



seriss

SYNERGIES FOR EUROPE'S
RESEARCH INFRASTRUCTURES
IN THE SOCIAL SCIENCES

Deliverable Number: 5.7

Deliverable Title: Report on the activities for statistical training under SERISS

Work Package: 5 Training and Dissemination

Deliverable type: Report

Dissemination status: Public

Submitted by: ESS ERIC (UL)

Date Submitted: March, 2019

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 654221.





www.seriss.eu  @SERISS_EU

SERISS (Synergies for Europe's Research Infrastructures in the Social Sciences) aims to exploit synergies, foster collaboration and develop shared standards between Europe's social science infrastructures in order to better equip these infrastructures to play a major role in addressing Europe's grand societal challenges and ensure that European policymaking is built on a solid base of the highest-quality socio-economic evidence.

The four year project (2015-19) is a collaboration between the three leading European Research Infrastructures in the social sciences – the European Social Survey (ESS ERIC), the Survey of Health Ageing and Retirement in Europe (SHARE ERIC) and the Consortium of European Social Science Data Archives (CESSDA AS) – and organisations representing the Generations and Gender Programme (GGP), European Values Study (EVS) and the WageIndicator Survey.

Work focuses on three key areas: Addressing key challenges for cross-national data collection, breaking down barriers between social science infrastructures and embracing the future of the social sciences.

Please cite this deliverable as: ESS ERIC (University of Ljubljana) Report on the activities for statistical training under SERISS Deliverable 5.7 of the SERISS project funded under the *European Union's Horizon 2020 research and innovation programme* GA No: 654221. Available at: www.seriss.eu/resources/deliverables



SERISS

WP5 Training and dissemination

Task 5.4 Statistical training for secondary users

1 Aims of the Task 5.4

The aim of this task is to equip secondary users of already collected survey data researchers with the skills and knowledge they need to analyse those data and improve the rigour of cross-national survey research in the European context and acquire a better understanding of the underlying methodology of comparative survey research. It consists of two parts:

1) Face-to-face training

A series of training courses with groups of about 25 participants per course were held over the period 2015-2019. The seminars were designed to enable intense, workshop-type courses, with participants interacting with trainers, receiving individualised feedback and doing practical exercises in small ad-hoc teams. Applicants were selected according to the published criteria by a panel consisting of senior scientists at UL and GESIS.

2) Online training

Using the same themes as in the workshops, short webinars enabling a wider train the trainers' action will be made between April and June 2019. Webinars will allow for wider audiences and are especially of benefit for those researchers with low travel budget. The webinars will be offered through the ESS website and will be conducted and archived by GESIS by the end of the SERISS project.

2 Report on face-to-face training activities

European infrastructures joined by the SERISS project have always recognized the need to spread excellence through the active dissemination of their results and methods. Under the SERISS project a series of four training courses, with groups of 25 participants per course was completed. The courses were designed to enable intense, workshop-type events, with participants interacting with trainers, receiving individualised feedback and doing practical exercises in small ad-hoc teams. In order to offset the reduced opportunity for direct participation the participants' recruitment process emphasises the 'training the trainers' principle of selection so that in the long run a much higher number of researchers will benefit.

The course instructors were drafted by GESIS while the online calls, communication with applicants and participants, as well as the course organization and reimbursements were handled by the UL

researchers. The participants were selected by a joint selection panel consisting of GESIS and UL researchers.

The number of direct beneficiaries of the four courses was 100, while indirectly (i.e. via personal diffusion the acquired knowledge and the forthcoming webinars) a much larger audience is expected to be reached. As planned, the four training courses followed the thematic format that went from data collection stages to data analysis stages: 1) Designing questionnaires for cross-cultural surveys, 2) Sampling, weighting and estimation for cross-national surveys, 3) Assessing data quality and comparability for cross-national surveys and 4) Analysing cross-national survey data.

2.1 An overview of the course content and experts

Course 1: Designing questionnaires for cross-cultural surveys (Ljubljana, 24 – 25 Oct. 2016)

Instructors: *Ana Villar* (Centre for Comparative Social Surveys, City University London) and *Dorothee Behr* (GESIS – Leibniz Institute for the Social Sciences in Mannheim/ Essex)

Course description: This course introduces students to designing questionnaires for cross-cultural surveys. We will approach this topic from two different but deeply interwoven angles: (1) designing a 'source' questionnaire and (2) producing equivalent target questionnaires based on the source questionnaire.

Day 1 will be dedicated to the latest developments in questionnaire design and pretesting in cross-cultural surveys. We will start reviewing existing models for design of questions in cross-cultural surveys and discussing the role of cross-cultural input in the process of designing, pretesting and evaluating questions. Using examples of actual questions from cross-cultural surveys, we will consider qualitative, quantitative, and mixed-method pretesting techniques available to researchers embarking on question design for cross-cultural surveys. We will review strategies to plan and manage cross-cultural question design efforts.

Day 2 will start with a general overview of requirements of good questionnaire translation. We then will move on to discuss methods employed to produce and assess questionnaire translations. Special emphasis will be placed on the TRAPD model (Harkness, 2003), which includes parallel translation, team-based review and adjudication stages, pretesting and thorough documentation. Practical aspects such as translation planning, translator recruitment and translator briefing will also receive attention.

The course will end with a synthesis of the two strands and its embedding into the overall survey context. The course will be a mixture between lectures, short practical exercises and discussions. Participants are encouraged to bring questions and materials designed for cross-cultural contexts for discussion at the class.

Course 2: Sampling, Weighting and Estimation in Survey Methodology (Ljubljana, 24 – 25 Apr. 2017)

Instructors: *Stefan Zins*, Leibniz Institute for the Social Science (GESIS), Mannheim; *Matthias Sand*, Leibniz Institute for the Social Science (GESIS), Mannheim

Course description: the course will cover three interrelated topics: methods of selecting complex samples, creation of analysis weights that adjust for nonresponse and undercoverage, and the analysis of data collected via complex weighted surveys.

The first day, will start with an introduction the framework for design based inference and some basic sampling designs will be introduced. Common features of sampling designs such as stratification, sampling of clusters and multi-stage sampling will be discussed. For each method, students will learn the relevant formulas for point estimates and variance estimates; however, the course will emphasize application over theoretical proofs of the formulas.

The second day will focus on estimation based on survey samples and inference. Furthermore, students will learn how complex designs and estimators alter the ways in which survey data should be analyzed. Traditional methods of analysis, usually taught in introductory statistics courses, are inapplicable to such data sets. There are several different methods that can be used to analyze complex weighted survey data.

Course 3: Measurement quality and correction for measurement error (Ljubljana, 26 – 27 Feb. 2018)

Instructors: *Diana Zavala-Rojas* and *Wiebke Weber*, Research and Expertise Centre for Survey Methodology (RECSM), Universitat Pompeu Fabra, Barcelona, Spain

Course description: The course discusses the problem of measurement error in survey research and its consequences in the results of the research. We present different methods to estimate and predict measurement error and show how is this information can be used for improvement of survey questions before data collection and for the correction for measurement error in data analysis.

Course 4: Cross-national comparative research: Longitudinal analysis of panel data (Ljubljana, 18 – 19 Feb. 2019)

Instructors: *Dr Gundi Knies* and *Dr Min Zhang*, University of Essex

Course description: This course will introduce participants to some of the many topics, data and analysis tools available for cross-national comparative research on the basis of panel data. Participants will gain hands-on experience working with some of the world's most powerful panel data: the UK Household Longitudinal Study (UKHLS), the German Socio-economic Panel Study (SOEP) and a cross-nationally harmonised data set of panel studies from around the world (CNEF). They will learn about the many steps involved in undertaking this type of analysis, covering the conceptual bases as well as the basic data management and modelling techniques.

The course is modelled on the success of the “Introduction to Understanding Society using Stata”-training course, which has been developed by Dr Knies and colleagues at ISER University of Essex. It is delivered through a series of overview lectures which are followed by hands-on sessions in the lab. Participants will be provided with a number of worked examples that they can work through during the course and use as the building block for their own analyses in the future. They have the opportunity to ask the trainers any questions they may have on the material. Examples will range in complexity from an introduction to the data management steps involved in setting up panel data to do transition analyses, to implementing and choosing between different panel estimators in within country comparisons, to testing hypotheses about effects in two different types of country, to implementing panel models across a larger number of countries.

2.2 Evaluation

All courses were oversubscribed, with a total of 332 applications received for 100 available places (course 1 received 60 applications; course two 63 applications, course 3 102 application and course 4 107 applications). The majority of participants came from a variety of European countries, while some participants came from outside Europe, when remaining within the limits of reimbursement rules. The interest in the courses suggests that the face-to-face training remains a popular form of acquiring expertise from European research infrastructures, as well as for informal networking and knowledge exchange.

Anonymous online evaluation forms consisting of eleven questions related to the content, instructions and organization were completed after each course by the participants. The results are generally very favourable in all three areas, exceeding the average of 3 on a 4 point scale. The topics seemed to have been well chosen as most participants felt that the courses enhanced their knowledge and helped them with their work. Nevertheless, some relatively lower marks indicate that in a the (likely) future generations of face-to-face infrastructures-based training, instructors should be specifically advised to allocate enough time for discussion, engage in intense communication with the participants (Q&A session) and give them encouragement.

	Scale: 1 strongly disagree ... strongly agree 4	1st Course (n=20)	2nd Course (n=22)	3rd Course (n=20)	4th Course (n=21)
q1	I was enhanced by the course	3.50	3.32	3.50	3.15
q2	It was helpful for my work	3.50	3.29	3.45	3.44
q3	It fulfilled my expectations	3.30	2.81	3.30	2.82
q4	The instructor(s) encouraged the students	3.45	2.81	3.85	2.89
q5	The instructor(s) answered the questions	3.45	3.76	3.80	2.85
q6	There was enough time for discussion	3.00	2.76	3.45	2.39
q7	The material was helpful	3.20	3.24	3.65	2.28
q8	It was well organized	3.60	3.52	3.75	3.17
q9	The facility was appropriate	3.42	3.24	3.68	3.56
q10	I felt welcome	3.47	3.57	3.90	3.67
q11	I was satisfied	3.68	3.10	3.80	2.98
	Overall average	3.41	3.22	3.64	3.02

APPENDIX 1: COURSE AGENDAS

COURSE 1: AGENDA

DAY 1	<i>MONDAY, 17th of October</i>
9.00 – 10.30	<i>Two-level regression: the random intercept model</i>
	This introductory session starts with a conceptual discussion on the specific character of nested data, and explains why conventional statistical tools (such as the classical regression model) are not appropriate to analyse this kind of data. The two-level regression model is introduced as an extension of classical regression.
10.30 – 11.00	Coffee Break
11.00 – 12.30	<i>Two-level regression: the random slopes model & cross-level interactions</i>
	The two-level regression model is further extended to the random slopes model and models with cross-level interactions. Real data examples are provided.
12.30 – 13.30	Lunch
13.30. – 15.00	<i>Two-level regression: additional topics</i>
	The following important additional topics are covered: model estimation - model evaluation - model comparison – assumptions of the multilevel model - categorical independent variables - standardization of parameters – centering of the data
15.00 – 15.30	Coffee Break
15.30 – 17.00	<i>Multilevel models for longitudinal data</i>
	In this session, it is illustrated how the two-level regression model can be a very useful tool to analyse longitudinal / repeated measurements data, and introduces the multilevel growth curve model.
19.00 –	Dinner at M Hotel
DAY 2	<i>TUESDAY, 18th of October</i>
9.00 – 10.30	<i>Complex data structures: three-level models & cross-classified models</i>
	Often, researchers are confronted with more complex data structures. The cross-classified multilevel model is introduced as a useful tool for situations where lower-level units are nested within a double hierarchical structure. The application of cross-classified models in order to disentangle Age, Period and Cohort effects is discussed. Furthermore, the multilevel model is extended to three-level situations.
10.30 – 11.00	Coffee Break
11.00 – 12.30	<i>Multilevel Structural Equation Modelling</i>
	Recent development in statistics and software have brought multilevel structural equation models within the reach of substantive researchers. Major advantages of multilevel SEM are the possibility to work with latent variables and to specify multilevel mediation effects. During this session, a brief introduction SEM is given.
12.30 – 13.30	Lunch
13.30. – 15.00	<i>Question and answer session</i>

COURSE 2: AGENDA

DAY 1 *MONDAY, 24th of April*

9.00 – 10.45 Introduction to Design Based Inference The framework for design based inference will introduced and central concepts like sampling variance and unbiased estimators will be explained.
10.45 – 11.15 Coffee Break
11.15 – 13.00 Complex Sampling Designs I Elements of complex sampling design will be discussed, starting with stratification and the allocation problem.
13.00 – 14.00 Lunch at Šumi (Faculty of Social Sciences)
14.00 – 15.45 Complex Sampling Designs II Elements of complex sampling will be continued with cluster and multistage sampling after which unequal probability sampling is covered. The chapter will conclude with the concept of the design effect and its estimation.
15.45 – 16.15 Coffee Break
16.15 – 18.00 Tutorial: Brief Introduction to the Analysis of Sample Surveys with R The first tutorial will provide a short introduction to the software R, that will be used throughout the two tutorials, as well as the relevant packages on sampling and estimation and how to install them. Furthermore, this tutorial will provide an overview on drawing samples using R.
19.30 – Dinner at M Hotel

DAY 2 *TUESDAY, 25th of April*

9.00 – 10.45 Calibration Weights I The concept and purposes of survey weights will be introduced, after which different weighting methods will be covered, starting with poststratification and raking.
10.45 – 11.15 Coffee Break
11.15 – 13.00 Calibration Weights II The chapter on calibrations weights will be continued and the General Regression Estimator will be introduced including the estimation of its sampling variance.
13.00 – 14.00 Lunch at Šumi (Faculty of Social Sciences)
14.00 – 16.00 Tutorial: Estimation with Complex Weights The second tutorial aims to provide an overview on how to incorporate design weights that stem from complex sample surveys into the estimation of inferences using Data from various countries and rounds of the ESS. Mostly, this tutorial will focus on how to use the survey package in R to estimate the correct inference for a particular population based on a given sampling design

COURSE 3: AGENDA

DAY 1 MONDAY, 26. 2. 2018 – Measurement error

9.00 – 10.45 Introduction to measurement quality What is measurement quality/error? Why and how does it matter?
10.45 – 11.15 Coffee Break
11.15 – 13.00 Estimating measurement error Different approaches to estimate measurement error
13.00 – 14.00 Lunch at M Hotel
14.00 – 15.45 Predicting measurement error Prediction of the measurement quality with SQP2.1. Statistical approach of the software, coding scheme of questions' characteristics and how it helps to predict the measurement error of a question
15.45 – 16.15 Coffee Break
16.15 – 18.00 Correction for measurement error in regression models ... with single indicators and sum scores
19.30 – Dinner at M Hotel

DAY 2 TUESDAY, 27. 2. 2018 – Invariance testing

8.30 – 10.15 Correction for measurement error in structural equation models (1) ... accounting for effects other than just direct effects with single indicators
10.15 – 10.45 Coffee Break
10.45 – 12.30 Correction for measurement error in structural equation models (2 - continuation) ... accounting for effects other than just direct effects with single indicators and sum scores
12.30 – 13.30 Lunch at M Hotel
13.30 – 14.45 Correction for measurement error in structural equation models (3) ...with multiple indicators
14.45 – 15.30 Concluding remarks Advantages and disadvantages of the introduced approaches Group discussion

COURSE 4: AGENDA**Day 1: 18.2.2019**

Time	Topic
9:00-9:30	Registration
9:30-10:15	Longitudinal studies for cross-national comparative research
10:15-11:00	Longitudinal analysis: Transition models, Fixed effects and random effects models
11:00-11:15	Short break
11:15-12:30	Hands-on session 1: Setting up the data and analysing transitions
12:30-13:30	Lunchbreak
13:30-15:30	Hands-on session 2: Implementing and interpreting panel models
15:30-15:45	Short break
15:45-17:00	Hands-on session 3: Doing a within-country comparison
19:00	Dinner, M Hotel

Day 2: 19. 2. 2019

Time	Topic
9:00-10:45	Hands-on session 4: Testing cross-national hypotheses
10:45-11:15	Short break
11:15-12:30	Hands-on session 5: Comparing a whole range of countries (1)
12:30-13:30	Lunchbreak
13:30-14:45	Hands-on session 6: Comparing a whole range of countries (2)
14:45-15:00	Close