Business Statistics I (STAT 220)
Excel Group Assignment – Spring 2021

Dr. Mohammed Ait Lahcen
Department of Finance and Economics

Through this assignment, you will explore the state of the covid-19 pandemic and the pace of the vaccination campaign around the world. The data we will use was collected by Our World in Data from various sources (see https://ourworldindata.org/coronavirus-data-explorer). We chose a selection of indicators in the accompanying Excel file that you would analyze in order to answer some interesting questions. Definitions of the variables used are available here.

Instructions:

- Select your own group.
- **The maximum number of students in a group is 5** (you can work alone if you want).
- Each group should submit **one electronic report** (word or pdf document) on Blackboard on **Thursday, April 15 (before Thursday night 11:59pm)**. This means that one person of the group should submit the document on Blackboard. Email submissions will not be accepted and **late submissions will lose 3 marks**.
- Please make sure you write the names of the group members at the top of the report.
- The submitted report should contain the answers with their corresponding question number and include all the answer elements such as graphics (copied from Excel), numbers and explanations. Please round your reported numbers to 2 decimals.
- You are not asked to submit the Excel file used for calculations.

Part I: Covid-19 pandemic impact (9 marks)

The objective of this part is explore the negative health impact of the covid-19 pandemic and understand whether it differs between rich and poor countries.

1. What is the share of the world population confirmed to have been infected by covid-19 (i.e. total cases in percentage of the world population) and the share confirmed to have died from covid-19 (i.e. total deaths in percentage of the world population)? **(2 mark)**
   Hint: to calculate each of these measures, compute the sum of their respective columns and then divide it by the sum of the population column.
2. Draw a scatter diagram of the total deaths and total cases per country (use a log scale to make it more readable). Explain what it tells us about the relationship between these two variables. **(1 mark)**

Hint: to use a log scale in Excel, once you create the scatter plot graphic, select the X-axis, right click on it and select Format Axis. Scroll down and check the box Logarithmic scale. Do the same with the Y-axis.

3. Calculate the average total cases per million across all countries (average of the column total cases per million). If we interpret this statistic as a probability, what would be the probability of being infected by covid-19 (in percentage)? **(1 mark)**
4. Calculate the average total deaths per million across all countries (total deaths per million). If we interpret this statistic as a probability, what is the probability of dying from covid-19 (in percentage)? **(1 mark)**
5. Plot the histogram of the distribution of total deaths per million across countries. Would you say that the distribution is bell-shaped, positively skewed or negatively skewed? Interpret. **(1 mark)**
6. The case-fatality rate is usually defined as the total number of deaths divided by the total number of cases. Under some simplifying assumptions, we can interpret the case-fatality rate as the probability of dying conditional on being infected by covid-19. Create a new column that computes for each country the case-fatality rate. What is the average value of that column? Interpret. **(1 mark)**

One would expect richer countries to deal better with the covid-19 pandemic. In order to explore this idea, we will look at the relationship between GDP per capita, a measure of the wealth of a country, and some covid-19 related variables.

7. Calculate the correlation between total cases per million and GDP per capita. Is the correlation positive or negative? Interpret. **(0.5 mark)**
8. Because the majority of covid-19 cases are asymptomatic, it could be that richer countries, with more resources, did more tests and hence detected more cases. Calculate the correlation between total tests per thousand and GDP per capita. What do you conclude? **(0.5 mark)**
9. Calculate the correlation between total deaths per million and GDP per capita. Is the correlation positive or negative? Interpret. **(0.5 mark)**
10. Next, calculate the correlation between the case fatality rate and GDP per capita. Is the correlation positive or negative? What do you conclude given how we interpreted the case fatality rate in question 7 above? **(0.5 mark)**

Part II: Covid-19 vaccination campaign (6 marks)

In this section, we will have a look at the vaccination campaign that is taking place around the world.

11. What is the share of the world population that is vaccinated (i.e. received at least one vaccine dose) and the share of the world population that is fully vaccinated (i.e. received all required doses)? **(2 mark)**
    Hint: to calculate each of these two measures, compute the sum of their respective columns and then divide it by the sum of the population column.
12. Given that vaccinations started in many countries in late January 2021 (in some before), give a rough estimate of the time it will take to give at least one vaccine dose to 70% of the world population using the number you computed above. Do you think that the vaccination campaign is going fast enough? Discuss. **(1 mark)**
    Hint: assume it took 2 months to vaccinate the share of the world population that you computed above.
13. What is the proportion of countries that have already vaccinated 10% or more of their population? **(1 mark)**
    Hint: use the COUNT and COUNTIF functions in Excel.
14. Calculate the coefficient of variation of total vaccinations per hundred. Calculate the coefficient of variation of total deaths per million. Compare the two. Would you say that access to the vaccine is more, less or as unequal between countries as the death toll from covid-19? (i.e. which one is more dispersed?) **(2 marks)**