STAT 222 Spring 2022 HW7

Matthew Zhao

```
pr8.4 = read.table("http://users.stat.umn.edu/~gary/book/fcdae.data/pr8.4", header=T)
pr8.4$Time = factor(pr8.4$temp, labels = c("8hr","13hr"))
pr8.4$Airflow = factor(pr8.4$flow, labels = c("L","H"))
pr8.4$Laser = factor(pr8.4$laser, labels = c("old","new"))
```

Q1 — 5 points

```
library(mosaic)
mean(y ~ 1, data=pr8.4)
##        1
## 0.488125
mean(y ~ Time, data=pr8.4)
##     8hr    13hr
## 0.78125 0.19500
mean(y ~ Airflow, data=pr8.4)
##       L       H
## 0.50125 0.47500
mean(y ~ Laser, data=pr8.4)
##     old     new
## 0.49750 0.47875
mean(y ~ Time+Airflow, data=pr8.4)
##  8hr.L 13hr.L  8hr.H 13hr.H
## 0.7850 0.2175 0.7775 0.1725
mean(y ~ Time+Laser, data=pr8.4)
##  8hr.old 13hr.old  8hr.new 13hr.new
##   0.7975   0.1975   0.7650   0.1925
mean(y ~ Airflow+Laser, data=pr8.4)
##  L.old  H.old  L.new  H.new
## 0.4875 0.5075 0.5150 0.4425
mean(y ~ Time+Airflow+Laser, data=pr8.4)
##  8hr.L.old 13hr.L.old  8hr.H.old 13hr.H.old  8hr.L.new 13hr.L.new  8hr.H.new
##      0.805      0.170      0.790      0.225      0.765      0.265      0.765
## 13hr.H.new
##      0.120
```

$\hat{\alpha}_{13hr} = \bar{y}_{13hr\cdots} - \bar{y}_{\cdots\cdot} = 0.19500 - 0.488125 = -0.293125$

$\hat{\beta\gamma}_{H,new} = \bar{y}_{\cdot H,new\cdot} - \bar{y}_{\cdot H\cdot\cdot} - \bar{y}_{\cdot\cdot new\cdot} + \bar{y}_{\cdot\cdot\cdot\cdot} = 0.4425 - 0.47500 - 0.47875 + 0.488125 = -0.023125$

$\alpha\hat{\beta}\gamma_{13hr,H,new} = \bar{y}_{13hr,H,new\cdot} - \bar{y}_{13hr,H\cdot\cdot} - \bar{y}_{13hr\cdot new\cdot} - \bar{y}_{\cdot H,new\cdot} + \bar{y}_{13hr\cdot\cdot\cdot} + \bar{y}_{\cdot H\cdot\cdot} + \bar{y}_{\cdot\cdot new\cdot} - \bar{y}_{\cdot\cdot\cdot\cdot} = 0.120 - 0.1725 - 0.1925 - 0.4425 + 0.19500 + 0.47500 + 0.47875 - 0.488125 = -0.026875$

Q2 — 5 points

```
contrasts(pr8.4$Time) = contr.sum(2)
contrasts(pr8.4$Airflow) = contr.sum(2)
contrasts(pr8.4$Laser) = contr.sum(2)
lm1 = lm(y ~ Time*Airflow*Laser, data=pr8.4)
lm1$coef
##          (Intercept)                Time1             Airflow1
##             0.488125             0.293125             0.013125
##               Laser1        Time1:Airflow1         Time1:Laser1
##             0.009375            -0.009375             0.006875
##       Airflow1:Laser1 Time1:Airflow1:Laser1
##            -0.023125             0.026875
options(digits=8) # increasing the number of digits in the output
anova(lm1)
## Analysis of Variance Table
##
## Response: y
##                    Df   Sum Sq  Mean Sq   F value        Pr(>F)
## Time                1 1.374756 1.374756 210.48900 0.00000049884 ***
## Airflow             1 0.002756 0.002756   0.42201       0.53414
## Laser               1 0.001406 0.001406   0.21531       0.65500
## Time:Airflow        1 0.001406 0.001406   0.21531       0.65500
## Time:Laser          1 0.000756 0.000756   0.11579       0.74241
## Airflow:Laser       1 0.008556 0.008556   1.31005       0.28547
## Time:Airflow:Laser  1 0.011556 0.011556   1.76938       0.22013
## Residuals           8 0.052250 0.006531
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$SS_{Time} = bcn\sum_{i=1}^{a}(\hat{\alpha}_i)^2 = bcn((\hat{\alpha}_{13hr})^2 + (\hat{\alpha}_{13hr})^2) = 2*2*2*((-0.293125)^2 + (0.293125)^2) = 1.374756$

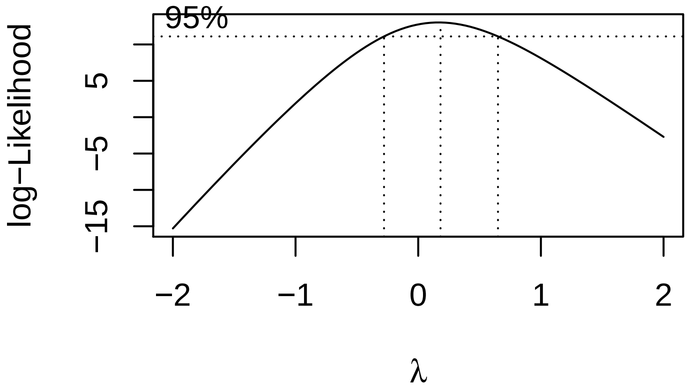
$SS_{Airflow-Laser} = an\sum_{jk}(\hat{\beta\gamma}_{jk})^2 = an((\hat{\beta\gamma}_{L,old})^2 + (\hat{\beta\gamma}_{H,old})^2 + (\hat{\beta\gamma}_{L,new} + (\hat{\beta\gamma}_{H,new})^2)^2) = 2*2*((-0.023125)^2 + (0.023125)^2 + (0.023125)^2 + (-0.023125)^2) = 0.00855625$

$SS_{Time-Airflow-Laser} = n\sum_{ijk}(\alpha\hat{\beta}\gamma_{ijk})^2 = n((\alpha\hat{\beta}\gamma_{8hr,L,old})^2 + (\alpha\hat{\beta}\gamma_{8hr,H,old})^2 + (\alpha\hat{\beta}\gamma_{8hr,L,new})^2 + (\alpha\hat{\beta}\gamma_{8hr,H,new})^2 + (\alpha\hat{\beta}\gamma_{13hr,L,old})^2 + (\alpha\hat{\beta}\gamma_{13hr,H,old})^2 + (\alpha\hat{\beta}\gamma_{13hr,L,new})^2 + (\alpha\hat{\beta}\gamma_{13hr,H,new})^2) = 2*(8*(0.026875)^2) = 0.01155625$

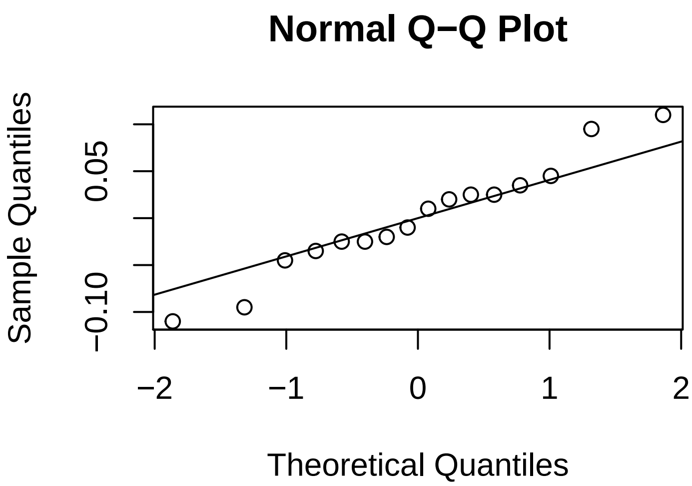
Q3 — 4 points

```
lm1 = lm(y ~ Time*Airflow*Laser, data=pr8.4)
anova(lm1)
## Analysis of Variance Table
##
## Response: y
##                    Df   Sum Sq  Mean Sq   F value        Pr(>F)
## Time                1 1.374756 1.374756 210.48900 0.00000049884 ***
## Airflow             1 0.002756 0.002756   0.42201       0.53414
## Laser               1 0.001406 0.001406   0.21531       0.65500
## Time:Airflow        1 0.001406 0.001406   0.21531       0.65500
## Time:Laser          1 0.000756 0.000756   0.11579       0.74241
## Airflow:Laser       1 0.008556 0.008556   1.31005       0.28547
## Time:Airflow:Laser  1 0.011556 0.011556   1.76938       0.22013
## Residuals           8 0.052250 0.006531
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
lmlog = lm(log(y) ~ Time*Airflow*Laser, data=pr8.4)
anova(lmlog)
## Analysis of Variance Table
##
## Response: log(y)
##                    Df  Sum Sq Mean Sq   F value         Pr(>F)
## Time                1 8.22129 8.22129 323.90613 0.000000093194 ***
## Airflow             1 0.07392 0.07392   2.91249      0.1262844
## Laser               1 0.01988 0.01988   0.78313      0.4019994
## Time:Airflow        1 0.06241 0.06241   2.45868      0.1555134
## Time:Laser          1 0.00344 0.00344   0.13543      0.7224093
## Airflow:Laser       1 0.26998 0.26998  10.63685      0.0115029 *
## Time:Airflow:Laser  1 0.30664 0.30664  12.08115      0.0083712 **
## Residuals           8 0.20305 0.02538
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

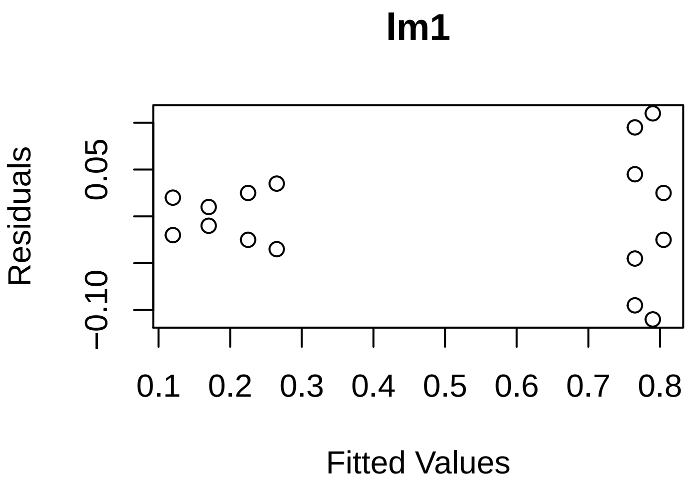
```
boxcox(lm1)
```



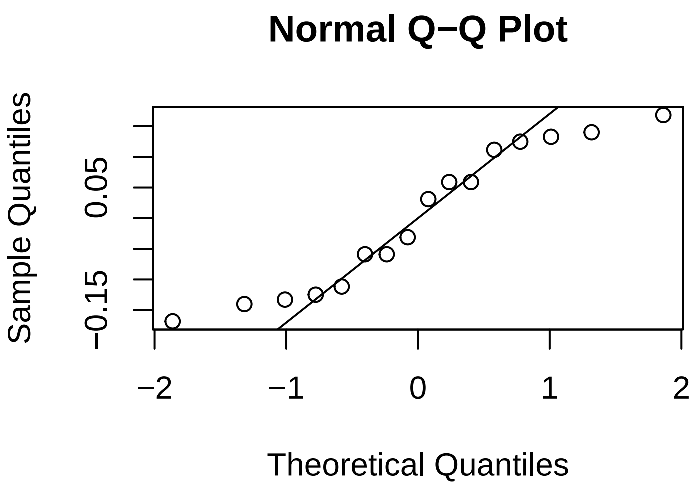

```
qqnorm(lm1$residuals)
qqline(lm1$residuals)
```



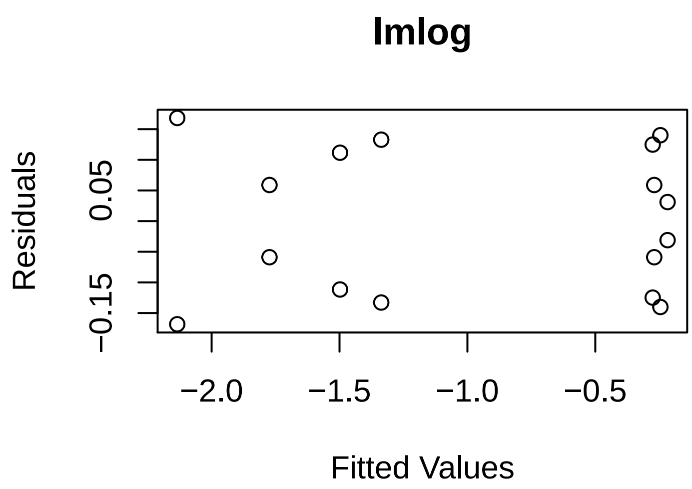

```
plot(lm1$fitted.values, lm1$residuals, xlab = "Fitted Values",
     ylab = "Residuals", main = "lm1")
```




```
qqnorm(lmlog$residuals)
qqline(lmlog$residuals)
```



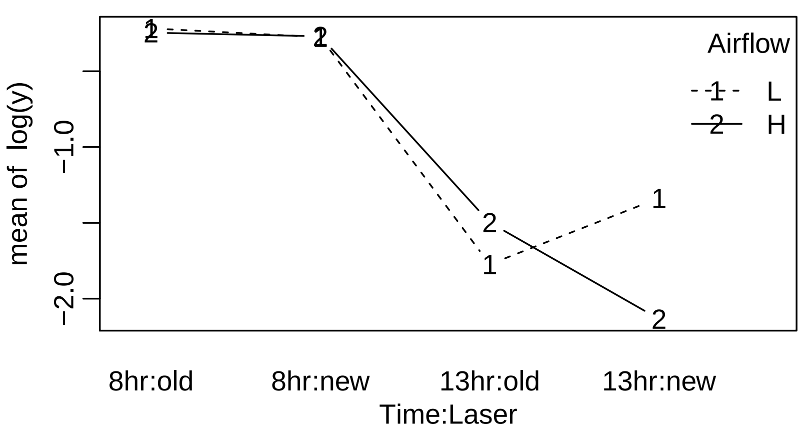

```
plot(lmlog$fitted.values, lmlog$residuals, xlab = "Fitted Values",
     ylab = "Residuals", main = "lmlog")
```



I believe that the lmlog ANOVA table is more trustworthy. This is because for the standard regression, the graph of fitted values against residuals reveals non-constant variance. Additionally, the qq plot shows that the data is light-tailed. The boxcox plot shows that 1 does lie within the 95% confidence interval; however, the interval is extremely wide and the other graphs provide evidence that the data violates our assumptions. Lastly, the fitted values vs residual plot shows constant variance, indicating that the ANOVA results should be more reliable than those of the base linear model.

Q4 — 6 points

```
par(mai=c(.6,.6,.01,.2),mgp=c(2,.6,0)) # for reducing the margin of the plot
with(pr8.4, interaction.plot(Time:Laser,Airflow,log(y), type="b"))
```



a) Using a new vs old laser and changing the furnance airflow both do not result in a significant difference on delamination rate. This can be seen in the graph where there is virtually no change in log of delamination rate from old to new and by the fact that the lines for low and high airflow are on top of each other. There also does not appear to be an interaction between laser and airflow since the lines are nearly the same and parallel.

b) Yes there does appear to be an interaction since the lines are not parallel. This means that the effect of changing the laser on the log of delamination rate changes with different levels of airflow. #### c) The three way interaction would be significant since the interaction between laser and airflow changes with different levels of firing profile time. Specifically, when firing profile time is 8hr, there does not appear to be any interaction between laser and furnance airflow, but when firing profile time is 13hr, there appears to be a significant interaction between laser and furnance airflow, supporting my initial statement.

d) Increasing the firing from 8hr to 13hr always reduces the delamination rate. Higher airflow reduces the delamination rate on average, but is not necessarily the case for each replicate. This is also true for using a new laser except for the case of 13hr firing and low airflow, where the average delamination rate is actually lower for the old laser compared with the new laser. As a result, our conclusions are not consistent.

Q5 — 5 points

```
q <- qtukey(1-0.05,8,2*2*2*(2-1))
mse <- 0.02538
hsd <- q/sqrt(2) * sqrt(mse*((1/2) + (1/2)))
hsd
## [1] 0.63040915
```

```
pr8.4 %>% group_by(Time,Airflow,Laser) %>% summarise(.groups = 'keep',group_means = mean(y))
## # A tibble: 8 x 4
## # Groups:   Time, Airflow, Laser [8]
##   Time  Airflow Laser group_means
##   <fct> <fct>   <fct>       <dbl>
## 1 8hr   L       old         0.805
## 2 8hr   H       old         0.79
## 3 8hr   L       new         0.765
## 4 8hr   H       new         0.765
## 5 13hr  L       new         0.265
## 6 13hr  H       old         0.225
## 7 13hr  L       old         0.17
## 8 13hr  H       new         0.12
```

13hr, high airflow, and new laser, along with 13hr, low airflow, and old laser, are significantly lower than the other log means. Since they are within the HSD of each other, they are not significantly different from each other. This means that we should make more replicates for these two so that we can find which combination is the best.