EDUCATIONAL ISSUES

An experiment on the effects of non-response reweighting on estimators' precision in a web survey

ENRICO FABRIZI *(Catholic University of Sacro Cuore, Italy, enrico.fabrizi@unicatt.it)*
SILVIA BIFFIGNANDI *(University of Bergamo, Italy, silvia.biffignandi@unibg.it)*
DANIELE TONINELLI *(University of Bergamo, Italy, daniele.toninelli@unibg.it)*

The paper analyzes reweighting adjustments for non-response in surveys carrying out a bootstrap evaluation of non-response adjusted estimators.
In our study we consider a population made of students from the University of Bergamo graduating between A and B.
This population has been surveyed twice (web mode in both cases), before and after graduation. The ante-graduation survey (hereinafter: AGS) is in fact a census, the post-graduation survey (PGS) aimed at being another "census", but as answering was not compelling anymore, we have a 56% non-response rate. Administrative (archive) data is available for all the students.
We apply the non-response process of the PGS in the analysis of AGS data. In this way, we have a controlled situation in which all survey variables, for both respondents and non-respondents are known. We avoid artificial assumptions on the non-response process.
AGS covers many topics, including student satisfaction for his/her university experience. Using the complete AGS data set, we calculate two satisfaction factors from a set of satisfaction questions. This is done using the factor analysis (via principal components) method. First factor may be interpreted as a "general satisfaction" index; the second is a contrast between software (teaching, exams, graduation organization) and hardware (classrooms, libraries, cafeterias) evaluation. Factors are not directly observable, but we treat them as they are, for simplicity's sake. They are continuous, approximately normal variables.
We have 459 students in the artificial population, 203 among "respondents". Is the non-response process MAR ("Missing at Random")? To say this, we regress the response indicators on all the administrative variables (including sex, age, number of years needed for graduating, graduation mark, high school final mark, faculty, type of course) and the two satisfaction factors. There is slight evidence that the respondents and non-respondents have different distribution of factor scores.
If we have to take any inference on the non-respondents we must assume that they are related to respondents in some way. The way is often the assumption that they are related through the auxiliary information, i.e. through variables known for both respondent and non-respondents.
We bootstrap the population to evaluate the ability of the calibration correction to improve the estimators of non-response. Firstly, we try to use the quasi-

randomization approach to estimate propensity, then use these weights as a basis for calibration. Several combinations for calibration variables are used. The faculty is always included as they are main subdivisions of the University and estimates by faculty are routinely required.
Calibration 1: only faculty and type of course (MoS, G)
Calibration 2: faculty and non-response explaining variables
Calibration 3: faculty and non-response explaining variables, variables significant for the factors
Calibration 4: faculty and non-response explaining variables, variables significant for the factors, but excluding the presence of personal email in the register
Calibration 5: faculty and factors explaining variables.

Keywords: web survey, non-response, bootstrap, calibration, quasi-randomization