4. Análisis de Varianza

4.1. ANOVA de una vía

COMPARACION DE MAS DE DOS MEDIAS

- **Cuando se comparan dos medias a nivel de significación α, la probabilidad de cometer un error de tipo I es α**
- **Cuando se comparan dos a dos, a medias, la esperanza de rechazos infundados de Ho es :** $\boxed{\alpha \times C_2^a}$

Por ej. : Comparando las medias de 7 grupos a nivel 0.05 la esperanza de decisiones equivocadas es:

$$0.05 \times C_2^7 = 0.05 \times 21 = 1.05$$

Una solución para este problema es la CORRECCION DE BONFERRONI :

$$\boxed{\alpha' = \frac{\alpha}{C_2^a}}$$

Suele ser excesivamente severa

En el ejemplo:

$$\boxed{\alpha' = \frac{0.05}{21} = 0.002}$$

HAY OTRAS ALTERNATIVAS, UNA DE ELLAS ES EL ANALISIS DE LA VARIANZA

ANOVA (ANalysis Of Variance)

Finalidad

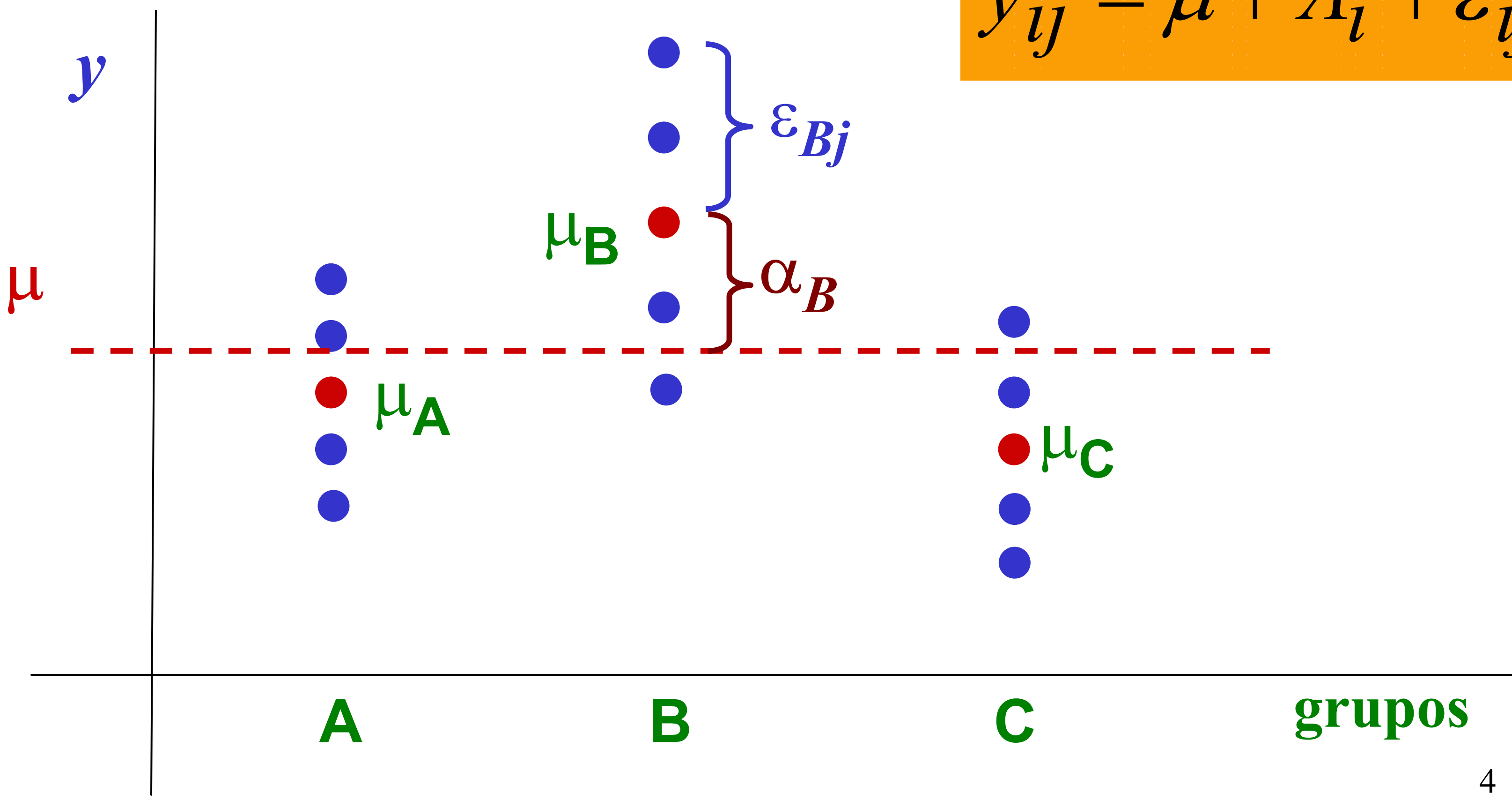
Comparar simultáneamente varias medias

Modelo I – efectos fijos

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

Modelo II – efectos aleatorios

$$y_{ij} = \mu + A_i + \varepsilon_{ij}$$



$$y_{ij} - \mu = \underbrace{(\mu_i - \mu)}_{\alpha_i} + \underbrace{(y_{ij} - \mu_i)}_{\varepsilon_{ij}}$$

En la población

$$y_{ij} - \bar{\bar{y}} = (\bar{y}_i - \bar{\bar{y}}) + (y_{ij} - \bar{y}_i)$$

En la muestra

Elevando al cuadrado:

$$(y_{ij} - \bar{\bar{y}})^2 = (\bar{y}_i - \bar{\bar{y}})^2 + (y_{ij} - \bar{y}_i)^2 + 2(\bar{y}_i - \bar{\bar{y}})(y_{ij} - \bar{y}_i)$$

Sumando:

$$\boxed{\sum_{ij} (y_{ij} - \bar{\bar{y}})^2 = \sum_{ij} (\bar{y}_i - \bar{\bar{y}})^2 + \sum_{ij} (y_{ij} - \bar{y}_i)^2}$$

SC TOTAL

SC ENTRE grupos

SC DENTRO de grupos (residual)

HIPOTESIS

Modelo I

$$H_0 : \forall_i : \alpha_i = 0$$

Modelo I I

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

En general

$$H_0 : \forall_i : \mu_i = \mu$$

MEDIAS DE CUADRADOS		ESTIMA
MC entre = SC entre/(a-1) **a = nº de grupos**	MI	$\sigma^2 + \bar{n}_i \sum \alpha_i^2 / (a-1)$
$\bar{n}_i =$ **tamaño medio del grupo**	MII	$\sigma^2 + \bar{n}_i \sigma_A^2$
MC dentro = SC dentro/(n-a) **n = tamaño de la muestra total**		σ^2

Si Ho es verdadera : MC entre = MC dentro en la población

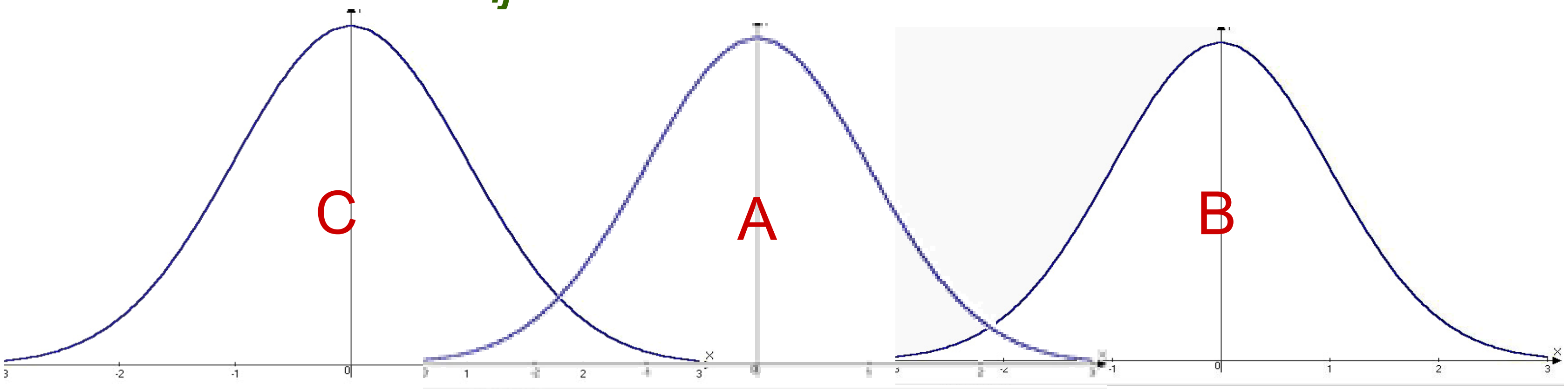
TEST DE HIPOTESIS

$$\boxed{F_{calc} = \text{MC entre/ MC dentro}}$$

se compara con F_{tab} (a-1) y (n-a) grados de libertad

Supuestos para la validez del test

- **Normalidad de los residuos (ε_{ij})**
- **Homocedasticidad de los residuos**



- **Independencia de las observaciones**

$$SC\ \ entre = \sum_i \frac{T_i^2}{n_i} - T^2 \Big/ n$$

$$SC\ \ total = \sum_{ij} y_{ij}^2 - T^2 \Big/ n$$

$$SC\ \ dentro = SC\ \ total - \ SC\ \ entre$$

Donde: $T_i = \sum_j y_{ij}$ **En el i-ésimo grupo**

$n_i =$ **Tamaño del i-ésimo grupo**

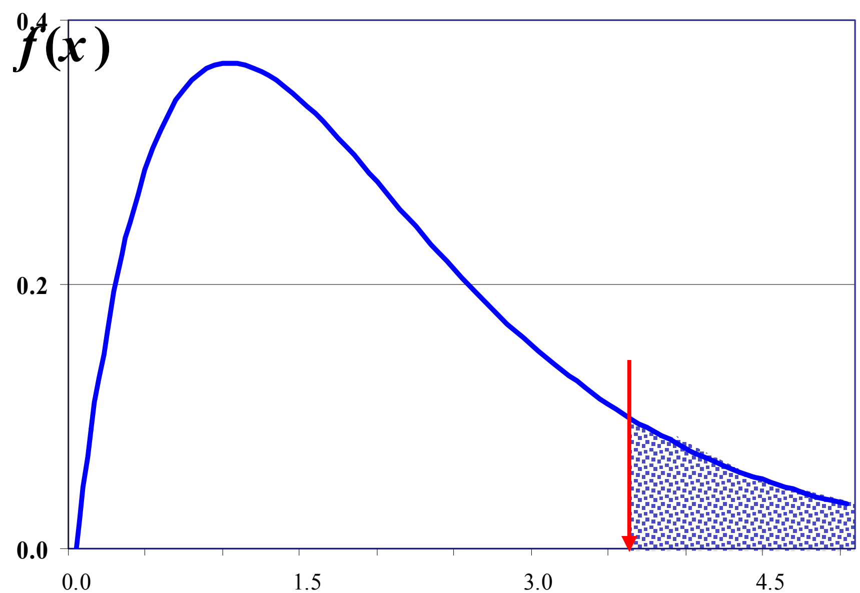
$T = \sum_{ij} y_{ij}$ **Gran total**

$n = \sum_i n_i$ **Tamaño total de la muestra**

FUENTE DE VARIACION	SUMA DE CUADRADOS	GL	MEDIA DE CUADRADOS	Fcalc
ENTRE GRUPOS	SC entre	a-1	$SC\ entre / (a-1)$	$\dfrac{MC\ entre}{MC\ dentro}$
DENTRO DE GRUPOS	SC dentro	n-a	$SC\ dentro / (n-a)$	
TOTAL	SC total	n-1		

El $F_{calculado}$ se compara con

el $F_{tabulado}$ con (a-1) y (n-a) GL



CALCULO DE LAS SUMAS DE CUADRADOS

	A	B	C	D	
	4.4	8.6	3.4	8.9	
	5.9	4.5	7.3	0.0	
	6.2	8.4	8.8	1.7	
	6.3	8.7	0.2		
			0.1		
Ti	22.8	30.2	19.8	10.6	T =83.4
ni	4	4	5	3	n = 16

$$H_0 : \forall_i : \mu_i = \mu$$

$$\sum_{ij} y_{ij}^2 = 597.2$$

$$n = \sum n_i = 16$$

$$a = 4$$

$$SC \quad total = 597.2 - 83.4^2 / 16 = 162.4775$$

$$SC \quad entre = \frac{22.8^2}{4} + \frac{30.2^2}{4} + \frac{19.8^2}{5} + \frac{10.6^2}{3} - \frac{83.4^2}{16} = 39.1088$$

$$SC \quad dentro = SC \quad total - SC \quad entre = 162.4775 - 39.1088 = 123.3687$$

FUENTE DE VARIACION	SUMA DE CUADRADOS	GL	MEDIA DE CUADRADOS	Fcalc
ENTRE GRUPOS	39.1088	3	13.036	12.27
DENTRO DE GRUPOS	123.3687	12	10.281	
TOTAL	162.4775	15		

$F_{0.95}(3, 12) = 3.49$

F_{calc} mayor que F_{tab} $\Longrightarrow$ Se rechaza Ho

$\Longrightarrow$ las medias difieren entre sí

ANOVA – Cálculos después del test de F

Modelo I – efectos fijos

Comparaciones previstas "a priori"

El ANOVA aporta una mejor estimación de $\sigma^2 : \hat{s}^2_{res} = MC_{dentro}$

$$t_c = \frac{\bar{x}_1 - \bar{x}_2}{\hat{s}_{res}\sqrt{1/n_1 + 1/n_2}}$$

provee un test válido a nivel α con (n - a) grados de libertad

Comparaciones previstas "a posteriori"

- t_c igual que el anterior si el número de grupos es menor que 4
- En otros casos, usar la corrección de Bonferroni u otros tests (Scheffé, Duncan, Tukey, etc).
- Ninguno de ellos es sistemáticamente mejor

ejemplo Comparación de los grupos **A** y **C**

$$Ho\!:\mu_A = \mu_C$$

$$n_A = 4 \qquad \bar{x}_A = 5.7 \qquad \hat{s}^2_{res} = MCdentro = 10.281$$

$$n_C = 5 \qquad \bar{x}_C = 3.96 \qquad \hat{s} = \sqrt{10.281} = 3.21$$

$$\boldsymbol{t}_c = \frac{5.7 - 3.96}{3.21\sqrt{1/4 + 1/5}} = 0.81$$

$t_{0.975;12}$ **= 2.18** $\Rightarrow$ $|t_c| < |t_t|$ $\Rightarrow$ **NO rechazar Ho**

ANOVA – Cálculos después del test de F

Modelo II – efectos aleatorios

Asociación entre la variable dependiente y los niveles del factor

$$\eta^2 = \frac{SCentre}{SCtotal}$$

η : *razón de correlación*

ejemplo

- SC_{entre} = 39.1088
- SC_{total} = 162.4775

$$\eta^2 = \frac{39.1088}{162.4775} = 0.241$$

COMPARACION DE VARIANZAS, VARIOS GRUPOS

Test de Bartlett

$$H_0 : \sigma_i^2 = \sigma_k^2 \quad \forall (i, j)$$

$$\hat{s}^2 = MC \ \ dentro$$

$$M = (n-1)\ln \hat{s}^2 - \sum_i (n_i - 1)\ln \hat{s}_i^2$$

$$C = 1 + \frac{1}{3(a-1)}\left[\sum_i \frac{1}{n_i - 1} - \frac{1}{n-a}\right]$$

$$\chi^2_{calc} = M / C; \quad \chi^2_{tab} = \chi^2_{a\text{-}1}$$

$$Si \ \ \chi^2_{calc} \geq \chi^2_{tab}$$

Se Rechaza Ho

El Test de Bartlett es sensible a la Kurtosis. Existen otros test que no son sistemáticamente mejores ni peores

COMPARACION DE VARIANZAS, VARIOS GRUPOS

ejemplo

$$H_0 : \sigma_i^2 = \sigma_k^2 \quad \forall (i, j)$$

$\hat{s}_A^2 = 0.78$ $\hat{s}_B^2 = 4.15$ $\hat{s}_C^2 = 15.983$ $\hat{s}_D^2 = 22.323$ $\hat{s}_{res}^2 = 10.281$

$n_A = 4$ $n_B = 4$ $n_C = 5$ $n_D = 3$ $n = 16$

$$M = (16-1)\ln 10.281 - [(4-1)\ln 0.78 + (4-1)\ln 4.1 +$$

$$+(5-1)\ln 15.983 + (3-1)\ln 22.323 = 14.17$$

$$C = 1 + \frac{1}{3(4-1)}\left(\frac{1}{4-1} + \frac{1}{4-1} + \frac{1}{5-1} + \frac{1}{3-1} - \frac{1}{16-4}\right) = 1.148$$

$$\chi^2_{calc} = 14.170/1.148 = 12.342;$$

$\chi^2_{0.95;3} = 7.81$ $\chi^2_{calc} \geq \chi^2_{tab}$ Se Rechaza Ho

SOLUCIONES A LA VIOLACION DE LOS SUPUESTOS DEL ANOVA

- Si los supuestos del ANOVA no se cumplen, la descomposición de la varianza sigue siendo válida. No así el test de F, por lo que no pueden obtenerse conclusiones confiables respecto a la Ho

- Existen 3 grandes grupos de soluciones

a) ANOVA ponderado	b) Transformar la variable	c) Test de Kruskal-Wallis

- La solucion a se verá en BEII

SOLUCIONES A LA VIOLACION DE LOS SUPUESTOS DEL ANOVA

- La solución b, para algunos casos particulares puede ser:

- recuentos números pequeños

$$x' = \sqrt{x} \quad o \quad bien \quad x' = \sqrt{x + k} \quad ; (k = 0.5;1)$$

- efectos multiplicativos

$$x' = \log x \,,\ o \quad bien \quad log(x + k)$$

$$(k > |\text{-}x_{min}|)$$

- Proporciones (p)

$$x' = arcsen \sqrt{p}$$

- Si la transformación es en grados sexagesimales

$$\sigma^2_{x'} = \frac{820.8}{n}$$

- Es innecesario si $0.3 \leq p \leq 0.7$

- Método de BOX-COX : será visto después de regresión

PRUEBA DE KRUSKAL-WALLIS

- Es el equivalente al ANOVA, con variable ordinal
- Es una generalización del test de Mann-Whitney para más de dos grupos independientes

- Asignar rangos a todas las observaciones como si fuera un solo grupo

- Sumar los rangos de cada grupo (Ri)

- Calcular

$$\boxed{k_c = \frac{12}{n(n+1)} \sum \frac{R_i^2}{n_i} - 3(n+1)}$$

- Si el número de grupos es a = 3, existen tablas especiales

Si a $>$3, comparar k_c con $\chi^2_{1-\alpha;a-1}$

Si $\chi^2_c > \chi^2_t$ rechazar Ho

ejemplo

valores originales

						n_i
A	4.4	5.9	6.2	6.3		4
B	8.6	4.5	8.4	8.7		4
C	3.4	7.3	8.8	0.2	0.1	5
D	8.9	0.0	1.7			3

$$n = \sum n_i = 16$$

$$a = 4$$

rangos

						R_i
A	6	8	9	10		4
B	13	7	12	14		4
C	5	11	15	3	2	5
D	16	1	4			3

$$k_c = \frac{12}{16(16+1)}\left[\frac{33^2}{4} + \frac{46^2}{4} + \frac{36^2}{5} + \frac{21^2}{3} + \right] - 3(16+1) = 2.27$$

$\chi^2_{1-\alpha;a-1} = \chi^2_{0.95;3} = 7.81$ $\chi^2_c < \chi^2_t$ NO rechazar Ho