Transportation Research Forum

Statistical and Econometric Methods for Transportation Data Analysis

By Brian W. Sloboda

Transportation statistics is an evolving field with a greater abundance of transportation data and improvements in computing power to handle the complexities in statistical methods that provide empirical results. *Statistical and Econometric Methods for Transportation Data Analysis* provides the reader a comprehensive presentation of statistical methods that can be applied in transportation.

The book does not overwhelm the reader with the analytical details for each of these methods, yet it still presents each of the statistical methods rather well. In addition, the description of the methods is detailed enough to provide a solid understanding of each technique. Also, the authors sprinkle throughout the book the pitfalls or oversights in the application of these methods. A major advantage of this book is the authors' inclusion of such statistical methods as Bayesian methods, logistic regression, ordered probability models, random parameter models, and additional time series methods.

The book is divided into 17 chapters, grouped by the authors into four sections. Section 1, which consists of chapters 1 and 2, provides a rudimentary review of statistical methods: descriptive statistics, confidence intervals, hypothesis testing for a single mean, hypothesis testing comparing two means, and some nonparametric methods. The authors clearly provide the reasoning for the use of nonparametric statistics rather than parametric methods. Because of the comprehensive nature of statistical methods covered in this book, the authors relegated the details of nonparametric methods to other references given in this book. The reader should consult these other references for those details.

The second section, comprising chapters 3 through 10, is dedicated to discussing *continuous* dependent variable models. Chapters 3 and 4 delve into the details of regression analysis, which forms the primary foundation of the methods used in transportation research. Chapter 3 provides a rudimentary introduction to regression analysis, but the authors stress the importance of checking the assumptions of regression analysis that are often mentioned briefly if at all in other books. Having covered the assumptions of regression analysis, the authors turn to variable manipulation, outlier identification, identification, goodness-of-fit measures, multicollinearity, model-building strategies, and causality.

Chapter 4 addresses violations of the regression assumptions, namely heteroscedasticity, serial correlation, and model specification errors. As each of the violations is presented, the authors discuss how to detect it and how to most appropriately correct it.

The remaining chapters in Section 2 cover more advanced methods pertaining to continuous variables. Chapter 5 delves into simultaneous equation models and when to use them. As the authors explain, the seemingly unrelated equations (SURE) method is used when there is contemporaneous correlation in the error terms. In an appendix, the authors present a brief discussion of generalized least squares (GLS), often used in econometric analysis. Chapter 6 deals with panel data methods, often used in microeconometric analysis. While not giving these methods a full treatment, the authors do provide a foundation for understanding their use.

Chapter 7 delves into the rudiments of time series data, which play an important role in transportation analysis. This chapter explains the descriptive assessment of time series data, smoothing techniques, and the concepts of stationarity and dependence. In the discussions of stationarity and dependence, the authors delve into unit root testing, namely the Dickey-Fuller test. There is also some discussion of fractional integration and long memory as applied to transportation data. The latter discussion is, however, rather brief, and the reader will need to consult references for greater details of this method. Chapter 8 continues the study of time series analysis through the autoregressive integrated moving average (ARIMA). Then the authors deal with variations of the ARIMA models and expand the discussion to nonlinear models as well as multivariate models. The final part of the chapter presents a discussion and brief descriptive analysis of neural networks, though with little attention focused on transportation.

Chapter 9 presents latent variable models, which are used in transportation analysis when there are measurement difficulties and unobservable variables. This chapter specifically covers the methods of principal component analysis, factor analysis, and structural equation modeling. Chapter 10 presents duration models, which deal with instances of elapsed time until the occurrence of an event or the duration of an event. Covered here are hazard-based duration models, nonparametric models, semi-parametric models, and fully parametric models.

Section 3, consisting of chapters 11 through 15, discusses discrete variable models, which have numerous applications to transportation analysis. Chapter 11 delves into count data models, covering the Poisson regression model, truncated Poisson regression models, the negative binomial regression model, and random effects count models. This chapter provides numerous examples as used in transportation analysis. Chapter 12 presents a discussion on logistic regression, that is, the use of binary outcome variables as a function of the predictor variables (regressors). This chapter is brief and provides a good introduction for the study of logistic regression. Chapter 13 covers discrete outcome models involving the application of discrete or nominal data; numerous transportation analyses deal with these types of data. More importantly, there is a discussion concerning multinomial logit models (MNL), which have numerous applications in transportation analysis. The final part of this chapter covers the nested logit model (also known as the Generalized Extreme Value Model), which is based on independence of irrelevant alternatives (IIA). Chapter 14 discusses ordered probability models, recognizing that many transportation applications use ordered discrete data. The presentation of standard or multinomial logit models in the preceding chapter does not account for the ordinal nature of the data which leads to a loss of information. The final chapter of this section does, though, deal with discrete/continuous models. This chapter is brief and presents instrumental variable methods, selectivity bias correction, and applications of discrete/ continuous models.

Section 4 of the book presents other statistical methods that can be applied to transportation data. Chapter 16 delves into random parameter models, which, unlike models taken up in the previous chapters, assume the parameters in regression analysis are not fixed. The fixed parameter assumption may not be correct in some transportation applications. In the final chapter of this section, the authors present Bayesian methods, which apply Bayes' Theorem to classical statistical models. Any statistical model that can be estimated using a classical approach can also be estimated using a Bayesian approach. More importantly, subjective prior probabilities play a role in the estimation of classical statistical models. This chapter provides a foundation for the application of these methods as used in transportation analysis. The method of Markov Chain Monte Carlo (MCMC), widely used in transportation research, is covered exclusively in this chapter.

The final sections of the book include various appendices that provide a review of matrix algebra, fundamentals of statistics, a glossary of statistical concepts, and variables transformations. The reader will be able to refer to these appendices on an as-needed basis for a refresher on these topics.

All in all, this book provides a good repertoire of the methods that can be used in transportation research. Also, this book includes numerous references for readers seeking additional technical details and applications of these statistical methods. Because of the applied focus of this book, it will serve as a valuable reference for transportation practitioners, policymakers, and researchers. Given the presentation of the statistical methods in this book, it would make a good textbook for research methods in the transportation discipline. In the end, this is a solid reference for those engaged in transportation modeling work used in transportation policymaking.

Brian W. Sloboda is currently a pricing economist with the U.S. Postal Service. Previously he was an economist with Bureau of Transportation Statistics in the U.S. Department of Transportation and with the Bureau of Economic Analysis in the U.S. Department of Commerce. He has served as president of the Society of Government Economists and on the board of the Federal Forecasters Consortium. He has been an adjunct faculty member at the University of Phoenix and the University of Maryland University College. He has published articles in the Journal of Economics, Advances in Econometrics, Tourism Economics, Journal of Transportation and Statistics, Pennsylvania Economic Review, International Journal of Transport Economics *and the* Virginia Economic Journal*. He holds a Ph.D. degree from Southern Illinois University at Carbondale.*