Because of the random assignments, this design allows for studying crossed interviewer and respondent effects, using cross-classified random effects models (e.g., Fielding and Goldstein 2006). Such models can estimate both interviewer and respondent random effects simultaneously. In a face-to-face design (see Figure 1) however, to control for unobserved respondent (and/or interviewer) effects, fixed effects models would have to be used. However, if there is no correlation between the regressors and respondent (and/or interviewer) specific errors (usually termed ui in the econometric literature) – as is the case in experimental designs - random effects models are consistent.

We present some examples (see reference list) which benefit from this randomized assignment to analyze interviewer, respondent, and timing effects. We make use of the contact file from the SHP (which include call(attempt) information such as time of call, call outcome, interviewer ID) and information from the yearly interviewer survey. Four papers (Lipps 2007, 2010; Lipps and Lutz 2010, forthcoming) make use of the randomized interviewer-respondent assignment over waves, and three papers analyze randomly assigned calls during the procedure to try to obtain contact and cooperation of the sample members (Lipps 2008, 2009, 2012).

References

Lipps, O. (2007) Interviewer and respondent survey quality effects in a CATI panel. Bulletin de Méthodologie Sociologique 95: 5-25.

Lipps, O. (2008) A note on interviewer performance measures in centralised CATI surveys. Survey Research Methods 2 (2): 61-73.

Lipps, O. (2009) Cooperation in centralised CATI household panel surveys - A contact-based multilevel analysis to examine interviewer, respondent, and fieldwork process effects. Journal of Official Statistics 25 (3): 323-338.

Lipps, O. (2010) Does interviewer-respondent socio-demographic matching increase cooperation in centralized CATI household panels? Survey Practice (August 2010).

Lipps, O. (2012): A Note on improving contact times in panel surveys. Field Methods 24 (1): 95-111.

Lipps, O. and G. Lutz (2010) How answers on political attitudes are shaped by interviewers: evidence from a panel survey. Special Issue (Lipps, O., R. Tillmann, U. Kuhn, D. Lillard, and M. Bergman (eds.): "Social science research with panel data in Switzerland") of the Swiss Journal of Sociology 2/2010: 345-358.

Lipps, O. and G. Lutz (forthcoming) Gender of interviewer effects in centralised CATI panel surveys. Unpublished manuscript, FORS, Lausanne.

The Dataset Project: Handling survey data in R

Emmanuel Rousseaux, NCCR LIVES Overcoming vulnerability, Life-course perspectives, Switzerland
Institute for Demographic and Life Course Studies, University of Geneva, Switzerland
Gilbert Ritschard, NCCR LIVES Overcoming vulnerability, Life-course perspectives, Switzerland
Institute for Demographic and Life Course Studies, University of Geneva, Switzerland

Population studies strongly rely on survey data. To meet the needs of recent research questions in social sciences, data collected have become in the past decades more and more complex, such as longitudinal data, network data and spatial data. These high volumes of structured data complicate the task of both documenting data and manipulating data, as for example when we want to prepare data for a specific study. There is a need for specific tools to assist the user in handling these complex data. The Dataset software is an effort in this direction, aiming at providing a framework for handling survey data in R, especially network and biographical data. More precisely, the software aims at facilitating the management of survey data by providing researchers in social sciences with high-level tools for storing, documenting, sharing, exploring and recoding survey data in a secure and efficient way. This initiative, conducted within the NCCR LIVES project, targets mainly life course data and especially data types collected and used within the NCCR LIVES project. Thus, the current roadmap includes the development of the framework to support (1) cross-sectional data, (2) network data with a specific handling of demographic data from people cited in the network of each respondent, and (3) panel data organized in successive waves.

The software comes in the form of an R package. R is a powerful statistical tool, freely available and multi-platform which is nowadays more and more often used in the social sciences as an alternative to classical commercial software (SPSS, SAS, Stata). As R is open-source, a lot of researchers in methods appreciate to be able to share their work through this software. As a consequence, most of the recent state-of-the-art methods are available in R and many of them in R only. This is especially the case for the newest tools for life course analysis (e.g. the *TraMineR* R package for life course sequence analysis and the *ltm* R package for latent class model). Further, working in R allows benefiting from the numerous statistical procedures already optimized in R and taking advantages of the R powerful graphical capability.

From a general point of view, the Dataset software follows three goals:

- *Providing an efficient framework for storing and documenting complex survey data.* As a key point, the software aims at storing in a single object data together with the design of the survey within which data were collected. The data and the user manual describing the data are thus merged together. For example, the package offers a structure for assigning short and long labels to variables and variable values, declaring user-defined types of missing values, and accounting for cross-sectional and longitudinal weights. Many important metadata can also be stored such as the population concerned by the survey, the used sampling method, the organization releasing the data, the user license type, etc. As all information is stored within the data object, we provide a method for such objects for generating a summary of the whole data base. The summary gives for each variable its long label, the percent of valid cases and basic descriptive statistics. This summary can be directly exported as a PDF file and serves as a basic user manual of the database that proves particularly useful for sharing data with others.
- *Saving the scientist's time spent on data processing in favor of time devoted to the research question*. Preparing data for a study is often a very burden task. The Dataset software is intended to help the analyst in this task, allowing her to focus more quickly on the analysis. For example, the package provides a search function allowing exploring the whole database and retrieving relevant variables, which is especially useful for huge databases. It provides efficient tools for recoding categorical and quantitative variables. The Dataset software also provides support for handling missing values and allowing for instance to easily turn a missing value into a valid case and vice-versa. Further, the software provides for some classical statistical methods front-ends especially designed for scientists in social sciences. These front-ends facilitate the scientist daily work within the R environment. As a key point, the software performs systematic data consistency checks to ensure that data were not altered by data preprocessing operations. When filtering out cases and using weights when available, the software also processes automatic checks to prevent the loss of representativeness with respect to control variables defined by the user.
- *Facilitating reproducible research*. Demographic and sociologic questions are generally complex and require a lot of work to be understood. Reproducible research, requires attaching sufficient information about the performed data analysis to allow anyone to retrieve the same results, is a helpful methodology when studying social dynamics. Having the possibility to rerun an experiment made by other researchers, or by oneself several months ago, gives the possibility to verify, better understand, and pursue an already done work. The Dataset software works in this direction by tracing operations made on data, so that the user can find back previously performed operations. Furthermore, for each method provided by the package, results can be printed in a PDF file which also provides all settings used for calibrating the method. Outputs are displayed with a "ready-to-publish" formatting, allowing to quickly focusing on the interpretation.

In addition of these tools for cross-sectional data, the proposed software solution provides efficient methods for handling panel data organized in successive waves such as in the Swiss Household Panel. The user can directly extract whole trajectories from the panel data without having to bother with extracting the same variable independently from each yearly wave. The software automatically checks for each variable that it shares the same missing values and valid cases across years. By specifying '..' in place of the two year digits in the variable names, the user can extract a whole sequence in a single step. Likewise, he can recode or merge some values, or turn a missing value into a valid case directly for all waves where the variable exists. There also is a method for exporting a trajectory as a sequence object ready to be analyzed with the *TraMineR* package.

Comparing different cross-sectional weights for children in the Swiss Household Panel

Martina Rothenbühler, FORS

Objectives

Currently, only children aged 14 and more have a cross-sectional and longitudinal individual weight in the Swiss Household Panel (SHP). For the approximately 1500 children younger than 14 years of age at each wave, information coming from proxy questionnaires is available, but no weight is provided for them. Until now, it is thus not possible to conduct weighted analysis including the children using the data of the SHP. The introduction of children's cross-sectional weights represents therefore a direct gain for the social research in Switzerland. From a methodological point of view, the weighting of children is interesting, as only few longitudinal panels include children in their weighting system and new developments are needed.

Methods

The weighting system of the SHP is based on different approaches. One of them is the Generalized Weight Share Method (GWSM) of Lavallée (2007), which is also used for the construction of the children's cross-sectional weights. The GWSM enables to allocate a weight to individuals who joined the household after the first wave of the panel and who thus have an unknown inclusion probability. In the context of panel surveys, the sampling frame U^A can be associated to the initial population at wave 1, while the target population U^B represents the