HELLENIC OPEN UNIVERSITY

MASTER'S DEGREE PROGRAMME IN BUSINESS ADMINISTRATION

Module: ***ADVANCED QUANTITATIVE METHODS FOR MANAGERS (MBA60)***
Academic Year: ***2021-22***

1st Written Assignment (WA1)

The Environmental Protection Agency (EPA) in New York collects daily data related to NY air quality and uses them as a basis for decision making. In this assignment, you are provided with a dataset, which contains daily air quality measurements in New York for the period between May and September of 1973. The data comprise of Ozone (parts per billion), Temperature (Fahrenheit degrees), Wind (average speed) and Solar Radiation measurements on a daily basis between 1st of May and 30th of September.

The dataset in the file **MBA60_2021_22_WA1_DATA.xlsx** contains153 observations on 6 variables. Variable metadata are as follows:

Variable	Unit	Description
Ozone	parts per billion	mean ozone concentration
Solar.R		solar radiation
Wind	miles per hour	average wind speed
Temp	Fahrenheit	maximum daily temperature
Month		month of observation
Day		day of month

Question 1: (5%)
Create the following new variables within the given dataset:

- **month**: From the original variable Month replace month numbers with labels as follows: 5=May, 6=June, 7=July, 8=August, 9=September.
- **TempCelsius**: Transform daily temperatures from Fahrenheit to Celsius using the appropriate conversion formula.
- **HotTemp**: A dichotomous variable with the value 1 if TempCelsius > 25 and 0 otherwise.
- **OzAlertLevel**: An ordinal variable with value **2** if Ozone > 65 & TempCelsius > 25, **1** if Ozone > 50 & Ozone $\leq$ 65 & TempCelsius > 25 and **0** in all other cases.

HELLENIC OPEN UNIVERSITY

MASTER'S DEGREE PROGRAMME IN BUSINESS ADMINISTRATION

Question 2: (25%)

2.1 Calculate the mean and the standard deviation of the temperature (in Celsius) for each month. Compare and comment on your results.

2.2 Calculate the proportion of the hot days for each month (hot days are defined as those with temperature >25 Celsius degrees). Comment on your findings. Are these in line with the ones from question 2.1?

2.3 Calculate the median, the 1st and the 3rd quartile of the variable **Wind** for each month. Interpret your findings to describe the distribution of the Wind data in August.

2.4 Compare the variability of **Ozone**, **TempCelsius**, **Solar.R** and **Wind** for the entire period, using the appropriate measure(s). Which variable shows the largest variability? Comment on your results.

2.5 Construct the frequency distribution and the percentage distribution of the variable **Solar.R** for the entire period, by choosing the appropriate number of classes (justify your choice). Calculate the mean twice using the raw and the grouped data. Compare and comment on the results. Which one will you use and why?

Question 3: (25%)

3.1 Plot the histogram and the boxplot of the **Ozone** variable for the entire period using the appropriate labels and titles. Interpret your data visualization findings describing the main characteristics of the distribution.

3.2 Plot the boxplots of the **Ozone** variable for each month, side by side, in one graph. Add a horizontal line to the graph at the value of the Grand Median (the Ozone median of the entire period). Discuss your findings as derived by the graph.

3.3 Construct scatter plots of the variable pairs (**Temp**, **Wind**), (**Temp**, **Ozone**), (**Ozone**, **Wind**). Detect and describe the possible relationship of the variables in each pair. Comment on your findings.

3.4 Create a clustered bar chart for the variable **OzAlertLevel** per month. Comment on the results.

Question 4: (15%)

When the Environmental Protection Agency (EPA) in New York detects Ozone level above 65, they issue the first warning (yellow flag).

4.1 Calculate the probability of at least one yellow flag within a week (7 days) between May and September.

4.2 How many yellow flags are expected during the same week?

For answering the above questions assume that the probability of a yellow flag in any given day is estimated from the percentage of yellow flags in the given data set. (Percentage of days with Ozone > 65).

HELLENIC OPEN UNIVERSITY

MASTER'S DEGREE PROGRAMME IN BUSINESS ADMINISTRATION

Question 5: (15%)
When the Environmental Protection Agency (EPA) in New York detects Ozone level above 65 and Temp above 90, they issue a second warning (red flag).

5.1 Calculate the probability of at least two red flags within a given month (31 days).
5.2 What is the probability of at least one but no more than three red flags during the same month?

For answering the above questions assume that the probability of a red flag in any given day is estimated from the percentage of red flags in the given data set. (Percentage of days with Ozone > 65 and Temp > 90). Use the Poisson approximation to the Binomial distribution.

Question 6: (15%)
The daily temperature between May and September follows approximately normal distribution. Using this fact and after estimating the **μ** and **σ** from the dataset as the mean and the standard deviation of the variable **TempCelsius,** perform the following calculations:
6.1 Calculate the probability that the temperature in a randomly selected day will be larger than 20 Celsius degrees.
6.2 Calculate the temperature that a day must have (at least) in order to be among the 10% of the warmest days of the season.

Note: *You can use:*

- *the Tables of Cumulative Binomial Distribution or the R function pbinom(x,n,p) or the Excel function: BINOMDIST(X;n;p;TRUE) or any other stat package relative command.*
- *the Tables of Poisson Distribution or the R function ppois(x,λ) or the Excel function : POISSON. DIST(X; mean value;TRUE) or any other stat package relative command.*
- *the tables of Normal Distribution or the R function pnorm(x,μ,σ) και qnorm(p,μ,σ) or Excel function : NORM.DIST(X; mean value; standard deviation;TRUE) or any other stat package relative command.*

Assignment guidelines

- *It is important that the coursework reflects your **knowledge** rather than it being simply an accumulation of information.*
- *The assignment should be **well structured** and **easy to read**.*

HELLENIC OPEN UNIVERSITY

MASTER'S DEGREE PROGRAMME IN BUSINESS ADMINISTRATION

- *The assignment should clearly present all aspects and perspectives of the subject area, i.e.:*
  - *efficiently develop all necessary elements*
  - *refer to actual case studies or statistics if required*
  - *present reasonable argumentation*
  - *omit irrelevant material*
- *All questions are compulsory. The assignment, including possible diagrams, tables, references etc., should not exceed 5000words. For every additional 1000 words there will be a penalty of 0.5 points.*
- *Each question accounts for a percentage of the total mark. This is clearly marked at the beginning of each question.*
- ***The assignment is due on 23/11/2021. Please note that no assignment will be acceptable after this date as the electronic submission system automatically locks at 23:59 on the last day of submission.***
  You should submit your assignment via http://study.eap.gr using your username and password.
- *The assignment should* ***be submitted for grading ONLY in MS Word format*** *.*
  - *You may optionally submit your calculations in any file format you have used (Excel, SPSS, R, Minitab, JMP).*

  Other document formats or read only file formats such as Portable Document Format (.pdf) are not acceptable formats for the submission of your assignment.*
- *Please pay attention to the proper naming of your assignment. The file should be named as follows:* ***Surname-Initial-WA1-YourClass****. For example, if your name is* ***Peter Drucker****, you are sending in your 1st assignment, and you are in ΗΛΕ15 Class, then you should name your file as follows:* ***Drucker-P-WA1-ΗΛΕ15****. Assignments that fail to comply with this requirement will receive a lower mark in the presentation grade.*
- *Copying is considered cheating and is not acceptable in any form.* ***Copying large parts or whole paragraphs of text found in any of the sources used for an assignment (printed books, academic articles, or electronic media of any kind) is totally unacceptable. It is considered plagiarism and leads to a severe penalty for the student(s) involved.*** *Students should cite all sources from which they take data, ideas or words, whether quoted directly or paraphrased.*

Good luck!!