involve

a journal of mathematics

Optional unrelated-question randomized response models

Sat Gupta, Anna Tuck, Tracy Spears Gill and Mary Crowe

msp

2013 vol. 6, no. 4

INVOLVE 6:4 (2013)
dx.doi.org/10.2140/involve.2013.6.483

Optional unrelated-question randomized response models

Sat Gupta, Anna Tuck, Tracy Spears Gill and Mary Crowe

(Communicated by Kenneth S. Berenhaut)

We propose a generalization of Greenberg's unrelated-question randomized response model allowing subjects the option of giving a correct response if they find the survey question nonsensitive, and to give a scrambled response if they find the question sensitive. Models are provided for both the binary response and the quantitative response situations. Mathematical properties of the proposed models are examined and validated with computer simulations.

1. Introduction

Obtaining accurate information is essential in all surveys, particularly in public health research where respondents often face sensitive and personal questions. Examples include surveys of sexual behavior, drug use, or illegal activities. Despite assurances of anonymity, subjects often give untruthful responses leading to problematic response bias.

One method of reducing this bias is the randomized response technique (RRT), originally introduced in [Warner 1965], and subsequently developed and generalized by many researchers [Greenberg et al. 1969; Gupta et al. 2002; 2010; Mehta et al. 2012; Sousa et al. 2010]. We will focus on the unrelated-question RRT method, developed in [Greenberg et al. 1969]. Compared to direct questioning methods, all RRT methods lead to more accurate estimates of sensitive behaviors, because of increased anonymity of the subject's response. In the unrelated-question model, a predetermined proportion of subjects are randomized to answer an innocuous unrelated question with known prevalence level. The researcher is unaware of which question (actual or innocuous) any particular respondent answered, although the mean of the research question can be estimated at the aggregate level. Unrelated-question RRT has been used extensively over the past fifty years to estimate

MSC2010: 62D05.
Keywords: parameter estimation, randomized response technique, unrelated-question model, optional scrambling.
Research supported by NSF grants DBI 0926288 and DMS 0850465.

prevalence of behaviors ranging from induced abortion [Chow et al. 1979] to software piracy [Kwan et al. 2010] and livestock disease prevalence [Cross et al. 2010]. This technique avoids the ethical issues associated with the bogus pipeline technique [Jones and Sigall 1971] and is not as lengthy as the Marlowe–Crowne social desirability scale method [Crowne and Marlowe 1960]. Here the increase in anonymity offered by the technique lessens respondent anxiety during the survey, resulting in more truthful responses [Stem and Bozman 1988].

The original unrelated-question RRT model makes no differentiation as to whether an individual actually considers the topic sensitive; every subject is assumed to find the research question sensitive, so all subjects utilize the randomization device to produce a scrambled response. However, a topic or question may be sensitive for one person, but not sensitive for another. Optional RRT models, introduced in [Gupta et al. 2002], take this into account by allowing subjects who do not find the question sensitive to answer it without utilizing the randomization step. Subjects who find the research question sensitive still use the randomization device prior to providing a response. In this optional model, the researcher remains unaware as to whether or not the subject used the scrambling device or provided a truthful response.

We propose a generalization of the unrelated-question RRT, which takes this difference into account by allowing the randomization step to be optional for the subjects. We deal with both the binary response and the quantitative response situations and estimate the prevalence (π) of the sensitive behavior and the mean response (μ) of the quantitative sensitive question. In addition, the model also estimates the sensitivity level (W) of the underlying question, which is the proportion of subjects who consider the question to be sensitive, and hence choose to provide a scrambled response. We provide the theoretical framework for the two models and examine their mathematical properties, which are also validated by computer simulations.

2. Proposed quantitative response model

We begin first with the quantitative response case, where the researcher is interested in estimating population mean. A randomization device provided to the respondent by the researcher determines whether the subject receives the sensitive research question or the innocuous, unrelated question.

Let X be the true sensitive variable of interest with unknown mean μ_X and unknown variance σ_X^2, and Y be a nonsensitive variable with known mean μ_Y and known variance σ_Y^2. Let p represent the probability of receiving the sensitive question from the randomization device.

The reported response Z is given by

$$Z = \begin{cases} X & \text{with probability } p, \\ Y & \text{with probability } 1-p. \end{cases}$$

Let W be the sensitivity level of the question. That is, a proportion W of the respondents considers the question sensitive and will choose to provide a scrambled response. Others will provide a direct response with probability $1 - W$. Then

$$Z = \begin{cases} X & \text{with probability } (1-W) + Wp, \\ Y & \text{with probability } W(1-p), \end{cases}$$

with

$$\begin{aligned} E(Z) &= (1-W)E(X) + W(pE(X) + (1-p)E(Y)), \\ \mathrm{Var}(Z) &= [(1-W)+Wp]E(X^2) + W(1-p)E(Y^2) - [E(Z)]^2. \end{aligned} \tag{2-1}$$

Here, both μ_X and W are unknown parameters. To solve the above equation for two unknowns, we use a split-sample approach where the total sample size may be split into two subsamples, each receiving a randomization device with a different probability (p_i, $i = 1, 2$) of receiving the sensitive question. The expected response in the i-th ($i = 1, 2$) subsample then is given by

$$E(Z_i) = (1-W)E(X) + W(p_i E(X) + (1-p_i)E(Y)), \quad \text{where } i = 1, 2. \tag{2-2}$$

2.1. ***Estimation of population mean.*** Solving the system of two equations (2-2) for the parameters of interest, we get

$$\frac{E(Z_1) - E(X)}{E(Z_2) - E(X)} = \frac{1-p_1}{1-p_2}.$$

Solving for $E(X)$, we get

$$E(X) = \frac{E(Z_1) - \lambda E(Z_2)}{1-\lambda}, \quad \text{where } \lambda = \frac{1-p_1}{1-p_2}.$$

This suggests estimating μ_X by

$$\hat{\mu}_X = \frac{\bar{Z}_1 - \lambda \bar{Z}_2}{1-\lambda}, \tag{2-3}$$

where $\bar{Z}_i$ is the sample mean of reported responses in the i-th subsample. The variance of this estimator is given by

$$\mathrm{Var}(\hat{\mu}_X) = \frac{\mathrm{Var}(\bar{Z}_1) + \lambda^2 \mathrm{Var}(\bar{Z}_2)}{(1-\lambda)^2}, \tag{2-4}$$

where

$$\begin{aligned} \mathrm{Var}(\bar{Z}_1) &= \frac{[(1-W)+Wp_1]E(X^2) + W(1-p_1)E(Y^2) - [E(Z_1)]^2}{n_1}, \\ \mathrm{Var}(\bar{Z}_2) &= \frac{[(1-W)+Wp_2]E(X^2) + W(1-p_2)E(Y^2) - [E(Z_2)]^2}{n_2}. \end{aligned}$$

It is easy to see that $E(\hat{\mu}_X) = \mu_X$, so the estimator $\hat{\mu}_X$ is unbiased. Also, $\hat{\mu}_X$ is a linear combination of independent sample means; hence it has an asymptotic normal distribution. More formally, we have the following asymptotic result:

Theorem 1. *The estimator $\hat{\mu}_X$ is distributed as $AN(\mu_X, V)$, where*

$$V = \frac{\operatorname{Var}(\bar{Z}_1) + \lambda^2 \operatorname{Var}(\bar{Z}_2)}{(1-\lambda)^2}$$

with

$$\operatorname{Var}(\bar{Z}_i) = \frac{[(1-W) + Wp_i]E(X^2) + W(1-p_i)E(Y^2) - [E(Z_2)]^2}{n_i}, \quad i = 1, 2$$

and

$$E(Z_i) = (1-W)E(X) + W(p_i E(X) + (1-p_i)E(Y)).$$

2.2. ***Optimal allocation of sample size.*** For the optimal sample split (n_1, n_2), we look at the first derivative of $\operatorname{Var}(\hat{\mu}_X)$ from (2-3), given by

$$\frac{\partial \operatorname{Var}(\hat{\mu}_X)}{\partial n_1} = \frac{1}{(1-\lambda)^2}\left\{\frac{-\sigma_1^2}{n_1^2} + \lambda^2 \frac{\sigma_2^2}{(n-n_1)^2}\right\}.$$

Setting this equal to zero, we get

$$0 = \frac{1}{(1-\lambda)^2}\left(\frac{-\sigma_1^2}{n_1^2} + \lambda^2\left(\frac{\sigma_2}{n-n_1}\right)^2\right),$$

$$\frac{\sigma_1^2}{n_1^2} = \lambda^2 \frac{\sigma_2^2}{(n-n_1)^2},$$

$$\frac{n-n_1}{n_1} = \sqrt{\lambda^2 \frac{\sigma_2^2}{\sigma_1^2}} = \left|\lambda \frac{\sigma_2}{\sigma_1}\right|.$$

Therefore,

$$\frac{n_2}{n_1} = \lambda \frac{\sigma_2}{\sigma_1} \tag{2-5}$$

gives the optimal ratio of subjects split in the two subsamples. This will result in the minimum variance of the estimator $\hat{\mu}_X$ since the second derivative of $\operatorname{Var}(\hat{\mu}_X)$ is positive. Equation (2-5) assumes rough preliminary estimates of σ_1 and σ_2 are available. These may be obtained through a pilot study.

2.3. ***Estimation of sensitivity level.*** In addition to estimating the mean $(\hat{\mu}_X)$, the proportion of subjects who scramble their response (W) is also estimated. We can easily solve (2-2) for W, which will lead to the possible estimator

$$\hat{W} = \frac{\bar{Z}_1 - \bar{Z}_2}{(p_2 - p_1)(\mu_Y - \hat{\mu}_X)}. \tag{2-6}$$

This representation of $\hat{W}$ as a ratio of two random variables presents difficulties in deriving its properties. We can, however, rewrite $\hat{W}$ in terms of $\bar{Z}_1$ and $\bar{Z}_2$ to get

$$\hat{W} = \frac{\bar{Z}_1 - \bar{Z}_2}{\mu_Y(p_2 - p_1) + (1 - p_2)\bar{Z}_1 - (1 - p_1)\bar{Z}_2}. \tag{2-7}$$

Using the first-order bivariate Taylor approximation, with $A = E(\bar{Z}_1)$ and $B = E(\bar{Z}_2)$, we get

$$\begin{aligned}
\hat{W} &\approx \hat{W}(A, B) + \frac{\partial \hat{W}(\bar{Z}_1, \bar{Z}_2)}{\partial \bar{Z}_1}\bigg|_{A,B}(\bar{Z}_1 - A) + \frac{\partial \hat{W}(\bar{Z}_1, \bar{Z}_2)}{\partial \bar{Z}_2}\bigg|_{A,B}(\bar{Z}_2 - B) \\
&= \frac{A - B}{\mu_Y(p_2 - p_1) + (1 - p_2)A - (1 - p_1)B} \\
&\quad + \frac{(p_2 - p_1)(\mu_Y - B)(\bar{Z}_1 - A)}{[\mu_Y(p_2 - p_1) + (1 - p_2)A - (1 - p_1)B]^2} \\
&\quad + \frac{(p_2 - p_1)(A - \mu_Y)(\bar{Z}_2 - B)}{[\mu_Y(p_2 - p_1) + (1 - p_2)A - (1 - p_1)B]^2} =: \hat{W}_1.
\end{aligned}$$

Taking the expected value, we get $(Z_1 - \mu_\gamma) \to (\Lambda - \mu_\gamma)$:

$$\begin{aligned}
E(\hat{W}_1) &= \frac{A - B}{\mu_Y(p_2 - p_1) + (1 - p_2)A - (1 - p_1)B} \\
&\quad + \frac{(p_2 - p_1)(\mu_Y - B)(E(\bar{Z}_1) - A)}{[\mu_Y(p_2 - p_1) + (1 - p_2)A - (1 - p_1)B]^2} \\
&\quad + \frac{(p_2 - p_1)(\bar{Z}_1 - \mu_Y)(E(\bar{Z}_2) - B)}{[\mu_Y(p_2 - p_1) + (1 - p_2)A - (1 - p_1)B]^2} \\
&= \frac{A - B}{\mu_Y(p_2 - p_1) + (1 - p_2)A - (1 - p_1)B} = W.
\end{aligned}$$

Thus $\hat{W}_1$, the first-order approximation of $\hat{W}$, is an unbiased estimator of W with variance given by

$$\begin{aligned}
\mathrm{Var}(\hat{W}_1) &= \left(\frac{(p_2 - p_1)(\mu_Y - B)}{[\mu_Y(p_2 - p_1) + (1 - p_2)A - (1 - p_1)B]}\right)^2 \frac{\sigma_1^2}{n_1} \\
&\quad + \left(\frac{(p_2 - p_1)(\mu_Y - B)(A - \mu_Y)}{[\mu_Y(p_2 - p_1) + (1 - p_2)A - (1 - p_1)B]}\right)^2 \frac{\sigma_2^2}{n_2},
\end{aligned} \tag{2-8}$$

where

$$\sigma_i^2 = [1 - W + Wp_i]E(X^2) + W(1 - p_i)E(Y^2) - [E(Z_i)]^2, \quad i = 1, 2.$$

Also, $\hat{W}_1$ is asymptotically normal since it is a linear combination of independent sample means $\bar{Z}_1$ and $\bar{Z}_2$. This property is later confirmed by simulation. This result is summarized in the following theorem.

Theorem 2. $\hat{W}_1 \sim AN(W, V_w)$, *where*

$$\operatorname{Var}(\hat{W}_1) = \left(\frac{(p_2 - p_1)(\mu_Y - B)}{[\mu_Y(p_2 - p_1) + (1 - p_2)A - (1 - p_1)B]}\right)^2 \frac{\sigma_1^2}{n_1} + \left(\frac{(p_2 - p_1)(A - \mu_Y)}{[\mu_Y(p_2 - p_1) + (1 - p_2)A - (1 - p_1)B]}\right)^2 \frac{\sigma_2^2}{n_2},$$

$$\sigma_i^2 = [1 - W + Wp_i]E(X^2) + W(1 - p_i)E(Y^2) - [E(Z_i)]^2, \quad i = 1, 2.$$

3. Proposed binary response model

The estimator proposed in the preceding section is used when an estimate of the population mean is needed. In many cases the main research interest is in the prevalence of a particular sensitive behavior or characteristic. In this case the research question demands a binary response, such as "yes" or "no". We modify the preceding estimator to be used in these cases.

3.1. ***Proposed model.*** Let X be a sensitive binary variable of interest with unknown mean π_X, and Y be a nonsensitive binary variable with known mean π_Y. Let p represent probability of receiving the sensitive question from the randomization device. Here the probability of a "yes" response (P_Y) is given by

$$P_Y = (1 - W)\pi_X + W[p\pi_X + (1 - p)\pi_Y].$$

Again, the sample is split into two subsamples to solve for both π_X and W. The probability of a "yes" response in the i-th ($i = 1, 2$) subsample is given by

$$P_{Y_i} = (1 - W)\pi_X + W[p_i\pi_X + (1 - p_i)\pi_Y], \quad i = 1, 2.$$

Solving this system of two equations for π_X gives

$$\pi_X = \frac{P_{Y_1} - \lambda P_{Y_2}}{1 - \lambda}, \qquad \text{where } \lambda = \frac{1 - p_1}{1 - p_2}. \tag{3-1}$$

3.2. ***Estimation of population proportion.*** Using (3-1), we obtain the estimate for the population proportion (π_X) of the sensitive characteristic as

$$\hat{\pi}_X = \frac{\hat{P}_{Y_1} - \lambda \hat{P}_{Y_2}}{1 - \lambda}, \tag{3-2}$$

with variance given by

$$\operatorname{Var}(\hat{\pi}_X) = \frac{\operatorname{Var}(\hat{P}_{Y_1}) + \lambda^2 \operatorname{Var}(\hat{P}_{Y_2})}{(1 - \lambda)^2}, \tag{3-3}$$

where

$$\mathrm{Var}(\hat{P}_{Y_1}) = \frac{P_{Y_1}(1-P_{Y_1})}{n_1} \quad \text{and} \quad \mathrm{Var}(\hat{P}_{Y_2}) = \frac{P_{Y_2}(1-P_{Y_2})}{n_2}.$$

Again, it can easily be seen that $E(\hat{\pi}_X) = \pi_X$, so the estimator $\hat{\pi}_X$ is unbiased. Also $\hat{\pi}_X$ is a linear combination of independent sample means, and hence has an asymptotic normal distribution.

3.3. ***Optimal allocation of sample size.*** Just as in the quantitative response case, the optimal sample split is given by

$$\frac{n_2}{n_1} = \lambda\sqrt{\frac{P_{Y_2}(1-P_{Y_2})}{P_{Y_1}(1-P_{Y_1})}}. \tag{3-4}$$

3.4. ***Estimation of sensitivity level.*** From (3-1), an estimator for the sensitivity level (W) in the binary case can be represented as

$$\hat{W}_\pi = \frac{\hat{P}_{Y_1} - \hat{P}_{Y_2}}{(p_2-p_1)(\pi_Y - \hat{\pi}_X)} = \frac{\hat{P}_{Y_1} - \hat{P}_{Y_2}}{\pi_Y(p_2-p_1) + (1-p_2)\hat{P}_{Y_1} - (1-p_1)\hat{P}_{Y_2}}. \tag{3-5}$$

Applying the first-order Taylor approximation expansion for a bivariate function, and assuming $A = P_{Y_1}$, $B = P_{Y_2}$, this can be approximated by

$$\begin{aligned}\hat{W}_\pi \approx{}& \frac{A-B}{\pi_Y(p_2-p_1)+(1-p_2)A-(1-p_1)B} \\ &+ \frac{(p_2-p_1)(\pi_Y-B)(\hat{P}_{Y_1}-A)}{[\pi_Y(p_2-p_1)+(1-p_2)A-(1-p_1)B]^2} \\ &+ \frac{(p_2-p_1)(A-\pi_Y)(\hat{P}_{Y_2}-B)}{[\pi_Y(p_2-p_1)+(1-p_2)A-(1-p_1)B]^2} =: \hat{W}_{\pi_1}.\end{aligned}$$

It can be verified that

$$E(\hat{W}_{\pi_1}) = \frac{A-B}{\mu_Y(p_2-p_1)+(1-p_2)A-(1-p_1)B} = W_\pi.$$

Thus, $\hat{W}_{\pi_1}$ is an unbiased estimator of W with variance given by

$$\begin{aligned}\mathrm{Var}(\hat{W}_{\pi_1}) ={}& \left(\frac{(p_2-p_1)(\pi_Y-B)}{[\pi_Y(p_2-p_1)+(1-p_2)A-(1-p_1)B]}\right)^2 \frac{\sigma_1^2}{n_1} \\ &+ \left(\frac{(p_2-p_1)(A-\mu_Y)}{[\pi_Y(p_2-p_1)+(1-p_2)A-(1-p_1)B]}\right)^2 \frac{\sigma_2^2}{n_2},\end{aligned} \tag{3-6}$$

where

$$\sigma_1^2 = \frac{P_{Y_1}(1-P_{Y_1})}{n_1} \quad \text{and} \quad \sigma_2^2 = \frac{P_{Y_2}(1-P_{Y_2})}{n_2}.$$

W	n_1	$\hat{\mu}_X$	$\mathrm{Var}(\hat{\mu}_X)$	$\widehat{\mathrm{Var}}(\hat{\mu}_X)$	$\hat{W}_1$	$\mathrm{Var}(\hat{W}_1)$	$\widehat{\mathrm{Var}}(\hat{W}_1)$
0.0	698	1.9988	0.0058	0.0058	−0.0001	0.0065	0.0067
0.1	674	2.0012	0.0066	0.0065	0.0982	0.0069	0.0069
0.2	680	2.0013	0.0068	0.0068	0.1982	0.0073	0.0073
0.3	690	2.0008	0.0072	0.0072	0.2984	0.0077	0.0077
0.4	699	2.0004	0.0075	0.0075	0.3989	0.0080	0.0080
0.5	710	2.0005	0.0079	0.0079	0.4991	0.0082	0.0082
0.6	722	2.0006	0.0081	0.0082	0.5996	0.0083	0.0084
0.7	737	1.9998	0.0084	0.0084	0.7004	0.0085	0.0085
0.8	753	1.9996	0.0086	0.0088	0.8005	0.0087	0.0087
0.9	774	1.9999	0.0089	0.0090	0.9002	0.0090	0.0090
1.0	800	1.9991	0.0091	0.0092	1.0001	0.0095	0.0097

Table 1. Estimates of μ_X and W with optimized subsamples. X and Y have Poisson distributions with $\mu_X = 2.0$, $\mu_Y = 4.0$. Total sample size is 1000, $p_1 = 0.8$, $p_2 = -0.2$.

Also, $\hat{W}_{\pi_1}$ clearly has an asymptotic normal distribution being a linear combination of independent sample means.

4. Simulation study

The preceding theoretical formulas are tested empirically through computer simulations. Poisson distribution is assumed for both X and Y. The subsample split (n_1, n_2) is obtained by the optimal split method described above. Table 1 and Table 2 present simulation results obtained with SAS.

The simulation results provide strong support for the theoretical results that $\hat{\mu}_X$ and $\hat{\pi}_X$ are unbiased. The theoretical and simulated variances of $\hat{\mu}_X$ and $\hat{\pi}_X$ can also be seen to be very close. The simulations also support that $\hat{W}_1$ and $\hat{W}_{\pi_1}$ are good estimators of W for the quantitative case and the binary case, respectively.

We also note that $\hat{W}_1$ and $\hat{W}_{\pi_1}$ may occasionally give estimates that are outside of the normal range [0, 1]. This happens when the true value of W is close to zero or 1. As in [Warner 1965], this is because our estimators are unconstrained. In such cases we recommend using an estimate of zero if $\hat{W}_1 < 0$, and 1 if $\hat{W}_1 > 1$.

The Kolmogorov–Smirnov normality test is used in SAS to check the sampling distributions of $\hat{\mu}_X$, $\hat{\pi}_X$, $\hat{W}_1$, and $\hat{W}_{\pi_1}$ against the normal distribution. The p-values for $\hat{\mu}_X$, $\hat{\pi}_X$, $\hat{W}_1$, and $\hat{W}_{\pi_1}$ are all greater than 0.15, indicating that their distributions are not significantly different from the normal distribution.

W	n_1	$\hat{\pi}_X$	$\mathrm{Var}(\hat{\pi}_X)$	$\widehat{\mathrm{Var}}(\hat{\pi}_X)$	$\hat{W}_{\pi_1}$	$\mathrm{Var}(\hat{W}_{\pi_1})$	$\widehat{\mathrm{Var}}(\hat{W}_{\pi_1})$
0.0	800	0.1508	0.0003	0.0003	−0.0037	0.0045	0.0047
0.1	786	0.1500	0.0004	0.0004	0.1005	0.0050	0.0050
0.2	777	0.1500	0.0004	0.0004	0.2007	0.0053	0.0052
0.3	772	0.1499	0.0005	0.0005	0.3008	0.0054	0.0055
0.4	770	0.1498	0.0005	0.0005	0.4009	0.0054	0.0055
0.5	770	0.1501	0.0005	0.0005	0.5008	0.0053	0.0055
0.6	772	0.1502	0.0005	0.0005	0.6005	0.0052	0.0052
0.7	776	0.1502	0.0006	0.0006	0.7007	0.0049	0.0050
0.8	782	0.1501	0.0006	0.0006	0.8006	0.0046	0.0046
0.9	790	0.1500	0.0006	0.0006	0.9008	0.0042	0.0042
1.0	800	0.1500	0.0006	0.0006	0.9996	0.0038	0.0038

Table 2. Estimates of π_X and W with optimized subsamples. The true values are $\pi_X = 0.15$, $\pi_Y = 0.85$. Total sample size is 1000, $p_1 = 0.8$, $p_2 = -0.2$.

5. Concluding remarks

The optional unrelated-question RRT proposed above provides models for simultaneously estimating both the mean and sensitivity level of a sensitive behavior. This is distinct from previous unrelated-question RRT models, which estimate only the mean. Estimators are derived for both the quantitative and binary response cases. In both cases, estimators of the mean ($\hat{\mu}_X$, $\hat{\pi}_X$) and first-order Taylor approximations of the sensitivity level ($\hat{W}_1$, $\hat{W}_{\pi_1}$) are shown to be asymptotically normal and unbiased.

Of note in Table 1, the variances of both $\hat{\mu}_X$ and $\hat{W}_1$ increase as W increases (when more subjects choose to provide scrambled responses). In Table 2 the variance of $\hat{\pi}_X$ increases slightly as W increases. When optionality is incorporated into this model, when even a small proportion of subjects do not find the question sensitive (and thus answer directly) the variance of the estimator is smaller than in a comparable model where all subjects must provide a scrambled response ($W = 1.0$).

References

[Chow et al. 1979] L. P. Chow, W. Gruhn, and W. P. Chang, "Feasibility of the randomized response technique in rural Ethiopia", *Amer. J. Public Health* **69**:3 (1979), 273–276.

[Cross et al. 2010] P. Cross, G. Edwards-Jones, H. Omed, and A. Williams, "Use of a randomized response technique to obtain sensitive information on animal disease prevalence", *Prev. Vet. Med.* **96**:3–4 (2010), 252–262.

[Crowne and Marlowe 1960] D. P. Crowne and D. Marlowe, "A new scale of social desirability independent of psychopathology", *J. Consult. Psychol.* **24**:4 (1960), 349–54.

[Greenberg et al. 1969] B. G. Greenberg, A.-L. A. Abul-Ela, W. R. Simmons, and D. G. Horvitz, "The unrelated question randomized response model: Theoretical framework", *J. Amer. Statist. Assoc.* **64** (1969), 520–539. MR 40 #982

[Gupta et al. 2002] S. Gupta, B. Gupta, and S. Singh, "Estimation of sensitivity level of personal interview survey questions", *J. Statist. Plann. Inference* **100**:2 (2002), 239–247. MR 1877192 Zbl 0985.62010

[Gupta et al. 2010] S. Gupta, J. Shabbir, and S. Sehra, "Mean and sensitivity estimation in optional randomized response models", *J. Statist. Plann. Inference* **140**:10 (2010), 2870–2874. MR 2011h:62027 Zbl 1191.62009

[Jones and Sigall 1971] E. E. Jones and H. Sigall, "The bogus pipeline: A new paradigm for measuring affect and attitude", *Psychological Bulletin* **76**:5 (1971), 349–364.

[Kwan et al. 2010] S. S. K. Kwan, M. K. P. So, and K. Y. Tam, "Applying the randomized response technique to elicit truthful responses to sensitive questions in IS research: The case of software piracy behavior", *Info. Sys. Research* **21**:4 (2010), 941–959.

[Mehta et al. 2012] S. Mehta, B. K. Dass, J. Shabbir, and S. Gupta, "A three stage optional randomized response model", *J. Stat. Theory Pract.* **6**:3 (2012), 417–427.

[Sousa et al. 2010] R. Sousa, J. Shabbir, P. C. Real, and S. Gupta, "Ratio estimation of the mean of a sensitive variable in the presence of auxiliary information", *J. Stat. Theory Pract.* **4**:3 (2010), 495–507. MR 2758690 Zbl 05902629

[Stem and Bozman 1988] D. E. Stem and C. S. Bozman, "Respondent anxiety reduction with the randomized response technique", *Adv. Consum. Res.* **15**:1 (1988), 595–599.

[Warner 1965] S. L. Warner, "Randomized response: A survey technique for eliminating evasive answer bias", *J. Am. Stat. Assoc.* **60**:309 (1965), 63–69.

Received: 2012-10-01 Revised: 2012-11-05 Accepted: 2012-11-08

sngupta@uncg.edu *Department of Mathematics and Statistics, University of North Carolina at Greensboro, 317 College Avenue, Greensboro, NC 27412, United States*

avmikh@uw.edu *Department of Biostatistics, University of Washington, 1705 NE Pacific Street, Seattle, Washington 98195, United States*

tgspears@uncg.edu *School of Nursing, University of North Carolina at Greensboro, P.O. Box 26170, Greensboro, NC 27402, United States*

tgspears@unc.edu *Department of Biostatistics, University of North Carolina at Chapel Hill, 135 Dauer Drive, Chapel Hill, NC 27599, United States*

involve

msp.org/involve

See inside back cover or msp.org/involve for submission instructions. The subscription price for 2013 is US $105/year for the electronic version, and $145/year (+$35, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to MSP.

Involve (ISSN 1944-4184 electronic, 1944-4176 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840, is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

Involve peer review and production are managed by EditFLOW® from Mathematical Sciences Publishers.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

involve

2013 vol. 6 no. 4

1944-4176(2013)6:4;1-7