Model Assisted Statistics and Applications 3 (2008) 87
IOS Press

Guest Editorial

Statistical inference with missing data

The problem of missing data is common to almost all investigations. Indeed, missing data happen when sampling units do not participate fully in the study because respondents refuse to answer some questions from the questionnaire, the interviewer is not able to contact all investigated units, or collected information is lost, obscure or unreliable through factors outside the immediate control of the investigator. As a result sample data are incomplete for some of the variables included in the study. Most classical methods of data analysis assume that all study characteristic values are obtained, and the use of these methods is compromised when any portion of the data is missing.

Dealing with missing data in surveys is no simple matter. In recent years this problem has become the focus of much research in several areas of application for which data sets with hundreds of thousands of variables are available. A variety of methods have been developed to compensate for missing data in a way that enables the survey's data file to be analyzed as if the data were complete. Incomplete data problems can and should inspire good scientific thinking.

I am happy that our call for papers has found such great response. From the papers submitted, we have selected eight for publication. They cover a broad range of methods, which actually should surprise no one because the missing values topic itself encompasses multifarious problems and approaches. I am inclined to believe that this special issue is a valuable contribution to the knowledge of missing data, and I hope the readers of *Model Assisted Statistics and Applications* will share this opinion.

Two papers in this issue deal with methods for imputing data. Zaizai, Jingyu, Junfeng and Hua propose a ratio imputation method based on scrambled randomized responses. This is a hot topic since the randomized response sampling finds its application in many practical surveys where sensitive questions are asked. Singh, Priyanka and Kozak suggest the imputation method to apply when surveys are conducted on repeated occasions.

Four papers discuss the estimation of parameters in the presence of missing data without employing imputation techniques. Agarwal and Gupta consider methods for estimating the population total, mean and variance when sampling frames are incomplete. Under a general sampling design, Dubey and Uprety propose a new class of estimators of the population mean using available auxiliary information in a new way to improve existing estimation techniques under non-response. Bouza investigates the estimation of the population mean and the difference of population means in the presence of missing observations, considering the basic ranked set sampling and its two variants, extreme ranked set sampling and median ranked set sampling. Gamrot suggests how an estimator of the population coefficient of variation can be adjusted for non-response.

Zhang analyses some models from a censored sample, and a general procedure is described based on estimating equations. Harel's paper considers the problem of measuring the influence of missing values on the estimation of a population quantity of interest. The author introduces a new measure called "outfluence" to identify influential values.

I would like to thank all authors who submitted their manuscripts to this Special Issue, and the external reviewers for their invaluable contributions to the reviewing process. Finally, I would like to thank the Editor-in-Chief, Sarjinder Singh, and the Co-Editor-in-Chief, Marcin Kozak, for giving me the opportunity to organize this special issue.

I wish you all pleasant and fruitful reading of this issue, and hope you will find it stimulating, valuable and helpful in your own research.

M[a] del Mar Rueda
Guest Editor
University of Granada
mrueda@ugr.es

ISSN 1574-1699/08/$17.00