Kiválasztás

Egymintás t-próba

Mann-Whitney U-test

Kétmintás t-próba

Varianciaanalízis

Fontos, hogy mindig a kérdésnek és a változónak megfelelőt!

De melyiket válasszam?

A változó szerint

paraméteres

Egy ismert eloszlás valamilyen paraméterére vonatkozó hipotézis vizsgálata.
Az ismert eloszlás leggyakrabban a normális eloszlás.

nem-paraméteres

Egy ismeretlen eloszlás paraméterére, típusára vonatkozó hipotézis vizsgálata.

Nem-paraméteres eljárások

Eloszlás-független eljárások. (distribution free methods)

Előnyük : nem kötöttek eloszláshoz.
Hátrányuk: általában kisebb teljesítményűek.

Rangsorolásos módszerek:
Az értékek helyett
az ún. ***rang***okat használjuk.

Rangok

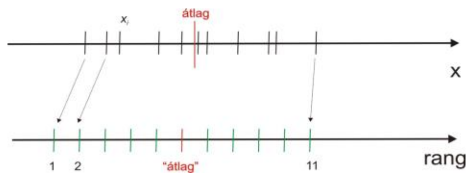
Rang: Egy valamilyen szabály szerint felállított sorban elfoglalt hely.

pl.:
Hadnagy
Őrnagy
Ezredes
Stb.

Kapcsolt rangok:
Azonos értékek esetében
az egyes értékek a rangok átlagát kapják.

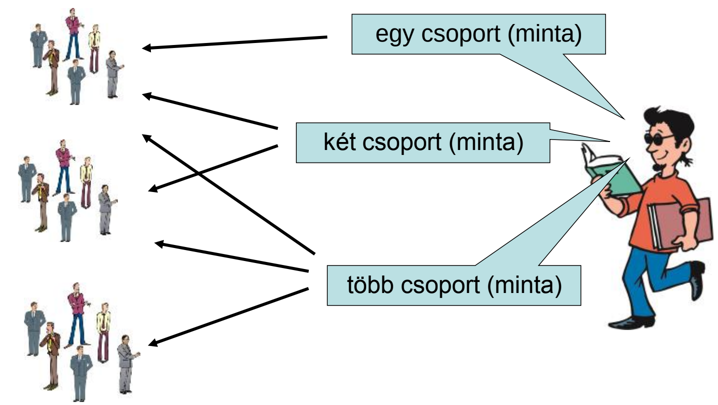
Érték:	1,2	2	2	3,5	4
Rang:	1	2,5	2,5	4	5

A rangok „átlaga” a medián



A medián veszi át az átlag szerepét.

A kérdés szerint



„Variációk egy témára”

	paraméteres	nem paraméteres
egy csoport	egymintás t-próba,	Wilcoxon-féle előjeles rangpróba, előjel-próba
két csoport	kétmintás t-próba	Mann-Whitney U-próba
több csoport	ANOVA	Kruskall-Wallis próba

(A teljesség igénye nélkül)

Vizsgálat egy csoportban

Kérdés: A minta alapján lehet-e a populáció jellemző értéke egy megadott érték?

paraméteres	nem paraméteres
μ = ?	medián = ?
Nullhipotézis: $\overline{x} = \mu_0$	Nullhipotézis: a medián egy megadott érték
egymintás t-próba	Wilcoxon-féle előjel-próba

Egymintás t-próba

A példa: Hatásos-e a lázcsillapító vagy sem?

↓

Nullhipotézis: nem! μ_0 = 0. De az átlag nem 0!

minta	átlag
1.	-0,2 °C
2.	-1 °C
3.	-1,5 °C

Ha az eltérés nagyobb, biztosabbnak tűnik az alternatív hipotézis (a gyógyszer hatásos)

Mit jelent a nagy eltérés?

Mi a mértéke az eltérésnek?

Standard hiba: az átlagok átlagos eltérése a μ-től.

$(\bar{x} \pm s_{\bar{x}})$ ~ 68% - konfidencia intervallum.

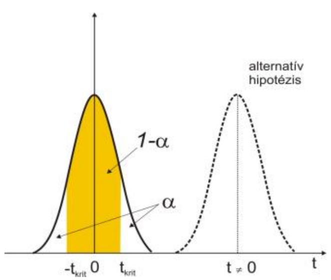
A t-érték

$$t = \frac{x - \mu_0}{s_x}$$

Viszonyítsuk az eltérést a standard hibához! (μ_0 igen gyakran = 0)

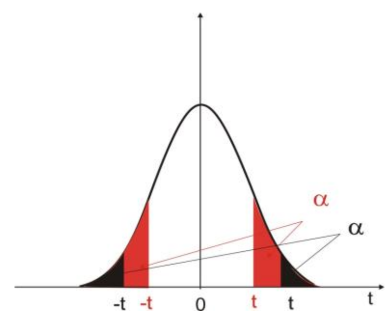
Mivel az átlagok a μ_0 körül ingadoznak, a *t*-értékek a 0 körül.

(feltéve, hogy a nullhipotézis igaz!)



Miért alkalmasabb a t-érték?

Képesek vagyunk kiszámolni ennek az eltérésnek a valószínűségét!!! (Student- vagy *t*-eloszlás)



Csak a *t*-értékek véletlen ingadozását írja le!

Az eloszlás alakja függ az elemszámtól.

Miért t-eloszlású?

$t = \dfrac{\bar{x} - \mu_0}{s_{\bar{x}}}$ ⟶ $\bar{x}$

Az átlagok ingadozása normális eloszlással írható le.
A számláló tehát egy normális eloszlású változó!

$s_{\bar{x}} = \dfrac{s}{\sqrt{n}}$ ⟶ $s = \sqrt{\dfrac{\sum_i (x_i - \bar{x})^2}{n-1}}$

A szórás pedig valószínűségi változók négyzetösszegéből vont négyzetgyök.

emlékeztető

t-eloszlás

$$\xi_n = \frac{\sqrt{n} \cdot \xi}{\sqrt{\sum_i \zeta_i^2}}$$

Q.E.D.
(Quad erat demonstrandum)

A ***t*** változó
(n-1) szabadságfokú
t-eloszlást követ.

A szabadsági fok

Gondoltam 3 számra! (minta)

A 3 szám átlaga: 8! (információ!)

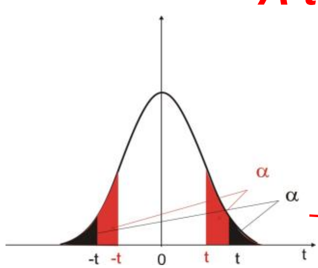
3, 12, 8 vagy 5, 7, 11 stb.

A szabadsági fok = n

3, 12, **9** vagy 5, 7, **12** stb.

A szabadsági fok = n-1

A t-táblázat



Különböző t_{krit} értékek tartoznak a különböző szigfikancia szinthez.

t-táblázat

szignifikancia szint

d.f.	0.1	0.05	0.02	0.01
	probability (p)			
1	6.31	12.7	31.8	63.7
2	2.92	4.3	6.96	9.92
3	2.35	3.18	4.54	5.84
4	2.13	2.78	3.75	4.6
5	2.02	2.57	3.37	4.03

szabadsági fok: (n-1)

Döntés t-táblázat alapján

t-táblázat

szignifikancia szint

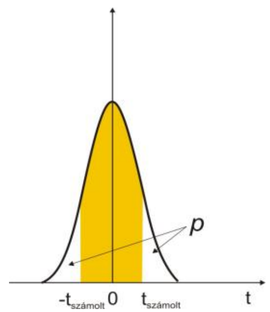
d.f.	0.1	0.05	0.02	0.01
	probability (p)			
1	6.31	12.7	31.8	63.7
2	2.92	4.3	6.96	9.92
3	2.35	3.18	4.54	5.84
4	2.13	2.78	3.75	4.6
5	2.02	2.57	3.37	4.03

szabadsági fok: (n-1)

Kiválasztunk egy alkalmas szignifikancia szintet!

Ha $t \geq 2{,}78$ elvetjük, ha kisebb megtartjuk a nullhipotézist.

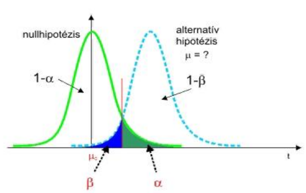
Döntés számítógép segítségével

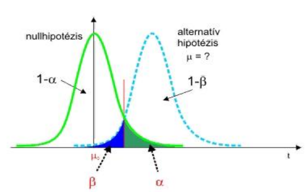


p: annak a valószínűsége, hogy véletlenül ilyen nagy a $t_{számolt}$.

A döntés

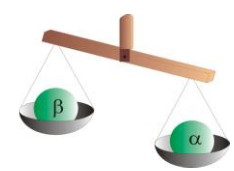
- 1. Ha a véletlen eltérés valószínűsége kicsi ($p(|t|\geq t_{krit}) \leq \alpha$) – ***elvetjük*** a nullhipotézist.
- 2. Ha a véletlen eltérés valószínűsége nagy ($p(|t|\geq t_{krit}) > \alpha$) – ***megtartjuk*** a nullhipotézist.





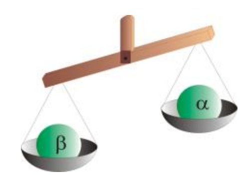
A hiba „mérlegelése”

Elvetjük a nullhipotézist



Az α a tévedés mértéke. Minél kisebb *p* érték a kedvezőbb.

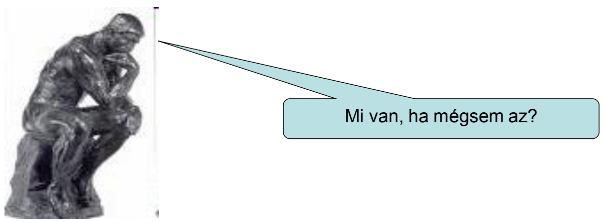
Megtartjuk a nullhipotézist



A β a tévedés mértéke. Minél nagyobb *p* érték a kedvezőbb.

Az egymintás t-próba feltétele

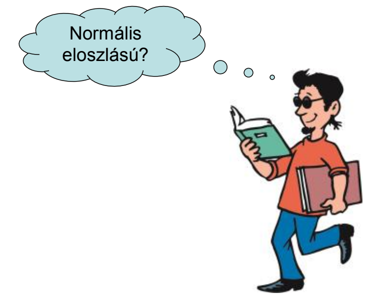
- A feladat: egy minta alapján döntés a μ értékéről.
- A változó **normális eloszlású** legyen.



Wilcoxon-féle előjel-próba

Példa: Van-e hatása egy szórakoztató film megtekintésének, a páciensek együttműködési hajlamára? (A számok pontértékek)

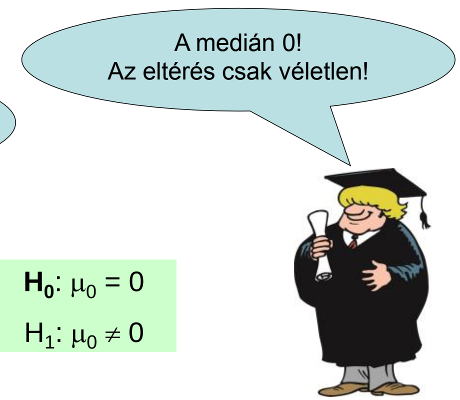
sorszám	előtte	utána	különbség
1	2	2	0
2	0	1	1
3	3	2	-1
4	2	4	2
5	1	3	2
6	3	3	0
7	1	4	3
8	1	5	4
9	5	2	-3
10	4	4	0



A rangok

különbség	abszolút érték	rang	Előjeles rang
0	0		
1	1	1,5	1,5
-1	1	1,5	-1,5
2	2	3,5	3,5
2	2	3,5	3,5
0	0		
3	3	5,5	5,5
4	4	7	7
-3	3	5,5	-5,5
0	0		

A nullhipotézis megfogalmazása



H_0: $\mu_0 = 0$

H_1: $\mu_0 \neq 0$

Ismert eloszlás

De a rangok eloszlása sem ismert!

$$t = \frac{\overline{R} - 0}{s/\sqrt{n}}$$

Ha n elég nagy!!!

$\overline{R}$ -	az előjeles rangok átlaga
s –	a rangok szórása

Emlékeztető!
A rangok átlaga = medián

Döntés

Ó! Innen már tudom!!!

Hát persze! Ez egy egymintás t-próba!!!

Párosított t-próba

Ha az adatok valamilyen szempontból párokba rendezhetőek!

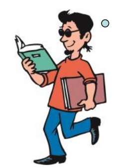
Egyazon egyeden, páros szerven (pl. vese) végzett két megfigyelés.

Ritkábban, szempontok alapján (kor, foglalkozás, stb.) párosítható adatok.

Lásd: lázcsillapító hatása.

Kísérlet-tervezés

Kísérlet-tervezés



Vizsgálat „hozott” anyagból?

Célszerű sorrend: Kérdés felvetés → kísérlet-tervezés → értékelés.

Már meglevő anyag.

Sok problémát vethet fel. (pl. kevés megfelelő adat)

„valódi” egymintás t-próba

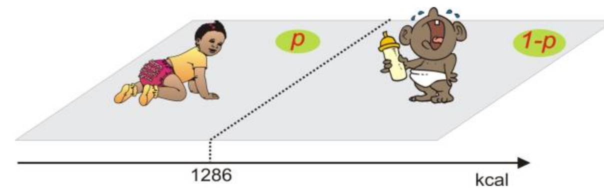
Lehet-e a várható érték egy megadott érték?

$$t = \frac{x - \mu_0}{s_x}$$

Ritkábban előforduló eset.

Előjel-próba

Példa: vizsgálat 2 éves gyerekek populációjában
az energia felvétel nagyságáról.
Kérdés: Lehet-e a medián
(egy másik felmérésből származó érték) 1286 kcal?
Nullhipotézis: a medián 1286 kcal, az eltérés csupán véletlen.



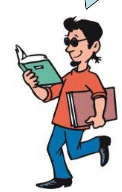
A vizsgálat

Kis elemszám esetében	Nagy elemszám esetében
$P(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$	$z = \frac{\lvert x - np \rvert - 1/2}{\sqrt{np(1-p)}}$
binomiális eloszlás	standard normális eloszlás

x – gyerekek száma 1286 kcal alatt
n – vizsgálatba bevont gyerekek száma
p – annak a valószínűsége, hogy véletlenül kisebb legyen (lásd: binomiális eloszlás)

A döntés

Kiszámoljuk a véletlen eltérés valószínűségét. (binomiális, vagy standard normális eloszlás)



Vége ennek a résznek!