2. Peruvian Recycling

Look at this article about encouraging recycling in Peru. The paper contains two experiments, a "participation study" and a "participation intensity study." In this problem, we will focus on the latter study, whose results are contained in Table 4 in this problem. You will need to read the relevant section of the paper (starting on page 20 of the manuscript) in order to understand the experimental design and variables. (*Note that "indicator variable" is a synonym for "dummy variable," in case you haven't seen this language before.*)

1. In Column 3 of Table 4A, what is the estimated ATE of providing a recycling bin on the average weight of recyclables turned in per household per week, during the six-week treatment period? Provide a 95% confidence interval.

   ATE is the same as what is mentioned as the coefficient estimate by the model at 0.187 Confidence interval 0.251 and 0.123

2. In Column 3 of Table 4A, what is the estimated ATE of sending a text message reminder on the average weight of recyclables turned in per household per week? Provide a 95% confidence interval.

   ATE is the same as what is mentioned as the coefficient estimate by the model at -0.024 Confidence interval 0.054 and 0.054

3. Which outcome measures in Table 4A show statistically significant effects (at the 5% level) of providing a recycling bin?
   - Percentage of visits turned in a bag
   - Avg. no. of bins turned in per week
   - Avg. weight (in kg) of recyclables turned in per week
4. Which outcome measures in Table 4A show statistically significant effects (at the 5% level) of sending text messages?

   None

5. Suppose that, during the two weeks before treatment, household A turns in 2kg per week more recyclables than household B does, and suppose that both households are otherwise identical (including being in the same treatment group). From the model, how much more recycling do we predict household A to have than household B, per week, during the six weeks of treatment? Provide only a point estimate, as the confidence interval would be a bit complicated. This question is designed to test your understanding of slope coefficients in regression.

   $Y_{AB} = [\beta_{A1}B + \beta_{A2}S + \lambda Ybl_A + P_A + \alpha_j + \epsilon_i] - [\beta_{B1}B + \beta_{B2}S + \lambda Ybl_j + P_B + \alpha_j + \epsilon_i]$ represents the difference in model output for households A and B

   α_j cancels out because of fixed street effects that are same for both households A and B

   ϵ_i cancels out because experiments are randomized and observations are identical and independently distributed (i.i.d), hence we can assume error terms are homoskedastic (constant variance of error terms).

   B, S, P_i are indicator variables, hence cancel out between A and B households (as they are identical being in the same treatment groups)

   Therefore, we only have difference in model prediction due to λYbl_{AB} only.

   Which can be calculated as 0.281×2 that comes to 0.562

6. Suppose that the variable "percentage of visits turned in bag, baseline" had been left out of the regression reported in Column 1. What would you expect to happen to the results on providing a recycling bin? Would you expect an increase or decrease in the estimated ATE? Would you expect an increase or decrease in the standard error? Explain our reasoning.

7. In column 1 of Table 4A, would you say the variable "has cell phone" is a bad control? Explain your reasoning.

8. If we were to remove the "has cell phone" variable from the regression, what would you expect to happen to the coefficient on "Any SMS message"? Would it go up or down? Explain your reasoning.