

No. of Printed Pages : 8 **MSTL-012(Set-I)**

M. SC. (APPLIED STATISTICS)
(MSCAST)

Term-End Practical Examination
December, 2024

MSTL-012(Set-I) : STATISTICAL COMPUTING
USING R-II

Time : 3 Hours

Maximum Marks : 75

Note : (i) Attempt any *three* questions.

(ii) Solve the questions in R software and create script file.

(iii) Mention necessary steps, hypotheses, interpretation, etc.

(iv) Symbols have their usual meanings.

1. A meteorological department recorded the rainfall (in cm) for 20 days in City A and 18 days in City B which are given in the following table :

Day	City A	City B	Day	City A	City B
1	106	116	11	106	115

2	102	129	12	103	131
3	105	138	13	94	124
4	100	118	14	114	130
5	112	111	15	118	124
6	100	143	16	106	102
7	105	134	17	98	122
8	96	123	18	99	116
9	104	136	19	107	
10	98	108	20	102	

(i) Determine all possible samples of sizes $n_1 = n_2 = 3$.

(ii) Obtain the sampling distribution of the ratio of two sample variances for $n_1 = n_2 = 3$.

(iii) Plot the histogram for the sampling distribution of the ratio of two variances and overlay the corresponding curve of its sampling distribution.

(iv) Assuming equal population variances, obtain the probability that the ratio of variances of the rainfall in a simple lies between 1 and 2 for $n_1 = n_2 = 10$.

- (v) Are variances of the rainfall in both cities equal at 1% level of significance, assuming the given data as a sampled data ?

3+5+7+3+7

2. To study the relationship between income and expenditure in a household survey, the monthly income and expenditure of 60 families are given in the following table :

S. No.	Income	Expenditure	S. No.	Income	Expenditure
1	25,600	12,700	31	14,600	9,400
2	14,600	7,600	32	14,000	6,500
3	5,800	3,800	33	20,800	13,000
4	28,800	23,900	34	27,800	16,600
5	9,600	7,300	35	9,600	5,900
6	14,600	7,900	36	18,200	10,000
7	20,200	14,700	37	22,800	12,200
8	20,400	12,600	38	14,400	7,700
9	12,600	8,300	39	6,400	3,800
10	7,600	5,000	40	20,600	16,300
11	27,800	26,300	41	4,000	3,000
12	19,600	9,300	42	9,400	5,400
13	23,400	14,900	43	21,600	18,900

[4]

MSTL-012(Set-I)

14	27,800	11,100	44	26,600	13,800
15	17,800	12,400	45	24,800	14,000
16	13,600	10,300	46	7,600	4,800
17	5,800	4,600	47	14,000	7,591
18	26,200	10,000	48	24,600	21,500
19	17,200	11,600	49	22,000	13,000
20	9,400	7,700	50	19,400	9,100
21	13,800	7,900	51	7,200	4,300
22	9,400	5,600	52	18,400	9,200
23	16,000	9,900	53	15,600	10,100
24	10,200	6,800	54	24,000	15,700
25	7,000	4,900	55	4,600	3,080
26	5,800	4,500	56	9,000	3,800
27	22,000	13,500	57	27,000	17,600
28	11,400	7,400	58	6,800	4,500
29	7,800	7,300	59	29,200	13,100
30	29,200	19,800	60	14,000	10,400

- (i) Fit a linear regression model for expenditure on income.
- (ii) Apply the logarithmic and square-root transformations to the response variable.
- (iii) Fit the regression models using the transformed response variable.

- (iv) Draw the residual and Q-Q plots for the original and transformed models.
- (v) Explain which transformation mentioned in part (ii) helped in meeting the assumptions of constant error variance and normality in regression analysis.

2+4+6+10+3

3. The following data, consider different nutritional contents present in the 250 gm bowl of 12 different food items :

Item	Protein (gm)	Carbohydrates (gm)	Fat (gm)	Energy (kcal)	Calcium (mg)
1	6	25	1	200	150
2	8	28	3	150	170
3	10	32	2	170	175
4	3	30	3	110	130
5	2	24	4	120	120
6	4	26	6	105	190
7	6	18	8	108	140
8	7	15	3	170	90
9	10	19	2	160	100
10	5	24	5	190	120
11	12	23	6	170	95
12	5	20	8	150	105

- (i) Cluster the given data using the K-means algorithm, taking $K = 2, 3$ and 4 .
 - (ii) Visualise the clusters formed for each of the K values considered in part (i).
 - (iii) Obtain the summary of the classifications obtained for each value of K and interpret the results.
 - (iv) Determine the ideal number of clusters to classify the data. $7+7+7+4$
4. (a) Consider the following data on length (mm), width (mm) and height (mm), of 12 males and 12 females painted turtles :

Male			Female		
Length	Width	Height	Length	Width	Height
93	74	37	98	81	38
96	80	35	103	86	42
102	85	38	109	88	44
104	83	39	123	95	46
107	82	38	133	102	51
113	88	40	134	100	48
116	90	43	138	98	51

117	91	41	141	105	53
120	89	40	149	107	55
121	95	42	155	115	63
127	96	45	158	115	62
131	95	46	162	124	61

Assume that the observations for male and female turtles are as 12 pairs of observations. Test whether the population mean vectors for male and female populations of painted turtles can be assumed to be equal at 5% level of significance.

- (b) The following data of ten products are recorded to explore the relationship between monthly sales and prices of the products :

Product	Price	Sale
1	75	4,740
2	32	2,640
3	112	21,000
4	54	4,700

[8]

MSTL–012(Set-I)

5	64	12,000
6	128	21,800
7	150	28,300
8	212	47,000
9	121	10,000
10	230	41,000

Detect the influence points using the :

11 + 14

(i) Cook's distances; and

(ii) DFBETAS

x x x x x x x