Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 71(3) 2017 201–205

Randomized Response Techniques with Multiple Responses

Raghunath Arnab, D.K. Shangodoyin and P.M. Kgosi
Department of Statistics, University of Botswana, Gaborone, Botswana
University of Kwazulu-Natal, Durban, South Africa

Received 13 April 2017; Accepted 11 May 2017

SUMMARY

The Odumade and Singh (2009) randomized response (RR) technique involving two decks of cards is used to estimate the proportion of individuals belonging to a certain sensitive group. In this paper, Odumade and Singh (2009) RR technique has been extended to *k*-decks of cards. The proposed alternative estimator for *k*-decks cards is more efficient than the existing Odumade and Singh (2009) estimator for $k = 2$. The proposed variance estimator of the RR technique is simple and nonnegative.

Keywords: Complex sampling designs; Randomized response; Relative efficiency.

1. INTRODUCTION

The data relating to issues that could lead to stigmatization on personality are tedious to obtain with degree of response rate because respondents very often report untrue values or even refuse to respond. To protect the privacy of respondents, prevent an unacceptable rate of non-response and increase quality of data, Warner (1965) introduced an ingenious technique known as the randomized response technique (RR). Warner (1965) technique was modified by Horvitz *et al.* (1967), Greenberg *et al.* (1969), Kim (1978), Franklin (1989), Mangat and Singh (1990), Arnab (2004), Arnab and Mothupi (2015) among other researchers to improve co-operation from the respondents and quality of data. A good review is given by Chaudhuri and Mukherjee (1988), Singh (2003) and Arnab (2017) amongst others.

Odumade and Singh (2009) proposed a RR technique where each of the selected respondents has to perform two independent randomized responses. Several researchers including Singh and Sedory (2011, 2012), Arnab *et al.* (2012), Lee *et al.* (2016) provided with further extensions. Arnab *et al.* (2012), Arnab and Shangodoyin (2015) and Arnab *et al.* (2016) extended the existing results for complex survey designs and a general class of estimators.

In this paper we present a method of estimation of the population proportion π based on *k*-independent randomized responses. The proposed estimator is more efficient than the estimators proposed by Odumade and Singh (2009) as a special case $k = 2$. A simple non-negative unbiased estimator of the variance of the proposed estimator is also provided. The RR techniques relevant to the present study are described below.

1.1 Warner's Technique

In this method a sample s of n units (respondents) is selected from a known population by the simple random sampling with replacement (SRSWR); and the information on a sensitive characteristic say A is obtained by using a RR technique. The RR technique consists of a deck of cards with identical in appearance having one of the following two statements: (i) "I belong to the sensitive group A" and (ii) "I do not belong to the sensitive group A". The statements occur with known relative frequencies P_0 and $(1 - P_0)$, respectively, in the deck of cards. Each respondent in the sample is asked to select one card at random from

Corresponding author: Raghunath Arnab
E-mail address: arnabr@mopipi.ub.bw

the well-shuffled deck. Without showing the card to the interviewer, the interviewer answers the question, "Is the statement true?" The respondent answers "Yes" or "No" truthfully to the interviewer. Confidentiality of the respondent is maintained because the interviewer will not know which question the respondent has answered (see Arnab *et al.* 2016). Such a RR trial will be termed as Warner's trial with parameter P_0.

Let n_1 be the total number of "Yes" answers obtained from the sampled respondents. Warner (1965) proposed the following unbiased estimator for π:

$$\hat{\pi}_w = \frac{\lambda_w - (1-P_0)}{2P_0 - 1} \tag{1.1}$$

where $\lambda_w = n_1/n$ = proportion of "Yes" answers and $P_0 \neq 0.5$.

The variance of the estimator $\hat{\pi}_w$ is given by:

$$V(\hat{\pi}_w) = \frac{\pi(1-\pi)}{n} + \frac{P_0(1-P_0)}{n(2P_0-1)^2} \tag{1.2}$$

1.2 Odumade and Singh's Strategy

In Odumade and Singh (2009), a sample of size n is selected by SRSWR. Each of the selected respondents in the sample is asked to select two cards, one card from Deck-I and the other from Deck-II. Each of the decks consists of two types of cards as in the Warner (1965) model. The proportion of cards bearing the statement "I belong to the sensitive group A" in Deck-I and Deck-II are P_1 and P_2, respectively. The respondent is asked to report his/her response as (*X, Y*) where *X* indicates response from the card selected from Deck-I while *Y* indicates response from the card selected from Deck-II. For example, if a respondent selects a card written "I belong to the sensitive group A" from the Deck-I and selects the other card written "I do not belong to the sensitive group *A*" from the Deck-II, then he/she will give a response (Yes, No) if he/she belongs to the sensitive group *A*. On the other hand if the respondent does not belong to the group A, he/she will supply (No, Yes) as his/her response. So, in Odumade and Singh's (2009) RR technique each of the respondents performs two Warner's (1965) trials independently with parameters P_1 and P_2 respectively. Let n_{11}, n_{10}, n_{01} and n_{00} denote, respectively, the frequencies of the responses (Yes, Yes), (Yes, No), (No, Yes) and (No, No). The cell frequencies and their respective probabilities (in brackets) are given in the following table (see Arnab *et al.* 2016).

		Deck-II		**Total**
		Yes	**No**	
Deck-1	Yes	$n_{11}(\theta_{11})$	$n_{10}(\theta_{10})$	$n_{1\bullet}$
	No	$n_{01}\ (\theta_{01})$	$n_{00}(\theta_{00})$	$n_{0\bullet}$
Total		$n_{\bullet 1}$	$n_{\bullet 0}$	n

$$\theta_{11} = \pi P_1 P_2 + (1-\pi)(1-P_1)(1-P_2);\ \theta_{10} = \pi P_1(1-P_2) + (1-\pi)(1-P_1)P_2$$
$$\theta_{01} = \pi(1-P_1)P_2 + (1-\pi)P_1(1-P_2);\ \theta_{00} = \pi(1-P_1)(1-P_2) + (1-\pi)P_1P_2$$

Odumade and Singh (2009) proposed an unbiased estimator for the population proportion π as

$$\hat{\pi}_{os} = \frac{1}{2} + \frac{(P_1+P_2-1)(n_{11}-n_{00}) + (P_1-P_2)(n_{10}-n_{00})}{2n\left\{(P_1+P_2-1)^2 + (P_1-P_2)^2\right\}} \tag{1.3}$$

The variance of the estimator $\hat{\pi}_{os}$ and an unbiased estimator of the variance of $\hat{\pi}_{os}$ are, respectively, given by:

$$V(\hat{\pi}_{os}) = \frac{(P_1+P_2-1)^2\{P_1P_2+(1-P_1)(1-P_2)\} + (P_1-P_2)^2\{P_2(1-P_1)+P_1(1-P_2)\}}{4n\left[(P_1+P_2-1)^2+(P_1-P_2)^2\right]^2} - \frac{(2\pi-1)^2}{4n} \tag{1.4}$$

and

$$\hat{V}(\hat{\pi}_{os}) = \frac{1}{4(n-1)}\left[\frac{(P_1+P_2-1)^2\{P_1P_2+(1-P_1)(1-P_2)\}+(P_1-P_2)^2\{P_2(1-P_1)+P_1(1-P_2)\}}{[(P_1+P_2-1)^2+(P_1-P_2)^2]^2} - (2\hat{\pi}_{os}-1)^2\right] \tag{1.5}$$

The extension of Odumade and Singh (2009) RR methodology for general *k*-decks of cards is not straight forward as it needs to prepare 2^k consistency table of responses like (Yes, Yes,…..). In the next subsection we propose an alternative method of estimation of π for *k*-decks of cards. The proposed estimator is found to be more efficient than the Odumade and Singh (2009) estimator. An unbiased estimator of the variance of the proposed estimator is provided. The proposed variance estimator is very simple and non-negative. The proposed method can be extended for general complex sampling design following Arnab *et al.* (2016).

2. PROPOSED STRATEGY

Assume a finite population $U = (1,..,N)$ of N identifiable units. Let y_i be the value of the variable under study y for the ith unit. Denote y_i = 1, if the ith unit belongs to the sensitive group A and y_i = 0 if the ith unit does not belong to A. So the proportion of individuals belonging to the sensitive group A in the population is $\pi = \sum_{i=1}^{N} y_i / N$. Let a sample s of size n be selected from the finite population of N units by SRSWR method and each of the selected respondents performs k RR trails independently by Warner's (1965) method with parameter P_i for the ith trial $i = 1,..,k$. Let us define

$$Z_i(j) = \begin{cases} 1 \text{ if the } i\text{th respondent answers "Yes" for the } j\text{th RR trial} \\ 0 \text{ if the } i\text{th respondent answers "No" for the } j\text{th RR trial} \end{cases}$$

$$i = 1,..,n;\ j = 1,..,k \quad (2.1)$$

Then,

$$\text{Prob} = \{Z_i(j) = 1\} = \begin{cases} P_j \text{ if the } i\text{th respondent } \in \text{A} \\ 1 - P_j \text{ if the } i\text{th respondent } \notin \text{A} \end{cases}$$

$$= y_i P_j + (1 - y_i)(1 - P_j) \quad (2.2)$$

Let $E_P(E_R)$, $V_P V_R$ and $C_P(C_R)$ denote expectation, variance and covariance with respect to the sampling design (randomized response) respectively. Then, we have

$$E_R\left(Z_i(j)\right) = y_i P_j + (1 - y_i)(1 - P_j),\ V_R\left(Z_i(j)\right) = P_j(1 - P_j)$$
and $C_R\left(Z_i(j), Z_{i'}(j')\right) = 0$ for $i, j \neq (i', j')$;
$$i, i' = 1,..,N;\ j, j' = 1,..,k \quad (2.3)$$

The equation (2.3) indicates $\hat{y}_i(j) = \dfrac{Z_i(j) - (1 - P_j)}{2P_j - 1}$ is an unbiased estimator of y_i with variance $\mathrm{P}_j(1 - \mathrm{P}_j) / \left(2\mathrm{P}_j - 1\right)^2$ in the sense

$$E_R\left(\hat{y}_i(j)\right) = y_i, \quad V_R\left(\hat{y}_i(j)\right) = \frac{P_j(1 - P_j)}{\left(2P_j - 1\right)^2} = \phi_j \quad \text{and}$$

$$C_R\left(\hat{y}_i(j), \hat{y}_{i'}(j)\right) = 0 \text{ for } (i, j) \neq (i', j') \quad (2.4)$$

Let

$$\hat{\lambda}_i = \frac{1}{k}\sum_{j=1}^{k} w_j \hat{y}_i(j) \quad (2.5)$$

where $w_j\left(0 < w_j < 1\right)$ are weights which make $\hat{\lambda}_i$ unbiased for y_i and minimize variance $V_R\left(\hat{\lambda}_i\right)$. Obviously optimum choices of w_j s are

$$w_{j0} = \left(1/\phi_j\right) / \left(\sum_{j=1}^{k} 1/\phi_j\right) \quad (2.6)$$

The optimum value of $\hat{\lambda}_i$ with $w_j = w_{j0}$ is given by

$$\hat{\lambda}_{i0} = \sum_{j=1}^{k} w_{j0} \hat{y}_i(j) \quad (2.7)$$

Theorem 2.1.

(i) $\hat{\pi}_w = \frac{1}{n}\sum_{i=1}^{n} \lambda_{i0}$ is an unbiased estimator of π

(ii) $V\left(\hat{\pi}_w\right) = \dfrac{\pi(1-\pi)}{n} + \dfrac{\bar{\phi}_w}{n}$

and

(iii) $\hat{V}\left(\hat{\pi}\right) = \dfrac{1}{n(n-1)}\sum_{i=1}^{n}\left(\hat{\lambda}_{i0} - \hat{\pi}_w\right)^2$ is an unbiased estimator of $V\left(\hat{\pi}_w\right)$

where $\bar{\phi}_w = 1 / \sum_{i=1}^{k}\left(1/\phi_i\right)$

Proof:

(i) $$E\left(\hat{\pi}_w\right) = E_p\left[\frac{1}{n}\sum_{i=1}^{n} E_R\left(\hat{\lambda}_{i0}\right)\right] = E_p\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = \pi$$

(ii) $$V\left(\hat{\pi}_w\right) = V_p\left[\frac{1}{n}\sum_{i=1}^{n} E_R\left(\hat{\lambda}_{i0}\right)\right] + E_p\left[\frac{1}{n^2}\sum_{i=1}^{n} V_R\left(\hat{\lambda}_{i0}\right)\right]$$
$$= V_p\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) + E_p\left(\frac{1}{n^2}\sum_{i=1}^{n} \bar{\phi}_w\right)$$
$$= \frac{\pi(1-\pi)}{n} + \frac{\bar{\phi}_w}{n} \quad (2.8)$$

(iii) $$E\left[\hat{V}\left(\hat{\pi}_w\right)\right] = \frac{1}{n(n-1)}\left[E_p\left\{\sum_{i=1}^{n} E_R\left(\hat{\lambda}_{i0}^2\right)\right\} - n\left\{V\left(\hat{\pi}_w\right) + \pi^2\right\}\right]$$
$$= \frac{1}{n(n-1)}\left[E_p\left\{\sum_{i=1}^{n}\left(y_i + \bar{\phi}_w\right)\right\} - n\left\{V\left(\hat{\pi}_w\right) + \pi^2\right\}\right]$$
$$= V\left(\hat{\pi}_w\right)$$

3. COMPARISON WITH THE ODUMADE AND SINGH (2009) STRATEGY

The proposed estimator $\hat{\pi}_w$ can be written as

$$\hat{\pi}_w = \frac{1}{n}\sum_{i=1}^{n}\lambda_{i0} = \frac{1}{k}\sum_{j=1}^{k} w_{j0}\hat{\pi}_j \tag{3.1}$$

where

$$\hat{\pi}_j = \left(\frac{1}{n}\sum_{i=1}^{n} \hat{y}_i(j)\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \frac{Z_i(j)-(1-P_j)}{2P_j-1}$$

$$= \frac{Q_j}{n}\frac{1}{2P_j-1} - \frac{1-P_j}{2P_j-1} \tag{3.2}$$

and Q_j is total number of "Yes" answers obtained from the jth deck of cards $j = 1,..,k$.

Odumande and Singh (2009) estimator can be written as

$$\hat{\pi}_{os} = \frac{1}{2} + \frac{(P_1+P_2-1)(n_{11}-n_{00}) + (P_1-P_2)(n_{10}-n_{01})}{2n\left\{(P_1+P_2-1)^2 + (P_1-P_2)^2\right\}}$$

$$= \frac{1}{2} + \frac{(P_1+P_2-1)(n_{1\bullet}-n_{10}-n_{00}) + (P_1-P_2)(n_{10}-n_{\bullet 1}+n_{11})}{2n\left\{(P_1+P_2-1)^2 + (P_1-P_2)^2\right\}}$$

($n_{1\bullet} = n_{11}+n_{10}$ and $n_{\bullet 1} = n_{11}+n_{01}$)

$$= \frac{1}{2} + \frac{(P_1+P_2-1)(n_{1\bullet}-n+n_{1\bullet}) + (P_1-P_2)(n_{1\bullet}-n_{\bullet 1})}{2n\left\{(P_1+P_2-1)^2 + (P_1-P_2)^2\right\}}$$

$$= \frac{1}{2} + \frac{n(2P_1-1)\{(2P_1-1)\hat{\pi}_1+(1-P_1)\} + n(2P_2-1)\{(2P_2-1)\hat{\pi}_2+(1-P_2)\} - (P_1+P_2-1)n}{2n\left\{(P_1+P_2-1)^2 + (P_1-P_2)^2\right\}}$$

$$= \frac{(2P_1-1)^2\hat{\pi}_1 + (2P_1-1)^2\hat{\pi}_2}{\left\{(2P_1-1)^2 + (2P_2-1)^2\right\}} \tag{3.3}$$

Further noting,

$$E(\hat{\pi}_j) = E_p\left[E_R(\hat{\pi}_j)\right]$$

$$= E_p\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)$$

$$= \pi,$$

$$V(\hat{\pi}_j) = E_p\left[V_R(\hat{\pi}_j)\right] + V_p\left[E_R(\hat{\pi}_j)\right]$$

$$= E_p\left(\frac{1}{n^2}\sum_{i=1}^{n}\phi_j\right) + V_p\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)$$

$$= \frac{\pi(1-\pi)}{n} + \frac{\phi_j}{n}$$

and

$$\text{Cov}(\hat{\pi}_j, \hat{\pi}_k) = E_p\left[C_R(\hat{\pi}_j, \hat{\pi}_k)\right] + C_p\left[E_R(\hat{\pi}_j), E_K(\hat{\pi}_k)\right]$$

$$= V_p\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)$$

$$= \frac{\pi(1-\pi)}{n},$$

We find the variance of $\hat{\pi}_{os}$ as

$$V(\hat{\pi}_{os}) = \frac{(2P_1-1)^4 V(\hat{\pi}_1) + (2P_2-1)^4 V(\hat{\pi}_2) + 2(2P_1-1)^2(2P_2-1)^2\pi(1-\pi)/n}{\left\{(2P_1-1)^2 + (2P_2-1)^2\right\}^2}$$

$$= \frac{\pi(1-\pi)}{n} + \frac{(2P_1-1)^4\phi_1 + (2P_2-1)^4\phi_2}{n\left\{(2P_1-1)^2 + (2P_2-1)^2\right\}^2} \tag{3.4}$$

Finally from (2.8) with $k = 2$ and (3.4) we find

$$V(\hat{\pi}_{os}) - V(\hat{\pi}_w) = \frac{(2P_1-1)^4\phi_1 + (2P_2-1)^4\phi_2}{n\left[(2P_1-1)^2 + (2P_2-1)^2\right]^2} - \frac{\phi_1\phi_2}{n(\phi_1+\phi_2)}$$

$$= \frac{1}{n}\left[\frac{(2P_1-1)^2\phi_1 - (2P_2-1)^2\phi_2}{(2P_1-1)^2 + (2P_2-1)^2}\right]^2 \frac{1}{\phi_1+\phi_2}$$

$$= \frac{1}{n}\left[\frac{P_1(1-P_1) - P_2(1-P_2)}{(2P_1-1)^2 + (2P_2-1)^2}\right]^2 \frac{1}{\phi_1+\phi_2} \geq 0 \tag{3.5}$$

The equality is attained if $P_1 = P_2$ or $P_1 = 1 - P_2$

4. CONCLUSION

Odumande and Singh (2009) proposed RR technique involving two decks of cards for estimating π, the proportion of individuals belonging to a certain sensitive group. Several researchers including Singh and Sedory (2011, 2012), Arnab *et al.* (2012), Lee *et al.* (2016), and Arnab *et al.* (2016) provided alternative estimators for Odumande and Singh (2009)'s RR technique and compared performances of their proposed estimators with Odumande and Singh (2009)'s estimator numerically but no theoretical meaningful conclusion was obtained. In this paper an extension of two decks to k-decks has been proposed. The proposed estimator is found to be more efficient than the original Odumande and Singh (2009) estimator for the special case $k=2$. An unbiased estimator of the variance of the newly proposed estimator has been proposed which is simple and always non-negative.

REFERENCES

Arnab, R. (2004). Optional randomized response techniques for complex survey designs. *Biom. J.*, **46**, 114-124.

Arnab, R. (2017). *Survey Sampling Theory and Applications*. Academic Press, U.K.

Arnab, R. Singh, S. and North, D. (2012). Use of two deck of cards in randomized response techniques for complex survey designs. *Comm. Statist.-Theory Methods*, **41**, 3198-3210.

Arnab, R. and Mothupi, T. (2015). Randomized response techniques: A case study of the risky behaviors' of students of a certain University. *Model Assist. Statist. Appl.*, **10**, 421-430.

Arnab, R. and Shangodoyoin, D.K. (2015). Randomized response techniques using maximum likelihood estimator. *Comm. Statist.–Theory Methods*, **44**, 3340-3349.

Arnab, R., Shangodoyin, D.K. and Dwivedi, V.K. (2016). Generalization of randomized response techniques for complex survey designs. *J. Ind. Soc. Agril. Statist.*, **70**, 131-138.

Chaudhuri, A. and Mukherjee, R. (1988). *Randomized Response: Theory and Techniques*. Marcel Dekker, New York.

Franklin, L.A. (1989). A comparison of estimators for randomized response sampling with continuous distribution from dichotomous populations. *Comm. Statist.-Theory Methods*, **18(2)**, 489-505.

Greenberg, B.D., Abul-Ela, A.L.A., Simmons, W.R., and Horvitz, D.G. (1969). The unrelated question randomized response model: Theoretical framework. *J. Amer. Statist. Assoc.*, **64**, 520-539.

Horvitz, D.G., Shah, B.V. and Simmons, W.R. (1967). The unrelated question randomized response model. *Proceedings of Social Statistics Section, Amer. Statist. Assoc.*, 65-72.

Kim, J.I. (1978). Randomized response technique for surveying human populations. Ph.D. Dissertation, Temple University, Philadelphia, USA.

Lee, C.S., Su, S.C., Mondragon, K., Salinas, V.I., Zamora, M.L., Sedory, S.A. and Singh, S. (2016). Comparison of Cramer–Rao lower bounds of variances for at least equal protection of respondents. *Statist. Neerlandica*, **70**, 80-89.

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

Odumade, O. and Singh, S. (2009). Efficient use of two decks of cards in randomized response sampling. *Comm. Statist.-Theory Methods*, **38**, 439-446.

Singh, S. (2003). *Advanced Sampling Theory with Applications*. Kluwer Academic Publishers, Netherlands.

Singh, S. and Sedory, S.A. (2011). Cramer-Rao lower bound of variance in randomized response sampling. *Sociol. Methods Res.*, **40(3)**, 536-546.

Singh, S. and Sedory, S.A. (2012). A true simulation study of three estimators at equal protection of respondents in randomized response sampling. *Statist. Neerlandica*, **66**, 442-451.

Warner, S.L. (1965). Randomize response: a survey technique for eliminating evasive answer bias. *J. Amer. Statist. Assoc.*, **60**, 63-69.