GESIS Summer School in Survey Methodology 2023

Syllabus for course:
“Advanced Survey Design”

Lecturers: Bella Struminskaya
                Angelo Moretti
Email:     b.struminskaya@uu.nl
                a.moretti@uu.nl
Homepage: https://bellastrum.com
                https://www.uu.nl/staff/AMoretti

Date: 14 – 18 August 2023
Time: Mo: 10:00-17:00 | Tu-Fr: 09:00-16:00
Venue: GESIS Cologne

About the lecturers
Bella Struminskaya is an assistant professor in methods and statistics at Utrecht University. Her research focuses on the design and implementation of online, mixed mode, smartphone surveys, smartphone sensor studies and data donation. She has published on various aspects of data quality, nonresponse and measurement error, including panel conditioning and device effects, passive data collection, smartphone sensor measurement and recruitment and maintenance of online panels. Her current research focuses on design of sensing studies, collecting digital trace data, and privacy concerns.

Angelo Moretti is an assistant professor in Statistics at Utrecht University in the Department of Methodology and Statistics. He is a survey statistician and an elected member of the International Statistical Institute (ISI). He has conducted research in small area estimation under multivariate mixed-models, survey calibration, mean squared error estimation based on bootstrap approaches, and data integration methods (statistical matching and probabilistic record linkage). He is also interested in applications related to understanding geographical differences in social exclusion, crime and public attitudes indicators.

Selected Publications:
Course description

This course in ‘advanced survey design’ takes students beyond the introductory courses offered in BA and MA programmes and will discuss current issues in one of the most important data collection methods: surveys. We focus on doing surveys in the Internet-era, the state of the art in both the design and the analysis of survey data. We discuss new ways to analyse modern surveys, including non-probability survey designs, surveys conducted via apps, and mixing surveys with Big Data such as augmenting survey data with auxiliary data (e.g., register data, digital trace data). The course combines 1-hour lectures with exercises on most of the topics discussed. The course is taught through lectures, group exercises as well as personal tutorial meetings that give participants the opportunity to discuss their own survey designs. Each day we will discuss a specific topic that each focuses on one or more aspects of survey design within the Total Survey Error framework. Course participants must be proficient working with a statistical software package (e.g., R, Stata, SPSS) at the level of at least knowing multiple linear regression. Most course materials are prepared for working with R.

Keywords
Survey methods, sensor data, big data, total survey error, questionnaire design

Course prerequisites
- We expect students to have quite extensive knowledge of survey research (for example, by using survey data or conducting survey research in their daily work) and
- We expect students to have knowledge of statistics at the MSc level for social scientists (the general linear model).
- We ask students to send us a brief motivation letter (up to 200 words), in which they write a short list of your experience with surveys and statistical analyses. Please include in the motivation letter what software you are proficient in, and what courses related to survey design you have taken.
- Participants should be prepared to share information about the survey they are planning on conducting, they are currently involved with or would like to conduct in the future.

Target group
The course will be interesting for PhD students or PhD candidates, postdocs, or advanced researchers with a background in a social science discipline or in Statistics, working on academic research projects. This course is focused on survey research within the academic (university/scientific institute) setting, as well as focused on current issues related to mobile surveys and Big Data.

Course and learning objectives
The aim of this course is to provide an overview of theory and practice of modern survey design and analysis, in particular focusing on modern methods of web surveys and the analysis of new types of data (e.g., sensor data). This course is useful for more experienced students.

By the end of the course participants will be ready to apply the learned methods and techniques towards their own surveys, are able to critically assess existing surveys and survey documentation and analyze survey data themselves successfully.

Organizational structure of the course
Lectures, exercise, group studies, tutoring sessions will be held daily between 10:00 and 17:00 on Monday, and the rest of the week between 9:00 and 16:00 including breaks.
Software and hardware requirements
Participants need to bring a laptop computer to the course with their preferred statistical software installed. Since most exercises are prepared for R, ideally with R (version 4.2.3 or higher, [https://cran.r-project.org/](https://cran.r-project.org/)) and RStudio installed ([https://www.rstudio.com/](https://www.rstudio.com/)). Both programs are free and open source. Participants need to be able to download and install packages (internet access is provided by GESIS).

Participants may also use Stata or SPSS for the exercises, but answers/solutions will be provided in R. Participants who wish to work in Stata but do not own a license, may request a short-term license valid for the duration of the course from GESIS Training.

Day-to-day Schedule and Literature:

<table>
<thead>
<tr>
<th>Day</th>
<th>Topic(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>Introduction to the Total Survey Error Paradigm</strong></td>
</tr>
<tr>
<td></td>
<td>- Types of data and their relation to Total Survey Error</td>
</tr>
<tr>
<td></td>
<td>o Designed data</td>
</tr>
<tr>
<td></td>
<td>o Organic data</td>
</tr>
<tr>
<td></td>
<td>o Designed big data</td>
</tr>
<tr>
<td></td>
<td>- Exercise: Study design and minimizing Total Survey error (in groups)</td>
</tr>
<tr>
<td></td>
<td>- Choosing an appropriate sampling frame and sampling design. Registers, geodata and digital trace data.</td>
</tr>
<tr>
<td></td>
<td>- Sampling designs: statistical efficiency, survey costs and survey practice.</td>
</tr>
<tr>
<td></td>
<td>- Exercise: Working out a sampling design (computer exercise)</td>
</tr>
</tbody>
</table>

**Suggested reading:**

| 2   | - **Advanced questionnaire design**  |
|     | - Mixing the modes  |
|     | - Designing for mixed mode surveys (exercise)  |
|     | - Mobile and mixed-device surveys  |
|     | - Questionnaire design for mixed-device surveys  |
|     | - Exercise: questionnaire design for mixed-device surveys  |

**Suggested reading:**
| 3 | - Weighting to correct for survey nonresponse  
- Paradata: what is it and how to use it?  
- Exercise: Creating poststratification weights (computer exercise)  
- Sampling, coverage and nonresponse weights  
- Exercise: Raking, combining weights (computer exercise)  
- Exercise: Imputation or weighting (computer exercise) |
| --- | --- |
http://iserwww.essex.ac.uk/home/plynn/downloads/Lynn%201996%20Weighting.pdf  
| 4 | - Surveys and big data  
- Passive data collection using mobiles (sensors)  
- Ethics, consent, willingness  
- Exercise: Introduction to working with geo-data or accelerometer data (choose 1) (computer exercise)  
- Sampling revisited: design-based vs. model based inference and effects on Total Survey Error?  
- Exercise on model-based inferences from (non) probability samples |

5

- **Data donation**
- **Working with text or picture data**
- **Data integration**
- **Your own project. Presentations and consultations with teachers of the course to discuss your survey questions in more depth. You may bring your own dataset, questionnaire or study design to discuss. Alternatively, there is time to finish some of the exercises earlier or read specific literature.**

**Suggested reading:**


**Preparatory Reading:**

No preparatory reading is necessary. We assume participants to have basic knowledge of statistics (descriptive, graphs, and inferential statistics at the level of at least multiple regression). During the course, students will find the suggested reading in the course manual helpful in explaining the topics covered in greater detail.

**Additional Recommended Literature:**

None