STAT616 - HW #3

Elliot Smith

3/2/2018

Problem 1

Part a

- As we can see from the ANOVA table (see code appendix, Problem 1, Part a), we have received an F-value of 7.8415, which is certainly greater than 1, this implies that there is a significant difference between at least one of the means of the different tire brands.
- I used the following assumptions in my analysis:
  - Errors are iid Normal(0, σ^2)
  - Variances are equal for all tire brands
- I have confirmed these assumptions using a QQ Plot for Error Normality and a Barlett Test for Variance Homogeneity (see code appendix, Problem 1, Part a):
  - My QQ Plot confirmed that my errors are approximately normal
  - My Bartlett Test got a p-value of 0.1362, so I will not reject my Ho that the variances are equal

Part b

- Standard Confidence Intervals (using t method):
  - A: 185.5131 - 190.9469
  - B: 188.8118 - 201.6482
  - C: 183.7718 - 191.0682
  - D: 187.222 - 195.198
  - E: 196.3295 - 204.3705
- Bonferroni Confidence Intervals (using t method):
  - A: 184.3269 - 192.1331
  - B: 186.0095 - 204.4505
  - C: 182.1789 - 192.6611
  - D: 185.4808 - 196.9392
  - E: 194.5741 - 206.1259
- The Bonferroni confidence intervals are much more conservative than our traditional t-statistic confidence intervals (ie its a wider range).
- I prefer the traditional confidence interval due to the fact that with Bonferroni we are less likely to reject Ho which means we would be more likely to commit a Type II error.

Part c

Tukey Output

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = value ~ type, data = data_1)
```

```
##
## $type
##       diff         lwr        upr     p adj
## B-A  7.00  -0.7002118 14.7002118 0.0907506
## C-A -0.81  -8.5102118  6.8902118 0.9981948
## D-A  2.98  -4.7202118 10.6802118 0.8057780
## E-A 12.12   4.4197882 19.8202118 0.0004798
## C-B -7.81 -15.5102118 -0.1097882 0.0453343
## D-B -4.02 -11.7202118  3.6802118 0.5783176
## E-B  5.12  -2.5802118 12.8202118 0.3377270
## D-C  3.79  -3.9102118 11.4902118 0.6317521
## E-C 12.93   5.2297882 20.6302118 0.0001837
## E-D  9.14   1.4397882 16.8402118 0.0127107
```

- The largest differences are:
  - E-C: 12.93
  - E-A: 12.12
  - E-D: 9.14
- This shows that E most certainly has a different mean than these other tire brands, this is confirmed by the p-values of all the afforementioned comparisons which are significant at the 95% level.
- C-B is also significant at the 95% level and we can reject that these two have the same mean as well.

Bonferroni Output

```
##
## 99 %-confidence intervals
##  Method:  Difference of means assuming Normal distribution, allowing unequal variances
##
##
##     estimate   lower  upper
## B-A     7.00  -2.393 16.393
## C-A    -0.81  -6.654  5.034
## D-A     2.98  -3.257  9.217
## E-A    12.12   5.844 18.396
## C-B    -7.81 -17.497  1.877
## D-B    -4.02 -13.858  5.819
## E-B     5.12  -4.734 14.974
## D-C     3.79  -3.094 10.674
## E-C    12.93   6.014 19.846
## E-D     9.14   1.934 16.346
##
##
```

- We get very similar results for the Bonferroni as the Tukey method, but as we can see the bounds are certainly larger:
  - Bonferroni B-A: -2.4 - 16.4
  - Tukey B-A: -0.7 - 14.7
- This shows us that the Bonferroni method will produce much wider bounds than the Tukey method

Overall Interpetation

- Again, I prefer the tighter, Tukey confidence interval due to the fact that with Bonferroni we are less likely to reject Ho which means we would be more likely to commit a Type II error.

Problem 2

- See Hand-Written Answers section after Code Appendix

Problem 3

- For each of the four columns (blocks) of a 5x4 matrix, starting with the first, randomly permute t_1 to t_5 (representing treatments) without replacement and assign them to the cells in the order specified.
- We now have a 5x4 table with all 5 treatments randomly assigned to each cell within each column.
- See Hand-Written Answers section after Code Appendix for the Sample Table

Problem 4

- No. Due to the fact that observational studies are not randomized due to self-selection, a Complete Randomized Block design cannot be observational. If the design of an experiment is not randomized, then the experiement cannot be a CRB.

Problem 5

Part a

ANOVA Table

```
## Analysis of Variance Table
##
## Response: value
##           Df Sum Sq Mean Sq F value Pr(>F)
## device     2    0.0   0.007   1e-04 0.9999
## Residuals 42 3699.4  88.082
```

- From the ANOVA table we can see that the variation attributed to within the groups is significantly lower than the variation between the groups.

F-Test

- The F-Test is performed by the ANOVA Table. We can see that the p-value associated is extremely high, almost 1, meaning that we must fail to reject Ho that the treatment devices have the same mean.

Estimate of Mean Blood Pressure

- This estimate, from the Regresssion Summary (see Code Appendix, Problem 5, Part a), is 128.08067.

Estimates of the two Variance Components

- $\hat{\sigma}_\epsilon = 88.082$
- $\hat{\sigma}^2_\alpha = \frac{0-88.082}{15} = -5.87$
- Since $\hat{\sigma}^2_\alpha$ is negative, $\hat{\sigma}_\alpha = 0$

95% Confidence Interval

- $128.08067 \pm 4.3027(0.014) = 128.0204, 128.1409$
- The confidence interval is quite tight in this situation mostly due to the small size of the standard error.

Part b

ANOVA Table

```
## Analysis of Variance Table
##
## Response: value
##           Df Sum Sq Mean Sq F value    Pr(>F)
## doctor     2 496.33 248.163   139.1 < 2.2e-16 ***
## Residuals 42  74.93   1.784
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- From the ANOVA table we can see that the variation attributed to between the groups is significantly lower than the variation within the groups.

F-Test

- The F-Test is performed by the ANOVA Table. We can see that the p-value associated is extremely low, almost 0, meaning that we must reject Ho that the treatment doctors have the same mean.

Estimate of Mean Blood Pressure

- This estimate, from the Regresssion Summary (see Code Appendix, Problem 5, Part a), is 125.9360.

Estimates of the two Variance Components

- $\hat{\sigma}_\epsilon = 1.784$
- $\hat{\sigma}^2_\alpha = \frac{248.163-1.784}{15} = 16.43$
- Since $\hat{\sigma}^2_\alpha$ is 16.43, $\hat{\sigma}_\alpha = 4.05$

95% Confidence Interval

- $125.9360 \pm 4.3027(0.014) = 125.8758, 125.9962$
- The confidence interval is quite tight in this situation mostly due to the small size of the standard error.

Part c

- I think there is an error in my calculation because the confidence interval should me larger for the doctors in Part b, perhaps I incorrectly calculated the standard error. Even though my calculation may be incorrect, I believe that the readings by the devices should be much more exact than the doctors, since so little of the variation in the devices comes from between the devices (it mostly comes from within); this is opposite for the doctors.

Problem 6

Part a

- Completely Randomized Block. Yes, because I would expect a different results on the training for recent graduates vs. those that have had several years elapse since they graduated. I think that time since graduation is a good blocking methodology because I would expect more experienced auditors to get better scores so the blocking here will removing the confounding from time since graduation.

Part b

- See Hand-Written Answers section after Code Appendix

Part c

- See the ANOVA Table below:

```
## Analysis of Variance Table
##
## Response: score
##           Df  Sum Sq Mean Sq  F value    Pr(>F)
## block      9  433.37   48.15   7.7157 0.0001316 ***
## method     2 1295.00  647.50 103.7537 1.315e-10 ***
## Residuals 18  112.33    6.24
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Part d

- Since the p-value for the blocks is 0.0001, we can reject the Ho that the blocks are all the same. So yes, blocking was needed.

Part e

- Since the p-value for methods, according the the ANOVA Table, is almost 0, we can reject our null hypothesis that the means are all the same for the different treatment methods.

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = score ~ method, data = data_6)
##
## $method
##     diff        lwr       upr     p adj
## 2-1  4.0 -0.9849383  8.984938 0.1341187
## 3-1 15.5 10.5150617 20.484938 0.0000001
## 3-2 11.5  6.5150617 16.484938 0.0000129
```

- According to the above Tukey pairwise comparison of the different methods, the following methods differ based on their associated p-values:
  - Method 3 and Method 1

    - Method 3 and Method 2
- I used the following assumptions in my analysis:
    - Errors are iid Normal(0, σ^2)
    - Variances are equal for all methods
- I have confirmed these assumptions using a QQ Plot for Error Normality and a Barlett Test for Variance Homogeneity (see code appendix, Problem 6, Part e):
    - My QQ Plot confirmed that my errors are approximately normal
    - My Bartlett Test got a p-value of 0.5041, so I will not reject my Ho that the variances are equal

Part f

- See the ANOVA Table below:

```
## Analysis of Variance Table
##
## Response: score
##           Df Sum Sq Mean Sq F value    Pr(>F)
## block      4  523.2  130.80  10.226 6.586e-05 ***
## method     2 1796.5  898.23  70.222 1.607e-10 ***
## Residuals 23  294.2   12.79
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- In the case of the new data, since the p-value for the blocks is extremely small, we can reject the Ho that the blocks are all the same. So yes, blocking was needed again.
- And again, since the p-value for methods, according the the ANOVA Table, is almost 0, we can reject our null hypothesis that the means are all the same for the different treatment methods according to our blocking protocol.

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = score ~ method, data = data_6f)
##
## $method
##     diff       lwr      upr     p adj
## 2-1  8.2  2.099013 14.30099 0.0068497
## 3-1 18.9 12.799013 25.00099 0.0000001
## 3-2 10.7  4.599013 16.80099 0.0004988
```

- According to the above Tukey pairwise comparison of the different methods, ALL of the methods are different at a significant level!
- I used the following assumptions in my analysis:
    - Errors are iid Normal(0, σ^2)
    - Variances are equal for all methods
- I have confirmed these assumptions using a QQ Plot for Error Normality and a Barlett Test for Variance Homogeneity (see code appendix, Problem 6, Part f):
    - My QQ Plot confirmed that my errors are approximately normal
    - My Bartlett Test got a p-value of 0.2923, so I will not reject my Ho that the variances are equal

Part g

- I prefer the second experiment because we rejected Ho in all cases including the differences in means for each of the methods. My biggest concern is making a Type II error, so I would rather be in the position to reject an Ho then falsely not reject it.

Problem 7

- By building a BIBD (See Hand-Written Answers section after Code Appendix) I will analyze the affect of the shape on the noice, blocking by the plate.

```
## Analysis of Variance Table
##
## Response: noise
##           Df  Sum Sq  Mean Sq F value  Pr(>F)
## plate      3 0.34737 0.115789  8.4548 0.02105 *
## shape      3 0.46506 0.155019 11.3194 0.01146 *
## Residuals  5 0.06847 0.013695
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- See the above ANOVA Table. Based on both the p-value for the plates (0.02) and the p-value for the shape (0.01) we can reject each null hypothesis: (1) that the mean values are the same across blocks (meaning blocking is a good idea) and (2) that the mean values are the same across shapes.

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = noise ~ plate + shape, data = data_7)
##
## $plate
##            diff           lwr       upr     p adj
## 2-1  0.33666667 -0.0159083249 0.6892417 0.0590411
## 3-1  0.45000000  0.0974250084 0.8025750 0.0194330
## 4-1  0.35333333  0.0007583418 0.7059083 0.0496095
## 3-2  0.11333333 -0.2392416582 0.4659083 0.6598457
## 4-2  0.01666667 -0.3359083249 0.3692417 0.9978627
## 4-3 -0.09666667 -0.4492416582 0.2559083 0.7507145
##
## $shape
##            diff         lwr         upr     p adj
## B-A -0.40444444 -0.75701944 -0.05186945 0.0297637
## C-A -0.13888889 -0.49146388  0.21368610 0.5231292
## D-A -0.44777778 -0.80035277 -0.09520279 0.0198291
## C-B  0.26555556 -0.08701944  0.61813055 0.1287196
## D-B -0.04333333 -0.39590832  0.30924166 0.9660591
## D-C -0.30888889 -0.66146388  0.04368610 0.0795257
```

- To figure out which are different, please refer to the above Tukey pairwise comparision tables.
- We may conclude the following:
  - Plate 1 is significantly different then the other plates; it is the only plate whose p-value shows significant difference when compared to the others
  - Comparing both shapes A and B, as well as shapes A and D yield different results and we may conclude that there difference is significant

STAT616 - Problems 1 and 2

- See Hand-Written Answers section after Code Appendix

Code Appendix

Problem 1

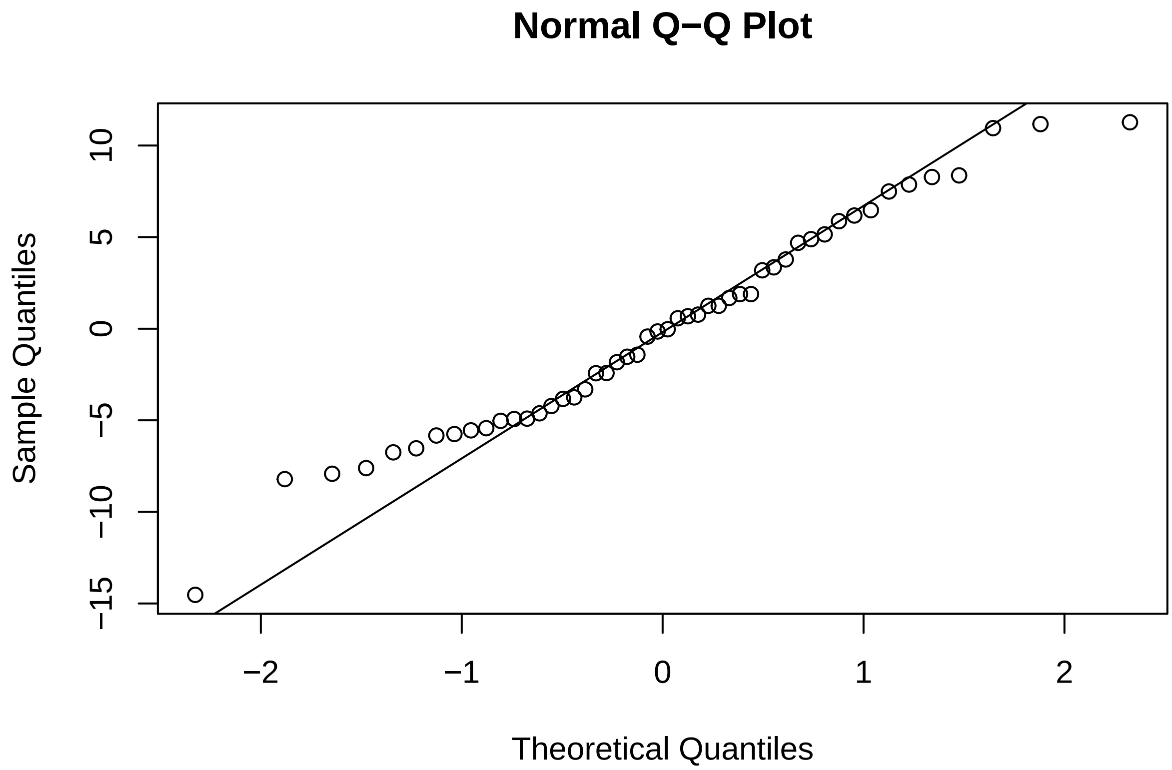
```
## Load in the data

data_1 <- read.csv("~/Documents/Rice_University/Spring_2018/STAT616/HW03/tire.csv")
data_1 <- data_1[,-1]
value <- c(data_1[,1], data_1[,2], data_1[,3], data_1[,4], data_1[,5])
type <- c(rep("A", 10), rep("B", 10), rep("C", 10), rep("D", 10), rep("E", 10))
data_1 <- data.frame(type = type, value = value)
```

Part a

```
## Build the regression model

lm_1 <- lm(value ~ type, data = data_1)
lm_1_summ <- summary(lm_1)
anova(lm_1)
```

```
## Analysis of Variance Table
##
## Response: value
##           Df Sum Sq Mean Sq F value    Pr(>F)
## type       4 1147.8 286.945  7.8145 7.168e-05 ***
## Residuals 45 1652.4  36.719
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Test my assumptions

resids_1 <- residuals(lm_1)

qqnorm(resids_1)
qqline(resids_1)
```




```
bartlett.test(value ~ type, data = data_1)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  value by type
## Bartlett's K-squared = 6.9941, df = 4, p-value = 0.1362
```

Part b

```
### Standard Confidence Intervals

### 1A - Confidence Interval

data_1A <- data_1[1:10,]
n <- length(data_1A$value)
mean_1A <- mean(data_1A$value)
sd_1A <- sd(data_1A$value)

error_1A <- qt(0.975, df = n - 1)*((sd_1A)/sqrt(n))

left_1A <- mean_1A - error_1A
right_1A <- mean_1A + error_1A

### 1B - Confidence Interval

data_1B <- data_1[11:20,]
n <- length(data_1B$value)
```

```r
mean_1B <- mean(data_1B$value)
sd_1B <- sd(data_1B$value)

error_1B <- qt(0.975, df = n - 1)*((sd_1B)/sqrt(n))

left_1B <- mean_1B - error_1B
right_1B <- mean_1B + error_1B

### 1C - Confidence Interval

data_1C <- data_1[21:30,]
n <- length(data_1C$value)
mean_1C <- mean(data_1C$value)
sd_1C <- sd(data_1C$value)

error_1C <- qt(0.975, df = n - 1)*((sd_1C)/sqrt(n))

left_1C <- mean_1C - error_1C
right_1C <- mean_1C + error_1C

### 1D - Confidence Interval

data_1D <- data_1[31:40,]
n <- length(data_1D$value)
mean_1D <- mean(data_1D$value)
sd_1D <- sd(data_1D$value)

error_1D <- qt(0.975, df = n - 1)*((sd_1D)/sqrt(n))

left_1D <- mean_1D - error_1D
right_1D <- mean_1D + error_1D

### 1E - Confidence Interval

data_1E <- data_1[41:50,]
n <- length(data_1E$value)
mean_1E <- mean(data_1E$value)
sd_1E <- sd(data_1E$value)

error_1E <- qt(0.975, df = n - 1)*((sd_1E)/sqrt(n))

left_1E <- mean_1E - error_1E
right_1E <- mean_1E + error_1E



## Bonferroni Confidence Intervals

### 1A - Confidence Interval

error_1A_bon <- qt(0.995, df = n - 1)*((sd_1A)/sqrt(n))

left_1A_bon <- mean_1A - error_1A_bon
```

```r
right_1A_bon <- mean_1A + error_1A_bon
c(left_1A_bon, right_1A_bon)
```

```
## [1] 184.3269 192.1331
```

```r
### 1B - Confidence Interval

error_1B_bon <- qt(0.995, df = n - 1)*((sd_1B)/sqrt(n))

left_1B_bon <- mean_1B - error_1B_bon
right_1B_bon <- mean_1B + error_1B_bon
c(left_1B_bon, right_1B_bon)
```

```
## [1] 186.0095 204.4505
```

```r
### 1C - Confidence Interval

error_1C_bon <- qt(0.995, df = n - 1)*((sd_1C)/sqrt(n))

left_1C_bon <- mean_1C - error_1C_bon
right_1C_bon <- mean_1C + error_1C_bon
c(left_1C_bon, right_1C_bon)
```

```
## [1] 182.1789 192.6611
```

```r
### 1D - Confidence Interval

error_1D_bon <- qt(0.995, df = n - 1)*((sd_1D)/sqrt(n))

left_1D_bon <- mean_1D - error_1D_bon
right_1D_bon <- mean_1D + error_1D_bon
c(left_1D_bon, right_1D_bon)
```

```
## [1] 185.4808 196.9392
```

```r
### 1E - Confidence Interval

error_1E_bon <- qt(0.995, df = n - 1)*((sd_1E)/sqrt(n))

left_1E_bon <- mean_1E - error_1E_bon
right_1E_bon <- mean_1E + error_1E_bon
c(left_1E_bon, right_1E_bon)
```

```
## [1] 194.5741 206.1259
```

Part c

```r
### Construct the ANOVA model

anova_1 <- aov(value ~ type, data = data_1)

### Construct the Tukey confidence intervals

TukeyHSD(anova_1)
```

```
##   Tukey multiple comparisons of means
```

```
##     95% family-wise confidence level
##
## Fit: aov(formula = value ~ type, data = data_1)
##
## $type
##      diff         lwr        upr     p adj
## B-A  7.00  -0.7002118 14.7002118 0.0907506
## C-A -0.81  -8.5102118  6.8902118 0.9981948
## D-A  2.98  -4.7202118 10.6802118 0.8057780
## E-A 12.12   4.4197882 19.8202118 0.0004798
## C-B -7.81 -15.5102118 -0.1097882 0.0453343
## D-B -4.02 -11.7202118  3.6802118 0.5783176
## E-B  5.12  -2.5802118 12.8202118 0.3377270
## D-C  3.79  -3.9102118 11.4902118 0.6317521
## E-C 12.93   5.2297882 20.6302118 0.0001837
## E-D  9.14   1.4397882 16.8402118 0.0127107
```

```
### Construct the Bonferroni confidence intervals

pairwiseCI(value ~ type, data = data_1, conf.level = 0.99)
```

```
##
## 99 %-confidence intervals
##  Method:  Difference of means assuming Normal distribution, allowing unequal variances
##
##
##     estimate   lower  upper
## B-A     7.00  -2.393 16.393
## C-A    -0.81  -6.654  5.034
## D-A     2.98  -3.257  9.217
## E-A    12.12   5.844 18.396
## C-B    -7.81 -17.497  1.877
## D-B    -4.02 -13.858  5.819
## E-B     5.12  -4.734 14.974
## D-C     3.79  -3.094 10.674
## E-C    12.93   6.014 19.846
## E-D     9.14   1.934 16.346
##
##
```

Problem 5

Part a

```
temp_dev <- rep(c("dev1", "dev2", "dev3"), each = 15)
temp_vals <- c(data_5[,2], data_5[,3], data_5[,4])

data_5a <- data.frame(device = temp_dev, value = temp_vals)

lm_5a <- lm(value ~ device, data = data_5a)
summary(lm_5a)
```

```
##
## Call:
```

```
## lm(formula = value ~ device, data = data_5a)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.192  -9.407   2.988   8.028  11.578
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 128.08067    2.42325  52.855   <2e-16 ***
## devicedev2   -0.03867    3.42699  -0.011    0.991
## devicedev3   -0.00400    3.42699  -0.001    0.999
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.385 on 42 degrees of freedom
## Multiple R-squared:  3.667e-06,  Adjusted R-squared:  -0.04762
## F-statistic: 7.7e-05 on 2 and 42 DF,  p-value: 0.9999
```

```r
anova(lm_5a)
```

```
## Analysis of Variance Table
##
## Response: value
##           Df Sum Sq Mean Sq F value Pr(>F)
## device     2    0.0   0.007   1e-04 0.9999
## Residuals 42 3699.4  88.082
```

Part b

```r
temp_doc <- rep(c("doc1", "doc2", "doc3"), each = 15)
temp_vals <- c(data_5[,5], data_5[,6], data_5[,7])

data_5b <- data.frame(doctor = temp_doc, value = temp_vals)

lm_5b <- lm(value ~ doctor, data = data_5b)
summary(lm_5b)
```

```
##
## Call:
## lm(formula = value ~ doctor, data = data_5b)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6973 -0.6907 -0.0360  0.6627  3.2993
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 125.9360     0.3449  365.16  < 2e-16 ***
## doctordoc2    4.1213     0.4877    8.45 1.33e-10 ***
## doctordoc3    8.1347     0.4877   16.68  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.336 on 42 degrees of freedom
```

```
## Multiple R-squared:  0.8688, Adjusted R-squared:  0.8626
## F-statistic: 139.1 on 2 and 42 DF,  p-value: < 2.2e-16
```

```r
anova(lm_5b)
```

```
## Analysis of Variance Table
##
## Response: value
##           Df Sum Sq Mean Sq F value    Pr(>F)
## doctor     2 496.33 248.163   139.1 < 2.2e-16 ***
## Residuals 42  74.93   1.784
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Problem 6

Part c

```r
data_6 <- read.csv("~/Documents/Rice_University/Spring_2018/STAT616/HW03/audit_b10.csv")

data_6$block <- as.factor(data_6$block)
data_6$method <- as.factor(data_6$method)
```

```r
lm_6c <- lm(score ~ block + method, data = data_6)
summary(lm_6c)
```

```
##
## Call:
## lm(formula = score ~ block + method, data = data_6)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.833 -1.125 -0.500  1.500  4.167
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   75.500      1.580  47.786  < 2e-16 ***
## block2        -1.000      2.040  -0.490 0.629872
## block3        -2.667      2.040  -1.307 0.207544
## block4        -1.667      2.040  -0.817 0.424554
## block5        -3.667      2.040  -1.798 0.089033 .
## block6        -4.000      2.040  -1.961 0.065533 .
## block7        -6.000      2.040  -2.942 0.008725 **
## block8        -8.667      2.040  -4.249 0.000483 ***
## block9        -9.000      2.040  -4.412 0.000336 ***
## block10      -12.333      2.040  -6.047 1.02e-05 ***
## method2        4.000      1.117   3.580 0.002139 **
## method3       15.500      1.117  13.874 4.72e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.498 on 18 degrees of freedom
## Multiple R-squared:  0.939,  Adjusted R-squared:  0.9017
## F-statistic: 25.18 on 11 and 18 DF,  p-value: 1.107e-08
```

```r
anova(lm_6c)
```

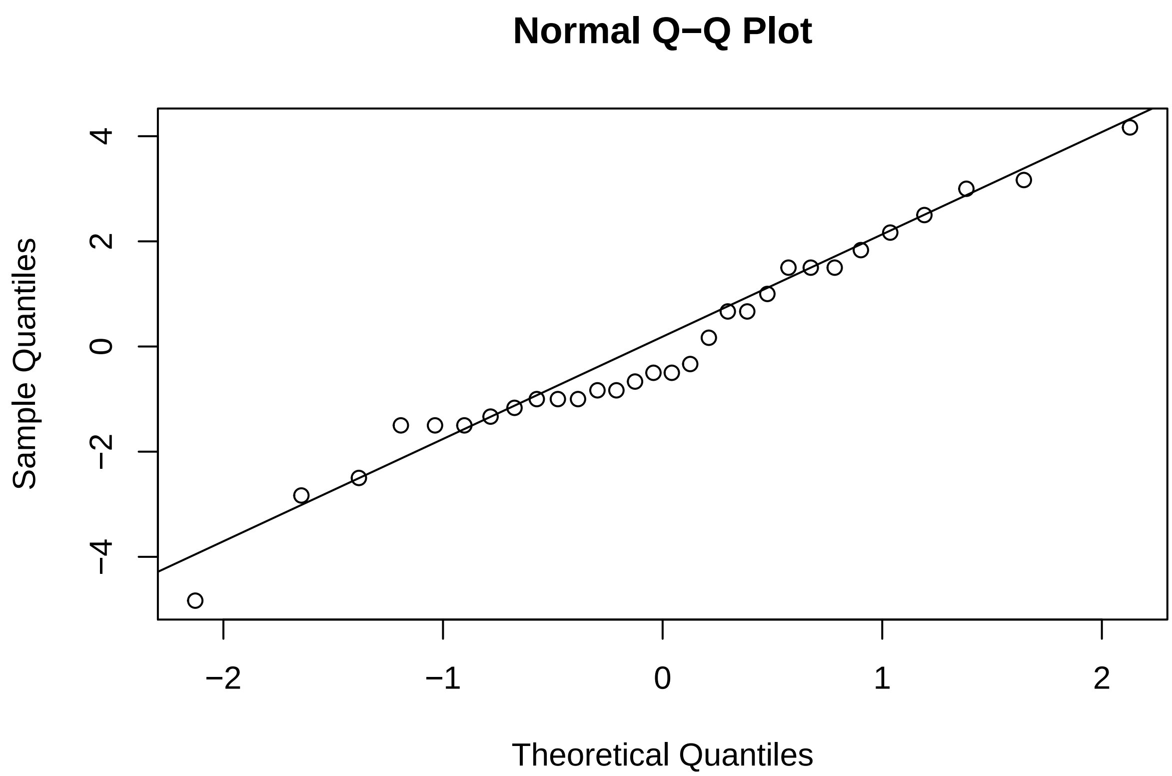
```
## Analysis of Variance Table
##
## Response: score
##           Df  Sum Sq Mean Sq  F value    Pr(>F)
## block      9  433.37   48.15   7.7157 0.0001316 ***
## method     2 1295.00  647.50 103.7537 1.315e-10 ***
## Residuals 18  112.33    6.24
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Part e

```r
anova_6e <- aov(score ~ method, data = data_6)
(tukey_6e <- TukeyHSD(anova_6e))
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = score ~ method, data = data_6)
##
## $method
##     diff        lwr       upr     p adj
## 2-1  4.0 -0.9849383  8.984938 0.1341187
## 3-1 15.5 10.5150617 20.484938 0.0000001
## 3-2 11.5  6.5150617 16.484938 0.0000129
```

```r
## Testing the assumptions

resids_6 <- residuals(lm_6c)

qqnorm(resids_6)
qqline(resids_6)
```




```r
bartlett.test(score ~ method, data = data_6)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  score by method
## Bartlett's K-squared = 1.3698, df = 2, p-value = 0.5041
```

Part f

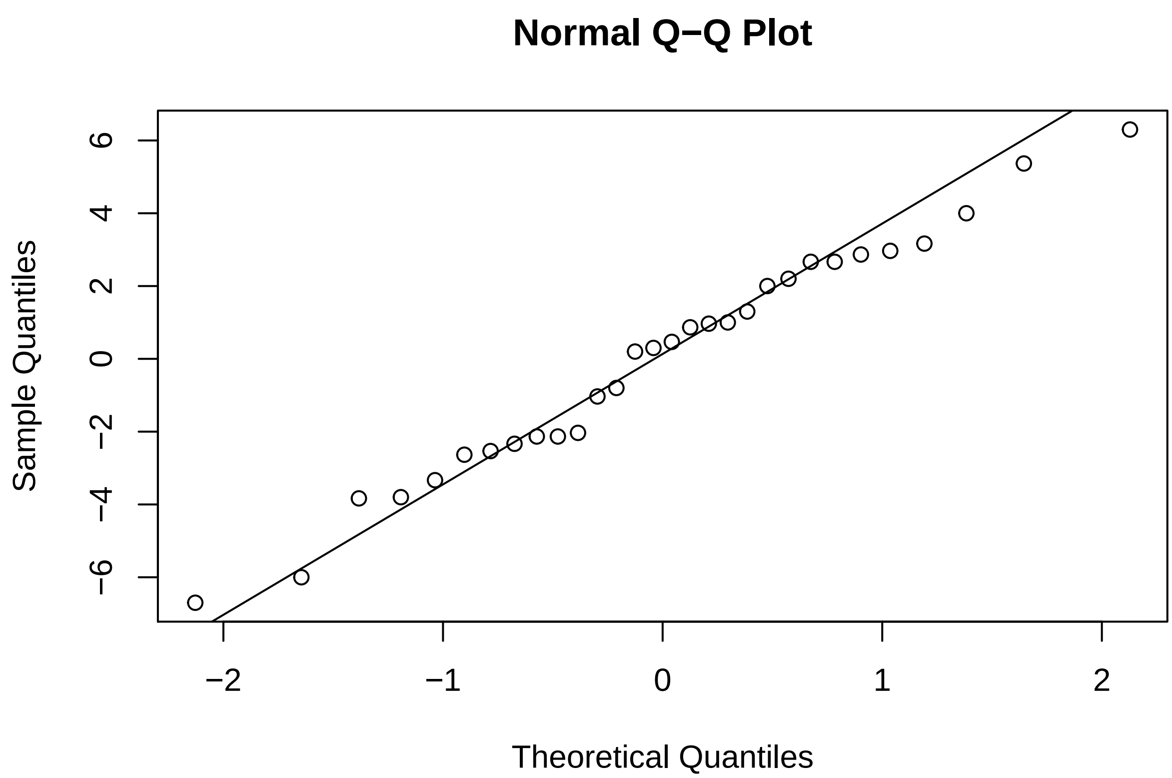
```r
data_6f <- read.csv("~/Documents/Rice_University/Spring_2018/STAT616/HW03/audit_b5.csv")

data_6f$block <- as.factor(data_6f$block)
data_6f$method <- as.factor(data_6f$method)

lm_6f <- lm(score ~ block + method, data = data_6f)
lm_6f_summ <- summary(lm_6f)
(lm_6f_anova <- anova(lm_6f))
```

```
## Analysis of Variance Table
##
## Response: score
##           Df Sum Sq Mean Sq F value    Pr(>F)
## block      4  523.2  130.80  10.226 6.586e-05 ***
## method     2 1796.5  898.23  70.222 1.607e-10 ***
## Residuals 23  294.2   12.79
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova_6f <- aov(score ~ method, data = data_6f)
(tukey_6f <- TukeyHSD(anova_6f))
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = score ~ method, data = data_6f)
##
## $method
##     diff       lwr      upr     p adj
## 2-1  8.2  2.099013 14.30099 0.0068497
## 3-1 18.9 12.799013 25.00099 0.0000001
## 3-2 10.7  4.599013 16.80099 0.0004988
```

```r
## Testing the assumptions

resids_6f <- residuals(lm_6f)

qqnorm(resids_6f)
qqline(resids_6f)
```




```r
bartlett.test(score ~ method, data = data_6f)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  score by method
## Bartlett's K-squared = 2.4602, df = 2, p-value = 0.2923
```

Problem 7

```r
data_7 <- read.csv("~/Documents/Rice_University/Spring_2018/STAT616/HW03/resistor.csv")

data_7$plate <- as.factor(data_7$plate)
data_7$shape <- rep(c("A", "B", "C", "D"), each = 4)
data_7$shape <- as.factor(data_7$shape)

lm_7 <- lm(noise ~ plate + shape, data = data_7)
summary(lm_7)
```

```
##
## Call:
## lm(formula = noise ~ plate + shape, data = data_7)
##
## Residuals:
##        1        2        3        6        7        8        9       11
## -0.07000  0.08375 -0.01375  0.05875 -0.04875 -0.01000 -0.07375  0.06250
##       12       13       14       16
##  0.01125  0.14375 -0.14250 -0.00125
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.18000    0.08938  13.202 4.45e-05 ***
## plate2       0.43625    0.10135   4.305  0.00768 **
## plate3       0.43375    0.10135   4.280  0.00786 **
## plate4       0.50500    0.10135   4.983  0.00417 **
## shapeB      -0.45500    0.10135  -4.490  0.00646 **
## shapeC      -0.15625    0.10135  -1.542  0.18378
## shapeD      -0.50375    0.10135  -4.971  0.00421 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.117 on 5 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.9223, Adjusted R-squared:  0.829
## F-statistic: 9.887 on 6 and 5 DF,  p-value: 0.01184
```

```r
(lm_7_anoca <- anova(lm_7))
```

```
## Analysis of Variance Table
##
## Response: noise
##           Df  Sum Sq  Mean Sq F value  Pr(>F)
## plate      3 0.34737 0.115789  8.4548 0.02105 *
## shape      3 0.46506 0.155019 11.3194 0.01146 *
## Residuals  5 0.06847 0.013695
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova_7 <- aov(noise ~ plate + shape, data = data_7)
TukeyHSD(anova_7)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
```

```
##
## Fit: aov(formula = noise ~ plate + shape, data = data_7)
##
## $plate
##            diff           lwr       upr     p adj
## 2-1  0.33666667 -0.0159083249 0.6892417 0.0590411
## 3-1  0.45000000  0.0974250084 0.8025750 0.0194330
## 4-1  0.35333333  0.0007583418 0.7059083 0.0496095
## 3-2  0.11333333 -0.2392416582 0.4659083 0.6598457
## 4-2  0.01666667 -0.3359083249 0.3692417 0.9978627
## 4-3 -0.09666667 -0.4492416582 0.2559083 0.7507145
##
## $shape
##            diff         lwr         upr     p adj
## B-A -0.40444444 -0.75701944 -0.05186945 0.0297637
## C-A -0.13888889 -0.49146388  0.21368610 0.5231292
## D-A -0.44777778 -0.80035277 -0.09520279 0.0198291
## C-B  0.26555556 -0.08701944  0.61813055 0.1287196
## D-B -0.04333333 -0.39590832  0.30924166 0.9660591
## D-C -0.30888889 -0.66146388  0.04368610 0.0795257
```