International Journal of Mathematical Analysis and Optimization: Theory and Applications
Vol. 2019, No. 1, pp. 512 - 520

A Generalized Regression Estimation of the Item Sum Technique in Sensitive Surveys

Alakija T. O. [1*], Adeleke I. A. [2], Adekeye K. S. [3], Adamu M. O. [4]

1*, Department of Statistics, Yaba College of Technology, Yaba, Lagos, Nigeria.
2, Department of Actuarial Sciences, University of Lagos, Lagos, Nigeria.
3, 4, Department of Mathematics, University of Lagos, Lagos, Nigeria.
*Corresponding author: lawaltope2003@yahoo.com

Article Info
Received: 22 May 2019 Revised: 30 July 2019
Accepted: 06 August 2019 Available online: 14 August 2019

Abstract

Survey researchers often find it difficult to collect reliable data of human populations, yet the validity of any research depends mainly on the accuracy of self-reported behavior especially when the respondents are to reflect about sensitive issues or highly personal matter. It is therefore important to develop methods of improving interviewees responses in any survey. The Item Sum Technique (IST) is the most recent indirect questioning method and it is a variant of the Item Count Technique (ICT) which can be used only for qualitative responses. The aim of this study is to estimate the sensitive characteristic when using the IST especially if two or more sensitive questions are investigated. It also focuses on the theoretical framework which includes the introduction of a classical method called the Generalized Regression model (GREG) using the IST. The efficiency of the GREG method was ascertained in comparison to the Calibration estimator by an extensive simulation study. Results from the statistical analysis indicates that the GREG estimator competes well with the calibration method and can further be used for a small sample size or data that is not normally distributed.

Keywords: Calibration estimator, Indirect questioning method, Item Count Technique, Sensitive questions, small sample.
MSC2010: 62D05

1 Introduction

The crucial role of survey depends solely on the accuracy and reliability of data from which vital information is gotten. However, collection of data on human populations by means of sample surveys is a very difficult task [14]. Survey researchers often find it difficult to collect reliable data due to various sources of non-sampling error [6]. The problems associated with the collection of valid data need not be ignored, since if not properly handled will result into inaccurate predictions and invalid inferences. Survey may often times contain stigmatizing issues of enquiry that it is hard to get a valid and reliable information. Several works have been done in an effort to tackle this problem (see [7] and [18]). Warner [19] introduced the first ingenious interviewing indirect questioning technique which is called the Randomized Response Technique (RRT). This technique aims to encourage truthful answers from respondents and also preserve respondents' confidentiality. The RRT deals with a simple principle of the question and answer method using a randomized device. Lensvelt-Mulders *et al.*, [10] noted that the RRT estimates is more valid than estimates gotten through the use of direct questioning. Review texts on randomized response techniques can be found in [5]. Despite the apparent advantages of the RRT questioning, not all studies have supported its alleged superiority over conventional questioning methods [2]. According to some studies, in the application of the RRT (e.g., [3] and [11]), estimates obtained by the RRT is not different from those obtained through direct questioning. In other studies, estimates obtained by the RRT were even lower from those obtained through direct questioning (e.g., Holbrook and Krosnick [8]). Arijit and Tasos [4] also noted that the RRT demand too much active cooperation and understanding from the respondents and that it also takes time. The ICT also known as the list experiment was originally proposed by [13] which provides much secrecy and confidentiality on sensitive attribute. In an item count design, there are k statements of behaviors which are referred to the non-key or non-sensitive items and one statement of behavior related to the sensitive characteristic which is referred to the key or sensitive item [7]. The gap however noticed in the ICT is its limitation to responses that are qualitative in nature as there could be some real-life situation that demands for a quantitative response rather than a dichotomous one. The IST method was developed to fill the gap in the ICT. The IST is a very recent indirect questioning method which was proposed by Chaudhuri & Christofides [5] as a generalization of the ICT to tackle situations whereby the sensitive variables are quantitative in nature and to provide maximum degree of privacy protection [20]. The IST has been used by few researchers e.g. (see [6], [12] and [15]). It has several advantages which include; a randomizing device not been required; the cognitive effort demanded from respondents is relatively low; implementation is easily possible in both interviewer and self-administered interviews [17]. In this paper, the work of Maria del Mar *et al.*,[12] on an advancement in methodologies using the Horvitz Thompson and calibration estimators was extended by introducing a GREG method of estimation for a two-stage sampling design. The limitation in the calibration method of estimation is its inability to perform well with large with small area estimation hence the need to introduce a method that fits well for both small area estimation and large domains.

2 The Horvitz Thompson Type Estimator for the IST.

The Horvitz Thompson estimator is use to estimate the total population of respondents as the technique is design-based [1]. The design based approach for estimation in any survey inference can be used for many designs. Two independent samples, S_1 and S_2, are selected from U according to the two-stage cluster sampling design. One of the samples, say S_1, is confronted with a long list (LL) of items containing $(\Lambda + 2)$ questions of which Λ refers to nonsensitive characteristics and 2 is related to the sensitive variables under study. The other sample S_2 receives a short list (SL) of items that only contains the Λ innocuous questions. All sensitive and nonsensitive items are quantitative variables. Total score of responses in both samples are recorded for all the items applicable to all independent respondents.
Let w be the variable denoting the total score applicable to the Λ nonsensitive questions.
Also, let y be the variable denoting the total score applicable to the sensitive questions, then

$$z = y + w \tag{2.1}$$

where z is the total score applicable to the nonsensitive questions and the sensitive question. Hence, the answer of the ith respondent will be

$$z_j = \{^{y_j+w_j, \quad if\ j\in S_1}_{w_j, \quad if\ j\in S_2} \tag{2.2}$$

We observe that for $\Lambda = 1$, the variable w simply denotes the innocuous variable and w_j its value on the jth unit.

Given

$$\hat{\hat{Z}}_{ht} = \frac{1}{N}\sum_{j\in S_1} d_j z_j \tag{2.3}$$

and

$$\hat{\hat{W}}_{ht} = \frac{1}{N}\sum_{j\in S_2} d_j w_j \tag{2.4}$$

Such that equation(3) and equation(4) be the unbiased Horvitz-Thompson estimators for LL and SL respectively. Hence, a Horvitz–Thompson-type estimator of $\bar{Y}$ can be immediately obtained as:

$$\hat{\hat{Y}}_{ht} = \hat{\hat{Z}}_{ht} - \hat{\hat{W}}_{ht}. \tag{2.5}$$

where d_j denotes the two stage sampling design basic weight for unit $j \in U$. The estimators in (3) and (4) are unbiased, therefore the estimator in (5) is also unbiased and its expectation is given as shown below,

$$E_p(\hat{\hat{Y}}_{ht}) = E_p(\hat{\hat{Z}}_{ht}) - E_p(\hat{\hat{W}}_{ht}). \tag{2.6}$$

The variance of (5) is given by

$$V_p(\hat{\hat{Y}}_{ht}) = V_p\left(\hat{\hat{Z}}_{ht} - \hat{\hat{W}}_{ht}\right)$$

$$= V_p(\hat{\hat{Z}}_{ht}) + V_p(\hat{\hat{W}}_{ht}) \tag{2.7}$$

$$= \frac{1}{N^2}\left(\sum_{j\in U}\sum_{j\in U}\Delta_{i,j}(d_i z_i)(d_j z_j) + \sum_{i\in U}\sum_{j\in U}\Delta_{i,j}(d_i w_i)(d_j w_j)\right) \tag{2.8}$$

where $\Delta_{ij} = \pi_{ij} - \pi_i\pi_j$. The unbiased estimator of (8) is given as

$$\hat{V}_p(\hat{\hat{Y}}_{ht}) = \frac{1}{N^2}\left(\sum_{i\in S_1}\sum_{j\in S_1}\hat{\Delta}_{i,j}(d_i z_i)(d_j z_j) + \sum_{i\in S_2}\sum_{j\in S_2}\hat{\Delta}_{i,j}(d_i w_i)(d_j w_j)\right) \tag{2.9}$$

where $\hat{\Delta}_{i,j} = \frac{\Delta_{ij}}{\pi_{ij}} = \frac{\pi_{ij}-\pi_i\pi_j}{\pi_{ij}} = 1 - \frac{\pi_i\pi_j}{\pi_{ij}}$, for $i \neq j$. However, if $i = j$, we have $\hat{\Delta}_{i,i} = 1 - \pi_i$

2.1 Generalized Regression(GREG) Estimator

The GREG estimator is a model assisted estimator which are mostly used to estimate the totals for population subgroups or domains. It is designed to improve the accuracy of the estimates by means of auxiliary information. GREG estimator ensures the coherence between sampling estimates and known totals of the auxiliary variables. The growing availability of parameters derived from population data, and previous surveys provides a wide range of variables that can be used to increase the efficiency of the statistic from sample data. The choice of a GREG estimator is as a result of its robustness to model choice since it is asymptotically design unbiased. The GREG approach,requires a working population model to find the efficient design-consistent estimators. Let the working model be a linear regression model of the form

$$y_i = x_i^{'}\beta + e_i,\ i = 1, ..., N \tag{2.10}$$

where,
e_i is the assumed uncorrelated model errors
and x_i is a vector of auxiliary variables with known population total X. The regression coefficient of (10) is given by

$$B = \left(\sum_{i\in U} \frac{x_i x_i^{'}}{\sigma_i^2}\right)^{-1} \left(\sum_{i\in U} \frac{x_i y_i}{\sigma_i^2}\right) \tag{2.11}$$

and the corresponding residuals are

$$e_i = y_i - \hat{y}_i = y_i - x_i^{'}\beta \tag{2.12}$$

Given the total as

$$Y = \sum_{i\in U} \hat{y}_i + \sum_{i\in U} e_i$$

where, $e_i = y_i - \hat{y}_i$ denotes the prediction error,then the estimator of Y is given by

$$\hat{Y} = \sum_{i\in U} \hat{y}_i + \sum_{i\in s} d_i e_i$$

This estimator can be expressed as a generalized regression (GREG) estimator given by,

$$\hat{Y}_{GREG} = \hat{Y} + \beta^{'}(X - \hat{X})$$

where $\hat{X} = \sum_{i\in s} d_i x_i$ (Sarndal, Swensson and Wretman, 1992).

In order to carry out the estimation, certain assumptions must be stated.

(i) Sub-sampling of selected i clusters is carried out.
(ii) The auxiliary variable, x_j is known only for elements j such that

$$j \in \bigcup_{j\in S} U_i$$

Consider the general case of auxiliary information. Suppose the GREG estimator is calculated at the cluster level under the model in equation (10) above, then the sample fit of the model in equation (10) is based on the data points $(y_i, x_i)\ \forall\ i \in S$, where s is the total sample given by

$$S = \bigcup_{j\in sl} s_j \tag{2.13}$$

The number of such points is

$$n_s = \sum_{sl} n_i \tag{2.14}$$

where n_i is ordinarily a random number. The estimate of equation (11) is given by

$$\hat{B} = \left(\sum_{j\in S} \frac{x_j x_j^{'}}{\sigma_j^2 \pi_j}\right)^{-1} \left(\sum_{j\in S} \frac{x_j y_j}{\sigma_j^2 \pi_j}\right) \tag{2.15}$$

where the sampling weights is $\frac{1}{\pi_j}$ The variance of equation (15) is then given as

$$Var(\hat{B}) = Var(y) \left(\sum_{j\in S} \frac{x_j x_j^{'}}{\sigma_j^2 \pi_j}\right)^{-1} \tag{2.16}$$

Theorem 1. The generalized regression estimator of modeling at the element level of the IST in a two stage sampling design is a weighted regression least square estimator.

$$\hat{\beta} = \left[\sum\sum_{j\in S} \frac{x_j x_j'}{\sigma_i^2 \pi_{ij}}\right]^{-1} \sum_{j\in S} \frac{x_j(z_{ij} - w_{ij})}{\sigma_j^2 \pi_j} \tag{2.17}$$

Proof:
Given

$$y_{ij} = x_{ij}' + e_{ij} \tag{2.18}$$

The distribution of y_{ij} is dependent on z_{ij} and w_{ij}
such that,

$$y_{ij} = z_{ij} - w_{ij}. \tag{2.19}$$

$$z_{ij} - w_{ij} = x_{ij}' + e_{ij}$$

$$E(z_{ij} - w_{ij}) = E[x_{ij}'\beta] + E(e_{ij}),\ E(e_{ij}) = 0$$

Then

$$E(z_{ij} - w_{ij}) = E(y_{ij}) = x_{ij}'\beta + e_{ij}$$

Minimizing e_{ij},

$$\sum_j\sum_i e_j' e_j = \sum_j\sum_i [(z_j - w_j) - x_{ij}'\beta]'[(z_j - w_j) - x_j'\beta] \tag{2.20}$$

To obtain an estimate for β, differentiate equation (20) with respect to β and equate the result to zero.
Thus we have,

$$\hat{\beta}\sum_j\sum_i x_j x_j' = \sum_j\sum_{i\in s_{ll}} x_j z_j - \sum_j\sum_{i\in s_{ll}} x_j w_j \tag{2.21}$$

Assuming that z_j, w_j, x_j are from two stage cluster design with probability that $i \in P(S)$ be π_i and that $i, j \in P(S)$ be π_{ij}. Then equ(21) becomes

$$\hat{\beta}\sum_{j\in S_{LL}}\sum_{i\in S_{SL}} \frac{x_{ij}x_{ij}'}{\sigma_i^2\pi_{ij}} = \sum_{j\in S_{LL}}\sum_{i\in S_{SL}} \frac{x_{ij}z_{ij}}{\sigma_i^2\pi_{ij}} - \sum_{j\in S_{LL}}\sum_{i\in S_{SL}} \frac{x_{ij}w_{ij}}{\sigma_i^2\pi_{ij}}$$

$$\hat{\beta} = \left[\sum_{j\in S_{LL}}\sum_{i\in S_{SL}} \frac{x_{ij}x_{ij}'}{\sigma_i^2\pi_{ij}}\right]^{-1} \left[\sum_{j\in S_{LL}}\sum_{i\in S_{SL}} \frac{x_{ij}z_{ij}}{\sigma_i^2\pi_{ij}} - \sum_{j\in S_{LL}}\sum_{i\in S_{SL}} \frac{x_{ij}w_{ij}}{\sigma_i^2\pi_{ij}}\right] \tag{2.22}$$

where σ_i^2 is the variance of the ith selected cluster, estimated by S_i^2

$$\hat{\beta} = \left[\sum_{j\in S_{LL}}\sum_{i\in S_{SL}} \frac{x_{ij}x_{ij}'}{\sigma_i^2\pi_{ij}}\right]^{-1} \left[\sum_{j\in S_{LL}}\sum_{i\in S_{SL}} \frac{x_{ij}y_{ij}}{\sigma_i^2\pi_{ij}}\right]. \tag{2.23}$$

3 Application

A study was carried out on 25 sampled respondents (students) using the Item Sum Technique (IST), consisting of two sensitive and two innocuous questions. The two sensitive questions focus on the use of substance abuse and internet gambling among students. The survey was carried out to know the distribution of the variables of interest where Y_1 denotes use of substance abuse, Y_2 denotes internet gambling and X is the auxiliary variable. The results of the Exploratory Data Analysis using R-language is presented in Table 1 .

Table 1: Summary statistic for Data from the Field Study

Variable	Min	Q1	Median	Mean	Q3	Max	Skewness	Kurtosis
Y_1	1.200	2.200	3.400	10.370	7.000	88.500	3.140	12.563
Y_2	1.500	3.333	5.833	7.020	9.000	22.333	1.421	4.282
X	1.000	3.000	4.667	4.439	6.000	7.667	-0.086	2.167

To further test for the normality of the data collected from the field, the Shapiro-Wilk test for normality was employed. The result is presented in Table 2.

Table 2: Shapiro-Wilk Normality Test from Field Study

Variable	W	P-value
Y_1	0.49904	3.66e-08
Y_2	0.83237	0.0008299
X	0.96615	0.5497

To measure the performance of the Generalized regression Estimation described in 2.1 and to compare the result with the existing calibration method, the data collected on abuse of substance and internet gambling was used. Furthermore, a simulation study was conducted to validate the results obtained with the real-life data. The parameter estimate along with their corresponding standard Error, P-value and Performance Index (using the Akaike Information Criteria and Mean Square Error) for the real life data are presented in Table 3.

Table 3: Performance of GREG and Calibration Methods for Data from Field Study

Model	Variables	Parameter		Standard Error		P-value		AIC	MSE
		β_0	β_1	β_0	β_1	β_0	β_1		
Model1	Y_{1GREG}	4.881	1.236	3.877	2.094	0.0135	0.5608	222.98	344.6402
Model2	Y_{2GREG}	2.721	0.9683	1.0532	0.5688	8.45e-07	0.101	156.83	24.4128
Model5	Y_{1CALIB}	10.368	1.236	10.072	2.094	0.633	0.561	223.09	345.7402
Model6	Y_{2CALIB}	7.020	0.9683	2.7361	0.5688	0.330	0.102	157.93	25.5128

3.1 Simulation Study

This section makes use of simulation studies to evaluate the performance of the GREG and Calibration methods.This is to determine the model that best fit the simulated data for small sample size and large sample size which ranges between n = 10 to 1000. To achieve this, n = 10 to n = 1000 artificial observations are generated. It is assumed that the data are observed from a bivariate normal distribution with mean and standard error given as μ = (10.370, 7.020) and

σ = (365.601, 29.924), respectively. The values generated are then used to define the total score variable $z = y + w$ and to obtain an estimate of y using the values of z and w in the Horvitz Thompson Estimator as defined in Equation(5). The results of the parameter estimates using the Akaike Information Criteria and Mean Square Error for the simulation study are presented in Table 4.

Table 4: Parameter Estimation from Simulation Study for n = 10, 250,1000

		Parameter		Standard Error		P-value			
n = 10									
Model	**Var Y_1**	β_0	β_1	β_0	β_1	β_0	β_1	**AIC**	**MSE**
Model1	GREG	0.01	0.08	0.04	0.01	0.06	5.78e-06	-28.03	0.002
Model3	CALIB	3.04	8.29	2.46	1.23	0.2517	0.0001	223.1	278.7
Model	**Var Y_2**	β_0	β_1	β_0	β_1	β_0	β_1	**AIC**	**MSE**
Model2	GREG	0.07	0.23	0.37	0.06	0.8633	0.0074	14.69	0.139
Model4	CALIB	6.37	2.57	0.71	0.36	1.96e-05	9.00e-05	157.93	47.76
n = 250									
		Parameter		Standard Error		P-value			
Model	**Var Y_1**	β_0	β_1	β_0	β_1	β_0	β_1	**AIC**	**MSE**
Model1	GREG	0.166	0.087	0.004	0.001	<2e-16	<2e-16	-1101.8	0.0007
Model3	CALIB	10.64	10.68	0.125	0.067	<2e-16	<2e-16	223.09	494.58
Model	**Var Y_2**	β_0	β_1	β_0	β_1	β_0	β_1	**AIC**	**MSE**
Model2	GREG	0.322	0.223	0.008	0.002	<2e-16	<2e-16	-765.69	0.0027
Model4	CALIB	7.667	2.952	0.037	0.019	<2e-16	<2e-16	157.93	66.697
n = 1000									
Model	**Var Y_1**	β_0	β_1	β_0	β_1	β_0	β_1	**AIC**	**MSE**
Model1	GREG	0.141	0.089	0.002	0.001	<2e-16	<2e-16	-4343.6	0.0008
Model3	CALIB	10.95	9.86	0.036	0.019	<2e-16	<2e-16	223.08	442.68
Model	**Var Y_2**	β_0	β_1	β_0	β_1	β_0	β_1	**AIC**	**MSE**
Model2	GREG	0.147	0.251	0.006	0.001	<2e-16	<2e-16	-2308.4	0.0058
Model4	CALIB	7.044	2.934	0.009	0.005	<2e-16	<2e-16	157.93	58.313

4 Discussion of Results

The results presented in Table 1 shows that the variables y_1 and y_2 are not Gaussian while the auxiliary variable x is approximately Gaussian as shown by their measure of skewness and kurtosis. Furthermore, Table 2 shows that the variables y_1 and y_2 are not Gaussian since their p-values are less than 0.05 while the auxiliary variable x is Gaussian, since its p-value is greater than 0.05. All tests were carried out at 5% level of significance in this study.

Four models were fitted using GREG and Calibration model, and the results are presented in Table 3. Model 1 from Table 3 is the regression of y_1 on x using GREG, Model 2 is the regression of y_2 on

x using GREG, Model 3 is the regression of y_1 on x using Calibration method and Model 4 is the regression of y_2 on x using Calibration method. GREG estimators used Gaussian family while for the Calibration method, the Gamma family was used.Table 4 shows that using AKaikie information criterion(AIC) and the Mean square error (MSE) criterion, the GREG performed better than the calibration model for small sample size.

The obtained results for n = 10 (small sample), n = 250 (medium sample) and n = 1000 (large sample) are presented in Table 4. The Table 4 further shows that Model 1 and Model 2 are the best model using AIC and MSE criteria for n = 10. It also shows that Model 1 and Model 2 are the best models using the two criteria for n = 250. For n = 1000, Model 1 and Model 2 which are the GREG method of estimation outperformed the calibration method for large sample size. The P-value in Table 4 indicates that the coefficients are highly significant and consistent for n = 10, 250 and 1000 at P-value less than 0.05 for model 1 , 2, 3 and 4..

5 Conclusion

This paper aims at finding the most appropriate method for asking sensitive questions in survey. The need of such methods is very crucial since sensitive questions could have a large impact on the analysis of real-life situation due to non-response or falsified information. Indirect questioning is one of the ways of collecting data on sensitive questions and should be adopted when sensitive questions are to be asked. This study extends the most recent method called the IST by introducing the GREG method for its finite population estimation. The study further compared the GREG to the current existing calibration method to determine its efficiency. Evaluation of the robustness of the methods was introduced using simulation study. The simulation study shows the consistency and the asymptotic properties of the standard errors of the parameter estimates. Result from the study shows that the GREG is a better method of estimation as compared to the Calibration for both small sample size and large domain. The performance of the estimators used also depends on the distribution of the data. If the data is not necessarily normally distributed then GREG the estimator better used.

Acknowledgment

We are grateful to the reviewers for some comments on an earlier version of the manuscript which led to this improved version. We also acknowledge Prof. Ray Okafor, whose major contribution brought this manuscript to the limelight.

Competing financial interests

The author(s) declare no competing financial interests.

References

[1] Adeleke, I.A., Okafor, R.O., Esan, E.O. Horvitz-Thompson Theorem as a tool for Generalization of Probability Sampling Technique. *Journal of Scientific research and development*. 10, 125-134, (2007).

[2] Adrian, H. and Jochen, M. Assessing the validity of two indirect questioning techniques: A Stochastic Lie Detector versus the Crosswise Model. *Psychonomic Society, Inc. Behaviour Res*, 48:1032–1046, (2015).

[3] Akers, R.L., Massey, J., Clarke, W. and Lauer, R. M. Are self-reports of adolescent deviance valid? Biochemical measures, randomized-response, and the bogus pipeline in smoking-behaviour. *Social Forces*, 62, 234–251, (1983).

[4] Arijit, C. and Tasos C. C. Indirect Questioning in Sample Surveys. *Springer-Verlag Berlin Heidelberg.* DOI 10.1007/978-3-642-36276-7 1, (2013).

[5] Chaudhuri, A., and Mukerjee, R. Randomized response: Theory and techniques. *Marcel Dekker Publication,*New York, (1988).

[6] Chaudhuri, A., and Christofides, T. C. Indirect questioning in sample surveys. *Springer-Verlag, Berlin, Heidelberg, DE*, (2013).

[7] Guo-Liang T. and Man-Lai T. Incomplete Categorical Data Design. Non-Randomized Response Techniques for Sensitive Questions in Surveys. *Chapman and Hall/CRC Statistics in the Social and Behavioral Sciences Series.* New York, USA. (2014).

[8] Holbrook, A. L. and Krosnick, J. A. Measuring voters turnout by using the randomized response technique: evidence calling into question, the method's validity. *Public Opinion Quarterly,* 74, 328–343. Doi:10.1093/Poq/Nfq012, (2010).

[9] Horvitz, D. G. and Thompson, D. J. A generalization of sampling without replacement from a finite universe. *Journal of American Statistical Assoc.* 47, 663–685, (1952).

[10] Lensvelt-Mulders, G.J., Van der Heijden, P.G., Laudy, O. and Van Gils, G. Meta-analysis of randomized response research: thirty-five years of validation. *Sociological Methods and Research*, 33, 319–348. Doi:10.1177/0049124104268664, (2005).

[11] Locander, W., Sudman, S. and Bradburn, N. An investigation of interview method, threat and response distortion. *Journal of the American Statistical Association,* 71, 269–275, (1976).

[12] María del Mar, G.R., Pier, F.P. and Beatriz, R. C. Advances in estimation by the item sum technique using auxiliary information in complex surveys. *Springer-Verlag Germany, part of Springer Nature*, Doi.org/10.1007/s10182-017-0315-2, (2017).

[13] Miller, J. D. A new survey technique for studying deviant behavior. PhD. thesis. The George Washington University, Washington, DC. (1984).

[14] Sarjinder Singh. Randomized response techniques. *Model Assisted Statistics and Applications.*, 9, 1-2, (2014).

[15] Pier, F.P., Maria del Mar, R.G., Beatriz, C.R. Multiple sensitive estimation and optimal sample size allocation in the item sum technique. *Biometrical Journal*, 1–19, (2017).

[16] Sarndal, C.E., Swensson, B. and Wretman, J. *Model Assisted Survey Sampling.* Springer, New York. MR1140409. (1992).

[17] Trappmann, M., Krumpal, I., Kirchner, A. and Jann, B. Item sum: a new technique for asking quantitative sensitive questions. *J. Surv. Stat. Methodol.*, 2, 58–77, (2014).

[18] Vakilian, K., Mousavi, S.A., and Keramat, A. Estimation of sexual behaviour in the 18-to-24-years-old Iranian youth based on a crosswise model study. *BMC Research Notes*, **7**(28), 1–4, (2014).

[19] Warner, S. L. Randomized response: A survey technique for eliminating evasive answer bias. *J. of Amer. Statistical Association.* 60, 63-69, (1965).

[20] Zawar, H. and Naila, S. On some new randomized item sum techniques. *Communications in Statistics — Theory and Methods.* 1–14, (2017).