SURVEYING THE HUMAN POPULATION: ERRORS AND THEIR CORRECTIONS

Outline of Ph.D. Thesis

BLANKA SZEITL

Supervisor: **Tamás Rudas**, Professor Emeritus of Statistics

Doctoral School of Mathematics
and Computer Science
University of Szeged, Bolyai Institute

Szeged

2024

Introduction

The thesis discusses human population surveys within the framework of finite population samples and presents approaches to improve the quality of estimates. It makes contributions to the fields of mathematical statistics and survey methodology by summarizing the mathematical foundations of human population surveys and proposing new methods for handling the two most pressing aspects: uncertainty in the sample composition and uncertainty in the measurement. The thesis presents theoretical approaches, an evaluation of a specific survey experiment, and simulations.

Survey statistics is an applied field of mathematical statistics, where data are obtained from questionnaire (survey) data collections. Within the field of survey statistics, human population surveys are data collections where the observed units are individuals. Sample surveys from human populations play a crucial role in providing a large portion of quantitative data about our economy and society. National statistical agencies frequently release estimates for various indicators, including unemployment rates, poverty rates, crop production, retail sales, and median family income. Although some statistics may be derived from complete enumerations (censuses), the majority are derived from samples taken from the relevant population [1]. In recent times, there has been a decrease in the precision of survey estimates. This is particularly evident in election forecasts, as they are one of the rare instances where the previously estimated population parameter becomes known. The prominent role of human population surveys and the quality concerns of data have made the statistical evaluation and development of survey-type data collections increasingly relevant. The thesis is a pioneering one as it aims to summarize the mathematical foundations of human population surveys and propose new approaches to improve the quality of estimates derived from human samples and responses. The thesis presents both theoretical and empirical findings of the author.

The thesis first outlines the background of human population surveys. We introduce the history of the field and highlight how this method developed within mathematical statistics, as well as show the relevance and magnitude of sample surveys. This thesis also outlines the current challenges of human population surveys and presents the main motivation to analyze uncertainties in sample composition and measurement.

We introduce the mathematical foundations with the relevant sampling methods and we show how sample survey theory can be dealt with within general statistical theory. Since in human population surveys, the observed units are individuals, the experiment results in errors and biases that are difficult to manage. Quality control in survey sampling, in general, is presented based on the total survey error (TSE) framework. Errors and biases are also categorized based on mathematical and non-mathematical factors.

We present a new allocation mechanism to handle errors in sample composition. The mechanism takes into account the expected response rates (ERRs). The relative performance of the ERR allocation is assessed by comparing the variances in the resulting estimates. Asymptotic variances are calculated using the δ-method and then initially compared by assuming correctly specified response rates. Variance comparison is made in terms of misspecified response rates and the results of an extensive evaluation using various combinations of specific population parameters are presented.

We also introduce a new scheme that investigates nonresponse and measurement uncertainties through replication surveys. We present how the general concept of assessing errors and biases can be reconsidered when comparing results with previous surveys. Our approach defines uncertainties regarding sample composition and measurement and decomposes total differences in theory as well as based on a case study. The main results are that the total uncertainty can be decomposed exclusively into nonresponse and measurement uncertainties. Measurement uncertainty is more relevant than nonresponse uncertainty. The respondents are generally inconsistent in their responses at the individual level, which implies uncertainties in the measurements.

Backgrounds

In the context of questionnaire surveys, individuals, or households are observed, which introduce human elements into the experiment. This means that the outcome of the experiment is not solely determined by the sampling procedure, but also influenced by factors such as whether the selected individuals are actually contacted, their willingness to respond to the survey, and to provide certain information, as well as the accuracy of the information they provide. These factors introduce both random errors and specific biases in the estimation of population parameters, and the combination of these errors and biases has an impact on the precision of the collected data.

One of these errors is unit-nonresponse, which refers to the discrepancy between the sample and the set of respondents, which is successfully observed [2]. It arises when individuals either cannot be contacted or choose not to participate in the survey, thus it can be related to the response rate. There is a consensus among survey researchers [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 3] that response rates are declining (i.e. nonresponse rates are increasing). Nonresponse can seriously distort the results and even lead to incorrect conclusions. Previous field experiences and the analysis of current survey meta-data indicate that the overall increase in survey nonresponse does not equally apply to different population subgroups [13, 15]. It has been found that single-person households, renters and individuals outside of the labour force are less likely to participate in surveys than members of other social groups [16, 13].

Measurement error is another highly relevant source of errors, which refers to the discrepancy between the ideal measurement and the actual responses obtained [2]. Measurement errors are even present in censuses, where theoretically the entire population is measured. It encompasses various factors, including interviewer effects, systematic errors, and random errors [17]. Measurement error occurs when the recorded or observed value deviates from the true value of the variable [18]. Several factors contribute to this disparity, such as the unclear or misleading phrasing of the questions and the context of the preceding questions [19]. Other important factors include changes in the mental state of the respondent, inconsistencies in their answers, social desirability, and the concealment of the true answer. For example, respondents may provide an answer that aligns with the perceived norm rather than their actual opinion due to socially desirable behavior or yea-saying [20].

Results

First, the thesis proposes a new sampling allocation method to handle unit-nonresponse. This new method allocates sample sizes in stratified sampling designs based on expected response rates (ERRs). To be able to determine the exact proportions during the allocation procedure, estimates regarding response rates from previous surveys are needed. In the case of unit-nonresponse, the contact data (or survey meta-data) are typically not available publicly, but survey organizations can use their own historical data.

Generally, the proportional-to-stratum size (PS) allocation method [21] is used, thus the ERRs allocation is assessed relative to the PS allocation procedure.

In the case of PS allocation, let N denote the population size and let N_h $(h = 1, 2, \ldots, H)$, be the sizes of the strata relevant to the stratified sampling procedure, with $N = N_1 + ... + N_H$. In a stratified random sample, a simple random sample of n_h elements is taken from each stratum h $(h = 1, 2, ..., H)$, with a total sample size of n elements.

When the survey aims to collect m responses, the response rate which characterizes the population needs to be taken into account in deciding about the attempted sample size. Of course, such decisions should be made based on the true response rate, but it is rarely known. Thus, the ERR, say r, is used which is based on former experience. Then, a total of $n = m/r$ observations are allocated.

In the case of allocation proportional to size (PS), let n_h^{PS} $(h = 1, 2, ..., H)$ denote the subsample size within stratum h. The sampling fraction n_h^{PS}/N_h is specified to be the same for each stratum and thus

$$n_h^{PS} = \frac{1}{r}\frac{N_h}{N}m \qquad h\text{=}1,\ldots,H, \tag{1}$$

which implies that the overall sampling fraction n/N is the same as the fraction taken from each stratum. The total allocated sample size is then as follows:

$$n^{PS} = m\sum_{h=1}^{H}\frac{N_h}{N}\frac{1}{r} = \frac{m}{r} \tag{2}$$

In the case of ERR allocation, let n_h^{ERR} $(h = 1, 2, ..., H)$ denote the allocated subsample size within stratum h. Let r_h $(h = 1, 2, ..., H)$ denote the stratum-specific ERRs, which are also assumed to be population parameters. Clearly,

$$r = \sum_{h=1}^{H} \frac{r_h N_h}{N}.$$

In ERR allocation, the allocated sample size in each stratum n_h^{ERR} is specified using, instead of the population level ERR, the stratum-specific ERRs. The allocated sample size in each stratum is

$$n_h^{ERR} = \frac{1}{r_h} \frac{N_h}{N} m \qquad h=1,...,H. \tag{3}$$

Consequently, the total allocated sample size is

$$n^{ERR} = m \sum_{h=1}^{H} \frac{N_h}{N} \frac{1}{r_h}. \tag{4}$$

Theorem 1 (Variance of the estimates)**.** *Let the population size be N, and let the population be divided into H strata of respective sizes of N_h, $(h = 1, .., H)$. Let m be the intended total sample size, r the ERR in the entire population and r_h the respective ERRs in the strata. The true population proportion of those possessing the characteristics of interest is denoted by q_h, which is the parameter to be estimated in each stratum h. Finally, let p_h be the true response rate in stratum h. Then, the asymptotic variances of the estimates obtained from samples based on PS and ERRs allocations, with post-stratification applied, are as follows.*

$$V^{PS}(\hat{q}) = \frac{1}{Nm} \sum_{h=1}^{H} N_h q_h (1 - q_h) \frac{r}{p_h} \tag{5}$$

$$V^{ERR}(\hat{q}) = \frac{1}{Nm} \sum_{h=1}^{H} N_h q_h (1 - q_h) \frac{r_h}{p_h} \tag{6}$$

Theorem 2 (Relationships among the variances)**.** *Let $\hat{V}^{PS}(\hat{q})$ be the total variance of the estimates based on a sample drawn via the PS allocation given in (5), and let $\hat{V}^{ERR}(\hat{q})$ be the total variance of the estimates based on a sample drawn by the allocation based on different ERRs, as given in (6). If the observed response rates are equal to the ERRs, then,*

$$\hat{V}^{ERR}(\hat{q}) \leq \hat{V}^{PS}(\hat{q}) \tag{7}$$

We compared the ERRs and PS allocation methods under misspecification that is, when the true response rates differ from the ERRs used in the sample allocation ($p_h \neq r_h$). The variances were compared for all combinations of parameter values with a fixed number of strata, $H = 3$. Specifically, all possible combinations of the following parameter values were considered: all possible combinations of the values $\{0.1, 0.3, 0.5, 0.7, 0.9\}$ for the true response rates $\{p_1, p_2, p_3\}$ and for the ERRs $\{r_1, r_2, r_3\}$. The parameter to be estimated in every stratum h $(h = 1, 2, 3)$ was given values between 0 and 1, with an increment of 0.05. The size of the population $N = 10^7$,the sizes of the strata $N_1 = 2{*}10^6$, $N_2 = 3{*}10^6$, $N_3 = 5{*}10^6$, and the desired total sample size $m = 1000$ were fixed. With the different choices, a total of 15.625.000 different sets of parameters were defined.

Figure 1 shows the comparison of the variances of the estimates obtained using ERRs and PS allocations. The comparison is given in terms of the total absolute misspecification of the response rates, $\sum_{h=1}^{H} |r_h - p_h|$ *(x-axis)* and of the total absolute distance of the ERRs $\{r_1, r_2, r_3\}$ from their weighted average, $\sum_{h=1}^{H} |r_h - r|$ *(y-axis).*

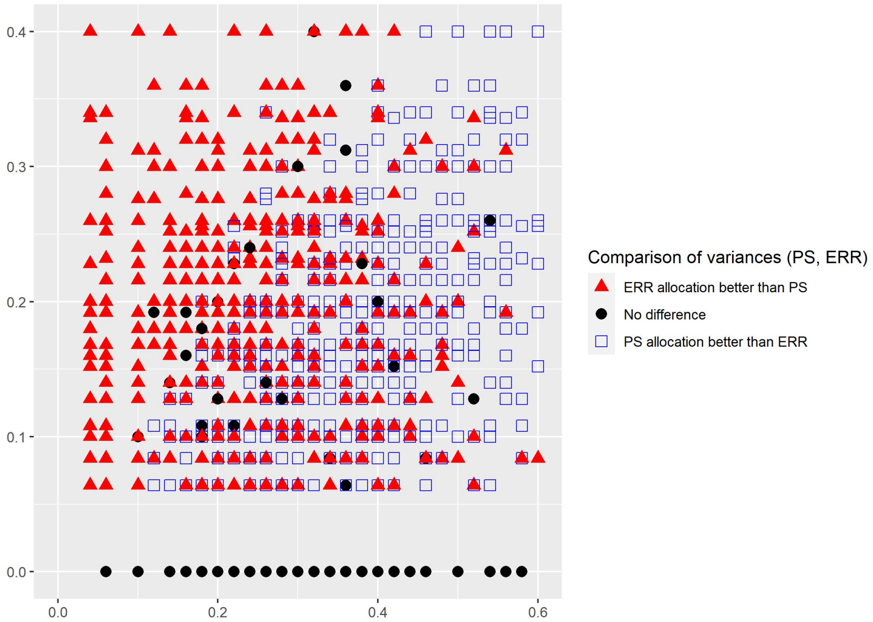

Figure 1: Comparison of the variances in the estimates obtained using ERR and PS allocations, in terms of the total absolute misspecification of the response rates *(x-axis:* $\sum_{h=1}^{H}|r_h - p_h|$*)* and the total absolute distance of the ERRs from one weighted average *(y-axis:* $\sum_{h=1}^{H}|r_h - r|$*)*.
Source: Own figure.

The results regarding the new allocation mechanism are as follows:

- The magnitude of the misspecification of the response rates appeared to have a greater impact on the relative performances of the two allocation procedures.
- When the total absolute misspecfication was less than 0.3, the ERR allocation almost always performed better.
- Meanwhile, the total absolute distance of the ERRs from their weighted average appears to have had a small and non-systematic effect.
- When the total absolute misspecification of the response rates was lower than 0.3, the ERR allocation yielded mostly smaller variances.
- Meanwhile, in the range of $0.3 - 0.4$, the two allocations performed equally well.
- Most notably, an equal precision can be expected in the extreme areas of the plot.
- When the difference between the total absolute deviations of the expected rates and the ERRs was less than approximately half of the latter, the ERR allocation always performed better, irrespective of whether or not the individual response rates were correctly predicted.

Second, the thesis introduce a new scheme for assessing survey uncertainties with a special focus on measurement problems. Our approach models the precision of the values found in a survey compared to a potential replication of the survey. We define nonresponse uncertainty (NU) and measurement uncertainty (MU), which refer to the sources of difference between two replications of surveys and can be linked to nonresponse and measurement error in the total survey error framework. Unlike general methods that assess the reliability and validity of a given question, this new scheme assesses the uncertainties of the survey as a whole. Instead of an illusion of a true population parameter, our method addresses the issue of survey quality through replication surveys.

In this chapter, we formally show the decomposition of the total difference of the answers from two replications of a survey. We consider the cases of continuous variables by decomposing the mean, and correlation coefficient, and discrete variables by decomposing the relative frequency of the i^{th} category, and χ^2-test statistics for independence. In the following, we present NU and MU in the joint attempted sample of potential first and second replications of a survey. The following notations are used for different groups of answers (Figure 2): set A denotes the answers of the total completed sample of the first replication of a survey, set B concerns the answers of the total completed sample of the second replication of a survey, set C denotes the answers of the group of those who responded only to the first replication of a survey, and set D denotes the answers of those who responded only to the second replication of a survey. Sets E and F refer to answers from those, who responded to both replications. Set E concerns the answers of the first survey, and set F concerns the answers of the second survey.

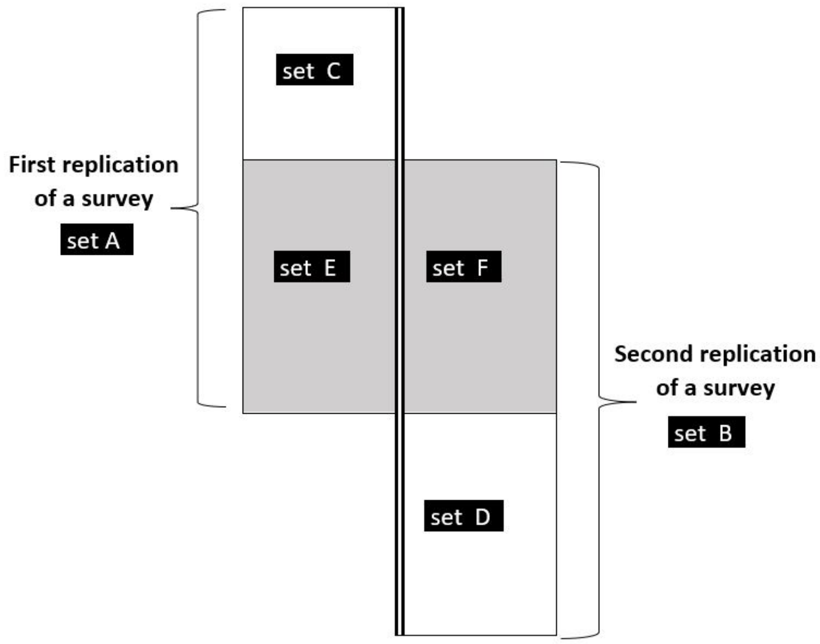

Figure 2: Different groups of answers in replication surveys

Source: Own figure.

NU depends on whether individuals who respond to the first replication of a survey give different answers from those who respond to the second replication of a survey, and thus NU captures the uncertainty of the base of responders/non-responders. Since there are respondents who did not participate in either replication of a survey and are therefore not included in this analysis, a particular kind of NU is studied: only in relation to the two replications of the survey. MU is the difference between the answers of the two replications of a survey from the same respondents. MU is defined both at the individual/respondent level and at the sample level. If MU occurs, there is an observation gap between the answers obtained in the first replication of a survey and the answers obtained in the second replication of a survey. Theorems 3 - 6 includes the decompositions of the mean, correlation coefficient, relative frequency of the i^{th} category of a discrete variable, and χ^2-test statistics for independence between two discrete variables.

Theorem 3 (Decomposition of the total difference of the mean)**.** *Let X denote a variable measured in both replications and $\bar{X}_A, \bar{X}_B, \bar{X}_C, \bar{X}_D, \bar{X}_E, \bar{X}_F$ be the mean of X in sets $A-F$ respectively. Let m_A, m_B, m_C, m_D, m_E, and m_F denote the sample sizes for each set, respectively. The decomposition of the difference of $\bar{X}$ between the first and second replications of a survey can be written as the weighted average of the differences between sets C and D and the weighted average of the differences between sets E and F. The decomposition of the difference is:*

$$\bar{X}_A - \bar{X}_B = \frac{m_C + m_D}{m_A + m_B}\left(\bar{X}_C - \bar{X}_D\right) + \frac{m_E + m_F}{m_A + m_B}\left(\bar{X}_E - \bar{X}_F\right), \tag{8}$$

where $\bar{X}_C - \bar{X}_D$ is the NU and $\bar{X}_E - \bar{X}_F$ is the MU.

Theorem 4 (Decomposition of the total difference of the correlation coefficient)**.** *Let X and Y denote two variables measured in both replications and let $r(X,Y)_A$, $r(X,Y)_B$, $r(X,Y)_C$, $r(X,Y)_D$, $r(X,Y)_E$, $r(X,Y)_F$ be the correlation coefficients of X and Y in sets A–F, respectively. The decomposition of the difference of $r(X,Y)$ between the first and the second replications is obtained as the weighted average of the Fisher's z-scores [22] of the differences between set C and set D and the difference between sets E and F. The Fisher's z-transformed correlation coefficients are denoted by $r'(X,Y)_A$, $r'(X,Y)_B$, $r'(X,Y)_C$, $r'(X,Y)_D$, $r'(X,Y)_E$, $r'(X,Y)_F$ in sets A–F, respectively. Following the standard Fisher's z-score method in Alexander (1990), the decomposition of the difference is:*

$$\begin{aligned} r(X,Y)_A - r(X,Y)_B =& \frac{m_C + m_D}{m_A + m_B}\left(r'(X,Y)_C - r'(X,Y)_D\right) \\ &+ \frac{m_E + m_F}{m_A + m_B}\left(r'(X,Y)_E - r'(X,Y)_F\right), \end{aligned} \tag{9}$$

where $r'(X,Y)_C - r'(X,Y)_D$ is the NU and $r'(X,Y)_E - r'(X,Y)_F$ is the MU.

Theorem 5 (Decomposition of the total difference in the relative frequency of the i^{th} category)**.** *Let g_{iA}, g_{iB} denote the relative frequencies in the two replications and let ν_{iA}, ν_{iB}, ν_{iC}, ν_{iD}, ν_{iE}, ν_{iF}, the number of cases of category i. The decomposition of the difference in the relative frequency of a given category between the first and second replications is obtained as the weighted average of the differences between set C and set D and the difference between sets E and F. The decomposition of the difference is:*

$$g_{iA} - g_{iB} = \frac{m_C + m_D}{m_A + m_B}\left(\frac{\nu_{iC}}{m_C} - \frac{\nu_{iD}}{m_D}\right) + \frac{m_E + m_F}{m_A + m_B}\left(\frac{\nu_{iE}}{m_E} - \frac{\nu_{iF}}{m_F}\right), \tag{10}$$

where $\left(\frac{\nu_{iC}}{m_C} - \frac{\nu_{iD}}{m_D}\right)$ is the NU and $\left(\frac{\nu_{iE}}{m_E} - \frac{\nu_{iF}}{m_F}\right)$ is the MU.

Theorem 6 (Decomposition of the total difference in the χ^2-test statistics for the independence)**.** *Let X and Y denote two variables measured in both replications, let r denote the number of response categories of variable X and let ν denote the number of response categories of variable Y. The observed frequencies of each cell (ij) in sets $A - F$ are denoted with O_{ij_A}, O_{ij_B}, O_{ij_C}, O_{ij_D}, O_{ij_E}, O_{ij_F}, respectively and the expected frequencies of each ij cell are denoted with E_{ij_A}, E_{ij_B}, E_{ij_C}, E_{ij_D}, E_{ij_E}, E_{ij_F} for all sets respectively. If identical marginal distributions are assumed for X and Y, between sets C and E and sets D and F, the decomposition of the difference is:*

$$\begin{aligned} {\chi^2}_A - {\chi^2}_B =& \sum_{i=1}^{r}\sum_{j=1}^{\nu}\Bigg(\left[\frac{1}{E_{ij_C}}\left(O_{ij_C}^2 - 2O_{ij_C}\right) - \frac{1}{E_{ij_D}}\left(O_{ij_D}^2 - 2O_{ij_D}\right)\right] + \\ &\left[\frac{1}{E_{ij_C}}\left(O_{ij_E}^2 - 2O_{ij_E}\right) - \frac{1}{E_{ij_D}}\left(O_{ij_F}^2 - 2O_{ij_F}\right)\right] + \\ &\left[\frac{1}{E_{ij_C}}\left(2O_{ij_C}O_{ij_E}\right) - \frac{1}{E_{ij_D}}\left(2O_{ij_D}O_{ij_F}\right)\right] + \\ &\left[\frac{1}{E_{ij_E}}\left(O_{ij_C}^2 - 2O_{ij_C}\right) - \frac{1}{E_{ij_F}}\left(O_{ij_D}^2 - 2O_{ij_D}\right)\right] + \\ &\left[\frac{1}{E_{ij_E}}\left(O_{ij_E}^2 - 2O_{ij_E}\right) - \frac{1}{E_{ij_F}}\left(O_{ij_F}^2 - 2O_{ij_F}\right)\right] + \\ &\left[\frac{1}{E_{ij_E}}\left(2O_{ij_C}O_{ij_E}\right) - \frac{1}{E_{ij_F}}\left(2O_{ij_D}O_{ij_F}\right)\right]\Bigg) \end{aligned} \tag{11}$$

If identical marginal distributions are not assumed, the expected frequencies of set $C + E$ and set $D + F$ cannot be given as a sum of the expected frequencies of the separate sets, but can be written as $E_{ij_C} + E_{ij_E} - \left(\frac{O_{i\cdot C}}{m_C} - \frac{O_{i\cdot C} + O_{i\cdot E}}{m_C + m_E}\right)$ and $E_{ij_D} + E_{ij_F} - \left(\frac{O_{i\cdot D}}{m_D} - \frac{O_{i\cdot D} + O_{i\cdot F}}{m_D + m_F}\right)$. In this case, the decomposition of the difference of the χ^2 test statistics becomes more complex, which will not be discussed further in this chapter.

It can be seen, that the total difference between the responses obtained in two replications of a survey can be decomposed exclusively into NU and MU in the case of continuous variables regarding the mean, and correlation coefficients, in the case of discrete variables regarding relative frequency and the χ^2-test statistics for independence. This means that if a survey is repeated, the total difference is due to a change in the respondent base and to the different answers of those who respond to both surveys.

The uncertainties are defined by comparing the answers of the first replication of the ESS and the second replication of the ESS. In the following, we present NU and MU in the common attempted sample of the first and second replications of the ESS ($n = 3,000$). Figure 3 presents the different sub-sets and their sample sizes in the replication of the ESS. Following the notations presented previously the sets are as follows: set A denotes the answers of the total completed sample of the first replication of the ESS, set B concerns the answers of the total completed sample of the second replication of the ESS, set C denotes the answers of the group of those who responded only to the first replication of the ESS, and set D denotes the answers of those who responded only to the second replication of the ESS. Sets E and F refer to answers from those, who responded to both replications. Set E concerns the answers of the first replication of the ESS, and set F concerns the answers of the second replication of the ESS. Compared to Figure 2, Figure 3 is supplemented with the unsuccessful addresses in both replications of the ESS.

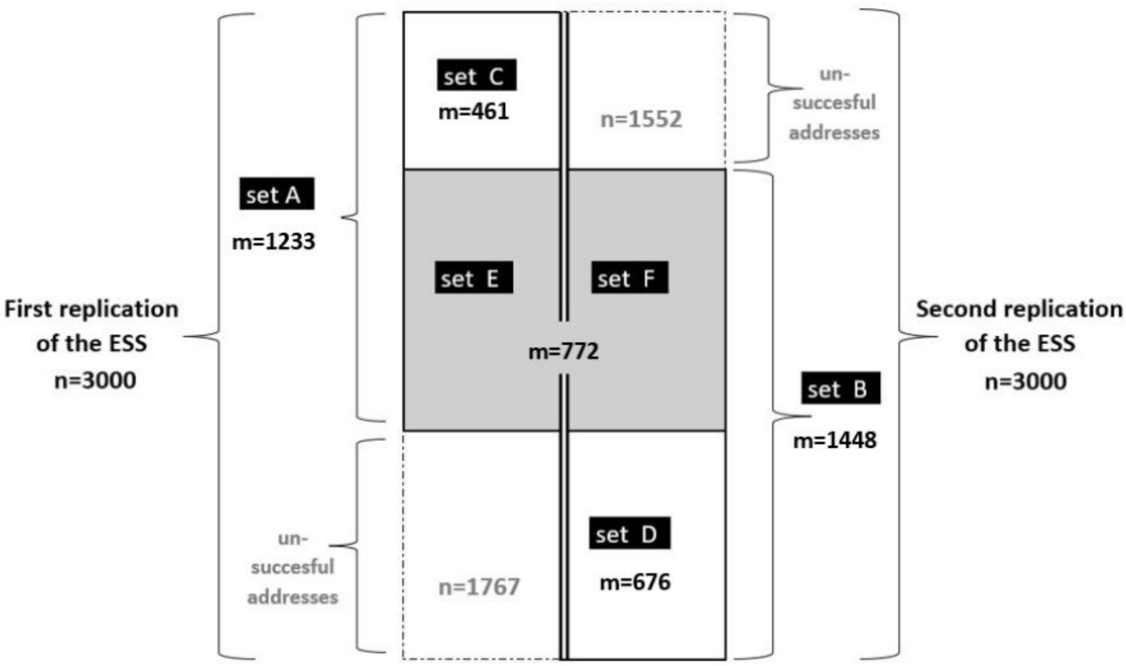

Figure 3: Different groups of answers and sample sizes in the replications of the ESS

Note: The columns on either side of the figure represent the two replications of the ESS in the same structure in which the different groups of answers of replication surveys in general are presented in Figure 2

The thesis presents the comparison of four selected variables: level of education of the respondent, level of education of the respondent's mother, general trust, and religiousness At the sample level, we find that between sets A and B and between C and D there are minor differences.

In the following, the difference is presented at the individual level. Even if there was a short time between the two replications of the ESS, it is hard to exclude the real changes in the answers. Out of the four variables under consideration, there are two factual variables for which the probability of a real change is very low: the level of education and the mother's level of education. Since 20 months of time passed between the two replications of the survey, it is unlikely that a respondent's or the mother's highest level of education would increase by one category (a category covers an average of 4 years). Moreover, it is logically impossible for respondents or their mothers to have a lower level of education a year and a half later.

Figure 4 presents the differences between the first and second answers relative to the answers from the first replication of the ESS. If there were a real change in answers, the distributions of the difference would be skewed towards the high values (positive changes in the answers). It can be seen that for each variable the distributions of the difference are symmetric. The figure also shows that reporting of the unlikely or logically impossible changes in answers to education questions is common not only for the full sample but also for the sub-sample of respondents over 45 years of age (yellow charts), for whom a change in the highest level of education, and mother's highest level of education is even less likely due to advancing age. This underlines the fact that, although a real change cannot be completely excluded, it can rather be said that the difference is mainly due to MU. This uncertainty measured for factual variables is assumed to be present as large or even larger volumes in the case of the attitude variables. The figure also shows that for each variable, the mean of the difference (μ) is around 0.

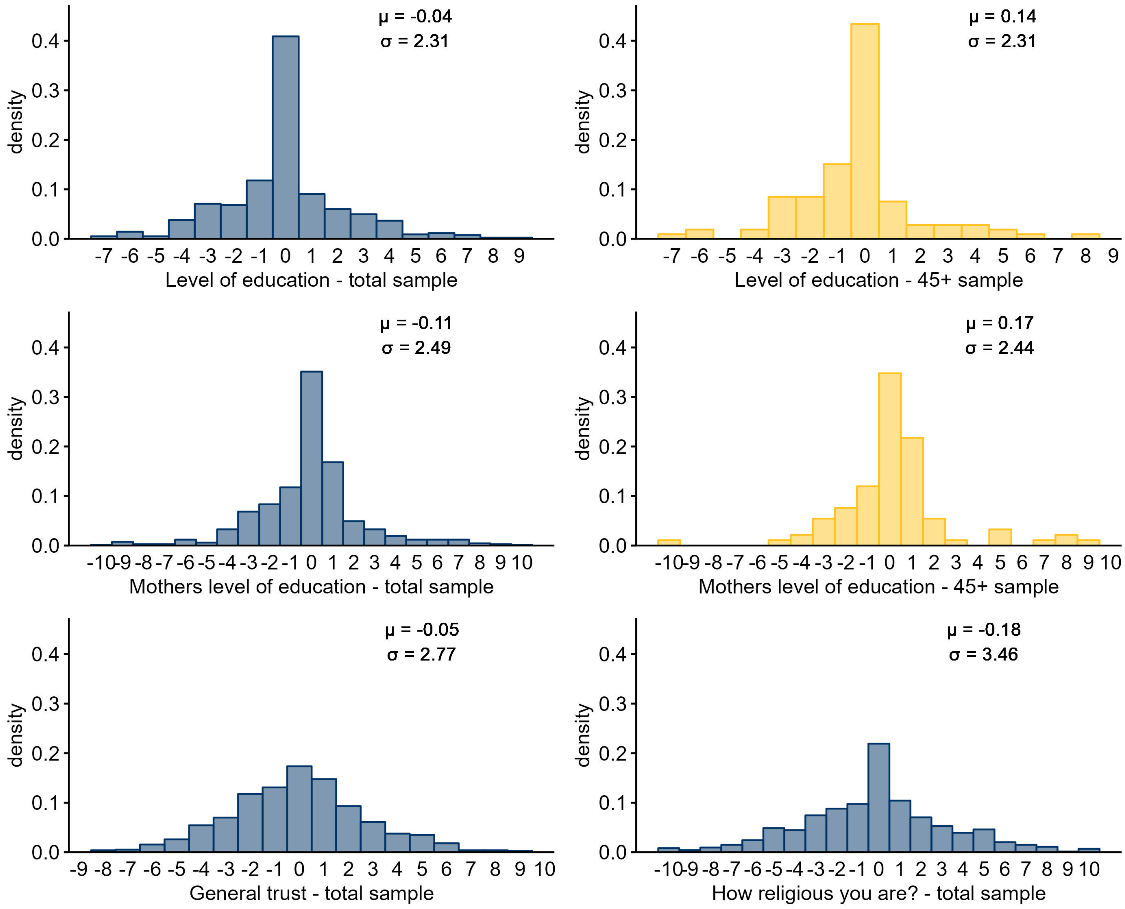

Figure 4: Difference between first and second answers (set E − set F)

Source: Own figure.

In the case of general trust and religiousness, the RTM phenomena is presented from a different perspective (Figure 5 and Figure 6). In the case of general trust, the mean is 4.54 and the mode is value 5. In observing the RTM phenomenon the mean as the reference point to which the observations may regress toward is considered to be 5. In Figure 5, the values on the x-axis represents the first answers' absolute difference from the mean value (0 represents value 5 as answer) and the values on the y-axis represents the second answers' absolute difference from the mean value (0 represents value 5 as answer). The size of the bubbles represent the proportion of the given pattern. It can be seen, that in the case of answers with a greater difference relative to the mean value (values 3, 4, 5 on the x-axis) there is a higher share of those regression back towrd the mean in the case of their second answers. In the case of religiousness, the mean is 4.02 and the mode is value 0. This distribution is skewed to the right, which make the picture concerning RTM more complex. However, in observing the RTM phenomenon the mean as the reference point to which the observations may regress toward is considered to be 4. In Figure 6, the values on the

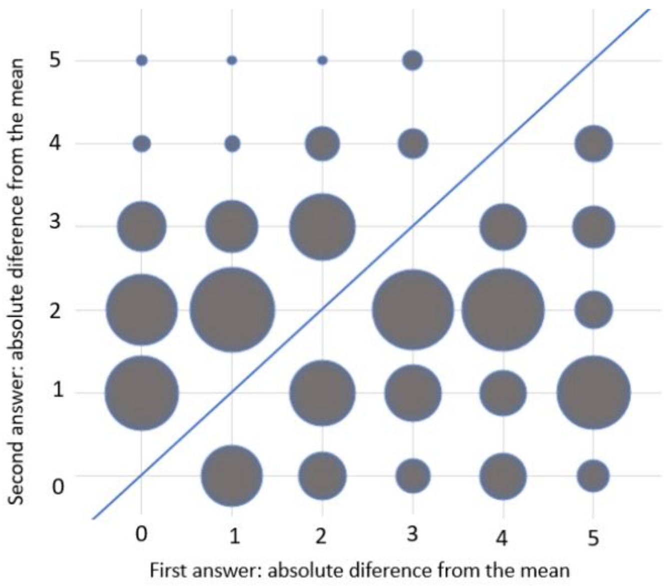

Figure 5: General trust: Responses relative to the mean values

Source: Own figure.

x-axis represents the first answers' absolute difference from the mean value (0 represents value 4 as answer) and the values on the y-axis represents the second answers' absolute difference from the mean value (0 represents value 4 as answer). The size of the bubbles represent the proportion of the given pattern. It can be seen, that in the case of answers with a greater difference relative to the mean value (values 4, 5, 6 on the x-axis) there is a higher share of those regression back towrd the mean in the case of their second answers.

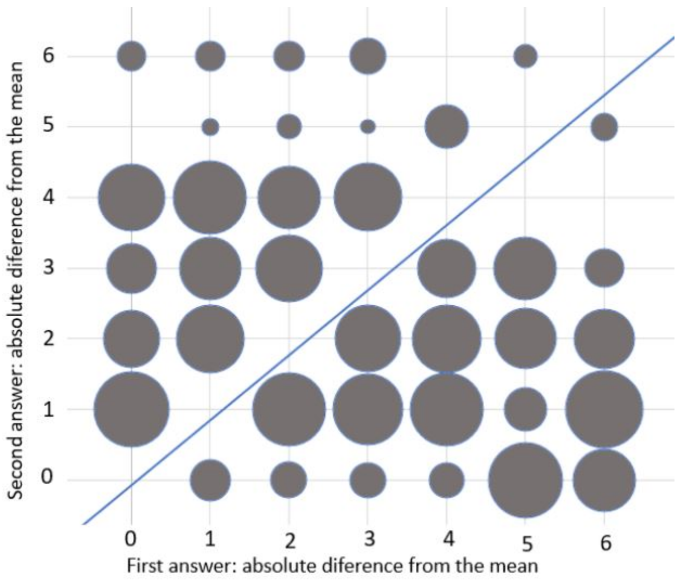

Figure 6: How religiousness you are? - Responses relative to the mean values

Source: Own figure.

The results regarding the new scheme for survey assessment are as follows:

- The total difference between two survey replications is the sum of NU and MU; therefore, the total difference was the combination of uncertainty about the respondent bases and uncertainty about the answers obtained.

- We found that for the univariate analysis, NU was negligible but relevant for the multivariate analysis.
- For MU, we compared the answers of those who responded to both surveys. The second finding of the study was that although the first and second answers generally resulted in the same distribution, on an individual basis, respondents appeared to be inconsistent with their answers.
- This phenomenon was explained with RTM, which occurs because values are observed with random errors.
- The third finding of the study was that, in multivariate analysis, both NU and MU are relevant, but their joint impact cause minor differences at the total sample level.

Összefoglalás

A matematikai statisztika általános elmélete végtelen alapsokaságot feltételez. Ezzel szemben a survey statisztika a véges sokaságból történő mintavétel alapjaira épül. A lakossági survey-ek esetében a mintavétel egységei emberek, ami humán tényezőket von be a kísérletbe. Az emberi természet és viselkedés nehézségeket okoz a lakossági mintavétel matematikai elméletének megvalósításában, és különböző módon befolyásolja az adatok minőségét. A survey adatok minősége az elmúlt évtizedben egyértelműen romlott, ami indokolja a jelenlegi módszertanok újraértékelését.

A dolgozat azzal járul hozzá a meglévő szakirodalomhoz, hogy újszerű megközelítéseket javasol a becslések minőségének javítására. A dolgozat a következő emberi tényezőkből eredő hiábak kezeli: (1) amikor a mintába került egyének úgy döntenek, visszautasítják a válaszadást, azaz a megfigyelést, ami bizonytalanságot eredményez a minta összetételét illetően, és (2) amikor az emberek inkonzistensen válaszolnak az egyes kérdésekre, ami bizonytalanságot eredményez a mérésben.

A minta összetételét illetően a dolgozat egy új minta-allokációs módszert vezet be, amely figyelembe veszi a várható válaszadási arányokat (ERRs). Az új módszer értékeléséhez a fő elméleti eszközt a δ-módszerrel végzett aszimptotikus számítások jelentik. Egy rétegzett mintavételi design esetében az ERR-allokáció alacsonyabb szórást eredményez, mint a hagyományos allokációs módszer, nem csak akkor, ha a válaszadási arányok helyesen vannak megadva, hanem a feltételek egy széles skálája mentén is. Az új mintavételi módszert szimulációkkal is szemléltetjük, amelyekben a konkrét populációs paraméterek különböző kombinációit használjuk.

A mérések bizonytalanságát illetően az értekezés egy új, az ismételt mérések logikájára épülő szempontrendszert mutat be. Ez a szempontrendszer abban tér el a klasszikus logikától, hogy a valódi populációs érték helyett a survey mérések eredményét ahhoz viszonyítja, hogy milyen eredményt kapnánk akkor, ha az adott mérést teljes egészében megismételnénk. Eredményeink szerint szerint két mérés különbsége felbontható a részvételből származó bizonytalanság (nonresponse uncertainty, NU) és a válaszok bizonytalanságának (measurement uncertainty, MU) összegére. Az új szemléletet egy esettanulmány, a European Social Survey (ESS) megismétlése alapján is bemutattuk. Azt találtuk, hogy az MU, a válaszadók szintjén rendkívül fontos, mert a válaszadók inkonzisztensnek tűnnek a válaszakban. Ezt az inkonzisztenciát a regression to the mean (RTM) jelenségeként mutatjuk be. Az eredmények azt mutatják, hogy ez a jelenség ordinális skálájú változók esetében is releváns.

A dolgozat számos olyan eredményt is közölt, amely indokolttá teszi a téma további kutatását. A szerző további tervei között szerepel: a megismételt survey keretrendszer vizsgálata kettőnél több ismétlés esetére, illetve a mérési bizonytalanság vizsgálata válaszok sorozatán; többváltozós modellek építése a mérési hibák elemzésére és ennek alapján új adatkorrekciós (utólagos rétegzési) eljárások fejlesztése; az adatgyűjtési mód bevonása a vizsgálatokba (leginkább a személyes, a telefonos és az online adatgyűjtések figyelembe vétele); valamint a bizonytalanságok modellezése az entrópia fogalmával.

Publications of the author

The thesis is based on the following three published papers:

1. B. Szeitl and T. Rudas (2022): Reducing Variance with Sample Allocation Based on Expected Response Rates in Stratified Sample Designs, Journal of Survey Statistics and Methodology, Volume 10, Issue 4, September 2022, Pages 1107–1120, https://doi.org/10.1093/jssam/smab021

2. B. Szeitl and T. Rudas (2024): Assessing survey quality with a replication survey: non-response uncertainty and measurement uncertainty in the ESS, Methods, data, analyses (MDA)

3. Messing, V., Ságvári, B., Szeitl, B. (2022): Is "push-to-web" an alternative to face-to-face survey?: Experiences from a "push-to-web" hybrid survey in Hungary. (In Hungarian) STATISZTIKAI SZEMLE (0039-0690): 100/3 pp 213-233 (2022)

Further publications of the author:

- Buda, J., Hajdu, G., Szeitl, B., Janky, B. (2023). A new method for the imputation of key indicators based on separate high-quality survey data. INTERNATIONAL JOURNAL OF PUBLIC OPINION RESEARCH (accepted)

- Messing, V., Ságvári, B., Szeitl, B. (2023): Respondings as expected? The effects of survey mode on estimates of sensitive attitudes in self-completion and face-to-face interviews of the European Social Survey. SURVEY RESEARCH METHODS (under review)

- Szeitl, B., Tóth, I. (2021): Revisiting the ESS R8 sample a year after – Lessons from a re-contact survey to test patterns of unit non-response in Hungary. Survey Methods: Insights from the Field. https://surveyinsights.org/?p=14864

- Simonovits, G., Kates, S., Szeitl, B. (2019). Local Economic Shocks and National Election Outcomes: Evidence from Hungarian Administrative Data. POLITICAL BEHAVIOR 41(2), 337–348.

Bibliography

[1] Wayne A. Fuller. *Sampling statistics.* John Wiley Sons, Inc., Hoboken, New Jersey, 2009.

[2] Robert Groves, Floyd J. Fowler, Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. *Survey Methodology*, volume 2. John Wiley Sons, 01 2009.

[3] Loosveldt G. Vandenplas C. Stoop I. Beullens, K. Response rates in the european social survey: Increasing, decreasing, or a matter of fieldwork efforts? *Survey Methods: Insights from the Field.*, 2018.

[4] Bates N. Burt G. Silberstein A. Atrostic, B. K. Nonresponse in u.s. government household surveys: Consistent measures, recent trends, and new insights. *Journal of Official Statistics*, 17(2):209–226, 1970.

[5] De Heer W. De Leeuw, E. D. *R. M. Groves, D. A. Dillman, E. Eltinge, R. J. A. Little (Eds.), Survey Nonresponse*, chapter Trends in Households Survey Nonresponse: A Longitudinal and International Comparison. New York: Wiley, 2002.

[6] Murtaugh M. A. Edwards S. Slattery M. L. Rogers, A. Contacting controls: Are we working harder for similar response rates, and does it make a difference? *American Journal of Epidemiology*, 160(1):85–90, 2004.

[7] Presser S. Singer E. Curtin, R. Changes in telephone survey nonresponse over the past quarter century. *Public Opinion Quarterly*, 69(1):87–98, 2005.

[8] E. Singer. Nonresponse bias in household surveys. *Public Opinion Quarterly*, 70(5):637–645, 2006.

[9] Handbook of Survey Research. *Survey Nonresponse. In P. Marsden J. D. Wright (Eds.).* Bingley: Emerald Group Publishing Limited, 2010.

[10] Cobben F. Schouten B. Bethlehem, J. G. *Handbook of Nonresponse in Household Surveys.* John Wiley, 2011.

[11] Williams D. Brick, J. M. Explaining rising nonresponse rates in cross-sectional surveys. *The ANNALS of the American Academy of Political and Social Science*, 645(1):36–59, 2013.

[12] F. Kreuter. Facing the nonresponse challenge. *The Annals of the American Academy of Political and Social Science*, 645(1):23–35, 2013.

[13] Bruce Meyer, Wallace Mok, and James Sullivan. Household surveys in crisis. *Journal of Economic Perspectives*, 29:199–226, 11 2015.

[14] Brick J. M. Williams, D. Trends in u.s. face-to-face household survey nonresponse and level of effort. *Journal of Survey Statistics and Methodology*, 9(1):1–26, 2017.

[15] Guillaume Osier. Unit non-response in household wealth surveys. Statistics Paper Series 15, European Central Bank, 2016.

[16] Katharine Abraham, Aaron Maitland, and Suzanne Bianchi. Non-response in the american time use survey: Who is missing from the data and how much does it matter? *Public Opinion Quarterly*, 70:676–703, 10 2006.

[17] D.T. Oberski. *Measurement errors in comparative surveys*. PhD thesis, University of Tilburg, 2011.

[18] Duane F. Alwin. *Margins of Error: A Study of Reliability in Survey Measurement*. New York: Wiley, 2007.

[19] W. Saris and I. Gallhofer. Estimation of the effects of measurement characteristics on the quality of survey questions. *Survey Research Methods*, 1:29–43, 2007.

[20] L. J. Rips Tourangeau, R. and K. Rasinski. *The Psychology of Survey Responses*. New York: Cambridge University Press, 2000.

[21] Michael D. Larsen. Proportional allocation to strata. In P. J. Lavrakas, editor, *Encyclopedia of Survey Research Methods*, pages 630–630. SAGE Publications, Inc., 2008.

[22] Alexander E. A. A note on averaging correlations. *Bulletin of the Psychonomic Society*, 28(4):335–336, 1990.