
Contents

Preface.....	xv
--------------	----

Part I Fundamentals

1. Statistical Inference I: Descriptive Statistics.....	3
1.1 Measures of Relative Standing.....	3
1.2 Measures of Central Tendency.....	4
1.3 Measures of Variability	5
1.4 Skewness and Kurtosis	9
1.5 Measures of Association	11
1.6 Properties of Estimators.....	14
1.6.1 Unbiasedness.....	14
1.6.2 Efficiency	15
1.6.3 Consistency	16
1.6.4 Sufficiency	16
1.7 Methods of Displaying Data	17
1.7.1 Histograms	17
1.7.2 Ogives	18
1.7.3 Box Plots.....	19
1.7.4 Scatter Diagrams.....	19
1.7.5 Bar and Line Charts.....	20
2. Statistical Inference II: Interval Estimation, Hypothesis Testing and Population Comparisons.....	25
2.1 Confidence Intervals.....	25
2.1.1 Confidence Interval for μ with Known σ^2	26
2.1.2 Confidence Interval for the Mean with Unknown Variance.....	28
2.1.3 Confidence Interval for a Population Proportion.....	28
2.1.4 Confidence Interval for the Population Variance.....	29
2.2 Hypothesis Testing	30
2.2.1 Mechanics of Hypothesis Testing.....	31
2.2.2 Formulating One- and Two-Tailed Hypothesis Tests.....	33
2.2.3 The p -Value of a Hypothesis Test	36
2.3 Inferences Regarding a Single Population	36
2.3.1 Testing the Population Mean with Unknown Variance.....	37

2.3.2	Testing the Population Variance	38
2.3.3	Testing for a Population Proportion.....	38
2.4	Comparing Two Populations.....	39
2.4.1	Testing Differences between Two Means: Independent Samples	39
2.4.2	Testing Differences between Two Means: Paired Observations	42
2.4.3	Testing Differences between Two Population Proportions	43
2.4.4	Testing the Equality of Two Population Variances	45
2.5	Nonparametric Methods.....	46
2.5.1	Sign Test	47
2.5.2	Median Test.....	52
2.5.3	Mann-Whitney U Test	52
2.5.4	Wilcoxon Signed-Rank Test for Matched Pairs	55
2.5.5	Kruskal-Wallis Test.....	56
2.5.6	Chi-Square Goodness-of-Fit Test.....	58

Part II Continuous Dependent Variable Models

3.	Linear Regression	63
3.1	Assumptions of the Linear Regression Model	63
3.1.1	Continuous Dependent Variable Y	64
3.1.2	Linear-in-Parameters Relationship between Y and X	64
3.1.3	Observations Independently and Randomly Sampled	65
3.1.4	Uncertain Relationship between Variables	65
3.1.5	Disturbance Term Independent of X and Expected Value Zero	65
3.1.6	Disturbance Terms Not Autocorrelated	66
3.1.7	Regressors and Disturbances Uncorrelated.....	66
3.1.8	Disturbances Approximately Normally Distributed	66
3.1.9	Summary.....	67
3.2	Regression Fundamentals.....	67
3.2.1	Least Squares Estimation.....	69
3.2.2	Maximum Likelihood Estimation.....	73
3.2.3	Properties of OLS and MLE Estimators.....	74
3.2.4	Inference in Regression Analysis	75
3.3	Manipulating Variables in Regression.....	79
3.3.1	Standardized Regression Models.....	79
3.3.2	Transformations	80
3.3.3	Indicator Variables	82

3.4	Estimate a Single Beta Parameter	83
3.5	Estimate Beta Parameter for Ranges of a Variable	83
3.6	Estimate a Single Beta Parameter for $m - 1$ of the m Levels of a Variable	84
3.6.1	Interactions in Regression Models	84
3.7	Checking Regression Assumptions	87
3.7.1	Linearity	88
3.7.2	Homoscedastic Disturbances	90
3.7.3	Uncorrelated Disturbances	93
3.7.4	Exogenous Independent Variables	93
3.7.5	Normally Distributed Disturbances	95
3.8	Regression Outliers	98
3.8.1	The Hat Matrix for Identifying Outlying Observations	99
3.8.2	Standard Measures for Quantifying Outlier Influence	101
3.8.3	Removing Influential Data Points from the Regression	101
3.9	Regression Model GOF Measures	106
3.10	Multicollinearity in the Regression	110
3.11	Regression Model-Building Strategies	112
3.11.1	Stepwise Regression	112
3.11.2	Best Subsets Regression	113
3.11.3	Iteratively Specified Tree-Based Regression	113
3.12	Estimating Elasticities	113
3.13	Censored Dependent Variables—Tobit Model	114
3.14	Box-Cox Regression	116
4.	Violations of Regression Assumptions	123
4.1	Zero Mean of the Disturbances Assumption	123
4.2	Normality of the Disturbances Assumption	124
4.3	Uncorrelatedness of Regressors and Disturbances Assumption	125
4.4	Homoscedasticity of the Disturbances Assumption	127
4.4.1	Detecting Heteroscedasticity	129
4.4.2	Correcting for Heteroscedasticity	131
4.5	No Serial Correlation in the Disturbances Assumption	135
4.5.1	Detecting Serial Correlation	137
4.5.2	Correcting for Serial Correlation	139
4.6	Model Specification Errors	142
5.	Simultaneous-Equation Models	145
5.1	Overview of the Simultaneous-Equations Problem	145
5.2	Reduced Form and the Identification Problem	146

5.3	Simultaneous-Equation Estimation.....	148
5.3.1	Single-Equation Methods	148
5.3.2	System-Equation Methods.....	149
5.4	Seemingly Unrelated Equations	155
5.5	Applications of Simultaneous Equations to Transportation Data	156
	Appendix 5A. A Note on GLS Estimation.....	159
6.	Panel Data Analysis	161
6.1	Issues in Panel Data Analysis.....	161
6.2	One-Way Error Component Models.....	163
6.2.1	Heteroscedasticity and Serial Correlation	166
6.3	Two-Way Error Component Models.....	167
6.4	Variable-Parameter Models	172
6.5	Additional Topics and Extensions	173
7.	Background and Exploration in Time Series.....	175
7.1	Exploring a Time Series	176
7.1.1	Trend Component.....	176
7.1.2	Seasonal Component.....	176
7.1.3	Irregular (Random) Component.....	179
7.1.4	Filtering of Time Series	179
7.1.5	Curve Fitting.....	179
7.1.6	Linear Filters and Simple Moving Averages.....	179
7.1.7	Exponential Smoothing Filters	180
7.1.8	Difference Filter.....	185
7.2	Basic Concepts: Stationarity and Dependence.....	188
7.2.1	Stationarity.....	188
7.2.2	Dependence	188
7.2.3	Addressing Nonstationarity	190
7.2.4	Differencing and Unit-Root Testing.....	191
7.2.5	Fractional Integration and Long Memory.....	194
7.3	Time Series in Regression.....	197
7.3.1	Serial Correlation.....	197
7.3.2	Dynamic Dependence.....	197
7.3.3	Volatility	198
7.3.4	Spurious Regression and Cointegration.....	200
7.3.5	Causality	202
8.	Forecasting in Time Series: Autoregressive Integrated Moving Average (ARIMA) Models and Extensions	207
8.1	Autoregressive Integrated Moving Average Models	207
8.2	Box-Jenkins Approach	210

8.2.1	Order Selection.....	210
8.2.2	Parameter Estimation.....	212
8.2.3	Diagnostic Checking.....	213
8.2.4	Forecasting.....	214
8.3	Autoregressive Integrated Moving Average Model Extensions.....	218
8.3.1	Random Parameter Autoregressive Models.....	219
8.3.2	Stochastic Volatility Models.....	222
8.3.3	Autoregressive Conditional Duration Models.....	224
8.3.4	Integer-Valued ARMA Models.....	224
8.4	Multivariate Models.....	225
8.5	Nonlinear Models.....	227
8.5.1	Testing for Nonlinearity.....	227
8.5.2	Bilinear Models.....	228
8.5.3	Threshold Autoregressive Models.....	229
8.5.4	Functional Parameter Autoregressive Models.....	230
8.5.5	Neural Networks.....	231
9.	Latent Variable Models.....	235
9.1	Principal Components Analysis.....	235
9.2	Factor Analysis.....	241
9.3	Structural Equation Modeling.....	244
9.3.1	Basic Concepts in Structural Equation Modeling.....	246
9.3.2	Fundamentals of Structural Equation Modeling.....	249
9.3.3	Nonideal Conditions in the Structural Equation Model.....	251
9.3.4	Model Goodness-of-Fit Measures.....	252
9.3.5	Guidelines for Structural Equation Modeling.....	255
10.	Duration Models.....	259
10.1	Hazard-Based Duration Models.....	259
10.2	Characteristics of Duration Data.....	263
10.3	Nonparametric Models.....	264
10.4	Semiparametric Models.....	265
10.5	Fully Parametric Models.....	268
10.6	Comparisons of Nonparametric, Semiparametric, and Fully Parametric Models.....	272
10.7	Heterogeneity.....	274
10.8	State Dependence.....	276
10.9	Time-Varying Covariates.....	277
10.10	Discrete-Time Hazard Models.....	277
10.11	Competing Risk Models.....	279

Part III Count and Discrete Dependent Variable Models

11. Count Data Models	283
11.1 Poisson Regression Model	283
11.2 Interpretation of Variables in the Poisson Regression Model	284
11.3 Poisson Regression Model Goodness-of-Fit Measures.....	286
11.4 Truncated Poisson Regression Model	290
11.5 Negative Binomial Regression Model.....	292
11.6 Zero-Inflated Poisson and Negative Binomial Regression Models	295
11.7 Random-Effects Count Models	300
12. Logistic Regression	303
12.1 Principles of Logistic Regression.....	303
12.2 Logistic Regression Model.....	304
13. Discrete Outcome Models	309
13.1 Models of Discrete Data	309
13.2 Binary and Multinomial Probit Models.....	310
13.3 Multinomial Logit Model	312
13.4 Discrete Data and Utility Theory	316
13.5 Properties and Estimation of MNL Models.....	318
13.5.1 Statistical Evaluation	321
13.5.2 Interpretation of Findings	323
13.5.3 Specification Errors.....	325
13.5.4 Data Sampling	330
13.5.5 Forecasting and Aggregation Bias	331
13.5.6 Transferability	333
13.6 Nested Logit Model (Generalized Extreme Value Models).....	334
13.7 Special Properties of Logit Models.....	342
14. Ordered Probability Models	345
14.1 Models for Ordered Discrete Data	345
14.2 Ordered Probability Models with Random Effects	352
14.3 Limitations of Ordered Probability Models.....	358
15. Discrete/Continuous Models	361
15.1 Overview of the Discrete/Continuous Modeling Problem	361
15.2 Econometric Corrections: Instrumental Variables and Expected Value Method	363
15.3 Econometric Corrections: Selectivity-Bias Correction Term.....	366
15.4 Discrete/Continuous Model Structures	368
15.5 Transportation Application of Discrete/Continuous Model Structures	372

Part IV Other Statistical Methods

16. Random-Parameter Models 375

16.1 Random-Parameter Multinomial Logit Model (Mixed Logit Model) 375

16.2 Random-Parameter Count Models 381

16.3 Random-Parameter Duration Models 384

17. Bayesian Models 387

17.1 Bayes' Theorem 387

17.2 MCMC Sampling-Based Estimation 389

17.3 Flexibility of Bayesian Statistical Models via MCMC Sampling-Based Estimation 395

17.4 Convergence and Identifiability Issues with MCMC Bayesian Models 396

17.5 Goodness-of-Fit, Sensitivity Analysis, and Model Selection Criterion Using MCMC Bayesian Models 399

Appendix A Statistical Fundamentals 403

A.1 Matrix Algebra Review 403

A.1.1 Matrix Multiplication 404

A.1.2 Linear Dependence and Rank of a Matrix 406

A.1.3 Matrix Inversion (Division) 406

A.1.4 Eigenvalues and Eigenvectors 408

A.1.5 Useful Matrices and Properties of Matrices 409

A.1.6 Matrix Algebra and Random Variables 410

A.2 Probability, Conditional Probability, and Statistical Independence 412

A.3 Estimating Parameters in Statistical Models—Least Squares and Maximum Likelihood 413

A.4 Useful Probability Distributions 415

A.4.1 The Z Distribution 416

A.4.2 The *t* Distribution 417

A.4.3 The χ^2 Distribution 418

A.4.4 The *F* Distribution 419

Appendix B Glossary of Terms 421

Appendix C Statistical Tables 459

Appendix D Variable Transformations 483

D.1 Purpose of Variable Transformations 483

D.2 Commonly Used Variable Transformations 484

D.2.1 Parabolic Transformations 484

D.2.2 Hyperbolic Transformations485

D.2.3 Exponential Functions485

D.2.4 Inverse Exponential Functions487

D.2.5 Power Functions.....488

References489

Index511