A COMPARISON OF THREE TECHNIQUES FOR ELICITING TRUTHFUL ANSWERS TO SENSITIVE QUESTIONS

Linda L. Smith, Walter T. Federer and Damaraju Raghavarao[1], Cornell University

1. Introduction

The problem of attempting to obtain truthful answers to sensitive, incriminating, or embarrassing questions has received increased attention since the appearance of Warner's [7] 1965 publication describing the randomized response (RR) technique. Some of this work may be found in the literature [1,2,3,4,8] cited at the end of this paper. An alternative to the RR method, utilizing balanced incomplete block design concepts, was presented by Raghavarao and Federer [5]; their method is designated as the block total response (BTR) procedure. A randomized form of the block total response (RBTR) technique is described by Smith [6]. She compared the three techniques in a designed experiment with v=7 questions in b=7 incomplete blocks of size k=3 questions each on a set of 7n=84 individuals enrolled in a statistics class. This paper includes the results obtained by Smith [6].

Briefly, the RR technique utilized here consisted of presenting the respondent with two questions, A a sensitive one or one of interest and B a nonsensitive one whose answer was known. The randomizing device used was such that 70% of the time the respondent, anonymously, answered question A and 30% of the time he answered question B; an affirmative answer to question B should occur 2/3 of the time. Each respondent was presented k=3 questions with a different randomizing device and a different nonsensitive question being used for each of the three questions presented.

The BTR method consists of using a balanced incomplete block design with parameters v=b, k=r, and λ. The respondent gives a total score for the particular set of k questions presented to him; responses to individual questions are not required and thus, under proper scoring of questions, the answers to individual questions are unknown to anyone but the respondent. From the set of b blocks it is possible to estimate average response to each question but not to obtain an individual's response. For this study a balanced incomplete block design with v=7=b, k=3=r, and λ=1 was utilized.

The RBTR procedure consists of randomly and anonymously allotting a block of k questions to the respondent. The respondent draws a cork from a jar containing corks numbered 1 to b, such that each number occurs an equal number of times. He keeps the cork and writes down a total for the block number on his cork. He then folds the questionnaire and places it in a sealed (or locked) box containing other respondents' answers. For the first respondent a few blank sheets could be inserted in the box to give the respondent a feeling of security regarding his anonymity. For this particular study, only seven corks were included in a jar at one time. This was done because the interviewing was done with groups of six to 14 students. Thus, in the RBTR procedure the interviewer does not know which block of questions was answered by the respondent and in addition the respondent's answers to individual questions are unknown. This double degree of security may be necessary in certain cases. The idea of using a locked box in which respondents placed their answers was described by Arnold J. King to one of the authors (WTF) and was used by King in a survey on income of Iowa farmers. The farmers had to be assured that the information they gave would not end up in the hands of the Internal Revenue Service.

In general survey practice with the RBTR technique, a plastic jar with n corks numbered 1, n corks numbered 2, n corks numbered 3, ..., n corks numbered 7, could be used. Then when one gets down to the last few corks, a check could be made to determine which numbers remained. If only one or two numbers remain, several corks with the same number could be added to give a sense of anonymity to the respondee. It was found that students felt much more secure about their anonymity with procedure III than with the other two procedures. This appears to show up in their answers to questions 2 and 4, where the percentage obtained by procedure III was much higher than for procedures I and II.

2. Procedure

It was decided to use 84 individuals in the study, which, as determined by the results, was too small a sample. The individual respondents were from an introductory statistics class and were interviewed in eight groups, ranging in size from six to 14 students. Since a comparison of the three techniques was being made, the question of representativeness of the sample is not involved. A copy of a four-page questionnaire presented to one student follows:

-page 1-

Introduction: This survey is designed for the purpose of comparing three methods of obtaining truthful answers to questions that an individual might feel are embarrassing, incriminating, or both. All three methods are aimed at keeping the individual's answers completely anonymous. We are not interested in how particular individuals answer specific questions, but rather in estimating the number of individuals responding in a certain way (either yes or no).

Results of this survey will be summarized and analyzed and then presented to you at a later date.

After finishing this survey, some of you may want to construct a similar survey using one or more of these three methods and your own questions for your term project.

When you have completed all three sections, please write down any comments you have concerning the three procedures. Did you feel as though your answers were really anonymous? Why or why not? Write comments on back of this page.

-page 2-

I. Randomized Response Procedure

In this method you are to answer one of two questions according to the procedure indicated. You give only one answer in each pair of questions, and because of the way this method is designed the interviewer cannot tell which question <u>you</u> have answered. The interviewer knows only the proportion of times the question is answered.

1. Answer question A if the last digit of your social security number is 0,1,2,3,4,5 or 6. Answer question B if the last digit is 7,8 or 9.

Question A: Does your parent earn more than $25,000 a year?

Question B: Is your number 7 or 8 (as opposed to 9)?

Answer: Yes_____ No_____

2. Answer question A if your birthday falls between January 1 and September 12. Answer B if your birthday falls between September 13 and December 31.

Question A: Have you smoked any marijuana during the past two weeks?

Question B: Is your birthday between September 13 and November 25?

Answer: Yes_____ No_____

3. Draw a cork and note its color, and if it is white, also note its number. Answer question A if it is red and question B if it is white.

Question A: Are you under 21 years of age?

Question B: Was the number on your cork 1?

Answer: Yes_____ No_____

-page 3-

II. Block Total Response Procedure

For this procedure you are asked to give only a total of your answers to the three questions. Please do not, unless you wish to, indicate what you answer for individual questions, but rather a total for all three. Because you give just the total, the interviewer cannot tell how you have answered individual questions. For example if your total was 4 and if you had been asked questions 1, 2, and 3, you could have answered in any of the following ways:

$$\begin{array}{lll} 1.\ \ 0\ \ \underline{1} & \quad 1.\ \ 0\ \ \underline{1} & \quad 1.\ \ \underline{0}\ \ 1 \\ 2.\ \ \underline{2}\ \ 3 & \quad 2.\ \ 2\ \ \underline{3} & \quad 2.\ \ 2\ \ \underline{3} \\ 3.\ \ \underline{1}\ \ 0 & \quad 3.\ \ 1\ \ \underline{0} & \quad 3.\ \ \underline{1}\ \ 0 \\ \quad\ \ = \ 4 & \quad\quad\ \ = \ 4 & \quad\quad\ \ = \ 4 \end{array}$$

	Possible Answers	
3. In general, are you happy with your decision to come to Cornell?	Yes (1)	No (0)
4. While at Cornell, have you ever stolen money or any other article worth more than $5.00 from a roommate, friend, employer, or anyone else?	Yes (3)	No (2)
6. Have you smoked any marijuana during the past two weeks?	Yes (2)	No (3)

Total for the three questions: _____

-page 4-

III. Randomized form of Block Total Response

Without looking, select a cork from the jar. Observe the number on the cork. This number will correspond to a certain set of questions that you will answer. The seven questions are:

1. Are you under 21 years of age?
   Yes (0) No (1)
2. Did you cheat in any way on the Stat 200 prelim that you took last week?
   Yes (2) No (3)
3. In general, are you happy with your decision to come to Cornell?
   Yes (1) No (0)
4. While at Cornell, have you ever stolen money or any other article worth over $5.00 from a roommate, friend, employer, or anyone else?
   Yes (3) No (2)
5. Does your parent earn more than $25,000 a year?
   Yes (0) No (1)
6. Have you smoked any marijuana during the past two weeks?
   Yes (2) No (3)
7. Are you enrolled in the College of Agriculture and Life Sciences?
   Yes (1) No (0)

Do not give answers to individual questions but rather a total of your three answers.

If the number on your cork was 1, give a total for questions 1,2,4 Answer_____

If number was 2, give total for 2,3,5 Answer_____

If number was 3, give total for 3,4,6 Answer_____

If number was 4, give total for 4,5,7 Answer_____

If number was 5, give total for 5,6,1 Answer_____

If number was 6, give total for 6,7,2 Answer_____

If number was 7, give total for 7,1,3 Answer_____

Now fold this page in half and place in the box provided. This is to insure that your answers remain anonymous.

From the above seven questions listed under III, it should be noted that questions 2 and 4 are highly sensitive ones; question 6 may or may not be sensitive; questions 3 and 5 may be sensitive for some individuals. Under the conditions of the study questions 1 and 7 should not be sensitive for any student. Note that all the scores are not zero and one. The scoring system used was such as to allow several ways in which at least some scores could be obtained. The problem of a good scoring procedure is a difficult and unsolved problem. Some block totals are obtainable in only one way which allows the interviewer to ascertain the respondent's answers to individual questions. Although the scoring procedure used is not perfect, it is a much better one than simply coding all yes answers zero and all no answers one. To improve the scoring procedure one could include a quantitative variable, for example age to nearest birthday, as one of the nonsensitive questions. Alternatively, one could include a nonsensitive question, whose answer is known, in every block.

The seven questions were grouped into v=7 subsets of k=3 questions each in a balanced incomplete block arrangement as follows:

$$\begin{array}{lll} & T_1 = Q_5, Q_6, Q_1 & \quad T_5 = Q_6, Q_7, Q_2 \\ (1) & T_2 = Q_4, Q_5, Q_7 & \quad T_6 = Q_1, Q_2, Q_4 \\ & T_3 = Q_3, Q_4, Q_6 & \quad T_7 = Q_2, Q_3, Q_5 \\ & T_4 = Q_7, Q_1, Q_3 & \end{array}$$

The three procedures were coded as follows:

I = Randomized Response Procedure (RR)

II = Block Total Response Procedure (BTR)

III = Randomized Form of Block Total Response Procedure (RBTR)

Each person answered a set of questions using each of the three methods. Since the order in which the techniques were presented might somehow affect the result, the following set of six sequences of the techniques was used to obtain a balanced arrangement among the groups and order of presentation:

$$S_1 = \text{I, II, III} \qquad S_4 = \text{II, III, I}$$
$$S_2 = \text{I, III, II} \qquad S_5 = \text{III, I, II}$$
$$S_3 = \text{II, I, III} \qquad S_6 = \text{III, II, I}$$

Then for 42 respondees the following setup was used where seven students did S_1, another seven did S_2, etc.

Sets of Seven People in Each S_i

	1	2	3	4	5	6	7
S_1	T_7T_6	T_1T_7	T_2T_1	T_3T_2	T_4T_3	T_5T_4	T_6T_5
S_2	T_2T_3	T_3T_4	T_4T_5	T_5T_6	T_6T_7	T_7T_1	T_1T_2
S_3	T_3T_5	T_4T_6	T_5T_7	T_6T_1	T_7T_2	T_1T_3	T_2T_4
S_4	T_4T_7	T_5T_1	T_6T_2	T_7T_3	T_1T_4	T_2T_5	T_3T_6
S_5	T_5T_2	T_6T_3	T_7T_4	T_1T_5	T_2T_6	T_3T_7	T_4T_1
S_6	T_6T_4	T_7T_5	T_1T_6	T_2T_7	T_3T_1	T_4T_2	T_5T_3

where the first T in each pair corresponds to procedure I and the second T corresponds to procedure II. Since it was not known which T_i an individual would select in procedure III, it was not possible to balance procedure III with the other two. For the RR and BTR procedures the same set of T_h questions for both groups was excluded. A pair of orthogonal latin squares of order seven was used to construct the above design by simply excluding the row of the pair of latin squares which contained the T_hT_h sets of questions. Note that under procedure III, a student may have answered one of the T_i in the pair T_hT_j for procedures I and II.

After the first six blocks of seven had been completed by the 42 respondees, another 42 students followed the same setup to obtain the 84 responses. The experiment was conducted during a one week period in March 1974.

3. Summarized Data and Calculations

The total number of "yes" responses for each question and for each S_j ($j=1,2,\ldots,6$) for the randomized response technique is:

	Number "Yes"			Number "Yes"	
Question	I.1	I.2	Sequence	I.1	I.2
1	12	13	1	10	10
2	3	5	2	8	13
3	15	13	3	9	7
4	3	4	4	8	13
5	4	6	5	11	12
6	6	7	6	9	7
7	12	15			

The total of the responses for each T_h from all 12 S_j for both sets of 42 students for the BTR (II) and the RBTR (III) procedures are:

Total	II	III
$Y_{1.}$	51	42
$Y_{2.}$	43	44
$Y_{3.}$	65	65
$Y_{4.}$	22	21
$Y_{5.}$	77	75
$Y_{6.}$	67	65
$Y_{7.}$	55	57

The calculations for the randomized response technique are:

$$\begin{aligned} P(\text{yes answer}) &= P(\text{Yes on 1st question}) + P(\text{Yes on 2nd question}) \\ &= P(\text{1st question chosen}) \cdot P(\text{Yes/1st question}) + P(\text{2nd question chosen}) \cdot P(\text{Yes/2nd question}) \end{aligned}$$

Rearranging and substituting in the known values, we have

$$P(\text{Yes/1st question}) = \frac{P(\text{Yes answer}) - (.3)(\frac{2}{3})}{0.7} = \hat{\bar{X}}_{h.},$$

where P(Yes answer) is the proportion of "Yes" responses for a particular question and P(Yes/2nd question) = 2/3 for all three nonsensitive questions used in the RR method.

The estimate for question one is obtained as $\hat{\bar{X}}_1 = [(12+13)/3(12) - .20]/0.7 = 0.706$. Note that there are 3(12) = 36 individuals who answered each question using the RR method. The proportions for the remaining questions are given in Table 1.

For a given question h asked of a population of individuals it is assumed that there is a true mean $\bar{X}_h$, that there is an individual response X_{hi} for the ith individual, and that $X_{hi} - \bar{X}_h = e_{hi}$ represents a deviation of individual i from the population mean. Then $E[e_{hi}|h] = 0$ and if the ith and i'th individuals' responses are independent then we can say that the e_{hi} are identically and independently distributed with zero mean and common variance. The responses to the seven questions may be represented as:

$$(2) \qquad \begin{aligned} &X_{1i} = \bar{X}_1 + e_{1i}, \quad X_{2i} = \bar{X}_2 + e_{2i}, \quad X_{3i} = \bar{X}_3 + e_{3i} \\ &X_{4i} = \bar{X}_4 + e_{4i}, \quad X_{5i} = \bar{X}_5 + e_{5i}, \quad X_{6i} = \bar{X}_6 + e_{6i} \\ &X_{7i} = \bar{X}_7 + e_{7i} \end{aligned}$$

where $X_{1i}, X_{2i}, X_{3i}, \ldots, X_{7i}$ are answers to $1,2,3,\ldots,7$ respectively, $\bar{X}_1, \bar{X}_2, \bar{X}_3, \ldots, \bar{X}_7$ are population means for questions $1,2,3,\ldots,7$, and e_{hi} is a deviation of an answer X_{hi} from the population mean $\bar{X}_h$. Let Y_{hj} be the total of the answers for the jth respondent answering the hth set of questions; then

$$(3) \qquad \begin{aligned} &Y_{1j} = X_{5j} + X_{6j} + X_{1j} \qquad Y_{5j} = X_{6j} + X_{7j} + X_{2j} \\ &Y_{2j} = X_{4j} + X_{5j} + X_{7j} \qquad Y_{6j} = X_{1j} + X_{2j} + X_{4j} \\ &Y_{3j} = X_{3j} + X_{4j} + X_{6j} \qquad Y_{7j} = X_{2j} + X_{3j} + X_{5j} \\ &Y_{4j} = X_{7j} + X_{1j} + X_{3j} \end{aligned}$$

Using (2) in (3), and omitting the e_{hi} terms we obtain estimates for individual questions as follows:

$$\begin{aligned} \hat{\bar{X}}_1 &= [Y_{1.}+Y_{4.}+Y_{6.}-(Y_{2.}+Y_{3.}+Y_{5.}+Y_{7.})/2]/3n \\ \hat{\bar{X}}_2 &= [Y_{5.}+Y_{6.}+Y_{7.}-(Y_{1.}+Y_{2.}+Y_{3.}+Y_{4.})/2]/3n \\ \hat{\bar{X}}_3 &= [Y_{3.}+Y_{4.}+Y_{7.}-(Y_{1.}+Y_{2.}+Y_{5.}+Y_{6.})/2]/3n \\ \hat{\bar{X}}_4 &= [Y_{2.}+Y_{3.}+Y_{6.}-(Y_{1.}+Y_{4.}+Y_{5.}+Y_{7.})/2]/3n \\ \hat{\bar{X}}_5 &= [Y_{1.}+Y_{2.}+Y_{7.}-(Y_{3.}+Y_{4.}+Y_{5.}+Y_{6.})/2]/3n \\ \hat{\bar{X}}_6 &= [Y_{1.}+Y_{3.}+Y_{5.}-(Y_{2.}+Y_{4.}+Y_{6.}+Y_{7.})/2]/3n \\ \hat{\bar{X}}_7 &= [Y_{2.}+Y_{4.}+Y_{5.}-(Y_{1.}+Y_{3.}+Y_{6.}+Y_{7.})/2]/3n \end{aligned}$$

where n is the number of people answering a given set of questions. Substitution of numerical values for $Y_{h.}$ in the above equations results in a set of values, some of which exceed unity. Subtracting the number to the left of the decimal results in the proportions given in column three of Table 1.

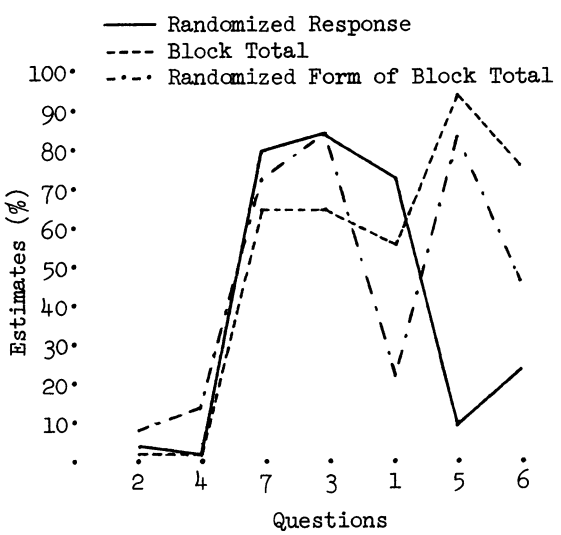
Table 1. Estimated proportions, $\hat{X}_h$, for each of seven questions by three different procedures

Estimate	I - RR	II - BTR	III - RBTR
$\hat{X}_1$	0.71	0.56	0.21
$\hat{X}_2$	0.03	0.01	0.08
$\hat{X}_3$	0.83	0.64	0.83
$\hat{X}_4$	0.01	0.01	0.12
$\hat{X}_5$	0.11	0.93	0.83
$\hat{X}_6$	0.23	0.76	0.46
$\hat{X}_7$	0.79	0.64	0.71

The method for obtaining the estimated proportions using the randomized form of the block total is the same as given above for the block total response. The only difference with this technique is in the manner in which an individual receives the set of three questions to be answered. With this method he chooses a number at random (without replacement) and answers a set of questions corresponding to that number. Thus, the interviewer not only does not know what the respondee has answered for a particular set of questions, but he does not even know which set of questions the respondee has answered.

From the class roster, we could check to find out the number of students who were actually enrolled in the College of Agriculture and Life Sciences (question 7 from the survey). Out of the 97 students enrolled in the course, 71, or 73%, of them were in the College. This value of 0.73 corresponds fairly closely to the estimates obtained by using our three techniques. The discrepancies are 0.06, -0.09, and -0.02 for methods I, II, and III respectively. The graphical representation of the above results is given in Figure 1.

Figure 1. Responses for seven questions from three methods



Considerable discrepancies were obtained for question 5 by the randomized response technique and the other two procedures. From discussions with students it would appear that a figure of 11% of parents with income over $25,000 is an unusually low figure, but that a figure of 93% by the block total response method is somewhat high. Likewise, it is doubtful if 76% of the students smoked marijuana during the past week (question 6 and the block total response technique).

The difference between block totals for question set 1,5, and 6 for the BTR and RBTR methods was relatively large, 51 versus 42. This difference resulted in quite different estimates for these questions by the two different methods. This could be explained as sampling variation and hence the conclusion that the sample size of n=12 was too small for comparing procedures. It was large enough to gain experience in conducting the three procedures in an experiment, and the total of 84 individuals allows a contrast of the three methods.

4. Estimated Variances for the Three Methods

The variance of an estimated percentage $\hat{X}$ from a binomial distribution would be $\hat{X}(1-\hat{X})/kn$. For the randomized response procedure, this variance is divided by π^2, where π is the fraction answering the sensitive question; in our case $\pi = 0.7$. The variance for each question was computed and is given in the second column of Table 3.

Two different methods have been devised for estimating the variances for the block total response procedure. The same methods may be used for the RBTR technique. The Y_{hi}, h=1,2,...,v=7 and i=1,2,...,n, are statistically independent since a simple random sample of n individuals was selected to answer each set T_h of questions and the different sets T_h were used on a different set of n individuals. Hence we may compute the variance of a given $\bar{Y}_{g\cdot} = \Sigma_{i=1}^{n} Y_{gi}/n$ as:

$$(5) \qquad V(\bar{Y}_{g\cdot}) = \left[\sum_{i=1}^{n} Y_{gi}^2 - \left(\sum_{i=1}^{n} Y_{gi}\right)^2 \Big/ n\right] \Big/ n(n-1) .$$

Using formula (5), the various variances were obtained and are presented in Table 2

Table 2. Estimated variances for each $Y_{g\cdot}$ for the BTR and the RBTR procedures

Question Set	II - BTR	III - RBTR
T_1	0.0170	0.0833
T_2	0.0827	0.0657
T_3	0.0372	0.0221
T_4	0.0884	0.0777
T_5	0.0221	0.0625
T_6	0.0221	0.0372
T_7	0.0221	0.0322
Sum	0.2916	0.3807

The first method of estimating the variance of the $\hat{X}_h$ in Table 1 for the BTR method is

(6) $V(\hat{X}_h)$ = [Sum of the variances of $\bar{Y}_{g\cdot}$ for the three blocks where question h occurred plus one-fourth of the variances of the $\bar{Y}_{g\cdot}$ in the blocks where question h did not occur]/9 = [Sum of all 7 variances plus 3(sum of variances of $\bar{Y}_{g\cdot}$ where question h occurred)]/36 .

For example, question 1 occurred in sets T_1, T_4, and T_6 with corresponding block means $\bar{Y}_{1\cdot}$, $\bar{Y}_{4\cdot}$, and $\bar{Y}_{6\cdot}$. The estimated variance of $\hat{X}_1$ is computed from (6) as:

$$
\begin{aligned}
V(\hat{X}_1) &= [V(\bar{Y}_{1.}) + V(\bar{Y}_{4.}) + V(\bar{Y}_{6.}) \\
&\quad + \{V(\bar{Y}_{2.}) + V(\bar{Y}_{3.}) + V(\bar{Y}_{5.}) + V(\bar{Y}_{7.})\}/4]/9 \\
&= [0.2916 + 3(0.0170 + 0.0884 + 0.0221)]/36 \\
&= 0.0187 \; .
\end{aligned}
$$

The remaining variances are computed from (6) and are given in the third column of Table 3. Likewise, the variances for the RBTR method are computed from (6) and are presented in the fourth column of Table 3.

The second method of computing the variances of the $\hat{X}_h$ is somewhat like a variance component procedure in that negative estimates of the variance can occur. The method is to use variances instead of block totals $Y_{s.}$ in equations (4) and solving. For example for $h = 1$,

$$
\begin{aligned}
V(\hat{X}_1) &= [V(\bar{Y}_{1.}) + V(\bar{Y}_{4.}) + V(\bar{Y}_{6.}) \\
&\quad - (V(\bar{Y}_{2.}) + V(\bar{Y}_{3.}) + V(\bar{Y}_{5.}) + V(\bar{Y}_{7.}))/2]/3 \\
&= [3\{V(\bar{Y}_{1.}) + V(\bar{Y}_{4.}) + V(\bar{Y}_{6.})\} - \sum_{s=1}^{7} V(\bar{Y}_{s.})]/6 \\
&= [3\{0.0170 + 0.0884 + 0.0221\} - 0.2916]/6 \\
&= 0.0152 \; .
\end{aligned}
$$

The remaining estimated variances for both the BTR and RBTR procedures may be computed in a similar manner and are presented in Table 3.

Table 3. Estimated variances of $\hat{X}_h$ times 10^4 for the RR, BTR, and RBTR procedures

		First method		Second method	
Variance	RR	BTR	RBTR	BTR	RBTR
$V(\hat{X}_1)$	117	187	271	152	356
$V(\hat{X}_2)$	16	136	216	-154	25
$V(\hat{X}_3)$	80	204	216	252	26
$V(\hat{X}_4)$	6	199	210	224	-10
$V(\hat{X}_5)$	55	182	257	123	272
$V(\hat{X}_6)$	100	145	246	-104	205
$V(\hat{X}_7)$	94	235	277	480	395
Average	67	184	241	139	181

Owing to relatively small sample size, $n=12$, comparisons of variances for individual questions are of little value. Instead, consider the average variances over all questions for the $7n=84$ individuals. By the first method of computing the estimated variances, the average of the variances for the BTR method, 0.0184, and the RBTR method, 0.0241, is roughly three times that for the RR procedure, 0.0067. For the second method of estimating variances for the BTR and RBTR methods, the average variance is roughly twice that for the RR method. Since the estimated $\hat{X}_h$ (Table 1) for the RR method are generally higher (or lower) than the corresponding ones for the BTR and RBTR methods, the average variance for the RR method is smaller than if the $\hat{X}_h$ from the other two procedures had been obtained. Hence, the average variance for the RR procedure is underestimated for comparisons with the other two procedures.

In comparing the estimated variances for the RR and BTR procedures, consider the simplified situation wherein the variances of all $\hat{X}_h$ are equal to σ^2, and further suppose that responses to all k questions in a block are independent. Then, for $k=3$, $V(\hat{X}_h) = 4\sigma^2/3n$ for the first method. Since the RR estimate is obtained on $kn=3n$ individuals, the variance is computed as $\sigma^2/3n\pi^2 = \sigma^2/3n(.7)^2 = \sigma^2/1.47n$. The ratio of the variances is $\sigma^2/3n(.7)^2/4\sigma^2/3n$ which is approximately equal to two. Thus, from variance considerations only the RR procedure is more efficient than the BTR method used here.

If cost is also considered, it will take $k=3$ or more times as long to administer the RR procedure as either of the other two procedures. Increasing the length of the interview and the length of time an interviewer spends with the respondent may be factors completely offsetting any gain in variance efficiency.

5. Discussion

After explaining the three techniques to the students and after attempting to convince them that their answers would truly be anonymous, several students remained skeptical. Some stated that no matter what an interviewer told them, they still would not answer a "sensitive" question truthfully, if a truthful answer could incriminate or embarrass them. They believed that if someone was ingenious enough to think of these techniques, they probably were ingenious enough to determine what an individual's answers had been. The problem of convincing the respondent of his anonymity appears to be the biggest problem associated with obtaining truthful answers. Interviewer training can help to some extent.

On questions 2 and 4, the most sensitive and incriminating ones, the responses for the RR and BTR procedures were much lower than for the RBTR method. Perhaps the reason for this is the increased anonymity obtained in the RBTR procedure, as students felt most anonymous with this method. This result brings out the fact that although the RR procedure may convince a proportion of the sample that their responses remain anonymous, a fraction remains unconvinced. The same is true for the BTR method. This would indicate that the RBTR method should be used in preference to the BTR method and that the technique of using a sealed box as in the RBTR method, with the RR procedure may be useful in practice to increase the truthfulness of responses.

It should be noted that birth dates are probably not uniformly distributed over a year. Hence, the assumption that 30% of the people have birth dates between September 13 and December 31 may not be correct. Since the RR procedure is affected by discrepancies in the proportion π, it might be better to have people whose birth dates fall on the 11, 12, ..., 19th day of a month, answer question B. Then, $1-\pi$ would be $108/365$ ignoring leap years or $[108/366 + 3(108)/365]/4$ including leap years. People may have birth dates uniformly distributed over the days of a month during the year but probably not throughout the year.

Lengthy interviews may be considerably shortened using the BTR or RBTR methods, especially if questions are of a sensitive nature and if the RR technique is used on each question. Hence, situations may arise in which the BTR method is more efficient when both cost and variance are considered. Regardless of efficiency, the surveyor should use whatever method produces the most truthful answers. In this connection the BTR and RBTR procedures can be considered as alternatives to the RR method. There may be, however, procedures yet to be devised, which will replace all these procedures.

References

[1] Abul-Ela, A. A., Greenberg, B. G., and Horvitz, D. G., "A Multi-Proportions Randomized Response Model," Journal of the American Statistical Association, 62 (September 1967), 990-1008.

[2] Campbell, C. and Joiner, B. L., "How to Get the Answer Without Being Sure You've Asked the Question," The American Statistician, 27, No. 5 (December 1973), 229-31.

[3] Greenberg, B. G., Abernathy, J. R., and Horvitz, D. G., "Application of the Randomized Response Technique in Obtaining Quantitative Data," Proceedings of Social Sciences Section, American Statistical Association (August 1969), 40-3.

[4] Greenberg, B. G., Abul-Ela, A. A., Simmons, W. R., and Horvitz, D. G., "The Unrelated Question Randomized Response Model Theoretical Framework," Journal of the American Statistical Association, 64 (June 1969), 520-39.

[5] Raghavarao, D. and Federer, W. T., "Application of BIB Designs as an Alternative to the Randomized Response Method in Survey Sampling," Number BU-490-M in the Mimeo Series of the Biometrics Unit, Cornell University (December 1973).

[6] Smith, L. L., "A Comparison of Three Techniques for Eliciting Answers to Sensitive Questions," B.S. Thesis in Statistics and Biometry, Cornell University (June 1974).

[7] Warner, S. L., "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," Journal of the American Statistical Association, 60 (March 1965), 63-9.

[8] Warner, S. L., "The Linear Randomized Response Model," Journal of the American Statistical Association, 66 (December 1971), 884-8.

[1]Located at Cornell University in Summer, 1973, at the University of Guelph in 1973-74, and at Temple University for 1974-75.