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

Homework #1

Instructor: Dr. Zafeirakis Zafeirakopoulos *Name: Baran Hasan Bozduman Student Id: 171044036*
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

pic.png

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.

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

4. How many deaths are confirmed for each country so far? Print pairwise results of country and total deaths.
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.)
For the first 18 questions I used python to implement program and I used panda and numpy libraries to classification and calculate the values for each country

To analyze data I used an online csv graph plotter https://www.csvplot.com, you can check the screenshots from zip file

Accoording to time number of new cases increases
ss.newcase-date
Number of newcases are increases according to date so that means even people get vaccinated still number of new cases increases

According to handwash facilities number of new cases decreases
ss.newcase-handwash
Handwashing facilities decrease the number of new cases

According to economic performance number of newcases decreases(?)
ss.newcase-gdp-per-capita
Countries that has low economic performances are has lower number of new cases but still some continents does not provide that such as Europe because of they have old aged peoples above 65

According to population number of cases increases
ss.newcase-population
The crowdeod continents has more new cases, although some continents has higher population density it still has low new cases

According to vaccinated people number of new cases decreases(?)
ss.newcases-people-vaccinated
Some continents has decreased their number with vaccinate but some asian countries still increases Europian countries has decreased their new cases with vaccacine

Totaldeaths increases with new cases
ss.totaldeath-newcases
If new cases increases total deaths are increases but thats also an expected result

Total deaths increases with icu parients
ss.totaldeath-icupatients
generally total deaths are icreases if the icu patients increases but as graph europian countries almost same it could be because of intense care more expensive in some countries such as USA(my own opinion)

According to date total death increases
ss.newdeaths-date
new deaths decreased and increased again it could be because of mutation or people do not care as first times