Start coding or generate with AI.

Multiple Choice Quiz: Crash Course in Causality

Question 1: What is the main goal of causal inference?

1. To identify correlations between variables.
2. To determine the mechanisms of cause-and-effect relationships.
3. To model random events in data.
4. To assign numerical scores to categorical data.

Correct Answers:

- **2. To determine the mechanisms of cause-and-effect relationships.**
  Explanation: The primary goal of causal inference is to establish and understand how one variable (the cause) directly impacts another (the effect). This goes beyond merely identifying patterns or correlations.

Incorrect Answers:

- **1. To identify correlations between variables.**
  Explanation: While identifying correlations is an important step in data analysis, correlation alone does not establish causation.
- **3. To model random events in data.**
  Explanation: Modeling random events is part of probability theory, not causal inference.
- **4. To assign numerical scores to categorical data.**
  Explanation: This describes encoding techniques like one-hot encoding, which are preprocessing steps unrelated to causal inference.

Question 2: What does a propensity score represent?

1. The probability of being in the treatment group given observed covariates.
2. The likelihood of achieving a certain outcome.
3. The odds ratio between two variables.
4. The similarity between two groups in observational data.

Correct Answers:

- **1. The probability of being in the treatment group given observed covariates.**
  Explanation: Propensity scores estimate the likelihood of receiving a treatment based on observable characteristics, enabling fair comparisons between treatment and control groups.

Incorrect Answers:

- **2. The likelihood of achieving a certain outcome.**
  Explanation: Propensity scores do not predict outcomes; they predict treatment assignment.
- **3. The odds ratio between two variables.**
  Explanation: Odds ratios compare the odds of an outcome across groups, but propensity scores focus on treatment assignment.
- **4. The similarity between two groups in observational data.**
  Explanation: While propensity scores help balance groups, they do not directly measure similarity.

Question 3: Why is randomization important in causal inference?

1. It eliminates all confounding variables.
2. It reduces bias in treatment assignment.
3. It ensures the sample is representative of the population.
4. It creates comparable groups for analysis.

Correct Answers:

- **2. It reduces bias in treatment assignment.**
  Explanation: Randomization ensures that treatment assignment is not influenced by confounding variables, reducing bias.
- **4. It creates comparable groups for analysis.**
  Explanation: By randomizing, groups are more likely to be balanced across observed and unobserved characteristics, facilitating fair comparisons.

Incorrect Answers:

- **1. It eliminates all confounding variables.**
  Explanation: Randomization minimizes confounding but cannot eliminate it entirely, especially with small sample sizes.
- **3. It ensures the sample is representative of the population.**
  Explanation: Randomization balances groups but does not guarantee representativeness unless the sampling process is also random.

Question 4: Which of the following are benefits of using Propensity Score Matching (PSM)?

1. It reduces the impact of observed confounders.
2. It accounts for unmeasured confounding variables.
3. It allows for fair comparisons in observational studies.

4. It guarantees causal inference.

Correct Answers:

- **1. It reduces the impact of observed confounders.**
  Explanation: PSM adjusts for confounding variables that are observed and included in the model.
- **3. It allows for fair comparisons in observational studies.**
  Explanation: PSM creates balanced treatment and control groups to mimic randomization.

Incorrect Answers:

- **2. It accounts for unmeasured confounding variables.**
  Explanation: PSM can only address confounders that are measured and included in the model.
- **4. It guarantees causal inference.**
  Explanation: PSM improves causal estimates but does not guarantee causation, as unmeasured confounders may still exist.

Question 5: What does the Average Treatment Effect (ATT) measure?

1. The overall impact of treatment across all individuals.
2. The average effect of treatment for those who received it.
3. The likelihood of receiving a treatment.
4. The difference in outcomes for matched groups.

Correct Answers:

- **2. The average effect of treatment for those who received it.**
  Explanation: ATT focuses on the causal effect of treatment on the treated group specifically.

Incorrect Answers:

- **1. The overall impact of treatment across all individuals.**
  Explanation: This describes the Average Treatment Effect (ATE), not ATT.
- **3. The likelihood of receiving a treatment.**
  Explanation: This is related to propensity scores, not ATT.
- **4. The difference in outcomes for matched groups.**
  Explanation: While ATT involves matched groups, it specifically estimates the average effect for the treated group.

Question 6: What is the purpose of Directed Acyclic Graphs (DAGs) in causal inference?

1. To estimate causal effects numerically.

2. To visualize the structure of causal relationships.
3. To identify confounding variables.
4. To perform statistical tests for causality.

Correct Answers:

- **2. To visualize the structure of causal relationships.**
  Explanation: DAGs provide a graphical representation of variables and their causal connections.
- **3. To identify confounding variables.**
  Explanation: DAGs help identify confounders by mapping paths between variables.

Incorrect Answers:

- **1. To estimate causal effects numerically.**
  Explanation: DAGs visualize relationships but do not compute causal effects directly.
- **4. To perform statistical tests for causality.**
  Explanation: DAGs guide analysis but are not used for statistical testing.

Question 7: Which of the following is NOT an assumption of Propensity Score Matching?

1. No unmeasured confounders.
2. Treatment assignment is random.
3. Overlap in propensity scores.
4. Correct model specification for propensity score estimation.

Correct Answer:

- **2. Treatment assignment is random.**
  Explanation: PSM is specifically designed for observational studies where treatment assignment is not random.

Incorrect Answers:

- **1. No unmeasured confounders.**
  Explanation: PSM assumes that all confounders are measured and included in the model.
- **3. Overlap in propensity scores.**
  Explanation: Overlap ensures matched treatment and control groups can be created.
- **4. Correct model specification for propensity score estimation.**
  Explanation: Accurate propensity scores require a well-specified model.

Question 8: Why do we use the logit function in logistic regression?

1. To handle non-linear relationships between variables.
2. To convert probabilities into odds.
3. To ensure the output is between 0 and 1.
4. To create balanced datasets.

Correct Answers:

- **3. To ensure the output is between 0 and 1.**
  Explanation: The logit function transforms odds into probabilities, constraining the output to a valid range.

Incorrect Answers:

- **1. To handle non-linear relationships between variables.**
  Explanation: While logistic regression can model non-linearities, this is not the logit function's primary purpose.
- **2. To convert probabilities into odds.**
  Explanation: The logit function converts odds into probabilities, not the other way around.
- **4. To create balanced datasets.**
  Explanation: Balancing datasets is not a function of the logit transformation.

Question 9: What is the main limitation of observational studies for causal inference?

1. Lack of control over treatment assignment.
2. Limited sample size.
3. Overlap between treatment and control groups.
4. Presence of random error in data.

Correct Answer:

- **1. Lack of control over treatment assignment.**
  Explanation: Observational studies lack randomization, making them prone to confounding and bias.

Incorrect Answers:

- **2. Limited sample size.**
  Explanation: Sample size affects statistical power but is not specific to observational studies.
- **3. Overlap between treatment and control groups.**
  Explanation: Overlap is a requirement, not a limitation, for methods like PSM.
- **4. Presence of random error in data.**
  Explanation: Random error is common in all types of studies, not just observational ones.

Question 10: What does the F1 score balance in model evaluation?

1. Precision and recall.
2. Accuracy and precision.
3. Recall and specificity.
4. False positives and false negatives.

Correct Answer:

- **1. Precision and recall.**
  Explanation: The F1 score is the harmonic mean of precision and recall, balancing the trade-off between false positives and false negatives.

Question 11: What is the purpose of balancing covariates in Propensity Score Matching (PSM)?

1. To eliminate confounding entirely.
2. To create comparable groups for estimating treatment effects.
3. To increase the sample size.
4. To improve the model's accuracy.

Correct Answer:

- **2. To create comparable groups for estimating treatment effects.**
  Explanation: Balancing covariates ensures that treatment and control groups are similar, isolating the treatment effect by reducing the influence of observed confounders.

Incorrect Answers:

- **1. To eliminate confounding entirely.**
  Explanation: PSM reduces but does not completely eliminate confounding, especially for unmeasured variables.
- **3. To increase the sample size.**
  Explanation: Matching often reduces the effective sample size by discarding unmatched individuals.
- **4. To improve the model's accuracy.**
  Explanation: PSM focuses on improving causal estimates rather than predictive accuracy.

Question 12: Which of the following is a limitation of Propensity Score Matching?

1. It does not address unmeasured confounders.

2. It cannot be used with categorical variables.
3. It assumes a linear relationship between variables.
4. It requires a large sample size to work effectively.

Correct Answer:

- **1. It does not address unmeasured confounders.**
  Explanation: PSM relies on observed variables and cannot adjust for confounding caused by unmeasured factors.

Incorrect Answers:

- **2. It cannot be used with categorical variables.**
  Explanation: PSM works with categorical and continuous variables, provided they are included in the propensity score model.
- **3. It assumes a linear relationship between variables.**
  Explanation: PSM does not assume linearity; the propensity score model can accommodate non-linear relationships.
- **4. It requires a large sample size to work effectively.**
  Explanation: While larger sample sizes improve the quality of matching, PSM can still be applied with smaller datasets.

Question 13: What does it mean when there is poor overlap in propensity scores between treatment and control groups?

1. The treatment effect cannot be estimated.
2. Matching will exclude individuals from analysis.
3. There is a lack of balance in covariates between the groups.
4. Treatment assignment is completely random.

Correct Answers:

- **2. Matching will exclude individuals from analysis.**
  Explanation: Poor overlap means fewer matches can be found, leading to a reduction in the effective sample size.
- **3. There is a lack of balance in covariates between the groups.**
  Explanation: Poor overlap indicates that treatment and control groups differ significantly in covariates, making causal inference difficult.

Incorrect Answers:

- **1. The treatment effect cannot be estimated.**
  Explanation: The effect can still be estimated but with limitations due to reduced sample size

and balance.

- **4. Treatment assignment is completely random.**
  Explanation: Poor overlap suggests systematic differences in groups, not random treatment assignment.

Question 14: Why is the Average Treatment Effect on the Treated (ATT) more relevant than the Average Treatment Effect (ATE) in some studies?

1. ATT is easier to calculate.
2. ATT focuses on the population that actually received the treatment.
3. ATT accounts for unmeasured confounders.
4. ATT provides better predictions for all individuals.

Correct Answer:

- **2. ATT focuses on the population that actually received the treatment.**
  Explanation: ATT estimates the causal effect for those who received the treatment, making it more relevant when the treated group is of primary interest.

Incorrect Answers:

- **1. ATT is easier to calculate.**
  Explanation: ATT and ATE require similar steps; the relevance depends on the research question.
- **3. ATT accounts for unmeasured confounders.**
  Explanation: Neither ATT nor ATE can address unmeasured confounders.
- **4. ATT provides better predictions for all individuals.**
  Explanation: ATT focuses on the treated group and does not generalize to the entire population.

Question 15: Which of the following are challenges when interpreting causal effects in observational studies?

1. Presence of unmeasured confounders.
2. Difficulty in randomizing treatment assignment.
3. Limited availability of causal inference techniques.
4. Over-reliance on large sample sizes.

Correct Answers:

- **1. Presence of unmeasured confounders.**
  Explanation: Unmeasured confounders can bias causal estimates, making interpretation

challenging.

- **2. Difficulty in randomizing treatment assignment.**
  Explanation: Observational studies lack randomization, making it harder to isolate causal effects.

Incorrect Answers:

- **3. Limited availability of causal inference techniques.**
  Explanation: Techniques like PSM, DAGs, and instrumental variables are widely available but require careful application.
- **4. Over-reliance on large sample sizes.**
  Explanation: Sample size is important but not a unique challenge for observational studies.

These additional questions are designed to test a deeper understanding of PSM and causal inference concepts while providing clear explanations for each option.