No. of Printed Pages : 6

# MSTE-004

## POST GRADUATE DIPLOMA IN APPLIED STATISTICS (PGDAST)

## Term-End Examination

## December, 2022

### 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 Tables 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) For the following data, the relative risk for the lung cancer among smokers is 2.67 :

| | Lung Cancer | No Lung Cancer |
|---|---|---|
| Smokers | 80 | 1420 |
| Non-smokers | 50 | 2450 |

(b) If the coefficient of determination for a multiple linear regression model with 3 regressor variables fitted on 15 observations is 0.89, then the value of adjusted $R^2$ will be 0.86.

(c) In case of ordinal categories of a response variable, we fit reference category logit models.

(d) Right censoring occurs when we know the exact time of occurrence of an event.

(e) The more weights are assigned to the categories which are distance apart to the diagonal in computing the weighted kappa statistic.

2. In a study, alcohol consumption is considered as exposure, liver disease as outcome and age as confounder. The following table shows an agewise classification of a cohort study :

| Level of Age | Alochol | Liver Disease | No Liver Disease |
|---|---|---|---|
| 1 | Yes | 18 | 247 |
| | No | 5 | 415 |
| 2 | Yes | 22 | 84 |
| | No | 30 | 514 |
| 3 | Yes | 15 | 21 |
| | No | 98 | 378 |

(i) Compute the relative risk of liver disease in alcohol drinkers after adjusting age. Interpret the result. 4

(ii) Construct 99% confidence interval for the relative risk computed in part (i). 6

3. A random sample of 15 women was selected to analyse the relationship of systolic blood pressure (SBP) with age and weight. The data

on SBP (in mm/Hg), age (in years) and weight (in kg) are given in the following table :

| SBP | Age | Weight |
|---|---|---|
| 124 | 30 | 71 |
| 134 | 38 | 82 |
| 135 | 39 | 98 |
| 121 | 26 | 72 |
| 122 | 29 | 70 |
| 119 | 27 | 72 |
| 128 | 32 | 76 |
| 118 | 25 | 54 |
| 120 | 26 | 58 |
| 123 | 31 | 68 |
| 129 | 37 | 63 |
| 117 | 25 | 62 |
| 131 | 35 | 92 |
| 126 | 34 | 75 |
| 134 | 40 | 89 |

(i) Fit a multiple regression model. 8

(ii) Determine the residuals. 2

4. For a fitted simple logistic model, the following values are provided :

$y_1 = 7, y_2 = 4, y_3 = 10, y_4 = 14$

$\pi_1 = 0.2188, \pi_2 = 0.25, \pi_3 = 0.4, \ \pi_4 = 0.4375,$

$\hat{\pi}_1 = 0.2188, \hat{\pi}_2 = 0.2845,$

$\hat{\pi}_3 = 0.3635, \hat{\pi}_4 = 0.4507,$

$n_1 = n_1^1 = 32, n_{2=16}^{1=n_2}, n_{3=25}^{1=n_3}$ and $n_{4=32}^{1=n_4}$

Test the goodness-of-fit of the fitted model at 5% level of significance using :

(i) Model deviance $D_F$ 6

(ii) Hosmer-Lameshow test 4

5. If the survival time T (in years) has the probability density function as : 10

$$f(t) = \begin{cases} \theta e^{-\theta t}; & \theta > 0, t > 0 \\ 0 \quad ; & \text{otherwise} \end{cases}$$

compute :

(i) Survival function

(ii) Cumulative distribution function

(iii) Hazard function

(iv) Median, when $\theta = 0.2$.

6. (a) Describe the Cox proportional hazard model with a suitable example. 4

(b) The following data show the number of families having diabetic patient in different socio-economic groups :

| Diabetic patient | Socio-economic Group | | | |
|---|---|---|---|---|
| | I | II | III | IV |
| Yes | 11 | 27 | 42 | 53 |
| No | 7 | 15 | 16 | 13 |

Test whether the proportions of families having diabetic patients are same in all socio-economic groups at 5% level of significance. 6

7. (a) Explain the residual plot. Also describe various types of residual plots. 6

(b) Differentiate between probit and complementary log-log models. 4