Graphic Era HILL UNIVERSITY
Established by an Act of the State Legislature of Uttarakhand (Achiriyam Sankhya 12 of 2011)
University under section 2(f) of UGC Act, 1956

End Term (Even) Semester Examination May-June 2025

Roll no. 2492113

Name of the Program and semester: **BCA/BCA AI & DS, 2nd Semester**
Name of the Course: **Probability and Statistics for Data Science**
Course Code: **TBC-204 /TBD-204**

Time: 3 hour Maximum Marks: 100

Note:
(i) All the questions are compulsory.
(ii) Answer any two sub questions from a, b and c in each main question.
(iii) Total marks for each question is 20 (twenty).
(iv) Each sub-question carries 10 marks.

Q1. (2X10=20 Marks)

a. What does 'understanding the business problem' mean in data science?
b. Define Exploratory Data Analysis (EDA) with suitable examples.
c. What are the different data types, and why are they important?

Q2. (2X10=20 Marks)

a. If the probability that an individual suffers a bad reaction from a certain injection is 0.001, determine that out of 2000 individuals,
   (i) Exactly three individuals
   (ii) More than two individuals
   (iii) More than one individual will suffer a bad reaction.
b. Define normal distribution.
c. Define Baye's theorem. In a bolt factory, machines A, B, and C manufacture, respectively, 25%, 35%, and 40% of the total. Of their output, 5, 4, and 2 percent are defective bolts. A bolt is drawn at random from the product and is found to be defective. What is the probability that it was manufactured by machine B?

Q3. (2X10=20 Marks)

a. Find the mean deviation for the following frequency distribution.

Class	0-6	6-12	12-18	18-24	24-30
Frequency	8	10	12	9	5

b. Calculate μ_1, μ_2, μ_3, μ_4 for the following frequency distribution:

x	0	1	2	3	4	5	6	7	8
f	1	8	28	56	70	56	28	8	1

c. Compute the Karl Pearson coefficient of skewness for the given data.

x	6	7	8	9	10	11	12
f	3	6	9	13	8	5	4

Q4. (2X10=20 Marks)

a. Define the following terms.
   (i) Type I, Type II Error (ii) p-value

End Term (Even) Semester Examination May-June 2025

b. The mean lifetime of a sample of 100 fluorescent light bulbs produced by a company is computed to be 1570 hours with a standard deviation of 120 hours. The company claims that the average life of the bulb produced by it is 1600 hours. Using the level of significance of 0.05 is 1.96, is the claim acceptable?

c. In one sample of 8 observations, the sum of the squares of the deviations of the sample values from the sample mean was 84.4, and in the other sample of the observations, it was 102.6. test whether this difference is significant at the 5% level, given that the five 5% point for $n_1 = 7$ & $n_2 = 9$ d.f is 3.29.

Q5. (2X10=20 Marks)

a. Define Principal Component Analysis (PCA) and explain its real-world applications in data science.

b. Find the Karl Pearson coefficient of correlation for the following data:

x	5	7	8	10	11	13	16
y	33	30	28	20	18	16	9

c. Find the regression line of y on x for the following data:

x	1	3	4	6	8	9	11	14
y	1	2	4	4	5	7	8	9

Estimate the value of y, when $x = 10$.