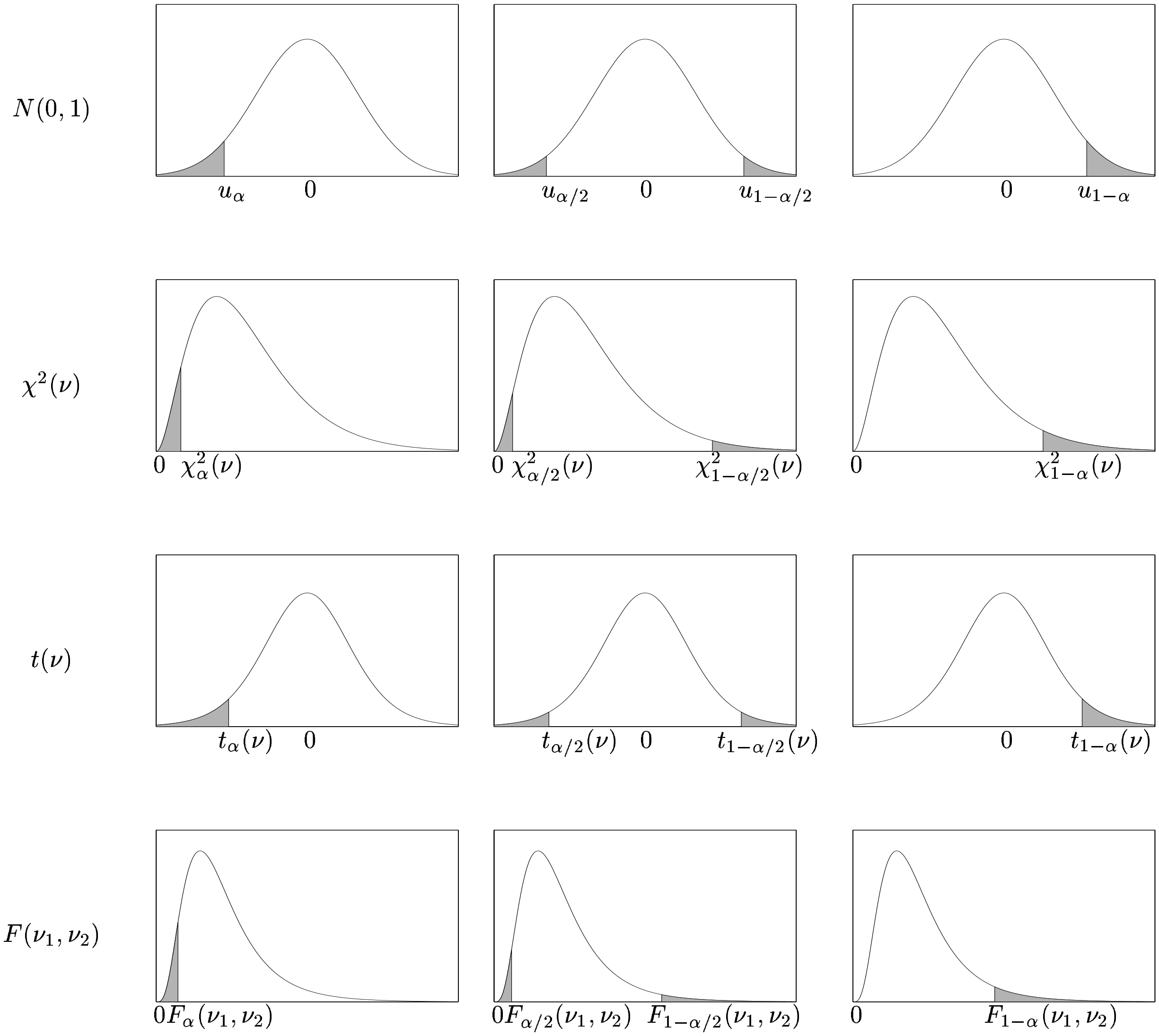
6 The statistical inferences based on a single sample from the normal distribution

It is not very difficult to find random variables which refer to natural or social phenomena and which are - or can be assumed to be - normally distributed. Then in case of not normally distributed random variables if the sample is large, we can invoke the central limit theorem. Thus we can obtain approximately normal distribution. Therefore it is necessary to pay great attention to random samples from normal distribution.
Normal distribution is fully specified by two parameters, mean μ and variance σ^2. Thus we are going to follow the tasks concerning with these parameters, e.g. forming the confidence intervals or hypothesis testing.

For simple random sample from normal distribution the following theorem states the list of common test statistics and their distributions:

Theorem 6.1
Let $X_1, \ldots, X_n$ be a random sample from normal distribution $N(\mu, \sigma^2)$. Then:

1. The sample mean $M = \sum_{i=1}^{n} X_i$ and the sample variance $S^2 = \sum_{i=1}^{n} (X_i - M)^2$ are mutually independent.
2. $U = \frac{M-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$, thus $M \sim N(\mu,\ \frac{\sigma^2}{n})$
   [Pivotal statistic U is instrumental towards inferences about μ, when σ^2 is known.]
3. $K = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$
   [Pivotal statistic K is instrumental towards inferences about σ^2, when μ is unknown.]
4. $T = \frac{M-\mu}{\frac{S}{\sqrt{n}}} \sim t(n-1)$
   [Pivotal statistic T is instrumental towards inferences about μ, when σ^2 is unknown.]
5. $\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n)$
   [This pivotal statistic is instrumental towards inferences about σ^2, when μ is known.]



$u_\alpha = -u_{1-\alpha}$ $\qquad$ $t_\alpha(\nu) = -t_{1-\alpha}(\nu)$ $\qquad$ $F_\alpha(\nu_1, \nu_2) = \frac{1}{F_{1-\alpha}(\nu_2, \nu_1)}$

Example 6.2
A weight of a packet of granulated sugar follows the normal distribution $N(1002\,g,\ 64\,g^2)$. The inspection draws randomly 9 packets of one series and is finding if the average weight is at least $999\,g$. Otherwise the enterprise has to pay a penalty. Find the probability that the enterprise will have to pay the penalty.

Solution
$X_1, \dots, X_9 \sim N(1002,\ 64)$, $M \sim N(1002,\ \frac{64}{9})$, $P(M \leq 999) =?$
$P(M \leq 999) = P(\frac{M-1002}{\sqrt{\frac{64}{9}}} \leq \frac{999-1002}{\sqrt{\frac{64}{9}}}) = P(U \leq \frac{-9}{8}) = 1 - \Phi(\frac{9}{8}) = 1 - \Phi(1,125) = 1 - 0,87076 = 0,12924$.
The probability, that the enterprise will pay the penalty is approximately 12,9%.

The common statistician's task is to derive confidence intervals for unknown parameters. In case of the normal distribution they are parameters μ and σ^2, thus four situations may occur: finding the confidence interval 1. for μ, when σ^2 is known; 2. for σ^2, when μ is unknown; 3. for μ, when σ^2 is unknown a 4. for σ^2, when μ is known. Doing confidence interval in accordance to one of four mentioned situations the appropriate pivotal statistic has to be selected . Then using the procedure 4.9, the construction of the confidence interval is easy. In case of the first situation it has been done in the example 4.11. The following theorem states the upper and lower limits of the confidence intervals for any mentioned situation.

Theorem 6.3
Let $X_1, \ldots, X_n$ be a random sample from normal distribution $N(\mu, \sigma^2)$. Let us consider $100(1-\alpha)\%$ empirical confidence interval.

1. The confidence interval for μ, when σ^2 is known is derived from pivotal statistic $U = \frac{M-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$. Thus the limits are for:

$$\begin{array}{lcl} \text{two-sided conf. int. } (d,h) & = & \left(m - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2} \quad , \quad m + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}\right) \\ \text{left-sided conf. int. } (d,\infty) & = & \left(m - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha} \quad , \quad \infty\right) \\ \text{right-sided conf. int. } (-\infty,h) & = & \left(-\infty \quad , \quad m + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha}\right) \end{array}$$

2. The confidence interval for σ^2, when μ is unknown is derived from pivotal statistic $K = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$. Thus the limits are for:

$$\begin{array}{lcl} \text{two-sided conf. int. } (d,h) & = & \left(\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)} \quad , \quad \frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)}\right) \\ \text{left-sided conf. int. } (d,\infty) & = & \left(\frac{(n-1)s^2}{\chi^2_{1-\alpha}(n-1)} \quad , \quad \infty\right) \\ \text{right-sided conf. int. } (-\infty,h) & = & \left(-\infty \quad , \quad \frac{(n-1)s^2}{\chi^2_{\alpha}(n-1)}\right) \end{array}$$

3. The confidence interval for μ, when σ^2 is unknown is derived from pivotal statistic $T = \frac{M-\mu}{\frac{s}{\sqrt{n}}} \sim t(n-1)$. Thus the limits are for:

$$\begin{array}{lcl} \text{two-sided conf. int. } (d,h) & = & \left(m - \frac{s}{\sqrt{n}} \cdot t_{1-\alpha/2}(n-1) \ , \ m + \frac{s}{\sqrt{n}} \cdot t_{1-\alpha/2}(n-1)\right) \\ \text{left-sided conf. int. } (d,\infty) & = & \left(m - \frac{s}{\sqrt{n}} \cdot t_{1-\alpha}(n-1) \quad , \quad \infty\right) \\ \text{right-sided conf. int. } (-\infty,h) & = & \left(-\infty \quad , \quad m + \frac{s}{\sqrt{n}} \cdot t_{1-\alpha}(n-1)\right) \end{array}$$

4. The confidence interval for σ^2, when μ is known is derived from pivotal statistic $\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n)$. Thus the limits are for:

$$\begin{array}{lcl} \text{two-sided conf. int. } (d,h) & = & \left(\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_{1-\alpha/2}(n)} \quad , \quad \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_{\alpha/2}(n)}\right) \\ \text{left-sided conf. int. } (d,\infty) & = & \left(\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_{1-\alpha}(n)} \quad , \quad \infty\right) \\ \text{right-sided conf. int. } (-\infty,h) & = & \left(-\infty \quad , \quad \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_{\alpha}(n)}\right) \end{array}$$

Example 6.4
The constant μ was measured 10 times independently. The results of measuring are:

$2 \quad 1,8 \quad 2,1 \quad 2,4 \quad 1,9 \quad 2,1 \quad 2 \quad 1,8 \quad 2,3 \quad 2,2$

These results are assumed to be the numerical realization of a random sample $X_1, \ldots, X_n$ from distribution $N(\mu, \sigma^2)$ where parameters μ, σ^2 are unknown. Find the 95% confidence interval for the parameter μ a) two-sided, b) left-sided, c) right-sided.

Solution

It is the confidence interval for μ when σ^2 is unknown. The statistic T is instrumental to deriving confidence limits, $T = \frac{M-\mu}{\frac{s}{\sqrt{n}}} \sim t(n-1)$ whose $\alpha-$ quantiles are looked up in table.

$n = 10 \qquad \alpha = 0,05 \qquad t_{1-\alpha/2}(n-1) = t_{0,975}(9) = 2,2622$

$t_{1-\alpha}(n-1) = t_{0,95}(9) = 1,8331$

$m = 2,06 \qquad s^2 = 0.0404 \qquad s = 0.2011$

---

ad a)

$d = m - \frac{s}{\sqrt{n}} \cdot t_{1-\alpha/2}(n-1) = 2,06 - \frac{0,2011}{\sqrt{10}} \cdot 2,2622 = 1,92$

$h = m + \frac{s}{\sqrt{n}} \cdot t_{1-\alpha/2}(n-1) = 2,06 + \frac{0,2011}{\sqrt{10}} \cdot 2,2622 = 2,20$

$$1,92 < \mu < 2,2 \text{ with the probability at least 0,95}$$

ad b)

$d = m - \frac{s}{\sqrt{n}} \cdot t_{1-\alpha}(n-1) = 2,06 - \frac{0,2011}{\sqrt{10}} \cdot 1,8331 = 1,94$

$$1,94 < \mu \text{ with the probability at least 0,95}$$

ad c)

$h = m + \frac{s}{\sqrt{n}} \cdot t_{1-\alpha/2}(n-1) = 2,06 + \frac{0,2011}{\sqrt{10}} \cdot 1,8331 = 2,18$

$$\mu < 2,18 \text{ with the probability at least 0,95}$$

So much for confidence intervals and now let us turn to hypothesis testing. We will follow the classical method using critical region; the other methods can be derived easily.

Definition 6.5

Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$, where σ^2 is known. Let $n \geq 2$ and c is a constant. Test $H_0: \ \mu = c$ versus $H_1: \ \mu \neq c$ (eventually $H_1: \ \mu < c$ eventually $H_1: \ \mu > c$) is called *z-test*.

Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$, where σ^2 is unknown. Let $n \geq 2$ and c is a constant. Test $H_0: \ \mu = c$ versus $H_1: \ \mu \neq c$ (eventually $H_1: \ \mu < c$ eventually $H_1: \ \mu > c$) is called one-sample *t-test*.

Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$, where μ is unknown. Let $n \geq 2$ and c is a constant. Test $H_0: \ \sigma^2 = c$ versus $H_1: \ \sigma^2 \neq c$ (eventually $H_1: \ \sigma^2 < c$ eventually $H_1: \ \sigma^2 > c$) is called *test about variance*.

Remark 6.6

The selection of an appropriate test statistic corresponding to particular test is analogous to the selection of an appropriate pivotal statistic in 6.3, thus for z-test the test statistic T_0 is derived from statistic U, for t-test it is derived from statistic T and for test about variance it is derived from statistic K.

Beware of ambiguity of a letter T. In general T_0 stands for any test statistic; in case of t-test T stands for statistic following Student's t-distribution. Under the null hypothesis it can be written $T_0 = U$, $T_0 = T$, $T_0 = K$

Theorem 6.7

Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, $c \in \mathbf{R}, \ n \geq 2$

1. Considering z-test at the significance level α the null hypothesis H_0 is rejected in favor of alternative hypothesis H_1, if the realization of the test statistic $T_0 = \frac{M-c}{\frac{\sigma}{\sqrt{n}}}$ falls within critical region W. According to the form of the alternative hypothesis the list of corresponding critical regions follows :

two-tailed test	$H_1: \ \mu \neq c$	$W = (-\infty, \ -u_{1-\alpha/2}\rangle \bigcup \langle u_{1-\alpha/2}, \ \infty)$
left-tailed test	$H_1: \ \mu < c$	$W = (-\infty, \ -u_{1-\alpha}\rangle$
right-tailed test	$H_1: \ \mu > c$	$W = \langle u_{1-\alpha}, \ \infty)$

2. Considering t-test at the significance level α the null hypothesis H_0 is rejected in favor of alternative hypothesis H_1, if the realization of the test statistic $T_0 = \frac{M-c}{\frac{S}{\sqrt{n}}}$ falls within critical region W.

two-tailed test	$H_1: \ \mu \neq c$	$W = (-\infty, \ -t_{1-\alpha/2}(n-1)\rangle \bigcup \langle t_{1-\alpha/2}(n-1), \ \infty)$
left-tailed test	$H_1: \ \mu < c$	$W = (-\infty, \ -t_{1-\alpha}(n-1)\rangle$
right-tailed test	$H_1: \ \mu > c$	$W = \langle t_{1-\alpha}(n-1), \ \infty)$

3. Considering test about variance at the significance level α the null hypothesis H_0 is rejected in favor of alternative hypothesis H_1, if the realization of the test statistic $T_0 = \frac{(n-1)S^2}{c}$ falls within critical region W.

two-tailed test	$H_1: \ \sigma^2 \neq c$	$W = (0, \ \chi^2_{\alpha/2}(n-1)\rangle \bigcup \langle \chi^2_{1-\alpha/2}(n-1), \ \infty)$
left-tailed test	$H_1: \ \sigma^2 < c$	$W = (0, \ \chi^2_{\alpha}(n-1)\rangle$
right-tailed test	$H_1: \ \sigma^2 > c$	$W = \langle \chi^2_{1-\alpha}(n-1), \ \infty)$

Example 6.8
According to the chocolate-wrapper, the net weight of chocolate should be 125 g. Manufacturer recorded buyers's complaints of lower weight then it was declared. For that reason the audit division drawn randomly 50 chocolates and found out the mean weight was 122g and the standard deviation was 8.6 g. Assuming that the weight of chocolates follows the normal distribution and using the significance level $\alpha = 0.01$ can we conclude that the buyer's complaints are true?

Solution
$X_1, \ldots, X_{50} \sim N(\mu, \sigma^2)$. We are testing $H_0 : \mu = 125$ versus $H_1 : \mu < 125$. Parameter σ^2 is unknown, thus the task leads to one-sample t-test.

The test statistic: $T_0 = \frac{M-c}{\frac{S}{\sqrt{n}}}$.
The numerical realization of it: $t_0 = \frac{122-125}{\frac{8,6}{\sqrt{50}}} = -2,4667$.
Critical region: $W = (-\infty, \ -t_{1-\alpha}(n-1)\rangle = (-\infty, \ -t_{0,99}(49)\rangle = (-\infty; \ -2,4049\rangle$
Since $t_0 \in W, \ H_0$ is rejected at the significance level 0,01.

The buyer's complaints can be concluded as true and the risk of an error is at most 1%.

Having random sample from two-dimensional normal distribution, this can be convert to single normal sample. Then the above stated inferences can be used.

Poznámka 6.9 on random sample from two-dimensional normal distribution
Let $\binom{X_1}{Y_1}, \ldots, \binom{X_n}{Y_n} \sim N_2\left(\binom{\mu_1}{\mu_2}, \ \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right), \ n \geq 2.$
Using linear transformation the random sample $\binom{X}{Y}$ is converted to scalar random variable $Z = (X - Y) \sim N((\mu_1 - \mu_2), \ (\sigma_1^2 - 2\sigma_{12} + \sigma_2^2))$
Let us denote $\mu = \mu_1 - \mu_2 \qquad \sigma^2 = \sigma_1^2 - 2\sigma_{12} + \sigma_2^2$

Thus the random sample $(X_1 - Y_1), \ldots, (X_n - Y_n) = Z_1, \ldots, Z_n$ follows the normal distribution $N(\mu, \sigma^2)$ and is called říkáme mu *differential random sample* .

Theorem 6.10
Let $\binom{X_1}{Y_1}, \ldots, \binom{X_n}{Y_n} \sim N_2\left(\binom{\mu_1}{\mu_2}, \begin{pmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{pmatrix}\right)$, $n \geq 2$ and variance-covariance matrix Σ is unknown. Considering $100(1-\alpha)\%$ empirical confidence interval the confidence limits for the parametric function $\mu = \mu_1 - \mu_2$ have the form:
$d = m - \frac{s}{\sqrt{n}} \cdot t_{1-\alpha/2}(n-1)$
$h = m + \frac{s}{\sqrt{n}} \cdot t_{1-\alpha/2}(n-1)$

Example 6.11
The chemical content in solution were tested by two laboratory measurements. (Data are in percentages.) The random sample consist of 5 specimen.:

the number of specimen	1	2	3	4	5
1. method	2.3	1.9	2.1	2.4	2.6
2. method	2.4	2.0	2.0	2.3	2.5

Assuming the sample is selected from two-dimensional normal distribution determine 90% empirical confidence interval for difference between expected values of considered methods.

Solution
At first we transform the given sample to the differential sample, where:
$z_1 = -0.1 \quad z_2 = -0.1 \quad z_3 = 0.1 \quad z_4 = 0.1 \quad z_5 = 0.1$
$m = 0.2 \qquad s^2 = 0.012 \qquad s = 0.109545 \qquad n = 5 \qquad t_{1-\alpha/2}(n-1) = t_{0.95}(4) = 2.1318$

$d = m - \frac{s}{\sqrt{n}} \cdot t_{1-\alpha/2}(n-1) = 0,02 - \frac{0,109545}{\sqrt{5}} \cdot 2,1318 = -0.0844$

$h = m + \frac{s}{\sqrt{n}} \cdot t_{1-\alpha/2}(n-1) = 0,02 + \frac{0,109545}{\sqrt{5}} \cdot 2,1318 = 0,1244$

The confidence interval $-0,0844 < \mu < 0,1244$ is true with the probability at least 0.95.

Definition 6.12
Let $\binom{X_1}{Y_1}, \ldots, \binom{X_n}{Y_n} \sim N_2\left(\binom{\mu_1}{\mu_2}, \begin{pmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{pmatrix}\right)$, $n \geq 2$.
The test $H_0 : \mu_1 - \mu_2 = 0$ versus $H_1 : \mu_1 - \mu_2 \neq 0$ is called *paired t-test*. Using differential random sample the paired t-test is converted to single sample t-test. Přechodem k rozdílovému náhodnému výběru převedeme párový t-test na jednovýběrový t-test.

Example 6.13
The following table lists the rate of return on investment (in percentages) of 12 randomly drawn companies, whose foreign investments are represented by random variable X and domestic investments are represented by random variable Y:

a number of company	1	2	3	4	5	6	7	8	9	10	11	12
X	10	12	14	12	12	17	9	15	9	11	7	15
Y	11	14	15	11	13	16	10	13	11	17	9	19

Assuming the sample is selected from two-dimensional normal distribution, at a significance level $\alpha = 0.1$ run the test that there is no difference between foreign and domestic investment. Use a)confidence interval method, b) classical method.

Solution
At first we transform the given sample to the differential sample $Z_i = X_i - Y_i,\ i = 1, \ldots 12$. Realization of sample characteristics follows: $m = -1,33, \quad s^2 = 4,\overline{78}$
We are testing hypothesis $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$,

needed quantile $t_{0.95}(11) = 1,7959$

ad a)

$d = m - \frac{s}{\sqrt{n}} \cdot t_{1-\alpha/2}(n-1) = -1,\overline{3} - \frac{\sqrt{4,\overline{78}}}{\sqrt{12}} \cdot 1,7959 = -2,4677$

$h = m + \frac{s}{\sqrt{n}} \cdot t_{1-\alpha/2}(n-1) = -1,\overline{3} + \frac{\sqrt{4,\overline{78}}}{\sqrt{12}} \cdot 1,7959 = -0,1989$

•Since $0 \notin (-2,4677\ ,\ -0,1989)$, H_0 is rejected on the significance level 0,1.

ad b)

The test statistic follows: $T_0 = \frac{M-c}{\frac{S}{\sqrt{n}}}$.

The numerical realization follows: $t_0 = \frac{-1,\overline{3}-0}{\frac{\sqrt{4,\overline{78}}}{\sqrt{12}}} = -2,11085$.

The critical region follows: $W = (-\infty,\ -t_{1-\alpha/2}(n-1)\rangle \cup \langle t_{1-\alpha/2}(n-1),\ \infty) =$
$= (-\infty,\ -1,7959\rangle \cup \langle 1,7959,\ \infty)$

•Since $t_0 \in W$, H_0 is rejected on the significance level 0,1.