

Assignment -2
Understanding Causality
Quiz Questions

Quiz Questions on Causal Inference

Question 1: What is the fundamental problem of causal inference?

- A) We cannot observe multiple treatments simultaneously
- B) We cannot observe the same unit under both treatment and control conditions at the same time
- C) We cannot randomize treatment assignment in observational studies
- D) We cannot measure confounding variables accurately

Correct Answer: B

Explanation: The fundamental problem of causal inference is that we can never observe what would have happened to the same individual both with and without treatment at the same time (the counterfactual). This is why we must rely on assumptions and statistical methods to estimate causal effects. Option A is incorrect because we can observe multiple treatments, just not on the same unit simultaneously. Option C describes a challenge but not the fundamental problem. Option D relates to measurement error, not the core identification problem.

Question 2: In a Directed Acyclic Graph (DAG), what does it mean when variable Z is a confounder of the relationship between treatment X and outcome Y?

- A) Z is caused by both X and Y
- B) Z causes both X and Y
- C) Z is on the causal path from X to Y
- D) Z blocks all paths from X to Y

Correct Answer: B

Explanation: A confounder is a common cause of both treatment and outcome ($Z \rightarrow X$ and $Z \rightarrow Y$), creating spurious association. This means we must control for Z to isolate the true causal effect of X on Y. Option A describes a collider, not a confounder. Option C describes a mediator. Option D describes a variable that would eliminate confounding but isn't the definition of a confounder itself.

Question 3: What happens when you condition on a collider in causal analysis?

- A) It removes confounding bias
- B) It opens a backdoor path and introduces bias
- C) It blocks all causal paths
- D) It has no effect on the causal estimate

Correct Answer: B

Explanation: Conditioning on a collider (a variable caused by both treatment and outcome, or their causes) opens a path between its parents, introducing collider bias or selection bias. This is a common mistake in causal analysis. Option A is incorrect—controlling for colliders creates bias rather than removing it. Options C and D are incorrect because conditioning on a collider has a specific harmful effect by creating spurious associations.

Question 4: Which of the following best describes propensity score matching?

- A) Matching units with similar outcome values
- B) Matching treated and control units with similar probabilities of receiving treatment
- C) Matching units with identical confounding variables
- D) Matching units randomly from treatment and control groups

Correct Answer: B

Explanation: Propensity score matching matches treated and control units based on their estimated probability of receiving treatment given covariates. This balances observed confounders between groups without requiring exact matching on all variables. Option A is incorrect as matching on outcomes would introduce bias. Option C describes exact matching, which is often impractical with many covariates. Option D describes random matching, which doesn't address confounding.

Question 5: What is the key assumption behind using instrumental variables (IV) for causal inference?

- A) The instrument is correlated with the outcome only through the treatment
- B) The instrument is randomly assigned
- C) The instrument is a strong predictor of the outcome
- D) The instrument eliminates all confounding

Correct Answer: A

Explanation: A valid instrument must affect the outcome only through its effect on the treatment (exclusion restriction), be correlated with treatment (relevance), and be independent of confounders (exogeneity). Option A captures the exclusion restriction, which is the key identifying assumption. Option B is helpful but not required if other IV assumptions hold. Option C incorrectly suggests the instrument should predict the outcome directly, which would violate the exclusion restriction. Option D overstates what IV achieves.

Question 6: In the context of causal inference, what does "exchangeability" mean?

- A) Treatment and control groups can be swapped without changing results
- B) The potential outcomes are independent of treatment assignment
- C) Confounders can be exchanged for instrumental variables
- D) Treatment effects are the same across all individuals

Correct Answer: B

Explanation: Exchangeability (or conditional exchangeability) means that potential outcomes are independent of treatment assignment, possibly conditional on covariates. This implies that treated and control groups would have had the same average outcome if they had received the same treatment. Option A misinterprets the concept as a symmetry property. Option C is nonsensical. Option D describes treatment effect homogeneity, not exchangeability.

Question 7: Which statement about the backdoor criterion is correct?

- A) It identifies variables that should NOT be controlled for
- B) It provides sufficient conditions for identifying causal effects by blocking non-causal paths
- C) It requires controlling for all variables in the DAG
- D) It only applies to randomized experiments

Correct Answer: B

Explanation: The backdoor criterion specifies a sufficient set of variables to control for to block all non-causal (backdoor) paths from treatment to outcome while not opening new biasing paths. This allows identification of causal effects in observational data. Option A is partially true (we shouldn't control for colliders or mediators) but doesn't capture what the criterion provides. Option C is incorrect—we shouldn't control for mediators or colliders. Option D is wrong; the criterion is specifically for observational studies.

Question 8: What is a mediator in causal analysis?

- A) A variable that confounds the treatment-outcome relationship
- B) A variable that lies on the causal path between treatment and outcome
- C) A variable that is caused by both treatment and outcome
- D) A variable that modifies the treatment effect

Correct Answer: B

Explanation: A mediator is a variable through which the treatment affects the outcome ($X \rightarrow M \rightarrow Y$). Mediators help us understand the mechanisms of causal effects. Controlling for mediators blocks part of the total causal effect. Option A describes a confounder. Option C describes a collider. Option D describes an effect modifier or moderator, not a mediator.

Question 9: Why is randomization considered the gold standard for causal inference?

- A) It ensures perfect balance on all measured confounders
- B) It ensures balance on both measured and unmeasured confounders in expectation
- C) It eliminates all bias in causal estimates
- D) It allows for the largest possible sample sizes

Correct Answer: B

Explanation: Randomization breaks the association between treatment and all potential confounders (both observed and unobserved) in expectation, achieving exchangeability without requiring knowledge of confounders. Option A is incomplete—randomization also balances unmeasured confounders. Option C overstates the case; randomization addresses confounding but not other sources of bias like measurement error or selection bias. Option D is unrelated to why randomization helps with causal inference.

Question 10: What is the Average Treatment Effect (ATE)?

- A) The effect of treatment on the treated individuals only
- B) The average difference in potential outcomes between treatment and control for the entire population
- C) The effect of treatment averaged across different time periods
- D) The weighted average of treatment effects across subgroups

Correct Answer: B

Explanation: $ATE = E[Y(1) - Y(0)]$ represents the expected difference in outcomes if everyone received treatment versus if everyone received control. This is the causal effect for the entire population. Option A describes ATT (Average Treatment Effect on the Treated). Option C

describes something like a time-averaged effect. Option D could describe various weighted estimands but not specifically ATE.

Question 11: In difference-in-differences (DiD) analysis, what is the key identifying assumption?

- A) Treatment assignment is random
- B) Parallel trends: treated and control groups would have followed similar trends without treatment
- C) There is no spillover between groups
- D) All confounders are measured

Correct Answer: B

Explanation: The parallel trends assumption states that in the absence of treatment, the treated and control groups would have experienced the same change over time. This allows DiD to difference out time-invariant confounding and common time trends. Option A would be ideal but isn't required for DiD. Option C (no spillover/SUTVA) is important but not the key identifying assumption for DiD. Option D is not required; DiD handles time-invariant unobserved confounders.

Question 12: What is selection bias in observational causal studies?

- A) Bias from selecting the wrong statistical model
- B) Bias from selecting the wrong outcome variable
- C) Bias arising when treatment assignment is related to potential outcomes
- D) Bias from selecting a non-representative sample

Correct Answer: C

Explanation: Selection bias occurs when the treatment assignment mechanism is related to potential outcomes, violating exchangeability. This creates systematic differences between groups beyond the treatment effect. Option A refers to model misspecification. Option B describes a different specification issue. Option D describes sampling bias, which can be a problem but is distinct from the causal selection bias arising from non-random treatment assignment.

Question 13: What does the Stable Unit Treatment Value Assumption (SUTVA) require?

- A) Treatment effects remain stable over time
- B) No interference between units and no hidden variations of treatment
- C) The sample is stable and representative
- D) Treatment values are measured without error

Correct Answer: B

Explanation: SUTVA has two parts: (1) no interference—one unit's treatment doesn't affect another unit's outcome, and (2) no hidden variations—treatment is well-defined with no different versions. SUTVA is crucial for defining individual causal effects. Option A describes temporal stability, not SUTVA. Option C relates to sampling. Option D relates to measurement error.

Question 14: In regression discontinuity design (RDD), causal identification relies on:

- A) Units just above and below the threshold being comparable
- B) Random assignment of the running variable
- C) Controlling for all confounders
- D) Having a large discontinuity in the outcome

Correct Answer: A

Explanation: RDD exploits a threshold rule for treatment assignment, assuming that units just above and below the cutoff are similar except for treatment status. This creates a quasi-random assignment near the threshold. Option B is incorrect—the running variable is typically not random, but treatment assignment around the threshold is as-if random. Option C isn't necessary due to the local randomization. Option D confuses the size of the treatment effect with the identification strategy.

Question 15: When should you include post-treatment variables as controls in a causal analysis?

- A) Always, to reduce bias
- B) Never, as they may be affected by treatment and introduce bias
- C) Only if they're strongly correlated with the outcome
- D) Only if they're measured before the outcome

Correct Answer: B

Explanation: Post-treatment variables (measured after treatment) can be mediators or colliders affected by treatment. Controlling for them can block causal pathways (bad collider bias), leading to biased causal estimates. We should generally only control for pre-treatment confounders. Option A is a common fallacy—more controls aren't always better. Options C and D miss the fundamental issue that post-treatment variables can be affected by treatment.

