STATISTICA, anno LXXI, n. 3, 2011

AN ALTERNATIVE RANDOMIZED RESPONSE MODEL USING TWO DECKS OF CARDS

S. Abdelfatah, R. Mazloum, S. Singh

1. INTRODUCTION

Randomized Response (RR) techniques first introduced by Warner (1965) provided a way to encourage honest answers on sensitive questions and increase the respondent's willingness for cooperation by maintaining the respondent's privacy through randomizing his response by making use of a random device.

Warner (1965) assumed that a proportion π of the population possessed a sensitive characteristic (A) while the remainder of the population did not possess this characteristic. He developed a model to estimate π without requiring the person to report his actual classification, whether it is A or not-A to the interviewer.

In this procedure, a simple random sample with replacement (SRSWR) of n persons is drawn from the population and each respondent is provided with a random device in order to choose one of two statements of the form:

I am a member of group A "selected with probability P_0"
I am not a member of group A "selected with probability (1-P_0)"

Without revealing to the interviewer which statement has been selected, the respondent is required to answer "Yes" or "No" according to his actual status and the statement chosen. The maximum likelihood estimator of π and its variance were derived.

Mangat and Singh (1990) developed a two-stage randomized response procedure which requires the use of two randomization devices in an attempt to propose a new procedure that is more efficient than the Warner (1965) model. In this method, each interviewee in the SRSWR of n respondents is provided with two random devices. The random device R_1 consists of two statements, namely (*i*) 'I belong to the sensitive group' and (*ii*) 'Go to random device R_2', represented with probabilities T_0 and (1-T_0), respectively. The random device R_2 which uses two statements, namely (*i*) 'I belong to the sensitive group' and (*ii*) 'I don't belong to the sensitive group', with known probabilities P_0 and (1-P_0) respectively, is exactly the same as used by Warner (1965). The maximum likelihood estimator of π and its variance were obtained.

Mangat (1994) developed a randomized response procedure which in addition of being more efficient than both Warner (1965) and Mangat and Singh (1990) models, it has the benefit of simplicity over that of Mangat and Singh (1990). In this procedure, each of *n* respondents assumed to be selected by equal probabilities with replacement sampling, is instructed to say "Yes" if he/she has the attribute *A*. If the respondent doesn't have the attribute *A*, then he/she is required to use the Warner randomization device consisting of two statements: (*i*) I am a member of group *A* "selected with probability P_0" and (*ii*) I am not a member of group *A* "selected with probability (1-P_0)". The maximum likelihood estimator of π and its variance were derived.

Odumade and Singh (OS) (2009) suggested an efficient use of two decks of cards in a randomized response model. Each respondent in a SRSWR of *n* respondents is provided with two decks of cards where deck-I includes the two statements: (a) I belong to group *A* and (b) I don't belong to group *A*, with probabilities *P* and (1-*P*) respectively. Deck-II includes the two statements as in deck-I with probabilities *T* and (1-*T*) respectively. Each respondent is requested to draw two cards simultaneously, one card from each deck of cards, and read the statements in order. The respondent first matches his/her status with the statement written on the card taken from deck-I, and then he/she matches his/her status with the statement written on the card taken from deck-II.

According to this procedure, the responses from the *n* respondents can be classified into a 2×2 contingency table as shown in table 1.

TABLE 1

Classification of the responses from deck-I and deck-II

Responses from deck-I	Responses from deck-II	
	Yes	No
Yes	n_{11}	n_{10}
No	n_{01}	n_{00}

The unknown population proportion π of the respondents belonging to group *A* is estimated by minimizing the squared distance between the observed proportions of (Yes, Yes), (Yes, No), (No, Yes) and (No, No) responses and the true proportions of such responses. Thus an unbiased estimator of the population proportion π is given by:

$$\hat{\pi}_{os} = \frac{1}{2} + \frac{(P+T-1)[n_{11}/n - n_{00}/n] + (P-T)[n_{10}/n - n_{01}/n]}{2[(P+T-1)^2 + (P-T)^2]}, \; P \text{ and } T \neq 0.5 \qquad (1)$$

The variance of the estimator $\hat{\pi}_{os}$ is given by:

$$V(\hat{\pi}_{os}) = \frac{(P+T-1)^2[PT + (1-P)(1-T)] + (P-T)^2[T(1-P) + P(1-T)]}{4n[(P+T-1)^2 + (P-T)^2]^2} - \frac{(2\pi-1)^2}{4n} \qquad (2)$$

where P and $T \neq 0.5$

If T=P=P_0, the variance of the OS estimator becomes the same variance obtained if each respondent was requested to use the Warner (1965) device twice.

An unbiased estimator of the variance of $\hat{\pi}_{os}$ is given by:

$$\hat{V}(\hat{\pi}_{os})=\frac{(P+T-1)^2[PT+(1-P)(1-T)]+(P-T)^2[T(1-P)+P(1-T)]}{4(n-1)[(P+T-1)^2+(P-T)^2]^2}-\frac{(2\hat{\pi}_{os}-1)^2}{4(n-1)} \quad (3)$$

where P and $T \neq 0.5$

An empirical study showed that if the proportion of the sensitive attribute is predominant (or rare) in the sample, the OS (2009) model is expected to perform better than both Warner (1965) and Mangat and Singh (1990) models. But if the proportion of the sensitive attribute is rare in the sample, the OS model is more efficient than the Mangat (1994) model.

In section 2, a new randomized response model is proposed based on two decks of cards. The proposed model differs from the OS model in the design of the decks, where one of the decks uses the same statements used by Odumade and Singh (2009) while the other deck includes forced yes and no statements. It will be shown that the proposed model is more efficient than the OS model.

2. THE PROPOSED PROCEDURE

In this paper, the idea of using two decks of cards proposed by Odumade and Singh (2009) is modified by using a different structure for one of the two decks in an attempt to obtain a more efficient estimator of π (the true proportion of respondents in the population that possess sensitive characteristic A).

According to this method, each interviewee in a SRSWR of n respondents is provided with two decks of cards as shown in figure 1.

Deck (I)
$I \in A$ with probability (W)
$I \in A^c$ with probability (1-W)

Deck (II)
Forced Yes with probability (Q)
Forced No with probability (1-Q)

Figure 1 – Statements of two decks of cards.

Each respondent is requested to draw two cards simultaneously; one card from each of the two decks "Deck (I) and Deck (II)" and read the statements in order.

The respondent first matches his/her actual status with the statement written on the card drawn from Deck (I), and then he/she is supposed to say "Yes" or "No" based on the card drawn from Deck (II) regardless of his/her actual status.

The whole procedure is done completely by the respondent, away from the interviewer.

The proposed procedure can increase the respondents' cooperation as they will be required to answer the question regarding the sensitive attribute only once because the statements in the second deck don't depend on the actual status of the respondent.

Consider a situation in which the selected respondent belongs to group A and his/her response is (Yes, Yes) which happens if the respondent draws the first card with the statement "$I \in A$" with the probability (W) from Deck (I) and the second card with the statement "Yes" with the probability (Q) from Deck (II).

Another situation in which the selected respondent belongs to group A^c and his/her response is also (Yes, Yes) occurs if the respondent draws the first card with the statement "$I \in A^c$" with the probability (1-W) from Deck (I) and the second card with the statement "Yes" with the probability (Q) from Deck (II).

As shown the response (Yes, Yes) can be obtained from respondents either belonging to group A or A^c and hence the confidentiality of the person reporting (Yes, Yes) will not be violated.

– The probability of getting a (Yes, Yes) response is given by:

$$P(Yes, Yes) = \theta_{11} = WQ\pi + (1-W)Q(1-\pi) = (2W-1)Q\pi + (1-W)Q \tag{4}$$

In the same way, the probability of getting a (Yes, No) response is given by:

$$P(Yes, No) = \theta_{10} = (2W-1)(1-Q)\pi + (1-W)(1-Q) \tag{5}$$

The probability of getting a (No, Yes) response is given by:

$$P(No, Yes) = \theta_{01} = (1-2W)Q\pi + WQ \tag{6}$$

And, the probability of getting a (No, No) response is given by:

$$P(No, No) = \theta_{00} = (1-2W)(1-Q)\pi + W(1-Q) \tag{7}$$

The responses from the n respondents can be classified into a (2×2) contingency table as shown in table 2.

TABLE 2

Classification of the responses from the two decks of cards

Responses from Deck (I)	Responses from Deck (II)		
	Yes	No	Σ
Yes	n_{11}	n_{10}	n_{1+}
No	n_{01}	n_{00}	n_{0+}
Σ	n_{+1}	n_{+0}	n

In order to estimate the unknown population proportion π of the respondents belonging to group A, let n_{11}/n, n_{10}/n, n_{01}/n and n_{00}/n be the observed proportions of (Yes, Yes), (Yes, No), (No, Yes) and (No, No) responses and they are unbiased estimators for θ_{11}, θ_{10}, θ_{01} and θ_{00} respectively where $\sum_{i=0}^{1}\sum_{j=0}^{1}\theta_{ij}=1$.

We define the squared distance between the observed proportions and the true proportions as:

$$D=\frac{1}{2}\sum_{i=0}^{1}\sum_{j=0}^{1}(\theta_{ij}-n_{ij}/n)^2$$

$$D=\frac{1}{2}\left[(2W-1)Q\pi+(1-W)Q-\frac{n_{11}}{n}\right]^2+\frac{1}{2}\left[(2W-1)(1-Q)\pi+(1-W)(1-Q)-\frac{n_{10}}{n}\right]^2$$

$$+\frac{1}{2}\left[(1-2W)Q\pi+WQ-\frac{n_{01}}{n}\right]^2+\frac{1}{2}\left[(1-2W)(1-Q)\pi+W(1-Q)-\frac{n_{00}}{n}\right]^2. \quad (8)$$

We want to choose π that minimizes the squared distance D in (8). Setting $\partial D/\partial\pi=0$, we have the following theorem.

Theorem 2.1. An unbiased estimator of the population proportion π is given by:

$$\hat{\pi}_f=\frac{1}{2}+\frac{Q(n_{11}/n-n_{01}/n)+(1-Q)(n_{10}/n-n_{00}/n)}{2(2W-1)[Q^2+(1-Q)^2]},\ W\neq 0.5 \quad (9)$$

Proof:

It follows from the fact that the observed proportions of (Yes, Yes), (Yes, No), (No, Yes), (No, No) responses (n_{ij}/n) are unbiased estimators for the true proportions of such responses (θ_{ij}), $i=0,1$; $j=0,1$.

i.e.: $\mathrm{E}(n_{ij}/n)=\theta_{ij}$ for all $i=0,1$; $j=0,1$.

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

$$V(\hat{\pi}_f)=\frac{Q^3+(1-Q)^3}{4n(2W-1)^2[Q^2+(1-Q)^2]^2}-\frac{(2\pi-1)^2}{4n},\ W\neq 0.5 \quad (10)$$

Proof:

Note that:

$$V(\hat{\pi}_f)=\frac{Q^2*V\left(\frac{n_{11}}{n}-\frac{n_{01}}{n}\right)+(1-Q)^2*V\left(\frac{n_{10}}{n}-\frac{n_{00}}{n}\right)+2Q(1-Q)*Cov\left(\frac{n_{11}-n_{01}}{n},\frac{n_{10}-n_{00}}{n}\right)}{4(2W-1)^2[Q^2+(1-Q)^2]^2} \quad (11)$$

Using the following results from the standard multinomial distribution in equation (11), we can prove theorem 2.2.

$$V(n_{11}/n)=\theta_{11}(1-\theta_{11})/n \qquad Cov(n_{11}/n,n_{01}/n)=-\theta_{11}\theta_{01}/n$$
$$V(n_{10}/n)=\theta_{10}(1-\theta_{10})/n \qquad Cov(n_{11}/n,n_{00}/n)=-\theta_{11}\theta_{00}/n$$
$$V(n_{01}/n)=\theta_{01}(1-\theta_{01})/n \qquad Cov(n_{10}/n,n_{01}/n)=-\theta_{10}\theta_{01}/n$$
$$V(n_{00}/n)=\theta_{00}(1-\theta_{00})/n \qquad Cov(n_{10}/n,n_{00}/n)=-\theta_{10}\theta_{00}/n$$
$$Cov(n_{11}/n,n_{10}/n)=-\theta_{11}\theta_{10}/n \qquad \text{and } Cov(n_{01}/n,n_{00}/n)=-\theta_{01}\theta_{00}/n$$

It is obvious from equation (10) that the variance of $\hat{\pi}_f$ is symmetric about 0.5 for each of the parameters (W, Q and π).

Theorem 2.3. An unbiased estimator of the variance of $\hat{\pi}_f$ is given by:

$$\hat{V}(\hat{\pi}_f)=\frac{1}{4(n-1)}\left[\frac{Q^3+(1-Q)^3}{(2W-1)^2[Q^2+(1-Q)^2]^2}-(2\hat{\pi}_f-1)^2\right],\ W\neq 0.5 \qquad (12)$$

The proof is immediate by taking the expected values on both sides of equation (12).

Efficiency comparison

The proposed estimator is more efficient than that proposed by Odumade and Singh (2009) if

$$\frac{Q^3+(1-Q)^3}{(2W-1)^2[Q^2+(1-Q)^2]^2}\prec\frac{(P+T-1)^2[PT+(1-P)(1-T)]+(P-T)^2[(1-P)T+P(1-T)]}{[(P+T-1)^2+(P-T)^2]^2} \qquad (13)$$

It is obvious that the above condition doesn't depend on π which is the parameter of interest

3. ANALYSIS OF DIFFERENT SITUATIONS

The relative efficiency (RE) of the proposed estimator $\hat{\pi}_f$ with respect to the Odumade and Singh (2009) estimator is given by:

$$\text{RE}(os)=\frac{V(\hat{\pi}_{os})}{V(\hat{\pi}_f)}\times 100\%$$

For each value of π where $\pi\in[0.1,\ 0.9]$; the relative efficiencies are calculated for all the possible combinations (5760 combinations) from the values of P,

T, W and Q where each of the parameters (P, T, W and Q) takes the values: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9. The following results were observed:

1) The value of the RE (*os*) is symmetric about 0.5 for each of the parameters (P, T, W, Q and π).
2) The value of the RE (*os*) for the combination $P=a$, $T=b$, $W=c$, $Q=d$ and $\pi=e$ is the same for the combination $P=b$, $T=a$, $W=c$, $Q=d$ and $\pi=e$, that is the value of the relative efficiency remains the same after exchanging the values of P and T.
3) It can be easily observed from table 3 that the value of the RE (*os*) increases for values of W and π close to 0 and values of Q, P and T close to 0.5.
4) The RE (*os*) reaches its maximum at P=0.4, T=0.5, W=0.1, Q=0.5 and π=0.1 as shown in table 3.
   Using results (1) and (2), the maximum value of the RE (*os*) will also be attained at other combinations of P, T, W, Q and π.
   As mentioned before, the RE (*os*) is symmetric around π=0.5. For specific values of P, T, W and Q, the RE (*os*) reaches its maximum as $\pi \to 0$ or $\pi \to 1$ which indicates that the proposed model can be safely used whether the proportion of the sensitive attribute is rare or predominant.
5) It is preferable to choose value of $W < P$ because if a high value of P (or W) is chosen then it is more likely that a respondent is being asked about his/her membership in the sensitive group A. Only the respondents belonging to group A are hesitant to report "Yes", the respondents belonging to the group A^c are not much worried in reporting "Yes" or "No" response. The choice of Q is made close to 0.5 because it will give almost equal chance for each respondent to say "Yes" or "No" without any hesitation. It is very interesting to note that the proposed alternative model shows maximum relative efficiency when Q is 0.5 and maximum cooperation is expected for this choice of Q. Also the relative efficiency of the proposed estimator remains higher when π is close to zero. In practice, the proportion π that possess sensitive characteristic in the population is likely to be close to zero.

Using the values of P and T that were proposed by Odumade and Singh (2009), it was found that the proposed model is more efficient than the OS model for values of W, π close to 0 or 1 and values of Q around 0.5. In this case the proposed model will not only be more efficient than the OS model but it will also be more efficient than the Warner (1965), Mangat and Singh (1990) and Mangat (1994) models.

TABLE 3

Percent RE of the proposed model over the OS model

P	*T*	*W*	*Q*	π				
				0.1	0.2	0.3	0.4	0.5
0.1	0.3	0.1	0.5	100.81	100.62	100.53	100.49	100.48
	0.4	0.1	0.5	102.04	101.56	101.34	101.24	101.2
0.2	0.2	0.1	0.1	118.23	114.41	112.53	111.62	111.35
			0.2	111.79	109.43	108.25	107.67	107.49
			0.3	115.79	112.54	110.93	110.15	109.91
			0.4	127.70	121.54	118.58	117.17	116.74
			0.5	135.38	127.14	123.27	121.44	120.89
	0.3	0.1	0.1	161.79	148.85	142.49	139.41	138.48
			0.2	152.99	142.37	137.06	134.47	133.69
			0.3	158.46	146.41	140.46	137.56	136.69
			0.4	174.76	158.12	150.15	146.33	145.19
			0.5	185.27	165.41	156.09	151.67	150.34
	0.4	0.1	0.1	193.12	173.61	164.03	159.39	157.99
			0.2	182.61	166.05	157.78	153.74	152.52
			0.3	189.14	170.77	161.69	157.28	155.94
			0.4	208.59	184.42	172.84	167.30	165.64
			0.5	221.14	192.93	179.68	173.4	171.52
	0.5	0.1	0.1	202.37	180.92	170.39	165.29	163.75
			0.2	191.36	173.05	163.90	159.44	158.09
			0.3	198.21	177.97	167.96	163.1	161.63
			0.4	218.59	192.2	179.55	173.5	171.68
			0.5	231.74	201.06	186.65	179.82	177.78
0.3	0.3	0.1	0.1	282.58	244.32	225.53	216.44	213.69
			0.2	267.20	233.69	216.95	208.78	206.3
			0.3	276.76	240.33	222.32	213.58	210.93
			0.4	305.22	259.54	237.66	227.19	224.05
			0.5	323.58	271.52	247.06	235.47	232
		0.2	0.1	125.65	122.94	121.33	120.47	120.2
			0.2	120.18	118.13	116.91	116.25	116.04
			0.3	123.59	121.14	119.68	118.9	118.65
			0.4	133.47	129.75	127.56	126.39	126.03
			0.5	139.63	135.04	132.36	130.95	130.5
	0.4	0.1	0.1	443.04	371.16	335.86	318.77	313.61
			0.2	418.93	355.01	323.08	307.09	302.77
			0.3	433.91	365.09	331.07	314.55	309.56
			0.4	478.54	394.28	353.92	334.61	328.81
			0.5	507.32	412.47	367.91	346.8	340.48
		0.2	0.1	196.99	186.77	180.69	177.44	176.41
			0.2	188.42	179.46	174.1	171.22	170.31
			0.3	193.77	184.03	178.22	175.11	174.13
			0.4	209.27	197.11	189.96	186.15	184.95
			0.5	218.92	205.15	197.11	192.86	191.52
	0.5	0.1	0.1	531.07	440.75	396.39	374.92	368.44
			0.2	502.18	421.58	381.31	361.65	355.69
			0.3	520.14	433.55	390.74	369.96	363.68
			0.4	573.64	468.21	417.70	393.54	386.29
			0.5	608.13	489.81	434.22	407.88	400
		0.2	0.1	236.14	221.78	213.26	208.69	207.25
			0.2	225.86	213.11	205.48	201.38	200.08
			0.3	232.28	218.53	210.35	205.95	204.57
			0.4	250.85	234.06	224.19	218.94	217.29
			0.5	262.42	243.61	232.64	226.83	225

P	*T*	*W*	*Q*	π				
				0.1	0.2	0.3	0.4	0.5
0.4	0.4	0.1	0.1	1170.07	945.86	835.75	782.44	766.35
			0.2	1106.4	904.71	803.94	754.74	739.84
			0.3	1145.97	930.39	823.84	772.09	756.45
			0.4	1263.84	1004.79	880.67	821.31	803.47
			0.5	1339.84	1051.14	915.51	851.23	832
		0.2	0.1	520.26	475.95	449.62	435.52	431.07
			0.2	497.62	457.34	433.23	420.26	416.16
			0.3	511.75	468.98	443.49	429.82	425.5
			0.4	552.67	502.31	472.68	456.92	451.95
			0.5	578.17	522.79	490.49	473.38	468
		0.3	0.1	201.13	196.72	193.8	192.13	191.59
			0.2	193.47	189.55	186.94	185.45	184.96
			0.3	198.26	194.04	191.23	189.63	189.11
			0.4	211.94	206.81	203.43	201.5	200.87
			0.5	220.32	214.6	210.84	208.7	208
	0.5	0.1	0.1	2306.06	1843.83	1616.82	1506.93	1473.75
			0.2	2180.57	1763.61	1555.28	1453.58	1422.77
			0.3	2258.56	1813.68	1593.77	1486.99	1454.7
			0.4	2490.87	1958.71	1703.73	1581.78	1545.14
			0.5	2460.65	2049.06	1771.12	1639.41	1600
		0.2	0.1	1025.37	927.8	869.82	838.78	828.99
			0.2	980.76	891.53	838.11	809.39	800.31
			0.3	1008.6	914.21	857.96	827.8	818.27
			0.4	1089.25	979.18	914.45	879.99	869.14
			0.5	1139.5	1019.12	948.9	911.69	900
		0.3	0.1	396.39	383.48	374.92	370.03	368.44
			0.2	381.31	369.5	361.65	357.16	355.69
			0.3	390.74	378.25	369.95	365.22	363.68
			0.4	417.7	403.15	393.54	388.07	386.29
			0.5	434.22	418.34	407.88	401.93	400

Department of Statistics SALLY ABDELFATAH
Faculty of Economics and Political Science REDA MAZLOUM
Cairo University

Department of Mathematics SARJINDER SINGH
Texas A&M University-Kingsville

ACKNOWLEDGEMENTS

The authors are thankful to the Executive Editor Daniela Grazia and a learned referee for their valuable comments to bring the original manuscript in the present form.

REFERENCES

N. S. MANGAT (1994), *An improved randomized response strategy*, "Journal of the Royal Statistical Society", 56, pp. 93-95.

N. S. MANGAT, R. SINGH (1990), *An alternative randomized response procedure*, "Biometrika", 77, pp. 439-442.

O. ODUMADE, S. SINGH (2009), *Efficient use of two decks of cards in randomized response sampling*, "Communication in Statistics-Theory and Methods", 38, pp. 439-446.

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

SUMMARY

An alternative randomized response model using two decks of cards

In an attempt to obtain trustworthy answers from the respondents regarding a sensitive question in order to estimate the population proportion possessing that sensitive attribute, Odumade and Singh (2009) have suggested a randomized response model based on the use of two decks of cards. In this paper, a modification to the structure of the two decks of cards used by Odumade and Singh (2009) is proposed. The condition when the proposed model is more efficient than the Odumade and Singh (2009) model has been obtained. The proposed model can be easily adjusted to be more efficient than the Warner (1965), Mangat and Singh (1990), Mangat (1994) and Odumade and Singh (2009) models. Another advantage of the proposed model is that it can increase the respondents' cooperation as they will be required to answer the sensitive question only once and not twice as in the case of the Odumade and Singh (2009) model.