

No. of Printed Pages : 9 **MSTL-012(Set-II)**

M. SC. (APPLIED STATISTICS)
(MSCAST)

Term-End Practical Examination
December, 2024

MSTL-012(Set-II) : 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. The responses of 50 male and female customers who purchased or did not purchase the mobile phone of a famous brand, in terms of 'Yes' or

‘No’, respectively, are collected separately. The recorded data are given as follows :

Customer	Male	Female	Customer	Male	Female
1	Yes	No	26	No	No
2	Yes	No	27	Yes	Yes
3	No	Yes	28	Yes	No
4	Yes	No	29	No	No
5	Yes	Yes	30	Yes	Yes
6	Yes	No	31	No	Yes
7	No	Yes	32	Yes	No
8	No	Yes	33	Yes	Yes
9	Yes	No	34	No	Yes
10	Yes	Yes	35	Yes	No
11	No	No	36	No	Yes
12	Yes	Yes	37	Yes	No
13	No	No	38	No	Yes
14	Yes	No	39	Yes	No
15	No	Yes	40	No	Yes
16	Yes	No	41	Yes	No
17	No	Yes	42	Yes	Yes
18	Yes	No	43	Yes	No
19	No	Yes	44	Yes	Yes

20	Yes	No	45	No	No
21	No	No	46	Yes	Yes
22	Yes	Yes	47	Yes	Yes
23	No	Yes	48	No	No
24	Yes	No	49	No	Yes
25	No	Yes	50	Yes	No

- (i) Determine the sampling distribution of the difference between two proportions for $n_1 = n_2 = 35$ (taking 10000 replications).
- (ii) Plot the histogram for the sampling distribution obtained in part (i) and overlay the normal curve on it.
- (iii) Obtain the probability that the difference between sample proportions is more than 0.
- (iv) Determine the 95% confidence interval for the difference between proportions of customers who purchased mobile phones in male and female groups, assuming the given data as a sampled data. 8+7+3+7

2. The total revenue (in ₹ lakhs) and operating cost (in ₹ thousands) of three Departments A, B and C for 40 months are given in the following table :

Department A		Department B		Department C	
Revenue	Operating Cost	Revenue	Operating Cost	Revenue	Operating Cost
290	140	400	260	440	310
310	140	370	200	450	300
270	100	330	200	510	340
340	160	330	210	410	300
350	190	400	250	370	260
350	180	380	220	420	280
390	180	440	260	500	320
350	160	440	250	520	360
210	100	380	230	420	290
340	160	400	250	480	340

280	100	410	230		
280	110	350	220		
390	190	380	210		
250	100	360	220		
360	170				
350	160				

- (i) Create indicator variable corresponding to the departments.
- (ii) Use the given data to develop a multiple linear regression model with interaction terms and then obtain the regression models for each department.
- (iii) Construct a scatter plot using different colours for departments and draw the fitted regression lines on it.
- (iv) Draw the residual and Q-Q plots to check the model's assumptions. Also, interpret the results.

2+7+8+8

3. The following data consider the texture measurements for 35 food items on 5 variables :

Oil (% oil), Density (product's density), Crispy (crispiness), Fracture (angle, in degrees, through which it slowly bent before it fractures) and hardness (sharp point to measure the force required before breakage occurs) :

Food Item	Oil	Density	Crispy	Fracture	Hardness
1	16.5	2955	10	23	97
2	17.7	2660	14	9	139
3	16.2	2870	12	17	143
4	16.7	2920	10	31	95
5	16.3	2975	11	26	143
6	19.1	2790	13	16	189
7	18.4	2750	13	17	114
8	17.5	2770	10	26	63
9	15.7	2955	11	23	123
10	16.4	2945	11	24	132
11	18.0	2830	12	15	121
12	17.4	2835	12	18	172
13	18.4	2860	14	11	170
14	13.9	2965	12	19	169
15	15.8	2930	9	26	65

16	16.4	2770	15	16	183
17	18.9	2650	14	20	114
18	17.3	2890	12	17	142
19	16.7	2695	13	13	111
20	19.1	2755	14	10	140
21	13.7	3000	10	27	177
22	14.7	2980	10	20	133
23	18.1	2780	13	14	150
24	17.2	2705	8	27	113
25	18.7	2825	13	20	166
26	18.1	2875	12	15	150
27	16.6	2945	10	25	100
28	17.1	2920	10	25	123
29	17.4	2845	13	19	129
30	19.4	2645	12	18	68
31	15.9	3080	10	23	106
32	17.1	2825	10	28	131
33	15.5	3125	7	33	92
34	17.7	2780	13	22	141
35	15.9	2900	12	21	192

Carry out the factor analysis using the number of factors equal to 2 and 3. Interpret the findings. 25

4. (a) Consider the following data on yearly sales (in thousands) and advertisement costs (in thousands) of 16 products :

S. No.	Sale	Advertisement Cost	S. No.	Sale	Advertisement Cost
1	120	15	9	124	21
2	115	16	10	125	22
3	118	18	11	126	23
4	116	19	12	126	24
5	118	19	13	127	25
6	119	19	14	127	25
7	121	20	15	128	26
8	123	21	16	129	27

Apply the Durbin-Watson test for testing the autocorrelation among the error terms of the regression model fitted on sales and advertisement cost.

- (b) The following matrix shows the distances (in '00 km) between 5 manufacturing units of a company :

	M_1	M_2	M_3	M_4	M_5
M_1	0				
M_2	6	0			
M_3	14	10	0		
M_4	8	18	24	0	
M_5	22	11	16	15	0

- (i) Cluster the manufacturing units M_1, M_2, M_3, M_4 and M_5 using the Average-Linkage method.
- (ii) Draw the dendrogram for this method. 10+15

× × × × × × ×