# MATH-650 Assignment 3

*Saket Choudhary (USCID: 2170058637) (skchoudh@usc.edu)*

*09/15/2015*

## Problem 24: Sex Discrimination

### Part (a)

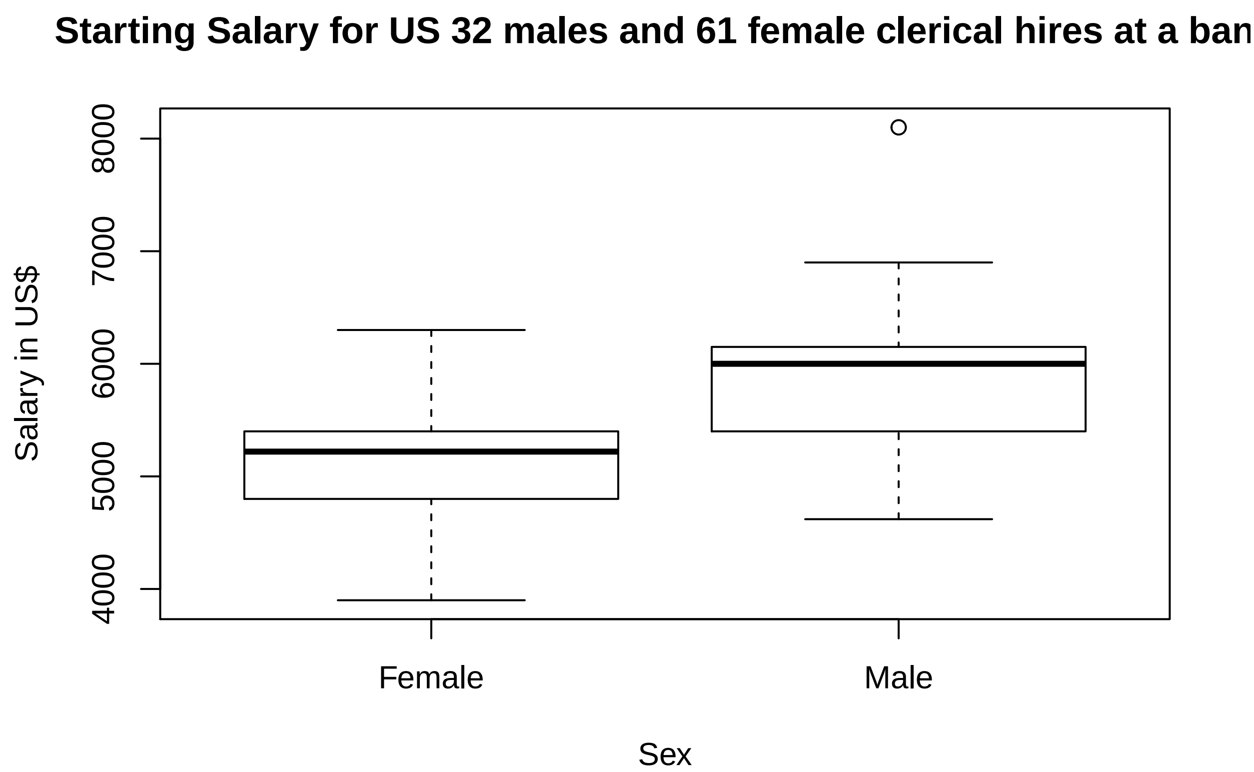
```
salary.data <- read.csv('data_24.csv')
head(salary.data)
```

```
##   Salary    Sex
## 1   3900 Female
## 2   4020 Female
## 3   4290 Female
## 4   4380 Female
## 5   4380 Female
## 6   4380 Female
```

```
salary.female <- salary.data[salary.data$Sex=="Female",]$Salary
salary.male <- salary.data[salary.data$Sex=="Male",]$Salary
```

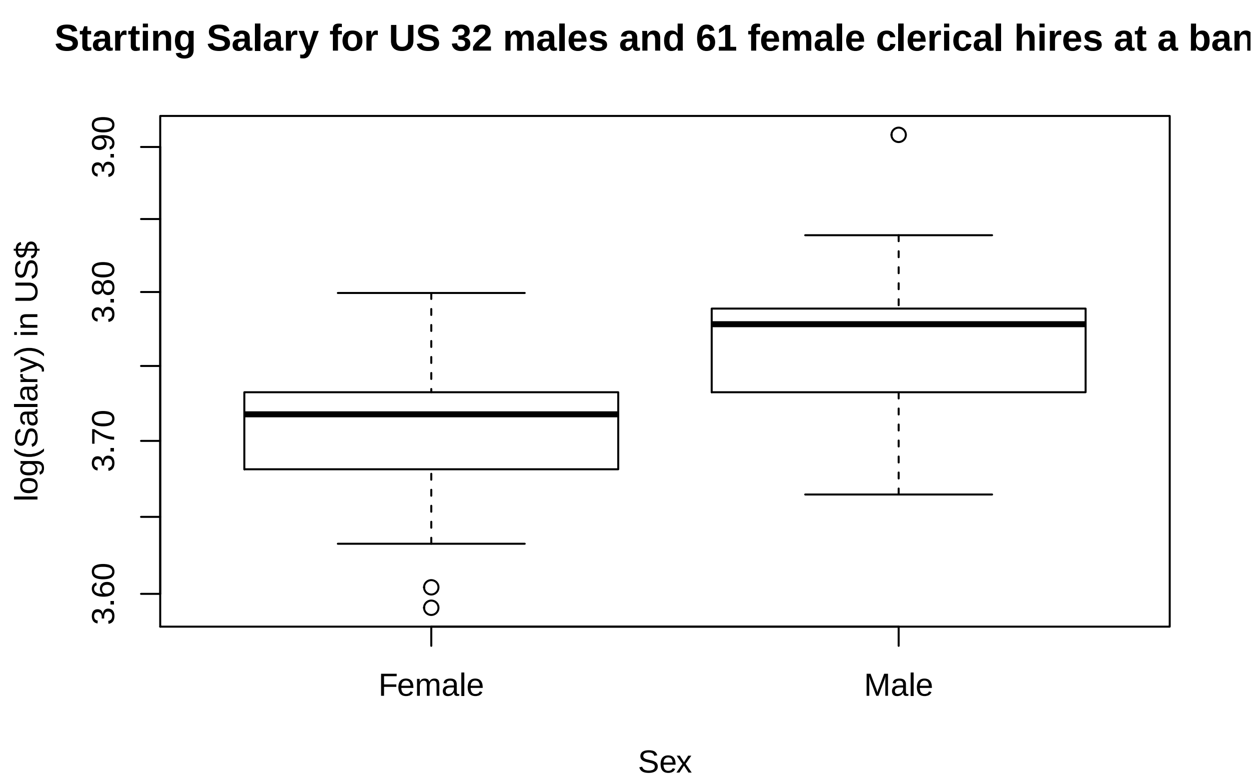
#### Boxplot of Salary *without* log adjustment

```
boxplot(Salary~Sex,data=salary.data, main="Starting Salary for US 32 males and 61 female clerical hires
    xlab="Sex", ylab="Salary in US$")
```




**Boxplot of log transformed salary**

```
boxplot(log10(Salary)~Sex,data=salary.data, main="Starting Salary for US 32 males and 61 female clerical
    xlab="Sex", ylab="log(Salary) in US$")
```



**Part (b)**

```
tt<-t.test(log10(salary.male), log10(salary.female), alternative = "greater")
dm <- tt$estimate[1]-tt$estimate[2]
dm <- unname(dm)
CI1<-tt$conf.int[1]
CI2<-2*dm-CI1
```

Difference of log means of Male-Female salaries: 0.0638168

Null hypothesis: mean of log transformed salary of males is the same as that of feamles

Alternative hypothesis: Mean of log transformed salary of males is greater than that of females

t-statistic: 6.0538724

p-value: $5.0434814 \times 10^{-8}$

CI of ratio of population medians: [0.0462048, 0.0814288]

**Part (c)**

```
estimate <- exp(dm)
lower.ci <- exp(CI1)
upper.ci <- exp(CI2)
estimate
```

```
## [1] 1.065897
```

```
lower.ci
```

```
## [1] 1.047289
```

```
upper.ci
```

```
## [1] 1.084836
```

95% Confidence interval for ratio of population medians: [1.0472889, 1.084836] At a significance level of 0.05, we can reject the null hypothesis that the mean salary of males is the same as that of females. The associated p-value is $5.0434814 \times 10^{-8}$

# Problem 25

```
data <- read.csv('data_25.csv')
vietnam <- data[data$Veteran=="Vietnam",]$Dioxin
other <- data[data$Veteran=="Other",]$Dioxin
vietnam.without1 <- vietnam[c(1:645)]
vietnam.without2 <- vietnam[c(1:644)]

tt.with <- t.test(vietnam, other, alternative="greater")
tt.with
```

```
##
##  Welch Two Sample t-test
##
## data:  vietnam and other
## t = 0.29122, df = 136.96, p-value = 0.3857
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -0.3491265        Inf
## sample estimates:
## mean of x mean of y
##  4.260062  4.185567
```

Doing a independent two sample t-test for equal population means for all observations, the associated p-value: 0.3856612(matches the value of 0.40 as in Display 3.7)

```r
tt.without1 <- t.test(vietnam.without1, other, alternative="greater")
tt.without1
```

```
##
##  Welch Two Sample t-test
##
## data:  vietnam.without1 and other
## t = 0.045709, df = 121.27, p-value = 0.4818
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -0.3996048        Inf
## sample estimates:
## mean of x mean of y
##  4.196899  4.185567
```

Doing a independent two sample t-test for equal population means by exluding the last outlier points(646), the associated p-value: 0.4818089 (matches the value of 0.48 as in Display 3.7)

```r
tt.without2 <- t.test(vietnam.without2, other, alternative="greater")
tt.without2
```

```
##
##  Welch Two Sample t-test
##
## data:  vietnam.without2 and other
## t = -0.0853, df = 117.33, p-value = 0.5339
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -0.4285707        Inf
## sample estimates:
## mean of x mean of y
##  4.164596  4.185567
```

Doing a independent two sample t-test for equal population means by exluding the last two outlier points(645, 646), the associated p-value: 0.5339159 (matches the value of 0.54 as in Display 3.7)