ARE 261
Fall 2021
Assignment 1 : Experimental Research Designs
Due Sept 20

PART 1: RCT/RED: post-treatment data only

Data set (stata): PS1_data.dta: Key variables: customer identifier; monthly electricity consumption (kwh); year/month id, 24 months of consumption data for 1000 customers.

You have been contacted by RCT Power and Light (RCTP&L) about helping them design a field experiment. Under a recent PUC ruling, RCTP&L is now required to reduce electricity usage in its service territory to 10 percent of forecast levels by 2025. If the utility can demonstrate that consumers reduce demand in response to intervention, they can count this program towards compliance with this regulatory compliance.

RCTP&L are proposing to run a pilot program to help them estimate the average effect of this intervention on monthly household electricity consumption. RCTP&L plans to provide their customers with following:

- An in-home monitor that captures and displays real time energy consumption (in terms of dollars and kWhs),
- An easy-to-use phone app that helps customers make sense of their consumption patterns and develop a strategy for reducing their energy consumption.

Pat the project manager has come to you with a research design and is hoping you can help him figure out how many customers he will need in the study. The proposed research design proceeds as follows:

(1) Randomly select N customers from the larger population of residential customers.
(2) Randomly assign these customers to either a control group or the treatment group.
(3) Contact treated households with the offer of free devices
(4) Collect monthly billing information for both the treated and control groups (i.e. N customers) over a two-year period following the introduction of the treatment in the treated group.
(5) Compare consumption patterns in the treated and control group.

Pat is asking you how big the sample should be in order to detect a treatment effect of 5% with reasonable confidence. He thinks customers would be crazy not to jump at this offer, and instructs you to assume all "treated" households will receive the intervention. Assume you will have 24 months of *post-treatment* data to work with.

1. Use the sample of customer billing data provided (and described at the end of this assignment) to estimate how big the sample of customers would need to be to detect a treatment effect of 5% with reasonable confidence.
    a. Begin by clearly stating the estimating equation you have in mind.

b. Given this equation, compute the required sample size using the analytic formulae provided in the lecture notes on power calculations.

2. Now repeat the power analysis you conducted in part (2) using the data provided and a **simulation-based approach**. In other words, simulate the estimation exercise using different sample sizes until you achieve the desired statistical power.
    a. Please explain the specific steps you take to conduct this analysis (and provide the associated code).
    b. Do you reach the same conclusion (i.e. the same sample size) as in part 2? If no, offer some explanation as to why not.

3. Pat takes the power analysis you did in parts 2 and 3 to the RCT P&L board. They are rightly concerned about Pat's assumption of 100% take up. They point to other exciting residential programs that have take-up rates around 10%. You explain that you can redo the power calculations to reflect a 10% take up rate. You also suggest an "opt-out" design where customers receive these devices in the mail by default. This could help engage a larger share of customers.

    Using the RED framework we discussed in class, modify the power analysis conducted above (you can choose to modify either the analytical or simulation-based approach) to account for the opt-in and opt-out alternatives, respectively. Assume an opt-in rate of 10%. Assume an opt-out rate of 50%. By how much does the required sample size increase under these two designs?

4. You present your cost and power calculations in part 4 to the energy efficiency division at RCTP&L. They want to know if you will be able to test whether the response from consumers who choose to participate is different from the customers who were nudged/defaulted into the program. Can you separately identify these effects with your design?

5. Think about some of the assumptions and complications we discussed in class: the monotonicity assumption; SUTVA; the Hawthorne effect. Which (if any) of these might you be concerned about in this empirical setting?

PART 2: Panel data!

Data sets(csv): pecanstreet_daily.csv; pecanstreet_monthly.csv (variables are hhid, year, month, day, kwh)

Another utility contacts you about designing a small field experiment. They have equipped a non-random sample or 97 households with smart meters. These households eagerly opted-into this experiment. They now want to equip these households with enabling devices that provide real-time feedback by displaying smart-meter data on an in-home device.

The electric data will be aggregated to the daily level, and they will give you one year of pre-treatment data and one year of post-treatment data. They want to know the minimum effect that they could detect with this sample. Assume full compliance with treatment (i.e. everyone who is offered the free monitor will take it).

a. You have data aggregated to the daily and monthly level. Write down your estimating equations for both the daily and monthly data, respectively.

b. Using the sample of customer billing data provided (daily and monthly), estimate the minimum detectable effect on electricity consumption using a simulation-based power analysis. Assume a sample of 97 households with one year of pre-treatment data and one year of post-treatment data. Comment on the relative magnitude of the MDEs you estimate for the daily versus monthly data.