No. of Printed Pages : 7 **MSTL–003(Set-I)**

POST GRADUATE DIPLOMA IN APPLIED STATISTICS (PGDAST)

Term-End Practical Examination

June, 2025

MSTL-003(Set-I) : BIOSTATISTICS LAB

Time : 3 Hours *Maximum Marks : 50*

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

(ii) Solve the questions in Microsoft Excel.

(iii) Use of Formulae and Statistical Tables, Booklet for PGDAST is allowed.

(iv) Mention necessary steps, hypotheses, interpretations, etc.

1. A random sample of 495 patients who visited to a hospital was taken and their sitting-working hours along with the status of lower back pain on 5-point Likert scale were recorded :

Sitting (in hours)	Lower Back Pain				
	No Pain	Mild	Moderate	Severe	Very Severe
< 2	18	24	8	7	6
2 to 4	22	21	17	15	7
4 to 6	20	19	15	13	11
6 to 8	18	19	16	18	14
8 to 10	8	17	21	21	23
10 to 12	6	15	28	23	25

Test whether sitting working hours and lower back pain are associated with each other or not at 5% level of significance. 15

(b) A researcher collected data on Obstructive Coronary Artery Disease (OCAD) and hypertension among subjects identified by a treadmill stress test, given as follows :

Risk Factor (Hypertension)	OCAD	
	Cases	Non-cases
Present	14	1
Absent	9	8

Test whether there is an association between hypertension and occurrence of OCAD at 1% level of significance. 10

2. A researcher is interested to determine the relationship between the serum creatinine (in mg/dL) with the weight (in kg) and gender ('0' for female and '1' for male). The data for 30 patients were collected from the hospital records and are given as follows :

S. No.	Serum Creatinine	Weight	Gender
1	0.7	45	1
2	1.3	64	1
3	1.0	58	1
4	1.5	83	0
5	1.7	91	1
6	1.5	78	1
7	1.0	53	0
8	0.7	49	1
9	0.5	42	0
10	1.6	86	0
11	1.1	52	1
12	0.8	53	0
13	1.3	64	1

14	1.1	61	1
15	1.2	71	0
16	1.1	54	1
17	0.9	54	0
18	0.9	61	0
19	1.1	64	0
20	0.8	53	0
21	0.5	44	0
22	0.6	44	1
23	1.0	61	0
24	0.6	40	0
25	1.0	57	1
26	0.5	64	1
27	0.9	66	1
28	1.3	41	1
29	1.1	61	0
30	1.4	64	1

(a) Fit a linear regression model and perform its related analysis at 5% level of significance.

(b) Check whether the fitted regression model satisfies the linearity and normality assumptions. 15+10

3. A group of cancer patients were randomly assigned either Treatment A or Treatment B. They were followed for 300 days to record their time of death. The survival data are as follows :

Patient ID	Survival Time	Treatment	Outcome (1-died, 0-consored)
001	193	1	1
002	97	2	1
003	235	2	1
004	300	1	0
005	203	1	1
006	76	2	0
007	43	2	0
008	98	1	1
009	131	2	0
010	265	1	1
011	62	2	0

012	200	1	1
013	261	1	1
014	300	1	1
015	220	2	1
016	254	1	1
017	37	2	0
018	267	1	1
019	36	2	0
020	267	1	1
021	36	2	0
022	111	1	0
023	67	2	1
024	175	2	0
025	192	1	0
026	102	2	1
027	300	1	0
028	250	1	0
029	145	1	1
030	90	2	1
031	149	2	1
032	206	1	0

(a) Estimate the Kaplan-Meier probabilities for both treatments.

(b) Construct the Kaplan-Meier survival curve for both treatments.

(c) Test whether there is a significant difference between the distributions of survival times of the patients under Treatment A and Treatment B at 5% level of significance. 10+10+5

× × × × ×