# Assignment 4: Collaborating Together
# Introduction to Applied Data Science
# 2022-2023

Ying Ying Tsai
y.tsai1@students.uu.nl
https://github.com/yingying171

April 2023

## 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.

[Fill in here]

### 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.

```
library(AER)
data(GrowthSW)
```

One of the variables in the dataset is `revolutions`, the number of revolutions, insurrections and coup d'etats 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)

# write your code here
# Creating the treat variable based on revolutions
GrowthSW$treat <- ifelse(GrowthSW$revolutions > 0, "More than 0 revolutions", "0 revolutions")

# Summarizing the variables growth and rgdp60 by treat groups
summary_table <- datasummary(GrowthSW,
                             var.labels = c(growth = "Growth", rgdp60 = "Real GDP 1960"),
                             by = "treat",
                             statistics = c("mean", "median", "sd", "min", "max"))
```

```
## Error in datasummary(GrowthSW, var.labels = c(growth = "Growth", rgdp60 = "Real GDP 1960"), : argume
```

```
# Writing the resulting dataset to memory
write.csv(summary_table, file = "summary_table.csv", row.names = FALSE)
```

```
## Error in eval(expr, p): object 'summary_table' not found
```

**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 reressions 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.

```
# write t test here
t_test_result <- t.test(growth ~ treat, data = GrowthSW)

t_test_result
```

```
##
##  Welch Two Sample t-test
##
## data:  growth by treat
## t = 1.8531, df = 61.015, p-value = 0.06871
## alternative hypothesis: true difference in means between group 0 revolutions and group More than 0 re
## 95 percent confidence interval:
##  -0.06182741  1.62566475
## sample estimates:
##           mean in group 0 revolutions mean in group More than 0 revolutions
##                              2.459985                              1.678066
```

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

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.

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.

```
# Model 1: Univariate regression
model1 <- lm(growth ~ treat, data = GrowthSW)

# Model 2: Adding rgdp60 as a control variable to model 1
model2 <- update(model1, . ~ . + rgdp60)

# Model 3: Adding tradeshare as a control variable to model 2
model3 <- update(model2, . ~ . + tradeshare)

# Model 4: Adding education as a control variable to model 3
model4 <- update(model3, . ~ . + education)

save(model1, model2, model3, model4, file = "stepwise_models.RData")
```

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

```
library(modelsummary)

# Putting the models in a list
model_list <- list(model1, model2, model3, model4)

# Applying modelsummary to the list of models
model_summary <- model_list |>
  modelsummary(stars = TRUE,
               statistics = c("rsq", "n"))
model_summary
```

**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.

```
# Loading the required packages
library(modelsummary)

# Putting the models in a list
model_list <- list(model1, model2, model3, model4)

# Applying modelsummary to the list of models
model_summary <- model_list |>
  modelsummary(stars = TRUE,
```

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| (Intercept) | 2.460*** | 2.854*** | 0.839 | −0.050 |
| | (0.400) | (0.751) | (1.045) | (0.967) |
| treatMore than 0 revolutions | −0.782 | −1.028 | −0.415 | −0.069 |
| | (0.491) | (0.633) | (0.647) | (0.589) |
| rgdp60 | | 0.000 | 0.000 | 0.000* |
| | | (0.000) | (0.000) | (0.000) |
| tradeshare | | | 2.233* | 1.813* |
| | | | (0.842) | (0.765) |
| education | | | | 0.564*** |
| | | | | (0.144) |
| Num.Obs. | 65 | 65 | 65 | 65 |
| R2 | 0.039 | 0.045 | 0.143 | 0.318 |
| R2 Adj. | 0.023 | 0.014 | 0.101 | 0.272 |
| AIC | 270.1 | 271.7 | 266.6 | 253.8 |
| BIC | 276.7 | 280.4 | 277.5 | 266.9 |
| Log.Lik. | −132.069 | −131.867 | −128.319 | −120.918 |
| F | 2.532 | 1.446 | 3.403 | 6.989 |
| RMSE | 1.85 | 1.84 | 1.74 | 1.55 |

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

```
               statistics = c("n", "rsq"),
               output = "default")

# Printing the model summary
model_summary
```

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

**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)
library(modelsummary)

# Generating the model summary table
model_table <- list(model1, model2, model3, model4) |>
  modelsummary(stars = TRUE, gof_map = c("nobs", "r.squared"))

# Applying kableExtra functions to modify the table
model_table_modified <- model_table %>%
  kable() %>%
  kable_styling() %>%
  row_spec(row = which(rownames(model_table) == "treat"),
           background = "red", color = "white") %>%
  row_spec(which(rownames(model_table) == "treat"),
           bold = TRUE, italic = TRUE) %>%
  column_spec(column = which(rownames(model_table) == "treat"),
              background = "red", color = "white")
```

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| (Intercept) | 2.460*** | 2.854*** | 0.839 | −0.050 |
| | (0.400) | (0.751) | (1.045) | (0.967) |
| treatMore than 0 revolutions | −0.782 | −1.028 | −0.415 | −0.069 |
| | (0.491) | (0.633) | (0.647) | (0.589) |
| rgdp60 | | 0.000 | 0.000 | 0.000* |
| | | (0.000) | (0.000) | (0.000) |
| tradeshare | | | 2.233* | 1.813* |
| | | | (0.842) | (0.765) |
| education | | | | 0.564*** |
| | | | | (0.144) |
| Num.Obs. | 65 | 65 | 65 | 65 |
| R2 | 0.039 | 0.045 | 0.143 | 0.318 |
| R2 Adj. | 0.023 | 0.014 | 0.101 | 0.272 |
| AIC | 270.1 | 271.7 | 266.6 | 253.8 |
| BIC | 276.7 | 280.4 | 277.5 | 266.9 |
| Log.Lik. | −132.069 | −131.867 | −128.319 | −120.918 |
| F | 2.532 | 1.446 | 3.403 | 6.989 |
| RMSE | 1.85 | 1.84 | 1.74 | 1.55 |

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

x

\begin{table} \centering \begin{tabular}[t]{lcccc} \toprule & (1) & (2) & (3) & (4)\\ \midrule (Intercept) & \num{2.460

```
# Printing the modified table
model_table_modified
```

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

```
library(pander)

# Convert the modified table to a markdown table
markdown_table <- pandoc.table(model_table_modified)
```

```
##
## ------------------------------------------------------------------------------
##                           \begin{table} \centering
##                         \begin{tabular}{l} \hline x\\
##                                       \hline
##                         \textbackslash{}begin\{table\}
##                           \textbackslash{}centering
##                  \textbackslash{}begin\{tabular\}[t]\{lcccc\}
##                          \textbackslash{}toprule   \&
##                              (1) \& (2) \& (3) \&
##                      (4)\textbackslash{}\textbackslash{}
##                            \textbackslash{}midrule
##                                (Intercept) \&
##                        \textbackslash{}num\{2.460\}***
##                                       \&
##                        \textbackslash{}num\{2.854\}***
##                                       \&
```

```
##                          \textbackslash{}num\{0.839\}
##                                      \&
##        \textbackslash{}num\{-0.050\}\textbackslash{}\textbackslash{}
##                                      \&
##                         (\textbackslash{}num\{0.400\})
##                                      \&
##                         (\textbackslash{}num\{0.751\})
##                                      \&
##                         (\textbackslash{}num\{1.045\})
##                                      \&
##        (\textbackslash{}num\{0.967\})\textbackslash{}\textbackslash{}
##                          treatMore than 0 revolutions
##                                      \&
##                         \textbackslash{}num\{-0.782\}
##                                      \&
##                         \textbackslash{}num\{-1.028\}
##                                      \&
##                         \textbackslash{}num\{-0.415\}
##                                      \&
##        \textbackslash{}num\{-0.069\}\textbackslash{}\textbackslash{}
##                                      \&
##                         (\textbackslash{}num\{0.491\})
##                                      \&
##                         (\textbackslash{}num\{0.633\})
##                                      \&
##                         (\textbackslash{}num\{0.647\})
##                                      \&
##        (\textbackslash{}num\{0.589\})\textbackslash{}\textbackslash{}
##                                rgdp60 \&  \&
##                          \textbackslash{}num\{0.000\}
##                                      \&
##                          \textbackslash{}num\{0.000\}
##                                      \&
##        \textbackslash{}num\{0.000\}*\textbackslash{}\textbackslash{}
##                                    \&  \&
##                         (\textbackslash{}num\{0.000\})
##                                      \&
##                         (\textbackslash{}num\{0.000\})
##                                      \&
##        (\textbackslash{}num\{0.000\})\textbackslash{}\textbackslash{}
##                             tradeshare \&  \&  \&
##                         \textbackslash{}num\{2.233\}*
##                                      \&
##        \textbackslash{}num\{1.813\}*\textbackslash{}\textbackslash{}
##                                   \&  \&  \&
##                         (\textbackslash{}num\{0.842\})
##                                      \&
##        (\textbackslash{}num\{0.765\})\textbackslash{}\textbackslash{}
##                           education \&  \&  \&  \&
##       \textbackslash{}num\{0.564\}***\textbackslash{}\textbackslash{}
##                                \&  \&  \&  \&
##        (\textbackslash{}num\{0.144\})\textbackslash{}\textbackslash{}
##                           \textbackslash{}midrule
##                                 Num.Obs. \&
```

```
##                           \textbackslash{}num\{65\} \&
##                           \textbackslash{}num\{65\} \&
##                           \textbackslash{}num\{65\} \&
##          \textbackslash{}num\{65\}\textbackslash{}\textbackslash{}
##                                       R2 \&
##                           \textbackslash{}num\{0.039\}
##                                       \&
##                           \textbackslash{}num\{0.045\}
##                                       \&
##                           \textbackslash{}num\{0.143\}
##                                       \&
##         \textbackslash{}num\{0.318\}\textbackslash{}\textbackslash{}
##                            \textbackslash{}bottomrule
##  \textbackslash{}multicolumn\{5\}\{l\}\{\textbackslash{}rule\{0pt\}\{1em\}+
##                           p \$<\$ 0.1, * p \$<\$ 0.05,
##                           ** p \$<\$ 0.01, *** p \$<\$
##                    0.001\}\textbackslash{}\textbackslash{}
##                         \textbackslash{}end\{tabular\}
##                         \textbackslash{}end\{table\}\\
##                              \hline \end{tabular}
##                                  \end{table}
##
## ------------------------------------------------------------------------------
```

```
# Write the markdown table to a Word document
writeLines(markdown_table, "model_summary.doc")
```

```
## Error in writeLines(markdown_table, "model_summary.doc"): can only write character objects
```

### The End