

**HOCHSCHULE RHEIN-WAAL**  
**RHINE-WAAL UNIVERSITY OF APPLIED SCIENCES**  
**FACULTY OF SOCIETY AND ECONOMICS**

**COUNTRY DEVELOPMENT REPORT**  
**of Singapore, Slovenia, Saudi Arabia**

**M-SD 2 7105 Quantitative Research Methods**

By

**Mulisa Kevine Umwiza**  
**32564**

Date: 2023-06-01

# Country Development Report of: Singapore, Slovenia, Saudi Arabia

## 1. Introduction

Obtain the data for all, in later problems required, variables for your three selected countries for the years 1980-2020 from the World Development Indicators-Database of the World Bank.

```
library(WDI)

MyIndicators <- c ("NY.GDP.MKTP.KD.ZG", "SP.POP.TOTL", "BX.KLT.DINV.CD.WD", "NE.TRD.GNFS.ZS",
                  "DT.ODA.ALLD.CD", "FP.CPI.TOTL.ZG", "SE.XPD.TOTL.GB.ZS", "LP.LPI.INFR.XQ",
                  "SH.XPD.GHED.GD.ZS" ) # <-- add all required indicators here!

??WDI

population <- WDIsearch("population")
FDI<- WDIsearch("FDI")
Trade<- WDIsearch("Trade")
ODA<- WDIsearch("ODA")
Inflation<- WDIsearch("Inflation")
Education<- WDIsearch("education")
Infrastructure<- WDIsearch("Infrastructure")
Health<- WDIsearch("Health expenditure")
```

As you will find missings (N/As) in your data, we impose spline interpolation per country and variable for obtaining a full dataset.

- a) Describe the obtained dataset by tables of summary statistics for each country separately, including the mean, quartiles and standard deviation of all variables required below for the given/available timeframe for all countries separately.

##Renamed the indicators

```
WDIfull_renamed <- WDIfull %>%
  rename(GDP = NY.GDP.MKTP.KD.ZG,
         Population = SP.POP.TOTL,
         FDI = BX.KLT.DINV.CD.WD,
         Trade = NE.TRD.GNFS.ZS,
         ODA = DT.ODA.ALLD.CD,
```

```

    Inflation = FP.CPI.TOTL.ZG,
    Education = SE.XPD.TOTL.GB.ZS,
    Infrastructure = LP.LPI.INFR.XQ,
    Health = SH.XPD.GHED.GD.ZS)
options(scipen = 999)

```

## Grouped and summarized data

```
library("stargazer")
```

##

## Please cite as:

## Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.

## R package version 5.2.3. <https://CRAN.R-project.org/package=stargazer>

```
WDISG <- WDIfull_renamed %>%
```

```
  filter(iso2c == "SG")
```

```
WDISG <- as.data.frame(WDISG)
```

```
stargazer(WDISG, type = "text", iqr = T, title = "Descriptive statistics Singapore")
```

##

## Descriptive statistics Singapore

## =====

## Statistic	N	Mean	St. Dev.	Min	Pctl(25)
--------------	---	------	----------	-----	----------

## -----

## year	41	2,000.000	11.979	1,980	1,990
---------	----	-----------	--------	-------	-------

## GDP	41	6.098	4.240	-4.143	3.661
--------	----	-------	-------	--------	-------

## Population	41	4,048,308.000	1,084,986.000	2,413,945	3,047,132
---------------	----	---------------	---------------	-----------	-----------

## FDI	41	27,484,725,959.000	30,527,465,959.000	1,046,746,813.000	3,654,794,589.000
--------	----	--------------------	--------------------	-------------------	-------------------

## Trade	41	350.567	36.354	294.826	323.864
----------	----	---------	--------	---------	---------

## ODA	41	12,748,198.000	16,421,233.000	-1,740,000.000	1,620,000.000
--------	----	----------------	----------------	----------------	---------------

## Inflation	41	1.915	2.214	-1.386	0.508
--------------	----	-------	-------	--------	-------

## Education	41	22.738	3.672	11.912	21.095
--------------	----	--------	-------	--------	--------

## Infrastructure	41	4.215	0.068	4.060	4.159
-------------------	----	-------	-------	-------	-------

## Health	41	1.291	0.474	0.882	1.001
-----------	----	-------	-------	-------	-------

```
## -----
```

```
WDISI <- WDIfull_renamed %>%  
  filter(iso2c == "SI")  
WDISI <- as.data.frame(WDISI)  
stargazer(WDISI, type = "text", iqr = T, title = "Descriptive statistics Slovenia")
```

```
##
```

```
## Descriptive statistics Slovenia
```

```
## =====
```

```
## Statistic      N      Mean      St. Dev.      Min      Pctl(25)      Pctl(75)
```

```
## -----
```

```
## year           41    2,000.000     11.979      1,980      1,990      2,010
```

```
## GDP            41       3.033       2.666     -7.548      3.146      4.350
```

```
## Population     41  2,003,924.000  49,349.270  1,901,315  1,988,628  2,048,924
```

```
## FDI            41 543,507,587.000 613,246,078.000 -346,971,373.000 116,900,000.000 875,946,373
```

```
## Trade          41    115.544     22.328     92.331     96.261    137.111
```

```
## ODA            41 60,982,698.000 29,858,259.000 7,140,000.000 43,078,368.000 73,860,998
```

```
## Inflation      41     73.986     215.837     -0.526      1.803     42.620
```

```
## Education      41     11.501       0.697      8.970     11.220     12.070
```

```
## Infrastructure 41       3.040       0.304      2.497      2.733      3.270
```

```
## Health         41       5.828       0.309      5.276      5.613      6.060
```

```
## -----
```

```
WDISA <- WDIfull_renamed %>%  
  filter(iso2c == "SA")  
WDISA <- as.data.frame(WDISA)  
stargazer(WDISA, type = "text", iqr = T, title = "Descriptive statistics Saudi Arabia")
```

```
##
```

```
## Descriptive statistics Saudi Arabia
```

```
## =====
```

```
## Statistic      N      Mean      St. Dev.      Min      Pctl(25)
```

```
## -----
```

```
## year           41    2,000.000     11.979      1,980      1,990
```

```
## GDP            41       1.989       7.457     -20.730     -1.211
```

## Population	41	22,411,643.000	7,888,233.000	10,171,710	16,004,763
## FDI	41	6,433,628,272.000	10,167,255,478.000	-3,192,312,800.000	-328,000,000.000
## Trade	41	73.545	11.571	52.077	64.593
## ODA	41	10,932,210.000	36,741,461.000	-131,070,007.000	13,610,000.000
## Inflation	41	1.479	2.669	-3.203	-0.077
## Education	41	19.286	4.089	11.735	15.252
## Infrastructure	41	3.220	0.095	2.950	3.188
## Health	41	2.938	0.651	1.889	2.507
## -----					

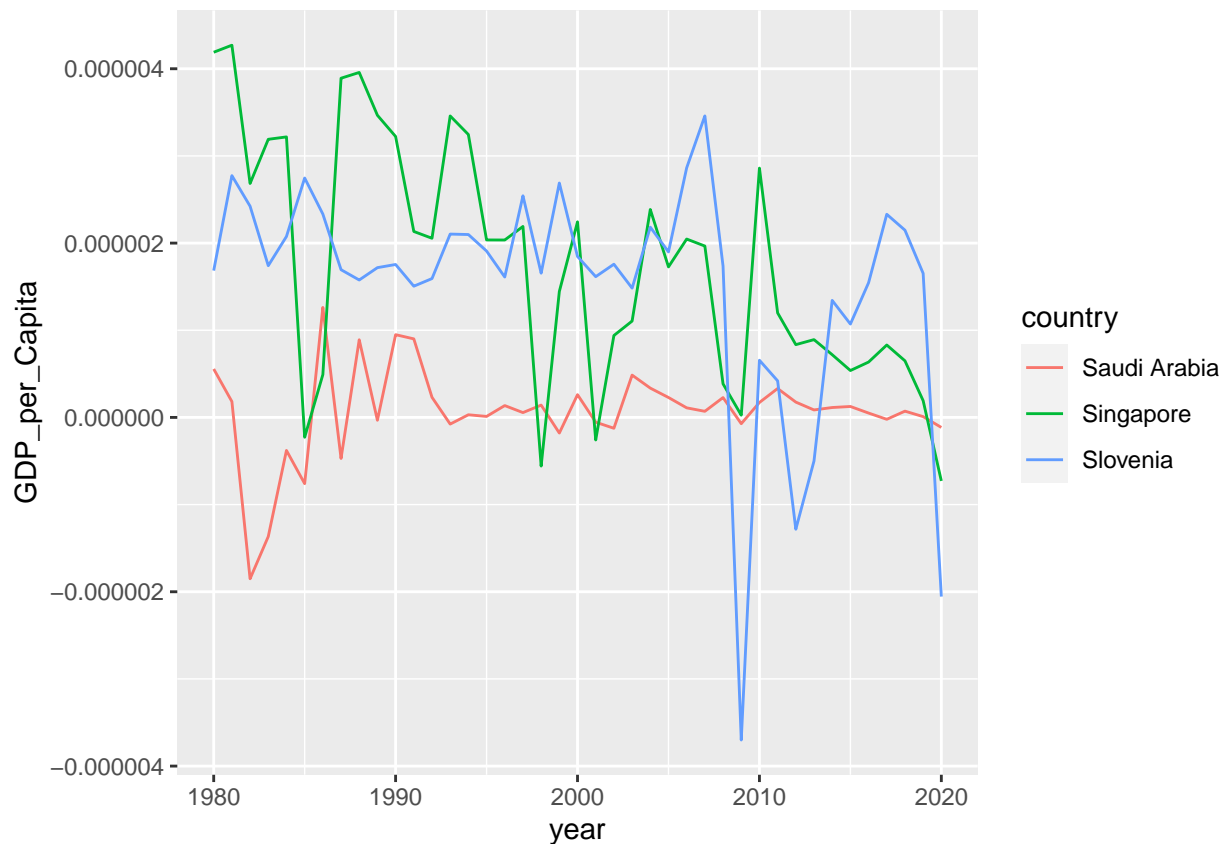
b) Calculate the GDP per capita (PPP) for all three countries and show its development from 1980 - 2020 in a single graph.

##Calculated GDP per capita

```
data <- WDIfull_renamed
filtered_data <- data %>%
  filter(country %in% c("Singapore", "Slovenia", "Saudi Arabia"),
         year >= 1980 & year <= 2020) %>%
  mutate(GDP_per_Capita = GDP / Population)

library(ggplot2)

ggplot(filtered_data, aes(x = year, y = GDP_per_Capita, color = country))+
  geom_line()
```



## 2. Determinants of Economic Development

From the literature we know that economic growth of a country is correlated with size of the labour force, investment, and involvement of the country in international trade. Moreover, it is often disputed that development aid makes a significant contribution. Thus, as model no. 1 (M1) run a pooled multiple linear regression (OLS) on the following equation:

$$(M1) \quad GDP = \alpha + \beta_1 Population + \beta_2 FDI + \beta_3 Trade + \beta_4 ODA + \beta_5 Inflation + \varepsilon$$

Find suitable variables that describe GDP, Population size, FDI (=Foreign Direct Investment), Trade (e.g. the degree of openness), ODA (Official Development Assistance). (M2) As model no. 2 estimate a log-linear version of M1. (M3) And a a log-log version of M1.

- a) Present your regression results for all 3 models in a single table. Discuss which measure you would consult in order to find the model with the “best fit”. Explain what is meant by “best fit”.

```
library(stargazer)
M1<- lm (GDP ~ Population + FDI + ODA + Trade + Inflation, data = filtered_data)
```

```
M2 <- lm (log(GDP) ~ Population + FDI + ODA + Trade + Inflation, data = filtered_data)
M3 <- lm (log(GDP) ~ log(Population) + log(FDI) + log (ODA) + log(Trade), data = filtered_data)
stargazer(M1, M2, M3, type = "text")
```

##	=====		
##	Dependent variable:		
##	-----		
##	GDP	log(GDP)	
##	(1)	(2)	(3)
##	-----		
##	Population	0.00000	-0.000
##		(0.00000)	(0.000)
##			
##	FDI	-0.000*	-0.000
##		(0.000)	(0.000)
##			
##	ODA	-0.000	-0.000
##		(0.00000)	(0.000)
##			
##	Trade	0.019***	0.002**
##		(0.006)	(0.001)
##			
##	Inflation	0.002	-0.0001
##		(0.004)	(0.001)
##			
##	log(Population)		0.244
##			(0.179)
##			
##	log(FDI)		-0.129
##			(0.091)
##			
##	log(ODA)		0.076
##			(0.089)

```
##
## log(Trade)                                0.711***
##                                           (0.262)
##
## Constant                0.693                1.177***                -4.391
##                        (1.844)                (0.305)                (3.563)
##
## -----
## Observations                123                102                88
## R2                        0.132                0.099                0.107
## Adjusted R2                0.095                0.052                0.064
## Residual Std. Error    5.169 (df = 117)    0.811 (df = 96)    0.778 (df = 83)
## F Statistic            3.564*** (df = 5; 117) 2.114* (df = 5; 96) 2.480* (df = 4; 83)
## =====
## Note:                                *p<0.1; **p<0.05; ***p<0.01
```

<R-squared and/or Adjusted R-squared can be used to find the model with the best fit. R-squared calculates the percentage of the dependent variable's variance that can be explained by the independent variables in the regression model. When comparing models with different numbers of independent variables, it is best to use adjusted R-squared, which takes into account the number of independent variables in the model by penalizing the addition of extra independent variables that do not significantly contribute to explaining the variation in the dependent variable(Montgomery et al., 2013). Since our models have the same number of independent variables, we will use R-squared to measure the goodness of fit. The better the model is in describing the variation in the dependent variable explained by the independent variable, the higher the value for R-squared (Hill et al., 2018).>

<M1 has r-squared value of 0.132 which means that the model explains 13.2% of the variation in the GDP caused by the independent variables which are Population, Trade, FDI, ODA, and Inflation. M2 has r-squared value of 0.099 which means that the model explains 9.9% of the variation in the log(GDP) caused by the independent variables which are Population, Trade, FDI, ODA, and Inflation. M3 has r-squared value of 0.107 which means that the model explains 10.7% of the variation in the GDP caused by the independent variables which are Population, Trade, FDI, ODA, and Inflation. Comparing



the r-squared values, we can conclude that M1 is the best fit because its R-squared value is the highest among other models which means that the model explains the higher variation in the GDP using the independent variables. However, it is important to mention that the R-squared values in the three models are noticeably low. This indicates that the independent variables used in the models only account for a small portion of the variation in the GDP or log(GDP). As a result, it raises the possibility that other variables or factors, not taken into account by the model, also contribute to the observed variation in the GDP or log(GDP).>

<‘best fit’ means finding a model that has minimal residuals meaning that the difference between observed values and predicted values should be small. Additionally, a best fit model is a model whose coefficients have statistical significance to indicate the effect they have on the dependent variable (Soetewey, 2021).>

- b) Calculate 95%-confidence intervals for all coefficients of model M2. (Save the results in ‘MyRESULTS\$P1b’.) Interpret the coefficients with respect to their effects (size, sign and significance) on the dependent variable.

```
confint_M2 <- confint(M2, level = 0.95)
stargazer(confint_M2, type = "text")
```

```
##
## =====
##           2.5 %   97.5 %
## -----
## (Intercept)  0.571    1.783
## Population  -0.00000  0.00000
## FDI          -0         0
## ODA          -0         0
## Trade        0.0003    0.004
## Inflation    -0.001    0.001
## -----
```

```
MyRESULTS$P1b <- confint_M2
```

< Confidence interval estimation is a technique used in data analysis to estimate the values of regression results with uncertainty measure caused by sampling variability and

it gives a range in which we can be confident to approve our assumptions (Hill et al., 2018).>

<Population: The coefficient for this variable is insignificant; therefore, there is no meaningful interpretation. However, if it were to be significant, the following would be the interpretation. The increase of total population by 1 unit would lead to the increase of log(annual GDP growth) by 0 and 0 holding other variables constant. The coefficient of Population at 2.5% confidence level is negative indicating that as population increases log(annual GDP growth) decreases but at the 97.5% confidence level the coefficient becomes positive which indicates that the increase in population leads to the increase of log(annual GDP growth).>

<FDI: The coefficient for this variable is insignificant; therefore, there is no meaningful interpretation. However, if it were to be significant, the following would be the interpretation. The increase of FDI by 1 USD would lead to the increase of log(annual GDP growth) by 0 and 0 holding other variables constant. The coefficient of FDI at 2.5% confidence level is negative indicating that as FDI increases log(annual GDP growth) decreases but at the 97.5% confidence level the coefficient becomes positive which indicates that the increase in FDI leads to the increase of log(annual GDP growth).>

<ODA: The coefficient for this variable is insignificant; therefore, there is no meaningful interpretation. However, if it were to be significant, the following would be the interpretation. The increase of ODA by 1 USD would lead to the increase of log(annual GDP growth) by 0 and 0 holding other variables constant. The coefficient of ODA at 2.5% confidence level is negative indicating that as ODA increases log(annual GDP growth) decreases but at the 97.5% confidence level the coefficient becomes positive which indicates that the increase in ODA leads to the increase of log(annual GDP growth).>

<Trade: The coefficient for this variable is significant at 5% significance level, therefore the following interpretation is meaningful. The increase of Trade by 1 % would lead to the increase of log(annual GDP growth) by 0.0002971 and 0.0039972 holding other variables constant. The coefficients of Trade at 2.5% confidence level and 97.5% confidence level are both positive which indicates that the increase in Trade leads to the increase of log(annual GDP growth).>

<Inflation: The coefficient for this variable is insignificant; therefore, there is no meaningful interpretation. However, if it were to be significant, the following would be the

interpretation. The increase of Inflation by 1 % would lead to the increase of  $\log(\text{annual GDP growth})$  by -0.0012703 and 0.0010987 holding other variables constant. The coefficient of Inflation at 2.5% confidence level is negative indicating that as Inflation increases  $\log(\text{annual GDP growth})$  decreases but at the 97.5% confidence level the coefficient becomes positive which indicates that the increase in Inflation leads to the increase of  $\log(\text{annual GDP growth})$ .>

- c) Estimate a model no. 4 (M4) by taking the “best fitting” model from (a) and add the interaction  $\beta_6 FDI * Trade$ . Present the results in a table below. Interpret the interaction coefficient  $\beta_6$ .

```
M4 <- lm (GDP ~ Population + FDI + ODA + Trade + Inflation + I(FDI*Trade), data = filtered_data)
summary(M4)
```

```
##
## Call:
## lm(formula = GDP ~ Population + FDI + ODA + Trade + Inflation +
##      I(FDI * Trade), data = filtered_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.460  -1.333   0.655   2.317  14.941
##
## Coefficients:
##              Estimate      Std. Error t value Pr(>|t|)
## (Intercept)  0.71009114063976542  1.85286496959809344   0.383  0.70224
## Population    0.00000003394491795  0.00000007318286697   0.464  0.64363
## FDI          -0.00000000007523856  0.00000000010622168  -0.708  0.48017
## ODA          -0.00000000636588454  0.00000001688390541  -0.377  0.70683
## Trade         0.01872409282069968  0.00591100555717838   3.168  0.00196 **
## Inflation     0.00194213287610974  0.00378255396458487   0.513  0.60862
## I(FDI * Trade) 0.00000000000008827  0.00000000000032854   0.269  0.78865
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.189 on 116 degrees of freedom
```

```
## Multiple R-squared:  0.1327, Adjusted R-squared:  0.08786
```

```
## F-statistic: 2.959 on 6 and 116 DF,  p-value: 0.01006
```

\_\_\_\_< A combined effect of two or more independent variables on the dependent variable that differs from what would be anticipated based on their individual effects alone is referred to as an interaction. It means that the relationship between the independent variables and the dependent variable is not additive but is instead affected by how the variables interact (James, 2017). Interpreting the interaction between FDI and Trade (I(FDI\*Trade)) will enable us to investigate how the interaction affects the relationship between the dependent - GDP and independent variables - Population, Trade, FDI, ODA, Inflation.>\_\_\_\_

\_\_\_\_< According to the coefficient of I(FDI\*Trade), we can say that if the interaction between FDI and Trade increases by 1 unit then the annual GDP growth will increase by 0 holding other variables constant. In order to determine if the coefficient is statistically significant we check the p-value. The P-value is 0.78865 which is higher than the significance level of 0.05, therefore we fail to reject the null hypothesis which states that “the interaction between FDI and Trade has no effect on GDP”. Hence we can conclude by saying that the interaction between FDI and Trade have no significant effect on the annual GDP growth.>\_\_\_\_

- d) Run a joint-hypothesis test on the null hypothesis that trade has no explanatory power for model M4. (Save the test result in ‘MyRESULTS\$P1d’). Explicitly formulate  $H_0$  and  $H_1$  and interpret the test result.

```
library(AER)
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

```
##
```

```
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
```

```
##
```

```
##      recode
```

```
## Loading required package: lmtest
```

```
## Loading required package: zoo
```

```
##
```

```
## Attaching package: 'zoo'

## The following object is masked from 'package:imputeTS':
##
##      na.locf

## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric

## Loading required package: sandwich

## Loading required package: survival

Hnull <- c("Trade = 0")
(jht <- linearHypothesis(M1, Hnull))

## Linear hypothesis test
##
## Hypothesis:
## Trade = 0
##
## Model 1: restricted model
## Model 2: GDP ~ Population + FDI + ODA + Trade + Inflation
##
##   Res.Df    RSS Df Sum of Sq    F   Pr(>F)
## 1     118 3423.4
## 2     117 3125.5   1    297.92 11.152 0.001126 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Results <- c()
summary(jht)
```

```
##      Res.Df      RSS      Df      Sum of Sq      F
## Min.      :117.0   Min.      :3125   Min.      :1   Min.      :297.9   Min.      :11.15
## 1st Qu.:117.2   1st Qu.:3200   1st Qu.:1   1st Qu.:297.9   1st Qu.:11.15
## Median :117.5   Median :3274   Median :1   Median :297.9   Median :11.15
```

```
## Mean :117.5 Mean :3274 Mean :1 Mean :297.9 Mean :11.15
## 3rd Qu.:117.8 3rd Qu.:3349 3rd Qu.:1 3rd Qu.:297.9 3rd Qu.:11.15
## Max. :118.0 Max. :3423 Max. :1 Max. :297.9 Max. :11.15
##
## NA's :1 NA's :1 NA's :1
## Pr(>F)
## Min. :0.001126
## 1st Qu.:0.001126
## Median :0.001126
## Mean :0.001126
## 3rd Qu.:0.001126
## Max. :0.001126
## NA's :1
```

```
MyRESULTS$P1d <- jht
```

<A joint hypothesis-test examines the combined effect of many independent variables on the dependent variable in order to assess the significance of a model (Greene, 2012). The test contrasts the variance in the dependent variable when the relevant variables are present in the model (the “restricted” model) and when they are not (the “unrestricted” or full model). It analyzes if the extra variance that is explained by the variables is statistically significant.>

<The null hypothesis for a joint hypothesis test is that all the coefficients associated with the variables under consideration are equal to zero simultaneously and the alternative hypothesis is that at least one of the coefficients is non-zero. In this context, we will be considering Trade only. Therefore, our null hypothesis states that “Trade has no explanatory power for model 4” and the alternative hypothesis states that “Trade has explanatory power for model 4”.>

<H0 : Trade has no explanatory power for model 4, H1 : Trade has explanatory power for model 4>

<Significance level is 5% or 0.05. In order to reject or accept the null hypothesis, we consider the p-value. When the p-value is smaller than the significance level, we can reject the null hypothesis but when the p-value is greater/higher than the significance level, we fail to reject the null hypothesis. The F-statistic is 11.152 p-value of our test is 0.001126, which is smaller than 0.05. Therefore, we can reject the null hypothesis (H0)

and conclude that ‘Trade’ has explanatory power for model 4. This means that ‘Trade’ provides valuable information for understanding and predicting the variation in GDP.>

- e) Add country fixed-effects to M4 and estimate it as model no. 5 (M5). Explain how the Goodness-of-fit has changed.

```
M5 <- lm(GDP ~ Population+FDI+ODA+Inflation+Trade + as.factor(country), data = filtered_data)
summary(M4)$r.squared
```

```
## [1] 0.132723
```

```
summary(M5)$r.squared
```

```
## [1] 0.1455648
```

```
print(paste("R-squared for M4: ", summary(M4)$r.squared))
```

```
## [1] "R-squared for M4: 0.13272301235164"
```

```
print(paste("R-squared for M5: ", summary(M5)$r.squared))
```

```
## [1] "R-squared for M5: 0.145564842488751"
```

< As mentioned above in 2a, R-squared is used to measure the goodness of fit. If we observe the change in R-squared we can determine the change in the goodness of fit. The higher the value of R-squared, the better the Goodness-of-fit becomes (Hill et al., 2018).>

<R-squared value for M4 is 0.1327 which means that 13.27% of the variation in GDP is explained by Population, FDI, ODA, Trade, and Inflation in the three countries in my data set(Slovenia, Singapore, and Saudi Arabia). And the R-squared value for M5 is 0.1455 which which means that 14.55% of the variation in GDP is explained by Population, FDI, ODA, Trade, Inflation, plus country fixed effects such as policies and unique characteristics associated with each country in my data set. And it is important to note that the country fixed effects are specific and may be different for each country.>

<We can observe that R-squared value has increased from 0.1327 to 0.1455 which means that M5 has a higher explanatory power of the variation in GDP. This change shows that including country fixed effects has improved the goodness of fit. >

- f) Test M5 for heteroskedasticity and interpret the results. Explain  $H_0$  and  $H_1$ . Save the test statistic as ‘MyRESULTS\$P2f’.

```

bp_test <- bptest (M5)
print(bp_test)

##
##  studentized Breusch-Pagan test
##
## data:  M5
## BP = 35.278, df = 7, p-value = 0.000009917

MyRESULTS$P2f <- bp_test

```

< Many tests can be used to examine the presence of heteroskedasticity in a regression model but I chose to use Breusch-Pagan test because it is simple and easy to understand. The Breusch-Pagan tests the explanatory power of the regression by regressing the squared residuals across all independent variable (Hill et al., 2018). The null hypothesis of the test states that “There is no heteroscedasticity in the regression model” which assumes that the variance of residuals is constant across different levels of the independent variables. The alternative hypothesis states that “There is heteroscedasticity in the regression model”. >

<H0 : There is no heteroscedasticity in the regression model, H1 : There is heteroscedasticity in the regression model>

<Significance level is 5% or 0.05. In order to reject or accept the null hypothesis, we consider the p-value. When the p-value is smaller than the significance level, we can reject the null hypothesis but when the p-value is greater/higher than the significance level, we fail to reject the null hypothesis. The P-value is 0.0000099 which is less than the significance level, therefore we can reject the null hypothesis (homoscedastic). Hence, the residuals are not distributed with equal variance at each level of the independent variables indicating the presence of heteroskedasticity in M5.>

### 3. Determinants of Human Development

In this problem, you will construct our own Human Development Index (HDI). In the WDI-Database find 1-2 indicators each that have meaningful explanatory power for: - Education, e.g. primary or secondary enrollment rates - Health, e.g. life expectancy or infant mortality rates - Infrastructure, e.g. basic or upgraded sanitation facilities, internet users Calculate an index (I) for each indicator with



range 0-1 by dividing by their indicator maximum (over all 3 countries):

$$Index = \frac{Indicator}{\max(Indicator)}$$

Define the HDI by the geometric mean of these three variables:

$$HDI = \sqrt[k]{I_{Education} * I_{Health} * I_{Infrastructure}}$$

with  $k$ : number of actual (3-6) Indices included.

For the following regression models, use HDI as dependent variable, and estimate linear-log models with the below mentioned explanatory variables.

- a) Estimate model no. 6 (M6) with GDP, FDI, Trade, and ODA as explanatory variables. Use a joint hypothesis test to detect whether all coefficients are zero. (Save the test-statistic as ‘MyRESULTS\$P3a’). Interpret the results.

```
HDI=filtered_data$index_Education <- filtered_data$Education / max(filtered_data$Education, na.rm = TRUE)
filtered_data$index_Health <- filtered_data$Health / max(filtered_data$Health, na.rm = TRUE)
filtered_data$index_Infratructure <- filtered_data$Infrastructure / max(filtered_data$Infrastructure, na.rm = TRUE)

k = rowSums(!is.na(filtered_data[,c("index_Education", "index_Health", "index_Infratructure")]))

filtered_data$HDI = (filtered_data$index_Education * filtered_data$index_Health * filtered_data$index_Infratructure)^(1/k)

library(stargazer)
WDIfull_renamed$FDI <- ifelse(WDIfull_renamed$FDI < 0, 0, WDIfull_renamed$FDI)
WDIfull_renamed$GDP <- ifelse(WDIfull_renamed$GDP < 0, 0, WDIfull_renamed$GDP)

WDIfull_renamed$FDI <- WDIfull_renamed$FDI +1
WDIfull_renamed$GDP <- WDIfull_renamed$GDP +1

M6 <- lm(HDI ~ log(GDP) + log(FDI) + log(Trade) + log(ODA), data = WDIfull_renamed)

stargazer (M6, type= "text")

##
```

```
## =====
##                               Dependent variable:
##                               -----
##                               HDI
## -----
## log(GDP)                      0.031*
##                               (0.017)
##
## log(FDI)                      -0.001
##                               (0.002)
##
## log(Trade)                   0.066***
##                               (0.024)
##
## log(ODA)                     -0.061***
##                               (0.011)
##
## Constant                     1.228***
##                               (0.234)
##
## -----
## Observations                  116
## R2                           0.362
## Adjusted R2                  0.339
## Residual Std. Error          0.145 (df = 111)
## F Statistic                  15.755*** (df = 4; 111)
## =====
## Note:                        *p<0.1; **p<0.05; ***p<0.01
```

```
Hnull <- c("log(GDP) = 0", "log(Trade)=0", "log(ODA)=0", "log(FDI)=0")
linearHypothesis(M6, Hnull)
```

```
## Linear hypothesis test
##
## Hypothesis:
```

```
## log(GDP) = 0
## log(Trade) = 0
## log(ODA) = 0
## log(FDI) = 0
##
## Model 1: restricted model
## Model 2: HDI ~ log(GDP) + log(FDI) + log(Trade) + log(ODA)
##
##   Res.Df    RSS Df Sum of Sq    F        Pr(>F)
## 1     115 3.6400
## 2     111 2.3218   4     1.3182 15.755 0.0000000003064 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Results <- c()
MyRESULTS$P3a <- M6

options(scipen = 999)
```

<A joint hypothesis-test examines the combined effect of many independent variables on the dependent variable in order to assess the significance of a model (Greene, 2012). The dependent variable in this context is HDI and the independent variables are GDP, FDI, ODA, and Trade.>

<The null hypothesis for a joint hypothesis test is that all the coefficients associated with the variables under consideration are equal to zero simultaneously and the alternative hypothesis is that at least one of the coefficients is non-zero.>

<H0 : All coefficient in the model are equal to zero, H1 : At least one coefficient in the model is not equal to zero>

<Since the determinant factor in rejecting or failing to reject the null hypothesis is the p-value, we have to look at our P-value. The F-statistic is 16.285 and the p-value for our test is 0.0000000003064 which is smaller than 0.05 therefore we can reject the null hypothesis which means that we have enough evidence to conclude that any of the independent variables (GDP, FDI, ODA, and Trade) have a significant effect on the dependent variable (HDI).>

<This is correct because as we observe in our results, the coefficients for the variables of  $\log(\text{GDP})$  and  $\log(\text{Trade})$  are significant at a 5% significance level and  $\log(\text{ODA})$  is highly significant at 1% significance level. Therefore the interpretation of the coefficients is meaningful.>

< $\log(\text{GDP})$  : The increase of GDP by 1 % would lead to the increase of HDI by 0.0308411 holding other variables constant. The effect of GDP on HDI is positive and statistically significant.>

< $\log(\text{Trade})$  : The increase of Trade by 1 % would lead to the increase of HDI by 0.0658141 holding other variables constant. The effect of Trade on HDI is positive and statistically significant.>

< $\log(\text{ODA})$ : The increase of ODA by 1 USD would lead to the increase of HDI by 0.0614487 holding other variables constant. The effect of ODA on HDI is negative and statistically significant.>

- b) You want to find out, whether there is a minimum or maximum effect of Official Development Assistance (ODA) on Human Development. Find a suitable functional form and estimate it as model no. 7 (M7). Calculate the minimum/maximum (and save it in 'MyRESULTS\$P3b'). Interpret your finding.

```
M7 <- lm(HDI ~ GDP + FDI + Trade + ODA + I(ODA^2), data = filtered_data)
summary(M7)

##
## Call:
## lm(formula = HDI ~ GDP + FDI + Trade + ODA + I(ODA^2), data = filtered_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.145284 -0.029454 -0.004294  0.031824  0.141627
##
## Coefficients:
##              Estimate      Std. Error t value
## (Intercept)  0.603223509063815832576  0.009188067753762355766  65.653
## GDP          -0.000559993608421676162  0.000829970780753761851  -0.675
```

```
## FDI          0.000000000000869651859  0.000000000000232776400   3.736
## Trade       -0.000344724173176056423  0.000041589646268635379  -8.289
## ODA          0.000000000178207697648  0.000000000126685216160   1.407
## I(ODA^2)     0.000000000000000003000  0.000000000000000001266   2.370
##
##              Pr(>|t|)
## (Intercept) < 0.00000000000000002 ***
## GDP          0.501188
## FDI          0.000291 ***
## Trade        0.0000000000000219 ***
## ODA          0.162168
## I(ODA^2)     0.019434 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0463 on 117 degrees of freedom
## Multiple R-squared:  0.4825, Adjusted R-squared:  0.4604
## F-statistic: 21.82 on 5 and 117 DF,  p-value: 0.0000000000000002186

MyRESULTS$P3b <- M7
```

< By analyzing and interpreting the coefficients of ODA and  $I(ODA^2)$  we will be able to find the maximum and minimum effect of ODA on HDI.>

< If ODA increases by 1 USD then the HDI will increase by 0 holding other variables constant. The P-value is 0.162168 which is greater than the significance level of 0.05, therefore we fail to reject the null hypothesis which states that “the interaction effect of ODA has no effect on HDI”. Hence we can conclude by saying that ODA has no significant effect on the HDI.>

< If the interaction effect of ODA increases by 1 USD then the HDI will increase by 0 holding other variables constant. The P-value is 0.019434 which is smaller than the significance level of 0.05, therefore we can reject the null hypothesis which states that “the interaction effect of ODA has no effect on HDI”. Hence we can conclude by saying that the interaction effect of ODA has a significant effect on the HDI.>

< Therefore we can conclude by saying that  $I(ODA^2)$  has a maximum effect on HDI because it is statistically significant and ODA has a minimum effect because its effect on

**HDI is not statistically significant.>**

- c) Carry out the Variance Inflation Factor (VIF) test on Model no. 7 (M7) to check for multicollinearity. Interpret if multicollinearity is an issue in the model.

```
Vif1 <- vif(M7)
```

```
Vif1
```

```
##      GDP      FDI      Trade      ODA I(ODA^2)
```

```
## 1.157489 1.463152 1.538103 1.246885 1.222243
```

< When one or more independent variables in a regression model have a strong correlation with one another, this is known as multi-collinearity. Due to the difficulty in analyzing and interpreting the impact of each independent variable on the dependent variable in the model, this can cause results to be misinterpreted. Hence, diminishing the statistical significance of the independent variable (Allen, 1997, pp.176-180). >

< Variance Inflation factor test for multicollinearity by quantifying the extent to which the variance of the estimated regression coefficient of an independent variable is inflated due to its correlation with other independent variables in the model. A high VIF proves the presence of multicollinearity.>

< In this model the VIFs are less than 5 (1-1.5) which means that all our variables pass the multi-collinearity test implying that there is no correlation between a given independent variable and other independent variables in the model. Therefore, multi-collinearity is not an issue in this model.>

- d) By using Ramsey's regression equation specification error (RESET) test, check to see if the functional form of M7 is adequate. Explain the test logic and interpret the result. (Save the test statistic as 'MyRESULTS\$P3d'.)

```
library(lmtest)
```

```
rrt <- resettest(M7, power = 2:3, type = "fitted" )
```

```
rrt
```

```
##
```

```
## RESET test
```

```
##
```

```
## data: M7
```

```
## RESET = 4.98, df1 = 2, df2 = 115, p-value = 0.008429
```

```
MyRESULTS$P3d <- rrt
```

< The RESET test determines whether the regression model appropriately represents the nonlinear relationship between the independent variables and the dependent variable in order to determine whether the functional forms are adequate (Wooldridge & Jeffrey, 2013). The null hypothesis of the reset test states that " the regression model has no omitted variables and the functional form is adequate" and the alternative hypothesis states that " the regression model has omitted variables and the functional form is not adequate".>

< H0: no omitted variables and the functional form is adequate, H1: omitted variables and the functional form is not adequate.>

< P-value is 0.0084295 which is smaller than 0.05, therefore we can reject the null hypothesis. Hence, the functional form is inadequate which means that there are likely omitted variables and specification errors that need to be addressed. >

- e) Use the Jarque-Bera-Test to test for normality of the residuals. Save the test statistic as 'MyRESULTS\$P3e'. Interpret all the results by explaining the test logic, including  $H_0$  and  $H_1$ .

```
library(tseries)
```

```
##
```

```
## Attaching package: 'tseries'
```

```
## The following object is masked from 'package:imputeTS':
```

```
##
```

```
##      na.remove
```

```
j.b.t <- jarque.bera.test(M7$residuals)
```

```
j.b.t
```

```
##
```

```
## Jarque Bera Test
```

```
##
```

```
## data: M7$residuals
```

```
## X-squared = 2.6999, df = 2, p-value = 0.2593
```

```
MyRESULTS$P3e <- j.b.t
```

< Jarque-Bera-Test test whether the residuals are normally distributed by examining for Kurtosis and skewness (Greene, 2012). The null hypothesis of the jarque-bera-test states that “Residuals are normally distributed” and the alternative hypothesis states that “Residuals are normally not distributed”. >

< H0 :Residuals are normally distributed, H1: Residuals are not normally distributed.>

< P-value is 0.2592592 which is greater than the significance level 0.05, therefore we fail to reject the null hypothesis. Hence, the data is normally distributed as there is no significant departure from normality based on the test.>

## 4. Summary and Conclusion

By including all explanatory variables used in models M1-M7, find the models with the largest explanatory power for: - economic development (M8) - human development (M9) by selecting an appropriate functional form. In M8 exclude GDP on the right hand-side and include a polynomial of third order of FDI. In M9 include the polynomial of third order for GDP.

- a) Present the estimation results of both models in a single table. Summarize and compare the main results for both models, and interpret the effects of changes in the explanatory variables on the dependent variable, as well as the statistical significance of their coefficients.

```
library(stargazer)
M8 <- lm(GDP ~ poly(FDI, 3) + Trade + Inflation + ODA+ Population + Education +Health + Infrast
M9 <- lm(HDI ~ poly(GDP, 3) + Trade + Inflation + ODA+ Population + Education +Health + Infrast
stargazer(M8, M9, type = "text")
```

```
##
```

```
## =====
```

```
##                               Dependent variable:
```

```
##                               -----
```

```
##                               GDP                HDI
```

```
##                               (1)                (2)
```

```
## -----
```

```
## poly(FDI, 3)1                -11.495*
```



##	(6.911)	
##		
## poly(FDI, 3)2	3.014	
##	(6.340)	
##		
## poly(FDI, 3)3	-5.075	
##	(5.752)	
##		
## poly(GDP, 3)1		-0.042**
##		(0.017)
##		
## poly(GDP, 3)2		0.013
##		(0.018)
##		
## poly(GDP, 3)3		0.010
##		(0.017)
##		
## Trade	0.015	-0.0002***
##	(0.015)	(0.00004)
##		
## Inflation	0.001	0.00000
##	(0.004)	(0.00001)
##		
## ODA	-0.000	-0.000
##	(0.00000)	(0.000)
##		
## Population	-0.00000	0.000***
##	(0.00000)	(0.000)
##		
## Education	0.149	0.008***
##	(0.180)	(0.001)
##		
## Health	-0.339	0.053***

##	(0.561)	(0.002)
##		
## Infrastructure	-1.105	0.066***
##	(2.969)	(0.009)
##		
## Constant	3.589	0.019
##	(9.990)	(0.032)
##		
## -----		
## Observations	123	123
## R2	0.149	0.939
## Adjusted R2	0.073	0.933
## Residual Std. Error (df = 112)	5.232	0.016
## F Statistic (df = 10; 112)	1.955**	171.782***
## =====		
## Note:	*p<0.1; **p<0.05; ***p<0.01	

<M8 used GDP as the dependent variable and M9 used HDI as the dependent variable.>

<M8 interpretation : The “poly(FDI,3)1, poly(FDI,3)2, and poly(FDI,3)3 indicate the coefficient estimate of the first order, second order, and third order respectively of the polynomial of FDI in the GDP model.>

<poly(FDI, 3)1: if poly(FDI, 3)1 increases by 1 unit, then the GDP decreases by 11.495 holding other variables constant in the model. This indicates that the increase of FDI has a negative effect on the GDP because it leads to its decrease. The coefficient of poly(FDI, 3)1 is only significant at 10% significance level.>

<poly(FDI, 3)2: The coefficient for this variable is insignificant; therefore, there is no meaningful interpretation. However, if it were to be significant, the following would be the interpretation. If poly(FDI, 3)2 increases by 1 unit, then the GDP increases by 3.014 holding other variables constant in the model. This indicates that the increase of FDI has a positive effect on the GDP because it leads to its increase.>

<poly(FDI, 3)3: The coefficient for this variable is insignificant; therefore, there is no meaningful interpretation.>

<Trade: The coefficient for this variable is insignificant; therefore, there is no meaningful interpretation. However, if it were to be significant, the following would be the interpretation. If Trade increases by 1% then the GDP will increase by 0.015 holding other variables constant.>

<Inflation: The coefficient for this variable is insignificant; therefore, there is no meaningful interpretation.>

<ODA: The coefficient for this variable is insignificant; therefore, there is no meaningful interpretation.>

<Population: The coefficient for this variable is insignificant; therefore, there is no meaningful interpretation.>

<Education: The coefficient for this variable is insignificant; therefore, there is no meaningful interpretation. However, if it were to be significant, the following would be the interpretation. If total government expenditure spent on education increases by 1 USD then the GDP will increase by 0.149 holding other variables constant.>

<Health: The coefficient for this variable is insignificant; therefore, there is no meaningful interpretation.>

<Infrastructure: The coefficient for this variable is insignificant; therefore, there is no meaningful interpretation.>

<M9 interpretation: The “poly(GDP,3)1, poly(GDP,3)2, and poly(GDP,3)3 indicate the coefficient estimate of the first order, second order, and third order respectively of the polynomial of GDP in the HDI model.>

<poly(GDP, 3)1: if poly(GDP, 3)1 increases by 1 unit, then the HDI decreases by 0.042 holding other variables constant in the model. The effect of poly(GDP, 3)1 on HDI is negative and statistically significant at 5% significance level.>

<poly(GDP, 3)2: The coefficient for this variable is insignificant; therefore, there is no meaningful interpretation.>

<poly(GDP, 3)3: The coefficient for this variable is insignificant; therefore, there is no meaningful interpretation.>

<Trade: The effect of Trade on HDI is positive and statistically significant at 1% significance level (highly significant). If Trade increases by 1% then the HDI will decrease

by 0.0001612 holding other variables constant.>

<Inflation: The coefficient for this variable is insignificant; therefore, there is no meaningful interpretation.>

<ODA: The coefficient for this variable is insignificant; therefore, there is no meaningful interpretation.>

<Population: The effect of population on HDI is positive and statistically significant at 1% significance level(highly significant). If Population increases by 1 unit then the HDI will increase by 0 holding other variables constant.>

<Education: The effect of government expenditure spent on education on HDI is positive and statistically significant at 1% significance level(highly significant). If government expenditure spent on education increases by 1 USD then the HDI will increase by 0.0082132 holding other variables constant.>

<Health: The effect of government expenditure on Health on HDI is positive and statistically significant at 1% significance level(highly significant). If total government expenditure on Health increases by 1 USD then the HDI will increase by 0.0529139 holding other variables constant.>

<Infrastructure: The effect of infrastructure quality, Logistics Performance Index on HDI is positive and statistically significant at 1% significance level(highly significant). If infrastructure quality, Logistics Performance Index increases by 1 unit then the HDI will increase by 0.0663943 holding other variables constant.>

<Therefore we can conclude by saying that M9 has coefficients that are more statistically significant than M8 (only one coefficient was statistically significant). This means that the relationship between the independent variables and the dependent variable HDI is more statistically reliable and strong compared to the relationship between the independent variables and dependent variable GDP.>

- b) Calculate the OLS residuals and report summary statistics (minimum, maximum, mean, median, variance) for both models. Explain in what ways these results support or violate the assumptions of the OLS-regression method.

```
M8_residuals <- M8$residuals
```

```
M9_residuals <- M9$residuals
```

```
summary(M8_residuals)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## -22.1480  -1.4540    0.6543    0.0000    2.3796   14.3226
```

```
summary(M9_residuals)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## -0.0353769 -0.0098343 -0.0005752  0.0000000  0.0063614  0.0512059
```

```
var(M8_residuals)
```

```
## [1] 25.13434
```

```
var(M9_residuals)
```

```
## [1] 0.000243139
```

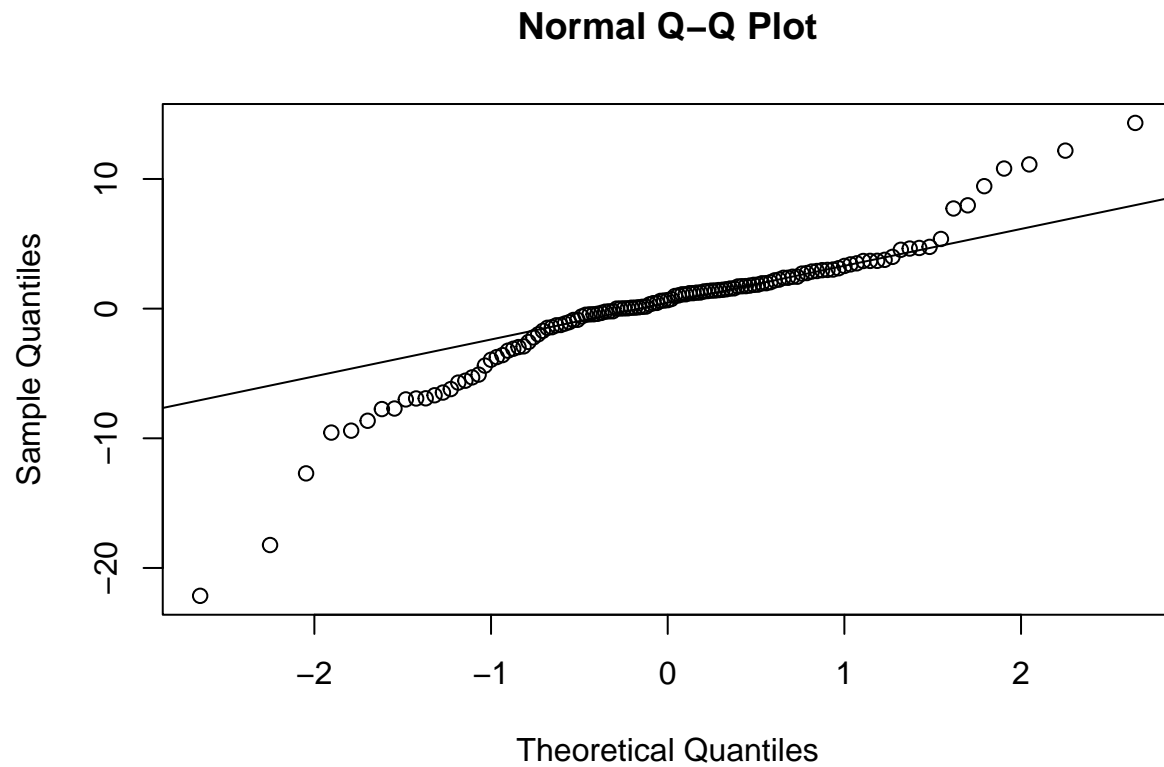
< The OLS (Ordinary Least squares) is a method used to determine the parameters of a regression model. The goal of the method is to minimize the sum of squared residuals from the linear equation. OLS has many assumptions and before we analyze how our results support or violate them, it is important to note which assumptions can be analyzed based on the results we have available and those are homoscedasticity and normality. Homoscedasticity assumes that the variance of residuals is constant across all levels of the independent variables and normality assumes that the residuals in a regression model follow a normal distribution and are distributed around zero (Hill et al., 2018)>

<The OLS residuals of M8 have a high variance which indicates that the residuals have a large amount of dispersion and can also indicate heteroscedasticity. This violates the assumption of Homoscedasticity of the OLS regression model. Additionally, the residuals are distributed around 0 because we can observe that the mean is close to zero supporting the OLS assumption of normality.>

<On the other hand, the OLS residuals of M9 have a low variance which indicates that the residuals have a small dispersion around the regression line which supports the OLS regression assumption of homoscedasticity. The residuals of M9 also support the OLS assumption of normality because they are distributed around 0. We can conclude by saying that the OLS regression assumptions are partially supported and supported in both models as indicated above.>

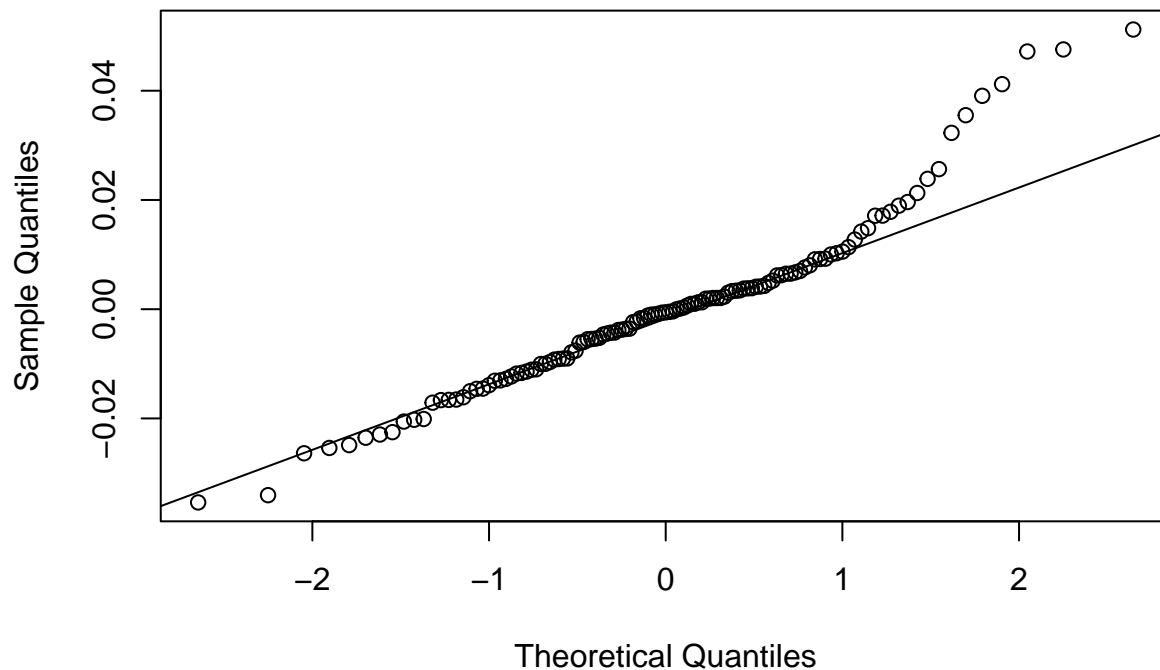
- c) Generate the Q-Q-plots for the residuals of both models. Explain the logic behind the plot and interpret the results.

```
M8_residuals <- residuals(M8)
qqnorm(M8_residuals)
qqline(M8_residuals)
```



```
M9_residuals <- residuals(M9)
qqnorm(M9_residuals)
qqline(M9_residuals)
```

## Normal Q-Q Plot



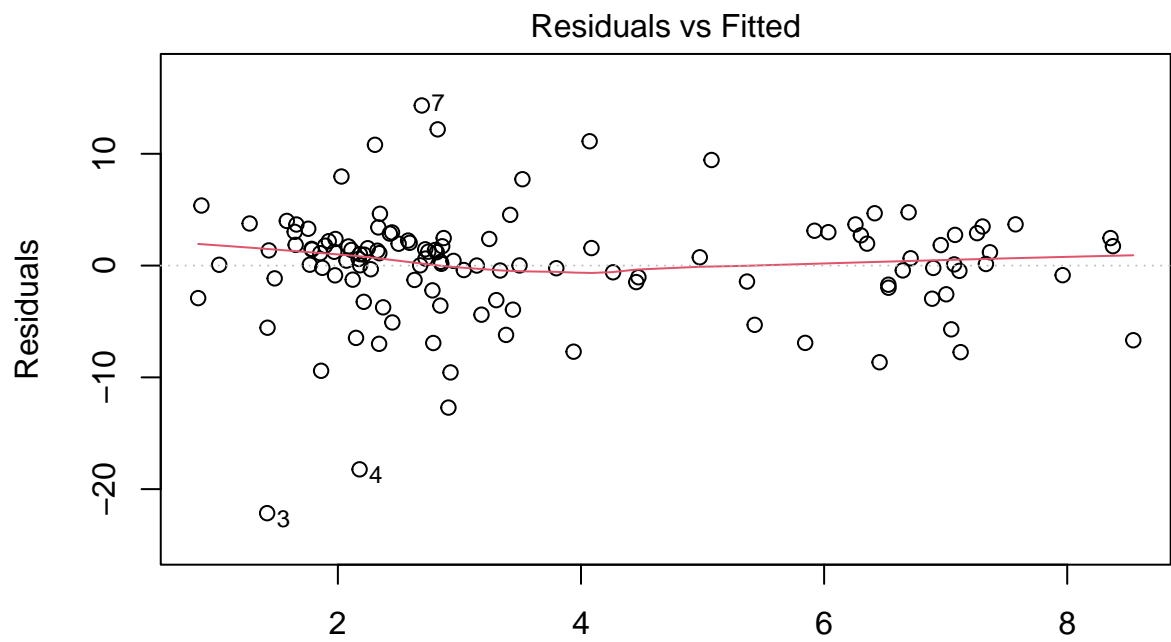
<Q-Q (Quantile-Quantile) plot is a graphical tool that helps to estimate the data at hand is in line with a certain theoretical distribution, the normal distribution typically. Quantile is a point in data, below which a certain share of the data falls (James, 2017). The logic behind the plot is to show the distribution of residuals of M8 AND M9.>

<The quantiles of residuals in M8 model are plotted against the normal distribution quantiles, the scatter plot roughly goes along the 45 degree line, therefore we can say that our residuals are fairly distributed but large dispersion is observed in the lower quantile compared to M9 model and a low dispersion in upper quantile.>

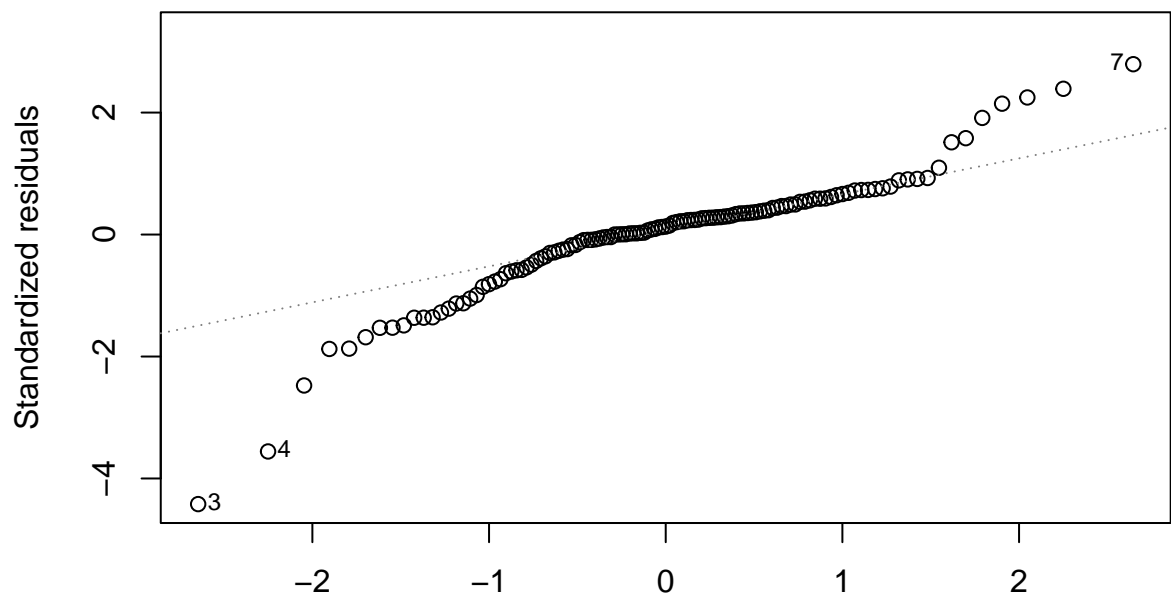
<For M9 we can conclude that our residuals are fairly distributed because the scatter plot roughly goes along the 45 degree line. However, we can see some deviations mostly in the upper quantiles and few in lower quantiles.>

- d) Generate the Scale-Location-plots for both models. Explain the logic behind the plot and interpret the results.

```
plot(M8)
```

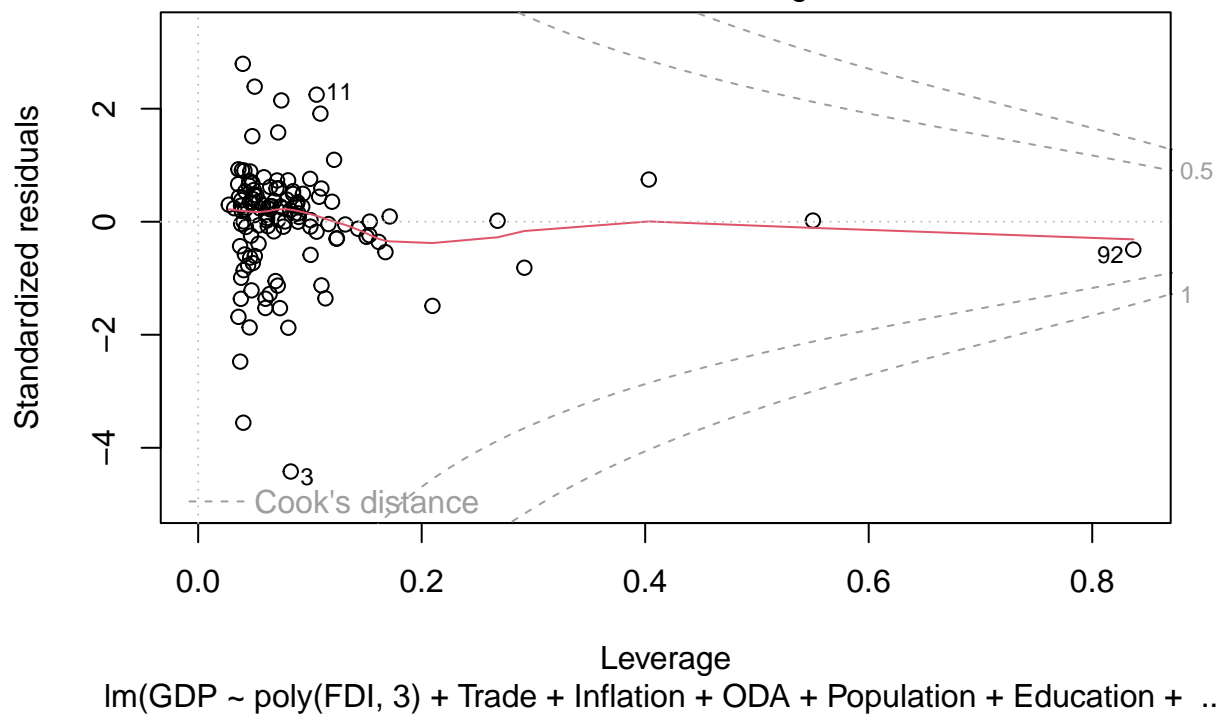
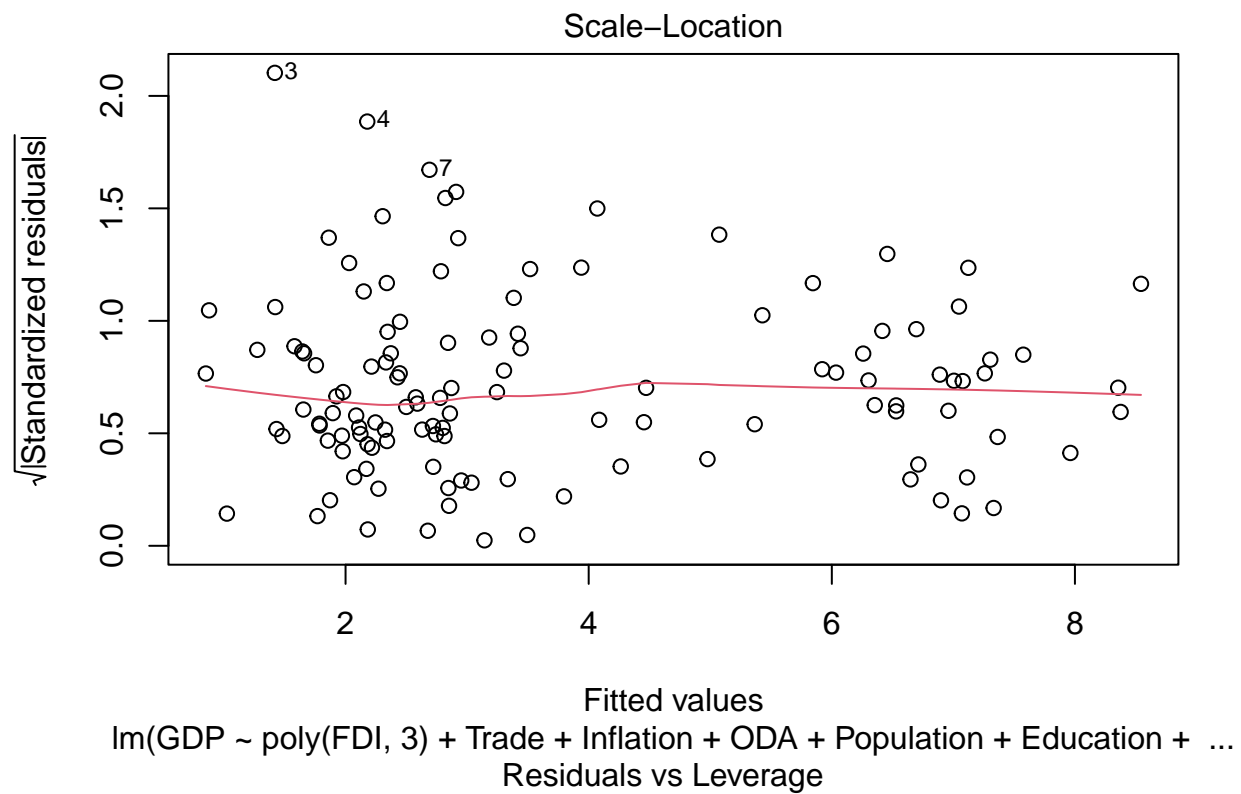


Fitted values  
 $\text{lm}(\text{GDP} \sim \text{poly}(\text{FDI}, 3) + \text{Trade} + \text{Inflation} + \text{ODA} + \text{Population} + \text{Education} + \dots)$   
 Normal Q-Q

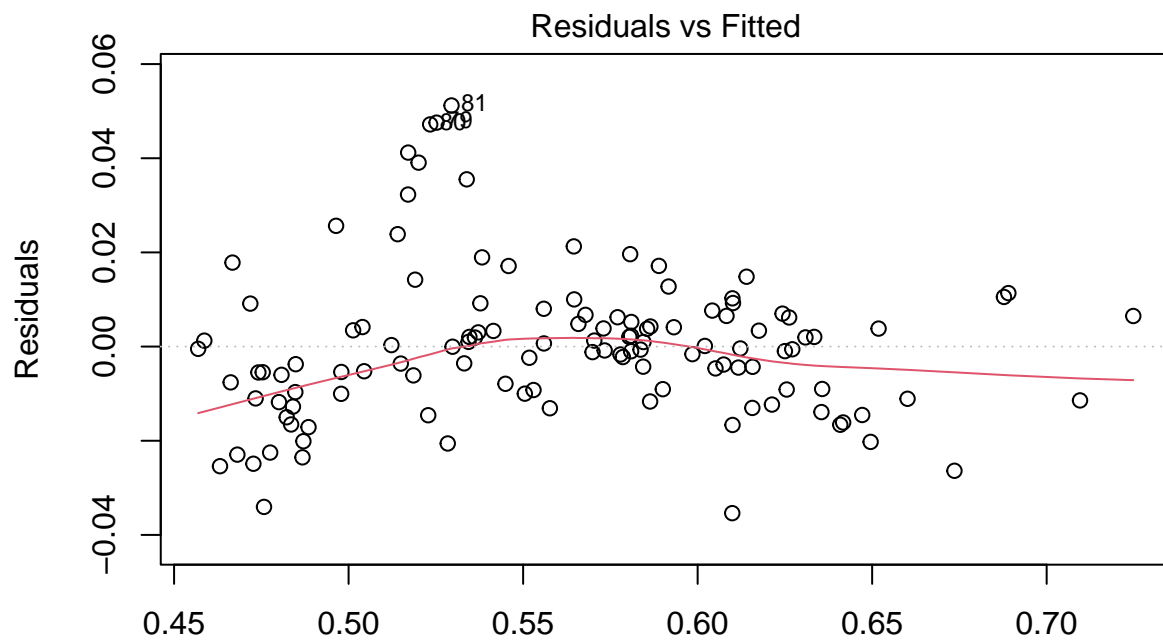


Theoretical Quantiles  
 $\text{lm}(\text{GDP} \sim \text{poly}(\text{FDI}, 3) + \text{Trade} + \text{Inflation} + \text{ODA} + \text{Population} + \text{Education} + \dots)$

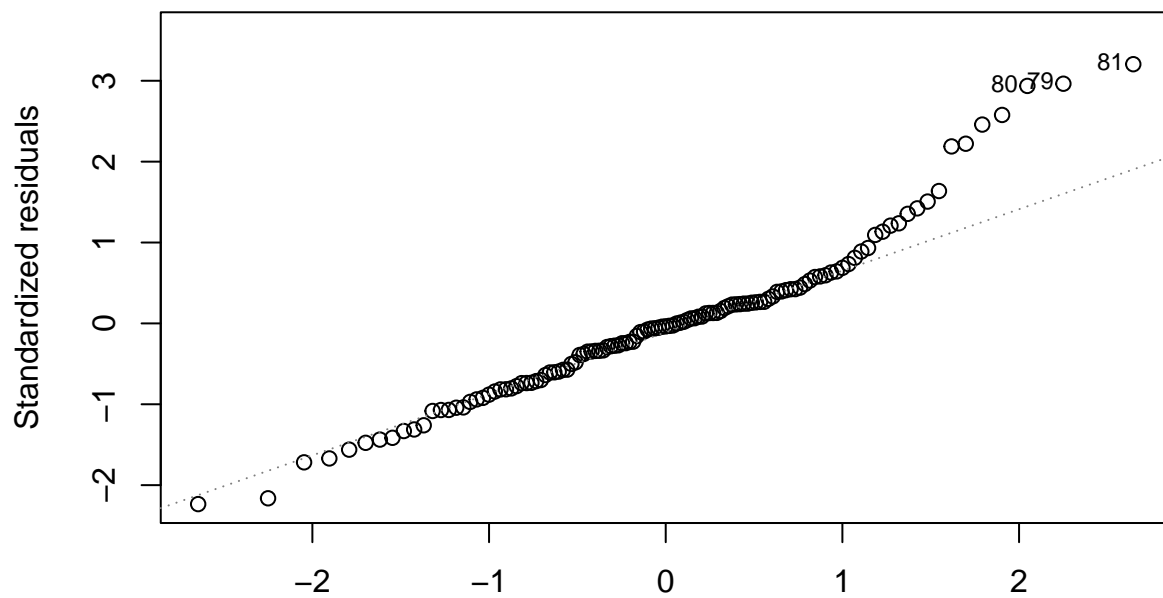




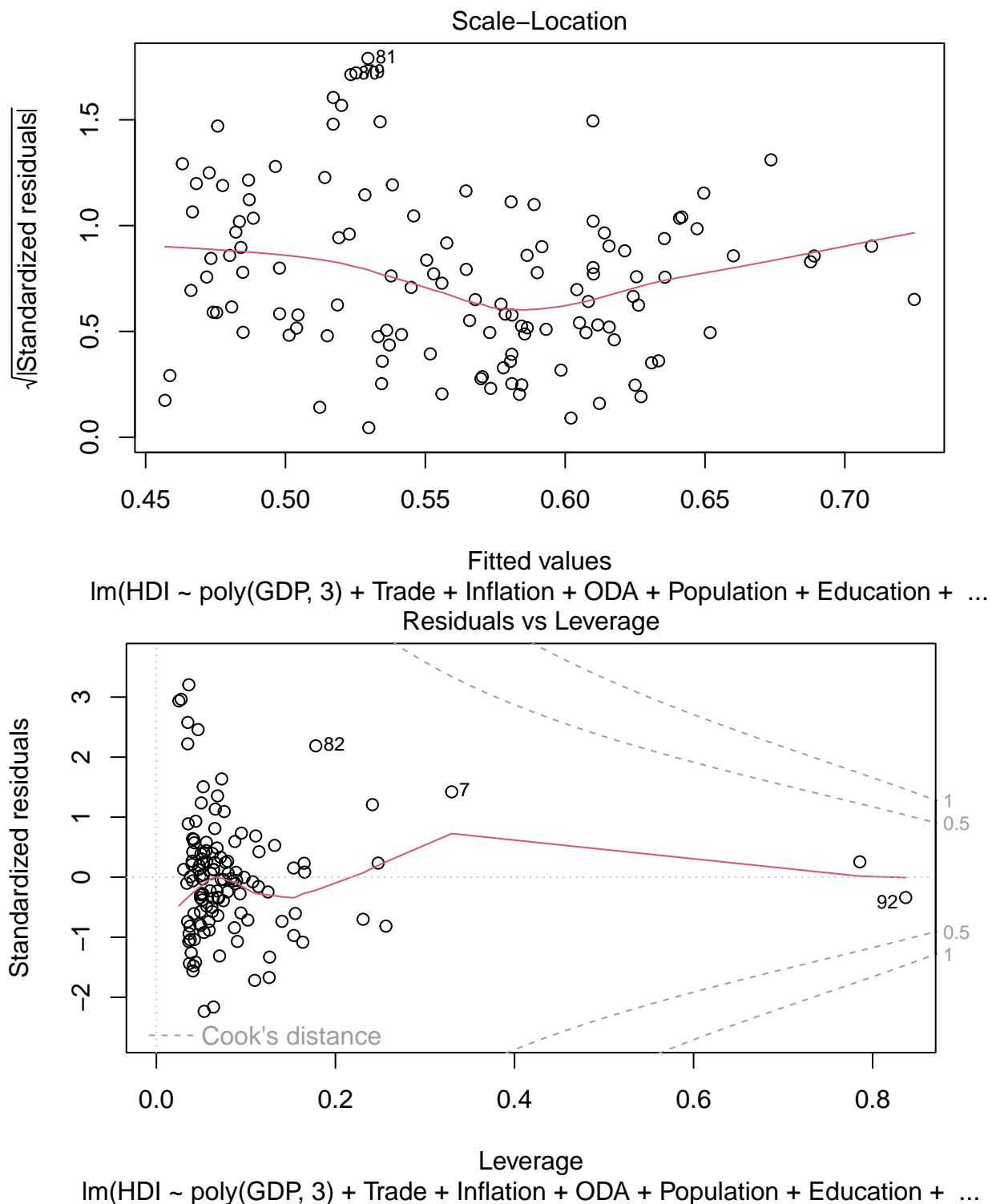
```
plot(M9)
```



Fitted values  
 $\text{lm}(\text{HDI} \sim \text{poly}(\text{GDP}, 3) + \text{Trade} + \text{Inflation} + \text{ODA} + \text{Population} + \text{Education} + \dots)$   
 Normal Q-Q



Theoretical Quantiles  
 $\text{lm}(\text{HDI} \sim \text{poly}(\text{GDP}, 3) + \text{Trade} + \text{Inflation} + \text{ODA} + \text{Population} + \text{Education} + \dots)$



<The logic behind the plot is to identify any errors in the residuals like heteroscedasticity, wrong functional form, autocorrelation. All these can be visualized by looking at the residuals plots (Hill et al., 2018).>

<For M8, the residuals do not always spread evenly across the horizontal line, and some of the empty spaces suggest that the residuals distribute dynamically and show a left-handed

bias. This violates the assumption of homoscedasticity and supports the presence of heteroscedasticity.>

<For M9, there is no constant even spread of the residuals across the horizontal line and a randomized spread of points is observed. Although the spread do not show any side bias trend, we can assume the presence of heteroscedasticity because the assumption of homoscedasticity was violated by the uneven spread of the residuals. >

## References

## Declaration of Authenticity

I, Mulisa Kevine Umwiza, hereby declare that the work presented herein is my own work completed without the use of any aids other than those listed. Any material from other sources or works done by others has been given due acknowledgement and listed in the reference section. Sentences or parts of sentences quoted literally are marked as quotations. The work presented herein has not been published or submitted elsewhere for assessment in the same or a similar form. I will retain a copy of this assignment until after the Board of Examiners has published the results, which I will make available on request.

This document was created by machine and is therefore valid without signature.