Homework 10

2024-12-04

1.

a.

```r
require(faraway)
```

```
## Loading required package: faraway
```

```r
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```r
require(dplyr)
```

```
## Loading required package: dplyr
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
data(rats)

rats %>%
  ggplot(aes(x = poison, y = time)) +
  geom_point() +
  stat_summary(fun.y = "mean", geom = "line", aes(group = treat, color = treat)) +
  theme(legend.position = "top", legend.direction = "horizontal")
```
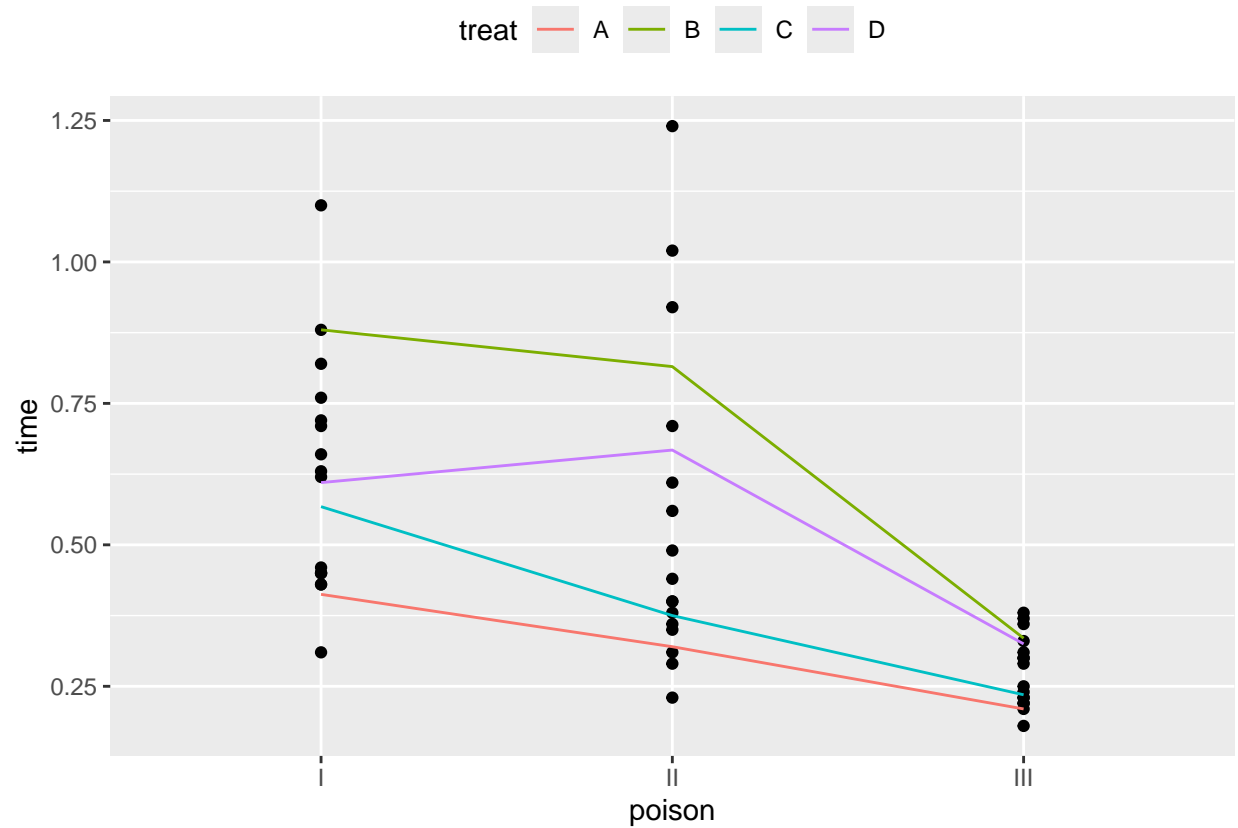
```
## Warning: The 'fun.y' argument of 'stat_summary()' is deprecated as of ggplot2 3.3.0.
## i Please use the 'fun' argument instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```
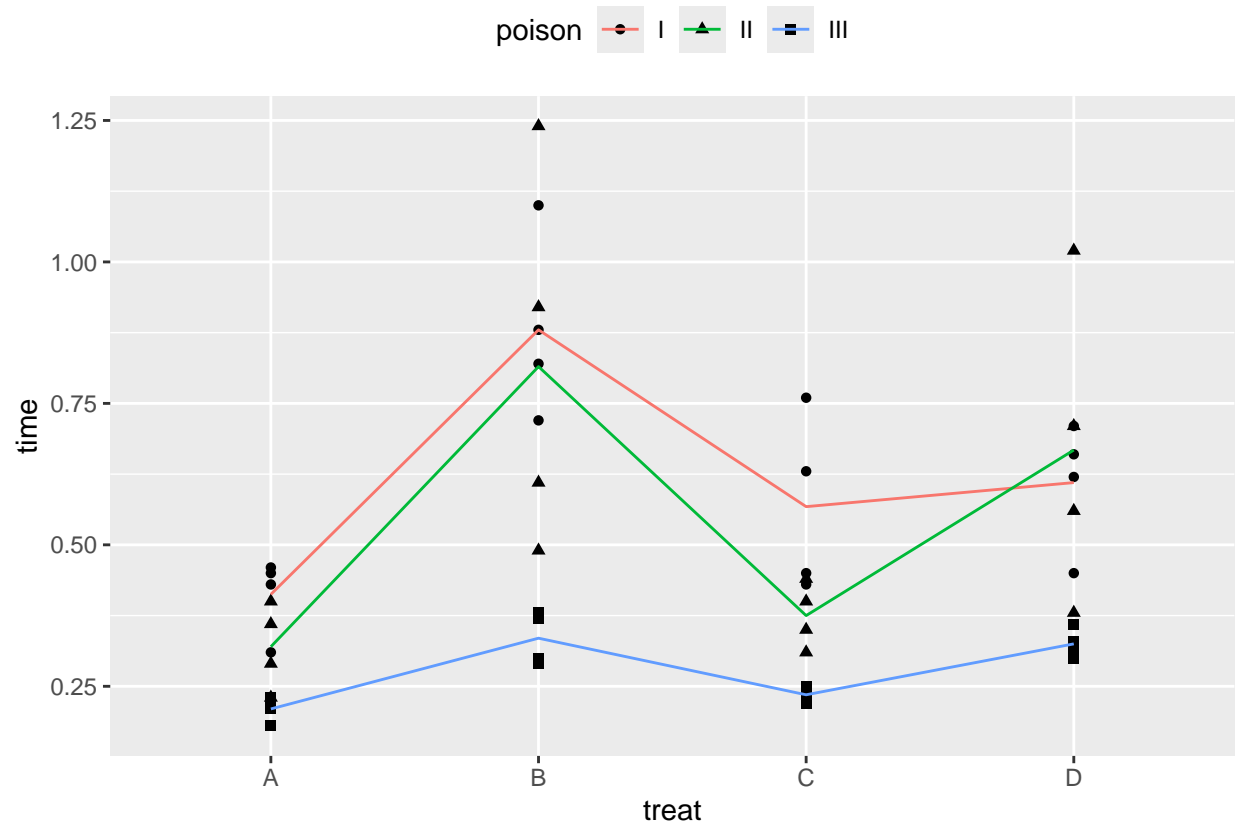
```
rats %>%
  ggplot(aes(x = treat, y = time, shape = poison)) +
  geom_point() +
  stat_summary(fun.y = "mean", geom = "line",
               aes(group = poison, color = poison)) +
  theme(legend.position = "top", legend.direction = "horizontal")
```

b. The interaction effect is not significant.

```
lmod <- lm(time ~ poison*treat, rats)
anova(lmod)
```

```
## Analysis of Variance Table
##
## Response: time
##              Df  Sum Sq Mean Sq F value    Pr(>F)
## poison        2 1.03301 0.51651 23.2217 3.331e-07 ***
## treat         3 0.92121 0.30707 13.8056 3.777e-06 ***
## poison:treat  6 0.25014 0.04169  1.8743    0.1123
## Residuals    36 0.80073 0.02224
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

c. There does appear to be statistical significance between poisons and between treatments.

```
anova(lm(time ~ treat + poison, rats))
```

```
## Analysis of Variance Table
##
## Response: time
##            Df  Sum Sq Mean Sq F value    Pr(>F)
```

```
## treat      3 0.92121 0.30707  12.273 6.697e-06 ***
## poison     2 1.03301 0.51651  20.643 5.704e-07 ***
## Residuals 42 1.05086 0.02502
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

    d. Diagnostics have not been met and give troublesome results. The Q-Q plot has long-tailed errors, and the plot of fitted values against residuals shows some concerning symmetry.
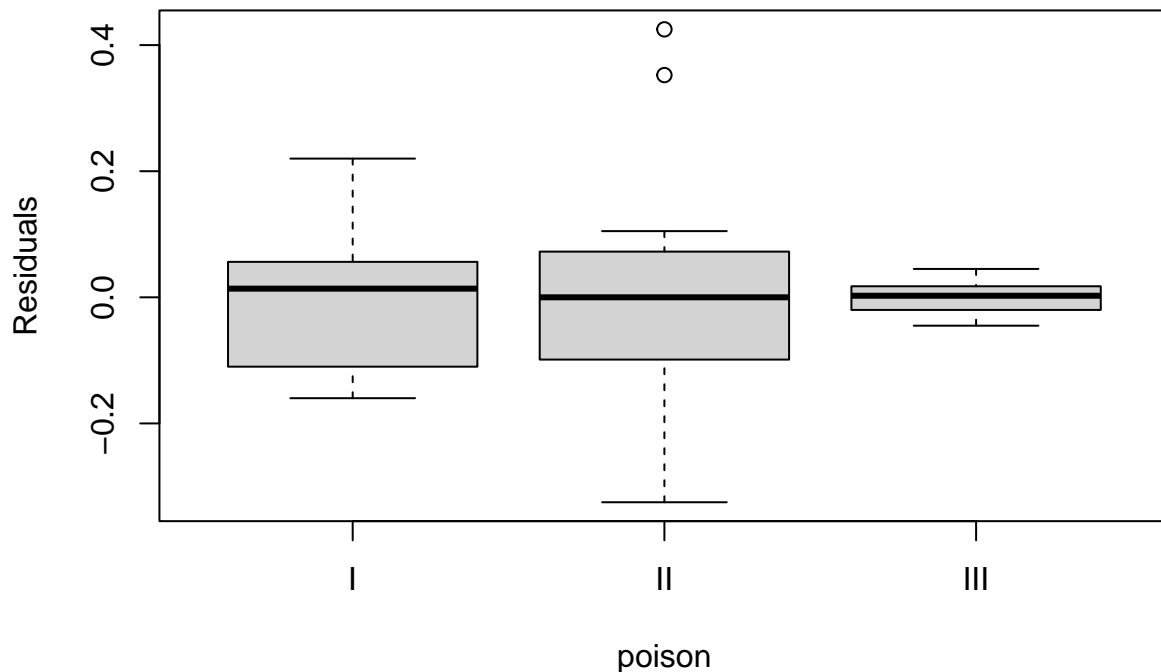
```r
qqnorm(residuals(lmod), main="")
qqline(residuals(lmod)) ## long-tailed errors
```
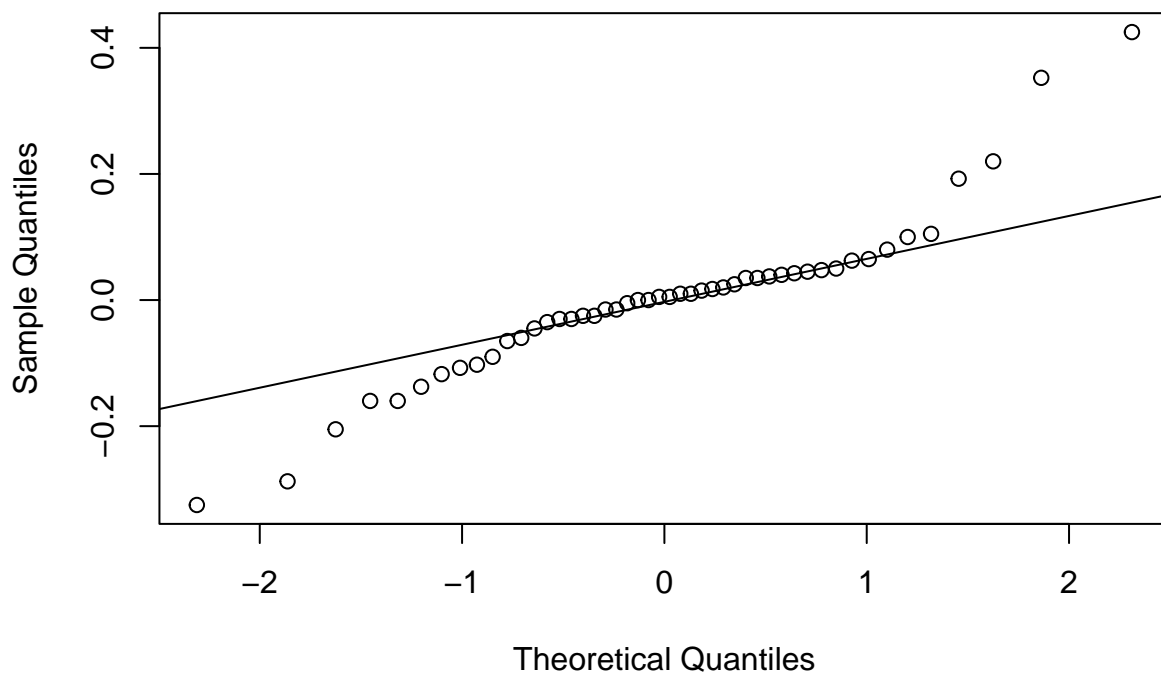


```r
plot(fitted(lmod), residuals(lmod), xlab = "Fitted",
     ylab = "Residuals") ## symmetry
abline(h=0)
```
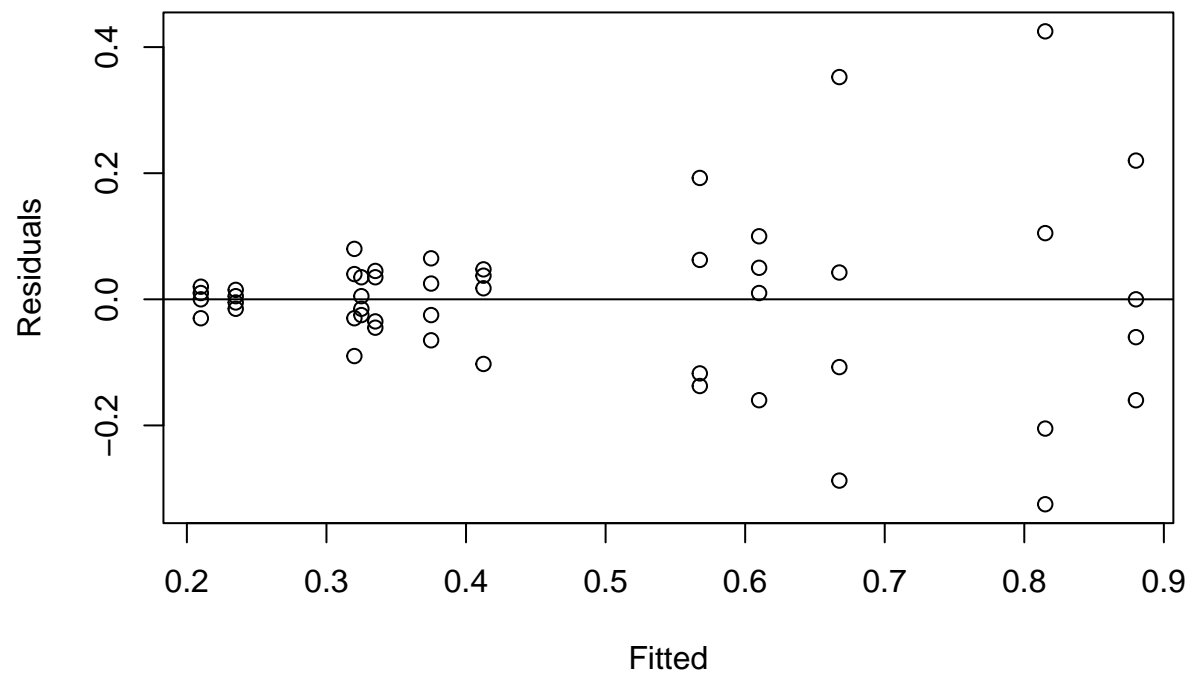
```
plot(residuals(lmod) ~ poison, rats, ylab = "Residuals")
```

```
## Test for difference in variance
ratse <- rats[(1:48),]
ratse$res <- sqrt(abs(residuals(lmod))[(1:48)])
vmod <- lm(res ~ poison + treat, ratse)
anova(vmod)
```

```
## Analysis of Variance Table
##
## Response: res
##           Df  Sum Sq  Mean Sq F value    Pr(>F)
## poison     2 0.35826 0.179129 12.7511 4.705e-05 ***
## treat      3 0.13525 0.045083  3.2092   0.03253 *
## Residuals 42 0.59002 0.014048
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

e. Looking at the plot from question 1 and the model without interaction, it is hard to say that one
treatment is *clearly* superior to the others. Treatment B appears to be the most effective, but treatment
D is close. And for poison III, they are about the same. I would say that treatment B appears to be
the best treatment but that it is not clearly superior to D.

```
rats %>%
  ggplot(aes(x = poison, y = time)) +
  geom_point() +
  stat_summary(fun.y = "mean", geom = "line", aes(group = treat, color = treat)) +
  theme(legend.position = "top", legend.direction = "horizontal")
```
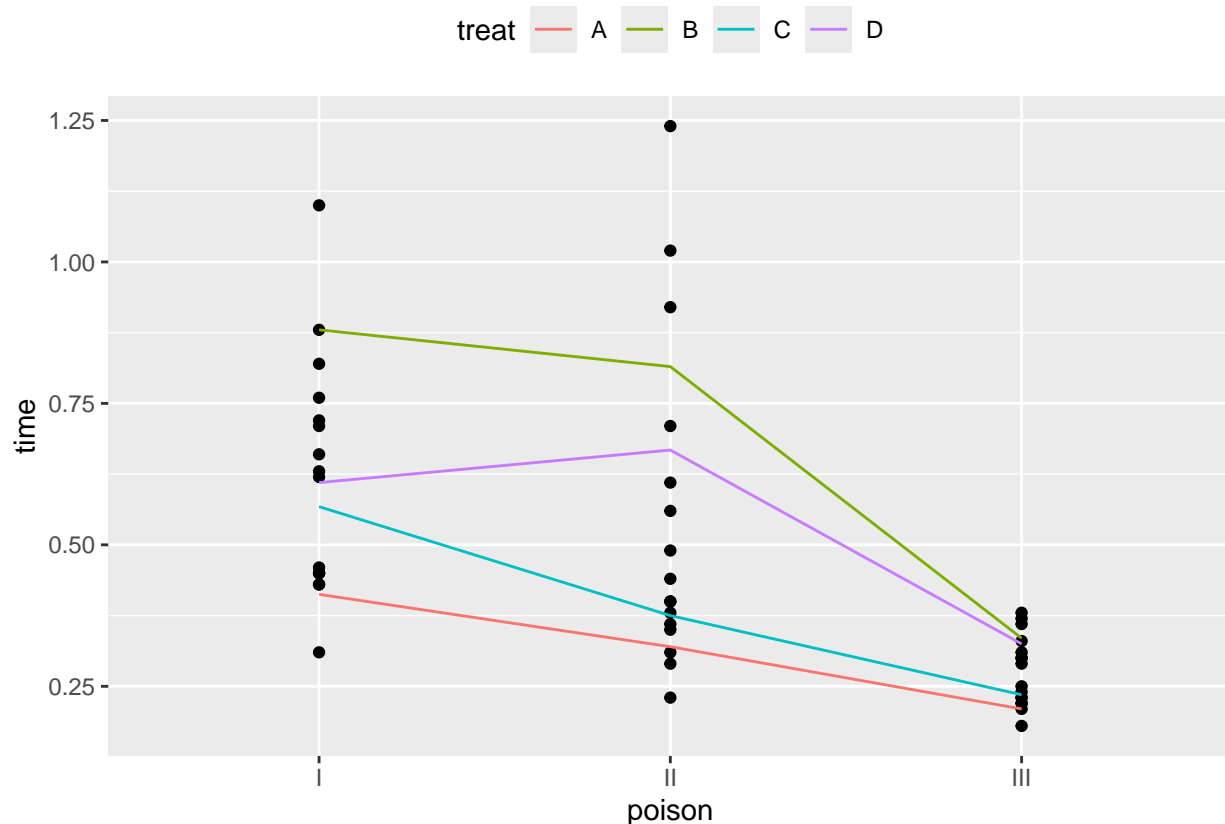
```r
summary(lm(time ~ treat + poison, rats))
```

```
##
## Call:
## lm(formula = time ~ treat + poison, data = rats)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.25167 -0.09625 -0.01490  0.06177  0.49833
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.45229    0.05592   8.088 4.22e-10 ***
## treatB       0.36250    0.06458   5.614 1.43e-06 ***
## treatC       0.07833    0.06458   1.213  0.23189
## treatD       0.22000    0.06458   3.407  0.00146 **
## poisonII    -0.07313    0.05592  -1.308  0.19813
## poisonIII   -0.34125    0.05592  -6.102 2.83e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1582 on 42 degrees of freedom
## Multiple R-squared:  0.6503, Adjusted R-squared:  0.6087
## F-statistic: 15.62 on 5 and 42 DF,  p-value: 1.123e-08
```

f. We find lambda to be -0.82828 and after transforming the data see better results. We see higher

significance in poison II and treatments C and D. We also notice that treatment B and D had large changes to their coefficients, as did poison III.

```
require(MASS)
```

```
## Loading required package: MASS
```
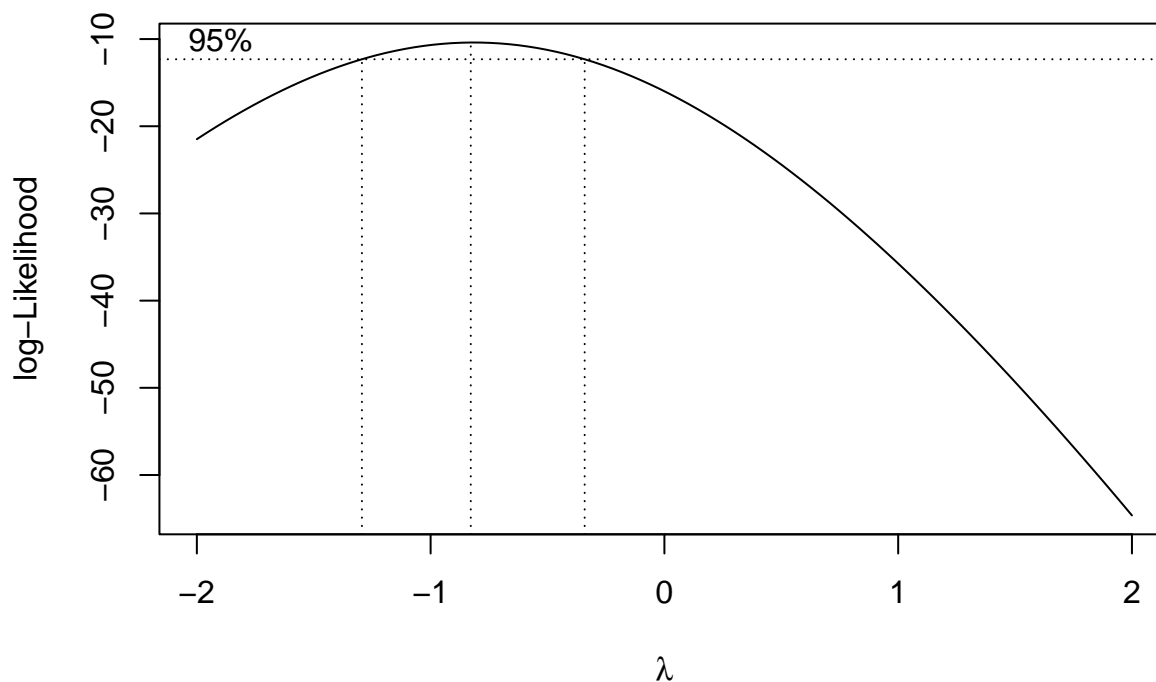
```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
bc <- boxcox(lmod)
```



```
lambda <- bc$x[which.max(bc$y)]

time_new <- (rats$time^lambda - 1) / lambda

lmod_main <- lm(time ~ poison + treat, rats)
lmod_transformed <- lm(time_new ~ poison + treat, rats)

summary(lmod_main)
```

```
##
## Call:
## lm(formula = time ~ poison + treat, data = rats)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.25167 -0.09625 -0.01490  0.06177  0.49833
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.45229    0.05592   8.088 4.22e-10 ***
## poisonII    -0.07313    0.05592  -1.308  0.19813
## poisonIII   -0.34125    0.05592  -6.102 2.83e-07 ***
## treatB       0.36250    0.06458   5.614 1.43e-06 ***
## treatC       0.07833    0.06458   1.213  0.23189
## treatD       0.22000    0.06458   3.407  0.00146 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1582 on 42 degrees of freedom
## Multiple R-squared:  0.6503, Adjusted R-squared:  0.6087
## F-statistic: 15.62 on 5 and 42 DF,  p-value: 1.123e-08
```

```r
summary(lmod_transformed)
```

```
##
## Call:
## lm(formula = time_new ~ poison + treat, data = rats)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.97378 -0.25075 -0.01478  0.28659  0.78178
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.4979     0.1483 -10.102 8.27e-13 ***
## poisonII     -0.3994     0.1483  -2.694  0.01011 *
## poisonIII    -1.6828     0.1483 -11.349 2.25e-14 ***
## treatB        1.4119     0.1712   8.247 2.54e-10 ***
## treatC        0.4714     0.1712   2.753  0.00868 **
## treatD        1.1352     0.1712   6.630 4.92e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4194 on 42 degrees of freedom
## Multiple R-squared:  0.8423, Adjusted R-squared:  0.8235
## F-statistic: 44.87 on 5 and 42 DF,  p-value: 8.808e-16
```

```r
qqnorm(residuals(lmod_transformed))
qqline(residuals(lmod_transformed))
```

Normal Q–Q Plot
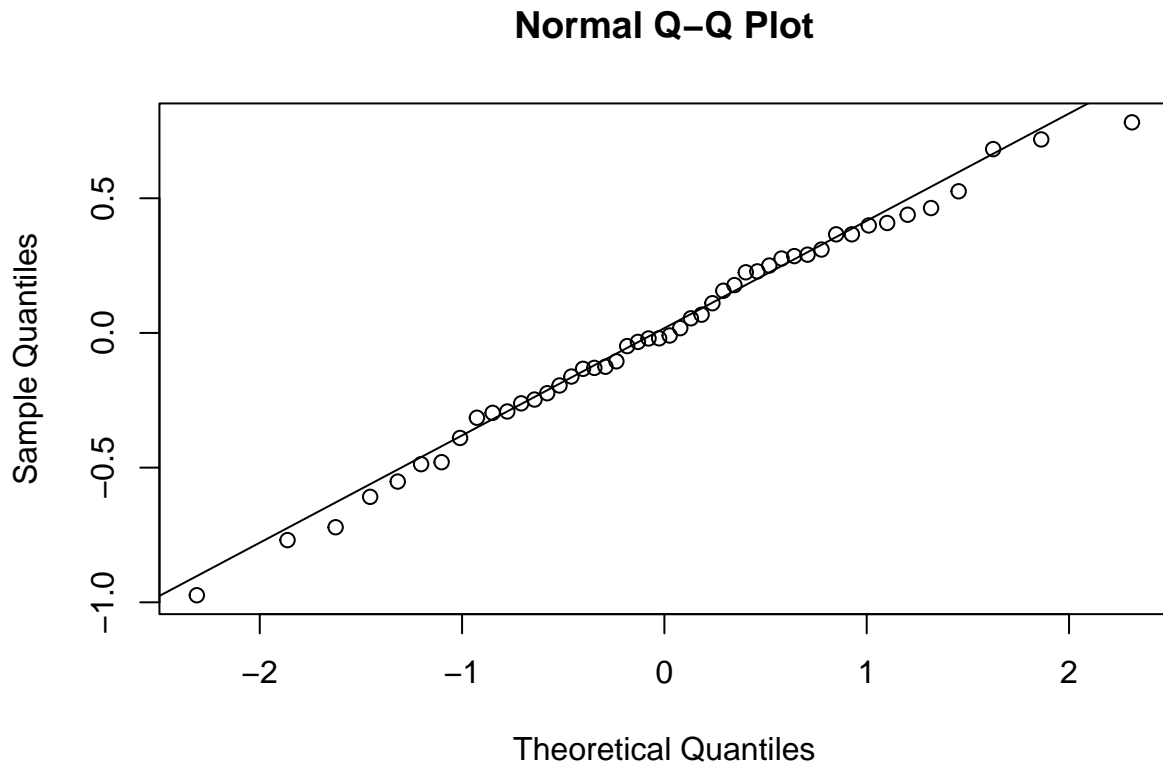
2.

We find the width of the band to be 31.10935 using the qtukey() function. Looking at a table of differences, it would appear that the differences between B and E, B and F, B and G, and B and H are significantly different at the 5% level. We also see that C and E, C and F, C and G, C and H, D and F, D and G, and D and H are significantly different.

```
data(OrchardSprays)
```

```
xtabs(decrease ~ rowpos + colpos, OrchardSprays)
```

```
##        colpos
## rowpos   1   2   3   4   5   6   7   8
##      1  57  84  87 130  43  12   8  80
##      2  95   6  72   4  28  29  72 114
##      3   8 127   5 114  60  44  13  39
##      4  69  36  39   9   5  77  57  14
##      5  92  51  22  20  17   4   4  86
##      6  90   2  16  24   7  27  81  55
##      7  15  69  72  10  81  47  20   3
##      8   2  71   4  51  71  76  61  19
```

```
matrix(OrchardSprays$treatment,8,8,byrow = TRUE)
```

```
##        [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
## [1,] "D"  "E"  "B"  "H"  "G"  "F"  "C"  "A"
## [2,] "C"  "B"  "H"  "D"  "E"  "A"  "F"  "G"
## [3,] "F"  "H"  "A"  "E"  "D"  "C"  "G"  "B"
## [4,] "H"  "A"  "E"  "C"  "F"  "G"  "B"  "D"
## [5,] "E"  "D"  "G"  "A"  "C"  "B"  "H"  "F"
## [6,] "A"  "C"  "F"  "G"  "B"  "D"  "E"  "H"
## [7,] "B"  "G"  "C"  "F"  "A"  "H"  "D"  "E"
## [8,] "G"  "F"  "D"  "B"  "H"  "E"  "A"  "C"
```

```r
lmodLatin <- lm(decrease ~ treatment + as.factor(rowpos) + as.factor(colpos), OrchardSprays)
drop1(lmodLatin, test="F")
```

```
## Single term deletions
##
## Model:
## decrease ~ treatment + as.factor(rowpos) + as.factor(colpos)
##                   Df Sum of Sq   RSS    AIC F value    Pr(>F)
## <none>                         15995 397.35
## treatment          7     56160 72155 479.77 21.0667 7.455e-12 ***
## as.factor(rowpos)  7      4767 20762 400.05  1.7884    0.1151
## as.factor(colpos)  7      2807 18802 393.70  1.0530    0.4100
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(lmodLatin)
```

```
##
## Call:
## lm(formula = decrease ~ treatment + as.factor(rowpos) + as.factor(colpos),
##     data = OrchardSprays)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -40.406  -9.094  -1.656   8.625  45.219
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)          29.906     11.442   2.614  0.01238 *
## treatmentB            3.000      9.757   0.307  0.76001
## treatmentC           20.625      9.757   2.114  0.04052 *
## treatmentD           30.375      9.757   3.113  0.00333 **
## treatmentE           58.500      9.757   5.995 4.03e-07 ***
## treatmentF           64.375      9.757   6.598 5.49e-08 ***
## treatmentG           63.875      9.757   6.546 6.50e-08 ***
## treatmentH           85.625      9.757   8.775 4.77e-11 ***
## as.factor(rowpos)2  -10.125      9.757  -1.038  0.30536
## as.factor(rowpos)3  -11.375      9.757  -1.166  0.25028
## as.factor(rowpos)4  -24.375      9.757  -2.498  0.01649 *
## as.factor(rowpos)5  -25.625      9.757  -2.626  0.01200 *
## as.factor(rowpos)6  -24.875      9.757  -2.549  0.01453 *
## as.factor(rowpos)7  -23.000      9.757  -2.357  0.02315 *
## as.factor(rowpos)8  -18.250      9.757  -1.870  0.06841 .
```

11

```
## as.factor(colpos)2    2.250      9.757    0.231  0.81875
## as.factor(colpos)3  -13.875      9.757   -1.422  0.16241
## as.factor(colpos)4   -8.250      9.757   -0.846  0.40262
## as.factor(colpos)5  -14.500      9.757   -1.486  0.14474
## as.factor(colpos)6  -14.000      9.757   -1.435  0.15875
## as.factor(colpos)7  -14.000      9.757   -1.435  0.15875
## as.factor(colpos)8   -2.250      9.757   -0.231  0.81875
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.51 on 42 degrees of freedom
## Multiple R-squared:  0.7994, Adjusted R-squared:  0.6991
## F-statistic: 7.969 on 21 and 42 DF,  p-value: 7.579e-09
```

```r
qtukey(0.95, 8, 42)*9.757/sqrt(2) ## 8 is no. of treatments, 42 deg of freedom
```

```
## [1] 31.10935
```

```r
scoefs <- c(0, coef(lmodLatin)[2:8])
outer(scoefs, scoefs, "-")
```

```
##              treatmentB treatmentC treatmentD treatmentE treatmentF
##                  0.000     -3.000    -20.625    -30.375    -58.500    -64.375
## treatmentB  3.000          0.000    -17.625    -27.375    -55.500    -61.375
## treatmentC 20.625         17.625      0.000     -9.750    -37.875    -43.750
## treatmentD 30.375         27.375      9.750      0.000    -28.125    -34.000
## treatmentE 58.500         55.500     37.875     28.125      0.000     -5.875
## treatmentF 64.375         61.375     43.750     34.000      5.875      0.000
## treatmentG 63.875         60.875     43.250     33.500      5.375     -0.500
## treatmentH 85.625         82.625     65.000     55.250     27.125     21.250
##            treatmentG treatmentH
##               -63.875    -85.625
## treatmentB    -60.875    -82.625
## treatmentC    -43.250    -65.000
## treatmentD    -33.500    -55.250
## treatmentE     -5.375    -27.125
## treatmentF      0.500    -21.250
## treatmentG      0.000    -21.750
## treatmentH     21.750      0.000
```