

## ABSTRACT

A coding approach, for handling sensitive questions in a group environment, is introduced. This is a variant of the direct questioning approach which still provides considerable protection to the respondents. An actual application of the coding approach is discussed.

## 1. INTRODUCTION

Many practical questionnaires contain questions which are of a sensitive nature. The respondent often either refuses to answer such questions or deliberately falsifies his answers. Such refusals to cooperate can cause large non-sampling errors which could lead to meaningless estimates of the desired parameters.

Warner [3] introduced a randomized response technique for increasing the rate of cooperation. The randomized response approach has recently received considerable attention. An excellent review paper on the subject has been written by Horvitz, et al. [1], where the properties of a number of randomized response techniques have been summarized. For the purpose of this note, it is sufficient to consider only Warner's technique, as the others have similar properties. Let  $A$  be a sensitive characteristic and  $\pi_A$  be the proportion of the population possessing property  $A$ . Warner suggested that the interviewees be provided with a randomization device and be instructed to answer one of the following two questions with probability  $p$  and  $1-p$ , based on the outcome of the randomization device, respectively:

1. I am an  $A$ .
2. I am not an  $A$ .

In this fashion, the interviewee's membership in  $A$  cannot be conclusively determined from the response. Warner's procedure leads to an unbiased maximum likelihood estimator  $\hat{\pi}_A$  of  $\pi_A$  with

$$\text{Var}(\hat{\pi}_A) = (\pi_A)(1-\pi_A)/n + (p)(1-p)/[n(2p-1)^2], \quad p \neq \frac{1}{2}. \quad (1)$$

Here  $n$  is the number of individuals interviewed. It can be seen that  $\text{Var}(\hat{\pi}_A)$  consists of the variance based on direct questioning plus a second component due to randomization. The second term in equation (1) can be quite large when  $p$  is close to  $\frac{1}{2}$ . The advantage of the randomized response technique occurs when the questions are truly sensitive. Randomization would increase the rate of cooperation for sensitive questions and, in this fashion, lead to a viable estimator. On the other hand, direct questioning would be preferred for nonsensitive

questions.

Even though the models incorporating randomized response techniques are theoretically appealing, they have not been widely used in practice. Obviously, the randomized approach of indirect questioning is not as practical to use as to state. In this note, a variant of direct questioning is introduced which continues to provide a reasonable degree of protection and yet is easy to use. This technique is designed for groups of people, such as classes of students. An example is discussed in this note.

## 2. THE CODING APPROACH FOR DICHOTOMOUS QUESTIONS

The author had two back to back statistics I classes, consisting largely of undergraduate business students with rather weak quantitative backgrounds, making it almost irresistible not to experiment on them. The question was how to run a survey consisting of a number of sensitive questions? One constraint was that the survey should consume only a small amount of class time. In addition, it was felt that the students would be hesitant to respond, not only because they were fearful of the interviewer, but also because fellow students could glance at the responses. The latter problem cannot be wholly eliminated by the randomized response approach, as a fellow respondent may possibly notice the color of the marble chosen, from the randomization device, in addition to the final response. For these reasons a coded direct question approach became more appealing.

The survey was conducted on a typical class day. The questionnaire contained two groups of questions: the first group consisted of four rather nonsensitive demographic questions while the second group consisted of five sensitive questions. The four nonsensitive questions were also used as an instructional illustration of the coding procedure. The students were told that the author could decode their answers given a minute or so. They were also informed that the questions were to be distributed randomly and that the questionnaires were to be returned face down and then shuffled in order to avoid any possibility of identifying the interviewees. In other words, a deliberate effort was made to guard the privacy of the respondents.

The five sensitive questions used were:

5. I have used marijuana.
6. I have never used L.S.D.
7. I have used heroine.
8. I have never cheated on an inclass examination at Temple University.

9. I have professionally seen a psychiatrist or psychologist at least ten times.

The first four questions were nonsensitive. It was explained that questions 6 and 8 are asked in the negative, for further confusion to roaming eyes, and discussed how to reply to a double negative. The code used was for a person to lie on any two questions and then compute  $K + 2L$  and  $2L + K$ , where  $L$  and  $K$  are the questions with numbers corresponding to the untruthful replies. One clearly now has two equations and two unknowns and can easily decode the responses. Furthermore, the respondents were asked to compute  $K + 3L$  and  $L + 3K$  in order to allow for a numerical error.

It is interesting to note that out of 66 students attending my class that day, 57 cooperated on the sensitive questions while nine did not. Failure to cooperate consisted of nonsensical answers, such as  $K + 2L = 100$  and  $L + 2K = 100$ . It should further be noted that these students were rather young with 66.67% responding that they were less than 21 years old. The decoded percentages for the sensitive questions, and their standard errors are given in Table 1.

Table 1

Responses to Sensitive Questions

Question	Percentage Of Yes Responses	Standard Error
Used Marijuana	77.2	5.56
Used L.S.D.	14.0	4.60
Used Heroine	3.5	2.43
Cheated on Exam	63.2	6.39
Seen Psychiatrist	10.5	4.06

In this procedure the students were given the choice of the two questions to falsify. The fact that about half the answers were false was especially appealing to the respondents. If a survey were to contain only one or two sensitive questions, then a number of nonsensitive questions should be added in order to be able to use the coded procedure. Similarly, a large number of questions should be subdivided into smaller groups.

One argument against this coding procedure is that if a person were to refuse to answer one dichotomous question, all information for the other questions, in the group, would be lost. In this regard, it can be noticed that if questions are asked individually, then

refusal to respond to a question would be incriminatory. For this reason, the person would either lie on this question or refuse to answer the entire questionnaire.

It should now be apparent that the direct questioning approach has some potential in sensitive question surveys. Comparisons between different survey techniques, such as the one performed by Smith, et al. [2], should include the direct questioning approach. In comparing randomized response techniques with direct questioning, the most ingenious direct questioning approaches should be used. The author hypothesized that in certain types of surveys direct questioning would prove superior, in terms of MSE, as compared to randomized response.

BIBLIOGRAPHY

- [1] Horvitz, Daniel, G., Greenberg, Bernard, G., and Abernathy, James, R., "Recent Developments in Randomized Response Designs," in J.N. Srivastava, ed., A Survey of Statistical Design and Linear Models, New York: North-Holland Publishing Company, 1975, 271-285.
- [2] Smith, L.L., Federer, W.T., and Raghavarao, D., "A Comparison of Three Techniques for Eliciting Truthful Answers to Sensitive Questions," Proceedings of the Social Statistics Section, American Statistical Association (1974).
- [3] Warner, S.L., "Randomized Response. A Survey Technique for Eliminating Evasive Answer Bias," Journal of the American Statistical Association, 60 (1965), 63-69.