
Homework # 3

A/B TESTING
FALL 2024

NAMES: NAMES OF TWO STUDENTS HERE
ANDREW IDs: ANDREW IDs OF TWO STUDENTS HERE

ONE STUDENT SUBMITS HOMEWORK ON GRADESCOPE
THEN ADD OTHER STUDENTS AS TEAM MEMBERS

DUE DECEMBER 10, 5.00PM EST

SHOW YOUR WORK AS PART OF YOUR ANSWERS.
INCLUDE CODE AND SCREEN CAPTURES OF THE GENERATED OUTPUT.
CITE ALL YOUR SOURCES HUMAN GENERATED AND OTHERWISE.
IF YOU USE A LLM TO HELP WITH YOUR HOMEWORK
YOU ARE REQUIRED TO LIST ALL PROMPTS THAT YOU USED TO DO SO.

Question 1 [70 points]

Define the following concepts using your own words [10 points each]:

- a) Treatment assignment and treatment
- b) Complier, defier, always taker, and never taker
- c) Intention to Treat (ITT)
- d) Local Average Treatment Effect (LATE)
- e) Stable Unit Treatment Value Assumption (SUTVA)
- f) Cluster-level Randomized Controlled Trial
- g) Heterogeneous causal effect

[Type your answer here!](#)

Question 2 [94 points]

As discussed in class, TELCO ran an RCT where 24k households were offered access to the Cinema Pack: 10 TV channels broadcasting movies and TV shows 24/7. Households were randomly assigned to two conditions: Cinema Pack with TSTV, where households could use TSTV to watch the cinema pack; Cinema Pack without TSTV, where households could only watch the Cinema Pack live. In class we found that giving access to the Cinema Pack with TSTV increased TV consumption and did not decrease live consumption.

However, it is likely that not all households offered access to the Cinema Pack with TSTV used TSTV to watch the Cinema Pack. A dataset from this experiment is shared with you in file HW-3-a.csv for the purpose of this homework. The covariates included in this dataset are:

week: the week in the experiment

after: an indicator for the weeks after the experiment started

treated: whether the household was offered access to the Cinema Pack with or without TSTV

used90: whether the household used TSTV to watch the Cinema Pack for at least 90 minutes

used60: whether the household used TSTV to watch the Cinema Pack for at least 60 minutes

used30: whether the household used TSTV to watch the Cinema Pack for at least 30 minutes

total: total TV viewership (live + TSTV)

live: TV viewership in live mode

tstv: TV viewership in TSTV mode

id: an anonymized ID for the household

Use this dataset to answer the following questions:

- a) How many households are included in this dataset and how many weeks? **[2+2 points]**
- b) Show the ITT effects of allowing TSTV to consume the Cinema Pack on the consumption of TV (total, live, and TSTV) **[20 points]**
- c) Tabulate the number of households per treatment condition and TSTV usage. Is this (close to) a case of one-side non-compliance? **[10+10 points]**
- d) Show the Local Average Treatment Effects (LATE) of using TSTV to consume the Cinema Pack on the consumption of TV (total, live, and TSTV) **[30 points]**
- e) Compare the estimates obtained in d) and b) and explain how they relate to each other **[20 points]**

[Type your answer here!](#)

Question 3 [86 points]

A team of research analysts at TELCO would like to test whether consumers are less price elastic to movies placed towards the left of the TV screen. The idea behind this hypothesis is that consumers perceive movies suggested towards the left of the screen as better, because their eyes scan the TV screen from left to right (and top to bottom). Consequently, TELCO run an RCT to measure how showing movies towards the left of the TV screen affects the consumers' price elasticity of demand. During this experiment, the price of the movies was randomized into a low and a high price. For each household, the slot on the TV screen where each movie was shown was also randomized (slot 1 was the one farthest to the left of the TV screen). A dataset was collected as part of this experiment, shared with you in file HW-3-b.csv for the purpose of this homework. The covariates included in this dataset are:

n_lease: the number of times the movie was leased

order: the average slot in which the movie was shown to households

price: the price charged to watch the movie

mac_type: the type of set-top box owned by the household

movie_group: the catalog group to which the movie belongs to

package_name: the identifier for the movie

start_date: the date when the movie was added to the catalog

The fundamental covariates for this study are n_lease, order and price. The other covariates are fixed effects that one may control for in the regression analyses of the data in this dataset. Use this dataset to answer the following questions:

- a) How many movies are included in this dataset? Show the distributions of order and n_leases. What do you conclude from these distributions? **[2+10+10 points]**
- b) Show how the price elasticity of demand changes with slot on the TV screen, i.e., show that the effect of price on leases is heterogeneous with respect to the slot on the TV screen where the movie is shown **[50 points]**

c) Does the empirical evidence support the hypothesis of the research analysts at TELCO? Why or why not? **[14 points]**

Note 1: to answer question b) you may want to use a Generalized Linear Model (GLM) instead of just a Linear Model (LM) because the dependent variable `n_lease` contains a significant number of zeros. In R, this can be accomplished by using `FEGLM` (using `family=Poisson()`) instead of `FELM`. It is okay to use only a Linear Model but if you do so your results will be quantitatively different (although likely qualitatively similar) from those obtained using the generalized approach.

Note 2: to have more statistical power when answering question b) you may want to consider bundling up all TV slots towards the left of the screen up to a certain TV slot. For example, you can test if the movies placed in the first 2 slots towards the left of the TV screen have a different price elasticity of demand compared to the movies placed in slots 3 and up, and then you can repeat this exercise for the first 3 slots towards the left of the TV screen, the first 4, 5, etc...

Type your answer here!