Assignment 5

For this assignment, use the file "Growth" that you used for the Assignments 3 and 4. 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).

Answer the following questions with outlier from last assignment removed.

a) One of the weakness of the previous assignment is the lack of flexibility in terms of model selection. Now that we can add more than one regressor, we first want to see if the relationship between gdp60 and gdpgrowth is nonlinear. To do so, estimate the following model:

$$gdpgrowth = \beta_0 + \beta_1(gdp60) + \beta_2(gdp60^2) + u$$

Answer the following questions:

- Test the null hypothesis that the relationship is linear against the alternative that it is nonlinear. Do it manually, and verify with summary().

- Predict the growth rate for countries with gdp60 equals to 2000, 3000, 4000, 5000, 6000, 7000, and 8000.

- Produce a scatterplot of gdp60 as a function of gdpgrowth and add the regression line. According to the regression line, what can you say about convergence (of GDP's)?

- According to your result, at which level of gdp60 the average growth is maximized? What is the maximum growth on the regression line?

- Explain why it is harder to test the Solow prediction using this model.

b) We saw in the previous assignment that the coefficient of gdp60 is different if we group countries according to some criteria. What we wanted to do is to test if countries with similar characteristics were converging to each other. We can accomplish that by adding regressors. Estimate the following models and answer the questions that follow.

$$gdpgrowth = \beta_0 + \beta_1(gdp60) + u$$

$$gdpgrowth = \beta_0 + \beta_1(gdp60) + \beta_2(literacy60) + u$$

$$gdpgrowth = \beta_0 + \beta_1(gdp60) + \beta_2(literacy60) + \beta_3(invest) + u$$

$$gdpgrowth = \beta_0 + \beta_1(gdp60) + \beta_2(literacy60) + \beta_3(invest) + \beta_4(oecd) + u$$

Note that you can add oecd in the regression even if it is equal to "yes" or "no". R will convert it into a dummy variable. We will learn about dummy variables in Chapter 7.

- Interpret the values of the coefficients in each model.

- For each model, test the hypothesis that the coefficient of gdp60 is zero against the alternative that it is negative. Compare the results and explain in what way they are related to what you found in the previous assignment.

- Compare the R^2 and the adjusted R^2 between models and explain what the differences imply.

- Using only the third model, produce a scatterplot of gdp60 as a function of gdpgrowth and add the following line:

$$\hat{Y} = (\hat{\beta}_0 + \hat{\beta}_2\overline{literacy60} + \hat{\beta}_3\overline{invest}) + \hat{\beta}_2 gdp60$$

- The number of observations in the first model is higher if you compare with the three others. Why? Do you think it is a problem? Explain.

c) Consider the following model:

$$gdpgrowth = \beta_0 + \beta_1(gdp60) + \beta_2(gdp60^2) + \beta_3(literacy60) + \beta_4(invest) + u$$

Answer all questions in a) using that model. For the graph and the predictions, evaluate invest and literacy60 at their respective mean. Compare your answers with the ones you obtained in a).