THE STATA JOURNAL

Department of Statistics
Texas A & M University
College Station, Texas 77843
979-845-3142; FAX 979-845-3144
jnewton@stata-journal.com

Department of Geography
University of Durham
South Road
Durham City DH1 3LE UK
n.j.cox@stata-journal.com

Review of Statistical Evaluation of Measurement Errors by Dunn

Nicholas J. Cox
Department of Geography, University of Durham, UK
n.j.cox@durham.ac.uk

Abstract. The new edition of the book by Dunn (2004) is reviewed.

Keywords: gn0015, measurement errors, linear models, mixed models, gllamm

1 Introduction

Almost all data are contaminated with measurement error. Even counting with small integers can be problematic. One legendary professor replied "about 5" when filling out the "Number of children" box on a questionnaire. That may be extreme, but consider asking a group of people how many cups of coffee they had yesterday. There will be some firm zeros from people who know they drank none yesterday, not least because they never drink it. If you like coffee, however, would your answer resemble "about 5", given some fuzziness of memory or definition (what precisely is a cup anyway? what about top-ups?). National censuses experience such problems multiplied many-fold, even on basics such as population present in specified places. Beyond counted variables, everyone remembers those school science experiments in which thermometers in boiling water did not all read $100°C$ or a class could not agree on the distance between two fixed points. Indeed, the hallmark of any exact science, as an old tag goes, is being ever open and self-critical about the inexactness of its data.

If science is well aware of measurement errors, statistical science is in two minds about them, often pretending that they do not exist or piously hoping that they are small relative to "natural" or "true" variability. At the same time, there are substantial literatures in various fields offering methods and models for evaluating measurement errors and their consequences. Graham Dunn's new book is a second edition of a text published in 1989 that gives a compact survey of this subject. Title and subtitle are reversed, itself a signal of a small shift in emphasis. The major chapters move from considering sources of variation (chapter 2) through method comparison given paired observations and given more informative designs (chapters 3 and 4, respectively) to methods for categorical (specifically binary) data (chapter 5).

2 Emphases and strengths

Although there is no clear indication of intended readership, Dunn's text presupposes a good general background in statistical theory and statistical methods, especially in linear models. Stata has been used throughout for routine data analyses and production of

graphs (page 191). For more specialized tasks, the reader is pointed to the Stata program `gllamm` (Rabe-Hesketh, Skrondal, and Pickles 2002), which shares honors with EQS and Mplus. Various appendices give examples of code. At the same time, this is not a software manual, and it is not tied strongly to any particular programs.

The strengths of this book stem from its breadth of coverage. The range is from simple graphics and numerics for comparing two measurement methods (although these are not discussed until well into chapter 3) to a variety of models used to represent the measurement process, which in several disciplines are the subject of lively current research. The treatment focuses mostly on measured variables, although with a major nod towards categorical data in the last chapter. There is a large and up-to-date bibliography of about 500 references, which itself is a major resource, especially for those new to the field. It appears strong in biomedical, behavioral, and social science literature, with comparatively little from some other natural sciences.

This is a text that goes beyond the technicalities to echo doubts and air disagreements openly, giving it an appealing frankness and freshness. The author's position is unequivocally in favor of modeling, and so is in sympathy with the general orientation of many researchers using Stata. Yet there are signs that in some problems the modeling is running free ahead of the pertinent science, or perhaps the pertinent scientists; for example, see comments on latent class models on page 177. Dunn's line is that "the process of trying and rejecting models will lead us to think more deeply about what might be going on (or not) during the measurement process" (page 51). Moreover, a key advantage is that "all of your assumptions have to be made explicit and open to scrutiny" (page 79). The defense is familiar and fair, but it remains true that modeling should not be just about traveling enjoyably but about getting somewhere.

Another example is Dunn's treatment of the work of Bland and Altman in medical statistics. They have repeatedly advocated (e.g., Altman and Bland 1983; Bland and Altman 1986; 1995a; 1999) that comparisons of methods start with (and often stop with) calculation of differences and plots of differences versus means. (Incidentally, Bland and Altman 1995b is a key contribution omitted from Dunn's bibliography.) Bland and Altman pull no punches in their papers, and Dunn cannot resist a few aimed in their direction, even before he has explained most of their arguments: "Although the simple methods that have been advocated by Bland and Altman ... are a big improvement over the naive use of, say, correlation coefficients or simple linear regression, they are not always appropriate or particularly informative and, given data from a sufficiently sophisticated design, are totally inadequate for the exploration of differences in the scaling and precision of the competing measurement methods" (pages 48–49).

3 Limitations

Dunn's text is concisely and well written with welcome lighter touches. It was not so well proofread. The most outrageous typo I spotted is that $\tanh^{-1}$ is rendered $\tan h^{-1}$ on page 62. As the text clearly indicates an inverse hyperbolic tangent transforma-

tion, most readers will not be misled. There are many worked examples based on real datasets, which are mostly given in full in text tables, but surprisingly often there is no specification of measurement units. Although units are not usually an issue in the examples, I expected a book on measurement to be more fastidious in this respect.

Authors who work hard to produce monographs of reasonable size are plagued by reviewers who insist on specifying important topics barely discussed or omitted completely. My own list is fourfold.

First, there is no discussion of resolution or rounding problems, even though they differ from issues of accuracy and precision (validity and reliability, if you will).

Second, despite the variety of error models on show, the assumption of normality (Gaussianity) of error terms is pervasive and largely unquestioned. My guess is that, although there is much literature querying the normality assumption, mostly by entertaining the possibilities of fatter tails, it does not square well with most of the modeling being discussed here.

Third, I wanted more advice on graphics, in particular, and exploration, in general, especially for multiple methods datasets. Scatterplot matrices are often used and, indeed, are a good starting point, but it follows from Bland and Altman-type arguments that they are not ideal for discerning fine structure; so what is recommended for that?

Finally, categorical data come off second best compared with measured data in the strict sense: ordinal data especially are pretty much ignored. A good defense would be that analyzing agreement of categorical data (e.g., kappa and its relatives) has been reviewed many times over in both general texts and other monographs. For the third edition, which will be needed long before 2019, my suggestion is therefore double or quits—expand this topic, or more likely cut it out altogether and focus on the continuous case.

4   Conclusion

This is a very useful book on a topic that is key but difficult and often neglected (and one in which Stata software has a vital role). It is concise and informative, and I expect to be pulling it off my shelves again and again.

5   References

Altman, D. G. and J. M. Bland. 1983. Measurement in medicine: the analysis of method comparison studies. *The Statistician* 32: 307–317.

Bland, J. M. and D. G. Altman. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* I: 307–310.

—. 1995a. Comparing two methods of clinical measurement: a personal history. *International Journal of Epidemiology* 24: S7–S14.

—. 1995b. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet* 346: 1085–1087.

—. 1999. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research* 8: 135–160.

Dunn, G. 1989. *Design and Analysis of Reliability Studies: The Statistical Evaluation of Measurement Errors*. London: Arnold.

—. 2004. *Statistical Evaluation of Measurement Errors: Design and Analysis of Reliability Studies*. London: Arnold.

Rabe-Hesketh, S., A. Skrondal, and A. Pickles. 2002. Reliable estimation of generalized linear mixed models using adaptive quadrature. *Stata Journal* 2(1): 1–21.

About the Author

Nicholas Cox is a statistically minded geographer at the University of Durham. He contributes talks, postings, FAQs, and programs to the Stata user community. He has also co-authored fifteen commands in official Stata. He was an author of several inserts in the *Stata Technical Bulletin* and is Executive Editor of the *Stata Journal*.