Teaching Mathematics and Statistics in Sciences, IPA HU-SRB/0901/221/088 - 2011

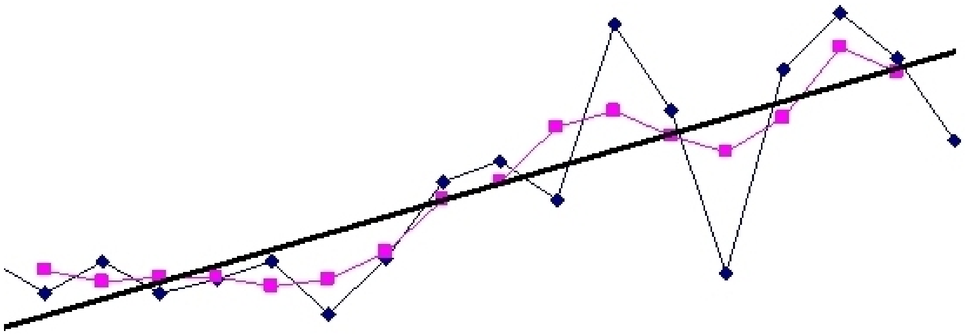
Author: **Tibor Nyári** PhD

University of Szeged
Department of Medical Physics and Informatics

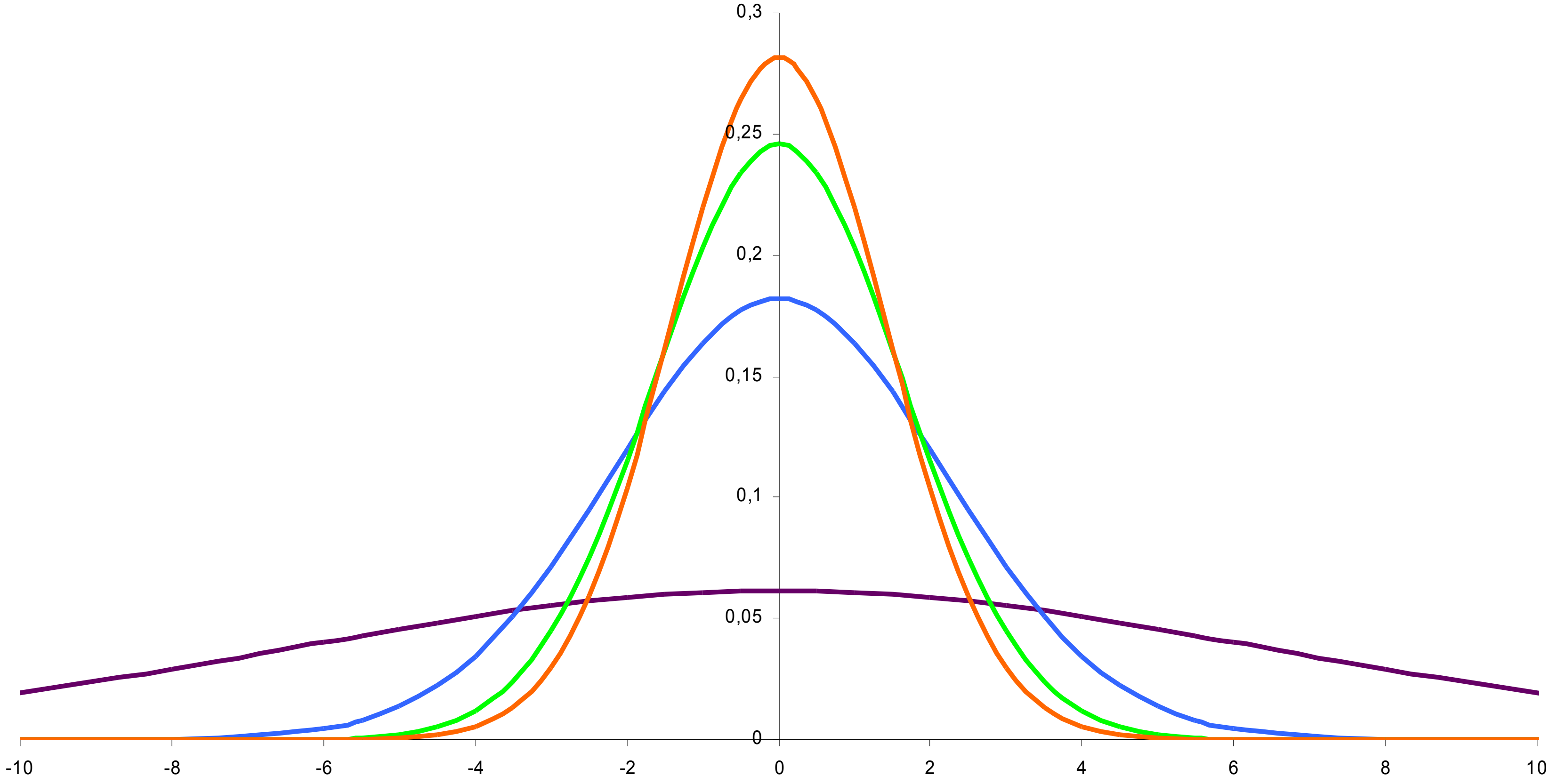
www.model.u-szeged.hu
www.szote.u-szeged.hu/dmi

Hypothesis testing

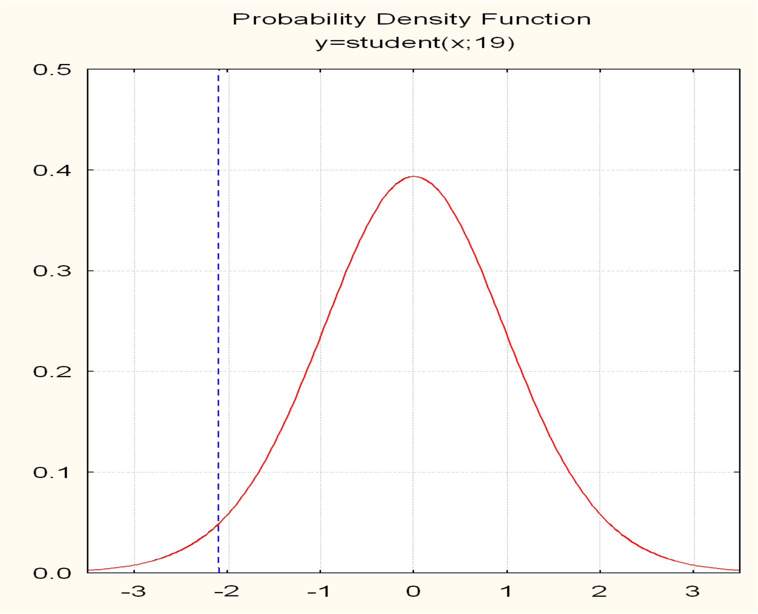
T-tests



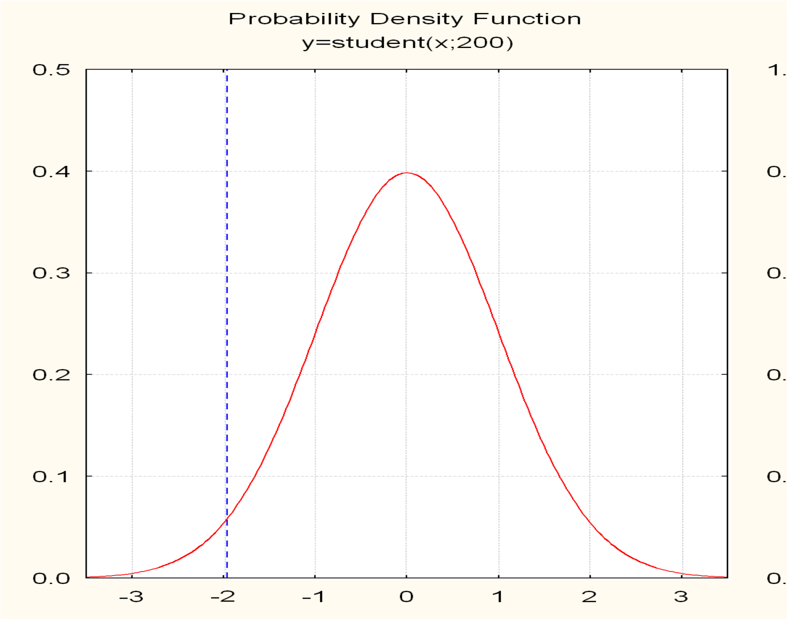
The *t*-distributions with 1-4 degrees of freedom



The *t*-distributions (Student's *t*-distributions)



df=19



df=200

TINV function (EXCEL)

df/α	0,1	0,05	0,025	0,01
1	6,314	12,706	25,452	63,657
2	2,920	4,303	6,205	9,925
3	2,353	3,182	4,177	5,841
4	2,132	2,776	3,495	4,604
5	2,015	2,571	3,163	4,032
6	1,943	2,447	2,969	3,707
7	1,895	2,365	2,841	3,499
8	1,860	**2,306**	2,752	3,355
9	1,833	2,262	2,685	3,250

Statistical inference: hypothesis testing

- Statisticians usually test the hypothesis which tells them what to expect by giving a specific value to work with.
  - They refer to this hypothesis as the **null hypothesis** and symbolize it as H_0. The null hypothesis is often the one that assumes fairness, honesty or equality.
  - The opposite hypothesis is called **alternative hypothesis** and is symbolized by H_a. This hypothesis, however, is often the one that is of interest. Some statisticians refer the H_a as the motivated hypothesis

Decision rules

- At the beginning of the experiment you should formulate the two opposing hypotheses. Then you should state what evidence will cause you to say that you think the alternative hypothesis is the true one. This statement is called your decision rule.
- When the evidence supports the alternative hypothesis, we say that we "reject the null hypothesis". When the evidence does not support the alternative , we say that we "fail to reject the null hypothesis"

Decision rule to test the mean of a normal distribution with unknown standard deviation
H_0: μ=c, H_a: μ ≠c

- **Decision using a confidence interval**
  - If the (1-α)100 % level confidence interval does not contain ***c***, we **reject** H_0 and say: the difference is significant at (1- α)100 % level.
  - If the (1- α)100 % level confidence interval contains ***c***, we **do not reject** H_0 and say: the difference is not significant at (1- α)100 % level.

Decision using critical values

- There is another way for finding the decision rule: we can use the so called critical points or rejection points instead of the confidence interval. If the null hypothesis is true, the statistic in has a t distribution with n-1 degrees of freedom

$$t = \frac{\bar{x} - c}{\frac{s}{\sqrt{n}}} = \sqrt{n}\,\frac{\bar{x} - c}{s}$$

$$P(|t| > t_{\alpha/2}) = \alpha$$

Example

- Average blood cholesterol of nine (9) patients were compared with the normal (5.2 mg/ml) value of blood cholesterol at 95% probability level?

5.4
4.3
8.4
4.9
5.8
5.3
5.5
5.4
6.3

Hypotheses

- H_0: The average cholesterol is equal to 5.2
  - H_0: $\mu=5.2$
- H_A: The average cholesterol differs from 5.2
  - H_a: $\mu\neq5.2$

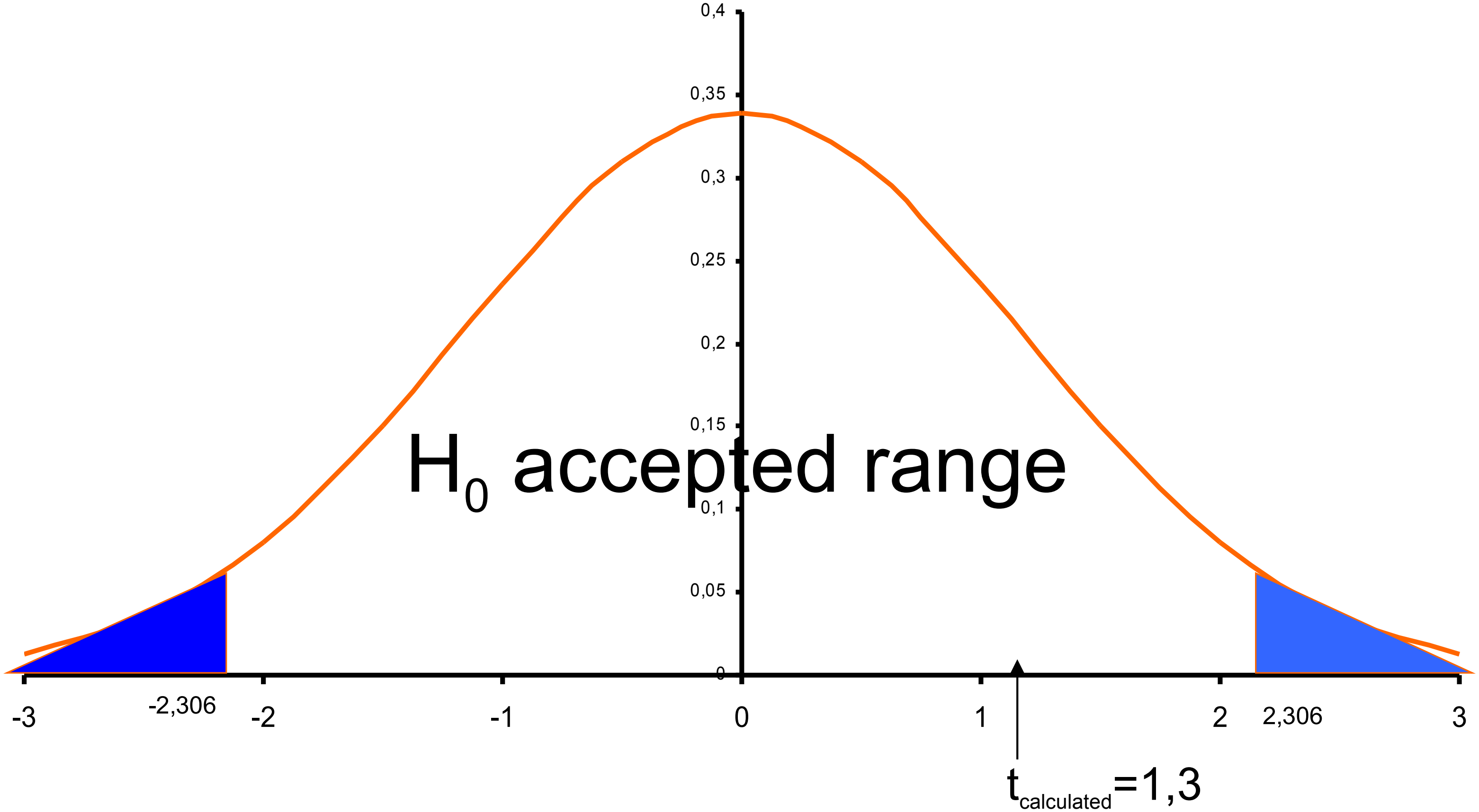
Results

Mean	5.7
SD	1.15
N	9
df	8
t - statistics	1.3
df	8
t_{table}	2.31
P(T<=t) 2-sided	0.23

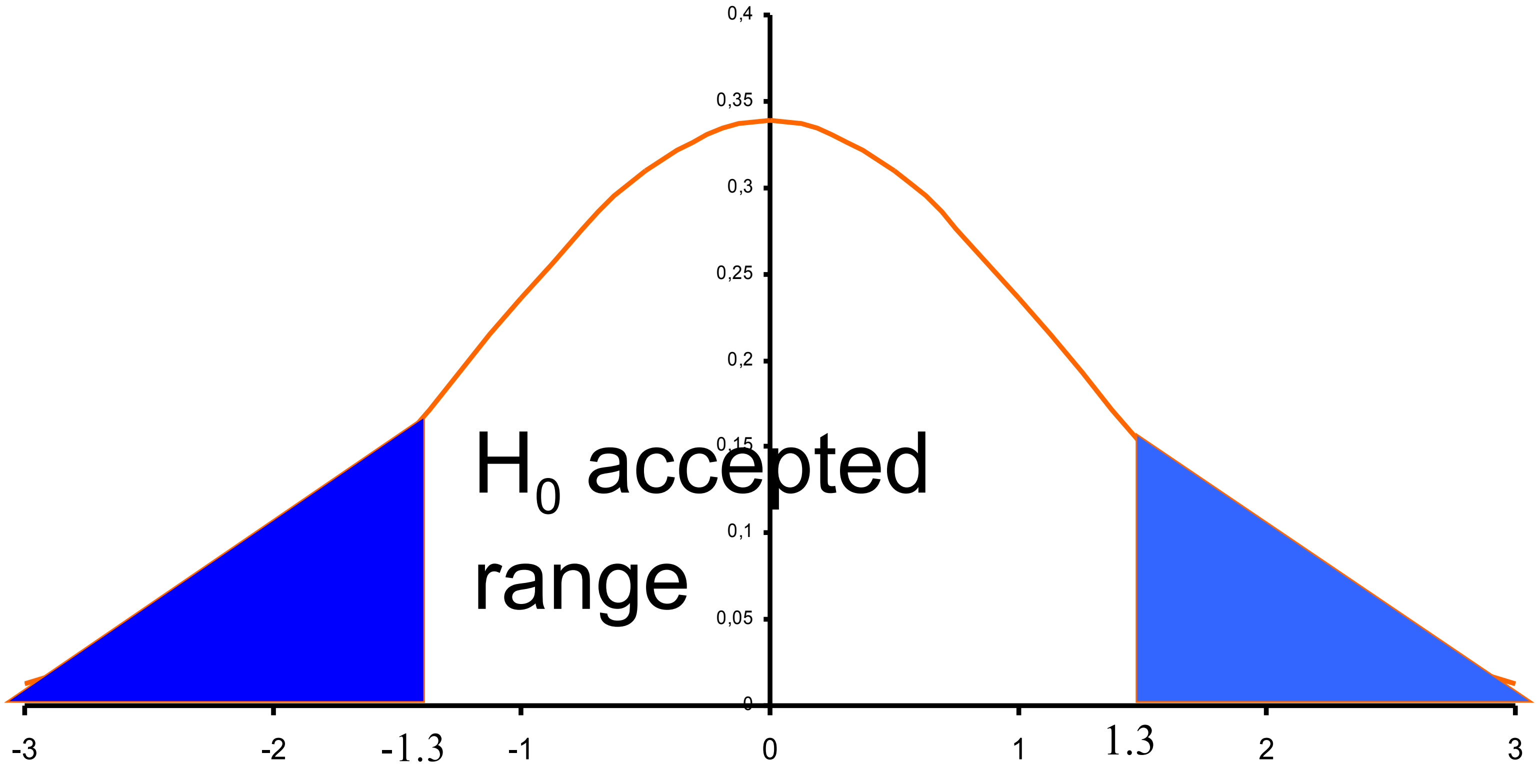
Table for t distribution

df/α	0,1	0,05	0,025	0,01
1	6,314	12,706	25,452	63,657
2	2,920	4,303	6,205	9,925
3	2,353	3,182	4,177	5,841
4	2,132	2,776	3,495	4,604
5	2,015	2,571	3,163	4,032
6	1,943	2,447	2,969	3,707
7	1,895	2,365	2,841	3,499
8	1,860	**2,306**	2,752	3,355
9	1,833	2,262	2,685	3,250

Decision using critical value from the table



Decision using the p-value (p=0.23)



SPSS results

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
koleszterin	9	5,7000	1,15326	,38442

One-Sample Test

	Test Value = 5.2					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
koleszterin	1,301	8	,230	,50000	-,3865	1,3865

Example II.

- Average blood glucose of nine (9) patients were compared with the normal (5.4 mg/ml) value of blood glucose at 95% probability level?

6.5
5.9
8.4
7.3
6.1
5.2
5.6
6.7
7.2

Hypotheses

- H_0: The average blood glucose is equal to 5.4
  - H_0: $\mu=5.4$
- H_A: The average blood glucose differs from 5.4
  - H_a: $\mu\neq5.4$

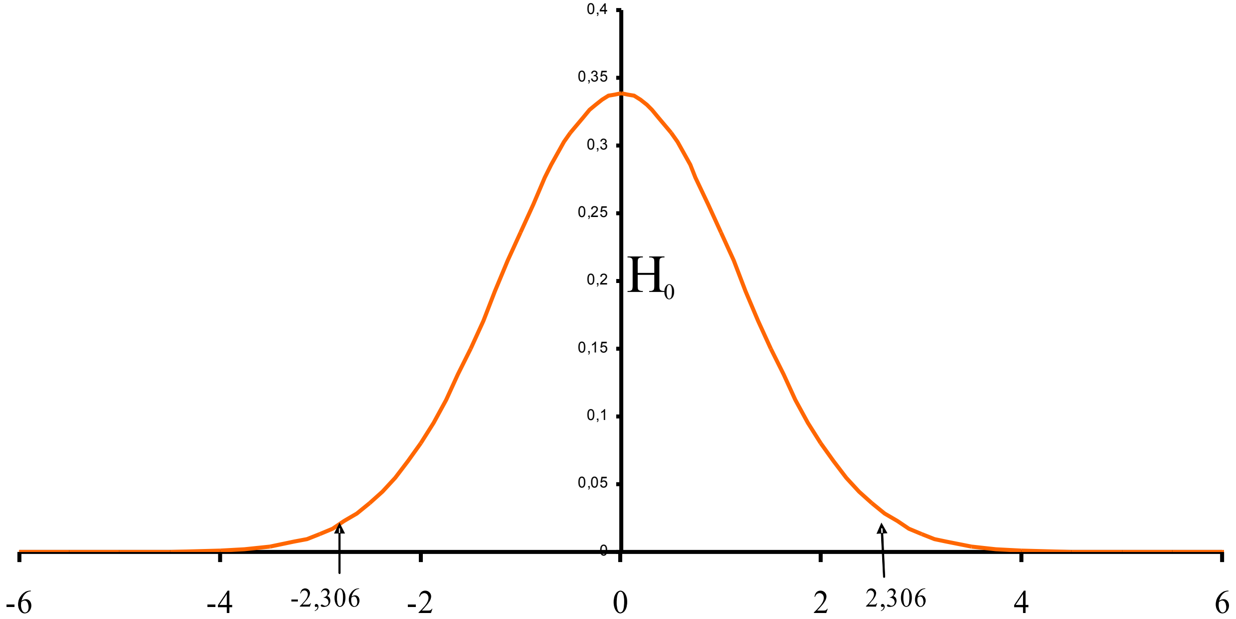
Results

Mean	6.544
SD	0.986
N	9
df	8
t calculated	3.481
P(T<=t) two-sided	0.008
t_{table}	2.306

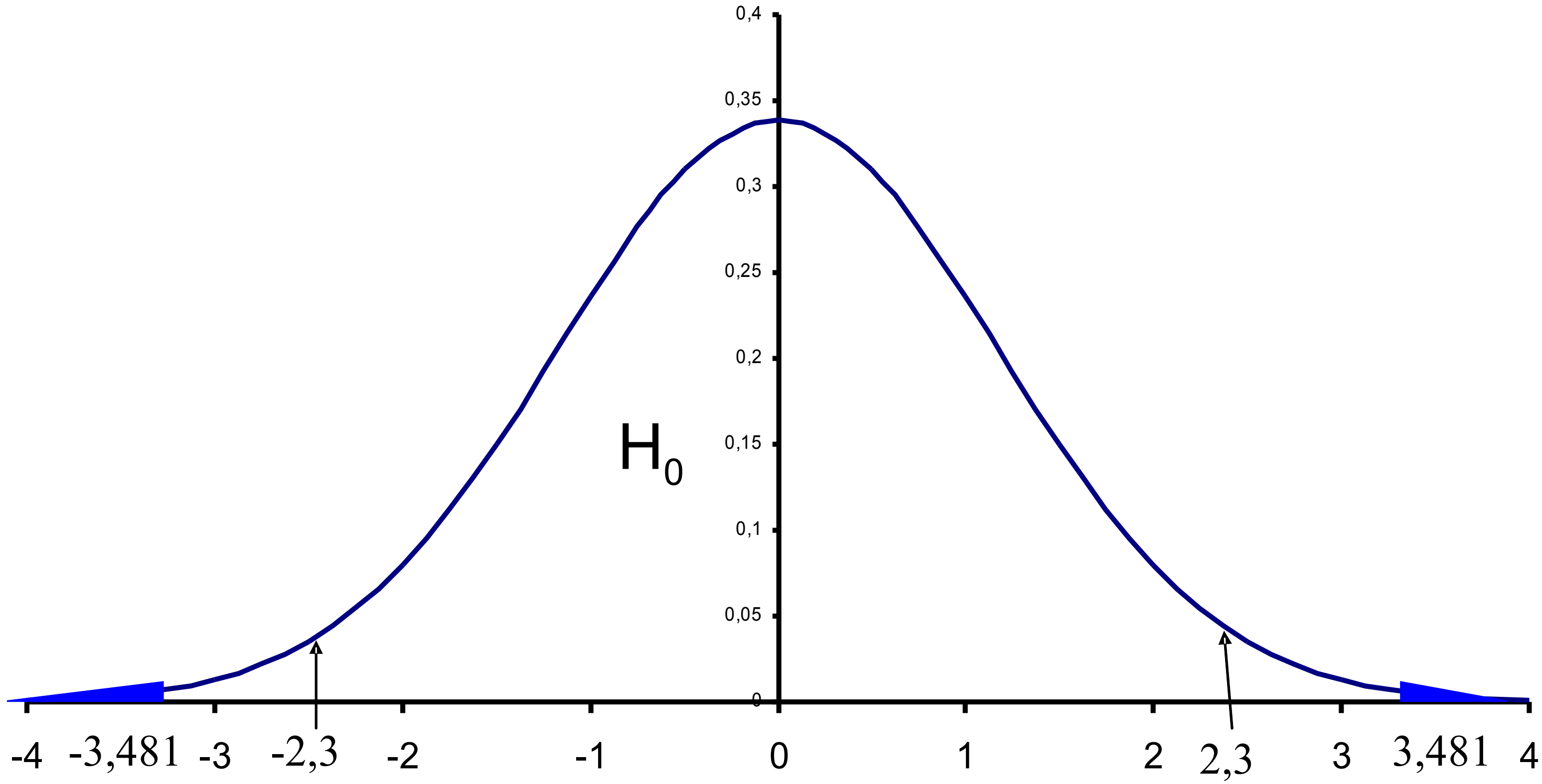
Table for t distribution

df/α	0,1	0,05	0,025	0,01
1	6,314	12,706	25,452	63,657
2	2,920	4,303	6,205	9,925
3	2,353	3,182	4,177	5,841
4	2,132	2,776	3,495	4,604
5	2,015	2,571	3,163	4,032
6	1,943	2,447	2,969	3,707
7	1,895	2,365	2,841	3,499
8	1,860	**2,306**	2,752	3,355
9	1,833	2,262	2,685	3,250

Decision using critical value from the table



Decision using the p-value (p=0.008)



SPSS results

- SPSS command:
  - Analyze/Compare Means/ One-sample t-test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
vercukor	9	6,5444	,98629	,32876

One-Sample Test

	Test Value = 5.4					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
vercukor	3,481	8	,008	1,14444	,3863	1,9026

Paired samples *t*-test I

- A special type of experiment is referred to as a **matched-pair** experiment, when we do "before-and after" comparisons, or when we compare "siblings".
- Let's suppose that an experimenter would like to prove that a special treatment decreases the blood pressure. To prove this, he first measures the blood pressure of a group of patients randomly selected from a group of people suffering disease with high blood pressure. This gives him a sample "before treatment". After the treatment the experimenter measures the blood pressure of the same patients, this results another sample "after treatment".
- To see the effect of the treatment, it is possible to compute the differences of the values "after treatment" and "before treatment" to each patient. To summarize the situation, we generally have to following data

Paired samples *t*-test II

- If the treatment has effect to the blood pressure, then the mean difference must be a number different from zero. If the treatment hasn't any effect to the blood pressure, than the mean difference must not be differ from zero. Let's formulate the two hypothesis:
- H_0: μ=0 (the mean of the population of differences is 0)
- H_a: μ≠0(the mean of the population of differences is not 0)
- This situation is a special case of the one sample t-test. If we suppose that the population of all d's is approximately normal, then our experiment reduces to a one-sample t-test, where the sample is no the sample of *d* differences. So we can test the null hypothesis by counting the value

$$t = \frac{\bar{d}}{s_d} \cdot \sqrt{n}, \quad \text{where } \bar{\mathrm{d}}$$

Paired samples *t*-test III

- is the mean of the sample of differences, s_d is the standard deviation of the sample of differences.
- The decision rule is the following: we can reject H_0 in favor of H_a setting the probability of a Type I error equal to α if and only if $|t| > t_{\alpha/2}$.
- For a given and for *n*-1 degrees of freedom we can find the value $t_{\alpha/2}$.

Example: Suppose that we have the following data for the the weights before and after a diet course measuring 9 persons

Before diet	After diet	difference
156	117	-39
111,4	85,9	-25,5
98,6	75,8	-22,8
104,3	82,9	-21,4
105,4	82,3	-23,1
100,4	77,7	-22,7
81,7	62,7	-19
89,5	69	-20,5
78,2	63,9	-14,3
	$\bar{d}$	-23.1444
	SD	6.7333
	SE	2.24444
	t=	-10.311

Paired t-test:

	Before	After
Mean	102.833	79.6889
Variance(sample)	520.483	264.881
N	9	9
df	8	
t-calculated	10.31	
P(T<=t) 2-sided	6.74E-06	
t_{table}	2.306	

Paired sample t-test

- SPSS command:
  - Analyze/Compare Means/ Paired-samples t-test

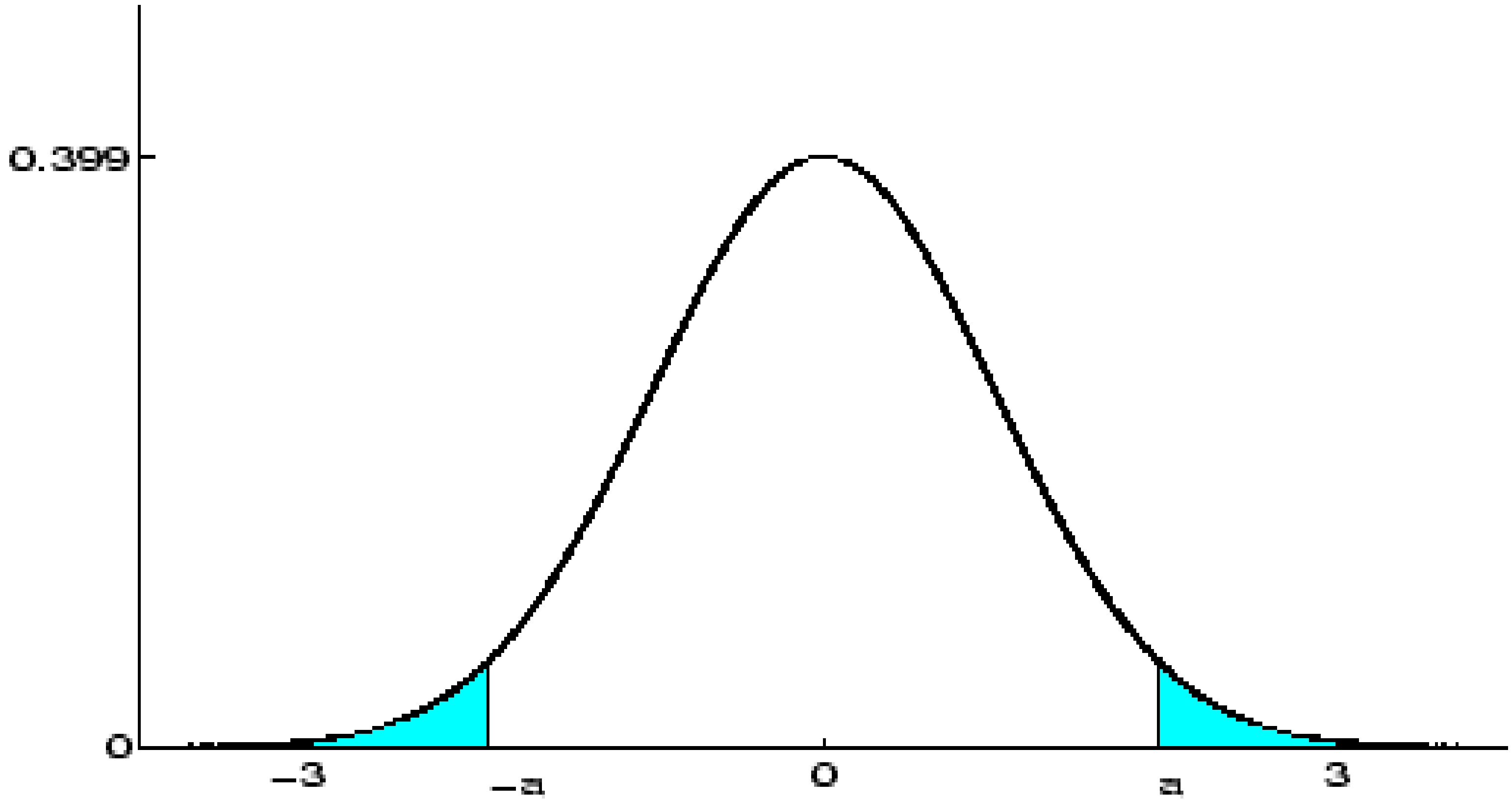
Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	before	102,8333	9	22,81409	7,60470
	after	79,6889	9	16,27525	5,42508

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	before - after	23,14444	6,73333	2,24444	17,96875	28,32014	10,312	8	,000

Decision using p-value



Example: Suppose that we have the following data for the blood pressure after measuring 8 persons

Before treatment	After treatment	Difference
170	150	20
160	120	40
150	150	0
150	160	-10
180	150	30
170	150	20
160	120	40
160	130	30

$\bar{d}=21.25$

$s_d=18.077$

$t=3.324$

t-Test: Paired Two Sample for Means

	Before treatment	*After treatment*
Mean	162,5	141,25
Variance	107,143	241,071428
Observations	8	8
Pearson Correlation	0,06667	
Hypothesized Mean Difference	0	
df	7	
t Stat	3,32486	
P(T<=t) one-tail	0,00634	
t Critical one-tail	1,89458	
P(T<=t) two-tail	0,01268	
t Critical two-tail	2,36462	

Confidence interval for the population's mean when σ is unknown

- Is σ, the standard deviation of the population is unknown, it can be approximated by the standard deviation of the sample

$$SD = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}}$$

- It can be shown that

$$P(\bar{x} - t_{\alpha/2} \frac{SD}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2} \frac{SD}{\sqrt{n}}) = 1 - \alpha$$

so

$$(\bar{x} - t_{\alpha/2} \frac{SD}{\sqrt{n}}, \quad \bar{x} + t_{\alpha/2} \frac{SD}{\sqrt{n}})$$

is a (1-α)100 confidence interval for μ.

- Here $t_{\alpha/2}$ can be found in tables of the Student's *t* distribution with n-1 degrees of freedom

Example

- We wish to estimate the average number of heartbeats per minute for a certain population.
- The mean for a sample of 36 subjects was found to be 90, the standard deviation of the sample was SD=15.5. Supposed that the population is normally distributed the 95 % confidence interval for μ:
- α=0.05, SD=15.5
- Degrees of freedom: df=n-1=36-1=35
- t $\alpha/2$=2.0301
- The lower limit is

  $90 - 2.0301 \cdot 15.5/\sqrt{36} = 90 - 2.0301 \cdot 2.5833 = 90 - 5.2444 = 84.755$
- The upper limit is

  $90 + 2.0301 \cdot 15.5/\sqrt{36} = 90 + 2.0301 \cdot 2.5833 = 90 + 5.2444 = 95.24$
- The 95% confidence interval for the population mean is

  (84.76, 95.24)
- It means that the true (but unknown) population means lies it the interval (84.76, 95.24) with 0.95 probability. We are 95% confident the true mean lies in that interval.

Statistical errors

- H_0 is true and the sample data lead you correctly to decide that it is true.
- H_0 is true but by bad luck the sample data lead you mistakenly to think that it is false.
- H_0 is false and the sample data lead you correctly to decide that it is false.
- H_0 is false but by bad luck the sample data lead you mistakenly to think that it could be true.

Statistical errors

	you fail to reject H0	you reject H0
H0 is true	correct	Type I error
H0 is false	Type II error	correct

Testing the mean of two independent samples: two-sample *t*-test

- Let's suppose that we have two independent samples with not necessarily equal sample sizes: . The problem of testing the difference between two means is simplest when we can make two additional assumptions.
- 1. Both populations are approximately normal.
- 2. The variances of the two populations are approximately equal.
- That is the *xi*-s are distributed as N(μ_1,σ) and the *yi*-s are distributed as N(μ_2,σ).
- Lets formulate the two hypotheses
  - H_0: $\mu_1 = \mu_2$,
  - H_a: $\mu_1 \neq \mu_2$

Quantity t has Student t distribution with n+m-2 degrees of freedom

$$t = \frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{\bar{x} - \bar{y}}{s_p} \cdot \sqrt{\frac{nm}{n+m}}$$

$$s_p^2 = \frac{(n-1) \cdot s_x^2 + (m-1) \cdot s_y^2}{n+m-2}$$

Testing the mean of two independent samples in the case of different standard deviations

- Let's suppose that we have two independent samples with not necessarily equal sample sizes: . The problem of testing the difference between two means is simplest when we can make two additional assumptions.
- 1. Both populations are approximately normal.
- 2. The variances of the two populations are different.
- That is the x_i-s are distributed as $N(\mu_1,\sigma_1)$ and the y_i-s are distributed as $N(\mu_2,\sigma_2)$.
- Lets formulate the two hypotheses
  - H_0: $\mu_1 = \mu_2$,
  - H_a: $\mu_1 \neq \mu_2$

36

Testing the mean of two independent samples in the case of different standard deviations

$$d = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}$$

$$df : \frac{(n-1)\cdot(m-1)}{g^2\cdot(m-1) + (1-g^2)\cdot(n-1)}$$

$$g = \frac{\frac{s_x^2}{n}}{\frac{s_x^2}{n} + \frac{s_y^2}{m}}$$

Comparison of the standard deviations of two normal populations: F-test

- Let's suppose that we have two independent samples with not necessarily equal sample sizes.
- The x_i-s are distributed as $N(\mu_1,\sigma_1)$ and the y_i-s are distributed as $N(\mu_2,\sigma_2)$.
- Lets formulate the two hypotheses
  - H_0: $\sigma_1 = \sigma_2$,
  - H_a: $\sigma_1 > \sigma_2$

$$F = \frac{\max(s_x^2, s_y^2)}{\min(s_x^2, s_y^2)}$$

F-distribution

- The F-distribution is not symmetrical and has two degrees of freedom. In our case the degrees of freedom are: the sample size of the nominator-1 and the sample size of the denominator -1.
- There are tables for the critical values of the F-distribution, these are one-tailed tables. After finding the critical F_{table} value, our decision is the following:
- In case of $F > F_{table}$ we reject the null hypothesis and claim that the variances are different at $(1-\alpha)100\%$ level,
- In case of $F < F_{table}$ we do not reject the null hypothesis and claim that the two variances are not different at $(1- \alpha)100\%$ level

(Student) t-test example

- Suppose that we measured the biomass (milligrams) produced by bacterium A and bacterium B, in shake flasks containing glucose as substrate. We had 4 replicate flasks of each bacterium

Data

	Bacterium A	Bacterium B
Replicate 1	520	230
Replicate 2	460	270
Replicate 3	500	250
Replicate 4	470	280
Σx	1950	1030
n	4	4
$\bar{x}$	487.5	257.5
Σx^2	952900	266700
$(\Sigma x)^2$	3802500	1060900
$\frac{(\Sigma x)^2}{n}$	950625	265225
Σd^2	2275	1475
σ^2	758.3	491.7

$$t = \frac{\bar{x}1 - \bar{x}2}{\sigma d}$$

Result of *F*-test using Excel

F-Test Two-Sample for Variances

	Bacterium A	*Bacterium B*
Mean	487,5	257,5
Variance	758,3333333	491,6666667
Observations	4	4
df	3	3
F	1,542372881	
P(F<=f) one-tail	0,365225092	
F Critical one-tail	9,276628154	

Result of *t*-test using Excel

t-Test: Two-Sample Assuming Equal Variances

	BactA	*BactB*
Mean	487,5	257,5
Variance	758,3333333	491,6666667
Observations	4	4
Pooled Variance	625	
Hypothesized Mean Difference	0,05	
df	6	
t Stat	13,00793635	
P(T<=t) one-tail	6,35743E-06	
t Critical one-tail	1,943180274	
P(T<=t) two-tail	1,27149E-05	
t Critical two-tail	2,446911846	

Summary of hypothesis testing

- **Step 1**. *State the motivated (alternative) hypothesis*. Based on some prior experience or idea, you are motivated to make a claim which you desire to prove by an experiment. That is, you state the alternative hypothesis H_a about some population.
- **Step 2**. *State the null hypothesis*. For statistical purposes you test the opposite hypothesis. Therefore, you state the null hypothesis H_0.
- **Step 3**. You *select the , the probability of Type I error, or the (1-α)100 % significance level*. That is, you are stating how large risk you are willing to take in making the error of claiming that your motivated hypothesis is true. You may select any significance level you wish. In most published statistics the authors have used α =0.05 or α =0.01. These are generally accepted standards.
- **Step 4**. You choose the *size n of the random sample*. This choice is often determined by the amount of the time and/or money that you have to do the experiment and the availability of subjects. Ease of computation might also be a factor in the selection of *n*.
- **Step 5.** *Select a random sample* from the appropriate population and obtain your data.
- **Step 6**. *Calculate the decision rule*. Your decision rule will have one or two critical points, depending on whether the motivated hypothesis is one-tailed or two-tailed.
- **Step 7**. *Decision* based on the experimental outcome and the previously calculated decision rule, you will make one of two decisions.
- a) Reject the null hypothesis and claim that your alternative hypothesis was correct.
- b) Fail to reject the null hypothesis: you have been unable to prove that the alternative hypothesis is correct. Since we have not determined the power of the test, we do not wish to state that the null hypothesis is true. If the power is low, there would be a correspondingly large probability of a Type II error. So we use the phrase "fail to reject H_0" rather than "accept H_0".