**Assignment 3: Online News Sharing @ Mashable**

The dataset D5.2 (described in C5.2) contains information of online news articles published by Mashable ([www.mashable.com](http://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. For tasks 1-3, the treatment indicator will a variable that equals 1 if the number of videos included in an article (num\_videos) is non-zero, and which equals 0 otherwise.

1. Task 1:
   1. Based on linear regression results, is the treatment associated to a typically larger or lower number of shares?
2. Task 2:
   1. Evaluate the propensity score overlap between treated and non-treated subsamples.
   2. Create a matched sample based on logistic propensity scores and in a way that accounts for overlap considerations
   3. Assess the matched sample in terms of covariate balancing. In your judgement, has the matching procedure been successful?
3. Task 3:
   1. 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 (ie, do not include other covariates).
   2. Provide a rationale that explains the sign of the difference 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).
   3. Suppose that the unconfoundedness assumption holds: what could then be the “fudge factor” (discussed in class) in this case? Explain.
4. Task 4:
   1. Propose a propensity score matching analysis that addresses the question of whether the specific number of videos included in an article (not whether a video is included) has a causal effect on the number of shares. For this analysis, you will need to make decisions about: (i) how to define the treatment indicator, (ii) what subsample of the full dataset to use. Note: there is no one single correct analysis.

Notes:

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

dropvars = c("x1","x2")

ds = ds[,!(names(ds) %in% dropvars)]