

MSTL-003 (Set-1)

POST GRADUATE DIPLOMA IN APPLIED STATISTICS (PGDAST)

Biostatistics Lab

Duration : 3 hours

Maximum Marks : 50

- Note: 1. Attempt any two questions.
2. Solve the questions in Microsoft Excel.
 3. Use of “Formulae and Statistical Tables Booklet of PGDAST is allowed.
 4. Mention necessary steps, hypotheses, interpretations, etc.

1. A study was conducted to investigate the relationship between stress and certain other variables. The following data were collected on a sample of 15 corporate executives:

Measure of Stress	Measures of Firm Size	No. of Years in Present Position	Annual Salary (in thousands)	Age
101	812	15	300	38
60	334	8	200	52
10	377	5	200	27
27	303	10	540	36
89	297	13	52	34

60	505	4	27	45
16	401	6	26	50
184	177	9	52	60
34	598	16	34	44
17	412	2	28	39
78	127	8	42	41
141	601	11	84	58
11	205	4	31	51
104	603	5	38	63
76	484	8	41	30

- (i) Prepare a scatter plot to get an idea about the relationship among the variables.
- (ii) Fit a linear regression model and perform its related analysis at 5% level of significance.
- (iii) Does the fitted regression model satisfy the linearity and normality assumptions? 25

2. (a) 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. 15

- (b) Suppose 20 patients suffering from brain tumor included in a study were assigned radiotherapy. To study the survival pattern, they all were followed upto death. The data so obtained are given as follows:

Patient ID	Survival Time (in months)	Patient ID	Survival Time (in months)
001	34	011	8
002	2	012	26
003	3	013	5
004	2	014	16
005	15	015	2
006	22	016	7
007	28	017	1
008	11	018	18
009	7	019	10
010	5	020	5

Estimate the survival function, CDF PDFG and hazard function. 10

3. A hypothetical data of 24 patients on age, weight and systolic blood pressure (SBP) (1-High SBP and 0-normal SBP) are given as follows:

Age (in years)	Weight (in kg)	SBP
52	60	0
60	68	1
50	54	0
62	74	1
52	62	0
50	67	0
51	66	0
54	65	1
59	71	1
51	87	1
47	49	0
42	52	0
45	50	0
56	83	1
50	62	0
55	94	1

54	87	1
50	56	0
64	70	1
44	60	0
52	85	1
54	45	0
65	75	1
40	68	1

For this data:

- (i) Fit an appropriate regression model.
- (ii) Test the significance of the individual coefficients at 5% level of significance.
- (iii) Determine the Nagelkerke's pseudo R-squared. 25
