

4.3. ANÀLISI DE LA VARIÀNCIA

1. ANÀLISI DE LA VARIÀNCIA (ANOVA). DEFINICIÓ I OBJECTIU.

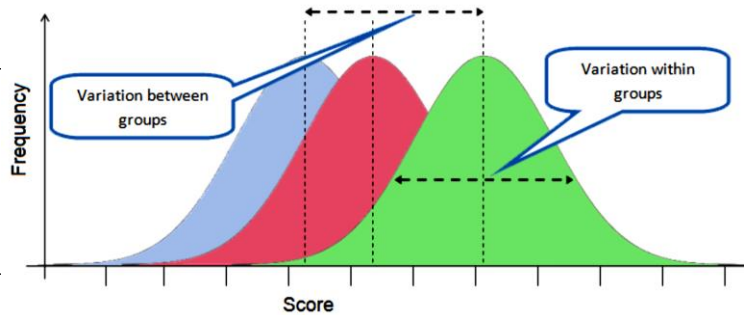
→ **ANOVA**: Eina estadística per a avaluar l'efecte d'un o més factors sobre la variació de dues o més sèries de resultats. És una comparació simultània de dos o més sets de dades a partir de la identificació i estimació de les fonts de variabilitat que afecten els resultats experimentals.

✓ Basat en la comparació de **valors mitjans**.

Fons de variabilitat:

- **Factor**: condició experimental (T, pH,...) o de procediment (laboratori, mètode, nivell de concentració...) que pot causar potencialment una variació en els resultats obtinguts.
 - ✓ L'efecte dels factors pot ser independent o pot existir interacció entre ells.
- **Error aleatori**

Un factor causa una variació significativa en els resultats si **la variància causada pel factor** és significativament més gran que la **variància associada a l'error aleatori** (residual).



Requeriments:

- La població de dades ha de tenir una distribució normal.
- Les variàncies de sèries comparades han de ser homogènies (Estadísticament iguals)

Concepte de variància: indica la dispersió de resultats respecte un valor mitjà

Limitacions: quan hi ha més d'una font de variació:

- **Ss i els dof sempre són additius.**

$$s^2 = \frac{\sum_i (x_i - \bar{x})^2}{n - 1} = \frac{ss}{d.o.f.}$$

2. ANOVA D'UN FACTOR AMB REPLICATS: els resultats obtinguts són significativament diferents?

- **Factor**: laboratori → **r** → **rn resultats**
- **Replicats**: d'una mostra → **n**

Càlculs:

1. Càlcul de la variància entre laboratoris (factor)
2. Càlcul de la variància entre replicats de cada laboratori (aleatòria).
3. Test F de comparació de variàncies

$$H_0 : s_{\text{factor}}^2 = s_{\text{aleatòria}}^2$$

$$H_1 : s_{\text{factor}}^2 > s_{\text{aleatòria}}^2$$

Laboratori	Replicats	
1	$x_{11} \ x_{12} \ x_{13} \ x_{14}$	M_1
2	$x_{21} \ x_{22} \ x_{23} \ x_{24}$	M_2
3	$x_{31} \ x_{32} \ x_{33} \ x_{34}$	M_3
4	$x_{41} \ x_{42} \ x_{43} \ x_{44}$	M_4
5	$x_{51} \ x_{52} \ x_{53} \ x_{54}$	M_5
		M_G

Nr. Laboratoris (r = 5)
Nr. replicats/laboratori (n = 4)

$$M_i = \frac{\sum_{k=1}^n x_{ik}}{n}$$

Mitjana dins un laboratori

$$M_G = \frac{\sum_{i=1}^r \sum_{k=1}^n x_{ik}}{n \cdot r}$$

$$M_G = \frac{\sum_{i=1}^r M_i}{r}$$

Mitjana global

Variància associada a la variabilitat aleatòria (dins de sèries): variància dels replicats, es calcula per a cada factor.

$$s_i^2 = \frac{\sum_{k=1}^n (x_{ik} - M_i)^2}{(n - 1)}$$

Variància dels replicats dins d'un laboratori

Variància aleatòria (variància conjunta de replicats^(*))

$$s_{\text{aleatòria}}^2 = \frac{\sum_{i=1}^r s_i^2}{r} \quad \rightarrow \quad s_{\text{aleat}}^2 = \frac{\sum_{i=1}^r \sum_{k=1}^n (x_{ik} - M_i)^2}{r(n - 1)}$$

$$ss_{\text{aleat}} = \sum_{i=1}^r \sum_{k=1}^n (x_{ik} - M_i)^2 \quad \text{Graus de llibertat: } r(n-1)$$

La **mitjana de variàncies** es calcula per tots els factors i estima la variabilitat aleatòria (*) s'assumeix que les variàncies de les sèries són comparables.

Variància associada al factor (variància entre sèries): informa de la variància de la mitjana de cada sèrie (Ex: variància de la mitjana dins de cada laboratori) respecte la mitjana global. Està afectat per l'error aleatori.

$$s_L^2 = \frac{\sum_{i=1}^r (M_i - M_G)^2}{r-1}$$

Variància associada a la diferència de mitjanes de laboratoris i mitjana global

En poblacions de mitjanes: $s_{mitjanes}^2 = \frac{s^2}{n}$

Variància associada al factor

$$s_L^2 = n \frac{\sum_{i=1}^r (M_i - M_G)^2}{r-1}$$

$ss_L = n \sum_{i=1}^r (M_i - M_G)^2$ **Graus de llibertat: r-1**

VARIÀNCIA GLOBAL:

Variàncies relacionades amb variabilitat aleatòria i de factor es comparen per tal d'estimar si hi ha una diferència significativa entre les mitjanes degut al factor. **TEST FISHER SEMPRE D'UNA CUA PERQUÈ SEMPRE SERÀ MÉS GRAN O IGUAL.**

$$s_O^2 = \frac{\sum_{i=1}^r \sum_{k=1}^n (x_{ik} - M_O)^2}{rn-1}$$

$$ss_O = \sum_{i=1}^r \sum_{k=1}^n (x_{ik} - M_O)^2 \quad \text{d.o.f.} = rn - 1$$

$$ss_O = ss_F + \text{SS}_{\text{global}} = \text{SS}_{\text{factor}} + \text{SS}_{\text{aleatoria}}$$

$$(x_{ik} - M_O = M_i - M_O + x_{ik} - M_i)$$

Factor residual Random residual

Variància global

$$s_G^2 = \frac{\sum_{i=1}^r \sum_{k=1}^n (x_{ik} - M_G)^2}{rn-1}$$

Avaluació de la significació del factor:

Font de variació	Suma de quadrats	Graus de llibertat	S ²	F _{calc}
Associada al factor (laboratoris)	$ss_L = n \sum_{i=1}^r (M_i - M_G)^2$	r-1	$s_L^2 = \frac{ss_L}{r-1}$	$\frac{s_L^2}{s_{aleat}^2}$
Error aleatori (replicats)	$ss_{aleat} = \sum_{i=1}^r \sum_{k=1}^n (x_{ik} - M_i)^2$	r(n-1)	$s_{aleat}^2 = \frac{ss_{aleat}}{r(n-1)}$	
Global	$ss_G = \sum_{i=1}^r \sum_{k=1}^n (x_{ik} - M_G)^2$	rn-1	$s_G^2 = \frac{ss_G}{rn-1}$	

Test F $H_0 : s_L^2 = s_{aleat}^2$
 $H_1 : s_L^2 > s_{aleat}^2$

$F_{calc} > F_{tab(\alpha, r-1, r(n-1))}$
 H_0 falsa \Rightarrow el factor és significatiu

3. ANOVA DE DOS FACTORS AMB REPLICATS

- Factor 1: r laboratoris
 - Factor 2: c mètodes
 - Replicats: n d'una mostra
- \rightarrow rcn resultats

Factor 1
 $H_0 : s_{fact1}^2 = s_{aleat}^2$
 $H_1 : s_{fact1}^2 > s_{aleat}^2$

Factor 2
 $H_0 : s_{fact2}^2 = s_{aleat}^2$
 $H_1 : s_{fact2}^2 > s_{aleat}^2$

Càlculs:

- Càlcul de la variància associada a cada factor
- Càlcul de la variància entre replicats de laboratori i mètode (aleatòria).
- Càlcul de la variància deguda a la interacció entre factors
- Tests F de comparació de variàncies

Interacció
 $H_0 : s_{interac}^2 = s_{aleat}^2$
 $H_1 : s_{interac}^2 > s_{aleat}^2$

	Met 1	Met j	Met c	
Lab 1	$x_{111} \dots x_{11n}$		$x_{1j1} \dots x_{1jn}$		$x_{1c1} \dots x_{1cn}$	M_1
⋮						⋮
Lab i	$x_{i11} \dots x_{i1n}$		$x_{ij1} \dots x_{ijn}$		$x_{ic1} \dots x_{icn}$	M_i
⋮						⋮
Lab r	$x_{r11} \dots x_{r1n}$		$x_{rj1} \dots x_{rjn}$		$x_{rc1} \dots x_{rcn}$	M_r
	M_1	M_j	M_c	M_G

Un lab té problemes amb un mètode. Efecte del factor depèn: **nivell** del factor + **interacció** entre factors

	Met 1	Met j	Met c	
Lab 1	$x_{111} \dots x_{11n}$		$x_{1j1} \dots x_{1jn}$		$x_{1c1} \dots x_{1cn}$	M_1
\vdots						\vdots
Lab i	$x_{i11} \dots x_{i1n}$		$x_{ij1} \dots x_{ijn}$		$x_{ic1} \dots x_{icn}$	M_i
\vdots						\vdots
Lab r	$x_{r11} \dots x_{r1n}$		$x_{rj1} \dots x_{rjn}$		$x_{rc1} \dots x_{rcn}$	M_r
	M_1	M_j	M_c	M_G

$$S_G^2 = \frac{\sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^n (x_{ijk} - M_G)^2}{rcn - 1}$$

Variància global

$$S_M^2 = \frac{\sum_{j=1}^c (M_{Mj} - M_G)^2}{c - 1}$$

Variància mitjanes

$$S_{aleat}^2 = \frac{\sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^n (x_{ijk} - M_{ij})^2}{rc(n - 1)}$$

Variància aleatòria

$$S_L^2 = \frac{\sum_{i=1}^r (M_{Li} - M_G)^2}{r - 1}$$

Variància mitjanes

$$S_M^2 = nr \frac{\sum_{j=1}^c (M_{Mj} - M_G)^2}{c - 1}$$

Variància factor 2 (mètode)

$$S_L^2 = nc \frac{\sum_{i=1}^r (M_{Li} - M_G)^2}{r - 1}$$

Variància factor 1 (laboratori)

Concepte d'interacció

Variació en l'efecte d'un factor en funció del nivell d'un altre (per exemple: les diferències entre laboratoris varien en funció del mètode que es compara)

En absència d'interacció:

$$SS_{global} = SS_{factor1} + SS_{factor2} + SS_{aleatoria}$$

En presència d'interacció:

$$SS_{global} = SS_{factor1} + SS_{factor2} + SS_{interacció} + SS_{aleatoria}$$

$$Dof = (r-1)(c-1)$$

La variància d'interacció i els graus de llibertat associats es calculen per diferència, un cop quantificades les altres causes de variació.

Font de variació	Suma de quadrats	Graus de llibertat	S^2	F_{calc}
Entre laboratoris (factor 1)	$SS_L = nc \sum_{i=1}^r (M_{Li} - M_G)^2$	$r-1$	$S_L^2 = \frac{SS_L}{r-1}$	$\frac{S_L^2}{S_{aleat}^2}$
Entre mètodes (factor 2)	$SS_M = nr \sum_{j=1}^c (M_{Mj} - M_G)^2$	$c-1$	$S_M^2 = \frac{SS_M}{c-1}$	$\frac{S_M^2}{S_{aleat}^2}$
Error aleatori	$SS_{aleat} = \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^n (x_{ijk} - M_{ij})^2$	$rc(n-1)$	$S_{aleat}^2 = \frac{SS_{aleat}}{rc(n-1)}$	
Interacció	$SS_{int} = SS_G - SS_L - SS_M - SS_{aleat}$	$(r-1)(c-1)$	$S_{int}^2 = \frac{SS_{int}}{(r-1)(c-1)}$	$\frac{S_{int}^2}{S_{aleat}^2}$
Global	$SS_G = \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^n (x_{ijk} - M_G)^2$	$rcn - 1$	$S_G^2 = \frac{SS_G}{rcn - 1}$	

$$H_0: S_{fact1}^2 = S_{aleatòria}^2$$

$$H_1: S_{fact1}^2 > S_{aleatòria}^2$$

Factor 1

$$H_0: S_{fact2}^2 = S_{aleatòria}^2$$

$$H_1: S_{fact2}^2 > S_{aleatòria}^2$$

Factor 2

$$H_0: S_{interac}^2 = S_{aleatòria}^2$$

$$H_1: S_{interac}^2 > S_{aleatòria}^2$$

Interacció

$$F_{calc} > F_{tab}$$

$$H_0 \text{ falsa} \Rightarrow$$

factor o interacció significatius

4. ANOVA DE DOS FACTORS SENSE REPLICATS

- Factor 1: r laboratoris
 - Factor 2: c mètodes
- rcn resultats

Sense replicats, no es pot calcular directament l'error aleatori i no es pot conèixer si hi ha interacció entre factors.

Càlculs:

- Càlcul de la variància associada a cada factor
- Càlcul de la variància residual (deguda a l'error aleatori i interacció).
- Tests F de comparació de variàncies

$$H_0: S_{fact1}^2 = S_{aleatòria}^2$$

$$H_1: S_{fact1}^2 > S_{aleatòria}^2$$

Factor 1

$$H_0: S_{fact2}^2 = S_{aleatòria}^2$$

$$H_1: S_{fact2}^2 > S_{aleatòria}^2$$

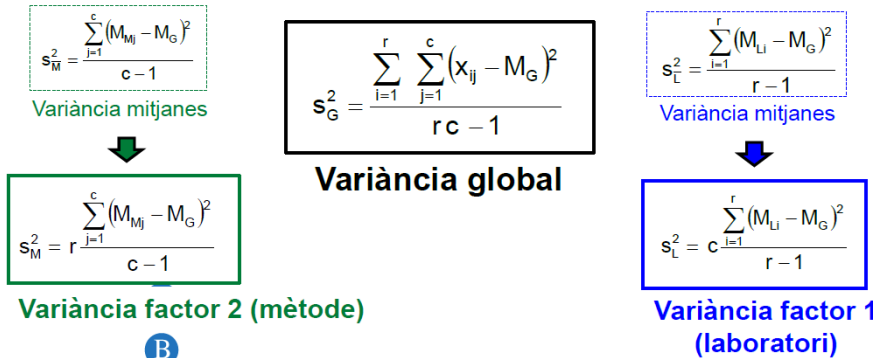
Factor 2

	Met 1	Met j	Met c	
Lab 1	X ₁₁		X _{1j}		X _{1c}	M ₁
⋮						⋮
Lab i	X _{i1}		X _{ij}		X _{ic}	M _i
⋮						⋮
Lab r	X _{r1}		X _{rj}		X _{rc}	M _r
	M ₁	M _j	M _c	M _G

Variància residual: és la variància aleatòria en absència de la interacció.

La variància residual i els graus de llibertat associats es calculen per diferència, un cop quantificades les altres causes de variació.

Conté la contribució de l'error aleatori i de la interacció si n'hi ha.



$$SS_{res} = SS_{global} - SS_{factor1} - SS_{factor2}$$

$$DOF_{res} = (r-1)(c-1)$$

Font de variació	Suma de quadrats	Graus de llibertat	S ²	F
Entre laboratoris	$SS_L = c \sum_{i=1}^r (M_{Li} - M_G)^2$	r-1	$S_L^2 = \frac{SS_L}{r-1}$	$\frac{S_L^2}{S_{res}^2}$
Entre mètodes	$SS_M = r \sum_{j=1}^c (M_{Mj} - M_G)^2$	c-1	$S_M^2 = \frac{SS_M}{c-1}$	$\frac{S_M^2}{S_{res}^2}$
Error residual	$SS_{res} = SS_G - SS_L - SS_M$	(r-1)(c-1)	$S_{res}^2 = \frac{SS_{res}}{(r-1)(c-1)}$	
Global	$SS_G = \sum_{i=1}^r \sum_{j=1}^c (x_{ij} - M_G)^2$	rc - 1	$S_G^2 = \frac{SS_G}{rc-1}$	

$$H_0 : S_{fact1}^2 = S_{residual}^2$$

$$H_1 : S_{fact1}^2 > S_{residual}^2$$

Factor 1

$$H_0 : S_{fact2}^2 = S_{residual}^2$$

$$H_1 : S_{fact2}^2 > S_{residual}^2$$

Factor 2

$$F_{calc} > F_{tab}$$

$$H_0 \text{ falsa} \Rightarrow$$

Factor significatiu

APLICACIONS PRÀCTIQUES PRINCIPALS DE ANOVA

- Descriure la **variància global** d'un procés mitjançant la estimació quantitativa de les fonts que la componen (per ex., **experiments de replicació**).
- Estudiar l'**efecte** d'un o més **factors** sobre el resultat d'una anàlisi.
- Determinar si dos factors són independents o si interaccionen.
- Comparar els resultats de diferents mètodes, laboratoris.
- Altres:
 - Optimitzar mètodes analítics mitjançant experiments dissenyats amb la variació adequada de factors i nivells.
 - Validació de models de regressió/calibratge.