COSC 311 Lab 3 (10 points)

Due Friday, 31 March 2023 at 11:59 PM

Please finish the following tasks and submit your lab report via MyClasses. Your submission must contain your source code file (one ".py" file for all code) and a PDF document. For each task, please include the output/results of your program (may use screenshot).

Dataset: Please use the "Bejaia Region Dataset" and the "Sidi-Bel Abbes Region Dataset" in lab 2 assignment. You may also download them from this lab assignment on MyClasses. There are two classes in these datasets: "not fire" and "fire".

Using "stats.py" file: You may import this file to your program to use any necessary one-dimensional statistics functions. This file is contained in the lecture note of "COSC311_Module4_1_One Dimensional Statistics". You may also download it from this lab assignment on MyClasses.

Task 1: Mean values of the "Bejaia Region Dataset" (2 points)

1) Using the "Bejaia Region Dataset", calculate and show the mean values of four attributes ("Temperature", "RH" (Relative Humidity), "Ws" (Wind speed) and "Rain", respectively) for each class (i.e. for "not fire" and "fire", respectively);
2) Draw a bar figure to show the mean values of these attributes for each class;
3) Describe your observations in the above figure;

Task 2: Using the " Sidi-Bel Abbes Region Dataset", calculate and show the median values of four attributes ("FFMC", "DMC", "DC" and "ISI", respectively) (1 point)

Task 3: Using the "Bejaia Region Dataset", calculate and show the 25-percent, 60-percent, and 75-percent quantiles of four attributes ("Temperature", "RH", "Ws" and "Rain", respectively) (1 point)

Task 4: Using the "Sidi-Bel Abbes Region Dataset", calculate and show the standard deviation values of four attributes ("Temperature", " Rain", "BUI" and "FWI", respectively) (1 point)

Task 5: Correlation between two attributes (2 points)

1) Using the "Bejaia Region Dataset", calculate and show the “correlation coefficient” between “RH” and each of the following attributes ("Temperature", "Ws", "Rain", "FFMC", "DMC", "DC", "ISI", "BUI" and "FWI"), respectively;
2) Describe, if there is one, which attribute has strongest positive correlation with attribute “RH”;
3) Describe, if there is one, which attribute has strongest negative correlation with attribute “RH”.

Task 6: (open question) Assume you need to select some attributes or design some new attributes to distinguish these two classes ("not fire" and "fire") as accurate as possible, which attributes you would like to select or what new attributes you would like to design? You need to use the data analysis results to support your conclusion. (3 points)

Policy

1. Each student MUST finish this lab independently. NO TEAM WORK and DISCUSSION are allowed. If you need any help, please feel free to contact the instructor.

2. You need to write your whole program in an editor and save your source code as a “.py” file, which will be submitted to MyClasses together with your PDF report.