

Contents

<i>List of figures</i>	xi
<i>List of tables</i>	xv
<i>Welcome!</i>	xvi
<i>Acknowledgements</i>	xvii

1 Numbers in business: the basics 1

- 1.1 Introduction 1
- 1.2 How this book is organised 2
- 1.3 Taking the first steps 4
 - 1.3.1 *The key terms you need to know* 4
 - 1.3.2 *The basic numerical skills you need* 5
- 1.4 Technological support 17
- 1.5 Vignette: how numbers can help get businesses going 18
- 1.6 Test yourself questions 19

2 Presenting data 23

- 2.1 Introduction 23
- 2.2 Types of data 24
- 2.3 Displaying qualitative data 26
 - 2.3.1 *Pictographs* 27
 - 2.3.2 *Pie charts* 29
 - 2.3.3 *Bar charts* 30
- 2.4 Displaying quantitative data 32
 - 2.4.1 *Grouped frequency distributions* 33
 - 2.4.2 *Histograms* 35
 - 2.4.3 *Cumulative frequency graphs* 38
 - 2.4.4 *Stem-and-leaf displays* 41
 - 2.4.5 *Presenting two quantitative variables* 46
 - 2.4.6 *Presenting time series data* 48
- 2.5 Vignette: taking data presentation further with infographics 50
- 2.6 Test yourself questions 50

3 Summarising values of a single variable 55

- 3.1 Introduction 55
- 3.2 Measures of location 56

- 3.2.1 *The mode* 56
- 3.2.2 *The median* 57
- 3.2.3 *The arithmetic mean* 58
- 3.2.4 *Choosing which measure of location to use* 59
- 3.2.5 *Finding measures of location from classified data* 63
- 3.3 Measures of spread 69
 - 3.3.1 *The range* 70
 - 3.3.2 *Quartiles and the semi-interquartile range* 71
 - 3.3.3 *The standard deviation* 76
 - 3.3.4 *Finding measures of spread from classified data* 80
- 3.4 Measuring quality and consistency 82
- 3.5 Vignette: minding the gender pay gap 84
- 3.6 Test yourself questions 84

4 Summarising bivariate data

91

- 4.1 Introduction 91
- 4.2 Correlation and regression 92
 - 4.2.1 *Correlation analysis* 92
 - 4.2.2 *The coefficient of determination* 100
 - 4.2.3 *Simple linear regression analysis* 101
- 4.3 Summarising data collected over time 106
 - 4.3.1 *Index numbers* 107
 - 4.3.2 *Basic time series analysis* 111
- 4.4 Vignette: a world of statistics 120
- 4.5 Test yourself questions 121

5 Assessing risk

127

- 5.1 Introduction 127
- 5.2 Measuring probability 128
- 5.3 Different types of probabilities 131
- 5.4 The rules of probability 134
 - 5.4.1 *The addition rule* 134
 - 5.4.2 *The multiplication rule* 137
 - 5.4.3 *Bayes's rule* 140
 - 5.4.4 *Applying the rules of probability* 142
- 5.5 Probability trees 144
- 5.6 Vignette: what drives the cost of car insurance? 146
- 5.7 Test yourself questions 146

6 Putting probability to work

153

- 6.1 Introduction 153
- 6.2 Simple probability distributions 153
- 6.3 The binomial distribution 158

- 6.4 The Poisson distribution 163
- 6.5 Expectation 165
- 6.6 Decision trees 167
- 6.7 Vignette: decisions, decisions, decisions! 170
- 6.8 Test yourself questions 170

7 Modelling populations

175

- 7.1 Introduction 175
- 7.2 The normal distribution 176
- 7.3 The standard normal distribution 181
 - 7.3.1 *Using the standard normal distribution* 190
- 7.4 Sampling distributions 192
 - 7.4.1 *Estimating the standard error* 195
- 7.5 The t distribution 196
- 7.6 Choosing the correct model for a sampling distribution 199
- 7.7 Vignette: do we perform normally? 199
- 7.8 Test yourself questions 201

8 Statistical decision-making

205

- 8.1 Introduction 205
- 8.2 Estimation 206
 - 8.2.1 *Determining sample size* 213
 - 8.2.2 *Estimating without σ* 214
 - 8.2.3 *Estimating with small samples* 215
- 8.3 Estimating population proportions 216
 - 8.3.1 *Determining sample size* 217
- 8.4 Hypothesis testing 218
 - 8.4.1 *Hypothesis testing without σ* 227
 - 8.4.2 *Hypothesis testing with small samples* 228
- 8.5 Testing hypotheses about two population means 229
 - 8.5.1 *Large independent samples* 230
 - 8.5.2 *Small independent samples* 231
 - 8.5.3 *Paired samples* 233
- 8.6 Testing hypotheses about population proportions 234
- 8.7 A hypothesis test for the population median 236
- 8.8 Vignette: sampling to solve brewing problems 238
- 8.9 Test yourself questions 239

9 Statistical decision-making with bivariate data

243

- 9.1 Introduction 243
- 9.2 Contingency tests 244
- 9.3 Testing and estimating with quantitative bivariate data 251
 - 9.3.1 *Testing correlation coefficients* 251

- 9.3.2 *Testing regression models* 253
- 9.3.3 *Constructing interval predictions* 258
- 9.3.4 *When simple linear models won't do the job* 262
- 9.4 *Vignette: going away green* 264
- 9.5 *Test yourself questions* 265

10 The role of data in business analytics

271

- 10.1 *Introduction* 271
- 10.2 *The importance of data* 272
 - 10.2.1 *Types of data sources* 272
 - 10.2.2 *Types of data* 273
 - 10.2.3 *The data lifecycle* 274
 - 10.2.4 *Data quality* 276
- 10.3 *Types of business analytics* 276
 - 10.3.1 *Descriptive analytics* 277
 - 10.3.2 *Predictive analytics* 277
 - 10.3.3 *Prescriptive analytics* 277
- 10.4 *Business analytics process* 278
- 10.5 *Data ethics* 279
 - 10.5.1 *Data ethics frameworks* 280
 - 10.5.2 *General Data Protection Regulation (GDPR)* 280
- 10.6 *Technologies and tools in business analytics* 282
 - 10.6.1 *Statistical software* 282
 - 10.6.2 *Data visualisation tools* 282
 - 10.6.3 *Programming languages* 283
 - 10.6.4 *Artificial intelligence, machine learning and data mining* 284
- 10.7 *Vignette: using business analytics to improve customers' experience* 285
- 10.8 *Test yourself questions* 286

11 Descriptive analytics

287

- 11.1 *Introduction* 287
- 11.2 *Data analysis warm-up* 288
 - 11.2.1 *Characteristics of a good question* 288
 - 11.2.2 *Types of questions to ask of the data* 289
- 11.3 *Data cleaning* 290
- 11.4 *Data summarisation* 292
 - 11.4.1 *Cross-tabulation* 292
 - 11.4.2 *Pivot tables* 293
- 11.5 *Data visualisation techniques* 298
 - 11.5.1 *Categorical data* 298
 - 11.5.2 *Numerical data* 299
 - 11.5.3 *Advanced graphical techniques* 304

- 11.6 Effective data visualisation 307
 - 11.6.1 *Mental models* 307
 - 11.6.2 *Gestalt principles of design* 308
- 11.7 Vignette: real-time interactive dashboard at Hilton 309
- 11.8 Test yourself questions 309

12 Predictive analytics

313

- 12.1 Introduction 313
- 12.2 The concept of machine learning 314
 - 12.2.1 *Supervised learning* 315
 - 12.2.2 *Unsupervised learning* 316
 - 12.2.3 *Reinforcement learning* 316
 - 12.2.4 *Machine learning process* 317
- 12.3 Introduction to Python programming 318
 - 12.3.1 *Programming environment* 319
 - 12.3.2 *Import Python libraries* 319
 - 12.3.3 *Recommended approach in learning Python in this book* 319
- 12.4 Predictive modelling 320
 - 12.4.1 *Regression models* 320
 - 12.4.2 *Decision trees* 330
 - 12.4.3 *Clustering* 339
 - 12.4.4 *Association rule mining* 344
 - 12.4.5 *Sentiment analysis* 350
 - 12.4.6 *Machine learning models with poor performance* 355
- 12.5 Vignette: Amazon recommender system 356
- 12.6 Test yourself questions 357

13 Prescriptive analytics

361

- 13.1 Introduction 361
- 13.2 Optimisation models 362
 - 13.2.1 *Linear programming* 362
 - 13.2.2 *Integer programming* 367
- 13.3 Simulation models 367
 - 13.3.1 *Monte Carlo simulation* 368
 - 13.3.2 *Probability distributions* 368
 - 13.3.3 *Monte Carlo simulation process* 369
- 13.4 Decision theory 373
 - 13.4.1 *Decision-making under risk* 373
 - 13.4.2 *Decision-making under uncertainty* 378
- 13.5 Vignette: maximising profitability – Marriott's revenue optimisation system 380
- 13.6 Test yourself questions 381

14 Managing statistical research	385
14.1 Introduction	385
14.2 Secondary data	387
14.3 Primary data	390
14.3.1 Selecting your sample	390
14.3.2 Choosing the size of your sample	393
14.3.3 Methods of collecting primary data	394
14.4 Presenting your analysis	399
14.5 Vignette: when <i>The Literary Digest</i> had to eat its words	403
 <i>Appendix 1 Statistical tables</i>	 405
<i>Appendix 2 Answers to test yourself questions</i>	414
<i>References</i>	434
<i>Index</i>	436