Construcción de modelos multifactoriales a partir de principios generales

Cálculo de estimadores de cuadrados medios y denominadores de F

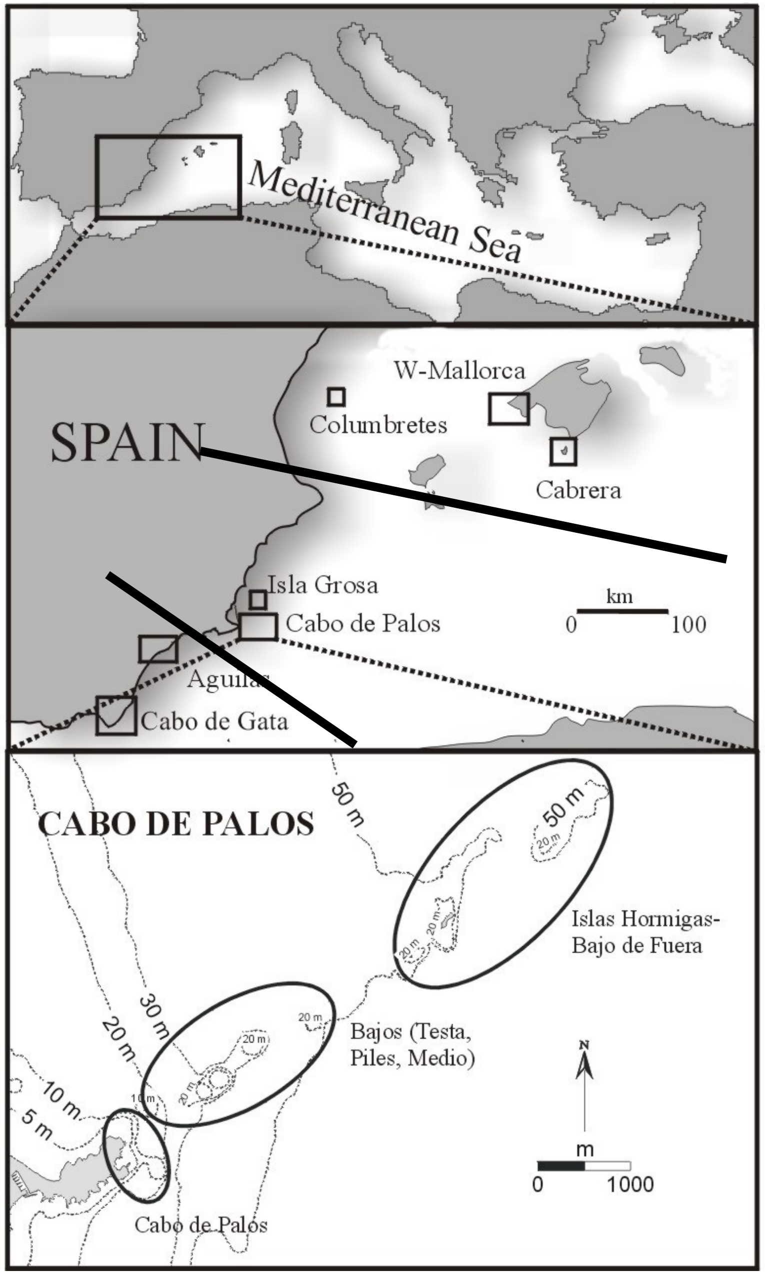
Universitat d'Alacant
Universidad de Alicante
Dept. of Marine Science and Applied Biology
Jose Jacobo Zubcoff

Diseños **mixtos**:

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

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

Hipótesis 3: pueden haber diferencias a una escala espacial inferior (en cada región).
->Factor Sitio (no hay interés especial)
(solo para detectar variabilidades a esta escala espacial)



Diseños **mixtos**:

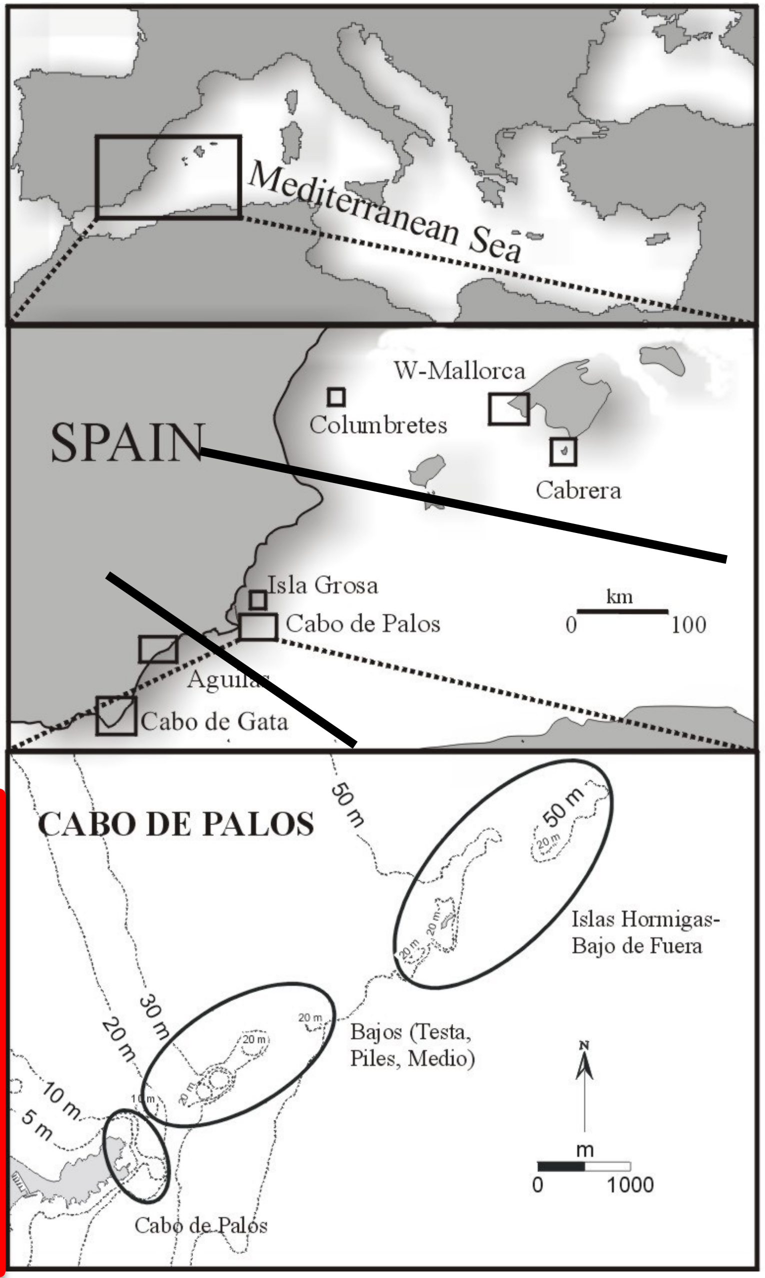
Hipótesis para la INTERACCION: la abundancia total de los peces litorales mediterráneos tiene un patrón distinto en cuanto a la Protección pesquera en las distintas Regiones.

→Interacción Protección x Region

Como son 2 factores FIJOS, y no hay más factores FIJOS, esta será la interacción de mayor interés (y de mayor nivel).

En el **supuesto** caso de tener 3 factores Ortogonales habría una triple interacción (sería la de mayor nivel) y 3 dobles interacciones:

Y = A + B + C + AxB + AxC + BxC + AxBxC +e



Diseños **mixtos**: parcialmente anidados

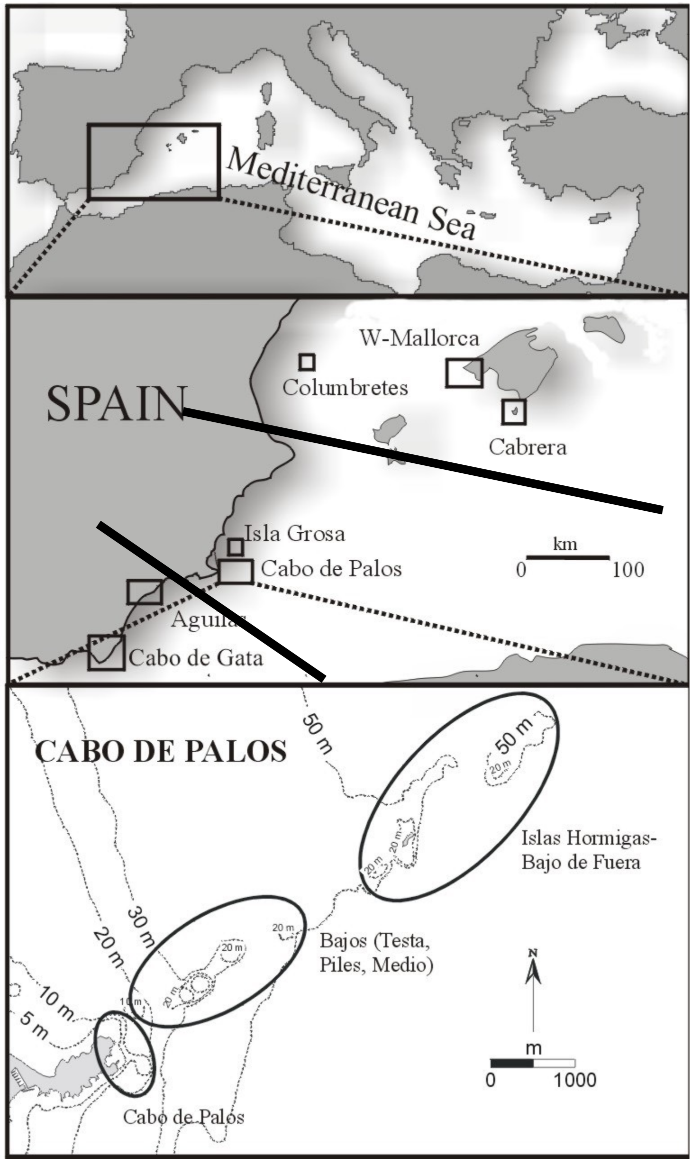
Ejemplo:

Factor **Protección** (FIJO, 2 niveles)

Factor **Región** (Ortogonal a Protección, FIJO, 3 niveles)

Factor **Sitio** (Anidado en la interacción, 3 niveles)

Réplicas: 3



Modelo lineal

$$X_{abcn} = \mu + P_a + R_b + P \times R_{ab} + S(P \times R)_{c(ab)} + e_{n(abc)}$$

Diseños **mixtos**: parcialmente anidados

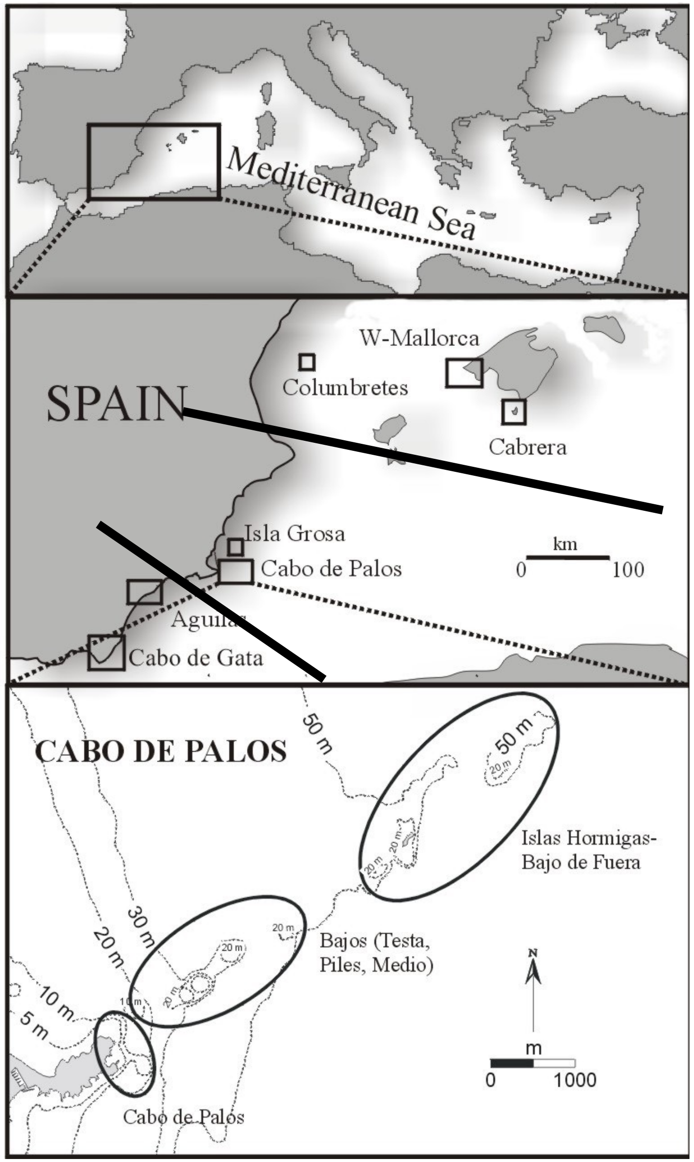
Ejemplo:

Factor **Protección** (FIJO, 2 niveles)

Factor **Región** (Ortogonal a Protección, FIJO, 3 niveles)

Factor **Sitio** (Anidado en la interacción, 3 niveles)

Réplicas: 3



Modelo lineal

$$X_{abcn} = \mu + P_{\boxed{a}} + R_{\boxed{b}} + P \times R_{\boxed{ab}} + S(P \times R)_{\boxed{c(ab)}} + e_{\boxed{n(abc)}}$$

Diseños **mixtos**: parcialmente anidados

$$X_{abcn} = \mu + P_a + R_b + P \times R_{ab} + S(P \times R)_{c(ab)} + e_{n(abc)}$$

Fuente de variación	Tipo	Nº de niveles
P	F	a = 2
R	F	b = 3
P x R		
S(P x R)	A	c = 3
Residual	A	n = 3
Total		

Cálculo de los grados de libertad (gl)

Ejemplo:

$$X_{abcn} = \mu + P_a + R_b + P \times R_{ab} + S(P \times R)_{c(ab)} + e_{n(abc)}$$

Reglas:

- Cada factor ortogonal: nº de niveles – 1
- Cada factor anidado: producto de
  - → nº de niveles del factor anidado – 1
  - → nº de combinaciones de tratamientos en los cuales ese factor está anidado
- Cada interacción: producto del nº de gl de cada término de la interacción
- Residual: producto del nº de niveles de cada factor en el experimento por el nº de réplicas – 1
- Total: producto del nº de niveles de cada factor (incluyendo réplicas) – 1

Cálculo de los grados de libertad (gl)

Ejemplo:

$$X_{abcn} = \mu + P_a + R_b + P{\times}R_{ab} + S(P{\times}R)_{c(ab)} + e_{n(abc)}$$

Fuente de variación	Tipo	Nº de niveles	gl	
P	F	p = 2		
R	F	r = 3		
P x R				
S(P x R)	A	s = 3		
Residual	A	n = 3		
Total				

Ejemplo:

$$X_{abcn} = \mu + P_a + R_b + P{\times}R_{\,ab} + S(P{\times}R)_{\,c(ab)} + e_{n(abc)}$$

Fuente de variación	**Tipo**	**Nº de niveles**	**gl**	
P	F	p = 2	(p -1)	1
R	F	r = 3	(r -1)	2
P x R			(p -1)(r -1)	2
S(P x R)	A	s = 3	pr(s – 1)	12
Residual	A	n = 3	prs(n – 1)	36
Total			prsn – 1	53

Estimas de Cuadrados Medios (CM) y contraste para el cálculo de las F

Ejemplo:

$$X_{abcn} = \mu + P_a + R_b + P \times R_{ab} + S(P \times R)_{c(ab)} + e_{n(abc)}$$

	f	f	a	a
Fuente de variación	***a***	***b***	***c***	***n***
P_a				
R_b				
$P \times R_{ab}$				
$S(P \times R)_{c(ab)}$				
$Residual_{n(c(ab))}$				

Estimas de Cuadrados Medios (CM) y contraste para el cálculo de las F

Ejemplo:

$$X_{abcn} = \mu + P_a + R_b + P \times R_{ab} + S(P \times R)_{c(ab)} + e_{n(abc)}$$

Reglas:

1. Si el subíndice situado en la columna se encuentra en la fila y:

   (a) el subíndice es **FIJO**
   - **NO** incluye componentes **ANIDADOS** (no está entre paréntesis) ➔ 0
   - **SI** incluye componentes **ANIDADOS** (está entre paréntesis) ➔ 1

   (b) el subíndice es **ALEATORIO** ➔ 1

2. Si el subíndice situado en la columna no se encuentra en la fila:
   ➔ el subíndice (nº de niveles del factor)

Estimas de Cuadrados Medios (CM) y contraste para el cálculo de las F

Ejemplo:

$$X_{abcn} = \mu + P_a + R_b + P \times R_{ab} + S(P \times R)_{c(ab)} + e_{n(abc)}$$

	f	f	a	a
Fuente de variación	***a***	***b***	***c***	***n***
P_a	0			
R_b	a			
$P \times R_{ab}$	0			
$S(P \times R)_{c(ab)}$	1			
$Residual_{n(c(ab))}$	1			

Estimas de Cuadrados Medios (CM) y contraste para el cálculo de las F

Ejemplo:

$$X_{abcn} = \mu + P_a + R_b + P\times R_{ab} + S(P\times R)_{c(ab)} + e_{n(abc)}$$

	f	f	a	a
Fuente de variación	***a***	***b***	***c***	***n***
P_a	0	b		
R_b	a	0		
$P \times R_{ab}$	0	0		
$S(P \times R)_{c(ab)}$	1	1		
$Residual_{n(c(ab))}$	1	1		

Estimas de Cuadrados Medios (CM) y contraste para el cálculo de las F

Ejemplo:

$$X_{abcn} = \mu + P_a + R_b + P \times R_{\,ab} + S(P \times R)_{\,c(ab)} + e_{n(abc)}$$

	f	f	a	a
Fuente de variación	***a***	***b***	***c***	***n***
P_a	0	b	c	
R_b	a	0	c	
$P \times R_{ab}$	0	0	c	
$S(P \times R)_{c(ab)}$	1	1	1	
$Residual_{n(c(ab))}$	1	1	1	

Estimas de Cuadrados Medios (CM) y contraste para el cálculo de las F

Ejemplo:

$$X_{abcn} = \mu + P_a + R_b + P \times R_{ab} + S(P \times R)_{c(ab)} + e_{n(abc)}$$

	f	f	a	a
Fuente de variación	***a***	***b***	***c***	***n***
P_a	0	b	c	n
R_b	a	0	c	n
$P \times R_{ab}$	0	0	c	n
$S(P \times R)_{c(ab)}$	1	1	1	n
$Residual_{n(c(ab))}$	1	1	1	1

Estimas de Cuadrados Medios (CM) y contraste para el cálculo de las F

Reglas:

1. Identificar en la fila objetivo todos los componentes de variación (σ^2_x) que pertenecen a ese CM. Estos componentes son los términos de las filas que contengan el subíndice o subíndices de la fila objetivo.

2. Cada componente identificado se ha de multiplicar por el producto de todas las entradas que aparecen en de la fila de ese componente, omitiendo las columnas que contienen el(los) subíndice(s) de la fila objetivo.

3. En la fila que contiene el residual el multiplicador es 1.

Estimas de Cuadrados Medios (CM) y contraste para el cálculo de las F

Ejemplo:

$$X_{abcn} = \mu + P_a + R_b + P \times R_{ab} + S(P \times R)_{c(ab)} + e_{n(abc)}$$

Omitiremos las Columnas con los subíndices de la fila objetivo …

Fuente de variación	*a*	*b*	*c*	*n*	Estimador de los CM
P_a	0	b	c	n	
R_b	a	0	c	n	
$P \times R_{ab}$	0	0	c	n	
$S(P \times R)_{c(ab)}$	1	1	1	n	$\sigma^2_e + n\sigma^2_{c(ab)}$
$Residual_{n(c(ab))}$	1	1	1	1	σ^2_e

… y las Filas que NO contienen los subíndices de la fila objetivo

Fila objetivo

Estimas de Cuadrados Medios (CM) y contraste para el cálculo de las F

Ejemplo:

$$X_{abcn} = \mu + P_a + R_b + P \times R_{ab} + S(P \times R)_{c(ab)} + e_{n(abc)}$$

Omitiremos las Columnas con los subíndices de la fila objetivo …

Fuente de variación	*a*	*b*	*c*	*n*	Estimador de los CM
P_a	0	b	c	n	
R_b	a	0	c	n	
$P \times R_{ab}$	0	0	c	n	
$S(P \times R)_{c(ab)}$	1	1	1	n	$\sigma^2_e + n\sigma^2_{c(ab)}$
$Residual_{n(c(ab))}$	1	1	1	1	σ^2_e

… y las Filas que NO contienen los subíndices de la fila objetivo

Fila objetivo

Estimas de Cuadrados Medios (CM) y contraste para el cálculo de las F

Ejemplo:

$$X_{abcn} = \mu + P_a + R_b + P \times R_{ab} + S(P \times R)_{c(ab)} + e_{n(abc)}$$

Omitiremos las Columnas con los subíndices de la fila objetivo…

Fuente de variación	*a*	*b*	*c*	*n*	Estimador de los CM
P_a	0	b	c	n	
R_b	a	0	c	n	
$P \times R_{ab}$	0	0	c	n	$\sigma^2_e + n\sigma^2_{c(ab)} + cn\sigma^2_{ab}$
$S(P \times R)_{c(ab)}$	1	1	1	n	$\sigma^2_e + n\sigma^2_{c(ab)}$
$\text{Residual}_{n(c(ab))}$	1	1	1	1	σ^2_e

… y las Filas que NO contienen los subíndices de la fila objetivo

Fila objetivo

Estimas de Cuadrados Medios (CM) y contraste para el cálculo de las F

Ejemplo:

$$X_{abcn} = \mu + P_a + R_b + P \times R_{ab} + S(P \times R)_{c(ab)} + e_{n(abc)}$$

Omitiremos las Columnas con los subíndices de la fila objetivo…

Fuente de variación	*a*	*b*	*c*	*n*	Estimador de los CM
P_a	0	b	c	n	
R_b	a	0	c	n	
$P \times R_{ab}$	0	0	c	n	$\sigma^2_e + n\sigma^2_{c(ab)} + cn\sigma^2_{ab}$
$S(P \times R)_{c(ab)}$	1	1	1	n	$\sigma^2_e + n\sigma^2_{c(ab)}$
$Residual_{n(c(ab))}$	1	1	1	1	σ^2_e

Fila objetivo

… y las Filas que NO contienen los subíndices de la fila objetivo

Estimas de Cuadrados Medios (CM) y contraste para el cálculo de las F

Ejemplo:

$$X_{abcn} = \mu + P_a + R_b + P \times R_{ab} + S(P \times R)_{c(ab)} + e_{n(abc)}$$

Fuente de variación	***a***	***b***	***c***	***n***	**Estimador de los CM**
P_a	0	b	c	n	$\sigma^2_e + n\sigma^2_{c(ab)} + bcn\sigma^2_a$
R_b	a	0	c	n	$\sigma^2_e + n\sigma^2_{c(ab)} + acn\sigma^2_b$
$P \times R_{ab}$	0	0	c	n	$\sigma^2_e + n\sigma^2_{c(ab)} + cn\sigma^2_{ab}$
$S(P \times R)_{c(ab)}$	1	1	1	n	$\sigma^2_e + n\sigma^2_{c(ab)}$
$\text{Residual}_{n(c(ab))}$	1	1	1	1	σ^2_e

Estimas de Cuadrados Medios (CM) y contraste para el cálculo de las F

Ejemplo:

$$X_{abcn} = \mu + P_a + R_b + P \times R_{ab} + S(P \times R)_{c(ab)} + e_{n(abc)}$$

Fuente de variación	*a*	*b*	*c*	*n*	Estimador de los CM	F-ratio versus
P_a	0	b	c	n	$\sigma^2_e + n\sigma^2_{c(ab)} + bcn\sigma^2_a$	
R_b	a	0	c	n	$\sigma^2_e + n\sigma^2_{c(ab)} + acn\sigma^2_b$	
$P \times R_{ab}$	0	0	c	n	$\sigma^2_e + n\sigma^2_{c(ab)} + cn\sigma^2_{ab}$	
$S(P \times R)_{c(ab)}$	1	1	1	n	$\sigma^2_e + n\sigma^2_{c(ab)}$	
$Residual_{n(c(ab))}$	1	1	1	1	σ^2_e	

Ahora debemos encontrar el denominador que ayude a eliminar las variabilidades debidas a otras fuentes que no sean la propia de la fila

$$F_{Fila} = \frac{\dfrac{SC_{Fila}}{gl_{Fila}}}{\dfrac{SC_{??}}{gl_{??}}}$$

Estimas de Cuadrados Medios (CM) y contraste para el cálculo de las F

Ejemplo:

$$X_{abcn} = \mu + P_a + R_b + P \times R_{ab} + S(P \times R)_{c(ab)} + e_{n(abc)}$$

Fuente de variación	*a*	*b*	*c*	*n*	Estimador de los CM	F-ratio versus
P_a	0	b	c	n	$[\sigma^2_e + n\sigma^2_{c(ab)}] + bcn\sigma^2_a$	**$S(P \times R)_{c(ab)}$**
R_b	a	0	c	n	$\sigma^2_e + n\sigma^2_{c(ab)} + acn\sigma^2_b$	
$P \times R_{ab}$	0	0	c	n	$\sigma^2_e + n\sigma^2_{c(ab)} + cn\sigma^2_{ab}$	
$S(P \times R)_{c(ab)}$	1	1	1	n	$[\sigma^2_e + n\sigma^2_{c(ab)}]$	
$Residual_{n(c(ab))}$	1	1	1	1	σ^2_e	

Cada F-ratio tendrá su denominador, tendremos que encontrar todos los denominadores.

$$F_{\Pr ot} = \frac{\dfrac{SC_{\Pr ot}}{gl_{\Pr ot}}}{\dfrac{SC_{Sitio(PxR)}}{gl_{Sitio(PxR)}}}$$

Estimas de Cuadrados Medios (CM) y contraste para el cálculo de las F

Ejemplo:

$$X_{abcn} = \mu + P_a + R_b + P \times R_{ab} + S(P \times R)_{c(ab)} + e_{n(abc)}$$

Fuente de variación	*a*	*b*	*c*	*n*	Estimador de los CM	F-ratio versus
P_a	0	b	c	n	$\sigma^2_e + n\sigma^2_{c(ab)} + bcn\sigma^2_a$	$S(P \times R)_{c(ab)}$
R_b	a	0	c	n	$\sigma^2_e + n\sigma^2_{c(ab)} + acn\sigma^2_b$	$S(P \times R)_{c(ab)}$
$P \times R_{ab}$	0	0	c	n	$\sigma^2_e + n\sigma^2_{c(ab)} + cn\sigma^2_{ab}$	
$S(P \times R)_{c(ab)}$	1	1	1	n	$\sigma^2_e + n\sigma^2_{c(ab)}$	
$Residual_{n(c(ab))}$	1	1	1	1	σ^2_e	

Estimas de Cuadrados Medios (CM) y contraste para el cálculo de las F

Ejemplo:

$$X_{abcn} = \mu + P_a + R_b + P \times R_{ab} + S(P \times R)_{c(ab)} + e_{n(abc)}$$

Fuente de variación	*a*	*b*	*c*	*n*	Estimador de los CM	F-ratio versus
P_a	0	b	c	n	$\sigma^2_e + n\sigma^2_{c(ab)} + bcn\sigma^2_a$	$S(P \times R)_{c(ab)}$
R_b	a	0	c	n	$\sigma^2_e + n\sigma^2_{c(ab)} + acn\sigma^2_b$	$S(P \times R)_{c(ab)}$
$P \times R_{ab}$	0	0	c	n	$\sigma^2_e + n\sigma^2_{c(ab)} + cn\sigma^2_{ab}$	$S(P \times R)_{c(ab)}$
$S(P \times R)_{c(ab)}$	1	1	1	n	$\sigma^2_e + n\sigma^2_{c(ab)}$	
$\text{Residual}_{n(c(ab))}$	1	1	1	1	σ^2_e	

$$F_{PxR} = \frac{\dfrac{SC_{PxR}}{gl_{PxR}}}{\dfrac{SC_{Sitio(PxR)}}{gl_{Sitio(PxR)}}}$$

Estimas de Cuadrados Medios (CM) y contraste para el cálculo de las F

Ejemplo:

$$X_{abcn} = \mu + P_a + R_b + P \times R_{ab} + S(P \times R)_{c(ab)} + e_{n(abc)}$$

Fuente de variación	*a*	*b*	*c*	*n*	Estimador de los CM	F-ratio versus
P_a	0	b	c	n	$\sigma^2_e + n\sigma^2_{c(ab)} + bcn\sigma^2_a$	$S(P \times R)_{c(ab)}$
R_b	a	0	c	n	$\sigma^2_e + n\sigma^2_{c(ab)} + acn\sigma^2_b$	$S(P \times R)_{c(ab)}$
$P \times R_{ab}$	0	0	c	n	$\sigma^2_e + n\sigma^2_{c(ab)} + cn\sigma^2_{ab}$	$S(P \times R)_{c(ab)}$
$S(P \times R)_{c(ab)}$	1	1	1	n	$\sigma^2_e + n\sigma^2_{c(ab)}$	$Residual_{n(c(ab))}$
$Residual_{n(c(ab))}$	1	1	1	1	σ^2_e	

Una vez encontrados los denominadores, optimizaremos la interacción de mayor nivel entre factores FIJOS.
Intentaremos aumentar los GL del denominador de dicha F-ratio

$$F_{PxR} = \frac{\dfrac{SC_{PxR}}{gl_{PxR}}}{\dfrac{SC_{Sitio(PxR)}}{gl_{Sitio(PxR)}}}$$

Contrastes para el cálculo de las F

$$X_{abcn} = \mu + P_a + R_b + P\times R_{ab} + S(P\times R)_{c(ab)} + e_{n(abc)}$$

Cada fuente de variabilidad tiene su Contraste de Hipótesis, su hipótesis científica, y su $F_{m,n}$ (donde m → GL Numerador, y n → GL Denominador)

$$F_P = \frac{CM_P}{CM_{S(PxR)}}$$

$$F_{PxR} = \frac{CM_{P\,x\,R}}{CM_{S(PxR)}}$$

$$F_R = \frac{CM_R}{CM_{S(PxR)}}$$

$$F_{S(PxR)} = \frac{CM_{S(PxR)}}{CM_{Residual}}$$

Contrastes para el cálculo de las F

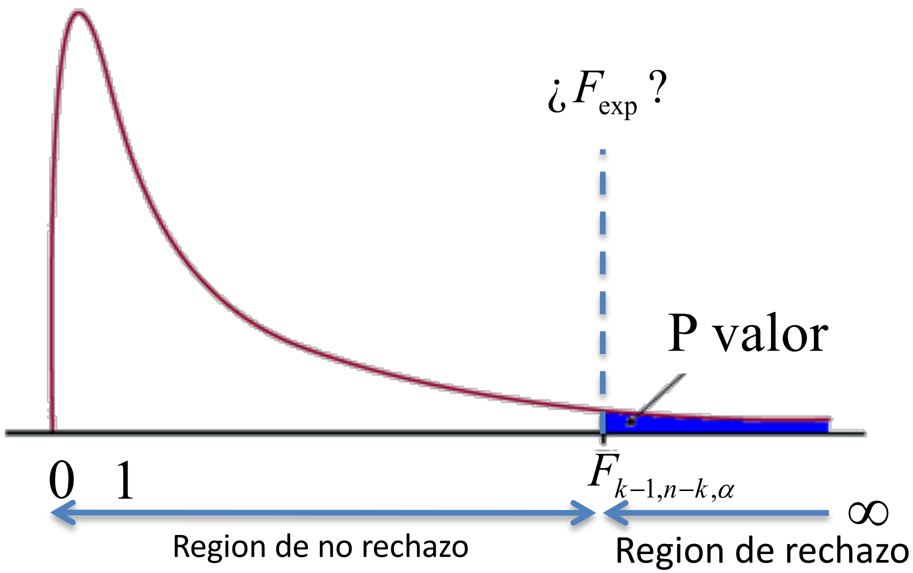
$$X_{abcn} = \mu + P_a + R_b + P \times R_{ab} + S(P \times R)_{c(ab)} + e_{n(abc)}$$

$$F_P = \frac{CM_P}{CM_{S\,(PxR)}}$$

$$F_{P\,xR} = \frac{\dfrac{SC_{P\,xR}}{gl_{P\,xR}}}{\dfrac{SC_{Sitio(PxR)}}{gl_{Sitio(PxR)}}}$$

$$F_R = \frac{CM_R}{CM_{S\,(PxR)}}$$

Distribución F



$$F_{S(PxR)} = \frac{CM_{S\,(PxR)}}{CM_{Residual}}$$

Contrastes para el cálculo de las F

$$X_{abcn} = \mu + P_a + R_b + P \times R_{ab} + S(P \times R)_{c(ab)} + e_{n(abc)}$$

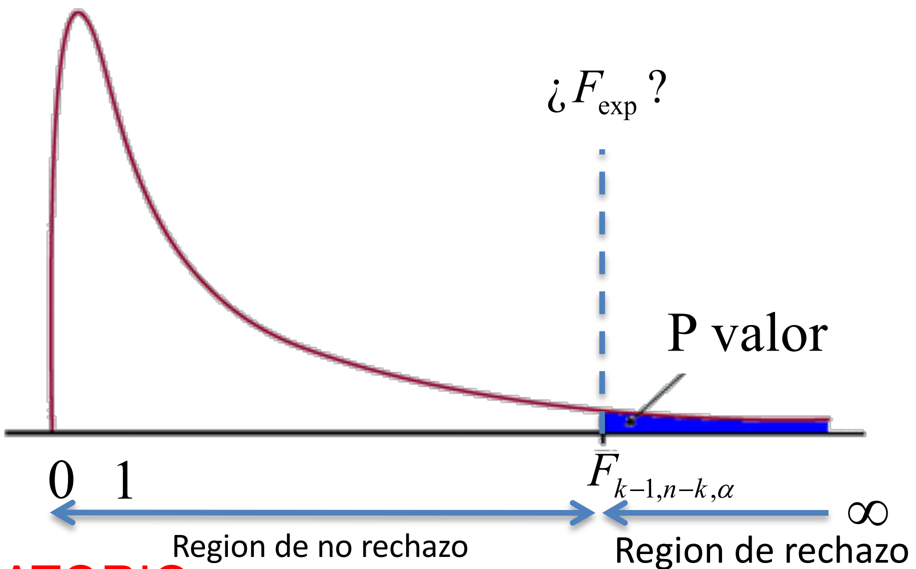
Optimizaremos la interacción de mayor nivel entre Factores FIJOS

$$F_{PxR} = \frac{\dfrac{SC_{PxR}}{gl_{PxR}}}{\dfrac{SC_{Sitio(PxR)}}{gl_{Sitio(PxR)}}}$$

Debemos aumentar la F para así maximizar la potencia del test F.

En este modelo, aumentando el número de Sitios (dado que es el DENOMINADOR de la F) se optimiza el experimento.

Distribución F



GL Sitio(PxrR) = pr(s – 1)
donde p=niveles de Protección
r=niveles de Región
s= niveles de Sitio

Se debería aumentar los niveles del factor ALEATORIO

Estrategia para el **análisis de ANOVA Multifactorial** o diseños mixtos

- Establecer los factores partir de la hipótesis científicas subyacentes como:
  - FIJO/ALEATORIO
  - ORTOGONAL/ANIDADO
- Verificar niveles de los factores y calcular Grados de Libertad
- Incluir en el modelo ANOVA todos los factores e interacciones posibles
- Calcular los Estimadores de MC
- Determinar los denominadores de cada F
- Revisar los requisitos para el modelo de ANOVA
- Si el diseño es balanceado, usaremos el paquete GAD, sino modificaremos los cálculos para las F "a mano" y calcularemos el p-valor para cada F
- Hacer gráficos necesarios para ver las posibles interacciones de mayor nivel*
- Usar los **gráficos** y la **tabla de ANOVA** correcta para **explicar los resultados**

Además:

- **Estudiar "*a posteriori*" (subconjuntos homogéneos) según las Prioridades:**

  1) Si es significativa la Interacción de factores FIJOS de mayor nivel
  2) Si no se cumple el (1), las de siguiente menor nivel
  3) Si no se cumple el (2), los factores FIJOS significativos.

Construcción de modelos mixtos: checklist

1. Definir los modelos, hipótesis e hipótesis nulas.
2. Definir las poblaciones y variables a muestrear.
3. Determinar cuáles son los factores relevantes (a partir de las hipótesis).
4. Definir si los factores son fijos o aleatorios (Hip. subyacentes e implicaciones).
5. Definir el nivel de replicación, de modo que:
   (a) sean posibles las generalizaciones
   (b) se asegure que no hay pseudo-replicación
6. Revisar amenazas a la independencia.
7. Revisar las implicaciones prácticas de factores anidados/ortogonales.
8. Calcular los grados de libertad (gl), los estimadores de los Cuadrados Medios (CM) y los contrastes apropiados para el cálculo de las F.
9. Determinar la potencia del análisis y optimizar el muestreo.
10. Si es necesario, re-diseñar el experimento.

Y recuerda, no hay limite para el análisis…

El objetivo es tener el control del diseño experimental para optimizar el análisis en base a las hipótesis científicas subyacentes

"Tienes que esforzarte para hacer las cosas mas simples, pero vale la pena, cuando lo consigues puedes mover montañas" (S. Jobs)

Universitat d'Alacant
Universidad de Alicante
Dept. of Marine Science and Applied Biology
Prof. Jose Jacobo Zubcoff