

# SKNCOE Comp 2021-22 Sem-II Unit-Test I

nikamabhi11@gmail.com [Switch account](#)

 Draft saved

\* Required

Email \*

nikamabhi11@gmail.com

Subject :Data science and Big Data Analytics  
:15/03/2022

Date



**Sinhgad Technical Education Society's  
Smt. Kashibai Navale College Of Engineering Pune – 41  
Department of Computer Engineering**

**Unit Test I (A.Y. 2021-22 Sem-II)**



**Sinhgad Institutes**

**Class : TE**

**Marks : 30**

Enter your Full Name : \*

Nikam Abhishek Madhukar

Enter Your Roll No (eg.19U101) : \*

20U133



Enter Your PRN No : \*

72027925k

Enter your Exam Seat No : \*

S190364470

Q.1 Mode refers to the value within a series that occurs \_\_\_\_\_ number of 1 point times.

- ☒ Maximum
- ☐ Minimum
- ☐ Zero
- ☐ Infinite

Clear selection

Q2. \_\_\_\_\_ is not a measure of central tendency. 1 point

- ☐ Mode
- ☐ Mean
- ☒ Range
- ☐ Median

Clear selection



Q3. The number of observations smaller than \_\_\_\_\_ is the same as the number of observations larger than it. 1 point

- ☒ Median
- ☐ Mode
- ☐ Mean
- ☐ None of the above

Clear selection

Q4. \_\_\_\_\_ divides the data into four equal parts. 1 point

- ☐ Median
- ☒ Quartiles
- ☐ Mean
- ☐ None of the above

Clear selection

Q5. The average of squared deviations from the arithmetic mean is known as \_\_\_\_\_. 1 point

- ☐ Quartile Deviation
- ☐ Standard Deviation
- ☒ Variance
- ☐ None of the above

Clear selection



Q6. The scatter within a distribution that is high on each side indicates

1 point

- \_\_\_\_\_
- ☐ High uniformity of data
  - ☒ Outliers of data
  - ☐ Low uniformity of data
  - ☐ None of the above

Clear selection

Q7. The data set below has an outlier of 42.2, 5, 12, 15, 19, 4, 6, 11, 16, 18, 12, 12, 42. What effect does removing the outlier have on the distribution of the data?

1 point

- ☒ The mean will decrease
- ☐ The median will decrease
- ☐ The mean will increase
- ☐ The median will increase

Clear selection



Q8. A survey was conducted by a chocolate manufacturing company, about the likes and dislikes of the flavors they marketed. Which measure of central tendency will help them most to make their choice of the most liked flavor? 1 point

- ☐ Mean
- ☐ Range
- ☒ Mode
- ☐ Median

Clear selection

Q9. Which hypothesis should be written as an inequality? 1 point

- ☒ Alternate Hypothesis
- ☐ Null Hypothesis
- ☐ either alternate or Null
- ☐ None of the above

Clear selection

Q10. When PCC value = 0, then this denotes 1 point

- ☐ positive linear correlation
- ☐ negative linear correlation
- ☒ no linear correlation
- ☐ none of the above

Clear selection



Q11. What is the assumption made for performing the hypothesis test with T distribution? a) b) c) d) 1 point

- ☐ the distribution is non-symmetric
- ☐ the distribution has more than one modal class
- ☐ the distribution has a constant variance
- ☒ the distribution follows a normal distribution

Clear selection

Q12. Which of the following is not a valid probability value? 1 point

- ☐ 0
- ☒  $\frac{4}{3}$
- ☐ 0.57
- ☐  $\frac{11}{13}$

Clear selection

Q13. Which of the following values of the chi-square distribution cannot occur? 1 point

- ☐ 0.61
- ☐ 38.4
- ☒ -2.45
- ☐ 100

Clear selection



Q14. Suppose a person has 8 red, 5 green, 12 orange, and 15 blue balls. Test the null hypothesis that the colors of the balls occur with equal frequency. What is the Chi Square value you get? 1 point

- ☐ 5.6
- ☐ 5.68
- ☐ 5.86
- ☒ 5.8

Clear selection

Q15. Degree of freedom for paired t test based on n pairs of observations is 1 point

- ☐  $2n-1$
- ☐  $n-2$
- ☐  $2(n-1)$
- ☒  $n-1$

Clear selection

Q16. If  $H_0$  is true and we reject it is called 1 point

- ☒ Type I error
- ☐ Type II error
- ☐ standard error
- ☐ sampling error

Clear selection



Q17. A statement made about a population for testing purpose is called? 1 point

- ☐ Statistic
- ☒ Hypothesis
- ☐ Level of Significance
- ☐ Test-Statistic

Clear selection

Q18. Which of the following function gives information about top level data? 1 point

- ☐ Tail
- ☐ Summary
- ☒ Head
- ☐ None of the above

Clear selection

Q19. Which of the following are not data reduction techniques? a) b) c) d) 1 point

- ☐ Data cube aggregation
- ☐ Numerosity reduction
- ☐ Attribute subset
- ☒ Decimal scaling

Clear selection





Q20. KDD stands for\_\_\_\_\_.

1 point

- ☐ Knowledge database in discovery
- ☐ Known discovery in databases
- ☐ Known distributed databases
- ☒ Knowledge discovery in databases

Clear selection

Q21. Removing Duplicate Record in Data Mining Process is called\_\_\_\_\_

1 point

- ☐ Recovery
- ☒ Data Cleaning
- ☐ Data reduction
- ☐ Data transformation

Clear selection

Q22. An outlier is a \_\_\_\_\_

1 point

- ☐ Description of records in the data
- ☒ Data point which is considered different from other data points
- ☐ Record with missing attributes
- ☐ Duplicate record

Clear selection



Q23. \_\_\_\_\_ is a process that reduced the amount of original data and represents it in a much smaller data. 1 point

- ☐ Data Warehousing
- ☐ Data Visualization
- ☐ Data Transformation
- ☒ Data Reduction

Clear selection

Q24. Which of the following statements is true about the collection of data? 1 point

- ☒ The data that is collected from the place of origin is known as primary data
- ☐ The data that is collected from the place of origin is known as secondary data
- ☐ The data that is collected from the place of origin is known as tertiary data
- ☐ None of the above

Clear selection

Q25. Which of the following is the main point of difference between primary data and secondary data? 1 point

- ☐ The collection of secondary data is costlier and more time consuming when compared to primary data
- ☐ The secondary data is always original whereas the primary data is not
- ☒ The primary data is much more reliable than secondary data because it is collected directly from respondents
- ☐ None of the above

Clear selection



Q26. Which of the following is not a data pre-processing methods?

1 point

- ☒ Data Visualization
- ☐ Data Discretization
- ☐ Data Cleaning
- ☐ Data Reduction

Clear selection

Q27. Which is not a data cleaning method?

1 point

- ☐ Binning
- ☐ regression
- ☐ clustering
- ☒ aggregation

Clear selection

Q28. The standard deviation of a set of 90 observations is 105. If the value of each observation is decreased by 9, then the new standard deviation of these observations would be \_\_\_\_\_.

1 point

- ☐ 96
- ☐ 100
- ☒ 105
- ☐ None of the above

Clear selection



Q29. Calculate Standard deviation for the following sample data 2,4,6,8,10 and 12. Is\_\_\_\_\_ 1 point

- ☐ 4.42
- ☐ 2.42
- ☒ 3.42
- ☐ 5.42

Clear selection

Q30. The measure of dispersion which uses only two observations is called\_\_\_\_\_ - 1 point

- ☐ mean
- ☐ median
- ☒ Range
- ☐ coefficient of variation

Clear selection

Submit

Clear form

Never submit passwords through Google Forms.

This form was created inside of Sinhgad Technical Education Society. [Report Abuse](#)

Google Forms

