

## GB740 Extra Credit Assignment

Carson Batchelor

### Module 1: Distinguishing correlation and causation; Causal vs. Prediction problems

- Question 1:
  - A company observes a positive correlation between employee satisfaction scores ( $y_i$ ) and productivity ( $y_i(0)$ ) but suspects this relationship might be influenced by a confounding factor such as manager quality ( $z_i$ ). Using the potential outcomes framework:
    - Propose a strategy to isolate the causal effect and describe how it addresses confounding.
- Answer:
  - Randomize manager assignments to create variation in employee satisfaction independent of confounders like manager quality.

### Module 2: Basics of RCTs and A/B tests; Randomization, Treatment Effects, Confidence Intervals

- Question 2: (Could use question 1 or 2)
  - An e-commerce company conducts an A/B test where the treatment group sees a redesigned homepage and the control group sees the original.
    1. Define the "units of randomization" and explain why their selection is crucial for the validity of the test.
    2. Before analyzing the difference in conversion rates, outline how you would use pretreatment variables to check for balance, and describe why balance is critical for interpreting the test results.
- Answer:
  - Units of randomization are customers who can be randomly assigned and have measurable outcomes.
  - Compare pretreatment variables between groups to ensure they are similar; balance ensures observed effects are due to the treatment, not confounders.

### Module 3: Experiments with multiple sub-groups; Analyzing experiments with regressions

- Question 3: (Could use question 1 or 2)
  - In an experiment testing a new feature across various customer age groups, you are tasked with analyzing whether the treatment effect varies by age.

1. What approach would you use to check for heterogeneous treatment effects (HTEs) by age group, and how does this help in understanding the impact of the treatment on different subgroups?
2. If you wanted to improve the precision of the treatment effect estimate for each age group, how could you apply demeaning to the outcome variable?

- Answer:

- Include an interaction term between age and the treatment variable in a regression model to estimate how treatment effects differ by age group.
- Subtract the mean outcome of each age group from individual outcomes to remove group-level variations, improving the precision of treatment effect estimates.

#### Module 4: What can you conclude? Sample size, variance, and power

- Question 4:

- Your experiment did not detect a significant treatment effect, but your sample size was small. How could this impact your conclusions, and what steps would you take next to ensure that the lack of significance isn't due to insufficient power?
  - What would you do to increase the power of your experiment and improve the precision of your estimate? How would increasing your sample size help with this?

- Answer:

- To increase power, increase the sample size to reduce the standard error, which in turn improves the precision of the estimate and lowers the risk of Type II error.

#### Module 5: Ethics, perception, and buy-in for experiments

- Question 5:

- Stakeholders are concerned about withholding a potentially beneficial treatment from the control group in an RCT. How would you address these ethical concerns, ensuring that the experiment respects legal and ethical principles?
  - How would you ensure informed consent is properly obtained, ensuring participants fully understand the risks and benefits before opting into the experiment?

- Answer:

- Informed consent should be obtained from all participants, providing them with full transparency about the study's goals, risks, and their right to withdraw without penalty.
- Question 6:
  - What ethical considerations should you evaluate when designing an experiment that involves vulnerable populations?
- Answer:
  - Ensure informed consent, minimize harm, and prioritize the well-being of participants over the experiment's goals.

## Module 6: Field experiments; Real-world challenges

- Question 7:
  - What would you do if participants in a field experiment start sharing information about the treatment with those in the control group, and how would you minimize potential contamination in the study?
    - In what ways does non-compliance affect the reliability of treatment effects?
- Answer:
  - Non-compliance could cause contamination, as participants in the treatment group might not fully comply with the treatment.
- Question 8:
  - If logistical challenges prevent you from randomizing treatment in a field experiment, what alternative approaches could you take to estimate causal effects, and how would you address potential biases?
    - What is the difference between treatment on the treated (ToT) and intention to treat (ITT), and when would you prefer to use each?
- Answer:
  - The difference between ToT and ITT is that ToT focuses on the individuals who actually received the treatment (compliers), while ITT includes all individuals who were randomized to the treatment group, regardless of whether they took the treatment.

## Module 7: Natural experiments; Event studies; Difference-in-Differences

- Question 9:
  - In a difference-in-differences (DID) study, how would you test whether the parallel trends assumption is valid before implementing the DID model?
- Answer:
  - Compare pre-treatment trends for the treatment and comparison groups to check if they are similar.
  
- Question 10:
  - In an RDD study, why might you focus only on observations close to the threshold?
- Answer:
  - Observations near the threshold are most likely to be comparable, minimizing bias from unobserved confounders.