Variables and the unit of observation

StatPREP Class Activity

Orientation

There's an expression about how to accomplish a complex, confusing, and difficult task: *One step at a time.* The purpose of this lesson is to provide you with a step-by-step procedure for working with data. Each step can be straightforward. Some steps involve informing yourself about the data at hand. Some steps require creativity and insight about the real-world system the data represents. Some steps require application of a particular technical procedure.

Here are steps for getting started with a data frame that's already available to you.

1. Find out what is the *unit of observation*, that is, what each row stands for.
2. Find out what are the *variables*, the name of each, and what each one stands for.
3. Choose a *response variable*.
4. Select one or more *explanatory variables* which you suspect might account for the response variable.
5. Examine *how much* of the variation in the response variable is accounted for by the explanatory variables.
6. Describe in words and numbers the *pattern of the relationship* between the explanatory variables and the response variable.

In this lesson, you're going to work on steps (1) and (2).

Activity

You'll use the variables-and-units Little App for this lesson. (See footnote[1]). Open the little app and select the NHANES data.

Steps (1) and (2) are to find out about the data frame: What is the unit of analysis and what are the variables. A standard place to get information about a data frame is the *codebook* for the data frame, which is the descriptive documentation about the data.

1. Go to the ***Codebook*** tab in the app. The documentation will appear. The *unit of observation* is the kind of "thing" each row of the data is about.

   - Read the description section of the codebook. This often contains clues about the unit of observation. Note the words "survey," "individuals," "interviewed."

   What do you think is the unit of observation? . .

[1] https://dtkaplan.shinyapps.io/LA_point_plot/

2. The variables are described further on in the codebook.

- Some variables have simple names like `Age`. Find a few other variables whose meaning is obvious to you even without looking at the description in the codebook.

Write down the names of a few obvious variables.   . .

- Some variables have names that are more like a codeword, like `BPSysAve`. Find a few variables with such names and see if the description of the variable's meaning helps you understand what the variable is about.

Are there any variables for which even the description in the codebook doesn't help you understand? Write down one and speculate what it might be about.   . .

Version 0.2, 2019-05-29, Danny Kaplan,