

Research Article

## An alternative approach for construction of strata using quantified sensitivity level

**Dejen Agegnehu\***

Department of Basic Science, College of Forestry, Dr Y S Parmar University of Horticulture and Forestry, Nauni - 173230 (Himachal Pradesh), India

**P. K. Mahajan**

Department of Basic Science, College of Forestry, Dr Y S Parmar University of Horticulture and Forestry, Nauni - 173230 (Himachal Pradesh), India

**R. K. Gupta**

Department of Basic Science, College of Forestry, Dr Y S Parmar University of Horticulture and Forestry, Nauni - 173230 (Himachal Pradesh), India

\*Corresponding author. Email: dej.agegne@gmail.com

### Article Info

<https://doi.org/10.31018/jans.v13i3.2679>

Received: April 26, 2021

Revised: July 25, 2021

Accepted: July 30, 2021

### How to Cite

Agegnehu, D. *et al.* (2021). An alternative approach for construction of strata using quantified sensitivity level. *Journal of Applied and Natural Science*, 13(3), 807 - 814. <https://doi.org/10.31018/jans.v13i3.2679>

### Abstract

The study is investigated on an alternative method for the construction of strata using sensitivity level when the samples are selected with simple random sampling with replacement (SRSWR) and the data are collected by scrambled optional randomization technique on the sensitive characters. Thus, the optional randomized response model  $Z = Y \cdot S^k$ , where  $k$  is a random variable having value 1 if the response is scrambled and 0 otherwise, was considered for finding out Approximate Optimum Strata Boundaries by minimizing the variance of the estimator  $\bar{y}_{st}$ . The cum.  $\sqrt[3]{D_2(x) f(x)}$  rule was proposed for finding out Approximate Optimum Strata Boundary in Neyman allocation for the optional scrambled response. This is applicable for wider classes of sampling design and estimators in stratification. The proposed rule on optional scrambled randomized response is efficient and can be used effectively for the construction of optimum strata boundary via Rectangular, Right triangular and Exponential distribution.

**Keywords:** Approximate optimum strata boundary, Minimal equation, Neyman allocation, Optional randomization, Scrambled response, Sensitivity level

### INTRODUCTION

In stratified random sampling, one of the main problems to gain the precision of the estimates is the determination for the construction of strata. Particularly, survey sampling on human populations in the non-sensitive topics usually receive a good response, while surveying with direct questions on sensitive topics such as tax dodging, induced abortion, taking the illegal drug and so on are often results either refusal to respond or falsification of the answer. To overcome this difficulty and ensure confidentiality for respondent, Warner (1965) initiated a randomized response technique (RRT) to procure trustworthy data. Since the introduction of Warner (1965) on estimating the proportion of people with a sensitive characteristic, randomized response

model has been extensively studied by several researchers. Among of them, one can refers to, Greenberg *et al.* (1971), Chaudhuri and Muherejee (1987), Chaudhuri and Chritofides (2013), Singh and Tarry (2014), Fox (2016), Arnab *et al.* (2017), and Kumar *et al.* (2020).

In the optional randomized response technique (ORRT), most of the people feel sensitive, but some are willing to answer directly Gupta *et al.* (2002) introduced the optional randomization technique, where the respondent can give a direct response  $Y$  or a scrambled response  $Y \cdot S$  with probability  $p$ . Even though the survey question tells us, there was a relative sensitivity compared to the other, but some respondents had willing to respond to the correct answer without scrambling. Therefore, the proposing of Gupta *et al.* (2002)

could it be determined and estimate the sensitivity level of the question using the multiplicative optional randomized model and it expressed as  $z_h = y_h \cdot s_h^k$  where  $k$  is a Bernoulli variate taking values 1 if the response is scrambled and 0 otherwise. Under Bernoulli distribution, let  $E(k) = p$  where  $p$  is the sensitivity level for the survey question. So that, the value of  $p$  is close to one indicates that the survey question is more sensitive, whereas the value of  $p$  is close to zero shows the survey question become less sensitive. The present study is used the Gupta et al. (2002) methods for collecting the data from different strata with SRSWR. Arnab (2004) was considered an alternative optional randomized response technique. Recently Singh and Gorey (2019) reviewed Gupta et al. (2002) model by suggesting a modified optional randomized response model. For more detail, see further work which is done by Gjestvang and Singh (2006), Tarray et al. (2015, 2017).

Dalenius (1950) first introduced the methods for constructing optimum strata boundaries (OSB) when the study variable itself was used as a stratification variable. Since there is difficulties on determining the exact point of optimum stratification in randomized response techniques, a lot of intensive work was made for obtaining approximate optimum solution (AOSB) to this problem. Mahajan et al. (1994) proposed the  $\sqrt[3]{P_1(\bar{x})}$  rule to obtain AOSB, where the sample was selected with SRSWR from different strata. The interviewee responded the multiplicative scrambled answer  $Z = Y \cdot S$ , where  $Y$  is sensitive variable and  $S$  is scrambled variable. In the present study, we attempt and proposed the methods for determining optimum stratification for scrambled optional randomized response techniques when sample in each stratum are selected with SRSWR and we collect the data using Gupta et al. (2002) methods. The study also has an approach to find minimal equations and limiting the form of the variance for the optional randomized response with Neyman allocation methods. Therefore, this study aimed to propose the rule for obtaining approximate optimum strata boundaries (AOSB) with the set of the minimal equation and limiting the expression of the variance. Furthermore, this study has an attempt and a comparison of percentage relative efficiency of stratification of the proposed rules over no stratification.

**MATERIALS AND METHODS**

Let a population of  $N$  units be divided into  $L$  strata, the  $h^{th}$  stratum containing  $N_h$  units with a total of  $Y_h$  for the character  $y$ . Thus, within each stratum, a random sample of size ' $n_h$ ' in the  $h^{th}$  stratum is selected so that  $\sum_{h=1}^L n_h = n$ . Sampling in one stratum is independent of that in another. For  $h^{th}$  stratum, let  $y_h$  denotes the value of the sensitive character and  $S_h$  be a scam-

bled random variable which is independent of  $Y_h$  and has a finite mean and variance.

In the proposed optional procedure, we have assumed that both  $S$  and  $Y$  are positive valued random variables and the model can be written as

$$Z_h = Y_h \cdot S_h^k \tag{1}$$

where  $k$  is random variable defined as

$$k = \begin{cases} 1 & \text{if the response is scrambled} \\ 0 & \text{otherwise} \end{cases}$$

In Gupta et al. (2002),  $k$  is a variate in Bernoulli distribution with  $E(k) = p$  and the probability that a person will report the scrambled response (or it called as sensitivity level) measured by  $p$ . However, more people have a scrambled response in a survey question which means the value of  $p$  approaches to one and the question become more sensitive. While, the less people report of survey response indicates is that the question is less sensitive and the value of  $p$  tends-to zero. Thus, the value of  $p$  could measure the level of sensitivity for the personal interview surveys question.

For theoretical approach, let us assume that,

$$E(S_h) = \vartheta_h = 1, \quad V(S_h) = \Lambda_h, \quad E(Y_h) = \mu_{hy} \text{ and } V(Y_h) = \sigma_{hy}^2.$$

In addition, coefficient of variation 
$$C_h = \frac{\sqrt{\Lambda_h}}{\vartheta_h}$$

Where  $\vartheta_h$  and  $\Lambda_h$  are known to the interviewer but  $\mu_{hy}$  and  $\sigma_{hy}^2$  are unknown. The particular values of  $S_h$  are unknown to the interviewer but its distribution is known. In this way the respondent's privacy is not violated.

Since  $Y_h$  and  $S_h$  are independent, then we have

$$E(Z_h) = \mu_{hy}$$

$$V(Z_h) = \sigma_{hy}^2 + \Lambda_h P_h (\sigma_{hy}^2 + \mu_{hy}^2) \tag{2}$$

If  $z_{hi}$  denotes the value of the scrambled variable  $z$  for  $i^{th}$  unit of the sample in  $h^{th}$  stratum and sampling within each stratum is SRSWR. Mahajan et al. (1994) showed that an unbiased estimator of the population mean

$$\mu = \sum_{h=1}^L W_h \mu_{hy} \text{ is}$$

$$\hat{\mu}_{st} = \sum_{h=1}^L W_h \hat{\mu}_{hy} \tag{3}$$

Where  $W_h$  is the proportion of units in the  $h^{th}$  stratum, and the variance obtained as

$$V(\hat{\mu}_{st}) = \sum_{i=1}^L W_h^2 \frac{1}{n_h} \{ \sigma_{hy}^2 (1 + \theta) + \theta \mu_{hy}^2 \}$$

where

$$\theta = \Lambda_h p \tag{4}$$

For instance, the sensitive study variable  $y$  (e.g., income understated in the time of return) and non-sensitive stratification variable  $x$  (e.g., eye estimated

value for the person) related as the regression of y on x in the population be of the form

$$y = \eta(x) + e \tag{5}$$

Where  $\eta(x)$  is a function of x and e is the error term such that  $E(e/x) = 0$  and  $V(e/x) = \omega(x) > 0$  for all x in the range (a, b) of x with  $(b - a) < \infty$ . If  $f(x)$  denotes the marginal density of x, then we have

$$W_h = \int_{x_{h-1}}^{x_h} f(x) dx; \mu_{hy} = \mu_{h\eta} = \frac{1}{W_h} \int_{x_{h-1}}^{x_h} \eta(x) f(x) dx$$

$$\text{and } \sigma_{hy}^2 = \sigma_{h\eta}^2 + \mu_{h\omega} \tag{6}$$

The series expansion for  $W_h$ ,  $\mu_{h\eta}$ ,  $\sigma_{h\eta}^2$  and  $\mu_{h\omega}$  can be obtained by using Taylor's theorem about both the upper and lower boundaries of h-th stratum on the lines of Singh and Sukhatme (1969) as well as Singh & Prakash (1975). For this purpose, we impose certain regularity conditions on the functions  $f(x)$ ,  $\eta(x)$  and  $\omega(x)$ . A function  $\tau(x)$  belongs to the class  $\Psi$  of functions if it satisfies the following conditions:

$$0 < \tau(x)$$

$$\tau(x) < \infty$$

$\tau(x)$ ,  $\tau'(x)$  and  $\tau''(x)$  exists and are continuous for all x in (a, b) where  $(b - a) < \infty$

Under this assumption, the stratification variable x is continuous in the super population and  $f(x)$  denotes probability density function of x and  $(x_{h-1}, x_h)$  are respectively the lower and upper boundaries of h<sup>th</sup> stratum. Using Taylor's series expansions of the function  $f(x)$ , Singh and Sukhatme (1969) have shown that the mean of the function  $\omega(x)$  in the h-th and i -th stratum are given by

$$\mu_{h\omega} = \omega \left[ 1 - \frac{\omega'}{2\omega} k_h + \frac{\omega' f' + 2f\omega''}{12f\omega} k_h^2 - \frac{ff''\omega' + ff'\omega'' + f^2\omega''' - f'^2\omega'}{24f^2\omega} k_h^3 + 0(k_h^4) \right] \tag{7}$$

$$\mu_{i\omega} = \omega \left[ 1 + \frac{\omega'}{2\omega} k_i + \frac{\omega' f' + 2f\omega''}{12f\omega} k_i^2 + \frac{ff''\omega' + ff'\omega'' + f^2\omega''' - f'^2\omega'}{24f^2\omega} k_i^3 + 0(k_i^4) \right] \tag{8}$$

And the partial derivative of  $\sigma_{h\eta}^2$  in the h-th stratum given that

$$\sigma_{h\eta}^2 = \frac{k_h^2}{12} [\eta'^2 - \eta' \eta'' k_i + 0(k_i^2)] \tag{9}$$

However, all the functions and their derivatives are evaluated about the common boundary  $x_h$ , whereas,  $\mu_{i\eta}$  can be obtained from  $\mu_{i\omega}$  by putting in  $\eta(x)$  place of  $\omega(x)$ . Similarly, the expression of  $\mu_{h\omega}$  can be used for obtaining  $\mu_{h\eta}$

Also, for obtaining the approximate solution of the minimal variance the expansion of  $\sqrt{\lambda f(y)}$  about the point  $t = y$  can be defined as

$$\begin{aligned} \left[ \int_y^x \sqrt{\lambda f(t)} dt \right]^\lambda &= \left[ \int_y^x \left\{ \sqrt{\lambda f(y)} + \frac{t-y}{\lambda} \frac{f'(y)}{f(y)} + 0(t-y)^2 \right\} dt \right]^\lambda \\ &= k^\lambda f(y) \left[ 1 + \frac{k}{2} \frac{f'(y)}{f(y)} + 0(k^2) \right] \end{aligned} \tag{10}$$

$$= k^{\lambda-1} \int_y^x f(t) dt [1 + 0(k^2)] \tag{11}$$

Where,  $k_i = x_{h+1} - x_h$  and

$$k_h = x_h - x_{h-1}$$

The effectiveness of the proposed method of finding the set [x] of AOSB has been explored numerically. For this purpose, the present study considers the following three common distribution of the auxiliary variable x, which are defined as

$$\begin{aligned} \text{Rectangular Distribution } f(x) &= 1 && [1,2] \\ \text{Right triangular Distribution } f(x) &= 2(2-x) && [1,2] \\ \text{Exponential Distribution } f(x) &= e^{-x+1} && [1, \infty) \end{aligned} \tag{12}$$

These density function of x has in some extent represent those usually encountered in practice. However, rectangular distribution is easiest to handle and right triangular density function gives a somewhat skewed distribution having a finite range, but the exponential distribution has infinite range. The study for the auxiliary variable x has been used one set of ranges. Thus, the multiplication of a constant with the auxiliary variable x will have the same distribution with different range values on the transformed variable. Moreover, this multiplication by a constant, however, do not change the basic conclusions obtained from the examination as it only amounts to measuring x on a different scale.

## RESULTS AND DISCUSSION

This section has been devoted to obtaining the variance of the estimator  $\hat{\mu}_{st}$ . Then, the proposed rule is found for obtaining the approximate optimum strata boundary. and the minimal solution, as well as we obtain the approximate and limiting expression for the variance. The relative efficiency for the proposed rule with numerical illustration is also discussed via rectangular, right triangular and exponential distribution.

In Neyman allocation methods, the allocation of sample size among different strata is based on a joint consideration of the stratum and the stratum variation. However, here we consider that the sampling cost per unit among different strata is the same and the size of the sample is fixed. However, we select  $n_h$  and define as

$$n_h = n \frac{w_h P_{hy}}{\sum_{h=1}^L w_h P_{hy}}$$

where

$$P_{hy} = \sqrt{(\sigma_{h\eta}^2 + \mu_{h\omega})(1 + \theta) + \theta \mu_{h\eta}^2} \tag{13}$$

Therefore, in relation to equation (5) and (13), the variance of the estimator  $\hat{\mu}_{st}$  expressed in equation (4) under Neyman allocation method again obtained as

$$V(\hat{\mu}_{st})_N = \frac{1}{n} \left\{ \sum_{h=1}^L W_h \sqrt{(\sigma_{h\eta}^2 + \mu_{h\omega})(1 + \theta) + \theta \mu_{h\eta}^2} \right\}^2 \tag{14}$$

The variance given in expression (14) is clearly shows that it is the function of the strata boundary on the auxiliary variable x. the variance can, therefore, further reduced by using the optimum strata boundary which correspond to the minimum of the variance in (14).

**Obtaining the minimal solution and their approximate expression**

For the purpose of obtaining minimal equations, we consider the approximate optimum points of stratification  $[x_h]$  on the range (a, b) of x. Then corresponding to these strata boundaries as determined by the optimum points of stratification  $[x_h]$ , we shall consider the variance of the estimator  $\hat{\mu}_{st}$  is minimum. However, the solutions which give the optimum points of stratification  $[x_h]$  for the minimal equations are obtained by the partial derivatives of  $V(\hat{\mu}_{st})_N$  in (14) with respect to  $[x_h]$ .

$$w_h \frac{\partial}{\partial x_h} \sqrt{(b)} + \sqrt{(b)} \frac{\partial}{\partial x_h} w_h = -w_i \frac{\partial}{\partial x_h} \sqrt{(d)} - \sqrt{(d)} \frac{\partial}{\partial x_h} w_i \tag{15}$$

where,  $i = h + 1, h = 1, 2, \dots, L - 1$

$$(b) = (\sigma_{h\eta}^2 + \mu_{h\omega})(1 + \theta) + \theta \mu_{h\eta}^2$$

$$(d) = (\sigma_{i\eta}^2 + \mu_{i\omega})(1 + \theta) + \theta \mu_{i\eta}^2$$

The expressions of the partial derivatives terms involved in (14) can be easily obtained on the lines of Singh and Sukhatme (1969). On inserting the values of the respective partial derivatives of the parameter in equation (15) and solving the required minimal equations as

$$\frac{(1+\theta)[(\eta(x_h)-\mu_{h\eta})^2 - \sigma_{h\eta}^2 + \phi(x_h)-\mu_{h\omega}] + 2\theta\mu_{h\eta}[\eta(x_h)-\mu_{h\eta}]}{\sqrt{(\sigma_{h\eta}^2 + \mu_{h\omega})(1+\theta) + \theta\mu_{h\eta}^2}} = \frac{(1+\theta)[(\eta(x_h)-\mu_{i\eta})^2 - \sigma_{i\eta}^2 + \phi(x_h)-\mu_{i\omega}] + 2\theta\mu_{i\eta}[\eta(x_h)-\mu_{i\eta}]}{\sqrt{(\sigma_{i\eta}^2 + \mu_{i\omega})(1+\theta) + \theta\mu_{i\eta}^2}} \tag{16}$$

Although the minimal equations obtained in (16) gives the approximate optimum strata boundaries  $[x_h]$  with Neyman allocation, but the parameter involved in equation (16) are a function of  $[x_h]$  themselves. However, it is difficult or rather not possible to obtained optimum points of stratification  $[x_h]$ , which corresponds to the minimum of the variance of the estimator  $\hat{\mu}_{st}$ . If and only if, it is desirable to find the approximate solution to this system of equation (16).

Now we expand the minimal equation (16) in both sides using the relation expressed in (7) to (9) at common

boundary points  $[x_h]$  for the h-th and i-th strata, then, it gives the approximate expression of the minimal equation of the variance.

Using the relation (7), (8) and (9), the system of equations (16) giving optimum points of stratification can, therefore, be reduced into

$$k_h^3 \left[ 1 - \frac{1}{2} \frac{N'_3(x_h)}{N_3(x_h)} k_h + 0(k_h^2) \right] = k_i^3 \left[ 1 + \frac{1}{2} \frac{N'_3(x_h)}{N_3(x_h)} k_i + 0(k_i^2) \right] \tag{17}$$

where

$$N_3(x_h) = \frac{f(x_h)\omega_1(x_h)^2 + 4f(x_h)\eta'(x_h)^2\omega^*(x_h)}{\omega^*(x_h)^{3/2}} \tag{18}$$

$$\omega^* = \omega + \theta\omega + \theta\eta^2 \tag{19}$$

$$\omega_1 = \frac{\partial \omega^*}{\partial x_h} = \omega' + \theta\omega' + 2\theta\eta\eta' \tag{20}$$

Now proceeding on the lines of Singh and Sukhatme (1969) and using the relation expressed in (17), equivalently, the system of equations (16) can also be put as

$$k_h^2 \int_{x_{h-1}}^{x_h} N_3(x_h) dx [1 + 0(k_h^2)] = k_i^2 \int_{x_h}^{x_{h+1}} N_3(x_h) dx [1 + 0(k_i^2)] \tag{21}$$

Therefore, if we have a large number of strata so that the strata width  $k_h$  are small and their higher powers in the expansion can be neglected, then the system of equations in (16) or equivalently the system of equations in (21) can be approximated as

$$k_h^2 \int_{x_{h-1}}^{x_h} N_3(x_h) dx = constant(say \mathfrak{S}_2) \tag{22}$$

If the terms of order  $O(m^4)$ ,  $m = \text{Sup}_{(a, b)}(K_h)$ , have been neglected on both sides of equation (21), since  $\int_{x_{h-1}}^{x_h} N_3(x) dx = O(m)$  in view of the fact that  $N_3(x)$  is bounded for all x in (a, b).

This method of finding the approximate optimum strata boundaries (AOSB) for Neyman allocation method shall be called the cumulative cube root rule denoted as  $\text{cum.} \sqrt[3]{D_2(x) f(x)}$  rule.

Cum.  $\sqrt[3]{D_2(x) f(x)}$  Rule

If the function  $N_3(x) = D_2(x) f(x)$  is bounded and possesses its first two derivatives for all x in (a, b) with,  $(b - a) < \infty$  then for a given value of L taking equal intervals on the cumulative of  $\sqrt[3]{D_2(x) f(x)}$  yields approximately optimum strata boundaries (AOSB) for Neyman allocation method.

**Limiting Expression for the Variance**

The limiting form for the variance is particularly important in optimum stratification as it gives an insight into the manner in which the variance of the estimator of the mean is reduced as the number of strata increases. For obtaining limiting expression of variance, as  $V(\hat{\mu}_{st})_N$  given by (14), we devoted the following lemma, which can be proved by using the series expansion of the various terms involved in it.

Lemma 4.1. For the  $i$ -th stratum, if  $(x_{h+1}, x_h)$  are the boundaries of the  $i$ -th stratum and  $k_i = x_{h+1} - x_h$ , then

$$w_i \sqrt{(\sigma_{i\eta}^2 + \mu_{i\omega})(1 + \theta) + \theta \mu_{i\eta}^2} - \int_{x_h}^{x_{h+1}} \sqrt{\omega^*(x)} f(x) dx$$

$$= \frac{1}{96} \left[ \int_{x_h}^{x_{h+1}} \sqrt[3]{D_2(x) f(x)} dx \right]^3 [1 + O(k_i^2)] \tag{23}$$

Where  $\omega^* = \omega + \theta\omega + \theta\eta^2$  and

$$D_2(x) = \frac{\omega_1^2(x) + 4\eta^2(x)\omega^*(x)}{\omega^*(x)^{3/2}} \tag{24}$$

Proof: Applying the series expansions in powers of interval width  $k_i = x_{h+1} - x_h$  from the relations given in (7) to (10) and on using the Taylor's theorem for the expansion of the integrand about point  $x = x_h$  the L.H.S. of expression (23) can be reduce to

$$\frac{k_i^2}{96} \left\{ D_2(x)f(x)k_i + \frac{\partial}{\partial x_h} [D_2(x)f(x)] \frac{k_i^2}{2} + O(k_i^3) \right\}$$

$$= \frac{k_i^2}{96} \int_{x_h}^{x_{h+1}} D_2(x)f(x)dx [1 + O(k_i^2)]$$

$$= \frac{1}{96} \left[ \int_{x_h}^{x_{h+1}} \sqrt[3]{D_2(x)f(x)} dx \right]^3 [1 + O(k_i^3)] \tag{25}$$

Hence the lemma is proved.

Now using the Lemma 4.1 in the expression of  $V(\hat{\mu}_{st})_N$ , the limiting expression for th variance can be obtained as

$$V(\hat{\mu}_{st})_N = \frac{1}{n} [\gamma + \frac{\delta}{L^2}]^2 \tag{26}$$

where

$$\gamma = \int_a^b \sqrt{\omega^*(x)} f(x) dx \tag{27}$$

$$\delta = \frac{1}{96} \left\{ \int_a^b \sqrt[3]{D_2(x) f(x)} dx \right\}^3 \tag{28}$$

**Theorem 4.1**

For the approximate optimum strata boundaries obtained from the cum.  $\sqrt[3]{D_2(x) f(x)}$  rule, therefore, taking  $L \rightarrow \infty$  and the limiting expression for the variance defined as

$$\lim_{L \rightarrow \infty} V(\hat{\mu}_{st})_N = \frac{\gamma^2}{n} \tag{29}$$

where

$$\gamma = \int_a^b \sqrt{\omega^*(x)} f(x) dx$$

The theorem 4.1 in (29) gives the exact manner for which the variance  $V(\hat{\mu}_{st})_N$  in (26) would approach to  $\frac{\gamma^2}{n}$  as the value of  $L$  is increased.

**Empirical approach**

The numerical analysis is devoted for the purpose of the usefulness of stratification for simple random sam-

pling and with a replacement scheme, the study considers three density function of  $x$ , namely rectangular, right triangular and exponential distribution (see expression 12).

For the sake of simplicity, the regression function is assumed to be linear with slope 45 degree, so the regression line takes the form  $y = c + x + e$ . The conditional variance of the error term  $V(e/x) = \omega(x)$  is assumed to be of the form  $\omega(x) = dx^g$ , where  $d > 0$  and  $g$  are constants. The values of 'g' are considered to be 1 and 2. The value of constant 'd' is computed by the following formula:

$$d = \frac{\sigma_x^2 (1 - \rho^2)}{\rho^2 E(x^g)}$$

Where  $\sigma_x^2$  is the variance of  $x$

The value of 'd' depends upon the different values of  $g$  and the squared correlation coefficient ( $\rho^2$ ). The different values of ( $\rho^2$ ) are taken as 0.90, 0.70 and 0.50. In addition, there is constant  $\theta$  which is the multiplication of scrambled variance ( $\Lambda_h$ ) and the probability that the person would report the scrambled response ( $p$ ). Hence, this multiplication is defined as  $\theta = 0.04 * p$ , where 0.04 is the value of the scrambled variance. While, based on Gupta et al. (2002) values of  $p$  (sensitivity level) are considered as 0.1 (lowest), 0.5 (medium), and 0.9 (high).

However, to calculate Approximate Optimum Strata Boundary (AOSB), the distribution of auxiliary variable  $x$  is divided into 10 classes which have equal intervals on the range of  $x$ . Then, for each distribution and using the different values of the number of strata 'L', the solutions of AOSB through the proposed cum.  $\sqrt[3]{D_2(x) f(x)}$  rule has been obtained by taking equal intervals.

The empirical review given in Tables (1), (2) and (3) shows that the relative efficiency of optimum stratification over no stratification via. rectangular, right triangular and exponential distributions, respectively. In addition, it gives a comparison of relative efficiency between Mahajan et al. (1994) (Cum.  $\sqrt[3]{P_1(x)}$ ) and proposed (cum.  $\sqrt[3]{D_2(x) f(x)}$ ) rule using different values of  $g$  (1 and 2), number of strata ( $L=1,2,3$ , and 4), sensitivity level ( $p=0.1, 0.5$ , and 0.9), and correlation coefficient ( $\rho^2 = 0.9$ ).

Based on the rule for computing the AOSB, it requires a finite range for the variable  $x$ . However, the exponential density function was truncated at  $x = 6$ . Besides to this, the probability values for variable  $x$  after the truncation point are very small.

The empirical analysis in Tables (1), (2) and (3) shown that the relative efficiency of the proposed rules with correlation coefficient  $\rho^2 = 0.9$  are increases with the increase in number of strata via. Rectangular, Right triangular and Exponential distribution. However, in Mahajan et al. (1994) cum.  $\sqrt[3]{P_1(x)}$  rule displayed that

the gain in relative efficiency varies from 53.86 up to 76.97 percent for  $g = 1$  and 2 with rectangular distribution, which were not considered the sensitivity level ( $p$ ). Likewise, the proposed rule in present study ( $\text{cum. } \sqrt[3]{D_2(x) f(x)}$ ) reveals that, the gain in relative efficiency varies from 47.63 up to 304.01 percent when the sensitivity level ( $p$ ) goes from 0.1 to 0.9 with rectangular distribution. Moreover, the increasing of relative efficiency is slow with the increasing of the value of  $g$  for all considered density function. Whereas, from the proposed rule we can show that the gain of relative efficiency decreases with increasing of the sensitivity level ( $p$ ). From table 2 & 3 showed that, the maximum gain in relative efficiency for Mahajan et al. (1994) rule was obtained 65.84 and 261.57 % with right triangular and exponential distribution, respectively. Keep in mind the sensitivity level (optional randomization), the numerical analysis given from table 2 & 3 reveals the maximum gain in relative efficiency for the proposed rule is found to be 303.03 and 507.59 % with right trian-

gular and exponential distribution, respectively. Therefore, the gain of relative efficiency obtained from table 1, 2, & 3 showed that it is convenient and efficient to use the proposed rule for construction of approximate optimum strata boundary for the scrambled optional randomized response. The proposed rule is more efficient than the rule proposed by Odumade and Singh (2009) as a special case  $k = 2$ .

Chandel et al. (2016) considers the problem of optimum stratification with an additive model of scrambled randomized response technique and cumulative cube root rule of finding approximately optimum strata boundaries (AOSB) has been proposed. Moreover, Verma et al. (2012) reviews the problem of optimum stratification for two sensitive quantitative variables when data on the sensitive variables are collected by scrambled randomized response techniques and cumulative cube root rule has been proposed. However, in both Chandel et al. (2016) and Verma et al. (2012) proposed rules do not consider the sensitivity level of the

**Table 1.** Variance and percentage relative efficiency using Rectangular distribution for Neyman allocation method.

g	Strata (L)	Cum. $\sqrt[3]{P_1(x)}$ rule		Cum. $\sqrt[3]{D_2(x) f(x)}$ rule	
		Efficiency (without p)	Efficiency (with p=0.1)	Efficiency (with p=0.5)	Efficiency (with p=0.9)
1	1	100.00	100.00	100.00	100.00
	2	153.12	249.68	171.65	147.63
	3	170.06	346.29	198.27	162.13
	4	176.97	400.71	209.71	167.92
2	1	100.00	100.00	100.00	100.00
	2	153.38	250.32	171.99	147.84
	3	170.47	348.34	198.87	162.45
	4	177.45	404.01	210.44	168.30

**Table 2.** Variance and percentage Relative efficiency using Right triangular distribution for Neyman allocation method.

g	Strata (L)	Cum. $\sqrt[3]{P_1(x)}$ rule		Cum. $\sqrt[3]{D_2(x) f(x)}$ rule	
		Efficiency (without p)	Efficiency (with p=0.1)	Efficiency (with p=0.5)	Efficiency (with p=0.9)
1	1	100.00	100.00	100.00	100.00
	2	144.95	241.56	171.72	148.51
	3	159.17	339.90	201.15	164.81
	4	164.94	399.01	214.52	171.67
2	1	100.00	100.00	100.00	100.00
	2	145.40	243.01	172.16	148.76
	3	159.78	342.84	201.84	165.18
	4	165.84	403.03	215.34	172.08

**Table 3.** Variance and percentage relative efficiency using Exponential distribution for Neyman allocation method.

g	Strata (L)	Cum. $\sqrt[3]{P_1(x)}$ rule		Cum. $\sqrt[3]{D_2(x) f(x)}$ rule	
		Efficiency (without p)	Efficiency (with p=0.1)	Efficiency (with p=0.5)	Efficiency (with p=0.9)
1	1	100.00	100.00	100.00	100.00
	2	227.28	273.86	248.78	230.11
	3	304.01	435.37	362.54	316.61
	4	344.65	555.28	435.14	366.87
2	1	100.00	100.00	100.00	100.00
	2	233.61	286.87	257.61	236.91
	3	316.83	468.32	382.84	330.98
	4	361.57	607.59	464.83	386.74

question and they comprise indirect survey methods only for the respondent, while the present study proposes a rule for stratification and sampling designed by considering the sensitivity level as well as both direct and indirect methods of survey (through optional randomized method). Therefore, the proposed rule is efficient and applicable for constructing optimum strata boundary in the different fields that are assessing to applied sampling design with stratified sampling.

**Conclusion**

The proposed rule has a simultaneous advantage of building optimum approximate strata boundary and measuring the sensitivity level using the optional randomization model at minimal approximate variance. Therefore, we can conclude that the proposed cum.  $\sqrt[3]{D_2(x)f(x)}$  rule of optimal stratification with the optional scrambled randomized response is efficient and can be used effectively via. Rectangular, Right triangular and Exponential distribution.

**ACKNOWLEDGEMENTS**

The authors wish to thank the Associate Editor for a suggestion that helped improve the paper's presentation. They are also thankful to the referees for their comments.

**Conflict of interest**

The authors declare that they have no conflict of interest.

**REFERENCES**

1. Arnab, R. (2004). Optional randomized response techniques for complex survey designs. *Biom. J.*, 46(1), 114–124. doi.org/10.1002/bimj.200210006

2. Arnab, R., Shangodoyin, D.K. & Kgosi P.M. (2017). Randomized response techniques with multiple responses. *J. Ind. Soc. Agril. Statist.*, 71(3), 201–205.

3. Chandel, A., Mahajan, P. K. & Baharti (2016). Minimum variance stratification using scrambled randomised response. *International Journal of Agricultural and Statistical Sciences*, 12(1), 53-58.

4. Chaudhuri, A. & Mukherjee, R. (1987). Randomized response techniques: A review. *Statistica Neerlandica*, 41 (1), 27-44. doi.org/10.1111/j.1467-9574.1987.tb01169.x

5. Chaudhuri, A. & Christofides, T C. 2013. Indirect Questioning in Sample Surveys. Heidelberg: *Springer*. doi.org/10.1007/978-3-642-36276-7

6. Dalenius T. (1950). The problem of optimum stratification. *Skand. Aktuarietid Skrift*, 33, 203-213. doi.org/10.1080/03461238.1950.10432042

7. Fox, J. A. 2016. Randomized Response and Related Methods. Second Edition. California: *SAGE Publications*. doi.org/10.4135/9781506300122

8. Gjestvang, C.R. & Singh, S. (2006). A new randomized response model. *Journal of The Royal Statistical Society B*, 68, 523–530.

9. Greenberg, B.G., Kuelber, R. R., Abernathy, J. R. & Horvitz, D. G. (1971). Application of the randomized response technique in obtaining quantitative data. *Journal of American Statistical Association*, 66, 243-250.

10. Gupta, S., Gupta, B. & Singh, S. (2002). Estimation of sensitivity level of personal interview survey questions. *J. Statist. Plann. Inference*, 100, 239–247. doi.org/10.1016/S0378-3758(01)00137-9

11. Kumar A, Vishwakarma G. K. & Singh G. N. (2020). An improved randomized response model for simultaneous estimation of means of two quantitative sensitive variables. *Comm. Statist. –simulation and computations*, 1-21. doi.org/10.1080/03610918.2020.1788587

12. Mahajan, P.K., Gupta, J.P. & Singh, R. (1994). Determination of optimum strata boundaries for scrambled randomized response. *Statistica Anno*, 54(3), 376-381. doi.org/10.6092/issn.1973-2201/1023

13. Singh, R. & Prakash, D. (1975). Optimum stratification for equal allocation. *Annals of Institute of Statistical Mathematics*, 27, 273-280.

14. Singh, R. & Sukhatme, B.V. (1969). Optimum stratification. *Annals of Institute of Statistical Mathematica*, 21, 515- 528.
15. Singh, H.P. & Gorey, S.M. (2019). A remark on Gupta, Gupta and Singh Optional Randomized Response Model, University of SS. *JAMSI*, 15 (1). doi.org/10.2478/jamsi-2019-0004
16. Singh, H. P. & Tarray, T A. 2014. An improved randomized response additive model. *Srilankan J. Appl. Stat.*, 15 (2), 131–138. doi.org/10.4038/sljastats.v15i2.7412
17. Tarray, T.A., Singh, H.P & Yan, Z. (2015). A dexterous optional randomized response model. *Sociological Methods and Research*, 46(3), 565-585. doi.org/10.1177/0049124115605332
18. Tarray, T.A. & Singh, H.P. (2017). An optional randomized response model for estimating a rare sensitive attribute using poisson distribution. *Communications in Statistics-Theory and Methods*, 46(6), 2638-2654. doi.org/10.1080/03610926.2015.1040506
19. Verma, M.R., Singh, S. & Pandey, R. 2012. Optimum stratification for sensitive quantitative variables using auxiliary information. *J. Ind. Soc. Agril. Statist.*, 66(3), 401-412.
20. Warner, S.L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60, 63-69.