W241 Final Project

Spring 2019

By Ava Rezvani, Jun Jun Peh, and Erico

1.0 Introduction

Craigslist[1] is an American website where users can post advertisements for jobs, items, housing, services, and more. Founded in 1995 by Craig Newmark, whose initial idea was to distribute email list amongst friends, the website is now home over 60 million users in over 70 countries.[2] It invites 80 million postings and has 50 billion page views per month. With its easy to post and free market space, where the users have control of what they post and when. Craigslist has become a market disruptor and now behemoth in the classified ads industry.

While users were previously restricted to character count in the Classified section of their local newspaper, they have now been given unlimited freedom to provide the exact amount and type of content that they wish to share, in the location and time that they want to. The power is in their hands, and therefore they have opportunities to market their items for their benefit.

1.1 Motivation

Given this freedom in posting content, it is in the users' best interest to create posts that will yield the most successful results. In this case, we are interested in one of the most common users: the merchant. Let's take the user who is trying to list an item for sale on Craigslist. Understanding the impact of all listing factors can help shed light into buyer interests in certain markets. Their goal is clear: maximize their sale (price, time to close deal, and in some cases, quantity). But they may have questions as to how to go about this. For example, how does a professional product description with full specs, proper grammar and clear photos increase buyer interest? How would one measure that to ensure that it is effective?

With this online classified platform comes the opportunity to more easily analyze and predict user behavior, and now is the time to leverage this in order to inform better results for this growing user base.

The outcome of this research might help users/merchants to understand the marketplace demand, how important is to write a full professional description, and how to use the right amount and quality of pictures. Moreover, it will help the Craigslist marketplace continue to grow as coveted platform for any user to list any service.

---

[1] www.craigslist.com

[2] https://en.wikipedia.org/wiki/Craigslist

1.2 Research Question

Our original research question is "*Do Listing Factors Affect Buyers Interest in Classified Advertisements Platform*?" As we move through the report, we will specify the listing factors of interest.

1.3 Hypothesis

For this research question, our null hypothesis is "No treatment effect for either photo quality or presence of a description for post an item for sale in an online marketplace." Our alternative hypothesis is "No treatment effect for either photo quality or presence of a description for post an item for sale in an online marketplace."

2.0 Experimental Design

This experiment was conducted via two different classified advertisement platforms, which are Craigslist and OfferUp[3]. The general idea is that several variation of listings and products were posted in these platforms and responses (buyers' interest) were aggregated based on these listings respectively.

2.1 Pilot Study

Before the actual advertisements are being posted for data collection, a mock Craigslist listing were posted to observe the potential outcome we would be getting. This listing, which is posted in Dallas TX, consists of a camera lens with some general descriptions and 2 typical photos of the product. Yet the posting includes incorrect details of the product were intentionally made to check if buyers actually care about the details of descriptions. Within the first 48 hours, the listing received 6 replies, in which 2 of the replies were asking for clarification on the camera lens. This pilot study allowed us to gauge the estimated responses or interest we will get if we ever increase the experiment to a larger scale and provided us an estimated idea on number of cities to perform the experiment on.

2.2 Methodology

The entire experiment comprises of 3 different data collections, due to circumstances faced (explained in latter section) when we posted the first product category (Electronics - TV) in Craigslist.

[3] https://offerup.com

2.2.1 Experiment 1

We selected Electronics - TV as our product category in Craigslist listing. This is because TV is a household electronics that can be generalized to the public regardless of the buyers age, gender, and locations. In this experiment, we created 4 different combinations of listings below:

1. *Control: bad quality photo and single line description.*
2. *Treatment 1: bad quality photo with full product description*
3. *Treatment 2: good quality photo with single line description*
4. *Treatment 3: good quality photo with full product description*

These listings were posted in 20 cities, where each city will be randomly assigned to receive one of the combinations in a fixed time period (a week from time of posting - 03/17/19 at 3pm respective time zone). At the end of the week, the responses (signified buyers' interest) were collected and counted based on respective listing groups as our outcome measure. Spam responses and responses that were purely asking for clarity were filtered and removed prior to data aggregation.

2.2.2 Experiment 2

There were several complications after Experiment 1, including many posts being marked as spam and being removed by Craigslist. Additionally, the Electronics - TV market was highlight saturated and therefore for posts that remained, we received few responses.

We decided to create a simpler study, which comprises only a control and a treatment listing using the same camera lens in pilot study. Instead of experimenting with the photo quality, that created 2 extra permutations in the treatments, we controlled the photos. This experiment studied the effect in an unsaturated market compared to TV, as camera lens listing is only targeting specific group of buyers. Variation on description below were used in this study:

1. *Control: Normal photos and no description*
2. *Treatment 1: Normal photos and full description*

These set of listings were posted in the same 20 cities, where 10 cities are randomly assigned to receive control listing while another 10 received treatment listing. Control group

2.2.3 Experiment 3

At the same time, we also created third experiment in OfferUp, which is a mobile-driven marketplace that allows user to buy and sell goods online. We decided to carry out this experiment due to the fact that OfferUp allows us to track buyers' view in addition to responses. Besides, this app allows us to post a listing agnostic of posting location(similar to how eBay[4]

[4] https://www.ebay.com

works), so that anyone in the US can view the listing. We chose to use the same camera lens as our product listing with the posting variations below:

1. *Control: Poor photos with single line description*
2. *Treatment 1: Normal photos with full professional description*

Since the same title was used, photo quality will determine the number of views on the listing while both photo and description quality will determine the number of responses received. The results obtained will then be compared to the previous 2 experiments above.

We considered that there will be a *spillover effect* in this experiment since the listing is posted nationwide, where people will be able to see multiple listing of same product (control and treatment). If this happens, buyer might be questioning the authenticity of the product and merchant, and they may choose to respond to just one of it and this might affect the outcome measure. Hence, we implemented *stepped-wedge desig*n by posting the 2nd listings of same product in different time-frame, while archiving the first one before that.

2.3 Treatment vs Control

A few variation of treatments were applied to each experiment described above to study the causal effect and treatment effect on the outcome measures as compared to control group's listings. From the 3 experiments above, variation between control versus treatment groups consists of:

1. Craigslist 1 (TV): Quality of photos and quality of descriptions
2. Craigslist 2 (Camera Lens): Presence of description or not
3. OfferUp (Camera Lens): Quality of photos and quality of descriptions

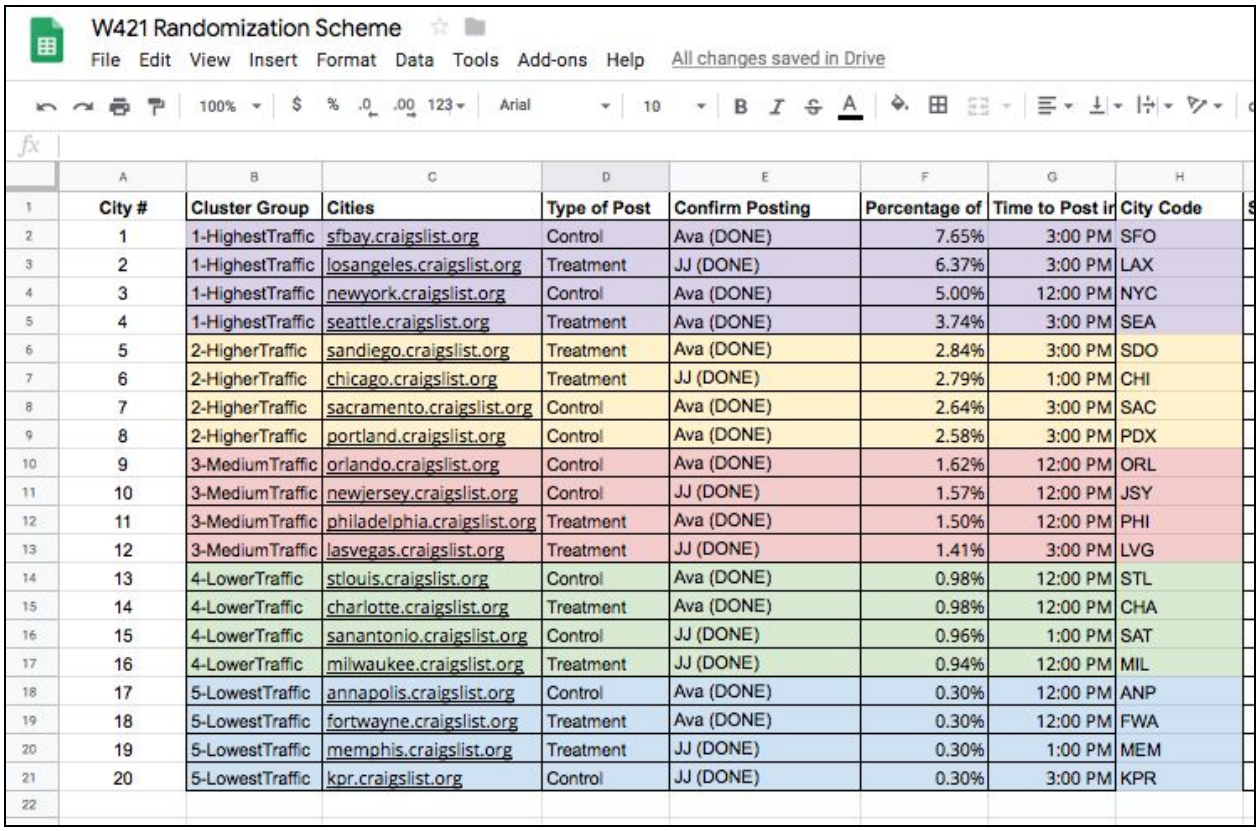
2.4 Assumptions and Controlled Variables

To avoid *selection bias*, several variables were being controlled in these experiments, such as time of the day of posting, product sale price, product category, title of listing, and product condition.

Since we did not proceed to respond to buyer’s messages due to ethical concern, we were not able to understand their real intention about certain messages. Hence, we made an assumption that people who responded about interest in buying and price negotiating are considered interested in making the purchase. However, responses purely asking for more product details were removed from data analysis.

2.5 Randomization

Randomization is required in this experiment to ensure that treatment and control listings are evenly distributed to cities chosen. According to study[5] on 2017 craigslist most trafficked cities, we used blocking method to group cities with similar Craigslist user percentage in groups of four. There was randomization within each group of four to receive control and treatment, evenly.



W421 Randomization Scheme

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

	City #	Cluster Group	Cities	Type of Post	Confirm Posting	Percentage of	Time to Post ir	City Code
2	1	1-HighestTraffic	sfbay.craigslist.org	Control	Ava (DONE)	7.65%	3:00 PM	SFO
3	2	1-HighestTraffic	losangeles.craigslist.org	Treatment	JJ (DONE)	6.37%	3:00 PM	LAX
4	3	1-HighestTraffic	newyork.craigslist.org	Control	Ava (DONE)	5.00%	12:00 PM	NYC
5	4	1-HighestTraffic	seattle.craigslist.org	Treatment	Ava (DONE)	3.74%	3:00 PM	SEA
6	5	2-HigherTraffic	sandiego.craigslist.org	Treatment	Ava (DONE)	2.84%	3:00 PM	SDO
7	6	2-HigherTraffic	chicago.craigslist.org	Treatment	JJ (DONE)	2.79%	1:00 PM	CHI
8	7	2-HigherTraffic	sacramento.craigslist.org	Control	Ava (DONE)	2.64%	3:00 PM	SAC
9	8	2-HigherTraffic	portland.craigslist.org	Control	Ava (DONE)	2.58%	3:00 PM	PDX
10	9	3-MediumTraffic	orlando.craigslist.org	Control	Ava (DONE)	1.62%	12:00 PM	ORL
11	10	3-MediumTraffic	newjersey.craigslist.org	Control	JJ (DONE)	1.57%	12:00 PM	JSY
12	11	3-MediumTraffic	philadelphia.craigslist.org	Treatment	Ava (DONE)	1.50%	12:00 PM	PHI
13	12	3-MediumTraffic	lasvegas.craigslist.org	Treatment	JJ (DONE)	1.41%	3:00 PM	LVG
14	13	4-LowerTraffic	stlouis.craigslist.org	Control	Ava (DONE)	0.98%	12:00 PM	STL
15	14	4-LowerTraffic	charlotte.craigslist.org	Treatment	Ava (DONE)	0.98%	12:00 PM	CHA
16	15	4-LowerTraffic	sanantonio.craigslist.org	Control	JJ (DONE)	0.96%	1:00 PM	SAT
17	16	4-LowerTraffic	milwaukee.craigslist.org	Treatment	JJ (DONE)	0.94%	12:00 PM	MIL
18	17	5-LowestTraffic	annapolis.craigslist.org	Control	Ava (DONE)	0.30%	12:00 PM	ANP
19	18	5-LowestTraffic	fortwayne.craigslist.org	Treatment	Ava (DONE)	0.30%	12:00 PM	FWA
20	19	5-LowestTraffic	memphis.craigslist.org	Treatment	JJ (DONE)	0.30%	1:00 PM	MEM
21	20	5-LowestTraffic	kpr.craigslist.org	Control	JJ (DONE)	0.30%	3:00 PM	KPR

Figure 2.5.1 Randomization Scheme for Experiment 2

2.6 Clustered Design

Our targeted population is clustered on a city-level, where each Craigslist listing is posted in selected cities as a whole, such that people in that city will be able to see the same post. The outcome will be measured on individual level, where we aggregate each of the buyers' responses.

[5] https://www.craigslistpostingservice.net/craigslist-most-trafficked-cities-2017-edition/

2.7 Experiment Result & Outcome Measures

The outcome that is measured in Experiment 2 is the number of responses received associated with each of the listings. The responses (after filtering out spams and product detail inquiries) signifies buyers interest on the product advertised. The data was aggregated to study the effect of treatment as compared to control in the field of classified advertisement platform. The overall evaluation criteria is if there's enough statistical power to reject the null hypothesis that description and photo quality has no effect on buyers' interest.

The outcome variables being measurement in Experiment 3 are both the number of responses received as well as the number of views. This is because number of views can tell us the effect of photo quality, as the number of views will purely come from the difference of photo quality of two different listings, while other variables such as product title and price being held constant. One drawback about this experiment is the number of responses will then be determined by both the photo and description quality. Therefore, this experiment will mainly allow us to see how these two treatment variables affect the responses together and compare it to the effect on Experiment 2, in which responses are purely based on description quality.

3.0 Data

3.1 Data Aggregation & Cleaning

As mentioned in 2.2 Methodology above, we created three different experiments to study different listing factors. The data collection was generated via the listings posted in both Craigslist and OfferUp accounts. When potential buyers contacted us via the listings, an email was sent to a dummy Gmail account we created for this experiment. At the end of the week, responses were tallied and grouped associated to the listings, which in turn classified based on their respective control or treatment groups.

In order to make sure that the responses gathered were a correct measure of buyer intent, we filtered out responses from people who are purely asking for clarification of product information or responses that appear fraudulence.

In experiment 2, the number of responses were tallied based on each of the cluster group and condition as shown in figure 3.1.1 below. Both number of responses and view counts were tallied in experiment 3 as shown in Figure 3.1.2. These are the important details we want to analyze on to deduce treatment effect; and the data from each of these experiments will give us different results to look at.

City #	Cluster Group	Cities	Condition	Percentage of Traffic	Time to Post in CA	City Code	Response_Count
1	1-HighestTraffic	sfbay.craigslist.org	Control	7.65%	3:00 PM	SFO	11
2	1-HighestTraffic	losangeles.craigslist.org	Treatment	6.37%	3:00 PM	LAX	6
3	1-HighestTraffic	newyork.craigslist.org	Control	5.00%	12:00 PM	NYC	3
4	1-HighestTraffic	seattle.craigslist.org	Treatment	3.74%	3:00 PM	SEA	4
5	2-HigherTraffic	sandiego.craigslist.org	Treatment	2.84%	3:00 PM	SDO	5
6	2-HigherTraffic	chicago.craigslist.org	Treatment	2.79%	1:00 PM	CHI	3
7	2-HigherTraffic	sacramento.craigslist.org	Control	2.64%	3:00 PM	SAC	0
8	2-HigherTraffic	portland.craigslist.org	Control	2.58%	3:00 PM	PDX	2
9	3-MediumTraffic	orlando.craigslist.org	Control	1.62%	12:00 PM	ORL	0
10	3-MediumTraffic	newjersey.craigslist.org	Control	1.57%	12:00 PM	JSY	0
11	3-MediumTraffic	philadelphia.craigslist.org	Treatment	1.50%	12:00 PM	PHI	0
12	3-MediumTraffic	lasvegas.craigslist.org	Treatment	1.41%	3:00 PM	LVG	6
13	4-LowerTraffic	stlouis.craigslist.org	Control	0.98%	12:00 PM	STL	0
14	4-LowerTraffic	charlotte.craigslist.org	Treatment	0.98%	12:00 PM	CHA	6
15	4-LowerTraffic	sanantonio.craigslist.org	Control	0.96%	1:00 PM	SAT	5
16	4-LowerTraffic	milwaukee.craigslist.org	Treatment	0.94%	12:00 PM	MIL	5
17	5-LowestTraffic	annapolis.craigslist.org	Control	0.30%	12:00 PM	ANP	0
18	5-LowestTraffic	fortwayne.craigslist.org	Treatment	0.30%	12:00 PM	FWA	2
19	5-LowestTraffic	memphis.craigslist.org	Treatment	0.30%	1:00 PM	MEM	0
20	5-LowestTraffic	kpr.craigslist.org	Control	0.30%	3:00 PM	KPR	1

Figure 3.1.1 Response count in Experiment 2

Product	PhotoQuality	Description	Condition	Response	Views
Lens	Good	Full	Treatment	24	982
Lens	Bad	Single	Control	7	338
Camera	Good	Full	Treatment	39	972
Camera	Bad	Single	Control	17	456

Figure 3.1.2 Response and View counts in Experiment 3

3.2 Data Analysis & Results

After data from experiments were collected and aggregated, we computed the data in R to study the treatment effect as well as several other statistical analyses, which are broken down into below parts:

1. Statistical power
2. Sample size consideration
3. Observation from control group
4. ATE analysis
5. Regression & Modeling
6. Calculate f1 score and robust standard error

We computed these analyses on each of the experiments we carried out, with specific focus on experiment 2 and experiment 3. Some visualizations and summaries were extracted from R scripts as shown below (Please refer to Rmarkdown file link here for full computational details of all 3 experiments).

3.3 Experiment 1

We first look at the statistical power in Experiment 1 based on the effect size (which represents the true population) and number of clustered block at n=10. We obtained a statistical power of 0.05321927 as shown in Figure 3.3.2, which signifies that our design is underpowered. If we want to have a design that is statistically significant, we would need a sample size of at least 60 or higher effect size comparing treatment and control groups.

```
pwr.f2.test(u=3, v=16, f2=pseudo_r2/(1-pseudo_r2), sig.level = .05)
```

```
##
##      Multiple regression power calculation
##
##               u = 3
##               v = 16
##              f2 = 0.1023904
##       sig.level = 0.05
##           power = 0.1622645
```

Figure 3.3.1: Power Test for Statistical Power Analysis

We then group the responses based on the control and treatment categories respectively in box plot to give a better representation of how each of the group responses. Overall, we obtained ATE as shown below (detailed calculations in R markdown files attached link in reference):

1. Treatment 1: 0.4
2. Treatment 2: 1.0
3. Treatment 3: 1.6

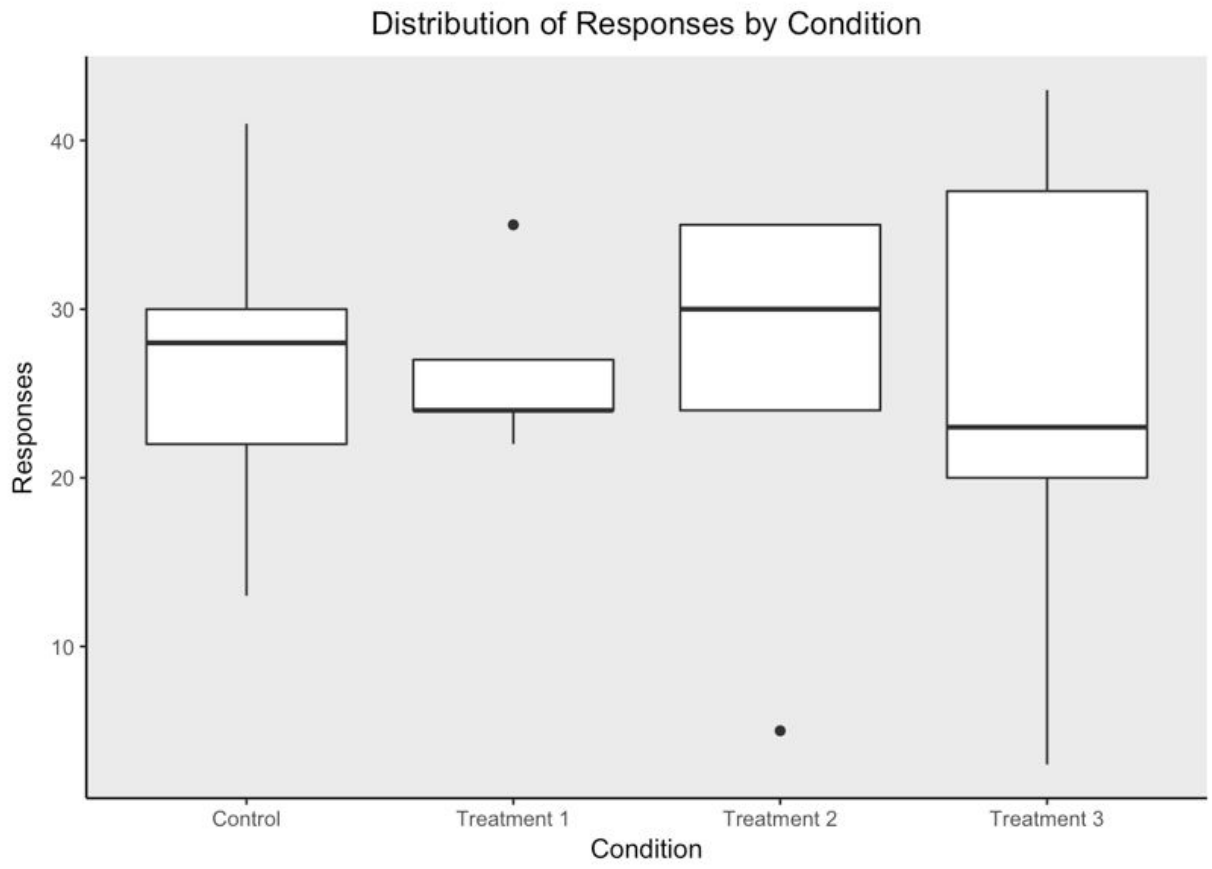


Figure 3.3.2: Box Plots Comparing 4 Experimental Groups in Experiment 1

Lastly, we created a regression model and calculated the R-squared score and robust standard error for this experiment. However, this experiment did not give us too much insight on the effect of photo and description qualities on number of responses. Hence, our analyses are geared towards experiment 2 and 3 detailed below.

3.4 Experiment 2

We used a two sample t-test to estimate the statistical power of our experiment. Below is a plot of a theoretical T-distribution in a two-tailed test. The rejection region is computed at +/- 2.10092. As seen in the figure, our obtained T-value falls outside of the rejection region, which means that we could not reject the null hypothesis (no evidence of significant effect).

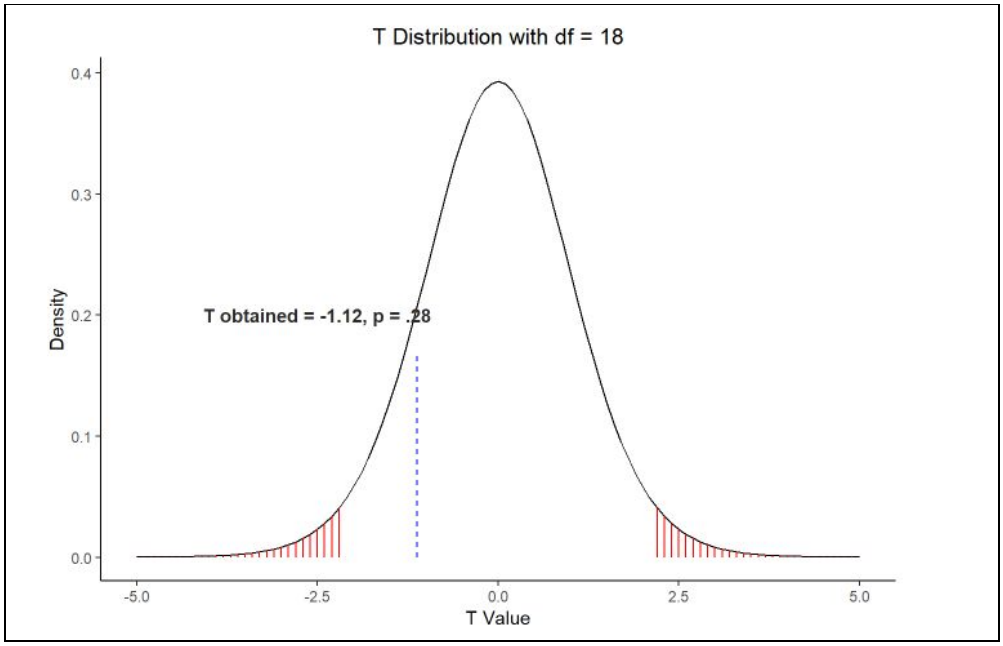


Figure 3.4.1: T Distribution of Experiment 2

In order to calculate power on a 2-group comparison, we calculated the Cohen's effect size and using the pwr.t.test function as shown in figure 3.4.2 below. The obtained power with 10 unique n in treatment and control groups is 0.1853525, which also shows that our experiment is underpowered. This would mainly be due to very little mean difference (ATE between control and treatment), termed by 'd' in our analysis. We would need to increase number of responses receive, number of groups, or the significance level to get enough statistical power.

```
# Calculate effect size
cohens_d <- function(x, y) {
  lx <- length(x)- 1
  ly <- length(y)- 1
  md  <- abs(mean(x) - mean(y))        ## mean difference (numerator)
  csd <- lx * var(x) + ly * var(y)
  csd <- csd/(lx + ly)
  csd <- sqrt(csd)                     ## common sd computation

  cd  <- md/csd                        ## cohen's d
}

exp2_cohens_d <- cohens_d(
  exp2_df[exp2_df$Condition == 'Control', 'Response_Count'],
  exp2_df[exp2_df$Condition == 'Treatment', 'Response_Count']
)
exp2_cohens_d
```

```
## [1] 0.5004636
```

```
pwr.t.test(n = 10, d = exp2_cohens_d, sig.level = 0.05,
           type = 'two.sample', alternative = 'two.sided')
```

```
##
##      Two-sample t test power calculation
##
##               n = 10
##               d = 0.5004636
##       sig.level = 0.05
##           power = 0.1853525
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

Figure 3.4.2: Statistical Power Estimation using pwr.t.test

Based on the effect size, d, of 0.5004636, we also plotted a power curve, assuming this effect size is the true population effect size. We understood that with our sample size of n=10 in each

group, our experiment is underpowered. The curve shows a recommendation of n=60 for each group in order to achieve the targeted power of 80%.

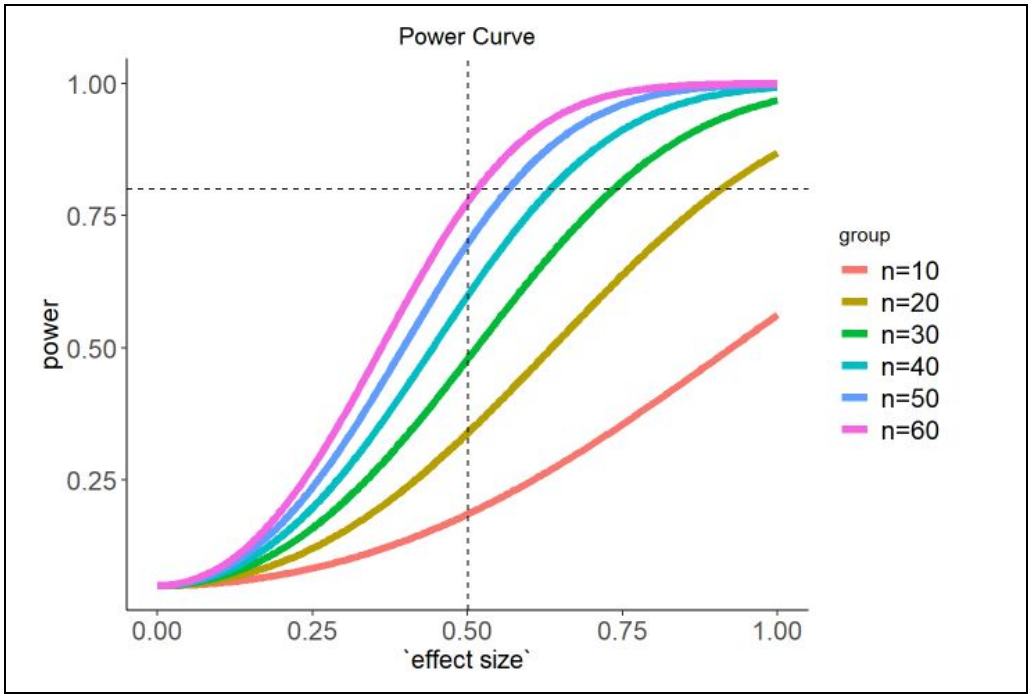


Figure 3.4.3: Power Curve Visualization for Experiment 2

Disregarding the fact that our experiment is lacking of statistical power (which could be improved if we had developed a better method for data collection), we saw an ATE of 1.5. This was computed from the control group effect of 2.2 and treatment group effect of 3.7, which translated to the number of responses received from the variation of Craigslist camera lens posts.

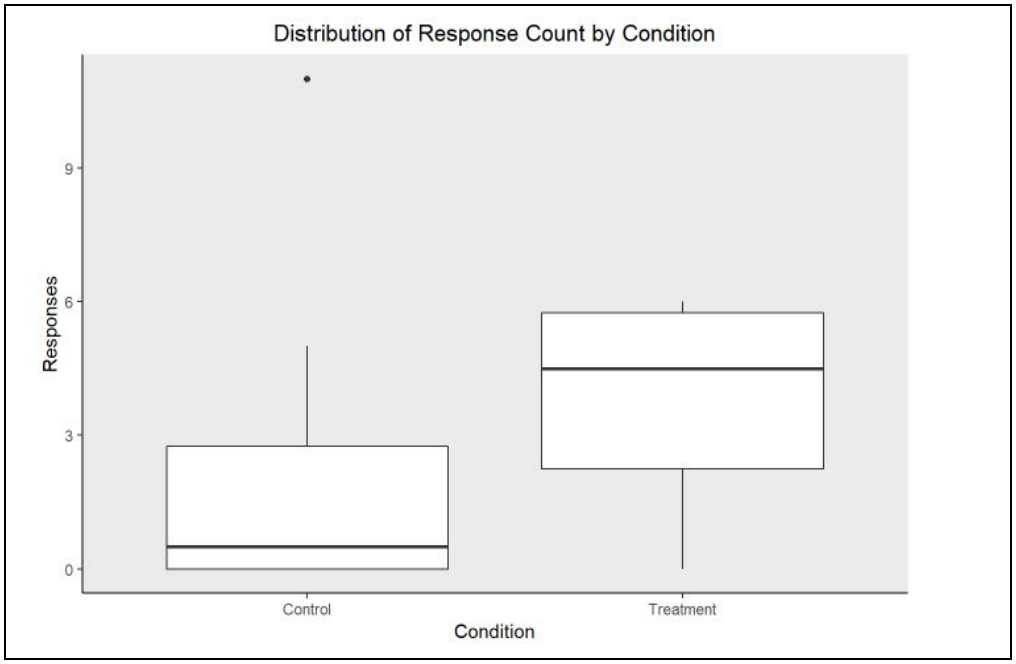


Figure 3.4.4: ATE of Experiment 2

Lastly, we created a regression model and calculated the F1 score and robust standard error for this experiment. The outcome is specified to be Poisson distributed, since it is the common error distribution for integer count. The mean of the response is mapped to the linear combination of features via a logarithmic link function. Robust standard error is shown via z-test of coefficients to be 0.3654 on treatment condition of the experiment.

```
##
## Call:
## glm(formula = Response_Count ~ Condition, family = poisson, data = exp2_df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.7203  -2.0976  -0.2567   0.7548   4.2199
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)          0.7885     0.2132   3.698 0.000217 ***
## ConditionTreatment   0.5199     0.2692   1.931 0.053481 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 67.706  on 19  degrees of freedom
## Residual deviance: 63.850  on 18  degrees of freedom
## AIC: 109.99
##
## Number of Fisher Scoring iterations: 6
```

Figure 3.4.5: Linear Regression Model of Experiment 2

```
##
## z test of coefficients:
##
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)          0.78846    0.53354  1.4778   0.1395
## ConditionTreatment   0.51988    0.57432  0.9052   0.3654
```

Figure 3.4.6: Z-test and Robust Standard Error of Experiment 2

3.5 Experiment 3

In experiment 3, we carried out similar two sample t-test to look at ATE and statistical power. The difference in experiment 3 is the capability to track page views associated to each listing, thus giving us a way to look at the combined effect of photo and description quality via monitoring page views and number of responses. The combined effect as shown in Figure 3.5.1 is deduced from the total responses each listing group gets in associated to the total views (in another way, number of people viewed the listing with respect to responses we obtained).

```
# combined_effect = (total views) / (total responses) for each control and treatment. this is the number of peo
ple view the listing wrt responses we get.
exp3_df$combined_effect <- exp3_df$Views / exp3_df$Response
exp3_df
```

```
##   Condition Response Views combined_effect
## 1 Treatment       24   982        40.91667
## 2   Control        7   338        48.28571
## 3 Treatment       39   972        24.92308
## 4   Control       17   456        26.82353
```

Figure 3.5.1: Table of Experiment 3 with Combined Effect

The statistical power of experiment 3 is calculated to be 0.2437954, which also appears to be underpowered. In order to get enough statistical power with significance level at 0.05, n has to be equal or greater than 8, which means that we will need to collect data of at least 8 products.

```r
# Calculate effect size
cohens_d <- function(x, y) {
  lx <- length(x)- 1
  ly <- length(y)- 1
  md  <- abs(mean(x) - mean(y))        ## mean difference (numerator)
  csd <- lx * var(x) + ly * var(y)
  csd <- csd/(lx + ly)
  csd <- sqrt(csd)                     ## common sd computation

  cd  <- md/csd                        ## cohen's d
}

# Description - Response
exp3_cohens_d <- cohens_d(
  exp3_df[exp3_df$Condition == 'Control', 'Response'],
  exp3_df[exp3_df$Condition == 'Treatment', 'Response']
)
exp3_cohens_d
```

```
## [1] 2.163331
```

```r
pwr.t.test(n=2, d = exp3_cohens_d, sig.level = 0.05,
           type = 'two.sample', alternative = 'two.sided')
```

```
##
##      Two-sample t test power calculation
##
##               n = 2
##               d = 2.163331
##       sig.level = 0.05
##           power = 0.2437954
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

Figure 3.5.2: Pwr.t.test for Statistical Power

We then compared the average treatment effect of experiment 3 from control and treatment group. Despite getting an expected positive treatment effect when assessing views (associated to photo quality) and responses (associated to description quality) separately, we observed a negative treatment effect when looking at combined effect (associated to both photo and description quality) on responses. This disobeyed the reasoning we had on ATE. One reason we can deduce is that the treatment effect for photo alone is large enough compared to description quality for the population who are already interested to purchase the product. Hence, as long as the people who view the listings belong to this potential buyer's group, they will want show the interest by response regardless the description quality.

```r
# Photo - Views
exp3_df %>%
  group_by(Condition) %>%
  summarize(avg = mean(Views),
            sd = sd(Views))
```

```
## # A tibble: 2 x 3
##   Condition   avg        sd
##       <fctr> <dbl>     <dbl>
## 1   Control   397 83.438600
## 2 Treatment   977  7.071068
```

```r
# Description - Response
exp3_df %>%
  group_by(Condition) %>%
  summarize(avg = mean(Response),
            sd = sd(Response))
```

```
## # A tibble: 2 x 3
##   Condition   avg        sd
##       <fctr> <dbl>     <dbl>
## 1   Control  12.0  7.071068
## 2 Treatment  31.5 10.606602
```

```r
# Description & View - combined_effect
exp3_df %>%
  group_by(Condition) %>%
  summarize(avg = mean(combined_effect),
            sd = sd(combined_effect))
```

```
## # A tibble: 2 x 3
##   Condition      avg       sd
##       <fctr>    <dbl>    <dbl>
## 1   Control 37.55462 15.17606
## 2 Treatment 32.91987 11.30918
```

Figure 3.5.3: Computations comparing 3 different ATEs based on photo, description, and combined effect

Lastly, we created a regression model and calculated the F1 score and robust standard error for this experiment. The outcome is specified to be Poisson distributed, since it is the common error distribution for integer count. The mean of the response is mapped to the linear combination of features via a logarithmic link function. Robust standard error is shown via t-test of coefficients to be 0.2657 on treatment condition of the experiment.

```
# 6) Regression

# CHECK: response is dependent on condition and views? or should be separated
summary(lm(Response ~ Condition + Views, data=exp3_df))
```

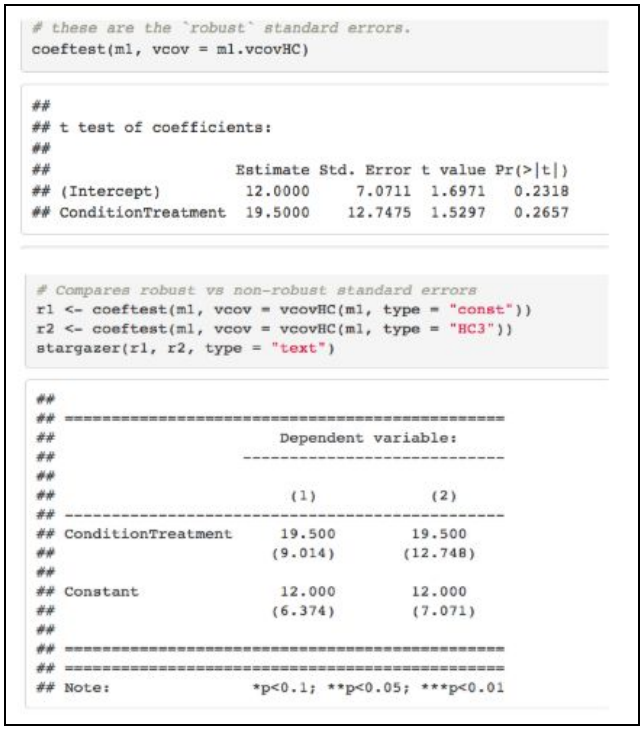
```
##
## Call:
## lm(formula = Response ~ Condition + Views, data = exp3_df)
##
## Residuals:
##       1       2       3       4
## -7.8672 -0.6667  7.8672  0.6667
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -17.15787   53.52266  -0.321    0.803
## ConditionTreatment -23.09840   78.14073  -0.296    0.817
## Views                0.07345    0.13334   0.551    0.679
##
## Residual standard error: 11.17 on 1 degrees of freedom
## Multiple R-squared:  0.7703, Adjusted R-squared:  0.3109
## F-statistic: 1.677 on 2 and 1 DF,  p-value: 0.4793
```

```
# You can model a t.test as a simple linear regression with
# a dummy-coded variable for the condition factor.

# This is equivalent to a two-sample T-test
t.test(exp3_df$Response ~ exp3_df$Condition, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  exp3_df$Response by exp3_df$Condition
## t = -2.1633, df = 2, p-value = 0.163
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -58.28359  19.28359
## sample estimates:
##   mean in group Control mean in group Treatment
##                    12.0                    31.5
```

Figure 3.5.4: Linear Regression Model of Experiment 3



```
# these are the `robust` standard errors.
coeftest(m1, vcov = m1.vcovHC)

##
## t test of coefficients:
##
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         12.0000     7.0711  1.6971   0.2318
## ConditionTreatment  19.5000    12.7475  1.5297   0.2657

# Compares robust vs non-robust standard errors
r1 <- coeftest(m1, vcov = vcovHC(m1, type = "const"))
r2 <- coeftest(m1, vcov = vcovHC(m1, type = "HC3"))
stargazer(r1, r2, type = "text")

##
## ================================================
##                         Dependent variable:
##                     ---------------------------
##
##                             (1)             (2)
## ------------------------------------------------
## ConditionTreatment       19.500          19.500
##                         (9.014)         (12.748)
##
## Constant                 12.000          12.000
##                         (6.374)         (7.071)
##
## ================================================
## ================================================
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

Figure 3.5.5: Robust Standard Error Comparison of Experiment 3

4.0 Questions & Future Discussions / Improvement

This project was a great learning experience in understanding the intricacies of data collection that can then compromise outcomes and conclusions. While we are limited in our current scope of the time of the course as well as the limitations in our data collection platforms, we do believe that our efforts can be applied for future usage.

4.1 Application

The following questions and explanations help scope the future applications of this project.

1. How would these methods and results translate across other product markets?
    a. This could include furniture, clothing etc. We would be interested in understanding this through saturated and unsaturated markets.
2. How would this translate in an auction like situation?
    a. This is another scenario where the power is in the hand of the user and they can customize their posting as they would without the prior limitations of advertisement. An ideal place for this would be Craigslist.
3. Why does Craigslist not include a page view counter for merchants?
    a. There were some workarounds for this but would have resulted in immediate spam markings yet we were interested as to why this was not an option.

4.2 Lesson Learned & Omitted Variable

Some unforeseen concerns that we were had were the major limitations using Craigslist as a method for our study. We ended up having to run multiple data collection efforts and reposting posts that had been marked as spam and removed. Additionally, our response windows to get emails were limited to 7 days because Craigslist posting windows in the more popular cities like San Francisco and New York City were limited to that. Finally, we had a low power due to the number of small responses and the limited time in our data collection.

4.3 Conclusion

In conclusion, due to the low statistical power of our experiments, we are not able to reject the null hypothesis that photo and description qualities have no effect on buyers' interest in an online marketplace. If we are given more time to properly plan out the experiment and data collection, we would want to run a more thorough pilot study experimenting different online marketplace platforms such as eBay and OfferUp. At the same time, we will need to experiment with more product categories in both saturated and unsaturated markets such that the results similar to what we obtained in experiment 3 will be valid. We also learned the importance to calculate the statistical power for our experiments prior to carrying out the data collection to avoid wasted effort. We do hope that in the future one of us are able to re-carryout this experiment in full functional scale since this is the kind of study that interest us to work on in the first place.

5.0 Notes

All data analyses were done in R markdown format and presentation slides were uploaded in github repository:https://github.com/jjpeh/W241.