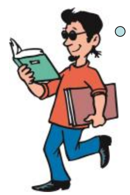
A szórások vizsgálata

Hogyan fogjunk hozzá?

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



De hiszen ez olyan, mint egy hipotézis-vizsgálat!

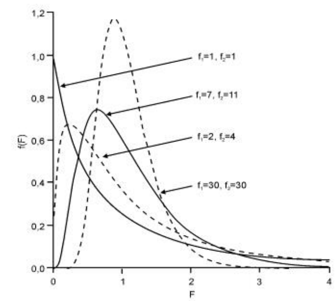
Az F-próba

A nullhipotézishez tartozik egy ún. F-eloszlás.

Szabadsági fokok:

számláló: n_1-1
nevező: n_2-1

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



Az F-próba szabadsági fokai

Számítógéppel számolva, bármelyik lehet.
Táblázatot használva, viszont mindig a nagyobb.
(Ennek megfelelően kell a sz.f.-okat figyelembe venni)

De melyik variancia legyen a számlálóban?

A döntés

- 1. Ha a véletlen eltérés valószínűsége kicsi ($p \leq \alpha$) – ***elvetjük*** a nullhipotézist.
- 2. Ha a véletlen eltérés valószínűsége nagy ($p > \alpha$) – ***megtartjuk*** a nullhipotézist.

F táblázat

Számítógép: F-próba

Ha a két szórás nem azonos!

Korrekciók:

A szabadsági fokok korrekciója.

a t-érték korrekciója.

Mann-Whitney U-próba

Példa: hatásos-e a fejfájás-csillapító?

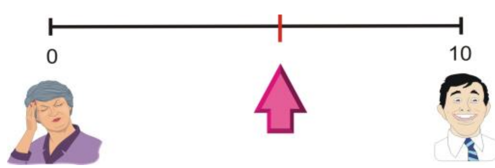
Doktor úr!
Csináljon valamit.

Hogyan mérhető a hatás?

Kísérlet

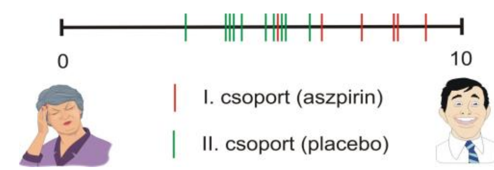
I. csoport:
(eset)
aszpirint kap

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 rangok összege

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

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

(n_1 elem, amelyek átlaga = $(n_1+n_2+1)/2$)

Nullhipotézis: az ettől való eltérés véletlen.

Kis n: egy U-eloszlás írja le a véletlen eltérés valószínűségét.

A „nagy átalakítás”

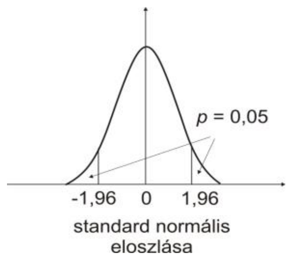
Ha n elég nagy:

A ***z*** változó standard normális eloszlású.

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

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

Döntés



A kiszámolt *z*-érték: 3,24.

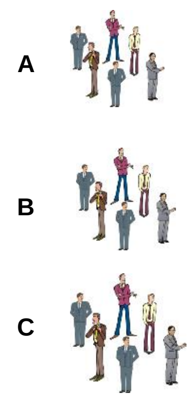
Ez nagyobb, mint az 1,96.

Következtetés: a nullhipotézist elvetjük.

Kiszámolt p-érték < 0,1%.

Következtetés hasonló a fentihez.

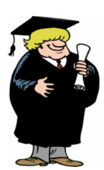
Variancia-analízis (ANOVA)



Nincs, az eltérés csak véletlen! Ez a nullhipotézis.

Több csoport

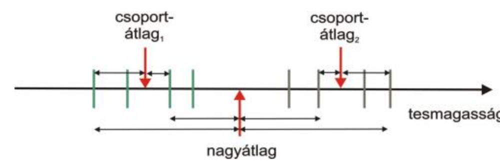
A csoportokat kezelhetjük külön-külön és együtt!



Csoportátlag: a csoport elemeiből számolt átlag.

Nagyátlag: a teljes adathalmazból számolt átlag.

A variancia összetevői



Emlékeztető:

A variancia arányos az átlagtól való eltérések négyzetösszegével!

$$(x_{i,j}-\bar{x}) = (x_{i,j}-\bar{x}_j)+(\bar{x}_j-\bar{x})$$

csoporton belüli (pl. véletlen) eltérés

csoportok közötti különbség

Ha a csoportok jelentősen különböznek egymástól, a nagyátlagtól való átlagos eltérések jóval nagyobbak, mint a csoporton belül a csoportátlagtól való eltérések!

$\bar{x}$ - nagyátlag

$\bar{x}_j$ - csoportátlag

A varianciák kiszámolása

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

$\bar{x}$ nagyátlag

$\bar{x}_j$ j-edik csoportátlag

N – összes elem száma

k – csoportok száma

A nullhipotézis

A csoportok között nincs különbség.

A csoportok közötti eltérés csupán a véletlen műve.

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

Hogyan hasonlítjuk össze?

Varianciák összehasonlítása? Ilyenről már volt szó!

Valóban, a kétmintás t-próba esetében.

$$F = \frac{MS_A}{MS_E}$$

A döntés

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

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

Az ANOVA feltétele

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

Kruskal-Wallis próba

Ha a változó
nem normális eloszlású!

Az adatokat a csoportoktól
függetlenül rangsoroljuk!

Rangsorolás

	1. csoport	2. csoport	3. csoport
1	173	170	175
2	175	163	174
3	169	165	171
4	168		172
5			172

elem	163	165	168	169	170	171	172	172	173	174	175	175
rang	1	2	3	4	5	6	7,5	7,5	9	10	11,5	11,5

csoport	elemszám	rangok összege
1	4	27,5
2	3	8
3	5	42,5

A nullhipotézis

A csoportok között nincs
különbség.

A rangok „átlaga” közötti
eltérés csupán a véletlen
műve.

Milyen eloszlást használjunk?

Akkor jön az átalakítás!

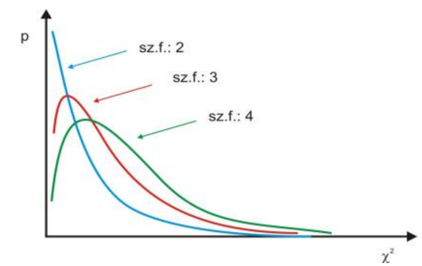
$$H = \frac{12}{N(N+1)}\sum_i \frac{R_i^2}{n_i} - 3(N+1)$$

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

A H értéke ≥ 0!

A χ^2-eloszlás

Emlékeztető:
χ^2-eloszlás – normális eloszlású változók négyzetösszege esetén lép fel.



A szabadsági fokok száma = csoportok száma -1

A döntés

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

Példa

ÖSSZESÍTÉS						
Csoportok	Darabszám	Összeg	Átlag	Variancia		
Oszlop 1	4	685	171,25	10,91667		
Oszlop 2	3	498	166	13		
Oszlop 3	5	864	172,8	2,7		
VARIANCIAANALÍZIS						
Tényezők	SS	df	MS	F	p-érték	F krit.
Csoportok között	89,36666667	2	44,68333	5,782171	0,02427	4,256495
Csoporton belül	69,55	9	7,727778			
Összesen	158,9166667	11				

α = 0,05
p = 0,024

Döntés:
elvetjük a nullhipotézist, a példa alapján a csoportok szignifikánsan különböznek egymástól.

Példa

Csoport	Elemszám (n_i)	Rangok összege (R_i)
1	4	27,5
2	3	8
3	5	42,5

N = 12

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

$$4{,}97=\frac{12}{12(12+1)}\left(\frac{27{,}5^2}{4}+\frac{8^2}{3}+\frac{42{,}5^2}{5}\right)-3(12+1)$$

sz.f. = 3 – 1 = 2

α = 0,05
p = 0,083

Döntés:

megtartjuk a nullhipotézist, a példa alapján a csoportok nem különböznek egymástól szignifikánsan.

Hasonlítsuk össze!

	1. csoport	2. csoport	3. csoport
1	173	170	175
2	175	163	174
3	169	165	171
4	168		172
5			172

ANOVA

α = 0,05
p = 0,024

Döntés:
elvetjük a nullhipotézist.

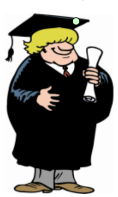
!!!

Kruskall-Wallis próba

α = 0,05
p = 0,083

Döntés:
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!