

Department of Politics

**ESSAY COVER SHEET**

|  |
| --- |
| **Date**: 10/07/2023 |
| **Student Number**: 13809430 |
| **Programme of Study**: International Security and Global Governance |
| **Module**: Advanced Quantitative Social Research |
| **Essay title**: Problem Set 1 – Multiple Linear Regression |
| **Word Count (including all footnotes, references and appendices)**: 1318 |
| **Disability and dyslexia support**: Do you have an Individual Student Support Agreement with the Birkbeck Disability Office that is relevant to this coursework?  **No** |
| **Plagiarism statement**:  Coursework is monitored for plagiarism and if detected may result in disciplinary action. In submitting this coursework, I hereby confirm that I have read Birkbeck's plagiarism guidelines and taken the online tutorial on avoiding plagiarism and on this basis declare that this coursework is free from plagiarism.  Plagiarism guidelines: <http://www.bbk.ac.uk/politics/current-students/essays/plagiarism>  Plagiarism tutorial: <http://www.bbk.ac.uk/mybirkbeck/get-ahead-stay-ahead/writing/referencing> |

**Problem set 1: Multiple Linear Regression**

**Question 1**

Visit the V-Dem Dataset’s website and read through the definitions of political corruption (v2x\_corr) and electoral democracy ( v2x\_polyarchy ). Do you agree with their definitions? Discuss.

**a)** Electoral democracy

In the case of electoral democracy, I think the definition used in the codebook is an accurate definition of electoral/polyarchic democracy as it takes into account the core dimensions of polyarchy. The measure would not accurately represent earlier conceptions of polyarchy, where polyarchy was considered a necessary but not sufficient condition for full democracy. However, Dahl’s definition of polyarchy evolved over time to become broader and with reference to more than simply contestation and participation as core dimensions (e.g. to include freedom of expression, access to alternative information and ‘associational autonomy’; Dahl, 1998). These additional dimensions expanded by Dahl are well represented, for example, by the freedom of the media, active civil society and *inclusive* suffrage included in V-Dem. However, aggregating these to a broad measure can disguise instances where some of the institutions of electoral democracy may be less than perfect. As a result, this measure may not entirely represent the nuanced reality of electoral democracy represented in the V-Dem definition.

**b)** Political corruption

For the corruption index, though I don’t disagree with the indicators included (broadly; executive, legislative, judicial and public sector corruption), I do think it could be expanded. The definition involves bribery and embezzlement; however, corruption can often be more pervasive and indirect than these obvious examples. Though this is accounted for by ‘corrupt activities’ of the legislature, corruption is limited to bribery and embezzlement in the executive. This doesn’t include subtler forms of abuse of power such as government deals for friends and donors, selection of friends and donors for important government appointed positions, or the revolving door between government and private companies. Corruption can take more insidious forms than simple exchanges of money, something I think this definition (and measure) of corruption could better represent.

**Question 2**

Use v2x\_polyarchy\_VDEM\_10 (the electoral democracy index) as the only independent variable to predict or explain a country’s GDP per capita (gdppc\_WDI\_PW).

**a)** Specify the model to be estimated (i.e., model specification with the error term).

The model I’ll be fitting using electoral democracy to explain GDP per capita:

Where:

Y is GDP per capita

X is polyarchy/electoral democracy

That is:



**b)** Show the scatterplot with the regression line.

A picture containing text, line, plot, screenshot

Description automatically generated

A picture containing text, font, line, screenshot

Description automatically generated

**c)** Run the model and interpret the estimated intercept and slope.

A picture containing text, screenshot, font, number

Description automatically generated

When electoral democracy is held at 0, GDP per capita is estimated to be -$2878. With a 1 unit increase in electoral democracy, GDP per capita is estimated to increase by $2594.6.

A screen shot of a computer code

Description automatically generated with low confidence

**d)** Calculate the 95% confidence interval of the estimated slope () and explain whether is statistically significant; if yes, discuss its substantive significance.

A picture containing text, font, screenshot, line

Description automatically generated

As the confidence intervals for (v2x\_polyarchy\_VDEM\_10) do not contain 0, it means that the estimated slope for the variable is statistically significant, as we’re confident that these intervals will contain the true population mean 95% of the time. These CIs suggest that 95% of the time, a 1-unit increase (/decrease) in polyarchy would generate an increase (/decrease) in GDP per capita of between $1774.9 and $3414.2. With a median GDP per capita of $3287.7, this means the estimated effect of increasing democracy would be noticeable for the median country in the dataset.

A picture containing text, font, white

Description automatically generated

**e)** Use stargazer to show b\_corruption and m\_mod together with b\_democracy; look across all model specifications and discuss whether democracy has an impact on development when we account for political corruption.

A picture containing text, screenshot, font, number

Description automatically generated

This table suggests that democracy does not have a significant impact on development when accounting for corruption. The model b\_corruption (2) shows that corruption alone is a significant predictor of development, and the model has a fairly good adjusted R2. Although polyarchy alone is statistically significant in model b\_democracy, the adjusted R2 of this model is poor. M\_mod reinforces that democracy has a negligible impact on development when accounting for corruption, as the addition of the democracy variable to a model with corruption renders the democracy variable non-significant. Moreover, the addition of democracy *reduces* the adjusted R2 value of m\_mod (compared A screenshot of a computer code

Description automatically generated with medium confidenceto b\_corruption), suggesting that it lessens the explanatory power of the model.

**f)** Use anova() to compare across different models to see which model performs the best.

A picture containing text, font, screenshot, white

Description automatically generated

The non-significant p-value in the table above table suggests that m\_mod (model 3 above) does not explain GDP per capita better than models 1 or 2. However, this is likely because b\_corruption (model 2 above) did explain GDP per capita well, meaning m\_mod did not perform significantly better than it with the addition of democracy (as mentioned in part E, above).

The p-values of the below comparisons demonstrate that when compared to b\_democracy alone, m\_mod does perform significantly better, however, when compared to b\_corruption alone, m\_mod does not perform significantly better. As good regression models need to be parsimonious, the fact that the addition of democracy does not improve the explanatory power of m\_mod suggests that b\_corruption is probably the best model.

A screenshot of a computer code

Description automatically generated with low confidence

A picture containing text, font, screenshot, white

Description automatically generated

A picture containing text, font, screenshot, algebra

Description automatically generated

**References**

Dahl, R. A. (1998). On Democracy. Yale University Press. https://doi.org/10.2307/j.ctv18zhcs4

**Appendix – R code**

Replication Exercise 1

SN: 13809430

## Assessment Problem Set 1 - Multiple Linear Regression  
#Student Number - 13809430  
  
##Load packages, set wd  
library(car) #<- for variance inflation factor - collinearity  
library(haven)  
library(tidyverse) #<- useful for easy stats, data cleaning  
library(stargazer) #<- to present (& combine) regression table  
library(dplyr)   
setwd("C:/Users/name/OneDrive/Documents/R/ATQSR/Week 2/")  
list.files()  
  
## Import the data, remove NAs, duplicates  
dta <- read.csv("./week2\_data1.csv")  
View(dta)  
  
#Using the same(ish) cut of data as in the problem set, otherwise anova won't work  
dta2 <- dta |>  
 dplyr::select(country, year, gdppc\_WDI\_PW, democracy\_DD, v2x\_polyarchy\_VDEM, v2x\_corr\_VDEM) |>  
 mutate(v2x\_polyarchy\_VDEM\_10 = v2x\_polyarchy\_VDEM\*10,  
 v2x\_corr\_VDEM\_10 = v2x\_corr\_VDEM\*10) |>  
 filter(year == 2000) |>  
 unique() |>  
 drop\_na()  
View(dta2)  
  
#######################################  
## Question 1 A/B  
#See word document  
  
#######################################  
## Question 2  
######################################  
##Part A - specify the model to be estimated  
b\_democracy <- lm(gdppc\_WDI\_PW ~ v2x\_polyarchy\_VDEM\_10, data = dta2)  
  
######################################  
##Part B - scatter plot with regression line  
plot(gdppc\_WDI\_PW ~ v2x\_polyarchy\_VDEM, data=dta2,  
 xlab = "Polyarchy",  
 ylab = "GDP per capita",  
 main = "Democracy and Development")  
abline(lm(gdppc\_WDI\_PW ~ v2x\_polyarchy\_VDEM, data=dta2),#adds regression line  
 col = "green", #colour  
 lwd = 2.5) #line weight  
  
## Prettier/easier to read (gg)plot  
ggplot(dta2, aes(x = v2x\_polyarchy\_VDEM, y = gdppc\_WDI\_PW, color = as.factor(democracy\_DD))) +  
 theme\_minimal() +  
 geom\_point() +  
 geom\_smooth(method = "lm", se = FALSE, color = "darkgreen", fullrange = TRUE, size = 0.9) +  
 xlim(c(-0.001,1)) + #regression line wouldn't extend to Y axis otherwise  
 geom\_hline(yintercept = 0, color = "black", size = 0.3) +  
 geom\_vline(xintercept = 0, color = "black", size = 0.3) +  
 theme(plot.title = element\_text(hjust = 0.5), legend.position = "bottom") +  
 scale\_color\_hue(name = "democracy\_DD" ,labels = c("0 (non-dem)", "1 (dem)")) +  
 labs(title = "Democracy and Development", x = "Polyarchy", y = "GDP per capita")  
  
######################################  
##Part C - model interpretation (see word document)  
#b\_democracy <- lm(gdppc\_WDI\_PW ~ v2x\_polyarchy\_VDEM, data = dta)  
summary(b\_democracy)  
stargazer(b\_democracy,  
 omit.stat = c("f", "rsq", "ser"),  
 covariate.labels = c("Democracy"),  
 type = "text",  
 digits = 3,   
 no.space = T,  
 intercept.bottom = TRUE,  
 star.cutoffs = c(0.05, 0.01, 0.001))  
  
median(dta2$gdppc\_WDI\_PW)  
######################################  
##Part D - 95% confidence interval of slope  
#'Manually'  
coefficients <- coef(b\_democracy)  
stderrs <- summary(b\_democracy)$coef[, "Std. Error"]  
cv <- qnorm(0.975) #critical value  
lower\_bound <- coefficients - cv \* stderrs  
upper\_bound <- coefficients + cv \* stderrs  
lower\_bound  
upper\_bound  
  
#With function  
confint(b\_democracy)  
  
######################################  
##Part E - regression model table  
#Specify m\_mod, b\_corruption  
m\_mod <- lm(gdppc\_WDI\_PW ~ v2x\_corr\_VDEM\_10 + v2x\_polyarchy\_VDEM\_10, data=dta2)  
b\_corruption <- lm(gdppc\_WDI\_PW ~ v2x\_corr\_VDEM\_10, data=dta2)  
  
#Create stargazer table  
stargazer(list(m\_mod, b\_corruption, b\_democracy),  
 omit.stat = c("f", "rsq", "ser"),  
 column.labels = c("mmod","bcorruption","bdemocracy"), #Stargazer apparently doesn't like underscores in column or covariate labels, so had to remove those  
 column.separate = c(1,1,1),  
 type = "text",  
 digits = 3,   
 no.space = T,  
 intercept.bottom = TRUE,  
 star.cutoffs = c(0.05, 0.01, 0.001))  
  
######################################  
##Part F - ANOVA  
#All three  
anova(b\_democracy, b\_corruption,m\_mod)  
  
#democracy vs corruption & democracy  
anova(b\_democracy,m\_mod)  
  
#corruption vs corruption & democracy  
anova(b\_corruption, m\_mod)  
  
anova(b\_corruption, b\_democracy)