

Salary Exploratory Data Analysis

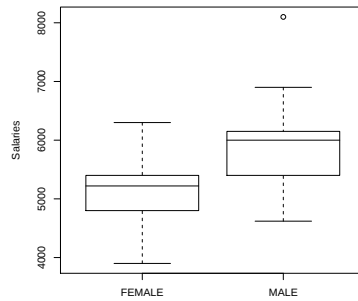


Figure 1: Salary Boxplots

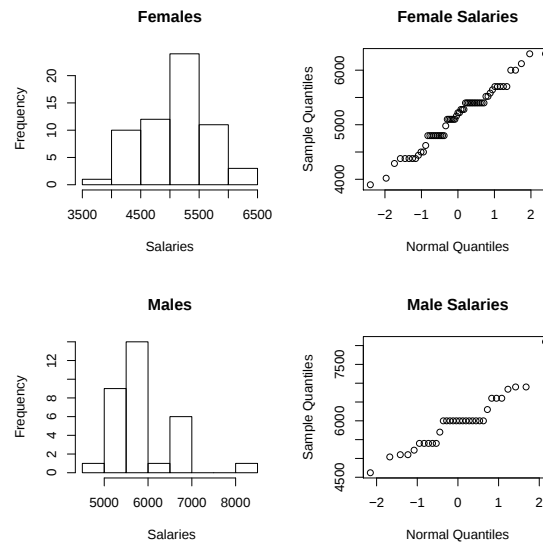


Figure 2: Normal Quantile Plots and Histograms

Validity of Normality assumptions in t-tests? Sensitivity?

Testing in R/S

```
> S.F <- SALARY[SEX=="FEMALE"]
> S.M <- SALARY[SEX=="MALE"]
> t.test(S.M, S.F, var.equal=T)
```

Two Sample t-test

```
data: S.M and S.F
t = 6.2926, df = 91, p-value = 1.076e-08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 559.7985 1076.2465
sample estimates:
mean of x mean of y
 5956.875  5138.852

> t.test(S.M, S.F, var.equal=F)
```

Welch Two Sample t-test

```
data: S.M and S.F
t = 5.83, df = 51.329, p-value = 3.71e-07
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 536.3758 1099.6693
sample estimates:
mean of x mean of y
 5956.875  5138.852

> wilcox.test(S.M, S.F)
```

Wilcoxon rank sum test with continuity correction

```
data: S.M and S.F
W = 1604, p-value = 3.272e-07
alternative hypothesis: true mu is not equal to 0
```

Conclusions? Interpretation?