J. Ind. Soc. Agril. Statist.
61(1), 2007 : 77-79

A Note on Unrelated Question Randomized Response Model

Raghunath Arnab and Sarjinder Singh[1]
University of Botswana, Botswana

(Received : July, 2006)

SUMMARY

It is shown that the derivations of most of the results in the randomized response technique proposed by Singh *et al.* (2000) are incorrect. Corrections of these results are given.

Key words: Randomized response, Sampling designs, Relative efficiency.

1. INTRODUCTION

Warner (1965) introduced an ingenious technique known as randomized response technique (RR) for estimating π_x, the proportion of population possessing certain stigmatized character x (say) by protecting the privacy of respondents and preventing the unacceptable rate of non-response. Since then Warner's (1965) technique has been modified by several researchers. A comprehensive review is available in Chaudhuri and Mukherjee (1988). Following the Moor's (1971) technique Singh *et al.* (2000) proposed two alternative RR techniques described as follows.

Singh *et al.* (2000) — Method 1: Two independent samples S_1 and S_2 were selected by simple random sampling without replacement (SRSWOR) method. Each respondent in the S_1 sample was asked to perform the randomized device R_1 while the respondents belonging to both the samples S_1 and S_2 were asked to perform randomized devise R_2 as described above. The respondents belonging to S_2 but not S_1 were directly asked whether or not they possess the neutral character y. The proposed estimator of π_x is given by

$$\hat{\pi}_p = w\hat{\pi}_1 + (1-w)\hat{\pi}_2 \qquad (1)$$

where $\hat{\pi}_1 = \dfrac{\hat{\theta}_1 - (1-p_1)\hat{\pi}_{2y}}{p_1}$, $\hat{\pi}_2 = \dfrac{\hat{\theta}_2 - (1-p_2)\hat{\pi}_{2y}}{p_2}$

$\hat{\theta}_i$ = proportion of "yes" answers in S_i, I = 1, 2

$\hat{\pi}_{2y}$ = proportion of the respondents belong to sample S_2 but not belong to S_1 possess the character y and W is a suitable weight.

Method 2: At first, an initially sample $\tilde{s}$ of size n was selected from the population U by SRSWOR method. The sample $\tilde{s}$ was divided at random into two sub samples $\tilde{s}_1$ and $\tilde{s}_2$ of sizes n_1 (to be determined appropriately) and n_2 (= n − n_1) respectively. Respondents belonging to the first sub-sample $\tilde{s}_1$, were asked to perform randomized device R_1 while respondents belonging to the sub-sample $\tilde{s}_2$ were asked directly to answer the question (ii) relating to possession of the neutral character y. The proposed estimator for π_x is given by

$$\tilde{\pi}_x = \frac{\tilde{\theta}_1 - (1-p_1)\tilde{\pi}_{2y}}{p_1}$$

where $\tilde{\theta}_1$ and $\tilde{\pi}_{2y}$ are the proportions of yes answers in the first and second samples.

2. CORRECTIONS OF SINGH *et al.* (2000) RESULTS

In this section, we will show that the following results obtained by Singh *et al.* (2000) are incorrect and we present corrections.

For the Method 1, let S_{21} be the sample of size n_{21} consisting of units belonging to both the samples

[1] St. Cloud State University, USA

S_1 and S_2, and S_{22} is a sample of size n_{22} ($= n_2 - n_{21}$) belonging to S_2 but disjoined to S_1 i.e $S_2 = S_{21} \cup S_{22}$. Let $z_i(z'_i)$ be the RR obtained from the ith unit if it belongs to $S_1(S_{21})$. Let $x_i = 1$ if ith unit possess the character x and $x_i = 0$ otherwise. Similarly, $y_i = 1$ if ith unit possess the neutral character y and $y_i = 0$ otherwise. Denoting $E_R(E_P)$ and $V_R(V_P)$ respectively as expectation and variance with respect to randomized response (sampling design) we note the following.

$$\pi_x = \sum_{i=1}^{N} x_i / N,\ \pi_y = \sum_{i=1}^{N} y_i / N$$

$$E_R(z_i) = p_1 x_i + (1 - p_1) y_i = w_i$$

$$E_R(z'_i) = p_2 x_i + (1 - p_2) y_i = \gamma_i$$

$$V_R(z_i) = w_i(1 - w_i) = \sigma_i^2$$

$$V_R(z'_i) = \gamma_i(1 - \gamma_i) = \sigma_i'^2$$

$$\hat{\theta}_1 = \frac{1}{n_1} \sum_{i \in S_1} z_i = \bar{z}(S_1)$$

$$\hat{\pi}_{2y} = \bar{y}(S_{22}) = \frac{1}{n_{22}} \sum_{i \in S_{22}} y_i$$

The incorrect results of Singh *et al.* (2000) paper are presented using notations of this paper as follows.

Result 1. (Lemma 3.3, page 247)

$$\text{Var}(\hat{\theta}_1) = \frac{\theta_1(1-\theta_1)}{n_1} = \frac{n_1 - 1}{n_1(N-1)} \pi_x(1 - \pi_x)$$

Result 2. (Lemma 3.4, page 247)

$$\text{Var}(\hat{\theta}_2) = \left(\theta_2(1-\theta_2) - \frac{\pi_x(1-\pi_x)}{N-1} \right) E(\frac{1}{n_{21}}) - \frac{\pi_x(1-\pi_x)}{N-1}$$

Result 3. (Lemma 3.5, page 248)

For uncorrelated x and y

$$\text{Cov}(\hat{\theta}_1, \hat{\pi}_{2y}) = \frac{N - n_1}{n_1(N-1)} (1 - p_1)\pi_y(1 - \pi_y)$$

Result 4. (Lemma 3.6, page 248)

For uncorrelated x and y

$$\text{Cov}(\hat{\theta}_2, \hat{\pi}_{2y}) = \frac{N - n_1}{n_1(N-1)} (1 - p_2)\pi_y(1 - \pi_y)$$

Result 5. (Lemma 3.7, page 249)

$$\text{Var}(\tilde{\theta}_1) = \left(\frac{\theta_1(1-\theta_1)}{n_1} - \frac{(n_1-1)\pi_x(1-\pi_x)}{n_1(N-1)} \right)$$

(for Method 2)

2.1 Corrections of the above Results

Result 1.

$$\text{Var}(\hat{\theta}_1) = \text{Var}[\bar{z}(S_1)] = E_p\left(V_R\left(\bar{z}(S_1)\right)\right) + V_p\left(E_R\left(\bar{z}(S_1)\right)\right)$$

$$= \frac{1}{n_1 N} \sum_{i=1}^{N} \sigma_i^2 + V_p\left(\frac{1}{n} \sum_{i \in s_1} w_i \right)$$

$$= \frac{\theta_1(1-\theta_1)}{n_1} - \frac{n_1 - 1}{n_1(N-1)} (p_1^2 \pi_x(1 - \pi_x) + (1-p_1)^2 \pi_y(1-\pi_y) + 2_{p1}(1-p_1)\pi_{xy}^*)$$

where $\pi_{xy}^* = \pi_{xy} - \pi_x \pi_y$, $\pi_{xy} = \sum_{i=1}^{N} x_i y_i / N$

In case x and y are independent $\pi_{xy}^* = 0$ and we get

$$\text{Var}(\hat{\theta}_1) = \frac{\theta_1(1-\theta_1)}{n_1} - \frac{n_1 - 1}{n_1(N-1)} (p_1^2 \pi_x(1-\pi_x) + (1-p_1)^2 \pi_y(1-\pi_y))$$

which is quite different from Result 1 obtained by Singh *et al.* (2000). It should be noted that the expression $\text{Var}(\hat{\theta}_1)$, obtained by Singh *et al.* (2000), is independent of π_y which is incorrect and can be checked from the fact that $z_i = y_i = w_i$ for $p_1 = 0$.

Result 2.

$$\text{Var}(\hat{\theta}_2) = \text{Var}\left(\bar{z}'(S_{21})\right) = E_p\left(V_R\left(\bar{z}'(S_{21})\right)\right) + V_p\left(E_R\left(\bar{z}'(S_{21})\right)\right)$$

Now writing $\underset{n_{21}}{E}$ as the unconditional expectation over n_{21}

$$\sigma_i'^2 = V_R(z'_i) = p_2 x_i + (1-p_2) y_i - \gamma_i^2$$

$$\gamma_i = E_R(z'_i) = p_2 x_i + (1-p_2) y_i$$

$$E_p\left(V_R\left(\bar{z}'(S_{21})\right)\right)=E_p\frac{1}{n_{21}^2}\sum_{i\in S_{21}}\sigma_i'^2$$

$$=\underset{n_{21}}{E}\left(E_p\frac{1}{n_{21}^2}\sum_{i\in S_{21}}\sigma_i'^2\mid n_{21}\right)$$

$$=\frac{1}{N}\sum_{i\in U}\sigma_i'^2E\left(\frac{1}{n_{21}}\right) \qquad (2)$$

and $V_p\left(E_R\left(\bar{z}'(S_{21})\right)\right)=V_p\left(\bar{\gamma}(S_{21})\right)$

$$=\underset{n_{21}}{E}\left(V_p\left(\bar{\gamma}(S_{21})\mid n_{21}\right)\right)+\underset{n_{21}}{V}\left(E_p\left(\bar{\gamma}(S_{21})\mid n_{21}\right)\right)$$

$$\doteq\underset{n_{21}}{E}\left(\frac{1}{n_{21}}-\frac{1}{N}\right)S_\gamma^2$$

$$=\left(E\left(\frac{1}{n_{21}}\right)-\frac{1}{N}\right)\frac{N}{(N-1)}\Pi_{xy}(p_2) \qquad (3)$$

where

$$\bar{\gamma}(S_1)=\sum_{i\in S_{21}}\gamma_i/n_{21}$$

$$(N-1)S_\gamma^2=\sum_{i\in U}(\gamma_i-\bar{\gamma})^2$$

$$\bar{\gamma}=\sum_{i\in U}\gamma_i/N$$

and

$$\Pi_{xy}(p)=\left(p^2\pi_x(1-\pi_x)+(1-p)^2\pi_y(1-\pi_y)+2p(1-p)\pi_{xy}^*\right)$$

$$p=p_1,p_2 \qquad (4)$$

From (2) and (3), we get

$$Var(\hat{\theta}_2)=\frac{1}{N}\sum_{i\in U}\sigma_i'^2E\left(\frac{1}{n_{21}}\right)+\left(E\left(\frac{1}{n_{21}}\right)-\frac{1}{N}\right)\frac{N}{(N-1)}\Pi_{xy}(p_2)$$

$$=\left(\theta_2(1-\theta_2)+\frac{\Pi_{xy}(p_2)}{N-1}\right)E\left(\frac{1}{n_{21}}\right)-\frac{\Pi_{xy}(p_2)}{N-1}$$

Result 3.

$$Cov(\hat{\theta}_1,\hat{\pi}_{2y})=Cov(\bar{z}(S_1),\bar{y}(S_{22}))$$

$$=\underset{n_{21}}{E}\left(Cov\left(\bar{w}(S_1),\bar{y}(S_{22})\mid n_{21}\right)\right)+\underset{n_{21}}{Cov}\left(E\left(\bar{w}(S_1\mid n_{21})\right),E\left(\bar{y}(S_{22}\mid n_{21})\right)\right)$$

$$=\underset{n_{21}}{E}\left(Cov[\bar{w}(S_1),\bar{y}(U-S_1)\mid n_{21}\right)$$

(since $\underset{n_{21}}{Cov}\left(E\left(\bar{w}(S_1\mid n_{21})\right),E\left(\bar{y}(S_{22}\mid n_{21})\right)\right)=0$)

$$=-\frac{n_1}{N-n_1}\underset{n_{21}}{E}\,Cov\left(\bar{w}(S_1),\bar{y}(S_1)\right)$$

$$=-\frac{1}{N}S_{wy}-\frac{1}{N-1}\left(p_1\pi_{xy}^*+(1-p_1)\pi_y(1-\pi_y)\right)$$

If x and y are uncorrelated, we get

$$Cov(\hat{\theta}_1,\hat{\pi}_{2y})=-\frac{(1-p_1)\pi_y(1-\pi_y)}{N-1} \qquad (5)$$

Result 4.

$Cov(\hat{\theta}_2,\hat{\pi}_{2y})=-\dfrac{(1-p_2)\pi_y(1-\pi_y)}{N-1}$ when x and y are independent.

(Proof of the Result 4 follows from (5))

Result 5.

It can be easily checked that

$$Var(\tilde{\theta}_1)=\frac{\theta_1(1-\theta_1)}{n_1}-\frac{n_1-1}{n_1(N-1)}\Pi_{xy}(p_1) \qquad (6)$$

For uncorrelated x and y, (6) reduces to

$$Var(\tilde{\theta}_1)=\frac{\theta_1(1-\theta_1)}{n_1}-\frac{n_1-1}{n_1(N-1)}\left(p_1^2\pi_x(1-\pi_x)+(1-p_1)^2\pi_y(1-\pi_y)\right) \qquad (7)$$

ACKNOWLEDGEMENTS

The authors are thankful to the referee for making valuable suggestions that led to the improvement of the paper.

REFERENCES

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

Moors, J.J.A. (1971). Optimization of the unrelated question randomized response models. *J. Amer. Statist. Assoc.,* **66**, 627-629.

Singh, S., Singh, R., Mangat, N.S. (2000). Some alternative strategies to Moor's model in randomized response sampling. *J. Statist. Plann. Inf.,* **83**, 243-255.

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