

Math Camp 2019 - Final Project

Your Friendly Instructors

8/10/2019

Card and Krueger (1994) - Data Analysis Project

You will be working data from a well-known 1994 study by David Card and Alan B. Krueger regarding the effect of raising minimum wage on employment. The data set contains information about the wages paid at a number of different fast food restaurants in Pennsylvania and New Jersey, recording them both before and after increases in New Jersey's minimum wage. The data set also contains information about the number of full-time and part-time workers at each location for the same time periods. Using this data, you will be exploring the effects of raising the minimum wage. As part of this, you will be using what you have learned about data analysis in **R** to explore the data set, visualize some of its features, fit a regression model, and make empirical claims concerning the results of your analysis.

When you have finished with the analysis, please write up your results in a short paper using RMarkdown and have it ready to submit by Sunday. Be sure to include `echo = TRUE` in your `knitr` options so that the resulting .pdf file includes your code chunks.

Variables in the Data Set

Variable Name	Description
chain	name of fast food restaurant
location	location of restaurant
wageBefore	wage before min. increased
wageAfter	wage after min. increased
fullBefore	# full-time employees before min. inc.
fullAfter	# full-time employees after min. inc.
partBefore	# part-time employees before min. inc.
partAfter	# part-time employees after min. inc.

Exploring the Data Set

Begin by reading in the data set and looking at its properties through the tasks below.

1. Use the appropriate function to read `minwage_final_project.csv`. Remember that on Windows machines, you may need to include `\\` instead of `\` in the filepath for R to find the file correctly.
 - How many variables are there in the data set? How many observations are there?
 - Use a function to check that the above variable names are correct.
 - Are there any missing values (NA entries) in the data set?
 - Create a table detailing how many observations there are for each state, and then for each coded location.
2. Using either base R or `ggplot2` functions, create a graph to visualize the distribution of wages across fast food chains *before* the minimum wage increase. Be sure to include appropriate labels/titles.
 - Discuss the wage distributions across fast food chains. Which chains differ significantly in terms of wage distribution when looking at the graph?
 - Make another graph, but this time for wages across fast food chains *after* the minimum wage increase. Discuss what comparing the two graphs suggests, if anything.

Manipulating the Data

Further exploring the data may require a little bit of manipulation, depending on the questions you wish to ask.

1. Calculate the mean wage in each state *both* before and after the minimum wage increase. It may be useful to recode the `location` variable to aggregate observations from the originally-coded locations, but be careful not to overwrite the `location` vector.
 - Which state experienced the largest relative change in wages based on comparison of mean wages before and after the minimum wage increase?
2. Calculate the mean number of full-time and part-time employees in each state *both* before and after the minimum wage increase.
 - What does comparing the number of full- and part-time employees in each state over the two time periods suggest about the effect of the minimum wage?
3. Using the original `location` vector to repeat this analysis by coded location, compare the difference in mean wage and numbers of full- and part-time workers before and after the minimum wage increase.
 - Does this help us understand anything new about the effect of the minimum wage increase?
4. It might be nice to know the difference in numbers of part- and full-time workers after the minimum wage increase. Create a new vector containing the difference between numbers of part- and full-time workers for each observation.

Hypothesis Testing and Linear Regression Modeling

You may have developed some intuition about what the minimum wage change appears to do to employment and wages. Now, use what you have learned to glean a more formal result, and see how it compares.

1. Based on what you know about the data set from the initial exploration, how would you characterize the effect of the increase in minimum wage? Be *specific* and *directional* (e.g. “Increasing the minimum wage will increase the number of full-time workers,” or “increasing the minimum wage will increase the mean wage in <location 1>” more than in <location 2>).
 - Although it is unlikely that these data will lead to anything groundbreaking, this is the basic structure of a hypothesis: a statement about the directional relationship between variables.
 - For brevity, we will consider the following two hypotheses:
 - H_1 : The increase in minimum wage will increase full-time employment in New Jersey.
 - H_2 : The increase in minimum wage will produce a larger increase in full-time employment for chains with higher pre-increase wages.

2. The nature of this minimum wage increase seems like it might be convenient. Code the minimum wage increase as an indicator (i.e. dichotomous) variable named `treatment`. `treatment` should be 1 for observations in New Jersey, as this is where the minimum wage increase occurred; conversely, it should be 0 for observations in Pennsylvania.
3. We are interested in knowing how the minimum wage increase affected overall wages and full-time employment, as well as the differences (if any) in these effects between fast food chain. Using linear regression, we can attempt to estimate these effects. Fit the following regression models, including covariates as appropriate.
 - Fit a linear regression model called `lmemp`. This model should estimate the effect of the minimum wage increase on full-time employment.
 - Fit a linear regression model called `lmwage`. This model should estimate the effect of the minimum wage increase on wages.
 - Are there any problems we might think of with using time-dependent variables in this kind of model?
4. Use `stargazer` to present these models in a succinct table of regression results with appropriate labels, title, etc.

Interpreting Results and Drawing Conclusions

Now that you have some regression coefficients to consider, use what you learned about interpreting regression results to evaluate the hypotheses you formed earlier. Be sure to reference p -values, α -levels, and relative magnitude of effect size in discussing whether or not these hypotheses are supported.

1. Is there any support for H_1 ?
2. Is there any support for H_2 ?
3. Based on (1) and (2), what appear to be the main consequences of increasing the minimum wage in New Jersey? Does the increase affect Pennsylvania in terms of wages or employment? Why? Do your results support the conclusions in Card and Krueger (1994)?