ISSN 2348-3156 (Print)
International Journal of Social Science and Humanities Research ISSN 2348-3164 (online)
Vol. 6, Issue 2, pp: (720-724), Month: April - June 2018, Available at: www.researchpublish.com

Post-stratification Weighting Adjustment Technique for Reducing Bias in Volunteer Panel Web Surveys

Dr. Md. Musa Khan

Assistant Professor, Department of Business Administration
International Islamic University Chittagong, Chittagong, Bangladesh.
E-mail: khanstatcu@gmail.com

Abstract: **Web surveys have become one of the most widely utilized and popular survey methods in recent years. Data collection on the Internet is faster than other methods such as paper-and-pencil, computer-assisted telephone interviews and personal interviews, because it is simple, cheap and provides quick access to the desired large group of respondents. However, in web surveys, bias may arise mainly due to limited coverage and self-selection. This study appraises characteristics and problems of web surveys and reduction techniques for reducing their biases. The post-stratification weighting is used for reducing these biases. In the application of this study, population estimates are compared to those from a volunteer panel web survey. The data are from the survey *"Using Social Networking Sites in the Education of Students of Open Education System, Anadolu University"*. It is shown that the weighting adjustment has reduced the bias substantially. When it is necessary to use volunteer panel web surveys, it is recommended to adjust the post-stratification weighting adjustment technique for reducing bias described in this study.**

Keywords: **Web surveys, Volunteer panel web surveys, Non-probability sampling design, Bias, Weighting adjustment techniques.**

1. INTRODUCTION

The Internet is a tool of data collection. It is usable for conducting web surveys. In web surveys, data are collected from the individuals via the World Wide Web. This method has become more popular. The survey research scenario has gone through rapid progress over the last two decades. The first most traditional data collection mode which is paper-and-pencil interviewing (PAPI) was replaced by the computer-assisted interviewing (CAI) method (Couper, 2005). Now, the most commonly used traditional data collection method comprises of face-to-face surveys (CAPI), computer-assisted telephone surveys (CATI), and mail surveys (CASI, CSAQ) have progressively substituted by web surveys (Bandilla et al. 2003, Couper et al. 2007, & Dever et al. 2008).

A web survey is an approach to reach a massive number of potential individuals via the World Wide Web. It includes a variety of techniques with diverse aims, target individuals, populations, targets groups etc. Questionnaires can be distributed among large number of individuals at low cost – interviewers or paper-based questioners are not required, and printing and/or mailing costs in the survey is reduced. Web surveys can start very rapidly. It demands less or no time between preparing the questionnaire and starting the fieldwork when compared to traditional studies. These surveys also offer up-to-date facilities, and thus attractive questionnaires can be produced using multimedia contents like pictures, sounds, animations, and movies.

Although web surveys appear to provide a lot more functionalities than traditional types, it is merely another mode or method of data collection. In the web surveys, interviewers are not required. Questionnaire is completed by the individuals over the Internet. However, some problems can make the outcomes of web surveys unreliable. In web surveys,

problems which can arise are under-coverage, self-selection, and measurement errors (Bethlehem, 2010). These issues can cause population characteristics (parameter) estimates to be biased. Therefore, wrong decisions can be made on the web survey.

Now, considering the web survey errors, firstly, under-coverage happens if the target population is more extensive than those members having the Internet access. Parameter estimates will be biased if the Internet access individuals do not differ from without Internet access individuals (Steinmetz et al. 2009). Secondly, self-selection means that there is a freedom to participate in the web surveys. Individuals are selected by themselves independently in the web surveys. The questionnaire of the study is just put on the web page or websites. Internet accessing respondents visit the website or web page and decide to participate in the surveys. These participants vary significantly from the nonparticipants. Finally, measurement errors are the difference between a measured quantity and its actual value. It involves random error and systematic error. The random error occurs if repeated measurements produce values that may vary around the real value of the quantity. This error may be caused by the limited precision of the instruments used for measuring (Bandilla et al. 2003, Fricker 2008, & Malhotra & Krosnick, 2007). It is said to have systematic errors if repeated measurement produces benefits that systematically occur. This may be caused by a miss-calibrated instrument that affects all measures in the survey.

Nowadays, it is more stringent to get information from individuals due to an increase in one-person households and dual-income households (Lee, 2011). The interviewers have difficulties in meeting those individuals during daylight hours, and growing concern for privacy is another primary concern involved in the traditional surveys.

This study reveals characteristics and problems of web surveys and evaluation of the performance of the implemented the weighting adjustment technique for reducing the bias in the volunteer panel web surveys. There exist many weighting adjustment techniques. Weighting adjustment is a set of techniques that take an attempt to increase the accuracy of estimates by using auxiliary information. Auxiliary information is defined as a set of variables that are measured in the survey.

Weighting adjustment technique can be applied to improve the precision of estimates. In addition, it turns weighted sample into representative with respect to some auxiliary variables as well as it is often used to correct the bias caused by non-response errors, coverage errors, and sampling errors.

Therefore, web surveys are less costly and offer a quicker method for data collection. The web surveys are becoming the main data collection method for its advantages. Data collection over web surveys is prone to many errors. Some of these are coverage problem, self-selection problem, and non-response problems. The post-stratification weighting adjustment technique can be applied to outweigh the biases.

2. DATA AND METHODOLOGY

The study conducts a volunteer panel web survey by self-selection—called a non-probability-based web survey. The survey title of this study *"Using Social Networking Sites in the Education of Students of Open Education System, Anadolu University"* has been conducted for an academic purpose to assist a study on the efficient use of the social networking sites in higher education at Anadolu University to enrich students' knowledge and learning. Volunteer panel web surveys are conducted based on panelists which consist of individuals who decide to participate voluntarily in surveys via websites. Respondents are members of a volunteer panel web of the Open Education System's students at Anadolu University (Buchanan et al., 2007). The panel web members (students) studies in one of the three faculties (Open Education, Business Administration, and Economics) at the Anadolu University and participated in the survey by self-selection**.** Social networking sites (SNSs) are online platforms those permit users to make a public profile and connect users on the websites. Social networking sites are also known as social networking websites or social websites. Social networking sites can be used as community-based websites, online discussions forums, chat rooms, and other social spheres online (Hill et al, 2014). Blogs, Twitter, Facebook, LinkedIn etc. are examples of social networking sites.

The target variable *"using social networking sites in the education"* is considered as a response variable. The selected auxiliary variables in this study are regarded as predictors. The questionnaire consists of 26 questions where 1-8 are demographic questions, and 9-26 are *"using social network sites in the education"* related questions. Total of 2920 respondents participated in the survey. The reliability of the volunteer panel web survey data is 62.3%. This study has only been used the seven demographic auxiliary variables to estimate the target variable. The selected auxiliary variables

ISSN 2348-3156 (Print)
International Journal of Social Science and Humanities Research ISSN 2348-3164 (online)
Vol. 6, Issue 2, pp: (720-724), Month: April - June 2018, Available at: www.researchpublish.com

in this study are*: Gender, Age, Region, Working status of respondents, Marital status, Level of the program in the study, and Faculty in the study***.** In this study, the population is the total students of the Open Education System, Anadolu University in Spring, 2016-2017. The students of Open Education System have been taken in the study because all the students have access to the Internet. The target population size is *N* = 1043283.

In web surveys, bias may arise mainly due to under coverage, self-selection and non-response errors. The data can be adjusted to correct these errors. Mainly, weighting adjustments are a potential resolution to improve the quality of web surveys (Bethlehem and Stoop 2007, & Dever et al. 2008). Weighting adjustments are techniques that used to reduce the bias of estimates by using auxiliary variables (Bethlehem, 2010). The utilized weighting adjustment techniques for reducing the bias in web surveys are post-stratification weighting, propensity score adjustment, rim weighting and generalized regression modeling (Lee, 2004; Bethlehem, 2010; & Steinmetz et al., 2009).

Post-stratification weighting is an adjustment estimation technique that reduce the non-coverage and non-response biases as well as variance of the estimates (Cervantes et al., 2009). It is utilized to adjust weight for demographic variable's differences between a sample and the population. Looseveldt and Sonck (2008) argued that the technique does not resolve the problem of selection bias since some response variables may be associated with variables apart from demographics characteristics. For example, attitudinal and behavioral differences may be observed even when applying the post-stratification weighting adjustment using demographic variables.

Post-stratification needs one or more auxiliary variables. An auxiliary variable is a variable that is measured in the survey, and that the distribution of the target population is available. In this study, the target variable is "*using social networking sites in the education".* Percentage distribution of auxiliary variables of the volunteer panel web survey are compared with its population to assess whether the conducted survey is representative to the population. If these distributions do not differ, it may be concluded that the conducted survey response is nonresponsive. Adjustment weights are calculated for this correction. Weights assess any or all register of observed elements. Population estimates can be computed utilizing the weighted values rather than the unweighted values.

The study has a volunteer panel web sample of size *n* = 2920. In stratum *h*, the number of sample elements is denoted by n_h, then $n = n_1 + n_2 + \cdots + n_L$. The values of the n_h are the outcome of a random selection procedure, so, they are random variables. It is noted that since the sample is collected from the Internet access population, only elements in the sub-strata $U_I \cup U_h$ are detected (*for h=1, 2, ..., L).*

For each stratum, post-stratification is allotted adjustment weights which are identical. The correction weight c_i for *h* stratum is equal to $c_i = \frac{N_h/N}{n_h/n}$, where, $N_h$is the size of stratum *h*; n_h is the sample size of stratum *h*; *N* is the total target population and *n* is the sample size. If the values of the inclusion weights d_i=*N/n*, then the post-stratification adjustment weights w_i are found by multiplying the correction weights c_i and the inclusion weights d_i as $w_i = c_i \times d_i$. The weighted estimate would be $c_i \times d_i \times n_h$.

Theoretically, a good estimator does not guarantee the bias reduction of an estimate in the sample because survey weights and application of adjustment have not been performed yet. Thus, the primary attention is on the performance of each implemented techniques according to the percentage of bias reduction. Each of the weightings models is assessed based on the target variable in terms of the percentage bias reduction. The computation of percentage bias reduction formula (Lee, 2011) is defined as follow:

$p.bias(\hat{\theta}^{\text{W. A)}} = \left[\frac{|bias(\hat{\theta}\text{W.U})| - |bias(\hat{\theta}\text{W.A})|}{|bias(\hat{\theta}\text{W.U})|}\right] \times 100,$ where, $bias(\hat{\theta}\text{W.U})$ is the unadjusted estimate and $bias(\hat{\theta}\text{W.A})$ is an adjusted estimate in the volunteer panel web survey.

3. RESULTS AND DISCUSSION

Now, it explores how above mentioned auxiliary variables can use for weighting adjustment. In the volunteer panel web survey, the target variable is the "*using social networking sites in the education"* where the volunteer panel web response percentage of the using SNSs in the education is 57.00% and the population percentage of the individuals using SNSs in the education is 58.80%. The difference between the volunteer panel web sample estimate and population estimate is significantly different—i.e., -1.88%. Therefore, the volunteer panel web sample estimate is a biased estimate. Now, the

study examines whether this estimate can be improved by weighting adjustment. It considers seven weighting models which are statistically significant. These weighting models have been created by using above mentioned auxiliary variables.

The first step of the analysis is to compare the percentage distribution of the response of the target variable with population distribution according to the selected auxiliary variables. It is shown as follows:

Table 1: The percentage of using SNSs in the education for the population and volunteer panel web sample

Population			**Volunteer panel web sample**		
	Frequency	**Percentage (%)**	**Frequency**	**Percentage (%)**	**Difference (%)**
No	429768	41.2	1255	43.0	+1.8
Yes	613515	58.8	1665	57.0	-1.8
Total	1043283	100	2920	100	

Table 1 evinces the overall using of social networking sites (SNSs) in the education of the population and volunteer panel web sample for the selected auxiliary variables. The volunteer panel web sample percentage (57.0%) of using the SNSs in the education is lower than the population percentage (58.8%) whereas not using the SNSs percentage (43.0%) is larger than the population percentage (41.2%). In the volunteer panel web sample, out of 2920 respondents, 1665 (57.0%) respondents use the SNSs in the education, but in the population, 613515 (58.8%) respondents consider the SNSs out of the1043283 respondents.

Table 2: Post-stratification weighting estimation of the target variable for reducing the bias in the volunteer panel web survey

Weighting model	**Estimate (%)**	**Standard error**
No weighting	**57.00**	**0.9160**
1. Gender	56.87	0.9667
2. Age	56.51	1.2217
3. Region	57.02	0.9156
4. Working status	57.04	0.9155
5. Marital status	57.02	0.8961
6. Level of the program in the study	**58.87**	**0.9025**
7. Faculty in the study	57.05	0.9244
Population	**58.80**	

Table 2 illustrates the post-stratification estimates of the target variable *"using social networking sites in the education"* to the weighting models with its standard error. There is almost no change in the estimate, and no reduction in the standard error of the estimate. For that reason, most of the weighting model's effects are not acceptable. The result is different in the model-6 for the variable of *Level of the program in the study*. The estimate of the model-6 has adjusted in correction, from 58.80% to 58.87%. The standard error of the adjusted estimate (0.9025) is also lower than unadjusted estimate. Therefore, the post-stratification estimate is almost unbiased for the variable, *Level of the program in the study*. It is recommended that the variable may be included in the post-stratification weighting model for reducing the bias in the volunteer panel web sample estimate.

Table 3 P: ercentage of bias reduction of the estimate of the target variable by the post-stratification weighting for the volunteer panel web sample

Weighting model	**Estimate (%)**	**Bias reduction (%)**
No weighting	**57.00**	**-**
1. Gender	56.87	-7.22
2. Age	56.51	-27.22
3. Region	57.02	1.11
4. Working status	57.04	2.22
5. Marital status	57.02	1.11
6. Level of program in the study	**58.87**	**96.11**
7. Faculty in the study	57.05	2.78
Population	**58.80**	

Table 3 reveals the percentage bias reduction of the implemented seven weighting models in the volunteer panel web sample for the post-stratification weighting technique. The total bias reduction ratio out of all seven weighting models in the study— by the post-stratification weighting of the target variable (bias reduction >0%)—is 71.43%, and the total ratio of considerable bias reduction (bias reduction >50%) is 14.23%. The worst-case bias reduction ratio is -27.22%. The best bias reduction 96.11% has found in the post-stratification weighting model-6 (*Level of the program in the study*). Therefore, the estimates of the post-stratification weighting model-6 should be used for calculating the volunteer panel web sample estimates.

4. CONCLUSION

Recently, web surveys have become familiar because of their attractive features in data collection. However, non-probability-based web surveys cause problems—coverage, self-selection and non-response. Numerous researches have been performed on web surveys, specifically, volunteer web panel surveys, to address the above issues and provide solutions. The post-stratification weighting method has been used for reducing the bias of the target variable. Bias has been reduced substantially. If any researcher would like to conduct a research considering web survey, volunteer panel web surveys would be the best option for the researcher. The study also recommends that in scientific research or commercial market research, volunteer panel web surveys would provide the high-quality data with minimum time and cost. Most importantly, if any bias occurs, it can be removed by applying the post-stratification weighting adjustment technique implemented in this study.

REFERENCES

[1] Bandilla, W., Bosnjak, M., & Altdorfer, P. (2003). Survey administration effects? *Social Science Computer Review*, 21(2), 235–243.

[2] Bauchanan, T., Paine, C., Joinson, A.N., & Reips, U. (2007). Development of Measures of Online Privacy Concern and Protection for Use on the Internet. *Journal of American Society for Information Science and Technology*, 58(2), 157-165.

[3] Bethlehem, J. G. (2007), Reducing the Bias of Web Survey Based Estimates. Discussion Paper 07001. Statistics Netherlands, Voorburg/Heerlen, Netherlands.

[4] Bethlehem, J. (2010). Selection Bias in Web Surveys. *International Statistical Review*, 78(2), 161–188.

[5] Cervantes, I. F., Brick, M. J., & Jones, M. (2009). Efficacy of post-stratification in complex sampling designs. *Methodology Series*, 4, 25-38.

[6] Couper, M.P., Kapteyn, A., Schonlau, M., & Winter, J. (2007). Noncoverage and nonresponse in an Internet survey. *Social Science Research*, 36(1), 131-148.

[7] Dever, J.A., Rafferty, A., & Valliant, R. (2008). Internet surveys: Can statistical adjustments eliminate coverage bias? *Survey Research Methods*, 2, 47-62.

[8] Fricker, R. D. J. (2008). Sampling Methods for Web and E-mail Surveys, *SAGE Publications, Ltd*, 195–216.

[9] Hill, C.A., Dean, E., & Murphy, J. (2014). *Social Media, Sociality, and Survey Research*. John Wiley & Sons, Hoboken, NJ, U.S.A.

[10] Lee, S. (2004). *Statistical estimation methods in volunteer panel web surveys.* Ph.D. thesis, University of Maryland.

[11] Loosveldt, G. & Sonck, N. (2008). An evaluation of the weighting procedures for an online access panel survey. *Survey Research Methods*, 2(2), 93-105.

[12] Lee, M.H. (2011). *Statistical methods for reducing bias in web surveys.* M.S. thesis, Simon Fraser University.

[13] Malhotra, N. & Krosnick, J. A. (2007). The effect of survey mode and sampling on inferences about political attitudes and behavior: Comparing the 2000 and 2004 anes to internet surveys with nonprobability samples. *Political Analysis*, 15(3), 286–323.

[14] Steinmetz, S., Tijdens, K., & de Pedraza, P. (2009). Comparing data from online and face-to-face surveys. Amsterdam Institute for Advanced Labor Studies, University of Amsterdam, 9-76.