

Crash Course in Causality Quiz Questions

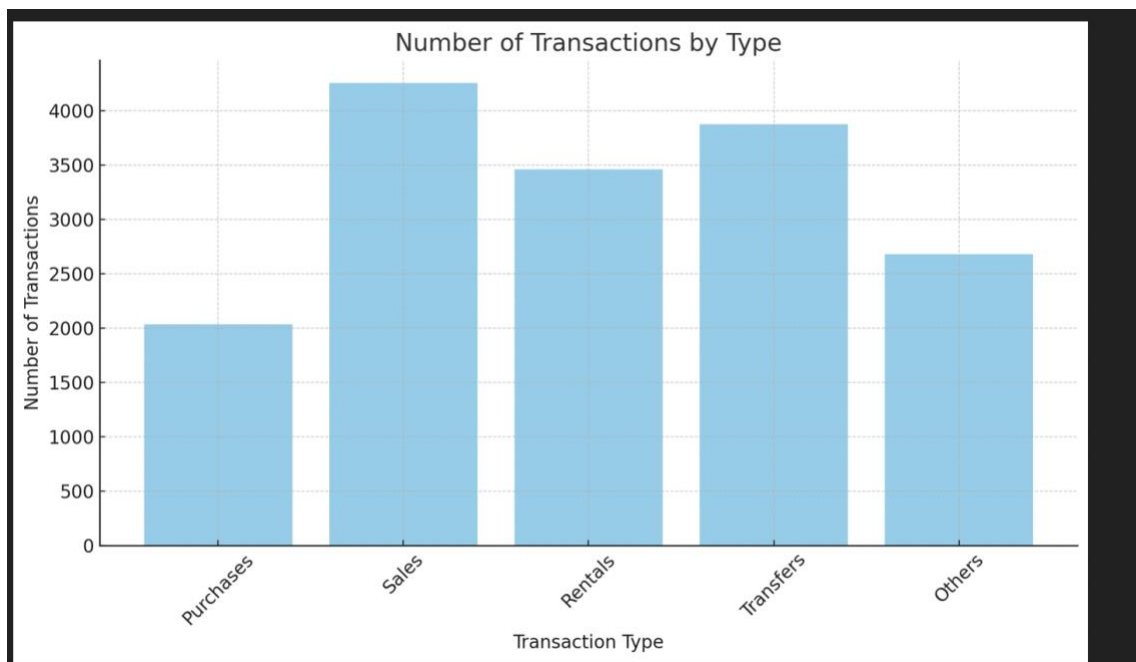
Question 1: What does causality aim to establish between two variables?

- A) Correlation
- B) Association
- C) Cause-and-effect relationship
- D) Similarity

Answer: C) Cause-and-effect relationship

Explanation: Causality aims to establish a cause-and-effect relationship, determining how one variable directly influences another. Unlike correlation or association, which only indicate that variables change together, causality explains the mechanism of influence.

Question 2: Based on the graph, which transaction type has the second highest number of transactions?



- A) Purchases
- B) Sales
- C) Rentals
- D) Transfers
- E) Others

Correct Answer: B) Sales

Explanation: According to the bar chart, 'Rentals' has the highest number of transactions as it has the tallest bar. The second tallest bar represents 'Sales', which indicates it has the second highest number of transactions among the options given.

Question 3: What is a confounding variable?

- A) A variable that is affected by the treatment
- B) A variable that has no impact on the outcome
- C) A variable that influences both the treatment and the outcome
- D) A variable that is used for randomization

Answer: C) A variable that influences both the treatment and the outcome

Explanation: A confounding variable is one that affects both the independent (treatment) and dependent (outcome) variables, potentially misleading the causal relationship unless appropriately controlled for.

Question 4: Which method can be used to adjust for confounding in observational studies?

- A) Correlation coefficients
- B) Propensity score matching
- C) Simple linear regression
- D) Pearson's chi-squared test

Answer: B) Propensity score matching

Explanation: Propensity score matching is a technique used to control for confounding in observational studies by matching subjects with similar characteristics (propensity scores) in treatment and control groups, thereby simulating randomization.

Question 5: What is the purpose of conducting a sensitivity analysis in causal inference studies?

- A) To test the robustness of the findings to different model specifications
- B) To increase the sample size of the study
- C) To reduce the computational complexity of the analysis

D) To identify the best machine learning model for prediction

Answer: A) To test the robustness of the findings to different model specifications

Explanation: Sensitivity analysis is conducted to assess how the results of a study might change when different assumptions or model specifications are applied, thereby testing the robustness and reliability of the causal findings.

Question 6: In the context of causal diagrams, what does a directed acyclic graph (DAG) represent?

- A) The correlation between variables
- B) The chronological order of variables
- C) The causal relationships among variables
- D) The distribution of variables

Answer: C) The causal relationships among variables

Explanation: A DAG is a graphical representation used to depict causal relationships among variables, where nodes represent variables and directed edges indicate causal directions, helping to visualize and reason about causality.

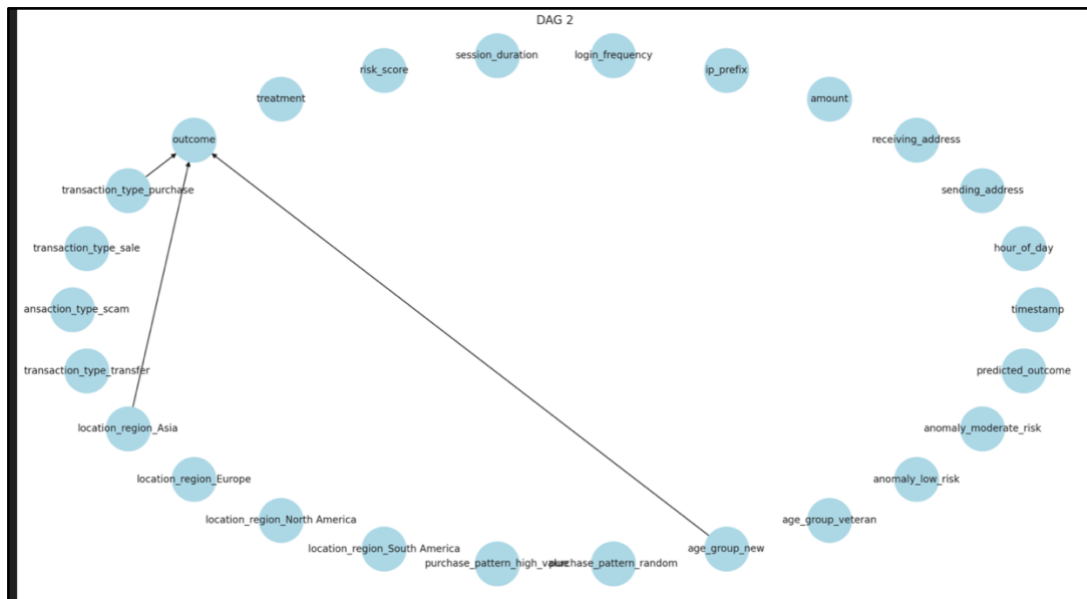
Question 7: Why is handling missing data important in causal analysis?

- A) Missing data can introduce bias and affect the validity of causal inferences
- B) Missing data increases the accuracy of the models
- C) Missing data simplifies the analysis
- D) Missing data enhances the computational efficiency of the analysis

Answer: A) Missing data can introduce bias and affect the validity of causal inferences

Explanation: Missing data, especially if not missing completely at random, can bias the results of an analysis by excluding or underrepresenting certain observations, thus affecting the validity and reliability of causal inferences.

Question 8: What does the graph primarily represent in relation to transactions?



- A) Transaction types and associated risks
- B) Product inventory levels over time
- C) Customer satisfaction ratings
- D) Vendor performance metrics

Answer: A) Transaction types and associated risks

Explanation:

The graph appears to be a Directed Acyclic Graph (DAG), which is often used to model dependencies and processes in various domains, including data science and risk assessment. In this particular graph, we see nodes that represent different transaction types (such as purchase, sale, scam, and transfer), different geographic regions (Asia, Europe, North America, South America), and various factors related to transaction behavior (such as purchase pattern and age group). These nodes are connected to an "outcome" node, suggesting that they are factors contributing to the outcome of a transaction. Additionally, there are nodes that indicate different levels of anomaly risk (low and moderate), which are indicative of a risk assessment process. This suggests that the graph is primarily representing transaction types and the factors that contribute to their associated risks, rather than inventory levels, customer satisfaction, or vendor

Question 9: What role does counterfactual reasoning play in causal inference?

- A) It predicts the future outcomes based on past data

- B) It estimates the causal effect by considering what would have happened in the absence of the treatment
- C) It classifies variables into different categories
- D) It visualizes the data for better understanding

Answer: B) It estimates the causal effect by considering what would have happened in the absence of the treatment

Explanation: Counterfactual reasoning involves considering the hypothetical scenario of what the outcome would have been had the treatment not been applied. This approach is fundamental in causal inference for estimating the causal effect of a treatment on an outcome. By comparing the actual outcome with the counterfactual outcome, researchers can isolate the effect of the treatment from other factors, providing a more accurate estimation of the causal impact.

Question 10: Which of the following is a key assumption in causal inference analysis to ensure valid conclusions?

- A) Independence assumption
- B) Multicollinearity assumption
- C) Homoscedasticity assumption
- D) Normality assumption

Answer: A) Independence assumption

Explanation: The independence assumption, often related to the assumption of no unmeasured confounders, is crucial in causal inference. It assumes that the treatment assignment is independent of the potential outcomes, a condition necessary to draw valid causal conclusions. This assumption helps ensure that the observed treatment effects are not biased by factors that influence both the treatment and the outcome. Multicollinearity, homoscedasticity, and normality are important in regression analysis but are not the key assumptions specifically for causal inference.