

Problem Set 6

Keith Hickman

October 8, 2017

R Markdown

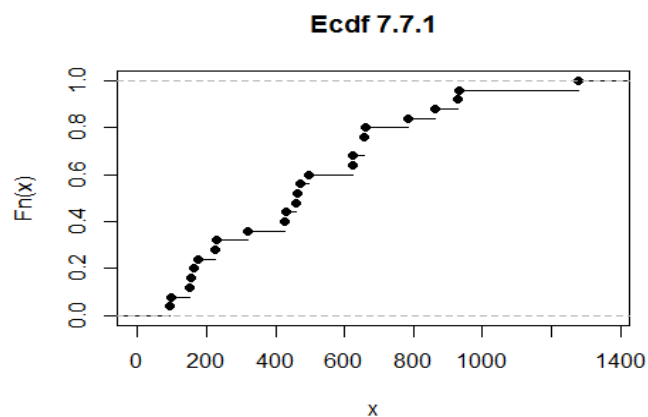
Problem 1

Trosset 7.7.1 a-e

a)

Graph the empirical cdf of x . *I copied the values from the text into a csv format*

```
ex7 <-  
read.table("c:/Users/khickman/Desktop/Personal/IUMSDS/StatsS520/Module6/ex7.csv")  
data <- ex7  
summary(data)  
  
##           V1  
##  Min.    : 92.0  
## 1st Qu.: 225.0  
##  Median : 462.0  
##   Mean  : 494.6  
## 3rd Qu.: 658.0  
##   Max.   :1277.0  
  
data <- data[,1]  
plot(ecdf(data), main = "Ecdf 7.7.1")
```



b)

Compute the plug-in estimates of the population mean and variance.

Mean:

```
lendata <- length(data)
meandata <- mean(data)
meandata
## [1] 494.6
```

Variance:

```
vardata <- (sum(data^2)/lendata) - meandata^2
vardata
## [1] 91078.72
```

(c)

Compute the plug-in estimates of the population median and interquartile range.

Median:

```
median(data)
## [1] 462
```

IQR:

```
quant75 <- quantile(data,.75)
quant25 <- quantile(data,.25)

IQR <- quant75 - quant25
IQR

## 75%
## 433

## or
iqr <- function(x)
  {q <- as.vector(quantile(x,probs=c(.25,.75)))
  return(q[2]-q[1])
  }
iqr(data)
## [1] 433
```

(d)

Compute the ratio of the plug-in estimate of the interquartile range to the square root of the plug-in estimate of the variance.

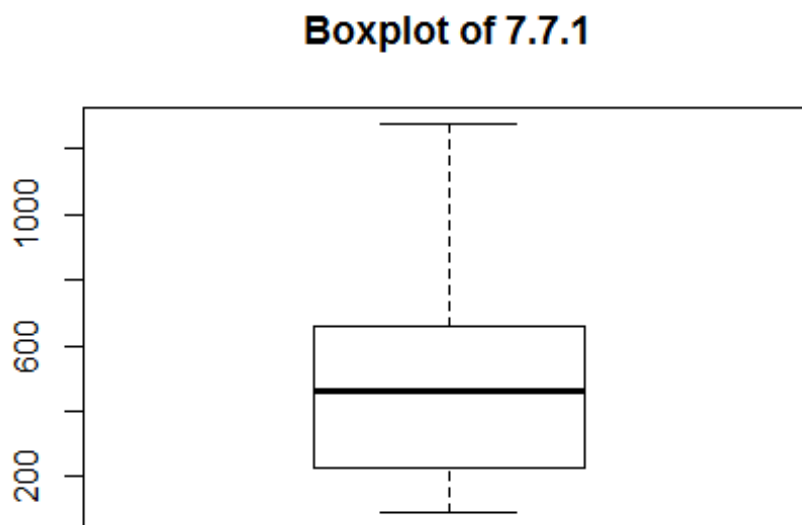
```
sqrt(vardata)
## [1] 301.7925
iqr(data) / (sqrt(vardata))
## [1] 1.434761
```

The ratio of IQR to standard deviation is 1.43

(e)

Construct a boxplot.

```
boxplot(data, main = "Boxplot of 7.7.1")
```



Problem 2

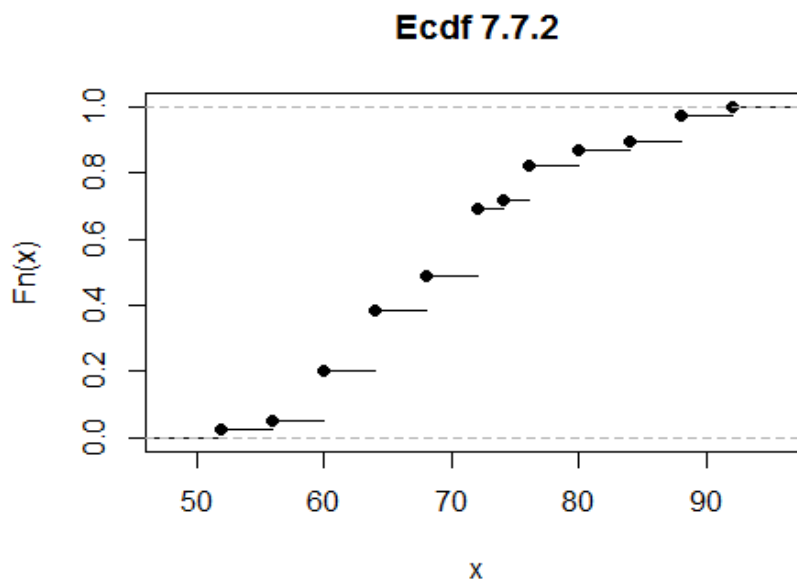
Trosset 7.7.2

```
pulse <-  
read.table("c:/Users/khickman/Desktop/Personal/IUMSDS/StatsS520/Module6/pulse  
rates.csv")  
summary(pulse)  
  
##           V1  
##  Min.    :52.00  
## 1st Qu.:64.00  
##  Median :72.00  
##   Mean  :70.31  
## 3rd Qu.:76.00  
##   Max.   :92.00  
  
pulse <- pulse[,1]
```

a)

Empirical cdf of X:

```
plot(ecdf(pulse), main = "Ecdf 7.7.2")
```



(b)

Compute the plug-in estimates of the population mean and variance. **Mean:**

```
mpulse <- mean(pulse)
mpulse
## [1] 70.30769
```

Variance

```
varpulse <- mean(pulse^2) - mpulse^2
varpulse
## [1] 87.90533
```

(c)

Compute the plug-in estimates of the population median and interquartile range. **Median**

```
median(pulse)
## [1] 72
```

IQR

```
iqrpulse <- iqr(pulse)
iqrpulse
## [1] 12
```

(d)

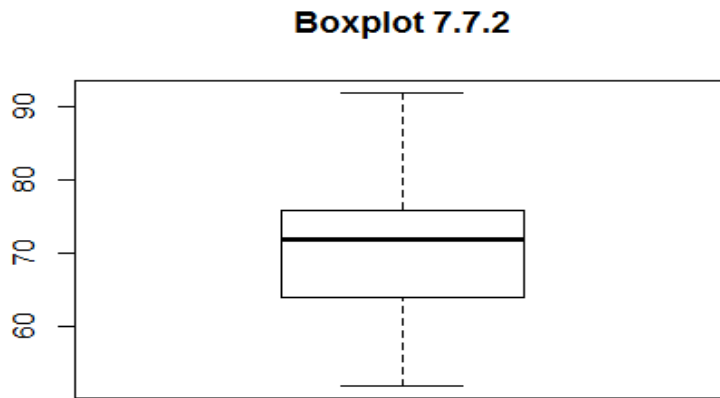
Compute the ratio of the plug-in estimate of the interquartile range to the square root of the plug-in estimate of the variance.

```
iqrpulse / sqrt(varpulse)
## [1] 1.279893
```

(e)

Construct a boxplot.

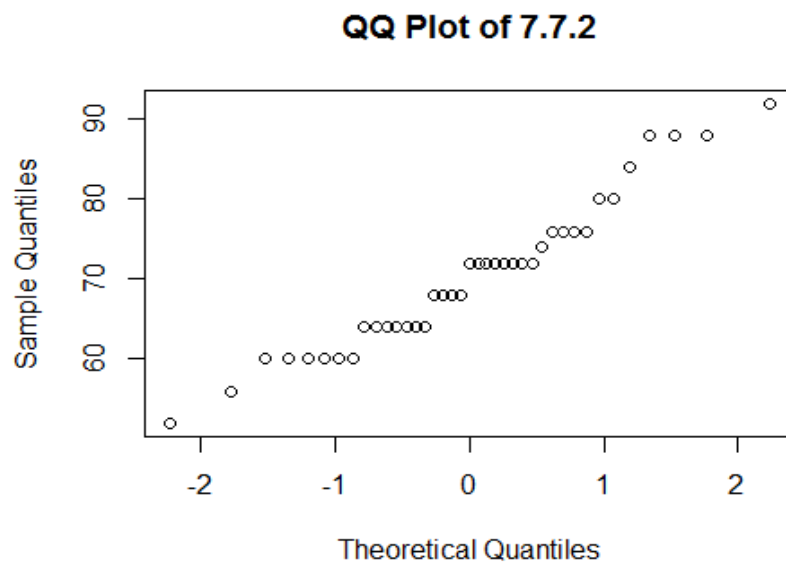
```
boxplot(pulse, main = "Boxplot 7.7.2")
```



(f)

Construct a normal probability plot.

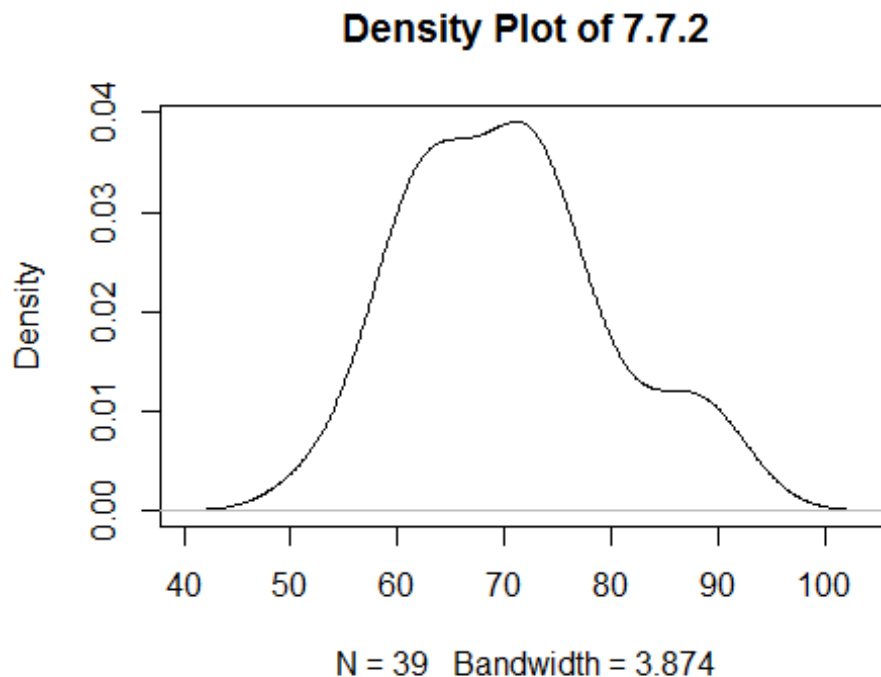
```
qqnorm(pulse, main = "QQ Plot of 7.7.2")
```



(g)

Construct a kernel density estimate.

```
plot(density(pulse), main = "Density Plot of 7.7.2")
```



(h)

Do you think that this sample was drawn from a normal distribution? Why or why not? Yes, this sample was drawn from a normal distribution, as evidenced by the various plots and metrics. First, the box plot shows no outlying values and normally distributed .25q, .75q, and median values. Second, the Normal QQ plot is linear and does not diverge significantly at either extreme. The density plot is normal as well. Additionally, mean and median are very close together, and the IQR/standard deviation ratio is close to 1 indicating a centrally and normally dispersed distribution.

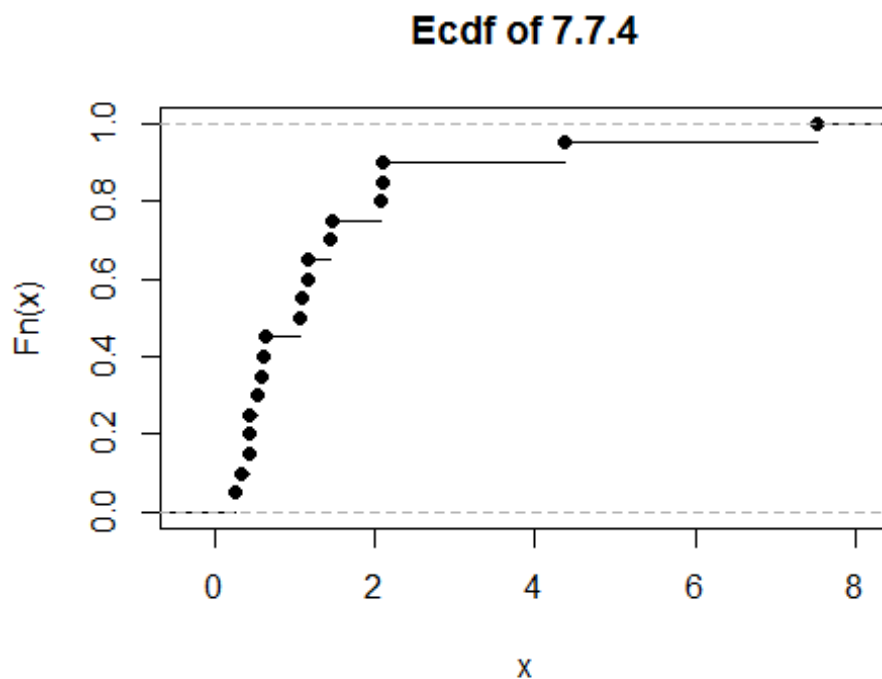
Problem 3

Trosset 7.7.4 #(a) Graph the empirical cdf of x.

```
ex74 <-  
read.table("c:/Users/khickman/Desktop/Personal/IUMSDS/StatsS520/Module6/sample7_4.csv")  
summary(ex74)
```

```
##          V1
##  Min.    :0.246
## 1st Qu.:0.506
##  Median :1.076
##   Mean   :1.488
## 3rd Qu.:1.614
##   Max.   :7.517

ex74 <- ex74[,1]
plot(ecdf(ex74), main = "Ecdf of 7.7.4")
```



(b) Calculate the plug-in estimates of the mean, the variance, the median, and the interquartile range.

Mean

```
mean74 <- mean(ex74)
mean74

## [1] 1.4876
```

Median:

```
median(ex74)

## [1] 1.076
```


Variance

```
var74 <- var(ex74)
var74

## [1] 2.934267
```

IQR

```
iqr74 <- iqr(ex74)
iqr74

## [1] 1.10775
```

(c)

Take the square root of the plug-in estimate of the variance and compare it to the plug-in estimate of the interquartile range. Do you think that x was drawn from a normal distribution? Why or why not?

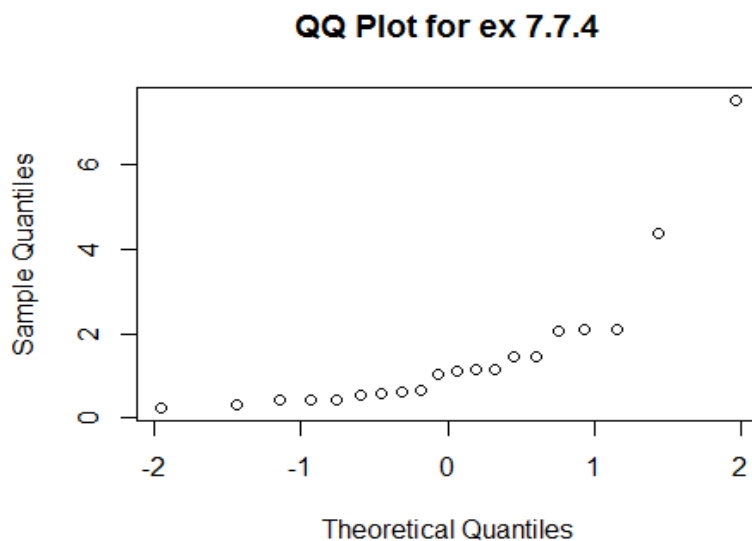
```
iqr74/(sqrt(var74))

## [1] 0.6466837
```

(d)

Use the `qqnorm` function to create a normal probability plot. Do you think that x was drawn from a normal distribution? Why or why not?

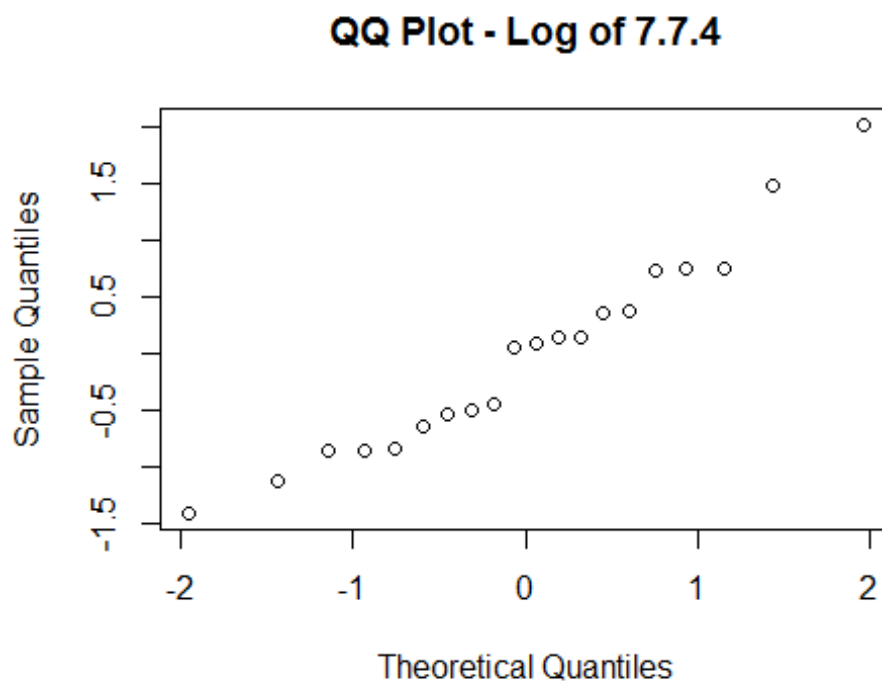
```
qqnorm(ex74, main = "QQ Plot for ex 7.7.4")
```



(e)

Do you think this sample was drawn from a normal distribution? Why or why not? This sample was not drawn from normally distributed data, and is right-skewed. The strongest evidence of this is possibly examining the QQ plot, which shows the extreme values moving well away from the central values at either end of the distribution graph. Additionally, the boxplot (below) shows two outliers well above 2 standard deviations above the .75 quantile. Finally, the IQR/standard deviation ratio is below 1, which is a signal that the data is more dispersed and could potentially be non-normal.

```
log74 <- log(ex74)
qqnorm(log74, main = "QQ Plot - Log of 7.7.4")
```



```
boxplot(ex74, main = "Boxplot 7.7.4")
```

