

Statistics and Econometrics: Methods and Applications

Orley Ashenfelter *Princeton University*

Phillip B. Levine Wellesley College

David J. Zimmerman Williams College



John Wiley & Sons, Inc.

Brief Contents

1. Introduction 1

2. Basic Probability Theory 9

3. Random Variables and Probability Distributions 27

4. Mathematics of Expectations 44

5. Multivariate Distributions 56

6. Sampling and Sampling Distributions 73

7. Estimation 91

8. Interval Estimation and Hypothesis Testing 104

9. Simple Linear Regression 122

10. Inference in Simple Linear Regression 140

11. Multiple Regression 158

12. Extending the Multiple Regression Model 170

13. Specification Error, Multicollinearity, and Measurement Error 186

14. Heteroskedasticity and Serial Correlation 204

15. Simultaneous Equations 220

16. Dummy Dependent Variable Models 235

17. Analysis of Time Series Data 245

18. Panel Data Models 262

Appendix A 275 Appendix B 284 Index 293

Contents

Chapter 1 Introduction 1

- 1.1 What is Econometrics? 1
 - 1.1.1 Quantifying Economic
 - Relationships 2 1.1.2 Testing Competing Hypotheses
 - 1.1.2 Testing Competing Hypotheses 2 1.1.3 Forecasting 2
 - 5
- 1.2 The Use of Models 3
- 1.3 Types of Data 3
 - 1.3.1 Cross-sectional Data 4
 - 1.3.2 Time Series Data 4
 - 1.3.3 Panel (or Longitudinal) Data 5
 - 1.3.4 Data Sources 6
- 1.4 Conducting an Empirical Project 6
 - 1.4.1 Pick a Topic 6
 - 1.4.2 Learn What Others Have Learned about This Topic 6
 - 1.4.3 Have an Empirical Strategy 7
 - 1.4.4 Interpret Your Results 7
 - 1.4.5 Write a Report 7
- 1.5 Overview of Text 8

Chapter 2 Basic Probability Theory 9

- 2.1 Introduction 9
- 2.2 Basic Definitions 9 2.2.1 Random Trial 9 2.2.2 Basic Outcomes 10
- Application 2.1 Nothing Better to Do? 10
 - 2.2.3 Sample Space 10
 - 2.2.4 Event 11
 - 2.2.5 Venn Diagrams 11
- 2.3 Set Theory 11 2.3.1 Unions, Intersections, and Complementation 11 2.3.2 Some Rules for Set Operations 13
- 2.4 Probability 15

- 2.4.1 Fundamental Postulates 15
- 2.4.2 Results 16
- 2.4.3 Interpretations of Probability 18
- 2.5 Conditional Probability 19
- Application 2.2 Me? Cheat? I Would Never Do (or, At Least, Admit to) That 20
- 2.6 Bayes' Theorem 21
- Application 2.3 You Gotta Relent If it's 95 Percent. Or Do You? 22
- 2.7 Independence 22
- Application 2.4 They Say It's Hot, but It's Probably Not 23
- 2.8 Problems 24

Chapter 3 Random Variables and Probability Distributions 27

- 3.1 Introduction 27
- 3.2 Basic Definitions 27
 - 3.2.1 Random Variables 27
 - 3.2.2 Discrete and Continuous Random Variables 27
- 3.3 Probability Distribution for A Discrete Random Variable 29
 - 3.3.1 Assigning Probabilities 29
 - 3.3.2 Discrete Probability Distribution 29
 - 3.3.3 Properties of Discrete Probability
 - Distributions 30
 - 3.3.4 Cumulative Probability Distributions 30
 - 3.3.5 Properties of Discrete Cumulative Probability Distributions 32
- Application 3.1 Where Have All the Girls Gone? 31
- Application 3.2 Saving a Little Extra for the Stretch Run 32

- xiv Contents
- 3.4 Probability Distribution for a Continuous Random Variable 33
 - 3.4.1 Assigning Probabilities 33
 - 3.4.2 Probability Density Function 33
 - 3.4.3 Properties of Continuous Probability Distributions 34
 - 3.4.4 Cumulative Probability Distributions 34
 - 3.4.5 Properties of Distribution Functions 35
- 3.5 Probability Distributions and Empirical Frequency Distributions 35
- 3.6 Some Popular Discrete Distributions 37
 - 3.6.1 The Bernoulli Distribution 37
 - 3.6.2 The Binomial Distribution 37
 - 3.6.3 Discrete Uniform Distribution 39
- 3.7 Some Popular Continuous Distributions 39 3.7.1 The Normal Distribution 39
- Application 3.3 Bringing Back Bad Memories 40
 - 3.7.2 The Exponential Distribution 41
- Application 3.4 Is Welfare Truly Addictive? 41
 - 3.7.3 The Uniform Distribution 42
- 3.8 PROBLEMS 42

Chapter 4 Mathematics of Expectations 44

- 4.1 Introduction 44
- 4.2 Summarizing a Distribution 44
- 4.3 Measures of Central Tendency 45 4.3.1 Mean or Expected Value 45
- Application 4.1 Where Is the Best Place to Place Your Bets? 46
 - 4.3.2 Properties of Expected Values 46
 - 4.3.3 Median 48
 - 4.3.4 Mode 48
 - 4.3.5 Mean, Median, or Mode: Which One? 48
- 4.4 Measures of Dispersion 49
 - 4.4.1 Range 49
 - 4.4.2 Variance 49
 - 4.4.3 Standard Deviation 50

4.4.4 Properties of the Variance and Standard Deviation 51

Application 4.2 The Risk from Russian Roulette 52

- 4.5 Chebyshev's Inequality 52
- 4.6 Problems 54

Chapter 5 Multivariate Distributions 56

- 5.1 Introduction 56
- 5.2 Discrete Bivariate Distributions 56
 - 5.2.1 Joint Probability Function 56 5.2.2 Properties of Joint Probability Function 58
 - 5.2.3 Cumulative Joint Probability Function 58
 - 5.2.4 Bivariate and Univariate Distributions 58
 - 5.2.5 Conditional Distributions 59
- 5.3 Continuous Bivariate Distributions 60
 - 5.3.1Joint Density Function605.3.2Properties of Joint Density
 - Function 60 5.3.3 Cumulative Joint Density Function 61
 - 5.3.4 Properties of Cumulative Joint Density Function 62
 - 5.3.5 Bivariate and Univariate Distributions 62
- 5.4 Independence 62
- 5.5 (More) Mathematics of Expectations 62
- Application 5.1 One Thumb, Two Thumbs, Drumming on a Drum 63
 - 5.5.1 Discrete Random Variables 64
 - 5.5.2 Continuous Random Variables 64
- 5.6 Expectations of Functions of Several Random Variables 64
- 5.7 Conditional Expectations 66
- 5.8 Measures of Association: Covariance and Correlation 66
 - 5.8.1 Covariance 66
 - 5.8.2 Properties of Covariance 67
 - 5.8.3 Correlation 68
 - 5.8.4 Properties of Correlation 69

Application 5.2 The Long and Variable Road 69

5.9 Problems 71

Chapter 6 Sampling and Sampling Distributions 73

- 6.1 Introduction 73
- 6.2 Populations 73
- 6.3 Samples 74 6.3.1 Simple Random Samples 74 6.3.2 Why Sample? 75
- Application 6.1 How Were They Supposed to Know? 75
- Application 6.2 Without the CPS, It Would Be Anybody's Guess 76
- 6.4 Populations, Probability Distributions, and Data 76
- Application 6.3 How Does the Boss Get Paid When His Songs Get Played? 77
- 6.5 Sample Statistics 78
- Application 6.4 It's Only Money 78
- 6.6 Sampling Distributions 79
- 6.7 The Normal Probability Distribution 80
 - 6.7.1 Properties of the Normal Distribution 80
 - 6.7.2 Standard Normal Random Variables 81
 - 6.7.3 Finding Probabilities for a Standard Normal Random Variable 81
 - 6.7.4 Some Standard Normal Probabilities Worth Remembering 82
- 6.8 The Distribution of the Sample Mean: X 83 6.8.1 The Central Limit Theorem 84
- Application 6.5 Dewey Beats Truman? 000
- 6.9 Probability Statements about $\bar{\mathbf{X}}$ 86
- 6.10 The Chi-Squared and T Distributions 86
 6.10.1 The Chi-Squared Distribution 87
 6.10.2 The T Distribution 88
- 6.11 Problems 89

Chapter 7 Estimation 91

- 7.1 Introduction 91
- 7.2 Point Estimation 91
- Application 7.1 How Do They Count Spotted Owls Anyway? 92
- 7.3 Finite Sample Properties 93
 7.3.1 Unbiasedness 93
 7.3.2 Efficiency 95
 7.3.3 Mean Squared Error 96
- 7.4 Asymptotic Properties 97
- 7.5 Methods of Finding Estimators 97
 7.5.1 Method of Moments 97
 7.5.2 Method of Maximum Likelihood 99
 7.5.3 Method of Least Squares 100
- Application 7.2 Who Knew? 101
- 7.6 Problems 102

Chapter 8 Interval Estimation and Hypothesis Testing 104

- 8.1 Introduction 104
- 8.2 Interval Estimation 104
 - 8.2.1 Interval Estimation of the Population Mean 104
 - 8.2.2 Interval Estimation of Difference Between Means 109
 - 8.2.3 Interval Estimation of a Population Proportion 109
 - 8.2.4 Interval Estimation of the Population Variance 110
- Application 8.1 How Can 1200 People Decide an Election? 111

8.2.5 Overview 111

- 8.3 Hypothesis Tests on the Population Mean 112
 - 8.3.1 Null and Alternative Hypotheses 112
 - 8.3.2 Test Statistic and Critical Region 112
- Application 8.2 What's Wrong with This Picture? 113
 - 8.3.3 Type I and Type II Errors 114
 - 8.3.4 Statistical Decision Rule for the Population Mean (Two-Sided Alternative) 115
 - 8.3.5 P-Values 116

xvi Contents

8.3.6	Statistical Decision Rule for the
	Population Mean (One-Sided
	Alternatives) 116

Application 8.3 If the Price Is High, Women Need Not Apply 117

8.4 Problems 118

Chapter 9 Simple Linear Regression 122

- 9.1 Introduction 122
- 9.2 Relating Two Variables 122
- 9.3 The Linear Regression Model 125
- Application 9.1 Wine Appreciation 126
- 9.4 Estimation (Three Ways) 128 9.4.1 Ordinary Least Squares 129
- Application 9.2 If You've Got the Data, You Can Get the Beta 132
 - 9.4.2 Method of Moments 132
 - 9.4.3 Maximum Likelihood 134
- 9.5 Interpretation of the Regression Slope Coefficient 135
 - 9.5.1 Units of Measurement and Regression Coefficients 135
 - 9.5.2 Models Including Logarithms 136

Application 9.3 It's More Than Just the Parties 137

9.6 Problems 137

Chapter 10 Inferences in Simple Linear Regression 140

- 10.1 Introduction 140
- 10.2 Sampling Distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$ Based on Simulation 140
- 10.3 Sampling Distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$ Based on Theory 143
 - 10.3.1 The Linearity of the OLS Estimators 143
 - 10.3.2 The Unbiasedness of the OLS Estimators 144
 - 10.3.3 The Variance of the OLS Estimators 145

10.3.4 The Normality of the OLS Estimators 146

- 10.4 Confidence Intervals 146
- 10.5 Hypothesis Testing 147
- 10.6 Goodness of Fit (R^2) 148
- Application 10.1 Learning the Lingo 149
- Application 10.2 Is It Hip to Be Square? 151
- 10.7 The Gauss-Markov Theorem 152
- 10.8 Problems 154

Chapter 11 Multiple Regression 158

- 11.1 Introduction 158
- 11.2 The Multiple Regression Model 158
- 11.3 Interpretation of the Regression Coefficients 159
- 11.4 Estimation (Three Ways) 160
 - 11.4.1 Ordinary Least Squares 160
 - 11.4.2 Method of Moments 160
 - 11.4.3 Maximum Likelihood 161
 - 11.4.4 Example: The Simple and Two-Variable Regression Models 162
- 11.5 R^2 Revisited 162
- 11.6 Inference in the Multiple Regression Model 163
 11.6.1 Confidence Intervals 163
 11.6.2 Hypothesis Tests on a Single Coefficient 164
- Application 11.1 It's a Tough Job, but Someone Has to Do It 165

11.6.3 The Rule of 2 165

- Application 11.2 If the Sun Don't Shine, Don't Buy the Wine 166
- 11.7 Problems 167

Chapter 12 Extending the Multiple Regression Model 170

- 12.1 Introduction 170
- 12.2 Dummy Explanatory Variables 170 12.2.1 Defining and Interpreting Dummy Variables 171

- Application 12.1 If the t is Greater then Two, The Verdict May Go against You 173
 - 12.2.2 Interaction Variables 172
 - 12.2.3 Experimental versus Observational Studies 175
- Application 12.2 Maybe It's Something in the Brownies 176
 - 12.2.4 Dummy Variables When There Are More than Two Groups 177
- 12.3 Hypothesis Tests on Several Regression Coefficients: F Tests 178
 12.3.1 Joint Tests on Several Regression Coefficients 179
- Application 12.3 It's Only a Piece of Paper (or Is It?) 182
 - 12.3.2 Testing Whether All of the Regression Slope Coefficients Are Zero: "The F Test" 183
- 12.4 Problems 184

Chapter 13 Specification Error, Multicollinearity, and Measurement Error 186

13.1 Introduction 186

- 13.2 Including an Irrelevant Variable 186
- 13.3 Excluding a Relevant Variable (Omitted Variable Bias) 188

Application 13.1 More Bad News for Smokers? 190

- 13.4 Nonlinearities 191
- Application 13.2 Less Gold in the Golden Years 192
- 13.5 Multicollinearity 192

Application 13.3 Some Like It Cold 193

- 13.6 Measurement Error 195 13.6.1 Probability Limits 196 13.6.2 The Classic Errors-in-Variables Model 196 13.6.3 Instrumental Variables 199
- Application 13.4 Double the Pleasure, Double the (Econometric) Fun 201

Application 13.5 Horatio Alger Meets the Data 202

13.7 Problems 202

Chapter 14 Heteroskedasticity and Serial Correlation 204

- 14.1 Introduction 204
- 14.2 Heteroskedasticity 204
 14.2.1 Definition 204
 14.2.2 Causes 205
 14.2.3 Consequences for OLS Estimates 205
 14.2.4 Detection 207
 14.2.5 Weighted Least Squares 208
 14.2.6 White Correction 210
- Application 14.1 Was Welfare Refom Worth the Weight? 211
- 14.3 Serial Correlation 212
 14.3.1 Definition 212
 14.3.2 Causes 213
 14.3.3 Consequences for OLS Estimates 214
 14.3.4 Correction: Quasi-Differencing 214
 14.3.5 Detection: The Durbin–Watson Test 216
- Application 14.2 Where Is the Inflation? 217

14.4 Problems 218

Chapter 15 Simultaneous Equations 220

- 15.1 Introduction 220
- 15.2 Structural and Reduced-Form Equations 220
- 15.3 Simultaneous Equations Bias 222
- 15.4 Identification 223
- 15.5 The Order Condition 226
- 15.6 Estimation 227 15.6.1 Exactly Identified Equations: Indirect Least Squares 227 15.6.2 Overidentified Equations: Two-Stage
 - Least Squares 227
- Application 15.1 Something Smells Fishy 228
- Application 15.2 Will Spending an Extra Dime Help Reduce the Rate of Crime? 230

xviii Contents

15.6.3 Hausman Specification Test 230

Application 15.3 When Your Number Comes Up 231

15.7 Problems 232

Chapter 16 Dummy Dependent Variable Models 235

16.1 Introduction 235

- 16.2 Linear Probability Model 236
- 16.3 Logit and Probit Models 237
 16.3.1 Logit Model 238
 16.3.2 Probit Model 239
 16.3.3 Which Is Better? Logit or Probit? 239
- Application 16.1 Do Banks Discriminate in Giving Out Mortagage Loans? 240
- 16.4 Censored Regression Model 240 16.4.1 Heckman Two-Step Estimator 241
- Application 16.2 Lotto Fever: Does It Make Economic Sense? 242
- 16.5 Problems 243

Chapter 17 Analysis of Time Series Data 245

17.1 Introduction 245

- 17.2 Describing Time Series Processes 246
 17.2.1 Moving Average Models 246
 17.2.2 Autoregressive Models 247
 17.2.3 ARMA Models 248
 17.2.4 Unit Roots and Random Walks 248
 17.2.5 Stationarity 250
 17.2.6 ARIMA Models 251
- 17.3 Estimating the Relationship Between Two Variables 251
- Application 17.1 Should You Use Darts to Pick Your Stocks? 252
 - 17.3.1 Spurious Regression 252
 - 17.3.2 Dickey–Fuller Tests 254
 - 17.3.3 Regression in First Differences 255
 - 17.3.4 Cointegration 256
- 17.4 Estimation and Forecasting 257
- 17.5 Vector Autoregression (VAR) 259

Application 17.2 Frustration in Forecasting Inflation 260

17.6 Problems 261

Chapter 18 Panel Data Models 262

- 18.1 Introduction 262
- 18.2 Unobservable Heterogeneity and Panel Data 262
 18.2.1 What Is Unobservable Heterogeneity? 262
 18.2.2 The Role of Panel Data 263
- 18.3 Fixed Effects Models 264
 18.3.1 First Differences 264
 18.3.2 Deviations from Means 264
- Application 18.1 Getting Your Monet's Worth 265

18.3.3 Individual Dummy Variables 266

- 18.4 Pooled Cross-Section, Time Series Data 266
 18.4.1 Fixed Effects Methods with Pooled Cross-Section, Time-Series Data 266
 18.4.2 Quasi-Experiments in Theory 267
 18.4.3 Quasi-Experiments in Practice 269
- 18.5 Random Effects Models 269
- Application 18.2 How Much Does Legal Abortion Reduce Births 270
- Application 18.3 If It Wasn't a Franchise Would They Ask If You Wanted Fries? 271
- 18.6 Fixed Effects or Random Effects? 272
- 18.7 Problems 273

Appendix A Review of Probability and Statistics 275

- A.1 Introduction 275
- A.2 Basic Probability Theory 276
 - A.2.1 Properties of Probabilities 276
 - A.2.2 Random Variables 276
 - A.2.3 Continuous Probability Distributions 276
- A.3 Descriptive Statistics 277 A.3.1 Measures of Central Tendency 277
 - A.3.2 Measures of Dispersion 278

- A.4 Calculating Probabilities from a Normal Distribution 278
- A.5 The Distribution of the Sample Mean 279
- A.6 Hypothesis Testing 280
 - A.6.1 Setting Up the Test 280
 - A.6.2 Intuition of the Test 280
 - A.6.3 Procedure for a Sample Mean 281
- A.7 The *t* Distribution versus the Standard Normal Distribution 281
- A.8 Measures of Association 281 A.8.1 Covariance 282 A.8.2 Correlation 282

Appendix B Statistical Tables 284

Index 293