STATISTICA, anno LXXIV, n. 2, 2014

POST-STRATIFICATION BASED ON A CHOICE OF A RANDOMIZATION DEVICE

Sarjinder Singh and Stephen A. Sedory

Department of Mathematics, Texas A&M University-Kingsville, Kingsville, TX 78363, USA

Housila P. Singh

School of Studies in Statistics, Vikram University, Ujjain-456010, M.P., India

Raghunath Arnab

Department of Statistics, University of Botswana, Botswana; and School of Statistics & Actuarial Sciences, University of Kwa-Zulu Natal, RSA

Inderjit Singh Grewal

Department of Mathematics, Statistics and Physics, Punjab Agricultural University, Ludhiana, India

1. INTRODUCTION

Warner, 1965 proposed an interviewing technique, called Randomized Response, to protect an interviewee's privacy and to reduce a major source of bias (evasive answers or refusing to respond) in estimating the prevalence of sensitive characteristics in surveys of human populations. Warner, 1965 designed a randomization device, for example a spinner or a deck of cards that consists of two mutually exclusive outcomes. In the case of cards, each card has one of the following statements: (i) I possess attribute *A*; (ii) I do not possess attribute *A*. An unbiased and the maximum likelihood estimator of π , the proportion of respondents in the population possessing the attribute *A*, is given by:

$$\hat{\pi}_w = \frac{(n_w / n) - (1 - P)}{2P - 1} \tag{1}$$

where n_w is the number of individuals responding "*Yes*", n is the number of respondents selected by a simple random and with replacement sample (SRSWR), and P is the probability of the statement, "I possess an attribute *A*." The variance of $\hat{\pi}_w$ is given by:

$$V(\hat{\pi}_w) = \frac{\pi(1-\pi)}{n} + \frac{P(1-P)}{n(2P-1)^2} = V_1 \quad \textit{(say)} \tag{2}$$

Greenberg *et al.*, 1969 introduced the idea of an Unrelated Question Model (U-model hereafter) in which the statement (ii) I do not possess attribute *A* in the Warner's model is simply replaced by the statement (ii) I belong to non-sensitive group Y . The membership in group Y is completely unrelated to the membership of being involved in group *A*. Assuming π_y the proportion of persons possessing the non-sensitive attribute Y in the population is known, Greenberg et al, 1969 proposed an unbiased estimator of π as:

$$\hat{\pi}_g = \frac{(n_g/n) - (1-P)\pi_y}{P} \tag{3}$$

where n_g is the number of individuals responding "Yes" out of n respondents. The variance of the estimator $\hat{\pi}_g$ is given by:

$$V(\hat{\pi}_g) = \frac{\theta_g(1-\theta_g)}{nP^2} = V_2 \quad \textit{(say)} \tag{4}$$

where $\theta_g = P\pi + (1-P)\pi_y$. Greenberg *et al.*, 1969 suggested choosing π_y as close to π as possible.

Kuk, 1990 introduced an ingenious randomized response model in which if a respondent belongs to a sensitive group A, then he/she is instructed to use a deck of cards having θ_1 proportion of cards with the statement, "I belong to group A" and the other cards bearing the statement, "I do not belong to group A". If the respondent belongs to non-sensitive group A^c then the respondent is requested to use a different deck of cards having θ_2 proportion of cards with the statement, "I do not belong to group A" and the other cards bearing the statement, "I belong to group A." Assume π be the true proportion of persons belonging to the sensitive group A. Obviously, the probability of 'Yes' answer in the Kuk, 1990 model is given by:

$$\theta_{kuk} = \theta_1\pi + (1-\pi)\theta_2 \tag{5}$$

Further assume a simple random with replacement sample of n respondents is selected from the population, and n_k be the number of observed "Yes" answers. The number of people n_1 that answer "Yes" is binomially distributed with parameters $\theta_{kuk} = \theta_1\pi + (1-\pi)\theta_2$ and n. For the Kuk, 1990 model, an unbiased estimator of the population proportion π is given by:

$$\hat{\pi}_{kuk} = \frac{n_k/n - \theta_2}{\theta_1 - \theta_2}, \quad \theta_1 \neq \theta_2 \tag{6}$$

where n_k is the number of individuals responding "Yes" out of n respondents. The variance of the estimator $\hat{\pi}_k$ is given by

$$V(\hat{\pi}_{kuk}) = \frac{\theta_{kuk}(1-\theta_{kuk})}{n(\theta_1 - \theta_2)^2} = V_3 \quad \textit{(say)} \tag{7}$$

Kuk, 1990 suggested choosing θ_1 and θ_2 such that the distance $|\theta_1 - \theta_2|$ should be as large as possible without jeopardizing the respondents' cooperation.

In the next section, we suggest a new method to use randomization devices in real practice which may increase respondent co-operation and, additionally, may increase in the relative efficiency of the resultant estimator.

2. POST-STATIFICATION BASED ON RANDOMIZATION DEVICE

Experience shows that in every face-to-face interview survey, there are respondents who are unwilling to trust the randomization device at hand but might be willing to trust a different device. Some respondents prefer the Warner model, some prefer the U-model, some prefer the Kuk's model and some would be willing to respond directly without using any randomization device. Suppose we selected a simple random with replacement (SRSWR) sample of n respondents. To every selected respondent in the sample, we ask for his/her choice of method of answering out of the above four options. Assume out of n respondents: n_1 respondents prefer the Warner model, n_2 respondents prefer the U-model, n_3 respondents prefer the Kuk's model and the rest of the n_4 respondents prefer to respond directly without using any randomization device. Suppose x_1 respondents reported 'Yes' out of the n_1 using the Warner's model, x_2 respondents reported 'Yes' out of the n_2 using the U-model, x_3 respondents reported 'Yes' out of the n_3 using the Kuk's model and x_4 respondents reported 'Yes' out of the n_4 using no randomization device. Note that here we know the preference of a respondent, unlike the optional randomization devices studied by Chaudhuri,1985, Mangat and Singh, 1991, Singh and Joarder, 1997 and Gupta *et. al.*, 2002 etc. among others.

Let $\hat{w}_1 = n_1/n$, $\hat{w}_2 = n_2/n$, $\hat{w}_3 = n_3/n$ and $\hat{w}_4 = n_4/n$ be the unbiased estimators of the true population proportions $W_1 = N_1/N$, $W_2 = N_2/N$, $W_3 = N_3/N$ and $W_4 = N_4/N$ who prefer to use the Warner model, U-model, Kuk model and respond directly, respectively. Unfortunately the values of W_h, $h = 1, 2, 3, 4$ remain unknown in a survey, thus their estimators $\hat{w}_h$ are used in constructing the estimators. The suffix h refers to what for will be the h[th] post-stratum. Also let $\pi_1 = A_1/N_1$ be the true proportion of respondents who belong to the group A and prefer the Warner model; $\pi_2 = A_2/N_2$ be the true proportion of respondents who belong to the group A and prefer the U-model; $\pi_3 = A_3/N_3$ be the true proportion of respondents who belong to the group A and prefer the Kuk model and $\pi_4 = A_4/N_4$ be the true proportion of respondents who belong to the group A and prefer to respond without any randomization device.

Now we have the following theorems:

THEOREM 1. An unbiased estimator of the population proportion π is given by:

$$\hat{\pi}_{post} = \sum_{h=1}^{4} \hat{w}_h \hat{\pi}_h \tag{8}$$

where $\hat{\pi}_1 = \dfrac{(x_1/n_1) - (1-P)}{2P-1}$, $\hat{\pi}_2 = \dfrac{(x_2/n_2) - (1-P)\pi_y}{P}$, $\hat{\pi}_3 = \dfrac{(x_3/n_3) - \theta_2}{\theta_1 - \theta_2}$ *and* $\hat{\pi}_4 = \dfrac{x_4}{n_4}$ *such that* $n_h > 0$ *are defined.*

PROOF. Let E_2 denote the expected value for a given partition of n into n_h, $h = 1,2,3,4$ over a given randomization device and E_1 is the expected values over all possible samples of size n, then we have

$$E\left(\hat{\pi}_{post}\right) = E_1E_2\left(\hat{\pi}_{post} \mid n_h > 0\right) = E_1E_2\left[\sum_{h=1}^{4}\hat{w}_h\hat{\pi}_h \mid n_h > 0\right]$$
$$= E_1\left[\sum_{h=1}^{4}\hat{w}_hE_2(\hat{\pi}_h \mid n_h > 0)\right] = E_1\left[\sum_{h=1}^{4}\hat{w}_h\pi_h\right]$$
$$= \left[\sum_{h=1}^{4}\pi_hE_1(\hat{w}_h)\right] = \left[\sum_{h=1}^{4}\pi_hW_h\right] = \pi$$

which proves the theorem.

THEOREM 2. *The variance of the unbiased estimator* $\hat{\pi}_{post}$ *is given by:*

$$V\left(\hat{\pi}_{\text{post}}\right) = \frac{1}{n}\sum_{h=1}^{4}W_h\left[\sigma_h^2 + \left(\pi_h - \pi\right)^2\right] \tag{9}$$

where

$$\sigma_1^2 = \pi_1(1-\pi_1) + P(1-P)\big/(2P-1)^2; \quad \sigma_2^2 = \theta_g^*(1-\theta_g^*)/P^2;$$
$$\sigma_3^2 = \theta_{kuk}^*(1-\theta_{kuk}^*)/(\theta_1-\theta_2)^2 \quad and \quad \sigma_4^2 = \pi_4(1-\pi_4) \qquad ()$$

with

$$\theta_g^* = P\pi_2 + (1-P)\pi_y \quad and \quad \theta_{kuk}^* = \theta_1\pi_3 + (1-\pi_3)\theta_2. \qquad ()$$

PROOF. Let V_2 denote the expected value for a given values of n_h, $h = 1, 2, 3, 4$ over a given randomization device and V_1 is the expected values over all possible samples of size n, then we have

$$V(\hat{\pi}_{post}) = E_1V_2(\hat{\pi}_{post} \mid n_h > 0) + V_1E_2(\hat{\pi}_{post} \mid n_h > 0)$$
$$= E_1V_2\left(\sum_{h=1}^{4}\hat{w}_h\hat{\pi}_h \mid n_h > 0\right) + V_1E_2\left(\sum_{h=1}^{4}\hat{w}_h\hat{\pi}_h \mid n_h > 0\right)$$
$$= E_1\left(\sum_{h=1}^{4}\hat{w}_h^2V_2(\hat{\pi}_h \mid n_h > 0)\right) + V_1\left(\sum_{h=1}^{4}\hat{w}_hE_2(\hat{\pi}_h \mid n_h > 0)\right)$$
$$= E_1\left(\sum_{h=1}^{4}\hat{w}_h^2\frac{\sigma_h^2}{n_h}\right) + V_1\left(\sum_{h=1}^{4}\hat{w}_h\pi_h\right) \qquad ()$$
$$= \frac{1}{n}E_1\left(\sum_{h=1}^{4}\hat{w}_h\sigma_h^2\right) + \left(\sum_{h=1}^{4}\pi_h^2V_1(\hat{w}_h) + \sum_{h}^{4}\sum_{\neq h'=1}^{4}\pi_h\pi_{h'}Cov_1\left(\hat{w}_h, \hat{w}_{h'}\right)\right)$$

$$
\begin{aligned}
&= \frac{1}{n}\left(\sum_{h=1}^{4}\sigma_h^2 E_1(\hat{w}_h)\right) + \frac{1}{n}\left[\sum_{h=1}^{4}\pi_h^2 W_h(1-W_h) - \sum_{h\neq}^{4}\sum_{h=1}^{4} W_h W_{h'}\pi_h\pi_{h'}\right] \\
&= \frac{1}{n}\left(\sum_{h=1}^{4}\sigma_h^2 E_1(\hat{w}_h)\right) + \frac{1}{n}\left[\sum_{h=1}^{4}\pi_h^2 W_h - \left(\sum_{h=1}^{4}\pi_h W_h\right)^2\right] \\
&= \frac{1}{n}\left(\sum_{h=1}^{4}\sigma_h^2 E_1(\hat{w}_h)\right) + \frac{1}{n}\left[\sum_{h=1}^{4}\pi_h^2 W_h - \pi^2\right] \textit{ where } \pi = \sum_{h=1}^{4}\pi_h W_h \\
&= \frac{1}{n}\left(\sum_{h=1}^{4}\sigma_h^2 W_h\right) + \frac{1}{n}\sum_{h=1}^{4} W_h\left(\pi_h - \pi\right)^2
\end{aligned} \qquad ()
$$

which proves the theorem.

Now we have the following corollary:

COROLLARY 1. An estimator of $V(\hat{\pi}_{post})$ is suggested as:

$$\hat{V}\left(\hat{\pi}_{post}\right) = \frac{1}{n}\sum_{h=1}^{4}\hat{w}_h\left[\hat{\sigma}_h^2 + \left(\hat{\pi}_h - \hat{\pi}\right)^2\right] \qquad (10)$$

where $\hat{\sigma}_h^2$ is an unbiased estimator of σ_h^2 based on information in the h^{th} post-stratum.

In the next section, we consider the comparison of the proposed estimator with the other competitors considered above.

3. RELATIVE EFFICIENCY

The percent relative efficiency of the proposed estimator $\hat{\pi}_{post}$ with respect to the Warner model, the U-model and the Kuk model is, respectively, defined as:

$$RE(j) = \frac{V_j}{V(\hat{\pi}_{post})} \times 100\%, \quad j = 1, 2, 3, \qquad (11)$$

In order to compute and investigate the percent relative efficiency values defined in (11), we wrote SAS codes (See Appendix A). In the SAS codes, for demonstration purposes, we choose $P = 0.7$ for the Warner model as well as for the U-model, $\theta_1 = 0.70$ and $\theta_2 = 0.2$ for the Kuk model. For unknown parameter π we used the values 0.1,0.3 and 0.5 with corresponding values of the known parameter π_y chosen to be 0.15, 0.35, and 0.55 respectively. These values were chosen because in practice the proportion of the population possessing the sensitive characteristic is typically moderately lower and one tries to use π_y values close to π (generally by an educated guess). Values of π_h and W_h, $h = 1, 2, 3, 4$ were allowed to range between 0.1 and 0.9 using a step of 0.1 while screening out these assumptions that violates the obvious constraints that $\sum W_h = 1$

and $\sum W_h \pi_h = \pi$. We created a variable 'freq', that, for a particular (π, π_y) pair, records the number of combinations of the π_h and W_h that satisfy the above constraints. Among these we kept these cases where the proposed estimator was more efficient than each of the competitors; i.e. where $RE(i) > 100\%$ for $i = 1, 2, 3$. Summary statistics of W_h , π_h and RE are given in Table 1 for each (π, π_y) pair. Note that the expressions for $RE(i)$, $i = 1, 2, 3$ are free from the value of the sample size.

TABLE 1
Descriptive statistics of various parameters involved in the simulation

Variable	*freq*	Mean	StDev	Minimum	Median	Maximum
$\pi = 0.1$, $\pi_y = 0.15$						
W_1	684	0.0500	0.0000	0.0500	0.0500	0.0500
W_2	684	0.0625	0.0245	0.0500	0.0500	0.1500
W_3	684	0.0500	0.0000	0.0500	0.0500	0.0500
W_4	684	0.8375	0.0245	0.7500	0.8500	0.8500
π_1	684	0.4950	0.2563	0.1000	0.5000	0.9000
π_2	684	0.3477	0.2267	0.1000	0.3000	0.9000
π_3	684	0.4083	0.2435	0.1000	0.4000	0.9000
π_4	684	0.0427	0.0219	0.0000	0.0411	0.1000
RE(1)	684	691.22	11.4300	675.87	689.35	721.04
RE(2)	684	102.37	1.6900	100.09	102.09	106.78
RE(3)	684	369.64	6.1100	361.43	368.64	385.58
$\pi = 0.3$, $\pi_y = 0.35$						
Variable	*freq*	Mean	StDev	Minimum	Median	Maximum
W_1	18880	0.0615	0.0211	0.0500	0.0500	0.1000
W_2	18880	0.1852	0.1254	0.0500	0.1500	0.6000
W_3	18880	0.0852	0.0419	0.0500	0.0500	0.2000
W_4	18880	0.6679	0.1119	0.3000	0.7000	0.8500
π_1	18880	0.4973	0.2582	0.1000	0.5000	0.9000
π_2	18880	0.4491	0.2531	0.1000	0.4000	0.9000
π_3	18880	0.4716	0.2576	0.1000	0.5000	0.9000
π_4	18880	0.2352	0.0955	0.0000	0.2375	0.7667
RE(1)	18880	384.79	32.10	345.77	376.77	477.49
RE(2)	18880	111.29	9.28	100.01	108.98	138.10
RE(3)	18880	229.99	19.19	206.67	225.20	285.39

$\pi = 0.5$, $\pi_y = 0.55$						
Variable	*freq*	Mean	StDev	Minimum	Median	Maximum
W_1	30421	0.0668	0.0217	0.0500	0.0500	0.1500
W_2	30421	0.2069	0.1370	0.0500	0.2000	0.6000
W_3	30421	0.0947	0.0505	0.3000	0.1000	0.2500
W_4	30421	0.6315	0.1198	0.1000	0.6500	0.8500
π_1	30421	0.4997	0.2582	0.1000	0.5000	0.9000
π_2	30421	0.5097	0.2557	0.1000	0.5000	0.9000
π_3	30421	0.4743	0.2575	0.1000	0.5000	0.9000
π_4	30421	0.4989	0.1357	0.0000	0.5000	1.0000
RE(1)	30421	341.59	29.69	306.55	333.49	433.12
RE(2)	30421	111.44	9.69	100.01	108.80	141.30
RE(3)	30421	216.43	18.81	194.23	211.30	174.42

In some practical situations, 10% to 15% non-response is expected in a survey, due to the nature of the question being asked in the interview. It is interesting to note here that if up to 25% of the respondents utilize one of the randomization devices (and presumably responded truthfully) and the remaining 75% of the respondents respond directly (and presumably truthfully) without using any randomization device., then the proposed estimator $\hat{\pi}_{post}$ has been observed, under many circumstances, to be more efficient than each of the three estimators individually. This helps to reduce the problem of non-response in a survey where a reasonably sensitive question is asked and only 10% to 30% non-response is expected due to the nature of the question.

4. DISCUSSION OF THE RESULTS

Note that in Table 1 the values of the descriptive indices (mean, standard deviation, minimum, median and maximum) are not results of a simulation but these are calculated on the basis of the number of cases or frequency "*freq*" obtained from different combinations of levels of parameters considered such that the considered criterion is met. Table 1 shows that for $\pi = 0.1$, $\pi_y = 0.15$, the proposed estimator could perform better than using a randomized response model if 5% of the population prefer to use Warner model, about 6.25% (with a standard deviation of 2.452%) prefer Greenberg et al model, 5% prefer Kuk's model, and about 83.75% (with a standard deviation of 2.452%) prefer to respond directly, and if about 49.503% (with a standard deviation of 25.635%) of those who prefer the Warner model belong to the sensitive group; about 34.766% (with a standard deviation of 22.678%) of those who prefer the U-model, belong to the sensitive group; about 40.833% (with a standard deviation of 24.358%) of those who prefer the Kuk model belong to the sensitive group , and about 4.2702% (with a standard deviation of 2.1914%) of those who prefer to answer directly belong to the sensitive group, then the average percent relative efficiency of the proposed estimator with respect to the Warner's model is 691.22% with a standard deviation of 11.43%, minimum of 675.87% and maximum of 721.04%; while that with respect to the Greenberg et al. model is 102.37% with a standard deviation of 1.69%, minimum of 100.09% and maximum of 106.78%; and finally that with respect to the Kuk's model is 369.64% with a standard deviation of 6.11%, minimum of 361.43% and maximum of 385.58%. These results are based on $freq = 684$ situations where the proposed estimator remains more efficient than the other estimators. By slightly modifying the SAS codes, it can also be observed that there is a total of 13,345 admissible combinations of W_h and π_h, thus in 12,661 cases the proposed estimator was less efficient than all three of the models considered when $\pi = 0.1$ and $\pi_y = 0.15$. In the same way, the rest of the results in Table 1 can be interpreted. It may be worth mentioning that when $\pi = 0.3$ and $\pi_y = 0.35$, there were 2,32,822 situations with admissible combination of W_h and π_h, but only for $freq = 18{,}880$ situations did the proposed post-stratified estimator performs better than all three of the randomized response models considered, and that when $\pi = 0.5$ and $\pi_y = 0.55$, there were 3,84,825 situations with admissible combinations W_h and π_h, but only for $freq = 30{,}421$ situations did the proposed post-stratified estimator performs better than all the three randomized response models considered.

Further note that a variety of randomized response models is available, as may be seen in a recent monograph by Chaudhuri, 2011, so post-stratification can be made based on any number of RR models; however in this demonstration we have limited ourselves to only three models, in addition to the direct response method. We conclude that while we cannot guarantee that the proposed post-stratification model is better from the efficiency point of view in every situation, we can guarantee that more cooperation is expected from the respondents with the proposed method.

A justification for the guarantee follows: assume a customer goes to a shoe-store to buy a pair of shoes. If there is only one brand of shoes is available in the store, it is very likely that the customer will leave the store without buying it. However, if a variety of shoes is available then the customer will start ask simple questions about the difference between the different varieties and shopkeeper has a greater chance to convince the customer to buy a pair of shoes from his store. In the same way, if an interviewer has a variety of randomization devices available at the interview time then a respondent is likely to take interest in learning about the options. We acknowledge that it could be a time consuming task for an interviewer, but there will be a greater chance to convince an interviewee to participate in the survey.

APPENDIX: SAS CODE USED IN PRODUCING RESULTS IN TABLE 1

```
DATA DATA1;
P = 0.7;
TH1 = 0.7;
TH2 = 0.2;
PI=0.5;
PY = PI+0.05;
DO W1 = 0.05 TO 0.9 BY 0.05;
DO W2 = 0.05 TO 0.9 BY 0.05;
DO W3 = 0.05 TO 0.9 BY 0.05;
W4 = 1-W1-W2-W3;
DO PI1 = 0.1 TO 0.9 BY 0.1;
DO PI2 = 0.1 TO 0.9 BY 0.1;
DO PI3 = 0.1 TO 0.9 BY 0.1;
PI4 = (PI-W1*PI1-W2*PI2-W3*PI3)/W4;
```

```
VARW = PI*(1-PI) +P*(1-P)/(2*P-1)**2;
VARWC = PI1*(1-PI1)+P*(1-P)/(2*P-1)**2;
THG = PI*P+(1-P)*PY;
VARG = THG*(1-THG)/P**2;
THGC = PI2*P+(1-P)*PY;
VARGC = THGC*(1-THGC)/P**2;
THK = PI*TH1+(1-PI)*TH2;
VARK = THK*(1-THK)/(TH1-TH2)**2;
THKC = PI3*TH1+(1-PI3)*TH2;
VARKC = THKC*(1-THKC)/(TH1-TH2)**2;
VARDC = PI4*(1-PI4);
COMPC = W1*(PI1-PI)**2+W2*(PI2-PI)**2+W3*(PI3-PI)**2+W4*(PI4-PI)**2;
VARNEW = W1*VARWC + W2*VARGC + W3*VARKC +W4*VARDC+COMPC;
RE_W = VARW*100/VARNEW;
RE_G = VARG*100/VARNEW;
RE_K = VARK*100/VARNEW;
OUTPUT;
END;
END;
END;
END;
END;
END;
DATA DATA2;
SET DATA1;
IF RE_W GT 100 AND RE_G GT 100 AND RE_K GT 100;
IF PI4 GT 0 AND PI4 LT 1.0;
IF W4 GT 0;
DROP THG THK VARW VARG VARK VARNEW THGC THKC VARWC VARGC VARKC VARDC
COMPC TH1 TH2 P;
PROC PRINT DATA = DATA2;
PROC EXPORT DATA=DATA2 OUTFILE='C:\SASDATAFILES\RESULTS3.XLS'
DBMS=EXCEL REPLACE;
RUN;
```

ACKNOWLEDGEMENTS

The authors would like to thank the Editor, a referee and Giovanna Galatà Executive Editor for their very valuable comments which helped a lot in bringing the original manuscript in the present form.

REFERENCES

A. CHAUDHURI (2011). *Randomized response and indirect questioning techniques in surveys.* Chapman & Hall, CRC Press, Boca Raton, FL.

A. CHAUDHURI, R. MUKHERJEE (1985). *Optionally randomized response technique.* Calcutta Statist. Assoc. Bull., 34, pp. 225-229.

B.G. GREENBERG, A.L.A. ABUL-ELA, W.R. SIMMONS, D.G. HORVITZ (1969). *The unrelated question randomized response model -- theoretical framework.* J. Amer. Statist. Assoc., 64, pp. 520--539.

S. GUPTA, B.C. GUPTA, S.SINGH (2002). *Estimation of sensitivity level of personal interview survey questions.* J. Statist. Plann. Infer., 100, pp. 209-220.

A.Y.C. KUK (1990). *Asking sensitive questions indirectly*. Biomerika, 77 no.2, pp. 436-438.

N.S. MANGAT, S. SINGH (1994).*Optional randomized response model*. J. Indian Stat. Assoc., 32, no. 3, pp.71-75.

S. SINGH, A.H. JOARDER (1997). *Optional randomized response technique for a quantitative sensitive character.* Metron, no.4, pp.151-157.

S.L. WARNER (1965). *Randomized response: a survey technique for eliminating evasive answer bias.* J. Amer. Stat. Assoc., 60, pp.63-69.

SUMMARY

Post-stratification based on a choice of a randomization device

In this paper, we use the idea of post-stratification based on the respondents' choice of a particular randomization device in order to estimate the population proportion of a sensitive characteristic. The proposed idea gives full freedom to the respondents and is expected to result in greater cooperation from them as well as to provide some increase in the relative efficiency of the newly proposed estimator.

Keywords: Respondents cooperation, post-stratification, sensitive characteristics, protection and efficiency.