One sample model. One sample model has no explanatory variable (although there could have been one such as sex or age), and it is determined by population parameters. We then assume that $y_1, \ldots, y_n$ are independent with

$$y_i \sim N(\mu, \sigma^2), \quad i = 1, \ldots, n$$

All observations are assumed to be sampled from the same distribution. Equivalently, we may write

$$y_i \sim \mu + e_i \quad i = 1, \ldots, n$$

where $e_1, \ldots, e_n$ are independent and $N(0, \sigma^2)$-distributed. The parameters are μ and σ^2, where μ is the expected value (or population mean) and σ is the average deviation from this value (or the population standard deviation).

Estimate of parameter. Consider one sample model, and assume that $y_1, \ldots, y_n$ are independent and distributed as $N(\mu, \sigma^2)$. The only parameter of interest is the common mean μ, which is estimated by

$$\hat{\mu} = \bar{y} = \frac{\sum_{i=1}^n y_i}{n} \sim N(\mu, \sigma^2/n)$$

The estimate $\hat{\mu}$ is *unbiased* (that is, it "hits" the correct value on average, if we repeated the experiment many times) and *consistent* (that is, it becomes more precise as the sample size increases since the variance decreases). Furthermore, it has a normal distribution, so we know how to calculate probability concerning the estimate.

Standard error (SE). The estimate $\hat{\sigma}$ of standard deviation σ is given by the sample standard deviation s; i.e., the square root of

$$s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}$$

The statistical model results in the estimate $\hat{\mu}$ of parameter. The standard deviation of the estimate is called the **standard error (SE)**, and given by

$$SE(\hat{\mu}) = s\sqrt{1/n}$$

The constant $\sqrt{1/n}$ depends on the model and the data structure, but not on the observed data. Thus, the value $\sqrt{1/n}$ could be computed even before the experiment was carried out. In particular, it decreases when the sample size n increases.

Example: Crab weights. For the crab weight data we get the sample mean $\bar{y}$ and the sample standard deviation s by

$$\bar{y} = 12.76, \quad s = 2.25$$

Hence, the estimate $\hat{\mu}$ of mean parameter and the estimate $\hat{\sigma}$ of standard deviation become

$$\hat{\mu} = 12.76, \quad \hat{\sigma} = 2.25$$

And we obtain the standard error of the mean parameter estimate by

$$SE(\hat{\mu}) = (2.25)\sqrt{1/162} = 0.177$$

Standardization and test statistic. Remember that if $y_1, \ldots, y_n$ are independent and $N(\mu, \sigma^2)$-distributed, then the sample mean is normally distributed by the central limit theorem. Furthermore, by **standardization** (the Z-score) we get

$$Z = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

However, the value of σ is unknown, so we cannot compute the Z-score. If we replace σ with its estimate s and consider instead

$$T = \frac{\bar{y} - \mu}{s/\sqrt{n}}$$

then extra uncertainty is introduced through the estimate s of σ, and the distribution is changed. T is called a **test statistic**.

Distribution of test statistic. Intuitively we expect the following properties to hold for the distribution of test statistic

$$T = \frac{\bar{y} - \mu}{s/\sqrt{n}} = \frac{\sqrt{n}(\bar{y} - \mu)}{s}$$

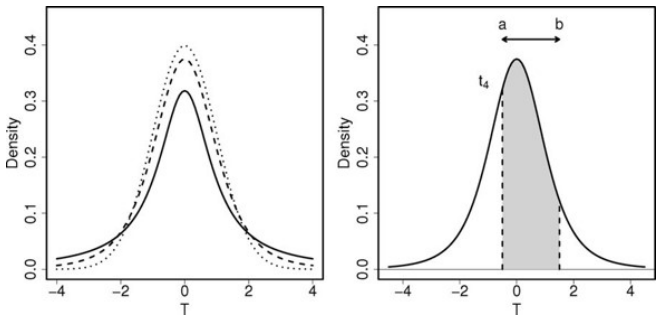
- **Symmetry.** The distribution of T is symmetric around zero, so positive and negative values are equally likely.
- **Dispersion.** Values far from zero are more likely for T than for Z due to the extra uncertainty. This implies that the interval should be wider for the probability to be retained at 0.95.
- **Large samples.** When the sample size increases then s is a more precise estimate of σ, and the distribution of T more closely resembles the standard normal distribution. In particular, the distribution of T should approach $N(0, 1)$ as n approaches infinity.

Student t distribution. It can be proven that these properties are indeed true. The distribution of T is called the t-distribution with $n - 1$ degrees of freedom and is denoted by t_{n-1}, so we write

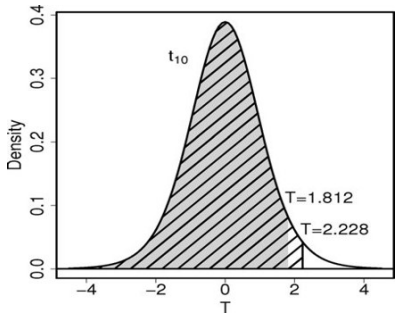
$$T = \frac{\sqrt{n}(\bar{y} - \mu)}{s} \sim t_{n-1}$$

The t-distribution is often called Student's t distribution because the distribution result was first published in 1908 under the pseudonym "Student." The author was a chemist, William S. Gosset, employed at the Guinness brewery in Dublin. Gosset worked with what we would today call quality control of the brewing process. Due to time constraints, small samples of 4 or 6 were used, and Gosset realized that the normal distribution was not the proper one to use. The Guinness brewery only let Gosset publish his results under pseudonym.

Density functions are compared for the t-distribution with $r = 1$ degree of freedom (solid) and $r = 4$ degrees of freedom (dashed) as well as for $N(0, 1)$ (dotted). The probability of an interval is the area under the density curve, illustrated in the figure below.



Critical values for t distribution. The graph below shows the density function of the t_{10}-distribution. The 95% quantile is 1.812, as illustrated by the gray region which has area 0.95. The 97.5% quantile is 2.228, illustrated by the dashed region with area 0.975. These quantiles are called **critical values** of level 0.05 and 0.025, respectively. The level α corresponds to the right tail probability of the $(1-\alpha)$-quantile, and the critical value of level α is denoted by $t_{\alpha,r}$ with r degrees of freedom.



The t distribution table. The following table shows the 95% and 97.5% quantiles for the t-distribution for a few selected degrees of freedom, denoted by t_r as illustrated for different degrees r of freedom. Equivalently they present the critical values $t_{0.05,r}$ of level 0.05 (the top row) and $t_{0.025,r}$ of level 0.025 (the bottom row). For data analyses where other degrees of freedom are in order, you should look up the relevant quantiles/critical values in a statistical table, called the **t-distribution table**.

Quantile	t_1	t_2	t_5	t_{10}	t_{20}	t_{50}	t_{100}	$N(0,1)$
95%	6.314	2.920	2.015	1.812	1.725	1.676	1.660	1.645
97.5%	12.706	4.303	2.571	2.228	2.086	2.009	1.984	1.960

Confidence interval. In one sample model let $y_1, \ldots, y_n$ be independent and $N(\mu, \sigma^2)$-distributed. The estimate of parameters μ and σ^2 are given by

$$\hat{\mu} = \bar{y} = \frac{\sum_{i=1}^n y_i}{n}, \quad \hat{\sigma}^2 = s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}$$

and hence the standard error of the mean parameter estimate $\hat{\mu}$ is given by

$$SE(\hat{\mu}) = s\sqrt{1/n}$$

By standardization we obtain

$$T = \frac{\bar{y} - \mu}{s/\sqrt{n}} = \frac{\hat{\mu} - \mu}{SE(\hat{\mu})} \sim t_{n-1}$$

If we denote the critical value in the t-distribution by $t_{\alpha/2,n-1}$ then we get

$$P\left(-t_{\alpha/2,n-1} < \frac{\hat{\mu} - \mu}{SE(\hat{\mu})} < t_{\alpha/2,n-1}\right) = 1 - \alpha$$

If we move around terms in order to isolate μ, we can derive

$$\begin{aligned} 1 - \alpha &= P\Big(-(t_{\alpha/2,n-1})SE(\hat{\mu}) < \hat{\mu} - \mu < (t_{\alpha/2,n-1})SE(\hat{\mu})\Big) \\ &= P\Big(\hat{\mu} - (t_{\alpha/2,n-1})SE(\hat{\mu}) < \mu < \hat{\mu} + (t_{\alpha/2,n-1})SE(\hat{\mu})\Big) \end{aligned}$$

Therefore, the interval

$$\Big(\hat{\mu} - (t_{\alpha/2,n-1})SE(\hat{\mu}),\ \hat{\mu} + (t_{\alpha/2,n-1})SE(\hat{\mu})\Big)$$

includes the true parameter value μ with a probability of $(1-\alpha)\%$. The interval is called a $(1-\alpha)\%$ **confidence interval** for μ.

Example: Crab weights. The interesting parameter for the crabs data is the population mean μ. Choose $(1-\alpha) = 0.95$. Then the critical value with level $\alpha/2 = 0.025$ in the t-distribution with $(n-1) = 161$ degrees of freedom becomes 1.975. From the estimates and standard error

$$\hat{\mu} = 12.76, \quad s = 2.25, \quad SE(\hat{\mu}) = (2.25)\sqrt{1/162} = 0.177$$

we compute the 95% confidence interval for μ by

$$12.76 \pm (1.975)(0.177) = (12.41,\ 13.11)$$

If n is large then the critical value will be close to 1.96 and it does not matter much which of the critical values, 1.975 or 1.96, is used. Moreover, s will be a precise estimate of σ when n is large, so it will be almost as if s was known. For small samples there is a difference, though, and the critical value $t_{0.025,n-1}$ is more correct to use.

Find the estimate and the standard error for the variable "wgt."

```
databox = read.csv(file.choose())
attach(databox)
est = mean(wgt)
se = sd(wgt) * sqrt(1/length(wgt))
```

Quantiles for the t-distribution are computed with the `qt()` function with degrees of freedom. For example, the 95% and the 97.5% quantiles with $n-1$ degrees of freedom are

```
qt(0.95, df=length(wgt)-1)
qt(0.975, df=length(wgt)-1)
```

The one sample model can be handled with the `lm()` function, acknowledging that there is only "one constant" parameter.

```
model = lm(wgt ~ 1)
summary(model)
confint(model, level=0.90)
confint(model, level=0.95)
```

Alternatively (and more easily in some cases) we can use the `t.test()` function:

```
t.test(wgt)
t.test(wgt, conf.level=0.90)
```

The output from `t.test()` is mainly concerned with t-test and will be explained later. But the estimate and the confidence interval are also produced. The confidence level can be specified with the option "conf.level=0.90."

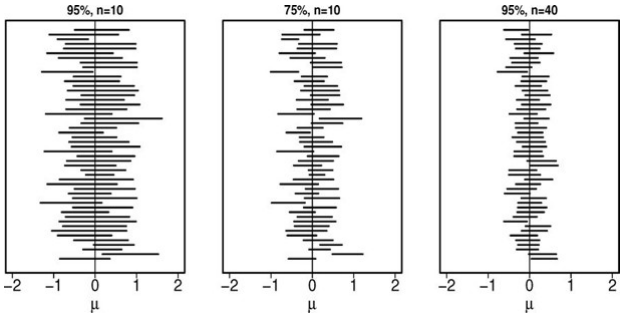
Confidence interval in general. If we repeated the experiment or data collection procedure many times and computed the interval

$$\hat{\mu} \pm (t_{0.025,n-1})SE(\hat{\mu})$$

then 95% of those intervals would include the true value of μ. We have drawn 50 samples of size 10 with $\mu = 0$, and for each of these 50 samples we have computed and plotted the confidence interval. The true value $\mu = 0$ is included in the confidence interval 95% of the time. The 75% confidence interval for μ is given by

$$\hat{\mu} \pm (t_{0.125,n-1})SE(\hat{\mu})$$

The 75% confidence intervals are more narrow such that the true value is excluded more often, with a probability of 25% rather than 5%. This reflects that our confidence in the 75% confidence interval is smaller compared to the 95% confidence interval.



Confidence intervals for 50 simulated data generated for the true value $\mu = 0$. Each shows 95% confidence intervals of size $n = 10$, 75% confidence intervals of size $n = 10$, and 95% confidence intervals of size $n = 40$, respectively.

Confidence interval in general. The true value μ is either inside the interval or it is not, but we will never know. We can, however, interpret the values in the confidence interval for the mean parameter μ by which it is reasonable to believe that they could have generated the data. If we use 95% confidence intervals,

1. the probability of observing data for which the corresponding confidence interval includes μ is 95%;

2. the probability of observing data for which the corresponding confidence interval does not include μ is 5%.

As a standard phrase we may say, "*the 95% confidence interval includes those values that are in agreement with the data on the 95% confidence level.*"

Example: Hormone concentration study. As part of a larger cattle study, the effect of a particular type of feed on the concentration of a certain hormone was investigated. Nine cows were given the feed for a period, and the hormone concentration was measured initially and at the end of the period. The purpose of the experiment was to examine if the feed changes the hormone concentration.

Cow	1	2	3	4	5	6	7	8	9
Initial (µg/ml)	207	196	217	210	202	201	214	223	190
Final (µg/ml)	216	199	256	234	203	214	225	255	182
Difference d	9	3	39	24	1	13	11	32	-8

We obtain the sample mean 13.778 and the sample standard deviation 15.238. Thus, the difference is positive for eight of the nine cows. But is it strong enough for us to conclude that the feed affects the concentration?

Hypothesis test. Initially, we may examine if the concentrations are generally increasing or decreasing from the start to the end of the experiment. The hormone concentration increases for eight of the nine cows, for some cows quite substantially, for some cows only a little bit. There is certainly a tendency, but is it strong enough for us to conclude that the feed affects the concentration? **Hypothesis tests** are used to investigate if the observed data contradict or support specific assumptions. In short, a hypothesis test evaluates how likely the observed data is if the assumptions under investigation are true. If the data is very unlikely to occur given the assumptions, then we do not believe in the assumptions. Hypothesis tests form the core of statistical inference, together with parameter estimation and confidence intervals, and involve important new concepts like **null and alternative hypotheses**, **test statistics**, and **p-values**.

In many cases, the interest is in identifying certain effects. This situation corresponds to the **alternative hypothesis**, whereas the **null hypothesis** corresponds to the situation of "no effect" or "no association." This may all seem a little counterintuitive, but the machinery works like this: with a hypothesis test we **reject null hypothesis** if the data and the hypothesis are in contradiction; that is, if the model under the null hypothesis fits poorly to the data. Hence, if we reject the null hypothesis then we believe in the alternative, which states that **there is an effect.** In principle we never accept the null hypothesis. If we fail to reject the null hypothesis we say that the data does not provide evidence against it. This is not a proof that the null hypothesis is true, but it only indicates that the model under the alternative hypothesis does not describe the data **significantly better** than the one under the null hypothesis.

Example: Hormone concentration study. We consider the differences of hormone concentration, denoted by $d_1, \ldots, d_9$, during the period and assume that they are independent. Then

μ is the expected change in hormone concentration for a random cow, or the average change in the population, and $\mu = 0$ corresponds to no effect of the feed on the hormone concentration. The mean of the nine differences is 13.78, and we would like to know if this reflects a real effect or if it might be due to chance. Here we are interested in the **null hypothesis**

$$H_0 : \mu = 0$$

corresponding to "no difference" between start and end measurements in the population.

Example: Hormone concentration study. Now, the average difference 13.78 is our best "guess" for μ, so it seems reasonable to believe in the null hypothesis if it is "close to zero" and not believe in the null hypothesis if it is "far away from zero." So we might ask: If μ is really zero (if the hypothesis is true), then how likely is it to get an estimate that is as far or even further away from zero than the 13.78 that we actually got? A **t-test** can answer that question. We can calculate the **test statistic** by

$$T_{\text{obs}} = \frac{\hat{\mu} - 0}{SE(\hat{\mu})} = \frac{13.78}{5.08} = 2.71$$

The numerator is just the difference between the estimate $\hat{\mu}$ and the value of μ if the null hypothesis H_0 is true. We then divide it by the standard error

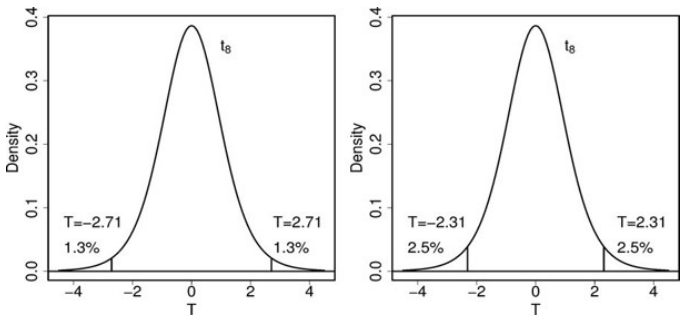
$$SE(\hat{\mu}) = (15.24)\sqrt{1/9} = 5.08$$

Mechanism of rejection. If $\mu = 0$ (the null hypothesis H_0 is true) then T_{obs} is an observation from the t_8 distribution (the number of degrees of freedom is $n-1$; here $9-1=8$). In particular, we can compute the probability of getting a test statistic T which is at least as extreme as 2.71 by

$$P(|T| \geq 2.71) = (2)P(T \geq 2.71) = (2)(0.013) = 0.026$$

Here T is a t_8-distributed random variable and the second equality follows from the symmetry of the t-distribution. This probability is called the **p-value**. If the p-value is small then the observed value $T_{\text{obs}} = 2.71$ is extreme, so we do not believe in the null hypothesis H_0 and reject it. If, on the other hand, the p-value is large then the observed value $T_{\text{obs}} = 2.71$ is quite likely, so we cannot reject the hypothesis. In this case the p-value is only 2.6%, so the observed test statistic of 2.71 is unlikely if the true value of μ is zero. Hence, we **reject** the null hypothesis.

Usually we use a **significance level** of 5%; that is, we reject the hypothesis if the p-value is less than 0.05 and fail to reject it otherwise. This means that the null hypothesis is rejected if $|T_{\text{obs}}|$ is larger than or equal to the critical value $t_{0.025,n-1}$ with level 0.025, which is 2.31 in this case. Hence, an observed test statistic T_{obs} outside $(-2.31, 2.31)$ leads to rejection of the null hypothesis.



Example: Hormone concentration study. We already used the `t.test()` function for computation of estimates and confidence intervals for one sample model. As the name suggests, the function also carries out complete t-tests.

```
databox = read.csv(file.choose())
attach(databox)
t.test(difference)
```

We recognize the estimate, the confidence interval, the t-test statistic, and the p-value from the example. Notice that the default "alternative" hypothesis is that the true mean is not equal to zero. If we are interested in the alternative hypothesis that the mean is greater than zero, then we need to specify it in the call to `t.test()`.

```
t.test(difference, alternative="greater")
```

Concepts of hypothesis test.

- **Null hypothesis.** A null hypothesis is a simplification of the statistical model and is as such always related to the statistical model. Hence, no null hypothesis exists without a corresponding statistical model. A null hypothesis typically describes the situation of "no effect" or "no relationship," such that rejection of the null hypothesis corresponds to evidence of an effect or relationship.
- **Alternative hypothesis.** There is a corresponding alternative hypothesis to every null hypothesis. The alternative hypothesis describes what is true if the null hypothesis is false. Usually the alternative hypothesis is simply the complement of the null hypothesis, called **two-sided**. If the alternative is **one-sided** then the one side of values expressed in the alternative is considered as evidence against the null hypothesis.

- **Test statistic.** A test statistic is a function of the data that measures the discrepancy between the data and the null hypothesis—with certain values contradicting the hypothesis and others supporting it. Values contradicting the hypothesis are called **critical** or **extreme**.
- **p-value.** The test statistic is translated to a p-value: the probability of observing data which fit as bad or even worse than the observed data if the null hypothesis is true. A small p-value indicates that the observed data are unusual if the null hypothesis is true, hence that the hypothesis is false.

- **Rejection.** The hypothesis is rejected if the p-value is small; namely, below (or equal to) the **significance level**, which is often taken to be 0.05. With statistics we can at best reject the null hypothesis with strong certainty, but we can never confirm the hypothesis. If we fail to reject the null hypothesis, then the only valid conclusion is that the data do not contradict the null hypothesis. A large p-value shows that the data are in fine accordance with the null hypothesis, but not that it is true.

- **Quantification of effects.** Having established a significant effect by a hypothesis test, it is of great importance to quantify the effect. For example, how much larger is the expected hormone concentration after a period of treatment? Moreover, what is the precision of the estimates in terms of standard errors and/or confidence intervals?

t-test in general. Consider the null hypothesis

$$H_0 : \mu = \mu_0$$

with a **null value** μ_0. Data for which the estimate $\hat{\mu}$ is close to μ_0 support the null hypothesis H_0, whereas data for which the estimate $\hat{\mu}$ is far from μ_0 contradict the null hypothesis; so it seems reasonable to consider the deviation.

$$T_{\text{obs}} = \frac{\hat{\mu} - \mu_0}{SE(\hat{\mu})}$$

can be used as a **test statistic**. An extreme value of T_{obs} is an indication that the data are unusual under the null hypothesis, and the p-value measures how extreme T_{obs} is compared to the t_{n-1}-distribution.

If the alternative is two-sided, $H_A : \mu \neq \mu_0$, then values of T_{obs} that are far from zero (both small and large values) are critical. Therefore, the p-value is

$$P(|T| \geq |T_{\text{obs}}|) = (2)P(T \geq T_{\text{obs}})$$

where T follows the t_{n-1}-distribution. If the alternative is one-sided, $H_A : \mu > \mu_0$, then large values of T_{obs} are critical, whereas negative values of T_{obs} are considered in favor of the null hypothesis rather than evidence against it. Hence the p-value is $P(T \geq T_{\text{obs}})$. Similarly if the alternative is one-sided, $H_A : \mu < \mu_0$, then only small values of T_{obs} are critical, so the p-value is $P(T \leq T_{\text{obs}})$. The significance level is usually denoted α, and it should be selected before the analysis. Tests are often carried out on the 5% level corresponding to $\alpha = 0.05$ (but $\alpha = 0.01$ and $\alpha = 0.10$ are not unusual).

Critical region. For a hypothesis with a two-sided alternative, the hypothesis is thus rejected on the 5% significance level if $|T_{\text{obs}}|$ is numerically larger than or equal to the critical value with level 0.025 in the t_{n-1}-distribution; that is, if $|T_{\text{obs}}| \geq t_{0.025,n-1}$. With a one-sided alternative, $H_A : \mu > \mu_0$, the null hypothesis H_0 is rejected if $T_{\text{obs}} \geq t_{0.05,n-1}$. Otherwise, we fail to reject the hypothesis, and the model under the alternative hypothesis does not describe the data significantly better than the model under the null hypothesis. Similarly with the choice of one-sided alternative $H_A : \mu < \mu_0$, the null hypothesis H_0 is rejected if $T_{\text{obs}} \leq -t_{0.05,n-1}$. In order to evaluate if the null hypothesis should be rejected or not, it is thus enough to compare T_{obs} or $|T_{\text{obs}}|$ to a certain critical value. But we recommend that the p-value is always reported.

Example: Production control. A dairy company bought a new machine for filling milk into cartons and wants to make sure that the machine is calibrated correctly. The aim is an average weight of 1070 grams per carton (including the carton). A sample of 100 cartons with milk is chosen at random from the production line and each carton is weighed. It is assumed that the weights are independent and $N(\mu, \sigma^2)$-distributed. The relevant hypothesis is $H_0 : \mu = 1070$. It turned out that

$$\bar{y} = 1072.9\text{grams}, \quad s = 15.8\text{grams},$$

and therefore, that $SE(\hat{\mu}) = 1.58$ and the t-test statistic becomes

$$T_{\text{obs}} = \frac{\hat{\mu} - 1070}{SE(\hat{\mu})} = \frac{2.9}{1.58} = 1.83$$

The corresponding p-value is

$$P(|T| \geq 1.83) = (2)P(T \geq 1.83) = 0.07$$

where T follows t_{99}-distribution. Thus, we fail to reject the hypothesis on the 5% significance level, but due to the low p-value we conclude nevertheless that there is a slight indication that the machine is calibrated incorrectly. Thus, the result has some significance since p-value is between 0.05 and 0.1.

Type I and type II errors. Four scenarios are possible as we carry out a hypothesis test: the null hypothesis is either true or false, and it is either rejected or not rejected. The conclusion is correct whenever we reject a false hypothesis or do not reject a true hypothesis. Rejection of a true hypothesis is called a **type I error**, whereas a **type II error** refers to not rejecting a false hypothesis; see the chart below.

	H_0 is true	H_0 is false
Reject	Type I error	Correct conclusion
Fail to reject	Correct conclusion	Type II error

We use a 5% significance level. Then we reject the hypothesis if p-value ≤ 0.05. This means that if the hypothesis is true, then we will reject it with a probability of 5%. In other words: The probability of committing a type I error is the significance level $\alpha = 0.05$.

The situation is analogous to the situation of a medical test: Assume for example that the concentration of some substance in the blood is measured in order to detect cancer. (Thus, the claim is that the patient has cancer, and the null hypothesis is that he or she is cancer-free.) If the concentration is larger than a certain threshold, then the "alarm goes off" and the patient is sent for further investigation. (That is, to reject the null hypothesis, and conclude that the patient has cancer.) But how large should the threshold be? If it is large, then some patients will not be classified as sick (failed to reject the null hypothesis) although they are sick due to cancer (type II error). On the other hand, if the threshold is low, then patients will be classified as sick (reject the null hypothesis) although they are not (type I error).

For a general significance level α, the probability of committing a type I error is α. Hence, by adjusting the significance level we can change the probability β of rejecting a true hypothesis. This is not for free, however. If we decrease α we make it harder to reject a hypothesis. Hence we will accept more false hypotheses, so the rate of type II errors will increase. The probability that a false hypothesis is rejected is called the power of the test, and it is given by $(1 - \beta)$. We would like the test to have large power $(1 - \beta)$ and at the same time a small significance level α, but these two goals contradict each other so there is a trade-off. As mentioned already, $\alpha = 0.05$ is the typical choice. Sometimes, however, the scientist wants to "make sure" that

false hypotheses are really detected; then α can be increased to 0.10, say. On the other hand, it is sometimes more important to "make sure" that rejection expresses real effects; then α can be decreased to 0.01.

t-tests and confidence intervals. The null hypothesis $H_0 : \mu = \mu_0$ is rejected on significance level α against the alternative $H_A : \mu \neq \mu_0$ if and only if the null value μ_0 is not included in the $(1 - \alpha)\%$ confidence interval. This relationship explains the formulation about confidence intervals; namely, that a confidence interval includes the values that are in accordance with the data. This now has a precise meaning in terms of hypothesis tests. If the only aim of the analysis is to conclude whether a hypothesis should be rejected or not at a certain level α, then we get that information from either the t-test or the confidence interval. On the other hand, they provide extra information on slightly different matters. The t-test provides a p-value explaining how extreme the observed data are if the hypothesis is true, whereas the confidence interval gives us the values of μ that are in agreement with the data.

Example: Heart rate reductions. A researcher is interested in how a new class of drug treating a patient actually affects the patient's heart rate reduction. The pairs of heart rate reduction of 40 participants under the standard drug and after taking the new drug are measured. The data set of heart rate reductions is prepared in a text file format.

```
databox <- read.csv(file.choose(), header=T)
attach(databox)
```

The paired sample test can be done by the `t.test()` command with the option "paired=T." Suppose that we want to test whether the true means of heart rate reductions with the new drug are higher. Then the t.test command can be used as follows.

```
t.test(StdDrug, NewDrug, alternative="less", paired=T)
```