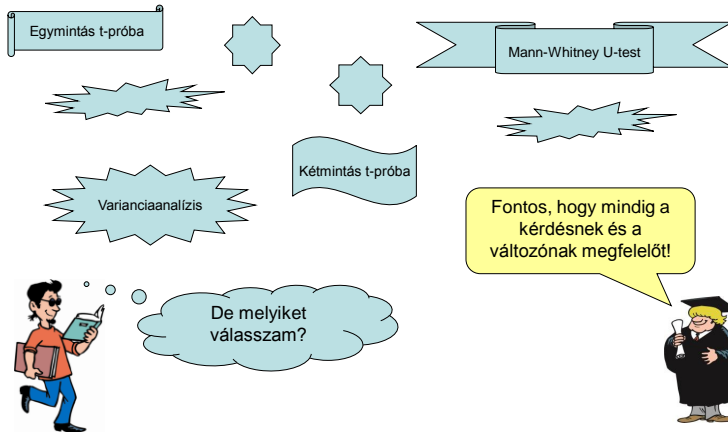


Kiválasztás



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.

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

Rangok



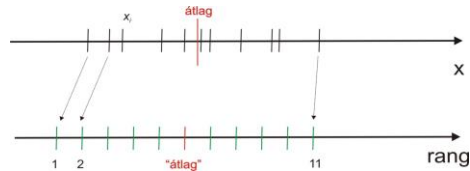
Rang: Egy valamilyen szabály szerint felállított sorban elfoglalt hely.

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

pl.:
Hadnagy
Őrmagy
Ezredes
Stb.

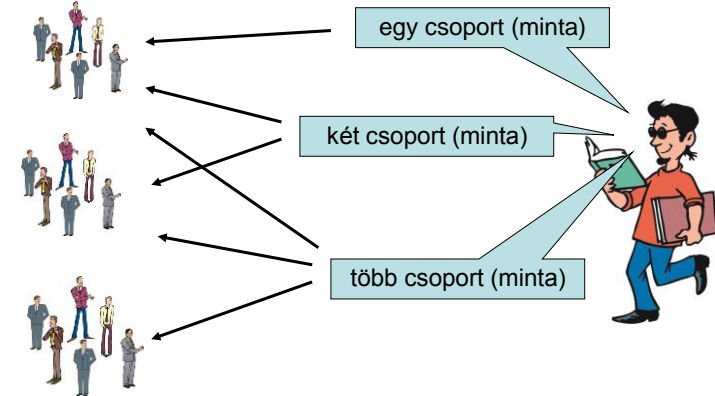
Érték:	1,2	2	2	3,5	4
Rang:	1	2,5	2,5	4	5

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

$\mu = ?$

Nullhipotézis: $\bar{x} = \mu_0$

egymintás t-próba

nem paraméteres

medián = ?

Nullhipotézis: a medián egy megadott érték

Wilcoxon-féle előjel-próba

Egymintás t-próba

A példa: Hatásos-e a lázcsillapító vagy sem?



Nullhipotézis: nem! $\mu_0 = 0$. 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 -tól.

$$(\bar{x} \pm s_{\bar{x}})$$

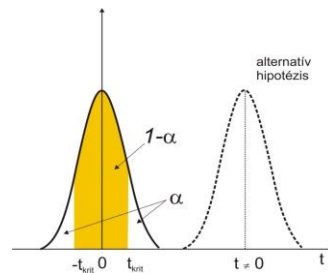
~ 68% - konfidencia intervallum.

A t-érték

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}}$$

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!)



A szabadsági fok

Gondoltam 3 számra!
(minta)

A 3 szám átlaga: 8!
(információ!)



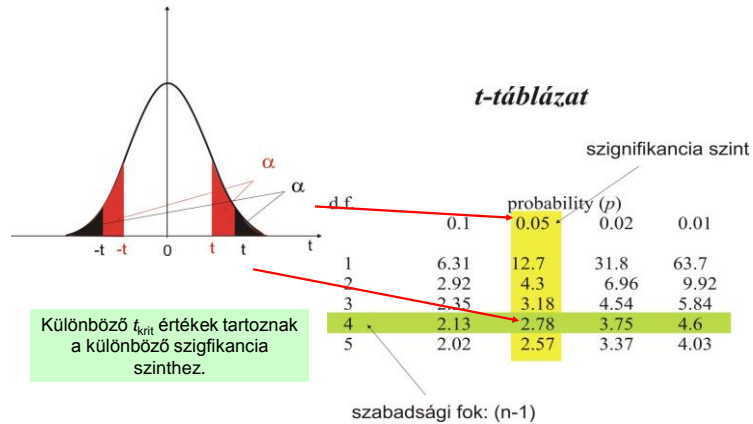
3, 12, 8 vagy 5, 7, 11 stb.

A szabadsági fok = n

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

A szabadsági fok = n-1

A t-táblázat



Döntés t-táblázat alapján

t-táblázat

szignifikancia szint

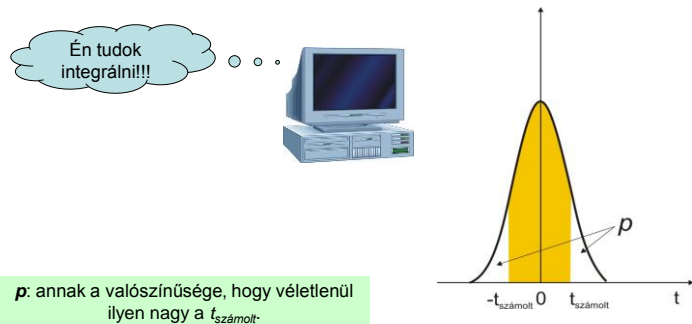
d.f.	0.1	0.05	0.02	0.01
1	6.31	12.7	31.8	63.7
2	2.92	4.3	6.96	9.92
3	2.35	3.18	4.54	5.84
4	2.13	2.78	3.75	4.6
5	2.02	2.57	3.37	4.03

szabadsági fok: (n-1)

Kiválasztunk egy alkalmas szignifikancia szintet!
Ha $\approx 2,78$ elvetjük, ha kisebb megtartjuk a nullhipotézist.



Döntés számítógép segítségével



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.

Az egymintás t-próba feltétele

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



Mi van, ha mégsem az?

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

Normális eloszlású?



A rangok

A különbségek abszolút értékét (kivéve a 0 értékeket) állítsuk sorba! A rangoknak adjuk vissza az előjelét!
Majd számoljuk ki a rangok átlagát és szórását.



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

Nincs hatása a filmnek!



A medián 0!
Az eltérés csak véletlen!



$$H_0: \mu_0 = 0$$

$$H_1: \mu_0 \neq 0$$

Ismert eloszlás

De a rangok eloszlása sem ismert!

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

Ha n elég nagy!!!

\bar{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 rendezhetők!

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

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

Lásd: lázcsillapító hatása.



Kísérlet-tervezés

„valódi” egymintás t-próba

Lehet-e a várható érték egy megadott érték?

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



Ritkábban előforduló eset.



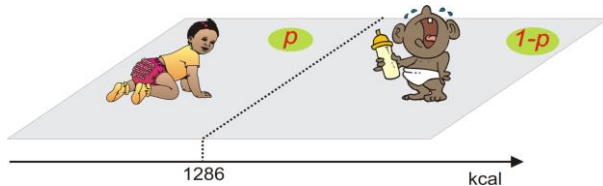
Előjel-próba

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

$$P(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

binomiális eloszlás

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



Nagy elemszám esetében

$$z = \frac{|x - np| - 1/2}{\sqrt{np(1-p)}}$$

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)

Vizsgálat két csoportban

Kérdés: A két minta származhat-e azonos populációból, vagy a két populáció paraméterei azonosak?

paraméteres

$$\mu_1 = \mu_2 ?$$

Nullhipotézis: $\mu_1 = \mu_2$

kétmintás t-próba

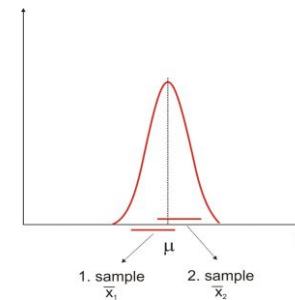
nem paraméteres

Nullhipotézis: a két minta azonos populációból származik.

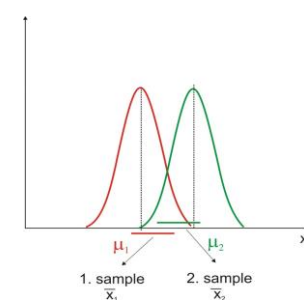
Mann-Whitney U-próba

Kétmintás t-próba

egy populáció
(a két átlag eltérése véletlen)



két populáció
(a két átlag eltérése nem véletlen)



Standard hiba

$$s_1 = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n_1 - 1}} \quad s_2 = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n_2 - 1}}$$

$$s_{\bar{x},1} = \frac{s_1}{\sqrt{n_1}} \quad s_{\bar{x},2} = \frac{s_2}{\sqrt{n_2}}$$

Közös standard hiba: a két standard hiba súlyozott átlaga.

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{Q_1 + Q_2}{n_1 + n_2 - 2} \cdot \frac{1}{n_1} + \frac{1}{n_2}}$$

Kétmintás t-próba

$$\bar{x}_1 \neq \bar{x}_2$$



Ismert eloszlású változóra van szükség!



$$t = \frac{\bar{x}_1 - \bar{x}_2}{s^* \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}}$$

$$s^* = \sqrt{\frac{Q_1 + Q_2}{n_1 + n_2 - 2}}$$

A kétmintás t-próba feltétele

- A feladat: két egymástól **független** csoport összehasonlítása.
- A változó **normális eloszlású** legyen.
- A **szórás** a két csoportban **azonos**nak tekinthető.



Ez utóbbi új!
Hogyan állapítható meg?