

Quiz - 3



Course: Biostatistics | Instructor: Dr. Gaurav

Duration: 60 minutes | Each question weights Two Marks

Wrong answer have negative marking -1

Name :

Roll no :

1) What graphical method is best suited for comparing distributions of different groups along a continuous variable?

A) Histogram **B) Box plot** C) Scatter plot D) Q-Q plot E) Pie chart

2) What statistic is used to measure the strength and direction of the linear relationship between two variables?

A) Mean B) Median C) Mode **D) Correlation coefficient** E) Standard deviation

3) Which of the following correlation coefficients indicates the strongest linear relationship?

A) 0.2 B) -0.6 C) 0 **D) 0.8** E) -0.1

4) Which sampling method involves selecting every nth element from the population to be included in the sample?

A) Simple random sampling B) Stratified sampling **C) Systematic sampling** D) Cluster sampling E) Convenience sampling

5) What statistical test is used to determine whether a sample comes from a normally distributed population?

A) ANOVA B) Chi-square test C) T-test **D) Shapiro-Wilk test** E) Kruskal-Wallis test

6) Which plot is used to visually assess if two data sets come from populations with the same distribution?

A) Box plot B) Scatter plot C) Histogram **D) Q-Q plot** E) Pie chart

7) What does the Central Limit Theorem state?

A) The mean of a sufficiently large sample approaches a normal distribution

B) The median of a sample will be normally distributed

C) The mode of a sample will be normally distributed

D) The range of a sample will be normally distributed

E) All statistical distributions converge to normal over time

8) Which test is used to determine if there is a significant difference between the means of two independent samples?

A) Paired t-test B) Mann-Whitney U test **C) Independent samples t-test** D) Wilcoxon signed-rank test E) Kruskal-Wallis test

16) In a time series analysis, a researcher applies a moving average method to smooth out fluctuations in the data. After performing data smoothing, the researcher notices that the resulting smoothed curve lags behind the original time series data. What could be a potential drawback or limitation of using the moving average method for data smoothing?

- A) It introduces bias into the smoothed data.
- B) It amplifies noise in the data.
- C) It may lead to overfitting of the model.
- D) It may cause the smoothed curve to lag behind the original data.**
- E) It requires setting a fixed window size, which may not capture short-term fluctuations adequately.

9) In a regression analysis, if the Shapiro-Wilk test indicates non-normality in the residuals with a significance level of 0.01, what would be the most appropriate action to address this issue?

- A) Proceed with the analysis without any adjustments, as regression is robust to violations of normality assumptions.
- B) Transform the response variable using a suitable data transformation technique.**
- C) Apply a non-parametric regression method instead of traditional linear regression.
- D) Exclude outliers from the dataset to improve normality.
- E) Ignore the Shapiro-Wilk test results and interpret the regression model cautiously.

10) In survival analysis, what does the Kaplan-Meier curve represent?

- A) Probability of survival over time**
- B) Cumulative hazard over time
- C) Probability of failure over time
- D) Cumulative risk over time
- E) None of the above

11) What does the ROC curve visualize?

- A) The relationship between sensitivity and specificity**
- B) The distribution of data points in a dataset
- C) The performance of a classifier at different thresholds
- D) The variability of a sample
- E) The distribution of errors in a regression model

12) What is the critical value used for in hypothesis testing?

- A) To determine the standard deviation of the sample
- B) To establish the confidence interval
- C) To define the rejection region for a hypothesis test**
- D) To calculate the p-value
- E) To determine the degrees of freedom

13) What is the main purpose of attribute construction in data preprocessing?

- A) To reduce the dimensionality of the dataset
- B) To remove outliers from the dataset
- C) To transform categorical variables into numerical ones
- D) To create new variables from existing ones**
- E) To standardize the scale of variables

14) What does the Shapiro-Wilk test assess?

- A) Normality of data distribution**
- B) Homogeneity of variance
- C) Linearity of relationship
- D) Independence of observations
- E) None of the above

16) In a time series analysis, a researcher applies a moving average method to smooth out fluctuations in the data. After performing data smoothing, the researcher notices that the resulting smoothed curve lags behind the original time series data. What could be a potential drawback or limitation of using the moving average method for data smoothing?

- A) It introduces bias into the smoothed data.
- B) It amplifies noise in the data.
- C) It may lead to overfitting of the model.
- D) It may cause the smoothed curve to lag behind the original data.**
- E) It requires setting a fixed window size, which may not capture short-term fluctuations adequately.

15) A researcher is conducting a study on customer satisfaction at a large shopping mall. The researcher wants to ensure that the sample is representative of the entire population of shoppers. Which probability sampling method would be most appropriate for this study?

- A) Simple random sampling
- B) Convenience sampling
- C) Systematic sampling
- D) Stratified sampling**
- E) Cluster sampling

17) In data discretization, what is the process of converting continuous variables into categorical ones?

- A) Binning**
- B) Scaling
- C) Normalizing
- D) Transforming
- E) Centering

18) In the context of data preprocessing, what is the primary objective of data generalization and data discretization?

- A) To increase the complexity of the dataset for more detailed analysis
- B) To reduce the dimensionality of the dataset by aggregating similar data points**
- C) To introduce outliers into the dataset to enhance variability
- D) To remove missing values from the dataset to improve data quality
- E) To standardize the scale of variables for better comparability

19) A researcher conducts a study to determine whether there is a significant difference in mean test scores between two groups. After performing a two-sample t-test, the researcher obtains a test statistic of 2.75 and a p-value of 0.009. What conclusion can be drawn from these results?

- A) There is insufficient evidence to conclude a difference in mean test scores between the two groups.
- B) There is evidence to support a difference in mean test scores between the two groups at a significance level of 0.05.**
- C) The test statistic indicates the magnitude of the difference in mean test scores.
- D) The p-value represents the probability of observing the data if the null hypothesis is true.
- E) The p-value indicates the probability of making a Type I error.

20) How does the variance (σ^2) affect the goodness of fit of a regression line?

- A) Higher variance leads to a better fit
- B) Lower variance leads to a better fit**
- C) Variance does not affect the fit of the regression line

16) In a time series analysis, a researcher applies a moving average method to smooth out fluctuations in the data. After performing data smoothing, the researcher notices that the resulting smoothed curve lags behind the original time series data. What could be a potential drawback or limitation of using the moving average method for data smoothing?

- A) It introduces bias into the smoothed data.
- B) It amplifies noise in the data.
- C) It may lead to overfitting of the model.
- D) It may cause the smoothed curve to lag behind the original data.**
- E) It requires setting a fixed window size, which may not capture short-term fluctuations adequately.

- D) Variance affects the slope of the regression line²
- E) Variance affects the intercept of the regression line

21) What does a negative β coefficient signify in regression analysis?

- A) There is no relationship between the variables
- B) There is a positive relationship between the variables
- C) As the independent variable increases, the dependent variable decreases**
- D) As the independent variable increases, the dependent variable also increases
- E) The regression line is vertical

22) In a survival analysis, what does the survival curve represent?

- A) The probability of experiencing an event over time
- B) The probability of surviving beyond a certain time point**
- C) The cumulative hazard over time
- D) The cumulative risk over time
- E) The probability density function

23) In a study examining the survival rates of patients diagnosed with a specific disease, the Kaplan-Meier estimator is used to estimate the survival function. After analyzing the data, the estimated survival probabilities at 1 year, 2 years, and 3 years are found to be 0.85, 0.70, and 0.55, respectively. What do these survival probabilities represent?

- A) The probability of experiencing the event of interest at each time point.
- B) The proportion of patients still alive at each time point after diagnosis.**
- C) The hazard ratio comparing survival rates between two treatment groups.
- D) The median survival time of patients diagnosed with the disease.
- E) The probability of being censored at each time point in the study.

24) In a hypothesis test comparing the mean IQ scores of two different populations, a researcher obtains a p-value of 0.03. What can be concluded from this result?

- A) There is sufficient evidence to reject the null hypothesis at a significance level of 0.05.**
- B) There is insufficient evidence to reject the null hypothesis at a significance level of 0.05.
- C) The populations have identical mean IQ scores.
- D) The p-value indicates the probability of making a Type I error.
- E) The p-value indicates the probability of making a Type II error.

16) In a time series analysis, a researcher applies a moving average method to smooth out fluctuations in the data. After performing data smoothing, the researcher notices that the resulting smoothed curve lags behind the original time series data. What could be a potential drawback or limitation of using the moving average method for data smoothing?

- A) It introduces bias into the smoothed data.
- B) It amplifies noise in the data.
- C) It may lead to overfitting of the model.
- D) It may cause the smoothed curve to lag behind the original data.**
- E) It requires setting a fixed window size, which may not capture short-term fluctuations adequately.

25) In regression analysis, if the p-value associated with a predictor variable is less than the chosen significance level (e.g., $\alpha = 0.05$), what does it indicate?

- A) The predictor variable has a significant effect on the response variable**
- B) The predictor variable has no effect on the response variable
- C) There is multicollinearity present among the predictor variables
- D) The regression model is overfitting the data
- E) The residuals are normally distributed