International Journal of Statistics and Applications 2015, 5(6): 263-267
DOI: 10.5923/j.statistics.20150506.01

An Improved Warner's Randomized Response Model

F. B. Adebola, O. O. Johnson*

Department of Statistics, Federal University of Technology, Akure, Nigeria

Abstract This paper presents a modification of Warner's [8] Randomized Response model. According to O'Muircheartaigh et al [7], non-response is inevitable in a survey; in view of this, our model further reduces the non-response bias by further sampling for the non-respondents. In this paper we performed an empirical practice of our model and we also performed the empirical comparison of our model with Warner [8] model. We discovered that our model is more efficient than the Warner [8] model.

Keywords Close supervision, Sensitive behavior, Non-respondent, Sub-sample, Randomized response techniques

1. Introduction

Warner [8] proposed the randomized response technique as a survey technique to reduce potential bias due to non-response and social desirability when asking questions about sensitive behaviors (see Warner [8], for a comprehensive review). The method asks respondents to use a randomization device, such as a coin, deck of cards, spinners whose outcome is not known by the enumerator. The outcome of the randomized device determines which of the two questions the respondent answers. A lot of improvements have been done to Warner's randomized response model, to mention few, Greenberg *et al.* [4], Gupta and Shabbir [5], Adebola and Adegoke [1], Adepetun and Adebola [3], Adebola *et al.* [2].

In this paper, we develop a Modification of Warner's Randomized Response Techniques by introducing the concept of sub-sample of non-respondent. Randomized Response Techniques helps to reduce response and non-response bias while our model further reduces the non-response bias. Item non-response occurs when the respondent refuses to answer the sensitive part of the question which is the major concern of the interviewer. In sections that follow, we present the Warner's [8] Randomized Response Model, propose Randomized Response Model and thereafter its relative efficiency over the existing one.

2. Warner's Randomized Response Model

Warner [7] gave a genius idea by using randomized device to encourage truthful answer from the respondent with respect to a sensitive behavior. The randomizing device, such as a spinning arrow, dice or coins is used to select one of the two questions; such as,

"*I am HIV positive*" (class A, presented with probability P)

"*I am HIV negative*" (class B, presented with probability 1-P)

The respondents have the options "Yes" or "No" presented to him or her. The interviewer does not know which question any respondent has answered but knows the probability P and 1-P with which the two statements are presented. Here, with a random sample of n respondents, the interviewer records a binomial estimate $\hat{\theta} = \frac{y}{n}$ of the proportion θ of "Yes" answers, where y is the number of yes answers.

If the questions are answered truthfully, the relation between θ and π in the population is given as:

$$\begin{aligned}\theta &= \pi p + (1-p)(1-\pi) \\ &= (2p-1)\pi + (1-p) \qquad (2.1)\end{aligned}$$

Where π is the proportion of people with the stigmatized or sensitive behavior using Warner's techniques and p is the probability of selecting the sensitive question.

$$\pi = \frac{\theta - (1-p)}{(2p-1)}, \quad p \neq \frac{1}{2} \qquad (2.2)$$

Proof

From equation (2.1), we have

$$\theta = \pi p + (1-p)(1-\pi)$$

Making π the subject of the relation, we have

$$\pi = \frac{\theta - (1-p)}{(2p-1)}$$

* Corresponding author:
teejay21111@gmail.com (O. O. Johnson)
Published online at http://journal.sapub.org/statistics

The unbiased estimator $\hat{\pi}$ of a sensitive proportion π is given by:

$$\hat{\pi} = \frac{\hat{\theta}-(1-p)}{(2p-1)} \tag{2.3}$$

The Variance is given by

$$V(\pi) = V\left(\frac{\hat{\theta}-(1-p)}{(2p-1)}\right)$$

$$V(\pi) = V\left(\frac{\hat{\theta}}{(2p-1)}\right) - V\left(\frac{(1-p)}{(2p-1)}\right)$$

Recall that $V(c)=0$, where c is a constant.

Then, $V\left(\frac{(1-p)}{(2p-1)}\right) = 0$

We have,

$$V(\pi) = V\left(\frac{\hat{\theta}}{(2p-1)}\right)$$

Thus,

$$V(\pi) = \frac{\theta(1-\theta)}{n(2p-1)^2} \tag{2.4}$$

Where $n\hat{\theta}$ follows a binomial distribution,
Now to find the unbiased estimator of the variance

$$\theta = \pi(2p-1)+(1-p)$$

$$1-\theta = \pi(1-2p)+p$$

$$V(\pi) = \frac{\left[\pi(2p-1)+(1-p)\right]\left[\pi(1-2p)+p\right]}{n(2p-1)^2}$$

$$= \frac{\left[4p\pi^2+4p^2\pi-4p^2\pi^2-4p\pi-\pi^2+\pi+p-p^2\right]}{n(2p-1)^2}$$

$$= \frac{p(1-p)}{n(2p-1)^2} + \frac{\left[\pi\left(4p^2-4p+1\right)-\pi^2\left(4p^2-4p+1\right)\right]}{n(2p-1)^2}$$

$$= \frac{p(1-p)}{n(2p-1)^2} + \frac{\left(\pi-\pi^2\right)(2p-1)^2}{n(2p-1)^2}$$

The Variance is given by

$$V(\pi) = \frac{p(1-p)}{n(2p-1)^2} + \frac{\pi(1-\pi)}{n}$$

The unbiased variance estimator $\hat{V}(\hat{\pi})$ of a sensitive proportion $\hat{\pi}$ is given by:

$$\hat{V}(\hat{\pi}) = \frac{p(1-p)}{n(2p-1)^2} + \frac{\hat{\pi}(1-\hat{\pi})}{n} \tag{2.5}$$

The second term in $\hat{V}(\hat{\pi})$ is the variance that $\hat{V}(\hat{\pi})$ would have if all n respondents answered truthfully a direct question about class A membership.

Except by chosen $\boldsymbol{\pi_A}$ near 0.5 and p > 0.85, the first term is greater than the second, often much greater. The method is thus quite imprecise in general. This might be expected since the interviewer does not know whether a "yes" answer implies membership in a class A or the opposite.

However, Warner's method may give a smaller mean square error (MSE) than a direct sensitive question would, if the latter produced numerous refusals or false answers.

3. Our Model

Several randomized response techniques has been developed, the models developed do not take into consideration of item non-response (refusal to answer the sensitive part of the question). Non-response is an important source of non-sampling error in survey sampling, it occurs when some but not all the required information is collected from the sample unit. The most damaging is unit non-response where a sampling unit refuses to answer the sensitive part of the randomized response techniques designed questionnaire.

In view of this, we proposed an improved Warner's randomized response model that is based on sub-sample of non-respondent so as to induce a better estimate of the proportion of people with the stigmatized or sensitive behavior. Questionnaires were sent out and the number of useable responses were recorded (useable responses at the first interview given as n_1) while the remaining were referred to non-response given as n_2. In order to further reduce the non-response bias then a survey on sub-sample of non-respondent come to place. It is assumed that the whole of the sub-sample responded to the survey and are useable responses which would be achieved by close supervision. Close supervision in this context doesn't mean the interviewer knows the question answered by the respondent but it means the respondent is properly instructed and monitored on a one to one basis. Let n_{y1} be the number "yes" response from the respondent at the first interview. Let n_{y2m} be the number of "yes" answer from the sub-sample of non-respondent. Let n be the sample size of the Survey. Let $m\left(=\frac{n_2}{k}\right)$ be the sub-sample size. Let k be the unit which is used to take the sub-sample. The proportion of "yes" response from our model is given by:

$$\theta = \frac{n_y}{n} = \frac{n_{y1} + k\, n_{y2m}}{n} \tag{3.1}$$

By simplifying, we have (see theorem 1 for proof):

$$\hat{\theta} = \hat{\theta}_1 + k\,\hat{\theta}_2 \tag{3.2}$$

Where $\hat{\theta}$ be the proportion of "yes" answer, $\hat{\theta}_1$ be the proportion of "yes" answer form the respondent at the first interview and $\hat{\theta}_2$ be the proportion of "yes" answer from the sub-sample of non-respondent.

From the Warner's randomized response techniques, the proportion of "yes" response from our model is given by:

$$\theta = \pi p + (1-p)(1-\pi) + k\left[\pi p + (1-p)(1-\pi)\right]$$
$$= (k+1)\left[\pi p + (1-p)(1-\pi)\right] \tag{3.3}$$

Let $B = k+1$, then we have:

$$\theta = B\left[\pi p + (1-p)(1-\pi)\right]$$

Solving for π, we have,

$$\pi = \frac{\theta - B(1-p)}{B(2p-1)}$$

Then we have,

$$\pi = \frac{\theta - (k+1)(1-p)}{(k+1)(2p-1)}$$

Recall that $\theta = \theta_1 + k\,\theta_2$

$$\pi = \frac{\theta_1 + k\,\theta_2 - (k+1)(1-p)}{(k+1)(2p-1)}$$

Hence, the unbiased estimator of $\hat{\pi}$ is given by:

$$\hat{\pi} = \frac{\theta_1 + k\,\theta_2 - (k+1)(1-p)}{(k+1)(2p-1)} \tag{3.4}$$

The variance of the estimator is given by:

$$V(\hat{\pi}) = \frac{v(\hat{\theta}) + k^2\, v(\hat{\theta})}{(k+1)^2 (2p-1)^2}$$

$$V(\hat{\pi}) = \frac{\dfrac{\theta_1(1-\theta_1)}{n_1} + k^2\, \dfrac{\theta_2(1-\theta_2)}{n_2}}{(k+1)^2 (2p-1)^2} \tag{3.5}$$

By simplifying, we have

$$V(\hat{\pi}) = \frac{p(1-p)}{n_1 (k+1)^2 (2p-1)^2} + \frac{\pi(1-\pi)}{n_1 (k+1)^2} + \frac{k^2 p(1-p)}{n_2 (k+1)^2 (2p-1)^2} + \frac{k^2 \pi(1-\pi)}{n_2 (k+1)^2}$$

By further simplification,

The unbiased variance estimator $\hat{V}(\hat{\pi})$ of a sensitive proportion π is given by:

$$\hat{V}(\hat{\pi}) = \left(\frac{k-1}{k+1}\right)\left[\frac{p(1-p)}{n(2p-1)} + \frac{\pi(1-\pi)}{n}\right] \tag{3.6}$$

Then,

$$\hat{V}(\hat{\pi}) = \left(\frac{k-1}{k+1}\right) V(\pi_w) \tag{3.7}$$

Where $V(\pi_w)$ is the variance for a Warner's model.

Theorem 1: The proportion of "yes' response, θ is given by:

$$\hat{\theta} = \hat{\theta}_1 + k\,\hat{\theta}_2$$

Proof: From the Hansen and Hurwitz [6], which introduces the concept of subsample of non-respondent, we have

$$\theta = w_1\theta_1 + w_2\theta_{2m}$$

Where $w_1 = \frac{n_1}{n}$, $w_2 = \frac{n_2}{n}$, $\theta_1 = \frac{n_{y1}}{n_1}$, $\theta_{2m} = \frac{n_{y2}}{m}$.

Then, we have

$$\theta = \frac{n_{y1}}{n} + \frac{n_2 n_y}{nm} \tag{3.8}$$

Recall that $m = \frac{n_2}{k}$

By substituting for *m* in equation (3.8), we have

$$\theta = \frac{n_{y1}}{n} + \frac{k\, n_{y2}}{n}$$

The unbiased estimator of θ is given by

$$\hat{\theta} = \hat{\theta}_1 + k\,\hat{\theta}_2$$

Where $\hat{\theta}_1 = \frac{n_{y1}}{n}$ and $\hat{\theta}_2 = \frac{n_{y2}}{n}$.

Theorem 2: Given that k is an integer value and k >1

Show that $\frac{k-1}{k+1} < 1$

Proof:

Given that, $k \succ 1$ Thus when $k = k^2 - k(k-1)$

That is,

$$k^2 - k(k-1) \succ 1$$

By re-arranging,

$$k^2 - 1 \succ k(k-1)$$

$$(k-1)(k+1) \succ k(k-1) \quad (1)$$

Divide through by $(k+1)^2$

$$\frac{k-1}{k+1} \succ \frac{k(k-1)}{(k+1)^2}$$

Multiply through by $\frac{k+1}{k}$

We have,

$$\frac{k-1}{k} > \frac{k-1}{k+1}$$

Then,

$$\frac{k-1}{k+1} \prec \frac{k-1}{k} \quad (2)$$

From equation (1), we have

$$(k-1)(k+1) \succ k(k-1)$$

Divide through by $(k-1)$, we have

$$k+1 \succ k$$

Subtract 1 from both sides, we have

$$k \succ k-1$$

Divide through by k, we have

$$1 \succ \frac{k-1}{k}$$

Then,

$$\frac{k-1}{k} \prec 1 \quad (3)$$

From equation (2) and (3), using transitivity law

We have,

$$\frac{k-1}{k+1} \prec 1$$

4. Comparative Study of our Model

Here we performed the comparative study of our model; this can be achieved mathematically and empirically. Mathematically, it follows that the proposed model is more efficient than the Warner's randomized response model if we have;

Relative efficiency (RE) =

$$\frac{Variance\ of\ proposed\ \mathrm{model}}{Variance\ of\ Warner's\ \mathrm{model}} < 1$$

$$RE = \frac{\left(\frac{k-1}{k+1}\right)\left[\frac{p(1-p)}{n(2p-1)} + \frac{\pi(1-\pi)}{n}\right]}{\frac{p(1-p)}{n(2p-1)^2} + \frac{\pi(1-\pi)}{n}} < 1$$

By simplifying, we have

$$RE = \left(\frac{k-1}{k+1}\right) < 1$$

Since theorem 2 holds, then the variance of our proposed model is less than the variance of Warner's RRT.

Empirically, to also validate our conclusion on the proposed model we present the tables below.

Table 1. Table showing the relative efficiency when n= 250, π = 0.1, p=0.7

N	π	P	K	Warner's Variance	Proposed Variance	Relative Efficiency (%)
250	0.1	0.7	2	0.0056	0.0019	33.33%
250	0.1	0.7	3	0.0056	0.0028	50%
250	0.1	0.7	5	0.0056	0.0037	66.67%
250	0.1	0.7	10	0.0056	0.0046	81.82%
250	0.1	0.7	15	0.0056	0.0049	87.50%

Table 2. Table showing the relative efficiency when n= 500, π = 0.1, p=0.7

N	π	P	K	Warner's Variance	Proposed Variance	Relative Efficiency (%)
500	0.1	0.7	2	0.0028	0.0009	33.33%
500	0.1	0.7	3	0.0028	0.0014	50%
500	0.1	0.7	5	0.0028	0.0019	66.67%
500	0.1	0.7	10	0.0028	0.0023	81.82%
500	0.1	0.7	15	0.0028	0.0025	87.50%

We can deduce from the empirical comparison that the choice of k plays a major role in the comparative study. It can be derived from the table that k = 2 gave the minimum variance in the proposed model; conclusively, the smaller the choice of k, the more efficient the proposed model is over the conventional Warner's model.

5. Conclusions

This paper presented an improved Warner's randomized response model; the proposed strategy further reduces the non-response bias by introducing the concept of sub-samples of non-respondent. The proposed model is likely to induce better estimate with a reduced variance. Moreover, the proposed model is more efficient than the Warner's model. Lastly, we are able to conclude that the smaller the choice of k (the unit which is used to divide the non-respondent so as to have the sub-sample size), the higher the gain in efficiency of the proposed model over the conventional Warner's model.

REFERENCES

[1] Adebola, F.B. and Adegoke, N.A. (2013): A Survey of Examination Malpractices using the Randomized Response Technique. *Journal of the Nigerian Association of Mathematical Physics,* 23, 375-388.

[2] Adebola, F. B., Johnson, O. O., & Adegoke, N. A. (2014): A Modified Stratified Randomized Response Techniques. *Mathematical Theory and Modeling*, *4*(13), 29-42.

[3] Adepetun, A.O. and Adebola, F.B. (2014): On the Relative Efficiency of the Proposed Reparametized Randomized Response Model. *International Journal of Mathematical Theory and Modeling*, 4, 58-67.

[4] Greenberg, B.G., Abul-Ela, A.A., Simmons, W.R. and Horvitz, D.G. (1969): The Unrelated Question Randomized Response: Theoretical Framework. *Journal of the American Statistical Association* 64, 520-539.

[5] Gupta, S.N. and Shabbir, J. (2006): An Alternative to Warner's Randomized Response Model. *Journal of Modern Applied Statistical Methods*, 5, 328-331.

[6] Hansen, M.H. and Hurwitz, W.N. (1946): The Problem of Non-Response in Sample Surveys, *Journal of the American Statistical Association* 41, 517-529.

[7] O'Muircheartaigh, C. and Campanelli, P. (1999): A multilevel exploration of the role of interviewers in survey non-response. *Journal of the Royal Statistical Society, Series A* 162, 437-46.

[8] Warner, S.L. (1965): Randomized response: A Survey Technique for Eliminating Evasive Answer Bias, *Journal of the American Statistical Association* 60, 63-69.