

Uncertainty Intervals and Sensitivity Analysis for Missing Data

Minna Genbäck

Akademisk avhandling

som med vederbörligt tillstånd av Rektor vid Umeå universitet för
avläggande av filosofie doktorsexamen framläggs till offentligt försvar i
Hörsal E, Humanisthuset, fredagen den 25 november, kl. 10:00.
Avhandlingen kommer att försvaras på engelska.

Fakultetsopponent: Docent, Arvid Sjölander,
Institutionen för medicinsk epidemiologi och biostatistik, Karolinska
Institutet, Stockholm, Sverige.



Department of statistics
Umeå School of Business and Economics
Umeå University
Umeå 2016

Organization
Umeå University
Department of statistics
Umeå School of Business and
Economics

Document type
Doctoral thesis

Date of publication
4 November 2016

Author
Minna Genbäck

Title
Uncertainty Intervals and Sensitivity Analysis for Missing Data

Abstract

In this thesis we develop methods for dealing with missing data in a univariate response variable when estimating regression parameters. Missing outcome data is a problem in a number of applications, one of which is follow-up studies. In follow-up studies data is collected at two (or more) occasions, and it is common that only some of the initial participants return at the second occasion. This is the case in Paper II, where we investigate predictors of decline in self reported health in older populations in Sweden, the Netherlands and Italy. In that study, around 50% of the study participants drop out. It is common that researchers rely on the assumption that the missingness is independent of the outcome given some observed covariates. This assumption is called *data missing at random* (MAR) or *ignorable missingness mechanism*. However, MAR cannot be tested from the data, and if it does not hold, the estimators based on this assumption are biased. In the study of Paper II, we suspect that some of the individuals drop out due to bad health. If this is the case the data is not MAR. One alternative to MAR, which we pursue, is to incorporate the uncertainty due to missing data into interval estimates instead of point estimates and uncertainty intervals instead of confidence intervals. An uncertainty interval is the analog of a confidence interval but wider due to a relaxation of assumptions on the missing data. These intervals can be used to visualize the consequences deviations from MAR have on the conclusions of the study. That is, they can be used to perform a sensitivity analysis of MAR.

The thesis covers different types of linear regression. In Paper I and III we have a continuous outcome, in Paper II a binary outcome, and in Paper IV we allow for mixed effects with a continuous outcome. In Paper III we estimate the effect of a treatment, which can be seen as an example of missing outcome data.

Keywords

missing data, missing not at random, non-ignorable, set identification, uncertainty intervals, sensitivity analysis, self reported health, average causal effect, average causal effect on the treated, mixed-effects models

Language
English

ISBN
978-91-7601-555-1

ISSN
1100-8989

Number of pages
13 + 4 papers