Statistics and Applications {ISSN 2452-7395 (online)}
Volume 15 Nos. 1&2, 2017 (New Series), pp 147-160

A Stratified Randomized Response Model for Sensitive Characteristics using Geometric Distribution of Order K

Housila P. Singh[1] , Swarangi M. Gorey[1] and Tanveer A. Tarray[2]
[1]*School of Studies in Statistics, Vikram University, Ujjain.*
[2]*Department of Computer Science and Engineering, Islamic University of Science and Technology-Awantipora-Kashmir.*

Received: December 23, 2016; Revised: January 10, 2017; Accepted: February 04, 2017

Abstract

Taking the clue from the work of Hussain et al.'s (2016) we have suggested a new stratified randomized response model. The properties of the suggested stratified randomized response model have been studied under proportional and "Neyman" allocations. Numerical illustrations are given in support of the present study.

Keywords: Randomized response model, Estimation of proportion, Stratified random sampling, Variance.

1. Introduction

In both clinical and community setting, survey research concerning sensitive questions such as gambling, alcoholism, sexual behavior, drug taking, tax evasion, illegal income and else, direct techniques for collecting information may induce interviewed people to refuse answering or to give untruthful or misleading responses. To reduce non respondents rates and biased responses arising from sensitive, embarrassing, threatening, or even incriminating questions, some special statistical techniques may be employed to ensure interviewee anonymity or, at least, a higher degree of confidence. Such techniques, known as randomized response methods, use a randomization device, such as a die or a deck of cards, rather than a direct response to collect reliable information on sensitive issues. Depending on the result produced by the randomization device, the interviewee gives an answer concerning his/her true status. Since the interviewer is unaware of the result of the device, the use of these methods ensures that respondents cannot be identified on the basis of their answers. Warner (1965) was first to develop an ingenious method of collecting information on sensitive characters. It provides the opportunity of reducing response biases due to dishonest answers to sensitive questions. According to the method, for estimating the population proportion π possessing the sensitive character "G", a simple random with replacement sample of n persons is drawn from the population. Each interviewee in the sample is furnished an identical randomization device where the outcome "I possess character G" occurs with probability P while its complement "I do not possess character G" occurs with probability (1-P). The respondent answers "Yes" if the outcome of the randomization device tallies with his actual status otherwise he/she answers "No". Some

Corresponding Author: Tanveer A. Tarray
E-mail: *tanveerstat@gmail.com*

modifications in the model have been suggested by Searls (1967), Fox and Tracy (1986), Kuk (1990), Mangat (1990), Mangat and Singh (1990), Singh and Mathur (2005), Shabir and Gupta (2005), Singh and Tarray (2012, 2013, 2014), Tarray and Singh (2014, 2015, 2016) and Tarray (2016). Greenberg et al. (1969) provided theoretical framework for a modification to the Warner's model. The proposed method consisted in modifying the randomization device where the second outcome "I do not possess the character G" was replaced by the outcome "I possess the character *Y*" where "*Y*" was unrelated to character "G". This modified model is now known as 'unrelated question model, or U- model'.

Hong et al. (1994) proposed a stratified RR technique under the proportional sampling assumption. Under Hong et al.'s (1994) proportional sampling assumption, it may be easy to derive the variance of the proposed estimator. However, it may come at a high cost in terms of time, effort and money. For example, obtaining a fixed number of samples from a rural country in India through a proportional sampling method may be very difficult compared to the researcher's time, effort and money.

To overcome this problem, Kim and Warde (2004), Kim and Warde (2005) and Kim and Elam (2005, 2007) suggested stratified RR techniques using an optimal allocation which are more efficient than a stratified RR technique using a proportional allocation. The extension of the randomized response technique to stratified random sampling may be useful if the investigator is interested in estimating the proportion of HIV/AIDS positively affected persons at different levels such as by rural areas or urban areas, age group or income group, for instance, see Kim and Elam (2007). The study related with Kuk (1990), Singh and Grewal's (2013), Hussain et al. (2014) and Hussain et al. (2016) randomized response models are given in subsequent subsections.

1.1. Kuk's (1990) Randomized Response Model

Kuk (1990) suggested a randomized response model in which respondents belonging to a sensitive group G are instructed to use a deck of cards having the proportion θ_1^* of cards with the statement, "I belong to group G" and if respondents belong to non – sensitive group $\overline{G}$ then they are instructed to use a different deck of cards having the proportion θ_2^* of cards with the statement, "I do not belong to group G". Let π_G be the true proportion of persons belonging to the sensitive group G. Then, the probability of a "Yes" answer in the Kuk's (1990) model is given by

$$\theta_k = \theta_1^* \pi_G + (1-\pi_G)\theta_2^* \tag{1}$$

Let a simple random sample with replacement (SRSWR) of n respondents be chosen from the population, and n_1 is the number of observed "Yes" answers. The number of people n_1 that answer "Yes" is binomially distributed with parameters $\theta_k = \theta_1^* \pi_G + (1-\pi_G)\theta_2^*$ and *n*. For the Kuk (1990) model, an unbiased estimator of the population proportion π_G is given by

$$\hat{\pi}_k = \frac{\hat{\theta} - \theta_2^*}{\theta_1^* - \theta_2^*}, \quad \theta_1^* \neq \theta_2^* \quad . \tag{2}$$

where $\hat{\theta} = \frac{n_1}{n}$ is the proportion of "yes" answers obtained from the n sampled respondents.

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

$$V(\hat{\pi}_k) = \frac{\theta_k(1-\theta_k)}{n(\theta_1^* - \theta_2^*)^2} \quad . \tag{3}$$

Introducing Geometric distribution as a randomization device, Singh and Grewal (2013) have suggested an improvement in the Kuk (1990) model. The description of Singh and Grewal (2013) randomized response technique is given below.

1.2 Singh and Grewal's (2013) Randomized Response Model

In this RRT, an individual respondent in the sample is provided with two decks of cards in the same way as in Kuk (1990) model. In the first deck of cards, let θ_1^* be the proportion of cards with the statement "I belong to a sensitive group G" and (1-θ_1^*) be the proportion of cards with the statement, "I do not belong to a sensitive group G". In the second deck of cards, let θ_2^* be the proportion of cards with the statement, "I do not belong to group G" and (1-θ_1^*) be the proportion of cards with the statement, " I belong to a sensitive group G". Up to here, it is same as the Kuk (1990) randomized response model. If a respondent belongs to a sensitive group G, he/she is instructed to draw cards, one-by-one with replacement, from the first deck of cards until he / she gets the first card bearing the statement of his / her own status, and requested to report the total number of cards, say X drawn by him / her to obtain the first card of his/ her own status. If a respondent belongs to group $\overline{G}$, he / she is instructed to draw cards, one-by-one using with replacement, from the second deck of cards until he / she gets the first card bearing the statement of his/ her own status, and requested to report the total number of cards, say Y, drawn by him/ her to obtain the first card of his / her own status. Since cards are drawn using with replacement sampling, it is clear that X and Y follow geometric distribution with parameters θ_1^* and θ_2^*, respectively [see Singh and Grewal (2013,pp 244-245)]. If Z_i denotes the number of cards reported by the i^{th} respondent then it can be expressed as

$$Z_i = \alpha_i X_i + (1-\alpha_i) Y_i , \tag{4}$$

where α_i is a Bernoulli random variable . An unbiased estimator of π_G due to Singh and Grewal (2013) is given by

$$\hat{\pi}_{G(SG)} = \frac{\theta_1^* \theta_2^* \overline{Z} - \theta_1^*}{\theta_2^* - \theta_1^*} , \quad \theta_1^* \neq \theta_2^* . \tag{5}$$

The variance of $\hat{\pi}_{G(SG)}$ is given by

$$V(\hat{\pi}_{G(SG)}) = \frac{\pi_G(1-\pi_G)}{n} + \frac{\left\{\theta_2^{*2}(1-\theta_1^*)\pi_G + \theta_1^{*2}(1-\theta_1^*)(1-\pi_G)\right\}}{n(\theta_2^* - \theta_1^*)^2} . \tag{6}$$

Further, Hussain et al. (2014) have suggested an alternative to the Singh and Grewal's (2013) model whose description is given below.

1.3 Hussain et al.'s (2014) Randomized Response Model

This model is same as that of Singh and Grewal's randomized response model (2013) RRT except that the respondent belonging to either using first deck or second deck of cards are instructed to report number of cards drawn to obtain $r(>1)$ cards of his/her own status. Then X and Y follow Negative Binomial (NB) distribution with parameters (r, θ_1^*) and (r, θ_2^*), respectively [see Hussain et al. (2014)]. If R_i denotes the number of cards reported by the i^{th} respondent then it can be expressed as randomized response model.

$$R_i = \alpha_i X_i + (1-\alpha_i)Y_i, \tag{7}$$

where α_i is same as in Singh and Grewal (2013) randomized response model. An unbiased estimator of π_G proposed by Hussain et al. (2014) is given by

$$\hat{\pi}_{G(H1)} = \frac{\theta_1^*\theta_2^*\overline{R} - r\theta_1^*}{r(\theta_2^* - \theta_1^*)}, \quad \theta_1^* \neq \theta_2^*, r > 1. \tag{8}$$

with variance given by

$$V(\hat{\pi}_{G(H1)}) = \frac{\pi_G(1-\pi_G)}{n} + \frac{\left\{\theta_2^{*2}(1-\theta_1^*)\pi_G + \theta_1^{*2}(1-\theta_1^*)(1-\pi_G)\right\}}{nr(\theta_2^* - \theta_1^*)^2}. \tag{9}$$

Recently, Hussain et al.(2016) have suggested a generalized randomized response procedure introducing the generalized geometric distribution of order k as a randomization device. The description of Hussain et al.'s (2016) RRT is as follows:

1.4 Hussain et al.'s (2016) Randomized Response Model

This model conceptually, is similar to Kuk (1990), Singh and Grewal's (2013) and Hussain et al. (2014) RRTs. In the procedure due to Hussain et al. (2016), the two decks of cards are exactly the same as used by Singh and Grewal (2013) and Hussain et al. (2014). The only difference is in the event of interest. A respondent belonging to group $G(\overline{G})$ is requested to report the number, say X (Y) of cards drawn to observe k_1 (k_2) consecutive cards for the first time, of his/her actual status. [see Hussain et al.'s (2016)]. Here the response $X(Y)$ follows $GD_{k_1}\left(GD_{k_2}\right)$. If v_i is reported randomized response then it can be expressed as.

$$v_i = \alpha_i X_i + (1-\alpha_i)Y_i, \tag{10}$$

where α_i is same as in Singh and Grewal (2013) and Hussain (2014) randomized response models. An unbiased estimator of π_G proposed by Hussain et al. (2016) is given by

$$\hat{\pi}_{G(H2)} = \frac{\overline{v} - \mu_y}{\mu_x - \mu_y}, \tag{11}$$

where $\mu_x = \dfrac{1-(\theta_1^*)^{k_1}}{(1-\theta_1^*)(\theta_1^*)^{k_1}}$ and $\mu_y = \dfrac{1-(\theta_2^*)^{k_2}}{(1-\theta_2^*)(\theta_2^*)^{k_2}}$.

The variance of $\hat{\pi}_{G(H2)}$ is given by

$$V(\hat{\pi}_{G(H2)}) = \frac{\pi_G(1-\pi_G)}{n} + \frac{\left\{\sigma_x^2\pi_G + \sigma_y^2(1-\pi_G)\right\}}{n(\mu_x - \mu_y)^2}. \qquad . \tag{12}$$

where, $\sigma_x^2 = \dfrac{\left\{1-(2k_1+1)(1-\theta_1^*)(\theta_1^*)^{k_1} - (\theta_1^*)^{2k_1+1}\right\}}{(1-\theta_1^*)^2(\theta_1^*)^{2k_1}}$ and

$$\sigma_y^2 = \frac{\left\{1-(2k_2+1)(1-\theta_2^*)(\theta_2^*)^{k_2} - (\theta_2^*)^{2k_2+1}\right\}}{(1-\theta_2^*)^2(\theta_2^*)^{2k_2}}.$$

In this paper we have suggested a new stratified randomized response model based on Hussain et al. (2016) model. The properties of the suggested stratified randomized response model have been studied under proportional and Neyman allocations. Numerical illustrations are given in support of the present study.

2. Suggested Stratified Randomized Response Model

Let $U = (u_1, u_2, ..., u_N)$ be the dichotomous population and every individual in the population belong either to a sensitive group (possessing a sensitive attribute) G, or to its complement $\overline{G}$. The population is partitioned into L non – overlapping groups such that $N = \sum_{h=1}^{L} N_h$, where N_h is number of units in the h^{th} stratum (h=1,2,…,L). Let $w_h = N_h / N$ be the weight of the h^{th} stratum. The problem is to estimate $\pi_G = \sum_{h=1}^{L} w_h \pi_{Gh}$ $(0 < \pi_G < 1)$, the unknown proportion of population members in group G, where π_{Gh} $(0 < \pi_{Gh} < 1)$ is proportion of respondents with the sensitive trait in a stratum h. To do so, a sample is selected by simple random sampling with replacement (SRSWR) sampling scheme. Let n_h denote the number of units in the sample from stratum h and n denote the total number of units in the samples from all strata so that $n = \sum_{h=1}^{L} n_h$. Now we below give the description of the proposed Randomized Response Technique (RRT):

An individual respondent in the sample of stratum h is provided with two decks of cards in the same way as in the Kuk (1990) RRT. In the first deck of cards θ_{1h}^* is the proportion of cards with the statement, "I$\in$G" and (1-θ_{1h}^*) is the proportion of cards with the statement, "I$\notin$G". In the second deck of cards θ_{2h}^* is the proportion of cards with the statement, "I$\notin$G" and (1- θ_{2h}^*) is the proportion of cards with the statement, "I $\notin$G" and (1-θ_{2h}^*) is the proportion of cards with the statement , "I$\in$G". If a respondent belongs to sensitive group G, s/he is instructed to draw cards, one by one using with replacement, from the first deck of cards. If a respondent belongs to non-sensitive group $\overline{G}$, s/he is instructed to draw cards, one by one using with replacement drawing, from the second deck of cards. Up to here, it is same as the Singh and Grewal (2013) and Hussain et al. (2014) RRTs. The only difference is in the event of interest. A respondent belonging to group $G(\overline{G})$ is requested to report the number, say X_h (Y_h) of cards drawn to observe k_{h1} (k_{h2}) consecutive cards for the first time, of his/her actual status. [see Hussain et al. (2016)].

Let v_{hi} be the number of cards reported by the i^{th} respondent in the h^{th} stratum, then it can be written as

$$v_{hi} = \alpha_{hi} X_{hi} + (1-\alpha_{hi}) Y_{hi} \tag{13}$$

where α_{hi} is a Bernoulli random variable with $E(\alpha_{hi}) = \pi_{Gh}$. In stratum h, the expected number of reported cards is given by

$$E(v_{hi}) = \pi_{Gh}\mu_{xh} + (1-\pi_{Gh})\mu_{xh} \tag{14}$$

Solving (14) for π_{Gh} we have

$$\hat{\pi}_{Gh} = \frac{E(v_{hi}) - \mu_{yh}}{(\mu_{xh} - \mu_{yh})}, \tag{15}$$

where $\mu_{xh} = \dfrac{1-(\theta_{1h}^*)^{k_{1h}}}{(1-\theta_{1h}^*)(\theta_{1h}^*)^{k_{1h}}}$ and $\mu_{yh} = \dfrac{1-(\theta_{2h}^*)^{k_{2h}}}{(1-\theta_{2h}^*)(\theta_{2h}^*)^{k_{2h}}}$.

Estimating $E(v_{hi})$ by $\bar{v}_h = \left[n_h^{-1} \sum_{i=1}^{n_h} v_{hi} \right]$ an unbiased estimator of $\hat{\pi}_G$ is proposed as

$$\hat{\pi}_{G(SGT)} = \sum_{h=1}^{L} w_h \hat{\pi}_{Gh} = \sum_{h=1}^{L} w_h \frac{\bar{v}_{hi} - \mu_{yh}}{\mu_{xh} - \mu_{yh}}. \tag{16}$$

The variance of the estimator $\hat{\pi}_{G(SGT)}$ is given by

$$V(\hat{\pi}_{G(SGT)}) = \sum_{h=1}^{L} \frac{w_h^2}{n_h} V_h, \tag{17}$$

where
$$V_h = \pi_{Gh}(1-\pi_{Gh}) + \frac{\{\sigma_{xh}^2 \pi_{Gh} + \sigma_{yh}^2 (1-\pi_{Gh})\}}{(\mu_{xh} - \mu_{yh})^2}. \tag{18}$$

with $\sigma_{xh}^2 = \dfrac{1-(2k_{1h}+1)(1-\theta_{1h}^*)(\theta_{1h}^*)^{k_{1h}} - (\theta_{1h}^*)^{2k_{1h}+1}}{(1-\theta_{1h}^*)^2(\theta_{1h}^*)^{2k_{1h}}}$ and

$$\sigma_{yh}^2 = \frac{1-(2k_{2h}+1)(1-\theta_{2h}^*)(\theta_{2h}^*)^{k_{2h}} - (\theta_{2h}^*)^{2k_{2h}+1}}{(1-\theta_{2h}^*)^2(\theta_{2h}^*)^{2k_{2h}}}, \quad h = 1, 2, \ldots, L.$$

Now, we derive the variance of the proposed estimator $\hat{\pi}_{G(SGT)}$ under
(i) Proportional allocation and (ii) Neyman allocation.

2.1 Proportional allocation

Under the proportional allocation $n_h = nW_h = n(N_h / N)$, the variance of the estimator in (17) reduces to:

$$V(\hat{\pi}_{G(SGT)})_P = \frac{1}{n} \sum_{h=1}^{L} w_h V_h \tag{19}$$

2.2 Neyman allocation

Information on π_{Gh} is usually not available. But, if prior information on π_{Gh} is available from the past experience or experience gathered in due course of time then it helps to derive the following Neyman allocation formula.

Theorem 2.2 The Neyman allocation of n to n_1, n_2, ..., n_{L-1} and n_L to derive the minimum variance of the estimator $\hat{\pi}_G$ subject to $n=\sum_{h=1}^{L} n_h$ is approximately, given by

$$\frac{n_h}{n}=\frac{w_h\sqrt{V_h}}{\sum_{h=1}^{L} w_h\sqrt{V_h}}. \tag{20}$$

The variance of the suggested estimator $\hat{\pi}_G$ under Neyman allocation is given by

$$V(\hat{\pi}_{G(SGT)})_N=\frac{1}{n}\left[\sum_{h=1}^{L} w_h\sqrt{V_h}\right]^2. \tag{21}$$

Proof is simple so omitted.

3. Efficiency Comparison

In this section we compare the proposed Stratified RRT with that of Hussain et al. (2016) RRT.

From Hussain et al. (2016,Sec.3, equation (3.5)) the variance of the Hussain et al.'s (2016) estimator $\hat{\pi}_{G(H2)}$ for two strata *(h*=2*)* in the population is given by

$$V\left(\hat{\pi}_{G(H2)}\right)=\frac{1}{n}\left[\pi_G(1-\pi_G)+\sum\ \right], \tag{22}$$

where $\pi_G=w_1\pi_{G1}+w_2\pi_{G2}$, $w_1=N_1/N$, $w_2=N_2/N$,

$$\sum=\frac{\left\{\sigma_x^2\pi_G+\sigma_y^2(1-\pi_G)\right\}}{(\mu_x-\mu_y)^2},\quad \sigma_x^2=\frac{\left[1-(2k_1+1)(1-\theta_1^*)(\theta_1^*)^{k_1}-(\theta_1^*)^{2k_1+1}\right]}{(1-\theta_1^*)^2(\theta_1^*)^{2k_1}},$$

$$\sigma_y^2=\frac{\left[1-(2k_2+1)(1-\theta_2^*)(\theta_2^*)^{k_2}-(\theta_2^*)^{2k_2+1}\right]}{(1-\theta_2^*)^2(\theta_2^*)^{2k_2}},\quad \mu_x=\frac{\left(1-\theta_1^{*k_1}\right)}{(1-\theta_1^*)\theta_1^{*k_1}},\quad \mu_y=\frac{\left(1-\theta_2^{*k_2}\right)}{(1-\theta_2^*)\theta_2^{*k_2}}.$$

The efficiency comparison of the suggested RRT under proportional allocation with that of Hussain et al. (2016) RRT for two strata (i.e. h=2) in population, is given in the following theorem.

Theorem 1: Assume that there are two strata (i.e. *h*=2) in the population, $n=n_1+n_2$, $\theta_{11}^*=\theta_{12}^*=\theta_1^*$, $\theta_{21}^*=\theta_{22}^*=\theta_2^*$, $k_{11}=k_{12}=k_1$, $k_{21}=k_{22}=k_2$, $\pi_G=w_1\pi_{G1}+w_2\pi_{G2}$, $\hat{\pi}_{G(SGT)}=w_1\hat{\pi}_{G1}+w_2\hat{\pi}_{G2}$ and $\pi_{G1}\neq\pi_{G2}$. The proposed estimator $\hat{\pi}_{G(SGT)}$ under proportional allocation $\left(i.e.\,\hat{\pi}_{G(SGT)P}\right)$ is more efficient than the Hussain et al. (2016) estimator $\hat{\pi}_{G(H2)}$.

Proof: The variance of the proposed estimator $\hat{\pi}_{G(SGT)}$ under the proportional allocation with the assumption given in the Theorem 1 is given by

$$V\left(\hat{\pi}_{G(SGT)P}\right)=\frac{1}{n}\left[w_1v_1^*+w_2v_2^*\right], \tag{23}$$

$$v_1^*=\left[\pi_{G1}(1-\pi_{G1})+\Sigma_1\right],\ v_2^*=\left[\pi_{G2}(1-\pi_{G2})+\Sigma_2\right],\ \Sigma_1=\frac{\left\{\sigma_x^2\pi_{G1}+\sigma_y^2(1-\pi_{G1})\right\}}{\left(\mu_x-\mu_y\right)^2},$$

$$\Sigma_2=\frac{\left\{\sigma_x^2\pi_{G2}+\sigma_y^2(1-\pi_{G2})\right\}}{\left(\mu_x-\mu_y\right)^2},$$

Now from (22) and (23) we have

$$n\left[V\left(\hat{\pi}_{G(H2)}\right)-V\left(\hat{\pi}_{G(SGT)}\right)_P\right]$$
$$=\left[\pi_G(1-\pi_G)-w_1\pi_{G1}(1-\pi_{G1})-w_2\pi_{G2}(1-\pi_{G2})+\Sigma-w_1\Sigma_1-w_2\Sigma_2\right]$$
$$=\Bigg[\left(w_1\pi_{G1}+w_2\pi_{G2}\right)-\left(w_1\pi_{G1}+w_2\pi_{G2}\right)^2-w_1\pi_{G1}(1-\pi_{G1})-w_2\pi_{G2}(1-\pi_{G2})$$
$$+\frac{1}{\left(\mu_x-\mu_y\right)^2}\Big\{\sigma_x^2\left(w_1\pi_{G1}+w_2\pi_{G2}\right)+\sigma_y^2\left(1-w_1\pi_{G1}-w_2\pi_{G2}\right)-\sigma_x^2w_1\pi_{G1}$$
$$-\sigma_x^2w_2\pi_{G2}-\sigma_y^2w_1(1-\pi_{G1})-\sigma_y^2w_2(1-\pi_{G2})\Big\}\Bigg]$$
$$=w_1w_2\left(\pi_{G1}-\pi_{G2}\right)^2$$

which is always positive. It follows that

$$V\left(\hat{\pi}_{G(SGT)}\right)_P<V\left(\hat{\pi}_{G(H2)}\right)$$

Thus the proposed estimator $\hat{\pi}_{G(SGT)_P}$ is more efficient than Hussain et al. (2016) estimator $\hat{\pi}_{G(H2)}$. Thus the theorem is proved.

The efficiency comparison of the proposed RRT under Nayman allocation with that of Hussain et al.'s (2016) RRT with two strata (i.e. h=2) in the population, is given in the following theorem.

Theorem 2: Suppose that there are two strata (i.e. *h=2*) in the population, $n=n_1+n_2$, $\theta_{11}^*=\theta_{12}^*=\theta_1^*$, $\theta_{21}^*=\theta_{22}^*=\theta_2^*$, $k_{11}=k_{12}=k_1$, $k_{21}=k_{22}=k_2$, $\pi_G=w_1\pi_{G1}+w_2\pi_{G2}$, $\hat{\pi}_{G(SGT)}=w_1\hat{\pi}_{G1}+w_2\hat{\pi}_{G2}$ and $\pi_{G1}\neq\pi_{G2}$. The suggested estimator $\hat{\pi}_{G(SGT)}$ under Neyman allocation is better than the Hussain et al. (2016) estimator $\hat{\pi}_{G(H2)}$.

Proof: For two strata (i.e. *h=2*) in the population and the assumptions given in the Theorem 2, we write the variance of the proposed estimator $\hat{\pi}_{G(SGT)}$ under Neyman allocation as

$$V\left(\hat{\pi}_{G(SGT)}\right)_N=\frac{1}{n}\left[w_1\sqrt{v_1^*}+w_2\sqrt{v_2^*}\right]^2 \tag{24}$$

where v_1^* and v_2^* are same as defined earlier.

From (22) and (24) we have

$$n\left[V\left(\hat{\pi}_{G(H2)}\right)-V\left(\hat{\pi}_{G(SGT)}\right)_N\right]$$

$$
\begin{aligned}
&= \pi_G(1-\pi_G) + \Sigma - \left[w_1^2 v_1^* + w_2^2 v_2^* + 2w_1 w_2\sqrt{v_1^* v_2^*}\right] \\
&= (w_1\pi_{G1} + w_2\pi_{G2})\{1-(w_1\pi_{G1}+w_2\pi_{G2})\} \\
&\quad -\begin{bmatrix} w_1^2\{\pi_{G1}(1-\pi_{G1}) + \Sigma_1\} \\ + w_2^2\{\pi_{G2}(1-\pi_{G2}) + \Sigma_2\} + 2w_1w_2\sqrt{v_1^* v_2^*} \end{bmatrix} \\
&= \left[w_1 w_2(\pi_{G1} + \pi_{G2} - 2\pi_{G1}\pi_{G2}) + \Sigma - w_1^2\Sigma_1 - w_2^2\Sigma_2 - 2w_1w_2\sqrt{v_1^* v_2^*}\right] \\
&= w_1 w_2\left[(\pi_{G1}+\pi_{G2})^2 + \left(\sqrt{v_1^*} - \sqrt{v_2^*}\right)^2\right] \\
&\quad - w_1w_2(\Sigma_1+\Sigma_2) - w_1^2\Sigma_1 - w_2^2\Sigma_2 + \Sigma \\
&= w_1 w_2\left[(\pi_{G1}+\pi_{G2})^2 + \left(\sqrt{v_1^*} - \sqrt{v_2^*}\right)^2\right] + (\Sigma - w_1\Sigma_1 - w_2\Sigma_2) \\
&= w_1 w_2\left[(\pi_{G1}+\pi_{G2})^2 + \left(\sqrt{v_1^*} - \sqrt{v_2^*}\right)^2\right] + 0 \\
&= w_1 w_2\left[(\pi_{G1}+\pi_{G2})^2 + \left(\sqrt{v_1^*} - \sqrt{v_2^*}\right)^2\right] \qquad (25)
\end{aligned}
$$

which is always positive.

It follows that $V(\hat{\pi}_{G(SGT)})_N < V(\hat{\pi}_{G(H2)})$.

Thus the proposed estimator $\hat{\pi}_{G(SGT)}$ under Neyman allocation is more efficient than the Hussain et al. (2016) estimator $\hat{\pi}_{G(H2)}$.

The efficiency comparison of the suggested estimator $\hat{\pi}_{G(SGT)}$ under Neyman allocation with that of under proportional allocation is presented in the following theorem.

Theorem 3: The envisaged estimator $\hat{\pi}_{G(SGT)}$ under Neyman allocation is more efficient than that of the suggested estimator $\hat{\pi}_{G(SGT)}$ under proportional allocation.

Proof: From (19) and (21) we have

$$V(\hat{\pi}_{G(SGT)})_P - V(\hat{\pi}_{G(SGT)})_N = \frac{1}{n}\sum_{h=1}^{L} w_h\left\{\sqrt{v_h} - \left(\sum_{h=1}^{L} w_h\sqrt{v_h}\right)\right\}^2 \qquad (26)$$

which is always positive. Thus from (26) we have

$$V(\hat{\pi}_{G(SGT)})_N < V(\hat{\pi}_{G(SGT)})_P$$

This completes the proof of the theorem.

4. Numerical Illustration

To have tangible idea about the performance of the proposed estimator $\hat{\pi}_G$ under different allocations over the $\hat{\pi}_{G(H2)}$ due to Hussain et al. (2016). We have computed the $PRE\left(\left(\hat{\pi}_{G(SGT)}\right)_P, \hat{\pi}_{G(H2)}\right)$, $PRE\left(\left(\hat{\pi}_{G(SGT)}\right)_N, \hat{\pi}_{G(H2)}\right)$ and $PRE\left(\left(\hat{\pi}_{G(SGT)}\right)_N, \left(\hat{\pi}_{G(SGT)}\right)_P\right)$ by using the formulae:

$$PRE\left(\hat{\pi}_{G(SGT)_P}, \hat{\pi}_{G(H2)}\right) = \frac{V\left(\hat{\pi}_{G(H2)}\right)}{V\left(\hat{\pi}_{G(SGT)}\right)_P} \times 100$$

$$= \frac{\left[\pi_G\left(1-\pi_G\right)+\Sigma\right]}{\left[w_1\pi_{G1}\left(1-\pi_{G1}\right)+w_2\pi_{G2}\left(1-\pi_{G2}\right)+w_1\Sigma_1+w_2\Sigma_2\right]} \times 100, \tag{27}$$

$$PRE\left(\hat{\pi}_{G(SGT)_N}, \hat{\pi}_{G(H2)}\right) = \frac{V\left(\hat{\pi}_{G(H2)}\right)}{V\left(\hat{\pi}_{G(SGT)}\right)_N} \times 100$$

$$= \frac{\left[\pi_G\left(1-\pi_G\right)+\Sigma\right]}{\left[w_1^2 v_1^* + w_2^2 v_2^* + 2w_1 w_2\sqrt{v_1^* v_2^*}\right]} \times 100, \tag{28}$$

$$PRE\left(\hat{\pi}_{G(SGT)_N}, \hat{\pi}_{G(SGT)_P}\right) = \frac{V\left(\hat{\pi}_{G(SGT)_P}\right)}{V\left(\left(\hat{\pi}_{G(SGT)}\right)_N\right)} \times 100$$

$$= \frac{\left[w_1\pi_{G1}\left(1-\pi_{G1}\right)+w_2\pi_{G2}\left(1-\pi_{G2}\right)+w_1\Sigma_1+w_2\Sigma_2\right]}{\left[w_1^2 v_1^* + w_2^2 v_2^* + 2w_1 w_2\sqrt{v_1^* v_2^*}\right]} \times 100 \tag{29}$$

where

$$\pi_G = \left(w_1\pi_{G1} + w_2\pi_{G2}\right),$$

$$\Sigma_1 = \frac{\sigma_x^2\pi_{G1} + \sigma_y^2\left(1-\pi_{G1}\right)}{\left(\mu_x - \mu_y\right)^2},$$

$$\Sigma_2 = \frac{\sigma_x^2\pi_{G2} + \sigma_y^2\left(1-\pi_{G2}\right)}{\left(\mu_x - \mu_y\right)^2},$$

$$\Sigma = \frac{\sigma_x^2\pi_G + \sigma_y^2\left(1-\pi_G\right)}{\left(\mu_x - \mu_y\right)^2},$$

$$\sigma_x^2 = \frac{\left[1-\left(2k_1+1\right)\left(1-\theta_1^*\right)\left(\theta_1^*\right)^{k_1} - \left(\theta_1^*\right)^{2k_1+1}\right]}{\left(1-\theta_1^*\right)^2\left(\theta_1^*\right)^{2k_1}},$$

$$\sigma_y^2 = \frac{\left[1-\left(2k_2+1\right)\left(1-\theta_2^*\right)\left(\theta_2^*\right)^{k_2} - \left(\theta_2^*\right)^{2k_2+1}\right]}{\left(1-\theta_2^*\right)^2\left(\theta_2^*\right)^{2k_2}},$$

$$v_1^* = \left[\pi_{G1}\left(1-\pi_{G1}\right)+\Sigma_1\right],\ v_1^* = \left[\pi_{G1}\left(1-\pi_{G1}\right)+\Sigma_1\right].$$

We have computed the PREs for different values of $\left(w_1, w_2\right)$, $\left(\pi_{G_1}, \pi_{G2}\right)$, $\left(\theta_1^*, \theta_2^*\right)$ and fixed values of $k_1 = k_2 = 2;$ and findings are displayed in Tables 1, 2 and 3.

It is observed from tables 1 to 2 that the percent relative efficiencies are larger than 100% which follows that the proposed estimator $\hat{\pi}_{G(SGT)}$ under proportional and Neyman allocations are more efficient than the Hussain et al.'s (2016) estimator $\hat{\pi}_{G(H2)}$ with substantial gain in efficiency. Table 3 exhibits that the proposed estimator $\hat{\pi}_{G(SGT)}$ under Neyman allocation performs better than that under proportional allocation. Thus the suggested RRT in stratified random sampling is to be preferred over Hussain's et al. (2016) RRT in simple random sampling with replacement (SRSWR) sampling scheme.

5. Conclusion

This paper advocates the problem of estimating the population proportion π_G of sensitive attribute based on stratified sampling scheme. An estimator $\hat{\pi}_{G(SGT)}$ for population proportion in Hussain et al. (2016) randomized response technique using stratified random sampling has been proposed. The variance of the suggested estimator $\hat{\pi}_{G(SGT)}$ are obtained under proportional and Neyman allocations. We have shown theoretically that the suggested estimator $\hat{\pi}_{G(SGT)}$ under both allocations proportional as well as Neyman is better than the Hussain et al. (2016) $\hat{\pi}_{G(H2)}$. We have further proved that the proposed estimator $\hat{\pi}_{G(SGT)}$under Neyman allocation is better than that under proportional allocation. So the proposed model is more efficient than the Hussain et al. (2016) randomized response model. We have also shown numerically that the proposed randomized response model is more efficient than Hussain et al. (2016) randomized response model. Thus our recommendation is to use proposed randomized response model in practice.

Acknowledgments

The authors are grateful to the editor in chief and to the learned referees for their valuable suggestions regarding improvement of this article.

References

Fox, J.A. and Tracy, P.E. (1986). *Randomized Response: A Method of Sensitive Surveys.* Newbury Park, CA: SEGE Publications.

Hong, K., Yum, J. and Lee, H. (1994). A stratified randomized response technique. *Korean Journal of Applied Statistics*, **7,** 141-147.

Hussain, Z., Al-Zahrani, B., Shabbir, J. and Khan, M. (2014). On using Negative Binomial Distribution as a randomization device in sensitive surveys. *Communication in Statistics Simulation and Computation,* **45(10)**, 3584-3596. DOI: 10.1080/03610918.2014.948193.

Hussain, Z., Shabbir, J., Pervez, Z., Shah, F.S. and Khan, M. (2016). Generalized Geometric distribution of order *k*: A flexible choice to randomize the response. *Communication in Statistics Simulation and Computation,* DOI: 10.1080/03610918.2015.1129410.

Kim, J.M. and Elam, M.E. (2005). A two-stage stratified Warner's randomized response model using optimal allocation. *Metrika*, **61**, 1-7.

Kim, J.M. and Warde, W.D. (2004). A stratified Warner randomized response model. *Journal of Statistics Planning and Inference,* **120**, 155-165.

Kim, J.M. and Warde, W.D. (2005). A mixed randomized response model. *Journal of Statistics Planning and Inference*, **133**, 211-221.

Kuk, A.Y.C. (1990). Asking sensitive questions indirectly. *Biometrika*, **77**, 436-438.

Mangat, N.S. (1994). An improved randomized response strategy. *Journal of Royal Statistical Society*, **B56(1)**, 93–95.

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

Searls, D.T. (1967). A note on the use of an approximately known coefficient of variation. *The American Statistician*, **21(2)**, 20-21.

Shabbir, J. and Gupta, S. (2005). Optimal allocation in stratified randomized response model. *Pakistan Journal of Statistics and Operation Research*, **1,** 15-22.

Singh, S. and Grewal, I.S. (2013). Geometric distribution as a randomization device: Implemented to the Kuk's model. *International Journal of Contempt Mathematical Sciences*, **8(5)**, 243-248.

Singh, H.P. and Mathur, N. (2005). Estimation of the population mean when the coefficient of variation is known using scrambled response technique. *Journal of Statistics Planning and Inference*, **131**, 135-144.

Singh, H.P. and Tarray, T.A. (2012). A Stratified Unknown repeated trials in randomized response sampling. *Communication of the Korean Statistical Society*, **19(6)**, 751-759.

Singh, H.P. and Tarray, T.A. (2013). An alternative to Kim and Warde's mixed randomized response model. *Statistics and Operation Research Transactions*, **37 (2)**, 189-210.

Singh, H.P. and Tarray, T.A. (2014). An alternative to stratified Kim and Warde's randomized response model using optimal (Neyman) allocation. *Model Assisted Statistical Applications*, **9**, 37-62.

Tarray, T.A. (2016). *Statistical Sample Survey Methods and Theory*. Elite Publishers (onlinegatha.com), INDIA.

Tarray, T.A. and Singh, H.P. (2014). A proficient randomized response model. *Istatistika: Journal of Turkey Statistical Association*,**7(3)**, 87-98.

Tarray, T.A. and Singh, H.P. (2015). Some improved additive randomized response models utilizing higher order moments ratios of scrambling variable. *Model Assisted Statistical Applications*, **10**, 361-383.

Tarray, T.A. and Singh, H.P. (2016). An adroit randomized response new additive scrambling model. *Gazi University Journal of Sciences*, **29(1)**, 159-165.

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

Table 1 The percent relative efficiency of $\left(\hat{\pi}_{G(SGT)}\right)_P$ with respect to $\hat{\pi}_{G(H2)}$

π_{G1}	π_{G2}	w_1	w_2	$\theta_1^*=0.1$ $\theta_2^*=0.5$	$\theta_1^*=0.2$ $\theta_2^*=0.6$	$\theta_1^*=0.3$ $\theta_2^*=0.7$	$\theta_1^*=0.4$ $\theta_2^*=0.8$
0.10	0.90	0.7	0.3	129.09	125.51	122.40	120.35
0.11	0.89	0.7	0.3	126.94	123.68	120.83	118.95
0.12	0.88	0.7	0.3	124.94	121.96	119.35	117.63
0.13	0.87	0.6	0.4	122.71	120.02	117.70	116.15
0.14	0.86	0.6	0.4	121.16	118.67	116.53	115.10
0.15	0.85	0.6	0.4	119.69	117.40	115.42	114.10
0.16	0.84	0.4	0.6	114.71	112.99	111.54	110.57
0.17	0.83	0.4	0.6	113.77	112.18	110.83	109.92
0.18	0.82	0.4	0.6	112.88	111.40	110.14	109.30
0.19	0.81	0.3	0.7	109.67	108.56	107.62	107.00
0.20	0.80	0.3	0.7	109.04	108.01	107.14	106.55
0.21	0.79	0.3	0.7	108.43	107.48	106.67	106.12
0.22	0.78	0.7	0.3	111.05	109.86	108.80	108.08
0.23	0.77	0.7	0.3	110.11	109.03	108.06	107.41
0.24	0.76	0.7	0.3	109.23	108.25	107.37	106.78
0.25	0.75	0.6	0.4	108.83	107.89	107.05	106.49
0.26	0.74	0.6	0.4	108.06	107.20	106.44	105.93
0.27	0.73	0.6	0.4	107.33	106.55	105.87	105.40
0.28	0.72	0.4	0.6	105.84	105.21	104.67	104.30
0.29	0.71	0.4	0.6	105.31	104.74	104.25	103.91
0.30	0.70	0.4	0.6	104.80	104.29	103.85	103.55
0.31	0.69	0.3	0.7	103.60	103.21	102.88	102.66
0.32	0.68	0.3	0.7	103.23	102.89	102.59	102.39
0.33	0.67	0.3	0.7	102.89	102.58	102.32	102.14

Table 2 The percent relative efficiency of $\left(\hat{\pi}_{G(SGT)}\right)_N$ with respect to $\hat{\pi}_{G(H2)}$

π_{G1}	π_{G2}	w_1	w_2	$\theta_1^*=0.1$ $\theta_2^*=0.5$	$\theta_1^*=0.2$ $\theta_2^*=0.6$	$\theta_1^*=0.3$ $\theta_2^*=0.7$	$\theta_1^*=0.4$ $\theta_2^*=0.8$
0.10	0.90	0.7	0.3	153.06	148.81	144.27	141.29
0.11	0.89	0.7	0.3	147.84	144.13	140.17	137.55
0.12	0.88	0.7	0.3	143.25	139.99	136.51	134.20
0.13	0.87	0.6	0.4	138.47	135.62	132.66	130.69
0.14	0.86	0.6	0.4	135.20	132.63	129.98	128.21
0.15	0.85	0.6	0.4	132.22	129.92	127.54	125.94
0.16	0.84	0.4	0.6	123.29	121.60	119.94	118.83
0.17	0.83	0.4	0.6	121.56	120.02	118.50	117.48
0.18	0.82	0.4	0.6	119.95	118.54	117.15	116.22
0.19	0.81	0.3	0.7	114.67	113.62	112.62	111.94
0.20	0.80	0.3	0.7	113.60	112.64	111.72	111.10
0.21	0.79	0.3	0.7	112.59	111.71	110.87	110.29
0.22	0.78	0.7	0.3	116.63	115.57	114.46	113.70
0.23	0.77	0.7	0.3	115.09	114.13	113.14	112.46
0.24	0.76	0.7	0.3	113.66	112.81	111.92	111.31
0.25	0.75	0.6	0.4	113.01	112.18	111.34	110.77
0.26	0.74	0.6	0.4	111.79	111.04	110.29	109.77
0.27	0.73	0.6	0.4	110.65	109.99	109.31	108.84
0.28	0.72	0.4	0.6	108.42	107.87	107.34	106.97
0.29	0.71	0.4	0.6	107.61	107.12	106.64	106.31
0.30	0.70	0.4	0.6	106.85	106.42	105.99	105.69
0.31	0.69	0.3	0.7	105.10	104.77	104.45	104.23
0.32	0.68	0.3	0.7	104.56	104.27	103.99	103.79
0.33	0.67	0.3	0.7	104.06	103.80	103.55	103.38

Table 3 The percent relative efficiency of $\left(\hat{\pi}_{G(SGT)}\right)_P$ with respect to $\left(\hat{\pi}_{G(SGT)}\right)_N$

π_{G1}	π_{G2}	w_1	w_2	$\theta_1^*=0.1$, $\theta_2^*=0.5$	$\theta_1^*=0.2$, $\theta_2^*=0.6$	$\theta_1^*=0.3$, $\theta_2^*=0.7$	$\theta_1^*=0.4$, $\theta_2^*=0.8$
0.10	0.90	0.7	0.3	118.56	118.56	117.87	117.40
0.11	0.89	0.7	0.3	116.46	116.54	116.01	115.64
0.12	0.88	0.7	0.3	114.65	114.78	114.38	114.09
0.13	0.87	0.6	0.4	112.84	113.00	112.71	112.51
0.14	0.86	0.6	0.4	111.59	111.76	111.55	111.39
0.15	0.85	0.6	0.4	110.47	110.66	110.50	110.38
0.16	0.84	0.4	0.6	107.48	107.61	107.53	107.47
0.17	0.83	0.4	0.6	106.84	106.98	106.92	106.87
0.18	0.82	0.4	0.6	106.26	106.41	106.36	106.33
0.19	0.81	0.3	0.7	104.56	104.66	104.64	104.63
0.20	0.80	0.3	0.7	104.18	104.29	104.28	104.27
0.21	0.79	0.3	0.7	103.84	103.94	103.94	103.93
0.22	0.78	0.7	0.3	105.03	105.20	105.20	105.20
0.23	0.77	0.7	0.3	104.52	104.68	104.70	104.70
0.24	0.76	0.7	0.3	104.06	104.22	104.23	104.24
0.25	0.75	0.6	0.4	103.84	103.98	104.00	104.01
0.26	0.74	0.6	0.4	103.45	103.59	103.61	103.62
0.27	0.73	0.6	0.4	103.10	103.22	103.25	103.26
0.28	0.72	0.4	0.6	102.43	102.53	102.55	102.56
0.29	0.71	0.4	0.6	102.19	102.27	102.29	102.31
0.30	0.70	0.4	0.6	101.96	102.04	102.06	102.07
0.31	0.69	0.3	0.7	101.45	101.51	101.52	101.53
0.32	0.68	0.3	0.7	101.29	101.34	101.36	101.37
0.33	0.67	0.3	0.7	101.14	101.19	101.21	101.22