MSTL-003 (Set-2)

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 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-Censored)
P01	194	1	1
P02	98	2	1
P03	236	2	1
P04	300	1	0
P05	204	1	1

P06	76	2	0
P07	44	2	0
P08	99	1	1
P09	132	2	0
P10	266	1	1
P11	62	2	0
P12	201	1	1
P13	262	1	1
P14	300	1	1
P15	221	2	1
P16	255	1	1
P17	37	2	0
P18	268	1	1
P19	37	2	0
P20	111	1	0
P21	67	2	1
P22	175	2	0
P23	192	1	0
P24	102	2	1
P25	300	1	0
P26	250	1	0
P27	145	1	1

P28	91	2	1
P29	150	2	1
P30	207	1	0
P31	62	2	1
P32	300	1	0
P33	101	1	0
P34	74	1	1
P35	275	2	0
P36	74	1	1
P37	300	2	0
P38	28	2	1

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

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

(iii) 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. 25

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:

Serum Creatinine	Weight	Gender
0.7	46	1
1.3	65	1
1	59	1
1.5	84	0
1.7	91	1
1.5	78	1
1	53	0
0.7	49	1
0.5	42	0
1.6	87	0
1.1	53	1
0.8	54	0
1.3	65	1
1.1	61	1
1.2	71	0
1.1	55	1
0.9	55	0

0.9	62	0
1.1	65	0
0.8	54	0
0.5	45	0
0.6	45	1
1	62	0
0.6	40	0
1	58	1
0.5	65	1
0.9	67	1
1.3	42	1
1.1	62	0
1.4	65	1

(i) Fit a linear regression model and perform its related analysis at 1% level of significance.

(ii) Check whether the fitted regression model satisfy the linearity and normality assumptions.

3. (a) A study was conducted to check the efficacy of a vaccine on a particular disease. The researcher selected 258 diabetic and 260 non-diabetic patients suffering from that disease for vaccination. The data so obtained are classified by taking 1 if the disease is cured and 0 if not cured and recorded the results as follows:

	Non-diabetic Patient	Diabetic Patient
Cured	135	120
Non-Cured	125	138

Test the hypothesis that the proportion of patients whose disease is cured are different in diabetic and non-diabetic groups of patients, at 5% level of significance. 10

(b) A researcher wants to study the effect of different doses (in mg) of hypertension. He gave different amount of doses to a number of patients and noted the response. The data so obtained are given as follows:

Amount of Dose	Total No. of Patients	Number of Cured Patients
5	60	24
10	48	18
15	40	12
20	80	20
25	104	26

For the given data:

(i) Fit a logistic regression model.

(ii) Determine the variance for the estimates of B_0 and B_1.

(iii) Test the significance of the model coefficients B_0 and B_1 at 5% level of significance.

(iv) Determine the McFadden, Cox and Snell pseudo R-squared. 15

******