UNIT-2

2 Marks Questions

1. Define probability.
2. Write the formula for mean.
3. Write the formula for variance.
4. What is a missing value in data?
5. Name two techniques to handle missing data.
6. Define an outlier.
7. State one impact of outliers on statistical analysis.
8. What is normalization in data preprocessing?
9. What is standardization (z-score scaling)?
10. Define null hypothesis (H_0).
11. Define Central Limit Theorem.
12. What is IDE (integrated Development Environment).
13. There are 20 men in a company in which 5 are graduates. If 3 men are picked out of the 20 at random, what is the probability that they are all graduates?
    If $P(A \cap B) = \frac{1}{2}$, $P(A) = \frac{12}{36}$, $P(B) = \frac{9}{36}$. Find $P(A \cap B)$.

3 Marks Questions

1. Explain with example: probability of drawing a red card from a deck.
2. Discuss the different types of curves of Kurtosis with their values of β_2.
3. Discuss about Regression.
4. Differentiate between mean, median, and mode.
5. Explain three methods of handling missing values.
6. Why is deleting missing values not always a good choice?
7. Explain the difference between global outliers and local outliers.
8. What are the steps to perform Min-Max Normalization?
9. Compare normalization and standardization with suitable examples.
10. Why is feature scaling important in machine learning algorithms?
11. Write the main steps in hypothesis testing.
12. Explain the role of p-value in hypothesis testing.
13. Discuss the steps of Data Operations.
14. Find the mode for the following distribution:

Class Interval	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	5	8	7	12	28	20	10	10

15. Discuss about Heatmaps. Differentiate what is application based and Statistics based Heatmaps.

5 Marks Questions

1. Explain the various techniques of handling missing values with examples.
2. Discuss in detail the different types of outliers and how to treat them.
3. With examples, explain normalization and standardization. Show the formula for each.
4. A coin is tossed 3 times. Find the probability of getting:
   a) 3 heads
   b) Exactly 2 heads
   c) At least 1 head
5. A die is rolled twice. Find the probability of:
   a) Sum = 7
   b) Both numbers even
   c) A number greater than 4 on at least one die
6. Explain hypothesis testing with a flow of steps and give a simple example. 7. Explain the difference between confidence interval and point estimate with examples. 8. Discuss the effect of missing values, outliers, and scaling on hypothesis testing results. 9. Explain in detail the importance of data preprocessing (handling missing values, scaling, outliers) before statistical analysis.
10. Discuss the difference between Type I error and Type II error in hypothesis testing with examples.
11. Obtain Karl Pearson's measure of skewness for the following data:

Values	5-10	10-15	15-20	20-25	25-30	30-35	35-40
Frequency	6	8	17	21	15	11	2

12. Define Hypothesis Test with null and alternative hypotheses. How are they used in hypothesis testing? Provide examples from real-life business cases.
13. A teacher claims that the average score of students in a test is 70. A sample of 15 students has a mean score of 74 with a standard deviation of 8. At 5% significance level, test whether the teacher's claim is correct using a **t-test**. (Given table value for $t_{0.05,14}=2.145$)
14. A company claims that the average lifetime of its batteries is 500 hours. A random sample of 100 batteries has a mean life of 490 hours with a standard deviation of 50 hours. At 5% significance level, test whether the company's claim is valid using a z test. (Given

table value for $z_{0.05} = 1.96$)

15. A new teaching method is claimed to improve student performance. A sample of 12 students taught with this method has a mean score of 68 with a standard deviation of 6. Test at 5% significance level if the average score is 65. (Given $t_{0.05,11} = 2.201$)

16. The average weekly study hours of students is claimed to be 25. A sample of 10 students shows a mean of **27** hours and a standard deviation of 5. Test the claim at 5% significance level. (Given $t_{0.05,9} = 2.262$).

17. A medicine manufacturer claims that the average recovery time is 8 days. A sample of 16 patients shows an average recovery time of 9 days with a standard deviation of 2 days. Test the claim at 5% significance level. (Given $t_{0.05,15} = 2.131$)

18. A light bulb manufacturer claims that the average life of bulbs is 1000 hours. A sample of 64 bulbs has a mean life of 980 hours and a standard deviation of 80 hours. Test at 5% significance level. (Given $z_{0.05} = 1.96$)

19. A company claims that the average salary of its workers is ₹20,000. A sample of 100 workers shows an average salary of ₹19,200 with a standard deviation of ₹3,000. Test the company's claim at 1% significance level. (Given $z_{0.01} = 2.33$)

20. It is claimed that the average weight of packets of chips is 200 g. A sample of 81 packets has an average weight of 195 g with a standard deviation of 15 g. Test the claim at 5% significance level. (Given $z_{0.05} = 1.96$)