2014

Review of Regression; Models, Methods and Applications

Jeffrey E. Jarrett

Book Review for Technometrics

Regression; Models, Methods and Applications by Ludwig Fahrmeir, Thomas Kneib, Stefan Lang and Brian Marx, Springer, 2013 (Hardback 3-642-34332-2 and 978-3-642-34333-9 (eBook)

Review by Jeffrey E Jarrett, of University of Rhode Island, Management Science and Finance

Books on *Regression* are not new and its use in applied fields is extremely important and common. Their importance in the applied fields is beyond comparison and is often taught by professors in applied fields who have a good deal of mathematical education. Often this is not the case, therefore, the appeal of this text is diminished. The book is written at an intermediate level of mathematics, that is, calculus, linear algebra and perhaps the mathematics of probability and statistics. The authors' aim to introduce the reader to parametric, nonparametric and semi-parametric regression by example but also introduces enough theory for an applied data analyst to interpret and base a decision on regression based methods.

The manuscript contains ten chapters of material on regression with two chapters in the appendix on Matrix Algebra and the calculus of probability and statistical inference.

Chapter 1 contains the initial steps in regression involving Univariate distributions and the graphical analysis of association. This include some plots (such as the Boxplot) which may or may not be familiar to students at the intermediate level unless they had a very thorough applied statistics class as a prerequisite.

Chapter 2 follows with introductory treatments of simple linear and multilinear regression models. An unusual but thorough introduction to other models follow. This is often unusual in a text of this nature at an intermediate level, however, this is the plan followed by the authors. Chapter 3 is really the start of the book for applied data analysts where the authors thoroughly introduce the classical linear model and its methods of estimation. The use of one chapter length example, i.e. the Munich Rent Index, is gratifying because the reader does not have to be

confused with non-statistical dimensions of many applications but may focus on only one application. I like the use of the Boxplot on page 153 to explain points of argument concerning the analysis. Too often, textbooks do not examine the residual analysis in a manner that is full and complete. Proof are given at the end of the chapter and can be easily ignored by most readers who are not inclined to have mathematical proof to establish the legitimacy of the methods.

Chapter 4 extends regression to consider the General Linear Model. The method of weighted least squares (WLS) to bypass some problems with ordinary least squares (OLS). Two stage regression follows with a discussion of determining heteroskedasticity (non-constant error variance) and the autocorrelation of errors. Bridge, Bayesian and other models are presented later followed by the mathematical proofs which may likely be avoided by many readers.

Chapter 5 entitled Generalized Linear Models considers cases whereby linear models may not reflect the necessary modeling required to analyze the data correctly. These models call probit and logit models provide methods by which one can properly models cases where linear models will fail to produce the most useful results. The chapter examples are well done and do appear to be valuable whereas some texts provide silly and not useful examples of these analyses. Also, I very much like the use of p-values and confidence limits in the tables of results. In simpler times, one would only see asterisks placed near results to indicate the so-called "level of significance."

Categorical variables are the substance of Chapter 6 and I am refreshed to see that the term categorical is employed rater then "dummy." Again the examples are very good as well as the tables of results. Chapter 7 involves mixed models which bring on illustrations that most applied researchers would likely not consider. However, advance researchers would love the opportunity to work with such models to improve their understanding of the nature of model building when simpler models do not fit the data.

Chapter 8, 9 and 10 discuss some additional topics in regression; nonparametric regression, structured additive regression and quantile regression. These chapters are handled in the same manner as earlier chapters. They are specialized topics that some but not all professors who

have taught regression over a lengthy period of time thank about. They do provide some insight into doctoral dissertations in applied fields. The data must be of a type that cannot be correctly, sufficiently nor easily accomplished by the more usual methods described in earlier chapters. Last, the two chapters in the appendix are useful for those who wish to study the mathematics underlying many of the models discussed in the chapters.

I do not see this text used in many applied fields since it does not fit the background of so many graduate students. Further, I would like better reproductions of the graphics in the text. The graphs are small in one color and often difficult to interpret by someone who has not been educated in graphical analysis. The authors who support the use of computing software should look at Minitab to see how graphics are presented to students. They ought also to use Minitab as a guide to presenting the computer results and hence the understanding of results of the computer analysis of regression. They need to be more student oriented in a revision because the market for regression analysis is very wide and most students would have difficulty understanding a great deal of this text. I love the examples. They are useful, real and wide in scope. Certainly this book and identifies some very useful applications of regression. A perfect book it is not.

In summary, the book "Regression" is a wonderful text with very good applications for those who wish to delve deep into the mathematical foundations of Regression, understand its useful applications and delve into its many additional methods not heretofore discussed often in one text. The book does not have the dimension of the use of color, graphics and computer output now seen in the modern books on regression applications. I believe the authors tried to put into much into the book that would appeal to all readers and in turn some readers would be put off. Last, I would have like to see output and instructions in the use of standard software, i.e., Minitab, SAS and perhaps SPSS.