

# 10<sup>th</sup> GESIS Summer School in Survey Methodology

## [2nd Virtual GESIS Summer School]

28 July – 20 August 2021

### Syllabus for Course 4: Survey Sampling and Weighting

Lecturers:	Dr. Simon Kühne	Zaza Zindel
E-mail:	simon.kuehne@uni-bielefeld.de	zaza.zindel@uni-bielefeld.de
Homepage:	<a href="https://simon-kuehne.de/">https://simon-kuehne.de/</a>	<a href="https://ekvv.uni-bielefeld.de/pers_publ/publ/PersonDetail.jsp?personId=84399219">https://ekvv.uni-bielefeld.de/pers_publ/publ/PersonDetail.jsp?personId=84399219</a>

Date: 02–06 August 2021  
 Time: 09:00–12:30 + 14:00–17:00  
 Time zone: CEST/CEDT, course starts Monday at 10:00 am  
 Venue: Online via Zoom

#### About the Instructors:

*Simon Kühne* is a postdoc at Bielefeld university. His research focus is on survey methodology, social data science, and social inequality. He is an expert in survey sampling and weighting. Prior to his position at Bielefeld University he worked at the Socio-Economic Panel Study (SOEP, DIW Berlin), where he was involved in the sampling and weighting of the latest migrant/refugee samples.

Zaza Zindel is a Researcher at Bielefeld university in the BMBF project "Gender and Sexual Diversity in the Focus: Participation and Diversity of Lifestyles (SOEP-GeSMIn)".

#### Selected Publications:

- Jacobsen, J., & Kühne, S., forthcoming: Using a Mobile App when Surveying Highly Mobile Populations: Panel Attrition, Consent and Interviewer Effects in a Survey of Refugees. *Social Science Computer Review*.
- Kühne, S., Jacobsen, J. & M. Kroh, 2019: Sampling in Times of High Immigration: The Survey Process of the IAB-BAMF-SOEP Survey of Refugees. In: *Survey Methods: Insights from the Field*. doi: 10.13094/SMIF-2019-00005.

#### Short Course Description:

This practical course will cover the process of probability sampling, weighting and estimation. This includes a) methods of sample selection, b) methods of weighting including design weighting, nonresponse weighting and poststratification techniques, and c) methods of parameter estimation for complex sample survey data. Please note that the course will only provide a short excursus to non-probability sampling. 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. Prior knowledge in how to use Stata is needed.

#### Keywords:

Probability Sampling, Weighting, Nonresponse

#### Course Prerequisites:

- Introductory course in statistics. No prior knowledge of sampling theory is assumed.
- Prior knowledge in Stata is required for this course.
- Basic understanding in survey methodology and how to handle survey data.

## Target Group:

Participants will find the course useful if:

- they have experience conducting surveys and/or analyzing survey data but have no experience with survey sampling and weighting.
- they plan their own survey data collection and need to sample and/or weight the data.

## Course and Learning Objectives:

By the end of the course participants will:

- know about the most commonly used sample designs including, clustered sampling, stratified sampling, and multi-stage sampling.
- know how to create design weights, nonresponse weights and apply poststratification techniques.
- know how the sample design can affect data analysis and how to incorporate complex survey designs and survey weights into parameter estimation.

## Organizational Structure of the Course:

A typical course day will consist of two thematic clusters. Each cluster starts with a lecture of around 1 hour. After that lecture, participants split into small groups and work on practical exercises using Stata. Simon will provide and discuss a solution for each exercise in-depth afterwards. Each day ends with a recapitulation of the two thematic clusters. The last half of the last course day will consist of 1-to-1 sessions in which Simon provides guidance for current sampling and weighting projects of the participants.

## Software and Hardware Requirements:

Stata, R and RStudio should be installed on participants' computers. The following Stata packages should be installed on all machines:

- gsample
- unique
- distinct
- survwgt
- fre
- coefplot
- moremata
- estout

## Long Course Description:

The majority of social science research projects rely on random samples. High quality surveys require elaborate sampling and recruiting procedures prior to the actual data collection.

Random sampling in practice is almost never "simple random sampling". Usually, more complex sample techniques such as stratified sampling, cluster sampling, multi-stage sampling, and their combination are used to draw samples of the general population or specific subgroups. This course will cover the most common sample designs and how to apply them for own survey research projects.

Complex sample designs usually cause sample probabilities to vary across sample elements (e.g., households or individuals). Survey weights are needed to account for these unequal probabilities, and thus, allow for unbiased survey estimates. Moreover, in almost every survey, some of the sampled individuals cannot be interviewed, either because they cannot be contacted, or they refuse to participate. In this regard, so-called nonresponse weights are needed to minimize potential bias in survey estimates due to systematic differences between respondents and nonrespondents. Finally, in some cases, we use weights so that sample properties match exactly to known distributions in the target population (gender or age, for instance). This course will cover survey weighting in all of these three aspects: design weighting, nonresponse weighting, and post-stratification techniques.

Finally, analyzing survey data based on complex sample designs requires specific techniques as traditional methods of statistical analysis cannot always be applied. Not applying these techniques can lead to biased estimates and incorrect standard errors, causing researchers to may come to incorrect scientific and policy conclusions. In our course, we will discuss the most important methods available to analyze complex, weighted survey data.

To sum up, in this course, participants will learn to plan and draw complex samples (1), to create survey weights (2), and to analyze survey data based on complex samples (3).

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

### Day-to-day Schedule and Literature:

Day	Topic(s)
1	Introduction to Survey Sampling <ul style="list-style-type: none"> <li>▪ Key concepts &amp; definitions</li> <li>▪ Sampling frames</li> <li>▪ Simple random sampling</li> <li>▪ Sample sizes</li> </ul>
	<u>Suggested reading:</u> <ul style="list-style-type: none"> <li>▪ Chapter 1, "Introduction" Lohr, S. L., 2009: Sampling: Design and Analysis. 2nd edition. Cengage.</li> </ul>
2	Complex Survey Sampling <ul style="list-style-type: none"> <li>▪ Stratified sampling</li> <li>▪ Cluster sampling</li> <li>▪ Multi-stage sampling</li> <li>▪ Design effects</li> </ul>
	<u>Suggested reading:</u> <ul style="list-style-type: none"> <li>▪ Chapter 3 and 5 in Lohr, S. L., 2009: Sampling: Design and Analysis. 2nd edition. Cengage.</li> </ul>
3	Weighting I <ul style="list-style-type: none"> <li>▪ Design weights</li> <li>▪ Nonresponse weights</li> </ul>
	<u>Suggested reading:</u> <ul style="list-style-type: none"> <li>▪ Chapter 13,14 in: Valliant, R., Dever, J. A., &amp; Kreuter, F. (2013) Practical Tools for Designing and Weighting Survey Samples. Statistics for Social and Behavioral Sciences: Springer.</li> <li>▪ Groves, R., 2006: Nonresponse Rates and Nonresponse Bias in Household Surveys, Public Opinion Quarterly, Volume 70(5): 646–675, <a href="https://doi.org/10.1093/pog/nfl033">https://doi.org/10.1093/pog/nfl033</a></li> </ul>
4	Weighting II <ul style="list-style-type: none"> <li>▪ Post-stratification &amp; Raking</li> <li>▪ Trimming</li> <li>▪ Evaluate weighting factors</li> </ul>
	<u>Suggested reading:</u> <ul style="list-style-type: none"> <li>▪ Kalton, G. and I. Flores-Cervantes, 2003: Weighting Methods. Journal of Official Statistics 19(2): 81-97.</li> </ul>
5	Analyzing Complex Sample Survey Data <ul style="list-style-type: none"> <li>▪ Applying weights</li> <li>▪ Multi-level models</li> <li>▪ Bootstrapping</li> </ul>
	<u>Suggested reading:</u> <ul style="list-style-type: none"> <li>▪ 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.</li> </ul>

### Preparatory Reading:

Any introductory sampling text covering hypothesis testing, standard errors, confidence intervals.

### Additional Recommended Literature:

- 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.
- Lohr, S. L., 2009: Sampling: Design and Analysis. 2nd edition. Cengage.
- Valliant, R. and Dever, J. (2018) Survey Weights: A Step-by-Step Guide to Calculation. Stata Press.
- Valliant, R., Dever, J. A., & Kreuter, F. (2013) Practical Tools for Designing and Weighting Survey Samples. Statistics for Social and Behavioral Sciences: Springer.