ch19output_beetles

February 18, 2020

```
########################### Example: Beetles  ############################################
beetles<- read.table(file="~/Desktop/jenn/teaching/stat445545/data/beetle.dat",header = TRUE)
#or you can read from Professor Erik Erhardt's website
#beetles <- read.table("https://statacumen.com/teach/ADA2/notes/ADA2_notes_Ch05_beetles.dat"
#                       , header = TRUE)
beetles
```

```
##    dose insecticide   t1   t2   t3   t4
## 1     1           A 0.31 0.45 0.46 0.43
## 2     1           B 0.82 1.10 0.88 0.72
## 3     1           C 0.43 0.45 0.63 0.76
## 4     1           D 0.45 0.71 0.66 0.62
## 5     2           A 0.36 0.29 0.40 0.23
## 6     2           B 0.92 0.61 0.49 1.24
## 7     2           C 0.44 0.35 0.31 0.40
## 8     2           D 0.56 1.02 0.71 0.38
## 9     3           A 0.22 0.21 0.18 0.23
## 10    3           B 0.30 0.37 0.38 0.29
## 11    3           C 0.23 0.25 0.24 0.22
## 12    3           D 0.30 0.36 0.31 0.33
```

```
# make dose a factor variable and label the levels
beetles$dose <- factor(beetles$dose, labels = c("low","medium","high"))
beetles$dose
```

```
## [1] low    low    low    low    medium medium medium medium high   high
## [11] high   high
## Levels: low medium high
```

```
library(reshape2)
beetles.long <- melt(beetles
                  , id.vars       = c("dose", "insecticide")
                  , variable.name = "number"
                  , value.name    = "hours10"
)
str(beetles.long)
```

```
## 'data.frame':    48 obs. of  4 variables:
##  $ dose       : Factor w/ 3 levels "low","medium",..: 1 1 1 1 2 2 2 2 3 3 ...
##  $ insecticide: Factor w/ 4 levels "A","B","C","D": 1 2 3 4 1 2 3 4 1 2 ...
##  $ number     : Factor w/ 4 levels "t1","t2","t3",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ hours10    : num  0.31 0.82 0.43 0.45 0.36 0.92 0.44 0.56 0.22 0.3 ...
```

```
head(beetles.long)
```

```
##     dose insecticide number hours10
## 1    low           A     t1    0.31
## 2    low           B     t1    0.82
## 3    low           C     t1    0.43
## 4    low           D     t1    0.45
## 5 medium           A     t1    0.36
```

```
## 6 medium          B     t1    0.92
library(plyr)
# Calculate the cell means for each (dose, insecticide) combination
mean(beetles.long[, "hours10"])
```

```
## [1] 0.479375
```

```
beetles.mean    <- ddply(beetles.long, .(), summarise, m = mean(hours10))
beetles.mean
```

```
##    .id        m
## 1 <NA> 0.479375
```

```
beetles.mean.d  <- ddply(beetles.long, .(dose), summarise, m = mean(hours10))
beetles.mean.d
```

```
##     dose        m
## 1    low 0.617500
## 2 medium 0.544375
## 3   high 0.276250
```

```
beetles.mean.i  <- ddply(beetles.long, .(insecticide), summarise, m = mean(hours10))
beetles.mean.i
```

```
##   insecticide         m
## 1           A 0.3141667
## 2           B 0.6766667
## 3           C 0.3925000
## 4           D 0.5341667
```

```
beetles.mean.di <- ddply(beetles.long, .(dose,insecticide), summarise, m = mean(hours10))
beetles.mean.di
```

```
##       dose insecticide      m
## 1      low           A 0.4125
## 2      low           B 0.8800
## 3      low           C 0.5675
## 4      low           D 0.6100
## 5   medium           A 0.3200
## 6   medium           B 0.8150
## 7   medium           C 0.3750
## 8   medium           D 0.6675
## 9     high           A 0.2100
## 10    high           B 0.3350
## 11    high           C 0.2350
## 12    high           D 0.3250
```

```
# Interaction plots, ggplot
library(ggplot2)
p <- ggplot(beetles.long, aes(x = dose, y = hours10, colour = insecticide, shape = insecticide))
p <- p + geom_hline(aes(yintercept = 0), colour = "black"
                    , linetype = "solid", size = 0.2, alpha = 0.3)
p <- p + geom_boxplot(alpha = 0.25, outlier.size=0.1)
p <- p + geom_point(alpha = 0.5, position=position_dodge(width=0.75))
p <- p + geom_point(data = beetles.mean.di, aes(y = m), size = 4)
p <- p + geom_line(data = beetles.mean.di, aes(y = m, group = insecticide), size = 1.5)
p <- p + labs(title = "Beetles interaction plot, insecticide by dose")
print(p)
```
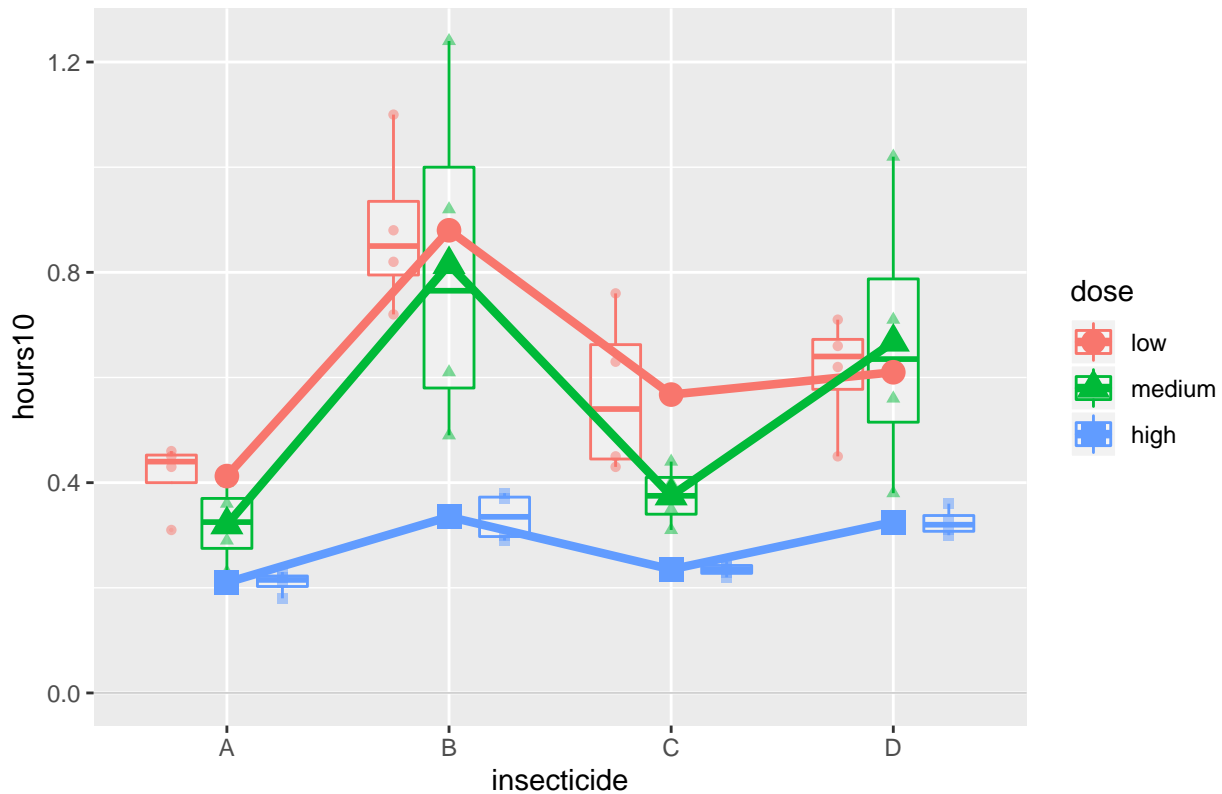
Beetles interaction plot, insecticide by dose
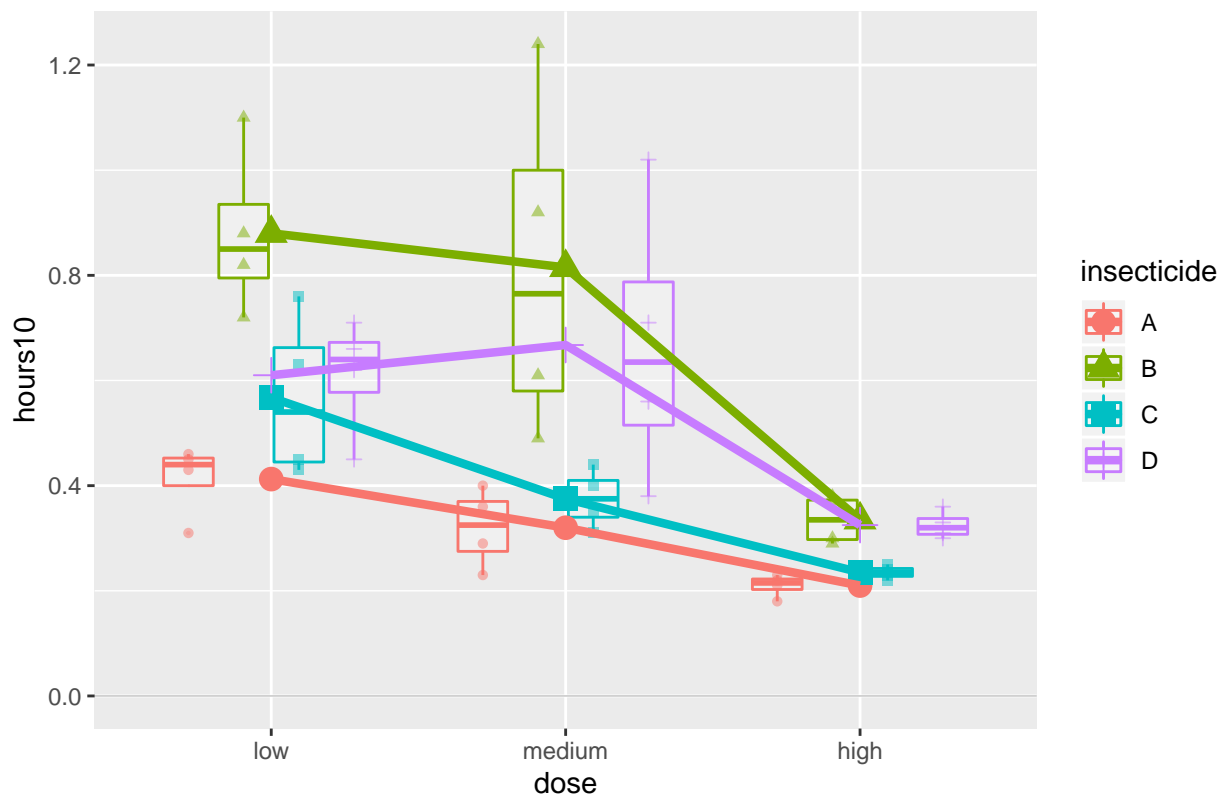

```
#dev.copy(jpeg,filename="~/Desktop/jenn/teaching/ADA2/lecture notes/plots/ch05plot3.jpg")
#dev.off()

p <- ggplot(beetles.long, aes(x = insecticide, y = hours10, colour = dose, shape = dose))
p <- p + geom_hline(aes(yintercept = 0), colour = "black"
                    , linetype = "solid", size = 0.2, alpha = 0.3)
p <- p + geom_boxplot(alpha = 0.25, outlier.size=0.1)
p <- p + geom_point(alpha = 0.5, position=position_dodge(width=0.75))
p <- p + geom_point(data = beetles.mean.di, aes(y = m), size = 4)
p <- p + geom_line(data = beetles.mean.di, aes(y = m, group = dose), size = 1.5)
p <- p + labs(title = "Beetles interaction plot, dose by insecticide")
print(p)
```

Beetles interaction plot, dose by insecticide


```
dev.copy(jpeg,filename="~/Desktop/jenn/teaching/ADA2/lecture notes/plots/ch05plot4.jpg")
```

```
## jpeg
##    3
```

```
dev.off()
```

```
## pdf
##   2
```

```
##fit ANOVA model
myfit <- aov(hours10~ dose*insecticide, data=beetles.long)
summary(myfit) #ANOVA table
```

```
##                  Df Sum Sq Mean Sq F value   Pr(>F)
## dose              2 1.0330  0.5165  23.222 3.33e-07 ***
## insecticide       3 0.9212  0.3071  13.806 3.78e-06 ***
## dose:insecticide  6 0.2501  0.0417   1.874    0.112
## Residuals        36 0.8007  0.0222
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##notice that interaction effect is not significant (p value =0.112)
#report row, column and cell means
print(model.tables(myfit,"means"),digits=3)  #you can also find cell mean, marginal mean etc using myfi
```

```
## Tables of means
## Grand mean
##
## 0.479375
```

```
## 
##   dose
## dose
##    low medium   high
##   0.618  0.544  0.276
## 
##   insecticide
## insecticide
##     A     B     C     D
## 0.314 0.677 0.393 0.534
## 
##   dose:insecticide
##         insecticide
## dose      A     B     C     D
##   low     0.413 0.880 0.568 0.610
##   medium  0.320 0.815 0.375 0.668
##   high    0.210 0.335 0.235 0.325
```

checking assumptions

```r
# mean vs sd plot
library(plyr)
# means and standard deviations for each dose/interaction cell
beetles.meansd.di <- ddply(beetles.long, .(dose,insecticide), summarise
                           , m = mean(hours10), s = sd(hours10))
beetles.meansd.di
```

```
##        dose insecticide      m          s
## 1    low           A 0.4125 0.06946222
## 2    low           B 0.8800 0.16083117
## 3    low           C 0.5675 0.15671099
## 4    low           D 0.6100 0.11284207
## 5  medium          A 0.3200 0.07527727
## 6  medium          B 0.8150 0.33630343
## 7  medium          C 0.3750 0.05686241
## 8  medium          D 0.6675 0.27097048
## 9    high          A 0.2100 0.02160247
## 10   high          B 0.3350 0.04654747
## 11   high          C 0.2350 0.01290994
## 12   high          D 0.3250 0.02645751
```

```r
p <- ggplot(beetles.meansd.di, aes(x = m, y = s, shape = dose, colour = insecticide))
p <- p + geom_point(size=4)
p <- p + labs(title = "Beetles standard deviation vs mean")
print(p)
```

Beetles standard deviation vs mean
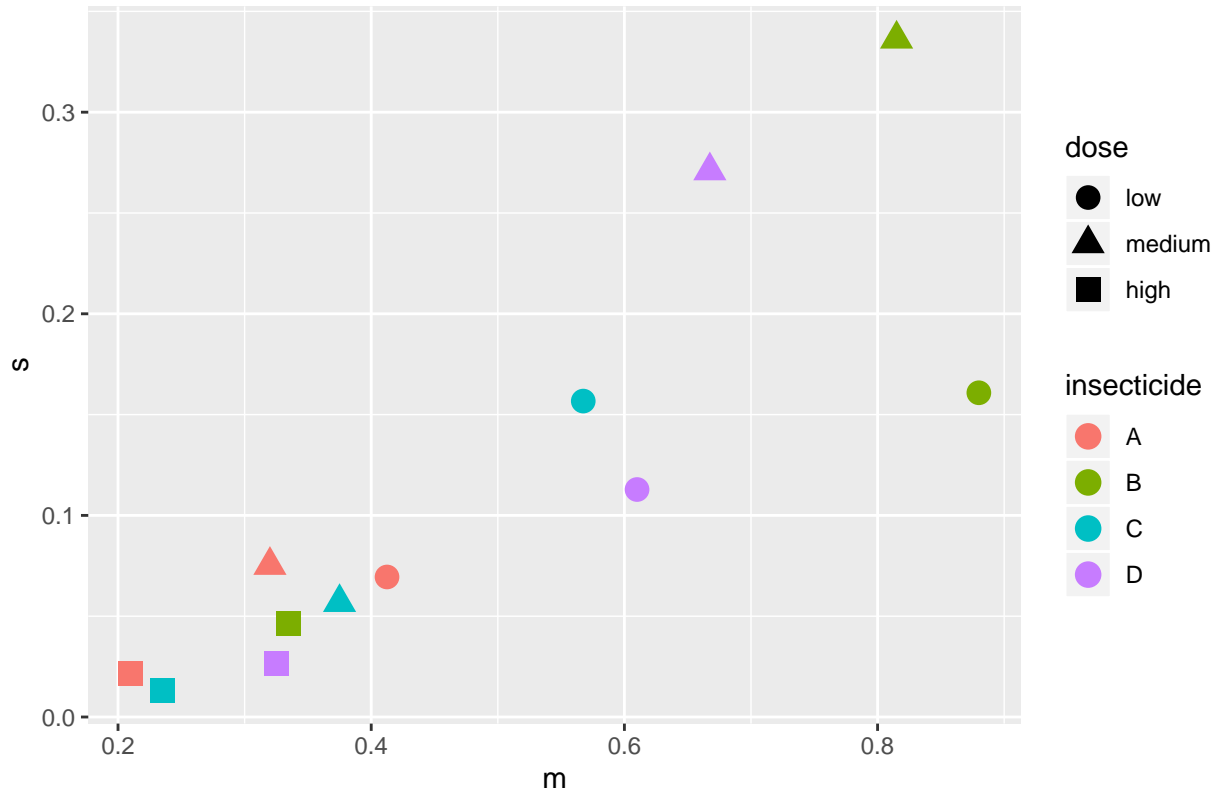

```r
dev.copy(jpeg,filename="~/Desktop/jenn/teaching/stat445545/notes/plots/ch05sdvsmean.jpg")
```

```
## jpeg
##    3
```

```r
dev.off()
```

```
## pdf
##    2
```

```r
# plot diagnistics
par(mfrow=c(2,2))

#residual vs fitted value
plot(myfit$fitted, myfit$residuals, main="Residuals vs Fitted Values")
abline(h = 0, col = "gray75")

plot(beetles.long$dose, myfit$residuals, main="Residuals vs dose")
# horizontal line at zero
abline(h = 0, col = "gray75")

plot(beetles.long$insecticide, myfit$residuals, main="Residuals vs insecticide")
# horizontal line at zero
abline(h = 0, col = "gray75")

# Normality of Residuals
library(car)
```
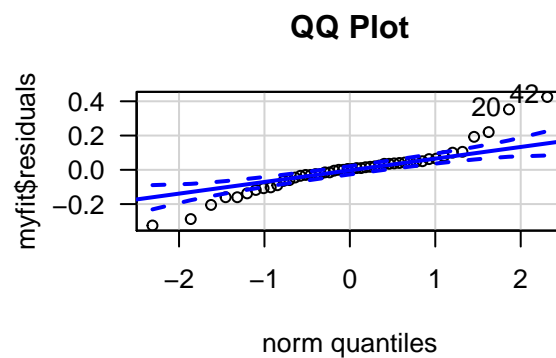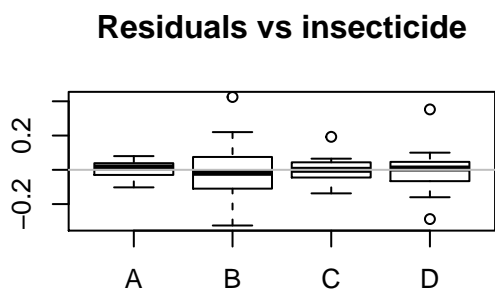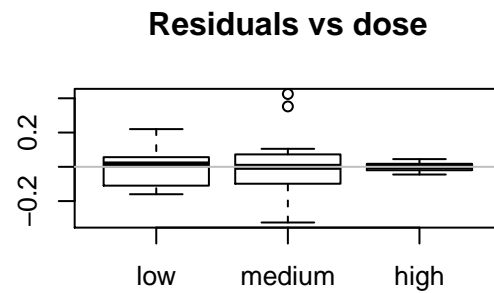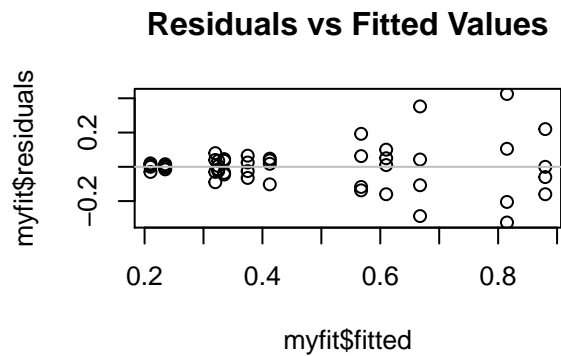
```
## Loading required package: carData
```

```
qqPlot(myfit$residuals, las = 1, main="QQ Plot")
```

Residuals vs Fitted Values

Residuals vs dose

Residuals vs insecticide

QQ Plot


```
## [1] 42 20
```

```
dev.copy(jpeg,filename="~/Desktop/jenn/teaching/stat445545/notes/plots/ch05diagfull.jpg")
```

```
## jpeg
##     3
```
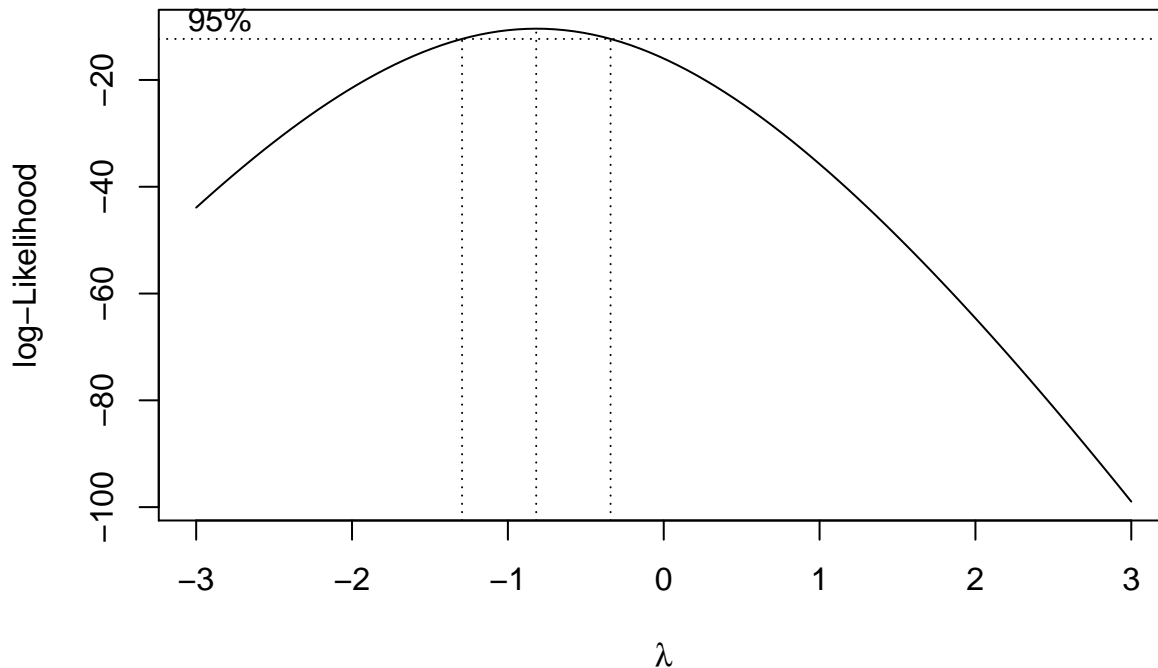
```
dev.off()
```

```
## pdf
##     2
```

```
#transformation
par(mfrow=c(1,1))
library(MASS)
boxcox(myfit, lambda = seq(-3, 3, length = 10),plotit = TRUE)
```

```
dev.copy(jpeg,filename="~/Desktop/jenn/teaching/stat445545/notes/plots/ch05boxcox.jpg")
```

```
## jpeg
##    3
```

```
dev.off()
```

```
## pdf
##    2
```

```
##transformations on $y$
myfit2 <- aov(1/hours10 ~ dose*insecticide, data = beetles.long)
summary(myfit2)
```

```
##                   Df Sum Sq Mean Sq F value   Pr(>F)
## dose               2  34.88  17.439   72.64 2.31e-13 ***
## insecticide        3  20.41   6.805   28.34 1.38e-09 ***
## dose:insecticide   6   1.57   0.262    1.09    0.387
## Residuals         36   8.64   0.240
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#plot diagnostics
par(mfrow=c(2,3))

#residual vs fitted value
plot(myfit2$fitted, myfit2$residuals, main="Residuals vs Fitted Values for transformed data")
abline(h = 0, col = "gray75")

plot(myfit2$fitted, rstudent(myfit2), main="Studentized Deleted Residuals vs Fitted Values for transform
abline(h = 0, col = "gray75")

plot(beetles.long$dose, myfit2$residuals, main="Residuals vs dose")
# horizontal line at zero
abline(h = 0, col = "gray75")
```

```
plot(beetles.long$insecticide, myfit2$residuals, main="Residuals vs insecticide")
# horizontal line at zero
abline(h = 0, col = "gray75")

# Normality of Residuals
library(car)
qqPlot(myfit2$residuals, las = 1, main="QQ Plot")
```

[1] 41 44

```
dev.copy(jpeg,filename="~/Desktop/jenn/teaching/stat445545/notes/plots/ch05diagtrans.jpg")
```

jpeg
3

```
dev.off()
```

pdf
2

##refit model without interaction term

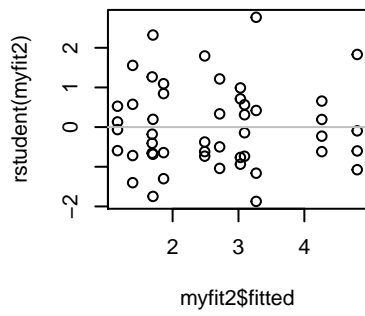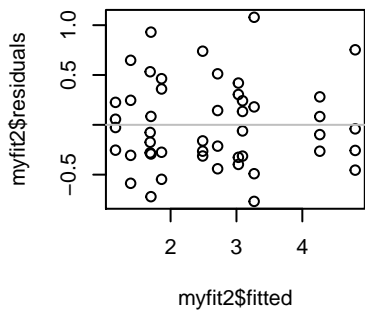```
myfit3 <- aov(1/hours10 ~ dose+insecticide, data = beetles.long)
summary(myfit3)
```

```
Df Sum Sq Mean Sq F value   Pr(>F)
dose   2 34.88 17.439   71.71 2.86e-14 ***
insecticide   3 20.41   6.805   27.98 4.19e-10 ***
Residuals   42 10.21   0.243

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
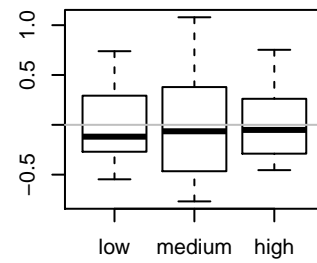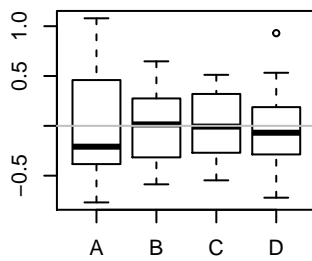
```
#plot diagnostics
par(mfrow=c(1,2))
```

**luals vs Fitted Values for transforeted Residuals vs Fitted Values f**     **Residuals vs dose**
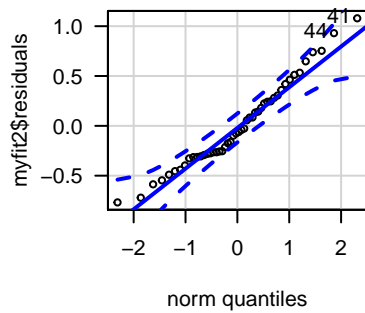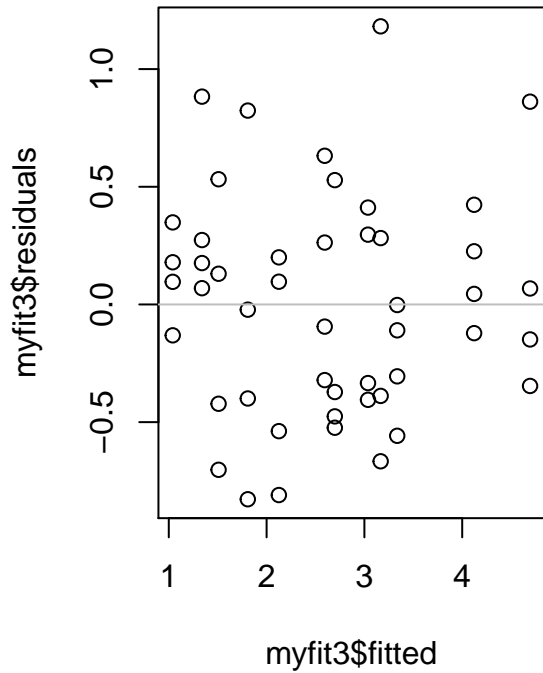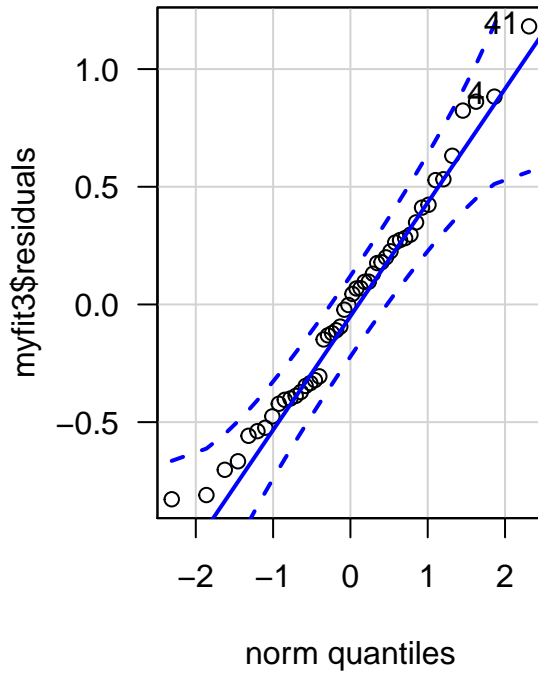
**Residuals vs insecticide**     **QQ Plot**

```
#residual vs fitted value
plot(myfit3$fitted, myfit3$residuals, main="Residuals vs Fitted Values for transformed reduced model")
abline(h = 0, col = "gray75")

# Normality of Residuals
library(car)
qqPlot(myfit3$residuals, las = 1, main="QQ Plot")
```

10

s vs Fitted Values for transformed re   **QQ Plot**


```
## [1] 41  4
dev.copy(jpeg,filename="~/Desktop/jenn/teaching/stat445545/notes/plots/ch05diagreduce.jpg")

## jpeg
##    3
dev.off()

## pdf
##    2
# Testing multiple factors is of interest here.
#multiple comparisons
#Tukey's multiple comparison for factor Dose
beetleT<-TukeyHSD(myfit3,"dose",conf.level=.95)
beetleT

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = 1/hours10 ~ dose + insecticide, data = beetles.long)
##
## $dose
##                  diff        lwr       upr      p adj
## medium-low  0.4686413 0.04505584 0.8922267 0.0271587
## high-low    1.9964249 1.57283950 2.4200103 0.0000000
## high-medium 1.5277837 1.10419824 1.9513691 0.0000000

plot(beetleT, sub="Tukey Honest Significant Differences")

#Tukey's multiple comparison for factor Insecticide
```

```
beetleT2<-TukeyHSD(myfit3,"insecticide",conf.level=.95)
beetleT2
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = 1/hours10 ~ dose + insecticide, data = beetles.long)
##
## $insecticide
##            diff         lwr         upr       p adj
## B-A -1.6574024 -2.1959343 -1.11887050 0.0000000
## C-A -0.5721354 -1.1106673 -0.03360355 0.0335202
## D-A -1.3583383 -1.8968702 -0.81980640 0.0000002
## C-B  1.0852669  0.5467351  1.62379883 0.0000172
## D-B  0.2990641 -0.2394678  0.83759598 0.4550931
## D-C -0.7862029 -1.3247347 -0.24767096 0.0018399
```

```
plot(beetleT2, sub="Tukey Honest Significant Differences")
```

```
#or you can use the default to derive all the comparisons, however, if you
#want to do simultaneous comparisons for both factors Dose and Insecticide,
#confidence coefficient is less than 0.95, at least 90% by using Bonferroni inequality
beetleT3<-TukeyHSD(myfit3,conf.level=.95)
beetleT3
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = 1/hours10 ~ dose + insecticide, data = beetles.long)
##
## $dose
##                   diff        lwr       upr       p adj
## medium-low  0.4686413 0.04505584 0.8922267 0.0271587
## high-low    1.9964249 1.57283950 2.4200103 0.0000000
## high-medium 1.5277837 1.10419824 1.9513691 0.0000000
##
## $insecticide
##            diff         lwr         upr       p adj
## B-A -1.6574024 -2.1959343 -1.11887050 0.0000000
## C-A -0.5721354 -1.1106673 -0.03360355 0.0335202
## D-A -1.3583383 -1.8968702 -0.81980640 0.0000002
## C-B  1.0852669  0.5467351  1.62379883 0.0000172
## D-B  0.2990641 -0.2394678  0.83759598 0.4550931
## D-C -0.7862029 -1.3247347 -0.24767096 0.0018399
```
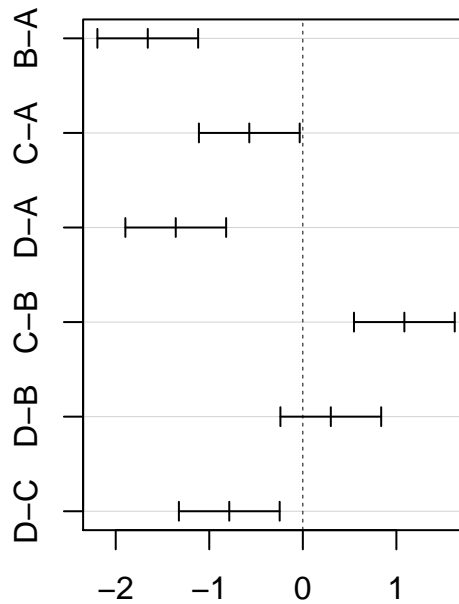
```
plot(beetleT3, sub="Tukey Honest Significant Differences")
```

95% family-wise confidence leve   ## 95% family-wise confidence leve
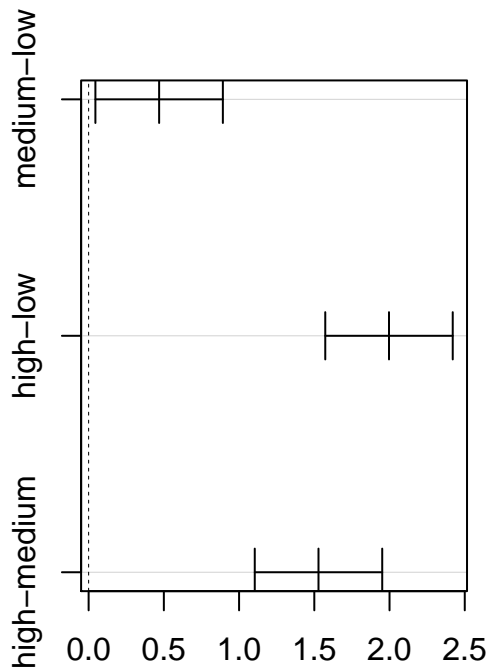
Differences in mean levels of dose
Tukey Honest Significant Differences

Differences in mean levels of insecticide
Tukey Honest Significant Differences

```
beetles.mean.d2  <- ddply(beetles.long, .(dose), summarise, m = mean(1/hours10))
beetles.mean.d2
```

```
##     dose        m
## 1    low 1.800688
## 2 medium 2.269329
## 3   high 3.797112
```

```
beetles.mean.i2  <- ddply(beetles.long, .(insecticide), summarise, m = mean(1/hours10))
beetles.mean.i2
```

```
##   insecticide        m
## 1           A 3.519345
## 2           B 1.861943
## 3           C 2.947210
## 4           D 2.161007
```

```
beetles.mean.di2 <- ddply(beetles.long, .(dose,insecticide), summarise, m = mean(1/hours10))
beetles.mean.di
```

```
##      dose insecticide        m
## 1     low           A 2.486881
## 2     low           B 1.163464
## 3     low           C 1.862724
## 4     low           D 1.689682
## 5  medium           A 3.268470
## 6  medium           B 1.393392
## 7  medium           C 2.713919
## 8  medium           D 1.701534
## 9    high           A 4.802685
```

```
## 10    high          B 3.028973
## 11    high          C 4.264987
## 12    high          D 3.091805
```

```
#with family confidence coefficient at least 90%, we conclude that
#(1)the population mean dying rate for the high dose
#(averaged over insecticides) is significantly faster than the population mean dying rate
#for the low and medium doses (averaged over insecticides).
#(2) the mean dying rate of the two lower doses are slightly significantly different from each other.
#(3) the population mean dying rate for the Insecticide A
# (averaged over doses) is slightly significantly different from the population mean dying rate for
#the C insecticide #(averaged over doses).
#(4) Both insectides A and C has significant faster dying rate than those of insectides B and D,
#regardless of the doses.
#(5) The population mean dying rate for the Insecticide B
#(averaged over doses) is not significantly different from the population mean dying rate
#for the D insecticide (averaged over doses).
```