

Variancia-analízis (ANOVA)



Miért nem csinálunk kétmintás t-próbákat?



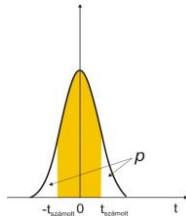
A – B és B – C valamint A – C összehasonlítása. (nem tranzitív)

csoportok száma	próbák száma
3	3
4	6
5	10

Mert a tévedés lehetősége nagyon megnőne!



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.



Kérdés: Mekkora az esélye, hogy legalább egy esetben tévedtünk?

Mekkora az esélye, hogy legalább egy kívül essen véletlenül?



Egy próba esetében ez p (legyen 5%). Több próba esetében a binomiális eloszlás adja meg.

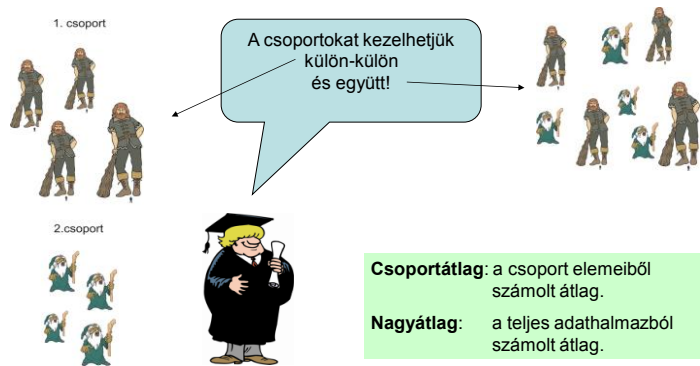
Ez ebben az esetben kb. 15%!!!

5% eléréséhez a p értékének kb. 5/3 %-nak kell lennie.

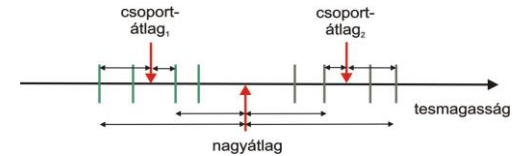
$$1 - (1 - p)^3$$



Több csoport



A variancia összetevői



Emlékeztető:
A variancia arányos az átlagtól való eltérések négyzetösszegével!

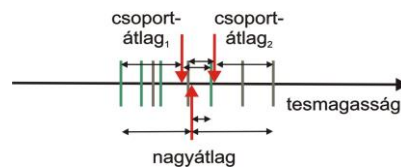


Ha a csoportok jelentősen különböznek egymástól, a nagyátlagtól való átlagos eltérések jóval nagyobbak, mint a csoporton belül a csoportátlagtól való eltérések!

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!



A variancia összetevői

	1. csoport	2. csoport	3. csoport
1	173	170	175
2	175	163	174
3	169	165	171
4	168		172
5			172
csoportátlag	171,25	166	172,8

nagyátlag = 170, 58

$$(170 - 170,58) = (170 - 166) + (166 - 170,58)$$

$$(175 - 170,58) = (175 - 172,8) + (172,8 - 170,58)$$

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

csoportok közötti különbség

csoporton belüli (pl. véletlen) eltérés

\bar{x} - nagyátlag

\bar{x}_j - csoportátlag

A variancia felbontása

(Nem kell tudni!)

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



$$(x_{i,j} - \bar{x})^2 = (x_{i,j} - \bar{x}_j)^2 + (\bar{x}_j - \bar{x})^2 + 2(x_{i,j} - \bar{x}_j)(\bar{x}_j - \bar{x})$$



$$\sum (x_{i,j} - \bar{x})^2 = \sum (x_{i,j} - \bar{x}_j)^2 + \sum (\bar{x}_j - \bar{x})^2 + 2 \sum (x_{i,j} - \bar{x}_j)(\bar{x}_j - \bar{x})$$



$$\sum (x_{i,j} - \bar{x})^2 = \sum (x_{i,j} - \bar{x}_j)^2 + \sum (\bar{x}_j - \bar{x})^2$$

SS_T

SS_E

SS_A

a csoporton belüli és a csoportok közötti eltérések kovarianciája. Ezeket függetlennek feltételezve, ez a tag nulla!

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}
 \bar{x}_j

nagyátlag

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

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

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		
VARIANCIANALÍZIS						
Tényezők	SS	df	MS	F	p-érték	F krit.
Csoportok között	89,3666667	2	44,68333	5,782171	0,02427	4,256495
Csoporton belül	69,55	9	7,727778			
Összesen	158,916667	11				

$\alpha = 0,05$
 $p = 0,024$

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

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?

A H változó χ^2 -eloszlást követ!

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

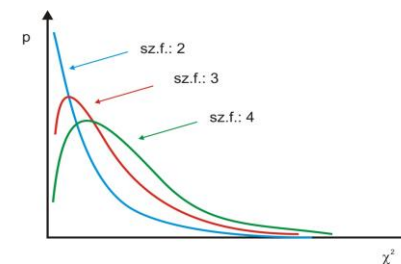
A H értéke ≥ 0 !



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

Példa

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

$$N = 12$$

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

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

$$\text{sz.f.} = 3 - 1 = 2$$

$$\alpha = 0,05$$

$$p = 0,083$$

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

$$\alpha = 0,05$$

$$p = 0,024$$

Döntés:
elvetjük a nullhipotézist.

Kruskal-Wallis próba

$$\alpha = 0,05$$

$$p = 0,083$$

Döntés:
megtartjuk a nullhipotézist.



Hipotézis
vizsgálat?



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

