>
</tbody>
</table>
Survey Design Checklist

Estimation of Parameter Variances (p126)

How are parameter variances to be estimated in the data analysis?

Conduct of Sampling (p142)

What procedure was used in selecting the sample?

Was random sampling used at all stages of sampling?
**Section 4   Survey Instrument Design**

<table>
<thead>
<tr>
<th>Question Content (p151)</th>
</tr>
</thead>
<tbody>
<tr>
<td>What types of information are being sought in the survey?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trip Recording Techniques (p155)</th>
</tr>
</thead>
<tbody>
<tr>
<td>How are trips and activities being sought from respondents?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Physical Nature of Forms (p159)</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is the physical nature of the survey forms?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>what paper size and weight was used?</th>
</tr>
</thead>
<tbody>
<tr>
<td>what colours and printing methods were used?</td>
</tr>
</tbody>
</table>

---
<table>
<thead>
<tr>
<th>Question Types (p166)</th>
</tr>
</thead>
<tbody>
<tr>
<td>What classification questions were asked?</td>
</tr>
<tr>
<td>where did the classification categories come from?</td>
</tr>
<tr>
<td>What attitude questions were asked?</td>
</tr>
<tr>
<td>what testing was performed on the attitude scales?</td>
</tr>
</tbody>
</table>
### Survey Design Checklist

#### Question Format (p187)

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which questions were asked as open questions (and why)?</td>
<td></td>
</tr>
<tr>
<td>Which questions were asked as closed questions (and why)?</td>
<td></td>
</tr>
<tr>
<td>where did the closed question categories come from?</td>
<td></td>
</tr>
</tbody>
</table>
Survey Design Checklist

Question Wording (p194)

How has the question wording been tested for:

- simple vocabulary  YES □
- words appropriate to the audience  YES □
- length of questions  YES □
- ambiguous questions (get someone else to read them)  YES □
- double-barrelled questions  YES □
- vague words  YES □
- loaded questions  YES □
- leading questions  YES □
- double negatives  YES □
- stressful questions  YES □
- grossly hypothetical questions  YES □
- the effect of response styles  YES □
- periodicity questions  YES □

Question Ordering (p205)

What reasons are there for the question ordering?
Survey Design Checklist

Question Instructions (p207)

What instructions were provided for respondents/interviewers?
# Section 5   Pilot Survey(s)

**Description of Pilot Surveys (p214)**

<table>
<thead>
<tr>
<th>What pilot testing was performed?</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>If no pilot testing was done, why not?</th>
</tr>
</thead>
</table>

**Size of the Pilot Survey (p222)**
### Survey Design Checklist

**Lessons from the Pilot Survey (p216)**

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>How adequate was the sampling frame?</td>
<td></td>
</tr>
<tr>
<td>What was the variability within the survey population?</td>
<td></td>
</tr>
<tr>
<td>What response rate was achieved?</td>
<td></td>
</tr>
<tr>
<td>How suitable was the survey method?</td>
<td></td>
</tr>
<tr>
<td>How well did the questionnaire perform?</td>
<td></td>
</tr>
<tr>
<td>How effective was the interviewer training?</td>
<td></td>
</tr>
<tr>
<td>Did the coding, data entry, editing and analysis procedures work satisfactorily?</td>
<td></td>
</tr>
</tbody>
</table>

**Cost and Duration of Pilot Surveys (p221)**

<table>
<thead>
<tr>
<th>Component</th>
<th>Cost</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Section 6  Administration of the Survey

### Survey Procedures

<table>
<thead>
<tr>
<th>Self-Completion Questionnaires (p239)</th>
</tr>
</thead>
<tbody>
<tr>
<td>- pre-contact procedures</td>
</tr>
<tr>
<td>- mail-out procedures</td>
</tr>
<tr>
<td>- response receipt procedures</td>
</tr>
<tr>
<td>- phone enquiry procedures</td>
</tr>
<tr>
<td>- postal reminder regime</td>
</tr>
<tr>
<td>- telephone follow-ups</td>
</tr>
<tr>
<td>- validation interviews</td>
</tr>
<tr>
<td>- non-response interviews</td>
</tr>
</tbody>
</table>
## Survey Design Checklist

<table>
<thead>
<tr>
<th>Personal Interviews (p246)</th>
</tr>
</thead>
<tbody>
<tr>
<td>- pre-contact procedures</td>
</tr>
</tbody>
</table>

| - call-back procedures   |
|                          |

<table>
<thead>
<tr>
<th>- maintenance of survey logs</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>- interviewer payment methods</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>- field supervisor tasks</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>- work distribution procedures</th>
</tr>
</thead>
</table>
**Telephone Interviews (p251)**

- sampling procedures

- dealing with non-response

- use of CATI systems
Survey Design Checklist

Intercept Surveys (p255)
- procedures for obtaining respondents

- distribution of surveys

- collection of surveys
## Survey Design Checklist

**In-depth Interview Surveys (p258)**

- pre-contact procedures

- call-back procedures

- maintenance of survey logs

- recording methods

- transcription methods

- interpretation of responses
## Section 7 Data Coding

### Selection of Coding Method (p266)

What physical method was used for data coding?

### Preparation of Code Format (p268)

What coding frame was used?

(Provide full coding frame in Appendix)

What location-coding method was used?

### Development of Data Entry Programs (p292)

What special data entry programs were developed?

(Provide screen-shots of data entry screens in Appendix)
## Survey Design Checklist

### Coder and Data Entry Training (p290)

What training was provided for coders and data enterers?

(Provide training manual in Appendix)

### Coding Administration (p290)

How was the coding administered?

What quality control procedures were implemented?

How were changes made to coding frames?
# Survey Design Checklist

## Section 8  Data Editing

### Initial Questionnaire Editing (p299)

<table>
<thead>
<tr>
<th>What in-field checking was done by interviewer/supervisor?</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>What checking was done on receipt in survey office?</th>
</tr>
</thead>
</table>

### Verification of Data Entry (p299)

| Was data entry verified for accuracy? YES ☐ NO ☐ |

### Development of Editing Computer Programs (p299)

| Were special data editing programs developed? YES ☐ NO ☐ |

<table>
<thead>
<tr>
<th>If so, in what language were they written?</th>
</tr>
</thead>
</table>

### Consistency and Range Checks (p299)

| What permissible range checks were applied? |

| (provide full list of checks in Appendix) |

<table>
<thead>
<tr>
<th>What logic checks were applied?</th>
</tr>
</thead>
</table>

| (provide full list of checks in Appendix) |
## Survey Design Checklist

### Missing Data (p299)

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>How was missing data coded?</td>
<td></td>
</tr>
<tr>
<td>Were estimates made of missing values?</td>
<td></td>
</tr>
</tbody>
</table>
## Section 9  Data Correction and Expansion

### Editing Check Corrections (p299)

What procedures were used for office edits?

### Secondary Data Comparisons (p307)

What secondary data was used for sample expansion?

What variables were used for expansion purposes?

Was expansion based on cross-tabulations or marginal totals?

- Cross-tabs ☐
- Marginals ☐

What were the final expansion factors?

(provide full list of expansion factors in Appendix)

How are they to be applied when using the data?
**Survey Design Checklist**

**Corrections for Internal Biases**

What recognition was there of non-reported data? (p313)

<table>
<thead>
<tr>
<th>Were non-reporting factors calculated?</th>
<th>YES □</th>
<th>NO □</th>
</tr>
</thead>
<tbody>
<tr>
<td>If so, how are they to be applied to the data?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

What recognition was there of non-response? (p321)

<table>
<thead>
<tr>
<th>Were non-response factors calculated?</th>
<th>YES □</th>
<th>NO □</th>
</tr>
</thead>
<tbody>
<tr>
<td>If so, how are they to be applied to the data?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Section 10  Data Analysis and Management

### Exploratory Data Analysis (p339)

What EDA methods were used?

### Model Building (p346)

Is the data to be used to build specific models?  
- YES ☐  
- NO ☐

If so, what type?

### Interpretation of Results (p395)

Are any limitations on the data clearly stated? YES ☐  NO ☐

How?

How is the sampling error expressed?
### Survey Design Checklist

#### Database Management (p268)

<table>
<thead>
<tr>
<th>Question</th>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is the structure of the datafiles clearly described?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are the relationships between data files clear?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Provide full list of datafiles in Appendix)

#### Provision of Data Support Services (p402)

<table>
<thead>
<tr>
<th>Question</th>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>What support is available for users of the data?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is it clear where such support can be obtained?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Provide full set of relationships in Appendix)
Section 11  Presentation of Results

Presentation of Results of Analysis (p396)

Are the major descriptive results presented:

- in a clear visual manner?  YES □  NO □

- with accompanying written explanations?  YES □  NO □

- with appropriate interpretations?  YES □  NO □

- and with clear statement of any qualifications?  YES □  NO □

Publication of Results (p402)

Are the results of the survey or the survey methodology written up in concise form, and available in the general literature?

YES □  NO □

If so, where?
## Section 12  Tidying-Up

### Storage and Archival of Data (p412)

<table>
<thead>
<tr>
<th>Question</th>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Where is the data stored?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Who is the contact person?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Position:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organisation:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Address:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are telephone, fax and e-mail numbers provided?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telephone:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fax:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e-mail:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is this documentation stored electronically with the data?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Has the data been lodged with any other archival service?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Have all survey staff been fully paid?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Have all outstanding bills been paid?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>What arrangements have been made for destroying original questionnaires?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>