02441 Applied Statistics and Statistical Software

Exercise 1B - Bodyfat

The dataset bodyfat contains measurements of bodyfat for a number of men and women

Variable name	Description
gender	gender (male/femal)
fatpct	Measurement of fat percentage

1. Can the data in each group (gender) be assumed to be normally distributed

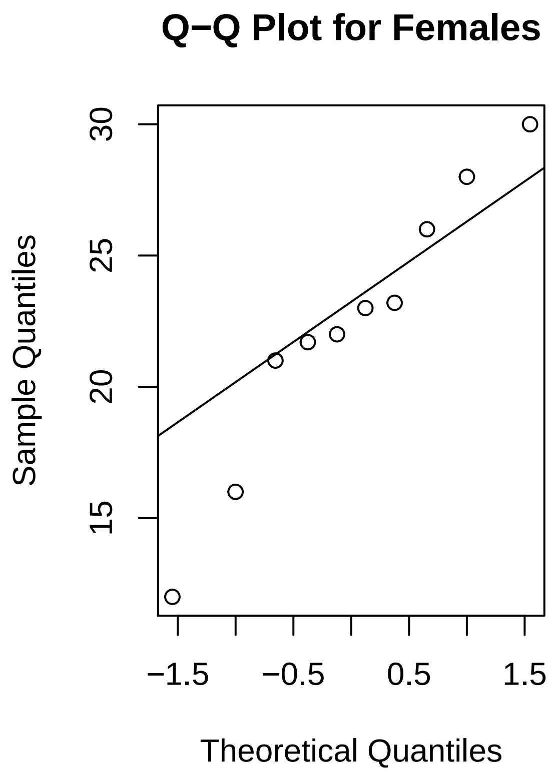
Start by loading the data (make sure datafile is present in working directory)

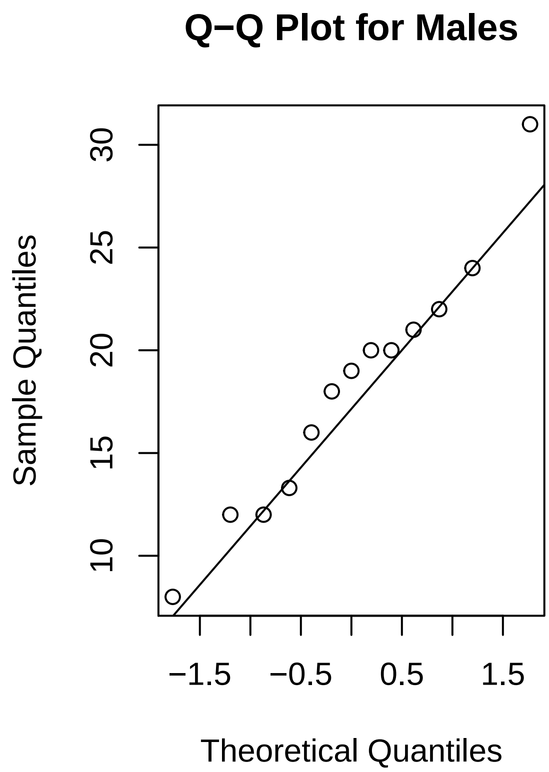
```
bodyfat <- read.table("bodyfat.txt", header=TRUE)
```

Use both visual aids as well as normality tests to check the normality of fat-percentages on each gender.

```
# Split data according to gender
bodyfat2 <- split(bodyfat$fatpct, bodyfat$gender)
names(bodyfat2) <- c("female","male")

# Check for normality (both samples) with Q-Q Plots
par(mfrow=c(1,2))
qqnorm(bodyfat2$female, main="Q-Q Plot for Females")
qqline(bodyfat2$female)
qqnorm(bodyfat2$male, main="Q-Q Plot for Males")
qqline(bodyfat2$male)
```

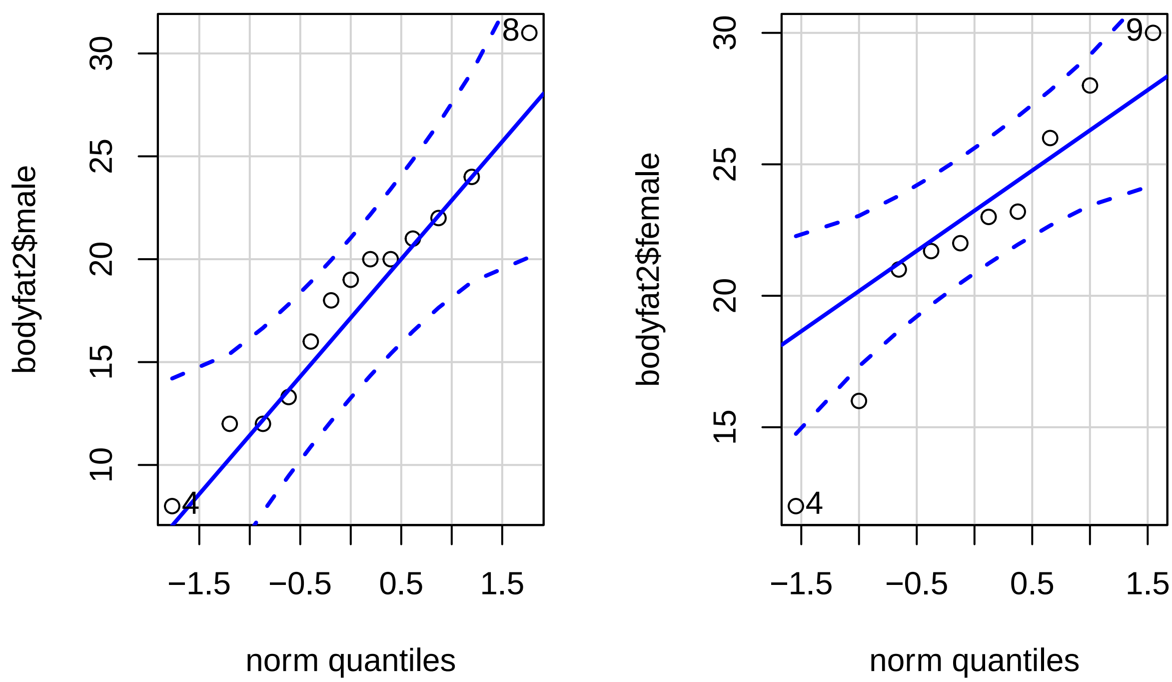





```r
# Plot with confidence intervals
qqPlot(bodyfat2$male)
```

```
## [1] 8 4
```

```r
qqPlot(bodyfat2$female)
```




```
## [1] 4 9
```

```r
# Normality test (Shapiro)
shapiro.test(bodyfat2$male)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  bodyfat2$male
## W = 0.97067, p-value = 0.9021
```

```r
shapiro.test(bodyfat2$female)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  bodyfat2$female
## W = 0.95191, p-value = 0.6911
```

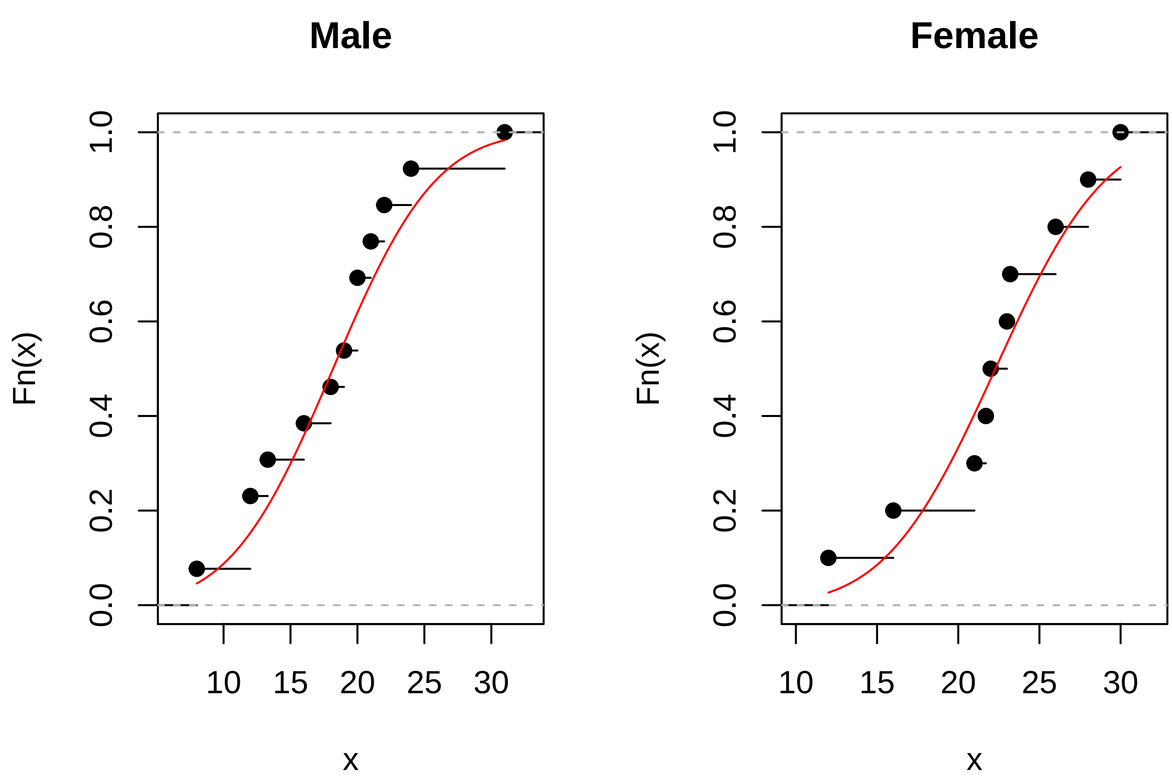
```r
# Visual normality test with ecdf and cdf

# Male parameters
xSeqMale <- seq(min(bodyfat2$male), max(bodyfat2$male), by=0.1)
muMale <- mean(bodyfat2$male)
```

```r
sMale <- sd(bodyfat2$male)

# Female parameters
xSeqFemale <- seq(min(bodyfat2$female), max(bodyfat2$female), by=0.1)
muFemale <- mean(bodyfat2$female)
sFemale <- sd(bodyfat2$female)

# Plot ecdf and cdf
par(mfrow=c(1,2))
plot(ecdf(bodyfat2$male), main="Male")
lines(xSeqMale, pnorm(xSeqMale, muMale, sMale), col="red")
plot(ecdf(bodyfat2$female), main="Female")
lines(xSeqFemale, pnorm(xSeqFemale, muFemale, sFemale), col="red")
```



Both male and female are normally distributed since p-values are $> \alpha = 0.05$, thus we cannot reject the null hypothesis.

2. Is there a difference in the percentage of body fat for men and women? Perform a t-test

```r
t.test(bodyfat2$male, bodyfat2$female)
```

```
##
```

```
##  Welch Two Sample t-test
##
## data:  bodyfat2$male and bodyfat2$female
## t = -1.7336, df = 20.539, p-value = 0.09798
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -9.0538774  0.8277235
## sample estimates:
## mean of x mean of y
##  18.17692  22.29000
```

There's no difference in the percentage of body fat for men and women since the p-value is $> \alpha = 0.05$, thus we cannot reject the null hypothesis.

3. Is there a difference in the percentage of body fat for men and women? Perform a non-parametric test

```r
wilcox.test(bodyfat2$female, bodyfat2$male, conf.int = TRUE)
```

```
## Warning in wilcox.test.default(bodyfat2$female, bodyfat2$male, conf.int = TRUE):
## cannot compute exact p-value with ties
```

```
## Warning in wilcox.test.default(bodyfat2$female, bodyfat2$male, conf.int = TRUE):
## cannot compute exact confidence intervals with ties
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  bodyfat2$female and bodyfat2$male
## W = 94.5, p-value = 0.07153
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  -0.2999953  9.7000328
## sample estimates:
## difference in location
##               4.000024
```

There's no difference in the percentage of body fat for men and women since the p-value is $> \alpha = 0.05$, thus we cannot reject the null hypothesis.