Employment Bias

Katelyn Patricio

2025-07-01

Goal: Analyze the presence and impact of racial, gender, and industry-related bias in callback rates using the ResumeNames dataset, with a focus on identifying which factors most influence interview opportunities.

First we need to transform the data into binary form for the model to work

```
ResumeNames$call <- ifelse(ResumeNames$call == "no", 0, 1)
```

Is there racial bias?

```
race <- lm(call ~ ethnicity, data = ResumeNames)
summary(race)
```

Looking at the model outcome, White applicants have a 9.65% chance of getting a callback while African Americans have 3.2% less of a chance (around 6.45%). With the p-value < 5 (statistically significant), there is an unlikely chance that this outcome happened by chance.

```
##
## Call:
## lm(formula = call ~ ethnicity, data = ResumeNames)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.09651 -0.09651 -0.06448 -0.06448  0.93552
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.096509   0.005505  17.532  < 2e-16 ***
## ethnicityafam -0.032033   0.007785  -4.115 3.94e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2716 on 4868 degrees of freedom
## Multiple R-squared:  0.003466,   Adjusted R-squared:  0.003261
## F-statistic: 16.93 on 1 and 4868 DF,  p-value: 3.941e-05
```

Is racial bias intersectional?

```
rg <- lm(call ~ gender * ethnicity, data = ResumeNames)
summary(rg)
```

There doesn't seem to be a significant intersectional bias. Overall, African Americans have a lower chance of receiving a callback with no difference between males and females.

```
##
## Call:
## lm(formula = call ~ gender * ethnicity, data = ResumeNames)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.09892 -0.09892 -0.06628 -0.06628  0.94171
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              0.088696   0.011329   7.829     6e-15 ***
## genderfemale             0.010229   0.012963   0.789    0.4301
## ethnicityafam           -0.030408   0.016211  -1.876    0.0607 .
## genderfemale:ethnicityafam -0.002239 0.018482  -0.121    0.9036
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2717 on 4866 degrees of freedom
## Multiple R-squared:  0.003669,   Adjusted R-squared:  0.003054
## F-statistic: 5.973 on 3 and 4866 DF,  p-value: 0.0004642
```

What is the easiest sector to get an interview in?

```
sector <- ResumeNames %>%
  group_by(industry) %>%
  summarise(avgcb = mean(call, na.rm = TRUE)) %>%
  arrange(desc(avgcb))

print(sector)
```

Before running a model, we can average the number of callbacks an applicant receives by industry to see which sector is the easiest to get an interview. According to the sector dataframe, transport/communication on average has the highest of callbacks.

```
## # A tibble: 7 x 2
##   industry                       avgcb
##   <fct>                          <dbl>
## 1 transport/communication        0.135
## 2 health/education/social services 0.103
## 3 business/personal services      0.0836
```

2

```
## 4 unknown                         0.0759
## 5 finance/insurance/real estate   0.0725
## 6 trade                           0.0691
## 7 manufacturing                   0.0545
```

```
ind <- lm(call ~ gender * ethnicity + industry + experience + college, data = ResumeNames)
summary(ind)
```

According to the model, experience has the largest impact on receiving a callback. There is some marginal evidence of racial bias as well as no indication of gender bias. Looking at those with college degrees, it seems to disadvantage applicants in receiving a callback. All of these outcomes could be influenced by industry, but similar to our aggregated dataframe, transport/communication applicants have the highest chance of receiving a callback (8.1%).

```
##
## Call:
## lm(formula = call ~ gender * ethnicity + industry + experience +
##     college, data = ResumeNames)
##
## Residuals:
##      Min      1Q   Median      3Q     Max
## -0.19880 -0.09417 -0.07534 -0.05389  0.97707
##
## Coefficients:
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                        0.0427179  0.0197621   2.162  0.03070
## genderfemale                       0.0053596  0.0133999   0.400  0.68920
## ethnicityafam                     -0.0290938  0.0161807  -1.798  0.07223
## industrytransport/communication    0.0811072  0.0260502   3.113  0.00186
## industryfinance/insurance/real estate 0.0164107 0.0190965  0.859  0.39019
## industrytrade                      0.0173360  0.0159393   1.088  0.27682
## industrybusiness/personal services 0.0291380  0.0154901   1.881  0.06002
## industryhealth/education/social services 0.0458818 0.0168891 2.717 0.00662
## industryunknown                    0.0219318  0.0165448   1.326  0.18503
## experience                         0.0031706  0.0007726   4.104 4.13e-05
## collegeyes                        -0.0018258  0.0088669  -0.206  0.83687
## genderfemale:ethnicityafam        -0.0037437  0.0184537  -0.203  0.83924
##
## (Intercept)                       *
## genderfemale
## ethnicityafam                     .
## industrytransport/communication   **
## industryfinance/insurance/real estate
## industrytrade
## industrybusiness/personal services .
## industryhealth/education/social services **
## industryunknown
## experience                        ***
## collegeyes
## genderfemale:ethnicityafam
## ---
```

3

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2709 on 4858 degrees of freedom
## Multiple R-squared:  0.01058,    Adjusted R-squared:  0.008335
## F-statistic:  4.72 on 11 and 4858 DF,  p-value: 3.093e-07
```