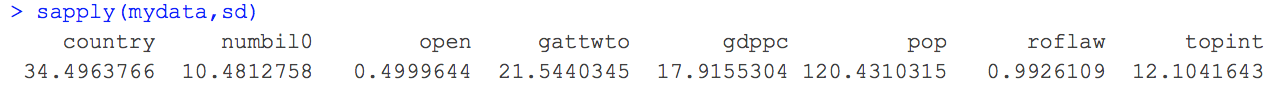
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name** | **Student ID Number** | **Tutor** | **Tutorial Day & Time** | **Tutorial Location** |
| Chan Jie Ho | 961948 | Richard Hayes | Monday 3:15 | The Spot 3031 |

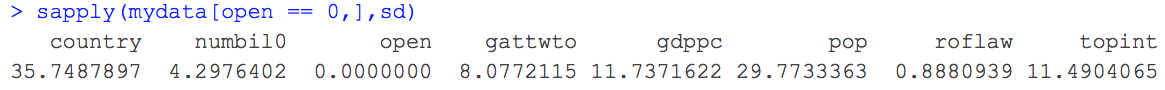
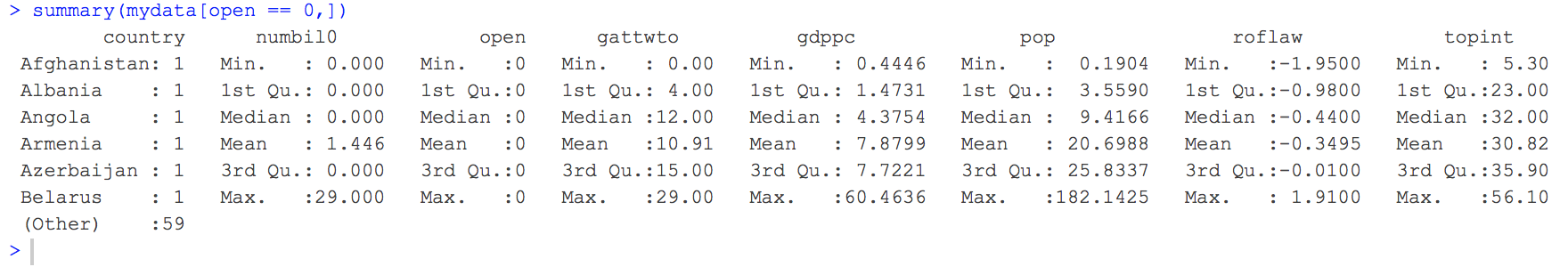
**ASSIGNMENT 1**

**Question 1**

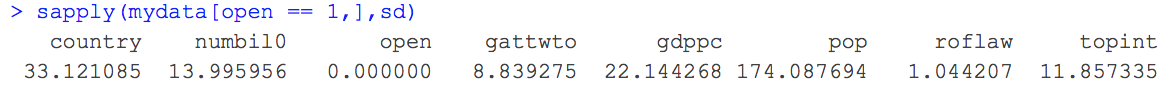
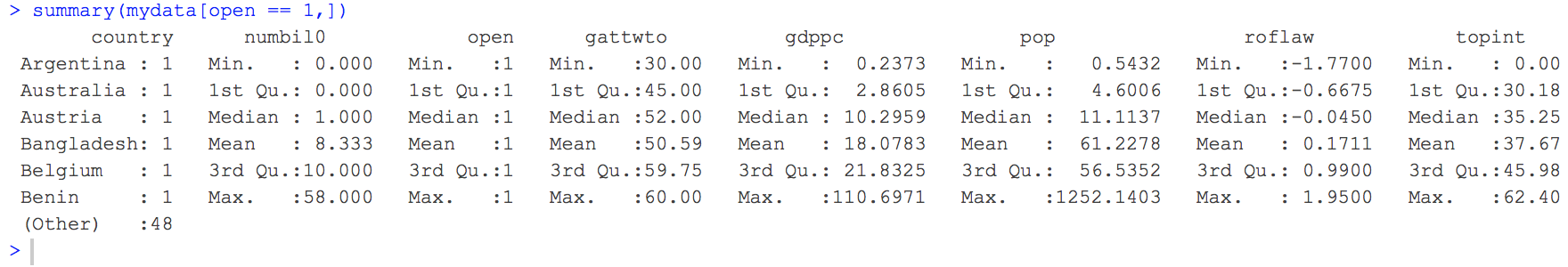
**Summary statistics for entire dataset:**



**Summary statistics for countries closed to trade:**



**Summary statistics for countries closed to trade:**



A typical country would follow the mean of the entire dataset and can therefore be said to have 4 to 5 billionaires, been a member of GATT or WTO for 28 to 29 years, a GDP per capita of $12 507.70, a population of 39.0901 million people, a rule of law index of -0.1133, a top marginal income tax rate of 33.93%, and is not open to trade as only 45.38% of countries are open to trade, but if we split them based on whether they are open or closed to trade, countries closed to trade has, on average, less billionaires, with an average of 1 to 2 billionaires per country, with a standard deviation of 4 to 5 billionaires between countries, than countries open to trade, which has an average of 8 to 9 billionaires, with a standard deviation of 13 to 14 billionaires between countries, as more than three quarters of countries closed to trade have no billionaires while half of those open to trade not only have at least one with 25% having at least 10, but they also have a higher max number of billionaires compared to countries closed to trade, which can be justified by the fact that those open to trade would likely have more business opportunities that allow people to have a higher chance of becoming a billionaire, and this could also explain the differences in population where at least 75% of those closed to trade have a lower population than at least half of those open to trade, where the mean population of those closed to trade of 20.6988 million people, which has a standard deviation of 29.733336 million people, is almost 3 times smaller than those open at 61.2278 million people, which has a standard deviation of 174.087694 million people, as the increase in business opportunities would result in people wanting to migrate to countries that are open to trade as a means to take advantage of the benefits of international trading in which people would often flock to the same few countries, thus resulting in countries having extremely large populations compared to all the other countries, which may explain the very large standard deviation. This can also explain why countries opened to trade has a higher average GDP per capita of US$18 078.30, with a standard deviation of US$22 144.27, while those closed to trade only has an average GDP per capita of US$7 879.90, with a standard deviation of US$11 737.16. Besides that, countries closed to trade with the greatest number of years as a member of GATT or WTO of 29 years has less years than the countries open to trade with the least, which is 30 years, where the average length of the membership of those open to trade is approximately 5 times longer than those closed to trade but has not much difference in standard deviations.

**Question 2**



The density function for the number of billionaires per country is very rightly skewed where almost a third of countries have no billionaires while most of the other countries have vastly different numbers of billionaires from one another, which is foreseeable as billionaires tend to be in the same few countries possibly due to the business opportunities that the country can provide them, thus leaving countries that are poor business-opportunities-wise, which also explains the few observations of a country have more than 30 billionaires, whereas the density function for the number of years spent as a GATT or WTO member, on the other hand, is bimodal, with both modes being either very high or very low in the range, which is justified as while many countries were part of the founding of the GATT or WTO, many countries have only recently become a member with the help of the Internet as GATT and WTO can better promote their work and heightened living standards to the people of other countries via the world wide web and thus encouraging them to work towards encouraging their countries to join. However, putting them together, the very high density of countries with no billionaires could imply that the number of years a country was a member may not have a very significant effect on the number of billionaires that the country houses.

**Question 3**

****

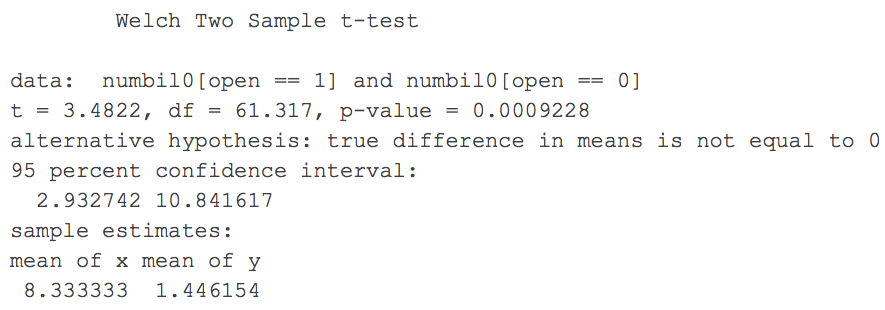
From the summary statistics, we know that the total number of countries are spread rather evenly between closed to trade and open to trade, but this density graph shows that not only are there almost three times the number of countries closed to trade that have no billionaires than those open to trade, but for every country closed to trade with a certain number of billionaires, there is almost twice as many countries open to trade that have that many billionaires. This ties in with the fact that we can also see that the max number of billionaires for those closed to trade falls very short of the max number of billionaires in open countries (about half), and we can therefore conclude that there is a higher probability that any country with at least a few billionaires is open to trade.

**Question 4**



Judging from the scatter plot (middle graph) of the distribution of billionaires per country of all countries based on number of years spent as a member of GATT or WTO, we can see that there is a moderately positive correlation between the two variables as shown by the positive gradient of the line of best fit. However, when we split them based on whether they are closed to trade (left graph) or open to trade (right graph), we can see that those closed to trade have a weak negative correlation between the two variables as shown by the slight downwards slope of the line of best fit, and while this may not be what we may have been expecting, this may just be due to an outlier where there is a country that did not spend any time a member yet has close to 30 billionaires, while those open to trade has a strong correlation as shown by the positive gradient of the line of best fit, but the strength of this correlation may not be very accurate as we can see that the variation of the number of billionaires a country has when they have been a member of GATT or WTO for over 55 years is very high. Furthermore, we can also see that there is no overlap in the number of years spent as member between open and close countries, and putting them both together, we can justify the moderately positive correlation between the two variables that we see in the middle graph as the strong positive correlation at the second half of the graph is more than enough to offset the weak negative correlation in the first half of the graph. We can therefore conclude that the longer a country is a member of GATT or WTO, the more billionaires the country may have, and this is understandable as people who have been in these countries from the beginning of their memberships would have a higher chance of becoming billionaires having been able to make use of the benefits of the lower trade barriers for a longer time.

**Question 5**



We can see that the t-stat that the mean difference in number of billionaires between countries open to trade and countries that are now is very high at 3.4822, which may be due to the large difference of 6.887179 between the two means, and this results in a very low p-value of 0.0009228, which means that the probability that we get a mean difference of 6.887179 when the true mean difference is equal to 0 is very low, and that is why we reject the null that the mean number of billionaires for countries open to trade is equal to the mean number of billionaires for countries closed to trade. The 95% confidence interval of [2.932742, 10.841617] implies that the true difference of the means has a 95% chance of lying within that interval, and since 0 is not included in this interval, it adds on to the reason as to why we rejected the null. This can imply that those open to trade should on average have a higher number of billionaires compared to those closed to trade.

**Question 6**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Billionaires and country trade openness** | | | | | |
|  | **(1)** | **(2)** | **(3)** | **(4)** | **(5)** |
| (Intercept) | -0.89 | 0.60 | -1.47 | -1.25 | -1.85 |
|  | (1.49) | (1.52) | (2.86) | (2.28) | (2.26) |
| gattwto | 0.19\*\* | 0.15\*\* | 0.14\*\* | 0.08\* | 0.07\* |
|  | (0.04) | (0.04) | (0.04) | (0.04) | (0.03) |
| roflaw |  | 2.82\*\* | 2.58\*\* | 3.15\*\* | 1.35 |
|  |  | (0.91) | (0.96) | (0.77) | (1.11) |
| topint |  |  | 0.07 | 0.06 | 0.02 |
|  |  |  | (0.08) | (0.06) | (0.06) |
| pop |  |  |  | 0.05\*\* | 0.05\*\* |
|  |  |  |  | (0.01) | (0.01) |
| gdppc |  |  |  |  | 0.14\* |
|  |  |  |  |  | (0.06) |
| Adj. R2 | 0.14 | 0.20 | 0.20 | 0.49 | 0.51 |
| Num. obs. | 119 | 119 | 119 | 119 | 119 |
| F statistic | 20.73 | 15.89 | 10.81 | 29.42 | 25.32 |
| *Dependent variable is numbil0, the number of billionaires in the country.*  *Homoscedastic standard errors in parentheses.*  *Statistical significance from two-sided tests of the null of no effect marked as \* for 5% and \*\* for 1%* | | | | | |

**Question 6a**

While most of the coefficient estimates that we see in the table are very small (<1) which is understandable seeing how few billionaires there truly are in the world, and, since such small values do not make much sense in the real world, we would usually have to scale it by 100, some of the intercepts are negative numbers and it is impossible for a country to have a negative number of billionaires, but this can be overseen when we see that countries will definitely not have zero coefficients for all the regressors, and this is why it is more understandable to have multiple regressors in a linear regression model as many regressors would not only be correlated to the dependent variable, but also to one another, such as the number of years spent as a member of GATT or WTO and the population, which had the biggest effect on the decrease on the coefficient of the number of years spent as a member. This is an example of an omitted variable bias, and this stands to show that the number of billionaires a country has is not just dependent on the number of years the country was a member of GATT or WTO but also on the population as well as many other variables. Having scaled the coefficients, we can interpret these coefficients as how many more billionaires a country would be expected to have if we increased the variables, ceteris paribus. For example, an increase of 100 million people in the population should have an increase of 5 billionaires. However, certain variables should not be scaled by 100 such as the number of years a country was a member of GATT or WTO as it does not make much sense to do so, so we should scale this by a more appropriate scale such as 10 or 50. We can also see that while all the coefficient estimates seem to decrease the more regressors we add, the rule of law index coefficients also seems to be decreasing until the population variable was introduced, which would imply a negative correlation between the two which is understandable as a higher population would imply more crimes that could happen.

**Question 6b**

The coefficient of 0.05 for pop with a homoscedastic standard error of 0.01 means that for every 1 million people increase in the population of a country, ceteris paribus, the number of billionaires the country has should increase by 0.05 where this may have a standard error of 0.01 billionaires, but, since 0.05 and 0.01 billionaires does not make much sense in the real world, it is better to interpret it as a 5 billionaires increase for every 100 million population increase, which has a standard error of 1 billionaire from one country to another, and this can be justified as an increase in the population would signify the increase of business opportunities as a business would be able to increase sales which would also encourage more businesses to open, thus creating more revenue and profits. On the other hand, the coefficient of 0.14 for gdppc with a standard error of 0.06 which means that if the GDP per capita were to increase by US$1000, ceteris paribus, then the number of billionaires would increase by 0.14 with a standard deviation of 0.06, and by scaling this, we get an increase of 14 billionaires with a standard deviation of 6 billionaires for every US$100 000 increase in the GDP per capita, ceteris paribus, and this can be linked to higher earnings and higher savings which would ultimately lead to some people becoming billionaires.

**Question 6c**

The adjusted R-squared of 0.51 means that 51% of the variation in the number of billionaires in a country is explained by the regressors present in the model, which in case is the country's number of years as a member of GATT or WTO, rule of law index, top marginal income tax rate, population, and GDP per capita, and looking at the other adjusted R-squares from the different models, they can be seen mostly increasing as we add more regressors which means that we are reducing the omitted variable biasness that could be found when we only regress the number of billionaires solely on the number of years spent as a member, and stands to show that each of the individual regressors not only has a correlation with the number of billionaires a country has but also with one another. However, the adjusted R^2 did not increase when adding the top marginal income tax rate, and this would mean that the correlation between the tax rate and the other regressors as well as the number of billionaires is very weak, which is unlike when we added population which is proven to have the highest share of the variation by the largest effect on the adjusted R-squared, which is expected as there should be more billionaires when there are more people in a country.

**APPENDIX – R CODE**

# Set working directory using setwd() function

setwd("~/Downloads/ECOM20001/Assignment")

# Read data from a csv file using read.csv() function

mydata=read.csv(file = "billionaires\_clean.csv")

# Create variables for each table attribute simplify script

country = mydata$country

numbil0 = mydata$numbil0

open = mydata$open

gattwto = mydata$gattwto

gdppc = mydata$gdppc

pop = mydata$pop

roflaw = mydata$roflaw

topint = mydata$topint

# ==========================================================================================================================

# QUESTION 1

# ----------

# Summary statistics for the entire dataset using the summary()

# function and the standard deviation using the sapply( ,sd) function

summary(mydata)

sapply(mydata,sd)

# Summary statistics for all attributes of countries that are closed (open == 0) to trade and open (open == 1) to trade

summary(mydata[open == 0,])

sapply(mydata[open == 0,],sd)

summary(mydata[open == 1,])

sapply(mydata[open == 1,],sd)

# ==========================================================================================================================

# QUESTION 2

# ----------

# Density function of the number of billionaires in a country with a main title (assigner to 'main') and axes labels

# (assigned to 'xlab' and 'ylab'), where the x-axis is limited to the range of the number of billionaires (assigned to

# 'xlim') with the graph coloured red (assigned to 'col') using the density() function, and plotting it using the plot()

# function into a pdf for easy extraction using the pdf() function at the beginning and dev.off() function at the end

pdf("Density of Billionaires per Country.pdf")

plot(density(numbil0),

main = "Density of Billionaires per Country",

xlab = "Number of Billionaires",

xlim = c(min(numbil0), max(numbil0)),

ylab = "Density",

col = "red")

dev.off()

# Density function of the number of years a country was a GATT or WTO member with a main title and axes labels

# witht the range of the x-axis limited to the range of the number of years a country was a GATT or WTO member

pdf("Density of Years as a Member of GATT or WTO per Country.pdf")

plot(density(gattwto),

main = "Density of Years Spent as a GATT or WTO Member per Country",

xlab = "Number of Years",

xlim = c(min(gattwto), max(gattwto)),

ylab = "Density",

col = "blue")

dev.off()

# ==========================================================================================================================

# QUESTION 3

# ----------

# Density function of the number of billionaires in a country closed (open == 0) to trade in red vs

# a country that was open (open == 1) in blue (appended using the lines() function) with a main title,

# axes labels where the x-axis is limited to the range of the number of billionaires in a country,

# and a legend to help differentiate the graphs at the topright using the legend() function

pdf("Density of Billionaires per Country as per Trade Openness.pdf")

plot(density(numbil0[open == 0]),

main = "Density of Billionaires per Country as per Trade Openness",

xlab = "Number of Billionaires",

xlim = c(min(numbil0), max(numbil0)),

col = "red",

lty = 1)

lines(density(numbil0[open == 1]), col = "blue", lty = 1)

legend("topright", legend = c("Not open", "Open"), col = c("red", "blue"), lty = c(1,1))

dev.off()

# ==========================================================================================================================

# QUESTION 4

# ----------

# Scatter graph of the number of billionaires in a country against the number of years that

# country was a member of GATT or WTO where the dots are filled in (assigned to 'pch') with

# a main title, axes labels, and a line of best fit, calculated using a linear model of the

# two variables using the lm() function, appended to the graph using the abline() function

pdf("Distribution of Billionaires per Country based on Years Spent as a GATT or WTO Member.pdf")

best\_fit\_1 = lm(numbil0 ~ gattwto, data = mydata)

plot(gattwto, numbil0,

main = "Distribution of Billionaires per Country \nbased on Years Spent as a GATT or WTO Member",

xlab = "Number of Years",

ylab = "Number of Billionaires",

col = "red",

pch = 16)

abline(best\_fit\_1, col = "forestgreen", lwd = 2)

dev.off()

# Scatter graph of the number of billionaires in a country closed to trade against the number of years

# that country was a member of GATT or WTO with a main title, axes labels, and a line of best fit

pdf("Distribution of Billionaires of Countries Closed to Trade based on Years Spent as a GATT or WTO Member.pdf")

best\_fit\_2 = lm(numbil0[open == 0] ~ gattwto[open == 0], data = mydata)

plot(gattwto[open == 0], numbil0[open == 0],

main = "Distribution of Billionaires of Countries Closed to Trade \nbased on Years Spent as a GATT or WTO Member",

xlab = "Number of Years",

ylab = "Number of Billionaires",

col = "red",

pch = 16)

abline(best\_fit\_2, col = "forestgreen", lwd = 2)

dev.off()

# Scatter graph of the number of billionaires in a country open to trade against the number of years

# country was a member of GATT or WTO with a main title, axes labels, and a line of best fit

pdf("Distribution of Billionaires of Countries Open to Trade based on Years Spent as a GATT or WTO Member.pdf")

best\_fit\_3 = lm(numbil0[open == 1] ~ gattwto[open == 1], data = mydata)

plot(gattwto[open == 1], numbil0[open == 1],

main = "Distribution of Billionaires of Countries Open to Trade \nbased on Years Spent as a GATT or WTO Member",

xlab = "Number of Years",

ylab = "Number of Billionaires",

col = "red",

pch = 16)

abline(best\_fit\_3, col = "forestgreen", lwd = 2)

dev.off()

# ==========================================================================================================================

# QUESTION 5

# ----------

# Two-sample t-test for the null that the mean number of billionaires in countries open to trade (open == 1) is

# equal to the mean number of billionaires in countries closed to trade (open == 0) using the t.test() function

t.test(numbil0[open == 1], numbil0[open == 0])

# ==========================================================================================================================

# QUESTION 6

# ----------

# Single linear regression model of the number of billionaires in a country (regressand)

# based on the number of years the country was a member of GATT or WTO (regressor)

reg1 = lm(numbil0 ~ gattwto, data = mydata)

# Multiple linear regression model of the number of billionaires in a country (regressand) based on (regressors):

# • Number of years as a member of GATT or WTO (gattwto)

# • Rule of law index (roflaw)

reg2 = lm(numbil0 ~ gattwto + roflaw, data = mydata)

# Multiple linear regression model of the number of billionaires in a country (regressand) based on (regressors):

# • Number of years as a member of GATT or WTO (gattwto)

# • Rule of law index (roflaw)

# • Top marginal income tax rate (topint)

reg3 = lm(numbil0 ~ gattwto + roflaw + topint, data = mydata)

# Multiple linear regression model of the number of billionaires in a country (regressand) based on (regressors):

# • Number of years as a member of GATT or WTO (gattwto)

# • Rule of law index (roflaw)

# • Top marginal income tax rate (topint)

# • Population (pop)

reg4 = lm(numbil0 ~ gattwto + roflaw + topint + pop, data = mydata)

# Multiple linear regression model of the number of billionaires in a country (regressand) based on (regressors):

# • Number of years as a member of GATT or WTO (gattwto)

# • Rule of law index (roflaw)

# • Top marginal income tax rate (topint)

# • Population (pop)

# • GDP per capita (gdppc)

reg5 = lm(numbil0 ~ gattwto + roflaw + topint + pop + gdppc, data = mydata)

# Footnote to explain the table of reports (below)

custom\_note = "Dependent variable is numbil0, the number of billionaires in the country.

Homoscedastic standard errors in parentheses.

Statistical significance from two-sided tests of the null of no effect marked as \* for 5% and \*\* for 1%"

# Install the texreg package as to use the screenreg function (below) using the library() function

library("texreg", lib.loc="/Library/Frameworks/R.framework/Versions/3.5/Resources/library")

# Print the coefficients of all the linear regression models (compiled in a list using the list() function) in

# a table form using the texreg() function where each column is numbered from (1) through (5) (assigned to

# 'custom.model.names') as to allow comparison between linear regression models as to help determine omitted

# variable biasness as well as performing two-sided t-tests on the coeffients at a 5% and 1% significance level

# using stars to denote the rejection of null of the coefficient being equal to 0 using stars (assigned to 'stars'),

# reporting the goodness of fit statistics at the bottom such as the number of observations, adjusted R^2, and

# the regression F-statistic (assigned to 'include.fstatistic') while not including the rmse (assigned to 'include.rmse')

# and the R^2 (assigned to 'include.rsquared') with the footnote at the bottom of the table (assigned to 'custom.note')

# and the table title (assigned to 'caption' and 'caption.above') and put them all in a file (assigned to 'file')

htmlreg(list(reg1, reg2, reg3, reg4, reg5),

file = "Regressions.doc",

caption = "Billionaires and country trade openness",

caption.above = TRUE,

stars = c(0.01, 0.05),

custom.note = custom\_note,

include.rsquared = FALSE,

include.rmse = FALSE,

include.fstatistic = TRUE,

custom.model.names = c("(1)", "(2)", "(3)", "(4)","(5)"))

# ==========================================================================================================================