Experimental Design Questions

1. What are the key components of a well-designed experiment in data science?
    - Discuss the importance of randomization, control, and replication.
2. Why is it essential to define a clear hypothesis before designing an experiment?
    - How does a hypothesis guide the choice of independent and dependent variables?
3. What is the difference between observational studies and experimental studies?
    - Give an example of each in the context of data science.
4. In what scenarios might you prefer a quasi-experimental design over a purely experimental one?
    - Explain the trade-offs in terms of control and feasibility.

Causality and Correlation Questions

1. "Correlation does not imply causation." Explain this statement with an example.
    - Why is it dangerous to infer causation without experimental evidence?
2. What is a confounding variable, and how can it affect the interpretation of results?
    - Provide an example of a confounder in a dataset.
3. How can experimental design help establish causality?
    - Discuss the role of randomization and control groups in isolating causal effects.
4. Suppose you find a strong positive correlation between ice cream sales and drowning incidents. How would you investigate if there is a causal relationship?
5. Why is it important to control for external variables when investigating causality?
    - What methods can you use to account for these variables?

A/B Testing Questions

1. What is the primary objective of an A/B test?
    - How does it help in decision-making for business or product improvements?
2. In an A/B test, what factors should you consider when splitting your users into groups?
    - Why is randomization critical in this process?
3. Suppose you are testing two landing page designs (A and B). How would you determine which one performs better?
    - What metrics would you use, and how would you analyze the results?
4. What are the risks of running an A/B test without a sufficient sample size?
    - How might this impact the validity of your conclusions?
5. Explain why statistical significance is important in interpreting A/B test results.
    - What is the role of p-values and confidence intervals in this context?
6. Describe a situation where running an A/B test might not be feasible or ethical.
    - How would you approach the problem instead?

Control Group and Experimental Group Questions

1. What is the purpose of a control group in an experiment?
    - Why is it essential for establishing causality?
2. How would you design an experiment with control and experimental groups to test the impact of a new ad campaign?
    - Define the independent and dependent variables clearly.
3. What is the difference between a control group and a placebo group?
    - Provide an example from a data science application.
4. Why is random assignment of participants to control and experimental groups important?
    - How does this help minimize bias?
5. In an experiment comparing two website layouts, one group sees Layout A and the other sees Layout B. Which group serves as the control, and why?
6. What are some challenges in creating control and experimental groups for observational data?
    - How can matching techniques (e.g., propensity score matching) address these challenges?

Scenario-Based Questions

1. Imagine you are testing a new recommendation system for an e-commerce site. How would you set up control and experimental groups?
    - What data would you collect to evaluate the system's effectiveness?
2. You observe a correlation between increased email frequency and higher customer churn. Design an experiment to test whether email frequency causes churn.
3. In an A/B test, you find that Group B performs better, but the results are not statistically significant. What steps would you take next?
4. Suppose you are running an A/B test to measure the impact of a discount offer on sales. How would you handle users who received both versions of the offer due to a technical issue?
5. A company wants to test the impact of a new employee training program on productivity. How would you design the experiment?
    - What metrics would you use, and how would you ensure fairness?