

Introduction to Analyzing NCES Data Using EdSurvey

Developed by Paul Bailey, Nancy Collins, Ahmad Emad, Thomas Fink, Huade Huo, Julian Gerez, Mich

March 22 2021

Overview of the EdSurvey Package

The **EdSurvey** package is designed to help users analyze data from the National Center for Education Statistics (NCES), including the National Assessment of Educational Progress (NAEP) datasets. Because of their scope and complexity, these datasets require special statistical methods to analyze. The **EdSurvey** package gives users functions to perform analyses that account for both complex sample survey designs and the use of plausible values.

The **EdSurvey** package also seamlessly takes advantage of the **LaF** package to read in data only when it is required for an analysis. Users with computers that have insufficient memory to read in the entire NAEP datasets can still do analyses without having to write special code to read in just the appropriate variables. This is all addressed directly in the **EdSurvey** package—behind the scenes and without additional work by the user.

Technical Papers

Book and Journal Publication

Bailey, P., Lee, M., Nguyen, T., & Zhang, T. (2020). *Using EdSurvey to Analyse PIAAC Data*. In *Large-Scale Cognitive Assessment* (pp. 209-237). Springer, Cham.

Data Set Specific Overviews

Documents that describe the analysis of specific survey data in the EdSurvey package include the following:

- *Using EdSurvey to Analyze ECLS-K:2011 Data* is an introduction to the methods used in the analysis of the large-scale child development study Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011) using the **EdSurvey** package. The vignette covers topics such as preparing the R environment for processing, creating summary tables, running linear regression models, and correlating variables.
- *Using EdSurvey to Analyze NCES Data: An Illustration of Analyzing NAEP Primer* is an introduction to the basics of using the **EdSurvey** package for analyzing NCES data, using the NAEP Primer as an example. The vignette covers topics such as preparing the R environment for processing, creating summary tables, running linear regression models, and correlating variables.
- *Using EdSurvey to Analyze TIMSS Data* is an introduction to the methods used in analysis of large-scale educational assessment programs such as Trends in International Mathematics and Science Study (TIMSS) using the **EdSurvey** package. The vignette covers topics such as preparing the R environment for processing, creating summary tables, running linear regression models, and correlating variables.

- *Using EdSurvey to Analyze NAEP Data With and Without Accommodations* provides an overview of the use of NAEP data with accommodations and describes methods used to analyze this data.

Task Specific Walkthroughs

Documents providing an overview of functions developed in the EdSurvey package include the following:

- *Installing the EdSurvey Package on a Restricted-Use Data Computer* provides guidance for how to install EdSurvey on a restricted-use data (RUD) computer without an Internet connection.
- *Converting Text Data File(s) With Companion SPSS Script to SPSS Data File Format* details the process of converting a data file and SPSS script to an SPSS Data File for use with EdSurvey.
- *Using the `getData` Function in EdSurvey* describes the use of the EdSurvey package when extensive data manipulation is required before analysis.
- *Using EdSurvey for Trend Analysis* describes the methods used in the EdSurvey package to conduct analyses of statistics that change over time in large-scale educational studies.
- *Exploratory Data Analysis on NCES Data* provides examples of conducting exploratory data analysis on NAEP data.
- *Calculating Adjusted p-Values From EdSurvey Results* describes the basics of adjusting p -values to account for multiple comparisons.
- *Producing L^AT_EX Tables From edsurveyTable Results With edsurveyTable2pdf* details the creation of pdf summary tables from summary results using the `edsurveyTable2pdf` function.

Methodology Resources

Documents that describe the statistical methodology used in the EdSurvey package include the following:

- *Statistical Methods Used* details the estimation of the statistics in the `lm.sdf`, `achievementLevel`, and `edsurveyTable` functions.
- *Analyses Using Achievement Levels Based on Plausible Values* describes the methodological approaches for analyses using NAEP achievement levels.
- *Gap Analysis* covers the methods comparing the gap analysis results of the EdSurvey package to the NAEP Data Explorer.
- *Estimating Percentiles* describes the methods used to estimate percentiles.
- *Estimating Mixed-Effects Models* describes the methods used to estimate mixed-effects models with plausible values and survey weights, and how to fit different types of mixed-effects models using the EdSurvey package.
- *Multivariate Regression* details the estimation of multivariate regression models using `mvr1m.sdf`.
- *Running Wald Tests* describes the use of the Wald test to jointly test regression coefficients estimated using `lm.sdf` and `glm.sdf`.
- *Weighted and Unweighted Correlation Methods for Large-Scale Educational Assessment: wCorr Formulas* introduces the methodology used by the wCorr R package for computing the Pearson, Spearman, polyserial, polychoric and tetrachoric correlations, with and without weights applied. Simulation evidence is presented to show correctness of the methods, including an examination of the bias and consistency.

Software Requirements

Unless you already have R version 3.2.0 or later, install the latest R version—which is available online at <https://cran.r-project.org/>. Users also may want to install RStudio desktop, which has an interface that many find easier to follow. RStudio is available online at <https://rstudio.com/products/rstudio/download/>.

Setting Up the Environment for Analyzing NCES Data

Installing and Loading EdSurvey

Inside R, run the following command to install **EdSurvey** as well as its package dependencies:

```
install.packages("EdSurvey")
```

Once the package is successfully installed, **EdSurvey** can be loaded with the following command:

```
library(EdSurvey)
```