Available online at www.isas.org.in/jisas
Journal of the Indian Society of Agricultural Statistics 72(2) 2018 141–145

An Improved Randomized Response Technique

Raghunath Arnab
University of Botswana, Botswana and University of KwaZulu-Natal, South Africa

Received 18 December 2017; Revised 29 May 2018; Accepted 02 June 2018

SUMMARY

Warner (1965) introduced the randomized response (RR) techniques for the collection of data relating to sensitive characteristics. Kuk (1990) proposed a modified RR technique which is more efficient than Warner's (1965) RR technique. In this paper an alternative RR technique is proposed which is more efficient than the Kuk's (1990) RR technique.

Keywords: Inclusion probabilities, Randomized response, Sampling design.

1. INTRODUCTION

In surveys related to sensitive issues such as domestic violence, induced abortions and use of the illegal drugs, respondents often give untrue responses because of social stigma and fear. To improve cooperation from respondents and get more truthful answers from them, Warner (1965) proposed the RR technique where respondents provide indirect responses. Kuk (1990) proposed an alternative RR technique which is more efficient than Warner's technique. Both RR techniques are based on simple random sampling with replacement (SRSWR) sampling scheme only. In this paper an alternative RR technique is proposed for estimating the population proportion π of a sensitive characteristic. The proposed RR technique is more efficient than Kuk's RR technique. The method can be used for any sampling design and yields elegant expressions of the unbiased estimator of π, its variance and unbiased estimator of the variance of the estimator.

1.1 Warner's RR Technique

In this technique, respondents draw a card at random from a pack of cards containing two types of cards with statements "I belong to the sensitive group A" and "I belong to the non-sensitive group $\overline{A}$" with proportions $P(\neq 1/2)$ and $1-P$ respectively. The respondents are asked to report "Yes" if the statement written on the card drawn matches theirstatuses. Otherwise, if the statement does not match with their statuses, the respondentsshould answer "No". The whole experiment is performed in the absence of the investigator. Thus the confidentiality of the response is maintained.Warner considered the situation where a sample *s* of size *n* is selected from a population by the simple random sampling with replacement (SRSWR) method. Let λ_w be the proportion of "yes" answers obtained from the respondents selected in the sample *s*. Warner derived the following results:

(i) $\hat{\pi}_w = \dfrac{\lambda_w - (1-P)}{2P-1}$ is an unbiased estimator of π

(ii) The variance of $\hat{\pi}_w$ is

$$\mathrm{Var}(\hat{\pi}_w) = \frac{\phi_w(1-\phi_w)}{n(2P-1)^2} = \frac{\pi(1-\pi)}{n} + \frac{P(1-P)}{n(2P-1)^2}$$

where π is the proportion of individuals belonging to the sensitive group A and $\phi_w = \pi P + (1-\pi)(1-P)$ = the probability of obtaining "Yes" answer from a respondent selected at random.

E-mail address: arnabr@mopipi.ub.bw

1.2 Kuk's RR Technique

The RR technique consists of two packs of cards: pack-1 and pack-2. Each pack consists of two types of cards such as black and red. The proportions of black cards for pack-1 and pack-2 are P and $T(\neq P)$ respectively. The respondents selected in the sample s are asked to draw $k(\geq 2)$ cards at random with replacement from pack-1 if they belong to the sensitive group A. Otherwise, if a respondent belongs to the non-sensitive group $\bar{A}$, he/she has to select k cards at random with replacement from pack-2. The respondents are asked to report the number of black cards selected as their randomized response. The whole experiment is performed in the absence of the interviewer,so, the confidentiality of the respondent is maintained. Clearly the Kuk's RR technique reduces to Warner's technique for $k=1$ and $T=1-Pz$.

Let a sample s of size n be selected from a population by SRSWR method and λ_k be the proportion of black cards selected from a total of nk cards drawn. Kuk derived the following result:

(i) $\hat{\pi}_{\mathrm{k}}=\dfrac{\lambda_{\mathrm{k}}-T}{P-T}$ is an unbiased estimator of π

(ii) The variance of $\hat{\pi}_{\mathrm{k}}$ is

$$\mathrm{Var}\left(\hat{\pi}_{\mathrm{k}}\right)=\frac{\phi_{\mathrm{k}}\left(1-\phi_{\mathrm{k}}\right)}{nk\left(P-T\right)^2}+\frac{\pi(1-\pi)}{n}\left(1-\frac{1}{k}\right)$$

$$=\frac{\pi\left(1-\pi\right)}{n}+\frac{1}{kn}\left[\frac{\left(P-T\right)\left(1-P-T\right)\pi+T\left(1-T\right)}{\left(P-T\right)^2}\right]$$

where $\phi_{\mathrm{k}}=\pi P+(1-\pi)T$

(iii) An unbiased estimator of $\mathrm{Var}\left(\hat{\pi}_{\mathrm{k}}\right)$ is

$$\hat{\mathrm{V}}\mathrm{ar}\left(\hat{\pi}_{\mathrm{k}}\right)=\frac{1}{n^2\left(P-T\right)^2}\Sigma_{i\in s}\frac{1}{k(k-1)}\left[\sum_{i=1}^{k}R_{ij}-\left(\sum_{i=1}^{k}R_{ij}\right)^2/k\right]$$

where $\Sigma_{i\in s}$ denotes sum of the units in s with repetition and $R_{ij}=1(0)$ if ith respondent draws a black (red) at $j(=1,..,k)$ th draw.

The expression $\hat{\mathrm{V}}\mathrm{ar}\left(\hat{\pi}_{\mathrm{k}}\right)$ was obtained by Chaudhuri (2011).

2. PROPOSED RR TECHNIQUE

The proposed RR technique comprises two sets of cards: Set-1 and Set-2. Each of the sets, Set-1 and Set-2, consists of k packs of cards. The $j(=1,..,k)$ th pack of Set-1 consists of two types of cards: black and red with proportion P_j and $1-P_j$ and the j th pack of Set-2 consists of black and red cards with proportions $T_j\left(\neq P_j\right)$ and $1-T_j$ respectively. Respondents selected in the sample were directed to draw one card from each packs of the Set-1 if they belong to the sensitive group A; otherwise if the respondents belonged to the non-sensitive group $\bar{A}$, they were to draw one card from each of the k packs from the Set-2 independently. The whole experiment is performed in the absence of the interviewer, so the confidentiality of the respondents is maintained. The proposed RR technique reduces to Kuk's RR technique if $P_j=P$ and $T_j=T$ for $j=1,..,k\left(\geq 2\right)$.

2.1 Sampling design and methods of Estimation

Consider a finite population $U=\left(1,..,i,..,N\right)$ of N identifiable units (respondents). Let a sample s of size n be selected from the population U with probability $p(s)$ using a sampling design p. Let the inclusion probabilities of the i-th, and i-th and $j(\neq i)$-th units be denoted by $\pi_i(>0)$ and π_{ij} respectively. The objective is to estimate π, the proportion of individuals belonging to a certain sensitive group A. Let $y_i=1$, if the ith unit belongs to the group A and $y_i=0$ if $i\in\bar{A}$ where $\bar{A}$ is the complementary of A. Then the population proportion of the persons belongingto the sensitive group A is

$$\pi=\sum_{i\in U}y_i/N \tag{2.1}$$

We define $z_i(j)=1$ if the ith respondent selected in the sample s draws a black card from the j-th pack and $z_i(j)=0$ if the drawn card is red; $j=1,..,k$. Hence,

$$\mathrm{Prob}\left(z_i(j)=1\right)=\begin{cases}P_j & \text{if the ith respondent } \in A\\ T_j & \text{if the ith respondent } \notin A\end{cases} \tag{2.2}$$

i.e. $\mathrm{Prob}\left(z_i(j)=1\right)=y_iP_j+(1-y_i)T_j$

$$=y_i\left(P_j-T_j\right)+T_j \tag{2.3}$$

Let $E_R(V_R)$ denotes expectation (variance) with respect to the RR model. Then,

$$E_R(z_i(j)) = y_i(P_j - T_j) + T_j \tag{2.4}$$

and

$$V_R(z_i(j)) = y_iP_j + (1-y_i)T_j - \{y_iP_j + (1-y_i)T_j\}^2$$
$$= y_iP_j + (1-y_i)T_j - y_iP_j^2 - (1-y_i)T_j^2$$

(noting $y_i^2 = y_i$ and $(1-y_i)^2 = 1-y_i$ as $y_i = 0,1$)

$$= y_iP_j(1-P_j) + (1-y_i)T_j(1-T_j)$$

$$= y_i(P_j - T_j)(1-P_j-T_j) + T_j(1-T_j) \tag{2.5}$$

The Eq. (2.4) yields

$$E_R(\bar{z}_i) = y_i(\bar{P} - \bar{T}) + \bar{T} \tag{2.6}$$

where $\bar{z}_i = \sum_{j=1}^{k} z_i(j)/k$, $\bar{T} = \sum_{j=1}^{k} T_j/k$ and $\bar{P} = \sum_{j=1}^{k} P_j/k$

From Eq. (2.6), one finds an unbiased estimator of y_i as

$$r_i = \frac{\bar{z}_i - \bar{T}}{\bar{P} - \bar{T}} \tag{2.7}$$

For a fixed effective size (n) sampling design with $\pi_i > 0$ for every $i \in U$, the Horvitz-Thompson estimator for the population π is $\hat{\pi}_{ht} = \frac{1}{N}\sum_{i\in s}\frac{r_i}{\pi_i}$. The properties of $\hat{\pi}_{ht} = \frac{1}{N}\sum_{i\in s}\frac{r_i}{\pi_i}$ are given the following theorem.

Theorem 2.1

(i) $\hat{\pi}_{ht} = \frac{1}{N}\sum_{i\in s}\frac{r_i}{\pi_i}$ is an unbiased estimator of π

(ii) $\mathrm{Var}(\hat{\pi}_{ht}) = \frac{1}{N^2}\left[\sum_{i\in A}\left(\frac{1}{\pi_i}-1\right) + \sum_{i\neq j\in A}\left(\frac{\pi_{ij}}{\pi_i\pi_j}-1\right) + \sum_{i\in U}\frac{\sigma_i^2}{\pi_i}\right]$

where $V_R(r_i) = \sigma_i^2 = \dfrac{y_i\sum_{j=1}^{k}(P_j - T_j)(1-P_j-T_j) + \sum_{j=1}^{k}T_j(1-T_j)}{k^2(\bar{P}-\bar{T})^2}$

(iii) $\hat{\mathrm{Var}}(\hat{\pi}_{ht}) = \frac{1}{N^2}\left[\sum_{i\in s}\frac{1}{\pi_i}\left(\frac{1}{\pi_i}-1\right)r_i^2 + \sum_{i\neq j\in s}\frac{1}{\pi_{ij}}\left(\frac{\pi_{ij}}{\pi_i\pi_j}-1\right)r_ir_j + \sum_{i\in s}\frac{r_i(r_i-1)}{\pi_i}\right]$

is an unbiased estimator of $\mathrm{Var}(\hat{\pi}_{ht})$.

Proof:

(i) $E(\hat{\pi}_{ht}) = \frac{1}{N}E_p\left[\sum_{i\in s}\frac{E_R(r_i)}{\pi_i}\right]$

$$= \frac{1}{N}E_p\left[\sum_{i\in s}\frac{y_i}{\pi_i}\right]z$$

$$= \frac{1}{N}\sum_{i\in U} y_i$$

$= \pi$ (noting $y_i = 1$ for $i \in A$ and $y_i = 0$ for $i \notin A$)

(ii) $\mathrm{Var}(\hat{\pi}_{ht}) = \frac{1}{N^2}\left[V_p\left\{\sum_{i\in s}\frac{E_R(r_i)}{\pi_i}\right\} + E_p\left\{\sum_{i\in s}\frac{V_R(r_i)}{\pi_i^2}\right\}\right]$

$$= \frac{1}{N^2}\left[V_p\left\{\sum_{i\in s}\frac{y_i}{\pi_i}\right\} + E_p\left\{\sum_{i\in s}\frac{\sigma_i^2}{\pi_i^2}\right\}\right]$$

$$= \frac{1}{N^2}\left[\sum_{i\in U}\left(\frac{1}{\pi_i}-1\right)y_i^2 + \sum_{i\neq j\in U}\sum\left(\frac{\pi_{ij}}{\pi_i\pi_j}-1\right)y_iy_j + \sum_{i\in U}\frac{\sigma_i^2}{\pi_i}\right]$$

$$= \frac{1}{N^2}\left[\sum_{i\in A}\left(\frac{1}{\pi_i}-1\right) + \sum_{i\neq j\in A}\sum\left(\frac{\pi_{ij}}{\pi_i\pi_j}-1\right) + \sum_{i\in U}\frac{\sigma_i^2}{\pi_i}\right]$$

(iii) $E\left[\hat{\mathrm{Var}}(\hat{\pi}_{ht})\right] = \frac{1}{N^2}E_p\left[\sum_{i\in s}\frac{1}{\pi_i}\left(\frac{1}{\pi_i}-1\right)E_R(r_i^2) + \sum_{i\neq j\in s}\sum\frac{1}{\pi_{ij}}\left(\frac{\pi_{ij}}{\pi_i\pi_j}-1\right)E_R(r_ir_j) + \sum_{i\in s}\frac{E_R\{r_i(r_i-1)\}}{\pi_i}\right]$

$$= \frac{1}{N^2}E_p\left[\sum_{i\in s}\frac{1}{\pi_i}\left(\frac{1}{\pi_i}-1\right)(y_i^2+\sigma_i^2) + \sum_{i\neq j\in s}\sum\frac{1}{\pi_{ij}}\left(\frac{\pi_{ij}}{\pi_i\pi_j}-1\right)y_iy_j + \sum_{i\in s}\frac{\sigma_i^2+y_i^2-y_i}{\pi_i}\right]$$

$$= \frac{1}{N^2}\left[\sum_{i\in U}\left(\frac{1}{\pi_i}-1\right)(y_i^2+\sigma_i^2) + \sum_{i\neq j\in U}\sum\left(\frac{\pi_{ij}}{\pi_i\pi_j}-1\right)y_iy_j + \sum_{i\in U}\sigma_i^2\right]$$

$$= \mathrm{Var}(\hat{\pi}_{ht})$$

For a simple random sampling without replacement (SRSWOR) sampling $\pi_i = n/N$ and $\pi_{ij} = n(n-1)/\{N(N-1)\}$. Hence by substituting $\pi_i = n/N$ and $\pi_{ij} = n(n-1)/\{N(N-1)\}$ in the above Theorem 2.1, we get

Theorem 2.2

For SRSWOR sampling,

(i) $\bar{r}_{wor} = \frac{1}{n}\sum_{i\in s} r_i$ is an unbiased estimator of π

(ii) $\mathrm{Var}(\bar{r}_{wor}) = \frac{(1-f)}{n}\frac{N}{N-1}\pi(1-\pi) + \frac{\pi\sum_{i=1}^{k} P_j(1-P_j) + (1-\pi)\sum_{i=1}^{k} T_j(1-T_j)}{n(\bar{P}-\bar{T})^2}$

where $f = n/N$

(iii) $\hat{\mathrm{V}}\mathrm{ar}(\bar{r}_{wor}) = \frac{(1-f)}{n}\frac{1}{n-1}\sum_{i\in s}(r_i - \bar{r}_{wor})^2 + \frac{1}{Nn}\sum_{i\in s} r_i(r_i-1)$

For simple random sampling with replacement (SRSWR) sampling design r_i's are independently and identically distributed, hence we have the following results:

Theorem 2.3

For SRSWR sampling,

(i) $\bar{r}_{wr} = \frac{1}{n}\Sigma_{i\in s} r_i$ is an unbiased estimator of π

where $\Sigma_{i\in s}$ denotes sum over the units in s including repetition.

(ii) $\mathrm{Var}(\bar{r}_{wr}) = \frac{\pi(1-\pi)}{n} + \frac{\pi\sum_{i=1}^{k} P_j(1-P_j) + (1-\pi)\sum_{i=1}^{k} T_j(1-T_j)}{n(\bar{P}-\bar{T})^2}$

(iii) $\hat{\mathrm{V}}\mathrm{ar}(\bar{r}_{wr}) = \frac{1}{n-1}\Sigma_{i\in s}(r_i - \bar{r}_{wr})^2$

Remark 2.1.

For $P_j = \bar{P}$ and $T_j = \bar{T}$, $\mathrm{Var}(\bar{r}_{wr})$ becomes equal to $\mathrm{Var}(\hat{\pi}_w)$.

3. COMPARISON WITH KUK'S RR TECHNIQUE

For Kuk's RR technique $P_i = P$ and $T_i = T$ for $i = 1,..,N$. Substituting $P_i = P$ and $T_i = T$ in (2.2) and (2.5), we get $E_R(z_i(j)) = y_i(P-T)+T$ and $V_R(z_i(j)) == y_i(P-T)(1-P-T)+T(1-T)$. In this case an unbiased estimator of y_i becomes

$$r_i(k) = \frac{\bar{z}_i - T}{P-T} \tag{3.1}$$

Now using Theorem (2.1), we note that $\hat{\pi}_{ht}(k) = \frac{1}{N}\sum_{i\in s}\frac{r_i(k)}{\pi_i}$ is an unbiased estimator of π with variance

$$\mathrm{Var}(\hat{\pi}_{ht}(k)) = \frac{1}{N^2}\left[\sum_{i\in A}\left(\frac{1}{\pi_i}-1\right) + \sum_{i\neq}\sum_{j\in A}\left(\frac{\pi_{ij}}{\pi_i\pi_j}-1\right) + \sum_{i\in U}\frac{\sigma_i^2(k)}{\pi_i}\right] \tag{3.2}$$

where $\sigma_i^2(k) = \frac{y_i(P-T)(1-P-T)+T(1-T)}{k(P-T)^2}$

Theorem 3.1.

The proposed RR technique is more efficient than Kuk's RR technique if $\bar{P} = P$ and $\bar{T} = T$.

Proof.

Now noting

$$\sigma_i^2(k) - \sigma_i^2 = \frac{y_i(\bar{P}-\bar{T})(1-\bar{P}-\bar{T})+\bar{T}(1-\bar{T})}{k(\bar{P}-\bar{T})^2} - \frac{y_i\sum_{j=1}^{k}(P_j-T_j)(1-P_j-T_j) + \sum_{j=1}^{k} T_j(1-T_j)}{k^2(\bar{P}-\bar{T})^2}$$

$$\frac{y_i\left[\frac{1}{k}\sum_{j=1}^{k}(P_j-T_j) - (\bar{P}-\bar{T})\right] + \left[\frac{1}{k}\sum_{j=1}^{k} T_j^2 - \bar{T}^2\right]}{k(\bar{P}\ \ \bar{T})}$$

≥ 0,

we find

$$\mathrm{Var}(\hat{\pi}_{ht}(k)) - \mathrm{Var}(\hat{\pi}_{ht}) = \frac{1}{N^2}\left[\sum_{i\in U}\frac{\sigma^2(k)-\sigma^2}{\pi_i}\right] \tag{3.3}$$

≥ 0

The relative percentage efficiency of the proposed strategy with respect to the Kuk's strategy is given by

$$E = \frac{\mathrm{Var}(\hat{\pi}_{ht}(k))}{\mathrm{Var}(\hat{\pi}_{ht})}\times 100 \tag{3.4}$$

For the SRSWR sampling the expression of E is

$$E^* = \frac{\mathrm{Var}\left(\hat{\pi}_{\mathrm{k}}\right)}{\mathrm{Var}\left(\bar{r}_{wr}\right)} \times 100 \qquad (3.5)$$

It should be noted that the relative efficiency E^* is symmetric over permutation of the coordinates of $\mathbf{P} = (P_1, P_2, P_3, P_4)$ and $\mathbf{T} = (T_1, T_2, T_3, T_4)$. The following Table 3.1 gives the relative efficiency E^* of the proposed RR technique for different values of **P**, **T** and π with respect to Kuk's RR technique with $P = \bar{P} = \sum P_i / 4$ and $T = \bar{T} = \sum T_i / 4$. The results show that the proposed strategy can bring substantial gain in efficiency upto 68.2% over Kuk's strategy.

Table 3.1. Relative efficiency (E^*) of the proposed strategy overKuk'sstrategy under SRSWR sampling Scheme

π	P = (0.1, 0.2, 0.3, 0.4) T = (0.1, 0.1, 0.2, 0.2) $\bar{P}$ = 0.25; $\bar{T}$ = 0.15	P = (0.1, 0.2, 0.3, 0.4) T = (0.1, 0.7, 0.2, 0.7) $\bar{P}$ = 0.25; $\bar{T}$ = 0.425	P = (0.1, 0.2, 0.3, 0.4) t = (0.1, 0.3, 0.2, 0.7) $\bar{P}$ = 0.25; $\bar{T}$ 0.325	P = (0.1, 0.2, 0.3, 0.4) T = (0.2, 0.3, 0.2, 0.8) $\bar{P}$ = 0.25; $\bar{T}$ = 0.375
0.05	102.3	142.4	129.5	133.8
0.10	102.6	139.3	128.2	131.9
0.20	103.2	133.9	125.5	128.4
0.30	103.7	129.4	123.0	125.3
0.40	104.2	125.6	120.5	122.3
0.45	104.4	123.8	119.4	121.0
	P = (0.1, 0.4, 0.7, 0.8) T = (0.2, 0.3, 0.2, 0.8) $\bar{P}$ = 0.5; $\bar{T}$ = 0.375	P = (0.2, 0.8, 0.2, 0.4) T = (0.2, 0.3, 0.2, 0.8) $\bar{P}$ = 0.4; $\bar{T}$ = 0.375	P = (0.1, 0.8, 0.2, 0.6) T = (0.2, 0.6, 0.3, 0.9) $\bar{P}$ = 0.425; $\bar{T}$ = 0.5	P = (0.1, 0.8, 0.2, 0.6) T = (0.4, 0.8, 0.9, 0.1) $\bar{P}$ = 0.425; $\bar{T}$ = 0.55
0.05	135.6	135.7	142.9	168.2
0.10	135.4	135.6	143.1	165.9
0.20	135.2	135.3	143.4	162.1
0.30	135.3	135.0	143.8	159.0
0.40	135.6	134.7	144.3	156.4
0.45	135.8	134.6	144.6	155.4

4. CONCLUSION

In Kuk's RR technique, the respondent belonging to the sensitive group $A\left(\bar{A}\right)$ performs k independent Bernoulli trials with the constant probability of success $P(T)$. Kuk's technique can be improved if the respondent belonging to the sensitive group $A\left(\bar{A}\right)$ performs k independent Bernoulli trials with unequal success probabilities $P_j\left(T_j\right)$, $j = 1,..,k$ keeping the average $\frac{1}{k}\sum_{j=1}^{k} P_j = P$ and $\frac{1}{k}\sum_{j=1}^{k} T_j = T$ fixed.

ACKNOWLEDGEMENTS

The author is grateful to the referee for his constructive suggestions that lead to improvement of the final manuscript.

REFERENCES

Kuk, A. (1990). Asking sensitive question directly. Biometrika, **77**, 436-438.

Chaudhuri, A. (2011). Randomized response and indirect questioning techniques in surveys.CRC Press, USA.

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