

QUANTITATIVE ANALYSIS

Data Paper Assignment

Your assignment is to test a hypothesis using statistics. For most of you, this will involve seeing if there is a relationship between two or more variables. Linear Regression is the technique that most of you will use to test your hypothesis. Depending on your question and your data, a few of you may end up using a “Chi Square” analysis (when your Y variable is categorical) or a logit or probit model (when your Y variable is discrete: 1,0...yes, no, etc). I recommend trying to avoid this technique unless you’ve seen it or used it before our class.

What does this mean in concrete terms?

Take the example of testing whether there is a relationship between a voluntary job training program for the unemployed and their future income. You would state a null and an alternative hypothesis. You would then test that hypothesis using data you found at the Bureau of Labor Statistics using simple linear regression. You might find that you can be 99% confident that those who used the program earned more 2 years after going through the program. Your predicted effect might be that a person going through the program earned \$3000 more per year than those who did not.

You will also want to add control variables such as education level, region of country and gender to see what happened to the effect of this job training program on income after employing the control variables. You will want to discuss and justify your use of these variables.

In addition to this analysis, you should:

- 1). Discuss the perfect experiment to get at your hypothesis
- 2). Discuss (besides the perfect experiment) what data you would like to have collected to test the hypothesis.
- 3). Discuss the different control variables you would like to have to answer the question ideally
 - 4) Discuss some of the issues below depending on their applicability to your project
 - A) Quality of the samples: random, how taken, etc
 - B) Measurement Error: was income measured correctly in this study, etc
 - C) Selection Effects: this job training program was voluntary, who joined it?
 - D). Policy vs. Statistical Significance

Other Guidelines and Thoughts

- 1) The paper part should be 2-5 pages (closer to 2 is better)

- 2) It should have an intro/background section, a results section and a theoretical/ideal analysis section
- 3) Somewhere in the paper, you need to make use of p-values, hypothesis tests, confidence intervals, R squared, and regression equations.
- 4) Be flexible between your ideal question and where data is readily available. Alter your question if you're really struggling to get data
- 5) You may use EXCEL, STATA, SPSS, SAS or any other program if you happen to know how to use them.

- 6). The minimum analysis needed to achieve a good grade is the basic hypothesis and then at least 2 control variables. That followed by a good write up of the ideal experiment and how a further analysis would work. This does not mean that you have to stop there. You're welcome to do more.

- 7). Besides the written part, you will do a 5-7 minute presentation of your project. You can either use powerpoint or you can hand out sheets to the class, or use no props if you'd like. The purpose is so we can learn from each other!

- 8). Have fun doing it! It can be annoying to find appropriate or decent data, but once you have your data, it should be fun to play around with it to create the best model you can.