MEME19803/MECG11503 Practical Assessment 1 (1 Hour)

Name: Student ID: Mark: /10

Course Code & Course Title:	MEME19803/MECG11503 Programming for Data Analytics		
Faculty:	LKC FES, UTAR	Course:	MAC, EEC
Session:	Jan 2022	Lecturer:	Dr Liew How Hui

Instruction: Write all your answers in Microsoft Word or LibreOffice document. Answers without working steps may receive ZERO mark. You can write your answers on a piece of paper, take photo and insert to Microsoft Word or LibreOffice writer. Submit both Word or LibreOffice document together with the PDF files.

CO2: Interpret data using exploratory data analysis .... C5 (10 marks)

1. Numerical data are usually stored as array in Python because Numpy provides a good syntax and platform for numerical data processing and exploratory data analysis. Suppose the following two-dimensional array needs to be stored in Python for investigation

$$A = \begin{bmatrix} 1 & 6 & 19 & 7 \\ 0 & 2 & 10 & 9 \\ 0 & 0 & 3 & 18 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

   (a) Write down (i) a single Python command to construct the above two-dimensional array A and (ii) a single Python command to construct a zero matrix Z of the same size/shape as A. (1 mark)

   (b) Write down the Python commands to find the minimum, maximum, range and mean of **each row** of the two-dimensional array A. Write down the output of the Python commands. (2 marks)

   (c) Write down a single line Python command with a string length less than 50 using the Numpy methods only to perform standard scaling on the columns of the matrix A and show the two-dimensional array after the column standardisation (round all numbers to 4 decimal places). (1 mark)

   (d) Write down a single line Python command with a string length less than 60 using the Numpy methods only to perform standard scaling on the rows of the matrix A and show the two-dimensional array after the row standardisation (round all numbers to 4 decimal places). (1 mark)

   (e) Let B be the matrix product of A four times $A^4 = AAAA$. Write down (i) the **Python command** to calculate A^4 and (ii) write down the command and the output of the **transpose of the first two rows of** B. (1 mark)

   (f) Write down the Python command of a string length less than 50 to take linear combinations of the first column $A_{:,1}$, third column $A_{:,3}$ and second column $A_{:,2}$, i.e. $A_{:,1} + A_{:,3} - A_{:,2}$ to add it to the right of the two-dimensional array A forming the new array

$$C = \begin{bmatrix} 1 & 6 & 19 & 7 & 14 \\ 0 & 2 & 10 & 9 & 8 \\ 0 & 0 & 3 & 18 & 3 \\ 0 & 0 & 0 & 4 & 0 \end{bmatrix}.$$

(0.5 mark)

(g) Given that the definition of correlation coefficient between two arrays $\boldsymbol{x}$ and $\boldsymbol{y}$ is

$$Corr(\boldsymbol{x}, \boldsymbol{y}) = \frac{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]}{\sqrt{\mathbb{E}[(X - \mathbb{E}[X])^2]\mathbb{E}[(Y - \mathbb{E}[Y])^2]}} = \frac{(\boldsymbol{x} - \overline{x}) \cdot (\boldsymbol{y} - \overline{y})}{n\sigma_x\sigma_y}$$

Find correlation coefficient between row 2 of A and row 3 of A by writing down the Python for calculation and the value of the correlation coefficient to 6 decimal places. (0.5 mark)

2. Given the following 20 univariate data:

$$9.4,\ 10,\ 11.7,\ 0,\ 5.1,\ 6.9,\ 9.3,\ 9.9,\ 10.2,\ 4.5,$$
$$6.5,\ 9.7,\ 4.1,\ 0.6,\ 3.3,\ 8.5,\ 10.6,\ 7,\ 10.5,\ 11$$

Use the Shapiro-Wilk normality test and the D'Agostino's K^2 normality test to determine if the univariate data follows a normal distribution. (i) Write down all the Python commands to perform the normal tests; (ii) Write down the outputs by running the Python commands.

(2 marks)

3. In exploratory data analysis, we sometimes need to fit data using linear regression which leads to a least square problem.

Table 1: x-y data.

x	y
1.0	7.3
2.0	15.0
3.5	18.5
4.0	19.7
6.0	29.3

For the data in Table 1, write down the **Python commands** to solve the least square solution for the data by fitting the data using the linear regression formula

$$y = ax + b.$$

The matrix representation of the least square problem are

$$\min_{a,b} \| \begin{bmatrix} 7.3 \\ 15.0 \\ 18.5 \\ 19.7 \\ 29.3 \end{bmatrix} - \begin{bmatrix} 1.0 & 1 \\ 2.0 & 1 \\ 3.5 & 1 \\ 4.0 & 1 \\ 6.0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \|_2^2.$$

(1 mark)