

April 2023

```
{r setup, include = FALSE} knitr::opts_chunk$set(warning = FALSE,  
message = FALSE, error=TRUE)
```

Assignment 4: Collaborating Together

Part 1: Contributing to another student's Github repository

In this assignment, you will create a Github repository, containing this document and the .pdf output, which analyzes a dataset individually using some of the tools we have developed.

This time, make sure to not only put your name and student e-mail in your Rmarkdown header, but also your Github account, as I have done myself.

However, you will also pair up with a class mate and contribute to each others' Github repository. Each student is supposed to contribute to another student's work by writing a short interpretation of 1 or 2 sentences at the designated place (this place is marked with **designated place**) in the other student's assignment.

This interpretation will not be graded, but a Github shows the contributors to a certain repository. This way, we can see whether you have contributed to a repository of a class mate.

Question 1.1: Fill in the **github username** of the class mate to whose repository you have contributed.

ElectronicJonah

Part 2: Analyzing various linear models

In this part, we will summarize a dataset and create a couple of customized tables. Then, we will compare a couple of linear models to each other, and see which linear model fits the data the best, and yields the most interesting results.

We will use a dataset called GrowthSW from the AER package. This is a dataset containing 65 observations on 6 variables and investigates the determinants of economic growth. First, we will try to summarize the data using the modelsummary package.

```
{r, warning=FALSE, message=FALSE} library(AER) data(GrowthSW)
```

One of the variables in the dataset is revolutions, the number of revolutions, insurrections and coup d'états in country i from 1965 to 1995.

Question 2.1: Using the function datasummary, summarize the mean, median, sd, min, and max of the variables growth, and rgdp60 between two groups: countries with revolutions equal to 0, and countries with more than 0 revolutions. Call this variable treat. Make sure to also write the resulting data set to memory. Hint: you can check some examples [here](#).

```
library(modelsummary); library(tidyverse)

GrowthSW <- GrowthSW %>%
  mutate(treat = ifelse(revolutions > 0, "more than 0", "equal to 0"))
datasummary <- datasummary(growth + rgdp60 ~ treat*(Mean + Median + SD
+ Min + Max),data=GrowthSW)
datasummary(growth + rgdp60 ~ treat*(Mean + Median + SD + Min +
Max),data=GrowthSW)
# write your code here
```

Designated place: type one or two sentences describing this table of a fellow student below. For example, comment on the mean and median growth of both groups. Then stage, commit and push it to their github repository.

Part 3: Make a table summarizing regressions using modelsummary and kable

In question 2, we have seen that growth rates differ markedly between countries that experienced at least one revolution/episode of political stability and countries that did not.

Question 3.1: Try to make this more precise this by performing a t-test on the variable growth according to the group variable you have created in the previous question.

```
result <- t.test(growth ~ treat, data = GrowthSW)

print(result)
# write t test here
```

Question 3.2: What is the p -value of the test, and what does that mean? Write down your answer below.

The p -value result is equal to 0.06871. Considering we use a 95% confidence interval, and we have a p -value greater than 0.05, this means that the observed data is likely to occur even if the null hypothesis is true. We fail to reject the null hypothesis

We can also control for other factors by including them in a linear model, for example:

$$\text{growth}_i = \beta_0 + \beta_1 \cdot \text{treat}_i + \beta_2 \cdot \text{rgdp60}_i + \beta_3 \cdot \text{tradeshare}_i + \beta_4 \cdot \text{education}_i + \epsilon_i$$

Question 3.3: What do you think the purpose of including the variable rgdp60 is? Look at ?GrowthSW to find out what the variables mean.

The purpose of having the variable rgdp60 is that it displays an effective comparison for GDP per capita among different countries.

We now want to estimate a stepwise model. Stepwise means that we first estimate a univariate regression $\text{growth}_i = \beta_0 + \beta_1 \cdot \text{treat}_i + \epsilon_i$, and in each subsequent model, we add one control variable.

Question 3.4: Write four models, titled `model1`, `model2`, `model3`, `model4` (using the `lm` function) to memory. Hint: you can also use the `update` function to add variables to an already existing specification.

```
model1 <- lm(growth ~ treat, data=GrowthSW)
model2 <- update(model1, growth ~ treat + rgdp60)
model3 <- update(model2, growth ~ treat + tradeshare)
model4 <- update(model3, growth ~ treat + education)
```

Now, we put the models in a list, and see what `modelsummary` gives us:

```
list(model1, model2, model3, model4) |>
  modelsummary(stars=TRUE,
               gof_map = c("nobs", "r.squared"),
               title = c("model1", "model2", "model3", "model4"))
```

Question 3.5: Edit the code chunk above to remove many statistics from the table, but keep only the number of observations N , and the R^2 statistic.

Question 3.6: According to this analysis, what is the main driver of economic growth? Why?

According to this analysis, both `tradeshare` and `education` are main drivers of economic growth in a country. `Tradeshare` is a positive coefficient with a statistical significance level at 1% which implies that the variable has a positive impact on economic growth. Similarly, `education` is a positive coefficient with a statistical significance level at 5%, also implying that this variable has a positive impact on growth. We know this because `tradeshare` has a "***" which implies the 1% statistical significance level. `Education` has a "*" which implies the 5% statistical significance level.

Question 3.7: In the code chunk below, edit the table such that the cells (including standard errors) corresponding to the variable `treat` have a red background and white text. Make sure to load the `kableExtra` library beforehand.

```
library(kableExtra)
list(model1, model2, model3, model4) |>
  modelsummary(stars=T, gof_map = c("nobs", "r.squared")) |>
  row_spec(3, color = "white", background = "red")
```

use functions from `modelsummary` to edit this table

Question 3.8: Write a piece of code that exports this table (without the formatting) to a Word document.

```
library(flextable)
modelsummary(list(model1,model2,model3,model4), gof_map = c("nobs",
```

```
"r.squared"), title = "Regression table", output =  
"table_growthSW.docx")
```

The End