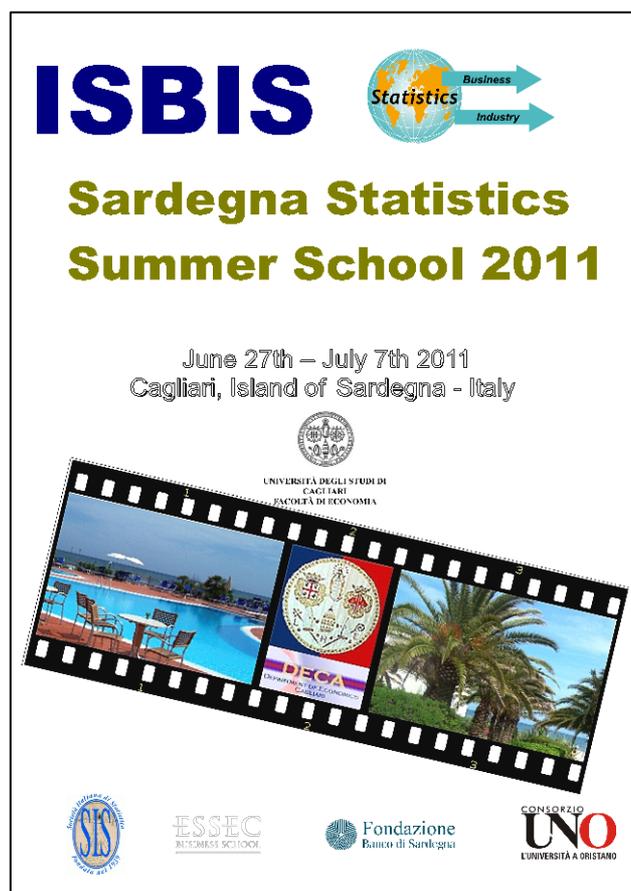


Final Report on the ISBIS Sardegna Statistics Summer School 2011



Statistical Methods and Software for Large and Complex Data Sets and Business and Industrial Statistics

*Organised by
Department of Economics at the University of Cagliari
ESSEC Business School of Paris
ISBIS - International Society for Business and Industrial Statistics*

**Cagliari, Island of Sardegna – Italy
June 27th – July 7th 2011**

School Director:

- Nicholas Fisher (ISBIS Past-President - ValueMetrics Australia)

Chief Organisers:

- Francesco Mola (University of Cagliari, Italy)
- Vincenzo Esposito Vinzi (ESSEC Business School of Paris, France)

Sponsored by:

- Fondazione Banco di Sardegna
- University of Cagliari
- ESSEC Business School of Paris
- Consorzio UNO – University of Oristano

Under the auspices of:

- Italian Statistical Society (SIS)

Lecturers:

- Jaromir Antoch (Charles University, Czech Republic)
- William S. Cleveland (Purdue University, USA)
- Vincenzo Esposito Vinzi (ESSEC Business School of Paris, France)
- Nicholas Fisher (ValueMetrics Australia)
- Jerome H. Friedman (Stanford University, USA)
- Ray Kordupleski (Customer Value Management, Inc., USA)
- Roberta Siciliano (University of Naples “Federico II”, Italy)

Participants: Twenty-five students attended the School, the majority from Italy, but there were also attendees from Denmark, Germany, Ireland, Norway and Turkey.

Scientific Program

The ISBIS Sardegna Statistics Summer School 2011 was devoted to Statistical Methods and Software for Large and Complex Data Sets (lectures in the first week), as well as to Business and Industrial Statistics (lectures in the second week).



The summer school was organized as it follows: during the two weeks, lessons were divided into six classes (four in the first week and two in the second week), one for each topic. All lessons were given in English.

The school was equipped with a Computer Lab, which allowed the

development of practical exercises performed with specialized software such as MATLAB, R and XLSTAT.

Every participant was provided with: a certificate of attendance, a timetable of the school, a list of participants and lectures with their e-mail addresses, a copy of the slides of the lessons on a USB stick, maps and tourist information of Sardinia, tickets for two excursions, a notebook, a personal badge and a welcoming letter.

Two excursions were arranged by



the organisation: the first one took place on Saturday July 2nd and it consisted in a tour of the cities of Cagliari and Pula, a flamingos bird-watching experience and a visit to Nora; the second one took place on Sunday July 3rd and consisted in a visit to Zuddas' Grottos and to the two wonderful beaches of Tuerredda and Chia.

Summary of the lectures

In the first class Prof. Cleveland spoke about LARGE AND COMPLEX DATA



SETS. With the increasing availability of massive amounts of data it is essential to develop new, practical approaches to extracting information from such data sets. Professor Cleveland presented innovative techniques that can simplify the analysis by means of a breakdown of the massive data set and its subsequent reconstruction. The last part of the class was devoted to a computer

application using the R language. Prof. Cleveland demonstrated the tools used to carry out this type of analysis and, in an interactive session, he also showed how to program the needed functions in the R environment. With exemplary professionalism he managed to make the comprehension of an advanced technology accessible to all the participants.

Prof. Antoch, who spoke of ROC CURVES, held the second class. Using a strictly mathematical approach he showed both the classical parametric approach and alternative non-parametric methods for binary segmentation. Thanks to his great passion for the subject he was able to evince interest from the students in even the most theoretical and technical aspects.

In the third class, Professor Siciliano focussed on DATA MINING, specifically the topic of EXPLORATORY TREES. These lectures concentrated on classification and segmentation methods within the statistical learning paradigm. After explaining the basic concepts and the algorithms, she spent considerable time on several aspects related to the application of the methods and interpretation of results. Using MatLab software she showed the enormous practical potential of these methods and was able to establish a good relationship with the students who actively participated in her class.



In the fourth class Professor Friedman spoke about STATISTICAL LEARNING. With a highly mathematical approach he explained the elements that

form the basis of this branch of Statistics. Several models were introduced and discussed: predictive learning, ensemble methods, kernel methods, gradient boosting, fast spread regression and classification among others. His great teaching skills were highly appreciated by the students as he managed to convey to them a strong sense of curiosity about these techniques.



The fifth class, conducted by Prof. Kordupleski and Prof. Fisher, concentrated on **WHAT LEADERS NEED TO RUN THEIR ENTERPRISES**. Gaining data and transforming them into useful information represent the main problems for managers today. Due to the immense amount of data that managers have at their disposal, the real problem is figuring out what is useful and what is not. Thanks to their

wonderful creativity and their vast experience in the field, the two professors were able to equip the students with a few basic principles intended to guide managers in their decisions. They explained the specific methodologies that help to recognize the types of data business men need, the different techniques used to collect these data, and how they can best use them.

In the last class Professor Vinzi exposed the approach of PLS (Partial Least Squares)

PATH MODELING for modelling relationships between latent variables and computing factor scores in presence of a multi-block data structure. This approach can be profitably exploited in many fields, one of which is Customer Satisfaction. During the explanation of the algorithm, its properties and interpretation rules, Professor Vinzi was outstanding in transmitting his passion and enthusiasm to the students who



were fascinated by this approach. Particular emphasis was also placed on practical applications. After explaining the operational functioning of the PLSPM module in the XLSTAT software, students were asked – with practical exercises – to put into practice all they had learned during the class.

This School was undoubtedly a very enriching and stimulating experience for the participants. They had the privilege of listening to internationally renowned lecturers and researchers who put emphasis on theoretical foundations, methodological developments and their application to real-world situation presented with varied teaching styles.

A special thank goes to Prof. Francesco Mola and his team of volunteers from the University of Cagliari who did a great job in taking care of all the details related to the local organization of the School, from the lecture room to the computer lab, from coffee breaks to dinners, from excursions to... summer school games (diving competition, water polo, surfing) that took place in the swimming pool of the Hotel Flamingo (the resort that hosted the School) and really helped to create a pleasant atmosphere as well as to build a



strong social network among the participants.

All the students wish that ISBIS would organize this kind of initiative on a regular basis in the future!

Luca Frigau
School Participant
25 July 2011