No. of Printed Pages : 6

MSTE-004

POST GRADUATE DIPLOMA IN APPLIED STATISTICS (PGDAST)

Term-End Examination

December, 2023

MSTE-004 : BIOSTATISTICS—II

Time : 3 Hours *Maximum Marks : 50*

Note : (*i*) *Question No.* ***1*** *is compulsory.*

(*ii*) *Attempt any* ***four*** *questions from the remaining question nos.* ***2*** *to* ***7****.*

(*iii*) *Use of scientific calculator (non-programmable) is allowed.*

(*iv*) *Use of Formulae and Statistical Table Booklet for PGDAST is allowed.*

(*v*) *Symbols have their usual meanings.*

---

1. State whether the following statements are True *or* False. Give reasons in support of your answers : 5×2=10

   (a) If there are 170 true positive cases out of 200 diseased cases, the sensitivity of the test will be 0.85.

(b) For fitting a multiple linear regression model, the regressor variables must be correlated.

(c) We use method of least squares for fitting the logistic model.

(d) The log-rank test for comparing survival functions is a parametric test.

(e) For a data, if $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 30$ and $\sum_{i=1}^{n}(y_i - \bar{y})^2 = 500$, the coefficient of determination is 0.94 to fit a regression model.

2. The following data were obtained to study the association between low birth weights of neonates and history of obesity in mothers :

Obesity	**Low Birth Weight**	
	Yes	No
Yes	10	02
No	03	05

Test whether there is an association between low birth weight and obesity in mothers at 5% level of significance using Fisher's exact test. 10

3. For a fitted multiple linear regression model of systolic blood pressure (y) on age (x_1) and weight (x_2), the following values are computed :

$n = 15$, $\Sigma y_i = 1881$, $\Sigma x_{1i} = 474$, $\Sigma x_{1i}^2 = 15372$,

$\Sigma x_{2i} = 1102$, $\Sigma x_{2i}^2 = 83140$, $\Sigma y_1^2 = 236403$,

$\Sigma x_{1i}x_{2i} = 35523$, $\Sigma y_i x_{1i} = 59880$, $\Sigma y_i x_{2i} = 139075$,

$\hat{\beta}_0 = 88.1732$, $\hat{\beta}_1 = 0.9266$ and $\hat{\beta}_2 = 0.1082$.

Test the significance of the fitted multiple linear regression model at 5% level of significance.

Also compute R^2 and R^2_{adj}. 10

4. For the fitted multiple logistic model, the following values are obtained :

$N = 100$, $N_1 = 44$, $y_1 = 15$, $y_2 = 18$, $y_3 = 6$,

$y_4 = 4$, $y_5 = 1$, $\hat{\pi}_1 = 0.6126$, $\hat{\pi}_2 = 0.5$,

$\hat{\pi}_3 = 0.3585$, $\hat{\pi}_4 = 0.2494$ and $\hat{\pi}_5 = 0.1649$.

Compute the (i) McFadden R^2, (ii) Cox and Snell R^2 and (iii) Nagelkerke pseudo-R^2. 10

5. A group of patients diagnosed with a disease and divided into two groups randomly to receive either standard or new treatments.

They are followed for 72 days to observed mortality experience. The survival data obtained from these groups are given as follows :

Patient Number	**Survival Time (days)**	**Outcome**	**Treatment**
1	2	Died	2
2	4	Died	2
3	5	Died	2
4	6	Unknown	1
5	9	Died	2
6	9	Unknown	2
7	12	Died	2
8	12	Died	1
9	15	Unknown	2
10	15	Unknown	1
11	22	Died	2
12	30	Died	1
13	37	Died	1
14	55	Died	1
15	72	Survived	1

Test whether there is a significant difference between the survival times of patients under standard and new treatments at 5% level of significance. 10

6. (a) Mention the assumptions of multiple linear regression model. 5

(b) In a small prospective study, ten participants were followed for the development of myocardial infarction (MI) over a period of 10 years. The follow-up data are given as follows :

Participant No.	MI Time
1	5
2	8
3	2
4	6 +
5	9
6	6
7	10 +
8	4
9	8
10	4 +

where, + represents censored observation. Estimate the survival function using the K-M method. 5

7. (a) State the assumptions for applying the Chi-square test for association. 3

(b) Write a short note on Polytomous logistic models with suitable examples. 7