Equal Opportunity between Sexes in the Deer Valley Utility Company: Data Exercise 3

Yehzee Ryoo

3/30/2022

Executive Summary

Dear the Director of the Office of Equal Opportunity,

In response to the complaint claimed by some employees in the Deer Valley Utility Company, I investigate whether there has been any discrimination against women. I examine if the company is not rewarding the job performances of women to the same degree as the performances of men. First, I look at descriptive statistics of the data of past trainees' salaries and job performances. According to the statistical table below, the mean salary of 60 employees in the data is $66,660. To compare if there is any difference between men and women, I subset the data by sex and look at the same summary statistics. Here, the mean value of the salary is lower in the subset of women, which is $62,429, compared to the mean salary of the men, $70,363. It seems true that the average amount of money earned by women has been less than that of men according to the data. The plot 'Salary by Sex' also shows that the salary of women is mostly distributed under the blue line which shows the average amount of salary of all employees.

However, the fact that women have received a lower salary on average compared to men does not confirm that there has been discrimination in the Deer Valley Utility Company. It is not a sign of sex discrimination if the higher salary of men simply reflects their better job performance. Thus, I test the hypothesis that there is no difference in how men and women are rewarded following their job performances. I compute a statistical model - linear regression - with various factors that might have affected the salary of the employees. I estimate how much each factor, sex, rating of the employees based on their job performance, number of credits earned in college or company courses, affects the salary they are rewarded. To estimate whether the degree to which the job performance affects the reward is conditional on employees' sexes, I include the interaction term of Sex and Rating into the regression model. A statistical explanation of this regression model is specified in the following pages.

According to my statistical result, I reject the null hypothesis that there is no difference between men and women in how their performances are rewarded. Men get rewarded better for the same amount of increase in their work performance compared to women. Yet, I would like to investigate some other factors more that might have also affected the impact of rating on the reward. For example, the past career of the employee might have affected the way he or she is rewarded. Moreover, which task the employee is currently in charge of can also matter. Since the data set I examine here only has one control variable, credits, other than sex and rating, it can be a hasty conclusion that the sex is the main reason why men and women are rewarded differently. Including these variables in a further investigation would help me better examine the existence of sex discrimination in the company.

Descriptive Statistics

```
data <- read.csv("DataExercise3.csv")
stargazer(data, type= 'text', title = "Summary Statistics")
```

```
##
```

```
## Summary Statistics
## ==========================================================
## Statistic N   Mean  St. Dev.  Min   Pctl(25) Pctl(75)  Max
## ----------------------------------------------------------
## Rating    60 72.033  16.442    44     58.8     86.2    100
## Credits   60 10.650  5.605     1       6        15      20
## Sex       60 0.533   0.503     0       0        1       1
## Salary    60 66.660  6.549   56.500  62.475   71.000  82.100
## p_Salary  60 66.660  6.489   56.500  62.375   71.000  82.100
## ----------------------------------------------------------
```

```
data$Sex <- as.factor(data$Sex)

d_women <- data %>% filter(Sex==0)
d_men <- data %>% filter(Sex==1)
stargazer(d_women, d_men, type= 'text',
          title = c("Summary Statistics for Women", "Summary Statistics for Men"))
```

```
##
## Summary Statistics for Women
## ==========================================================
## Statistic N   Mean  St. Dev.  Min   Pctl(25) Pctl(75)  Max
## ----------------------------------------------------------
## Rating    28 70.250  17.118    44      55      83.8    100
## Credits   28 10.607  5.473     1      5.5       15      19
## Salary    28 62.429  4.222   56.500  57.950   65.900  69.000
## p_Salary  28 62.429  4.157   56.500  58.277   66.025  69.000
## ----------------------------------------------------------
##
## Summary Statistics for Men
## ==========================================================
## Statistic N   Mean  St. Dev.  Min   Pctl(25) Pctl(75)  Max
## ----------------------------------------------------------
## Rating    32 73.594  15.935    49     59.8      88      99
## Credits   32 10.688  5.806     1       6       15.2     20
## Salary    32 70.363  5.982   59.900  64.975   75.200  82.100
## p_Salary  32 70.363  5.896   61.619  65.075   75.050  82.100
## ----------------------------------------------------------
```

```
plot(data$Salary,
     col = data$Sex,
     main="Salary by Sex",
        ylab="Salary")
abline(h=mean(data$Salary), col="blue") #Mean of the Salary


legend("topright",
       legend=c("Women", "Men"),
       col=data$Sex, lty=1:2, cex=0.9,
       box.lty=0)
```

Regression Model

To test whether there is a differential impact of performance on reward, I test the following hypothesis.

H0: There is no difference in how the performance of men and the performance of women are rewarded. In other words, the relationship between salary and rating does not differ by sex.

H1: Job performances of women are rewarded differently by those of men. In other words, the relationship between salary and rating differs by sex.

Simply including Sex and Rating variable to explain Salary cannot fully answer the relationship of our interest. To examine whether the impact of Rating on Salary is conditional on Sex, I include an interaction term of Sex and Rating. Moreover, I add Credits as a control variable since higher education can be related to performance and salary. I present both models of regression with and without the interaction term here. Model 2 in the table below shows the relationship of our interest.

$$Salary = \alpha + \beta_1 * Rating + \beta_2 * Sex + \beta_3 * Rating * Sex + \beta_4 * Credits + \epsilon$$

The table below shows the statistical result of the regression models. According to the regression result, the p-value of the interaction effect (Rating * Sex) is statistically significant. Looking closely, Men and women respectively get the following interaction effect on their salary:

For men,

$$Salary = 46.597 + (0.227 + 0.135) * Rating + (-2.772) * 1 + (-0.010) * Credits + \epsilon$$

For women,

$$Salary = 46.597 + 0.227 * Rating + (-0.010) * Credits + \epsilon$$

It means that the contribution of rating to salary for women is \$227 while for men it is \$362. Therefore, the regression analysis here including the interaction term of sex and performance shows that the increase in performance can lead to a greater increase in salary for men compared to women.

Thus, I reject the null hypothesis (H0) that there is no difference in how the performance of men and the performance of women are rewarded. The impact of rating on salary seems to be conditional on the employee's sex.

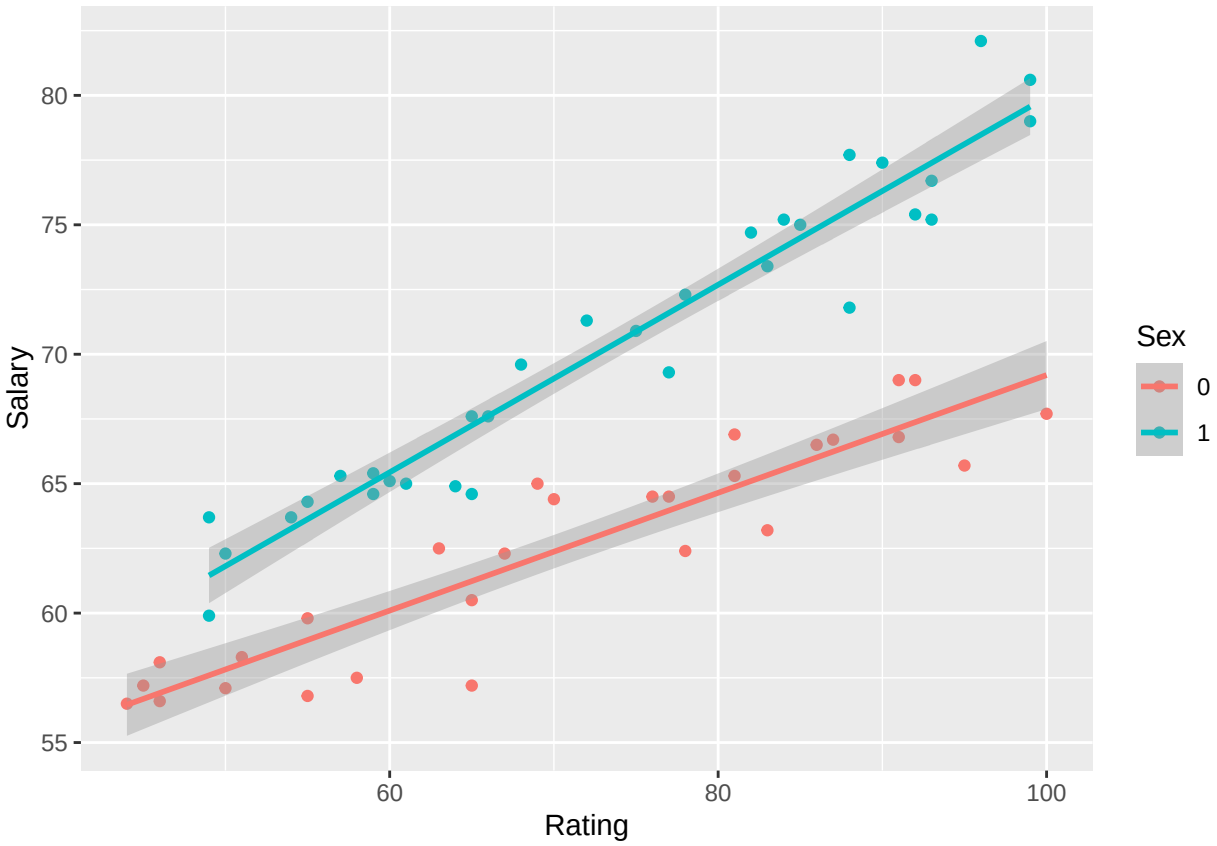
```
reg0 <- lm (Salary ~ Sex + Rating + Credits, data = data)
reg1 <- lm (Salary ~ Sex + Rating + Credits + Sex*Rating, data= data)
stargazer(reg0, reg1, type = 'text')
```

```
##
## ===============================================================
##                                 Dependent variable:
##                     ------------------------------------------
##                                        Salary
##                              (1)                    (2)
## ---------------------------------------------------------------
## Sex1                      6.948***               -2.772
##                            (0.516)                (1.930)
##
## Rating                    0.295***               0.227***
##                            (0.016)                (0.019)
##
## Credits                    0.001                  -0.010
##                            (0.046)                (0.039)
##
## Sex1:Rating                                       0.135***
##                                                    (0.026)
##
## Constant                  41.710***              46.597***
##                            (1.336)                (1.456)
##
## ---------------------------------------------------------------
## Observations                 60                     60
## R2                           0.913                  0.941
## Adjusted R2                  0.908                  0.937
## Residual Std. Error     1.983 (df = 56)        1.642 (df = 55)
## F Statistic         195.937*** (df = 3; 56) 220.992*** (df = 4; 55)
## ===============================================================
## Note:                                *p<0.1; **p<0.05; ***p<0.01
```

Plot of Relationship between Salay, Rating, and Sex

```
ggplot(data = data, aes(x=Rating, y=Salary, col = Sex)) +
  geom_point() +
  geom_smooth(method = 'lm')
```

```
## `geom_smooth()` using formula 'y ~ x'
```



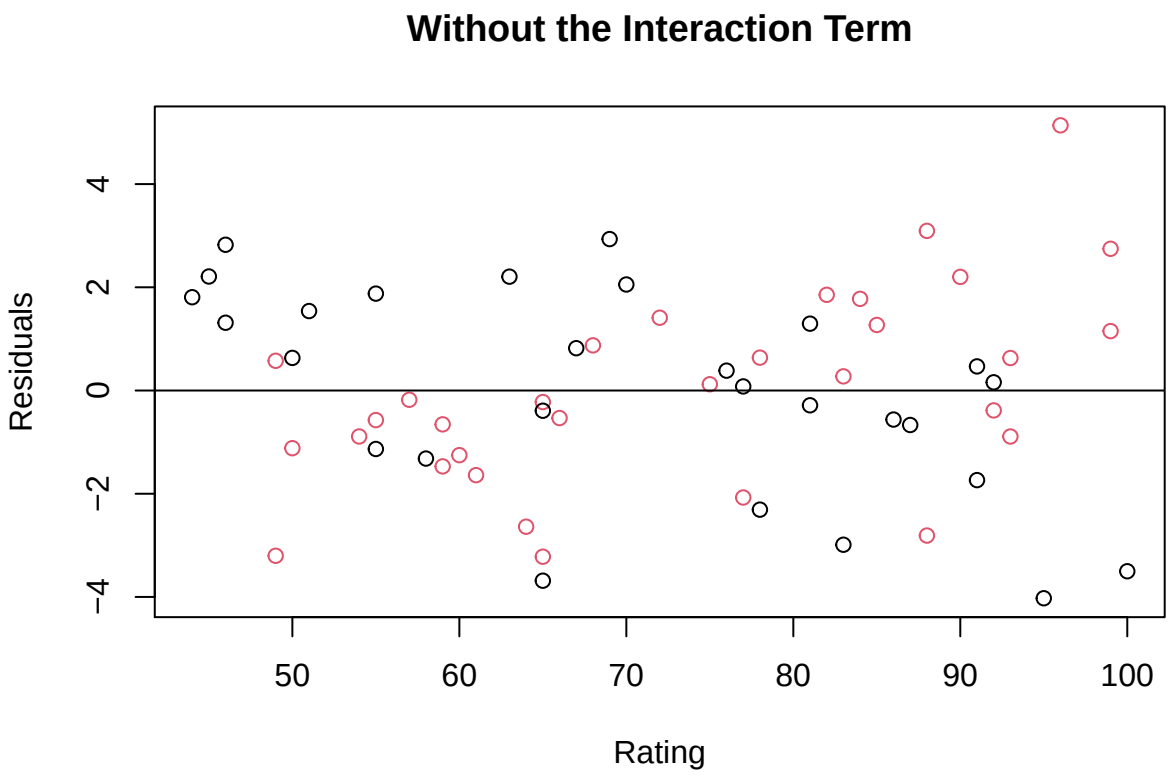
Here, we can see that the slope of the regression line of the rating on salary is different between men and women. Men, the blue line, gets bigger increase in their salary as they show better performance on their work.

Residual plots

Below is the residual plots of the regression with and without the interaction term of sex and rating. The color of the dots show the sex of the observation employee, red dots for women. As you can see in the first plot of residuals against rating, residuals of women are not evenly distributed.

```
res_0 <- resid(reg0) #without the interaction term
res <- resid(reg1) #with the interaction term

res01 <- plot(data$Rating, res_0, col=data$Sex,
              ylab="Residuals", xlab="Rating", main = "Without the Interaction Term") +abline(0,0)
```




```
res02 <-  plot(data$Credits, res_0, col=data$Sex,
               ylab="Residuals", xlab="Creidts", main ="Without the Interaction Term") + abline(0,0)
```

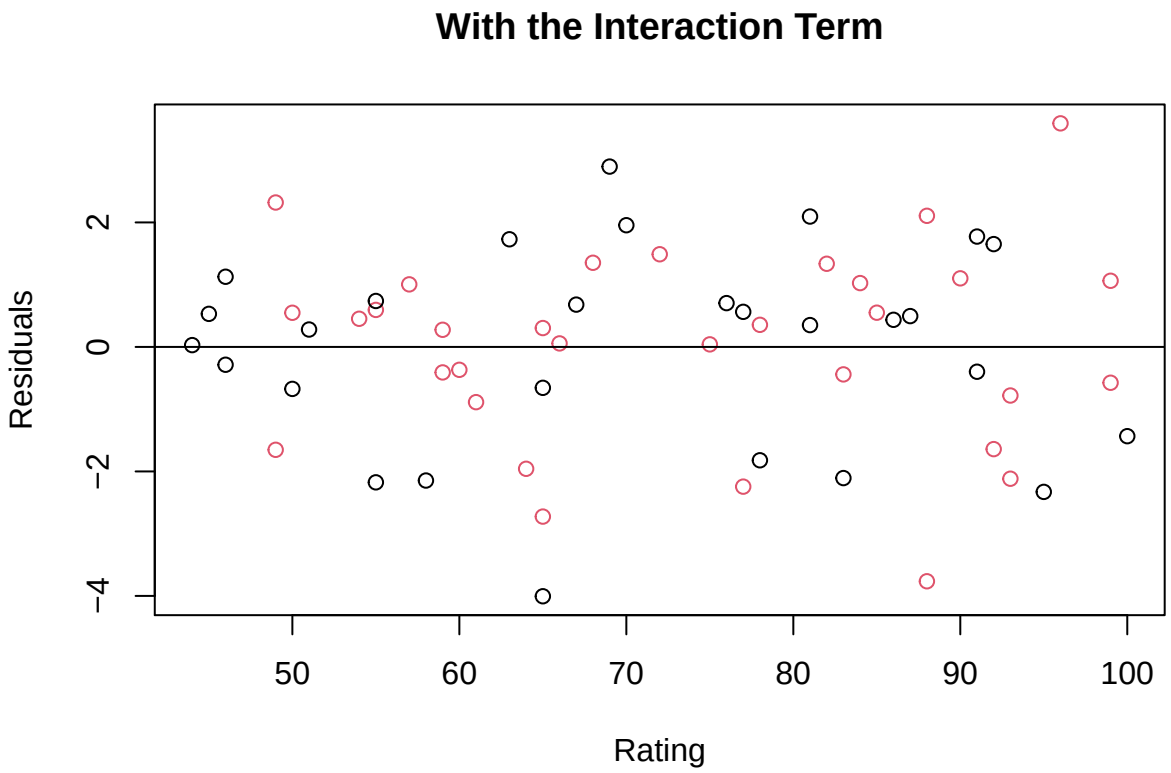

```
res01
```

```
## integer(0)
```

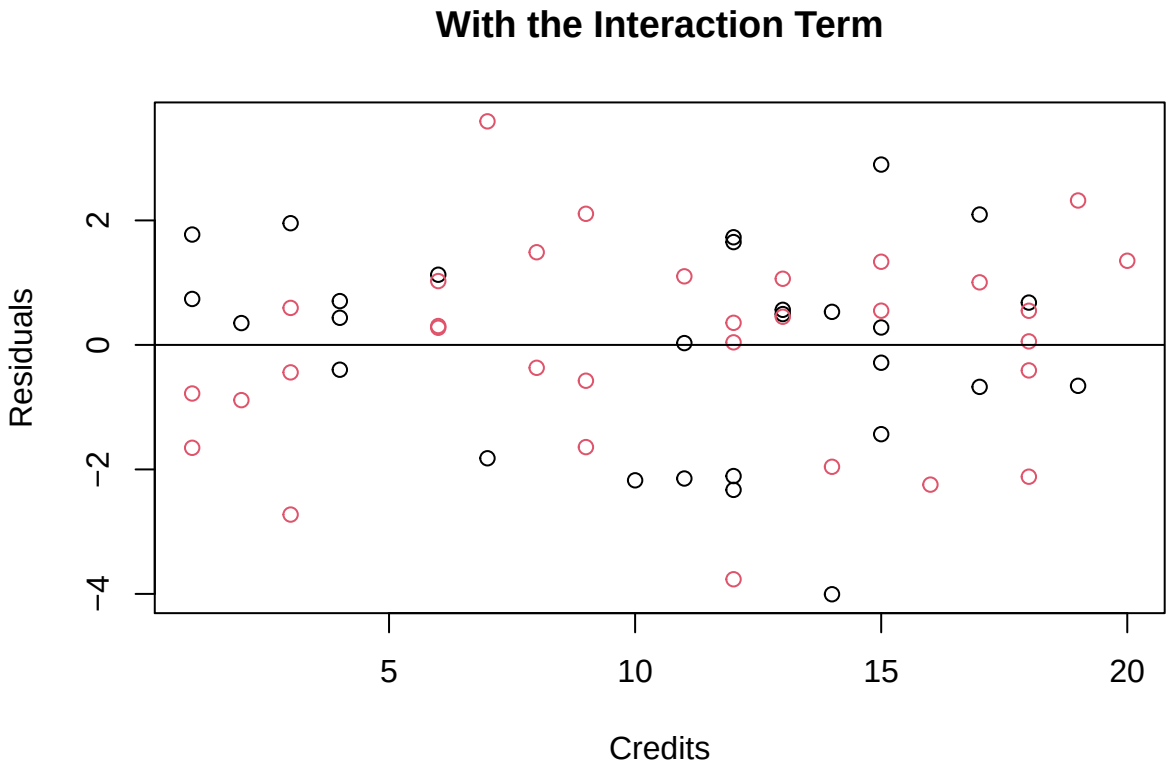
```
res02
```

```
## integer(0)
```

```
res1<- plot(data$Rating, res, col=data$Sex,
            ylab="Residuals", xlab="Rating", main ="With the Interaction Term") + abline(0,0)
```




```
res2 <- plot(data$Credits, res, col=data$Sex,
             ylab="Residuals", xlab="Credits", main ="With the Interaction Term") + abline(0,0)
```




```
res1
```

```
## integer(0)
```

```
res2
```

```
## integer(0)
```