

SPS Workshop 3rd term 2012-2013

Workshop on Missing Data and Multiple Imputation

Instructor: Raul Gomez

Dates: 17-18 June 2013

Seminar Room 2, Badia Fiesolana

Please register with: Monika.Rzemieniecka@EUI.eu

Abstract: This workshop deals with missing data in quantitative research, especially (but not exclusively) in survey data analysis. Missing data refers to cases for which information is unknown, often because respondents do not know the answer or prefer not to say –or because that particular piece of information is simply not available for some particular cases (schools, regions, countries, etc). The existence of missing data is a serious challenge to statistical analysis, and ignoring them may lead to biased results. Even though strategies to deal with missing data have been greatly improved over the past few years, this knowledge is not always available to social scientists.

The aim of the workshop is threefold. First, it should allow participants to learn the theory behind missing data, learn how to identify the patterns in their own data and discuss different methods to deal with this problem. Second, participants will learn the reasoning behind Full-Information Maximum-Likelihood (FIML) and Multiple Imputation (MI) techniques. Third, the workshop will mainly focus on the theory and practical execution of MI techniques in STATA, which will be explained and conducted using examples. In summary, the workshop should allow participants to think carefully about the missing data problem in their own research and make them able to implement appropriate techniques.

If you want to write a term paper for this workshop, please send a copy by email to the seminar's professor as well as to the organizing secretary. Once the paper is approved, she will update your credit award in your academic records.

The deadline for submission is: 14 June 2013.

Programme *:

17 June: Missing data theory and methods. Introduction to FIML approaches.

10: 00 – 12:00: Introduction to missing data: a) missing data theory; b) analysing missing data patterns using Stata.

12:00 – 13:00: Break

Badia Fiesolana ■ Via dei Roccettini 9 ■ 50014 San Domenico di Fiesole (FI) ■ Italy

Tel. +39 055 4685 036 ■ Fax +39 055 4685 201 ■ www.eui.eu

13:00 – 14: 30: Dealing with missing data: an overview of different methods and their pros and cons.

14:30 – 15:00: Break

15:00 – 16:00: Introduction to Full-Information Maximum-Likelihood and its implementation in Stata.

16:00 – 17:00: Data clinic: bring in your data and your questions.

18 June: Multiple Imputation

10: 00 – 12:00: Multiple imputation: theory and FAQs.

12:00 – 13:00: Break

13:00 – 14: 30: Imputation methods: MNV and Chained Equations. Implementation in Stata.

14:30 – 15:00: Break

15:00 – 16:00: Practical issues and data clinic.

(*NB: This programme is subject to changes based on the needs and requirements of both instructor and attendants)

Essential readings:

Allison, Paul D. 2001. *Missing Data*. Sage University Papers Series on Quantitative Applications in the Social Sciences. Thousand Oaks, CA: Sage.

de Leeuw, E. (2001). Reducing missing data in surveys: an overview of methods. *Quality and Quantity* 35, 147–160.

Graham, J. W., A. E. Olchowski, and T. D. Gilreath (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science* 8, 206–213.

Little, R. and D. Rubin (1987). *Statistical Analysis with Missing Data*. Wiley New York.

Rubin, D. (1987). *Multiple Imputation for Non-response in Surveys*. Wiley.

Schafer, J. (1997). *Analysis of Incomplete Multivariate Data*. Chapman & Hall/CRC.

Schafer, Joseph L., and John W. Graham. 2002. “Missing Data: Our View of the State of the Art.” *Psychological Methods* 7(2): 147-177.

Stata Corp (2011). *Multiple Imputation Reference Manual*. Stata Press.

von Hippel, P. (2009). How to impute interactions, squares, and other transformed variables. *Sociological Methodology* 39, 265–291.