

**IM.Sc. DS Sem.-8 Examination  
CC-411**

**Multivariate Analysis**

**April-2025**

**Time : 2-30 Hours]**

**[Max. Marks : 70**

**Instructions:** All questions are compulsory.

- Q.1 (a)** A company wants to analyze whether customer satisfaction is related to the type of product they purchased. A survey was conducted, and customers were asked to rate their satisfaction level (Satisfied, Neutral, Dissatisfied) based on the product category (Electronics, Clothing, Groceries). (07)

	Satisfied	Neutral	Dissatisfied
<b>Electronics</b>	30	15	10
<b>Clothing</b>	25	20	15
<b>Groceries</b>	20	25	30

Perform the Chi Square test with 5% level of significance and check satisfaction level is independent of product category. Chi Square Table value is 9.488

- (b)** Identify and explain the key statistics used in factor analysis, including the KMO measure, Bartlett's test, and communalities. (07)

**OR**

- (a)** Find Pearson's Correlation Coefficient Between AI/ML Model Accuracy and Data Size (07)

<b>Observation</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>
<b>Data Size (X)</b>	5.8	10.7	15.9	20.5	25.6	30.4
<b>Accuracy (Y)</b>	65.4	70.1	75.3	78.8	80.4	83.2
<b>Observation</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>
<b>Data Size (X)</b>	36	40.2	45.9	50.5	55.6	60.6
<b>Accuracy (Y)</b>	86	87.4	88.7	90.8	91.5	93.8

- (b)** Explain the differences between descriptive and inferential statistics and discuss how they complement each other in data analysis. (07)

- Q.2 (a)** What are Structural Coefficients, and how are they interpreted in Discriminant Analysis? (07)

- (b)** A bank wants to predict whether a customer will default (1) or not (0) on a loan based on their income and credit score. (07)

Customer	Income (\$1000s)	Credit Score	Default (1=Yes, 0=No)
1	50	700	0
2	40	650	0
3	35	600	1
4	60	720	0
5	30	580	1

Intercept	Income (\$1000s)	Credit Score
11.2254	0.0676	-0.0211

predict whether a customer with Income = 32 and Credit Score = 570 will default, based on the given logistic regression model. Probability of cutoff value is 0.70

OR

- (a) Explain the assumptions and limitations of Logit Analysis in classification problems (07)
- (b) A University wants to classify students into two groups: **Pass** or **Fail**, based on their scores in **Math** and **English**. (07)

Student	Math Score ( $X_1$ )	English Score ( $X_2$ )	Class (Y)
A	85	78	Pass
B	72	65	Pass
C	90	80	Pass
D	60	55	Fail
E	45	50	Fail
F	50	48	Fail

Classifying Students Based on Test Scores using Discriminant Analysis.

Below is coefficient for Discriminant Analysis

Intercept (c)	Math Score ( $X_1$ ) ( $w_1$ )	English Score ( $X_2$ ) ( $w_2$ )
-1.6090	0.0097	0.0233

- Q.3 (a) What statistical measures are commonly used to evaluate the quality of clusters? (07)
- (b) What are the key assumptions in cluster analysis, and how can they be stabilized to ensure meaningful results? (07)

OR

- (a) Explain the role of distance metrics (e.g., Euclidean, Manhattan) in cluster analysis (07)
- (b) Compare and contrast hierarchical clustering, k-means clustering, and fuzzy clustering (07)
- Q.4 (a) How is preference data used in MDS to analyze consumer behavior? (07)
- (b) What are the common sources of similarity or dissimilarity data in MDS applications? (07)
- OR**
- (a) Explain the concept of stress function in MDS and its role in measuring the goodness of fit. (07)
- (b) How is Conjoint Analysis applied in marketing research to understand consumer preferences? (07)
- Q.5 Attempt any **SEVEN** out of **FOURTEEN (True or False)** (14)
- (1) Conjoint analysis is primarily used for clustering similar objects
- (2) K-means clustering is a type of hierarchical clustering
- (3) Discriminant analysis and ANOVA are unrelated statistical techniques
- (4) Conjoint analysis is useful for understanding how consumers value different features of a product
- (5) Data transformation is required only in factor analysis, not in bivariate analysis
- (6) Discriminant analysis can be used for classifying observations into predefined groups.
- (7) Spearman's Rank Order Correlation Coefficient is used for measuring linear relationships between numerical variables.
- (8) Multidimensional scaling (MDS) is used to visualize similarity or dissimilarity data in a lower-dimensional space
- (9) Factor rotation helps in simplifying the interpretation of factors in factor analysis
- (10) Binary Logit Analysis cannot be used when the dependent variable has only two categories.
- (11) Model-based clustering assumes that data is generated from a probabilistic model
- (12) The standardized discriminant function coefficients represent the relative importance of each predictor variable
- (13) Multidimensional scaling can only be used with numerical data
- (14) Hierarchical clustering always requires defining the number of clusters in advance.
-