Thailand Statistician
July 2023; 21(3): 466-483
http://statassoc.or.th
Contributed paper

Further Thoughts on Applications of Block Total Response Techniques in Case of One or Two Sensitive Binary Features

Opendra Salam*[a] and Bikas K. Sinha [b]
[a] Manipur University, Imphal, Manipur, India
[b] Retired Professor, Indian Statistical Institute, Kolkata, India
* Corresponding author; Email: sopendra@manipuruniv.ac.in

Received: 16 March 2020
Revised: 20 July 2020
Accepted: 8 September 2020

Abstract

In this paper, we propose to study block total response techniques in case of one or two sensitive binary features to estimate the population proportion(s) of incidence of the sensitive binary features by giving more flexibility to the sampled respondents to truthfully respond to the sensitive feature(s) with increased sense/perception of anonymity.

Keywords: Randomized response technique, population proportion, sensitive features, binary version, designing the survey.

1. Introduction

In sample survey, the study variable may often be sensitive in nature, as for example it may be related to 'backdoor' income, addiction of a drug, being a habitual gambler, having a history of abortion, extramarital affairs and the like. The respondent may refuse to give a truthful answer by protecting his/her privacy or even furnish an evasive answer. Data thus obtained by direct questionnaire procedure is difficult to use for making a valid inference about the population parameter. To elicit truthful response from the respondents by providing a satisfactory degree of privacy and thereby make a valid inference about the population parameter by reducing or eliminating the non-sampling bias, in 1965, Warner introduced a randomized response technique/methodology (RRT/RRM). We refer to the books / book chapters on RRT by Fox and Tracy (1986), Chaudhuri and Mukerjee (1988), Hedayat and Sinha (1991), Chaudhuri (2011), Chaudhuri and Christofides (2013), and Mukherjee et al. (2018).

After the introduction of RRT, survey sampling practitioners/theoreticians have given more attention to this area of survey methodological research. The basic concept of RRT is to elicit the true response on the sensitive feature(s) from the sampled respondents by protecting their privacy so that the population proportion of incidence of the sensitive feature can be unbiasedly estimated. For increasing the degree of protection of confidentiality of the respondent's response, Raghavarao and Federer (1979) introduced a novel technique/methodology termed block total response technique/methodology (BTRT/BTRM). In this approach, the authors introduced and incorporated basic experimental design elements in the preparation of the questionnaire framework. The technique is elaborated below.

Consider a collection of v, otherwise non-sensitive, binary features (NSBFs) and only one extra sensitive binary feature (SBF) Q^*. So, the total number of binary features (BFs) is $(v + 1)$. A block of questions (which contains some NSBFs and the single SBFQ^*) is presented to the respondent. The respondent is to provide only the block total response (BTR) in terms of the overall score (i.e., only the total number of yes answers) without divulging any response to any specific Qs – be it NSBF or SBF. It is believed that this BTR technique (BTRT) will adequately protect the privacy of the respondent, and hence, the correct response to the SBF will emerge. BTRT is an alternative method of RRT to increase respondent's anonymity and enable estimation of the parameter (i.e., proportion of yes respondents in the population) involving sensitive binary feature. In this approach, there is nothing special about the choice of the NSBFs. These may be associated with socio-economic survey features of the population under consideration. In the context of a survey design, a 'block' of questions 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).

Since the introduction of BTRT/BTRM almost 40 years back, studies have emerged to examine the benefits of this 'new' method in contrast to the usual RRT. 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. These favorable merits encourage the survey sampling experts to deal with BTRT/BRTM. Recently, Nandy et al. (2015) undertook various meaningful versions and generalizations of the block total response technique with the introduction of empirical Bayes estimators.

Nandy and Sinha (2020) provided an innovative interpretation of the estimator of the population mean of a sensitive feature using a permutation mechanism in the BTRT. However, when the block size (k) is small, respondents may not feel comfortable responding truthfully since responding to Q^* is compulsory. To overcome this problem, they have extended the BTRT, which provides more flexibility in increasing the privacy of the respondents. Variations in the block compositions have been suggested. Further to this, they also introduced BTRT for quantitative features and extended the BTRT when k is small to protect the privacy of the respondents.

In this paper, by changing the sampling frame, we study the BTRT with one and two qualitative sensitive features to estimate the population proportion of incidence of these features by giving more flexibility to the sampled respondents to provide truthful answer(s) for the sensitive feature(s). It is perceived notion of preservation of anonymity that works well for the respondents. Moreover, this flexibility [introduced via BTRT mechanism] supports the notions of increased respondents' trust, better understanding of the interview instructions, lesser time to answer as well as lower non-response rates" (Coutts and Jann, 2011). In fine, our paper provides extensions to Nandy and Sinha (2020) work in two directions : (i) by providing more perceived sense of anonymity and (ii) by simultaneously dealing with two sensitive binary features.

2. Block Total Response Technique For One Binary Sensitive Feature

In this section, we narrate the basic feature of block total response technique (BTRT) and its extension, as formulated by Nandy and Sinha (2020). Consider a collection of v non-sensitive binary features (NSBFs) $[Q_1, Q_2, \ldots, Q_v]$ and one sensitive binary feature (SBF) Q^*. Denote by 'b' the number of blocks (i.e., sets of questions), each containing some 'k' distinct NSBFs and the SBF. In this collection of b blocks, each NSBF is replicated 'r' times. Further to this, there is one block 'B_0' containing all NSBFs. Thus, on the whole, there are $k+1$ BFs in each of the b blocks and all the v NSBFs appear in the additional block 'B_0'. The respondents in the sample are split into $(b+1)$ sets of sizes $n^*, n^*, \ldots, n^*, n_0$ such that $n = (bn^* + n_0)$. Each of the b blocks receives n^* respondents and the block 'B_0' receives n_0 respondents. Taking, for example, $v = 10, b = 5, k = 4, r = 2$, let the blocks so formed be taken as

$B_1 : [Q_1, Q_2, Q_3, Q_4; Q^*]$
$B_2 : [Q_5, Q_6, Q_7, Q_8; Q^*]$
$B_3 : [Q_9, Q_{10}, Q_1, Q_2; Q^*]$

$B_4 : [Q_3, Q_4, Q_5, Q_6; Q^*]$
$B_5 : [Q_7, Q_8, Q_9, Q_{10}; Q^*]$
$B_0 : [Q_1, Q_2, Q_3, Q_4, Q_5, \ldots, Q_{10}]$.

In this technique, in every block $[B_1$ to $B_b]$, we utilize only k of the v NSBFs, while the rest $(v - k)$ NSBFs are left unutilized. When k is small, respondents may feel uncomfortable responding truthfully since responding to Q^* is compulsory in each of the b blocks.

Nandy and Sinha (2020) extended the above technique by bringing variations in the block compositions as:

1. List of k 'must respond' NSBFs are kept in Part A.
2. Remaining $(v - k)$ NSBFs and Q^* (SBF) are all kept in Part B.

A respondent is to choose one question from $(v - k + 1)$ Qs in Part B and mix with the questions in Part A and supply BTR without divulging the identity of the question selected from Part B.

To select exactly one question from Part B, the rule suggested is: select Q^* with probability δ and any one of the remaining Q's with probability $(1 - \delta))/(v - k)$. This shows that there is constant probability for selection of Q^* and constant probability for selection of any one of the NSBFs in Part B.

Before proceeding further, we undertake an illustrative example to explain the application of Nandy-Sinha (2020) scheme.

Example 1 We start with the same design parameters i.e., $b = 5, v = 10, k = 4, r = 2$. We also choose $\delta = 0.10$. Suppose a random sample of $n(= 280)$ individuals have been selected. We randomly split them into 6 sets, taking $n^* = 50$ and $n_0 = 30$. We adopt the same block compositions as are displayed above. In order to implement the above scheme for Block 1, for example, we prepare a set of 20 identical cards of the same shape [square, say] and size. At the back of the cards, we write the Questions by adopting the following rule: Q_5 (3 cards), Q_6 (3 cards), ..., Q_{10} (3 cards), Q^* (2 cards). These refer to the Qs in Part B.

The procedure is : A respondent belonging to Block 1 is to draw one card at random from the collection of 20 cards and look at the back and identify the specific question (in Part B) to be answered truthfully (binary : $1/0$) and add this to the responses of the 4 questions in Part A of Block 1. The respondent gives a total, say 3, of the scores accrued from a total of 5 questions – 4 compulsory in Part A and one randomly selected from Part B. The respondent must not divulge any details –only provide the total score (block total response). This procedure continues for all 50 respondents in Block 1 with the same set of cards. Likewise, we prepare 20 cards for use of the respondents in Block 2 and so on. Of course, each time we study the block composition before labeling the cards. For the last block B_0, we do not need any cards. All NSBFs are compulsory.

Having implemented the data-gathering tools, we end up with raw scores of each of the 280 respondents – classified into 6 distinct groups. In each group, we calculate the group average of the scores and these are called summary statistics. Assume that at the end we end up with the following results:

Table 1 Summary statistics of example 1

Block	Total Score	No. of Respondents	Average Score
1	70	50	1.4
2	120	50	2.4
3	70	50	1.4
4	90	50	1.8
5	100	50	2.0
6	108	30	3.6

Following Nandy and Sinha (2020), we now prepare the following table.

Table 2 Data analysis: theory

Block	Sample Size	Expected Block Average	Block Average Score
1 (B_1)	n^*	EBA(1)	$\bar{x}_1$
2 (B_2)	n^*	EBA(2)	$\bar{x}_2$
3 (B_3)	n^*	EBA(3)	$\bar{x}_3$
4 (B_4)	n^*	EBA(4)	$\bar{x}_4$
5 (B_5)	n^*	EBA(5)	$\bar{x}_5$
6 (B_0)	n_0	$\mathbf{\Delta}$	$\bar{x}_0$

In the above,

$$EBA(1) = p_1 + p_2 + p_3 + p_4 + \delta P^* + (1-\delta)/(v-k)(\mathbf{\Delta} - p_1 - p_2 - p_3 - p_4);$$
$$EBA(2) = p_5 + p_6 + p_7 + p_8 + \delta P^* + (1-\delta)/(v-k)(\mathbf{\Delta} - p_5 - p_6 - p_7 - p_8);$$
$$EBA(3) = p_9 + p_{10} + p_1 + p_2 + \delta P^* + (1-\delta)/(v-k)(\mathbf{\Delta} - p_9 - p_{10} - p_1 - p_2);$$
$$EBA(4) = p_3 + p_4 + p_5 + p_6 + \delta P^* + (1-\delta)/(v-k)(\mathbf{\Delta} - p_3 - p_4 - p_5 - p_6);$$
$$EBA(5) = p_7 + p_8 + p_9 + p_{10} + \delta P^* + (1-\delta)/(v-k)(\mathbf{\Delta} - p_7 - p_8 - p_9 - p_{10});$$
$$EBA(6) = p_1 + p_2 + p_3 + p_4 + p_5 + p_6 + p_7 + p_8 + p_9 + p_{10} = \mathbf{\Delta}, say.$$

Now, summing over all the block average scores and using the 'method of moments', we set

$$\sum_{i=1}^{5} \bar{x}_i = 2\mathbf{\Delta} + 5\delta P^* + \frac{1-\delta}{(v-k)} 3\mathbf{\Delta}.$$

Now we substitute $\mathbf{\Delta}$ by $\hat{\mathbf{\Delta}} = \bar{x}_0$ and solve for P^* as:

$$\hat{P}^* = (\sum_{i=1}^{5} \bar{x}_i - 2.45\hat{\mathbf{\Delta}})/5\delta.$$

Computations yield: $\hat{\mathbf{\Delta}} = 3.60, \delta = 0.10, \sum_i = 1^5 \bar{x}_i = 9$ and, finally, from the above equation: we derive $\hat{P}^* = 0.36$.

Remark 1 For variance estimation, we may rework on this exercise by using 'half-sample' technique. We divide the total number of respondents into two broad groups: Group I and Group II. The entire exercise is carried out independently in the two groups. There we have the liberty of using different 'design parameters', sample sizes, δ-parameters etc – keeping the same SBF Q^* in mind. Therefore, we arrive at two independent estimates of P^*, say $\hat{P}^{*1}$ and $\hat{P}^{*2}$. Then we propose $\hat{P}^{**} = (\hat{P}^{*1} + \hat{P}^{*2})/2$ as the 'pooled' estimate of P^* and its variance estimate is given by $(\hat{P}^{*1} - \hat{P}^{*2})^2/4$.

2.1. An extension of the above technique to increase the privacy of the respondents

Let us change the selection rule in Part B as

1. Choose randomly two questions from $(v-k+1)$ BFs in Part B, following a selection rule.
2. The selected pair may or may not include Q^*.

We allot a chance of α to each of the pairs $[(Q^*, Q_i), Q_i$ is the NSBF question in Part $B]$ to be included in the selected sample pair; each of other pairs $[(Q_i, Q_j)]$ etc. get a chance of β such that $(v-k)\alpha + [(v-k)(v-k-1)/2]\beta = 1$. After the selection of two questions from Part B, we blend them with k questions [which are mandatory in Part A] and provide the BTR without divulging any information about the 2 questions selected in Part B. This is to be carried out by each of the n^* respondents of Block 1 and by the respondents in all the subsequent blocks labeled 2 to b. For the last block, labeled B_0, all the n_0 respondents are to provide total response to all the NSBFs. That part is easy since the SBF is not included in this special block. Now, we denote by $\bar{x}_1$ the first block total response, averaged over all the n^* respondents. Then its model expectation provides the 'estimating equation' stated below. It is assumed that the block sizes in Part A are each 'k'. Moreover, it is tacitly assumed that Part A in the first block contains the NSBFs labeled $(1, 2, \ldots, k)$. Then we may write the estimating equation arising out of this block as

$$\bar{x}_1 = [p_1 + p_2 + \ldots + p_k] + [(v-k)\alpha P^*] + (\mathbf{\Delta} - \sum_{i=1}^{i=k} p_i)[\alpha + (v-k-1)\beta].$$

As before, $[p_1, p_2, \ldots, p_v]$ are the proportions of NSBFs in the population and P^* is the proportion for the SBF Q^* in the population. Here for all the features, 'proportion' refers to the proportion of 'yes' responses. Note, further, that in the above, $\mathbf{\Delta} = \sum_{i=1}^{i=v} p_i$.

Summing over all the b blocks and noting that each NSBF has a replication of 'r'' in the collection, we obtain, for the over-all estimating equation,

$$\sum_{i=1}^{i=b} \bar{x}_i = r\mathbf{\Delta} + b(v-k)\alpha P^* + [\alpha + (v-k-1)\beta](b-r)\mathbf{\Delta}.$$

From this we can estimate P^* noting that $\hat{\Delta} = \bar{x}_0$ which stands for the sample mean of BTRs of the n_0 respondents in the block B_0. Routine simplification yields:

$$\hat{P}^* = \frac{\sum_{i=1}^{i=b} \bar{x}_i - r\hat{\Delta} - [\alpha + (v-k-1)\beta](b-r)\hat{\Delta}}{b(v-k)\alpha}.$$

Remark 2 If we use $\alpha = \beta = 2/(v-k+1)(v-k)$, the result coincides with that based on $SRSWOR(v-k+1, 2)$. This scheme suggests drawing a random sample of 2 questions from the pool in Part B. Of course, subsequently, these are combined with all the 'k'' NSBFs of Part A.

We now undertake an illustrative example.

Example 2 Consider the same design parameters i.e., $b = 5, v = 10, k = 4, r = 2$ and the same experimental set-up with $n = 280$ respondents. We randomly split 280 respondents into 6 sets, taking as before $n^* = 50$ and $n_0 = 30$. We adopt the same block compositions as are displayed above. We have $k = 4$ and suppose, we choose, $\alpha = 0.1$. Then we have $\beta = 4/15 = 0.02666$. In order to implement the above scheme for Block 1, for example, we prepare a set of 30 identical cards of the same shape (square, say) and size. At the back of the cards, we write the following:
Category (i) : Three cards for $[5, Q^*]$, 3 for $[6, Q^*]$, ..., 3 for $[10, Q^*]$, thus taking care of 18 cards.
Category (ii) : All the remaining 12 cards read : [any two out of 5 to 10].

We also keep a set of10 Bridge Cards : Ace $[1]$ to 10 ready for subsequent use, if needed. A respondent from block 1 picks up a card at random out of 30 cards. If it belongs to the Category

(i) above, then he/she already has the pair of questions selected from Part B. Otherwise, the card is from Category (ii) and in that case he/she turns to the set of 10 Bridge Cards and picks up two cards at random – noting that only the Bridge Cards with numbers 5 to 10 are valid. Once 2 Qs are thus chosen from Part B, these are to be combined with the 4 NSBFs of Part A and BTR is to be provided for all the 6 questions together. All the respondents from Block 1 have to follow this rule and provide BTRs. For Block 2, Part A contains the NSBFs $5-6-7-8$ and the rest are placed in Part B. Similar data-gathering exercise is carried out in each of the blocks 2 to 5. For block B_0, we proceed as before – taking all the 10 NSBFs together.

We must remember that only the BTR scores are supposed to be reported – without divulging any details. Note that any specific respondent may / may not have chosen the SBF (Q^*). Of course, he/she must respond truthfully and provide the BTR score – even if this has been selected. Having implemented the data-gathering tools, we end up with 'raw' scores of each of the 280 respondents – classified into 6 distinct groups. In each group, we calculate the group average of the scores and these are called 'summary statistics'. Assume that at the end we end up with the following results:

Table 3 Summary statistics of example 2

Block	Total Score	No. of Respondents	Average Score
1	120	50	2.4
2	125	50	2.5
3	114	50	2.28
4	121	50	2.42
5	118	50	2.36
6	114	30	3.8

As in the above, following Nandy and Sinha (2020), we now prepare the following table.

Table 4 Data analysis : theory

Block	No. of Respondents	Expected Block Average	Block Average Score
$1(B_1)$	n^*	EBA(1)	$\bar{x}_1$
2 (B_2)	n^*	EBA(2)	$\bar{x}_2$
3 (B_3)	n^*	EBA(3)	$\bar{x}_3$
4 (B_4)	n^*	EBA(4)	$\bar{x}_4$
5 (B_5)	n^*	EBA(5)	$\bar{x}_5$
6 (B_0)	n_0	$\boldsymbol{\Delta}$	$\bar{x}_0$

In the above,

$$EBA(1) = p_1 + p_2 + p_3 + p_4 + 0.6P^* + 0.2333(\boldsymbol{\Delta} - p_1 - p_2 - p_3 - p_4);$$
$$EBA(2) = p_5 + p_6 + p_7 + p_8 + 0.6P^* + 0.2333(\boldsymbol{\Delta} - p_5 - p_6 - p_7 - p_8);$$
$$EBA(3) = p_9 + p_{10} + p_1 + p_2 + 0.6P^* + 0.2333(\boldsymbol{\Delta} - p_9 - p_{10} - p_1 - p_2);$$
$$EBA(4) = p_3 + p_4 + p_5 + p_6 + 0.6P^* + 0.2333(\boldsymbol{\Delta} - p_3 - p_4 - p_5 - p_6);$$
$$EBA(5) = p_7 + p_8 + p_9 + p_{10} + 0.6P^* + 0.2333(\boldsymbol{\Delta} - p_7 - p_8 - p_9 - p_{10});$$
$$EBA(6) = p_1 + p_2 + p_3 + p_4 + p_5 + p_6 + p_7 + p_8 + p_9 + p_{10} = \boldsymbol{\Delta}, say.$$

Now, summing over all the block average scores and using the 'method of moments', we set

$$\sum_{i=1}^{5} \bar{x}_i = 2\boldsymbol{\Delta} + 3P^* + 0.6999\boldsymbol{\Delta}.$$

Now we substitute $\mathbf{\Delta}$ by $\hat{\mathbf{\Delta}} = \bar{x}_0$ and solve for P^* as:

$$\hat{P}^* = (\sum_{i=1}^{5} \bar{x}_i - 2.6999 \times 3.8)/3 = 0.566.$$

For variance estimation, we may take recourse to 'split-half' method, as explained above.

2.2. Case of purely random choice from both parts

Suppose there are k_1 NSBFs in Part A and $k_2 = v - k_1 + 1$ BFs, including the sensitive question Q^* in Part B. Arrangement of k_1 NSBFs in Part A is the same as above. Respondent is to blend randomly selected s_1 NSBFs from k_1 NSBFs and s_2 from $k_2 = v - k_1 + 1$ BFs [which includes the sensitive question Q^*) and supply the BTR of $(s_1 + s_2)$ BFs (which may or may not include the sensitive question in the chosen sample], without divulging any information about the selected questions. Let π_1 and π_2 denote respectively, the inclusion probabilities of i^{th} unit $(i = 1, 2, \ldots, k_1)$ NSBF from Part A out of s_1 and of i^{th} BF $(i = 1, 2, \ldots, (v - k_1 + 1))$ from Part B respectively. Therefore, every question within the Parts A and B have an equal chance of inclusion viz., $\pi_1 = s_1/k_1, \pi_2 = s_2/k_2$, these being explicit expressions for the questions in the two parts in s_1 and s_2 respectively. Now, as before, we work out the sample mean of all the n^* BTRs arising out of the n^* respondents of block 1 and denote it by $\bar{x}_1$. Then the estimating equation will be

$$\bar{x}_1 = \sum_{i=1}^{i=k_1} p_i(s_1/k_1) + P^*(s_2/k_2) + (\mathbf{\Delta} - \sum_{i=1}^{i=k_1} p_i)(s_2/k_2).$$

Note that the notations have already been explained. And, summing over all the b blocks, we derive the single estimating equation for P^* as

$$\sum_{i=1}^{i=b} \bar{x}_i = r\mathbf{\Delta}(s_1/k_1) + bP^*(s_2/k_2) + (b - r)\mathbf{\Delta}(s_2/k_2).$$

Therefore, the population proportion P^* of the sensitive qualitative question can be estimated by using the formula

$$P^* = \frac{\sum_{i=1}^{i=b} \bar{x}_i - (s_1/k_1)2\mathbf{\Delta} - (b - r)\mathbf{\Delta} \times (s_2/k_2)}{b(s_2/k_2)}.$$

Note that, as before, $\mathbf{\Delta}$ is estimated from the data in the last block B_0.

Remark 3 If n_0 is large, B_0 can be split into two sets : B_{01} and B_{02} with the purpose of splitting v NSBFs into two mutually exclusive and exhaustive subsets of v_1 and v_2 such that $\mathbf{\Delta} = \mathbf{\Delta}_1 + \mathbf{\Delta}_2$. The two can now be comfortably combined to derive an estimate of $\mathbf{\Delta}$.

We now take up an illustrative example.

Example 3 Let us start with the same design parameters i.e., $b = 5, v = 10, k = 4, r = 2$ and the same experimental set-up with $n = 280$ respondents. As before, we randomly split them into 6 sets, taking $n^* = 50$ and $n_0 = 30$. We adopt the same block compositions as are displayed above. We have $k_1 = k = 4$ and we select $s_1 = 3$. We also select $s_2 = 3$. In order to implement the above scheme for Block 1, for example, we prepare a set of 11 identical cards of the same shape [square, say] and size. At the back of the cards, we write the numbers $1, 2, \ldots, 10$ and the symbol $(^*)$ – one card for each.

The procedure is : A respondent belonging to Block 1 is to draw three cards at random from the collection of first 4 cards $(1 to 4)$. Note that this is as good as selecting one card at random and discarding the same, and thereby, taking the rest at hand! Out of the remaining 7 cards, the respondent has to select any 3. Thus he/she will have a collection of 6 cards altogether from the two sets. Next, he/she will respond truthfully to all the 6 binary $(1-0)$ features selected and arrive at the Block Total Response and only the BTR score is supposed to be reported – without divulging any details. Note that the respondent may / may not have chosen the SBF (Q^*). Of course, he/she must respond truthfully and provide the BTR score – even if this has been selected. Likewise, we prepare 11 cards for use of the respondents in block 2 and so on. Of course, each time we study the block composition before using the cards to form two designated sets. For the last block B_0, we do not need any cards. All NSBFs are compulsory.

Having implemented the data-gathering tools, we end up with 'raw' scores of each of the 280 respondents – classified into 6 distinct groups. In each group, we calculate the group average of the scores and these are called 'summary statistics'. Assume that at the end we end up with the following results:

Table 5 Summary statistics of example 3

Block	Total Score	No. of Respondents	Average Score
1	110	50	2.2
2	135	50	2.7
3	104	50	2.08
4	112	50	2.24
5	124	50	2.48
6	114	30	3.8

Following Nandy and Sinha (2020), we may prepare the following table.

Table 6 Data analysis : theory

Block	No. of Respondents	Expected Block Average	Block Average Score
1 (B_1)	n^*	EBA(1)	$\bar{x}_1$
2 (B_2)	n^*	EBA(2)	$\bar{x}_2$
3 (B_3)	n^*	EBA(3)	$\bar{x}_3$
4 (B_4)	n^*	EBA(4)	$\bar{x}_4$
5 (B_5)	n^*	EBA(5)	$\bar{x}_5$
6 (B_0)	n_0	$\mathbf{\Delta}$	$\bar{x}_0$

In the above,

$$
\begin{aligned}
EBA(1) &= [(3/4)(p_1+p_2+p_3+p_4) + (3/7)[P^* + (\mathbf{\Delta} - p_1 - p_2 - p_3 - p_4)],\\
EBA(2) &= [(3/4)(p_5+p_6+p_7+p_8) + (3/7)[P^* + (\mathbf{\Delta} - p_5 - p_6 - p_7 - p_8)],\\
EBA(3) &= [(3/4)(p_9+p_{10}+p_1+p_2) + (3/7)[P^* + (\mathbf{\Delta} - p_1 - p_2 - p_9 - p_{10})],\\
EBA(4) &= [(3/4)(p_3+p_4+p_5+p_6) + (3/7)[P^* + (\mathbf{\Delta} - p_3 - p_4 - p_5 - p_6)],\\
EBA(5) &= [(3/4)(p_7+p_8+p_9+p_{10}) + (3/7)[P^* + (\mathbf{\Delta} - p_7 - p_8 - p_9 - p_{10})],\\
EBA(6) &= \mathbf{\Delta} = \sum_{i=1}^{i=10} p_i.
\end{aligned}
$$

Summing over all the first five block means, we obtain the estimating equation :

$$\sum_{i=1}^{i=5} \bar{x}_i = (3/4)2\mathbf{\Delta} + (3/7)5P^* + (3/7) \times 3\mathbf{\Delta}.$$

Now we replace $\mathbf{\Delta}$ by its estimate $\bar{x}_0$ and derive

$$\hat{P}^* = [\sum_{i=1}^{i=5} \bar{x}_i - (39/14)\bar{x}_0] \times (7/15) = 0.52.$$

Remark 4 For variance estimation, we may take recourse to 'split-half' method, as explained above.

2.3. Inclusion probability of Q^* is different from other NSBFs

Since Q^* is a sensitive question, we may consider a special status of SBF Q^* and accordingly, its inclusion probability may be taken to be different from the other NSBFs. Considering the same procedure as in the above, we are giving a special status to the sensitive question in the sample s_2 to be drawn from Part B. For this purpose, we propose a scheme to select s_2 questions from k_2 questions as follows:

(i) In the first draw : probability of selection of the sensitive question Q^* is θ_1^* and each of the remaining questions $(k_2 - 1 = v - k_1)$ has selection probability $(1 - \theta_1^*)/(v - k_1)$.

(a) If the sensitive question Q^* is selected at the 1^{st} draw, select remaining questions by adopting $SRSWOR(k_2 - 1, s_2 - 1)$.

(b) If the sensitive question Q^* is not selected at 1^{st} draw, then at the 2^{nd} draw, select the sensitive question Q^* with probability θ_1^{**} and select the remaining questions with probability $(1 - \theta_1^{**})/(v - k_1 - 1)$ each.

In case Q^* has not been selected in the first and second steps, we adopt $SRSWOR(v - k_1 - 1, s_2 - 2)$ to draw the remaining $(s_2 - 2)$ questions from Part B. This time we do not differentiate Q^* from the other left-over NSBFs in the lot.

Now, let π^* and $(\pi_1 = \pi_2 = \ldots = \pi_{(k_2-1)} = \pi_0)$ denote the inclusion probability of sensitive question Q^* and other NSBFs in Part B respectively. It may be verified that

$$\pi^* = \theta_1^* + (1-\theta_1^*)\theta_1^{**} + (1-\theta_1^*)(1-\theta_1^{**})\frac{(s_2-2)}{(v-k_1-1)}.$$

$$\pi_0 = \frac{(1-\theta_1^*)}{(v-k_1)} + \frac{\theta_1^*(s_2-1)}{(v-k_1)} + \frac{(1-\theta_1^*)(1-\theta_1^{**})}{(v-k_1)} + \frac{(1-\theta_1^*)\theta_1^{**}(s_2-2)}{(v-k_1)} + \frac{(1-\theta_1^*)}{(v-k_1)}(1-\theta_1^{**})\frac{(v-k_1-2)(s_2-2)}{(v-k_1-1)}.$$

Since this is a fixed size sampling design, therefore, we have $(v - k_1)\pi_0 + \pi^* = s_2$, which is readily verified. Now, proceeding as before, we derive the estimating equation based on the average of first block total responses:

$$\bar{x}_1 = \sum_{i \in A} p_i + \pi^* P^* + \sum_{i \in A^c} p_i(s_2 - \pi^*)/(k_2 - 1)),$$

and summing over all the blocks, we obtain:

$$\sum_{i=1}^{i=b} \bar{x}_i = r\mathbf{\Delta} + b\pi^* P^* + \frac{(s_2 - \pi^*)(b - r)\mathbf{\Delta}}{(k_2 - 1)}.$$

All the notations are already explained above.

Therefore, the population proportion P^* of sensitive question can be estimated by using the formula

$$\hat{P}^* = \frac{\sum_{i=1}^{i=b} \bar{x}_i - r\hat{\mathbf{\Delta}} - \frac{(s_2-\pi^*)}{(k_2-1)}(b-r)\hat{\mathbf{\Delta}}}{b\pi^*}.$$

Once $\mathbf{\Delta}$ is estimated from the data in the last block, we have the estimate. We skip any illustrative example here.

2.4. Introducing block total response technique involving two sensitive questions

In the above, we discussed block total response technique having only one sensitive question. In this section, we provide an extension of the above technique with two sensitive questions Q_1^* and Q_2^*. There is one more sensitive question in part B of the above case; so total number of questions in Part B is $k_2 = v - k_1 + 2$. Let α, β and γ be the inclusion probabilities of Q_1^*, Q_2^* and each of the rest of the questions respectively in a sample of size s_2 out of k_2. The inclusion probabilities satisfy the identity:$\alpha + \beta + (v - k_1)\gamma = s_2$. There are b blocks; in every block this kind of choice can be made. Suppose P_1^* and P_2^* be the proportions of incidence of the sensitive questions Q_1^* and Q_2^* respectively in the over-all population.

Then we can write

$$\bar{x}_1 = \sum_{i \in A_1} p_i + \alpha_1 P_1^* + \beta_1 P_2^* + (\mathbf{\Delta} - \sum_{i \in A_1} p_i)\gamma_1,$$

and similar relations hold for all the other blocks. In the above equation, A_1 is the collection of all $k = k_1$ NSBFs in Part A of the Q_s in block 1. In each block, Part A deals with k NSBFs and all these are made 'compulsory' for respondents. From Part B, in each block, we select s_2 Q_s out of the collection of k_2 Q_s, including Q^* and Q^{**}.

It is tacitly assumed that we are adopting sampling plans with the following features in Part B:

Table 7 Sampling design: plan parameters

Block	choice of α	choice of β	choice of γ
1 (B_1)	α_1	β_1	γ
2 (B_2)	α_2	β_2	γ
3 (B_3)	α_3	β_3	γ
...	...	...	...
b (B_b)	α_b	β_b	γ

Now, we deduce, adding all the 'b' block averages:

$$\sum_{i=1}^{i=b} \bar{x}_i = r\mathbf{\Delta} + P_1^* \sum_{i=1}^{i=b} \alpha_i + P_2^* \sum_{i=1}^{i=b} \beta_i + \mathbf{\Delta} b\gamma - \gamma(\sum_{i \in A_1} p_i + \sum_{i \in A_2} p_i + \ldots + \sum_{i \in A_b} p_i).$$

The above is further simplified to:

$$\sum_{i=1}^{i=b} \bar{x}_i = r\mathbf{\Delta} + P_1^* \sum_{i=1}^{i=b} \alpha_i + P_2^* \sum_{i=1}^{i=b} \beta_i + (b-r)\mathbf{\Delta}\gamma. \quad (1)$$

From Eqn. (1) we have to find the values of P_1^* and P_2^* under the condition

$$\alpha_i + \beta_i + (v - k_i)\gamma_i = s_2, \text{for all } i = 1, 2, \ldots, b.$$

To solve for P_1^* and P_2^*, we need two separate sets of blocks such that the condition of non-singularity mentioned below is satisfied. From (1), for the two sets, we have

$$\sum_{i=1}^{i=b_1} \bar{x}_i = r_1\mathbf{\Delta} + P_1^* \textstyle\sum_{i=1}^{i=b_1} \alpha_i + P_2^* \sum_{i=1}^{i=b_1} \beta_i + (b_1 - r_1)\mathbf{\Delta}\gamma_1 \tag{2}$$

$$\sum_{i=1}^{i=b_2} \bar{x}_i = r_2\mathbf{\Delta} + P_1^* \textstyle\sum_{i=1}^{i=b_2} \eta_i + P_2^* \sum_{i=1}^{i=b_2} \delta_i + (b_2 - r_2)\mathbf{\Delta}\gamma_2 \tag{3}$$

In (2) and (3), all the notations have their usual significance based on the two separate BTRDs. The non-singularity condition comes from the left-side coefficients of P_1^* and P_2^* in Eqns. (4) and (5) mentioned below.

$$P_1^* \sum_{i=1}^{i=b_1} \alpha_i + P_2^* \sum_{i=1}^{i=b_1} \beta_i = \textstyle\sum_{i=1}^{i=b_1} \bar{x}_i - r_1\mathbf{\Delta} - (b_1 - r_1)\mathbf{\Delta}\gamma_1 \tag{4}$$

$$P_1^* \sum_{i=1}^{i=b_2} \eta_i + P_2^* \sum_{i=1}^{i=b_2} \delta_i = \textstyle\sum_{i=1}^{i=b_2} \bar{x}_i - r_2\mathbf{\Delta} - (b_2 - r_2)\mathbf{\Delta}\gamma_2 \tag{5}$$

In fine, we require

$$(\sum_{i=1}^{i=b_1} \alpha_i)(\sum_{i=1}^{i=b_2} \delta_i) \neq (\sum_{i=1}^{i=b_1} \beta_i)(\sum_{i=1}^{i=b_2} \eta_i) \tag{6}$$

Example 4 We consider $n = 560$ respondents. Let us split the respondents into two sets of 280 respondents each.

Set I

Let us start with the same 'design parameters' i.e., $b_1 = 5, v = 10, k_1 = 4, r_1 = 2$ and the experimental set-up with $n = 280$ respondents. As before, we randomly split them into 6 sets, taking $n^* = 50$ and $n_0 = 30$. We adopt the same block compositions as are displayed above. We have $k_1 = k = 4$ questions in Part A and all these NSBFs are compulsory to respond. We also select $s_2 = 3$ out of $6+2 = 8$ questions in Part B. For Part B, we have to be careful in the sampling design selection. This is explained below.

For implementation of sampling in Part B, we have to make choices of the 'design parameters' i.e., α, β and γ, satisfying the requirement : $\alpha+\beta+6\gamma = 3$. Suppose we make a choice of $\gamma_1 = 0.25$. We have $s_2 = 3$ and hence $\alpha_1 + \beta_1 = 3 - 6\gamma Y_1 = 1.5$. We adopt the following 5 choices for these parameters :

Table 8 Sampling design parameters

Sl. No.	choice of α	choice of β
1	0.5	1.0
2	1.0	0.5
3	0.6	0.9
4	0.9	0.6
5	0.75	0.75

The above choices will serve our purpose for selection of respondents from the 5 distinct blocks. Of course, we are not bothered about the block B_0. Now we have to ensure choice of a sampling design for sample of size 3 from the subpopulation of size 8, which will attain (α, β, γ) as stated in the above table.

Table 9 Sampling design I : α=0.5 , β=1.0 , γ=0.25

sample	probability
(5, 6, Q^{**})	0.25
(7, 8,Q^{**})	0.25
(9,Q^{*}, Q^{**})	0.25
(10, Q^{*}, Q^{**})	0.25

Table 10 Sampling design II : α=1.0 , β=0.5 , γ=0.25

sample	probability
(5, 6, Q^{*})	0.25
(7, 8, Q^{*})	0.25
(9, Q^{*}, Q^{**})	0.25
(10, Q^{*}, Q^{**})	0.25

Table 11 Sampling design III : α=0.6 , β=0.9 , γ=0.25

sample	probability
(5, Q^{*}, Q^{**})	0.10
(6, Q^{*}, Q^{**})	0.10
(7, Q^{*}, Q^{**})	0.10
(8, Q^{*}, Q^{**})	0.10
(9, Q^{*}, Q^{**})	0.10
(10, Q^{*}, Q^{**})	0.10
(5, 6, Q^{**})	0.10
(7, 8, Q^{**})	0.10
(9, 10, Q^{**})	0.10
(5, 7, 9)	0.05
(6, 8, 10)	0.05

Sampling requires 20 cards and one card is to be selected at random. Description of 20 cards – all of same size and shape, identical in appearance from 'exterior'.
Inside:

$$\text{Cards } 1-2 : [5, *, **] \text{ Cards } 3-4 : [6, *, **] \ \ldots \ \text{Cards } 11-12 : [10, *, **]$$

$$\text{Card}s \ 13-14 : [5, 6, **] \text{ Cards } 15-16 : [7, 8, **] \text{ Cards } 17-18 : [9, 10, **]$$

$$\text{Card } 19 : [5, 7, 9] \ \ \text{Card } 20 : [6, 8, 10]$$

Sampling operation involves picking up one card at random and selecting the questions described therein for providing responses. These 3 questions are to be combined with the 4 questions in the Part A and then BTR has to be provided for the all the 7 questions combined without divulging any details. Likewise, we can suggest sampling designs for the other two choices:

$$(\alpha, \beta, \gamma) = (0.9, 0.6, 0.25); = (0.75, 0.75, 0.25).$$

Having implemented the data-gathering tools, we end up with 'raw' scores of each of the 280 respondents – classified into 6 distinct groups. In each group, we calculate the group average of the scores and these are called 'summary statistics'. Assume that at the end we end up with the following results:

Table 12 Summary statistics of example 4 (Set I)

Block	Total Score	No. of Respondents	Average Score
1	120	50	2.4
2	152	50	3.04
3	111	50	2.22
4	140	50	2.80
5	165	50	3.30
6	110	30	3.67

Following Nandy and Sinha (2020), we may prepare the following table.

Table 13 Data analysis : theory

Block	Sample Size	Expected Block Average	Block Average Score
1 (B_1)	n^*	EBA(1)	$\bar{x}_1$
2 (B_2)	n^*	EBA(2)	$\bar{x}_2$
3 (B_3)	n^*	EBA(3)	$\bar{x}_3$
4 (B_4)	n^*	EBA(4)	$\bar{x}_4$
5 (B_5)	n^*	EBA(5)	$\bar{x}_5$
6 (B_0)	n_0	$\boldsymbol{\Delta}$	$\bar{x}_0$

where $n^* = 50, n_0 = 30$ and

$$EBA(1) = \sum_{i=1}^{i=4} p_i + \alpha_1 P_1^* + \beta_1 P_2^* + \gamma_1[\mathbf{\Delta}_1 - \sum_{i=1}^{i=4} p_i]$$
$$EBA(2) = \sum_{i=5}^{i=8} p_i + \alpha_2 P_1^* + \beta_2 P_2^* + \gamma_1[\mathbf{\Delta}_1 - \sum_{i=5}^{i=8} p_i]$$
$$EBA(3) = (p_9 + p_{10} + p_1 + p_2) + \alpha_3 P_1^* + \beta_3 P_2^* + \gamma_1[\mathbf{\Delta}_1 - p_1 - p_2 - p_9 - p_{10}]$$
$$EBA(4) = \sum_{i=3}^{i=6} p_i + \alpha_4 P_1^* + \beta_4 P_2^* + \gamma_1[\mathbf{\Delta}_1 - \sum_{i=3}^{i=6} p_i]$$
$$EBA(5) = \sum_{i=7}^{i=10} p_i + \alpha_5 P_1^* + \beta_5 P_2^* + \gamma_1[\mathbf{\Delta}_1 - \sum_{i=7}^{i=10} p_i]$$
$$EBA(6) = \mathbf{\Delta} = \sum_{i=1}^{i=10} p_i.$$

Now, using the 'method of moments' and summing over all the 5 block means, we obtain :

$$\sum_{i=1}^{i=5} \bar{x}_i = 2\mathbf{\Delta}_1 + P_1^* \sum_{i=1}^{i=5} \alpha_i + P_2^* \sum_{i=1}^{i=5} \beta_i + 3\mathbf{\Delta}_1\gamma_1.$$

Set II

Let us consider the 'design parameters' $b_2 = 4, v = 10, k_2 = 5, r_2 = 2$ and the experimental set-up with $n = 280$ respondents. As before, we randomly split them into $4 + 1 = 5$ sets, taking $n^* = 60$ and $n_0 = 40$.

We adopt the same block compositions as are displayed above. We have $k_2 = k = 5$ questions in Part A and all these NSBFs are compulsory to respond. We also select $s_2 = 4$ out of $5 + 2 = 7$ questions in Part B. For Part B, we have to be careful in the sampling design selection. This is explained below.

For implementation of sampling in Part B, we have to make choices of the 'design parameters' i.e., η, δ, γ, satisfying the requirement $\eta + \delta + 5\gamma_2 = 4$.

Suppose, we choose $\gamma_2 = 0.5$. Since $s_2 = 4$, our choices for the η and δ-parameters may be taken as

$$(\eta, \delta) = (0.5, 1.0); (0.7, 0.8); (0.8, 0.7); (0.9, 0.6).$$

The above 4 choices will serve our purpose for selection of respondents from the 4 distinct blocks. Of course, we are not bothered about the block B_0.

Now we have to ensure choice of a sampling design for sample of size 4 from the subpopulation of size 7, which will attain (η, δ) and γ_2 as stated above.

Table 14 Sampling design I : $\eta = 0.5, \delta = 1.0, \gamma_2 = 0.5$

Sample	Probability
(6, 7, Q^*, Q^{**})	0.25
(8, 9, Q^*, Q^{**})	0.25
(6, 8, 10, Q^{**})	0.25
(7, 9, 10, Q^{**})	0.25

Sampling requires only 10 cards and one card is to be selected at random.

Table 15 Sampling design II : $\eta = 0.7,\ \delta = 0.8,\ \gamma_2 = 0.5$

Sample	Probability
(7, 8, Q^*, Q^{**})	0.1
(9,10, Q^*, Q^{**})	0.1
(6, 7, Q^*, Q^{**})	0.1
(8, 9, Q^*, Q^{**})	0.1
(10, 6, Q^*, Q^{**})	0.1
(7, 8, 9, Q^*)	0.1
(10, 6, 7, Q^*)	0.1
(8, 9, 10, Q^{**})	0.1
(6, 7, 8, Q^{**})	0.1
(9,10, 6, Q^{**})	0.1

Sampling requires only 10 cards and one card is to be selected at random

Table 16 Sampling design III : $\eta = 0.8, \delta = 0.7, \gamma_2 = 0.5$

Sample	Probability
(7, 8, Q^*, Q^{**})	0.1
(9,10, Q^*, Q^{**})	0.1
(6, 7, Q^*, Q^{**})	0.1
(8, 9, Q^*, Q^{**})	0.1
(10, 6, Q^*, Q^{**})	0.1
(7, 8, 9, Q^*)	0.1
(10, 6, 7, Q^*)	0.1
(8, 9, 10, Q^*)	0.1
(6, 7, 8, Q^{**})	0.1
(9,10, 6, Q^{**})	0.1

Sampling requires only 10 cards and one card is to be selected at random

Table 17 Sampling Design IV : $\eta = 0.9, \delta = 0.6, \gamma_2 = 0.5$

sample	probability
(7, 8, Q^*, Q^{**})	0.1
(9,10, Q^*, Q^{**})	0.1
(6, 7, Q^*, Q^{**})	0.1
(8, 9, Q^*, Q^{**})	0.1
(10, 6, Q^*, Q^{**})	0.1
(7, 8, 9, Q^*)	0.1
(10, 6, 7, Q^*)	0.1
(8, 9, 10, Q^*)	0.1
(6, 7, 8, Q^*)	0.1
(9,10, 6, Q^{**})	0.1

Sampling requires only 10 cards and one card is to be selected at random. In each of the above, sampling operation involves picking up one card at random and selecting the questions described therein for providing responses. These 4 questions are to be combined with the 5 questions in the

Part A and then BTR has to be provided for all the 9 questions combined – without divulging any details. Having implemented the data-gathering tools, we end up with 'raw' scores of each of the 280 respondents – classified into 5 distinct groups. In each group, we calculate the group average of the scores and these are called 'summary statistics'. Assume that at the end we end up with the following results:

Table 18 Summary statistics of example 4 (Set II)

Block	Total Score	No. of Respondents	Average Score
1	190	60	3.17
2	218	60	3.63
3	225	60	3.75
4	226	60	3.77
5	152	40	3.80

Following Nandy and Sinha (2020), we may prepare the following table.

Table 19 Data analysis : theory

Block	Sample Size	Expected Block Average	Block Average Score
1 (B_1)	n^*	EBA(1)	$\bar{x}_1$
2 (B_2)	n^*	EBA(2)	$\bar{x}_2$
3 (B_3)	n^*	EBA(3)	$\bar{x}_3$
4 (B_4)	n^*	EBA(4)	$\bar{x}_4$
5 (B_0)	n_0	$\mathbf{\Delta}$	$\bar{x}_0$

In the above, $n^* = 60, n_0 = 40$ and

$$EBA(1) = \sum_{i=1}^{i=5} p_i + \eta_1 P_1^* + \delta_1 P_2^* + \gamma_2[\mathbf{\Delta}_2 - \sum_{i=1}^{i=5} p_i],$$

$$EBA(2) = \sum_{i=6}^{i=10} p_i + \eta_2 P_1^* + \delta_2 P_2^* + \gamma_2[\mathbf{\Delta}_2 - \sum_{i=6}^{i=10} p_i],$$

$$EBA(3) = \sum_{i=2}^{i=6} p_i + \eta_3 P_1^* + \delta_3 P_2^* + \gamma_2[\mathbf{\Delta}_2 - \sum_{i=2}^{i=6} p_i],$$

$$EBA(4) = (\sum_{i=7}^{i=10} p_i + p_1) + \eta_4 P_1^* + \delta_4 P_2^* + \gamma_2[\mathbf{\Delta}_2 - p_1 - \sum_{i=7}^{i=10}],$$

$$EBA(5) = \sum_{i=1}^{i=10} p_i = \mathbf{\Delta}.$$

Summing over all these 4 block means and using the method of moments, we obtain

$$\sum_{i=1}^{i=4} \bar{x}_i = 2\mathbf{\Delta}_2 + P_1^* \sum_{i=1}^{i=4} \eta_i + P_2^* \sum_{i=1}^{i=4} \delta_i + (b_2 - r_2)\mathbf{\Delta}_2\gamma_2.$$

Now from Set I, we have already derived

$$\sum_{i=1}^{i=5} \bar{x}_i = 2\boldsymbol{\Delta}_1 + P_1^* \sum_{i=1}^{i=5} \alpha_i + P_2^* \sum_{i=1}^{i=5} \beta_i + 3\boldsymbol{\Delta}_1 \gamma_1,$$

which, upon using the accrued data, simplifies to

$$13.76 = 2 \times 3.67 + 3.75 P_1^* + 3.75 P_2^* + 3 \times 3.67 \times 0.25. \tag{7}$$

And, upon carrying out similar exercise based on data from Set II, we obtain

$$14.32 = 2 \times 3.8 + 2.9 P_1^* + 3.10 P_2^* + 2 \times 3.8 \times 0.5. \tag{8}$$

The coefficient matrix for P_1^* and P_2^* is non-singular and we readily solve the system to derive

$$\hat{P}_1^* = 0.56; \hat{P}_2^* = 0.42.$$

Remark 5 Any choice for (α, β, γ) is ok subject to the conditions concerning the first order inclusion probabilities. Completely symmetric choice will not do as singularity will prevail.

Remark 6 In finite population inference, to study the nature of incidence of a qualitative attribute, we provide 'point estimate' of the underlying population proportion P and also an interval estimate, under large sample assumption and in simple random sampling. To study the distribution of the estimate of P, we need repeated sampling which is not the case. At best, the over-all sample of size n can be decomposed randomly into k components, each of size m [taking $n = mk$], thereby providing k independent estimates of P and, hence, studying the distribution. We need large k for study of the sampling distribution of $\hat{P}$ and large m for stability of the estimates of P.

Subsampling extends naturally to the study of joint distribution of two or more features as well. Clearly, non-sensitive features can be ackled comfortably. However, for sensitive features, such repeated subsampling operations in large numbers are prohibitive.

3. Conclusions

For a general fixed size sampling design, we developed a version of BTRT having one and two sensitive questions and the results have been deduced under fixed-size unequal probability sampling schemes. The plans have been amply illustrated by examples.

Acknowledgements

We are highly thankful to the anonymous referees for carefully reading the earlier version of the manuscript and offering very useful suggestions which improved the current version substantially. We are also extremely grateful to Dr Sobita Sapam of the Department of Statistics, Manipur University, Imphal for fruitful discussions from time to time with extreme courtesy.

References

Chaudhuri A, Mukerjee R. Randomized response: Theory and applications. New York: Marcel and Dekker; 1988.

Chaudhuri A. Randomized response and indirect questioning techniques in Surveys. Boca Raton, Florida, USA: CRC Press; 2011.

Chaudhuri A, Christofides TC. Indirect questioning in sample surveys. Germany: Springer-Verlag; 2013.

Coutts E, Jann B. Sensitive questions in online surveys: Experimental results for the randomized response technique (RRT) and the unmatched count technique (UCT). Socio Meth Res. 2011; 40(1): 169 - 193.

Fox JA, Tracy PE. A method for sensitive surveys. Quantitative applications in the social sciences; 07-058. California: Sage Publications, Inc; 1986.

Hedayat AS, Sinha BK. Design and inference in finite population sampling. New York: Wiley; 1991.

Mukherjee SP, Sinha BK, Chattopadhyay AK. Statistical methods in social science research. Singapore: Springer Nature; 2018.

Nandy K, Sinha BK. Block total response technique for quantitative sensitive features in a finite population. Stat Appl. 2020; 18(2): 88-95.

Nandy K, Marcovitz M, Sinha BK. Eliciting information on sensitive features: Block total response technique and related inference. In: Chaudhuri R, Christofides TC, Rao CR, eds. Handbook of Statistics, Volume 34: Data Gathering, analysis, and protection of privacy through randomized response techniques: Qualitative and quantitative human traits. Amsterdam: North Holland; 2015. p. 317-328.

Raghavarao D, Federer WT. Block total response as an alternative to the randomized response method in surveys. J Roy Stat Soc B Met. 1979; 41(1): 40-45.

Warner SL. Randomized response: A survey technique for eliminating evasive answer bias. J Am Stat Assoc. 1965; 60: 63-69.

APPENDIX

To ascertain : $(v-k_1)\pi_0+\pi^*=s_2$.
We know

$$\pi_0=\frac{(1-\theta_1^*)}{(v-k_1)}+\frac{\theta_1^*(s_2-1)}{(v-k_1)}+\frac{(1-\theta_1^*)(1-\theta_1^{**})}{(v-k_1)}+\frac{(1-\theta_1^*)\theta_1^{**}(s_2-2)}{(v-k_1)}+\frac{(1-\theta_1^*)(1-\theta_1^{**})(s_2-2)(v-k_1-2)}{(v-k_1)(v-k_1-1)}.$$

Further,

$$\pi^*=\theta_1^*+(1-\theta_1^*)\theta_1^{**}+(1-\theta_1^*)(1-\theta_1^{**})\frac{(s_2-2)}{(v-k_1-1)}.$$

We first simplify $(v-k_1)\pi_0$ as equal to

$$(1-\theta_1^*)+\theta_1^*(s_2-1)+(1-\theta_1^*)(1-\theta_1^{**})+(1-\theta_1^*)\theta_1^{**}(s_2-2)+(1-\theta_1^*)(1-\theta_1^{**})(s_2-2)\frac{(v-k_1-2)}{(v-k_1-1)}.$$

Therefore, we readily have an expression for $(v-k_1)\pi_0+\pi^*$ and upon algebraic simplification, it reduces to s_2.