NAEP Linking Error

March 22, 2021

Introduction

Students who take paper bases assessments (PBA) and digital-based assessments (DBA) may experience items differently and respond to them as if the item parameters are different–for example, an item may be more difficult in DBA format than in PBA format. The effect this has on the estimated contrasts between PBA and DBA versions of the assessment is called *linking error*; NCES tries to capture the effect change has on student responses in the variance estimate between the two modes. While it is possible to conceptualize the linking error as a contrast shared across respondents, NCES instead conceptualizes linking error as an individual level property that is heterogeneous across students.

NCES uses three sets of plausible values to capture this:

- A first set of plausible values are used to estimate the quantity of interest
- A second set of plausible values are used to estimate sampling variance
- A third set of plausible values are used to estimate imputation variance

NCES distributes the data with only the first set of plausible values (PVs) present, so these second and third set must be calculated before any variance can be calculated. If there are no subscales in the subject, there are two sets, one on set that are scaled to the reporting scale and another that remains unscaled in the *theta* scale that came from the Item Response Theory (IRT) results. When there are subscales in the subject, then there is a set of scaled and a set of unscaled (theta) plausible values per subscale as well as composites. When there are subscale, the subscales are used to form the composite values in the second and third set of plausible values used for variance estimation.

Estimation

A statistic is estimated using the original scale score plausible values by calculating the statistic with the full sample weights and averaging the statistic across the plausible values.

$$Q = \frac{\sum_{j=1}^{J} Q_j}{J} \tag{1}$$

where Q is the estimated statistic Q_j is the statistic estimated with the jth set of plausible values, of which there are J, and all estimates use the full sample weights (`origwt` on NAEP).

Overview

For this exercise we will use the DBA student's unscaled scores and scale them with an equation of the form

$$X = a + b\theta \tag{2}$$

where X is the scaled score a is the location parameter, b is the scale parameter, and θ is the unscaled (theta scale) score.

The values a and b are chosen so that students who took the DBA will have their θ scores mapped to have the same mean and standard deviation as students who took the PBA. This scaling is performed by imputation index for the imputation variance and by replicate weight for the sampling variance, as described in the next two sections.

As is typical with NAEP, scores are scaled by subscale and then the composite score is a weighted combination of those

$$X_c = \sum_{k=1}^{K} \beta_k \cdot X_k (\#eq : composite) \tag{3}$$

where X_c is the composite score, β_k is the weight applied the the kth subscale, and X_k is the kth subscale score. Note that the sum of the β values is always one. If there are no subscales, the overall score can be used in the following sections and the number of subscales is simply one.

Students who took the DBA will have the variable `dbapba` set to `"DBA"` while students who took the PBA will have the variable set to `"PBA"`.

Imputation variance Ideally this would happen by summing over all possible combinations of the 20 PBA and 20 DBA imputations, so 20! possible combinations of plausible values would be calculated. To minimize computation time, only five permutations of the plausible values are chosen, they are shown in the permutation table (Table @ref(tab:PERM)) where the row number (not shown) is the index of the DBA plausible value to use and the cell value is the index of the PBA plausible value to use.

The data already has the plausible values used for estimation on it but needs the imputation variance and sampling variance plausible values added.

Algorithm for generating the imputation variance plauislbe values:

- define w_{DBA} as the vector of weights for for students who took the DBA. The vector has length n_{DBA}.
- define w_{PBA} as the vector of weights for for students who took the PBA. The vector has length n_{PBA}.
- for each column (n) of the permutation table
  - for each row (j) of the permutation table
    - for each subscale (k)
      - define X_k as the vector of the kth subscale, scaled score PV for all students who took the PBA. Select the PV based on the element in the permutation table for this row and column (the j, nth element from the permutation table).
      - define θ_k as the vector of the kth subscale, PV j for all students who took the DBA.
      - define $\mu_t = \frac{\sum w_{DBA}\theta_k}{\sum w_{DBA}}$
      - define $s_t = \sqrt{\frac{\sum w_{DBA}(\theta_k - \mu_t)^2}{\sum w_{DBA}}}$
      - define $\mu_x = \frac{\sum w_{PBA}X_k}{\sum w_{PBA}}$
      - define $s_x = \sqrt{\frac{\sum w_{PBA}(X_k - \mu_x)^2}{\sum w_{PBA}}}$
      - generate a new scaled score for the DBA students, centered at the mean for PBA students using $X'_k = b + a\theta_k$ where $a = \frac{s_x}{s_t}$ and $b = \mu_x - a\mu_t$.
      - the j, nth *imputation variance PV* for subscale k is X'_k for students who took the DBA and X_k for students who took the PBA.
    - generate the composite for DBA students using the composite formula $X'_c = \sum_k \beta_k X'_k$.
    - the j, nth *imputation variance PV* for the composite is X'_c for students who took the DBA and X_c for students who took the PBA.

```
## Warning in kable_pipe(x = structure(c("1", "2", "3", "4", "5", "6", "7", : The
## table should have a header (column names)
```

Table 1: Permutation table

1	14	6	2	19
2	11	4	14	10
3	18	3	7	13
4	15	9	9	8
5	20	11	20	17
6	17	10	13	12
7	3	2	16	1
8	16	17	8	11
9	12	19	4	5
10	6	13	11	6
11	19	7	19	18
12	2	8	1	9
13	1	15	18	15
14	13	18	10	2
15	5	20	6	3
16	7	12	12	7
17	4	5	15	14
18	9	16	3	20
19	8	14	5	16
20	10	1	17	4

The above steps need to be performed only once. Then, for each statistic the imputation variance is calculated as follows. Using the imputation variance plausible values, the linking imputation variance is then calculated per permutation (column) and then averaged acrosss the permutations.

- for each column (n) of the permutation table
  - for each row (j) of the permutation table
    * calculate the statistic in question using the j, nth imputation variance PV and the full sample weights, call this Q_{jn}
  - Calculate $Q_{\cdot n} = \frac{1}{J}\sum_{j=1}^{J} Q_{jn}$ where J is the number of permutation table rows
  - calculate $V_{imp,n} = \frac{J+1}{J(J-1)}\sum_{j=1}^{J}\left(Q_{jn} - Q_{\cdot n}\right)^2$
- Calculate $V_{imp} = \frac{1}{N}\sum_{n}^{N} V_{imp,n}$ where N is the number of columns in the permutable table

The value V_{imp} is the estimated imputation variance for the statistic in question.

Sampling variance Then generate the sampling variance plausible values, using only the first plausible value but reweighting using the replicate weights.

- for each subscale (k)
  - define X_k as the vector of the kth subscale, scaled score PV for all students who took the PBA. Select the first plausible value.
  - define θ_k as the vector of unscaled scores for the kth subscale for students who took the DBA. Select the first plausible value.
  - for each replicate weight (i)
    * define $w_{DBA}^{(i)}$ as the vector of the ith replicate weights for for students who took the DBA. The vector has length n_{DBA}.

* define $w_{PBA}^{(i)}$ as the vector of the ith replicate weights for for students who took the PBA. The vector has length n_{PBA}.
* define $\mu_t^{(i)} = \frac{\sum w_{DBA}^{(i)} \theta_k}{\sum w_{DBA}^{(i)}}$
* define $s_t^{(i)} = \sqrt{\frac{\sum w_{DBA}^{(i)} (\theta_k - \mu_t^{(i)})^2}{\sum w_{DBA}^{(i)}}}$
* define $\mu_x^{(i)} = \frac{\sum w_{PBA}^{(i)} X_k}{\sum w_{PBA}^{(i)}}$
* define $s_x^{(i)} = \sqrt{\frac{\sum w_{PBA}^{(i)} (X_k - \mu_x^{(i)})^2}{\sum w_{PBA}^{(i)}}}$
* generate a new scaled score for the DBA students, centered at the mean for PBA students using $X_k'^{(i)} = b^{(i)} + a^{(i)} \theta_k$ where $a^{(i)} = \frac{s_x^{(i)}}{s_t^{(i)}}$ and $b^{(i)} = \mu_x^{(i)} - a^{(i)} \mu_t^{(i)}$.
* the ith *sampling variance PV* is then $X_k'^{(i)}$ for students who took the DBA and X_k for students who took the PBA.

The above steps need to be performed only once. Then, for each statistic the sampling variance is calculated as follows. Using the sampling variance plausible values, the linking sampling variance is then calculated per replicate weight and summed.

- for each replicate weight (i)
  - calculate Q_i as the statistic in question using the ith sampling variance PV and the ith replicate weight
- calculate the mean $Q_{\cdot} = \frac{1}{I} \sum_{i=1}^{I} Q_i$ where I is the number of replicate weights
- calculate $V_{samp} = \sum_{i=1}^{I} \left(Q_i - Q_{\cdot}\right)^2$

The value V_{samp} is the estimated sampling variance for the statistic in question.