Team 5 Ladonde Study Project

10/2/2019

SUMMARY

For this assignment we were working in our MIDS Assigned Teams, looking at a Dataset to determine whether vocational training through the National Support Work (NSW) to answer several key research questions on its effectiveness. To do this we used a series of methods, including Teamwork, Critical Thinking, Exploratory Data Analysis, and Logistic Regression which ultimately lead to the model below.

Once we had built our model we were able to answer key research questions purposed by the NSW.

Research Quesions

Is there evidence that workers who receive job training tend to earn higher wages than workers who do not receive job training?

Quantify the effect of the treatment, that is, receiving job training, on real annual earnings.

What is a likely range for the effect of training?

Is there any evidence that the effects differ by demographic groups?

Are there other interesting associations with wages that are worth mentioning?

INTRODUCTION

In the 1970s, researchers in the United States ran several randomized experiments intended to evaluate public policy programs. One of these was the National Supported Work (NSW) Demonstration, in which researchers wanted to assess whether job training for disadvantaged workers influenced their wages. Eligible workers were randomly assigned either to receive job training or not to receive job training. Candidates eligible for the NSW were randomized into the program between March 1975 and July 1977.

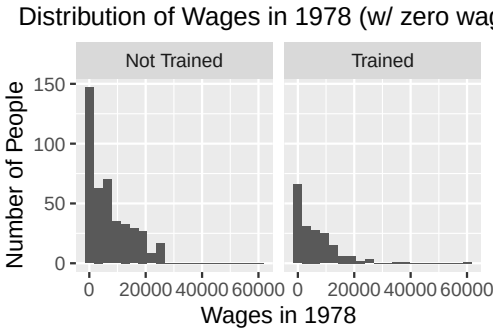
From here we read in the data from the Lalonde Study and decide to use 1978 as a response variable. The reasoning behind this is subjects that participated in the study received their job training during 1975. This meant that by 1978 they had had ample time to adjust and apply the skills they had learned in training to their work. This would allow us to truly see the effectiveness the NSW Demonstration had had.

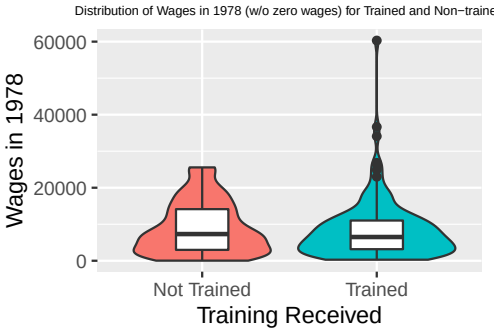
DATA AND EDA (Incomplete)

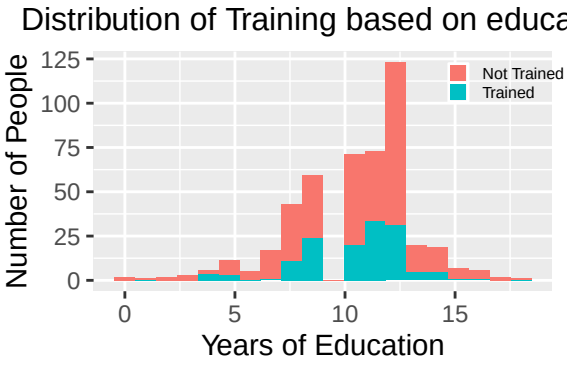
Before we could build our model, we first had to do some Exploratory Data Analysis (EDA). Here we could iteratively inspect variables and determine whether they were likely to be added to our final model.

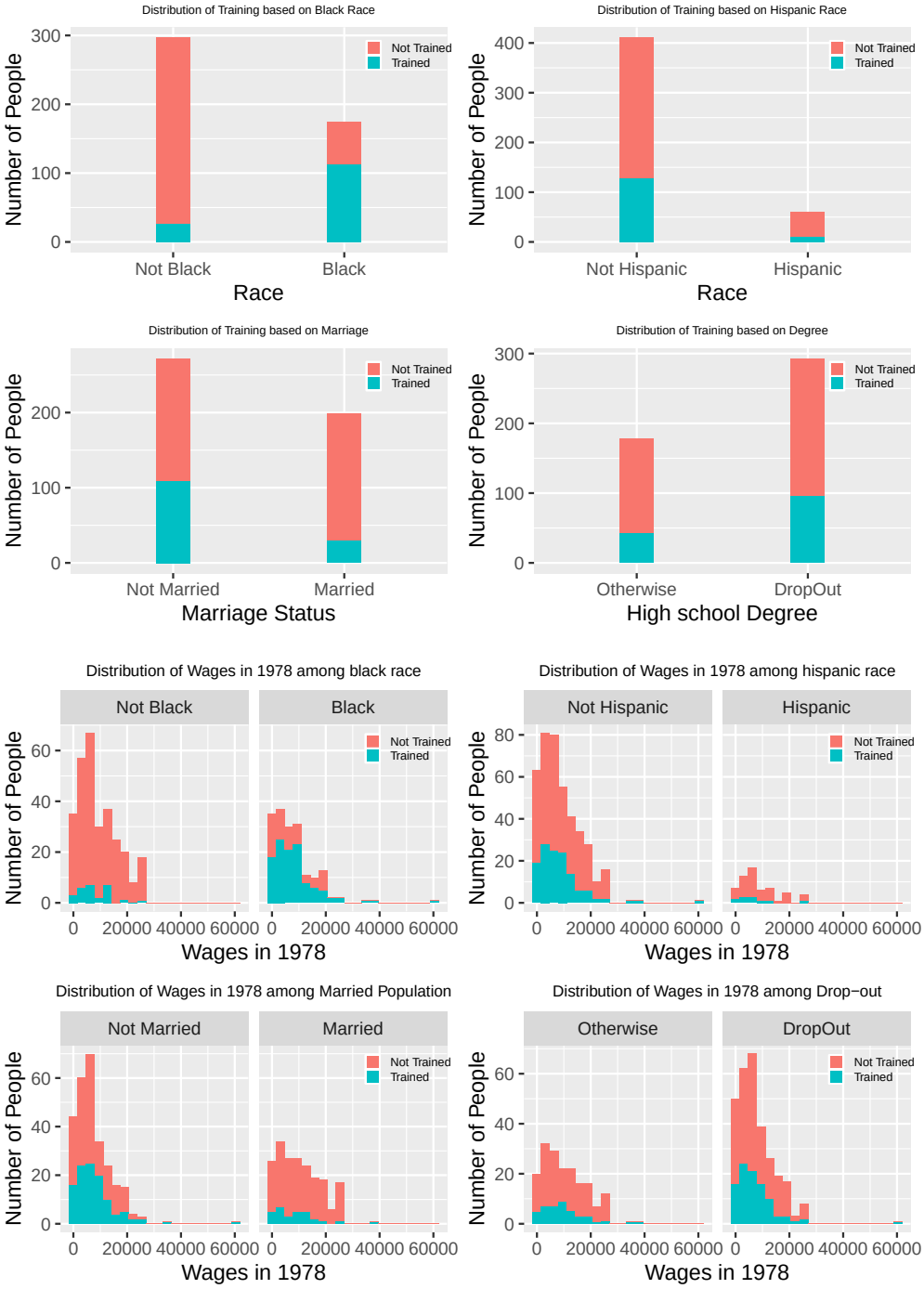
For this we primarily focused on things that we felt may have an effect on the income changing in 1978, and overlaying them with whether or not that individual had been treated. From this we could make some inferences on what factors would best fit into the final model.

There is an issue with this sections code







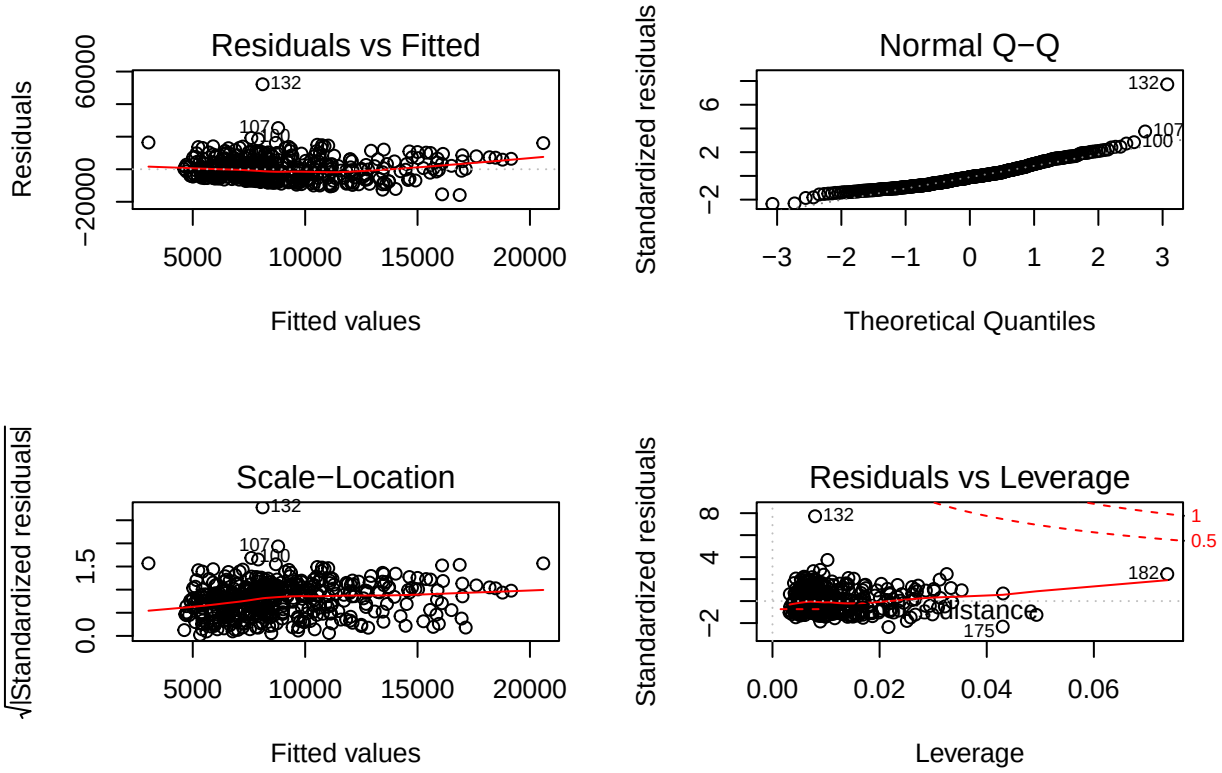


MODELING

$$Incomein1978 = \beta_0 + \beta_1 Treatment + \beta_2 Education + \beta_3 Incomein1974 + \epsilon, \epsilon \sim N(0, variance^2)$$

term	estimate	std.error	statistic	p.value
(Intercept)	294.2086004	1706.5336199	0.1724013	0.8631969
treatTrained	730.8885156	713.2212346	1.0247711	0.3060029

term	estimate	std.error	statistic	p.value
educ	395.5808344	125.5721876	3.1502265	0.0017364
re74	0.3308532	0.0536771	6.1637647	0.0000000
age	97.4638979	38.0796967	2.5594715	0.0107977




```
##
## Call:
## lm(formula = re78 ~ treat + educ + re74 + age, data = dataset[which(dataset$re78 >
##     0), ])
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -15855  -5098   -698   3614  52202
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.942e+02  1.707e+03   0.172  0.86320
## treatTrained 7.309e+02  7.132e+02   1.025  0.30600
## educ         3.956e+02  1.256e+02   3.150  0.00174 **
## re74         3.308e-01  5.368e-02   6.164 1.54e-09 ***
## age          9.746e+01  3.808e+01   2.559  0.01080 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6792 on 466 degrees of freedom
```

```
## Multiple R-squared:  0.1607, Adjusted R-squared:  0.1535
## F-statistic:  22.3 on 4 and 466 DF,  p-value: < 2.2e-16
```

```
##                      2.5 %       97.5 %
## (Intercept)  -3050.5358330 3638.9530337
## treatTrained  -666.9994171 2128.7764484
## educ           149.4638692  641.6977995
## re74             0.2256479    0.4360584
## age             22.8290639  172.0987320
```