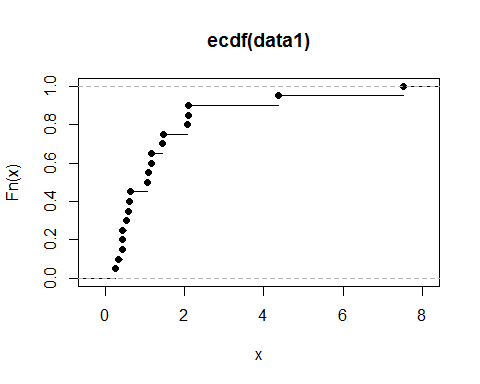
Assignment 6 (S520)

FNU Anirudh

October 15, 2015

# Solution 1

data1<- c(0.246,0.530,1.098,2.063,0.327,0.583,1.158,2.105,0.423,0.613,1.163,  
 2.106,0.425,0.641,1.439,4.363,0.434,1.054,1.464,7.517)  
# A) Empirical CDF  
fn=ecdf(data1)  
plot(fn)



# B) Mean, Variance, Median and IQR of data  
mean(data1)

## [1] 1.4876

var(data1)

## [1] 2.934267

median(data1)

## [1] 1.076

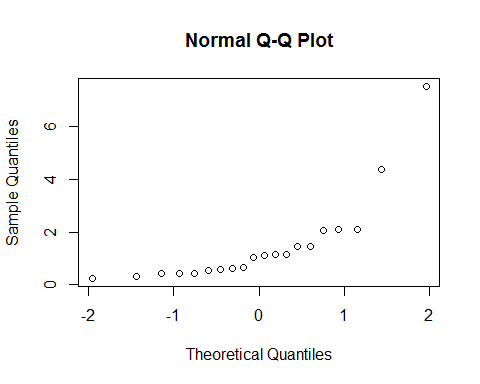
quantile(fn)

## 0% 25% 50% 75% 100%   
## 0.24600 0.50600 1.07600 1.61375 7.51700

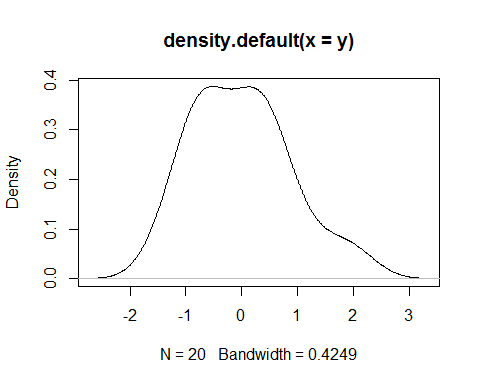
# IQR= 1.61375- 0.50600= 1.10775  
# C)   
IQR(data1)/sqrt(var(data1))

## [1] 0.6466837

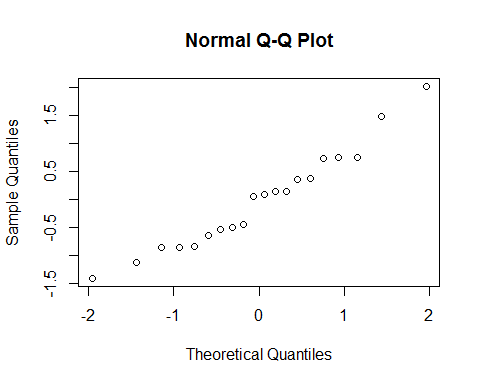
# 0.6466837 shows that data is not drawn from normal distribution as ratio does not conform to that of normal distribution it can also be seen from plotting kernel density function or box plot.  
# D)  
qqnorm(data1)



#The normal probability plot does not conform to a normal distribution.   
#Only looking at this piece of information, this data does not seem to   
#be drawn from a normal distribution.  
# E)  
y<-log(data1)  
plot(density.default(x=y))



qqnorm(y)



quantile(y)

## 0% 25% 50% 75% 100%   
## -1.4024237 -0.6848364 0.0730414 0.4669196 2.0171671

IQR(y)

## [1] 1.151756

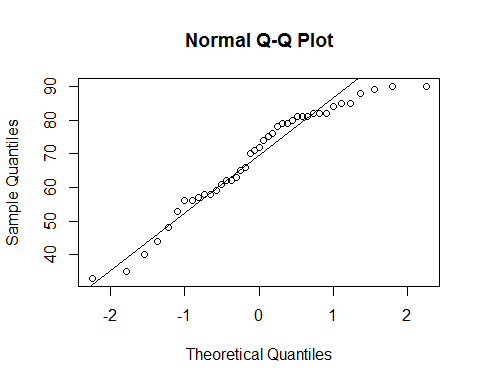
IQR(y)/sqrt(var(y))

## [1] 1.308589

# From looking at kernel density plot, normal plot and IQR to stdev ratio, We can say that y is not drawn from normal distribution.

# Solution 2

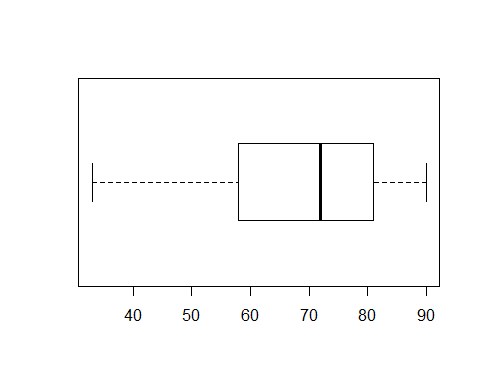
score<- c(90,81,72,58,90,81,71,58,89,81,70,57,88,80,66,56,85,79,65,56,85,79,  
 63,53,84,78,62,48,82,76,62,44,82,75,61,40,82,74,59,35,33)  
qqnorm(score)  
qqline(score)



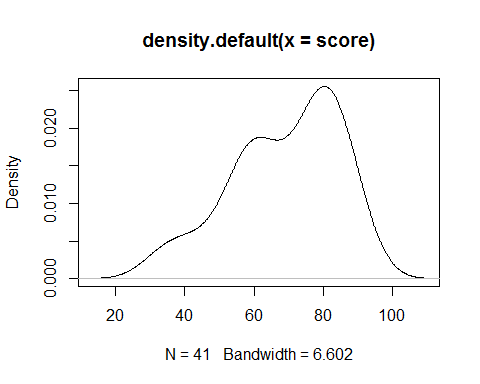
ratio<- IQR(score)/sqrt(var(score))  
ratio

## [1] 1.491889

# a) Since most of the data is concentrated around identity line and #ratio is around 1.491, we can say that data is normally distributed.  
  
boxplot(score,horizontal = TRUE)



plot(density(score))



median(score)

## [1] 72

mean(score)

## [1] 69.02439

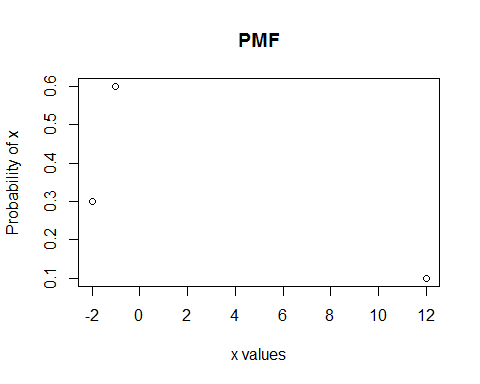
# b) After seeing the density plot we can say that this is bimodal with  
# two peaks, Most of the data is concentrated around mean and median #between 60 and 80

Solution 5

# a) EX of X  
library(Hmisc)

## Loading required package: grid  
## Loading required package: lattice  
## Loading required package: survival  
## Loading required package: Formula  
## Loading required package: ggplot2  
##   
## Attaching package: 'Hmisc'  
##   
## The following objects are masked from 'package:base':  
##   
## format.pval, round.POSIXt, trunc.POSIXt, units

x<-c(-2,-1,12)  
px<-c(0.3,0.6,0.1)  
plot(x,px ,xlab="x values",ylab="Probability of x",main="PMF")



EX<-wtd.mean(x,px)  
EX

## [1] 2.220446e-16

# b) Variance OF X  
Varx<-wtd.mean(x^2,px)-(EX^2)  
Varx

## [1] 16.2

# c) Expected value of ???= µ  
# E[???]= 0  
# d) Variance of ???  
Varx\_b<-Varx/3  
Varx\_b

## [1] 5.4

# e) Given that n=100  
y<-rnorm(100,0,sqrt(5.4))  
# P(???>0.5) is  
1-pnorm(0.5,mean(y),sd(y))

## [1] 0.4142607

Solution 6

# a) Sample Mean of our data  
library(Hmisc)  
hs<-c(1,2,3,4,5,6,7)  
hno<-c(27,54,16,13,6,3,1)  
data<-data.frame(hs,hno)  
sm<-weighted.mean(data$hs,data$hno)  
sm

## [1] 2.416667

# b)Standard Deviation of Household size  
varData<-weighted.mean(data$hs^2,data$hno)-sm^2  
sdData<-sqrt(varData)  
sdData

## [1] 1.294754

# c) Standard Error  
se<- sdData/sqrt(7)  
se

## [1] 0.4893712

# d) approximate probability that the absolute value of the error in a   
#survey of this form  
#and size is less than 0  
pnorm(0,0,0.49)

## [1] 0.5

# e) Yescoz the sample mean pf household size is coming out to be 2.41   
#with standard error around 0.48 which is mostly between 2 and 3.