ANOVA multifactorial

Diseños ortogonales

Universitat d'Alacant
Universidad de Alicante

Dept. of Marine Science and Applied Biology
Jose Jacobo Zubcoff

ANOVA

En el caso de 1 factor:

Compara la distribución de una variable continua normal en dos o más poblaciones

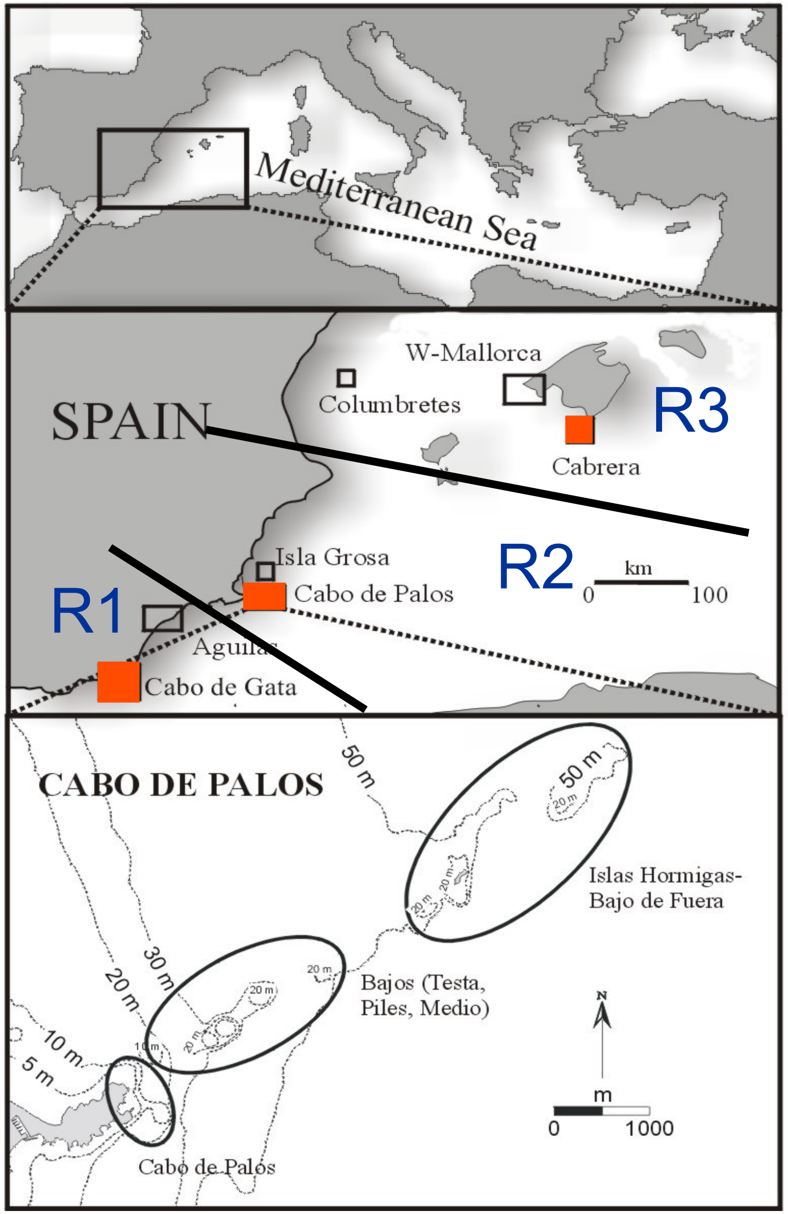
(niveles o categorías)

Pruebas de contraste para dos o más grupos independientes (ANOVA entre sujetos): un factor completamente aleatorizado.

Se puede examinar **más de un factor** simultáneamente (ANOVA de 2 factores, de 3 factores, etc.)

¿Por qué un único análisis para todos los factores, en lugar de ~~dos o más análisis separados~~?

- Eficacia
- Disminuye el efecto "achacado" al error aleatorio: (disminuir la probabilidad del error Tipo I)
- Mayor información (efecto combinado de los factores)



Ejemplo
Diseño con 2 factores

Hipótesis 1: la abundancia total de los peces litorales mediterráneos responde a la protección pesquera en las reservas marinas

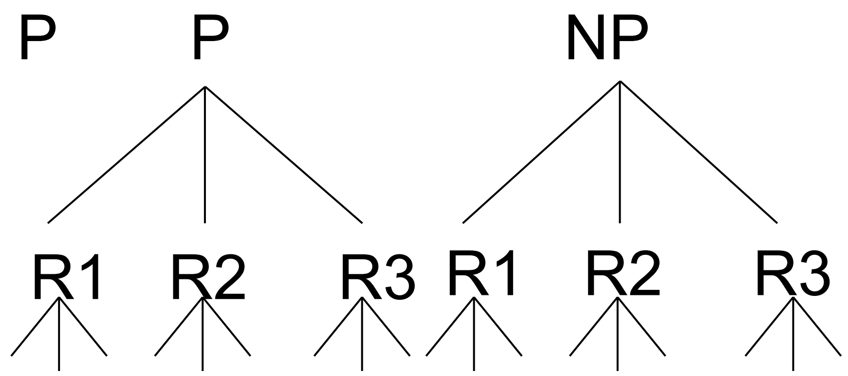
Hipótesis 2: pueden haber diferencias regionales de abundancia, achacables a variaciones puramente espaciales (debidas a otros factores además de la protección)

Tenemos dos factores experimentales:

nivel de protección P

región R

réplicas



$p = 2$

$r = 3$

$n = 3$

	R1	R2	R3
P	n	n	n
NP	n	n	n

DISEÑO ORTOGONAL

Tenemos dos factores experimentales: DISEÑO ORTOGONAL

- Modelo lineal para dos factores con interacción:

$$X_{ijk} = \mu + A_i + B_j + AB_{ij} + e_{k(ij)}$$

- Modelo lineal para dos factores SIN interacción:

$$X_{ijk} = \mu + A_i + B_j + e_{k(ij)}$$

- Modelo lineal para UN solo factor:

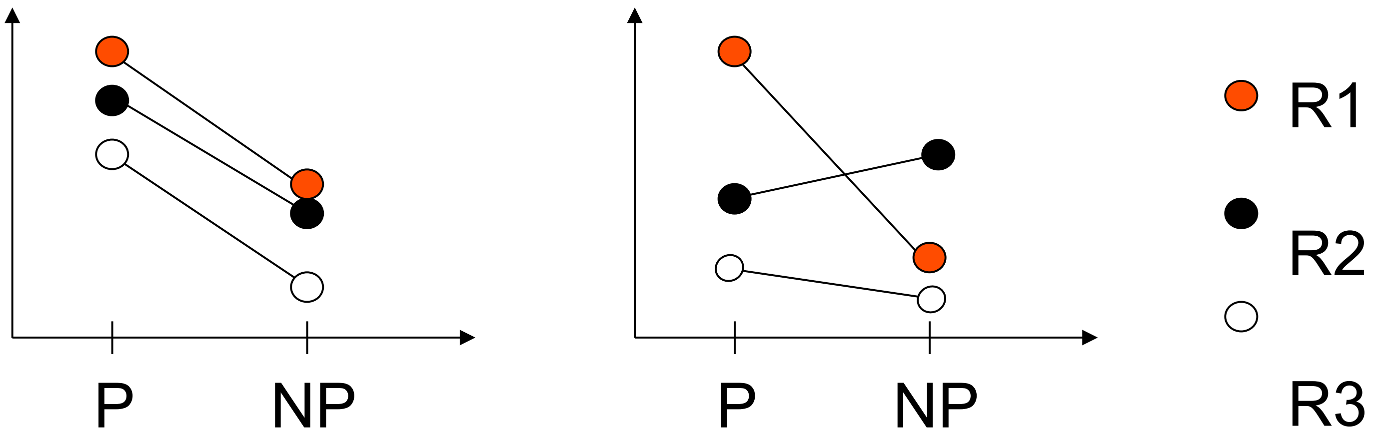
$$X_{ij} = \mu + A_i + e_{j(i)}$$

<u>Diseño ortogonal</u>

• Buscamos explorar más de un factor experimental simultáneamente y en combinación

• Cada nivel de un factor está presente en el experimento en combinación con cada nivel del otro factor (**ortogonalidad**)

• Debemos asegurar la ortogonalidad con el fin de investigar las <u>interacciones</u> entre factores



Diseños Factoriales

- Comparaciones entre los trat. -> afectados por las condiciones en las que ocurren.
- Los efectos de un factor deben tener en cuenta los efectos de otros factores. (mas real)
- Mas de un factor a la vez -> Diseño Factorial
- Son experimentos mas eficientes-> Cada observación proporciona información sobre todos los factores
- Las respuestas de un factor en diferentes niveles de otro
- Interacción-> ocurre cuando su actuación no es independiente.

Análisis para 2 factores

Efectos de un factor:

- Es un cambio en la respuesta medida ocasionado por un cambio en el nivel de ese factor
- Efectos Simples: comparaciones entre niveles de un factor (contrastes)
- Efectos principales: de un factor son comparaciones entre los niveles de un factor promediados para todos los niveles de otro factor
- Efectos de Interacción: son las diferencias entre efectos simples

Ejemplo para Diseño factorial 2 x 2

	B		
A	B1	B2	Medias del Factor A
A1	68	60	64
A2	65	97	81
Medias del Factor B	66.5	78.5	

EFECTO SIMPLE

EFECTO PRINCIPA

Ejemplo para Diseño factorial 2 x 2

- **Efectos Simples:** (para factor A)

$$\ell_1 = \mu_{21} - \mu_{11} = 65 - 68 = -3$$

$$\ell_2 = \mu_{22} - \mu_{12} = 97 - 60 = 37$$

	B		
A	B1	B2	Medias del Factor A
A1	68	60	64.0
A2	65	97	81.0
Medias del Factor B	66.5	78.5	

Ef. simple

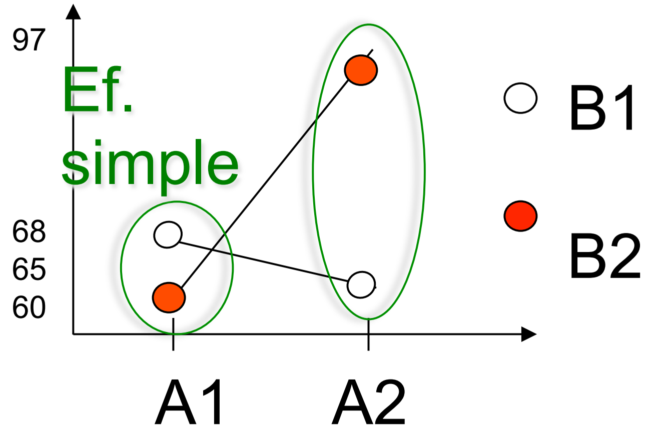
Ef. princ

- **Efectos principales:** (promedio de los dos efectos)

$$\ell_3 = \mu_{2.} - \mu_{1.} = 81 - 64 = 17$$

- **Efectos de Interacción:**

$$\ell_4 = \ell_2 - \ell_1 = 37 - (-3) = 40$$



Y el efecto del error aleatorio ¿cómo se mide?

	B		
A	B1	B2	Medias del Factor A
A1	68	60	64
A2	65	97	81
Medias del Factor B	66.5	78.5	

Error aleatorio r_1 r_2 r_n

EFECTO PRINCIPA

Diseño ortogonal

¿Cuándo un factor es fijo?

Y ¿cuándo es aleatorio?

Ejemplos: Niveles de protección, Regiones, Estaciones, Años, etc.

Un factor será **fijo** si nos interesa saber si existen subconjuntos homogéneos (hacer un contraste a posteriori).
En otro caso, será **aleatorio**.

El análisis con factores fijos es más potente, pero requiere más análisis a posteriori, no siempre justificado y no siempre racional.

Diseño ortogonal

- Hipótesis nula:

2º) Efectos de los factores

H_{01}: $\mu_1 = \mu_2 = \ldots = \mu_i = \ldots = \mu_a$

H_{02}: $\mu_1 = \mu_2 = \ldots = \mu_j = \ldots = \mu_b$

1º) Ef. de la Interacción

H_{03}: las diferencias entre niveles del factor P son independientes de las diferencias entre niveles del factor R

- En nuestro ejemplo:

H_{01}: $\mu_1 = \mu_2 = \mu_3$

H_{02}: $\mu_1 = \mu_2$

H_{03}: las diferencias entre niveles del factor P son independientes de las diferencias entre niveles del factor R

Diseño ortogonal

- Modelo lineal:

$$X_{ijk} = \mu + A_i + B_j + \cancel{AB_{ij}} + e_{k(ij)}$$

Aditividad y Efectos de los factores:

- Si **no hubiera interacción**

$$\mu_{ij} = \mu_{..} + \alpha_i + \beta_j$$

Siendo: $\alpha_i = \mu_{i.} - \mu_{..}$

$$\beta_j = \mu_{.j} - \mu_{..}$$

- Los Efectos de los factores son aditivos en ausencia de Interacción

Diseño ortogonal

- Modelo lineal:

$$X_{ijk} = \mu + A_i + B_j + \cancel{AB_{ij}} + e_{k(ij)}$$

- Si **no hubiera interacción**

$$\overbrace{\underbrace{\sum_{i=1}^{I}\sum_{j=1}^{J}(y_{ij}-\bar{y}_{..})^2}_{g.l.=IJ-1}}^{\substack{Suma\ de\ cuadrados \\ Global\ (scG)}} = \overbrace{\underbrace{J\sum_{i=1}^{I}(y_{i.}-\bar{y}_{..})^2}_{g.l.=I-1}}^{\substack{Suma\ de\ cuadrados \\ Explicada\ por\ T\alpha\ (scT)}} + \overbrace{\underbrace{I\sum_{j=1}^{J}(y_{.j}-\bar{y}_{..})^2}_{g.l.=J-1}}^{\substack{Suma\ de\ cuadrados \\ Explicada\ por\ B\beta\ (scB)}}$$

$$+ \overbrace{\underbrace{e_{ij}^2}_{g.l.=(I-1)(J-1)}}^{\substack{Suma\ de\ cuadrados \\ Residual\ (scR)}}$$

Diseño ortogonal

- Modelo lineal:

$$X_{ijk} = \mu + A_i + B_j + AB_{ij} + e_{k(ij)}$$

Si **existe** la **interacción**

$$\mu_{ij} = \mu_{..} + \alpha_i + \beta_j + \alpha\beta_{ij}$$

$$\alpha\beta_{ij} = (\mu_{ij} - \mu_{..}) - (\mu_{i.} - \mu_{..}) - (\mu_{.j} - \mu_{..})$$

$$\alpha\beta_{ij} = (\mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..})$$

Suma de Cuadrados para los efectos de los factores

$$\sum\sum\sum(y_{ijk} - y_{...})^2 = r\sum\sum(y_{ij.} - y_{...})^2 + \sum\sum\sum(y_{ijk} - y_{ij.})^2$$

a b r SC Total SC Trat. SC Error

Diseño ortogonal

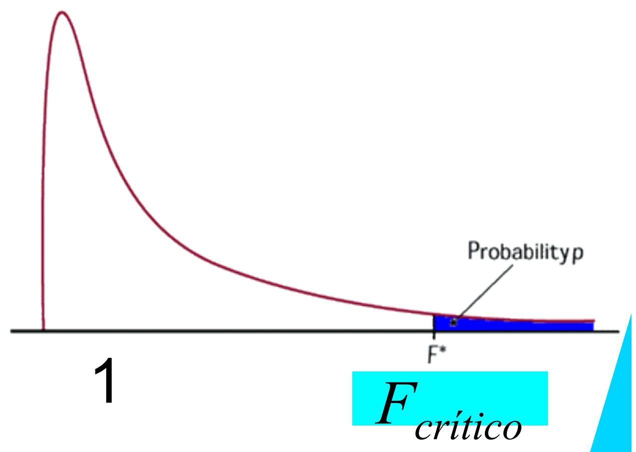
- Modelo lineal **con interacción**:

$$X_{ijk} = \mu + A_i + B_j + AB_{ij} + e_{k(ij)}$$

$$X_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$ *Modelo I de ANOVA de dos factores puro*

Prueba de NO Aditividad

$$\left.\begin{array}{l} Ho: (\alpha\beta)_{ij} = \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..} = 0 \\ H_1: (\alpha\beta)_{ij} = \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..} \neq 0 \end{array}\right\}$$



$$\mathcal{F}_{interacción} = \mathrm{CM(AB)} / \mathrm{CME}$$

Diseño ortogonal

- Modelo lineal **con interacción**:

$$X_{ijk} = \mu + A_i + B_j + AB_{ij} + e_{k(ij)}$$

$$\overbrace{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk}-\overline{y}_{...})^2}^{\substack{Suma\ de\ cuadrados \\ Global\ (scG)}}_{\underbrace{}_{g.l.=IJK-1}} = \underbrace{\overbrace{JK\sum_{i=1}^{I}(\overline{y}_{i..}-\overline{y}_{...})^2}^{\substack{Suma\ de\ cuadrados \\ Explicada\ por\ T\alpha\ (scT\alpha)}}}_{g.l.=I-1} + \underbrace{\overbrace{IK\sum_{j=1}^{J}\left(\overline{y}_{.j.}-\overline{y}_{...}\right)^2}^{\substack{Suma\ de\ cuadrados \\ Explicada\ por\ T\beta\ (scT\beta)}}}_{g.l.=J-1}$$

$$+ \underbrace{\overbrace{K\sum_{i=1}^{I}\sum_{j=1}^{J}\left(y_{ijk}-\overline{y}_{i..}-\overline{y}_{.j.}+\overline{y}_{...}\right)^2}^{\substack{Suma\ de\ cuadrados \\ Explicada\ por\ interac.\ (sc\alpha\beta)}}}_{g.l.=(I-1)(J-1)} + \underbrace{\overbrace{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(y_{ijk}-\overline{y}_{ij.}\right)^2}^{\substack{Suma\ de\ cuadrados \\ Residual\ (scR)}}}_{g.l.=IJ(K-1)}$$

Diseño ortogonal

A y B Fijos

- Modelo lineal:

$$X_{ijk} = \mu + A_i + B_j + AB_{ij} + e_{k(ij)}$$

Fuente de variación	**SC**
A	$\sum \sum \sum (X_i - X)^2$
B	$\sum \sum \sum (X_j - X)^2$
AB	$\sum \sum \sum (X_{ij} - X)^2 - \sum \sum \sum (X_i - X)^2 - \sum \sum \sum (X_j - X)^2$
Residual	$\sum \sum \sum (X_{ijk} - X_{ij})^2$
Total	$\sum \sum \sum (X_{ijk} - X)^2$

Diseño ortogonal

A y B Fijos

- Modelo lineal:

$$X_{ijk} = \mu + A_i + B_j + AB_{ij} + e_{k(ij)}$$

Fuente de variación	SC	g.l.
A	$\sum\sum\sum(X_{i..} - X_{...})^2$	
B	$\sum\sum\sum(X_{j..} - X_{...})^2$	
AB	$\sum\sum\sum(X_{ij.} - X_{i..} - X_{j..} - X_{...})^2$	
Residual	$\sum\sum\sum(X_{ijk} - X_{ij.})^2$	
Total	$\sum\sum\sum(X_{ijk} - X_{...})^2$	

Cálculo grados de libertad

$$[ab - a - (b - 1)] =$$
$$(ab - a - b + 1) =$$
$$(a - 1)(b - 1)$$

Diseño ortogonal

A y B Fijos

- Modelo lineal:

$$X_{ijk} = \mu + A_i + B_j + AB_{ij} + e_{k(ij)}$$

Fuente de variación	SC	g.l.
A	$\sum\sum\sum(X_i - X)^2$	(a–1)
B	$\sum\sum\sum(X_j - X)^2$	(b–1)
AB	$\sum\sum\sum(X_{ij} - X_i - X_j - X)^2$	(a–1) (b–1)
Residual	$\sum\sum\sum(X_{ijk} - X_{ij})^2$	ab(n–1)
Total	$\sum\sum\sum(X_{ijk} - X)^2$	abn–1

Diseño **ortogonal**

Fuente de variación	g.l.	**Est. SC**
A	(a–1)	$(a-1)\sigma_e^2 + bn \sum (AB_i - AB)^2 + bn \sum (A_i - A)^2$
B	(b–1)	$(b-1)\sigma_e^2 + an \sum (AB_i - AB)^2 + an \sum (B_i - B)^2$
AB	(a–1) (b–1)	σ_e^2 + interacción
Residual	ab(n–1)	σ_e^2
Total	abn–1	

Diseño **ortogonal**

• Importancia de definir si los factores considerados son **fijos o aleatorios:**

Matemáticamente se puede demostrar que el 2º término del estimador de la SC es 0 cuando el factor es fijo, puesto que…

$$\sum AB_{ij} = 0$$

En cambio, cuando uno de los factores (p.ej. B) es aleatorio, los valores B_j constituyen una muestra de todos los posibles valores, y el sumatorio de las interacciones será $\neq 0$

En nuestro ejemplo?

Nivel de protección: FIJO

Región: ALEATORIO

Diseño ortogonal: A fijo
B aleatorio

Fuente de var.	g.l.	Est. MC	Est. MC
A	(a–1)	$\sigma^2_e + n\sigma^2_{AB} + bn \sum (A_i - A)^2 / (a-1)$	$\sigma_e^2 + n\sigma^2_{AB} + bnk^2_A$
B	(b–1)	$\sigma^2_e + an\sigma^2_B$	$\sigma_e^2 + an\sigma^2_B$
AB	(a–1) (b–1)	$\sigma^2_e + n\sigma^2_{AB}$	$\sigma_e^2 + n\sigma^2_{AB}$
Residual	ab(n–1)	σ^2_e	σ^2_e
Total	abn–1		

Diseño ortogonal: A fijo B fijo

Fuente de var.	g.l.	Est. MC	Est. MC
A	(a–1)	$\sigma^2_e + bn \sum (A_i - A)^2 / (a-1)$	$\sigma_e^2 + bnk^2_A$
B	(b–1)	$\sigma^2_e + an \sum (B_i - B)^2 / (b-1)$	$\sigma_e^2 + ank^2_B$
AB	(a–1) (b–1)	σ^2_e + n [var. Interacc.] / (a-1)(b–1)	$\sigma_e^2 + nk^2_{AB}$
Residual	ab(n–1)	σ^2_e	σ_e^2
Total	abn–1		

Diseño ortogonal: A fijo B fijo

Fuente de var.	g.l.	Est. MC	Est. MC
A	(a–1)	$\sigma^2_e + bn \sum (A_i–A)^2 / (a–1)$	$\sigma_e^2 + bnk^2_A$
B	(b–1)	$\sigma^2_e + an \sum (B_i–B)^2 / (b–1)$	$\sigma_e^2 + ank^2_B$
AB	(a–1) (b–1)	σ^2_e + n [var. Interacc.] / (a-1)(b–1)	$\sigma_e^2 + nk^2_{AB}$
Residual	ab(n–1)	σ^2_e	σ_e^2
Total	abn–1		

Diseño ortogonal: A aleatorio B aleatorio

Fuente de var.	g.l.	Est. MC
A	(a–1)	$\sigma_e^2 + n\sigma^2_{AB} + bn\sigma^2_A$
B	(b–1)	$\sigma_e^2 + n\sigma^2_{AB} + an\sigma^2_B$
AB	(a–1) (b–1)	$\sigma_e^2 + n\sigma^2_{AB}$
Residual	ab(n–1)	σ^2_e
Total	abn–1	

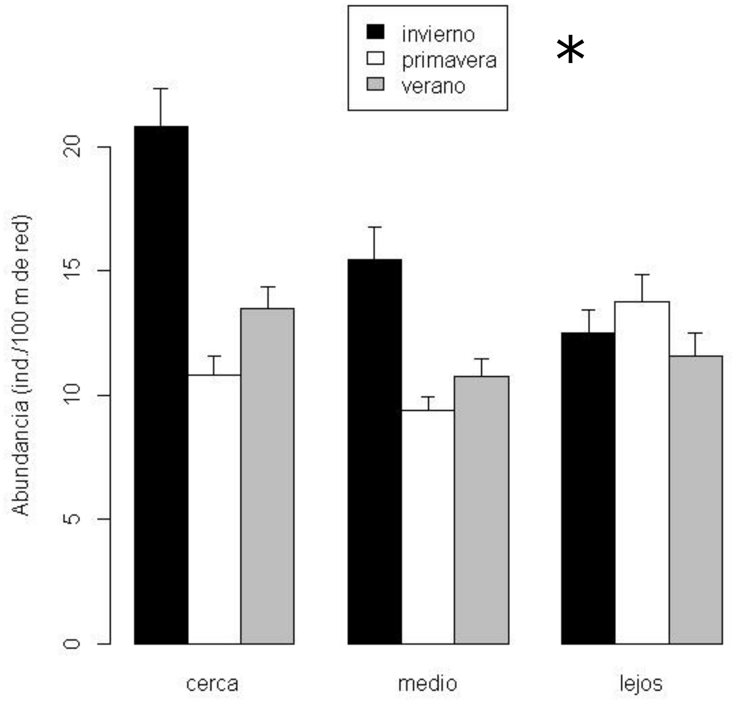
Diseño ortogonal:

Fuente de var.	g.l.	Est. MC	
A	(a–1)	$\sigma_e^2 + n\sigma^2_{AB} + bnk^2_A$	A fijo
B	(b–1)	$\sigma_e^2 + an\sigma^2_B$	B aleatorio
AB	(a–1) (b–1)	$\sigma_e^2 + n\sigma^2_{AB}$	
Residual	ab(n–1)	σ^2_e	
Total	abn–1		
A	(a–1)	$\sigma_e^2 + bnk^2_A$	A fijo
B	(b–1)	$\sigma_e^2 + ank^2_B$	B fijo
AB	(a–1) (b–1)	$\sigma_e^2 + nk^2_{AB}$	
Residual	ab(n–1)	σ_e^2	
Total	abn–1		
A	(a–1)	$\sigma_e^2 + n\sigma^2_{AB} + bn\sigma^2_A$	A aleatorio
B	(b–1)	$\sigma_e^2 + n\sigma^2_{AB} + an\sigma^2_B$	B aleatorio
AB	(a–1) (b–1)	$\sigma_e^2 + n\sigma^2_{AB}$	
Residual	ab(n–1)	σ^2_e	
Total	abn–1		

(distancia entre g.l.)

Diseño ortogonal:

Comparaciones múltiples a posteriori



Interacción NO significativa

- Se buscan las diferencias para los factores principales

Interacción significativa (*)

- Solo deben buscarse las diferencias a posteriori para la interacción
- Se deben analizar las combinaciones con al menos 1 nivel común

Diseño ortogonal:

Comparaciones múltiples a posteriori (p.ej. SNK)

ANOVA → A x B ns → A * → SNK para A (bn casos)

ANOVA → A x B ns → B * → SNK para B (an casos)

ANOVA → A x B * → SNK para todas las combinaciones de A y B

STOP

Diseño ortogonal

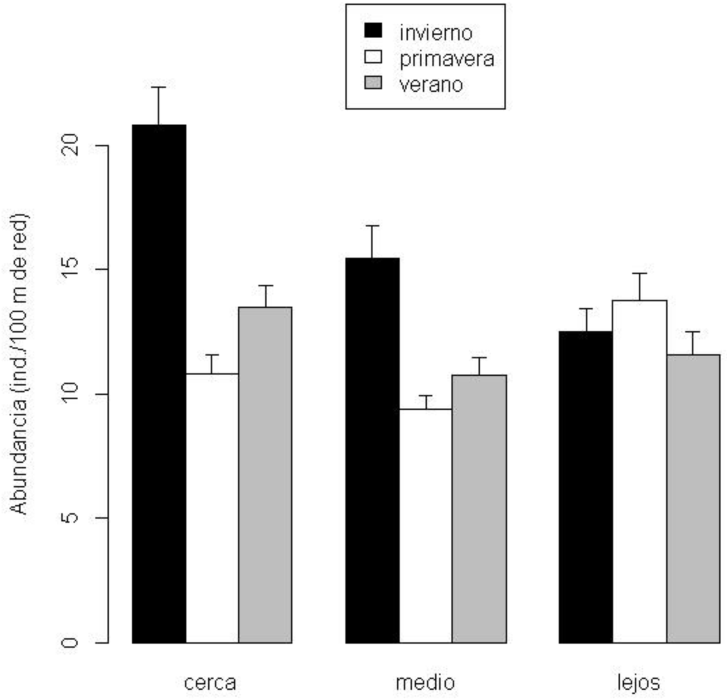
Comparaciones múltiples a posteriori

```
TukeyHSD(aov(PADINA~Zona*Profundidad))
  Tukey multiple comparisons of means
    95% family-wise confidence level
Fit: aov(formula = PADINA ~ Zona * Profundidad)

$Zona
        diff       lwr      upr     p adj
N-C -2.314815 -12.03046 7.400825 0.8395177
S-C -4.259259 -13.97490 5.456381 0.5544780
S-N -1.944444 -11.66008 7.771196 0.8838268

$Profundidad
        diff       lwr       upr p adj
2-1 -17.22222 -23.84418 -10.60026 8e-07

$`Zona:Profundidad`
             diff       lwr       upr     p adj
N:1-C:1  -6.666667 -23.42123 10.087898 0.8603058
S:1-C:1  -5.740741 -22.49530 11.013823 0.9209766
C:2-C:1 -21.111111 -37.86568 -4.356547 0.0049711
N:2-C:1 -19.074074 -35.82864 -2.319510 0.0156529
S:2-C:1 -23.888889 -40.64345 -7.134325 0.0008761
S:1-N:1   0.925926 -15.82864 17.680490 0.9999854
C:2-N:1 -14.444444 -31.19901  2.310120 0.1341259
N:2-N:1 -12.407407 -29.16197  4.347157 0.2740355
S:2-N:1 -17.222222 -33.97679 -0.467658 0.0400863
C:2-S:1 -15.370370 -32.12493  1.384194 0.0922796
N:2-S:1 -13.333333 -30.08790  3.421231 0.2018392
S:2-S:1 -18.148148 -34.90271 -1.393584 0.0253708
```



Diseños ortogonales

- No hay límites **matemáticos** a la adición de factores ortogonales

$$X_{ijkl} = \mu + A_i + B_j + C_k + AB_{ij} + AC_{ik} + BC_{jk} + ABC_{ijk} + e_{l(ijk)}$$

8 combinaciones de tipos de factores:

fff
ffa
faf
aff
faa
afa
aaf
aaa

- Los límites están en… … imaginar las hipótesis subyacentes
… interpretar las interacciones