HW04

Elliot Smith

4/2/2018

Problem 1

Part a

Estimates

- $\sigma_\epsilon^2 = 1.784$
- $\sigma_\alpha^2 = \frac{248.163 - 1.784}{15} = 16.42527$
- $SE(\hat{\mu}) = \sqrt{\frac{\sigma_\epsilon^2}{45}} = 0.1991091$
- Confidence Interval = $[129.6195, 130.4232]$

Part b

Estimates

- $\sigma_\epsilon^2 = 1.784$
- $\sigma_\alpha^2 = 10.911$
- $SE(\hat{\mu}) = \sqrt{\frac{\sigma_\epsilon^2 + 15\sigma_\alpha^2}{45}} = 1.917458$
- Confidence Interval = $[124.668177, 135.374484]$

Part c

Estimates

- $\sigma_\epsilon^2 = 1.784$
- $\sigma_\alpha^2 = 16.425$
- $SE(\hat{\mu}) = \sqrt{\frac{\sigma_\epsilon^2 + 15\sigma_\alpha^2}{45}} = 2.348328$
- Confidence Interval = $[124.668177, 135.374484]$

Part d

- All of the methods have the same σ_ϵ^2
- ANOVA and REML have the same σ_α^2 of 16.425, while the ML method's is 10.911
- ANOVA has smallest $SE(\hat{\mu})$ at 0.1991, ML is next smallest at 1.917 and REML is the largest at 2.348
- The Confidence Interval for ANOVA is the tightest by far, where REML and ML are the same and much wider
- 1a and 1c are biased estimators, 1b estimators are unbiases
- I prefer 1b (the REML method) because the estimators are unbiased

Part e - ANOVA Method

Estimates

- $\sigma_\epsilon^2 = 88.082$
- $\sigma_\alpha^2 = \frac{0.007 - 88.082}{15} = -5.871667 = 0$
- $SE(\hat{\mu}) = \sqrt{\frac{\sigma_\epsilon^2}{45}} = 1.399063$
- Confidence Interval = $[125.243023, 130.889866]$

Part e - LM Method

Estimates

- $\sigma_\epsilon^2 = 82.21$
- $\sigma_\alpha^2 = 0$
- $SE(\hat{\mu}) = \sqrt{\frac{\sigma_\epsilon^2 + 15\sigma_\alpha^2}{45}} = 1.351625$
- Confidence Interval = $[125.359749, 130.773140]$

Part e - REML Method

Estimates

- $\sigma_\epsilon^2 = 84.08$
- $\sigma_\alpha^2 = 0$
- $SE(\hat{\mu}) = \sqrt{\frac{\sigma_\epsilon^2 + 15\sigma_\alpha^2}{45}} = 1.366911$
- Confidence Interval = $[125.359749, 130.773140]$

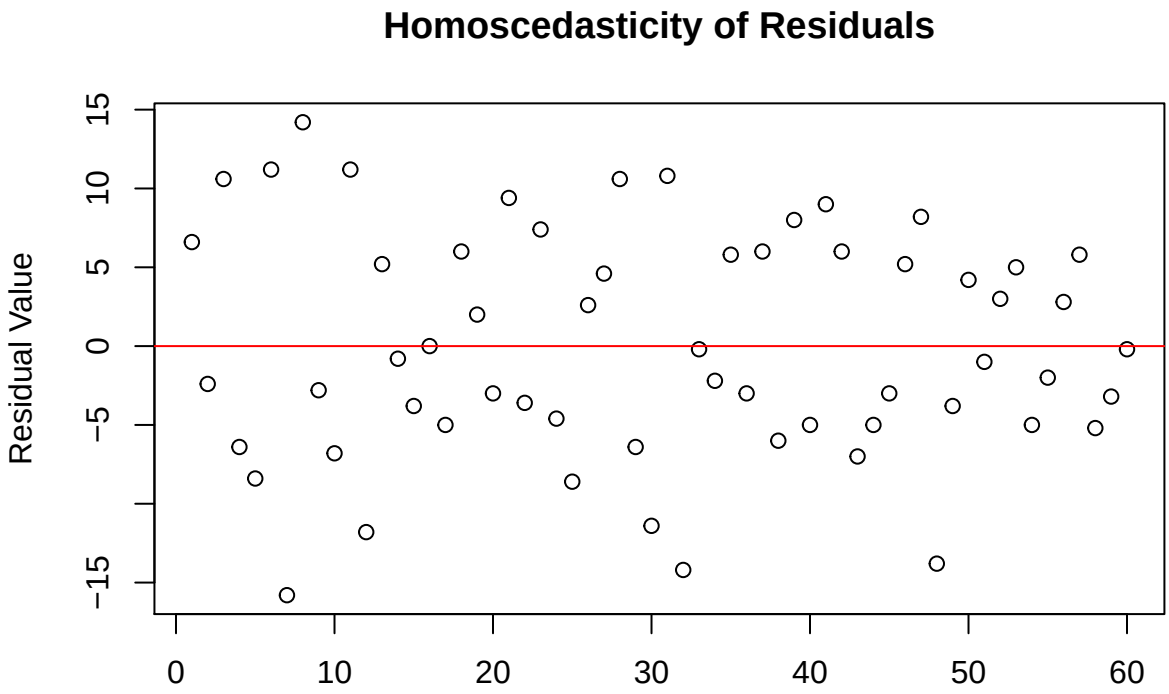
Part e - Results Comparison

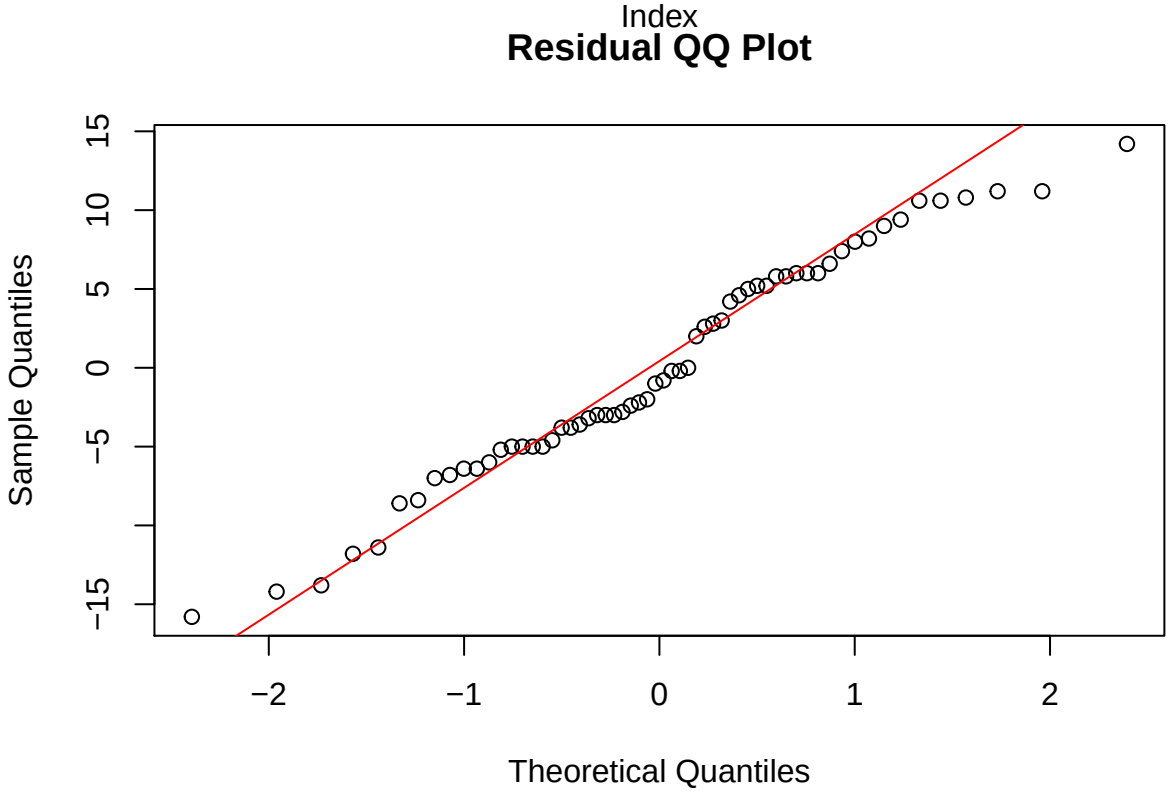
- The largest σ_ϵ^2 is from ANOVA with 88.082, followed by REML with 84.08 and finally ML with 82.21
- ANOVA, ML and REML all have a σ_α^2 of 0
- ML has smallest $SE(\hat{\mu})$ at 1.352, REML is next smallest at 1.367 and ANOVA is the largest at 1.399
- All of the Confidence Intervals are about equally as tight, where REML and ML are the same again
- ANOVA and ML are biased estimators, REML estimators are unbiased
- I prefer the REML method because the estimators are unbiased

Problem 2

Interaction Effects

Diagnostics





Summary

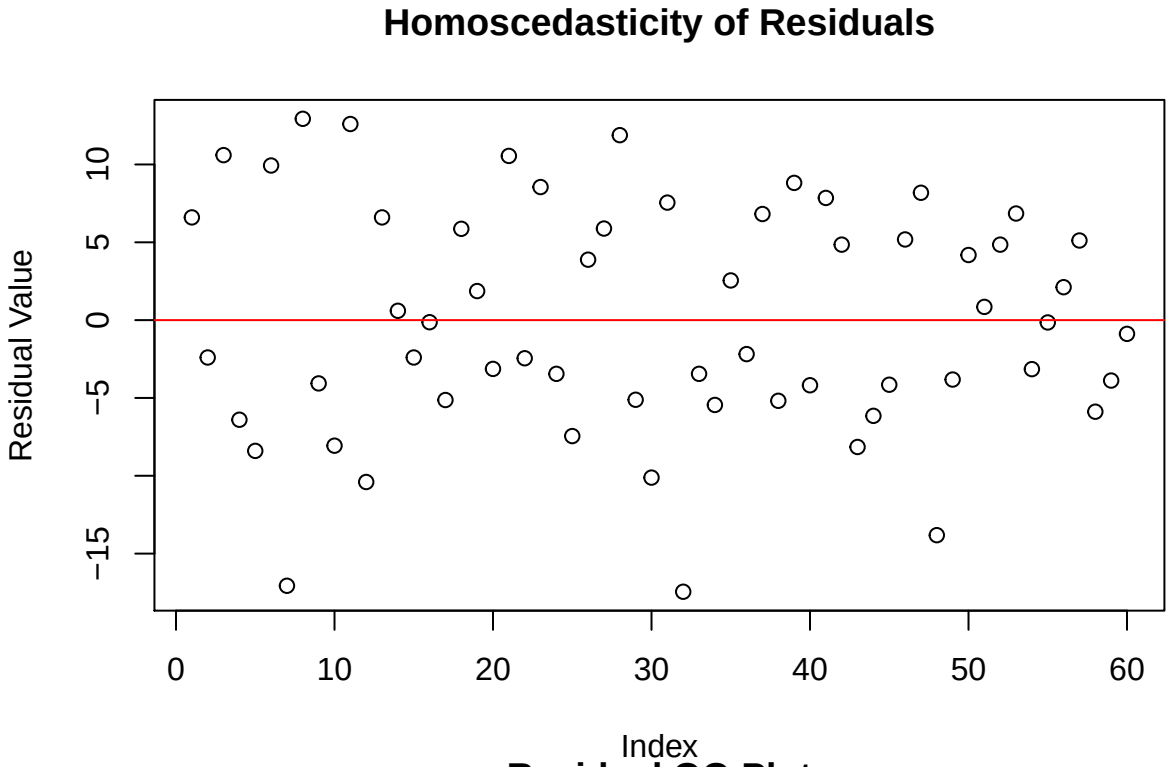
- Significant Location p-values: Ground, Lower, Middle
- Significant Trap p-values: None
- Significant Interaction p-values: None
- Based on ANOVA Table, Location is significant and a large portion of the variance is explained by Location
- Based on ANOVA Table, an even larger amount of the variance is explained by the error component
- The confidence interval is wider when incorporating interaction affects, accounting for increased uncertainty
- According to our diagnostic plots, residuals are approximately normal and have equal variance

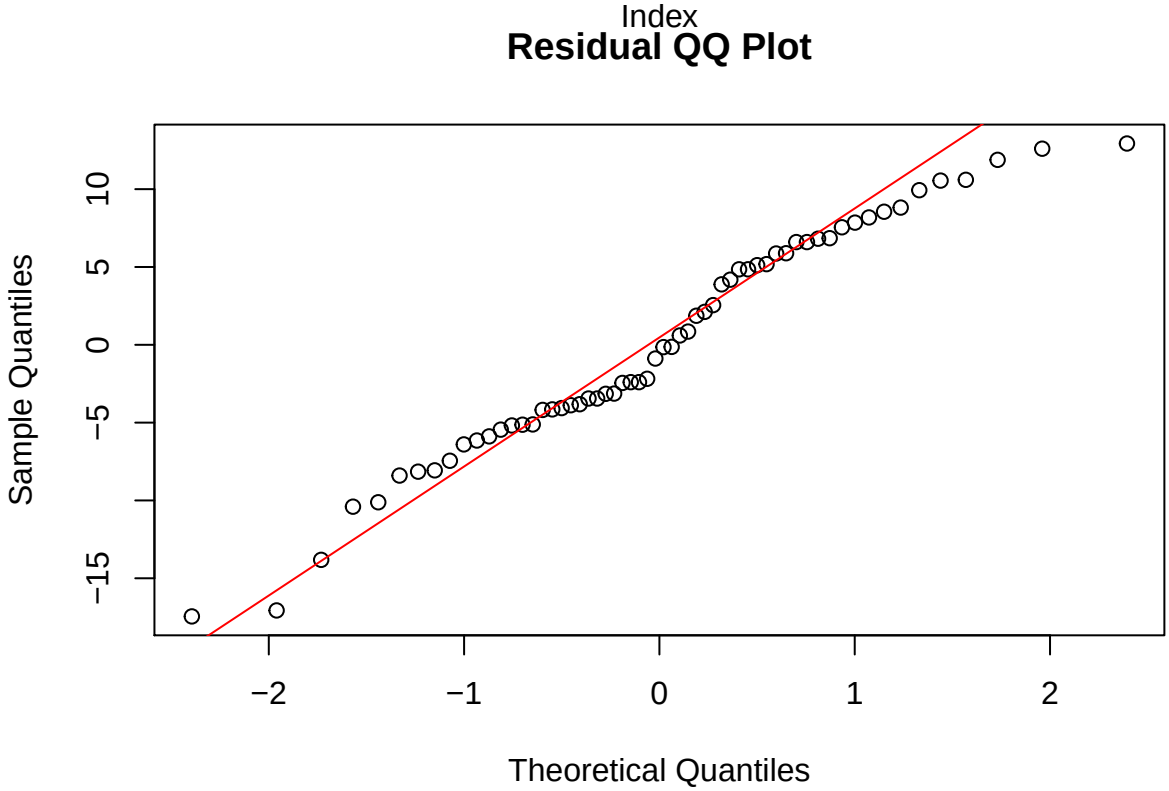
ANOVA Table

```
## Analysis of Variance Table
##
## Response: count
##               Df  Sum Sq Mean Sq F value    Pr(>F)
## location       3 1981.38  660.46 10.4503 2.094e-05 ***
## trap           2  113.03   56.52  0.8943    0.4156
## location:trap  6  114.97   19.16  0.3032    0.9322
## Residuals     48 3033.60   63.20
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

No Interaction Effects

Diagnostics





Summary

- Significant Location p-values: Ground, Lower, Middle
- Significant Trap p-values: None
- Based on ANOVA Table, Location is significant and a large portion of the variance is explained by Location
- Based on ANOVA Table, an even larger amount of the variance is explained by the error component
- The confidence interval is tighter when ignoring interaction affects, accounting for decreased uncertainty
- According to our diagnostic plots, residuals are approximately normal and have equal variance

ANOVA Table

```
## Analysis of Variance Table
##
## Response: count
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## location   3 1981.38  660.46 11.3273 7.167e-06 ***
## trap       2  113.03   56.52  0.9693    0.3859
## Residuals 54 3148.57   58.31
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Problem 3

Part a

ANOVA Table

```
##
## Call:
## lm(formula = value ~ oper + part + (oper * part), data = df_3)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0035364 -0.0011341  0.0004091  0.0011818  0.0022818
##
## Coefficients:
##                  Estimate Std. Error  t value Pr(>|t|)
## (Intercept)     9.555e-01  7.293e-04 1310.101   <2e-16 ***
## operthree       1.800e-03  1.031e-03    1.745   0.0866 .
## opertwo         6.667e-04  1.031e-03    0.646   0.5208
## part            3.636e-05  1.175e-04    0.309   0.7582
## operthree:part -1.545e-04  1.662e-04   -0.930   0.3567
## opertwo:part   -5.758e-05  1.662e-04   -0.346   0.7304
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.00151 on 54 degrees of freedom
## Multiple R-squared:  0.08768,    Adjusted R-squared:  0.003201
## F-statistic: 1.038 on 5 and 54 DF,  p-value: 0.405
```

```
## Analysis of Variance Table
##
## Response: value
##           Df    Sum Sq    Mean Sq F value Pr(>F)
## oper       2 9.233e-06 4.6167e-06  2.0251 0.1419
## part       1 5.840e-07 5.8380e-07  0.2561 0.6149
## oper:part  2 2.013e-06 1.0066e-06  0.4415 0.6453
## Residuals 54 1.231e-04 2.2797e-06
```

Estimates

- $\sigma^2_\epsilon = 0.0000022797$
- $\sigma^2_{oper/part} = \frac{MS_{oper/part} - MSE}{n} = \frac{0.0000010066 - 0.0000022797}{2} = -0.00000063655 = 0$
- $\sigma^2_{part} = \frac{MS_{part} - MS_{oper/part}}{n*3} = \frac{0.0000005838 - 0}{6} = 0.0000000973$
- $\sigma^2_{oper} = \frac{MS_{oper} - MS_{oper/part}}{n*10} = \frac{0.0000046167 - 0}{20} = 0.000000230835$

Conclusion

- We can get valid estimates from the ANOVA table because we have balance!

Part b

ML Method

- Results obtained from the LM Summary Table (see Code/Table Appendix)
- $\sigma^2_\epsilon = 0.0000004329$
- $\sigma^2_{oper/part} = 0.00000000000000000536$
- $\sigma^2_{part} = 0.000001722$
- $\sigma^2_{oper} = 0.0000001807$

REML MEthod

- Results obtained from the LM Summary Table (see Code/Table Appendix)
- $\sigma^2_\epsilon = 0.0000004326$
- $\sigma^2_{oper/part} = 0$
- $\sigma^2_{part} = 0.000001871$
- $\sigma^2_{oper} = 0.0000002092$

Comparisons

- The σ^2_ϵ are extremely small in all cases, however largest with ANOVA and approximately equal with ML and REML
- The estimates for $\sigma^2_{oper/part}$ are approximately 0 in all cases
- And again with variance estimates, σ^2_{part} and σ^2_{oper}, they are all extremely small and approximately 0
- Again, we know the ANOVA and ML are biased, while REML is unbiased

Part c

- According to the ANOVA table from Part a, there are no significant sources of variation; all p-values (based on F-statistics) are > 0.05

Problem 4

Part a

ANOVA Table

```
## Analysis of Variance Table
##
## Response: decrease
##           Df Sum Sq Mean Sq F value Pr(>F)
## treatment  1  245.0  245.00  1.7492 0.2025
## Residuals 18 2521.2  140.07
```

Results

- According to our ANOVA Table, our p-value of 0.2025 means we must not reject our null hypothesis of equality of treatments

Part b

ANOVA Table

```
## Analysis of Variance Table
##
## Response: decrease
##                   Df  Sum Sq Mean Sq F value    Pr(>F)
## treatment          1  245.00  245.00  4.3971   0.05225 .
## initial            1 1490.08 1490.08 26.7428 9.279e-05 ***
## treatment:initial  1  139.62  139.62  2.5058   0.13299
## Residuals         16  891.50   55.72
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results

- According to our ANOVA Table, our p-value of 0.0606 means we must still not reject our null hypothesis of equality of treatments, however, the introduction of the initial weights clearly shows that it does have an effect, evident by the significant drop in our p-value.

Part c

- Please see each results section for results comparision. The biggest takeaway here is that when we account for the initial pressure, we see a marked drop in our p-value, which means it contributes strongly

to the treatment differences. I prefer the results from Part b because the ANOVA table from Part b tells us that the initial weight is significant and this is important infomration to know.

STAT616 Problems

Code/Table Appendix

```
########## Workspace Prep ##########

## Load in the necessary libraries

library(lme4)

## For later use

# options(contrasts = rep("contr.treatment", 2))
# options(contrasts = rep("contr.sum", 2))



########## Problem 1 ##########



## Load in the data

data_1 <- read.csv("~/Documents/Rice_University/Spring_2018/STAT616/HW04/blood.csv")

##### Part a #####

data_1a <- setNames(data.frame(matrix(NA, ncol = 2, nrow = 45)), c("doctor", "value"))
data_1a$doctor <- c(rep("doc1", times = 15), rep("doc2", times = 15), rep("doc3", times = 15))
data_1a$value <- c(data_1$doc1, data_1$doc2, data_1$doc3)

mu_doc <- mean(data_1a$value)

lm_1a <- lm(value ~ doctor, data = data_1a)
# lm_1a <- lm(value ~ doctor, data = data_1a, contrasts = list(doctor = contr.sum))
anova(lm_1a)
```

```
## 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
```

```
summary(lm_1a)
```

```
##
## Call:
## lm(formula = value ~ doctor, data = data_1a)
##
## 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
```

```
sigma_alpha_1a <- (248.163 - 1.784) / 15
sigma_epsilon_1a <- 1.784
# For standard error, check against results from model fit by contra-sum
se_mu_hat_1a <- sqrt(sigma_epsilon_1a / 45)
# For confidence interval, check against results from model fit by contra-sum
ci_mu_hat_1a <- confint(lm_1a)
# c(mu_doc - qt(0.975, df = 42) * se_mu_hat_1a, mu_doc + qt(0.975, df = 42) * se_mu_hat_1a)

##### Part b #####

lm_1b <- lmer(value ~ (1|doctor), data = data_1a, REML = FALSE)
summary(lm_1b)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: value ~ (1 | doctor)
##    Data: data_1a
##
##      AIC      BIC   logLik deviance df.resid
##    173.3    178.8    -83.7    167.3       42
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.01913 -0.53908 -0.04267  0.49641  2.50281
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  doctor   (Intercept) 10.911   3.303
##  Residual              1.784   1.336
## Number of obs: 45, groups:  doctor, 3
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  130.021      1.917   67.81
```

```
sigma_alpha_1b <- 10.911
sigma_epsilon_1b <- 1.784
se_mu_hat_1b <- sqrt((sigma_epsilon_1b + (15 * sigma_alpha_1b)) / 45)
ci_mu_hat_1b <- confint(lm_1b)
```

```
## Computing profile confidence intervals ...
```

```
##### Part c #####

lm_1c <- lmer(value ~ (1|doctor), data = data_1a, REML = TRUE)
```

```r
summary(lm_1c)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: value ~ (1 | doctor)
##    Data: data_1a
##
## REML criterion at convergence: 164
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.01922 -0.52809 -0.04894  0.49631  2.49191
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  doctor   (Intercept) 16.425   4.053
##  Residual              1.784   1.336
## Number of obs: 45, groups:  doctor, 3
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  130.021      2.348   55.37
```

```r
sigma_alpha_1c <- 16.425
sigma_epsilon_1c <- 1.784
se_mu_hat_1c <- sqrt((sigma_epsilon_1c + (15 * sigma_alpha_1c)) / 45)
ci_mu_hat_1c <- confint(lm_1c)
```

```
## Computing profile confidence intervals ...
```

```r
##### Part d #####

# All of the methods have the same sigma_square_epsilon
# ANOVA and REML have the same sigma_square_alpha of 16.425, while the ML method's is 10.911
# ANOVA has smallest SE of mu_hat at 0.1991, ML is next smallest at 1.917 and REML is the largest at 2.
# The CI for ANOVA is the tightest by far, where REML and ML are the same and much wider
# 1a and 1c are biased estimators, 1b is unbiased unbiased
# I prefer 1b (the REML method) because the estimators are unbiased

##### Part e #####

data_1e <- setNames(data.frame(matrix(NA, ncol = 2, nrow = 45)), c("device", "value"))
data_1e$device <- c(rep("dev1", times = 15), rep("dev2", times = 15), rep("dev3", times = 15))
data_1e$value <- c(data_1$dev1, data_1$dev2, data_1$dev3)

mu_dev <- mean(data_1e$value)

## ANOVA

lm_1e_anova <- lm(value ~ device, data = data_1e)
# lm_1e_anova <- lm(value ~ device, data = data_1e, contrasts = list(device = contr.sum))
anova(lm_1e_anova)
```

```
## 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
```

```
summary(lm_1e_anova)
```

```
##
## Call:
## lm(formula = value ~ device, data = data_1e)
##
## 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
```

```
sigma_alpha_1e_anova <- (0.007 - 88.082) / 15
sigma_alpha_1e_anova <- 0
sigma_epsilon_1e_anova <- 88.082
# For standard error, check against results from model fit by contra-sum
se_mu_hat_1e_anova <- sqrt((sigma_epsilon_1e_anova + (15 * sigma_alpha_1e_anova)) / 45)
# For confidence interval, check against results from model fit by contra-sum
ci_mu_hat_1e_anova <- confint(lm_1e_anova)
# c(mu_dev - qt(0.975, df = 42) * se_mu_hat_1e_anova, mu_dev + qt(0.975, df = 42) * se_mu_hat_1e_anova)

## ML

lm_1e_ml <- lmer(value ~ (1|device), data = data_1e, REML = FALSE)
summary(lm_1e_ml)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: value ~ (1 | device)
##    Data: data_1e
##
##      AIC      BIC   logLik deviance df.resid
##    332.1    337.5   -163.1    326.1       42
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8988 -1.0363  0.3268  0.8827  1.2743
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  device   (Intercept)  0.00    0.000
##  Residual             82.21    9.067
```

```
## Number of obs: 45, groups:  device, 3
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  128.066      1.352   94.75
```

```
sigma_alpha_1e_ml <- 0
sigma_epsilon_1e_ml <- 82.21
se_mu_hat_1e_ml <- sqrt((sigma_epsilon_1e_ml + (15 * sigma_alpha_1e_ml)) / 45)
ci_mu_hat_1e_ml <- confint(lm_1e_ml)
```

```
## Computing profile confidence intervals ...
```

```
## Warning in optwrap(optimizer, par = start, fn = function(x)
## dd(mkpar(npar1, : convergence code 3 from bobyqa: bobyqa -- a trust region
## step failed to reduce q
```

```
## REML

lm_1e_reml <- lmer(value ~ (1|device), data = data_1e, REML = TRUE)
```

```
## Warning in optwrap(optimizer, devfun, getStart(start, rho$lower, rho$pp), :
## convergence code 3 from bobyqa: bobyqa -- a trust region step failed to
## reduce q
```

```
summary(lm_1e_reml)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: value ~ (1 | device)
##    Data: data_1e
##
## REML criterion at convergence: 323.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8776 -1.0248  0.3232  0.8729  1.2600
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  device   (Intercept)  0.00    0.000
##  Residual             84.08    9.169
## Number of obs: 45, groups:  device, 3
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  128.066      1.367   93.69
## convergence code: 3
```

```
sigma_alpha_1e_reml <- 0
sigma_epsilon_1e_reml <- 84.08
se_mu_hat_1e_reml <- sqrt((sigma_epsilon_1e_reml + (15 * sigma_alpha_1e_reml)) / 45)
ci_mu_hat_1e_reml <- confint(lm_1e_reml)
```

```
## Computing profile confidence intervals ...
```

```
## Warning in optwrap(optimizer, par = start, fn = function(x)
## dd(mkpar(npar1, : convergence code 3 from bobyqa: bobyqa -- a trust region
## step failed to reduce q
```

```
## Results Comparison

# The largest sigma_square_epsilon is from ANOVA with 88.082, followed by REML with 84.08 and finally M
# ANOVA, ML and REML all have a sigma_square_alpha of 0
# ML has smallest SE of mu_hat at 1.352, REML is next smallest at 1.367 and ANOVA is the largest at 1.3
# All of the CIs are about equally as tight, where REML and ML are the same again
# 1a and 1c are biased estimators, 1b is unbiased unbiased
# I prefer 1b (the REML method) because the estimators are unbiased



########## Problem 2 ##########



## Load in the data

data_2 <- read.table("~/Documents/Rice_University/Spring_2018/STAT616/HW04/moth.txt", header = TRUE)

## Interaction effects model

lm_interact_2 <- lm(count ~ location + trap + (location * trap), data = data_2)
summary(lm_interact_2)
```

```
##
## Call:
## lm(formula = count ~ location + trap + (location * trap), data = data_2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -15.80  -5.00  -0.90   5.85  14.20
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 19.200      3.555   5.400 2.04e-06 ***
## locationLower               16.800      5.028   3.341  0.00162 **
## locationMiddle              12.600      5.028   2.506  0.01565 *
## locationTop                  3.800      5.028   0.756  0.45347
## trapScent                   -2.200      5.028  -0.438  0.66367
## trapSugar                    1.800      5.028   0.358  0.72191
## locationLower:trapScent     -1.000      7.111  -0.141  0.88875
## locationMiddle:trapScent    -1.800      7.111  -0.253  0.80124
## locationTop:trapScent        0.600      7.111   0.084  0.93310
## locationLower:trapSugar     -6.600      7.111  -0.928  0.35795
## locationMiddle:trapSugar    -0.200      7.111  -0.028  0.97768
## locationTop:trapSugar        0.800      7.111   0.113  0.91089
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.95 on 48 degrees of freedom
## Multiple R-squared:  0.4214, Adjusted R-squared:  0.2888
## F-statistic: 3.178 on 11 and 48 DF,  p-value: 0.002653
```

```r
anova(lm_interact_2)
```

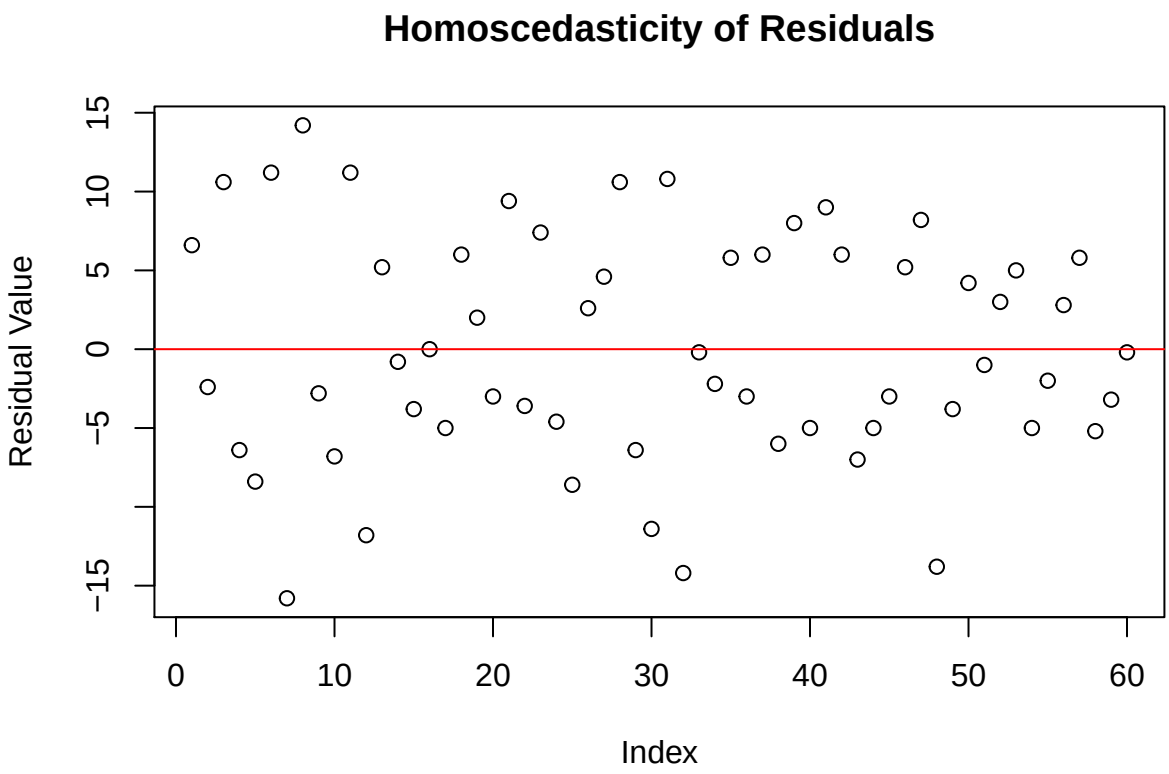
```
## Analysis of Variance Table
##
## Response: count
##               Df  Sum Sq Mean Sq F value    Pr(>F)
## location       3 1981.38  660.46 10.4503 2.094e-05 ***
## trap           2  113.03   56.52  0.8943    0.4156
## location:trap  6  114.97   19.16  0.3032    0.9322
## Residuals     48 3033.60   63.20
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
confint(lm_interact_2)
```

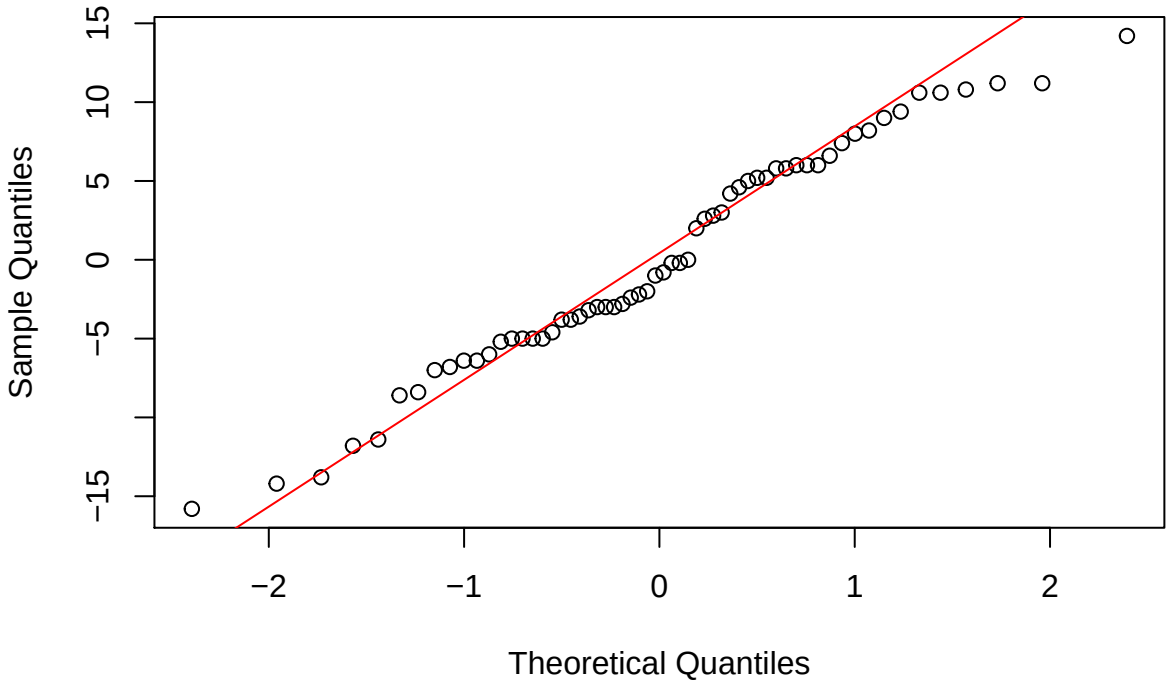
```
##                                2.5 %    97.5 %
## (Intercept)                12.051635 26.348365
## locationLower               6.690685 26.909315
## locationMiddle              2.490685 22.709315
## locationTop                -6.309315 13.909315
## trapScent                 -12.309315  7.909315
## trapSugar                  -8.309315 11.909315
## locationLower:trapScent   -15.296730 13.296730
## locationMiddle:trapScent  -16.096730 12.496730
## locationTop:trapScent     -13.696730 14.896730
## locationLower:trapSugar   -20.896730  7.696730
## locationMiddle:trapSugar  -14.496730 14.096730
## locationTop:trapSugar     -13.496730 15.096730
```

```r
## Diagnostics

plot(resid(lm_interact_2), xlab = "Index", ylab = "Residual Value", main = "Homoscedasticity of Residual
abline(h = 0, col = "red")
```




```r
qqnorm(resid(lm_interact_2), main = "Residual QQ Plot")
qqline(resid(lm_interact_2), col = "red")
```




```
## Summary
```

```r
# Significant Location p-values: Ground, Lower, Middle
# Significant Trap p-values: None
# Significant Interaction p-values: None
# Based on ANOVA Table, Location is significant and a large portion of the variance is explained by Loc
# Based on ANOVA Table, an even larger amount of the variance is explained by the error component
# The confidence interval is wider when incorporating interaction affects, accounting for increased unc
# According to our diagnostic plots, residuals are approximately normal and have equal variance

## No interaction effects model

lm_no_interact_2 <- lm(count ~ location + trap, data = data_2)
summary(lm_no_interact_2)
```

```
##
## Call:
## lm(formula = count ~ location + trap, data = data_2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.4500  -5.1208  -0.5167   6.0625  12.9333
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      19.883      2.415   8.234 4.14e-11 ***
## locationLower    14.267      2.788   5.117 4.23e-06 ***
## locationMiddle   11.933      2.788   4.280 7.70e-05 ***
## locationTop       4.267      2.788   1.530    0.132
## trapScent        -2.750      2.415  -1.139    0.260
## trapSugar         0.300      2.415   0.124    0.902
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.636 on 54 degrees of freedom
## Multiple R-squared:  0.3995, Adjusted R-squared:  0.3439
## F-statistic: 7.184 on 5 and 54 DF,  p-value: 3.236e-05
```

```r
anova(lm_no_interact_2)
```

```
## Analysis of Variance Table
##
## Response: count
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## location   3 1981.38  660.46 11.3273 7.167e-06 ***
## trap       2  113.03   56.52  0.9693    0.3859
## Residuals 54 3148.57   58.31
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

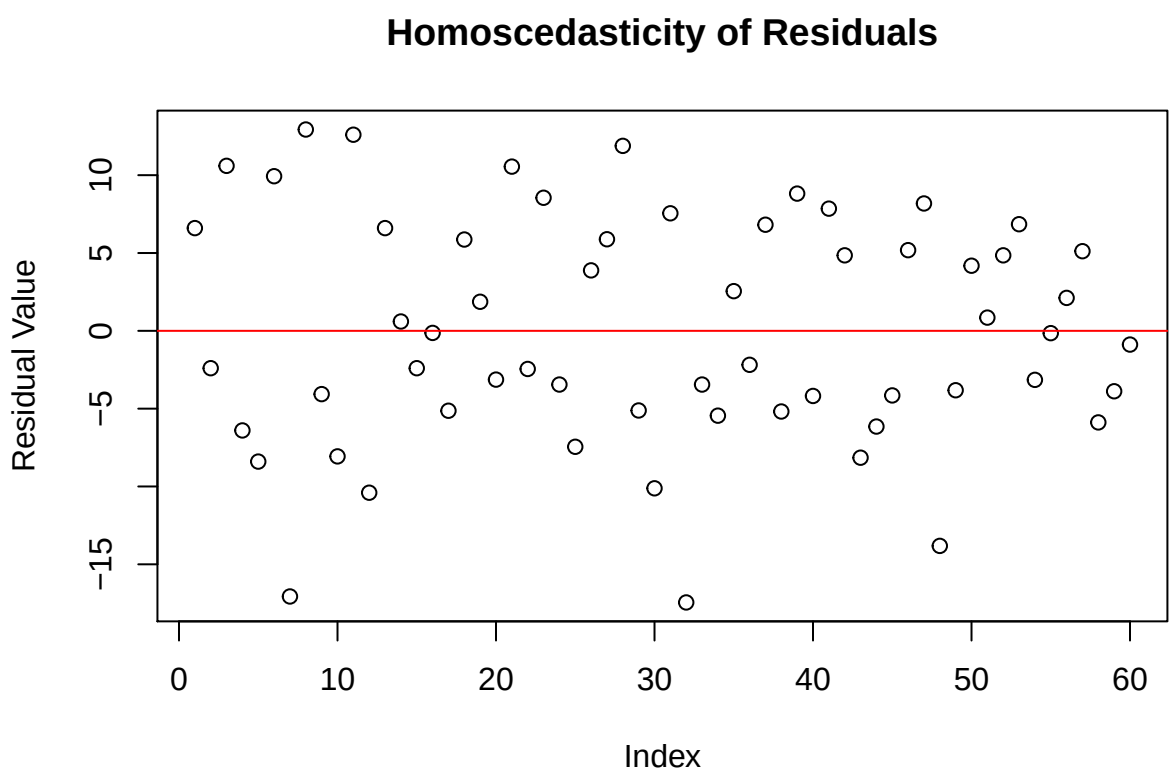
```r
confint(lm_no_interact_2)
```

```
##                    2.5 %    97.5 %
## (Intercept)    15.042192 24.724475
## locationLower   8.676598 19.856736
## locationMiddle  6.343264 17.523402
## locationTop    -1.323402  9.856736
```

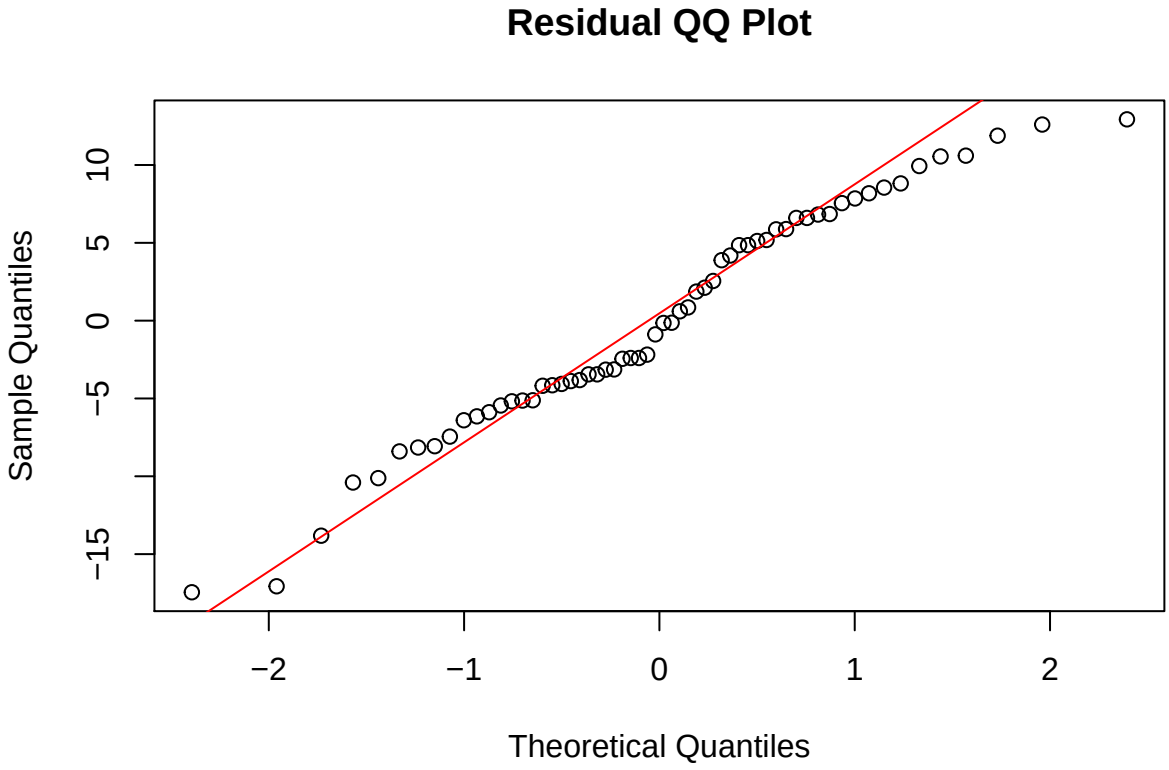
```
## trapScent      -7.591142  2.091142
## trapSugar      -4.541142  5.141142
```

```
## Diagnostics

plot(resid(lm_no_interact_2), xlab = "Index", ylab = "Residual Value", main = "Homoscedasticity of Resi
abline(h = 0, col = "red")
```




```
qqnorm(resid(lm_no_interact_2), main = "Residual QQ Plot")
qqline(resid(lm_no_interact_2), col = "red")
```




```
## Summary

# Significant Location p-values: Ground, Lower, Middle
# Significant Trap p-values: None
# Based on ANOVA Table, Location is significant and a large portion of the variance is explained by Loc
# Based on ANOVA Table, an even larger amount of the variance is explained by the error component
# The confidence interval is tighter when ignoring interaction affects, accounting for decreased uncert
# According to our diagnostic plots, residuals are approximately normal and have equal variance



########## Problem 3 ##########



## Load in the data

data_3 <- read.table("~/Documents/Rice_University/Spring_2018/STAT616/HW04/thick_guage.txt", header = T
value <- c(data_3[,2], data_3[,3], data_3[,4], data_3[,5], data_3[,6], data_3[,7])
oper <- c(rep("one", times = 20), rep("two", times = 20), rep("three", times = 20))
part <- rep(data_3[,1], times = 6)

df_3 <- data.frame(part, oper, value)

##### Part a #####

# format(scientific_value, scientific = FALSE)
format(9.73e-08, scientific = FALSE)
```

```
## [1] "0.0000000973"
```

```r
lm_3a <- lm(value ~ oper + part + (oper * part), data = df_3)
summary(lm_3a)
```

```
##
## Call:
## lm(formula = value ~ oper + part + (oper * part), data = df_3)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0035364 -0.0011341  0.0004091  0.0011818  0.0022818
##
## Coefficients:
##                  Estimate Std. Error  t value Pr(>|t|)
## (Intercept)     9.555e-01  7.293e-04 1310.101   <2e-16 ***
## operthree       1.800e-03  1.031e-03    1.745   0.0866 .
## opertwo         6.667e-04  1.031e-03    0.646   0.5208
## part            3.636e-05  1.175e-04    0.309   0.7582
## operthree:part -1.545e-04  1.662e-04   -0.930   0.3567
## opertwo:part   -5.758e-05  1.662e-04   -0.346   0.7304
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.00151 on 54 degrees of freedom
## Multiple R-squared:  0.08768,    Adjusted R-squared:  0.003201
## F-statistic: 1.038 on 5 and 54 DF,  p-value: 0.405
```

```r
anova(lm_3a)
```

```
## Analysis of Variance Table
##
## Response: value
##           Df    Sum Sq    Mean Sq F value Pr(>F)
## oper       2 9.233e-06 4.6167e-06  2.0251 0.1419
## part       1 5.840e-07 5.8380e-07  0.2561 0.6149
## oper:part  2 2.013e-06 1.0066e-06  0.4415 0.6453
## Residuals 54 1.231e-04 2.2797e-06
```

```r
sigma_epsilon_3a <- 0.0000022797
sigma_oper_part_3a <- (0.0000010066 - sigma_epsilon_3a) / 2
sigma_oper_part_3a <- 0
sigma_part_3a <- (0.0000005838 - sigma_oper_part_3a) / (2 * 3)
sigma_oper_3a <- (0.0000046167 - sigma_oper_part_3a) / (2 * 10)

# We can get valid estimates from the ANOVA table because we have balance!

##### Part b #####

## ML

lm_3b_ml <- lmer(value ~ (1|oper) + (1|part) + (1|(oper:part)), data = df_3, REML = FALSE)
summary(lm_3b_ml)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: value ~ (1 | oper) + (1 | part) + (1 | (oper:part))
##    Data: df_3
```

```
##
##      AIC      BIC   logLik deviance df.resid
##   -662.0   -651.5    336.0   -672.0       55
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.33883 -0.53319 -0.02849  0.38850  2.45116
##
## Random effects:
##  Groups      Name        Variance  Std.Dev.
##  (oper:part) (Intercept) 5.360e-19 7.321e-10
##  part        (Intercept) 1.722e-06 1.312e-03
##  oper        (Intercept) 1.807e-07 4.251e-04
##  Residual                4.329e-07 6.580e-04
## Number of obs: 60, groups:  (oper:part), 30; part, 10; oper, 3
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 0.9561333  0.0004895    1953
```

```
sigma_epsilon_3b_ml <- 0.0000004329
sigma_oper_part_3b_ml <- 0.00000000000000000536
# sigma_oper_part_3b_ml <- 0
sigma_part_3b_ml <- 0.000001722
sigma_oper_3b_ml <- 0.0000001807

## REML

lm_3b_reml <- lmer(value ~ (1|oper) + (1|part) + (1|(oper:part)), data = df_3, REML = TRUE)
summary(lm_3b_reml)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: value ~ (1 | oper) + (1 | part) + (1 | (oper:part))
##    Data: df_3
##
## REML criterion at convergence: -658.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.31995 -0.52148 -0.01125  0.37751  2.45240
##
## Random effects:
##  Groups      Name        Variance  Std.Dev.
##  (oper:part) (Intercept) 0.000e+00 0.0000000
##  part        (Intercept) 1.871e-06 0.0013679
##  oper        (Intercept) 2.092e-07 0.0004574
##  Residual                4.326e-07 0.0006578
## Number of obs: 60, groups:  (oper:part), 30; part, 10; oper, 3
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 0.9561333  0.0005139    1861
```

```
sigma_epsilon_3b_reml <- 0.0000004326
sigma_oper_part_3b_reml <- 0
```

```r
# sigma_oper_part_3b_reml <- 0
sigma_part_3b_reml <- 0.000001871
sigma_oper_3b_reml <- 0.0000002092

## Comparisons

# The sigma_square_epsilon are extremely small in all cases, however largest with ANOVA and approximate
# The estimates for sigma_square_interact is approximately 0 in all cases
# And again with variance estimates for oper and part, they are all extremely small and approximately 0
# Again, we know the ANOVA and ML are biased, while REML is unbiased

##### Part c #####

# According to the ANOVA table from Part a, there are no significant sources of variation; all p-values



########## Problem 4 ##########



## Load in the data

data_4 <- read.table("~/Documents/Rice_University/Spring_2018/STAT616/HW04/bloodpressure.txt", header =

df_4 <- setNames(data.frame(matrix(NA, ncol = 2, nrow = 20)), c("treatment", "decrease"))
df_4$treatment <- c(rep("A", times = 10), rep("B", 10))
df_4$decrease <- c(data_4[, 3], data_4[, 6])

##### Part a #####

lm_4a <- lm(decrease ~ treatment, data = df_4)
anova(lm_4a)
```

```
## Analysis of Variance Table
##
## Response: decrease
##           Df Sum Sq Mean Sq F value Pr(>F)
## treatment  1  245.0  245.00  1.7492 0.2025
## Residuals 18 2521.2  140.07
```

```r
summary(lm_4a)
```

```
##
## Call:
## lm(formula = decrease ~ treatment, data = df_4)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -18.2   -8.7    1.8   10.3   16.8
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   33.200      3.743   8.871 5.46e-08 ***
```

```
## treatmentB     7.000      5.293   1.323    0.203
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.83 on 18 degrees of freedom
## Multiple R-squared:  0.08857,    Adjusted R-squared:  0.03793
## F-statistic: 1.749 on 1 and 18 DF,  p-value: 0.2025
```

```
# Due the p-value of 0.2025, I will not reject my null hypothesis that two samples are equal.

##### Part b #####

df_4b <- setNames(data.frame(matrix(NA, ncol = 3, nrow = 20)), c("treatment", "decrease", "initial"))
df_4b$treatment <- c(rep("A", times = 10), rep("B", 10))
df_4b$decrease <- c(data_4[, 3], data_4[, 6])
df_4b$initial <- c(data_4[, 2], data_4[, 5])

lm_4b <- lm(decrease ~ treatment + initial, data = df_4b)
anova(lm_4b)
```

```
## Analysis of Variance Table
##
## Response: decrease
##           Df Sum Sq Mean Sq F value  Pr(>F)
## treatment  1  245.0  245.00  4.0393 0.06060 .
## initial    1 1490.1 1490.08 24.5668 0.00012 ***
## Residuals 17 1031.1   60.65
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(lm_4b)
```

```
##
## Call:
## lm(formula = decrease ~ treatment + initial, data = df_4b)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.1497  -4.4518  -0.0662   4.7338  12.2199
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -28.5373    12.6970  -2.248  0.03817 *
## treatmentB    9.7427     3.5266   2.763  0.01331 *
## initial       0.5175     0.1044   4.956  0.00012 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.788 on 17 degrees of freedom
## Multiple R-squared:  0.6272, Adjusted R-squared:  0.5834
## F-statistic:  14.3 on 2 and 17 DF,  p-value: 0.0002276
```

```
##### Part c #####

# I prefer the results from Part b because the ANOVA table from Part b tells us that the initial weight
```