A STRATIFIED RANDOMIZED RESPONSE MODEL FOR SENSITIVE CHARACTERISTICS USING NEGATIVE BINOMIAL DISTRIBUTION

Tanveer A. Tarray and Housila P. Singh
School of Studies in Statistics,
Vikram University, Ujjain – India, 456010.

ABSTRACT
Taking the clue from the pioneer work of Hussain et al.'s (2014) 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. The study has been also carried out using modified Neyman allocation in the presence of crude prior estimates. Numerical illustrations are also given in support of the present study.

KEYWORDS: Randomized response sampling, Estimation of proportion, Respondents protection, Negative Binomial Distribution.

MSC: 62D05.

RESUMEN
Tomando la señalización hecha en el trabajo pionero de Hussain et al. (2014) sugerimos un nuevo modelo de respuestas aleatorizadas estratificado. Las propiedades del modelo de respuestas aleatorizadas sugerido han sido estudiadas bajo la afijación proporcional y la de "Neyman". El estudio ha sido realizado usando una modificación de la afijación de Neyman en la presencia de un estimador a prior crudo. Ilustraciones numéricas también son dadas para soportar el presente estudio.

1. INTRODUCTION

One of the leading paraphernalia for obtaining data pertaining to human populations is the social survey. To measure opinions, attitudes, and behaviors that cover a wide band of interests, the social survey has been established as being tremendously practical. The surveys are conducted due to many reasons, non – availability of certain facts / information in the archives being the most understandable and apparent. For instance, if one is interested in knowing crime rate, information about unseen crimes or unreported victimization experience is not available in formal records on crimes. Sometimes the facts about the individuals (in a population) are inaccessible to the investigators for legal reasons. Questionnaires, in particular social surveys, generally consist of many items. Some of the items may be about sensitive / high risk behavior, due to the social stigma carried by them. One problem with research on high – risk behavior is that respondents may consciously or unconsciously give wrong information. In psychological surveys, a social desirability bias has been observed as a major cause of distortion in standardized personality measures. Survey researchers have similar concerns about the truth of survey results/ findings about such topics as drunk driving, use of marijuana, tax evasion, illicit drug use, induced abortion, shop lifting, child abuse, family disturbances, cheating in exams, HIV/AIDS, and sexual behavior. The most serious problem in studying certain social problems that are sensitive in nature (e.g. induced abortion, drug usage, tax evasion, etc.) is lack of reliable measure of their incidence or prevalence. Thus to obtain trustworthy data on such confidential matters, especially the sensitive ones, instead of open surveys alternative procedures are required. Such an alternative procedure known as "randomized response technique" (RRT) was first introduced by Warner (1965). 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 modifications in the model have been suggested by Singh (2003), Mahajan (2005-2006), Mahajan et al. (2007), Hong (2005-2006), Grewal et al. (2005-2006), Sidhu and Bansal (2008), Perri (2008), Zaizai et al. (2008) and Singh and Tarray (2012,2013,2014). Greenberg et al. (1969) provided theoretical framework for a modification to the Warner's model proposed by Horvitz et al. (1969). 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) envisaged 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) 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, p. 216). The study related with Kuk (1990), Singh and Grewal's (2013) and Hussain et al.'s (2014) randomized response models whose description are given in subsequent subsections.

1.1 Kuk (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

$$\theta_k = \theta_1^* \pi_G + (1-\pi_G)\theta_2^* \tag{1.1}$$

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

$$\hat{\pi}_k = \frac{\hat{\theta} - \theta_2^*}{\theta_1^* - \theta_2^*}, \quad \theta_1^* \neq \theta_2^* \quad . \tag{1.2}$$

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

$$V(\hat{\pi}_k) = \frac{\theta_k (1-\theta_k)}{n(\theta_1^* - \theta_2^*)^2} \quad . \tag{1.3}$$

Singh and Grewal (2013) have suggested improvement in the Kuk (1990) model Geometric distribution as a randomization device. They have claimed that their method is more protective and efficient than the Kuk (1990) model while doing surveys in practice. The description of Singh and Grewal (2013) randomized response technique is given below.

1.2 Singh and Grewal (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

$$Z_i = \alpha_i X_i + (1-\alpha_i) Y_i \quad , \tag{1.4}$$

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

$$\hat{\pi}_{G(SG)} = \frac{\theta_1^* \theta_2^* \overline{Z} - \theta_1^*}{\theta_2^* - \theta_1^*}, \quad \theta_1^* \neq \theta_2^* . \tag{1.5}$$

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

$$V(\hat{\pi}_{G(SG)}) = \frac{\pi_G (1-\pi_G)}{n} + \frac{\left\{\theta_2^{*2}(1-\theta_1^*)\pi_G + \theta_1^{*2}(1-\theta_1^*)(1-\pi_G)\right\}}{n(\theta_2^* - \theta_1^*)^2}. \tag{1.6}$$

Recently, Hussain et al.’s (2014) have suggested improvement in the Singh and Grewal’s (2013) model. They have claimed that their method is more protective and efficient than the Singh and Grewal (2013) model while doing surveys in practice. The description of Hussain et al.’s (2014) RRT is as follows in the next section.

1.2 Hussain et al.’s (2014) randomized response model

This model is the 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.’s (2014)].)]. If R_i denotes the number of cards reported by the i[th] respondent then it can be expressed as randomized response model.

$$R_i = \alpha_i X_i + (1-\alpha_i) Y_i \ , \tag{1.7}$$

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.’s (2014) is given by

$$\hat{\pi}_{G(P)} = \frac{\theta_1^* \theta_2^* \overline{R} - r\theta_1^*}{r(\theta_2^* - \theta_1^*)}, \quad \theta_1^* \neq \theta_2^*, r > 1. \tag{1.8}$$

with variance given by

$$V(\hat{\pi}_{G(P)}) = \frac{\pi_G (1-\pi_G)}{n} + \frac{\left\{\theta_2^{*2}(1-\theta_1^*)\pi_G + \theta_1^{*2}(1-\theta_1^*)(1-\pi_G)\right\}}{nr(\theta_2^* - \theta_1^*)^2}. \tag{1.9}$$

In this paper we have suggested a new stratified randomized response model based on Hussain et al.’s (2014) model. The properties of the suggested stratified randomized response model have been studied under proportional and Neyman allocations. A modified Neyman allocation procedure is also given in presence of crude prior estimates. Numerical illustrations are also given in support of the present study.

2. SUGGESTED STRATIFIED RANDOMIZED RESPONSE

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 , "$I \in G$". Up to here, it is same as the Kuk (1990) RRT. If a respondent belongs to sensitive group G, he / she 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 $\overline{G}$, he / she 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) RRT. The respondents belonging to either using first deck or second deck of cards are instructed to report number of cards drawn to obtain $r_h (> 1)$ cards of his/her own status. Let X_h and Y_h be the number of cards drawn by the respondent from the first and second deck of cards respectively to obtain $r_h (> 1)$ cards of his/her own status. Then X_h and Y_h follow a Negative Binomial (NB) distribution with parameters (r_h, θ_{1h}^*) and (r_h, θ_{2h}^*), respectively [see, Hussain et al. (2014)].

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

$$R_{hi} = \alpha_{hi} X_{hi} + (1 - \alpha_{hi}) Y_{hi} \quad , \tag{2.1}$$

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

$$E(R_{hi}) = E(\alpha_{hi})E(X_{hi}) + E(1 - \alpha_{hi})E(Y_{hi})$$
$$= \left[\frac{r_h \pi_{Gh}}{\theta_{1h}^*} + \frac{r_h (1 - \pi_{Gh})}{\theta_{2h}^*} \right] = \left[\frac{r_h \pi_{Gh} (\theta_{2h}^* - \theta_{1h}^*) + r_h \theta_{1h}^*}{\theta_{1h}^* \theta_{2h}^*} \right]. \tag{2.2}$$

Let $\overline{R}_h$ be the sample mean of reported response in stratum 'h'. Then an unbiased estimator of π_{Gh} is given by

$$\hat{\pi}_{Gh} = \left[\frac{(\theta^*_{1h}\theta^*_{2h}\overline{R}_h - r_h\theta^*_{1h})}{r_h(\theta^*_{2h} - \theta^*_{1h})} \right], \theta^*_{1h} \neq \theta^*_{2h}, r_h > 1. \quad (2.3)$$

Thus, an unbiased estimator of the population proportion $\pi_G = \sum_{h=1}^{L} w_h \pi_{Gh}$ is given by

$$\hat{\pi}_G = \sum_{h=1}^{L} w_h \hat{\pi}_{Gh} = \sum_{h=1}^{L} \frac{w_h(\theta^*_{1h}\theta^*_{2h}\overline{R}_h - r_h\theta^*_{1h})}{r_h(\theta^*_{2h} - \theta^*_{1h})}, \theta^*_{1h} \neq \theta^*_{2h}, r_h > 1. \quad (2.4)$$

The variance of $\hat{\pi}_G$ is given in the following theorem.

Theorem 2.1 The variance of the estimator $\hat{\pi}_G$ is given by

$$V(\hat{\pi}_G) = \sum_{h=1}^{L} \frac{w_h^2}{n_h} V_h \quad (2.5)$$

where

$$V_h = \pi_{Gh}(1-\pi_{Gh}) + \frac{\left\{\theta^{*2}_{2h}(1-\theta^*_{1h})\pi_{Gh} + \theta^{*2}_{1h}(1-\theta^*_{1h})(1-\pi_{Gh})\right\}}{r_h(\theta^*_{2h} - \theta^*_{1h})^2}. \quad (2.6)$$

Proof- Since the selections are made independently from stratum to stratum, therefore, the variance of $\hat{\pi}_G$ is given by

$$V(\hat{\pi}_G) = \sum_{h=1}^{L} w_h^2 V(\hat{\pi}_{Gh})$$

$$= \sum_{h=1}^{L} \frac{w_h^2 \theta^{*2}_{1h}\theta^{*2}_{2h} V(R_{hi})}{n_h r_h^2 (\theta^*_{2h} - \theta^*_{1h})^2}, \quad (2.7)$$

The variance of R_{hi} is obtained as follows:

$$V(R_{hi}) = E(R^2_{hi}) - (E(R_{hi}))^2$$

$$= \left[\frac{r_h(r_h + 1 - \theta^*_{1h})\pi_{Gh}}{\theta^{*2}_{1h}} + \frac{r_h(r_h + 1 - \theta^*_{2h})(1-\pi_{Gh})}{\theta^{*2}_{2h}} - \left\{ \frac{r_h\pi_{Gh}}{\theta^*_{1h}} + \frac{r_h(1-\pi_{Gh})}{\theta^*_{2h}} \right\}^2 \right]$$

$$= \frac{1}{\theta^{*2}_{1h}\theta^{*2}_{2h}} \left[\pi_{Gh}(1-\pi_{Gh}) r_h^2 (\theta^*_{2h} - \theta^*_{1h})^2 + r_h \left\{ \theta^{*2}_{2h}(1-\theta^*_{1h})\pi_{Gh} + \theta^{*2}_{1h}(1-\theta^*_{2h})(1-\pi_{Gh}) \right\} \right] \quad (2.8)$$

Putting (2.8) in (2.7) we get the desired result in (2.5). This completes the proof of the theorem.

Now, we derive the variance of the proposed estimator $\hat{\pi}_G$ in

(i) Proportional and (ii) Neyman allocations.

2.1 Proportional allocation

Under the proportional allocation $n_h = nW_h = n(N_h / N)$, the variance of the estimator $\hat{\pi}_G$ in (2.5) reduces to:

$$V(\hat{\pi}_G)_P = \frac{1}{n} \sum_{h=1}^{L} w_h V_h \quad (2.9)$$

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 $n_1, n_2, \ldots, n_{L-1}$ 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

$$\frac{n_h}{n} = \frac{w_h\sqrt{V_h}}{\sum_{h=1}^{L} w_h\sqrt{V_h}} \quad . \tag{2.10}$$

Proof- is simple so omitted.

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

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

(2.11)

3. EFFICIENCY COMPARISON

From (2.9) and (2.11) we have

$$V(\hat{\pi}_G)_P - V(\hat{\pi}_G)_N = \frac{1}{n}\left[\sum_{h=1}^{L} w_h V_h - \left(\sum_{h=1}^{L} w_h\sqrt{V_h}\right)^2\right]$$

$$= \frac{1}{n}\sum_{h=1}^{L} w_h\left[\sqrt{V_h} - \left(\sum_{h=1}^{L} w_h\sqrt{V_h}\right)\right]^2 \tag{3.1}$$

which is always positive. It follows that the proposed estimator $\hat{\pi}_G$ under Neyman allocation is more efficient than that under proportional allocation, (i.e. $V(\hat{\pi}_G)_N \le V(\hat{\pi}_G)_P$). However, Neyman allocation needs prior information about π_{Gh} (h=1,2,…,L) which can be obtained either through pilot sample surveys or past data or past experience or the experienced gathered in due course of time.

We, now, compare the efficiency of the proposed stratified randomized response technique and the Hussain et al.'s (2014) randomized response technique.

To compare the proposed estimator $\hat{\pi}_G$ with that of Hussain et al.'s (2014) estimator $\hat{\pi}_{G(P)}$ we have made the following assumptions:

{L(number of strata) =2, $\theta_{11}^* = \theta_{12}^* = \theta_1^*, \theta_{21}^* = \theta_{22}^* = \theta_2^*,$ $r_1=r_2=r$; and

$$\hat{\pi}_G = w_1\hat{\pi}_{G_1} + w_2\hat{\pi}_{G_2}\,;\, \pi_{G1} \ne \pi_{G2}.\ \} \tag{3.2}$$

Under the above assumption we write the variance of Hussain et al.'s (2014) estimator $\hat{\pi}_{G(P)}$ in stratified random sampling and variance of the proposed estimator $\hat{\pi}_G$ under proportional and Neyman allocations respectively as

$$V(\hat{\pi}_{(G(P)}) = \frac{1}{n}\left[(w_1\pi_{G1} + w_2\pi_{G2})(1 - w_1\pi_{G1} - w_2\pi_{G2}) + (a-b)(w_1\pi_{G1} + w_2\pi_{G2}) + b\right], \tag{3.3}$$

$$V(\hat{\pi}_G)_P = \frac{1}{n}\left[(w_1V_1 + w_2V_2)\right] \ , \tag{3.4}$$

$$V(\hat{\pi}_G)_N = \frac{1}{n}\left(w_1\sqrt{V_1} + w_2\sqrt{V_2}\right)^2 \ , \tag{3.5}$$

where

$$a = \frac{\theta_2^{*2}(1-\theta_1^*)}{r(\theta_2^* - \theta_1^*)^2}, b = \frac{\theta_1^{*2}(1-\theta_2^*)}{r(\theta_2^* - \theta_1^*)^2},$$

$$V_1 = [\pi_{G1}(1-\pi_{G1}) + (a-b)\pi_{G1} + b],$$

$$V_2 = [\pi_{G2}(1-\pi_{G2}) + (a-b)\pi_{G2} + b],$$

From (3.3), (3.4) and (3.5), we have

$$n(V(\hat{\pi}_{G(P)}) - V(\hat{\pi}_G)_P) = w_1 w_2 [\pi_{G1} - \pi_{G2}]^2 > 0 \tag{3.6}$$

$$n(V(\hat{\pi}_{G(P)}) - V(\hat{\pi}_G)_N) = w_1 w_2 \left[(\pi_{G1} - \pi_{G2})^2 + \left(\sqrt{V_1} - \sqrt{V_2}\right)^2 \right] > 0 \tag{3.7}$$

From (3.6) and (3.7) we state the following theorem.

Theorem 3.1 Under the assumptions given by (3.2), the proposed estimator $\hat{\pi}_G$ (with proportional and Neyman allocations) is always superior to the estimator $\hat{\pi}_{G(P)}$ (in stratified random sampling) recently proposed by Hussain et al.'s (2014).

If prior information on $\pi_{G1}, \pi_{G2}, w_1, w_2$ and n can be roughly obtained and θ_1^*, θ_2^* and 'r' are chosen by the researcher, we can compute the percent relative efficiencies (PREs) of the proposed estimator $\hat{\pi}_G$ under proportional and Neyman allocations with respect to Hussain et al.'s (2014) estimator $\hat{\pi}_{G(P)}$ and the percent relative efficiency of the proposed estimator $\hat{\pi}_G$ under Neyman allocation estimator $(\hat{\pi}_G)_N$ with respect to proportional allocation estimator $(\hat{\pi}_G)_P$ by using the formulae:

$$e_1 = PRE((\hat{\pi}_G)_P, \hat{\pi}_{G(P)}) = \frac{[(w_1\pi_{G1} + w_2\pi_{G2})(1 - w_1\pi_{G1} - w_2\pi_{G2}) + (a-b)(w_1\pi_{G1} + w_2\pi_{G2}) + b]}{(w_1 V_1 + w_2 V_2)} \times 100, \tag{3.8}$$

$$e_2 = PRE((\hat{\pi}_G)_N, \hat{\pi}_{G(P)}) = \frac{[(w_1\pi_{G1} + w_2\pi_{G2})(1 - w_1\pi_{G1} - w_2\pi_{G2}) + (a-b)(w_1\pi_{G1} + w_2\pi_{G2}) + b]}{\left(w_1\sqrt{V_1} + w_2\sqrt{V_2}\right)^2} \times 100, \tag{3.9}$$

$$e_3 = PRE((\hat{\pi}_G)_N, (\hat{\pi}_{G(P)}) = \frac{(w_1 V_1 + w_2 V_2)}{\left(w_1\sqrt{V_1} + w_2\sqrt{V_2}\right)^2} \times 100, \tag{3.10}$$

where a,b,V_1 and V_2 are same as defined just above the equation (3.6). We have computed $e_1 = (PRE((\hat{\pi}_G)_P, \hat{\pi}_{G(P)}))$, $e_2 = (PRE((\hat{\pi}_G)_N, \hat{\pi}_{G(P)}))$ and $e_3 = (PRE((\hat{\pi}_G)_N, (\hat{\pi}_{G(P)}))$ for different values of $\pi_{G1}, \pi_{G2}, \theta_1^*,, \theta_2^*, r, w_1$ and w_2 and findings are displayed in Tables 3.1, 3.2 and 3.3.

Tables 1-3 exhibits that the values of the $e_1 = (PRE((\hat{\pi}_G)_P, \hat{\pi}_{G(P)}))$, $e_2 = (PRE((\hat{\pi}_G)_N, \hat{\pi}_{G(P)}))$ and $e_3 = (PRE((\hat{\pi}_G)_N, (\hat{\pi}_{G(P)}))$ are greater than 100. Thus, we conclude that the proposed new randomization device can be efficiently and cooperatively used in real practice to estimate the proportion of a sensitive characteristic.

Table 3.1 The percent relative efficiency of $(\hat{\pi}_G)_P$ with respect to $\hat{\pi}_{G(P)}$.

π_{G1}	π_{G2}	w_1	w_2	r	θ_1^*	θ_2^*	PRE's
0.8	0.13	0.9	0.1	3.00	0.7	0.3	110.43
0.8	0.13	0.8	0.2	3.00	0.7	0.3	116.83
0.8	0.13	0.7	0.3	3.00	0.7	0.3	120.22
0.8	0.13	0.6	0.4	3.00	0.7	0.3	121.31
0.8	0.13	0.5	0.5	3.00	0.7	0.3	120.59
0.8	0.13	0.4	0.6	3.00	0.7	0.3	118.43
0.8	0.13	0.3	0.7	3.00	0.7	0.3	115.11
0.8	0.13	0.2	0.8	3.00	0.7	0.3	110.83
0.8	0.13	0.1	0.9	3.00	0.7	0.3	105.75
0.9	0.14	0.9	0.1	3.50	0.7	0.3	121.61
0.9	0.14	0.8	0.2	3.50	0.7	0.3	132.26
0.9	0.14	0.7	0.3	3.50	0.7	0.3	136.49
0.9	0.14	0.6	0.4	3.50	0.7	0.3	136.64
0.9	0.14	0.5	0.5	3.50	0.7	0.3	134.03
0.9	0.14	0.4	0.6	3.50	0.7	0.3	129.48
0.9	0.14	0.3	0.7	3.50	0.7	0.3	123.50
0.9	0.14	0.2	0.8	3.50	0.7	0.3	116.44
0.9	0.14	0.1	0.9	3.50	0.7	0.3	108.55

Table 3.2 The percent relative efficiency of $(\hat{\pi}_G)_N$ with respect to $\hat{\pi}_{G(P)}$.

π_{G1}	π_{G2}	w_1	w_2	r	θ_1^*	θ_2^*	PRE's
0.8	0.13	0.9	0.1	3.00	0.7	0.3	112.36
0.8	0.13	0.8	0.2	3.00	0.7	0.3	120.15
0.8	0.13	0.7	0.3	3.00	0.7	0.3	124.35
0.8	0.13	0.6	0.4	3.00	0.7	0.3	125.71
0.8	0.13	0.5	0.5	3.00	0.7	0.3	124.82
0.8	0.13	0.4	0.6	3.00	0.7	0.3	122.13
0.8	0.13	0.3	0.7	3.00	0.7	0.3	118.04
0.8	0.13	0.2	0.8	3.00	0.7	0.3	112.84
0.8	0.13	0.1	0.9	3.00	0.7	0.3	106.76
0.9	0.14	0.9	0.1	3.50	0.7	0.3	128.08
0.9	0.14	0.8	0.2	3.50	0.7	0.3	143.04
0.9	0.14	0.7	0.3	3.50	0.7	0.3	149.22
0.9	0.14	0.6	0.4	3.50	0.7	0.3	149.44
0.9	0.14	0.5	0.5	3.50	0.7	0.3	145.62
0.9	0.14	0.4	0.6	3.50	0.7	0.3	139.07
0.9	0.14	0.3	0.7	3.50	0.7	0.3	130.68
0.9	0.14	0.2	0.8	3.50	0.7	0.3	121.10
0.9	0.14	0.1	0.9	3.50	0.7	0.3	110.76

Table 3.3 The percent relative efficiency of $(\hat{\pi}_G)_N$ with respect to $(\hat{\pi}_G)_P$.

π_{G1}	π_{G2}	w_1	w_2	r	θ_1^*	θ_2^*	PRE's
0.8	0.13	0.9	0.1	3.00	0.7	0.3	101.74
0.8	0.13	0.8	0.2	3.00	0.7	0.3	102.84
0.8	0.13	0.7	0.3	3.00	0.7	0.3	103.44
0.8	0.13	0.6	0.4	3.00	0.7	0.3	103.63
0.8	0.13	0.5	0.5	3.00	0.7	0.3	103.50
0.8	0.13	0.4	0.6	3.00	0.7	0.3	103.12
0.8	0.13	0.3	0.7	3.00	0.7	0.3	102.55
0.8	0.13	0.2	0.8	3.00	0.7	0.3	101.81
0.8	0.13	0.1	0.9	3.00	0.7	0.3	100.95
0.9	0.14	0.9	0.1	3.50	0.7	0.3	105.32
0.9	0.14	0.8	0.2	3.50	0.7	0.3	108.16
0.9	0.14	0.7	0.3	3.50	0.7	0.3	109.33
0.9	0.14	0.6	0.4	3.50	0.7	0.3	109.37
0.9	0.14	0.5	0.5	3.50	0.7	0.3	108.64
0.9	0.14	0.4	0.6	3.50	0.7	0.3	107.40
0.9	0.14	0.3	0.7	3.50	0.7	0.3	105.81
0.9	0.14	0.2	0.8	3.50	0.7	0.3	104.00
0.9	0.14	0.1	0.9	3.50	0.7	0.3	102.04

4. MODIFIED NEYMAN ALLOCATION

There is difficulty in using Neyman allocation, as the values of V_h in (2.10) will usually be unknown. However, the stratum variance V_h of stratum h may be obtained from previous surveys or from a specially planned pilot survey. The other alternative is to conduct the main survey in a phased manner and utilize the data in the first phase for ensuring better allocation in the second phase.
In case the value of the stratum variance V_h is not known at all. Then this problem can be resolved using the approaches of Searls (1967) and Singh and Mathur (2005) when the guessed value (or the crude prior estimate) of π_{Gh} is available. Hence, the gussed value (or the crude prior estimate) of V_h is available as V_h depends on π_{Gh}. Let V_h^* be the guessed or prior estimate of V_h. Then the value of n_h in (2.10) based on the guessed value $V_h^* = \tau_h V_h$ of V_h is given by

$$\frac{n_h^*}{n} = \frac{w_h\sqrt{V_h^*}}{\sum_{h=1}^{L} w_h\sqrt{V_h^*}} \tag{4.1}$$

where τ_h ($> 0\ \forall$ h=1,2,…,L) is departure from the true value of V_h. The modified allocation in (4.1) may be named as modified Neyman allocation. Putting (4.1) in (2.5), we get the variance of $\hat{\pi}_G$ under modified Neyman (MN) allocation as

$$V(\hat{\pi}_S)_{MN} = \sum_{h=1}^{L}\frac{w_h^2}{n_h^*}V_h \quad = \sum_{h=1}^{L}\frac{w_h^2 V_h}{n w_h\sqrt{V_h^*}}\sum_{h=1}^{L} w_h\sqrt{V_h^*}$$

$$= \frac{1}{n}\left(\sum_{h=1}^{L} w_h\left(\frac{V_h}{\sqrt{V_h^*}}\right)\right)\left(\sum_{h=1}^{L} w_h\sqrt{V_h^*}\right) = \frac{1}{n}\left(\sum_{h=1}^{L} w_h\left(\sqrt{\frac{V_h}{\tau_h}}\right)\right)\left(\sum_{h=1}^{L} w_h\sqrt{\tau_h V_h}\right) \tag{4.2}$$

4.1 Efficiency comparison of the proposed estimator $\hat{\pi}_G$ under modified Neyman allocation with that of the estimator $\hat{\pi}_G$ under proportional allocation

From (2.9) and (4.2), we have

$$n[V(\hat{\pi}_G)_P - V(\hat{\pi}_G)_{MN}] = \sum_{h=1}^{L} w_h \sqrt{V_h}\left(\sqrt{V_h} - \frac{1}{\sqrt{\tau_h}}\sum_{h=1}^{L} w_h \sqrt{\tau_h V_h}\right)$$

which is always positive if

$$\left[\sqrt{V_h} - \frac{1}{\sqrt{\tau_h}}\sum_{h=1}^{L} w_h \sqrt{\tau_h V_h}\right] > 0\,,\ \forall h = 1,2,\ldots,L\ .$$

(4.3)

In the case of two strata (i.e. L=2) and under the assumption $\tau_2 = \tau\tau_1$ (τ, being a constant), the expression (4.3) reduces to

$$n[V(\hat{\pi}_G)_P - V(\hat{\pi}_G)_{MN}] = \left[\sum_{h=1}^{2} w_h (1 - w_h) V_h - w_1 w_2 \sqrt{V_1 V_2}\,\frac{(1+\tau)}{\sqrt{\tau}}\right] > 0 \text{ if}$$

$$\frac{(1+\tau)}{\sqrt{\tau}} < \frac{\sum_{h=1}^{2} w_h (1 - w_h) V_h}{w_1 w_2 \sqrt{V_1 V_2}}$$

i.e. if $\dfrac{(1+\tau)}{\sqrt{\tau}} < \dfrac{(V_1 + V_2)}{\sqrt{V_1 V_2}}$

i.e. if $\left\{\dfrac{(1+\tau)}{\sqrt{\tau}} - 2\right\} < \left\{\dfrac{(V_1 + V_2)}{\sqrt{V_1 V_2}} - 2\right\}$

i.e. if $\left\{\dfrac{(1+\tau - 2\sqrt{\tau})}{\sqrt{\tau}}\right\} < \left\{\dfrac{(V_1 + V_2 - 2\sqrt{V_1 V_2})}{\sqrt{V_1 V_2}}\right\}$

i.e. if $\left\{\dfrac{(1+\tau - 2\sqrt{\tau})}{\sqrt{\tau}}\right\} < \left\{\dfrac{(\sqrt{V_1} - \sqrt{V_2})^2}{\sqrt{V_1 V_2}}\right\}$

i.e. if $\left\{\tau - \sqrt{\tau}\left[2 + \dfrac{(\sqrt{V_1} - \sqrt{V_2})^2}{\sqrt{V_1 V_2}}\right] + 1\right\} < 0$

i.e. if $\left\{\tau - \sqrt{\tau}(2+\phi) + 1\right\} < 0$

i.e. if $\left[1 + \dfrac{1}{2}\left\{\phi - \sqrt{\phi(\phi+4)}\right\}\right] < \sqrt{\tau} < \left[1 + \dfrac{1}{2}\left\{\phi + \sqrt{\phi(\phi+4)}\right\}\right]$

i.e. if $\left[1 + \dfrac{1}{2}\left\{\phi - \sqrt{\phi(\phi+4)}\right\}\right]^2 < \tau < \left[1 + \dfrac{1}{2}\left\{\phi + \sqrt{\phi(\phi+4)}\right\}\right]^2$,

(4.5)

where

$$\phi = \frac{\left(\sqrt{V_1} - \sqrt{V_2}\right)^2}{\sqrt{V_1 V_2}}\ . \tag{4.6}$$

Thus in the case of two strata (i.e. L=2) and under the assumption $\tau_2 = \tau\tau_1$, the suggested estimator $(\hat{\pi}_G)_{MN}$ with modified Neyman allocation is more efficient than the proposed estimator $(\hat{\pi}_G)_P$ under proportional allocation as long as the condition (4.5) is satisfied.

The range of τ can be calculated from (4.5) in which the proposed estimator $(\hat{\pi}_G)_{MN}$ with modified Neyman is more efficient than the proposed estimator $(\hat{\pi}_G)_P$ under proportional allocation, for different values of $\pi_{G1}, \pi_{G2}, \theta_1^*, \theta_2^*, r, w_1$ and w_2 .

4.2 Efficiency comparison of the proposed estimator $\hat{\pi}_G$ under modified Neyman allocation with that of the Hussain et al.'s (2014) estimator $\hat{\pi}_{G(P)}$ in stratified random sampling

We write the variance of the Hussain et al.'s (2014) estimator $\hat{\pi}_{G(P)}$ in stratified sampling (for L strata) as

$$V(\hat{\pi}_{G(P)}) = \frac{1}{n}\left[\pi_G(1-\pi_G) + \pi_G a + (1-\pi_G)b\right], \qquad (4.7)$$

[see Hussain et al.'s (2014), Theorem 3.2, equation (3.4)],

where $\pi_G = \sum_{h=1}^{L} w_h \pi_{Gh}$, a and b are same as defined earlier.

From (4.2) and (4.7), we have

$$n[V(\hat{\pi}_{G(P)}) - V(\hat{\pi}_G)_{MN}] = \left[\pi_G(1-\pi_G) + \pi_G a + (1-\pi_G)b - \left(\sum_{h=1}^{L} \frac{w_h\sqrt{V_h}}{\sqrt{\tau_h}}\right)\left(\sum_{h=1}^{L} w_h\sqrt{\tau_h V_h}\right)\right]$$

(4.8)

which is positive if

$$\left(\sum_{h=1}^{L} \frac{w_h\sqrt{V_h}}{\sqrt{\tau_h}}\right)\left(\sum_{h=1}^{L} w_h\sqrt{\tau_h V_h}\right) < \left[\pi_G(1-\pi_G) + \pi_G a + (1-\pi_G)b\right]$$

(4.9)

In the case of two strata (i.e. L=2) and under the assumption $\tau_2 = \tau\tau_1$, the expression in (4.8) reduces to:

$$n[V(\hat{\pi}_{G(P)}) - V(\hat{\pi}_G)_{MN}] = w_1 w_2\left[(\pi_{G1}-\pi_{G2})^2 + \left(\sqrt{V_1}-\sqrt{V_2}\right)^2 + \sqrt{V_1V_2}\left\{2 - \frac{(1+\tau)}{\sqrt{\tau}}\right\}\right]$$

$$= \frac{w_1 w_2\sqrt{V_1V_2}}{\sqrt{\tau}}\left[\frac{(\pi_{G1}-\pi_{G2})^2 + \left(\sqrt{V_1}-\sqrt{V_2}\right)^2\sqrt{\tau}}{\sqrt{V_1V_2}} + \left\{2\sqrt{\tau} - 1 - \tau\right\}\right]$$

$$= \frac{w_1 w_2\sqrt{V_1V_2}}{\sqrt{\tau}}\left[(\phi+2)\sqrt{\tau} - 1 - \tau\right]$$

which is non – negative if

$$\left\{1 + \tau - (\phi^* + 2)\sqrt{\tau}\right\} < 0$$

i.e. if $\left[1 + \frac{1}{2}\left\{\phi^* - \sqrt{\phi^*(\phi^*+4)}\right\}\right] < \sqrt{\tau} < \left[1 + \frac{1}{2}\left\{\phi^* + \sqrt{\phi^*(\phi^*+4)}\right\}\right]$

i.e. if $\left[1 + \frac{1}{2}\left\{\phi^* - \sqrt{\phi^*(\phi^*+4)}\right\}\right]^2 < \tau < \left[1 + \frac{1}{2}\left\{\phi^* + \sqrt{\phi^*(\phi^*+4)}\right\}\right]^2$, (4.10)

where $\phi^* = \dfrac{\left[(\pi_{G1}-\pi_{G2})^2 + \left(\sqrt{V_1}-\sqrt{V_2}\right)^2\right]}{\sqrt{V_1V_2}}$.

The range of τ can be calculated from (4.10) for different values of $\pi_{G1}, \pi_{G2}, \theta_1^*, , \theta_2^*, r, w_1$ and w_2, in which the proposed estimator $(\hat{\pi}_G)_{MN}$ with modified Neyman allocation is better than the Hussain et al.'s (2014) estimator $\hat{\pi}_{G(P)}$ in stratified random sampling.

5. NUMERICAL ILLUSTRATION

To examine the performance of the proposed estimator $(\hat{\pi}_G)_{MN}$ under modified Neyman allocation with that of the proposed estimator $\hat{\pi}_{G(P)}$ under proportional allocation and with the estimator $\hat{\pi}_{G(P)}$ due to Hussain et al.'s (2014) in stratified random sampling, we have computed the percent relative efficiencies of the proposed estimator $(\hat{\pi}_G)_{MN}$ with respect to $(\hat{\pi}_G)_P$ and $\hat{\pi}_{G(P)}$, by using the following formulae:

$$PRE((\hat{\pi}_G)_{MN}, (\hat{\pi}_G)_P) = \frac{(w_1V_1 + w_2V_2)}{w_1^2V_1 + w_2^2V_2 + w_1w_2\sqrt{V_1V_2}\,\frac{(1+\tau)}{\sqrt{\tau}}} \times 100, \quad (5.1)$$

$$PRE((\hat{\pi}_G)_{MN}, \hat{\pi}_{G(P)}) = \frac{(w_1\pi_{G1} + w_2\pi_{G2})(1 - w_1\pi_{G1} - w_2\pi_{G2}) + a(w_1\pi_{G1} + w_2\pi_{G2}) + b(1 - w_1\pi_{G1} - w_2\pi_{G2})}{w_1^2V_1 + w_2^2V_2 + w_1w_2\sqrt{V_1V_2}\,\frac{(1+\tau)}{\sqrt{\tau}}} \times 100 \quad (5.2)$$

for different values of $\pi_{G1}, \pi_{G2}, \theta_1^*, , \theta_2^*, r, w_1$ and w_2. Findings are shown in Tables 5.1 and 5.2. We have also computed the range of τ using the formulae (4.5) and (4.10) for the parametric values considered in Table 5.3.

It is observed from Tables 5.1 and 5.2 that there is enough scope of selecting the value of τ in order to get the efficiency larger than 100 percent i.e. the suggested estimator $\hat{\pi}_G$ with modified Neyman allocation is more efficient than the Hussain et al.'s (2014) estimator $\hat{\pi}_{G(P)}$ in stratified random sampling in certain range of τ. The length of the interval of τ in which the suggested estimator $\hat{\pi}_G$ with modified Neyman allocation is better than the Hussain et al.'s (2014) estimator $\hat{\pi}_{G(P)}$, depends on the actual values of π_{G1} and π_{G2}. Thus, the proposed estimator $\hat{\pi}_G$ with modified Neyman allocation is also recommended for its use in practice.

Table 5.1 The percent relative efficiency of $(\hat{\pi}_G)_{MN}$ with respect to $(\hat{\pi}_G)_P$

π_{G1}	π_{G2}	W_1	W_2	r	θ_1^*	θ_2^*	V_1	V_2	τ	τ_1	π^*_{G1}	π^*_{G2}	$PRE((\hat{\pi}_G)_N, (\hat{\pi}_G)_P)$	$PRE((\hat{\pi}_G)_{MN}, (\hat{\pi}_G)_P)$
	0.13	**0.90**	**0.10**	**3.00**	**0.70**	**0.30**	0.35	0.74	0.70	0.20	0.56	0.02		101.36
							0.35	0.74	0.90	0.20	0.72	0.02	**101.74**	101.71
0.8							0.35	0.74	**1.00**	0.20	0.80	0.03		**101.74**
	0.14	**0.90**	**0.10**	**3.50**	**0.70**	**0.30**	0.19	0.65	0.70	0.20	0.63	0.02		104.85
							0.19	0.65	0.90	0.20	0.81	0.03		105.28
0.9							0.19	0.65	**1.00**	0.20	0.90	0.03	**105.32**	**105.32**

Table 5.2 The percent relative efficiency of $(\hat{\pi}_G)_{MN}$ with respect to $(\hat{\pi}_{G(P)})$

π_{G1}	π_{G2}	W_1	W_2	r	θ_1^*	θ_2^*	V_1	V_2	τ	τ_1	π^*_{G1}	π^*_{G2}	$PRE((\hat{\pi}_G)_N, (\hat{\pi}_{G(P)}))$	$PRE((\hat{\pi}_G)_{MN}, (\hat{\pi}_{G(P)}))$
	0.13	**0.90**	**0.10**	**3.00**	**0.70**	**0.30**	0.35	0.74	0.70	0.20	0.56	0.02		111.93
							0.35	0.74	0.90	0.20	0.72	0.02	**112.36**	112.32
0.8							0.35	0.74	**1.00**	0.20	0.80	0.03		**112.36**
	0.14	**0.90**	**0.10**	**3.50**	**0.70**	**0.30**	0.19	0.65	0.70	0.20	0.63	0.02		127.51
							0.19	0.65	0.90	0.20	0.81	0.03	**128.08**	128.03
0.9							0.19	0.65	**1.00**	0.20	0.90	0.03		**128.08**

Table 5.3 Range of τ for the values of parameters considered for Tables 5.1 and 5.2.

π_{G1}	π_{G2}	W_1	W_2	r	θ_1^*	θ_2^*	V_1	V_2	τ lower range	τ upper range
	0.13	**0.90**	**0.10**	**3.00**	**0.70**	**0.30**	0.70	0.20	0.142	7.030
							0.90	0.20	0.142	7.030
0.8							1.00	0.20	0.142	7.030
	0.14	**0.90**	**0.10**	**3.50**	**0.70**	**0.30**	0.70	0.20	0.072	13.908
							0.90	0.20	0.072	13.908
0.9							1.00	0.20	0.072	13.908

7. DISCUSSION

This paper addresses the problem of estimating the population proportion of sensitive attribute π_G based on stratified sampling scheme. It has been shown theoretically as well as numerically that the proposed model is more efficient than the Hussain et al.'s (2014) randomized response model. The properties of the developed stratified randomized response model have been studied under proportional and Neyman allocations. The modified Neyman allocation has been developed when the crude prior estimates of V_1 and V_2 are available. A theoretical comparison is also given of the proposed estimator $\hat{\pi}_G$ under modified Neyman allocation with the Hussain et al.'s (2014) estimator $\hat{\pi}_{G(P)}$ in stratified random sampling. Numerical illustrations are also given in support of the present study.

RECEIVED JANUARY, 2015

REVISED JUNE, 2015

REFERENCES

[1] FOX J.A. and TRACY P.E. (1986): **Randomized Response: A method of Sensitive Surveys**.: SAGE Publication Newbury Park, CA s.

[2] HONG K., YUM J. and LEE H. (1994): A stratified randomized response technique. **Korean Jour. Appl. Statist**., 7, 141-147.

[3] KIM J. and WARDE W. (2004): A stratified Warner randomized response model. **Jour. Statist. Plan. Infer**., 120, 155-165.

[4] KIM J. and WARDE W.. (2005): A mixed randomized response model. **Jour. Statist. Plan. Infer**., 133, 211-221.

[5] KIM J.M. and ELAM M.E. (2005): A two – stage stratified Warner's randomized response model using optimal allocation . **Metrika**, 61, 1-7.

[6] KUK A.Y.C. (1990): Asking sensitive questions inderectely. **Biometrika** 77, 436-438.
[7] MANGAT N.S. and SINGH, R. (1990): An alternative randomized procedure. **Biometrika**, 77, 439-442.
[8] MANGAT N.S. (1994): An improved randomized response strategy**. Jour. Roy. Statist. Soc**., B, 56, 93–95.
[9] SEARLS D.T. (1967): A note on the use of an approximately known coefficient of variation. **The Ameri. Statist**. 21, 20-21.
[10] SHABBIR J. and GUPTA S. (2005): Optimal allocation in stratified randomized response model. **Pak. Jour. Statist. Oper. Res**., 1,15-22.
[11] SINGH H.P. and MATHUR N (2005): Estimation of the population mean when the coefficient of variation is known using scrambled response technique. **Jour. Statist. Plann. Infere**. 131, 135-144.
[12] SINGH H. P. and TARRAY T. A. (2012): A Stratified Unknown repeated trials in randomized response sampling. **Commun. Korean Statist. Soc**., 19, 751-759.
[13] SINGH H. P. and TARRAY T. A.. (2013): An alternative to Kim and Warde's mixed randomized response model. **Statist. Oper. Res.Trans**., 37, 189-210.
[14] SINGH H. P. and TARRAY T. A.. (2014); An alternative to stratified Kim and Warde's randomized response model using optimal (Neyman) allocation. **Model Assist. Statist. Appl**. , 9, 37-62.
[15] SINGH S. and SEDORY S.A. (2013): A new randomized response device for sensitive characteristics: an application of the negative hypergeometric distribution. Metron,71,3-8.
[16] WARNER S.L. (1965): Randomized response: A survey technique for eliminating evasive answer bias. **Jour. Amer. Statist. Assoc**.,60, 63-69.