Work on Linear Models

Global Development Indicators

**Introduction**

The World Development Indicators (WDIs) are a vital source of information for monitoring a country's progress, identifying trends, and formulating policies. The WDIs include a wide range of socio-economic indicators spanning a variety of sectors such as education, health, environment, economy and governance, providing a comprehensive picture of global development dynamics.

In this study we will use the World Bank's WDI database[[1]](#footnote-1) . Specifically, a subset of the data (data\_tidy.csv) is given, consisting of 27 variables for 120 countries, in the year 2018.

**Work Structure**

The task must be implemented in the R programming language. You are given an RStudio Project named WorDevInd. The project has the following structure:

* data: data folder
  + data\_tidy.csv: The main data file.
  + data\_new.csv: A file with 8 more countries that you will need in Section 2, Question 8.
  + metadata.pdf: A file with information about the data you will need in Module 1.
* doc: work file
  + assignment.docx: This file with information about the assignment.
  + report.docx: A template/framework for presenting results used in academic papers.
* src: Code folder
  + analysis.R: A code file (blank) in which you are asked to implement the analysis according to the questions of the task.

The deliverables of the exercise are the WorDevInd project you received, with changes to two files:

* The analysis.R with the code you wrote.
* A report presenting the results in pdf format. You can use word, latex, or any other text editor you prefer. For guidance, the report.docx file is provided with the typical skeleton of an academic paper. Its use is optional.

You can upload the RStudio project as a compressed file in the "Assignments" tab of e-class.

**Module 1η : Descriptive Statistics**

In this section you will describe and analyse the data from the project.

1. A file with many variables can easily become chaotic. Therefore, it is very important that any data file is accompanied by "metadata", a file that describes each of the variables in detail (what it indicates, how it was measured, etc.). Open the accompanying "metadata.pdf" file, study it, and transfer this description to the main text of your paper so that readers know what each variable is about.
2. Create two new categorical variables:
   1. The "Fertility Category" (give it the name FertCat), which will take the values 1, 2, and 3 if the Adolescent Fertility Rate (AdolFertRate) of the country in question is in the ranges (0, 50], (50, 100], or (100, , respectively.
   2. The "Inflation Category" (give the name InflCat), which will take the value 0 if the country has Inflation less than or equal to 2(%), and 1 otherwise.
3. Study and comment on the descriptive statistical measures for the variables Corruption Control (CorControl), Political Stability (PolStab), and Voice and Accountability (VoiceAcc), as well as FertCat, InflCat. For the qualitative variables, make the frequency table and a barplot. For the quantitative variables, report some summary statistics (sample mean, median, standard deviation, percentiles, etc.), a histogram and a boxplot.
4. A colleague informs you that there is a package called psych, which allows us to use the describeBy function. Look for information on how this function works (on the help page, on the web, and so on). Use it to calculate basic descriptors (sample mean and standard deviation) for the variables Corruption Control (CorControl), Political Stability (PolStab), and Voice and Accountability (VoiceAcc) by Fertility Category (FertCat).

**Section 2η : Linear Model**

The three indicators Corruption Control (CorControl), Political Stability (PolStab), and Voice and Accountability (VoiceAcc) are difficult to measure. It would therefore be useful to develop a multiple regression model capable of estimating them through the other indicators.

Feel free to choose one of the three to play the role of the dependent variable (y) in your paper. The other two variables are prohibited from being used in the models you build as independent variables (x).

1. Create the full multiple linear model, i.e. including all indicators (except the two you excluded) as independent variables.
2. Starting from the complete model you have built, follow a Backward Stepwise procedure with AIC as a criterion to arrive at the "optimal" model.
3. Calculate the and for the resulting model and comment on them.
4. For the resulting model, interpret the coefficients of the independent variables.
5. Select two of the independent variables in the resulting model and perform the statistical significance test at the 5% level:
6. Calculate the 99% confidence intervals for the coefficients of the resulting model.
7. Test the assumptions of the model by checking the residuals. Are they satisfied? If not, try a transformation on the variables.
8. The data\_new.csv file includes 8 more countries for which the estimates of Corruption Control (CorControl), Political Stability (PolStab), and Voice and Accountability (VoiceAcc) are not available. Based on your model,
   1. Give a point forecast for the dependent variable for each of the 8 countries.
   2. Construct mean and individual prediction intervals for the dependent variable of each of the 8 countries.

**Section 3η : ANOVA**

1. Perform a variance analysis with the independent variables "Fertility Category" (FertCat) and "Inflation Category" (Inflation Category -InflCat) created in the 1η Section, and the variable selected above as the dependent variable.
2. State the assumptions of the ANOVA model and fit such a model that studies the effect of the dependent variables on the response variable. Comment on the model.
3. Test the interaction and main effects of the factors on the response variable at the 5% level of statistical significance.
4. Test the assumptions of the model by checking the residuals.
5. Investigate whether there is an effect of each level of factors on the response variable.

1. https://databank.worldbank.org/reports.aspx?source=world-development-indicators [↑](#footnote-ref-1)