Experiments and Causality Final Project Progress Report

Chris Mcpherson, Leslie Teo, Nicholas Chen, Paul Varjan
March 7, 2017

**Research Question**
Do exclamation points signal sincerity and friendliness in an email to a prospective landlord? Operationalized, does including an exclamation point (or more) make you more likely to receive a positive response when applying to an apartment?

**Progress Update**
Our research to investigate the effects of punctuation on the success or failure of renters' first communications with potential new landlords has progressed along several dimensions since we first formulated our research question. As will be described in more detail below, we have operationalized phase one of data gathering, researched potential strategies for phase two of data collection, and made slight modifications to our process pipeline.

For phase one of data collection we have developed a web scraper that collects apartment listings from Craigslist for a specified price range, bedroom range, and list of cities. This data collection web scraper has been fully tested and is operational. The scraper is easily capable of pulling thousands of listings, but for reasons to be described below, we have decided to go a different direction than collecting as many thousands of data points as possible.

Our original intent was to automate as much of the data collection and administration of treatment, emailing landlords with varying levels of sincerity (exclamation points), as possible. Ultimately, we did not want to spend hours manually sending emails to thousands of landlords, and with a budget of $500 to spend, there are more efficient options available to us. To automate the administration of treatment, we were considering leveraging Amazon Mechanical Turk as it provides a platform to commision Human Intelligence Tasks (HITs), such as sending thousands of emails while avoiding the limitations of employing robots (captchas). Through some research, we found a Python library for the Mechanical Turk API called Boto. Through this API we would be able to send the thousands of listings identified by the scraper and clear instructions on what email to send to human workers without a single manual intervention on our part. However, it was not until meeting with David Reiley did the risks of this approach become clear.

A brief conversation with David Reiley resulted in a dramatic lowering of the intended number of samples for our project, from thousands to hundreds. His primary concern with our initial intent to send thousands of email inquiries was balancing the inconvenience to thousands of innocent real estate listers with the benefit derived from understanding the impact of excessive punctuation on an email's appeal and persuasiveness. Additionally Dr. Reiley cited concerns that Mechanical Turk was not going to be reliable enough to faithfully execute the intended tasks in a consistent enough fashion. Lastly, as we as a group had anticipated, there was the potential stumbling block of spam filters and similar bot defenses of both

the Craigslist site and the email provider we ultimately would choose for our execution. As a result of our discussions the our target number of subjects was reduced from a few thousand to 400 - 500 total subjects. In order to determine the operational efficiency of our experiment we decided to run an initial test scenario of 20 - 30 subjects. The goal is to evaluate how our intended process actually runs, and to see how frequently we would get a real response from the Craigslist posters.

**Action Items for Next Steps**

Going forward, we will continue pushing our work plan forward. We plan to conduct a full test run of our data collection phase to identify any weak points in our pipeline that require further refinement before our final data collection begins. We plan to do a test run of about 20 - 30 observations to test the functioning of our data collection process. Apartment listings have been gathered for this test run (and may need to be collected again given how quickly listings seem to go stale), but in addition to this, we will need to accomplish the following:

1. Draft inquiry emails for each treatment group.

Hi there (, or !)

I am writing to express interest in the apartment you posted for rent on Craigslist . It looks like exactly what I'm looking for(. or !)

If the apartment is still available, I would love to set up a time to come check it out - can you let me know if you have any availability in the next week to set up a time?

Looking forward to hearing from you( ! or !!!!!!!!!!!!!!!!!!!!!!!!)

-Jane / John

2. Set up an email account from which inquiry emails will be sent. - the nature of the usernames or domains of the email account(s) has been considered as another source treatment, but we are concerned about diluting the pool of subjects too much
3. Set up a log to track record responses received. The email client should help with some of this.
4. Send out inquiries for initial test run.

Once we've completed our test run, we will reflect on whether our data collection pipeline needs improvement and make the necessary modifications. Once these modifications have been implemented, we will proceed to our operational data collection phase in which we will collect and send inquiries for 400 - 500 apartment listings. Data will be collected by each of the group members, probably to be split up by city (about 125 in each city)

**Research Design**

Data will be blocked on city, price and room number. We plan to run a linear regression of responses to treatment i.e.

$$Y(i) = a + b0\text{treatment}0 + b1\text{treatment}1 + \text{factors (cities, price, rooms)} + \text{gender} + \text{gender} * \text{treatment} + e_i$$

Where treatment are dummy variables and our covariates are converted to factors. While logistic regression might be a more appropriate model if we were interested in predicting the probability of response for individual listings in the future, since we are interested in the difference in overall response rate to treatment group emails compared to control group emails, a linear regression and associated tests for b0, b1 will be sufficient for the purposes of our study.(See Annex for more details)

Our hypothesis is a test of the statistical significance of these coefficients. If exclamation points were to have an effect, we would expect that coefficients to be statistically significant.

**Questions:**

- We don't think this is a cluster design but do you agree?
- Is there a good way for us to calibrate the treatment (number of exclamation points) during the pilot. We want to have the largest effect but obviously too many could become counterproductive
- How many covariates do we want?
- Is it ok to sift through the postings and filter out duplicates and postings without email addresses? We would like to select out the best candidates prior to randomization and treatment to increase our likelihood of genuinely reaching unique posters.
- We want to bucket rent levels in 4 categories, but the cities all vary in cost of living. Is it ok to bucket out the emails after they have been collected, based upon what rent quartile each email falls in for each given city or is that somehow circular?
- Given the reduced number of subjects, would it be advisable to drop our inquiries to just 1br or 2br listings rather than the 1, 2, and 3br listings we initially intended?