The EdSurvey `suggestWeights` Function

Huade Huo, Paul Bailey, and Cameron McPhee[*][†]

March 13, 2019

Introduction

The `suggestWeights` function in EdSurvey allows users to specify the list of variables they intend to use in an analysis and suggests sample weights that may be appropriate—final weight variable selection is up to the user and should be based on a detailed understanding of the available analytic weights.

Sampling weights designed for use with data from a complex sample survey, such as the NCES longitudinal studies, serve two primary purposes. When used in analyses, the main sampling weight weights the sample size up to the population total of interest. For NCES studies, weighting produces national-level estimates. Also, the main sampling weight adjusts for differential nonresponse patterns that can lead to bias in the estimates. If people with certain characteristics are systematically less likely than others to respond to a survey, the collected data may not accurately reflect the characteristics and experiences of the nonrespondents, which can lead to bias. To adjust for this, respondents are assigned weights that, when applied, result in respondents representing their own characteristics and experiences as well as those of nonrespondents with similar attributes.

A sample weight could be produced for use with data from every component of the study (e.g., student questionnaires, parent surveys, student assessment data, transcript data) at every round of data collection and for every combination of components and rounds. However, creating all possible weights for a study with as many components as the NCES longitudinal studies would be impractical. Therefore, there may not be a weight corresponding exactly to the components included in an analysis.

When no weight corresponds exactly to the combination of components included in the analysis, researchers might prefer to use a weight with more components included. Choosing a weight that requires more responses will result in some of the respondents with complete data for the necessary components being removed while choosing a weight that requires less complete data for the necessary components will result in missing data and potentially nonresponse bias. The former slightly reduces the power, the latter could lead to biased estimates.

Please refer to study-specific documentation for more comprehensive information about the analytic weights available for each study and the analysis for which each is best suited. For more information, please visit https://nces.ed.gov/.

Currently the `suggestWeights` function has been implemented for The Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011), sponsored by the National Center for Education Statistics (NCES) within the Institute of Education Sciences (IES) of the U.S. Department of Education.

Rationale for Using Weights

It is important to weight the data in your analyses using an appropriate sampling weight:

- Weighted analyses provide population estimates.
- Weights adjust for differential selection probabilities (e.g., oversampling).
- Weights reduce bias associated with nonresponse by adjusting for differential nonresponse.

Researchers need to choose a weight that is adjusted for nonresponse at the rounds analyzed and is adjusted for nonresponse to the components providing data for the analysis.

Using the `suggestWeights` Function to Determine the Weight

To use `suggestWeights` function, the first step is to attach `EdSurvey` package and load ECLS-K:2011 data. The following code attach `EdSurvey` package and read-in ECLS-K:2011 data file with parameters specified:

```
library(EdSurvey)
eclsk11 <- readECLS_K2011(file.path("C:/EdSurveyData/", "ECLS_K", "2011"),
                          filename = "childK4p.dat",
                          layoutFilename = "ECLSK2011_K4PUF.sps",
                          verbose = FALSE)
```

Next, provide variable(s) intended to use in the further analysis. The `suggestWeights` can only suggest weight from variables provided in ECLS-K:2011 K-4 Public Use File. Source variables need to be used if any new variables were generated or derived for analysis.

For example, weight `w8cf8p_80` was suggested if the research question focus on child questionnaire data from all eight rounds from kindergarten through fourth grade:

```
suggestWeights(c("x1mscalk4", "x2mscalk4", "x3mscalk4", "x4mscalk4", "x5mscalk4",
    "x6mscalk4", "x7mscalk4", "x8mscalk4"), eclsk11)
```

```
## Weight suggestion is experimental. Please consult the data file documentation
## or data file user's manual for complete descriptions of the analytic weights
## and their adjustments.
```

```
## [1] "w8cf8p_80"
```

To print out additional information on weight suggestion, such as the number of cases with a valid main sampling weight, description, variables(s) used in weight suggestions and variables(s) not used, use `verbose=TRUE` option:

```
suggestWeights(c("x1mscalk4", "x2mscalk4", "x3mscalk4", "x4mscalk4", "x5mscalk4",
    "x6mscalk4", "x7mscalk4", "x8mscalk4"), eclsk11, verbose = TRUE)
```

```
## Weight suggestion is experimental. Please consult the data file documentation
## or data file user's manual for complete descriptions of the analytic weights
## and their adjustments.
## The following weight is suggested:
## Weight: w8cf8p_80
## Number of cases: 2,354
## Description: Child base weight adjusted for nonresponse associated with child
## assessment/child questionnaire data from all eight rounds from kindergarten
## through fourth grade, as well as parent data from fall kindergarten or spring
## kindergarten, and parent data from all rounds from fall first grade through
## spring fourth grade. (C1C2C3C4C5C6C7C8)(P1_P2)(P3P4P5P6P7P8)
##
##
## Weight selection based on the inclusion of the following variables:
##    varnames                                  source
## 1 x1mscalk4    Fall 2010 composite/derived variables
## 2 x2mscalk4  Spring 2011 composite/derived variables
## 3 x3mscalk4    Fall 2011 composite/derived variables
## 4 x4mscalk4  Spring 2012 composite/derived variables
```

```
## 5 x5mscalk4    Fall 2012 composite/derived variables
## 6 x6mscalk4  Spring 2013 composite/derived variables
## 7 x7mscalk4  Spring 2014 composite/derived variables
## 8 x8mscalk4  Spring 2015 composite/derived variables
```

```
## [1] "w8cf8p_80"
```

To display all applicable weights that covers more components, which typically are more conservative with smaller sample size, use `showAllWeightSuggestions=TRUE` option:

```
suggestWeights(c("x1mscalk4", "x2mscalk4", "x3mscalk4", "x4mscalk4", "x5mscalk4",
    "x6mscalk4", "x7mscalk4", "x8mscalk4"), eclsk11, showAllWeightSuggestions = TRUE,
    verbose = TRUE)
```

```
## Weight suggestion is experimental. Please consult the data file documentation
## or data file user's manual for complete descriptions of the analytic weights
## and their adjustments.
## The following weights are suggested:
## Weight: w8cf8p_80
## Number of cases: 2,354
## Description: Child base weight adjusted for nonresponse associated with child
## assessment/child questionnaire data from all eight rounds from kindergarten
## through fourth grade, as well as parent data from fall kindergarten or spring
## kindergarten, and parent data from all rounds from fall first grade through
## spring fourth grade. (C1C2C3C4C5C6C7C8)(P1_P2)(P3P4P5P6P7P8)
##
##
## Weight: w8cf8p_2t180
## Number of cases: 2,715
## Description: Child base weight adjusted for nonresponse associated with child
## assessment/child questionnaire data from all eight rounds from kindergarten
## through fourth grade, as well as parent data from fall kindergarten or spring
## kindergarten, and teacher data from fall and spring kindergarten, fall and
## spring first grade, fall and spring second grade, spring third grade, and spring
## fourth grade (C1C2C3C4C5C6C7C8)(P1_P2)(T1T2T3T4T5T6T7T8) Note: This weight was
## created with nonresponse adjustments for the reading teacher only. There is no
## similar weight with nonresponse adjustments for the mathematics or science
## teacher.
##
##
## Weight selection based on the inclusion of the following variables:
##    varnames                                source
## 1 x1mscalk4    Fall 2010 composite/derived variables
## 2 x2mscalk4  Spring 2011 composite/derived variables
## 3 x3mscalk4    Fall 2011 composite/derived variables
## 4 x4mscalk4  Spring 2012 composite/derived variables
## 5 x5mscalk4    Fall 2012 composite/derived variables
## 6 x6mscalk4  Spring 2013 composite/derived variables
## 7 x7mscalk4  Spring 2014 composite/derived variables
## 8 x8mscalk4  Spring 2015 composite/derived variables
```

```
## [1] "w8cf8p_80"    "w8cf8p_2t180"
```

The `suggestWeights` function will only return weights if there is any. If the function "cannot determine weights for variables provided" or returns weights that are too conservative, try to remove the variable(s) from the variable list. This is especially helpful in the exploratory data analysis stage. For example, after removing `x8mscalk4` (X8 Math IRT Scale Score-K4), weight `w7cf7p_70` was suggested instead of `w8cf8p_80`,

increasing the sample size from 2,354 to 2,538:

```r
suggestWeights(c("x1mscalk4", "x2mscalk4", "x3mscalk4", "x4mscalk4", "x5mscalk4",
    "x6mscalk4", "x7mscalk4"), eclsk11, verbose = TRUE)
```

```
## Weight suggestion is experimental. Please consult the data file documentation
## or data file user's manual for complete descriptions of the analytic weights
## and their adjustments.
## The following weight is suggested:
## Weight: w7cf7p_70
## Number of cases: 2,538
## Description: Child base weight adjusted for nonresponse associated with child
## assessment/child questionnaire data from all seven rounds from kindergarten
## through third grade, as well as parent data from all seven rounds from
## kindergarten through third grade (C1C2C3C4C5C6C7)(P1_P2)(P3P4P5P6P7)
##
##
## Weight selection based on the inclusion of the following variables:
##    varnames                              source
## 1 x1mscalk4    Fall 2010 composite/derived variables
## 2 x2mscalk4  Spring 2011 composite/derived variables
## 3 x3mscalk4    Fall 2011 composite/derived variables
## 4 x4mscalk4  Spring 2012 composite/derived variables
## 5 x5mscalk4    Fall 2012 composite/derived variables
## 6 x6mscalk4  Spring 2013 composite/derived variables
## 7 x7mscalk4  Spring 2014 composite/derived variables

## [1] "w7cf7p_70"
```

General Considerations When Selecting Weights

Within a research study, we recommend using the same full sample weight for all analyses so that the underlying sample used to produce the estimates is the same across analyses. Therefore, the analytic weight should be selected based on all variables included in all key research questions. The set of individual with positive weights then becomes the analytic sample for all analyses within the study.

Inside `suggestWeights`

The `suggestWeights` function first to determine the source and round of incoming variables. This is determined by the prefixes in the variable names. For example, variable `p8games`, the frequency of playing games in Spring 2015, starts with "p8" which stands for "parent" component and round "8" (spring 4th grade). The ECLS-K:2011 Data File User Manuel provide an exhaustive list of prefixes. See Exhibit 7-1. Prefixes for fourth-grade variables in NCES 2018-032 as an example.

For most individual pairs of components and rounds, there will be one or more weights that adjust for nonresponse to those elements. This is summarized in the Data File User Manuals . Weights like "w8c18p_80, w8c28p_8a0, w8c28p_8b0, w8cf8p_80. . . " are all applicable to variables starts with "p8".

Depending on the number of incoming variables, multiple lists of possible weights will be generated. An intersection of this list will provide the common set of weights that apply to all incoming variables. In case there is no intersection, an error message "no weight suggestions for the following variables: . . . " will be raised. Consider consulting the data file documentation or data file user's manual for complete descriptions of the analytic weights and their adjustments.

Once this list is obtained, the list is ranked by the number of components included in the analysis for which a weight adjusts. The fewer components for which a weight can adjust, the higher the chance a specific weight will be suggested. For example, `W8C8P_2` adjusts for 3 components (i.e., P1, P2, and C8) and `W8C18P_2` adjusts 5 components (i.e., C1, P1, C2, P2, C8). If only "C8" variables were used in the analysis, `W8C8P_2` would be prioritized since it adjusts for the component included in the analysis, but also adjusts for fewer components not included in the analysis.

Limitations

First, `suggestWeights` function does not distinguish the "or" condition among components. This affects weight suggestion when `W12P0` is returned. Researchers need to examine if child-level weight `W1_2P0` is applicable when a child had to either have a fall kindergarten parent interview **or** a spring kindergarten parent interview.

Second, the classification of round-specific derived variables (i.e., variables with prefix X#) is based on their variable description. For example, the component of all round-specific derived variables with "teacher report" in their description were classified as "teacher" and variable descriptions that contains "parent report" was classified as "parent." The `suggestWeights` function may not yield appropriate weight if a variable is derived from multiple components, such as an ECLS-K2010-11 variable that could be taken from parent or child components.

References

Tourangeau, K., Nord, C., Lê, T., Sorongon, A.G., Hagedorn, M.C., Daly, P., and Najarian, M. (2015). *Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011), User's Manual for the ECLS-K:2011 Kindergarten Data File and Electronic Codebook, Public Version* (NCES 2015-074). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Tourangeau, K., Nord, C., Lê, T., Wallner-Allen, K., Hagedorn, M.C., Leggitt, J., and Najarian, M. (2015). *Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011), User's Manual for the ECLS-K:2011 Kindergarten-First Grade Data File and Electronic Codebook, Public Version* (NCES 2015-078). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Tourangeau, K., Nord, C., Lê, T., Wallner-Allen, K., Vaden-Kiernan, N., Blaker, L. and Najarian, M. (2017). *Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011) User's Manual for the ECLS-K:2011 Kindergarten-Second Grade Data File and Electronic Codebook, Public Version* (NCES 2017-285). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Tourangeau, K., Nord, C., Lê, T., Wallner-Allen, K., Vaden-Kiernan, N., Blaker, L. and Najarian, M. (2018). *Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS -K:2011) User's Manual for the ECLS-K:2011 Kindergarten-Third G rade Data File and Electronic Codebook, Public Version* (NCES 2018-034). U.S. Department of Education. Washington, DC: National Center for Education S tatistics

Tourangeau, K., Nord, C., Lê, T., Wallner-Allen, K., Vaden-Kiernan, N., Blaker, L. and Najarian, M. (2018). *Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011) User's Manual for the ECLS-K:2011 Kindergarten-Fourth Grade Data File and Electronic Codebook, Public Version* (NCES 2018-032). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

APPENDIX: ECLS-K Weight Notes

The ECLS-K:2011 is a longitudinal data set that is weighted to compensate for unequal probabilities of selection at each sampling stage and to adjust for the effects of nonresponse to various components on

generated estimates, including parent, , teacher, before- and after-school care provider, school administrator, and student surveys and direct child assessments.[1] Non-response can occur on any combination of these components across rounds of data collection (of which there are nine). Because users may select components from various rounds and information sources, many possible weights exist. Not all possible patterns of component response are covered by the weights, so the process is not simple. The `suggestWeights` function in the `EdSurvey` R package can serve as the starting point of determining appropriate weight.

The rounds of data in the analysis and components providing the data in the analysis can be determined by reviewing the weight name and label as well as the description of the weights provided in the Data File User's Manual for each round:

- Weight names and labels indicate the round(s) of data included in the construction of the weight. For example, round 1 stands for Fall Kindergarten, round 2 stands for Spring Kindergarten, and round 9 stands for Spring Fifth Grade.
- Weight names and labels indicate the component(s) of data for which the weight is adjusted for nonresponse. For example, C stands for Child Assessment, P stands for Parent Interview, T stands for Teacher Questionnaires.

---

[1]Mulligan, G. M., McCaroll, J. C., Flanagan, K. D., & Potter, D. (2018). *Findings from the Fourth-Grade Round of the Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K: 2011).* NCES 2018-094. National Center for Education Statistics.