# BRAINWARE UNIVERSITY

### Term End Examination 2024-2025
### Programme – MBA-2023
### Course Name – Data Management and Analytics
### Course Code - SM401
### ( Semester IV )

**Full Marks : 60**                                                      **Time : 2:30 Hours**

[The figure in the margin indicates full marks. Candidates are required to give their answers in their own words as far as practicable.]

### Group-A
(Multiple Choice Type Question)                                    1 x 15=15

1.   *Choose the correct alternative from the following :*

(i)   An e-commerce website wants to reduce cart abandonment rates. Which analytics approach should they employ to achieve this?

a) Descriptive analytics to analyze past abandonment instances

b) Predictive analytics to anticipate potential abandonment

c) Diagnostic analytics to identify reasons behind abandonment

d) Prescriptive analytics to recommend actions to prevent abandonmen

(ii)  Select the type of analytics that focuses on analyzing past data to understand historical trends and patterns.

a) Predictive analytics

b) Diagnostic analytics

c) Descriptive analytics

d) Prescriptive analytics

(iii) Select why is it important to critically assess the validity of data analysis results before making business decisions.

a) To ensure alignment with company values

b) To avoid legal repercussions

c) To mitigate risks associated with decision-making

d) To reduce reliance on data analytics tools

(iv)  Choose the Davenport's article, what are some potential challenges for companies aiming to compete on analytics?

a) Limited availability of data

b) High costs associated with analytics implementation

c) Resistance to change within the organization

d) Lack of skilled personnel in analytics

(v)   If a company wants to emulate the success stories highlighted in the article, what should be their primary focus?

a) Investing in outdated technology

b) Relying solely on intuition

c) Building a robust analytics capability

d) Ignoring market trends

(vi)  Select the following statements that are true regarding the mean and median.

a) The mean is always greater than the median
b) The median is always greater than the mean
c) The mean and median are equal in a symmetric distribution
d) The mean and median are not related to each other

(vii) Select how does the interquartile range (IQR) is represented?

a) The difference between the highest and lowest values
b) The range of the middle 50% of the data
c) The average deviation from the mean
d) The spread of data around the median

(viii) Calculate the variance for the following data set: 5, 10, 15, 20, 25.

a) 62.5
b) 75
c) 100
d) 125

(ix) Select the following statements that are true regarding the relationship between mean, median, and mode in a symmetric distribution.

a) Mean = Median = Mode
b) Mean > Median > Mode
c) Mean < Median < Mode
d) Mean = Median, Mode may or may not be equal to them

(x) What does the sd() function in R compute?

a) Standard deviation
b) Range
c) Median
d) Interquartile range

(xi) Compare and contrast the summary() function and the describe() function in R for summarizing a dataset.

a) The summary() function provides basic statistics, while the describe() function provides more detailed information
b) The describe() function provides basic statistics, while the summary() function provides more detailed information
c) Both functions provide the same summary statistics but differ in syntax
d) Neither function is available in R

(xii) Develop an R script to generate a scatter plot of two variables, "x" and "y" stored in separate vectors.

a) plot(x, y)
b) scatterplot(x, y)
c) scatter(x, y)
d) scatter_plot(x, y)

(xiii) Select the statistical test is used to determine if there is a significant difference between the means of two independent groups.

a) Chi-Square test
b) t-Test
c) ANOVA
d) Correlation Analysis

(xiv) Which R function is used to perform a one-way ANOVA?

a) chisq.test()
b) t.test()
c) anova()
d) cor.test()

(xv) Which statistical test is used to compare observed frequencies with expected frequencies in categorical data?

a) t-Tes
b) ANOVA
c) Chi-Square test
d) Correlation Analysis

## Group-B
### (Short Answer Type Questions)    3 x 5=15

2. A researcher wants to determine if there is a relationship between gender (male, female) and smoking status (smoker, non-smoker). Predict the type of Chi-Square test should the researcher use, and why?    (3)
3. Explain businesses apply predictive analytics to solve operational challenges.    (3)
4. Analyze the potential risks and challenges associated with competing on analytics, as discussed in Davenports article.    (3)
5. Describe the data wrangling.    (3)

6. You have a dataset containing features such as age, income, and education level, and the     (3)
   target variable is loan approval (yes/no). How would you apply a classification technique to
   predict loan approval based on these features?

**OR**

Summarize the procedure of utilizing data analytics to shape pivotal business decisions.     (3)

## Group-C
### (Long Answer Type Questions)                                          5 x 6=30

7. Compare the implications of a high standard deviation for datasets with different     (5)
   distributions, such as normal, skewed, and uniform distributions.
8. Examine a Chi-Square test is conducted on a contingency table with 3 rows and 4 columns.  (5)
   Calculate the degrees of freedom for this test.
9. You have a dataset containing students' number of hours studied (independent variable)     (5)
   and exam scores (dependent variable). How would you use regression analysis to predict
   exam scores based on the hours studied?
10. You have transaction data from a retail store containing information about items purchased (5)
    together. How would you apply association rules analysis to identify frequent item sets and
    generate rules?
11. Explain a sensitivity analysis plan for a supply chain optimization model, considering factors (5)
    such as transportation costs, lead times, and demand variability.
12. Explain a hypothesis test to determine whether there is a significant difference in mean     (5)
    exam scores between two groups of students (e.g., group A and group B).

**OR**

Contrast in situations with small expected frequencies, the Chi-Square test may not be     (5)
appropriate due to its reliance on large sample sizes. In such cases, alternative tests or
methods, such as Fisher's exact test, should be considered.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*