

Quantitative Methods II

Assignment 3

September 11, 2011

This is the third homework assignment for the course. Its purpose is to continue increasing your proficiency in using a statistical program, producing presentation-worthy graphics, and writing to get your point across.

Remember that all statistics are an attempt to gather information about a process (or population) from a sample of data generated by the process. As such, we will never be able to answer a statistical question with absolute certainty, only with statements of confidence ranges and of expected Type I Error rates.

This assignment covers everything we have had assignments thus far, plus simple linear models. Remember: If you need assistance using R, *do not* hesitate to ask for it. To get the most out of such assistance, you will need to explicitly explain your issue, attach the **code** you have already written, and start asking earlier than Sunday.

When you hand in this assignment, you will email to me two separate files, your **typed solutions** to the questions asked in the homework and a separate **script file**. The script file allows me to check that you did the correct analysis. The solution file allows me to see that you can answer the questions in complete and coherent sentences, weaving in graphics and statistics appropriately.

The email must include, as its subject line:

POLS6123: Assignment 3

Note. *Make sure you include neither code nor raw results in the write-up. The code needs to be attached to the email in the separate script file.*

PROBLEM: GDPs PER CAPITA BY REGION

[[25]]

The data file of interest for this problem is the `gdpcap` data file. Please download it and explore it. Get a feel for it. Now, do the following on separate pages.

- (1) Create a table of summary statistics for these four variables: world region (`region`), OPEC membership (`OPEC`), the level of honesty in the government (`hig`), and the GDP per capita in the state (`gdpcap`). The summary statistics should be mean, standard deviation, minimum value, and maximum value. Make the table look nice. Also, calculate and nicely report the number of countries are in each region.

- (2) Calculate the (linear) correlations between the four variables: world region (`region`), OPEC membership (`OPEC`), the level of honesty in the government (`hig`), and the GDP per capita in the state (`gdpcap`). Report these correlation in a correlations table. Make the table look nice. Also, determine which of these correlations is statistically significant.

- (3) Appropriately test the null hypothesis:

H_0 : The average GDP per capita in Africa is \$5000.

Make sure you state the name of the test you will perform, test the assumption(s) and their relevant statistics, and the appropriate three-part conclusion. Include a boxplot summarizing the GDPs per capita in Africa. As this boxplot will be used for utilitarian purposes, the default style is acceptable.

(4) Appropriately test the null hypothesis:

H_0 : The average GDP per capita in Africa is equal to that of Latin America.

Make sure you state the name of the test you will perform, test the assumption(s) and their relevant statistics, and the appropriate three-part conclusion. Include a boxplot comparing the GDPs per capita for Africa and Latin America. Again, this will be for utilitarian purposes, so the default is acceptable.

(5) Appropriately test the null hypothesis:

H_0 : The average GDP per capita in Africa is equal to that of Latin America.

Make sure you state the name of the test you will perform, test the assumption(s) and their relevant statistics, and the appropriate three-part conclusion. Include a boxplot comparing the GDPs per capita for Africa and Latin America. Again, this will be for utilitarian purposes, so the default is acceptable.

(6) You earlier found that there is a statistically significant relationship between GDP per capita (`gdpcap`) and the level of honesty in the government (`hig`). Let us perform linear regression to see how robust this finding is.

Let our dependent variable be the GDP per capita (`gdpcap`). Let us have two independent variables: the level of honesty in the government (`hig`) and world region (`region`). Thus, our research formula will be

$$\text{gdpcap} \sim \text{hig} + \text{region}$$

Note. Observe that `region` is a categorical variable; its estimated coefficients are effect levels as measured relative to some base category.

Finally, let us perform linear regression (`lm`) to simultaneously estimate the effects of the level of honesty in government and the region on the GDP per capita.

Create a summary table of the regression results. Make sure you include the variable name, the estimated effect of that variable, its standard error, the test statistic, and the p-value.

With that information, test these null hypotheses and write the three-part conclusion for each:

- H_0 : The level of honesty in the government has no effect on the GDP per capita when the region of the world it taken into consideration.
- H_0 : When taking the level of honesty in the government into account, the six regions of the world have the same GDP per capita.