MISSING DATA AND SMALL-AREA ESTIMATION

Longford, N.T.

The analysis of large scale surveys usually present the problem of dealing with incomplete data jointly with making inference in a detailed geographical division of the country. Specific statistical methods are necessary to address each one of these topics, which are extensively covered by the present book.

Three differentiated parts divide the eleven chapters of this book. The first and the second parts are focused on the missing data and the small-area estimation problems, respectively. The third part, a single chapter, addresses the problem of model uncertainty presenting a solution inspired on the basis of small-area estimation.

Chapters 1 through 5 constitute the first part of the book devoted to the missing data problem. The book starts with an introductory chapter of survey sampling terminology. The concepts of type of estimator and efficiency are defined in this first chapter. The problem of incompleteness of data is described in Chapter 2. The definitions of complete, incomplete dataset and data analysis are depicted with a detailed example, which leads to introduce the nature of the nonresponse process in surveys and their mechanisms. To deal with those, the single imputation methods, their related models and the EM algorithm are presented in Chapter 3, whereas the method of multiple imputation is detailed in Chapter 4. This chapter makes emphasis on whether to apply multiple or single imputation in an incomplete dataset. Alternative applications of multiple imputation such as measurement error or data editing are posed in the last sections of this chapter. Finally, Chapter 5 presents four case studies of missing data.

Chapters 6 through 10 address the small area estimation problem developing the key idea of similarity among small-areas. Chapter 6 introduces the concept of similarity, discusses the selection procedure between two estimators, suggests composite estimators and finishes with the concept of spatial similarity. Chapter 7 presents models for small area estimation, describes their computational procedures and discusses the

model selection issues. Chapter 8 develops methods that allow to estimate the quantity of interest when auxiliary information allows for greater or equal precision than without it. Chapter 9 highlights on the non-asymptotic nature of the small-area estimators and notes their small-sample variances and precisions. The small-area estimation section of the book finishes in Chapter 10 with four case studies.

The third section is developed in Chapter 11 and argues against the current practice of adopting a model due to failing to find evidence against it. This chapter presents limitations of model selection and introduces the concept of synthetic estimation. In brief, the selection of model and estimators is replaced by linear combinations or predictors based on alternative models. Applications of synthetic estimation are discussed for the analysis of variance and linear regression. This chapter concludes with other applications of synthetic estimation such as meta-analysis.

Every chapter includes suggested readings and exercises, and the author points out that code for the data analyses is available through request. Although computing preferences of the author are S-plus and R, section 5.5 describes other software available for missing data analysis.

To sum up, I think this is an excellent book and it thoroughly covers methods to deal with incomplete data problems and small-area estimation. It is a useful and suitable book for survey statisticians, as well as for researchers and graduate students interested on sampling designs.

Ramon Clèries Soler[1,2]

---

[1] Servei d'Epidemiologia i Registre del Càncer de l'Institut Català d'Oncologia. Av. Gran Via Km 2,7. HOSPITALET DE LLOBREGAT 08907.

[2] Dept. Economia de l'Empresa. EU Empresarials. Universitat Autònoma de Barcelona Edifici S. C/ dels Emprius, 2. SABADELL 08202.