Boris Garber

W241 – Field Experiments

Essay 2: Experimental Research Proposal

Introduction:

This experiment proposal is motivated by my personal experiences trying to find an apartment. Although it is always challenging to look for a new apartment, it is even more frustrating to do so with a pet. Even after finding advertisements for pet friendly rentals, the tenants/landlords would often change their minds and decide that pets were no longer allowed. Not everyone is an animal lover and pets can certainly cause major damage to property, so it is understandable why a landlord would be hesitant. Especially since pet ownership is not legally protected. However, federal law does prohibit discrimination based on many factors. Even though direct discrimination is difficult to measure and prove, I am curious if an experiment can detect a bias in the rental market.

I am particularly interested in if having a child or being a senior would affect the outcome. Many buildings in NYC are prewar apartment buildings with rent stabilized units. Although both these statuses are protected by housing law, there is strong motivation for landlords to prefer other types of tenants. Having a child in the building poses issues with noise and can pose unique risk due to lead poisoning. Furthermore, families and seniors have strong protection from eviction.

Research Question:

Do specific attributes disclosed in a rental application affect the response rate from the landlord?

Data Collection:

There are many resources online for listing apartments for rent. One of the most popular is craigslist.org, which has postings for many types of apartments for many cities. In order to avoid spam postings the data collection will focus on “by-owner apartments only”. Furthermore the listings requiring a call/txt will be excluded. Finally, listings that are duplicates will be detected and resolved. At this time craigslist does not offer an API for data collection so a web scraping script will be utilized to collect listing data. The data collected will include the following for each posting:

1. Date
2. Subject
3. Neighborhood
4. Contact email

Treatment:

The experiment will include a control email indicating interest in renting the apartment and requesting to set up a viewing. The treatment will include the same bare bones email as the control but will also have a statement identifying the subject as having a treatment condition. The main treatment of interest will be having a child. The experiment can be run with one treatment or with multiple groups of treatments if the sample size is large enough. Another treatment option could be being a senior. Craigslist has updated their search/posting feature to specify if the landlord will accept pets so that treatment is no longer viable. Care must be taken to craft the treatment email so that it is perceived as a plausible inquiry and not a contrived experiment.

Control email example:

Hi, my name is John. I am very interested in renting the apartment. When are you available to set up a viewing?

Treatment examples:

1. I have a baby.
2. I am a senior.

Outcomes:

Five different outcomes are possible:

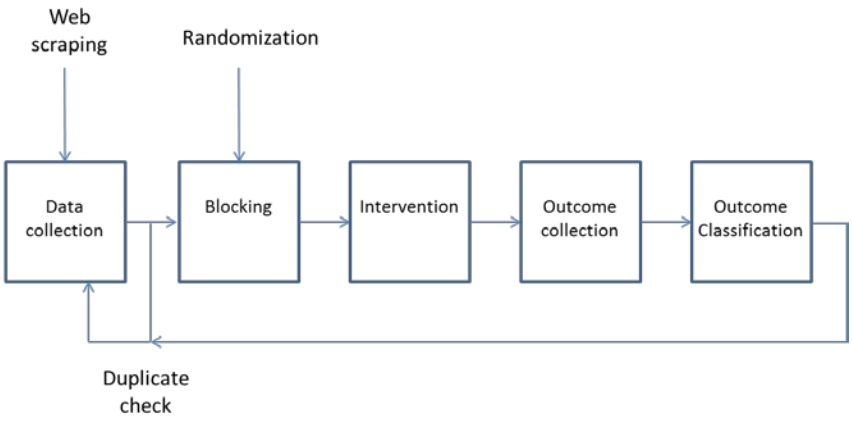
1. Positive reply setting up a viewing
2. Negative reply due to discrimination
3. Negative reply due to reasons other than discrimination
4. No reply due to discrimination
5. No reply due to reasons other than discrimination

Since it will often be impossible to observe if a negative reply or no reply was truly due to discrimination, the positive reply will have to be used as the outcome. A sentiment analysis or logistic regression can be utilized to identify the response emails as positive replies.

Blocking/Clustering:

Blocking will be performed on listings in the same neighborhoods. Since some neighborhoods seem to be trendier than others, landlords may be seeking particular types of tenants. This can cause higher variance in outcomes if it is not addressed. Assigning randomized treatments within blocks would reduce this variance. Since we will be able to individually assign a treatment to each listing, clustering should not be an issue.

Pilot study:



Since this is a group project, the effort can be split into separate functions:

1. Web scraping raw data and data cleaning
2. Blocking, randomization, treatment and emailing
3. Collection and classification of outcomes

Prior to the actual experiment a pilot study should be conducted to verify that all functions of the experimental design can be integrated without errors. Initial data can be collected and used to train the classification models for analyzing emails. Once all data is collected a statistical analysis can be performed.

Statistical Analysis:

Since this experiment will use different email addresses to send out the control and treatment emails, it is conceivable that each subject can be made to receive both emails. In this case both potential outcomes will be observable. However if experimental limitations allow for only one email per subject a randomization inference analysis will be conducted to test the sharp null hypothesis that the treatment effect for all subjects is zero.

Possible issues:

One major issue is the possibility that the poster will not believe the treatment emails and consider them either as fraudulent or entrapment. This may impact their behavior and bias the results. Another issue may be spam posts that are created by listing agents. These can request further contact by phone or text or may automatically schedule anyone who contacts them. In order to address these issues, the treatment emails should be written as naturally as possible so that the poster does not recognize that they are part of an experiment. Spam listings should be avoided before randomization is accomplished otherwise blocking will be less effective.