Short Courses Programme at EESW19

EUROPEAN NETWORK FOR BETTER ESTABLISHMENT STATISTICS ENBES-EESW EUROPEAN ESTABLISHMENT STATISTICS WORKSHOP

EESW19 is offering three one-day short courses on 24 September 2019, intended for business survey methodologists, researchers and other professionals in the field:

- **Topics in Business Data Collection Methodology: On web questionnaire design and response enhancement** by Ger Snijkers and Gustav Haraldsen
- **Statistical Data Cleaning for Business Statistics with R** by Mark van der Loo
- **Machine Learning** by José L. Cervera-Ferri

For more on these courses, please read below. Fees apply for attending the courses. Click here to register for one of them. The number of places is limited, registration closes on July 31, 2019.

A pdf version of the short course programme is also available.

An ENBES-ECAS short course

Topics in Business Data Collection: On web questionnaire design and response enhancement by Ger Snijkers and Gustav Haraldsen

About the course

Two important elements in any survey are the contact & recruitment strategy and the measurement instrument. In business surveys web is the dominant data collection mode. This course will focus on these elements, taking tailoring considerations into account.

Tailoring is an important cornerstone in business survey design, i.e. adapting the design to how businesses operate in the context of a survey. E.g. in business surveys, respondents do not report about themselves, but act as informants about the enterprise or establishments within the selected enterprises. From this follows that the contact & recruitment strategy needs to address both the company management and the respondent(s) chosen by the management. Furthermore, responding is typically a two-step process where administrative business information needs to be collected and processed before the questionnaire can be completed. It should also be taken into account that business surveys normally are parts of an ongoing relationship with more than one survey to be completed at the same time and the same questionnaire often repeated several times.

The course will discuss what implications these characteristics have and how modern business surveys should address them. The latest insights in designing business web questionnaires and contact & recruitment strategies will be presented. Dynamic web questionnaires will be illustrated by a number of video clips. Recent empirical studies on response enhancement, non-response reduction measures and responsive design will be discussed.

Prerequisites for the course

Participants are expected to have basic knowledge in survey methodology and preferably some experience with or familiarity with business surveys.

Recommended readings to become familiar with business surveys:

- *Chapter 1: Surveys and Business Surveys* in the textbook *Designing and Conducting Business Surveys* (Snijkers, G., G. Haraldsen, J. Jones, and D. Willimack, 2013, Wiley, Hoboken). The textbook serves as a basis for this course.
- Snijkers, G., 2016, Achieving Quality in Organizational Surveys: A Holistic Approach. In Liebig, S., and W. Matiaske (eds.), Methodische Probleme in der empirischen Organisationsforschung, pp. 33-59. Springer, Wiesbaden.

About the instructors

Ger Snijkers (Statistics Netherlands) and Gustav Haraldsen (Statistics Norway) have a longstanding experience in business survey methodology. As senior survey methodologists they have been involved in many business survey design projects in their agencies, and advisors in international expert panels. They have a longstanding experience in lecturing around the world. Together with Jacqui Jones (UK ONS, now ABS, Australia) and Diane Willimack (US Census Bureau) they co-authored the first coherent international textbook on business survey methodology (Wiley, 2013).

An ENBES-ECAS short course

Statistical Data Cleaning for Business Statistics with R by Mark van der Loo

About the course

The course will demonstrate how data quality can be systematically defined and improved using R. It will focus on data validation (data checking), locating errors, and imputing missing or erroneous values under restrictions. Examples will be drawn from the Structural Business Statistics survey where common restrictions include nonnegativity rules and record-wise balance checks. The instructor will present short introduction to the main principles, provide quizzes and discussions for the audience, and give short R-based exercises.

A number of R packages that the instructor (co-)authored will be demonstrated, including 'validate' (for data quality checks), 'errorlocate' (for error localization), 'simputation' (for imputation methods), 'rspa' (for value adjustment), and 'lumberjack' (for keeping track of changes in data). Special attention will be paid on how to combine the various data processing steps, and how to analyze and visualize the results. At the end of the course, participants will have insight into some of the methods common in data editing for business surveys as well as an overview of how to implement that with free and open source R and the mentioned packages.

Prerequisites for the course

As there will be some practical exercises participants are required to bring a laptop with a recent version of R (3.5.x or higher) and RStudio (1.0 or higher).

Participants are expected to have beginners' proficiency in R. Specifically, it is expected that participants can read simple text files (csv) into R and perform a few simple analyses such as summary statistics. It is also assumed that participants are comfortable working with scripts in RStudio.

About the instructor

Mark van der Loo works at the department of methodology at Statistics Netherlands. His main area of expertise is statistical data cleaning (data editing) and statistical computing in general. He has (co)authored several publicly available and widely used R packages. In 2018 he and Edwin de Jonge published the book "Statistical Data Cleaning with Applications in R" (John Wiley & Sons, Inc).

Machine Learning by José L. Cervera-Ferri

About the course

Machine Learning (ML) methods are becoming increasingly relevant for National Statistical Institutes (NSIs) as they are considering use of new data sources known under the name Big Data (satellite images, mobile call records, sensors and IoT, etc). The short course will present, at an introductory level, the most well-known ML methods, including linear and non-linear regression models, trees and random forests. Examples of actual application of ML methods in official statistics will be presented, as well as their potential in specific data treatment processes such as missing value imputation, classification and small area estimation.

Introduction of ML methods implies moving from design-based or model-based statistical inference to algorithm-based inference, with important consequences for the interpretability and transparency of methods for data users. A major barrier to the application of ML in NSIs is the difficulty of combining statistical skills, IT expertise, subject-matter knowledge and communication skills to explain the methods and results to users. During the course, the participants will have the opportunity to share experiences in their institutes about establishing Data Science teams, and identifying ways to attract and retain talent for these activities.

Prerequisites for the course

Experience with model-based inference (such as econometric modelling or mathematical statistics) would be beneficial.

About the instructor

José L. Cervera-Ferri, CEO of DevStat, is an international consultant in official statistics, with more than 25 years of experience in governmental and private institutions. José was recently involved as consultant in the preparation of the conference on Modernization of European Statistics (Valencia 2014) and the ESS Big Data Event (Rome 2014) by supporting Eurostat for the preparation of the scientific programme, contributing as facilitator and rapporteur, and editing the technical reports, which have been published in the ESS research portal (CROS). He also participates in projects related to the modernization of statistics from the legal viewpoint (impact assessment of the European framework legislation on social statistics). He currently coordinates the courses on Machine Learning for Eurostat and the EU National Statistical Institutes under the European Statistical Training Programme (ESTP).

Fees

Early bird fees (by May 31, 2019)	**£**
Regular	90 (approx. 100€)
Reduced (*)	72 (approx. 80€)
Full fees (by July 31, 2019)	
Regular	112 (approx. 125€)
Reduced (*)	90 (approx. 100€)

(*) EFTA citizen, IASS member, RSS member, student

Note

ENBES confirms that all three short courses will be held, as there is sufficient interest for them. However, a course may be not given due to unforeseen circumstances, the so-called *force majeur*. In that case, the responsibility of the organiser is limited to paying back the received registration fees in full.

Sponsored by:

Endorsed by:

To subscribe to ENBES emailing list, please visit www.1ka.si/a/175375.
