Malaysian Journal of Mathematical Sciences 15(3): 477–488 (2021)

Malaysian Journal of Mathematical Sciences

Journal homepage: https://einspem.upm.edu.my/journal

Population Mean Estimation Using Weight Adjustment for Unit Non-Responses

Jayamanne, I. T.* and Ramanayake, K. P. A.

Department of Statistics, University of Colombo, Sri Lanka

E-mail: imali@stat.cmb.ac.lk
**Corresponding author*

Received: 26 December 2020
Accepted: 26 April 2021

Abstract

The population mean is often estimated using the sample mean with no consideration to the design of the survey, the survey errors and biases. Although surveys are properly designed to reduce errors, unit non-response and coverage errors are mostly unavoidable. Thus, non-response adjustment for the sampling weights is essential when estimating the population mean. A survey of Sri Lankan university graduates was conducted in 2016 for a random sample of Art graduates who had graduated in 2012 in all state universities. This study aims to estimate the mean waiting time for the first employment after graduation for all 2012 Arts graduates based on these data. The response rate of the survey was 48% and it was noticed that the response rate varied with university, gender and ethnicity of the graduates. The sampling weights were adjusted using the individual propensities and the class propensities determined by the propensity adjustment score model to select the best non-response adjustment weights for the data. Next, the final weight adjustment was done using post-stratification, raking and calibration using census data that was available, taking university, ethnicity and gender as auxiliary variables. The model with the individual propensity adjustment when calibrated using university and gender cross-cell counts and marginal counts for ethnicity provided the smallest standard error for the population mean. Finally, the mean waiting time for the first employment was estimated using these adjusted weights as 15.19 months with a standard error of 0.657.

Keywords: calibration; post-stratification; propensity model; raking; trim weights.

1 Introduction

The sample mean is commonly used as a point estimate of the finite population mean when the sample is from a simple random sample (SRS) of the population. However, if the sample is drawn from a complex survey design, the weighted mean is one of the preferred estimators instead of the sample mean. When a complex survey design is used, the inclusion probabilities of the observations are often unequal. Thus, the sampling weights, which are the inverse of the inclusion probabilities become unequal too. The sample weight represents the number of units in the population that each individual represents. Even when a survey is properly designed to reduce sampling errors and bias, the non-response-bias is unavoidable due to non-response [1]. Hence, incorporating weights in the inferential analysis is essential to compensate for unequal probability of selection, nonresponse and non-coverage. Since the weights are required to provide estimates that represent the target population, proper weighting procedure may produce unbiased parameter estimators and confidence intervals [3]. Only a few pieces of evidence were about the waiting time for the first employment in Sri Lankan graduate context and none had attempted to carry out population parameter estimations [4]. However, a substantial amount of research has been conducted a few decades back for developing and refining methods for compensating for missing survey data.

A survey of Sri Lankan university graduates was conducted in 2016 for a random sample of Art graduates who had graduated in 2012 in all state universities to identify the changes in their employment over time. Stratified random sampling design was used to collect the information, considering the University as the strata. After a considerable effort, a 48% response rate was achieved. The waiting time for the first employment of the graduates were measured based on the date of first employment and the effective date of the degree. The sample mean waiting time was 15.25 months. Census had also been conducted for the same group in 2012, a few weeks before their official convocation [8]. Hence, the population information for certain variables could be computed when required. The main objective of this study is to estimate the population mean waiting time for the first employment after graduation. To achieve this objective, the best weight adjustment method for non-responses and the best method to adjust the non-coverage is investigated.

The rest of the paper is arranged as follows. Section 2 provides the necessary theoretical framework for weight adjustments for the population mean estimation, while Section 3 provides the applications of the theories to the Arts graduates'dataset. Finally, Section 4 and 5 provide the discussion and conclusions of the study respectively.

2 Materials and Methods

In design-based inference, let π_i be the sampling probability of (inclusion probability) of the i^{th} individual. The sampling weight or the design weight is denoted by w_i[6]

$$w_i = 1/\pi_i. \tag{1}$$

Thus, w_i represents the number of individuals in the population that is represented by the i^{th} individual. The Horvitz-Thompson estimator $\hat{T}_y$, an estimator of the population total T_y of Y response variable is given by,

$$\hat{T}_y = \sum_{i=1}^{n} \frac{1}{\pi_i} Y_i = \sum_{i=1}^{n} w_i Y_i. \tag{2}$$

Using the Horvitz-Thompson estimator, the population mean μ_y is estimated as,

$$\hat{\mu}_y = \frac{1}{N}\hat{T}_y. \tag{3}$$

A SRS of size n from a population of size N, will have equal inclusion probabilities (n/N) for all individuals. When the design is stratified random sampling without replacement, the population is divided into $h = 1, ..., H$ mutually exclusive strata. A sample size of $n_h$is selected in each stratum from a population of size N_h.. Then the inclusion probability of unit i in the stratum h is $\Pi_h i$. This is the same for each sample unit in stratum h, but the sampling rates may be different from one stratum to another.
Thus,

$$\pi_h = \frac{n_h}{N_h}, h = 1, ..., H. \tag{4}$$

2.1 Non-Response Bias

Bias of an estimate can be measured using either a deterministic or a stochastic approach. Only the stochastic approach is discussed here. In the stochastic approach, it is assumed that each individual has a random choice of participating in the survey with a non-zero probability of responding ([10]).

Let

$$I_t = \begin{cases} 1 & \text{if unit i is selected for sample,} \\ 0 & \text{if not.} \end{cases} \tag{5}$$

Thus, the probability of being in the sample can be written as,

$$Pr(I_i = 1) = \pi_i. \tag{6}$$

Next, let

$$R_i = \begin{cases} 1 & \text{if unit i responds given that it is in the sample,} \\ 0 & \text{if unit i does not respond.} \end{cases} \tag{7}$$

Let ϕ_i, be the propensity score for unit i, that is the probability that i^{th} unit will respond given that it is in the sample is

$$Pr(R_i = 1 \,|I_i = 1) = \phi_i. \tag{8}$$

Then the bias of $\hat{\mu_y}$ can be written as

$$Bias\,(\hat{\mu}_y) = \frac{1}{N\phi} \sum (y_i - \hat{\mu}_y)(\phi_i - \bar{\phi}) \tag{9}$$

where $\bar{\phi}$ is the average population probability of responding [10].

That is, the bias of $\hat{\mu_y}$ depends on the covariance of the response variable and its response propensity. If Yi and $\bar{\phi}_i$ are unrelated there is no bias. However, if non-response bias presents, to reduce or to eliminate it, the weights are adjusted as w_i^*, where

$$w_i^* = \frac{w_i}{\phi_i}. \tag{10}$$

2.2 Adjustments for Nonresponse

Since ϕ_i, is unobservable, it could be predicted using a set of auxiliary variables available in the sample which is known as paradata(data available for respondents and non-respondents). weighting class adjustment, propensity score adjustments and classification algorithms are the commonly used methods to predict ϕ_i, [10].In this paper, only the propensity score adjustment method was used since it is the only statistical oriented method that pertains to predicting ϕ_i, using the available para data. Here, it assumes that the missing data are missing at random (MAR) and hence ϕ_i, will depend only on the set of x_i(auxiliary variables) and not on the response variable Y_i Therefore,

$$\phi_i = \phi\left(x_i\right). \tag{11}$$

Thus, a binary regression model can be fitted for the response indicator R_i as,

$$E_R\left(R_i|I_i = 1\right) = Pr(R_i = 1\,|I_i = 1) = \phi\left(x_i\right). \tag{12}$$

Here responding to the survey can be modelled as a realization of a latent variable process. Assume R_i^* is a latent variable that is unobserved and continuous. When the value of R_i^* exceeds a threshold, the observation i responds; otherwise it does not respond. Hence, the latent variable follows a linear model as follows [10],

$$R_i^* = x_i^T\beta + u_i \tag{13}$$

where u_i has distribution function F given by,

$$F^{-1}\left[\phi\left(x_i\right)\right] = x_i^T\beta. \tag{14}$$

Table 1: Link functions

Model	**Link Function** $F^{-1}[\phi(x_i)]$
Probit	$\Phi^{-1}\left[\phi\left(x_i\right)\right]$
Logit	$log\left(\frac{\phi(x_i)}{1-\phi(x_i)}\right)$
Complementary log-log	$log{-}log[1 - \ \phi\left(x_i\right)]$

Different F distribution leads to different binary regressions. Different link functions can be formed as a generalized linear model taking the link function as logistic, probit and complementary log–log, as shown in Table 1. After fitting binary regression models, propensities can be predicted and these can be used for non-response adjustments either individually or by grouping units into classes using equation 10 and 11 by substituting $\hat{\phi}_i$ for ϕ_i [10].

1. Propensity weighting - Adjust the weight for an individual responding by $1/\hat{\phi}_i$ computed from a binary regression.
2. Propensity stratification - Use $\hat{\phi}_i$ to create classes and make a common adjustment for each class $\hat{\phi}_c$. Five classes are usually recommended. General steps to create classes are;
   (a) Calculate $\hat{\phi}_i$ for each unit in the sample.
   (b) Sort it from low to high.
   (c) Form classes with about same number of sample units in each The common adjustment for the classes is selected among mean, weighted mean, response rate, weighted repose rate and median.

2.3 Adjustments for Coverage Errors

After adjusting weights for the survey design and non-response, the final step is to use the auxiliary data to correct the coverage errors and to reduce the standard errors. When marginal totals of grouping auxiliary variables were known for the population, it can be used to reduce coverage errors using methods such as post-stratification, raking and calibration. These three methods are briefly discussed. It should be noted that these final weights do not depend on the response variable and thus it would be the same for estimating the population mean for any variable [6].

2.3.1 Post-Stratification

Post-stratification adjusts the sampling weights so that the estimated population group sizes are correct, as they would be in stratified sampling. If N_h were known, but the sample was not stratified, the estimated population group sizes would not be correct. The postStratify( ) function in the R survey package [5] was used to compute these estimates.

2.3.2 Raking

Raking is a method that could be used when post-strata are formed using more than one variable, but only the marginal population totals are known. It also can be applied when partial cross-classifications of the variables are available. Raking involves post-stratifying on each set of variables in sequence and repeating this process until the weights stop changing. The rake( ) function in the R survey package performs the computations for raking by repeatedly calling the function postStratify( ).

2.3.3 Calibration

Calibration to reduce bias [9]. Calibration weights are computed using a calibration equation and a distance function. Calibrate( ) function in the R survey package was used to compute calibrated weights to estimate the population mean.

Let g_i be the final weight adjustment from these three methods. Then, the final weight of the i_{th} individual is,

$$w_i^f = \frac{1}{\pi_i}\frac{1}{\hat{\phi}_i}g_i. \tag{15}$$

Having survey weights that vary is common; however, highly differential weights can increase the variances of estimates even if they decrease bias. Several procedures are often used to trim extreme weights [7]. Trim weights approach was considered where large weights are arbitrarily trimmed back to an upper bound. The total weight trimmed away is then spread among the other sample units. The next section illustrates the application of the theories and methods discussed in this section.

3 Results: Weight Adjustments

3.1 Sampling Weights

A stratified sampling scheme was used to obtain the sample from the target population where the universities were chosen as the strata. The sample size for each stratum was computed using Neyman allocation method. According to equation 1 and 4, the sampling weights were calculated. As depicted in Table 2, these sampling weights are quite similar to one another.

Table 2: Sampling weights.

University	n_h	N_h	$W_h = \frac{1}{\Pi_h}$
University of Colombo (CMB)	83	581	7
Eastern University, Sri Lanka(EUSL)	56	373	6.66
University of Jaffna(JFN)	61	454	7.44
University of Kelaniya (KLN)	168	1195	7.11
University of Peradeniya (PDN)	116	817	7.04
University of Ruhuna (RHN)	101	662	6.55
Rajarata University of Sri Lanka(RUSL)	28	187	6.68
South Eastern University of Sri Lanka(SEUSL)	35	237	6.77
University of Sri Jayewardenepura (SJP)	135	914	6.77
Sabaragamuwa University of Sri Lanka(SUSL)	27	197	7.3

3.2 Weight Adjustments for Non-Response

The response rate of 48% was achieved only after several follow up calls. Even though it is somewhat high compared to most similar studies, it is possible that non-response bias may exist in this survey data. University, gender and ethnicity were the only variables that were available as para data to adjust for non-response bias.

3.2.1 Propensity Weighting

The response rate of the sample is varied with gender, ethnicity and university. Therefore, a propensity adjustment score model was fitted for the para data by taking the response indicator as the response variable and these three variables as explanatory variables. The Probit, logit and complementary log -log is depicted in Table 1 were used as the link functions. The AIC values for the three models were: logistic, 1361.2; probit, 1361.21; and c–log–log, 1361.9. All three did fit equally well. AIC measures and the model deviance was also significant implying the model fitted well.

Hence, the logit model was chosen as the best model with the least AIC. While the coefficient of gender was significant at 10%, the coefficients of Ethnicity were significant at 12%. Except for the coefficient of RUSL and KLN universities, others were significant at 10%. When the full model with interaction terms is considered, all interactions were insignificant. Hence, the logit model with only the main effects gender, ethnicity and university was considered as the final model. Individual predicted response probabilities were obtained and they were used to calculate the adjusted weights for non-response.

3.2.2 Propensity Stratification

Next, these predicted individual propensities were classified into five classes as recommended by [2]; Figure 1, depicts the boxplots of these predicted probabilities based on logistic regression after sorting into five propensity classes. The predicted probabilities of all propensity classes have skewed distributions. Table 3, represents the values for the five estimates for the class propensities. Weighted and unweighted estimates are quite similar. Since the data are skewed, mean is not a proper estimate for the class propensity. According to the boxplots, median too does not seem to be suitable since the outliers are not captured well. Hence, the response rate is the best estimate since it uses the actual response rate of the class as the class propensity. The weighted and unweighted response rate are approximately equal and since the observations are weighted using sampling weights earlier, unweight response rate was considered as the class propensity to adjust the weights for non-response.

Table 3: Five estimation of class propensities.

Propensity Class	Mean	Weighted Mean	Median	Response Rate	Weighted Response Rate
[0.138,0.412]	0.362	0.362	0.392	0.368	0.366
(0.412,0.474]	0.454	0.454	0.474	0.445	0.446
(0.474,0.487]	0.481	0.481	0.486	0.496	0.495
(0.487,0.539]	0.538	0.538	0.539	0.531	0.532
(0.539,0.748]	0.572	0.572	0.548	0.578	0.577

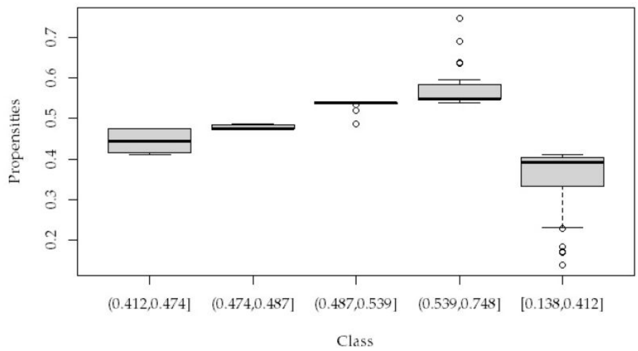

Figure 1: Predicted probabilities by propensity classes.

3.3 Weight Adjustments for Coverage Errors

Although the population information was available from the census data, there were item non–responses for the almost all of the variables except University. Thus, it was difficult to extract necessary information from the census data. However, the sampling frame had information about the gender and ethnicity of the population. Hence, the marginal population totals of ethnicity, gender and University and the cross information of the three variables were available. Therefore, population information about university, ethnicity and gender were used to adjust weights for coverage errors.

The last step, which is extremely important in many surveys, is to use auxiliary data to correct coverage problems and to reduce standard errors. Post-stratification, raking and calibration are the most frequently used approaches to deal with the final step of adjustments. Three types of initial weights were tested to adjust for the coverage errors. Those are,

i. sampling weights without adjusting for non-response.

ii. sampling weights adjust for non-response through propensity weights.

iii. sampling weights adjust for non-response through propensity stratification.

Several combinations of the university, gender, ethnicity variables were considered for the adjustments. For post-stratification, marginal population counts of each variable were considered individually. However, for raking and calibration following combinations were tested.

- Marginal gender counts
- Marginal ethnicity counts
- Marginal university counts
- Marginal gender, marginal ethnicity and marginal university
- Marginal gender and marginal ethnicity
- Gender and ethnicity cross cell counts

- Marginal gender and marginal university
- Gender and university cross cell counts
- Marginal university and marginal ethnicity
- Gender and university cross cell counts and marginal ethnicity

It should be noted that 'university and ethnicity cross cell counts'and 'gender, ethnicity and university cross cell counts'were not considered due to zero counts for several cells for population and sample. After adjusting weights using these approaches, the final weights were used to estimate the mean waiting time for the first employment response variables.

Table 4: Mean waiting time for the first employment estimates and standard errors.

Variables used for Adjustments	Sampling weights		Sampling weight-adjusted for Non-response using propensity Stratification		Sampling weight adjusted for Non-response using propensity weight	
	mean	SE	mean	SE	mean	SE
-	15.229	0.6503	15.159	0.6605	15.25	0.6679
Post-stratification						
Gender	15.392	0.6623	15.253	0.6657	15.347	0.6742
Ethnicity	15.229	0.6668	15.203	0.6802	15.15	0.6784
University	15.045	0.6572	15.13	0.6647	15.189	0.668
Raking						
Gender	15.391	0.6622	15.252	0.6656	15.346	0.6742
Ethnicity	15.224	0.6672	15.199	0.683	15.189	0.6823
University	15.045	0.6572	15.13	0.6649	15.147	0.6685
Gender, Ethnicity, University	15.225	0.6669	15.2	0.6649	15.241	0.6661
Gender, Ethnicity	15.437	0.6791	15.319	0.6858	15.267	0.6853
Gender, Ethnicity cross	15.42	0.6891	15.331	0.6955	15.284	0.6943
Gender and University	15.228	0.6723	15.235	0.6713	15.288	0.6752
Gender and University cross	15.166	0.6609	15.167	0.6616	15.213	0.663
University and Ethnicity	15.041	0.6519	15.102	0.6596	15.152	0.661
Gender and University cross and marginal Ethnicity	15.227	0.6599	15.213	0.6612	15.239	0.6613
Calibration						
Gender	15.392	0.6623	15.253	0.6657	15.347	0.6742
Ethnicity	15.229	0.6668	15.203	0.6802	15.15	0.6784
University	17.612	1.1131	17.714	1.1273	17.831	1.1301
Gender, Ethnicity, University	15.233	0.6669	15.209	0.6646	15.252	0.666
Gender, Ethnicity	15.437	0.679	15.319	0.6839	15.267	0.6823
Gender, Ethnicity cross	15.437	0.679	15.319	0.6839	15.267	0.682
Gender and University	15.228	0.6722	15.235	0.6709	15.288	0.6745
Gender and University cross	15.166	0.6609	15.167	0.6614	15.213	0.6626
University and Ethnicity	15.047	0.6517	15.108	0.6591	15.158	0.6607
Gender and University cross and marginal Ethnicity	15.208	0.6581	15.19	0.6584	15.224	0.6567

Table 4 depicts the estimates and the standard error for all of these approaches under the three initial weights. According to the Table 4, standard errors of the estimates increase when the sampling weights are adjusted for the non-response (yellow colored cells). Even though standard errors should be reduced if post-stratification, raking and calibration are well performed, the standard errors of the estimates (green colored cells) were reduced only for a few of the combinations. Only the raking and calibration with two variable combinations reduced standard errors. It means only those auxiliary variable combinations are related to the waiting time for the first employment and also less extreme weights are received for those variable combinations. Furthermore, it implies that the use of gender, ethnicity and university as auxiliary variables makes the final weights to be more reflected the target population characteristics. Among all the final weight adjustments, the least error was given for the calibration by gender and university cross-cell counts and marginal ethnicity counts for the initial weight of sampling weight adjusted for non-response using propensity weight. Thus, this least standard error given weights were selected as the final weight adjustment. However, when comparing the least standard error (0.6567) obtained for the mean waiting time for the first employment after adjusting for stratification, non-response and coverage error is higher than the standard error received after adjusting only for stratification (0. 6503). Which implies that non-response bias and coverage error has underestimated the standard error of the response variable.

It is vital to recognize the weight distribution of the final weights of the selected approach. Figure 2, illustrates the histogram and boxplot of the calibrated weights by gender and university cross-cell counts and marginal ethnicity counts for the initial weight of sampling weight adjusted for non-response using propensity weight. While the majority of the weights are less than 25, there were two extreme weight values.

Due to these extreme weights, it deemed necessary to trim these extreme weights. However, when the weights are trimmed, weights did not match with the population totals. As a solution, a recalibration was done to the trimmed weights and all the final weights after recalibration were below 37. Then the mean waiting time was estimated as 15.19 months with a standard error of 0.657.

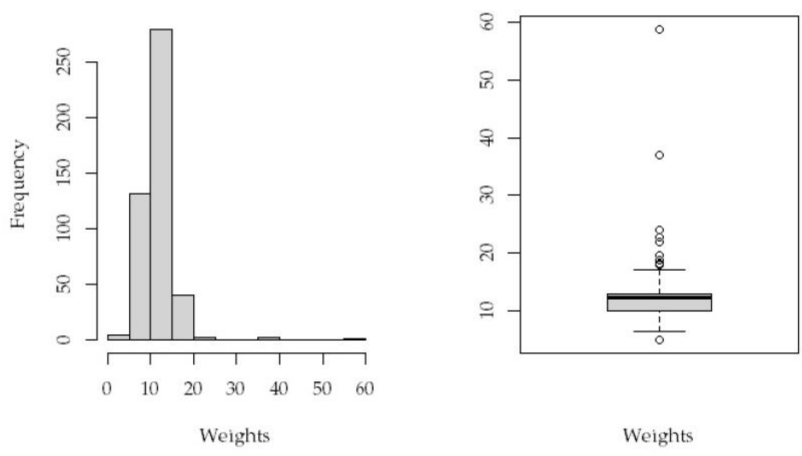

Figure 2: Weight distribution of the calibrated weights.

4 Discussion

The main objective of this study is to estimate the population mean waiting time for the first employment after graduation. As discussed in the introduction section, for complex survey designs, the weighted sample mean is one of the best estimators for the population mean. To achieve the objective of this study, weights were calculated in three main steps. Initial weights consisted of the sampling weights. Since the survey was carried out using a stratified random sampling taking

University as the strata, weights were identified for each university. After much effort, only 48% of the sample responded to the survey. Hence, a non-response bias and coverage error may exist in the data. Therefore, the initial weights were adjusted for the non-response bias and coverage errors in two steps. First, the propensity adjusted score method was performed fitting a logistic regression model to the para data by using university, gender and ethnicity as explanatory variables and response indicator as the response variable. The predicted response rate was used to create propensity classes. Thus non-response adjustment was performed using two approaches using the propensity weight and propensity stratification. Secondly, to overcome the coverage errors, auxiliary variables were used to match with the population totals. Post-stratification, raking and calibration were the methods used to adjust weights for the population totals of the university, gender and ethnicity. Only raking and calibration for a few auxiliary variable combinations reduced standard errors.

The main challenge faced was to find para data to compensate for non-response bias. After much effort, only three variables were obtained as para data. When the logistic model was fitted to the para data all interaction terms were insignificant while only the main effects were significant. However, all the main effects were significant at 12% but not at 5%. Due to the nonexistence of true population estimates to calculate the actual bias, only the standard error was used to identify the best weighting method. Few extreme weights were given for Tamil graduates for some of the calibrated weights since only a few Tamils responded from the universities, which consists of Sinhala graduates mostly. In addition, the weights of the University of Colombo was higher compared to other universities because the response rate was the lowest for it. Since there were no similar studies in local universities or in foreign universities in the world no comparisons could be performed. Hence, this study provides an initial foundation for future population mean estimations based on weight adjustments.

5 Concluding Remarks

When considering the weight adjustment for non-response, the standard error was less for the propensity stratification than propensity weight suggesting the class propensity prediction over individual propensity prediction. However, if the population totals of auxiliary variables are available, these weights can be adjusted again to compensate for the coverage errors. It showed that calibration is the best method to reduce standard errors of mean estimation over post-stratification and raking. Among all the final weight adjustments, the least standard error was given for the calibration by gender and university cross-cell counts and marginal ethnicity counts for the initial weight of sampling weight adjusted for non-response using propensity weight. When using many auxiliary variable totals for calibration it may provide a less standard error of estimates and captures the population variation well. If there are a few extreme weights, it is better to recalibrate from the same method after trimming those extreme weights. Finally, the waiting time was estimated from the final weights, as 15.19 months with a standard error of 0.657.

Acknowledgement The authors would like to thank the University of Colombo for funding this research under its Research Grant 2016 and also the National Center for Advanced Study in Humanities and Social Sciences (NCAS) for postgraduate research grant 2016. Further, the authors are grateful to all the Arts graduates responded to this survey.

Conflicts of Interest The authors declare no conflict of interest.

References

[1] J. M. Brick (2013). Unit nonresponse and weighting adjustments: A critical review. *Journal of Official Statistics*, *29*(3), 329–353.

[2] W. G. Cochran (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics*, *24*(2), 295–313.

[3] D. L. Hahs-Vaughn (2005). A primer for using and understanding weights with national datasets. *The Journal of Experimental Education*, *73*(3), 221–248.

[4] I. T. Jayamanne & K. A. Ramanayake (2017). A study on the waiting time for the first employment of arts graduates in Sri Lanka. *International Journal of Computer and Information Engineering*, *11*(12), 1153–1161.

[5] T. Lumley (2004). Analysis of complex survey samples. *Journal of Statistical Software*, *9*(8), 1–19.

[6] T. Lumley (2010). *Complex surveys: A guide to analysis using R*. Wiley, Hoboken, New Jersey.

[7] F. Potter & Y. Zheng (2015). Methods and issues in trimming extreme weights in sample surveys. In *Proceedings of the American Statistical Association, Section on Survey Research Methods*, pp. 2707–2719. American Statistical Association, Alexandria.

[8] A. Ramanayake, I. T. Jayamanne, Y. Ramyadevipriya & K. L. Perera (2013). *Graduand employment census 2012*. Ministry of Higher Education, Colombo.

[9] C. E. Särndal & S. Lundström (2005). *Estimation in surveys with nonresponse*. Wiley, Chichester, West Sussex.

[10] R. Valliant, J. A. Dever & F. Kreuter (2013). *Practical tools for designing and weighting survey samples*. Springer, Switzerland AG.