Block Party Project

The data set is from a leading coalition loyalty program (LP). The LP has over 10 million members representing over 67% of the households in country where it operates. As a coalition loyalty program, members collect miles at over 100 sponsors categories covering all aspects of purchases ranging from groceries to gasoline to apparel and credit card purchases. Collected miles can be exchanged for rewards ranging from travel to merchandize to discount coupons. The LP receives money from sponsors whenever a mile is issued, and so a core goal of the LP is to have members accumulate as many miles as possible.

The LP launched a social media website for members to discuss the program and benefits. Member posts can be linked to their purchases, providing a unique opportunity to measure the effect of participation in social media forums on purchase behaviors. Posting on the social media site is sporadic, with few posts on most days–the median number of posts per day is only 7 and the third quartile is 27. There are, however, large “spikes” in activity: the maximum number of posts on a single day was 6465. The spikes are driven by email promotions, e.g., on a certain date the LP sent an email announcing the one-year anniversary “Block Party,” in which members were offered up to 4 chances to win 25,000 reward miles by becoming a member, posting a tip, uploading a picture, or making an “I like this” thumbs up to the community site. This contest gave an opportunity for members to share their personal experience with the LP or listen to other member’s experiences, which transforms customers from “passive audiences” to “active players.”

You are to answer the following questions:

* Does participation in the Block Party contest increase subsequent purchases (i.e., mile accumulation)?
* Does the amount of *elaboration* (number of words written) affect subsequent purchases (mile accumulation)?
* For how long does the participation / elaboration affect persist?

The LP provided the mile accumulation history for a sample of 143,000 members, with all participants in the Block Party contest and a random sample of non-participants. Block Party participants form the treatment group and non-participants are the control group. Having a control group enables us to control for threats to internal validity such as history. We divide time into two periods: the **pre-period** is prior to the Block Party and the **post-period** is after the contest was announced. You have the following variables for 12302 LP members:

* Participate: dummy equaling 1 if member participated in Block Party, 0 if not
* Elaboration: number of words written in Block Party entry (0 for non-participants)
* Prefood: number of miles accumulated in pre-period from food category, e.g., grocery
* Pregas: number of miles in pre-period from gas stations
* Prebank: number of miles in pre-period using LP bank cards (e.g., branded credit card)
* Preretail: number of miles in pre-period from retail stores
* Preother: number of miles in pre-period from all other LP partners
* Rfreq: number of reward redemptions in pre-period
* Mile1: number of miles in the first week of the post-period.
* Mile2: number of miles in the second week of the post-period.
* Mile3: number of miles in the third week of the post-period.
* Mile4: number of miles in the fourth week of the post-period.
* Basemile: total miles in pre-period (sum of pre-food, …. Preother)

The dependent variables in your analysis should be mile1-mile4. Pre-measures account for customer heterogeneity when comparing those who participate with those who do not. You will want to include them as controls.

Although members self-select into participating in the contests, I did further matching with a *propensity scoring model*, where, for each participant I found a “twin” who did not participate but was a similar as possible to the participant in terms of their mile accumulations *prior to the Block Party contest.* In general, those who participated were better members of the LP, and our “twin matching” helps us to avoid this selection bias (where participants and non-participants are systematically different prior to the contest).

You may work in groups between one and five students on this project. Perform the analysis and write it up as a report. Follow the siegel.pdf chapter on how to do the writing. The project is due on May 5. Here are a few hints:

* Remember Tukey’s “first-aid” (usually good to log counts and amounts). You should especially worry about variance stabilizing transformations.
* You cannot use both the individual pre-period categories (preretail, prebank, etc.) and basemiles, since including both would create perfect multicollinearity. You should try the categories and basemiles in separate models to see how robust your findings are.
* Do not tell me everything you did. Tell me only the analyses/results to answer the three questions.
* Do not share this data with anyone and delete it after the class is over.