

No. of Printed Pages : 8 **MSTL-003(Set-II)**

**POST GRADUATE DIPLOMA IN
APPLIED STATISTICS (PGDAST)**

Term-End Practical Examination

December, 2024

MSTL-003 : 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,
interpretation, etc.

1. (a) A group of 75 women was participated in a study which were divided into two groups : W_1 and W_2 according to their nature of work. The data so obtained with 2×2 contingency table based on their anxiety

level and working group was as given below :

Anxiety	Working Group		Total
	W_1	W_2	
Yes	3	19	22
No	31	22	53
	34	41	75

Test whether there is association between anxiety level and working group at 1% level of significance. 10

- (b) A hypothetical data of 40 leukaemia patients who received stem cell transplant were followed up from transplant to 20 months. The data are given as follows : 15

Patient ID	Survival Time	Status
P01	6	Event
P02	3	Unknown
P03	4	Unknown
P04	11	Unknown
P05	3	Unknown
P06	1	Event
P07	15	Event

P08	20	Unknown
P09	2	Unknown
P10	2	Event
P11	20	Unknown
P12	17	Unknown
P13	1	Event
P14	2	Event
P15	20	Unknown
P16	1	Event
P17	16	Unknown
P18	2	Event
P19	2	Event
P20	7	Event
P21	4	Event
P22	18	Event
P23	7	Unknown
P24	15	Event
P25	2	Event
P26	17	Event
P27	20	Unknown
P28	3	Event
P29	9	Unknown

P30	2	Event
P31	4	Event
P32	4	Event
P33	5	Unknown
P34	7	Event
P35	12	Unknown
P36	12	Unknown
P37	3	Event
P38	6	Event
P39	8	Unknown
P40	10	Unknown

- (i) Construct the K-M survival curve.
 - (ii) Find the K-M survival probabilities.
 - (iii) Find the standard error of the estimate of the survival function.
 - (iv) Obtain 95% confidence interval for the survival function.
2. A study was conducted on three different groups of participants to evaluate the relationship between cholesterol and weight. The data on serum cholesterol (in mg/dL) and

weight (in kg) were obtained for three groups (A, B and C) of 40 participants as given below :

25

S. No.	Cholesterol (mg/dL)	Weight (kg)	Group
1	470	82	A
2	420	68	A
3	465	82	A
4	510	99	B
5	525	95	A
6	430	72	A
7	470	90	B
8	470	82	A
9	440	84	B
10	390	72	B
11	445	74	A
12	500	95	B
13	490	86	A
14	370	56	A
15	310	51	B
16	470	95	C
17	310	59	C

18	570	102	A
19	410	68	A
20	420	84	C
21	510	98	B
22	370	70	C
23	450	90	C
24	440	88	C
25	390	80	C
26	400	82	C
27	358	70	C
28	310	60	C
29	320	62	C
30	400	80	C
31	430	78	B
32	530	104	B
33	370	64	B
34	350	60	B
35	410	76	B
36	550	106	B
37	280	55	C
38	400	72	B
39	380	74	C
40	370	72	C

- (i) Prepare a scatter plot to get an idea about the relationship among the variables.
 - (ii) Fit a linear regression model and its related analysis at 5% level of significance.
 - (iii) Check the linearity and normality assumptions for the regression analysis.
3. A researcher wants to study the dependence of random blood sugar level on the average duration of walk and serum creatinine. The data on the average duration of walk (in minutes), serum creatinine (in mg/dL) and random blood sugar level (in mg/dL) (1 is assigned for high blood sugar level and 0 for normal sugar level) of 70 diabetic patients are recorded as given below :

25

S. No.	Duration of walk (minutes)	Serum Creatinine	No. of patients having high blood sugar	Total number of patients
1	25	0.6	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

Then :

- (i) Fit a multiple logistic regression model.
- (ii) Determine the variance for the estimators of β_0, β_1 and β_2 .
- (iii) Test the significance of the individual model coefficients β_0, β_1 and β_2 at 1% level of significance.
- (iv) Obtain the 99% confidence intervals for β_0, β_1 and β_2 .

x x x x x x x