
Contents

Preface	ix
1 Introduction	1
1.1 Before You Start	1
1.2 Initial Data Analysis	2
1.3 When to Use Regression Analysis	6
1.4 History	7
2 Estimation	11
2.1 Linear Model	11
2.2 Matrix Representation	12
2.3 Estimating β	12
2.4 Least Squares Estimation	13
2.5 Examples of Calculating $\hat{\beta}$	14
2.6 Gauss–Markov Theorem	15
2.7 Goodness of Fit	16
2.8 Example	18
2.9 Identifiability	21
3 Inference	25
3.1 Hypothesis Tests To Compare Models	25
3.2 Testing Examples	27
3.3 Permutation Tests	32
3.4 Confidence Intervals for β	34
3.5 Confidence Intervals for Predictions	36
3.6 Designed Experiments	39
3.7 Observational Data	43
3.8 Practical Difficulties	47
4 Diagnostics	53
4.1 Checking Error Assumptions	53
4.2 Finding Unusual Observations	64
4.3 Checking the Structure of the Model	71
5 Problems with the Predictors	75
5.1 Errors in the Predictors	75

5.2	Changes of Scale	79
5.3	Collinearity	81
6	Problems with the Error	87
6.1	Generalized Least Squares	87
6.2	Weighted Least Squares	90
6.3	Testing for Lack of Fit	92
6.4	Robust Regression	96
7	Transformation	107
7.1	Transforming the Response	107
7.2	Transforming the Predictors	110
8	Variable Selection	119
8.1	Hierarchical Models	119
8.2	Testing-Based Procedures	120
8.3	Criterion-Based Procedures	123
8.4	Summary	128
9	Shrinkage Methods	131
9.1	Principal Components	131
9.2	Partial Least Squares	138
9.3	Ridge Regression	141
10	Statistical Strategy and Model Uncertainty	145
10.1	Strategy	145
10.2	An Experiment in Model Building	146
10.3	Discussion	147
11	Insurance Redlining — A Complete Example	149
11.1	Ecological Correlation	149
11.2	Initial Data Analysis	151
11.3	Initial Model and Diagnostics	154
11.4	Transformation and Variable Selection	156
11.5	Discussion	159
12	Missing Data	161
13	Analysis of Covariance	165
13.1	A Two-Level Example	166
13.2	Coding Qualitative Predictors	170
13.3	A Multi-Level Factor Example	172
14	One-Way Analysis of Variance	179
14.1	The Model	179

CONTENTS	vii
14.2 An Example	180
14.3 Diagnostics	183
14.4 Pairwise Comparisons	184
15 Factorial Designs	187
15.1 Two-Way ANOVA	187
15.2 Two-Way ANOVA with One Observation per Cell	188
15.3 Two-Way ANOVA with More than One Observation per Cell	191
15.4 Larger Factorial Experiments	195
16 Block Designs	201
16.1 Randomized Block Design	202
16.2 Latin Squares	206
16.3 Balanced Incomplete Block Design	210
A R Installation, Functions and Data	215
B Quick Introduction to R	217
B.1 Reading the Data In	217
B.2 Numerical Summaries	217
B.3 Graphical Summaries	218
B.4 Selecting Subsets of the Data	219
B.5 Learning More about R	220
Bibliography	221
Index	225