

# CONTENTS

---

<b>Preface</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 What Is Regression Analysis?	1
1.2 Publicly Available Data Sets	2
1.3 Selected Applications of Regression Analysis	3
1.3.1 Agricultural Sciences	3
1.3.2 Industrial and Labor Relations	3
1.3.3 History	4
1.3.4 Government	6
1.3.5 Environmental Sciences	6
1.4 Steps in Regression Analysis	7
1.4.1 Statement of the Problem	11
1.4.2 Selection of Potentially Relevant Variables	11
1.4.3 Data Collection	11
1.4.4 Model Specification	12
1.4.5 Method of Fitting	14
1.4.6 Model Fitting	14
1.4.7 Model Criticism and Selection	16
1.4.8 Objectives of Regression Analysis	16
1.5 Scope and Organization of the Book	17
Exercises	18
	<b>vii</b>

<b>2</b>	<b>Simple Linear Regression</b>	<b>21</b>
2.1	Introduction	21
2.2	Covariance and Correlation Coefficient	21
2.3	Example: Computer Repair Data	26
2.4	The Simple Linear Regression Model	28
2.5	Parameter Estimation	29
2.6	Tests of Hypotheses	32
2.7	Confidence Intervals	37
2.8	Predictions	37
2.9	Measuring the Quality of Fit	39
2.10	Regression Line Through the Origin	42
2.11	Trivial Regression Models	44
2.12	Bibliographic Notes	45
	Exercises	45
<b>3</b>	<b>Multiple Linear Regression</b>	<b>53</b>
3.1	Introduction	53
3.2	Description of the Data and Model	53
3.3	Example: Supervisor Performance Data	54
3.4	Parameter Estimation	57
3.5	Interpretations of Regression Coefficients	58
3.6	Properties of the Least Squares Estimators	60
3.7	Multiple Correlation Coefficient	61
3.8	Inference for Individual Regression Coefficients	62
3.9	Tests of Hypotheses in a Linear Model	64
3.9.1	Testing All Regression Coefficients Equal to Zero	66
3.9.2	Testing a Subset of Regression Coefficients Equal to Zero	69
3.9.3	Testing the Equality of Regression Coefficients	71
3.9.4	Estimating and Testing of Regression Parameters Under Constraints	73
3.10	Predictions	74
3.11	Summary	75
	Exercises	75
	Appendix: Multiple Regression in Matrix Notation	82
<b>4</b>	<b>Regression Diagnostics: Detection of Model Violations</b>	<b>85</b>
4.1	Introduction	85
4.2	The Standard Regression Assumptions	86
4.3	Various Types of Residuals	88
4.4	Graphical Methods	90
4.5	Graphs Before Fitting a Model	93

4.5.1	One-Dimensional Graphs	93
4.5.2	Two-Dimensional Graphs	93
4.5.3	Rotating Plots	96
4.5.4	Dynamic Graphs	96
4.6	Graphs After Fitting a Model	97
4.7	Checking Linearity and Normality Assumptions	97
4.8	Leverage, Influence, and Outliers	98
4.8.1	Outliers in the Response Variable	100
4.8.2	Outliers in the Predictors	100
4.8.3	Masking and Swamping Problems	100
4.9	Measures of Influence	103
4.9.1	Cook's Distance	103
4.9.2	Welsch and Kuh Measure	104
4.9.3	Hadi's Influence Measure	105
4.10	The Potential-Residual Plot	107
4.11	What to Do with the Outliers?	108
4.12	Role of Variables in a Regression Equation	109
4.12.1	Added-Variable Plot	109
4.12.2	Residual Plus Component Plot	110
4.13	Effects of an Additional Predictor	114
4.14	Robust Regression	115
	Exercises	115
<b>5</b>	<b>Qualitative Variables as Predictors</b>	<b>121</b>
5.1	Introduction	121
5.2	Salary Survey Data	122
5.3	Interaction Variables	125
5.4	Systems of Regression Equations	128
5.4.1	Models with Different Slopes and Different Intercepts	130
5.4.2	Models with Same Slope and Different Intercepts	137
5.4.3	Models with Same Intercept and Different Slopes	138
5.5	Other Applications of Indicator Variables	139
5.6	Seasonality	140
5.7	Stability of Regression Parameters Over Time	141
	Exercises	143
<b>6</b>	<b>Transformation of Variables</b>	<b>151</b>
6.1	Introduction	151
6.2	Transformations to Achieve Linearity	153
6.3	Bacteria Deaths Due to X-Ray Radiation	155
6.3.1	Inadequacy of a Linear Model	156
6.3.2	Logarithmic Transformation for Achieving Linearity	158

6.4	Transformations to Stabilize Variance	159
6.5	Detection of Heteroscedastic Errors	164
6.6	Removal of Heteroscedasticity	166
6.7	Weighted Least Squares	167
6.8	Logarithmic Transformation of Data	168
6.9	Power Transformation	169
6.10	Summary	173
	Exercises	174
<b>7</b>	<b>Weighted Least Squares</b>	<b>179</b>
7.1	Introduction	179
7.2	Heteroscedastic Models	180
	7.2.1 Supervisors Data	180
	7.2.2 College Expense Data	182
7.3	Two-Stage Estimation	183
7.4	Education Expenditure Data	185
7.5	Fitting a Dose-Response Relationship Curve	194
	Exercises	196
<b>8</b>	<b>The Problem of Correlated Errors</b>	<b>197</b>
8.1	Introduction: Autocorrelation	197
8.2	Consumer Expenditure and Money Stock	198
8.3	Durbin-Watson Statistic	200
8.4	Removal of Autocorrelation by Transformation	202
8.5	Iterative Estimation With Autocorrelated Errors	204
8.6	Autocorrelation and Missing Variables	205
8.7	Analysis of Housing Starts	206
8.8	Limitations of Durbin-Watson Statistic	210
8.9	Indicator Variables to Remove Seasonality	211
8.10	Regressing Two Time Series	214
	Exercises	216
<b>9</b>	<b>Analysis of Collinear Data</b>	<b>221</b>
9.1	Introduction	221
9.2	Effects on Inference	222
9.3	Effects on Forecasting	228
9.4	Detection of Multicollinearity	233
9.5	Centering and Scaling	239
	9.5.1 Centering and Scaling in Intercept Models	240
	9.5.2 Scaling in No-Intercept Models	241
9.6	Principal Components Approach	243
9.7	Imposing Constraints	246

9.8	Searching for Linear Functions of the $\beta$ 's	248
9.9	Computations Using Principal Components	252
9.10	Bibliographic Notes	254
	Exercises	254
	Appendix: Principal Components	255
<b>10</b>	<b>Biased Estimation of Regression Coefficients</b>	<b>259</b>
10.1	Introduction	259
10.2	Principal Components Regression	260
10.3	Removing Dependence Among the Predictors	262
10.4	Constraints on the Regression Coefficients	264
10.5	Principal Components Regression: A Caution	265
10.6	Ridge Regression	268
10.7	Estimation by the Ridge Method	269
10.8	Ridge Regression: Some Remarks	272
10.9	Summary	275
	Exercises	275
	Appendix: Ridge Regression	277
<b>11</b>	<b>Variable Selection Procedures</b>	<b>281</b>
11.1	Introduction	281
11.2	Formulation of the Problem	282
11.3	Consequences of Variables Deletion	282
11.4	Uses of Regression Equations	284
	11.4.1 Description and Model Building	284
	11.4.2 Estimation and Prediction	284
	11.4.3 Control	284
11.5	Criteria for Evaluating Equations	285
	11.5.1 Residual Mean Square	285
	11.5.2 Mallows $C_p$	286
	11.5.3 Information Criteria: Akaike and Other Modified Forms	287
11.6	Multicollinearity and Variable Selection	288
11.7	Evaluating All Possible Equations	288
11.8	Variable Selection Procedures	289
	11.8.1 Forward Selection Procedure	289
	11.8.2 Backward Elimination Procedure	290
	11.8.3 Stepwise Method	290
11.9	General Remarks on Variable Selection Methods	291
11.10	A Study of Supervisor Performance	292
11.11	Variable Selection With Collinear Data	296
11.12	The Homicide Data	296

11.13	Variable Selection Using Ridge Regression	299
11.14	Selection of Variables in an Air Pollution Study	300
11.15	A Possible Strategy for Fitting Regression Models	307
11.16	Bibliographic Notes	308
	Exercises	308
	Appendix: Effects of Incorrect Model Specifications	313
<b>12</b>	<b>Logistic Regression</b>	<b>317</b>
12.1	Introduction	317
12.2	Modeling Qualitative Data	318
12.3	The Logit Model	318
12.4	Example: Estimating Probability of Bankruptcies	320
12.5	Logistic Regression Diagnostics	323
12.6	Determination of Variables to Retain	324
12.7	Judging the Fit of a Logistic Regression	327
12.8	The Multinomial Logit Model	329
	12.8.1 Multinomial Logistic Regression	329
	12.8.2 Example: Determining Chemical Diabetes	330
	12.8.3 Ordered Response Category: Ordinal Logistic Regression	334
	12.8.4 Example: Determining Chemical Diabetes Revisited	335
12.9	Classification Problem: Another Approach	336
	Exercises	337
<b>13</b>	<b>Further Topics</b>	<b>341</b>
13.1	Introduction	341
13.2	Generalized Linear Model	341
13.3	Poisson Regression Model	342
13.4	Introduction of New Drugs	343
13.5	Robust Regression	345
13.6	Fitting a Quadratic Model	346
13.7	Distribution of PCB in U.S. Bays	348
	Exercises	352
	<b>Appendix A: Statistical Tables</b>	<b>353</b>
	<b>References</b>	<b>363</b>
	<b>Index</b>	<b>371</b>