Jour. Ind. Soc. Ag. Statistics
***44** (1), 1992 : 82–87*

Two Stage Randomized Response Sampling Procedure Using Unrelated question

N.S. Mangat
Punjab Agricultural University
Ludhiana - 141 004.
(Received : Sept., 1990)

Summary

The present paper considers an alternative unrelated question randomized response (RR) procedure. Throughout the exposition, the proportion of population belonging to non-sensitive attribute has been assumed to be known. The proposed strategy is generalization of the usual unrelated question RR model with known π_v and is always more efficient than this method.

Key words : Equal probabilities with replacement sampling, Estimation of proportions, Randomized response technique.

Introduction

The randomized response (RR) technique to procure trust-worthy data for estimating the proportion of the population belonging to a sensitive attribute was first introduced by Warner [4]. Feeling that the confidence of the respondent provided by RR technique might be further enhanced if one of the two questions is referred to a non-stigmatized attribute, Horvitz *et al.* [2] developed an alternative procedure. We call it U-model. While developing theory for this U-model, Greenberg *et al.* [1] dealt with both the situations when π_v the proportion of innocuous character in population is known and when π_v is not known.

For the situation when π_v is known, they considered the following estimator of π :

$$\hat{\pi}_u = \frac{\frac{n'}{n} - (1-p)\pi_v}{p}$$

Here n′ is the number of "yes" answers obtained from the n respondents selected by using equal probabilities with replacement sampling. The p and (1–p) are the proportions representing the sensitive and non-sensitive attributes, respectively in the random device.

As n'/n has binomial distribution with parameters (n, θ), the estimator $\hat{\pi}_u$ is, therefore, unbiased and its variance is given by

$$V(\hat{\pi}_u) = \frac{\theta(1-\theta)}{np^2} \quad (1.1)$$

Where θ is the probability of "yes" answer such that

$$\theta = \pi p + (1-p)\pi_v$$

In order to obtain minimum value of variance $V(\hat{\pi}_u)$ Greenberg *et al.* [1] recommended to choose p close to 1 and π_v close to 0 or 1 according as $\pi < .5$ or $> .5$. If $\pi = .5$, then $|\pi_v - .5|$ could be maximum on either side.

In the present study, an attempt has been made to modify the above said U-model to two stage RR procedure. This modification is presented below.

2. *The Two Stage Procedure*

The proposed two stage RR unrelated question strategy works in the same manner with a slight change, as the two stage procedure given by Mangat and Singh [3] for the Warner's model. The difference is that in the random device R_2, the statement (ii) "I do not belong to the sensitive group" is replaced by the statement "I belong to non-sensitive group". The rest of the procedure remains unchanged. Therefore, θ_1, the probability of "yes" answer for each respondent by using this procedure follows on replacing $(1-\pi)$ by π_v in (2.1) of Mangat and Singh [3]. By doing so, one gets

$$\theta_1 = T\pi + (1-T)[\pi p + (1-p)\pi_v] \quad (2.1)$$

Solving it for and then replacing θ_1 by its observed estimate n'/n, we get estimator of π as

$$\hat{\pi}_d = \frac{\frac{n'}{n} - (1-T)(1-p)\pi_v}{T + p(1-T)} \tag{2.2}$$

Where n′ is the number of "yes" answers obtained by using the proposed procedure and π_v is known.

Now, n′/n being the binomial random variable with parameters (n, θ_1), is an unbiased estimator of θ_1. This leads to the results stated in Theorems 2.1 and 2.2, the proofs for which are obvious.

Theorem 2.1 : The estimator $\hat{\pi}_d$ is an unbiased estimator of population proportion π.

Theorem 2.2 : The variance of the estimator $\hat{\pi}_d$ is given by

$$V(\hat{\pi}_d) = \frac{\theta_1(1-\theta_1)}{n\{T+p(1-T)\}^2} \tag{2.3}$$

Where θ_1 has been defined in (2.1).

Now an unbiased estimator of the variance $V(\hat{\pi}_d)$ is obtained in the theorem below.

Theorem 2.3 : An unbiased estimator of the variance $V(\hat{\pi}_d)$ is given by

$$V(\hat{\pi}_d) = \frac{\frac{n'}{n}\left(1-\frac{n'}{n}\right)}{(n-1)\{T+p(1-T)\}^2}$$

Proof : We have

$$E[v(\hat{\pi}_d)] = \frac{E\left(\frac{n'}{n}\right) - E\left(\frac{n'}{n}\right)^2}{(n-1)\{T+p(1-T)\}^2}$$

On using $E\left(\frac{n'}{n}\right) = \theta_1$ and $E\left(\frac{n'}{n}\right)^2 = V\frac{n'}{n} + \theta_1^2$, one gets

$$E[v(\hat{\pi}_d)] = \frac{\theta_1 - \theta_1^2 - V\frac{n'}{n}}{(n-1)\{T+p(1-T)\}^2} = V(\hat{\pi}_d)$$

This proves the theorem.

We now look in to the efficiency aspect of the proposed procedure.

3. *Efficiency Comparison*

The relative efficiency of the proposed estimator $\hat{\pi}_d$ with respect to the usual estimator $\hat{\pi}_u$ is defined as

$$RE = \frac{V(\hat{\pi}_u)}{V(\hat{\pi}_d)}$$

Now the estimator $\hat{\pi}_d$ will be superior to estimator $\hat{\pi}_u$ if the RE defined above is greater than 1 i.e.

$$V(\hat{\pi}_d) < V(\hat{\pi}_u)$$

On substituting the values of variances $V(\hat{\pi}_d)$ and $V(\hat{\pi}_u)$ from (2.3) and (1.1) respectively, the above inequality after some algebraic simplifications, reduces to

$$p(-p\pi + p\pi_v - 2\pi_v + 2p\pi\pi_v + 2\pi_v^2 - 2p\pi_v^2) - T(1-p)$$

$$\left[2p\pi\pi_v - p\pi - \pi_v(1-p) + \pi_v^2(1-2p)\right] < 0$$

On rearranging the terms, the inequality becomes

$$D_4[p - T(1-p)] + p\pi_v(\pi_v - 1) < 0$$

Where

$$D_4 = p\left[-(\pi - \pi_v)^2 + \pi(\pi - 1)\right] + (1-p)\pi_v(\pi_v - 1)$$

The expression D_4 is always negative. As the choice of p is close to 1, the inequality (3.1), therefore, always holds. This leads to the statement given in the theorem below.

Theorem 3.1 : The estimator $\hat{\pi}_d$ based on the proposed two stage strategy will always be more efficient than the estimator $\hat{\pi}_u$ for the original u-model with known π_v.

Remark : For T = 0, the proposed strategy reduces to Greenberg *et al.*'s [1] usual U-model with known π_v.

We now give some numerical results to have an idea of RE achieved by using the proposed procedure.

Some Numerical Results :

The RE of the estimator $\hat{\pi}_d$ with respect to $\hat{\pi}_u$ has been worked out for various values of π by taking different values of T. The optimal values of P and π_v have been chosen following recommendations of Greenberg *et al.* [1] given in Section 1. The results are presented in Table 1 for $\pi <$ .5 only, as the symmetry prevails for $\pi >$ .5. The results obtained showed that the two stage procedure is always more

Table 1. Percent RE of the proposed procedure with respect to usual U-model with known π_v

π	T	Relative Efficiency			
		π_v = .1		π_v = .2	
		p = .7	p = .9	p = .7	p = .9
	.1	108.8	102.2	110.9	103.1
	.3	127.4	106.8	135.4	109.4
.1	.5	147.4	111.4	163.9	116.1
	.7	169.0	116.2	196.8	123.1
	.9	192.0	121.0	235.1	130.5
	.1	107.3	101.8	108.8	102.2
	.3	122.3	105.5	127.4	106.8
.2	.5	138.0	109.2	147.4	111.4
	.7	154.3	113.0	169.0	116.2
	.9	171.3	116.8	192.0	121.0
	.1	106.9	101.8	108.0	102.0
	.3	121.4	105.4	124.8	106.1
.3	.5	136.5	109.1	142.6	110.3
	.7	152.2	112.8	161.4	114.5
	.9	168.7	116.6	181.3	118.9
	.1	107.1	101.9	107.8	102.0
	.3	121.9	105.4	124.3	106.1
.4	.5	137.7	109.7	141.9	110.3
	.7	154.6	113.7	160.6	114.6
	.9	172.2	117.8	180.5	119.0
	.1	107.5	102.1	108.1	102.2
	.3	123.5	106.4	125.2	106.6
.5	.5	141.0	110.9	143.8	111.1
	.7	160.1	115.5	164.0	115.8
	.9	181.1	120.2	185.8	120.6

efficient than the original U-model and the RE goes on increasing as T increases. The efficiency of the proposed procedure can be, therefore, increased by selecting the value of T as large as the respondents are likely to accept.

ACKNOWLEDGEMENT

The author is grateful to the referee for several constructive suggestions which helped in bringing the manuscript to its present form.

REFERENCES

[1] Greenberg, B.G., Abul-Ela, A.L.A., Simmons, W.R. and Horvitz, D.G., 1969. The unrelated question randomized response model : theoretical framework., *J. Amer. Statist. Assoc.*, **64**, 520-539.

[2] Horvitz, D.G., Shah, B.V. and Simmons, W.R., 1967. The unrelated question randomized response model. *Proc. of Social Statistics Section, Amer. Statist. Assoc.*, 65-72.

[3] Mangat, N.S. and Singh, R., 1990. An alternative randomized response procedure. *Biometrika*, **77** (2), 439-42.

[4] Warner, S.L., 1965. Randomized response : a survey technique for eliminating evasive answer bias. *J. Amer. Statist. Assoc.*, **60**, 63-69.