## Contents

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

-			
$\mathbf{P}$	rei	•	$\alpha$
		a	LC

		X
A rev	iew of multiple regression by least squares	1
1.1	Introduction	
1.2	Two examples	1 2 7
1.3	Basic theory	
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
	Discussion	15
	Exercises 1	15
	References	16
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
	Exercise 2	40
	References	41
	Further reading	42
Obtai	ning least squares solutions	43
3.1	Introduction	43
	1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 Data 2.1 2.2 2.3 2.4 2.5 2.6 2.7	<ul> <li>1.2 Two examples</li> <li>1.3 Basic theory</li> <li>1.4 Sums of squares</li> <li>1.5 The extra sum of squares principle</li> <li>1.6 Scaled and centred variables</li> <li>1.7 Regression through the origin</li> <li>1.8 Problems and pitfalls      Discussion     Exercises 1     References</li> <li>Data exploration</li> <li>2.1 Introduction</li> <li>2.2 Objectives</li> <li>2.3 Detecting errors</li> <li>2.4 Exploring the features of the data</li> <li>2.5 Basic statistics</li> <li>2.6 Graphical methods</li> <li>2.7 Further exploratory data analysis techniques     Exercise 2     References     Further reading</li> <li>Obtaining least squares solutions</li> </ul>

vi	CONTENTS
V I	CONTENT

## Exercises 3.2  3.3 The LDU factorization of a matrix		3.2	Gaussian elimination	44
## Exercises 3.3   48  3.4 The Gauss-Jordan method for the inverse of a matrix   50			Exercises 3.2	46
3.4 The Gauss-Jordan method for the inverse of a matrix  Exercises 3.4  3.5 Application to regression  Exercise 3.5  3.6 Orthogonality. The QR factorization of a matrix  3.7 Orthogonal matrices. The uniqueness of the QR factorization  3.8 Updating regression  3.9 The eigenvalue decomposition of a symmetric matrix  69  3.10 The singular value decomposition (SVD) of a matrix  Further reading  References  4 Multicollinearity  4.1 Introduction  4.2 Effects of multicollinearities  4.3 Multicollinearity measures  4.4 Solutions to multicollinearity  4.5 Principal component regression  4.6 Latent root regression  4.7 Ridge regression  4.8 A suggested procedure  Exercise 4  References  5 Generalized inverse regression  5.1 Introduction  108		3.3	The LDU factorization of a matrix	
matrix         50           Exercises 3.4         52           3.5         Application to regression         52           Exercise 3.5         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           Further reading         81           References         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         94           4.7         Ridge regression         102           4.8         A suggested procedure         105           Exercise 4         106			Exercises 3.3	48
## Exercises 3.4   52  3.5 Application to regression		3.4	The Gauss-Jordan method for the inverse of a	
3.5       Application to regression			matrix	
Exercise 3.5  3.6 Orthogonality. The QR factorization of a matrix 3.7 Orthogonal matrices. The uniqueness of the QR factorization 3.8 Updating regression 3.9 The eigenvalue decomposition of a symmetric matrix 3.10 The singular value decomposition (SVD) of a matrix  Further reading References  4 Multicollinearity 4.1 Introduction 4.2 Effects of multicollinearities 4.3 Multicollinearity measures 4.4 Solutions to multicollinearity 4.5 Principal component regression 4.6 Latent root regression 4.7 Ridge regression 4.8 A suggested procedure Exercise 4 References 5 Generalized inverse regression 5.1 Introduction 108			Exercises 3.4	
Exercise 3.5  3.6 Orthogonality. The QR factorization of a matrix 3.7 Orthogonal matrices. The uniqueness of the QR factorization 3.8 Updating regression 3.9 The eigenvalue decomposition of a symmetric matrix 3.10 The singular value decomposition (SVD) of a matrix  Further reading References  4 Multicollinearity 4.1 Introduction 4.2 Effects of multicollinearities 4.3 Multicollinearity measures 4.4 Solutions to multicollinearity 3.5 Principal component regression 4.6 Latent root regression 4.7 Ridge regression 4.8 A suggested procedure Exercise 4 References 5 Generalized inverse regression 5.1 Introduction 108		3.5	Application to regression	52
3.7 Orthogonal matrices. The uniqueness of the QR factorization  3.8 Updating regression  3.9 The eigenvalue decomposition of a symmetric matrix  3.10 The singular value decomposition (SVD) of a matrix  Further reading  References  4 Multicollinearity  4.1 Introduction  4.2 Effects of multicollinearities  4.3 Multicollinearity measures  4.4 Solutions to multicollinearity  4.5 Principal component regression  4.6 Latent root regression  4.7 Ridge regression  4.8 A suggested procedure  Exercise 4  References  5 Generalized inverse regression  5.1 Introduction  63  63  63  64  65  65  67  68  68  69  69  81  82  40  81  82  41  82  42  43  44  84  85  86  44  86  45  97  46  46  47  87  88  88  89  40  40  40  40  40  40  40  40  40  4				58
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         Further reading       81         References       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         Exercise 4       106         References       106         5       Generalized inverse regression       108         5.1       Introduction       108		3.6	Orthogonality. The QR factorization of a matrix	58
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         Further reading       81         References       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         Exercise 4       106         References       106         5       Generalized inverse regression       108         5.1       Introduction       108		3.7	Orthogonal matrices. The uniqueness of the QR	
3.9 The eigenvalue decomposition of a symmetric matrix  3.10 The singular value decomposition (SVD) of a matrix  Further reading References  4 Multicollinearity 4.1 Introduction 4.2 Effects of multicollinearities 4.3 Multicollinearity measures 4.4 Solutions to multicollinearity 4.5 Principal component regression 4.6 Latent root regression 4.7 Ridge regression 4.8 A suggested procedure Exercise 4 References  5 Generalized inverse regression 5.1 Introduction  108				63
3.9 The eigenvalue decomposition of a symmetric matrix  3.10 The singular value decomposition (SVD) of a matrix  Further reading References  4.1 Introduction 4.2 Effects of multicollinearities 4.3 Multicollinearity measures 4.4 Solutions to multicollinearity 4.5 Principal component regression 4.6 Latent root regression 4.7 Ridge regression 4.8 A suggested procedure Exercise 4 References  5.1 Introduction  108		3.8	Updating regression	65
3.10 The singular value decomposition (SVD) of a matrix  Further reading  References  4 Multicollinearity  4.1 Introduction  4.2 Effects of multicollinearities  4.3 Multicollinearity measures  4.4 Solutions to multicollinearity  93  4.5 Principal component regression  4.6 Latent root regression  4.7 Ridge regression  4.8 A suggested procedure  Exercise 4  References  5 Generalized inverse regression  5.1 Introduction  78  78  78  78  78  78  78  78  78  7		3.9		
matrix         78           Further reading         81           References         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           Exercise 4         106           References         106           5 Generalized inverse regression         108           5.1 Introduction         108			•	69
### Further reading References ## 81  ### A 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  ### Exercise ## 106  ### References ## 106  ### 5 Generalized inverse regression ## 108  ### 5.1 Introduction ### 108		3.10	The singular value decomposition (SVD) of a	
References       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         Exercise 4       106         References       106         5 Generalized inverse regression       108         5.1 Introduction       108			matrix	
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         Exercise 4       106         References       106         5       Generalized inverse regression       108         5.1       Introduction       108			Further reading	81
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         Exercise 4       106         References       106         5       Generalized inverse regression       108         5.1       Introduction       108			References	81
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         Exercise 4       106         References       106         5       Generalized inverse regression       108         5.1       Introduction       108	4	Multio	collinearity	82
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         Exercise 4       106         References       106         5       Generalized inverse regression       108         5.1       Introduction       108				82
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		4.2	Effects of multicollinearities	83
4.5 Principal component regression 94 4.6 Latent root regression 99 4.7 Ridge regression 102 4.8 A suggested procedure 105		4.3	Multicollinearity measures	86
4.6 Latent root regression 99 4.7 Ridge regression 102 4.8 A suggested procedure 105		4.4	Solutions to multicollinearity	93
4.6       Latent root regression       99         4.7       Ridge regression       102         4.8       A suggested procedure       105         Exercise 4       106         References       106         5       Generalized inverse regression       108         5.1       Introduction       108		4.5		94
4.7       Ridge regression       102         4.8       A suggested procedure       105         Exercise 4       106         References       106         5       Generalized inverse regression       108         5.1       Introduction       108		4.6		99
4.8 A suggested procedure       105         Exercise 4       106         References       106         5 Generalized inverse regression       108         5.1 Introduction       108		4.7	Ridge regression	102
Exercise 4       106         References       106         5 Generalized inverse regression       108         5.1 Introduction       108		4.8		105
5 Generalized inverse regression 5.1 Introduction 108			20 1	106
5.1 Introduction 108			References	106
5.1 Introduction 108	5	Gener	alized inverse regression	108
			<u> </u>	108
		5.2	Generalized inverses	110
Exercises 5.2 113				113
5.3 The Moore–Penrose generalized inverse 114		53		114
Exercise 5.3 117		0.0	<del>_</del>	117
5.4 Some theory involving generalized inverses 117		5.4		117
5.5 Least squares with a generalized inverse 119			• • •	119
5.6 Algorithms for constructing a generalized inverse 126				126

CONTENT	S	Vii
5.7	Testing reduced models	129

	5.7 5.8*	Testing reduced models Analysis of variance References	129 132 137
6	Outlie	·	138
U	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	1-1-7
	0.5	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 related to submoder analysis  Some identities useful in the analysis of residuals	137
	0.9	to detect outliers and influential observations in	
		the regression	158
	6 10*	Likelihood ratio test of no outliers in the	150
	0.10	regression model	159
	611*	The Bonferroni inequality and the distribution of	100
	0.11	$d_i = t_i^2/(n-k)$	160
		$u_i - t_i/(n-k)$ Exercise 6	163
		References	163
		Further reading	164
		1 wither remaining	
7	Testing	g 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
		References	178
8	Testing	g 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

riii	CONTENTS

	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
		References	197
		Further reading	198
9	Hetero	oscedasticity 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
		References	213
		Further reading	214
10	Predic	ctions 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
		Exercises 10	228
		References	229
11		ing 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
		Exercises 11	246
		References	246

NTENTS	ix
NTENTS	iz

12	The a	nalysis 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
		References	265
13	Altern	native error structures	266
	13.1	Introduction	266
	13.2	The effect of errors in the explanatory variables	268
		Exercises 13.2	273
	13.3	Working rules for bias due to errors in	
		explanatory variables	273
	13.4	Functional relationship model	276
		Exercises 13.4	281
	13.5	Structural relationship model	283
		Exercises 13.5	285
	13.6	Models with autocorrelated errors	286
	13.7	Discussion	290
		References	290
App	endix	Some test data sets	293
	<b>A</b> .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
Aut	hor ind	ex	299
Sub	Subject index		303