Time series Analysis and Modeling

DATS 6313

Homework # 1

Pearson Correlation coefficient

In this LAB you will develop Pearson correlation coefficient program from the ground and will apply for different dataset. You may use the following packages for this LAB: (display all calculated numbers with 2-digit decimal precision):

```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import seaborn as sns
```

Write a python program for each problem below:

1- Write a python function called "correlation_coefficent_cal(x,y)" that implements the correlation coefficient between random variable x and y. The formula for correlation coefficient is given below. The function should be written in a general form that can work for any dataset x and dataset y. The return value for this function is r. Test the developed code with the following make-up dataset for x, y, z, h & g.
   a. The correlation coefficient between x and y. Display the answer on the console.
   b. The correlation coefficient between x and z. Display the answer on the console.
   c. The correlation coefficient between g and h. Display the answer on the console.

   Verify a, b and c using a python program.

```
x = [1,2,3,4,5]
y = [1,2,3,4,5]
z = [-1,-2,-3,-4,-5]
g = [1,1,0,-1,-1,0,1
h = [0,1,1,1,-1,-1,-1]
```

$$r = \frac{\sum(x_t - \overline{x})(y_t - \overline{y})}{\sqrt{\sum(x_t - \overline{x})^2}\sqrt{\sum(y_t - \overline{y})^2}}$$

2- Without use of python, calculate the following correlation coefficients. You need to show all your manual calculations. Compare the answer for this question and the answer to the previous question and show that they are identical.
   a. The correlation coefficient between x and y.
   b. The correlation coefficient between x and z.
   c. The correlation coefficient between g and h.

3- Load the time series data called tute1.csv from the course GitHub. Graph the scatter plot between Sales & GDP. Calculate the correlation coefficient between Sales and GDP [using the developed code in question 1] and update the graph title between calculated correlation coefficients. Update the x and y axis with an appropriate label. Does the calculated correlation coefficient make sense with respect to the scatter plot? Justify your answer.

4- Graph the scatter plot between Sales and AdBudget. Calculate the correlation coefficient between Sales and AdBudget [using the developed code in question 1] and update the graph title between calculated correlation coefficients. Update the x and y axis with an appropriate label. Does the calculated correlation coefficient make sense with respect to the scatter plot? Justify your answer.

5- Graph the scatter plot between GDP & AdBudget. Calculate the correlation coefficient between GDP & AdBudget [using the developed code in question 1] and update the graph title between calculated correlation coefficients. Update the x and y axis with an appropriate label. Does the calculated correlation coefficient make sense with respect to the scatter plot? Justify your answer.

6- Using the Seaborn package and pairplot() function, graph the correlation matrix for the tute1.csv dataset. Plot the Dataframe using the following options. Explain the graphs and justify the cross correlations.
   a. `kind="kde"`
   b. `kind="hist"`
   c. `diag_kind="hist"`

7- Using the Seaborn package and heatmap() function, graph the correlation matrix for the tute1.csv dataset. Explain the depicted correlation matrix.

8- Develop a python program that asks a user to input numerical numbers: mean, variance & number of observations [default values : mean = 0, variance = 1, observations = 1000]. Then generates a random variable x that is normally distributed with above statistics. Create the following random variables:
   a. $y = x^2$
   b. $z = x^3$

   Using the developed code in question 1 calculate the following correlation coefficients:
   - The correlation coefficient between x & y. Display the answer on the console.
   - The correlation coefficient between x & z. Display the answer on the console.
   - Justify your answer to the above questions using the plot of x versus y and z.
   - Are x and y independent? Are x & y correlated? Justify your answer.
   - Are x and z independent? Are x & z correlated? Justify your answer.

Upload the **solution <u>report (as a single pdf</u>**) plus **<u>the .py file</u>** through BB by the due date.