2818-4571

PRACTICAL STATISTICS FOR MEDICAL RESEARCH

Douglas G. Altman

Head of Medical Statistical Laboratory
Imperial Cancer Research Fund
London

CHAPMAN & HALL/CRC

Boca Raton London New York Washington, D.C.

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

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

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	$2 \times k$ tables – comparison of several proportions	259
10.9	Large tables with ordered categories	265

10.10	$k \times k$ tables – analysis of matched variables	266
10.11	Comparing risks	266
10.12	Presentation of results	271
10.13	Summary	271
	Exercises	272
11	**Relation between two continuous variables**	**277**
11.1	Association, prediction and agreement	277
11.2	Correlation	278
11.3	Use and misuse of correlation	282
11.4	Rank correlation	285
11.5	Adjusting a correlation for another variable	288
11.6	Use of the correlation coefficient in assessing non-Normality	291
11.7	Correlation – mathematics and worked examples	293
11.8	Interpretation of correlation	297
11.9	Presentation of correlation	300
11.10	Regression	300
11.11	Use of regression	306
11.12	Extensions	309
11.13	Regression – mathematics and worked example	311
11.14	Interpretation of regression	316
11.15	Relation to other analyses	318
11.16	Presentation of regression	319
11.17	Regression or correlation?	320
	Exercises	321
12	**Relation between several variables**	**325**
12.1	Introduction	325
12.2	Analysis of variance and multiple regression	325
12.3	Two way analysis of variance	326
12.4	Multiple regression	336
12.5	Logistic regression	351
12.6	Discriminant analysis	358
12.7	Other methods	360
	Exercises	361
13	**Analysis of survival times**	**365**
13.1	Introduction	365
13.2	Survival probabilities	367
13.3	Comparing survival curves in two groups	371
13.4	Mathematical calculations and worked examples	377
13.5	Incorrect analyses	385
13.6	Modelling survival – the Cox regression model	387
13.7	Design of survival studies	393

13.8 Presentation of results 393
Exercises 394

14 Some common problems in medical research 396
14.1 Introduction 396
14.2 Method comparison studies 396
14.3 Inter-rater agreement 403
14.4 Diagnostic tests 409
14.5 Reference intervals 419
14.6 Serial measurements 426
14.7 Cyclic variation 433
Exercises 435

15 Clinical trials 440
15.1 Introduction 440
15.2 Design of clinical trials 441
15.3 Sample size 455
15.4 Analysis 461
15.5 Interpretation of results 471
15.6 Writing up and assessing clinical trials 473
Exercises 474

16 The medical literature 477
16.1 Introduction 477
16.2 The growth of statistics in medical research 478
16.3 Statistics in published papers 481
16.4 Reading a scientific paper 493
16.5 Writing a scientific paper 498
Exercises 499

Appendix A Mathematical notation 505
A1.1 Introduction 505
A1.2 Basic ideas 505
A1.3 Mathematical symbols 509
A1.4 Functions 510
A1.5 Glossary of notation 510

Appendix B Statistical tables 514

Answers to exercises 546
References 575
Index 589