Course title
WA110 Weighting techniques to handle survey nonresponse

Instructor details
First name, last name: Hans Walter Steinhauer
Department/Unit: Methods Department & Department of Statistics and Econometrics
Institution: Leibniz Institute for Educational Trajectories & University of Bamberg
Full postal address for ECPR correspondence:
Leibniz-Institut für Bildungsverläufe e.V.
z. H. Hans Walter Steinhauer,
Wilhelmsplatz 3,
96047 Bamberg
Phone: +49 951 – 863 3450
Fax: +49 951 – 863 3411
E-mail: Hans-Walter.Steinhauer@lifbi.de

Short Bio
Hans Walter Steinhauer is a research assistant at the Leibniz Institute for Educational Trajectories (LIfBi) working on sampling designs and weighting strategies for complex survey designs. He is also a lecturer at the Department of Statistics and Econometrics at the University of Bamberg teaching sampling techniques and applied statistics using R.

Prerequisite knowledge
Note from the Academic Convenors to prospective participants: by registering to this course, you certify that you possess the prerequisite knowledge that is requested to be able to follow this course. The instructor will not teach again these prerequisite items. If you doubt whether you possess that knowledge to a sufficient extent, we suggest you contact the instructor before you proceed to your registration.
[do indicate if no specific prerequisite knowledge is requested. Otherwise, in short, list the core prerequisite items (conceptual, technical, software, ...). Not much detail is needed.]
- Introductory statistics
- Basic knowledge of sampling (refresher given in the course according to the reading list)
- Basic knowledge of linear and binary regression models
- Prior experience with at least one general syntax-based statistical software such as for example R, SPSS, Stata, etc.
- Basic knowledge of R is helpful but not necessary

Short course outline
The course will give insights on the derivation of weights, their successive nonresponse adjustments and their application to selected analyses. Therefore the course starts with a short refresher on sampling techniques and the corresponding derivation of design weights. The first part will introduce the fundamentals of weighting and how weights should be applied. Because most surveys are subject

1 Disclaimer: the information contained in this course description form may be subject to subsequent adaptations (e.g. taking into account new developments in the field, specific participant demands, group size etc.). Registered participants will be informed in due time in case of adaptations.
to unit nonresponse, that is, people refuse to participate, the design weights need to be adjusted to account for nonparticipation. The impact of nonresponse on survey estimates will be discussed in detail and examples will be given on how to handle this nonresponse by using weighting adjustments. Different methods and their appropriateness will be covered in the second part of the course. The last part of the course will cover the topic of applying weights in estimation routines. Examples will be given in R (www.r-project.org).

**Long course outline**

Usually weighting is performed in three steps. In the first step design weights (also known as base weights) are usually computed as the inverse of the inclusion probability. Thus, for most designs they are directly available after sampling. Design weights compensate for unequal probabilities of selection, unequal sampling fractions in stratified samples, that is, oversampling, or for subsampling. In the second step the design weights are adjusted to correct for unit nonresponse. This is often referred to as sample weighting adjustment. This step needs to be considered wherever unit nonresponse occurs. Sample weighting adjustments correcting for unit nonresponse usually result in increasingly varying weights and thereby lower the precision of survey estimates. The third step, referred to as population weighting adjustment, calibrates weights such that estimates conform to known parameters (for example totals or ratios) of the population. This last step corrects for potential bias due to incomplete coverage of the population and sampling error.

The course will cover each of the three steps emphasizing the sample weighting adjustments for unit nonresponse. Starting with sampling techniques and giving a refresher on simple random sampling designs as well as on unequal probability designs the course will illustrate their usage based on small object lessons. This part will serve as a basis to understand how design weights are derived. After this refresher it will become clear what the purpose of design weights is and how they are generally used. In most surveys unit nonresponse occurs for a variety of reasons ranging from not being able to contact target persons to targets that refuse to participate in the survey. At this point a short distinction will be made between cross-sectional and for panel surveys. Building on this the impact of unit nonresponse will be addressed. Different mechanisms of unit nonresponse will be covered and it will be made clear in which cases weighting adjustments for unit nonresponse makes sense and in which it does not.

Having the basic knowledge at hand the course will proceed with different methods for sample weighting adjustments for unit nonresponse. These methods will cover cell weighting, raking, iterative proportional fitting, post-stratification, generalized regression weighting as well as propensity score weighting methods. Each of these methods is suitable for different settings according to the data at hand. Thus, the decision towards which method to use heavily depends on the data available for unit nonresponse adjustments. It is not always possible to collect data for nonrespondents, and the only data available for adjustments in this case will be marginal distributions for certain characteristics of the population. For example, the percentages of men and women as well as the distribution of age in the population are in general available from official statistics sources. Having a joint distribution for two or more characteristics the methods of the previous settings can be used as well as method more appropriate using the richer set of information. An example for this joint distribution would be to have a cross tabulation of gender and age groups in the population. In this data setting the joint as well as the marginal distributions are available. Having data on respondents and nonrespondents allows for using the full set of adjustment methods. Modelling the decision to participate given auxiliary information is referred to as response propensity modelling. Here the dichotomous participation status, that is, responding to the survey or not, is regressed on auxiliary variables available for both groups using mostly logit or probit models for binary data. The data available for both groups vary from design information, for example geographical region, to individual level information such as gender, age, etc. Again, a distinction between cross-sectional and longitudinal survey will be addressed. Some of the methods introduced for sample weighting adjustments are also suitable for population weighting adjustments and thus will be discussed for both steps in the adjustment process. Corresponding examples referring to each of the steps will be given within the course.
In the last part of the course the attendees are expected to have the knowledge on how weights are derived and to understand the methods applied in sample weighting adjustments. Based on this knowledge the course will conclude with suggestions for using weights in estimation. This triggers the decision whether or not weights have to be used. Three convenient methods will be introduced covering this topic. Throughout the course examples will be given in R (www.r-project.org). Attendees do not necessarily have to be familiar with the R language but we recommend basic skill in at least one syntax based statistical software. Alternatives to weighting adjustments will be addressed shortly at the end of the course but not covered in more detail.)

Day-to-day schedule (Friday 13 February to Saturday 14 February)

<table>
<thead>
<tr>
<th>Topic(s)</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Friday afternoon</strong></td>
<td>Refresher in survey sampling and derivation of design weights</td>
</tr>
<tr>
<td><strong>Saturday morning</strong></td>
<td>The impact of unit nonresponse and weighting adjustments</td>
</tr>
<tr>
<td><strong>Saturday afternoon</strong></td>
<td>Taking care of weights in analysis of survey data</td>
</tr>
</tbody>
</table>

Day-to-day reading list

<table>
<thead>
<tr>
<th>Readings (please list at least the compulsory reading for the scheduled day)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Friday afternoon</strong></td>
<td>Lohr (2010) Kish (1990, 1992)</td>
</tr>
</tbody>
</table>

9. Software and hardware requirements
R, RStudio 3.x.y with the following packages: sampling, simFrame, SDaA, survey, pps, ipfp, weights

**Hardware requirements**
Beamer, overhead projector

**Course Literature**


Further reading


Lab requirement

Beamer and overhead projector

12. Other recommended courses (before or after this course)

The following other ECPR Methods School courses could be useful in combination with this one in a ‘training track’. NB this is an indicative list.
### Before this course:

<table>
<thead>
<tr>
<th>Course title</th>
<th>Summer School</th>
<th>Winter School</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### After this course:

<table>
<thead>
<tr>
<th>Course title</th>
<th>Summer School</th>
<th>Winter School</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>