UNIVERSITÀ DEGLI STUDI DI PARMA

Jointly with

European Commission

Organize the workshop

Robust analysis of multivariate data: from theory to software implementation and distribution to the user community

April 29th 2014, University of Parma, Department of Economics.

Room C, 1st floor of the Dept. of Economics, Via Kennedy, 6, 43125, Parma

Organized jointly by the Division of Statistics of the Department of Economics of the University of Parma and the Institute for the Protection and Security of the Citizen of the Joint Research Centre of the European Commission

Sommario: L'obiettivo del workshop è quello di rivisitare i moderni metodi statistici per l'analisi robusta dei dati di regressione o multivariati. Particolare attenzione sarà dedicata all'illustrazione del toolbox FSDA sviluppato congiuntamente dall'Area di Statistica e Informatica del Dipartimento di Economia dell'Università degli Studi di Parma e dall'Institute for the Protection and Security of the Citizen of the Joint Research Centre of the European Commission. Nella prima parte del workshop gli interventi saranno focalizzati sull'analisi di dati reali multivariati affetti da diversi fonti di errore. Nella seconda parte del workshop le relazioni si concentreranno sulla creazione e distribuzione di software professionale. In particolare verrà illustrato il software SVN che consente di gestire il versioning dei file quando più utenti accedono agli stessi file in maniera anche contemporanea. Infine verrà illustrata e discussa la creazione di file eseguibili e la gestione del calcolo parallelo su più processori di diversi

computer collegati in rete. Le relazioni saranno in Italiano se l'audience sarà composta solo di Italiani, in caso contrario si utilizzerà la lingua inglese.

Partecipazione gratuita. Per una corretta gestione dei partecipanti è gradita l'invio di una mail all'indirizzo toolboxfs@unipr.it

Abstract: The purpose of the workshop is to review modern methods of robust analysis of regression and multivariate data. Particular attention will be given to illustration of the FSDA toolbox for Matlab developed jointly by the Division of Statistics of the Department of Economics of the University of Parma and by the Institute for the Protection and Security of the Citizen of the Joint Research Centre of the European Commission. Talks in the first part of the workshop will be focused on the analysis of real datasets affected by different sources of heterogeneity. In the second part of the workshop the talks will be more focused on the creation and distribution of professional software.

Participation is free of charge. Persons willing to attend are kindly requested to send an email to toolboxfs@unipr.it.

TIMETABLE

9.15.9.30 Welcome by Luca Di Nella (Head of the Department of Economics).

First session: Robust Analysis of Regression and Multivariate Data

9.30-9.55: Andrea Cerioli, Univ. of Parma "Outlier detection in regression and multivariate data"

The talk will give an overview of robust statistical techniques in regression and multivariate contexts: LTS and LMS (Rousseeuw 1984), MCD (Rousseeuw 1985), Forward Search (Atkinson and Riani, 2000, Atkinson, Riani and Cerioli 2004, Riani Atkinson and Cerioli 2009, Cerioli 2010, Cerioli, Farcomeni, Riani, 2014).

9.55-10.00 Floor discussion

10.00-10.25: Marco Riani, Univ. of Parma "Outlier detection in practice: the FSDA toolbox"

The FSDA (Flexible Statistics for Data Analysis) Matlab toolbox (Perrotta, Riani and Torti, 2009; Riani et al. 2012, Riani, Atkinson and Perrotta 2014) will be presented. Concrete examples of its use will be shown on real datasets.

10.25-10.30 Floor discussion

10.30-10.55: Domenico Perrotta and Francesca Torti, JRC, "Application to large collections of datasets"

The talk describes an extensive application of the FSDA outlier detection functions to large trade data collections (Cerioli and Perrotta, 2014).

10.55-11.00 *Floor discussion*

11.00-11.30 *Coffee Break*

Second session: Software implementation and distribution

11.30 11.55: Alessandra Reale, Gianpiero Bianchi, ISTAT. "Robust methods for editing and imputation of Italian Agriculture Census data"

ISTAT has used a robust procedure based on the forward search to detect cases in which the information collected by the census was not in agreement with that coming from the General Agency for Agricultural Subsidies (AGEA). This talk summarizes how ISTAT tackled the problems of data correction and control.

11.55 12.00: *Floor discussion*

12.00-12.25: Emmanuele Sordini JRC, "Collaborative software development for geographically distributed teams"

The talk will describe the use of the version management tool (Apache Subversion, or SVN for short) and other built automatic tools to support collaborative software development for geographically distributed teams.

12.25-12.30 *Floor discussion*

12.30-12.55: Ivano Azzini JRC, "Running MATLAB applications on a cluster of computers"

The talk will focus on how stand-alone applications can be built from Matlab files. Moreover the use of mex files and a free parallel toolbox within Matlab code will be discussed.

12.55-13.00 Floor discussion

13.00-13:15: General discussion

14.30-16.30: Discussion on the future of FSDA (restricted to FSDA developers)