

Comments on Analyzing Survey Error with Latent Class Models by Paul Biemer
By Michael D. Larsen
March 18, 2005, NISS Workshop on Total Survey Error

Acknowledgement

I would like to thank Alan Karr and NISS for sponsoring this workshop on total survey error and Paul Biemer and Jerry Reiter for organizing the sessions. Thanks to them also for the opportunity to discuss the presentation by Paul Biemer.

Comments on the presentation

Dr. Paul Biemer of RTI International and the University of North Carolina presented a very interesting and promising method for addressing measurement error in sample surveys. Latent class analysis can be used to look for groups in a set of cases or individuals that are not directly observable. The characteristics of the groups and the interpretation of what they represent depend on the subjects associated with them and the values of parameters describing them. In the primary example of his presentation, Biemer uses latent class analysis to find respondents who provide inconsistent information related to marijuana use. In particular, some respondents report no recent use of marijuana. They also report, however, behaviors and experiences that likely are associated with ongoing marijuana use. Consequently, one of the latent classes found by the method is a class of probable users. The parameters in this class indicate high probabilities of the behaviors and experiences that are predictive of use. Respondents with high probabilities of being associated with this class are the individuals who report some or many of the predictive behaviors.

Basic latent class modeling can be used with multivariate categorical data and does not require the pre-specification of descriptions of classes or probabilities within the classes. Biemer's models are appropriate for more complex data that include responses over several time periods and covariate information. Despite the added complexity of the models, the principle of looking for strong patterns of responses among the respondents remains the same. Given the plausible interpretations of the latent classes in terms of inaccurate reporting of marijuana use, the latent class models prove useful for studying this source of measurement error. As a result, one suggestion from Biemer in his presentation is to word questions differently to reduce the error. This suggestion seems well justified in light of the data analysis. Before the analysis and fitting of the model it might have been more challenging to justify making the effort to alter questions in the longitudinal survey.

I would like to make three comments about possible topics for future research.

1. Gold standard

It would be interesting to compare findings from latent class models to responses to "gold standard" questions. In a survey with questions on a sensitive topic, one wants to know the truth. In Biemer's application one wants to know whether the respondent used marijuana in the last twelve months. If it were possible to test hair and perform other medical tests and if those tests were highly accurate, then it would be possible to compare latent classes formed using less expensive and less sensitive questions to the lab results. Of course in many situations there is no accurate measurement to use as a "gold standard" or the best

measurements are too intrusive or expensive to obtain. As a result in most studies latent class results will have to be judged based on their reasonableness and consistency with other sources. The topic of imprecise diagnoses is not unique to the survey world. A search on ISI Web of Science using the key words “gold standard and latent class” yielded 42 research articles from a variety of fields. Wes Johnson of UC Davis and co-authors also have written a series of papers concerning imperfect diagnoses in wildlife applications. A suggestion for survey research then is to try to collect gold standard data or simply better data on a subset of respondents when possible for comparison to latent class model results. In some studies administrative data might be available for this purpose. In others, further data collection and verification of answers on a sample of respondents might be attempted. In any case, one could contribute to this effort by looking for opportunities to collect the data needed for evaluation purposes.

2. Comparisons to other measurement error methods

Latent class analysis seems very well suited to the application presented in Biemer’s talk. Researchers studying tobacco, alcohol, and drug consumption and treatment have developed several other methods for collecting sensitive information. Others studying illegal immigration in the U.S. have developed methods as well. Techniques include randomized response methods, self-administered questionnaires, and methods involving multiple cards (see, for example, <http://www.unece.org/stats/documents/2003/04/migration/wp.3.e.pdf>). Research for the future could compare and contrast alternative methods, including asking less sensitive questions for use in latent class models, for gathering data on sensitive subjects.

3. Methodological developments for latent class modeling

Research to expand the range of latent class models is expected to contribute to their usefulness for study of measurement error in sample surveys. Biemer mentions several areas of possible methodological development in his comments on issues for future research. Methods for estimating interactions between response variables conditional on class membership and assessing model fit could be useful in studies in which the relationships among questions are fairly well understood. Methods for estimating models when some response variables are missing observations would be of great utility in many surveys. Methods for incorporating prior information into the fit of latent class models also could prove effective in assessing the seriousness of measurement error problems. One such approach is Bayesian statistical modeling with prior distributions on probabilities within latent classes of responses to questions that reflect substantive knowledge and pilot study data. These and the other areas mentioned by Biemer have the potential to improve and expand the use of latent class models in survey measurement error research.

Contact Information

Michael D. Larsen
220 Snedecor Hall
Department of Statistics
Iowa State University
Ames, Iowa 50011

Phone: (515) 294-5189
Fax: (515) 294-4433
Email: larsen@iastate.edu
Webpage: <http://www.public.iastate.edu/~larsen>