Statistical Tests

Computer Lab 3

Your Name

2025-03-16

The purpose of today's lab is for you to:

- practice the basic concepts of hypothesis testing (e.g., significance level and power) and what conclusions can be drawn from the analysis
- familiarize yourself with some of the functions available in R regarding various basic statistical tests
- practice constructing a reasonable statistical model using collected data, as well as critically reviewing the model and its ability to describe reality
- apply your knowledge and analyse a biostatistical dataset using R
- practice presenting assumptions, models, and conclusions from a statistical analysis in writing

Table of contents

Preparation Tasks

Review the concepts of *hypotheses, significance level, power function,* and *model with matched data and model with two independent samples.*

You should have completed the following tasks *before* you come to the lab.

Homework 1

To practice the basic concepts in hypothesis testing, complete exercises **Dig:4.4.1(error rates)_3, _4, and _5 in exercise 7**.

Homework 2

To practice different models, do **Dig:5.2_1 and _2**.

Before You Start

Load R functions needed for the lab. These functions come from the R package **Räkna med variation**, but to avoid installation problems, we have extracted the ones you will use and saved them in a file. You "inform" R that these functions exist by running the following code:

```
source("kod/funktioner_raknamedvariation_light.R")
```

1. Basic Concepts in Hypothesis Testing

Medication can reduce salivary gland function, which is a risk factor for cavities and other oral diseases. Saliva stimulated by chewing was measured for 5 minutes in 7 randomly selected patients, all receiving the same medication. The normal saliva amount under these conditions is 1 ml/min, and dry mouth is considered when the saliva amount is below 0.7 ml/min. The model assumed that the saliva amount is normally distributed with mean μ and standard deviation σ, where σ is considered to be 0.5 ml/min. Interesting questions include:

- Do the data support our suspicion that the medication lowers saliva production?
- If the medication results in an average saliva production of 0.8 ml/min, how likely are we to miss the reduced saliva production with our test?
- How many patients should we measure if we want the test to detect reduced saliva production at 0.7 ml/min with a probability of 0.95?

On the course website, you will find data in the file `saliv.RData`. Concise answers to the questions posed in the tasks are provided at the end of this part of the guide.

Task 1.1

First, we want to investigate if the data from the 7 patients support our suspicion that the medication lowers saliva production. Set up a model and appropriate hypotheses.

Model:

Hypotheses:

Task 1.2

Calculate the mean of the measurements.

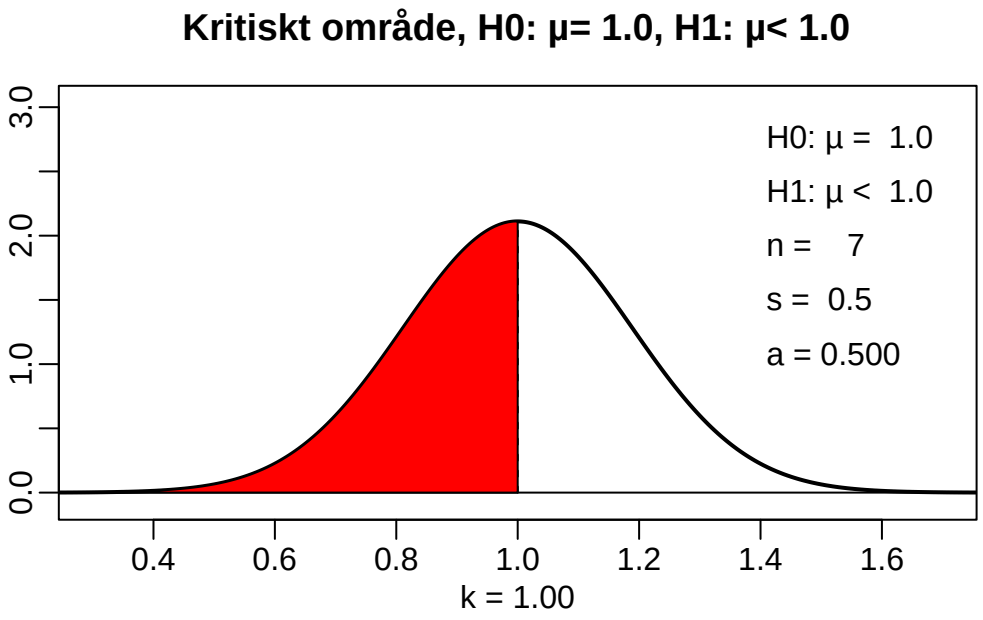
```
# Write your R code here
```

Answer:

Task 1.3

Use the routine `hypotes` to illustrate the test's critical region when the test is performed at a significance level of $\alpha = 0.05$. The relevant command is `hypotes(`σ, n, μ_0, α, `H1-direction)`. With hypotheses $H_0 : \mu = 1$ and $H_1 : \mu < 1$, the command becomes `hypotes(0.5, 7, 1, 0.05, '<')`. Use the routine with the hypotheses you have formulated. (Ignore any error messages that may occur and look at the figure.)

```
hypotes(0.5, 7, 1, 0.5, '<')
```



Task 1.4

The routine marks the critical region and provides a value k which is the boundary of the region. How has k been calculated? Set up the mathematical formula for k and use it to calculate k yourself.

Equation for k:

Own calculation of k:

```
# Write your R code here
```

Task 1.5

Use your calculated sample mean to conduct the test. What is your conclusion about the null hypothesis, H_0?

Answer:

Task 1.6

What is your concrete interpretation of the significance level $\alpha = 0.05$ in this example?

Answer:

Task 1.7

Examine how the critical region changes when you change the significance level to $\alpha = 0.01$. What is your conclusion now?

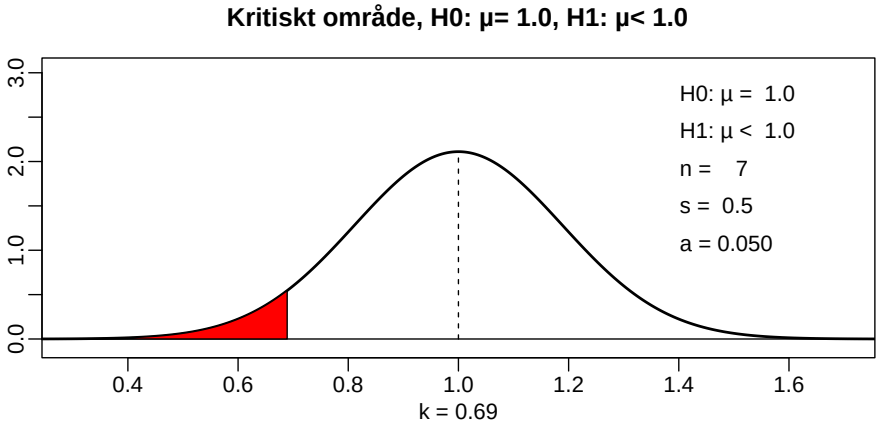
```
# write your R code here
```

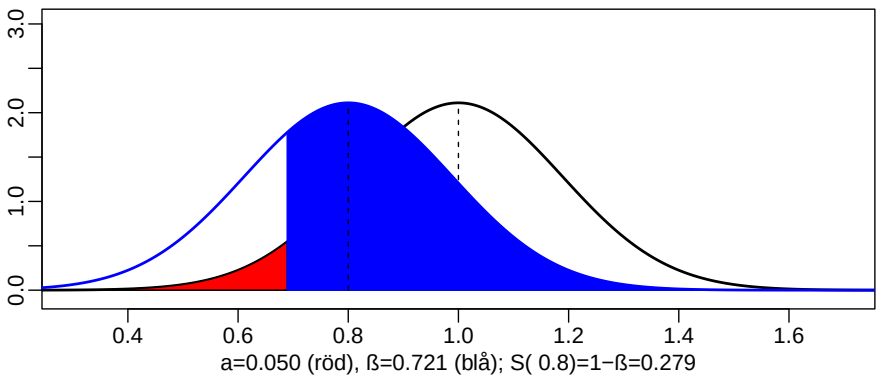
Answer:

2. Test Power and Power Function

Now assume that the average saliva secretion in the risk group is 0.8. Then, of course, $H_0 : \mu = 1$ is false, and we want our test to detect this and reject this hypothesis in favor of the hypothesis $H_1 : \mu < 1$. The probability that the test actually accomplishes this is called the test power at the point 0.8. Use the routine `hypotes` to illustrate the test power at the point 0.8. The command now is `hypotes(`σ, n, μ_0, α, H1-direction, true μ`), so in this case, you write`hypotes(0.5, 7, 1, 0.05, '<', 0.8)'.

```
hypotes(0.5, 7, 1, 0.05, '<', 0.8)
```





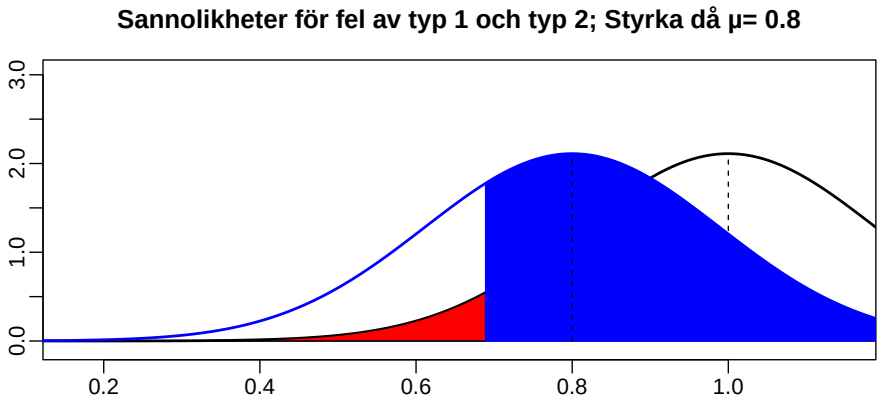
Task 2.1

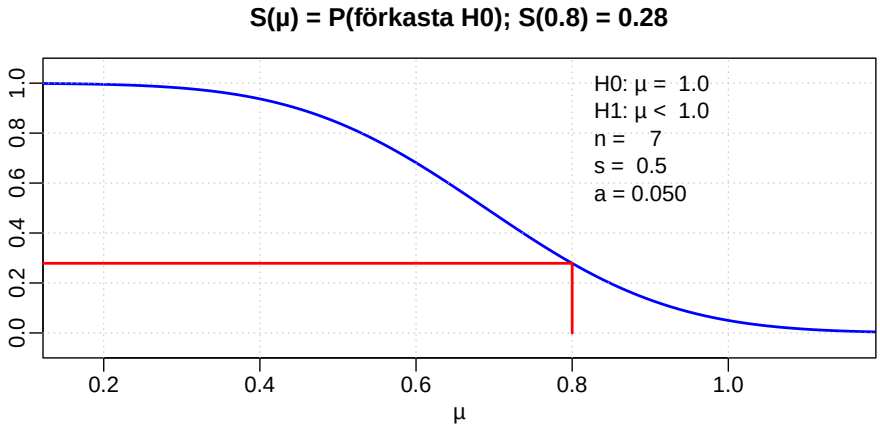
The routine gives you another figure which, besides the significance level α (type I error risk), also shows β (type II error risk). What is the concrete interpretation of β in this example? How does β relate to the test power?
Generally, the test power at the point μ is the probability that the null hypothesis is rejected given that μ is the true average saliva secretion in the risk group, i.e.,

$$S(\mu) = P(\text{reject } H_0|\mu)$$

Note that the power depends on the value of μ. In this example, the smaller μ is relative to $\mu_0 = 1$, the greater the chance that the test will detect that H_0 is not true. Therefore, it is interesting to study the power as a function of μ, and this function is often denoted $S(\mu)$.
The command `styrka(0.5, 7, 1, 0.05, '<', 0.8)` gives you the previous figure plus the power as a function of μ.

```
styrka(0.5, 7, 1, 0.05, '<', 0.8)
```





Task 2.2

Based on the power function, estimate the probability that with our test, we will detect that a group that should be classified as having dry mouth ($\mu = 0.7$) has reduced saliva production.

```
# write your R code here
```

Answer:

Task 2.3

How many patients should we measure if we want to have a 0.95 probability of detecting that individuals with dry mouth have reduced saliva production? Hint: Try different values of n in `styrkefkn`.

```
# write your R code here
```

Answer:

Answer to the example with dry mouth

1. Part 1
   1.1 $H_0 : \mu = 1$; $H_1 : \mu < 1$
   1.2 $k = \mu_0 - z_{1-\alpha} \frac{\sigma}{\sqrt{n}} = 1 - 1.6445 \cdot \frac{0.05}{\sqrt{7}} = 0.689$
   1.5 Since the mean is < 0.689, H_0 is rejected at the 0.05 level
   1.6 There is a 5% chance that we claim a person in the risk group has reduced saliva production when it is actually normal 1.7 H_0 cannot be rejected at the 0.01 level
2. Part 2
   2.1 $\beta = P(\text{not reject } H_0|\mu)$, i.e., β when the true expected value is 0.8 is 1 minus the power at the point 0.8 = 1- S(0.8)
   2.2 The power at the point 0.7, S(0.7), can be estimated as 0.48 according to the figure
   2.3 $n = 30$ patients are required for the power to be 0.95 at the point 0.7

3. Test of Population Mean (t-test)

> A mixture of blood serum contains exactly 42 g of albumin per liter. Two laboratories (A and B) make six determinations each of the concentration. We want to investigate if there is any systematic deviation from the true value (42 g/l) in each of these two groups.

```
load("data/lab3_filer/albumin.RData")
```

> Exercise 3.1
>
> Specify the model for A and B and specify hypotheses to test if there is a systematic deviation from the true value in each group. Hint: it should be a two-sided alternative hypothesis.

Model for group A: Let X be the concentration of g of albumin per liter in measurements from laboratory A. The mean and variance for X are μ_x and σ_x^2. Assume that measurements are independent and identically distributed. The mean is estimated by the sample mean $\hat{\mu}_x = \bar{x}$. According to the central limit theorem, $\bar{x} \overset{A}{\sim} N(\mu_x, \frac{\sigma_x}{\sqrt{n_x}})$.

Hypotheses for group A:

Model for group B: Let Y be the concentration of g of albumin per liter in measurements from laboratory B.

Hypotheses for group B:

> Exercise 3.2
>
> In R, perform these two t-tests using the following routine `t-test`. Interpret the outputs when you run the command. What are the p-values and what are the conclusions

for each group? What are the confidence intervals for the expected concentrations?

```
t.test(AlbuminA, mu=42)
```

```
    One Sample t-test

data:  AlbuminA
t = 2.9194, df = 5, p-value = 0.03304
alternative hypothesis: true mean is not equal to 42
95 percent confidence interval:
 42.05974 42.94026
sample estimates:
mean of x
     42.5
```

```
t.test(AlbuminB, mu=42)
```

```
    One Sample t-test

data:  AlbuminB
t = -2.1843, df = 5, p-value = 0.08067
alternative hypothesis: true mean is not equal to 42
95 percent confidence interval:
 35.68716 42.51284
sample estimates:
mean of x
     39.1
```

Answer:

Exercise 3.3

Suppose we want to investigate if **laboratory B** underestimates or overestimates concentrations. In R, this is done as a one-sided test using the routine `t-test`.
Interpret the output for the one-sided test for laboratory B. What happens to the p-values and intervals when moving from one-sided to two-sided alternatives?

```
t.test(AlbuminB, mu=42, alternative="less")
```

```
    One Sample t-test

data:  AlbuminB
t = -2.1843, df = 5, p-value = 0.04034
alternative hypothesis: true mean is less than 42
```

```
95 percent confidence interval:
     -Inf 41.77529
sample estimates:
mean of x
     39.1
```

```
t.test(AlbuminB, mu=42, alternative="greater")
```

```

    One Sample t-test

data:  AlbuminB
t = -2.1843, df = 5, p-value = 0.9597
alternative hypothesis: true mean is greater than 42
95 percent confidence interval:
 36.42471      Inf
sample estimates:
mean of x
     39.1
```

Answer:

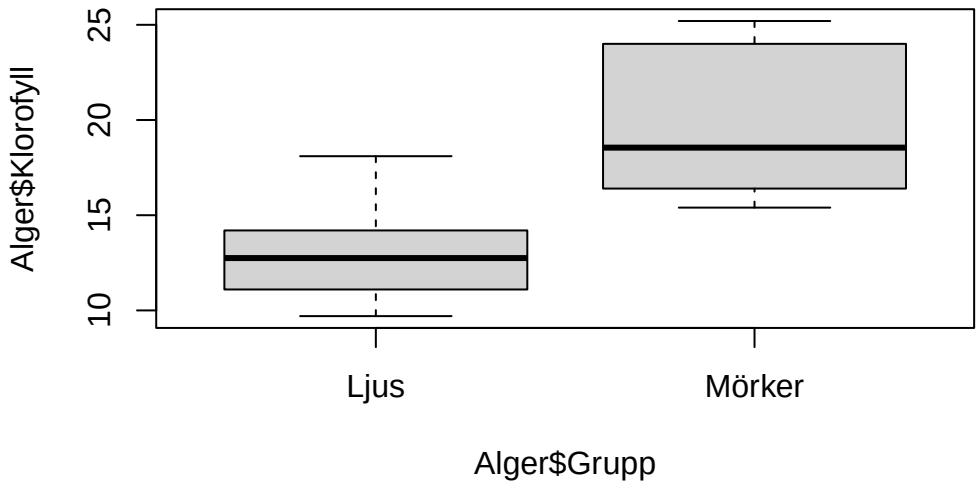
4. Comparison of Means in Two Populations (t-test for Two Independent Samples)

> Algae were allowed to grow under light and dark conditions, and then the chlorophyll content was measured.

```
load("data/lab3_filer/alger.RData")
```

> Exercise 4.1
>
> Make a graphical description of the differences. What does the boxplot show?

```
boxplot(Alger$Klorofyll ~ Alger$Grupp)
```



Answer:

> Exercise 4.2
>
> Set up the model and hypotheses, and investigate with a t-test if there are differences in the expected chlorophyll content between the two groups.
> This can be done by adding the group variable in the `t.test` routine. Interpret the output.

Model: Let X be the growth of algae in dark conditions with mean μ_x and variance σ_x^2, and Y be the growth of algae in light conditions with mean μ_y and variance σ_y^2. The means are estimated by sample means $\hat{\mu_x} = \bar{x}$ and $\hat{\mu_y} = \bar{y}$. According to the central limit theorem, the estimation of each mean will be approximately normally distributed, i.e., $\bar{x} \overset{A}{\sim} N(\mu_x, \frac{\sigma_x}{\sqrt{n_x}})$ and $\bar{y} \overset{A}{\sim} N(\mu_y, \frac{\sigma_y}{\sqrt{n_y}})$.

Hypotheses: $H_0 : \mu_x = \mu_y$ vs. $H_1 : \mu_x \neq \mu_y$

```
t.test(Alger$Klorofyll ~ Alger$Grupp)
```

```
    Welch Two Sample t-test

data:  Alger$Klorofyll by Alger$Grupp
t = -3.2089, df = 9.0652, p-value = 0.01058
alternative hypothesis: true difference in means between group Ljus and group Mörker is not
95 percent confidence interval:
 -11.219229  -1.947438
```

```
sample estimates:
  mean in group Ljus mean in group Mörker
            13.10000             19.68333
```

Answer:

> Exercise 4.3
>
> If we don't specify otherwise, the `t.test` routine assumes that the variances in the two groups, i.e., σ_x^2 and σ_y^2, are **unequal** and adjusts for it. If we know (or assume) that the variances are equal, we can utilize that to obtain higher degrees of freedom since there's one less parameter to estimate.
> Investigate if the variances are equal using a so-called χ^2-test. This is done in R using the `var.test` routine. Interpret the output!

Hypotheses: $H_0 : \sigma_x^2 = \sigma_y^2$ vs. $H_1 : \sigma_x^2 \neq \sigma_y^2$

```
var.test(Alger$Klorofyll ~ Alger$Grupp)
```

```

    F test to compare two variances

data:  Alger$Klorofyll by Alger$Grupp
F = 0.51386, num df = 5, denom df = 5, p-value = 0.4825
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.07190457 3.67222210
sample estimates:
ratio of variances
         0.5138575
```

Answer:

> Exercise 4.4
>
> Repeat the t-test assuming equal variances by adding `var.equal=TRUE` in the R routine. What are your conclusions now?

Assumption: $\sigma^2 = \sigma_x^2 = \sigma_y^2$

```
t.test(Alger$Klorofyll ~ Alger$Grupp, var.equal=TRUE)
```

```

    Two Sample t-test

data:  Alger$Klorofyll by Alger$Grupp
t = -3.2089, df = 10, p-value = 0.009349
```

```
alternative hypothesis: true difference in means between group Ljus and group Mörker is not
95 percent confidence interval:
 -11.154519  -2.012148
sample estimates:
  mean in group Ljus mean in group Mörker
            13.10000             19.68333
```

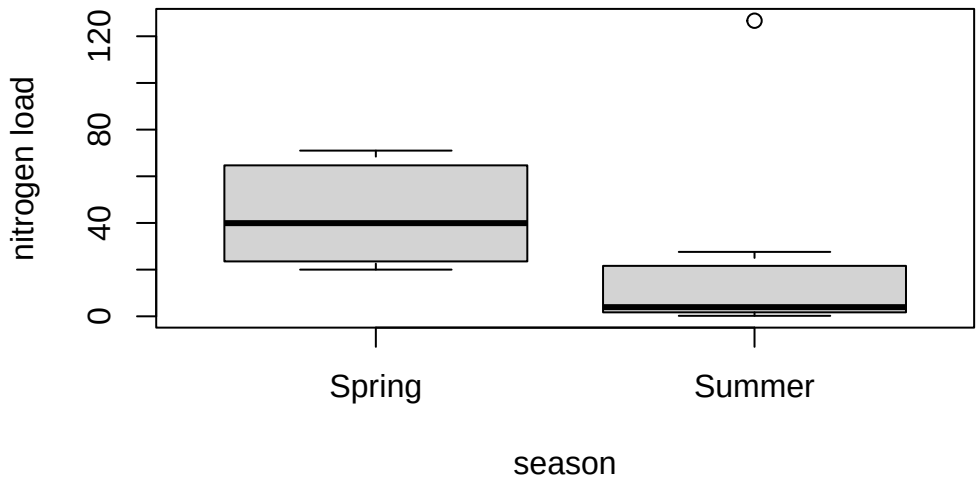
Answer:

5. Test with Matched Data (Paired Samples)

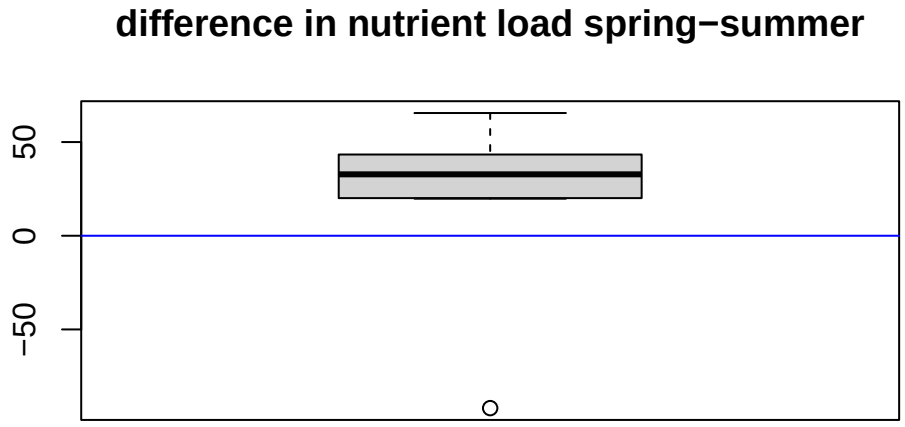
> Measurements of nitrogen load have been taken during spring and summer in a number of ponds. If one wants to compare the nitrogen load between the two seasons, a reasonable model is "paired samples."

```
load("data/lab3_filer/dammar.RData")
```

```
boxplot(Dammar$N_belast_V, Dammar$N_belast_S,
        ylab = "nitrogen load",
        xlab = "season",
        names = c("Spring","Summer"))
```




```
boxplot(Dammar$N_belast_V-Dammar$N_belast_S,
        main = "difference in nutrient load spring-summer")
abline(a=0,b=0,col='blue')
```



Model: Let D be the difference in measurements between spring and summer for a pond. The random variable D has mean μ and variance σ^2. We create a sample $(d_1, ..., d_n)$ for differences in measurement values. The mean is estimated by the sample mean, $\hat{\mu} = \bar{d}$. According to the central limit theorem, $\bar{d} \overset{A}{\sim} N(\mu, \frac{\sigma}{\sqrt{n}})$.

> Exercise 5.1
>
> Is there a difference in means between spring and summer?
> We test it with the following **hypotheses:** $H_0 : \mu = 0$ vs. $H_1 : \mu \neq 0$
> This is a common test, and one does not need to calculate a new sample. To inform `t.test` that the data is matched in pairs, one adds `paired=TRUE` as an argument.

```
t.test(Dammar$N_belast_V, Dammar$N_belast_S, paired=TRUE)
```

```
    Paired t-test

data:  Dammar$N_belast_V and Dammar$N_belast_S
t = 1.2189, df = 7, p-value = 0.2624
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -19.49477  60.97227
sample estimates:
mean difference
       20.73875
```

Answer:

Answers to Some of the Questions in This Section:

1. Answer
   3.2 A: p-value= 0.033; B: p-value= 0.081; We can detect a difference for A but not for B. 95% confidence intervals: A: (42.06, 42.94); B: (35.69, 42.51)
   4.3 The variances are not different (p-value= 0.48)
   4.4 t-test: p-value= 0.009; We can detect a difference in chlorophyll.
   5.1 p-value= 0.262 We cannot detect any difference.

6. Had our activation program an effect?

Problem specification in the study

Most researchers believe that high cholesterol levels in the blood are a risk factor for heart and vascular diseases. In a study, we wanted to investigate whether a multi-factor activation program (smoking cessation, mental, and physical training) could reduce cholesterol levels. We started with a group of 40 smoking men, all of whom had slightly elevated blood cholesterol levels. From these 40, we randomly selected 20 (Group A) to undergo our activation program.

The remaining 20 (Group B) continued with their usual lifestyle, at least as far as we perceived it. After six months, we measured the cholesterol levels (mmol/l) in all 40 men again.

In the file `blodprov.Rdata` there is data with the variables ´Afore´, ´Aefter´, ´Bfore´ and ´Befter´.

Questions

Question 6.1

Since we randomly selected the 20 individuals to undergo the activation program, there should be no differences between Group A and Group B regarding average cholesterol levels **before** the study begins. However, we want to ensure this explicitly so that we do not introduce a systematic difference between the groups from the start.
Is there a difference between the groups? Investigate with a suitable statistical test.

Question 6.2

Has the expected cholesterol level decreased in Group A? Investigate with a suitable statistical test.

Question 6.3

It is not unlikely that patients in Group B, even though they do not undergo the activation program, may still be affected in their cholesterol levels because attention to the issues may have an effect (a placebo effect).
Does this seem to be the case in our study? Study if there is a difference in cholesterol levels in patients in group B before and after the program. Investigate with a suitable statistical test.

Question 6.4

Did the activation program reduce the expected cholesterol level in Group A? Investigate with a suitable statistical test if the activation program affects the two groups differently?

Instruction for reporting

Please specify (if appropriate):

- what assumptions you make about data,
- which models you are specifying,
- which hypothesis you use,
- which theorems you use.
- Present the results of the analysis and what interpretations and conclusions you make.
- To pass task 6, avoid referring to the R-code to describe what you have done.