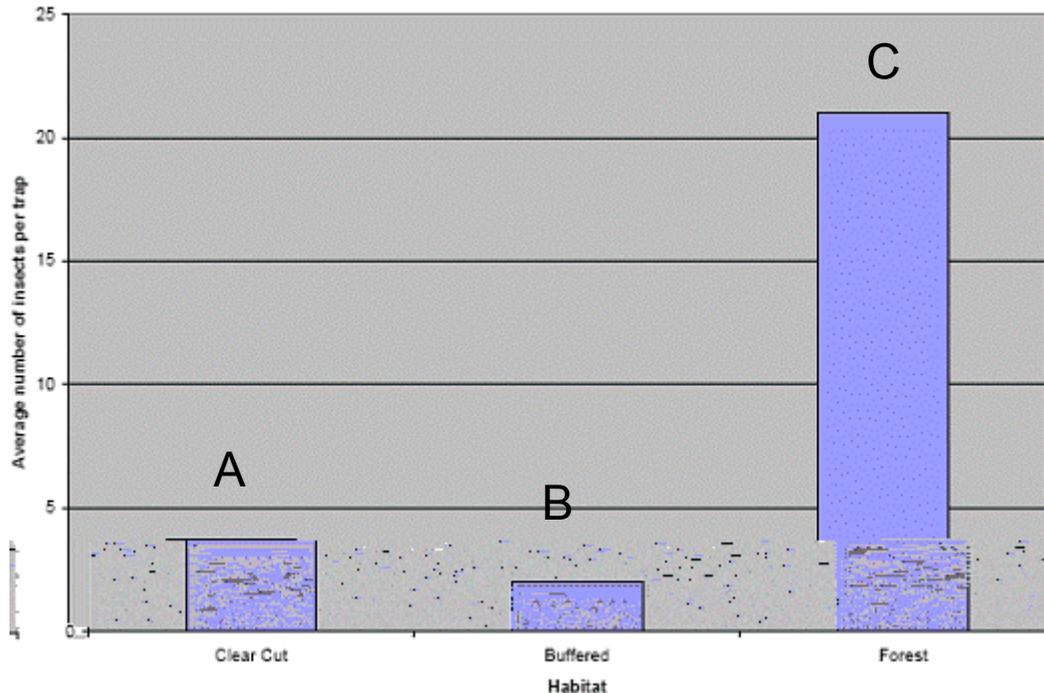




4.1. La estrategia del Análisis de varianza:

- Los test t múltiples (múltiples tratamientos); corrección a posteriori
- La mejora del ANOVA: necesidad de análisis a posteriori



Test t

A versus B
A versus C
B versus C

Si A, B y C son muestras de tamaño n de la misma población, entonces:

$$V_A = V_B = V_C = V_w \text{ (estima } V_p)$$

$$V_{ABC} = V_b \text{ (estima } 1/n V_p)$$

Distribución F



Entonces debería ocurrir que:

$$\frac{(n \times V_b)}{V_w} = 1, \text{ que se distribuye como una } F \text{ con } n_1 \text{ y } n_2 \text{ gl}$$

Tema IV. EL ANOVA de un factor

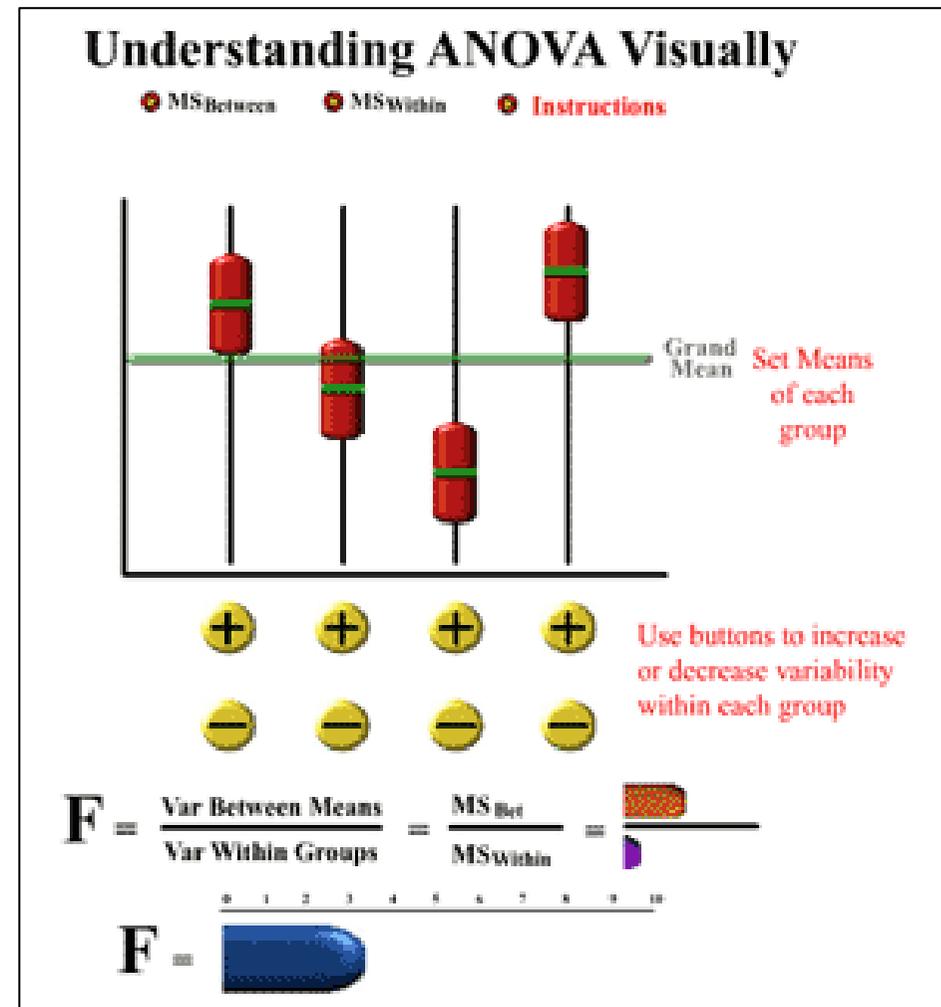
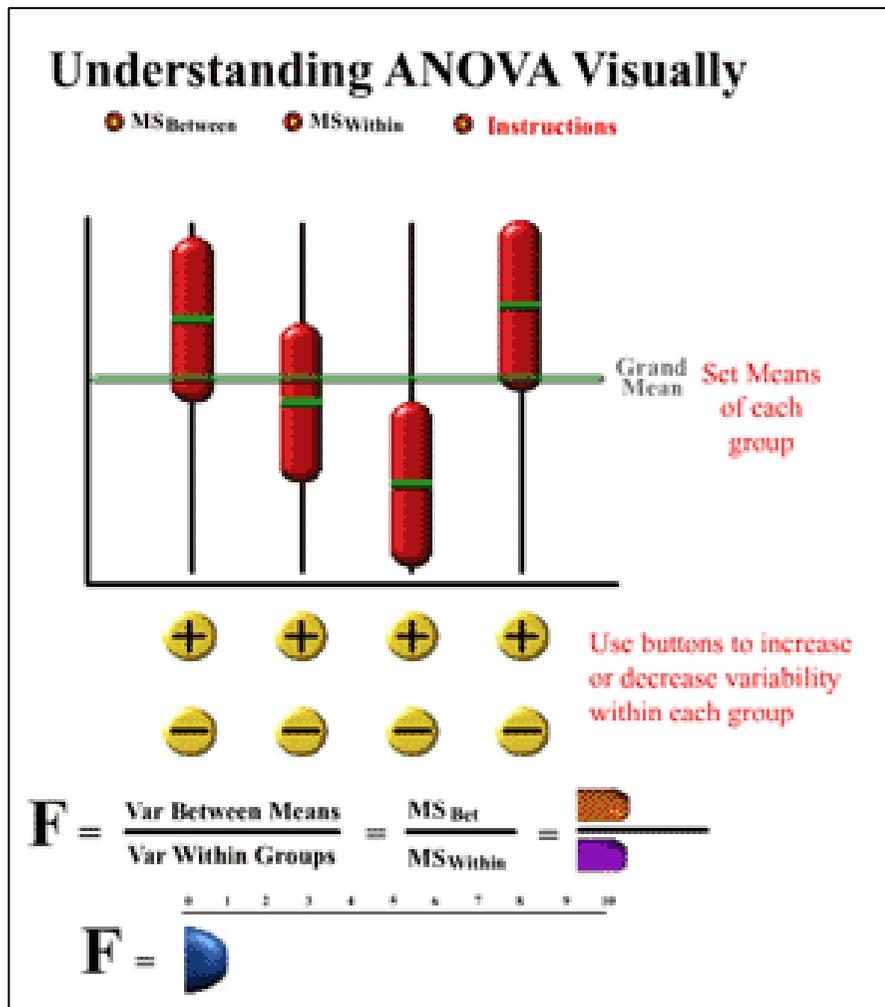
4.1. La estrategia del Análisis de varianza:

- Tenemos i grupos (tratamientos) y j valores dentro de grupo
- La variación total se descompone en dos partes:
 - . La varianza entre grupos (V_B)
 - . La varianza dentro de grupos (V_W)



4.1. La estrategia del Análisis de varianza:

- La interpretación del ANOVA (magnitud relativa de las diferencias entre medias)





Tema IV. EL ANOVA de un factor

4.2. Los dos modelos de ANOVA:

- Tratamientos Fijos (modelo I):

- . Estudiamos todos los tratamientos del factor (se evalúa su efecto diferencial)
- . Ejemplo: uso del teléfono de 10 individuos (doctores y enfermer@s)
- . Los sexos, las estaciones, etc...

Enfermer@s



Doctores



Enric Maguel Serrano 2010



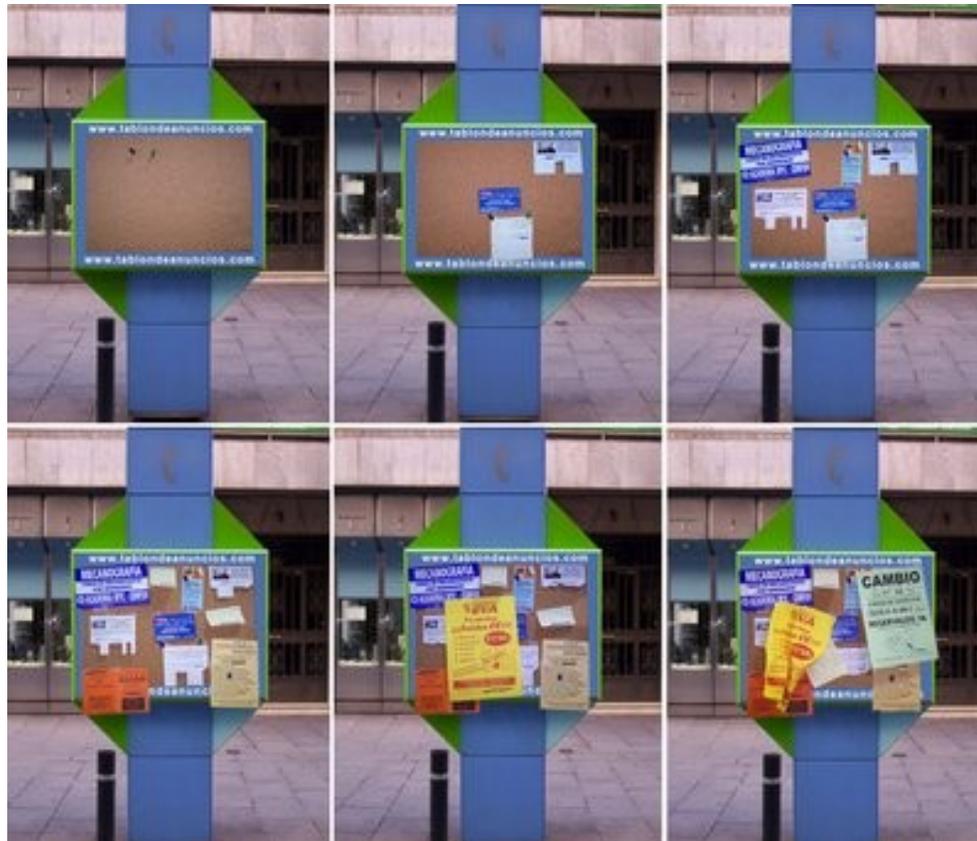
Tema IV. EL ANOVA de un factor

4.2. Los dos modelos de ANOVA:

- Tratamientos aleatorios (modelo II):

- . Solo disponemos de una muestra (aleatoria) de los tratamientos
- . Uso del teléfono a distintas horas: 9-10; 13-14; 16-17
- . Los tratamientos pueden ser temperaturas, alturas (variables continuas)...

Uso del tablón de una cabina a lo largo de una semana





Tema IV. EL ANOVA de un factor

4.3. Los cálculos del modelo I de ANOVA:

- En la práctica para un factor es igual modelo I y II (equilibrado)
- Los cálculos:

G_1	G_2	G_i
valor ₁₁
valor ₁₂
...		
valor _{1j}	...	valor _{ij}
X_1		X_i
		X_{ii}

$$X_1 = \sum x_{1j},$$

$$X_2 = \sum x_{2j}$$

y así sucesivamente

$$X_{ii} = \sum X_i$$

Fuente	SC (SS)	GL(DF)	CM(MS)	F
Entre (b)	$\frac{\sum X_i^2}{i} - \frac{X_{ii}^2}{ij}$	$i - 1$	$\frac{SC_b}{(i - 1)}$	$\frac{CM_b}{CM_w}$
Dentro (w)	$\frac{\sum x_{ij}^2}{i} - \frac{\sum X_i^2}{ij}$	$i(j-1)$	$\frac{SC_w}{i(j-1)}$	



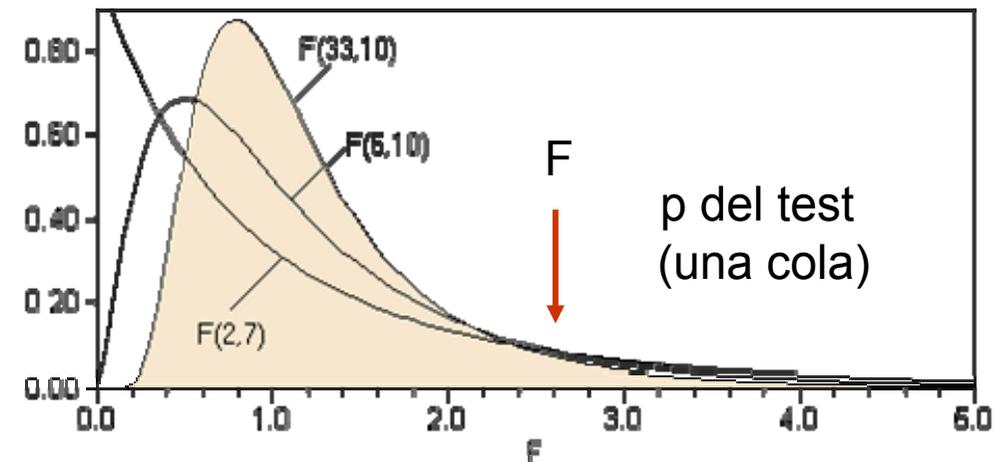
Tema IV. EL ANOVA de un factor

4.3. Los cálculos del modelo I de ANOVA:

- Ejemplo numérico: (evaluación de eficiencia de marcas de calmantes)
 - . 3 marcas se prueban en 5 pacientes (con similar dolencia)
 - . Se evalúa el tiempo necesario hasta la siguiente dosis
 - . Toma de decisión (equivalente a un test t de dos muestras)

$$H_0 = A = B = C$$

	A	B	C
	120	130	79
	110	105	99
	140	160	100
	135	123	88
	129	140	99
\bar{X}	126,8	131,6	93



PASOS

- . se calcula el valor de F
- . se evalúa probabilidad en tablas
- . se rechaza o no la H_0



Tema IV. EL ANOVA de un factor

4.3. Los cálculos del modelo I de ANOVA:

- Ejemplo numérico: (evaluación de eficiencia de marcas de calmantes)
 - . Tres marcas: A, B y C
 - . 5 pacientes (similar dolencia) probados por marca
 - . Tiempo de espera hasta reclamo de nueva dosis

	A	B	C	
	120	130	79	
	110	105	99	
	140	160	100	
	135	123	88	
	129	140	99	
X	634	658	465	1757

Fuente	SC (SS)	GL(DF)	CM(MS)	F
Entre (b)	$\frac{\sum X_i^2}{i} - \frac{X_{ii}^2}{ij}$	$i - 1$	$\frac{SC_b}{(i - 1)}$	$\frac{CM_b}{CM_w}$
Dentro (w)	$\frac{\sum x_{ij}^2}{i} - \frac{\sum X_i^2}{ij}$	$i(j-1)$	$\frac{SC_w}{i(j-1)}$	

$$\begin{aligned} \sum x_{ij}^2 &= 120^2 + 130^2 + 79^2 + \dots = 212807 \\ X_i^2 &= 634^2 + 658^2 + 465^2 = 1051145 \\ X_{ii} &= 1757^2 = 3087049 \end{aligned}$$

Fuente	SC	GL	CM	$F_{(2,12)}$
Entre (b)	$\frac{1051145}{5} - \frac{3087049}{15} = 4425,7$	2	2212,9	10,3
Dentro (w)	$\frac{212807}{5} - \frac{1051145}{15} = 2578$	12	214,8	
TOTAL	$= 7003,7$	14	500,3	

$p(F(2,12) = 10,3) < 0,01$: se rechaza H_0



Tema IV. EL ANOVA de un factor

4.4. La estimación de los componentes de varianza:

- Cuando el ANOVA es significativo:
 - . Comparación cuantitativa entre factores (η^2)
 - . Estimación del incremento de varianza causado por el factor (σ^2_{FACTOR})
- Depende del modelo del ANOVA:
 - . Modelo I (no veremos demostración)
 - . Modelo II (no veremos demostración)

A	B	C
120	130	79
110	105	99
140	160	100
135	123	88
129	140	99

Fuente	SC (SS)	GL(DF)	CM(MS)	F
Entre (b)	$\frac{\sum X_i^2}{i} - \frac{\sum x_{ij}^2}{ij}$	$i - 1$	$\frac{SC_b}{(i - 1)}$	$\frac{CM_b}{CM_w}$
Dentro (w)	$\frac{\sum x_{ij}^2}{i} - \frac{\sum X_i^2}{i}$	$i(j-1)$	$\frac{SC_w}{i(j-1)}$	

Modelo I	
Componentes	
Entre	$\sigma^2_w + \frac{j}{(i-1)} \times \eta^2$
Dentro	σ^2_w

Modelo II	
Componentes	
Entre	$\sigma^2_w + j\sigma^2_b$
Dentro	σ^2_w

Fuente	SC	GL	CM	F _(2,12)
Entre (b)	4425,7	2	2212,9	10,3
Dentro (w)	2578	12	214,8	
TOTAL	7003,7	14		

Estimación de η^2	
$\sigma^2_w = CM_w$	= 214,8
$\eta^2 = \frac{(CM_b - CM_w) \times (i-1)}{j}$	= 799,2

La deducción de los componentes de varianza es clave para determinar el test estadístico F



Tema IV. EL ANOVA de un factor

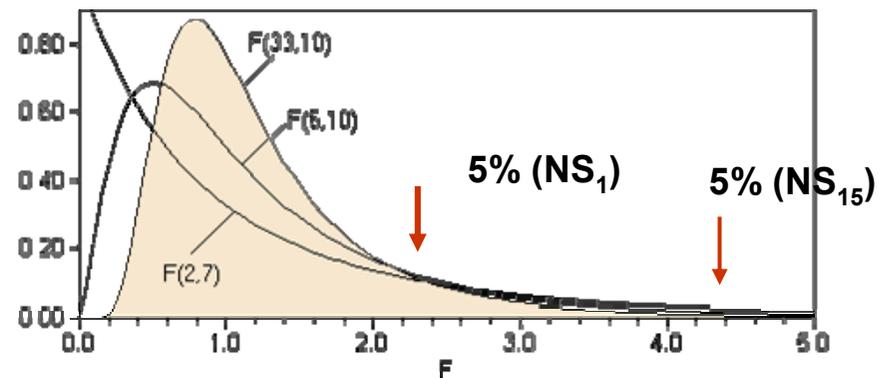
4.5. La estimación de los efectos (los test *a posteriori*):

- El problema de los (n) tests múltiples:
 - . Bonferroni (α/n) y otros (discutible su aplicación general)

Si tenemos 3 tratamientos A, B y C, tenemos 3 tests posibles:

$$\begin{array}{l}
 \cdot \text{A con B} \\
 \cdot \text{A con C} \\
 \cdot \text{C con C}
 \end{array}
 = \binom{3}{2} = 3$$

Si tenemos 15 tratamientos,
el número de test será de **105**



Si en un solo test el porcentaje de error es del 5%
en 105 test el número de ellos que saldrán significativos
por error será de $105 \times 0,05 = 5,25$

Para 105 test la p debería corregirse como $0.05/105 = 0,0005$

Se necesita por lo tanto corregir la probabilidad en las comparaciones múltiples



Tema IV. EL ANOVA de un factor

4.5. La estimación de los efectos (los test a posteriori):

- Las comparaciones *a priori* ó *a posteriori*
- Los métodos de comparación a posteriori:
 - . El Student-Newman-Keuls (SNK)
 - . Muchos otros (Bartlet,
 - . Pueden producir incoherencias (interpretación conservadora)

A	B	C
120	130	79
110	105	99
140	160	100
135	123	88
129	140	99

Fuente	SC	GL	CM	$F_{(2,12)}$
Entre (b)	4425,7	2	2212,9	10,3
Dentro (w)	2578	12	214,8	
TOTAL	7003,7	14		

Student-Newman-Keuls

Calmantes	N	Subconjunto para alfa = 0.05	
		1	2
C	5	93.0000	
A	5		126.8000
B	5		131.6000
Significación		1.000	0.614

El grupo C es el causante de las diferencias



Tema IV. EL ANOVA de un factor

4.5. La estimación de los efectos (los test a posteriori):

- Las comparaciones *a priori* ó *a posteriori*
- Los métodos de comparación a posteriori (se eligen *a priori*):
 - . El Student-Newman-Keuls (SNK)
 - . Muchos otros (Sidak, Scheffe, Tukey, Duncan, etc)
 - . Pueden producir incoherencias (interpretación conservadora)

A	B	C
120	130	79
110	105	99
140	160	100
135	123	88
129	140	99

Fuente	SC	GL	CM	F _(2,12)
Entre (b)	4425,7	2	2212,9	10,3
Dentro (w)	2578	12	214,8	
TOTAL	7003,7	14		

Comparaciones múltiples

Variable dependiente: tiempo

Sidak

(I) calmantes	(J) calmantes	Diferencia de medias (I-J)	Error típico	Sig.	Intervalo de confianza al 95%	
					Límite inferior	Límite superior
1.00	2.00	-4.80000	9.27002	.942	-30.4808	20.8808
	3.00	33.80000*	9.27002	.010	8.1192	59.4808
2.00	1.00	4.80000	9.27002	.942	-20.8808	30.4808
	3.00	38.60000*	9.27002	.004	12.9192	64.2808
3.00	1.00	-33.80000*	9.27002	.010	-59.4808	-8.1192
	2.00	-38.60000*	9.27002	.004	-64.2808	-12.9192

*. La diferencia de medias es significativa al nivel .05.

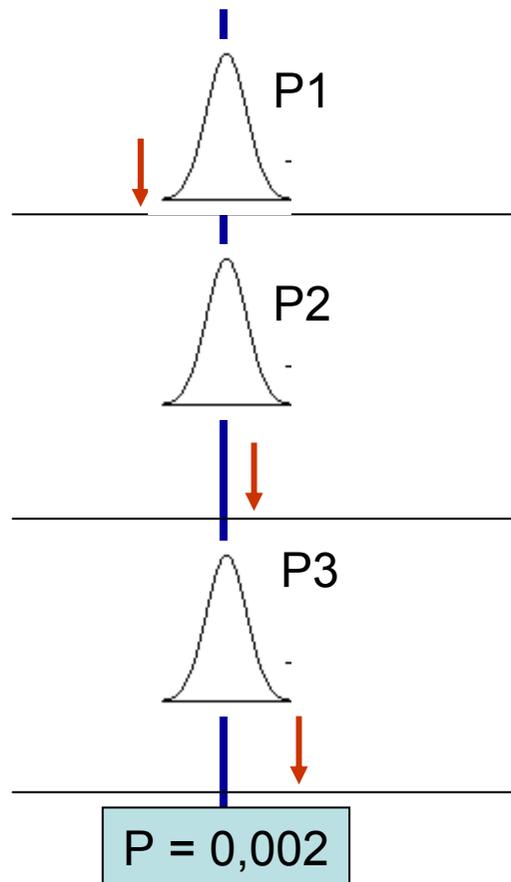


Tema IV. EL ANOVA de un factor

4.6. Las asunciones del ANOVA:

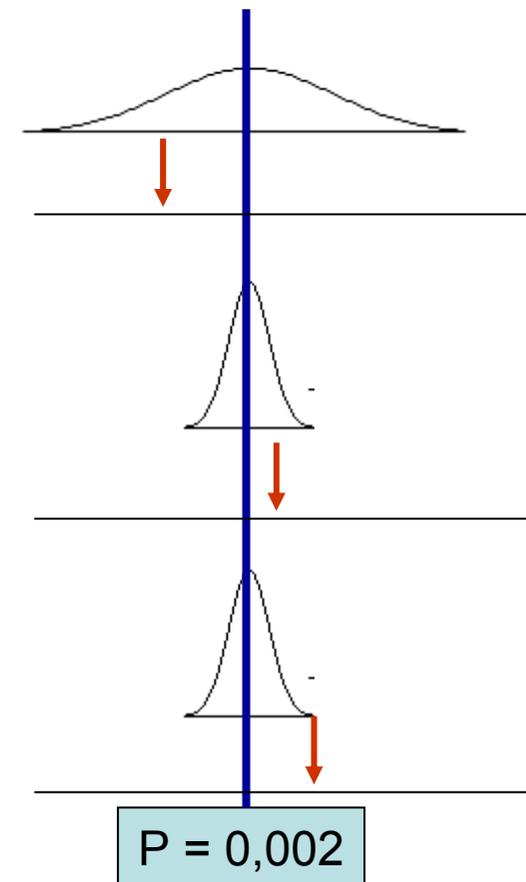
- Muestreo aleatorio (sino ANOVA de aleatorización)
- Distribución Normal (poco importante)
- Homogeneidad de varianzas (importante)

Varianzas Homogéneas



En el 2º caso
 $p \gg 0,002$

Varianzas Heterogéneas





Tema IV. EL ANOVA de un factor

4.6. Las asunciones del ANOVA:

- Los test de evaluación de varianzas:
 - . Test de Levine (varianzas distintas)
 - . Test de Cochran (una varianza más grande)

m_i = mediana de cada grupo
 Z_i = promedio de la desviación a la mediana
 s_i^2 = varianza muestral

$$Z_{ij} = |X_{ij} - m_i|$$

$$\bar{Z}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Z_{ij}$$

$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Z_{ij} - \bar{Z}_i)^2$$

se utiliza un test t clásico

$$R_c = \frac{(\text{máx. Var})^n}{\sum (\text{varianzas})}$$

se compara R_c con valor en tabla



Tema IV. EL ANOVA de un factor

4.7. Transformaciones de los datos:

- Pueden corregir Normalidad (no es importante)
- Pueden corregir heterogeneidad (si es importante):
 - . Raíz cuadrada (cuando los datos siguen una Poisson; hay muchos ceros)
 - . El logaritmo (cuando hay efectos mutiplicativos; distribuciones exponenciales, etc)
 - . Etc
- Aumento de complejidad en la interpretación

Transformación Raíz Cuadrada

$\sqrt{(x + 1/2)}$ = mayoría de valores entre 2 y 10

$\sqrt{(x + 1)}$ = muchos ceros

Transformación Logarítmica

$\log(x)$ = distribuciones muy asimétricas hacia la derecha



Tema IV. EL ANOVA de un factor

4.8. El ANOVA de aleatorización:

- Cuando el paramétrico no se puede aplicar (normalidad, heterogeneidad, otros)
- Procedimiento (ejemplo WEB):

	A	B	C
	120	130	79
	110	105	99
	140	160	100
	135	123	88
	129	140	99
\bar{X}	126,8	131,6	93

para cada nueva tabla aleatorizada se obtiene un valor de F

se obtiene la distribución de F por aleatorización

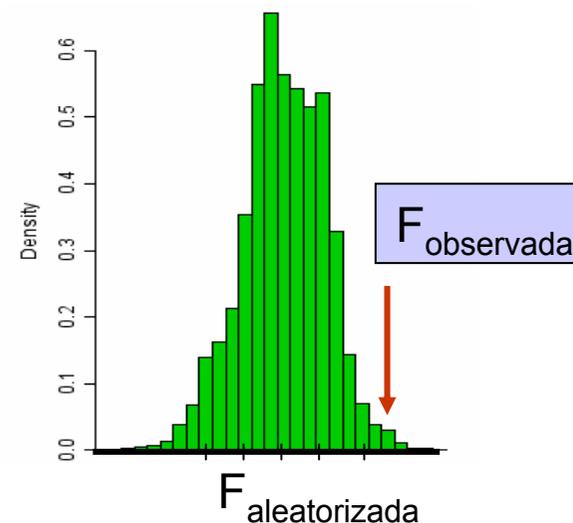
se obtiene la p ($F_{\text{observada}}$)

15 valores distribuidos en 3 grupos:

Val1 =120
Val2 =110
Val3 =140
Val4 =135.....

se aleatorizan las posiciones de los valores:

val12 val12 val9
val1 val5 val15...





Referencias Bibliográficas

LIBROS:

Sokal, R.R., Rohlf, F.J. 1995. Biometry. Freeman and co., New York

Underwood, A.J. 1981. Techniques of analysis of variance in experimental marine biology and ecology. *Oceanogr. Mar. Biol. Ann. Rev.*, 19: 513-605.

PÁGINAS WEB:

<http://stattrek.com/Tables/F.aspx>

(se dan valores de probabilidad de test F, y otros)

<http://faculty.vassar.edu/lowry/tabs.html>

(otro sitio para obtener la probabilidad de F, y otras distribuciones)

<http://webs.uvigo.es/c03/webc03/XENETICA/XB2/software.htm>

(página Web con ANOVA de aleatorización freeware)