
Homework # 2

A/B TESTING
FALL 2023

NAMES: NAMES OF ALL STUDENTS HERE
ANDREW IDs: ANDREW IDs OF ALL STUDENTS HERE

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

DUE NOVEMBER 21, 5.00PM EST

SHOW YOUR WORK AS PART OF YOUR ANSWERS.
INCLUDE CODE AND OUTPUTS AS NEEDED (SCREEN CAPTURES ARE ENOUGH).

CITE ALL YOUR SOURCES (HUMAN GENERATED AND OTHERWISE).

Question 1 [90 points]

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

- a) Panel data
- b) Serial correlation
- c) Time dummy
- d) Fixed effect
- e) Instrumental variable
- f) First-stage regression
- g) Exclusion restriction
- h) Differences-in-differences
- i) Propensity score matching

[Type your answer here!](#)

Question 2 [94 points]

A think tank has been hired to study the effect of streaming on piracy. For that purpose, they compiled the dataset available to you for the purpose of this homework in file HW-2-a.csv. The following covariates are included:

month: the month to which the data pertains

id: the identifier of the movie analyzed

available: an indicator for whether the movie was available in streaming platforms

log_IPs: the log of the number of IP addresses seen illegally exchanging the movie over the Internet

Use this dataset to answer the following questions:

- a) how many movies and which months are included in this dataset [2+2 points]?
- b) plot the average number of IP addresses per movie exchanging the movie illegally over time. From this plot, what do you conclude about how the availability of movies on streaming platforms affect piracy? Show regression results using plain OLS to support your claim [10+5+15 points]
- c) improve your empirical model using time dummies and show the results obtained. Is there a time trend in the number of IP addresses that are seen illegally exchanging a movie? Does this time trend make intuitive sense? [15+5+5 points]
- d) improve your empirical model by adding movie fixed effects and show the results obtained. What do you conclude for how the availability of movies on streaming platforms affect piracy? If you conclusions in b) and d) are different, which one you believe more and Why? [15+10+10 points]

Type your answer here!

Question 3 [66 points]

A few years ago a new technology was made available for racing cars in the pre-competitive bracket. A consulting firm is now studying the impact of that technology on race outcomes and gathered a dataset on a selected number of cars shared with you for the purposes of this homework in file HW-2-b.csv. This data set includes the following covariates:

id: the id of the car

tech: indicator for whether the car adopted the new technology

after: an indicator for whether the technology was already available or not

time: the identifier of the race

safety: a normalized rating for the safety of the car

speed: a normalized rating for the speed of the car

weight: the weight of the usual driver of the car

ability: an estimate of the ability of the usual driver of the car

points: the points obtained by the usual driver of the car

Use this dataset to answer the following questions:

- a) How many races and how many cars are included in this dataset? [**2+2 points**]
- b) Are there are significant relationships between points, safety, speed, ability and weight? [**10 points**]
- c) What is the effect of the new technology on points, safety and ability? [**30 points**]
- d) Are any of your estimates in c) capturing a causal effect? Show evidence to support your claim [**10+12 points**]

Type your answer here!