

STAT 210
Applied Statistics and Data Analysis
Problem List 4 - Solution
(Due on Week 5)

Fall 2025

Problem 1

Researchers in an oil company want to test a new fuel additive that is supposed to improve the average miles per gallon (mpg). The additive is coded A1. To test this, they conduct an experiment measuring the fuel efficiency of 20 identical cars driven under the same conditions. Previous experiments show that the vehicle gets an average of 20.5 mpg without the additive. The experiment results are stored in the mpg1 variable in the file 25Fp14.csv.

Read the data and store it in a data frame. Explore the structure

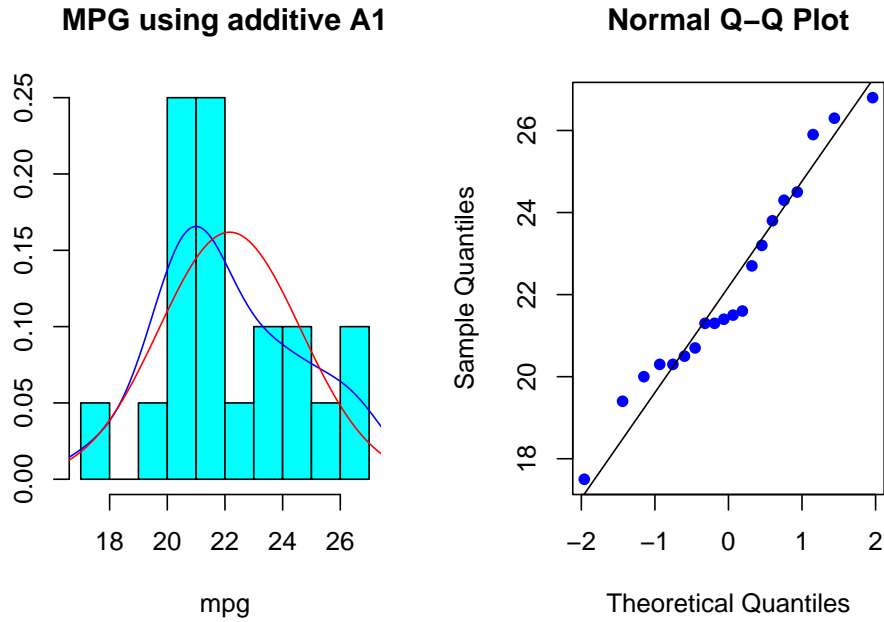
```
pl4df <- read.csv('25Fp14.csv')
str(pl4df)
```

```
## 'data.frame':    20 obs. of  3 variables:
## $ mpg1: num  19.4 23.8 26.3 17.5 24.3 24.5 21.3 21.4 21.3 20.3 ...
## $ mpg2: num  18.6 22.4 25 15.6 22.7 22.2 21 19.5 20.4 18.5 ...
## $ mpg3: num  20.7 26.4 15.9 19.7 23.3 24.7 25.4 20.3 21.2 16.6 ...
```

- (a) Divide the plotting window into two regions. On the left, plot a histogram of the data. Add a graph of the estimated density in blue and a graph of the normal density with parameters estimated from the sample, in red. Use appropriate names for the axis labels and the title. On the right, do a quantile plot and add a reference line. Comment on what you observe in the plots. Do you think that the assumption of normality for the data is valid?

Solution:

```
par(mfrow = c(1,2))
library(MASS)
truehist(pl4df$mpg1, h = 1, prob = T, xlab = 'mpg', main = 'MPG using additive A1')
lines(density(pl4df$mpg1), col = 'blue')
curve(dnorm(x, mean(pl4df$mpg1), sd(pl4df$mpg1)), 16, 28, col = 'red', add = T)
qqnorm(pl4df$mpg1, pch = 16, col = 'blue')
qqline(pl4df$mpg1)
```



```
par(mfrow = c(1,1))
```

The histogram does not look close to a normal density, but the estimated density (blue) is close to the normal density with estimated parameters (red). The differences we observe could be due to sampling variability, since the sample size is small. The quantile plot is reasonable, considering that we only have twenty observations. In conclusion, I don't see evidence in the plots against the assumption of normality.

- (b) Assuming that the variable `mpg1` values follow a Gaussian distribution, write down a formula for the upper one-sided confidence interval for the mean at level $(100 - \alpha)\%$. This interval is bounded below but unbounded above. Calculate this confidence interval for the mean for the case $\alpha = 0.02$, check whether the reference value of 20.5 falls inside or outside and give an interpretation.

Solution:

Define $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ and let μ be the population mean. We saw in the videos that when the population is normal but we have to estimate the standard deviation from the sample, the sampling distribution is

$$\frac{\bar{X}_n - \mu}{s_n/\sqrt{n}} \sim t_{n-1}$$

where s_n is the sample standard deviation, and n is the sample size, $n = 20$ in this case. Define $t_{n,\alpha}$ to be the quantile for the t distribution with n degrees of freedom, i.e., if T_n is a random variable with t_n distribution,

$$P(T_n \leq t_{n,\alpha}) = \alpha.$$

Using this, we have that

$$P\left(\frac{\hat{\mu}_n - \mu}{s_n/\sqrt{n}} < t_{n-1,1-\alpha}\right) = 1 - \alpha$$

Multiplying by s_n/\sqrt{n} inside the probability, we get

$$P\left(\hat{\mu}_n - \mu < \frac{s_n t_{n-1,1-\alpha}}{\sqrt{n}}\right) = 1 - \alpha$$

And from this, we get that

$$P\left(\hat{\mu}_n - \frac{s_n t_{n-1,1-\alpha}}{\sqrt{n}} < \mu\right) = 1 - \alpha.$$

This says that the $(1 - \alpha) \cdot 100\%$ upper one-sided confidence interval is given by

$$\left(\hat{\mu}_n - \frac{s_n t_{n-1, 1-\alpha}}{\sqrt{n}}, \infty\right)$$

We now use this formula to get the interval:

```
(mn1 <- mean(pl4df$mpg1))

## [1] 22.165

(sd1 <- sd(pl4df$mpg1))

## [1] 2.46433

mn1 - (sd1*qt(0.98, df=19))/sqrt(20)

## [1] 20.95012
```

The interval is $(20.94997, \infty)$. We see that the reference value is not included in this interval. This means that, with a confidence of 98%, the average mpg using A1 is higher than 20.5.

- (c) What parametric test would be adequate for testing whether the additive A1 increases the mpg for the car? State clearly what hypotheses you are testing and which assumptions are needed for the test. Explain why you think they are satisfied. Give a formula for the test statistic and calculate its value. Describe the sampling distribution and explicitly identify the type I and type II errors. Carry out this test and discuss the results.

Solution:

We want to test the following hypothesis

$$H_0 : \mu = 20.5 \quad vs. \quad H_1 : \mu > 20.5$$

If we can assume that the distribution of the miles per gallon is normal, we can use the one-sample t -test. The quantile plot above seems to support this assumption, but we do a Shapiro-Wilk test to confirm this

```
shapiro.test(pl4df$mpg1)

##
## Shapiro-Wilk normality test
##
## data:  pl4df$mpg1
## W = 0.95342, p-value = 0.422
```

The p -value is large and we do not have evidence to reject the null hypothesis of normality.

The test statistic is given by

$$\hat{t} = \frac{\hat{\mu}_n - \mu}{s_n / \sqrt{n}}$$

which has a t_{19} (sampling) distribution. The value for this statistic is

```
(tstat <- (mean(pl4df$mpg1) - 20.5)/(sd(pl4df$mpg1)/sqrt(20)))

## [1] 3.021554
```

A type I error would be to conclude that the additive improves the mileage of the cars when, in fact, it does not. A type II error would be to conclude that the additive does not improve the mileage when, in fact, it does.

To do the test in R, we use the `t.test` function:

```
t.test(pl4df$mpg1, mu = 20.5, conf.level = .98, alternative = 'greater')
```

```
##
## One Sample t-test
##
## data: pl4df$mpg1
## t = 3.0216, df = 19, p-value = 0.003509
## alternative hypothesis: true mean is greater than 20.5
## 98 percent confidence interval:
## 20.95012      Inf
## sample estimates:
## mean of x
## 22.165
```

We see that the value for the test statistic and for the confidence interval coincide with our previous results. The p -value is below the 2% level, and we reject the null hypothesis: at the 98% level, the additive improves the mileage of the cars.

- (d) What non-parametric tests will be adequate for the problem in (c)? What assumptions are needed, and why do you think they are satisfied? Perform this test, discuss the results, and compare them with your previous results.

Solution:

We can use Wilcoxon's test, which assumes that the distribution is continuous and symmetric with respect to the mean value. Since we previously accepted that the distribution was normal, this assumptions seems to be valid.

For the test in (c) we have

```
wilcox.test(pl4df$mpg1, mu = 20.5, conf.level = .98, alternative = 'greater')
```

```
## Warning in wilcox.test.default(pl4df$mpg1, mu = 20.5, conf.level = 0.98, :
## cannot compute exact p-value with ties

## Warning in wilcox.test.default(pl4df$mpg1, mu = 20.5, conf.level = 0.98, :
## cannot compute exact p-value with zeroes

##
## Wilcoxon signed rank test with continuity correction
##
## data: pl4df$mpg1
## V = 159.5, p-value = 0.004982
## alternative hypothesis: true location is greater than 20.5
```

The p -value is of the same order of magnitude as in the t -test and we reach the same conclusion: we reject the null hypothesis and conclude that the additive improves the mileage of the cars.

Problem 2

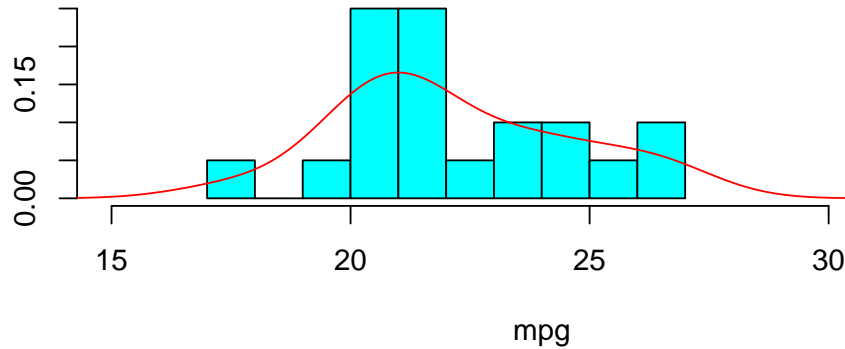
The oil company has also developed a second additive for improving fuel efficiency, coded A2. This new additive is tested using the same cars as in the previous experiment, and the results are stored in the variable `mpg2`, in the same order as for the first experiment.

- (a) Plot histograms for `mpg1` and `mpg2`, one above the other, using the same scale on the x -axis and bins of width 1. On each histogram, add a curve for the corresponding estimated density. Comment on what you observe.

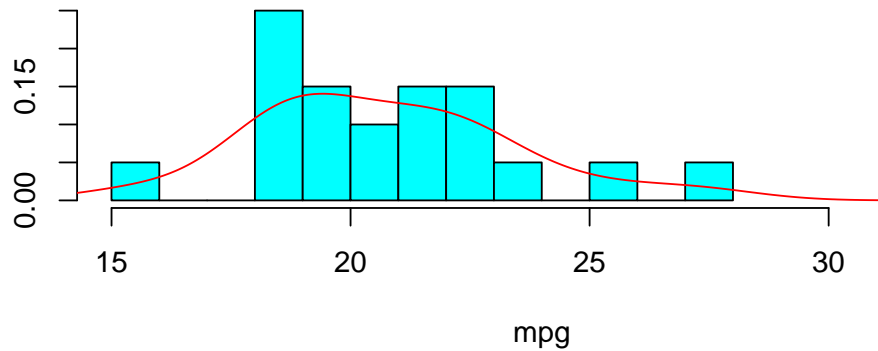
Solution:

```
old.par = par(no.readonly = TRUE)
par(mfrow = c(2,1)); #par(mar = c(3,4,3,2)+0.1)
truehist(pl4df$mpg1, h = 1, xlim = c(15, 33), main = 'A1', xlab = 'mpg')
lines(density(pl4df$mpg1), col = 'red')
truehist(pl4df$mpg2, h = 1, xlim = c(15, 33), main = 'A2', xlab = 'mpg')
lines(density(pl4df$mpg2), col = 'red')
```

A1



A2



```
par(mfrow = old.par)
```

Looking at the plots, the differences are not striking. Overall, additive A2 seems to produce smaller values for mpg than additive A1 and to have more variability as well. Observe that both the smallest and largest values come from using A2.

- (b) What parametric test would be adequate for comparing the average performance of the two additives? State clearly what hypothesis you are testing and which assumptions are needed for the test, and explain why you think they are satisfied. Identify the type I and type II errors. Carry out this test and discuss the results.

Solution:

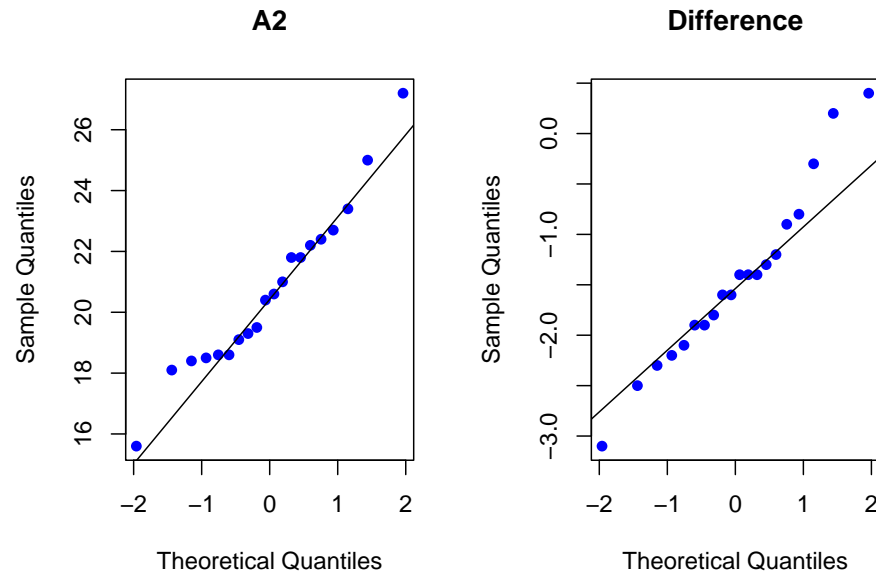
We want to test

$$H_0 : \mu_1 = \mu_2 \quad vs. \quad H_1 : \mu_1 \neq \mu_2$$

where μ_i is the average mpg using additive A_i , $i = 1, 2$.

Assuming that the samples follow a normal distribution, we can use a two-sample t -test. Since the two additives are tested on the same cars, the measurements are not independent, and we must use a paired test. To verify the assumptions, we can either see whether `mpg2` follows a normal distribution (since we already dealt with `mpg1`) or we can look at the difference between the two. We do quantile plots for both cases:

```
par(mfrow = c(1,2))
qqnorm(pl4df$mpg2, pch = 16, col = 'blue', main = 'A2')
qqline(pl4df$mpg2)
qqnorm(pl4df$mpg2-pl4df$mpg1, pch = 16, col = 'blue', main = 'Difference')
qqline(pl4df$mpg2-pl4df$mpg1)
```



```
par(mfrow = c(1,1))
```

In both cases we see that the fit is reasonable. We check using the Shapiro-Wilk test:

```
shapiro.test(pl4df$mpg2)

##
##  Shapiro-Wilk normality test
##
## data:  pl4df$mpg2
## W = 0.96287, p-value = 0.6026
shapiro.test(pl4df$mpg2 - pl4df$mpg1)

##
##  Shapiro-Wilk normality test
##
## data:  pl4df$mpg2 - pl4df$mpg1
## W = 0.96428, p-value = 0.6324
```

In both cases the p -values are large, and we cannot reject the null hypothesis of normality.

A type I error would be to reject the hypothesis that the two additives have the same average mileage per gallon when this is true, and a type II error would be to accept that they are equivalent when they are not.

The test is

```
t.test(pl4df$mpg1, pl4df$mpg2, paired = T)

##
## Paired t-test
##
## data: pl4df$mpg1 and pl4df$mpg2
## t = 7.4771, df = 19, p-value = 4.506e-07
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  1.047709 1.862291
## sample estimates:
## mean difference
##      1.455
```

The p -value is very small and we reject the null hypothesis of equal means. Observe that if we do not do a paired test, we would reach the wrong conclusion:

```
t.test(pl4df$mpg1, pl4df$mpg2)

##
## Welch Two Sample t-test
##
## data: pl4df$mpg1 and pl4df$mpg2
## t = 1.7837, df = 37.713, p-value = 0.08252
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.1967255  3.1067255
## sample estimates:
## mean of x mean of y
##    22.165    20.710
```

- (c) A different car model that uses a motor with the same specifications is also used for another experiment, in which the fuel efficiency is measured using additive A2. In the experiment, 20 cars were used, and the results are stored in the variable `mpg3`. You want to determine whether the two car models have the same average mpg. What parametric test would be adequate for comparing the fuel efficiency of the two models? State clearly what hypothesis you are testing and which assumptions are needed for the test. Explain why you think they are satisfied. Carry out this test and discuss the results.

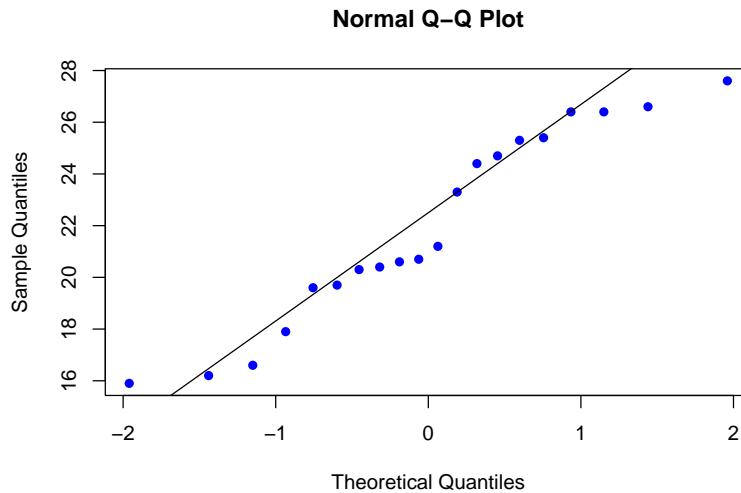
Solution:

In this case we want to test

$$\mu_a = \mu_b \quad \text{vs.} \quad \mu_a \neq \mu_b$$

where μ_a is the average miles per gallon with the first type of car while μ_b corresponds to the second type of car. Assuming that `mpg3` follows a normal distribution, we can use the t -test. To check this we do a quantile plot

```
qqnorm(pl4df$mpg3, pch = 16, col = 'blue')
qqline(pl4df$mpg3)
```



The plot does not look very convincing, and we do a Shapiro-Wilk test to check

```
shapiro.test(pl4df$mpg3)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  pl4df$mpg3
## W = 0.93132, p-value = 0.1637
```

The p -value is large and we do not reject the null hypothesis of normality. The test is

```
t.test(pl4df$mpg2, pl4df$mpg3)
```

```
##
##  Welch Two Sample t-test
##
## data:  pl4df$mpg2 and pl4df$mpg3
## t = -1.2131, df = 34.498, p-value = 0.2333
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.3429401  0.8429401
## sample estimates:
## mean of x mean of y
##    20.71    21.96
```

The p -value is large and we do not reject the null hypothesis of equal means.

- (d) What non-parametric tests will be adequate for the problems in (b) and (c)? What assumptions are needed, and why do you think they are satisfied? Perform these tests, discuss the results, and compare them with your previous results.

Solution:

In both cases we can use Wilcoxon's test, which assumes that the distribution is continuous and symmetric with respect to the mean value. Since we previously accepted that the distribution was normal, this assumptions seem to be valid.

For the test in (b) we have to use the `paired = T` option, since we want a paired test:

```
wilcox.test(pl4df$mpg1, pl4df$mpg2, conf.level = .98, paired = T)
```

```
## Warning in wilcox.test.default(pl4df$mpg1, pl4df$mpg2, conf.level = 0.98, :
```



```
## cannot compute exact p-value with ties
##
## Wilcoxon signed rank test with continuity correction
##
## data: pl4df$mpg1 and pl4df$mpg2
## V = 206, p-value = 0.0001752
## alternative hypothesis: true location shift is not equal to 0
```

The p value is small and, as before, we reject the null hypothesis of equal means.

For the test in (c) we have

```
wilcox.test(pl4df$mpg2, pl4df$mpg3, conf.level = .98)
```

```
## Warning in wilcox.test.default(pl4df$mpg2, pl4df$mpg3, conf.level = 0.98):
## cannot compute exact p-value with ties
##
## Wilcoxon rank sum test with continuity correction
##
## data: pl4df$mpg2 and pl4df$mpg3
## W = 158, p-value = 0.2615
## alternative hypothesis: true location shift is not equal to 0
```

Here the p -value is large and we do not reject the null hypothesis of equal means. This is the same conclusion as before.

Problem 3

- (a) A chocolate manufacturer claims that the medium size chocolate bars they produce have a caloric content of 190 calories. The variable `barA` in the file `calories.csv` has the result of calory measurements in 25 randomly chosen bars produced by the manufacturer. What parametric test would be adequate for testing whether the measurements are consistent with the manufacturer's claim? State clearly what hypotheses you are testing and which assumptions are needed for the test. Explain why they are satisfied. Describe the test statistic and calculate its value. Describe the sampling distribution, and explicitly identify the type I and type II errors. Carry out this test and discuss the results.

Solution:

```
data1 <- read.csv('calories.csv')
str(data1)

## 'data.frame': 25 obs. of 4 variables:
## $ barA: num 186 187 194 191 187 ...
## $ barB: num 186 187 194 191 187 ...
## $ barC: num 101 100 110 102 102 ...
## $ barD: num 102.2 98.5 106 104.7 101 ...

head(data1)

##   barA barB barC barD
## 1 186.1 185.5 100.6 102.2
## 2 187.4 186.8 100.5  98.5
## 3 194.2 193.8 109.9 106.0
## 4 191.2 190.7 101.7 104.7
## 5 187.0 187.0 102.0 101.0
```

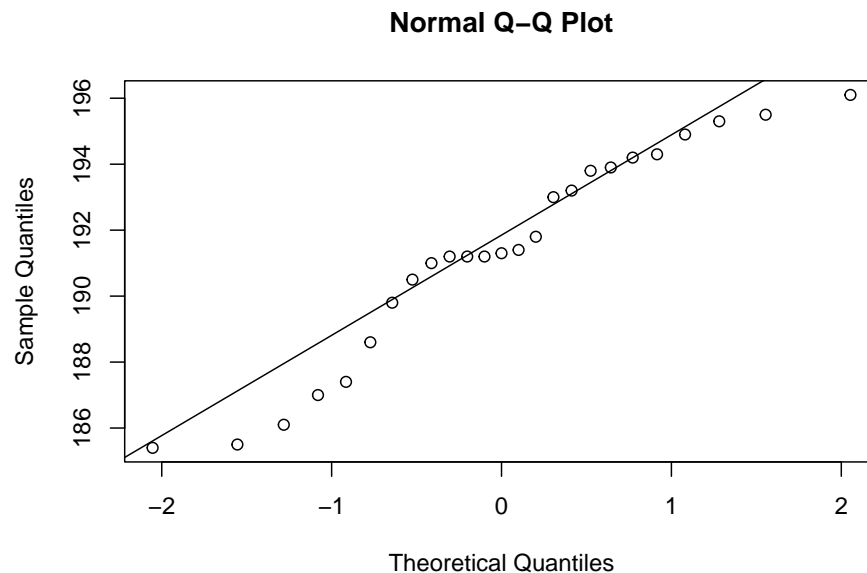
```
## 6 191.2 191.0 105.9 103.2
```

The parametric test to determine whether the average caloric content in the sample is consistent with the manufacturer's claim is the one-sample t-test. Let μ be the average caloric content of the bars produced by the manufacturer, the hypotheses we want to test are

$$H_0 : \mu = 190 \quad \text{vs.} \quad \mu \neq 190$$

The test assumes that either the population distribution is normal or that the sample size is large enough for the sampling distribution of the sample mean to be well approximated by the normal distribution due to the Central Limit Theorem. In this case, the sample size is not large, so we need to verify that the assumption of normality is reasonable. We do a normal quantile plot for the sample

```
qqnorm(data1$barA)
qqline(data1$barA)
```



The quantile plot does not look too good so we perform a Shapiro-Wilk test

```
shapiro.test(data1$barA)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data1$barA
## W = 0.93714, p-value = 0.1272
```

The test has a p -value above the significance level of 0.02, and we do not reject the null hypothesis of normality.

The other assumptions are that the observations are independent and have all the same distribution. Since the bars on which the measurements were made were chosen randomly, and since no important disruption on the production process is mentioned in the statement of the question, we may assume that these assumptions are satisfied.

The test statistic in this case is the standardized sample mean

$$\frac{\hat{\mu}_n - 190}{s_n / \sqrt{25}}$$

where s_n is the empirical standard deviation. This test statistic has a t distribution with $n - 1 = 24$ degrees of freedom. The value for the test statistic is

```
(tn <- (mean(data1$barA-190)/(sd(data1$barA)/sqrt(25))))
```

```
## [1] 2.09786
```

The type I error in this test would be to reject the null hypothesis that the average value of the caloric content of the bars is 190 when this is, in fact, true. The type II error would be not to reject the null hypothesis that the average value of the caloric content of the bars is 190 when this is false.

We now do this test:

```
t.test(data1$barA, mu = 190, conf.level = 0.98)
```

```
##
## One Sample t-test
##
## data: data1$barA
## t = 2.0979, df = 24, p-value = 0.04663
## alternative hypothesis: true mean is not equal to 190
## 98 percent confidence interval:
## 189.7474 192.9406
## sample estimates:
## mean of x
## 191.344
```

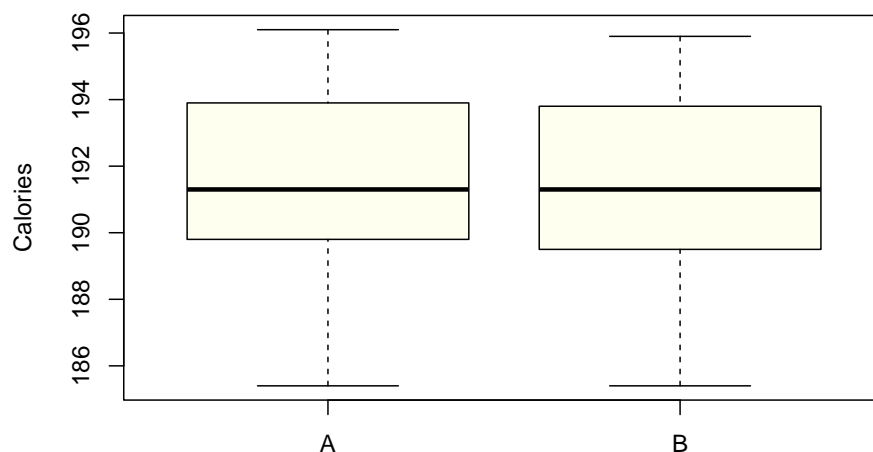
The p -value for the test is 0.047, bigger than $\alpha = 0.02$, and thus we do not reject the null hypothesis that the true value for the mean caloric content is 190.

- (b) The quality control department for the chocolate manufacturer wants to introduce a new and cheaper method for measuring the caloric content of the bars. Samples from the same chocolate bars as in question 1(a) were used to test the new process. The results are stored in the variable `barB` in the file `calories.csv`, in the same order as the original measurements in variable `barA`. Draw boxplots to compare the results for the two measurement methods and comment on the results.

Solution:

Since we want to use the boxplots to compare the measurements, we want the scales to be the same. One way of doing this is

```
boxplot(c(data1$barA,data1$barB) ~ rep(c('A','B'), each = 25),
        col = 'ivory', ylab = 'Calories', xlab = '')
```



We see that the plots are very similar. All the quartiles are approximately the same. There are no important differences in the plots.

- (c) What parametric test would be adequate for comparing the two measurement methods? State clearly what hypothesis you are testing and which assumptions are needed for the test, and explain why they are satisfied. Identify the type I and type II errors. Carry out this test and discuss the results.

Solution:

Since the measurements are done using the same chocolate bars, the data are paired and we should use a paired t -test. If μ_A is the mean for the caloric content according to the first method, and μ_B is the mean according to the second method, we want to consider

$$H_0 : \mu_A = \mu_B \quad \text{vs.} \quad H_A : \mu_A \neq \mu_B$$

In this case the two samples are not independent, and we can verify this by calculating the correlation between them:

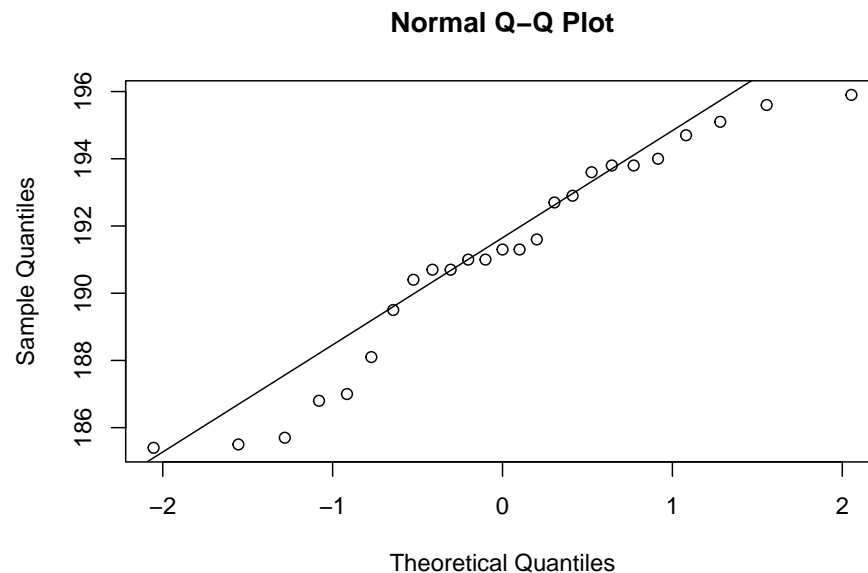
```
cor(data1$barA, data1$barB)
```

```
## [1] 0.9962476
```

The correlation is very high and shows that the samples are not independent.

The other assumption is that the samples come from normal populations, since the samples sizes are not large enough to use the CLT. We have already checked normality for `barA`, let's check it for `barB`:

```
qqnorm(data1$barB)
qqline(data1$barB)
```



The plot is not very good, so we do a Shapiro-Wilk test:

```
shapiro.test(data1$barB)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data1$barB
## W = 0.93639, p-value = 0.1221
```

The p -value is above the significance level and we do not reject the null hypothesis of normality.

In this test, a type I error would reject that the two measuring methods produce the same average values in favor of the alternative that they do not when, in fact, the methods are equivalent. A type II error would be

to conclude that the two methods are equivalent when they are not.

The test is

```
t.test(data1$barA, data1$barB, paired = TRUE, conf.level = 0.98)
```

```
##
## Paired t-test
##
## data: data1$barA and data1$barB
## t = 3.9513, df = 24, p-value = 0.0005958
## alternative hypothesis: true mean difference is not equal to 0
## 98 percent confidence interval:
## 0.08124243 0.35875757
## sample estimates:
## mean difference
## 0.22
```

The p -value is small and we conclude that the two methods are not equivalent. If we had used a standard (non-paired) test

```
t.test(data1$barA, data1$barB)
```

```
##
## Welch Two Sample t-test
##
## data: data1$barA and data1$barB
## t = 0.24232, df = 47.999, p-value = 0.8096
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -1.605452 2.045452
## sample estimates:
## mean of x mean of y
## 191.344 191.124
```

the conclusion would have been different (and wrong!).

- (d) What non-parametric tests will be adequate for the problems in (a) and (c)? What assumptions are needed, and why do you think they are satisfied? Perform these tests, discuss the results, and compare them with your previous results.

For (a) we can use the Wilcoxon signed-ranks test. This test assumes that the sample comes from a continuous distribution that is symmetric about the mean. Since we did not reject that the sample comes from a normal distribution, which is symmetric about its means, we have grounds to think that this assumption is satisfied.

For the sample in part (a) we have

```
wilcox.test(data1$barA, mu = 190, conf.level = 0.98)
```

```
## Warning in wilcox.test.default(data1$barA, mu = 190, conf.level = 0.98): cannot
## compute exact p-value with ties
##
## Wilcoxon signed rank test with continuity correction
##
## data: data1$barA
## V = 234.5, p-value = 0.0543
## alternative hypothesis: true location is not equal to 190
```

For (b) the non-parametric alternative is the Wilcoxon rank sums test. This test assumes that the data come

from a continuous distribution which is symmetric about its mean. Since in both samples the assumption that the data come from a normal distribution was not rejected, these assumptions seem to be satisfied. We need to use a paired test

```
wilcox.test(data1$barA, data1$barB, paired = TRUE, conf.level = 0.98)
```

```
## Warning in wilcox.test.default(data1$barA, data1$barB, paired = TRUE,
## conf.level = 0.98): cannot compute exact p-value with ties

## Warning in wilcox.test.default(data1$barA, data1$barB, paired = TRUE,
## conf.level = 0.98): cannot compute exact p-value with zeroes

##
## Wilcoxon signed rank test with continuity correction
##
## data: data1$barA and data1$barB
## V = 207, p-value = 0.001516
## alternative hypothesis: true location shift is not equal to 0
```

The p -value is below the significance level and we would reject the null hypothesis that the two measurement processes are equivalent.

- (e) The chocolate manufacturer also produces a ‘light’ version of the chocolate bar that only has 100 calories for bars of the same size. The variable `barC` in the file `calories.csv` has the result of calory measurements in 25 randomly chosen light bars produced by the manufacturer. For the company, it is important that the average calories in the bars do not exceed the amount stated in the packing. What parametric test would be adequate in this case? State clearly what hypotheses you are testing and which assumptions are needed for the test. Explain why they are satisfied. Describe the test statistic and calculate its value. Identify the type I and type II errors explicitly. Carry out this test and discuss the results.

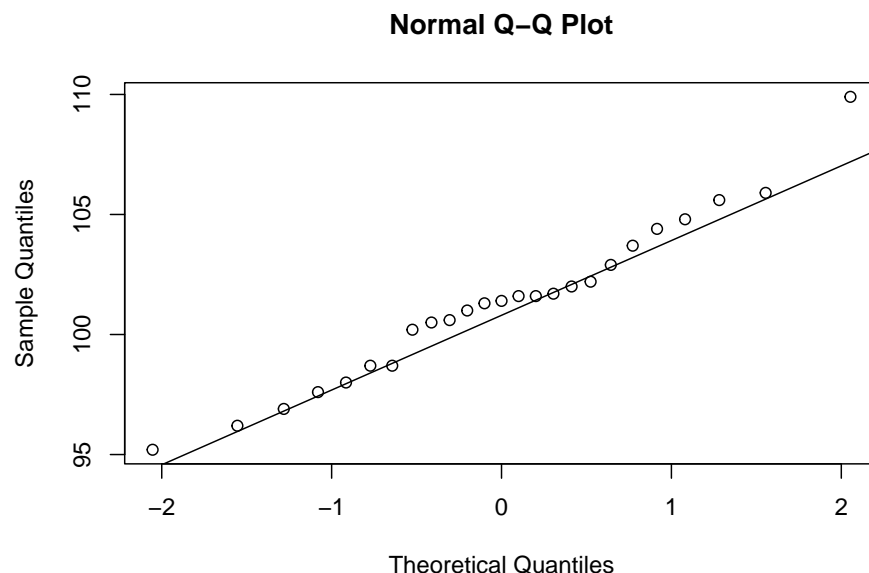
Solution:

In this case we also use the one-sample t -test but we have a one-sided alternative, because the company wants to be sure that the calories are not above 100. We are testing

$$H_0 : \mu = 100 \quad \text{vs.} \quad \mu > 100.$$

The assumptions are the same as in Q1(a) and, since the sample size is not large, we need to test whether the sample comes from a normal distribution. We start with a normal quantile plot

```
qqnorm(data1$barC)
qqline(data1$barC)
```



The plot looks better this time, but we have a large value at the upper tail. We confirm with a Shapiro-Wilk test:

```
shapiro.test(data1$barC)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data1$barC
## W = 0.97508, p-value = 0.7737
```

The p -value is large and we do not reject the null hypothesis of normality.

The test statistic in this case is the standardized sample mean

$$\frac{\hat{\mu}_n - 100}{s_n / \sqrt{25}}$$

where s_n is the empirical standard deviation. This test statistic has a t distribution with $n - 1 = 24$ degrees of freedom. The value for the test statistic is

```
(tn <- (mean(data1$barC-100)/(sd(data1$barC)/sqrt(25))))
```

```
## [1] 1.949131
```

The type I error would be to reject the null hypothesis in favor of the alternative that says that the caloric content is more than 100 when, in fact, this is not true. The type II error is to accept that the caloric content is 100 when it is more than 100.

We now do the test

```
t.test(data1$barC, mu = 100, alternative = 'greater', conf.level = 0.98)
```

```
##
##  One Sample t-test
##
## data:  data1$barC
## t = 1.9491, df = 24, p-value = 0.03153
## alternative hypothesis: true mean is greater than 100
## 98 percent confidence interval:
```

```
## 99.8512      Inf
## sample estimates:
## mean of x
## 101.304
```

The p -value is above the significance level of 2%, so we do not reject the null hypothesis that the caloric content is 100 calories.

- (f) A different factory of the same company produces the same chocolate bars and measures calories using the same method as in part (a) of this question. A sample of size 25 is obtained from the production of this factory, and the measured caloric content is stored in variable `barD`, in the same file. You want to determine whether the bars produced in both factories have the same average calories. What parametric test would be adequate for comparing the production of the two factories? State clearly what hypothesis you are testing and which assumptions are needed for the test. Explain why they are satisfied. Carry out this test and discuss the results.

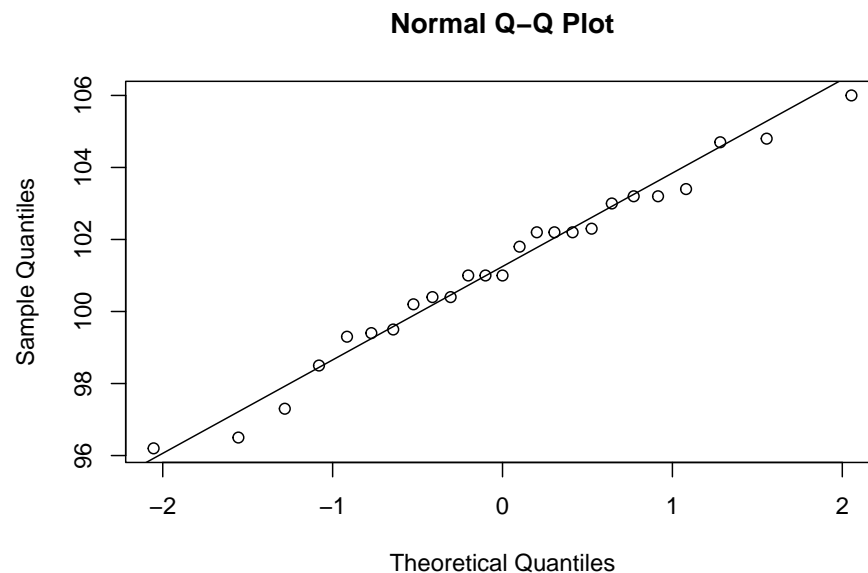
Solution:

In this case we can use a two-sample t -test to compare whether the two factories produce chocolate bars having, on average, the same caloric content. The hypothesis to be tested are

$$H_0 : \mu_C = \mu_D \quad \text{vs.} \quad \mu_C \neq \mu_D$$

In this test we assume that the two samples are independent, which is reasonable since they come from different factories, and that the two populations are normal, which needs checking. We have already considered `barC`, we need to look at `barD`, and we use quantile plots for this:

```
qqnorm(data1$barD)
qqline(data1$barD)
```



The plot is very good, and a Shapiro-Wilk test

```
shapiro.test(data1$barD)
```

```
##
## Shapiro-Wilk normality test
##
## data:  data1$barD
## W = 0.97803, p-value = 0.8434
```


gives a high p -value. We do not reject the null hypothesis of normality.

The test is

```
t.test(data1$barC, data1$barD, conf.level = 0.98)

##
## Welch Two Sample t-test
##
## data: data1$barC and data1$barD
## t = 0.13878, df = 44.48, p-value = 0.8902
## alternative hypothesis: true difference in means is not equal to 0
## 98 percent confidence interval:
## -1.901008 2.133008
## sample estimates:
## mean of x mean of y
## 101.304 101.188
```

and since the p -value is above 0.02, we do not reject the null hypothesis that the two factories produce chocolate bars with the same calories.

- (g) What non-parametric tests will be adequate for the problems in (a) and (b)? What assumptions are needed, and why do you think they are satisfied? Perform these tests, discuss the results, and compare them with your previous results.

For (a) we can use the Wilcoxon signed-ranks test. This test assumes that the sample comes from a continuous distribution that is symmetric about the mean. Since we did not reject that the sample comes from a normal distribution, which are symmetric about their means, we have grounds to think that this assumption is satisfied.

```
wilcox.test(data1$barC, mu = 100, alternative = 'greater', conf.level = 0.98)

## Warning in wilcox.test.default(data1$barC, mu = 100, alternative = "greater", :
## cannot compute exact p-value with ties
##
## Wilcoxon signed rank test with continuity correction
##
## data: data1$barC
## V = 228, p-value = 0.0401
## alternative hypothesis: true location is greater than 100
```

For (b) the non-parametric alternative is the Wilcoxon rank sums test. This test assumes that the data come from a continuous distribution which is symmetric about its mean. Since in both samples the assumption that the data come from a normal distribution was not rejected, these assumptions seem to be satisfied.

```
wilcox.test(data1$barC, data1$barD, conf.level = 0.98)

## Warning in wilcox.test.default(data1$barC, data1$barD, conf.level = 0.98):
## cannot compute exact p-value with ties
##
## Wilcoxon rank sum test with continuity correction
##
## data: data1$barC and data1$barD
## W = 308.5, p-value = 0.9458
## alternative hypothesis: true location shift is not equal to 0
```

In both cases we reach the same conclusions.

Problem 4

A new drug for reducing cholesterol levels in the blood is tested in an experiment involving 15 volunteers. The subjects were tested for cholesterol before and after taking the drug, and the results are stored in the file `p14-25_q4.csv`. The variable `chl` stands for cholesterol level in milligrams per deciliter (mg/dl), while `type` denotes whether the value corresponds to the measurement before (C1) or after (C2) taking the drug.

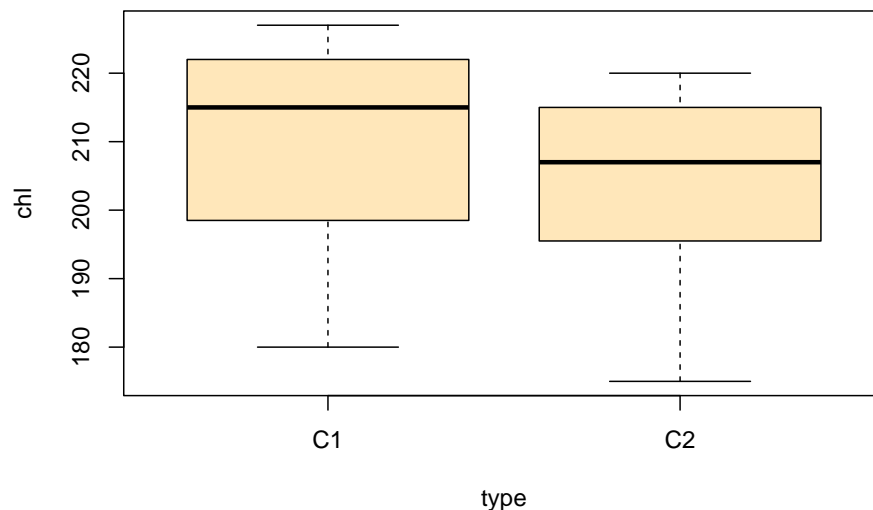
```
data2 <- read.csv('25Fp14Q4.csv')
str(data2)

## 'data.frame': 30 obs. of 3 variables:
## $ X : int 1 2 3 4 5 6 7 8 9 10 ...
## $ chl : int 180 226 208 221 199 189 224 219 197 215 ...
## $ type: chr "C1" "C1" "C1" "C1" ...

data2$type <- factor(data2$type)
```

- (a) Use graphical tools to compare the cholesterol level in the blood before and after taking the drug and comment on what you observe.

```
boxplot(chl ~ type, data = data2, col = 'wheat1')
```



We see that the values after taking the treatment are lower than the values before. The two boxes have similar sizes and the difference between the maximum and minimum values is also similar. There are no outliers in the plot.

- (b) You want to test whether the drug effectively reduces the cholesterol level in the blood using this data. What hypotheses do you want to test? What test or tests could be appropriate here? What are the assumptions? Why do you think they are satisfied in this case? Carry out the test(s) and comment on your results.

The test we are interested in is

$$H_0 : \mu_{C1} = \mu_{C2} \quad vs. \quad \mu_{C1} > \mu_{C2}$$

where μ_{C1} and μ_{C2} represent the average cholesterol level in the blood for the two groups. The null hypothesis is that there is no effect, while the alternative is that the treatment reduces the cholesterol level.

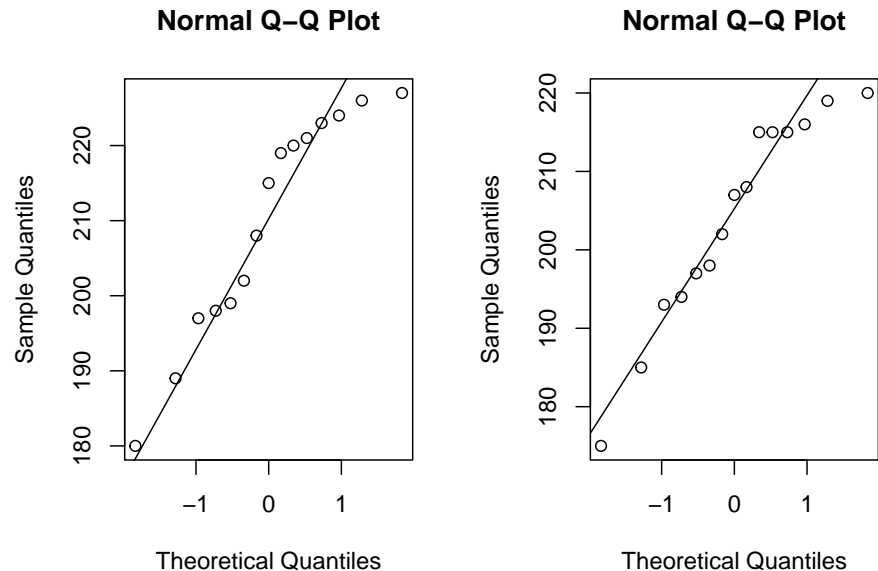
Since the measurements are made on the same subjects, before and after treatment, we have paired data. Therefore, if we want to use a parametric test we should use a *t*-test for paired data. This test assumes that the population distribution is normal, and that the measurements are correlated. The latter assumption is verified because the measurements are paired. We can calculate the correlation:

```
with(data2,cor(chl[type == 'C1'], chl[type == 'C2']))
```

```
## [1] 0.9742868
```

To check normality we use a quantile plot

```
par(mfrow = c(1,2))
qqnorm(data2$chl[data2$type=='C1']); qqline(data2$chl[data2$type=='C1'])
qqnorm(data2$chl[data2$type=='C2']); qqline(data2$chl[data2$type=='C2'])
```



```
par(mfrow = c(1,1))
```

The quantile plots look good and support the assumption of normality. We do a Shapiro-Wilk test

```
shapiro.test(data2$chl[data2$type=='C1'])
```

```
##
## Shapiro-Wilk normality test
##
## data: data2$chl[data2$type == "C1"]
## W = 0.90908, p-value = 0.1311
```

```
shapiro.test(data2$chl[data2$type=='C2'])
```

```
##
## Shapiro-Wilk normality test
##
## data: data2$chl[data2$type == "C2"]
## W = 0.92375, p-value = 0.2197
```

In both cases the p -value is above the significance level and we do not reject the null hypothesis of normality.

We do the test with the following command. Observe that we want a one-sided test since we want to determine if the treatment reduces the cholesterol level.

```
with(data2,t.test(chl[type == 'C1'], chl[type == 'C2'] , paired = T,
                  alternative = 'greater'))
```

```
##
## Paired t-test
```

```
##
## data:  chl[type == "C1"] and chl[type == "C2"]
## t = 6.5765, df = 14, p-value = 6.18e-06
## alternative hypothesis: true mean difference is greater than 0
## 95 percent confidence interval:
##  4.344276      Inf
## sample estimates:
## mean difference
##      5.933333
```

The p -value is small and we reject the null hypothesis of no effect. Observe that if we don't do a paired test, the conclusion is different:

```
with(data2,t.test(chl[type == 'C1'], chl[type == 'C2'] , alternative = 'greater'))
```

```
##
## Welch Two Sample t-test
##
## data:  chl[type == "C1"] and chl[type == "C2"]
## t = 1.1485, df = 27.728, p-value = 0.1303
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## -2.857875      Inf
## sample estimates:
## mean of x mean of y
##  209.8667  203.9333
```

Another possibility is the Wilcoxon test for paired data, which is a non-parametric alternative.

```
with(data2,wilcox.test(chl[type == 'C1'], chl[type == 'C2'], paired = T,
                        alternative = 'greater'))
```

```
## Warning in wilcox.test.default(chl[type == "C1"], chl[type == "C2"], paired =
## T, : cannot compute exact p-value with ties
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data:  chl[type == "C1"] and chl[type == "C2"]
## V = 120, p-value = 0.0003404
## alternative hypothesis: true location shift is greater than 0
```

In the output we get a warning about ties in the data. The p -value is also small and we would reach the same conclusion.

- (c) The pharmaceutical company is only interested in producing this drug if the reduction in cholesterol level in the blood is more than 10 mg/dl. To simplify the test, suppose you want to test that the reduction in cholesterol is 10 mg/dl versus the alternative that it is more. How would you carry out this test with the data that you have? What assumptions are needed? Are they justified in this case? Carry out the test or tests and comment on your results.

Let's look at the difference between the cholesterol level before and after treatment, and let's call this variable dif :

$$dif = CL(before) - CL(after)$$

where CL stands for cholesterol level. Then we want to test whether dif is equal to 10 versus the alternative that it is bigger:

$$H_0 : dif = 10 \quad vs. \quad H_A : dif > 10$$

We start by calculating the difference between the cholesterol levels before and after the treatment:

```
dif <- data2$chl[data2$type=='C1'] - data2$chl[data2$type=='C2']
```

The test is a one-sample t -test:

```
t.test(dif, mu = 10, alternative = 'greater')
```

```
##
##  One Sample t-test
##
## data:  dif
## t = -4.5075, df = 14, p-value = 0.9998
## alternative hypothesis: true mean is greater than 10
## 95 percent confidence interval:
##  4.344276      Inf
## sample estimates:
## mean of x
##  5.933333
```

The p -value is large and we cannot reject the null hypothesis. The reduction is not bigger than 10 units.