Burris_HW4.R

christieburris

Thu Jun 7 08:12:51 2018

```
################ Christie Burris  ##################
################ Biometry Homework 4 ##################


################## Question 1 ########################

# Prompt: The diameters of three species of pine trees were compared
# at each of four locations using five randomly selected trees per species.
# a.     Assume both factors are fixed.
# Make an ANOVA table to test for interaction and main effects.
# b.     Make an interaction plot; what does it tell you?

########## Response ############

# a. ANOVA assuming both factors are fixed: the differences in diameter
# means are significant (p=.0003) however the significance of the location
# and interaction between species and location are both insignificant
# (location p-value = .4779, interaction p-value = .4128)

# b. The interaction plot seems to indicate that the locations impact the
# diameter of trees across species.
# The three species have higher differences in means in location 1 than in location 4.
# The shape of the lines across species are different.
# Species B has a drop in mean at location 3 whereas species A and C have an increase in mean.

# c. the residuals do not look normal.
# The plot of the residuals appears to have a logarithmic shape.

diam <- read.table("~/Desktop/Biometry/hw4/ch19q09.txt",header=T,sep = ",")
colnames(diam) <- c("species","location", "diameter")
attach(diam)

# view data
plot(species,diameter)
```
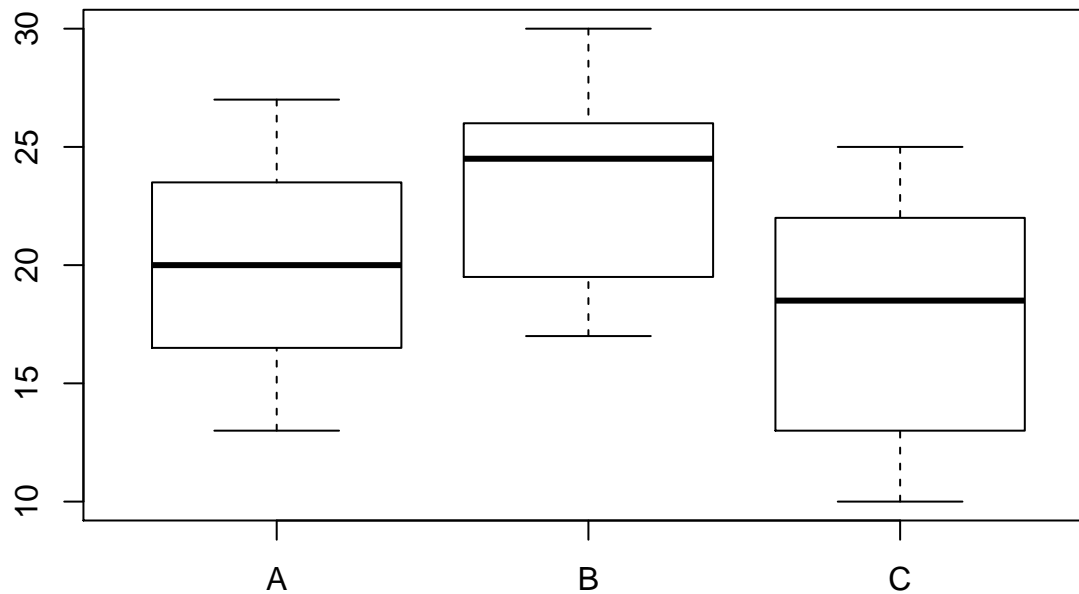
```r
speciesF = as.factor(species)
localF = as.factor(location)

# anova assuming factors are fixed. Including interactions
fit = lm(diameter~speciesF+localF+speciesF:localF,data=diam)
anova(fit)
```

```
## Analysis of Variance Table
##
## Response: diameter
##               Df Sum Sq Mean Sq F value    Pr(>F)
## speciesF       2 344.93 172.467  9.4545 0.0003453 ***
## localF         3  46.05  15.350  0.8415 0.4779000
## speciesF:localF 6 113.60  18.933  1.0379 0.4127844
## Residuals     48 875.60  18.242
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# interaction plot
interaction.plot(location,species,diameter,fun=mean)
```
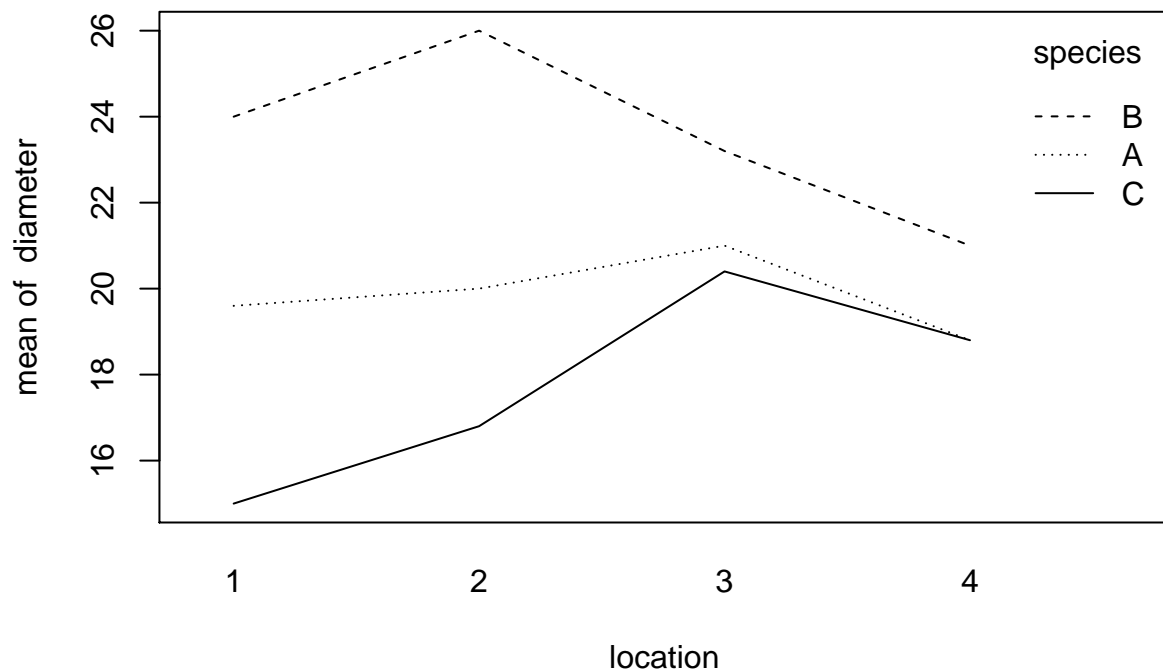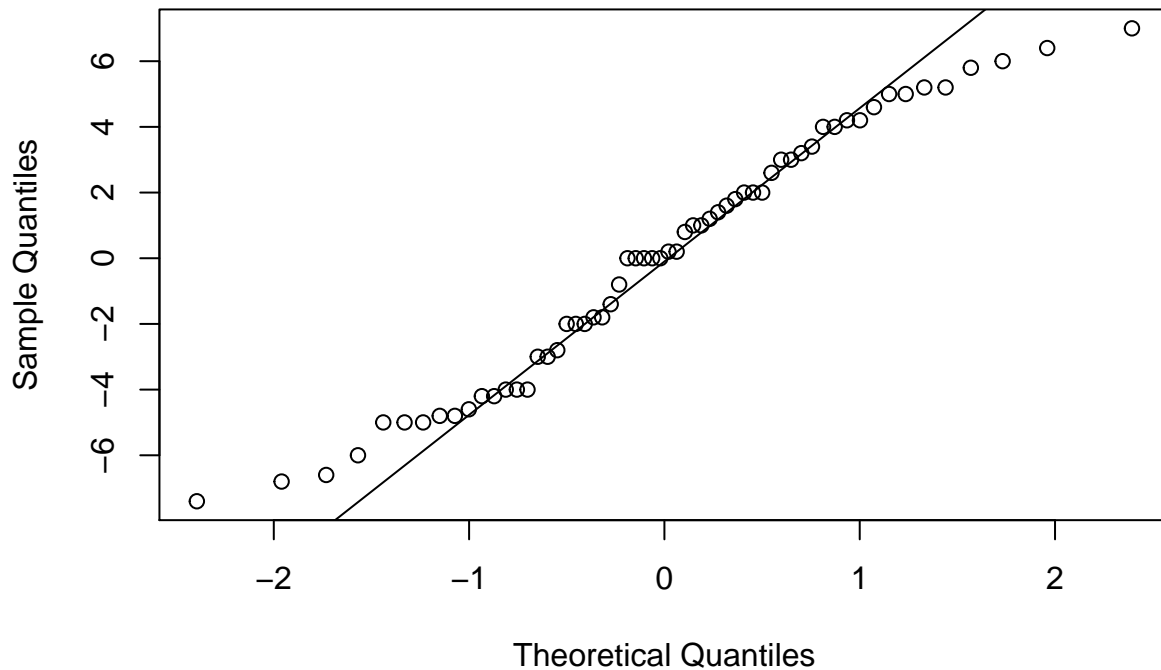
```
qqnorm(fit$residuals) # check assumptions
qqline(fit$residuals) # compare



################## Question 2 #######################
# The Tukey comparison indicates that species A and C are not significantly different, while species B

#install.packages("emmeans")
#install.packages("multcompView")
library(multcompView)
```

Normal Q–Q Plot


```r
library(emmeans)
```

```
## Warning: package 'emmeans' was built under R version 3.4.3
```

```r
fitA_e <- emmeans(fit, "speciesF") #estimate means from fit and factor species
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```r
pairs(fitA_e)  # intervals for pairs
```

```
##  contrast estimate       SE df t.ratio p.value
##  A - B        -3.7 1.350617 48  -2.739  0.0230
##  A - C         2.1 1.350617 48   1.555  0.2750
##  B - C         5.8 1.350617 48   4.294  0.0002
##
## Results are averaged over the levels of: localF
## P value adjustment: tukey method for comparing a family of 3 estimates
```
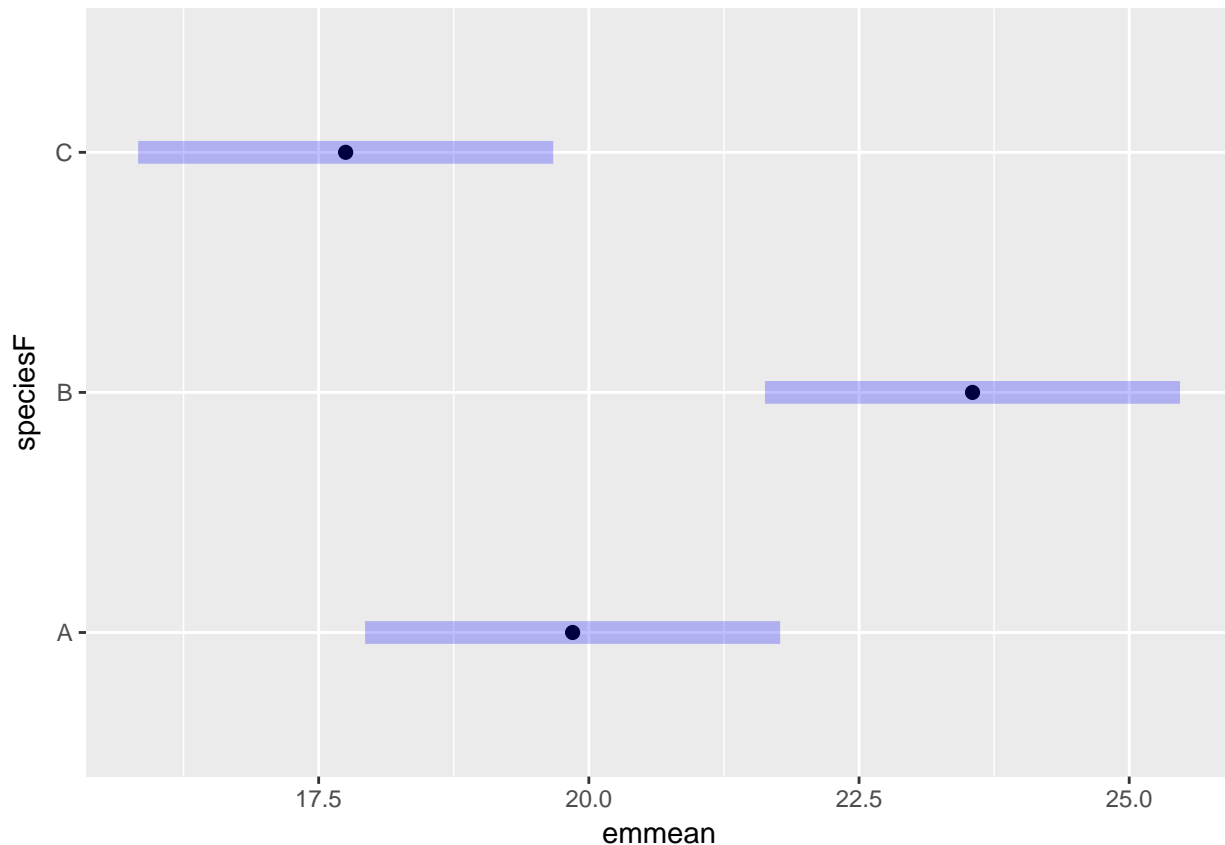
```r
plot(fitA_e)   # plot means and std errors
```

```r
cld(fitA_e,adjust="Tukey")
```

```
##  speciesF emmean        SE df lower.CL upper.CL .group
##  C          17.75 0.9550305 48 15.38733 20.11267  1
##  A          19.85 0.9550305 48 17.48733 22.21267  1
##  B          23.55 0.9550305 48 21.18733 25.91267   2
##
## Results are averaged over the levels of: localF
## Confidence level used: 0.95
## Conf-level adjustment: sidak method for 3 estimates
## P value adjustment: tukey method for comparing a family of 3 estimates
## significance level used: alpha = 0.05
```

```r
#install.packages("phia")
library(phia)
```

```
## Loading required package: car

## Warning: package 'car' was built under R version 3.4.3
```

```r
means.fit <- interactionMeans(fit)
means.fit
```

```
##    speciesF localF adjusted mean std. error
## 1        A      1          19.6   1.910061
## 2        B      1          24.0   1.910061
## 3        C      1          15.0   1.910061
## 4        A      2          20.0   1.910061
## 5        B      2          26.0   1.910061
```

```
## 6           C      2          16.8   1.910061
## 7           A      3          21.0   1.910061
## 8           B      3          23.2   1.910061
## 9           C      3          20.4   1.910061
## 10          A      4          18.8   1.910061
## 11          B      4          21.0   1.910061
## 12          C      4          18.8   1.910061
```

```r
# In the top right plot we see that Species A and C overlap in
# variance intervals and neither overlap with B as indicated
# in the Tukey comparison analysis.
# Also, the bottom left plot indicates variation in location is
# not significant whatsoever.
# We see almost complete overlap in every interval for species A, B and C.

plot(means.fit)
```

```r
################### Question 3 #########################

# Analyse whether a transformation is needed in the dbiomass variable of data.
# To do so, we check whether variance increases with mean.
# If it does, then a transformation will help to reduce the variance at higher mean values.
# We plot means vs. variance for the interaction species/water with a boxplot
# and see that in fact variance does increase with mean.
# This indicates that we should transform the data.
# We use boxcox to get the optimal power transformation.
# The optimal transformation falls in the range (,).
# We choose the square root transformation as this is the most relevant
# transformation we know within the interval.
# The new boxplot indicates less variation in variance as mean increases.
# There is however still high variation at the highest means.

library(readr)
d <- read_csv("~/Desktop/Biometry/L4 two factor models/data/ex13-48plants2.csv")
```

```
## Parsed with column specification:
## cols(
##    obs = col_integer(),
##    species = col_integer(),
##    water = col_integer(),
##    fbiomass = col_double(),
##    dbiomass = col_double()
## )
```

```r
attach(d)
```

```
## The following object is masked from diam:
##
##      species
```

```r
species <- as.factor(species)
water <- as.factor(water)
fit <- lm(dbiomass ~ water*species, data=d)
anova(fit)
```

```
## Analysis of Variance Table
##
## Response: dbiomass
##                Df Sum Sq Mean Sq F value     Pr(>F)
## water           1  54151   54151 207.985 < 2.2e-16 ***
## species         1  46525   46525 178.693 < 2.2e-16 ***
## water:species   1   4470    4470  17.167 6.824e-05 ***
## Residuals     108  28119     260
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
library(car)
Anova(fit,Type=III)
```

```
## Anova Table (Type II tests)
##
## Response: dbiomass
##                Sum Sq  Df F value     Pr(>F)
## water           54151   1 207.985 < 2.2e-16 ***
## species         46525   1 178.693 < 2.2e-16 ***
## water:species    4470   1  17.167 6.824e-05 ***
## Residuals       28119 108
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# check assumptions
qqnorm(fit$residuals)
qqline(fit$residuals)
```


adjusted mean
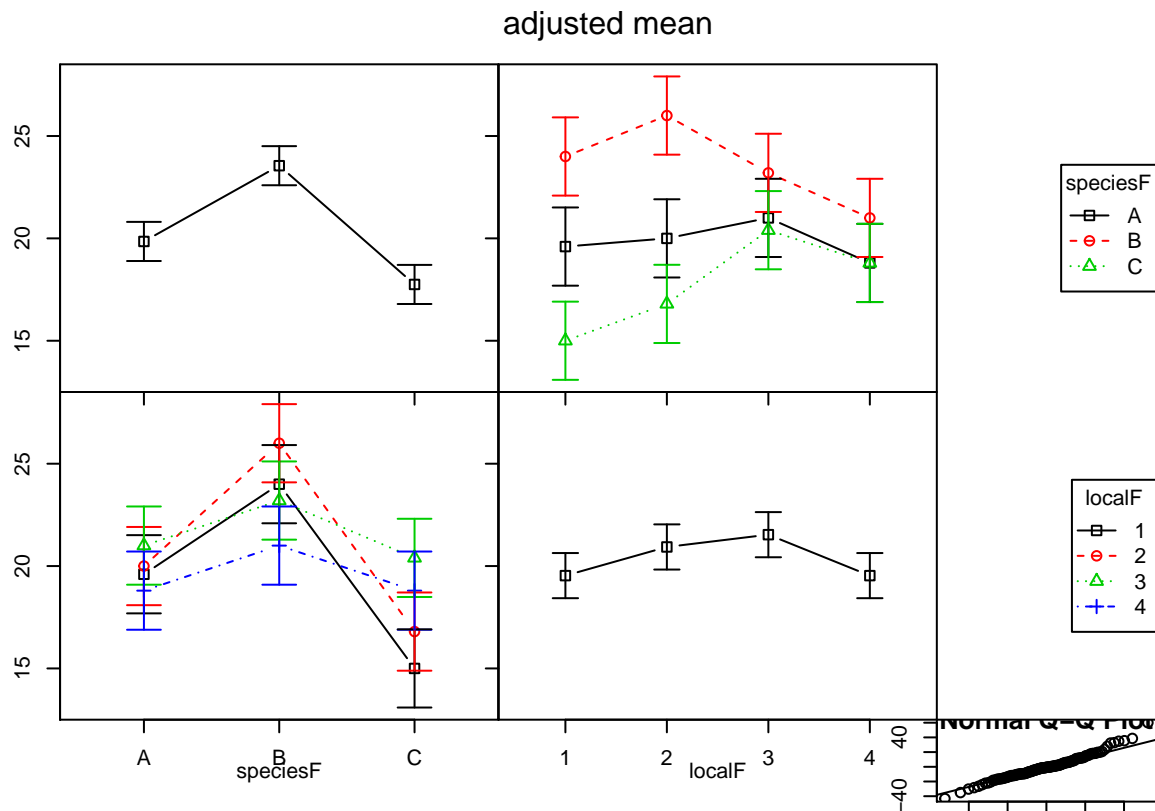
```r
# checking variance vs. mean
specieswater  <- factor(interaction(d$species,d$water))
par(mfrow=c(1,1))
boxplot(dbiomass~specieswater)
```
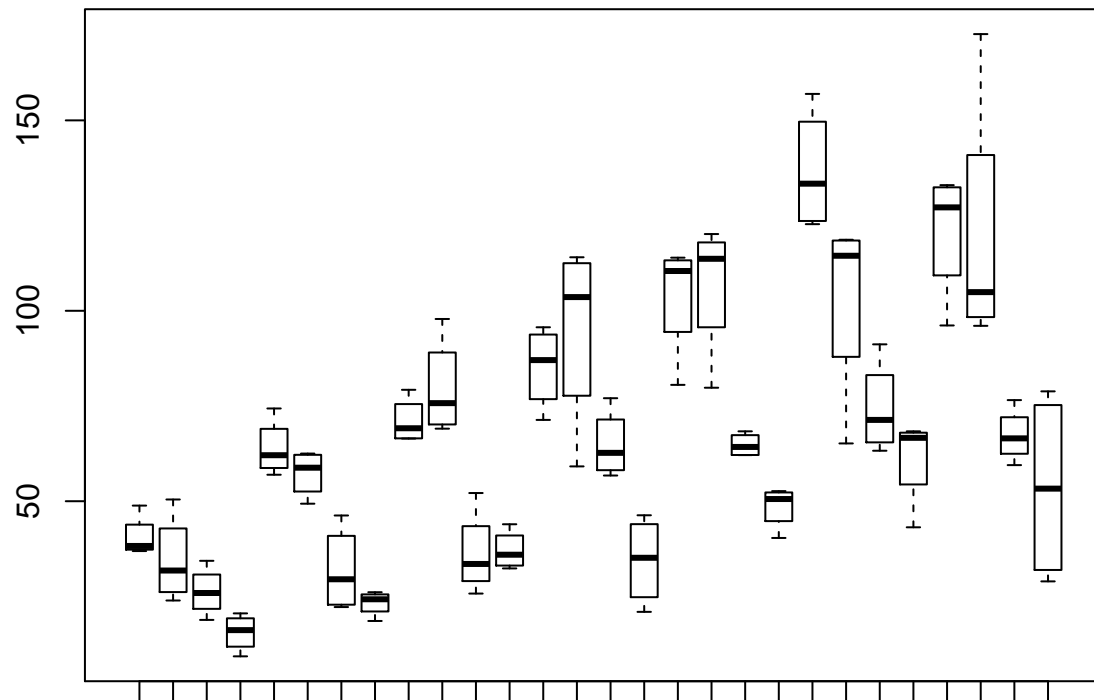


```r
library(MASS)

# use boxcox to find optimal transformation
# optimal transformation is the maximum on the plot,
# good to choose a relevant transformation that's close to the maximum

boxcox(dbiomass~specieswater, lambda = seq(-.25,1,length=100))
```
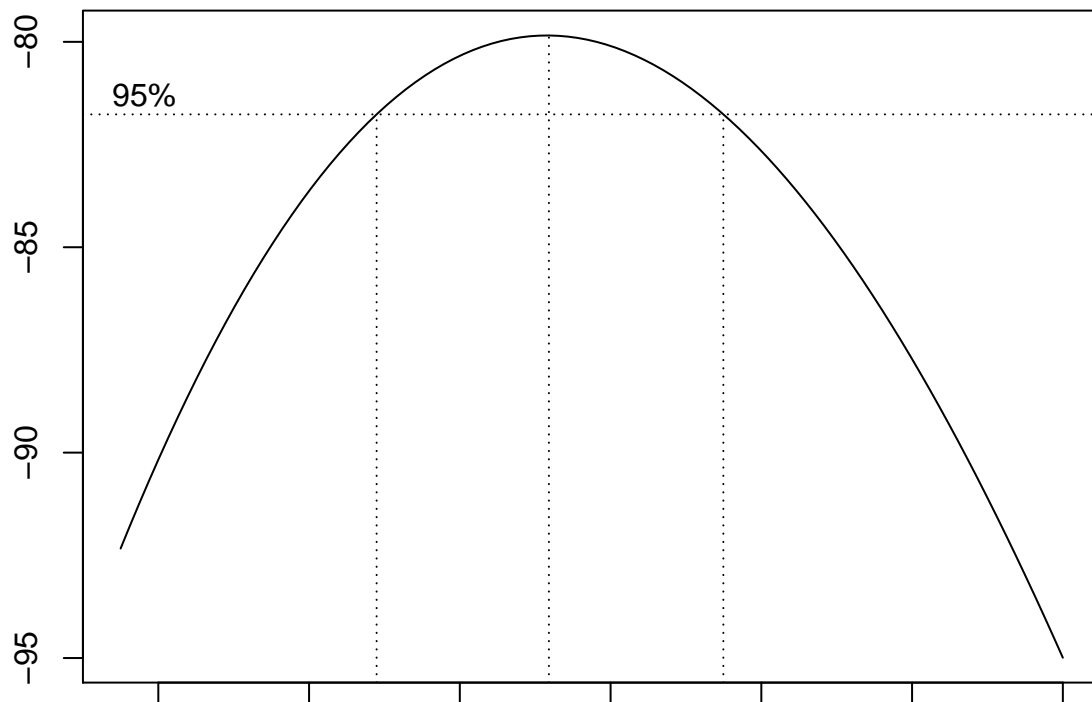
```
# this is a power transformation
d$droot = dbiomass^0.5

# recheck mean vs variance
boxplot(d$droot~specieswater)
```