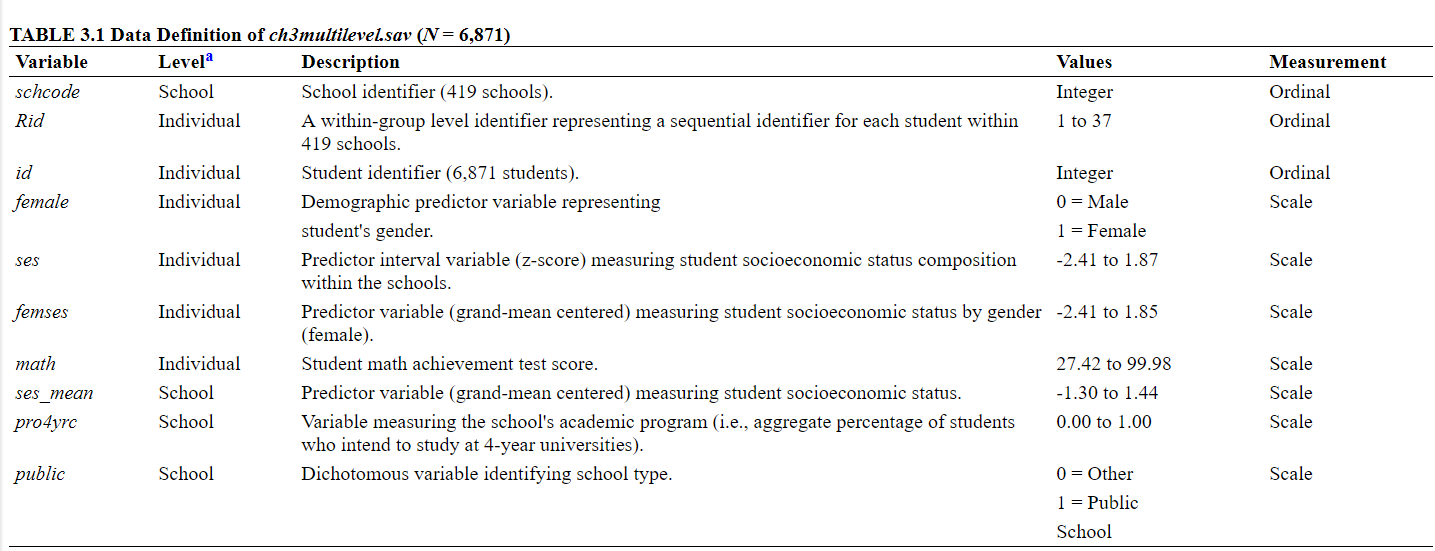
In Module 2, we will first review setting up an R session, then review simple and multiple linear regression.

The data for Module 2 were taken from chapter 3 of Heck, R. H., Thomas, S. L., & Tabata, L. N. (2011). *Multilevel and Longitudinal Modeling with IBM SPSS*: Taylor & Francis. These data have a multilevel structure, which we will work with in Module 3, but for this module we will ignore the clustering structure and conduct regular regression. The following is a description of the variables in the dataset:



1. Open R and create a new project file for this module.
2. Load the packages we will use in this data demo: ggplot2, magrittr.
3. Read in the data using an absolute or relative file path.
4. Run descriptive statistics to do a basic check that data were read in properly and match the following descriptive statistics:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Mean | SD | Minimum | Maximum |
| schcode | 209.43 | 121.44 | 1.00 | 419.00 |
| Rid | 9.20 | 5.71 | 1.00 | 37.00 |
| id | 3436.00 | 1983.63 | 1.00 | 6871.00 |
| female | 0.50 | 0.50 | 0.00 | 1.00 |
| ses | 0.03 | 0.78 | -2.41 | 1.87 |
| femses | 9.84e-05 | 0.55 | -2.41 | 1.85 |
| math | 57.73 | 8.78 | 27.42 | 99.98 |
| ses\_mean | 0.03 | 0.49 | -1.30 | 1.44 |
| pro4yrc | 0,87 | 0.27 | 0.00 | 1.00 |
| public | 0.73 | 0.44 | 0.00 | 1.00 |

1. Run a linear regression predicting math achievement (“math”) from socioeconomic status (“ses”).
2. Interpret the regression coefficients and R-squared value.
3. Visualize the relationship by graphing a scatterplot with math on the y-axis and ses on the x-axis.
4. Add the available sex variable (“female”, 0 = male, 1 = female) as a predictor to the regression from Question 5 and interpret the coefficients and R-squared.
5. Add an interaction term between sex and socioeconomic status and interpret the coefficients and R-squared.

In Module 3, we will consider the clustered data structure we ignored in this module.