Unit -1

Statistical Inference I

1. Which is a method for testing assertion or assumption about a parameter in a population using data measured in sample?

a) Hypothesis Testing.
b) Mean statistical method
c) Type 1 type 2 errors
d) None of the above

Ans: A.

2. Which is a hypothesis of equality between population parameters?

a) Hypothesis Testing.
b) Null Hypothesis
c) Alternative Hypothesis
d) None of the above

Ans: B.

3. Which is a hypothesis of difference between population parameters?

a) Hypothesis Testing.
b) Null Hypothesis
c) Alternative Hypothesis
d) None of the above

Ans: C.

4. Which hypothesis proposes that no statistical significance exists in a set of given observations?

a) Hypothesis Testing.
b) Null Hypothesis
c) Alternative Hypothesis
d) None of the above

Ans: B.

5. Which hypothesis proposes that existing model is better than proposed model ?

a) Hypothesis Testing.
b) Null Hypothesis
c) Alternative Hypothesis
d) None of the above

Ans: B.

6. Which hypothesis proposes that proposed model is better than existing model?

a) Hypothesis Testing.
b) Null Hypothesis
c) Alternative Hypothesis
d) None of the above

Ans: C.

7. When can Student's t-test can be applied ?

a) When the scaling terms of two populations are known.
b) When Population follows the normal distribution
c) Both a & b
d) None of the above

Ans: C.

8. Which Assumption is made by Student's t-test?

a) groups of data are sampled from populations that follow a normal distribution and that both populations have the same variance.
b) groups of data are sampled from populations that follow a normal distribution, *but it does not assume that those two populations have the same variance*.
c) No Assumptions.
d) None of the above

Ans: A.

9. Which Assumption is made by Welch's t-test?

a) groups of data are sampled from populations that follow a normal distribution and that both populations have the same variance.
b) groups of data are sampled from populations that follow a normal distribution, *but it does not assume that those two populations have the same variance*.
c) No Assumptions.
d) None of the above

Ans: B.

10. Which test is preferred when sample sizes and variances are unequal between groups ?

a) Student's t-test
b) Welch's t-test
c) Both can be prefered
d) None of the above

Ans: A.

11. Test preferred when sample sizes are variances are equal ?
a) Student's t-test
b) Welch's t-test
c) Both can be prefered
d) None of the above

Ans: C.

12. when one rejection of the null hypothesis when it is true then it is called ?
a) **type I error**
b) **type II error**
c) hypothesis error
d) None of the above

Ans: A.

13. when one rejection of the null hypothesis when it is true then it is called ?
a) **False positive**
b) **False negative**
c) hypothesis error
d) None of the above

Ans: A.

14. when one accepts the null hypothesis when it is false then it is called ?
a) **type I error**
b) **type II error**
c) hypothesis error
d) None of the above

Ans: B.

15. when one accepts the null hypothesis when it is false then it is called ?
a) **False positive**
b) **False negative**
c) hypothesis error
d) None of the above

Ans: B.

16. The -------- Test is often described as the non-parametric version of the two-sample t-test.
Wilcoxon Rank Sum
Type 1
Type 2
Non-of the above.

Ans : Wilcoxon Rank Sum

17. -------- is used to determine whether there are any statistically significant differences between the means of three or more independent (unrelated) groups.
    Power
    Clustering
    ANOVA
    Use Cases
    Ans : ANOVA.

18. A use case is a methodology used in system analysis to identify, clarify and organize system requirements.
    Use cases
    Test cases
    Test data
    Analysis
    Ans : Use cases

19. Depending upon the acceptance and rejection of null hypothesis which two types of error produces during the test.
    Type 3 and Type 6
    Type 1 and Type 2
    Type 5 and Type 4
    Type 8 and Type 9
    Ans: Type 1 and Type 2

20. --------- is the systematic computational analysis of data or statistics.
    Analytics
    Searching
    Checking
    Clustering
    Ans: Analytics

21. The K-means algorithm identifies -------- number of centroids.
    K
    R
    N
    M
    Ans:K

22. There is a popular method known as elbow method which is used to determine the optimal value of K to perform the K-Means Clustering Algorithm.
    Test
    Elbow
    ANNOVA

List
Ans: Elbow

23. The ----------means clustering algorithm is used to find groups which have not been explicitly labeled in the data.
    R
    K
    L
    E
    Ans : K

24. A t-test is used when the population parameters (mean and standard deviation) are-----.
    Not Known.
    Known.
    Both
    Non of the above.
    Ans: Not known

25. A ----------is a prediction of the relationship between two variables: the independent variable and the dependent variable.
    Simple hypothesis
    Large hypothesis
    Small hypothesis
    All of the above
    Ans: Simple hypothesis.

26. -------error, in statistical hypothesis testing, is the error caused by rejecting a null hypothesis when it is true.
    Type 1
    Type 2
    Non of the above
    All of the above
    Ans: Type 1

27. --------error is the error that occurs when the null hypothesis is accepted when it is not true.

    Type 1
    Type 2
    Non of the above
    All of the above
    Ans: Type 2

28. A conclusion is drawn that the null hypothesis is false when, in fact, it is true. Therefore,-------- errors are generally considered more serious than ---------- errors.

Type 2, Type 1
Type 1, Type 2
Non of the above
All of the above
Ans : Type 1,Type 2

29. a ---------error occurs when you get the right answer to the wrong question.
Type 1
Type 2
Type 3
All of the above
Ans: Type 3

30. The--------- is a method that determines whether two populations are statistically different from each other.
t-test.
ANOVA
m-test
n-test
Ans : t-test

31. ---------is the task of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups .
Clustering
Wilcoxon rank sum-Test
Searching
Testing
Ans : Clustering

32. How many types of clusters are there?
1 Types
2 Types
3 Types
4 Types
Ans : 3 Types

33. ------- analytics is a form of advanced analytics that examines data or content to answer the question.
Descriptive
Diagnostics
Predictive
Prescriptive
Ans : Diagnostics

34. ---------- is used to test whether two samples are likely to derive from the same population.
Clustering
Wilcoxon rank sum-Test
Searching
Testing
Ans : Wilcoxon rank sum-test

35. --------- is the probability of detecting a "true" effect when it exists.
Size
Power
Volume
Ratio
Ans : Power

36. Choose the correct option:

I. Assuming no difference between samples means Alternative Hypothesis.
II. Assumption of difference between samples means Null Hypothesis.

a. Both I & II are true.
b. Only I is true.
c. Only II is true.
d. Neither I Nor II is True.

Ans: d

37. Choose the correct option:

I. Student's t-test assumes two normally distributed populations, having equal variance.
II. Welch's t-test assumes two normally distributed populations, not necessarily having equal variance.

a. Both I & II are true.
b. Only I is true.
c. Only II is true.
d. Neither I Nor II is True.

Ans: a

38. Choose the correct option:

I. Type I error – rejection of the null hypothesis when the null hypothesis is TRUE
II. Type II error – rejection of the null hypothesis when the null hypothesis is FALSE

a. Both I & II are true.
b. Only I is true.
c. Only II is true.
d. Neither I Nor II is True.

Ans: b

39. Choose the correct option:

I. Type I error – acceptance of the null hypothesis when the null hypothesis is TRUE

II. Type II error – acceptance of the null hypothesis when the null hypothesis is FALSE

a. Both I & II are true.
b. Only I is true.
c. Only II is true.
d. Neither I Nor II is True.

Ans: c

40. Power of a test _____.

a. is the probability of correctly rejecting the null hypothesis.
b. increases as the sample size increases.
c. decreases as the sample size increases.
d. Both a & b.

Ans: d

41. ANOVA stands for _____.

a. Annotation of Variance
b. Analysis of Variance
c. Analysis of Variables
d. None of the above

Ans: b

42. ANOVA was developed by _____.

a. Ronald Conway
b. Ronald Fisher
c. James MacQueen
d. Stuart Lloyd

Ans: b

43. The expected value or _____ of a random variable is the center of its distribution.

a) mode

b) median

c) mean

d) bayesian inference

Answer: c

Explanation: A probability model connects the data to the population using assumptions.

44. The square root of the variance is called the _____ deviation.

a) empirical

b) mean

c) continuous

d) standard

Answer: d

Explanation: Standard Deviation (SD) is the measure of spread of the numbers in a set of data from its mean value.

45. Which measure of spread indicates variation about the mean?
    a) Median
    b) Standard deviation
    c) Mode
    d) Variance
46. Mode is the
    a) middle most frequent value
    b) least frequent value
    c) **maximum frequent value**
    d) none of these
47. Which of the following can not be determined graphically?
    a) Mean
    b) Median
    c) Mode
    d) None of these
48. For example: A cricketer's scores in five ODI matches are as follows: 12, 34, 45, 50, 24 find the mean of the data points.

a) 30
b) 33
c) 34
d) 32

49. If the heights of 5 people are 142 cm, 150 cm, 149 cm, 156 cm and 153 cm.Find the mean height.

a) 152
b) 140
c) 150
d) 149

50. For example, consider the data: 4, 4, 6, 3, 2. Find the median

a) 2
b) 4
c) 5

d) 6

51. Let's consider the data: 56, 67, 54, 34, 78, 43, 23. What is the median?

a) 67
b) 54
c) 34
d) 43