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)
1-) The number of cases in countries is not directly proportional to the number of people. For example, although Africa's population is more than San Marino, the number of cases in San Marino is higher than in Africa.

In Africa : population : 1.264 billion and total cases per million : 3288.8
In San Marino : population : 33938 and total cases per million : 147799

2-)Smoking rates are not related to the number of cases. Ex:
In Timor : male smokers : 78.1 and total cases per million : 904
In Panama : male smokers : 9.9 and total cases per million : 84k

3-)Hand washing rates of countries do not reduce the possibility of receiving a lawsuit. For example:
In Kazakhstan : hand washing rate : 98.99 and total cases per million : 17.887. However
In Liberia : hand washing rate : 1.118 and total cases per million 403.

4-)The average population is high where living conditions are good. And in these places, the corona virus spreads faster. For example: average population, living conditions and cases per million in Italy, Portugal and Spain respectively;
Italy: 47.9, 83.51, 63545,
Portugal: 46.2, 82.05, 81390,
Spain: 45.5, 83.56, 72875.

5-)Corona virus is more common in places where human development is high. Ex:
In Switzerland: human development index : 0.955 and total cases per million : 73070.7
In South Sudan: human development index : 0.433 and total cases per million : 931.95

6-)In countries where the corona virus is high, the number of beds in hospitals is high. Ex:
In Tanzania : hospital beds per thousand : 0.7 and total cases per million : 8.521. However,
In Monaco : hospital beds per thousand : 13.8 and total cases per million : 60926.5.

7-)Corona virus is less common in poor countries. For example, although Madagascar's poverty is more than United States, the number of cases in Us is higher than in Madagascar.
In Madagascar : extreme poverty : 77.6 and total cases per million : 1120.9. However,
In United States : extreme poverty : 1.2 and total cases per million : 95393.9.

8-)Vaccinating people does not affect the corona increase. For example, although there are approximately 1 billion 350 million vaccines in America, most cases are in America.
In United States : total cases per million : 95393.9