Communications in Advanced Mathematical Sciences
Vol. II, No. 1, 48-54, 2019
Research Article
e-ISSN:2651-4001
DOI: 10.33434/cams.443055

An Agile Optimal Orthogonal Additive Randomized Response Model

Tanveer A. Tarray[1]*, Housila P. Singh[2]

Abstract
In this paper, a new additive randomized response model has been proposed. The properties of the proposed model have been studied. It has been shown theoretically that the suggested additive model is better than the one envisaged by [1] under very realistic conditions. Numerical illustrations are also given in support of the present study.

Keywords: Estimation of mean, Randomized response sampling, Respondents protection, Sensitive quantitative variable.
2010 AMS: Primary 62D05, Secondary 62D99

[1] *Department of Mathematical Sciences, Islamic University of Science and Technology – Awantipora – Pulwama – Kashmir – India – 192122*
[2] *School of Studies in Statistics , Vikram University Ujjain, M.P. - India -456010*
***Corresponding author**: tanveerstat@gmail.com
Received: 12 July 2018, **Accepted:** 29 November 2018, **Available online:** 22 March 2019

1. Introduction

One problem with research on high – risk behavior is that respondents may consciously or unconsciously provide incorrect 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 of 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 the 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 [2]. It provides the opportunity of reducing response biases due to dishonest answers to sensitive questions. As a result, the technique assures a considerable degree of privacy protection in many contexts. Following the pioneering work of [2], many modifications are proposed in the literature. A good exposition of developments on randomized response techniques could refer to [3]-[18]. We below give the description of the model due to [1]

1.1 Additive model[1]:

Let there be k scrambling variables denoted by $S_j, j=1,2,\ldots,k$ whose mean $\theta_j (i.e. E(S_j)=\theta_j)$ and variance $\gamma_j^2 (i.e. V(S_j)=\gamma_j^2)$ are known. In [1] proposed optimal new orthogonal additive model named as (POONAM), each respondent selected in the sample is requested to rotate a spinner, as shown in Fig. 1.1, in which the proportion of the k shaded areas, say $P_1, P_2, \ldots P_k$ are

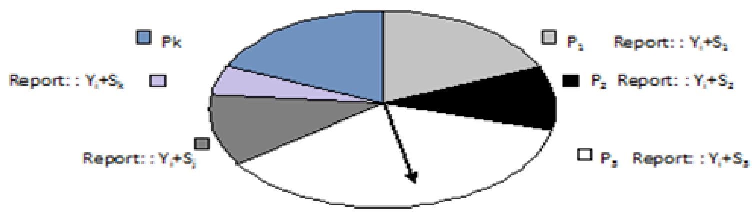

Figure 1.1. Spinner of POONAM[1]

orthogonal to the means of the k scrambling variables, say $\theta_1, \theta_2, \ldots\theta_k$ such that:

$$\sum_{i=1}^{k} P_j\theta_j = 0$$

and

$$\sum_{i=1}^{k} P_j = 1$$

Now if the pointer stops in the j^{th} shaded area, then the i^{th} respondent with the value of the sensitive variable, say Y_i, is requested report the scrambled response Z_i as:

$$Z_i = Y_i + S_i$$

Assuming that the sample of size n is drawn from the population using simple random sampling with replacement (SRSWR). [1] suggested an unbiased estimator of the population mean μ_Y as

$$\hat{\mu}_Y = \frac{1}{n}\sum_{i=1}^{n} Z_i$$

with variance

$$V(\hat{\mu}_Y) = \frac{1}{n}[\sigma_y^2 + \sum_{i=1}^{n} P_j(\theta_j^2 + \gamma_j^2)],$$

2. The proposed procedure:

Let $S_j, j = 1, 2, \ldots, k$ be k scrambling variables such that their distribution are known. In brief, let $E(S_j) = \theta_j)$ and variance $V(S_j) = \gamma_j^2)$ are known. Then, in the proposed additive model, each respondent selected in the sample is requested to rotate a spinner, as depicted in Fig. 2.1, in which the proportion of the $k$shaded areas, say $P_1, P_2, \ldots P_k$ are orthogonal to the means of the k scrambling variables, say $\theta_1, \theta_2, \ldots\theta_k$ such that:

$$\sum_{i=1}^{k} P_j\theta_j = 0$$

and

$$\sum_{i=1}^{k} P_j = 1$$

If the pointer stops in the j^{th} shaded area, then the i^{th} respondent with the value of the sensitive variable, say Y_i, is requested report the scrambled response Z_i^* as:

$$Z_i^* = Y_i + S_j^*$$

where $S_j^* = \dfrac{(a_j S_j + b_j \theta_j)}{(a_j + b_j)}$ and (a_j, b_j) being suitably chosen constants which may take real values and the functions of known parameters of scrambling variable S_j such as $\gamma_j, \theta_j, C_j (= \gamma_j/\theta_j), \beta_2(S_j) = \dfrac{\mu_4(S_j)}{\gamma_j^4}$ (coefficient of kurtosis), $G_1(S_j) = \dfrac{\mu_3(S_j)}{\gamma_j^3}$ is the Fisher's measure of skewness, $\mu_3(S_j)$ and $\mu_4(S_j)$ are third and fourth central moments of the scrambling variable S_j etc. Let a sample of size n be drawn from the population using the simple random sampling with replacement (SRSWR). Then we prove the following theorems.

Theorem 2.1. *An unbiased estimator of the population mean μ_Y is given by*

$$\hat{\mu}_{ST} = \frac{1}{n}\sum_{i=1}^{n} Z_i^*$$

Proof. Let E_1 and E_2 denote the expectation over the sampling design and the randomization device respectively, we have

$$E(\hat{\mu}_{ST}) = E_1 E_2[\frac{1}{n}\sum_{i=1}^{n} Z_i^*] = E_1[\frac{1}{n}\sum_{i=1}^{n} E_2(Z_i^*)] = E_1[\frac{1}{n}\sum_{i=1}^{n}(Y_i \sum_{i=1}^{n} P_j + \sum_{i=1}^{n} P_j\theta_j)]$$

$$E_1[\frac{1}{n}\sum_{i=1}^{n} Y_i] = \mu_Y,$$

since

$$\sum_{i=1}^{k} P_j\theta_j = 0$$

and

$$\sum_{i=1}^{k} P_j = 1$$

which completes the theorem. The variance of the proposed estimator $\hat{\mu}_{ST}$ is given in the following theorem. □

Theorem 2.2. *The variance of the proposed estimator $\hat{\mu}_{ST}$ is given by*

$$V(\hat{\mu}_{ST}) = \frac{1}{n}[\sigma_y^2 + \sum_{j=1}^{k} P_j\theta_j^2(1 + \eta_j^2 C_j^2)],$$

where $\eta_j = a_j/(a_j + b_j)$ and $C_j = \gamma_j/\theta_j; j = 1, 2, ...k$.

Proof. Let V_1 and V_2 denote the variance over the sampling design and over the proposed randomization device, respectively, then we have

$$\begin{aligned} V(\hat{\mu}_Y) &= E_1 V_2(\hat{\mu}_Y) + V_1 E_2(\hat{\mu}_Y) = E_1[V_2[\frac{1}{n}\sum_{i=1}^{n}(Z_i^*)] + V_1[E_2(\frac{1}{n}\sum_{i=1}^{n}(Z_i^*))] \\ &= E_1[\frac{1}{n^2}\sum_{i=1}^{n} V_2(Z_i^*)] + V_1[(\frac{1}{n}\sum_{i=1}^{n} E_2(Z_i^*))] = [\frac{\sigma_y^2}{n} + E_1[\frac{1}{n^2}\sum_{i=1}^{n} V_2(Z_i^*)]. \end{aligned}$$

Note that

$$V_2(Z_i^*) = \sum_{j=1}^{k} P_j E_2(Y_i + S_j^*)^2 - Y_i^2 = Y_i^2 + \sum_{j=1}^{k} P_j\theta_j^2(1 + \eta_j^2 C_j^2) - Y_i^2;$$

since

$$\sum_{i=1}^{k} P_j\theta_j = 0$$

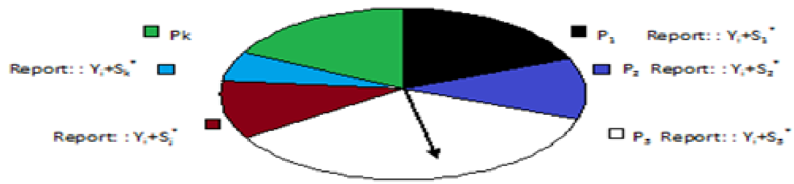

Figure 2.1. Spinner of Proposed Procedure

and

$$\sum_{i=1}^{k} P_j = 1$$

$$V_2(Z_i^*) = \sum_{j=1}^{k} P_j \theta_j^2 (1+\eta_j^2 C_j^2)$$

where $\eta_j = a_j/(a_j+b_j)$ and $C_j = \gamma_j/\theta_j; j = 1,2,...k.$
Solving the above equations , we get

$$V(\hat{\mu}_{ST}) = \frac{1}{n}[\sigma_y^2 + \sum_{j=1}^{k} P_j \theta_j^2 (1+\eta_j^2 C_j^2)],$$

This completes the proof of the theorem. □

3. Efficiency comparison

The proposed estimator $\hat{\mu}_{(ST)}$ will be more efficient than the estimator $\hat{\mu}_{(Y)}$ if

$$V(\hat{\mu}_{ST}) < V(\hat{\mu}_Y), if$$

i.e. if

$$\frac{1}{n}[\sigma_y^2 + \sum_{j=1}^{k} P_j \theta_j^2 (1+\eta_j^2 C_j^2)] < \frac{1}{n}[\sigma_y^2 + \sum_{j=1}^{k} P_j \theta_j^2 (1+C_j^2)]$$

i.e. if

$$[\sum_{j=1}^{k} P_j \theta_j^2 (1+\eta_j^2 C_j^2)] < [\sum_{j=1}^{k} P_j \theta_j^2 (1+C_j^2)]$$

i.e. if

$$[\sum_{j=1}^{k} P_j \theta_j^2 (\eta_j^2 - 1)] C_j^2 < 0$$

i.e. if

$$|\eta_j^2| < 1 \forall j = 1,2,...k$$

It follows from the above equation that we should choose the value of (a_j, b_j) such a way that

$$|\frac{a_j}{a_j+b_j}| < 1$$

σ_Y^2	θ_1	θ_2	θ_3	θ_4	$a_j = \frac{b_j}{4} and \eta_j = \frac{1}{5}$
25	1	5	10	-10.25	538.02
25	5	10	15	-17.50	310.22
25	10	15	20	-25.00	215.74
25	15	20	25	-32.50	171.97
125	1	5	10	-10.25	343.28
125	5	10	15	-17.50	251.87
125	10	15	20	-25.00	195.53
125	15	20	25	-32.50	163.61

Table 1. Percent relative efficiencies of the proposed estimator $\hat{\mu}_{(ST)}$ over the Singh (2010) estimator $\hat{\mu}_{(Y)}$.

σ_Y^2	$a_j = \frac{b_j}{4} and \eta_j = \frac{1}{5}$
25	1244.77
125	470.23
225	320.82
325	257.33
425	222.20
525	199.89
625	184.47
725	173.17
825	164.54

Table 2. Percent relative efficiencies of the proposed estimator $\hat{\mu}_{(ST)}$ over the Singh (2010) estimator $\hat{\mu}_{(Y)}$ for $\theta_j, j = 0,1,2...k$.

We have computed the percent relative efficiency (PRE) of the proposed estimator $\hat{\mu}_{(ST)}$ with respect to Singh's estimator $\hat{\mu}_Y$ by using the formula:

$$PRE(\hat{\mu}_{(ST)}, \hat{\mu}_{(Y)}) = \frac{[\sigma_y^2 + \sum_{j=1}^{k} P_j(\theta_j^2 + \gamma_j^2)]}{[\sigma_y^2 + \sum_{j=1}^{k} P_j(\theta_j^2 + \eta_j^2\gamma_j^2)]} \times 100$$

By keeping the respondents cooperation in mind, we decided to choose $\gamma = 40, \gamma_1 = 30, \gamma_2 = 40, \gamma_3 = 20, \gamma_4 = 10, P_1 = 0.01, P_2 = 0.02, P_3 = 0.03, P_4 = 0.04$ with $k = 4$. In addition we choose different values $\sigma_y^2, \theta_1, \theta_2, \theta_3, \theta_4$.

It is observed that the values of $PRE(\hat{\mu}_{(ST)}, \hat{\mu}_{(Y)})$ are greater than 100. It follows that the proposed estimator $\hat{\mu}_{(ST)}$ is more efficient than the estimator $\hat{\mu}_{(Y)}$ due to [1] with a substantial gain in efficiency. Thus, based on our simulation results, the use of the proposed estimator $\hat{\mu}_{(ST)}$ over [1] estimator $\hat{\mu}_{(Y)}$ is recommended for all situations. We also consider a situation where $\theta_j = 0 for j = 1,2,3,4$, and rest of the parameters are kept the same.The percent relative efficiency of the proposed estimator $\hat{\mu}_{(ST)}$ over [1] estimators $\hat{\mu}_{(Y)}$ has been shown. Numerical illustration clearly show that the percent relative efficiencies remain higher if the value of σ_y^2 is small. We have further considered the case $k = 2$ and computed the $PRE(\hat{\mu}_{(ST)}, \hat{\mu}_{(Y)})$ for different choices of parameters. Thus, based on our numerical findings, the proposed estimator $\hat{\mu}_{(ST)}$ is to be preferred over [1] estimator $\hat{\mu}_{(Y)}$ is recommended for all situations in real practice. It should be noted here that the experience is must in real surveys while making a choice of randomization device to be used in practice.

Discussion

In this article, we have suggested a new additive randomized response model and its properties are studied. We have proved the superiority of the proposed randomized response model over [1] randomized response models both theoretically and empirically.

σ_Y^2	θ_1	θ_2	θ_3	θ_4	$a_j = \frac{b_j}{4} and \eta_j = \frac{1}{5}$
25	1	1	2	-2.25	728.03
25	1	2	2	-2.75	505.45
25	2	1	3	-3.25	340.55
25	2	2	3	-3.75	239.00
125	1	1	2	-2.25	383.36
125	1	2	2	-2.75	320.09
125	2	1	3	-3.25	253.68
125	2	2	3	-3.75	199.62

Table 3. Percent relative efficiencies of the proposed estimator $\hat{\mu}_{(ST)}$ over the Singh (2010) estimator $\hat{\mu}_{(Y)}$.

σ_Y^2	$PRE's$
25	2380.00
125	556.00
225	353.33
325	275.38
425	234.12
525	208.57
625	191.20
725	178.62
825	169.09

Table 4. Percent relative efficiencies of the proposed estimator $\hat{\mu}_{(ST)}$ over the Singh (2010) estimator $\hat{\mu}_{(Y)}$ for $\theta_j, j = 0,1,2...k$.

σ_Y^2	θ_1	θ_2	θ_3	θ_4	$a_j = \frac{b_j}{4} and \eta_j = \frac{1}{5}$
25	1	5	10	-10.25	946.45
25	5	10	15	-17.50	715.96
25	10	15	20	-25.00	508.30
25	15	20	25	-32.50	373.52
125	1	5	10	-10.25	368.67
125	5	10	15	-17.50	340.15
125	10	15	20	-25.00	300.41
125	15	20	25	-32.50	261.38

Table 5. Percent relative efficiencies of the proposed estimator $\hat{\mu}_{(ST)}$ over the Singh (2010) estimator $\hat{\mu}_{(Y)}$ with $k = 2$.

σ_Y^2	θ_1	θ_2	θ_3	θ_4	$a_j = \theta_j and b_j = \gamma_j$ with $k = 2$
25	1	1	2	-2.25	1676.13
25	1	2	2	-2.75	1599.26
25	2	1	3	-3.25	1587.15
25	2	2	3	-3.75	1518.21
125	1	1	2	-2.25	424.94
125	1	2	2	-2.75	421.12
125	2	1	3	-3.25	420.26
125	2	2	3	-3.75	416.53

Table 6. Percent relative efficiencies of the proposed estimator $\hat{\mu}_{(ST)}$ over the Singh (2010) estimator $\hat{\mu}_{(Y)}$.

Acknowledgement

The authors are thankful to the Editor – in- Chief, and to the anonymous learned referees for their valuable suggestions regarding improvement of the paper.

References

[1] S. Singh, *Proposed optimal orthogonal new additive model (POONAM)*, Statist., LXX(1) (2010), 73–81.

[2] S. L. Warner, *Randomized response: A survey technique for eliminating evasive answer bias*, J. Amer. Statist. Assoc., **60** (1965), 63-69.

[3] S. K. Bar –Lev, E. Bobovitch, B. Boukai, *A note on Randomized response models for quantitative data*, Metrika, **60** (2004), 225-250.

[4] J. A. Fox, P. E. Tracy, *Randomized Response: A method of Sensitive Surveys*, Newbury Park, CA: SEGE Publications 1986.

[5] C. R. Gjestvang, S. Singh, *A new randomized response model*, J. Roy. Statist. Soc., **68** (2006), 523-530.

[6] C. R. Gjestvang, S. Singh, *An improved randomized response model: Estimation of mean*, J. Appl. Statist., **36**(12) (2009), 1361-1367.

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

[8] H. P. Singh, N. Mathur, *Estimation of population mean when coefficient of variation is known using scrambled response technique*, J. Statist. Plann. Infer., **131**(1) (2005), 135-144.

[9] H. P. Singh, T. A. Tarray, *An alternative to Kim and Warde's mixed randomized response technique*, Statistica, **73**(3) (2013), 379-402.

[10] H. P. Singh, T. A. Tarray, *An alternative to stratified Kim and Warde's randomized response model using optimal (Neyman) allocation*, Model Assist. Stat. Appl., **9** (2014), 37-62.

[11] H. P. Singh, T. A. Tarray, *A stratified Mangat and Singh's optional randomized response model using proportional and optimal allocation*, Statistica, **74**(1) (2014),65-83.

[12] H. P. Singh, T. A. Tarray, *An improved randomized response additive model*, Sri Lan. J. Appl. Statist., **15**(2) (2014) ,131-138.

[13] T. A. Tarray, H. P. Singh, Y. Zaizai, *A stratified optional randomized response model*, Socio. Meth. Res., **45** (2015), 1-15.

[14] T. A. Tarray, H. P. Singh, *An improved new additive model*, Gazi Uni. J. Sci., **29**(1) (2016), 159-165.

[15] T. A. Tarray, H. P. Singh, Y. Zaizai, *A dexterous optional randomized response model*, Socio. Meth. Res., **46**(3) (2017),565-585.

[16] T. A. Tarray, H. P. Singh, *A simple way of improving the Bar–Lev, Bobovitch and Boukai Randomized response model*, Kuwait J. Sci., **44**(4) (2017), 83-90.

[17] T. A. Tarray, H. P. Singh, *A randomization device for estimating a rare sensitive attribute in stratified sampling using Poisson distribution*, Afri. Mat., **29**(3) (2018), 407-423.

[18] T. A. Tarray, H.P. Singh, *Missing data in clinical trials: stratified Singh and Grewal's randomized response model using geometric distribution*, Trends Bioinfor., **11**(1) (2018), 63-69