**STOR 455 Homework 6**

**15 points - Due Wednesday 10/27 5:00pm**

**Are Emily and Greg More Employable Than Lakisha and Jamal?**

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review, 94*(4), pp. 991-1013.

*Abstract*

We perform a field experiment to measure racial discrimination in the labor market. We respond with fictitious resumes to help-wanted ads in Boston and Chicago newspapers. To manipulate perception of race, each resume is randomly assigned either a very African American sounding name or a very White sounding name. The results show significant discrimination against African-American names: White names receive 50 percent more callbacks for interviews. We also find that race affects the benefits of a better resume. For White names, a higher quality resume elicits 30 percent more callbacks whereas for African Americans, it elicits a far smaller increase. Applicants living in better neighborhoods receive more callbacks but, interestingly, this effect does not differ by race. The amount of discrimination is uniform across occupations and industries. Federal contractors and employers who list “Equal Opportunity Employer” in their ad discriminate as much as other employers. We find little evidence that our results are driven by employers inferring something other than race, such as social class, from the names. These results suggest that racial discrimination is still a prominent feature of the labor market.

| **Variables** | **Descriptions** |
| --- | --- |
| *call* | Was the applicant called back? (1 = yes; 0 = no) |
| *ethnicity* | indicating ethnicity (i.e., “Caucasian-sounding” vs. “African-American sounding” first name) |
| *sex* | indicating sex |
| *quality* | Indicating quality of resume. |
| *experience* | Number of years of work experience on the resume |
| *equal* | Is the employer EOE (equal opportunity employment)? |

Use the *ResumeNames455* found at the address below:

<https://raw.githubusercontent.com/JA-McLean/STOR455/master/data/ResumeNames455.csv>

library(readr)

ResumeNames455 = read\_csv("https://raw.githubusercontent.com/JA-McLean/STOR455/master/data/ResumeNames455.csv")

1. Construct a logistic model to predict if the job applicant was called back using *experience* as the predictor variable.

mod1 = glm(call~experience, data=ResumeNames455, family = binomial)

summary(mod1)

##

## Call:

## glm(formula = call ~ experience, family = binomial, data = ResumeNames455)

##

## Deviance Residuals:

## Min 1Q Median 3Q Max

## -0.7780 -0.4075 -0.3924 -0.3779 2.3598

##

## Coefficients:

## Estimate Std. Error z value Pr(>|z|)

## (Intercept) -2.75960 0.09620 -28.687 < 2e-16 \*\*\*

## experience 0.03908 0.00918 4.257 2.07e-05 \*\*\*

## ---

## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##

## (Dispersion parameter for binomial family taken to be 1)

##

## Null deviance: 2726.9 on 4869 degrees of freedom

## Residual deviance: 2710.2 on 4868 degrees of freedom

## AIC: 2714.2

##

## Number of Fisher Scoring iterations: 5

1. Plot the raw data and the logistic curve on the same axes.

plot(jitter(call, amount=0.1)~experience, data=ResumeNames455)

logit = function(B0, B1, x){

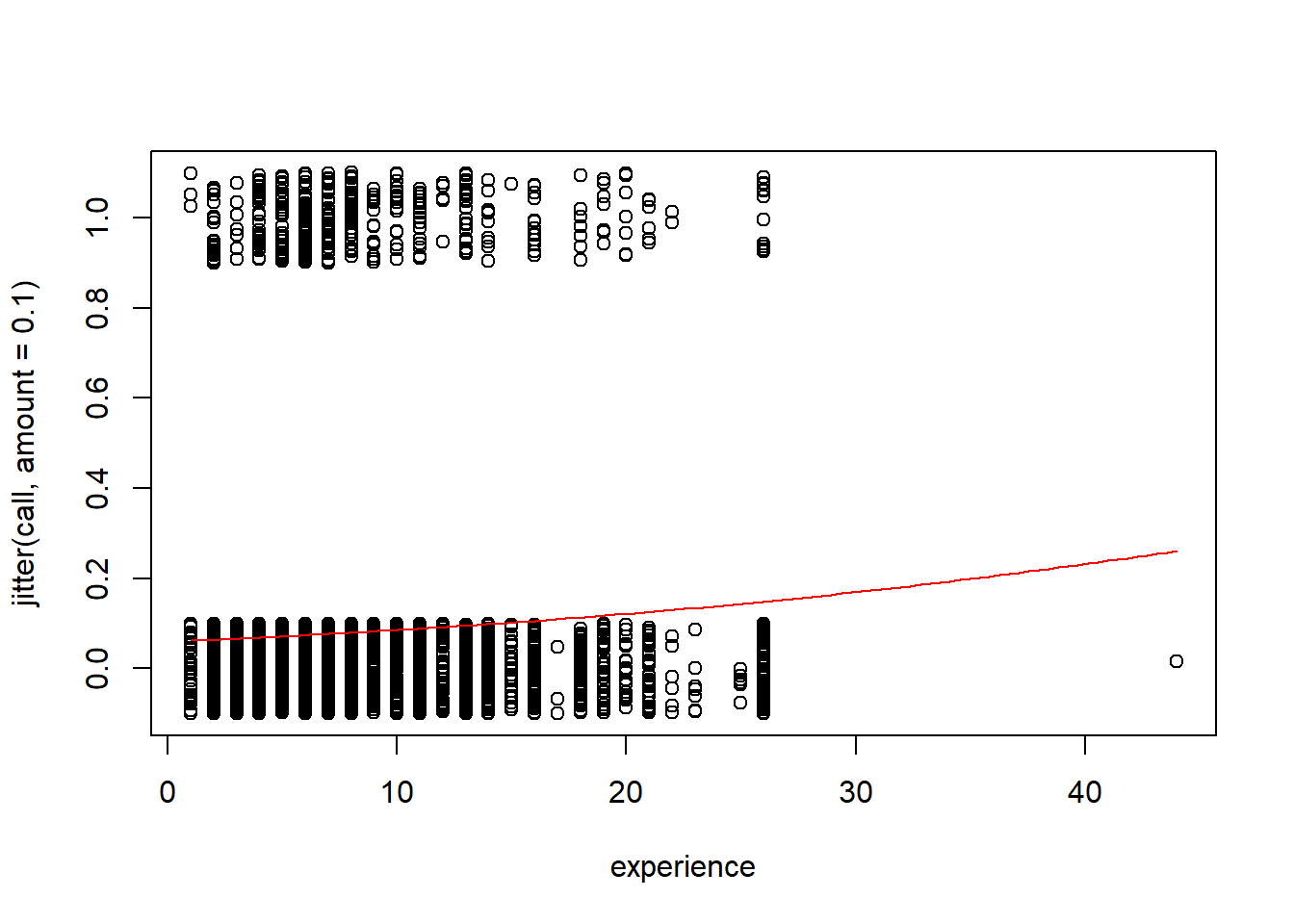
exp(B0 + B1\*x)/(1 + exp(B0 + B1\*x))

}

B0 = summary(mod1)$coef[1]

B1 = summary(mod1)$coef[2]

curve(logit(B0, B1, x), add=TRUE, col="red")



1. For an applicant with 3 years of experience, what does your model predict is the probability of this applicant getting called back?

# without type='response' you are predicting the log(odds)

applicant = data.frame(experience = 3)

predict(mod1, applicant, type="response")

## 1

## 0.06646115

1. Construct an empirical logit plot and comment on the linearity of the data.

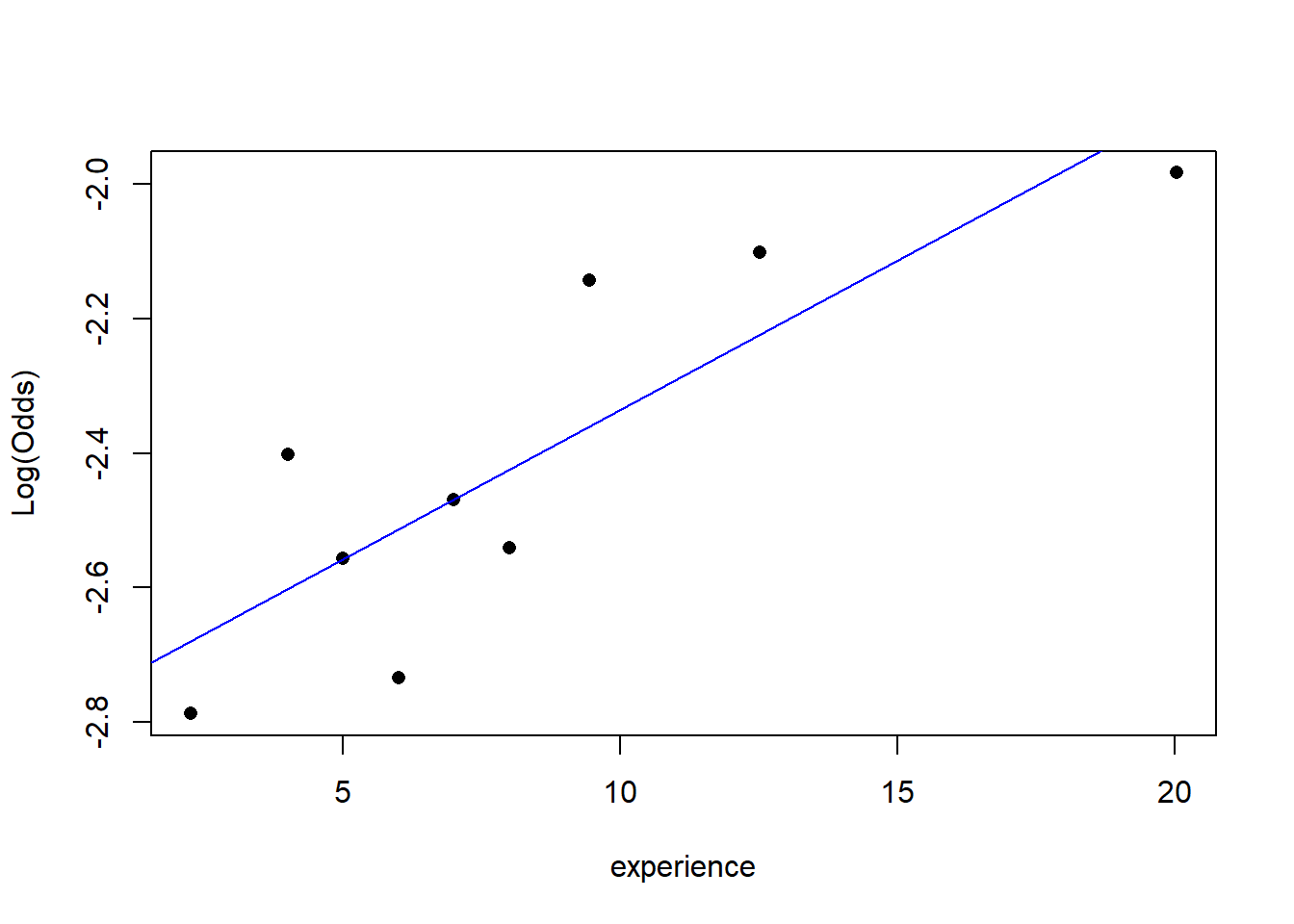
There are no clear nonlinear patterns in the data, so the logistic model seems appropriate

library(Stat2Data)

# ngroups=9 is arbitrarily chosen.

# This data was causes error in the function for many choices of breaks

emplogitplot1(call~experience, data=ResumeNames455, ngroups = 9)



1. Use the model from question #1 to perform a hypothesis test to determine if there is significant evidence of a relationship between *call* and *experience*. Cite your hypotheses, p-value, and conclusion in context.

H0: β1 = 0  
HA: β1 ≠ 0

Since the p-value (2.07e-05 using the summary or 4.298e-05 using log likelihood from anova) is less than 0.05, there is evidence to suggest that the coefficient of the experience term in the binary logistic model is nonzero.

# Either method could be used

summary(mod1)

##

## Call:

## glm(formula = call ~ experience, family = binomial, data = ResumeNames455)

##

## Deviance Residuals:

## Min 1Q Median 3Q Max

## -0.7780 -0.4075 -0.3924 -0.3779 2.3598

##

## Coefficients:

## Estimate Std. Error z value Pr(>|z|)

## (Intercept) -2.75960 0.09620 -28.687 < 2e-16 \*\*\*

## experience 0.03908 0.00918 4.257 2.07e-05 \*\*\*

## ---

## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##

## (Dispersion parameter for binomial family taken to be 1)

##

## Null deviance: 2726.9 on 4869 degrees of freedom

## Residual deviance: 2710.2 on 4868 degrees of freedom

## AIC: 2714.2

##

## Number of Fisher Scoring iterations: 5

anova(mod1, test="Chisq")

|  |
| --- |
|  |

|  | **Df**  **<int>** | **Deviance**  **<dbl>** | **Resid. Df**  **<int>** | **Resid. Dev**  **<dbl>** | **Pr(>Chi)**  **<dbl>** |
| --- | --- | --- | --- | --- | --- |
| NULL | NA | NA | 4869 | 2726.921 | NA |
| experience | 1 | 16.73516 | 4868 | 2710.186 | 4.2977e-05 |

2 rows

1. Construct a confidence interval for the odds ratio for your model and include a sentence interpreting the interval in the context.

We are 95% confident that for each 1 year increase in experience, the odds of getting called back will increase by a factor between approximately 1.02 and 1.06.

exp(confint(mod1))

## Waiting for profiling to be done...

## 2.5 % 97.5 %

## (Intercept) 0.05235403 0.07634629

## experience 1.02097262 1.05841947

1. For each 5-year increase in *experience*, how does your model predict the odds will change for the applicant getting called back?

For each 5 year increase in experience, the odds of getting called back will increase by a factor of approximately 1.22.

exp(summary(mod1)$coef[2,1]) ^ 5

## [1] 1.215796

# or by multiplying before using exp()

exp(5 \* summary(mod1)$coef[2,1])

## [1] 1.215796

In homework #7 we will continue with this data to investigate how the other variables impact an applicant’s chances of being called back.