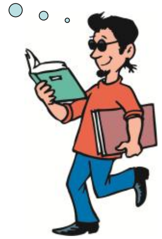
Wilcoxon-féle előjel-próba

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 rendezhetőek!

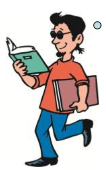
Lásd: lázcsillapító hatása.

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

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

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

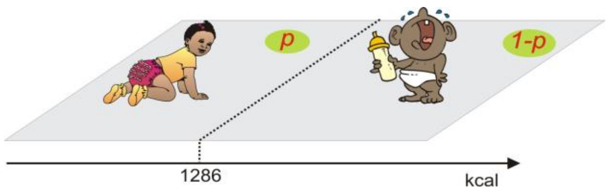
Lehet-e a várható érték egy megadott érték?

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

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

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

binomiális 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)

A döntés

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

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

?

Ismert eloszlású változóra van szükség!

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s^* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

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

A próba

A t-érték az t-érték!

Akkor meg tudom csinálni! Pardon, mennyi a szabadsági fokok száma?

sz.f. = $n_1 + n_2 - 2$

$((n_1-1)+(n_2-1))$

A kétmintás t-próba feltétele

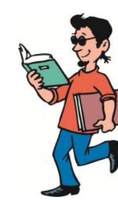
- A feladat: két egymástól **<u>független</u>** csoport összehasonlítása.
- A változó **<u>normális eloszlású</u>** legyen.
- A **<u>szórás</u>** a két csoportban **<u>azonos</u>**nak tekinthető.

Ez utóbbi új! Hogyan állapítható meg?

A szórások vizsgálata

Hogyan fogjunk hozzá?

Nullhipotézis: a két szórás azonos, az eltérés véletlen (mintavétel).



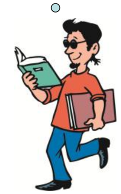
De hiszen ez olyan, mint egy hipotézis-vizsgálat!

Az F-próba

$$F = \frac{s_1^2}{s_2^2}$$

A nullhipotézishez tartozik egy ún. F-eloszlás.

De melyik variancia legyen a számlálóban?



A számlálóban mindig a nagyobb variancia van! ($F \geq 1$)

A döntés

- 1. Ha a véletlen eltérés valószínűsége kicsi ($p(F>F_{krit}) \leq 5\%$) – ***<u>elvetjük</u>*** a nullhipotézist.
- 2. Ha a véletlen eltérés valószínűsége nagy ($p(F>F_{krit}) > 5\%$) – ***<u>megtartjuk</u>*** a nullhipotézist.

Mann-Whitney U-próba

Példa: hatásos-e a fejfájás-csillapító?

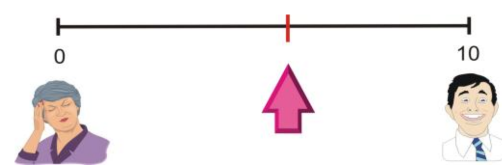
Doktor úr! Csináljon valamit.

Hogyan mérhető a hatás?

Kísérlet

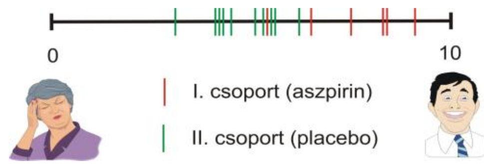
I. csoport: (eset) aszpirint kap

II. csoport: (kontroll) placebo-t kap (hatóanyag nélküli tabletta)



Ez egy önkényes, folytonos skála.

Eredmények



érték	3,1	4,1	4,2	4,3	4,5	5,1	5,3	5,4	5,5
rang	1	2	3	4	5	6	7	8	9
érték	5,6	6,2	6,2	6,5	7,5	8,3	8,3	8,4	9,1
rang	10	11,5	11,5	13	14	15,5	15,5	17	18

A nullhipotézis megfogalmazása

a „gyógyszer” nem hatásos.

A két csoport azonos populációhoz tartozik.

A rangok összege

(avagy a kis Gauss esete a tanárral)

Gyerekek! Adjátok össze a számokat 1-től százig.

Miért adjam össze? Könnyebben is kiszámolható!

1 + 100 = 101
2 + 99 = 101 …

$$\sum_{i=1}^{n} i = \frac{n}{2}\cdot(n+1)$$

A „nagy átalakítás”

T – a rangok összege az I. csoportban, véletlen eloszlás esetén a várható értéke:

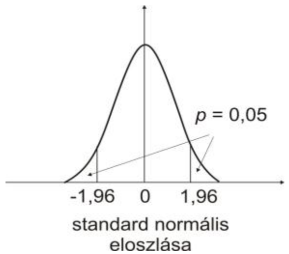
A ***z*** változó standard normális eloszlású. (ha n elég nagy)

$$n_1 \cdot \frac{n_1+n_2+1}{2}$$

$$z = \frac{T - n_1(n_1+n_2+1)/2}{s}$$

$$s = \sqrt{\frac{n_1 \cdot n_2 \cdot (n_1+n_2+1)}{12}}$$

Döntés



A kiszámolt *z*-érték: 3,24.

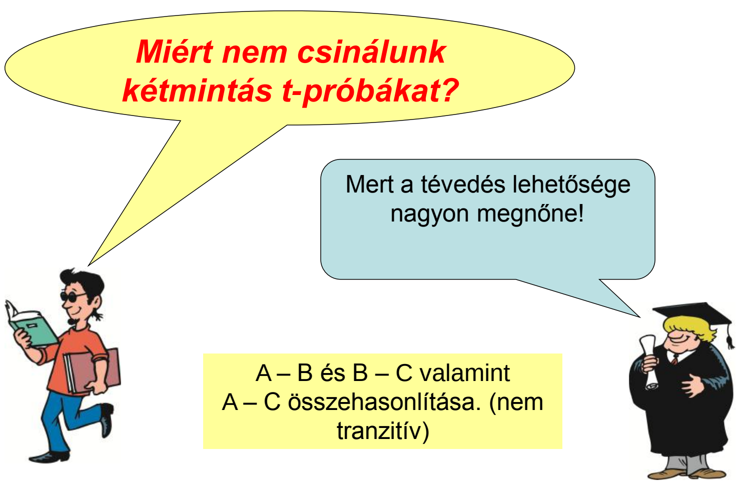
Ez nagyobb, mint az 1,96.

Következtetés: a nullhipotézist elvetjük.

Kiszámolt p-érték < 0,1%.

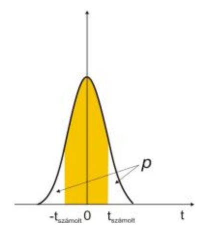
Következtetés hasonló a fentihez.

Variancia-analízis (ANOVA)



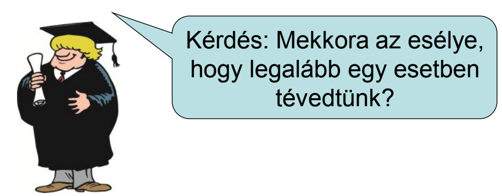
A – B és B – C valamint A – C összehasonlítása. (nem tranzitív)

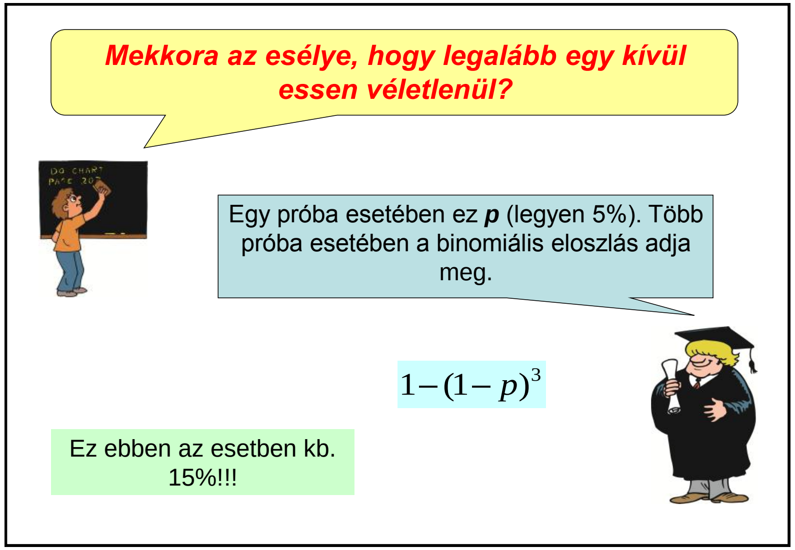
Mekkora a tévedés esélye?

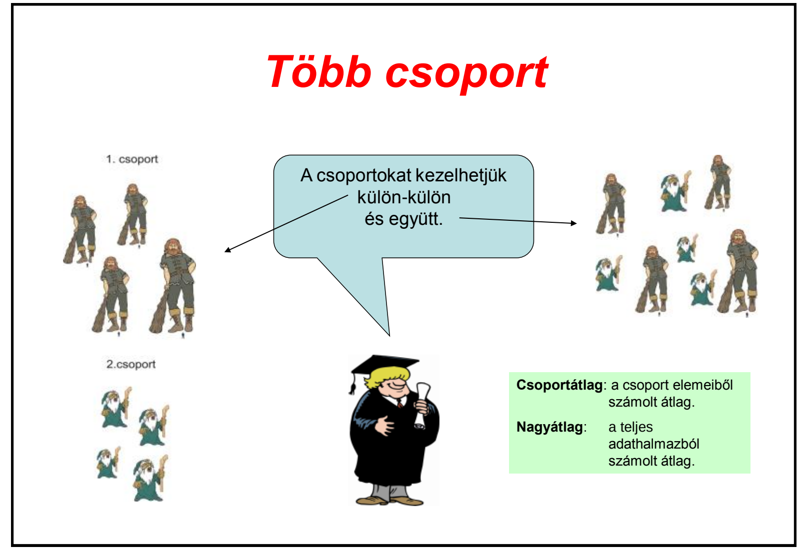


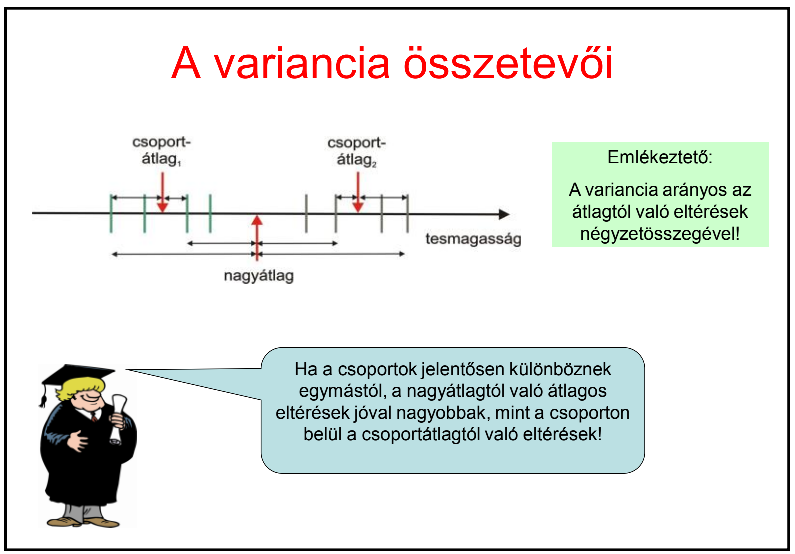
Tételezzük fel, hogy mindhárom esetben elvetjük a nullhipotézist.

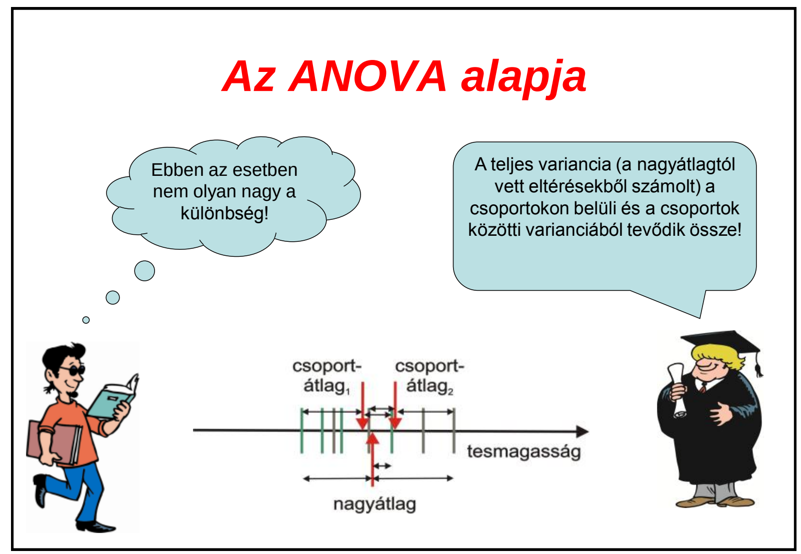
p – annak a valószínűsége, hogy kívül van (**1-p**) – annak, hogy belül van az érték véletlenül.




Mekkora az esélye, hogy legalább egy kívül essen véletlenül?
Egy próba esetében ez ***p*** (legyen 5%). Több próba esetében a binomiális eloszlás adja meg.
$1-(1-p)^3$
Ez ebben az esetben kb. 15%!!!


Több csoport
1. csoport
2.csoport
A csoportokat kezelhetjük külön-külön és együtt.
Csoportátlag: a csoport elemeiből számolt átlag.
Nagyátlag: a teljes adathalmazból számolt átlag.


A variancia összetevői
csoport-átlag$_1$
csoport-átlag$_2$
tesmagasság
nagyátlag
Emlékeztető:
A variancia arányos az átlagtól való eltérések négyzetösszegével!
Ha a csoportok jelentősen különböznek egymástól, a nagyátlagtól való átlagos eltérések jóval nagyobbak, mint a csoporton belül a csoportátlagtól való eltérések!


Az ANOVA alapja
Ebben az esetben nem olyan nagy a különbség!
A teljes variancia (a nagyátlagtól vett eltérésekből számolt) a csoportokon belüli és a csoportok közötti varianciából tevődik össze!
csoport-átlag$_1$
csoport-átlag$_2$
tesmagasság
nagyátlag

A variancia felbontása

(nem kell tudni)

Legyen *N* adat *k* csoportban.

x_{ij} – a j-edik adat az i-edik csoportban,

n_i – az elemek száma az i-edik csoportban

nagyátlag $x = \frac{\sum_{i,j} x_{ij}}{N}$

csoportátlag $x_i = \frac{\sum_j x_{ij}}{n_i}$

$$(x - x_{ij}) = (x - x_i) + (x_i - x_{ij})$$

az adat eltérése a nagyátlagtól

a csoportátlag eltérése

az adat eltérése a csoportátlagtól

Emeljük négyzetre!

$$(x - x_{ij})^2 = (x - x_i)^2 + (x_i - x_{ij})^2 + 2(x - x_i)(x_i - x_{ij})$$

$$\sum (x - x_{ij})^2 = \sum (x - x_i)^2 + \sum (x_i - x_{ij})^2 + \sum 2(x - x_i)(x_i - x_{ij})$$

kovariancia = 0 (függetlenek)

$$\sum 2(x - x_i)(x_i - x_{ij}) = 0$$

$$\sum (x - x_{ij})^2 = \sum (x - x_i)^2 + \sum (x_i - x_{ij})^2$$

Csoportok közötti

Csoporton belüli

A variancia felbontható komponensekre!

A nullhipotézis

A csoportok között nincs különbség.

A csoportok átlagai közötti eltérés csupán a véletlen műve.

Döntés: a csoportok közötti és a csoporton belüli varianciák összehasonlítása alapján.

A varianciák kiszámolása

A variancia	négyzetösszeg	szab. fok		F érték
csoportok között	$SS_A = \sum_j n_j (\bar{x}_j - \bar{x})^2$	k-1	$MS_A = \frac{SS_A}{k-1}$	$F = \frac{MS_A}{MS_E}$
csoporton belül	$SS_E = SS_T - SS_A$	N-k	$MS_E = \frac{SS_E}{N-k}$	
teljes	$SS_T = \sum_{i,j} (x_{i,j} - \bar{x})^2$	N-1		

k – a csoportok száma

A döntés

- 1. Ha a véletlen eltérés valószínűsége kicsi ($p(F>F_{krit}) \leq 5\%$) – ***elvetjük*** a nullhipotézist.
- 2. Ha a véletlen eltérés valószínűsége nagy ($p(F>F_{krit}) > 5\%$) – ***megtartjuk*** a nullhipotézist.

(A döntés után, ha szükségesnek tartjuk, csinálhatunk *t*-próbákat)

Kruskal-Wallis próba

Az adatokat a csoportoktól függetlenül rangsoroljuk! Az így kapott rangokat a csoportokon belül összegezzük!

Ha a változó nem normális eloszlású!

A nullhipotézis

A csoportok között nincs különbség.

A rangok összege közötti eltérés csupán a véletlen műve.

Milyen eloszlást használjunk?

$$H = \frac{12}{N(N+1)} \sum_i \frac{R_i^2}{n_i} - 3(N+1)$$

N – az elemek száma
R_i – a rangok összege az i-edik csoportban
n_i – az elemek száma az i-edik csoportban

Közjáték *(nem kell megtanulni)*

$$H = \frac{12}{N(N+1)}\sum_i \frac{R_i^2}{n_i} - 3(N+1) \longrightarrow \sum_i \frac{R_i^2}{n_i} = \sum_i n_i\left(\frac{R_i}{n_i}\right)^2 = \sum_i n_i R_i^2$$

A rangok átlaga: $R_i = R + \Delta_i$

Δ_i – az i-edik csoport átlagának eltérése a főátlagtól ($\Sigma\Delta_i = 0$)

$$R = \frac{N(N+1)}{2N} = \frac{N+1}{2} \longrightarrow \sum_i n_i R_i^2 = R^2\sum_i n_i + \sum_i n_i\Delta_i^2 + 2R\sum_i n_i\Delta_i (=0)$$

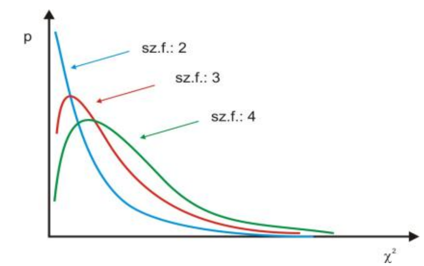
$$\sum_i n_i R_i^2 = \frac{(N+1)^2}{4}N + \sum_i n_i\Delta_i^2$$

$$H = \frac{12}{N(N+1)}\left(N\frac{(N+1)^2}{4} + \sum_i n_i\Delta_i^2\right) - 3(N+1) \longrightarrow H = \frac{12}{N(N+1)}\sum_i n_i\Delta_i^2$$

A χ^2-eloszlás

A *H* értéke ≥ 0!

A ***H*** változó χ^2-eloszlást követ!



A döntés

- 1. Ha a véletlen eltérés valószínűsége kicsi ($p(\chi^2 > \chi^2_{krit}) \leq 5\%$) – ***elvetjük*** a nullhipotézist.
- 2. Ha a véletlen eltérés valószínűsége nagy ($p(\chi^2 > \chi^2_{krit}) > 5\%$) – ***megtartjuk*** a nullhipotézist.

Hipotézis vizsgálat?

- Felállítjuk a **nullhipotézis**t.
- Keresünk egy **ismert eloszlású változó**t.
- Az eloszlás alapján kiszámoljuk a **véletlen eltérés valószínűségét**.
- Ha ez kisebb mint a szignifikancia szint **elvetjük**, ellenkező esetben **megtartjuk a nullhipotézist**.
- Ennyi!