MATH 118: Statistics and Probability **(Due: 26/04/21)**

Homework #1

Instructor: Dr. Zafeirakis Zafeirakopoulos *Name:* *Student Id:*
Assistant: Gizem Süngü

Course Policy: Read all the instructions below carefully before you start working on the assignment, and before you make a submission.

- It is not a group homework. Do not share your answers to anyone in any circumstance. Any cheating means at least -100 for both sides.
- Do not take any information from the Internet.
- No late homework will be accepted.
- For any questions about the homework, come to my office hour.
- After the office hour, no questions about the homework by email will be responded.
- Submit your homework (both your latex and pdf files in a zip file) into the course page of Moodle.
- Save your latex, pdf and zip files as "Name_Surname_StudentId".{tex, pdf, zip}.
- The deadline of the homework is 22/04/21 23:55.

Problem 1 (100 points)

Homework 1 considers a Covid-19 dataset which is published on Github. Please download any document type that you prefer of the dataset from the links which are shown in Figure 1. The dataset is updated daily and

Figure 1: The complete dataset links

includes data on confirmed cases, deaths, hospitalizations, testing, and vaccinations as well as other variables of potential interest. The data set has the following basic columns:

- iso_code: Short name of the country
- continent: The continent where the country exists
- location: The country name
- date: The date when the data about various variables are taken.

You are responsible to implement a program which reads the given dataset from the file and computes the data for the following questions. Any programming language that you prefer will be accepted. Putting comments on your functions that you implement is must. Each question must be appended to a file which is called "output{.csv, .txt}". The file contains the first 18 questions listed below. The 18th question will be written in this document.

1. How many countries the dataset has?
2. When is the earliest date data are taken for a country? Which country is it?
3. How many cases are confirmed for each country so far? Print pairwise results of country and total cases.
4. How many deaths are confirmed for each country so far? Print pairwise results of country and total deaths.

Table 1: The format of the output for the questions 5, 6, 7, 8, 9, 10, 12, 13.

Country	minimum	maximum	average	variation
value	value	value	value	value

5. What are the average, minimum, maximum and variation values of the reproduction rates for each country?

6. What are the average, minimum, maximum and variation values of the icu_patients (intensive care unit patients) for each country?

7. What are the average, minimum, maximum and variation values of the hosp_patients (hospital patients) for each country?

8. What are the average, minimum, maximum and variation values of the weekly icu (intensive care unit) admissions for each country?

9. What are the average, minimum, maximum and variation values of the weekly hospital admissions for each country?

10. What are the average, minimum, maximum and variation values of new tests per day for each country?

11. How many tests are conducted in total for each country so far?

12. What are the average, minimum, maximum and variation values of the positive rates of the tests for each country?

13. What are the average, minimum, maximum and variation values of the tests per case for each country?

14. How many people are vaccinated by at least one dose in each country?

15. How many people are vaccinated fully in each country?

16. How many vaccinations are administered in each country so far?

17. List information about population, median age, # of people aged 65 older, # of people aged 70 older, economic performance, death rates due to heart disease, diabetes prevalence, # of female smokers, # of male smokers, handwashing facilities, hospital beds per thousand people, life expectancy and human development index.

Table 2: The format of the output for the question 17

Country	population	median age	# of people aged 65 older
value	value	value	value

18. Summarize all the results that you obtain by the first 17 questions (except question 2).

Table 3: The format of the output for the question 18

Country	q#3	q#4	q#5_min	q#5_max	q#5_avg	q#5_var
value	value	value	value	value	value	value

19. Comment the results based on your observations. Write your opinions about the reasons of increasing infection rates by giving examples from the results. Feel free to explain any situation that you observe. More observations more opportunities will bring you for the second homework.
(Solution) (Write your observations here.)

I commented according to the data on 19/04/2021.

- Hypothesis 1 : The number of hospital beds per thousand people increases, deaths decrease.
  For Germany and Italy, which have the same conditions in most of the variables, the number of patient beds per thousand people in Germany is 8, while in Italy it is 3.18. Likewise, while the total number of deaths calculated in Germany is 79894, this number is 116366 in Italy. The number of deaths is less because there are more hospital beds in Germany.

- Hypothesis 2 : The median age of countries increases, the number of deaths increases.
  For Albania and Bahrain, which have similar data, the average age is 38 in Albania and 32 in Bahrein. The total number of deaths calculated for Albania is 2337, while for Bahrain it is 581. Because Albania has a lower average age, the death rate is less.

- Hypothesis 3 : According to the economic performance of the countries, the values of new tests per day it makes increases.
  For Denmark and Slovenia, which have similar data, the economic performance of Denmark is 46682.52 while Slovenia's is 31400,84. The values of new tests per day for Denmark's is 57322.9 and Slovenia's is 2669,2.Denmark, with its high economic performance, has more tests than Chile.

- Hypothesis 4 : The number of male and female smokers in the countries increases,values of the weekly hospital admissions increase.
  For Israel and Portugal, which have similar data, there are 15.4 female smokers and 35.4 male smokers in Israel, while this value is 16.3 female smokers and 30 male smokers in Portugal.The values of the weekly hospital admissions amount in Israel is 741.1356 while it is 278.779 in Portugal. As Smokers increased, the values of the weekly hospital admissions increased.

- Hypothesis 5 : Diabetes prevalence increases in countries, the amount of weekly icu patients also increases.
  For Greece and Lithuania with similar data, the rate of diabetes prevalence is 4,55 in Greece and 3,67 in Lithuania. While the number of weekly icu patients is 94,85315 in Greece, this value is 67,01041 for Lithuania. Since diabetes prevalence is higher in Greece, the number of patients is higher than in Lithuania.

- Hypothesis 6 : The amount of vaccination made by countries increases, the values of new tests per day decreases.
  For Morocco and Slovakia, which are similar to each other, the total amount of vaccination made in Morocco is 8763819 while it is 1269305 in Slovakia. While the average number of new tests per day is 13787.66 in Morocco, it is 73868.12 in Slovakia. Since Morocco vaccinates more, the number of daily tests is less than in Slovakia.