Contents

Preface xi

1 Statistics in medical research 1
  1.1 Statistics at large 1
  1.2 Statistics in medicine 3
  1.3 Statistics in medical research 4
  1.4 What does statistics cover? 5
  1.5 The scope of this book 8

2 Types of data 10
  2.1 Introduction 10
  2.2 Categorical data 10
  2.3 Numerical data 11
  2.4 Other types of data 13
  2.5 Censored data 16
  2.6 Variability 17
  2.7 Importance of the type of data 17
  2.8 Dealing with numbers 17

3 Describing data 19
  3.1 Introduction 19
  3.2 Averages 21
  3.3 Describing variability 22
  3.4 Quantifying variability 31
  3.5 Two variables 38
  3.6 The effect of transforming the data 41
  3.7 Data presentation 42

  Exercises 45

4 Theoretical distributions 48
  4.1 Introduction 48
  4.2 Probability 49
  4.3 Samples and populations 50
  4.4 Probability distributions 50
  4.5 The Normal distribution 51
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.6</td>
<td>The Lognormal distributions</td>
<td>60</td>
</tr>
<tr>
<td>4.7</td>
<td>The Binomial distribution</td>
<td>63</td>
</tr>
<tr>
<td>4.8</td>
<td>The Poisson distribution</td>
<td>66</td>
</tr>
<tr>
<td>4.9</td>
<td>Mathematical calculations</td>
<td>68</td>
</tr>
<tr>
<td>4.10</td>
<td>The Uniform distribution</td>
<td>71</td>
</tr>
<tr>
<td>4.11</td>
<td>Concluding remarks</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>Exercises</td>
<td>71</td>
</tr>
<tr>
<td>5</td>
<td>Designing research</td>
<td>74</td>
</tr>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>74</td>
</tr>
<tr>
<td>5.2</td>
<td>Categories of research design</td>
<td>75</td>
</tr>
<tr>
<td>5.3</td>
<td>Sources of variation</td>
<td>78</td>
</tr>
<tr>
<td>5.4</td>
<td>An experiment: is the blood pressure the same in both arms?</td>
<td>79</td>
</tr>
<tr>
<td>5.5</td>
<td>The design of experiments</td>
<td>80</td>
</tr>
<tr>
<td>5.6</td>
<td>The structure of an experiment</td>
<td>83</td>
</tr>
<tr>
<td>5.7</td>
<td>Random allocation</td>
<td>85</td>
</tr>
<tr>
<td>5.8</td>
<td>Minimization</td>
<td>91</td>
</tr>
<tr>
<td>5.9</td>
<td>Observational studies</td>
<td>91</td>
</tr>
<tr>
<td>5.10</td>
<td>The case-control study</td>
<td>93</td>
</tr>
<tr>
<td>5.11</td>
<td>The cohort study</td>
<td>96</td>
</tr>
<tr>
<td>5.12</td>
<td>The cross-sectional study</td>
<td>99</td>
</tr>
<tr>
<td>5.13</td>
<td>Studies of change over time</td>
<td>101</td>
</tr>
<tr>
<td>5.14</td>
<td>Choosing a study design</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>Exercises</td>
<td>103</td>
</tr>
<tr>
<td>6</td>
<td>Using a computer</td>
<td>107</td>
</tr>
<tr>
<td>6.1</td>
<td>Introduction</td>
<td>107</td>
</tr>
<tr>
<td>6.2</td>
<td>Advantages of using a computer</td>
<td>107</td>
</tr>
<tr>
<td>6.3</td>
<td>Disadvantages of using a computer</td>
<td>108</td>
</tr>
<tr>
<td>6.4</td>
<td>Types of statistical program</td>
<td>110</td>
</tr>
<tr>
<td>6.5</td>
<td>Evaluating a statistical package</td>
<td>111</td>
</tr>
<tr>
<td>6.6</td>
<td>Strategy for computer-aided analysis</td>
<td>112</td>
</tr>
<tr>
<td>6.7</td>
<td>Forms for data collection</td>
<td>114</td>
</tr>
<tr>
<td>6.8</td>
<td>Plotting</td>
<td>119</td>
</tr>
<tr>
<td>6.9</td>
<td>Other uses of computers</td>
<td>120</td>
</tr>
<tr>
<td>6.10</td>
<td>Misuses of the computer</td>
<td>120</td>
</tr>
<tr>
<td>6.11</td>
<td>Concluding remarks</td>
<td>121</td>
</tr>
<tr>
<td>7</td>
<td>Preparing to analyse data</td>
<td>122</td>
</tr>
<tr>
<td>7.1</td>
<td>Introduction</td>
<td>122</td>
</tr>
<tr>
<td>7.2</td>
<td>Data checking</td>
<td>122</td>
</tr>
<tr>
<td>7.3</td>
<td>Outliers</td>
<td>126</td>
</tr>
<tr>
<td>7.4</td>
<td>Missing data</td>
<td>130</td>
</tr>
<tr>
<td>7.5</td>
<td>Data screening</td>
<td>132</td>
</tr>
</tbody>
</table>
Contents

7.6 Why transform data? 143
7.7 Other features of the data 146
7.8 Concluding remarks 149
Exercises 149

8 Principles of statistical analysis 152
8.1 Introduction 152
8.2 Sampling distributions 153
8.3 A demonstration of the distribution of sample means 155
8.4 Estimation 160
8.5 Hypothesis testing 165
8.6 Non-parametric methods 171
8.7 Statistical modelling 173
8.8 Estimation or hypothesis testing? 174
8.9 Strategy for analysing data 175
8.10 Presentation of results 176
8.11 Summary 177
Exercises 177

9 Comparing groups – continuous data 179
9.1 Introduction 179
9.2 Choosing an appropriate method of analysis 179
9.3 The t distribution 181
9.4 One group of observations 183
9.5 Two groups of paired observations 189
9.6 Two independent groups of observations 191
9.7 Analysis of skewed data 199
9.8 Three or more independent groups of observations 205
9.9 One way analysis of variance – mathematics and worked example 218
9.10 Presentation of results 220
9.11 Summary 222
Exercises 223

10 Comparing groups – categorical data 229
10.1 Introduction 229
10.2 One proportion 230
10.3 Proportions in two independent groups 232
10.4 Two paired proportions 235
10.5 Comparing several proportions 241
10.6 The analysis of frequency tables 241
10.7 2 × 2 frequency tables – comparison of two proportions 250
10.8 2 × k tables – comparison of several proportions 259
10.9 Large tables with ordered categories 265
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.10</td>
<td>$k \times k$ tables – analysis of matched variables</td>
<td>266</td>
</tr>
<tr>
<td>10.11</td>
<td>Comparing risks</td>
<td>266</td>
</tr>
<tr>
<td>10.12</td>
<td>Presentation of results</td>
<td>271</td>
</tr>
<tr>
<td>10.13</td>
<td>Summary</td>
<td>271</td>
</tr>
<tr>
<td></td>
<td>Exercises</td>
<td>272</td>
</tr>
<tr>
<td>11</td>
<td><strong>Relation between two continuous variables</strong></td>
<td>277</td>
</tr>
<tr>
<td>11.1</td>
<td>Association, prediction and agreement</td>
<td>277</td>
</tr>
<tr>
<td>11.2</td>
<td>Correlation</td>
<td>278</td>
</tr>
<tr>
<td>11.3</td>
<td>Use and misuse of correlation</td>
<td>282</td>
</tr>
<tr>
<td>11.4</td>
<td>Rank correlation</td>
<td>285</td>
</tr>
<tr>
<td>11.5</td>
<td>Adjusting a correlation for another variable</td>
<td>288</td>
</tr>
<tr>
<td>11.6</td>
<td>Use of the correlation coefficient in assessing non-Normality</td>
<td>291</td>
</tr>
<tr>
<td>11.7</td>
<td>Correlation – mathematics and worked examples</td>
<td>293</td>
</tr>
<tr>
<td>11.8</td>
<td>Interpretation of correlation</td>
<td>297</td>
</tr>
<tr>
<td>11.9</td>
<td>Presentation of correlation</td>
<td>300</td>
</tr>
<tr>
<td>11.10</td>
<td>Regression</td>
<td>300</td>
</tr>
<tr>
<td>11.11</td>
<td>Use of regression</td>
<td>306</td>
</tr>
<tr>
<td>11.12</td>
<td>Extensions</td>
<td>309</td>
</tr>
<tr>
<td>11.13</td>
<td>Regression – mathematics and worked example</td>
<td>311</td>
</tr>
<tr>
<td>11.14</td>
<td>Interpretation of regression</td>
<td>316</td>
</tr>
<tr>
<td>11.15</td>
<td>Relation to other analyses</td>
<td>318</td>
</tr>
<tr>
<td>11.16</td>
<td>Presentation of regression</td>
<td>319</td>
</tr>
<tr>
<td>11.17</td>
<td>Regression or correlation?</td>
<td>320</td>
</tr>
<tr>
<td></td>
<td>Exercises</td>
<td>321</td>
</tr>
<tr>
<td>12</td>
<td><strong>Relation between several variables</strong></td>
<td>325</td>
</tr>
<tr>
<td>12.1</td>
<td>Introduction</td>
<td>325</td>
</tr>
<tr>
<td>12.2</td>
<td>Analysis of variance and multiple regression</td>
<td>325</td>
</tr>
<tr>
<td>12.3</td>
<td>Two way analysis of variance</td>
<td>326</td>
</tr>
<tr>
<td>12.4</td>
<td>Multiple regression</td>
<td>336</td>
</tr>
<tr>
<td>12.5</td>
<td>Logistic regression</td>
<td>351</td>
</tr>
<tr>
<td>12.6</td>
<td>Discriminant analysis</td>
<td>358</td>
</tr>
<tr>
<td>12.7</td>
<td>Other methods</td>
<td>360</td>
</tr>
<tr>
<td></td>
<td>Exercises</td>
<td>361</td>
</tr>
<tr>
<td>13</td>
<td><strong>Analysis of survival times</strong></td>
<td>365</td>
</tr>
<tr>
<td>13.1</td>
<td>Introduction</td>
<td>365</td>
</tr>
<tr>
<td>13.2</td>
<td>Survival probabilities</td>
<td>367</td>
</tr>
<tr>
<td>13.3</td>
<td>Comparing survival curves in two groups</td>
<td>371</td>
</tr>
<tr>
<td>13.4</td>
<td>Mathematical calculations and worked examples</td>
<td>377</td>
</tr>
<tr>
<td>13.5</td>
<td>Incorrect analyses</td>
<td>385</td>
</tr>
<tr>
<td>13.6</td>
<td>Modelling survival – the Cox regression model</td>
<td>387</td>
</tr>
<tr>
<td>13.7</td>
<td>Design of survival studies</td>
<td>393</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>13.8</td>
<td>Presentation of results</td>
<td>393</td>
</tr>
<tr>
<td></td>
<td>Exercises</td>
<td>394</td>
</tr>
<tr>
<td>14</td>
<td>Some common problems in medical research</td>
<td>396</td>
</tr>
<tr>
<td>14.1</td>
<td>Introduction</td>
<td>396</td>
</tr>
<tr>
<td>14.2</td>
<td>Method comparison studies</td>
<td>396</td>
</tr>
<tr>
<td>14.3</td>
<td>Inter-rater agreement</td>
<td>403</td>
</tr>
<tr>
<td>14.4</td>
<td>Diagnostic tests</td>
<td>409</td>
</tr>
<tr>
<td>14.5</td>
<td>Reference intervals</td>
<td>419</td>
</tr>
<tr>
<td>14.6</td>
<td>Serial measurements</td>
<td>426</td>
</tr>
<tr>
<td>14.7</td>
<td>Cyclic variation</td>
<td>433</td>
</tr>
<tr>
<td></td>
<td>Exercises</td>
<td>435</td>
</tr>
<tr>
<td>15</td>
<td>Clinical trials</td>
<td>440</td>
</tr>
<tr>
<td>15.1</td>
<td>Introduction</td>
<td>440</td>
</tr>
<tr>
<td>15.2</td>
<td>Design of clinical trials</td>
<td>441</td>
</tr>
<tr>
<td>15.3</td>
<td>Sample size</td>
<td>455</td>
</tr>
<tr>
<td>15.4</td>
<td>Analysis</td>
<td>461</td>
</tr>
<tr>
<td>15.5</td>
<td>Interpretation of results</td>
<td>471</td>
</tr>
<tr>
<td>15.6</td>
<td>Writing up and assessing clinical trials</td>
<td>473</td>
</tr>
<tr>
<td></td>
<td>Exercises</td>
<td>474</td>
</tr>
<tr>
<td>16</td>
<td>The medical literature</td>
<td>477</td>
</tr>
<tr>
<td>16.1</td>
<td>Introduction</td>
<td>477</td>
</tr>
<tr>
<td>16.2</td>
<td>The growth of statistics in medical research</td>
<td>478</td>
</tr>
<tr>
<td>16.3</td>
<td>Statistics in published papers</td>
<td>481</td>
</tr>
<tr>
<td>16.4</td>
<td>Reading a scientific paper</td>
<td>493</td>
</tr>
<tr>
<td>16.5</td>
<td>Writing a scientific paper</td>
<td>498</td>
</tr>
<tr>
<td></td>
<td>Exercises</td>
<td>499</td>
</tr>
<tr>
<td>Appendix A</td>
<td>Mathematical notation</td>
<td>505</td>
</tr>
<tr>
<td>A1.1</td>
<td>Introduction</td>
<td>505</td>
</tr>
<tr>
<td>A1.2</td>
<td>Basic ideas</td>
<td>505</td>
</tr>
<tr>
<td>A1.3</td>
<td>Mathematical symbols</td>
<td>509</td>
</tr>
<tr>
<td>A1.4</td>
<td>Functions</td>
<td>510</td>
</tr>
<tr>
<td>A1.5</td>
<td>Glossary of notation</td>
<td>510</td>
</tr>
<tr>
<td>Appendix B</td>
<td>Statistical tables</td>
<td>514</td>
</tr>
<tr>
<td>Answers to exercises</td>
<td>546</td>
<td></td>
</tr>
<tr>
<td>References</td>
<td>575</td>
<td></td>
</tr>
<tr>
<td>Index</td>
<td>589</td>
<td></td>
</tr>
</tbody>
</table>