

STAT-2104 Lab (Programming with R and Python Lab)

1. Create the following matrices:

$$A = \begin{pmatrix} 10 & 9 & 20 \\ 2 & -8 & 14 \\ 30 & 15 & 40 \end{pmatrix} \text{ and } B = \begin{pmatrix} 8 & 12 & 1 \\ -2 & 10 & 18 \\ 25 & 46 & 6 \end{pmatrix}$$

Using R and Python programs solve the following questions:

- Find the value in the cell [2,], [3,] and [2,3] of the matrix A.
- Find $A+B$, $A-B$, $A*B$, A/B .
- Find the transpose, determinant and inverse of A. Also show that $A^T A = I$.
- Suppose you are given a vector b consisting of numbers 12, 18 and 20. Solve the linear equation $Ax = b$ for x . Then find the value of the quadratic form $x^T A^{-1} x$.
- Create a new matrix C combining rows of the matrices A and B. Also create another matrix D combining columns of the matrices A and B. Guess what would be the first column of C and what will be the last column of D.

2. The information of 10 patients are shown in the following table:

Patient ID	Sex	Age	Weight (in kg)	Diabetes	Blood group	Medical Expense
1	M	23	56	Type I	O (+ve)	20000
2	F	25	48	No	B (+ve)	15000
3	M	39	60	Type 2	O (+ve)	30000
4	M	NA	65	Type I	AB (+ve)	22000
5	F	36	54	Type 2	O (+ve)	24000
6	M	25	57	No	B (+ve)	10000
7	F	22	45	Type I	AB (-ve)	14000
8	F	NA	50	No	O (-ve)	30000
9	M	60	54	Type 2	A (+ve)	40000
10	F	46	53	Type I	B (+ve)	32000

Using R software solve the following questions:

- Create a data frame for the variable Patient ID, Sex, Weight, Diabetes, Blood group and medical expense with name "patient.data".
- Display the variable Age and Blood group.
- Construct a cross table using the variables Age and Blood group.
- Remove the variable "weight" from the data frame and add the variable Age in the existing data frame with name "New.data".
- Replace the missing value of age of the patient by its mean and find the five number of summary for age of the patients.
- Draw two suitable graph for blood group of the patients.

3. Draw a random sample of size 1000 from

- Normal distribution with mean 50 and standard deviation 5
- Binomial distribution with $n=50$ and $p=0.4$
- Poisson distribution with mean 10

Plot density curve, histogram and cumulative distribution function curve for the generated data using R software.

4. Suppose y represent the production of wheat (in metric ton per acre), x_1 the amount of fertilizer (in pound per acre) and x_2 the amount of rainfall.

Y	1.49	1.87	1.86	2.61	2.43	2.44	3.00	3.12
x_1	100	200	300	400	500	600	700	800
x_2	26	50	26	76	51	50	75	74

Using R and Python programs solve the following questions:

- Plot y on x_1 and x_2 and check is there any linear relation among the three variables.
 - Compute the correlation coefficient between y and x_1 and comment on your result.
 - Fit the regression model of y on x_1 and x_2 . Interpret the parameters of the model. Also compute the residulas.
 - Find the coefficient of determination using the above data and comment on your result.
5. Using nested if condition in R and Python, print the greatest number among the three numbers.
6. Program using nested if condition in R and Python what will enter total marks and assign grades according to the following:

Marks	Grade	Marks	Grade
<40	F	>=60 and <65	B
>= 40 and <45	D	>=65 and <70	B+
>=45 and <50	C	>=70 and <75	A-
>=50 and <55	C+	>=75 and <80	A
>=55 and <60	B-	>=80	A+

If a student gets 52 marks what will the grades of the student.

7. Write an R function that prints out mean, median, mode, standard deviation, skewness, and kurtosis of a set of observations. Compute the result for 1st 50 natural numbers. Also solve these problems using Python program.
8. Write an R and Python functions that will calculate $1^3 + 2^3 + \dots + 99^3$ and $\frac{1}{1^3} \times \frac{1}{3^3} \times \dots \times \frac{1}{99^3}$.
9. Hypothetical data on weekly family consumption expenditure (Y) and weekly family income (X) are given below:

Y	70	65	90	95	110	115	120	140	155	150
X	80	100	120	140	160	180	200	220	240	260

- Write an R function for computing (a) correlation coefficient between X and Y; and (b) regression equation of Y on X.
- Write a python program computing (a) correlation coefficient between X and Y; and (b) regression equation of Y on X.

10. Table below gives body height (in cm) and body weight (in kg) of 10 males (sex=1) and 10 female (sex=2) students of a biometry and statistics course.

SI	Height	Weight	Sex	SI	Height	Weight	Sex
1	197	93	1	1	167	59	2
2	165	59	1	2	176	70	2
3	179	71	1	3	161	57	2
4	191	78	1	4	168	60	2
5	177	72	1	5	164	66	2
6	153	61	1	6	181	67	2
7	169	72	1	7	182	71	2
8	178	29	1	8	143	46	2
9	184	85	1	9	169	53	2
10	177	75	1	10	175	66	2

Compute body mass index (BMI) both for male and female using mapply function. Also compute BMI both for male and female using Python program.

11. Write a Python program to compute the element-wise sum of given tuples.
Original lists:
(1, 2, 3, 4)
(3, 5, 2, 1)
(2, 2, 3, 1)
12. Write a Python program to calculate the average value of the numbers in a given tuple of tuples.
Original Tuple:
((10, 10, 10, 12), (30, 45, 56, 45), (81, 80, 39, 32), (1, 2, 3, 4))
13. Write a Python program for checking (i) prime number, and (ii) even or odd number.
14. Write a R and python program to determine the series:
0+1+1+2+3+5+8+13+21+34+55+89+144
15. Write a Python program to find (i) the circumference and area of a circle, (ii) area of triangle, (iii) area of rectangle use return value function with argument.
16. Write a Python function that takes two lists and returns True if they have at least one common member.
17. Write a Python program to display Dept. of Statistics, PUST 10 times on the screen.
18. Write a python program to print all the months of given year.

19. Generate 50 random numbers from (i) Normal distribution with mean 20 and variance 25, (ii) Exponential distribution with parameter 20 and (iii) Uniform distribution between 0 and 10.

Using Pandas library solve the following questions:

- i. Create a data frame from the above information.
- ii. Show the column two of the data frame and row 6 of the data frame.
- iii. Calculate five number summaries for three distributions.
- iv. Calculate mean, median, mode, skewness and kurtosis for the above data.

20. Create the following NumPy arrays:

- (i) A 1-D array called zeros having 10 elements and all the elements are set to zero.
- (ii) A 1-D array called vowels having the elements 'a', 'e', 'i', 'o' and 'u'.
- (iii) A 2-D array called ones having 2 rows and 5 columns and all the elements are set to 1 and type as int.
- (iv) Use nested Python lists to create a 2-D array called myarray having 3 rows and 3 columns and store the following data:

2.7, -2, -19

0, 3.4, 99.9

10.6, 0, 13

- (v) Using the array created in (iv), solve the following problem using NumPy
 - (a) Find the transpose of myarray.
 - (b) Find the sum of all elements in each row and column.
 - (c) Find the five number summary.