*STATISTICA, anno LXIV, n. 3, 2004*

# IMPROVED ESTIMATION OF POPULATION PROPORTION POSSESSING SENSITIVE ATTRIBUTE WITH UNKNOWN REPEATED TRIALS IN RANDOMIZED RESPONSE SAMPLING

H.P. Singh, N. Mathur

## 1. INTRODUCTION

In many surveys involving human population a serious problem is untruthful reporting and refusal to respond. Such lack of co-operation is owing to sensitivity of certain questions as discerned by the respondent and his desire for privacy. Warner (1965) developed an interviewing procedure, popularly known as randomized response technique (RRT), designed to eliminate evasive answer bias. According to the model, for estimating the population proportion π possessing the sensitive attribute A, a simple random sample of *n* persons is selected with replacement from the population. Each interviewee in the sample is furnished with an identical randomization device when the outcome "I possesses attribute A" occurs with probability *p* while its complement "I do not possess attribute A" occurs with probability (1-*p*). The respondent answers "yes" if the outcome of the randomization device tallies with his/her actual status otherwise he/she answers "no". If π is the proportion of sensitive group 'A' in the population, the probability of "yes" reply will be

$$\theta = p\pi + (1-p)(1-\pi). \tag{1}$$

An unbiased estimator of π, the proportion of population belonging to the sensitive group A, considered by Warner is given by

$$\hat{\pi} = \frac{\{(n'/n) - 1 + p\}}{(2p-1)}, p \neq 0.5 \tag{2}$$

with variance

$$\mathrm{V}(\hat{\pi}) = \frac{\theta(1-\theta)}{n(2p-1)^2}. \tag{3}$$

where $n'/n$ is the proportion of 'yes' answers reported by $n$ individuals in the sample.

Several modification of Warner's (1965) have been suggested by various authors applicable to different situations as cited in Singh (1994).

Singh and Joarder (1997) suggested an alternative randomized response technique known as unknown repeated trial model. In this method, if a respondent belongs to group A, then he/she is requested to repeat the trial in the Warner's (1965) randomization device if in the first trial he/she does not get the statement according to his/her status. The rest of the procedure remains the same. The repetition of the trial is known to the interviewee but remains unknown to the interviewer. Assuming completely truthful reporting by the respondents, the probability of "yes" answer is given by

$$\theta_1 = \pi[p + (1-p)p] + (1-\pi)(1-p) \ . \tag{4}$$

An unbiased estimator of π is given by

$$\hat{\pi}_s = \frac{\{\hat{\theta}_1 - (1-p)\}}{[2p - 1 + p(1-p)]} \ . \tag{5}$$

which is due to Singh and Joarder (1997), where $\hat{\theta}_1$ is the sample proportion of "yes" responses in the procedure suggested by Singh and Joarder (1997).

The variance of $\hat{\pi}_s$ is given by

$$\mathrm{V}(\hat{\pi}_s) = \frac{\theta_1(1-\theta_1)}{n[2p - 1 + p(1-p)]^2} \ . \tag{6}$$

In this paper, motivated by Searls (1964) and Searls and Intarapanich (1990), we have suggested a family of estimators of π. Later estimators for π (based on estimated values) have been proposed with their properties. Numerical illustrations are given to show the merits of the suggested estimators.

## 2. THE SUGGESTED ESTIMATOR

We define a family of estimators of π as

$$\hat{\pi}_\alpha = \alpha \hat{\pi}_s \tag{7}$$

where α is a scalar to be chosen suitably.

The expressions for bias and MSE of $\hat{\pi}_\alpha$ are respectively given by

$$\mathrm{B}(\hat{\pi}_\alpha) = (\alpha - 1)\ \pi \tag{8}$$

and

$$\text{MSE}(\hat{\pi}_\alpha) = \alpha^2 \text{Var}(\hat{\pi}_s) + (\alpha - 1)^2 \pi^2 \tag{9}$$

where Var ($\hat{\pi}_s$) is given at (6).

From (9) it follows that MSE($\hat{\pi}_\alpha$) < Var($\hat{\pi}_s$) if either

$$\frac{\{\pi^2 - \text{Var}(\hat{\pi}_s)\}}{\{\pi^2 + \text{Var}(\hat{\pi}_s)\}} < \alpha < 1 \tag{10}$$

Using (10), we have computed the range of $\alpha$ for different values of $p$, π and $n$ and displayed in Table 2.1.

TABLE 2.1

| | $\pi = 0.05$ | | | | $\pi = 0.1$ | | | |
|---|---|---|---|---|---|---|---|---|
| *p* *n* | 2 | 10 | 20 | 50 | 2 | 10 | 20 | 50 |
| 0.6 | 0~1 | 0~1 | 0~1 | 0~1 | 0~1 | 0~1 | 0~1 | 0~1 |
| 0.7 | 0~1 | 0~1 | 0~1 | 0~1 | 0~1 | 0~1 | 0~1 | 0~1 |
| 0.8 | 0~1 | 0~1 | 0~1 | 0~1 | 0~1 | 0~1 | 0~1 | 0.18~1 |
| 0.9 | 0~1 | 0~1 | 0~1 | 0~1 | 0~1 | 0~1 | 0.01~1 | 0.44~1 |

Table 2.1 clearly indicates that the suggested estimator $\hat{\pi}_\alpha$ is always better than Singh and Joarder (1997) estimator $\hat{\pi}_s$ for full range of α except when ($\pi$=0.10, $p$=0.80, $n$=50) and ($\pi$=0.10, $p$=0.90, $n \geq 20$).

Further, minimization of (9) with respect to α yields the optimum value of α as

$$\alpha = \frac{\pi^2}{\text{Var}(\hat{\pi}_s) + \pi^2} = \alpha_0 \text{(say)} \tag{11}$$

Substitution (11) in (9) yields the minimum MSE of $\hat{\pi}_\alpha$ as

$$\text{min MSE}(\hat{\pi}_\alpha) = \frac{\pi^2 \text{Var}(\hat{\pi}_s)}{\pi^2 + \text{Var}(\hat{\pi}_s)} \tag{12}$$

which is always less than the variance of $\hat{\pi}_s$ i.e.

$$\text{min MSE}(\hat{\pi}_\alpha) < \text{Var}(\hat{\pi}_s)$$

Thus the proposed estimator $\hat{\pi}_\alpha$ is always better than Singh and Joarder (1997) estimator $\hat{\pi}_s$ at its optimum condition.

It is to be mentioned that estimator $\hat{\pi}_\alpha$ with $\alpha = \alpha_0$ can not be used in practice as it is based on unknown parameter π, which is under investigation. This led authors to suggest estimators depend on estimated optimum values of parameter and discuss their properties.

## 3. ESTIMATOR BASED ON ESTIMATED OPTIMUM VALUES

A consistent estimate of optimum $\alpha_0$ at (11) is given by

$$\hat{\alpha}_0^{(1)} = \frac{\hat{\pi}_s^2}{\left[\hat{\pi}_s^2 + \dfrac{\hat{\theta}_1(1-\hat{\theta}_1)}{n\{2p-1+p(1-p)\}^2}\right]}. \tag{13}$$

Thus the resulting estimator of π is given by

$$\hat{\pi}_{\hat{\alpha}_0}^{(1)} = \frac{\hat{\pi}_s^3}{\left[\hat{\pi}_s^2 + \dfrac{\hat{\theta}_1(1-\hat{\theta}_1)}{n\{2p-1+p(1-p)\}^2}\right]}. \tag{14}$$

Replacing $\pi^2$ by $\hat{\pi}_s^2$ and Var($\hat{\pi}_s$) by its unbiased estimator

$$\hat{\text{V}}\text{ar}(\hat{\pi}_s) = \frac{\hat{\theta}_1(1-\hat{\theta}_1)}{(n-1)\{2p-1+p(1-p)\}^2} \tag{15}$$

in (11) we get another consistent estimate of $\alpha_0$ as

$$\hat{\alpha}_0^{(2)} = \frac{\hat{\pi}_s^2}{\left[\hat{\pi}_s^2 + \dfrac{\hat{\theta}_1(1-\hat{\theta}_1)}{(n-1)\{2p-1+p(1-p)\}^2}\right]}. \tag{16}$$

Substitution of (16) in (7) yields another estimator of π as

$$\hat{\pi}_{\hat{\alpha}_0}^{(2)} = \frac{\hat{\pi}_s^3}{\left[\hat{\pi}_s^2 + \dfrac{\hat{\theta}_1(1-\hat{\theta}_1)}{(n-1)\{2p-1+p(1-p)\}^2}\right]}. \tag{17}$$

A more flexible form of the estimator π is given by

$$\hat{\pi}_{\hat{\alpha}_0}^{(k)} = \frac{\hat{\pi}_s^3}{\left[\hat{\pi}_s^2 + \dfrac{k\hat{\theta}_1(1-\hat{\theta}_1)}{n\{2p-1+p(1-p)\}^2}\right]} \tag{18}$$

where $k \geq 0$ is a scalar.

For $k$ = 0, $\hat{\pi}_{\hat{\alpha}_0}^{(k)}$ reduces to Singh and Joarder (1997) estimator $\hat{\pi}_s$ while for

$k$=1 it reduces to the estimator $\hat{\pi}_{\hat{\alpha}_0}^{(1)}$. If we get $k = n/(n-1)$ in (18), then $\hat{\pi}_{\hat{\alpha}_0}^{(k)}$ yields the estimator $\hat{\pi}_{\hat{\alpha}_0}^{(2)}$ in (17).

The exact MSE of an estimator $T$= $\hat{\pi}_{\hat{\alpha}_0}^{(1)}, \hat{\pi}_{\hat{\alpha}_0}^{(2)}, \hat{\pi}_{\hat{\alpha}_0}^{(k)}$ under unknown repeated model is given by

$$\text{MSE}(T) = \sum_{n_1=0}^{n} (T-\pi)^2 \; {}^nC_{n_1} \, \theta_1^{n_1} (1-\theta_1)^{n-n_1} . \tag{19}$$

The percent relative efficiency of an estimator $T$ with respect to Singh and Joarder (1997) estimator $\hat{\pi}_s$ is given by

$$\text{PRE}(T, \hat{\pi}_s) = \frac{\theta_1(1-\theta_1)}{n\{2p-1+p(1-p)\}^2} \left[ \sum_{n_1=0}^{n} (T-\pi)^2 \; {}^nC_{n_1} \, \theta_1^{n_1} (1-\theta_1)^{n-n_1} \right]^{-1} \times 100 . \tag{20}$$

The percent relative efficiencies of different estimators $\hat{\pi}_{\hat{\alpha}_0}^{(1)}$, $\hat{\pi}_{\hat{\alpha}_0}^{(2)}$ and $\hat{\pi}_{\hat{\alpha}_0}^{(k)}$ with respect to $\hat{\pi}_s$ have been computed for different values of $n$, $p$, π and $k$ and presented in Table 3.1.

To obtained the approximate expression of MSE($\hat{\pi}_{\hat{\alpha}_0}^{(k)}$), we write

$$\hat{\theta}_1 = \theta_1(1+\varepsilon) \text{ such that } \mathrm{E}(\varepsilon) = 0 \text{ and } \mathrm{E}(\varepsilon^2) = (1-\theta_1)/n\theta_1$$

Expressing $\hat{\pi}_{\hat{\alpha}_0}^{(k)}$ in terms of $\varepsilon$, we have

$$\hat{\pi}_{\hat{\alpha}_0}^{(k)} = \pi\left(1+\frac{\theta_1}{p^*\pi}\varepsilon\right)\left[1+\frac{k\theta_1(1-\theta_1)}{np^{*2}\pi^2}(1+\varepsilon)\left(1-\frac{\theta_1}{(1-\theta_1)}\varepsilon\right)\left(1+\frac{\theta_1}{p^*\pi}\varepsilon\right)^{-2}\right]^{-1}$$

or $$(\hat{\pi}_{\hat{\alpha}_0}^{(k)} - \pi) \cong \frac{\theta_1\varepsilon}{p^*} - \frac{k\theta_1(1-\theta_1)}{np^{*2}\pi}\left[1+\frac{\theta_1\varepsilon}{p^*\pi}\left\{\frac{(1-2\theta_1)p^*\pi}{\theta_1(1-\theta_1)}-1\right\}\right]$$

where $p^* = [2p-1+p(1-p)]$.

Squaring both sides of the above expression and then taking expectations, we get the MSE of $\hat{\pi}_{\hat{\alpha}_0}^{(k)}$ to terms of order $n^{-2}$ as

$$\text{MSE}(\hat{\pi}_{\hat{\alpha}_0}^{(k)}) = \text{Var}(\hat{\pi}_s) + \frac{k\{\text{Var}(\hat{\pi}_s)\}^2}{\pi^2}\left[k-2\left\{\frac{(1-2\theta_1)p^*\pi}{\theta_1(1-\theta_1)}-1\right\}\right] \tag{21}$$

It follows from (22) that MSE($\hat{\pi}_{\hat{\alpha}_0}^{(k)}$) < Var($\hat{\pi}_s$) if either

$$0 < k < 2\left[\frac{(1-2\theta_1)\,p^*\pi}{\theta_1(1-\theta_1)} - 1\right] \qquad (22)$$

It is observed from Table 3.1 that

*(i)* for $\pi$ =0.05 and $n \leq 150$, the proposed estimator $\hat{\pi}_{\hat{\alpha}_0}^{(k)}$ is more efficient than $\hat{\pi}_s$ with substantial gain in efficiency except when ($n$=100, $k$=15, $p$=0.9) and ($n$=150, $p$=0.9);

*(ii)* for $\pi$ =0.1 and $n \leq 150$, the suggested estimator $\hat{\pi}_{\hat{\alpha}_0}^{(k)}$ is more efficient than $\hat{\pi}_s$ with considerable gain in efficiency except for ($n$=50, $k$=15, $p$=0.8), ($n$=50, $p$=0.9), ($n$=100, $k$=15, $p$=0.7), ($n$=100, $0.8 \leq p \leq 0.9$) and ($n$=150, $0.7 \leq p \leq 0.9$);

TABLE 3.1

*Percent Relative Efficiency of* $\hat{\pi}_{\hat{\alpha}_0}^{(1)}$, $\hat{\pi}_{\hat{\alpha}_0}^{(2)}$, $\hat{\pi}_{\hat{\alpha}_0}^{(k)}$ *with respect to* $\hat{\pi}_s$

| | π = 0.05 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $p$ | Estimators \ $n$ | 2 | 10 | 50 | 100 | 150 | 200 | 300 | 400 | 500 |
| 0.6 | $\hat{\pi}_{\hat{\alpha}_0}^{(1)}$ | 102.40 | 184.11 | 193.41 | 182.76 | 172.39 | 0.63 | 0.42 | 0.31 | 0.25 |
| | $\hat{\pi}_{\hat{\alpha}_0}^{(2)}$ | 102.36 | 191.91 | 195.09 | 183.46 | 172.78 | 0.63 | 0.42 | 0.31 | 0.25 |
| | $\hat{\pi}_{\hat{\alpha}_0}^{(k=5)}$ | 102.32 | 436.59 | 483.74 | 384.57 | 316.51 | 0.63 | 0.42 | 0.31 | 0.25 |
| | $\hat{\pi}_{\hat{\alpha}_0}^{(k=15)}$ | 102.30 | 932.14 | 868.77 | 533.27 | 383.49 | 0.63 | 0.42 | 0.31 | 0.25 |
| 0.7 | $\hat{\pi}_{\hat{\alpha}_0}^{(1)}$ | 112.82 | 180.18 | 179.60 | 160.73 | 145.87 | 0.30 | 0.20 | 0.15 | 0.12 |
| | $\hat{\pi}_{\hat{\alpha}_0}^{(2)}$ | 112.98 | 187.38 | 180.96 | 161.20 | 146.08 | 0.30 | 0.20 | 0.15 | 0.12 |
| | $\hat{\pi}_{\hat{\alpha}_0}^{(k=5)}$ | 112.81 | 383.83 | 366.91 | 258.79 | 200.77 | 0.30 | 0.20 | 0.15 | 0.12 |
| | $\hat{\pi}_{\hat{\alpha}_0}^{(k=15)}$ | 112.63 | 617.13 | 498.13 | 283.08 | 198.79 | 0.30 | 0.20 | 0.15 | 0.12 |
| 0.8 | $\hat{\pi}_{\hat{\alpha}_0}^{(1)}$ | 132.26 | 172.67 | 161.26 | 136.39 | 120.35 | 0.16 | 0.10 | 0.08 | 0.06 |
| | $\hat{\pi}_{\hat{\alpha}_0}^{(2)}$ | 136.56 | 178.70 | 162.22 | 136.64 | 120.44 | 0.16 | 0.10 | 0.08 | 0.06 |
| | $\hat{\pi}_{\hat{\alpha}_0}^{(k=5)}$ | 137.85 | 293.97 | 262.38 | 170.95 | 129.56 | 0.16 | 0.10 | 0.08 | 0.06 |
| | $\hat{\pi}_{\hat{\alpha}_0}^{(k=15)}$ | 137.40 | 353.33 | 290.16 | 160.25 | 112.46 | 0.16 | 0.10 | 0.08 | 0.06 |
| 0.9 | $\hat{\pi}_{\hat{\alpha}_0}^{(1)}$ | 160.21 | 150.74 | 135.26 | 109.68 | 097.08 | 0.08 | 0.05 | 0.04 | 0.03 |
| | $\hat{\pi}_{\hat{\alpha}_0}^{(2)}$ | 183.72 | 154.30 | 135.75 | 109.73 | 097.06 | 0.08 | 0.05 | 0.04 | 0.03 |
| | $\hat{\pi}_{\hat{\alpha}_0}^{(k=5)}$ | 200.01 | 206.08 | 166.52 | 105.36 | 081.37 | 0.08 | 0.05 | 0.04 | 0.03 |
| | $\hat{\pi}_{\hat{\alpha}_0}^{(k=15)}$ | 201.95 | 217.93 | 154.63 | 086.02 | 061.73 | 0.08 | 0.05 | 0.04 | 0.03 |

*Table 3.1 continued*

| $\pi = 0.1$ | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $p$ | $n$ Estimators | 2 | 10 | 50 | 100 | 150 | 200 | 300 | 400 | 500 |
| 0.6 | $\hat{\pi}_{\hat{\alpha}_0}^{(1)}$ | 100.72 | 177.32 | 161.47 | 136.63 | 120.59 | 0.64 | 0.43 | 0.32 | 0.26 |
| | $\hat{\pi}_{\hat{\alpha}_0}^{(2)}$ | 100.61 | 184.29 | 162.42 | 136.87 | 120.68 | 0.64 | 0.43 | 0.32 | 0.26 |
| | $\hat{\pi}_{\hat{\alpha}_0}^{(k=5)}$ | 100.54 | 382.78 | 265.69 | 172.85 | 131.08 | 0.64 | 0.43 | 0.32 | 0.26 |
| | $\hat{\pi}_{\hat{\alpha}_0}^{(k=15)}$ | 100.50 | 680.46 | 300.38 | 164.49 | 115.20 | 0.64 | 0.43 | 0.32 | 0.26 |
| 0.7 | $\hat{\pi}_{\hat{\alpha}_0}^{(1)}$ | 108.30 | 168.01 | 134.78 | 108.72 | 095.91 | 0.31 | 0.21 | 0.15 | 0.12 |
| | $\hat{\pi}_{\hat{\alpha}_0}^{(2)}$ | 107.85 | 173.87 | 135.25 | 108.77 | 095.89 | 0.31 | 0.21 | 0.15 | 0.12 |
| | $\hat{\pi}_{\hat{\alpha}_0}^{(k=5)}$ | 107.26 | 313.11 | 168.25 | 105.39 | 081.17 | 0.31 | 0.21 | 0.15 | 0.12 |
| | $\hat{\pi}_{\hat{\alpha}_0}^{(k=15)}$ | 106.88 | 429.68 | 159.03 | 087.25 | 062.42 | 0.31 | 0.21 | 0.15 | 0.12 |
| 0.8 | $\hat{\pi}_{\hat{\alpha}_0}^{(1)}$ | 123.59 | 155.06 | 112.48 | 091.80 | 084.62 | 0.17 | 0.12 | 0.09 | 0.07 |
| | $\hat{\pi}_{\hat{\alpha}_0}^{(2)}$ | 125.10 | 159.45 | 112.63 | 091.75 | 084.55 | 0.17 | 0.12 | 0.09 | 0.07 |
| | $\hat{\pi}_{\hat{\alpha}_0}^{(k=5)}$ | 124.20 | 234.21 | 112.76 | 072.98 | 059.45 | 0.17 | 0.12 | 0.09 | 0.07 |
| | $\hat{\pi}_{\hat{\alpha}_0}^{(k=15)}$ | 122.73 | 257.00 | 094.75 | 053.86 | 040.17 | 0.17 | 0.12 | 0.09 | 0.07 |
| 0.9 | $\hat{\pi}_{\hat{\alpha}_0}^{(1)}$ | 145.64 | 132.76 | 095.37 | 084.71 | 083.92 | 0.10 | 0.06 | 0.05 | 0.04 |
| | $\hat{\pi}_{\hat{\alpha}_0}^{(2)}$ | 158.45 | 134.93 | 095.30 | 084.59 | 083.83 | 0.10 | 0.06 | 0.05 | 0.04 |
| | $\hat{\pi}_{\hat{\alpha}_0}^{(k=5)}$ | 163.04 | 160.22 | 077.57 | 055.92 | 050.35 | 0.10 | 0.06 | 0.05 | 0.04 |
| | $\hat{\pi}_{\hat{\alpha}_0}^{(k=15)}$ | 159.65 | 155.69 | 057.85 | 035.54 | 028.52 | 0.10 | 0.06 | 0.05 | 0.04 |

*(iii)* as the value of $k$ increases percent relative efficiency (PRE) increases except in few cases; and

*(iv)* as the value of $\pi$ increases PRE decreases when $n \leq 150$.

For large sample size $n$ (*i.e.* $n$>150), the performance of the proposed estimator $\hat{\pi}_{\hat{\alpha}_0}^{(k)}$ is very poor.

Thus the constant '$k$' play an important role in enhancing the efficiency of the suggested estimator $\hat{\pi}_{\hat{\alpha}_0}^{(k)}$ only when sample size $n \leq 150$. Hence we conclude that the suggested estimator $\hat{\pi}_{\hat{\alpha}_0}^{(k)}$ is to be preferred when the sample size $n$ is moderate ($n \leq 150$).

*School of Studies in Statistics,* H.P. SINGH
*Vikram University, India* N. MATHUR

ACKNOWLEDGEMENTS

Authors are thankful to the referee for his valuable suggestions regarding the improvement of the paper.

## REFERENCES

D. T. SEARLS (1964), *The utilization of a known coefficient of variation in the estimation procedure,* "Journal of American Statistical Association", 59, 1225-1226.

S. L.WARNER (1965), *Randomized response : A survey technique for eliminating evasive answer bias,* "Journal of American Statistical Association", 60, 63-69.

D. T. SEARLS, P. INTARAPANICH (1990), *A Note on an Estimator for the Variance that utilises the Kurtosis,* "The American Statistician", 44(4), 295-296.

S. SINGH (1994), *Unrelated question randomized response sampling using continuous distributions,* "Journal of Indian Society of Agricultural Statistics", 46, 349-361.

S. SINGH, A. H. JOARDER (1997), *Unknown repeated trials in randomized response sampling,* "Journal of Indian Society of Agricultural Statistics", 50(1), 103-105.

## RIASSUNTO

*Stimatore per la proporzione di un attributo sensibile nel caso di prove ripetute nel campionamento randomizzato*

Questo lavoro propone uno stimatore per la proporzione $\pi$, con cui si presenta un attributo sensibile, nel caso di modello di prove ripetute affrontato da Singh e Joarder (1977) degli stimatori proposti. Sono studiati la distorsione esatta e l'errore quadratico medio. La superiorità dello stimatore proposto rispetto a quello di Warner (1965) e di Singh e Joarder viene discussa attraverso un esempio numerico. Viene inoltre riportata un'espressione per l'errore quadratico medio.

## SUMMARY

*Improved estimation of population proportion possessing sensitive attribute with unknown repeated trials in randomized response sampling*

This paper proposes an estimator for population proportion $\pi$ possessing sensitive attribute under unknown repeated trials model envisaged by Singh and Joarder (1997). The exact bias and mean square error of the proposed estimators are worked out. The superiority of the suggested estimator over Warner (1965) estimator and Singh and Joarder (1997) estimator have been discussed through numerical illustrations. An approximate expression for MSE is also given.