606 Chapter 3 Lab

Munkhnaran Gankhuyag

September 17, 2017

```
load("more/bdims.RData")
```

```
mdims <- subset(bdims, sex == 1)
fdims <- subset(bdims, sex == 0)

head(bdims)
```

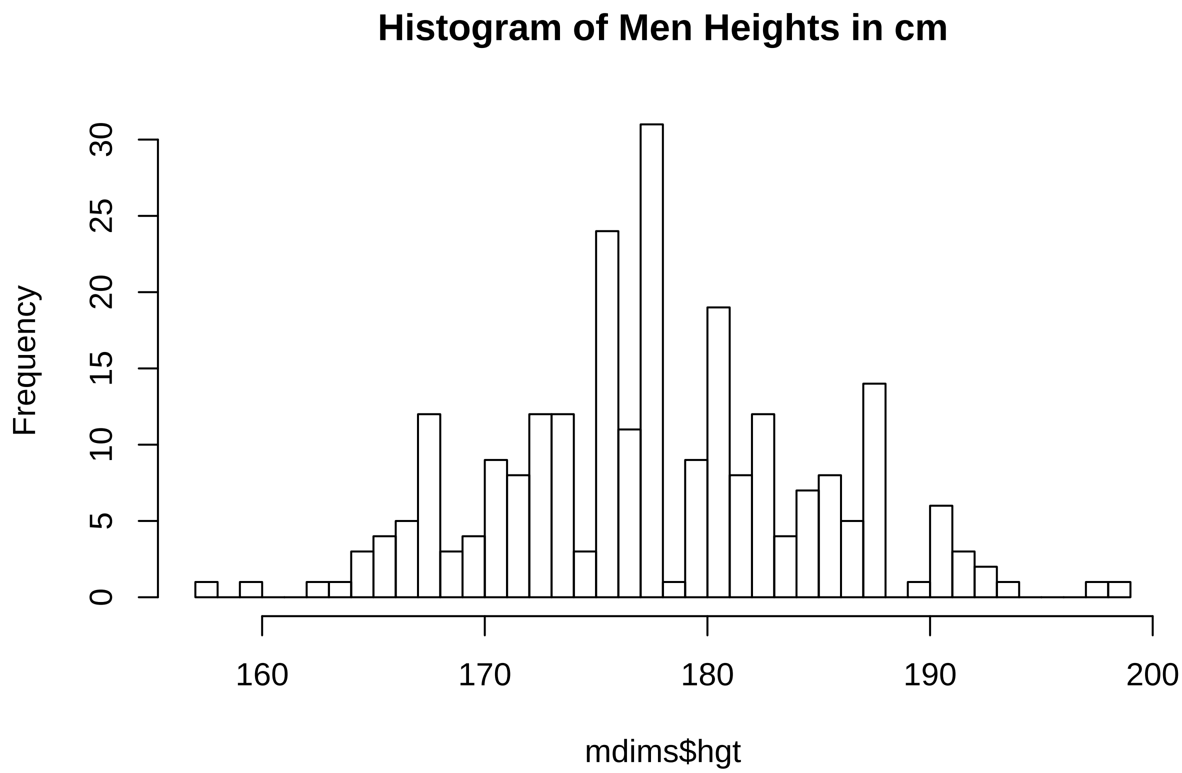
```
##   bia.di bii.di bit.di che.de che.di elb.di wri.di kne.di ank.di sho.gi
## 1   42.9   26.0   31.5   17.7   28.0   13.1   10.4   18.8   14.1  106.2
## 2   43.7   28.5   33.5   16.9   30.8   14.0   11.8   20.6   15.1  110.5
## 3   40.1   28.2   33.3   20.9   31.7   13.9   10.9   19.7   14.1  115.1
## 4   44.3   29.9   34.0   18.4   28.2   13.9   11.2   20.9   15.0  104.5
## 5   42.5   29.9   34.0   21.5   29.4   15.2   11.6   20.7   14.9  107.5
## 6   43.3   27.0   31.5   19.6   31.3   14.0   11.5   18.8   13.9  119.8
##   che.gi wai.gi nav.gi hip.gi thi.gi bic.gi for.gi kne.gi cal.gi ank.gi
## 1   89.5   71.5   74.5   93.5   51.5   32.5   26.0   34.5   36.5   23.5
## 2   97.0   79.0   86.5   94.8   51.5   34.4   28.0   36.5   37.5   24.5
## 3   97.5   83.2   82.9   95.0   57.3   33.4   28.8   37.0   37.3   21.9
## 4   97.0   77.8   78.8   94.0   53.0   31.0   26.2   37.0   34.8   23.0
## 5   97.5   80.0   82.5   98.5   55.4   32.0   28.4   37.7   38.6   24.4
## 6   99.9   82.5   80.1   95.3   57.5   33.0   28.0   36.6   36.1   23.5
##   wri.gi age  wgt   hgt sex
## 1   16.5  21 65.6 174.0   1
## 2   17.0  23 71.8 175.3   1
## 3   16.9  28 80.7 193.5   1
## 4   16.6  23 72.6 186.5   1
## 5   18.0  22 78.8 187.2   1
## 6   16.9  21 74.8 181.5   1
```

The normal distribution

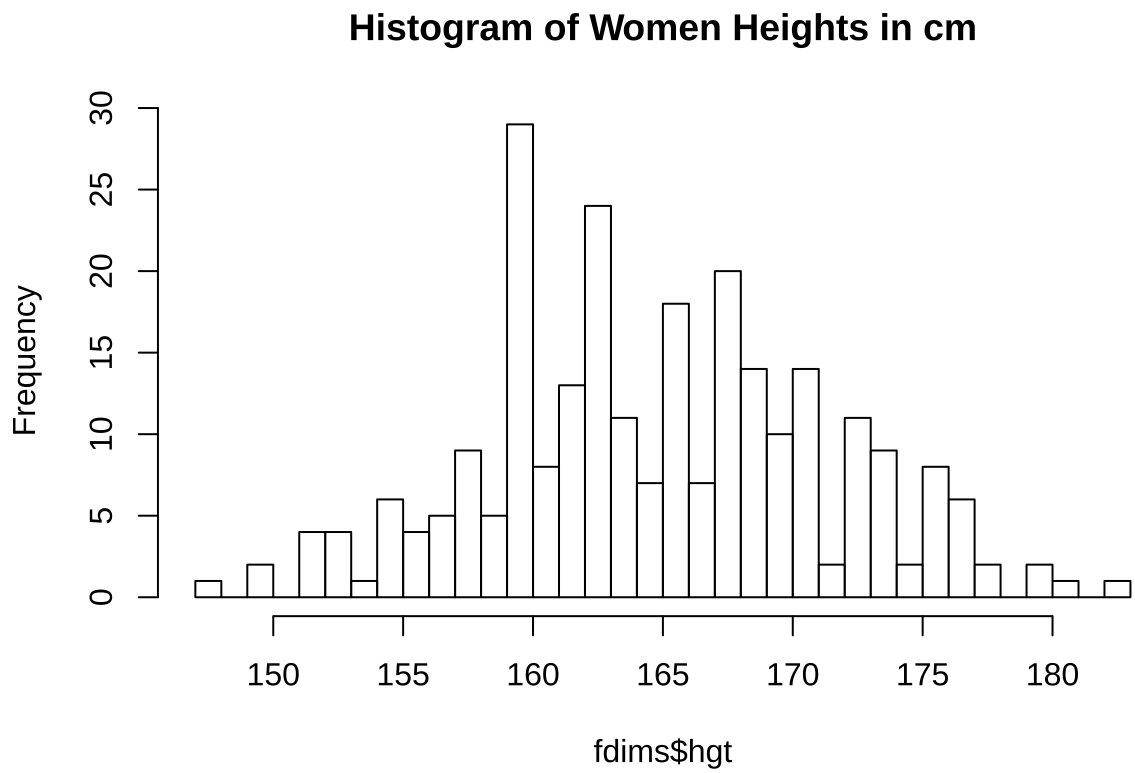
1. Make a histogram of men's heights and a histogram of women's heights. How would you compare the various aspects of the two distributions?

Based on the histograms, most frequent height for men and women is around 178 and 158.

```
hist(mdims$hgt, breaks = 50,  main = "Histogram of Men Heights in cm")
```




```r
hist(fdims$hgt, breaks = 50, main = "Histogram of Women Heights in cm")
```




```
fhgtmean <- mean(fdims$hgt)
fhgtsd   <- sd(fdims$hgt)
fwgtmean <- mean(fdims$wgt)
fwgtsd   <- sd(fdims$wgt)

fhgtmean
```

```
## [1] 164.8723
```

```
fhgtsd
```

```
## [1] 6.544602
```

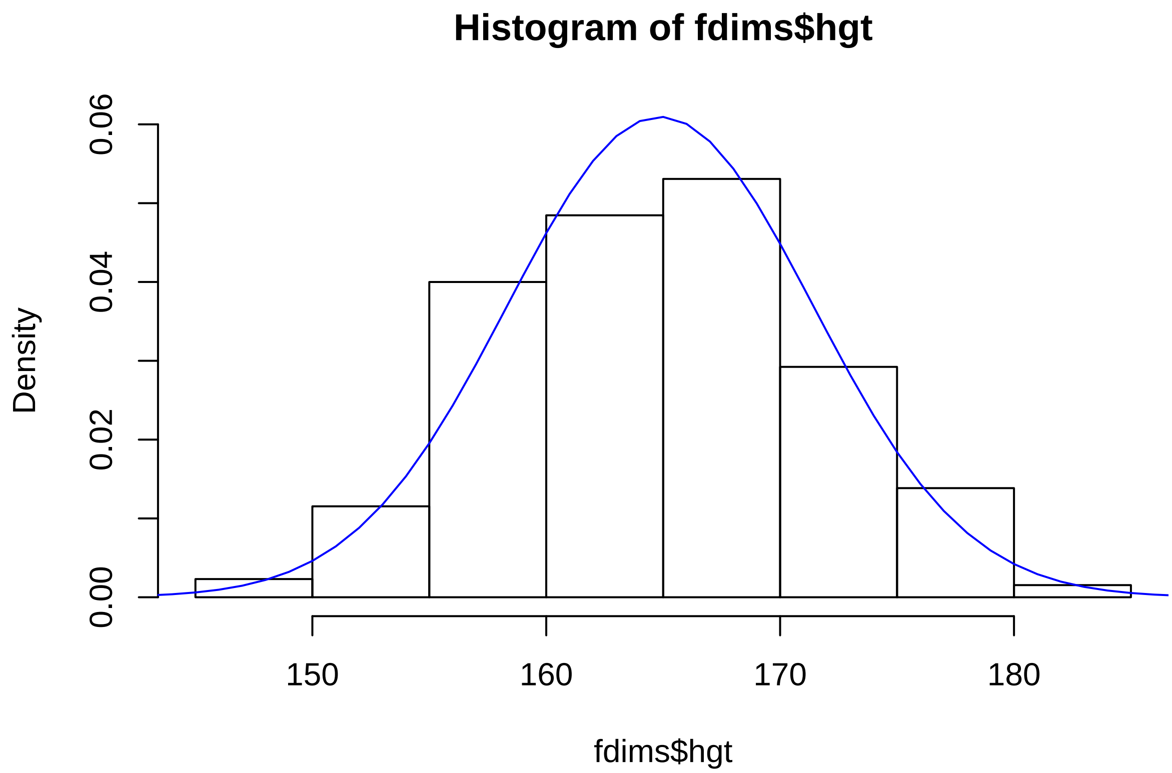
```
fwgtmean
```

```
## [1] 60.60038
```

```
fwgtsd
```

```
## [1] 9.615699
```

```
hist(fdims$hgt, probability = TRUE, ylim = c(0, 0.06))
x <- 140:190
y <- dnorm(x = x, mean = fhgtmean, sd = fhgtsd)
lines(x = x, y = y, col = "blue")
```

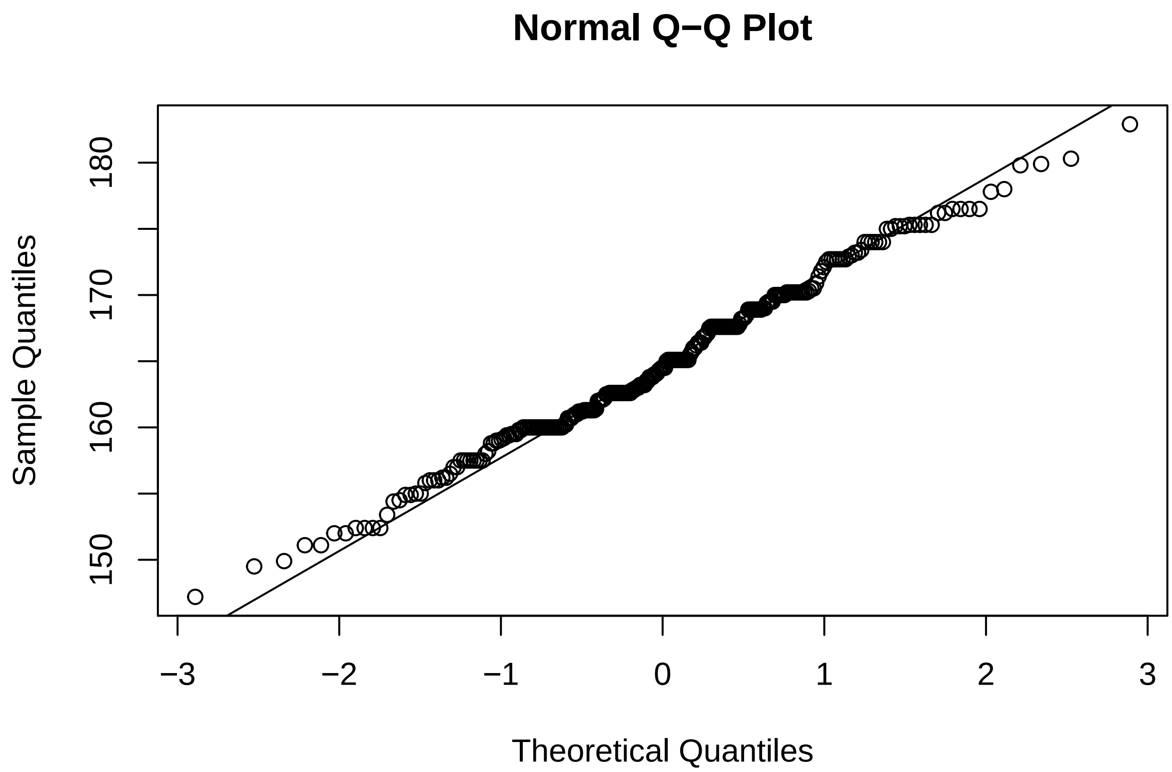


2. Based on the this plot, does it appear that the data follow a nearly normal distribution?

```
Yes, based on the plot it appears to follow a normal distribution.
```

Evaluating the normal distribution

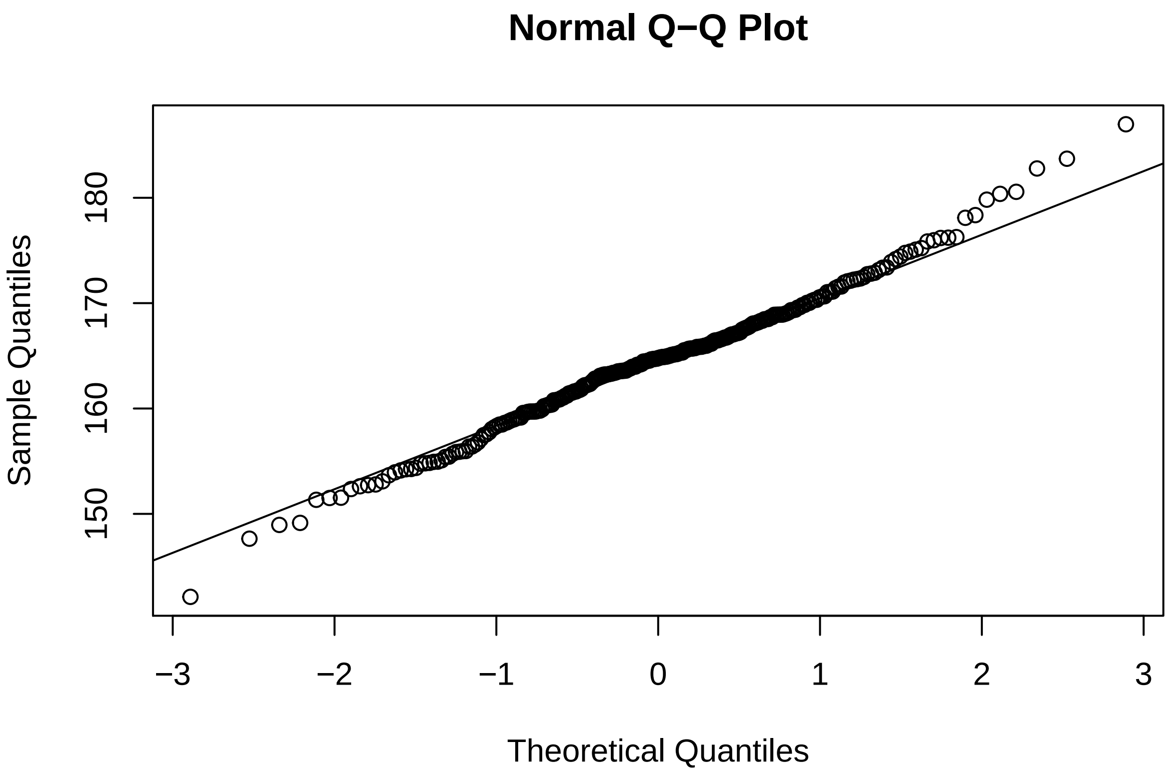
```
qqnorm(fdims$hgt)
qqline(fdims$hgt)
```




```
sim_norm <- rnorm(n = length(fdims$hgt), mean = fhgtmean, sd = fhgtsd)
```

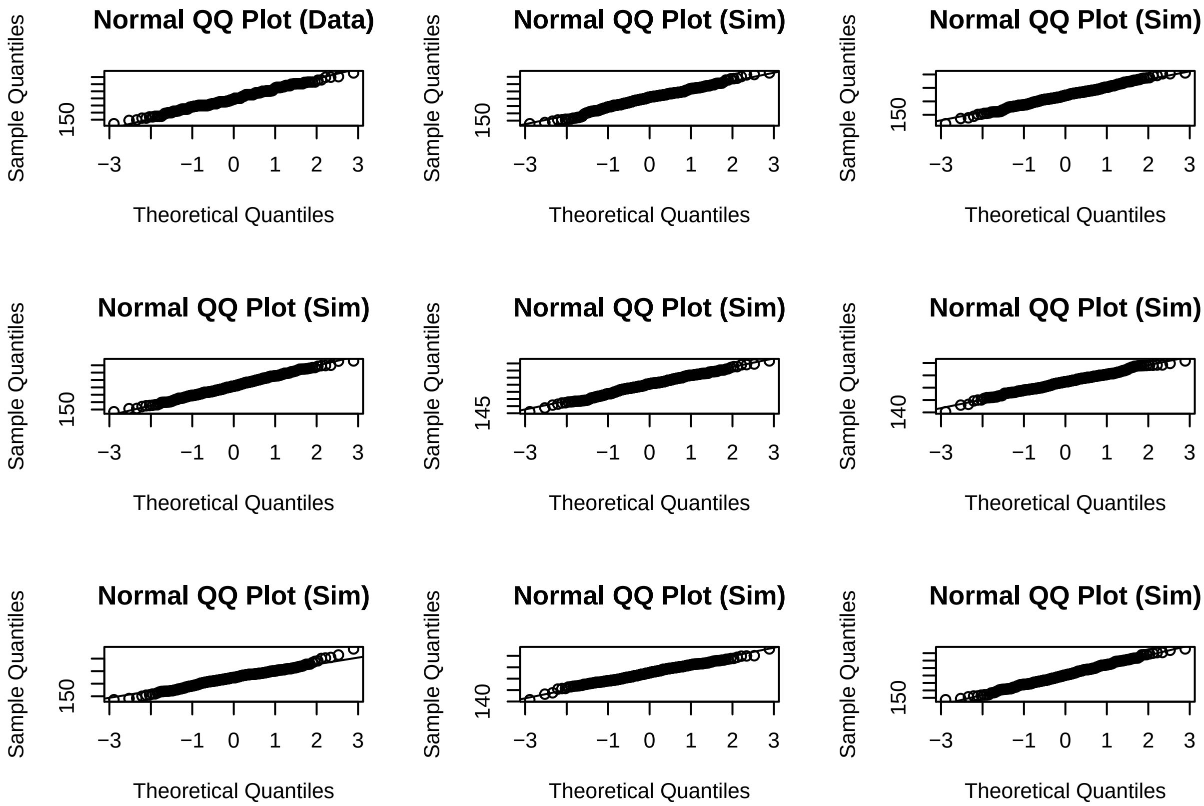
3. Make a normal probability plot of `sim_norm`. Do all of the points fall on the line? How does this plot compare to the probability plot for the real data?

```
qqnorm(sim_norm)
qqline(sim_norm)
```



the simulated model is similar to the probability plot of the real data where most of the data is on the line except at the tails.

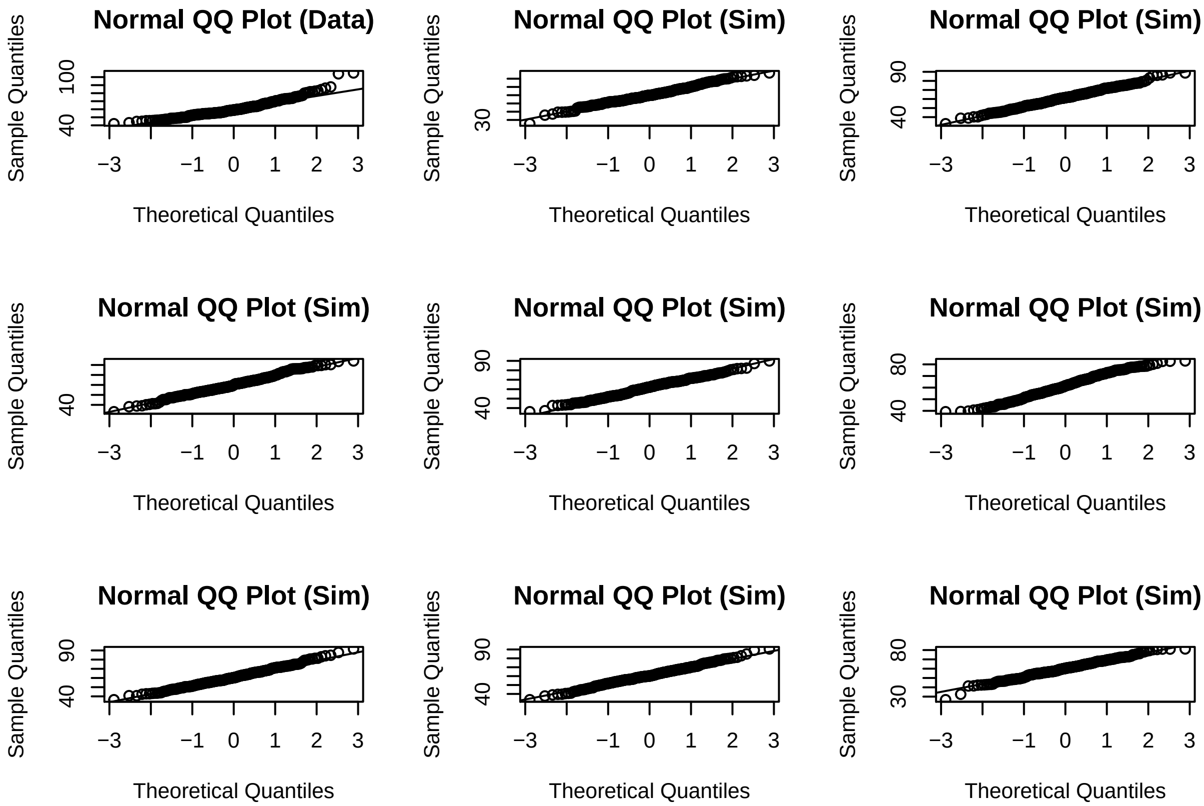
```
qqnormsim(fdims$hgt)
```



4. Does the normal probability plot for `fdims$hgt` look similar to the plots created for the simulated data? That is, do plots provide evidence that the female heights are nearly normal?

   Yes, the normal probablity plot for female heights look similar the plots created for the simulated data. Its follows the line except for at the tails of the line.

```
F_Weight <- qqnormsim(fdims$wgt)
```




```
F_Weight
```

```
## $mfrow
## [1] 3 3
```

5. Using the same technique, determine whether or not female weights appear to come from a normal distribution.

Female weight also appear to come from a normal distribution except at the tails but, the simulations do not have the outliers that are in the real data .

Normal probabilities

```r
1 - pnorm(q = 182, mean = fhgtmean, sd = fhgtsd)
```

```
## [1] 0.004434387
```

```r
sum(fdims$hgt > 182) / length(fdims$hgt)
```

```
## [1] 0.003846154
```

6. Write out two probability questions that you would like to answer; one regarding female heights and one regarding female weights. Calculate the those probabilities using both the theoretical normal distribution as well as the empirical distribution (four probabilities in all). Which variable, height or weight, had a closer agreement between the two methods?

What is the probablity that randomly chosen female is taller than (5'11) 180 cm?

```
1 - pnorm(q = 180, mean = fhgtmean, sd =fhgtsd)
```

```
## [1] 0.01040328
```

```
sum(fdims$hgt > 180) / length(fdims$hgt)
```

```
## [1] 0.007692308
```

What is the probablity that randomly chosen female is under 175 lbs?

```
pnorm(q = 79, mean = fwgtmean, sd =fwgtsd)
```

```
## [1] 0.9721578
```

```
sum(fdims$wgt < 79) / length(fdims$wgt)
```

```
## [1] 0.9538462
```

Between the probability models, the weight had a closer agreement between the two methods. The theoretical normal distribution predicts that there are 97.2% chance that a random chosen female is under 175 lbs while the empirical distribution says that there are is a 95.4% chance.

On Your Own

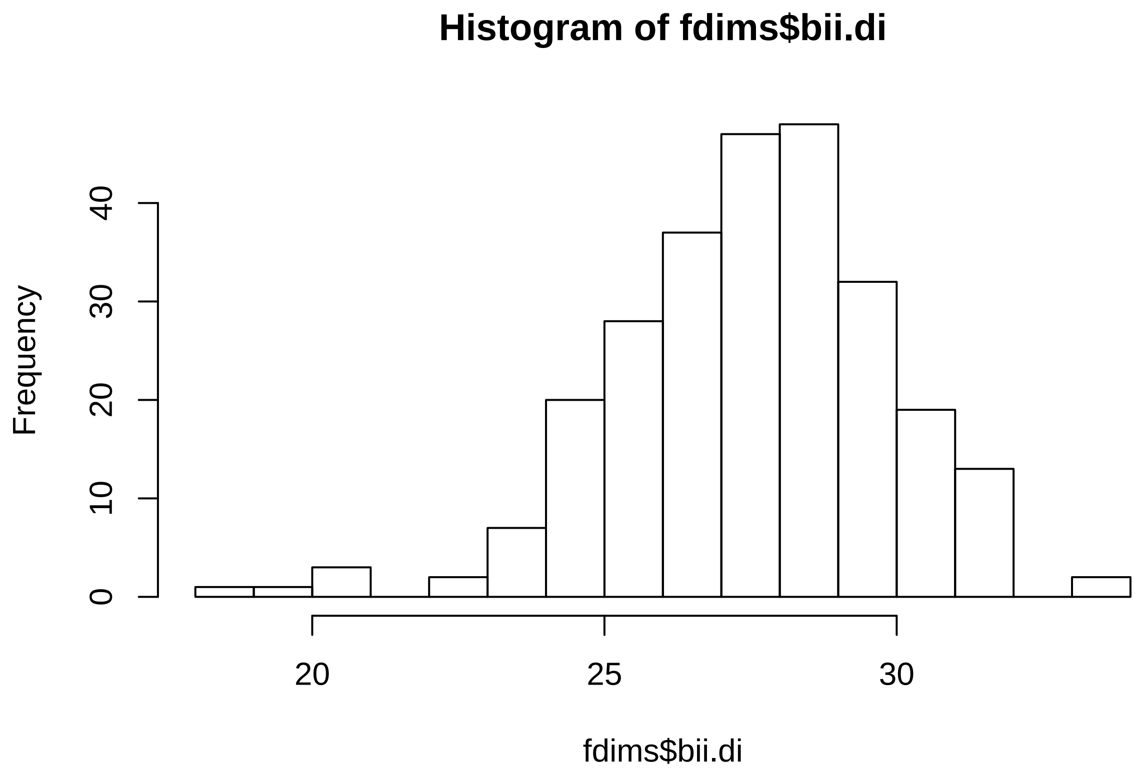
- Now let's consider some of the other variables in the body dimensions data set. Using the figures at the end of the exercises, match the histogram to its normal probability plot. All of the variables have been standardized (first subtract the mean, then divide by the standard deviation), so the units won't be of any help. If you are uncertain based on these figures, generate the plots in R to check.

```
fdims[1:20 ,]
```

```
##     bia.di bii.di bit.di che.de che.di elb.di wri.di kne.di ank.di sho.gi
## 248   37.6   25.0   31.3   16.2   24.9   11.2    9.2   17.0   12.3   95.0
## 249   36.7   26.4   31.0   16.8   24.5   12.1    9.9   19.3   12.8   99.5
## 250   34.8   25.9   30.2   16.4   24.2   11.3    8.9   17.0   12.2   88.0
## 251   36.6   27.9   31.8   19.3   24.9   12.3    9.5   18.6   13.0   97.0
## 252   35.5   28.2   31.0   18.2   26.2   11.5    9.1   17.2   12.4  103.3
## 253   37.0   28.0   32.0   15.1   25.7   12.5   10.0   17.2   13.2   93.5
## 254   35.5   26.5   29.2   15.4   24.5   12.3    9.4   17.2   12.0   93.3
## 255   37.4   30.2   33.2   18.8   26.6   13.3   10.7   19.8   13.8   94.5
## 256   37.8   29.0   32.6   18.6   25.0   12.1    9.8   17.8   12.7   98.6
## 257   38.6   28.8   33.2   19.7   29.4   13.4   11.5   20.9   13.2  115.5
## 258   37.6   28.5   32.2   15.5   24.3   11.8    8.6   17.1   11.9   97.9
## 259   36.0   25.6   31.5   15.4   25.5   12.8    9.7   17.6   13.2   97.7
## 260   39.5   30.0   31.7   17.3   27.3   12.8    9.2   18.1   12.4  100.5
## 261   34.0   25.0   27.0   16.9   22.6   10.6    8.3   15.9   11.6   88.7
## 262   35.0   26.5   31.6   18.3   23.7   11.5    8.6   16.8   12.2   96.6
## 263   35.6   25.8   32.0   16.2   25.7   11.5    9.0   17.2   11.8   92.0
## 264   36.2   27.4   29.5   14.6   23.9   11.2    9.6   16.7   12.6   90.0
## 265   39.0   28.4   34.9   19.6   26.7   13.4   11.0   18.9   13.6  104.0
## 266   32.6   25.6   30.0   15.3   22.6   10.3    8.1   16.2   11.6   90.1
## 267   37.6   30.0   33.9   19.1   28.8   13.4   10.5   19.2   13.2  104.0
##     che.gi wai.gi nav.gi hip.gi thi.gi bic.gi for.gi kne.gi cal.gi ank.gi
## 248   83.0   66.5   79.0   92.0   53.5   24.3   20.5   32.0   32.2   21.0
## 249   78.5   61.5   70.5   90.5   57.7   27.8   24.0   38.5   38.5   22.5
## 250   75.0   61.2   66.5   91.0   53.0   24.0   22.0   32.5   32.5   19.0
## 251   86.5   78.0   91.0   99.5   61.5   28.0   24.0   35.2   36.7   23.0
```

```
## 252   91.0   70.5   80.5   91.5   55.0   26.9   22.7   33.0   33.3   19.9
## 253   79.5   66.5   78.5   94.0   54.0   26.5   22.5   34.0   35.0   23.0
## 254   77.0   58.0   64.0   85.5   49.5   24.1   22.0   32.5   32.0   19.0
## 255   88.0   74.5   87.0  104.0   64.0   29.2   26.2   38.5   38.0   22.0
## 256   85.0   73.5   92.0  104.1   65.3   29.0   23.4   35.3   37.4   21.6
## 257   98.8   90.5  103.5  108.1   61.1   33.6   26.6   37.2   35.8   22.6
## 258   79.0   66.5   74.0   90.3   52.0   24.8   21.0   32.2   32.5   19.9
## 259   77.6   61.0   71.8   91.6   53.0   25.4   22.6   34.0   34.5   20.5
## 260   85.0   69.5   81.5   94.4   55.8   25.9   22.9   36.1   35.3   20.9
## 261   76.7   62.0   74.1   80.9   48.8   24.0   20.5   30.8   30.4   17.9
## 262   76.7   63.4   69.0   87.7   54.0   25.6   21.6   34.4   32.8   19.1
## 263   82.0   71.0   69.0   88.5   54.5   26.0   21.8   33.5   35.0   21.0
## 264   79.0   59.0   79.0   88.5   51.2   23.5   21.0   32.5   29.6   18.5
## 265   89.5   74.0   92.0  101.0   61.5   31.0   26.5   38.6   38.6   23.8
## 266   73.5   60.5   68.3   88.5   54.0   24.6   20.6   29.0   33.0   19.0
## 267   90.0   75.0   80.5   99.0   59.0   28.7   24.9   37.0   36.5   20.6
##     wri.gi age  wgt   hgt sex
## 248   13.5  22 51.6 161.2   0
## 249   15.0  20 59.0 167.5   0
## 250   14.0  19 49.2 159.5   0
## 251   15.0  25 63.0 157.0   0
## 252   14.5  21 53.6 155.8   0
## 253   14.5  23 59.0 170.0   0
## 254   13.9  26 47.6 159.1   0
## 255   16.8  22 69.8 166.0   0
## 256   15.2  28 66.8 176.2   0
## 257   16.3  40 75.2 160.2   0
## 258   13.8  32 55.2 172.5   0
## 259   15.3  25 54.2 170.9   0
## 260   14.4  25 62.5 172.9   0
## 261   13.2  29 42.0 153.4   0
## 262   13.8  22 50.0 160.0   0
## 263   14.0  25 49.8 147.2   0
## 264   14.5  23 49.2 168.2   0
## 265   17.0  37 73.2 175.0   0
## 266   13.2  19 47.8 157.0   0
## 267   15.9  23 68.8 167.6   0
```

```
hist(fdims$bii.di, breaks = 20)
```




```
**a.** The histogram for female biiliac (pelvic) diameter (`bii.di`) belongs
to normal probability plot letter D .

**b.** The histogram for female elbow diameter (`elb.di`) belongs to normal
probability plot letter C.

**c.** The histogram for general age (`age`) belongs to normal probability
plot letter B.

**d.** The histogram for female chest depth (`che.de`) belongs to normal
probability plot letter A.
```

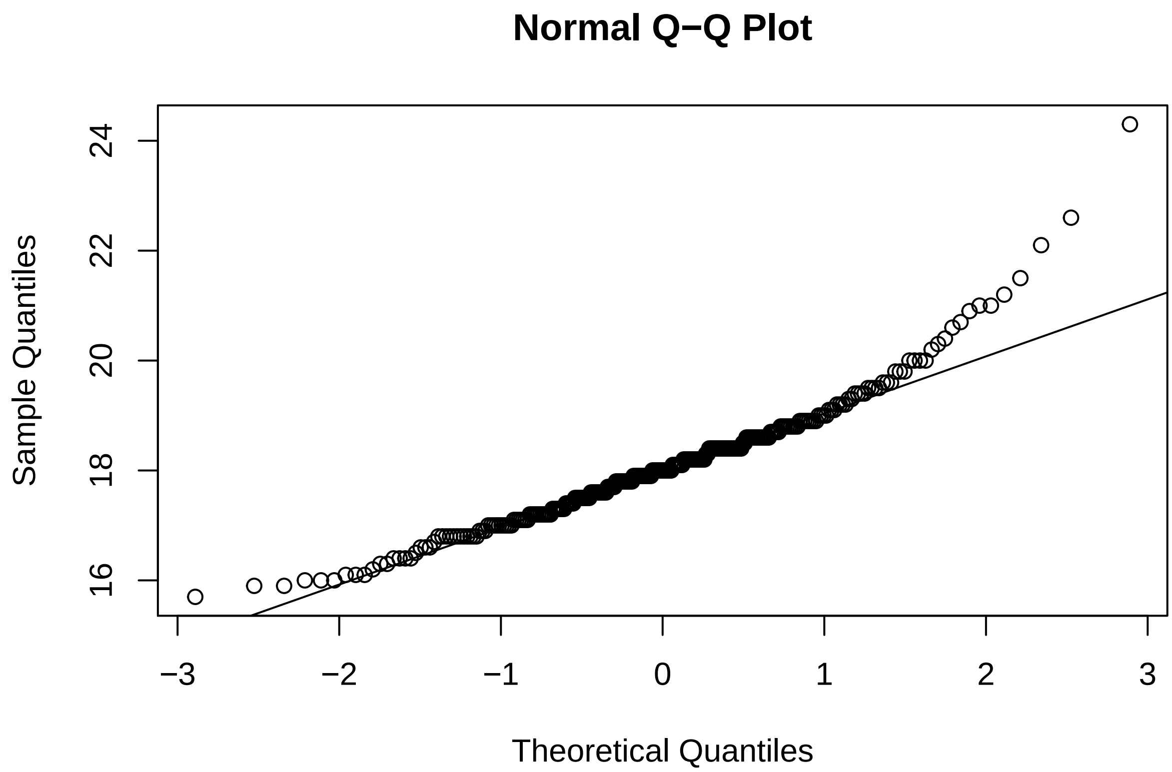
- Note that normal probability plots C and D have a slight stepwise pattern. Why do you think this is the case?

The are measurements of diameters of the pelvic and elbow, and there they're all around the same size.

- As you can see, normal probability plots can be used both to assess normality and visualize skewness. Make a normal probability plot for female knee diameter (`kne.di`). Based on this normal probability plot, is this variable left skewed, symmetric, or right skewed? Use a histogram to confirm your findings.

The normal probability plot for female knee diameter looks like it is skewed right.

```
qqnorm(fdims$kne.di)
qqline(fdims$kne.di)
```

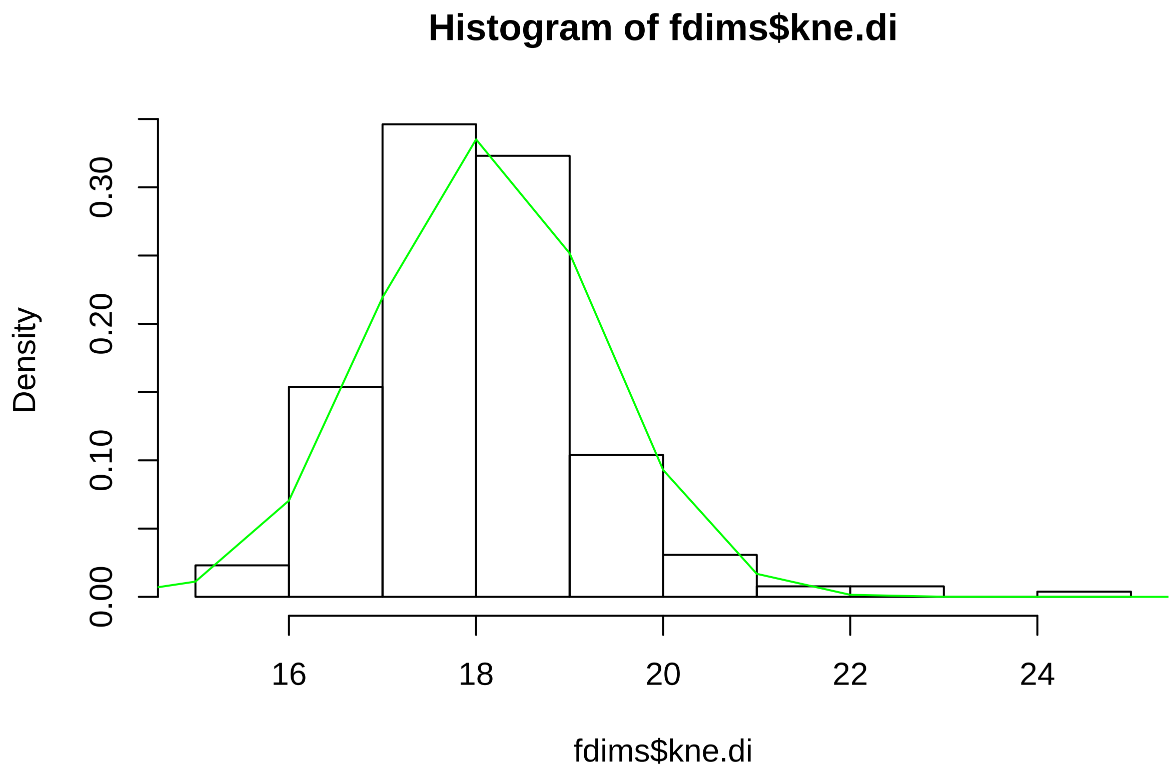



```r
range(fdims$kne.di)
```

```
## [1] 15.7 24.3
```

```r
fknee.dimean <- mean(fdims$kne.di)
fknee.disd <-sd(fdims$kne.di)

hist(fdims$kne.di, probability = TRUE,)
x <- 1:100
y <- dnorm(x = x, mean = fknee.dimean, sd = fknee.disd)
lines(x = x, y = y, col = "green")
```



Based on the historgram, it looks like the data is skewed right.