

9.11 #64

Jin Kweon
SID: 3032235207

a) I assessed normality of each male
b) female, both gender for
both body temperature and heart
rate data, by using 'ggnorm'
function in R.

Found all of them are decent to
be said normal.
With the means and SD I got
from original data, I simulate
five samples to judge the inherent
variability.

I concluded that samples have
small variability, meaning consistent.

* Another way of checking normality
is doing $X_{(k)}$ vs $F^{-1}\left(\frac{k}{n+1}\right)$ in R.

Get MLE of $\hat{\mu} = \bar{X}$ and $\hat{\sigma} = \sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2}$
and use 'ggnorm' (get inverse CDF quantiles)
to check whether data is reasonably
argued as normal

* We try to compare with data sampled
from normal dist and our
original data, to judge variability.

c)
 H_0 : Mean body temperature is 98.6°
vs.
 H_A : Mean body temperature is not 98.6° .
(standard deviation is not known).

So, I need to use two-sided t-test.
For each male/female test,
degree of freedom is $65-1 (=64)$.

As p-value is small, we know we
are more likely to reject
the null hypothesis. (or, I can say extreme: large/small values of T is rejected)
For male, we reject the null all
the times.

For female, we reject the null
at $\alpha = 0.05$, but not
reject at $\alpha = 0.01$.

Jin Kweon_3032235207_Lab2

Jin Kweon

4/7/2017

Data import

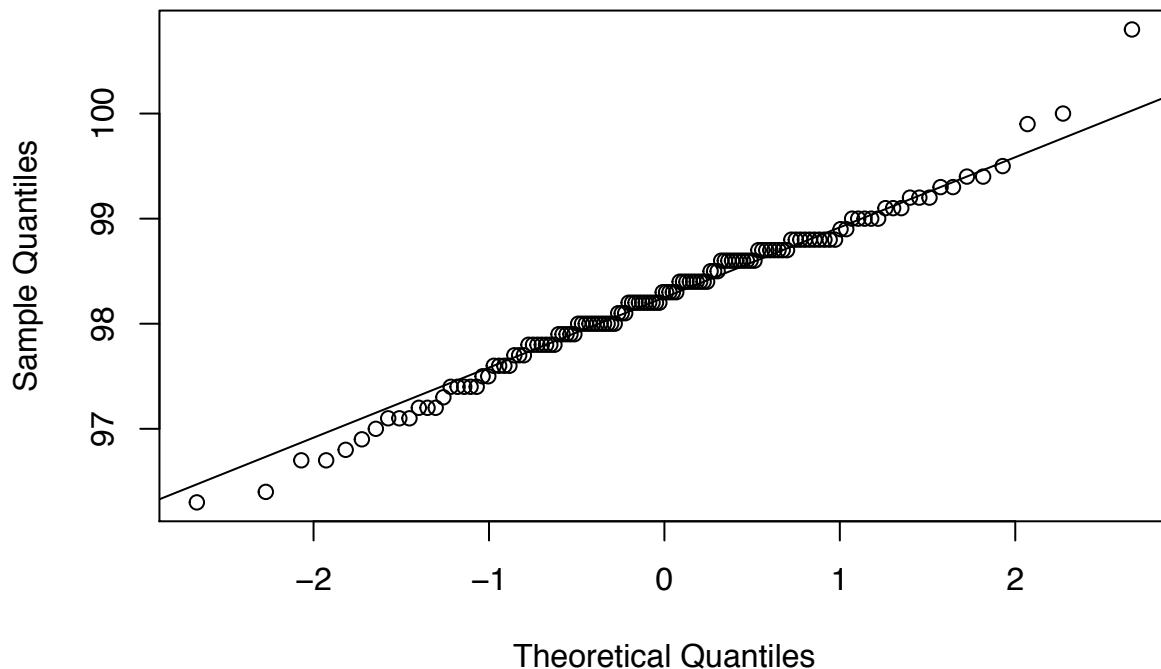
```
bodydata <- read.csv("/bodytemp.txt", header = TRUE, dec = ",", stringsAsFactors = FALSE)
options(digits=9, stringsAsFactors = FALSE)

# My default of the laptop is changing every data imported to characters, so I manually
# changed all the data to numeric, as you can see it below.
for (i in 1:3){
  bodydata[,i] <- as.numeric(bodydata[,i])
}
```

Check normality of data: 1. temperature (mixed) 2. temperature (male) 3. temperature (female) 4. heart rate (mixed) 5. heart rate (male) 6. heart rate (female)

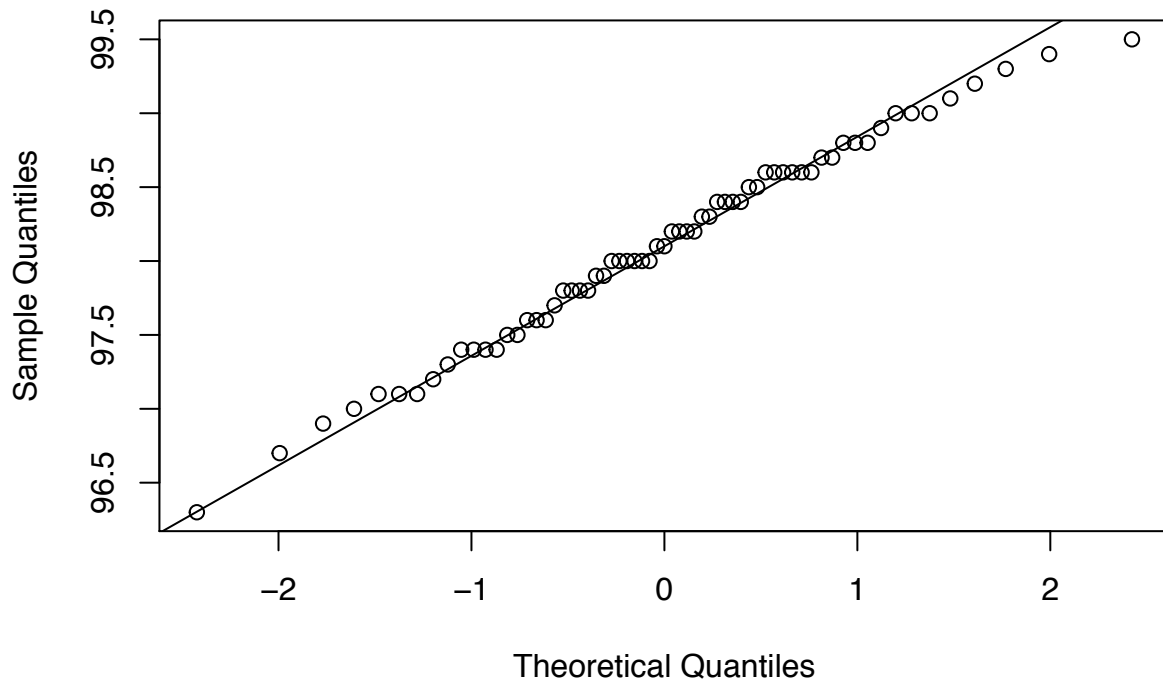
```
qqnorm(bodydata$temperature) #1
qqline(bodydata$temperature)
```

Normal Q-Q Plot



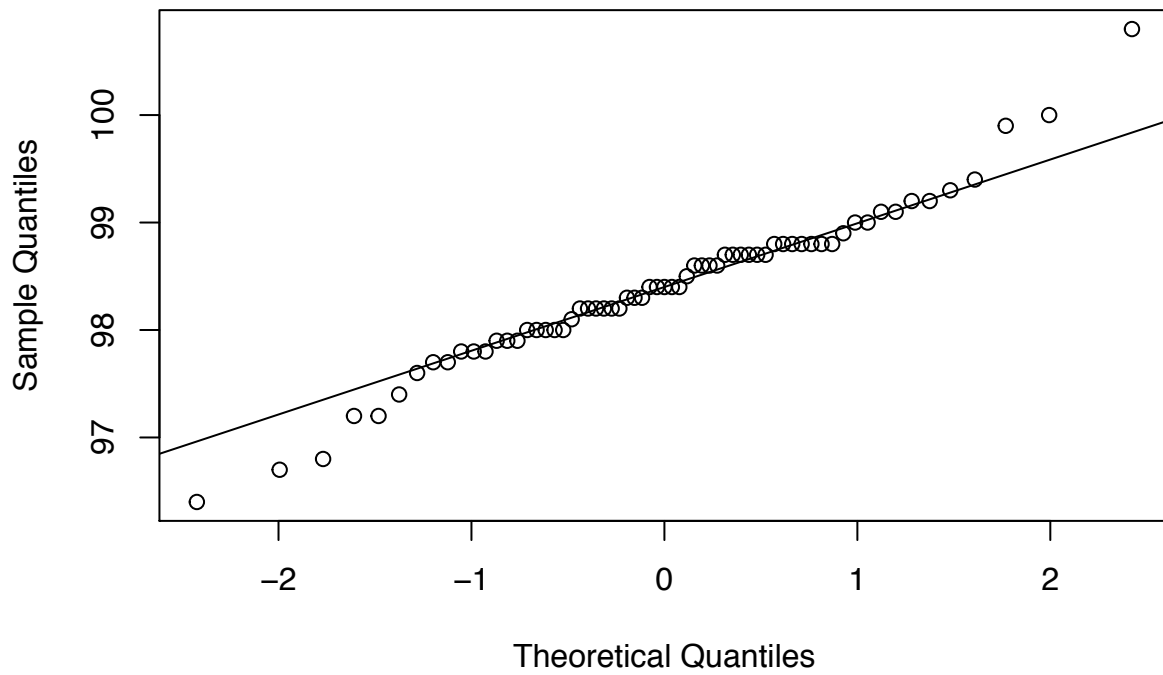
```
qqnorm(bodydata$temperature[1:65]) #2
qqline(bodydata$temperature[1:65])
```

Normal Q-Q Plot



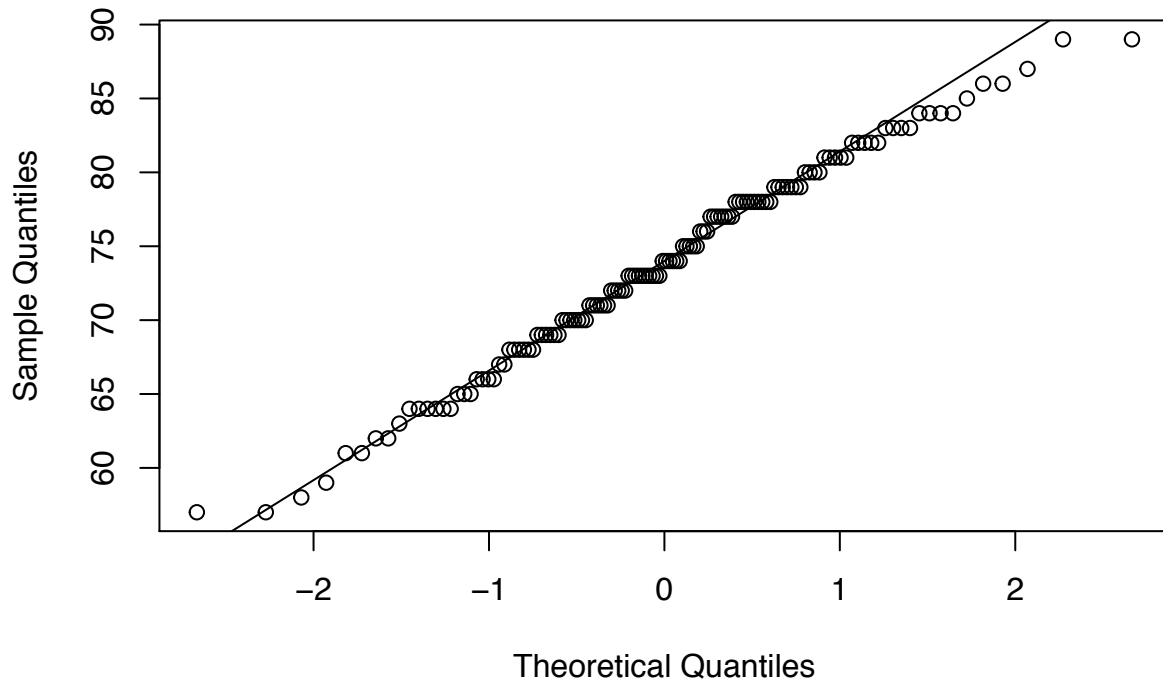
```
qqnorm(bodydata$temperature[66:130]) #3  
qqline(bodydata$temperature[66:130])
```

Normal Q-Q Plot



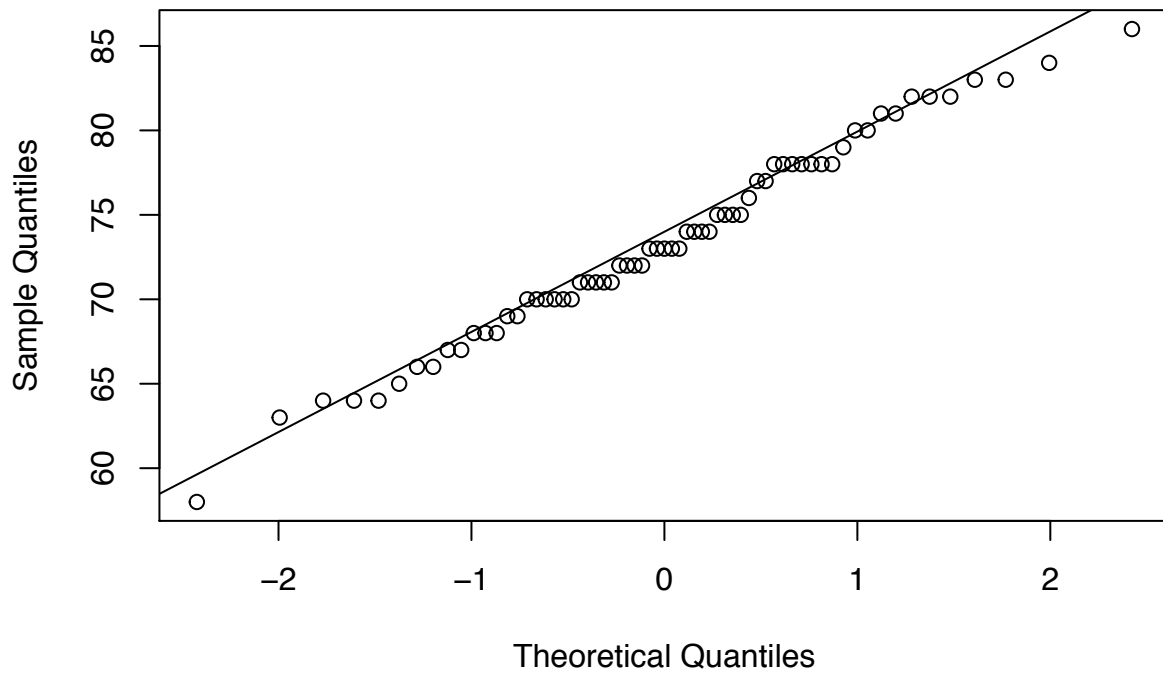
```
qqnorm(bodydata$rate) #4  
qqline(bodydata$rate)
```

Normal Q-Q Plot



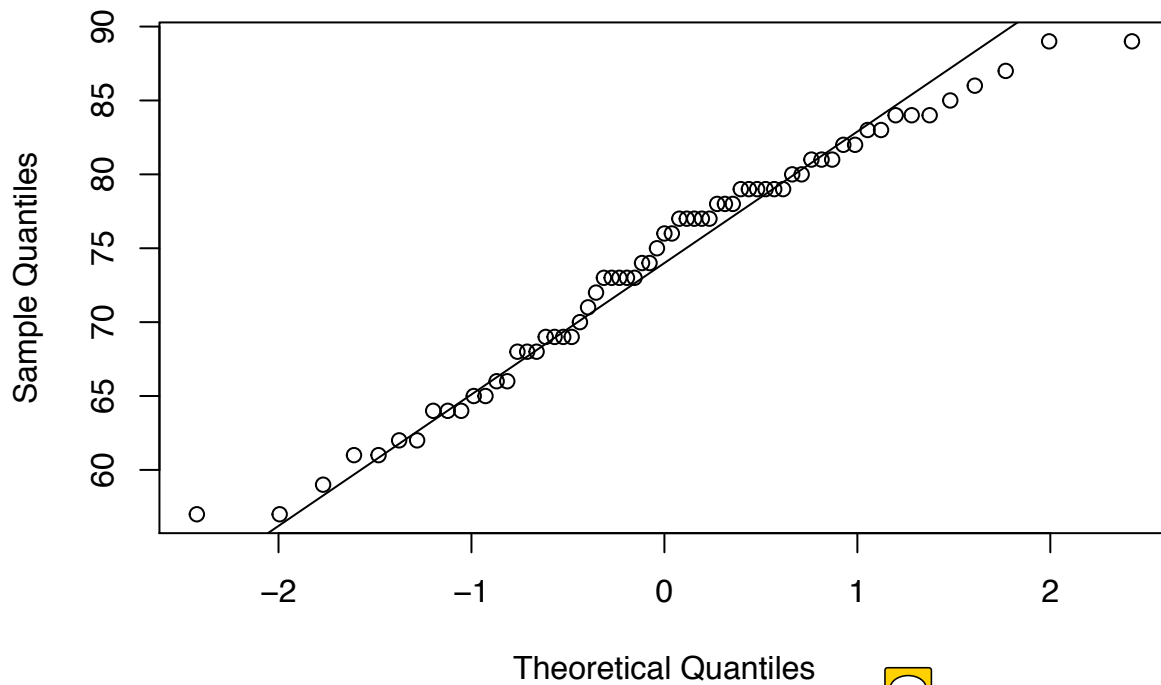
```
qqnorm(bodydata$rate[1:65]) #5  
qqline(bodydata$rate[1:65])
```

Normal Q-Q Plot



```
qqnorm(bodydata$rate[66:130]) #6  
qqline(bodydata$rate[66:130])
```

Normal Q-Q Plot



Generally, female's data (for both heart rate and temperature) are skewed.

A

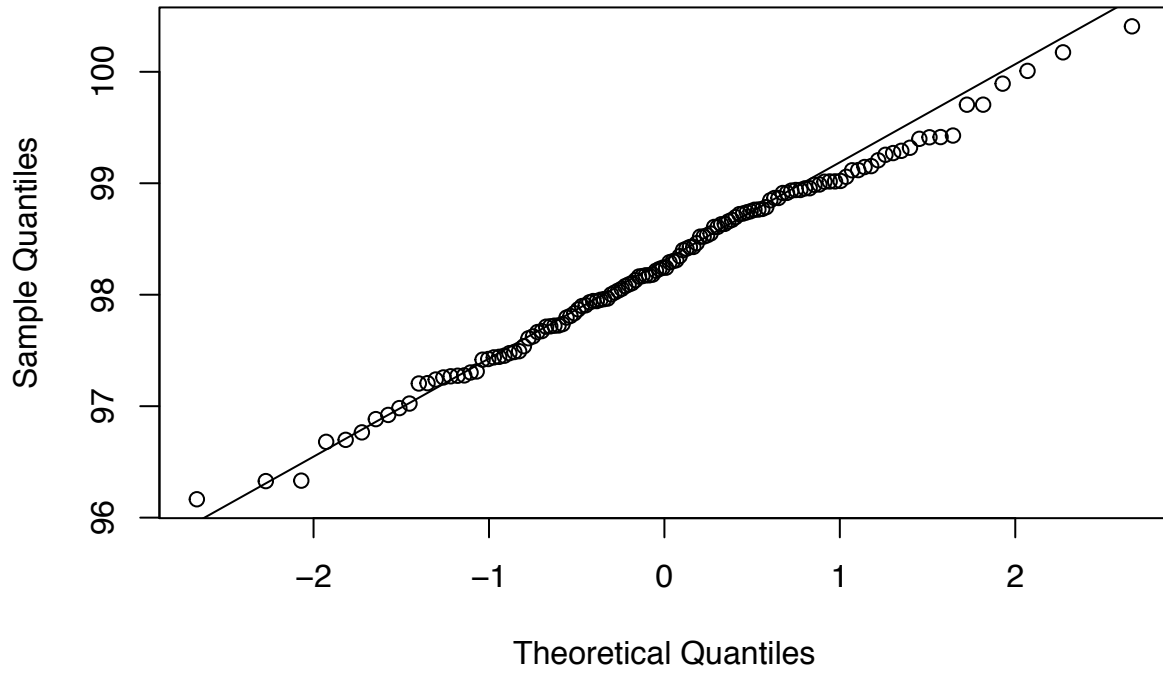
```
temp_mean <- mean(bodydata$temperature) #Get mean of temperature
temp_sd <- sd(bodydata$temperature) #Get standard deviation of temperature

simulation_provider <- function(){
  mean_simulation1 <- rnorm(n = length(bodydata$temperature), mean = temp_mean, sd = temp_sd)

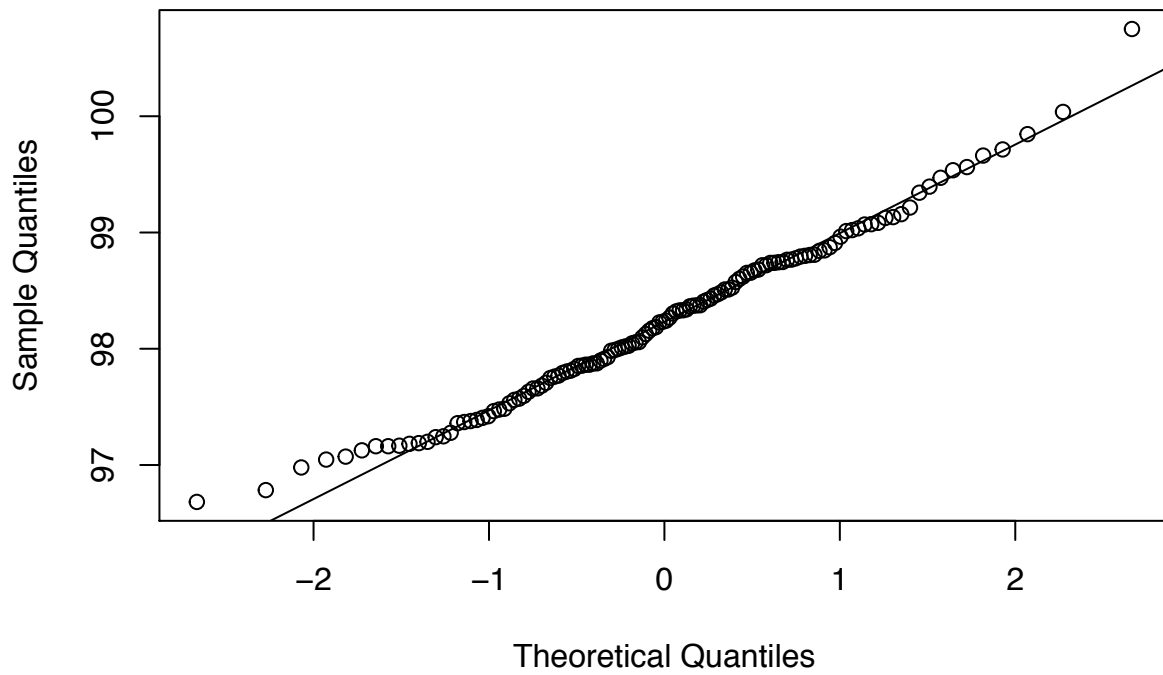
  qqnorm(sort(mean_simulation1), main = "Temperature Normality of male and female")
  #do not have to "sort"!
  qqline(mean_simulation1)
}

#Simulate samples 5 times
for (i in 1:5){
  simulation_provider()
}
```

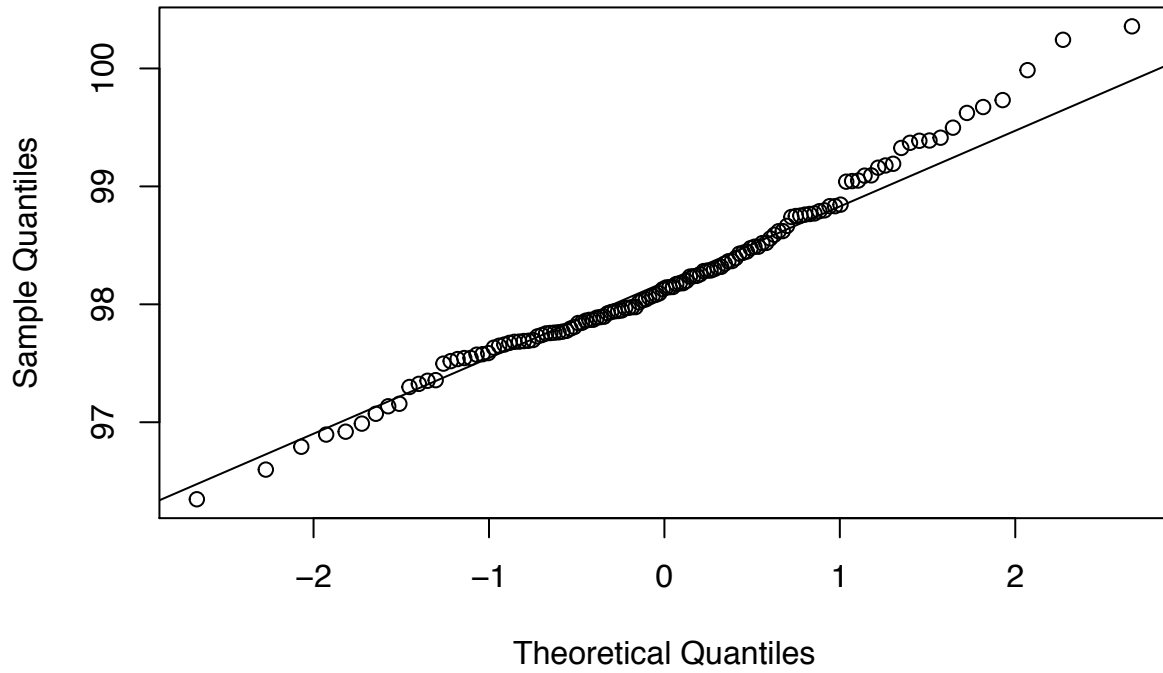
Temperature Normality of male and female



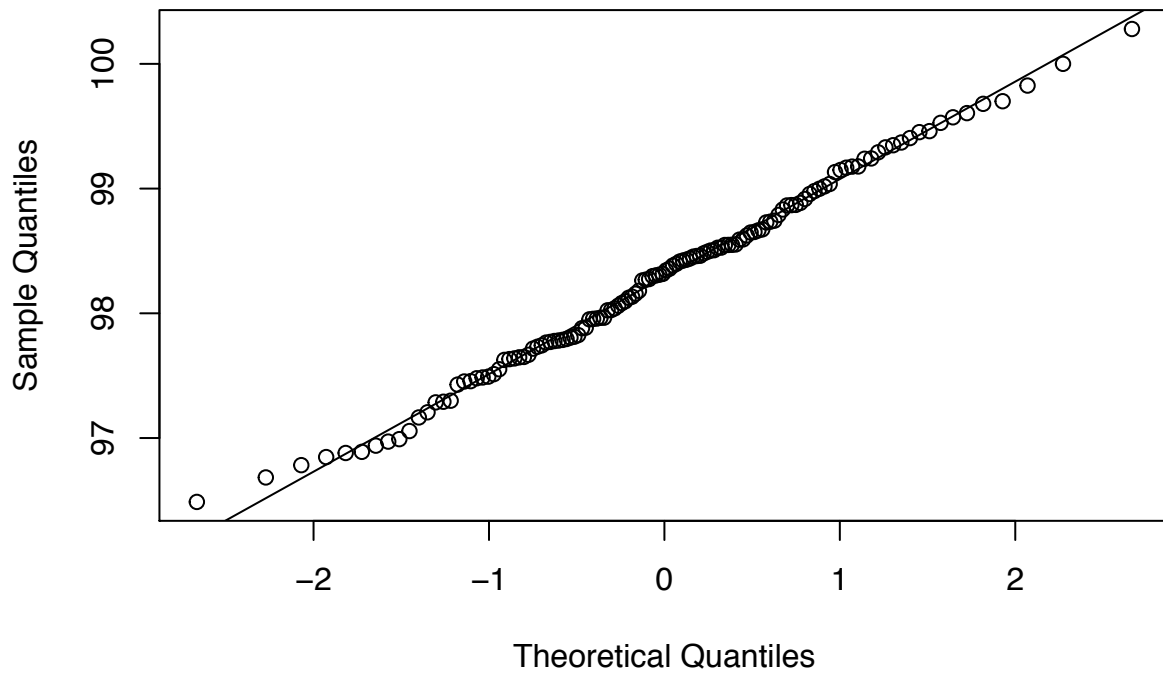
Temperature Normality of male and female



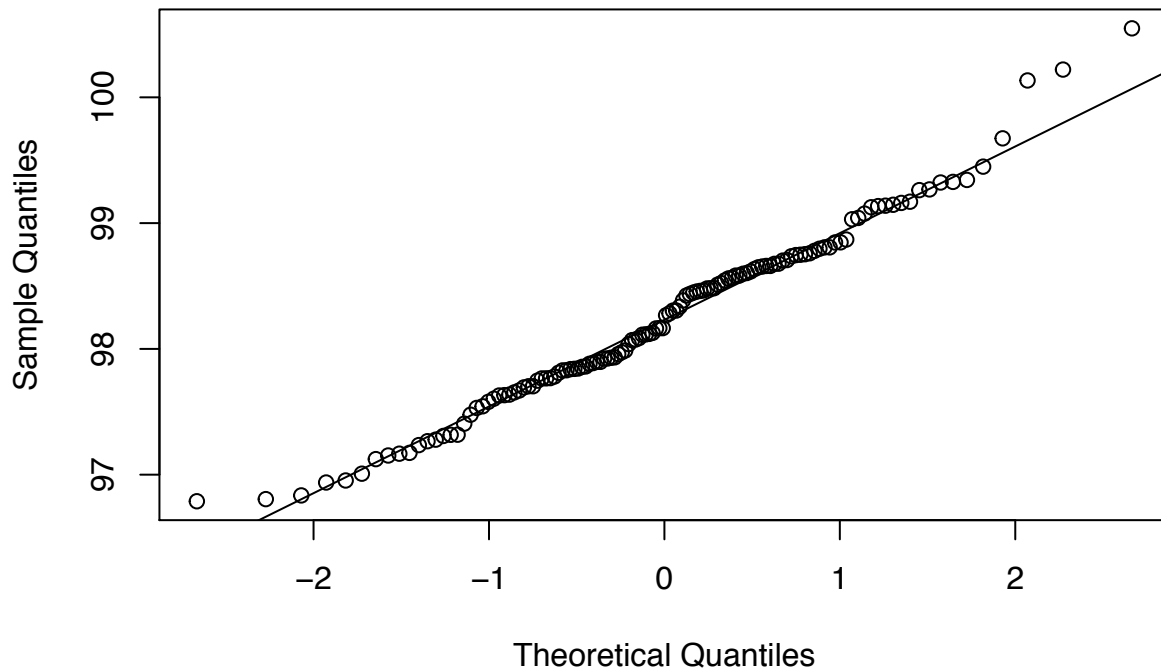
Temparature Normality of male and female



Temparature Normality of male and female



Temperature Normality of male and female



First of all, the most important thing for judging the variability of these plots is to compare those five samples with the normal qq plots with the original data. We saw the original data was following the normal quite well, and we got mean and variance from original data and do qqnorm five times. Then, those five times of qqnorm samples should look similar with qqnorm with the original data, to conclude that our original data has pretty decent variability. And, our samples turned out to be! There are two ways of justifying these. First (easy one), is just simply comparing the graphs by just observing them. Second (hard way), is to divide by sections and get observations of the points (from original data) and get points from one sample (theoretical) and check goodness-of-fit test/Pearson's chi square test five times (since I decide to have five samples), but the GSIs told me this test is not required for this question, so I just concluding the answer of variability visual way only!

So this is how I did. To check the normality, I get the mean of temperatures for both female and male's body temperatures. (since i need matching parameters from my original data) First, I use rnorm to with those mean and sample derived from the simulation. Second, just to visually check whether the temperatures are normally distributed, I used qqnorm and qqline with the points from rnorm.

By just seeing the five normal probability plots, I concluded that it is decent to say that data of the male and female body temperatures is close enough to be normal. However, the two side (both left and right) tails are little bit skewed (away from qqline). Also, I will conclude this data is heavy-tailed- Reference website is below: (<http://stats.stackexchange.com/questions/101274/how-to-interpret-a-qq-plot>).

I explained the formula and explanations in more detail on separate paper.

A -b)

Get male temperature only

```

temp_mean2 <- mean(bodydata$temperature[1:65]) #Get mean of temperature
temp_sd2 <- sd(bodydata$temperature[1:65]) #Get standard deviation of temeperature

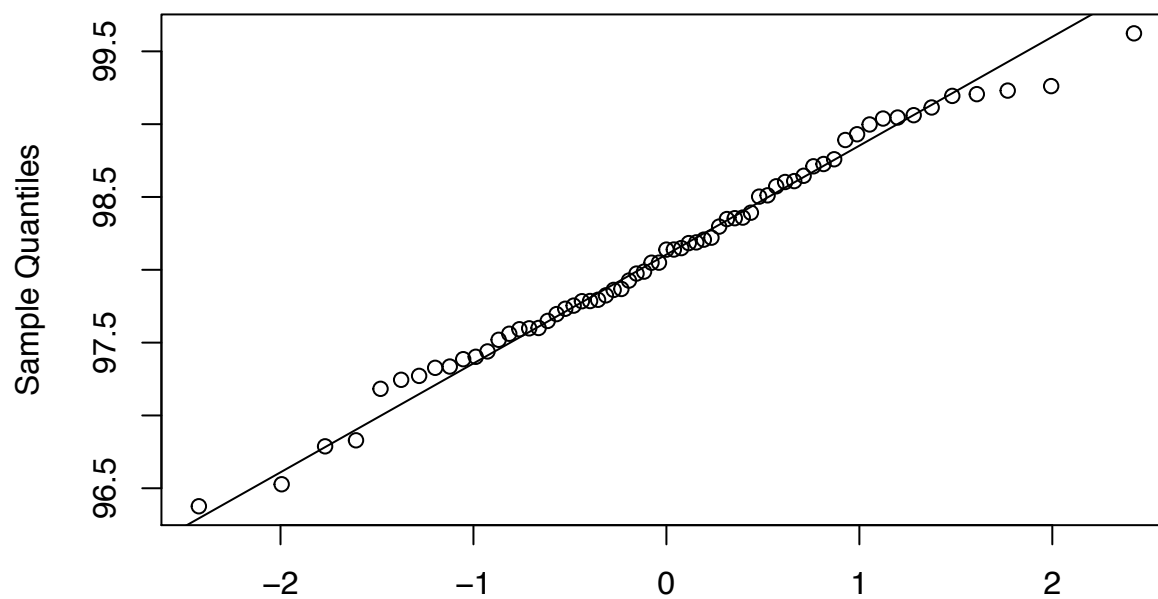
simulation_provider1b <- function(){
  mean_simulation1b <- rnorm(n = length(bodydata$temperature[1:65]), mean = temp_mean2, sd = temp_sd2)

  qqnorm(sort(mean_simulation1b), main = "Temparature Normality of male")
  qqline(mean_simulation1b)
}

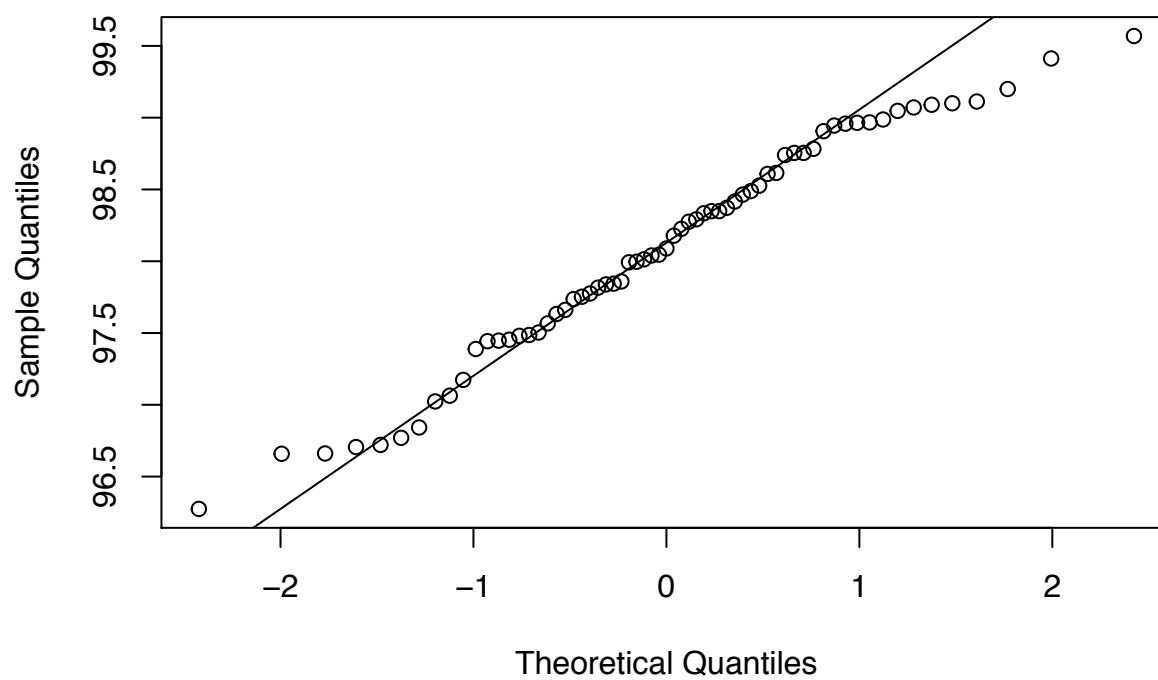
#Simulate samples 5 times
for (i in 1:5){
  simulation_provider1b()
}

```

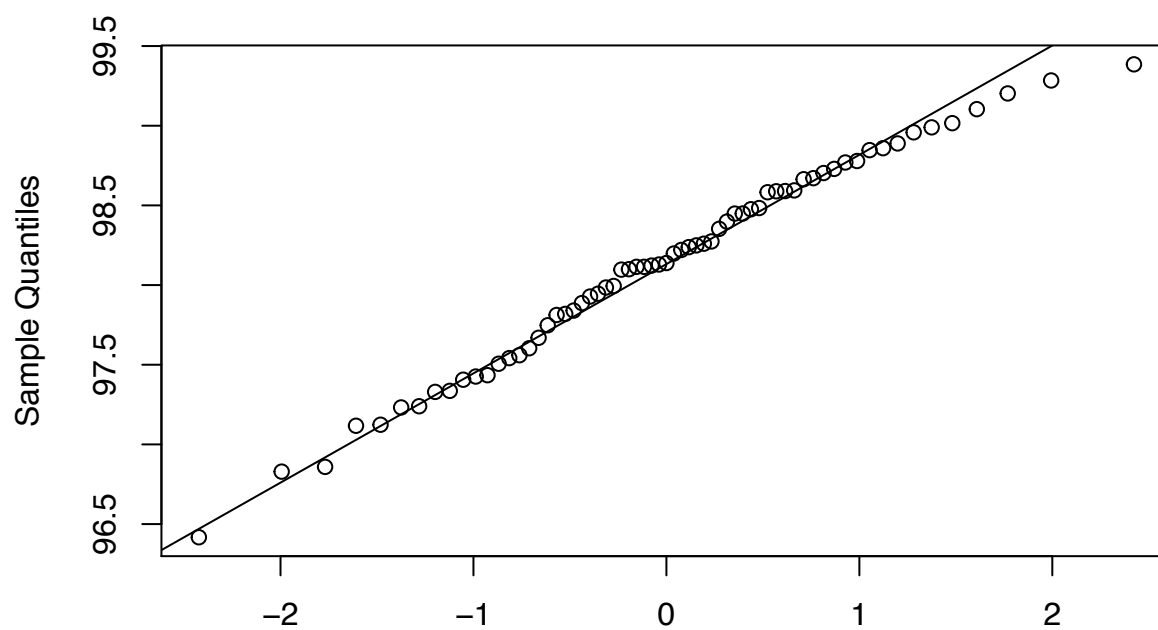
Temperature Normality of male



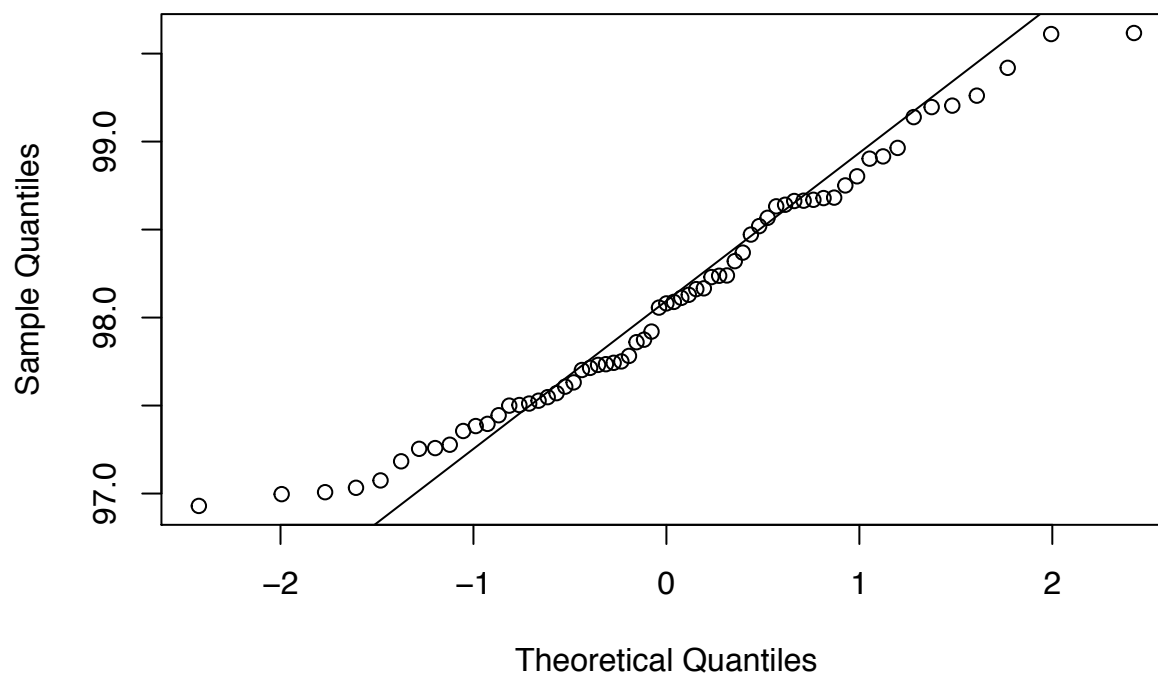
Temperature Normality of male



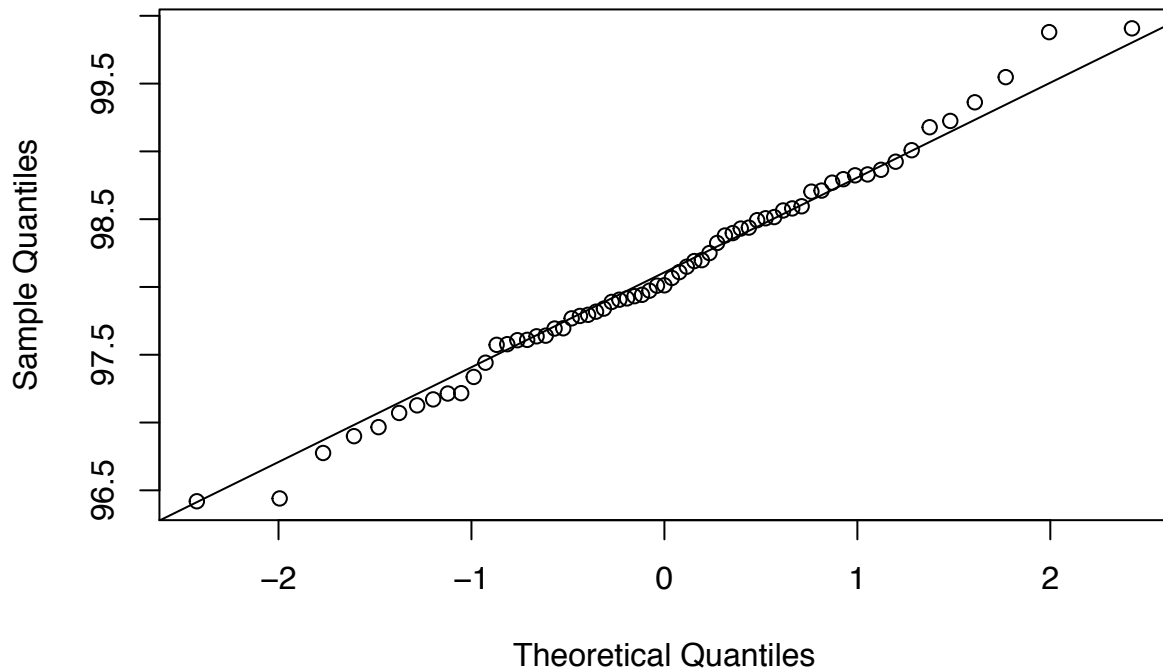
Temperature Normality of male



Temperature Normality of male



Temperature Normality of male



See more variability than gender combined, but still reasonable enough to prove normality (slight skeness on both tails)

A -c)

Get female temperature only

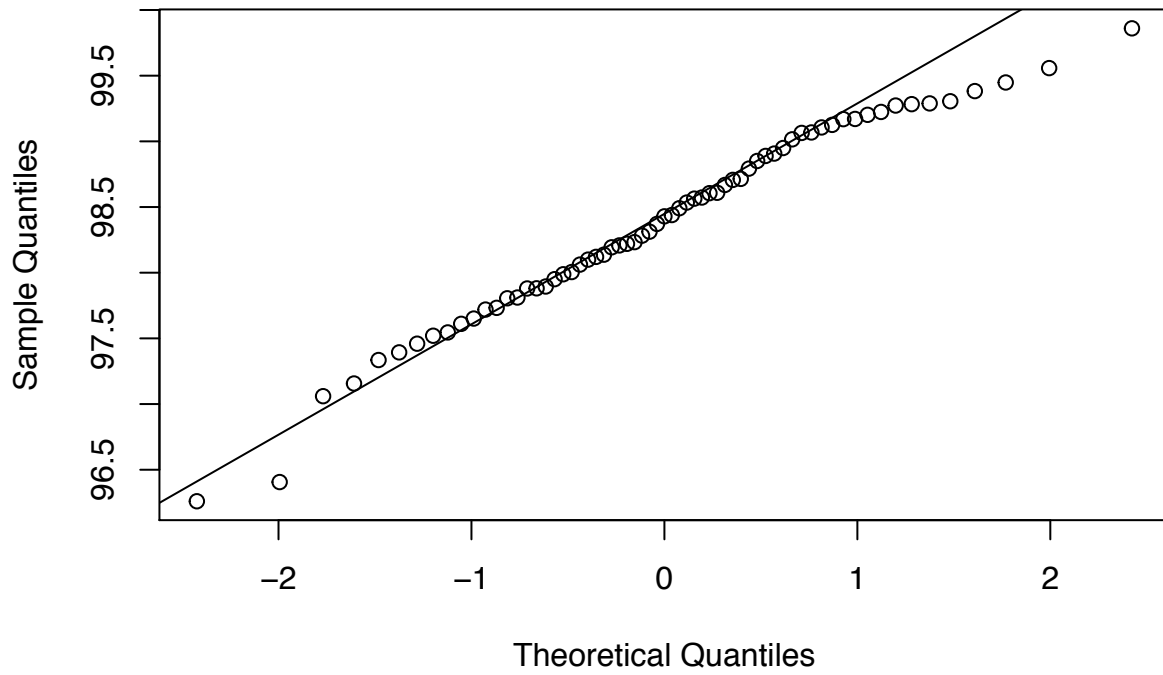
```
temp_mean3 <- mean(bodydata$temperature[66:130]) #Get mean of temperature
temp_sd3 <- sd(bodydata$temperature[66:130]) #Get standard deviation of temeperature

simulation_provider1c <- function(){
  mean_simulation1c <- rnorm(n = length(bodydata$temperature[66:130]), mean = temp_mean3,
                           sd = temp_sd3)

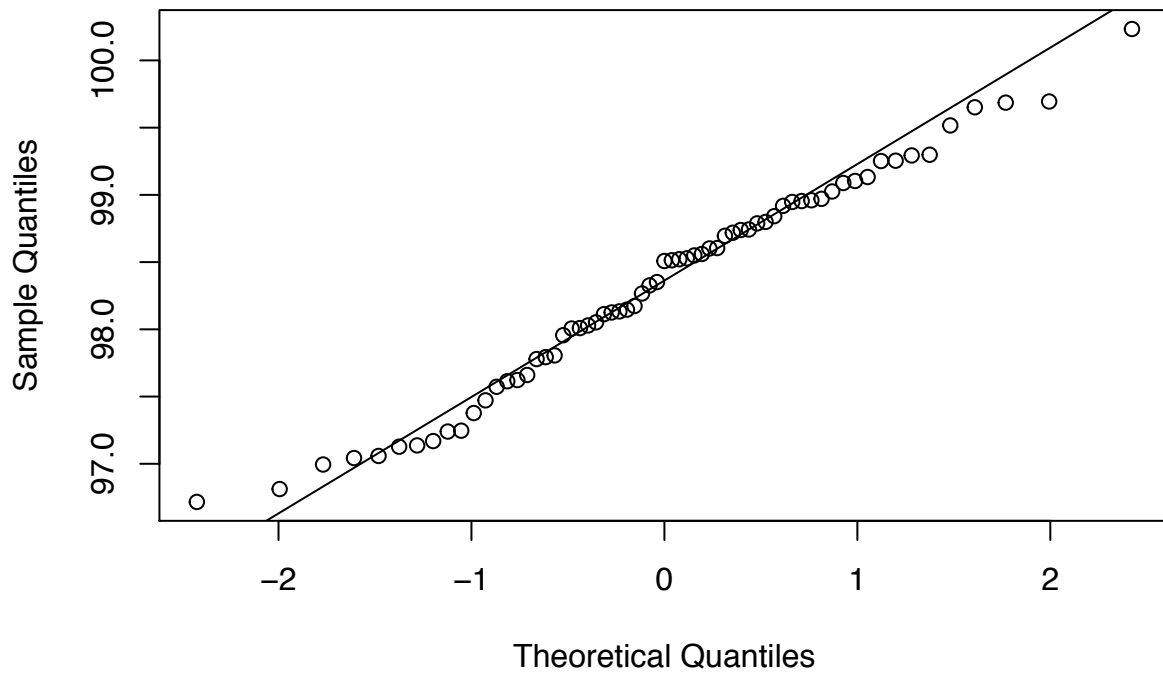
  qqnorm(sort(mean_simulation1c), main = "Temperature Normality of female")
  qqline(mean_simulation1c)
}

#Simulate samples 5 times
for (i in 1:5){
  simulation_provider1c()
}
```

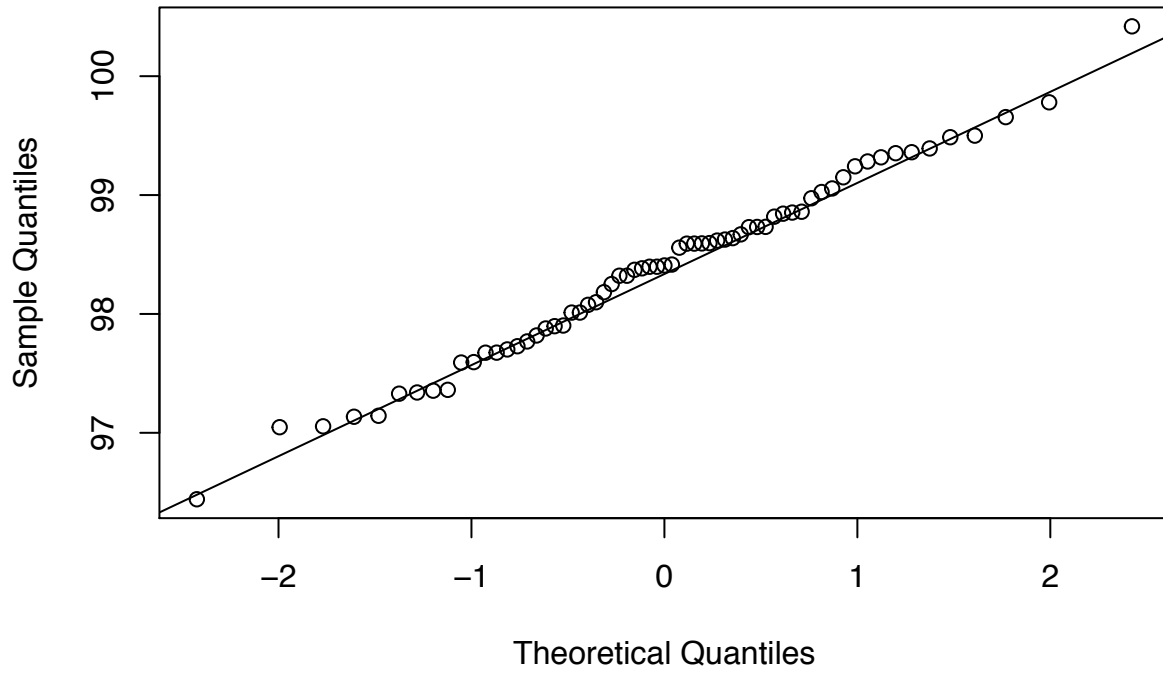
Temperature Normality of female



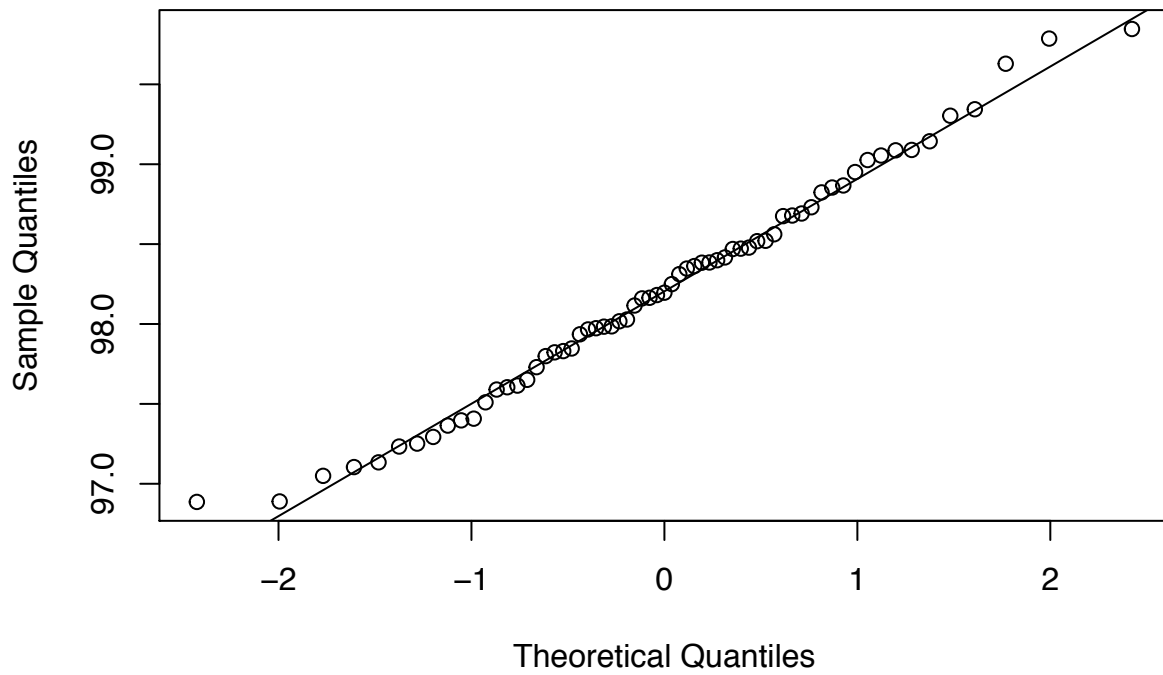
Temperature Normality of female



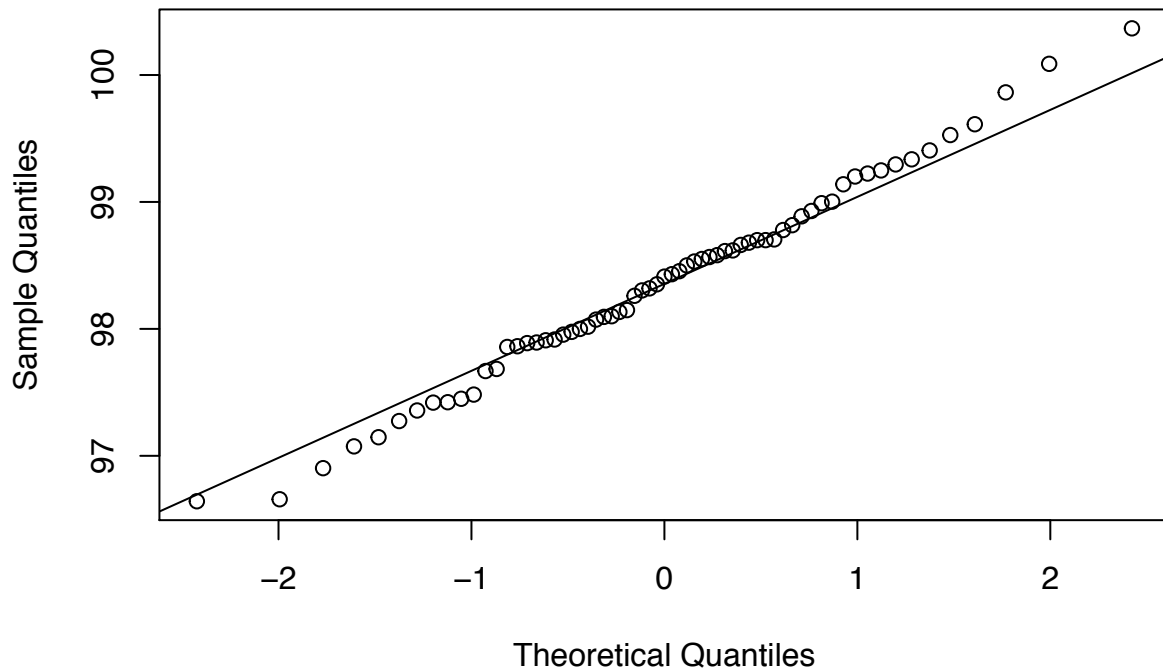
Temperature Normality of female



Temperature Normality of female



Temperature Normality of female



See more variability than gender combined, but still reasonable enough to prove normality (slight skeness on both tails)

B

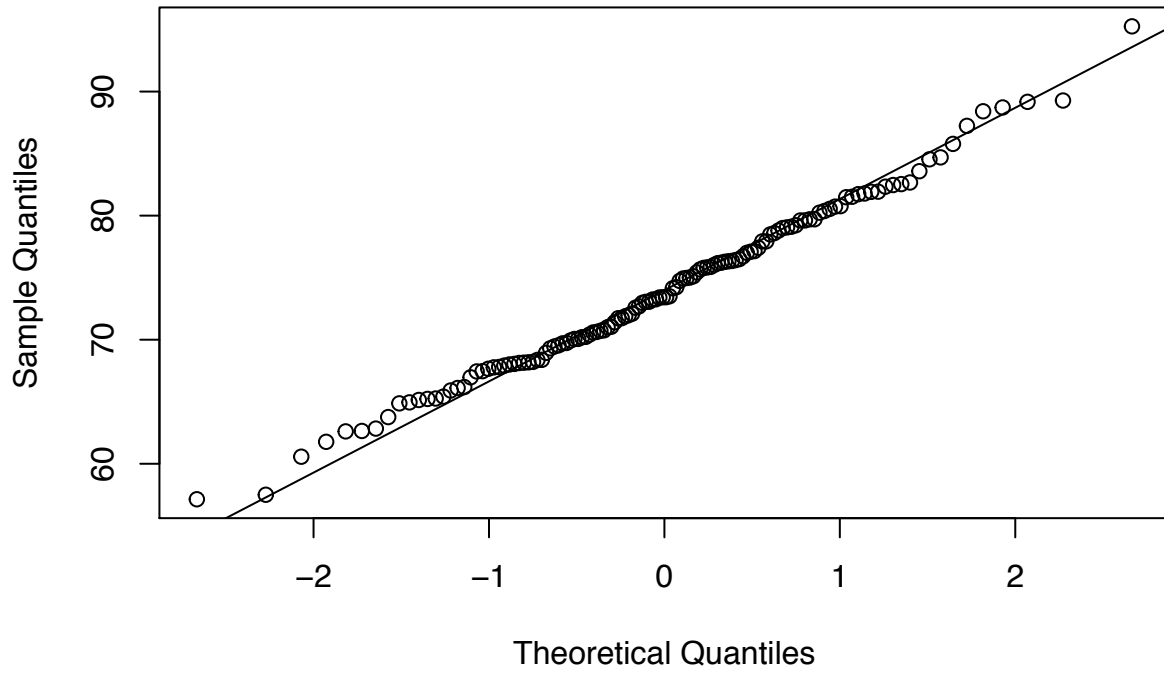
```
rate_mean <- mean(bodydata$rate) #Get mean of heart rate
rate_sd <- sd(bodydata$rate) #Get standard deviation of heart rate

simulation_provider2 <- function(){
  mean_simulation2 <- rnorm(n = length(bodydata$rate), mean = rate_mean, sd = rate_sd)

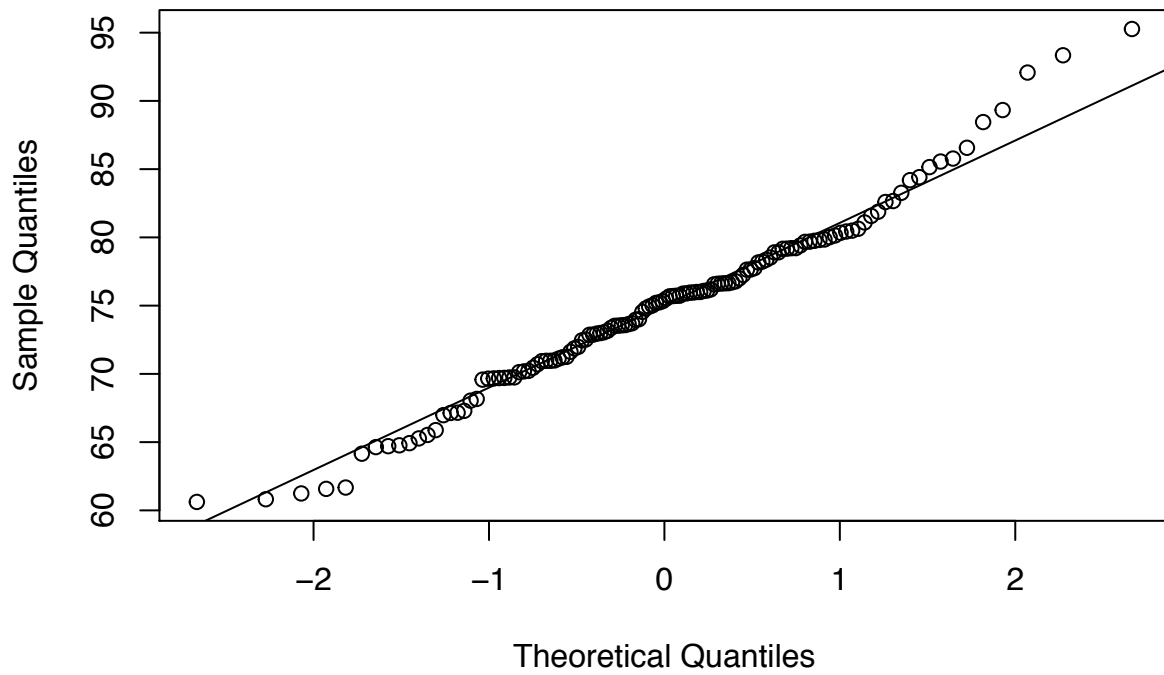
  qqnorm(sort(mean_simulation2), main = "Rate Normality of male and female")
  qqline(mean_simulation2)
}

#Simulate samples 5 times
for (i in 1:5){
  simulation_provider2()
}
```

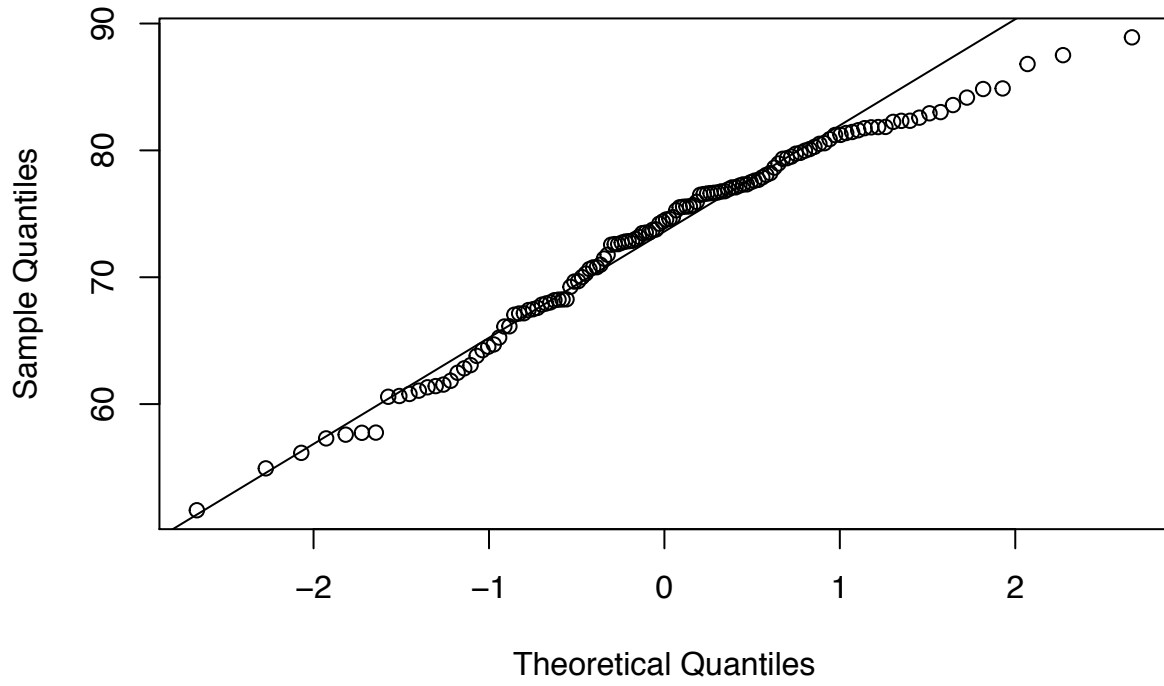

Rate Normality of male and female



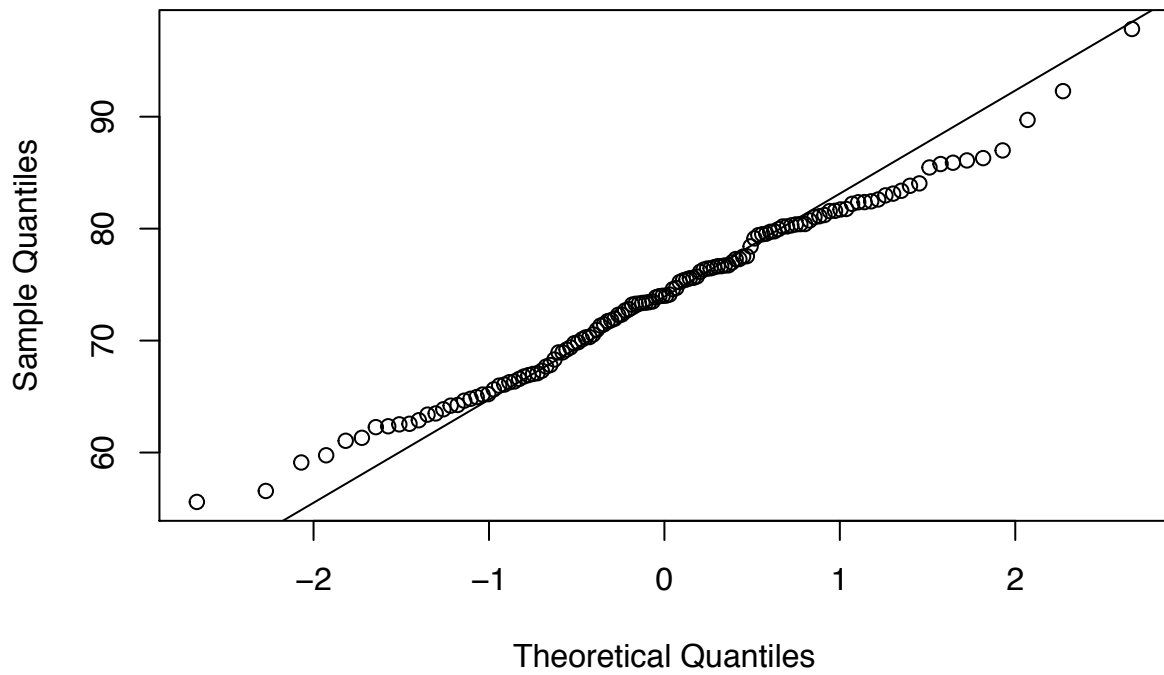
Rate Normality of male and female



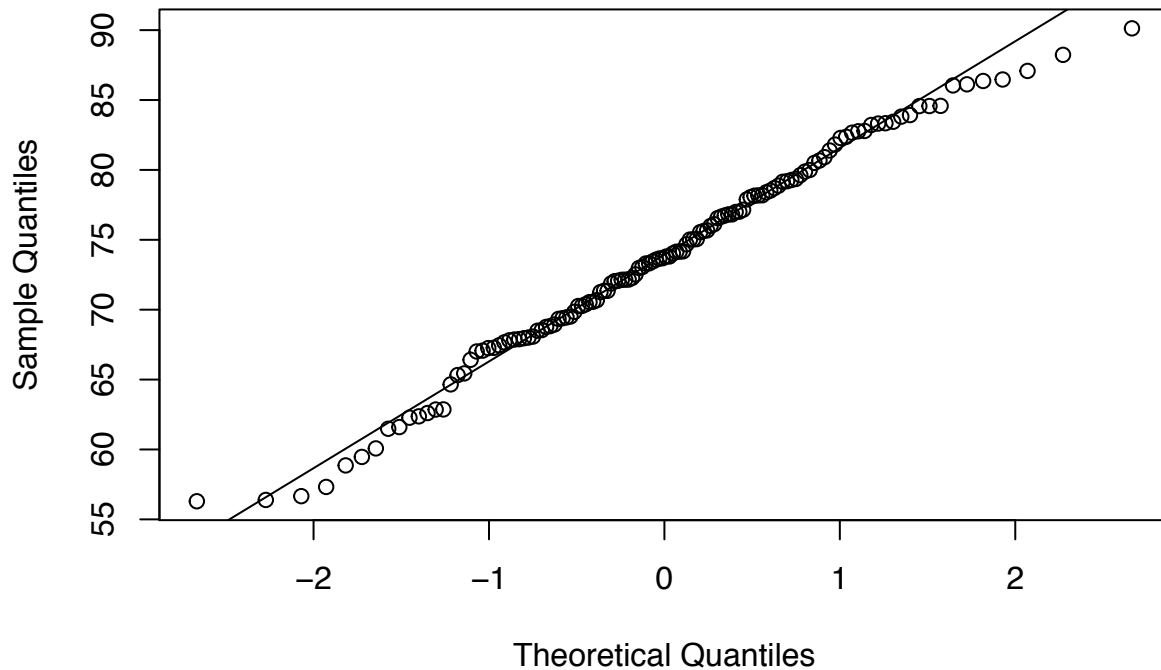
Rate Normality of male and female



Rate Normality of male and female



Rate Normality of male and female



I took the same procedure as in #A.

By just seeing the five normal probability plots, I concluded that it is decent to say that data of the male and female body temperatures is closed enough to be normal. However, the two side (both left and right) tails are little bit skewed (away from qqline). Also, I will conclude this data is heavy-tailed- Reference website is below: (<http://stats.stackexchange.com/questions/101274/how-to-interpret-a-qq-plot>).

B -b)

Get male rate only

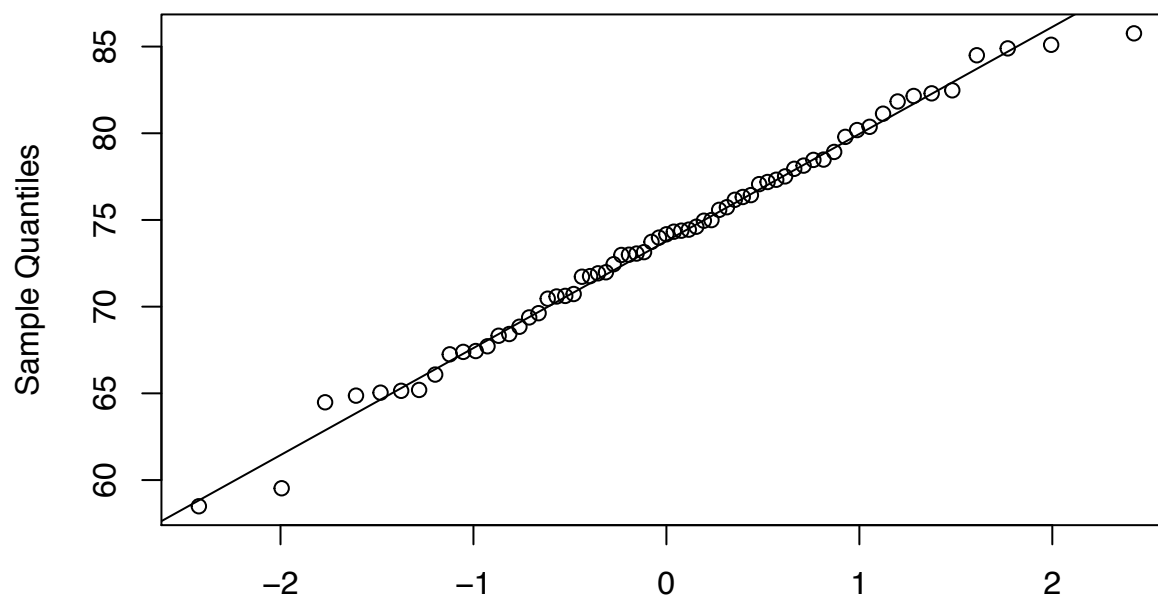
```
rate_mean2 <- mean(bodydata$rate[1:65]) #Get mean of heart rate
rate_sd2 <- sd(bodydata$rate[1:65]) #Get standard deviation of heart rate

simulation_provider2b <- function(){
  mean_simulation2b <- rnorm(n = length(bodydata$rate[1:65]), mean = rate_mean2, sd = rate_sd2)

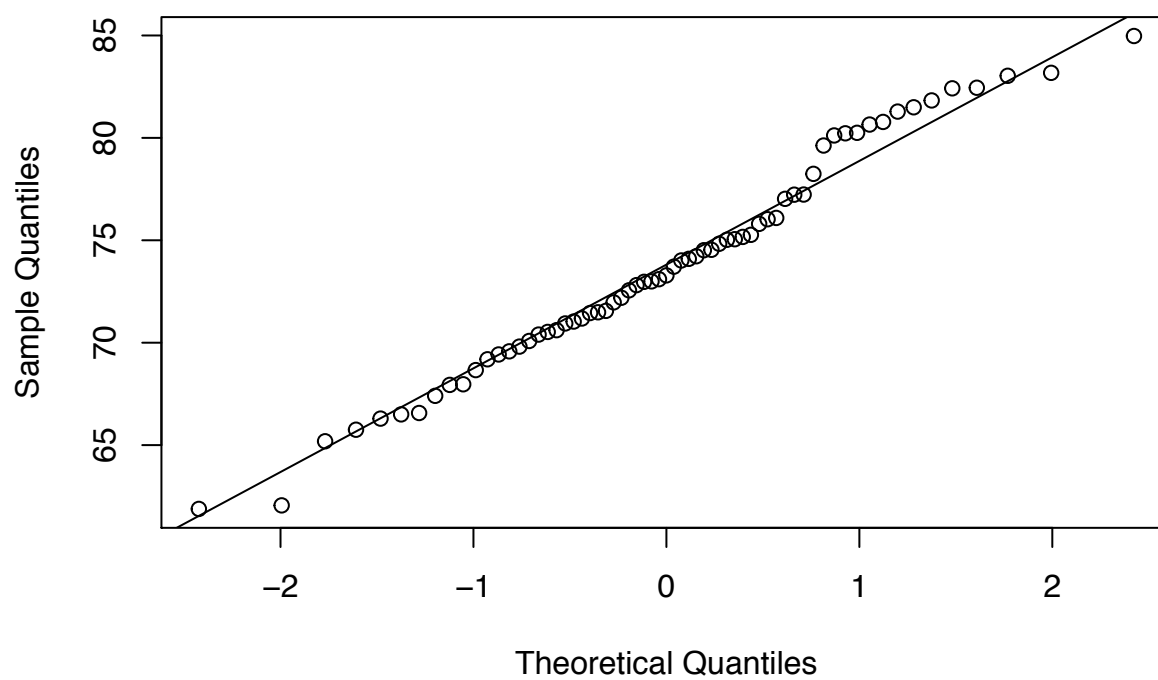
  qqnorm(sort(mean_simulation2b), main = "Rate Normality of male")
  qqline(mean_simulation2b)
}

#Simulate samples 5 times
for (i in 1:5){
  simulation_provider2b()
}
```

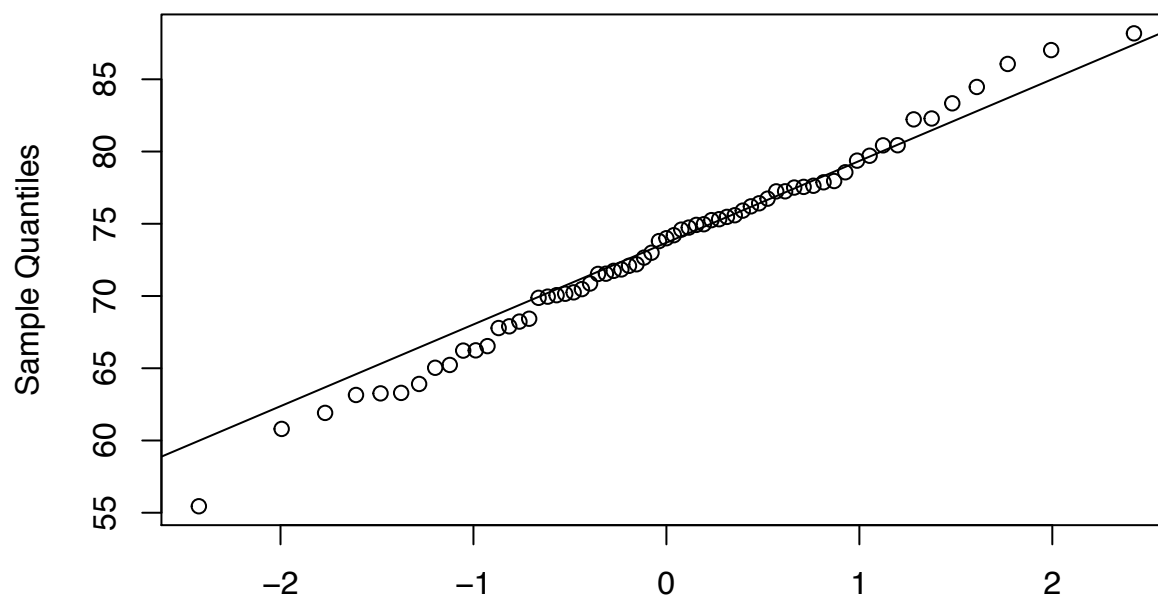
Rate Normality of male



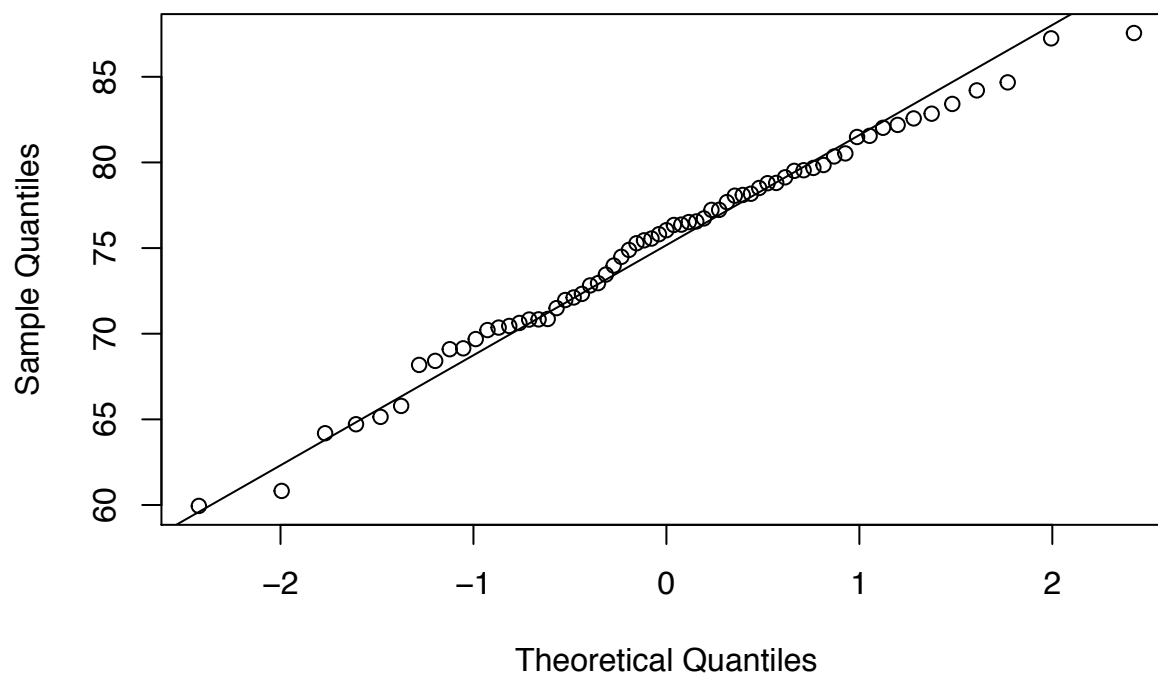
Rate Normality of male



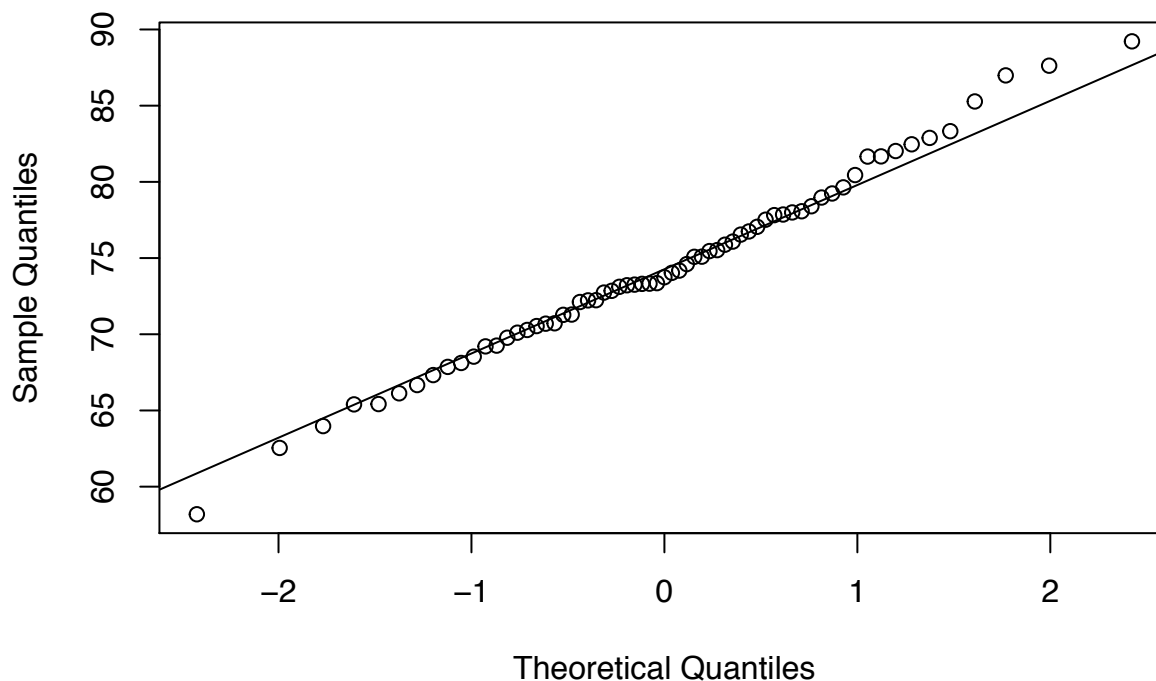
Rate Normality of male



Rate Normality of male



Rate Normality of male



See more variability than gender combined, but still reasonable enough to prove normality (slight skeness on both tails - more skewed than the temeprature one)

B -c)

Get female rate only

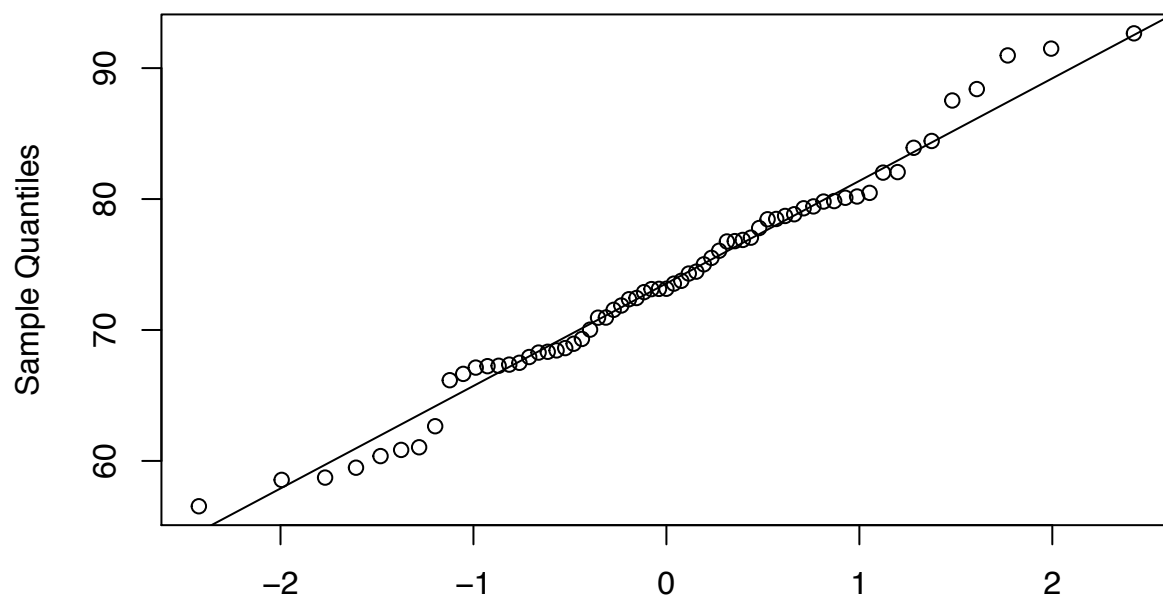
```
rate_mean3 <- mean(bodydata$rate[66:130]) #Get mean of heart rate
rate_sd3 <- sd(bodydata$rate[66:130]) #Get standard deviation of heart rate

simulation_provider2c <- function(){
  mean_simulation2c <- rnorm(n = length(bodydata$rate[66:130]), mean = rate_mean3, sd = rate_sd3)

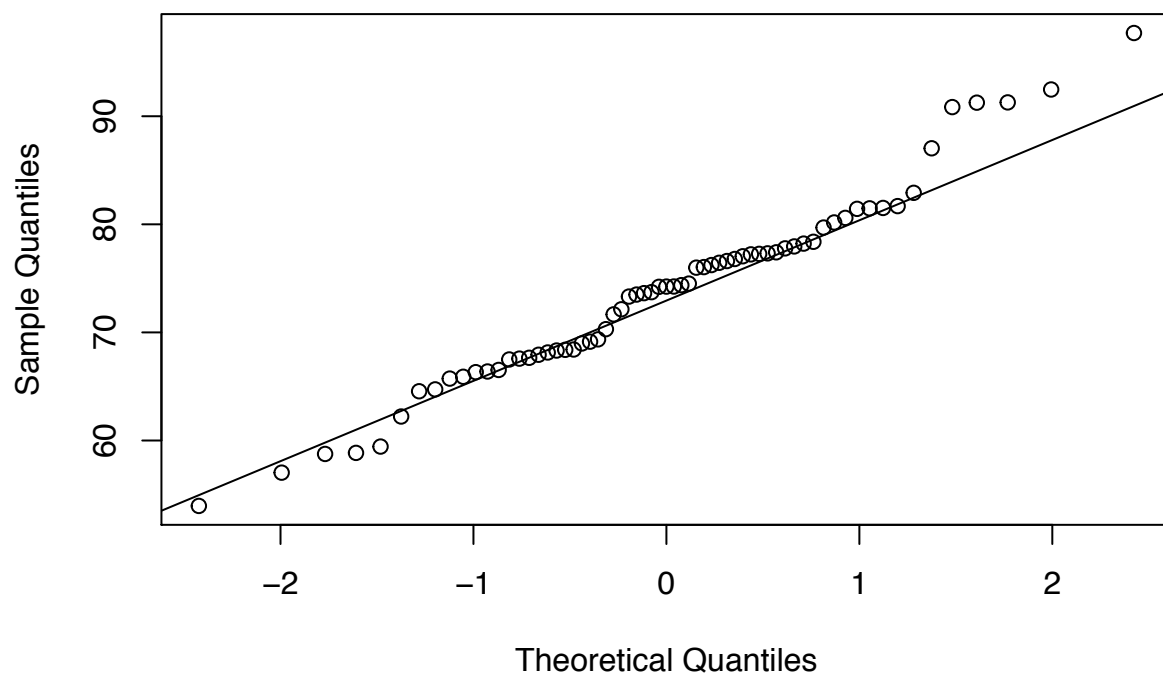
  qqnorm(sort(mean_simulation2c), main = "Rate Normality of female")
  qqline(mean_simulation2c)
}

#Simulate samples 5 times
for (i in 1:5){
  simulation_provider2c()
}
```

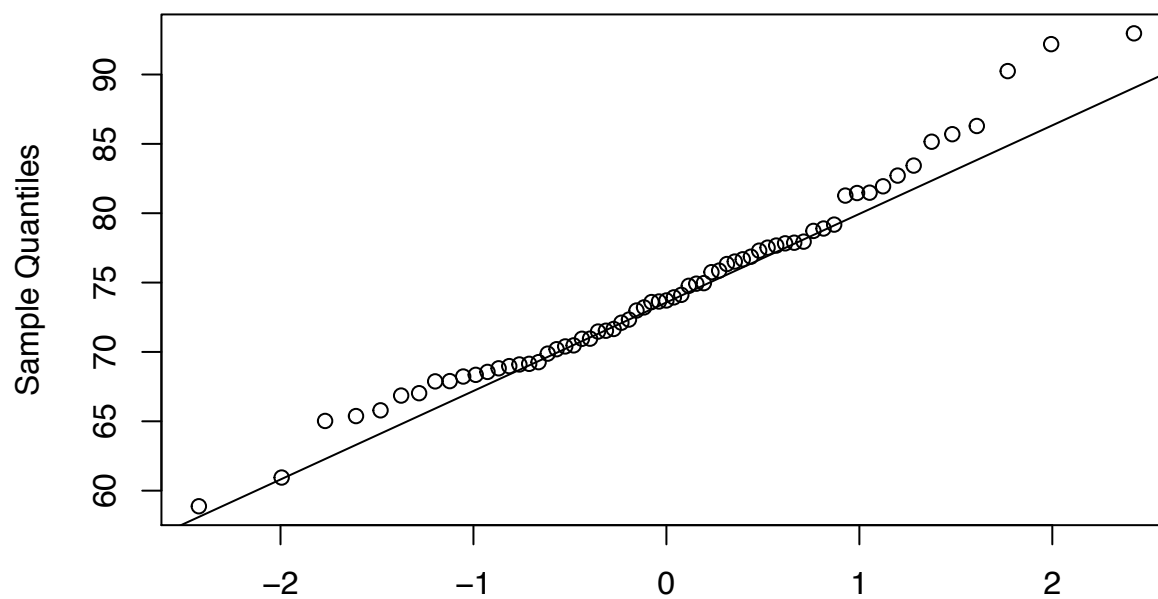
Rate Normality of female



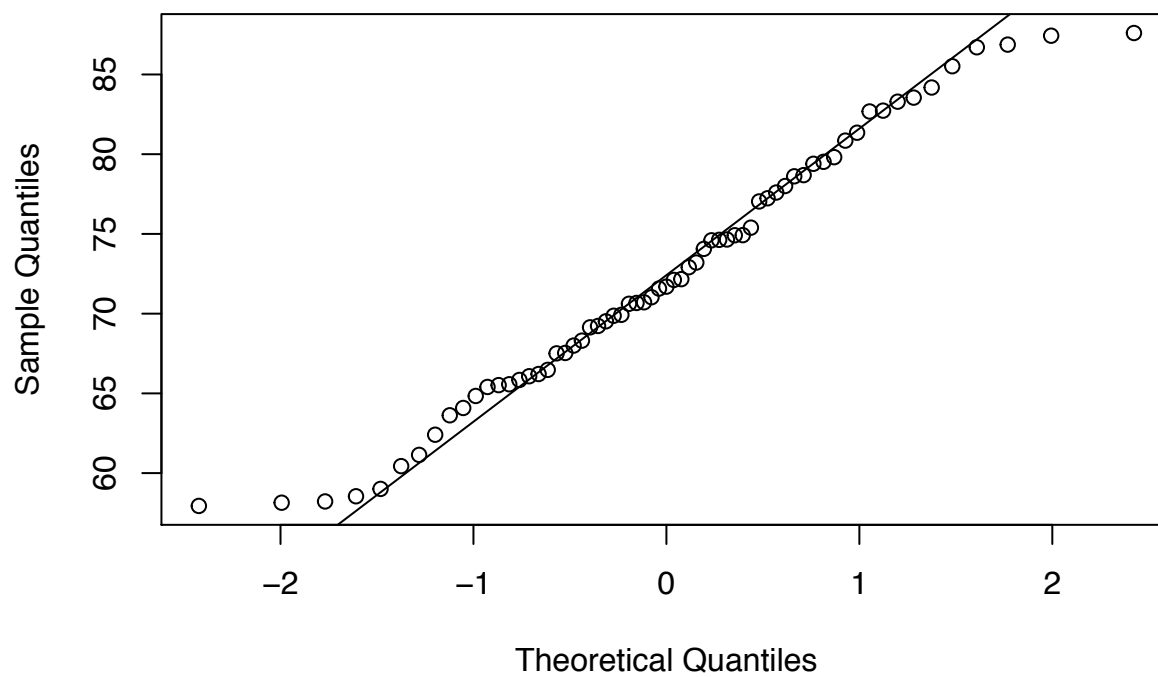
Rate Normality of female



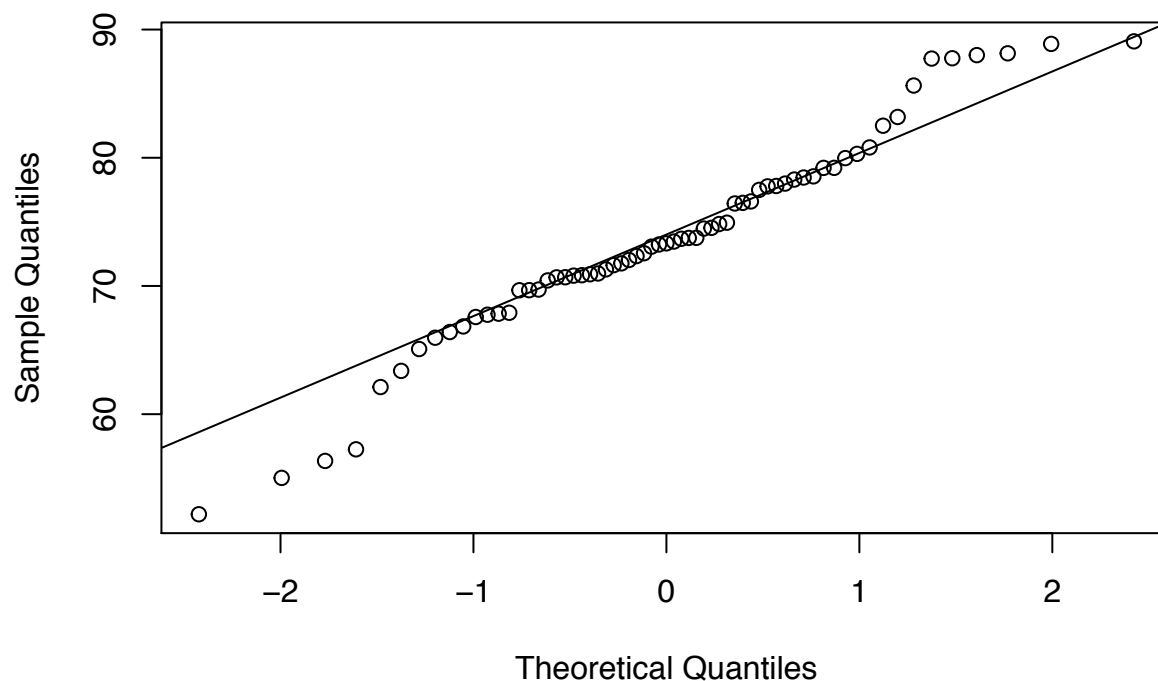
Rate Normality of female



Rate Normality of female



Rate Normality of female



See more variability than gender combined, but still reasonable enough to prove normality (slight skeness on both tails)

C -a)

T-test is $\frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$. “t.test” is a great function embedded in R to get t-test of our data.

male body temeperature

```
t.test(x=bodydata$temperature[1:65], alternative = "two.sided", mu = 98.6, paired=F, var.equal = T,
       conf.level = 0.95)
```

```
##
## One Sample t-test
##
## data: bodydata$temperature[1:65]
## t = -5.715757, df = 64, p-value = 3.08384e-07
## alternative hypothesis: true mean is not equal to 98.6
## 95 percent confidence interval:
## 97.9314722 98.2777586
## sample estimates:
## mean of x
## 98.1046154
```

As we knew the data is the approximately normal, we try to use two-sided t-test for our statistics, and also

get the 95% confidence interval of the data. Since the data is pretty big, it is almost closed to the value from z-test. So, $t \approx -5.7$ and the p value is super small ($\approx 3.08384e-07$), so, p-value is nearly zero. Thus, we reject the null hypothesis. And, as I can see from the 95% confidence interval, 98.6 is not in the C.I.

C -b)

female body temperature

```
t.test(x=bodydata$temperature[66:130], alternative = "two.sided", mu = 98.6, paired=F, var.equal = T,
       conf.level = 0.95)
```

```
##
## One Sample t-test
##
## data: bodydata$temperature[66:130]
## t = -2.235498, df = 64, p-value = 0.0288805
## alternative hypothesis: true mean is not equal to 98.6
## 95 percent confidence interval:
## 98.2096189 98.5780734
## sample estimates:
## mean of x
## 98.3938462
```

As we knew the data is the approximately normal, we try to use two-sided t-test for our statistics, and also get the 95% confidence interval of the data. Since the data is pretty big, it is almost closed to the value from z-test. So, $t \approx -2.24$ and the p value is ≈ 0.029 , we reject the null hypothesis at $\alpha = 0.05$, but do not reject the null hypothesis at $\alpha = 0.01$.