Biostatistics and Biometrics
Open Access Journal
ISSN: 2573-2633

Research Article
Volume 11 Issue 2 - May 2023
DOI: 10.19080/BBOAJ.2023.11.555808

Biostat Biom Open Access J

An alternative model to the generalized forced quantitative randomized response model that uses data collection

Altaf Ahmad Bhat[1], Tanveer A Tarray[2] and Zahoor A Ganie[3]

[1]*Department of General Requirements, University of Technology and Applied Sciences, Salalah-Oman*

[2]*Department of Mathematical Sciences, Islamic University of Science and Technology , Kashmir – India*

[3]*Department of Electrical Engineering, Islamic University of Science and Technology , Kashmir – India*

Submission: April 13, 2023; **Published:** May 03, 2023

***Corresponding author:** Altaf Ahmad Bhat, Department of General Requirements, University of Technology and Applied Sciences, Salalah-Oman

Abstract

As alternatives to the generalized forced quantitative randomized response model that is provided in this research, we have offered two other randomized response models. We have looked into the characteristics of these suggested models and discovered that they are more effective than the ones that are already in use. The previous estimate relied on two independent samples and randomized response models; however, this method raises survey expenses. As a result, we provide two different randomized response models based on a single sample that solves this problem without raising the survey's expenses.

Keywords: Randomized Response Model; Simple Random Sampling With Replacement (SRSWR); Mean Squared Error; Scrambled Response; Sensitive Variables.

AMS Subject Classification: 62D05

Introduction

Politicians and social scientists are regularly interested in gathering data on stigmatizing issues, and many surveys' aim parameter ends up being the population fraction of people who have a sensitive trait. Sometimes respondents are asked delicate questions, including as on their normatively charged beliefs and if they engage in humiliating behaviors like browsing pornographic websites or criminal ones like tax fraud. In certain surveys, participants may be questioned whether they have ever had an abortion, about their history of drug misuse, whether they have ever engaged in homosexual behavior, or whether they have ever been infected with AIDS. Respondents may decide not to answer a question or may knowingly give incorrect information if they believe their responses may be used against them. Direct questioning techniques may result in denials and slanted responses that are likely to result in a significant underestimating of . The randomized response method (RRT), developed by Warner in 1965, is frequently used by researchers since it has been shown in several field studies to significantly boost respondent participation. These researchers include [1-17], who suggested a multiplicative model to gather data on sensitive quantitative variables like income, tax evasion, the amount of drug used, etc. Think about the population. $\Omega = (\Omega_1, \Omega_2, ...\Omega_N)$ of N units. Let a sample of size n units be drawn from the population Ω using simple random sampling with replacement (SRSWR) design. Let X_i be the real value of the sensitive quantitative variable X under consideration for the i^{th} unit $\Omega_i (i = 1, 2, ..., N)$ of the population Ω . According to the Eichhorn and Hayre model, each respondent in the sample is requested to report the scrambled response $Y_i = ZX_i$, where $X_i (i = 1, 2, ..., n)$ is the real value of the sensitive quantitative variable X for the i^{th} unit of the sample and Z is the scrambling variable, whose distribution is assumed to be known. In other words, $E_R(Z) = \mu_z$ and $V_R(Z) = \sigma_z^2$ are assumed to be known and positive, where E_R and V_R denote the expected value and variance over the randomization device. Then an unbiased estimator of the population mean $\mu_x = \frac{1}{N}\sum_{i=1}^{N} X_i$ of the sensitive quantitative variable X under SRSWR design is given by $\hat{\mu}_{x(EH)} = \frac{1}{n}\sum_{i=1}^{n} Y_i$

with variance

$$V\left(\hat{\mu}_{x(EH)}\right) = \frac{\mu_x^2}{n}\left[C_x^2 + C_z^2\left(1 + C_x^2\right)\right],$$

where $C_z^2 = \sigma_z^2 / \mu_z^2$, and $C_x^2 = \sigma_x^2 / \mu_x^2$.

Existing models 1st and 2nd

An unbiased estimator of the population mean is given by:

$$\hat{\mu}_{x(BBB)} = \frac{1}{n\{(1-p)\mu_z + p\}} \sum_{i=1}^{n} Y_i$$

with variance under SRSWR sampling given by

$$V\left(\hat{\mu}_{x(BBB)}\right) = \frac{\mu_x^2}{n}\left[C_x^2 + (1+C_x^2)C_p^2\right]$$

and

$$\hat{\mu}_{x(F)} = \frac{\left(\frac{1}{n}\sum_{i=1}^{n} Y_i - p_3 F\right)}{(p_1 + p_2 F)}$$

with the variance $\hat{\mu}_{x(F)}$ of is given by

$$V\left(\hat{\mu}_{x(F)}\right) = \frac{\mu_x^2}{n}\left[\frac{\left(1+C_x^2\right)\left\{p_1 + p_2\mu_z^2\left(1+C_z^2\right)\right\}}{(p_1+p_2\mu_z)^2} - 1 + \frac{p_3\left\{(1-p_3)F^2 - 2F(p_1+p_2\mu_z)\mu_x\right\}}{(p_1+p_2\mu_z)^2\mu_x^2}\right],$$

The Suggested Randomized Response Model

Observing the optimum value of F, we select two appropriate values of F as

$$\hat{F}_{(1)} = \frac{(p_1 + p_2\mu_z)}{(1-p_3)}, \quad (\ \mu_z \text{ is known})$$

and $\hat{F}_{(2)} = \dfrac{(p_1 + p_2 Z)}{(1-p_3)}$.

Thus, we have suggested randomized response model.

In the suggested model RRM-I, each respondent chosen from the SRSWR sample is given access to a randomization tool, such as a spinner (or deck-D of cards) with three different sorts of statements: (i) Report the true value of the sensitive variable, say X_i ; (ii) Report the scrambled response, $X_i Z$; (iii) Report the scrambled response $\hat{F}_{(1)} = \frac{(p_1 + p_2\mu_z)X_i}{(1-p_3)}$.

with proportion p_1, p_2 and p_3 respectively, such that $p_1 + p_2 + p_3 = 1$ and Z is a scrambling variable which is defined previously (Figure 1).

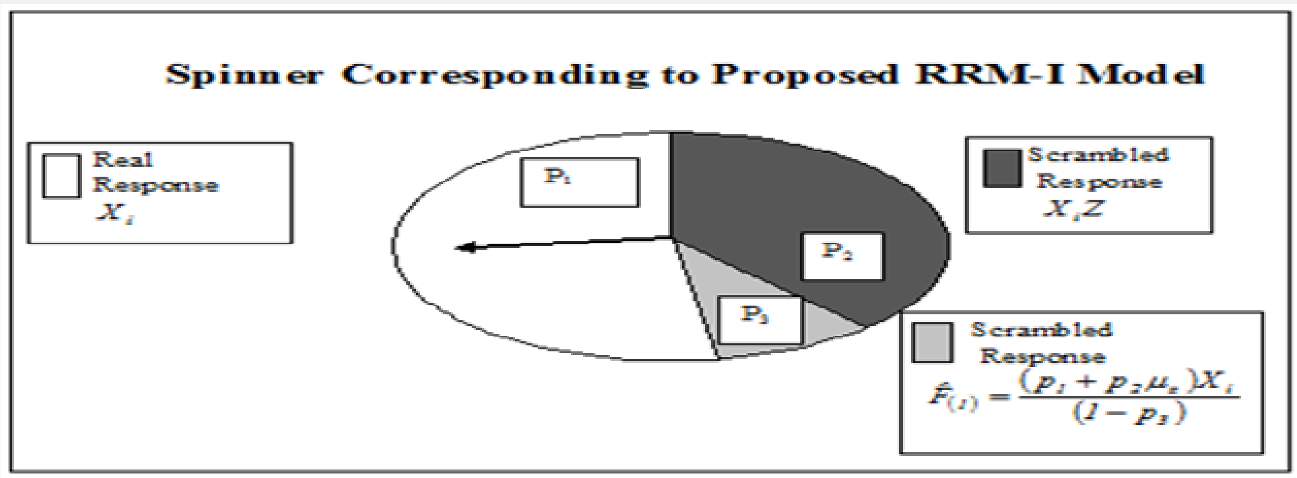

Figure 1: Proposed RRM-I randomized response device.

Mathematically, the distribution of the responses will

$$Y_i = \begin{cases} X_i & \text{with probability } p_1 \\ X_i Z & \text{with probability } p_2 \\ \dfrac{(p_1 + p_2\mu_z)X_i}{(1-p_3)} & \text{with probability } p_3, \end{cases} \quad (14)$$

where X_i denotes the true response, $X_i Z$ denotes the scrambled response and $\frac{(p_1 + p_2\mu_z)X_i}{(1-p_3)}$ also denotes the scrambled response.

Thus we have the following theorems:

Theorem- 5.1- An unbiased estimator of the population mean μ_x is given by

$$\hat{\mu}_{x(1)} = \frac{\frac{1}{n}\sum_{i=1}^{n} Y_i (1-p_3)}{(p_1 + p_2\mu_z)} = \frac{1}{n}\sum_{i=1}^{n}(Y_i/Q)$$

How to cite this article: Altaf Ahmad Bhat,Tanveer A Tarray and Zahoor A Ganie. An alternative model to the generalized forced quantitative randomized response model that uses data collection. Biostat Biom Open Access J. 2023; 11(2): 555808. DOI: 10.19080/BBOAJ.2023.11.555808

where $Q = \frac{(p_1 + p_2\mu_z)}{(1-p_3)}$.

Proof- Let E_p be the expected value over all possible samples and E_R be the expected value over the randomization device. Then the expected of $\hat{\mu}_{x(I)}$ is given by

$$E\left[\hat{\mu}_{x(1)}\right] = E_p E_R\left(\hat{\mu}_{x(1)}\right) = E_p E_R\left(\frac{1}{Qn}\sum_{i=1}^{n} Y_i\right)$$

$$= E_p \frac{1}{Q}\sum_{i=1}^{n} E_R(Y_i)$$

$$= E_p \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{N}\sum_{i=1}^{n} X_i = \mu_x.$$

Thus is an unbiased estimator of the population mean

Theorem-5.2- The variance of the estimator $\hat{\mu}_{x(I)}$ is given by

$$V(\hat{\mu}_{x(I)}) = \frac{\mu_x^2}{n}\left[\frac{(1-p_3)^2(1+C_x^2)\{p_1 + p_2\mu_z^2(1+C_z^2)\}}{(p_1+p_2\mu_z)^2} - 1 + p_3(1+C_x^2)\right].$$

Proof- Let be the variance over all possible samples and be the variance over the randomization device, then we have

$$V(\hat{\mu}_{x(I)}) = E_p V_R(\hat{\mu}_{x(I)}) + V_p E_R(\hat{\mu}_{x(I)})$$

$$= \frac{\mu_x^2}{n}\left[\frac{(1-p_3)(1+C_x^2)\{p_1 + p_2\mu_z^2(1+C_z^2)\}}{(p_1+p_2\mu_z)^2} - 1 + p_3(1+C_x^2)\right]$$

which proves the theorem.

Remark- 5.1- As the quantities $p_1, p_2\ p_3$, and μ_z are known, therefore the proposed randomized response model (RRM-I) and then the proposed estimator can be used in practice without any difficulty.

Special Cases

Case-I: If $p_1 = 0, p_2 = 1, p_3 = 0$ then the proposed model turns out to be the existing model.

Case-II: If $p_1 = p, p_2 = (1-p), p_3 = 0$ then the proposed model becomes one of the already suggested model.

Efficiency Comparison:

This section provides the condition in which the proposed estimator $\hat{\mu}_{x(1)}$ is more efficient than the existing estimators.

From the above equations, we have that

$V(\hat{\mu}_{x(I)}) < V(\hat{\mu}_{x(BBB)})$ if

$$\frac{\left[p_1 + p_2\mu_z^2(1+C_z^2)\right]}{(p_1+p_2\mu_z)^2} < \left[1 + \frac{\{(1-p)\mu_z^2(1+C_z^2) + p\}}{(1-p_3)\{(1-p)\mu_z + p\}^2}\right]$$

It is to be observed from above that

$V(\hat{\mu}_{x(I)}) < min.\, V(\hat{\mu}_{x(F)})$ if

$$\frac{\{1 + (1+C_x^2)(1-p_3)\}}{(1+C_x^2)(1-p_3)(2-p_3)} < \frac{\left[p_1 + p_2\mu_z^2(1+C_z^2)\right]}{(p_1+p_2\mu_z)^2}$$

Thus, the suggested estimator $\hat{\mu}_{x(I)}$ is more efficient than the existing estimator $\hat{\mu}_{x(BBB)}$ and estimator $\hat{\mu}_{x(F)}$ as long as the conditions are satisfied. In order to see the performance of the proposed estimator $\hat{\mu}_{x(I)}$ with respect to $\hat{\mu}_{x(BBB)}$ and $\hat{\mu}_{x(F)}$ by using following formulae:

$$PRE(\hat{\mu}_{x(I)}, \hat{\mu}_{x(BBB)}) = \frac{\left[C_x^2 + C_p^2(1+C_x^2)\right]}{\left[A(1-p_3)^2 + p_3(1+C_x^2) - 1\right]} \times 100,$$

$$PRE(\hat{\mu}_{x(I)}, \hat{\mu}_{x(F)}) = \frac{\left[A - 1 - p_3(1-p_3)^{-1}\right]}{\left[A(1-p_3)^2 + p_3(1+C_x^2) - 1\right]} \times 100,$$

where

$$A = \frac{(1+C_x^2)\left[p_1 + p_2\mu_z^2(1+C_z^2)\right]}{(p_1+p_2\mu_z)^2}.$$

We have taken the parameter values:

(i) $p_1 = p = 0.6, p_2 = 0.3, \mu_z = 2$;

(ii) $p_1 = p = 0.45, p_2 = 0.45, \mu_z = 3$;

$C_x = 0.1(0.2)1.5$,

$C_z = 0.1(0.2)1.5$,

for computing PREs of the proposed estimator $\hat{\mu}_{x(I)}$ with respect to existing models. Findings are given in (Table 1 & 2).

It is observed from Tables 1 and 2 that the gain in efficiency is achieved by using the suggested estimator $\hat{\mu}_{x(I)}$ over the existing estimator $\hat{\mu}_{x(F)}$ is larger.

It is further observed from Tables 1 and 2 that the percent relative efficiency of the suggested estimator $\hat{\mu}_{x(I)}$ with respect to existing are larger than 100% for the selected parametric values as given in (Table 1 & 2).

How to cite this article: Altaf Ahmad Bhat,Tanveer A Tarray and Zahoor A Ganie. An alternative model to the generalized forced quantitative randomized response model that uses data collection. Biostat Biom Open Access J. 2023; 11(2): 555808. DOI: 10.19080/BBOAJ.2023.11.555808

Table 1: The percent relative efficiency of the proposed estimator $\hat{\mu}_{x(I)}$ with respect to existing model.

$p_1 = p = 0.6 \quad p_2 = 0.3 \quad p_3 = 0.0 \quad \mu_z = 2$			$p_1 = p = 0.5 \quad p_2 = 0.5 \quad p_3 = 0.0 \quad \mu_z = 3$		
C_z	C_x	$PRE\left(\hat{\mu}_{x(I)}, \hat{\mu}_{x(BBB)}\right)$	C_z	C_x	$PRE\left(\hat{\mu}_{x(I)}, \hat{\mu}_{x(BBB)}\right)$
0.1	0.1	108.8	0.1	0.1	103.15
	0.3	105.63		0.3	102.55
	0.5	103.56		0.5	101.95
	0.7	102.54		0.7	101.55
	0.9	102		0.9	101.3
	1.1	101.7		1.1	101.15
	1.3	101.52		1.3	101.05
	1.5	101.4		1.5	100.98
0.3	0.1	112.38	0.3	0.1	105.4
	0.3	108.86		0.3	104.54
	0.5	106.07		0.5	103.62
	0.7	104.51		0.7	102.96
	0.9	103.65		0.9	102.53
	1.1	103.15		1.1	102.26
	1.3	102.83		1.3	102.07
	1.5	102.62		1.5	101.95
0.5	0.1	115.55	0.5	0.1	107.96
	0.3	112.45		0.3	107.02
	0.5	109.41		0.5	105.89

Table 2: The percent relative efficiency of the proposed estimator $\hat{\mu}_{x(I)}$ with respect to existing model.

$p_1 = p = 0.6 \quad p_2 = 0.3 \quad p_3 = 0.0 \quad \mu_z = 2$			$p_1 = p = 0.5 \quad p_2 = 0.5 \quad p_3 = 0.0 \quad \mu_z = 3$		
C_z	C_x	$PRE\left(\hat{\mu}_{x(I)}, \hat{\mu}_{x(F)}\right)$	C_z	C_x	$PRE\left(\hat{\mu}_{x(I)}, \hat{\mu}_{x(F)}\right)$
0.1	0.1	122.51	0.1	0.1	123.06
	0.3	118.41		0.3	120.78
	0.5	115.72		0.5	118.5
	0.7	114.4		0.7	116.98
	0.9	113.71		0.9	116.04
	1.1	113.32		1.1	115.46
	1.3	113.08		1.3	115.08
	1.5	112.93		1.5	114.82
0.3	0.1	122.79	0.3	0.1	123.14
	0.3	119.47		0.3	121.23
	0.5	116.84		0.5	119.17
	0.7	115.37		0.7	117.69
	0.9	114.56		0.9	116.74
	1.1	114.08		1.1	116.13
	1.3	113.78		1.3	115.73
	1.5	113.58		1.5	115.45

How to cite this article: Altaf Ahmad Bhat,Tanveer A Tarray and Zahoor A Ganie. An alternative model to the generalized forced quantitative randomized response model that uses data collection. Biostat Biom Open Access J. 2023; 11(2): 555808. DOI: 10.19080/BBOAJ.2023.11.555808

Conclusion

It has been thought about how to estimate the population mean of the sensitive quantitative variable. This study explains how the randomized response model might be enhanced. The unbiased estimate of the mean of a sensitive quantitative variable has been proposed, and the characteristics have been examined. Recommended estimator has been demonstrated quantitatively to be more effective than the existing estimator. It's noteworthy to note that there are no issues using the proposed estimator more effectively in practice.

References

1. Bar Lev SK, Bobovitch E, and Boukai B (2004) A Note on Randomized Response Models For Quantitative Data. Metrika 60: 255-260.
2. Eichhorn BH, Hayre LS (1983) Scrambled Randomized Response Methods For Obtaining Sensitive Quantitative Data. Journal of Statistical Planning and Inference 7(4) : 307-316.
3. Gjestvang CR, Singh S (2006) A new randomized response model. Journal of the Royal Statistical Society Series B 68(3): 523–530.
4. Gjestvang CR, Singh S (2007) Forced quantitative randomized response model: a new device. Metrika 66: 243–257.
5. Sing S, Singh R, Mangat NS (2000) Some alternative strategies to Moor's model in randomized response model. Journal of Statistical Planning and Inference 83(1): 243-255.
6. Singh HP, Tarray TA (2012) A Stratified Unknown repeated trials in randomized response sampling. Communications for Statistical Applications and Methods 19(6): 751-759.
7. Singh HP, Tarray TA (2013) An alternative to Kim and Warde's mixed randomized response model. Statist. Oper. Res. Trans., 37(2): 189-210.
8. Singh HP, Tarray TA (2014) An alternative to stratified Kim and Warde's randomized response model using optimal (Neyman) allocation. Model Assisted Statistics and Applications 9(1): 37-62.
9. Singh HP, Tarray TA (2022) Two - stage stratified partial randomized response strategies. Communications in Statistics - Theory and Methods.
10. Singh HP, Tarray TA (2016) An improved Bar – Lev, Bobovitch and Boukai randomized response model using moments ratios of scrambling variable. Hacettepe Journal of Mathematics and Statistics 45(2): 593-608.
11. Singh S, Sedory SA (2013) A new randomized response device for sensitive characteristics: an application of the negative hypergeometric distribution. Metron 71: 3-8.
12. Tarray TA, Singh HP (2015) A stratified randomized response model for sensitivity characteristics using negative binomial distribution . Investigacion Operacional 36(3): 268-279.
13. Tarray TA, Singh HP (2016) New procedures of estimating proportion and sensitivity using randomized response in a dichotomous finite population. Journal of Model Applied Statistical Methods 15(1): 635-669.
14. Tarray TA, Ganie ZA (2022) A New Data Gathering Exponential Type Ratio Estimator for the Population Mean. Journal of Scientific Research and Reports 28(10): 148-153.
15. Tarray TA, Ganie ZA, Baziga Y (2019) An improved mathematical model applying practicable algorithms. Journal of Applied Science, Engineering, Technology, and Education 1(2): 114-118.
16. Tarray TA, Ganie ZA and Baziga Y (2021) Stratified Optional Mathematical Model. Renewable Power for Sustainable Growth. Springer 723: pp. 307-313.
17. Warner SL (1965) Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. J. Amer. Statist. Assoc., 60(309):63-69.

DOI: 10.19080/BBOAJ.2023.11.555808

How to cite this article: Altaf Ahmad Bhat,Tanveer A Tarray and Zahoor A Ganie. An alternative model to the generalized forced quantitative randomized response model that uses data collection. Biostat Biom Open Access J. 2023; 11(2): 555808. DOI: 10.19080/BBOAJ.2023.11.555808