Statistics and Applications {ISSN 2452-7395(online)}
Volume 18, No. 2, 2020 (New Series), pp 85–95

Block Total Response Technique for Quantitative Sensitive Features in a Finite Population

Karabi Nandy[1] **and Bikas K. Sinha**[2]
[1] *Department of Population and Data Sciences, UT Southwestern Medical Center, Dallas, U.S.A.*
[2] *Formerly at Indian Statistical Institute, Kolkata, India.*

Received: 05 June 2020; Revised: 24 June 2020; Accepted: 26 June 2020

Abstract

The issue of eliciting truthful answers from survey respondents on sensitive questions has always been a challenge. Survey statisticians have developed various techniques to address this issue. Randomized response technique (RRT), originating in 1965 due to Warner, is a popular method in this area.

Block total response technique (BTRT), due to Raghavrao and Federer in 1979, is a method that incorporates experimental design features into RRT with the goal of increasing respondents' anonymity, in addition to producing unbiased estimators of parameters involving sensitive features. In this paper, we have developed an innovative estimator of the population mean of a sensitive feature using a permutation mechanism in the BTRT framework. This enables us also to compute an unbiased estimate of the variance of the proposed estimator.

Key words: Sensitive qualitative feature(s); Sensitive quantitative feature(s); Randomized response; Block total response technique; Random permutations.

1. Measuring Sensitive Characteristics Through Surveys: A Brief Review

Eliciting truthful responses on sensitive issues/characteristics from survey respondents has always been a challenge. During the latter half of the twentieth century, survey statisticians have proposed various methods of conducting surveys that provide anonymity to respondents and encourage them to answer truthfully to such sensitive issues. This enables gauging the level of the sensitive issue in the population, overcoming the biases that arise from false reporting. At the head of these survey methods is the popular randomized response technique (RRT) of interviewing, proposed by Warner in 1965. At the heart of this technique is a randomizing device, which is used to direct the respondent to provide answer to either a random (non-sensitive) question or to the sensitive question. It is of paramount importance that only the respondent knows how the randomization device directed him to respond and he provides only his answer to the interviewer without letting on how the device

Corresponding Author: Karabi Nandy
Email: Karabi.Nandy@UTSouthwestern.edu

directed him to respond. The underlying assumption is that since the respondent provides an answer to the interviewer without exposing his personal situation, any stigma associated with the sensitive question will be abated and the respondent will feel encouraged to respond truthfully. Warner showed that the responses obtained from this process will enable obtaining reliable estimates of the population parameter without direct knowledge of the responses obtained from individual respondents.

The RRT literature has focused on developing innovative randomization devices for both qualitative and quantitative characteristics. Examples of sensitive qualitative characteristics, particularly in the realm of public health, are use of contraceptive methods during sexual activity (yes/no), illicit substance use (yes/no), ever had an induced abortion (yes/no), ever had suicidal thoughts (yes/no) etc. Examples of sensitive quantitative features are the number of sexual partners one has, the number of times a person used illicit substance in the last month, the amount of time spent in a correctional facility etc. In addition to Warner's seminal work, we refer to a few book chapters and journal publications on RRT for details: Chaudhuri and Mukerjee (1987, 1988), Hedayat and Sinha (1991), Chaudhuri (2011) and Chaudhuri and Christofides (2013). Fifty years since it was first introduced, a celebratory Golden Jubilee Volume on RRT was compiled by Chaudhuri *et al.* in 2016 in a volume of the Handbook of Statistics.

In this paper, we are interested in quantitative sensitive characteristics; so we will focus our discussion henceforth for such sensitive characteristics only.

Greensberg *et al.* (1971) presented the first work involving RRT for continuous sensitive characteristics. Several others followed since, such as Eriksson (1973), Pollock and Bek (1976), Anderson (1977), *etc.* In recent years, Diana and Perri (2011) showcased the use of auxiliary information for estimating the mean of quantitative sensitive data and compared different models from both the perspectives of gaining efficient estimators as well as protecting respondents' anonymity. In 2015, Bose's work dealt with estimating the population mean of a sensitive feature wherein it is assumed that the true population values are captured by possibly a superset of M known quantities $[T_1, T_2, \ldots, T_M]$.

In the various approaches to extracting truthful responses on sensitive issues in surveys, an alternative method was proposed by Raghavarao and Federer (1979) where the idea was to incorporate basic experimental design elements in this framework. Block total response technique (BTRT) was suggested, based on the use of supplemented block designs / balanced incomplete block designs / spring balance weighing designs. In the context of a survey design, a "block" may be thought of as a questionnaire, containing a subset of the total number of questions, selected from a pool of questions which includes the sensitive question(s) as well. Of course, a given block may or may not contain the sensitive question(s).

Henceforth, we will closely follow the methodology suggested by Raghavarao and Federer (1979) and adopted in Nandy *et al.* (2016) and elsewhere. We shall use the BTRT framework to develop an estimator of the population mean for a quantitative sensitive characteristic. The entire exercise of sampling and estimation is geared towards unbiased estimation of the parameter under consideration. Layout of the rest of the paper is as follow. In section 2, we present some of the BTRT literature on quantitative sensitive characteristics.

In section 3, we introduce a BTRT version for estimating the mean for a sensitive quantitative item, similar to the methodology presented in Nandy *et al.* (2016) for qualitative items. In section 4, we provide an extension of our methodology that may potentially provide increased protection of respondents' privacy. Finally in section 5, we present some concluding remarks.

2. Methods to Obtain Truthful Responses for Sensitive Quantitative Items Through Survey

2.1. The development of BTRTS for sensitive quantitative items

As mentioned in the introduction, the BTRT method was suggested by Raghavarao and Federer (1979) as an alternative to the RRT. What stands out about this method is the increased protection of respondents' privacy when answering to sensitive questions. In fact, Coutts and Jann (2011) compared various RRT methods to BTR and showed that BTR outperformed the RRTs in terms of increased respondents' trust, better understanding of the interview instructions, lesser time to answer as well as lower non-response rates.

After its introduction, subsequent works in BTR focused on how to incorporate multiple sensitive questions into the design as well as development of various scoring mechanisms, i.e., how to score questions, other than in a binary fashion, so as to further increase respondents' privacy. The latter is rooted in the idea that a total score could be incriminating if that score could only be achieved by answering "yes" to at least one sensitive question. The works of Smith and Street (2003) and Smith (2005) are some examples of these. In 2016, we undertook various meaningful versions/generalizations of the BTRT and introduced empirical Bayes estimators.

All advancements in this area, however, have been for qualitative sensitive characteristics. In this work, we propose to present BTRT for quantitative sensitive characteristics. For this, our starting points are our own work in Nandy *et al.* (2016), Bose (2015) and Mukherjee *et al.* (2018). We present some details related to the latter two in the following subsection.

2.2. Study of a sensitive quantitative item under general sampling scheme

According to Bose (2015), we assume that truthful unknown responses $[Y_1, Y_2, \ldots, Y_N]$ are captured by possibly a superset of M known quantities $[T_1, T_2, \ldots, T_M]$. Therefore, quantitative nature of the sensitive feature is only to the extent of being discrete-valued. In effect, therefore, a finite population Y-distribution refers to a frequency distribution of the T's such as $[N_1, N_2, \ldots, N_M; \sum_i N_i = N]$ or, in other words, it refers to a probability distribution $[w_1, w_2, \ldots, w_M]$ where $w_i = N_i/N; i = 1, 2, \ldots, M$. An RRT is now geared towards unbiased estimation of the w's - using a suitable randomization device as described below.

Choose a fraction δ and a total of R chips such that $R_0 = \delta R$ chips [among the R chips] read as "Report T" in case the respondent happens to choose any of these R_0 chips. Further to this, a set of $R_i = (1-\delta)R/M$ chips reads as "Report T_i", if the respondent happens to choose any one of these R_i chips $(i = 1, 2, \ldots, M)$. It is now clear that each respondent is supposed to select at random one chip and act accordingly by responding truthfully - without divulging the type of chip selected. The chosen chip is returned back to the collection each time. That is how we generate the data under RRT.

The observed proportions of the Ts, say $p_1, p_2, \ldots, p_M$ are random and it turns out that

$$E(p_i) = \delta w_i + (1 - \delta)/M; i = 1, 2, \ldots, M.$$

This suggests that we can unbiasedly estimate w_i as $[p_i - (1 - \delta)/M]/\delta$ for $i = 1, 2, \ldots, M$. Therefore, the population mean $[=\sum_i w_i T_i]$ is unbiasedly estimated. It can be seen that SRSWR sampling of respondents entails us to regard the responses as being independently and identically distributed (iid). This simplifies the data analysis significantly.

Bose also gave an expression for variance of the estimate of the population mean. Variance estimation is not considered there. It follows that this method has an inherent limitation in that it does not address the estimation problem in case the sampling design is fixed size (N, n) sampling design such as SRSWOR (N, n) or any arbitrary sampling design. The iid nature of the responses is highly restrictive to do away with.

Next, Mukherjee *et al.*. (2018) undertook this study in its most general form. They provided formulae for mean estimation, expression for variance of the estimate and a method for variance estimation as well.

The idea is to provide an unbiased estimate of Y_i^* - the true Y-value on the sensitive feature, associated with the i-th respondent - for every selected respondent i in the sample. Once the respondent i has been selected and has been asked to provide RR $[R_i]$ by making a random choice of one chip and acting accordingly, it turns out that

$$E(R_i) = \delta Y_i^* + (1 - \delta) \sum_i T_i/M.$$

Therefore,

$$\hat{Y}_i^* = [R_i - (1 - \delta) \sum_i T_i/M]/\delta.$$

It is now simple to obtain the Horvitz Thomson Estimator (HTE) $\sum_i \hat{Y}_i^*/N\pi_i$ for the population mean for any arbitrary fixed size (N, n) sampling design - the choice of the sampling design being subject to providing positive first and second order inclusion probabilities *etc.* For the proof of unbiasedness, deduction of the expression for variance as well as estimated variance, we may refer to Mukherjee *et al.* (2018).

It must, however, be noted that the discrete nature of the quantitative feature is still maintained. That is a major limitation of the study so far described.

3. Introducing Block Total Response Technique for Quantitative Sensitive Feature

We now proceed to discuss BTR Technique for unbiased estimation of the population mean for a sensitive quantitative feature Q^* with true values $Y(Q^*)$, apriori known to the respondents. Also are apriori known values of each of a set of v other quantitative ordinary [non-sensitive] features $Q_1, Q_2, \ldots, Q_v$ to the respondents. We draw a random sample of n respondents, following SRSWOR(N, n) sampling. We no longer require that the Y-values be discrete. At best, it may be convenient to make a choice of the ordinary features Qs such that their ranges broadly cover the range of values of the sentitive feature Q^*.

We employ BTR technique in the following manner. We start with a Binary Proper Equireplicate Block Design [BPEBD] involving b blocks, each of size k, with equal replication number r of each of the v non-sensitive Qs. We then supplement each such block with one additional question, *viz.*, the sensitive question Q^*. Thus each of the b blocks of size k is 'extended' to one of size $(k+1)$. We also introduce an additional block B_0 of size v - incorporating all the v non-sensitive Qs.

The respondents in the sample are randomly split into $(b+1)$ sets of sizes $n^*, n^*, \ldots, n^*, n_0$. We assume that the sample size n has a convenient integer decomposition $n = (bn^* + n_0)$ for suitably chosen integers n^*, n_0. The b blocks each receive n^* respondents and each respondent provides only the sum total of responses to $(k+1)$ questions, the k non-sensitive questions included in the specific block in which the respondent belongs, along with the sensitive feature Q^*. The same is also true of the last block B_0 - although all the v features in this block are non-sensitive in nature. Our goal is to obtain an estimate of $\bar{Y}(Q^*)$.

In this context, we are tempted to quote Raghavarao and Federer (1979): "One early anonymous-direct-question method that was used successfully (e.g. by A. J. King and others at Iowa State University) was to have the respondent complete an unmarked questionnaire in secret and to deposit the questionnaire in a large locked box in which other questionnaires had been deposited; then, the respondent observed that the contents of the box were thoroughly mixed. We shall call this method the 'black box' (BB) method."

In our context, we may refer to $(b+1)$ such black boxes in a meaningful manner.

3.1. Estimation of $\bar{Y}(Q^*)$

At this stage, let us consider an illustrative example with $b = 5, v = 10, r = 2, k = 4, n = 350, n^* = 50, n_0 = 100$. Let further the blocks of the BPEBD be formed as seen in Table 1 below.

Table 1: Blocks in the BPEBD

Block	Non-sensitive Features	Sensitive Feature
B_1	Q_1, Q_2, Q_3, Q_4	Q^*
B_2	Q_5, Q_6, Q_7, Q_8	Q^*
B_3	Q_9, Q_{10}, Q_1, Q_2	Q^*
B_4	Q_3, Q_4, Q_5, Q_6	Q^*
B_5	Q_7, Q_8, Q_9, Q_{10}	Q^*

In the first block, let us now compute the average of Block Total Responses - averaged over all the n^* respondents' BTR scores. Let us denote it by $B\bar{T}R(B_1)$. It follows that its expectation is given by $E[B\bar{T}R(B_1)] = \bar{Y}(Q_1) + \bar{Y}(Q_2) + \bar{Y}(Q_3) + \bar{Y}(Q_4) + \bar{Y}(Q^*)$ where $\bar{Y}(Q)$ refers to the population average of true values for the feature identified through Q. Likewise, we carry out the same for all blocks B_1 to B_5. Additionally, we work it out for the last block B_0 as well.

Adding the results for the first 5 blocks, we obtain

$$E[\sum_i B\bar{T}R(B_i)] = 2[\bar{Y}(Q_1) + \bar{Y}(Q_2) + \ldots + \bar{Y}(Q_{10})] + 5\bar{Y}(Q^*)$$

while

$$E[B\bar{T}R(B_0)] = \bar{Y}(Q_1) + \bar{Y}(Q_2) + \ldots + \bar{Y}(Q_{10}).$$

From the above, we deduce

$$\hat{\bar{Y}}(Q^*) = \frac{[\sum_i B\bar{T}R(B_i)] - 2[B\bar{T}R(B_0)]}{5}.$$

3.2. Estimation of $V(\hat{\bar{Y}}))$

In order to work out variance estimate of this estimate of the population average of the sensitive feature Q^*, we propose to develop an important representation of the estimate derived above. For this, we assume that the respondents' responses are not associated with others' responses and that, to most extent, the respondents behave independently - so far as the responses are concerned.

We provide below an extensive use of permutation groups. Let P_1 denote a random permutation of size n^* of the integers $1, 2, \ldots, n^*$ associated with the labels of the respondents in Block 1. Likewise, we develop independently all other permutations P_2 to P_b and lastly, P_0 of size n_0 for the block B_0.

Now we group the responses across the $b+1$ blocks in sets of $b+1$ - taking one from each of the b blocks and 2 from the last block B_0. Once more we illustrate this feature by referring to the above example. We choose, for example, $n^* = 50$ and $n_0 = 100$ so that $n = 350$.

$P_1 =$[44 18 17 14 26 38 19 34 30 37 7 1 20 39 11 3 31 22 46 23 9 28 10 8 12
4 16 27 32 40 29 49 21 48 5 13 15 43 50 2 41 25 35 45 33 36 47 42 6 24];
$P_2 =$[4 30 16 14 38 46 21 39 32 13 49 19 20 2 48 47 17 31 9 50 27 44 35 6 40
3 10 12 37 11 8 29 1 22 26 24 33 7 34 18 45 23 42 36 43 5 28 15 25 41];
$P_3 =$[7 8 44 21 39 38 4 43 19 11 45 48 26 3 10 31 15 49 30 25 16 17 46 14 2
5 28 32 1 41 47 40 20 34 27 18 9 13 24 50 36 37 23 33 42 22 12 29 35 6];
$P_4 =$[13 29 41 11 36 40 46 31 3 48 50 30 7 14 23 21 25 8 9 32 2 37 28 1 42 33
20 45 49 19 12 16 44 43 38 15 39 24 26 4 22 10 17 27 34 6 18 35 5 47];
$P_5 =$[2 37 13 47 27 21 32 1 22 43 20 33 36 24 28 16 9 35 19 15 31 44 23 41
30 29 5 14 4 49 34 42 48 12 18 6 10 46 17 26 39 7 3 45 25 38 50 11 8 40];

P_0 =[9 10 95 26 18 61 21 60 57 8 67 70 73 2 46 54 100 80 17 40 5 4 77 19 87 52 76 25 81 35 55 14 50 37 29 69 38 89 98 90 59 12 68 7 53 16 75 39 94 48 42 32 56 36 41 96 82 65 78 62 74 93 86 3 97 13 47 49 63 88 85 43 51 30 91 15 58 22 64 71 33 1 27 45 28 20 84 11 23 44 99 34 24 6 31 66 79 92 72 83];

According to the above permutations applied to different blocks, Set 1 comprises of responses of 7 respondents labeled (1) 44 in B_1, (2) 4 in B_2, (3) 7 in B_3, (4) 13 in B_4, (5) 2 in B_5 and (6,7) $9, 10$ in B_0. An estimator for $\bar{Y}(Q^*)$ based on this data Set 1 is given by

$$\begin{aligned}1/5 \times [&\{Y(44;Q_1)+Y(44;Q_2)+Y(44;Q_3)+Y(44;Q_4)\}\\ &+\{Y(4;Q_5)+Y(4;Q_6)+Y(4;Q_7)+Y(4;Q_8)\}\\ &+\{Y(7;Q_9)+Y(7;Q_{10})+Y(7;Q_1)+Y(7;Q_2)\}\\ &+\{Y(13;Q_3)+Y(13;Q_4)+Y(13;Q_5)+Y(13;Q_6)\}\\ &+\{Y(2;Q_7)+Y(2;Q_8)+Y(2;Q_9)+Y(2;Q_{10})\}\\ &-\{Y(9;Q_1)+Y(9;Q_2)+\ldots+Y(9;Q10)\}\\ &-\{Y(10;Q_1)+Y(10;Q_2)+\ldots+Y(10;Q10)\}].\end{aligned}$$

We proceed in this manner and obtain 50 estimates of $\bar{Y}(Q^*)$ based on the 50 sets as defined above. Because of the underlying permutation principle, these estimates are also exchangeable in nature. This characterization of the individual estimates lends itself to easy computation of their average, which is the estimate of the population mean. Further, variance estimation becomes a trivial task: $\sum_i(e_i-\bar{e})^2/n(n-1)$ is an unbiased variance estimate of $\bar{e}$ based on iid estimates es.

Even though the respondents are selected according to $SRSWOR(N,n)$, use of permutations within blocks enables us to justify the assumption of iid nature of the estimates based on different sets of data. It is not however clear if the same holds true for any arbitrary fixed size (n) sampling design.

In the above, we assumed the condition: $n = bn^* + n_0$ for suitably chosen integers n^*, n_0. It is possible to relax this condition and instead work with another representation. We reconsider the above example to illustrate this point.

Once again, we start with $n = 350$ but assume the representation: $350 = 30+40+50+60+70+100$. Note that there is a common divisor of 10 among all the respondent group sizes. This time we can assemble the sets so that we have 10 iid estimates of the parameter of interest, *e.g.*, mean of the sensitive feature Q^*. Once these formations are done, the rest is routine in terms of computation of mean and variance of iid estimates.

We describe the essential step below with reference to the first of the 10 sets of estimates. The sizes of the blocks will be the highest common factor, which is 10 in this case. This suggests (i) deriving random permutations of the respondent labels within each block; (ii) forming 10 subsets of equal size within each block. Note that subset sizes will vary across the blocks; (iii) forming unbiased estimates for the mean of the sensitive feature [Q^*] from subsets collected serially across all the blocks; (iv) using iid sample estimates to arrive at the over-all average *etc*.

We carry out the exercise below. The subsets within each block, after random permutation, are shown within parenthesis. Also the block sizes are indicated in parenthesis.

$P_1(1-30) =$[(30 7 1); (20 11 3); (22 23 9); (28 10 8); (12 4 16);
(27 29 21); (5 13 15); (2 25 6); (24 18 17); (14 26 19)];
$P_2(1-40) =$[(13 19 20 2); (17 31 9 27); (35 6 40 3); (10 12 37 11); (8 29 1 22);
(26 24 33 7); (34 18 23 36); (5 28 15 25); (4 30 16 14); (38 21 39 32)];
$P_3(1-50) =$[(7 8 44 21 39); (38 4 43 19 11); (45 48 26 3 10); (31 15 49 30 25);
(16 17 46 14 2); (5 28 32 1 41); (47 40 20 34 27); (18 9 13 24 50);
(36 37 23 33 42); (22 12 29 35 6)];
$P_4(1-60) =$[(16 39 48 42 32 56); (36 41 3 13 47 49); (43 51 30 15 58 22);
(33 1 27 45 28 20); (11 23 44 34 24 6 31); (35 55 14 50 37 29);
(38 59 12 7 53 9); (10 26 18 21 60 57); (8 2 46 54 17 40); (5 4 77 19 52 25)]
$P_5(1-70) =$[(3 13 47 49 63 43 51); (30 15 58 22 64 35 55); (14 50 37 29 69 38 59);
(12 68 7 53 33 1 27); (45 28 20 11 23 44 34); (24 6 31 66 9 10 26);
(18 61 21 60 57 8 67); (70 2 46 54 17 40 5); (4 62 19 52 25 16 39);
(48 42 32 56 36 41 65)].

$P_0(1-100)$ is the 10 subsets formed taking 10 permutations at a time and serially - starting from the left corner.

We display the result based on data analysis for *Set* 1 across all the 6 blocks. For B_1, we consider the first set of 3 respondents labeled $(1, 7, 30)$ and average out the BTRs collected from them. So,

$$E(Set1) = \bar{Y}(Q_1) + \bar{Y}(Q_2) + \bar{Y}(Q_3) + \bar{Y}(Q_4) + \bar{Y}(Q^*).$$

Likewise, we have similar results from the first set of all other blocks. We denote these averaged responses by $R(B_1; S_1), R(B_2 : S_1), \ldots, R(B_6; S_1)$. It now follows that

$$\hat{\hat{Y}}(Q^*; S_1) = [R(B_1; S_1) + R(B_2; S_1) + \ldots + R(B_5; S_1) - 2R(B_6; S_1)]/5.$$

We will be referring to 10 such sample estimates and proceed to compute the combined estimate of the population mean of Q^* and its estimated standard error.

So,

$$\hat{\hat{Y}}(Q^*; S_2) = [R(B_1; S_2) + R(B_2; S_2) + \ldots + R(B_5; S_2) - 2R(B_6; S_2)]/5,$$
$$\hat{\hat{Y}}(Q^*; S_i) = \ldots.$$
$$\hat{\hat{Y}}(Q^*; S_{10}) = [R(B_1; S_{10}) + R(B_2; S_{10}) + \ldots + R(B_5; S_{10}) - 2R(B_6; S_{10})]/5.$$

Therefore,

$$\hat{\hat{Y}}(Q^*) = 1/10[\hat{\hat{Y}}(Q^*; S_1) + \hat{\hat{Y}}(Q^*; S_2) + \ldots + \hat{\hat{Y}}(Q^*; S_{10})].$$

Further, estimated standard error is computed as usual by taking square root of

$$\sum(\hat{\hat{Y}}(Q^*; S_i) - \hat{\hat{Y}}(Q^*))^2/10 \times 9.$$

4. An Extention of the BTRT Method

In the technique presented above, in every block $[B_1$ to $B_b]$, we are utilizing some k of the v Qs - leaving the rest unutilized. When k is small, respondents may not feel comfortable responding truthfully since responding to Q^* is mandatory. In this section, we provide an extension of the above technique, as follows.

For every block out of B_1 to B_b, we bring a variation in the block composition as: 1. List of k "must respond" Q's - these are kept in Part A. This is the same as before. 2. Remaining $(v-k)$ Q's and Q^* are kept in Part B. From Part B, a respondent is to make a random choice of exactly one question from the total $(v-k+1)$ questions; next the respondent will blend the selected question with those in Part A and provide BTR - without divulging the nature of the question selected from Part B. To simplify the data analysis, it may be assumed that selection from Part B is governed by the rule: Select Q^* with probability δ and any one of the remaining Q's with probability $(1-\delta)/(v-k)$. Further, we use the same selection mechanism in each such block.

Once more, we can study the formation of estimates based on the sets separately and then combine them. We display the result for Set 1 below for the example considered above and with the choice $\delta = 0.4$. Accrued Block Totals provide for the first 5 blocks, the following expressions for their expectations under random choice of the question selected from Part B:

$$[\{Y(44; Q_1)+Y(44; Q_2)+Y(44; Q_3)+Y(44; Q_4)\}+0.1\{Y(44; Q_5)+\ldots+Y(44; Q_{10})\}+0.4Y(44; Q^*);$$

$$\{Y(4; Q_5) + Y(4; Q_6) + Y(4; Q_7) + Y(4; Q_8)\} + 0.1\{Y(4; Q_9) + \ldots + Y(4; Q_4)\} + 0.4Y(4; Q^*);$$

$$\{Y(7; Q_9) + Y(7; Q_{10}) + Y(7; Q_1) + Y(7; Q_2)\} + 0.1\{Y(7; Q_3) + \ldots + Y(7, Q_8)\} + 0.4Y(7; Q^*);$$

$$\{Y(13; Q_3)+Y(13; Q_4)+Y(13; Q_5)+Y(13; Q_6)\}+0.1\{Y(13; Q_7)+\ldots+Y(13; Q_2)\}+0.4Y(13; Q^*);$$

$$\{Y(2; Q_7) + Y(2; Q_8) + Y(2; Q_9) + Y(2; Q_{10})\}] + 0.1\{Y(2; Q_1) + \ldots + Y(2; Q_6)\} + 0.4Y(2; Q^*)$$

As for block B_0, we obtain $[Y(9; Q_1) + Y(9; Q_2) + \ldots + Y(9; Q_{10}) + Y(10; Q_1) + Y(10; Q_2) + \ldots + Y(10; Q_{10})]$.

From the above, it is routine to obtain an estimate for the average $Y(Q^*)$. Once such estimates are computed from each set, they may be treated as iid sample estimates and hence mean estimation and variance estimation are immediate.

5. Concluding Remarks

In the context of sensitive quantitative features, we have proposed a version of block total response technique which has flexibility in terms of implementation. We expect that the scheme in section 4 will likely provide increased privacy protection to respondents, compared to the BPEBD scheme in Section 3 which involves only the sensitive item Q^* in the supplementary part. We plan to quantify this increase in future work. It may be noted that in most practical surveys, collection of data is on several variables, which are then used to estimate not just marginal distributions but also joint distributions, correlations, regressions *etc.* Deriving joint inferences from data gathered using BTRT would be interesting. We plan to explore this in future study. The innovativeness of the method studied here lies in procuring an estimate as well as estimated standard error by exploiting a permutation method to generate exchangable observations. The results have been deduced under SRSWOR sampling. For a general fixed size sampling design, we have yet to develop a version of BTRT. This is true for both qualitative and quantitative sensitive features.

References

Anderson, H. (1977). Efficiency versus protection in a general randomized response model. *Scandinavian Journal of Statistics*, **4**, 11–19.

Bose, M. (2015). Respondent privacy and estimation efficiency in randomized response surveys for discrete-valued sensitive variables. *Statistical Papers*, **56**, 1055-1069.

Chaudhuri, A. (2011). *Randomized Response and Indirect Questioning Techniques in Surveys.* CRC Press, Boca Raton, FL.

Chaudhuri, A. and Mukerjee, R. (1987). Randomized response techniques: a review. *Statistica Neerlandica*, **41(1)**, 27–44.

Chaudhuri, A. and Mukerjee, R. (1988). *Randomized Responses: Theory and Techniques.* Marcel Dekker, New York, NY.

Chaudhuri, A. and Christofides, T. C. (2013). *Indirect Questioning in Sample Surveys.* Springer, Germany.

Series Volume Editors: Chaudhuri, A., Christofides T. C. and Rao C. R. (2015). *Data Gathering, Analysis and Protection of Privacy Through Randomized Response Techniques: Qualitative and Quantitative Human Traits.* Handbook of Statistics; volume **4**.

Chaudhuri A. and Roy D. (1997). Model assisted survey sampling strategies with randomized response. *Journal of Statistical Planning and Inference*, **60**, 61–68.

Coutts, E. and Jann, B. (2011). Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT). *Sociological Methods and Research*, **40(1)**, 169–193.

Diana G. and Perri P. F. (2011). A class of estimators for quantitative sensitive data. Statistical Papers, **52**, 633–650.

Eriksson S. A. (1973). A new model for randomized response. *International Statistical Review*, **41**, 101–113.

Guerriero M. and Sandri M. F. (2007). A note on the comparison of some randomized response procedures. *Journal of Statistical Planning and Inference*, **137**, 2184–2190.

Pollock K. H. and Bek Y. (1976). A comparison of three randomized response models for quantitative data. *Journal of the American Statistical Association*, **71**, 884–886.

Hedayat, A. S. and Sinha, B. K. (1991). *Design and Inference in Finite Population Sampling.* Wiley, New York, NY.

Mahajan P. K., Gupta J. P. and Singh R. (1994). Determination of optimum strata boundaries for scrambled response. *Statistica*, **54**, 375–381.

Mukherjee, S., Sinha, B. K. and Chatterjee, A. (2018). *Statistical Methods in Social Science Research.* Springer.

Nandy, K., Marcovitz, M. and Sinha, B. K. (2016). Eliciting Information on Sensitive Features: Block Total Response Technique and Related Inference. In 'Handbook of Statistics, Special Volume on Golden Jubilee Celebration on Randomized Response Technique'. Chapter 19, pp 317-330.

Nayak, T. K. and Adeshiyan, S. A. (2009). A unified framework for analysis and comparison of randomized response surveys of binary characteristics. *Journal of Statistical Planning and Inference*, **139**, 2757–2766.

Raghavarao, D. and Federer, W. T. (1979). Block total response as an alternative to the randomized response method in surveys. *Journal of Royal Statistical Society B Methodology*, **41**, 40–45.

Singh S., Joarder A. H. and Kinh M. L. (1996). Regression analysis using scrambled responses. *Australian Journal of Statistics*, **38**, 201–211.

Sinha B. K. (2017). Some refinements of block total response technique in the context of RRT methodology. *Statistics and Applications*, **15(1)**, pp 167–171.

Smith N. F. (2005). The design of scoring schemes for surveys using the block total response method. *Communications in Statistics - Theory and Methods*, **34(11)**, 2157–2168.

Smith, N. F. and Street, D. J. (2003). The use of balanced incomplete block designs in designing randomized response surveys. *Australian and New Zealand Journal of Statistics*, **45(2)**, 181–194.

Warner, S. L. (1965). Randomized response: a survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, **60**, 63–69.