IS4300 Final Project

Bonnie Ritchie

5/3/2021

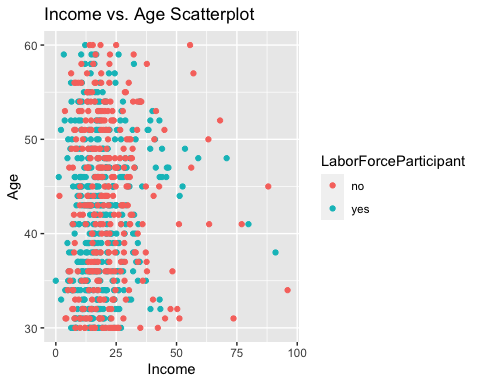
## Introduction

The data set that we are going to look at is titled “U.S. Women’s Labor Force Participation”. The data set can be found on GitHub. This data set includes 753 observations with eight variables. The variable *lfp* which was changed to *LaborForceParticipant* is a factor variable on whether the woman is a participant in the labor force. *k5* was changed to *KidsUnder5* and is how many children five years or younger the woman has. *k618* was changed to *Kids\_6to18* and is how many children ages 6 to 18 the woman has. Then there’s *age* and the variables *Wife\_College* and *Husband\_College* which detail if the wife and husband attended college. The final two variables are *Expected\_Wage\_Rate* which is a log expected wage rate: for women participating in the labor force, it is their actual wage rate, if they are not a participant, then it is an imputed value. The head of the data set can be seen below.

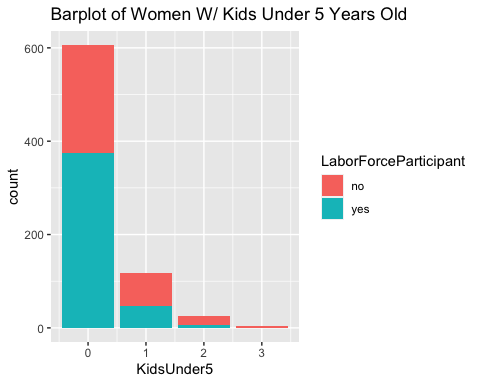
## LaborForceParticipant KidsUnder5 Kids\_6to18 Age Wife\_College Husband\_College  
## 1 yes 1 0 32 no no  
## 2 yes 0 2 30 no no  
## 3 yes 1 3 35 no no  
## 4 yes 0 3 34 no no  
## 5 yes 1 2 31 yes no  
## 6 yes 0 0 54 no no  
## Expected\_Wage\_Rate Income  
## 1 1.2101647 10.910  
## 2 0.3285041 19.500  
## 3 1.5141279 12.040  
## 4 0.0921151 6.800  
## 5 1.5242802 20.100  
## 6 1.5564855 9.859

## Visualizations

The first visualization of this data set shown below is a scatter plot of the income versus the age colored by whether or not the woman is a participant in the labor force. As one can see from the visualization, there is a relatively even spread of labor force participants and non-participants.



The next visualization gives a broad overview of how women who participate in the workforce tend to have less children. This bar plot detailing the number of kids under five each woman has colored by whether they are a labor force participant or not. One can see that more than fifty percent of those with zero children under five years old are labor force participants. Then as you look at the bars for one, two, and three kids, the majority becomes non-labor force participants.



## Analysis

The first test conducted on this data set was a two sample t-test. The data was split by whether the wife went to college or not and the t-test conducted to see if the means are statistically different. After filtering the data, the mean expected wage rate for women who did and didn’t college was as follows:

## [1] 1.389064

## [1] 0.9827096

Then when you run the t-test, the p-value it returns is 1.346e-14. Since this is less than .05, we can reject the null hypothesis that the means for expected wage rate of women who did and did not attend college are the same.

##   
## Welch Two Sample t-test  
##   
## data: x$Expected\_Wage\_Rate and y$Expected\_Wage\_Rate  
## t = -8.0821, df = 317.66, p-value = 1.346e-14  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.5052752 -0.3074337  
## sample estimates:  
## mean of x mean of y   
## 0.9827096 1.3890640

The next test to run is whether the mean income for households where the husband attended college is the same as the mean income for households where the husband did not attend college. The data was split again into a set where *Husband\_College* equals “yes” and a set where *Husband\_College* equals “no”. The mean income amount for households whose husband did and did not attend college are as follows:

## [1] 25.36128

## [1] 16.75881

The t-test for this comes out to be 2.2e-16 and since it is less than .05, we can reject the null hypothesis that the mean income for households whose husband attended college and the mean income for households whose husband did not attend college are statistically different.

##   
## Welch Two Sample t-test  
##   
## data: x$Income and y$Income  
## t = -9.7992, df = 468.48, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -10.327528 -6.877412  
## sample estimates:  
## mean of x mean of y   
## 16.75881 25.36128

## Conclusion

Based on the findings of the t-tests done on this data, one can see that attending college, whether that be the wife or husband, does significantly affect the income and wage rate. The difference between mean expected wage rates of women who did and did not attend college is .406 while the difference in income of men who did and did not attend college is 8.602. Hopefully, one can see through this data set how helpful a college education can be.