

Contents

<i>Preface</i>	<i>page xv</i>
New to the Second Edition	xv
Guiding Principles Underlying Our Approach	xvi
Overview of Content Coverage and Intended Audience	xvii
<i>Acknowledgments</i>	xix
1 INTRODUCTION	1
The Role of Statistical Software in Data Analysis	1
Statistics: Descriptive and Inferential	2
Variables and Constants	3
The Measurement of Variables	3
Nominal Level	4
Ordinal Level	4
Interval Level	5
Ratio Level	6
Choosing a Scale of Measurement	6
Discrete and Continuous Variables	8
Setting a Context with Real Data	11
Exercises	14
2 EXAMINING UNIVARIATE DISTRIBUTIONS	26
Counting the Occurrence of Data Values	26
When Variables are Measured at the Nominal Level	26
Frequency and Percent Distribution Tables	27
Bar Charts	28
Pie Charts	31
When Variables are Measured at the Ordinal, Interval, or Ratio Level	32
Frequency and Percent Distribution Tables	32
Stem-and-Leaf Displays	35
Histograms	38
Line Graphs	40
Describing the Shape of a Distribution	42
Accumulating Data	44
Cumulative Percent Distributions	44
Ogive Curves	45
Percentile Ranks	46
Percentiles	47
Five-Number Summaries and Boxplots	51
Modifying the Appearance of Graphs	56

	Summary of Graphical Selection	56
	Summary of Stata Commands	56
	Exercises	58
3	MEASURES OF LOCATION, SPREAD, AND SKEWNESS	74
	Characterizing the Location of a Distribution	74
	The Mode	74
	The Median	78
	The Arithmetic Mean	80
	<i>Interpreting the Mean of a Dichotomous Variable</i>	82
	<i>The Weighted Mean</i>	83
	Comparing the Mode, Median, and Mean	84
	Characterizing the Spread of a Distribution	86
	The Range and Interquartile Range	89
	The Variance	91
	The Standard Deviation	93
	Characterizing the Skewness of a Distribution	95
	Selecting Measures of Location and Spread	99
	Applying What We Have Learned	99
	Summary of Stata Commands	104
	Helpful Hints When Using Stata	105
	<i>Online Resources</i>	106
	<i>The Stata Command</i>	106
	<i>Stata Tips</i>	108
	Exercises	109
4	RE-EXPRESSING VARIABLES	118
	Linear and Nonlinear Transformations	118
	Linear Transformations: Addition, Subtraction, Multiplication, and Division	119
	The Effect on the Shape of a Distribution	121
	The Effect on Summary Statistics of a Distribution	121
	Common Linear Transformations	124
	Standard Scores	126
	z-Scores	127
	<i>Using z-Scores to Detect Outliers</i>	130
	<i>Using z-Scores to Compare Scores in Different Distributions</i>	133
	<i>Relating z-Scores to Percentile Ranks</i>	134
	Nonlinear Transformations: Square Roots and Logarithms	135
	Nonlinear Transformations: Ranking Variables	142
	Other Transformations: Recoding and Combining Variables	144
	Recoding Variables	144
	Combining Variables	147
	Data Management Fundamentals: The Do-File	147
	Summary of Stata Commands	150
	Exercises	151
5	EXPLORING RELATIONSHIPS BETWEEN TWO VARIABLES	159
	When Both Variables are at Least Interval-Leveled	159
	Scatterplots	160

CONTENTS

	The Pearson Product-Moment Correlation Coefficient	166
	Interpreting the Pearson Correlation Coefficient	170
	<i>Judging the Strength of the Linear Relationship</i>	170
	<i>The Correlation Scale Itself Is Ordinal</i>	171
	<i>Correlation Does Not Imply Causation</i>	172
	<i>The Effect of Linear Transformations</i>	172
	<i>Restriction of Range</i>	173
	<i>The Shape of the Underlying Distributions</i>	174
	<i>The Reliability of the Data</i>	174
	When at Least One Variable Is Ordinal and the Other Is at Least Ordinal:	
	The Spearman Rank Correlation Coefficient	174
	When at Least One Variable Is Dichotomous: Other Special Cases of the	
	Pearson Correlation Coefficient	176
	The Point Biserial Correlation Coefficient: The Case of One at Least	
	Interval and One Dichotomous Variable	176
	The Phi Coefficient: The Case of Two Dichotomous Variables	181
	Other Visual Displays of Bivariate Relationships	185
	Selection of Appropriate Statistic or Graph to Summarize a Relationship	188
	Summary of Stata Commands	189
	Exercises	189
6	SIMPLE LINEAR REGRESSION	202
	The “Best-Fitting” Linear Equation	202
	The Accuracy of Prediction Using the Linear Regression Model	209
	The Standardized Regression Equation	210
	<i>R</i> As a Measure of the Overall Fit of the Linear Regression Model	210
	Simple Linear Regression When the Independent Variable Is	
	Dichotomous	214
	Using <i>r</i> and <i>R</i> As Measures of Effect Size	217
	Emphasizing the Importance of the Scatterplot	217
	Summary of Stata Commands	219
	Exercises	219
7	PROBABILITY FUNDAMENTALS	228
	The Discrete Case	228
	The Complement Rule of Probability	230
	The Additive Rules of Probability	231
	First Additive Rule of Probability	231
	Second Additive Rule of Probability	232
	The Multiplicative Rule of Probability	233
	The Relationship between Independence and Mutual Exclusivity	236
	Conditional Probability	236
	The Law of Total Probability	239
	Bayes’ Theorem	239
	The Law of Large Numbers	240
	Exercises	240
8	THEORETICAL PROBABILITY MODELS	244
	The Binomial Probability Model and Distribution	244
	The Applicability of the Binomial Probability Model	249

CONTENTS

	The Normal Probability Model and Distribution	254
	Using the Normal Distribution to Approximate the Binomial Distribution	260
	Summary of Stata Commands	260
	Exercises	261
9	THE ROLE OF SAMPLING IN INFERENTIAL STATISTICS	269
	Samples and Populations	269
	Random Samples	270
	Obtaining a Simple Random Sample	271
	Sampling with and without Replacement	273
	Sampling Distributions	275
	Describing the Sampling Distribution of Means Empirically	275
	Describing the Sampling Distribution of Means Theoretically	280
	The Central Limit Theorem	281
	Estimators and Bias	285
	Summary of Stata Commands	286
	Exercises	287
10	INFERENCES INVOLVING THE MEAN OF A SINGLE POPULATION WHEN σ IS KNOWN	291
	Estimating the Population Mean, μ , When the Population Standard Deviation, σ , Is Known	291
	Interval Estimation	293
	Relating the Length of a Confidence Interval, the Level of Confidence, and the Sample Size	296
	Hypothesis Testing	296
	The Relationship between Hypothesis Testing and Interval Estimation	305
	Effect Size	306
	Type II Error and the Concept of Power	307
	Increasing the Level of Significance, α	310
	Increasing the Effect Size, δ	310
	Decreasing the Standard Error of the Mean, $\sigma_{\bar{x}}$	311
	Closing Remarks	312
	Summary of Stata Commands	313
	Exercises	314
11	INFERENCES INVOLVING THE MEAN WHEN σ IS NOT KNOWN: ONE- AND TWO-SAMPLE DESIGNS	319
	Single Sample Designs When the Parameter of Interest Is the Mean and σ Is Not Known	319
	The t -Distribution	320
	Degrees of Freedom for the One-Sample t -Test	321
	Violating the Assumption of a Normally Distributed Parent Population in the One-Sample t -Test	322
	Confidence Intervals for the One-Sample t -Test	323
	Hypothesis Tests: The One-Sample t -Test	330
	Effect Size for the One-Sample t -Test	333
	Two-Sample Designs When the Parameter of Interest Is μ , and σ Is Not Known	336
	Independent (or Unrelated) and Dependent (or Related) Samples	337

CONTENTS

Independent Samples t -Test and Confidence Interval	338
The Assumptions of the Independent Samples t -Test	340
Effect Size for the Independent Samples t -Test	349
Paired Samples t -Test and Confidence Interval	353
The Assumptions of the Paired Samples t -Test	354
Effect Size for the Paired Samples t -Test	359
The Bootstrap	360
Conducting Power Analyses for t -Tests on Means	364
Summary	369
Summary of Stata Commands	372
Exercises	374
12 RESEARCH DESIGN: INTRODUCTION AND OVERVIEW	391
Questions and their Link to Descriptive, Relational, and Causal Research Studies	391
The Need for a Good Measure of our Construct: Weight	391
The Descriptive Study	392
From Descriptive to Relational Studies	393
From Relational to Causal Studies	393
The Gold Standard of Causal Studies: The True Experiment and Random Assignment	395
Comparing Two Kidney Stone Treatments Using a Non-Randomized Controlled Study	396
Including Blocking in a Research Design	397
Underscoring the Importance of Having a True Control Group Using Randomization	398
Analytic Methods for Bolstering Claims of Causality from Observational Data	402
Quasi-Experimental Designs	404
Threats to the Internal Validity of a Quasi-Experimental Design	404
Threats to the External Validity of a Quasi-Experimental Design	405
Threats to the Validity of a Study: Some Clarifications and Caveats	406
Threats to the Validity of a Study: Some Examples	407
Exercises	408
13 ONE-WAY ANALYSIS OF VARIANCE	412
The Disadvantage of Multiple t -Tests	412
The One-Way Analysis of Variance	414
A Graphical Illustration of the Role of Variance in Tests on Means	414
ANOVA As an Extension of the Independent Samples t -Test	416
Developing an Index of Separation for the Analysis of Variance	416
Carrying Out the ANOVA Computation	417
The Between Group Variance (MS_B)	418
The Within Group Variance (MS_W)	418
The Assumptions of the One-Way ANOVA	419
Testing the Equality of Population Means: The F -Ratio	420
How to Read the Tables and Use Stata Functions for the F -Distribution	422
ANOVA Summary Table	425
Measuring the Effect Size	426
Post-Hoc Multiple Comparison Tests	431

	The Bonferroni Adjustment: Testing Planned Comparisons	444
	The Bonferroni Tests on Multiple Measures	446
	Conducting Power Analyses for One-Way ANOVA	447
	Summary of Stata Commands	450
	Exercises	451
14	TWO-WAY ANALYSIS OF VARIANCE	457
	The Two-Factor Design	457
	The Concept of Interaction	460
	The Hypotheses That are Tested by a Two-Way Analysis of Variance	465
	Assumptions of the Two-Way Analysis of Variance	466
	Balanced versus Unbalanced Factorial Designs	467
	Partitioning the Total Sum of Squares	468
	Using the F-Ratio to Test the Effects in Two-Way ANOVA	469
	Carrying Out the Two-Way ANOVA Computation by Hand	469
	Decomposing Score Deviations about the Grand Mean	474
	Modeling Each Score As a Sum of Component Parts	475
	Explaining the Interaction As a Joint (or Multiplicative) Effect	475
	Measuring Effect Size	476
	Fixed versus Random Factors	479
	Post-Hoc Multiple Comparison Tests	479
	Simple Effects and Pairwise Comparisons	482
	Summary of Steps to Be Taken in a Two-Way ANOVA Procedure	487
	Conducting Power Analyses for Two-Way ANOVA	491
	Summary of Stata Commands	493
	Exercises	495
15	CORRELATION AND SIMPLE REGRESSION AS INFERENTIAL TECHNIQUES	503
	The Bivariate Normal Distribution	503
	Testing whether the Population Pearson Product-Moment Correlation Equals Zero	506
	Using a Confidence Interval to Estimate the Size of the Population Correlation Coefficient, ρ	509
	Revisiting Simple Linear Regression for Prediction	512
	Estimating the Population Standard Error of Prediction, $\sigma_{Y X}$	513
	Testing the b -Weight for Statistical Significance	514
	Explaining Simple Regression Using an Analysis of Variance Framework	518
	Measuring the Fit of the Overall Regression Equation: Using R and R^2	520
	Relating R^2 to $\sigma^2_{Y X}$	521
	Testing R^2 for Statistical Significance	522
	Estimating the True Population R^2 : The Adjusted R^2	523
	Exploring the Goodness of Fit of the Regression Equation: Using Regression Diagnostics	524
	Residual Plots: Evaluating the Assumptions Underlying Regression	526
	Detecting Influential Observations: Discrepancy and Leverage	529
	Using Stata to Obtain Leverage	530
	Using Stata to Obtain Discrepancy	531
	Using Stata to Obtain Influence	531
	Using Diagnostics to Evaluate the Ice Cream Sales Example	533

Using the Prediction Model to Predict Ice Cream Sales	536
Simple Regression When the Predictor Is Dichotomous	536
Conducting Power Analyses for Correlation and Simple Regression	538
Summary of Stata Commands	540
Exercises	541
16 AN INTRODUCTION TO MULTIPLE REGRESSION	553
The Basic Equation with Two Predictors	554
Equations for b , β and $R_{Y.12}$ When the Predictors Are Not Correlated	555
Equations for b , β , and $R_{Y.12}$ When the Predictors Are Correlated	556
Summarizing and Expanding on Some Important Principles of Multiple Regression	558
Testing the b -Weights for Statistical Significance	563
Assessing the Relative Importance of the Predictors in the Equation	565
Measuring the Drop in R^2 Directly: An Alternative to the Squared Semipartial Correlation	566
Evaluating the Statistical Significance of the Change in R^2	566
The b -Weight As a Partial Slope in Multiple Regression	568
Multiple Regression When One of the Two Independent Variables Is Dichotomous	571
Controlling Variables Statistically: A Closer Look	576
A Hypothetical Example	577
Conducting Power Analyses for Multiple Regression	580
Summary of Stata Commands	582
Exercises	583
17 TWO-WAY INTERACTIONS IN MULTIPLE REGRESSION	590
Testing the Statistical Significance of an Interaction Using Stata	593
Comparing the Y-Hat Values from the Additive and Interaction Models	598
Centering First-Order Effects if the Equation Has an Interaction	599
Probing the Nature of a Two-Way Interaction	600
Interaction When One of the Independent Variables Is Dichotomous and the Other Is Continuous	603
Methods Useful for Model Selection	610
Conducting a Power Analysis to Detect an Interaction	613
Summary of Stata Commands	614
Exercises	617
18 NONPARAMETRIC METHODS	622
Parametric versus Nonparametric Methods	622
Nonparametric Methods When the Dependent Variable Is at the Nominal Level	623
The Chi-Square Distribution (χ^2)	623
The Chi-Square Goodness-of-Fit Test	625
The Chi-Square Test of Independence	630
<i>Assumptions of the Chi-Square Test of Independence</i>	633
Fisher's Exact Test	635
<i>Calculating the Fisher's Exact Test by Hand Using the Hypergeometric Distribution</i>	637

CONTENTS

Nonparametric Methods When the Dependent Variable Is Ordinal-Leveled	639
Wilcoxon Sign Test	640
The Mann–Whitney <i>U</i> -Test or Wilcoxon’s Rank-Sum Test	642
The Kruskal–Wallis Analysis of Variance	647
Summary of Stata Commands	649
Exercises	650
19 COMMUNICATING YOUR STATA RESULTS VIA EXCEL	655
Setting the Working Directory	655
Reproducing a Table of Univariate Summary Statistics in Excel	656
Using estpost and esttab	656
Using putexcel	657
Reproducing a Correlation Matrix As a Table in Excel	661
Using estpost and esttab	661
Using putexcel	662
Reproducing Regression Output As a Table in Excel	663
Using outreg2 to obtain a table of model statistics in Excel	663
Using eststo and esttab to obtain a table of model statistics in Excel	663
Using putexcel to reproduce a table of regression coefficients in Excel	664
Reproducing a Graph in Excel (Using putexcel)	666
Conclusion	668
Summary of Stata Commands	668
Exercises	671
<i>Appendix A Data Set Descriptions</i>	673
<i>Appendix B Stata .Do-files and Data Sets in Stata Format</i>	686
<i>Appendix C Statistical Tables</i>	688
<i>Appendix D Solutions</i>	708
<i>References</i>	709
<i>Index</i>	713