

# Answers (Problem set 6)

## Online S520

1. *Trosset exercise 7.7.1 parts (a)–(e)*. Here's some R code for the question.

```
x = scan("http://mypage.iu.edu/~mtrosset/StatInfer/Data/sample771.dat")
plot.ecdf(x)
mean(x)
# Plug-in variance
mean(x^2) - mean(x)^2
median(x)
IQR = quantile(x, 0.75) - quantile(x, 0.25)
IQR / sqrt(mean(x^2) - mean(x)^2)
boxplot(x, main="Boxplot for Trosset exercise 7.7.1", ylab="x")
```

- (a) See Figure 1.

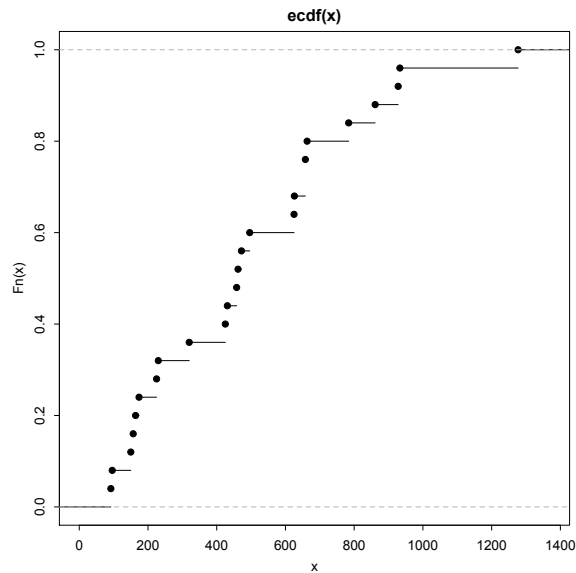


Figure 1: ECDF plot for Trosset exercise 7.7.1.

- (b) Mean is 494.6, plug-in variance is 91079, or 94974 if you use `var(x)` (which technically is not a plug-in estimate.)
- (c) Median is 462, IQR is  $658 - 225 = 433$
- (d) The ratio is 1.43 (or 1.41 if you use the `sd` function.)
- (e) See Figure 2.



Figure 2: Boxplot for Trosset exercise 7.7.1.

2. *Trosset exercise 7.7.2* Here's R code to produce the required graphs and and estimates.

```
pulses = scan("http://mypage.iu.edu/~mtrosset/StatInfer/Data/pulses.dat")
plot(ecdf(pulses), main="ECDF plot")
summary(pulses)
mean(pulses^2) - mean(pulses)^2 # Plug-in variance
IQR = quantile(pulses, 0.75) - quantile(pulses, 0.25)
IQR / sqrt(mean(pulses^2) - mean(pulses)^2) # IQR/SD
boxplot(pulses, main="Peruvian pulse rate boxplot")
qqnorm(pulses)
plot(density(pulses), main="Density plot")
```

- (a) ECDF. See Figure.
- (b) The plug-in mean is 70.3 and the plug-in variance 87.9 (the sample variance, which is 90.2, is also acceptable).
- (c) The median is 72 and the IQR is 12.
- (d) The ratio of the IQR to the plug-in SD is 1.28 (1.26 if you use the sample SD), about the same as the normal.
- (e) Boxplot. See Figure.
- (f) QQ plot (normal probability plot). See Figure.
- (g) Kernel density plot. See Figure.
- (h) The QQ plot is close to a straight line, so the distribution looks pretty close to normal (though there may or may not be a bump on the right-hand side of the distribution in the high 80s).

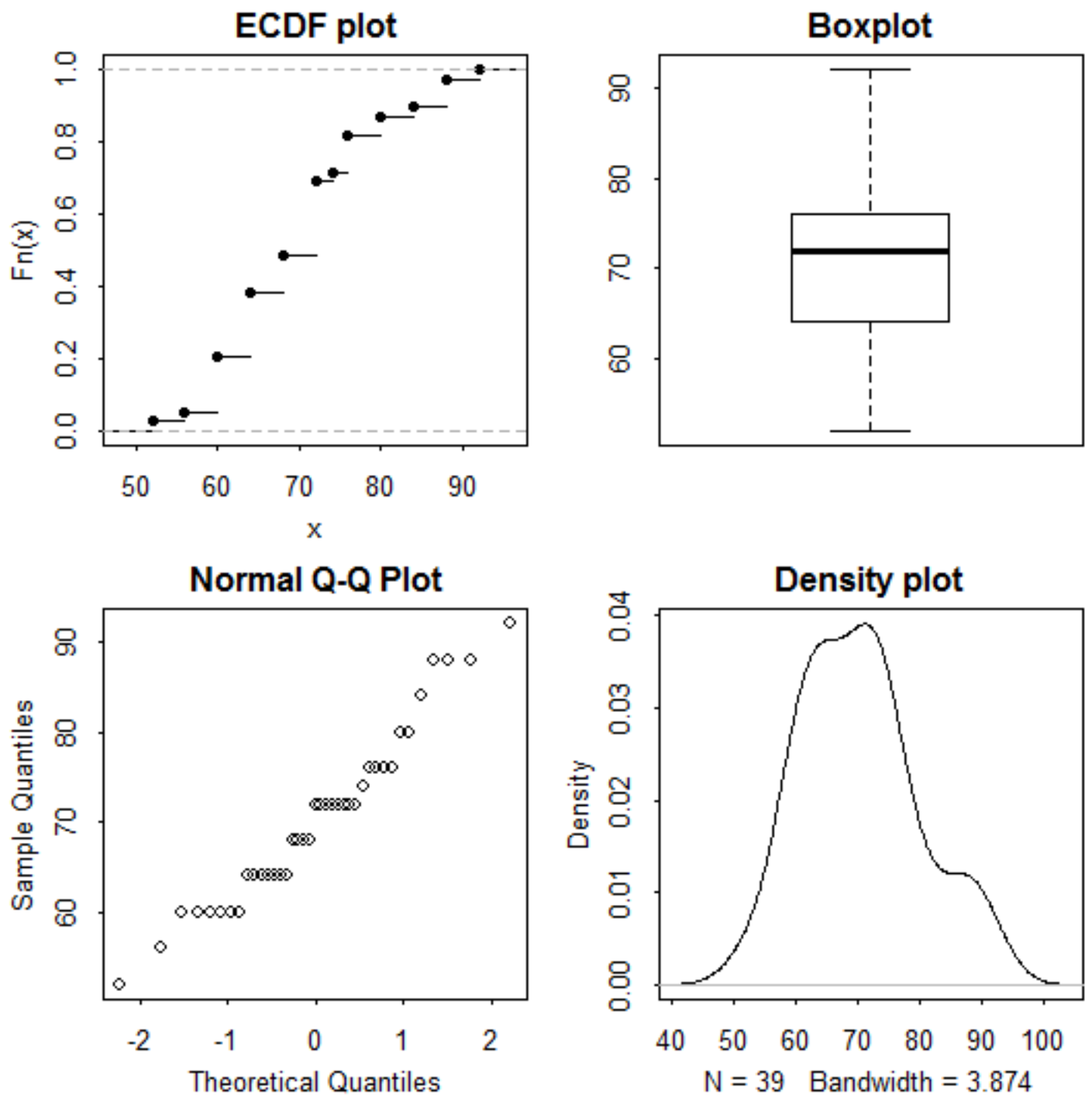


Figure 3: Plots of the distribution of Peruvian Indian pulse rates.

3. (Trosset exercise 7.7.4.)

```
data = scan("http://mypage.iu.edu/~mtrosset/StatInfer/Data/sample774.dat")
plot(ecdf(data), main="ECDF of data")
summary(data)
mean(data^2) - mean(data)^2 # Plug-in variance
sort(data)
(1.464+2.063)/2 - (.434+.530)/2 # Plug-in IQR
1.2815 / sqrt(mean(data^2) - mean(data)^2) # IQR/SD
qqnorm(data)
plot(density(log(data)), main="Density of log data")
qqnorm(log(data))
```

- (a) ECDF. See Figure.
- (b) Plug-in estimates are mean 1.49, variance 2.79 (2.93 for sample variance), median 1.076, IQR 1.28 (1.11 if you use R's default method for quartiles).
- (c) The ratio of IQR to SD is somewhere from 0.65 to 0.77 depending on which methods you use; in any, it's far from the 1.35 you expect from a normal distribution.
- (d) The QQ plot bends upwards, which confirms the distribution isn't normal.
- (e) The QQ plot of the logged data is very close to a straight line, so the logged data is approximately normal.

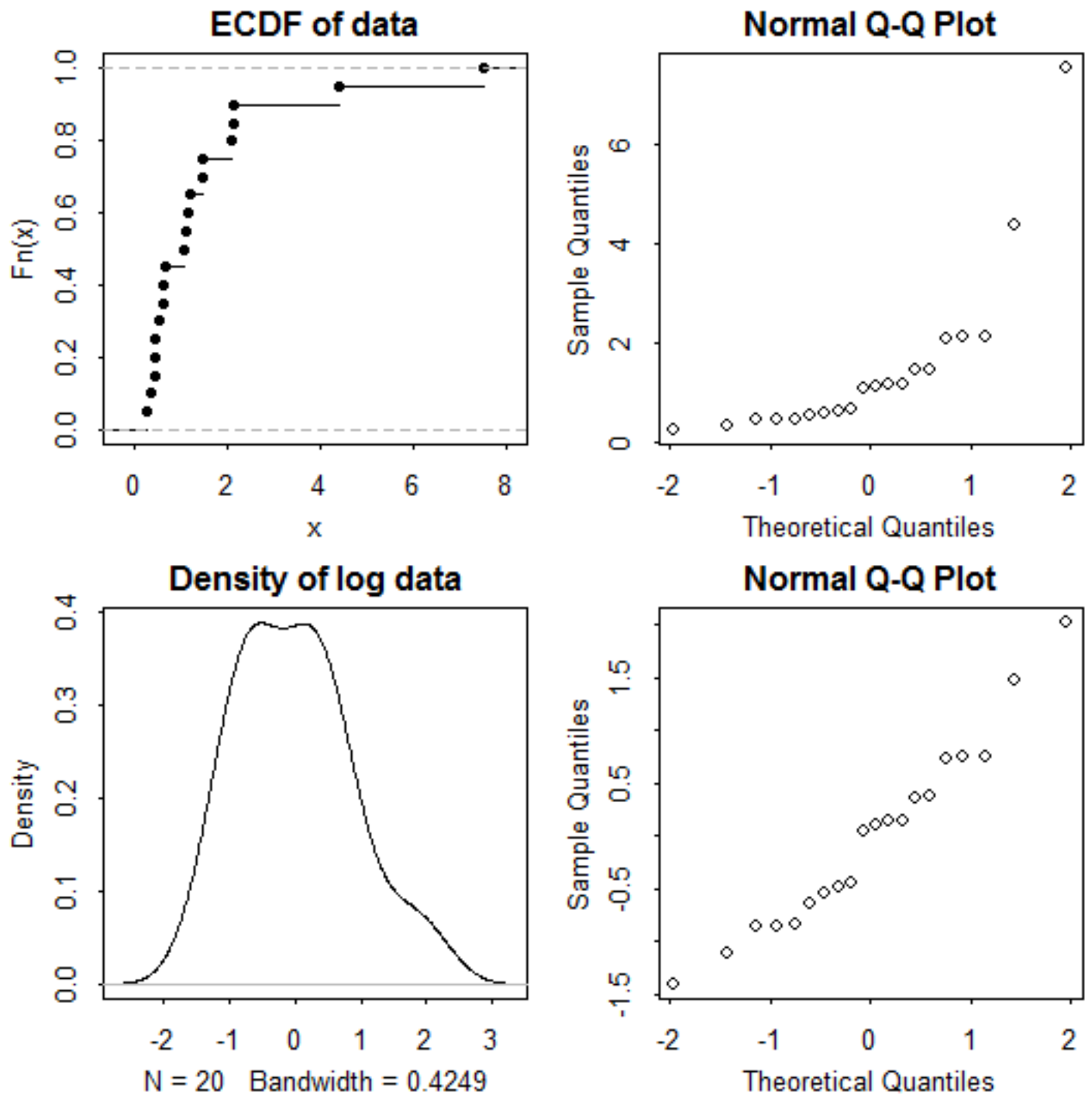


Figure 4: Plots of data for Trosset ex. 7.7.4. Top two are for raw data, bottom two are for logged data.