

# Contents

---

Series Editor's Introduction	xi
Preface	xvii
<b>1. Prologue: Regression Analysis as Problematic</b>	<b>1</b>
<b>2. A Grounded Introduction to Regression Analysis</b>	<b>5</b>
2.1 Some Examples of Regression Analysis	5
2.1.1 Abortion and Subsequent Crime	5
2.1.2 Mandatory Basic Education for Welfare Recipients	6
2.1.3 Gender and Academic Salaries	7
2.1.4 Climate Change and Water Resources in India	8
2.1.5 Deforestation and Soil Erosion in the Yangtze River Valley	8
2.1.6 Epidemics of Hepatitis C	9
2.1.7 Onward and Upward	9
2.2 What Is Regression Analysis?	10
2.2.1 A Simple Illustration	10
2.2.2 Controlling for a Third Variable	13
2.2.3 Imposing a Smoother	16
2.3 Getting From Data to Stories	18
<b>3. Simple Linear Regression</b>	<b>21</b>
3.1 Introduction	21
3.2 Describing a Conditional Relationship With a Straight Line	22
3.3 Defining the "Best" Line	24
3.4 Some Useful Formulas	27
3.5 Standardized Slopes	28
3.6 Using Transformations for a Nonlinear Fit	30
3.7 What About the Variance Function?	35
3.8 Summary and Conclusions	37
<b>4. Statistical Inference for Simple Linear Regression</b>	<b>39</b>
4.1 The Role of Sampling	39
4.1.1 Random Sampling	39
4.1.2 Strategy I: Treating the Data as Population	42

4.1.3	Strategy II: Treating the Data <i>as If</i> They Were Generated by Random Sampling From a Population	44
4.1.4	Strategy III: Inventing an Imaginary Population	51
4.1.5	Strategy IV: Model-Based Sampling—Inventing a Friendly Natural Process Responsible for the Data	53
4.1.6	A Note on Randomization Inference	56
4.1.7	Summing Up	58
4.2	Simple Linear Regression Under Random Sampling	58
4.2.1	Estimating the Population Regression Line	58
4.2.2	Estimating the Standard Errors	61
4.2.3	Estimation Under Model-Based Sampling	62
4.2.4	Some Things That Can Go Wrong	62
4.2.5	Tests and Confidence Intervals	65
4.3	Statistical Power	69
4.4	Stochastic Predictors	69
4.5	Measurement Error	73
4.6	Can Resampling Techniques Help?	74
4.6.1	Percentile Confidence Intervals	76
4.6.2	Hypothesis Testing	77
4.6.3	Bootstrapping Regression	77
4.6.4	Possible Benefits From Resampling	78
4.7	Summary and Conclusions	79
<b>5.</b>	<b>Causal Inference for the Simple Linear Model</b>	<b>81</b>
5.1	Introduction	81
5.2	Some Definitions: What's a Causal Effect?	82
5.2.1	The Neyman-Rubin Model	84
5.2.2	Thinking About Causal Effects as Response Schedules	88
5.2.3	What's an Intervention?	90
5.3	Studying Cause and Effect With Data	97
5.3.1	Using Nonstatistical Solutions for Making Causal Inferences	97
5.3.2	Using Statistical Solutions for Making Causal Inferences	98
5.3.3	Using the Simple Linear Model for Making Causal Inferences	99
5.4	Summary and Conclusions	101
<b>6.</b>	<b>The Formalities of Multiple Regression</b>	<b>103</b>
6.1	Introduction	103
6.2	Terms and Predictors	103
6.3	Some Notation for Multiple Regression	105

6.4	Estimation	105
6.5	How Multiple Regression “Holds Constant”	107
6.6	Summary and Conclusions	110
<b>7.</b>	<b>Using and Interpreting Multiple Regression</b>	<b>111</b>
7.1	Introduction	111
7.2	Another Formal Perspective on Holding Constant	111
7.3	When Does Holding Constant Make Sense?	113
7.4	Standardized Regression Coefficients: Once More With Feeling	117
7.5	Variances of the Coefficient Estimates	119
7.6	Summary and Conclusions	122
<b>8.</b>	<b>Some Popular Extensions of Multiple Regression</b>	<b>125</b>
8.1	Introduction	125
8.2	Model Selection and Stepwise Regression	126
8.2.1	Model Selection by Removing Terms	127
8.2.2	Tests to Compare Models	128
8.2.3	Selecting Terms Without Testing	130
8.2.4	Stepwise Selection Methods	132
8.2.5	Some Implications	133
8.3	Using Categorical Terms: Analysis of Variance and Analysis of Covariance	135
8.3.1	An Extended Example	135
8.4	Back to the Variance Function: Weighted Least Squares	141
8.4.1	Visualizing Lack of Fit	141
8.4.2	Weighted Least Squares as a Possible Fix	142
8.4.3	Evaluating the Mean Function	145
8.5	Locally Weighted Regression Smoother	147
8.6	Summary and Conclusions	148
<b>9.</b>	<b>Some Regression Diagnostics</b>	<b>151</b>
9.1	Introduction	151
9.2	Transformations of the Response Variable	152
9.2.1	Box-Cox Procedures	152
9.2.2	Inverse Fitted Value Response Plots	153
9.3	Leverage and Influence	159
9.3.1	Influential Cases and Cook’s Distance	159
9.4	Cross-Validation	162
9.5	Misspecification Tests	163
9.5.1	Instrumental Variables	164
9.5.2	Tests for Exogeneity	167
9.6	Conclusions	168

<b>10. Further Extensions of Regression Analysis</b>	<b>171</b>
10.1 Regression Models for Longitudinal Data	171
10.1.1 Multiple Linear Regression for Time Series Data	172
10.2 Regression Analysis With Multiple Time Series Data	177
10.2.1 Fixed Effects Models	178
10.2.2 Random Effects Models	178
10.2.3 Estimation	180
10.3 Multilevel Models	180
10.4 The Generalized Linear Model	183
10.4.1 GLM Structure	183
10.4.2 Normal Models	184
10.4.3 Poisson Models	184
10.4.4 Poisson Models for Contingency Tables	186
10.4.5 Binomial Regression	186
10.5 Multiple Equation Models	188
10.5.1 Causal Inference Once Again	191
10.5.2 A Final Observation	196
10.6 Meta-Analysis	196
10.7 Conclusions	200
<b>11. What to Do</b>	<b>203</b>
11.1 How Did We Get Into This Mess?	203
11.2 Three Cheers for Description	206
11.2.1 What's Description?	207
11.2.2 Advocacy Settings	207
11.2.3 Descriptive Regressions as Part of a Broad Research Program	209
11.2.4 Spotting Provocative Associations	210
11.2.5 Some Other Benefits of Description	212
11.3 Two Cheers for Statistical Inference	218
11.3.1 Working With Near-Random Samples	220
11.3.2 Working With Data From Nature	222
11.3.3 Working With a Nearly Correct Model	222
11.4 One Cheer for Causal Inference	223
11.4.1 Special-Purpose Estimators	226
11.4.2 Propensity Scores	230
11.4.3 Sensitivity Analysis of the Selection Process	231
11.4.4 Bounding Treatment Effects	232
11.4.5 Some Forecasts	234
11.5 Some Final Observations	234
11.5.1 A Police Story	234
11.5.2 Regression Analysis as Too Little, Too Late	237
<b>References</b>	<b>239</b>
<b>Index</b>	<b>251</b>
<b>About the Author</b>	<b>259</b>