An improved Quantitative Randomized Response Model

Zawar Hussain[*]
Muhammad Hamraz[*]
Javid Shabbir[*]
[*]Department of Statistics, Quaid-i-Azam University
45320, Islamabad 44000, Pakistan
Phone: 925190642185, E-mail: zhlangah@yahoo.com

Abstract

Using the two responses from a single respondent an improvement of the Bar-Lev et al. (2004) randomized response model has been proposed. The proposed model remains more efficient than Bar-Lev et al. (2004) randomized response model. The relative efficiency of the proposed model with respect to the Bar-Lev et al. (2004) model has been investigated under different situations.

Keywords: Sensitive variable, randomized response model, evasive answer bias, simple random sampling.

1. Introduction

It has been established through many social surveys that erroneous reporting and denial to respond an item are the two main sources of non-sampling errors. A bias creeps into the estimators due to these non-sampling errors and is called evasive answer bias. This bias may lead, seriously, to misleading results when the study characteristic is of stigmatizing nature or socially unacceptable. In an attempt to reduce evasive answer bias and increase respondent cooperation, Warner (1965) suggested an ingenuous method to estimate the proportion of a sensitive characteristic like induced abortion, drug usage, tax evasion, shoplifting, cheating in exam, etc. To maintain the anonymity of the respondents Warner (1965) proposed to use a randomization device such as a deck of cards or a spinner. Greenberg et al. (1971) borrowed the idea and extended it to the estimation of mean of sensitive quantitative variables. A lot of work has been done in this direction and for a thorough discussion interested readers may be referred to Fox and Tracy (1986), Chaudhuri and Mukerjee (1988), Hedayat and Sinha (1991), Singh (2003), Eichhorn and Hayre (1983), Gupta et al. (2002), Gupta and Shabbir (2004), Bar-Lev et al. (2004) (*BBB*), Ryu et al. (2005-2006), Hussain et al. (2007), Odumade and Singh (2009).

It is easy to perceive that the information contents in multiple responses from a single respondent may be more than in a single response. This idea has been explored in many studies (cf. Christofides (2003), Gupta and Shabbir (2006), and Hussain and Shabbir (2008)). Suppose m responses from the i^{th} respondent are obtained through a given randomization device (RR) then we have m different set of responses. Based on these set of responses m unbiased estimators of the parameter of interest may be defined and finally a weighted estimator of these m estimators may be obtained. In addition, if all the m estimators have equal variances then obviously weighted estimator will have smaller variance than all of these m estimators. It is interesting to note that this idea has not been explored in quantitative Randomized Response Models (*RRM*). To explore this idea in quantitative *RRM* we studied the *BBB* model. The reason of studying the *BBB* model is that it is relatively simpler and efficient than many of the existing quantitative *RRM*s. In the sections to follow we present a brief summary of the *BBB* model, our proposal, efficiency condition of our proposal, numerical results and our findings in this study.

2. BBB Model

In the *BBB* model, each respondent is requested to rotate a spinner unobserved by the interviewer. If the spinner stops in the shaded area then the respondent is requested to report the real response on the sensitive variable, say X_i. If the spinner stops in the non-shaded area then the respondent is requested to report the scrambled response, say X_iZ , where Z is any scrambling variable. Let P be the proportion of the shaded area of the spinner and Y_i be the response from the i^{th} respondent then it can be written as

$$Y_i = \alpha_i X_i + (1-\alpha_i) X_i Z_i , \quad (2.1)$$

where $$\alpha_i = \begin{cases} 1 \text{ with probability } P \\ 0 \text{ with probability } (1-P). \end{cases}$$

An unbiased estimator of the population mean μ_x under Simple Random Sampling With Replacement (*SRSWR*) is given by

$$\mu_{x(BBB)} = \frac{1}{n\{(1-P)\mu_z + P\}}\sum_{i=1}^{n} Y_i \,. \qquad (2.2)$$

Its variance is given by

$$V\{\hat{\mu}_{x(BBB)}\} = \frac{\mu_x^2}{n}\{C_x^2 + (1+C_x^2)C_z^2(p)\}, \qquad (2.3)$$

where

$$C_z^2(p) = \frac{P+(1-P)E(Z^2)}{(P+(1-P)\mu_z)^2} - 1,\ C_z^2 = \frac{\sigma_z^2}{\mu_z^2},\ \mu_x = \frac{X}{n},\ X = \sum_{i=1}^{N} X_i \text{ and } C_x^2 = \frac{\sigma_x^2}{\mu_x^2}.$$

3. proposed Scheme

Following the idea of Hussain et al. (2007), in the proposed scheme, each selected respondent is requested to use Bar-Lev et al. (2004) randomizing device twice to report two responses. To obtain the first response the probability of reporting the true response is set at P and for the second response probability of reporting the true response is set at $(1-T)$. That is, the two statements in the *BBB* model for obtaining the first response are

(1) Report the true value of the sensitive variable, say X_i with probability P.

(2) Report the scrambled response $X_i Z$ with probability $(1-P)$.

The statements to obtain the second response are same as above but the probability of reporting the true response is $(1-T)$ instead of P. Let Y_{1i} and Y_{2i} be the two responses from the i^{th} respondent then we may write them as

$$Y_{1i} = \beta_i X_i + (1-\beta_i) Z_i X_i \qquad (3.1) \text{ and}$$

$$Y_{2i} = (1-\gamma_i) X_i + \gamma_i Z_i X_i, \qquad (3.2)$$

where β_i and. γ_i are the Bernoulli random variables with means P and T respectively.

An unbiased estimator of the population mean using first set of responses is given by (2.2) with variance given in (2.3). Similarly from second set of responses we have another unbiased estimator of the population mean defined as

$$\hat{\mu}_{2x} = \frac{\overline{Y}_2}{T\mu_z + (1-T)}, \qquad (3.3)$$

with variance

$$Var(\hat{\mu}_{2x}) = \frac{1}{n}\{\sigma_x^2 + \mu_x^2(1+C_x^2)C_z^2(t)\}, \qquad (3.4)$$

where $C_z^2(t) = \dfrac{TE(Z^2)+(1-T)}{\{T\mu_z + (1-T)\}^2} - 1$.

By imposing the condition that $P+T=1$, variances of the both $\hat{\mu}_{1x}$ and $\hat{\mu}_{2x}$ become equal since $C_z^2(t) = C_z^2(p)$. To take the advantage of equal variances we define the weighted estimator of $\hat{\mu}_{1x}$ and $\hat{\mu}_{2x}$ as

$$\hat{\mu}_{xw} = W_1\hat{\mu}_{1x} + W_2\hat{\mu}_{2x}, (W_1 + W_2 = 1). \qquad (3.5)$$

Now by the applying the variance operator on (3.5) we get

$$Var(\hat{\mu}_{xw}) = W_1^2 Var(\hat{\mu}_{1x}) + W_2^2 Var(\hat{\mu}_{2x}) + 2W_1W_2 Cov(\hat{\mu}_{1x}, \hat{\mu}_{2x}). \qquad (3.6)$$

To find the weights W_1 and W_2, we differentiate $Var(\hat{\mu}_{xw})$ with respect to W_1 and equate it to zero i.e. $\dfrac{\partial Var(\hat{\mu}_{xw})}{\partial W_1} = 0$ and get $W_1 = W_2 = \dfrac{1}{2}$. Thus the variance of $\hat{\mu}_{xw}$ can be written as

$$Var(\hat{\mu}_{xw}) = \frac{1}{2}Var(\hat{\mu}_{1x}) + \frac{1}{2}Cov(\hat{\mu}_{1x}, \hat{\mu}_{2x}), \qquad (3.7) \text{ since}$$

$Var(\hat{\mu}_{1x}) = Var(\hat{\mu}_{2x})$.

In order to have the variance expression of the weighted estimator in terms of the design parameters we need the following lemma.

Lemma 3.1: under the condition that $P+T=1$, the $Cov\left(\hat{\mu}_{1x},\hat{\mu}_{2x}\right)$ is given by

$$Cov\left(\hat{\mu}_{1x},\hat{\mu}_{2x}\right)=Var\left(\hat{\mu}_{1x}\right)+\frac{D}{n\left\{P+(1-P)\mu_z\right\}^2}, \quad (3.8)$$

where

$$D=\left[(1-P)^2\sigma_z^2\left(\mu_x^2+\sigma_x^2\right)-\left(\mu_x^2+\sigma_x^2\right)\left\{P+(1-P)E\left(Z^2\right)-\left\{P+(1-P)\mu_z\right\}^2\right\}\right]$$

proof: By definition

$$Cov\left(\hat{\mu}_{1x},\hat{\mu}_{2x}\right)=\frac{1}{n^2\left\{P+(1-P)\mu_z\right\}^2}\sum_{i=1}^{n}Cov\left(Y_{1i},Y_{2i}\right). \quad (3.9)$$

Now

$$\begin{aligned}Cov\left(Y_{1i},Y_{2i}\right)&=E\left(Y_{1i}Y_{2i}\right)-E\left(Y_{1i}\right)E\left(Y_{2i}\right)\\&=\left(\mu_x^2+\sigma_x^2\right)\left\{(1-P)^2\left(\mu_z^2+\sigma_z^2\right)+2P(1-P)\mu_z+P^2\right\}\\&-\left\{(1-P)^2\mu_z^2\mu_x^2+P^2\mu_x^2+2P(1-P)\mu_x^2\mu_z\right\}\\&=(1-P)^2\sigma_z^2\left(\mu_x^2+\sigma_x^2\right)+\sigma_x^2\left\{P+(1-P)\mu_z\right\}^2\end{aligned}$$

Adding and subtracting $C_z^*(p)$ in the above expression and then using (3.9) we get the final expression for $Cov\left(\hat{\mu}_{1x},\hat{\mu}_{2x}\right)$ as given in (3.8).

Using the Lemma 3.1, it can be easily shown that the variance of the weighted estimator is given by

$$Var\left(\hat{\mu}_{xw}\right)=Var\left(\hat{\mu}_{1x}\right)+\frac{D}{2n\left\{P+(1-P)\mu_z\right\}^2}. \quad (3.10)$$

4. *Efficiency Comparison:*

The proposed model will be efficient than the *BBB* model if,

$$Var\left(\hat{\mu}_{xw}\right)\le Var\left\{\hat{\mu}_{x(BBB)}\right\}$$

Or $$Var\left(\hat{\mu}_{xw}\right)-Var\left\{\hat{\mu}_{x(BBB)}\right\}\le 0,$$

Or $$\frac{\sigma_z^2}{\left(\mu_z-1\right)^2}\ge -1.$$

This is always true.

Numerical efficiency results of the Proposed Model

For different values of $\mu_x,\sigma_x^2,\mu_z,\sigma_z^2$ and P the Relative Efficiency (*RE*) results of the proposed model are given below in Tables 1-7,

Table 1: *RE* of $\hat{\mu}_{xw}$ relative to $\hat{\mu}_{x(BBB)}$ for $\mu_Z=1,\sigma_Z=1$

μ_x	σ_x	P								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.5	0.5	1.033	1.066	1.096	1.122	1.143	1.154	1.151	1.129	1.081
	1.0	1.027	1.053	1.075	1.094	1.106	1.111	1.106	1.087	1.053
	1.5	1.026	1.049	1.07	1.087	1.098	1.102	1.096	1.078	1.047
	2.0	1.025	1.048	1.068	1.084	1.095	1.098	1.092	1.075	1.045
	2.5	1.025	1.048	1.067	1.083	1.094	1.097	1.091	1.074	1.044
1.0	0.5	1.043	1.087	1.132	1.176	1.217	1.25	1.266	1.25	1.176
	1.0	1.033	1.066	1.096	1.122	1.143	1.154	1.151	1.129	1.081
	1.5	1.029	1.057	1.082	1.102	1.117	1.123	1.118	1.098	1.06
	2.0	1.027	1.053	1.075	1.094	1.106	1.111	1.106	1.087	1.053
	2.5	1.026	1.051	1.072	1.089	1.101	1.105	1.099	1.081	1.049

Table 2: *RE* of $\hat{\mu}_{xw}$ relative to $\hat{\mu}_{x(BBB)}$ for $\mu_Z = 2, \sigma_Z = 1$

μ_x	σ_x	P								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.5	0.5	1.033	1.066	1.098	1.128	1.154	1.174	1.183	1.174	1.128
	1.0	1.024	1.047	1.07	1.091	1.109	1.122	1.127	1.118	1.084
	1.5	1.022	1.043	1.064	1.083	1.099	1.111	1.115	1.107	1.076
	2.0	1.021	1.042	1.061	1.08	1.096	1.107	1.111	1.103	1.073
	2.5	1.021	1.041	1.06	1.078	1.094	1.105	1.109	1.101	1.071
1.0	0.5	1.055	1.11	1.164	1.216	1.263	1.303	1.329	1.328	1.263
	1.0	1.033	1.066	1.098	1.128	1.154	1.174	1.183	1.174	1.128
	1.5	1.026	1.053	1.078	1.101	1.121	1.137	1.143	1.134	1.096
	2.0	1.024	1.047	1.07	1.091	1.109	1.122	1.127	1.118	1.084
	2.5	1.022	1.045	1.066	1.086	1.102	1.115	1.12	1.111	1.079

Table 3: *RE* of $\hat{\mu}_{xw}$ relative to $\hat{\mu}_{x(BBB)}$ for $\mu_Z = 2.5, \sigma_Z = 1$

μ_x	σ_x	P								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.5	0.5	1.039	1.078	1.116	1.153	1.186	1.213	1.23	1.227	1.179
	1.0	1.027	1.054	1.081	1.107	1.131	1.15	1.161	1.157	1.12
	1.5	1.025	1.049	1.074	1.098	1.119	1.137	1.147	1.143	1.109
	2.0	1.024	1.048	1.071	1.094	1.115	1.132	1.142	1.138	1.104
	2.5	1.023	1.047	1.07	1.092	1.113	1.13	1.139	1.135	1.102
1.0	0.5	1.071	1.139	1.204	1.265	1.32	1.367	1.401	1.408	1.348
	1.0	1.039	1.078	1.116	1.153	1.186	1.213	1.23	1.227	1.179
	1.5	1.031	1.061	1.091	1.12	1.146	1.168	1.181	1.177	1.136
	2.0	1.027	1.054	1.081	1.107	1.131	1.15	1.161	1.157	1.12
	2.5	1.026	1.051	1.077	1.101	1.123	1.141	1.152	1.148	1.113

Table 4: *RE* of $\hat{\mu}_{xw}$ relative to $\hat{\mu}_{x(BBB)}$ for $\mu_Z = 3, \sigma_Z = 1$

μ_x	σ_x	P								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.5	0.5	1.045	1.09	1.135	1.178	1.217	1.252	1.277	1.282	1.236
	1.0	1.031	1.062	1.093	1.124	1.153	1.179	1.197	1.199	1.162
	1.5	1.028	1.056	1.085	1.113	1.14	1.163	1.18	1.181	1.147
	2.0	1.027	1.054	1.082	1.109	1.135	1.158	1.174	1.175	1.141
	2.5	1.026	1.053	1.08	1.107	1.132	1.155	1.171	1.172	1.139
1.0	0.5	1.086	1.167	1.242	1.311	1.373	1.426	1.466	1.481	1.43
	1.0	1.045	1.09	1.135	1.178	1.217	1.252	1.277	1.282	1.236
	1.5	1.035	1.07	1.105	1.139	1.172	1.2	1.22	1.223	1.183
	2.0	1.031	1.062	1.093	1.124	1.153	1.179	1.197	1.199	1.162
	2.5	1.029	1.058	1.088	1.117	1.145	1.169	1.186	1.188	1.152

Table 5: *RE* of $\hat{\mu}_{xw}$ relative to $\hat{\mu}_{x(BBB)}$ for $\mu_Z = 4, \sigma_Z = 1$

μ_x	σ_x	*P*								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.5	0.5	1.056	1.111	1.165	1.219	1.27	1.317	1.357	1.377	1.345
	1.0	1.037	1.075	1.114	1.153	1.192	1.229	1.261	1.277	1.248
	1.5	1.033	1.068	1.103	1.139	1.175	1.21	1.239	1.254	1.227
	2.0	1.032	1.065	1.099	1.134	1.169	1.203	1.231	1.246	1.219
	2.5	1.031	1.064	1.097	1.132	1.166	1.2	1.228	1.242	1.216
1.0	0.5	1.113	1.213	1.303	1.384	1.455	1.515	1.564	1.592	1.564
	1.0	1.056	1.111	1.165	1.219	1.27	1.317	1.357	1.377	1.345
	1.5	1.042	1.085	1.128	1.172	1.215	1.255	1.289	1.306	1.276
	2.0	1.037	1.075	1.114	1.153	1.192	1.229	1.261	1.277	1.248
	2.5	1.034	1.07	1.107	1.144	1.181	1.217	1.247	1.263	1.235

Table 6: *RE* of $\hat{\mu}_{xw}$ relative to $\hat{\mu}_{x(BBB)}$ for $\mu_Z = 1, \sigma_Z = 1.5$

μ_x	σ_x	*P*								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.5	0.5	1.042	1.085	1.128	1.171	1.209	1.239	1.252	1.234	1.162
	1.0	1.037	1.074	1.11	1.144	1.171	1.189	1.191	1.168	1.11
	1.5	1.036	1.071	1.106	1.136	1.161	1.176	1.176	1.154	1.099
	2.0	1.035	1.07	1.104	1.134	1.158	1.172	1.171	1.149	1.095
	2.5	1.035	1.07	1.103	1.132	1.156	1.17	1.169	1.146	1.093
1.0	0.5	1.048	1.099	1.154	1.211	1.269	1.325	1.37	1.383	1.313
	1.0	1.042	1.085	1.128	1.171	1.209	1.239	1.252	1.234	1.162
	1.5	1.039	1.078	1.116	1.152	1.183	1.204	1.209	1.187	1.124
	2.0	1.037	1.074	1.11	1.144	1.171	1.189	1.191	1.168	1.11
	2.5	1.036	1.073	1.107	1.139	1.165	1.181	1.182	1.159	1.103

Table7: RE of $\hat{\mu}_{xw}$ relative to $\hat{\mu}_{x(BBB)}$ for $\mu_Z = 1, \sigma_Z = 2$

μ_x	σ_x	*P*								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.5	0.5	1.046	1.095	1.146	1.198	1.25	1.296	1.328	1.327	1.25
	1.0	1.043	1.087	1.132	1.176	1.217	1.25	1.266	1.25	1.176
	1.5	1.042	1.085	1.128	1.17	1.208	1.238	1.25	1.232	1.161
	2.0	1.041	1.084	1.126	1.168	1.205	1.233	1.244	1.225	1.155
	2.5	1.041	1.083	1.126	1.167	1.203	1.231	1.241	1.222	1.152
1.0	0.5	1.05	1.104	1.163	1.226	1.294	1.364	1.429	1.471	1.429
	1.0	1.046	1.095	1.146	1.198	1.25	1.296	1.328	1.327	1.25
	1.5	1.044	1.09	1.137	1.184	1.228	1.265	1.285	1.273	1.197
	2.0	1.043	1.087	1.132	1.176	1.217	1.25	1.266	1.25	1.176
	2.5	1.042	1.086	1.13	1.173	1.212	1.242	1.256	1.238	1.166

5. Conclusions

In the present investigation we have used Bar-Lev et al (2004) twice, and for different values of $\mu_x, \sigma_x^2, \mu_z, \sigma_z^2$ and P we have found the results of the proposed model in the above tables which shows that the proposed scheme gives better results than the corresponding *BBB* model. The reason is that we have utilized more information in the present investigation. Also by imposing the condition that $P+T=1$, the variances of both devices become equal because in this case we have $C_z^2(t)=C_z^2(p)$. The efficiency is almost same for different values of the parameters and the selection probabilities and remains between 1 and 2, the reason for that is, most of the terms in numerator and denominator are same and they cancel out the effect of each other.

References:

1. *Bar-Lev, S. K., Bobovitch, E. and Boukai, B. (2004) A note on randomized response models for quantitative data. Metrika 60:255-260.*
2. *Chaudhuri, A. & Mukerjee, R. (1988) Randomized response: theory and techniques. Marcel Dekker, Inc., New York.*
3. *Eichhorn, B. H., & Hayre, L. S. (1983) Scrambled randomized response methods for obtaining sensitive quantitative data. J Stat Plan Infer 7:307-316.*
4. *Fox, J. A., & Tracy, P. E. (1986) Randomized response: a method for sensitive surveys. SAGE Publications, London.*
5. *Greenberg, B.G., Kubler, R. R., and Horvitz, D.G. (1971): Applications of RR technique in obtaining quantitative data. J. Am. Stat. Assoc. 66. 243-250.*
6. *Gupta, S., Gupta, B., and Singh, S. (2002): Estimation of sensitivity level of personal interview survey questions. J. Stat. Plann Infer. 100. 239-247.*
7. *Gupta, S., and Shabbir, J. (2004) Sensitivity estimation for personal interview survey Questions. Statistica, anno LXIV, n. 4. 643-653.*
8. *Gupta, S and Shabbir, J. (2006): An alternative to Warner's randomized response Model, Journal of Modern Applied Statistical Methods, 5(2), 328-331.*
9. *Hedayat, A. S., & Sinha, B. K. (1991) Design and inference in finite population sampling. Wiley, New York.*
10. *Hussain, Z., Shabbir, J., and Gupta, S. (2007) An alternative to Ryu. et. al. randomized response model, 10(4), 511-517.*
11. *Odumade, O., & Singh, S. (2009) Improved Bar-Lev, Bobovitch, and Boukai Randomized Response Models, Communication in Statistics- Simulation and Computation, 38:3,473-502.*
12. *Singh, S. (2003) Advanced sampling theory with applications, vol 2. Kluwer, Dordrecht.*
13. *Warner, S. L. (1965) Randomized response: a survey technique for eliminating evasive answer bias. J Am Stat Assoc 60:63-69.*