Assignment #5: Experiments

Sanittawan Nikki Tan

11/10/2018

Part 1

a) The experiment that interests me is an investment experiment posted by Fynn Gerken. The task involves reading some provided unspecified materials and answering related questions. Because I am interested in financial investment and have had experience in investing, this task seems to suit me.

(b) The full payment structure comprises of a flat rate wage of USD 0.5 per 10 minutes plus a bonus of USD 0.25 per correct answer of a comprehension question. There are three of them, so a bonus is up to USD 0.75 in total.

(c) There are four qualifications. First, the HIT approval rate must be more than 95 percent. Second, a worker must not be granted previous participation. Third, a worker must pass a test that determines his/her investment experience. The test has three questions asking if the participant is older than 18 years old, if the participant has invested in bond or stocks, and if a participant has ever analyzed a stock using financial statements. The final requirement is that the participant must be in the US.

(d) The job takes approximately 10 minutes to complete. The implied hour rate is USD 3 per hour $(60 * 0.5/10)$.

(e) The job expires on Monday, November 26, 2018.

(f) The most that this project can cost the experiment creator is $(0.5 + 0.75) * 1000000 = 1,250,000$. Assume that everybody receives a flat rate payment of USD 0.5 and a full bonus of USD 0.75. If we assume that nobody answers the comprehension questions correctly, the researcher is expected to spend USD $0.5 * 1000000 = 500,000$.

Please refer to this link: https://worker.mturk.com/requesters/A1PAW2HAVVOXLU/projects?ref=w_pl_prvw for the HIT and the photo below.
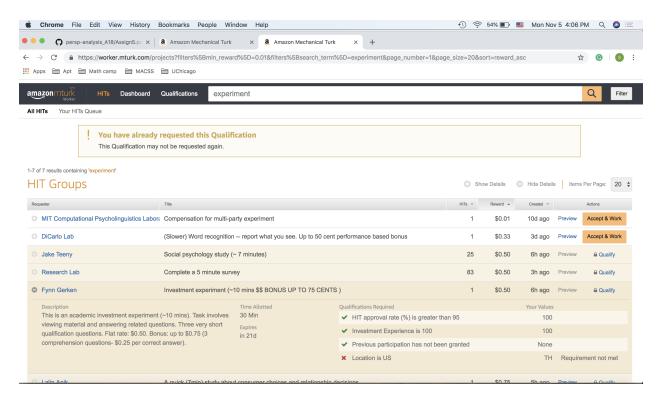
Figure 1: Fynn Gerken's Investment Experiment

Part 2

Building off of previous studies, Costa and Kahn (2013) asked "Does effectiveness of energy conservation"nudges" in the form of Home Energy Report (HER) vary among people with different political ideologies?" in their paper. Following along this question, the authors would like to test two hypotheses: whether liberals who presumably practice voluntary-restraint will reduce energy consumption by a fewer percentage than political conservatives and whether there is a "boomerang effect" among anti-environmentalist conservatives in a sense that they *increase* their consumption to defy their peers. To answer their question and testing the hypotheses, they obtain residential billing data between January 2007 and October 2009 from OPower which details information on households' energy choices and consumption. The second data source is the treatment and control data collected from households that receive or not receive the HERs. The third data source is individual voter registration and marketing data detailing an individual's political affiliation and whether he/she donates to environmental organizations. The authors purchased this data set from third-party aggregator aristotle.com. Costa and Kahn then linked individuals in each data set to construct a merge data set that has the treatment, energy consumption, and political ideology information.

Costa and Kahn leveraged data collected from the HER experiment between March 14 and May 9, 2008, when the utility company sent HERs to approximately 35,000 randomly-selected households. The households that received HERs are in the treatment group (HERs are the treatment) while roughly 49,000 of randomized households, which are in the control group, never received such reports. According to Costa and Kahn, their paper contributes a new perspective to related OPower studies because they examine household's political and environmental ideology which can affect how responsive a household is to the nudges. They included variables such as party registration and green indicators which mean that the household donates to environmental organizations and purchase energy from renewable sources or not. They went beyond Schultz et al. study in 2007 that only looked at the role of descriptive and injunctive norms in affecting above and below average energy users and other follow-up studies that examine home attributes and individual's socio-economic characteristics. Costa and Kahn also included standard characteristics in their model (such as individual, block and house characteristics).

Costa and Kahn found that the effectiveness of nudges to spur energy conservation depends on political and environmental ideologies. Although the energy consumption decreases in both groups, liberals and environmentalists are more responsive than conservatives and anti-environmentalists to change their behaviors after receiving HERs. They also found that households with liberal and environmental ideology tends to decrease their energy consumption by a larger percentage than conservatives. Conservatives households, on the other hand, are more likely to report disliking HERs and drop out of the experiment than their liberal counterparts. It appears that they did not observe boomerang effect in any group as they initially hypothesized.

Part 3

(a) In experimental research design, there are two core assumptions of the potential outcome framework making sure that any observed effects can be attributed to the treatment as Gerber and Green (2012, 38) puts it "each potential outcome depends *solely* on whether the subject *itself* receives the treatment." The first assumption is "no interference" or "no spillovers." The second assumption is excludability. Regardless of the number of clinics chosen into a study, researchers have to avoid violating these assumptions. However, it is often tricky to control for variations in the real field experiments. The location of the clinics could be one condition that determines whether a researcher should focus resources on a small number of clinics rather than spreading them out or vice versa. For example, if the location of the clinics is in a small town with a close-knit community, it might be hard to avoid the interference of treatment effect if the researchers choose to focus on a small number of clinics which means that they can obtain more patients in the treatment and control groups per clinic. Because it is safe to assume that patients know each other quite well in a small community, there may be scenarios where patients in the treatment group unintentionally show text messages to another patient in the control group who is a friend and expose patients in the control group with treatment. In a case like this, researchers may consider spreading the clinics out geographically and reduce the number of patients per clinic so as to minimize the chance of violating no spillover assumption. In contrast, if the clinics are in a large urban area like New York city and geographically spread out across neighborhoods, it may be practical for researchers to focus on a small number of clinics instead. Another condition is the representativeness of the samples compared with the population. Take a case where researchers want to also study how the text messages affect people from different races differently in a highly segregated city where a neighborhood mostly includes people from the same race. If the researcher were to choose only a few clinics to work with, they might end up with clinics with predominantly white or Hispanic patients which are not representative of the population. Another condition that puts focusing resources on a small number of clinics at the advantage is what Gerber and Green (2012, 39) called "asymmetries in measurement." According to Gerber and Green, a different standard of measuring the outcomes, gathering data or handling of the control and treatment groups could jeopardize the excludability assumption. In this case, each clinic that the researcher will select is willing to measure the outcome for free. However, if each clinic were to have different standard of accounting for patients in the treatment group who receive vaccination and the researchers were to choose to focus on a large number of clinics spreading out geographically, the results obtained from each clinic might be systemically different since it is possible that some clinics may overcount or undercount their patients, thus violating the excludibility assumption. Hence, choosing to focus on a small or large number of clinics depends on a variety of conditions, but the goal is to avoid violating the core assumptions of the potential outcome framework.

(b) I think there are two factors that determine the smallest effect size that can reliably be observed given a budget limitation. Firstly, the number of subjects in the treatment and control groups can affect precision which, in turn, affects the effect size. To make the effect size as small and reliable as possible, the researchers has to obtain large enough samples in order to reduce the standard errors which will increase precision. Increasing precision will help us detect smallest effect size because, as Salganik mentioned in his book (Salganik 2018, 209), relatively small effect can be difficult to detect if the data is very noisy, i.e. have large standard errors. Given budget constraints and asymmetric costs of treatment and control, the researchers, in this case, may have to choose to work with fewer clinics so that they can observe a larger number of patients who are in the treatment and control groups in each

clinic to observe a small effect size. The second factor is the choice of estimators. Salganik discussed that difference-in-differences estimator allows researchers to detect a smaller sized effect because the standard errors of difference-in-differences research design are smaller than that of difference-in-means design. In addition, difference-in-differences also control for or difference-out other natural variations among subjects. Thus, the choice of research design is also a factor that determines the smallest effect size.

Gerber, Alan S., and Donald P. Green. Field Experiments: Design, Analysis, and Interpretation. 1st ed. New York: W. W. Norton, 2012.