LPS ECON-103-601, Fall 2014
Due 12.10.14

Statistical Discrimination Project

Her Majesty, Shonda Rhimes, the Queen of Shondaland, has made her mission to end economic gender discrimination in her land. Your project consists of two tasks. First, you must determine whether economic gender discrimination exists and to what extent. Second, you will assess the underlying causes of this discrimination or its absence using your statistical and economic knowledge. You will be assessed equally on well you apply the concepts we learned in this setting, perform data analysis in R, complete the two tasks above, present your results, and clearly/coherently communicate your ideas. For this project, I want you to think about the tools we learned and apply them while demonstrating basic skills in R. The guidelines for this project follow.

- Your project write up must be typed, no more than 25 pages including figures, 12 point font, 1in margins, double-spaced and include a cover page (which I will provide and does not count toward the page count)
- Your project write up must be in an essay/report form. You may break the writing into the sections below, but it should have a clear introduction and conclusion stating your objectives, what you do, and your findings.
- You must send your R code electronically as a $.R$ file to econstats103@gmail.com (Subject Line: Project-R-Code-yourNameHere). I must be able to replicate all the results in your project with your code. Any results, statistics, figures or table without supporting code will not count towards your mark.
- You may not consult with your classmates on the project or copy code. It will be considered a violation of academic integrity and result in a zero on the project.

Here is some advice for the writing the project.

- Clearly label all figures and tables
- Avoid run-on sentences and keep the subject and verb close in sentences
- Write the introduction and conclusion last
- Justify choices of tests and statistics

Here is some advice for the R Code.

- Use comments to separate sections of your code and make clear what your code does. A "#" before text indicates that it is a comment (e.g. # Mean Wage by Gender)
- Use white space to space out your code

Getting started with R

1. Go to *http://cran.r-project.org*, download and install R. Go to *http://www.rstudio.com*, download and install $R - Studio$.
2. R is the base program and *R-Studio* is the user interface
3. Complete the R-Tutorials 1-5 at *http://fditraglia.github.io/Econ103Public/*

Loading the Data

1. Download the data from the course site.
2. Use ProjectSampleCode.R to see how to load the .csv file. You must change the directory to where you put the data file.
3. Read ShondalandSurveyCodebook.rtf on the course site to become familiar with the variables in the dataset, the values they take on, and what these numbers mean.

Preliminaries

What is economic discrimination? Our definition is different pay or employment status to identical workers based on some characteristic unrelated to productivity like gender or ethnicity. Wage disparities are noticeable differences in the pay structure across characteristics unrelated to productivity, which could be due to gender-based discrimination or other factors. There may also be employment disparities like higher unemployment rates for women. We have a random sample from 2013 consisting of 80,000+ Shonalandnians that includes some demographic information, employment information, and wage information. Answering all questions in the survey was mandatory for all those randomly selected. Not all information about Shondalandnians is in the survey!

Week 1: Measuring Economic Disparities

Suggested Date of Completion: 11.09.14

Suggested Reading:

- Chapter 2: Descriptive Statistics
- http://www.statmethods.net/stats/withby.html
- http://www.statmethods.net/stats/descriptives.html
- http://www.statmethods.net/stats/frequencies.html
- http://www.r-bloggers.com/basics-of-histograms/

- http://www.statmethods.net/graphs/

We need to develop a sense for whether wages between men and women look different, so we begin by looking at wage disparities and unemployment rates. Choose the variables and statistics you want to present, justify your choices, and explain what the statistic means in terms of the potential for economic gender discrimination. For example, "I find that the average wage of men well exceeds the average wage of women, which leads me to hypothesis economic gender discrimination exists to a large extend." You will want to summarise variables by categories of variables. In which groups, are the wage disparities are largest? Be precise about the population you are considering. For example, you may only look at wages for full-time workers or workers age 16-65. Use the data to determine this information. If the sample is small, then explain why or why not this statistic be subject to small sample bias. One useful tool to present the distribution of a variable may be a histogram. It shows the frequency of each outcome of a variable in a bar graph (see the suggested reading above). Another useful tool may be a boxplot. The goal of this section is to compute statistics, interpret them, and present them in a concise, readable way to help the reader understand the extent of economic disparities in Shondaland (we will try to understand what underlies these disparities next). Do not forget to label your graphs and tables!

Week 2: Sources of Wage Disparities

Suggested Date of Completion: 11.16.14

Suggested Reading:

- http://www.statmethods.net/stats/correlations.html
- http://www.statmethods.net/graphs/line.html

Wage disparities may not necessarily be evidence of statistical discrimination. Why not? Suppose 60% of working women only work part-time and part-time wages are lower than full-time wages on average. Most men on the other hand work full-time. Then we should expect a disparity in wages due to the fact that women work mostly part-time, perhaps due to childcare purposes. Part-time work is a confounder for gender-based wage discrimination. The survey is a random sample, but there may also be issues with selection bias. Examine the relationship between being female and wages along with the relationships between being female, wages and other possible confounders. Use your economic knowledge to explain why a factor might be a confounder in addition to whatever statistical evidence (correlations, covariance) you provide. Also, explain why there may be a positive or negative, strong or weak relationship between variables (correlation close to 0 is weak, correlation close to 1 is strong). Be selective about the variables you consider and explain your choices (e.g. log daily wage vs. raw daily wage). What kinds of relationships might correlations be missing out on? (Hint: Take the average log dailywage by age and plot it against age) Plots may be useful for showing relationships between variables.

Week 3: Statistical Evidence

Suggested Date of Completion: 11.23.14

Now that we have documented the situation in Shondaland, we want to get more precise about our statements. For example, we can formally test whether the average wages of men exceed those of women. Use the tools at your disposal including the Central Limit Theorem, confidence intervals and hypothesis testing to verify your previous hypotheses. For example, do the confidence intervals for the wages of men and women overlap in this sample? If so, we cannot infer that average wages of men and women differ much. We can use a hypothesis test to test equality of average wages or equality of wage variance. Perhaps, *i.i.d.* is a bad assumption to apply the Central Limit Theorem across all male or female wages but a better assumption if we look at wages within age, occupation, and education groups (e.g. men and women in their 20's working in the service sector with a high school diploma). You may want to redefine the education groups to just a few (e.g. below secondary school, high school diploma or equivalent, some college, bachelor's, masters or professional degree, Ph.D.). What happens to your confidence intervals and the results of your hypothesis tests once you look within groups? Do the confidence intervals become more narrow or wider? Do you draw different conclusions from your hypothesis test when zooming in on wage and employment disparities with groups? Do you find other wage disparities on other dimensions like race?

Week 4: Regression Analysis

Suggested Date of Completion: 11.30.14

Looking at the previous statistical results, it should be evident that confounders are everywhere. So we want to model wages statistical allowing for all kinds of variables to affect wages. Typically, labour economist model log wages as linear in on experience (or age) and education in the following way.

$$\log w_i = \beta_0 + \beta_1 age_i + \beta_2 age_i^2 + \beta_3 education_i + \varepsilon_i$$

ε_i is the error term. Add to this regression. In particular, add variables which you think determine wages and variables on which discrimination may occur. If you still find a significant relationship between being a woman and wages after controlling for all relevant variables, then you can confidently say whether economic gender discrimination is present in Shondaland. You may not have all the relevant variables in the data. Which might be important that you do not have? Why might they be important? How would their omission in the regression bias your estimates of the marginal effects (i.e. the coefficient) of a change in the variables? Be selective in the variables you choose. Perhaps, you will want to use the F-Test to select between different regression specifications. How much variation in wages do you chosen variables explain?

Remark With log wages, the coefficients have a different interpretation. A one unit increase in an independent variable changes the dependent variable by the coefficient in percentage terms. For example, one year increase in age changes wages by $(\beta_1 + 2\beta_2 \cdot age)\%$.

Week 5: Eliminating Discrimination

Suggested Date of Completion: 12.07.14

In light of the evidence you presented, write your introduction and conclusion. You should state your thesis or key claim in a clear, concise manner. You should briefly summarise to what extent economic gender discrimination exists and along what dimensions as support for your thesis statement. Mention other dimensions seem to be important factors in determining wages and detecting discrimination. Make suggestions for how to eliminate wage disparities or discrimination between the genders given your thesis and evidence. For example, perhaps Queen Shonda encourage women to work full-time to eliminate wage disparities, because part-time workers are paid less (despite working less) and part-time work confounds different pay by gender. Your introduction and conclusion should consists of meaningful (economic) statements supported by statistical evidence. You will earn no points by merely summarizing your results. You must interpret them in an economic/policy framework and connect your results to your "big picture" or thesis.