



Roll No:

| | | | | | | | | | | | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| | | | | | | | | | | | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|

BTECH
(SEM V) THEORY EXAMINATION 2023-24
STATISTICAL COMPUTING

TIME: 3 HRS**M.MARKS: 100****Note: 1.** Attempt all Sections. If require any missing data; then choose suitably.**SECTION A****1. Attempt all questions in brief.**

| Q no. | Question | Marks |
|-------|--|-------|
| a. | Discuss the significance of measures of dispersion in a dataset. Provide examples of two measures of dispersion and explain how they differ. | 2 |
| b. | Define the concept of "mean" as a measure of central tendency. How is it calculated, and what are its advantages and limitations? | 2 |
| c. | Explain the concept of correlation and its significance in statistical analysis. Provide an example of a positive and a negative correlation. | 2 |
| d. | Outline the steps involved in conducting an inference procedure for a correlation coefficient. Why is it important in statistical analysis? | 2 |
| e. | Differentiate between bivariate correlation and simple correlation. Provide an example scenario where bivariate correlation would be more appropriate. | 2 |
| f. | Explain the concept of linear regression. How is the regression line determined, and what does the slope of the line represent in a regression analysis? | 2 |
| g. | Describe the key differences between simple linear regression and multiple regression. Why might a researcher choose to use multiple regression? | 2 |
| h. | Compute the correlation coefficient for the following paired data: X (hours of study) - 10, 15, 20, 25; Y (exam scores) - 60, 75, 80, 90. | 2 |
| i. | Given the dataset: X - 2, 4, 6, 8; Y - 5, 8, 11, 14, find the equation of the regression line (Y on X). | 2 |
| j. | In a deck of 52 cards, what is the probability of drawing a queen given that the card drawn is a face card? | 2 |

SECTION B**2. Attempt any three of the following:**

| | | |
|----|---|----|
| a. | Apply singular value decomposition (SVD) to a given matrix B. Discuss the significance of the singular values and how they contribute to linear dimension reduction. | 10 |
| b. | Conduct a multiple regression analysis using a dataset with two independent variables. Interpret the coefficients and assess the overall fit of the model. | 10 |
| c. | In a study comparing two teaching methods, Group A has 15 students taught with Method 1, and Group B has 20 students taught with Method 2. The exam scores for both groups are given. Use a randomization test to determine if there is a significant difference in the mean scores between the two groups. | 10 |
| d. | (i) Define principal component analysis (PCA) and its purpose in linear dimension reduction. Discuss the steps involved in PCA. (ii) Apply PCA to a dataset with five variables. Interpret the results and explain how PCA aids in dimensionality reduction. | 10 |
| e. | A random sample of 50 students is selected from a population. The average hours of study per week for the sample is 15 hours, with a standard deviation of 3 hours. Calculate a 95% confidence interval for the mean hours of study for the population. | 10 |



Roll No:

| | | | | | | | | | | | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| | | | | | | | | | | | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|

BTECH
(SEM V) THEORY EXAMINATION 2023-24
STATISTICAL COMPUTING

TIME: 3 HRS

M.MARKS: 100

SECTION C

3. Attempt any *one* part of the following:

| | | |
|----|---|----|
| a. | Perform a Monte Carlo simulation to test the hypothesis that the mean of a population is 50, based on a sample of 30 observations with a known standard deviation of 10. Use 1000 simulated samples and a significance level of 0.05. | 10 |
| b. | Define Markov chains and explain their role in Markov Chain Monte Carlo (McMC) methods. Discuss the importance of convergence in McMC simulations. | 10 |

4. Attempt any *one* part of the following:

| | | |
|----|--|----|
| a. | Explain how Monte Carlo simulations can be used for hypothesis testing. Discuss the advantages of Monte Carlo hypothesis testing over traditional methods. | 10 |
| b. | Explain the concept of the jackknife resampling method. Apply the jackknife to estimate the variance of the mean for a dataset with 30 observations. | 10 |

5. Attempt any *one* part of the following:

| | | |
|----|--|----|
| a. | Perform a permutation test to compare the means of two independent groups (Group A: 25 observations, Group B: 30 observations). The observed difference in means is 2.5, and the permutation distribution yields differences of -1.8, -0.5, 1.0, ... | 10 |
| b. | Use the jackknife resampling method to estimate the bias and variance of the mean for a dataset with 40 observations. Present the jackknife-estimated mean, bias, and variance. | 10 |

6. Attempt any *one* part of the following:

| | | |
|----|---|----|
| a. | Perform 5-fold cross-validation on a linear regression model. The dataset has 80 observations, and the mean squared error for each fold is as follows: 12, 15, 10, 18, 14. Compute the average mean squared error across the folds. | 10 |
| b. | Briefly discuss the history and origin of the R programming language. Highlight key milestones and contributors. | 10 |

7. Attempt any *one* part of the following:

| | | |
|----|---|----|
| a. | Save a vector of numbers (e.g., c(2, 4, 6, 8, 10)) to the R workspace. Show how to inspect the variables in the workspace and display their values. | 10 |
| b. | Define a vector with the elements 1 to 5. Create a matrix with 3 rows and 2 columns using the vector. Display both the vector and the matrix. | 10 |