300 STATISTICS MCQ Q&A FOR DATA SCIENCE

EASY-LEVEL

1. **Which of the following is a measure of central tendency?**
   A) Variance
   B) Standard deviation
   C) Mean
   D) Range
   Answer: C) Mean

2. **Which measure describes the spread of data?**
   A) Median
   B) Mode
   C) Range
   D) Mean
   Answer: C) Range

3. **The most frequent value in a dataset is called the:**
   A) Mean
   B) Median
   C) Mode
   D) Standard deviation
   Answer: C) Mode

4. **The probability of an event occurring is always between:**
   A) 0 and 100
   B) 0 and 1
   C) -1 and 1
   D) -100 and 100
   Answer: B) 0 and 1

5. **Which measure is not affected by extreme values in the dataset?**
   A) Mean
   B) Median
   C) Range
   D) Standard deviation
   Answer: B) Median

6. **Which statistical method involves making predictions or inferences about a population based on a sample?**
   A) Descriptive Statistics
   B) Inferential Statistics
   C) Experimental Statistics
   D) Predictive Analytics
   Answer: B) Inferential Statistics

7. **In a normal distribution, what percent of the data lies within one standard deviation of the mean?**
   A) 50%
   B) 68%
   C) 95%
   D) 99%
   Answer: B) 68%

8. **What is the median of the following set of numbers: 2, 4, 6, 8, 10?**
   A) 4

B) 5
C) 6
D) 8
Answer: C) 6

9. **Which is an example of qualitative data?**
   A) Height
   B) Age
   C) Gender
   D) Weight
   Answer: C) Gender

10. **If you roll a fair six-sided die, what is the probability of rolling a 3?**
    A) 1/3
    B) 1/6
    C) 1/2
    D) 1/12
    Answer: B) 1/6

11. **What does the term "population" refer to in statistics?**
    A) A subset of individuals from a larger group
    B) All individuals or items in a defined group
    C) Only the average of a dataset
    D) Data collected from a single experiment
    Answer: B) All individuals or items in a defined group

12. **In a normal distribution, the mean, median, and mode are typically:**
    A) All different
    B) All equal
    C) Mode is always higher than the mean
    D) Mean is always higher than the median
    Answer: B) All equal

13. **The range of a dataset is calculated by:**
    A) Adding the smallest and largest values
    B) Subtracting the smallest value from the largest value
    C) Dividing the sum by the number of values
    D) Finding the middle value of the dataset
    Answer: B) Subtracting the smallest value from the largest value

14. **What is the average of the squared deviations from the mean called?**
    A) Standard deviation
    B) Mean
    C) Median
    D) Variance
    Answer: D) Variance

15. **Which graphical representation is best for showing the distribution of a single variable?**
    A) Scatterplot
    B) Histogram
    C) Line chart
    D) Pie chart
    Answer: B) Histogram

16. **If a dataset is "right-skewed," which measure of central tendency is typically the highest?**
    A) Mean

B) Median
C) Mode
D) Range
Answer: A) Mean

17. **What type of variable has categories without any specific order?**
A) Nominal
B) Ordinal
C) Interval
D) Ratio
Answer: A) Nominal

18. **A box plot is commonly used to display:**
A) Only the mean of the dataset
B) Measures of central tendency and spread
C) Frequency of each data point
D) Only the maximum and minimum values
Answer: B) Measures of central tendency and spread

19. **The probability of a coin landing on heads in a single flip is:**
A) 0
B) 0.5
C) 1
D) 2
Answer: B) 0.5

20. **The interquartile range (IQR) is defined as:**
A) Q1 - Q3
B) Q3 - Q1
C) Q3 + Q1
D) The range of all data points
Answer: B) Q3 - Q1

21. **A sample is:**
A) All items in a population
B) A summary of a population
C) A subset of a population
D) A parameter
Answer: C) A subset of a population

22. **Which of these is a measure of variability?**
A) Median
B) Mode
C) Variance
D) Mean
Answer: C) Variance

23. **If the variance of a dataset is 16, what is the standard deviation?**
A) 4
B) 8
C) 16
D) 32
Answer: A) 4

24. **In hypothesis testing, the null hypothesis (H0) represents:**
A) The conclusion to reject

B) A statement of no effect or no difference
C) A statement that there is an effect
D) None of the above
Answer: B) A statement of no effect or no difference

25. **Which of the following is true in a right-skewed distribution?**
A) Mean < Median < Mode
B) Mean > Median > Mode
C) Median > Mean > Mode
D) Mode > Mean > Median
Answer: B) Mean > Median > Mode

26. **What type of data can have only whole numbers?**
A) Continuous
B) Categorical
C) Discrete
D) None of the above
Answer: C) Discrete

27. **In probability, the sum of all possible outcomes is:**
A) 1
B) 0
C) 100
D) Infinity
Answer: A) 1

28. **A survey that measures people's satisfaction with a service on a scale of 1 to 5 is collecting:**
A) Nominal data
B) Ordinal data
C) Interval data
D) Ratio data
Answer: B) Ordinal data

29. **What is the mode of the following set of numbers: 2, 4, 4, 5, 7?**
A) 2
B) 4
C) 5
D) 7
Answer: B) 4

30. **Which type of graph is used to show the frequency of categorical data?**
A) Histogram
B) Scatterplot
C) Pie chart
D) Line graph
Answer: C) Pie chart

31. **The median is a measure of:**
A) Variability
B) Central tendency
C) Frequency
D) Probability
Answer: B) Central tendency

32. **Which type of data is measured on a continuous scale?**
A) Nominal

B) Ordinal
C) Interval
D) Binary
Answer: C) Interval

33. **The sum of squares is used to measure:**
A) Central tendency
B) Spread or dispersion
C) The median
D) Sample size
Answer: B) Spread or dispersion

34. **The standard deviation is:**
A) The average of the dataset
B) The square root of the variance
C) Twice the variance
D) The square of the variance
Answer: B) The square root of the variance

35. **A z-score represents:**
A) A measure of central tendency
B) The probability of a data point occurring
C) How many standard deviations a value is from the mean
D) The mean of a dataset
Answer: C) How many standard deviations a value is from the mean

36. **The likelihood of a single specific outcome occurring is known as:**
A) Frequency
B) Probability
C) Distribution
D) Variability
Answer: B) Probability

37. **Which of the following is a parametric test?**
A) Chi-square test
B) T-test
C) Mann-Whitney test
D) Kruskal-Wallis test
Answer: B) T-test

38. **If the p-value is less than the significance level (α), we:**
A) Accept the null hypothesis
B) Reject the null hypothesis
C) Accept the alternative hypothesis
D) Both B and C
Answer: D) Both B and C

39. **When data points tend to cluster around the mean, the dataset has a:**
A) High standard deviation
B) Low standard deviation
C) High mean
D) High range
Answer: B) Low standard deviation

40. **What is the range of probabilities?**
A) -1 to +1

B) 0 to 1
C) -∞ to +∞
D) 0 to 100
Answer: B) 0 to 1

41. **Which graphical tool is best for visualizing the spread and central tendency of data?**
A) Histogram
B) Box plot
C) Scatterplot
D) Line chart
Answer: B) Box plot

42. **A two-way table is commonly used to summarize:**
A) Two categorical variables
B) Two continuous variables
C) The range and mean
D) Variability and central tendency
Answer: A) Two categorical variables

43. **The area under the normal curve between two z-scores represents:**
A) A measure of central tendency
B) A probability
C) The mean of a dataset
D) Standard deviation
Answer: B) A probability

44. **In a sample, the mean height of students is 160 cm. This value is called:**
A) Parameter
B) Statistic
C) Variable
D) Data
Answer: B) Statistic

45. **The mean, median, and mode are all types of:**
A) Hypotheses
B) Data transformations
C) Measures of central tendency
D) Sample sizes
Answer: C) Measures of central tendency

46. **If the mean and median are equal, the data distribution is likely to be:**
A) Skewed left
B) Symmetrical
C) Skewed right
D) Bimodal
Answer: B) Symmetrical

47. **What type of data can only take specific values (like whole numbers)?**
A) Continuous
B) Discrete
C) Interval
D) Ordinal
Answer: B) Discrete

48. **Which of the following describes an outlier?**
A) A value near the mean

B) A value far from the mean
C) The smallest value in a dataset
D) The median of a dataset
Answer: B) A value far from the mean

49. **In probability, the likelihood of two independent events both occurring is found by:**
A) Adding their probabilities
B) Subtracting their probabilities
C) Multiplying their probabilities
D) Dividing their probabilities
Answer: C) Multiplying their probabilities

50. **The probability of the complement of an event happening is given by:**
A) 1 + Probability of the event
B) 1 - Probability of the event
C) Probability of the event - 1
D) The square root of the probability of the event
Answer: B) 1 - Probability of the event

51. **A bar chart is typically used to represent:**
A) Continuous data
B) Discrete data
C) Only ordinal data
D) Frequency of a single continuous variable
Answer: B) Discrete data

52. **The mean of a dataset is also known as:**
A) The median
B) The central tendency
C) The average
D) The mode
Answer: C) The average

53. **If two events are mutually exclusive, then the probability of both events occurring is:**
A) 1
B) 0
C) Equal to the sum of each event's probability
D) Equal to the product of each event's probability
Answer: B) 0

54. **What is the standard deviation of a set of identical numbers?**
A) 1
B) 0
C) It varies based on the mean
D) Equal to the range
Answer: B) 0

55. **In a dataset with an odd number of values, the median is the:**
A) Mean of the dataset
B) Value in the center
C) Sum of the largest and smallest values
D) Sum of all values divided by two
Answer: B) Value in the center

56. **If a dataset has two modes, it is called:**
A) Bimodal

B) Trimodal
C) Unimodal
D) Multimodal
Answer: A) Bimodal

57. **Which of these is a continuous variable?**
A) Number of children in a family
B) Number of pets a person has
C) Temperature
D) Number of books on a shelf
Answer: C) Temperature

58. **In a positively skewed distribution, the median is typically:**
A) Greater than the mean
B) Less than the mean
C) Equal to the mean
D) Equal to the mode
Answer: B) Less than the mean

59. **In hypothesis testing, the significance level is often set at:**
A) 10%
B) 5%
C) 50%
D) 0%
Answer: B) 5%

60. **The main purpose of a scatterplot is to:**
A) Show the distribution of a single variable
B) Compare means of two variables
C) Show the relationship between two variables
D) Display percentages of a whole
Answer: C) Show the relationship between two variables

61. **The square of the standard deviation is called the:**
A) Variance
B) Mean
C) Median
D) Range
Answer: A) Variance

62. **If the p-value is greater than the significance level, we:**
A) Reject the null hypothesis
B) Accept the null hypothesis
C) Accept the alternative hypothesis
D) Reject both hypotheses
Answer: B) Accept the null hypothesis

63. **In a dataset, the first quartile (Q1) is:**
A) The middle value of the data
B) The value below which 25% of data lies
C) The average of all data points
D) The value below which 50% of data lies
Answer: B) The value below which 25% of data lies

64. **A statistic that measures the strength and direction of a linear relationship between two variables is called the:**

A) Standard deviation
B) Range
C) Correlation coefficient
D) Variance
Answer: C) Correlation coefficient

65. **The result of dividing the sum of a set of data by the number of data points is called the:**
A) Median
B) Mode
C) Mean
D) Range
Answer: C) Mean

66. **The concept that larger samples give a more accurate estimation of a population parameter is known as:**
A) The law of averages
B) The law of large numbers
C) Central tendency
D) Probability theorem
Answer: B) The law of large numbers

67. **A test that compares the means of two independent groups is called a:**
A) Paired t-test
B) Independent t-test
C) Z-test
D) Chi-square test
Answer: B) Independent t-test

68. **Which measure of central tendency is most appropriate for categorical data?**
A) Mean
B) Median
C) Mode
D) Range
Answer: C) Mode

69. **Which of the following is NOT a measure of dispersion?**
A) Mean
B) Range
C) Variance
D) Standard deviation
Answer: A) Mean

70. **If the data is negatively skewed, the mean is typically:**
A) Less than the median
B) Greater than the median
C) Equal to the median
D) Greater than the mode
Answer: A) Less than the median

71. **Which measure is affected by extreme values in a dataset?**
A) Mode
B) Mean
C) Median
D) Interquartile Range
Answer: B) Mean

72. **What is the probability of drawing a red card from a standard deck of 52 playing cards?**
    A) 1/2
    B) 1/4
    C) 1/3
    D) 1/13
    Answer: A) 1/2

73. **What does a z-score of 0 indicate?**
    A) The value is at the mean
    B) The value is one standard deviation above the mean
    C) The value is one standard deviation below the mean
    D) The value is an outlier
    Answer: A) The value is at the mean

74. **Which statistical measure divides data into 100 equal parts?**
    A) Quartile
    B) Decile
    C) Percentile
    D) Range
    Answer: C) Percentile

75. **In probability, an event that has a probability of 1 will:**
    A) Never occur
    B) Sometimes occur
    C) Always occur
    D) Be uncertain to occur
    Answer: C) Always occur

76. **In a normal distribution, approximately what percentage of data lies within one standard deviation of the mean?**
    A) 50%
    B) 68%
    C) 75%
    D) 95%
    Answer: B) 68%

77. **Which term describes the "middle" value of a dataset?**
    A) Mean
    B) Mode
    C) Median
    D) Range
    Answer: C) Median

78. **A sample statistic is used to estimate a:**
    A) Parameter
    B) Variable
    C) Population
    D) Data point
    Answer: A) Parameter

79. **If data is "left-skewed," then the mean is typically:**
    A) Less than the median
    B) Greater than the median
    C) Equal to the median
    D) Less than the mode
    Answer: A) Less than the median

80. **Which type of error occurs when the null hypothesis is incorrectly rejected?**
    A) Type I error
    B) Type II error
    C) Standard error
    D) Probability error
    Answer: A) Type I error

81. **The interquartile range (IQR) measures the spread of the:**
    A) Entire dataset
    B) Middle 50% of data
    C) Top 25% of data
    D) Bottom 50% of data
    Answer: B) Middle 50% of data

82. **The null hypothesis typically represents:**
    A) A statement of change
    B) A statement of no effect or no difference
    C) The desired outcome
    D) The expected sample mean
    Answer: B) A statement of no effect or no difference

83. **Which graphical method is most appropriate for showing a frequency distribution of a continuous variable?**
    A) Bar chart
    B) Histogram
    C) Pie chart
    D) Line graph
    Answer: B) Histogram

84. **A small p-value (e.g., less than 0.05) indicates that:**
    A) The null hypothesis is likely true
    B) The null hypothesis should be rejected
    C) The sample size is too small
    D) The effect size is large
    Answer: B) The null hypothesis should be rejected

85. **Which of the following values is not affected by outliers?**
    A) Mean
    B) Median
    C) Range
    D) Standard deviation
    Answer: B) Median

86. **A confidence interval provides:**
    A) The probability of a single event
    B) A range of values likely to contain a population parameter
    C) The mean of a sample
    D) The standard deviation of a sample
    Answer: B) A range of values likely to contain a population parameter

87. **If two events cannot happen at the same time, they are:**
    A) Independent
    B) Mutually exclusive
    C) Dependent
    D) Complementary
    Answer: B) Mutually exclusive

88. **Which of these is not a measure of central tendency?**
    A) Median
    B) Mean
    C) Mode
    D) Range
    Answer: D) Range

89. **What does "statistically significant" mean?**
    A) The result occurred by random chance
    B) The result is unlikely to occur by chance alone
    C) The result is a large value
    D) The result is important
    Answer: B) The result is unlikely to occur by chance alone

90. **The mean is best used as a measure of central tendency when the data is:**
    A) Skewed
    B) Symmetrical
    C) Categorical
    D) Multimodal
    Answer: B) Symmetrical

91. **If the probability of an event is 0.8, then the probability of its complement is:**
    A) 0.2
    B) 0.5
    C) 1.8
    D) 0.8
    Answer: A) 0.2

92. **Which hypothesis test compares the means of two paired groups?**
    A) Independent t-test
    B) Chi-square test
    C) Paired t-test
    D) ANOVA
    Answer: C) Paired t-test

93. **What is the term for the likelihood of observing a result as extreme as, or more extreme than, the observed result?**
    A) Confidence interval
    B) P-value
    C) Null hypothesis
    D) Effect size
    Answer: B) P-value

94. **If all values in a dataset are the same, the variance is:**
    A) High
    B) Low
    C) Zero
    D) Infinite
    Answer: C) Zero

95. **The central limit theorem states that the distribution of the sample mean:**
    A) Will be bimodal
    B) Will always be skewed
    C) Approaches normality as the sample size increases
    D) Decreases with larger sample sizes
    Answer: C) Approaches normality as the sample size increases

96. **Which graphical method is best for comparing parts of a whole?**
    A) Pie chart
    B) Box plot
    C) Histogram
    D) Scatterplot
    Answer: A) Pie chart

97. **The term "parameter" refers to:**
    A) A characteristic of a sample
    B) A summary measure for a population
    C) The mean of a dataset
    D) The standard deviation of a dataset
    Answer: B) A summary measure for a population

98. **The purpose of inferential statistics is to:**
    A) Summarize data from a sample
    B) Make predictions or inferences about a population based on a sample
    C) Calculate the mean and median
    D) Display data visually
    Answer: B) Make predictions or inferences about a population based on a sample

99. **Which of these is a measure of spread?**
    A) Median
    B) Mode
    C) Range
    D) Mean
    Answer: C) Range

100. **The type of variable that can take on an infinite number of values within a range is called:**
    A) Categorical
    B) Continuous
    C) Ordinal
    D) Discrete
    Answer: B) Continuous

MEDIUM-LEVEL

1. **Which of the following represents a measure of relative standing within a data set?**
   A) Range
   B) Z-score
   C) Mean
   D) Variance
   Answer: B) Z-score

2. **When constructing a 95% confidence interval, we are stating that:**
   A) The interval will always contain the true population parameter
   B) The interval will contain the true population parameter 95% of the time
   C) The sample mean is equal to the population mean
   D) The interval includes all values of the dataset
   Answer: B) The interval will contain the true population parameter 95% of the time

3. **In a regression analysis, the coefficient of determination (R^2) measures:**
   A) The probability of the null hypothesis
   B) The proportion of variance explained by the model
   C) The slope of the regression line
   D) The residual error in predictions
   Answer: B) The proportion of variance explained by the model

4. **Which of the following is an example of a non-parametric test?**
   A) ANOVA
   B) Chi-square test
   C) Paired t-test
   D) Z-test
   Answer: B) Chi-square test

5. **The shape of a chi-square distribution is determined by:**
   A) The sample size
   B) The degrees of freedom
   C) The significance level
   D) The median
   Answer: B) The degrees of freedom

6. **The null hypothesis for a chi-square test of independence states that:**
   A) The variables are independent
   B) The variables are dependent
   C) The data is normally distributed
   D) There is a correlation between the variables
   Answer: A) The variables are independent

7. **If a distribution has a positive skew, then:**
   A) The median is greater than the mean
   B) The mean is greater than the median
   C) The median is equal to the mode
   D) The mean is less than the mode
   Answer: B) The mean is greater than the median

8. **In ANOVA, what does a large F-statistic indicate?**
   A) Low variability between groups
   B) High variability within groups
   C) A significant difference between group means

D) No difference between group means
Answer: C) A significant difference between group means

9. **Which test would you use to compare more than two related groups?**
   A) Paired t-test
   B) Independent t-test
   C) One-way ANOVA
   D) Repeated measures ANOVA
   Answer: D) Repeated measures ANOVA

10. **What does a high p-value (e.g., 0.9) suggest about the null hypothesis?**
    A) It should be rejected
    B) There is weak evidence against the null hypothesis
    C) It should be accepted as a fact
    D) The test needs to be repeated
    Answer: B) There is weak evidence against the null hypothesis

11. **Which of the following describes the "standard error of the mean"?**
    A) The square root of the variance
    B) The standard deviation divided by the square root of the sample size
    C) The sum of all sample means
    D) The range of the sample
    Answer: B) The standard deviation divided by the square root of the sample size

12. **When using a 1-way ANOVA, what is the primary assumption about the groups?**
    A) They have equal means
    B) They are normally distributed with equal variances
    C) They have different variances
    D) They are related samples
    Answer: B) They are normally distributed with equal variances

13. **If you want to test the relationship between two continuous variables, which method should you use?**
    A) Chi-square test
    B) Pearson correlation
    C) Independent t-test
    D) Kruskal-Wallis test
    Answer: B) Pearson correlation

14. **The purpose of the Tukey HSD (Honestly Significant Difference) test is to:**
    A) Test for normality
    B) Compare variances
    C) Determine which group means are significantly different
    D) Test the goodness-of-fit
    Answer: C) Determine which group means are significantly different

15. **Which of the following does NOT assume a normal distribution?**
    A) Paired t-test
    B) Mann-Whitney U test
    C) Z-test
    D) One-way ANOVA
    Answer: B) Mann-Whitney U test

16. **A Type II error occurs when:**
    A) The null hypothesis is incorrectly rejected
    B) The null hypothesis is correctly accepted
    C) A false null hypothesis is not rejected

D) A true null hypothesis is rejected
Answer: C) A false null hypothesis is not rejected

17. **The interquartile range (IQR) is calculated as:**
    A) Q1 + Q3
    B) Q3 - Q1
    C) Mean - Median
    D) Median - Mode
    Answer: B) Q3 - Q1

18. **The sampling distribution of the sample mean has:**
    A) The same mean as the population mean
    B) A larger variance than the population
    C) A lower mean than the population mean
    D) No relationship to the population mean
    Answer: A) The same mean as the population mean

19. **Which of these statements is TRUE about the F-distribution?**
    A) It is symmetric like the normal distribution
    B) It is used to compare variances between two populations
    C) It is centered around zero
    D) It is always positively skewed
    Answer: D) It is always positively skewed

20. **If a correlation coefficient is 0, this indicates that:**
    A) There is no linear relationship between the variables
    B) The variables are independent
    C) The variables are highly correlated
    D) The sample size is too small
    Answer: A) There is no linear relationship between the variables

21. **When comparing two independent groups with unequal variances, which test is appropriate?**
    A) Paired t-test
    B) Independent t-test with Welch's correction
    C) Z-test
    D) Kruskal-Wallis test
    Answer: B) Independent t-test with Welch's correction

22. **Which test can be used to determine if a sample mean significantly differs from a known population mean when the population standard deviation is unknown?**
    A) Z-test
    B) Chi-square test
    C) T-test
    D) ANOVA
    Answer: C) T-test

23. **A significant result in a hypothesis test implies that:**
    A) The null hypothesis is true
    B) The sample mean equals the population mean
    C) The sample provides sufficient evidence to reject the null hypothesis
    D) The population is normally distributed
    Answer: C) The sample provides sufficient evidence to reject the null hypothesis

24. **If the 95% confidence interval for the mean difference between two groups includes zero, then:**
    A) The two groups have equal variances
    B) The null hypothesis cannot be rejected

C) The alternative hypothesis should be accepted
D) The p-value is less than 0.05
Answer: B) The null hypothesis cannot be rejected

25. **What does the "power" of a statistical test measure?**
A) The probability of making a Type I error
B) The probability of making a Type II error
C) The probability of correctly rejecting a false null hypothesis
D) The significance level of the test
Answer: C) The probability of correctly rejecting a false null hypothesis

26. **If two variables have a correlation of -0.8, this indicates:**
A) A strong negative linear relationship
B) A weak positive linear relationship
C) No relationship
D) A strong positive linear relationship
Answer: A) A strong negative linear relationship

27. **A standard normal distribution has a mean of ___ and a standard deviation of ___.**
A) 1, 0
B) 0, 1
C) 1, 1
D) 0, 0
Answer: B) 0, 1

28. **Which of the following is NOT a type of correlation coefficient?**
A) Pearson's
B) Spearman's
C) Kendall's
D) ANOVA
Answer: D) ANOVA

29. **In hypothesis testing, the “alternative hypothesis” is:**
A) The hypothesis that there is no effect
B) A hypothesis that contradicts the null hypothesis
C) Always accepted if the p-value is high
D) The hypothesis that describes sample statistics
Answer: B) A hypothesis that contradicts the null hypothesis

30. **In a normal distribution, approximately what percent of data lies within two standard deviations of the mean?**
A) 68%
B) 75%
C) 95%
D) 99.7%
Answer: C) 95%

31. **What does an R-squared value of 0.85 imply?**
A) 85% of the variance in the dependent variable is explained by the independent variable(s)
B) There is a weak correlation
C) 85% of the data is above the mean
D) 15% of the data is unexplained
Answer: A) 85% of the variance in the dependent variable is explained by the independent variable(s)

32. **Which statistical test is appropriate for comparing the means of more than two independent groups?**
A) Paired t-test

B) Chi-square test
C) One-way ANOVA
D) Pearson correlation
Answer: C) One-way ANOVA

33. **When performing a regression analysis, multicollinearity occurs when:**
A) The residuals are normally distributed
B) The predictors are highly correlated with each other
C) The response variable is normally distributed
D) The predictors are independent
Answer: B) The predictors are highly correlated with each other

34. **If a dataset is negatively skewed, the mean is generally:**
A) Less than the median
B) Greater than the median
C) Equal to the median
D) Equal to the mode
Answer: A) Less than the median

35. **Which of these tests can be used to assess the normality of a dataset?**
A) Shapiro-Wilk test
B) T-test
C) ANOVA
D) Pearson correlation
Answer: A) Shapiro-Wilk test

36. **A regression model has the form Y=2+3XY = 2 + 3XY=2+3X. What does the coefficient 3 represent?**
A) The intercept
B) The slope of the line
C) The error term
D) The predicted Y-value
Answer: B) The slope of the line

37. **The Kruskal-Wallis test is a non-parametric alternative to:**
A) Chi-square test
B) One-way ANOVA
C) Paired t-test
D) Z-test
Answer: B) One-way ANOVA

38. **When conducting a two-tailed test at a 0.05 significance level, which p-value would lead you to reject the null hypothesis?**
A) 0.10
B) 0.06
C) 0.04
D) 0.15
Answer: C) 0.04

39. **The null hypothesis for a correlation test states that:**
A) There is no linear relationship between the variables
B) The variables are independent
C) The variables are normally distributed
D) There is a strong relationship between the variables
Answer: A) There is no linear relationship between the variables

40. **The Bonferroni correction is applied in order to:**
    A) Increase the p-value
    B) Control the Type I error rate when conducting multiple tests
    C) Control the Type II error rate
    D) Determine the median value
    Answer: B) Control the Type I error rate when conducting multiple tests

41. **In regression, heteroscedasticity refers to:**
    A) Constant variance of residuals
    B) Variable variance of residuals
    C) Correlation between predictors
    D) Normal distribution of residuals
    Answer: B) Variable variance of residuals

42. **Which test would you use to determine if two groups have the same variance?**
    A) T-test
    B) Chi-square test
    C) F-test
    D) ANOVA
    Answer: C) F-test

43. **If the p-value is 0.03 and the significance level is 0.05, then:**
    A) You fail to reject the null hypothesis
    B) You reject the null hypothesis
    C) You accept the null hypothesis
    D) The sample size is too small
    Answer: B) You reject the null hypothesis

44. **A histogram of data that has two peaks is called:**
    A) Symmetrical
    B) Bimodal
    C) Positively skewed
    D) Negatively skewed
    Answer: B) Bimodal

45. **If the correlation coefficient is -0.2, this indicates a:**
    A) Strong negative correlation
    B) Weak negative correlation
    C) Strong positive correlation
    D) Weak positive correlation
    Answer: B) Weak negative correlation

46. **The Mann-Whitney U test is a non-parametric test equivalent to:**
    A) One-way ANOVA
    B) Independent t-test
    C) Chi-square test
    D) Pearson correlation
    Answer: B) Independent t-test

47. **The variance of a dataset is 25. What is the standard deviation?**
    A) 5
    B) 25
    C) 10
    D) 50
    Answer: A) 5

48. **A paired t-test is used when:**
    A) Comparing the means of two independent samples
    B) Comparing the means of two related samples
    C) Analyzing categorical data
    D) Testing for independence
    Answer: B) Comparing the means of two related samples

49. **Which type of data cannot be analyzed with an independent t-test?**
    A) Continuous data
    B) Ordinal data
    C) Categorical data
    D) Normally distributed data
    Answer: C) Categorical data

50. **What is the primary purpose of calculating effect size in hypothesis testing?**
    A) To determine sample size
    B) To measure the strength of a relationship or difference
    C) To calculate the p-value
    D) To decide whether to use a one-tailed or two-tailed test
    Answer: B) To measure the strength of a relationship or difference

51. **The standard deviation of a sample is calculated to be 10. What is the variance?**
    A) 100
    B) 10
    C) 5
    D) 20
    Answer: A) 100

52. **Which measure is most appropriate to describe the center of a skewed distribution?**
    A) Mean
    B) Median
    C) Mode
    D) Range
    Answer: B) Median

53. **If the null hypothesis is rejected at a 0.01 significance level, it means:**
    A) The probability of making a Type I error is 0.01
    B) The probability of making a Type II error is 0.01
    C) There is no relationship between variables
    D) The test is inconclusive
    Answer: A) The probability of making a Type I error is 0.01

54. **Which statistical test would be most appropriate to compare the medians of two independent groups?**
    A) Independent t-test
    B) Mann-Whitney U test
    C) Paired t-test
    D) Chi-square test
    Answer: B) Mann-Whitney U test

55. **Which of the following is NOT a measure of dispersion?**
    A) Range
    B) Mean
    C) Variance
    D) Standard deviation
    Answer: B) Mean

56. **If the sample mean is equal to the population mean, then the bias of the estimator is:**
    A) Positive
    B) Negative
    C) Zero
    D) Undefined
    Answer: C) Zero

57. **The "Law of Large Numbers" states that as sample size increases:**
    A) Sample mean approaches the population mean
    B) Variance increases
    C) Mean remains constant
    D) Probability decreases
    Answer: A) Sample mean approaches the population mean

58. **A Type I error occurs when:**
    A) The null hypothesis is incorrectly rejected
    B) The null hypothesis is accepted
    C) The sample mean equals the population mean
    D) The sample size is too small
    Answer: A) The null hypothesis is incorrectly rejected

59. **When calculating a 99% confidence interval, the interval will be:**
    A) Wider than a 95% confidence interval
    B) Narrower than a 95% confidence interval
    C) Equal to a 95% confidence interval
    D) Impossible to determine
    Answer: A) Wider than a 95% confidence interval

60. **The Central Limit Theorem implies that as sample size increases, the sampling distribution of the sample mean approaches:**
    A) A skewed distribution
    B) A normal distribution
    C) A uniform distribution
    D) An exponential distribution
    Answer: B) A normal distribution

61. **In regression, which assumption states that the mean of the residuals is zero?**
    A) Linearity
    B) Homoscedasticity
    C) Independence
    D) Zero-mean assumption
    Answer: D) Zero-mean assumption

62. **The correlation coefficient ranges from:**
    A) 0 to 1
    B) -1 to 1
    C) -1 to 0
    D) 0 to ∞
    Answer: B) -1 to 1

63. **If the F-statistic in an ANOVA test is large, it indicates that:**
    A) The variance within groups is large
    B) The variance between groups is large compared to within groups
    C) The means of groups are equal
    D) The sample size is large
    Answer: B) The variance between groups is large compared to within groups

64. **A boxplot provides a visual representation of:**
    A) The mean and standard deviation
    B) The range and mean
    C) The median, quartiles, and outliers
    D) The variance and mode
    Answer: C) The median, quartiles, and outliers

65. **An outlier is typically defined as a data point that lies beyond:**
    A) One standard deviation from the mean
    B) Two standard deviations from the mean
    C) 1.5 times the interquartile range from the quartiles
    D) The range
    Answer: C) 1.5 times the interquartile range from the quartiles

66. **Which distribution is used to describe the number of occurrences of an event in a fixed interval of time or space?**
    A) Normal distribution
    B) Binomial distribution
    C) Poisson distribution
    D) Uniform distribution
    Answer: C) Poisson distribution

67. **The null hypothesis in a t-test for independent means states that:**
    A) The population means are different
    B) The sample means are equal
    C) The population means are equal
    D) The variances are equal
    Answer: C) The population means are equal

68. **What does a residual plot with a clear pattern indicate in regression analysis?**
    A) The model is correctly specified
    B) The model is not correctly specified
    C) The residuals are normally distributed
    D) The data is heteroscedastic
    Answer: B) The model is not correctly specified

69. **In a two-tailed test, if the calculated test statistic falls into the rejection region, we should:**
    A) Fail to reject the null hypothesis
    B) Reject the null hypothesis
    C) Decrease the sample size
    D) Increase the significance level
    Answer: B) Reject the null hypothesis

70. **A positive skewed distribution has a tail that is:**
    A) On the left
    B) On the right
    C) Centered
    D) Symmetric
    Answer: B) On the right

71. **When plotting a scatterplot of two variables, a curvilinear relationship indicates that:**
    A) The relationship between variables is linear
    B) The correlation coefficient is zero
    C) The relationship between variables is non-linear
    D) There is no relationship
    Answer: C) The relationship between variables is non-linear

72. **Which measure is unaffected by extreme values?**
    A) Mean
    B) Standard deviation
    C) Median
    D) Variance
    Answer: C) Median

73. **The degrees of freedom in a chi-square test are calculated by:**
    A) The sample size minus one
    B) The sum of the row totals
    C) (Number of rows - 1) × (Number of columns - 1)
    D) The square root of the mean
    Answer: C) (Number of rows - 1) × (Number of columns - 1)

74. **If the p-value is 0.07 and the significance level is 0.05, we should:**
    A) Reject the null hypothesis
    B) Fail to reject the null hypothesis
    C) Accept the alternative hypothesis
    D) Increase the significance level
    Answer: B) Fail to reject the null hypothesis

75. **When a sample size is increased, which of the following is true?**
    A) The standard error increases
    B) The standard error decreases
    C) The standard deviation increases
    D) The mean increases
    Answer: B) The standard error decreases

76. **Which test would you use to check if there is a significant association between two categorical variables?**
    A) T-test
    B) Chi-square test
    C) ANOVA
    D) Correlation
    Answer: B) Chi-square test

77. **What does a 95% confidence interval mean?**
    A) 95% of the data lies within the interval
    B) The interval contains the population parameter 95% of the time
    C) The sample mean falls within the interval 95% of the time
    D) The population mean is exactly at the midpoint of the interval
    Answer: B) The interval contains the population parameter 95% of the time

78. **When using the F-test in ANOVA, which of the following would indicate a significant difference between group means?**
    A) F-value close to zero
    B) Large F-value
    C) F-value equal to 1
    D) F-value less than 1
    Answer: B) Large F-value

79. **In regression analysis, which of the following describes multicollinearity?**
    A) When the independent variables are uncorrelated
    B) When the independent variables are highly correlated with each other
    C) When the residuals are heteroscedastic
    D) When the sample size is large
    Answer: B) When the independent variables are highly correlated with each other

80. **Which of the following is NOT an assumption of linear regression?**
    A) Linearity
    B) Homoscedasticity
    C) Independence
    D) Categorical dependent variable
    Answer: D) Categorical dependent variable

81. **If a distribution is symmetric, which of the following is true about the mean and median?**
    A) Mean > Median
    B) Mean < Median
    C) Mean = Median
    D) Mean > Mode
    Answer: C) Mean = Median

82. **A z-score of 2 indicates that a value is:**
    A) Two standard deviations below the mean
    B) Two units above the mean
    C) Two standard deviations above the mean
    D) Equal to the mean
    Answer: C) Two standard deviations above the mean

83. **The purpose of calculating the coefficient of determination (R^2) is to:**
    A) Test the significance of a correlation
    B) Measure the strength of association between variables
    C) Assess the variability in the dependent variable explained by the independent variable
    D) Determine the probability of an event
    Answer: C) Assess the variability in the dependent variable explained by the independent variable

84. **The degrees of freedom for a t-distribution with a sample size of 25 is:**
    A) 24
    B) 25
    C) 26
    D) 50
    Answer: A) 24

85. **In a hypothesis test, the critical value is:**
    A) The observed test statistic
    B) The point that separates the rejection region from the non-rejection region
    C) Always equal to 0.05
    D) The probability of a Type II error
    Answer: B) The point that separates the rejection region from the non-rejection region

86. **If the p-value in a test is greater than the significance level, you should:**
    A) Reject the null hypothesis
    B) Accept the alternative hypothesis
    C) Fail to reject the null hypothesis
    D) Reduce the sample size
    Answer: C) Fail to reject the null hypothesis

87. **Which measure would most likely be used to describe the spread of a normally distributed dataset?**
    A) Mode
    B) Range
    C) Standard deviation
    D) Median
    Answer: C) Standard deviation

88. **In simple linear regression, the slope represents:**
    A) The intercept of the line
    B) The change in the dependent variable for a one-unit change in the independent variable
    C) The residual of the line
    D) The predicted value
    Answer: B) The change in the dependent variable for a one-unit change in the independent variable

89. **The power of a statistical test is the probability of:**
    A) Rejecting the null hypothesis when it is false
    B) Accepting the null hypothesis when it is false
    C) Rejecting the null hypothesis when it is true
    D) Making a Type I error
    Answer: A) Rejecting the null hypothesis when it is false

90. **A high p-value indicates that:**
    A) The observed result is statistically significant
    B) The null hypothesis is likely true
    C) There is a high probability of a Type I error
    D) The sample size is too large
    Answer: B) The null hypothesis is likely true

91. **In a normal distribution, the mean and median are:**
    A) The same
    B) Different
    C) Unrelated
    D) Negative
    Answer: A) The same

92. **Which test would be suitable for comparing the means of two related samples?**
    A) Independent t-test
    B) Paired t-test
    C) Chi-square test
    D) ANOVA
    Answer: B) Paired t-test

93. **Which term describes the likelihood of observing a sample statistic as extreme as the one observed under the null hypothesis?**
    A) Significance level
    B) Confidence interval
    C) P-value
    D) Power
    Answer: C) P-value

94. **In hypothesis testing, failing to reject the null hypothesis means that:**
    A) The null hypothesis is proven to be true
    B) There is not enough evidence to support the alternative hypothesis
    C) The sample is biased
    D) The significance level is too high
    Answer: B) There is not enough evidence to support the alternative hypothesis

95. **Which of the following distributions is defined by two parameters, mean and standard deviation?**
    A) Normal distribution
    B) Poisson distribution
    C) Exponential distribution
    D) Binomial distribution
    Answer: A) Normal distribution

96. **Which statistical test is used to assess whether there is a significant difference between expected and observed frequencies?**
    A) T-test
    B) Z-test
    C) Chi-square test
    D) F-test
    Answer: C) Chi-square test

97. **If the 95% confidence interval for a mean difference includes 0, this suggests:**
    A) A significant difference exists
    B) A significant difference does not exist
    C) The sample is biased
    D) The sample size is too small
    Answer: B) A significant difference does not exist

98. **The sampling distribution of a statistic is the distribution of:**
    A) The sample statistic calculated from different samples
    B) Individual observations in the sample
    C) The population parameters
    D) The p-values from multiple tests
    Answer: A) The sample statistic calculated from different samples

99. **Which of the following measures is most appropriate for describing the spread of a dataset with outliers?**
    A) Mean
    B) Median
    C) Range
    D) Interquartile range (IQR)
    Answer: D) Interquartile range (IQR)

100. **A scatterplot is most useful for visualizing the relationship between:**
    A) Two categorical variables
    B) One categorical and one continuous variable
    C) Two continuous variables
    D) Two independent samples
    Answer: C) Two continuous variables

HARD-LEVEL

1. **Which type of regression should be used when the dependent variable is categorical with more than two levels?**
   A) Simple linear regression
   B) Logistic regression
   C) Polynomial regression
   D) Multinomial logistic regression
   Answer: D) Multinomial logistic regression

2. **In a time series analysis, which process describes a series that has constant mean and variance over time?**
   A) Non-stationary
   B) Stationary
   C) Seasonal
   D) Random walk
   Answer: B) Stationary

3. **The Durbin-Watson statistic tests for:**
   A) Multicollinearity
   B) Homoscedasticity
   C) Serial correlation
   D) Normality of residuals
   Answer: C) Serial correlation

4. **Which of the following methods is used to estimate parameters in generalized linear models?**
   A) Ordinary Least Squares
   B) Ridge regression
   C) Maximum Likelihood Estimation
   D) Principal Component Analysis
   Answer: C) Maximum Likelihood Estimation

5. **In hypothesis testing, if we want to decrease the probability of a Type I error, we should:**
   A) Increase the significance level
   B) Decrease the significance level
   C) Increase the sample size
   D) Decrease the sample size
   Answer: B) Decrease the significance level

6. **In an ANOVA test, what does it mean if the between-group variance is greater than the within-group variance?**
   A) There is likely no significant difference between group means
   B) The group means are likely significantly different
   C) The test cannot be performed
   D) The sample size is insufficient
   Answer: B) The group means are likely significantly different

7. **Which of the following would be used to model the relationship between a categorical outcome with more than two levels and multiple predictors?**
   A) Simple logistic regression
   B) Linear regression
   C) Multinomial regression
   D) Poisson regression
   Answer: C) Multinomial regression

8. **A significant interaction effect in a factorial ANOVA indicates:**
   A) The main effects are also significant

B) There is a combined effect of the factors on the dependent variable
C) The factors are independent
D) There is no effect of the factors on the dependent variable
Answer: B) There is a combined effect of the factors on the dependent variable

9. **In hypothesis testing, the probability of a Type II error increases as:**
A) Sample size increases
B) Sample size decreases
C) Significance level decreases
D) Effect size increases
Answer: B) Sample size decreases

10. **The Akaike Information Criterion (AIC) is used to:**
A) Test the significance of the model
B) Determine the likelihood of Type I error
C) Compare models with different numbers of predictors
D) Assess the normality of residuals
Answer: C) Compare models with different numbers of predictors

11. **When testing for autocorrelation in a regression model, which statistic is commonly used?**
A) Pearson correlation
B) Variance inflation factor
C) Durbin-Watson statistic
D) T-test
Answer: C) Durbin-Watson statistic

12. **Which technique is suitable for reducing the number of correlated predictors in a model?**
A) Logistic regression
B) Factor analysis
C) Poisson regression
D) Chi-square test
Answer: B) Factor analysis

13. **The coefficient of determination (R^2) in multiple regression can be adjusted by:**
A) Increasing sample size
B) Adjusting for the number of predictors
C) Normalizing the data
D) Removing multicollinearity
Answer: B) Adjusting for the number of predictors

14. **Which of the following is not a requirement for valid results in the chi-square test?**
A) Large sample size
B) Expected frequencies greater than 5 in each cell
C) Normally distributed variables
D) Independence of observations
Answer: C) Normally distributed variables

15. **The purpose of bootstrapping in statistics is to:**
A) Increase the power of the test
B) Estimate the sampling distribution of a statistic
C) Reduce bias in small samples
D) Test for normality
Answer: B) Estimate the sampling distribution of a statistic

16. **Which assumption is violated if residuals from a regression model show a funnel shape when plotted against predicted values?**

A) Linearity
B) Homoscedasticity
C) Normality
D) Independence
Answer: B) Homoscedasticity

17. **The likelihood ratio test is commonly used to compare:**
A) One model with no predictors
B) Nested models
C) Non-nested models
D) The significance of coefficients in one model
Answer: B) Nested models

18. **If the residuals in a regression model are not normally distributed, this most likely affects:**
A) The interpretation of coefficients
B) The calculation of standard errors
C) The distribution of predictors
D) The sample size requirement
Answer: B) The calculation of standard errors

19. **A nonparametric test that compares the ranks of two independent samples is:**
A) Kruskal-Wallis test
B) Mann-Whitney U test
C) Chi-square test
D) Paired t-test
Answer: B) Mann-Whitney U test

20. **Multicollinearity in a multiple regression model is often assessed by:**
A) The residual plot
B) Durbin-Watson statistic
C) Variance Inflation Factor (VIF)
D) T-statistics of coefficients
Answer: C) Variance Inflation Factor (VIF)

21. **Which of the following is true of Bayesian inference compared to frequentist inference?**
A) It does not require prior knowledge
B) It combines prior knowledge with observed data
C) It relies solely on sampling distributions
D) It uses the p-value as the primary measure of evidence
Answer: B) It combines prior knowledge with observed data

22. **In logistic regression, the odds ratio is used to interpret:**
A) The change in odds for a one-unit increase in the predictor
B) The change in the probability of the predictor
C) The variance explained by the model
D) The goodness-of-fit of the model
Answer: A) The change in odds for a one-unit increase in the predictor

23. **Which test would be suitable to assess if the variances of two or more populations are equal?**
A) T-test
B) ANOVA
C) Levene's test
D) Kruskal-Wallis test
Answer: C) Levene's test

24. **The power of a test is defined as:**
    A) The probability of rejecting the null hypothesis when it is true
    B) The probability of rejecting the null hypothesis when it is false
    C) The probability of a Type I error
    D) The probability of making a Type II error
    Answer: B) The probability of rejecting the null hypothesis when it is false

26. **In a logistic regression model, the logit function is used to transform:**
    A) The predictor variable to a binary outcome
    B) The probability of the outcome to a continuous scale
    C) The predictor variable to an odds ratio
    D) The log-likelihood to a normal distribution
    Answer: B) The probability of the outcome to a continuous scale

27. **Which of the following is a primary difference between ANCOVA and ANOVA?**
    A) ANCOVA includes covariates to adjust the dependent variable
    B) ANOVA is only used for categorical data
    C) ANCOVA assumes normality of covariates
    D) ANOVA includes covariates as predictors
    Answer: A) ANCOVA includes covariates to adjust the dependent variable

28. **Which of the following statements about principal component analysis (PCA) is true?**
    A) PCA maximizes the explained variance by orthogonal components
    B) PCA is primarily used for hypothesis testing
    C) PCA can only be used with categorical data
    D) PCA does not require any assumptions about the data
    Answer: A) PCA maximizes the explained variance by orthogonal components

29. **In a multiple regression model, if two predictor variables have a high Variance Inflation Factor (VIF), this indicates:**
    A) Multicollinearity is present
    B) The predictors are irrelevant
    C) The model has high predictive power
    D) The response variable has low variance
    Answer: A) Multicollinearity is present

30. **Which assumption is violated if the residuals in a regression model show a pattern when plotted against fitted values?**
    A) Linearity
    B) Normality
    C) Homoscedasticity
    D) Independence
    Answer: A) Linearity

31. **When conducting a hypothesis test, a researcher decides to use a two-tailed test instead of a one-tailed test. This decision will:**
    A) Increase the likelihood of a Type I error
    B) Increase the p-value
    C) Decrease the critical value for rejection
    D) Increase the degrees of freedom
    Answer: B) Increase the p-value

32. **Which of the following is a nonparametric alternative to the one-way ANOVA test?**
    A) Kruskal-Wallis test
    B) Mann-Whitney U test
    C) Paired t-test

D) Fisher's Exact test
Answer: A) Kruskal-Wallis test

33. **In hypothesis testing, a one-sided hypothesis test has greater power than a two-sided test because:**
A) It uses a larger sample size
B) It has a larger rejection region for the same significance level
C) It has a smaller rejection region
D) It has a higher probability of Type II error
Answer: B) It has a larger rejection region for the same significance level

34. **Which of the following is true for residuals in a well-fitted linear regression model?**
A) They are uniformly distributed
B) They have constant variance
C) They are correlated with predictors
D) They have a mean greater than zero
Answer: B) They have constant variance

35. **When interpreting the results of a logistic regression model, a coefficient of 0.5 for a predictor indicates:**
A) A 50% increase in the odds of the outcome for a one-unit increase in the predictor
B) The predictor does not significantly affect the outcome
C) The log odds of the outcome increases by 0.5 for a one-unit increase in the predictor
D) The odds of the outcome increase by 0.5
Answer: C) The log odds of the outcome increases by 0.5 for a one-unit increase in the predictor

36. **The term "heteroscedasticity" in regression analysis refers to:**
A) The non-normality of residuals
B) The presence of correlated residuals
C) Variance of residuals that changes with fitted values
D) Independent observations in the dataset
Answer: C) Variance of residuals that changes with fitted values

37. **Which of the following is a key assumption for valid application of the Central Limit Theorem?**
A) The population distribution is normal
B) The sample size is large enough
C) The sample mean equals the population mean
D) The sample variance equals the population variance
Answer: B) The sample size is large enough

38. **Which method is suitable for comparing the variances of more than two groups?**
A) T-test
B) Kruskal-Wallis test
C) Bartlett's test
D) Mann-Whitney U test
Answer: C) Bartlett's test

39. **The result of an F-test in ANOVA is significant, which implies:**
A) The variances of the groups are equal
B) At least one group mean is significantly different
C) All group means are significantly different
D) The test cannot be interpreted
Answer: B) At least one group mean is significantly different

40. **If a residual plot for a regression model shows a systematic curve, this suggests a violation of which assumption?**
A) Normality
B) Linearity

C) Homoscedasticity
D) Independence
Answer: B) Linearity

41. **The use of a Bonferroni correction is primarily intended to:**
A) Increase statistical power
B) Control for Type II error
C) Adjust for multiple comparisons
D) Reduce multicollinearity
Answer: C) Adjust for multiple comparisons

42. **In a normal distribution, which percentile is closest to the mean?**
A) 5th percentile
B) 50th percentile
C) 75th percentile
D) 95th percentile
Answer: B) 50th percentile

43. **A mixed ANOVA design is characterized by:**
A) Multiple dependent variables
B) A mix of random and fixed effects
C) Both within-subjects and between-subjects factors
D) A combination of ANOVA and ANCOVA elements
Answer: C) Both within-subjects and between-subjects factors

44. **Which statistical test compares the means of two paired samples?**
A) Independent t-test
B) Paired t-test
C) Chi-square test
D) Mann-Whitney U test
Answer: B) Paired t-test

45. **The Box-Cox transformation is used to:**
A) Test for independence of observations
B) Normalize the distribution of residuals
C) Adjust for Type I error
D) Increase multicollinearity
Answer: B) Normalize the distribution of residuals

46. **The null hypothesis in an F-test for overall significance in regression states that:**
A) All predictors have significant effects
B) At least one predictor has no effect on the outcome
C) All coefficients are zero
D) The intercept is zero
Answer: C) All coefficients are zero

47. **In multiple regression, the adjusted R-squared value:**
A) Always increases as more predictors are added
B) Decreases as predictors with little explanatory power are added
C) Is the square root of the R-squared
D) Does not account for the number of predictors
Answer: B) Decreases as predictors with little explanatory power are added

48. **A partial correlation measures:**
A) The correlation between two variables while controlling for a third variable
B) The sum of squared deviations from the mean

C) The total variance explained by a variable
D) The strength of a single predictor in isolation
Answer: A) The correlation between two variables while controlling for a third variable

49. **Which distribution is used as the basis for the chi-square goodness-of-fit test?**
A) T-distribution
B) F-distribution
C) Normal distribution
D) Chi-square distribution
Answer: D) Chi-square distribution

50. **In an experiment, the p-value of 0.03 suggests that:**
A) There is a 3% chance that the results are due to random variation
B) The null hypothesis should be rejected at the 0.05 significance level
C) The null hypothesis is proven to be false
D) The alternative hypothesis has a 3% probability of being true
Answer: B) The null hypothesis should be rejected at the 0.05 significance level

51. **Which of the following is true about residuals in a multiple regression model?**
A) They must be normally distributed for the predictors to be significant
B) They must be independent and have constant variance
C) They must be positively correlated with predictors
D) They must have an R-squared value greater than 0.5
Answer: B) They must be independent and have constant variance

52. **In hypothesis testing, the term "power" refers to the probability of:**
A) Rejecting the null hypothesis when it is true
B) Rejecting the null hypothesis when it is false
C) Accepting the null hypothesis when it is false
D) Accepting the alternative hypothesis when it is true
Answer: B) Rejecting the null hypothesis when it is false

53. **Which statistical test would be most appropriate for determining if there is an association between two categorical variables?**
A) ANOVA
B) T-test
C) Chi-square test
D) F-test
Answer: C) Chi-square test

54. **In time series analysis, what does ARIMA stand for?**
A) Autoregressive Integrated Moving Average
B) Autoregressive Independent Moving Analysis
C) Average Integrated Median Analysis
D) Advanced Independent Moving Average
Answer: A) Autoregressive Integrated Moving Average

55. **Which of the following is an assumption of the two-way ANOVA without replication?**
A) Homogeneity of variances
B) Covariate inclusion
C) Heteroscedasticity
D) Dependence of observations
Answer: A) Homogeneity of variances

56. **A t-test for dependent means is commonly used for which of the following situations?**
A) Comparing the means of two unrelated groups

B) Comparing the means of the same group measured twice
C) Comparing the variances of two independent groups
D) Comparing proportions across groups
Answer: B) Comparing the means of the same group measured twice

57. **When performing a linear regression analysis, multicollinearity is problematic because it can:**
A) Decrease the R-squared value
B) Inflate the variance of the estimated coefficients
C) Cause heteroscedasticity in residuals
D) Violate the assumption of linearity
Answer: B) Inflate the variance of the estimated coefficients

58. **If a confidence interval for a mean difference includes zero, we would conclude that:**
A) The difference is statistically significant
B) The difference is not statistically significant
C) The sample size is too small
D) The variance of the difference is zero
Answer: B) The difference is not statistically significant

59. **In a multiple regression model, an interaction term between two predictors indicates:**
A) Both predictors have independent effects
B) The effect of one predictor depends on the level of the other
C) The predictors have collinearity issues
D) The model has a random effect component
Answer: B) The effect of one predictor depends on the level of the other

60. **A sampling distribution refers to:**
A) The distribution of a sample from a population
B) The distribution of a statistic over many samples
C) The spread of individual data points in a sample
D) The proportion of different samples in a population
Answer: B) The distribution of a statistic over many samples

61. **A Type I error occurs when:**
A) A true null hypothesis is rejected
B) A false null hypothesis is accepted
C) A true null hypothesis is accepted
D) A true alternative hypothesis is rejected
Answer: A) A true null hypothesis is rejected

62. **A significant Levene's test in ANOVA suggests that:**
A) The groups have equal variances
B) The groups have unequal variances
C) The means of the groups are equal
D) There is no effect of the predictor on the outcome
Answer: B) The groups have unequal variances

63. **Which test would you use to compare the means of three or more related groups?**
A) One-way ANOVA
B) Two-way ANOVA
C) Repeated measures ANOVA
D) Chi-square test
Answer: C) Repeated measures ANOVA

64. **Which of the following is a non-parametric test for comparing more than two independent groups?**
A) Paired t-test

B) Kruskal-Wallis test
C) Friedman test
D) One-way ANOVA
Answer: B) Kruskal-Wallis test

65. **The Shapiro-Wilk test is commonly used to assess:**
A) The equality of variances
B) The normality of a data distribution
C) The independence of observations
D) The homogeneity of regression slopes
Answer: B) The normality of a data distribution

66. **The term "homoscedasticity" in regression analysis refers to:**
A) The residuals having constant variance across levels of a predictor
B) Residuals being normally distributed
C) The response variable having constant mean
D) Residuals increasing as the predictor increases
Answer: A) The residuals having constant variance across levels of a predictor

67. **The difference between parametric and non-parametric tests primarily relates to:**
A) Sample size requirements
B) The types of hypotheses tested
C) Assumptions about the population distribution
D) The number of predictor variables
Answer: C) Assumptions about the population distribution

68. **A correlation coefficient of -0.85 indicates:**
A) A weak positive relationship
B) A strong negative relationship
C) A moderate positive relationship
D) No relationship
Answer: B) A strong negative relationship

69. **Which of the following methods is suitable for identifying outliers in a dataset?**
A) Calculating the mean
B) Creating a scatterplot
C) Checking standardized residuals
D) Performing a t-test
Answer: C) Checking standardized residuals

70. **In logistic regression, multicollinearity can be problematic because it:**
A) Causes model overfitting
B) Inflates the variance of the estimated coefficients
C) Violates the assumption of homoscedasticity
D) Reduces the sample size
Answer: B) Inflates the variance of the estimated coefficients

71. **Which of the following methods can handle multicollinearity in regression analysis?**
A) Stepwise regression
B) Principal Component Analysis (PCA)
C) T-test
D) Z-test
Answer: B) Principal Component Analysis (PCA)

72. **In hypothesis testing, if the p-value is greater than the significance level, we:**
A) Accept the null hypothesis

B) Reject the null hypothesis
C) Fail to reject the null hypothesis
D) Conclude the alternative hypothesis is true
Answer: C) Fail to reject the null hypothesis

73. **When conducting a two-way ANOVA, an interaction effect means that:**
A) The effects of one factor are consistent across all levels of the other factor
B) The effects of one factor vary depending on the level of the other factor
C) Both factors are irrelevant to the dependent variable
D) There is no main effect for either factor
Answer: B) The effects of one factor vary depending on the level of the other factor

74. **The purpose of using stratified sampling is to:**
A) Reduce the sample size
B) Achieve a greater degree of representation of subgroups
C) Introduce variability into the sample
D) Maximize sampling error
Answer: B) Achieve a greater degree of representation of subgroups

75. **A high F-statistic in ANOVA suggests that:**
A) The null hypothesis should be accepted
B) The means of the groups are significantly different
C) The sample size is too small
D) The data violates assumptions of normality
Answer: B) The means of the groups are significantly different

76. **The Durbin-Watson statistic is used to test for:**
A) Homoscedasticity
B) Normality of residuals
C) Autocorrelation in residuals
D) Multicollinearity in predictors
Answer: C) Autocorrelation in residuals

77. **The F-distribution is used primarily in which type of statistical test?**
A) T-tests
B) Chi-square tests
C) ANOVA and regression tests
D) Mann-Whitney U tests
Answer: C) ANOVA and regression tests

78. **In a normal distribution, approximately what percentage of data falls within three standard deviations from the mean?**
A) 50%
B) 95%
C) 99%
D) 99.7%
Answer: D) 99.7%

79. **The Wilcoxon signed-rank test is a non-parametric alternative to which test?**
A) Paired t-test
B) Independent t-test
C) Chi-square test
D) One-way ANOVA
Answer: A) Paired t-test

80. **Which of the following is used to assess goodness-of-fit in logistic regression?**
    A) Adjusted R-squared
    B) Akaike Information Criterion (AIC)
    C) Pearson correlation coefficient
    D) Mann-Whitney U test
    Answer: B) Akaike Information Criterion (AIC)

81. **A Type II error occurs when:**
    A) A false null hypothesis is rejected
    B) A true null hypothesis is accepted
    C) A true null hypothesis is rejected
    D) A false null hypothesis is accepted
    Answer: D) A false null hypothesis is accepted

82. **In regression analysis, the term "variance inflation factor" (VIF) is used to measure:**
    A) The increase in R-squared when a predictor is added
    B) The increase in the standard error of a predictor due to multicollinearity
    C) The goodness-of-fit of the model
    D) The residual standard error
    Answer: B) The increase in the standard error of a predictor due to multicollinearity

83. **The concept of “degrees of freedom” in statistics generally refers to:**
    A) The sample size divided by the number of parameters
    B) The amount of information available for estimating a parameter
    C) The number of observations in a sample
    D) The level of confidence in the model
    Answer: B) The amount of information available for estimating a parameter

84. **Which of the following distributions is typically used to model count data?**
    A) Normal distribution
    B) Binomial distribution
    C) Poisson distribution
    D) T-distribution
    Answer: C) Poisson distribution

85. **In the context of hypothesis testing, the p-value represents:**
    A) The probability of rejecting the null hypothesis
    B) The probability of obtaining the observed result, or something more extreme, if the null hypothesis is true
    C) The probability that the null hypothesis is true
    D) The probability that the alternative hypothesis is true
    Answer: B) The probability of obtaining the observed result, or something more extreme, if the null hypothesis is true

86. **The null hypothesis in a chi-square test for independence states that:**
    A) The observed frequencies equal the expected frequencies
    B) The variables are independent of each other
    C) The variables are dependent on each other
    D) The sample sizes are equal
    Answer: B) The variables are independent of each other

87. **The term "sample variance" refers to:**
    A) The average of all observed values
    B) The sum of squares of each observed value
    C) The average of the squared deviations from the mean
    D) The square root of the standard deviation
    Answer: C) The average of the squared deviations from the mean

88. **Which of the following is a characteristic of the chi-square distribution?**
    A) It is symmetric
    B) It is positively skewed
    C) It is normally distributed
    D) It has both positive and negative values
    Answer: B) It is positively skewed

89. **A logistic regression model is typically used when:**
    A) The outcome variable is continuous
    B) The predictor variable is binary
    C) The outcome variable is binary
    D) The predictors are normally distributed
    Answer: C) The outcome variable is binary

90. **In the context of hypothesis testing, "alpha" (α) represents:**
    A) The probability of making a Type II error
    B) The significance level, or the probability of a Type I error
    C) The effect size
    D) The power of the test
    Answer: B) The significance level, or the probability of a Type I error

91. **Which of the following assumptions is violated if residuals are correlated with predictor variables in a linear regression?**
    A) Linearity
    B) Independence of errors
    C) Homoscedasticity
    D) Normality of residuals
    Answer: B) Independence of errors

92. **A QQ plot is used to assess:**
    A) The independence of observations
    B) The equality of variances
    C) The distributional assumptions of residuals
    D) The sample size adequacy
    Answer: C) The distributional assumptions of residuals

93. **In a multiple regression model, the adjusted R-squared value:**
    A) Is always greater than R-squared
    B) Adjusts for the number of predictors in the model
    C) Increases as more predictors are added
    D) Is the square of the R-squared value
    Answer: B) Adjusts for the number of predictors in the model

94. **When using a one-tailed test, the rejection region is:**
    A) On both tails of the distribution
    B) On only one tail of the distribution
    C) Based on the standard deviation alone
    D) Determined by the sample size
    Answer: B) On only one tail of the distribution

95. **Multicollinearity in a regression model occurs when:**
    A) The predictors are too weak to explain variance in the response
    B) Two or more predictors are highly correlated with each other
    C) The residuals are correlated
    D) The predictors are unrelated to each other
    Answer: B) Two or more predictors are highly correlated with each other

96. **The Kruskal-Wallis test is a non-parametric alternative to which of the following tests?**
    A) Independent t-test
    B) One-way ANOVA
    C) Paired t-test
    D) Pearson correlation
    Answer: B) One-way ANOVA

97. **Which of the following can be concluded if the p-value is less than 0.05 in a hypothesis test?**
    A) The null hypothesis is true
    B) The result is statistically significant
    C) There is a 5% chance the null hypothesis is false
    D) The sample size is too large
    Answer: B) The result is statistically significant

98. **The effect size in hypothesis testing helps to:**
    A) Increase the sample size
    B) Measure the strength or magnitude of the effect
    C) Increase the p-value
    D) Control for Type II errors
    Answer: B) Measure the strength or magnitude of the effect

99. **The assumption of homogeneity of variance is assessed by:**
    A) Bartlett's test or Levene's test
    B) T-test
    C) Paired sample t-test
    D) Pearson correlation
    Answer: A) Bartlett's test or Levene's test

100. **In ANOVA, the total variance is partitioned into which components?**
    A) Explained and unexplained variance
    B) Population and sample variance
    C) Within-group and between-group variance
    D) Correlation and regression variance
    Answer: C) Within-group and between-group variance