A Study on Weight Truncation for the National Assessment of Educational Progress (NAEP)

Hyunshik Lee,
Westat,
1650 Research Blvd., Rockville, MD 20850, USA
e-mail: LeeH1@westat.com

Jae-Kwang Kim,
Westat,
1650 Research Blvd., Rockville, MD 20850, USA
e-mail: KimJ@westat.com

Sample survey data are weighted to obtain unbiased survey estimates by reflecting the selection probability and other factors that affect the outcome of the survey. The weighting procedure starts with the base sampling weight, which is usually the reciprocal of the selection probability. This base weight is almost always modified to handle survey nonresponse problem, and also to incorporate external information to improve the efficiency of the survey estimates through post-stratification, raking or other forms of calibration. The final weight used for estimation is often highly variable and the variance of a survey estimate can be substantially inflated depending on the relationship between the weights and the survey variable. When it is judged that excessive variation in the weights is harmful for estimation, weight truncation (or trimming) is frequently employed to control the weight variation and in turn the variance inflation. Weight truncation (more appropriately Winsorization) is a kind of weight modification strategy often used for robust estimation in the presence of outliers (Lee, 1995).

In the literature, many different weight truncation methods have been proposed. Among others, Potter (1990) discussed four methods: (1) the Taylor Series Procedure; (2) Weight Distribution Procedure; (3) Estimated MSE Procedure; and (4) the NAEP Procedure. The first and the third are based on the mean square error (MSE) criterion, which establish the truncation point to minimize the estimated MSE of chosen variables. The first procedure further tries to minimize the relative estimated bias at the same time. The second and fourth procedures reduce the effect of large weights by truncating the large weights and redistributing the excess weights in order to keep all weights below a predetermined level without considering any survey variable. All these are intrinsically design-based since no model is assumed for the survey variables. However, there are also model-based procedures as studied in Elliot and Little (2000). They considered weight modification methods through weight pooling or smoothing, which are formulated under the random effect model for the stratum means of a single variable, where the strata are weight strata of homogeneous weights.

In this paper we consider weight truncation procedures that are applicable to the NAEP. This specifically intended application imposes a fundamental restriction on the choice since the procedure must conform to the survey's operational constraints.

The NAEP is a biennial survey to assess the academic performance of elementary and secondary students in US. The NAEP consists of the main (national) NAEP and the state NAEP. Both programs use a multi-stage sample design aiming at self-weighting of sample students. The base sampling weight undergoes a complex weight adjustment process, of which the main components are school and student level nonresponse adjustments, post-stratification for the main program and raking ratio adjustment for the state program. Although the NAEP design strives for self-weighting (i.e., equal weights), the final weights after all these weight adjustments are very variable. Sometimes, a certain sector of the population is over-sampled and thus, the self-weighting feature is deliberately disturbed. In any case, unusually large weights are truncated using some weight truncation procedures to curtail the effect of the large weights, one of which was studied in Potter (1990).

It is important to understand that the weighting operation of the NAEP data is carried out without having the survey data, which are processed by a different organization simultaneously while weighting is done. Therefore, weight truncation procedures that require survey data cannot be used. Moreover, it is generally believed that the association between the survey variables (test scores) and the weights is weak, and therefore, not much benefit would be realized even when the data are available.

The goal of the NAEP weight truncation procedure is of course to reduce the variance of the survey estimates that is inflated by large variation of the survey weights without incurring a large bias caused by weight truncation. The increase of the variance due to weight variation is often approximated by $1+c_w^2$ where c_w^2 is the relative variance of the weights, which is also an approximate formula for the design effect. The NAEP weight truncation procedures are primarily designed to control this measure. The procedure for the main NAEP checks individual school's contribution to the total variation of the weights. If its contribution is too large, its weight is truncated as described in the following. Let $v_i = \sum_{j\in s_i} w_{ij}$ be the sum of student weights w_{ij} of school i. If $(v_i - \overline{v})^2 > d\sum_{k\in S}(v_k - \overline{v})^2 \big/ n$, where d is an appropriately chosen constant, S is the NAEP sample of schools, and n is the school sample size, then the weight contribution of school i is considered too large and its weight is iteratively truncated until the inequality is reversed. In the 1998 NAEP, d was 10.

Although the weight truncation procedure of the state NAEP has the same goal as the main NAEP, its procedure is quite different. It uses the three-median rule, where an individual student weight larger than three times the median of the weights within each state is truncated at that point. No iteration is needed in this procedure because the truncation rule is satisfied by just one application of the rule.

Alternatives to the current procedures have to satisfy the primary objective of controlling the variance inflation (or design effect) at an acceptable level without causing a large bias. Under this situation, we use a simple weight model given by $w_k = \mathbf{x}_k b + e_k$ to systematize the weight truncation procedure, where $\mathbf{x}_k$ are vectors of auxiliary variables, which include design variables among others and possibly domain indicators. For other applications, the survey variables can be included in the model. The term $E(w_k) = \mathbf{x}_k b$ represents the desirable weights and the error e_k is to take various weighting adjustments into account, which cause the outlying weight problem. Assuming the normality of the error term, we use t-test to decide truncation of an individual weight and its truncation point if needed. This approach provides a flexible way of accommodating different truncation points for different parts of the sample, which is particularly useful if we want to control domain level design effects. We employ transformation when the error distribution is skewed and robust estimation to fit the model to avoid the damaging influence of outlying weights. Other alternatives we study include the weight distribution procedure of Potter's (1990), and the current and other weight truncation methods that establish the truncation point differently. We evaluate the current methods and alternatives using the 1998 NAEP main and state data with respect to the variance and bias, and make recommendation on the choice of the procedures.

References

Elliott, R.E., and Little, R.J.A. (2000). Model-Based Alternatives to Trimming Survey Weights. *Journal of Official Statistics*, 16, 191-209.

Lee, H. (1995). Chapter 26. Outliers in Business Surveys. *Business Survey Methods*, B. Cox, D. Binder, N. Chinnappa, A. Christianson, M. Colledge, and P. Kott (eds.), pp. 503-526. New York: Wiley.

Potter, F. (1990). A Study of Procedures to Identify and Trim Extreme Sample Weights. *Proceedings of the Survey Methods Section*, American Statistical Association, pp. 225-230.

Resume

En échantillonage, des poids inégaux, quoique nécessaire afin d'avoir des estimations non biais éees, peuvent parfois avoir un effect dramatique sur les estimés car une grande variation de ces poids po urrait augmenter substantiellement la variance de ces estimés. Des procédures visant à tronquer les poid s sont fréquemment utilisées pour réduire les effects négatifs d'une grande variation dans les poids d'éha ntillon. L'enquête du National Assessment of Educational Progress (NAEP) utilise des procédures basée s sur un principe simple pour tronquer les poids les plus larges. Cet article évalue les procédures de tron cation des poids déjà connues et étudie aussi bien des procédures alternatives, incluant une procédure no uvellement proposée qui présente une façon systématique de tronquer des poids larges.