Tema 2

**Comanda ajutatoare pentru realizarea proiectului:**

*install.packages('e1071', repos='http://cran.us.r-project.org')*

**>faithful**

*Se afiseaza valorile din “faithful”*

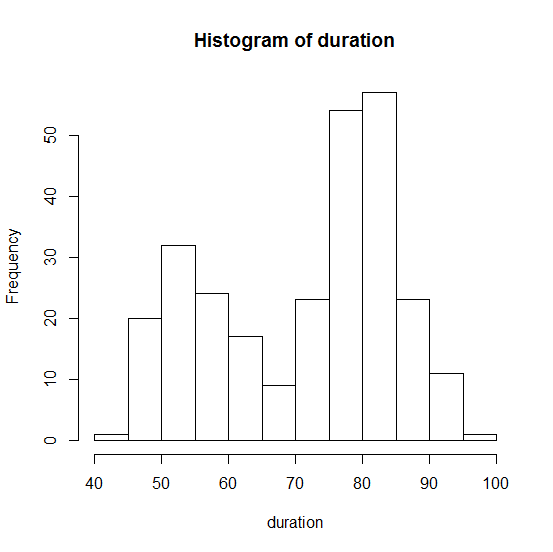
**>F=faithful**

*Stocam in variabila F datele din faithful*

**>duration=F$waiting**

**>hist(duration,right=FALSE)**

*Realizam histograma pentru coloana waiting*

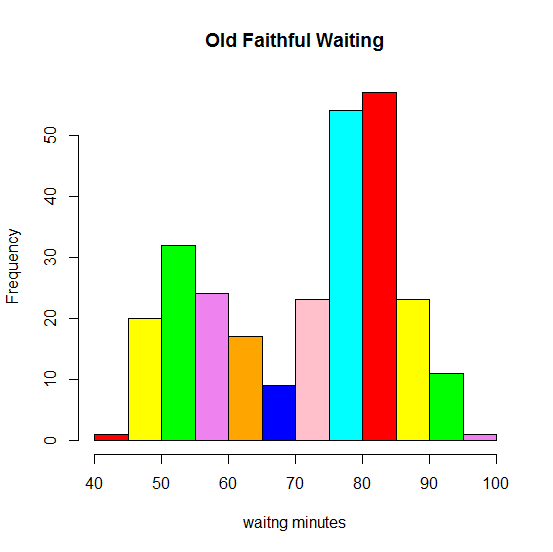


>

colors=c("red","yellow","green","violet","orange","blue","pink","cyan")

>hist(duration,right=FALSE,col=colors,main="Old Faithful Waiting",xlab="waitng minutes")

Coloram histograma.



***Cumulative frequency distribution***

>duration=F$waiting

> breaks=seq(10,110,by=10) #creez intervalul

> duration.cut=cut(duration,breaks,right=FALSE)

#divizam valorile din waiting conform intervalului,le punem la locul corespunzator.

> duration.freq=table(duration.cut)

> duration.cumfreq=cumsum(duration.freq)

#aplicam cumsum pt a calcula distributia cumulative de frecventa

>cbind(duration.cumfreq) #afisam sub forma de matrice

duration.cumfreq

[10,20) 0

[20,30) 0

[30,40) 0

[40,50) 21

[50,60) 77

[60,70) 103

[70,80) 180

[80,90) 260

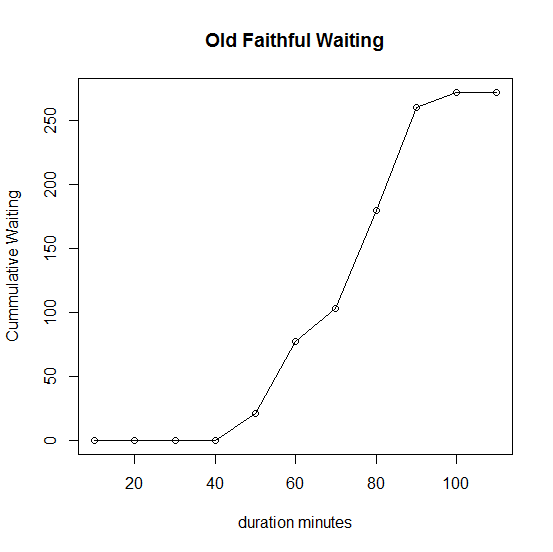
[90,100) 272

[100,110) 272

>cumfreq0=c(0,cumsum(duration.freq))

> plot(breaks, cumfreq0,main="Old Faithful Waiting", xlab="duration minutes", ylab="Cummulative Waiting")

> lines(breaks,cumfreq0) #unim prin linie



Media asteptarii:

> mean(duration)

[1] 70.89706

x(i)-perioada de asteptare intre eruptii

Valoarea centrala:

Mediana unei variabile aleatoare este t pt care are loc functia:

>median(duration)

[1] 76

quantile:

Quantilele sunt valori numerice care ımpart setul de date ın q grupe egale. Constanta q se numeste ordinul quantilei.

>quantile(duration)

0% 25% 50% 75% 100%

43 58 76 82 96

>quantile(duration, c(.22,.59, .97))

22% 59% 97%

56 78 90

Range:

Diferenta dintre valoarea minima si maxima

>max(duration)-min(duration)

[1] 53

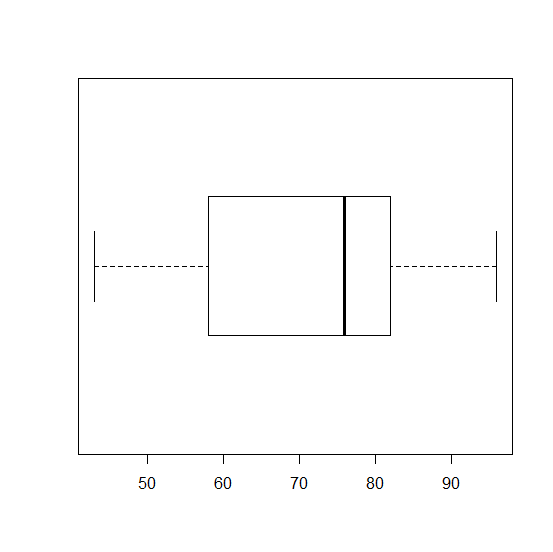
Interquantila:

Diferenta dintre Quantila superioara si Quantila inferioara

>IQR(duration)

[1] 24

> boxplot(duration, horizontal=TRUE)



Varianta:

Ne ajuta sa vedem cum o variabila se disperseaza in jurul mediei

var(duration)

[1] 184.8233

Deviatia Standard:

sd(duration)

[1] 13.59497

Covarianta:

)

cov(duration,eruptions)

[1] 13.97781

Corelatia:

Un indice de corelatie apropiat de 1 inseamna ca variabilele sunt corelate linear pozitiv.

Un indice de corelatie apropiat de 1 inseamna ca variabilele sunt corelate linear negativ.

Daca indicele de corelatie este apropiat de 0, nu putem spune ca exista o corelatie intre cele 2 variabile.

Sxy=s{x|y}/s{x}s{y}

cor(duration,eruptions )

[1] 0.9008112

Al treilea moment central:

M(k)=

>library(moments)

> duration=F$waiting

> moment(duration, order=3, central= TRUE)

[1] -1040.307

Asimetria:

O asimetrie negativa semnifica faptul ca media este mai mica decit mediana, rezultind o asimetrie stinga.

O asimetrie pozitiva semnifica faptul ca media este mai mare decit mediana, rezultind astfel o asimetrie dreapta.

>skewness(duration)

[1] -0.4163188

Coeficient de aplatizare:

Un coeficient negativ este specific unei distributii uniforme(fara punct de extrem global). - distributie platikurtica

Un coeficient pozitiv ne va indica prezenta unui punct de extrem global. - distributie leptokurtica

Z=

>kurtosis(duration)

[1] 1.857369

>summary(duration)

Min. 1st Qu. Median Mean 3rd Qu. Max.

43.0 58.0 76.0 70.9 82.0 96.0