Problem Set 3

Predictive Analytics for Business Strategy

Spring 2023

McDermott

Include everyone’s name on the submission and paste any relevant output into the problem set.

1. In this problem, you will use the Pulse data in Canvas to build a model for passive prediction. You will passively predict whether someone has gotten a Covid 19 Vaccine.
   1. What would you include and why? (include at minimum 3 independent variables)
   2. How can you tell if you did a good job.
   3. Give an example of a passive prediction. Show and explain the results (both the regression model estimates and how you calculate what you passively predict for Y or for the change in Y).
   4. Explain whether any of the variables you included in your model are control variables and why/why not.
   5. Does your model have a variable of interest? Explain.
2. Suppose you have observational data on student exam scores in worksheet 1 of PS3.xlsx.
   1. Report your results of this simple regression with a screenshot.
   2. Discuss the results.
   3. Discuss possible confounding factors.
   4. Which confounding factors are you likely to be able to use as control variables (meaning they wouldn’t actually be CFs) and which ones are you likely not able to (assume you aren’t necessarily limited to the data for the latter).
   5. Let’s say determination is a confounding factor (you won’t get credit for citing it above). Propose a proxy variable and discuss how it satisfies all 3 assumptions/requirements of a good proxy variable.
3. Suppose you have monthly observational data on sales of a specific children’s toy in worksheet 2 of PS3.xlsx.
   1. Report your results of this simple regression with a screenshot.
   2. Clearly explain and discuss the results (there are lots of categories so focus on month11, month12, and price).
   3. Discuss possible confounding factors.
   4. Which confounding factors are you likely to be able to use as control variables (meaning they wouldn’t actually be CFs) and which ones are you likely not able to (assume you aren’t necessarily limited to the data here).
4. Can you use estimates from an RCT to make a passive prediction? Explain why or why not.