**Quiz for Causal Analysis of Workplace Accidents**

**Asawari Kadam – 002711195**

**Question 1:**

What does causality primarily focus on?

1. Identifying relationships between two variables.
2. Determining whether one event causes another.
3. Observing patterns in data.
4. Understanding correlation between variables.

**Correct Answers:**

* **2. Determining whether one event causes another.**  
  *Explanation:* Causality is about understanding cause-and-effect relationships, where a change in one variable (cause) results in a change in another (effect). This is the essence of causal analysis.

**Incorrect Answers:**

* **1. Identifying relationships between two variables.**  
  *Explanation:* While causality involves understanding relationships, it is more specific—it seeks to determine if one variable causes changes in another. Merely identifying relationships is descriptive and not causal.
* **3. Observing patterns in data.**  
  *Explanation:* Observing patterns is part of exploratory data analysis but does not establish causation. Patterns alone can be misleading without causal interpretation.
* **4. Understanding correlation between variables.**  
  *Explanation:* Correlation measures association, not causation. Causality builds on correlation but adds directionality and a mechanism.

**Question 2:**

Which of the following methods are used for causal inference?

1. Regression Discontinuity Design (RDD).
2. Propensity Score Matching (PSM).
3. Descriptive Statistics.
4. Instrumental Variables (IV).

**Correct Answers:**

* **1. Regression Discontinuity Design (RDD).**  
  *Explanation:* RDD is used in causal inference to estimate the treatment effect at a cutoff point, such as before and after a policy change.
* **2. Propensity Score Matching (PSM).**  
  *Explanation:* PSM balances the treated and control groups based on the probability of receiving treatment, reducing bias in causal estimation.
* **4. Instrumental Variables (IV).**  
  *Explanation:* IV is used when treatment is endogenous (correlated with unobserved factors), leveraging an external variable to isolate the causal effect.

**Incorrect Answer:**

* **3. Descriptive Statistics.**  
  *Explanation:* Descriptive statistics summarize data but do not infer causal relationships. They are useful for exploratory analysis but not for establishing causality.

**Question 3:**

Which of the following are assumptions of Propensity Score Matching (PSM)?

1. All confounders are measured.
2. There are no unmeasured confounders.
3. Treatment assignment is random.
4. The data follows a linear distribution.

**Correct Answers:**

* **1. All confounders are measured.**  
  *Explanation:* PSM requires that all potential confounders are accounted for to achieve balance between treated and control groups.
* **2. There are no unmeasured confounders.**  
  *Explanation:* Unmeasured confounders can bias results, violating the assumption that treated and control groups are comparable.

**Incorrect Answers:**

* **3. Treatment assignment is random.**  
  *Explanation:* PSM is used when treatment assignment is *not* random. It tries to mimic randomization by balancing covariates.
* **4. The data follows a linear distribution.**  
  *Explanation:* PSM does not assume any specific distribution of the data; it focuses on balancing covariates.

**Question 4:**

What is the primary role of a Directed Acyclic Graph (DAG)?

1. To show temporal relationships between variables.
2. To identify confounding variables.
3. To test statistical significance.
4. To visualize causal relationships.

**Correct Answers:**

* **2. To identify confounding variables.**  
  *Explanation:* DAGs help detect variables that influence both the treatment and the outcome, which need to be adjusted to avoid bias.
* **4. To visualize causal relationships.**  
  *Explanation:* DAGs graphically represent assumed causal links, helping researchers clarify their hypotheses.

**Incorrect Answers:**

* **1. To show temporal relationships between variables.**  
  *Explanation:* DAGs represent causation, not necessarily time-based relationships. Temporal order is sometimes implied but not always depicted.
* **3. To test statistical significance.**  
  *Explanation:* DAGs do not test significance; they are conceptual tools to guide analysis.

**Question 5:**

Which of the following are challenges in causal inference?

1. Measurement error.
2. Confounding variables.
3. Selection bias.
4. Lack of statistical software.

**Correct Answers:**

* **1. Measurement error.**  
  *Explanation:* Incorrect or inconsistent data can distort causal relationships, leading to biased estimates.
* **2. Confounding variables.**  
  *Explanation:* Confounders influence both the treatment and outcome, making it hard to isolate the true causal effect.
* **3. Selection bias.**  
  *Explanation:* Bias occurs when the sample is not representative of the population, invalidating causal conclusions.

**Incorrect Answer:**

* **4. Lack of statistical software.**  
  *Explanation:* While tools are essential, the primary challenges in causal inference arise from data and methodology, not software availability.

**Question 6:**

What is the primary benefit of using Instrumental Variables (IV)?

1. It eliminates unobserved confounding.
2. It ensures random treatment assignment.
3. It balances the treatment and control groups.
4. It predicts future outcomes.

**Correct Answer:**

* **1. It eliminates unobserved confounding.**  
  *Explanation:* IV addresses endogeneity by isolating the causal effect of treatment through an external instrument.

**Incorrect Answers:**

* **2. It ensures random treatment assignment.**  
  *Explanation:* IV does not involve randomization but rather adjusts for non-random treatment assignment.
* **3. It balances the treatment and control groups.**  
  *Explanation:* Balancing is the focus of PSM, not IV analysis.
* **4. It predicts future outcomes.**  
  *Explanation:* IV estimates causal effects, not predictions.

**Question 7:**

Which of the following best describes the term "endogeneity"?

1. Random assignment of treatment.
2. A variable correlated with both treatment and outcome.
3. Measurement error in the outcome variable.
4. A perfectly balanced dataset.

**Correct Answer:**

* **2. A variable correlated with both treatment and outcome.**  
  *Explanation:* Endogeneity arises when unobserved factors influence both treatment and outcome, leading to biased estimates.

**Incorrect Answers:**

* **1. Random assignment of treatment.**  
  *Explanation:* Random assignment eliminates endogeneity by balancing confounders.
* **3. Measurement error in the outcome variable.**  
  *Explanation:* Measurement error can affect analysis but is not synonymous with endogeneity.
* **4. A perfectly balanced dataset.**  
  *Explanation:* Balance helps reduce bias but does not define endogeneity.

**Question 8:**

What does the term "counterfactual" refer to in causal inference?

1. Observed data from the treated group.
2. Hypothetical outcomes in the absence of treatment.
3. The statistical model used to predict outcomes.
4. Temporal trends in the data.

**Correct Answer:**

* **2. Hypothetical outcomes in the absence of treatment.**  
  *Explanation:* Counterfactuals estimate what would have happened if the treatment had not been applied.

**Incorrect Answers:**

* **1. Observed data from the treated group.**  
  *Explanation:* Counterfactuals are hypothetical, not observed data.
* **3. The statistical model used to predict outcomes.**  
  *Explanation:* Models are tools to estimate counterfactuals, but they are not counterfactuals themselves.
* **4. Temporal trends in the data.**  
  *Explanation:* Counterfactuals are not related to time trends but to alternative scenarios.

**Question 9:**

Why is randomization critical in causal inference?

1. It ensures balanced treatment and control groups.
2. It eliminates the need for adjustment.
3. It guarantees statistical significance.
4. It reduces unobserved confounding.

**Correct Answers:**

* **1. It ensures balanced treatment and control groups.**  
  *Explanation:* Randomization creates comparable groups, making causal effects easier to identify.
* **4. It reduces unobserved confounding.**  
  *Explanation:* By balancing unobserved confounders, randomization reduces bias in causal estimates.

**Incorrect Answers:**

* **2. It eliminates the need for adjustment.**  
  *Explanation:* While randomization reduces bias, adjustment may still be needed for precision.
* **3. It guarantees statistical significance.**  
  *Explanation:* Randomization does not guarantee significance; it only improves validity.

**Question 10:**

Which of the following scenarios demonstrates a confounding variable?

1. Studying the effect of smoking on lung cancer, but age is also associated with both.
2. Observing that ice cream sales and drowning incidents are correlated without considering temperature.
3. Using a placebo to separate treatment effects from psychological effects.
4. Comparing accident rates between two factories, one of which has better safety measures.

**Correct Answers:**

* **1. Studying the effect of smoking on lung cancer, but age is also associated with both.**  
  *Explanation:* Age influences both smoking habits and cancer risk, confounding the relationship between smoking and lung cancer.
* **2. Observing that ice cream sales and drowning incidents are correlated without considering temperature.**  
  *Explanation:* Temperature is a confounding variable that impacts both ice cream sales and drowning rates.

**Incorrect Answers:**

* **3. Using a placebo to separate treatment effects from psychological effects.**  
  *Explanation:* This is an example of controlling for a placebo effect, not confounding.
* **4. Comparing accident rates between two factories, one of which has better safety measures.**  
  *Explanation:* This example lacks randomization, but safety measures are not confounders; they are part of the treatment effect.

**Question 11:**

What is the primary assumption of the Potential Outcomes Framework in causal inference?

1. Outcomes under treatment and control are independent of each other.
2. Treatment assignment is random.
3. Counterfactuals are observable.
4. The same treatment effect applies to all individuals.

**Correct Answer:**

* **1. Outcomes under treatment and control are independent of each other.**  
  *Explanation:* The framework assumes that the potential outcomes for one individual are not affected by the treatment status of another.

**Incorrect Answers:**

* **2. Treatment assignment is random.**  
  *Explanation:* While randomization simplifies causal inference, it is not an assumption of the Potential Outcomes Framework itself.
* **3. Counterfactuals are observable.**  
  *Explanation:* Counterfactuals are hypothetical and cannot be directly observed.
* **4. The same treatment effect applies to all individuals.**  
  *Explanation:* This assumption is not required; treatment effects can vary across individuals.

**Question 12:**

Which of the following is an example of a placebo effect in causal studies?

1. Patients receiving a sugar pill report reduced symptoms.
2. A control group experiences no change in symptoms.
3. A new drug has the same effect as the placebo.
4. Random assignment eliminates bias in the study.

**Correct Answer:**

* **1. Patients receiving a sugar pill report reduced symptoms.**  
  *Explanation:* The placebo effect occurs when individuals experience an effect due to their belief in the treatment, not the treatment itself.

**Incorrect Answers:**

* **2. A control group experiences no change in symptoms.**  
  *Explanation:* This describes a lack of effect in the control group, not the placebo effect.
* **3. A new drug has the same effect as the placebo.**  
  *Explanation:* This suggests the drug is ineffective but does not define the placebo effect.
* **4. Random assignment eliminates bias in the study.**  
  *Explanation:* Random assignment improves study validity but is unrelated to the placebo effect.

**Question 13:**

When should Regression Discontinuity Design (RDD) be used?

1. When there is a clear cutoff for treatment assignment.
2. When randomization is not possible.
3. When treatment effects are linear.
4. When there are multiple confounders.

**Correct Answers:**

* **1. When there is a clear cutoff for treatment assignment.**  
  *Explanation:* RDD analyzes outcomes at a specific cutoff, such as age or income thresholds.
* **2. When randomization is not possible.**  
  *Explanation:* RDD is a quasi-experimental design used when randomization cannot be implemented.

**Incorrect Answers:**

* **3. When treatment effects are linear.**  
  *Explanation:* RDD does not assume linear effects; it can handle non-linear relationships.
* **4. When there are multiple confounders.**  
  *Explanation:* While RDD addresses confounding around the cutoff, it is not specifically designed to handle multiple confounders.

**Question 14:**

Which of the following outcomes can result from selection bias?

1. Overestimation of treatment effects.
2. Underestimation of treatment effects.
3. Accurate estimation of causal effects.
4. Reduced generalizability of results.

**Correct Answers:**

* **1. Overestimation of treatment effects.**  
  *Explanation:* Selection bias can inflate the apparent effect of treatment if more favorable cases are included in the treatment group.
* **2. Underestimation of treatment effects.**  
  *Explanation:* Bias can also underestimate effects if key groups are excluded or underrepresented.
* **4. Reduced generalizability of results.**  
  *Explanation:* Selection bias limits the ability to generalize findings to the broader population.

**Incorrect Answer:**

* **3. Accurate estimation of causal effects.**  
  *Explanation:* Selection bias distorts estimates, making accurate causal inference unlikely.

**Question 15:**

What is the key role of backdoor adjustments in causal inference?

1. To control for confounding variables.
2. To establish a causal pathway.
3. To simplify data visualization.
4. To randomize treatment assignment.

**Correct Answer:**

* **1. To control for confounding variables.**  
  *Explanation:* Backdoor adjustments involve conditioning on confounders to block spurious paths between treatment and outcome.

**Incorrect Answers:**

* **2. To establish a causal pathway.**  
  *Explanation:* Backdoor adjustments remove spurious paths, not establish new ones.
* **3. To simplify data visualization.**  
  *Explanation:* Adjustments are analytical techniques, not visualization tools.
* **4. To randomize treatment assignment.**  
  *Explanation:* Randomization is separate from backdoor adjustments.