Slide 1 of 13

Considerations for Analysis of Postsecondary Education Data

Slide 2 of 13

Module Objectives

Describe the analysis considerations for postsecondary education data regarding

- Using the derived files
- Using the source files
- Considering the "Applies To" fields and reviewing frequencies
- Generalizing beyond defined sampling strata
- Noting study changes across administrations
- Considering trend analyses
- Using postsecondary education transcripts
- Understanding weight component files

Slide 3 of 13

Using the Derived File

- NCES datasets are the product of many data sources and you should be working with the derived file
  - Longitudinal products will contain multiple derived files, one for each phase of the study
  - For example, the BPS:04/09 product will contain derived files corresponding to study members' NPSAS:04 (base year) interview, BPS:06 (first follow-up) interview, and the BPS:09 (final follow-up) interview
- File names typically contain information about the administration year (e.g., N4 is the code for the NPSAS:04 derived file, with F6 and F9 corresponding to BPS follow-ups in 2006 and 2009, respectively)
- **Since the derived files are cumulative, only the file listed first within the README file should be of concern to researchers**

Slide 4 of 13

Selecting Variables & Combining Files

- Users who enter the source file should use caution in selecting variables for analysis
- These variables are used to prepare derived files and are typically included on the data products without edits to
  - Variable names
  - Labels
  - Descriptions
  - Data values
- Note: users should carefully review syntax and results associated with merge commands

Slide 5 of 13

Carefully Read "Applies To" Fields and Always Review Frequencies

- NCES derived variables include an “Applies To” note
  - Specifies the subpopulation of study members for whom a valid value should be expected
- For variables that represent a dollar value, consider the meaning of a zero value
  - For example, the average of a “total student aid” variable excluding zeros indicates the average among the aided students
  - Including zeros would create an average across all students, irrespective of whether they were aid recipients
- Considerations when using the restricted-use CD
  - Check for unweighted frequencies of each variable
  - Be sure to identify special data values that should be excluded from statistics (e.g., missing data coded as a -3 or a -9)
  - Ensure that a sufficient number of cases are present for analysis

Slide 6 of 13

Use Caution Generalizing Beyond Defined Sampling Strata

- Studies are designed to generate estimates that are representative of our explicit strata, such as institution type or, in certain years of NPSAS, a subset of states
  - Sampling strata are detailed in each study's Methodology Report
- We implicitly stratify to other important characteristics (e.g., institutions' status as an Historically Black Colleges and Universities (HBCU)) to help ensure coverage in the final sample
  - While other interesting groups will be found in the data, many were not explicitly sampled, and estimates may be based on relatively low sample sizes
- Therefore, consider the use of confidence intervals to provide the reader information about the precision of your estimates and be prudent in describing findings
- Never use any sample survey product to make statements about specific institutions

Slide 7 of 13

Changes in NPSAS and Related Studies Across Administrations

- Institutional and Student Eligibility
  - Most significant change in institutional eligibility occurred in 2000, which limited institutions to those participating in Title IV Federal Student Aid Programs
  - Most significant change in student eligibility occurred in 2012 when NCES clarified that institutions should consider eligible students enrolled entirely in remedial coursework so long as their program of study was otherwise Title IV eligible
- Disproportionate Sampling Rates
  - Specific policy needs in a given year may motivate student oversamples, such as a focus on SMART grant recipients in NPSAS:08 or sub-baccalaureate students in NPSAS:12
  - Some studies always oversample to increase the likelihood of yielding key student populations, such as potential teachers in Baccalaureate and Beyond (B&B)
- Puerto Rican Institutions
  - In 2012, Puerto Rican institutions were not sampled
  - For consistent estimates across time, analysts should use the COMPTO87 variable in prior NPSAS studies to ensure only institutions in the 50 States and the District of Columbia are included
- See each most recent study's methodology report for additional information

Slide 8 of 13

Cautions in Trend Analysis

- When examining dollar amounts over time (e.g., tuition costs or loan amounts), analysts are cautioned to adjust all estimates for inflation to the current (or other relevant) year
  - The Bureau of Labor Statistics (BLS) has a publicly available tool to assist users, with detailed tables available on the BLS website
- Analysts should be cognizant of changes in federal financial aid policy over the duration of their time series analysis and consider whether those changes may have a material influence on their results

Slide 9 of 13

Cautions in Analyzing Subpopulations

- Many research questions focus on a specific subpopulation
- Do not delete cases from the data set or use standard subsetting commands (e.g., if)
    - Your software may require all the cases to calculate the proper standard errors
- Instead, use your software's subpopulation command

Slide 10 of 13

Postsecondary Education Transcripts

B&B

- The most recent available transcript data for B&B students accompanies the 2007-08 cohort
- For each sample member, NCES sought to collect the final transcript from the institution awarding the baccalaureate degree
- No transcripts were collected for the 2015-16 cohort.

BPS

- The first set of transcript data made available for BPS students accompanies the 2003-04 cohort. Transcripts are also available for the 2011-12 BPS cohort.
- For each sample member, NCES sought to collect all undergraduate transcripts
- Analysts using BPS transcripts must select one of two weights
    - WTC, when analyzing only data contained in the transcript file (that is, not linked to any student data)
    - WTD, when analyzing transcript data in combination with other BPS: data elements

Slide 11 of 13

Weight Component Files

- Restricted-use data includes the final analysis weight and weight components
  - Final analysis weight is the weight that includes the probability of institutional and student selection and nonresponse and post-stratification adjustment
  - Each weight component is multiplied together to create the final analysis weight
- Each data product's WEIGHT file contains weight components, labeled WT1, WT2, and so forth, as well as the final analysis weight
- Additional information can be found in each study's methodology report

Slide 12 of 13

Module Summary

Described the analysis considerations for postsecondary education data regarding

- Using the derived files
- Using the source files
- Considering the "Applies To" fields and reviewing frequencies
- Generalizing beyond defined sampling strata
- Noting study changes across administrations
- Considering trend analyses
- Using postsecondary education transcripts
- Understanding weight component files

Module Resources

- Derived files
- README files
- Reserve codes
- Sampling strata
- Weights
- Bureau of Labor Statistics (BLS) website
- BLS publicly available tool
- Subpopulation Command Resource Document
- Study-specific methodological report(s) of interest
    - NPSAS
    - B&B
    - BPS