

- Instructions :** (1) This paper contains **FIVE** questions.  
(2) All questions are compulsory.  
(3) Question No. **2, 3, 4** have internal options.  
(4) Figures in the right side in parenthesis indicate marks.

**Q:1** Data constraints significantly impact the scope and accuracy of analysis -Explain the statement in light of potential of data analysis and importance of data management practices, including careful data collection, preprocessing, and validation, to mitigate the limitations on analysis. (14)

**Q:2** How would you apply multivariate data analysis techniques to a real-world dataset? Describe the objectives of your analysis and the assumptions you would need to consider. (14)

**OR**

**Q:2** Provide a comprehensive overview of Multivariate Data Analysis techniques. For each technique, explain: (14)

- Its fundamental principles
- Its suitability for different types of data
- Real-world examples of its application
- How it contributes to reducing familywise error rate in statistical inference

**Q:3** Missing data are a nuisance and primarily result from errors in data collection – Discuss challenges, types of missing data and its impact (14)

**OR**

**Q:3** Explain MANOVA and its purpose in statistical analysis. Also answer following questions. (14)

- What is Wilks' Lambda in MANOVA, and how is it interpreted?
- Describe Hotelling's  $T^2$ , How does it differ from Wilks' Lambda?
- What is Pillai's Trace, and why is it considered robust in MANOVA?

Explain Roy's Largest Root in context of MANOVA and when it is most useful.

**Q:4** What are KMO, communalities, the anti-image correlation matrix, eigenvalues, the scree plot, and the rotated component matrix? Which multivariate technique uses these measures? (14)

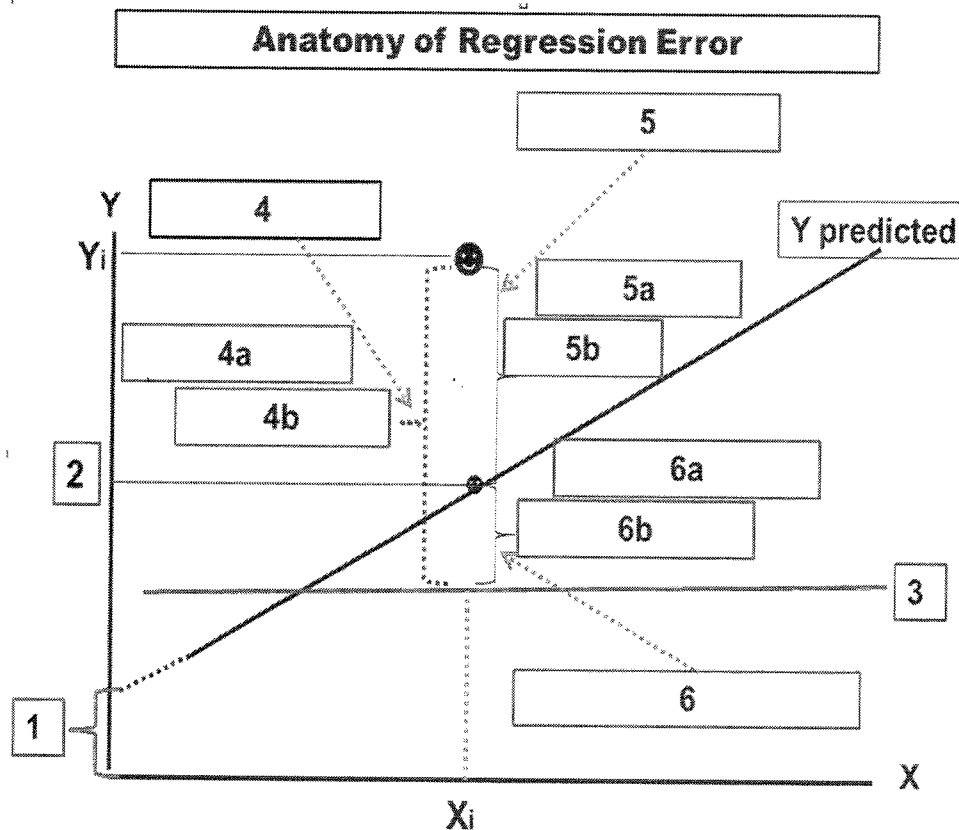
OR

**Q:4** You are provided with the standardized canonical discriminant function coefficients (14) as follows:

- X1: 0.6
- X2: 0.8

If a new observation has  $X_1=6$  and  $X_2=5$ , calculate the discriminant score for this observation

**Q:5** The diagram below illustrates the concept of regression error. Please identify the correct terms for each box labeled with the number. (14)



Fill the blanks based on identified correct term

1. The vertical axis represents \_\_\_\_\_ of the response variable.
2. The horizontal axis represents the \_\_\_\_\_ variable.
3. The actual observed value at  $X_i$  is represented by the term \_\_\_\_\_.

4. Predicted value on regression line at  $X_i$  is represented by the term \_\_\_\_\_.
5. The line that shows the trend of predicted values is called the \_\_\_\_\_.
6. Difference between observed value  $Y_i$  & predicted value is called \_\_\_\_\_.
7. Write equation in the box labelled with 4b, in relation with for the term indicated by box labeled with 4a. \_\_\_\_\_ (use appropriate notations)
8. Write equation in the box labelled with 5b, in relation with for the term indicated by box labeled with 5a. \_\_\_\_\_ (use appropriate notations)
9. Write equation in the box labelled with 6b, in relation with for the term indicated by box labeled with 6a. \_\_\_\_\_ (use appropriate notations)
10. Write appropriate term for box labelled with 1. \_\_\_\_\_

---