A New Mixed Randomized Response Model

Ayesha Nazuk
NUST Business School
Islamabad, Pakstan
E-mail: Ayesha.nazuk@nbs.edu.pk
Phone: 0092-51-9085-3267

Javid Shabbir
Department of Statistics
Quaid-i-Azam University
Islamabad, Pakistan

Abstract

In this study we present a modification of Kim and Warde (2004) model to estimate the proportion of a qualitative sensitive variable. It has been numerically shown that the proposed model performs better than the model of Kim and Warde (2004) and Moors (1971).

Keywords: Randomized response technique (RRT), simple random sampling with replacement, sensitive variable, and innocuous variable.

1.1 Introduction

Sensitive study variables are often dealt in survey research such as proportion of adulterated milk packs of a company, proportion of illicit drugs usage, etc. Warner (1965) introduced a technique to estimate the true proportion of qualitative sensitive variable. Greenberg et al. (1971) presented a revised version of Warner (1965) technique for qualitative variables. As far as qualitative sensitive variables are concerned, many researchers have modified the Warner (1965) model, some of them include Mangat et al. (1997), Singh et al. (2000), Chang and Huang (2001), Chang et al. (2004), and Gupta et al. (2006).

Kim and Warde (2004) presented a new randomized response model using simple random sample with replacement sampling scheme which improves the privacy of respondents. In this paper we have modified the Kim and Warde (2004) model to estimate the proportion of qualitative sensitive variable. Main aim of modification has been to reduce the variance of estimator, for proportion of qualitative sensitive variable, and to improve the privacy protection of the respondents.

2.1 Proposed Model

Let a random sample of size n be selected using simple random sampling with replacement. Each respondent in the sample is instructed to answer an innocuous question "I possess the innocuous characteristic Y". If the answer to the initial direct question is "Yes" then the respondent is instructed to go to randomization device R_1. otherwise to R_2. Where R_1 consists of two statements (i) "I belong to sensitive group" and (ii) "I belong to the innocuous group", with respective probability p_1 and $(1-p_1)$. While R_2 consists of the same pair of statements as in R_1 but with respective probability p_2 and $(1-p_2)$. In order to offer privacy to the respondents they are not required to tell that which randomizing device they have used. Let n_1 and n_2 be the number of respondents using R_1 and R_2 respectively such that $(n_1+n_2)=n$.

Note that the respondents coming to R_1 have reported a "Yes" to the initial direct question therefore $\pi_y = 1$ in R_1. Denote by X_1 the probability of "Yes" from the respondents using R_1.

Then

$$X_1 = p_1\pi + (1 - p_1)\pi_y = p_1\pi + (1 - p_1). \quad (2.1)$$

An unbiased estimator for the true proportion of the sensitive trait X is as follows

$$\hat{\pi}_{prop(1)} = \frac{\hat{X}_1 - (1 - p_1)}{p_1}, \quad (2.2)$$

where $\hat{X}_1$ is the sample proportion of "Yes" response from the randomizing device R_1.

The variance of $\hat{\pi}_{prop(1)}$ is given by

$$V(\hat{\pi}_{prop(1)}) = \frac{X_1(1-X_1)}{n_1 p_1^2} = \frac{(1-\pi)\left[p_1\pi + (1-p_1)\right]}{n_1 p_1}. \quad (2.3)$$

Note that the respondents using R_2 have reported a "No" to the initial direct question therefore $\pi_y = 0$ in R_2. Denote by X_2 the probability of "Yes" from the respondents using R_2, which is given by

$$X_2 = p_2\pi + (1 - p_2)\pi_y = p_2\pi. \quad (2.4)$$

Proceeding similarly as we have done for R_1, we get

$$\hat{\pi}_{prop(2)} = \frac{\hat{X}_2 - (1 - p_2)}{p_2}, \quad (2.5)$$

where $\hat{X}_2$ is the sample proportion of "Yes" response from the randomizing device R_2.

The variance of $\hat{\pi}_{prop(2)}$ is given by,

$$V(\hat{\pi}_{prop(2)}) = \frac{X_2(1-X_2)}{n_2 p_2^2} = \frac{\pi(1-p_2\pi)}{n_2 p_2}. \quad (2.6)$$

Now we shall pool the two estimators using weights, to form an estimator for π.

$$\hat{\pi}_{prop} = \left(\frac{n_1}{n}\right)\hat{\pi}_{prop(1)} + \left(\frac{n_2}{n}\right)\hat{\pi}_{prop(2)}. \quad (2.7)$$

Now the variance of $\hat{\pi}_{prop}$ is given by,

$$V\left(\hat{\pi}_{prop}\right) = \left(\frac{n_1}{n}\right)^2\left(\frac{(1-\pi)\left[p_1\pi + (1-p_1)\right]}{n_1 p_1}\right) + \left(\frac{n_2}{n}\right)^2\left(\frac{\pi(2-p_1-\pi)}{n_2}\right).$$ where $p_2 = \frac{1}{2-p_1}$ (see Lanke (1975) (2.8)

2.2 Testing the Hypothesis of Truthful Reporting

Various researchers including Chang and Huang (2001) and Chang et al. (2004) incorporated the possibility of less than completely reporting, even with RRT. The proposed model has been equipped with the probabilities of truthful responses. We rewrite the proportion of "Yes" from the two devic R_1 and R_2, incorporating the probability of truthful reporting. Let probability of truthful reporting is denoted by T, where $0 \le T \le 1$ and it is assumed that

i. Let T_i, $i = 1, 2$, be the probability of truthful response in the first and second devices respectively.
ii. Respondents do not lie about the innocuous trait; they may lie for X.

Note that the respondents coming to R_1 have reported a "Yes" to the initial direct question therefore $\pi_y = 1$ in R_1. Denote by X_1^* the probability of "Yes" from the respondents using R_1, which is given by

$$X_1^* = p_1\pi T_1 + (1-p_1)\pi_y = p_1\pi T_1 + (1-p_1). \qquad (2.9)$$

An unbiased estimator for the true proportion of the sensitive trait X is as follows

$$\hat{\pi}_{t(1)} = \frac{\hat{X}_1^* - (1-p_1)}{p_1 T_1}, \qquad (2.10)$$

where $\hat{X}_1^*$ is the sample proportion of "Yes" response from the randomized response R_1.

The variance of $\hat{\pi}_{t(1)}$ is given by

$$V\left(\hat{\pi}_{t(1)}\right) = \frac{X_1^*\left(1-X_1^*\right)}{n_1(T_1 p_1)^2} = \frac{(1-T_1\pi)\left[1-p_1(1-T_1\pi)\right]}{n_1 p_1 T_1^2}. \qquad (2.11)$$

Proceeding similarly as we have done for the first device, we get,

$$X_2^* = p_2\pi T_2 + (1-p_2)\pi_y = p_2\pi T_2 \qquad (2.12)$$

$$\hat{\pi}_{t(2)} = \frac{\hat{X}_2^*}{p_2 T_2}. \qquad (2.13)$$

The variance of $\hat{\pi}_{t(2)}$ is given by

$$V\left(\hat{\pi}_{t(2)}\right) = \frac{\pi(1-p_2\pi T_2)}{n_2 p_2 T_2^2}. \qquad (2.14)$$

Now we formulate the weighted estimator for π

$$\hat{\pi}_t = \left(\frac{n_1}{n}\right)^2 \hat{\pi}_{t(1)} + \left(\frac{n_2}{n}\right)^2 \hat{\pi}_{t(2)}, \qquad (2.15)$$

The variance of $\hat{\pi}_t$ is given by

$$V(\hat{\pi}_t) = \left(\frac{n_1}{n}\right)^2 \left(\frac{(1-T_1\pi)\left[1-p_1(1-T_1\pi)\right]}{n_1 p_1 T_1^2}\right) +$$

$$\left(\frac{n_2}{n}\right)^2 \left(\frac{\pi(2-p_1-\pi T_2)}{n_2 T_2^2}\right) \qquad (2.16)$$

(where $p_2 = 1/(2-p_1)$, (see Lanke (1975))

We can test that whether the probability of truthful reporting is one or less than one in first randomizing device, by testing

$$H_0 : T_1 = 1 \text{ vs } H_1 : T_1 < 0$$

The associated critical region is

$$\frac{X_1^*}{\left[\dfrac{X_1^*\left(1-X_1^*\right)}{\sqrt{n_1}}\right]} \geq z_{\alpha/2}$$

where $z_{\alpha/2}$ is the α_{th} quintile point of the standard normal distribution. Similar analysis can be done for R_2..

2.3 Estimating the True Probability of Truthful Reporting

One may be interested in estimating the true probabilities of truthful reporting in the population belonging to the two random devices.

From (2.9) we construct an unbiased estimator of T_1 as

$$\hat{T}_1 = \frac{X_1^* - (1-p_1)}{p_1\pi}. \qquad (2.17)$$

where $\hat{X}_1^*$ is the sample proportion of "Yes" response from the randomizing device R_1.

The variance of $\hat{T}_1$ is given by

$$V\left(\hat{T}_1\right) = \frac{V\left(X_1^*\right)}{p_1^2\pi^2} = \frac{(1-T_1\pi)\left[1-p_1(1-T_1\pi)\right]}{n_1 p_1\pi^2}. \qquad (2.18)$$

Similarly the estimator for T_2 is obtained from (2.12) as follows

$$\hat{T}_2 = \frac{\hat{X}_2^*}{p_2\pi}.$$

The variance of $\hat{T}_2$ is given by

$$V\left(\hat{T}_2\right) = \frac{T_2(1-p_2\pi T_2)}{n_2 p_2\pi}. \qquad (2.19)$$

3. Efficiency Comparison

An efficiency comparison of the proposed model, under completely truthful responding case, has been done with the Kim and Warde (2004) model.

From Kim and Warde (2004), we have

$$V(\hat{\pi}_{KW}) = \frac{\pi(1-\pi)}{n} + \frac{(1-p_1)\left[\lambda p_1(1-\pi)+(1-\lambda)\right]}{np_1^2}, \qquad (3.1)$$

The PRE of the $\hat{\pi}_{prop}$ with respect to $\hat{\pi}_{KW}$, is as follows,

$$PRE = \frac{V(\hat{\pi}_i)}{V(\hat{\pi}_{prop})} \times 100.$$

$$PRE = \frac{\left[\frac{\pi(1-\pi)}{n} + \frac{(1-p_1)\left[\lambda p_1+(1-\lambda)\right]}{np_1^2}\right]}{\left(\frac{n_1}{n}\right)^2\left(\frac{(1-\pi)\left[p_1\pi+(1-p_1)\right]}{n_1 p_1}\right) + \left(\frac{n_2}{n}\right)^2\left(\frac{\pi(2-p_1-\pi)}{n_2}\right)} \times 100. \qquad (3.2)$$

In Tables 1 and 2, one may easily see that the proposed estimator $\hat{\pi}_{prop}$ is always, at least numerically, efficient than the Kim and Warde's (2004) estimator $\hat{\pi}_{KW}$.

5. Conclusion

In this paper we have modified the Kim and Warde (2004) model. We have also worked in the less than completely truthful reporting but it has been assumed that the respondents may lie about the sensitive study variable but not the innocuous variable.

Acknowledgement

The authors are especially appreciative for Dr. Zawar Hussain of Quaid e University, Islamabad, Pakistan.

References

Chang, H.J and Huang, K.C. (2001): Estimation of proportion and sensitivity of a qualitative character. Metrika, Vol. 53, pp 269-280.

Chang, H.J. , Wang, C,L. and Huang, K.C. (2004): On estimating the proportion of a qualitative sensitive character using randomized response sampling, Quality and Quantity, Vol. 38, pp. 675-680.

Chang, H.J. , Wang, C,L. and Huang, K.C. (2004): Using randomized response to estimate the proportion and truthful reporting probability in a dichotomous finite population, Journal of Applied Statistics, Vol. 31, pp. 565-573.

Gupta, S. , Shabbir, J. and Iembo, R. (2006): Modifications to Warner's model using blank cards, American Journal of Mathematical and Management Sciences, Vol. 26, pp. 185-196.

Greenberg, B.G. , Kubler, R.R. , Abernathy, J.R. and Horvitz, D.G. (1971): Applications of the RR technique in obtaining quantitative data, Journal of the American Statistical Association, Vol. 66, pp. 243-250.

Kim, J. and Warde, W.D. (2004): A mixed randomized response model. Journal of Statistical Planning and Inference, Vol. 133, pp. 211-221

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

Mangat, N.S. , Singh, R. and Singh, S. (1997): Violation of respondent's privacy in Moors' model—its rectification through a random group strategy response model. Communications in Statistics Theory and Methods, Vol. 26 pp. 243–255.

Singh, S. , Singh, R. and Mangat. N.S. (2000): Some alternative strategies to Moor's model in randomized response sampling- a survey technique for eliminating evasive answer bias. Journal of Statistical Planning and Inference, Vol. 83, pp. 243-255.

Table 1

π	$n=100$		PRE				
	n_1	n_2	$p_1=0.1$	$p_1=0.3$	$p_1=0.5$	$p_1=0.7$	$p_1=0.9$
	10	90	8348.62	2011.01	880.000	436.835	193.679
0.1	50	50	1174.19	427.990	272.566	197.739	139.123
	70	30	566.096	246.257	179.591	147.337	120.881
	10	90	7857.96	1554.30	942.010	300.180	149.821
0.2	50	50	1266.49	428.271	321.341	176.397	124.122
	70	30	612.866	251.878	205.115	139.634	113.599
	10	90	7556.78	1311.20	501.204	246.423	134.362
0.3	50	50	1383.69	434.364	245.669	164.474	117.707
	70	30	672.622	260.296	174.496	135.050	110.253
	10	90	7407.60	1167.72	437.500	218.545	126.572
0.4	50	50	1536.53	447.119	240.625	157.428	114.225
	70	30	751.453	272.494	175.000	132.384	108.382
	10	90	7393.66	1080.68	400.000	202.360	121.996
0.5	50	50	1743.11	468.319	240.000	153.482	112.130
	70	30	859.971	290.184	177.777	131.162	107.251
	10	90	7513.81	1030.92	378.181	192.81	119.136
0.6	50	50	2036.36	501.333	244.067	151.988	110.857
	70	30	1018.48	316.546	183.606	131.322	106.584