

## Contents

<i>List of figures</i>	<i>page</i> xiii
<i>List of tables</i>	xvii
<i>List of panels</i>	xix
<i>Preface</i>	xxi
<b>Part I Elementary statistical analysis</b>	
<b>Chapter 1 Introduction</b>	<b>3</b>
1.1 Aims of the book	3
1.2 The four case studies and the data sets	6
1.3 Types of measurement	7
1.3.1 Cases, variables, and values	8
1.3.2 Cross-section and time-series variables	8
1.3.3 Levels of measurement	9
1.3.4 Populations and samples	10
1.3.5 Dummy variables	11
1.4 Simple notation	12
1.5 The equation of a straight line	13
1.6 Powers and logarithms	14
1.6.1 Powers	14
1.6.2 Logarithms	16
1.6.3 Logs as proportions and ratios	17
1.7 Index numbers	19
1.8 Trends and fluctuations	21
1.8.1 Measuring the trend by a moving average	22
1.8.2 Measuring the fluctuations by the ratio to the trend	25
<b>Chapter 2 Descriptive statistics</b>	<b>33</b>
2.1 Presentation of numerical data	33
2.1.1 Frequency distributions	33

2.1.2	Bar charts and histograms	36
2.1.3	Frequency curves	40
2.2	Measures of central tendency	42
2.2.1	The mean, the median, and the mode	43
2.2.2	Weighted averages and standardized rates	44
2.2.3	Percentiles, deciles, and quartiles	46
2.3	Measures of dispersion	47
2.3.1	The range and the quartile deviation	47
2.3.2	The mean deviation, the variance, and the standard deviation	47
2.3.3	The coefficient of variation	50
2.4	Measures of central tendency and dispersion with grouped data	51
2.4.1	The median and mean with grouped data	51
2.4.2	The variance and standard deviation with grouped data	52
2.5	The shape of the distribution	53
2.5.1	Symmetrical and skewed distributions	53
2.5.2	Measures of skewness	55
2.6	The normal distribution	56
2.6.1	Properties of the normal distribution	56
2.6.2	The standard normal distribution	61
2.7	Exercises for chapter 2	66
<b>Chapter 3</b>	<b>Correlation</b>	<b>71</b>
3.1	The concept of correlation	71
3.1.1	Correlation is not causation	72
3.1.2	Scatter diagrams and correlation	72
3.1.3	Outliers	74
3.2	The correlation coefficient	76
3.2.1	Measuring the strength of the association	76
3.2.2	Derivation of the correlation coefficient, $r$	77
3.2.3	Interpreting a correlation coefficient	82
3.2.4	An illustration of the use of correlation coefficients	83
3.3	Spearman's rank correlation coefficient	86
3.4	Exercises for chapter 3	90
<b>Chapter 4</b>	<b>Simple linear regression</b>	<b>93</b>
4.1	The concept of regression	93
4.1.1	Explanatory and dependent variables	93
4.1.2	The questions addressed by regression	94
4.2	Fitting the regression line	95
4.2.1	How do we define a line?	95
4.2.2	What criterion should be adopted for fitting the line?	96
4.2.3	How do we find this best fitting line?	98

## CONTENTS

vii

4.2.4	The regression of Y on X versus the regression of X on Y	99
4.2.5	Fitting a linear trend to a time series	100
4.3	Measuring the goodness of fit	102
4.3.1	Explained variations and unexplained variations (residuals)	102
4.3.2	The coefficient of determination, $r^2$	105
4.3.3	Some intuitive implications	106
4.4	Introducing the Ballantine	110
4.5	Exercises for chapter 4	112

**Part II Samples and inductive statistics**

<b>Chapter 5</b>	<b>Standard errors and confidence intervals</b>	<b>117</b>
5.1	Introduction	117
5.2	Sampling distributions	120
5.2.1	Sampling distributions and point estimates of the population mean	120
5.3	Four fundamental statistical propositions	122
5.3.1	The Central Limit Theorem	122
5.4	Theoretical probability distributions	125
5.4.1	Empirical and theoretical frequency distributions	125
5.4.2	The Z- and t-distributions	131
5.5	Confidence intervals for a sample mean	136
5.5.1	Interval estimates and margins of error	137
5.5.2	Confidence interval for an estimate of a population mean	137
5.6	Confidence intervals for other sample statistics	140
5.6.1	Sample proportions	141
5.6.2	Correlation coefficients	141
5.6.3	Regression coefficients	143
5.7	Exercises for chapter 5	145
<b>Chapter 6</b>	<b>Hypothesis testing</b>	<b>149</b>
6.1	Testing hypotheses	149
6.1.1	A simple illustration	150
6.1.2	The five stages of hypothesis testing	151
6.2	The null hypothesis	151
6.2.1	Other assumptions of the model	152
6.3	How confident do we want to be about the results?	153
6.3.1	Two types of error and the null hypothesis	153
6.3.2	Significance levels and the critical region	155
6.3.3	One- and two-tailed tests	158
6.3.4	Statistical versus historical significance	160

6.3.5	Conventional significance levels and prob-values	160
6.3.6	The critical region and the four cases of theft	162
6.4	Test statistics and their sampling distributions	163
6.4.1	The basic ideas	163
6.4.2	Z- and <i>t</i> -tests	164
6.5	Making a decision	167
6.6	Interpreting the results	170
6.7	Three illustrations of the use of <i>t</i> -tests	170
6.7.1	A <i>t</i> -test for a difference of means	170
6.7.2	A <i>t</i> -test for a correlation coefficient when the null hypothesis is zero correlation	173
6.7.3	A <i>t</i> -test for a regression coefficient	175
6.8	Exercises for chapter 6	181
<b>Chapter 7</b>	<b>Non-parametric tests</b>	<b>185</b>
7.1	Introduction	185
7.2	Non-parametric tests for two independent samples	187
7.2.1	The Wald–Wolfowitz runs test	188
7.2.2	The Mann–Whitney <i>U</i> -test	193
7.2.3	The Kolmogorov–Smirnov test	198
7.2.4	Criteria for choosing between alternative tests	201
7.3	The $\chi^2$ -test	202
7.3.1	Contingency tables or cross-tabulations	203
7.3.2	The $\chi^2$ -test for a relationship between two variables	204
7.4	The one-sample runs test of randomness	210
7.5	One-sample non-parametric tests of goodness of fit	216
7.5.1	The one-sample $\chi^2$ -test	216
7.5.2	The one-sample Kolmogorov–Smirnov test	216
7.6	Non-parametric measures of strength of association	218
7.6.1	Four measures of association: $\phi$ , Cramer's V, the contingency coefficient C, and Goodman and Kruskal's tau	218
7.7	Exercises for chapter 7	224
<b>Part III</b>	<b>Multiple linear regression</b>	
<b>Chapter 8</b>	<b>Multiple relationships</b>	<b>231</b>
8.1	The inclusion of additional explanatory variables	231
8.2	Multiple regression	232
8.2.1	The coefficient of multiple determination, $R^2$	232
8.2.2	Partial regression coefficients	234
8.2.3	Controlling for the influence of a second explanatory variable	235

## CONTENTS

ix

8.2.4	Understanding the impact of additional explanatory variables	238
8.2.5	The Ballantine for multiple regression	240
8.2.6	The least squares criterion in multiple regression	242
8.2.7	Standardized beta coefficients	243
8.2.8	The adjusted coefficient of multiple determination, $\bar{R}^2$	245
8.2.9	Interpreting the intercept term	246
8.3	Partial and multiple correlation	248
8.3.1	The interpretation and order of partial correlation coefficients	248
8.3.2	Removing the influence of the control variable	252
8.3.3	Calculation of partial correlation coefficients	253
8.3.4	The coefficient of multiple correlation, $R$	254
8.4	Exercises for chapter 8	255
<b>Chapter 9</b>	<b>The classical linear regression model</b>	<b>258</b>
9.1	Historical research and models of relationships between variables	258
9.1.1	Quantitative methods and historical judgement	260
9.2	Deviations of the dependent variable from the estimated regression line	261
9.2.1	Errors of measurement and specification	261
9.2.2	Stochastic (random) factors and the error term, $e$	262
9.2.3	Sampling and the regression line	263
9.2.4	Assumptions about the error term	264
9.3	A significance test for the regression as a whole	268
9.3.1	The $F$ -test and the $F$ -distribution	269
9.3.2	$F$ -tests of the statistical significance of the regression as a whole	269
9.3.3	The relationship of $F$ and $R^2$	271
9.4	The standard error of the estimate, SEE	272
9.4.1	Confidence interval for a prediction of the mean value of a dependent variable	273
9.4.2	Confidence interval for a prediction of a specific value of a dependent variable	276
9.5	Exercises for chapter 9	277
<b>Chapter 10</b>	<b>Dummy variables and lagged values</b>	<b>280</b>
10.1	Dummy variables	280
10.1.1	Dummy variables with two or more categories	280
10.1.2	Dummy variables as a change in the intercept	283
10.1.3	Additional dummy variables	285

	10.1.4 Interaction between two dummy variables	286
	10.1.5 Interaction between dummy and numerical variables	287
10.2	Lagged variables	289
	10.2.1 Lagged dependent variables	291
	10.2.2 Distributed lag models	292
10.3	Exercises for chapter 10	295
<b>Chapter 11</b>	<b>Violating the assumptions of the classical linear regression model</b>	<b>300</b>
11.1	The assumptions of the classical linear regression model	300
	11.1.1 The model is correctly specified	301
	11.1.2 The error term is appropriately specified	301
	11.1.3 The variables are correctly measured and appropriately specified	301
11.2	Problems of model specification	302
	11.2.1 Non-linear models	302
	11.2.2 Omitted and redundant variables	303
	11.2.3 Unstable parameter values	305
11.3	Problems of error specification	308
	11.3.1 Non-zero errors	308
	11.3.2 Heteroscedasticity	309
	11.3.3 Autocorrelation	311
	11.3.4 Outliers	316
11.4	Problems of variable specification	319
	11.4.1 Errors in variables	319
	11.4.2 Multicollinearity	321
	11.4.3 Simultaneity	324
11.5	Exercises for chapter 11	326
<b>Part IV</b>	<b>Further topics in regression analysis</b>	
<b>Chapter 12</b>	<b>Non-linear models and functional forms</b>	<b>333</b>
12.1	Different forms of non-linear relationships	334
12.2	Estimation of non-linear regression models	341
	12.2.1 Two groups of non-linear models	342
	12.2.2 Estimating the double logarithmic and semi-logarithmic models	345
12.3	Interpretation of non-linear relationships	347
	12.3.1 Interpretation of the regression coefficients for logarithmic models	347
	12.3.2 Elasticities	349
	12.3.3 Long-run multipliers	351

## CONTENTS

xi

12.4	Fitting a trend to a time series	351
12.4.1	Fitting a log-linear trend	351
12.4.2	Fitting non-linear trends	352
12.5	Modelling a relationship	359
12.5.1	Choice of functional form	360
12.5.2	Some general principles	362
12.5.3	Implementing the principles: some examples from economic history	363
12.5.4	Implementing the principles: some examples from demographic, social, and political history	367
12.6	Model specification	372
12.7	Exercises for chapter 12	380
<b>Chapter 13</b>	<b>Logit, probit, and tobit models</b>	<b>384</b>
13.1	Limited dependent variables	385
13.1.1	Three cases of limited dependent variables	385
13.1.2	Why linear models are not appropriate for limited dependent variables	387
13.2	Estimating the logit model: coefficients, absolute effects, and elasticities	390
13.2.1	The logit regression model	390
13.2.2	Interpreting logit coefficients	391
13.2.3	Absolute changes: marginal effects and impact effects	399
13.2.4	Calculating the marginal and impact effects	402
13.2.5	Proportionate effects: logit elasticities	406
13.2.6	Non-linearity and the measurement of change	412
13.2.7	Goodness of fit	413
13.3	Estimating the logit model: an illustration	415
13.3.1	An illustration of the logit model	415
13.3.2	Estimating the logit model using grouped data	418
13.4	Estimating the probit model	419
13.4.1	The probit model	419
13.4.2	Interpreting the results from a probit model	420
13.5	Multi-response models	422
13.6	Censored dependent variables	423
13.6.1	A censored continuous dependent variable	423
13.6.2	The tobit model	424
13.6.3	Two-part censored models	427
13.7	Exercises for chapter 13	432

## Part V Specifying and interpreting models: four case studies

<b>Chapter 14</b>	<b>Case studies 1 and 2: unemployment in Britain and emigration from Ireland</b>	<b>437</b>
14.1	Inter-war benefits and unemployment in Great Britain	438
14.1.1	The Benjamin and Kochin model	438
14.1.2	Parameter stability and outliers in the Benjamin and Kochin model	440
14.2	Emigration from Ireland	445
14.2.1	The time-series model of Irish emigration	446
14.2.2	The panel model of Irish emigration	451
14.2.3	The impact of the explanatory variables on changes in emigration over time	458
<b>Chapter 15</b>	<b>Case studies 3 and 4: the Old Poor Law in England and leaving home in the United States, 1850–1860</b>	<b>463</b>
15.1	The effects of the Old Poor Law	463
15.1.1	Relief payments, wages, and unemployment	464
15.1.2	The Old Poor Law and the birth rate	476
15.2	Leaving home in the United States, 1850–1860	480
15.2.1	Influences on leaving home	482
15.2.2	The logit model	483
Appendix to §15.1.1	Derivation of reduced-form equations from a simultaneous-equations model	491
<b>Appendix A</b>	<b>The four data sets</b>	<b>496</b>
A.1	The iniquitous effects of the Old Poor Law	496
A.2	What explains the scale and timing of the massive emigration from Ireland after the famine?	497
A.3	To what extent were generous unemployment benefits responsible for the high level of unemployment in inter-war Britain?	501
A.4	What influenced children's decision to leave home in the United States in the mid-nineteenth century?	502
<b>Appendix B</b>	<b>Index numbers</b>	<b>507</b>
B.1	Price and quantity relatives	507
B.2	Two types of index number	508
B.3	Laspeyres index numbers with fixed weights	510
B.4	The index number problem and chained indices	513
B.5	Paasche index numbers with changing (given year) weights	520
	<i>Bibliography</i>	529
	<i>Index of subjects</i>	539
	<i>Index of names</i>	545