FACULTAD DE BIOLOGÍA

UNIVERSIDAD MICHOACANA DE SAN NICOLÁS DE HIDALGO

DISEÑO Y ANÁLISIS EXPERIMENTAL

ANDEVA

El análisis de varianza, con siglas ANDEVA es la técnica que compara las medias de 3 ó más poblaciones.

Las principales premisas de ANDEVA son:

1) Poblaciones normalmente distribuidas
2) Varianzas poblacionales iguales ó casi iguales.
3) Muestreo aleatorio extraído de las poblaciones

CONCEPTOS PRELIMINARES:

UNIDAD EXPERIMENTAL:

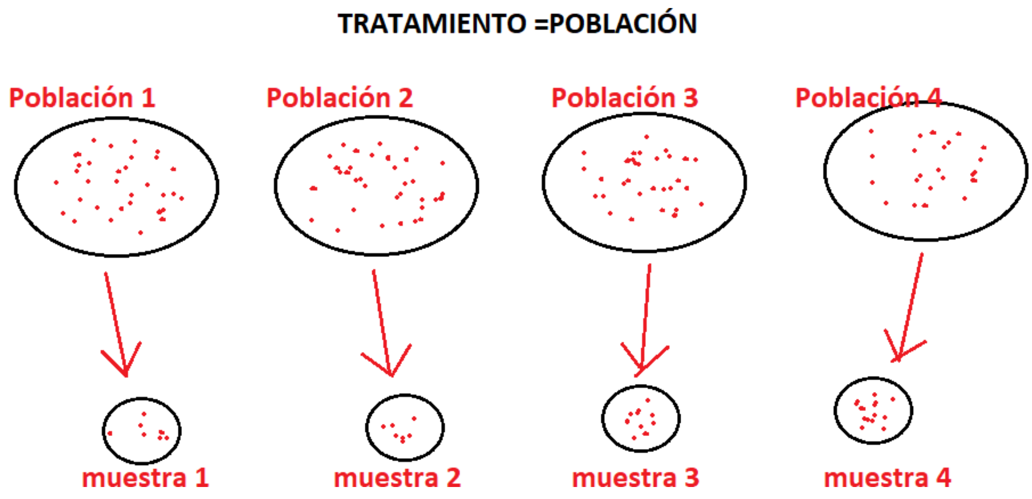
Un elemento cualquiera de los muchos que se someten a la comparación. Cada elemento viene a estar representado por un solo dato en las tablas de datos.

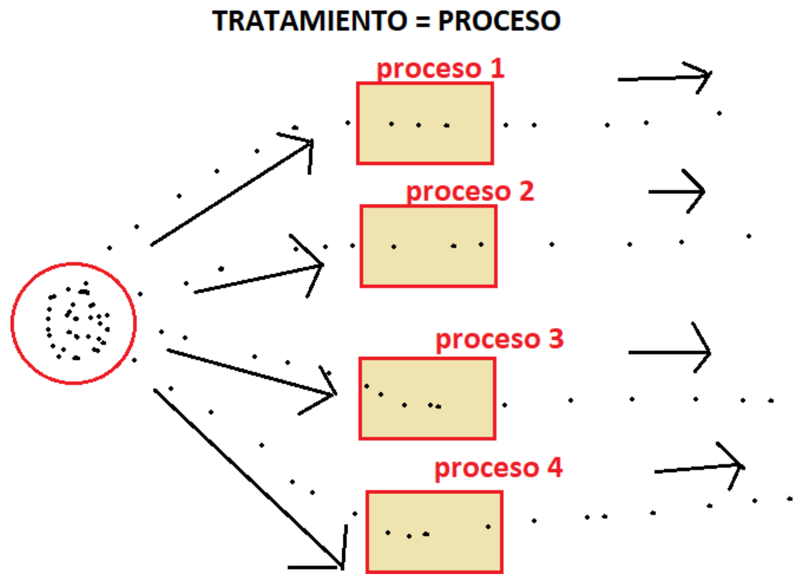
TRATAMIENTOS:

El significado de tratamiento tiene 2 acepciones, según el tipo de experimento:

1. (*Tratamiento*= *Población*). Cuando en ANDEVA comparamos poblaciones, de cada población se extrae una muestra. La comparación de poblaciones se efectúa a través de la comparación entre las muestras respectivas. En estas condiciones, a cada población se le denomina "*tratamiento*". Es común que en dichos experimentos se busque identificar el tratamiento con medidas mayores (o menores) de un atributo (variable).
2. (*Tratamiento*= *Proceso*). A veces los tratamientos son los procesos distintivos a los que se someten y son "tratadas" las unidades experimentales. Las unidades experimentales pueden sufrir daños o alteraciones durante el proceso. ". Es común que en dichos experimentos se busque identificar el tratamiento de mayor (o menor) eficacia o calidad.

En cualquier caso, basta decir que *los tratamientos son los entes que se comparan entre sí*, sea para determinar cuál es más efectivo, cual arroja resultados mejores, cuál tiene medidas mayores(o menores) de un atributo o variable.





En ANDEVA se utilizan varios diseños, el diseño más simple en ANDEVA es:

DISEÑO COMPLETAMENTE ALEATORIZADO

Cuando "*Tratamiento= Población*" este diseño se suele aplicar cuando las unidades experimentales son homogéneas y/o cuando los efectos ambientales (bajo los cuales se comparan los tratamientos) son similares. El nombre deriva a razón de que cada muestra se

extrae mediante muestreo totalmente **aleatorio** de la población respectiva (es lo deseable, pero no siempre se extrae de esta manera, a veces por limitaciones prácticas).

Cuándo "*Tratamiento= Proceso*" cada unidad experimental de las disponibles para el experimento se asigna **al azar** a cualquiera de los tratamientos.

Sus principales ventajas son:

1. Es muy flexible, ya que se puede utilizar cualquier número de tratamientos

y de réplicas (réplicas es el número de unidades experimentales en cada tratamiento).
2. El análisis estadístico es muy sencillo.
3. Aún cuando los datos de algunos tratamientos se hayan perdido, el análisis sigue siendo posible.

La principal objeción contra este diseño es un menor grado de precisión, cuando llega a haber cabida al uso de otro diseño que también pueda utilizarse en el experimento (bloques, cuadrado latino, etc.)

Resolución del diseño (de forma manual)

1. Plantear:

H_o: Los procesos no difieren significativamente en

efectividad. ó

Las medias de las poblaciones son todas iguales.

H_a: Los procesos difieren significativamente en efectividad.

ó

Las medias de las poblaciones no son todas iguales.

2. Elegir α . Si rechazamos H_o (si aceptamos H_a) al finalizar la prueba, entonces la probabilidad de haber tomado una decisión errónea, es decir la probabilidad de que H_o sea en realidad falsa, es una probabilidad menor que α .

En particular le asignaremos a α el valor de 0.05 = 5%. Es el valor más usado.

3. Calcular los tamaños de las muestras (n_j) y las sumas de los tratamientos ($X_{.j}$) en la siguiente tabla de datos:

Tratamiento 1	**Tratamiento 2**	**.........**	**Tratamiento k**
X_{11}	X_{12}	....	X_{1k}
X_{21}	X_{22}	....	X_{2k}
X_{31}	X_{32}	....	K_{3k}
....	....	....	....
....	....	....	....
$X_{n1,1}$	$X_{n2,2}$	....	$X_{nk,k}$
n_1 (elementos en la muestra)	n_2 (elementos en la muestra)	....	n_K (elementos en la muestra)
$X_{\bullet 1}$ (suma de la muestra)	$X_{\bullet 2}$ (suma de la muestra)	……..	$X_{\bullet k}$ (suma de la muestra)

Donde:

$X_{\bullet\bullet} = X_{\bullet .1} + X_{\bullet .2} + X_{\bullet .3} + ...$

$N = n_1 + n_2 + n_3 + \ldots\ldots$

k=número de tratamientos (o sea número de columnas en la tabla anterior)

4. Construir la siguiente tabla ANDEVA :

FUENTE DE VARIACIÓN	**GRADOS DE LIBERTAD**	**SUMA DE CUADRADOS (SC)**	**CUADRADOS MEDIOS (CM)**
Tratamientos	k-1	$X_{\bullet 1}^2/n_1 + X_{\bullet 2}^2/n_2 + X_{\bullet 3}^2/n_3 + \ldots\ldots - X_{\bullet\bullet}^2/N$	SC_{trat} / k-1
error	N-k	$X_{11}^2 + X_{12}^2 + X_{13}^2 + \ldots\ldots + X_{21}^2 + X_{22}^2 + X_{23}^2 + \ldots + X_{31}^2 + X_{32}^2 + X_{33}^2 + \ldots\ldots - (X_{\bullet 1}^2/n_1 + X_{\bullet 2}^2/n_2 + X_{\bullet 3}^2/n_3 + \ldots)$	SC_{error} / N-k

Además:

$F_{calc.} = CM_{trat} / CM_{error}$

$F_{teórica} = F_{1-\alpha,\ k-1,\ N-k}$

La $F_{teórica}$ es llamada F de Fisher, la puedes consultar en internet o en los sigs. enlaces: http://www.normaltable.com/ftable.html

5. Si $F_{calc.}$ mayor que $F_{teórica}$ rechazar H_o. (aceptar H_a) Si $F_{calc.}$ menor o igual que $F_{teórica}$ aceptar H_o.

Ejemplo:

En un experimento, montado mediante un diseño totalmente aleatorizado, se tuvo como objetivo comparar el consumo de oxígeno entre 4 variedades de salmones canadienses de aproximadamente el mismo peso (2 kgs.) y sometidos a condiciones de *pH de 8.4 y oxígeno disuelto de 5.6 ppm*. Se pretende determinar si existen diferencias entre las variedades en cuanto al oxígeno que consumen esos peces. La siguiente tabla de datos presenta las variedades consideradas (tratamientos) y los datos representan el consumo de oxígeno (µl/h) registrados por cada salmón. Comparar los tratamientos.

Variedad 1	Variedad 2	Variedad 3	Variedad 4

15	18	55	50
16	20	51	47
12			

1. Plantear:

   H_o: Las medias de las poblaciones son todas

   iguales. H_a: Las medias de las poblaciones no son

   todas iguales.

2. $\alpha = 0.05$
3. Calcular los tamaños de las muestras (n_j) y las sumas de los tratamientos ($X_{.j}$) en la siguiente tabla de datos:

Variedad 1	Variedad 2	Variedad 3	Variedad 4
15	18	55	50
16	20	51	47
12			

$X_{\bullet\bullet}$= 43+ 38+ 106+ 50 =.

237 n_1= 3, n_2=2, n_3=2, n_1=1

N= 8

k=4

4. Construir la siguiente tabla ANDEVA :

FUENTE DE VARIACIÓN	**GRADOS DE LIBERTAD**	**SUMA DE CUADRADOS (SC)**	**CUADRADOS MEDIOS (CM)**
Tratamientos	k-1= 4-1= 3	$X_{\bullet 1}^2/n_1 + X_{\bullet 2}^2/n_2 + X_{\bullet 3}^2/n_3 + \ldots - X_{\bullet\bullet}^2/N$ $= 43^2/3 + 38^2/2 + 106^2/2 + 50^2/1 - 237^2/8 =$	SC $_{trat}$ / k-1 =2435.2 / 3 = 811.7

		9456.3-7021.1= 2435.2	
Error	N-k= 8-4= 4	$X_{11}^2 + X_{12}^2 + X_{13}^2 + \dots + X_{21}^2 + X_{22}^2 + X_{23}^2 + \dots + X_{31}^2 + X_{32}^2 + X_{33}^2 + \dots$ $- (X_{\cdot 1}^2/n_1 + X_{\cdot 2}^2/n_2 + X_{\cdot 3}^2/n_3 + ..)$ $= 15^2 + 18^2 + 55^2 + 50^2.$ $+16^2 + 20^2 + 51^2 + 12^2$ $- (43^2/3 + 38^2/2 + 106^2/2 + 50^2/1)$ $= 9475 - 9456.3 = 18.7$	$SC_{error} / N\text{-}k$ $=18.7 / 4$ $= 4.675$

$F_{calc.} = CM_{trat} / CM_{error}$

$F_{calc.} = 811.7 / 4.675 = 173.6$ F teórica $= F_{1-\alpha,\ k-1,\ N-k}$

$F_{0.95,\ 3,\ 4}$

$= 6.59$

http://www.normaltable.com/ftable.html

5. Se acepta Ha

PRUEBA DE SCHEFFÉ

Esta prueba, como algunas otras, identifica los pares de tratamientos que difieren significativamente en cualquier diseño de ANDEVA donde se ha concluido por aceptar H_a.

Pasos sugeridos:

1 Calcular la media armónica n de los tamaños de las muestras:

$$n = k \; /(1/ n_1 + 1/ n_2 + 1/ n_3 + \ldots\ldots.)$$

2 Calcular el estadígrafo S:

$$S = \sqrt{[\; 2\,(k-1) \;\; F_{1-\alpha,\; k-1,\; gl\ error} \; CM_{error} \;\; / \;\; n \;]}$$

3. Calcular las medias de las muestras y las diferencias (positivas) entre todas las combinaciones posibles de pares de medias, nos podemos auxiliar en la siguiente tabla:

	$\bar{x}1$	$\bar{x}2$	$\bar{x}3$	$\bar{x}4$	$\bar{x}5$
$\bar{x}1$					
$\bar{x}2$					
$\bar{x}3$					
$\bar{x}4$					
$\bar{x}5$					

4. Las diferencias de medias muestrales que resultan mayores que S corresponden a los pares de tratamientos que difieren significativamente.

Notas para la interpretación:

Definiciones:

Sea M_A la media del tratamiento A y M_B la media del tratamiento B. Más específicamente:

Si "*Tratamiento= Población*" entonces M_A (la media del tratamiento A) es la media de la población A.

Si "*Tratamiento= Proceso*" entonces M_A (la media del tratamiento A) es la eficacia (o ineficacia, según el contexto) del proceso A.

La interpretación del paso 4 sigue las siguientes reglas:

Regla 1. Si el tratamiento A resultó significativamente distinto al tratamiento B y si por otro lado se tiene que la media muestral de A es mayor que la media muestral de B entonces existe evidencia suficiente para concluir que M_A es mayor que M_B.

Regla 2. Si el tratamiento A no resultó significativamente distinto al tratamiento B y aunque se tenga que la media muestral de A es mayor que la media muestral de B entonces no existe evidencia suficiente para concluir que M_A sea mayor que M_B.

Cuando "*Tratamiento= Población",* si llegásemos a determinar que M_A sea mayor que M_B entonces estaríamos concluyendo que la media de la población A es mayor que la media de la población B. El concepto de "media" tiene, en este contexto, repercusiones bastante útiles, pues decir que *la media de la población A es mayor que la media de la población B* es decir también que *la población A tiene muchos (o la mayoría) de sus elementos con mayores medidas que los elementos de la población B*.

Cuando "*Tratamiento= Proceso"*, si llegásemos a determinar que M_A sea mayor que M_B entonces estaríamos concluyendo que la eficacia (o ineficacia) del proceso A es mayor que la eficacia (o ineficacia) del proceso B.

Ejemplo:

Aplicando *Scheffe* al problema de los salmones:

a) Calcular la media armónica n de los tamaños de las muestras:

$$n = k \;/\; (1/n_1 + 1/n_2 + 1/n_3 + \ldots)$$

$$n = k \;/\; (1/3 + 1/2 + 1/2 + 1/1)$$

$$n = 4 \;/\; 2.33$$

$$n = 1.71$$

b) Calcular el estadígrafo S:

$$S = \sqrt{[\; 2\,(k-1)\; F_{1-\alpha,\, k-1,\, gl\ error}\; CM_{error} \;/\; n \;]}$$

$$S = \sqrt{[\; 2\;(3)\; F_{0.95,\; 3,\; 4}\; 4.675 \;/\; 1.71 \;]}$$

$$S = \sqrt{[\; 2\;(3)\; (6.59)\; 4.675 \;/\; 1.71 \;]}$$

$$S = \sqrt{[\; 2\;(3)\; (6.59)\; 4.675 \;/\; 1.71 \;]}$$

$$S = \sqrt{[\; 184.84 \;/\; 1.71 \;]}$$

$$S = \sqrt{108.09}$$

S= 10.39

c) medias muestrales:

$\bar{x}_1 = 14.33$	$\bar{x}_2 = 19$	$\bar{x}_3 = 53$	$\bar{x}_4 = 48.5$

	$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$	$\bar{x}_4$
$\bar{x}_1$		4.67	**38.67**	**34.17**
$\bar{x}_2$			**34**	**10.5**
$\bar{x}_3$				4.5
$\bar{x}_4$				

d)

Trats. 1 y 3 difieren signif.

Trats. 1 y 4 difieren signif.

Trats. 2 y 3 difieren signif.

Trats. 2 y 4 difieren signif.

Trats. 1 y 2 no difieren signif. Trats. 3 y 4 no difieren signif.

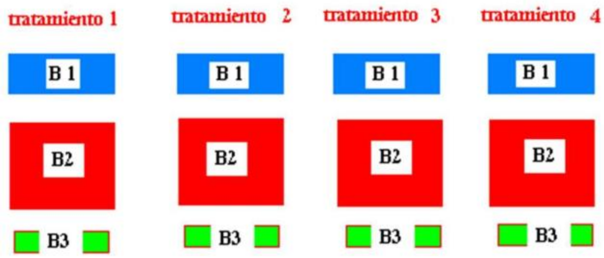
Como las medias muestrales eran:

$\bar{x}_1=14.33$	$\bar{x}_2=19$	$\bar{x}_3=53$	$\bar{x}_4=48.5$

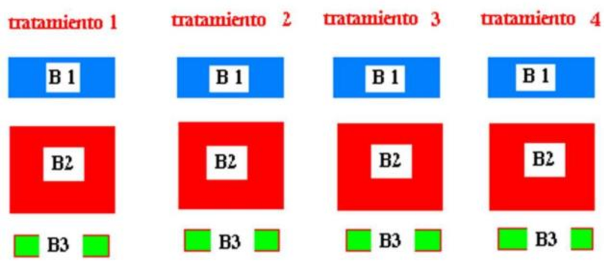
Existe evidencia suficiente para concluir que M_3 sea mayor que M_1. Existe evidencia suficiente para concluir que M_3 sea mayor que M_2. No existe evidencia suficiente para concluir que M_2 sea mayor que M_1. No existe evidencia suficiente para concluir que M_3 sea mayor que M_4.

DISEÑO DE BLOQUES ALEATORIZADOS

Cuando "tratamiento=población" las unidades experimentales de cada muestra se clasifican en grupos homogéneos llamados bloques de forma tal que los bloques sean del mismo tamaño en todas las muestras. (se pueden desechar unidades sobrantes).



Cuando "tratamiento= proceso" las unidades disponibles para el experimento se clasifican en grupos homogéneos (bloques) y luego las unidades de cada bloque se asignan al azar en cualquiera de los tratamientos, con la condición de que al asignarse todas, los bloques sean del mismo tamaño en todos los tratamientos. (se pueden desechar unidades sobrantes).



El objetivo de este diseño es evitar que en la comparación entre tratamientos intervenga la variabilidad entre bloques. En experimento con animales los bloques se pueden conformar a partir de características tales como sexo, edad, peso inicial, estadio de lactancia. Los bloques también pueden representar laboratorios, días, operadores, etc. Puede utilizarse cualquier

número de bloques y de tratamientos. Este diseño, cuando es posible implementarse, puede ser más preciso que el diseño totalmente aleatorizado.

Resolución del diseño (de forma manual)

1. Plantear:

H_o: Los procesos no difieren significativamente en efectividad.

Ó

Las medias de las poblaciones son todas iguales.

H_a: Los procesos difieren significativamente en efectividad.

Ó

Las medias de las poblaciones no son todas

2. Elegir α. Si rechazamos H_o al finalizar la prueba, entonces la probabilidad de haber tomado una decisión errónea, es decir la probabilidad de que H_o sea en realidad cierta, es una probabilidad menor que α .

En particular le asignaremos a α el valor de 0.05 = 5%

3. Calcular las sumas de los tratamientos $X_{.j}$ y las sumas de los bloques $X_{i\,.}$ en la siguiente tabla

	Tratamiento 1	Tratamiento 2	....	Tratamiento k	SUMAS:
Bloque 1	X_{11}	X_{12}	...	X_{1k}	$X_{1\bullet}$
Bloque 2	X_{21}	X_{22}	....	X_{2k}	$X_{2\bullet}$
Bloque 3	X_{31}	X_{32}	...	K_{3k}	$X_{3\bullet}$
…..	....				
……	....				
Bloque n	$X_{n,1}$	$X_{n,2}$		$X_{n,k}$	$X_{n\bullet}$
SUMAS:	$X_{\bullet 1}$	$X_{\bullet 2}$	……..	$X_{\bullet k}$	$X_{\bullet\bullet}$

Considere que n es el número de bloques y k es el número de

tratamientos, Además:

$X_{\bullet\bullet} = X_{\bullet 1} + X_{\bullet 2} + X_{\bullet 3} + \ldots..$

N= k n (N es el total de datos)

4. Construir la siguiente tabla ANDEVA :

F.V.	**G. L.**	**S.C.**	**C.M.**
Tratamientos	k-1	$X_{\bullet 1}^2/n + X_{\bullet 2}^2/n + X_{\bullet 3}^2/n + \ldots\ldots. - X_{\bullet\bullet}^2/N$	SC_{trat} / (k-1)
Bloques	n-1	$X_{1\bullet}^2/k + X_{2\bullet}^2/k + X_{3\bullet}^2/k + \ldots\ldots - X_{\bullet\bullet}^2/N$	$SC_{bloques}$ / (n-1)
Error	(k-1) (n-1)	$SC_{total} - SC_{trat} - SC_{bloques}$	SC_{error} / (k-1) (n-1)
Total	N-1	$X_{11}^2 + X_{12}^2 + X_{13}^2 + \ldots\ldots. + X_{21}^2 + X_{22}^2 + X_{23}^2 + \ldots. + X_{31}^2 + X_{32}^2 + X_{33}^2 + \ldots\ldots..$	

		$- X_{..}^2 / N$	

Además:

$F_{calc.} = CM_{trat} / CM_{error}$

$F_{teórica} = F_{1-\alpha,\ k-1,\ (k-1)(n-1)}$

5. Si $F_{calc.}$ mayor que $F_{teórica}$ rechazar H_o. (o sea aceptar H_a) Si $F_{calc.}$ menor o igual que $F_{teórica}$ aceptar H_o.

Nota: si deseamos comparar la variabilidad entre bloques,

compare CM_{trat} / CM_{error} contra $F_{1-\alpha,\ n-1,\ (k-1)(n-1)}$

Ejemplo no resuelto: Se realiza un estudio del efecto de la luz sobre el crecimiento de los helechos. Puesto que las plantas crecen con velocidad distinta a edades diferentes, se controla esta variable mediante bloques. En el estudio se utilizan 4 plantas neonatas (plantas crecidas en la oscuridad durante cuatro días), cuatro plantas jóvenes (plantas crecidas en la oscuridad durante ocho días) y cuatro plantas más viejas (plantas crecidas en la oscuridad durante doce días). Resultaron los siguientes datos (el crecimiento viene dado en cronómetros cuadrados). Realizar la comparación.

Tratamientos → *bloques ↓*	*420 nm*	*460 nm*	*600 nm*	*720 nm*
Neonata	1412	1001	1027	1112
Joven	1217	929	839	1081
Adulta	954	689	741	797

DISEÑO FACTORIAL

En este diseño a cada uno de los dos conjuntos de tratamientos se le acostumbra decir “factores”, y a cada tratamiento se le llama “nivel”. Este diseño tiene como principal finalidad contrastar de forma simultánea los efectos de dos factores para probar si existe interacción entre ellos

La interacción se puede presentar de alguna de las siguientes dos modalidades: como un sinergismo ó una interferencia.

El “sinergismo” lo podemos identificar cuando es mayor los efectos de dos niveles que operan juntos que la suma de los efectos de los dos niveles cuando operan por separado. Por ejemplo, podríamos suponer que a determinada persona una taza de café por sí sola le robe una hora de sueño por la noche, y que un refresco de cola le reste dos horas de sueño. Pero si la persona consume ambas bebidas antes de dormir entonces las horas de sueño perdidas no serían 1 +2 =3 horas, sino digamos 5 las horas de sueño perdidas. En este ejemplo el refresco y el café “potencializan” sus efectos contra el sueño cuando operan juntos. El mismo fenómeno se presenta en la *gráfica A,* donde podríamos decir que la temperatura de 30 ºC asociada con la concentración de 40 mg /L potencializan sus efectos, y por tanto esa “combinación” de niveles determina una alta capacidad de mortandad en los peces.

La interferencia es lo opuesto al sinergismo. La “interferencia” la descubrimos cuando es menor los efectos de dos niveles que operan juntos que la suma de los efectos de los dos niveles cuando operan por separado. Un alumno pierde en promedio 10 sesiones mensuales de clase si trabaja extra-escolarmente medio tiempo, y pierde en promedio 5 sesiones mensuales si está casado; pero si está comprometido en ambos casos entonces no pierde 10+5=15 sesiones mensuales sino sólo 12. El trabajo y el matrimonio “inhiben” mutuamente sus efectos.

Una forma gráfica de descubrir la interacción es la siguiente:

Construya una gráfica donde en el eje horizontal se registren los niveles de un factor y el eje vertical las frecuencias. Para un nivel cualquiera del otro factor, ubique las medias de ese nivel, medidas en cada uno de los niveles del primer factor, una las medias con rectas. Repita el procedimiento para los demás niveles.

Luego repita todo el procedimiento en una nueva gráfica, pero invirtiendo los

factores. Entonces:

1. No existe interacción si las líneas son paralelas ó casi paralelas en ambas gráficas
2. Existe interacción si las líneas distan mucho de ser paralelas en una ó en ambas gráficas.

Resolución del diseño (de forma manual)

1. Plantear:

H_o: No existe interacción entre los factores.

H_a: Existe interacción entre los factores.

2. Elegir α

3. Calcular las sumas y tamaños en las combinaciones de celdas de la siguiente tabla:

	B_1	B_2	………………		
A_1	X_{111}	X_{121}	………………		
	X_{112}	X_{122}	……………..		
	……. …….	……. …….	…………….		
	N_{11} (tamaños)	N_{12} (tamaños)	…………….	$N_{1\bullet}$	
	$X_{11\bullet}$ (sumas)	$X_{12\bullet}$ (sumas)	…………….		$X_{1\bullet\bullet}$
A_2	X_{211}	X_{221}	…………..		
	X_{212}	X_{222}	…………….		
	……. …….	……. …….	…………..		
	N_{21} (tamaños)	N_{22} (tamaños)	…………….	$N_{2\bullet}$	
	$X_{21\bullet}$ (sumas)	$X_{22\bullet}$ (sumas)	…………….		$X_{2\bullet\bullet}$
………… ………… ………… ………… ………… …………					
$N_{\bullet J}$ $X_{\bullet J\bullet}$	$N_{\bullet 1}$ $X_{\bullet 1\bullet}$	$N_{\bullet 2}$ $X_{\bullet 2\bullet}$		N	$X_{\bullet\bullet\bullet}$

Donde:

A_I es un nivel cualquiera de

A B_J es un nivel cualquiera

de B X_{IJK} es cada dato

N_{IJ}. Es el número de datos en la intersección de los niveles A_I y

B_J X_{IJ}. Es la suma de datos en la intersección de los niveles A_I

y B_J

N_I. Es el número de datos en el nivel A_I, por ejemplo, $N_{1\bullet} = N_{11} + N_{12} + \ldots\ldots$

$N_{.J}$ es el número de datos en el nivel B_J

$X_{I..}$ es la suma de los datos en el nivel A_I, por ejemplo, $X_{1\bullet\bullet} = X_{11\bullet} + X_{12\bullet} + \ldots\ldots$

$X_{.J.}$ es la suma de datos en el nivel B_J, por ejemplo, $X_{\bullet 1\bullet} = X_{11\bullet} + X_{21\bullet} + \ldots\ldots$

N es el total de datos: $N = N_{\bullet 1} + N_{\bullet 2} + \ldots\ldots$

$X_{\ldots}$ es la suma de todos los datos: $X_{\bullet\bullet\bullet} = X_{\bullet 1\bullet} + X_{\bullet 2\bullet} + \ldots$

4. Construir la siguiente tabla ANDEVA:

FUENTE DE VARIACIÓN	**GRADOS DE LIBERTAD**	**SUMA DE CUADRADOS (SC)**	**CUADRADOS MEDIOS (CM)**
Filas (A)	I-1	$X_{1\bullet\bullet}^2/N_{1\bullet} + X_{2\bullet\bullet}^2/N_{2\bullet} + X_{3\bullet\bullet}^2/N_{3\bullet} + \ldots\ldots - X_{\bullet\bullet\bullet}^2/N$	SC_A / I-1
Columnas (B)	J-1	$X_{\bullet 1\bullet}^2/N_{\bullet 1} + X_{\bullet 2\bullet}^2/N_{\bullet 2} + X_{\bullet 3\bullet}^2/N_{\bullet 3} + \ldots\ldots - X_{\bullet\bullet\bullet}^2/N$	SC_B/ J-1
Interacción (AB)	(I-1)(J-1)	$SC_{SUBTOTAL} - SC_A - SC_B$	SC_{AB} / (I-1)(J-1)
Subtotal	IJ-1	$X_{11\bullet}^2/N_{11} + X_{12\bullet}^2/N_{12} + X_{13\bullet}^2/N_{13} + \ldots\ldots$ $X_{21\bullet}^2/N_{21} + X_{22\bullet}^2/N_{22} + X_{23\bullet}^2/N_{23} + \ldots$ $X_{31\bullet}^2/N_{31} + X_{32\bullet}^2/N_{32} + X_{33\bullet}^2/N_{33} + \ldots\ldots$ $+ \ldots\ldots - X_{\bullet\bullet\bullet}^2/N$	

Error Residual	N-IJ	$SC_{TOTAL} - SC_{SUBTOTAL}$	SC_{ERROR} / (N-IJ)
Total	N-1	$X_{111}^2 + X_{121}^2 + X_{112}^2 + \dots + X_{211}^2 + X_{221}^2 + X_{212}^2 + \dots + X_{311}^2 + X_{321}^2 + X_{312}^2 + \dots - X_{\cdots}^2/N$	

Además:

$F_{calc.}(AB) = CM_{AB} / CM_{error}$

$F_{teórica}(AB) = F_{1-\alpha, (I-1)(J-1), N-IJ}$

5. Entonces

Si $F_{calc.}$ mayor que $F_{teórica}$ rechazar H_o. (aceptar H_a) Si $F_{calc.}$ menor o igual que $F_{teórica}$ aceptar H_o.

Ejemplo.

Se realiza un experimento donde se registra el tiempo de sobrevivencia de una especie de pez sometida a distintas temperaturas y a distintas concentraciones de ión cianuro. Los datos (en minutos) son los siguientes, cada dato proviene de un pez distinto. Probar si existe interacción entre temperatura y concentración de cianuro.

	20 ºC	25 ºC	30 ºC
10 mg /L	10	12	15
	12	15	18
	12	15	18
	15	15	16
20 mg /L	8	10	12
	9	11	10
	8	12	12
	10	10	12
40 mg / L	5	5	3
	7	7	2
	7	7	3
	6	7	1

Si los efectos de cada temperatura y concentración se estimaran matemáticamente, éstos vienen a ser, aproximadamente:

-

nivel	*efecto*
10 mg /L	6
20 mg /L	3
40 mg /L	0
20 C	5
25 C	7
30 C	10

Y en realidad cada dato de la tabla se puede descomponer como una suma de términos, de la siguiente forma:

Potencialización:

Efecto de A y B juntos= Efecto de A + efecto de B+ efecto de la interacción (positivo).+ término aleatorio o intrínseco o inherente (pequeño, positivo o negativo)

Inhibición:

Efecto de A y B juntos= Efecto de A + efecto de B+ efecto de la interacción (negativo).+ término aleatorio o intrínseco o inherente (pequeño, positivo o negativo)

Si no hay interacción:

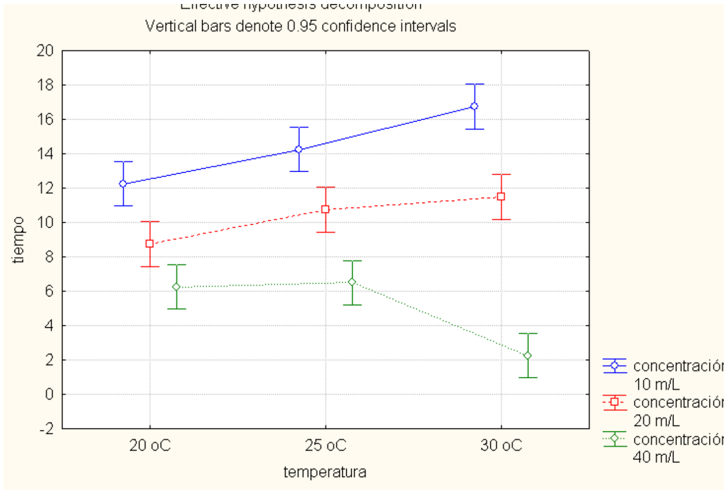
Efecto de A y B juntos= Efecto de A + efecto de B.+ término aleatorio o intrínseco o inherente (pequeño, positivo o negativo).

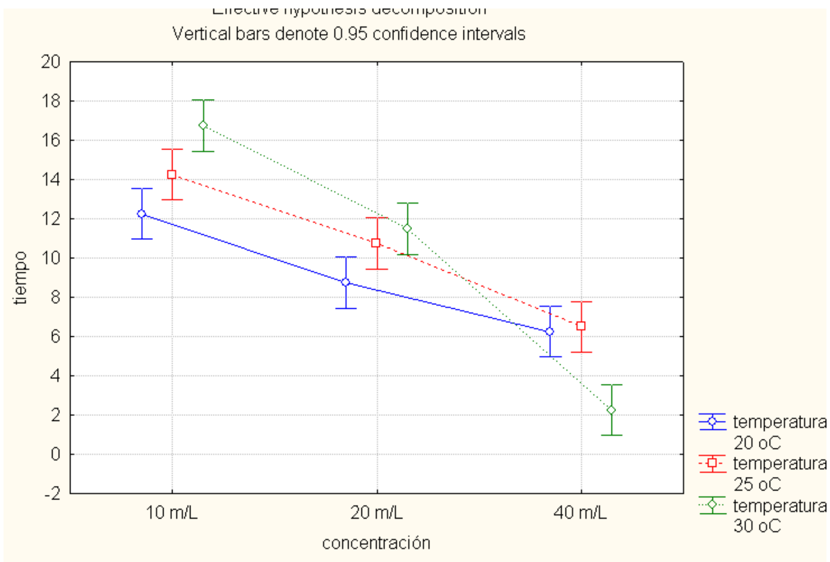
Así, el primer dato: 10 es descompuesto de la forma:

10= 5(efecto de 20 C) +6(efecto de 10 mg/L) + 0(no hay interacción en esa combinación de temp. y concentración) +(-1) (término intrínseco al pez)

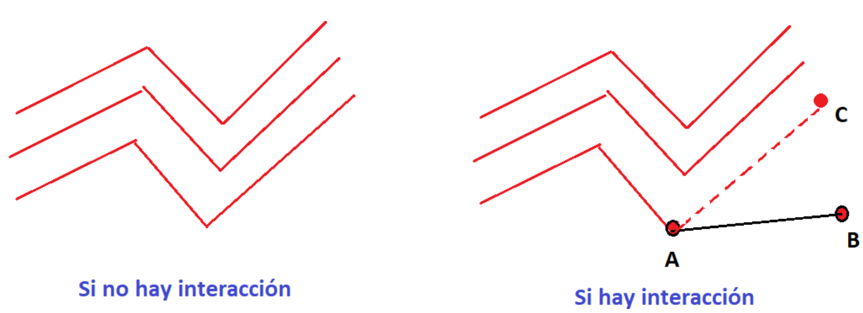
El 3, penúltimo dato de todos se descompone como:

3= 10(efecto de 20 C) +0(efecto de 10 mg/L) + -8(hay interacción de tipo inhibición en esa combinación de temp. y concentración) +1 (término intrínseco al pez) *Nota*: las gráficas antes referidas son:





Los software despliegan gráfica como las siguientes, las cuales nos conducen a ubicar dónde se encuentra alguna eventual interacción, identificando la "barrita rebelde "que no sigue la dirección del resto de barras de la gráfica desplegada:



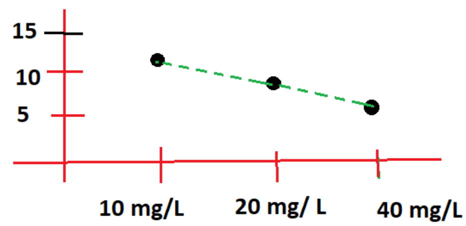
Cuando existe interacción entre 2 factores, es útil especificar el nivel de uno de los factores cuando se hagan comparaciones entre los niveles del otro factor. Supongamos que un investigador realiza las comparaciones entre las 3 concentraciones mediante un diseño totalmente aleatorizado en un ambiente donde la temperatura ambiente es de 20 º C. Posiblemente concluya que no existe diferencia significativa entre las 3 concentraciones.

10 mg/L	20 mg/L	40 mg/L
13	8	7
12	9	6
12	9	7
13	7	8
11	8	9
13	9	6

No existe evid. suf. para concluir que M1 mayor que M2

No.........M2 mayor que M3

No.........M1 mayor que M3



20º C

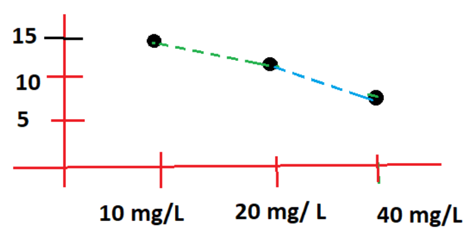
Si otro investigador realiza las comparaciones de las 3 concentraciones a una temperatura ambiente de 25 º C, posiblemente concluya que no existe diferencia significativa entre las 3 concentraciones.

10 mg/L	20 mg/L	40 mg/L
14	11	7
13	10	6
14	11	7
15	12	7
13	10	6
14	11	7

No existe evid. suf. para concluir que M1 mayor que M2

No.........M2 mayor que M3

No.........M1 mayor que M3

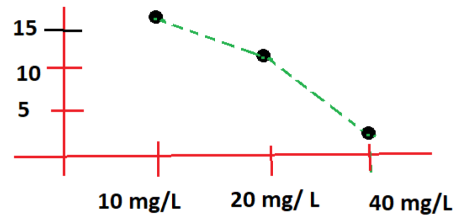


25º C

Si otro investigador realiza las comparaciones a una temperatura ambiente de 30 º C, concluirá posiblemente que sí existe diferencia significativa entre las concentraciones;

10 mg/L	20 mg/L	40 mg/L
17	12	2
16	13	1
17	12	2
18	12	3
16	11	1
17	13	2

No existe evid. suf. para concluir que
M1 mayor que M2
No.........M2 mayor que M3
Sí ....M1 mayor que M3



30 ^{o} C

Esta discrepancia en los 3 experimentos se atribuye a la interacción entre temperatura y concentración. Si no hubiese interacción, entonces los 3 investigadores concluirían todos que no existe diferencia significativa entre las concentraciones, o todos concluirían que sí existe diferencia significativa entre las concentraciones (*)

Por tanto, si en un experimento se concluye diferencia o no diferencia entre ciertos tratamientos y en otro experimento se concluye en el sentido contrario; entonces valdría la pena considerar la posibilidad si algún factor ambiental o externo fuera el responsable al interaccionar con los tratamientos.

(*(*) Más aún, arribarían a las mismas conclusiones en la prueba de*

Scheffé.) En síntesis:

Si no hay interacción entre temperatura y concentración, todos los investigadores aceptarán Ho, o todos aceptarán Ha. Más aún, si aplican Scheffe todos llegarán exactamente a todas las mismas conclusiones: " (Sí, no) existe evid. sif. que M..."

Si hay interacción, se tendrá alguno de los siguientes contrastes:

a) **Uno(s) aceptará(n) Ho y otro(s) aceptará(n) Ha**
b) **Puede ser que todos acepten Ha, pero en la prueba de Scheffe uno(s) diferirá(n) respecto a los demás.**

DISEÑO DEL CUADRADO LATINO

Este diseño es una ampliación del diseño de bloques, el cuadro latino permite la comparación de los tratamientos cuando están sujetos simultáneamente a dos conjuntos de bloques. Un conjunto de bloques se asigna a las filas del cuadrado latino, el otro conjunto de bloques se asigna a las columnas del cuadrado. Luego los tratamientos, generalmente representados con letras latinas, se asignan al azar en cada celda de las intersecciones filas- columnas bajo la única condición de que cada tratamiento aparezca una vez y sólo una vez en cada fila y en cada columna. Éste diseño adolece de la condición de que el número de bloques de cada conjunto debe ser igual al número de tratamientos.

Resolución del diseño (de forma manual)

1. Plantear:

H_o: Los procesos no difieren significativamente en efectividad.

Ó

Las medias de las poblaciones son todas iguales.

H_a: Los procesos difieren significativamente en efectividad.

Ó

Las medias de las poblaciones no son todas iguales

2. Elegir α Si rechazamos H_o al finalizar la prueba, entonces la probabilidad de haber tomado una decisión errónea, es decir la probabilidad de que H_o sea en realidad cierta, es una probabilidad menor que α .

En particular le asignaremos a α el valor de 0.05 = 5%

3. Calcular las sumas de las filas y columnas del siguiente cuadrado:

Columnas: Filas:	**1**	**2**	**.........**	**M**	**Sumas**
1	X_{11}	X_{12}		X_{1M}	$X_{1\bullet}$
2	X_{21}	X_{22}		X_{2M}	$X_{2\bullet}$
....					
M	X_{M1}	X_{M2}		X_{MM}	$X_{M\bullet}$
Sumas	$X_{\bullet 1}$	$X_{\bullet 2}$		$X_{\bullet M}$	$X_{\bullet\bullet}$

Además:

X_1= suma de valores del primer

tratamiento X_2= suma de valores del

segundo tratamiento X_3= suma de valores

del tercer tratamiento

etc.

4. Construir la siguiente tabla ANDEVA:

F. V.	G. L.	SC	CM
Filas	M-1	$X_{1\bullet}^2/M + X_{2\bullet}^2/M + X_{3\bullet}^2/M + \ldots - X_{\bullet\bullet}^2/M^2$	$SC_{FILAS} / (M-1)$
Columnas	M-1	$X_{\bullet 1}^2/M + X_{\bullet 2}^2/M + X_{\bullet 3}^2/M + \ldots - X_{\bullet\bullet}^2/M^2$	$SC_{COLUMNAS} / (M-1)$
Tratamientos	M-1	$X_1^2/M + X_2^2/M + X_3^2/M + \ldots - X_{\bullet\bullet}^2/M$	$SC_{TRATAM} / (M-1)$
Error	(M-1) (M-2)	$X_{11}^2 + X_{12}^2 + X_{13}^2 + \ldots + X_{21}^2 + X_{22}^2 + X_{23}^2 + \ldots + X_{31}^2 + X_{32}^2 + X_{33}^2 + \ldots - (X_{\bullet 1}^2/M + X_{\bullet 2}^2/M + X_{\bullet 3}^2/M + \ldots)$	$SC_{ERROR} / (M-1)(M-2)$
Total	M^2-1		

Además:

$F_{calc.} = CM_{trat} / CM_{error}$

$F_{teórica} = F_{1-\alpha,\ M-1,\ (M-1)(M-2)}$

5. Si $F_{calc.}$ mayor que $F_{teórica}$ rechazar H_o. (aceptar H_a) Si $F_{calc.}$ menor o igual que $F_{teórica}$ aceptar H_o.

Ejemplo: Se realiza un estudio para comparar los índices de monóxido de carbono en 5 puntos estratégicos de una ciudad.(N-norte, S-sur, E-este, O-oeste, C-centro). Los conjuntos de bloques involucrados son determinados días de la semana y distintos horarios del día. El cuadrado obtenido y los datos (*ppm*) registrados se presenta a continuación. Realizar la comparación.

	Lunes	Miércoles	Viernes	Sábado	Domingo
08:00	N (124)	S (124)	C (124)	O (122)	E (124)
11:00	E (112)	C (100)	N (130)	S (131)	O (114)
14:00	S (123)	N (133)	O (112)	E (121)	C (133)
17:00	O (118)	E (112)	S (133)	C (124)	N (134)
20:00	C (102)	O (122)	E (118)	N (131)	S (133)

Diseño de parcelas divididas

En este diseño se utilizan bloques, un mismo conjunto de tratamientos entran en cada bloque, y ese conjunto da origen a otro conjunto de tratamientos. El primer conjunto se le denomina tratamientos principales o parcelas completas o parcelas principales, el segundo conjunto se le denomina tratamientos secundarios o subparcelas o parcelas divididas. Es un diseño comúnmente utilizado cuando se tienen partes grandes de cierta materia prima (parcela principal) y cada parte grande se divide en partes pequeñas (subparcelas). El término tiene antecedentes agrícolas porque las parcelas grandes eran grandes extensiones de tierra y las parcelas pequeñas eran pequeñas áreas de tierra. Los bloques generalmente corresponden a repeticiones del experimento o a distintos entornos para llevarlo a cabo. Una tabla de datos del diseño de 3 bloques, 3 parcelas principales y 3 subparcelas tiene entonces la siguiente estructura:

Bloques:	1	1	1	2	2	2	3	3	3
Parcela principal:	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
Subparcela:	dato	Dato	Dato	Dato	Dato	Dato	Dato	Dato	Dato
1	dato	Dato	Dato	Dato	Dato	Dato	Dato	Dato	Dato
2	dato	Dato	Dato	Dato	Dato	Dato	Dato	Dato	Dato
3	dato	Dato	Dato	Dato	Dato	Dato	Dato	Dato	Dato

El modelo estadístico para este diseño es:

$$y_{ijkl} = \mu + \tau_i + \beta_{)} + (\tau\beta)_{ij} + \gamma_k + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\gamma\beta)_{ijk} + (e)_{ijk}$$

donde τ_i, $\beta_{)}$ y $(\tau\beta)_{ij}$ corresponden a la parcela completa y se refieren respectivamente a los bloques (factor A), los tratamientos principales (factor B) y al error de la parcela completa (AB), en tanto γ_k, $(\tau\gamma)_{ik}$, $(\beta\gamma)_{jk}$, $(\tau\gamma\beta)_{ijk}$ y $(e)_{ijk}$ corresponden a las subparcelas y se refieren respectivamente al tratamiento de la subparcela (factor C), y a las interacciones AC y BC y al error de la subparcela.
En este diseño, tanto las parcelas como las subparcelas generalmente corresponden a la definición de “tratamiento=proceso”, y por otro lado, los bloques generalmente son de efectos aleatorios.

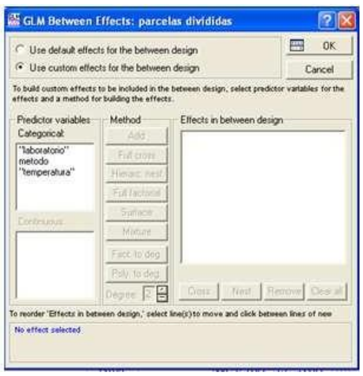
Ejemplo

Personal de la FAO está interesado en producir una pasta de maní que se conserve el máximo de tiempo, la cual se suministrará a la población en hambruna en diversas partes del planeta. Disponen de 3 métodos diferentes para preparar la pasta y de cuatro diferentes temperaturas de cocción. Los investigadores han decidido usar 3 laboratorios distintos como bloques. Llevado a cabo el experimento, se obtuvieron los siguientes datos (conservación en días). Así pues, las porciones grandes de pasta se destinaban para ser tratadas con distintos métodos de preparación, pero distintas partes de esas porciones se sometían a distintas temperaturas de cocción.

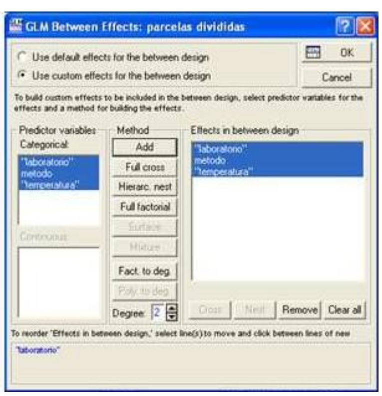
Laboratorio:	1	1	1	2	2	2	3	3	3
Método de preparación:	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
Temperatura:									

200	30	34	29	28	31	31	31	35	32
225	35	41	26	32	36	30	37	40	34
250	37	38	33	40	42	32	41	39	39
275	36	42	33	41	40	40	40	44	45

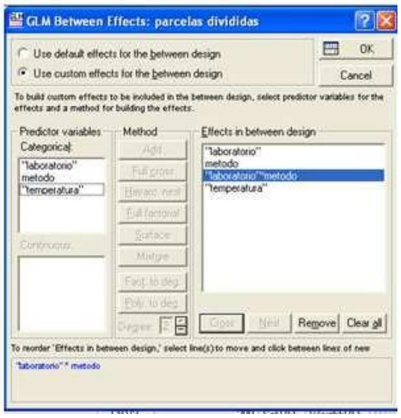
Los pasos para trabajar con este diseño en *Statistica* son:
Statistic → Advanced Linear/Nonlinear Models→ General Linear Models → General Linear Models→Ok →Variables→(en var. dependiente elegir *tiempo*, en categorical predictors elegir las demás) →*Ok*→*Options*→*Random Factors*→(elegir laboratorio) →*Ok*→*Quick*→*Between effects*→*Use custom effects for the between design,* entonces está presente el siguiente cuadro de diálogo:



Del modelo del diseño de las parcelas divididas, nos propondremos obtener los efectos de A (bloques: laboratorio) , B (parcela completa: método) , C (subparcela: temperatura), y las interacciones AB, AC y BC.
Elegir *laboratorio, método* y *temperatura*, presionar *Add*, entonces el cuadro es::

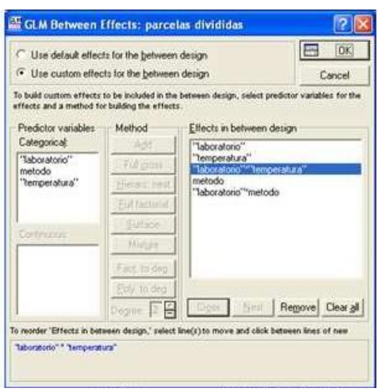


Elegimos *laboratorio* y *método* de la lista derecha y presionamos *cross* como en el

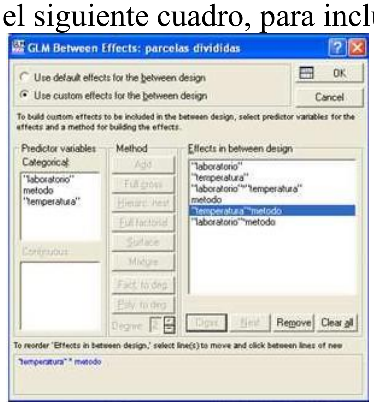


siguiente cuadro, para incluir la interacción de esos 2 factores:

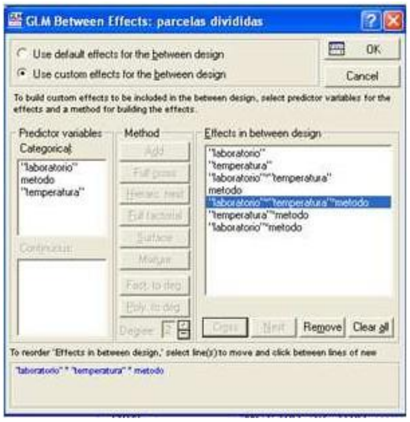
Elegimos *laboratorio* y *temperatura* de la lista derecha y presionamos *cross* como en el siguiente cuadro, para incluir la interacción de esos 2 factores:



Elegimos método y *temperatura* de la lista derecha y presionamos *cross* como en el siguiente cuadro, para incluir la interacción de esos 2 factores:



Por último, para incluir la interacción de los 3 factores, elegimos en la lista derecha cualquier interacción de 2 factores (digamos *laboratorio*temperatura*) y



elegimos también *método*, y presionamos *Cross*:

Presionemos *OK, Ok* y *All effects*, aparecerán los resultados del análisis de varianza: Por el valor (rojo) de *p* de temperaturas, hay diferencia significativa entre las temperaturas, lo mismo entre los métodos. No tiene sentido interesarnos en comparar los bloques, puesto que no son tratamientos, y peor aún, son de efectos aleatorios.
Si presionamos *All efects Graph* , elegimos *método- temperatura* y continuamos el proceso, podemos percatarnos para cual combinación de método y temperatura se obtienen los mayores tiempos de conservación.

Intente contestar las siguientes preguntas, presionando los botones de los cuadros de diálogo que aparezcan en el contorno del actual procedimiento::
a) Hay diferencia significativa entre los métodos? (Se acepta H_o o H_a
b) Existe un método significativamente mejor que los otros? (Se entiende que para maximizar los días de conservación, que es la variable de los datos)
c) Hay diferencia significativa entre las temperaturas? (Se acepta H_o o H_a)
d)) Existe una temperatura significativamente mejor que las otras? (Se entiende que para maximizar los días de conservación, que es la variable de los datos)
e) Hay interacción método-temperatura?
f) Para cual combinación de método-temperatura se obtienen los mayores tiempos de conservación?

Diseño Jerárquico

Antecedentes
Esta es la tabla de un problema ya contemplado de parcelas divididas:

Laboratorio:	1	1	1	2	2	2	3	3	3
Método de preparación:	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
Temperatura:									
200	30	34	29	28	31	31	31	35	32
225	35	41	26	32	36	30	37	40	34
250	37	38	33	40	42	32	41	39	39

275	36	42	33	41	40	40	40	44	45

La cual pudo haberse presentado también de la siguiente forma:

Método de preparación	**1**	**1**	**1**	**1**	**2**	**2**	**2**	**2**	**3**	**3**	**3**	**3**
Temperatura:	**200**	**225**	**250**	**275**	**200**	**225**	**250**	**275**	**200**	**225**	**250**	**275**
Labor. 1	30	35	37	36	34	41	38	42	29	26	33	33
Labor.2	….	….										
Labor.3	….											

De forma tal que arriba aparecen los dos conjuntos de tratamientos: (método y temperatura). Nadie dudaría que la temperatura de 200 en el primer método guarda una correspondencia especial con la temperatura de 200 del segundo método. (La misma explicación es válida para las demás temperaturas). Imagine que en lugar de la tabla anterior, la tabla hubiese sido la siguiente:

Método de preparación	**1**	**1**	**1**	**1**	**2**	**2**	**2**	**2**	**3**	**3**	**3**	**3**
Ingrediente especial:	**maní**	**chocolate**	**fresa**	**Durazno**	**crema**	**uva**	**frambuesa**	**salsa**	**arroz**	**yogurt**	**crema**	**Atún**
	dato	dato	dato	…	…	…	…					
	…	…	….	…	…							

El primer ingrediente (maní) dentro del primer método, no guarda ninguna correspondencia especial con el ingrediente (crema) dentro del segundo método, (La misma explicación es válida para los demás ingredientes) . Una tabla como esta corresponde al diseño jerárquico.

Diseño Jerárquico de 2 etapas

En algunos experimentos de varios factores, los niveles de un factor (digamos el B), no son idénticos en los diferentes niveles de otra factor (digamos el A). A tal diseño se le conoce como jerárquico o anidado. Cuando son sólo 2 factores los involucrados es un diseño jerárquico de 2 etapas. Ilustrémoslo con un ejemplo:

Ejemplo.

Un acuario que compra ingredientes orgánicos a tres diferentes proveedores. El acuario desea determinar si la calidad de los ingredientes es la misma de cada proveedor. Hay cuatro lotes de ingredientes de cada proveedor, y se hacen determinaciones de la calidad de cada lote. Si los factores estuvieran cruzados, el lote 1 del primer proveedor estaría relacionado con el lote 1 del segundo proveedor, pero no es así en este caso. Lo mismo sucede con los demás lotes. Otra manera de decirlo es que los lotes del segundo proveedor bien podrían llamárseles lotes 4, 5 y 6; los del tercer proveedor serían 7, 8 y 9; toda vez que no guardan relación los lotes de cada proveedor con respecto a los demás proveedores.

Las medidas de calidad están expuestas en la siguiente tabla:

Proveedor 1					**Proveedor 2**				**Proveedor 3**			
lotes	***1***	***2***	***3***	***4***	***1***	***2***	***3***	***4***	***1***	***2***	***3***	***4***
	1	-2	-2	1	1	0	-1	0	2	-2	1	3
	-1	-3	0	4	-2	4	0	3	4	0	-1	2
	0	-4	1	0	-3	2	-2	2	0	2	2	1

Los lotes de ingredientes se eligieron al azar de cada proveedor. La calidad varía considerablemente, lo que genera problemas con los clientes cuando éstos ven variaciones en las distintas compras.

Se desea saber si las diferencias de calidad obedecen a los distintos proveedores o si más bien corresponden a los distintos lotes. Los datos vaciados en *Statistica* exhiben la siguiente estructura:

	1 proveedor	2 lotes	3 calidad
1	proveedor 1	lote 1	1
2	proveedor 1	lote 1	-1
3	proveedor 1	lote 1	0
4	proveedor 1	lote 2	-2
5	proveedor 1	lote 2	-3
6	proveedor 1	lote 2	-4
7	proveedor 1	lote 3	-2
8	proveedor 1	lote 3	0
9	proveedor 1	lote 3	1
10	proveedor 1	lote 4	1
11	proveedor 1	lote 4	4
12	proveedor 1	lote 4	0
13	proveedor 2	lote 1	1
14	proveedor 2	lote 1	-2
15	proveedor 2	lote 1	-3
16	proveedor 2	lote 2	0
17	proveedor 2	lote 2	4
18	proveedor 2	lote 2	2
19	proveedor 2	lote 3	-1
20	proveedor 2	lote 3	0
21	proveedor 2	lote 3	-2
22	proveedor 2	lote 4	0
23	proveedor 2	lote 4	3
24	proveedor 2	lote 4	2
25	proveedor 3	lote 1	2
26	proveedor 3	lote 1	4
27	proveedor 3	lote 1	0
28	proveedor 3	lote 2	-2
29	proveedor 3	lote 2	0
30	proveedor 3	lote 2	2

El modelo estadístico para el diseño jerárquico es:

$$y_{ij} = \mu + \tau_j + \beta_{j(i)} + e_{(ij)k}.$$

i= 1, 2, …,a
j= 1, 2, …,b
k= 1, 2, …,n

Entonces, existen a niveles para el factor *A*, y *b* niveles para el factor *B*, jerarquizados bajo cada nivel de *A*, y n réplicas. El subíndice *j(i)* dice que el j-ésimo nivel del factor *B* está anidado en el i-ésimo nivel del factor *A*. las réplicas están anidadas dentro de las combinaciones de los niveles de *A* y *B*. el subíndice *(ij)k* se usa para el término del error.

Pasos para el diseño jerárquico:
Statistic → Advanced Linear/Nonlinear Models→ General Linear Models →Nested Design Anova→Ok→Variables→(en var. dependiente elegir *calidad*, en categorical predictors elegir *lote* y *proveedor*) →Ok→Options →Random factors→(elegir lote) →Ok→Ok→All effects.

Univariate Tests of Significance for calidad (jerárquico 2 niveles)
Over-parameterized model
Type III decomposition

Effect	Effect (F/R)	SS	Degr. of Freedom	MS	Den.Syn. Error df	Den.Syn. Error MS	F	p
Intercept	Fixed	4.69444	1	4.694444	9.00000	7.768519	0.604291	0.456892
proveedor	Fixed	15.05556	2	7.527778	9.00000	7.768519	0.969011	0.415783
lote(proveedor)	Random	69.91667	9	7.768519	24.00000	2.638889	2.943860	0.016674
Error		63.33333	24	2.638889				

Intente contestar las siguientes preguntas, presionando los botones de los cuadros de diálogo que aparezcan en el contorno del procedimiento:
a) Hay diferencia significativa entre los proveedores? (Se acepta H_o o H_a), considere tratamiento (proveedor) =proceso
b) Existe un proveedor significativamente mejor que los otros?
c) Hay diferencia significativa entre los lotes? (Se acepta H_o o H_a)?

Nota: en el diseño los lotes son aleatorios, si hubiesen sido fijos, se seguirían los siguientes pasos:

Statistic → Advanced Linear/Nonlinear Models→ General Linear Models →Nested Design Anova→Ok→Variables→(en var. dependiente elegir *calidad*, en categorical predictors elegir *lote* y *proveedor*) →Ok→All effcts

Cabe señalar que los lotes son aleatorios (no fijos) porque se eligieron al azar de cada proveedor. Esto significa que cada proveedor manejaba más de 4 lotes, y los 4 referidos fueron escogidos al azar de todos los lotes pertenecientes del proveedor. Hubiesen sido fijos si cada proveedor hubiese manejado sólo 4 lotes, los cuales todos estarían representados en la tabla.

El análisis de varianza se presenta en la tabla anterior. No hay diferencia significativa en la calidad que se atribuya a los proveedores., pero sí la hay en los lotes de dentro de cada proveedor. Las implicaciones de éstos resultados son interesantes. Si la diferencia de calidad se hubiese atribuido a los proveedores, entonces el acuario hubiese considerado seriamente en elegir al "mejor" proveedor. Pero la diferencia de calidad fué atribuida a los lotes, esto quiere decir que el acuario, si desea mejorar la calidad, debe trabajar con todos los proveedores en sus procesos de producción o en el sistema de control de calidad interno.

Factores cruzados.
En ocasiones, es posible que el diseño en cuestión tenga algunos factores cruzados (como en el diseño factorial) y otros anidados (como en el jerárquico). Inspirémonos en

el siguiente ejemplo para profundizar mejor en estos tipos de diseños mixtos:

Ejemplo:
Un ecólogo está estudiando los tiempos para llevar a cabo determinadas mediciones de la densidad de cierta especie de árbol. Las mediciones se llevan a cabo usando tres distintos software, se trabaja en dos distintos ecosistemas y él se apoya en 4 operadores para cada ecosistema. Dado que los ecosistemas están distanciados entre sí, los 4 operadores de un ecosistema no son los mismos que en el otro ecosistema. Se obtiene dos réplicas en cada combinación de tratamientos, los tiempos son los siguientes:

Ecosistema	***1***	***1***	***1***	***1***	***2***	***2***	***2***	***2***
Operador	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
Software 1	22	23	28	25	26	27	28	24
Software 1	24	24	29	23	28	25	25	23
Software 2	30	29	30	27	29	30	24	28
Software 2	27	28	32	25	28	27	23	30
Software 3	25	24	27	26	27	26	24	28
Software 3	21	22	25	23	25	24	27	27

Notemos que los operadores están jerarquizados en los 2 ecosistemas. Mientras que los software y los operadores pueden estar cruzados (esto es podría haber interacción entre ellos). Además, no puede haber interacción entre ecosistemas y operadores porque no todos los operadores operan en todos los ecosistemas, y por ende, no puede existir tampoco la interacción operador-ecosistema-software.

El modelo estadístico para este diseño es:

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_{k(j)} + (\tau\beta)_{ij} + (\tau\gamma)_{ik(j)} + e_{(ijk)l}.$$

Para nuestro ejemplo:
τ_i es el efecto del i-ésimo software
β_j es el efecto del j- ésimo ecosistema
$\gamma_{k(j)}$ es el efecto del k-ésimo operador dentro del j-ésimo ecosistema
$(\tau\beta)_{ij}$ es la interacción software por ecosistema
$(\tau\gamma)_{ik(j)}$ es la interacción de software por el operador dentro del ecosistema
$e_{(ijk)l.}$ es el término del error.

Así:
i= 1, 2,3,
j= 1, 2.
k= 1, 2,3,4
l=1,2

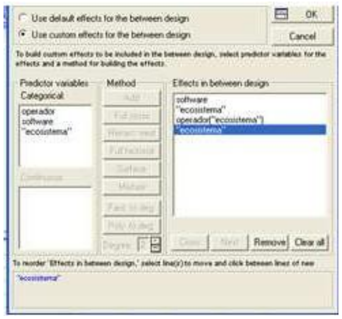
Los pasos son:
Statistic → Advanced Linear/Nonlinear Models→ General Linear Models → General Linear Models→Ok →Variables→(en var. dependiente elegir *tiempos*, en categorical predictors elegir las demás) *→Ok→Options→Random Factors*→(elegir operador) *→Ok→Quick→Between effects→Use custom effects for the between design,* entonces está presente el siguiente cuadro de diálogo:

Elegir *software* y *ecosistemas*, presionar *Add*, entonces el cuadro es::

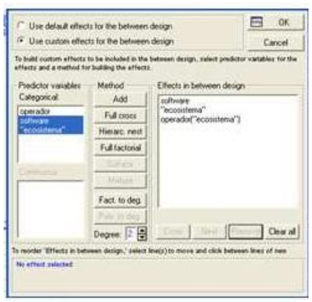
Elegir *operador* y *ecosistema* de la lista izquierda y presionar *Hierarc Nest*, entonces aparece el cuadro, donde elegir ecosistema y presionamos *Ok*: (esto es, porque

los operadores están jerarquizados en los ecosistemas, no al revés)

En el sig. cuadro, elegimos *ecosistema* de la lista derecha y presionamos *remove*, esto es, porque no queremos incluir en el modelo al efecto del ecosistema 2 veces:

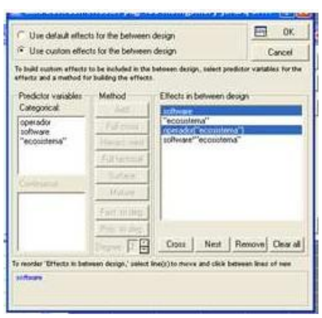


Elegimos *software* y *ecosistema* del la lista izquierda y presionamos *full cross* como en

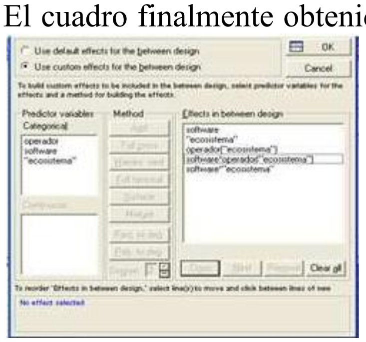


el siguiente cuadro, para incluir la interacción de esos 2 factores:

Por último, elegimos, en la lista derecha, *software* y además *operador (“ecosistema”)* como en el siguiente cuadro y presionamos después *cross*



El cuadro finalmente obtenido es:



Presionemos *OK, Ok* y *All effects*, aparecerán los resultados del análisis de varianza.

Cabe señalar que los operadores son aleatorios (no fijos) porque se eligieron al azar para cada software. Hubiesen sido fijos si para cada software existieran en el mundo entero 4 operadores capaces de manipularlos (los representados en la tabla), lo cual se antoja imposible. Esa condición de "aleatorios" se condicionó en el procedimiento cuando se eligió "operador" en *"Random Factors"*.

Considere que los datos representan tiempos, entonces a menores tiempos de tardanza se considera mayor eficacia de los procesos respectivos.

Preguntas:

a) Hay diferencia significativa entre los software? (Se acepta H_o o H_a)
b) Existe un software significativamente mejor que los otros?
c) Hay diferencia significativa entre los ecosistemas? (Se acepta H_o o H_a)
d) Existe un software significativamente mejor que los otros?
e) Hay interacción software –ecosistema?
(de haberla, significaría que unos software son más eficaces que otros dependiendo del ecosistema en cuestión.
f) Hay diferencia significativa entre los operadores? (Se acepta H_o o H_a). De haberla significaría que unos operadores son más habilidosos que otros para realizar las mediciones en el menor tiempo posible. Para ello vea la *p* de "*operador(ecosistema)*"

Compendio de ejercicios

1. Se evaluó el efecto de los tratamientos de almacenamiento en frío durante tres semanas sobre la calidad de la naranja Valencia, C*itrus sinensis*. La siguiente tabla de datos exhibe los puntajes de daños en las naranjas sometidas al experimento. Cada dato procede de una naranja. Los puntajes se manejan en una escala donde a puntajes mayores corresponden daños mayores. Resuelve la comparación de las tres temperaturas.

38 C	***46 C***	***54 C***
5	8	12
3	7	7
2	7	

2. Se registran con chips y se comparan los kilómetros que en un año recorren elefantes adultos machos de 3 distintas especies, no obstante, es sabido que el tipo de vegetación puede ser un factor que influya en las distancias y por tanto se le controla manejándolo como bloques; así entonces fueron elegidos 6 animales, de distintas especies y de distinto hábitat. Realiza la comparación de las especies. (*los datos están en miles de kilómetros*)

	Africano	*Asiático Hindú*	*Asiáticode Indonesia*
Tropical	4.2	2.1	1.5
Semi árido	3.1	1.3	0.6

3. Se estudiará la permanencia de huevecillos de cierta especie de rana en dos

hábitats (tratamientos), los datos serán los días que se mantengan antes de romperse en esas condiciones, (si algún huevo nunca se rompe se desechará). Comúnmente existen dos colores de huevecillos: blancos y amarillos. Se sospecha que el color incide en el tiempo de sobrevivencia. Si las letras indican el color de cada huevo, y cada letra es un huevo realiza la asignación de los huevos en los hábitats.

AAAAAAAAAAAAAAAAAABBBBBBBBBBBBBB

hábitat 1	hábitat 2

a) ¿Cuántos huevos se asignaron en el primer bloque?

b) ¿Cuántos huevos se asignaron en el segundo bloque?

4. Si en cambio, se comparan los tiempos de permanencia antes de romperse huevecillos de 2 especies de ranas en un mismo hábitat, y el color puede incidir en ese tiempo, y si las letras indican los colores de los huevos disponibles, determina cuantos huevos se considerarán en la comparación.

Especie 1: AAAAABBBB, Especie 2: AAAAAABBBBB

especie 1	especie 2

a) ¿Cuántos huevos se asignaron en el primer bloque?
b) ¿Cuántos huevos se asignaron en el segundo bloque?

5. Explica cuándo es conveniente implementar el diseño de bloques (en lugar del diseño totalmente aleatorio), y/o por qué es útil dicho diseño.

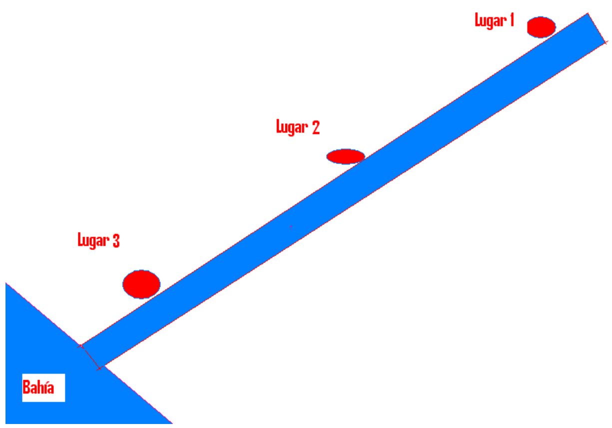
6. *Canavalia ensiformis* es una leguminosa nativa de América Tropical, la cual está siendo estudiada con el objetivo de incorporarla a la producción en agroecosistemas mexicanos. Se construyó un cuadrado latino donde un conjunto de tratamientos son las diferentes posiciones de siembra con respecto a los camellones (sobre camellón SC, en mitad de camellón MC, en plano Pl y en surcos S) las cuales se asocian a diferentes disponibilidades de agua. Un conjunto de bloques se vino a ser los distintos potenciales de presión del xilema (sistema que conduce las sales minerales y el agua de las raíces a las hojas). El otro conjunto de bloques fué la dosis de fósforo y potasio. Los rendimientos promedio de grano en kilogramo por hectárea de cada combinación aparecen en el siguiente cuadrado latino. Determina que posición de siembra maximiza la producción en cualquier condición.

	Dosis baja	**Dosis moderada**	**Dosis semi alta**	**Dosis alta**
Potencial bajo	SC (1119)	PL (1234)	S (1334)	MC (1754)

Potencial moderado	S (1445)	MC (1875)	SC (773)	PL (1210)
Potencial semialto	MC (1778)	SC (984)	PL (1114)	S (1224)
Potencial alto	PL (1224)	S (1447)	MC (1997)	SC (1225)

7. En un ejercicio anterior donde se comparaban 3 variedades de maiz: dos de tipo transgénico (A y B) y una variedad no transgénica (C) se construyó un cuadrado latino parecido al siguiente, suponga que se llevó a cabo el experimento y se reportaron con datos la producción media en toneladas por hectárea por año. Realiza la comparación de los tratamientos.

	Abono orgánico	*Abono artificial*	*Sin abono*
Riego artificial	A 3.9	B 2.2	C 8.4
Sin riego artificial	B 2.0	C 8.3	A 8.2
Mixto	C 8.2	A 2.2	B 3.8



8. Se quiere probar la efectividad de un novedoso sistema de contención anticontaminante. Accidentalmente se arrojó material tóxico a un río que entra en una gran área de pesca comercial en agua salada. Los ambientalistas saben que el agua transporta el material tóxico, midiendo la cantidad de material tóxico (cmt) en partes por millón, hallado en las ostras escogidas en tres lugares diferentes, desde la salida del estuario hasta la bahía donde se realiza la mayor parte de la pesca comercial. En condiciones normales, tras el accidente, las ostras estarían uniformemente contaminadas. Pero se implementó a lo largo del río, oportunamente, el sistema anticontaminante que opera con una mecánica de contención. Se presentan los resultados en la siguiente tabla, donde cada dato corresponde a una ostra hallada en el lugar correspondiente. Es claro que si se demuestra menos material tóxico en el lugar 3 que en el lugar 1 entonces se determinaría que fué efectivo el sistema de contención.

Lugar 1 (estuario)	**Lugar 2 (lejos de la bahía)**	**Lugar 3 (cerca de la bahía)**
15	11	12

26	15	6
20	10	4
20	20	6
29	11	5
28	5	
21		

Tabla de contaminación de las ostras

a) ¿Qué tipo de diseño es ?

b) Algunos ambientalistas proponen que la comparación mediante la pasada tabla no es adecuada, toda vez que unas ostras son "amarillas" , otras "verdes claro" y otras "verdes", las blancas son más repelentes a la adhesión de contaminantes que las amarillas, y la distribución de ellas en la tabla es:

Lugar 1 (estuario)	Lugar 2 (lejos de la bahía)	Lugar 3 (cerca de la bahía)
Verde	Amarilla	Amarilla
Verde	Amarilla	Amarilla
Verde	Verde claro	Verde claro
Verde	Verde claro	Verde
Verde	Verde	Verde
Verde claro	Verde	
Amarilla		

Tabla de colores de las ostras

Lo cual haría una comparación sesgada (inequitativa) de los lugares. Entonces que diseño propones para comparar los lugares?

c) Luego, los investigadores reciben información en el sentido de que las ostras pequeñas tienden a repeler menos la contaminación que las grandes, entonces, la edad (Así como lo fue antes el "color") es otro factor que haría una comparación sesgada (inequitativa) de los lugares, qué diseño propones para compara los lugares, donde se remuevan simultáneamente los efectos del "color" y tamaño" ?

Lugar 1 (estuario)	Lugar 2 (lejos de la bahía)	Lugar 3 (cerca de la bahía)
Grande	Grande	mediana
pequeña	Pequeña	Pequeña
Mediana	Pequeña	Pequeña
Grande	Pequeña	grande
Grande	Mediana	Mediana
Grande	Grande	
Grande		

Tabla de tamaños de las ostras

d) El diseño que propusiste en c), es factible de construirse con los datos representados en las tres tablas?, o es necesario recolectar más ostras?

9. Diseño de subparcelas divididas

A veces las parcelas divididas se dividen en otros tratamientos, las llamadas subparcelas, el siguiente relato lo ilustra: Un veterinario está estudiando el tiempo de absorción de un antibiótico en cápsulas. Hay tres ayudantes de laboratorio, tres niveles de dosis y cuatro espesores de la cápsula que son de interés para el investigador. Se tiene la disposición de llevar a cabo el experimento en cuatro ocasiones (días) distintos, estos días pueden representar los bloques. La tabla obtenida fue la siguiente:

		Dia	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4
		Esp	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	2	3	4	1	2	3	4
Téc	*Dosis*																	
1	*1*		95	104	101	108	95	106	103	109	96	105	106	113	90	100	102	114
1	*2*		71	82	85	85	78	84	86	84	70	81	88	90	68	84	85	88
1	*3*		108	115	117	116	110	109	116	110	107	106	112	117	109	112	115	118
2	*1*		96	99	95	97	100	101	99	112	94	100	104	121	98	102	100	118
2	*2*		70	84	83	85	72	79	80	86	66	84	87	90	68	81	85	85
2	*3*		108	100	105	109	104	102	108	109	100	101	109	117	106	103	110	116
3	*1*		95	102	105	107	92	100	101	108	90	97	100	110	98	102	105	110
3	*2*		70	81	84	87	69	76	80	86	73	75	82	91	72	78	80	95
3	*3*		100	106	113	115	101	104	109	113	98	100	104	112	101	105	110	120

Considerando que los bloques (dias) se presentan con A, los tratamientos principales (parcelas): espesor, con B; subparcelas (técnico) con C, y dosis (subsubparcela):dosis con D, y considerando que se obtendrán los efectos:
A
B
C
D
A
B
A
C
BC
ABC
AD
BD
ABD
CD
ACD
BCD

Los técnicos (C), y los días (bloques, A) son efectos aleatorios (random factors). Llevar a cabo el procedimiento de comparación e intentar contestar las siguientes preguntas, presionando los botones de los cuadros de diálogo que aparezcan en el contorno del procedimiento:
a) Hay diferencia significativa entre los espesores? (Se acepta H_o o H_a)
b) Existe un espesor significativamente mejor que los otros?
c) Hay diferencia significativa entre las dosis? (Se acepta H_o o H_a)
d)) Existe una dosis significativamente mejor que las otras?
e) Hay interacción espesor-.dosis?
f) Qué combinación de espesor-dosis es la mejor?