

Statistical AND Econometric Methods FOR Transportation Data Analysis



Simon P. Washington
Matthew G. Karlaftis
Fred L. Mannering



CHAPMAN & HALL/CRC

Contents

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	8
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	17
1.7	Methods of Displaying Data	17
1.7.1	Histograms	17
1.7.2	Ogives	17
1.7.3	Box Plots	18
1.7.4	Scatter Diagrams	19
1.7.5	Bar and Line Charts.....	19
2	Statistical Inference II: Interval Estimation, Hypothesis Testing, and Population Comparisons.....	23
2.1	Confidence Intervals.....	23
2.1.1	Confidence Interval for m with Known σ^2	24
2.1.2	Confidence Interval for the Mean with Unknown Variance	26
2.1.3	Confidence Interval for a Population Proportion.....	27
2.1.4	Confidence Interval for the Population Variance.....	27
2.2	Hypothesis Testing.....	28
2.2.1	Mechanics of Hypothesis Testing.....	30
2.2.2	Formulating One- and Two-Tailed Hypothesis Tests	32
2.2.3	The p -Value of a Hypothesis Test.....	34
2.3	Inferences Regarding a Single Population.....	35
2.3.1	Testing the Population Mean with Unknown Variance ...	35
2.3.2	Testing the Population Variance	36
2.3.3	Testing for a Population Proportion	37
2.4	Comparing Two Populations	38
2.4.1	Testing Differences between Two Means: Independent Samples	38

2.4.2	Testing Differences between Two Means: Paired Observations.....	41
2.4.3	Testing Differences between Two Population Proportions	42
2.4.4	Testing the Equality of Two Population Variances	43
2.5	Nonparametric Methods.....	45
2.5.1	The Sign Test.....	46
2.5.2	The Median Test.....	50
2.5.3	The Mann-Whitney <i>U</i> Test	51
2.5.4	The Wilcoxon Signed-Rank Test for Matched Pairs.....	54
2.5.5	The Kruskal-Wallis Test.....	55
2.5.6	The Chi-Square Goodness-of-Fit Test	56

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 <i>Y</i>	64
3.1.2	Linear-in-Parameters Relationship between <i>Y</i> and <i>X</i>	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 <i>X</i> 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.....	68
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
3.3.3.1	Estimate a Single Beta Parameter.....	83
3.3.3.2	Estimate Beta Parameter for Ranges of the Variable.....	83
3.3.3.3	Estimate a Single Beta Parameter for $m - 1$ of the m Levels of the Variable	84
3.3.4	Interactions in Regression Models.....	84
3.4	Checking Regression Assumptions.....	87
3.4.1	Linearity.....	88

3.4.2	Homoscedastic Disturbances	91
3.4.3	Uncorrelated Disturbances	93
3.4.4	Exogenous Independent Variables.....	93
3.4.5	Normally Distributed Disturbances.....	96
3.5	Regression Outliers.....	99
3.5.1	The Hat Matrix for Identifying Outlying Observations.....	99
3.5.2	Standard Measures for Quantifying Outlier Influence	101
3.5.3	Removing Influential Data Points from the Regression	102
3.6	Regression Model Goodness-of-Fit Measures	106
3.7	Multicollinearity in the Regression.....	110
3.8	Regression Model-Building Strategies.....	112
3.8.1	Stepwise Regression	113
3.8.2	Best Subsets Regression	113
3.8.3	Iteratively Specified Tree-Based Regression.....	113
3.9	Logistic Regression	114
3.10	Lags and Lag Structure.....	115
3.11	Investigating Causality in the Regression	117
3.12	Limited Dependent Variable Models.....	118
3.13	Box–Cox Regression	119
3.14	Estimating Elasticities.....	120
4	Violations of Regression Assumptions	121
4.1	Zero Mean of the Disturbances Assumption	121
4.2	Normality of the Disturbances Assumption	122
4.3	Uncorrelatedness of Regressors and Disturbances Assumption	123
4.4	Homoscedasticity of the Disturbances Assumption	125
4.4.1	Detecting Heteroscedasticity	127
4.4.2	Correcting for Heteroscedasticity.....	128
4.5	No Serial Correlation in the Disturbances Assumption.....	132
4.5.1	Detecting Serial Correlation	135
4.5.2	Correcting for Serial Correlation.....	136
4.6	Model Specification Errors	139
5	Simultaneous Equation Models	143
5.1	Overview of the Simultaneous Equations Problem.....	143
5.2	Reduced Form and the Identification Problem.....	144
5.3	Simultaneous Equation Estimation.....	146
5.3.1	Single-Equation Methods.....	146
5.3.2	System Equation Methods.....	147
5.4	Seemingly Unrelated Equations	150
5.5	Applications of Simultaneous Equations to Transportation Data.....	152
	Appendix 5A: A Note on Generalized Least Squares Estimation	152

6	Panel Data Analysis	155
6.1	Issues in Panel Data Analysis	155
6.2	One-Way Error Component Models	157
6.2.1	Heteroscedasticity and Serial Correlation.....	160
6.3	Two-Way Error Component Models.....	161
6.4	Variable Coefficient Models	166
6.5	Additional Topics and Extensions.....	168
7	Time-Series Analysis	169
7.1	Characteristics of Time Series	169
7.1.1	Long-Term Movements	170
7.1.2	Seasonal Movements	170
7.1.3	Cyclic Movements.....	171
7.1.4	Irregular or Random Movements	171
7.2	Smoothing Methodologies.....	171
7.2.1	Simple Moving Averages.....	172
7.2.2	Exponential Smoothing	173
7.3	The ARIMA Family of Models	177
7.3.1	The ARIMA Models.....	180
7.3.2	Estimating ARIMA Models	182
7.4	Nonlinear Time-Series Models.....	184
7.4.1	Conditional Mean Models.....	184
7.4.2	Conditional Variance Models.....	185
7.4.3	Mixed Models.....	187
7.4.4	Regime Models.....	187
7.5	Multivariate Time-Series Models.....	188
7.6	Measures of Forecasting Accuracy.....	189
8	Latent Variable Models	193
8.1	Principal Components Analysis	194
8.2	Factor Analysis	199
8.3	Structural Equation Modeling	204
8.3.1	Basic Concepts in Structural Equation Modeling.....	204
8.3.2	The Structural Equation Model	208
8.3.3	Non-Ideal Conditions in the Structural Equation Model.....	210
8.3.4	Model Goodness-of-Fit Measures	211
8.3.5	Guidelines for Structural Equation Modeling.....	213
9	Duration Models	217
9.1	Hazard-Based Duration Models.....	217
9.2	Characteristics of Duration Data.....	221
9.3	Nonparametric Models	222
9.4	Semiparametric Models	223

9.5	Fully Parametric Models.....	226
9.6	Comparisons of Nonparametric, Semiparametric, and Fully Parametric Models.....	229
9.7	Heterogeneity.....	231
9.8	State Dependence.....	233
9.9	Time-Varying Covariates	234
9.10	Discrete-Time Hazard Models	235
9.11	Competing Risk Models	236

Part III Count and Discrete Dependent Variable Models

10	Count Data Models.....	241
10.1	Poisson Regression Model.....	241
10.2	Poisson Regression Model Goodness-of-Fit Measures	243
10.3	Truncated Poisson Regression Model.....	247
10.4	Negative Binomial Regression Model	248
10.5	Zero-Inflated Poisson and Negative Binomial Regression Models	250
10.6	Panel Data and Count Models.....	254
11	Discrete Outcome Models.....	257
11.1	Models of Discrete Data.....	257
11.2	Binary and Multinomial Probit Models	258
11.3	Multinomial Logit Model	260
11.4	Discrete Data and Utility Theory	264
11.5	Properties and Estimation of Multinomial Logit Models	266
11.5.1	Statistical Evaluation.....	269
11.5.2	Interpretation of Findings.....	271
11.5.3	Specification Errors	273
11.5.4	Data Sampling	278
11.5.5	Forecasting and Aggregation Bias.....	280
11.5.6	Transferability	282
11.6	Nested Logit Model (Generalized Extreme Value Model).....	283
11.7	Special Properties of Logit Models	286
11.8	Mixed MNL Models	287
11.9	Models of Ordered Discrete Data	288
12	Discrete/Continuous Models	297
12.1	Overview of the Discrete/Continuous Modeling Problem	297
12.2	Econometric Corrections: Instrumental Variables and Expected Value Method	299
12.3	Econometric Corrections: Selectivity-Bias Correction Term.....	302
12.4	Discrete/Continuous Model Structures	305

Appendix A: Statistical Fundamentals	309
Appendix B: Glossary of Terms	327
Appendix C: Statistical Tables	367
Appendix D: Variable Transformations.....	389
References.....	395
Index	413