

Assignment 4: Online News Sharing @ Mashable

The dataset D5.2 (described in C5.2) contains information of online news articles published by Mashable (www.mashable.com). Some of these articles contains videos. Our primary question in this assignment is whether including at least one video in an article leads to the article being shared more in social media. Accordingly, the key outcome is “shares”—the number of social media shares for each article. The treatment indicator will be a variable that equals 1 if the number of videos included in an article (num_videos) is non-zero, and equals 0 otherwise.

1. Task 1:

- a. [2 points] Based on linear regression results, is the treatment associated with a typically larger or lower number of shares? For simplicity, base your answer on a regression of the outcome on the treatment indicator (i.e., do not include other covariates).

2. Task 2:

- a. [2 points] Evaluate the propensity score overlap between treated and non-treated subsamples.
- b. [3 points] Create a matched sample based on logistic propensity scores and in a way that accounts for overlap considerations
- c. [2 points] Assess the matched sample in terms of covariate balancing. In your judgement, has the matching procedure been successful?

3. Task 3:

- a. [2 points] Based on your analysis above, provide a matching ATE estimate. Do videos increase the number of shares? By how much? For simplicity, base your answer on a regression of the outcome on the treatment indicator (i.e., do not include other covariates).
- b. [2 points] Provide a rationale that explains the disparity between the estimate of 3.a and 1.a (i.e., your rationale must describe some form of behavior for why one estimate is larger than the other).
- c. [2 points] Suppose that the unconfoundedness assumption (discussed in class) holds: what could then be the “fudge factor” in this case? Explain.

Notes:

- If there are variables (e.g., x1, x2) that you want to drop from the data frame (ds), you can use the following codes:

```
drop = c("x1", "x2")  
ds = ds[,!(names(ds) %in% drop)]
```

Submission guidelines

- Submit via Canvas, 9PM EST on the day of class 6
 - Late submissions will be penalized
 - Late corrections will not be accepted
- Note that assignments are automatically checked for similarity—it is ok to discuss with other students, it is not ok to copy
- Submit **two** files (one submission per individual):
 1. Slide Deck (converted to pdf)
 - In the slide deck, I expect you to present results in an executive way.
 - Use as many slides as you need.
 - Please include screenshots of the R command lines (with command line #) in your slide deck to demonstrate your key steps.
 - The title page must include your name.
 - If you have worked/discussed with someone else, please also include their name(s) in a separate line on the title page.
 2. R script file containing the codes that you used for your analysis.
 - Include comments in the script to help the TA follow your procedures.
 - The script file should be understood as a companion. This way TAs can easily go back and double check that your answers in the ppt are well supported.