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

M. Sc. (APPLIED STATISTICS)

(M. SC. AST)

Term-End Practical Examination

June, 2025

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 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 technology on obtaining better quality of potatoes, a researcher has applied old(O) and new(N) technologies on two distinct fields. Following table presents weights of potatoes (in kgs)

obtained by using the old and new technologies for cultivation :

S. No.	New	Old
1	98	74
2	86	73
3	81	88
4	99	81
5	91	84
6	87	83
7	90	80
8	93	78
9	85	87
10	94	79
11	91	85
12	90	88
13	89	83
14	91	90
15	83	76
16	98	93
17	–	91
18	–	80

If population variances are not known for the above data, then : 8+7+3+7

(i) Find all possible samples with replacement of size $n = 3$ for each technique.

(ii) Obtain sampling distribution of the differences between the two techniques for $n = 3$.

(iii) Compute standard error of the sampling distribution obtained in (ii), in case of unequal population variances.

(iv) Assuming given data as the sampled data, test whether the average weights of potatoes are equal for the two techniques at 5% level of significance.

2. A study was conducted on 1000 patients suffering from a known disease D. Some of them were exposed to a particular bacteria. These sample patients were classified based on different stages of disease D, for exposed

and non-exposed groups tabulated as follows : 25

Exposed	Disease	Count
Yes	Stage I	230
Yes	Stage II	155
Yes	Stage III	265
Yes	Stage IV	110
No	Stage I	55
No	Stage II	40
No	Stage III	85
No	Stage IV	60

(i) Fit the (i) adjacent category and (ii) cumulative logit models.

(ii) Also compute odds ratios for the above fitted models.

3. Consider the "Iris" data from R environment consisting of measurements on four variables : Sepal length, Sepal width, Petal Length and Petal width. Sample data was collected for 150 flowers belonging to one of

the above three species of flowers. Answer the following : 10+5+5+5

(i) Obtain Fisher's linear discriminate functions manually to classify a given flower into one of the three known species.

(ii) Classify the observation

$$x_0^1 = (5.1, 3.5, 1.4, 0.2)$$

by using the Fisher's classification criteria.

(iii) Carry out linear discriminate analysis by using a built-in function in R. Hence interpret the results.

(iv) Visualize how well the above obtained discriminant function has separated three different species.

4. (a) Data on the length (in cm, height in cms), and weight (in gms), for bricks produced at two different factories are given as follows : 15

First Factory			
S. No.	Length (X_1)	Height (X_2)	Weight (X_3)
1	20	6	300
2	22	7	320
3	21	6.5	315
4	22	6.8	330
5	27	7.1	380
6	25	6.4	400
7	20	6.9	290
8	24	6	340
9	24	6.5	450
10	26	7	390

Second Factory			
S. No.	Length (X_1)	Height (X_2)	Weight (X_3)
1	20	6.5	315
2	24	6.8	380
3	21	5.9	320

4	22	6.7	330
5	24	7	310
6	26	6.9	400
7	21	6	370
8	22	6.2	360
9	24	5.8	350
10	25	6.7	375
11	20	6	350

Test whether the population mean vectors for the bricks produced in these factories can be assumed to be equal at 5% level of significance. It is given that the brick production in the two factories is independent.

(b) In *"mtcars"* dataset of the R environment, consider variables : mileage per gallon (*mpg*), cylinder displacement (*disp*), and number of cylinders (*cyl*) to ascertain the relationship between the response

variable (*mpg*) and explanatory variables (*disp* and *cyl*). Answer the following : 10

(i) Create indicator variables relating to explanatory variable *cyl*.

(ii) Use this data to fit a linear regression model with interaction.

(iii) Test significance of the fitted model at 1% level of significance.

× × × × ×