An Alternative Nonresponse Adjustment Estimator

Imbi Traat
University of Tartu, Estonia, imbi.traat@ut.ee

Nonresponse is a severe problem in nowadays surveys. The resulting estimators are biased. In the presence of auxiliary information, several adjustment methods are created. A review is given e.g. in Brick (2013). Nevertheless, bias remains in the estimators, hopefully smaller. The auxiliary variables are required to explain nonresponse and study variables. Often, their explanation power remains low for several study variables. For some study variables, some adjustment methods may even be harmful (Haziza, Lesage, 2016).

In this paper we present one more adjustment estimator for the population mean $\bar{Y} = \sum_{k \in U} y_k / N$, the f-estimator. Regressing y-variable on the auxiliary vector x in sample *s*, and then estimating regression coefficient vector in response set *r*, we end up with the estimator

$$\bar{y}_{\mathrm{f}} = \frac{\sum_{k \in r} d_k f_k y_k}{\sum_{k \in r} d_k}. \quad (1)$$

Here d_k is the sampling weight (inverse of the inclusion probability) and $f_k = \bar{\mathrm{x}}_r' \Sigma_s^{-1} \mathrm{x}_k$, where $\bar{\mathrm{x}}_r = \sum_{k \in r} d_k \mathrm{x}_k / \sum_{k \in r} d_k$ is the mean of x-vector in *r*, and $\Sigma_s = \sum_{k \in s} d_k \mathrm{x}_k \mathrm{x}_k' / \sum_{k \in s} d_k$ is its second moment in *s*.

The factor f_k, when multiplied by response rate, estimates the response probability. The f-estimator is related to the calibration and two-phase nonresponse adjustment estimators, but in some sense, is opposite to them. If to replace f_k in (1) by its inverse f_k^{-1}, we end up with the two-phase estimator. If to replace f_k by $g_k = \bar{\mathrm{x}}_s' \Sigma_r^{-1} \mathrm{x}_k$, involving mean of the x-vector in *s* and its second moment in *r*, we end up with the linear calibration estimator. In fact, g_k is an inverse of f_k in general sense: $\sum_{k \in s} d_k f_k g_k / \sum_{k \in s} d_k = \sum_{k \in r} d_k f_k g_k / \sum_{k \in r} d_k = 1$. In the special case, where x is a group vector, $g_k = f_k^{-1}$, and then linear calibration estimator is equal to the two-phase estimator, but f-estimator differs from them.

We compare the f-estimator, the scaled f-estimator, the calibration estimator, the two-phase estimator and the unadjusted estimator in a simulation study. We have many study variables from a real survey. Different estimators perform best for different study variables. The f-estimator and its scaled form are competitive. The f-estimator works well for the study variables that affect nonresponse directly (the informative nonresponse).

References

Brick, J. M. (2013). Unit nonresponse and weighting adjustments: A critical review. *Journal of Official Statistics*, 29: 329-353.

Haziza, D., Lesage, Eric. (2016). A Discussion of Weighting Procedures for Unit Nonresponse. *Journal of Official statistics*, 32: 129-145.