The R EdSurvey Package

Using R and Generative AI for NAEP Data Analysis: Part Two

Presenters: Sinan Yavuz, Ting Zhang, Paul Bailey, Blue Webb and Emmanuel Sikali

April 2024

NAEP
NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Goal

Provide participants with an overview of the methods used to analyze large-scale assessment data using the R package `EdSurvey` and generative AI chatbot-EdSurveyGPT

Follow along in edsurvey_training_part2.R

Outline of EdSurvey Session: Part Two

- Summary Statistics: `summary2` and `edsurveyTable`
- Gap Analysis
- Linear Regression
- Logistic Regression

Loading NAEP

Read NAEP Primer data:

```
library(EdSurvey)
```

```
## Loading required package: car
## Loading required package: carData
## Loading required package: lfactors
## lfactors v1.0.4
## Loading required package: Dire
## Dire v2.2.0
## EdSurvey v4.0.4
##
## Attaching package: 'EdSurvey'
## The following objects are masked from 'package:base':
##
##     cbind, rbind
```

```
sdf <- readNAEP(system.file("extdata/data", "M36NT2PM.dat",
                            package="NAEPprimer"))
```

Summary Statistics

NAEP
NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Summary Statistics

`summary2()` produces both weighted and unweighted descriptive statistics for a variable. `summary2()` takes the four following arguments in order:

- `data`: an `EdSurvey` object
- `variable`: name of the variable you want to produce statistics on
- `weightVar`: name of the weight variable, or `NULL` if users want to produce unweighted statistics
- `dropOmittedLevels`: if `TRUE`, the function will remove omitted levels for the specified variable before producing descriptive statistics. If `FALSE`, the function will include omitted levels in the output statistics.

Summary Statistics

For a continuous variable (i.e., composite Math score):

```
summary2(data = sdf, variable = "composite")
```

```
## Estimates are weighted using the weight variable 'origwt'
##    Variable     N Weighted N   Min.  1st Qu.   Median     Mean  3rd Qu.    Max.      SD NA's Zero weights
## 1 composite 16915   16932.46 126.11 251.9626 277.4784 275.8892 301.1827 404.184 36.5713    0            0
```

- For NAEP data and other datasets that have a default weight variable, `summary2` produces weighted statistics by default.
- If a specified `variable` is a scale name (e.g., `composite`) and the weight option is selected, `summary2` statistics account for both plausible value pooling and weighting.

Summary Statistics

For a continuous variable (i.e., composite Math score):

- By specifying `weightVar = NULL`, the function prints out unweighted descriptive statistics for the `variable`, or each plausible value of the specified scale (e.g., `composite`).

```
summary2(sdf, "composite", weightVar = NULL)
```

```
## Estimates are not weighted.
##   Variable     N   Min.  1st Qu. Median     Mean  3rd Qu.   Max.       SD NA's
## 1   mrpcm1 16915 130.53 252.0600 277.33 275.8606 300.7200 410.80 35.89864    0
## 2   mrpcm2 16915 124.16 252.2100 277.33 275.6399 300.6900 408.58 36.08483    0
## 3   mrpcm3 16915 115.09 252.0017 277.19 275.6570 300.5600 398.17 36.09278    0
## 4   mrpcm4 16915 137.19 252.4717 277.44 275.7451 300.5767 407.41 35.91078    0
## 5   mrpcm5 16915 123.58 252.4900 277.16 275.6965 300.5000 395.96 36.10905    0
```

Summary Statistics

For a categorical variable (i.e., frequency of students talking about studies at home):

```
summary2(sdf, "b017451")
```

```
## Estimates are weighted using the weight variable 'origwt'
##                b017451    N Weighted N Weighted Percent Weighted Percent SE
## 1 Never or hardly ever 3837  3952.4529      23.34245648           0.4318975
## 2 Once every few weeks 3147  3190.8945      18.84483329           0.3740648
## 3    About once a week 2853  2937.7148      17.34960077           0.3414566
## 4  2 or 3 times a week 3362  3425.8950      20.23270282           0.3156289
## 5            Every day 3132  3223.8074      19.03921080           0.4442216
## 6              Omitted  575   194.3312       1.14768416           0.1272462
## 7             Multiple    9     7.3676       0.04351168           0.0191187
```

- By default, `dropOmittedLevels` is set to `FALSE`. That is, the function includes omitted levels of the variable `b017451` in the output statistics.

Summary Statistics

For a categorical variable (i.e., frequency of students talking about studies at home):

- By specifying `dropOmittedLevels = TRUE`, the function removes omitted levels from the output statistics.

```
summary2(sdf, "b017451", dropOmittedLevels = TRUE)
```

```
## Estimates are weighted using the weight variable 'origwt'
##                b017451    N Weighted N Weighted Percent Weighted Percent SE
## 1 Never or hardly ever 3837   3952.453         23.62386           0.4367548
## 2 Once every few weeks 3147   3190.894         19.07202           0.3749868
## 3    About once a week 2853   2937.715         17.55876           0.3486008
## 4  2 or 3 times a week 3362   3425.895         20.47662           0.3196719
## 5            Every day 3132   3223.807         19.26874           0.4467063
```

Cross Tabulation - Let's ask EdSurveyGPT

- Which function should I use to get crosstab in EdSurvey?
- Feel free to ask follow up questions; examples, arguments, etc.

Cross Tabulation

`edsurveyTable()`: creates a summary table of outcome and categorical variables. There are three important arguments:

- `formula`: typically written as `a ~ b + c`, in which:
  - `a`: a continuous variable (optional) that the function will return the weighted mean for
  - `b` and `c`: categorical variable(s) that the function will run cross-tabulation on; multiple cross-tab categorical variables can be separated using the `+` symbol
- `data`: an `EdSurvey` object
- `pctAggregationLevel`: a numeric value (i.e., 0, 1, 2) that indicates the level of aggregation in the cross-tabulation result's percentage column.

Cross Tabulation

- Summary table of NAEP composite mathematics performance scale scores (`composite`) of eighth-grade students by two student factors:
  - `dsex`: gender
  - `b017451`: frequency of talk about studies at home
- `pctAggregationLevel` is set by default to `NULL` (or `1`). That is, the `PCT` column adds up to 100 within each level of the first categorical variable `dsex`.

```
es1 <- edsurveyTable(composite ~ dsex + b017451, data = sdf)
```

dsex	b017451	N	WTD_N	PCT	SE(PCT)	MEAN	SE(MEAN)
Male	Never or hardly ever	2350	2434.844	29.00978	0.6959418	270.8243	1.057078
Male	Once every few weeks	1603	1638.745	19.52472	0.5020657	275.0807	1.305922
Male	About once a week	1384	1423.312	16.95795	0.5057265	281.5612	1.409587
Male	2 or 3 times a week	1535	1563.393	18.62694	0.4811497	284.9066	1.546072

Cross Tabulation

- By specifying `pctAggregationLevel = 0`, the `PCT` column adds up to 100 across the entire sample.

```
es2 <- edsurveyTable(composite ~ dsex + b017451, data = sdf, pctAggr
```

dsex	b017451	N	WTD_N	PCT	SE(PCT)	MEAN	SE(MEAN)
Male	Never or hardly ever	2350	2434.844	14.553095	0.3738531	270.8243	1.057078
Male	Once every few weeks	1603	1638.745	9.794803	0.2651368	275.0807	1.305922
Male	About once a week	1384	1423.312	8.507154	0.2770233	281.5612	1.409587
Male	2 or 3 times a week	1535	1563.393	9.344421	0.2670298	284.9066	1.546072
Male	Every day	1291	1332.890	7.966700	0.3000579	277.2597	1.795784
Female	Never or hardly ever	1487	1517.609	9.070768	0.2984443	266.7897	1.519020
Female	Once every few weeks	1544	1552.149	9.277216	0.2498498	271.2255	1.205528
Female	About once a week	1469	1514.403	9.051606	0.2899668	278.7502	1.719778

- Related documentation: EdSurvey-LaTeXtables.pdf

Think, Pair, Share: edsurveyTable

Could you identify the number of plausible values, the replicate weight settings, and the variance method in the `edsurveyTable` output?

Why is the n used smaller than the full-data n?

```
> edsurveyTable(composite ~ dsex + b017451, data = sdf, pctAggregationLevel = 0)

Formula: composite ~ dsex + b017451

Plausible values: 5
jrrIMax: 1
Weight variable: 'origwt'
Variance method: jackknife
JK replicates: 62
full data n: 17606
n used: 16331
```

Think, Pair, Share: edsurveyTable

How would you use `edsurveyTable` to create a summary table with these parameters:

- overall math performance across subscales (`composite`)
- a variable that has to do with IEP status
- a variable that has to do with the number of books at home

Feel free use EdSurveyGPT:

- I want to create a crosstab using math composite score with IEP status and number of books at home on NAEP primer data set. Provide the necessary codes.

Think, Pair, Share: edsurveyTable

Scenario result:

```
edexercise <- edsurveyTable(composite ~ iep + b013801,
                            weightVar = 'origwt', data = sdf)
edexercise
```

```
##
## Formula: composite ~ iep + b013801
##
## Plausible values: 5
## jrrIMax: 1
## Weight variable: 'origwt'
## Variance method: jackknife
## JK replicates: 62
## full data n: 17606
## n used: 16351
##
##
## Summary Table:
##  iep b013801    N     WTD_N      PCT   SE(PCT)     MEAN  SE(MEAN)
##  Yes    0-10  304  297.1972 17.33406 1.0388812 226.1623 2.3075125
##  Yes   11-25  430  429.6252 25.05794 1.4034976 231.8103 2.3796081
##  Yes  26-100  517  530.9539 30.96795 1.5297784 249.2306 2.4682667
##  Yes    >100  457  456.7507 26.64004 1.6556494 257.6787 2.8205193
##   No    0-10 1720 1890.3037 12.56502 0.4765198 257.6975 1.2861579
##   No   11-25 2936 3170.9954 21.07789 0.5632689 266.0401 0.9908671
##   No  26-100 5330 5350.4978 35.56524 0.6242526 281.5820 0.8305656
##   No    >100 4657 4632.3807 30.79185 0.8511616 296.2606 1.0533164
```

Achievement Level Analysis and Percentile Analyses

- EdSurvey presents many other features, two of them are Achievement Level Analysis and Percentile Analyses. See the related documentation for more information and example codes.
- Analyses Using Achievement Levels Based on Plausible Values
- EdSurvey-Percentiles.pdf

Gap Analysis: Estimating the Difference between Two Groups

NAEP
NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Gap Analysis

`gap()`: estimate the difference in a statistic for two groups in the population. A gap occurs when one group outperforms another group and the difference in the statistics are statistically significant.

- Statistics can be:
  - mean scores
  - student group percentages
  - achievement level percentages
  - percentiles
- Variance estimation:

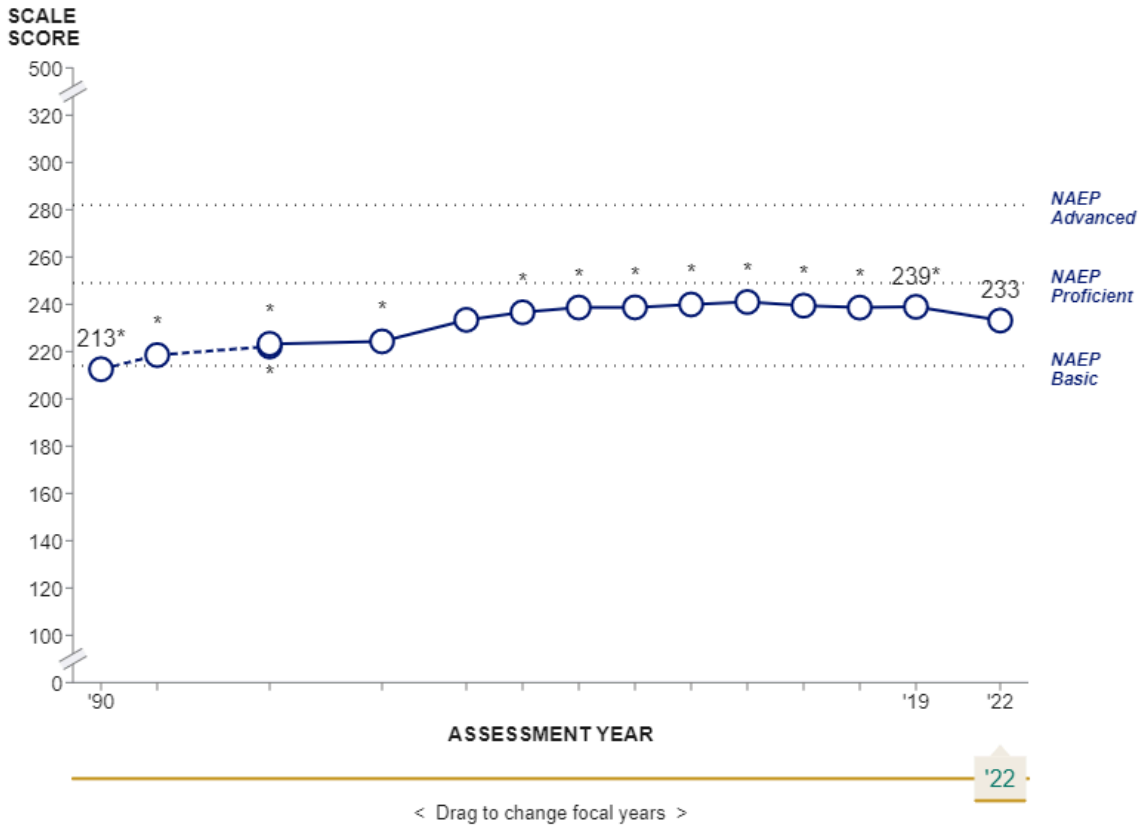
$$Var(\theta_A - \theta_B) = Var(\theta_A) + Var(\theta_B) - 2Cov(\theta_A, \theta_B)$$

- Related methodology documentation: EdSurvey-Gap.pdf

Typical Gap Comparisons

- Comparisons of the same group/jurisdiction between years

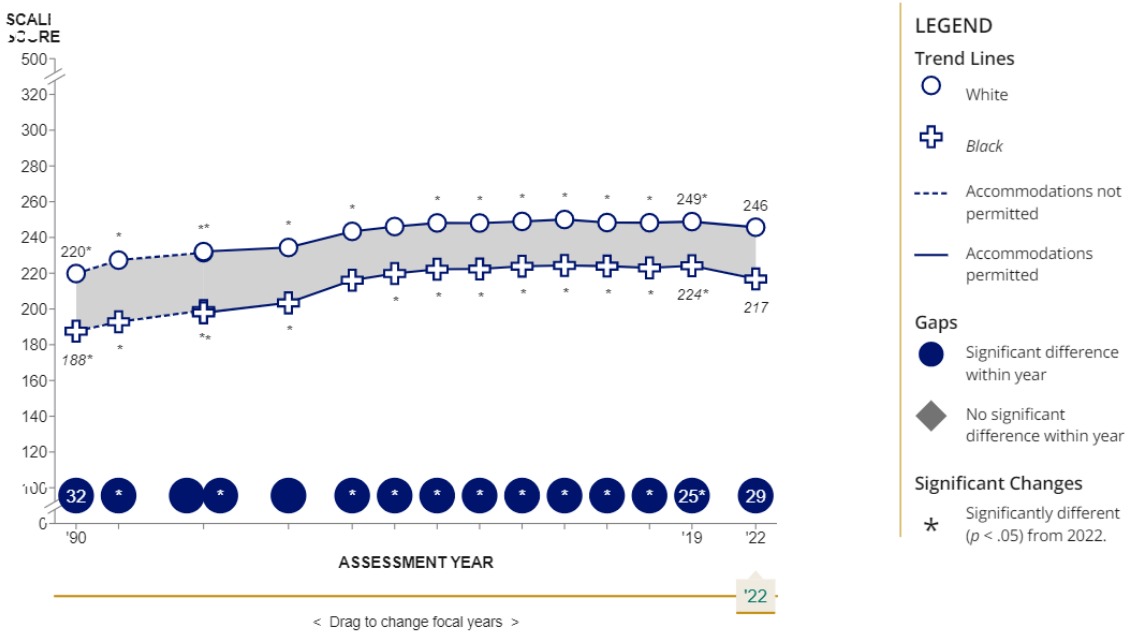
FIGURE | Trend in fourth-grade NAEP mathematics average scores for female students



Typical Gap Comparisons (cont.)

- Comparisons of the gap of different groups/jurisdictions between years
  - Among racial/ethnic groups, the average mathematics score for White fourth-grade students was 29 points higher than their Black peers (compared to 25 points in 2019).

FIGURE | Trend in fourth-grade NAEP mathematics average scores and score gaps, by race/ethnicity



Within-Year Comparisons

Between students groups

```
mathGap <- gap(variable = "composite", data = sdf,
               groupA = dsex %in% "Male",
               groupB = dsex %in% "Female")
```

- `groupA`: defines a condition to subset data
  - `dsex %in% "Male"`
- `groupB`: defines a condition to subset data to compare to `groupA`
  - `dsex %in% "Female"`

Within-Year Comparisons (cont.)

- Mean score results returned with `mathGap$results`

```
mathGap$results
```

```
##   estimateA estimateAse estimateB estimateBse   diffAB     covAB  diffABse diffABpValue    dofAB
## 1  276.7235   0.8207151  275.0458   0.9402535 1.677756 0.5676583 0.6498719   0.01259479 53.70969
```

- `estimateA`/`estimateB`: value of estimate
- `estimateAse`/`estimateBse`: standard error of estimates
- `diffAB/diffABse`: difference between estimateA and estimateB and the standard error of the difference in group estimates
- `covAB`: the covariance used in calculating diffABse
- `diffABpValue`: the p-value associated with the t-test used for the hypothesis test that diffAB is zero
- `dofAB`: the degrees of freedom used in calculating diffABpValue

Gap Analysis: Summary

Analyses:

- mean scores
- student group percentages
- achievement level percentages
- percentiles

Comparison types:

- within year
- between years (uses `edsurvey.data.frame.list`)

Learn more from the EdSurvey User Guide

Think, Pair, Share: Gap Analysis

A. Select Your Own Variables: use EdSurvey functions to select variables of interest to calculate `gap` results

```
# for reference
help(package = "EdSurvey")
searchSDF("text", sdf)
levelsSDF("myvar", sdf)
```

B. Replicate This Scenario: use EdSurvey functions to calculate `gap` results with these parameters:

- composite math performance at the proficient level or higher
- within-year comparisons of English language learners

Think, Pair, Share: Gap Analysis

Scenario result:

```
exerciseGap <- gap(variable = "composite", data = sdf,
                   groupA = ell3 %in% "Yes", groupB = ell3 %in% "No"
                   achievementLevel = c("Proficient"))
exerciseGap$results
```

```
##         achievementLevel estimateA estimateAse estimateB estimateBse    diffAB      covAB diffABse diffABpValue    dofAB
## 1 At or Above Proficient  4.853979    1.372415  28.91917   0.8365519 -24.06519 -0.0565243 1.642069 1.014523e-07 9.263124
```

Linear Regression

NAEP
NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Linear Regression: lm.sdf()

`lm.sdf()`: fits a linear model formula using sampling weights and a variance estimation method.

The format is:

```
myfit <- lm.sdf(formula, data, weightVar, varMethod, relevels)
```

- `formula`: model to be fit.
- `data`: an `EdSurvey` object containing the data to be used in fitting the model
- `weightVar`: indicates the weight variable to use.
- `varMethod`: the variance estimation method (Jackknife or Taylor series) with Jackknife as the default
- `relevels`: used when the user wants to change the reference level of a categorical variable

Linear Regression: lm.sdf()

The resulting object (`myfit` in this case) is a list containing extensive information about the fitted model.

Formula notation is typically written as:

```
Y ~ X1 + X2 + ... + Xk
```

- The `~` separates the response variable on the left from the predictor variables on the right.
- The `+` sign separates the predictor variables.

Linear Regression: lm.sdf()

Example of bivariate regression:

```
lm1 <- lm.sdf(composite ~ b017451,
              weightVar = 'origwt', data = sdf)
summary(lm1)
```

```
##
## Formula: composite ~ b017451
##
## Weight variable: 'origwt'
## Variance method: jackknife
## JK replicates: 62
## Plausible values: 5
## jrrIMax: 1
## full data n: 17606
## n used: 16331
##
## Coefficients:
##                                coef        se        t    dof  Pr(>|t|)
```

Linear Regression: lm.sdf()

Example of multiple regression:

```
lm2 <- lm.sdf(composite ~ dsex + b017451,
              weightVar = 'origwt', data = sdf)
```

- The sampling weight for this regression: `origwt`
- How do these estimates account for the short test form?
- How do these estimates account for correlation within schools?

Linear Regression: lm.sdf()

```
summary(lm2)
```

```
##
## Formula: composite ~ dsex + b017451
##
## Weight variable: 'origwt'
## Variance method: jackknife
## JK replicates: 62
## Plausible values: 5
## jrrIMax: 1
## full data n: 17606
## n used: 16331
##
## Coefficients:
##                                 coef        se        t    dof  Pr(>|t|)
## (Intercept)                270.41112   1.02443 263.9615 54.670 < 2.2e-16 ***
## dsexFemale                  -2.95858   0.60423  -4.8965 54.991 8.947e-06 ***
## b017451Once every few weeks  4.23341   1.18327   3.5777 57.316 0.0007131 ***
## b017451About once a week    11.22612   1.25854   8.9200 54.683 2.983e-12 ***
## b0174512 or 3 times a week  14.94591   1.18665  12.5951 72.582 < 2.2e-16 ***
## b017451Every day             7.52998   1.30846   5.7549 48.470 5.755e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Multiple R-squared: 0.0224
```

Linear Regression: lm.sdf()

Adding `src = TRUE` displays standardized regression coefficients.

```
summary(lm2, src = TRUE)
```

```
##
## Formula: composite ~ dsex + b017451
##
## Weight variable: 'origwt'
## Variance method: jackknife
## JK replicates: 62
## Plausible values: 5
## jrrIMax: 1
## full data n: 17606
## n used: 16331
##
## Coefficients:
##                                 coef          se        t    dof   Pr(>|t|) stdCoef    stdSE
## (Intercept)                270.4111210   1.0244340 263.9615 54.670 0.0000e+00      NA       NA
## dsexFemale                  -2.9585783   0.6042285  -4.8965 54.991 8.9474e-06 -0.0407 0.008313 **
## b017451Once every few weeks  4.2334144   1.1832671   3.5777 57.316 7.1311e-04  0.0458 0.012791 *
## b017451About once a week    11.2261232   1.2585369   8.9200 54.683 2.9834e-12  0.1175 0.013175 *
## b0174512 or 3 times a week  14.9459085   1.1866461  12.5951 72.582 0.0000e+00  0.1659 0.013175 *
## b017451Every day             7.5299837   1.3084558   5.7549 48.470 5.7550e-07  0.0817 0.014200 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Multiple R-squared: 0.0224
```

Linear Regression: lm.sdf()

Use `relevels` to set omitted/reference level of `dsex` to "Female":

```
lm3 <- lm.sdf(composite ~ dsex + b017451,
              weightVar = 'origwt',
              relevels = list(dsex = "Female"), data = sdf)
```

- What does relevels argument do in lm.sdf? Provide example.

Linear Regression: lm.sdf()

```
summary(lm3)
```

```
##
## Formula: composite ~ dsex + b017451
##
## Weight variable: 'origwt'
## Variance method: jackknife
## JK replicates: 62
## Plausible values: 5
## jrrIMax: 1
## full data n: 17606
## n used: 16331
##
## Coefficients:
##                                 coef        se        t    dof  Pr(>|t|)
## (Intercept)                 267.45254   1.13187 236.2919 76.454 < 2.2e-16 ***
## dsexMale                      2.95858   0.60423   4.8965 54.991 8.947e-06 ***
## b017451Once every few weeks   4.23341   1.18327   3.5777 57.316 0.0007131 ***
## b017451About once a week     11.22612   1.25854   8.9200 54.683 2.983e-12 ***
## b0174512 or 3 times a week   14.94591   1.18665  12.5951 72.582 < 2.2e-16 ***
## b017451Every day              7.52998   1.30846   5.7549 48.470 5.755e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Multiple R-squared: 0.0224
```

Think, Pair, Share: lm.sdf

How would you use `lm.sdf` to perform a regression with multiple predictors using these parameters:

- overall math performance across subscales (`composite`)
- variable that has to do with computers at home
- variable that has to do with language other than English spoken in home

Feel free to use EdSurveyGPT.

Think, Pair, Share: lm.sdf

Scenario result:

```
lmexercise2 <- lm.sdf(composite ~ b017101 + b018201,
                      weightVar = 'origwt', data = sdf)
summary(lmexercise2)
```

```
##
## Formula: composite ~ b017101 + b018201
##
## Weight variable: 'origwt'
## Variance method: jackknife
## JK replicates: 62
## Plausible values: 5
## jrrIMax: 1
## full data n: 17606
## n used: 15884
##
## Coefficients:
##                               coef        se         t    dof  Pr(>|t|)
## (Intercept)                281.95112   0.80871 348.64281 43.827 < 2.2e-16 ***
## b017101No                  -22.44306   1.36521 -16.43932 42.935 < 2.2e-16 ***
## b018201Once in a while       0.63672   0.90717   0.70188 61.423    0.4854
## b018201Half the time        -7.32985   1.58448  -4.62604 50.514 2.624e-05 ***
## b018201All or most of time -12.61417   1.27458  -9.89675 29.860 6.121e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Multiple R-squared: 0.0658
```

Ask EdSurveyGPT

- Can I run multilevel analyses with lm.sdf in edsurvey?
- Does EdSurvey account for the ICC between students in the cluster sampled schools?

Logistic Regression

NAEP
NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Logistic Regression

`logit.sdf()` and `probit.sdf()`: predict binary outcomes from a set of continuous predictor variables (sampling weights and variance estimates)

- `I()` is used to specify the outcome level of the `b013801` variable (books in home)
- `I()` treats certain characters or operators as mathematical operations rather than as part of the formula syntax, we can also use to create a dichotomous variable

```
logit1 <- logit.sdf(I(b013801 %in% ">100") ~ dsex,
                    weightVar = 'origwt', data = sdf)
```

Logistic Regression

```
summary(logit1)
```

```
##
## Formula: b013801 ~ dsex
## Family: binomial (logit)
##
## Weight variable: 'origwt'
## Variance method: jackknife
## JK replicates: 62
## full data n: 17606
## n used: 16359
##
## Coefficients:
##                   coef         se          t    dof  Pr(>|t|)
## (Intercept)  -0.920421   0.046355 -19.855835 60.636 < 2.2e-16 ***
## dsexFemale    0.178274   0.050129   3.556331 54.578 0.0007863 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The *log* odds of having more than 100 books in the home (versus less than or equal to 100 books) increases by 0.178274 for female students compared with male students.

Logistic Regression

```
oddsRatio(logit1)
```

```
##                    OR      2.5%     97.5%
## (Intercept) 0.3983511 0.3630823 0.4370459
## dsexFemale  1.1951531 1.0809029 1.3214796
```

Alternatively, the odds of having more than 100 books in the home (versus less than or equal to 100 books) increases by 1.1951531 for female students, compared with male students.

Bonus point: The Wald test is available for `logit.sdf` and `lm.sdf` models. See details in `?waldTest`.

- Related documentation: EdSurvey-Wald Test.pdf

Think, Pair, Share: Logistic Regression

Use EdSurvey functions to perform a logistic regression and `logit.sdf` using these parameters:

- use an outcome variable that has to do with English Langauge Learners (ELL)
  - code it to 1 when `%in% "Yes"`
- use a predictor variable that has to do with language spoken at home

Think, Pair, Share: Logistic Regression

Scenario result:

```
logitexercise1 <- logit.sdf(I(lep %in% "Yes") ~ b018201,
                            weightVar = 'origwt', data = sdf)
summary(logitexercise1)
```

```
##
## Formula: lep ~ b018201
## Family: binomial (logit)
##
## Weight variable: 'origwt'
## Variance method: jackknife
## JK replicates: 62
## full data n: 17606
## n used: 16159
##
## Coefficients:
##                               coef        se         t     dof  Pr(>|t|)
## (Intercept)                -4.78197   0.19709 -24.26310  9.8977 3.796e-10 ***
## b018201Once in a while      1.94536   0.20702   9.39713 26.0267 7.539e-10 ***
## b018201Half the time        3.13919   0.15354  20.44574 38.9522 < 2.2e-16 ***
## b018201All or most of time  3.63098   0.17657  20.56340 26.7407 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Let's Try EdSurveyGPT

- What are the differences between logit.sdf, glm.sdf and probit.sdf?
- Can EdSurvey run Wald test? What is Wald test and what is it useful for?

Wrap Up

NAEP
NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Learning EdSurvey

- Reading vignettes provided in training materials

```
vignette("introduction", package="EdSurvey")

# There are additional functions that we couldn't cover!
gap() #gap analysis
cor.sdf() # Bivariate correlations using "Pearson", "Spearman", "pol
edsurveyTable2pdf() # creating production ready summary tables
cbind(), rbind(), append(), merge() # useful functions in processing
```

- R help

```
help(package = "EdSurvey")
```

- EdSurvey Website
- EdSurvey Github
- NAEP Data Training Workshop

Under Development

- Package is still under development
  - Subsequent releases of the EdSurvey package will provide additional functionality
- Your feedback is important to us!

Contact Information

EdSurvey Package Help

- https://github.com/American-Institutes-for-Research/EdSurvey

EdSurvey Package Help on NCES.ed.gov

- http://nces.ed.gov/nationsreportcard/contactus.aspx

Ting Zhang

- tzhang@air.org

Paul Bailey

- pbailey@air.org

Emmanuel Sikali

- Emmanuel.Sikali@ed.gov