

Rcode for GV900 Homework 3

GV900 Political Explanation HW3

My student ID: 2107091

04-02-2022

Firstly, it will be advisable to clean the console and the environment from previously used data and values by using the following functions:

```
rm(list=ls(all=TRUE))  
cat("\014")
```

1. INSTITUTIONAL PERFORMANCE AND CIVIC COMMUNITY (PUTNAM DATA)

1.1 Load packages

I load the required packages in order to enable the following code to draw graphs and analyse regressions by using the "library" function since I have already installed them (using the "install.packages" function):

```
library(ggplot2)
```

To create graphs and visualise data.

```
library(foreign)
```

```
## Warning: package 'foreign' was built under R version 4.1.2
```

To read different types of extensions.

```
library(stargazer)
```

```
##
```

```
## Please cite as:
```

```
## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

To produce LaTeX code, HTML code and ASCII text for well-formatted tables with regression analysis results.

For this, I will cite: "Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables."

```
library(carData)
```

```
library(effects)
```

```
## Warning: package 'effects' was built under R version 4.1.2
```

```
## lattice theme set by effectsTheme()
```

```
## See ?effectsTheme for details.
```

To construct an "eff" object for a term in a regression that models a response as a linear function of main effects and interactions of factors and covariates. However, I needed to load also the "carData" packages since required to correctly run the gridExtra package.

(sources: <https://www.educba.com/list-of-r-packages/>;

<https://www.rdocumentation.org/packages/stargazer/versions/5.2.2/topics/stargazer>

azer; <https://www.rdocumentation.org/packages/effects/versions/4.2-0/topics/effect>).

1.2 Load the data set

Now, I can import the data set in my console by using the "read.csv" function to read the Comma Separated Values file. Moreover, I will use the "paste0" function to concatenate all elements without a separator (source: <https://r-lang.com/paste0-function-in-r-with-example/>). The path to finding the data is stored in an object called "myPath":

```
put <- read.csv(paste0(myPath, "putnam.csv"))
```

1.3 Regression models

(a) DV Institutional Performance, IDV Civic Community Index

To regress the institutional performance on civic community, I will use the "lm" function since it is a linear model regression. In the argument, I will put first the dependent variable (InstPerform) and then the independent variable (CivicCommunity). Moreover, I will store the regression in an object named "reg_a1". Thus:

```
reg_a1 <- lm(InstPerform ~ CivicCommunity, data = put)
```

(b) Additive model for North-South

First, I will create a new dummy variable called "North". I will add the variable North to the table, and I will use the "ifelse" function to automatically assign the value of 1 if it is North and 0 if it is South. Indeed, with "ifelse", I will code first the test to see if an object can be coerced to logical mode (true or false based if has same shape or different shape), and it will give different values based on the results of the test, and then I will put it as a factor (source: <https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/ifelse>)

```
put$North <- ifelse(put$NorthSouth == "South", 0, 1)
put$North <- as.factor(put$North)
```

To estimate the additive regression model, I will use the "lm" function and add the second variable "North" in the equation. Further, I will store it as an object called "add_reg_b1":

```
add_reg_b1 <- lm(InstPerform ~ CivicCommunity + North, data = put)
```

(c) Interactive model

To estimate the interactive regression model, I will use the "lm" function and insert the control between CC and NS. Further, I will store it as an object called "int_reg_c1":

```
int_reg_c1 <- lm(InstPerform ~ CivicCommunity + North + CivicCommunity:North,
data = put)
```

Finally, I will use the "stargazer" function to produce a regression table. Moreover, I will put "text" as the "type" argument. To display also the previous regression models, I will add them inside the brackets to show all the results in a unique table. Hence:

```
stargazer(reg_a1, add_reg_b1,int_reg_c1, type = "text")
```

```
##
##
=====
=====
##                                     Dependent variable:
##                                     -----
##                                     (1)          InstPerform
##                                     (2)
##                                     -----
##
## CivicCommunity          0.567***          0.571**
0.540*
##                          (0.066)          (0.215)
(0.269)
##
## North1                  -0.048
-1.194
##                          (2.678)
(6.396)
##
## CivicCommunity:North1
0.094
##
## (0.472)
##
## Constant                2.711***          2.698**
2.828**
##                          (0.844)          (1.121)
(1.326)
##
## -----
##
## Observations            20                20
```

```

20
## R2                0.806                0.806
0.807
## Adjusted R2       0.796                0.784
0.771
## Residual Std. Error    1.789 (df = 18)    1.841 (df = 17)
1.895 (df = 16)
## F Statistic          74.967*** (df = 1; 18) 35.402*** (df = 2; 17)
22.281*** (df = 3; 16)
##
=====
=====
## Note:                *p<0.1;
**p<0.05; ***p<0.01

```

1.4 Model fit statistics

First of all, checking the p-value for each model, we notice that only Model 1a and 2a are statistically significant at 95%, while Model 3a is only statistically significant at 90%. Besides, Model 1a has a lower p-value than Model 2a, giving more assurances in rejecting our null hypothesis.

Secondly, checking the adjusted R2, the best model is Model 1a since it has the larger adjusted R2. In other words, Model 1a explain more of the variations of the dependent variables caused by the independent variable. However, the difference between the three models is smallish, especially regarding R2 (source: [https://www.tutorialspoint.com/statistics/adjusted_r_squared.htm#:~:text=Adjusted%20R-squared%20adjusts%20the%20statistic%20based%20on%20the,for%20the%20number%20of%20terms%20in%20a%20model](https://www.tutorialspoint.com/statistics/adjusted_r_squared.htm#:~:text=Adjusted%20R-squared%20adjusts%20the%20statistic%20based%20on%20the,for%20the%20number%20of%20terms%20in%20a%20model;); <https://www.theanalysisfactor.com/assessing-the-fit-of-regression-models/>; <https://www.wallstreetmojo.com/adjusted-r-squared/>)

Thirdly, we checked the Residual Standard Error, which measures the standard deviation of the residuals in a regression. Also, in this case, the lowest RSE is in Model 1a, showing that the linear regression better fits the dataset compared to the other models. (source: <https://www.statology.org/how-to-interpret-residual-standard-error/>)

Therefore, among the three models, I choose the linear regression model (Model 1a) because it fits better the dataset according to the Adjusted R2 and the RSE.

1.5 Implied regression lines

```

# (a) Model(1a):
# In general, the regression equation is:
# DV = a*IDV + c

```

```

#  $IP = c + a*CC$ 

# The implied regression line:
#  $IP = 2.711 + 0.567 * CC$ 

# (b) Model(2a):
# In general, the regression equation is:
#  $DV = a*IDV1 + b*IDV2 + c$ 
#  $IP = (b*N + c) + a*CC$ 

# The implied regression line for Southern (0) regions:
#  $IP = 2.698 + 0.571 * CC$ 

# The implied regression line for Northern (1) regions:
#  $IP = 2.650 + 0.571 * CC$ 

# (c) Model(3a):
# In general, the regression equation is:
#  $DV = a*IDV1 + b*IDV2 + d*IDV1*IDV2 + c$ 
#  $IP = (b*N + c) + (a + d*N)*CC$ 

# The implied regression line for Southern (0) regions:
#  $IP = 2.828 + 0.540 * CC$ 

# The implied regression line for Northern (1) regions:
#  $IP = 1.634 + 0.634 * CC$ 

# I collected the previous coefficients only by substituting the general
# equation variables with the actual number represented in the regression.
# However, more formally, I could have also created a new regression model that
# would have displayed the numbers. I will do it in comments to not interfere
# with the other code, and I will use it to check my previous answers:

# checkN_int_c1 <- lm(InstPerform ~ CivicCommunity, data = put[put$NorthSouth
# == "North",])
# checkS_int_c1 <- lm(InstPerform ~ CivicCommunity, data = put[put$NorthSouth
# == "South",])
# stargazer(checkN_int_c1, checkS_int_c1, type = "text")

```

1.6 Interactive model?

```

# Based on previous models and answers 1.4 and 1.5, an interaction model is
# not needed in this case. Indeed, the value of interaction
# ("CivicCommunity:North1") is not statistically significant, demonstrating no
# strong correlation between the two independent variables inside this model.

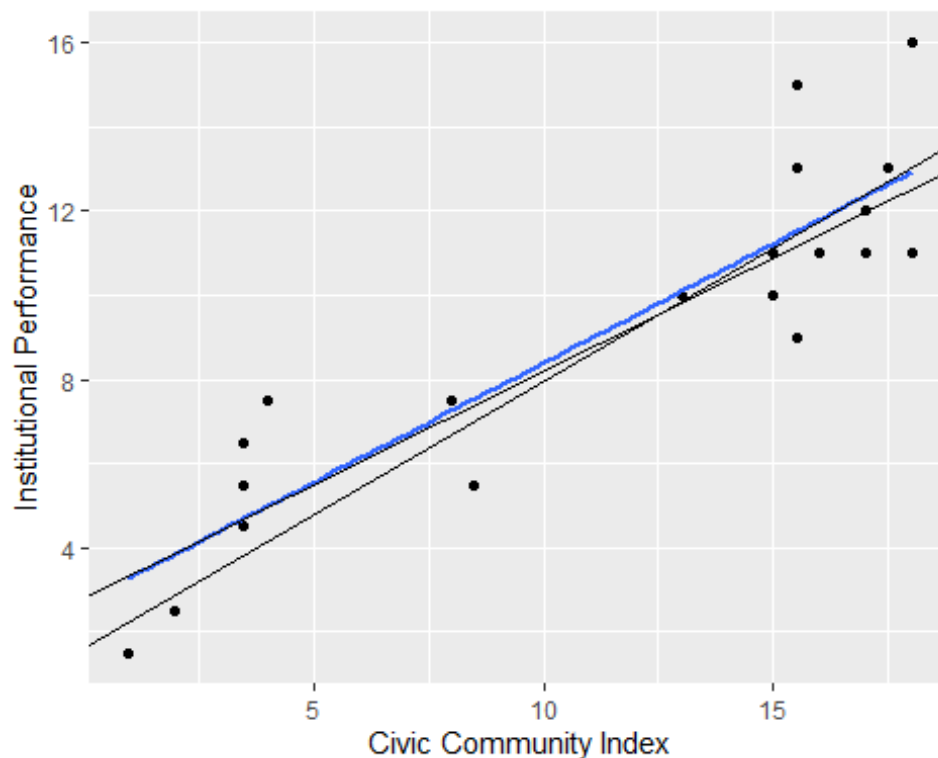
```

Besides, the differences in slope and intercept of the two regression lines in the interactive model are almost similar. Therefore, in this case, an interactive model is not needed.

(source: <https://quantifyinghealth.com/why-and-when-to-include-interactions-in-a-regression-model/>)

However, to check for my statement, I also created a proof-graph to display my findings. The data visualisation further proved my assertion that an interactive model was unnecessary since there is no significant difference between the Southern and Northern regions results and no significant difference with the linear regression model.

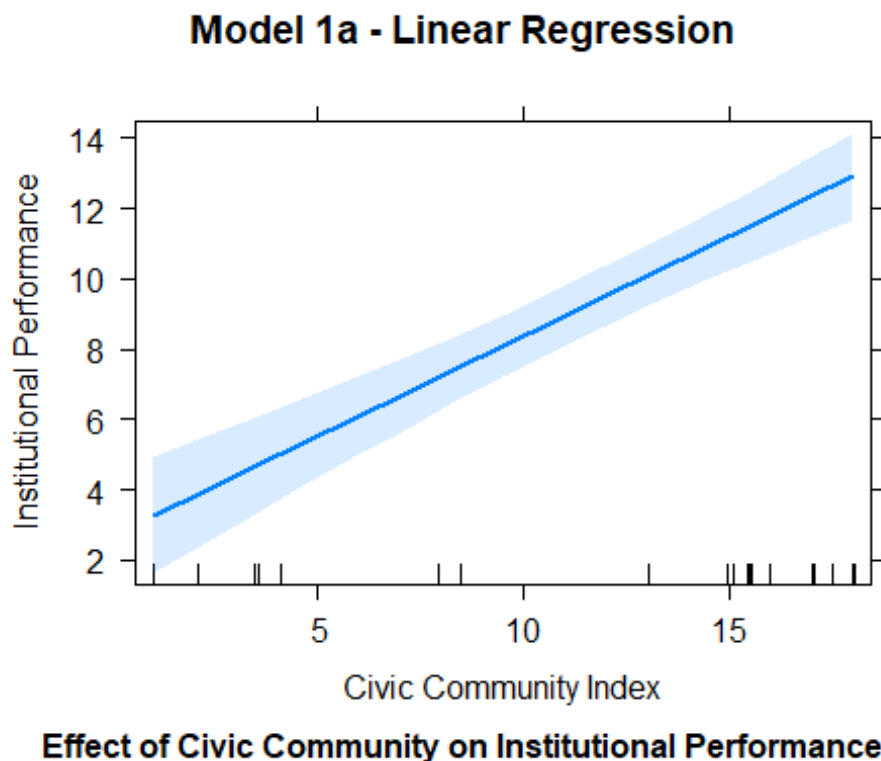
```
prof_1 <- ggplot(put, aes(x = CivicCommunity, y = InstPerform)) +  
  geom_point()  
prof_1 <- prof_1 + ylab("Institutional Performance") + xlab("Civic Community  
Index")  
prof_1 <- prof_1 + geom_smooth(method = lm, alpha = 0)  
prof_1 <- prof_1 + geom_abline(intercept = 1.634, slope = 0.634) # Line for  
Northern regions  
prof_1 <- prof_1 + geom_abline(intercept = 2.828, slope = 0.540) # Line for  
Southern Regions  
print(prof_1)  
## `geom_smooth()` using formula 'y ~ x'
```



1.7 Graph Model(1)

To draw a graph of the regression model, I will use the plot() function for the first linear regression model and the effect() function to insert the regression data. Besides, I will improve the design of the graph by adding some labels using various arguments, hence:

```
plot(effect(term = "CivicCommunity", mod = reg_a1),  
      main = "Model 1a - Linear Regression",  
      sub = "Effect of Civic Community on Institutional Performance",  
      xlab = "Civic Community Index",  
      ylab = "Institutional Performance")
```



1.8 Graphs Model(2)

To graphically display the relationship between the civic community and institutional performance, based on region position, I will use the "effect" function, inserting the additive regression model (add_reg_b1) as the "mod" argument.

Firstly, I will use the "effect" function to separate the regression results based on geography. For Northern regions, I will store the result in an object called "n_eff_b1", hence:

```
n_eff_b1 <- effect(term = "CivicCommunity",  
                   mod = add_reg_b1,
```



```
given.values = c("North1" = 1))
```

For Southern regions, I will store the result in an object called "s_eff_b1", thus:

```
s_eff_b1 <- effect(term = "CivicCommunity",  
                  mod = add_reg_b1,  
                  given.values = c("North1" = 0))
```

Then, I will translate the result into graphs using the "plot" function. I will store the plot for Northern regions in an object called "n_gr_b1". More than asked in this assignment, in order to increase the comprehension of the data displayed in the graph, I will use the "main", "xlab", and "ylab" arguments to improve them. Further, I will use the "ylim" function to set the two graphs at an equal level. Therefore, I can graphically display the difference of the two intercept measures between the two regions. Indeed, "ylim" denotes the y limits (y1, y2) of the plot, and I chose -5 and 20 based on the visualisation of the plots without the "ylim" argument (source: <https://thepracticalr.wordpress.com/tag/ylim/>). Similarly, I will also implement the "xlim" to display the graph with identical y-axis and x-axis ranges. I will insert the x-axis range as c(0, 20) since all the Civic Community Index values are inside this range. Hence:

```
n_gr_b1 <- plot(n_eff_b1,  
               main = "Northern Regions",  
               sub = "Effect of Civic Community  
on Institutional Performance (North)",  
               xlab = "Civic Community Index",  
               ylab = "Institutional Performance",  
               ylim = c(-5, 20),  
               xlim = c(0, 20))
```

While I will store the graph for Southern regions in an object called "s_gr_b1", and I will do the same improvements, thus:

```
s_gr_b1 <- plot(s_eff_b1,  
               main = "Southern Regions",  
               sub = "Effect of Civic Community  
on Institutional Performance (South)",  
               xlab = "Civic Community Index",  
               ylab = "Institutional Performance",  
               ylim = c(-5, 20),  
               xlim = c(0, 20))
```

Then, I will use the "grid.arrange" function (after calling the gridExtra and dplyr packages) to display the two graphs together. Moreover, I will use the "ncol" option to represent them one alongside the other and not one above and below, thus:

```

library(gridExtra)
library(dplyr)

##
## Attaching package: 'dplyr'

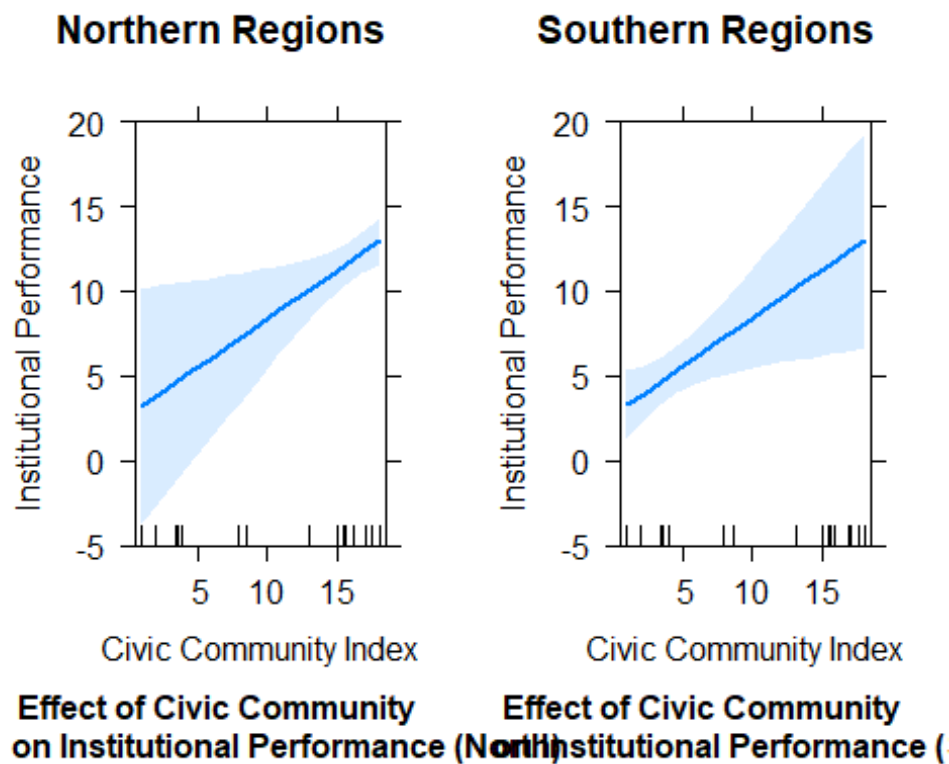
## The following object is masked from 'package:gridExtra':
##
##      combine

## The following objects are masked from 'package:stats':
##
##      filter, lag

## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union

grid.arrange(n_gr_b1, s_gr_b1, ncol = 2)

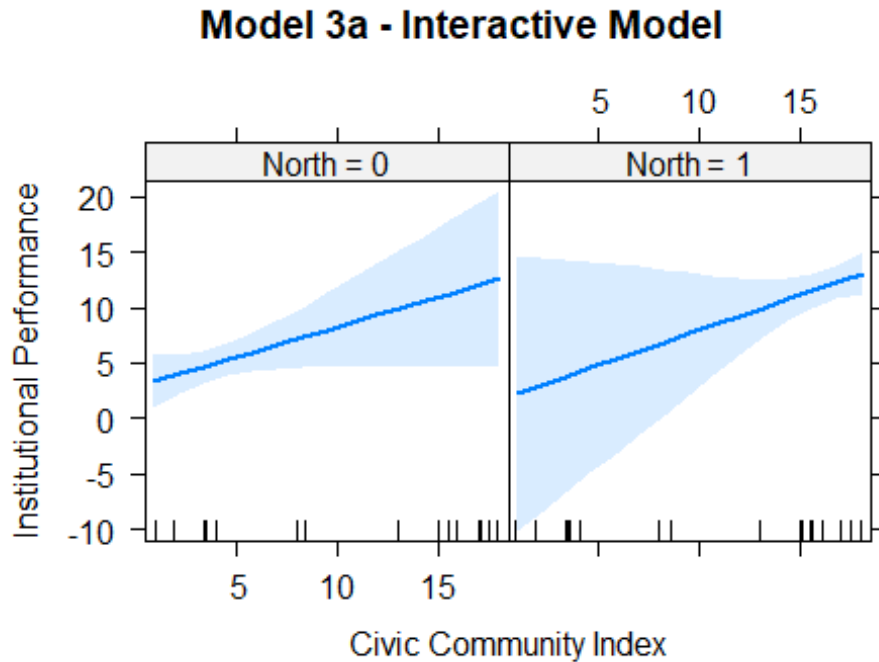
```



1.9 Graphs Model (3)

To graphically display the relationship between the civic community and institutional performance, based on region position, I will use the "effect" function, inserting the interactive regression model (int_reg_c1) as the "mod" argument. Further, I insert the interaction between the IDV and the control variable as "term" argument.

```
plot(effect(term = "CivicCommunity:North", mod = int_reg_c1),
      main = "Model 3a - Interactive Model",
      sub = "Effect of Civic Community on Institutional Performance
North on left, South on right",
      xlab = "Civic Community Index",
      ylab = "Institutional Performance")
```



1.10 Spuriousness

Spuriousness refers to a situation in which long-lasting trends in variables produce false evidence of a statistical relationship between those variables when none truly exists (Kellstedt and Whitten, 2018). For this reason, the relationship between the civic community and institutional performance is not spurious since testing region position they are still correlated. However, inserting regional differences decreased the p-value of the correlation, thus decreasing the probability at which we can reject the null hypothesis (no causal relationships between the two variables). However, the regional position is never statistically significant.

Therefore, the relationship is not spurious, although other control variables could interfere with the results.

2. INSTITUTIONAL PERFORMANCE AND ECONOMIC MODERNISATION (PUTNAM DATA)

2.1 Regression models

*# (a) DV Institutional Performance, IDV Economic Modernisation
To regress the institutional performance on economic modernisation, I will use the "lm" function since it is a linear model regression. In the argument, I will put first the dependent variable (InstPerform) and then the independent variable (EconModern). Moreover, I will store the regression in an object named "reg_a1". Thus:*

```
reg_a2 <- lm(InstPerform ~ EconModern, data = put)
```

*# (b) Additive model North-South
To estimate an additive regression model, I will use the "lm" function. Further, I will store it as an object called "add_reg_b1":*

```
add_reg_b2 <- lm(InstPerform ~ EconModern + North, data = put)
```

*# (c) Interactive model
To estimate an interactive regression model, I will use the "lm" function. Further, I will store it as an object called "int_reg_c1":*

```
int_reg_c2 <- lm(InstPerform ~ EconModern + North + EconModern:North, data = put)
```

Finally, I will use the "stargazer" function to produce a regression table. Moreover, I will put "text" as the "type" argument. To display also the previous regression models, I will add them inside the brackets to show all the results in a unique table. Hence:

```
stargazer(reg_a2, add_reg_b2, int_reg_c2, type = "text")
```

```
##
##
=====
=====
##                               Dependent variable:
##                               -----
##                               (1)          (2)
##                               -----
##                               -----
## EconModern                    0.589***          -0.019
##                               0.015
```

```

##                                (0.120)                (0.220)
(0.387)
##
## North1                                6.884***
7.306
##                                (2.229)
(4.506)
##
## EconModern:North1
-0.052
##
(0.478)
##
## Constant                3.011**                5.222***
5.051**
##                                (1.385)                (1.347)
(2.093)
##
## -----
-----
## Observations                20                20
20
## R2                0.572                0.726
0.726
## Adjusted R2                0.549                0.694
0.675
## Residual Std. Error    2.659 (df = 18)        2.190 (df = 17)        2.256
(df = 16)
## F Statistic            24.097*** (df = 1; 18)  22.531*** (df = 2; 17)
14.152*** (df = 3; 16)
##
=====
=====
## Note:                                *p<0.1;
**p<0.05; ***p<0.01

```

2.2 Model fit statistics

First of all, checking the p-value for each model, we notice that only Model 1b and 2b are statistically significant at 95%, while Model 3b is not statistically significant at all. However, Model 1b is significant at 99% for our IV (EM), while Model 2b is significant at 99% for the dummy regional variable (North1).

Secondly, checking the adjusted R2, the best model is Model 2b since it has the largest adjusted R2. In other words, Model 2b explain more of the variations of the dependent variables caused by the independent variable. Further, the difference with Model 1b is noteworthy (almost 15%), while with Model 3b is less significant (source:

[https://www.tutorialspoint.com/statistics/adjusted_r_squared.htm#:~:text=Adjusted%20R-squared%20adjusts%20the%20statistic%20based%20on%20the,for%20the%20number%20of%20terms%20in%20a%20model](https://www.tutorialspoint.com/statistics/adjusted_r_squared.htm#:~:text=Adjusted%20R-squared%20adjusts%20the%20statistic%20based%20on%20the,for%20the%20number%20of%20terms%20in%20a%20model;); <https://www.theanalysisfactor.com/assessing-the-fit-of-regression-models/>; <https://www.wallstreetmojo.com/adjusted-r-squared/>)

Thirdly, we checked the Residual Standard Error measuring the standard deviation of the residuals in a regression. Also, in this case, the lowest RSE appears in Model 2b, demonstrating that the additive regression better fits the dataset compared to the other models. (source: <https://www.statology.org/how-to-interpret-residual-standard-error/>)

Therefore, between the three models, I choose the additive regression model (Model 2b) because it fits better the dataset according to the Adjusted R² and the RSE.

2.3 Implied regression lines

(a) Model(1b):

In general, the regression equation is:

$DV = a \cdot IDV + c$

$IP = c + a \cdot EM$

The implied regression line:

$IP = 3.011 + 0.589 \cdot EM$

(b) Model(2b):

In general, the regression equation is:

$DV = a \cdot IDV1 + b \cdot IDV2 + c$

$IP = (b \cdot N + c) + a \cdot EM$

The implied regression line for Southern (0) regions:

$IP = 5.222 - 0.019 \cdot EM$

The implied regression line for Northern (1) regions:

$IP = 12.106 - 0.019 \cdot EM$

(c) Model(3b):

In general, the regression equation is:

$DV = a \cdot IDV1 + b \cdot IDV2 + d \cdot IDV1 \cdot IDV2 + c$

$IP = (b \cdot N + c) + (a + d \cdot N) \cdot EM$

The implied regression line for Southern (0) regions:

$IP = 5.051 + 0.015 \cdot EM$

```
# The implied regression line for Northern (1) regions:  
#  $IP = 12.357 - 0.037 * EM$ 
```

I collected the previous coefficients only by substituting the general interactive regression equation variables with the actual number represented in the regression. However, more formally, I could have also created a new regression model that would have displayed the numbers. I will do it in comments to not interfere with the other code, and I will use it to check my previous answers:

```
# checkN_int_c2 <- lm(InstPerform ~ EconModern, data = put[put$NorthSouth ==  
"North",])  
# checkS_int_c2 <- lm(InstPerform ~ EconModern, data = put[put$NorthSouth ==  
"South",])  
# stargazer(checkN_int_c2, checkS_int_c2, type = "text")
```

2.4 Interactive model?

Based on previous models and answers 2.2 and 2.3, an interactive model is not needed in this case. Indeed, the value of the interaction ("EconModern:North1") is not statistically significant, demonstrating no strong correlation between the two independent variables inside this model. However, the differences in slope and intercept of the two regression lines in the interactive model are quite significant a relationship demonstrated also by the additive model. Therefore, in this case, an interactive model is not needed since an additive model will display the difference in intercept while the difference in the slope is not significant. Further, there is no theoretical explanation that can confirm or request the use of the interactive model. Finally, the adjusted R2 is better in the additive model, giving more explanatory power to the additive model rather than the interactive one.

(source: <https://quantifyinghealth.com/why-and-when-to-include-interactions-in-a-regression-model/>)

However, to check for my statement, I also created a proof-graph to display my findings. The data visualisation further proved my assertion that an interactive model (green lines) was not needed since there is not a significant difference between the results of the Southern and Northern regions with the additive model (red lines) but there is a significant difference with the linear regression model.

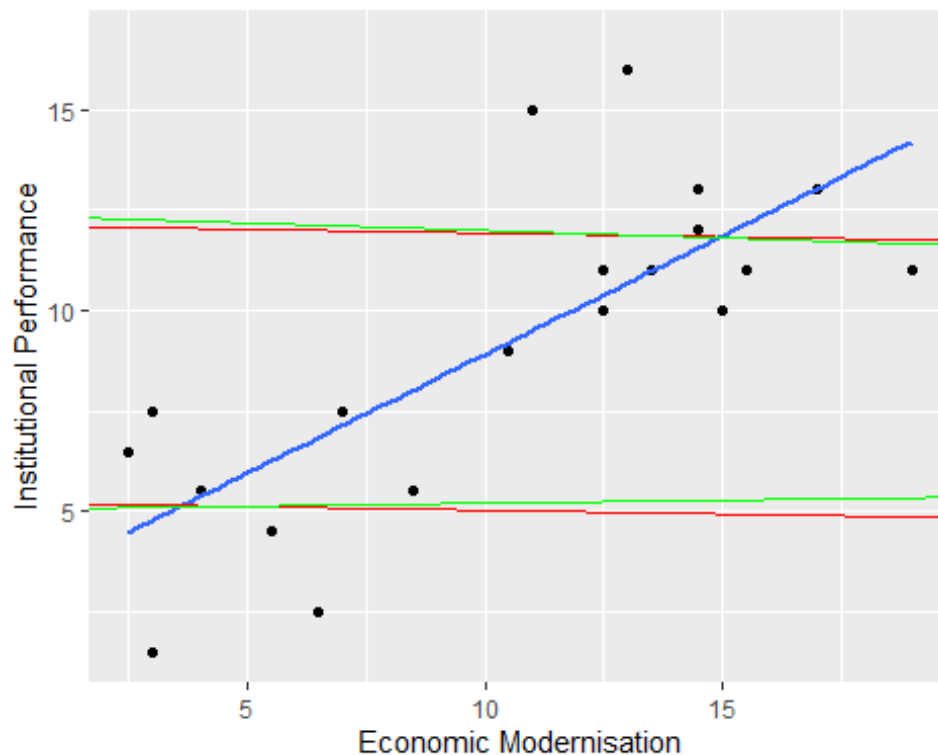
```
prof_2 <- ggplot(put, aes(x = EconModern, y = InstPerform)) + geom_point()  
prof_2 <- prof_2 + ylab("Institutional Performance") + xlab("Economic  
Modernisation")  
prof_2 <- prof_2 + geom_smooth(method = lm, alpha = 0)  
prof_2 <- prof_2 + geom_abline(intercept = 12.106, slope = -0.019, colour =  
"red") # Line for Northern Regions (additive)  
prof_2 <- prof_2 + geom_abline(intercept = 12.358, slope = -0.037, colour =
```

```

"green") # Line for Northern Regions (interactive)
prof_2 <- prof_2 + geom_abline(intercept = 5.222, slope = -0.019, colour =
"red") # Line fo Southern Regions (additive)
prof_2 <- prof_2 + geom_abline(intercept = 5.051, slope = 0.015, colour =
"green") # Line for Southern Regions (interactive)
print(prof_2)

## `geom_smooth()` using formula 'y ~ x'

```



As can be seen by the graph, there is no major difference between the two models in terms of slope and intercepts. Further, there is no sign of theoretical background to confirm the need for an interactive model. Thus, I conclude that the interactive model is not needed.

2.5 Graph model(1)

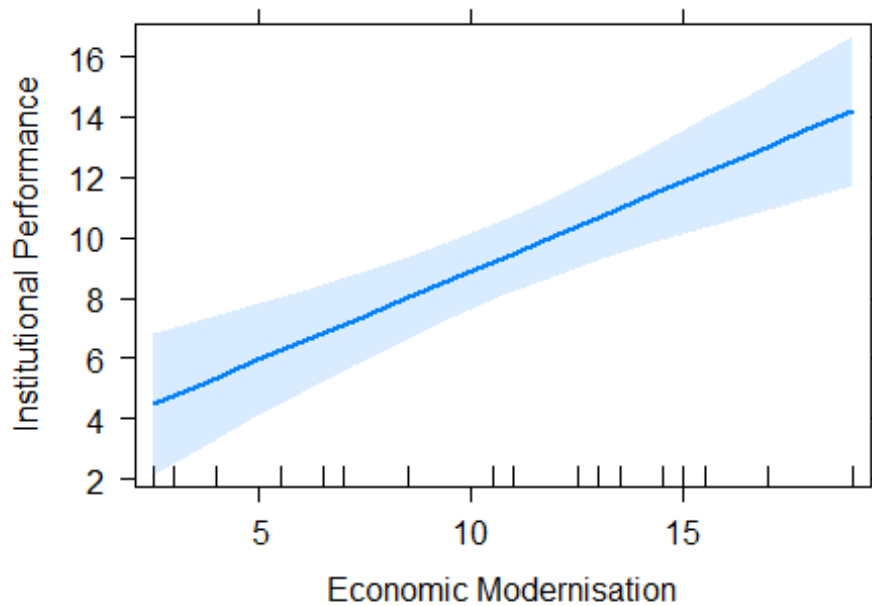
I will use the `plot()` function to draw the graph of the first Linear regression model (Model 1b) and the `effect()` function to insert the regression data. Besides, I will improve the design of the plot by adding some labels using various arguments, hence:

```

plot(effect(term = "EconModern", mod = reg_a2),
      main = "Model 1b - Linear Regression",
      sub = "Effect of Economic Modernisation on Institutional Performance",
      xlab = "Economic Modernisation",
      ylab = "Institutional Performance")

```


Model 1b - Linear Regression



Effect of Economic Modernisation on Institutional Performance

2.6 Graphs model(2)

To graphically display the relationship between economic modernisation and institutional performance, based on region position, I will use the "effect" function, inserting the additive regression model (add_reg_b2) as the "mod" argument.

Firstly, I will use the "effect" function to separate the regression results based on geography. For Northern regions, I will store the result in an object called "n_eff_b2", hence:

```
n_eff_b2 <- effect(term = "EconModern",  
                  mod = add_reg_b2,  
                  given.values = c("North1" = 1))
```

For Southern regions, I will store the result in an object called "s_eff_b2", thus:

```
s_eff_b2 <- effect(term = "EconModern",  
                  mod = add_reg_b2,  
                  given.values = c("North1" = 0))
```

Then, I will translate the result into graphs using the "plot" function. I will store the plot for Northern regions in an object called "n_gr_b2". More than asked in this assignment, in order to increase the comprehension of the

data displayed in the graph, I will use the "main", "xlab", and "ylab" arguments to improve them. Further, I will use the "ylim" function to set the two graphs at an equal level. Therefore, I can graphically display the difference of the two intercept measures between the two regions. Indeed, "ylim" denotes the y limits (y1, y2) of the plot, and I chose -5 and 20 based on the visualisation of the plots without the "ylim" argument (source: <https://thepracticalr.wordpress.com/tag/ylim/>). Similarly, I will also implement the "xlim" to display the graph with identical y-axis and x-axis ranges. I will insert the x-axis range as c(0, 20) since all the Economic Modernisation values are inside this range. Hence:

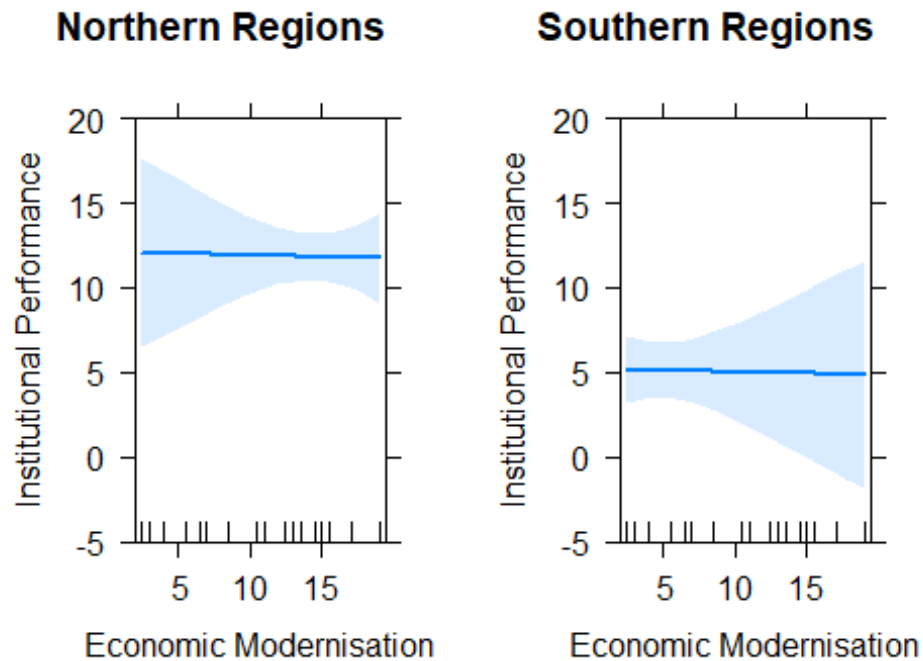
```
n_gr_b2 <- plot(n_eff_b2,
               main = "Northern Regions",
               sub = "Effect of Economic Modernisation
on Institutional Performance (North)",
               xlab = "Economic Modernisation",
               ylab = "Institutional Performance",
               ylim = c(-5, 20),
               xlim = c(0, 20))
```

While I will store the graph for Southern regions in an object called "s_gr_b1", and I will do the same improvements, thus:

```
s_gr_b2 <- plot(s_eff_b2,
               main = "Southern Regions",
               sub = "Effect of Economic Modernisation
on Institutional Performance (South)",
               xlab = "Economic Modernisation",
               ylab = "Institutional Performance",
               ylim = c(-5, 20),
               xlim = c(0, 20))
```

Then, I will use the "grid.arrange" function to display the two graphs together. Moreover, I will use the "ncol" option to represent them one alongside the other and not one above and below, thus:

```
grid.arrange(n_gr_b2, s_gr_b2, ncol = 2)
```



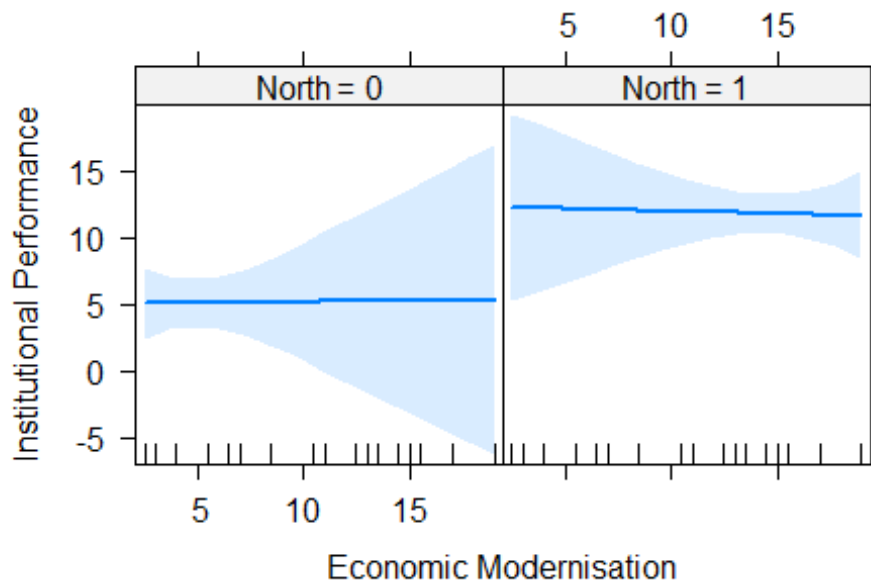
Effect of Economic Modernisation on Institutional Performance (North) **Effect of Economic Modernisation on Institutional Performance (South)**

2.7 Graphs model(3)

To graphically display the relationship between the civic community and institutional performance, based on region position, I will use the "effect" function, inserting the interactive regression model (int_reg_c1) as the "mod" argument. Further, I insert the interaction between the IDV and the control variable as "term" argument.

```
plot(effect(term = "EconModern:North", mod = int_reg_c2),
     main = "Model 3b - Interactive Regression",
     sub = "Effect of Economic Modernisation on Institutional Performance
           (North on left, South on right)",
     xlab = "Economic Modernisation",
     ylab = "Institutional Performance")
```

Model 3b - Interactive Regression



**Effect of Economic Modernisation on Institutional Performance
(North on left, South on right)**

2.8 Spuriousness

The relationship between economic modernisation and institutional performance is spurious. Indeed, after testing for region position, the two variables are not correlated anymore. In fact, the inclusion of regional differences drastically reduced the p -value (the probability at which we can reject the null hypothesis). Thus, there are no causal relationships between the two studied variables after controlling for region location. Indeed, the regional position is statistically significant in the additive regression, and also in the interactive model, the regression coefficient is almost zero.

Therefore, the relationship between economic modernisation and institutional performance is spurious after controlling for region location.