Stat 311: HW 5, Chapters 6 & 7, Solutions

Fritz Scholz

Ch. 6, Problem 1. The area under the density to the left of $x \in [1,2]$ is

$$\text{base} \times \text{height} \times (1/2) = (x-1) \times 2(x-1) \times (1/2) = (x-1)^2$$

Thus we need to solve $(x-1)^2 = 0.25$, $(x-1)^2 = 0.5$, $(x-1)^2 = 0.75$ and get

$$\begin{aligned} q(X;0.25) &= q_1(X) = 1 + \texttt{sqrt}(.25) = 1.5, \\ \text{population median} &= q_2(X) = q(X;0.25) = 1 + \texttt{sqrt}(.5) = 1.707107, \\ \text{and} \quad q(X;0.75) &= q_3(X) = 1 + \texttt{sqrt}(.75) = 1.866025 \\ \text{with} \quad \text{iqr} &= q(X;0.75) - q(X;0.25) = q_3(X) - q_1(X) = 1.866025 - 1.5 = 0.366025 \end{aligned}$$

Ch. 6, Problem 7. (a) The answer is TRUE, provided one treats quantiles symmetrically around the center of symmetry θ (= median) when they are ambiguous. When the p-quantiles are unique, then $q(X;p)$ and $q(X,1-p)$ are positioned symmetrically around θ, i.e., $q(X;p) - \theta = -(q(X,1-p) - \theta)$. To see this we use that $X - \theta$ and $-(X - \theta)$ have the same distribution and argue

$$\begin{array}{rcl} P(X < q(X;p)) \le & p & \le P(X \le q(X;p)) \\ P(X - \theta < q(X;p) - \theta) \le & p & \le P(X - \theta \le q(X;p) - \theta) \\ P(-(X-\theta) < q(X;p) - \theta) \le & p & \le P(-(X-\theta) \le q(X;p) - \theta) \\ P(X - \theta > -(q(X;p) - \theta)) \le & p & \le P(X - \theta \ge -(q(X;p) - \theta)) \\ 1 - P(X - \theta \le -(q(X;p) - \theta)) \le & p & \le 1 - P(X - \theta < -(q(X;p) - \theta)) \\ P(X - \theta < -(q(X;p) - \theta)) \le & 1-p & \le P(X - \theta \le -(q(X;p) - \theta)) \\ P(X - \theta < q(X;1-p) - \theta) \le & 1-p & \le P(X - \theta \le q(X;1-p) - \theta) \end{array}$$

The uniqueness of quantiles $\Longrightarrow \quad q(X;1-p) - \theta = -(q(X,p) - \theta)$ and thus $(q(X;p) + q(X,1-p))/2 = \theta$. The median then is $q(X;0.5) = \theta$ and the first and third quartile ($p = 0.25$ and $1 - p = 0.75$) average to θ. When the quantiles are ambiguous then the same chain of inequalities above shows that for any $q(X;p)$ as p-quantile we can choose $q(X;1-p)$ as $(1-p)$-quantile when it satisfies $q(X;1-p) - \theta = -(q(X,p) - \theta)$, i.e., $q(X;1-p) = \theta - (q(X,p) - \theta)$, with the same conclusion.

(b) The upper and lower distribution tails (up to just less than 25% at either end) have no effect on the first and third quartile, i.e., no effect on the iqr. However, the distribution tails could be very long in those upper and lower 25% of the distribution. If we focus on symmetric distributions, then the center = mean = μ is not affected by the tails but the variance $E(X-\mu)^2$ could become arbitrarily large. Thus the anser is: FALSE.

(c) If we move slightly less than the upper 25% of the distribution to very high values, then the expected value or mean can be made arbitrarily large without affecting the quartiles. Thus the answer is: FALSE

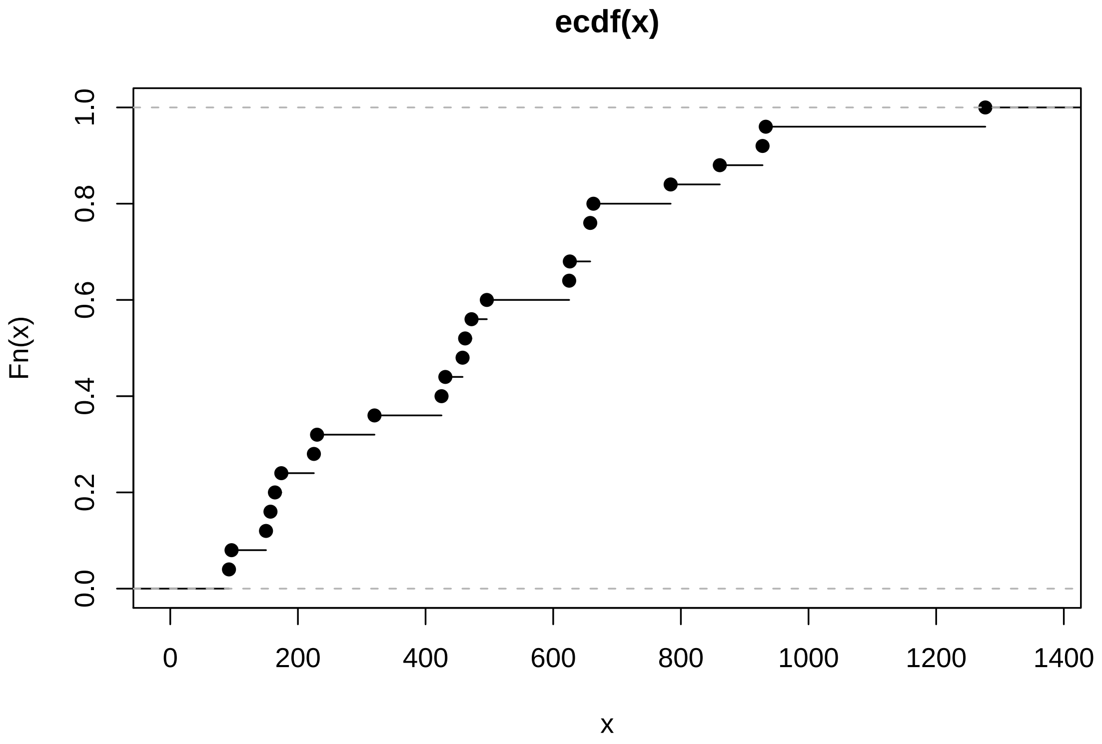
(d) The answer is TRUE, since a zero variance can only occur if the random variable is a constant c, in which case the iqr = 0.

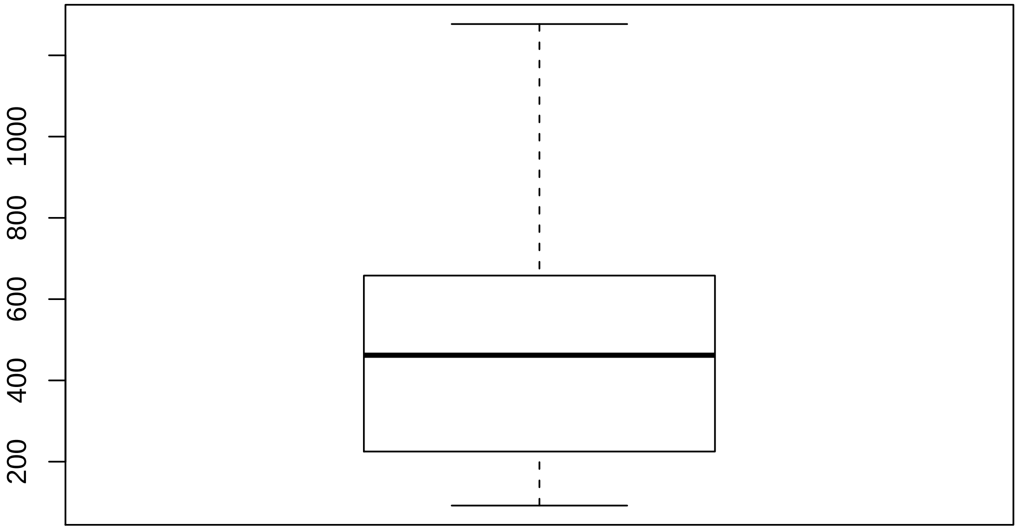
(e) We can keep the mean or expectation of a random variable constant at the median as long as we balance half the probability mass on either side of the median in such a way that the center of gravity is at the median, but such balancing does not require symmetry. The answer is: FALSE.

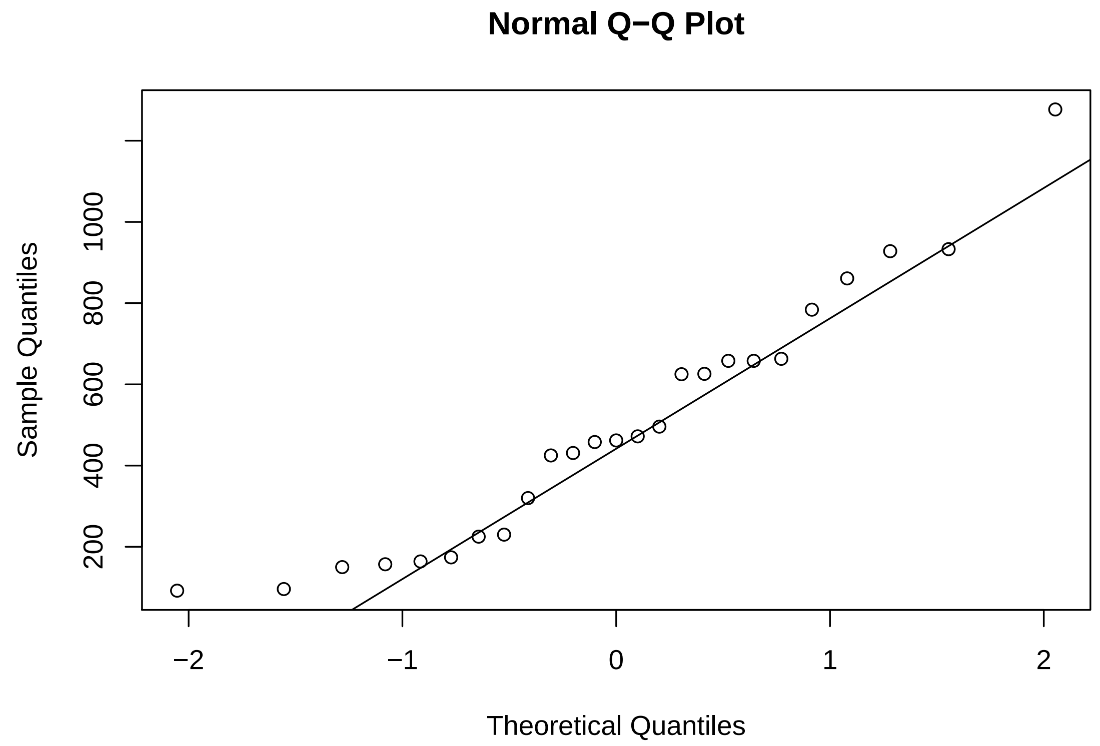
Ch. 7, Problem 1.

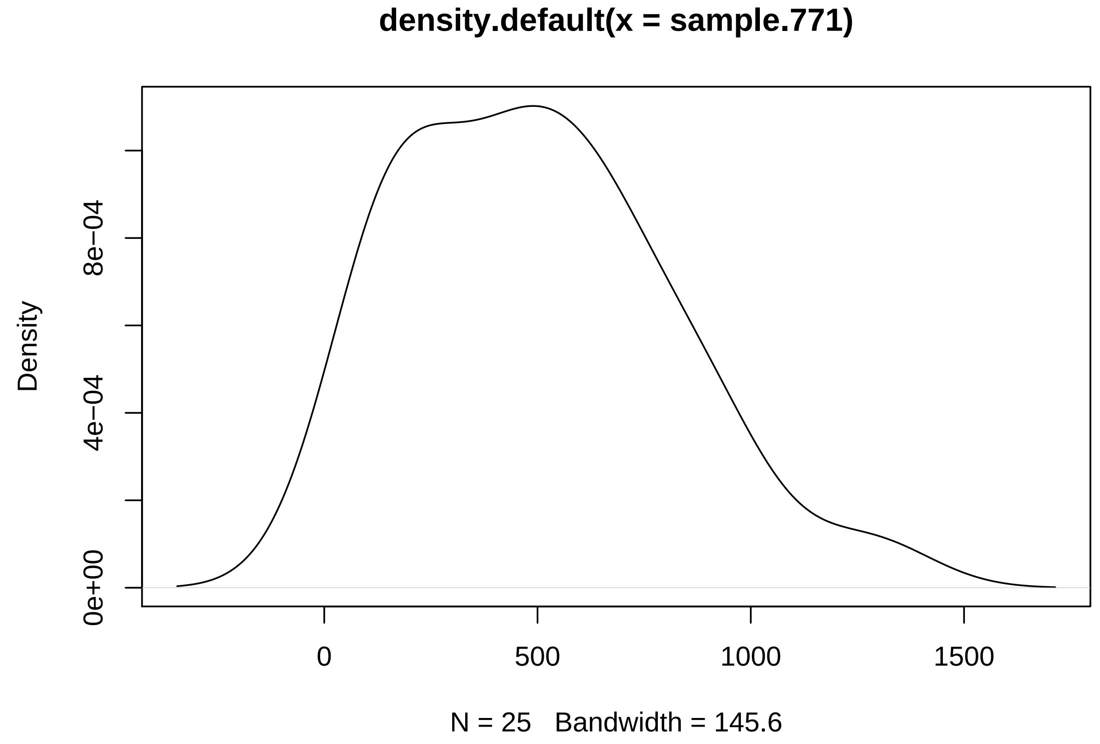
```
> sample.771 <- scan("sample771.dat") # reading the data from file sample771.dat
# (a) plotting the ecdf, see plots below
> plot.ecdf(sample.771)
# (b) getting mean and variance plug-in estimates
> mean(sample.771) #
[1] 494.6
> length(sample.771)
[1] 25
> var(sample.771)*24/25
[1] 91078.72
# (c) plug-in estimates of median and iqr
> out <- quantile(sample.771,prob=c(.25,.5,.75))
> out
25% 50% 75%
225 462 658
> as.vector(out[3]-out[1])
[1] 433
# (d) ratio of plug-in iqr to sqrt(plug-in variance) estimates
> 433/sqrt(91078.72)
[1] 1.434761
# (e) boxplot, see plots below
> boxplot(sample.771)
# (f) normal probability plot, see plots below
> qqnorm(sample.771)
> qqline(sample.771)
# (g) kernel density estimate, see plots below
> plot(density(sample.771))
```

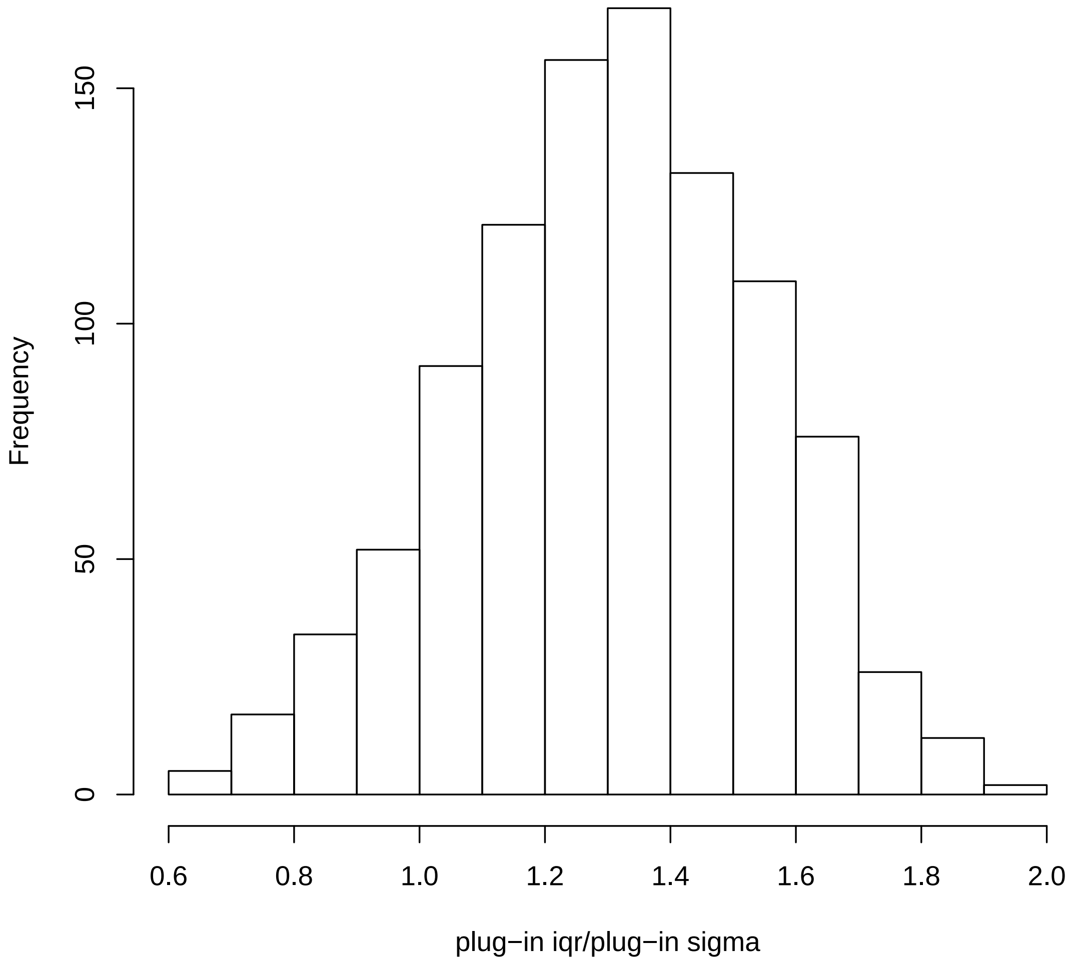
(h) For a normal population we have iqr/$\sigma = 1.348980$.
Our estimated value 1.434761 could be well within sampling variation of that value, as the histogram below shows..
The probability plot looks fairly linear in the middle, with a bit of curving upward at the low end
(indicating shorter tail at the low end).
The boxplot looks fairly symmetric in the middle 50% but has a bit of skew toward the higher values.
That is also indicated by the density plot.
Since the sample is relatively small ($n = 25$) we cannot make a strong case against normality.


ecdf(x)
Fn(x)
0.0
0.2
0.4
0.6
0.8
1.0
0
200
400
600
800
1000
1200
1400
x


200
400
600
800
1000


Normal Q−Q Plot
Sample Quantiles
200
400
600
800
1000
−2
−1
0
1
2
Theoretical Quantiles

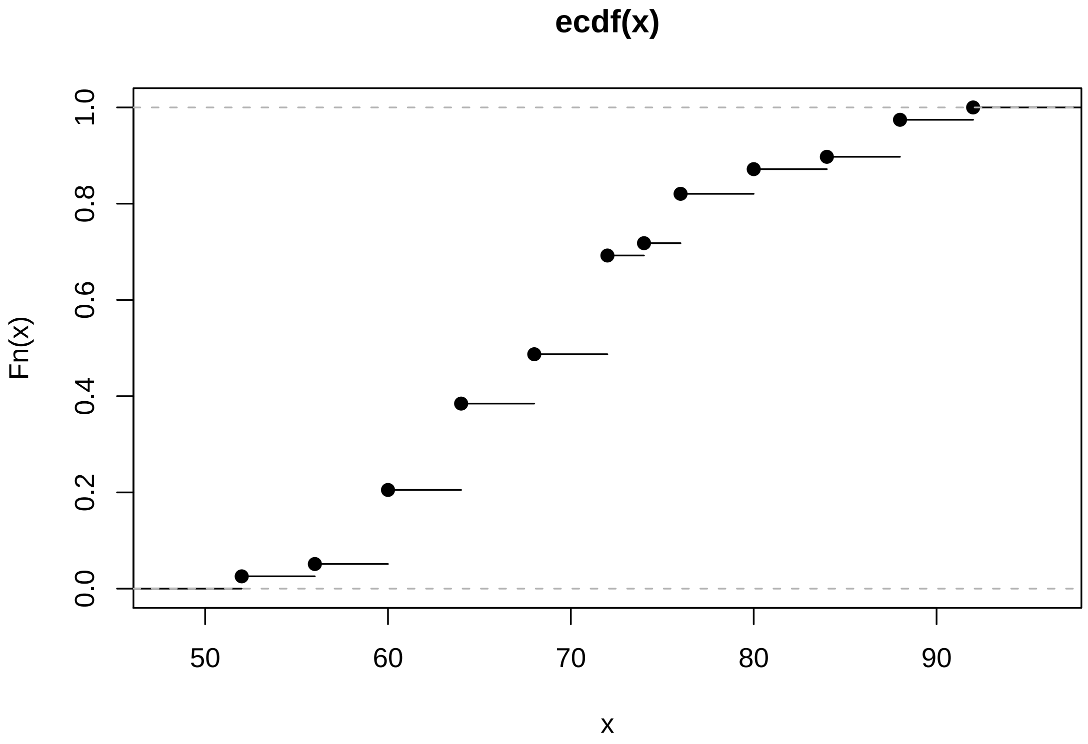

density.default(x = sample.771)
Density
0e+00
4e−04
8e−04
0
500
1000
1500
N = 25   Bandwidth = 145.6


Frequency
0
50
100
150
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
plug−in iqr/plug−in sigma

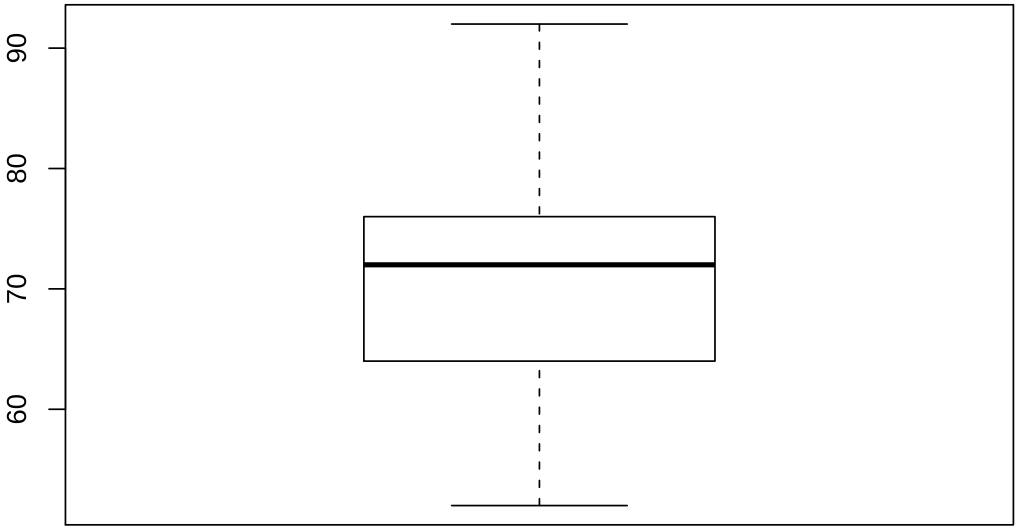
Ch. 7, Problem 2.

```
> pulses <- scan("pulses.dat") # reading the data from file pulses.dat
# (a) plotting the ecdf, see plot below
> plot.ecdf(pulses)
# (b) getting mean and variance plug-in estimates
> mean(pulses) #
[1] 70.30769
> length(pulses)
[1] 39
> var(pulses)*38/39
[1] 87.90533
# (c) plug-in estimates of median and iqr
> out <- quantile(pulses,prob=c(.25,.5,.75))
> out
25% 50% 75%
 64  72  76
> as.vector(out[3]-out[1])
[1] 12
# (d) ratio of plug-in iqr to sqrt(plug-in variance) estimates
> 12/sqrt(87.90533)
[1] 1.279893
# (e) boxplot
> boxplot(pulses)
# (f) normal probability plot
> qqnorm(pulses)
> qqline(pulses)
# (g) kernel density estimate
> plot(density(pulses))
```

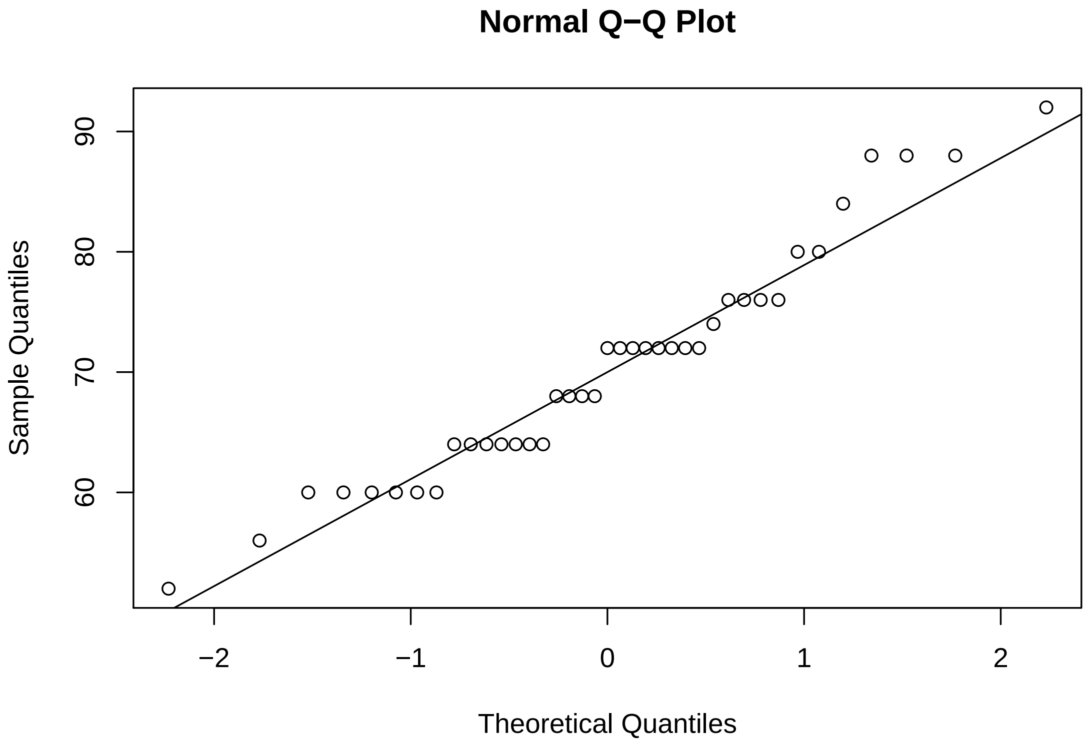
(h) For a normal population we have iqr/$\sigma = 1.348980$.
Our estimated value 1.279893 could be well within sampling variation of that value, as the histogram below shows.
The probability plot looks fairly linear.
The boxplot looks mildly asymmetric in the middle 50% (more spread out in the lower half of the box than in the upper part). The whiskers look almost symmetric.
The density plot looks roughly bell shaped, with a quirky shoulder on the right, could be a sampling idiosynchracy.
Since the sample is relatively small ($n = 39$) we cannot make a strong case against normality.

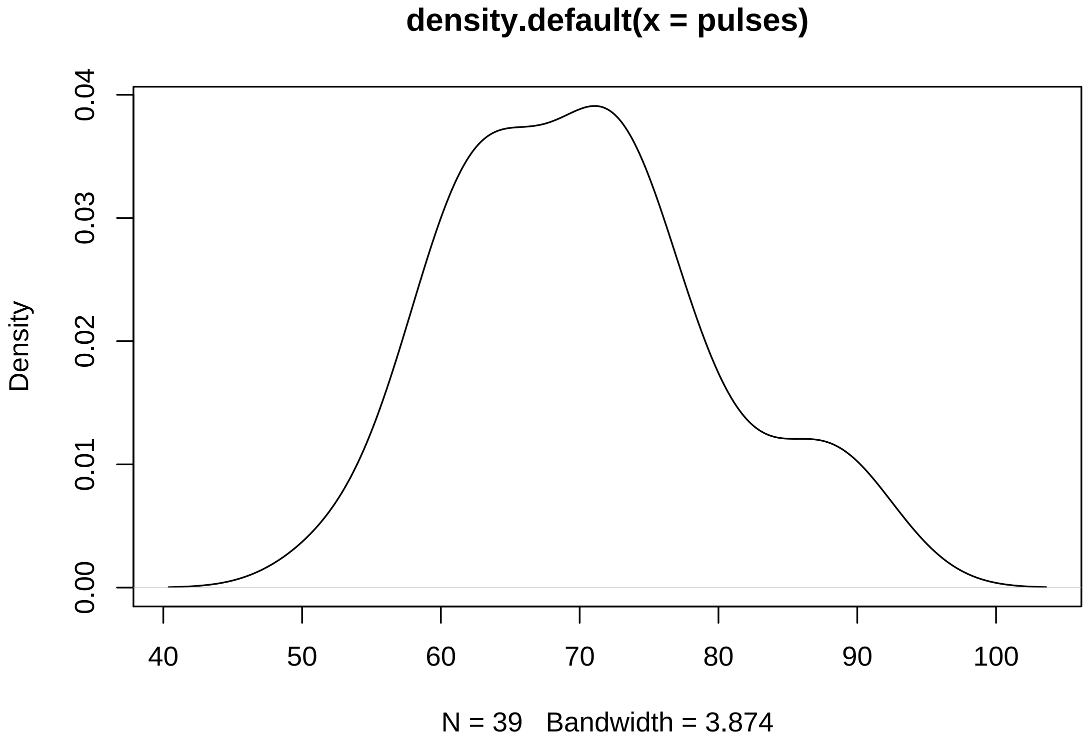


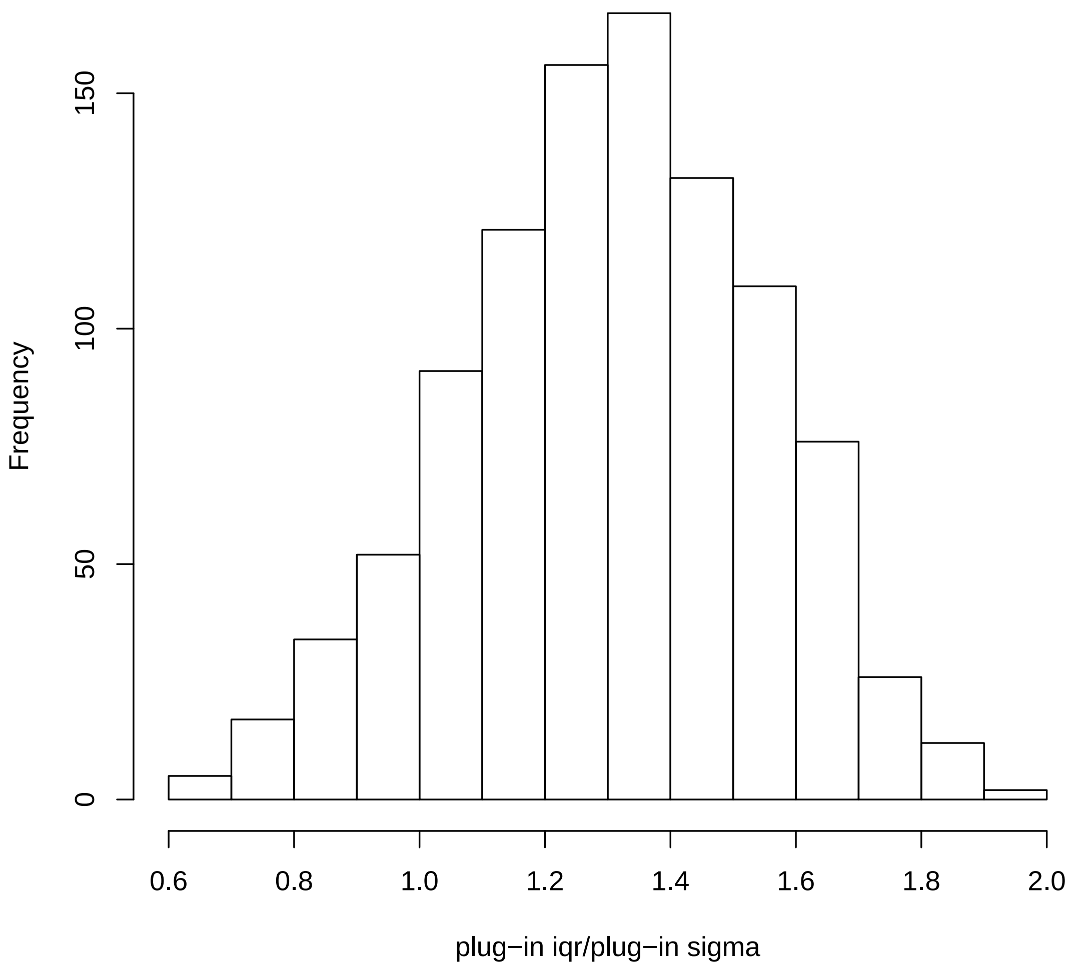
ecdf(x)
Fn(x)
0.0
0.2
0.4
0.6
0.8
1.0
50
60
70
80
90
x



60
70
80
90


Normal Q−Q Plot
Sample Quantiles
60
70
80
90
Theoretical Quantiles
−2
−1
0
1
2

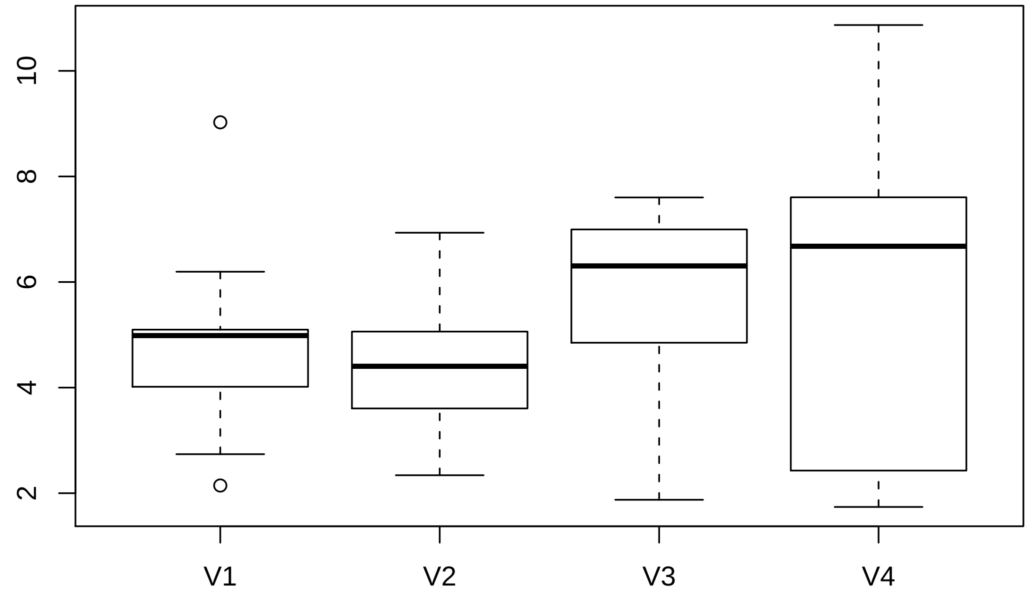

density.default(x = pulses)
Density
0.00
0.01
0.02
0.03
0.04
40
50
60
70
80
90
100
N = 39   Bandwidth = 3.874


Frequency
0
50
100
150
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
plug−in iqr/plug−in sigma

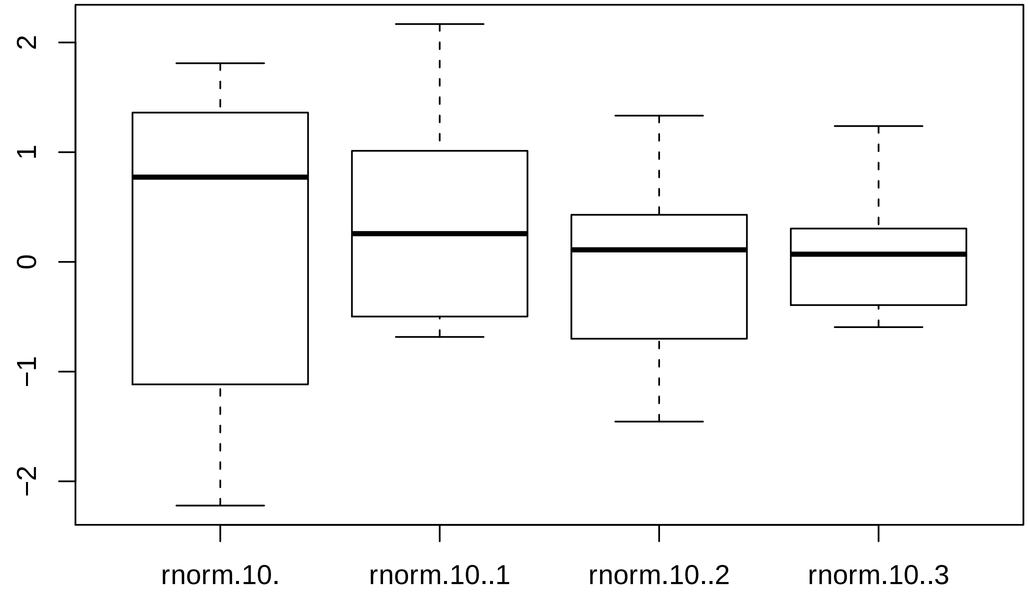
Ch. 7, Problem 3.

```
> sample773 <- read.table("sample773.dat")
# running the following function several times with
# plotsample773() at the command line
# should give us an idea of what to expect for nromal samples.
# Three such sets of simulated normal sample boxplots are shown below.
plotsample773 <- function(){
boxplot(sample773)
readline("hit return\n")
simsample <- data.frame(rnorm(10),rnorm(10),rnorm(10),rnorm(10))
boxplot(simsample)
}
```

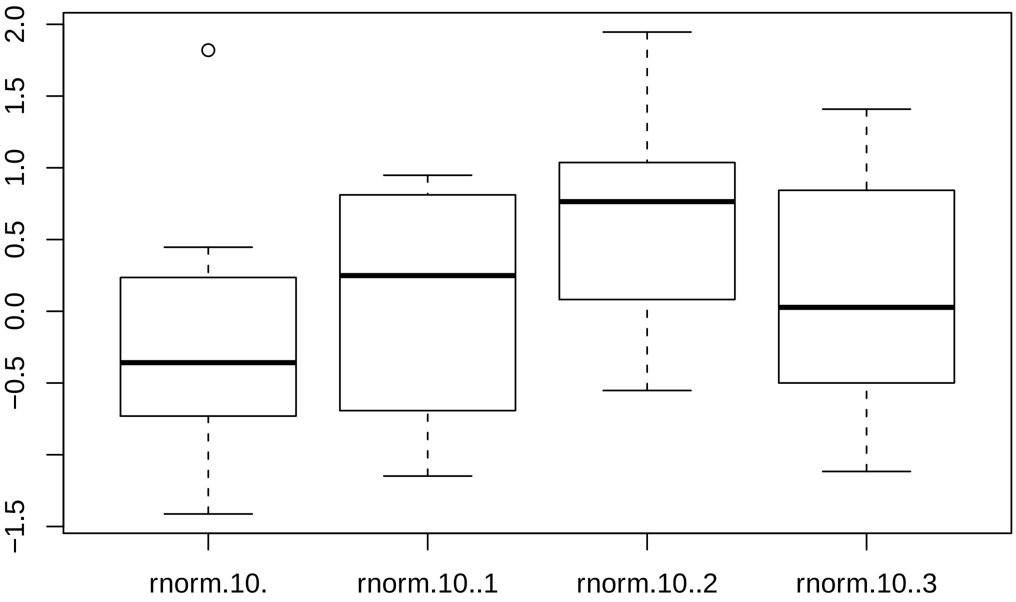
Based on the four plots shown below we should not rule out that these samples came from one common normal population. This is due to the samples sizes $n = 10$ being quite low, causing a lot of sampling variation.

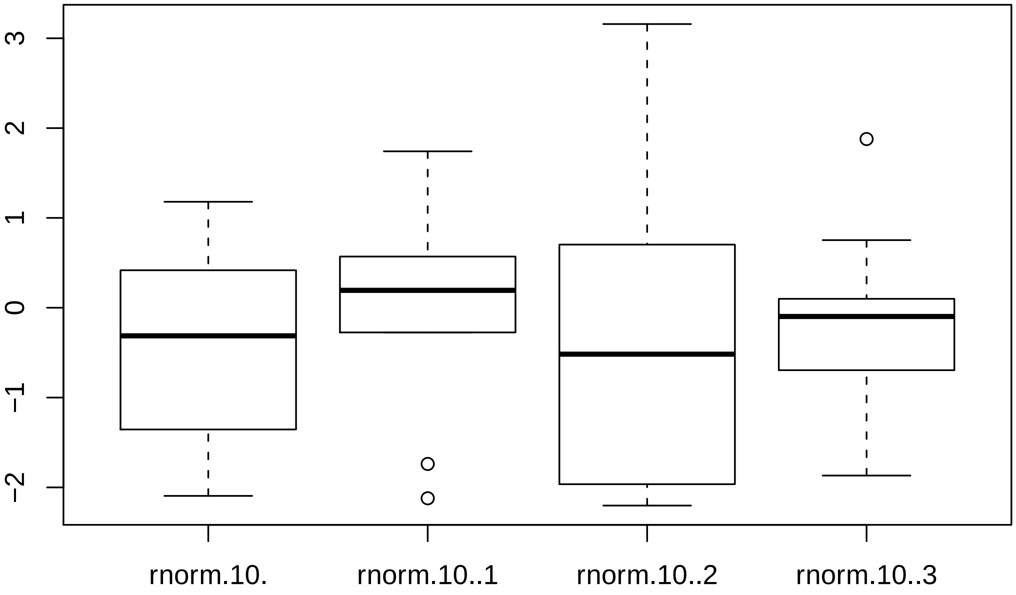


10
8
6
4
2
V1
V2
V3
V4



2
1
0
−1
−2
rnorm.10.
rnorm.10..1
rnorm.10..2
rnorm.10..3


2.0
1.5
1.0
0.5
0.0
−0.5
−1.5
rnorm.10.
rnorm.10..1
rnorm.10..2
rnorm.10..3


3
2
1
0
−1
−2
rnorm.10.
rnorm.10..1
rnorm.10..2
rnorm.10..3