Session 2: Practice Problems

Mark Williamson

November 7, 2017

Review Questions

Problem 1

1. Load the civil war dataset.
2. Choose three numeric variables and calculate their means and medians.

Problem 2

1. Using the original data, create a "dummy" variable for observations where the value is 1 if the country is located in Sub-Saharan Africa and 0 otherwise.
2. Select a numeric variables and calculate its average separately for countries in Sub-Saharan Africa and countries elsewhere. What is the difference in the two means?

Practice Problems

Problem 1

1. There is an issue with the mountainous terrain variable, *mtn*. Instead of NA values, missing data was coded as -99. Ask R for a summary of the *mtn* variable to verify this.
2. Create a new variables called *mtn2* that is the same as *mtn* except with the -99 values replaced with NA's.
3. Bonus: there are two ways to solve this problem - one using dplyr and the other using the basic functions from last week. Try solving the problem using whichever method you didn't use the first time.

Problem 2

1. Create a dataframe containing only oil-exporting countries. When creating this dataset, select only the country, year and log GDP per capita variables.
2. Using dplyr's mutate() function, create a new variable in this dataset that scores 1 if the country is Indonesia and 0 otherwise.

Problem 3

1. Import the National Military Capabilities dataset.
2. Select only the following variables: *cow*, *year* and *milex* from the NMC dataset.
3. Now, join the NMC dataset to the original civil war dataset in a new dataframe called "cw2".