
Contents

Sections with an asterisk contain proofs and other more mathematical material and may be omitted by readers interested only in results.

Preface

	x
1 A review of multiple regression by least squares	1
1.1 Introduction	1
1.2 Two examples	2
1.3 Basic theory	7
1.4 Sums of squares	8
1.5 The extra sum of squares principle	10
1.6 Scaled and centred variables	11
1.7 Regression through the origin	12
1.8 Problems and pitfalls	14
<i>Discussion</i>	15
<i>Exercises 1</i>	15
<i>References</i>	16
2 Data exploration	17
2.1 Introduction	17
2.2 Objectives	17
2.3 Detecting errors	18
2.4 Exploring the features of the data	20
2.5 Basic statistics	22
2.6 Graphical methods	25
2.7 Further exploratory data analysis techniques	39
<i>Exercise 2</i>	40
<i>References</i>	41
<i>Further reading</i>	42
3 Obtaining least squares solutions	43
3.1 Introduction	43

3.2	Gaussian elimination	44
	<i>Exercises 3.2</i>	46
3.3	The LDU factorization of a matrix	47
	<i>Exercises 3.3</i>	48
3.4	The Gauss–Jordan method for the inverse of a matrix	50
	<i>Exercises 3.4</i>	52
3.5	Application to regression	52
	<i>Exercise 3.5</i>	58
3.6	Orthogonality. The QR factorization of a matrix	58
3.7	Orthogonal matrices. The uniqueness of the QR factorization	63
3.8	Updating regression	65
3.9	The eigenvalue decomposition of a symmetric matrix	69
3.10	The singular value decomposition (SVD) of a matrix	78
	<i>Further reading</i>	81
	<i>References</i>	81
4	Multicollinearity	82
4.1	Introduction	82
4.2	Effects of multicollinearities	83
4.3	Multicollinearity measures	86
4.4	Solutions to multicollinearity	93
4.5	Principal component regression	94
4.6	Latent root regression	99
4.7	Ridge regression	102
4.8	A suggested procedure	105
	<i>Exercise 4</i>	106
	<i>References</i>	106
5	Generalized inverse regression	108
5.1	Introduction	108
5.2	Generalized inverses	110
	<i>Exercises 5.2</i>	113
5.3	The Moore–Penrose generalized inverse	114
	<i>Exercise 5.3</i>	117
5.4	Some theory involving generalized inverses	117
5.5	Least squares with a generalized inverse	119
5.6	Algorithms for constructing a generalized inverse	126

5.7	Testing reduced models	129
5.8*	Analysis of variance	132
	<i>References</i>	137
6	Outliers	138
6.1	The problem of outliers	138
6.2	The standardized residuals and an outlier test	139
6.3	The hat matrix	141
6.4	Influence measure	144
6.5	Illustration of the hat matrix, standardized residual and the influence measure	145
6.6	Multiple unusual points and 'How to detect them'	148
6.7	Examples	149
6.8*	Some identities related to submodel analysis	154
6.9*	Some identities useful in the analysis of residuals to detect outliers and influential observations in the regression	158
6.10*	Likelihood ratio test of no outliers in the regression model	159
6.11*	The Bonferroni inequality and the distribution of $d_i = t_i^2/(n - k)$	160
	<i>Exercise 6</i>	163
	<i>References</i>	163
	<i>Further reading</i>	164
7	Testing for transformations	165
7.1	Introduction	165
7.2	The Box-Cox transformation	166
7.3	Testing for transformations of the explanatory variables	170
7.4*	Ramsey's tests	171
7.5*	Implementing the test	174
7.6	Summary of Ramsey's test	176
	<i>References</i>	178
8	Testing for normality	180
8.1	Introduction	180
8.2	Some tests for normality	181
8.3	The Shapiro-Wilk test	184
8.4	D'Agostino's test	187

8.5	Empirical distribution function tests	192
8.6	Tests based on skewness and kurtosis	194
8.7	Power studies	195
8.8	Effect of explanatory variables	196
	<i>References</i>	197
	<i>Further reading</i>	198
9	Heteroscedasticity and serial correlation	199
9.1	Introduction	199
9.2	Heteroscedasticity	200
9.3	Anscombe's test and its variants	201
9.4*	Constructive tests	204
9.5*	Non-constructive tests	206
9.6	Serial correlation: The Durbin-Watson test	209
9.7	Weighted least squares (WLS)	210
9.8	Generalized least squares (GLS)	212
	<i>References</i>	213
	<i>Further reading</i>	214
10	Predictions from regression	215
10.1	Underfitting	215
10.2	The effect of overfitting on prediction	217
10.3	Variances of predicted means and observations	220
10.4	Confidence intervals	223
10.5	Tolerance intervals	225
10.6	The C_p -statistic	226
10.7*	Adding a variable to a regression (result for Section 10.2)	227
	<i>Exercises 10</i>	228
	<i>References</i>	229
11	Choosing a regression model	230
11.1	Preliminaries	230
11.2	Computational techniques: stepwise procedures	234
11.3	Computational techniques: other procedures	237
11.4	Criteria for selecting models	240
11.5	Freedman's simulation	245
11.6	Discussion	245
	<i>Exercises 11</i>	246
	<i>References</i>	246

12	The analysis of response surface data	249
12.1	Introduction	249
12.2	Example	250
12.3	Model selection	254
12.4	Interpretation of the fitted surface	258
12.5	Further reading	264
	<i>References</i>	265
13	Alternative error structures	266
13.1	Introduction	266
13.2	The effect of errors in the explanatory variables	268
	<i>Exercises 13.2</i>	273
13.3	Working rules for bias due to errors in explanatory variables	273
13.4	Functional relationship model	276
	<i>Exercises 13.4</i>	281
13.5	Structural relationship model	283
	<i>Exercises 13.5</i>	285
13.6	Models with autocorrelated errors	286
13.7	Discussion	290
	<i>References</i>	290
Appendix	Some test data sets	293
A.1	Longley's data	293
A.2	Black cherry tree data	295
A.3	Hocking and Pendleton's problem	296
A.4	Salinity forecasting	297
	Author index	299
	Subject index	303