

**SIS 600 - International Affairs Statistics and Methods**  
**Problem Set 3**  
**Due via email by May 1**

Go to the Problem Set 3 Folder in the Assignments section of Blackboard. There, you'll find data, an article, and an Excel file outlining and describing the data in the dataset. Download the data and answer the questions below. Include all Stata outputs with your answers. Be sure Stata outputs are readable, clear and professional (such as a screenshot, regression table or other easily readable format).

In the same file as your document, copy and paste your annotated do file.

## Oil Wealth and Democracies

1. Read the article by Fains (2019) with particular attention to the Empirical Approach section and what the polyarchy index means.
2. Familiarize yourself with the codebook.
3. Import `problemset3.csv` file into Stata.
4. Familiarize yourself with the dataset.
5. What is the unit of analysis? How many observations are there?
6. Find the countries that are in the top decile for **Growth in Income, Per Capita**. Which countries have the highest per capita growth? Find the countries that are in the bottom decile for **Growth in Income, Per Capita**. Which countries have the lowest per capita income growth? Do countries in the top decile have anything in common? Do countries in the bottom decile have anything in common?
7. Generate a variable `avg_oil` and populate all observations with missing values. Replace these values with a '1' if the observation had at or above average (mean) oil income per capita (`oilincome`). Replace these values with a '0' if the observation had lower than average (mean) oil income per capita. Do places with a higher oil income per capita on average have lower income growth per capita? What is the null hypothesis that you are testing? Report your results and be sure to discuss the confidence with which you report them.
8. Is there a systematic relationship between the polyarchy index `poly5` and countries with above average oil income per capita? Specify a null and alternative hypothesis with respect to this question. Investigate the data to identify the test best suited to answer this question. Report your results and be sure to discuss the confidence with which you report them.
9. Is there a systematic relationship between being a democracy (`dem`) and oil income per capita (`oilincome`)? Specify a null and alternative hypothesis with respect to this question. Investigate the data to identify the test best suited to answer this question. Report your results and be sure to discuss the confidence with which you report them.
10. How correlated are income growth per capita and polity scores? Is this correlation statistically significant? Interpret these results as if you were explaining them to a smart person who had never taken a statistics class.
11. Produce a scatter plot for the relationship found in question 9. Include the best fit line. Interpret this figure.
12. Produce a scatter plot of polity scores and the variable `poly5`. Include the best fit line. Interpret this figure.

13. Test the following null hypothesis using OLS Regression: oil income per capita has no effect on the how democratic a country is. Interpret this substantively by explaining how many standard deviations polity scores change as a result of a one-standard deviation increase in oil income per capita (logged). Additionally, explain how much polity scores change when moving from the 25th percentile to the 75th percentile of oil income per capita.
14. Using the same variable as your dependent variable from the previous question, estimate a theoretically informed model by selecting from the remaining variables in the dataset to use as your independent variables. What estimator (OLS regression, logistic, t-test, cross-tabs, correlation) will you use? Why did you include these variable(s)? Are there any additional variables **not** in the dataset that you would like to include but cannot? Report the results of your model.
15. Interpret the results of your model. Which variables are significant? Do these variables have a positive or negative effect on the dependent variable you selected? Interpret the substantive effects of each of the variables you included such that a smart person who had never taken a statistics class could understand your results.

## Data Analysis

In this section we will work with the authors' data to develop our own models. Review the "Empirical Approach" section and the "Results" sections. Use the `poly5dv` variable for your dependent variable.

16. How is the `poly5dv` variable constructed (categorical, binary, continuous, ordinal)?
17. Estimate a theoretically informed model by selecting from the remaining variables in the dataset to use as your independent variables. What estimator (OLS regression, logistic, t-test, cross-tabs, correlation) will you use? Why did you include these variable(s)? Are there any additional variables **not** in the dataset that you would like to include but cannot? Report the results of your model.
18. Interpret the results of your model. Which variables are significant? Do these variables have a positive or negative effect on the dependent variable you selected? Interpret the substantive effects of each of the variables you included such that a person who had never taken a statistics course but was very smart could understand your model.
19. Look at Model I, of Table I. We will use most of the same control variables from this model. Using the same dependent variable you used in Question 17, include the following control variables in your model: `oilincome`, `lincome_new`, `WBgrow_new`, `poly5`, `pastpoly5`, `aclp_region`, `vtim`, `vtim2`, `vtim3`. Report the results.
20. Compare the model above (Question 19) to the model you created in Question 17. Assess whether your model or the authors' model is better. Why? If you produced the same model, explain how the model could be improved.