An Improved Randomized Response Additive Model

Housila P. Singh and Tanveer A. Tarray*
School of Studies in Statistics,
Vikram University, Ujjain – India, 456010.
Corresponding Author: tanveerstat@gmail.com

Received: 23, January 2014 / Revised: 27, June 2014 / Accepted: 12, August 2014

ABSTRACT

In this paper a new randomized response model has been proposed. The theoretical properties of the proposed model have been studied. Comparisons of the proposed model with the existing additive models have been made. It is found that the proposed model is superior to the existing additive model under very realistic condition. Numerical illustrations are also given in support of the present study.

Keywords: Randomized response sampling, Estimation of proportion, Respondents protection, sensitive quantitative variable.

1. Introduction

Himmelfarb and Edgel (1980) have considered the usual additive model for gathering information on quantitative sensitive variables. Their model permits the interviewer to hide personal information using a scrambling variable to the response. Let *Y* be the true response variable for which we desire to estimate its mean μ_Y . Suppose S is a scrambling variable whose distribution is known, that is, its population mean θ and variance γ^2 are known. For estimating μ_Y , a simple random sample of *n* respondents is selected with replacement from the population. Each respondent selected in the sample is requested to draw a value from the distribution of the scrambling variable, add it to the real response and report back to the interviewer. Thus, the observed i^{th} scrambled response is given by

$$Z_i = [Y_i + S] \tag{1}$$

Thus an unbiased estimator of the population mean μ_Y is given by

$$\hat{\mu}_{Y(1)} = \frac{1}{n}\sum_{i=1}^{n}\left(Z_i - \theta\right) \tag{2}$$

and the variance of the estimator $\hat{\mu}_{Y(1)}$ is given by

$$V\left(\hat{\mu}_{Y(1)}\right) = \frac{1}{n}\left(\sigma_y^2 + \gamma^2\right) \tag{3}$$

Gjestvang and Singh (2009) have mentioned that "the practical application of an additive model is much easier than the multiplicative model, that is respondents may like to add two numbers rather than doing painstaking work of multiplying two numbers or dividing two numbers; thus the improvement of the additive model has its importance in the literature." This led Gjestvang and Singh (2009) to suggest an alternative additive randomized response model. The description of the additive model due to Gjestvang and Singh (2009) is given below.

Let α and β be two known positive real numbers [see Gjestvang and Singh (2006)]. Consider a deck of cards in which *P* is the proportion of cards bearing the statement: 'Multiply scrambling variable *S* with α and add to the real value of the sensitive variable *Y*' and *(1-P)* be the proportion of the cards bearing the statement: 'Multiply scrambling variable *S* with β and subtract it from the real value of the sensitive variable *Y*'. Let $P = \beta/\left(\alpha+\beta\right)$ be known. Each respondent is asked to draw one card secretly and report the scrambled response accordingly. Mathematically, we have

$$Z_i = \begin{cases} Y_i + \alpha S & with \;\; probability \;\; P = \dfrac{\beta}{\left(\alpha+\beta\right)} \\ Y_i - \beta S & with \;\; probability \left(1-P\right) = \dfrac{\alpha}{\left(\alpha+\beta\right)} \end{cases} \tag{4}$$

Then, Gjestvang and Singh (2009) proposed an unbiased estimator of the population mean μ_Y as

$$\hat{\mu}_{Y(2)} = \frac{1}{n}\sum_{i=1}^{n} Z_i \tag{5}$$

The variance of $\hat{\mu}_{Y(2)}$ is given by

$$V\left(\hat{\mu}_{Y(2)}\right) = \frac{1}{n}\left[\sigma_y^2 + \alpha\beta\left(\gamma^2 + \theta^2\right)\right] \tag{6}$$

It is not surprise to mention that the mean (θ) and variance (γ^2) of the scrambling variable *S* are known, but as far as our knowledge goes no one has used this information in defining the estimator of the population mean μ_Y of the true response variable *Y*. It is well known that the main advantage of standardization is the reduction in the size of the raw data so that the data can be handled easily. Here we have used the standardized scrambling variable to utilize the prior knowledge of mean *(θ)* and standard deviation *(γ)* of the scrambling variable. It is also well recognized that the use of prior knowledge of the parameters such as mean *(θ)* and standard deviation *(γ)* either before the start of the survey or at the estimation stage to improve the precision of the estimate of the parameter [see, Thompson (1968), Mehta and Srinivasan (1971) and Hirano (1977)]. Thus in our study we have used the standardized scrambling variable so that the proposed model perform better (or to increase the efficiency of the proposed resulting estimator) than some existing models. This fact has been shown both theoretically and empirically. For real situations where such models can be used, the reader is referred to Eichhorn and Hayre (1983), Ahsanullah and Eichhorn (1988), Bar – Lev et al. (2004) and Gjestvang and Singh (2009) etc.

2. The Suggested Additive Model

The procedure is exactly the same as Gjestvang and Singh (2009) the only difference in the proposed procedure is to use standardized scrambled variable $S^* = (S-\theta)/\gamma$ in place of the original scramble variable *S*. Replacing *S* by S^* in (4), we have the observed ith scrambled response as

$$Z_i = \begin{cases} Y_i + \alpha S^* & with \;\; probability \qquad P = \beta/(\alpha+\beta) \\ Y_i - \beta S^* & with \;\; probability\,(1-P) = \alpha/(\alpha+\beta) \end{cases} \tag{7}$$

Then we suggest an unbiased estimator of the population mean μ_Y , as

$$\hat{\mu}_{Y(3)} = \frac{1}{n}\sum_{i=1}^{n} Z_i \tag{8}$$

and the variance of $\hat{\mu}_{Y(3)}$ is given by

$$V(\hat{\mu}_{Y(3)}) = \frac{1}{n}(\sigma_y^2 + \alpha\beta) \tag{9}$$

From (6) and (9) we have

$$V(\hat{\mu}_{Y(2)}) - V(\hat{\mu}_{Y(3)}) = \frac{\alpha\beta}{n}(\gamma^2 + \theta^2 - 1). \tag{10}$$

which is greater than 'zero' if

$$\gamma^2 > (1 - \theta^2) \tag{11}$$

The condition (11) is always true if $|\theta| < 1$ as γ^2 is always positive.
Further from (3) and (9) we have

$$V(\hat{\mu}_{Y(1)}) - V(\hat{\mu}_{Y(3)}) = \frac{1}{n}(\gamma^2 - \alpha\beta)$$

which is positive if

$$\gamma^2 > \alpha\beta \tag{12}$$

3. Relative Efficiency With Respect to Different Additive Model

The percent relative efficiencies of the proposed estimator $\hat{\mu}_{Y(3)}$ with respect to $\hat{\mu}_{Y(1)}$ and $\hat{\mu}_{Y(2)}$ are respectively given by the formula:

$$PRE(\hat{\mu}_{Y(3)}, \hat{\mu}_{Y(1)}) = \frac{(\sigma_y^2 + \gamma^2)}{(\sigma_y^2 + \alpha\beta)} \times 100 \tag{13}$$

$$PRE(\hat{\mu}_{Y(3)}, \hat{\mu}_{Y(2)}) = \frac{[\sigma_y^2 + \alpha\beta(\gamma^2 + \theta^2)]}{(\sigma_y^2 + \alpha\beta)} \times 100 \tag{14}$$

We have computed the percent relative efficiencies $PRE(\hat{\mu}_{Y(3)}, \hat{\mu}_{Y(i)}), (i = 1,2)$ for different values of $\sigma_y^2, \gamma^2, \alpha$ *and* β and findings are displayed in Tables 1 and 2.

Table 1: The PRE ($\hat{\mu}_{Y(3)}, \hat{\mu}_{Y(1)}$)

σ_Y^2	α	β	γ	PRE
	0.03	0.01	10	499.94
	0.03	0.01	20	1699.80
	0.03	0.01	30	3699.56
25	0.03	0.01	40	6499.22
	0.06	0.02	10	179.98
	0.06	0.02	20	419.96
	0.06	0.02	30	819.92
125	0.06	0.02	40	1379.87
	0.09	0.03	10	144.43
	0.09	0.03	20	277.74
	0.09	0.03	30	499.94
225	0.09	0.03	40	811.01
	0.12	0.04	10	130.75
	0.12	0.04	20	223.04
	0.12	0.04	30	376.87
325	0.12	0.04	40	592.22
	0.15	0.05	10	123.51
	0.15	0.05	20	194.08
	0.15	0.05	30	311.71
425	0.15	0.05	40	476.39
	0.18	0.06	10	119.02
	0.18	0.06	20	176.15
	0.18	0.06	30	271.37
525	0.18	0.06	40	404.68
	0.21	0.07	10	115.97
	0.21	0.07	20	163.96
	0.21	0.07	30	243.94
625	0.21	0.07	40	355.92
	0.24	0.08	10	113.76
	0.24	0.08	20	155.13
	0.24	0.08	30	224.08
725	0.24	0.08	40	320.60
	0.27	0.09	10	112.09
	0.27	0.09	20	148.44
	0.27	0.09	30	209.03
825	0.27	0.09	40	293.85

Table 2: The PRE ($\hat{\mu}_{Y(3)}, \hat{\mu}_{Y(2)}$)

σ_Y^2	θ	α	β	γ	PRE
	100	0.03	0.01	10	221.17
	150	0.03	0.01	20	374.76
	200	0.03	0.01	30	590.73
25	250	0.03	0.01	40	869.10
	100	0.06	0.02	10	196.94
	150	0.06	0.02	20	319.81
	200	0.06	0.02	30	492.59
125	250	0.06	0.02	40	715.29
	100	0.09	0.03	10	221.17
	150	0.09	0.03	20	374.76
	200	0.09	0.03	30	590.73
225	250	0.09	0.03	40	869.10
	100	0.12	0.04	10	249.13
	150	0.12	0.04	20	438.15
	200	0.12	0.04	30	703.96
325	250	0.12	0.04	40	1046.55
	100	0.15	0.05	10	278.19
	150	0.15	0.05	20	504.03
	200	0.15	0.05	30	821.62
425	250	0.15	0.05	40	1230.96
	100	0.18	0.06	10	307.71
	150	0.18	0.06	20	570.97
	200	0.18	0.06	30	941.18
525	250	0.18	0.06	40	1418.34
	100	0.21	0.07	10	337.47
	150	0.21	0.07	20	638.46
	200	0.21	0.07	30	1061.72
625	250	0.21	0.07	40	1607.25
	100	0.24	0.08	10	367.38
	150	0.24	0.08	20	706.27
	200	0.24	0.08	30	1182.83
725	250	0.24	0.08	40	1797.07
	100	0.27	0.09	10	397.37
	150	0.27	0.09	20	774.28
	200	0.27	0.09	30	1304.31
825	250	0.27	0.09	40	1987.45

By keeping the respondents cooperation in mind, we decided to choose α = 0.03(0.03)0.27, β = 0.01(0.01)0.09, γ = 10 (10) 40.

It is observed from Table 1 and 2 that the values of $PRE\left(\hat{\mu}_{Y(3)}, \hat{\mu}_{Y(2)}\right)$ and $PRE\left(\hat{\mu}_{Y(3)}, \hat{\mu}_{Y(1)}\right)$ are greater than 100. It follows that the proposed estimator $\hat{\mu}_{Y(3)}$ (based on standardized variate S^*) is more efficient than the usual additive model estimator $\hat{\mu}_{Y(1)}$ and the estimator $\hat{\mu}_{Y(2)}$ due to Gjestvang and Singh (2009) with substantial gain in efficiency. Table 1 and 2 exhibit that the values of $PRE\left(\hat{\mu}_{Y(3)}, \hat{\mu}_{Y(2)}\right)$ and $PRE\left(\hat{\mu}_{Y(3)}, \hat{\mu}_{Y(1)}\right)$ remain higher if the values of variance σ_y^2 of the sensitive variable *y* is larger. Thus, based on our simulation results, the use of the proposed estimator $\hat{\mu}_{Y(3)}$ over usual additive model estimator and the Gjestvang and Singh (2009) estimator $\hat{\mu}_{Y(2)}$ is recommended for all situations close to Table 1 and 2. It should be mentioned here that the experience is must in real surveys while making a choice of randomization device to be used in practice.

4. Conclusion

This paper illustrates enrichment over the Gjestvang and Singh’s (2009) randomized response model. We have suggested the new additive randomized response model utilizing the prior knowledge of mean *(θ)* and standard deviation *(γ)* of scrambling variable *S*. The proposed model is found to be more efficient both theoretically as well as numerically than the additive randomized response model studied by Gjestvang and Singh (2009) and the additive model due to Himmelfarb and Edgell’s (1980). Thus the proposed randomized response procedure is therefore recommended for its use in practice as an alternative to Gjestvang and Singh’s (2009) model.

Acknowledgements

The authors are heartily thankful to the learned referees and the Associate - Editor for their fruitful and constrictive comments to bring the original manuscript to the present form.

References

1. Ahsanullah M. and Eichhorn B.H. (1988). On estimation of response from scrambled quantitative data. *Pakistan Journal of Statistics, 4(2): 83-91.*

2. Bar –Lev S.K.E., Bobovitch and Boukai B. (2004). A note on Randomized response models for quantitative data. *Metrika, 60: 225-25,.* DOI: 10.1007/s001840300308

3. Eichhorn B.H. and Hayre L.S. (1983). Scrambled randomized response methods for obtaining sensitive quantitative data. *Journal of Statistical Planning and Inference, 7: 307-316,* DOI: 10.1016/0378-3758(83)90002-2

4. Gjestvang C.R. and Singh S. (2006). A new randomized response model. *Journal of Royal Statistical Society, 68: 523-530,* DOI: 10.1111/j.1467-9868.2006.00554.x

5. Gjestvang C.R. and Singh S. (2009). An improved randomized response model: Estimation of mean. *Journal of Applied Statistics, 36(12): 1361-1367,* DOI: 10.1080/02664760802684151

6. Himmelfarb S. and Edgell S.E. (1980). Additive constant model: a randomized response technique for eliminating evasiveness to quantitative response questions. *Psychological Bulletin, 87: 525-530,* DOI: 10.1037/0033-2909.87.3.525

7. Hirano K. (1977). Estimation procedures based on preliminary test, shrinkage technique and information criterion. *Annals of Institutional Statistics and Mathematics, 29: 21-34,* DOI: 10.1007/bf02532771

8. Mehta J.S. and Srinivasan R. (1971). Estimation of the mean by shrinkage to a point. *Journal of American Statistics Association. 66:86-90,* DOI: 10.1080/01621459.1971.10482224

9. Thompson J.R. (1968): Some shrinkage techniques for estimating the mean. *Journal of American Statistics Association. 63:113-122,* DOI: 10.2307/2283832