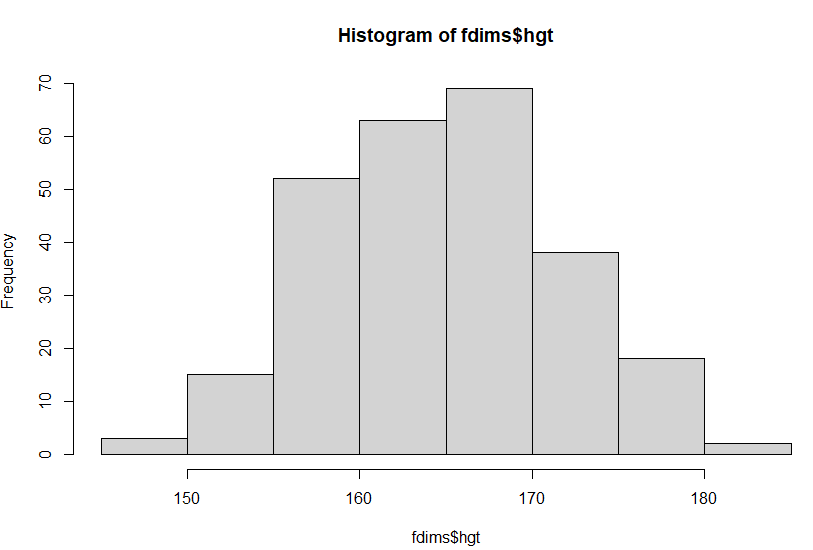
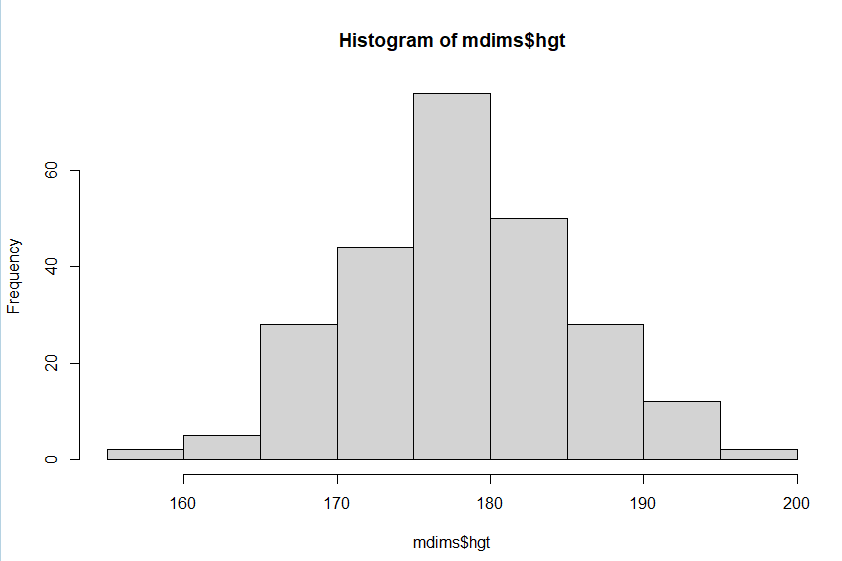
Lab 3

**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?**

Hist(mdims$hgt)

Hist(fdims$hgt)



Both seem normally distributed and unimodal, but the men’s histogram appears more symmetric, while the female plot have a smaller range and slight right skew. Men have a larger mean and range.

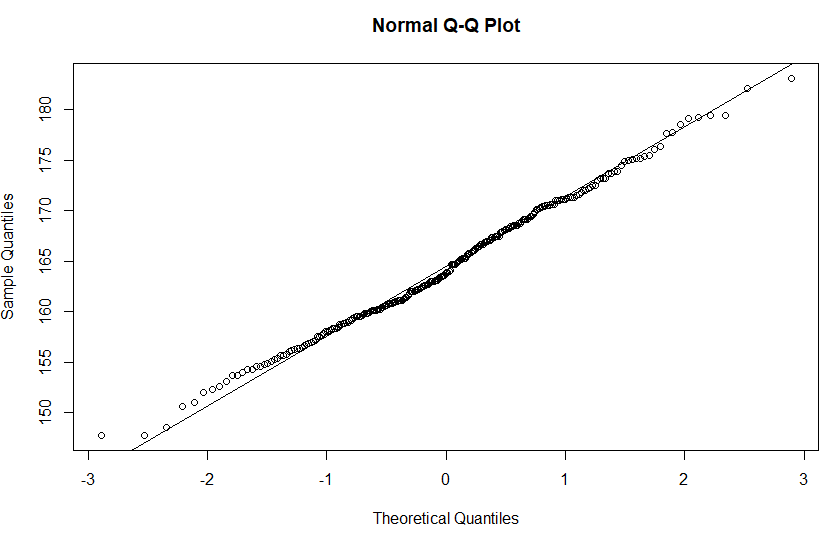
**2) Based on this plot, does it appear that the data follow a nearly normal distribution?**

Yes, it appears that the data follows a nearly normal distribution, as the plot is unimodal and symmetric and strongly correlates to the curve of the line

**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)

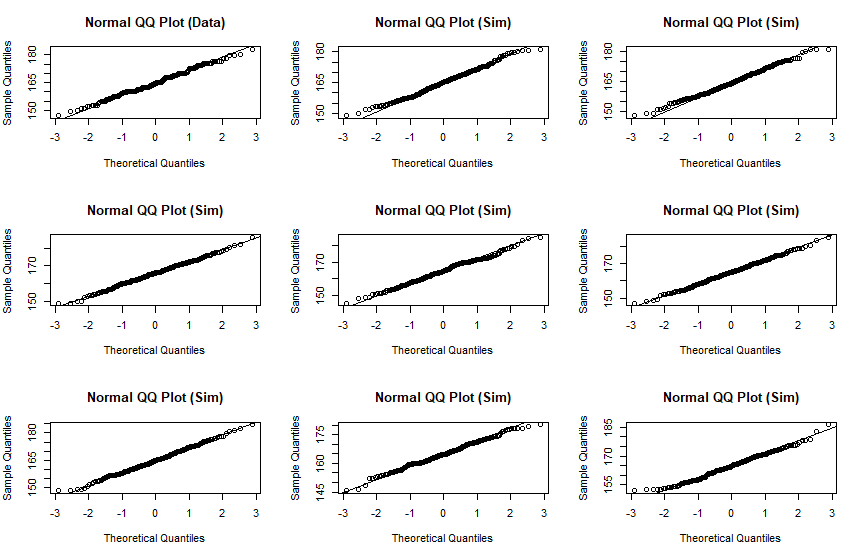
qqline(sim\_norm)



No, not all the plots on the head and tail fall on the line. They more closely follow the line than the real data, especially in the middle of the line. It is more linear than the real data.

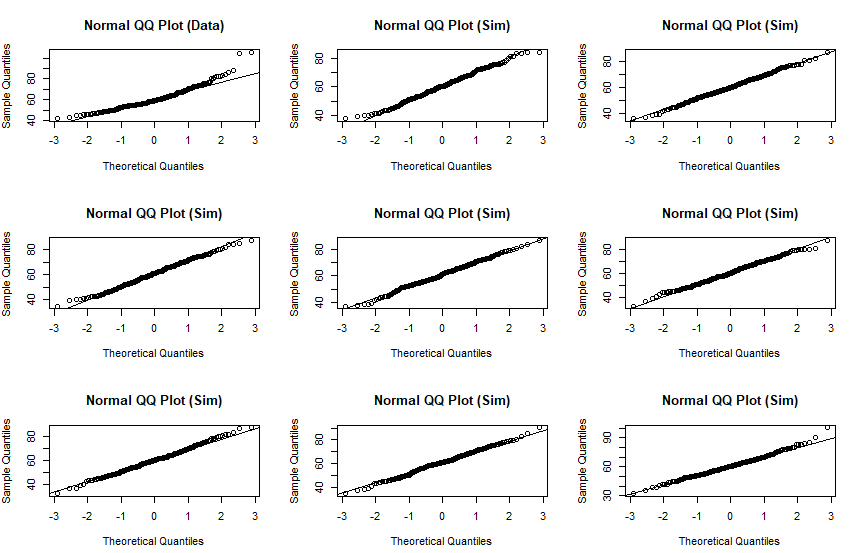
**4) Does the normal probability plot for fdims$hgt look similar to the plots created for the simulated data? That is, do plots provide evidence that the female heights are nearly normal?**

qqnormsim(fdims$hgt)



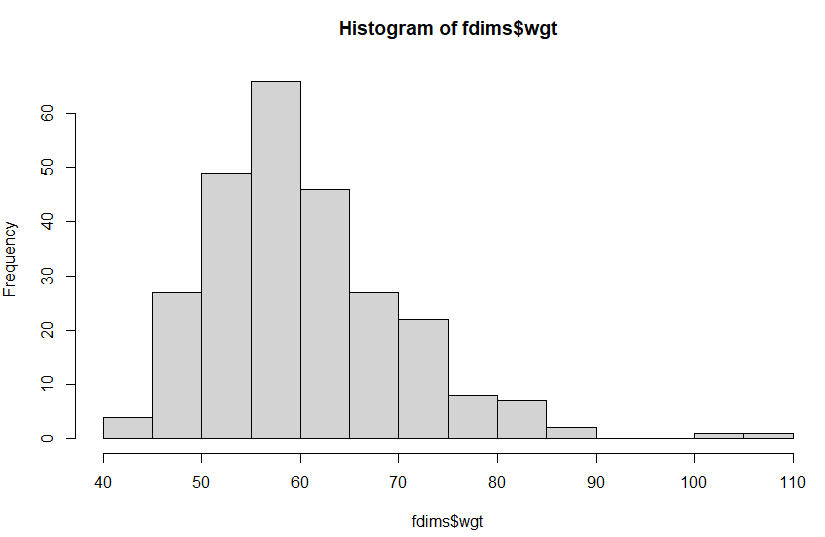
Yes, the normal plot looks very similar to the simulated data, but it appears slightly choppier in the body of the line than the simulated plots. The maximums and minimums appear similar for both types. The simulated plots all appear a little more linear or closer to the line, but not by much. These plots do provide evidence that the female heights are nearly normal, as they are very similar.

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

qqnormsim(fdims$wgt)

Weight appears less normal and more skewed to the right than the simulated plots, as the top of the normal plot veers off the curve near the maximum. This may lead us to believe that the weights do not come from a normal distribution. The histogram plotted below shows there is a right skew.

Hist(fdims$wgt)



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?**

1) What is the probability that a female’s height is under 168 cm, based on the data?

pnorm(q = 168, mean = fhgtmean, sd = fhgtsd)

= 0.6836408

sum(fdims$hgt < 168) / length(fdims$hgt)

= 0.6846154

2) What is the probability that a female’s weight is over 70 kg based on the data?

1 - pnorm(q = 70, mean = fwgtmean, sd = fwgtsd)

= 0.1641539

> sum(fdims$wgt > 70) / length(fdims$wgt)

= 0.1576923

Height had a closer agreement between the two methods, as they are virtually identical, while weight had a slightly bigger difference with the two methods. This can likely be explained by the fact that weight can be affected by many more variables than height, which cannot be changed.

**ON YOUR OWN**

1) **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.**

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

qqnorm(fdims$bii.di)

qqline(fdims$bii.di)

**b.** The histogram for female elbow diameter (elb.di) belongs to normal probability plot letter \_\_C\_\_.

qqnorm(fdims$elb.di)

qqline(fdims$elb.di)

**c.** The histogram for general age (age) belongs to normal probability plot letter \_\_D\_\_.

qqnorm(fdims$age)

qqline(fdims$age)

**d.** The histogram for female chest depth (che.de) belongs to normal probability plot letter \_A\_\_\_.

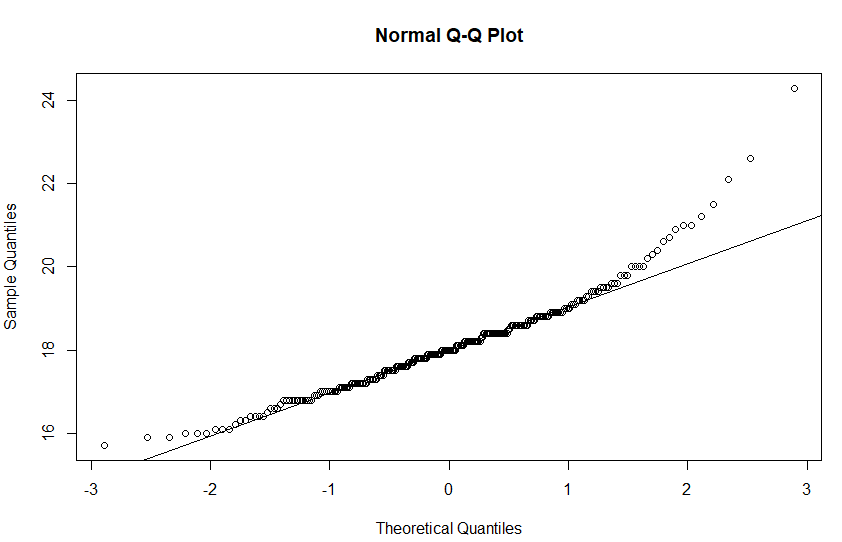
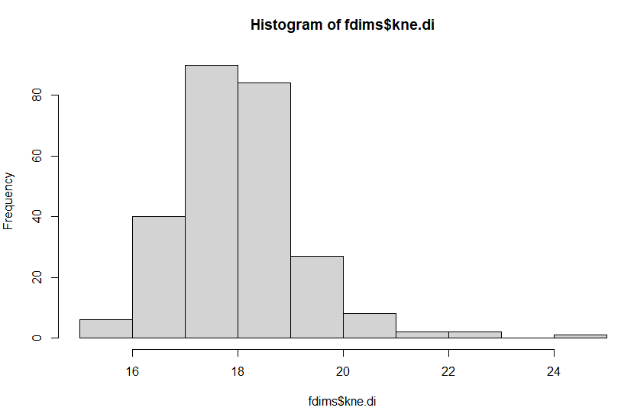
qqnorm(fdims$che.de)

qqline(fdims$che.de)

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

Plots C and D have a stepwise pattern because they have many reoccurring data points, or modes, that repeat the same values more than others. This is visible by the histograms, which show a few values have much higher frequency than others. They are less normally distributed and more skewed. Age is right skewed, and elbow diameter is slightly left skewed. This means the plot points will be less likely to follow the linear line and be a little more sporadically placed.

**3) 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.**



qqnorm(fdims$kne.di)

qqline(fdims$kne.di)

hist(fdims$kne.di)

Based on this normal probability plot, the data appears to be right skewed, as the right end of the plot curves up away from the normal line, indicating the variable is right skewed. The histogram confirms this, as most of the data falls on the left side of the graph, leading to outliers on the right, and giving a right skew.