
STATISTICS II

Third activity

Year 2021-2022

DIRECTIONS

1. This is an individual activity.
2. The deadline for submitting your results is January 7th, 2022 BEFORE 23:59.
3. The data file that you have to use is named with your NIA, and can be downloaded from the link below.

Data File

4. You have to use RStudio in this activity.
5. You must use the form that you will find in the link below to enter the results that you obtain.

Answers Form

6. In that form you have to enter:
 - a) Your personal data: Name, last names, NIA and class group (the group you are registered in).
 - b) The results that you have obtained when solving the activity in RStudio. Please always **use all the decimal digits** that RStudio computes.
 - c) If one of the values looks something like "13674." (with no decimal digits) DO NOT enter the final dot, enter the number only
 - d) **Don't** "copy and paste" from RStudio into the answers form. RStudio uses some non standard characters that compromises its reading.
 - e) If the outcome of any of your replies is a value of the type "4.73E-13", you must enter this value in the answers form **in this exact format**.
 - f) Please, **read** all the questions, comments, and directions **with care**. Follow all those instructions so that the answers that you sent can be properly processed.

You will have only one opportunity to enter your answers. Please make sure that the results that you are sending are the correct ones.

SUBMISSIONS THAT DO NOT MEET THESE GUIDELINES WILL NOT BE CONSIDERED

The database provided combines data from the World Bank, the United Nations (UN), and the Division of Environmental Sciences of the Oak Ridge National Laboratory (Tennessee, USA), on income, population, and environmental variables. Corresponding to the year 2018, we have selected a set of variables from a total of 60 countries that you will find in the file <your_your_NIA>.csv

The variable considered are:

- **COUNTRYNAME**: Name of the country
- **REGION**: Region of the world where the country is located
- **INCOMEGROUP**: Country Income Level (Grouped Variable)
 1. High Income
 2. Upper Middle Income
 3. Lower Middle Income
 4. Low Income
- **REN_ENG_CON**: Renewable energy consumption (% of total energy consumption)
- **CO2_EM**: CO2 emission (in kilotons, kt)
- **GHE_GAS_EM**: Greenhouse Effect Gas (GHE) (in kt CO2 equivalent)
- **POP_TOT**: Total population of the country
- **POP_URB**: Urban population of the country

YOU ARE ASKED TO:

1. Test whether there is a significant relationship between the **REGION** and **INCOMEGROUP** variables. (significance level of 5%)
 - (a) What is the p_value in this test? [0.5 POINTS]
 - (b) Based on the outcome of the test, would you conclude that there is a relationship between these two variables? [0.75 POINTS]
 - (c) What is your interpretation of this result? [0.5 POINTS]
2. Analyze and interpret the correlation between the variables **REN_ENG_CON**, **CO2_EM**, **GHE_GAS_EM**, **POP_TOT** and **POP_URB**.
Based on the results obtained:
 - (a) Is there a clear linear relationship (above 0.7 in absolute value) between CO2 emissions and GHE emissions? [0.5 POINTS]
 - (b) Is the GHE gas emissions more related to the total population or to the urban population? [0.5 POINTS]
 - (c) Do you observe an increasing relationship between renewable energy consumption and the total population? [0.5 POINTS]
 - (d) Based on the correlations obtained above, what variable is most correlated with GHE emissions? [0.5 POINTS]
3. Conduct an OLS regression of the variable **GHE_GAS_EM** as a function of the variable **POP_TOT**. Based on the results obtained,
 - (a) What is the goodness-of-fit of the model? [0.5 POINTS]

- (b) Is the coefficient that multiplies the variable POP_TOT significantly different from zero for a significance level of 1%? [0.5 POINTS]
 - (c) If one expects a growth of 100,000 people in the total population, by how much will GHE emissions increase? [0.75 POINTS]
4. Conduct an OLS regression of the variable GHE_GAS_EM as a function of the variable POP_URB. Based on the results obtained.
- (a) What is the goodness-of-fit of the model? [0.5 POINTS]
 - (b) Is the coefficient that multiplies the variable POP_URB significantly different from zero for a significance level of 1%? [0.5 POINTS]
 - (c) If a change in the country's demographic policy produces a decrease in the urban population of 75,000, what effect would it have on GHE emissions? [0.75 POINTS]
5. Conduct an OLS regression of the variable CO2_EM as a function of the variable REN_ENG_CON. Based on the results obtained,
- (a) What is the goodness-of-fit of the model? [0.5 POINTS]
 - (b) Is the coefficient that multiplies the variable REN_ENG_CON significantly different from zero for a significance level of 1%? [0.5 POINTS]
 - (c) If a change in the country's energy policy achieves a 0.5% increase in renewable energy consumption, what effect would it have on CO2 emissions? [0.75 POINTS]
6. Based on the models estimated in questions 3 and 4 above, which situation would have the greatest impact on GHE emissions: an increase in the total population of 1,000,000 people or an increase in the urban population of 500,000 people? [1 POINT]