|  |  |
| --- | --- |
| **SESSION** | **FEB-MARCH 2025** |
| **PROGRAM** | **MASTER OF BUSINESS ADMINISTRATION (MBA)** |
| **SEMESTER** | **III** |
| **COURSE CODE & NAME** | **DADS302 EXPLORATORY DATA ANALYSIS** |
|  |  |
|  |  |

**Assignment Set – 1**

**Q1. Explain various measures of dispersion in detail using specific examples. 10**

**Ans 1.**

**Measures of Dispersion**

Measures of dispersion refer to statistical techniques used to describe the spread or variability within a data set. While measures of central tendency like mean and median give a single value summary, dispersion measures how much data points differ from this central value. Understanding dispersion is essential in statistical analysis, particularly in evaluating risk, consistency, and variability of datasets.

**Range as a Basic Measure**

The range is the simplest measure of dispersion, calculated by subtracting the smallest value from the largest. For example, in a dataset of test scores: 45, 50, 60, 70, 80, the range is 80 –

Its Half solved only

Buy Complete assignment from us

**Price – 190/ assignment**

**MUJ Manipal University Complete SolvedAssignments MARCH 2025**

buy cheap assignment help online from us easily

we are here to help you with the best and cheap help

**Contact No – 8791514139 (WhatsApp)**

**OR**

**Mail us-** bestassignment247@gmail.com

**Our website -** [www.assignmentsupport.in](http://www.assignmentsupport.in)

**Q2. What is Data Science? Discuss the role of Data Science in various Domains.2 + 8**

#### Ans 2.

Data Science is an interdisciplinary field that combines scientific methods, statistical analysis, algorithms, and machine learning techniques to extract meaningful insights from structured and unstructured data. It is often referred to as the future of decision-making, as it empowers organizations to make data-driven decisions rather than relying on intuition. At its core, Data Science integrates knowledge from mathematics, statistics, computer science, and domain-specific expertise to uncover patterns and solve complex problems.

The lifecycle of a Data Science project typically includes data collection, data cleaning,

**Q3. Discuss various techniques used for Data Visualization. 10**

#### Ans 3.

#### ****Importance of Data Visualization****

Data visualization is the graphical representation of information and data using visual elements like charts, graphs, and maps. It is a core component of exploratory data analysis, as it allows users to identify trends, patterns, and outliers quickly and intuitively. Visualization helps simplify complex datasets by transforming them into an easily understandable visual format. It is widely used in business intelligence, reporting dashboards, academic research,

**Assignment Set – 2**

**Q4. What is feature selection? Discuss any two feature selection techniques used to get optimal feature combinations. 2+4+4**

**Ans 4.**

**Feature Selection**

Feature selection is a critical step in the data preprocessing phase of machine learning and statistical modeling. It involves selecting a subset of the most relevant features (variables or predictors) from the original dataset that contribute the most to the prediction or classification outcome. The main goal is to improve model performance by eliminating irrelevant or redundant features, thereby reducing complexity, overfitting, and training time.

Feature selection differs from dimensionality reduction, as it selects existing features without transforming them. It enhances model interpretability by retaining the most significant

**Q5. Discuss in detail the concept of Factor Analysis**

**Ans 5.**

**Factor Analysis**

Factor analysis is a statistical technique used to identify underlying relationships among a large set of observed variables. The main goal is to reduce data complexity by grouping related variables into latent factors that represent common themes or constructs. These hidden factors help explain the patterns of correlations within the dataset. Factor analysis is often used in social sciences, psychology, market research, and behavioral studies where abstract

**Q6. Differentiate between Principal Component Analysis and and Linear Discriminant Analysis 10**

**Ans 6.**

**Dimensionality Reduction Techniques**

Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are two widely used dimensionality reduction techniques in machine learning and statistics. Although they serve the common purpose of reducing the number of features in a dataset, they differ significantly in their objectives, methodology, and applications. Understanding these differences is crucial for selecting the appropriate technique for a given problem.

Dimensionality reduction is often necessary in datasets with high numbers of features, as it helps reduce noise, prevent overfitting, improve model performance, and simplify