

# VARIANCIANALÍZIS (ANOVA)

## Véletlen faktor esetén

Varianciakomponens-elemzés

## **Rögzített faktorok:**

szintjeiket a kísérletekhez megválaszthatjuk és beállíthatjuk

Szakmai kérdés:

van-e különbség a faktor különböző szintjei között, melyik közülük a legjobb?

## **Véletlen faktor:**

szintjeit egy elképzelt sokaságból véletlenszerűen választjuk ki

Szakmai kérdés:

a faktornak van-e hatása az ingadozásra, több véletlen faktor közül melyik milyen mértékben járul hozzá az ingadozáshoz, a jövőben mekkora ingadozás várható?

# Egy véletlen faktor szerinti varianciaanalízis

## 3. példa (Napszem.sta)

Egy elemzést három napon kétszer-kétszer végeztek el.

Okoz-e többletingadozást az, hogy különböző napokon végezték a méréseket?

	1. nap	2. nap	3. nap	
	96.897	96.905	97.495	
	96.963	97.567	97.195	
$y_{i.}$	96.930	97.236	97.345	$y_{..}=97.1705$

**A modell:**  $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$

$\alpha_i$  a faktor  $i$ -edik szintjének ( $i$ -edik nap) hatása  
 $\mu$  közös érték;  $r+1$  paraméter

rögzített faktornál  $H_0 : \alpha_i = 0, \quad i = 1, \dots, r$

véletlen faktornál  $\alpha_i \sim N(0, \sigma_A^2)$

$$E(\alpha) = 0 \quad \text{Var}(\alpha) = \sigma_A^2$$

$$H_0 : \sigma_A^2 = 0$$

$$\sum_i^r p_i \alpha_i \neq 0$$

# ANOVA-táblázat egy véletlen faktorra

az eltérés forrása	eltérés-négyzetösszeg	szabadsági fokszám	szórás-négyzet	szórásnégyzet várható értéke	$F_0$
A hatása (csoportok közötti)	$S_A = p \sum_i (y_{i.} - y_{..})^2$	$r-1$	$s_A^2 = \frac{S_A}{r-1}$	$\sigma_e^2 + p\sigma_A^2$	$\frac{s_A^2}{s_R^2}$
Ismétlések (csoportokon belüli)	$S_R = \sum_i \sum_j (y_{ij} - y_{i.})^2$	$r(p-1)$	$s_R^2 = \frac{S_R}{r(p-1)}$	$\sigma_e^2$	
Teljes	$S_0 = \sum_i \sum_j (y_{ij} - y_{..})^2$	$rp-1$			

$$\begin{array}{l}
 r-1 \quad \longrightarrow \quad \frac{s_A^2}{\sigma_e^2 + p\sigma_A^2} \\
 \\
 r(p-1) \quad \longrightarrow \quad \frac{s_R^2}{\sigma_e^2}
 \end{array}
 \quad
 H_0 : \sigma_A^2 = 0
 \quad
 F_0 = \frac{s_A^2}{s_R^2}$$

### 3. példa ANOVA táblája

Univariate Tests of Significance for Y (Napszem)								
Over-parameterized model								
Type III decomposition								
Include condition: szem=1								
Effect	Effect (F/R)	SS	Degr. of Freedom	MS	Den.Syn. Error df	Den.Syn. Error MS	F	p
Intercept	Fixed	56652.44	1	56652.442	2.000000	0.092581	611925.190	0.000002
NAP	Random	0.19	2	0.093	3.000000	0.088767	1.043	0.453029
Error		0.27	3	0.089				

Elfogadjuk a nullhipotézist.

$$F_0 = \frac{S_A^2}{S_R^2}$$

$$H_0 : \sigma_A^2 = 0$$

# Varianciakomponens-elemzés

**Ha** a  $H_0 : \sigma_A^2 = 0$  hipotézist elutasítjuk, becsülnünk kell a  $\sigma_A^2$  varianciát:

$$E(s_A^2) = \sigma_e^2 + p\sigma_A^2$$

$$E(s_R^2) = \sigma_e^2$$

$$\widehat{(\sigma_e^2 + p\sigma_A^2)} = s_A^2$$

$$\widehat{\sigma_e^2} = s_R^2$$

$$\widehat{\sigma_A^2} = \frac{s_A^2 - s_R^2}{p}$$

*Summary fölön: Random effects > Var. comp.*

Components of Variance (Napszem Over-parameterized model Type III decomposition)	
Effect	Y
NAP	0.001907
Error	0.088767

## Keresztosztályozás két véletlen faktor szerint

### 4. példa (Napszem.sta)

Egy elemzést nemcsak különböző napokon végeztek el, hanem különböző személyek is.

Az, hogy a mérést különböző napokon és különböző személyek végzik, okoz-e többletingadozást az egy nap egy személy végezte ismétlések szóródásához képest?



## 4. példa adatai:

	1. nap	2. nap	3. nap	$y_{.j}$
1. személy	96.897 96.963	96.905 97.567	97.495 97.195	97.170
2. személy	97.232 97.184	97.241 97.025	97.215 97.581	97.247
3. személy	96.988 96.797	97.202 97.324	97.352 97.283	97.158
4. személy	97.035 97.095	97.339 97.318	97.388 97.168	97.224
$y_{i..}$	97.024	97.240	97.335	$y_{...}=97.200$

# Modell

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{k(ij)}$$

nap

személy

kölcsönhatás

ismétlési hiba

$$\alpha_i \sim N(0, \sigma_A^2)$$

$$\beta_j \sim N(0, \sigma_B^2)$$

$$\alpha\beta_{ij} \sim N(0, \sigma_{AB}^2)$$

$$\varepsilon \sim N(0, \sigma_e^2)$$

függetlenek!

$$i=1, \dots, r; \quad j=1, \dots, q; \quad k=1, \dots, p$$

A 4. példában  $r=3$ ,  $q=4$ ,  $p=2$

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{k(ij)}$$

(nap, személy, kölcsönhatás, hiba)

## Nullhipotézisek:

$$H_0^A : \sigma_A^2 = 0$$

$$H_0^B : \sigma_B^2 = 0$$

$$H_0^{AB} : \sigma_{AB}^2 = 0$$

**Szakmai kérdés:** Növelik az ingadozást? Mennyire?

$$\sigma_{ytotal}^2 = \sigma_A^2 + \sigma_B^2 + \sigma_{AB}^2 + \sigma_e^2$$

# ANOVA-táblázat két véletlen faktorra

az eltérés forrása	eltérés-négyzetösszeg	szabadsági fokszám	szórásnégyzet	szórásnégyzet várható értéke	F
A hatása	$S_A = qp \sum_i (y_{i..} - y_{...})^2$	$r-1$	$s_A^2 = \frac{S_A}{r-1}$	$qp\sigma_A^2 + p\sigma_{AB}^2 + \sigma_e^2$	$s_A^2 / s_{AB}^2$
B hatása	$S_B = rp \sum_j (y_{.j.} - y_{...})^2$	$q-1$	$s_B^2 = \frac{S_B}{q-1}$	$pr\sigma_B^2 + p\sigma_{AB}^2 + \sigma_e^2$	$s_B^2 / s_{AB}^2$
AB kölcssh.	$S_{AB} =$ $= p \sum_i \sum_j (y_{ij.} - y_{i..} - y_{.j.} + y_{...})^2$	$(r-1)(q-1)$	$s_{AB}^2 =$ $= \frac{S_{AB}}{(r-1)(q-1)}$	$p\sigma_{AB}^2 + \sigma_e^2$	$s_{AB}^2 / s_R^2$
Ismétlések	$S_R = \sum_i \sum_j \sum_k (y_{ijk} - y_{ij.})^2$	$rq(p-1)$	$s_R^2 = \frac{S_R}{rq(p-1)}$	$\sigma_e^2$	
Teljes	$S_0 = \sum_i \sum_j (y_{ij} - y_{..})^2$	$rqp-1$			

*Statistics > Advanced Linear/Nonlinear Models > General Linear Models > Factorial ANOVA*

*Options fülön: Random → Nap, Szem*

Univariate Tests of Significance for Y (Napszem in Napszem)  
 Over-parameterized model  
 Type III decomposition; Std. Error of Estimate: 0,1853

Effect	Effect (F/R)	SS	Degr. of Freedom	MS	Den.Syn. Error df	Den.Syn. Error MS	F	p
Intercept	Fixed	226746,0	1	226746,0	1,73047	0,189467	1196759	0,000005
NAP	Random	0,406	2	0,203	6,00000	0,024319	8,348	0,018476
SZEM	Random	0,032	3	0,011	6,00000	0,024319	0,443	0,730944
NAP*SZEM	Random	0,146	6	0,024	12,00000	0,034337	0,708	0,649658
Error		0,412	12	0,034				

Expected Mean Square Coefficients (Napszem)

Over-parameterized model

Type III decomposition

Effect	Effect (F/R)	Intercpt	NAP	SZEM	NAP*SZEM	Error
Intercept	Fixed	24.000000	8.000000	6.000000	2.000000	1.000000
NAP	Random		8.000000		2.000000	1.000000
SZEM	Random			6.000000	2.000000	1.000000
NAP*SZEM	Random				2.000000	1.000000
Error						1.000000

Az eltérés forrása	df	E(MS)	E(MS)
A: nap	2	$qp\sigma_A^2 + p\sigma_{AB}^2 + \sigma_e^2$	$8\sigma_A^2 + 2\sigma_{AB}^2 + \sigma_e^2$
B: személy	3	$pr\sigma_B^2 + p\sigma_{AB}^2 + \sigma_e^2$	$6\sigma_B^2 + 2\sigma_{AB}^2 + \sigma_e^2$
AB: kölcsönhatás	6	$p\sigma_{AB}^2 + \sigma_e^2$	$2\sigma_{AB}^2 + \sigma_e^2$
csoportokon belüli	12	$\sigma_e^2$	$\sigma_e^2$

# Varianciakomponens-elemzés

Az eltérés forrása	df	E(MS)	MS	MS
A: nap	2	$qp\sigma_A^2 + p\sigma_{AB}^2 + \sigma_e^2$	$s_A^2$	0,20317
B: személy	3	$pr\sigma_B^2 + p\sigma_{AB}^2 + \sigma_e^2$	$s_B^2$	0,01078
AB: kölcsönhatás	6	$p\sigma_{AB}^2 + \sigma_e^2$	$s_{AB}^2$	0,02431
csoportokon belüli	12	$\sigma_e^2$	$s_R^2$	0,03426

Components of Variance (Napszem)  
Over-parameterized model  
Type III decomposition

Effect	Y
NAP	0.0223
SZEM	-0.0023
NAP*SZEM	-0.0050
Error	0.0343

$$\hat{\sigma}_A^2 = \frac{s_A^2 - s_{AB}^2}{qp} = \frac{0.20317 - 0.02431}{8} = 0.0223$$

$$\hat{\sigma}_B^2 = \frac{s_B^2 - s_{AB}^2}{rp} = \frac{0.0108 - 0.02431}{6} = -0.0023$$

$$\hat{\sigma}_e^2 = s_R^2 = 0.0343$$

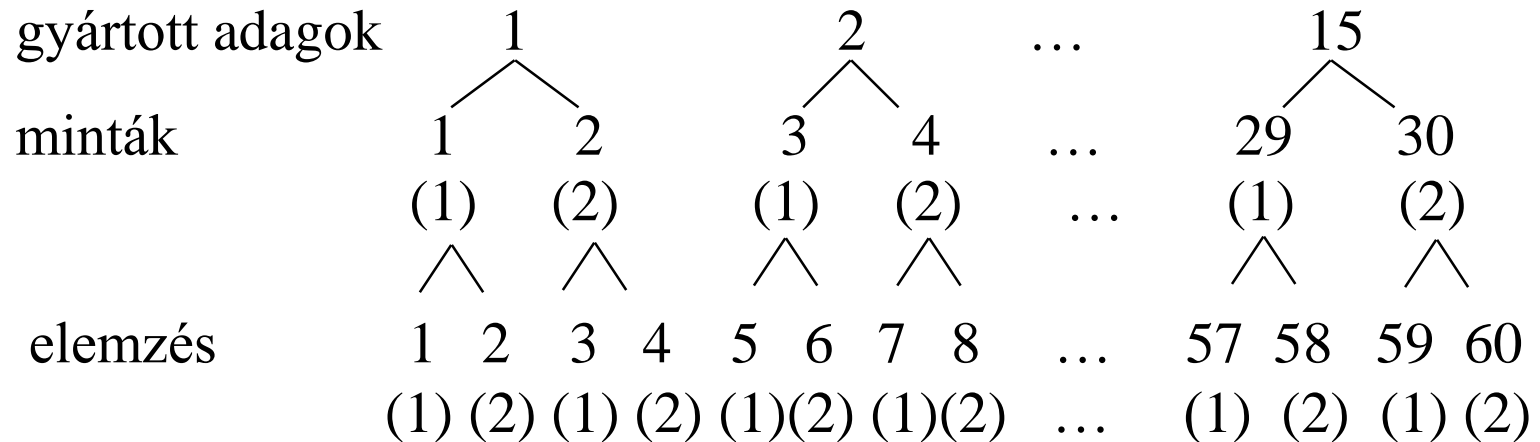
# Hierarchikus osztályozás

## 5. példa (Moisture.sta)

Festékgyári nedvesség-tartalom meghatározása:

15 gyártott adagból két-két mintát vesznek, mindkét minta víztartalmát kétszer-kétszer megméri.

(Box-Hunter-Hunter: Statistics for Experimenters, J. Wiley, 1978, p. 571)





## Az 5. példa adattáblázatának egy részlete:

adag	minta	elemzés		minta átlaga	adag átlaga
1	1	40.0	39.0	39.5	34.75
	2	30.0	30.0	30.0	
2	3	26.0	28.0	27.0	26.25
	4	25.0	26.0	25.5	
3	5	29.0	28.0	28.5	21.5
	6	14.0	15.0	14.5	
15	29	39.0	37.0	38.0	32.50
	30	26.0	28.0	27.0	

# Hierarchikus (beágyazott) modell

A modell: 
$$y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \varepsilon_{k(ij)}$$

adag                      minta                      analízis

A  $\beta$  faktor (minta) szintjei különböznek az  $\alpha$  faktor (adag) különböző szintjein,  
azaz a  $\beta$  faktor (minta) szintjei az  $\alpha$  faktor (adag) szintjeibe vannak ágyazva.

**A modellt a kísérlet elrendezése, a kísérleti terv szerkezete dönti el!**

## Az 5. példa megoldásának folytatása:

### *Milyen típusúak a faktorok?*

Ezt (ahogy tanultuk) a *szakmai kérdés alapján* döntjük el.

Itt: mindkét faktor (adag, minta) véletlen.

$$\alpha_i \sim N(0, \sigma_A^2) \quad \beta_j \sim N(0, \sigma_B^2) \quad \varepsilon_{k(ij)} \sim N(0, \sigma_e^2)$$

Ennek megfelelően a nullhipotézisek:

$$H_0 : \sigma_A^2 = 0 \quad H_0 : \sigma_{B(A)}^2 = 0$$

# Hierarchikus osztályozás ANOVA-táblázata

az eltérés forrása	eltérés-négyzetösszeg	szab. fok	szórásnégyzet	szórásnégyzet várható értéke	$F$
A hatása	$S_A = qp \sum_i (y_{i..} - y_{...})^2$	$r-1$	$s_A^2 = \frac{S_A}{r-1}$	$qp\sigma_A^2 + p\sigma_B^2 + \sigma_e^2$	$s_A^2 / s_{B(A)}^2$
B(A) hatása	$S_{B(A)} = p \sum_i \sum_j (y_{ij.} - y_{i..})^2$	$r(q-1)$	$s_{B(A)}^2 = \frac{S_{B(A)}}{r(q-1)}$	$p\sigma_B^2 + \sigma_e^2$	$s_{B(A)}^2 / s_R^2$
Ismétlések	$S_R = \sum_i \sum_j \sum_k (y_{ijk} - y_{ij.})^2$	$rq(p-1)$	$s_R^2 = \frac{S_R}{rq(p-1)}$	$\sigma_e^2$	

**Próbastatisztika:**

$$H_0^A : \sigma_A^2 = 0$$

$$s_A^2 / s_{B(A)}^2$$

$$H_0^B : \sigma_{B(A)}^2 = 0$$

$$s_{B(A)}^2 / s_R^2$$

**Variancia-komponens elemzés:**

$$\hat{\sigma}_A^2 = \frac{s_A^2 - s_{B(A)}^2}{qp}$$

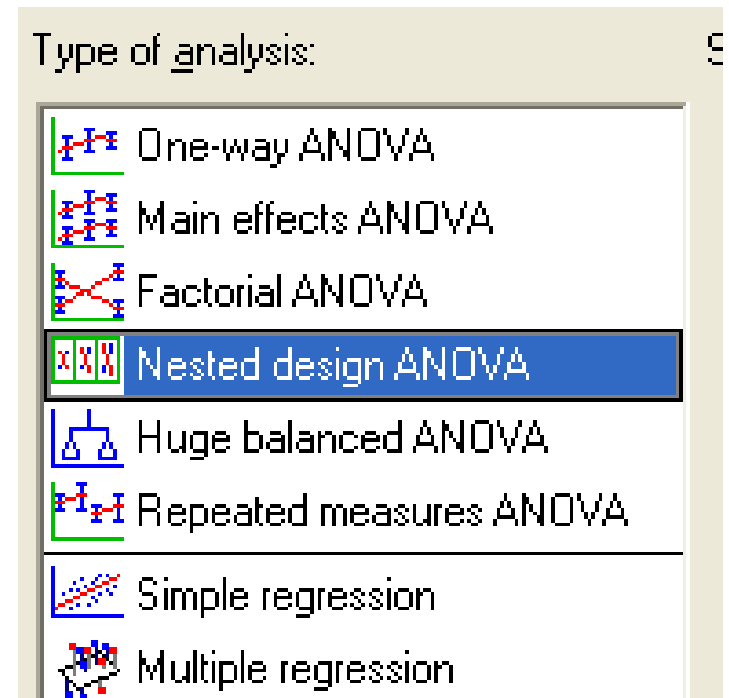
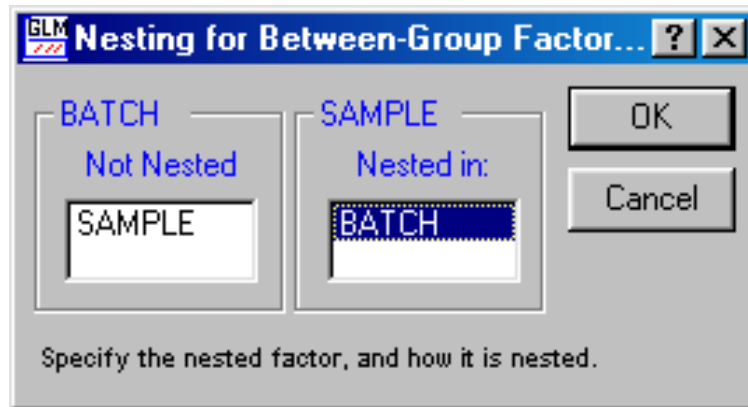
$$\hat{\sigma}_B^2 = \frac{s_{B(A)}^2 - s_R^2}{p}$$

## Az 5. példa megoldása STATISTICA szoftverben:

Statistics > Advanced Linear/Nonlinear Models > General Linear Models >  
> **Nested design ANOVA**

*Options* fülön: Random batch, sample

*Between effects*



## Az 5. példa megoldásának eredménytáblázatai

Univariate Tests of Significance for MOISTURE (Moisture)								
Over-parameterized model								
Type III decomposition								
Effect	Effect (F/R)	SS	Degr. of Freedom	MS	Den.Syn. Error df	Den.Syn. Error MS	F	p
Intercept	Fixed	43040.82	1	43040.82	14.0	86.495	497.61	0.0000
BATCH	Random	1210.93	14	86.50	15.0	57.983	1.49	0.2256
MSAMPLE(BATCH)	Random	869.75	15	57.98	30.0	0.917	63.25	0.0000
Error		27.50	30	0.92				

Components of Variance (Moisture)	
Over-parameterized model	
Type III decomposition	
Effect	MOISTURE
BATCH	7.13
MSAMPLE(BATCH)	28.53
Error	0.92

Értelmezzük a kapott eredményeket!

**Gondoljuk át**, hogy hogyan kellene módosítani a kísérleti elrendezést, hogy a **4. példa** (amikor ugyanazt az analitikai elemzést több személy több különböző napon végezte el) **hierarchikus modellel** kellene megoldani!

- a) A nap faktor legyen a személy faktorba ágyazva!
- b) A személy faktor legyen a nap faktorba ágyazva!