Assignments 3 and 4

This assignment counts for 2 assignments. The marking scheme is 0,2,4 (as opposed to 0,1,2). If you answer all questions but get less than 50% of the answers right, you get 3. The assignment is due on June 2nd in Learn dropbox.

In this assignment, we want to test the prediction of the Solow growth model. This model predicts that poor countries (or countries with low level of capital) grow faster than rich countries. As a result, we should see a convergence of the standard of leaving between countries. In other words, every countries should eventually have the same level of GDP (gross domestic product).

We can test the hypothesis by running the following regression:

$$G = \beta_0 + \beta_1 GDP_0 + u$$

where G if the average GDP growth between year t_0 and year t and GDP_0 is the GDP at t_0. According to the Solow model, if the initial GDP is low, the growth should be high. The coefficient β_1 should therefore be negative.

To answer the questions, use the datset GrowthDJ from the library AER. The data is taken from a research paper published in 1995 in the Journal of Applied Econometrics by S.N. Durlauf and P.A. Johnson. The title of the paper is "Multiple Regimes and Cross-Country Growth Behavior". The variables are

- oil: Is the country an oil-producing country (1 if yes 0 otherwise)?
- inter: Does the country have a better quality data (1 if yes 0 otherwise)?
- oecd: Is the country a member of the OECD (1 if yes 0 otherwise)?
- gdp60: Per capita GDP in 1960
- gdp85: Per capita GDP in 1985
- gdpgrowth: Average growth rate of per capita GDP from 1960 to 1985 (in percent)
- popgrowth: Average growth rate of working-age population 1960 to 1985 (in percent).
- invest: Average ratio of investment (including Government Investment) to GDP from 1960 to 1985 (in percent)
- school: Average fraction of working-age population enrolled in secondary school from 1960 to 1985 (in percent).
- literacy60: Fraction of the population over 15 years old that is able to read and write in 1960 (in percent).

The dataet contains some missing values, so you may have to use what you learned in the tutorial about missing values.

1. Estimate the model
$$gdpgrowth = \beta_0 + \beta_1 gdp60 + u$$
by OLS, and report the results. Is β_1 negative and significant? Do the test manually and verify that you get the right answer by looking at the summay() output (the p-value will be different because you have to do a one sided test). Print the result using stargazer, and interpret the value of the coefficients.

2. Plot gdpgrowth as a function of gdp60, using a scatter plot. Add the regression line from the previous question. Discuss what you see in relation with the prediction of the Solow growth model, in a paragraph of 3 to 5 lines.

3. You have probably noticed on the graph of previous question, that one country seems to be coming from out of space. Locate that point, remove it, and answer 1) and 2) without it. Hint: You want to drop the observation number with the highest gdp60. Suppose it is the 30th observation. Then you would run the regression using data without its 30th observation this way (I use x and y just for this example):

   ```
   fit <- lm(y~x, data[-30,])
   ```

   If you know that the 30th is the only one above 10,000, you can use subset:

   ```
   fit <- lm(y~x, data, subset=x<10000)
   ```

   Don't forget to also remove the observation for the scatter plot.

For the following questions, assume that the observation you dropped is an outlier that you do not want in your dataset, so redefine "data" without that observation. For the Hint above, that would mean run the following:

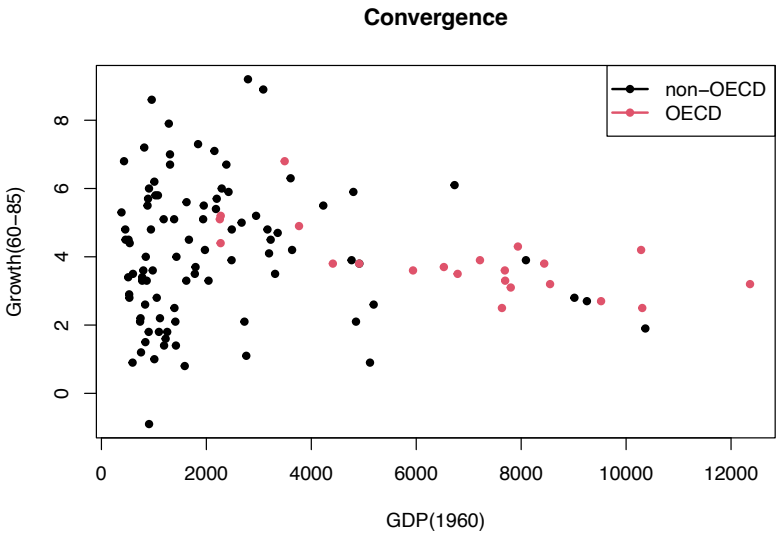
```
data <- data[-30,]
```

An extension to the benchmark Solow growth model, which states that convergence occurs independently of the structure of each country, is to assume instead that convergence occurs within groups of countries with similar structures. In the following questions, we will try to estimate the model using different grouping strategies, and test if similar countries, in terms of specific characteristics, seem to converge to same standard of living.

4. First, we will divide the sample into countries that are members of the OECD (Organisation for Economic Co-operation and Development), and countries that are not. Answer questions 1) and 2) for each group and return the result in one stargazer table. Do you see any change in the R^2? Hint: the variable oecd is equal to "yes" when the country is a member, and "no" if it is not.

   Here is a trick for putting both lines on the same graph. It is only for those of you who are interested to learn. First, you can give a different color to each point, by setting col to a vector of ones and twos. In that case, observations with ones will be black, and observation with twos will be red. You can then run abline() with the two fitted models to have the two lines on the same graph.

```r
## col=2 for oecd and 1 otherwise
col <- (data$oecd=="yes")+1
## the pch=20 creates bullets instead of empty points
plot(gdpgrowth~gdp60, data, col=col, main="Convergence",
     xlab="GDP(1960)", ylab="Growth(60-85)", pch=20, lwd=2)
legend("topright", c("non-OECD","OECD"), col=1:2, lty=1,
       lwd=2, pch=20)
```



If fit1 and fit2 are the fitted model for OECD and non-OECD countries respectively, you can add the two lines this way:

```r
abline(fit1, col=2, lwd=2)
abline(fit2, col=1, lwd=2)
```

5. Test if we have convergence among oil producers.
6. A lack of education can prevent countries from growing. Restrict your sample to countries with 1960 literacy being greater than 70%. Then, answer questions 1) and 2).
7. The model predicts that countries with higher investment rate should, everything else being equal, grow at a faster rate. Divide your sample into low investment countries (invest<15) and high investment countries (invest>23) and answer questions 1 and 2 for both groups.
8. Based on all regression results, write a conclusion in 5 to 10 lines.
9. For the last question, I want you to have a small reflection on causality. First, estimate the following models, print the results on one stagazer table, and interpret the value of the coefficients. As for the previous questions, produce a scatter plot and add the regression line for each model.

$$gdp60 = \beta_0 + \beta_1 literacy60 + u$$

$$\log(gdp60) = \beta_0 + \beta_1 literacy60 + u$$

I want you to make up a story (a plausible one) that could explain the forward causality (from literacy60 to dgp60) that is compatible with what you see. Then, find an equally plausible story that could explain reverse causality (from gdp60 to literacy60). Finally, find a plausible story that could explain why there may be no relationship between literacy60 and gdp60, despite what we see on the graphs.