

Biostatistics (BIO545)

Mid Semester Exam

Duration: 60 min

Each question carries 4 marks

Biostatistics - Mixed MCQ and Short Answer Question Exam

Instructions: This exam consists of multiple-choice questions (MCQs) and short-answer questions (SAQs). For the MCQs, choose the best answer and provide a concise explanation (1-2 sentences) supporting your choice.

1. You are comparing the effectiveness of two drugs for lowering cholesterol. Drug A lowers cholesterol by an average of 20 mg/dL with a standard deviation of 5 mg/dL. Drug B lowers cholesterol by an average of 25 mg/dL with a standard deviation of 10 mg/dL. Which drug exhibits more relative variability in cholesterol lowering?

(a) Drug A (b) Drug B (c) They have the same relative variability. (d) Cannot be determined without the sample size.

Answer : (b) Drug B

Explanation: Relative variability is measured by the coefficient of variation (CV), which is calculated as: (standard deviation / mean) * 100%. A higher CV indicates greater relative variability.

For Drug A: $CV = (5/20) * 100\% = 25\%$

For Drug B: $CV = (10/25) * 100\% = 40\%$

Drug B has a higher CV, indicating more relative variability in its effect on cholesterol lowering.

2. A researcher wants to determine if there's a correlation between daily hours of sunlight and vitamin D levels. They find a correlation coefficient of 0.6. This means:

- (a) More sunlight causes higher vitamin D levels.
(b) There is a moderately strong positive linear relationship between sunlight and vitamin D.
(c) There is a weak positive linear relationship between sunlight and vitamin D.
(d) There is no relationship between sunlight and vitamin D.

Answer: (b) There is a moderately strong positive linear relationship between sunlight and vitamin D.

Explanation: A correlation coefficient measures the strength and direction of a linear relationship between two variables.

The value of 0.6 indicates a positive relationship, meaning that as sunlight exposure increases, vitamin D levels tend to increase as well.

A correlation coefficient of 0.6 is considered a moderately strong positive relationship.

3. A screening test for a disease has a sensitivity of 90% and a specificity of 80%. If 1000 people are tested, and 100 actually have the disease, approximately how many false positives would you expect? (a) 80 (b) 100 (c) 180 (d) 200

Ans: (c) 180

Explanation:

Sensitivity: The proportion of true positives correctly identified by the test (90% in this case).

Specificity: The proportion of true negatives correctly identified by the test (80% in this case).

Out of 1000 people, 100 have the disease, and 900 do not.

With 80% specificity, 20% of the 900 people without the disease will be incorrectly identified as having the disease (false positives).

False Positives = $0.20 \times 900 = 180$

4. Which of the following is NOT a potential confounding variable when studying the relationship between coffee consumption and heart disease?

(a) Age (b) Smoking status (c) Time of day the coffee is consumed (d) Exercise habits

Ans: **(c) Time of day the coffee is consumed**

Explanation: A confounding variable is a factor that is related to both the exposure (coffee consumption) and the outcome (heart disease) and can distort the true relationship between them. Age, smoking status, and exercise habits are all known to be related to both coffee consumption and heart disease, making them potential confounders.

The time of day coffee is consumed is less likely to have a significant impact on heart disease risk compared to the other factors.

5. Which of the following graphical methods is MOST appropriate for comparing the distributions of two independent groups of continuous data?

(a) Scatter plot (b) Histogram (c) Box plot (d) Pie chart

Ans: **(c) Box plot**

Explanation: Box plots are a great way to visualize the distribution of continuous data, especially for comparing groups. They show the median, quartiles, and potential outliers, providing a good summary of the data's spread and central tendency.

11. Which best describes the relationship between sample size and the width of a confidence interval? (a) Larger sample size leads to wider intervals. (b) Smaller sample size leads to narrower intervals. (c) Larger sample size leads to narrower intervals. (d) Sample size has no effect on confidence interval width.

Answer: **(c) Larger sample size leads to narrower intervals.**

Explanation: A larger sample size provides more information about the population, leading to a more precise estimate of the population parameter. This increased precision results in a smaller margin of error and a narrower confidence interval.

13. What does the area under the receiver operating characteristic (ROC) curve (AUC) represent?

(a) Sensitivity of a diagnostic test (b) Specificity of a diagnostic test (c) Accuracy of a diagnostic test (d) Prevalence of a disease

Answer: **(c) Accuracy of a diagnostic test**

Explanation: The area under the ROC curve (AUC) is a measure of the overall ability of a diagnostic test to discriminate between individuals with and without the disease. A higher AUC indicates better accuracy.

14. Which type of bias occurs when participants in a study are systematically different from the population they are meant to represent?

(a) Selection bias (b) Recall bias (c) Confounding bias (d) Information bias

Answer: **(a) Selection bias**

Explanation: Selection bias occurs when the participants in a study are not representative of the population of interest. This can lead to inaccurate results and conclusions.

16. What is the purpose of blinding in a clinical trial?

(a) To reduce bias by preventing participants and/or researchers from knowing which treatment group participants are in.

(b) To ensure that the sample size is large enough to detect a statistically significant difference. (c) To control for confounding variables that might affect the outcome of the study.

(d) To make sure that the data is analyzed correctly.

Answer: **(a) To reduce bias by preventing participants and/or researchers from knowing which treatment group participants are in.**

Explanation: Blinding helps to prevent bias by ensuring that participants and researchers are unaware of the treatment assignment. This prevents their expectations or beliefs from influencing the outcome of the study.

17. In hypothesis testing, if the null hypothesis (H_0) is rejected, what does it imply?

A) The alternative hypothesis (H_1) is proven true. B) There is sufficient evidence to support the alternative hypothesis. C) The null hypothesis is absolutely false. D) The null hypothesis must always be accepted.

Answer: **(b) There is sufficient evidence to support the alternative hypothesis**

Explanation: In hypothesis testing, the null hypothesis (H_0) is a statement of no effect or no difference. If the null hypothesis is rejected, it means that there is enough evidence to support the alternative hypothesis (H_1), which is the statement that there is an effect or difference.

18. What is the difference between precision and accuracy?

(a) Precision refers to how close measurements are to each other, while accuracy refers to how close measurements are to the true value.

(b) Precision and accuracy are interchangeable terms.

(c) Accuracy refers to how close measurements are to each other, while precision refers to how close measurements are to the true value.

(d) Precision refers to the ability to get the same result multiple times, while accuracy refers to the ability to get the correct result.

Answer: **(a) Precision refers to how close measurements are to each other, while accuracy refers to how close measurements are to the true value.**

Explanation: Precision refers to the consistency or reproducibility of measurements. Accuracy refers to how close a measurement is to the actual or true value.

19. Which of the following is a measure of spread or dispersion in a dataset?

(a) Mean (b) Median (c) Standard deviation (d) Mode

Answer: **(c) Standard deviation**

Explanation: The standard deviation is a measure of the spread or dispersion of data around the mean. A higher standard deviation indicates greater variability in the data.

20. In a normally distributed dataset, approximately what percentage of the data falls within two standard deviations of the mean?

(a) 68% (b) 95% (c) 99.7% (d) 50%

Answer: **(b) 95%**

Explanation: In a normal distribution, approximately 95% of the data falls within two standard deviations of the mean. This is known as the 68-95-99.7 rule (or the empirical rule), which describes the percentage of data within one, two, and three standard deviations of the mean, respectively.