Perspectives on Computational Analysis (Fall 2018)

Assignment 5

Problem 1

a) I selected the experiment named ’20-25 Minute Auction Experiment’, which was posted by Wladislaw Mill. Many of the other HIT tasks classified themselves as ‘experiments’, but resembled surveys. I could discern no clear ‘treatment’ being introduced. Auctions provide an interesting avenue to assess hypotheses in behavioural research. Hence, I selected this task.

b) As part of the payment structure, the worker is paid $0.50 for completing the experiment. An additional ‘average’ bonus of $2.10 may be provided based on performance.
The specific meaning of performance is not provided. The time allotted is 300 minutes. This may imply that performance is the number of times the worker can complete the experiment within the time limit.
Furthermore, the researcher uses the term ‘average’ bonus to mean the average of the minimum and maximum values ($0.43 and $10 respectively)- which is not the statistical meaning of the term.
The payments will be determined and paid out via Amazon Mechanical Turk within a week of task completion.

c) The experiment involves specific eligibility requirements. Firstly, HIT Approval Rate must be greater than 97%. Secondly, the worker must be located in the US. Thirdly, total approved HITs is greater than 500.
The first two restrictions are relatively straightforward and can be seen on other similar tasks. The third one seems to imply that the researcher wants to ensure a certain level of past experience and observed work quality to complete this task.

d) The task itself takes between 20-25 minutes. Thus the implied hourly rate is thus between 0.50*3 and 0.50*2.4, which amounts to between $1.2 and $1.5. This is not taking into account the bonus payment structure outlined earlier.

e) The expiration date of the task is 7 days from now, which is 11/19/2018.

f) If million people participated in the study, and finished only the core task of the experiment, it would cost 1million*0.50, or $500,000. However, if all of those workers reached the maximum possible bonus ($10), the cost could rise to $10 million.

Problem 2

These answer draw extensively on material from Costas and Kahn (2013).

(a) **State the research question of this paper in the form of a question and in one sentence**

The research question in this paper would be-"Are political liberals/ environmentalists more likely than political conservatives/non-environmentalists respectively to respond to energy conservation nudges?"

Phrased as a sentence, this paper's focus has been to estimate the differential response of environmentalists and non-environmentalists to receiving a Home Energy Report, as well as the differential response between political liberals and conservatives to the normative message that was included with the first report.

(b) **The data for this study came from at least two sources. Name the sources, and describe the data**.

The principal sources were the residential billing data (from January 2007 to October 2009, provided by Positive Energy/OPOWER), treatment and control data as well as individual voter registration and marketing data (for March 2009, purchased from www.aristotle.com). Though not explicitly stated, the paper suggests that the researchers had availed of data at the census block level. Finally, they also had access to an ancillary survey conducted by Positive Energy in 2009.

The billing data information provided information kilowatt hours purchased per billing cycle, the length of the billing cycle (measured in days), whether the house uses electric heat, and whether the household is enrolled in the electric utility's program to purchase energy from renewable sources.

The Treatment and Control Data contained information on the start date from when household began to receive the Home Energy Reports, information on square footage of the house, on whether the home uses electricity or natural gas for heating, and the age of the house.

The voter registration and marketing data revealed (whenever available) each individual's party affiliation, and whether the individual donated to environmental organizations.

The data at the census block level included information on income, education level, the number of families in a single home and age of the head of the household. Finally, the ancillary Positive Energy survey asked questions about the Home Energy Report, and was limited to 1,375 households who had actually received such reports.

(c) **Define and describe the control group and the treatment group in this study. What was the treatment?**

The study was conducted on household-level outcomes, but determined its treatment and control groups at the block level. Given the experimental nature of the study, the treatment and control groups were defined deliberately by the researchers in collaboration with Positive Energy (rather than through some natural experiment or post-hoc analysis). Further, as a randomized experiment, the allocation of a household to either group was not based on their pre-existing socioeconomic or demographic characteristics, etc. Nonetheless, summary statistics for these groups (in terms of all the variables described in the previous answer) are available in Costas and Kahn (2013: 687).

Randomly selected batches of five contiguous census blocks were randomly assigned in alternating moves to either the treatment or control group. In this manner, 35,000 homes were assigned to each. The remaining 14,000 homes were then assigned to the control group. After subtracting 1,976 households whose residential billing data could not be matched with pilot and control data, the final compiled data set included 33,664 and 48,058 and households in the treatment and control group respectively.

Both the treatment and control households shared specific characteristics so as to be eligible for this study. At the block level, households were drawn from 85 census tracts with a high density of single-family homes. At the household level itself, all households were required to have a current account with the electric utility that had been active for at least one year, could not be living in apartment buildings, and had to be living in a house with square footage between 250 and 99,998 square feet.

The treatment in this study was a two-page Home Energy Report (HER) sent by Positive Energy (the electric utility company) to its client households. Each report included the absolute level of consumption as well as one of three normative messages: “great”, “good”, or “room for improvement”, based on the household’s consumption in comparison to that of 100 neighbors living in similar-sized houses. Perhaps one of the lacunae in this study is that the treatment was not delivered in an identical manner for all members of the treatment group- 24,028 received a monthly report while 9,636 received a quarterly report

(d) **Beyond the previous work of Schultz et al. (2007), what extra layer of participant heterogeneity did Costa and Kahn control for in order to answer their research question?**

Schultz et al (2007) had only considered the different methods of delivering its ‘nudge’ (with injunctive norms and emoticons) and estimated differential effects thereof based on participant heterogeneity in terms of their energy consumption before the treatment. In this regard, they had detected a boomerang effect for the two groups- those above and below the median level of consumption.

For Costa and Kahn (2013), the extra level of heterogeneity arose from the ideologies of participants in the experiment - both political (liberal or conservative) and towards the environment- - whichwas not controlled for in Schultz et al (2007).

(e) **What was Costa and Kahn's finding?**

Costa and Kahn (2013) found that own ideology- whether measured by political party affiliation, donations to environmental organizations, or the purchase of green energy- was associated with differential treatment effects. Their findings suggest that environmental nudges are most effective in relatively liberal communities.

Liberal households were seen to be both less likely to drop out of the experiment and more likely to report that they like receiving the report than political conservatives. As a response to the nudge, both liberals and conservatives reduced their electricity consumption. However liberals were observed to reduce their electricity consumption by a larger proportion than the conservatives.

Similarly, users who had previously purchased energy from renewable resources and donated to environmental organizations reduced their consumption by 0.9% and 1.1% respectively in response to the nudge relative to those did not.

PROBLEM 3

A new experiment aims to estimate the effect of receiving text message reminders on vaccination uptake. One hundred and fifty clinics, each with 600 eligible patients, are willing to participate. There is a fixed cost of $100 for each clinic you want to work with, and it costs $1 for each text message that you want to send. Further, any clinics that you are working with will measure the outcome (whether someone received a vaccination) for free. Assume that you have a budget of $1,000.

(a) **Under what conditions might it be better to focus your resources on a small number of clinics and under what conditions might it be better to spread them more widely?**

Salganik (2018) speaks of four key considerations when designing an experiment- cost, control, realism and ethics. The approach we are using here relies on fits the 'Partner with the Powerful' model. This particular question focuses on the cost dimension, keeping the other three in consideration. We see that there are only fixed costs (adding a clinic, sending an SMS for the treatment group) and no major variable costs. Hence, our focus will remain primarily on control while staying within the budget.

Furthermore, while designing this study, it is important to state our **assumptions**. We refer to the SUTVA (Stable Unit Treatment Value Assumptions)

Firstly, there are no Spillovers. We assume that an individual is not impacted by the treatment given to other people. In the context of our study, this assumption could prove highly tenuous and contestable. Patients at the same clinic are likely to directly forward the SMS or discuss its contents to friends or peers if they find its content compelling or relevant enough.

Secondly, there are no Hidden Treatments. In other words, the only relevant treatment is the one that we as the researchers are delivering. We could assess Gerder and Green's (2012) question on 'breakdown in symmetry', i.e. whether anything other than the treatment itself causes people in the treatment and control groups to be treated differently. Specifically in our study, there does not seem to be any such feature to the study

Other than this, there are also other assumptions that all participants have working mobile phones, they check them regularly and that the SMS is not automatically relegated to spam folders where it will most likely go unnoticed.

Using the Potential Outcomes Framework, we can focus on estimating the Average Treatment Effect, without needing to estimate the treatment effect for any particular participant in the study.

If we were to involve all eligible patients and spend primarily on the SMS, then 2-3 clinics alone would suffice. We would have up to 1800 (600*3) eligible participants, with $700-$800 available for SMS costs ($200-$300 spent on the SMS). This would make sense if less than half of the 1800 participants would receive the SMS. This is possible, since as Salganik (2018) emphasizes that a 50% split between Control and Treatment group is ideal if the cost of delivering both the treatment and control group is the same. With $1 spent only on the Treatment Group, this scenario does not merit the concern of an equal split between the groups.

Thus, if all the above assumptions were met, we could focus on a smaller number of clinics. Further, if we were to select a smaller number of clinics, we could control for the clinic-level fixed effects at each individual clinic.

However, spreading the study across may be preferable under specific conditions. Firstly, we would do so if a key assumption from those above was violated. For example, if we expected a high rate of spillover within the participants at one clinic, and were more interested in the effect of the SMS in isolation (rather than due to social contagion), we could potentially mitigate the spillover effects by spreading our participants across several clinics.

Secondly, if we were interested in an analysis at the level of clinics themselves, they could only be detected with a sufficiently large ‘sample’ of clinics. For example, we might want to assess differences in effectiveness of the SMS in different neighbourhoods, in which case a sufficient number of clinics would be needed in each neighbourhood.

(b) What factors would determine the smallest effect size that you will be able to reliably detect with your budget?

We have already committed ourselves to measuring the Average Treatment Effect (ATE). Detection of small effects would be challenging if there already exists a large degree of variability (across the target population) in the outcome that we seek to measure.

If the treatment is very predictive of the outcome variable (both in the treatment and control groups), then a difference-in-differences approach would provide a more precise estimate than a simple difference-in-means model.

As outlined in Salganik, this choice would depend on the level of ‘noise’ or variability in the outcome as already observed in the participants. So if we see immense variability in the vaccination rates across clinics, we would be able to detect small effects only through a difference-in-difference approach, without necessarily increasing the budget.

Furthermore, if any of the SUTVA assumptions are violated- such as the presence of spillover effects and excludability- that are not explicitly included in experimental design, the effect size estimation would be inaccurate and may be missed entirely.

BIBLIOGRAPHY

Alcott, Hunt and Todd Rogers, \The Short-run and Long-run E_ects of Behavioral Interventions: Experimental Evidence from Energy Conservation," American Economic Review, 2014, 104 (10), 3003{3037.

Gerber and Green (2012). *Field Experiments: Design, Analysis, and Interpretation*. New York: W. W. Norton.

Costa, Dora L. and Matthew E. Kahn, \Energy Conservation Nudges and Environmentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment," Journal of the European Economic Association, June 2013, 11 (3), 680{702.

Salganik, Matthew J., Bit by Bit: Social Research in the Digital Age, Princeton University Press, 2018.