

Problem Set #2: Empirical Methods

Due by 11:59 ET 11/9/20

1. From the lectures, we learned how to run a regression in both Excel and R and how to interpret the results both using normal variables, and after some variable transformations. Let's use those tools to analyze some data on our own! I've provided you with some data on Cigarette consumption (our dependent variable or Y variable) and we'll try to explain that using a number of different independent variables or X variables. Its under PS2_Data.xlsx. Using Excel (for now, I'm still being nice and not forcing R just yet), follow the instructions to see what happens when we try to control many factors impacting cigarette consumption and do some variable manipulation.

a. Using Excel, run a simple regression. where the number of cigarettes smoked is explained ONLY by the person's annual income (column D). You don't need the residuals or the plots at all so if they're checked, uncheck them. Have it put the results in N2. Interpret the coefficient on price (coefficient and p-value).

b. Now create a new variable in column E that is annual income squared. Run a second regression where Cigarettes smoked is explained by income and income². Put the results in N22. Interpret the results (and don't worry about the p-value here, the squared term won't be statistically significant).

c. Now create a new variable in column J that is the inverse of Miles from nearest store (so 1/Miles from Nearest Store). Run a regression where cigarettes smoked is explained by Miles from Store and the inverse of that. Put the results in N43. Interpret those results (and don't worry about the p-value here, the inverse term won't be statistically significant).

d. One last regression! Run a regression where the cigarettes smoked is explained by the information campaign. What kind of variable is the information campaign? Put the results in N64. Interpret the coefficient.

2. Listen to the Freakonomics podcast, "Nurses to the Rescue!" and answer the following questions.

<http://freakonomics.com/podcast/nurses-to-the-rescue/>

a. Turns out nurses are the most trusted profession and have been for a long time. Why do you think that is?

b. Why was the parental-leave program such a good opportunity to study the effectiveness of nurses? In other words, briefly explain what the research design/identification strategy of the study was?

c. Part of the healthcare issues in our country is the shortage of general practice doctors, specifically in rural areas. Do you think this is a compensation issue (the they don't get paid enough) or is it something else?

Bonus: Any solutions beyond the podcast's suggestions?

d. From an economic standpoint, what's the rationale behind their argument that we should plug the rural doctor shortage issue with nurses?

HINT: It rhymes with community lost but you can explain it without the term too!

If you could save your excel sheet from 1. with your regressions on it and upload it to Blackboard along with your Word doc written responses to 2., that would be great! Feel free to add your interpretations for 1. in the excel document if you like.