Name: Enrolment No:	

UPES

End Semester Examination, May 2025

Course: Business Statistics **Semester: II**

Program: BBA/B.COM/INT-BBA-MBA/INT-B.COM-MBA **Time: 03 hrs.**

Course Code: DSQT 1004 **Max. Marks: 100**

Instructions:

1. **All Questions are Compulsory.**
2. **Simple/Scientific both calculators are allowed.**

SECTION A

10Qx2M=20Marks

S. No.		Marks	CO
Q 1	**Multiple choice questions**		
(i)	In a simple linear regression model, which of the following best describes the slope coefficient (β_1)? A. The value of Y when X is zero B. The change in Y for a one-unit change in X C. The average of the dependent variable D. The standard error of the regression line	**2**	**CO1**
(ii)	Which of the following best describes the main use of a scatter diagram in marketing data analysis? A. To determine the strength and direction of a linear relationship between two variables B. To visually assess the possible pattern or trend between two numeric variables C. To test the statistical significance of the correlation coefficient D. To identify the dependent and independent variables for regression analysis	**2**	**CO1**
(iii)	What does the height of a bar represent in a histogram? a) Class interval b) Frequency density c) Frequency d) Cumulative frequency	**2**	**CO1**
(iv)	Which visualization tool is best for showing cumulative frequencies? a) Histogram b) Pie chart c) Ogive curve d) Scatter plot	**2**	**CO1**
(v)	What is the probability of selecting a vowel from the word **"PROBABILITY"**?	**2**	**CO1**

	A) 4/11 B) 3/11 C) 2/11 D) 5/11		
(vi)	The probability of an impossible event is: A) 1 B) 0.5 C) 0 D) Undefined	**2**	**CO1**
(vii)	In a perfectly symmetric distribution, which of the following is true? A) Mean > Median > Mode B) Mean < Median < Mode C) Mean = Median = Mode D) Median = Mode > Mean	**2**	**CO1**
(viii)	What is the range of data sets? (a) The difference between the highest and lowest values in the data set. (b) The most frequently occurring value in the data set. (c) The middle value of the data set. (d) The average value of the data se	**2**	**CO1**
(ix)	A company's HR department analyzed the ages of its employees and found that the mean age is 30 years, and the median age is 28 years. Determine the mode of the age distribution. What does the result suggest about the skewness of the data? a) Mode is 24 years, indicating a positively skewed distribution. b) Mode is 26 years, indicating a negatively skewed distribution. c) Mode is 28 years, indicating a perfectly symmetrical distribution. d) Mode is 30 years, indicating a uniform distribution.	**2**	**CO1**
(x)	The degree of peaked ness or flatness of a distribution is called: (a) Skewness (b) Symmetry (c) Dispersion (d) Kurtosis	**2**	**CO1**
SECTION B **4Qx5M= 20 Marks**			
Q 2.			
(i)	What role do skewness and kurtosis play in analyzing business metrics, and what insights do they offer about the shape of data?	**5**	**CO2**
(ii)	In a supply chain setting, products are transported across several legs of a route at different speeds. To estimate the average speed over the entire trip, which type of mean is most suitable, and why?	**5**	**CO2**
(iii)	What is the purpose of a scatter diagram in marketing analysis? How can it help in understanding the relationship between customer satisfaction (X) and customer loyalty (Y)? Sketch a rough diagram showing a positive linear relationship and explain your interpretation.	**5**	**CO2**

<table>
<tr><td>(iv)</td><td>Differentiate between exclusive and inclusive class intervals by providing examples for each</td><td>5</td><td>CO2</td></tr>
<tr><td colspan="4">SECTION-C
3Qx10M=30 Marks</td></tr>
<tr><td>Q 3.</td><td></td><td></td><td></td></tr>
<tr><td>(i)</td><td>Two dice are rolled. Find the probability of getting
(a) The sum is at least 10
(b) The sum is an even number
(c) The sum is prime number
(d) Same number on dice
(e) A total sum of 7</td><td>10</td><td>CO3</td></tr>
<tr><td>(ii)</td><td>The following distribution gives the pattern of overtime work per week done by 120 employees of a company. Calculate median, mode, and quartile deviation of following data.

<table><tr><td>Overtime hours</td><td>10-15</td><td>15-20</td><td>20-25</td><td>25-30</td><td>30-35</td><td>35-40</td></tr><tr><td>No. of employees</td><td>20</td><td>22</td><td>45</td><td>18</td><td>5</td><td>10</td></tr></table></td><td>10</td><td>CO3</td></tr>
<tr><td>(iii)</td><td>Explain the idea of simple linear regression. Why is it called "simple" and "linear"? Given the following data on the weekly social media advertising budget (X in ₹'000) and weekly website traffic (Y in visits) for five weeks, fit a simple linear regression model to find the equation of the line. Then, interpret the equation. What will be the expected website traffic if social media advertising budget is 12.

<table><tr><td>Week</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr><tr><td>Social Media Advertising Budget (X)</td><td>5</td><td>10</td><td>15</td><td>20</td><td>25</td></tr><tr><td>Website Traffic (Y)</td><td>2000</td><td>2500</td><td>3000</td><td>3500</td><td>4000</td></tr></table></td><td>10</td><td>CO3</td></tr>
<tr><td colspan="4">SECTION-D
2Qx15M= 30 Marks</td></tr>
<tr><td>Q 4.</td><td>Attempt any two.</td><td></td><td></td></tr>
<tr><td>(i)</td><td>The number of support tickets closed each week by three agents over six weeks is shown:
Agent A: 34, 36, 35, 33, 35, 34
Agent B: 32, 38, 34, 36, 33, 35
Agent C: 35, 34, 35, 35, 34, 35</td><td>15</td><td>CO4</td></tr>
</table>

<table>
<tr><td></td><td>Based on this information, which agent demonstrated the most consistent performance?</td><td></td><td></td></tr>
<tr><td>(ii)</td><td>Discuss the concept of Karl Pearson's Correlation and Spearman's Rank Correlation. Why are both used in statistical analysis? When would you prefer to use one over the other?
A small online retail company wants to analyze the relationship between Customer Satisfaction and Product Quality. The company has collected data from 5 customers, and each customer was ranked by two judges on Customer Satisfaction and Product Quality. The ranks were given based on their feedback and product reviews. The company is interested in understanding whether there is a correlation between Customer Satisfaction and Product Quality.

<table>
<tr><th>Customer</th><th>Rank of Customer Satisfaction</th><th>Rank of Product quality</th></tr>
<tr><td>A</td><td>1</td><td>3</td></tr>
<tr><td>B</td><td>2</td><td>1</td></tr>
<tr><td>C</td><td>3</td><td>4</td></tr>
<tr><td>D</td><td>4</td><td>2</td></tr>
<tr><td>E</td><td>5</td><td>5</td></tr>
</table>

Suggest and calculate appropriate measure of correlation coefficient, based on your calculation what will you conclude?</td><td>15</td><td>CO4</td></tr>
<tr><td>(iii)</td><td>A supermarket manager recorded how long (in minutes) customers spent shopping.

<table>
<tr><th>Time Spent (minutes)</th><td>0-10</td><td>10-20</td><td>20-30</td><td>30-40</td><td>40-50</td></tr>
<tr><th>Number of Customers</th><td>5</td><td>8</td><td>12</td><td>7</td><td>3</td></tr>
</table>

(i) Draw the ogive curve.
(ii) Estimate the median using graph.
(iii) Verify using the formula for the median</td><td>15</td><td>CO4</td></tr>
</table>