

**ΔΙΑΧΕΙΡΙΣΗ ΕΣΩΤΕΡΙΚΩΝ ΚΑΙ ΠΑΡΑΚΤΙΩΝ
ΥΔΑΤΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ - ΕΡΓΑΣΤΗΡΙΟ**

Σύντομη Επανάληψη Συναρτήσεων Ανάλυσης Δεδομένων

$\text{mean}(x)$ – αριθμητικός μέσος

$\text{mean}(x, 0.10)$ – τετριμμένος μέσος

$\text{median}(x)$ – διάμεσος

$\text{mIn}(x)$ – συχνότερα εμφανιζόμενη τιμή – απαιτεί εγκατάσταση του πακέτου `modeest`

$\text{min}(x)$ – ελάχιστη τιμή

$\text{max}(x)$ – μέγιστη τιμή

$\text{range}(x)$ – ελάχιστη και μέγιστη τιμή

$\text{sd}(x)$ – τυπική απόκλιση

$\text{var}(x)$ – διακύμανση

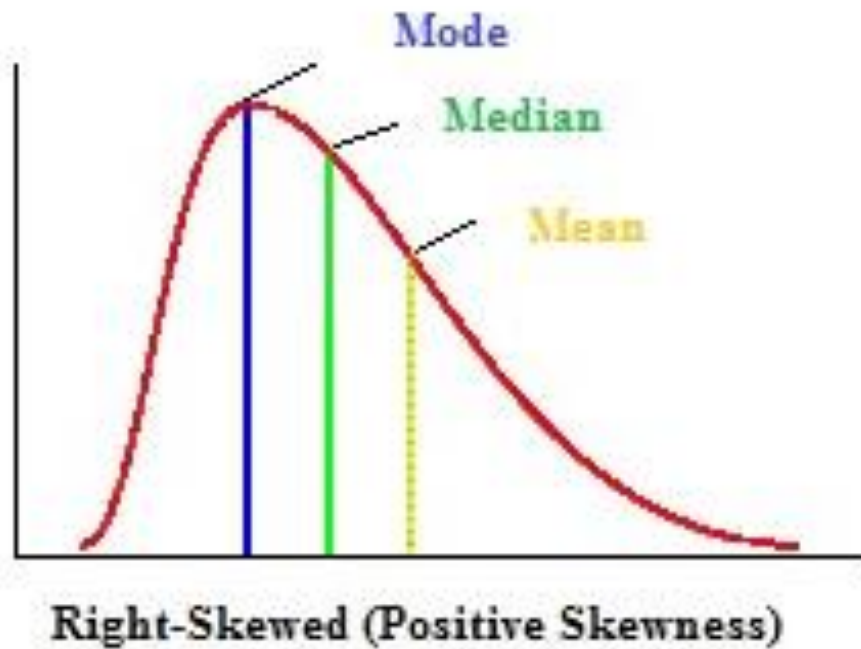
$\text{IQR}(x)$ – ενδοτεταρτομοριακό εύρος

$\text{sum}(x)$ – άθροισμα όρων διανύσματος

$\text{length}(x)$ – μέγεθος διανύσματος

$\text{sqrt}(x)$ – τετραγωνική ρίζα

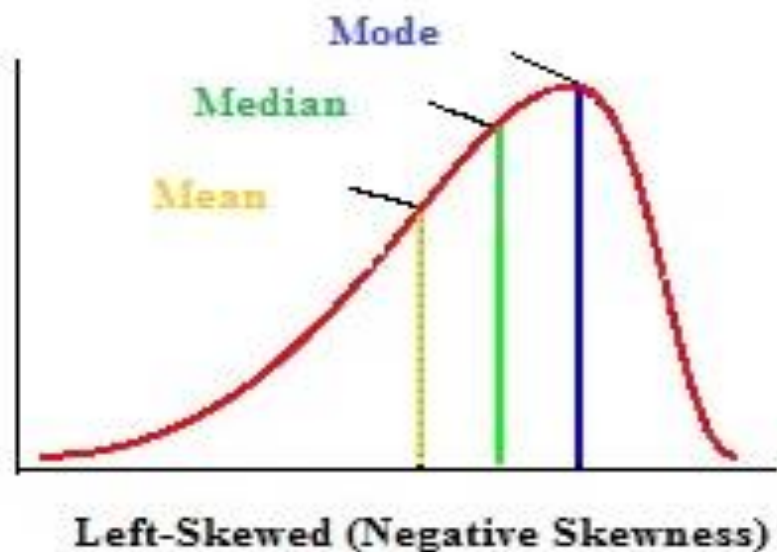
$\text{summary}(x)$ – περίληψη στατιστικών χαρακτηριστικών δεδομένων



Θετική ασυμμετρία δεδομένων

Αν η κατανομή παρουσιάζει θετική ασυμμετρία, τότε:

- A) έχει μία μεγάλη δεξιά ουρά, και
- B) ο αριθμητικός μέσος είναι μεγαλύτερος από τη διάμεσο και την επικρατούσα τιμή.



Αρνητική ασυμμετρία δεδομένων

Αν η κατανομή παρουσιάζει αρνητική ασυμμετρία, τότε:

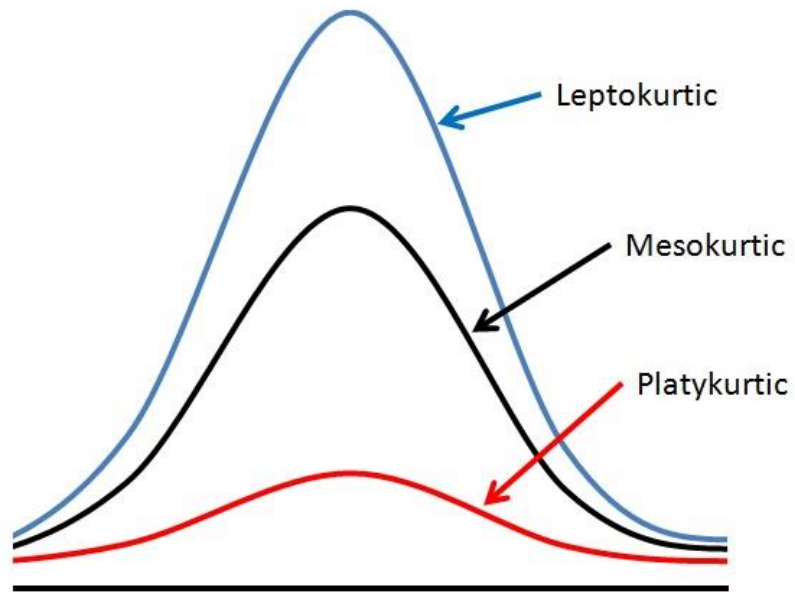
- A) έχει μία μεγάλη αριστερή ουρά, και
- B) ο αριθμητικός μέσος είναι μικρότερος από τη διάμεσο και την επικρατούσα τιμή.

Αυτό σημαίνει ότι αν γνωρίζουμε τις τιμές του μέσου, της διαμέσου και της επικρατούσας τιμής για μία κατανομή, μπορούμε να αποφανθούμε προς ποια διεύθυνση κινείται η ασυμμετρία της.

```
term1 = (fish-mean.fish)
term2 = term1^3
term3 = sum(term2)
term4 = 1/(length(fish))*term3
term5 = term1^2
term6 = sum(term5)
term7 = 1/length(fish)*term6
term8 = (term7)^(3/2)
skewness1 = term4/term8
```

```
library(e1071)
```

```
skewness(fish,type=1)
```



Μία κατανομή μπορεί να έχει μία από τις τρεις παρακάτω μορφές:

- Λεπτόκυρτη, όταν η κατανομή έχει σχετική υψηλή κορυφή,
- Μεσόκυρτη, όταν δεν είναι ούτε πολύ υψηλή ούτε πολύ χαμηλή.
- Πλατύκυρτη, όταν η κατανομή είναι επίπεδη και η κορυφή χαμηλή.

`library(e1071)`

`kurtosis(fish, type=1)`

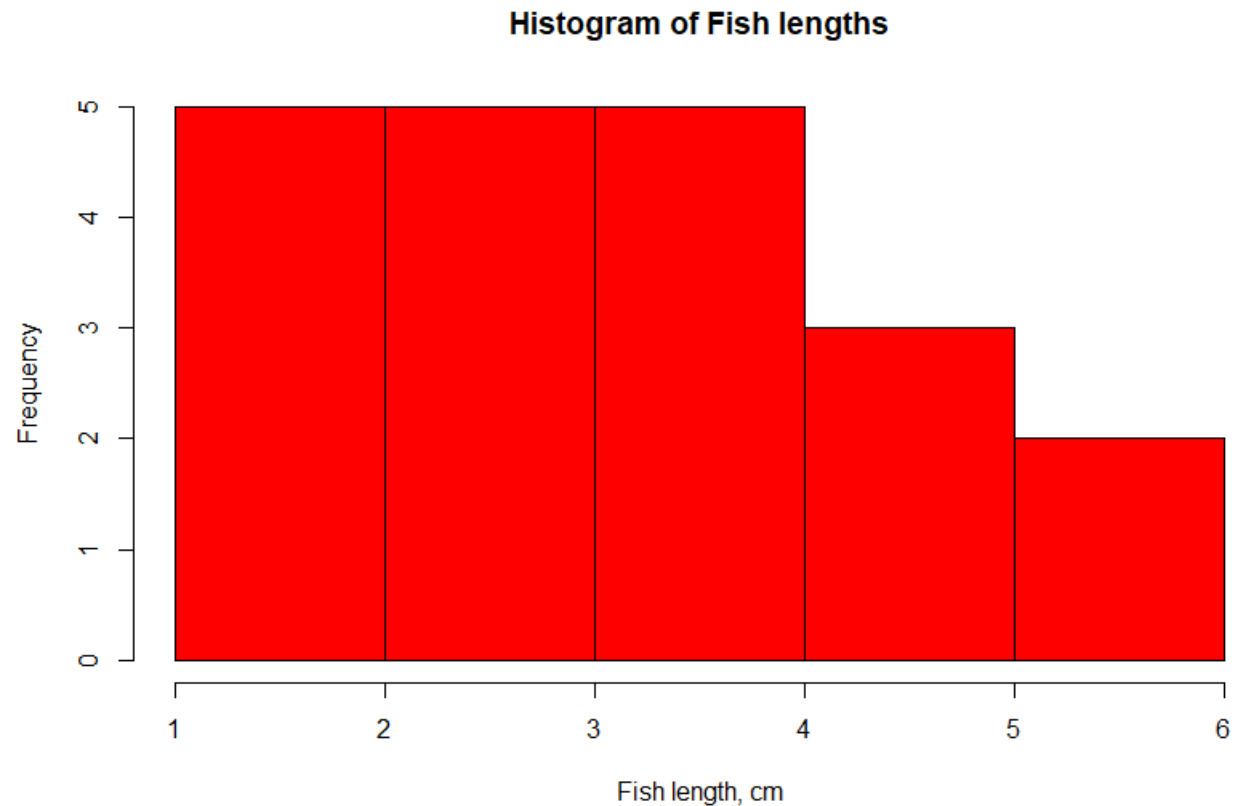
2. Το Ιστόγραμμα (Histogram) με πλήρεις παραμέτρους και καθορισμένες τάξεις

#Input fish length data

```
fish = c(1.1,1.3,1.7,1.8,1.9,2.1,2.3,2.4,2.5,2.8,3.1,3.3,3.6,3.7,3.9,4.1,4.5,4.8,5.1,5.3)
```

#Create a histogram

```
hist(fish,  
     breaks = c(1,2,3,4,5,6),  
     main="Histogram of Fish lengths",  
     xlab="Fish length, cm",  
     xlim=c(1,6),  
     col="red")
```



2. Το Διάγραμμα Θηκόγραμμα (Box-Whisker) στην R

#Input the data

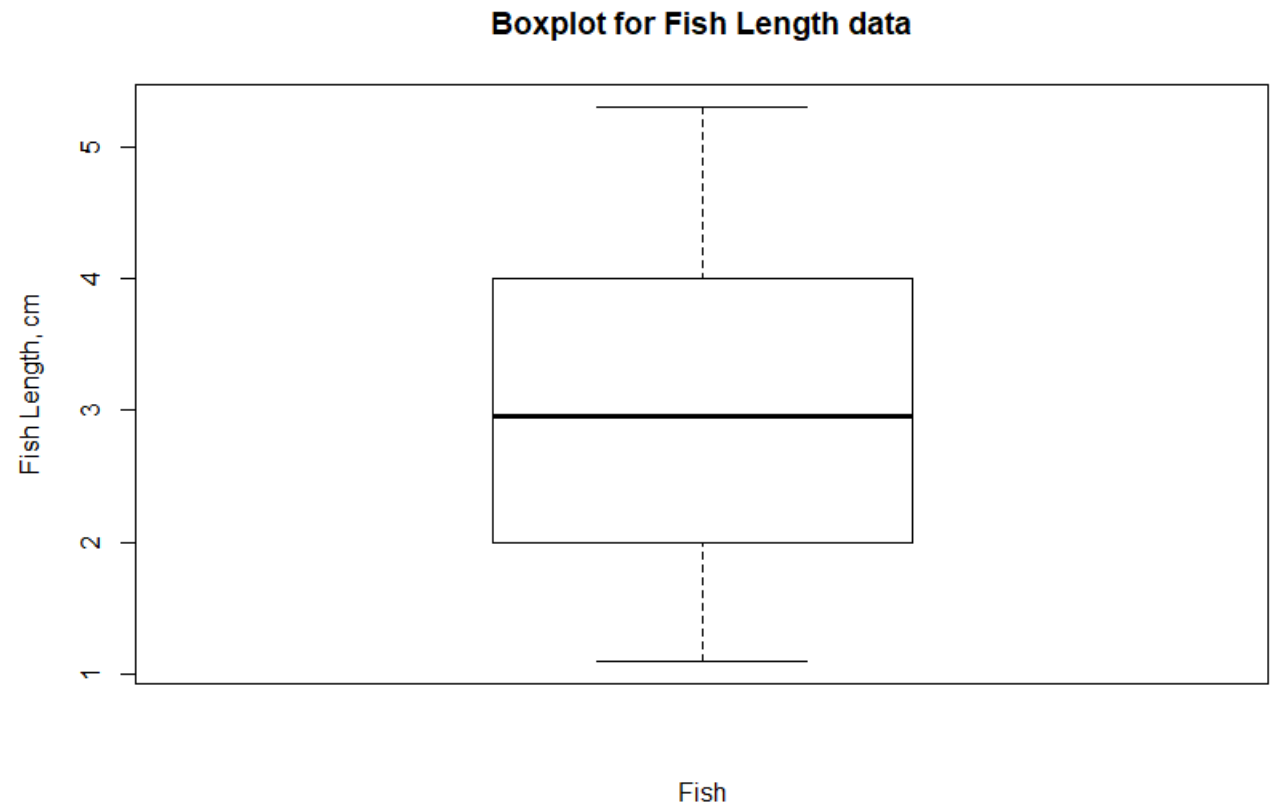
```
fish = c(1.1,1.3,1.7,1.8,1.9,2.1,2.3,2.4,2.5,2.8,3.1,3.3,3.6,3.7,3.9,4.1,4.5,4.8,5.1,5.3)
```

#Create a box-whisker plot

```
boxplot(fish,  
        main = "Boxplot for Fish Length data",  
        xlab = "Fish",  
        ylab = "Fish Length, cm"  
)
```

upper whisker = $\min(\max(x), Q_3 + 1.5 * IQR)$

lower whisker = $\max(\min(x), Q_1 - 1.5 * IQR)$



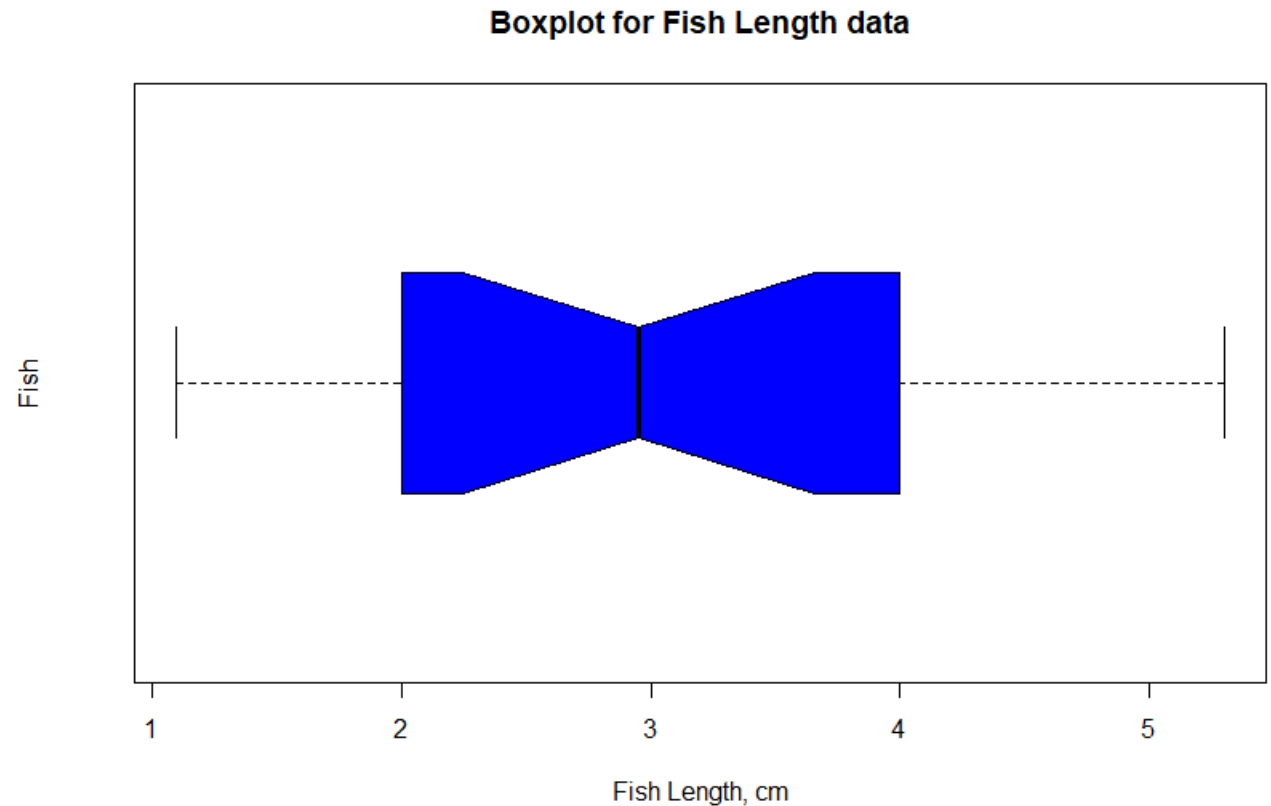
2. Βελτιώνοντας το Θηκόγραμμα (Box-Whisker) στην R

```
#Input the data
```

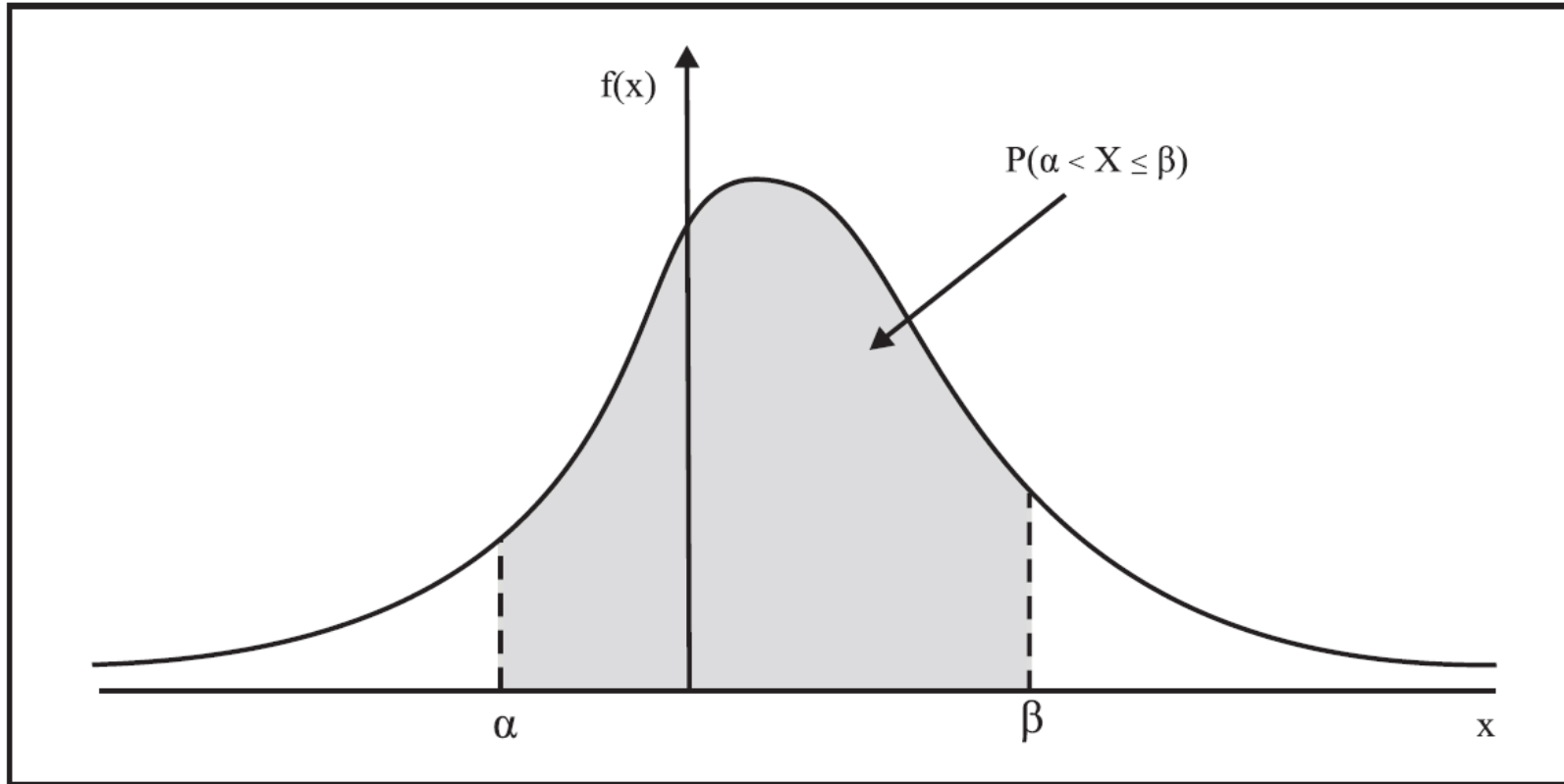
```
fish = c(1.1,1.3,1.7,1.8,1.9,2.1,2.3,2.4,2.5,2.8,3.1,3.3,3.6,3.7,3.9,4.1,4.5,4.8,5.1,5.3)
```

```
#Create a box-whisker plot
```

```
boxplot(fish,  
        main = "Boxplot for Fish Length data",  
        xlab = "Fish Length, cm",  
        ylab = "Fish",  
        col = "blue",  
        horizontal = TRUE,  
        notch = TRUE  
)
```



Διάγραμμα Συνάρτησης Πυκνότητας Πιθανότητας



Η πιθανότητα $P(\alpha < X \leq \beta)$ απεικονίζεται γραφικά ως το εμβαδόν της επιφάνειας που περιέχεται μεταξύ της καμπύλης που ορίζει η συνάρτηση πυκνότητας πιθανότητας $f(x)$ και του άξονα των x .

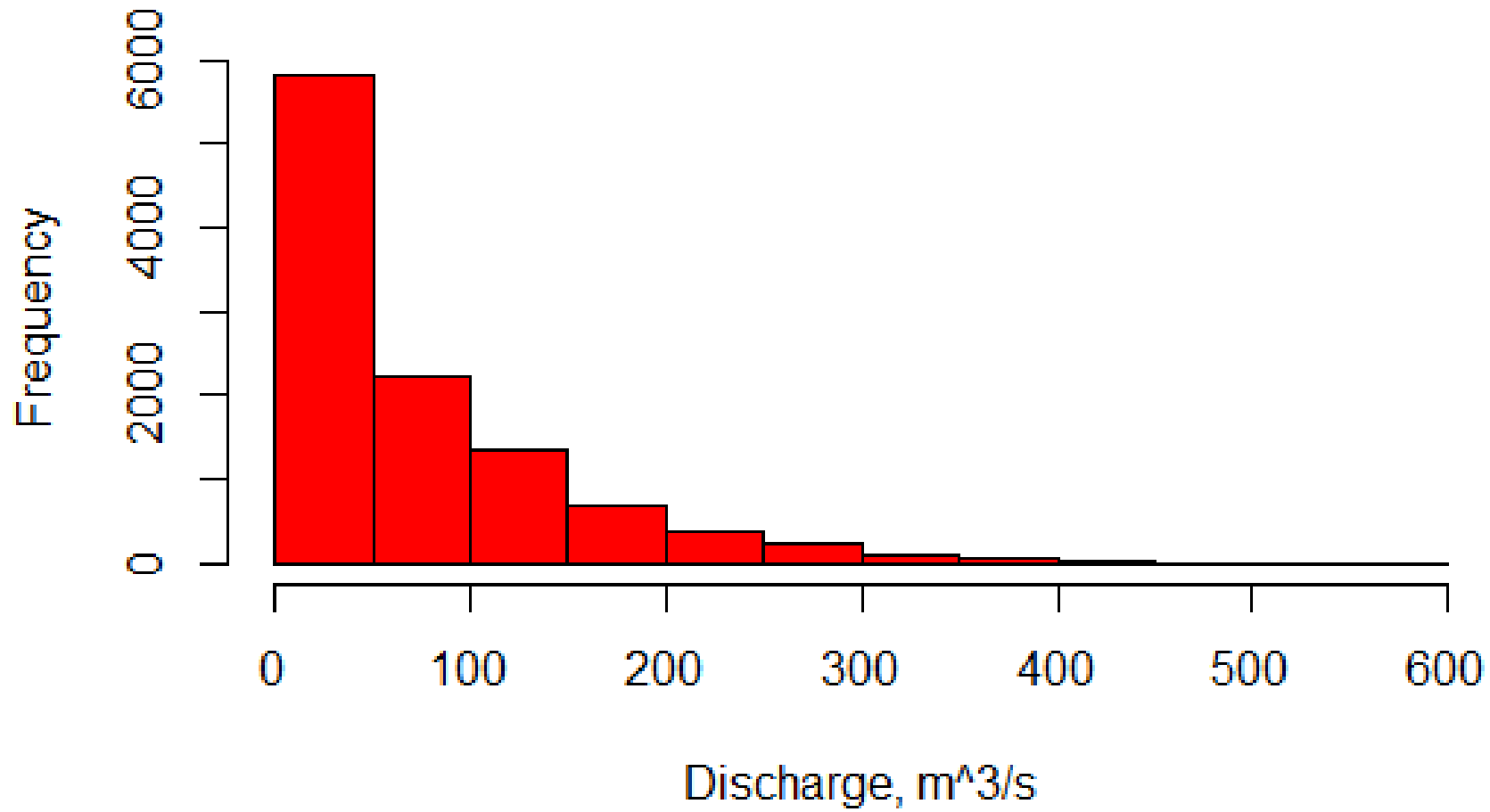
```
library(fitdistrplus)
```

```
plotdist(flow, histo = TRUE, demp = TRUE)
```

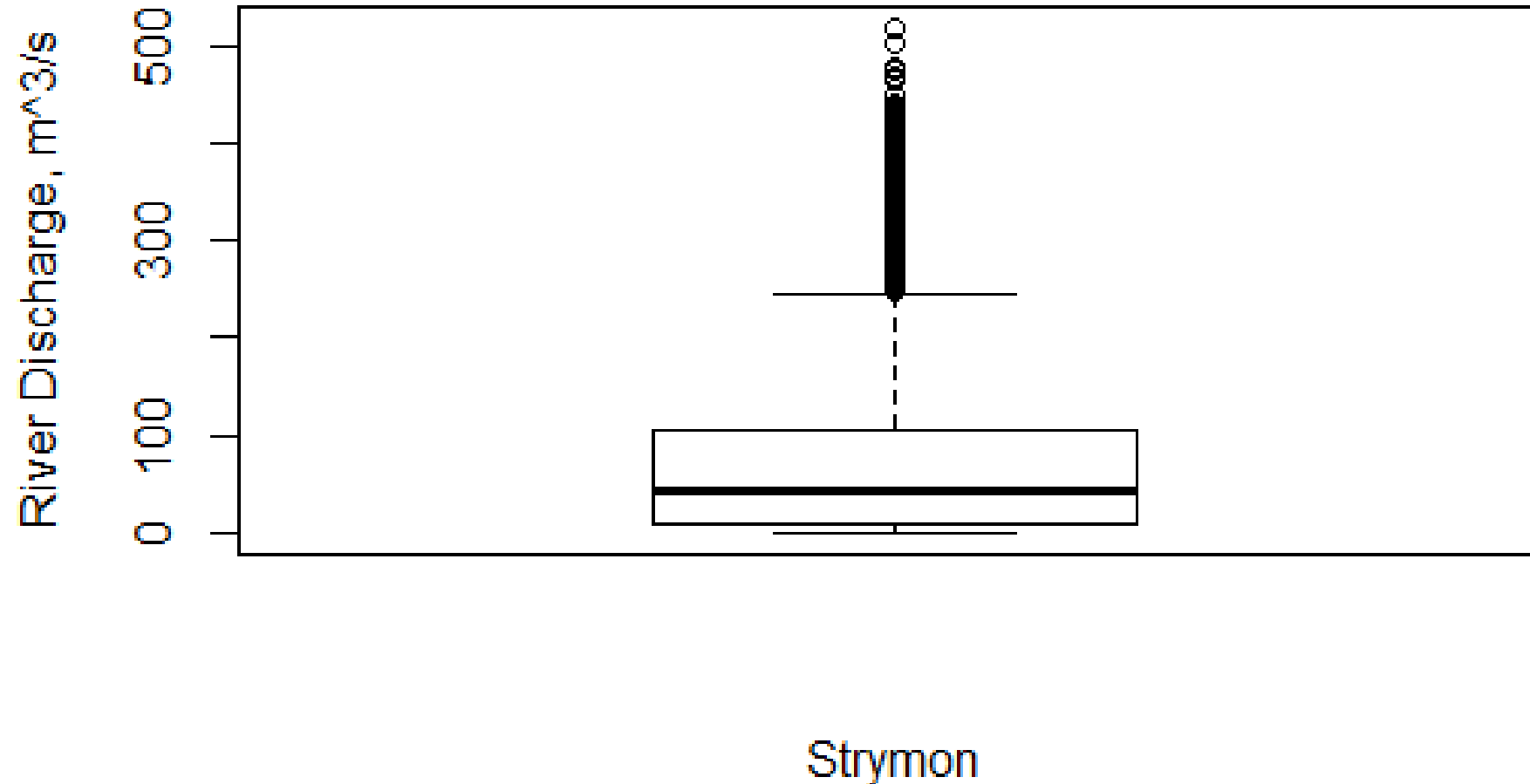
```
dweibull(seq(0,100,by=10),  
          shape=1, scale = 70.48704,  
          log = FALSE)
```

```
pweibull(100,  
          shape=1, scale = 70.48704,  
          log = FALSE,  
          lower.tail=TRUE)
```

Histogram of Strymon River Discharge



Boxplot for Strymon Discharge data

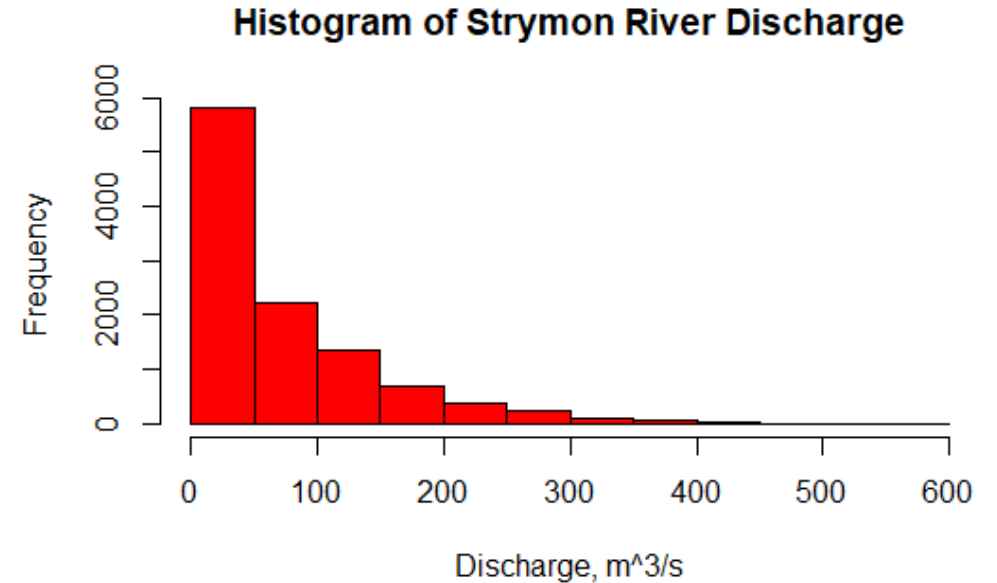


upper whisker = $\min(\max(x), Q_3 + 1.5 * IQR)$
lower whisker = $\max(\min(x), Q_1 - 1.5 * IQR)$

Weibull probability density function (PDF)

$$f(T) = \frac{\beta}{\eta} \left(\frac{T - \gamma}{\eta} \right)^{\beta-1} e^{-\left(\frac{T - \gamma}{\eta} \right)^\beta}$$

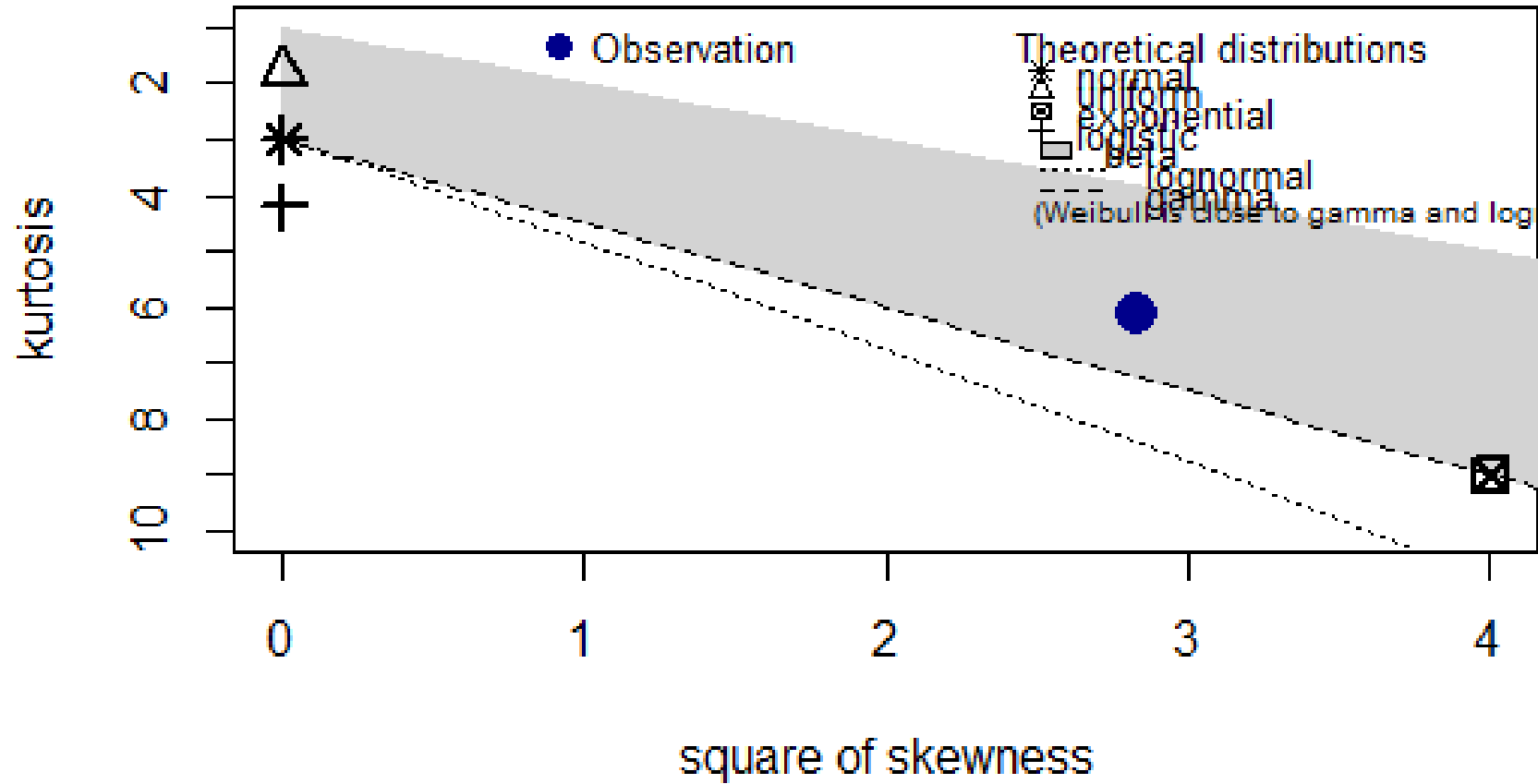
- β is the shape parameter, also known as the Weibull slope
- η is the scale parameter
- γ is the location parameter



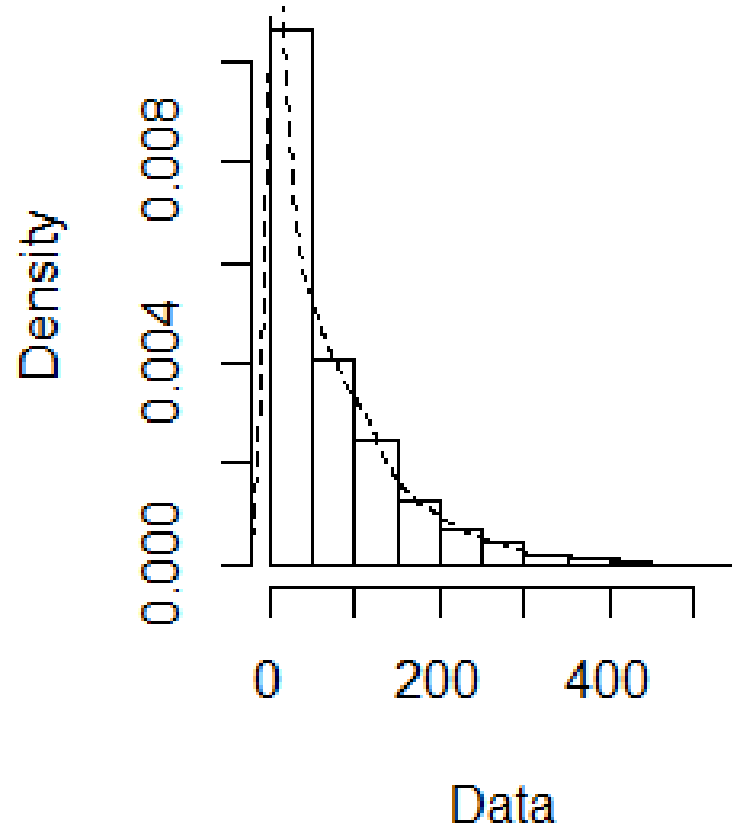
Συνήθως η παράμετρος θέσης δεν χρησιμοποιείται οπότε βάζουμε $\gamma = 0$.

Τότε η κατανομή λέγεται «κατανομή Weibull δύο παραμέτρων»

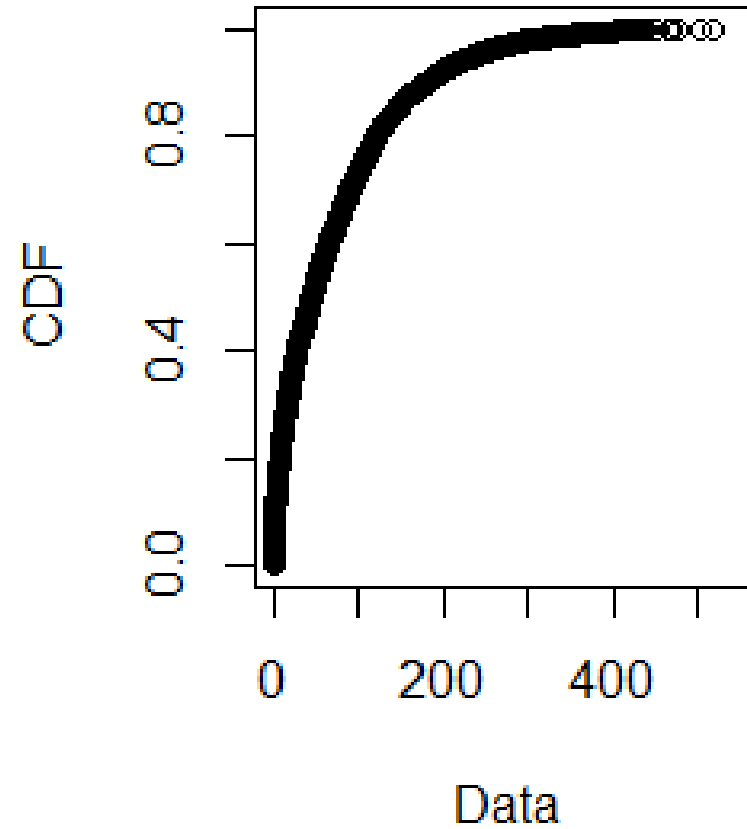
Cullen and Frey graph



Empirical density



Cumulative distribution



Κώδικας στην R για τον υπολογισμό περιγραφικών στατιστικών παραμέτρων, την δημιουργία στατιστικών γραφημάτων και την εύρεση του βέλτιστου μοντέλου συνάρτησης πυκνότητας πιθανότητας που εφαρμόζεται στα υδρολογικά δεδομένα.

```
#Define the working directory  
setwd("C:/Users/user/OneDrive/Notes/Χανθι/ΔιαχείρισηΥδατικώνΠόρων")
```

```
#Install required packages  
install.packages("e1071")  
install.packages("fitdistrplus")
```

```
#Call package libraries  
library(e1071)  
library(fitdistrplus)
```

```
#read the external data file  
flow1 = read.table("Strymon_discharge.txt", header=TRUE, dec=".")
```

```
#transform list to vector  
flow = flow1$Discharge
```

```
#Compute descriptive statistics
```

```
mean.flow = mean(flow)
```

```
median.flow = median(flow)
```

```
range = range(flow)
```

```
range.flow = range[2]-range[1]
```

```
sd.flow = sd(flow)
```

```
var.flow = var(flow)
```

```
IQR.flow = IQR(flow)
```

```
summary(flow)
```

```
#Compute skewness and kurtosis
```

```
skewness.flow = skewness(flow,type=1)
```

```
kurtosis.flow = kurtosis(flow,type=1)
```

```
#Plot histogram of Strymon data
```

```
hist(flow,
```

```
  main="Histogram of Strymon River Discharge",
```

```
  xlab="Discharge, m3/s",
```

```
  xlim=c(0,3000),
```

```
  breaks = seq(0,3000, by=100),
```

```
  col="red")
```

```
#Plot boxplot of Strymon data
boxplot(flow,
  main = "Boxplot for Strymon Discharge data",
  xlab = "Strymon",
  ylab = "River Discharge, m3/s")

#Plot histogram, probability density function
#and cumulative density distribution
plotdist(flow, histo = TRUE, demp = TRUE)

#Plot the location of Strymon data in relation to other
#theoretical probability density distributions
descdist(flow)

#Fit the most appropriate Weibull probability density
#distribution model to Strymon data
fw <- fitdist(flow1$Discharge,
  "weibull",
  method = "mle",
  lower=c(0,0),
  start=list(scale=1,shape=1))
```

```
#Get summary of Weibull model output  
summary(fw)
```

```
#Get the probability of an event  
#with discharge from 0 to 100 m3/s with step 50 m3/s  
dweibull(seq(0,100,by=50),  
          shape=1, scale = 70.48704,  
          log = FALSE)
```

```
#Get the cumulative probability of an event  
#with discharge higher than 100 m3/s  
pweibull(100,  
          shape=1, scale = 70.48704,  
          log = FALSE,  
          lower.tail=FALSE)
```

```
#Create 10 random values sampled from a population  
#of Weibull probability density distribution  
rweibull(n=10, shape=1, scale = 70.48704)
```