Dear Student,

Please read the information on assignments in the Programme Guide that we have sent youafter your enrolment. A weightage of 30%, as you are aware, has been earmarked for continuousevaluation, **which would consist of one tutor-marked assignment** for this course. The assignmentsfor the theory courses MST-001 to MSTE-004 have been given inthis booklet.

Instructions for Formatting Your Assignments

Before attempting the assignment, please read the following instructions carefully:

1) On top of the first page of your answer sheet, please write the details exactly in the following format:

ENROLLMENT NO :……

NAME :……

ADDRESS :……

……

……

PROGRAMME CODE: ……..

COURSE CODE: …….

COURSE TITLE: ……

STUDY CENTRE: ……..……. DATE: …….……...

PLEASE FOLLOW THE ABOVE FORMAT STRICTLY TO FACILITATE EVALUATION ANDTO AVOID DELAY.

2) Use only foolscap size writing paper (but not of very thin variety) for writing your answers.

3) Leave 4 cm margin on the left, top and bottom of your answer sheet.

4) Your answers should be precise.

5) This assignment is to be submitted at the Study Centre.

   We strongly suggest that you should retain a copy of your answer sheets.

6) This assignment is valid from January 1st, 2025up toDecember 31, 2025.

7) The latest assignments should be submitted by the candidate.

8) **You cannot fill the Exam Form for this course** till you have submitted this assignment. So solve itand **submit it to your study centre at the earliest.**If you wish to appear in the **TEE, June 2025,** you should submit your TMAs by**March 31, 2025.** Similarly, If you wish to appear in the **TEE, December 2025,** you should submit your TMAs by**September 30, 2025.**

We wish you good luck.

TUTOR MARKED ASSIGNMENT

MSTE-004: Biostatistics-II

Course Code: MSTE-004 **Assignment Code: MSTE-004/TMA/2022**

Maximum Marks: 100

Note: All questions are compulsory. Answer in your own words.

1. State whether the following statements are **True** or **False**. Give reason in support of your answer: **(2×5=10)**

(a) The value of sensitivity of the following results of a diagnostic test is 0.85.

Disease	**Test result**		**Total**
	+	–	
Present	170	30	200
Absent	20	280	300

(b) For the following cohort study, the relative risk for the lung cancer among smokers is 3.5.

	Lung Cancer	**No Lung Cancer**	**Total**
Smokers	100	1220	1320
Non-smokers	50	2260	2310

(c) The logit link function is $\log\left[-\log(1-\pi)\right]$.

(d) We define three indicator/dummy variables for a regressor variable with three categories.

(e) Left censoring occurs whenever the exact time of occurrence of an event is not known.

2. Differentiate between Chi-square tests for association and homogeneity of proportions. Also mention the assumptions of these tests. **(10)**

3. A random sample of 250 patients was selected and their workout timing and diabetes status were recorded. The following table shows the workout timing and severity of diabetes:

Workout (in minutes)	**Severity of diabetes status**		
	Low	**Moderate**	**High**
0 −15	06	27	19
15 to 30	08	36	17
30 to 45	21	45	33
≥ 45	14	18	06

Test at 5% level of significance whether workout habit and diabetes are associated with to each other or not. **(10)**

4. **(a)** Explain the assumptions underlying multiple linear regression model.

(b) Suppose a researcher wants to evaluate the effect of cholesterol on the blood pressure. The following data on serum cholesterol (in mg/dL) and systolic blood pressure (in

mm/Hg) were obtained for 15 patients to explore the relationship between cholesterol and blood pressure:

S. No.	Cholesterol (mg/dL)	SBP (mm/Hg)
1	300	150
2	410	270
3	380	210
4	530	310
5	570	350
6	490	310
7	340	210
8	320	150
9	280	110
10	550	320
11	340	220
12	350	170
13	410	260
14	390	230
15	450	270

(i) Fit a linear regression model using the method of least squares.

(ii) Construct the normal probability plot for the data on serum cholesterol and systolic blood pressure.

(iii) Test the significance of the fitted regression model. **(5+15)**

5. Write a short note on the following:

(i) Polytomous logistic models

(ii) Poisson regression

(iii) Kaplan and Meier method **(12)**

6. The following data on diagnosis of coronary heart disease (where 0 indicating absence and 1 indicating presence), serum cholesterol (in mg/dl), resting blood pressure (in mmHg) and weight (in kg) were obtained for 80 patients to explore the relationship of coronary heart disease with cholesterol and weight.

S. No.	Serum Cholesterol (mg/dl)	Weight (kg)	Number of Patients having CHD	Total Number of Patients
1	420	60	10	20
2	450	68	15	30
3	400	54	4	15

4	510	74	2	10
5	480	62	1	5

(i) Fit a multiple logistic model for the dependence of coronary heart disease on the average serum cholesterol and weight considering $\hat{\beta}_0^0 = 4.279$, $\hat{\beta}_1^0 = -0.035$ and $\hat{\beta}_2^0 = 0.172$ as the initial values of the parameters (solve only for one Iteration).

(ii) Test the significance of the fitted model using Hosmer-Lemeshow test at 5% level of significance. **(12+8)**

7.(a) Describe censoring and differentiate between different types of censoring with the help of examples which are not considered in Block 4 of MSTE-004.

(b) A study was conducted on 185 patients aged more than 45 years which are followed until the time of death or up to 10 years, whichever comes first. The patients have different covariates: age, gender (male/female), systolic blood pressure, smoking (yes/no), total serum cholesterol and diabetes (yes/no). The objective of this study is to determine which covariate influences the survival time. An analysis is conducted to investigate differences in all-cause mortality between men and women participating in the study. Suppose we obtain the following results after applying the Cox regression hazard model analyses:

Risk Factor	**Parameter Estimate**	**SE**
Age	0.150	0.010
Gender	0.450	0.150
Systolic Blood Pressure	0.015	0.008
Smoking	0.650	0.170
Total Serum Cholesterol	0.002	0.004
Diabetes	-0.350	0.250

(i) Obtain hazard ratio and interpret the results.

(ii) Find the 99% confidence interval for the hazard ratio.

(iii) Test whether the covariates are significant or not at 1% level of significance. **(8+10)**