PSS Weights, Variance, and Missing Data

Slide 1 of 9

Title Slide: PSS Weights, Variance, and Missing Data

Slide 2 of 9

Although PSS is a universe survey, it used a complex sample design to construct the area frame component of schools from which data were collected to augment the list frame. As discussed in the common modules, when analyzing data from complex surveys, certain procedures must be used to ensure that estimates made from the data are representative of the population and that hypothesis tests are accurate. Specifically, a weight must be applied and standard errors must be calculated using appropriate procedures. This module discusses these issues specifically in relation to analyses of data from PSS.

For more general information about weighting and standard error calculation, please view the Common Modules titled 'Analyzing NCES Complex Survey Data' and 'Statistical Analysis of NCES Datasets Employing a Complex Sample Design,' if you have not already done so. These modules can be accessed by clicking the underlined text, 'weights that must be applied,' and 'calculating appropriate standard errors'.

Additionally, this module describes how to handle missing data in the PSS data files, specifically which missing data codes are used and how to handle missing data to ensure accurate data analysis.

Slide 3 of 9

As a universe survey, PSS attempts to include every private school in the United States. Therefore, a universe list of private schools meeting the PSS definition must be created. The definition includes those schools that are not supported primarily by public funds, that provide classroom instruction for one or more of grades kindergarten through 12th or comparable ungraded levels, and that have one or more teachers.

NCES used a dual frame approach for building the private school universe in 2011-2012. The dual frame consists of a list frame and an area frame. The list frame was the primary means for improving coverage of private schools. The list frame was constructed using the 2009-2010 PSS universe list matched to private schools obtained from information from state departments of education and private school associations. To identify private schools that might have been overlooked in the list frame, a group of geographic areas selected through stratified cluster sampling was searched for private schools.

The combination of the list frame schools and additional schools identified in the area search comprised the schools included in the 2011-2012 PSS universe. Additional information in the data file users manual about the PSS dual frame approach can be accessed by clicking the corresponding underlined screen text.

Slide 4 of 9

As was discussed in the Common Module, 'Analyzing NCES Complex Survey Data,' weighting should be used in analyses of data from surveys such as PSS in order to make estimates produced from the sample representative of the target population. There is a single weight variable in the PSS data file (PFNLWT, spelled "P-F-N-L-W-T"), which should be used in analyses of PSS data. The final weight assigned to each school respondent accounts for the unequal probabilities of selection of geographic areas sampled for the area frame of schools, as well as nonresponse. It is the product of the Base Weight and the Nonresponse Adjustment Factor. The Base Weight is the inverse of the probability of selection of the school. It is equal to one for all list-frame schools. For area-frame schools, it is equal to the inverse of the probability of selecting the geographic area in which the school is located. The Nonresponse Adjustment Factor is the weighted ratio of the total eligible in-scope schools to the total responding in-scope schools within cells. For schools in the list frame, for which there is more information, the cells used to compute the nonresponse adjustment were defined by affiliation, locale type, grade level, Census region, and enrollment. The nonresponse adjustment cells for area-frame schools were defined by three-level typology and grade level. More information on the cells used to compute the nonresponse adjustment is available in the Data File Users' Manuals, which can be accessed by clicking on the underlined screen text, 'Nonresponse Adjustment Factor,' and scrolling to the PSS User's Manuals.

Slide 5 of 9

Weights matter in PSS because in computing estimates without the use of the final weight, any statistical software package is assuming a simple random sample (or SRS) with an equal probability of selection for each respondent. PSS does not use an equal probability of selection, so the final weight value varies from one respondent record to another. Generating an estimate from PSS data without the use of the final weight yields a result that is not accurate because it does not take into account the unequal probabilities of selection used in PSS.

Slide 6 of 9

In the common module titled, 'Statistical Analysis of NCES Datasets Employing a Complex Sample Design,' two standard error calculation procedures were discussed: Replication Techniques and Taylor Series linearization. Replication is a method that calculates appropriate standard errors based on differences between estimates from the full sample and a series of created subsamples, or replicates. In this method, you need to select replicate weights that are associated with the PSS data file's final weight. The replicate weights associated with the final weight PFNLWT are REPW1 (spelled R-E-P-W-1) through REPW88 (spelled R-E-P-W-8-8). It is important to include the entire set of replicate weights in the weight statement. PSS replication weights were produced using a Balanced Repeated Replication (or BRR) method so that is the method that should be specified within the statistical software used for analysis. It is important to note that Balanced Repeated Replication is the only accurate method for calculating standard errors in PSS, as no Primary Sampling Unit or PSU codes and strata identifiers, which are required for Taylor series linearization, are provided in the PSS data files.

Slide 7 of 9

Categorical estimates, those with Yes/No response categories, Likert scales, or with a limited number of discrete categories use PROC CROSSTABS. Continuous data estimates, such as the total student enrollment, require PROC DESCRIPT. Percentage estimates require PROC RATIO. In addition to the resource documents provided in the Common Module titled, 'Statistical Analysis of NCES Datasets Employing a Complex Sample Design,' you can click the underlined screen text 'SAS Code' to view a resource document that provides examples of using PROC DESCRIPT and PROC RATIO with PSS data.

It is important to note that using this code for your own research purposes will necessitate substituting your variables of interest.

Slide 8 of 9

Since the 1991-1992 PSS, a complex imputation process has been used to compensate for item nonresponse. All items that are missing data after the editing process are imputed. Two approaches to imputation are used. Missing values are first imputed using donor imputation, where values are created by extracting data from a record with similar characteristics using a hot-deck imputation methodology. For records that still have missing values for some items after donor imputation, the data are manually imputed. A value consistent with information from the data record, sample file record, and questionnaire is identified. These imputation methods are described in detail in the 'Data File Users' Manuals, which can be accessed by clicking on the underlined screen text, 'imputation methods,' and scrolling to the PSS User's Manuals. For each data item with imputed values, an imputation flag variable is created to allow users to identify imputed values. These flags are named F_ variable name. If the response for the item was not imputed, the imputation flag is set equal to 0. The flag is set to 4 if the missing value was imputed using a donor value. If an item was imputed manually, the flag is set to 5. Users can employ the imputation flag to delete the imputed values, use alternative imputation procedures, or account for the imputation in computations of the reliability of the estimates produced from the dataset.

Slide 9 of 9

This module provided information about the universe design, weighting, and variables associated with calculating estimates and standard errors for PSS data. The module also detailed how the final weight must be applied and how standard errors must be calculated to ensure that estimates made from PSS data are representative of the population and that hypothesis tests are accurate. Additionally, this module described how to handle missing data in the PSS data files. Specifically which missing data codes are used and some tips for handling missing data to ensure accurate data analysis. Important resources that have been provided throughout the module are also summarized in this slide along with the module's objectives for your reference.

You may now proceed to the next module in the series, or click the exit button to return to the landing page.