UT 2 Solution

1 a) The following stem-and-leaf plot shows the ages of a group of people in a room.

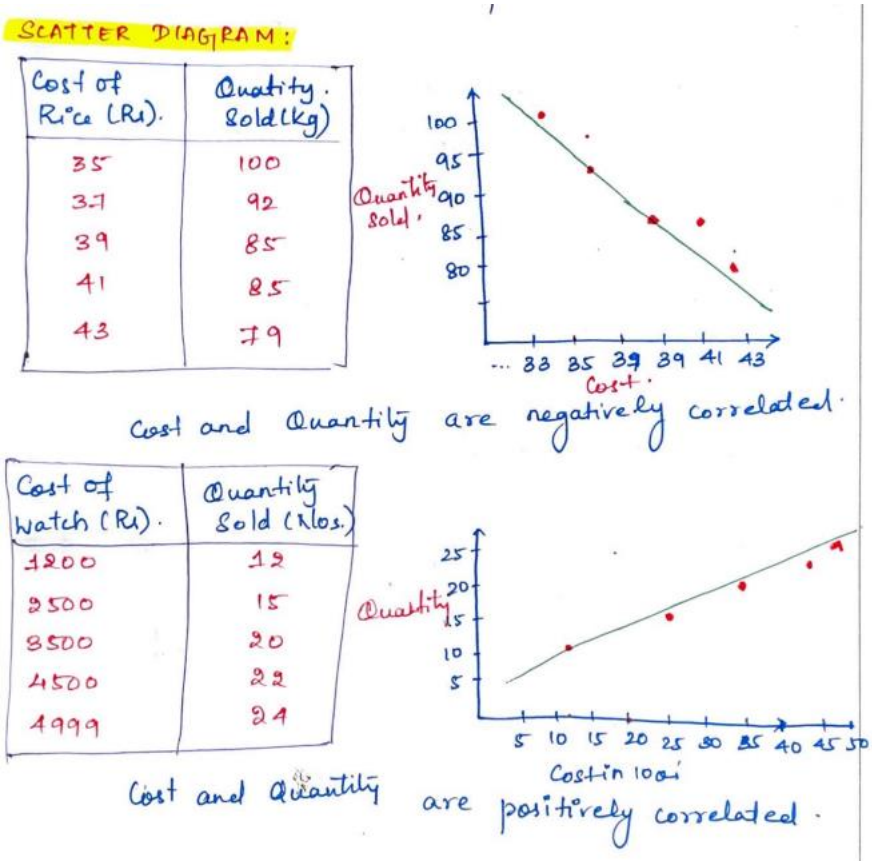
i) How many people were there in the room? - 12

ii) Two people have the same age. What is that age? 22

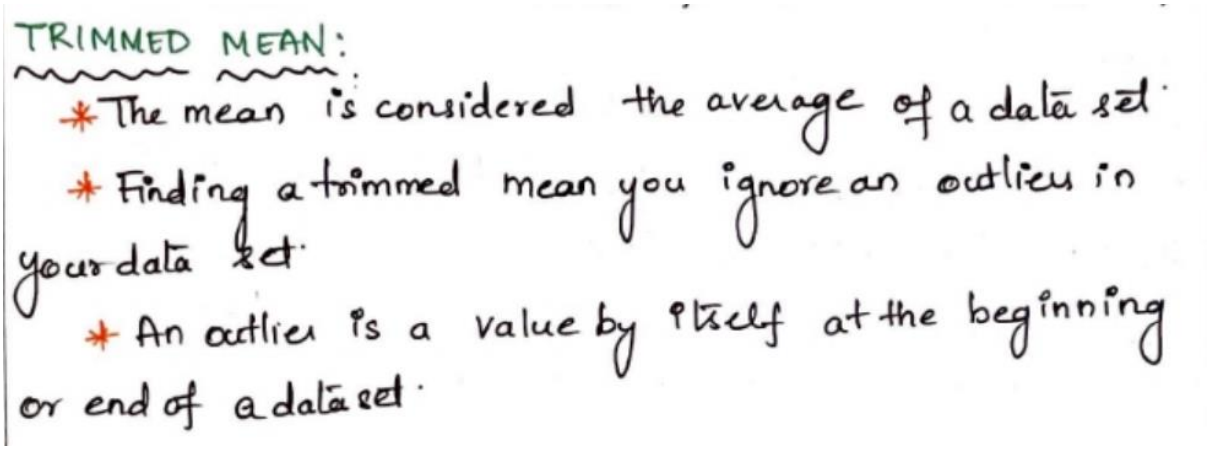
iii) What is the mode, median and mean of the ages? Mean = 26.66 Median = 23 Mode = 22

b) Give the use of scatter plots. Below is a table of 11 student's scores out of 100 on their Maths and English tests. Plot a scatter graph from this data and specify its relation.

Maths mark	38	62	18	75	38	59	66	92	52	75	48
English mark	74	44	85	19	88	69	50	33	29	32	56



c) Define trimmed mean. Charlie recorded the number of pushups he completed each day for 10 days as follows: 5, 4, 7, 6, 8, 10, 11, 0, 7, 18. Calculate the 20% trimmed mean.



Answer = 7.3

d) A garden contains 39 plants. The following plants were chosen at random, and their heights were recorded in cm: 38, 51, 46, 79, and 57. Calculate their heights' standard deviation.

Answer = 15.5

2 a) A trucking company wishes to test the average life of each of the four brands of tyres. The company uses all the brands on randomly selected trucks. The records showing the lives (thousands of miles) of tyres are as given in the table: Test the hypothesis using one-way ANOVA that the average life for each brand of tyres is the same. (Critical value = 5.56)

Brand 1	**Brand 2**	**Brand 3**	**Brand 4**
20	19	21	15
23	15	19	17
18	17	20	16
17	20	17	18

b) To study the performance of 3 detergents and 3 different water temperature, the following readings were obtained with specially designed equipment:

	A	B	C
Cold Water	47	45	50
Warm Water	39	42	52
Hot Water	44	36	48

Perform a two way ANOVA using 5% level of significance. (Critical Value = 6.94).

Solution:

Detergent / Water Temperature	A	B	C
Cold Water (C)	47	45	50
Warm Water (W)	39	42	52
Hot Water (H)	44	36	48.

(1) Calculation of Grand Total and Correction factor:
Data is coded by subtracting any guessed mid value (ie. 40) for easy calculation.

	A	B	C	Total
C	7	+5	+10	+22
W	-1	+2	+12	+13
H	+4	-4	+8	+8
T.	+10	3	30	43

T → Grand Total = 43.

Correction factor = $\frac{T^2}{N}$.

$= \frac{(43)^2}{9} = \frac{1849}{9} = \boxed{205.44}$.

2 Way Anova — Table for Calculation:-

Source of Variation	Sum of Squares	Degree of Freedom	Mean sum of squares	Ratio of F
B/w the Columns.	SSC = 130.89	$\nu = (C-1) = 2$.	MSC = SSC/ν = 130.89/2 = 65.45	MSC/MSE = 65.45/12.28 = 5.32
B/w the rows	SSR = 33.55	$\nu = (r-1) = 2$	MSR = SSR/ν = 33.55/2 = 16.78	MSR/MSE = 16.78/12.28 = 1.37
Residual Error.	SSE = 49.12.	$\nu = (C-1)(r-1) = (2)(2) = 4$	MSE = SSE/ν = 49.12/4 = 12.28.	

(2) Calculation of SSC.

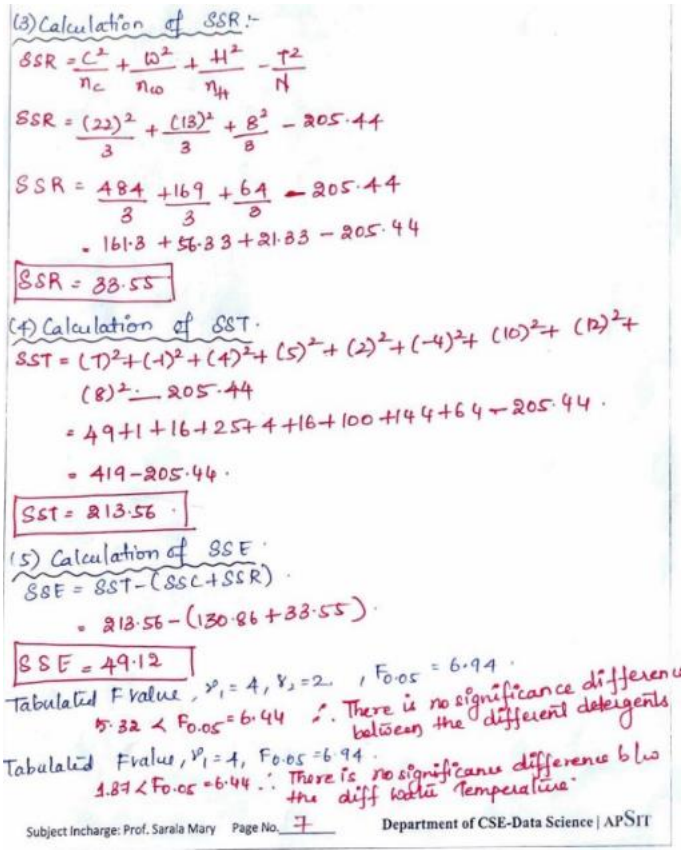
$$SSC = \frac{A^2}{n_A} + \frac{B^2}{n_B} + \frac{C^2}{n_C} - \frac{T^2}{N}$$

$$= \frac{(10)^2}{3} + \frac{(3)^2}{3} + \frac{(30)^2}{3} - 205.44.$$

$$= 100/3 + 9/3 + 900/3 - 205.44$$

$$= 33.33 + 3 + 300 - 205.44$$

$$\boxed{SSC = 130.86}$$



c) In order to the following data represents the number of units of tablet production(in thousands) per day by five different technicians by using 3 different type of machines.

Technicians	Machine X	Machine Y	Machine Z
A	54	48	57
B	56	50	62
C	44	46	54
D	53	48	56
E	48	52	59

Conduct a Friedman Test with the given data and judge whether there is any difference among the machines. (Tabulated value = 5.99)

Solution:

H_0 = There is no difference among machines.

H_a = There is difference between machines.

$H_0 : X = Y = Z$

$H_a : X \neq Y \neq Z$.

Technicians	Machine X	Machine Y	Machine Z	R_1	R_2	R_3
A	54	48	57	2	1	3
B	56	50	62	2	1	3
C	44	46	54	1	2	3
D	58	48	56	2	1	3
E	48	52	59	1	2	3
	R_1			8	7	15
	R_1^2			64	49	225

$$FM = \left(\frac{12}{(n \times k \times (k+1)}\right) \times \Sigma R^2 - [3 \times n \times (k+1)]$$

$n = 5, k = 3, \Sigma R^2 = 64 + 49 + 225 = 338$

$$FM = \left(\frac{12}{5 \times 3 \times (3+1)}\right) \times 338 - [3 \times 5 \times (3+1)]$$

$$FM = \left(\frac{12}{60}\right) \times 338 - [15 \times 1]$$

$FM = 0.2 \times 338 - 60$

$FM = 67.6 - 60 = \boxed{7.6}$

Degree of freedom $r = k - 1 = 3 - 1 = 2$

$\chi^2_{0.05,2} = 5.99$.

$FM_{cal} = 7.6 > \chi^2_{0.05,2} = 5.99$.

H_0 is rejected. There is difference between the 3 machines.

d) Consider there are three groups and their reaction time is measured. Check whether there is difference between the groups using Kruskal Wallis Test at 5% level of significance.

A	B	C
34	44	35
36	37	39
41	45	42
43	33	46

Solution:

H_0: There is no difference between the groups. } Step 2.

H_a: There is difference between the groups.

Step 3 : Assign Ranks

Group A	R_1	Group B	R_2	Group C	R_3
34	2	44	10	35	3
36	4	37	5	39	6
41	7	45	11	42	8
43	9	33	1	46	12
$n_1 = 4$	$\Sigma R_1 = 22$	$n_2 = 4$	$\Sigma R_2 = 27$	$n_3 = 4$	$\Sigma R_3 = 29$

$N = n_1 + n_2 + n_3 = 4 + 4 + 4 = 12$

Step 3: Calculate H value.

$$H = \frac{12}{N(N+1)} \times \left(\frac{\Sigma R_1^2}{n_1} + \frac{\Sigma R_2^2}{n_2} + \frac{\Sigma R_3^2}{n_3} + \cdots + \frac{\Sigma R_k^2}{n_k} \right) - 3(N+1)$$

$$= \frac{12}{12(12+1)} \times \left(\frac{(22)^2}{4} + \frac{(27)^2}{4} + \frac{(29)^2}{4} \right) - 3(12+1)$$

$$= 0.076 \times \left(\frac{484}{4} + \frac{729}{4} + \frac{841}{4} \right) - 3(13)$$

$$= 0.076 \times (121 + 182.25 + 210.25) - 39$$

$$= 0.076 (513.5) - 39$$

$$= 39.026 - 39$$

$H = 0.026$

Degree of freedom = 3-1 = 2

$0.026 < \chi^2_{0.05,2} = 5.991$.

Null hypothesis is accepted.

3 a) Find the linear regression of the data of weekend product sales(in Thousands) given in table. Use Linear regression in matrix form. Predict the 7th week sale.

X(Week)	Y(Sales in thousands)
1	1
2	3
3	4
4	8

Solution:

Here the independent variable X is be given as:

$$X^T = [1\ 2\ 3\ 4]$$

The dependent vasiables is given as follows:

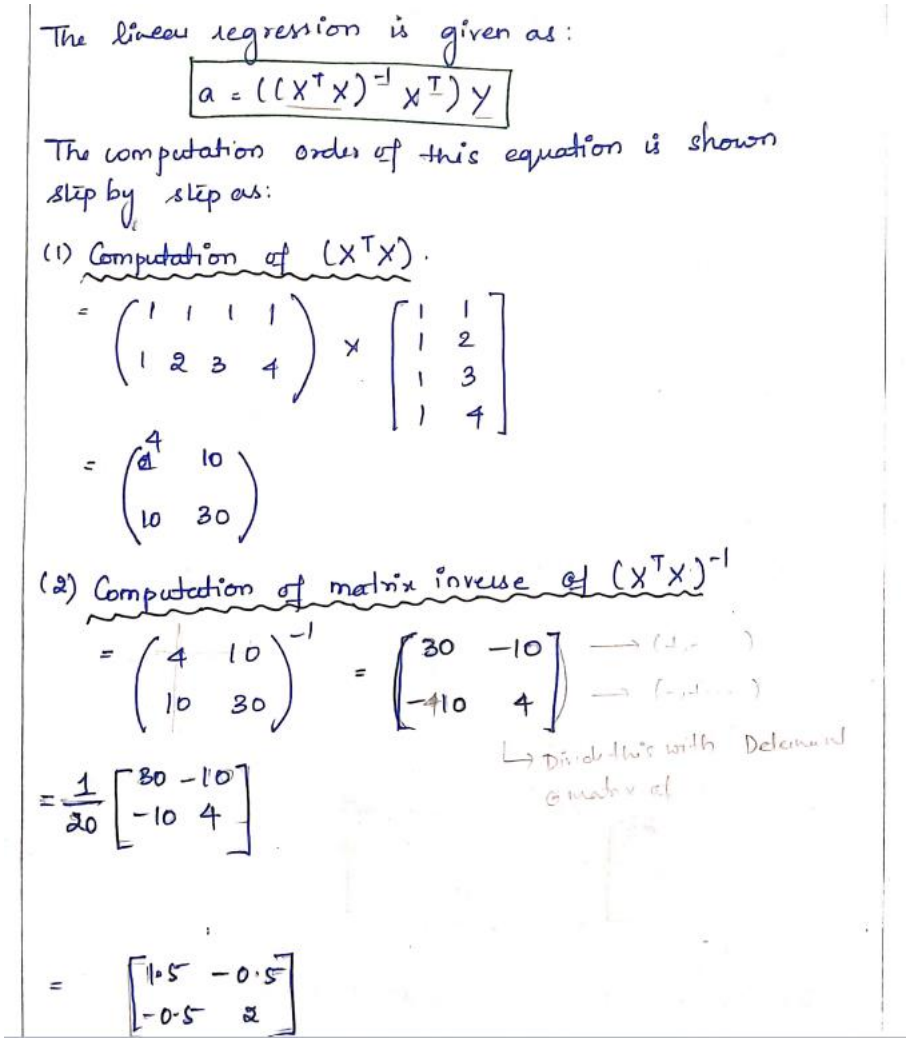
$$Y^T = [1\ 3\ 4\ 8]$$

The data can be given in matrix form as follows.

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \quad Y = \begin{bmatrix} 1 \\ 3 \\ 4 \\ 8 \end{bmatrix}$$

↓

The first coloumn is used for setting bias.



b) Find the multiple regression equation using the below data:

Subject	Y	X1	X2
1	-3.7	3	8
2	3.5	4	5
3	2.5	5	7
4	11.5	6	3
5	5.7	2	1

c) Find the simple linear regression equation using the below data:

Hour	Temp
2	21
4	27
6	29
8	86
10	86
12	92

d) Find the value of the correlation coefficient from the data given in the following table:

Subject	Age(X)	Glucose Level(B)
1	43	99
2	21	65
3	25	79
4	42	75
5	57	87
6	59	81