
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	72
5 Problems with the Predictors	77
5.1 Errors in the Predictors	77

5.2	Changes of Scale	81
5.3	Collinearity	83
6	Problems with the Error	89
6.1	Generalized Least Squares	89
6.2	Weighted Least Squares	92
6.3	Testing for Lack of Fit	94
6.4	Robust Regression	98
7	Transformation	109
7.1	Transforming the Response	109
7.2	Transforming the Predictors	112
8	Variable Selection	121
8.1	Hierarchical Models	121
8.2	Testing-Based Procedures	122
8.3	Criterion-Based Procedures	125
8.4	Summary	130
9	Shrinkage Methods	133
9.1	Principal Components	133
9.2	Partial Least Squares	140
9.3	Ridge Regression	143
10	Statistical Strategy and Model Uncertainty	147
10.1	Strategy	147
10.2	An Experiment in Model Building	148
10.3	Discussion	149
11	Insurance Redlining — A Complete Example	151
11.1	Ecological Correlation	151
11.2	Initial Data Analysis	153
11.3	Initial Model and Diagnostics	156
11.4	Transformation and Variable Selection	158
11.5	Discussion	161
12	Missing Data	163
13	Analysis of Covariance	167
13.1	A Two-Level Example	168
13.2	Coding Qualitative Predictors	172
13.3	A Multilevel Factor Example	174
14	One-Way Analysis of Variance	181
14.1	The Model	181

CONTENTS	vii
14.2 An Example	182
14.3 Diagnostics	185
14.4 Pairwise Comparisons	186
15 Factorial Designs	189
15.1 Two-Way ANOVA	189
15.2 Two-Way ANOVA with One Observation per Cell	190
15.3 Two-Way ANOVA with More than One Observation per Cell	193
15.4 Larger Factorial Experiments	197
16 Block Designs	203
16.1 Randomized Block Design	204
16.2 Latin Squares	208
16.3 Balanced Incomplete Block Design	212
A R Installation, Functions and Data	217
B Quick Introduction to R	219
B.1 Reading the Data In	219
B.2 Numerical Summaries	219
B.3 Graphical Summaries	220
B.4 Selecting Subsets of the Data	221
B.5 Learning More about R	222
Bibliography	223
Index	227