

# Problem Set #4

## IRIII.2 - Quantitative Methods in the Study of International Relations

Steven V. Miller

This homework is aimed at replication and doing a full linear regression in R. The data used will come by way of {stevedata}. The student should be explicit about the R packages used and a failure to do this may result in zero-credit for the assignment.

### Optimal Obfuscation: Democracy and Trade Policy Transparency

This homework assignment will use the OODTPT data included in {stevedata}. This data set is based on a replication data set provided by Daniel Kono for his 2006 article in *American Political Science Review*, which is also included in the course plan as additional reading.<sup>1</sup> Answer the following questions based on the data and you're reading of the article.

```
?OODTPT
```

Here are the data.

```
OODTPT
```

```
## # A tibble: 75 x 16
##   country    isocode tariff corecov qualcov polity   iec lngdppc lngdp lnexpgdp
##   <chr>      <chr>   <dbl>  <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Albania    ALB     15.9    0        2.02  5      14    6.60  21.6  2.36
## 2 Algeria    DZA     23.5    3.47    20.6  -5      9     7.34  24.5  3.11
## 3 Argentina ARG     13.4    28.2    51.0  7.33   14    8.98  26.4  2.26
## 4 Australia AUS      4.96    1.21    17.8  10     14   10.0  26.8  2.96
## 5 Austria    AUT     12.7    74.0     5.13  10     14   10.2  26.1  3.58
## 6 Bangladesh BGD     52.4    21.5    25.5  6      14    5.81  24.4  2.41
## 7 Belarus    BLR     12.6    0.0269  15.9  -7     13    7.74  23.9  4.09
## 8 Bhutan     BTN     15.3    0.776    6.12  -8      6     6.24  19.8  3.37
## 9 Bolivia    BOL      9.74    6.01    11.8  9      14    6.83  22.7  2.99
## 10 Brazil     BRA     14.5    14.4    47.2  8      14    8.40  27.3  2.19
## # i 65 more rows
## # i 6 more variables: reer <dbl>, growth <dbl>, dimpgdp <dbl>, lngovcons <dbl>,
## #   gatt <dbl>, avgstar <dbl>
```

<sup>1</sup>Kono, Daniel. 2006. "Optimal Obfuscation: Democracy and Trade Policy Transparency" *American Political Science Review* 100(3): 369-384.

Answer these questions/do these commands. A successful answer of these question must include the R code you used to help you answer the question. Each question is worth a point.

1. In two-to-three complete sentences, describe the hypothesis that Kono (2006) is interested in testing. Be careful to identify what is the purported cause and what is the purported effect and why Kono has reason to believe there is this relationship. Notice that the issue in this prompt is conceptual rather than operational, so your treatment here should not reference R code or even the data.
2. In the data that Kono uses, and is available for you in OODTPT in `{stevedata}`, identify one of the three variables/column names that coincide with the “effect” of interest to Kono (2006) and one of the two variables/column names that coincide with the *main* “cause” of interest.<sup>2</sup>
3. For the two variables you identified in question 2, what is the Pearson’s *r* correlation coefficient? Your answer should include R code.
4. For the two variables you identified in question 2, run a simple bivariate linear regression that regresses the variable that is the “effect” (dependent variable) on the variable that is the “cause” (independent variable). Identify the coefficient size, direction, and statistical significance in at least one complete sentence. Your answer should include the R code used to run the regression.
5. Revisit question 2, for which you chose one of the two variables that Kono (2006) uses as proxies for the main “cause”. Here, use the other proxy of the main “cause” that Kono (2006) uses. Run a simple bivariate linear regression that regresses the variable that is the “effect” (dependent variable) on this other variable that is the main “cause” (independent variable). Identify the coefficient size, direction, and statistical significance in at least one complete sentence. Your answer should include the R code used to run the regression.

### Extra Credit

6. [1 POINT] A copy of this data set, in Stata format, is available for download on Athena. Download it to your personal computer and provide the R code necessary to load it into R. Assign it to an object name of your choosing but be mindful to use the full file path.
7. [3 POINTS] Observe Model 1 in Table 1 (which is the left-most regression model reported in the table) and replicate it using the `lm()` function in R. For the *main* independent variable of interest in the analysis, tell me the size of the coefficient (to four-decimal points to prove to me you’re not just mimicking the article).

---

<sup>2</sup>An eagle-eyed student will notice you cannot treat the empirical analysis section of an article as a page-turner in order to answer this question.