JAMES E. REINMUTH and MICHAEL D. GEURTS*

This article extends the randomized response sampling design to find the intensity of positive action on a sensitive topic among those who have taken a positive action. Sampling properties of the ratio estimate are explored and the model is used to estimate the intensity of shoplifting among shoplifters in a Honolulu shopping center.

The Collection of Sensitive Information Using a Two-Stage, Randomized Response Model

INTRODUCTION

In sampling of consumer populations, two nonsampling errors which frequently distort the research findings involve some respondents refusing to answer all of the questions or deliberately providing incorrect information. Green and Tull [6] have noted that when the respondent is afraid of losing prestige or of becoming embarrassed by offering truthful responses to sensitive questions such distortions may result. The bias produced by these nonsampling errors is sometimes large enough to make the sample estimates seriously misleading.

Nevertheless, personal opinions, controversial topics, and intimate behavior are frequently relevant topics for the marketing researcher. But such topics are very difficult to explore accurately using traditional survey research methods because of their sensitive, often self-incriminating nature.

In 1965, Warner [13] developed an interviewing procedure designed to reduce errors caused by nonresponse and untruthful answers. His procedure is called the *randomized response* technique because the respondent answers one of several questions selected at random. The interviewer is given an answer but is unaware of the question which is being answered by the respondent. (See [2] for an elementary overview of Warner's model.) Warner's original model considered the problem where responses are dichotomous, requiring either a "Yes" or a "No" answer to each of several questions. This model has been modified by a number of different researchers [1, 8, 9]. The most useful version of the randomized response model is the *unrelated question model* which pairs the sensitive question with an innocuous one, as opposed to the grouping of a set of questions all related to some sensitive characteristic. For example, in a study to explore the incidence of default on credit accounts, one might pair the sensitive question, "Have you ever defaulted on a credit account?" with an innocuous one such as, "Does your birthday occur during the month of December?" The unrelated question model is based upon the assumption that the subject is more likely to respond accurately if at least one of the possible questions is not sensitive. He can then avoid self-incrimination as the interviewer is unaware of the question being answered.

Greenberg et al. [7] modified the Warner model allowing the interviewer to ask questions requiring a quantitative response to a sensitive topic. The Greenberg study was carried out to estimate the induced abortion rate in an urban setting in North Carolina. A sensitive question, "How many induced abortions have you had during your lifetime?" was paired with the innocuous question, "How many children do you believe a working woman should have?"

The purpose of this study is to introduce a two-stage, randomized response model and to demonstrate its properties with an empirical example. Two-stage, randomized response sampling combines the results of a qualitative and a quantitative randomized response survey and adds an additional dimension to the sampling of sensitive information.

*James E. Reinmuth is Associate Professor of Quantitative Methods, University of Oregon; Michael D. Guerts is Associate Professor of Marketing, Brigham Young University. The authors wish to thank an unnamed reviewer for his valuable comments on an earlier version of this article.

Journal of Marketing Research
Vol. XII (November 1975), 402-7

TWO-STAGE, RANDOMIZED RESPONSE SAMPLING

The need for a modification to currently available survey techniques arose as a by-product of a study conducted by the authors for the Honolulu Retail Association. The purpose of the Honolulu study was to examine the intensity of shoplifting in a large retail shopping center in Honolulu. More specifically, its objective was to determine the frequency of shoplifting among the sector of shoppers who can be considered shoplifters, i.e., the frequency of repeated acts of shoplifting among shoplifters.

The assumed proportion of shoplifters in the population of all shoppers in the Honolulu shopping center can be estimated by a dichotomous, qualitative randomized response model. The result is a proportion, $\hat{\pi}_s$, estimating the proportion of shoplifters.

Frequency of shoplifting is a quantitative response. However, a typical quantitative, randomized response survey would record a large number of zeros, indicative of the fact that most shoppers have never indulged in the practice of shoplifting. The results of the survey would provide a mean estimate, $\hat{\mu}_s$, representing the average frequency of shoplifting per shopper. In order to estimate the average frequency of shoplifting among those indulging in shoplifting, the qualitative and quantitative responses must be combined in a two-stage ratio estimate, $\hat{\mu}_s/\hat{\pi}_s$. It is the estimate, $\hat{\theta} = \hat{\mu}_s/\hat{\pi}_s$, and its sampling properties that are the central focus of this article.

Using the two-stage approach, it is essential that the qualitative measure, $\hat{\pi}_s$, and the quantitative measure, $\hat{\mu}_s$, be determined from independent, non-overlapping samples. This requirement is necessary for two reasons. First, a subject may be much more reluctant to respond when more than one sensitive question confronts him. Furthermore, if the qualitative and quantitative shoplifting questions are each paired with innocuous ones, the respondent may see an absurdity if, for instance, he first answers that he has never shoplifted and is then asked for the number of times he has shoplifted. Much more reliable results should be expected if different respondent groups provide estimates for π_s and μ_s. The second reason for using separate, independent samples is a theoretical one. Such a design results in estimates $\hat{\pi}_s$ and $\hat{\mu}_s$ which are statistically independent, simplifying distributional properties of the ratio estimate, $\hat{\mu}_s/\hat{\pi}_s$.

The Qualitative Randomized Response Model

Using the unrelated questions model, the observed responses are a set of "Yes" and "No" answers forming a mixture of two distributions—sensitive and innocuous. Under the assumption that the respondents will truthfully answer the question they select, the proportion of "Yes" answers in sample is the mixture:

$$\lambda = p\pi_s + (1 - p)\pi_a, \tag{1}$$

where:

λ = the total proportion of "Yes" responses to both questions,
p = the probability that the sensitive question is selected,
$1 - p$ = the probability that the innocuous question is selected,
π_s = the proportion of "Yes" responses to the sensitive question, and
π_a = the proportion of "Yes" responses to the innocuous question.

As π_a or its estimate $\hat{\pi}_a$ may be known on the basis of past empirical studies, we find:

$$\hat{\pi}_s = \frac{\lambda - (1 - p)\hat{\pi}_a}{p} \tag{2}$$

as the estimate of the proportion of "Yes" responses to the sensitive question.

The variance of the estimate, $\hat{\pi}_s$, is:

$$V(\hat{\pi}_s) = \frac{\lambda(1 - \lambda)}{np^2}, \tag{3}$$

where n is the size of the sample used to find $\hat{\pi}_s$.

If the proportion responding "Yes" to the innocuous question, π_a, is unknown, the sample must be partitioned into two independent, non-overlapping samples of sizes n_1 and n_2, where $n = n_1 + n_2$. We then establish two equations in two unknowns, π_s and π_a, of the form:

$$\begin{aligned} \lambda_1 &= p_1\pi_s + (1 - p_1)\pi_a, \\ \lambda_2 &= p_2\pi_s + (1 - p_2)\pi_a. \end{aligned} \tag{4}$$

Solving the second equation for π_a, we find:

$$\pi_a = (\lambda_2 - p_2\pi_s)/(1 - p_2).$$

Inserting π_a into the first equation, we obtain:

$$\lambda_1 = p_1\pi_s + (\lambda_2 - p_2\pi_s)k,$$

where $k = (1 - p_1)/(1 - p_2)$. The above equation can be rewritten

$$\lambda_1 - k\lambda_2 = p_1\pi_s - p_2k\pi_s.$$

Solving for π_s, we thus find:

$$\hat{\pi}_s = \frac{\lambda_1 - k\lambda_2}{p_1 - kp_2}. \tag{5}$$

Moors [10] has shown that when the probability the sensitive question is chosen in sample i is p_i, $i = 1, 2$, the variance of $\hat{\pi}_s$ computed from (4) is:

$$V(\hat{\pi}_s) = \frac{1}{(p_1 - p_2)^2}\left[\frac{\lambda_1(1 - \lambda_1)(1 - p_2)^2}{n_1} + \frac{\lambda_2(1 - \lambda_2)(1 - p_1)^2}{n_2}\right]. \tag{6}$$

The Quantitative Randomized Response Model

The model outlined by Greenberg et al. [7] allows for the sampling of sensitive information requiring a quantitative response. To apply their model, two independent, non-overlapping samples of sizes n_1 and n_2 are selected. A randomization device is created such that a respondent in the ith sample selects the sensitive question with probability p_i and the innocuous question with probability $1 - p_i$. Under the assumption that the respondents will answer truthfully, an expected response from an individual from either sample is a function of one of the mixture distributions

$$E(Y_1) = p_1\mu_s + (1 - p_1)\mu_a, \text{ for sample 1,}$$

or

$$E(Y_2) = p_2\mu_s + (1 - p_2)\mu_a, \text{ for sample 2.}$$

Again, we have two equations in two unknowns, μ_s and μ_a, the mean response for those answering, respectively, the sensitive and the innocuous questions. Substituting the observed sample means, $\bar{Y}_1$ and $\bar{Y}_2$, for $E(Y_1)$ and $E(Y_2)$, we can find the unbiased estimator for μ_s as:

$$\hat{\mu}_s = \frac{(1 - p_2)\bar{Y}_1 - (1 - p_1)\bar{Y}_2}{p_1 - p_2}. \tag{7}$$

From [7], we find the variance of (7) is:

$$V(\hat{\mu}_s) = \frac{1}{(p_1 - p_2)^2}[(1 - p_2)^2 V(\bar{Y}_1) + (1 - p_1)^2 V(\bar{Y}_2)], \tag{8}$$

an estimate of which can be obtained if we use the sample variances:

$$\hat{V}(Y_i) = s_i^2/n_i, \qquad i = 1,2,$$

where:

$$s_i^2 = \sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_i)^2/n_i.$$

Designing the Survey

Greenberg et al. [7] note that one must be very judicious in selecting the sensitive question relative frequencies, p_1 and p_2. As we can see from the variance expressions (6) and (8), the variances $V(\hat{\pi}_s)$ and $V(\hat{\mu}_s)$ are minimized when p_1 and p_2 are as far apart as possible. In the limit, this would mean that one would select $p_1 = 1$ and $p_2 = 0$ (or vice versa). But such a design would clearly defeat the basic intent of the randomized response model, likely giving results no better than could be expected using direct sampling methods.

Studies evidenced in [7] suggest that p_1 should be chosen as far from 0.5 as possible without creating suspicion on the part of the respondent. Satisfactory results were generally observed with p_1 between 0.7 and 0.8 and $p_2 = 1 - p_1$.

According to [7], the variances of the estimates $\hat{\mu}_s$ and $\hat{\pi}_s$ are minimized when the sample size n is partitioned into independent, non-overlapping subsamples according to the ratio $n_1/n_2 = p_1/p_2$.

When constructing an innocuous question to pair with a sensitive one, it is essential that the distribution of its responses closely coincides with the distribution of responses to the sensitive question. If the means μ_s and μ_a (or the proportions π_s and π_a) differ by a wide margin, the respondent may become suspicious. Similarly, suspicion may become aroused if the sensitive response and innocuous response variances are widely different. For instance, suppose the shoplifting frequency question is paired with one asking the number of glasses of beer drunk by the respondent during the past week. The second question would likely generate responses of zero for nondrinkers or a number of large order for drinkers, causing suspicion in the discerning respondent. General recommendations thus suggest an innocuous question with similar response distribution to that of the sensitive question without requiring that the two questions be related in meaning.

The Two-Stage Model

The two-stage model focuses attention upon the ratio estimate

$$\hat{\theta} = \hat{\mu}_s/\hat{\pi}_s, \tag{9}$$

the expected frequency for those belonging to the "sensitive" category of the population. The estimates $\hat{\mu}_s$ and $\hat{\pi}_s$ are obtained from separate, independent samples. Therefore, the expectation of the ratio estimator (9) is:

$$E(\hat{\theta}) = E[\hat{\mu}_s/\hat{\pi}_s] = E(\hat{\mu}_s)E(1/\hat{\pi}_s).$$

Greenberg et al. [7] show that $\hat{\mu}_s$ and $\hat{\pi}_s$ are unbiased estimators of μ_s and π_s, respectively. But from [11], we find that for a positive random variable x,

$$E(1/x) \geq 1/E(x). \tag{10}$$

Therefore,

$$\begin{aligned} E(\hat{\theta}) &= E(\hat{\mu}_s)E(1/\hat{\pi}_s) \\ &\geq \mu_s/\pi_s. \end{aligned}$$

Cochran [3] makes two notes about the bias inherent in ratio sample estimators. First, he notes that the ratio estimate has a bias of order $1/n$. Thus, for large samples the bias becomes negligible. Second, he proves that if the estimated quantities in the ratio estimate are linearly related through the origin, the ratio estimator is unbiased.

Examining this latter criterion, we are specifying that:

$$E(\mu_s|\pi_s) = \beta\pi_s,$$

for all possible values of π_s. Relating this to our shoplifting application, clearly π_s and $E(\hat{\mu}_s|\pi_s)$ are related through the origin. That is, if the proportion of shoppers who are shoplifters is $\pi_s = 0$, the expected frequency of shoplifting per shopper is zero. The linearity assumption is not as easily justified. As the proportion of shoplifters, π_s, increases, the expected increase in shoplifting per shopper might be assumed to also be increasing. If a shoplifter knows that many others are also indulging in this feloneous behavior, he may feel "comfort in numbers" and be less concerned about apprehension. On the other hand, if he believes π_s to be rather small, the shoplifter may be exhibiting a more cautious behavior.

From (15) in the Appendix, we find that the bias inherent in our ratio estimate is approximately equal to:

$$B(\hat{\theta}) = E(\hat{\theta} - \theta) \doteq (\theta/\pi_s^2)V(\hat{\pi}_s)$$

which is approximated by the term:

$$\hat{B}(\hat{\theta}) = (\hat{\theta}/\hat{\pi}_s^2)V(\hat{\pi}_s). \tag{11}$$

An estimate of θ corrected for bias is then:

$$\begin{aligned}\theta' &= \hat{\theta} - \hat{B}(\hat{\theta}) \\ &= \hat{\theta}(1 - V(\hat{\pi}_s)/\hat{\pi}_s^2).\end{aligned} \tag{12}$$

Notice that the bias and π_s are inversely related. Thus, the smaller the proportion of shoplifters among all shoppers, the more their behavior will be hidden when observing all shoppers, and the fewer will be our measured responses from actual shoplifters. Also from the Appendix, we find that a computational form for an approximate variance for (9) is:

$$\hat{V}(\hat{\theta}) = \frac{1}{\hat{\pi}_s^2}[V(\hat{\mu}_s) + \theta'^2 V(\hat{\pi}_s)]. \tag{13}$$

RESULTS OF THE HONOLULU STUDY

In the Honolulu study, 342 respondents (shoppers) were randomly selected from the Ala Moana Shopping Center of Honolulu, widely acclaimed as the world's largest shopping center. The study was conducted over a period of five consecutive days by two trained interviewers. The randomization device was a black cloth bag containing 75 black and 25 white marbles. Each respondent was given careful instructions on the use of the randomization device to determine the question to be answered. Naturally, his actual selection was confidential.

The first stage in the analysis was to estimate π_s, the proportion of shoppers who have shoplifted. A randomly selected sample of $n = 184$ was divided into subsamples of $n_1 = 138$ and $n_2 = 46$. Each respondent was asked to answer either: (S) "Have you purposefully shoplifted from any Ala Moana retail establishment during the past 12 months?" or (A) "Other than today, have you shopped at Ala Moana during the past week?" The first subsample was instructed to answer question (S) if a black ball is chosen from the bag; the second subsample was instructed to answer (S) after selecting a white ball. The results showed 29 "Yes" answers from the first subsample, 11 "Yes" answers from the second.

To determine the intensity of shoplifting in the population, another non-overlapping sample of $n =$ 158 was selected and divided into subsamples of sizes $n_1 = 126$, $n_2 = 42$. They were asked to answer either: (S) "How many times have you shoplifted from any Ala Moana retail establishment during the past 12 months?" or (A) "Other than today, how many times have you shopped at Ala Moana during the past month?" As before, the first subsample associated question (S) with the selection of a black ball, while the second answered (S) upon the selection of a white ball. The results of the quantitative, randomized response survey are shown in Table 1.

With both samples, the randomization device insured the sensitive selection probabilities, $p_1 = .75$ and $p_2 = .25$. The results can then be derived in a straightforward manner using the computational formulas presented in earlier sections. These results are shown in Table 2.

From Table 2, we can see that about 20% of all shoppers indicate they have shoplifted at Ala Moana. The quantitative survey indicates that the average incidence of shoplifting is about 1.7 such feloneous acts per shopper. We find that shoppers who have shoplifted have done so an average of about 7.9 times each. The wide confidence limits on θ suggest that the average incidence of shoplifting per shoplifter may be as low as 1 incident a year or as high as 15. These limits may, however, be biased on the high side even though they were corrected for estimated bias (12).

Table 1

FREQUENCY OF RESPONSES TO THE QUANTITATIVE RANDOMIZED-RESPONSE SHOPLIFTING SURVEY

Response	*Subsample 1 frequency*	*Subsample 2 frequency*
0	82	13
1	2	4
2	8	3
3	5	4
4	3	5
5	1	3
6	0	2
7	2	0
8	2	1
9	6	2
10	11	5
11	3	0
12	1	0
Totals	$n_1 = 126$	$n_2 = 42$
Means	$\bar{Y}_1 = 2.2936$	$\bar{Y}_2 = 3.4524$
Variances	$s_1^2 = 14.465$	$s_2^2 = 12.303$

Table 2
RESULTS OBTAINED FROM THE RANDOMIZED RESPONSE SURVEYS

	Qualitative survey	*Quantitative survey*	*Ratio estimator*
Mean	$\hat{\pi}_s = .19565$	$\hat{\mu}_s = 1.7142$	$\theta' = 7.9117$
Variance	$V(\hat{\pi}_s) = .00369$	$V(\hat{\mu}_s) = .3315$	$\hat{V}(\hat{\theta}) = 14.6941$
Approximate 95% confidence intervals	.07565 to .31565	0.5642 to 2.8642	0.2517 to 15.5717

It is significant to note that only three of the respondents refused to supply an answer to the randomized response survey. The randomized response technique would thus appear to be eliminating nonresponse bias due to subject sensitivity. However, as in most cases where the randomized response model is used, measures of external verification are not available to monitor the accuracy of our results.

The two-stage survey results provide information on frequency of repetition of shoplifting to law enforcement agencies as well as providing a better base for estimating losses resulting from shoplifting. In further studies, it may be of interest to detect reasons for the deviant behavior of shoplifters. Do they believe the retail stores to be "ripping off" society? Do they steal to support a vice? Do they do so just for the thrill of the risk involved? Identification of the appropriate causes may be a first step in creating effective deterrents to shoplifting.

CONCLUDING REMARKS

Certain evidence exists to show that the randomized response sampling model reduces bias caused by nonsampling errors which might otherwise result when using direct sampling methods to survey sensitive topics. Dowling and Shachtman [4] show that the unrelated question model (4) or (7) is more efficient than direct sampling procedures for any reasonable value of the randomization parameter. In fact, they show the unrelated question model is better if the $\max(p_1, p_2)$ is greater than one-third, a sufficient condition easily established by the model. Folsom et al. [5] suggest a design where the sensitive question and two innocuous questions are separately paired in different samples, yielding two unbiased estimates of μ_s or π_s. The result of their design is a weighted average of the two sample estimates. They found that with μ_a or π_a unknown, their two alternative question model will never be more efficient than the single alternative model. However, sizable gains in efficiency are obtained using the two alternative question model when μ_a or π_a are unknown.

In order for the methods of randomized response sampling to become useful to the marketing researcher, procedures must be designed for externally verifying the results of randomized response surveys. In order to do this, one would have to know in advance π_s or μ_s to see if these values can be verified in selective randomized response surveys. Perhaps the incidence of failing course grades by students, where official transcripts offer π_s and μ_s, could provide such a verification procedure.

Randomized response sampling designs offer the marketing researcher an opportunity to obtain reliable information on topics formerly considered inappropriate for sample surveys. In addition, two-stage randomized response sampling adds an additional dimension to the sampling of sensitive information. Apart from the use illustrated in this article, the two-stage model can prove useful for investigating problems such as:

1. estimating the frequency of drug usage among drug users;
2. estimating the average number of purchases of a sensitive product among regular users of that product;
3. estimating the incidence of white collar crime among employees who have committed such crimes.

APPENDIX

To obtain an estimate of the bias inherent in (9), we shall consider the leading term in the bias when it is expanded in a Taylor series [12]. Let $\theta = \mu_s/\pi_s$. Therefore,

$$\text{(14)}\quad \hat{\theta} - \theta \doteq \frac{\hat{\mu}_s - \theta\hat{\pi}_s}{\pi_s}\left(1 - \frac{\hat{\pi}_s - \pi_s}{\pi_s}\right)$$

$$= \frac{\hat{\mu}_s - \theta\hat{\pi}_s}{\pi_s} - \frac{\hat{\mu}_s(\hat{\pi}_s - \pi_s)}{\pi_s^2} + \frac{\theta\hat{\pi}_s(\hat{\pi}_s - \pi_s)}{\pi_s^2}.$$

The bias $B(\hat{\theta}) = E(\hat{\theta} - \theta)$ is thus approximately:

$$\begin{aligned} E(\hat{\theta} - \theta) &= (1/\pi_s)E(\hat{\mu}_s - \theta\hat{\pi}_s) \\ &\quad - (1/\pi_s^2)E(\hat{\mu}_s(\hat{\pi}_s - \pi)) \\ &\quad + (\theta/\pi_s^2)E(\hat{\pi}_s(\hat{\pi}_s - \pi_s)) \\ &= 0 - (1/\pi_s^2)E(\hat{\mu}_s - \mu_s)(\hat{\pi}_s - \pi_s) \\ &\quad + (\theta/\pi_s^2)E(\hat{\pi}_s - \pi_s)^2. \end{aligned}$$

Now, the term $E(\hat{\mu}_s = \mu_s)(\hat{\pi}_s - \pi_s) = 0$ since μ_s and π_s were estimated from independent, non-overlapping samples. An approximation of the bias is thus:

$$B(\hat{\theta}) = E(\hat{\theta} - \theta) \doteq (\theta/\pi_s^2)E(\hat{\pi}_s - \pi_s)^2$$

$$\text{(15)}\qquad = (\theta/\pi_s^2)V(\hat{\pi}_s).$$

An approximation to the variance of (9) can be obtained using a suggestion offered by Cochran [3]. Since $\theta = \mu_s/\pi_s$,

$$\text{(16)}\qquad (\hat{\theta} - \theta) = \frac{\hat{\mu}_s}{\hat{\pi}_s} - \theta = \frac{\hat{\mu}_s - \theta\hat{\pi}_s}{\hat{\pi}_s}.$$

Cochran [3] points out that if the sample size used in determining $\hat{\pi}_s$ is sufficiently large, $\hat{\pi}_s$ should not differ greatly from π_s. Using π_s for $\hat{\pi}_s$ in the denominator of (16), we find:

$$(\hat{\theta} - \theta) \doteq \frac{\hat{\mu}_s - \theta\hat{\pi}_s}{\pi_s} \tag{17}$$

thus avoiding the problem posed by (10).

Taking the expectation of (17), we find

$$E(\hat{\theta} - \theta) \doteq E\frac{(\hat{\mu}_s - \theta\hat{\pi}_s)}{\pi_s} = \frac{\mu_s - \theta\pi_s}{\pi_s} = 0,$$

since $\hat{\mu}_s$ and $\hat{\pi}_s$ are unbiased estimators of μ_s and π_s, respectively. An approximate variance for (9) is obtained as

$$V(\hat{\theta}) = E(\hat{\theta} - \theta)^2 \doteq \frac{1}{\pi_s^2}E(\hat{\mu}_s - \theta\hat{\pi}_s)^2. \tag{18}$$

Expression (18) does not offer a computational form for the variance. To find one, note that we can write:

$$\begin{aligned} E(\hat{\mu}_s - \theta\hat{\pi}_s)^2 = E[&(\hat{\mu}_s - \mu_s) - \theta(\hat{\pi}_s - \pi_s) \\ &+ (\mu_s - \theta\pi_s)]^2. \end{aligned}$$

However,

$$\mu_s - \theta\pi_s = 0,$$

so that we can write:

$$\begin{aligned} E(\hat{\mu}_s - \theta\hat{\pi}_s)^2 &= E[(\hat{\mu}_s - \mu_s) - \theta(\hat{\pi}_s - \pi_s)]^2 \\ &= E(\hat{\mu}_s - \mu_s)^2 + \theta^2E(\hat{\pi}_s - \pi_s)^2 \\ &\quad - 2\theta E(\hat{\mu}_s - \mu_s)(\hat{\pi}_s - \pi_s). \end{aligned}$$

The latter term in this expression reduces to zero since μ_s and π_s are estimated from independent, non-overlapping samples. Therefore,

$$\begin{aligned} E(\hat{\mu}_s - \theta\hat{\pi}_s)^2 &= E(\hat{\mu}_s - \mu_s)^2 + \theta^2E(\hat{\pi}_s - \pi_s)^2 \\ &= V(\hat{\mu}_s) + \theta^2V(\hat{\pi}_s). \end{aligned} \tag{19}$$

Substituting (19) into expression (18) and using the bias-corrected θ' from (12) for θ, we find a computational form for an approximate variance for (9) as:

$$\hat{V}(\hat{\theta}) = \frac{1}{\hat{\pi}_s^2}[V(\hat{\mu}_s) + \theta'^2V(\hat{\pi}_s)].$$

Approximate $(1 - \alpha)$% confidence limits for θ are then:

$$\theta' \pm z_{\alpha/2}[\hat{V}(\hat{\theta})]^{1/2} \tag{20}$$

bearing in mind that the upward bias in $\hat{\theta}$ may be "shifting" the limits of (20) slightly to the right.

REFERENCES

1. Abdul-Ela, A. Abdel-Latif, Bernard G. Greenberg, and Daniel G. Horvitz. "A Multi-Proportions Randomized Response Model," *Journal of the American Statistical Association*, 62 (September 1967), 990-1008.
2. Campbell, Cathy and Brian L. Joiner. "How to Get the Answer Without Being Sure You've Asked the Question," *The American Statistician*, 26 (December 1973), 229-31.
3. Cochran, William G. *Sampling Techniques*, second edition. New York: John Wiley & Sons, 1963.
4. Dowling, T. A. and Richard Shachtman. "On the Relative Efficiency of Randomized Response Models," *Journal of the American Statistical Association*, 70 (March 1975), 84-8.
5. Folsom, R. E., B. G. Greenberg, D. G. Horvitz, and J. R. Abernathy. "The Two Alternative Questions Randomized Response Model for Human Surveys," *Journal of the American Statistical Association*, 68 (September 1973), 525-31.
6. Green, Paul and Donald S. Tull. *Research for Marketing Decisions*. Englewood Cliffs, N.J.: Prentice Hall, 1970.
7. Greenberg, Bernard G., A. Abdel-Latif Abdul-Ela, Walt R. Simmons, and D. G. Horvitz. "The Unrelated Question Randomized Response Model, Theoretical Framework," *Journal of the American Statistical Association*, 64 (June 1969), 520-39.
8. Greenberg, Bernard G., Roy T. Kuebler, Jr., James R. Abernathy, and D. G. Horvitz. "Application of the Randomized Response Technique in Obtaining Quantitative Data," *Journal of the American Statistical Association*, 66 (June 1971), 243-50.
9. Horvitz, Daniel G., B. V. Shah, and Walt R. Simmons. "The Unrelated Question Randomized Response Model," *Proceedings*. Social Statistics Section, American Statistical Association, 1969, 65-72.
10. Moors, J. A. A. "Optimization of the Unrelated Question Randomized Response Model," *Journal of the American Statistical Association*, 66 (September 1971), 627-9.
11. Sclove, Stanley L., Gordon Simons, and J. Van Ryzin. "Further Remarks on the Expectation of the Reciprocal of a Positive Random Variable," *The American Statistician*, 21 (October 1967), 33-4.
12. Sukhatme, P. V. *Sampling Theory of Surveys with Applications*. Ames: Iowa State College Press, 1954.
13. Warner, Stanley L. "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," *Journal of the American Statistical Association*, 60 (March 1965), 63-9.