No. of Printed Pages : 6 **MSTL–012(Set-I)**

M. SC. (APPLIED STATISTICS)
(MSCAST)
Term-End Practical Examination
June, 2025

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 using R software and create script file.

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

(iv) Symbols have their usual meanings.

1. To study impact of a new fertilizer for obtaining better quality in fruits, a scientist has applied old (O) and new(N) fertilizers on two parts of an orchard. The following table shows weights of fruits (in grams) which are grown by separately using the two fertilizers :

Fruit, Veriety	New Fertilizer (O)	Old Fertilizer (N)
1	87	74
2	99	75
3	82	89
4	100	82
5	92	85
6	90	80
7	93	78
8	85	87
9	94	79
10	91	85
11	90	88
12	89	83
13	91	90
14	88	76
15	–	93
16	–	84

Compute the following in case of known population variances :

(i) Find all possible samples with replacement of size $n = 4$ for each fertilizer O and N.

(ii) Obtain the sampling distribution of the difference between means for each fertilizer.

(iii) Plot the histogram for the sampling distribution obtained in part (ii) and plot the corresponding frequency curve overlaid with appropriately choosen probability distribution.

(iv) Assuming the given data as sampled data, test whether the average weights of fruits for both fertilizers are equal at 5% level of significance. 4+10+4+7

2. To study impact of age (in years), weight (in kgs), stress level and walking time (in minutes) on blood sugar (in mg/dL) of human beings, a researcher has recorded the following data on 15 patients :

Blood Sugar (B)	**Age (A)**	**Weight (W)**	**Stress (S)**	**Walking Time (T)**
130	38	55	11	25
135	39	60	13	30
140	43	64	14	29

126	41	56	15	31
150	42	65	7	35
160	44	70	5	27
170	46	75	14	25
180	47	77	16	35
190	51	79	17	40
200	54	74	19	45
210	57	71	22	30
230	58	73	21	35
160	61	68	23	55
170	52	69	24	50
185	63	73	25	53

Fit an appropriate linear model for the response variable (B) using stepwise regression method first (i) manually and then by (ii) using a built-in function in R. 25

3. Consider the "Iris" dataset from the R environment, in which the data of 50 flowers for each of the three species (Setosa, Versicolor and Virginica) on four variables (sepal-length, sepal-width, petal-length and petal-width) are given.

Assuming multivariate normality and equal variance-covariance matrices for the above three species, test equality of mean vectors of four variables for three species at 5% level of significance. 25

4. (a) Suppose four random variables are jointly distributed as a Multivariate Normal Distribution, with mean vector $\underset{\sim}{\mu} = (4, 12, 15, 20)$ and variance-covariance matrix given as :

$$\Sigma = \begin{bmatrix} 3 & 1 & 2 & -1 \\ 1 & 2 & 3 & -2 \\ 2 & 3 & 2 & 1 \\ -1 & -2 & 1 & 4 \end{bmatrix}$$

(i) Generate 20 observations for each of *four* random variables using the Cholesky-Decomposition method.

(ii) Obtain eigen values and eigen vectors from the above variance-covariance matrix.

(b) Data on an average duration of walk (in minutes), serum creatinine (in mg/dL) and random blood sugar level

(in mg/dL) is assigned '1' for high and '0' for normal blood sugar level. Data for 70 diabetic patients are given as follows :

S. No.	Duration of Walk	Serum Creatinine	No. of Patients Having High Sugar	Total Number of Patients
1	25	0.5	15	20
2	30	0.6	4	8
3	35	0.7	3	10
4	40	0.4	9	12
5	45	0.5	8	16
6	50	0.7	1	4

Fit a multiple logistic regression model to the above data. Also, estimate variance for the estimators of the model coefficients β_0, β_1 and β_2. 10+15

× × × × ×