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 si la distribución de una variable continua normal se ve afectada por el efecto de un FACTOR

H1: Algún nivel del factor tiene distinta media

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)

Factores de un experimento $\left\{ \begin{array}{l} \text{FIJOS} \\ \\ \text{ALEATORIOS} \end{array} \right.$

Depende de la **hipótesis biológica subyacente**
No es una propiedad intrínseca del análisis de la varianza

Factor fijo:

- Tiene **hipótesis subyacente**
- Todos los niveles relevantes del factor están incluidos en el experimento.

Ejemplos:

- estaciones del año
- temperatura (p.ej. cálida vs. fría)
- profundidad (p.ej. somera vs. profunda)
- hábitat (p.ej. sebadal vs. arena)
- determinada perturbación de origen antrópico (p.ej. el impacto de un vertido)
- efecto de la protección pesquera en una reserva marina

Factor aleatorio:

- **No hay hipótesis biológica subyacente**
- Solo se incluye para medir alguna variabilidad achacable a un factor no relevante (medir variabilidad espacial, temporal, etc.)
- No todos los niveles relevantes del factor están incluidos en el experimento, únicamente una fracción (elegida al azar)
- No tenemos interés en dar conclusiones **"a posteriori"** para estos factores o sus interacciones

Ejemplos:

- variaciones espaciales (p.ej. entre cabos o bahías)
- variaciones temporales (p.ej. entre años, o a lo largo del año)
- temperatura
- profundidad

Asegurarnos de que se han tomado *n* réplicas en todos los niveles del factor !!!

Comparaciones múltiples "a posteriori"

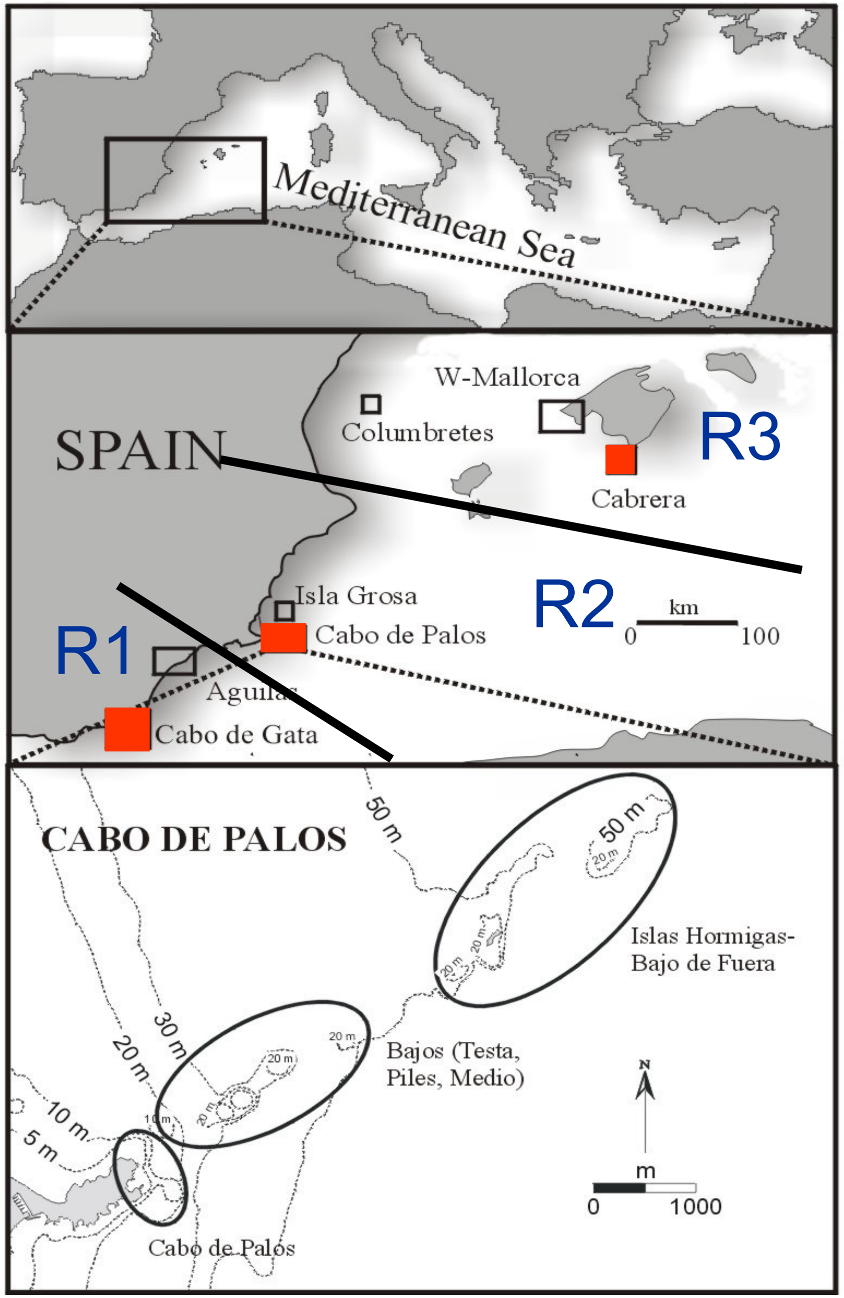
Para un factor **aleatorio**, con rechazar la H_0 **se acaba el análisis!!**
(no es necesario ir más allá, porque no hay una hipótesis subyacente)

Factor **fijo**: si hay más de dos tratamientos (o niveles del factor) es **necesario**

saber qué media(s) es (son) significativamente(s) diferente del resto,

es decir:

identificar la hipótesis alternativa a H_0



Ejemplo
Diseño con 2 factores

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

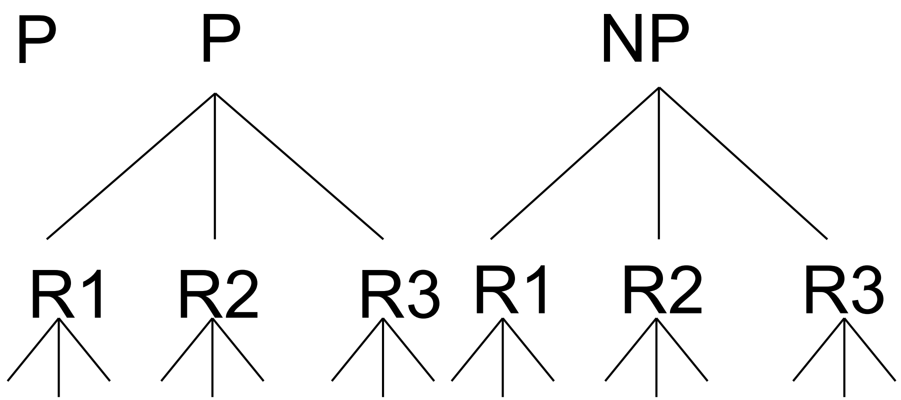
Hipótesis 2: pueden haber diferencias regionales de abundancia, achacables a variaciones puramente espaciales (debidas a otros factores además de la protección)

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 **UN solo factor**:

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

- Modelo lineal para **dos factores SIN interacción**:

$$X_{ijk} = \mu + A_i + B_j + e'_{k(ij)}$$

- Modelo lineal para **dos factores con interacción**:

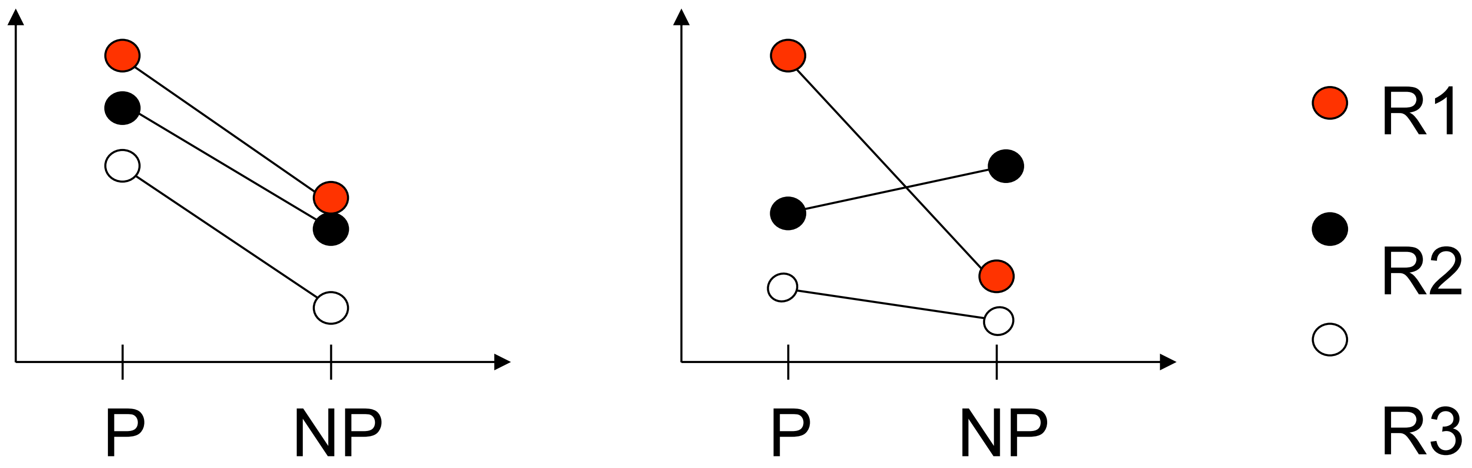
$$X_{ijk} = \mu + A_i + B_j + AB_{ij} + e''_{k(ij)}$$

Diseño ortogonal

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

• **Ortogonalidad:** *Cada nivel de un factor está presente en combinación con cada nivel del otro factor*

• Debemos asegurar la ortogonalidad para estudiar las interacciones



Diseños Factoriales

- Comparaciones entre los tratamientos – en las condiciones en las que ocurren
- Los efectos de un factor tienen en cuenta los efectos de otros factores (mas real)
- Experimentos mas eficientes
  - Cada observación proporciona información sobre **todos** los factores
  - Las respuestas de un factor en diferentes niveles de otro

- **Se puede comprobar la Interacción**
  - El efecto de un factor no es independiente de los niveles de otro factor

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

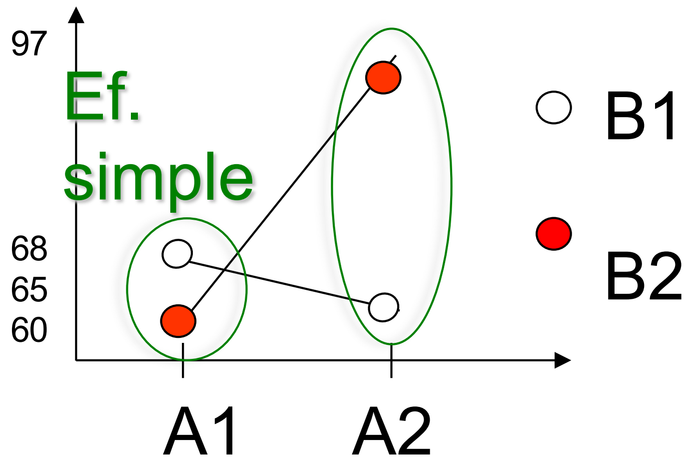
Ef. princ

- 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.

Un poco de "background"

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 **con interacción**:

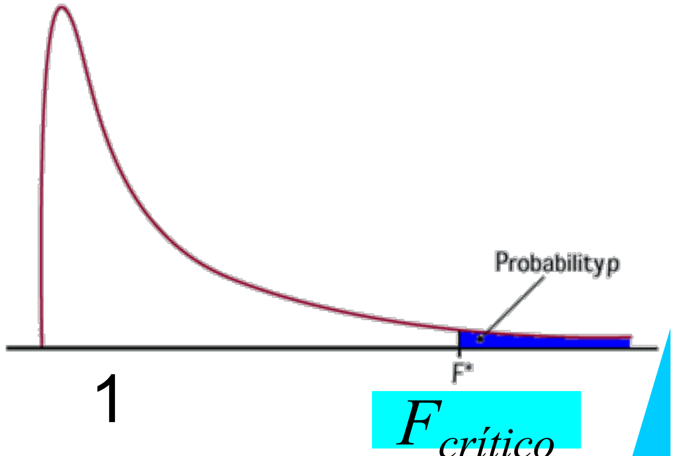
$$X_{ijk} = \mu + A_i + B_j + AB_{ij} + e_{k(ij)}$$

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

Prueba de NO Aditividad

$$Ho: (\alpha\beta)_{ij} = \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..} = 0$$

$$H_1: (\alpha\beta)_{ij} = \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..} \neq 0$$



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

Diseño ortogonal

- Modelo lineal **con interacción**:

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

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

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

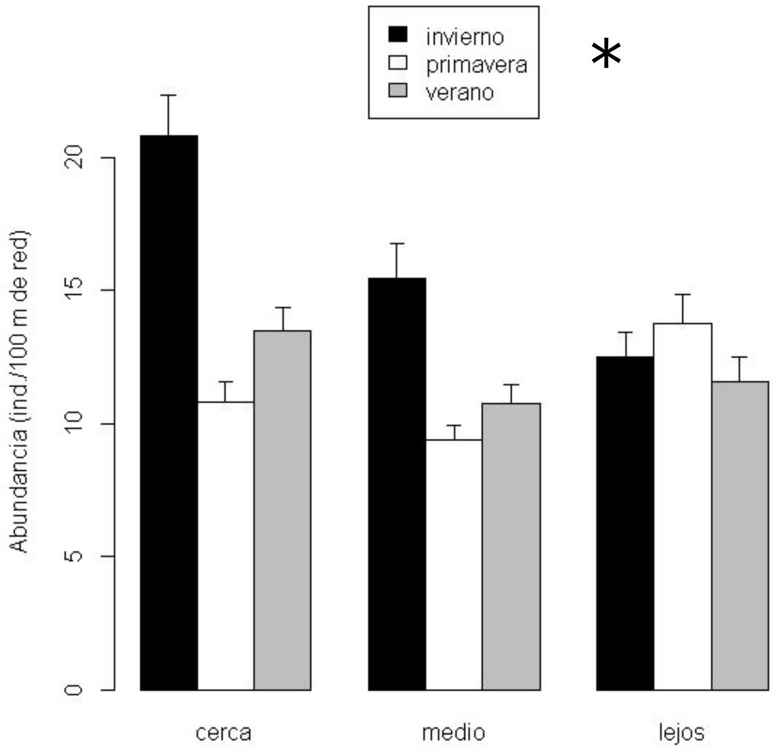
Diseño ortogonal:

Fuente de var.	g.l.	Est. MC	
A	(a–1)	$\boxed{\sigma_e^2 + n\sigma^2_{AB}} + \boxed{bnk^2_A}$	A fijo
B	(b–1)	$\sigma_e^2 + an\sigma^2_B$	B aleatorio
AB	(a–1) (b–1)	$\boxed{\sigma_e^2 + n\sigma^2_{AB}}$	
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



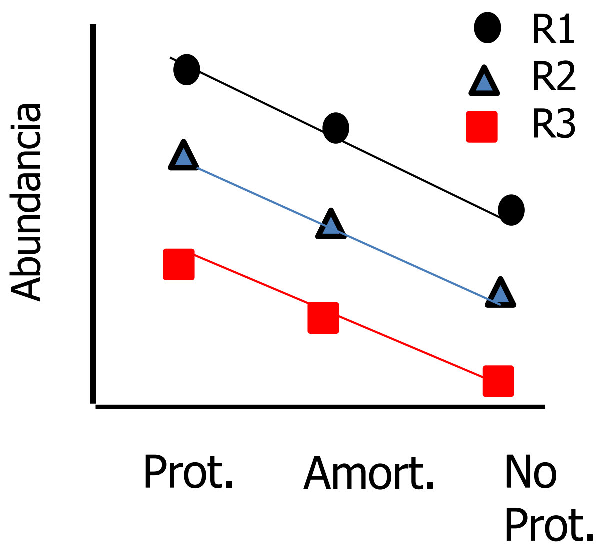
Interacción NO significativa

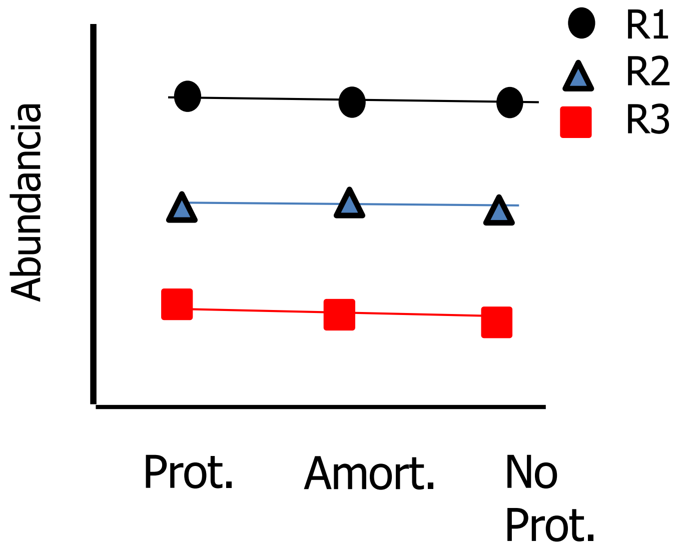
Se buscan las diferencias para los factores principales

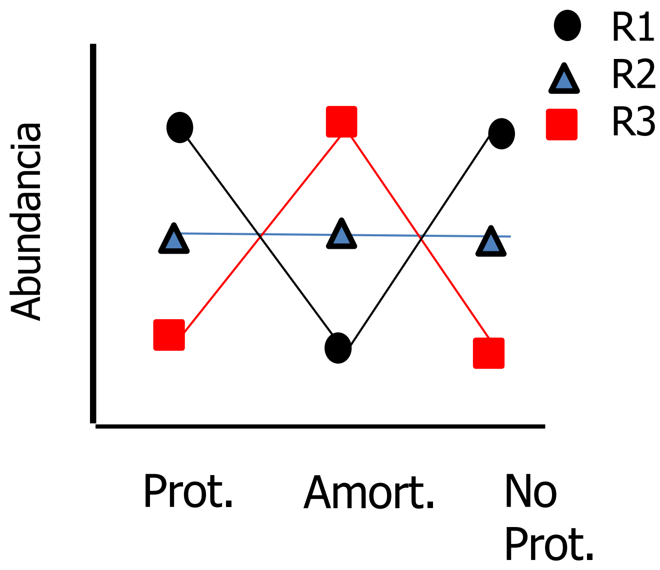
Interacción significativa (*)

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

Diseño ortogonal: la importancia de estudiar la interacción



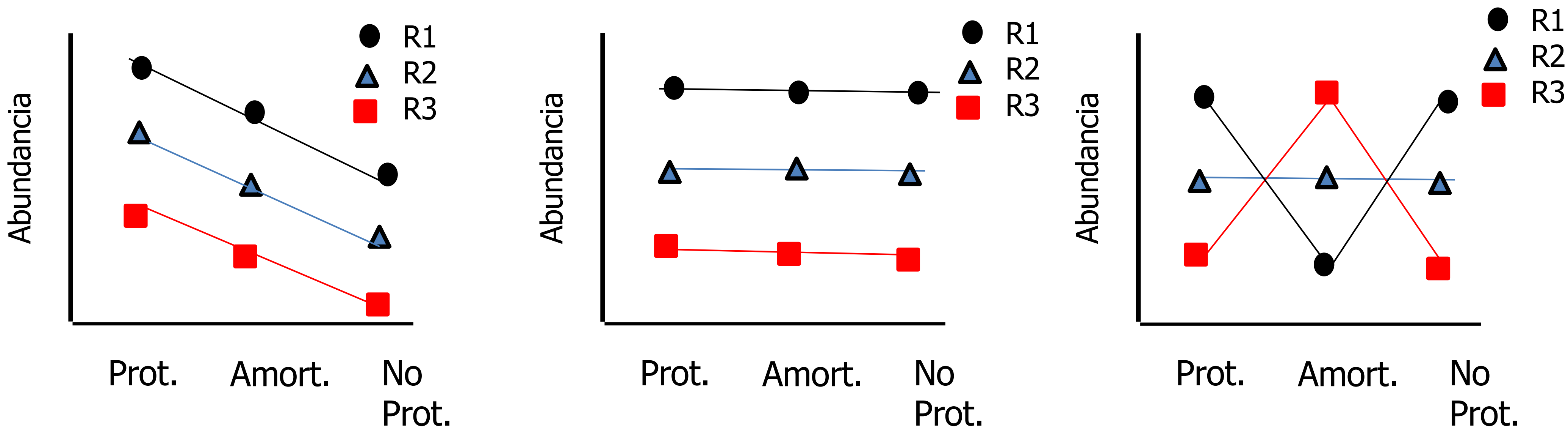




Interpretación VISUAL

¿Hay efecto de la Interacción? *(si hay es lo más importante)
¿Hay efecto de la Protección?
¿Hay efecto de la Región?

Diseño ortogonal: la importancia de estudiar la interacción



Interpretación VISUAL

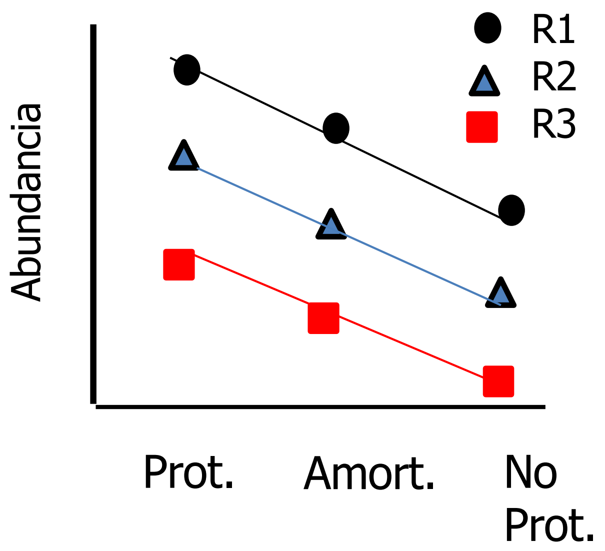
Interacción: mismo patrón en las 3 Regiones, NO hay interacción
Protección: sí hay efecto
Región: parece haber efecto

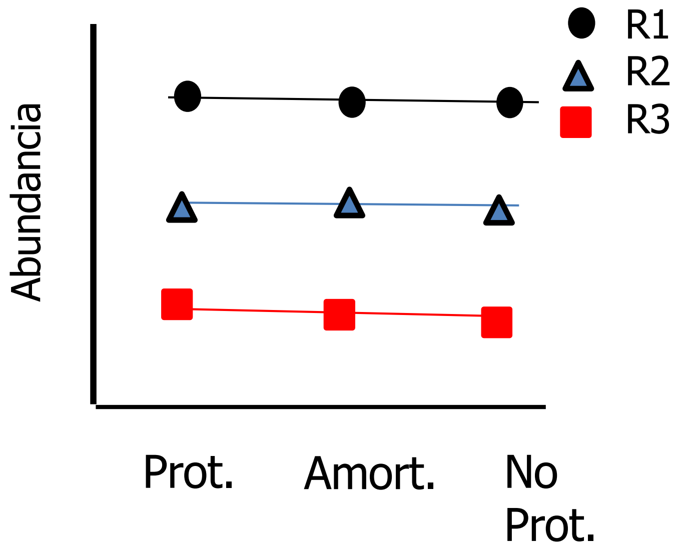
Interacción: mismo patrón en 3 Regiones, NO hay interacción
Protección: No hay efecto
Región: Sí hay efecto

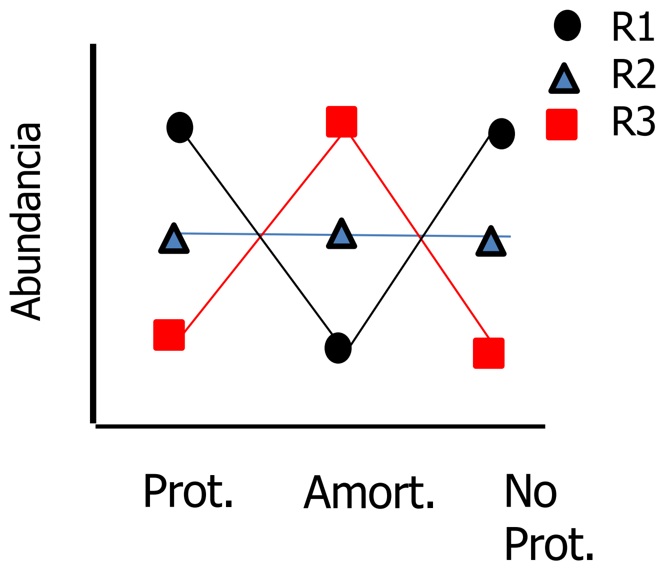
Interacción: distinto patrón en las 3 Regiones, SI hay interacción
Protección: …
Región: …

Estas tendencias habrá que comprobarlas en la tabla de ANOVA

Diseño ortogonal: la importancia de estudiar la interacción







Interpretación VISUAL

Test a posteriori para Interacción

Interacción: mismo patrón en las 3 Regiones, NO hay interacción
Protección: sí hay efecto
Región: parece haber efecto

Interacción: mismo patrón en 3 Regiones, NO hay interacción
Protección: No hay efecto
Región: Sí hay efecto

Interacción: distinto patrón en las 3 Regiones, SI hay interacción
Protección: …
Región: …

(Test a posteriori para factores individuales si son significativos en ANOVA)

Estas tendencias habrá que comprobarlas en la tabla de ANOVA

<u>Diseño ortogonal</u>:

Comparaciones múltiples a posteriori (p.ej. SNK)

ANOVA

A x B ns

A * → SNK para A (bn casos)

B * → SNK para B (an casos)

A x B * → SNK para todas las combinaciones de A y B (Interacción)

STOP

Diseño ortogonal

Comparaciones múltiples a posteriori

```
> TukeyHSD(modelo2)
$epoca
                         diff       lwr       upr     p adj
primavera-invierno -4.9351852 -6.871113 -2.999258 0.0000000
verano-invierno    -4.3055556 -6.241483 -2.369628 0.0000009
verano-primavera    0.6296296 -1.306298  2.565557 0.7242195

$`distancia:epoca`
                                    diff          lwr        upr     p adj
medio:invierno-cerca:invierno        .3                           0.0065641
lejos:invierno-cerca:invierno        .2                           0.0000005
cerca:primavera-cerca:invierno    -10.0                           0.0000000
medio:primavera-cerca:invierno    -11.3                           0.0000000
lejos:primavera-cerca:invierno     -7.0                           0.0000453
cerca:verano-cerca:invierno        -7.2                           0.0000196
medio:verano-cerca:invierno       -10.0                           0.0000000
lejos:verano-cerca:invierno        -9.2                           0.0000000
lejos:invierno-medio:invierno                                     0.4973201
cerca:primavera-medio:invierno     -4.6                           0.0315494
medio:primavera-medio:invierno     -6.0                           0.0009242
lejos:primavera-medio:invierno     -1.6                           0.9582046
cerca:verano-medio:invierno        -1.9                           0.9098410
medio:verano-medio:invierno        -4.6                           0.0296993
lejos:verano-medio:invierno        -3.8                           0.1412464
cerca:primavera-lejos:invierno     -1.7                           0.9540438
medio:primavera-lejos:invierno     -3.1                           0.4181157
lejos:primavera-lejos:invierno      1.2                           0.9939690
cerca:verano-lejos:invierno         1.0                           0.9987375
medio:verano-lejos:invierno        -1.7                           0.9495963
lejos:verano-lejos:invierno        -0.9                           0.9991657
medio:primavera-cerca:primavera    -1.3                           0.9878424
lejos:primavera-cerca:primavera     2.9                           0.4838689
cerca:verano-cerca:primavera        2.7                           0.6061070
medio:verano-cerca:primavera       -0.02777778 -4.47541864 4.4198631 1.0000000
lejos:verano-cerca:primavera        0.77777778 -3.66986308 5.2254186 0.9998019
```

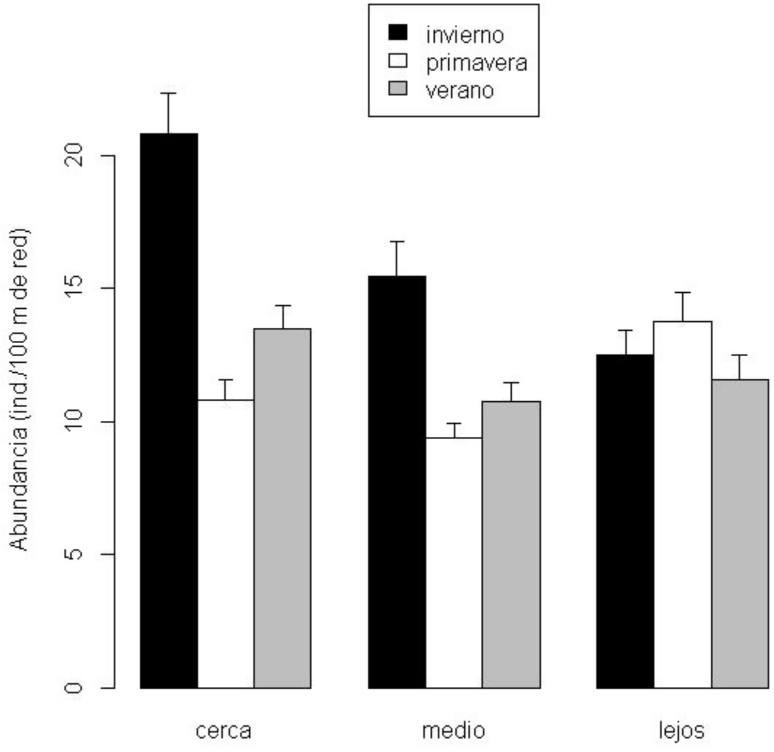
Interacción

Subconjuntos homogéneos

Por ej.: Fijamos el nivel Invierno.
Todas las comparaciones que tengan Invierno en ambos lados.
Y explicamos lo que pasa entre las distancias. La Abundancia en Invierno es:

C > M=L

Se deben obtener los subconjuntos homogéneos



Diseños ortogonales

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

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

- SOLO se hace análisis a posteriori de las interacciones entre factores FIJOS si han sido significativas en ANOVA
  - La estrategia es fijar un nivel de un factor y explicar lo que pasa en cada nivel con el resto de factores
- Si no hay interacciones significativas (de factores fijos),se estudiarán los factores FIJOS por separado (si son significativos en ANOVA).

- Los límites están en…

… imaginar las **hipótesis subyacentes**

… interpretar las interacciones

…escribir de manera clara los resultados