Missing Data, PLS and Bootstrap: A Magical Recipe?

Clara Cordeiro
FCT/UALG, Departamento de Matemática
ccordei@ualg.pt

Alexandra Machás
ESCS/IPL, Departamento de Economia
amachas@netcabo.pt

M. Manuela Neves
ISA/UTL, Departamento de Matemática
manela@isa.utl.pt

February 10, 2006

Abstract

The problem of missing data or incomplete data is frequently found in many data bases. The amounts of missing data create difficulty in statistical analysis because the techniques used are not designed for them. Therefore, missing data reduces statistical power because the statistical methods presume that the data bases has information on all variables.

In order to solve the problem of missing data we use some missing data techniques or data imputation algorithms for reconstructing the incomplete data to a complete data set. These algorithms fill out the missing data values, by examining the range of probable values for each variable and calculates many future values randomly. So, using these methods we end up with a credible data set and the results often produce more accurate estimates.

Ensuring the good quality of data, methods as Structural Equation Models (SEM) and Customer Satisfaction Models (CSM) can be considered a true strategic instrument for the organizations and the base for the definition of action marketing planning.

The main objective of this work is to bring up in discussion a problem that could affect the quality of estimators and the validation of the models- the missing data. This issue will be applied to a CSM using data from a market survey conducted for the mobile telecommunication sector in Portugal. Using these results we intend to obtain estimates for the missing data and also to achieve better results combining this procedure with the PLS, reducing the biases of estimators. An extensive computer work is perform and a large number of estimates are calculated using R Software and their packages. Overall, it was concluded that for a higher non-response rates(50%) bootstrap is the best method to be adopted in case of missing data completely at random.

Author Keywords: PLS; Bootstrap; Missing Data; Missing Data methods;