

# Contents

<i>Preface</i>	v
<i>List of Figures</i>	xv
<i>List of Tables</i>	xvii
1. Introduction	1
1.1 Regression Model . . . . .	1
1.2 Goals of Regression Analysis . . . . .	4
1.3 Statistical Computing in Regression Analysis . . . . .	5
2. Simple Linear Regression	9
2.1 Introduction . . . . .	9
2.2 Least Squares Estimation . . . . .	10
2.3 Statistical Properties of the Least Squares Estimation . . . . .	13
2.4 Maximum Likelihood Estimation . . . . .	18
2.5 Confidence Interval on Regression Mean and Regression Prediction . . . . .	19
2.6 Statistical Inference on Regression Parameters . . . . .	21
2.7 Residual Analysis and Model Diagnosis . . . . .	25
2.8 Example . . . . .	28
3. Multiple Linear Regression	41
3.1 Vector Space and Projection . . . . .	41
3.1.1 Vector Space . . . . .	41
3.1.2 Linearly Independent Vectors . . . . .	44
3.1.3 Dot Product and Projection . . . . .	44

3.2	Matrix Form of Multiple Linear Regression . . . . .	48
3.3	Quadratic Form of Random Variables . . . . .	49
3.4	Idempotent Matrices . . . . .	50
3.5	Multivariate Normal Distribution . . . . .	54
3.6	Quadratic Form of the Multivariate Normal Variables . .	56
3.7	Least Squares Estimates of the Multiple Regression Parameters . . . . .	58
3.8	Matrix Form of the Simple Linear Regression . . . . .	62
3.9	Test for Full Model and Reduced Model . . . . .	64
3.10	Test for General Linear Hypothesis . . . . .	66
3.11	The Least Squares Estimates of Multiple Regression Parameters Under Linear Restrictions . . . . .	67
3.12	Confidence Intervals of Mean and Prediction in Multiple Regression . . . . .	69
3.13	Simultaneous Test for Regression Parameters . . . . .	70
3.14	Bonferroni Confidence Region for Regression Parameters .	71
3.15	Interaction and Confounding . . . . .	72
	3.15.1 Interaction . . . . .	73
	3.15.2 Confounding . . . . .	75
3.16	Regression with Dummy Variables . . . . .	77
3.17	Collinearity in Multiple Linear Regression . . . . .	81
	3.17.1 Collinearity . . . . .	81
	3.17.2 Variance Inflation . . . . .	85
3.18	Linear Model in Centered Form . . . . .	87
3.19	Numerical Computation of LSE via QR Decomposition .	92
	3.19.1 Orthogonalization . . . . .	92
	3.19.2 QR Decomposition and LSE . . . . .	94
3.20	Analysis of Regression Residual . . . . .	96
	3.20.1 Purpose of the Residual Analysis . . . . .	96
	3.20.2 Residual Plot . . . . .	97
	3.20.3 Studentized Residuals . . . . .	103
	3.20.4 PRESS Residual . . . . .	103
	3.20.5 Identify Outlier Using PRESS Residual . . . . .	106
	3.20.6 Test for Mean Shift Outlier . . . . .	108
3.21	Check for Normality of the Error Term in Multiple Regression . . . . .	115
3.22	Example . . . . .	115

4.	Detection of Outliers and Influential Observations in Multiple Linear Regression	129
4.1	Model Diagnosis for Multiple Linear Regression . . . . .	130
4.1.1	Simple Criteria for Model Comparison . . . . .	130
4.1.2	Bias in Error Estimate from Under-specified Model	131
4.1.3	Cross Validation . . . . .	132
4.2	Detection of Outliers in Multiple Linear Regression . . . . .	133
4.3	Detection of Influential Observations in Multiple Linear Regression . . . . .	134
4.3.1	Influential Observation . . . . .	134
4.3.2	Notes on Outlier and Influential Observation . . . . .	136
4.3.3	Residual Mean Square Error for Over-fitted Regression Model . . . . .	137
4.4	Test for Mean-shift Outliers . . . . .	139
4.5	Graphical Display of Regression Diagnosis . . . . .	142
4.5.1	Partial Residual Plot . . . . .	142
4.5.2	Component-plus-residual Plot . . . . .	146
4.5.3	Augmented Partial Residual Plot . . . . .	147
4.6	Test for Inferential Observations . . . . .	147
4.7	Example . . . . .	150
5.	Model Selection	157
5.1	Effect of Underfitting and Overfitting . . . . .	157
5.2	All Possible Regressions . . . . .	165
5.2.1	Some Naive Criteria . . . . .	165
5.2.2	PRESS and GCV . . . . .	166
5.2.3	Mallow's $C_P$ . . . . .	167
5.2.4	AIC, $AIC_C$ , and BIC . . . . .	169
5.3	Stepwise Selection . . . . .	171
5.3.1	Backward Elimination . . . . .	171
5.3.2	Forward Addition . . . . .	172
5.3.3	Stepwise Search . . . . .	172
5.4	Examples . . . . .	173
5.5	Other Related Issues . . . . .	179
5.5.1	Variance Importance or Relevance . . . . .	180
5.5.2	PCA and SIR . . . . .	186

6.	Model Diagnostics	195
6.1	Test Heteroscedasticity . . . . .	197
6.1.1	Heteroscedasticity . . . . .	197
6.1.2	Likelihood Ratio Test, Wald, and Lagrange Multiplier Test . . . . .	198
6.1.3	Tests for Heteroscedasticity . . . . .	201
6.2	Detection of Regression Functional Form . . . . .	204
6.2.1	Box-Cox Power Transformation . . . . .	205
6.2.2	Additive Models . . . . .	207
6.2.3	ACE and AVAS . . . . .	210
6.2.4	Example . . . . .	211
7.	Extensions of Least Squares	219
7.1	Non-Full-Rank Linear Regression Models . . . . .	219
7.1.1	Generalized Inverse . . . . .	221
7.1.2	Statistical Inference on Null-Full-Rank Regression Models . . . . .	223
7.2	Generalized Least Squares . . . . .	229
7.2.1	Estimation of $(\beta, \sigma^2)$ . . . . .	230
7.2.2	Statistical Inference . . . . .	231
7.2.3	Misspecification of the Error Variance Structure . . . . .	232
7.2.4	Typical Error Variance Structures . . . . .	233
7.2.5	Example . . . . .	236
7.3	Ridge Regression and LASSO . . . . .	238
7.3.1	Ridge Shrinkage Estimator . . . . .	239
7.3.2	Connection with PCA . . . . .	243
7.3.3	LASSO and Other Extensions . . . . .	246
7.3.4	Example . . . . .	250
7.4	Parametric Nonlinear Regression . . . . .	259
7.4.1	Least Squares Estimation in Nonlinear Regression . . . . .	261
7.4.2	Example . . . . .	263
8.	Generalized Linear Models	269
8.1	Introduction: A Motivating Example . . . . .	269
8.2	Components of GLM . . . . .	272
8.2.1	Exponential Family . . . . .	272
8.2.2	Linear Predictor and Link Functions . . . . .	273
8.3	Maximum Likelihood Estimation of GLM . . . . .	274

8.3.1	Likelihood Equations . . . . .	274
8.3.2	Fisher’s Information Matrix . . . . .	275
8.3.3	Optimization of the Likelihood . . . . .	276
8.4	Statistical Inference and Other Issues in GLM . . . . .	278
8.4.1	Wald, Likelihood Ratio, and Score Test . . . . .	278
8.4.2	Other Model Fitting Issues . . . . .	281
8.5	Logistic Regression for Binary Data . . . . .	282
8.5.1	Interpreting the Logistic Model . . . . .	282
8.5.2	Estimation of the Logistic Model . . . . .	284
8.5.3	Example . . . . .	285
8.6	Poisson Regression for Count Data . . . . .	287
8.6.1	The Loglinear Model . . . . .	287
8.6.2	Example . . . . .	288
9.	Bayesian Linear Regression . . . . .	297
9.1	Bayesian Linear Models . . . . .	297
9.1.1	Bayesian Inference in General . . . . .	297
9.1.2	Conjugate Normal-Gamma Priors . . . . .	299
9.1.3	Inference in Bayesian Linear Model . . . . .	302
9.1.4	Bayesian Inference via MCMC . . . . .	303
9.1.5	Prediction . . . . .	306
9.1.6	Example . . . . .	307
9.2	Bayesian Model Averaging . . . . .	309
	<i>Bibliography</i> . . . . .	317
	<i>Index</i> . . . . .	325