PROPOSED OPTIMAL ORTHOGONAL NEW ADDITIVE MODEL (POONAM)

Sarjinder Singh

1. INTRODUCTION

The problem of estimation of a proportion of a sensitive character using a randomization device in survey sampling is well known since Warner (1965). A detailed review and applications of such techniques can be found in Fox and Tracy (1986). Following Gjestvang and Singh (2006), let α and β be two known positive real numbers. Gjestvang and Singh (2009) considered a new additive model in which each respondent in the sample is requested to draw a card secretly from a well-shuffled deck of cards. In the deck, let p be the proportion of cards bearing the statement, "Multiply scrambling variable S with α and add to the real value of the sensitive variable Y_i", and $(1-p)$ be the proportion of cards bearing the statement, "Multiply scrambling variable S with β and subtract it from the real value of the sensitive variable Y_i." Mathematically, each respondent is requested to report the scrambled response Z_i as:

$$Z_i = \begin{cases} Y_i + \alpha S & \text{with probability } p = \beta/(\alpha+\beta) \\ Y_i - \beta S & \text{with probability } (1\text{-p}) = \alpha/(\alpha+\beta) \end{cases} \tag{1}$$

Gjestvang and Singh (2009) defined an unbiased estimator of the population mean $\bar{Y}$ as:

$$\hat{\bar{Y}}_{GS} = \frac{1}{n}\sum_{i=1}^{n} Z_i \tag{2}$$

with variance given by

$$V(\hat{\bar{Y}}_{GS}) = \frac{1}{n}[\sigma_y^2 + \alpha\beta(\theta^2 + \gamma^2)] \tag{3}$$

where $V(S)=\gamma^2$ and $E(S)=\theta$ are known, $\sigma_y^2 = N^{-1}\sum_{i=1}^{N}(Y_i-\overline{Y})^2$ be the variance of the sensitive variable Y and N be the population size. Some recent contribution to randomized response sampling is given by Odumade and Singh (2008, 2009a, 2009b) and Singh and Chen (2009).

2. POONAM

Let S_j, $j=1,2,...,k$ be k scrambling variables such that their distributions are known. In short, let $E(S_j)=\theta_j$ and $V(S_j)=\gamma_j^2$ for $j=1,2,...,k$ be known. Then, in the proposed optimal orthogonal new additive model (POONAM), each respondent selected in the sample is requested to rotate a spinner, as shown in Fig. 1, in which the proportions of the k shaded areas, say p_1, p_2, ..., p_k are orthogonal to the means of the k scrambling variables, say θ_1, θ_2, ..., θ_k such that:

$$\sum_{j=1}^{k} p_j\theta_j = 0 \tag{4}$$

and

$$\sum_{j=1}^{k} p_j = 1 \tag{5}$$

Now if the pointer stops in the jth shaded area, then the ith respondent with real value of the sensitive variable, say Y_i, is requested to report the scrambled response Z_i as:

$$Z_i = Y_i + S_j \tag{6}$$

One of the easiest method to make such a randomization device is to choose the values of p_j, $j=1,2,...k$ subject to (5) and θ_j, $j=1,2,...(k-1)$ as you like, but make the choice of θ_k, such that (4) is satisfied, so:

$$\theta_k = \frac{-\sum_{j=1}^{k-1} p_j\theta_j}{p_k} \tag{7}$$

Notice that for making the orthogonal randomization devices, at least one of the scrambling variables is assumed to have negative mean value. As reported by

Gjestvang and Singh (2006) that negative responses help in randomized response sampling, and we also notice that a choice of such a scrambling variable remains useful.

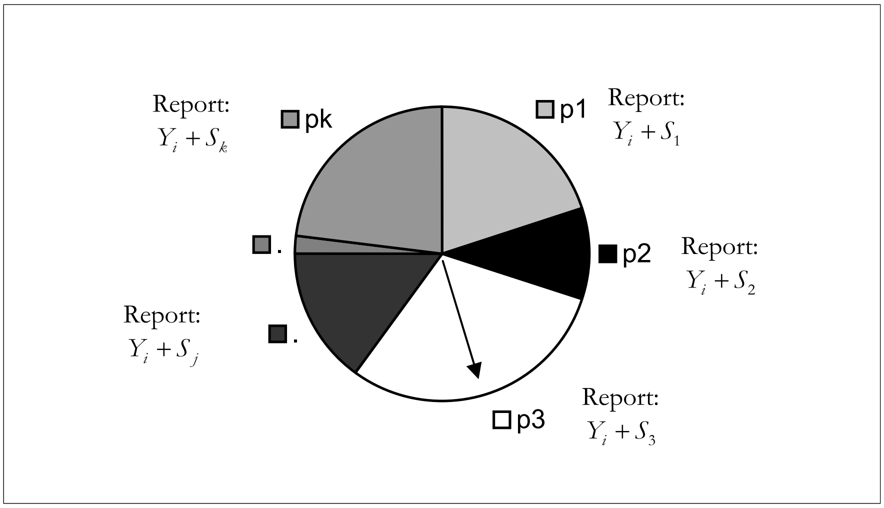

Figure 1 – Spinner for POONAM.

Note that the parameters of both the randomization devices, the spinner and the means of the scrambling variables, are orthogonal to each other and hence we named it the proposed optimal orthogonal new additive model (POONAM). Assuming that the sample of size n is selected using the simple random and with replacement (SRSWR) sampling, we prove the following theorems:

Theorem 1. An unbiased estimator of the population mean $\overline{Y}$ is given by

$$\hat{\overline{Y}}_p = \frac{1}{n}\sum_{i=1}^{n} Z_i \tag{8}$$

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

$$E(\hat{\overline{Y}}_p) = E_1 E_2\left[\frac{1}{n}\sum_{i=1}^{n} Z_i\right] = E_1\left[\frac{1}{n}\sum_{i=1}^{n} E_2(Z_i)\right] = E_1\left[\frac{1}{n}\sum_{i=1}^{n}\left\{Y_i \sum_{j=1}^{k} p_j + \sum_{j=1}^{k} p_j\theta_j\right\}\right]$$

$$= E_1\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right] = \overline{Y}$$

which proves the theorem.

Theorem 2. The variance of the proposed estimator $\hat{\hat{Y}}_p$ is given by

$$V(\hat{\hat{Y}}_p) = \frac{1}{n}\left[\sigma_y^2 + \sum_{j=1}^{k} p_j(\theta_j^2 + \gamma_j^2)\right] \tag{9}$$

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

$$V(\hat{\hat{Y}}_p) = E_1V_2(\hat{\hat{Y}}_p) + V_1E_2(\hat{\hat{Y}}_p) = E_1\left[V_2\left\{\frac{1}{n}\sum_{i=1}^{n} Z_i\right\}\right] + V_1\left[E_2\left\{\frac{1}{n}\sum_{i=1}^{n} Z_i\right\}\right] \tag{10}$$

Note that:

$$V_2(Z_i) = \sum_{j=1}^{k} p_j E_2(Y_i + S_j)^2 - Y_i^2 = Y_i^2\sum_{j=1}^{k} p_j + 2Y_i\sum_{j=1}^{k} p_j\theta_j + \sum_{j=1}^{k} p_j(\gamma_j^2 + \theta_j^2) - Y_i^2$$

$$= \sum_{j=1}^{k} p_j(\theta_j^2 + \gamma_j^2) \tag{11}$$

On using (11) in (10), we have the theorem.

Remarks: **(a)** One choice of p_j could be considered as:

$$p_j = \frac{1/(\theta_j^2 + \gamma_j^2)}{\sum_{j=1}^{k}\{1/(\theta_j^2 + \gamma_j^2)\}} \tag{12}$$

Then, the variance of the POONAM estimator $\hat{\hat{Y}}_p$ becomes:

$$V(\hat{\hat{Y}}_p)_{\text{Remark}} = \frac{1}{n}\left[\sigma_y^2 + \frac{1}{\sum_{j=1}^{k}\{1/(\theta_j^2 + \gamma_j^2)\}}\right] \tag{13}$$

(b) One obvious choice of $\theta_j = 0$ for all $j = 1, 2, 3..., k$, will also satisfy the condition (4) for any choice of p_j satisfying (5).

3. EFFICIENCY COMPARISONS

The proposed estimator POONAM $\hat{\bar{Y}}_p$ will be more efficient than the estimator $\hat{\bar{Y}}_{GS}$ if:

$$V(\hat{\bar{Y}}_p) < V(\hat{\bar{Y}}_{GS}) \tag{14}$$

or if

$$\frac{1}{n}\left[\sigma_y^2 + \sum_{j=1}^{k} p_j(\theta_j^2 + \gamma_j^2)\right] < \frac{1}{n}[\sigma_y^2 + \alpha\beta(\theta^2 + \gamma^2)]$$

or if

$$\sum_{j=1}^{k} p_j(\theta_j^2 + \gamma_j^2) < \alpha\beta(\theta^2 + \gamma^2) \tag{15}$$

The condition (15) depends only on the randomization devices parameters, and it could be always possible to adjust the randomization device parameters such that (15) is satisfied. The relative efficiency of the POONAM estimator $\hat{\bar{Y}}_p$ with respect to the recent estimator $\hat{\bar{Y}}_{GS}$ is given by:

$$RE = \frac{\sigma_y^2 + \alpha\beta(\theta^2 + \gamma^2)}{\sigma_y^2 + \sum_{j=1}^{k} p_j(\theta_j^2 + \gamma_j^2)} \times 100\% \tag{16}$$

By keeping the respondents' cooperation in mind, we decided to choose $\alpha = 0.4$, $\beta = 0.6$ (similarly to Gjestvang and Singh (2009)), $\gamma = 40$, $\gamma_1 = 30$, $\gamma_2 = 40$, $\gamma_3 = 20$, $\gamma_4 = 10$, $p_1 = 0.02$, $p_2 = 0.05$, $p_3 = 0.06$ and $p_4 = 0.87$ with $k = 4$. In addition, we choose different values of σ_y^2, θ, θ_1, θ_2, θ_3 and θ_4 as listed in Table 1.

The value of θ was allowed to change between 200 to 1700, the value θ_1 was allowed to change between 300 to 1800, the value of θ_2 was allowed to change between 200 to 1700 and the value of θ_3 was allowed to change between 100 to 1600. Then the values of θ_4 were computed so that θ_j and p_j for $j = 1,2,3,4$ are orthogonal. The computed values of θ_4 ranged between -249.94 to –25.20. The relative efficiency (RE) values have been presented in the 7th and 14th columns of Table 1, which indicates that the POONAM estimator remains more ef-

TABLE 1

Relative efficiency of the POONAM estimator over the GS estimator

σ_y^2	θ	θ_1	θ_2	θ_3	θ_4	*RE*	σ_y^2	θ	θ_1	θ_2	θ_3	θ_4	*RE*
25	200	300	200	100	-25.2	192.84	1225	200	300	200	100	-25.3	175.41
	700	800	700	600	-100.0	173.97		700	800	700	600	-100.0	172.68
	1200	1300	1200	1100	-174.7	168.63		1200	1300	1200	1100	-174.7	168.23
	1700	1800	1700	1600	-249.4	166.33		1700	1800	1700	1600	-249.4	166.14
125	200	300	200	100	-25.3	191.08	1325	200	300	200	100	-25.3	174.24
	700	800	700	600	-100.0	173.86		700	800	700	600	-100.0	172.58
	1200	1300	1200	1100	-174.7	168.59		1200	1300	1200	1100	-174.7	168.20
	1700	1800	1700	1600	-249.4	166.31		1700	1800	1700	1600	-249.4	166.12
225	200	300	200	100	-25.3	189.40	1425	200	300	200	100	-25.3	173.12
	700	800	700	600	-100.0	173.75		700	800	700	600	-100.0	172.47
	1200	1300	1200	1100	-174.7	168.56		1200	1300	1200	1100	-174.7	168.16
	1700	1800	1700	1600	-249.4	166.29		1700	1800	1700	1600	-249.4	166.10
325	200	300	200	100	-25.3	187.77	1525	200	300	200	100	-25.3	172.02
	700	800	700	600	-100.0	173.64		700	800	700	600	-100.0	172.37
	1200	1300	1200	1100	-174.7	168.53		1200	1300	1200	1100	-174.7	168.13
	1700	1800	1700	1600	-249.4	166.28		1700	1800	1700	1600	-249.4	166.09
425	200	300	200	100	-25.3	186.20	1625	200	300	200	100	-25.3	170.96
	700	800	700	600	-100.0	173.53		700	800	700	600	-100.0	172.26
	1200	1300	1200	1100	-174.7	168.49		1200	1300	1200	1100	-174.7	168.10
	1700	1800	1700	1600	-249.4	166.26		1700	1800	1700	1600	-249.4	166.07
525	200	300	200	100	-25.3	184.68	1725	200	300	200	100	-25.3	169.93
	700	800	700	600	-100.0	173.43		700	800	700	600	-100.0	172.16
	1200	1300	1200	1100	-174.7	168.46		1200	1300	1200	1100	-174.7	168.06
	1700	1800	1700	1600	-249.4	166.25		1700	1800	1700	1600	-249.4	166.06
625	200	300	200	100	-25.3	183.22	1825	200	300	200	100	-25.3	168.93
	700	800	700	600	-100.0	173.32		700	800	700	600	-100.0	172.06
	1200	1300	1200	1100	-174.7	168.43		1200	1300	1200	1100	-174.7	168.03
	1700	1800	1700	1600	-249.4	166.23		1700	1800	1700	1600	-249.4	166.04
725	200	300	200	100	-25.3	181.81	1925	200	300	200	100	-25.3	167.96
	700	800	700	600	-100.0	173.21		700	800	700	600	-100.0	171.95
	1200	1300	1200	1100	-174.7	168.40		1200	1300	1200	1100	-174.7	168.00
	1700	1800	1700	1600	-249.4	166.21		1700	1800	1700	1600	-249.4	166.02
825	200	300	200	100	-25.3	180.44	2025	200	300	200	100	-25.3	167.02
	700	800	700	600	-100.0	173.11		700	800	700	600	-100.0	171.85
	1200	1300	1200	1100	-174.7	168.36		1200	1300	1200	1100	-174.7	167.97
	1700	1800	1700	1600	-249.4	166.20		1700	1800	1700	1600	-249.4	166.01
925	200	300	200	100	-25.3	179.12	2125	200	300	200	100	-25.3	166.10
	700	800	700	600	-100.0	173.00		700	800	700	600	-100.0	171.75
	1200	1300	1200	1100	-174.7	168.33		1200	1300	1200	1100	-174.7	167.93
	1700	1800	1700	1600	-249.4	166.18		1700	1800	1700	1600	-249.4	165.99
1025	200	300	200	100	-25.3	177.84	2225	200	300	200	100	-25.3	165.20
	700	800	700	600	-100.0	172.89		700	800	700	600	-100.0	171.64
	1200	1300	1200	1100	-174.7	168.30		1200	1300	1200	1100	-174.7	167.90
	1700	1800	1700	1600	-249.4	166.17		1700	1800	1700	1600	-249.4	165.98
1125	200	300	200	100	-25.3	176.60	2425	200	300	200	100	-25.3	163.48
	700	800	700	600	-100.0	172.79		700	800	700	600	-100.0	171.44
	1200	1300	1200	1100	-174.7	168.26		1200	1300	1200	1100	-174.7	167.83
	1700	1800	1700	1600	-249.4	166.15		1700	1800	1700	1600	-249.4	165.95

ficient than the Gjestvang and Singh (2009) estimator in all situations simulated in the present investigation. A more depth study of the relative efficiency results in Table 1 indicates that the mean relative efficiency value remains 171.00% with standard deviation of 6.01%. The minimum value of the relative efficiency in Table 1 is observed as 163.48% and maximum 192.82% with a median of 168.47% based on 96 situations investigated in Table 1 for different choice of parameters.

In the next case, we consider a situation where $\theta = 0$ as well as $\theta_j = 0$ for $j = 1,2,3,4$, and rest of the parameters are kept same as in Table 1. The relative efficiency of the POONAM estimator over the Gjestvang and Singh (2009) estimator has been quoted in Table 2.

TABLE 2

RE of the POONAM estimator over the GS estimator

σ_y^2	25	125	225	325	425	525	625	725	825	925	1025	1125
RE	174.79	152.40	140.32	132.77	127.60	123.84	120.98	118.74	116.92	115.43	114.18	113.12
σ_y^2	1225	1325	1425	1525	1625	1725	1825	1925	2025	2125	2225	2425
RE	112.20	111.41	110.71	110.09	109.54	109.05	108.60	108.20	107.83	107.50	107.19	106.64

The mean relative efficiency value remains 191.17% with standard deviation of 16.50%. The minimum value of the relative efficiency in Table 2 is observed as 106.64% and maximum 174.79% with a median of 112.66% based on 24 situations investigated in Table 2 for different choice of parameters.

From Table 2, we learned that the RE value remains higher if the value of σ_y^2 is small. In order to look as the maximum gain we also investigated lower values of σ_y^2 given that in practice, for example, the number of abortions by a woman could vary from 0 to 3 or 4, because it may not be practical for a woman to go for more than 3 or 4 abortions. In that case the value of σ_y^2 will be around 0.5 to 5.0. We observed that the relative efficiency value decreases from 183.53% to 181.78% as the value of σ_y^2 increases from 0.5 to 5 when all the means of the scrambling variables are at zero level.

In Table 3, we provide different choice of parameters for $k = 2$ such that the POONAM estimator remains more efficient than the Gjestvang and Singh (2009) estimator. For $25 \le \sigma_y^2 \le 2425$, $p_1 = 0.2$, $p_2 = 1 - p_1 = 0.8$, $\theta = 1700$, $\theta_1 = 1300$, and $\theta_2 = -325$, the RE values remain almost equal to 163%; for $p_1 = 0.4$, $p_2 = 1 - p_1 = 0.6$, $\theta = 700$, $\theta_1 = 300$, and $\theta_2 = -200$, the RE values remains in the range 188% to 192%. Thus, based on our simulation results, the use of POONAM over the Gjestvang and Singh (2009) estimator is recommended for all situations close to Table 1, Table 2 and Table 3 in real practice. Note that experience is must in real surveys while making a choice of randomization device to be used in practice.

TABLE 3

RE of the POONAM estimator over the GS estimator with $k = 2$

p_1	θ	θ_1	θ_2	σ_y^2	RE	p_1	θ	θ_1	θ_2	σ_y^2	RE
0.2	1700	1300	-325.0	25	163.69	0.4	1700	800	-533.3	25	162.15
				125	163.67					125	162.13
				225	163.66					225	162.12
				325	163.64					325	162.10
				425	163.63					425	162.09
				525	163.61					525	162.07
				625	163.60					625	162.06
				725	163.58					725	162.05
				825	163.57					825	162.03
				925	163.55					925	162.02
				1025	163.54					1025	162.00
				1125	163.52					1125	161.99
				1225	163.51					1225	161.97
				1325	163.49					1325	161.96
				1425	163.48					1425	161.94
				1525	163.46					1525	161.93
				1625	163.45					1625	161.92
				1725	163.43					1725	161.90
				1825	163.42					1825	161.89
				1925	163.40					1925	161.87
				2025	163.39					2025	161.86
				2125	163.37					2125	161.84
				2225	163.36					2225	161.83
				2325	163.34					2325	161.82
				2425	163.33					2425	161.80
0.4	700	300	-200.0	25	192.37	0.8	1700	300	-1200.0	25	192.21
				125	192.22					125	192.19
				225	192.07					225	192.16
				325	191.92					325	192.14
				425	191.77					425	192.11
				525	191.62					525	192.08
				625	191.47					625	192.06
				725	191.33					725	192.03
				825	191.18					825	192.01
				925	191.03					925	191.98
				1025	190.89					1025	191.96
				1125	190.74					1125	191.93
				1225	190.60					1225	191.91
				1325	190.45					1325	191.88
				1425	190.31					1425	191.86
				1525	190.16					1525	191.83
				1625	190.02					1625	191.80
				1725	189.88					1725	191.78
				1825	189.74					1825	191.75
				1925	189.59					1925	191.73
				2025	189.45					2025	191.70
				2125	189.31					2125	191.68
				2225	189.17					2225	191.65
				2325	189.03					2325	191.63
				2425	188.89					2425	191.60

Acknowledgements: The author is thankful to the Executive Editor D.M. Daniela Grazia and an educated referee to bring the original manuscript in the present form.

Department of Mathematics
Texas A&M University-Kingsville
Kingsville, TX 78363, USA

SARJINDER SINGH

REFERENCES

J. A. FOX, P.E. TRACY (1986), *Randomized response: A method for sensitive surveys*, "SAGE Publications."

C. R. GJESTVANG, S. SINGH (2006), *A new randomized response model*, "Journal of the Royal Statistical Society" B, 68, 523-530.

C. R. GJESTVANG, S. SINGH (2009), *An improved randomized response model: Estimation of mean*, "Journal of Applied Statistics" 36(12), 1361-1367.

O. ODUMADE, S. SINGH (2008), *Generalized forced quantitative randomized response model: A unified approach*, "Journal of the Indian Society of Agricultural Statistics" 62(3), 244-252.

O. ODUMADE, S. SINGH (2009a), *Efficient use of two decks of cards in randomized response Sampling*, "Communications in Statistics-Theory and Methods" 38, (3-5), 439-446.

O. ODUMADE, S. SINGH (2009b), *Improved Bar-Lev, Bobovitch, and Boukai randomized response models*, "Communications in Statistics- Simulation and Computation" 38(3), 473-502.

S. SINGH, C.C. CHEN (2009), *Utilization of higher order moments of scrambling variables in randomized response sampling*, "Journal of Statistical Planning and Inference" 139(9), 3377-3380.

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

SUMMARY

Proposed Optimal Orthogonal New Additive Model (POONAM)

In this paper, the proposed optimal orthogonal new additive model (POONAM) is shown to remain more efficient than the recent additive model introduced by Gjestvang and Singh (2009). Several situations where the POONAM estimator shows efficiency over the Gjestvang and Singh (2009) model are simulated and investigated.