Interview Case: Solution

Dr. Avery Haviv

Load and Explore the Data

1. A 'factor' variable is a categorical variable. It means that a variable can take on several nonnumeric values
2. Minimum was stored as a factor variable because it had several categories, including 'none' and 'some'
3. The treatment the authors are interested in is `ethnicity` and `quality`, which is why those variables were randomized. The outcome variable is `call`, which indicates whether there was a call back for the position

Benefits of Experimental Variation

1. we can test another variable we can add it as an independent variable in the regression. For example, if we wanted to test how military experience affected getting a callback, we would run

```
summary(lm(call~ethnicity+experience+city+honors+school+
           college+volunteer+quality+military,data=resumeData))
```

```
##
## Call:
## lm(formula = call ~ ethnicity + experience + city + honors +
##     school + college + volunteer + quality + military, data = resumeData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.23827 -0.09430 -0.07446 -0.04885  0.98188
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.1070895  0.0198878   5.385 7.60e-08 ***
## ethnicitycauc  0.0313477  0.0077462   4.047 5.27e-05 ***
## experience     0.0015543  0.0008446   1.840 0.065780 .
## citychicago   -0.0314400  0.0083072  -3.785 0.000156 ***
## honorsTRUE     0.0778109  0.0176812   4.401 1.10e-05 ***
## schoolTRUE    -0.0130640  0.0084039  -1.555 0.120128
## collegeTRUE   -0.0112240  0.0087981  -1.276 0.202112
## volunteerTRUE -0.0183953  0.0129344  -1.422 0.155032
## qualitylow    -0.0322845  0.0132173  -2.443 0.014617 *
## militaryTRUE  -0.0214785  0.0145423  -1.477 0.139748
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2702 on 4860 degrees of freedom
## Multiple R-squared:  0.01547,    Adjusted R-squared:  0.01365
## F-statistic: 8.485 on 9 and 4860 DF,  p-value: 1.086e-12
```

2. We cannot tell. However, based on the coefficient and the standard error, it does not seem to have a large impact. Even though it is statistically insignificant, the coefficient might not be zero.

3. The chance of getting a call back increases by 2×0.0015543, or 0.0031086

Non-experimental variation

1. Relative to not having an honors degree, an honors degree increases the chance of being called back by 7.07%

2. While initially counterintuitive, notice that we are also controlling for `experience`. Therefore, you need to think about what the effect of a change in `jobs` has holding `experience` fixed. In that case, the applicant has more jobs, but the same experience, meaning they were at each job for less time. This is intuitively not desirable to employers.

Randomization Check

1. We can run the analysis as follows:

```
summary(lm(volunteer~ethnicity,data=resumeData))
```

```
##
## Call:
## lm(formula = volunteer ~ ethnicity, data = resumeData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4144 -0.4144 -0.4086  0.5856  0.5914
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.414374   0.009974  41.543   <2e-16 ***
## ethnicitycauc -0.005749   0.014106  -0.408    0.684
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4922 on 4868 degrees of freedom
## Multiple R-squared:  3.413e-05,  Adjusted R-squared:  -0.0001713
## F-statistic: 0.1661 on 1 and 4868 DF,  p-value: 0.6836
```

   There is no significant difference. This is expected because `ethnicity` was randomized, and so it should not be correlated with any other independent variables

2. We can run the analysis as follows:

```
summary(lm(volunteer~school,data=resumeData))
```

```
##
## Call:
## lm(formula = volunteer ~ school, data = resumeData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5303 -0.5303 -0.2606  0.4697  0.7394
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)  0.26061    0.01023   25.48   <2e-16 ***
## schoolTRUE   0.26967    0.01367   19.73   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4736 on 4868 degrees of freedom
## Multiple R-squared:  0.07401,    Adjusted R-squared:  0.07382
## F-statistic: 389.1 on 1 and 4868 DF,  p-value: < 2.2e-16
```

There is a significant difference in the amount of volunteering applicants who completed high school complete. This is not surprising because these variables were not randomized, and therefore they might be correlated.