Name: _____

In-Class Activity: Experiments II

PS200C: Causal Inference in the Social Sciences
Spring 2026

Scenario. You are running a small GOTV experiment in a county. You have $N = 800$ registered voters, organized into 200 households of 4 voters each. For each voter you observe `past_turnout`, an index of voting in recent elections that is a strong predictor of whether they will vote next time. You will randomize a mailer treatment D and measure $Y \in \{0, 1\}$, whether each voter votes.

1. **Variance, not bias.** A colleague says: "Since random assignment makes the difference in means unbiased, we don't need to bother adjusting for `past_turnout`." In one or two sentences, are they right? Why or why not?

2. **Block, adjust, or both?** Would also *blocking* on quartiles of `past_turnout` help, on top of regression adjustment? When does blocking add the most?

3. **Clustering.** Suppose for logistical reasons you have to send the mailer to whole households at a time — everyone in a household either gets $D = 1$ or everyone gets $D = 0$.
   (a) Should you cluster your standard errors? On what unit?

   (b) Why? Tie your answer to the principle from lecture.

4. **What does an SE actually tell you?** If a standard error is properly estimated, what real thing about your experiment and analysis is it telling you? (One or two sentences — connect the SE to the distribution of $\hat{\tau}$ across hypothetical repetitions of the same experiment.)