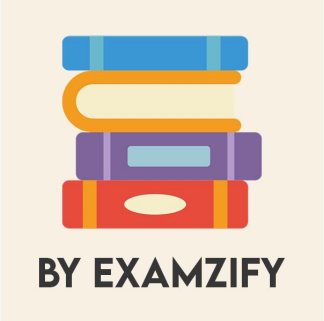
University of Central Florida (UCF) GEB4522 Data Driven Decision Making Final Exam Practice (Sample)

Study Guide



Everything you need from our exam experts!

Questions

1. Why is cross-validation important in model assessment?

   A. It increases the complexity of the model

   B. It prevents overfitting by evaluating model performance

   C. It focuses on speed rather than accuracy of predictions

   D. It allows easy comparison between multiple models

2. If, on average, more umbrella sales occur on days with more rainfall, then the strongest statement we can make is:

   A. Increased umbrella sales causes increased rainfall.

   B. Rainfall and umbrella sales are associated with one another.

   C. Increased rainfall causes increased umbrella sales.

   D. More data is needed to make a statement on the scenario.

3. True or False: From an employee viewpoint, high quality data is data that is in conformance to standards.

   A. True

   B. False

   C. Not applicable

   D. Depends on the context

4. If a report states that 200 is twice as much as 100, what type of data is being used?

   A. ordinal

   B. ratio

   C. nominal

   D. interval

5. Which measure of central tendency is defined as the value for which half the observations are higher and half are lower?

   A. Median

   B. Mean

   C. Mode

   D. Variance

6. What should be included in your executive summary?

A. Specific financial information

B. The recommended course of action

C. Extensive background information

D. A direct address to the client of record

7. Which of the following is most accurate to describe writing letters and writing e-mail?

A. letters have unreliable formatting, e-mail has reliable formatting

B. a greeting by the first name is always appropriate for both e-mail and letters

C. letters follow more established rules than e-mail

D. email is formal, letters are informal

8. Which of the following is an example of qualitative data?

A. Customer feedback

B. Sales revenue

C. Employee turnover rates

D. Annual profit margins

9. Which of the following most likely describes an appropriate use of a balanced scorecard?

A. Identify the causes for any gap between targets and actual values for each indicator

B. Determining the causes for delays between leading and lagging indicators

C. Work backward from skills and capabilities up to owners to determine the chain of cause-and-effect

D. All of the above

10. What can be a result of focusing on unimportant data during analysis?

A. Enhanced decision-making

B. Increased confusion

C. Motivated staff

D. Cost savings

Answers

1. B
2. B
3. A
4. B
5. A
6. B
7. C
8. A
9. A
10. B

Explanations

1. Why is cross-validation important in model assessment?

A. It increases the complexity of the model

B. It prevents overfitting by evaluating model performance

C. It focuses on speed rather than accuracy of predictions

D. It allows easy comparison between multiple models

Cross-validation is a crucial technique in model assessment because it plays a significant role in preventing overfitting. Overfitting occurs when a model learns the training data too well, capturing noise and outliers instead of the underlying patterns. This often leads to a model that performs excellently on the training data but poorly on unseen data. By using cross-validation, data is split into different subsets multiple times, allowing the model to be trained and validated on various combinations of training and validation sets. This provides a more reliable estimate of a model's performance on new, unseen data. It helps ensure that the insights gained from the training process are generalizable rather than tailored to the specificities of the training dataset. Consequently, cross-validation gives a better assessment of how the model will perform in real-world situations, improving model robustness and accuracy. In contrast, options discussing complexity and speed do not relate to the main advantage of cross-validation, which is to evaluate effectiveness in a way that guards against overfitting. The focus on comparison among models, while beneficial, is a secondary aspect; the primary reason for cross-validation is its role in ensuring the model's ability to generalize beyond the training data.

2. If, on average, more umbrella sales occur on days with more rainfall, then the strongest statement we can make is:

A. Increased umbrella sales causes increased rainfall.

B. Rainfall and umbrella sales are associated with one another.

C. Increased rainfall causes increased umbrella sales.

D. More data is needed to make a statement on the scenario.

The strongest statement we can make in this scenario is that rainfall and umbrella sales are associated with one another. This conclusion is based on the observation of a correlation between the two variables. When analyzing data where one variable shows a consistent pattern with another, such as increased umbrella sales coinciding with heightened rainfall, it suggests a relationship. However, it's crucial to understand that correlation does not imply causation. Although umbrella sales increase with more rainfall, we cannot definitively conclude that changes in rainfall directly cause changes in umbrella sales or vice versa, which is why choices that imply a direct cause-and-effect relationship are not appropriate here. Additionally, while seeking more data can be beneficial for a more robust analysis, the existing data already indicates a noticeable association between rainfall and umbrella sales, making the statement regarding their association the strongest and most accurate conclusion that can be drawn at this stage.

3. True or False: From an employee viewpoint, high quality data is data that is in conformance to standards.

A. True

B. False

C. Not applicable

D. Depends on the context

From an employee's viewpoint, high-quality data is indeed considered to be data that conforms to established standards. This perspective is grounded in the need for reliability, accuracy, and consistency in the data that employees use to perform their duties effectively. When data meets predefined standards, it ensures that employees can trust the information for decision-making processes, analysis, and operational functions. Conformance to standards involves aspects such as accuracy, completeness, consistency, and timeliness, all of which are essential for maintaining the integrity of the data. Therefore, high-quality data supports employees in achieving their objectives efficiently, leading to better overall performance within the organization.

4. If a report states that 200 is twice as much as 100, what type of data is being used?

A. ordinal

B. ratio

C. nominal

D. interval

The correct identification of the data as ratio is based on the inherent characteristics of ratio data, which include a meaningful zero point and the ability to make comparisons in terms of multiples. In this case, stating that 200 is twice as much as 100 exemplifies a clear and absolute measurement that indicates the quantity of one value relative to another. Ratio data allows for the comparison of absolute magnitudes and ratios. For instance, if you consider the value of zero in a ratio scale (such as 0 denotes no quantity of the measured attribute), this provides a foundation for defining relationships between different magnitudes. In this instance, you can confidently say that 200 is not only greater than 100, but also exactly twice 100, reinforcing the ratio aspect. In contrast, other types of data such as ordinal, nominal, and interval does not support such multiplicative comparisons in the same meaningful way. Ordinal data provides a ranking but does not quantify differences between ranks. Nominal data categorizes without a quantitative relation, and interval data allows for addition and subtraction but lacks a true zero, making the concept of ratios inapplicable. Thus, the statement reflects the properties of ratio data correctly.

5. Which measure of central tendency is defined as the value for which half the observations are higher and half are lower?

A. Median

B. Mean

C. Mode

D. Variance

The median is the measure of central tendency that represents the middle value in a dataset when it is ordered from least to greatest. This means that when you arrange all the observations, the median is the point that divides the dataset into two equal halves: one half has values greater than the median, and the other half has values lower. This characteristic of the median makes it particularly useful in datasets that may contain outliers or skewed distributions, as it provides a better central location than the mean in such cases. When the data is not symmetrically distributed, the mean can be significantly affected by extreme values, whereas the median remains a more robust indicator of central tendency. In contrast, the mean is the average of all observations and can be influenced heavily by outliers, while the mode represents the most frequently occurring value and does not provide information about the distribution of all values. Variance, on the other hand, measures the spread or dispersion of the dataset rather than a central tendency. Therefore, the median is correctly identified as the value for which half the observations are higher and half are lower.

6. What should be included in your executive summary?

A. Specific financial information

B. The recommended course of action

C. Extensive background information

D. A direct address to the client of record

The inclusion of the recommended course of action in an executive summary is essential for effectively communicating the key takeaway from a report or proposal. The executive summary serves as a distillation of the main points, allowing decision-makers to quickly grasp the strategic direction being suggested. By clearly outlining the recommended actions, readers can understand the rationale behind the analysis and the proposed solutions without delving into the full document. This clarity ensures that stakeholders can efficiently assess the recommendations and their potential implications for the organization. It creates a focused narrative that aligns with the data and analyses presented throughout the document, thereby enhancing the decision-making process.

7. Which of the following is most accurate to describe writing letters and writing e-mail?

A. letters have unreliable formatting, e-mail has reliable formatting

B. a greeting by the first name is always appropriate for both e-mail and letters

C. letters follow more established rules than e-mail

D. email is formal, letters are informal

The assertion that letters follow more established rules than e-mail is accurate because traditional letter writing has a long history with clearly defined formats, structures, and conventions. This includes aspects such as salutation, body, closing, and signature, which are often standardized across different contexts and purposes. In contrast, e-mail communication is more dynamic and can vary widely depending on the individual sender's style, the nature of the correspondence, and the recipient. While e-mails can adopt formal structures, they generally offer more flexibility in formatting and tone, which can lead to a less consistent adherence to established rules. The traditional nature of letter writing means that it is often perceived as more formal and respectful, which impacts the way we design and format these communications. In the digital age, while e-mail has become a primary mode of communication, it still lacks the uniformity that is inherent in letter formats. This distinction highlights the structured nature of letters compared to the varied practices found in e-mail exchanges.

8. Which of the following is an example of qualitative data?

A. Customer feedback

B. Sales revenue

C. Employee turnover rates

D. Annual profit margins

Qualitative data refers to non-numeric information that describes qualities or characteristics. Customer feedback is a prime example of qualitative data because it often consists of subjective insights, opinions, and descriptions that cannot be quantified numerically. This type of data is often used to understand customer experiences, preferences, and perceptions, providing rich context that numerical data alone cannot offer. In contrast, the other options represent quantitative data, which can be measured and expressed numerically. Sales revenue, employee turnover rates, and annual profit margins all involve specific numerical values that can be calculated and analyzed statistically. These figures provide hard data that can be used for analysis, forecasting, and decision-making, but they do not capture the descriptive insights that qualitative data offers.

9. Which of the following most likely describes an appropriate use of a balanced scorecard?

A. Identify the causes for any gap between targets and actual values for each indicator

B. Determining the causes for delays between leading and lagging indicators

C. Work backward from skills and capabilities up to owners to determine the chain of cause-and-effect

D. All of the above

The correct answer highlights how the balanced scorecard can effectively identify gaps between performance targets and actual outcomes. This tool is used to assess an organization's strategic performance by translating its vision and strategy into a comprehensive set of performance metrics. In practice, identifying the causes of gaps is essential because it enables management to understand where performance does not meet expectations and allows for targeted interventions to address these deficiencies. The other options, while they may relate to performance measurement and improvement, do not directly reflect the primary purpose and most common application of the balanced scorecard. The identification of causes for delays between leading and lagging indicators pertains to a more tactical analysis rather than the broader strategic perspective that the balanced scorecard encompasses. Similarly, working backward from skills and capabilities to determine cause-and-effect relationships may be relevant for skills assessment or training effectiveness but does not align as directly with the balanced scorecard's holistic approach to performance management. Thus, the focus on identifying gaps and understanding their causes as part of a balanced scorecard framework is fundamental, making the first option the most appropriate use of this strategic tool.

10. What can be a result of focusing on unimportant data during analysis?

A. Enhanced decision-making

B. Increased confusion

C. Motivated staff

D. Cost savings

Focusing on unimportant data during analysis can lead to increased confusion because it distracts from the core issues and key insights that drive meaningful decision-making. When analysts sift through irrelevant or less significant information, it can create a muddled understanding of the actual situation. This can result in decision-makers drawing incorrect conclusions or making uninformed choices based on noise rather than actionable insights. Clarity is crucial in data-driven decision-making, and when the analysis is clouded by unimportant data, it tends to hinder the overall effectiveness of the decision-making process.