CONTENTS

		page
	Preface	xi
1	**Why do linguists need statistics?**	1
2	**Tables and graphs**	8
2.1	Categorical data	8
2.2	Numerical data	13
2.3	Multi-way tables	19
2.4	Special cases	20
	Summary	22
	Exercises	23
3	**Summary measures**	25
3.1	The median	27
3.2	The arithmetic mean	29
3.3	The mean and the median compared	30
3.4	Means of proportions and percentages	34
3.5	Variability or dispersion	37
3.6	Central intervals	37
3.7	The variance and the standard deviation	40
3.8	Standardising test scores	43
	Summary	45
	Exercises	46
4	**Statistical inference**	48
4.1	The problem	48
4.2	Populations	49
4.3	The theoretical solution	52
4.4	The pragmatic solution	54
	Summary	57
	Exercises	58

Contents

5	**Probability**	**59**
5.1	Probability	59
5.2	Statistical independence and conditional probability	61
5.3	Probability and discrete numerical random variables	66
5.4	Probability and continuous random variables	68
5.5	Random sampling and random number tables	72
	Summary	75
	Exercises	75
6	**Modelling statistical populations**	**77**
6.1	A simple statistical model	77
6.2	The sample mean and the importance of sample size	80
6.3	A model of random variation: the normal distribution	86
6.4	Using tables of the normal distribution	89
	Summary	93
	Exercises	93
7	**Estimating from samples**	**95**
7.1	Point estimators for population parameters	95
7.2	Confidence intervals	96
7.3	Estimating a proportion	99
7.4	Confidence intervals based on small samples	101
7.5	Sample size	103
7.5.1	Central Limit Theorem	103
7.5.2	When the data are not independent	104
7.5.3	Confidence intervals	105
7.5.4	More than one level of sampling	106
7.5.5	Sample size to obtain a required precision	107
7.6	Different confidence levels	110
	Summary	111
	Exercises	112
8	**Testing hypotheses about population values**	**113**
8.1	Using the confidence interval to test a hypothesis	113
8.2	The concept of a test statistic	117
8.3	The classical hypothesis test and an example	120
8.4	How to use statistical tests of hypotheses: is significance significant?	127
8.4.1	The value of the test statistic is significant at the 1% level	129

8.4.2	The value of the test statistic is not significant	130
	Summary	130
	Exercises	131
9	**Testing the fit of models to data**	**132**
9.1	Testing how well a complete model fits the data	132
9.2	Testing how well a type of model fits the data	137
9.3	Testing the model of independence	139
9.4	Problems and pitfalls of the chi-squared test	144
9.4.1	Small expected frequencies	144
9.4.2	The 2 × 2 contingency table	146
9.4.3	Independence of the observations	147
9.4.4	Testing several tables from the same study	149
9.4.5	The use of percentages	150
	Summary	151
	Exercises	152
10	**Measuring the degree of interdependence between two variables**	**154**
10.1	The concept of covariance	154
10.2	The correlation coefficient	160
10.3	Testing hypotheses about the correlation coefficient	162
10.4	A confidence interval for a correlation coefficient	163
10.5	Comparing correlations	165
10.6	Interpreting the sample correlation coefficient	167
10.7	Rank correlations	169
	Summary	174
	Exercises	174
11	**Testing for differences between two populations**	**176**
11.1	Independent samples: testing for differences between means	176
11.2	Independent samples: comparing two variances	182
11.3	Independent samples: comparing two proportions	182
11.4	Paired samples: comparing two means	184
11.5	Relaxing the assumptions of normality and equal variance: nonparametric tests	188
11.6	The power of different tests	191
	Summary	192
	Exercises	193

Contents

12	**Analysis of variance – ANOVA**	
12.1	Comparing several means simultaneously: one-way ANOVA	194
12.2	Two-way ANOVA: randomised blocks	200
12.3	Two-way ANOVA: factorial experiments	202
12.4	ANOVA: main effects only	206
12.5	ANOVA: factorial experiments	211
12.6	Fixed and random effects	212
12.7	Test score reliability and ANOVA	215
12.8	Further comments on ANOVA	219
12.8.1	Transforming the data	220
12.8.2	'Within-subject' ANOVAs	221
	Summary	222
	Exercises	222
13	**Linear regression**	**224**
13.1	The simple linear regression model	226
13.2	Estimating the parameters in a linear regression	229
13.3	The benefits from fitting a linear regression	230
13.4	Testing the significance of a linear regression	233
13.5	Confidence intervals for predicted values	234
13.6	Assumptions made when fitting a linear regression	235
13.7	Extrapolating from linear models	237
13.8	Using more than one independent variable: multiple regression	237
13.9	Deciding on the number of independent variables	242
13.10	The correlation matrix and partial correlation	244
13.11	Linearising relationships by transforming the data	245
13.12	Generalised linear models	247
	Summary	247
	Exercises	248
14	**Searching for groups and clusters**	**249**
14.1	Multivariate analysis	249
14.2	The dissimilarity matrix	252
14.3	Hierarchical cluster analysis	254
14.4	General remarks about hierarchical clustering	259
14.5	Non-hierarchical clustering	261
14.6	Multidimensional scaling	262
14.7	Further comments on multidimensional scaling	265

14.8	Linear discriminant analysis	265
14.9	The linear discriminant function for two groups	268
14.10	Probabilities of misclassification	269
	Summary	271
	Exercises	271
15	**Principal components analysis and factor analysis**	**273**
15.1	Reducing the dimensionality of multivariate data	273
15.2	Principal components analysis	275
15.3	A principal components analysis of language test scores	278
15.4	Deciding on the dimensionality of the data	282
15.5	Interpreting the principal components	284
15.6	Principal components of the correlation matrix	287
15.7	Covariance matrix or correlation matrix?	287
15.8	Factor analysis	290
	Summary	295
	Appendix A Statistical tables	296
	Appendix B Statistical computation	307
	Appendix C Answers to some of the exercises	314
	References	316
	Index	319