Two Randomized Response Models of Sensitive Questions Survey

With Quantity Character

Yu Chunquan
Statistics Department of Shandong Economic University, Jinan, P. R. China; Postcode: 250014
E-mail: sdeu.sts@263.net

In the social—economic survey, some items are sensitive, so it is difficult to obtain the real value from direct survey method. In order to obtain the real result, some special survey methods should be employed. Randomized response survey is an effective method, because it can keep the privacy for the people who are investigated and get the real result at the same time.

Sensitive question survey can be divided into two types according to statistics. One is for attribute characters, its purpose is to obtain the proportion of the population; the other is for quantity characters, the purpose of it is to achieve the population mean or the population total. Based on the discrete uniform distribution and the uniform distribution, in this paper, we design two randomized response model for the sensitive questions survey with quantity characters.

\[ Y \bar{V} \text{ and } S_Y^2 \] are assumed as the survey characteristic, the population mean and the population variance respectively. The purpose of the survey is to estimate \( \bar{V} \). Let \( X \) be a random variable with a discrete uniform distribution or a uniform distribution. Note: \( Z = X \ Y \), where \( Z \) is the product of \( X \) and \( Y \). The distribution of \( X \) is a symmetric discrete uniform distribution with 1 as the mean or a uniform distribution within the interval of \([1 - a, 1 + a]\), and \( 0 < a \leq 1 \). The latter’s mean is 1, too. Assumed \( n \) people are sampled to be investigated, selecting one of the two probabilistic models, everyone partakes one random experiment based on the model. Assume that \( x_i \) is the observational value for the \( i \)th person’s experiment and \( y_i \) is the true value of the survey characteristic, and they are only to be known by the people who will be investigated, not by the investigator, then the person being investigated can obtain \( z_i = x_i \ y_i \), and return the value \( z_i \) to the investigator. Thus the person’s privacy can be kept for him (or her). Then the investigator can get all the values of \( z_1, z_2, \ldots, z_n \) and take the estimator of \( \bar{z} = \frac{1}{n} \sum_{i=1}^{n} z_i \) and \( S_z^2 = \frac{1}{n-1} \sum (z_i - \bar{z})^2 \) as the sample mean and the sample variance respectively.

Assumed \( \bar{Z} = V(\bar{z}) = \bar{y} \), and \( f = \frac{n}{N} \), as the population mean of characteristic \( Z \) the variance of \( \bar{z} \) the estimator of \( V(\bar{z}) \) and the sampling fraction respectively, where \( N \) is the population size. In this paper it can be proved that:

1. \( \bar{Z} = \bar{V} \) and \( E(\bar{z}) = \bar{V} \) are both true based on either the discrete uniform distribution model or the uniform distribution model. Thus \( \bar{z} \) can be regarded as the unbiased estimator of \( \bar{V} \).

2. Under the discrete uniform distribution model, if \( P\{X = m\} = \frac{1}{5}, m = 0.6, 0.8, 1, 1.2, 1.4 \). Note \( \hat{V}_1 = \bar{z} \), then: \( V(\hat{V}_1) = \frac{1-f}{n} \left( 0.08S_y^2 + 0.08\bar{V}^2 \right) \). Its unbiased estimator is: \( V(\hat{V}_1) = \frac{1-f}{n} S_z^2 \).
If $P\{X = m\} = \frac{1}{7}, m = 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6$. Note $\hat{Y}_2 = \bar{z}$, then: $V\left(\hat{Y}_2\right) = \frac{1-f}{n} \left(1.16S_y^2 + 0.16\bar{Y}^2\right)$. Its unbiased estimator is: $v\left(\hat{y}_2\right) = \frac{1-f}{n} s_z^2$

3 Under the uniform distribution model, if the density function of $X$ is:

$$f(x) = \begin{cases} \frac{1}{2a}, & 1-a \leq x \leq 1+a \\ 0, & \text{others} \end{cases} \quad 0 < a \leq 1$$

Note $\hat{Y}_3 = \bar{z}$, then $V\left(\hat{y}_3\right) = \frac{1-f}{n} \left(\frac{3+a^2}{3}S_y^2 + \frac{a^2}{3}\bar{Y}^2\right)$

Its unbiased estimator is: $v\left(\hat{y}_3\right) = \frac{1-f}{n} s_z^2$

In this paper, it is assumed that the distribution of $X$ follows a discrete uniform distribution or a uniform distribution with 1 as the mean, the purpose of it is to make the unbiased estimator $\bar{z}$ of $\bar{Y}$ as the unbiased estimator of $\bar{Y}$ directly and to get the variance of the estimator easily, so the model is laconic. Even if undertherandomizedresponse,thoughthevarianceoftheunbiasedestimatorisbiggerundertherandomized response and the precision is lower compared with the direct answer method, the direct answer may not be gotten or the estimator has systematic bias. Therefore we can get the estimator with some precision under the randomized response method. In addition the parameters of the models are selective, therefore the precision of the model is adjustable. Then, in practice, we can select proper parameters based on the requirements of the precision and the secret of the sensitive questions survey.

References:
3 Jin Ying, Liang Xiaojun, An improved survey and estimation to the quantitative sensitivity questions, statistical research, 2000(11).

**DEUX MODELES DE REPONSE ALEAALOIRE POUR L’ENQUETE SUR LES PROBLEMES SENSIBLES DE CARACTERISTIQUE QUANTITATIF**

Dans cet article, l’auteur étudie l’enquête sur désir chantillons sur les problèmes sensibles de caractère ristique quantitatif et conçoit deux modèles de réponse basé sur la répartition régulière dispersée et la répartition régulière.