

Slide 1 of 11

Title Slide: EdSurvey: Analysis of the Early Childhood Longitudinal Study, Kindergarten Class of 2010-11

Slide 2 of 11

This module reintroduces EdSurvey to users while providing basic information on installing and loading EdSurvey. However, users who have not reviewed the first two EdSurvey modules should do so now. Those modules are entitled Introduction to Using EdSurvey R for Analyzing Large-Scale Data Sets and Analyzing Large-Scale Assessment Data – Using NAEP and TIMSS Data as Examples.

This will show users how to use EdSurvey to analyze the ECLS-K:2011 data file.

Throughout this module, underlined blue screen text indicates a link to additional resources.

Please note: To fully understand this module, you must have an understanding of complex surveys. Additionally, you must have knowledge of the R statistical package.

Slide 3 of 11

For detailed instruction and documentation of the EdSurvey package, including how to report bugs and other issues, please visit the EdSurvey home page.

Next, we will walk through a few key steps in using the EdSurvey R package to analyze ECLS-K:2011 data.

Once inside R, users should run the “install.packages” command to install the EdSurvey R package as well as its package dependencies. When the package is successfully installed, users should run the “library” command to load the package.

Slide 4 of 11

Data can be downloaded via an internet connection by the download ECLS_K function by specifying the years and root location where it should be stored on your machine.

After the download completes, unzip the ChildK5p.zip file in “~/ECLSK/2011” folder location.

The unzipped data can be read in with the readECLS_K2011 function. In this example, we assign the read-in data object with the name “eclsk11”.

Slide 5 of 11

Before conducting data analysis, users have the options of retrieving the rows and columns of the data to do data manipulation, such as subsetting the data based on certain conditions, recoding the levels of a categorical variable, or renaming variables. To extract the data with selected variables, users need to use the “getData” function. In

EdSurvey: Analysis of the Early Childhood Longitudinal Study, Kindergarten Class of 2010-11

the example, `x_chsex_r` (“Sex of Students”), `p5sumsch` (“Child attended summer school”), `p5nhrprm` (“Hours per day child attended summer school”), and the weight variable `w5cf5pf_50` in `eclsk11` are called via `getData`.

A subset of a dataset can be created with EdSurvey package functions. In this example, a subset of data is created after filtering the sample to include only female students and those who are 1: WHITE, NON-HISPANIC or 2: BLACK/AFRICAN AMERICAN, NON-HISPANIC.

Slide 6 of 11

Let’s suppose you were interested in exploring student performance in mathematics based on the number of hours or days a parent reported their child attended summer school (`p5nhrprm`). We’d first need to recode that variable so that students who didn’t attend summer school (where `p5sumsch` coded as a 2: NO) are included in the analytic subset with zero minutes.

The `table` function is a simple method of ascertaining the number of values for each level of a variable in a data set. Using the `table` function for the `p5nhrprm` variable indicates that parents reported their child attending summer school anywhere from 2 to 7 hours per day. The “NA” values contain special values of different sorts including survey skip codes such as those for parents who indicated “NO” on item `p5nhrprm`.

In order to include children who attended summer school for zero hours per day - those that were skipped by the design of the survey - recode `p5nhrprm` values to zero where `p5sumsch == “2: NO”`.

More examples of recoding variables in ECLS-K:2011 can be found in this vignette [linked here](#).

Slide 7 of 11

Users can create a survey weighted summary table using the “`edsurveyTable`” function. In the example shown, “`edsurveyTable`” is called to create a table that shows the number and percentage of students by gender and their parent’s current marital status. The `edsurveytable` function makes use of the survey attributes, such as the sample weights and its replicate weights, to produce the correct standard errors and weighted n counts.

Slide 8 of 11

In this example, the correlation between the variable “`x9mscalk5`” and “`x9povty_i`” is calculated that takes into account the sample weight “`w9c29p_9a0`”, and uses the Pearson method. Other methods supported by the “`cor.sdf`” functions include, Spearman, Polychoric, and Polyserial correlations.

Slide 9 of 11

Users can also conduct regression analyses via EdSurvey regression functions, which accounts for the complex sample design used for the ECLS-K:2011 data. The most

EdSurvey: Analysis of the Early Childhood Longitudinal Study, Kindergarten Class of 2010-11

common regressions supported by EdSurvey include linear regressions and logistic regressions. In the example, linear regression analysis was conducted using the “lm.sdf” function, and the regression results are saved in an R object named “lm1”. The summary function was used to show the contents of the lm1 object.

Slide 10 of 11

This module provided a brief introduction of using EdSurvey to analyze longitudinal data. We used the ECLS-K:2011 data file as the example data.

This module supplemented the first two modules in the EdSurvey module. Those modules are entitled Introduction to EdSurvey and EdSurvey: Analyzing Large-Scale Assessment Data- Using NAEP and TIMSS Data as Examples. Users should review those two modules before reviewing this module. Again, the EdSurvey module series is for users who are familiar with the R statistical software.

Slide 11 of 11

Use the following links for more EdSurvey resources.

You have now completed the EdSurvey: Analysis of the Early Childhood Longitudinal Study, Kindergarten Class of 2010-11. Click the exit button to return to the landing page.