No. of Printed Pages : 4 **MSTL–014(Set-I)**

M. SC. (APPLIED STATISTICS)

(MSCAST)

Term-End Practical Examination

June, 2025

MSTL-014(Set-I) : DATA ANALYSIS WITH PYTHON LAB

Time : 2 Hours *Maximum Marks : 25*

Note : *(i) Attempt any **one** question.*

(ii) Solve the question using Python software and create script file.

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

(iv) Symbols have their usual meanings.

1. (a) A packaging process claims to fill small boxes with approximately 50 raisins. However, the number of raisins in each box may vary. Suppose 100 boxes are

randomly sampled, and the number of raisins in each box is counted. The recorded data are given in the following table :

57	51	53	52	50	60	51	51	52	52
44	53	45	57	39	53	58	47	51	48
49	49	44	54	46	52	55	54	47	53
49	52	49	54	57	52	52	53	49	47
51	48	55	53	55	47	53	43	48	46
54	46	51	48	53	56	48	47	49	57
55	53	50	47	57	49	43	58	52	44
46	59	57	47	61	60	49	53	41	48
59	53	45	45	56	40	46	49	50	57
47	52	48	50	45	56	47	47	48	46

Answer the following :

(i) Construct a continuous frequency distribution for the given data. 8

(ii) Also, construct a histogram. What does the histogram reveal about the filled boxes ? 7

(b) Suppose a researcher wants to evaluate the effect of age on the presence or absence of Cervical Spondylosis (CS). For this purpose, the following data on CS (where 0 indicates absence and 1 indicates presence of CS) and age (in years) were obtained for 15 employees of an organisation : 10

Age	CS
25	0
30	0
35	0
28	0
40	1
45	0
50	1
24	1
34	1
48	1
34	1
42	1
38	1
22	1
49	0

Fit a logistic regression model.

2. Load the iris dataset in Python, in which the data on 50 flowers are given for each of the three species, namely setosa, versicolor, and virginica, corresponding to four variables : Sepal-Length, Sepal-Width, Petal-Length and Petal-Width.

   (i) Create a pairwise scatter plot for the given variables. 8

   (ii) Carry out PCA using the given four variables and interpret. 8

   (iii) Construct a scree plot and interpret it. How many principal components should be retained ? 9

× × × × ×