

Homework 1

Quantitative Methods 2

Due Wednesday, January 28

.Do File due 1 hour before section in QM2 Homework Folder

Write up due at beginning of class on lecturn

.Do file runs properly = 20 points

1. Create a .do file in your C:\data\QM2 folder (5 points). (*Mac Users, use the file you have set-up for class lectures*).

2.

Call the file: "LastName_ProblemSet1"

1.1. Set up your .do file:

- a. Set the delimiter function.
- b. Set the pause function.
- c. Clear the memory.
- d. Set your new memory amount.
- e. Set the central directory to C:\data\QM2
- f. Open your dataset. Use the data in the file "transition_cross2.DTA" for this analysis. This is a dataset of Eastern European and Former Soviet states. It measures their progress on economic reform.
- g. Save your data as a new dataset.
- h. Open a log file.

2. Description (*describe*¹) of Dataset (5 points: 1.25 points for each sub-question).

- 2.1. How many observations are in your data?
- 2.2. How many variables?
- 2.3. What is the memory size taken up by this file?
- 2.4. What is the meaning of the variable "postcomseats"?

3. Creation of Dependent Variable – Progress toward Complete Reform (5 points)

"Hard Reform" is a variable created by the European Bank of Reconstruction and Development (EBRD) that measures each country's performance on four types of hard, institution building economic reforms. These include: privatization of large state owned enterprises, the enactment of antitrust law (competition policy), bank reform (the creation of an independent two-tier banking sector, and non-bank financial reform (the creation of bond and stock markets). Each of these reforms is measured on a scale ranging from 1 to 4.3. By summing them up, we have a new index of hard reforms with a possible range of 4-17.2.

Create (*generate*) a new variable that equals the percentage progress a country has made toward completion of hard economic reforms. Make sure this variable is on a 100 pt.

¹ Stata commands are written in *italics*.

scale. Call this variable “progress.” *Label* this variable “Progress toward Complete Reform.”

4. Facility with Descriptive Statistics (5 points: 1.25 points for each sub-question).

- 4.1. Calculate the mean and standard deviation of your progress variable “by hand”. In other words, use only *generate*, *tabstat*, and *display* functions.
- 4.2. Verify your answers using the *summary* function.
- 4.3. Calculate one standard deviation from the mean progress in both directions.
- 4.4. Which countries represent the maximum and minimum values in the sample?

5. Detailed Descriptive Statistics (5 points: 1.25 points for each sub-question).

Do not peek at the data or use the *list* function to find this out (use a single command). What is the second largest value of progress in the sample? What is the third smallest value? How much progress constitutes the 75th percentile? The 95th percentile?

6. Ability to count and list (2.5 points: 1.25 points for each sub-question).

How many countries have completed exactly 77.32558% of their hard reforms? Which countries are those? (Hint: Don’t count the countries with missing values.)

7. Graphical depiction of normality (2.5 points: 1.25 points for each sub-question).

- 7.1. Perform a histogram of time since the first independent election (“time”).
- 7.2. In your opinion can we safely assume that “time” has a normal distribution?
- 7.3. Repeat the two operations above for distance from the European Union in 1000s of kilometers (*dist_1000*).

8. Bivariate correlation (5 points).

- 8.1. Calculate the bivariate correlation between progress and *distance_1000* by hand (2.5 points).
- 8.2. Now, use the code for generating a pair-wise correlation coefficient to verify your answer (1.25 points).
- 8.3. Is this correlation statistically significant at the .05 level? (Hint: It involves a sub-command called *star*) (1.25 points).

9. Visual analysis of scatter plot (5 points: 1.25 points for each sub-question).

- 9.1. Visually observe the above relationship in a scatter plot.
- 9.2. Which variable do you think is the dependent and which do you think should be the independent variable? (Hint: You need a visit from Mr. Theory).
- 9.3. Defend your reasoning behind this choice.
- 9.4. Which observations do you think represent outliers on the plot?

10. Visual analysis of linear fit (5 points: 2.5 points for each sub-question).

- 10.1. Draw a regression line through that scatter plot in question 9 with a confidence interval around it.
- 10.2. Eyeballing the line what do you think is its intercept? Its slope?

11. Bivariate Linear Regression (10 points: 2 points for each sub-question).

11.1. Perform a linear regression of the function $\hat{y}_i = \hat{B}_0 + \hat{B}_1 x_i + \hat{u}$

where y is progress and x is distance_1000.

11.2. What are the slope and intercept of the equation?

11.3. Interpret the results by filling in the blank to this sentence: “A 1000 km increase in distance leads to a _____% change in the predicted progress toward complete economic reform?”

11.4. What is the International Political Economy logic behind this finding?

11.5. Can this finding be trusted? Why or why not?

12. Bivariate Linear Regression (10 points: 2 points for each sub-question).

12.1. Perform a linear regression of the function $\hat{y}_i = \hat{B}_0 + \hat{B}_1 x_i + \hat{u}$

where y is progress and x is “polcon3.”

12.2. What are the slope and intercept of the equation?

12.3. Interpret the results by filling in the blank to this sentence: “A 1 point change in the political constraints on the executive (veto players) leads to a _____% change in the predicted progress toward complete economic reform?”

12.4. Based on your study of Policy Making Processes, does this finding make sense to you? Why or why not?

12.5. Can this finding be trusted? Why or why not?

13. Analysis of unexplained variance (5 points: 1.25 points for each sub-question).

13.1. What is Sum of Squared Residuals (SSR) of the model in problem 11?

13.2. Verify this number, calculating by hand.

13.3. What are the total degrees of freedom in the model? What are the residual degrees of freedom?

13.4. What is the Mean Squared Error (Residual)? How do you calculate the Mean Squared Error?

14. Analysis of explained variance (5 points: 2.5 points for each sub-question).

What is the percentage variation of progress explained by the polcon3 variable? How do you know?

15. Understanding of predicted values and residuals (5 points: 1 point for each sub-question)

Re-run the model used in question 11.

15.1. Using the formula $\hat{y}_i = \hat{B}_0 + \hat{B}_1 x_i$ for the model calculate the predicted value for a hypothetical country that is 2000 kilometers away from the EU.

15.2. Generate the predicted values and residuals from the model in question 13 for all observations using the STATA *predict* command?

15.3. Using the *list* command, observe the predicted values and residuals next to the observed progress on economic reform for every country.

15.4. Verify that the predicted value \hat{y}_i plus the residual \hat{u}_i equals the observed progress (y_i) in three different observations?

15.5. Plot the observed progress and the predicted progress on the same scatter plot, where distance_1000 is the independent variable? Label the observation by the country name.