International Mathematical Forum, Vol. 9, 2014, no. 29, 1377 - 1381
HIKARI Ltd, www.m-hikari.com
http://dx.doi.org/10.12988/imf.2014.311223

Estimation of Survey Accuracy for Percentile Estimators in Exponential Populations

Roshanak Alimohammadi

Department of Mathematics
Alzahra University, Tehran, Iran

Abstract

Determining accuracy of survey results is important due to their application in planning. In the previous work, estimation of survey accuracy for median estimator in populations with normal distribution is assessed. In this paper, a right skewed population that is exponential distribution is considered. Besides, instead of median, percentile estimators is studied. Furthermore, under some assumptions, accuracy of percentiles as well as median estimator is assessed under some assumptions.

Mathematics Subject Classification: 62D05

Keywords: Data Quality, Percentile Estimator, Precision, Survey Accuracy

1 Introduction

Estimation of survey precision and survey accuracy of median and the other percentiles is different from many of the other nonlinear estimators. Alimohammadi [1] present some equations to compute survey accuracy of median estimator in surveys for normal populations. Alimohammadi [2] assess survey precision of nonlinear estimators and present some relations to calculate survey precision based on taylor's approximation and the proposed response error model.

In this paper, survey accuracy of percentile estimators as well as median estimator in exponential distribution is studied. Therefore, total variance, including sampling as well as nonsampling variance should be estimated. Response error is an unavoidable part of nonsampling error. In the next sections, some response error model is considered. One of the applications of the model is to quantify response error in surveys.

2 Response Error Models and Survey Accuracy

To study of total survey error, Mean Square Error (MSE) criterion may be applied. Then bias and variance of the estimators should be determined. Total survey variance has an inverse relation with survey precision. Therefore, survey precision is an essential component of survey accuracy.

In this study, precision and accuracy of percentile estimators in exponential populations is computed under some assumptions. These assumptions included the cases that the other nonsampling errors are ignorable except that response error. In the other words, processing error may be ignorable now, because of technology's development. If a proper frame is applied in the survey, then coverage error does not occurred. On the other hand, there are several methods to reduce nonresponse rate. Usually response error is an unavoidable part of nonsampling error.

Biemer and Trewin [1] have considered a general models for response error. Alimohammadi and Navvabpour [2] proposed a model of response error in face to face surveys. In this paper, the following response error model is considered:

$$y_{ls} = \mu_{ls} + D_l + R_{ls} \tag{1}$$

Where μ_{ls} is the true value and y_{ls} is the observed value of lsth unit, D_l is the effect of lth interviewer on any interview and R_{ls} is respondent error. Usually, interviewer effect D_l and respondent effect R_{ls} are random effects in surveys. In model (1), effect of respondents nested in interviewers. Assume that these random variables distributed as $N(0, \sigma_D^2)$ and $N(0, \sigma_R^2)$ respectively and $\mu_{ls} \sim N(\mu, \sigma_M^2)$.

In this paper, variance of $\bar{y}$ under model (1) is called response variance of $\bar{y}$. It is denoted by $Var_z(\bar{y})$. Two-stage sampling design is a commonly used method in surveys. It can be shown that, for two-stage cluster sampling design, $Var_z(\bar{y})$ equals to:

$$Var_z(\bar{y}) = \frac{1}{n}[\sigma_D^2(h(1-\frac{1}{m}) + \frac{1}{m}) + \frac{t}{m}] \tag{2}$$

Where n is the number of sampled clusters, L is the number of interviewers, h is the number of clusters assigned to each interviewer such that n=hL, m is the number of sampled units from each of the sampled clusters and $t = \sigma_R^2 + \sigma_M^2$. σ_D^2 and σ_R^2 are variance of effects D and R in model (1), respectively.

3 Survey Accuracy for Percentiles as well as Median Estimator

Percentiles of some survey variables may have many applications. Median as a percentile, is used quite often in social science and social policies. For example, median of population income is a more proper central tendency criterion than the mean. Because of asymmetric distribution of income and it is well known that income has a right skewed distribution, in this paper exponential distribution is considered.

The method of estimating survey accuracy is based on the straight definition of percentiles in continues populations.

For a continues variable, kth percentile (indicated by α) can be written based on distribution function (F) as: $F(\alpha) = \frac{k}{100}$. Then $F^{-1}(k/100) = \alpha$. Where F^{-1} is the inverse function of distribution function F. That is:

$$F_Y(\alpha) = \int_{-\infty}^{\alpha} f(t)dt$$

Then $F(\alpha) = \frac{k}{100}$. Therefore, median can be written as: $F(med) = 0.5$. Then $med = F^{-1}(0.5)$.

In exponential distribution function with parameter λ, as follows:

$$F(y) = \int_0^y \lambda \exp(-\lambda t)dt = 1 - \exp(-\lambda y)$$

kth percentile equals to:

$$\alpha = \frac{\ln(100) - \ln(100-k)}{\lambda} \tag{3}$$

For $k = 50$ in equation (3), median may be obtained as:

$$med = \frac{\ln(2)}{\lambda}$$

On the other hand, $E(X) = \mu = \frac{1}{\lambda}$. Easily Maximum Likelihood estimator of kth percentile may be obtained as $\hat{\alpha}_{ML} = (\ln(100) - \ln(100-k))\bar{Y}$. Where $\bar{Y}$ is the sample mean estimator of observations. Maximum Likelihood estimator of median estimator is $\hat{med}_{ML} = \ln(2)\bar{Y}$.

Total variance of percentile estimator is

$$Var_T(\hat{\alpha_{ML}}) = (\ln(100) - \ln(100-k))^2 Var_T(\bar{Y}) \tag{4}$$

For median, $k = 50$ and $Var_T(\hat{med}_{ML}) = (\ln(2))^2 Var_T(\bar{Y})$.

Alimohammadi [4] estimates Maximum Likelihood (ML) estimator for effects of a proposed response error model.

In this paper, ML estimator of model (1) effects are obtained as follows:

For $(1 - \frac{1}{L})MSD \geq MSR$:

$$\hat{\sigma}_D^2 = \frac{(1 - 1/L)MSD - MSR}{S},\ \hat{t} = MSR$$

And for $(1 - \frac{1}{L})MSD < MSR$:

$$\hat{\sigma}_D^2 = \frac{SSR + SSD}{LS(= N)},\ \hat{t} = 0$$

Where L is the number of interviewers, N is the sample size, SSD and SSR are the sum of squares of effects D and R respectively.

By replacing the above relations in equation (3), ML estimator of response variance of kth percentile estimator in exponential distribution may be obtained as follow:

For $(1 - \frac{1}{L})MSD \geq MSR$:

$$\begin{aligned} Var_z(\hat{\alpha_{ML}}) &= (\ln(100) - \ln(100 - k))^2 \hat{V}ar_z(\bar{y}) \\ &= (\ln(100) - \ln(100 - k))^2 \frac{1}{n}[\hat{\sigma}_D^2(h(1 - \frac{1}{m}) + \frac{1}{m}) + \frac{\hat{t}}{m}] \\ &= (\ln(100) - \ln(100 - k))^2 \frac{1}{n}[\frac{(1 - 1/L)MSD - MSR}{S}(h(1 - \frac{1}{m}) \\ &+ \frac{1}{m}) + \frac{MSR}{m}] \end{aligned}$$

For $(1 - \frac{1}{L})MSD < MSR$:

$$\hat{V}ar_z(\bar{y}) = (\ln(100) - \ln(100 - k))^2 \frac{SSR + SSD}{LS(= N)n}(h(1 - \frac{1}{m}) + \frac{1}{m})$$

Where MSD and MSR are mean squares of effects D and R respectively.

Under the mentioned assumptions about nonsampling errors, and approximating it by response error as the main part of nonsampling error, sampling variance may be added to response variance for computing total variance.

On the other hand, ML estimators of kth percentile in exponential distribution is an unbiased estimator. Since,

$$\begin{aligned} E(\hat{\alpha}_{ML}) &= (\ln(100) - \ln(100 - k))E(\bar{Y}) \\ &= (\ln(100) - \ln(100 - k))\frac{1}{\lambda} = \alpha \end{aligned}$$

Besides, computing expectation under model (1) indicates that, ML estimator of percentile is unbiased. Then, total error with MSE criterion equals to total variance. Finally, the obtained relations may be applied to estimate survey accuracy of percentile estimators under the mentioned assumptions.

4 Conclusion

Estimation of survey precision and accuracy of survey results is important due to their application in planning and decision making. Median and some percentiles have considered in many area. In this paper, accuracy of percentiles estimators is assessed under some assumptions. Besides, maximum likelihood estimators of the considered response error model is obtained. Furthermore, some relations is presented to estimate precision and accuracy in surveys based on the considered response error model and two stage cluster sampling design.

The results of this paper may be applied to estimate precision and accuracy of median estimator and the other percentiles in surveys.

ACKNOWLEDGEMENTS.
This research is supported financially by Alzahra University.

References

[1] R. Alimohammadi, Assessment of Survey Accuracy for Median Estimator, *International Journal of Contemporary Mathematical SciencesScience*, 7: 41-44 (2012), 2111-2115.

[2] R. Alimohammadi, Estimation of Survey Precision for Nonlinear Estimators, Applied Mathematical Sciences, *Journal of Tarbiat Moallem University*, 10: 1 (2010) , 2405 - 2410.

[3] R. Alimohammadi and H. Navvabpour, Response error modeling in face to face surveys, *Research Journal of Science, Isfahan University*, 33 (2008), 1-14.

[4] R. Alimohammadi, Comparison of Variance Components Estimation Methods of Response Error Model in Surveys, *Applied Mathematical Sciences*, 5: 48 (2011), 2405 - 2410.

[5] P.P. Biemer and D. Trewin, A review of measurement error effects on the analysis of survey data. In: L. Lyberg, P. P. Biemer et al. (Eds) *Survey measurement and process quality,* 603-632, New York, Wiley, 1997.

Received: October 1, 2013