Estimation and calibration of response probabilities

Natalja Lepik and Imbi Traat

University of Tartu, Estonia, natalja.lepik@ut.ee, imbi.traat@ut.ee

Keywords: sampling, nonresponse, response probability, backward calibration

Common feature of nowadays sample surveys is low response rate. As a result, adjustments have to be made in estimators to guarantee (nearly) unbiasedness. If the response probabilities θ_k were known then unbiased estimator for the population total $t = \sum_U y_k$ is

$$\hat{t} = \sum_r \frac{y_k}{\pi_k \theta_k},$$

where r is the response set and π_k the inclusion probability of unit k. A review on nonresponse weighting adjustments is given in [1].

We concentrate on the backward calibration property of estimates $\hat{\theta}_k$. Broadly speaking, this property forces $\hat{\theta}_k$ to be close to the sample-based response proportions, e.g. in the groups of units with given auxiliary vector values. If the true response probability is constant in that group then the response proportion is its unbiased estimate. We show that the estimates $\hat{\theta}_k$ resulting from the regression and logistic regression modeling obey the backward calibration property.

Nowadays, there are many new statistical learning methods for classification that could be used for prediction of response probabilities [2]. We compare $\hat{\theta}_k$ of different estimation methods and study the backward calibration property of these methods.

References

[1] Brick, J.M. (2013). Unit nonresponse and weighting adjustments: a critical review. *Journal of Official Statistics* **29**, 329–353.

[2] Valliant, R., Dever, J.A., Kreuter, F. (2013). *Practical Tools for Designing and Weighting Survey Samples*. Springer, New York.