



Analyzing NCES Complex Survey Data

Module Objectives

- Introduce the design and analysis of NCES data
- Identify different types of NCES surveys and studies
- Explain the essentials of sampling and sampling designs used by NCES
- Describe how to use sampling weights and design variables in data analysis

Using NCES Studies

- NCES studies are designed to collect data on a variety of educational topics that are of interest to researchers and policy makers
- NCES relies on the processes detailed in the [NCES Statistical Standards](#) in order to ensure that the data collected are consistent, reliable, complete, and accurate
- NCES monitors and documents the methodologies associated with each dataset. It also analyzes and reports on the data in a variety of products and publications
- NCES sample surveys provide access to nationally representative data that are generalizable to the population and can be used to answer a variety of research questions
- Existing NCES products, including web tables and published reports, may already address some or all of your research questions
- Publicly available tools can assist you in answering simple questions based on several NCES datasets

Analyzing NCES Data



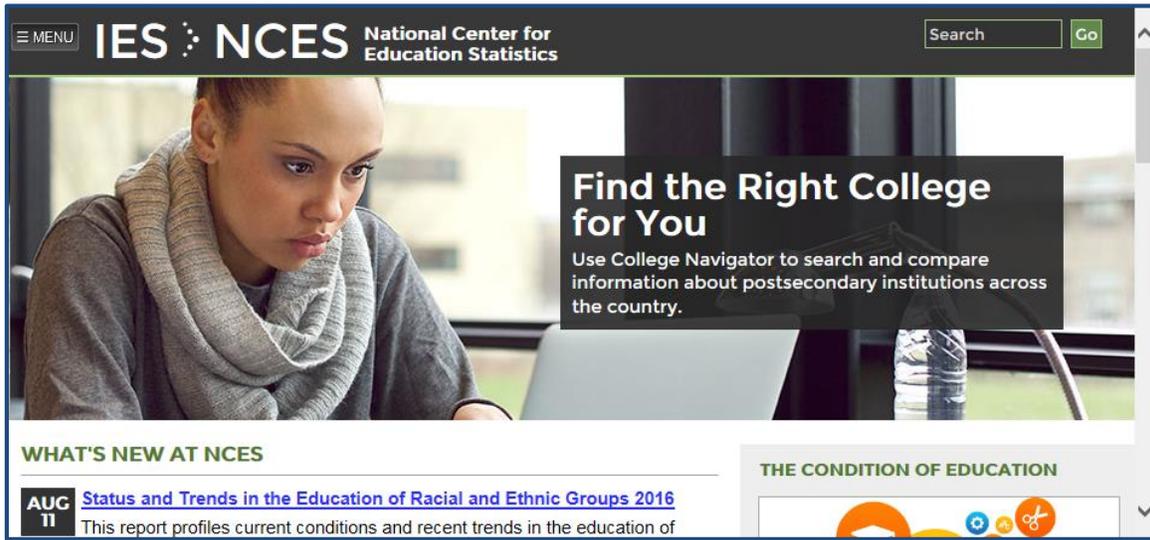
Limitations of Using NCES Studies to Address your Research Question(s)

- Not all education research questions can be answered using these data
- Considerable time and energy must be invested to understand
 - What questions the data can and cannot answer
 - How the data may or may not be appropriate for your analytic purposes
 - How to analyze complex sample survey data

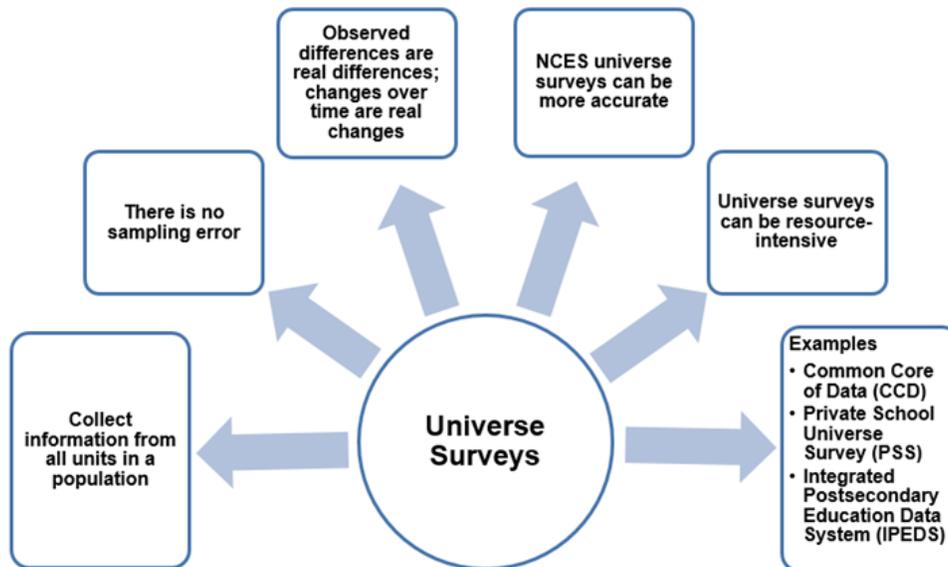
Where to Start with Analysis of NCES data

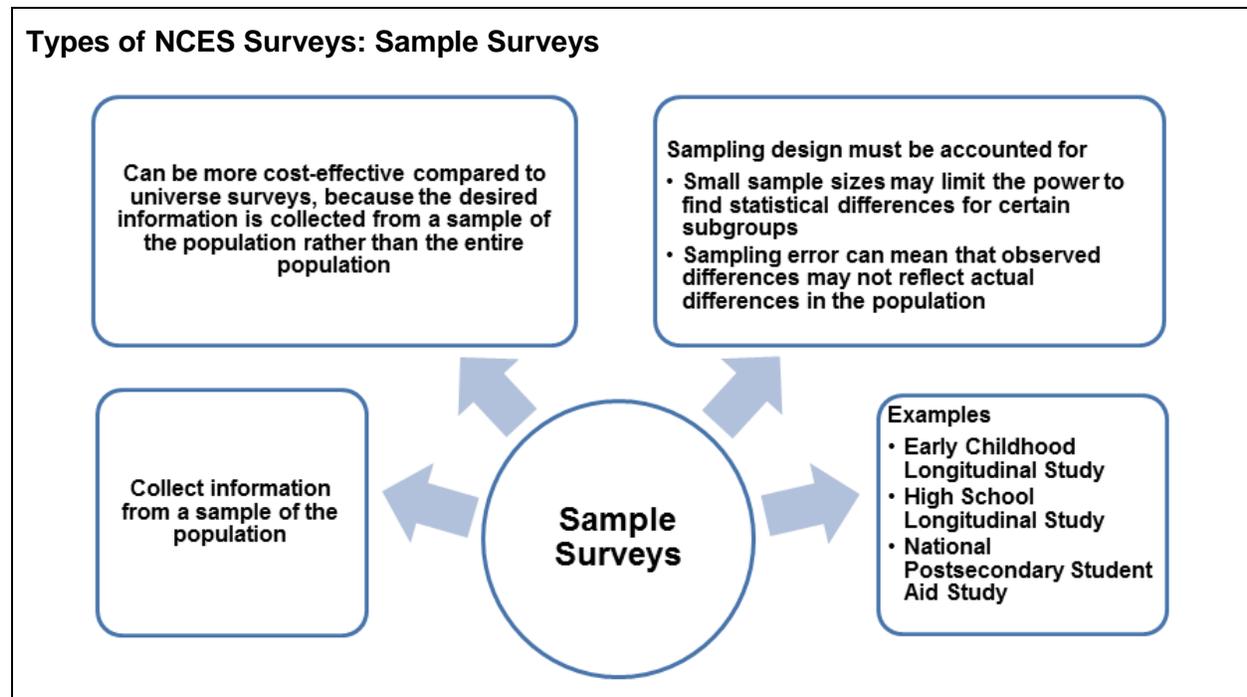
- Familiarize yourself with NCES studies to identify which may be appropriate to address your research question
- Review the available studies, methodology reports, and technical documentation
- Use web tools and publicly available micro-level data for use in the statistical package of your choice to begin your exploration of the data

NCES Data Publications, Data Tools, and Data Products



Types of NCES Surveys: Universe Surveys

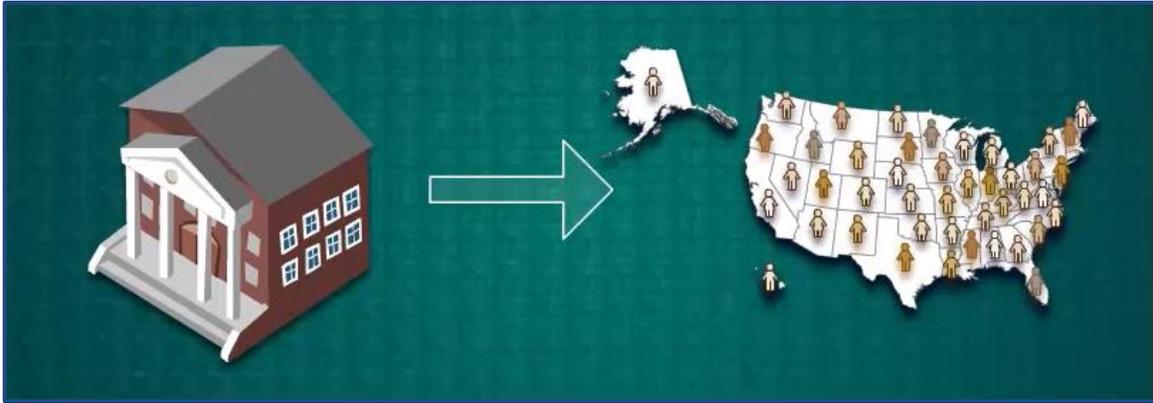




- ### How NCES Selects a Sample
- NCES studies are grounded in a scientific approach
 - They are designed to address specific research goals or questions within populations of interest
 - A sampling design is employed to obtain a representative sample
 - This allows researchers to use smaller numbers of participants in studies while still maintaining validity, reliability, and accuracy

An Example of NCES Sampling

Research goal: We are interested in high school graduates who immediately enroll in a postsecondary education program

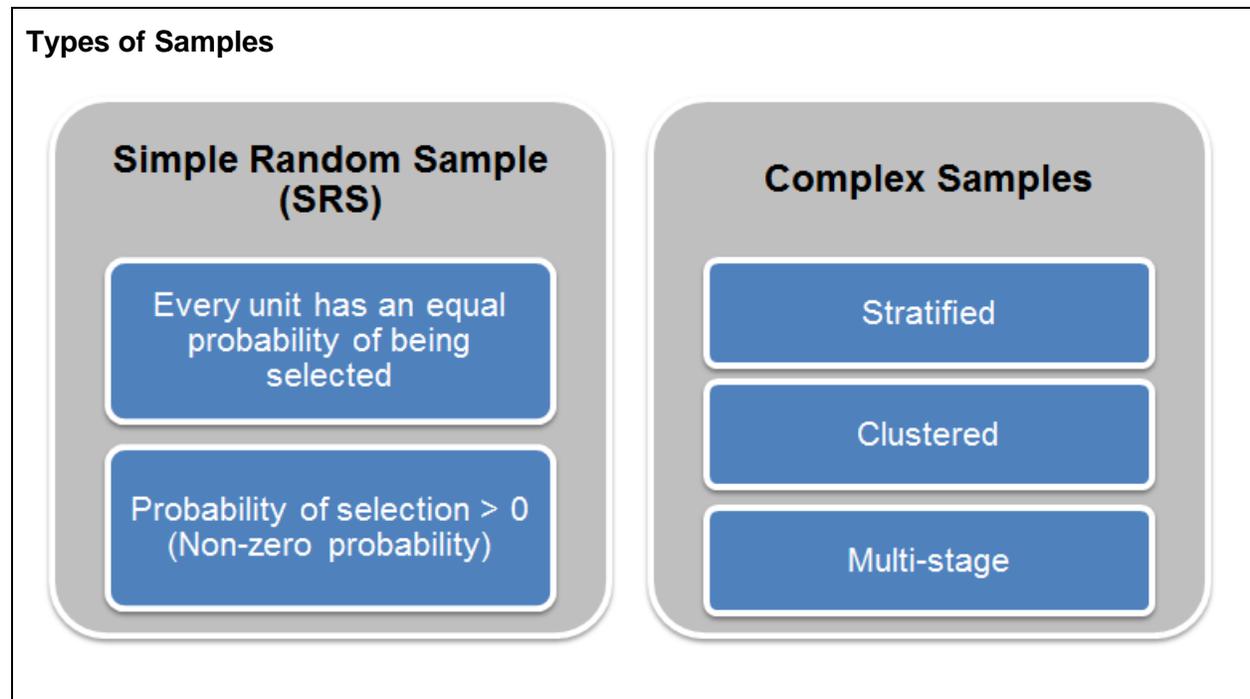


Why NCES Uses Probability Samples

NCES sample surveys use probability samples to ensure

- Representativeness of populations and sub-populations (minimal bias)
- Sample sizes are large enough to detect differences within/between key populations and items (sufficient statistical power)
- Known, non-zero chances of selection from the population of interest





Simple Random Sampling (SRS)

Simple: Each sample member is selected independently of all others

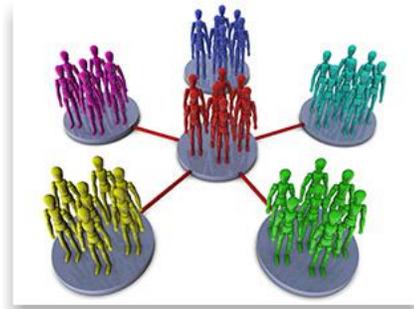
Random: Every unit in the population has an equal non-zero probability of being selected for the sample

Simple Random Samples are easy to implement and analyze. However, they

- May yield too few cases for certain groups
- Can be expensive and thus may not be practical

Complex Sample Designs

- Stratification is used to ensure that different subgroups are adequately represented in the sample by dividing the sampling frame into relevant subgroups prior to sample selection
- Clustering involves the selection of groups of units or clusters (schools within a district or students within schools)
- Multi-stage involves multiple stages of sub-sampling (select schools, then students)



Stratification

Stratification: Used to ensure that different subgroups are adequately represented in the sample. Sampling frame is divided into relevant subgroups prior to sample selection



Complex Sample Designs: Clustering

Clustering and multi-stage cluster sampling can reduce the time and cost of data collection over large areas

- Clustering - Selecting a sample of grouped sampling units, or clusters
 - One-stage Clustering - all the sampling units in the selected clusters are included in the sample
 - Two-stage Clustering - only a sample of units is taken from each selected cluster
- Multi-stage Cluster Sampling - Selecting a sample of respondents from within the clusters
 - Can be employed to reduce the time and cost of data collection

NCES Datasets Based on Studies with Complex Sampling Designs

- NCES datasets are derived from studies with complex sample designs
- When using NCES datasets, appropriate sampling weights **must** be used to produce population-level estimates
- Due to the complex nature of the sample, the variance of estimates **cannot** be accurately calculated via standard SRS statistical methods. You must use a variance estimation technique that is appropriate for the complex sample designs used in NCES surveys
 - NCES provides the sampling design information needed to use complex variance estimation software to compute estimates of variance that reflect the complex sample design of the data collection
 - If this information is not used, or if it is not used correctly, the results of hypothesis testing, or the p values, will be **incorrect**



Sampling Weights

- Used to indicate the relative contribution of an observation
- Equal to the inverse of the probability of a unit being selected into the sample
- ($w = 1/\pi$)
- Adjusted to help account for nonresponse that can alter the relative distributions of the data compared to the target population
 - This adjustment is important because when there are differential patterns of nonresponse, the data can be biased, or not representative of either the population or subgroups of interest
 - Weighting the data alone, without accounting for the complex sample design, will result in many findings that are statistically significant due to the large weighted population estimates that are really population counts

Weights Developed for NCES Studies

NCES develops weights that must be used with NCES data to compute estimates of population parameters that reflect the sampling design. Each dataset provides weights to be used with data from the study.

Cross-Sectional Studies

- Weights are used to make the sample data representative of a given population at one point in time, such as the population of children attending kindergarten through grade 12 in 2007
- Some cross-sectional NCES studies have collected data about the same population over time
- The groups of people included in each data collection of a cross-sectional study are different, so the data from each sample must be weighted with a distinct cross-sectional weight

Weights Developed for NCES Studies (Continued)

Longitudinal Studies

- Weights can also be used to make the sample data representative of a given population at one point in time, or to look at change over time
- Longitudinal studies follow the same group of people over time
- The weights are designed to make the data representative of the target population, or cohort
- NCES longitudinal studies provide weights that can be used to analyze data from one or more rounds of data collection

*More information about the weights available for each study is provided within the [study-specific DLDT modules](#)

Critical Reminders

- Sampling weights must always be used with data from complex samples to produce correct population estimates
- Your statistical software will provide guidance on appropriate syntax for conducting a weighted analysis
- Study-specific examples and simulations of how to apply weights will be presented within the survey-specific modules within this DLDT system

*General examples and simulations of how to use weights will be presented in [Statistical Analysis of NCES Datasets Employing a Complex Sample Design](#)

Module Summary and Resources

Summary

This module provided information about the research methods and statistical techniques that NCES uses to provide high-quality data by describing:

- NCES study designs
- Sampling designs and weights
- Analysis of NCES data

Resources

- [NCES Statistical Standards Program](#)
- [NCES DLDT](#)
- [Statistical Analysis of NCES Datasets Employing a Complex Sample Design](#)