

# Worksheet2

STAT414

2024-09-15

```
# #1 - Use the DownloadFestival data set for this part. The dataset and its analysis is available  
# in DSUR book.
```

```
#install.packages("EnvStats")
```

```
#read in the data using read.table, delimited by whitespace, filling blank values
```

```
data <- read.table("C:/Users/criss/Desktop/STAT414/Assignments/9-21/DownloadFestival.dat",  
                  header = TRUE, sep = " ", stringsAsFactors = FALSE, fill=TRUE)
```

```
head(data,10)
```

```
##      ticknumb gender day1 day2 day3  
## 1         2111   Male 2.64 1.35 1.61  
## 2         2229 Female 0.97 1.41 0.29  
## 3         2338   Male 0.84  NA   NA  
## 4         2384 Female 3.03  NA   NA  
## 5         2401 Female 0.88 0.08  NA  
## 6         2405   Male 0.85  NA   NA  
## 7         2467 Female 1.56  NA   NA  
## 8         2478 Female 3.02  NA   NA  
## 9         2490   Male 2.29  NA   NA  
## 10        2504 Female 1.11 0.44 0.55
```

```
max(data$day1)
```

```
## [1] 20.02
```

```
max(data$day2,na.rm = TRUE)
```

```
## [1] 3.44
```

```
max(data$day3,na.rm = TRUE)
```

```
## [1] 3.41
```

```
sd(data$day1,na.rm = TRUE)
```

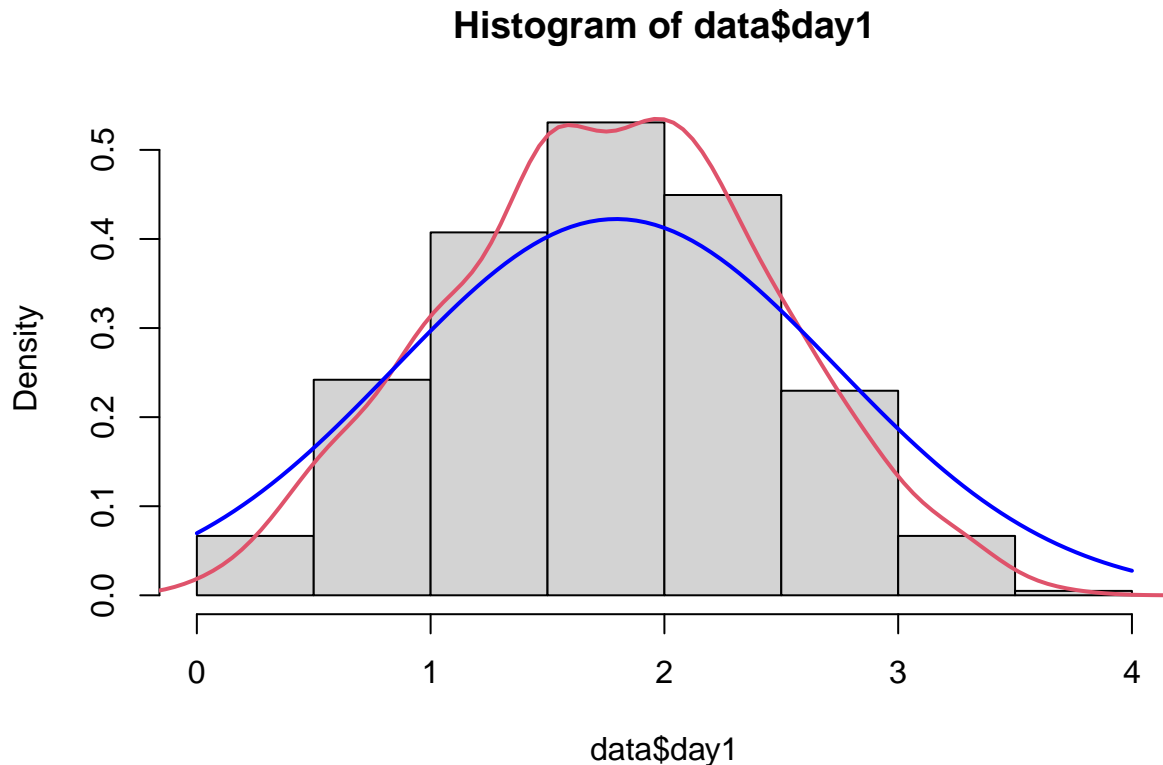
```
## [1] 0.9444949
```

```

day2_sd <- sd(data$day2,na.rm = TRUE)
day3_sd <- sd(data$day3,na.rm = TRUE)

#Generate histogram, histogram line, and superimpose normal line for day 1 data
hist(data$day1,probability=TRUE,xlim = c(0, 4),breaks= c(0,0.5,1,1.5,2,2.5,3,3.5,4,4.5,22))
lines(density(data$day1), col = 2, lwd = 2)
curve(dnorm(x, mean=mean(data$day1), sd=sd(data$day1)), lwd=2, col="blue", add=TRUE)

```



```

cat("The normal line is symmetric and matches the actual line representation of 'day 1' well,
therefore we can say that a Normal Distribution is appropriate for this dataset.")

```

```

## The normal line is symmetric and matches the actual line representation of 'day 1' well,
## therefore we can say that a Normal Distribution is appropriate for this dataset.

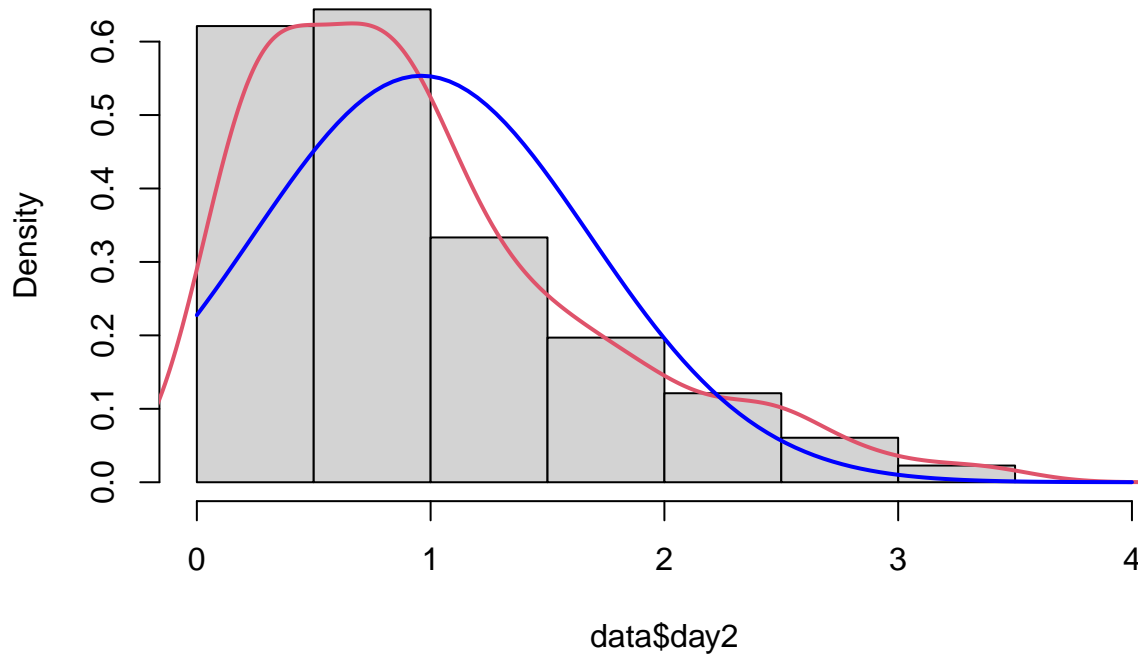
```

```

#Generate histogram, histogram line, and superimpose normal line for day 2 data, remove any NA values
hist(data$day2,probability=TRUE,breaks= c(0,0.5,1,1.5,2,2.5,3,3.5,4))
lines(density(data$day2, na.rm = TRUE), col = 2, lwd = 2)
curve(dnorm(x, mean=mean(data$day2, na.rm = TRUE), sd=sd(data$day2,na.rm = TRUE)), lwd=2,
      col="blue", add=TRUE)

```

## Histogram of data\$day2

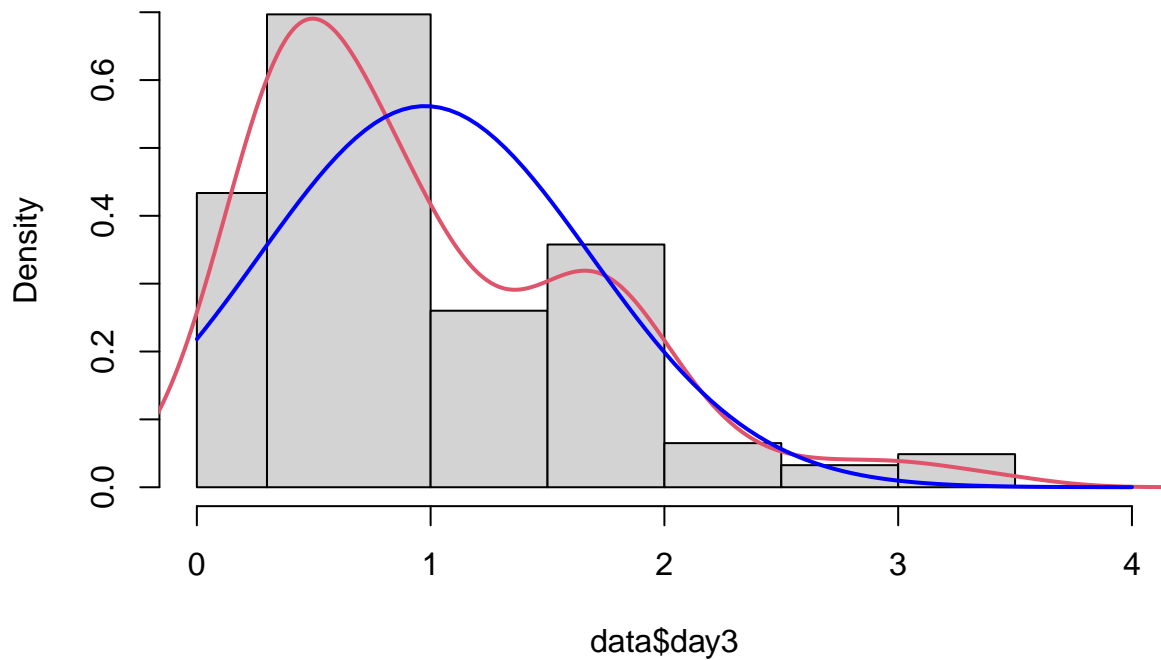


```
cat("The normal line is heavily skewed with a right-skew and a long tail, the line is definitely  
not symmetric, therefore we can say that a Normal Distribution is not appropriate for 'day 2'")
```

```
## The normal line is heavily skewed with a right-skew and a long tail, the line is definitely  
## not symmetric, therefore we can say that a Normal Distribution is not appropriate for 'day 2'
```

```
#Generate histogram, histogram line, and superimpose normal line for day 3 data, remove any NA values  
hist(data$day3,probability=TRUE,breaks= c(0,0.3,1,1.5,2,2.5,3,3.5,4))  
lines(density(data$day3, na.rm = TRUE), col = 2, lwd = 2)  
curve(dnorm(x, mean=mean(data$day3, na.rm = TRUE), sd=sd(data$day3,na.rm = TRUE)), lwd=2,  
      col="blue", add=TRUE)
```

## Histogram of data\$day3

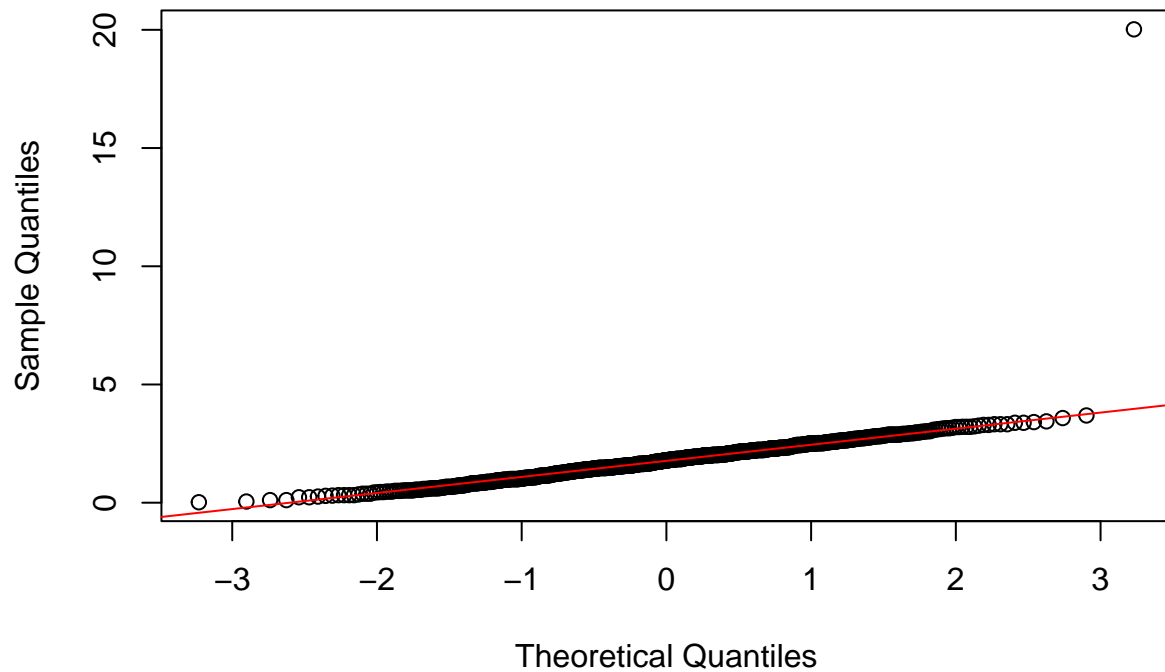


```
cat("The normal line is heavily skewed with a right-skew and a long tail, the line is definitely  
not symmetric, therefore we can say that a Normal Distribution is not appropriate for 'day 3'")
```

```
## The normal line is heavily skewed with a right-skew and a long tail, the line is definitely  
## not symmetric, therefore we can say that a Normal Distribution is not appropriate for 'day 3'
```

```
#Create qqplots of the 3 days, as well as a reference line that shows normal fit  
qqnorm(data$day1, main="day1 qq plot")  
qqline(data$day1, col="red")
```

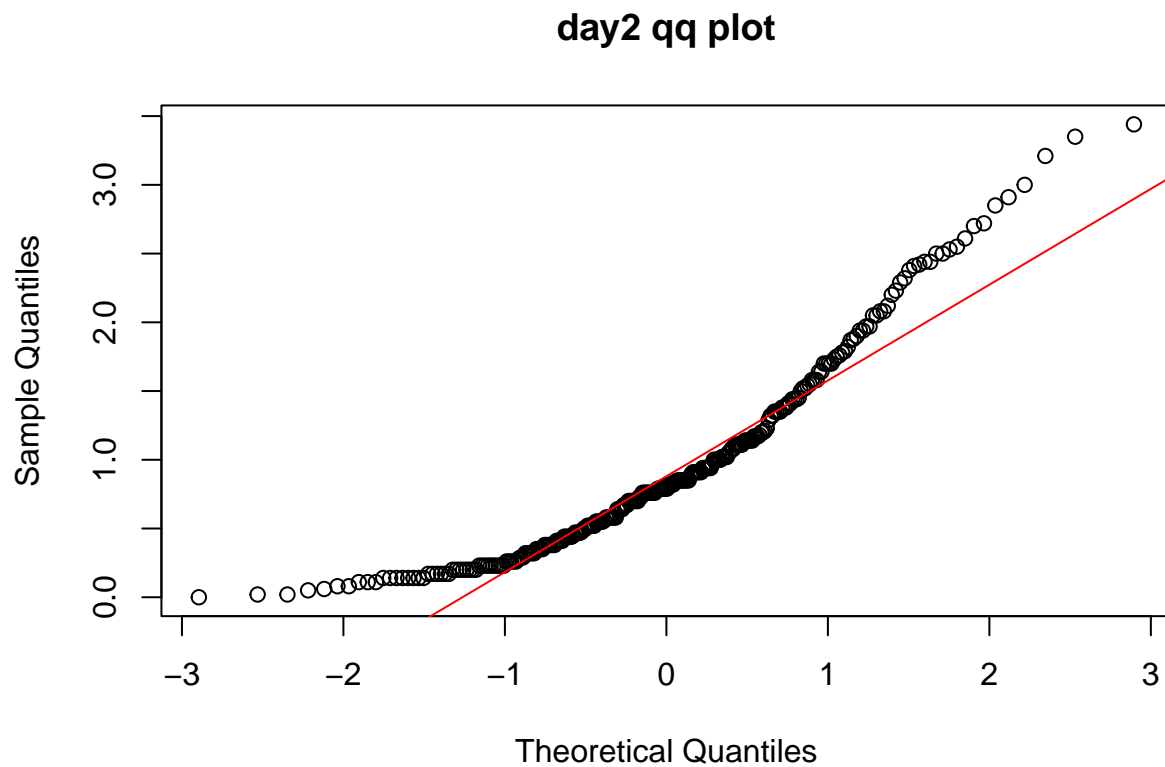
### day1 qq plot



```
cat("The datapoints of day 1 fall along the reference line, therefore we can say  
    Normal distribution is appropriate to model this dataset")
```

```
## The datapoints of day 1 fall along the reference line, therefore we can say  
##     Normal distribution is appropriate to model this dataset
```

```
qqnorm(data$day2, main="day2 qq plot")  
qqline(data$day2, col="red")
```

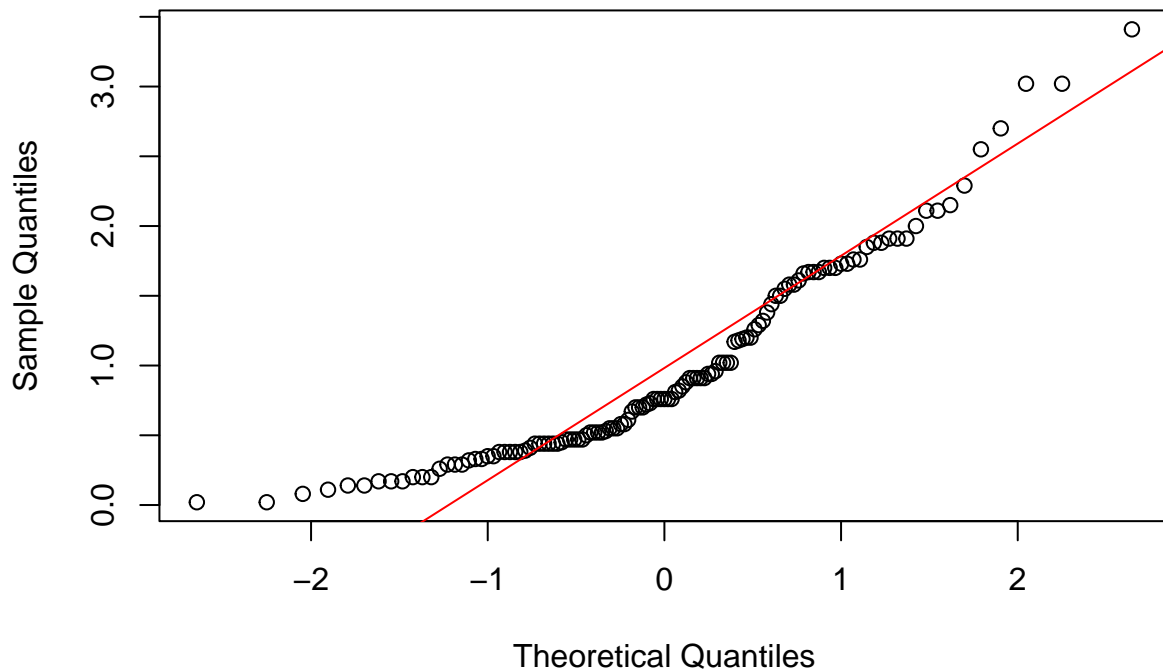


```
cat("The datapoints of day 2 do not fall along the reference line, therefore  
we can say Normal distribution is not appropriate to model this dataset")
```

```
## The datapoints of day 2 do not fall along the reference line, therefore  
## we can say Normal distribution is not appropriate to model this dataset
```

```
qqnorm(data$day3, main="day3 qq plot")  
qqline(data$day3, col="red")
```

### day3 qq plot



```
cat("The datapoints of day 3 do not fall along the reference line, therefore
we can say Normal distribution is not appropriate to model this dataset")
```

```
## The datapoints of day 3 do not fall along the reference line, therefore
## we can say Normal distribution is not appropriate to model this dataset
```

```
# #2 - The dataset halibut (available in the R package EnvStats) has two variables: Annual Catch
# Per Unit Effort (CPUE), and biomass and exploitable biomass of Pacific halibut for the years
# 1935 through 1989. Dataset is analyzed in the Millard and Neerchal book. (Warning: It uses
# SPLUS, a precursor to the R package. Syntax is very similar, but not identical.)
```

```
library(readxl)
library(ggplot2)
```

```
data <- read_excel("halibut_df.xls") #import excel data
```

```
halibut <- data.frame(data) #convert excel sheet to dataframe
head(halibut,10)
```

```
##      cpue biomass
## 1    132  171.19
## 2    143  179.71
## 3    160  182.77
## 4    171  190.01
## 5    145  199.12
## 6    155  208.18
```

```
## 7    178  214.28
## 8    179  222.83
## 9    177  239.13
## 10   221  258.25
```

```
#Compute summary statistics of cpue and biomass
```

```
summary(halibut$cpue)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    108.0   174.0   202.0   201.9   242.0   290.0
```

```
summary(halibut$biomass)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    124.0   181.2   241.3   236.5   300.5   314.9
```

```
#loop to generate breaks for the histogram of cpue
```

```
cpue_breaks <- c()
i <- 100
count <- 1
while(i <= max(halibut$cpue)+50){
  cpue_breaks[count] = i
  i <- i + 10
  count <- count + 1
}
cpue_breaks
```

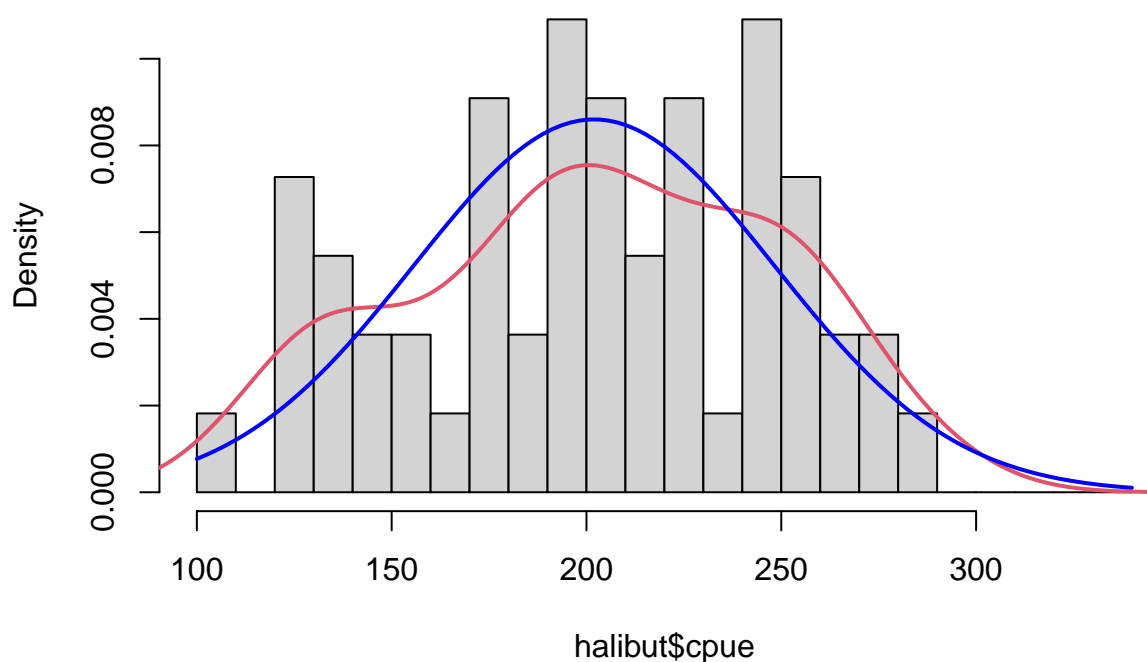
```
## [1] 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280
## [20] 290 300 310 320 330 340
```

```
#Histogram of cpue, with normal curve
```

```
hist(halibut$cpue,probability=TRUE,breaks=cpue_breaks)
lines(density(halibut$cpue), col = 2, lwd = 2)
curve(dnorm(x, mean=mean(halibut$cpue), sd=sd(halibut$cpue)), lwd=2,
      col="blue", add=TRUE)
```



## Histogram of halibut\$cpue



```
#loop to generate breaks for the histogram of biomass
```

```
bio_breaks <- c()
```

```
j <- 100
```

```
count2 <- 1
```

```
while(j <= max(halibut$biomass)+50){
  bio_breaks[count2] = j
  j <- j + 10
  count2 <- count2 + 1
}
```

```
bio_breaks
```

```
## [1] 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280
```

```
## [20] 290 300 310 320 330 340 350 360
```

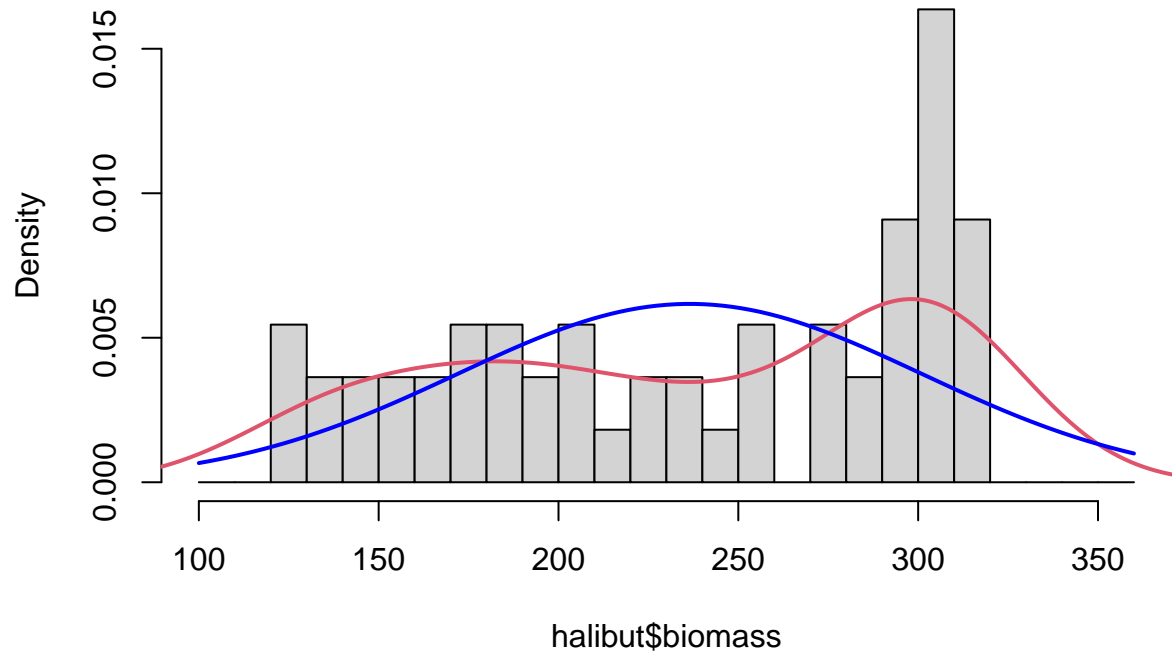
```
#Histogram of cpue, with normal curve
```

```
hist(halibut$biomass,probability=TRUE,breaks=bio_breaks)
```

```
lines(density(halibut$biomass), col = 2, lwd = 2)
```

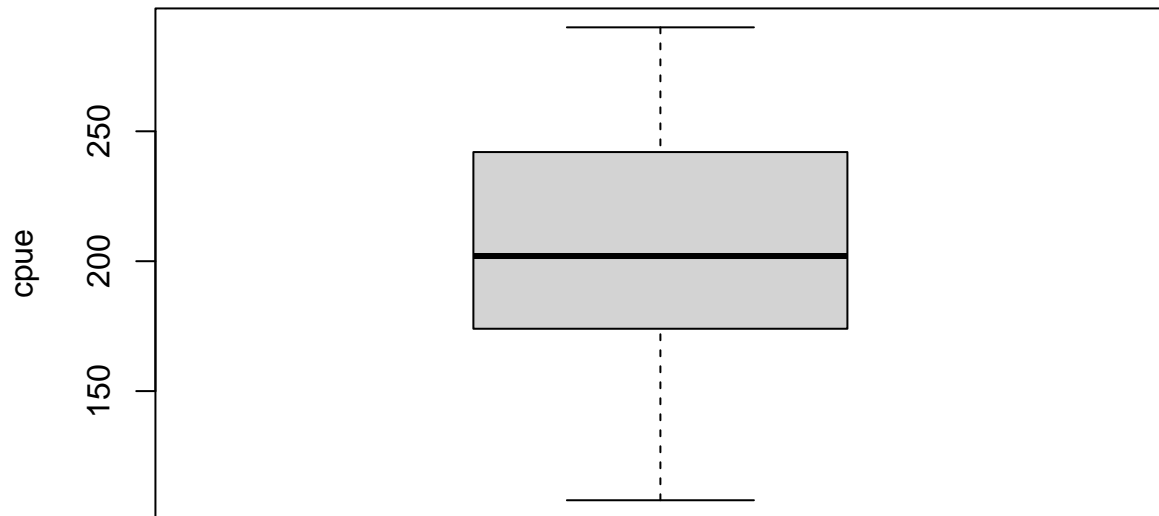
```
curve(dnorm(x, mean=mean(halibut$biomass), sd=sd(halibut$biomass)), lwd=2,
      col="blue", add=TRUE)
```

## Histogram of halibut\$biomass



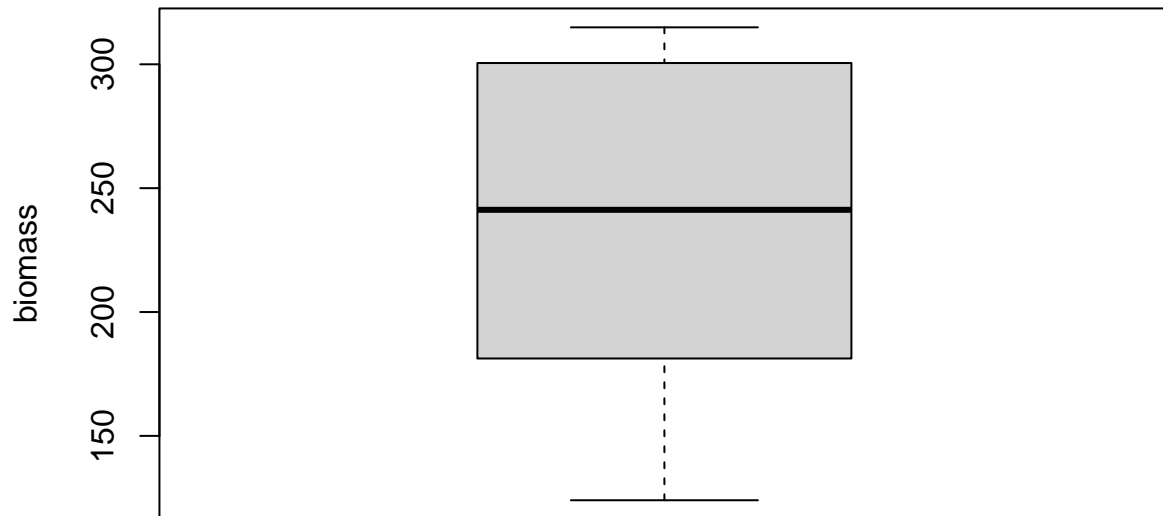
```
#Boxplot of cpue  
boxplot(data$cpue,main="Boxplot of cpue", ylab="cpue")
```

**Boxplot of cpue**



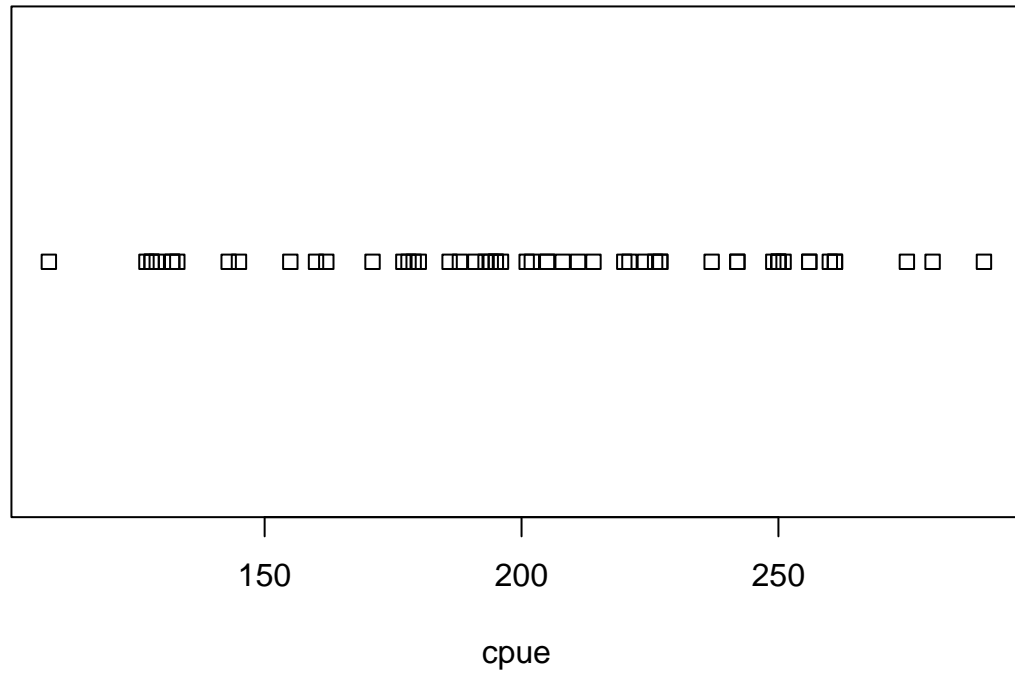
```
#Boxplot of biomass  
boxplot(data$biomass,main="Boxplot of biomass", ylab="biomass")
```

**Boxplot of biomass**



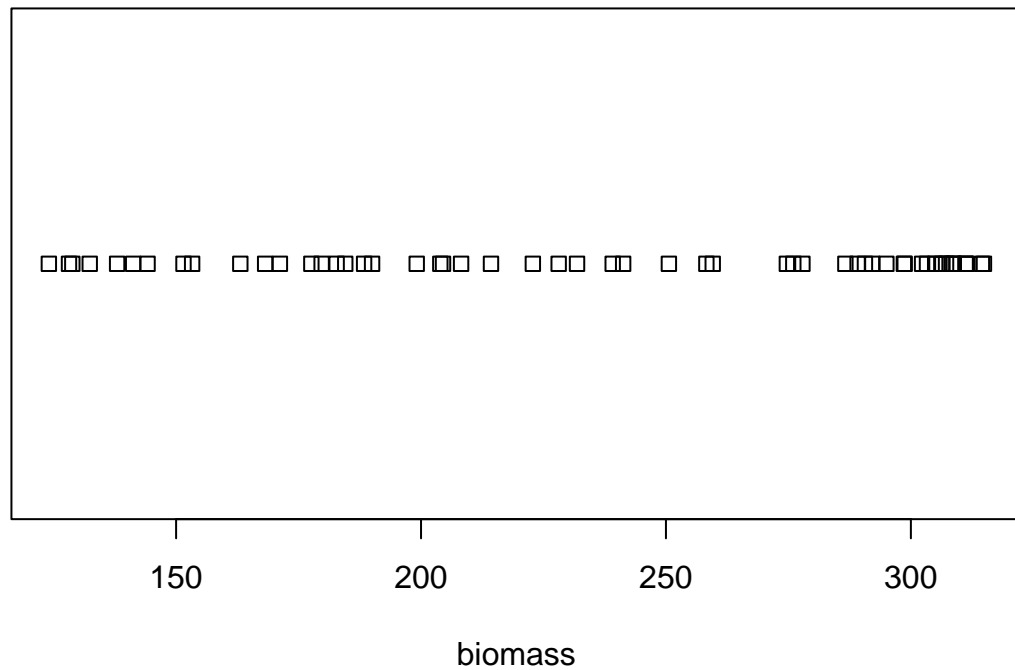
```
#Strip plot of cpue  
stripchart(data$cpue,main="Strip plot of cpue", xlab="cpue")
```

## Strip plot of cpue



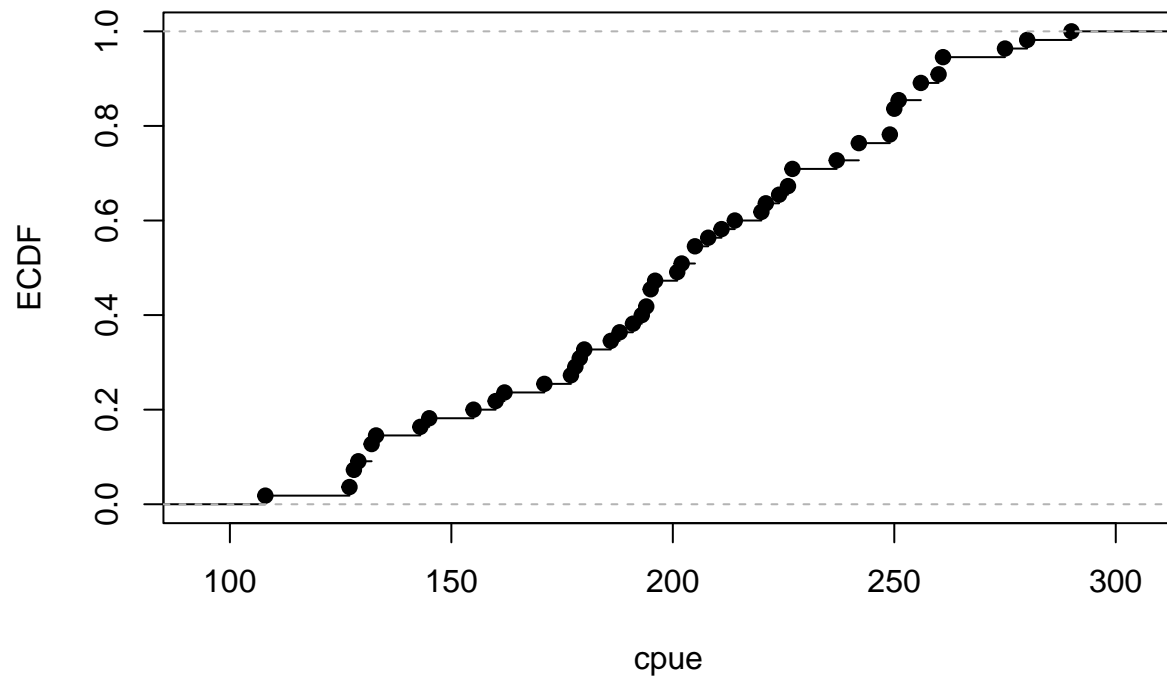
```
#Strip plot of biomass  
stripchart(data$biomass,main="Strip plot of biomass", xlab="biomass")
```

## Strip plot of biomass



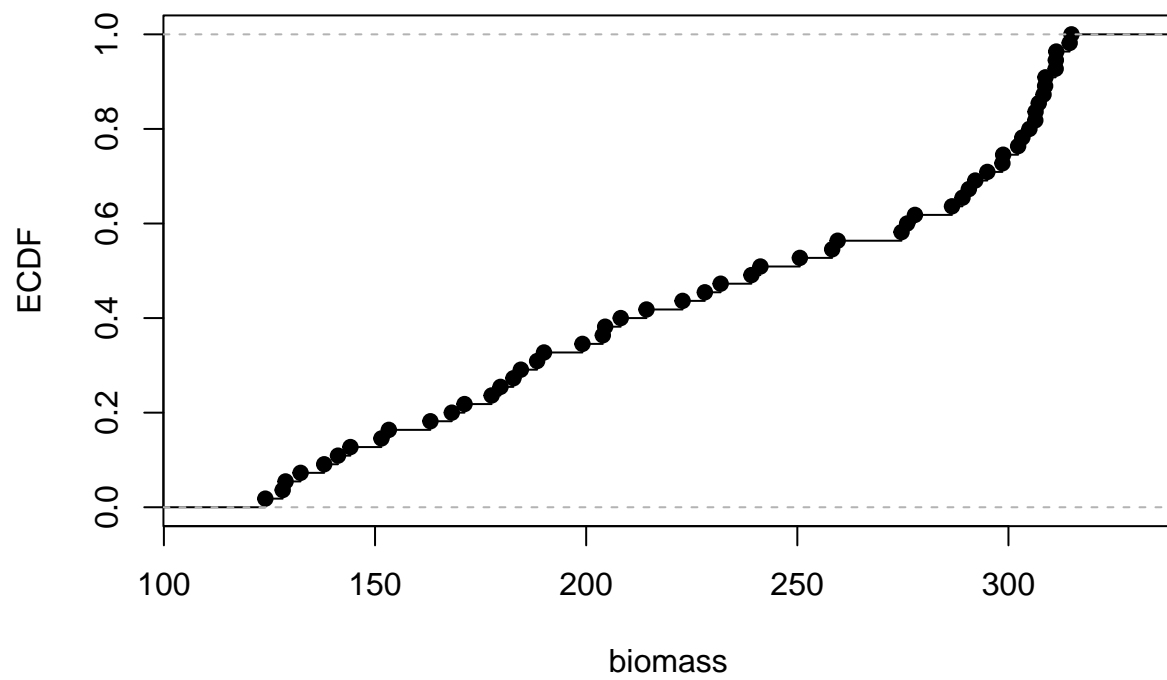
```
#Quantile (Empirical CDF / ECDF) plot of cpue  
ecdf_cpue <- ecdf(halibut$cpue) #use ecdf function  
plot(ecdf_cpue, main = "Quantile plot of cpue", xlab = "cpue", ylab = "ECDF")
```

### Quantile plot of cpue



```
#Quantile (Empirical CDF / ECDF) plot of cpue  
ecdf_biomass <- ecdf(halibut$biomass) #use ecdf function  
plot(ecdf_biomass, main = "Quantile plot of biomass", xlab = "biomass", ylab = "ECDF")
```

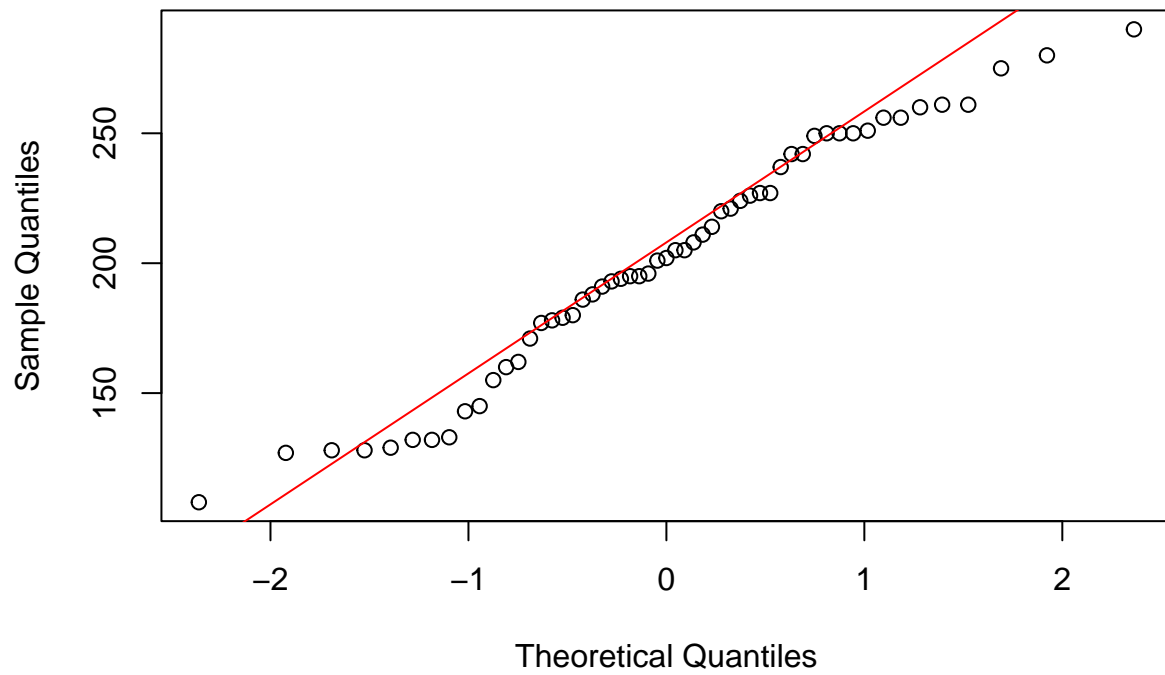
### Quantile plot of biomass



```
#qqplot and reference line of cpue  
qqnorm(halibut$cpue, main="cpue qq plot")  
qqline(halibut$cpue, col="red")
```

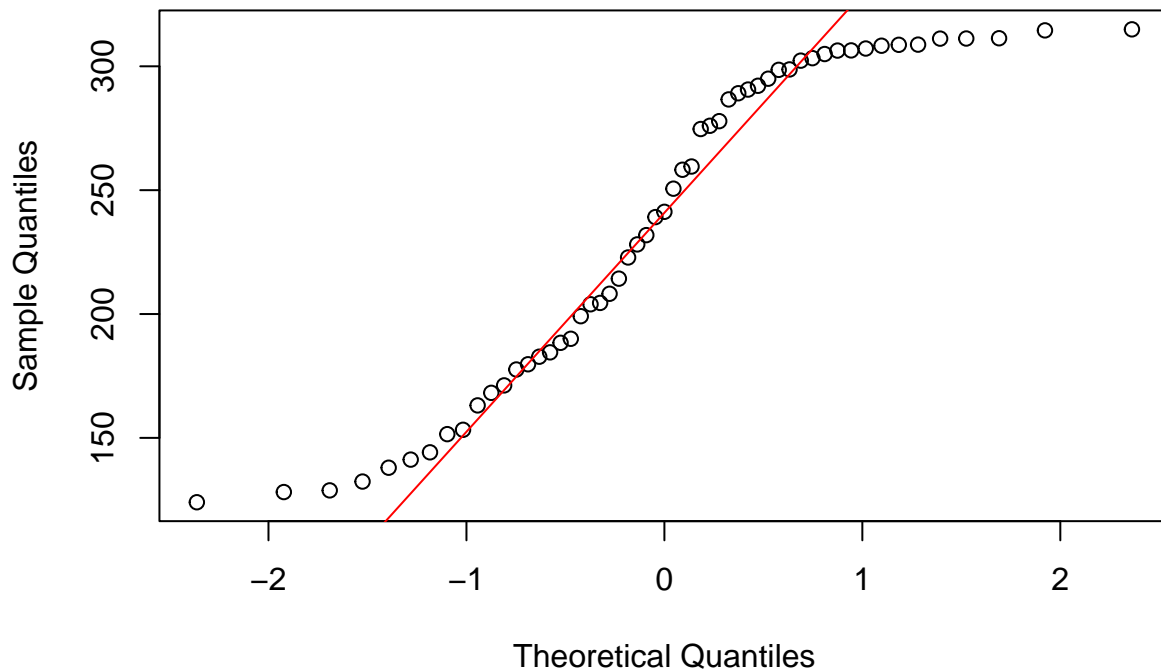


cpue qq plot



```
#qqplot and reference line of biomass  
qqnorm(halibut$biomass, main="biomass qq plot")  
qqline(halibut$biomass, col="red")
```

## biomass qq plot



*#For the tukey mean difference plot, there are no 2 groups to compare cpue and biomass  
#therefore simply divide the dataset in half and compare the two halves.*

*#nrow(halibut) = 55*

```
cpue1 <- halibut$cpue[1:(55/2)]
```

```
cpue2 <- halibut$cpue[ceiling((55/2)):55]
```

```
length(cpue1)
```

```
## [1] 27
```

```
length(cpue2)
```

```
## [1] 28
```

```
bio1 <- halibut$biomass[1:(55/2)]
```

```
bio2 <- halibut$biomass[ceiling((55/2)):55]
```

*#Tukey mean difference plot of cpue*

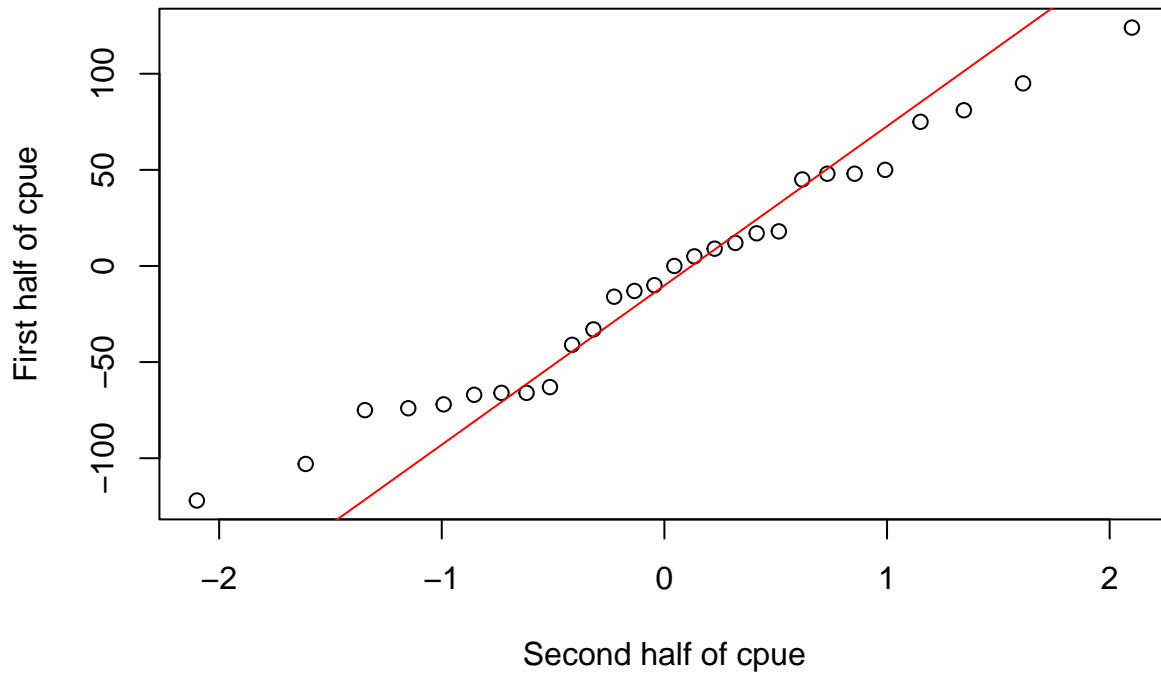
```
qqnorm(cpue2 - cpue1, main="Tukey Mean-Difference Q-Q Plot for cpue",  
       xlab="Second half of cpue", ylab="First half of cpue")
```

```
## Warning in cpue2 - cpue1: longer object length is not a multiple of shorter  
## object length
```

```
qqline(cpue2 - cpue1, col = "red")
```

```
## Warning in cpue2 - cpue1: longer object length is not a multiple of shorter
## object length
```

### Tukey Mean-Difference Q-Q Plot for cpue



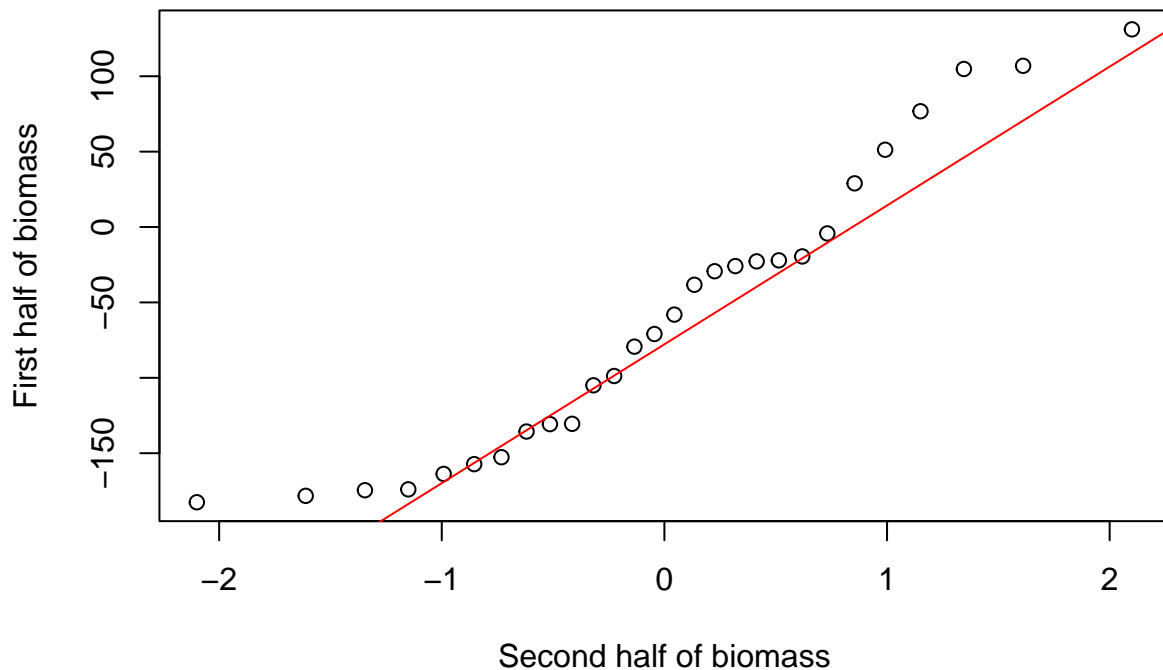
```
#Tukey mean difference plot of biomass
qqnorm(bio2 - bio1, main="Tukey Mean-Difference Q-Q Plot for biomass",
        xlab="Second half of biomass", ylab="First half of biomass")
```

```
## Warning in bio2 - bio1: longer object length is not a multiple of shorter
## object length
```

```
qqline(bio2 - bio1, col = "red")
```

```
## Warning in bio2 - bio1: longer object length is not a multiple of shorter
## object length
```

## Tukey Mean-Difference Q-Q Plot for biomass



```
cat("Cpue appears to be normally distributed if we analyze the features of all the
visual plots we have generated for cpue.
* The histogram has a clear symmetrical curve that the super-imposed normal curve
is very similar to. The peak is centered.
* The boxplot is mostly symmetrical about the median, and the whiskers are of
equal length, supporting a normal distribution.
* In the strip plot, the datapoints are not symmetrical about the mean, the
concentration of the datapoints are left-skewed, and the spread of the datapoints
are uneven.
* In the quantile plot, the curve somewhat resembles a S-shape, but the curve is
not symmetrical about the median.
* In the qqplot, the datapoints stray away from the reference line.
* In the tukey mean-difference qqplot, the datapoints again do not follow the
reference line accurately.
")
```

```
## Cpue appears to be normally distributed if we analyze the features of all the
## visual plots we have generated for cpue.
## * The histogram has a clear symmetrical curve that the super-imposed normal curve
## is very similar to. The peak is centered.
## * The boxplot is mostly symmetrical about the median, and the whiskers are of
## equal length, supporting a normal distribution.
## * In the strip plot, the datapoints are not symmetrical about the mean, the
## concentration of the datapoints are left-skewed, and the spread of the datapoints
## are uneven.
## * In the quantile plot, the curve somewhat resembles a S-shape, but the curve is
```

```

##      not symmetrical about the median.
##      * In the qqplot, the datapoints stray away from the refernce line.
##      * In the tukey mean-difference qqplot, the datapoints again do not follow the
##      reference line accurately.
##

cat("Biomass appears to not be normally distributed if we analyze the features of all the
visual plots we have generated.
* The histogram does not have a symmetrical curve and the peak is skewed to the left.
* The boxplot is symmetrical about the median, however the wiskers are of extremely varying
length, which does not support normal distribution.
* In the strip plot, the datapoints are mostly symmetrical about the mean, the
concentration of the datapoints are most prevalent around the mean, and the
spread of the datapoints are even on both sides from the mean.
* In the quantile plot, the curve resembles a S-shape, and the curve is
symmetrical about the median.
* In the qqplot, the datapoints follow the refernce line accurately.
* In the tukey mean-difference qqplot, the datapoints follow the reference
line accurately.
")

## Biomass appears to not be normally distributed if we analyze the features of all the
## visual plots we have generated.
##      * The histogram does not have a symmetrical curve and the peak is skewed to the left.
##      * The boxplot is symmetrical about the median, however the wiskers are of extremely varying
##      length, which does not support normal distribution.
##      * In the strip plot, the datapoints are mostly symmetrical about the mean, the
##      concentration of the datapoints are most prevalent around the mean, and the
##      spread of the datapoints are even on both sides from the mean.
##      * In the quantile plot, the curve resembles a S-shape, and the curve is
##      symmetrical about the median.
##      * In the qqplot, the datapoints follow the refernce line accurately.
##      * In the tukey mean-difference qqplot, the datapoints follow the reference
##      line accurately.
##

library(EnvStats)

##
## Attaching package: 'EnvStats'

## The following objects are masked from 'package:stats':
##
##      predict, predict.lm

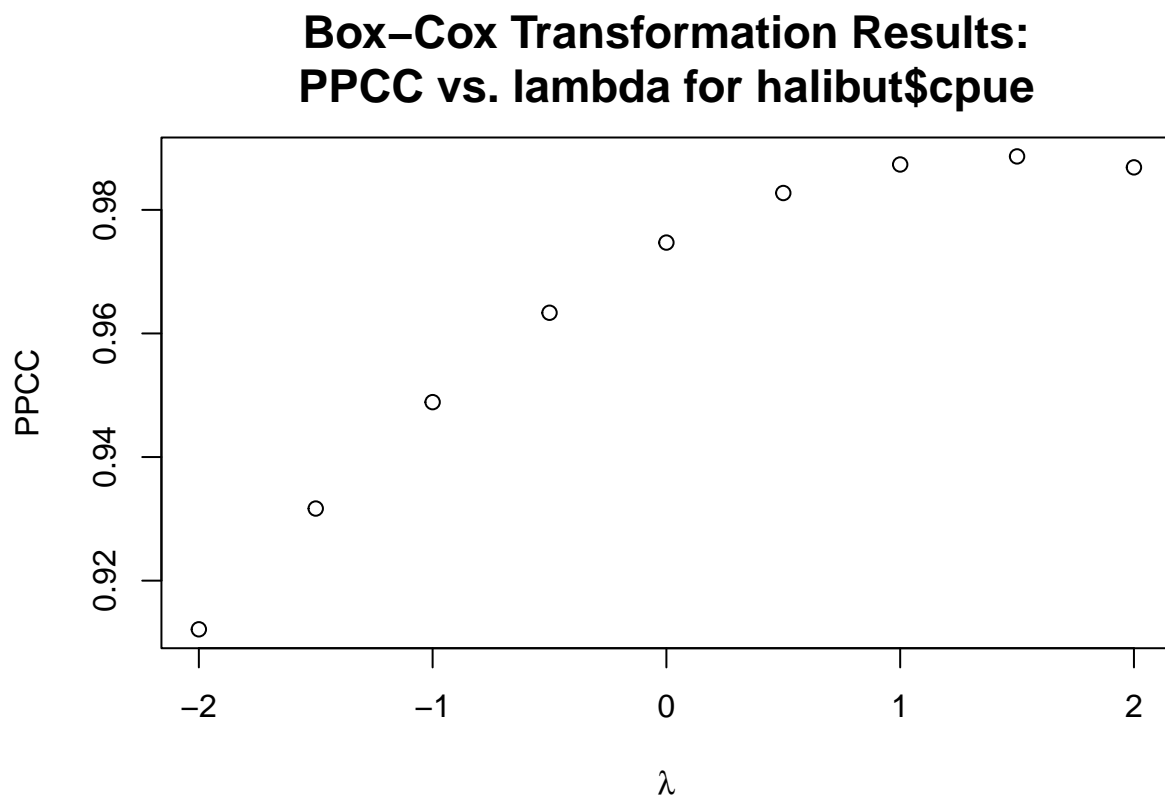
#perform boxcox transformation on cpue for (-2,2) incrementing by .5
box_cpue <- boxcox(halibut$cpue, lambda = seq(-2, 2, by=0.5), plot = TRUE,
                  optimize = FALSE)
#plot lambda vs PPCC
box_cpue

##

```

```
## Results of Box-Cox Transformation
## -----
##
## Objective Name:          PPCC
##
## Data:                   halibut$cpue
##
## Sample Size:            55
##
## lambda      PPCC
##   -2.0 0.9121241
##   -1.5 0.9316654
##   -1.0 0.9488856
##   -0.5 0.9633503
##    0.0 0.9747060
##    0.5 0.9827267
##    1.0 0.9873434
##    1.5 0.9886486
##    2.0 0.9868768
```

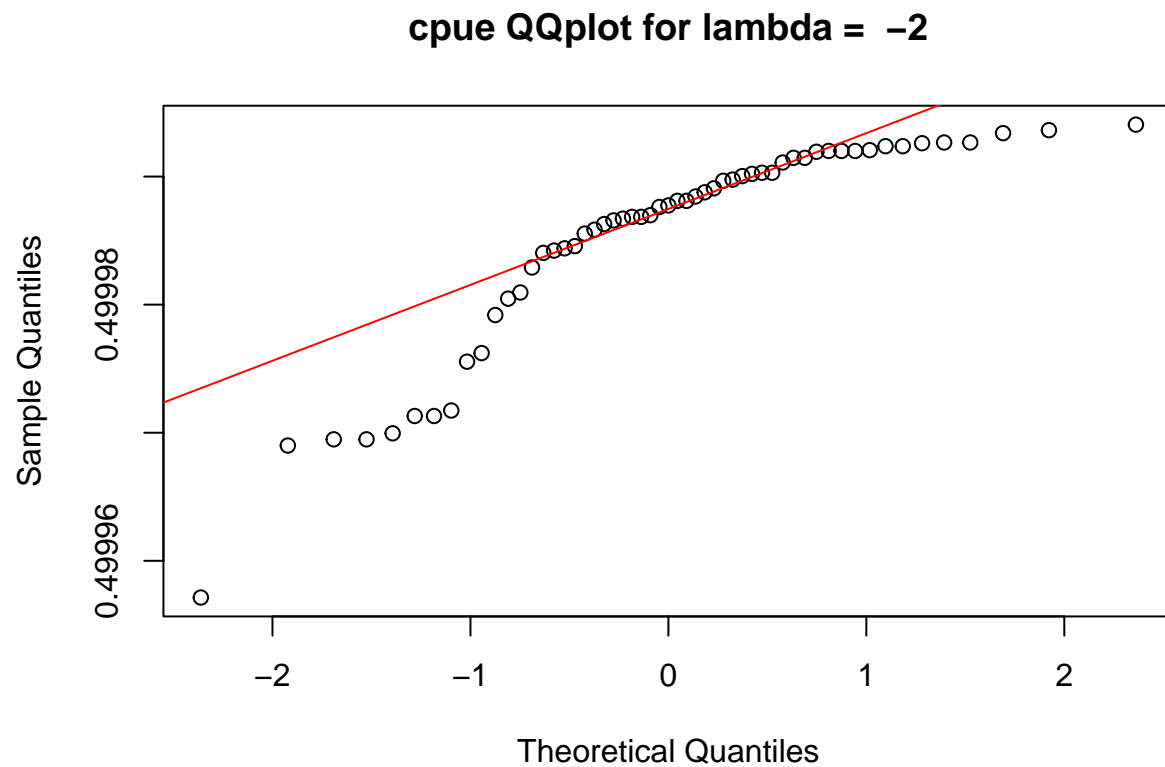
```
plot(box_cpue)
```



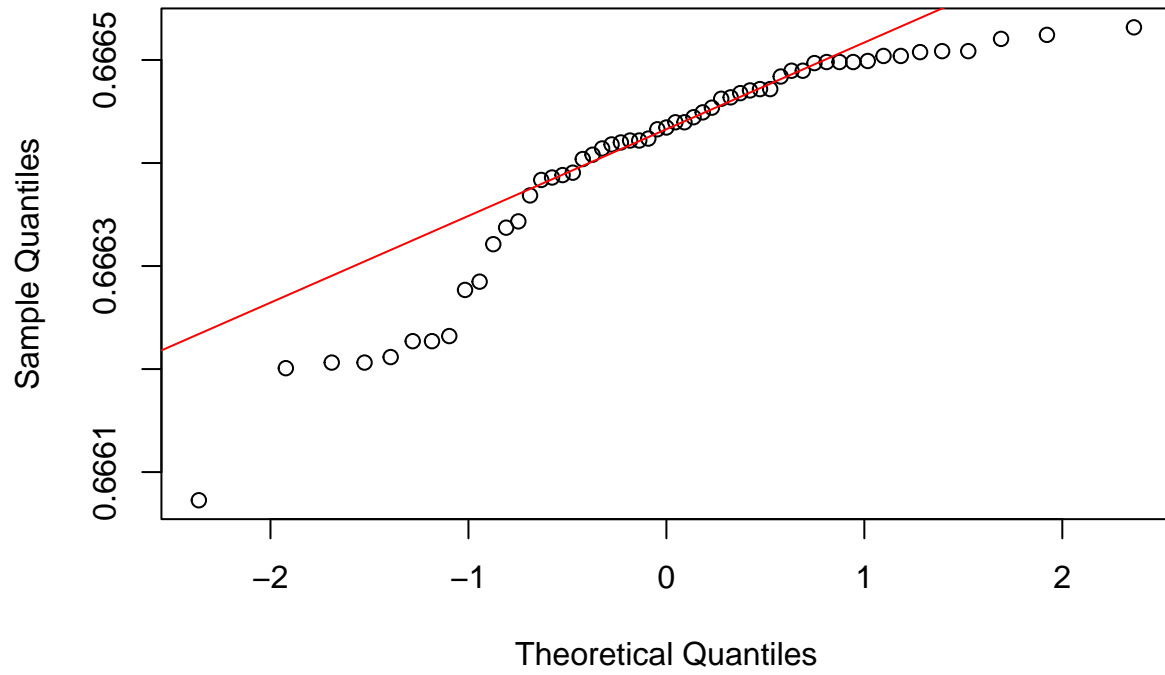
```
#Retrieve the optimal lambda with the highest PPCC value
cat("From the box_cpue output and plot, it is read that 1.5 produces the highest PPCC  
value. Thus select 1.5 to be the optimal lambda")
```

```
## From the box_cpue output and plot, it is read that 1.5 produces the highest PPCC
## value. Thus select 1.5 to be the optimal lambda
```

```
#transform cpue with every value of lambda and qqplot every transformation
for(i in seq(-2, 2, by=0.5)){
  cpue_transform <- boxcoxTransform(halibut$cpue, lambda = i)
  qqnorm(cpue_transform, main=paste("cpue QQplot for lambda = ", i))
  qqline(cpue_transform, col = "red")
}
```

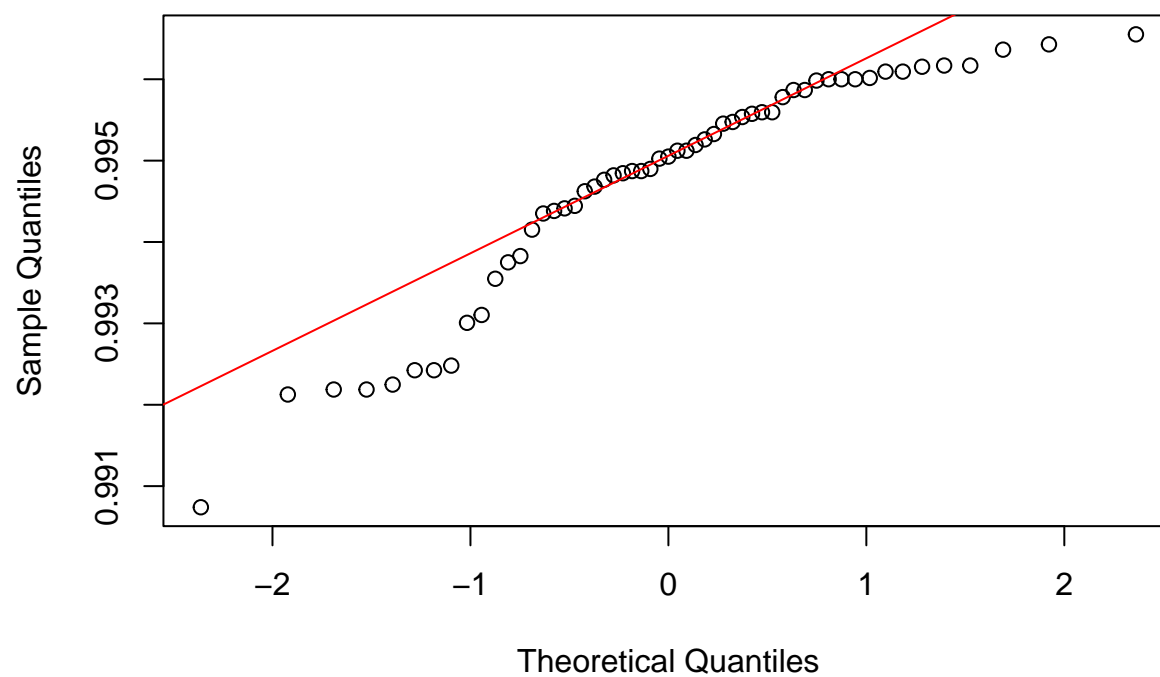


**cpue QQplot for lambda = -1.5**

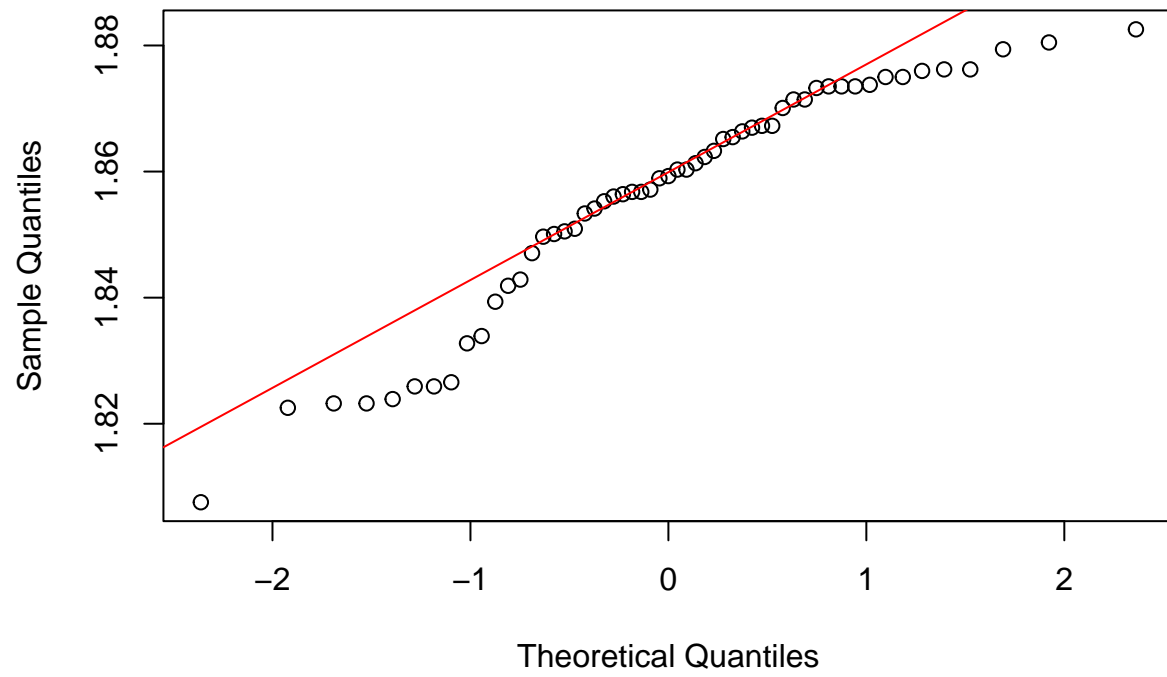




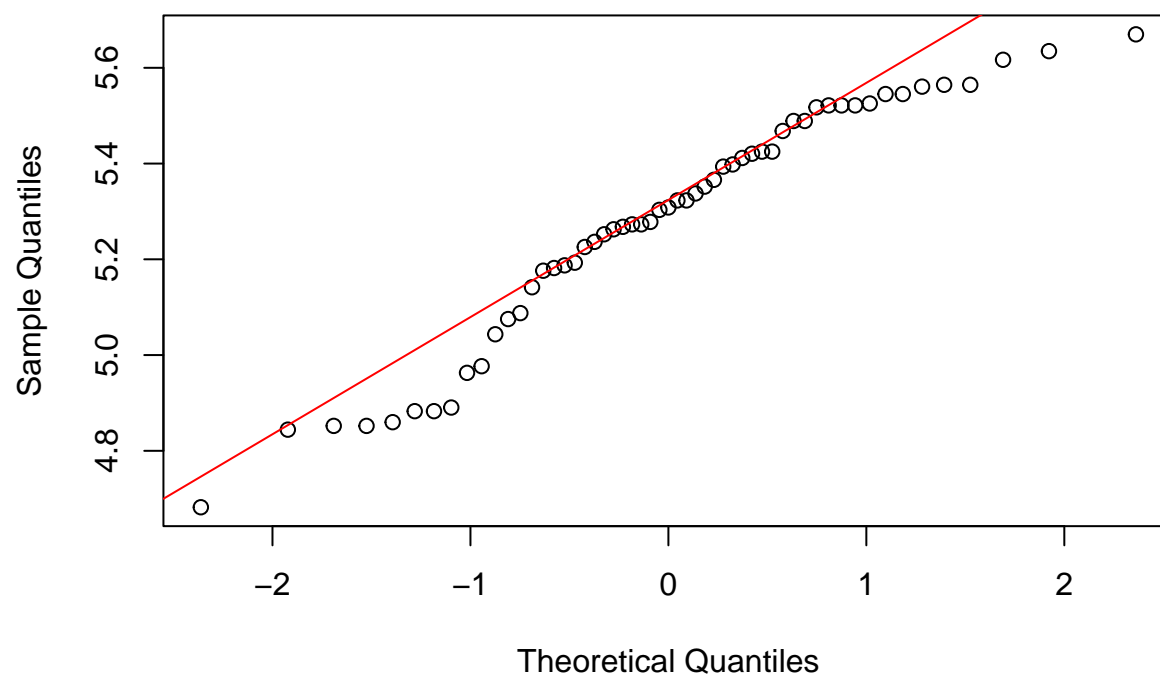
**cpue QQplot for lambda = -1**



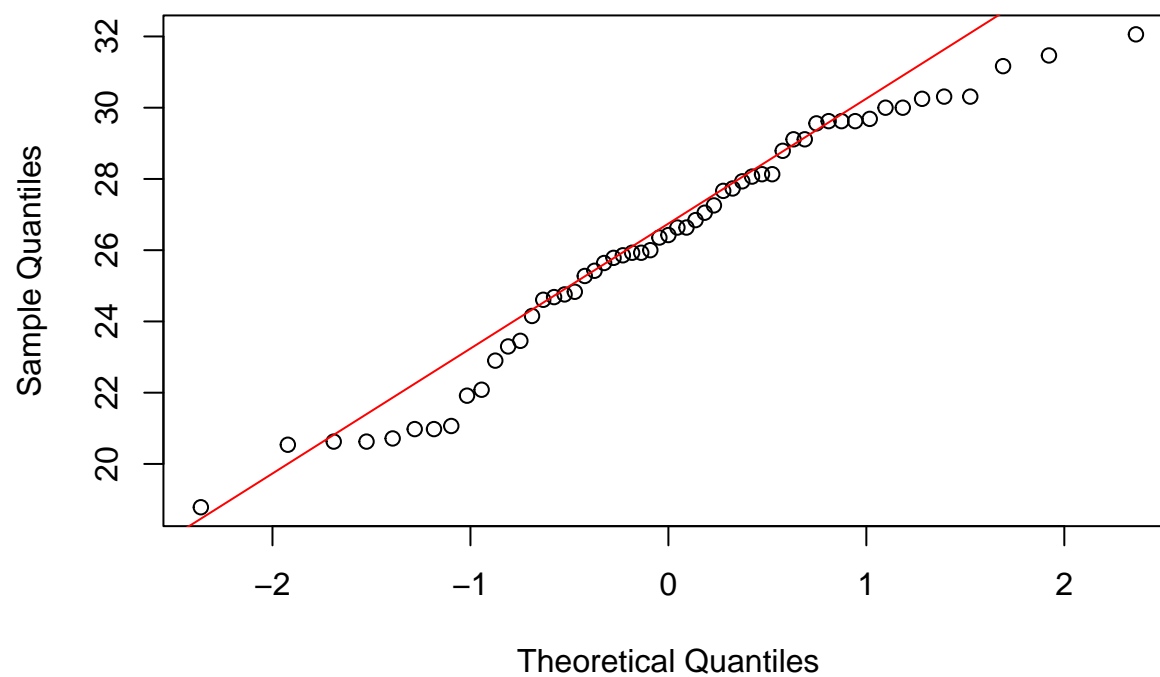
**cpue QQplot for lambda = -0.5**



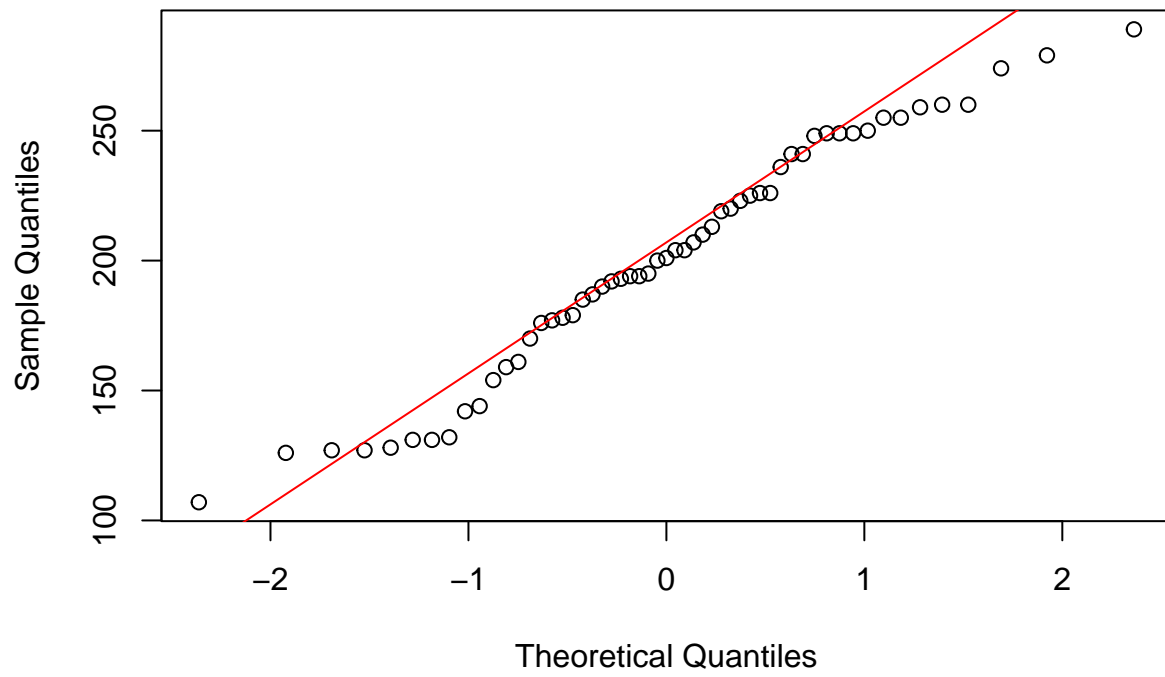
**cpue QQplot for lambda = 0**



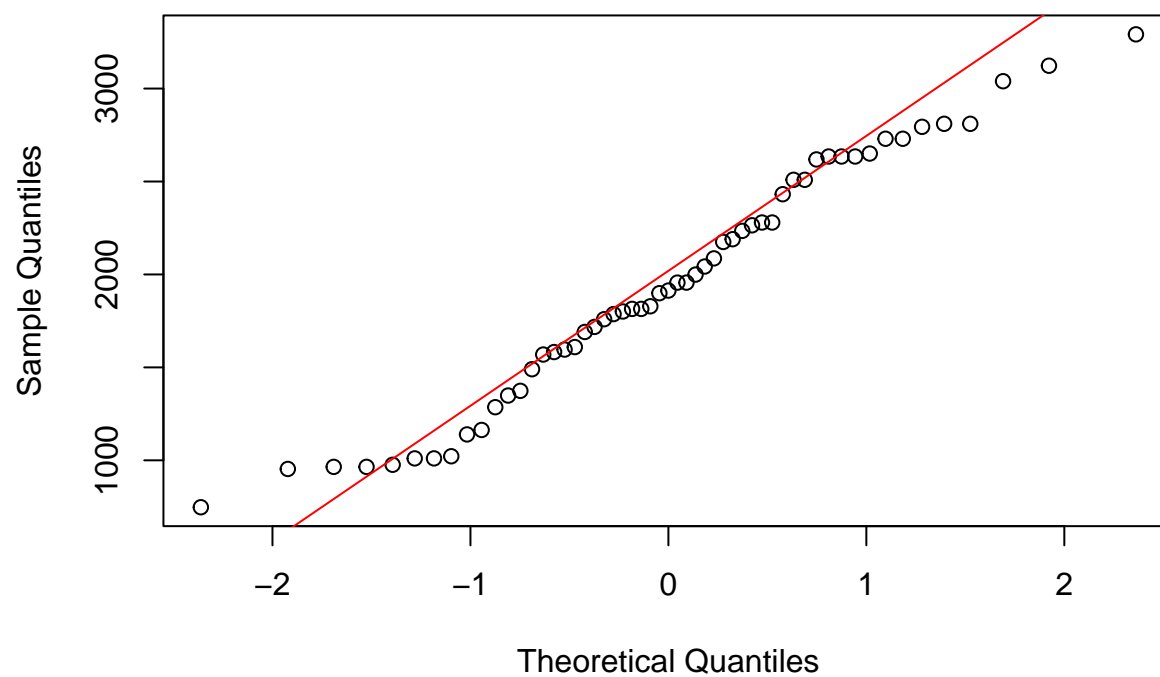
**cpue QQplot for lambda = 0.5**



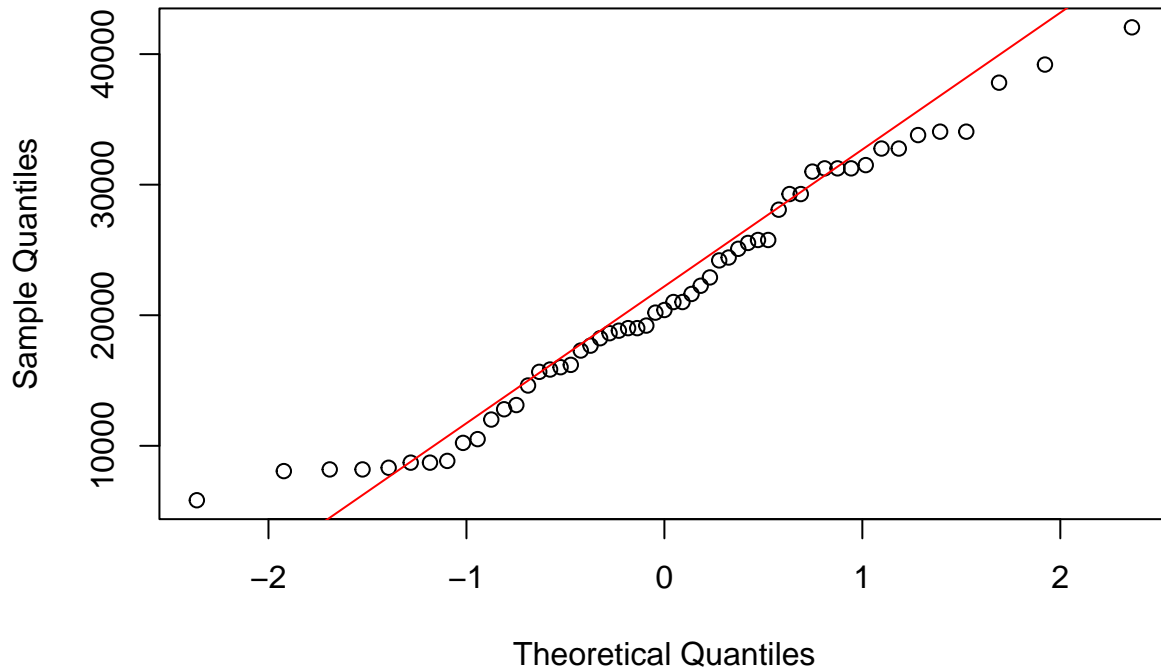
**cpue QQplot for lambda = 1**



**cpue QQplot for lambda = 1.5**



## cpue QQplot for lambda = 2



```
cat("Based on the results of the lambda vs PPCC plot, and the qqplots of
all the transformations of cpue for each lambda, lambda=1.5 produced the
highest value of PPCC, and the plots for the transformations of lambdas 0-2 of
cpue were very similar, therefore lambda=1.5 with the highest PPCC value
appears to be the best transformation. A higher PPCC value implies
a closer match for the transformed data to be normally distributed.")
```

```
## Based on the results of the lambda vs PPCC plot, and the qqplots of
## all the transformations of cpue for each lambda, lambda=1.5 produced the
## highest value of PPCC, and the plots for the transformations of lambdas 0-2 of
## cpue were very similar, therefore lambda=1.5 with the highest PPCC value
## appears to be the best transformation. A higher PPCC value implies
## a closer match for the transformed data to be normally distributed.
```

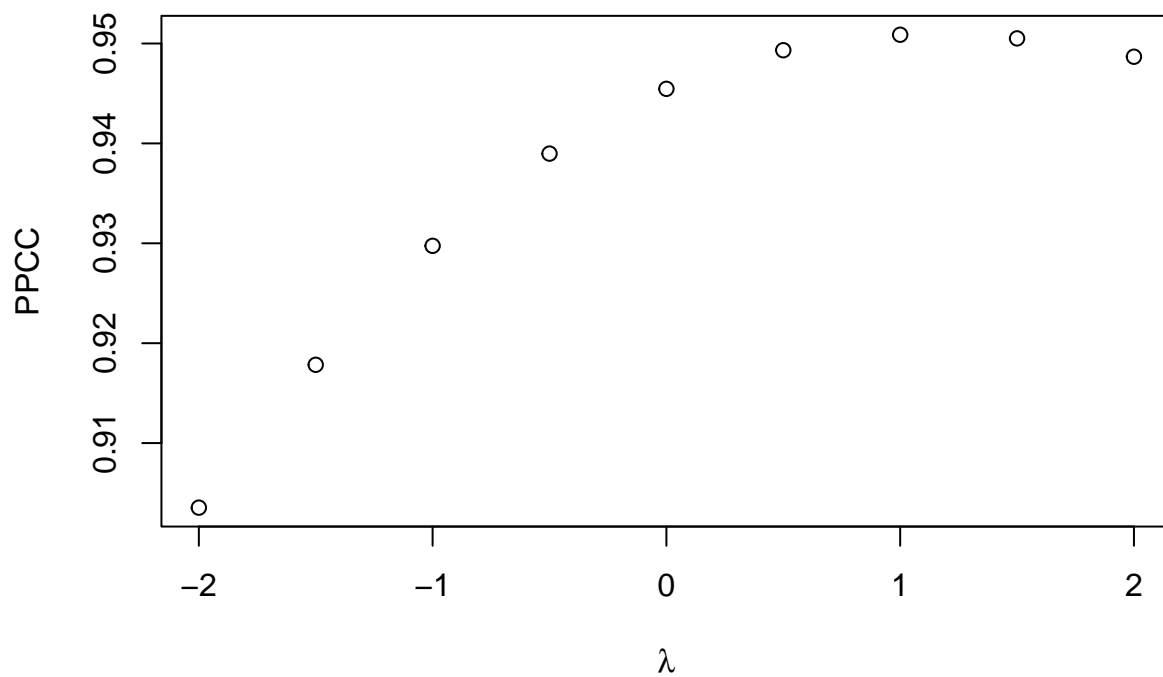
```
#perform boxcox transformation on cpue for (-2,2) incrementing by .5
box_bio <- boxcox(halibut$biomass, lambda = seq(-2, 2, by=0.5), plot = TRUE,
                 optimize = FALSE)
#plot lambda vs PPCC
box_bio
```

```
##
## Results of Box-Cox Transformation
## -----
##
## Objective Name:                PPCC
```

```
##
## Data:                  halibut$biomass
##
## Sample Size:          55
##
## lambda      PPCC
##   -2.0 0.9035375
##   -1.5 0.9178377
##   -1.0 0.9297465
##   -0.5 0.9389792
##    0.0 0.9454617
##    0.5 0.9493274
##    1.0 0.9508767
##    1.5 0.9505129
##    2.0 0.9486756
```

```
plot(box_bio)
```

### Box-Cox Transformation Results: PPCC vs. lambda for halibut\$biomass



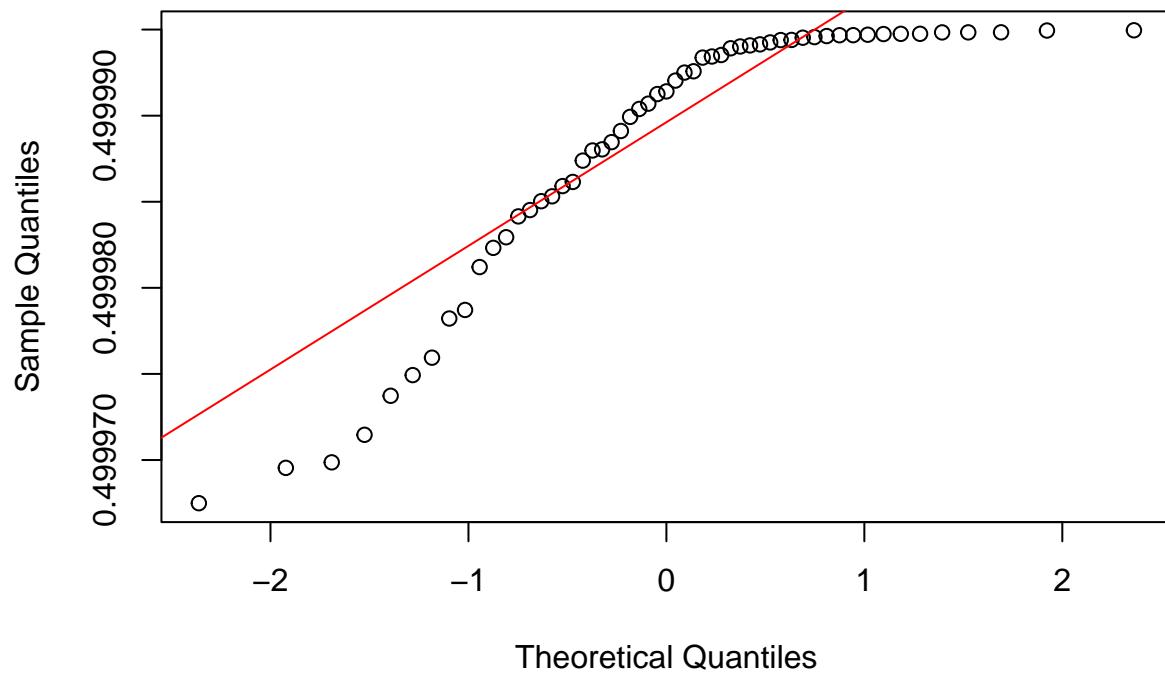
```
#Retrieve the optimal lambda with the highest PPCC value
cat("From the box_bio output and plot, it is read that 1 produces the highest PPCC  
value. Thus select 1 to be the optimal lambda")
```

```
## From the box_bio output and plot, it is read that 1 produces the highest PPCC  
## value. Thus select 1 to be the optimal lambda
```

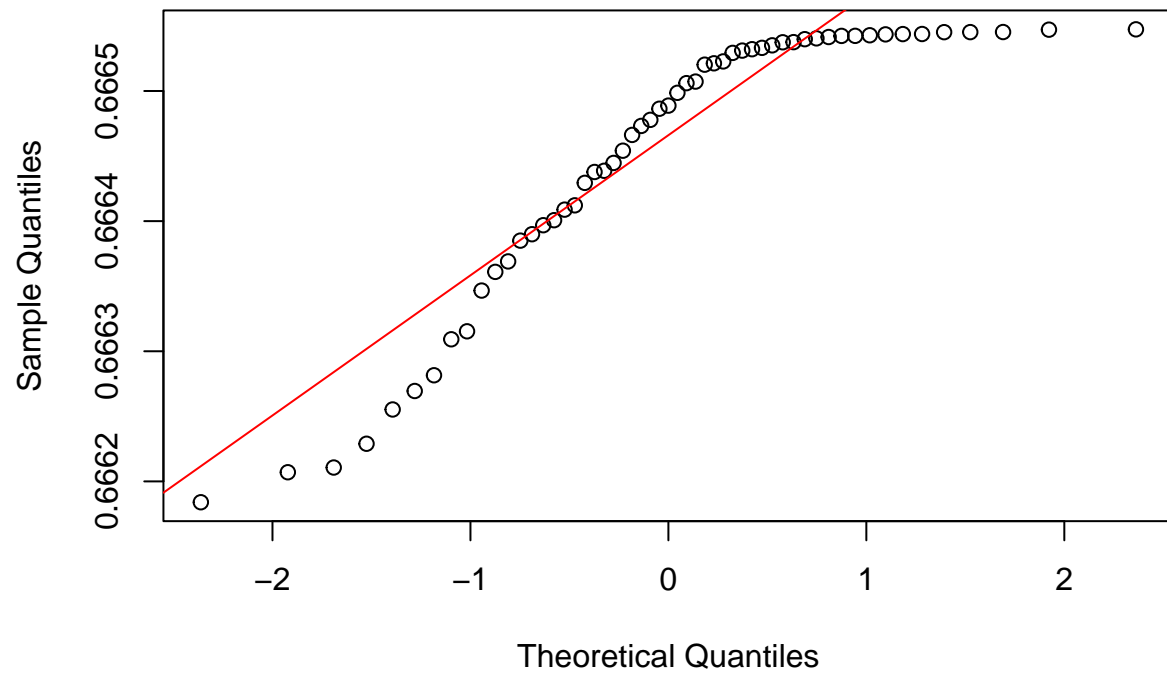


```
#transform biomass with every value of lambda and qqplot every transformation
for(i in seq(-2, 2, by=0.5)){
  bio_transform <- boxcoxTransform(halibut$biomass, lambda = i,)
  qqnorm(bio_transform, main=paste("biomass QQplot for lambda = ", i))
  qqline(bio_transform, col = "red")
}
```

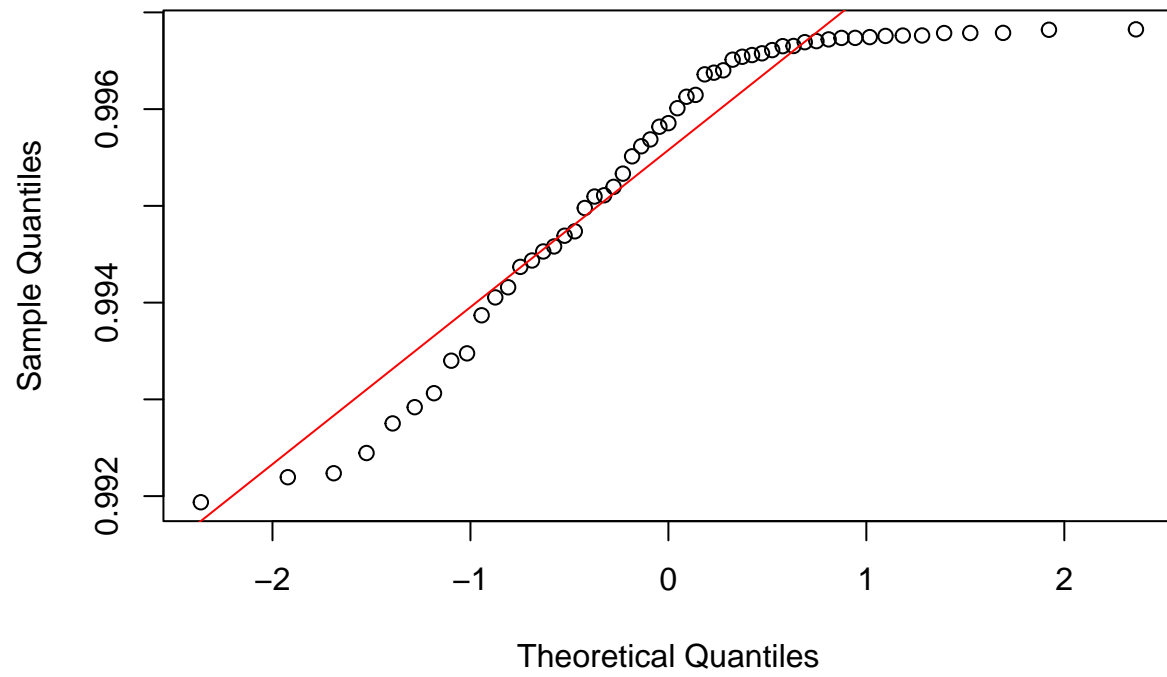
**biomass QQplot for lambda = -2**



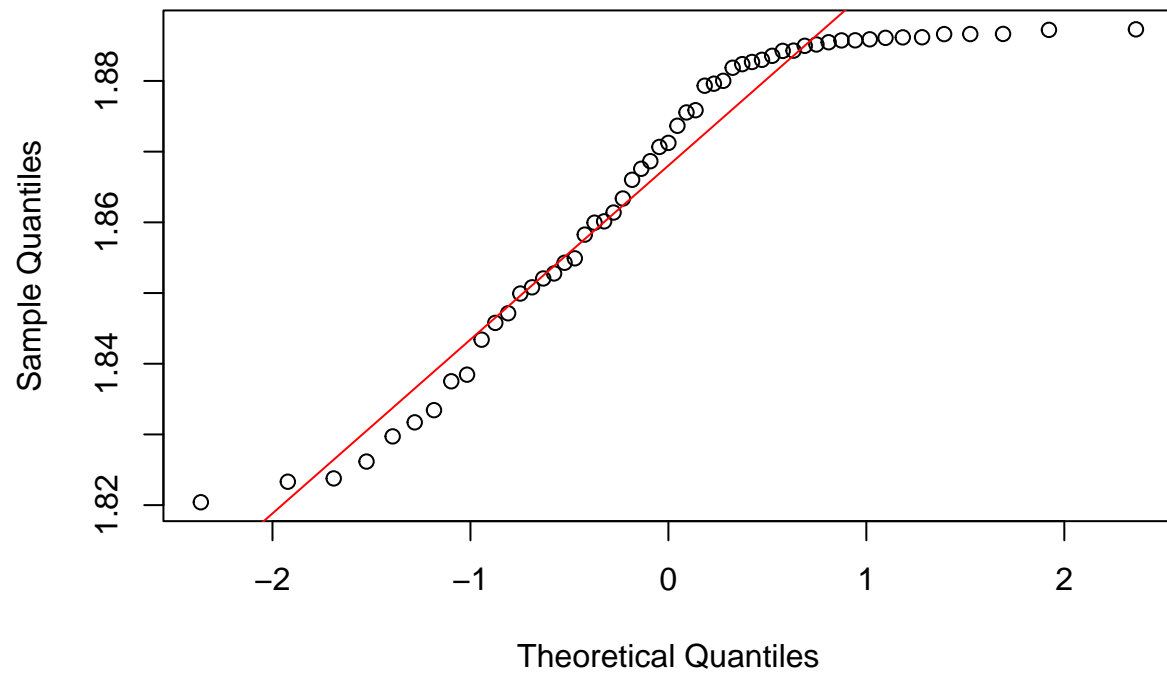
**biomass QQplot for lambda = -1.5**



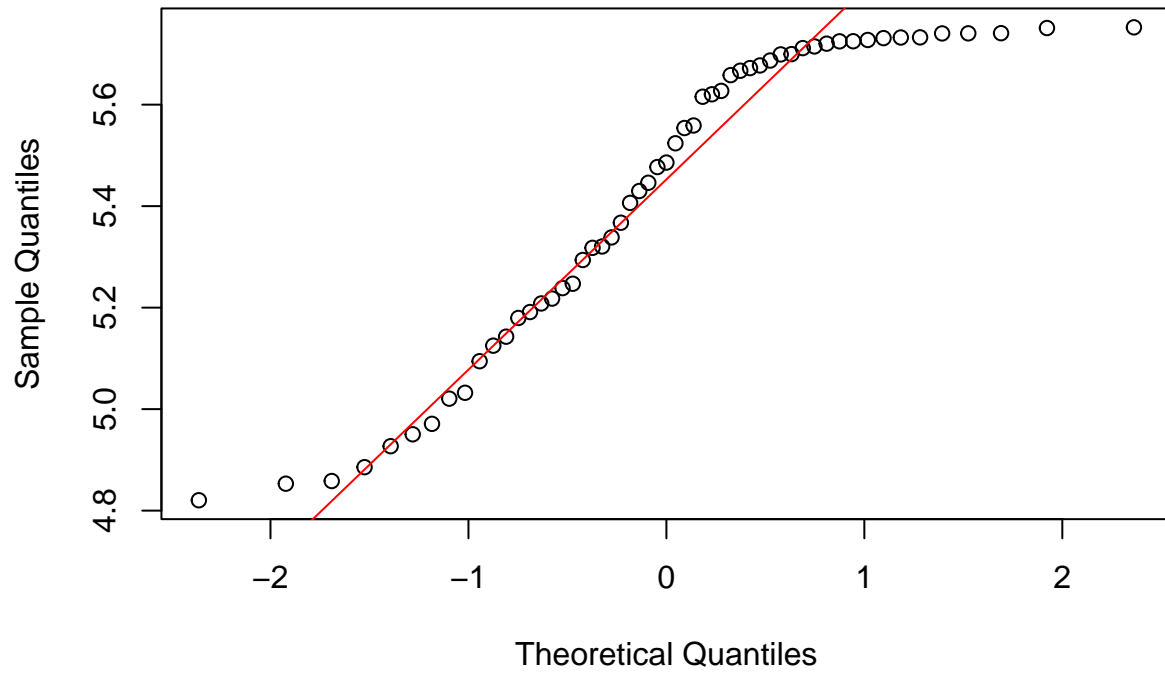
**biomass QQplot for lambda = -1**



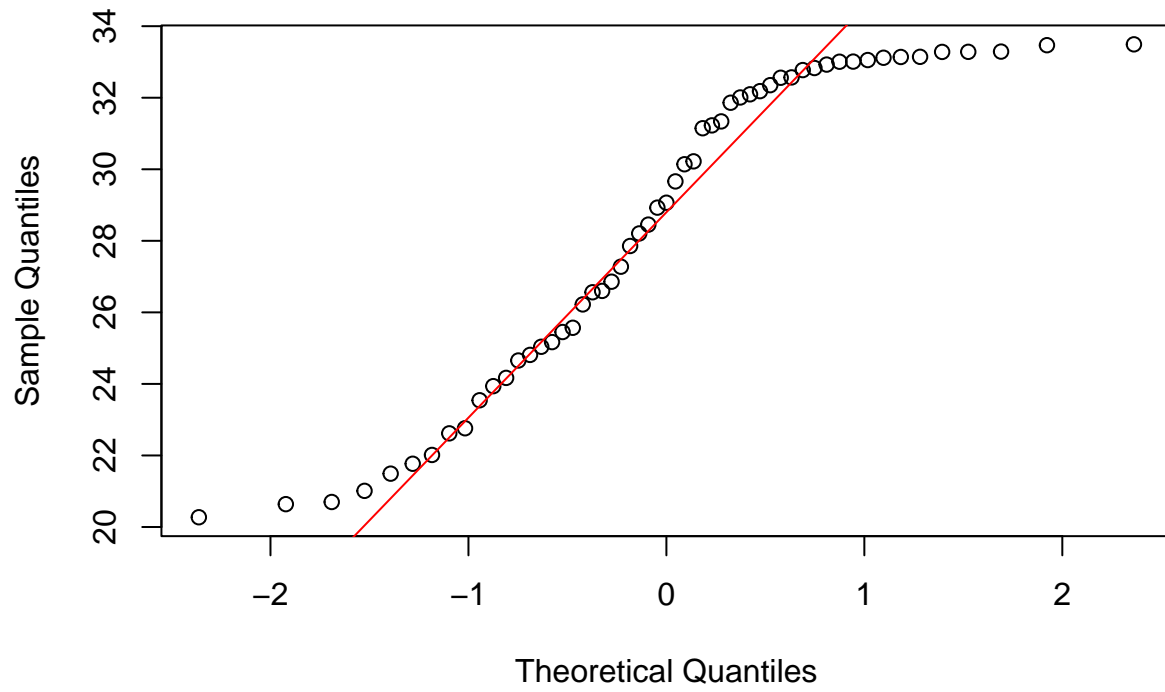
**biomass QQplot for lambda = -0.5**



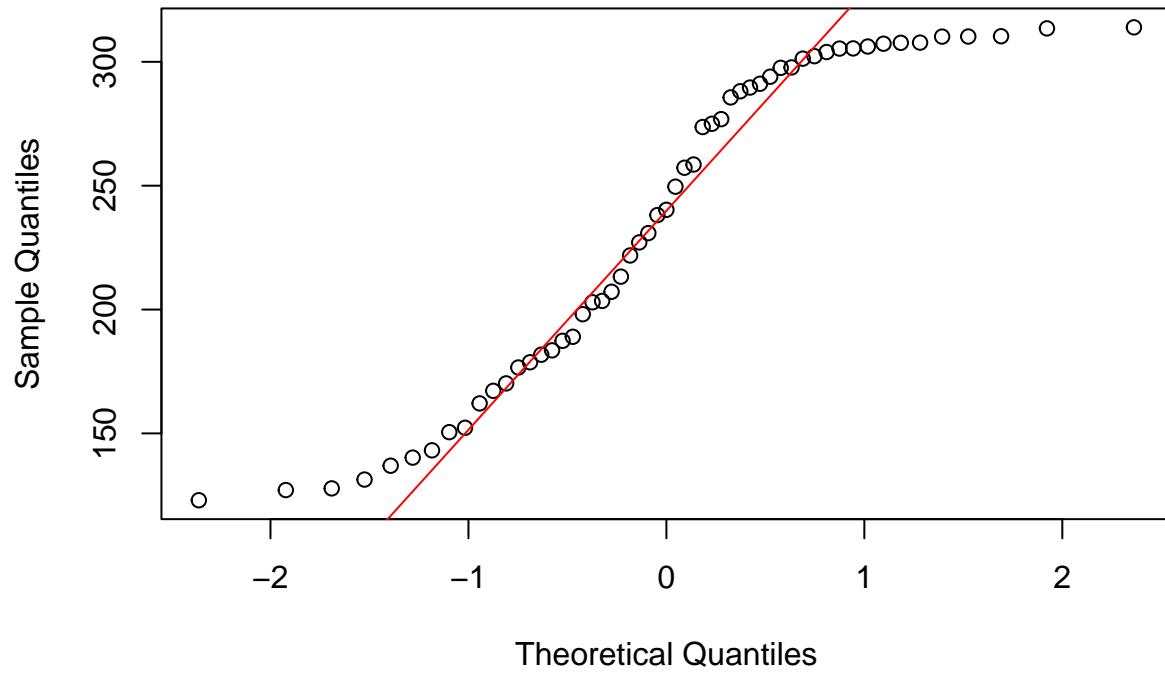
**biomass QQplot for lambda = 0**



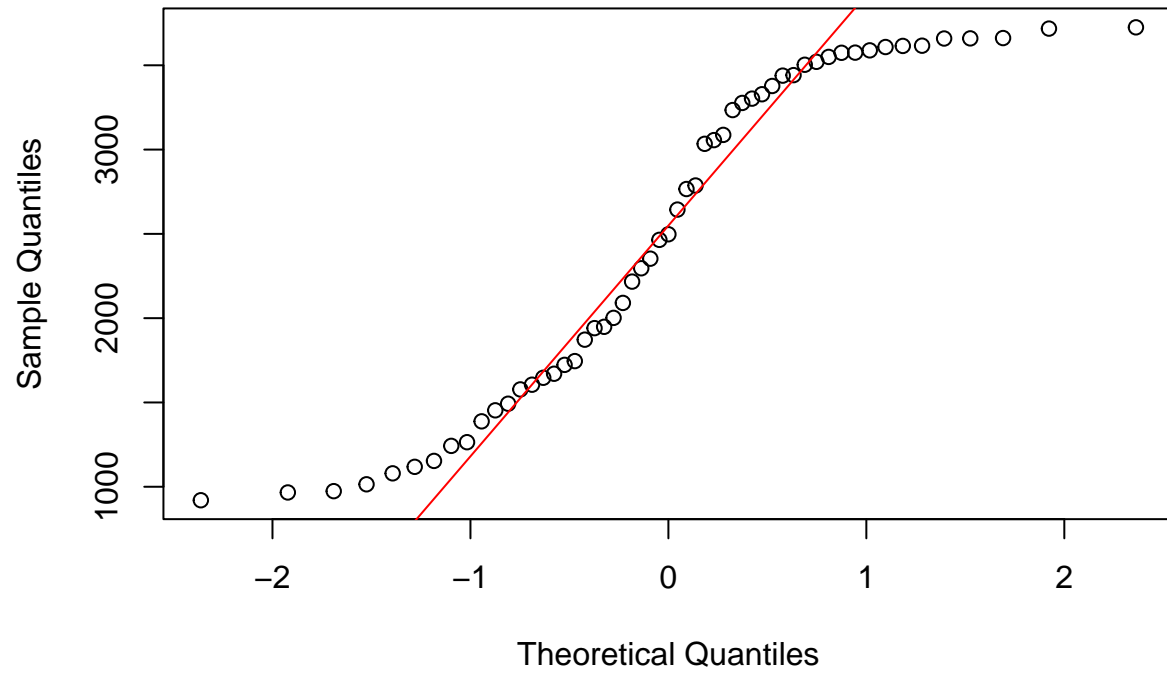
**biomass QQplot for lambda = 0.5**



**biomass QQplot for lambda = 1**

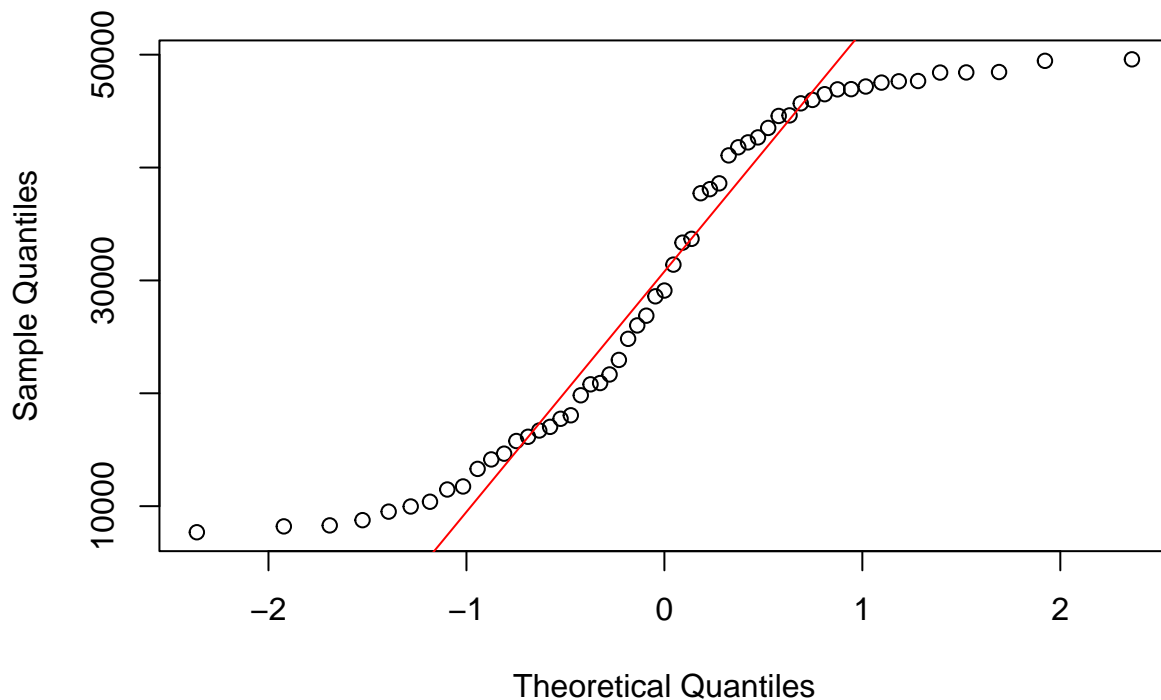


**biomass QQplot for lambda = 1.5**





## biomass QQplot for lambda = 2



```
cat("Based on the results of the lambda vs PPCC plot, and the qqplots of
all the transformations of biomass for each lambda, lambda=1 produced the
highest value of PPCC, and the plots for transformations of lambdas 0-2 of biomass
were very similar, therefore lambda=1 with the highest PPCC value appears to be
the best transformation. A higher PPCC value implies a closer match for
the transformed data to be normally distributed.")
```

```
## Based on the results of the lambda vs PPCC plot, and the qqplots of
## all the transformations of biomass for each lambda, lambda=1 produced the
## highest value of PPCC, and the plots for transformations of lambdas 0-2 of biomass
## were very similar, therefore lambda=1 with the highest PPCC value appears to be
## the best transformation. A higher PPCC value implies a closer match for
## the transformed data to be normally distributed.
```

```
# #3 - Use the function stat.desc() of R (install the relevant R package as necessary) and obtain
# the output of page 57 of the Exploring Data, Exploring Assumptions, Graphs.pdf for
# the variables CPUE and biomass. Provide the definition/formula for each of the statistic
# computed in this output. (These formulas, for the most part, are available in Millard and
# Neerchal book. You should be able to look up the index and find the formula.) Draw
# conclusions on the normality (or lack of) of the variables.
```

```
#install.packages("pastecs")
library(pastecs)
statdesc <- stat.desc(cbind(halibut$cpue, halibut$biomass), basic = FALSE, norm = TRUE)
```

```
colnames(statdesc) <- c("cpue","biomass")
statdesc
```

```
##              cpue      biomass
## median      202.000000  2.412600e+02
## mean        201.9272727  2.364964e+02
## SE.mean      6.2541490  8.716848e+00
## CI.mean.0.95 12.5388139  1.747623e+01
## var         2151.2909091  4.179089e+03
## std.dev      46.3820106  6.464588e+01
## coef.var     0.2296966  2.733483e-01
## skewness     -0.1648767 -2.789871e-01
## skew.2SE     -0.2562250 -4.335570e-01
## kurtosis     -0.9698153 -1.452879e+00
## kurt.2SE     -0.7654336 -1.146695e+00
## normtest.W   0.9676733  8.908189e-01
## normtest.p   0.1445546  1.211366e-04
```

```
cat("Skewness: Measure of a data's skew, a skewness=0 means there is no skew,
positive skew is right skew, negative skew is left skew.")
```

```
## Skewness: Measure of a data's skew, a skewness=0 means there is no skew,
## positive skew is right skew, negative skew is left skew.
```

$$\text{Skewness} = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s} \right)^3$$

```
cat("Skew.2SE is skewness divided by 2 times its standard error.")
```

```
## Skew.2SE is skewness divided by 2 times its standard error.
```

$$\text{Skew.2SE} = \frac{\text{Skewness}}{(2)(\text{SE})}$$

```
cat("Kurtosis describes the distribution around the mean of the data, and the
tailedness(shape and tail length). A negative kurtosis implies small tails
and less outliers, a positive kurtosis implies large tails and more outliers")
```

```
## Kurtosis describes the distribution around the mean of the data, and the
## tailedness(shape and tail length). A negative kurtosis implies small tails
## and less outliers, a positive kurtosis implies large tails and more outliers
```

$$\text{Kurtosis} = \frac{\mu_4}{s^4}$$

```
cat("kurt.2SE is kurtosis divided by 2 times its standard error.")
```

```
## kurt.2SE is kurtosis divided by 2 times its standard error.
```

$$\text{kurt.2SE} = \frac{\frac{\mu_4}{s^4}}{(2)(\text{SE})}$$

```
cat("We can draw conclusions about the normality of cpue and biomass from their
skewness and kurtosis statistics.
cpue has a skewness of -0.1648767 and a kurtosis of -0.9698153, which shows that
the skewness is not very far from normality (no skew = 0), and the kurtosis
is not far from 0 (normality), and implies that the tails are light. Thus we can
say that there is a high probability of cpue being normally distributed.
biomass has a skewness of -.2789871 and a kurtosis of -1.452879. This also shows
that there is a slight left-skew for biomass and the skew is close to normality(0).
The kurtosis is also not extremely far from 0, but less than -1 or greater than 1
indicates tail behavior that does not match normal distribution. Therefore we
can say that biomass is less likely to be normally distributed than cpue, and is
more likely to not be normally distributed.
")
```

```
## We can draw conclusions about the normality of cpue and biomass from their
## skewness and kurtosis statistics.
## cpue has a skewness of -0.1648767 and a kurtosis of -0.9698153, which shows that
## the skewness is not very far from normality (no skew = 0), and the kurtosis
## is not far from 0 (normality), and implies that the tails are light. Thus we can
## say that there is a high probability of cpue being normally distributed.
## biomass has a skewness of -.2789871 and a kurtosis of -1.452879. This also shows
## that there is a slight left-skew for biomass and the skew is close to normality(0).
## The kurtosis is also not extremely far from 0, but less than -1 or greater than 1
## indicates tail behavior that does not match normal distribution. Therefore we
## can say that biomass is less likely to be normally distributed than cpue, and is
## more likely to not be normally distributed.
```