Journal of Modern Applied Statistical Methods

Volume 5 | Issue 2 Article 6

11-1-2005

An Alternative to Warner's Randomized Response Model

Sat Gupta
University of North Carolina at Greensboro

Javid Shabbir
Quaid-I-Azam University, jsqau@yahoo.com

Journal of Modern Applied Statistical Methods
November, 2006, Vol. 5, No. 2, 328-331

1538 – 9472/06/$95.00

An Alternative to Warner's Randomized Response Model

Sat Gupta
University of North Carolina at Greensboro

Javid Shabbir
Quaid-I-Azam University

A modification to Warner's (1965) Randomized Response Model is suggested. The suggested model is more efficient than the original model.

Key words: randomized response model, survey bias

Introduction

Warner (1965) suggested an Indirect Questioning method to circumvent social desirability response bias in surveys involving sensitive questions. If π is the proportion of subjects in a population who have a sensitive characteristic A, then Warner's method recommends using a randomization device, such as a deck of cards, to scramble the true response. A known proportion (p) of the cards in the deck have the statement "I have characteristic A" and the remaining cards in the deck have the statement "I do not have characteristic A". A participant in the survey draws a card randomly from the deck and reports his/her agreement/disagreement with the statement on the card. Thus, a respondent who actually has the characteristic A, but draws a "I do not have characteristic A" card, will give a response "no" indicating lack of agreement with the statement on the card. Probability of a "yes" response (p_y) is given by

$$p_y = p\pi + (1-p)(1-\pi). \qquad (1)$$

Sat Gupta is Professor of statistics. He received Ph. D. degree in Mathematics from University of Delhi and Ph. D. degree in Statistics from Colorado State University. Javid Shabbir is Associate Professor of Statistics. Dr Shabbir received a Ph. D. degree in Statistics from Kent University at Canterbury, UK.

Equation (1) can be rewritten as

$$\pi = \frac{p_y - (1-p)}{2p-1}. \qquad (2)$$

Equation (2) suggests estimating π by $\hat{\pi}_w$ where

$$\hat{\pi}_w = \frac{\frac{n_1}{n} - (1-p)}{2p-1}, \; p \neq .5 \qquad (3)$$

where n_1 is proportion of "yes" responses in a simple random sample with replacement of size n. The fact that n_1 has a binomial distribution with parameters (n, p_y) can be used to prove that $\hat{\pi}_w$ is a maximum likelihood estimator of π. Its variance is given by

$$V(\hat{\pi}_w) = \frac{\pi(1-\pi)}{n} + \frac{p(1-p)}{n(1-2p)^2}. \qquad (4)$$

The second term in the above expression is the penalty due to indirect responding. Note that the penalty is smallest when p is closest to zero or one.

Several variations of Warner's model have been proposed in the literature. These include models by Greenberg et. al (1969), Mangat and Siingh (1990) and Christofides (2002). Gupta and Thornton (2002) have attempted to validate some of these models with actual survey data.

Proposed Alternative Strategy

It is clear from (4) that Warner's model works best when p is very close to zero or to one. But, both of these cases make the scrambling deck look very suspicious because almost all of the cards will be of the same type. Using two decks of the type described above is proposed, one with a low value of p (say p_1) and the other with a high value of p (say p_2). This will increase cooperation because the respondent is less suspicious in using decks of both kinds – one with a high value of p and one with a low value of p.

A simple random sample with replacement of size n is selected and each respondent is asked to give a response using each of the two decks. Let $(Z_{1i}, Z_{2i}), i = 1, 2, ..., n$ be the responses where $Z_{ki} = 1$ if the response using the k^{th} deck ($k = 1,2$) is "yes" and $Z_{ki} = 0$ if the response is "no". Let $n_{1i}, i = 1,2$, be the number of "yes" responses from the two decks. Then one can construct two estimators of the type (3). These are given by

$$\hat{\pi}_{w1} = \frac{\frac{n_{11}}{n} - (1 - p_1)}{2p_1 - 1}, \; p_1 \neq .5, \tag{5}$$

$$\hat{\pi}_{w2} = \frac{\frac{n_{12}}{n} - (1 - p_2)}{2p_2 - 1}, \; p_2 \neq .5 \tag{6}$$

It is easy to note from (4) that both of these estimators have the same variance if p_1 and p_2 are symmetric about .5.

We now propose the estimator

$$\hat{\pi}_p = k_1 \hat{\pi}_{w1} + k_2 \hat{\pi}_{w2}, \; k_1 + k_2 = 1. \tag{7}$$

Obviously $\hat{\pi}_p$ is unbiased because both $\hat{\pi}_{w1}$ and $\hat{\pi}_{w2}$ are unbiased. Also, variance of $\hat{\pi}_p$ is given by

$$\begin{aligned} V(\hat{\pi}_p) &= k_1^2 V(\hat{\pi}_{w1}) + k_2^2 V(\hat{\pi}_{w2}) \\ &+ 2k_1 k_2 Cov(\hat{\pi}_{w1}, \hat{\pi}_{w2}) \end{aligned}. \tag{8}$$

The following lemma is proven before exploring this variance further.

Lemma 1:

$$Cov(\hat{\pi}_{w1}, \hat{\pi}_{w2}) = \frac{\pi(1-\pi)}{n} \tag{9}$$

Proof:

Note that $n_{1i} = \sum_i Z_{1i}$.

Hence,

$$\begin{aligned} & Cov(n_{11}, n_{12}) \\ &= Cov(\sum_i Z_{1i}, \sum_i Z_{2i}) \\ &= \sum_i Cov(Z_{1i}, Z_{2i}) \end{aligned}$$

In the above block in the middle equation, please change the second summation to $\sum_j Z_{2j}$

because Z_{1i} and Z_{2j} are independent for $i \neq j$. Hence,

$$\begin{aligned} & Cov(n_{11}, n_{12}) = \\ & nCov(Z_{11}, Z_{12}) = n\{E(Z_{11}.Z_{12}) - E(Z_{11})E(Z_{12})\} \\ & = n\{p_{yy} - p_{y1}p_{y2}\}, \end{aligned} \tag{10}$$

where p_{yy} is the probability of a "yes" response with both decks, p_{y1} is the probability of a "yes" response with Deck 1 and p_{y2} is the probability of a "yes" response with Deck 2. The following is provided as in (1).

$$p_{yi} = p_i \pi + (1 - p_i)(1 - \pi), i = 1,2 \tag{11}$$

and

$$p_{yy} = \pi\, p_1\, p_2 + (1-\pi)(1-p_1)(1-p_2)\,. \quad (12)$$

Substituting (11) and (12) in (10), one can easily obtain

$$Cov(n_{11}, n_{12}) = n\pi(1-\pi)(2p_1-1)(2p_2-1) \quad (13)$$

The lemma follows easily from (5), (6) and (13). Also, it is easy to verify that when $p_1 + p_2 = 1, p_1 \neq p_2$, the optimum values of (k_1, k_2) in (8) are (.5, .5). This is because $V(\hat{\pi}_{w1}) = V(\hat{\pi}_{w2})$ if $p_2 = 1 - p_1$. With these choices for (k_1, k_2), our proposed estimator becomes

$$\hat{\pi}_p = \frac{\hat{\pi}_{w1} + \hat{\pi}_{w2}}{2}\,. \quad (14)$$

As remarked earlier, $\hat{\pi}_p$ is unbiased because both $\hat{\pi}_{w1}$ and $\hat{\pi}_{w2}$ are unbiased.

Theorem 1:

When $p_1 + p_2 = 1$ and $p_1 \neq p_2$, estimator $\hat{\pi}_p$ is more efficient than Warner's estimators $\hat{\pi}_{w1}$ and $\hat{\pi}_{w2}$.

Proof:

Note that

$$V(\hat{\pi}_p) = \frac{1}{4}\{V(\hat{\pi}_{w1}) + V(\hat{\pi}_{w2}) + 2\,Cov(\hat{\pi}_{w1}, \hat{\pi}_{w2})$$

$$= \frac{1}{2}V(\hat{\pi}_{w1}) + \frac{1}{2}\frac{\pi(1-\pi)}{n},$$

because $V(\hat{\pi}_{w1}) = V(\hat{\pi}_{w2})$.

$$< V(\hat{\pi}_{w1})$$

because

$$\frac{\pi(1-\pi)}{n} < V(\hat{\pi}_{w1}), \text{ from (4).}$$

Numerical Examples

In this section, the efficiency of the proposed estimator is compared with Warner's estimator for various choices of π, p_1 and p_2 $(p_2 = 1 - p_1)$. Note that the proposed estimator is more efficient than Warner's estimator, as expected, for all choices of the parameters.

Table1: Efficiency of the proposed estimator compared to Warner's estimator

π	$p_1 = 0.1$	$p_1 = 0.2$	$p_1 = 0.3$	$p_1 = 0.4$
0.1	143.860	171.171	187.940	197.087
0.2	130.529	158.140	180.398	194.937
0.3	125.084	151.414	175.758	193.458
0.4	122.659	148.077	173.222	192.593
0.5	121.951	147.059	172.414	192.308
0.6	122.659	148.077	173.222	192.593
0.7	125.084	151.414	175.758	193.458
0.8	130.529	158.140	180.398	194.937
0.9	143.860	171.171	187.940	197.087

Conclusion

The proposed strategy is likely to induce greater cooperation from the survey participants because it provides greater diversity in the scrambling process. Moreover, the proposed strategy is clearly more efficient than Warner's model, particularly for higher values of p_1.

References

Christofides, T. C. (2002). A generalized randomized response technique, *Metrika, 57*, 195-200

Greenberg, B. G, Abul-ela, A. L. A., Simmons, W. R., & Horvitz, D. G. (1969). The unrelated question randomized response model: Theoretical frame work. *Journal of American Statistical Association, 64*, 520-539.

Gupta, S. & Thornton, B. (2002). Circumventing social desirability response bias in personal interview surveys. *American Journal of Mathematical and Management Sciences, 22*, 369-383.

Mangat, N. S. & Singh, R. (1990), An alternative randomized response procedure. *Biometrika*, *77*, 439-442.

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