

Causal and Predictive Analytics – Homework 2

Individual Assignment

This assignment is based on a Facebook advertising dataset. It contains data from a large-scale field experiment of advertisements for various clothing brands on the facebook platform. The dataset was initially provided to a previous MS student, and was the subject of several questions in the hiring process. In this assignment, we will use this data to explore both omitted variable bias and interaction effects. The advertisements were all run in the same week. The business goal is to evaluate which attributes constitute successful advertisements.

You will be graded both on your code, and the written answers you provide. When evaluating the code, the grader will take on the role of a co-worker. Code will be evaluated both in terms of how correct and how clear it is. By correctness, I mean that the code fulfills the requirements of the question. By clarity, I mean that the grader should be able to understand what your code does within 30 seconds of reading it. As discussed in class, this is aided by clear comments, good variable names, proper indentation, and short lines.

The written portions will be evaluated the use of data and analysis to support your statements, and the quality of the writing.

Assignment Materials for Download:

1. An Rmarkdown template titled 'Assignment2Template-RenameToBlackboardUsername.Rmd'
2. A data file in .csv format titled 'Facebook Data.csv'

Submission Checklist:

To help us grade the assignments efficiently and correctly, we ask that you submit your assignments in a specific format. A complete submission for the following to blackboard:

- A .rmd Rmarkdown file, based on the template for this assignment.
- A .html file, generated by knitting the .rmd file in RStudio.
- Please make sure all files have the correct extensions
- Please zip all files into a single archive

Data Dictionary:

- `date` to the week the advertising campaign was run,
- `adType` refers to whether the advertisement appeared in a link or a photo post
- `category` refers to the type of retailer referenced in the advertisement
- `placement` represents whether the ad appeared on the desktop or mobile client
- `keyword` is a categorical variable that represents which store the purchase was made at
- `body` represents the actual text of the ad
- `ageMean` represents the average age of the targeted consumers
- `clickPerDollar` represents the number of clicks the ad achieved for each dollar spent

Part 1: Regression and Categorical Variables (12 marks)

In this section you will implement and interpret the results of a linear regression. You can complete this section using exclusively the `summary` and `lm` functions.

- Run a regression with the following formula:

```
clickPerDollar~factory(adType)+factor(category)+factor(placement)+  
factor(keyword)+factor(body)+ageMean
```
- Adding `date` to the set of independent variables in the regression will result in an error. Why did this not work? Answer in 30 words or less. (4 marks).
- In quantitative terms, what is the interpretation of the `factor(category)General` coefficient? Answer in 30 words or less. (4 marks).
- In quantitative terms, what is the interpretation of the `ageMean` coefficient? Answer in 30 words or less. (4 marks).

Part 2: Interaction Effects (8 marks)

In this section we will see how adding interaction effects changes the results of our analysis. You can complete this section using exclusively the `summary` and `lm` functions.

- a) Add an interaction between `category` and `ageMean` as independent variables to the regression in 1a. What can we conclude from the `factor(category)eCom:ageMean` coefficient? Answer in 30 words or less. *Hint: The best answers here will discuss what this coefficient means in qualitative terms.* (4 marks).
- b) Compare the regression in part 2a with the regression in 1a. What happens to the coefficient of `factor(category)eCom`? Why do you think this happened? Answer in 30 words or less. (4 marks).