Methods Used for Gap Analysis in EdSurvey

Developed by Ting Zhang, Paul Bailey, Michael Lee, Michael Cohen, and Jiao Yu[*†]

May 11, 2018

Introduction

Gap analysis is a function in the `EdSurvey` package that calculates the difference between two statistics (e.g., mean scores, achievement level percentages, percentiles, and student group percentages) for two groups in a population. When computing the variance estimation of a difference, the assumption about two groups decides how covariances are handled and how the degrees of freedom are calculated. In a simple random sample data, an independent t-test often is employed to calculate the gap of the estimates and its variance, where a covariance is assumed to be zero and not accounted in the variance estimation. On the other hand, in a complex sampling designed assessment that involves plausible values, even two sampled groups are not overlapping, they can still have covariance because of the way that they were sampled (e.g., they attended the same school or were in the same classroom when sampled) and the way that their scores were obtained (plausible values that involve multiple imputation). In other words, this covariance arises from the fact that students in the same school have test scores and background factors that are correlated with each other and is analogous to how the NCES calculates the overall scale scores and their standard errors for the National Assessment of Educational Progress (NAEP). Therefore, when two groups are from the same assessment sample (same assessment and year), `EdSurvey` assumes a dependence of the groups when calculating the gap and accounts for covariance at the stratum level, as well as covariance in imputation that is shared nationally. For observed values from different assessment samples (e.g., 2011 Mathematics and 2015 Mathematics), `EdSurvey` assumes the groups do not have a covariance and treats them independently.

This document first compares gap analysis results under the two different assumptions—showing a difference in variance estimation and p-values—and then describes the methodology `EdSurvey` uses to arrive at the results.

Comparison of Values

Table 1 displays a comparison of relevant statistics of the average scale score gap between Black and White students in 2011 and 2015 NAEP Mathematics Grade 4 assessments. The first block shows the outcomes from the NAEP Data Explorer (NDE); the second block displays t-test outcomes from SAS when assuming the two groups have no covariance; and the third block shows `EdSurvey` gap analysis results under the dependent assumption that accounts for the covariance between the two groups.

Table 1: Black-White Mean Score Gap

Software	Assumption	Year	Black Avg.	White Avg.	B-W Diff.	SE Diff.	DoF Diff.	p-Value Diff.
NDE		2015	223.969	248.323				0.000
NDE		2011	223.921	248.959				0.000
NDE		2015−2011						0.306
SAS	Independent	2015	223.969	248.323	−24.354	0.502	81.389	0.000
SAS	Independent	2011	223.921	248.959	−25.039	0.438	90.102	0.000
SAS	Independent	2015−2011	0.049	−0.636	0.685	0.667	165.779	0.306
EdSurvey	Dependent	2015	223.969	248.323	−24.354	0.462	45.768	0.000
EdSurvey	Dependent	2011	223.921	248.959	−25.039	0.434	62.786	0.000
EdSurvey	Dependent	2015−2011	0.049	−0.636	0.685	0.634	103.513	0.283

[*]This publication was prepared for NCES under Contract No. ED-IES-12-D-0002 with the American Institutes for Research. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

[†]The authors would like to thank Dan Sherman and Victor Bandeira de Mello for reviewing this document.

When assuming the two groups have no covariance and using independent t-tests, SAS obtained results that match with NDE in terms of mean estimates and p-values.[1]

Under the dependent assumption and after accounting for the covariance between the two groups, we observed an agreement between SAS and `EdSurvey` in terms of the mean estimate, but they disagreed in standard errors, degrees of freedom, and p-values.[2]

The results also showed that `EdSurvey` agreed with NDE in terms of mean estimates but not the p-values, which suggests that the `EdSurvey` and NDE methodologies differ in some way—most likely the calculation of the standard error and/or the calculation of the degrees of freedom resulting from the different assumptions regarding covariance.

Variance Estimation Strategy

This section describes the gap analysis method used by `EdSurvey` for how covariances are handled and how the degrees of freedom are calculated. The method for calculating the variance of a difference and the degrees of freedom are mathematically equivalent to the "Dependent via Differences" method described by NCES (2017).

Calculation of Variances and Covariances

`EdSurvey` assumes observed values from different assessment samples (e.g., 2011 Mathematics and 2015 Mathematics) do not have a covariance and treats them independently.

When two values are from the same sample (same assessment and year), the groups could overlap in several different ways, and `EdSurvey` uses the same variance estimator for every type of overlap. This estimator accounts for covariance at the stratum level, even when the groups are not overlapping (e.g., Males and Females can still have covariances when they attend the same schools), as well as covariance in imputation that is shared nationally.

The goal is then to calculate variance of the difference between a statistic calculated on two populations A and B; call these statistics θ_A and θ_B. Then the variance of the difference is given by

$$Var(\theta_A - \theta_B) = Var(\theta_A) + Var(\theta_B) - 2Cov(\theta_A, \theta_B)$$

where $Var(\cdot)$ is the variance and $Cov(\cdot, \cdot)$ is the covariance. This section explains the calculation of the covariance, which also suffices to explain the calculation of the variances using the formula $Cov(\theta_A, \theta_A) = Var(\theta_A)$. Thus, the reader who wishes to know how $Var(\theta)$ is calculated can simply find $Cov(\theta, \theta)$.

The covariance term ($Cov(\theta_A, \theta_B)$) is given by the sum of the sampling variance (Cov_{jrr}) and the imputation variance (Cov_{imp}):

$$Cov(\theta_A, \theta_B) = Cov_{jrr}(\theta_A, \theta_B) + Cov_{imp}(\theta_A, \theta_B)$$

The sampling covariance is

$$Cov_{jrr}(\theta_A, \theta_B) = \frac{1}{M} \sum_{p=1}^{M} \sum_{j=1}^{J} (\theta_{Ajp} - \theta_{A0p})(\theta_{Bjp} - \theta_{B0p})$$

where θ_{Ajp} is the estimate of θ_A using the jth jackknife replicate weights and the pth plausible value, θ_{A0p} is the estimate of θ_A using the full sample weights and the pth plausible value, M is the number of plausible values used in the calculation, and J is the number of jackknife replicates.

[1]Note because the p-value for the Black-White score gap is too close to zero to meaningfully compare between the three calculations, the difference between the change in the Black average score from 2011 to 2015 and the change in the White average score from 2011 to 2015 also is computed.

[2]Note that `B-W Diff.`, `SE Diff.`, and `DoF Diff.` are unavailable in NDE gap analyses; see https://bit.ly/2vNTzJH.

The imputation variance is estimated according to the covariance analog to the Rubin (1987) variance estimate:

$$Cov_{imp}(\theta_A, \theta_B) = \frac{M+1}{M(M-1)} \sum_{p=1}^{M} (\theta_{Ap} - \theta_{A0})(\theta_{Bp} - \theta_{B0})$$

where θ_{Ap} is the estimate of θ_A using the pth set of plausible values, and θ_{A0} is the average of θ_{Ap} across the plausible values.

Calculation of the Degrees of Freedom

These formulas allow for the calculation of the variance of $\theta_A - \theta_B$ and the calculation of the t-statistic:

$$t_{\theta_A - \theta_B} = \frac{\theta_A - \theta_B}{\sqrt{Var(\theta_A - \theta_B)}}$$

When a difference is computed within a sample, this t-statistic then has degrees of freedom given by the Johnson and Rust (1992) corrected degrees of freedom (dof_{JR})

$$dof_{JR} = \left(3.16 - \frac{2.77}{\sqrt{J}}\right) dof_{WS}$$

where dof_{WS} is the Welch-Satterthwaite degrees of freedom estimate:

$$dof_{WS} = \frac{1}{M} \sum_{p=1}^{M} \frac{\left[\sum_{j=1}^{J} \left[(\theta_{Ajp} - \theta_{A0p}) - (\theta_{Bjp} - \theta_{B0p})\right]^2\right]^2}{\sum_{j=1}^{J} \left[(\theta_{Ajp} - \theta_{A0p}) - (\theta_{Bjp} - \theta_{B0p})\right]^4}$$

When a difference is computed across samples, the traditional Welch-Satterthwaite equation is used. This is the case for the rows labeled "2015$-$2011" in Table 1.

References

Johnson, E.G., and Rust, K. F. (1992). Population inferences and variance estimation for NAEP data. *Journal of Educational Statistics, 17*(2), 175–190.

National Center for Education Statistics. (2017). *Dependent* T-*Tests: NDE Statistical Specification.* Washington, DC: Author. Retrieved from https://nces.ed.gov/nationsreportcard/NDEHelp/WebHelp/dependent_t-tests___nde_statistical_specification.htm

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.* New York, NY: Wiley.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin, 2*, 110–114.

Welch, B. L. (1947). The generalization of 'student's' problem when several different population variances are involved. *Biometrika, 34*, 28–35.