Introduction to Data Analytics 2

Pradeep Paladugula

2020-02-02

Part 1: Experiment Design

Title: Consumer Pseudo-Showrooming and Omni-Channel Product Placement Strategies

Abstract: Recent advances in information technologies (IT) have powered the merger of online and offline retail channels into one single platform. Modern consumers frequently switch between online and offline channels when they navigate through various stages of the decision journey, motivating multichannel sellers to develop omni-channel strategies that optimize their overall profit. This study examines consumers' cross-channel search behavior of "pseudo-showrooming," or the consumer behavior of inspecting one product at a seller's physical store before buying a related but different product at the same seller's online store, and investigates how such consumer behavior allows a multichannel seller to achieve better coordination between its online and offline arms through optimal product placement strategies.

Participants in the study were grouped into the following categories:

- Where_bought: Where they ended up purchasing an item: bought at the store, bought online.
- Who_bought: If they bought from the same or a different retailer.

Each participant was then measured on:

- Money: how much money they spent in dollars on the product.
- Time: how much time (in minutes) they spent looking at the product online.

1) What would be one possible null hypothesis based on this study?

   study resource: https://study.com/academy/lesson/what-is-a-null-hypothesis-definition-examples.html

   Solution: Based on the above the theory, their is no significancant relationship that consumer spent time on the searching a product online and buying the same product in same store or a different sotre.

2) What would be one possible alternative hypothesis based on this study?

   Solution: It is a inverse of the null hypothesis, the significancant relationship that consumer spent time on the searching a product online and ended up buying the same product in same store or a different sotre.

3) Who are they sampling in this study?

   resource: https://aisel.aisnet.org/misq/vol41/iss2/13/ Solutions: The above theory is smsapling the “Consumers”.

4) Who is the intended population in this study?

   Solution: Consumers who are making purchases in in-store and online purchases.

5) Give an example of type 1 error based on this study (do not just define, explain in context how it might have happened here).

   Solution: Type 1 error is a rejection of a true null hypothesis. the bad quantitative analysis between mutiple factors, the amount of time spent on searching for a product, the money spent on a product, whehter that product is bought from the same or a different store.

6) Give an example of type 2 error based on this study (do not just define, explain in context how it might have happened here).

   Solution: type 2 error is a nono-rejection of false null hypothesis. The good quatative analysis between the amount of time spent on the searching for a product, and having relationship in the prurchaing the product through the same store or a different and vice versa.

Part 2: Use the 04_data.csv to complete this portion.

```
data = read.csv('04_data.csv')
```

1) For each IV list the levels (next to a, b):

   a. Where bought: store versus online

   b. Who bought: same store versus different store

2) What are the conditions in this experiment?

   Solution:

   1. Consumer bought the product from the same store
   2. Consumer bought the product from the different store
   3. Consumer bought the product from the same online store
   4. Consumer bought the product from the different online store

3) For each condition list the means, standard deviations, and standard error for the conditions for time and money spent. Please note that means you should have several sets of M, SD, and SE. Be sure you name the sets of means, sd, and se different things so you can use them later.

```
n1 = tapply(data$money, list(data$where_bought, data$who_bought), length)
n1
```

```
##        different same
## online        50   50
## store         50   50
```

```
mm = tapply(data$money, list(data$where_bought, data$who_bought), mean)
mm
```

```
##        different     same
## online  25.38076 25.13133
## store   34.70278 34.88091
```

```
msd = tapply(data$money, list(data$where_bought, data$who_bought), sd,
simplify = TRUE)
msd
```

```
##        different     same
## online  2.805442 3.256608
## store   3.051608 2.738662
```

```
se1 = msd / sqrt(n1)
se1
```

```
##        different      same
## online 0.3967494 0.4605540
## store  0.4315625 0.3873052
```

```
n2 = tapply(data$time, list(data$where_bought, data$who_bought), length,
simplify = TRUE)
n2
```

```
##        different same
## online        50   50
## store         50   50
```

```
tm = tapply(data$time, list(data$where_bought, data$who_bought), mean,
simplify = TRUE)
tm
```

```
##        different     same
## online  10.04745 10.51706
## store   15.33488 20.10002
```

```
tsd = tapply(data$time, list(data$where_bought, data$who_bought), sd,
simplify = TRUE)
tsd
```

```
##        different     same
## online  2.747142 3.029583
## store   2.657094 2.910464
```

```
se2 = tsd / sqrt(n2)
se2
```

```
##        different      same
## online 0.3885046 0.4284477
## store  0.3757699 0.4116018
```

4) Which condition appears to have the best model fit using the mean as the model (i.e. smallest error) for time?

Solution: "Bought in store from a different retailer" has the smallest error for time

5) What are the df for each condition?

```
#degrees of freedom: total number of observations in the market data to be tested minus the number of observations used by indicators, signals, and rules.
dataFrameM = data.frame(n1, mm, msd, se1, check.names = TRUE, fix.empty.names = TRUE)
dfm = N1 -1 #if df is degrees of freedom for money

## Error in eval(expr, envir, enclos): object 'N1' not found

dfm

## Error in eval(expr, envir, enclos): object 'dfm' not found

dataFrameT = data.frame(n2, tm, tsd, se2, check.names = TRUE, fix.empty.names = TRUE)
dft = N2 - 1 #if df is degrees of freedom for time

## Error in eval(expr, envir, enclos): object 'N2' not found

dft

## Error in eval(expr, envir, enclos): object 'dft' not found
```

6) What is the confidence interval (95%) for the means?

```
#Resourse: http://onlinestatbook.com/2/estimation/mean.html
mm - 1.96  * se1

##        different     same
## online  24.60313 24.22864
## store   33.85692 34.12179

mm + 1.96  * se1

##        different     same
## online  26.15839 26.03401
## store   35.54865 35.64003

tm - 1.96  * se1

##        different      same
## online  9.269819  9.614369
## store  14.489019 19.340900
```

```
tm + 1.96  * se1
```

```
##        different     same
## online  10.82508 11.41974
## store   16.18074 20.85914
```

7) Use the MOTE library to calculate the effect size for the difference between money spent for the following comparisons (that means you'll have to do this twice):

```
#Resourse:
https://www.rdocumentation.org/packages/MOTE/versions/1.0.2/topics/d.ind.t.t
library('MOTE')
```

```
## Registered S3 methods overwritten by 'lme4':
##   method                          from
##   cooks.distance.influence.merMod car
##   influence.merMod                car
##   dfbeta.influence.merMod         car
##   dfbetas.influence.merMod        car
```

```
sameRetailerEffect = d.ind.t(mm[2,2], mm[1,2], msd[2,2], msd[1,2], n1[2,2],
n1[1,2])
sameRetailerEffect$d
```

```
## [1] 3.240354
```

```
library('MOTE')
differentRetailerEffect = d.ind.t(mm[2, 1], mm[1, 1], msd[2, 1], msd[1, 1],
n1[2, 1], n2[1, 1])
differentRetailerEffect$d
```

```
## [1] 3.180373
```

8) What can you determine about the effect size in the experiment - is it small, medium or large?

   #resourse: https://www.simplypsychology.org/effect-size.html #Cohen's effective size resourse: https://opentextbc.ca/introbusinessstatopenstax/chapter/cohens-standards-for-small-medium-and-large-effect-sizes/

   Solution: As Cohen's standars for the both masurements are more than the 1.4 Effect size which is infact a very larger relative size for a significance level (a = 0.05)

   sameRetailerEffect: 3.240354 differentRetialerEffect: 3.180373

9) How many people did we need in the study for each comparison?

   Resource: https://cran.r-project.org/web/packages/pwr/vignettes/pwr-vignette.html

```
library('pwr')
##Store versus online when bought at the same retailer
```

```r
# As we have take n = NULl, as we got find that value.
pwr.t.test(n = NULL, d = sameRetailerEffect$d, sig.level = .05, power = 0.80)
```

```
##
##      Two-sample t test power calculation
##
##               n = 2.852122
##               d = 3.240354
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```r
##Store versus online when bought at a different retailer
pwr.t.test(n = NULL, d = differentRetailerEffect$d, sig.level = .05, power =
0.80)
```

```
##
##      Two-sample t test power calculation
##
##               n = 2.901572
##               d = 3.180373
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```