Regression and Multivariate Data Analysis Assignment

Determinants of Economic Growth

Final Report

Business Analytics Specialization

Prof. Drº. Pedro Silva

20/05/2024

Alexandre Brito – 355423006

Francisco Nieto Guimarães – 356321002

José Miguel Guedes - 355423064

Table of Contents

1. Introduction

During the Regression and Multivariate Data Analysis course, lectured in the 2nd semester of the 1st year in the Master in Management, the class was given tools, procedures, and both general and in-depth knowledge regarding the appropriate interpretation and management of data analysis. The proposed report aims to affirm concretely the previous statement, in order to further solidify the usage of data as an asset that grants competitive advantage via leveraging value from itself. The importance of the analysis lies in the fundamental aspect of comprehending the correlative behavior between variables and factors that change the result of GDP growth. The appointed time period (1960-1985) makes it so that, regardless of the country it's analyzed. To achieve the intended goal, the report will begin with cleaning and processing the data. Following up, a multiple linear regression model will be used to measure the strength of the relationship between each variable. To further strengthen the group's goal, numerous tests shall be conducted to ensure maximum model quality, relevance, and pertinence of conclusions made.

By analyzing the results of the complete model, it will be possible to compute the most significant factors affecting the general economic growth of the time period, as well as provide recommendations and commentary regarding the final proposed model.

2. Variable Analysis

Before we analyze the variable is in an individual way, all the records that had NA were removed and the attributes were characterized by numerical or factor depending on their nature. The analysis of the variables was based on values obtained by the *summary ()* function and are explicit in figure 19 of the annex. Due to the large part of values removed for not presenting values having in the oil attribute the result yes, it had therefore a drastic decrease in the sample of countries that are oil producers. Thus, due to the significant sampling difference of data from oil-producing countries compared to non-oil-producing countries, we cannot make a reliable comparison about the relationship of variables when there is or is not oil production. Therefore, a specific analysis of the numerical variables with the influence of oil production was not made.

The following graphs (figure 1) present the relationship between *gdp60* and *gdpgrowth* taking into consideration other factors such as if the production of oil is present if it possesses an OECD membership and if the country has stronger and more reliable data quality.

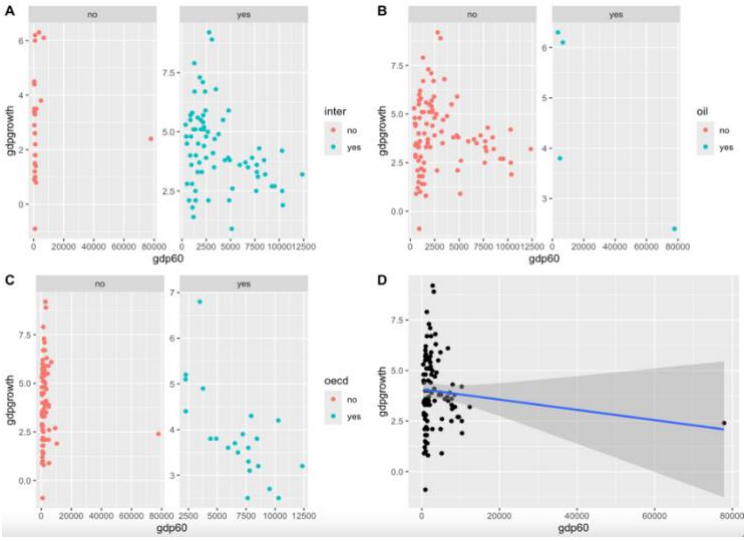

Figure 1 - Relationship between gdp60 and gdpgrowth

Analysing the set of graphs referring to the *gdp60* attribute it is possible to notice the presence of an extreme value and therefore an outlier influenced the structure of the graphs. In graph A, there is a higher incidence of data for countries with good data quality (*inter = yes*), having a wider distribution of values between 0 and 12500 dollars and generally higher GDP growth rates. On the other hand, countries lacking quality data generally had lower GDP *per capita* in 1960, with much of the data concentrated from 0 to 2000 dollars. In graph C, it is possible to verify a behaviour very close to graph A, in which OECD members countries, have higher values of *gdp60* and *gdpgrowth*. Graph D shows that there is a negative correlation between these two variables. This suggests that when GDP increases, GDP growth decreases. The standard deviation is 8005.468973, indicating that the GDP60 values are spread out over a wider range. Regarding density, it is possible to verify from it (figure 2), that most of the data are close to the first quartile, this having a value of 1001. Thus, the median (1945) is lower than the mean (3837) and therefore, we can state that the distribution is right skewed (positive skew). The third quartile in turn has a value of 4776, resulting in an inter-quartile distance of 3775.250.

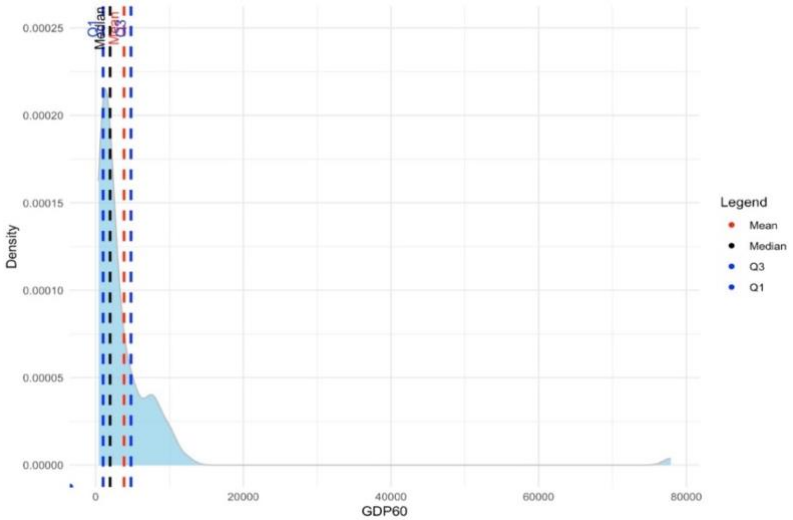

Figure 2 - Density plot of gdp60

The graphs (figure 3) that follow illustrate the relationship between *gdp85* and *gdpgrowth*. Graph A shows that countries with stronger and more reliable data may be associated with higher GDP *per capita* in 1985, as well as with an increase in the GDP growth rate. Chart C suggests that OECD countries generally have higher GDP and more stable GDP growth rates. Regarding graph D, there is a slightly positive correlation between the two variables due to data dispersion. Finally, there is a clear higher data density (figure 4) near the first quartile (Q1 = 1182). Based on this information, the distribution is also skewed to the right, resulting in an average higher than the median by about 2150 units. The third quartile has a value of 7719 creating an inter-quartile value of 6536.5.

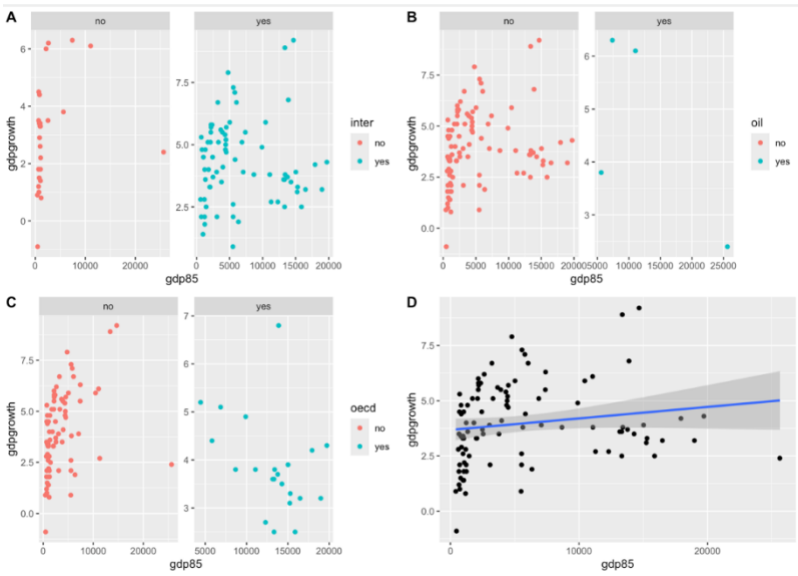

Figure 3 - Relationship between gdp85 and gdpgrowth

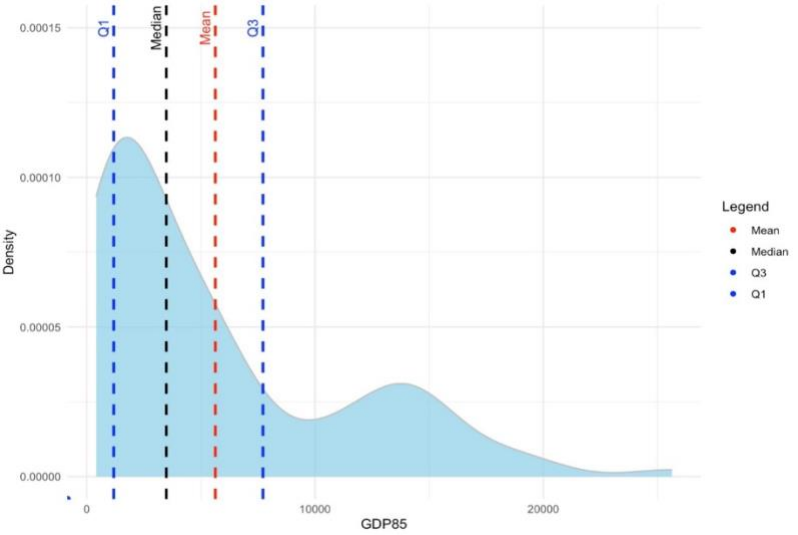

Figure 4 - Density plot of gdp85

The following graphs (figure 5) show the relationship between *gdpgrowth* and *popgrowth*. According to chart A, countries without good data quality have population growth rates between approximately 1.5% and 2.5%, with GDP growth rates ranging from 0% to about 6%. While other countries have a wider range of population growth rates, about 1% to 4%, and GDP growth rates of 0% to 7.5%. About graph C, we can see that non-OECD countries are more concentrated between a population growth rate between 2% and 4%, while OECD countries have lower population growth rates. Analysis of graph D shows that these variables have a positive correlation and suggests that higher GDP growth is mainly associated with higher population growth. Finally, regarding density (figure 6), it is possible to verify that the highest data density is close to the third quartile, with this result of 2,900. The value of the mean and median are very close to each other; however, the

median is higher having a value of 2.274 and 2.400 respectively. Thus, the distribution has a negative skew. In conclusion, the first quartile has a value of 1.7 and the inter-quartile result is 1.2.

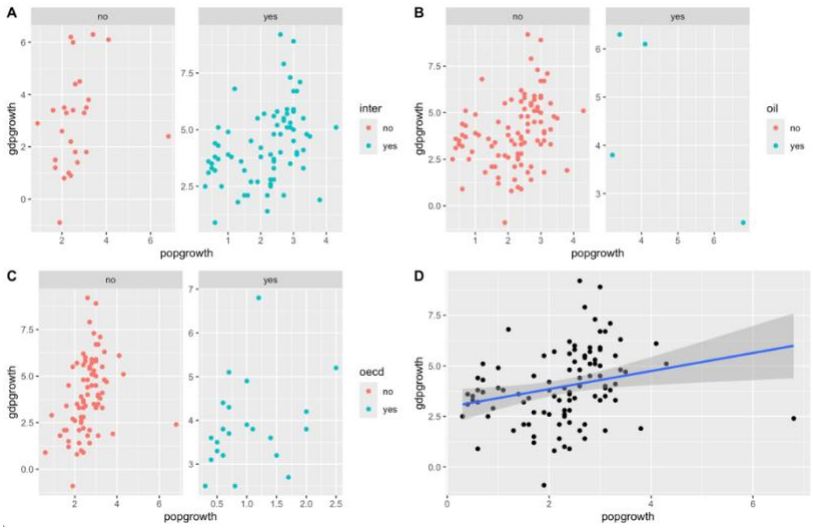

Figure 4 - Relationship between popgrowth and gdpgrowth

Discussing the relationship between *gdpgrowth* and *invest* (figure 10), the image concludes that countries with additional data quality invest more than the ones with fewer data quality. According to graph B, oil-producing nations don't invest as much as non-oil producing countries, with non-oil producing states possessing a wider range of *invest* and *gdpgrowth* rates. Graph C displays *invest* values spanning from about 5 to 30 and *gdpgrowth* rates vary between 0% to 7,5% in non-OECD countries. Regarding OECD countries, the invest values are higher, and as Graph D shows these variables have a positive correlation, implying that higher investments lead to higher GDP growth. According to the density plot of *invest*, we concluded that the mean is higher than the median, 17.43 and 16.55, respectively. From figure 5, we can see that there is a higher density near the first quartile.

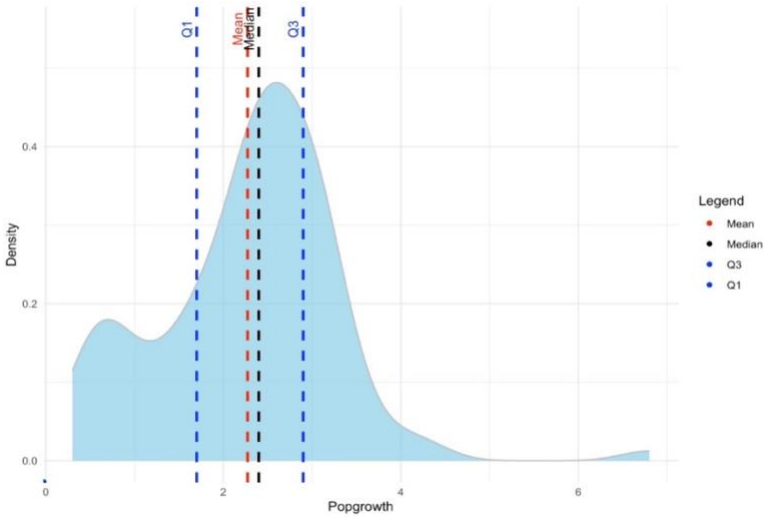

Figure 5 - Density plot of popgrowth

Addressing *gdpgrowth* and *literacy60*, the graphs (figure 9) dictate that countries with proper quality data have their data points more distributed, indicating that nations with a bigger percentage of population over 15 years of age that can read and write have a bigger GDP growth rate. Graph C shows that almost all OECD countries have 100% of people over 15 years who can read and write, while in non-OECD nations the data points are much more dispersed. Graph D exhibits that these variables have a slightly positive correlation, suggesting that higher rates of literacy lead to increased GDP growth rates. The value of the first quartile is 15 and the value of the third is 84. The mean has a value of 49.43, which is higher than the median that presents a value of 46. As it is possible to verify from the density analysis (figure 13), there is

a higher density of the data in the first quartile and after the third quartile, and between the two quartiles the density diminishes.

Regarding the relationship between *gdpgrowth* and *school* (figure 8), countries that display superior quality data present generally possess a higher fraction of the working-age population enrolled in high school. According to graph C, in non-OECD countries display a wide range of *gdpgrowth*, from 0% to around 7,5% with a broad range fraction of the working-age population enrolled in secondary school. The OECD states exhibit a lower range of '*gdpgrowth'* and *school'*, with the last graph unveiling a small but positive correlation between the variables, insinuating that an increased fraction of the working-age population enrolled in high school leads to greater GDP growth. On the topic of density (figure 12), distribution of the average fraction of working-age population enrolled in high school shows that the value of the first quartile is 2.4, the median is 4.85, and the third quartile is 8.075. As such, the inter-quartile distance is 5.675, with the mean being higher than the median, assuming a value of 5,453, with a higher data density being present near the first quartile.

Finally, mentioning the dependant variable (figure 14), its distribution demonstrates once again that the mean is higher than the median. From the minimum value of -0.9 for the first quartile there is a difference of 3.6, and on the other hand, the difference from the maximum value for the third quartile is 4.075. The quartile difference is 2,425. Since the first quartile has a value of 2.7 and the median has a result of 3.8, there is a higher incidence of data near the first quartile.

To analyze the relationships among variables, a correlation matrix (figure 7) was conducted. This analysis was crucial for identifying strong and potential relationships between the independent variables and the dependent variable, with it indicating that population growth and investment are the factors that have the greatest impact on GDP growth. Even though *gdp60* has a small correlation, it is the only variable that leads *gdpgrowth* to a decrease. All the other independent variables have a correlation higher than zero, which means that when GDP growth increases, they increase too, in different proportions. It was also discovered that some great relationships between independent variables, such as *school* and *literacy60*, with the highest correlation of 0.81; and *gdp85* and *literacy60*, with a

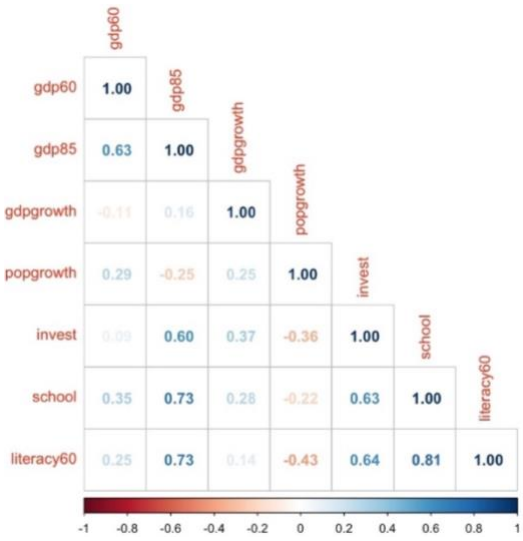

Figure 7 - Correlation Matrix

0.73 correlation. To build an accurate model, the independent variables can't be very correlated, which represents a need to avoid multicollinearity.

3. Regression Model Development

From a theoretical point of view, the measures used to check the quality of the models and compare them were the adjusted R-squared and the AIC. The hypothesis tests used to study both the individual significance of the coefficients and the global significance of the model had as a significance level the value of 0.05, so if the p-value assumes values less than 0.05, the null hypotheses are rejected.

Initially, two first models were created, one with only the numerical variables and the other adding the binary variables. The objective was to analyze the impact of the dummy in the first regression model, and therefore, we analyzed the impact of the *oil, oecd* and *inter* attributes separately.

Starting with the impact of the *oil* variable, when it presents a true value, there are no values in the *invest, school* and *literacy60* coefficients. As it is possible to verify in Table 1, there is a great impact when the variable *oil* assumes values 0 (no), being its coefficients, in general, higher than those existing when *oil* assumes value 1 (yes). The main change, in addition to adding the 3 variables, was in *popgrowth* registering the largest increase, going from a negative coefficient to a positive coefficient. In relation to the impact of the variable *inter*, most of the variables, except for *literacy60 and invest*, which increased, and *gdp85*, which remained constant, there was a decrease in value in the coefficients when inter assumes values of 0 compared to when it assumes values of 1 (Table 2). Finally, the OECD binary variable also has an increase in the coefficients when it assumes the value no (Table 3), however, it is not so significant compared to the previous categorical variables analyzed. To finalize the analysis of the impact of dummy variables in the regression *Mod0reg*, based on Graph 17, it is possible to highlight that the variable that has a considerably greater impact is inter.

That said, the *Mod1reg* is created with all variables, presenting an adjusted R-squared of 0.4994 higher than the *mod0reg* adjusted R-squared of 0.4712. Continuing with the analysis of the function *summary(Mod1Reg)* it is possible to verify that the variables *school, oil and OECD* present an individual significance test (p-value) higher than the significance level. As

such, these variables are considered statistically non-significant and so, *Mod2reg* was created without these three variables.

Moving now to the analysis of the third model, again using the summary function, it is possible to obtain an improvement in the adjusted R-squared (0.5014), becoming a better model compared to the rest. However, the variable *literacy60* was no longer theoretically statistically significant in the model, as a result of a p-value greater than 0.05. Thus, Mod3reg was created by removing the variable *literacy60*, nevertheless, the model lost quality since the adjusted R-squared decreased having the lowest value compared to Mod1reg and *Mod2Reg*, being 0.4899. So, in order not to lose value in adjusted R-squared and AIC, it was decided that we will keep the variable *literacy60*.

Therefore, model *Mod4Reg* was created, being the optimization of the coefficients (*step() function*) of the model with better adjusted R-squared and lower AIC, hence *Mod2Reg*. *Mod4reg* has as independent *variables GDP60, GDP85, inter, invest and literacy60*. Finally, to obtain the final model, the impacts of the outliers in the *Mod4Reg* model were identified and analyzed. From the Graph 15 and 16, it was possible to obtain a view that the records of line 47 are a possible outlier and the impact of this line on the regression. Applying the outlier test and comparing with the records near line 47, it was proven that in fact it is an outlier.

After removing the values considered as outliers and repeating the tests to verify the presence and impact of other possible outliers, the *Mod5res* is obtained. Analyzing the model *Mod5reg* the variable *literacy60* presents a p-value quite far from 0.05. In fact, analyzing the adjusted R-squared with and without this variable, it was found that the maximization of adjusted R-squared is obtained by removing this variable, ceasing to be significant for the model. As such, the *Mod6Reg* model was created without *literacy60*, with the highest adjusted R-squared value obtained compared to the previous models, having a value of 0.7146.

4. Final Model

$$GDPGrowth = -3.807e-01 - GDP60 * 7.761e-04 + GDP85 * 4.738e-04 + popgrowth * 1.227 + invest * 4.546e-02 + inter * 8.854e-01$$

The coefficient *gdp60* is the only coefficient that has a decreasing impact on *gdpgrowth*, and when it increases by one unit, GDP decreases by an average of 0.3807 units. Regarding the only binary variable in the model, countries that have higher data quality have an average *gdpgrowth* higher than countries that do not have better data quality by 0.8854. Looking at a

more detailed perspective for each coefficient, the variable that has a greater impact on *gdpgrowth* is *popgrowth*, since when *popgrowth* increases by one unit, *gdpgrowth* increases, on average, by 1.227 keeping the other variables constant. The variable that has the least impact is *gdp85* and when it increases one unit, *gdpgrowth* grows by an average of 0.0004738 units. Finally, when the variable *invest* increases by one unit, GDP grows by an average of 0.04546 units while keeping the remaining variables constant. The values of the coefficients are among the confidence intervals reported in Table 3. Respecting the individual significance test, all variables of the model are considered statistically significant for the prediction of *gdpgrowth*, reflecting a p-value lower than the significance level. The model possesses variables with substantially low numbers, ranging from 4.21e-16 to 0.012732 which, to a predetermined significance level of 0.05, reflects strong evidence against the null hypothesis, indicating that the independent variables have a wide and significant effect on the dependent variable gdpgrowth, and as such, a very positive relationship in-between them in the model.

Addressing the R-squared of the model, the value obtained is 0.7292 on the 'Multiple R-squared', and of 0.7146 on the adjusted R-squared, a significant increase on both indicators comparatively from the previous analysis made in April. These results can be interpreted as an indicator of the fact that around 71,46% to 72,92% of the variability in the dependent variable *gdpgrowth* is explained by the independent variables presented. The values obtained are well above the middle line and near the upper quartile, justifying with solid strength that the variables starting at *gdp60* all the way to *inter* explain to a vast extent the variability of *gdpgrowth*, representing a strong effectiveness. Complemented with the AIC value, being another method to assess the quality of the model, the lowest possible was obtained compared to the previously explored models, being 282.9347 units.

In conclusion, on the topic of F-statistic p-value, the regression model has a substantially low value, assuming a result less than 2.2e-16 which indicating that the model is globally significant.

5. Regression Model Results

Regarding the hypothetical violation of the five assumptions of linear regressions, none have been breached on the final developed model, and thus, it's the model that best fits the data presented. Complementing the previous statement, this further validates the regression model and the reliability of the estimates and inferences drawn from it, allowing the group to have

confidence in the conclusions drawn in the report's analysis. Going further into the details and specifics of these results, a succinct explanation and demonstration of the values obtained in each verification of the assumptions will be provided.

Addressing linearity, the changes in the model's dependent variable are proportional to the same impacts in each of the independent ones, as can be seen in the relevant scatterplots provided in the script, which show that there are no clear patterns in them. Adding to this, a straight line can be seen when plotting the residuals against the fitted values or against each independent variable. With this, we confirm the first assumption.

On the topic of independence of errors, the residuals are, in fact, independent of each other, as can be seen due to the lack of autocorrelation between them. By plotting the residuals against fitted values, as previously mentioned on the first assumption, we verify that there are no patterns identified. Utilizing the Durbin-Watson to test it in the residuals, its value comes up to 2.0871, as well as high p-value of 0.6246, which further indicates the lack of autocorrelation being present (had it been significantly closer to 4 or 0, it would indicate a positive or negative autocorrelation). All said and done, there is no significant evidence to reject the null hypothesis of no autocorrelation.

The third assumption of homoscedasticity is also verified, with the variance of the errors being constant across all levels of the independent variables. Despite being present in the start of the analysis, after being identified with the Breusch-Pagan test, its correction was tested via the adoption and usage of Robust Standard Errors. After conducting said test, the group arrived at a 'Prob > Chi2' value of 0.2202, and based on this p-value, there is no strong indication of heteroscedasticity in the residuals, and the assumption of homoscedasticity holds for the regression model developed.

Discussing the normality of errors, they are distributed accordingly. Via examination of the relevant histograms, normal Q-Q plot analysis and conducting the Shapiro-Wilk test of the residuals, it can be visually verified that the points on the plot fall approximately along a straight line, translating in them being normally distributed. Adding to this, the SW value of 0.98802 with a p-value of 0.5171 being greater than the chosen significance level, there is a failure in rejecting the null hypothesis, suggesting that the data is normally distributed.

Finally, regarding the fifth and final assumption of the multicollinearity issue, there should be no perfect linear relationship between independent variables in the final model, and the ideal status occurs when one or more of these can be exactly predicted from the others. Upon arriving at the VIF values of 4.775068 for *gdp60,* 7.088974 for *gdp85,* 1.343647 for

popgrowth, 2.035165 for *invest* and 1.222162 for *inter* and considering that VIF being close to 1 is the ideal scenario, lower than 5 indicating low multicollinearity and between 5 and 10 indicating moderate multicollinearity, the group was able to surmise that only *gdp85* represented a weaker output. Nevertheless, with additional research, the conclusion made was that despite the value being quite higher than its peers, its effects weren't deemed problematic in regard to the quality of the model nor its results, as it didn't generate any unstable estimates of the regression coefficients.

6. Conclusion

The fundamental aspect of the elaboration of the present report lied in the proper data management of the samples, model construction and respective analysis of the results. Following the intermediary report feedback, as well as delving deeper into personal research and the course's fundamental teachings, adequate adjustments were made to support both the theoretical and practical dimensions of the proposed model.

Based on the 'Determinants of Economic Growth' dataset, in order to explain and try to predict GDP growth with the other variables, an analysis of variables and construction of the model was created, with the strongest and most reliable model being the 'Mod6Reg' with coefficients *gdp60*, *gdp85*, *popgrowth*, *invest* and *inter*. In general, the model becomes useful for predicting the outcome due to expressing a adjusted R-squared of 0.7146. The main attribute influencing GDP growth was popgrowth. As this variable reflects the average growth rate of the working-age population from 1960 to 1985, its impact was to be expected since the working population is a source of wealth creation for a country. On the other hand, the impact of the invest variable was a surprise for the group, since investment is essential for the growth of the group.

The model's ability to produce reliable and interpretable results has significantly contributed to the understanding of the macro economical phenomena under study of a predetermined set of countries' behavior and ability to propel its GDP. By fulfilling key assumptions and yielding solidpredictive accuracy, 'Mod6Reg' has demonstrated its efficacy in uncovering meaningful insights, identifying correlation strength, and calculating statistical relevance. Despite all this, moving forward, further research could explore potential extensions or adjustments to the model, as well as its application to different datasets or scenarios due to the utility obtained through the model's versatility and adaptability.

7. Attachments

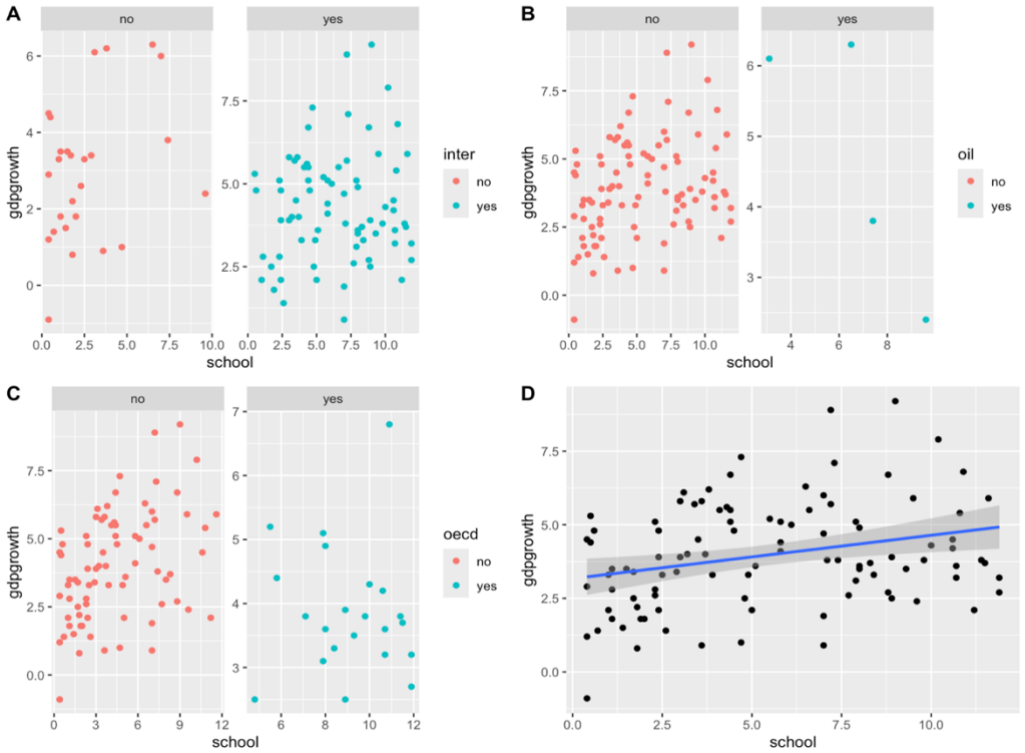

Figure 8 - Relationship between school and gdpgrowth

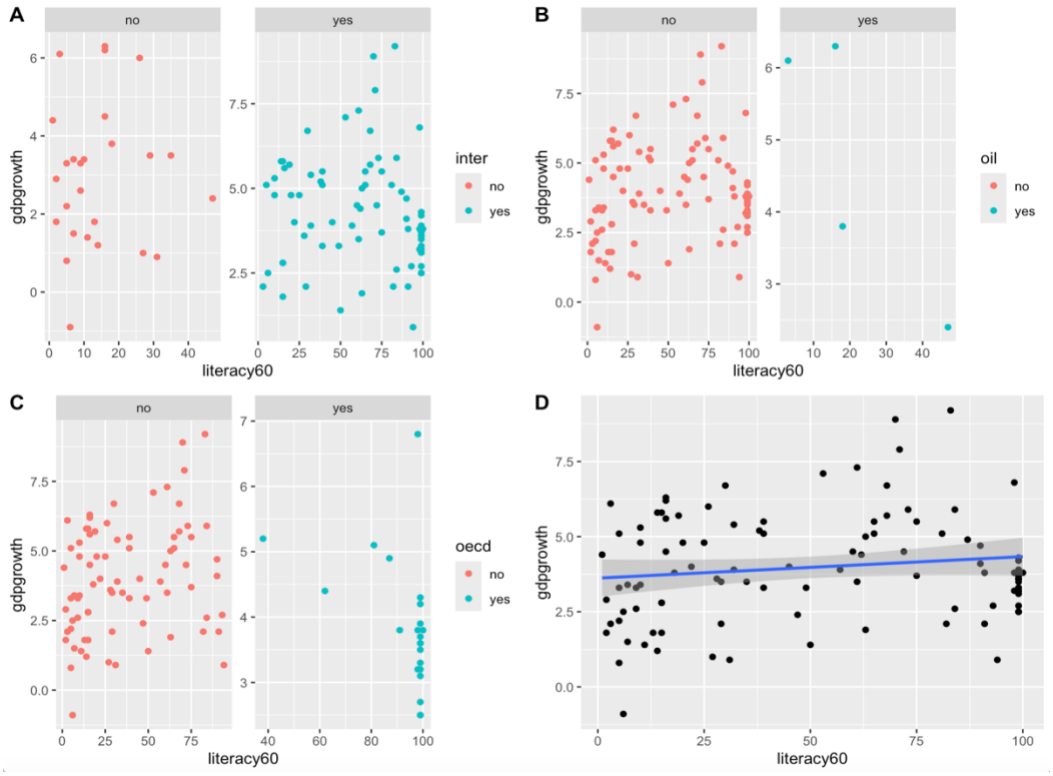

Figure 9 - Relationship between literacy60 and gdpgrowth

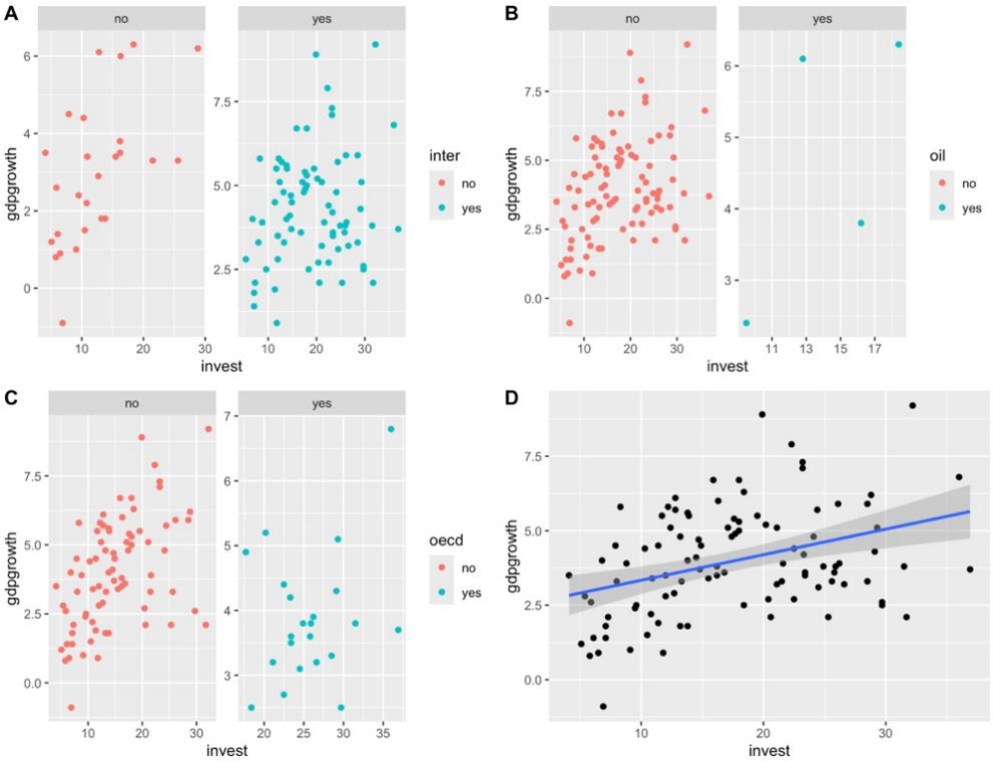

Figure 10: Relationship between invest and gdpgrowth

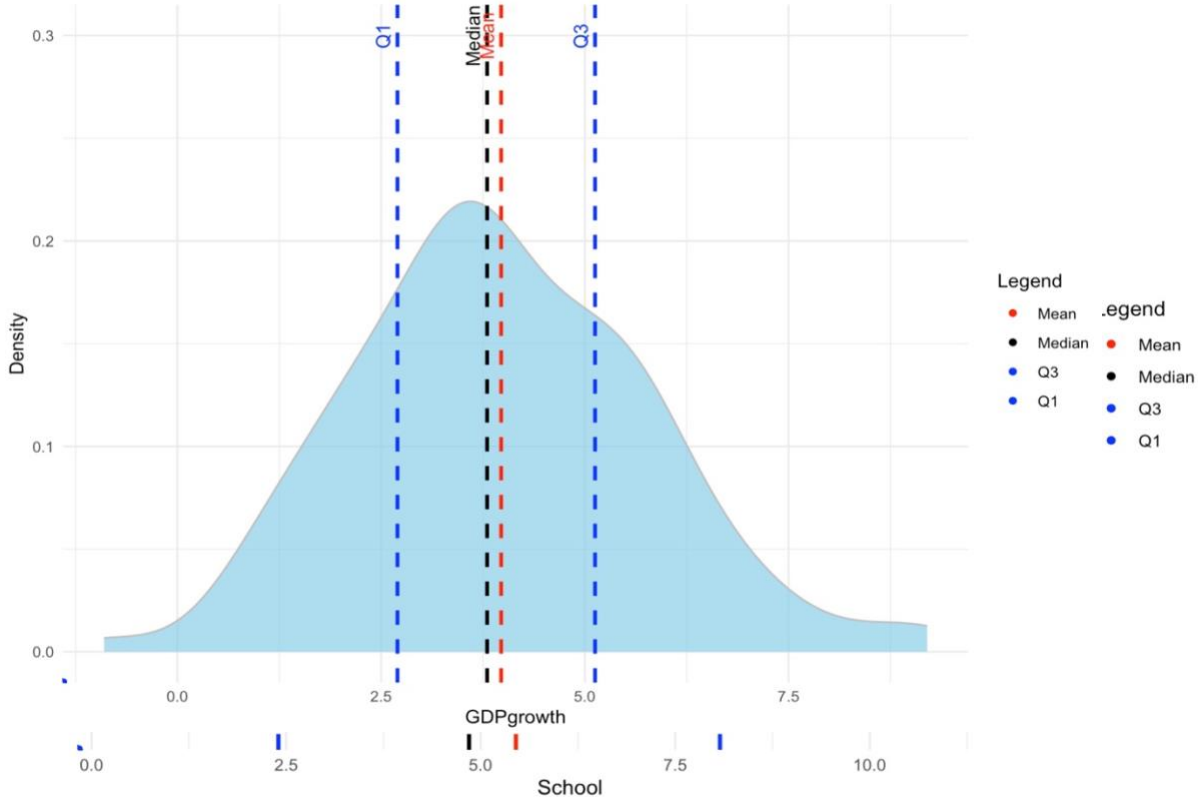

Figure 11 - Density plot of School

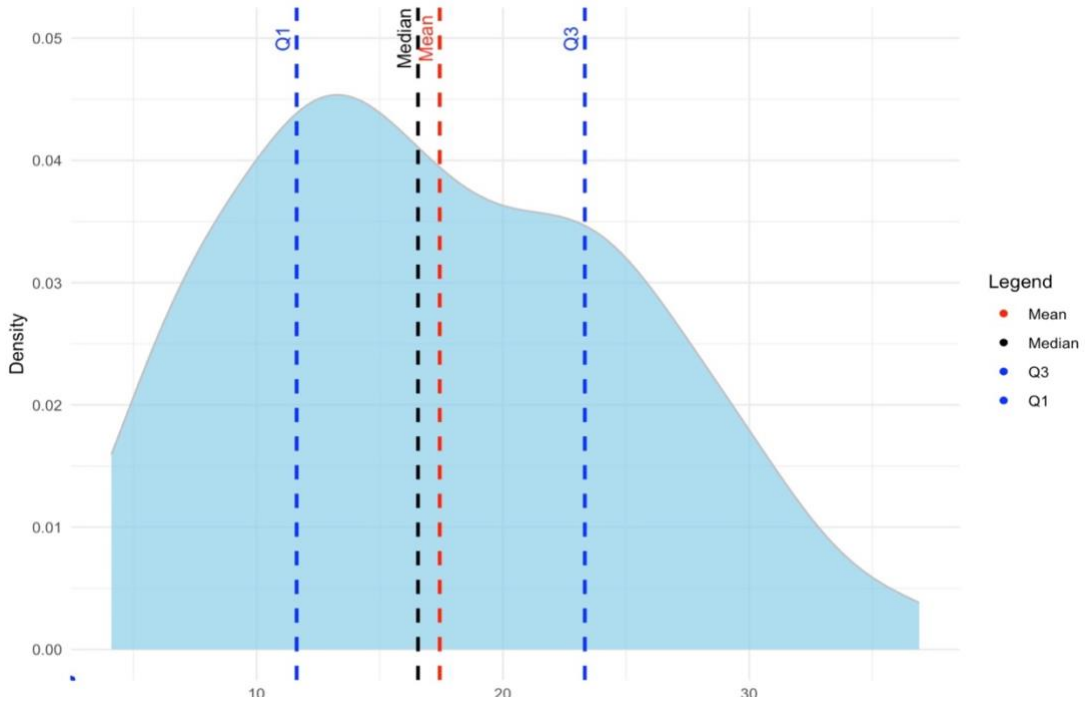

Figure 12 - Density plot of Invest

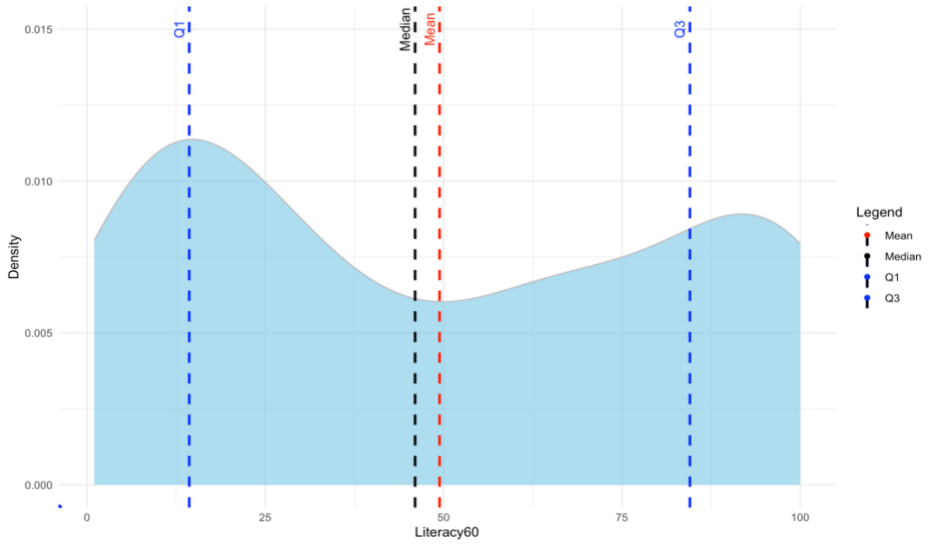

Figure 13 - Density plot of Literacy60

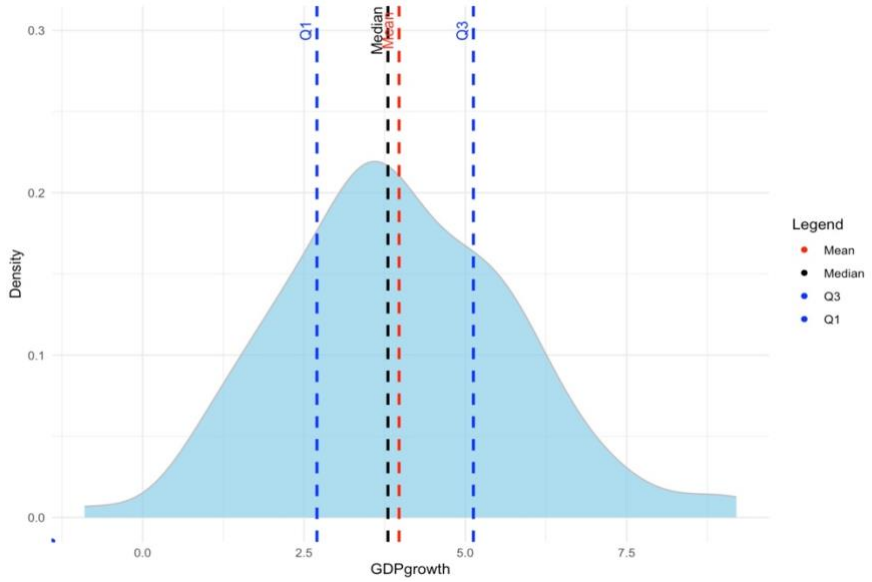

Figure 14: Density plot of gdpgrowth

Table 1 - Impact of the variable oil in the numeric variables

Oil = Yes						
(Intercept)	*gdp60*	*gdp85*	*popgrowth*	*invest*	*school*	*literacy60*
4.036e+01	-7.576e-05	3.278e-03	-1.707e+01	NA	NA	NA
Oil = No						
(Intercept)	*gdp60*	*gdp85*	*popgrowth*	*invest*	*school*	*literacy60*
-0.1137503	-0.0008000	0.0004463	1.2045604	0.0420153	0.0957114	0.0040071

Table 2 -Impact of the variable inter in the numeric variables

Inter = Yes						
(Intercept)	*gdp60*	*gdp85*	*popgrowth*	*invest*	*school*	*literacy60*
0.9949831	-0.0007759	0.0004517	1.1344665	0.0144767	0.0757683	-0.0005714
Inter = No						
(Intercept)	*gdp60*	*gdp85*	*popgrowth*	*invest*	*school*	*literacy60*
-0.2857293	-0.0001724	0.0004528	0.3058394	0.1302851	-0.0474910	0.0323280

Table 3 -Impact of the variable inter in the numeric variables

Inter = Yes						
(Intercept)	*gdp60*	*gdp85*	*popgrowth*	*invest*	*school*	*literacy60*
0.9949831	-0.0007759	0.0004517	1.1344665	0.0144767	0.0757683	-0.0005714
Inter = No						
(Intercept)	*gdp60*	*gdp85*	*popgrowth*	*invest*	*school*	*literacy60*
-0.2857293	-0.0001724	0.0004528	0.3058394	0.1302851	-0.0474910	0.0323280

Table 4 - Impact of the variable *oecd* in the numeric variables

OECD = Yes						
(Intercept)	*gdp60*	*gdp85*	*popgrowth*	*invest*	*school*	*literacy60*
3.0531972	-0.0005888	0.0002704	0.9969305	0.0094007	0.0065454	-0.0008629
OECD = No						
(Intercept)	*gdp60*	*gdp85*	*popgrowth*	*invest*	*school*	*literacy60*
-0.4542822	-0.0002008	0.0003435	1.1192241	0.0568531	0.0922137	-0.0095543

Table 5 - Outliers test result

	StudRes	Unadjusted p-value	Bonferroni p	Hat	CookD
47	8.271223	9.7974e-13	9.7974e-11	0.9631609	1.481406e+02

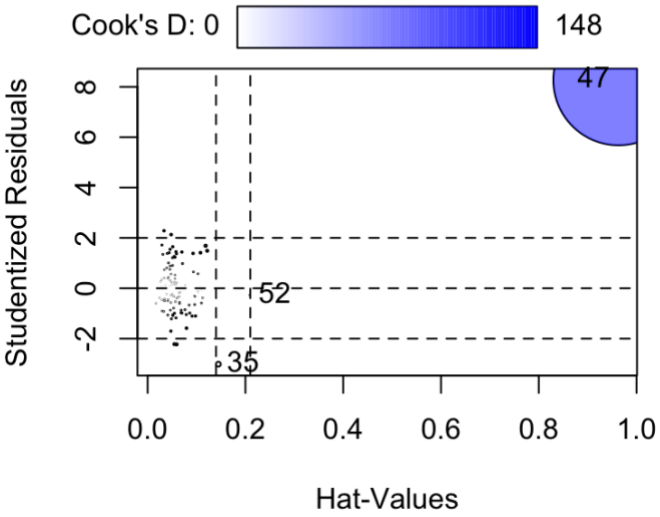

Figure 15 - Influence Plot

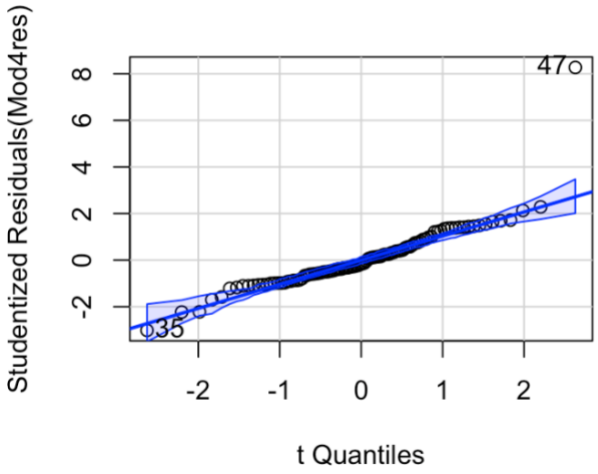

Figure 16 - QQPlot

Table 6 - Confidence intervals of the coefficients

	2.5 %	97.5 %
(Intercept)	-1.2369096255	0.4754952898
gdp60	-0.0009244065	-0.0006277177
gdp85	0.0003755698	0.0005721163
popgrowth	0.9799697810	1.4747702468
invest	0.0099227725	0.0809918761
inter	0.3922306366	1.3784740786

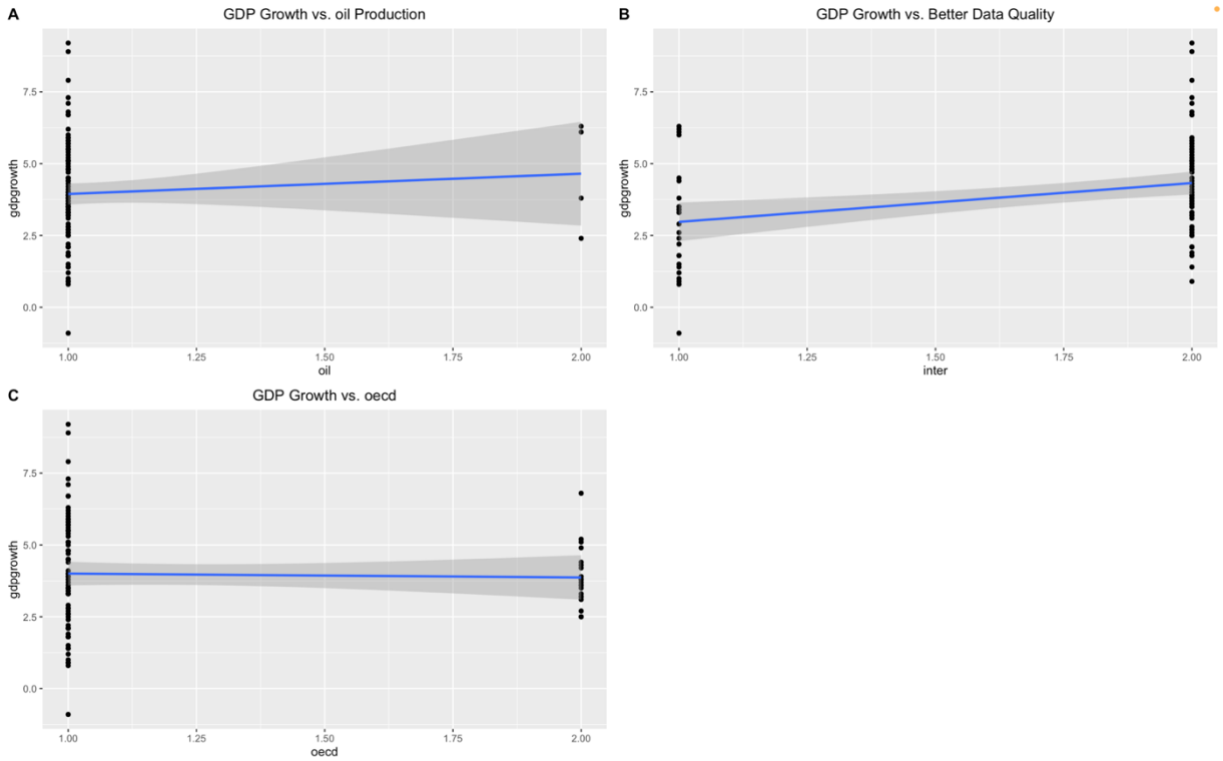

Figure 17 - Impact of binary variables in gdpgrowth

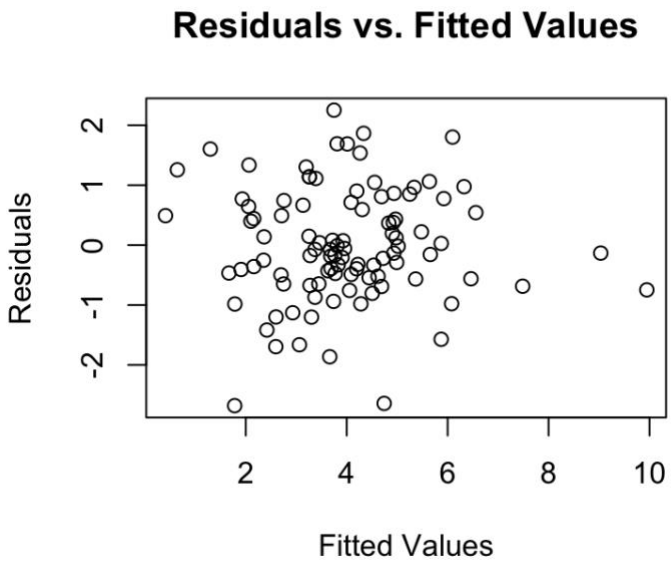

Figure 18 - Residuals vs. Fitted Values

```
> summary(df)
  oil      inter      oecd        gdp60           gdp85         gdpgrowth       popgrowth        invest          school
 no :96   no :26   no :78   Min.   :  383   Min.   :  412   Min.   :-0.900   Min.   :0.300   Min.   : 4.10   Min.   : 0.400
 yes: 4   yes:74   yes:22   1st Qu.: 1001   1st Qu.: 1182   1st Qu.: 2.700   1st Qu.:1.700   1st Qu.:11.62   1st Qu.: 2.400
                            Median : 1945   Median : 3484   Median : 3.800   Median :2.400   Median :16.55   Median : 4.850
                            Mean   : 3837   Mean   : 5634   Mean   : 3.973   Mean   :2.274   Mean   :17.43   Mean   : 5.453
                            3rd Qu.: 4776   3rd Qu.: 7719   3rd Qu.: 5.125   3rd Qu.:2.900   3rd Qu.:23.32   3rd Qu.: 8.075
                            Max.   :77881   Max.   :25635   Max.   : 9.200   Max.   :6.800   Max.   :36.90   Max.   :11.900
   literacy60
 Min.   :  1.00
 1st Qu.: 15.00
 Median : 46.00
 Mean   : 49.43
 3rd Qu.: 84.00
 Max.   :100.00
```

Figure 19 – Summary

```
> apply(df[,numericas],2,sd)
      gdp60       gdp85   gdpgrowth   popgrowth      invest      school  literacy60
8005.468973 5625.719893    1.815470    1.016592    7.821341    3.460147   35.102233
```

Figure 20 - Standard Deviation

```
> summary(Mod6res)

Call:
lm(formula = gdpgrowth ~ gdp60 + gdp85 + popgrowth + invest +
    interyes, data = df_new)

Residuals:
     Min       1Q   Median       3Q      Max
-2.67920 -0.55299 -0.07245  0.68938  2.25172

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.807e-01  4.312e-01  -0.883 0.379524
gdp60       -7.761e-04  7.470e-05 -10.389  < 2e-16 ***
gdp85        4.738e-04  4.949e-05   9.575 1.62e-15 ***
popgrowth    1.227e+00  1.246e-01   9.852 4.21e-16 ***
invest       4.546e-02  1.789e-02   2.540 0.012732 *
interyes     8.854e-01  2.483e-01   3.565 0.000576 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.971 on 93 degrees of freedom
Multiple R-squared:  0.7292,    Adjusted R-squared:  0.7146
F-statistic: 50.08 on 5 and 93 DF,  p-value: < 2.2e-16
```

Figure 21 - Summary(Mod6res)