Unit 5

Nikolaos Kornilakis Rodrigo Viale Jakub Trnan Aleksandra Daneva
Luis Diego Pena Monge

2024-04-24

Exercise 85

a)

For a binomial random variable X with n trials and probability p of success, the likelihood function under a particular value of p is given by:

$$L(p) = P(X = x|p) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ with log-likelihood}$$

$$LL(p) = \ln \binom{n}{x} + x\ln(p) + (n-x)\ln(1-p), \text{ maximized at}$$

$$\frac{\partial}{\partial p} LL(p) = \frac{x}{p} - \frac{n-x}{1-p} = \frac{x-np}{p(1-p)} = 0 \implies \hat{p} = \frac{x}{n}$$

For a two-sided test where the null hypothesis $H_0 : p = 0.5$ and the alternative hypothesis $H_A : p \neq 0.5$, the generalized likelihood ratio test statistic is:

$$\Lambda = \frac{L(p = 0.5)}{L(\hat{p})}$$

where $L(\hat{p})$ is the likelihood function evaluated at the maximum likelihood estimator $\hat{p} = \frac{X}{n}$.

Since the binomial coefficient $\binom{n}{x}$ is constant for a given x and n, it cancels out when we take the ratio. Thus, the test statistic simplifies to:

$$\Lambda = \frac{0.5^x (1-0.5)^{n-x}}{\hat{p}^x (1-\hat{p})^{n-x}} = \frac{0.5^n}{\left(\frac{X}{n}\right)^X \left(1 - \frac{X}{n}\right)^{n-X}}$$

For $X = x$ and $\hat{p} = x/n$, this becomes:

$$\Lambda = \frac{0.5^n}{\left(\frac{x}{n}\right)^x \left(1 - \frac{x}{n}\right)^{n-x}}$$

b)

The test statistic for a two-sided binomial test is $|X - n/2|$.

For large values of $|X - n/2|$, the probability of success under the null hypothesis $H_0 : p = 0.5$ is low, indicating that it is more likely that p deviates from 0.5, which leads to rejection of H_0.

The binomial distribution under H_0 is symmetric about $n/2$, so the rejection region for the two-sided test at a significance level α would include the extreme tails of the distribution:

$$P(|X - n/2| > c) = \alpha$$

To find the critical value c, we look for the value such that the probability of X falling more than c away from $n/2$ is α, the significance level of the test:

$$P(X < n/2 - c) + P(X > n/2 + c) = \alpha$$

Since the distribution is symmetric, this is equivalent to:

$$2P(X > n/2 + c) = \alpha$$

This allows for determining the critical value c for a given significance level α, thus defining the rejection region for the test.

c)

The binomial random variable X has a probability mass function given by

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Under the null hypothesis $H_0 : p = 0.5$, this simplifies to

$$P(X = k) = \binom{n}{k} (0.5)^n$$

The test rejects the null hypothesis if $|X - n/2| > c$, which corresponds to the two-tailed rejection regions $X < n/2 - c$ and $X > n/2 + c$. To find the significance level α of this test, we need to find the probability of X falling in either tail, hence:

$$\begin{aligned}
\alpha &= P(|X - n/2| > c) \\
&= P(X < n/2 - c) + P(X > n/2 + c) \\
&= P(X < n/2 - c) + 1 - P(X \leq n/2 + c) \\
&= 1 + P(X < n/2 - c) - P(X \leq n/2 + c)
\end{aligned}$$

d)

Given that X is binomially distributed with $n = 10$ and $p = 0.5$, the probabilities can be calculated using the binomial probability mass function:

$$P(X = k) = \binom{10}{k} \left(\frac{1}{2}\right)^{10}$$

The rejection region for our test is $|X - 5| > 2$, which translates to $X < 3$ or $X > 7$ for $n = 10$. The significance level α is the sum of the probabilities of these two events:

$$\begin{aligned}
\alpha &= P(|X - 5| > 2) \\
&= P(X < 3) + P(X > 7) \\
&= \sum_{k=0}^{2} \binom{10}{k} \left(\frac{1}{2}\right)^{10} + \sum_{k=8}^{10} \binom{10}{k} \left(\frac{1}{2}\right)^{10} \\
&= 0.109375
\end{aligned}$$

Exercise 86

a)

True. The generalized likelihood ratio is defined as:

$$\Lambda = \frac{\text{Likelihood under } H_0}{\text{Likelihood under } H_A}$$

By definition, the likelihood under H_0 cannot be greater than the unrestricted maximum likelihood because the maximum likelihood estimate is the value that maximizes the likelihood function given the data, without any constraint imposed by the null hypothesis.

Therefore, Λ ranges from 0 to 1. If the likelihoods under both hypotheses are the same, Λ would be exactly 1. In all other cases, where the likelihood under the alternative hypothesis is greater, Λ would be less than 1. Hence, Λ never exceeds 1.

b)

False.

The p-value is the smallest level of significance at which the null hypothesis can be rejected. If a p-value is .03, the null hypothesis can be rejected at any significance level greater than .03. The p-value must be less than or equal to the significance level for the null hypothesis to be rejected.

c)

True.

The p-value is the smallest level of significance at which the null hypothesis can be rejected. If the test rejects at a significance level 0.06, then by definition of the p-value, it is less or equal than 0.06.

d)

False.

The p-value is the probability of observing test results at least as extreme as the ones observed during the test, assuming that the null hypothesis is true. It is not the probability that the null hypothesis is correct. The p-value quantifies how well the observed data support the null hypothesis, but it does not provide a direct probability of the null hypothesis itself.

e)

False.

The p-value, is the probability that the observed data (or more extreme) would occur if the null hypothesis were true. It is not equivalent to the likelihood ratio. The likelihood ratio itself does not provide the probability of the observed data under the null hypothesis; it is a measure of relative likelihood between two hypotheses.

f)

Testing the statement in R:

```r
pchisq(8.5, 4, lower.tail = FALSE)
```

```
## [1] 0.07488723
```

The p-value for the described test statistic is $0.07488723 > 0.05$, hence, the statement is false.

Exercise 87

a)

If the test rejects for large $|T|$, it implies a two sided test, as T can be large or small (away from 0). Therefore, the p-value will be:

$$P(T \geq 1.5) + P(T \leq -1.5) = P(T \leq -1.5) + P(T \leq -1.5) = 2 \cdot \Phi(-1.5)$$

Implementing this in R:

```r
cat("The p-value will be: ", 2 * pnorm(-1.5))
```

```
## The p-value will be:  0.1336144
```

b)

Since the test rejects for large T, it is a one-sided test, with the p-value given by:

$$P(T \geq 1.5) = 1 - \Phi(1.5)$$

Computing it in R:

```r
cat("The p-value will be: ", 1 - pnorm(1.5))
```

```
## The p-value will be:  0.0668072
```

Exercise 88

In hypothesis testing, a level α test means that making a Type I error, is α. The test statistic T is used to decide whether to reject the null hypothesis.

Because g is monotone-increasing, for any t such that $T > t$, we will have $S = g(T) > g(t)$. Therefore, the event that $T > t_0$ is equivalent to the event that $S > g(t_0)$, since g preserves the order due to its monotonicity.

Since the level α test for T rejects when $T > t_0$, and since this is equivalent to $S > g(t_0)$, the test that rejects when $S > g(t_0)$ will have the same probability of rejecting the null hypothesis when it is true as the original test in terms of T. Thus, the new test based on S will also be a level α test.

Exercise 89

a)

For a given observed value t of the test statistic T, the p-value V is the probability of observing a test statistic as extreme or more extreme than t under the null hypothesis. For a right-tailed test, this is $P(T \geq t)$.

Since $F(t)$ is the probability that T is less than or equal to t, $1 - F(t)$ is the probability that T is greater than t. This means that

$$V = P(T \geq t) = 1 - P(T < t) = 1 - F(t)$$

Therefore, the p-value V for the test statistic T is $1 - F(T)$.

b)

As covered in (a), the p-value V is defined as $V = 1 - F(T)$, where F is the cumulative distribution function (CDF) of the test statistic T under the null hypothesis. We want to show that V is uniformly distributed on the interval $[0, 1]$. Let $v \in [0, 1]$

$$\begin{aligned}
P(V \leq v) &= P(1 - F(T) \leq v) \\
&= P(F(T) \geq 1 - v) \\
&= 1 - P(F(T) \leq 1 - v) \\
&= 1 - (1 - v) \\
&= v
\end{aligned}$$

This is exactly the definition of a uniform distribution on the interval $[0, 1]$.

c)

As we showed in (b) that , we can say that $P(V \geq 0.1) = 1 - P(V \leq 0.1) = 1 - 0.1 = 0.9$.

d)

The significance level α of a test is the probability of rejecting the null hypothesis when it is actually true (Type I error). For a test that rejects when the p-value $V < \alpha$, we can show it has a significance level of α by demonstrating that the probability of such an event occurring under the null hypothesis is exactly α. Since we proved in (b) that $V \sim U(0,1)$, it follows that $P(V \leq \alpha) = \alpha$.

Exercise 90

Let $X_1, ..., X_n$ be independent such that $X_i \sim Poisson(\lambda_i)$. Let's first calculate the MLE under the null hypothesis that $\lambda := \lambda_1 = ... = \lambda_n$:

$$L = \prod_{i=1}^{n} \frac{\lambda^{X_i} e^{-\lambda}}{X_i!} \implies LL = \sum_{i=1}^{n} [X_i ln(\lambda) - \lambda - ln(X_i!)]$$

$$\implies LL' = \sum_{i=1}^{n} \left(\frac{X_i}{\lambda} - 1 \right) = \frac{1}{\lambda} \sum_{i=1}^{n} X_i - n = 0 \iff \lambda = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$$

Then the MLE is given by $\tilde{\lambda} = \bar{X}$. Next, let's compute the Fisher information of the parameter:

$$\begin{aligned} I(\lambda) &= -E \left[\frac{\partial^2}{\partial \lambda^2} ln(f(X|\lambda)) \right] \\ &= -E \left[\frac{\partial^2}{\partial \lambda^2} X ln(\lambda) - \lambda - ln(X!) \right] \\ &= -E \left[\frac{\partial}{\partial \lambda} \left(\frac{X}{\lambda} - 1 \right) \right] \\ &= -E \left[\frac{-X}{\lambda^2} \right] \\ &= \frac{1}{\lambda} \end{aligned}$$

Then we have that

$$I(\tilde{\lambda}) = \frac{1}{\tilde{\lambda}} \implies I^{-1}(\tilde{\lambda}) = \tilde{\lambda} = \bar{X}$$

Now, notice that:

$$\nabla_\theta (LL) = \begin{pmatrix} \frac{X_1}{\lambda_1} - 1 \\ \vdots \\ \frac{X_n}{\lambda_n} - 1 \end{pmatrix} \implies S(\tilde{\lambda}) = \begin{pmatrix} \frac{X_1}{\bar{X}} - 1 \\ \vdots \\ \frac{X_n}{\bar{X}} - 1 \end{pmatrix}$$

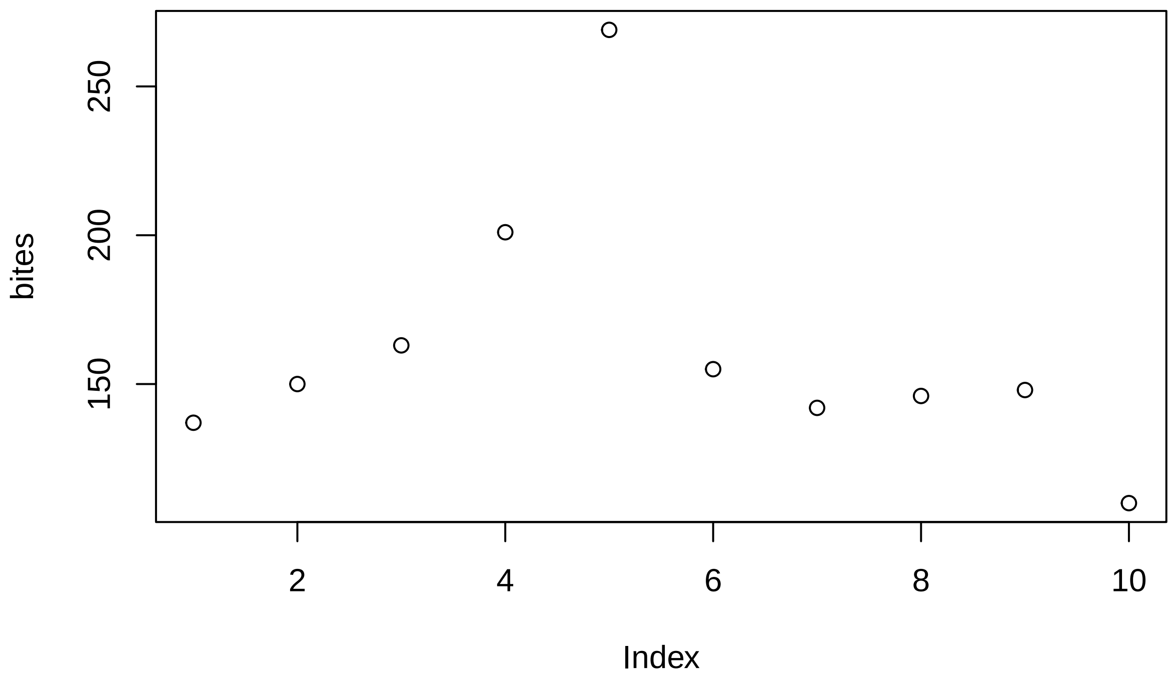
Which means we can compute the test statistic as:

$$
\begin{aligned}
T_S &= S(\tilde{\lambda})' I^{-1}(\tilde{\lambda}) S(\tilde{\lambda}) \\
&= \bar{X} S(\tilde{\lambda})' S(\tilde{\lambda}) \\
&= \bar{X} \sum_{i=1}^{n} \left(\frac{X_i}{\bar{X}} - 1 \right)^2 \\
&= \bar{X} \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\bar{X}} \right)^2 \\
&= \sum_{i=1}^{n} \frac{\left(X_i - \bar{X}\right)^2}{\bar{X}}
\end{aligned}
$$

Exercise 91

We perform a chi-squared test for equal versos distinct probabilities along the specified dates:

```r
bites <- c(137,
           150,
           163,
           201,
           269,
           155,
           142,
           146,
           148,
           110)

plot(bites)
```



```
n <- sum(bites)
expected <- c(rep(3/29, 9), 2/29)
observed <- bites/sum(bites)

chisq.statistic <- sum((bites - n*expected)^2/(n*expected))

chisq.statistic
```

```
## [1] 85.47975
```

```
1 - pchisq(chisq.statistic, df = 9)
```

```
## [1] 1.310063e-14
```

We observe that we very strongly reject the null hypothesis of homogeneous rate of bites, which does imply the presence of a temporal trend in the incidence. As seen in the plot, the peak occurs in the group that contains days 28, 29 and 1. Recall 29 is the full moon.

Exercise 92

Notice that:

$$
\begin{aligned}
X^2 &= \frac{(X_1 - np_1)^2}{np_1} - \frac{(X_2 - np_2)^2}{np_2} \\
&= \frac{(X_1 - np_1)^2}{np_1} - \frac{(n - X_1 - n(1 - p1))^2}{n(1 - p_1)} \\
&= \frac{(X_1 - np_1)^2}{np_1} - \frac{(np_1 - X_1)^2}{n(1 - p_1)} \\
&= \frac{(X_1 - np_1)^2}{n} \left(\frac{1}{p_1} - \frac{1}{1 - p_1} \right) \\
&= \frac{(X_1 - np_1)^2}{np_1(1 - p_1)}
\end{aligned}
$$

Now, under the null hypothesis X_1 follows a binomial distribution with parameters n, p_1, which means that $E[X_1|H_0] = np_1$ and $Var[X_1|H_0] = np_1(1 - p_1)$. By the central limit theorem, we have that asymptotically:

$$\frac{X_1 - np_1}{\sqrt{np_1(1 - p_1)}} \sim N(0, 1)$$

This means that we can approximate $\frac{X_1 - np_1}{\sqrt{np_1(1-p_1)}}$ using a standard normal distribution (note that the goodness of this approximation will depend on n), and as we know from class, the square of a standard normal corresponds to a chi-squared distribution with 1 degree of freedom. This means we can approximate $\frac{(X_1 - np_1)^2}{np_1(1-p_1)}$ through a chi-squared distribution with 1 degree of freedom.

Exercise 93

Let's derive a likelihood ratio test. Under the null, we assume that all probabilities are the same. Note that in this case, since X_i are independent and share a probability, then $\sum_{i=1}^m X_i \sim Binomial(\sum_{i=1}^m n_i, p)$. Define $n := \sum_{i=1}^m n_i$. Then the MLE of p under the null is given by:

$$\tilde{p}_0 = \frac{\sum_{i=1}^m X_i}{n}$$

Similarly, under the alternative hypothesis we look not at the sum's MLE as we don't have an explicit distribution for this, but rather at each individual sample's MLE for its respective p_i:

$$\tilde{p}_i = \frac{X_i}{n_i} \implies X_i = n_i \tilde{p}_i$$

Then we can define our LRT statistic as:

$$
\begin{aligned}
\Lambda &= \frac{\tilde{p}_0^{\sum_{i=1}^m X_i} (1 - \tilde{p}_0)^{n - \sum_{i=1}^m X_i}}{\prod_{i=1}^m \tilde{p}_i^{X_i} (1 - \tilde{p}_i)^{n_i - X_i}} \\
&= \frac{\tilde{p}_0^{\sum_{i=1}^m n_i \tilde{p}_i} (1 - \tilde{p}_0)^{\sum_{i=1}^m n_i (1 - \tilde{p}_i)}}{\prod_{i=1}^m \tilde{p}_i^{n_i \tilde{p}_i} (1 - \tilde{p}_i)^{n_i (1 - \tilde{p}_i)}}
\end{aligned}
$$

Then we can follow the same logic used in class:

$$\begin{aligned} ln(\Lambda) &= \left[\sum_{i=1}^{m} \tilde{p}_i n_i ln(\tilde{p}_0) + \sum_{i=1}^{m} \tilde{(}1-p_i) n_i ln(1-\tilde{p}_0)\right] - \left[\sum_{i=1}^{m} \tilde{p}_i n_i ln(\tilde{p}_i) + \sum_{i=1}^{m} \tilde{(}1-p_i) n_i ln(1-\tilde{p}_i)\right] \\ &= \sum_{i=1}^{m} \tilde{p}_i n_i ln\left(\frac{\tilde{p}_0}{\tilde{p}_i}\right) + \sum_{i=1}^{m} (1-\tilde{p}_i) n_i ln\left(\frac{1-\tilde{p}_0}{1-\tilde{p}_i}\right) \end{aligned}$$

Then:

$$-ln(\Lambda) = \sum_{i=1}^{m} \tilde{p}_i n_i ln\left(\frac{\tilde{p}_i}{\tilde{p}_0}\right) + \sum_{i=1}^{m} (1-\tilde{p}_i) n_i ln\left(\frac{1-\tilde{p}_i}{1-\tilde{p}_0}\right)$$

We will now focus on the first sum as the second one's results are completely analogous. Using the Taylor expansion of $xln(x/x_0)$, we have that:

$$\tilde{p}_i ln\left(\frac{\tilde{p}_i}{\tilde{p}_0}\right) \approx (\tilde{p}_i - \tilde{p}_0) + \frac{1}{2\tilde{p}_0}(\tilde{p}_i - \tilde{p}_0)^2$$

Then:

$$\begin{aligned} \sum_{i=1}^{m} \tilde{p}_i n_i ln\left(\frac{\tilde{p}_i}{\tilde{p}_0}\right) &= \sum_{i=1}^{m} n_i \left((\tilde{p}_i - \tilde{p}_0) + \frac{1}{2\tilde{p}_0}(\tilde{p}_i - \tilde{p}_0)^2\right) \\ &\text{The first term cancels as probabilities add up to 1} \\ &= \sum_{i=1}^{m} \frac{(n_i\tilde{p}_i - n_i\tilde{p}_0)^2}{2n_i\tilde{p}_0} \\ &= \sum_{i=1}^{m} \frac{(X_i - n_i\tilde{p}_0)^2}{2n_i\tilde{p}_0} \end{aligned}$$

And through completely analogous logic we get that:

$$\sum_{i=1}^{m} (1-\tilde{p}_i) n_i ln\left(\frac{1-\tilde{p}_i}{1-\tilde{p}_0}\right) = \sum_{i=1}^{m} \frac{(n_i - X_i - n_i(1-\tilde{p}_0))^2}{2n_i(1-\tilde{p}_0)}$$

Combining the terms in the sum of sums and multiplying by 2, we get that:

$$\begin{aligned} -2ln(\Lambda) &= \sum_{i=1}^{m} \frac{(X_i - n_i\tilde{p}_0)^2}{n_i\tilde{p}_0} + \frac{(n_i - X_i - n_i(1-\tilde{p}_0))^2}{n_i(1-\tilde{p}_0)} \\ &\text{Through the same procedure as in exercise 92} \\ &= \sum_{i=1}^{m} \frac{(X_i - n_i\tilde{p}_0)^2}{n_i p_0(1-p_0)} \end{aligned}$$

Now, we know from exercise 92 that each inside term of the sum has a chi-squared distribution with 1 degree of freedom. However, notice that in this case we are adding up m of them, but conditional on fixed margins, if you know the first $m-1$ values of n_i then you know also the m-th, which is also the case with X_i. This means that, under H_0, asymptotically $-2ln(\Lambda)$ has a chi-squared distribution with $m-1$ degrees of freedom.

Exercise 98

```
bodytemp<-read.table(paste(myPath,"bodytemp.txt", sep="/"),header=TRUE,sep=",")

head(bodytemp)
```

```
##   temperature gender rate
## 1        96.3      1   70
## 2        96.7      1   71
## 3        96.9      1   74
## 4        97.0      1   80
## 5        97.1      1   73
## 6        97.1      1   75
```

a)

First we calculate the means and standard deviations by gender:

```
means <- aggregate(bodytemp[,-2],
                   list(gender = bodytemp$gender),
                   mean)

means
```

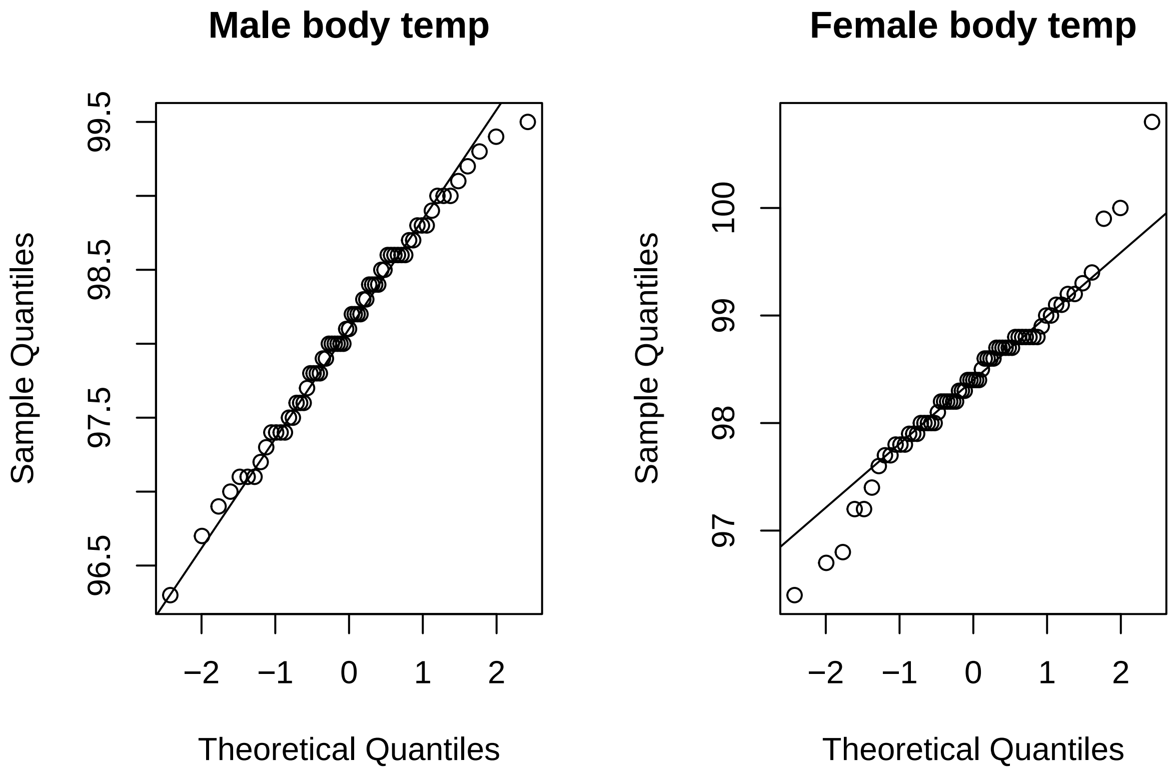
```
##   gender temperature     rate
## 1      1    98.10462 73.36923
## 2      2    98.39385 74.15385
```

```
sds <- aggregate(bodytemp[,-2],
                 list(gender = bodytemp$gender),
                 sd)

sds
```

```
##   gender temperature     rate
## 1      1   0.6987558 5.875184
## 2      2   0.7434878 8.105227
```

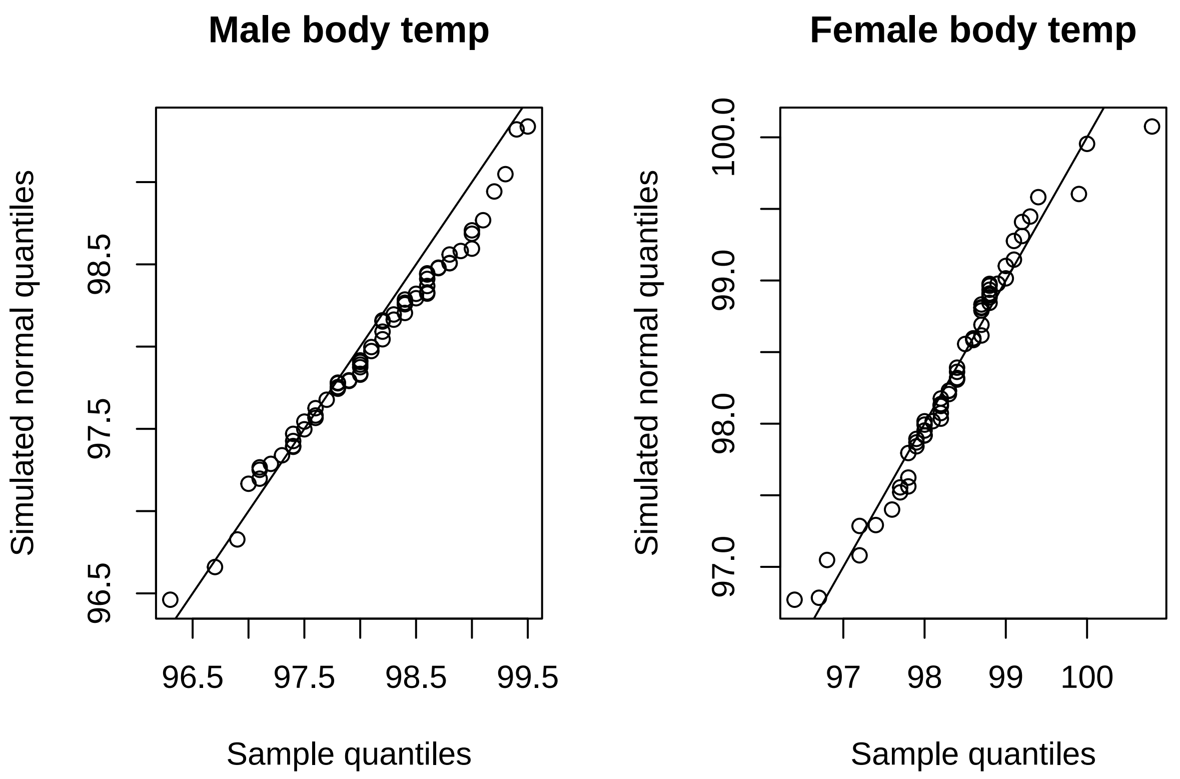
We start by plotting a regular normal quantile plot:

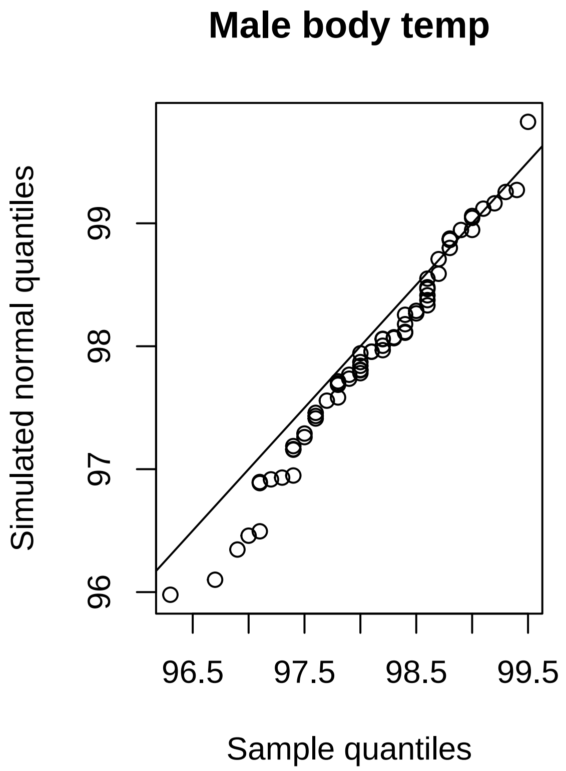
```
par(mfrow = c(1,2))
qqnorm(bodytemp$temperature[bodytemp$gender == 1],
       main = "Male body temp")
qqline(bodytemp$temperature[bodytemp$gender == 1])

qqnorm(bodytemp$temperature[bodytemp$gender == 2],
       main = "Female body temp")
qqline(bodytemp$temperature[bodytemp$gender == 2])
```

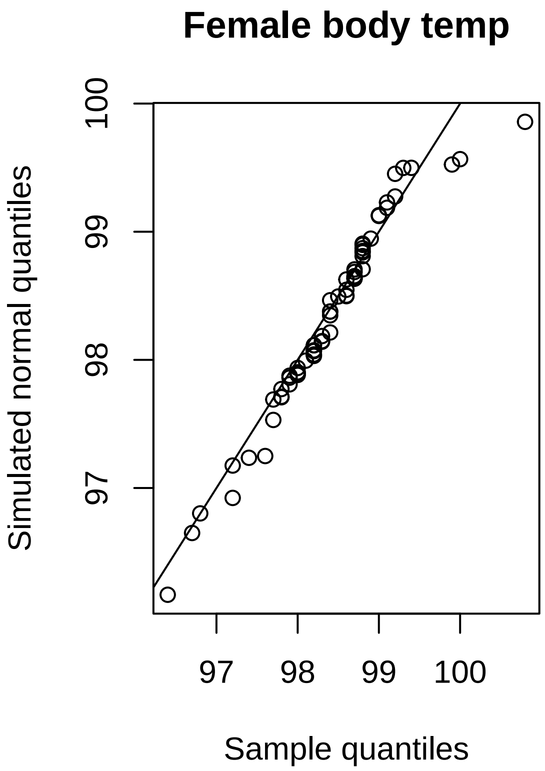


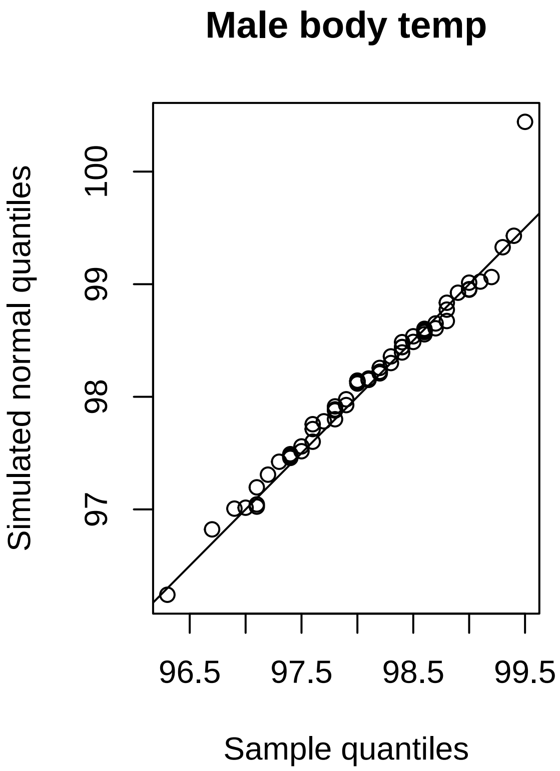
We observe enough cisual evidence to believe that (especially in the male case) the observations follow a normal-ish distribution. In the female case we can observe deviation in the tails. Now we simulate normal samples for both genders and repeat the exercise to see the plot variability:

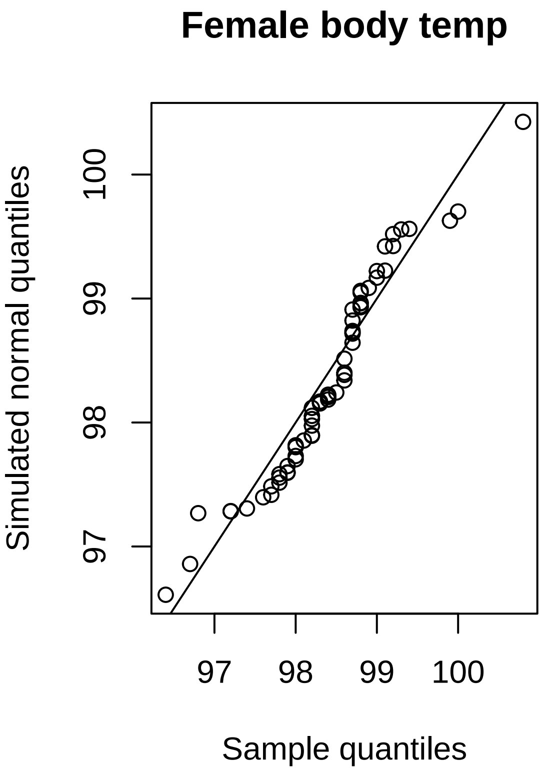
```
for(i in 1:10){
  norm_sim_m <- rnorm(sum(bodytemp$gender == 1),
                      mean = means$temperature[1],
                      sd = sds$temperature[1])

  norm_sim_f <- rnorm(sum(bodytemp$gender == 2),
                      mean = means$temperature[2],
                      sd = sds$temperature[2])
  par(mfrow = c(1,2))
  qqplot(bodytemp$temperature[bodytemp$gender == 1],
         norm_sim_m,
         xlab = "Sample quantiles",
         ylab = "Simulated normal quantiles",
         main = "Male body temp")
  abline(a = 0, b = 1)

  qqplot(bodytemp$temperature[bodytemp$gender == 2],
         norm_sim_f,
         xlab = "Sample quantiles",
         ylab = "Simulated normal quantiles",
         main = "Female body temp")
  abline(a = 0, b = 1)
}
```

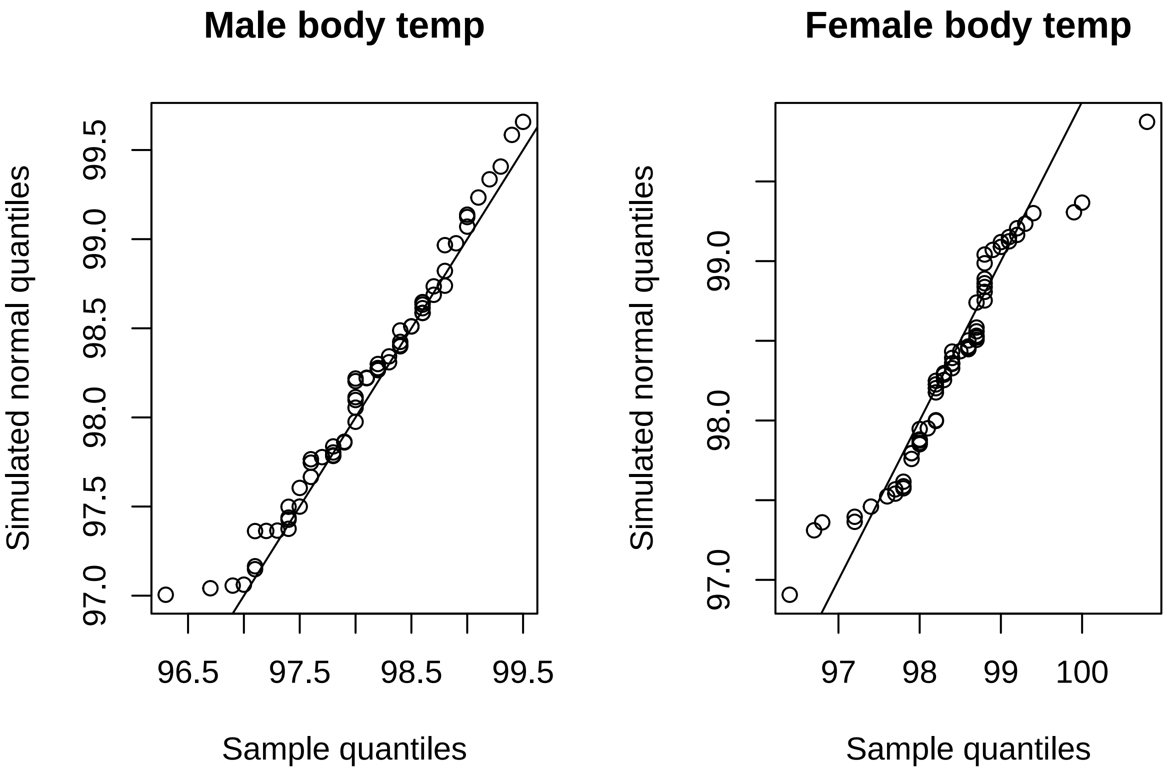

Male body temp
Simulated normal quantiles
96.5
97.5
98.5
96.5
97.5
98.5
99.5
Sample quantiles
Female body temp
Simulated normal quantiles
97.0
98.0
99.0
100.0
97
98
99
100
Sample quantiles

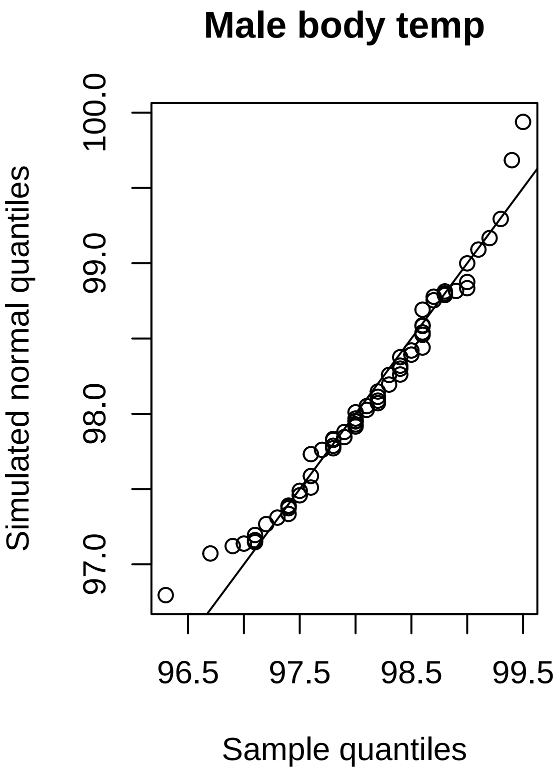

Male body temp
Simulated normal quantiles
96
97
98
99
96.5
97.5
98.5
99.5
Sample quantiles

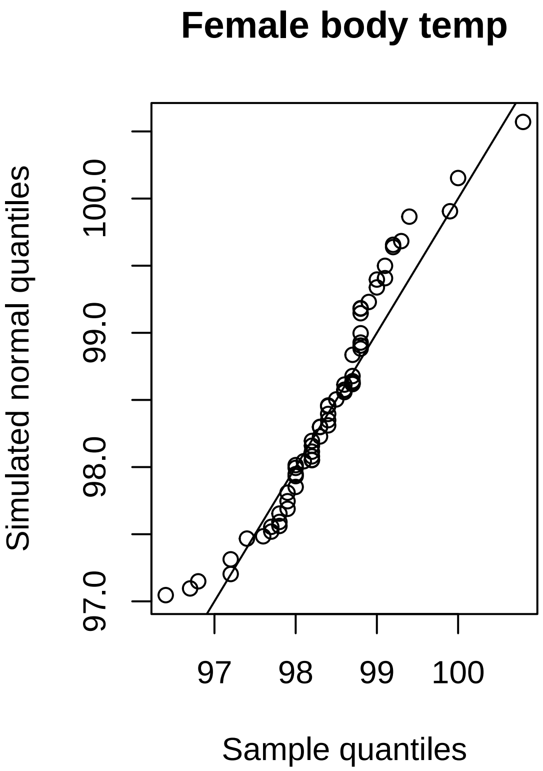

Female body temp
Simulated normal quantiles
97
98
99
100
97
98
99
100
Sample quantiles

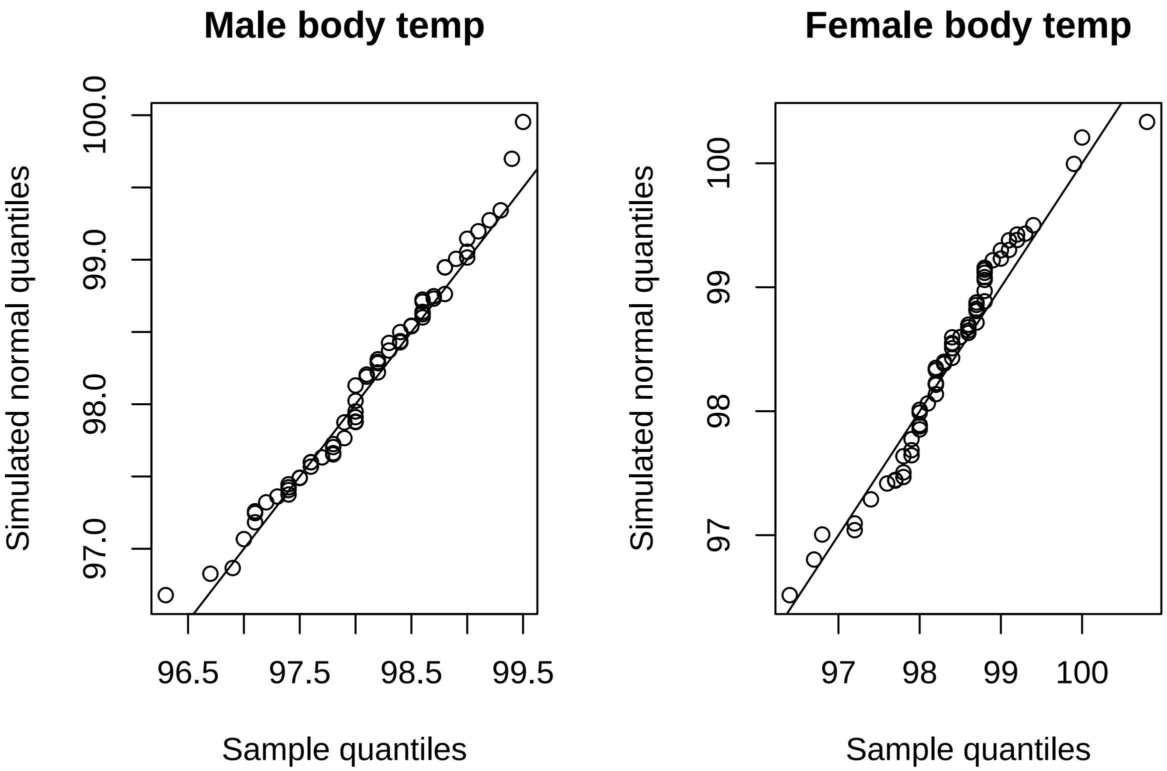

Male body temp
Simulated normal quantiles
97
98
99
100
96.5
97.5
98.5
99.5
Sample quantiles

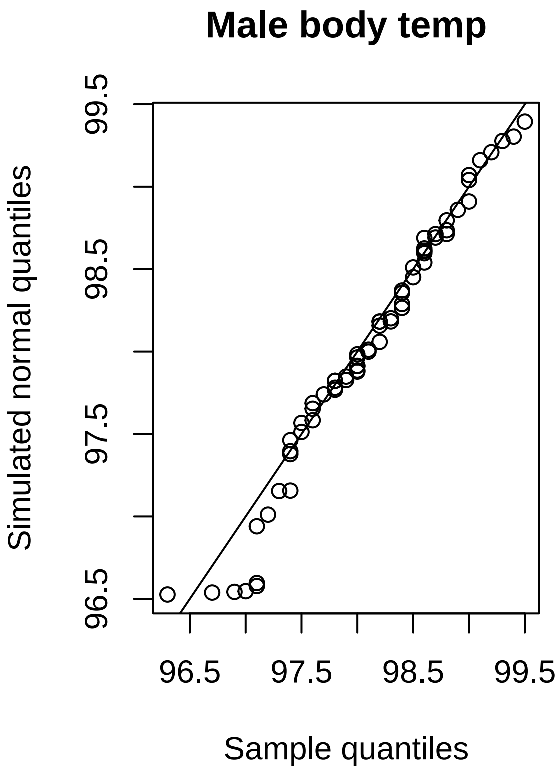

Female body temp
Simulated normal quantiles
97
98
99
100
97
98
99
100
Sample quantiles

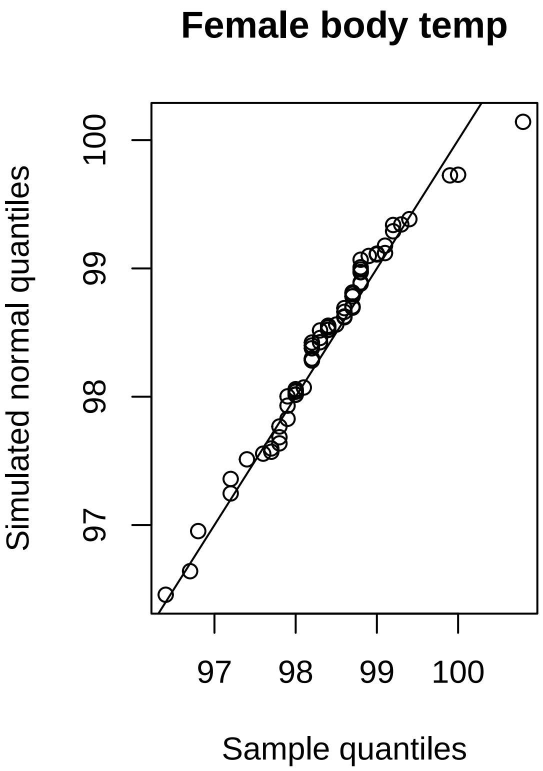

Male body temp
Simulated normal quantiles
97.0
97.5
98.0
98.5
99.0
99.5
96.5
97.5
98.5
99.5
Sample quantiles
Female body temp
Simulated normal quantiles
97.0
98.0
99.0
97
98
99
100
Sample quantiles

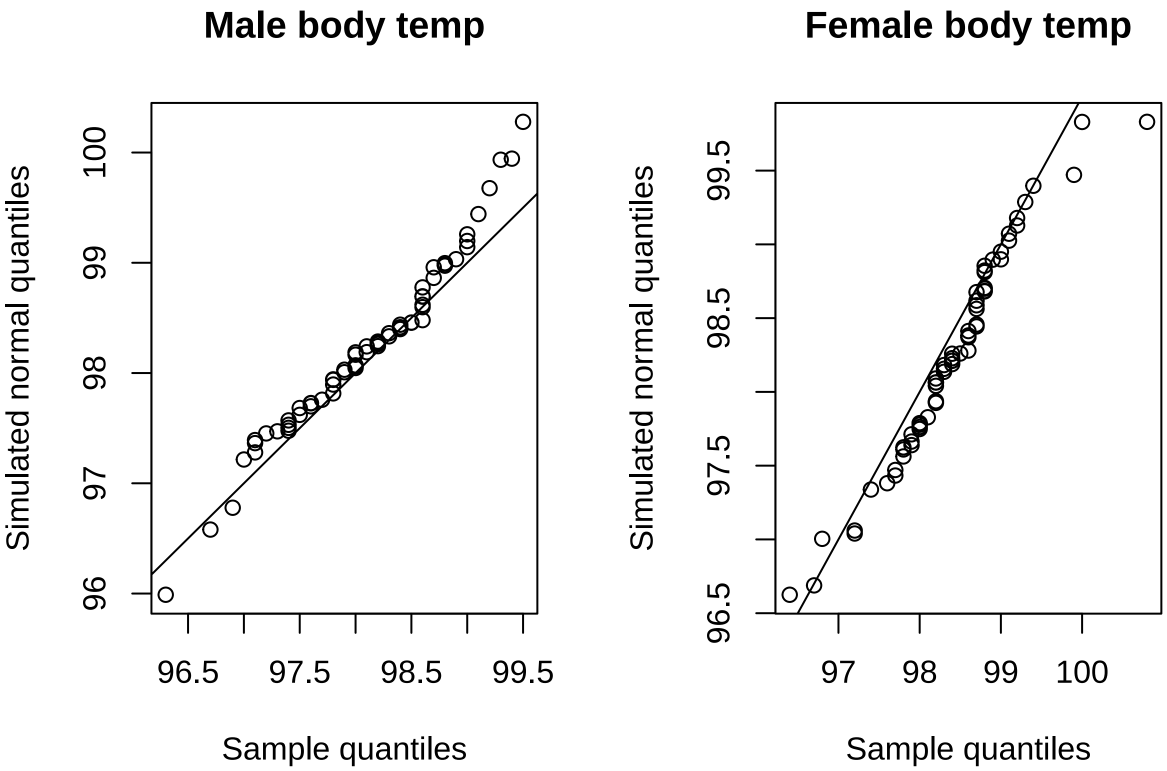

Male body temp
Simulated normal quantiles
97.0
98.0
99.0
100.0
96.5
97.5
98.5
99.5
Sample quantiles

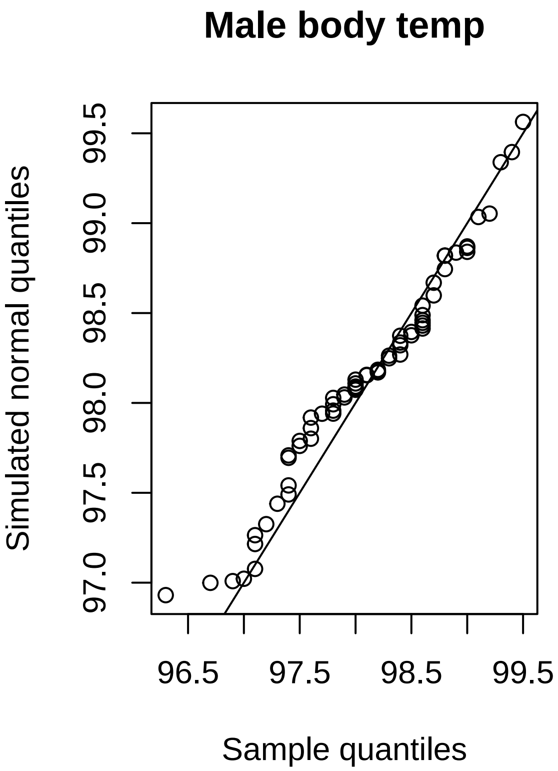

Female body temp
Simulated normal quantiles
97.0
98.0
99.0
100.0
97
98
99
100
Sample quantiles

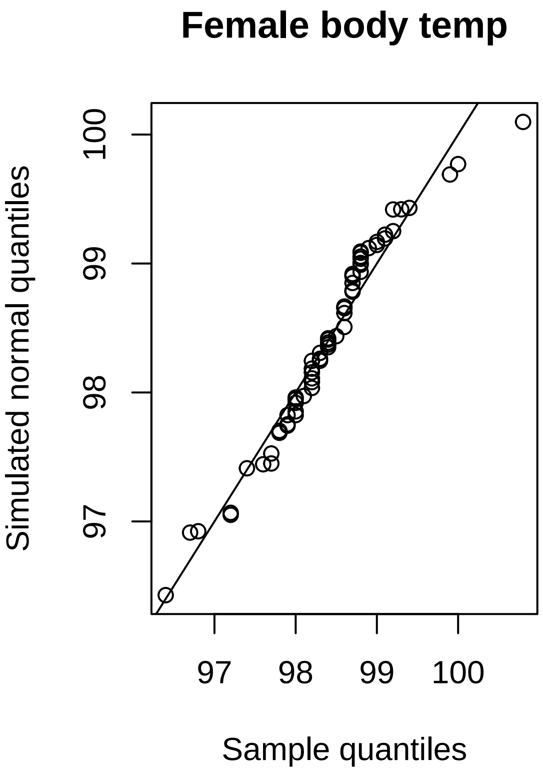

Male body temp
Simulated normal quantiles
97.0
98.0
99.0
100.0
96.5
97.5
98.5
99.5
Sample quantiles
Female body temp
Simulated normal quantiles
97
98
99
100
97
98
99
100
Sample quantiles

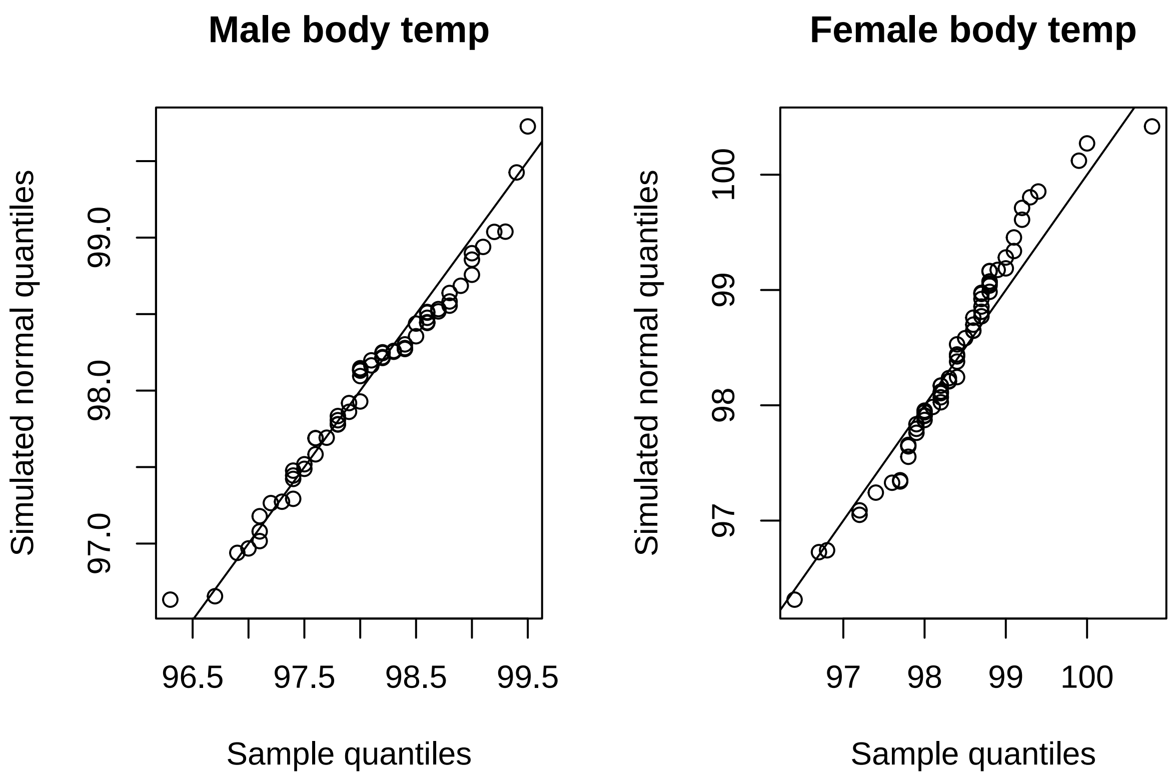

Male body temp
Simulated normal quantiles
96.5
97.5
98.5
99.5
96.5
97.5
98.5
99.5
Sample quantiles


Female body temp
Simulated normal quantiles
97
98
99
100
97
98
99
100
Sample quantiles


Male body temp
Simulated normal quantiles
96
97
98
99
100
96.5
97.5
98.5
99.5
Sample quantiles
Female body temp
Simulated normal quantiles
96.5
97.5
98.5
99.5
97
98
99
100
Sample quantiles


Male body temp
Simulated normal quantiles
97.0
97.5
98.0
98.5
99.0
99.5
96.5
97.5
98.5
99.5
Sample quantiles


Female body temp
Simulated normal quantiles
97
98
99
100
97
98
99
100
Sample quantiles



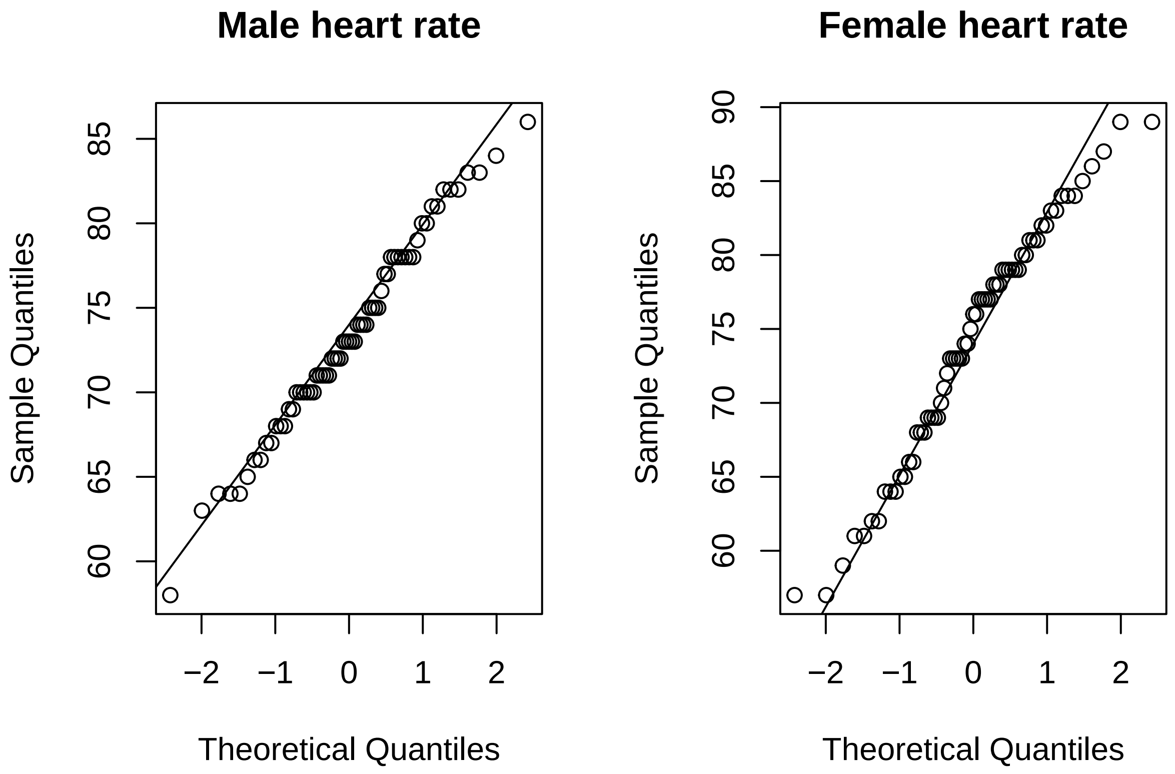
We observe that while male body temperatures seem to follow a more normal distribution than female body temperatures, especially around the tails, it does not seem unreasonable to assume normality for the observed sample. The simulated plots show some variability, but on average the fit seems relatively good. At least acceptable enough to do statistical inference.

b)

We repeat part a) but with heart rates:

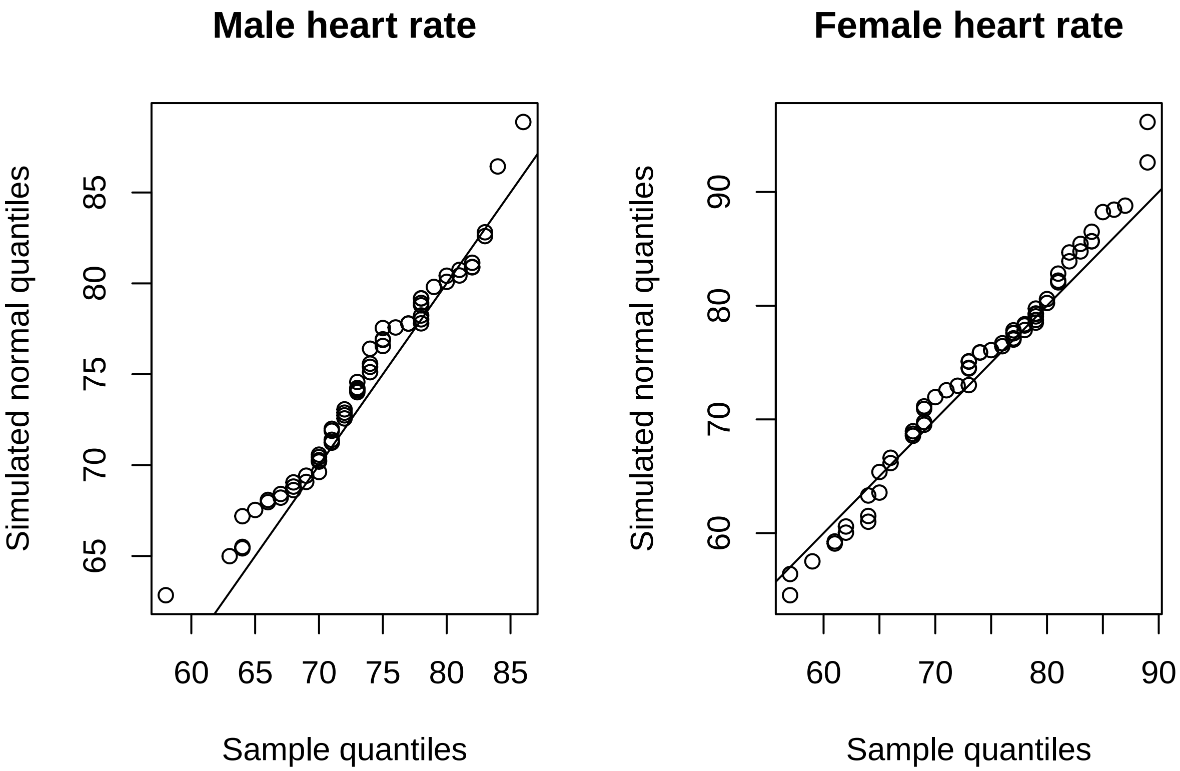
We start by plotting a regular normal quantile plot:

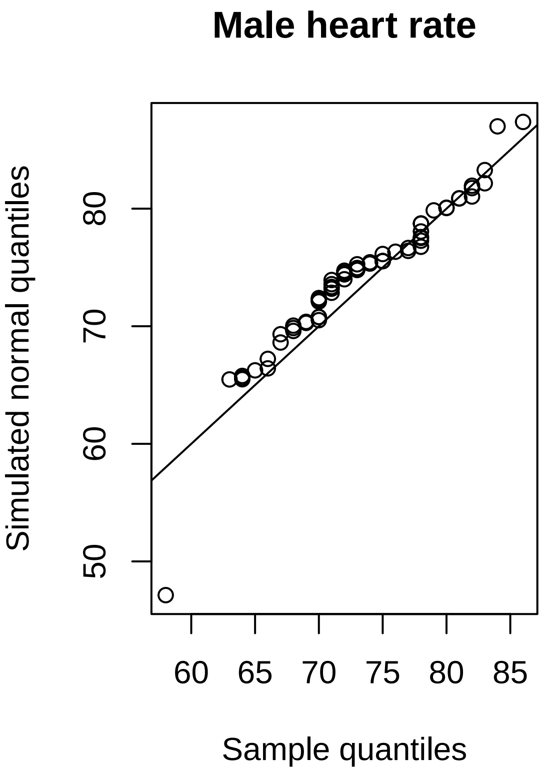
```r
par(mfrow = c(1,2))
qqnorm(bodytemp$rate[bodytemp$gender == 1],
       main = "Male heart rate")
qqline(bodytemp$rate[bodytemp$gender == 1])

qqnorm(bodytemp$rate[bodytemp$gender == 2],
       main = "Female heart rate")
qqline(bodytemp$rate[bodytemp$gender == 2])
```

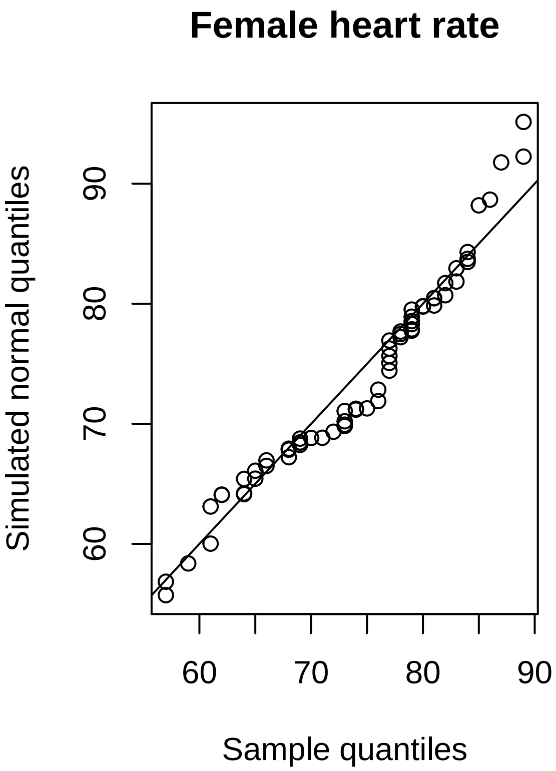


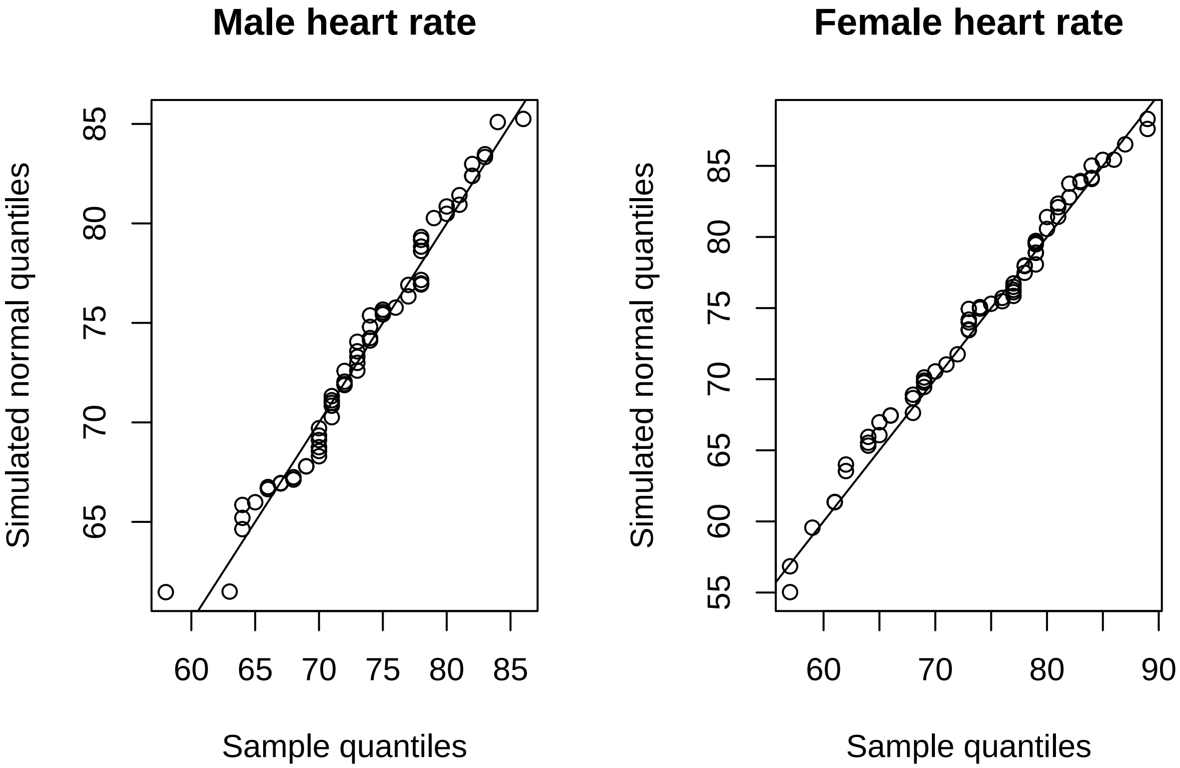
Just like before, the fit is not perfect, but there is enough visual evidence for the assumption of normality not to be outlandish. Now we simulate normal samples for both genders and repeat the exercise to see the plot variability:

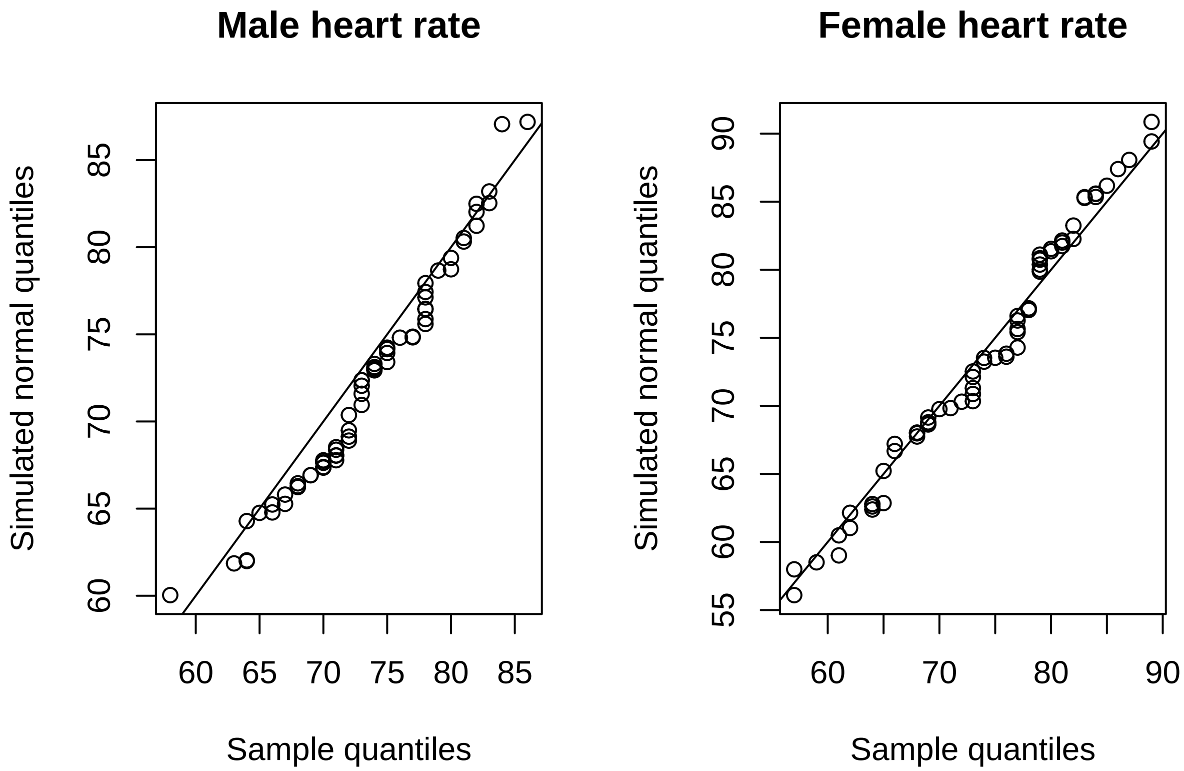
```
for(i in 1:10){
  norm_sim_m <- rnorm(sum(bodytemp$gender == 1),
                      mean = means$rate[1],
                      sd = sds$rate[1])

  norm_sim_f <- rnorm(sum(bodytemp$gender == 2),
                      mean = means$rate[2],
                      sd = sds$rate[2])
  par(mfrow = c(1,2))
  qqplot(bodytemp$rate[bodytemp$gender == 1],
         norm_sim_m,
         xlab = "Sample quantiles",
         ylab = "Simulated normal quantiles",
         main = "Male heart rate")
  abline(a = 0, b = 1)

  qqplot(bodytemp$rate[bodytemp$gender == 2],
         norm_sim_f,
         xlab = "Sample quantiles",
         ylab = "Simulated normal quantiles",
         main = "Female heart rate")
  abline(a = 0, b = 1)
}
```

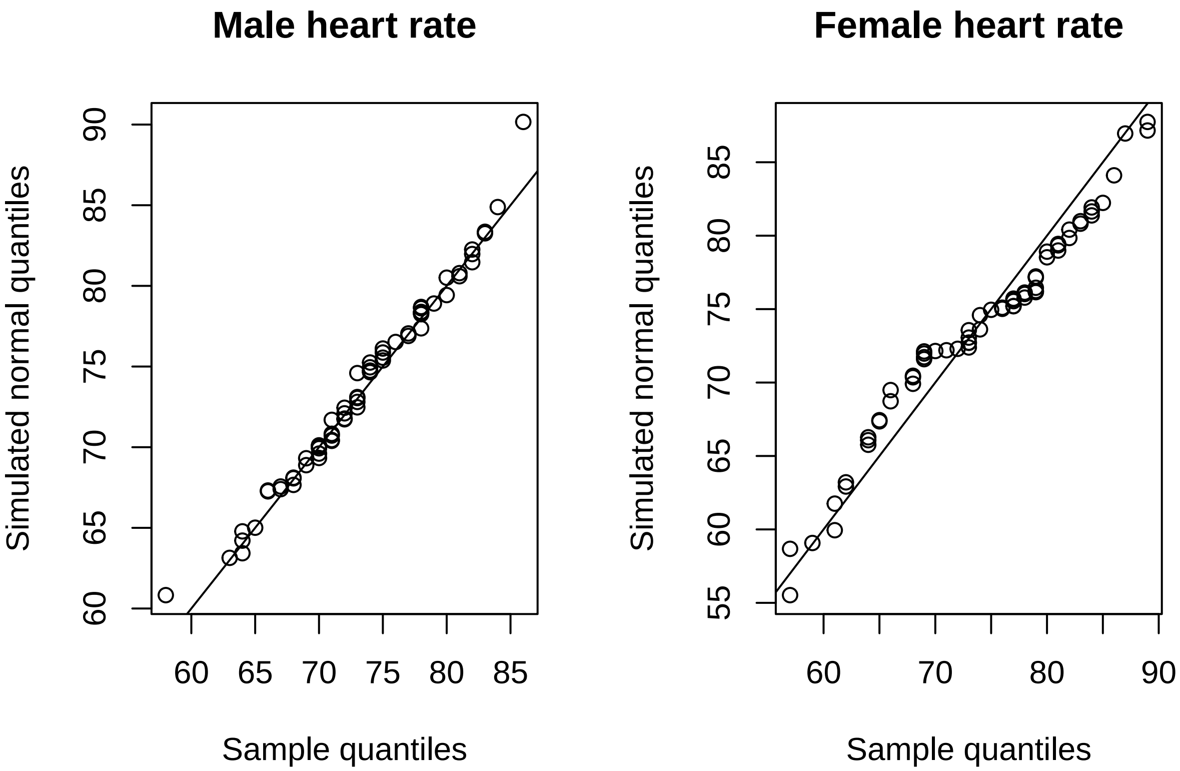

Male heart rate
Simulated normal quantiles
65
70
75
80
85
60
65
70
75
80
85
Sample quantiles
Female heart rate
Simulated normal quantiles
60
70
80
90
60
70
80
90
Sample quantiles

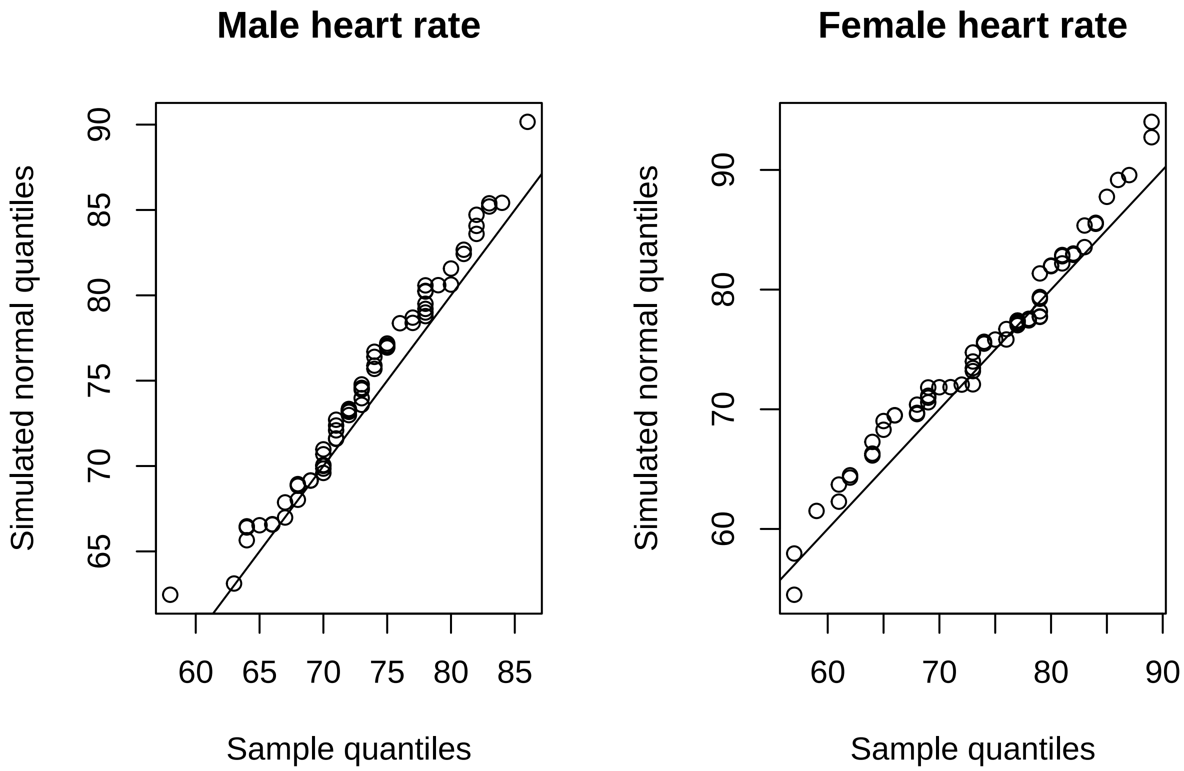

Male heart rate
Simulated normal quantiles
50
60
70
80
60
65
70
75
80
85
Sample quantiles

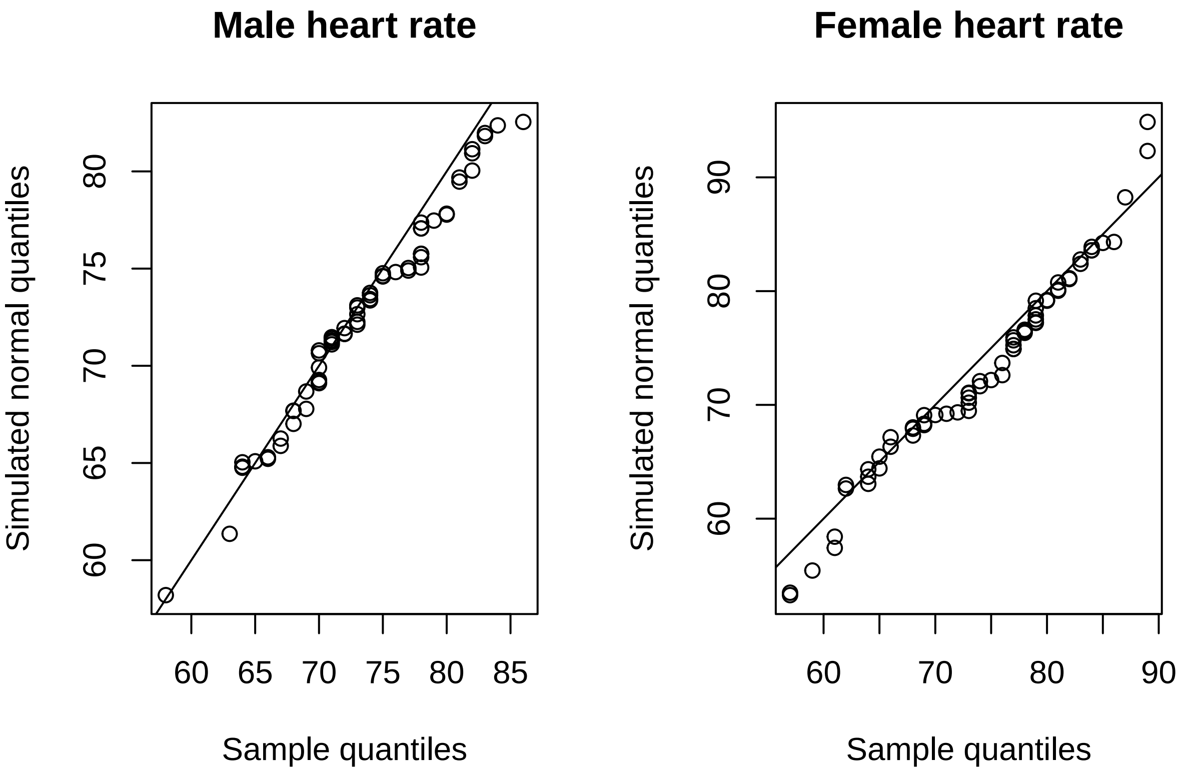

Female heart rate
Simulated normal quantiles
60
70
80
90
60
70
80
90
Sample quantiles

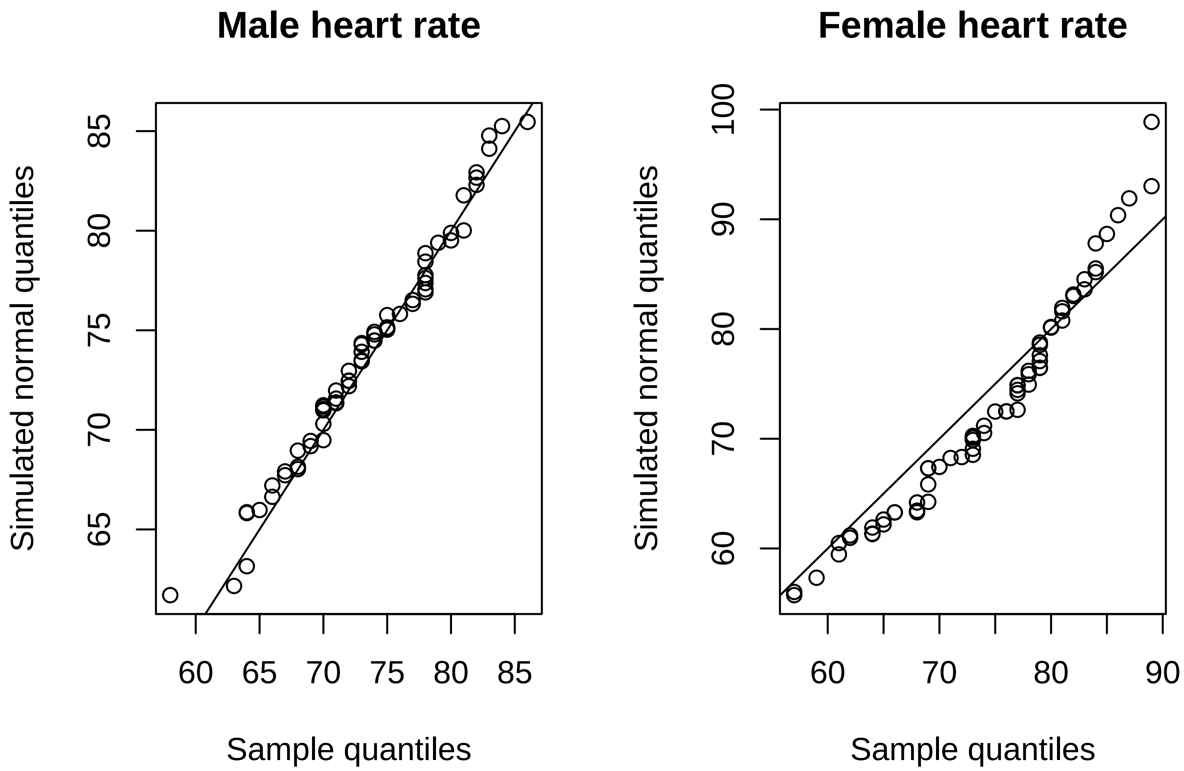

Male heart rate
Simulated normal quantiles
65
70
75
80
85
60
65
70
75
80
85
Sample quantiles
Female heart rate
Simulated normal quantiles
55
60
65
70
75
80
85
60
70
80
90
Sample quantiles

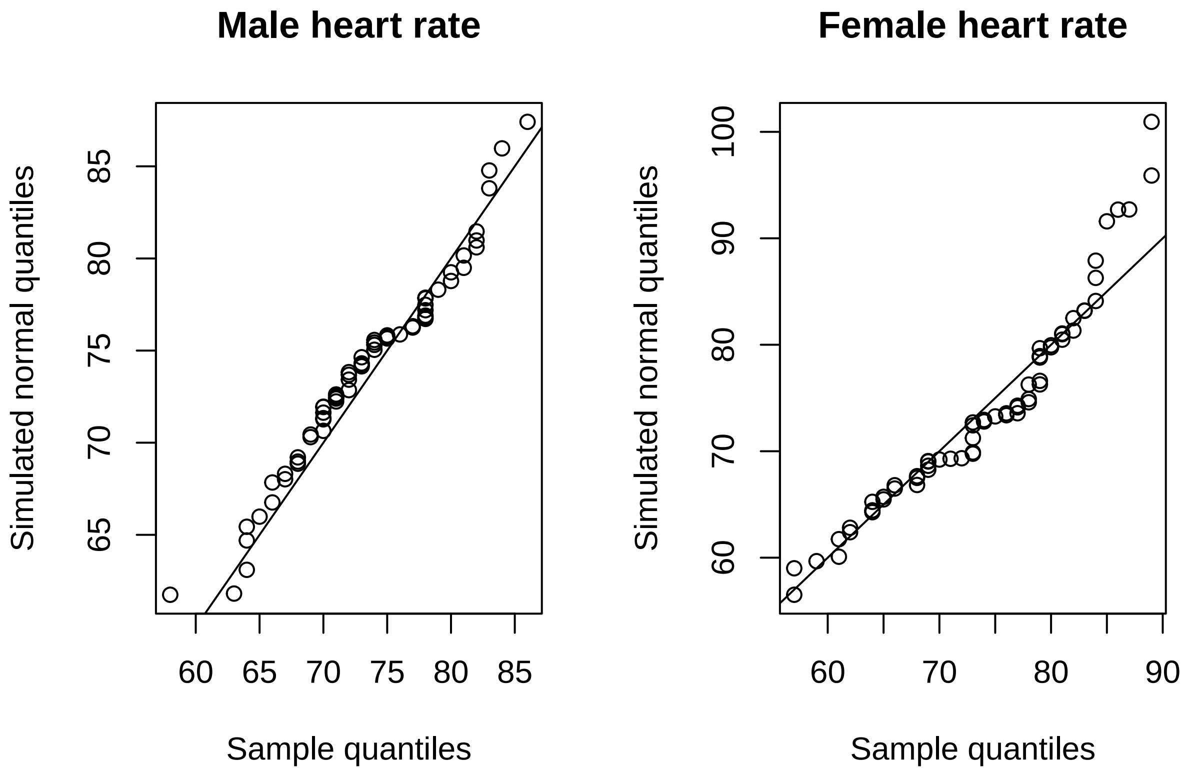

Male heart rate
Simulated normal quantiles
60
65
70
75
80
85
Sample quantiles
60
65
70
75
80
85
Female heart rate
Simulated normal quantiles
55
60
65
70
75
80
85
90
Sample quantiles
60
70
80
90

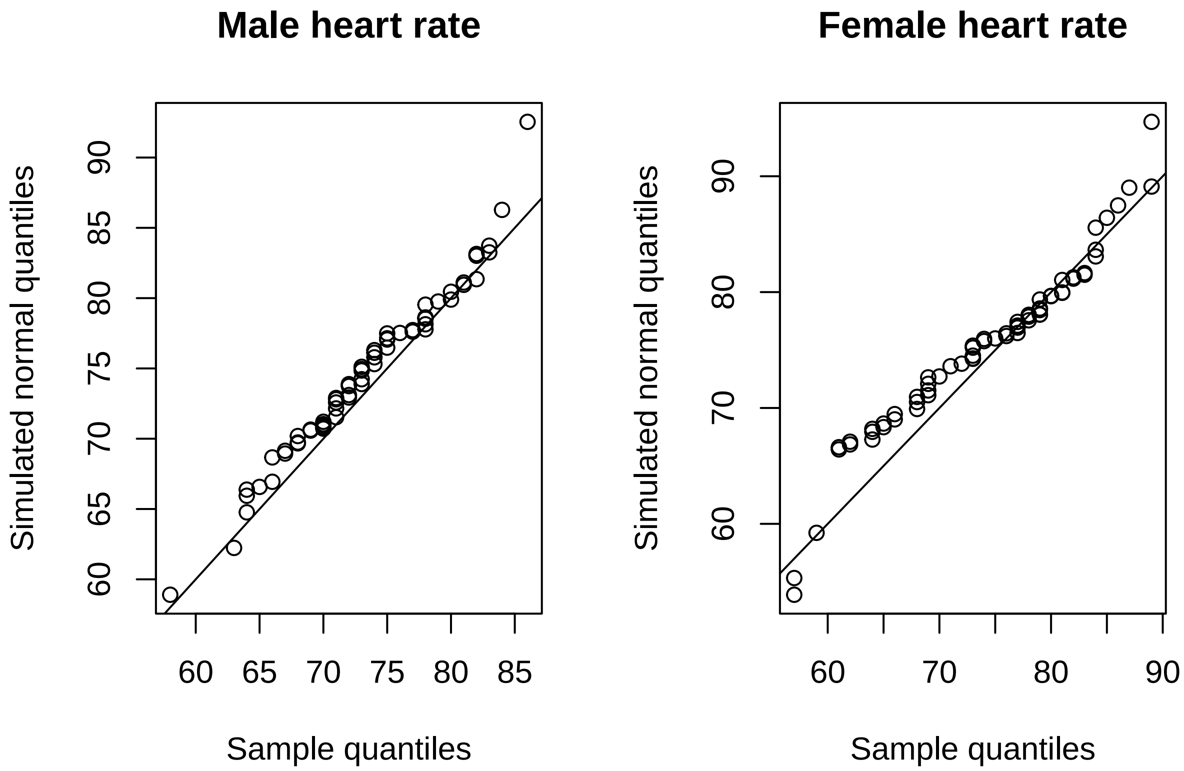

Male heart rate
Female heart rate
Simulated normal quantiles
Sample quantiles
60
65
70
75
80
85
90
55


Male heart rate
Simulated normal quantiles
65
70
75
80
85
90
60
65
70
75
80
85
Sample quantiles
Female heart rate
Simulated normal quantiles
60
70
80
90
60
70
80
90
Sample quantiles


Male heart rate
Simulated normal quantiles
60
65
70
75
80
60
65
70
75
80
85
Sample quantiles
Female heart rate
Simulated normal quantiles
60
70
80
90
60
70
80
90
Sample quantiles


Male heart rate
Simulated normal quantiles
65
70
75
80
85
60
65
70
75
80
85
Sample quantiles
Female heart rate
Simulated normal quantiles
60
70
80
90
100
60
70
80
90
Sample quantiles


Male heart rate
Simulated normal quantiles
65
70
75
80
85
60
65
70
75
80
85
Sample quantiles
Female heart rate
Simulated normal quantiles
60
70
80
90
100
60
70
80
90
Sample quantiles



In this case we observe a much better normal fit than the previous point in most of the cases. However, there seems to be even more variability between the plots, particularly in the female case. Still, it does not appear unreasonable to assume normality for statistical inference.

c)

We start with the males:

```
t.test(bodytemp$temperature[bodytemp$gender == 1],
       mu = 98.6,
       alternative = "two.sided")
```

```
##
##  One Sample t-test
##
## data:  bodytemp$temperature[bodytemp$gender == 1]
## t = -5.7158, df = 64, p-value = 3.084e-07
## alternative hypothesis: true mean is not equal to 98.6
## 95 percent confidence interval:
##  97.93147 98.27776
## sample estimates:
## mean of x
##  98.10462
```

We observe that the t-test quite strongly rejects the null that the mean body temperature is 98.6 degrees. We now do the same for women:

```r
t.test(bodytemp$temperature[bodytemp$gender == 2],
       mu = 98.6,
       alternative = "two.sided")
```

```
##
##  One Sample t-test
##
## data:  bodytemp$temperature[bodytemp$gender == 2]
## t = -2.2355, df = 64, p-value = 0.02888
## alternative hypothesis: true mean is not equal to 98.6
## 95 percent confidence interval:
##  98.20962 98.57807
## sample estimates:
## mean of x
##  98.39385
```

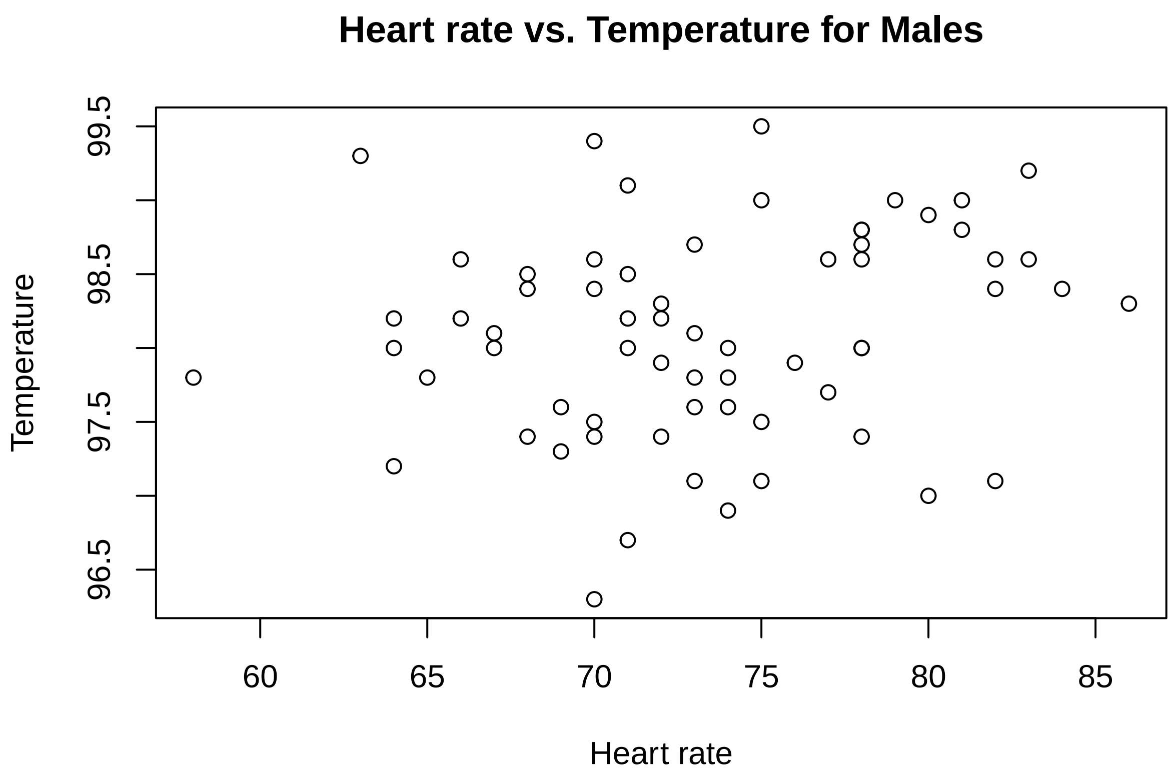
For females once again we get strong evidence against the mean being 98.6 degrees. However in this case the evidence is not as strong as in males and thus we don't reject the null at all significance levels. At 5% we do, but not at 1%, for example.

Exrcise 102

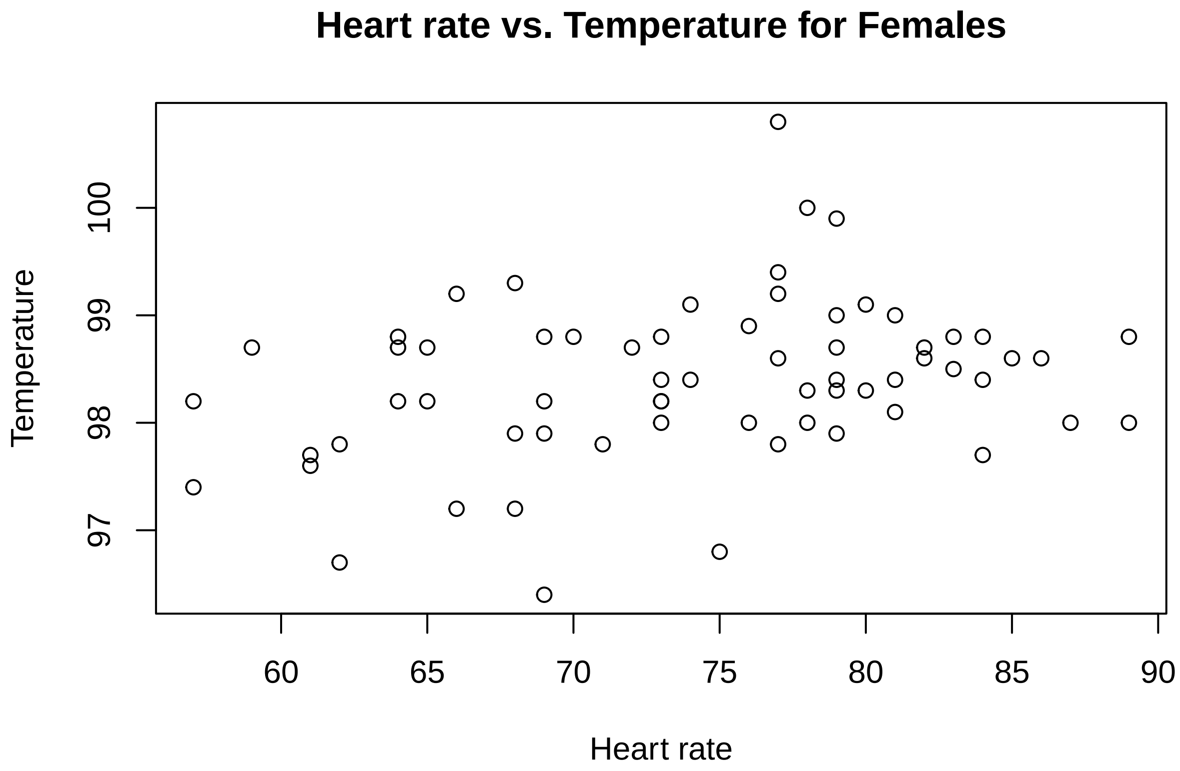
a)

We start creating the scatterplots:

```r
plot(bodytemp$rate[bodytemp$gender == 1],
     bodytemp$temperature[bodytemp$gender == 1],
     main = "Heart rate vs. Temperature for Males",
     xlab = "Heart rate",
     ylab = "Temperature")
```



```
plot(bodytemp$rate[bodytemp$gender == 2],
     bodytemp$temperature[bodytemp$gender == 2],
     main = "Heart rate vs. Temperature for Females",
     xlab = "Heart rate",
     ylab = "Temperature")
```



There seems to be no discernible pattern or relationship between the two variables in either of the genders.

b)

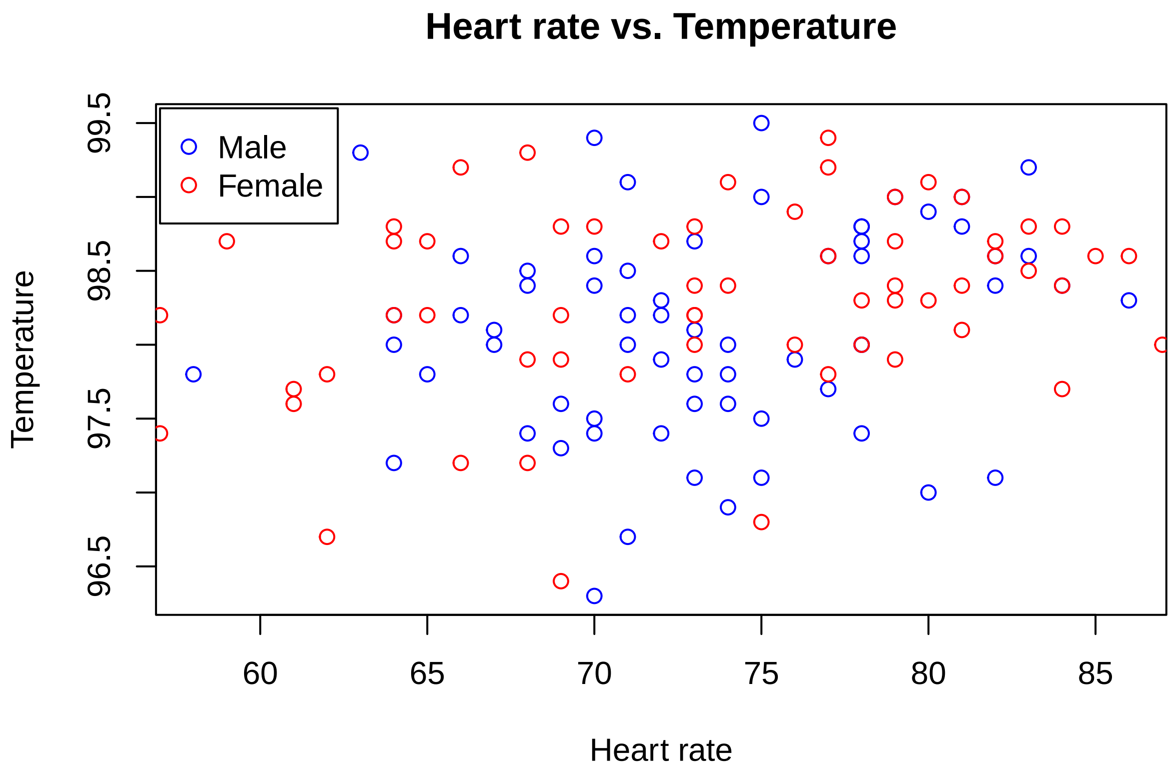
```
do.call(rbind, by(bodytemp,
                  INDICES = list(bodytemp$gender),
                  FUN = function(x) data.frame(Gender = unique(x$gender),
                                               Pearson.Corr =  cor(x$temperature, x$rate, method = "pea
                                               Rank.Corr = cor(x$temperature, x$rate, method = "spearma
```

```
##   Gender Pearson.Corr Rank.Corr
## 1      1    0.1955894 0.2239387
## 2      2    0.2869312 0.2655711
```

We observe that while there is a positive correlation between the variables, both using the Pearson and Spearman rank methods, however it is not high.

c)

```
plot(bodytemp$rate[bodytemp$gender == 1],
     bodytemp$temperature[bodytemp$gender == 1],
     col = "blue",
     main = "Heart rate vs. Temperature",
     xlab = "Heart rate",
     ylab = "Temperature")
points(bodytemp$rate[bodytemp$gender == 2],
       bodytemp$temperature[bodytemp$gender == 2],
       col = "red")
legend(57, 99.6, legend=c("Male", "Female"),
       col=c("blue","red"), pch = 1)
```



Yes, the relationship (or rather the lack thereof) seems to be the same for both males and females. In both cases there is no discernible pattern whatsoever, and if not by the color of the dots, the observations would not be distinguishable between groups.

Exercise 104

a)

Suppose we allocate s_x elements to group X. Then $s_y = s - s_x$ will go to group Y. Observe we have two cases: if the variance is known and if it is not known. Let's start with the known case, as it is simpler.

By normality of the samples, we have that:

$$\bar{X} - \bar{Y} \sim N\left(\mu_x - \mu_y, \sigma^2\left(\frac{1}{s_x} + \frac{1}{s_y}\right)\right)$$

Then a $100(1-\alpha)\%$ confidence interval for this specific scenario is given by:

$$(\bar{X}-\bar{Y}) \pm z_{1-\alpha/2}\sigma\sqrt{\frac{1}{s_x}+\frac{1}{s_y}}$$

Notice the mean difference is just a constant that centers the confidence interval, and neither the normal quantile nor the variance depend on the allocation. Also, the interval is symmetric around the sample mean difference. Therefore, the only thing that affects its width is the term inside the root, which we now minimize:

$$\frac{1}{s_x}+\frac{1}{s_y}=\frac{1}{s_x}+\frac{1}{s-s_x}$$

Differentiating with respect to s_x we obtain:

$$\frac{1}{(s-s_x)^2}-\frac{1}{s_x^2}=0 \iff s_x=\frac{s}{2}$$

We know this minimizes, as taking the second derivative with respect to s_x yields:

$$\frac{2}{(s-s_x)^3}+\frac{2}{s_x^3}>0$$

Therefore we minimize the interval length when the samples are symmetric. Of course this only makes sense when s is even. If not, it is enough for the difference between sample sizes to be 1.

If σ^2 is not known, we will change notation a bit to avoid confusion. Denote n_x, n_y what we denoted s_x, s_y respectively in the previous case. Recall from corollary seen in class, we have that the $100(1-\alpha)\%$ confidence interval for this specific scenario is given by:

$$(\bar{X}-\bar{Y}) \pm t_{n_x+s-n_x-2,1-\alpha/2}\cdot s_p\sqrt{\frac{1}{n_x}+\frac{1}{s-n_x}}$$

Notice two things: first, the t-quantile does not depend on the allocation as n_x cancels out. Second, just like before, the interval's length does not depend on $\bar{X}-\bar{Y}$, as this is just a constant that is its center. However, we now have the added complication that the width depends on $\bar{X}$ and $\bar{Y}$ through the pooled sample standard deviation s_p. However, we can still reduce the interval's length by simply minimizing the expression in the root, which is exactly the same as in the previous case. For that reason, we also minimize the confidence interval length when $n_x = s/2$.

b)

Recall due to the duality between confidence intervals and hypothesis testing, we can reject the null hypothesis with a level of α when our observation lies outside of the confidence interval under the null hypothesis. By minimizing the length of the confidence interval, we also maximize the rejection region, as it is its complement. By maximizing the rejection region, we also maximize the power of the test. That means that the allocation that maximizes the test power is exactly the allocation found in part a).

Exercise 108

Let $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$ be independent. Then we have that:

$$P(X<Y)=P(X-Y<0)=P(X+(-1)\cdot Y<0)$$

Since X, Y are independent, then:

$$X + (-Y) \sim N(\mu_x - \mu_y, \sigma_x^2 + \sigma_y^2)$$

And,

$$X - Y < 0 \iff \frac{(X-Y) - (\mu_x - \mu_y)}{\sqrt{\sigma_x^2 + \sigma_y^2}} < \frac{-(\mu_x - \mu_y)}{\sqrt{\sigma_x^2 + \sigma_y^2}}$$

Notice that

$$\frac{(X-Y) - (\mu_x - \mu_y)}{\sqrt{\sigma_x^2 + \sigma_y^2}} \sim N(0,1)$$

Then we have that:

$$\begin{aligned}
P(X < Y) &= P(X - Y < 0) \\
&= P\left(\frac{(X-Y) - (\mu_x - \mu_y)}{\sqrt{\sigma_x^2 + \sigma_y^2}} < \frac{-(\mu_x - \mu_y)}{\sqrt{\sigma_x^2 + \sigma_y^2}} \right) \\
&= \Phi\left(\frac{-(\mu_x - \mu_y)}{\sqrt{\sigma_x^2 + \sigma_y^2}} \right)
\end{aligned}$$

Exercise 109

For this exercise we will use methods based on normal distributions as we can obtain explicit forms for the power. We will also be computing the case for a two-sided test.

a)

Let's start with the paired case. Define $D = X - Y$. Note that:

$$\begin{aligned}
Var(D) &= \sigma_x^2 + \sigma_y^2 - 2Cov(X, Y) \\
&= 100 + 100 - 2 \cdot 50 \\
&= 100
\end{aligned}$$

Now define $\bar{D} = \bar{X} - \bar{Y}$ and notice that $E[\bar{D}] = \mu_x - \mu_y =: \Delta$, and $Var(\bar{D}) = \frac{100}{25} = 4$

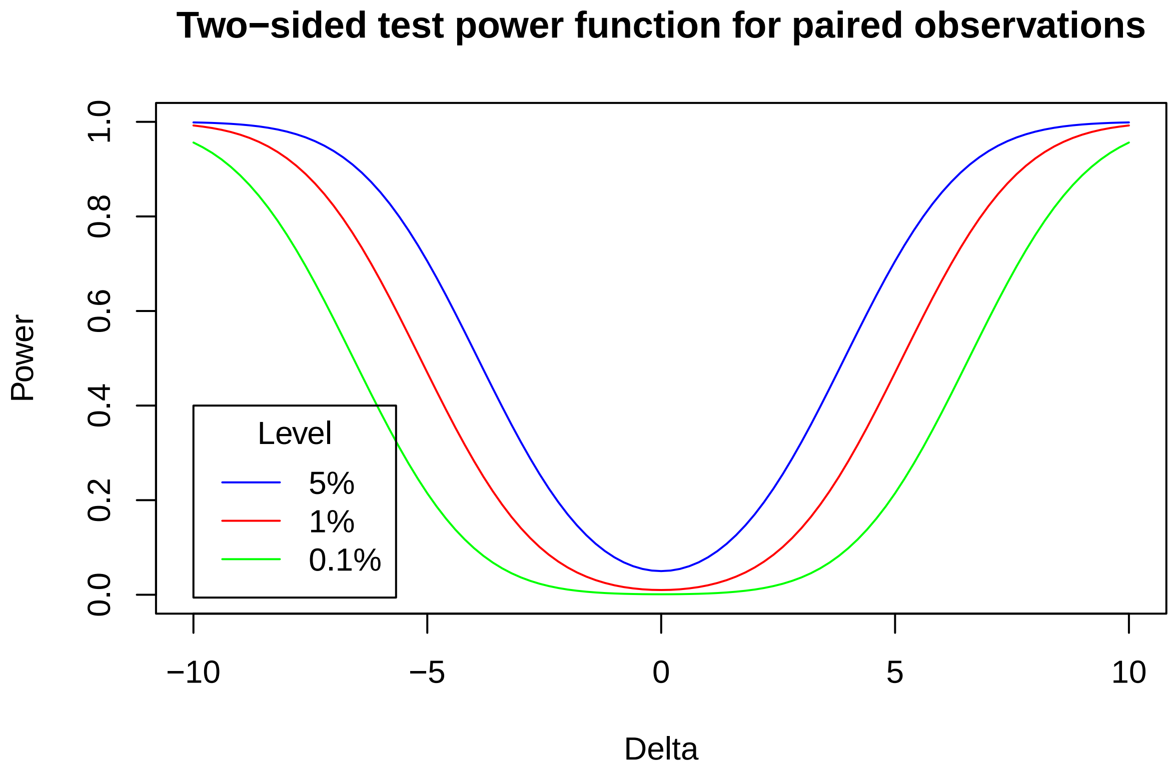
That means that $\bar{D} \sim N(\Delta, 4)$, then $\frac{\bar{D} - \Delta}{2} \sim N(0,1)$.

Under the null hypothesis $\Delta = 0$, and given a level α, which means we reject if $\left|\frac{\bar{D}}{2}\right| > z_{1-\alpha/2}$. However, let's assume the alternative hypothesis is true, and $\Delta \neq 0$. Then:

$$
\begin{aligned}
P\left(\left|\frac{\bar{D}}{2}\right| > z_{1-\alpha/2}\right) &= P\left(\frac{\bar{D}}{2} > z_{1-\alpha/2}\right) + P\left(\frac{\bar{D}}{2} < z_{\alpha/2}\right) \\
&= P(\bar{D} > 2z_{1-\alpha/2}) + P(\bar{D} < 2z_{\alpha/2}) \\
&= P\left(\frac{\bar{D}-\Delta}{2} > \frac{2z_{1-\alpha/2}-\Delta}{2}\right) + P\left(\frac{\bar{D}-\Delta}{2} < \frac{2z_{\alpha/2}-\Delta}{2}\right) \\
&= 1 - \Phi\left(\frac{2z_{1-\alpha/2}-\Delta}{2}\right) + \Phi\left(\frac{2z_{\alpha/2}-\Delta}{2}\right)
\end{aligned}
$$

Let's plot this, for various levels of significance:

```
power_paired <- function(delta, alpha){
  1 - pnorm( (2*qnorm(1 - alpha/2) - delta)/2 ) + pnorm( (2*qnorm(alpha/2) - delta)/2 )
}

curve(power_paired(delta = x, alpha = 0.05),
      from = -10,
      to = 10,
      xname = "x",
      col = "blue",
      ylim = c(0,1),
      xlab = "Delta",
      ylab = "Power",
      main = "Two-sided test power function for paired observations")
curve(power_paired(delta = x, alpha = 0.01),
      from = -10,
      to = 10,
      xname = "x",
      col = "red",
      add = T)
curve(power_paired(delta = x, alpha = 0.001),
      from = -10,
      to = 10,
      xname = "x",
      col = "green",
      add = T)
legend(-10, 0.4, legend=c("5%", "1%", "0.1%"),
       col=c("blue","red", "green"), lty = 1,
       title = "Level")
```



b)

Now let's do the same but for unpaired observations. Again define D and $\bar{D}$ as above. In this case, $Var(D) = \sigma_x^2 + \sigma_y^2 = 100 + 100 = 200$, then $Var(\bar{D}) = \frac{200}{25} = 8$.

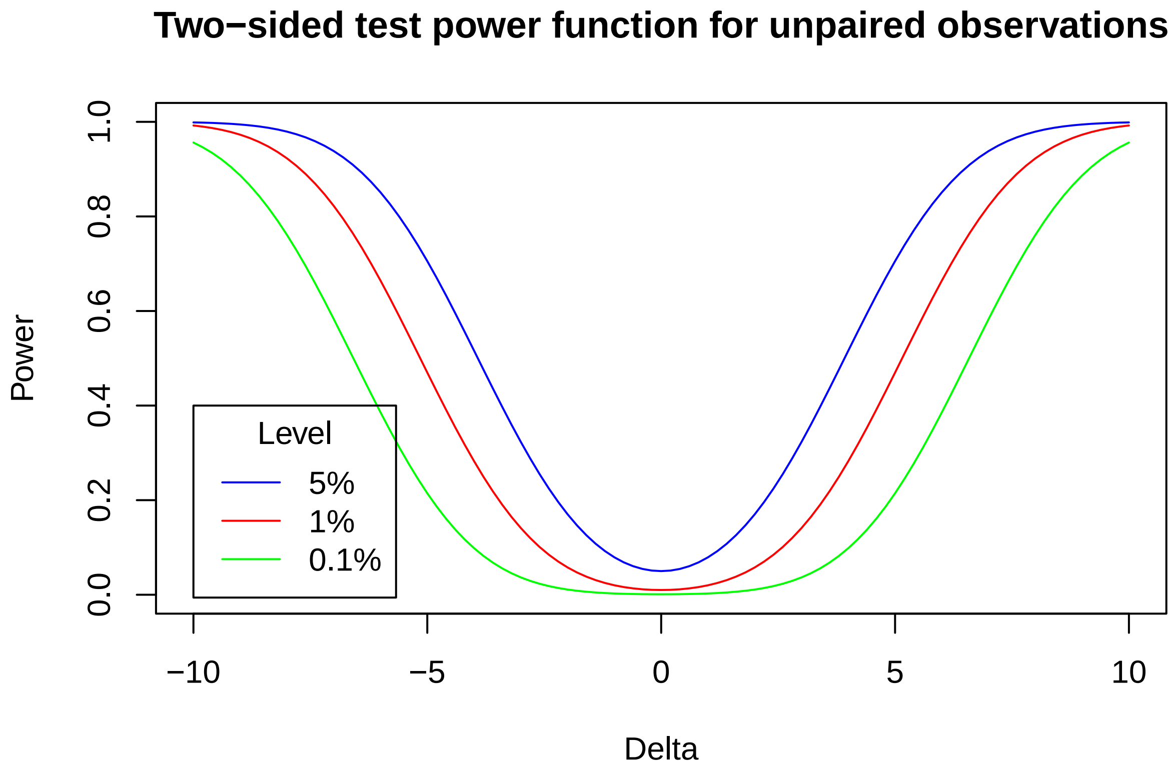
That means that $\bar{D} \sim N(\Delta, 8)$, then $\frac{\bar{D}-\Delta}{\sqrt{8}} \sim N(0,1)$.

Under the null hypothesis $\Delta = 0$, and given a level α, which means we reject if $\left|\frac{\bar{D}}{\sqrt{8}}\right| > z_{1-\alpha/2}$. However, let's assume the alternative hypothesis is true, and $\Delta \neq 0$. Then:

$$
\begin{aligned}
P\left(\left|\frac{\bar{D}}{\sqrt{8}}\right| > z_{1-\alpha/2}\right) &= P\left(\frac{\bar{D}}{\sqrt{8}} > z_{1-\alpha/2}\right) + P\left(\frac{\bar{D}}{\sqrt{8}} < z_{\alpha/2}\right) \\
&= P(\bar{D} > \sqrt{8}z_{1-\alpha/2}) + P(\bar{D} < \sqrt{8}z_{\alpha/2}) \\
&= P\left(\frac{\bar{D}-\Delta}{\sqrt{8}} > \frac{\sqrt{8}z_{1-\alpha/2}-\Delta}{\sqrt{8}}\right) + P\left(\frac{\bar{D}-\Delta}{\sqrt{8}} < \frac{\sqrt{8}z_{\alpha/2}-\Delta}{\sqrt{8}}\right) \\
&= 1 - \Phi\left(\frac{\sqrt{8}z_{1-\alpha/2}-\Delta}{\sqrt{8}}\right) + \Phi\left(\frac{\sqrt{8}z_{\alpha/2}-\Delta}{\sqrt{8}}\right)
\end{aligned}
$$

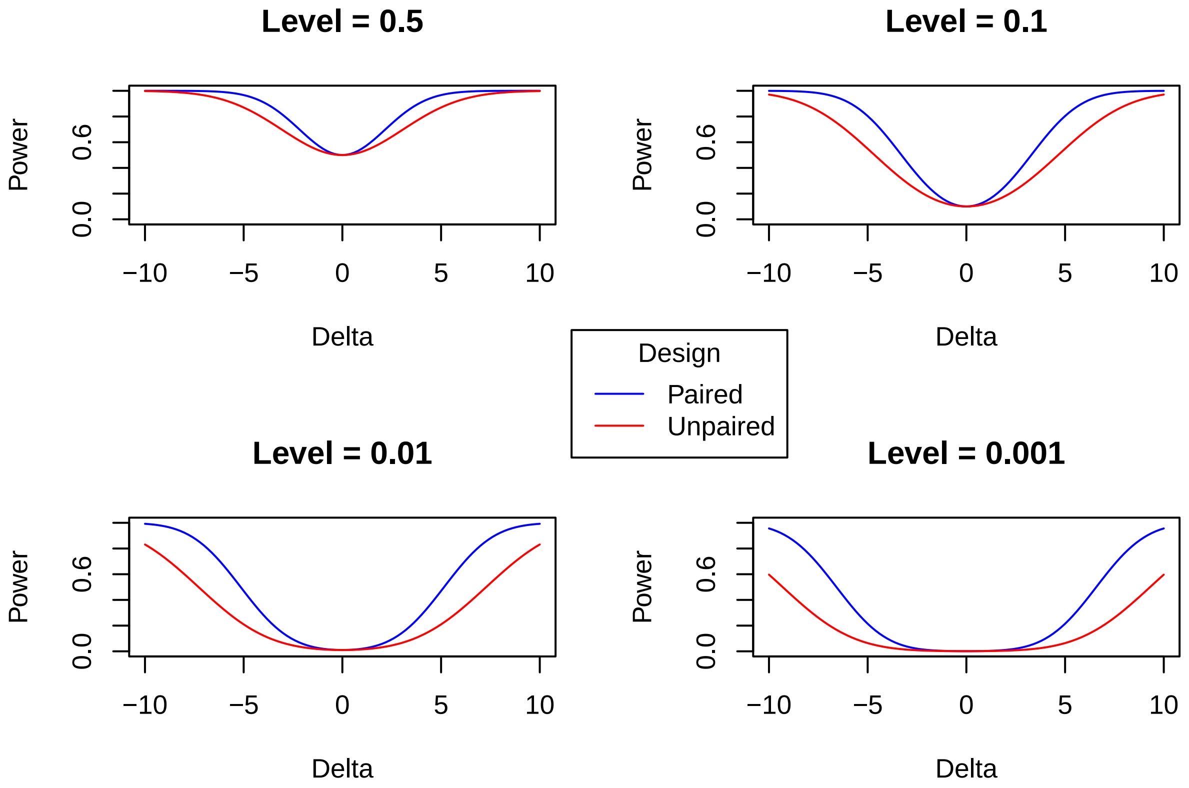
We proceed to plot for various significance levels:

```
power_unpaired <- function(delta, alpha){
  1 - pnorm( (sqrt(8)*qnorm(1 - alpha/2) - delta)/sqrt(8) ) + pnorm( (sqrt(8)*qnorm(alpha/2) - delta)/s
}

curve(power_paired(delta = x, alpha = 0.05),
      from = -10,
      to = 10,
      xname = "x",
      col = "blue",
      ylim = c(0,1),
      xlab = "Delta",
      ylab = "Power",
      main = "Two-sided test power function for unpaired observations")
curve(power_paired(delta = x, alpha = 0.01),
      from = -10,
      to = 10,
      xname = "x",
      col = "red",
      add = T)
curve(power_paired(delta = x, alpha = 0.001),
      from = -10,
      to = 10,
      xname = "x",
      col = "green",
      add = T)
legend(-10, 0.4, legend=c("5%", "1%", "0.1%"),
       col=c("blue","red", "green"), lty = 1,
       title = "Level")
```



Now let's compare the curves:

```
par(mfrow = c(2,2))
for(alpha in c(0.5, 0.1, 0.01, 0.001)){
  curve(power_paired(delta = x, alpha = alpha),
        from = -10,
        to = 10,
        xname = "x",
        col = "blue",
        ylim = c(0,1),
        xlab = "Delta",
        ylab = "Power",
        main = paste0('Level = ', alpha))
  curve(power_unpaired(delta = x, alpha = alpha),
        from = -10,
        to = 10,
        xname = "x",
        col = "red",
        add = T)
}
par(xpd=NA)
#legend(locator(1), legend=as.numeric(levels(factor(mtcars$cyl))), pch=19, col= as.numeric(levels(facto
legend(-20,2.5, legend=c("Paired", "Unpaired"),
       col=c("blue","red"), lty = 1,
       title = "Design")
```



As we can see, the paired test is consistently more powerful than the unpaired one. This makes sense in our case due to the fact that, as discussed in class, here we have a positive covariance between the variables in the paired design, resulting a decrease of the variance of D. This decrease in variance is not present in the unpaired case, which makes the test less powerful. Also, as expected, as the level decreases the test rejects less often, resulting in a generally less powerful test. The reason for this can be seen in last homework's True/False exercise.

Exercise 113

a)

As per slide 7 in the Lecture slides, we can make use of a two sample-t test with the assumption of normality and same variance for both samples to test the difference in means in body temperatures between male and female. Note that the pooled variance for two groups with the same size can be expressed as:

$$\begin{aligned} s_p^2 &= \frac{(n-1)s_1^2 + (m-1)s_2^2}{m+n-2} \qquad \text{(note that m=n)} \\ &= \frac{(m-1)(s_1^2 + s_2^2)}{2(m-1)} \\ &= \frac{(s_1^2 + s_2^2)}{2} \end{aligned}$$

The standard error then will be:

$$\begin{aligned} SE &= \sqrt{s_p^2}\sqrt{\frac{1}{m}+\frac{1}{m}} \\ &= \sqrt{\frac{(s_1^2+s_2^2)}{2}}\sqrt{\frac{2}{m}} \\ &= \sqrt{\frac{s_1^2}{m}+\frac{s_2^2}{m}} \end{aligned}$$

Therefore, we can use the following statistic to form the confidence intervals:

$$t = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{m}+\frac{s_2^2}{m}}}$$

Which will be student t-distributed with m+n+2 degrees if freedom. The confidence interval for the difference in means will be

$$CI = (\bar{x}_1-\bar{x}_2) \pm t_{1-\frac{\alpha}{2}} \times \sqrt{\frac{s_1^2}{m}+\frac{s_2^2}{m}}$$

```
setwd(myPath)
bodytemp<-read.table(paste(myPath,"bodytemp.txt", sep="/"),header=TRUE,sep=",")


maletemp<-bodytemp$temperature[bodytemp$gender==1]
femaletemp<-bodytemp$temperature[bodytemp$gender==2]

# Calculating mean and standard deviation for male and female temperatures
x1 <- mean(maletemp)
s1 <- sd(maletemp)
x2 <- mean(femaletemp)
s2 <- sd(femaletemp)

# Determining sample sizes
n <- length(maletemp)
m <- length(femaletemp)

# Calculating the difference in means and its standard deviation
Difference_mean <- x1 - x2
SE <- sqrt(s1^2 / n + s2^2 / m)

# Confidence intervals
alpha <- 0.05
t_quantile <- qt(1 - alpha / 2, n + m - 2)
conf_int <- c("lower bound" = Difference_mean - t_quantile * SE,
              "upper bound" = Difference_mean + t_quantile * SE)

# Displaying confidence intervals
conf_int
```

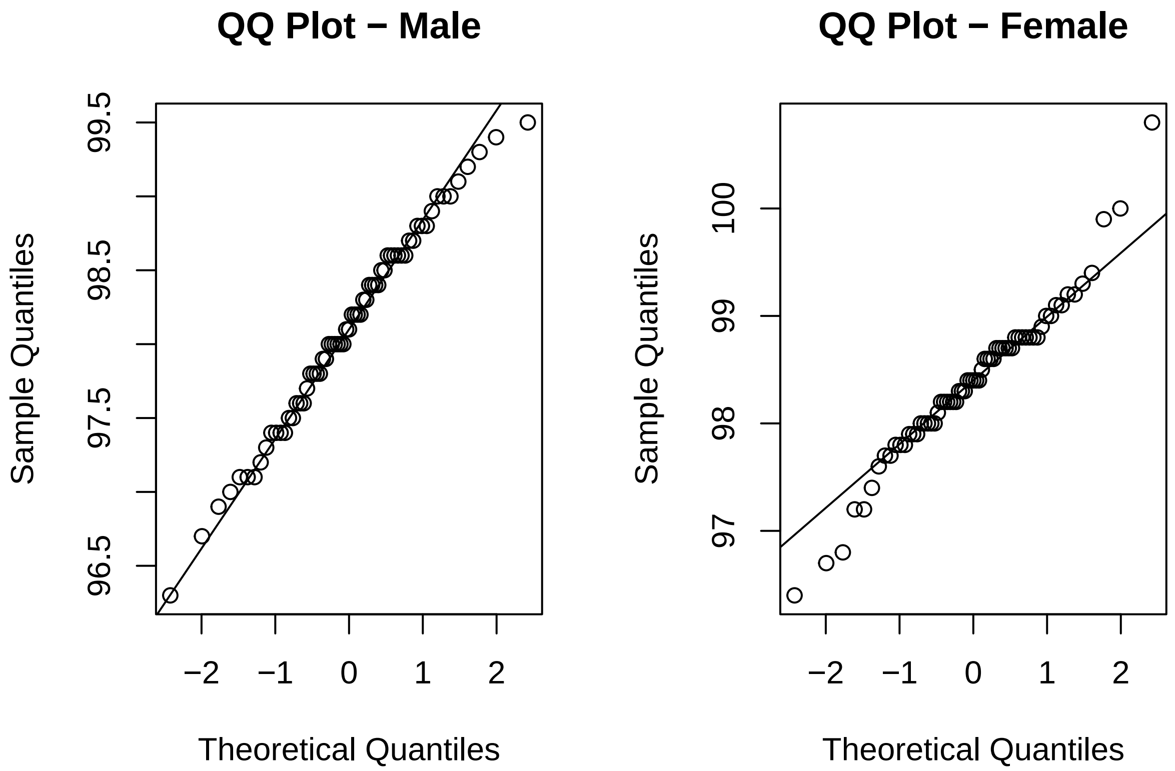
```
## lower bound upper bound
## -0.53963938 -0.03882216
```

```
t_test_result <- t.test(maletemp, femaletemp, var.equal = TRUE)
t_test_result
```

```
##
##  Two Sample t-test
##
## data:  maletemp and femaletemp
## t = -2.2854, df = 128, p-value = 0.02393
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.53963938 -0.03882216
## sample estimates:
## mean of x mean of y
##  98.10462  98.39385
```

At 95% confidence level, the difference is significant and the assumption of normality is tested using the QQ-plots:

```
#QQ plots:
par(mfrow=c(1,2))
qqnorm(maletemp, main="QQ Plot - Male")
qqline(maletemp)
qqnorm(femaletemp, main="QQ Plot - Female")
qqline(femaletemp)
```



Additionally, we test without the assumption of same variances for the two samples:

```
t_test_result <- t.test(maletemp, femaletemp, var.equal = FALSE)
t_test_result
```

```
##
##  Welch Two Sample t-test
##
## data:  maletemp and femaletemp
## t = -2.2854, df = 127.51, p-value = 0.02394
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.53964856 -0.03881298
## sample estimates:
## mean of x mean of y
##  98.10462  98.39385
```

We see that even without the assumption of equal variances, the result is significant and not very different.

b)

Now we test for the differences in heart rates for the groups male and female

```
malerate<-bodytemp$rate[bodytemp$gender==1]
femalerate<-bodytemp$rate[bodytemp$gender==2]

# Calculating mean and standard deviation for male and female temperatures
x1 <- mean(malerate)
s1 <- sd(malerate)
x2 <- mean(femalerate)
s2 <- sd(femalerate)

# Determining sample sizes
m <- length(malerate)
n <- length(femalerate)

# Calculating the difference in means and its standard deviation
Delta <- x1 - x2
SE <- sqrt(s1^2 / m + s2^2 / n)

# Confidence intervals
alpha <- 0.05
t_quantile <- qt(1 - alpha / 2, m + n - 2)
conf_int <- c("lower bound" = Delta - t_quantile * SE,"upper bound" = Delta + t_quantile * SE)
conf_int
```

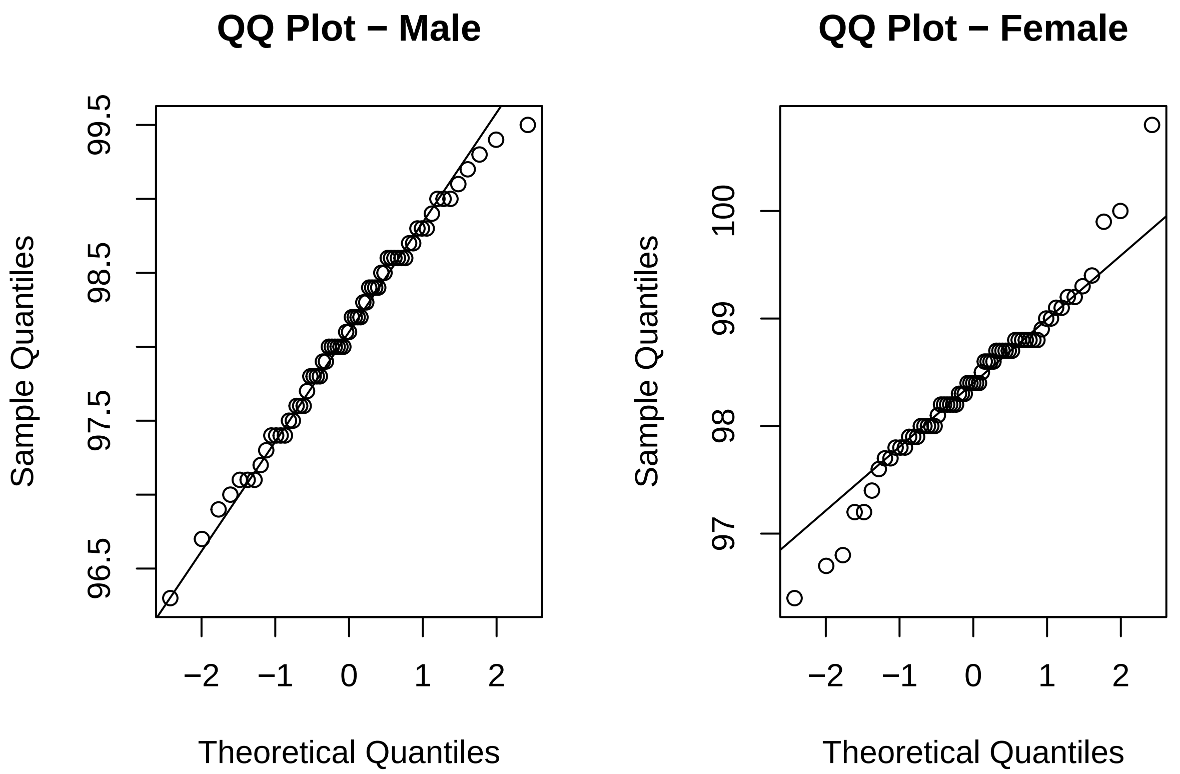
```
## lower bound upper bound
##   -3.241461    1.672230
```

```
t.test(malerate,femalerate,var.equal = TRUE)
```

```
##
```

```
##  Two Sample t-test
##
## data:  malerate and femalerate
## t = -0.63191, df = 128, p-value = 0.5286
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.241461  1.672230
## sample estimates:
## mean of x mean of y
##  73.36923  74.15385
```

```
#QQ plots:
par(mfrow=c(1,2))
qqnorm(maletemp, main="QQ Plot - Male")
qqline(maletemp)
qqnorm(femaletemp, main="QQ Plot - Female")
qqline(femaletemp)
```



Additionally, we test without the assumption of same variances for the two samples:

```
t.test(malerate,femalerate,var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
```

```
## data:  malerate and femalerate
## t = -0.63191, df = 116.7, p-value = 0.5287
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.243732  1.674501
## sample estimates:
## mean of x mean of y
##  73.36923  74.15385
```

c)

Recall from part a), we already have the results for parametric t-test:

```
# Confirming with t.test function in R
t_test_result <- t.test(maletemp, femaletemp, var.equal = TRUE)
t_test_result
```

```
##
##  Two Sample t-test
##
## data:  maletemp and femaletemp
## t = -2.2854, df = 128, p-value = 0.02393
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.53963938 -0.03882216
## sample estimates:
## mean of x mean of y
##  98.10462  98.39385
```

For the non-parametric test we perform a Wilcoxon rank sum test.

```
wilcox.test(maletemp,femaletemp)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  maletemp and femaletemp
## W = 1637, p-value = 0.02676
## alternative hypothesis: true location shift is not equal to 0
```

In both the parametric and non-parametric tests, we find that the null hypothesis is rejected at the 5% significance level. This consistent outcome suggests that the average body temperature is higher among women than men at 95% confidence level.

Now, we repeat the procedure for the differences in heart rates among genders:

```
# t-test for heart rates
t_test_result <- t.test(malerate, femalerate, var.equal = TRUE)
t_test_result
```

```
##
##  Two Sample t-test
##
## data:  malerate and femalerate
## t = -0.63191, df = 128, p-value = 0.5286
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.241461  1.672230
## sample estimates:
## mean of x mean of y
##  73.36923  74.15385
```

```
# Wilcoxon rank sum test for heart rates
wilcox_test_result <- wilcox.test(malerate, femalerate)
wilcox_test_result
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  malerate and femalerate
## W = 1927.5, p-value = 0.3898
## alternative hypothesis: true location shift is not equal to 0
```

Regarding heart rates, the results from the t-test indicate that the null hypothesis of equal mean heart rates between men and women cannot be rejected (p-value 0.5286). Similarly, the non-parametric Wilcoxon rank sum test also yields a high p-value (0.3898), supporting the conclusion that the mean heart rate does not significantly differ between genders.

Exercise 114

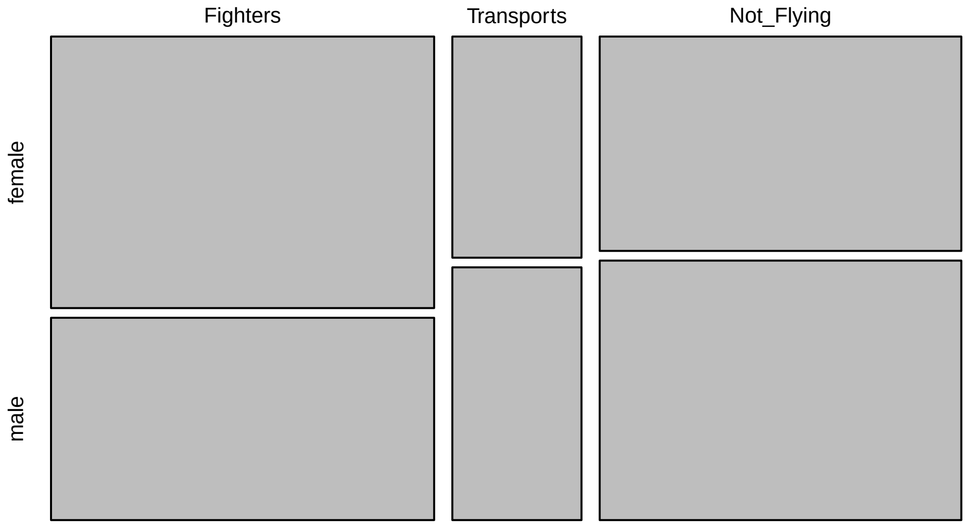
We create a table as given in the exercise:

```
data <- matrix(c(51, 14, 38, 38, 16, 46), 3, byrow = FALSE, dimnames
               = list(c("Fighters", "Transports", "Not_Flying"), c("female", "male")))
data
```

```
##            female male
## Fighters       51   38
## Transports     14   16
## Not_Flying     38   46
```

We can also visualize the data in the mosaic plot:

```
mosaicplot(data, main = "")
```



First we perform the Fisher's exact test to test differences between the groups.

```
#Fisher's exact test:
fisher.test(data)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  data
## p-value = 0.2594
## alternative hypothesis: two.sided
```

In Fisher's exact test, we did not find enough evidence to reject the null hypothesis that the true odds ratio equals 1, even at a significance level of 10% (p-value 0.2594). This suggests that there are no significant differences between the groups.

Then, we perform the chi-square test.

```
#Fisher's exact test:
chisq.test(data)
```

```
##
##  Pearson's Chi-squared test
##
## data:  data
## X-squared = 2.7504, df = 2, p-value = 0.2528
```

The Chi-Squared test did not detect significant differences between the groups, as it failed to reject the null hypothesis that the distribution of cell counts in the contingency table follows the expected pattern based on the row and column totals, even at a 10% significance level (p-value 0.2528).

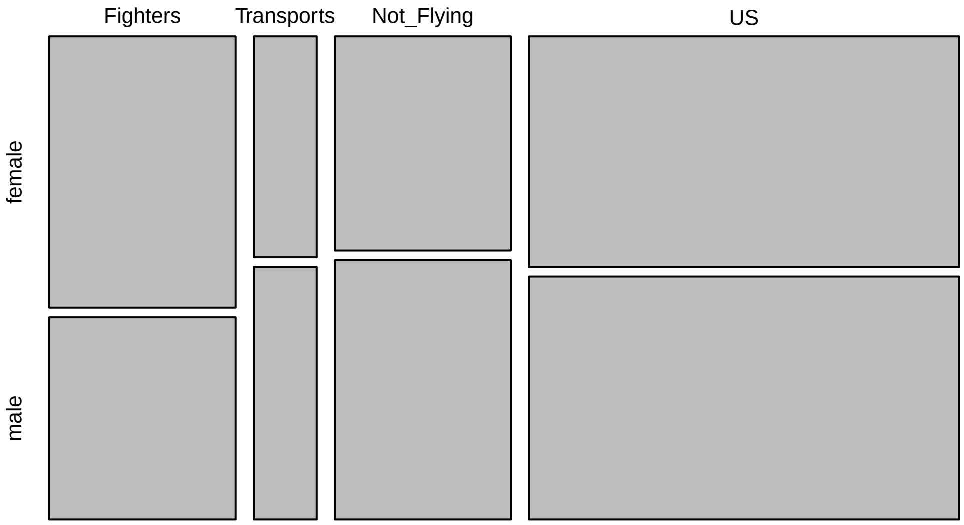
To investigate whether the sex ratios in our dataset match those of the entire US population in 1950 (105.37 males to 100 females), we conducted the following:

We perform the Fisher's exact test and the Chi-Squared test with an additional row in the contingency table representing the population's sex ratio. This allowed us to test for differences between any of the three groups and the population.

```
#First analysis:
dataExtended<-rbind(data,c(100,105.37))

rownames(dataExtended)<-c("Fighters","Transports","Not_Flying","US")
dataExtended
```

```
##              female   male
## Fighters         51  38.00
## Transports       14  16.00
## Not_Flying       38  46.00
## US              100 105.37
```

```
mosaicplot(dataExtended, main = "")
```



```
fisher.test(dataExtended)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  dataExtended
## p-value = 0.3772
## alternative hypothesis: two.sided
```

```
chisq.test(dataExtended)
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataExtended
## X-squared = 3.0885, df = 3, p-value = 0.3782
```

There are no significant differences between groups.

Then, we compared the combined sample against the population ratio using the same tests. This enabled us to assess whether the whole dataset corresponds to the observed sex ratio in the United States during 1950.

```
#Second analysis:
combinedSample<-rbind(colSums(data),c(10000,10537)) # add row
combinedSample
```

```
##      female  male
## [1,]    103   100
## [2,]  10000 10537
```

```
fisher.test(combinedSample)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  combinedSample
## p-value = 0.5731
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.8149901 1.4457525
## sample estimates:
## odds ratio
##   1.085288
```

```
chisq.test(combinedSample)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  combinedSample
## X-squared = 0.25998, df = 1, p-value = 0.6101
```

In both cases, the p-values are relatively high, indicating that we fail to reject the null hypothesis (even at 10% significance). Therefore, we conclude that there are no significant differences between the categorical samples and the data is consistent with the sex ratio in the US.

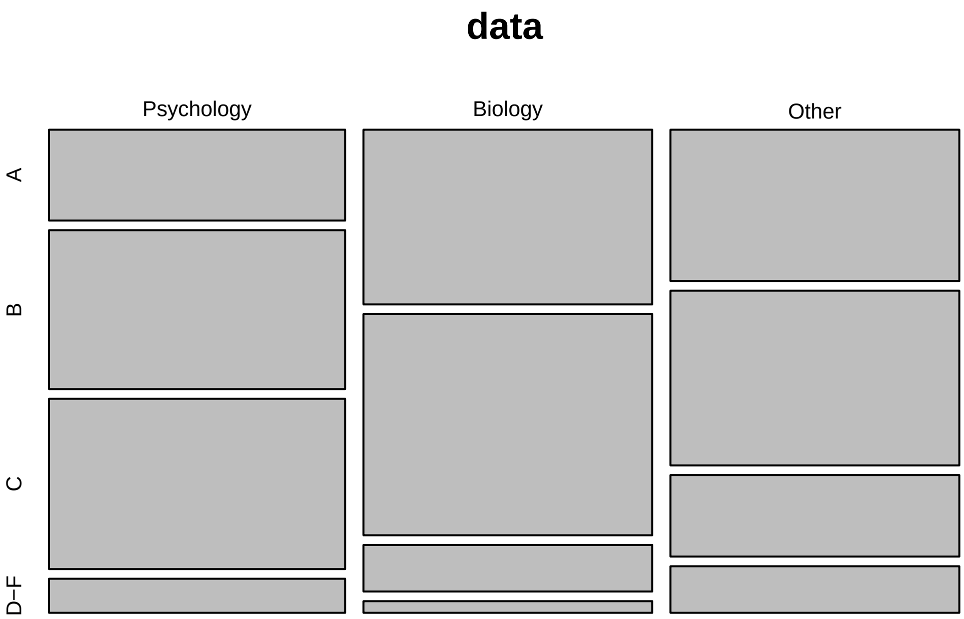
Exercise 115

```
data <- matrix(c(8,15,13,14,19,15,15,4,7,3,1,4), 3, 4, dimnames =
                 list(c("Psychology", "Biology", "Other"), c("A", "B", "C", "D-F")))

data
```

```
##               A  B  C D-F
## Psychology  8 14 15   3
## Biology    15 19  4   1
## Other      13 15  7   4
```

We visualize the data in a mosaic plot:

```
mosaicplot(data)
```



To investigate whether there are any relationships between grade and major, we perform a chi-squared test of independence.

```r
chisq.test(data)
```

```
## Warning in chisq.test(data): Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  data
## X-squared = 12.183, df = 6, p-value = 0.058
```

Considering the warning message, and looking into the documentation of chisq.test() function , the parameter `simlate.p.value` is a logical indicating whether to compute p-values by Monte Carlo simulation. Therefore we run it with simulation to trust the p-value estimate.

```r
chisq.test(data,simulate.p.value = TRUE)
```

```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000
##  replicates)
##
## data:  data
## X-squared = 12.183, df = NA, p-value = 0.05247
```

Resulting in around 0.06 p-value, we can only reject the null hypothesis of independence at 10% significance level.

Since having a small sample size in one category can be detrimental for a chi-square test we also strengthen our inference by performing Fisher's exact test:

```r
fisher.test(data)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  data
## p-value = 0.05858
## alternative hypothesis: two.sided
```

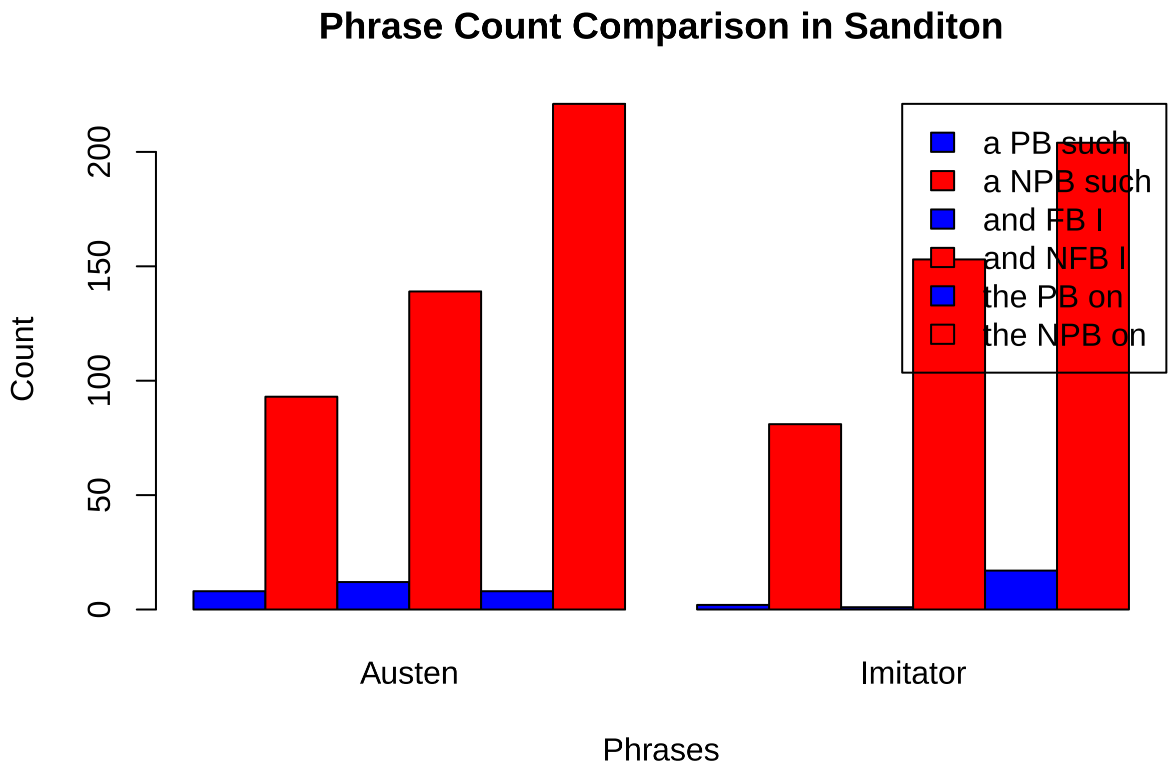
With p-value of 0.058 we do not have enough evidence to reject the null hypothesis of independence between major and grade.

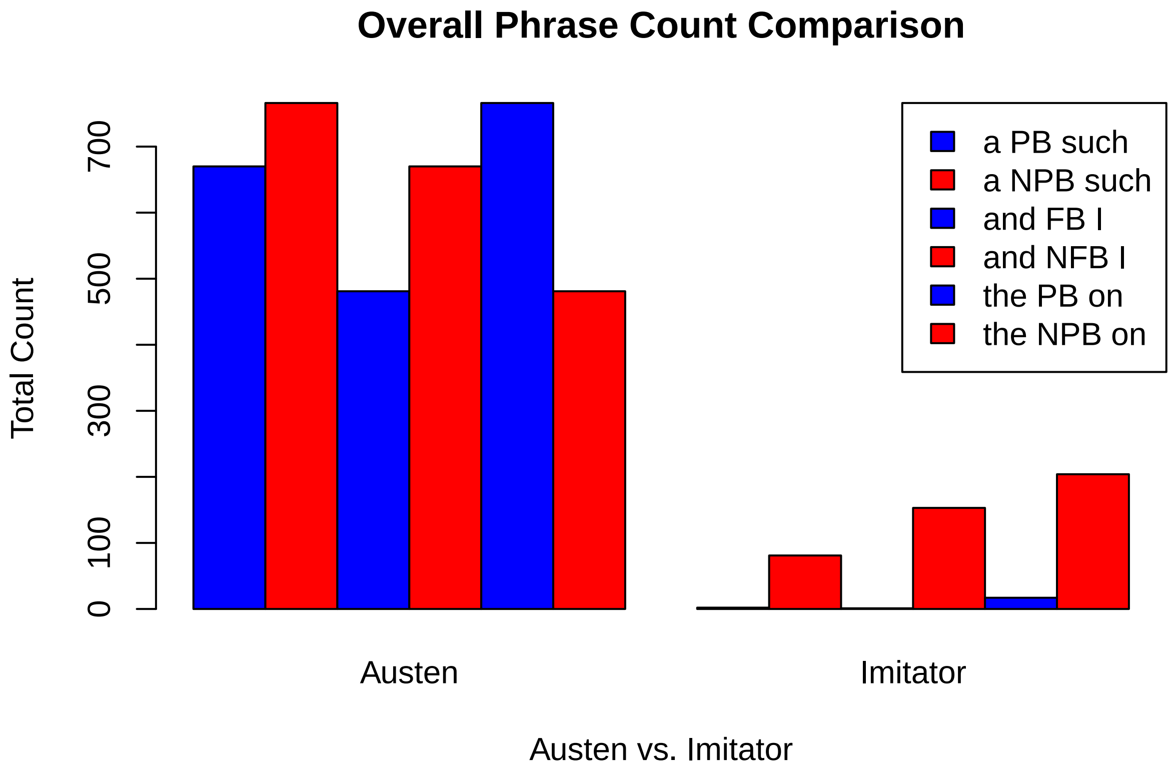
Exercise 116

In this analysis, we explore Jane Austen's writing style consistency across different works and compare it to an imitator's style in the novels "Sanditon I" and "Sanditon II." We employ a chi-squared test to statistically evaluate the frequency of certain phrases.

We constructed a matrix *df_lit* with the counts of selected phrases in four of Austen's works. The rows represent the phrases, and the columns represent the novels "Sense and Sensibility," "Emma," "Sanditon I" (authored by Austen), and "Sanditon II" (completed by an imitator).

```r
# Creating the data frame with literary analysis data
df_lit <- matrix(
  c(14, 16, 8, 2,
    133, 180, 93, 81,
    12, 14, 12, 1,
    241, 285, 139, 153,
    11, 6, 8, 17,
    259, 265, 221, 204),
  nrow = 6, ncol = 4, byrow = TRUE
)
rownames(df_lit) <- c("a PB such", "a NPB such", "and FB I", "and NFB I", "the PB on", "the NPB on")
colnames(df_lit) <- c("Sense and Sensibility", "Emma", "Sanditon I", "Sanditon II")
```





```
# Chi-squared tests to analyze differences
chi_squared_test_sanditon <- chisq.test(au_vs_imi_san)
```

```
## Warning in chisq.test(au_vs_imi_san): Chi-squared approximation may be
## incorrect
```

```
chi_squared_test_sanditon
```

```
##
##  Pearson's Chi-squared test
##
## data:  au_vs_imi_san
## X-squared = 17.774, df = 5, p-value = 0.003244
```

The comparison between Austen's "Sanditon I" and the imitator's "Sanditon II" yielded a chi-squared value of 17.774 with a p-value of 0.003244.

We then performed a chi-squared test to assess the consistency within Austen's works and to compare Austen's original work against the imitator:

```
# Statistical test to evaluate Jane Austen's consistency across her works (excluding the imitator's)
chi_squared_test_austen <- chisq.test(df_lit[, 1:3])
chi_squared_test_austen
```

```
##
##  Pearson's Chi-squared test
##
## data:  df_lit[, 1:3]
## X-squared = 23.287, df = 10, p-value = 0.009735
```

The test for consistency within Austen's works produced a chi-squared value of 23.287 with a p-value of 0.009735, leading us to reject the null hypothesis that there is no difference in the usage of phrases across the three Austen novels.

```
chi_squared_test_overall <- chisq.test(au_vs_imi)
print(chi_squared_test_overall)
```

```
##
##  Pearson's Chi-squared test
##
## data:  au_vs_imi
## X-squared = 508.24, df = 5, p-value < 2.2e-16
```

In comparing the sum of phrases from all Austen's works against the imitator, we obtained a chi-squared value of 508.24, which is highly significant ($p < 2.2e\text{-}16$).

Our findings lead us to reject the null hypothesis at the 0.05 significance level, concluding that the imitator did not capture the nuances of Austen's stylistic preferences. The statistical evidence suggests a significant difference in the use of selected phrases, indicating an inconsistency with Austen's authentic style.

Exercise 117

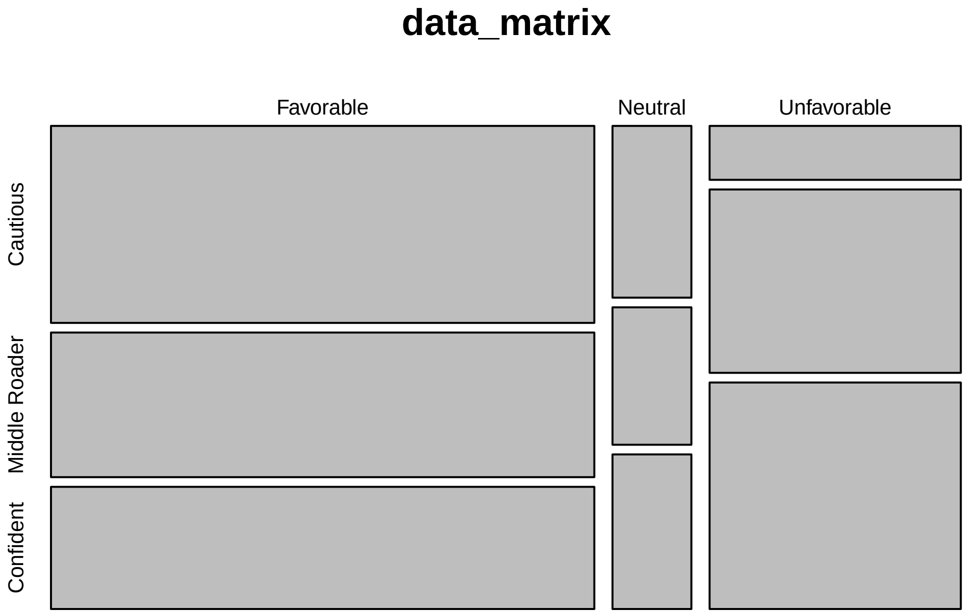
For this exercise, in order to determine if there is a relationship between personality and attitude towards small cars, we will employ the chi-square statistical test given that we have categorical data. Through this test, we can determine if the observed data of attitude (favorable, neutral, and unfavorable) is independent of the personality types (cautious conservative, middle-of-the-roader, confident explorer), or if there is a relationship between the the two variables.

```
data_matrix<- matrix(c(79,58,49,10,8,9,10,34,42), nrow=3, byrow=T)
rownames(data_matrix)<-c("Favorable","Neutral","Unfavorable")
colnames(data_matrix)<-c("Cautious","Middle Roader", "Confident")

chi_square<- chisq.test(data_matrix)
chi_square
```

```
##
##  Pearson's Chi-squared test
##
## data:  data_matrix
## X-squared = 27.289, df = 4, p-value = 1.737e-05
```

From the results of the chi-squared test, we can reject the null hypothesis that there is no relationship between personality and attitude towards small cars. Hence, we have a significant relationship between the two variables. The test does not imply the nature of the relationship, it just suggests that the two categories are related. To assess the nature of the relationship, we can visualize the data through a mosaic plot.

```
mosaicplot(data_matrix)
```



Based on the differences of the column widths we can observe that the most common overall opinion is favorable attitude towards small cars, followed by unfavorable and neutral, respectively. The rectangle in the "Favorable" column and "Cautious" row is quite large, suggesting that a relatively high number of cautious individuals have a favorable opinion of small cars. The "Unfavorable" rectangle is quite large for Confident/Explorer individuals, suggesting that a substantial number of them have an unfavorable opinion of small cars. This is disproportionate to their favorable and neutral opinions, which may indicate a tendency for the Confident/Explorer personality to be critical of small cars.'