STATISTICS IN TRANSITION new series, June 2023
Vol. 24, No. 3, pp. 139–149, https://doi.org/10.59170/stattrans-2023-040
Received – 10.03.2022; accepted – 16.02.2023

Models for survey nonresponse and bias adjustment techniques

H. Öztaş Ayhan[1]

Abstract

Survey statisticians have been dealing with the issues of nonresponse in sample surveys for many years. Due to the complex nature of the mechanism, so far it has not been easy to find a general solution to this problem. In this paper, several aspects of this topic will be elaborated on: the survey unit nonresponse bias has been examined alternatively by taking response amounts which are fixed initially and also by taking the response amounts as random variables. An overview of the components of the bias due to nonresponse will be performed. Nonresponse bias components are illustrated for each alternative approach and the amount of bias was computed for each case.

Key words: response rate, nonresponse bias, nonresponse rate, sample survey, survey error.

1. Introduction

During many past studies, the evaluation of the nonresponse bias was based on presenting the nonresponse error in the form of nonresponse rates. However, nonresponse bias may not be related to the response or nonresponse rates of a given study. Increasing response rate (decreasing nonresponse rate) may not always correspond to decreasing nonresponse bias for a given study (Groves and Couper, 1998).

On the other hand, in many studies in the past, the term "bias" was interpreted differently from, how we evaluate the "statistical bias". The arguments have gone even to far to suggest that, the bias can be obtained within the available inside information on a given sample. However, there are few studies that mentioned statistical bias, which is based on the differences between the "expected value of all possible sample estimates" from the "corresponding parameter" Bethlehem and Kersten (1985), Groves *et al.* (2002), Keser (2011), Kish (1995), Moser and Kalton (1979), Lindström *et al.* (1979), and Lindström (1983). Since, we can only afford to select one sample for a given study,

[1] Professor Emeritus, Department of Statistics, METU, Ankara, Turkey. E-mail: oayhan@metu.edu.tr. ORCID: https://orcid.org/0000-0003-3818-483X.

in this case "the parameter" and "other sample estimates" will also be unknown. Furthermore, by using certain rules during the field operation, the amount of nonresponse can be determined after the fieldwork. Consequently, the amount of nonresponse is only fixed when the field operation is completed.

Alternatively, the amount of nonresponse is unknown before the fieldwork and therefore initially it can be evaluated as a random variable.

The objective of this research is to formulate basic computation of response and nonresponse models. The study also aims to present and discuss alternative response/ nonresponse models (fixed response model and random response model). In addition to these, it is aimed to present and compare the alternative bias adjustment techniques for diffferent models.

The impact of nonresponse on the estimators have been examined under two alternative approaches. These are the "*fixed response model*" and the "*random response model*" (Lindström *et al.* 1979). Most of the past research is based on assuming that the response and nonresponse amounts are fixed before the survey. Therefore, these studies have used the following nonresponse bias evaluation.

2. Taking response amounts which are fixed initially

The fixed response model assumes the population to consist of two mutually exclusive and exhaustive strata: the response stratum and the nonresponse stratum. If selected in the sample, elements in the response stratum will participate in the survey with certainty and elements in the nonresponse stratum will not participate with certainty (Bethlehem, 2009).

The population size of *N* can artificially be divided into response and nonresponse stratum. We can use the following form ($R_i = N_i / N$) of the rate and the size ($N_i = \sum_{j=1}^{J} N_{ij}$) to illustrate the mechanism, where *i = 1, 2.*

Response rate: $R_1 = N_1 / N$ and Nonresponse rate: $R_2 = N_2 / N$ (1)

$$N_1 + N_2 = N, \quad R_1 + R_2 = 1, \quad (1 - R_1) = R_2 \tag{2}$$

The survey data will only be collected for the response strata. The response strata will have the mean μ_1 which is based on the N_1 observations.

Response stratum mean will be, $\mu_1 = N_1^{-1} \left[\sum_{j=1}^{N_1} X_{1j} \right]$. (3)

Nonresponse stratum mean will be, $\mu_2 = N_2^{-1}\left[\sum_{j=1}^{N_2} X_{2j}\right]$. (4)

Population mean will be, $\mu = N^{-1}\left[\sum_{j=1}^{N} X_j\right]$. (5)

Household and individual response and nonresponse rates are computed on the basis of the methodology which was proposed by Ayhan (2017).

In a similar way, the sample size of *n* can artificially be divided into response and nonresponse stratum. We can use the following form ($r_i = n_i/n$) of the rate and the size ($n_i = \sum_{j=1}^{J} n_{ij}$) to illustrate the mechanism, where *i = 1, 2*.

For ***one-stage sample selection***, which can be based on household selection;

Response rate: $r_1 = n_1/n$ (6)

Nonresponse rate: $r_2 = n_2/n$ (7)

$n_1 + n_2 = n$, $r_1 + r_2 = 1$, $(1 - r_1) = r_2$ (8)

Household response rate (*HRR*) is computed as the ratio of (n_1/n) from the selected sample. *Household nonresponse rate* (*HNRR*) can be taken as the complement of the household response rate (*HRR*), for first stage sample selection.

$Household\ SurveyRR = \frac{n_1}{n}$ (9)

$HNRR = (n_2/n) = 1 - HRR = [1 - (n_1/n)]$ (10)

For ***two-stage sample selection***, which can be based on household survey (n_i/n) and individual person (m_i/m) selections, as a product;

Household response rate, $HRR = n_1/n$ (11)

Individual response component, $IRC = m_1/m$ (12)

Individual response rate, $IRR = (HRR)(IRC) = (n_1/n)(m_1/m)$ (13)

Individual nonresponse rate (*INRR*) is calculated by the multiplication of *household nonresponse rate* and *individual nonresponse component*. Individual nonresponse component is calculated by taking *nonrespondent individuals* (m_2) over *enumerated individuals* (m).

Household response rate (*HRR*) is computed as the ratio of (n_1/n) from the selected sample. *Individual response rate* (IRR) is calculated by the multiplication of household response rates and individual response component. Individual response

component is calculated as respondent individuals (m_1) over, enumerated individuals (m). These calculations are given with the following formulas.

$$Individual\ Survey\ RR = \frac{n_1}{n}\frac{m_1}{m} \tag{14}$$

$$Individual\ Survey\ NRR = \frac{n_2}{n}\frac{m_2}{m} \tag{15}$$

Individual nonresponse component, $INRC = m_2/m$ (16)

Individual nonresponse rate, $INRR = (HNRR)(INRC) = (n_2/n)(m_2/m)$ (17)

For two-stage sample selection, individual survey response rate and individual survey nonresponse rate cannot be simple complements.

That is, $[(n_1/n)(m_1/m)] \neq 1 - [(n_2/n)(m_2/m)]$. (18)

When n_1 and n_2 are taken as fixed, where $n_1 = \frac{N_1}{N} n$ and $n_2 = \frac{N_2}{N} n$ then

$$E(n_1)\,E(\bar{x}_1 \mid n_1) = \frac{N_1}{N} n \quad \text{and} \quad E(n_2)\,E(\bar{x}_2 \mid n_2) = \frac{N_2}{N} n \tag{19}$$

Moser and Kalton (1979), Ayhan (1981), and Bethlehem and Kersten (1985) stated that, the bias of nonresponse occurs when the response stratum mean μ_1 is used instead of the total population mean μ. The source of nonresponse bias is based on the use of

$$Lim_{n\to N} E(\bar{x}_1) = \mu \text{ instead of } Lim_{n\to N} E(\bar{x}) = \mu, \tag{20}$$

$$\text{where } Lim_{n\to N} E(\bar{x}_1) \neq \mu \text{ but } Lim_{n\to N} E(\bar{x}_1) = \mu_1. \tag{21}$$

The *nonresponse bias* due to the use of response stratum mean will be,

$$B(\bar{x}_1) = \mu_1 - \mu = \mu_1 - (R_1\mu_1 + R_2\mu_2) \tag{22}$$

$$= \mu_1(1 - R_1) - R_2\mu_2 = R_2(\mu_1 - \mu_2) \tag{23}$$

where $\mu = (R_1\mu_1 + R_2\mu_2)$ and $(1 - R_1) = R_2$

The effect of bias is based on the *amount of nonresponse rate* and the *difference between the response and nonresponse strata means*. Detailed derivations of the proof are available by Moser and Kalton (1979) and Ayhan (1981).

3. Taking response amounts as random variables

Survey nonresponse components and issues of bias have been examined by Bethlehem and Kersten (1985), Bethlehem and Keller (1987) and Bethlehem (2002 & 2009).

The random response model assumes every element in the population to have an unknown response probability. If an element is selected in the sample, a random mechanism is activated that results with a given probability in response and with a complement probability in nonresponse (Bethlehem, 2009).

In order to consider the response amounts as random variables we have proposed the following set of formulations;

$$\text{Define } \mu = \frac{N_1}{N}\mu_1 + \frac{N_2}{N}\mu_2 \ , \ \ N = N_1 + N_2 \tag{24}$$

$$\text{Let } \ \hat{\mu} = \frac{n_1}{n}\bar{x}_1 + \frac{n_2}{n}\bar{x}_2 \ , \ \ n = n_1 + n_2 \tag{25}$$

$$E(\hat{\mu}) = \frac{1}{n}\ E(n_1)E(\bar{x}_1 \mid n_1) + \frac{1}{n}\ E(n_2)E(\bar{x}_2 \mid n_2) \tag{26}$$

$$= \frac{1}{n}\left[n\ \frac{N_1}{N}\ \mu_1 + n\ \ \frac{N_2}{N}\ \mu_2 \right] = \ \frac{N_1}{N}\ \mu_1 + \ \ \frac{N_2}{N}\ \mu_2 \tag{27}$$

Putting the two strata together and draw a random sample of size *n*. Let n_1 fall into stratum 1, and n_2 fall in stratum 2, $n_1 + n_2 = n$.

$$\text{Then } E\left(\frac{n_1\ \bar{x}_1 + n_2\ \bar{x}_2}{n} \right) = \frac{n\ \dfrac{N_1}{N}\ \mu_1 + n\ \dfrac{N_2}{N}\ \mu_2}{n} = \frac{N_1\ \mu_1 + N_2\ \mu_2}{N} = \mu \tag{28}$$

$$E(\bar{x}_1 \mid n_1) = \mu_1 \quad \text{and} \quad E(\bar{x}_2 \mid n_2) = \mu_2 \tag{29}$$

When nonresponse occurs at random, it reduces to a single sample situation with sample size *n* in which case $\bar{x}_1$ is estimating μ.

There is a real problem with the methodology as follows: Let $\bar{x}$ be the sample mean.

$$\text{Then, } E(\bar{x}) = \mu \quad \text{or } E(\bar{x}) = \bar{X} \ \text{ and} \tag{30}$$

$$V(\bar{x}) = \frac{\sigma^2}{n} \quad \text{or} \quad V(\bar{x}) = \frac{S^2}{n}\left(1 - \frac{n}{N}\right) \tag{31}$$

Only if n is fixed a priori. If n is a random, the results are different:

$$E(\bar{x}) = E(n)\, E(\bar{x} \mid n) \tag{32}$$

$$E(\bar{x}^2) = E(n^2)\, E(\bar{x}^2 \mid n) \tag{33}$$

In this case, n_1 (*responses*) and n_2 (*nonresponses*) and

$$n_1 + n_2 = n \tag{34}$$

n is the *sample size*, are both subject to (34). Therefore, the results in (31) are not applicable. In fact, n_1 (or n_2) has a binomial distribution with,

$$E(n_1) = n(N_1/N) \tag{35}$$

Here, N_1/N is the proportion of responses in the population.

The difficulty is that, the value of N_1 is not known. If one estimates N_1/N by n_1/n, then $E(n_1) \cong n\,(n_1/n) = n_1$ (*does not make sense, since the expected value of a random variable cannot be the random variable itself*), and that is where the difficulty is. Knowing n_1 a priori which is untenable.

$N_1 + N_2 = N$ Here, N_2 are presumed to be nonresponses;

$$\mu = (N_1/N)\mu_1 + (N_2/N)\mu_2 . \tag{36}$$

A sample of size *n* is available with n_1 responses (random n_1) and n_2 nonresponses;

$$n = n_1 + n_2 . \tag{37}$$

Assuming *N* is very large and sampling is done without replacement, n_1/n is a binomial variate with $E(n_1/n) = N_1/N$ (Tiku, 1964).

$$E\left(\frac{n_1}{n}\bar{x}_1\right) = E\left(\frac{n_1}{n}\right) E(\bar{x}_1/n_1) = \frac{N_1}{N}\mu_1 = \mu - \frac{N_2}{N}\mu_2 \tag{38}$$

$$Bias(\bar{x}_1) = -\frac{N_2}{N}\,\mu_2 = R_2\ \mu_2 . \tag{39}$$

If we replace the random variable n_1/n by its expected value N_1/N which is mathematically naive,

$$E\left(\frac{n_1}{n}\bar{x}_1\right) \cong E\left(\frac{N_1}{N}\bar{x}_1\right) = \frac{N_1}{N}\mu_1 \tag{40}$$

Consequently, $E(\bar{x}_1) \cong \mu_1$;

The bias is, $Bias(\bar{x}_1) = E(\bar{x}_1) - \mu \;\cong\; \mu_1 - \left(\frac{N_1}{N}\mu_1 + \frac{N_2}{N}\mu_2\right)$

$$= \frac{N_2}{N}(\mu_1 - \mu_2) = R_2(\mu_1 - \mu_2); \tag{41}$$

This is however a very naive approximation.

The variance of $\frac{n_1}{n}\bar{x}_1$ is $V\left(\frac{\mathrm{n}_1}{\mathrm{n}}\bar{\mathrm{x}}_1\right) = \frac{N_1}{N}\frac{S_1^2}{n}\left(1 - \frac{N_1}{N}\right)$. (42)

***Remark*:** You may notice that, equations (39) and (41) are very different from one another. While equation (39) is mathematically correct, equation (41) is suspicious. The only common ground is when $N_2/N = 0,$ *i.e.*, $N_2 = 0,$ in which case both equations (39) and (41) are equal to zero. Since n_1 is a random variable, the sampling variance of the mean for response stratum is,

$$V(\bar{\mathrm{x}}_1) = \frac{S_1^2}{n}\left(1 - \frac{N_1}{N}\right)E\left(\frac{1}{n_1}\right). \tag{43}$$

Thus $\bar{x}_1$ is not an attractive estimator since $n_1 = 0$ has to be included in which case the Binomial has to be truncated.

$$\mu = \left(\frac{N_1}{N}\mu_1 + \frac{N_2}{N}\mu_2\right) \quad \text{and} \quad E\left(\frac{n_1}{n}\bar{x}_1\right) = \frac{N_1}{N}\mu_1 \tag{44}$$

$$\frac{N}{N_1}\frac{n_1}{n}\bar{x}_1 \text{ is an unbiased estimator of } \mu_1. \tag{45}$$

$$E\left(\frac{\mathrm{N}}{\mathrm{N}_1}\frac{\mathrm{n}_1}{\mathrm{n}}\bar{\mathrm{x}}_1 - \mu\right) \quad \text{is} \quad \mu_1 - \mu = \mu_1 - \left(\frac{N_1}{N}\mu_1 + \frac{N_2}{N}\mu_2\right) = \frac{N_2}{N}(\mu_1 - \mu_2). \tag{46}$$

$$\text{Bias in } \frac{N}{N_1}\frac{n_1}{n}\bar{x}_1 \text{ is } \frac{N_2}{N}(\mu_1 - \mu_2). \tag{47}$$

4. Bias Adjustment Procedures

A recent research on the survey nonresponse bias adjustment has been proposed by Ayhan (2017). The results have shown the effect of nonresponse and callbacks on the estimation of survey nonresponse bias. The following remedies can also be used to adjust the nonresponse error which has occurred in a given survey. Due to the available means, we cannot elaborate any further for the all possible survey situations.

4.1. Use of auxiliary information from subsampled nonrespondents

The nonresponse bias of the stratum mean estimator is,

$$B(\bar{x}_1) = \mu_1 - (R_1\mu_1 + R_2\mu_2) \tag{48}$$

The design mean can be evaluated as,

$$\hat{\mu} = R_1\bar{x}_1 + R_2\mu_2 \tag{49}$$

Since μ_2 is not known, the sample estimator will take the following form,

$$\bar{x}_w = \sum_{i=1}^{2} R_i x_i = R_1\bar{x}_1 + R_2\bar{x}_2^* \tag{50}$$

where $\bar{x}_1 = n_1^{-1}\left[\sum_{j=1}^{J} X_{1j}\right]$ and $\bar{x}_2 = n_2^{-1}\left[\sum_{j=1}^{J} X_{2j}\right]$ is unknown. By taking a random subsample of size m_2, a new estimator of the nonresponse stratum mean will take the following form, where $m_2 = f_b\,(n_2)$.

$$\bar{x}_2^* = m_2^{-1}\left[\sum_{j=1}^{J} X_{2j}\right] = \hat{\mu}_2 \text{ and } E(\bar{x}_2^*) = \mu_2 \tag{51}$$

Here f_b is the subsampling rate from the nonresponse stratum and can be taken as $f_b = 0.05$. The expected value of the subsample estimator will be, $Lim_{n\to N} E(\bar{x}_2^*) = \mu_2$.

On the other hand, the desired estimator of the sample mean is, $\bar{x} = n^{-1}\left[\sum_{j=1}^{J} X_j\right]$

4.2. Domain based weighting adjustments for nonresponse

For the domain–based weighting adjustments for nonresponse, we have proposed the following set of formulations. The probability of selection of the overall sample is

obtained simply by the sampling fraction of the selected sample $f = x/X = 1/F$ for the total sample. On the other hand, after using some method of stratification, the sampling fraction of any strata is $f_i = x_i / X_i = 1/F_i$.

Design weights (Ayhan 1991 and Verma 1991) for non self-weighting sample designs can be computed for each domain *i* with the same probability of selection p_i .

For a *combined ratio mean* $\theta = Y/X = \sum_i^H Y_i \Big/ \sum_i^H X_i$, which is estimated by

$$\hat{\theta} = y/x = \sum_i^H y_i \Big/ \sum_i^H x_i \quad . \tag{52}$$

On the other hand, for a *separate ratio mean* $\theta_w = \sum_i^H W_i \ \theta_i$, estimated by

$$\hat{\theta}_w = \sum_i^H W_i \ \hat{\theta}_i = \sum_i^H W_i \ \left[y_i / x_i \right] . \tag{53}$$

The weight $W_i = \left[\sum_{i=1}^H x_i \Big/ \sum_{i=1}^H \left\{ x_i / \left[(X/x)\, p_i \right] \right\} \right] \Big/ \left[(X/x)\, p_i \right] = P_0 / P_i$ (54)

where $\sum_{i=1}^H \left(W_i \ x_i \right) = x$

Here, P_0 has been computed to adjust the overall weighted and unweighted sample to be the same. In addition, a weighting procedure for nonresponse is also essential for self-weighting and nonself-weighting sample design outcomes (Ayhan 2003).

Here $W_i^* = R_0 / R_i$ where $R_i = x_i^* / x_i$ is the response rate in domain *i*. (55)

The overall response rate (R_0) for the design can be computed as,

$$R_0 = \sum_{i=1}^I \left(W_i \ x_i \right) \Big/ \sum_{i=1}^I \left(W_i \ x_i \ / \ R_i \right) \tag{56}$$

where R_0 is used to adjust the sample sizes to be the same, $\sum_{i=1}^I \left(W_i \ \ W_i^* \ x_i \right) = x$. (57)

5. Conclusions

The evaluation of the nonresponse bias as nonresponse error or nonresponse rate was misleading. The nonresponse bias may seem to be related to the response rates for a given study. Increasing response rate may not always correspond to decreasing

nonresponse bias for a given study. This paper has shown alternative approaches to nonresponse bias. In addition to this, the causes of the nonresponse bias can also be obtained from empirical studies of components and models relating to the covariates of survey participation and non-participation.

The current research examined the response amounts as fixed initially. The proposed methodology has shown the effect of bias of nonresponse which is based as the product of "amount of nonresponse rate" and the "difference between the response and nonresponse strata means" [$B(\bar{x}_1) = R_2(\mu_1 - \mu_2)$].

When the response amounts are taken as random variables, the nonresponse bias has provided the same solution [$B(\bar{x}_1) = N_2/N\ (\mu_1 - \mu_2)$].

A recent research on the survey nonresponse bias adjustment has been proposed by Ayhan (2017). The current study has examined the nonresponse bias adjustment by using additional auxiliary information from the subsampled nonrespondents. An alternative approach was also used by domain–based weighting adjustments for nonresponse.

References

Ayhan, H. Ö., (1981). Sources of Nonresponse Bias in 1978 Turkish Fertility Survey. *Turkish Journal of Population Studies*, 2–3, pp. 104-148.

Ayhan, H. Ö., (1991). Post Stratification and Weighting in Sample Surveys. *Research Symposium '91*. State Institute of Statistics, Ankara, 11 pp.

Ayhan, H. Ö., (2003). Combined Weighting Procedures for Post–Survey Adjustment in Complex Sample Surveys. *Bulletin of the International Statistical Institute*, 60(1), pp. 53-54.

Ayhan, H. Ö., (2017). Effect of Nonresponse and Callbacks on the Estimation of Survey Nonresponse Bias. *Turkish Journal of Population Studies*, 39, pp. 91-107.

Bethlehem, J. G., Kersten, H. M. P., (1985). On the Treatment of Nonresponse in Sample Surveys. *Journal of Official Statistics*, 1(3), pp. 287-300.

Bethlehem, J. G., Keller, W. J., (1987). Linear Weighting of Sample Survey Data. *Journal of Official Statistics*, 3(2), pp. 141-154.

Bethlehem, J. G., (2002). Weighting Nonresponse Adjustments Based on Auxiliary Information. In: Groves, R. M., Dillman, D. A., Eltinge, J. L. & Little, R. J. A. (Eds), *Survey Nonresponse*. New York: John Wiley & Sons.

Bethlehem, J. G., (2009). *Applied Survey Methods: A Statistical Perspective*. Hoboken, NJ: John Wiley & Sons.

Groves, R. M., Couper, M. P., (1998). *Nonresponse in Household Interview Surveys*. New York: John Wiley & Sons.

Groves, R. M., D. A. Dillman, J. L. Eltinge and R. J. A. Little, Eds., (2002). *Survey Nonresponse*. New York: John Wiley & Sons.

Keser, İ. K., (2011). The History of Survey Sampling. *TurkStat, Journal of Statistical Research*, 8(3), pp. 54-74.

Kish, L., (1995). *Survey Sampling*. New York: John Wiley & Sons.

Lindström, H., (1983). *Non-Response Errors in Sample Surveys*. Statistics Sweden, Urval Number 16, 94 pp.

Lindström, H., J. Wretman, G. Forsman, and Cassel, C., (1979). *Standard Methods for Non-Response Treatment in Statistical Estimation*. National Central Bureau of Statistics, Sweden, 60 pp.

Moser, C. A., Kalton, G., (1979). *Survey Methods in Social Investigation*. London: Heinemann Educational Books.

Tiku, M. L,. (1964). A Note on the Negative Moments of a Truncated Poisson Variate. *Journal of the American Statistical Association*, 59, pp. 1220-1224.

Verma, V. K., (1991). *Sampling Methods*. United Nations, Statistical Institute for Asia and the Pacific, Manual for Statistical Trainers, Number 2. Tokyo, Japan.