Unit-II

1. A statement about a population developed for the purpose of testing is called:

(a) Hypothesis

(b) Hypothesis testing

(c) Level of significance

(d) Test-statistic

Answer : a

2. Any hypothesis which is tested for the purpose of rejection under the assumption that it is true is called:

(a) Null hypothesis

(b) Alternative hypothesis

(c) Statistical hypothesis

(d) Composite hypothesis

Answer : a

3. A statement that is accepted if the sample data provide sufficient evidence that the null hypothesis is false is called:

(a) Simple hypothesis

(b) Composite hypothesis

(c) Statistical hypothesis

(d) Alternative hypothesis

Answer : d

4. The alternative hypothesis is also called:

(a) Null hypothesis

(b) Statistical hypothesis

(c) Research hypothesis

(d) Simple hypothesis

Answer : c

5. The probability of rejecting the null hypothesis when it is true is called:

(a) Level of confidence

(b) Level of significance

(c) Power of the test

(d) Difficult to tell

Answer : b

6. If the critical region is located equally in both sides of the sampling distribution of test-statistic, the test is called:

(a) One tailed

(b) Two tailed

(c) Right tailed

(d) Left tailed

Answer : b

7. The choice of one-tailed test and two-tailed test depends upon:

(a) Null hypothesis

(b) Alternative hypothesis

(c) None of these

(d) Composite hypotheses

Answer : b

8. Test of hypothesis Ho: $\mu = 50$ against H1: $\mu > 50$ leads to:

(a) Left-tailed test

(b) Right-tailed test

(c) Two-tailed test

(d) Difficult to tell

Answer : b

9. Testing Ho: $\mu = 25$ against H1: $\mu \neq 25$ leads to:

(a) Two-tailed test

(b) Left-tailed test

(c) Right-tailed test

(d) Neither (a), (b) and (c)

Answer : a

10. A formula that provides a basis for testing a null hypothesis is called:

(a) Test-statistic

(b) Population statistic

(c) Both of these

(d) None of the above

Answer : a

11. $1 - \alpha$ is also called:

(a) Confidence coefficient

(b) Power of the test

(c) Size of the test

(d) Level of significance

Answer : a

12. Area of the rejection region depends on:

(a) Size of α

(b) Size of β

(c) Test-statistic

(d) Number of values

Answer : a

13. Student's t-test is applicable only when:

(a) $n \leq 30$ and σ is known

(b) $n > 30$ and σ is unknown

(c) $n = 30$ and σ is known

(d) All of the above

Answer : a

14. In an unpaired samples t-test with sample sizes n1= 11 and n2= 11, the value of tabulated t should be obtained for:

(a) 10 degrees of freedom

(b) 21 degrees of freedom

(c) 22 degrees of freedom

(d) 20 degrees of freedom

Answer : d

15. The purpose of statistical inference is:

(a) To collect sample data and use them to formulate hypotheses about a population

(b) To draw conclusion about populations and then collect sample data to support the conclusions (c) To draw conclusions about populations from sample data

(d) To draw conclusions about the known value of population parameter

Answer : c

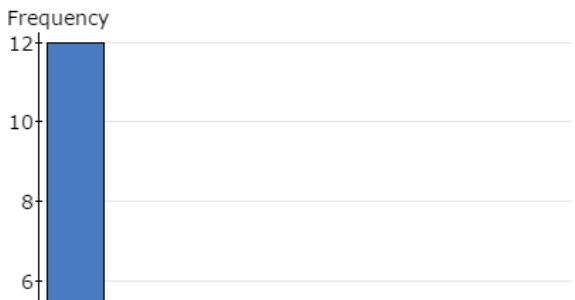
16. The histogram to the right represents the hospital length of stay (in days) for patients at a nearby medical facility. How many patients are included in the histogram?

a. 5

b. 21

c. 17

d. 9



Answer : b

17. Using the histogram to the right that represents the hospital lengths of stay (in days) for patients at a nearby medical facility, determine the relationship between the mean and the median.

a. Mean = Median

b. Mean ≈ Median

c. Mean < Median

d. Mean > Median

Answer : d

18. The statement “If there is sufficient evidence to reject a null hypothesis at the 10%

significance level, then there is sufficient evidence to reject it at the 5% significance level” :

Please select the best answer of those provided below.

a. Always True

b. Never True

c. Sometimes True; the p-value for the statistical test needs to be provided for a conclusion

d. Not Enough Information; this would depend on the type of statistical test used

Answer : c

19.Analysis of variance in short form is?

a) ANOV

b) AVA

c) ANOVA

d) ANVA

Ans:c

20) Which of the following is required by K-means clustering?
a) defined distance metric
b) number of clusters
c) initial guess as to cluster centroids
d) all of the mentioned
Ans: defined distance metric, number of clusters, initial guess as to cluster centroids

21) Hierarchical clustering should be primarily used for exploration.
a) True
b) False
Ans: True

22) Which of the following function is used for k-means clustering?
a) k-means
b) k-mean
c) heatmap
d) none of the mentioned
Ans: k-means

23) The goal of clustering a set of data is to
a)divide them into groups of data that are near each other
b)choose the best data from the set
c)determine the nearest neighbors of each of the data
d)predict the class of data
Ans: divide them into groups of data that are near each other

24) The k-means algorithm...
a)always converges to a clustering that minimizes the mean-square vector-representative distance
b)can converge to different final clustering, depending on initial choice of representatives
c)is widely used in practice
d)is typically done by hand, using paper and pencil
e)should only be attempted by trained professionals
Ans: an converge to different final clustering, depending on initial choice of representatives, is widely used in practice

25) Considering the K-means algorithm, after current iteration, we have 3 centroids (0, 1) (2, 1), (-1, 2). Will points (2, 3) and (2, 0.5) be assigned to the same cluster in the next iteration?
a) Yes
b) No
Ans: Yes

26) What are the two types of Hierarchical Clustering?
a)Top-Down Clustering (Divisive)
b)Bottom-Top Clustering (Agglomerative)
c)Dendrogram

d)K-means
Ans: Top-Down Clustering (Divisive), Bottom-Top Clustering (Agglomerative)

27) The most commonly used measure of similarity is the _____ or its square.
a)euclidean distance
b)city-block distance
c)Chebychev's distance
d)Manhattan distance
Ans: euclidean distance

29) Which of the following is required by K-means clustering?
a)defined distance metric
b)number of clusters
c)initial guess as to cluster centroids
Ans: defined distance metric, number of clusters, initial guess as to cluster centroids

30) Clustering is a-
A. Supervised learning
B. Unsupervised learning
C. Reinforcement learning
D. None
Ans: Unsupervised learning

31) Which of the following clustering algorithms suffers from the problem of convergence at local optima?
A. K- Means clustering
B. Hierarchical clustering
C. Diverse clustering
D. All of the above
Ans: K- Means clustering, Hierarchical clustering, Diverse clustering

32) Which version of the clustering algorithm is most sensitive to outliers?
A. K-means clustering algorithm
B. K-modes clustering algorithm
C. K-medians clustering algorithm
D. None
Ans: K-means clustering algorithm

33) Which of the following is a bad characteristic of a dataset for clustering analysis-
A. Data points with outliers
B. Data points with different densities
C. Data points with non-convex shapes
D. All of the above
Ans: Data points with outliers, Data points with different densities, Data points with non-convex shapes

34) For clustering, we do not require-
A. Labeled data
B. Unlabeled data
C. Numerical data
D. Categorical data
Ans: Labeled Data

35) Which of the following is an application of clustering?
A. Biological network analysis
B. Market trend prediction
C. Topic modeling
D. All of the above
Ans: Biological network analysis, Market trend prediction, Topic modeling

36) The final output of Hierarchical clustering is-
A. The number of cluster centroids
B. The tree representing how close the data points are to each other
C. A map defining the similar data points into individual groups
D. All of the above
Ans: The tree representing how close the data points are to each other

37. Which type of test is the Wilcoxon rank sum test?

a. Parametric

b. non parametric

c. Distributed

d. Normal

38. Input data for Wilcoxon test is normally distributed, True or False?

39. What is the null hypothesis for a Wilcoxon test?

a.Two group means are equal.

b.Two or more group means are equal.

c.Two mean groups are not equal.

d. None of these

40 Which of following test statics is used in Wilcoxon Rank Sum Test?

a. test statistics <= critical value, Ho will be Rejected

b. if test statistics > critical value, Ho will be Rejected

c. if test statistics >critical value, Ho will be accepted

d. none of these.

40. What must you include when applying Wilcoxon Rank sum test?

a. variance

b. Critical Value

c. Rank sum

e. standard deviation

Type1 and type 2 error

41. Type 1 is also called as

a. False Positive

b. false negative

c. True Positive

d. True negative

42. Type 2 is also called as

a. False Positive

b. False negative

c. True Positive

d. True negative

43. Type 1 error occurs when_____

a. Null hypothesis rejected when it is true.
b. Null hypothesis is accepted when it is false
c. Null hypothesis rejected when it is false

d. None of Above

44. Type 2 error occurs when_____

a. Null hypothesis rejected when it is true.
b. Null hypothesis is accepted when it is false
c. Null hypothesis rejected when it is false
d. All of above

ANOVA

44. Analysis of Variance is statistical method of comparing____of several populations.

a. Means

b. variance

c. standard Deviation

d. None of above.

45. ANOVA is used when____

a. If more than two population

b for two population

c. for Three population

d. for any populations

46. What is Null Hypothesis in ANOVA?

a. all group means are equal

b. Three group means are equal

c. atleast one pair of group means unequal.

d. all group means are unequal.

47. What do ANOVA calculate?

a. Z-score

b. F ratio

c. T-score

d. Chi Square

Q.25 What are the two types of variance which can occur in your data?

a. Independent and Dependent

b. Between and within groups

c. Personal and interpersonal

d. Anova and Anoca

Q.26 If between group mean sum of square variability increases value of F statistics_____

a. Increases
b. Decreases
c. Neutral
d. None of these

Q.27 What must you include when applying ANOVA test?

a. Means

b. Critical Value

c. degree of freedom

d. F statistics

e. All of above

Q.28 How many dependent variables are there in a two-way ANOVA?

a.1

b.3

c.2

d.any

Q.29 Which of following test statics is used in ANOVA?

a.if critical value > F ratio, Ho will be Rejected

b.if critical value < F ratio, Ho will be Rejected

c.if critical value > F ratio, Ho will be accepted

d.None of these

Q.30 Various types of ANOVA are___.

a.Two way ANOVA

b.ANCOVA

c.MANOVA

d.ZANOVA