Biostatistics

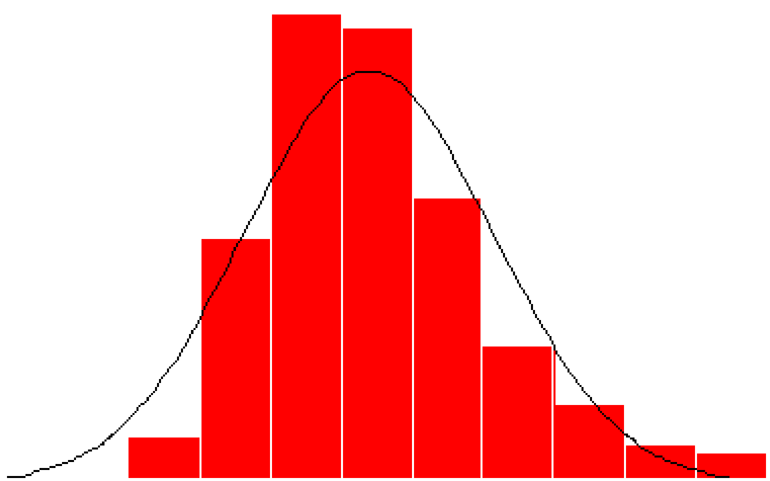
Author: **Krisztina Boda** PhD

University of Szeged
Department of Medical Physics and Informatics

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

Hypothesis tests III.

Statistical errors, one-and two sided tests. One-way analysis of variance.



Student’s t-tests

- **General purpose**. Student’s t-tests examine the **mean** of normal populations. To test hypotheses about the population mean, they use a test-statistic ***t*** that follows Student’s t distribution with a given degrees of freedom if the nullhypothesis is true.
- **One-sample t-test.** There is one sample supposed to be drawn from a normal distribtuion. We test whether the mean of a normal population is a given constant:
  - H0: μ=c
- **Paired t-test (=one-sample t-test for paired differences).** There is only one sample that has been tested twice (before and after the treatment) or when there are two samples that have been matched or "paired".We test whether the mean difference between paired observations is zero:
  - H0: $\mu_{differerence}$=0
- **Two sample t-test (or independent samples t-test).** There are two independent samples, coming from two normal populations. We test whether the two population means are equal:
  - H0: $\mu_1 = \mu_2$

Experimental design of *t*-tests

- **Paired t-test**
- **(related samples)**
- Each subject are measured twice

1st	2nd
x_1	y_1
x_2	y_2
…	…
x_n	y_n

- **Two-sample t-test**
- **(independent samples)**
- Each subject is measured once, and belongs to one group .

Group	Measurement
1	x_1
1	x_2
…	…
1	x_n
2	y_1
2	y_2
…	…
2	y_m

Sample size is not necessarily equal

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

- Independent samples:
  - Control group, treatment group
  - Male, female
  - Ill, healthy
  - Young, old
  - etc.
- Assumptions:
  - Independent samples : $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_m$
  - the x_i-s are distributed as $N(\mu_1,\sigma_1)$ and the y_i-s are distributed as $N(\mu_2,\sigma_2)$.
- H_0: $\mu_1=\mu_2$, H_a: $\mu_1\neq\mu_2$

Evaluation of two sample t-test depends on equality of variances;
To compare the means, there are two different formulas with different degrees of freedom depending on equality of variances

Comparison of the means (t-test)

- If H_0 is true and the variances are equal, then

$$t = \frac{\bar{x}-\bar{y}}{SD_p\sqrt{\frac{1}{n}+\frac{1}{m}}} = \frac{\bar{x}-\bar{y}}{SD_p}\cdot\sqrt{\frac{nm}{n+m}} \qquad SD_p^2 = \frac{(n-1)\cdot SD_x^2 + (m-1)\cdot SD_y^2}{n+m-2}$$

has Student's t distribution with **n+m-2** degrees of freedom.

- If H_0 is true and the variances are not equal, then

$$d = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_x^2}{n}+\frac{s_y^2}{m}}} \qquad df = \frac{(n-1)\cdot(m-1)}{g^2\cdot(m-1)+(1-g^2)\cdot(n-1)} \qquad g = \frac{\frac{s_x^2}{n}}{\frac{s_x^2}{n}+\frac{s_y^2}{m}}$$

has Student's t distribution with **df** degrees of freedom.

- Decision
  - If $|t|>t_{\alpha,df}$, the difference is significant at α level, we reject H0
  - If $|t|<t_{\alpha,df}$, the difference is not significant at α level, we do not reject H0

Comparison of the variances of two normal populations: quick F-test

- H_0: $\sigma_1^2=\sigma_2^2$
- H_a:$\sigma_1^{\,2} > \sigma_2^{\,2}$ (one sided test)
- F: the higher variance divided by the smaller variance:

$$F=\frac{\max(s_x^2,s_y^2)}{\min(s_x^2,s_y^2)}=\frac{higher\ sample\ \mathrm{var}iance}{smaller\ sample\ \mathrm{var}iance}$$

- Degrees of freedom:
  - 1. Sample size of the nominator-1
  - 2. Sample size of the denominator-1
- Decision based on F-table
  - If $F>F_{\alpha,table}$, the two variances are significantly different at α level

Table of the F-distribution α=0.05

Nominator->

Denominator	1	2	3	4	5	6	7	8	9	10
|	161.4476	199.5	215.7073	224.5832	230.1619	233.986	236.7684	238.8827	240.5433	241.8817
2	18.51282	19	19.16429	19.24679	19.29641	19.32953	19.35322	19.37099	19.38483	19.3959
3	10.12796	9.552094	9.276628	9.117182	9.013455	8.940645	8.886743	8.845238	8.8123	8.785525
4	7.708647	6.944272	6.591382	6.388233	6.256057	6.163132	6.094211	6.041044	5.998779	5.964371
5	6.607891	5.786135	5.409451	5.192168	5.050329	4.950288	4.875872	4.81832	4.772466	4.735063
6	5.987378	5.143253	4.757063	4.533677	4.387374	4.283866	4.206658	4.146804	4.099016	4.059963
7	5.591448	4.737414	4.346831	4.120312	3.971523	3.865969	3.787044	3.725725	3.676675	3.636523
8	5.317655	4.45897	4.066181	3.837853	3.687499	3.58058	3.500464	3.438101	3.38813	3.347163
9	5.117355	4.256495	3.862548	3.633089	3.481659	3.373754	3.292746	3.229583	3.178893	3.13728
10	4.964603	4.102821	3.708265	3.47805	3.325835	3.217175	3.135465	3.071658	3.020383	2.978237

Example

Control group	Treated group
170	120
160	130
150	120
150	130
180	110
170	130
160	140
160	150
	130
	120
n=8	n=10
$\bar{x}=162.5$	$\bar{y}=128$
$s_x=10.351$	$s_y=11.35$
$s_x^2=107.14$	$s_y^2=128.88$

$$F=\frac{128.88}{107.14}=1.2029,$$

Degrees of freedom 10-1=9, 8-1=7, critical value int he F-table is $F_{\alpha,9,7}=3.68$. As 1.2029<3.68, the two variances are considered to be equal, the difference is not significanr.

$$s_p^2=\frac{7\cdot 107.14+9\cdot 128.88}{10+8-2}=\frac{749.98+1160}{16}=119.37$$

$$t=\frac{162.5-128}{\sqrt{119.37}}\cdot\sqrt{\frac{10\cdot 8}{18}}=\frac{34.5}{10.92}\cdot\sqrt{4.444}=6.6569$$

Our computed test statistic t = 6.6569 , the critical value int he table $t_{0.025,16}=2.12$. As 6.6569>2.12, we reject the null hypothesis and we say that the difference of the two treatment means is significant at 5% level

Result of SPSS

Group Statistics

	csoport	N	Mean	Std. Deviation	Std. Error Mean
BP	Kontroll	8	162.5000	10.35098	3.65963
	Kezelt	10	128.0000	11.35292	3.59011

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
BP	Equal variances assumed	.008	.930	6.657	16	.000	34.50000	5.18260	23.51337	45.48663
	Equal variances not assumed			6.730	15.669	.000	34.50000	5.12657	23.61347	45.38653

Two sample t-test, example 2.

- A study was conducted to determine weight loss, body composition, etc. in obese women before and after 12 weeks in two groups:
- Group I. treatment with a very-low-calorie diet .
- Group II. no diet
- Volunteers were randomly assigned to one of these groups.
- We wish to know if these data provide sufficient evidence to allow us to conclude that the treatment is effective in causing weight reduction in obese women compared to no treatment.

Two sample t-test, cont.
Data

Group	Patient	Change in body weight
Diet	1	-1
	2	5
	3	3
	4	10
	5	6
	6	4
	7	0
	8	1
	9	6
	10	6
Mean		4.
SD		3.333
No diet	11	2
	12	0
	13	1
	14	0
	15	3
	16	1
	17	5
	18	0
	19	-2
	20	-2
	21	3
Mean		1
SD		2.145

Two sample t-test, example, cont.

- HO: $\mu_{diet}=\mu_{control,}$ (the mean change in body weights are the same in populations)
- H_a: $\mu_{diet} \neq \mu_{control}$ (the mean change in body weights are different in the populations)
- Assumptions:
  - normality (now it cannot be checked because of small sample size)
  - Equality of variances (check: visually compare the two standard deviations)

Two sample t-test, example, cont.

- Assuming equal variances, compute the *t* test- statistic: t=2.477

$$t=\frac{\bar{x}-\bar{y}}{s_p\sqrt{\frac{1}{n}+\frac{1}{m}}}=\frac{\bar{x}-\bar{y}}{s_p}\cdot\sqrt{\frac{nm}{n+m}}=\frac{4-1}{\sqrt{\frac{9\cdot 3.3333^2+10\cdot 2.145^2}{9+10}}}\sqrt{\frac{10\cdot 11}{10+11}}=\frac{3}{\sqrt{\frac{99.999+46.01025}{19}}}\sqrt{5.238}=2.477$$

- Degrees of freedom: 10+11-2=19
- Critical t-value: $t_{0.05,19}$=2.093
- Comparison and decision:
  - |t|=2.477>2.093(=$t_{0.05,19}$), the difference is significant at 5% level
- p=0.023<0.05 the difference is significant at 5% level

SPSS results

Group Statistics

	group	N	Mean	Std. Deviation	Std. Error Mean
Change in body mass	Diet	10	4.0000	3.33333	1.05409
	Control	11	1.0000	2.14476	.64667

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Change in body mass	Equal variances assumed	1.888	.185	2.477	19	.023	3.00000	1.21119	.46495	5.53505
	Equal variances not assumed			2.426	15.122	.028	3.00000	1.23665	.36600	5.63400

Comparison of variances.
p=0.185>0.05, not significant.
We accept the equality of variances

Comparison of means (t-test).
1st row: equal variances assumed.
t=2.477, df=19, p=0.023
The difference in mean weight loss is significant at 5% level

Comparison of means (t-test). 2nd row: equal variances not assumed.
As the equality of variances was accepted, we do not use the results from this row.

Motivating example

- Two lecturers argue about the mean age of the first year medical students. Is the mean age for boys and girls the same or not?
  - Lecturer#1 claims that the mean age boys and girls is the same.
  - Lecturer#2 does not agree.
  - Who is right?
- Statistically speaking: there are two populations:
  - the set of ALL first year boy medical students (anywhere, any time)
  - the set of ALL first year girl medical students (anywhere, any time)
- Lecturer#1 claims that the population means are equal: $\mu_{boys} = \mu_{girls}$.
- Lecturer#2 claims that the population means are not equal: $\mu_{boyys} \neq \mu_{girls}$.

Answer to the motivated example (mean age of boys and girls)

Group Statistics

	Sex	N	Mean	Std. Deviation	Std. Error Mean
Age in years	Male	84	21.18	3.025	.330
	Female	53	20.38	3.108	.427

- The mean age of boys is a litlle bit higher than the mean age of girls. The standard deviations are similar.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Age in years	Equal variances assumed	.109	.741	1.505	135	.135	.807	.536	-.253	1.868
	Equal variances not assumed			1.496	108.444	.138	.807	.540	-.262	1.877

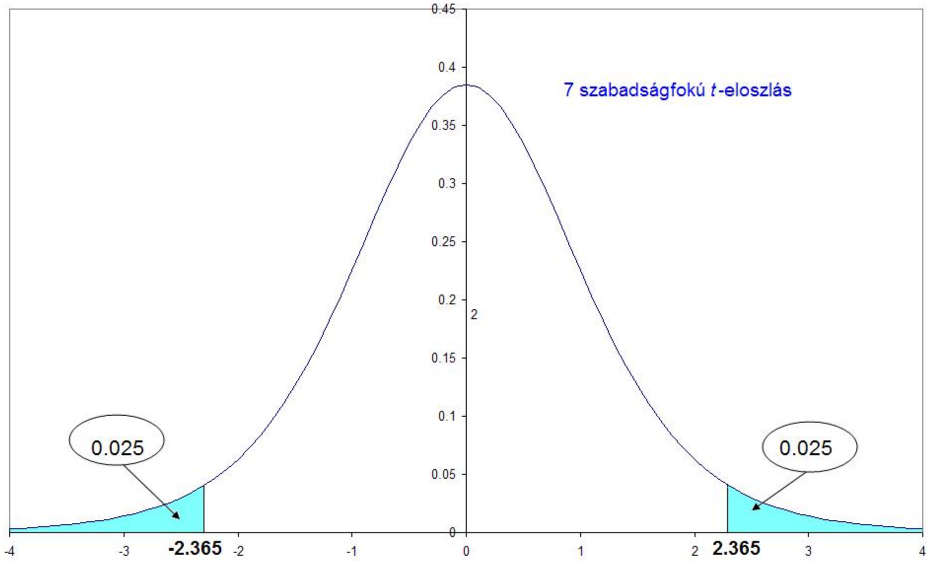
- Comparison of variances (F test for the equality of variances): p=0.741>0.05, not significant, we accept the equality of variances.
- Comparison of means: according to the formula for equal variances, t=1.505. df=135, p=0.135. So p>0.05, the difference is not significant. Althogh the experiencedd difference between the mean age of boys and girls is 0.816 years, this is statistically not significant at 5% level. We cannot show that the mean age of boay and girls are different.

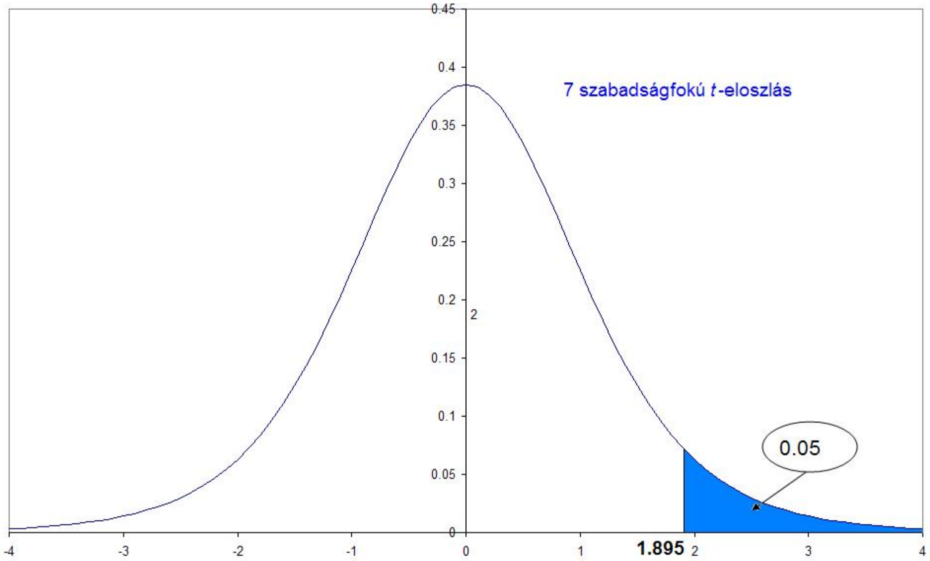
Other aspects of statistical tests

One- and two tailed (sided) tests

- Two tailed test
- H_0: there is no change $\mu_1=\mu_2$,
- H_a: There is change (in either direction) $\mu_1 \neq \mu_2$

- One-tailed test
- H_0: the change is negative or zero $\mu_1 \leq \mu_2$
- H_a: the change is positive (in one direction) $\mu_1 > \mu_2$





Critical values are different. ***p*-values:** p(one-tailed)=p(two-tailed)/2

Significance

- Significant difference – if we claim that there is a difference (effect), the probability of mistake is small (maximum α- Type I error).
- Not significant difference – we say that there is not enough information to show difference. Perhaps
  - there is no difference
  - There is a difference but the sample size is small
  - The dispersion is big
  - The method was wrong
- Even is case of a statistically significant difference one has to think about its biological meaning

Statistical errors

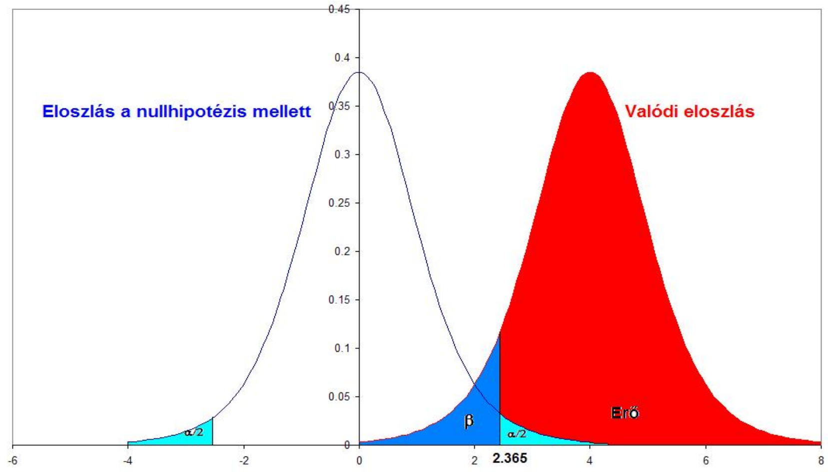
Truth	Decision	
	do not reject H_0	reject H_0 (significance)
H_0 is true	correct	Type I. error its probability: α
H_a is true	Type II. error its probability: β	correct

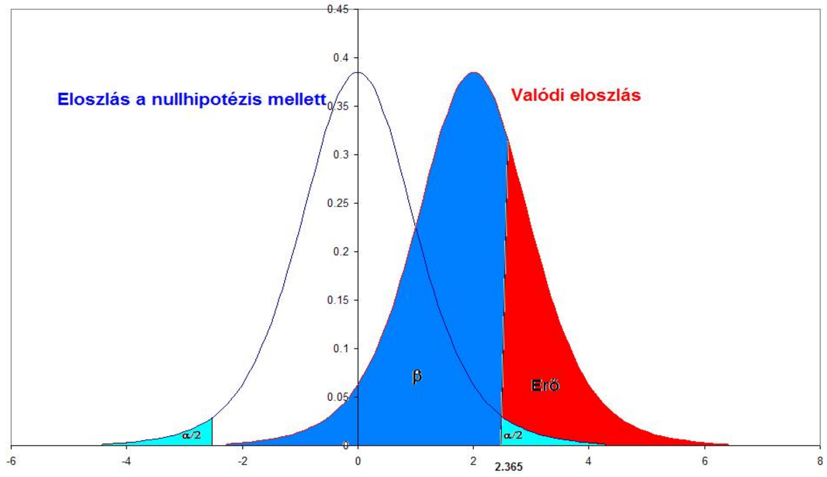
Error probabilities

- The probability of type I error is known (α).
- The probability of type II error is not known (β)
- It depends on
  - The significance level (α),
  - Sample size,
  - The standard deviation(s)
  - The true difference between populations
  - others (type of the test, assumptions, design, ..)
- The power of a test: 1- β

  It is the ability to detect a real effect

The power of a test in case of fixed sample size and α, with two alternative hypotheses

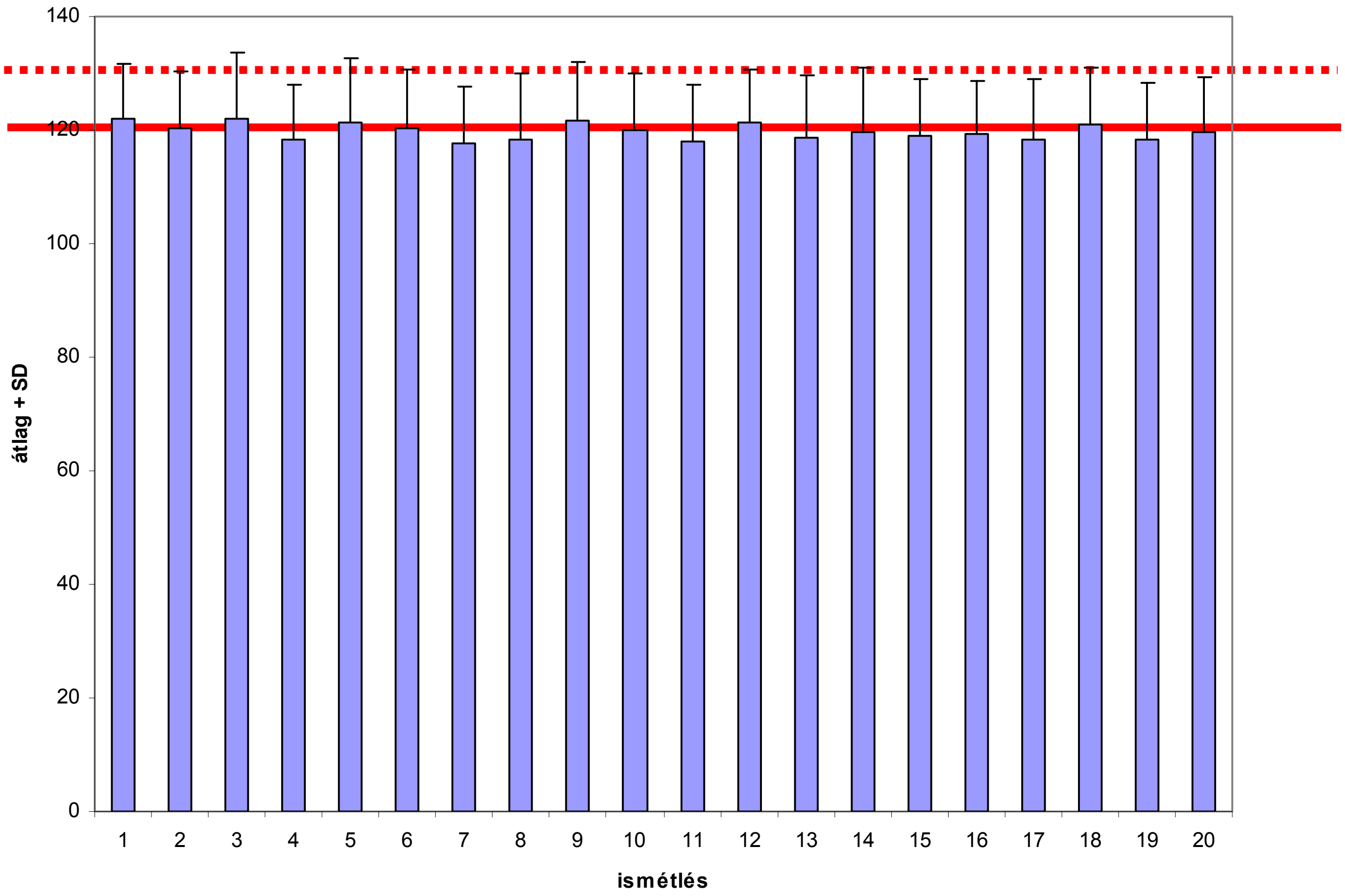




Comparison of several samples

The repeated use of t-tests is not appropriate

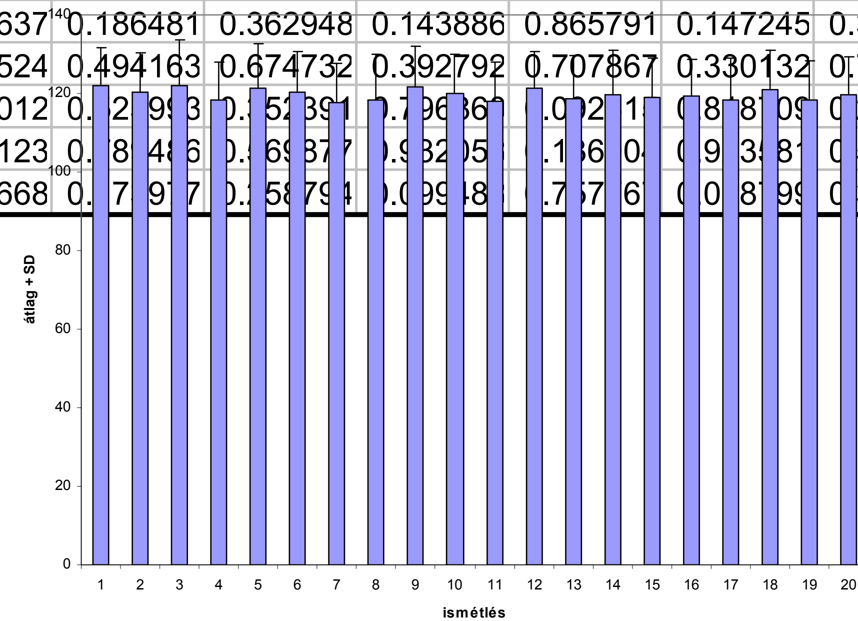
Mean and SD of samples drawn from a normal population N(120, 10^2), (i.e. μ=120 and σ=10)



Pair-wise comparison of samples drawn from the same distribution using *t*-tests

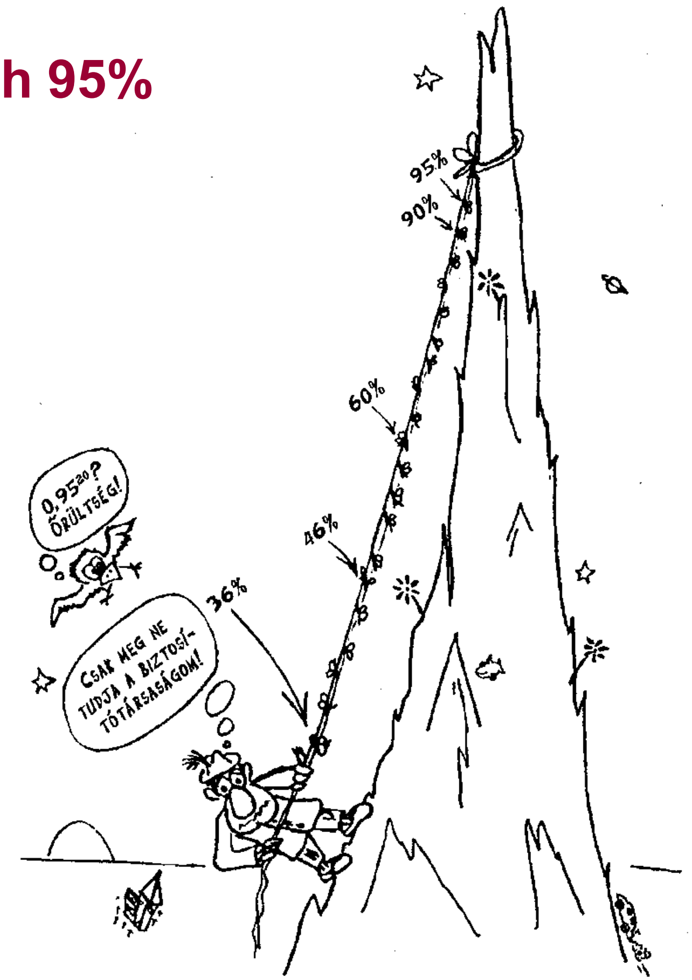
Variable	T-test for Dependent Samples: p-levels (veletlen) Marked differences are significant at p < .05000										
	s10	s11	s12	s13	s14	s15	s16	s17	s18	s19	s20
s1	0.304079	0.074848	0.781733	0.158725	0.222719	0.151234	0.211068	0.028262	0.656754	0.048789	0.223011
s2	0.943854	0.326930	0.445107	0.450032	0.799243	0.468494	0.732896	0.351088	0.589838	0.312418	0.842927
s3	0.364699	0.100137	0.834580	0.151618	0.300773	0.152977	0.201040	0.136636	0.712107	0.092788	0.348997
s4	0.335090	0.912599	0.069544	0.811846	0.490904	0.646731	0.521377	0.994535	0.172866	0.977253	0.338436
s5	0.492617	0.139655	0.998307	0.236234	0.420637	0.186481	0.362948	0.143886	0.865791	0.147245	0.399857
s6	0.904803	0.285200	0.592160	0.429882	0.774524	0.494163	0.674732	0.392792	0.707867	0.330132	0.796021
s7	0.157564	0.877797	0.053752	0.631788	0.361012	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.263511
s8	0.462223	0.858911	0.156711	0.878890	0.624123	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.564532
s9	0.419912	0.040189	0.875361	0.167441	0.357668	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.371769

***p*-values (detail)**



Knotted ropes: each knot is safe with 95% probability

- The probability that two knots are „safe” =0.95*0.95 =0.9025~90%
- The probability that 20 knots are „safe” $=0.95^{20}=0.358\sim36\%$
- The probability of a crash in case of 20 knots ~64%

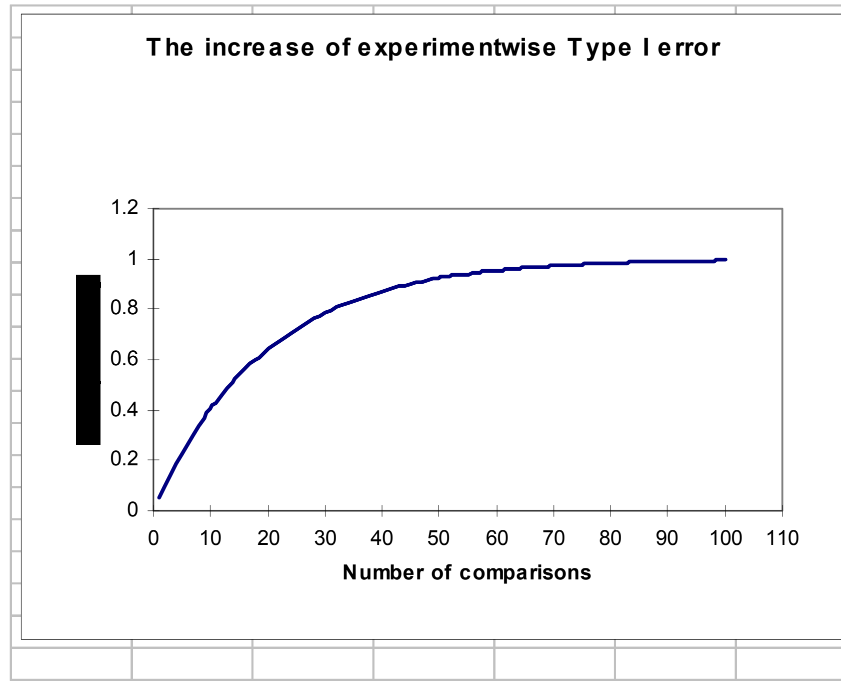


10. ábra. Nemtörődöm doktor, amint a nemzetközi szakirodalom által javasolt számos, egyenként meglehetősen biztonságos csomóval összekötözött mászókötélen függ. Ez az utolsó felvétel Nemtörődöm doktorról. Egy naiv elképzelésnek esett áldozatul, azt hitte, hogy a tudomány megbízhatósági kritériumait a hegymászásra is alkalmazni lehet

The increase of type I error

- It can be shown that when t tests are used to test for differences between multiple groups, the chance of mistakenly declaring significance (Type I Error) is increasing. For example, in the case of 5 groups, if no overall differences exist between any of the groups, using two-sample t tests pair wise, we would have about 30% chance of declaring at least one difference significant, instead of 5% chance.
- *In general, the t test can be used to test the hypothesis that **two** group means are not different. To test the hypothesis that **three ore more** group means are not different, analysis of variance should be used.*

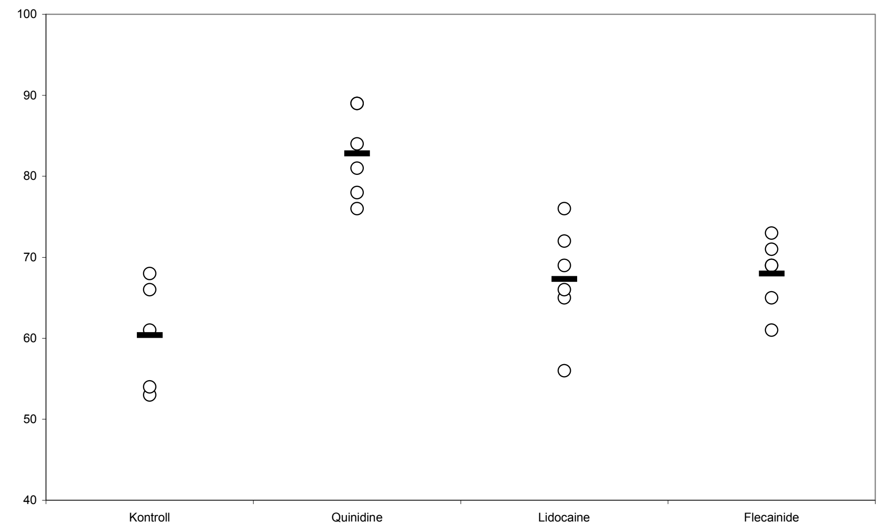
- False positive rate for each test = 0.05
- Probability of incorrectly rejecting ≥ 1 hypothesis out of *N* testings
- $= 1 - (1-0.05)^N$



Motivating example

- In a study (Farkas *et al*, 2003.) the effects of three Na+ channel–blocking drugs—quinidine, lidocaine and flecainide— was examined on length of QT interval and on the heart rate before and during regional ischemia in isolated rat hearts.
- The table and the figure show the length of the QT intervals measured in the 4 groups. Is there a significant difference between the means?

	Control	Quinidine	Lidocaine	Flecainide
	61	76	65	69
	53	84	56	65
	68	89	76	73
	66	78	72	71
	54	81	66	61
		89	69	69
mean	60.4	82.8	67.3	68.0
SD	6.80	5.49	6.86	4.34



One-Way ANOVA (Analysis of Variance)
Comparison of the mean of several normal populations

- Let us suppose that we have *t* independent samples (*t* "treatment" groups) drawn from normal populations with equal variances ~N(μ_i,σ).
- Assumptions:
  - Independent samples
  - normality
  - Equal variances
- Null hypothesis: population means are equal, $\mu_1 = \mu_2 = .. = \mu_t$

Method

- If the null hypothesis is true, then the populations are the same: they are normal, and they have the same mean and the same variance. This common variance is estimated in two distinct ways:
  - between-groups variance
  - within-groups variance
- If the null hypothesis is true, then these two distinct estimates of the variance should be equal
- ‘New’ (and equivalent) null hypothesis: $\sigma^2_{between} = \sigma^2_{within}$
- their equality can be tested by an F ratio test
- The p-value of this test:
  - if $p > 0.05$, then we accept H_0. The analysis is complete.
  - if $p < 0.05$, then we reject H_0 at 0.05 level. There is at least one group-mean different from one of the others

The ANOVA table

Source of variation	Sum of squares	Degrees of freedom	Variance	F	p
Between groups	$Q_k = \sum_{i=1}^{t} n_i(\bar{x}_i - \bar{x})^2$	t-1	$s_k^2 = \frac{Q_k}{t-1}$	$F = \frac{s_k^2}{s_b^2}$	p
Within groups	$Q_b = \sum_{i=1}^{t} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$	N-t	$s_b^2 = \frac{Q_b}{N-t}$		
Total	$Q = \sum_{i=1}^{t} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2$	N-1			

ANOVA

QT

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1515.590	3	505.197	14.426	.000
Within Groups	665.367	19	35.019		
Total	2180.957	22			

$F(3,19)=14.426$, $p<0.001$, the difference is significant at 5% level, There are one or more different group-means

Following-up ANOVA

- If the F-test of the ANOVA is not significant, we are ready
- If the F-test of ANOVA is significant, we might be interested in pairwise comparisons (but t-tests are NOT appropriate!)

Pair wise comparisons

- As the two-sample t-test is inappropriate to do this, there are special tests for multiple comparisons that keep the probability of Type I error as α. The most often used multiple comparisons are the modified t-tests.
- Modified t-tests (LSD)
  - Bonferroni: α/(number of comparisons)
  - Scheffé
  - Tukey
  - Dunnett: a test comparing a given group (control) with the others
  - ….

	Mean difference	p
Control – Quinidine	22.4333	.000
Control – Lidocaine	6.9333	.158
Control – Flecainide	7.6000	.113

Resutl of the Dunnett test

Review questions and problems

- The null- and alternative hypothesis of the two-sample t-test
- The assumption of the two-sample t-test
- Comparison of variances
- F-test
- Testing significance based on t-statistic
- Testing significance based on p-value
- Meaning of the p-value
- One-and two tailed tests
- Type I error and its probability
- Type II error and its probability
- The power of a test
- In a study, the effect of Calcium was examined to the blood pressure. The decrease of the blood pressure was compared in two groups. Interpret the SPSS results

Group Statistics

	treat	N	Mean	Std. Deviation	Std. Error Mean
decr	Calcium	10	5.0000	8.74325	2.76486
	Placebo	11	-.2727	5.90069	1.77913

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
decr	Equal variances assumed	4.351	.051	1.634	19	.119	5.27273	3.22667	-1.48077	12.02622
	Equal variances not assumed			1.604	15.591	.129	5.27273	3.28782	-1.71204	12.25749

Review questions and exercises

- One-and two tailed tests
- The type I error and its probability
- The type II error and its probability
- The increase of Type I. error
- The aim and the nullhypothesis of one-way ANOVA
- The assumptions of one-way ANOVA
- The ANOVA table
- Pair-wise comparisons