gesis Leibniz Institute for the Social Sciences

8th GESIS Summer School in Survey Methodology
Cologne, August 2019

Syllabus for Course 10: "Sampling, Weighting and Estimation"

Instructor: Stephanie Eckman, Ph.D.
E-mail: steph.eckman@gmail.com
Homepage: www.stepheckman.com

Date: 19.-23. August 2019
Time: 09:00-13:00, 14:00-16:00
Course starts Monday morning at 09:00

About the Instructor:

Stephanie Eckman, Ph.D. is a Fellow at RTI International in Washington DC where she designs samples and consults on all aspects of survey design. Her research interests include undercoverage in household surveys and motivated underreporting by interviewers and respondents in survey interactions. She has taught specialized courses on survey methodology, weighting and data analysis at the University of Mannheim, Technical University of Trondheim, Ludwig Maximilian University in Munich, the Joint Program in Survey Methodology at the University of Maryland, and at the National Institute of Statistics in Islamabad, Pakistan.

Selected Publications:

- Eckman, S, Kreuter, F, Kirchner, A, Jäckle, A, Tourangeau, R and Presser, S. (2014) "Assessing the Mechanisms of Misreporting to Filter Questions in Surveys" Public Opinion Quarterly.
- Biemer, P, de Leeuw, E, Eckman, S, Edwards, B, Kreuter, F, Lyberg, L, Tucker, C and West, B. (editors). 2017. Total Survey Error in Practice. Wiley.
- Eckman, S, Himelein, K and Dever, J. (2018) "Innovative Sample Designs using GIS Technology" in Johnson, TP, Pennell, B, Stoop, I and Dorer, B. Advances in Comparative Survey Methods: Multicultural, Multinational and Multiregional Context. Wiley.

Short Course Description:

This course will cover: methods of sample selection; calculation of weights that adjust for nonresponse and undercoverage; and analysis of complex survey data. We will also discuss analysis of nonprobability surveys. The emphasis of the course is more applied than theoretical, but students are expected to be comfortable with statistics and to have some experience with data analysis. For each topic, students will do exercises in Stata that apply the techniques learned in the lectures. Students will get the most out of the class if they have prior experience with Stata.

Keywords:

Sampling, Analysis, Variance, Weighting, Nonresponse

Course Prerequisites:

- Introductory course in statistics. No prior knowledge of sampling theory is assumed, but students should be comfortable with statistical concepts such as hypothesis testing, variance, standard errors, confidence intervals, etc.;
- Prior knowledge of Stata is required for this course;
- Basic understanding of survey methodology (this could be gained in the course "Introduction to Survey Design" in the first week);
- Experience in handling survey data is helpful but not necessary.

Target Group:

Participants will find the course useful if:

- they have some experience conducting surveys and/or analyzing social science data, but have not yet studied sampling;
- they are conducting their own survey or are analyzing survey data.

Course and Learning Objectives:

By the end of the course participants will

- have a sound understanding of the most frequently used sample designs (one- and two-stage sampling, clustered sampling, stratified sampling, and related designs);
- know how to create probability weights, and have experience with several methods for adjusting such weights for nonresponse and undercoverage;
- understand how and why the design of a sample survey affects the analysis of the data;
- know the appropriate methods to use in Stata to analyze complex survey data, and the pros and cons of each method.

Organizational Structure of the Course:

The course will consist of 4 hours of lecture in the morning. Students will spend the afternoons working on exercises (most of them in Stata) designed to deepen their understanding of the material. The instructor will be available in the afternoons to answer questions.

Software and Hardware Requirements:

None. This course takes place in a computer lab. GESIS will provide participants with access to the statistical software package used in the course (i.e., Stata) as well as other packages (e.g., SPSS, Mplus, R). Students may also use their own laptop, if Stata installed.

Instructor will use own laptop.

Stata ados to install on student laptops:
Gsample
Unique
Survwgt
Svr
Coefplot
Grc1leg
Dirtools
Moremata
Egenmore
Kdens
samplepps
estout
ipfraking

R packages to install on student laptops:
foreign
MASS
dplyr
survey
PracTools

Long Course Description:

The data analysis techniques often taught in introductory statistics courses rely on the assumption that the data come from a simple random sample. However, many of the data sets that we use as social science researchers are based on complex samples: that is, those that include stratification and/or cluster sampling. In addition, the cases may have unequal weights due to sample selection or adjustment for nonresponse and undercoverage. Data from such complex samples cannot be analyzed in quite the same way as data from simple random samples. This course discusses how these sample design elements affect estimates and provides some guidance on how to do proper analyses accounting for complex sample designs.

We will cover three interrelated topics: methods of selecting samples, creation of analysis weights that adjust for nonresponse and undercoverage, and the analysis of data collected via complex weighted surveys. In each section of the course, students will have ample time to apply the concepts and techniques learned in the lecture to exercises in Stata.

The first section, on sampling, will begin with simple random sampling and move onto stratified, clustered and multi-stage sampling. 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 section will cover the calculation of sampling weights that adjust for unequal probabilities of selection and the ways in which such weights can be adjusted for nonresponse and undercoverage. Each adjustment method relies on assumptions, and students will learn the pros and cons of each method. During this section we will also learn how to create weights for nonprobability studies. Again, the course will involve several practical exercises, so that students can practice the methods learned in class.

The third section will teach students how complex designs and weights alter the ways in which survey data should be analyzed. Traditional methods of statistical analysis, usually taught in introductory statistics courses, do not apply to such data sets. Researchers who use this in appropriate methods will estimate incorrect standard errors and may come to incorrect scientific and policy conclusions. There are several different methods that can be used to analyze complex weighted survey data, such as jackknife and bootstrap approaches to estimation of standard errors. Students will learn both why we need new methods and how the methods can be applied, as well as the advantages and disadvantages of each method.

The course will be applicable to surveys of individuals, schools, households and organizations.

Day-to-day Schedule and Literature:

Day	Topic(s)
1	Sample designs ▪ Different types of sample designs: simple random, stratification, clustering ▪ Probability proportional to size sampling, systematic sampling ▪ Selecting samples in Stata ▪ Motivation of effects of stratification and clustering on standard errors
	Suggested reading: ▪ Chapter 2 in: Heeringa, S. G., West, B., & Berglund, P. A. (2017) Applied Survey Data Analysis. Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences. Boca Raton; London; New York: CRC Press.
2	Complex samples, design effects ▪ Sample size calculations ▪ Design effects for stratified, clustered designs ▪ Effects of cluster homogeneity on standard errors ▪ Effective sample size
	Suggested reading: ▪ Chapter 9,10 in: Valliant, R., Dever, J. A., & Kreuter, F. (2013) Practical Tools for Designing and Weighting Survey Samples. Statistics for Social and Behavioral Sciences: Springer.

3	Weighting methods ▪ How to calculate response rates ▪ Baseweights ▪ Adjusting for nonresponse using sample based weighting ▪ Adjusting for nonresponse, coverage and sampling error using population based weighting ▪ Weighting effects ▪ How to calculate weights in Stata
	Suggested reading: ▪ Chapter 13,14 in: Valliant, R., Dever, J. A., & Kreuter, F. (2013) Practical Tools for Designing and Weighting Survey Samples. Statistics for Social and Behavioral Sciences: Springer.
4	Analysis of complex samples ▪ Taylor Series linearization ▪ Bootstrap replication ▪ Other replication methods ▪ How to apply weights in Stata
	Suggested reading: ▪ Chapter 15 in: Valliant, R., Dever, J. A., & Kreuter, F. (2013) Practical Tools for Designing and Weighting Survey Samples. Statistics for Social and Behavioral Sciences: Springer. ▪ Chapter 3 in: Heeringa, S. G., West, B., & Berglund, P. A. (2017) Applied Survey Data Analysis. Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences. Boca Raton; London; New York: CRC Press.
5	Analysis of complex samples ▪ Subpopulation analysis ▪ Regression ▪ Multi-level models ▪ Cross-country analyses ▪ How to do all these things in Stata
	Suggested reading: ▪ Chapter 15 in: Valliant, R., Dever, J. A., & Kreuter, F. (2013) Practical Tools for Designing and Weighting Survey Samples. Statistics for Social and Behavioral Sciences: Springer. ▪ Chapter 3 in: Heeringa, S. G., West, B., & Berglund, P. A. (2017) Applied Survey Data Analysis. Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences. Boca Raton; London; New York: CRC Press.

Preparatory reading:

- Any introductory sampling text covering hypothesis testing, standard errors, confidence intervals.
- Eckman, S and West, B. (2016) Analysis of Data from Stratified and Clustered Surveys in Handbook of Survey Methodology, Wolf, C, Joye, D, Smith, T, and Fu, Y (Eds). Sage.

Additional Recommended Literature:

- Kalton, G. and Kasprzyk, D. (1986) The Treatment of Missing Survey Data. Survey Methodology 12(1), pp.1-16.
- Kreuter, F. and Valliant, R. (2007) A survey on survey statistics: What is done and can be done in Stata. The Stata Journal 7(1), pp. 1 -21.
- Valliant, R. and Dever, J. (2018) Survey Weights: A Step-by-Steph Guide to Calculation. Stata Press.