I Don't Like My Data

Note from the Guest Editor about the Special Issue of *MLRV*

T. Mark Beasley, Guest Editor
St. John's University

This Special Issue of *MLRV* was conceived during the 1998 AERA meeting in San Diego, CA. Isadore Newman and Keith McNeil approached me about guest editing this issue after I had served as the Discussant for a paper session sponsored by the MLR: GLM SIG. The original idea was to include the four papers from that session with my discussion notes as editorial commentary. After accepting this challenge I contacted each of those authors and invited two other papers that I felt would help complete two coherent themes: (1) Analysis of Missing Values and (2) Alternative Regression Procedures. I think you will find the articles enlightening at both the applied and theoretical levels. I can only hope that my comments are equally insightful.

My first general point is that it seems that all data sets have problems, hence the title. "What can be done when these problems arise?" is the central theme to all the articles in this special issue. One problem in particular is that researchers often encounter missing data. In my discussion with many data analysts, the norm seems to be discarding the missing cases. This of course is a loss of information which may bias the results. Another approach involves estimating what the missing value would have been if the subject had actually responded. Of course, the estimation of this replacement value can be biased by many factors. How much bias is created by these two general approaches (i.e., discarding data and imputing missing values) is the underlying theme for the first three articles (Orsak et al.; Mundfrom & Whitcomb; Brockmeier et al.).

The other general problem is that many data sets do not seem to conform to the assumptions of Ordinary Least Squares Regression. Alternative approaches include: (a) transforming the data in some manner or (b) computing parameter estimates in an entirely different manner (i.e., Long; Nevitt & Tam). Some robust methods such as "Trimming" suggest discarding (or downweighting) outliers that may result from a nonnormal error distribution (Nevitt & Tam). It is ironic that purposely deleting values is suggested when assumptions are not met while other researchers are trying to find a way to replace data that is missing.

In this vein of alternative analytic strategies, Kromrey and Hogarty investigate different statistical tests for analyzing the same data without transformation. Thus, even a simple research situation can be approached from several perspectives. The major issue is that different approaches tend to give different interpretations and possibly that is why they remain "alternatives." This is not to say that alternative methods are somehow inferior, but as researchers we have a tendency to rely on more established methods with which we are familiar.

Concerning the reliance on familiar methods, I feel it necessary to comment on the "controversy" surrounding statistical significance testing. I agree with Joel Levin (e.g., 1993) in that until a better alternative to significance testing is developed researchers should continue its use. In the interpretation of results, however, researchers should also understand and state explicitly the precise meaning and limitations of significance testing. To revamp statistical significance testing, researchers and statisticians alike might benefit from using a confidence interval approach. Moreover, the social science research community should consider the perspective of Ron Serlin (1993) and employ a "range null hypothesis." Subscribing to the confidence interval approach has particular implications for investigations that compare methodologies and simulation studies. One issue is that methodological researchers should consider is the accuracy of parameter estimates rather than simply investigate Type I error rate and subsequently power. This concern for accuracy is evident in several of the studies in this issue. Yet, one should not overlook Cliff's (1993) perspective that social science data is typically ordinal in nature and that exact parameter estimates may not be extremely meaningful. One reason that statistical significance testing has been so prevalent in the social sciences has been the scales of measurement issue. To elaborate, when constructing a confidence interval for a parameter estimate for variable measured on an arbitrary scale sometimes the only meaningful value covered (or not covered) by the interval is the null value.

In terms of Monte Carlo studies, statistical hypothesis testing, and therefore investigating whether Type I error rates remain near an expected nominal alpha level, has been the bread-and-butter of simulation researchers. Furthermore, given that statistical hypothesis testing is not going away any time soon, coverage probabilities for confidence intervals should be reported. To elaborate, if a 95% confidence interval is constructed in multiple replications, the confidence interval should cover the population parameter 95% of the time regardless of its value (i.e., whether it is a null or non-null structure). By taking this approach, one can examine

the potential biases in: (a) coverage probabilities (i.e., Does the confidence interval cover the population parameter at the specified level?); (b) power (i.e., How often does the confidence interval cover 0 with a non-null structure?); and (c) Type I error rate (i.e., How often does the confidence interval cover 0 with a null structure?).

In summary, as the popular adage goes, "Necessity is the mother of invention." Two notable trends have led to inventions that have increased statistical sophistication among social science researchers but have also resulted in more problematic data sets for most research projects. First of all, research problems, policy analyses, and educational evaluations have increasingly employed a quantitative perspective. This has resulted in more quantitative analyses of "real-life" data. Anyone who has collected their own data in an experiment, but especially those who have collected their own survey data, and those who have analyzed a national data base (e.g., NAEP, NELS) knows that real data have real problems. Secondly, technology has allowed researchers to handle these real data problems but also to view research issues in a more complex manner and subsequently to employ more complex and sophisticated methods.

From my experience in the graduate education of statistics and data analysis, the analogy, "You don't have to be a mechanic to drive a car," has been used to sickening extent. To expand this analogy, researchers are "driving" some very technologically sophisticated machines these days. What happens when there are problems? Today's automobiles are becoming so sophisticated that the "average driver" cannot work on them. (By the way that is not just a coincidence of technology, it is purposeful goal of car manufacturers). Likewise with sophisticated statistical software, the problem is that they will run any data you put in to them and do it very quickly. You will get results; they may just be meaningless. Furthermore, the speed of statistical software has perpetuated a certain level of sloppiness in dealing with quantitative analysis. So the purpose of this special issue is to "look under the hood" of these machines and see what happens if we throw a wrench into it. Sometimes we find that this new machine (i.e., methodology) is just a "souped-up" version of an older model and that it has the same basic problems. Occasionally, we will find that these new machines are true innovations and that either: (a) they have superior performance or (b) they operate in an entirely different manner. Most of all what should be taken from these articles is, "How do these approaches and techniques integrate with what is already known about statistics and data analysis?"

References

Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*, *114*, 494-509.

Levin, J. R. (1993). Statistical significance testing from three perspective. *Journal of Experimental Education*, *61*, 378-382.

Serlin, R. C. (1993). Confidence intervals and the scientific method: A case for Holm on the range. *Journal of Experimental Education*, *61*, 350-360.

MLRV is the oldest AERA SIG sponsored journal. It has been both a challenge and a pleasure, but I could not have done this alone. First of all, I want to thank former *MLRV* editor Isadore Newman (University of Akron) and current *MLRV* co-editor Keith McNeil (New Mexico State University) for allowing me this opportunity. I also thank the authors for their quick response time in getting their manuscripts to me and their availability when last minute changes were necessary. I also sincerely appreciate the efforts of the most recent former editor (and the person who taught me more about regression than my students care to know), John Pohlmann (Southern Illinois University-Carbondale). His insight in constructing, formatting, and completing this issue was invaluable. I also want to thank the other current co-editor Randall Schumacker (University of North Texas) for his continued support in the area of research in statistics. Special thanks go to former Executive Secretary, Steven Spaner (University of Missouri-St. Louis), and current Executive Secretary, Susan Tracz (California State University-Fresno), for their assistance and guidance in the reproduction and distribution of this Special Issue.

I would like to remind you that I will be the MLR: GLM SIG Program Chair for the 2000 AERA meeting in New Orleans. **START GETTING YOUR PROPOSAL IDEAS TOGETHER NOW. RECRUIT YOUR COLLEAGUES AND STUDENTS**. There is an application form in the back of this issue. I want a BIG SIG in the BIG EASY. I look forward to seeing you in Montreal.

Happy Reading and Regressing,

T. Mark Beasley
St. John's University, New York