# DSBDA InSem Sample Bank

1. What is big data? Justify its need. Explain the characteristics of big data.
2. What is data science? Justify the need of data science. Provide any three applications of data science.
3. Explain the data analytics life cycle with examples appropriately.
4. Compare between data science and information science. Justify their relationship.
5. What is Business Intelligence? Justify the relation between Business intelligence and data science.
6. What are the different phases of the data science life cycle? Illustrate each phase.
7. What do you understand about data preparation and the model building phase in data science?
8. What is the need of data science? state and explain applications of data science.
9. Write a short note on 1) data integration 2) data transformation 3) data discretization 4) data cleaning.
10. Describe data transformation techniques in detail.
11. Describe data reduction techniques in detail
12. Illustrate data discretization.
13. What do you understand about Data Standardization? Explain with examples.
14. What are various ways to clean the data
15. What do you mean by data wrangling?
16. What other methods are there for data smoothing?
17. Normalize AGE attribute values in range[0.0-1.0] using Min-Max normalization , and decimal scaling method. AGE: 10,20,30,45,70.

18. calculate Z-Score of the following data : 8,10,15,20.

19. Consider the following data: 10,10,12,15,15,20,25,26,28,40,45,40,20,30.
    a) Use smoothing by bin means to smooth these data, using a bin depth of 3..
    (b) are there any outliers in the data?
    (c) What other methods are there for data smoothing?

20. Analyze the needs of Data Science. Identify five applications of Data Science.

21. Given data for analysis is age values for the data tuples as following :
    20, 21, 22,13, 15, 16, 16, 19, 225, 13, 15, 16, 16, 19, 2, 5
    Calculate Mean, Median,Mode, Range, Midrange
22. Explain the common measures used in descriptive statistics analysis ?
23. What is the difference between mean, median, and mode? Which one do you prefer to use and why?
24. Calculate Mean.Median,Mode and standard deviation for following data. 10, 12, 23, 23, 16, 23, 21, 16
25. Analyze the Descriptive Analysis and its objective to perform for big data analytics.
26. Analyze the measure of relationship and write the properties of correlation.
27. Applying measure of variability, the driving details of two drivers Rohit and Mohit are given below. Which drive is more consistent?

| Drive No | Rohit | Mohit |
|----------|-------|-------|
| 1 | 30 | 30 |
| 2 | 22 | 27 |
| 3 | 26 | 27 |
| 4 | 18 | 28 |
| 5 | 15 | 26 |

28. Illustrate the following : Percentiles, Quartiles , Standard score or Z-score.

29. Apply students' t test,  a random sample of size 20 from a normal population gives a sample mean of 40. Standard deviation of 6 .Test the hypothesis is population mean is 44.Check whether there is any difference between mean. For Df = 19 and  Significance level is 5 % , given T -table  value is 2.093.

30. State the need of statistics and hypothesis testing in data science.
31. Analyze the type of Statistical Analysis, with the importance of statistics in Data Science
32. Find the Pearson's correlation coefficient for the following pairs of observations from a population.

| Age | 43 | 21 | 25 | 42 | 57 | 59 |
|-----|----|----|----|----|----|----|
| Glucose level | 99 | 65 | 79 | 75 | 87 | 81 |

      a) Calculate the Pearson Correlation Coefficient.
      b) Interpret the result

33. An experiment was conducted to test the efficacy of chloromycetin in checking typhoid. In a certain hospital chloromycetin was given to 285 out of the 392 patients suffering from typhoid. The number of typhoid cases were as follows:

| | Typhoid | No Typhoid | Total |
|--|---------|-----------|-------|
| Chloromycetin | 35 | 250 | 285 |
| No chloromycetin | 50 | 57 | 107 |
| Total | 85 | 307 | 392 |

With the help of Chi-square , test the effectiveness of chloromycetin in checking typhoid. (The chi-square value at 5 percent level of significance for one degree of freedom is 3.841).

34. Explain Bayes Theorem with an example
35. Explain the steps of Hypothesis testing with an Example
36. A pharmaceutical company is interested in testing whether or not their new drug relieves pain more than their current drug. They run a statistical hypothesis test with the outcome being the mean score on a pain relief scale. Write Null and Alternative hypothesis
37. List common measures used in descriptive statistics ?

38. The following nine observations were drawn from a normal population:

   9 20 24 23 29 21 17 27

   (i) Test the null hypothesis H0: m = 26 against the alternative hypothesis Ha: m < 26. At what level of significance can H0 be rejected?

   (ii) At what level of significance can H0 : m = 26 be rejected when tested against Ha : m < 26

39. The following sample was taken from a normally distributed population with a known standard deviation $\sigma = 4$. Test the hypothesis that the mean $\mu = 20$ using a level of significance of 0.05 and the alternative that $\mu > 20$:

   Sample data : 23, 32, 22, 31, 27, 25, 21, 24, 20, 18

40. An aptitude test has been conducted to test the aptitude of graduate engineers in the country. The test is conducted so that scores are normally distributed with a standard deviation of 10. A statistical test administered to a random sample of size 500 examinees. The sample yields a mean of 51.07.

   a. Test the hypothesis that the population mean is 50. Consider the level of confidence is 5%. You should clearly show all the five steps in your calculation

41. Justify the importance of hypothesis testing? Formulate the hypothesis and select an appropriate type of hypothesis test for following examples:

42. Example: Assume a beauty-product company believes that spending more money on digital advertising leads to increased sales. To test this, the company may increase money spent on digital advertising during a two-month period and collect data to see if overall sales have increased.