Code Appendix

Elliot Smith

2/13/2018

```r
options(scipen = 999)
```

Problem 2

Part b

```r
# Problem 2

## Part b

values <- c(1,3,5,9,5,5,5,6,6,3,3,3,0,6,14,10,18)
labels <- c("A","A","A","B","B","B","B","C","C","C","D","D","D","D","E","E","E")

data <- data.frame(labels, values)

aov_out <- aov(values ~ labels, data = data)
summary(aov_out)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## labels       4    258   64.50   10.18 0.000782 ***
## Residuals   12     76    6.33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
f_stat <- as.vector(summary.lm(aov_out)$fstatistic["value"])



f_stats <- numeric()
n <- 5000

for (i in 1:n) {

    labels_temp <- sample(labels)
    data_temp <- data.frame(labels_temp, values)
    aov_out_temp <- aov(values ~ labels_temp, data = data_temp)
    f_stats[i] <- as.vector(summary.lm(aov_out_temp)$fstatistic["value"])

}

count_greater <- sum(f_stats >= f_stat)

count_greater/n
```

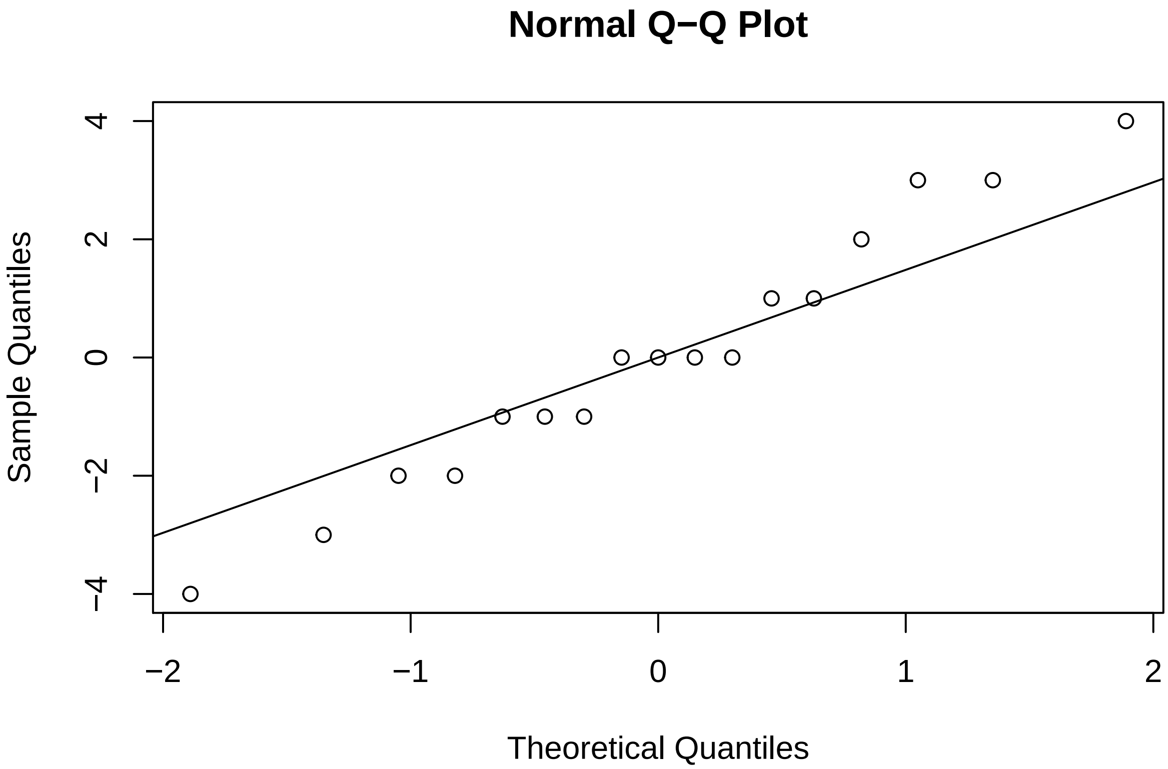
```
## [1] 0.0012
```

Part c

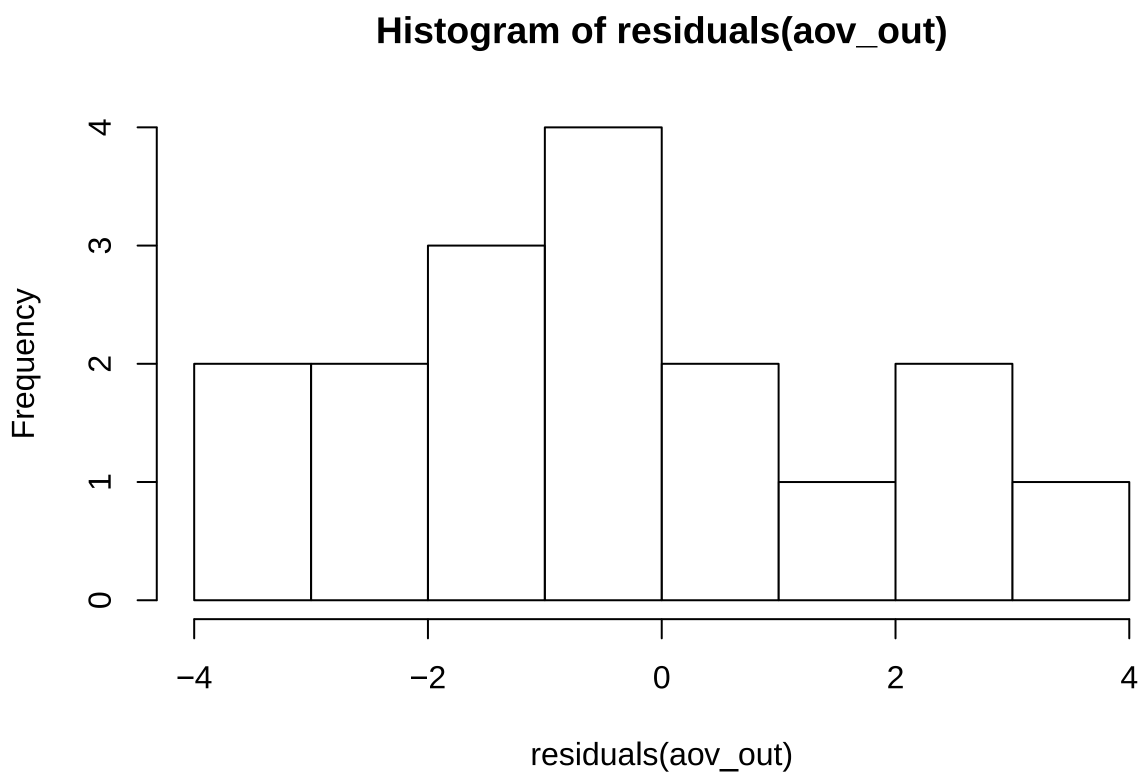
```r
## Part c

### Check for error normality

qqnorm(residuals(aov_out))
qqline(residuals(aov_out))
```

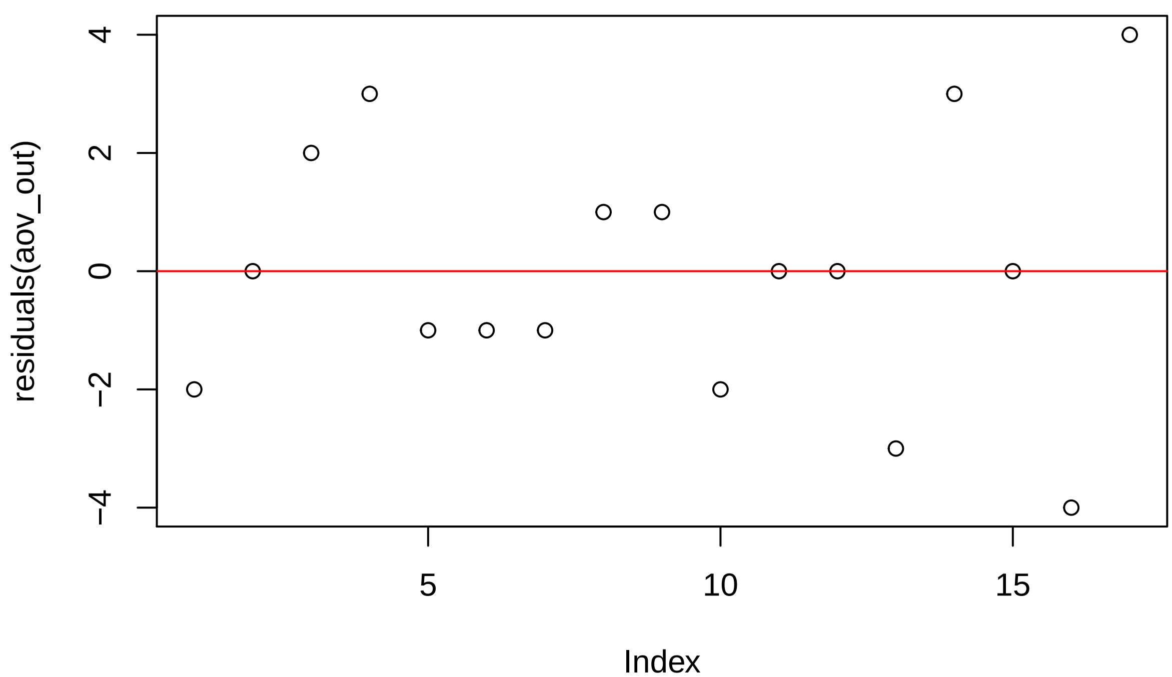



```r
hist(residuals(aov_out))
```




```
### Check residuals for equal variance

plot(residuals(aov_out))
abline(h = 0, col = "red")
```



Part e

```r
## Part e

### Compare B and C

t.test(x = data[4:7,2], y = data[8:10,2])
```

```
##
##  Welch Two Sample t-test
##
## data:  data[4:7, 2] and data[8:10, 2]
## t = 0.70711, df = 4.8, p-value = 0.5123
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.681387  4.681387
## sample estimates:
## mean of x mean of y
##         6         5
```

```r
### Compare B and E

t.test(x = data[4:7,2], y = data[14:17,2])
```

```
##
##  Welch Two Sample t-test
##
## data:  data[4:7, 2] and data[14:17, 2]
## t = -2.1669, df = 3.8802, p-value = 0.09825
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.78215   1.78215
## sample estimates:
## mean of x mean of y
##         6        12
```

STAT616 - Problem 1

```r
# STAT616 - Problem 1

values <- c(1,2,3,4,60000001,60000000)
labels <- c("A","A","B","B","C","C")

data <- data.frame(labels, values)

aov_out <- aov(values ~ labels, data = data)
summary(aov_out)
```

```
##             Df            Sum Sq          Mean Sq          F value
## labels       2 4799999680000010 2399999840000005 4799999696203265
## Residuals    3                1                0
##                         Pr(>F)
## labels      <0.0000000000000002 ***
```

```
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
f_stat <- as.vector(summary.lm(aov_out)$fstatistic["value"])



f_stats <- numeric()
n <- 5000

for (i in 1:n) {

    labels_temp <- sample(labels)
    data_temp <- data.frame(labels_temp, values)
    aov_out_temp <- aov(values ~ labels_temp, data = data_temp)
    f_stats[i] <- as.vector(summary.lm(aov_out_temp)$fstatistic["value"])

}

count_greater <- sum(f_stats >= f_stat)

count_greater/n
```

```
## [1] 0.0214
```