

Dakshay Ahuja - 2010990178

STATISTICS AND PROBABILITY

Dated- 19/09/2023

Class Test

1. In a regression analysis, what does the coefficient of determination (R-squared) measure?
 - A) The strength of the relationship between the dependent and independent variables
 - B) The percentage of variation in the dependent variable explained by the independent variable(s)**
 - C) The probability of a type I error
 - D) The standard error of the regression coefficient
2. What is the central limit theorem?
 - A) It states that the population mean is equal to the sample mean.
 - B) It states that the distribution of the sample mean approaches a normal distribution as the sample size increases, regardless of the shape of the population distribution.**
 - C) It is a measure of the skewness of a dataset.
 - D) It is a method for estimating the population standard deviation.
3. When conducting a hypothesis test, what is the p-value?
 - A) The probability of making a Type I error
 - B) The probability of making a Type II error
 - C) The significance level of the test
 - D) The probability of observing a test statistic as extreme as, or more extreme than, the one computed from the sample data, assuming the null hypothesis is true.**
4. What is the purpose of the chi-squared test of independence?
 - A) To test if the means of two independent samples are equal
 - B) To test if the variances of two independent samples are equal
 - C) To test if two categorical variables are independent of each other**
 - D) To test if two continuous variables are correlated
5. What is the difference between stratified sampling and cluster sampling?
 - A) Stratified sampling divides the population into mutually exclusive groups, while cluster sampling divides the population into random clusters.**
 - B) Stratified sampling divides the population into random clusters, while cluster sampling divides the population into mutually exclusive groups.
 - C) Stratified sampling involves selecting a random sample from each cluster, while cluster sampling involves selecting a random sample from each stratum.
 - D) There is no difference between stratified sampling and cluster sampling.
6. What is the formula for calculating the standard error of the sample mean in a simple random sample?
 - A) Standard Error = (Population Standard Deviation) / $\sqrt{\text{Sample Size}}$
 - B) Standard Error = (Sample Standard Deviation) / $\sqrt{\text{Sample Size}}$**
 - C) Standard Error = (Population Standard Deviation) / Sample Size
 - D) Standard Error = (Sample Standard Deviation) / Sample Size
7. What is the difference between correlation and causation in statistics?
 - A) Correlation implies a cause-and-effect relationship, while causation measures the strength of association between two variables.
 - B) Correlation measures the strength and direction of a relationship between two variables, while causation implies that one variable directly causes the other.**
 - C) Correlation and causation are interchangeable terms in statistics.
 - D) Correlation is only applicable to categorical data, while causation applies to continuous data.

8. What does the term "pooled variance" refer to in the context of statistical hypothesis testing?
- A) The variance of the null hypothesis distribution
 - B) The weighted average of the variances from two independent samples**
 - C) The variance of the sample mean
 - D) The variance of the population
9. What is the term for the measure of central tendency that represents the middle value in a dataset when it is arranged in ascending order?
- A) Mean
 - B) Median**
 - C) Mode
 - D) Range
10. Which of the following is a measure of dispersion in a dataset?
- A) Mean
 - B) Median
 - C) Standard Deviation**
 - D) Mode
11. In a standard normal distribution, what is the mean and standard deviation?
- A) Mean = 0, Standard Deviation = 1**
 - B) Mean = 1, Standard Deviation = 0
 - C) Mean = 0, Standard Deviation = 0
 - D) Mean = 1, Standard Deviation = 1
12. If two events are mutually exclusive, what is the probability of both events occurring?
- A) 0**
 - B) 0.25
 - C) 0.5
 - D) 1
13. In a binomial distribution, what are the two key parameters?
- A) Mean and Median
 - B) Standard Deviation and Variance
 - C) Probability of success and number of trials**
 - D) Mode and Range
14. What is the formula for calculating the variance of a dataset?
- A) Sum of squares of the data points
 - B) Mean divided by the number of data points
 - C) Standard deviation squared**
 - D) Range of the data
15. What is the range of a dataset?
- A) The difference between the largest and smallest data points**
 - B) The average of the data points
 - C) The middle value in the dataset
 - D) The most frequently occurring value
16. In a normal distribution, what percentage of data falls within one standard deviation from the mean?
- A) 25%
 - B) 50%

C) 68%

D) 95%

17. What is the formula for calculating the probability of an event occurring given the odds in favor of the event?

A) Probability = Odds / (Odds + 1)

B) Probability = 1 / (Odds + 1)

C) Probability = Odds / (Odds - 1)

D) Probability = 1 / Odds

18. If two events are independent, what is the probability of both events occurring?

A) The sum of their probabilities

B) The product of their probabilities

C) The difference of their probabilities

D) The ratio of their probabilities

19. What is the mode of the following dataset: 3, 5, 5, 7, 9, 9, 9?

A) 3

B) 5

C) 7

D) 9

20. Which of the following is a measure of the spread of data points around the mean?

A) Median

B) Mode

C) Range

D) Standard Deviation

21. In a normal distribution, what is the approximate range that contains 99.7% of the data?

A) Within 1 standard deviation from the mean

B) Within 2 standard deviations from the mean

C) Within 3 standard deviations from the mean

D) Within 4 standard deviations from the mean

22. What does the term "expected value" represent in probability theory?

A) The most likely outcome of an experiment

B) The mean of a probability distribution

C) The maximum value in a dataset

D) The mode of a probability distribution

23. Which probability distribution is used to model the number of successes in a fixed number of independent Bernoulli trials?

A) Binomial distribution

B) Normal distribution

C) Exponential distribution

D) Poisson distribution

24. What is the formula for calculating the probability of an event A and event B occurring when they are independent?

A) $P(A \text{ and } B) = P(A) * P(B)$

B) $P(A \text{ and } B) = P(A) + P(B)$

C) $P(A \text{ and } B) = P(A) / P(B)$

D) $P(A \text{ and } B) = P(A) - P(B)$

25. Which measure of central tendency is affected by outliers?

- A) Mean
- B) Median
- C) Mode
- D) Range

26. What is the formula for calculating the standard deviation of a sample?

- A) Sum of squares of the data points divided by n
- B) Square root of the sum of squares of the data points divided by n
- C) Square root of the sum of squares of the data points divided by (n-1)
- D) Mean of the data points divided by n

27. In a normal distribution, what is the skewness of the data?

- A) Positively skewed
- B) Negatively skewed
- C) Symmetric
- D) Cannot be determined from the information given

28. Which probability distribution is used to model the time between events in a Poisson process?

- A) Binomial distribution
- B) Normal distribution
- C) Exponential distribution
- D) Poisson distribution

29. What is the formula for calculating the variance of a sample?

- A) Sum of squares of the data points divided by n
- B) Sum of squares of the data points divided by (n-1)
- C) Square root of the sum of squares of the data points divided by n
- D) Mean of the data points divided by n

30. If two events are mutually exclusive, what is the probability of either event A or event B occurring?

- A) $P(A \text{ or } B) = P(A) * P(B)$
- B) $P(A \text{ or } B) = P(A) + P(B)$
- C) $P(A \text{ or } B) = P(A) / P(B)$
- D) $P(A \text{ or } B) = P(A) - P(B)$

31. Which probability distribution is commonly used to model the time until an event occurs, such as the time until a customer arrives at a store?

- A) Normal distribution
- B) Exponential distribution
- C) Poisson distribution
- D) Binomial distribution

32. Which statistical test is used to determine if there is a significant difference between the means of three or more groups?

- A) Chi-squared test
- B) ANOVA (Analysis of Variance)
- C) T-test
- D) Regression analysis

33. When should you use a one-tailed hypothesis test instead of a two-tailed test?

- A) When you want to be more conservative in your conclusions.
- B) When you have a directional hypothesis, and you are only interested in one side of the distribution.

- C) When you have a non-directional hypothesis.
- D) When you want to increase the power of the test.

34. What is the purpose of a confidence interval in statistics?

- A) To provide a range of values where the population parameter is likely to fall.**
- B) To test the null hypothesis.
- C) To determine the effect size of an experiment.
- D) To calculate the p-value.

35. Which of the following is NOT an assumption of linear regression?

- A) Independence of errors.
- B) Homoscedasticity (constant variance of errors).
- C) Normality of predictor variables.**
- D) Linearity of the relationship between predictors and the response variable.

36. What is the primary use of a contingency table in statistics?

- A) To display the distribution of a single categorical variable.
- B) To compare two or more quantitative variables.
- C) To analyse the relationship between two or more categorical variables.**
- D) To calculate the median of a dataset.

37. Which statistical test is appropriate for testing the association between two categorical variables?

- A) Chi-squared test**
- B) ANOVA
- C) Pearson correlation
- D) T-test

38. What is the mean of a Poisson distribution with rate parameter λ ?

- A) λ**
- B) λ^2
- C) $1/\lambda$
- D) $1/\sqrt{\lambda}$

39. In linear regression, which method is used to handle multicollinearity among predictor variables?

- A) Principal Component Analysis (PCA)
- B) Ridge Regression**
- C) Lasso Regression
- D) Logistic Regression

40. What is the formula for calculating the Akaike Information Criterion (AIC) for model selection?

- A) $AIC = 2k - 2\ln(L)$**
- B) $AIC = 2k + 2\ln(L)$
- C) $AIC = -2k - 2\ln(L)$
- D) $AIC = -2k + 2\ln(L)$