Trends in

Applied Sciences Research

ISSN 1819-3579

Academic
Journals Inc.

www.academicjournals.com

OPEN ACCESS Trends in Applied Sciences Research

ISSN 1819-3579
DOI: 10.3923/tasr.2019.1.6

Research Article
An Endowed Randomized Response Model for Estimating a Rare Sensitive Attribute Using Poisson Distribution

[1]Tanveer Ahmad Tarray, [2]Housila Prasad Singh and [3]Saadia Masood

[1]Department of Mathematical Sciences, Islamic University of Science and Technology, Awantipora, 192122 Kashmir, India
[2]School of Studies in Statistics, Vikram University Ujjain, India
[3]Department of Mathematics and Statistics, PMAS-University of Arid Agriculture, Rawalpindi, Pakistan

Abstract

Background and Objective: The crux of this paper is to develop a new a distinct question randomized response model for estimating a rare sensitive attribute using Poisson distribution. **Methodology:** The equilibrium point of the model was investigated and a new stratified sampling and stratified sampling randomized response model is proposed. **Results:** It has suggested an unbiased estimator of the mean number of persons possessing the rare sensitive attribute in presence of the unknown proportion of persons possessing a rare unrelated attribute. The addressed issue is resolved by using Lagrange multipliers technique and the optimum allocation is acquired in the form of fuzzy numbers. **Conclusion:** Anew dexterous stratified randomized response model has been proposed and properties of the proposed randomized response model have been studied. Numerical it has shown that the suggested randomized response model is superior one.

Key words: Randomized response technique, estimation of proportion, rare sensitive characteristics, unrelated randomized response technique

Citation: Tanveer Ahmad Tarray, Housila Prasad Singh and Saadia Masood, 2019. An endowed randomized response model for estimating a rare sensitive attribute using poisson distribution. Trends Applied Sci. Res., 14: 1-6.

Corresponding Author: Tanveer Ahmad Tarray, Department of Mathematical Sciences, Islamic University of Science and Technology, Awantipora, 192122 Kashmir, India Tel: +9596560866

Copyright: © 2019 Tanveer Ahmad Tarray *et al.* This is an open access article distributed under the terms of the creative commons attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

The randomized response (RR) data-gathering device to procure trustworthy data on sensitive issues by protecting privacy of the respondent was first developed by Warner[1]. Feeling that the co-operation of the respondent might be further enhanced if one of the two questions referred to a non-sensitive, innocuous attribute, say Y, unrelated to the sensitive attribute A. Horvitz *et al.*[2] proposed an unrelated question randomized response model. Greenberg *et al.*[3] provided theoretical framework for a rectification to the Warner's[1] model envisaged by Horvitz *et al.*[2]. Numerous randomized response techniques have been developed for reducing non sampling errors in sample surveys, protecting a respondent's privacy and increasing response rates. The key references of the randomized response model are Singh and Mathur[4,5], Singh *et al.*[6], Kim and Warde[7], Kim and Elam[8,9] and Singh and Tarray[10,11].

Land *et al.*[12] have considered a different and unique problem where the number of persons possessing a rare sensitive attribute is very small and a huge sample size is required to estimate this number. They proposed a method to estimate the mean of the number of persons possessing a rare sensitive attribute by utilizing the Poisson distribution in survey sampling. Land *et al.*[12] have discussed two different situations that when the proportion of persons possessing a rare unrelated attribute is known and that when it is unknown. Lee *et al.*[13] has extended the studies of Land *et al.*[12] to the stratified sampling.

Singh and Tarray[14] have further considered the problem of estimating the mean of the number of persons possessing a rare sensitive attribute using the Poisson distribution in the situation where the proportion of persons possessing a rare unrelated attribute is known. Singh and Tarray[14] have suggested an alternative randomized response model based on Singh *et al.*[15] model and studied its properties in presence of known population proportion of rare unrelated attributes. The main problem with the use of the method due to Singh and Tarray[14] is that sometimes the mean value of the rare unrelated attribute remains unknown.

ESTIMATION OF PROPORTION OF A RARE SENSITIVE ATTRIBUTE WHEN PROPORTION OF A RARE UNRELATED ATTRIBUTE IS UNKNOWN

Let π_1 be the true proportion of the rare sensitive attribute A_1 in the population U. For example, the proportion of AIDS/HIV patients who continue having affairs with strangers, the proportion of persons who have witnessed a murder, the proportion of persons who are told by the doctors that they will not survive long due to a ghastly disease, for more examples, the reader is referred to Land *et al.*[12]. Consider selecting a large sample of n persons from the population such that as $n \to \infty$ and $\pi_1 \to 0$ then, $\lim (n\,\pi_1) = \delta_1$ (finite). Let π_2 be the true proportion of the population having the rare unrelated attribute A_2 such that as $n \to \infty$ and $\pi_2 \to 0$ then, $\lim (n\,\pi_2) = \delta_2$ (finite and known). For instance, π_2 might be the proportion of persons who are born exactly at 12:00 o'clock, the proportion of babies born blind, see Land *et al.*[12].

In the proposed procedure, each respondent in the sample of n persons, selecting using simple random sampling with replacement (SRSWR) from the given population, is requested to use the deck of cards marked as Deck-I and Deck-II. Each respondent in the sample is requested to use Deck-I consists of three types of cards bearing statements:

- Do you possess the rare sensitive attribute A_1?
- Do you possess the rare unrelated attribute A_2?
- Draw one more card

with probabilities P_1, P_2 and P_3, respectively such that $\sum_{i=1}^{3} P_i = 1$. The respondent is required to draw one card randomly from Deck-I and give answer in term of "Yes" or "No" according to his/her actual status and the statement, (i) or (ii), drawn. However if the statement (iii) is drawn, he/she is required to repeat the above process without replacing that card. If the statement (iii) is drawn in the second phase, he/she is directed to report "No". If m is the total number of cards in the Deck-I, the probability of a "Yes" answer is given by:

$$\theta_1 = [\pi_1 P_1 + P_2 \pi_2] \frac{1 + P_3 m}{m^{-1}} \tag{1}$$

Note that both attributes A_1 and A_2 are very rare in population. Assuming that, as $\to \infty$ and $\theta_1 \to 0$ such that $\lim_{n\to\infty} (n\,\theta_1) = \delta_1^*$ (finite), thus it is clear that:

$$\delta_1^* = [\delta_1 P_1 + P_2 \delta_2] \frac{1 + P_3 m}{m^{-1}} \tag{2}$$

Next, the respondent is requested to use Deck-II consists of three types of cards bearing statements:

- Do you possess the rare sensitive attribute A_1?
- Do you possess the rare unrelated attribute A_2?
- Draw one more card

with probabilities T_1, T_2 and T_3, respectively such that $\sum_{i=1}^{3} T_i = 1$. The respondent is required to draw one card randomly from Deck-II and give answer in term of "Yes" or "No" according to his/her actual status and the statement, (i) or (ii) drawn. However, if the statement (iii) is drawn, he/she is required to repeat the above process without replacing that card. If the statement (iii) is drawn in the second phase, he/she is instructed to report "No". If m is the total number of cards in the Deck - II, the probability of a "Yes" answer is given by:

$$\theta_2 = [T_1\pi_1 + T_2\pi_2]\left[1 + T_3\left(\frac{m}{(m^{-1})}\right)\right] \tag{3}$$

As before assuming that as $\rightarrow \infty$ and $\theta_2 \rightarrow 0$ such that $\lim_{n\to\infty}(n\theta_2) = \delta_2^*$ (finite). Thus it is obvious that:

$$\delta_2^* = [\delta_1 T_1 + T_2\delta_2]\frac{1 + T_3 m}{m^{-1}} \tag{4}$$

By following the procedure as adopted by Singh and Tarray[14], we have:

$$[P_1\hat{\delta}_1 + P_2\hat{\delta}_2]\left[1 + P_3\left(\frac{m}{(m^{-1})}\right)\right] = \frac{1}{n}\sum_{i=1}^{n} y_{1i} \tag{5}$$

and:

$$[T_1\hat{\delta}_1 + T_2\hat{\delta}_2]\left[1 + T_3\left(\frac{m}{(m^{-1})}\right)\right] = \frac{1}{n}\sum_{i=1}^{n} y_{2i} \tag{6}$$

where, y_{1i} and y_{2i} denotes the observed values in the first and the second response from the ith respondent, respectively. Solving Eq. 5 and 6 for $\hat{\delta}$, it has established the following theorems.

Theorem 1: An unbiased estimator of the parameter δ_1 for the rare sensitive attribute A_1 is given by:

$$\hat{\delta}_1 = \frac{1}{n(P_1T_2 - P_2T_1)}\left[\frac{T_2}{P^*}\sum_{i=1}^{n} y_{1i} - \frac{P_2}{T^*}\sum_{i=1}^{n} y_{2i}\right] \tag{7}$$

Where:

$$P^* = \left[1 + P_3\left(\frac{m}{(m^{-1})}\right)\right]$$

and:

$$T^* = \left[1 + T_3\left(\frac{m}{(m^{-1})}\right)\right]$$

Proof: Since y_{1i} ~ iid Poisson (δ^*_1) and y_{2i} ~ iid Poisson (δ^*_2), thus by taking expected value on both sides of the Eq. 7, we have:

$$\begin{aligned}
E(\hat{\delta}_1) &= \frac{1}{n(P_1T_2 - P_2T_1)}\left[\frac{T_2}{P^*}\sum_{i=1}^{n} E(y_{1i}) - \frac{P_2}{T^*}\sum_{i=1}^{n} E(y_{2i})\right] \\
&= \frac{1}{(P_1T_2 - P_2T_1)}\left[\frac{T_2}{P^*}\{P_1\delta_1 + P_2\delta_2\}P^* - \frac{P_2}{T^*}\{T_1\delta_1 + T_2\delta_2\}T^*\right] \\
&= \frac{1}{(P_1T_2 - P_2T_1)}[T_2P_1\delta_1 + T_2P_2\delta_2 - P_2T_1\delta_1 - P_2T_2\delta_2] = \delta_1
\end{aligned}$$

which proves the theorem.

Theorem 2: The variance of the unbiased estimator $\hat{\delta}^*_1$ of the parameter δ^*_1 is given by:

$$V(\hat{\delta}_1) = \frac{1}{n(P_1T_2 - P_2T_1)^2}\left[\begin{matrix}\left\{\frac{P_1T_2^2}{P^*} + \frac{T_1P_2^2}{T^*} - 2T_1T_2P_1P_2\right\}\delta_1 \\ + \left\{\frac{P_2T_2^2}{P^*} + \frac{P_2^2T_2}{T^*} - 2T_2^2P_2^2\right\}\delta_2\end{matrix}\right] \tag{8}$$

Proof: Since y_{1i}~ iid Poisson (δ^*_1) and y_{2i}~ iid Poisson (δ^*_2), thus $V(y_{1i}) = \delta^*_1$ and $V(y_{2i}) = \delta^*_2$. It is to be mentioned that both responses are not independent, thus we have Eq. 9 where:

$$\begin{aligned}
V(\hat{\delta}_1) &= \frac{1}{n^2(P_1T_2 - P_2T_1)^2}\left[\frac{T_2^2}{P^*}\sum_{i=1}^{n} V(y_{1i}) + \frac{P_2^2}{T^*}\sum_{i=1}^{n} V(y_{2i}) - \frac{2T_2P_2}{P^*T^*}\sum_{i=1}^{n}Cov(y_{1i}, y_{2i})\right] \\
&= \frac{1}{n^2(P_1T_2 - P_2T_1)^2}\left[\frac{T_2^2}{P^*}\sum_{i=1}^{n}\delta_1^* + \frac{P_2^2}{T^*}\sum_{i=1}^{n}\delta_2^* - \frac{2T_2P_2}{P^*T^*}\sum_{i=1}^{n}\delta_{12}^*\right]
\end{aligned} \tag{9}$$

$$V(y_{1i}) = E(y_{1i}) = \delta_1^* = (P_1\delta_1 + P_2\,\delta_2)P^* \tag{10}$$

$$V(y_{2i}) = E(y_{2i}) = \delta_2^* = (T_1\delta_1 + T_2\,\delta_2)T^* \tag{11}$$

and $Cov\,V(y_{1i}, y_{2i}) = \delta_{12}^* = E(y_{1i}\,y_{2i}) - E(y_{1i})(y_{2i})$

$$\begin{aligned}
&= [P_1T_1(\delta_1^2 + \delta_1) + P_2T_2(\delta_2^2 + \delta_2) + P_1T_2\delta_1\delta_2 + P_2T_1\delta_1\delta_2 \\
&\quad -(P_1\delta_1 + P_2\delta_2)(T_1\delta_1 + T_2\delta_2)]P^*T^* \\
&= [P_1T_1\,\delta_1 + P_2T_2\delta_2]P^*T^*
\end{aligned} \tag{12}$$

Putting Eq. 10-12 in 9 it established the theorem.

Corollary 1: An unbiased estimator to estimate the parameter δ_2 for rare unrelated attribute A_2 is given by:

$$\hat{\delta}_2 = \frac{1}{n(P_2T_1 - P_1T_2)}\left[\frac{T_1}{P^*}\sum_{i=1}^{n} y_{1i} - \frac{P_1}{T^*}\sum_{i=1}^{n} y_{2i}\right] \tag{13}$$

with the variance:

$$V(\hat{\delta}_2) = \frac{1}{n(P_2T_1 - P_1T_2)^2}\left[\begin{matrix}\left\{\frac{P_1T_1^2}{P^*} + \frac{T_1P_1^2}{T^*} - 2T_1^2P_1^2\right\}\delta_1 \\ + \left\{\frac{P_2T_1^2}{P^*} + \frac{P_1^2T_2}{T^*} - 2P_1P_2T_1T_2\right\}\delta_2\end{matrix}\right] \tag{14}$$

Proof: Analogous to the proof of the theorems 1 and 2.

Corollary 2: An unbiased estimator of the variance of the estimator $\hat{\delta}_1$ is given by:

$$\hat{V}(\hat{\delta}_1)=\frac{1}{n(P_1T_2-P_2T_1)^2}\left[\begin{array}{l}\left\{\frac{P_1T_2^2}{P^*}+\frac{T_1P_2^2}{T^*}-2T_1T_2P_1P_2\right\}\hat{\delta}_1\\+\left\{\frac{P_2T_2^2}{P^*}+\frac{P_2^2T_2}{T^*}-2T_2^2P_2^2\right\}\hat{\delta}_2\end{array}\right] \quad (15)$$

and an unbiased estimator of the variance of the estimator $\hat{\delta}_2$ is given by:

$$\hat{V}(\hat{\delta}_2)=\frac{1}{n(P_2T_1-P_1T_2)^2}\left[\begin{array}{l}\left\{\frac{P_1T_1^2}{P^*}+\frac{T_1P_1^2}{T^*}-2T_1^2P_1^2\right\}\hat{\delta}_1+\\\left\{\frac{P_2T_1^2}{P^*}+\frac{P_1^2T_2}{T^*}-2P_1P_2T_1T_2\right\}\hat{\delta}_2\end{array}\right] \quad (16)$$

where, $\hat{\delta}_1$ and $\hat{\delta}_2$ are, respectively defined in Eq. 7 and 13, respectively.

RELATIVE EFFICIENCY

The percentage of relative efficiency of the proposed estimator $\hat{\delta}_1$ with respect to the Land *et al.*[12] estimator δ_1^* is given by:

$$\mathrm{PRE}(\hat{\delta}_1,\hat{\delta}_L^*)=\frac{(P_1T_2-P_2T_1)^2\left[\begin{array}{l}\{P_1(1-T_1)^2+T_1(1-P_1)^2-\\2P_1T_1(1-P_1)(1-T_1)\}\delta_1\\+\{(1-P_1)(1-T_1)(2-P_1-T_1)\\-2(1-P_1)^2(1-T_1)^2\}\delta_2\end{array}\right]}{(P_1-T_1)^2\left[\begin{array}{l}\left\{\frac{P_1T_2^2}{P^*}+\frac{T_1P_2^2}{T^*}-2T_1T_2P_1P_2\right\}\delta_1+\\\left\{\frac{P_2T_2^2}{P^*}+\frac{P_2^2T_2}{T^*}-2T_2^2P_2^2\right\}\delta_2\end{array}\right]}\times 100 \quad (17)$$

Where:

$$V(\hat{\delta}_L^*)=\frac{1}{n(P_1-T_1)^2}\left\{\begin{array}{l}[P_1(1-T_1)^2+T_1(1-P_1)^2-2P_1T_1(1-P_1)(1-T_1)]\delta_1\\+[(1-P_1)(1-T_1)(2-P_1-T_1)-2(1-P_1)^2(1-T_1)^2]\delta_2\end{array}\right\}$$

See Land *et al.*[12], Eq. 14, p. 7.

It is observed from Eq. 17 that the percentage of relative efficiency of the proposed estimator $\hat{\delta}_1$ with respect to Land *et al.*[12] estimator $\hat{\delta}_L^*$ is free from the sample size n. To see the performance of the proposed estimator $\hat{\delta}_1$ relative to Land *et al.*[12] estimator $\hat{\delta}_L^*$, it has computed the values of PRE ($\hat{\delta}_1$, $\hat{\delta}_L^*$) using the formula given in Eq. 17 for fixed (m =100) and different parametric values as given in Table 1. The resulting values of PRE ($\hat{\delta}_1$, $\hat{\delta}_L^*$) are shown in Table 1.

Table 1 exhibited that the values of PREs are greater than 100 for all the parametric values considered here. Thus the proposed procedure is better than the Land *et al.*[12] procedure.

Table 1: Percentage of relative efficiency of the proposed estimator $\hat{\delta}_1$ with respect to Land *et al.*[12] estimator $\hat{\delta}_L^*$

P_1	P_2	P_3	T_1	T_2	T3	δ_1	δ_2	PRE
0.60	0.20	0.20	0.10	0.45	0.45	0.50	0.50	135.80
							1.00	143.42
							1.50	148.55
						1.00	0.50	130.41
							1.00	135.80
							1.50	140.03
						1.50	0.50	128.27
							1.00	132.37
							1.50	135.80
0.60	0.20	0.20	0.20	0.40	0.40	0.50	0.50	141.51
							1.00	149.32
							1.50	154.35
						1.00	0.50	135.74
							1.00	141.51
							1.50	145.89
						1.50	0.50	133.39
							1.00	137.86
							1.50	141.51
0.60	0.20	0.20	0.30	0.35	0.35	0.50	0.50	149.23
							1.00	156.81
							1.50	161.45
						1.00	0.50	143.28

Table 1: Continue

P_1	P_2	P_3	T_1	T_2	T3	δ_1	δ_2	PRE
							1.00	149.23
							1.50	153.54
						1.50	0.50	140.77
							1.00	145.50
							1.50	149.23
0.60	0.20	0.20	0.40	0.30	0.30	0.50	0.50	159.52
							1.00	166.10
							1.50	169.87
						1.00	0.50	153.95
							1.00	159.52
							1.50	163.33
						1.50	0.50	151.48
							1.00	156.07
							1.50	159.52

For the choice of δ_1, δ_2 as 0.5, 1.50, the percentage of relative efficiency remains considerably larger than the other two cases, which reveals that it is appropriate to use the rare unrelated attribute Y, one with a mean value greater than that of the rare sensitive attribute A without affecting the cooperation of the respondents in using the suggested randomization device. The choice of (P_i, T_i), I = 1, 2 should be made in such a way that the respondents should not feel that their privacy is threatened, while the difference (P_1T_2-P_2T_1) should be kept large as compared to P_1-T_1. Finally, our recommendation is to use the suggested estimator $\hat{\delta}_1$ in practice.

CONCLUSION

This study advocates the problem where the number of persons possessing a rare sensitive attribute is very small and huge sample size is required to estimate. A more practical situation is discussed, when the proportion of persons possessing a rare unrelated attributes is unknown. Properties of the proposed randomized response model have been studied along with recommendations. Efficiency comparison is worked out to investigate the performance of the suggested procedures. It is interesting to mention that the proposed procedure is superior.

SIGNIFICANCE STATEMENT

This study discovers a new Stratified randomized response model and random sampling is generally obtained by dividing the population into non-overlapping groups called strata and selecting a simple random sample from each stratum. An RR technique using a stratified random sampling gives the group characteristics related to each stratum estimator. Also, stratified sample protect a researcher from the possibility of obtaining a poor sample. This study will help the researchers to uncover the critical areas related to randomized response technique. For the future research, researcher can be considering a new theory for randomized response model.

ACKNOWLEDGMENT

The authors are grateful to the referees for there valuable suggestions.

REFERENCES

1. Warner, S.L., 1965. Randomized response: A survey technique for eliminating evasive answer bias. J. Am. Stat. Assoc., 60: 63-69.
2. Horvitz, D.G., B.V. Shah and W.R. Simmons, 1967. The unrelated question randomized response model. Proceedings of the Social Statistics Section, (SSS'67), American Statistical Association, Washington, pp: 65-72.
3. Greenberg, B.G., A.L.A. Abul-Ela, W.R. Simmons and D.G. Horvitz, 1969. The unrelated question randomized response model: Theoretical framework. J. Am. Statist. Assoc., 64: 520-539.
4. Singh, H.P. and N. Mathur, 2004. Unknown repeated trials in the unrelated question randomized response model. Biometr. J., 46: 375-378.
5. Singh, H.P. and N. Mathur, 2005. Estimation of population mean when coefficient of variation is known using scrambled response technique. J. Stat. Plan. Inference, 131: 135-144.
6. Singh, S., S. Horn, R. Singh and N.S. Mangat, 2003. On the use of modified randomization device for estimating the prevalence of a sensitive attribute. Stat. Trans., 6: 515-522.
7. Kim, J.M. and W.D. Warde, 2004. A stratified Warner's randomized response model. J. Statist. Plann. Inference, 120: 155-165.

8. Kim, J.M. and M.E. Elam, 2005. A two-stage stratified Warner's randomized response model using optimal allocation. Metrika, 61: 1-7.
9. Kim, J.M. and M.E. Elam, 2007. A stratified unrelated question randomized response model. Stat. Pap., 48: 215-233.
10. Singh, H.P. and T.A. Tarray, 2012. A stratified unknown repeated trials in randomized response sampling. Commun. Stat. Applic. Methods, 19: 751-759.
11. Singh, H.P. and T.A. Tarray, 2013. A modified survey technique for estimating the proportion and sensitivity in a dichotomous finite population. Int. J. Adv. Sci. Tech. Res., 3: 459-472.
12. Land, M., S. Singh and S.A. Sedory, 2012. Estimation of a rare sensitive attribute using poisson distribution. Statistics: J. Theor. Applied Stat., 46: 351-360.
13. Lee, G.S., D. Uhm and J.M. Kim, 2013. Estimation of a rare sensitive attribute in a stratified sample using poisson distribution. Statistics: J. Theor. Applied Stat., 47: 575-589.
14. Singh, H.P. and T.A. Tarray, 2014. A dexterous randomized response model for estimating a rare sensitive attribute using Poisson distribution. Stat. Probab. Lett., 90: 42-45.
15. Singh, S., R. Singh, N.S. Mangat and D.S. Tracy, 1994. An alternative device for randomized responses. Statistica, 54: 233-243.