

# **CONTENTS**

## **CHAPTER 1 Introducing Stata 1**

- 1.1 Starting Stata 2
- 1.2 The opening display 2
- 1.3 Exiting Stata 3
- 1.4 Stata data files for Principles of Econometrics 3
  - 1.4.1 A working directory 4
- 1.5 Opening Stata data files 5
  - 1.5.1 The use command 5
  - 1.5.2 Using the toolbar 6
  - 1.5.3 Using files on the internet 6
  - 1.5.4 Locating book files on the internet 7
- 1.6 The variables window 7
  - 1.6.1 Using the data editor for a single label 7
  - 1.6.2 Using the data utility for a single label 8
  - 1.6.3 Using variables manager 9
- 1.7 Describing data and obtaining summary statistics 11
- 1.8 The Stata help system 12
  - 1.8.1 Using keyword search 13
  - 1.8.2 Using command search 14
  - 1.8.3 Opening a dialog box 15
  - 1.8.4 Complete documentation in Stata manuals 16
- 1.9 Stata command syntax 16
  - 1.9.1 Syntax of summarize 16
  - 1.9.2 Learning syntax using the review window 17
- 1.10 Saving your work 20
  - 1.10.1 Copying and pasting 20
  - 1.10.2 Using a log file 21
- 1.11 Using the data browser 25
  - 1.12.1 Histograms 25
  - 1.12.2 Scatter diagrams 28
- 1.13 Using Stata Do-files 29
- 1.14 Creating and managing variables 32
  - 1.14.1 Creating (generating) new variables 32
  - 1.14.2 Using the expression

builder 34

- 1.14.3 Dropping or keeping variables and observations 35
- 1.14.4 Using arithmetic operators 38
- 1.14.5 Using Stata math functions 38
- 1.15 Using Stata density functions 39
  - 1.15.1 Cumulative distribution functions 39
  - 1.15.2 Inverse cumulative distribution functions 40
- 1.16 Using and displaying scalars 41
  - 1.16.1 Example of standard normal cdf 41
  - 1.16.2 Example of t-distribution tail-cdf 42
  - 1.16.3 Example computing percentile of the standard normal 42
  - 1.16.4 Example computing percentile of the t-distribution 42
- 1.17 A scalar dialog box 43
- 1.18 Using factor variables 45
  - 1.18.1 Creating indicator variables using a logical operator 47
  - 1.18.2 Creating indicator variables using tabulate 48

Key Terms 49

Chapter 1 Do-file 50

## **CHAPTER 2 Simple Linear Regression 53**

- 2.1 The food expenditure data 53
  - 2.1.1 Starting a new problem 54
  - 2.1.2 Starting a log file 54
  - 2.1.3 Opening a Stata data file 54
  - 2.1.4 Browsing and listing the data 55
- 2.2 Computing summary statistics 57
- 2.3 Creating a scatter diagram 58
  - 2.3.1 Enhancing the plot 60
- 2.4 Regression 62
  - 2.4.1 Fitted values and residuals 63
  - 2.4.2 Computing an elasticity 66
  - 2.4.3 Plotting the fitted regression line 69

2.4.4	Estimating the variance of the error term	70			combinations of parameters	113
2.4.5	Viewing estimated variances and covariance	71			Appendix 3A Graphical tools	114
2.5	Using Stata to obtain predicted values	73			Appendix 3B Monte Carlo simulation	116
2.5.1	Saving the Stata data file	74			Key Terms	119
2.6	Estimating nonlinear relationships	75			Chapter 3 Do-file	119
2.6.1	A quadratic model	75				
2.6.2	A log-linear model	80				
2.7	Regression with indicator variables	84				
Appendix 2A	Average marginal effects	89				
2A.1	Elasticity in a linear relationship	89				
2A.2	Elasticity in a quadratic relationship	91				
2A.3	Slope in a log-linear model	92				
Appendix 2B	A simulation experiment	93				
Key Terms		96				
Chapter 2 Do-file		97				
<b>CHAPTER 3 Interval Estimation and Hypothesis Testing 103</b>						
3.1	Interval estimates	103				
3.1.1	Critical values from the t-distribution	104				
3.1.2	Creating an interval estimate	105				
3.2	Hypothesis tests	106				
3.2.1	Right-tail test of significance	106				
3.2.2	Right-tail test of an economic hypothesis	108				
3.2.3	Left-tail test of an economic hypothesis	109				
3.2.4	Two-tail test of an economic hypothesis	109				
3.3	p-values	110				
3.3.1	p-value of a right-tail test	111				
3.3.2	p-value of a left-tail test	112				
3.3.3	p-value for a two-tail test	112				
3.3.4	p-values in Stata output	113				
3.3.5	Testing and estimating linear combinations of parameters	113				
<b>CHAPTER 4 Prediction, Goodness-of-Fit and Modeling Issues 123</b>						
4.1	Least squares prediction	123				
4.1.1	Editing the data	124				
4.1.2	Estimate the regression and obtain postestimation results	124				
4.1.3	Creating the prediction interval	125				
4.2	Measuring goodness-of-fit	126				
4.2.1	Correlations and $R^2$	127				
4.3	The effects of scaling and transforming the data	128				
4.3.1	The linear-log functional form	129				
4.3.2	Plotting the fitted linear-log model	131				
4.3.3	Editing graphs	132				
4.4	Analyzing the residuals	134				
4.4.1	The Jarque-Bera test	135				
4.4.2	Chi-square distribution critical values	136				
4.4.3	Chi-square distribution p-values	137				
4.5	Polynomial models	137				
4.5.1	Estimating and checking the linear relationship	138				
4.5.2	Estimating and checking a cubic relationship	141				
4.5.3	Estimating a log-linear yield growth model	143				
4.6	Estimating a log-linear wage equation	144				
4.6.1	The log-linear model	145				
4.6.2	Calculating wage predictions	148				
4.6.3	Constructing wage plots	149				
4.6.4	Generalized $R^2$	150				
4.6.5	Prediction intervals in the log-linear model	150				

- 4.7 A log-log model 151
- Key Terms 154
- Chapter 4 Do-file 154

**CHAPTER 5 Multiple Linear Regression 160**

- 5.1 Big Andy's Burger Barn 160
- 5.2 Least squares prediction 163
- 5.3 Sampling precision 163
- 5.4 Confidence Intervals 165
  - 5.4.1 Confidence interval for a linear combination of parameters 166
- 5.5 Hypothesis Tests 167
  - 5.5.1 Two-sided tests 167
  - 5.5.2 One-sided tests 168
  - 5.5.3 Testing a linear combination 169
- 5.6 Polynomial Equations 170
  - 5.6.1 Optimal advertising: nonlinear combinations of parameters 171
  - 5.6.2 Using factor variables for interactions 172
- 5.7 Interactions 174
- 5.8 Goodness-of-fit 176
- Key Terms 178
- Chapter 5 Do-file 178

**CHAPTER 6 Further Inference in the Multiple Regression Model 181**

- 6.1 The F-Test 181
  - 6.1.1 Testing the significance of the model 185
  - 6.1.2 Relationship between t- and F-tests 187
  - 6.1.3 More general F-tests 188
- 6.2 Nonsample information 191
- 6.3 Model specification 193
  - 6.3.1 Omitted variables 193
  - 6.3.2 Irrelevant variables 195
  - 6.3.3 Choosing the model 196
- 6.4 Poor data, collinearity, and insignificance 203
- Key Terms 206
- Chapter 6 Do-file 206

**CHAPTER 7 Using Indicator Variables 211**

- 7.1 Indicator variables 211
  - 7.1.1 Creating indicator variables 212
  - 7.1.2 Estimating an indicator variable regression 213
  - 7.1.3 Testing the significance of the indicator variables 214
  - 7.1.4 Further calculations 214
  - 7.1.5 Computing average marginal effects 215
- 7.2 Applying indicator variables 217
  - 7.2.1 Interactions between qualitative factors 218
  - 7.2.2 Adding regional indicators 220
  - 7.2.3 Testing the equivalence of two regressions 221
  - 7.2.4 Estimating separate regressions 223
  - 7.2.5 Indicator variables in log-linear models 225
- 7.3 The linear probability model 227
- 7.4 Treatment effects 229
- 7.5 Differences-in-differences estimation 236
- Key Terms 241
- Chapter 7 Do-file 241

**CHAPTER 8 Heteroskedasticity 247**

- 8.1 The nature of heteroskedasticity 247
- 8.2 Detecting heteroskedasticity 249
  - 8.2.1 Residual plots 249
  - 8.2.2 Lagrange multiplier tests 250
  - 8.2.3 The Goldfeld-Quandt test 253
- 8.3 Heteroskedastic-consistent standard errors 255
- 8.4 The generalized least squares estimator 257
  - 8.4.1 GLS using grouped data 259
  - 8.4.2 Feasible GLS-a more general case 260
- 8.5 Heteroskedasticity in the linear probability model 262
- Key Terms 264

Chapter 8 Do-file 264

**CHAPTER 9 Regression with Time-Series  
Data: Stationary Variables 269**

9.1 Introduction 269

9.1.1 Defining time-series in  
Stata 270

9.1.2 Time-series plots 272

9.1.3 Stata's lag and difference  
operators 273

9.2 Finite distributed lags 276

9.3 Serial correlation 277

9.4 Other tests for serial correlation 283

9.5 Estimation with serially correlated  
errors 285

9.5.1 Least squares and HAC  
standard errors 285

9.5.2 Nonlinear least squares 287

9.5.3 A more general model 288

9.6 Autoregressive distributed lag  
models 290

9.6.1 Phillips curve 292

9.6.2 Okun's law 294

9.6.3 Autoregressive models 297

9.7 Forecasting 299

9.7.1 Forecasting with an AR  
model 299

9.7.2 Exponential smoothing 301

9.8 Multiplier analysis 305

9.9 Appendix 308

9.9.1 Durbin-Watson test 308

9.9.2 Prais-Winsten FGLS 309

Key Terms 311

Chapter 9 Do-file 312

**CHAPTER 10 Random Regressors and  
Moment Based Estimation 319**

10.1 Least squares estimation of a wage  
equation 319

10.2 Two-stage least squares 321

10.3 IV estimation with surplus  
instruments 326

10.3.1 Illustrating partial  
correlations 329

10.4 The Hausman test for endogeneity 332

10.5 Testing the validity of surplus  
instruments 335

10.6 Testing for weak instruments 336

10.7 Calculating the Cragg-Donald  
F-statistic 340

10.8 A simulation experiment 342

Key Terms 349

Chapter 10 Do-file 349

**CHAPTER 11 Simultaneous Equations  
Models 357**

11.1 Truffle supply and demand 357

11.2 Estimating the reduced form  
equations 358

11.3 2SLS estimates of truffle demand 359

11.4 2SLS estimates of truffle supply 364

11.5 Supply and demand of fish 365

11.6 Reduced forms for fish price and  
quantity 366

11.7 2SLS estimates of fish demand 368

11.8 2SLS alternatives 369

11.9 Monte Carlo simulation 374

Key Terms 379

Chapter 11 Do-file 379

**CHAPTER 12 Regression with Time-Series  
Data: Nonstationary Variables 385**

12.1 Stationary and nonstationary data 385

12.1.1 Review: generating dates in  
Stata 386

12.1.2 Extracting dates 387

12.1.3 Graphing the data 387

12.2 Spurious regressions 391

12.3 Unit root tests for stationarity 393

12.4 Integration and cointegration 401

12.4.1 Engle-Granger test 401

12.4.2 Error-correction model 403

Key Terms 404

Chapter 12 Do-file 404

**CHAPTER 13 Vector Error Correction and  
Vector Autoregressive Models 407**

13.1 VEC and VAR models 407

13.2 Estimating a VEC model 408

13.3 Estimating a VAR 412

- 13.4 Impulse responses and variance decompositions 419
- Key Terms 423
- Chapter 13 Do-file 423

**CHAPTER 14 Time-Varying Volatility and ARCH Models 426**

- 14.1 ARCH model and time-varying volatility 426
- 14.2 Estimating, testing, and forecasting 428
- 14.3 Extensions 435
  - 14.3.1 GARCH 435
  - 14.3.2 T-GARCH 436
  - 14.3.3 GARCH-in-mean 437
- Key Terms 439
- Chapter 14 Do-file 439

**CHAPTER 15 Panel Data Models 442**

- 15.1 A microeconomic panel 442
- 15.2 A pooled model 444
  - 15.2.1 Cluster-robust standard errors 445
- 15.3 The fixed effects model 446
  - 15.3.1 The fixed effects estimator 448
  - 15.3.2 The fixed effects estimator using xtreg 453
  - 15.3.3 Fixed effects using the complete panel 454
- 15.4 Random effects estimation 458
  - 15.4.1 The GLS transformation 459
  - 15.4.2 The Breusch-Pagan test 462
  - 15.4.3 The Hausman test 463
  - 15.4.4 The Hausman-Taylor model 466
- 15.5 Sets of regression equations 467
  - 15.5.1 Seemingly unrelated regressions 472
  - 15.5.2 SUR with wide data 475
- 15.6 Mixed models 478
- Key Terms 483
- Chapter 15 Do-file 483

**CHAPTER 16 Qualitative and Limited Dependent Variable Models 489**

- 16.1 Models with binary dependent

- variables 489
  - 16.1.1 Average marginal effects 493
  - 16.1.2 Probit marginal effects: details 497
  - 16.1.3 Standard error of average marginal effect 499
- 16.2 The logit model for binary choice 501
  - 16.2.1 Wald tests 504
  - 16.2.2 Likelihood ratio tests 505
  - 16.2.3 Logit estimation 507
  - 16.2.4 Out-of-sample prediction 508
- 16.3 Multinomial logit 509
- 16.4 Conditional logit 514
  - 16.4.1 Estimation using asclogit 517
- 16.5 Ordered choice models 520
- 16.6 Models for count data 523
- 16.7 Censored data models 526
  - 16.7.1 Simulated data example 526
  - 16.7.2 Mroz data example 528
- 16.8 Selection bias 533
- Key Terms 539
- Chapter 16 Do-file 539

**Appendix A Review of Math Essentials 547**

- A.1 Stata math and logical operators 547
- A.2 Math functions 548
- A.3 Extensions to generate 549
- A.4 The calculator 550
- A.5 Scientific notation 550
- A.6 Numerical derivatives and integrals 551
- Key Terms 554
- Appendix A Do-file 554

**Appendix B Review of Probability Concepts 555**

- B.1 Stata probability functions 555
- B.2 Binomial distribution 557
- B.3 Normal distribution 558
  - B.3.1 Normal density plots 558
  - B.3.2 Normal probability

- calculations 559
- B.4 Student's t-distribution 560
  - B.4.1 Plot of standard normal and  $t(3)$  560
  - B.4.2 t-distribution probabilities 561
  - B.4.3 Graphing tail probabilities 562
- B.5 F-distribution 563
  - B.5.1 Plotting the F-density 563
  - B.5.2 F-distribution probability calculations 563
- B.6 Chi-square distribution 564
  - B.6.1 Plotting the chi-square density 564
  - B.6.2 Chi-square probability calculations 565
- B.7 Random numbers 566
  - B.7.1 Using inversion method 567
  - B.7.2 Creating uniform random numbers 569

Key Terms 570

Appendix B Do-file 570

#### **Appendix C Review of Statistical Inference 574**

- C.1 Examining the hip data 574
  - C.1.1 Constructing a histogram 574
  - C.1.2 Obtaining summary statistics 576
  - C.1.3 Estimating the population mean 577
- C.2 Using simulated data values 578
- C.3 The central limit theorem 581
- C.4 Interval estimation 584
  - C.4.1 Using simulated data 585
  - C.4.2 Using the hip data 586
- C.5 Testing the mean of a normal population 587
  - C.5.1 Right-tail test 588
  - C.5.2 Two-tail test 590
- C.6 Testing the variance of a normal population 591
- C.7 Testing the equality of two normal population means 592

- C.7.1 Population variances are equal 592
- C.7.2 Population variances are unequal 592
- C.8 Testing the equality of two normal population variances 594
- C.9 Testing normality 595
- C.10 Maximum likelihood estimation 597
- C.11 Kernel density estimator 597
- Key Terms 602
- Appendix C Do-file 602

**Index 608**