STOR 455 Homework 7

20 points - Due Wednesday 11/3 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

1) An Equal Opportunity Employer (EOE) is an employer who agrees not to discriminate against any employee or job applicant because of race, color, religion, national origin, sex, physical or mental disability, or age. Construct a logistic model to predict if the job

applicant was called back using *ethnicity*, *equal*, and the interaction between *ethnicity* and *equal* as the predictor variables.

```
resume <- read.csv("https://raw.githubusercontent.com/JA-McLean/STOR455/master/data/ResumeNames455.csv")
library(Stat2Data)
library(readr)
library(TTR)
library(bestglm)

## Loading required package: leaps

library(leaps)

employ.mod = glm(call~factor(ethnicity)+equal+factor(ethnicity)*equal,
data=resume, family=binomial)
summary(employ.mod)

##
## Call:
## glm(formula = call ~ factor(ethnicity) + equal + factor(ethnicity) *
##     equal, family = binomial, data = resume)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.4662  -0.4440  -0.3697  -0.3697   2.3676
##
## Coefficients:
##                                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)                      -2.64903    0.09691 -27.334  < 2e-16 ***
## factor(ethnicity)cauc             0.38163    0.12730   2.998  0.00272 **
## equalyes                         -0.09106    0.18479  -0.493  0.62219
## factor(ethnicity)cauc:equalyes  0.19372    0.23713   0.817  0.41396
## ---
## 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: 2709.2  on 4866  degrees of freedom
## AIC: 2717.2
##
## Number of Fisher Scoring iterations: 5
```

2) Conduct a drop in deviance hypothesis test to determine the effectiveness of the *equal* terms in the model constructed in the previous question. Cite your hypotheses, p-value, and conclusion in context.

Ho: Xequals = 0 Ha: Xequals =/= 0

P-values = 0.8291

Conclusion: Because we recieved a p-value of 0.8291, we do not have enough evidence to reject the null hypothesis that the coeffecient of the “equals” variable is equal to 0.

```
anova(employ.mod, test = "Chisq")

## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: call
##
## Terms added sequentially (first to last)
##
##
##                          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                                      4869     2726.9
## factor(ethnicity)         1  16.9832      4868     2709.9 3.771e-05 ***
## equal                     1   0.0466      4867     2709.9    0.8291
## factor(ethnicity):equal  1   0.6714      4866     2709.2    0.4126
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3) Based on your model from question 1, What is the probability of an applicant with a “Caucasian-sounding” name getting a call back from an Equal Opportunity Employer (EOE). What is the probability of an applicant with an “African-American sounding” name getting a call back from an Equal Opportunity Employer (EOE)?

There is a 10.29% chance that someone with a caucasian sounding name will receive a call back from an EOE and there is a 9.38% chance that an African-American sounding name will receive a call from an EOE.

```
summary(employ.mod)

##
## Call:
## glm(formula = call ~ factor(ethnicity) + equal + factor(ethnicity) *
##     equal, family = binomial, data = resume)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.4662  -0.4440  -0.3697  -0.3697   2.3676
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   -2.64903    0.09691 -27.334  < 2e-16 ***
## factor(ethnicity)cauc          0.38163    0.12730   2.998  0.00272 **
## equalyes                      -0.09106    0.18479  -0.493  0.62219
## factor(ethnicity)cauc:equalyes  0.19372    0.23713   0.817  0.41396
## ---
## 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: 2709.2  on 4866  degrees of freedom
## AIC: 2717.2
##
## Number of Fisher Scoring iterations: 5
```

```
B0 <- summary(employ.mod)$coeff[1,1]
B1 <- summary(employ.mod)$coeff[2,1]
B2 <- summary(employ.mod)$coeff[3,1]
B3 <- summary(employ.mod)$coeff[4,1]

# Caucasian-sounding name
x = 1
y = 1
z = x*y
# Q = exp(B0+(B1*x)+(B2*y)+(B3*z))
# Q
P = exp(B0+(B1*x)+(B2*y)+(B3*z))/(exp(B0+(B1*x)+(B2*y)+(B3*z))+1)
P
```

```
## [1] 0.1029619
```

```
# African-American Sounding Name
a = 1
b = 0
c = a*b
# O = exp(B0+(B1*a)+(B2*b)+(B3*c))
# O
D = exp(B0+(B1*a)+(B2*b)+(B3*c))/(exp(B0+(B1*a)+(B2*b)+(B3*c))+1)
D
```

```
## [1] 0.09385863
```

4) Does the number of years of work experience impact the relationship between ethnicity and an applicant getting called back? Construct a logistic model to predict if the job applicant was called back using *ethnicity*, *experience*, and the interaction between *ethnicity* and *experience* as the predictor variables.

```
employ.mod.q4 =
glm(call~factor(ethnicity)+experience+factor(ethnicity)*experience,
data=resume, family=binomial)
summary(employ.mod.q4)
```

```
##
## Call:
## glm(formula = call ~ factor(ethnicity) + experience + factor(ethnicity) *
##     experience, family = binomial, data = resume)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
```

```
## -0.7558  -0.4337  -0.4052  -0.3472   2.4686
##
## Coefficients:
##                                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)                      -3.042323   0.150186 -20.257  < 2e-16 ***
## factor(ethnicity)cauc             0.512792   0.195743   2.620  0.00880 **
## experience                        0.043988   0.014052   3.130  0.00175 **
## factor(ethnicity)cauc:experience -0.008465   0.018570  -0.456  0.64852
## ---
## 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: 2693.0  on 4866  degrees of freedom
## AIC: 2701
##
## Number of Fisher Scoring iterations: 5
```

5) Conduct a drop in deviance hypothesis test to determine the effectiveness of the *experience* term in the model constructed in the previous question. Cite your hypotheses, p-value, and conclusion in context.

Ho: Xexperiences = 0 Ha: Xexperiences =/= 0

P-values = 0.8291

Conclusion: Because we recieved a p-value of 4.406e-05, we have enough evidence to reject the null hypothesis that the coeffecient of the “experiences” variable is equal to 0.

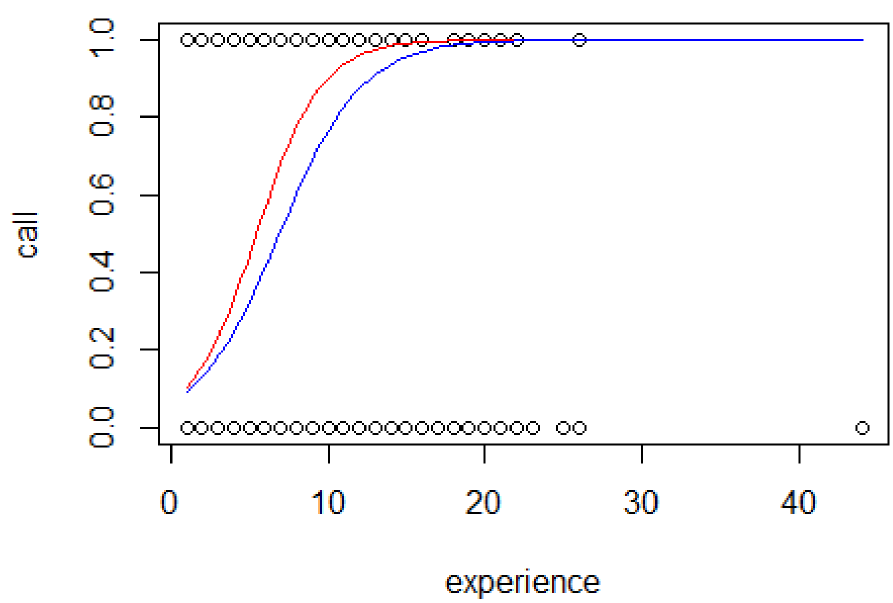
```
anova(employ.mod.q4, test = "Chisq")

## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: call
##
## Terms added sequentially (first to last)
##
##
##                             Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                                         4869     2726.9
## factor(ethnicity)            1  16.9832      4868     2709.9 3.771e-05
***
## experience                   1  16.6881      4867     2693.2 4.406e-05
***
## factor(ethnicity):experience  1   0.2071      4866     2693.0    0.6491
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

6) Construct a plot with *experience* on the horizontal axis and *call* on the vertical axis. Add to this plot two curves, made from the model constructed in question 4. For an applicant with a "Caucasian-sounding" name, plot a red logistic curve showing the probability of getting a call back based on experience. For an applicant with an "African-American sounding" name, plot a blue logistic curve showing the probability of getting a call back based on experience. Comment on the similarities or differences between the two models.

Based on the two models, it appears that caucasian sounding names have a higher rate of call back. Overall, as experience increases, the chance of recieving a call back also appears to increase.

```
plot(call~experience, resume)
#Caucasian
curve(exp(B0+(B1*x)+(B2*x)+(B3*x))/(exp(B0+(B1*x)+(B2*x)+(B3*x))+1), col =
"red", add = TRUE)
# African American
curve(exp(B0+(B1*x))/(exp(B0+(B1*x))+1), col = "blue", add = TRUE)
```



7) Use an appropriate model selection method to construct a best model to predict if the job applicant was called back using any of the other variables as predictors (except for *name*). You may also use interaction terms. Why would you not want to use *name* as a predictor?

You don't want to use name as a predictor because you can't logically level the names of the people in the dataset. Furthermore, it doesn't make sense to want to use name as a predictor because the rate of something happening to someone with a specific name isn't very useful to know. Based on the best models output, the one with the lowest criteria says to include ethnicity and experience.

```
resume <- read.csv("https://raw.githubusercontent.com/JA-McLean/STOR455/master/data/ResumeNames455.csv", stringsAsFactors = TRUE)
resume.2 = within(resume, {name = NULL})
head(resume.2)
```

```
##      sex ethnicity quality call experience equal
## 1 female      cauc     low    0          6   yes
## 2 female      cauc    high    0          6   yes
## 3 female      afam     low    0          6   yes
## 4 female      afam    high    0          6   yes
## 5 female      cauc    high    0         22   yes
## 6   male      cauc     low    0          6   yes
```

```
resume.2 <- resume.2[,c(1:3,5:6,4)]
head(resume.2)
```

```
##      sex ethnicity quality experience equal call
## 1 female      cauc     low          6   yes    0
## 2 female      cauc    high          6   yes    0
## 3 female      afam     low          6   yes    0
## 4 female      afam    high          6   yes    0
## 5 female      cauc    high         22   yes    0
## 6   male      cauc     low          6   yes    0
```

```
resume.2.bestglm <- bestglm(resume.2, family=binomial)
```

```
## Morgan-Tatar search since family is non-gaussian.
```

```
resume.2.bestglm$BestModels
```

```
##     sex ethnicity quality experience equal Criterion
## 1 FALSE      TRUE   FALSE       TRUE FALSE  2710.231
## 2 FALSE      TRUE    TRUE       TRUE FALSE  2716.616
## 3  TRUE      TRUE   FALSE       TRUE FALSE  2717.918
## 4 FALSE      TRUE   FALSE      FALSE FALSE  2718.429
## 5 FALSE     FALSE   FALSE       TRUE FALSE  2718.677
```

```
employ.mod.2 =
glm(call~factor(ethnicity)+experience+factor(ethnicity)*experience,
data=resume.2, family=binomial)
summary(employ.mod.2)
```

```
##
## Call:
## glm(formula = call ~ factor(ethnicity) + experience + factor(ethnicity) *
##     experience, family = binomial, data = resume.2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
```

```
## -0.7558  -0.4337  -0.4052  -0.3472   2.4686
##
## Coefficients:
##                                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)                      -3.042323   0.150186 -20.257  < 2e-16 ***
## factor(ethnicity)cauc             0.512792   0.195743   2.620  0.00880 **
## experience                        0.043988   0.014052   3.130  0.00175 **
## factor(ethnicity)cauc:experience -0.008465   0.018570  -0.456  0.64852
## ---
## 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: 2693.0  on 4866  degrees of freedom
## AIC: 2701
##
## Number of Fisher Scoring iterations: 5
```