응용통계연구 제 7권 1호
1994년 3월 pp. 141 - 147

A Stratified Randomized Response Technique

Kihak Hong[1), Junkeun Yum,[2) Hwayoung Lee[3)

ABSTRACT

In the present paper an attempt has been made to develop a stratified randomized response technique when the respondents are selected using simple random sampling without replacement(SRSWOR) as well as simple random sampling with replacement(SRSWR). The conditions under which the proposed technique will be more efficient than the corresponding Warner's technique have been obtained.

1. Introduction

The randomized response(RR) technique to produce trust-worthy data for estimating proportion π of the population possesing sensitive characteristic was first introduced by Warner(1965). In this procedure, each respondent is provided with a randomization device by which he or she chooses one of the two questions "Do you belong to sensitive group A ?" or "Do you belong to nonsensitive group $\overline{A}$?" with respective probabilities p and 1-p(0<p<1) and replies "yes" or "no" to the question chosen. Assuming truthful reporting, clearly the probability of getting "yes" response is

$$\lambda = \pi + (1-\pi)(1-p) \\ = (1-p) + (2p-1)\pi . \tag{1.1}$$

He considered the following maximum likelihood estimator (MLE) of π :

$$\widehat{\pi_w} = (n'/n \ -1+p)/(2p-1) \ , \ p \neq 0.5 . \tag{1.2}$$

where n' is the number of "yes" answers obtained from the n respondents selected by SRSWR .

Warner showed that the estimator $\widehat{\pi_w}$ is unbiased and the variance is given by

$$V(\widehat{\pi_w}) = \pi(1-\pi)/n + p(1-p)/[n(2p-1)^2] . \tag{1.3}$$

1) Department of statistics, Dongshin University, Naju , Chonnam,520-714, Korea.
2) Department of statistics, Dongguk University, Seoul,100-715,Korea.
3) Department of statistics, Dongshin University, Naju , Chonnam,520-714, Korea.

Many researchers have modified and suggested alternative RR procedures applicable to different situations.

Kim & Fleuk(1978) modified the Warner model to the case of selecting the respondents by SRSWOR method .

The estimator $\widehat{\pi_w}$ is unbiased and the variance is given by

$$V(\widehat{\pi_w}) = [\pi(1-\pi)/n] \cdot [(N-n)/(N-1)] + p(1-p)/[n(2p-1)^2] . \qquad (1.4)$$

2. The proposed technique

In this technique, we follow Kim & Fleuk(1978) and utilize the following notation :

$$X_{hi} = \begin{cases} 1 & ith\ respondent\ \ in\ \ stratum\ h \in A_h \\ 0 & otherwise \end{cases}$$

$$Y_{hi} = \begin{cases} 1 & ith\ respondent\ of\ stratum\ h\ selects\ sensitive\ question \\ 0 & otherwise \end{cases}$$

$$Z_{hi} = \begin{cases} 1 & ith\ respondent\ of\ stratum\ h\ say\ yes \\ 0 & otherwise \end{cases}$$

N : *size of population*
N_h : *the population size of saratum h*
A : *numbers of sensitive group of the population*
A_h : *numbers of sensitive group of stratum h*
n_h : *the sample size of stratum h*
$W_h = N_h/N$: *the stratum weight*
$f_h = n_h/n$: *sampling fraction of stratum h*
π : *the population proportion of sensitive group*
π_h : *the population proportion of sensitive group of stratum h*
p : *the probability of selecting the sensitive question at randomizing device*
λ_h : *the probability of getting yes at stratum h*

We assume $f_h = f = \frac{n}{N}$ i.e. $n_h = n(N_h/N)$ and the same structure of the randomization device for every stratum. We ask the Warner-type questions in each stratum .

Then,

$$Z_{hi} = X_{hi} \cdot Y_{hi} + (1-X_{hi})(1-Y_{hi}) \ . \tag{2.1}$$

We also define $Z_h = \sum_i^{n_h} Z_{hi}$, $X_h = \sum_i^{n_h} X_{hi}$ and $Y_h = \sum_i^{n_h} Y_{hi}$.

2-1. Respondents are selected with SRSWR

X_h, Y_h and Z_h are binomially distributed with parameters n_h , π_h, p and λ_h , respectively , where $\lambda_h = \pi_h p + (1-\pi_h)(1-p)$.

The ML estimator $\hat{\pi}_{st}$ of π

$$\hat{\pi}_{st} = \sum W_h \cdot \hat{\pi}_h \tag{2.2}$$

is the unbiased estimator of π_h , where

$$\hat{\pi}_h = [\hat{\lambda}_h - (1-p)]/(2p-1) \quad \text{and} \quad \hat{\lambda}_h = Z_h/n_h \ .$$

THEOREM 2.1 In the case of SRSWR sample the variance of $\hat{\pi}_{st}$ is given by

$$\begin{aligned} V(\hat{\pi}_{st}) &= V(\sum W_h \hat{\pi}_h) = \sum W_h^2 \cdot V(\hat{\pi}_h) \\ &= \frac{1}{n} \sum W_h \cdot \pi_h(1-\pi_h) + \frac{p(1-p)}{n(2p-1)^2} \ . \end{aligned} \tag{2.3}$$

Proof We have

$$\begin{aligned} V(\hat{\pi}_h) &= V[(\frac{Z_h}{n_h} - (1-p))/(2p-1)] \\ &= \frac{1}{n_h^2(2p-1)^2} V(Z_h) \ , \end{aligned}$$

where

$$V(Z_h) = V(\sum Z_{hi}) = E(\sum Z_{hi}^2) + E(\sum_{i \neq j} Z_{hi} Z_{hj}) - [E(\sum Z_{hi})]^2 \ , \tag{2.4}$$

$$E(\sum Z_{hi}^2) = E(\sum Z_{hi}) = n_h(\pi_h p + (1-\pi_h)(1-p)) \ , \tag{2.5}$$

and

$$\begin{aligned} E(\sum_{i \neq j} Z_{hi} Z_{hj}) &= n_h(n_h - 1) \Pr(Z_{hi}=1, Z_{hj}=1) \\ &= n_h(n_h-1)(\pi_h p + (1-\pi_h)(1-p))^2 \ . \end{aligned} \tag{2.6}$$

Using (2.2) and after some simplication , one gets the expression for $V(\hat{\pi}_{st})$ as given by (2.3). This proves the theorem. ▮

2-2 Respondents are selected with SRSWOR

X_h is hyper-geometrically distributed (n_h, N_h, A_h) , and Y_h and Z_h are binomially distributed with parameters p and λ_h, respectively.

The MLE $\hat{\pi}_{st}$ of π

$$\hat{\pi}_{st} = \sum W_h \hat{\pi}_h$$

is the unbiased estimator of π_h , where

$$\hat{\pi}_h = [\hat{\lambda}_h-(1-p)]/(2p-1) \ , \ \text{and} \ \hat{\lambda}_h = Z_h/n_h \ .$$

THEOREM 2.2 In the case of SRSWOR sample the variance of $\hat{\pi}_{st}$ is given by

$$\begin{aligned} V(\hat{\pi}_{st}) &= V(\sum W_h \hat{\pi}_h) = \sum W_h^2 \cdot V(\hat{\pi}_h) \\ &= \sum W_h [\pi_h(1-\pi_h)/n \cdot \frac{1}{N_h-1} \cdot \frac{N_h(N-n)}{N} + \frac{p(1-p)}{n(2p-1)^2}] \qquad (2.7) \\ &\fallingdotseq \sum W_h [\pi_h(1-\pi_h)/n \cdot \frac{(N-n)}{N} + \frac{p(1-p)}{n(2p-1)^2}] \ . \end{aligned}$$

Proof We can follow the proof of theorem 2.1 with the following modification for the equation (2.6).

$$\begin{aligned} E(\sum_{i \neq j} Z_{hi} Z_{hj}) &= n_h(n_h-1)\Pr(Z_{hi}=1, Z_{hj}=1) \\ &= n_h(n_h-1)\left\{ \frac{\pi_h(N_h\pi_h-1)}{(N_h-1)(2p-1)^2} + (1-2p)[2\pi_h \cdot (p-1)+1] \right\} \ . \end{aligned}$$ ▮

2-3. Efficiency comparison

For the case of SRSWR, the proposed estimator $\hat{\pi}_{st}$ will be more efficient than the usual Warner's estimator if $V(\hat{\pi}_{st}) \le V(\widehat{\pi_w})$. BY using the equation (1.3) and (2.3), the above inequality reduces to the theorem 2.3 .

THEOREM 2.3 $$[\pi(1-\pi) - \sum W_h \pi_h (1-\pi_h)]/n \ \geq \ 0 \qquad (2.8)$$

Proof Since $\pi = \sum W_h \pi_h$ and $\sum W_h = 1$,

$$\pi(1-\pi) - \sum W_h \pi_h (1-\pi_h) = \pi - \pi^2 - \sum W_h \pi_h + \sum W_h \pi_h^2$$
$$= \sum W_h \pi_h^2 - \pi^2$$
$$= \sum W_h \pi_h^2 - (\sum W_h \pi)^2$$
$$= \sum W_h (\pi_h - \pi)^2 \geq 0. \quad \blacksquare$$

In (2.8) the equality holds if $\pi = \pi_h$ for all h. The same method is applicable to the case of SRSWOR. By using the equation (1.4) and (2.7), the above inequality reduces approximately to

$$[(N-n)/(nN)] \cdot [\pi(1-\pi) - \sum W_h \pi_h (1-\pi_h)] \geq 0 \qquad (2.9)$$

In (2.9) the inequality is clear by theorem 2.3 and holds the equality if $\pi = \pi_h$ for all h .

3. Concluding Remarks

We can see that the stratified randomized response technique is always more efficient than the Warner's technique if π_h is not equal to π for all h no matter how the respondents are selected .

In the application of Warner's RR technique to the stratified random sampling, it must be noted that the randomization estimate $\hat{\pi}_h$ may have negative values when the stratum size n_h is small.

The following shrinkage-type estimator

$$\hat{\pi}_r = \frac{[\exp(\frac{25\,\hat{\pi}_w}{6} - \frac{25}{12})]}{[1 + \exp(\frac{25\,\hat{\pi}_w}{6} - \frac{25}{12})]}$$

suggested by Raghavarao(1978) shrinks the estimator $\widehat{\pi_w}$ into the admissible range when $\widehat{\pi_w} \neq [0, 1]$ and compares favorably with $\hat{\pi}_w$ when π is 0.1 to 0.4 or 0.6 to 0.9 .

An improvement of the stratified randomized response technique by using the estimator $\widehat{\pi_r}$ instead of $\widehat{\pi_w}$ will be developed before long.

Acknowledgements

We are grateful to the referees for their valuable suggestions and comments on improving this paper.

References

[1] Kim,J.,and Fleuk,J.A.(1978), Modifications of the randomized response technique for sampling without replacement, *Proceedings of Survey Research Section, American Statistical Association,* 346-350.

[2] Cohran,W.G.(1977), *Sampling Techniques,* 3rd ed., John Wiley & Sons, New York.

[3] Raghavarao, D.(1978), On the estimation problem in Warner's RR technique, *Biometrics,* Vol. 34, 87-90.

[4] Warner,S.L.(1965), Randomized response : Asurvey technique for eliminating evasive answer bias, *Journal of American Statistical Association,* Vol. 60, 63-69.

층화 확률화 응답 기법

홍기학[4], 염준근[5], 이화영[6]

요약

범죄의 성향이나 도박, 마약 복용 실태등과 같은 사회적으로나 개인적으로 매우 민감한 문제에 대한 조사에서 세대별 또는 계층별로 상당한 차이가 나는 경우에 단순임의 추출법에 의한 Warner의 확률화 응답 기법보다 효율적인 층화 임의 추출법에 의한 층화 확률화 응답 기법을 제시하고 그 효율성을 증명하였다.

4) (520-714) 전라남도 나주시 동신대학교 자연과학대학 통계학과
5) (100-715) 서울특별시 중구 필동 동국대학교 이과대학 통계학과
6) (520-714) 전라남도 나주시 동신대학교 자연과학대학 통계학과