EXPERIMENTOS FACTORIALES

En los modelos de los experimentos factoriales los parámetros τi que hacen referencia a los efectos de tratamientos se descompone en un conjunto de parámetros que dan cuenta de cada uno de los factores intervinientes y se agrega según sea necesario, conveniente y posible, los términos correspondientes a las interacciones.

Modelos aditivos

Los modelos factoriales aditivos son aquellos en los que los términos que modelan la interacción están ausentes. Para ejemplificar este caso se presenta un experimento factorial 2x2 (dos factores con dos niveles cada uno) en el que la interacción se supone ausente y montado en un diseño completamente aleatorizado. Los Factores se han designado como A y B y sus niveles como A1,A2 y B1,B2. Como existen 4 tratamientos (A1B1, A1B2, A2B1, A2B2) y estos no están repetidos, se necesitan sólo cuatro parcelas experimentales. Dado que el diseño es completamente aleatorizado la asignación de las parcelas a cada uno de los tratamientos es al azar. Un arreglo posible se presenta en la siguiente figura

A_2B_1	A_1B_2
A_1B_1	A_2B_2

Figura 1: Experimento bifactorial sin repeticiones montado en un diseño completamente aleatorizado.

El modelo para este experimento es el siguiente:

$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$ con i=1,2; j =1,2

En este modelo Y_{ij} representa la respuesta al i-ésimo nivel del factor A y j-ésimo nivel de factor B, μ representa una media general, α_i el efecto que produce el i-ésimo nivel del factor A, β_j corresponde al j-ésimo nivel del factor B y ε_{ij} es el error asociado a la observación ij-ésima que como siempre se suponen normales, independientes, con esperanza cero y varianza común σ^2.

El cuadro de Análisis de la Varianza para este diseño se calcula según las expresiones provistas en la Tabla 1.

Tabla 1:Expresiones para el cálculo del cuadro de análisis de la varianza de un experimento bifactorial con diseño completamente aleatorizado.

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F
Factor A	$SCF=\sum_{i=1}^{a} \frac{(y_{i\bullet})^2}{b} - \frac{(y_{\bullet\bullet})^2}{a \cdot b}$	gla=a-1	$CMA=\frac{SCA}{gla}$	$\frac{CMA}{CMD}$
Factor B	$SCC=\sum_{j=1}^{b} \frac{(y_{\bullet j})^2}{a} - \frac{(y_{\bullet\bullet})^2}{a \cdot b}$	glb=b-1	$CMB=\frac{SCB}{glb}$	$\frac{CMB}{CMD}$
Dentro (Error Experimental)	SCD=SCT-SCA-SCB	gld=(a-1)(b-1)	$CMD=\frac{SCD}{gld}$	
Total	$SCT=\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2-\frac{(y_{\bullet\bullet})^2}{a \cdot b}$	glt=a·b-1		

Ejemplo 1

En un ensayo comparativo del efecto del estrés hídrico y salino sobre la germinación de Atriplex cordobensis, se sometieron lotes de semillas a cuatro niveles de potencial agua: 0, -0.5, -1.0 y –1.5 Mpa obtenidos mediante la aplicación al medio de dos osmolitos: polietilenglicol (PEG) o cloruro de sodio (ClNa). El experimento se montó en un diseño completamente aleatorizado sin repeticiones cuyos resultados se presentan en la siguiente tabla.

Tabla 2: Resultados de un ensayo comparativo del efecto de distintos potenciales agua del substrato obtenido con dos osmolitos: polietilenglicol(PEG) y cloruro de sodio (ClNa) sobre el porcentaje de germinación en A. cordobensis.

Mpa	*0*	*-0.5*	*-1.0*	*-1.5*
ClNa	85	78	54	14
PEG	83	76	43	9

Cuando los experimentos factoriales no tienen repeticiones, el analista debe suponer que los factores no interactúan para poder estimar la varianza del error experimental. Si este supuesto no se cumple entonces el experimento está deficientemente diseñado y las conclusiones del análisis pueden ser completamente erróneas. Existen algunas pruebas para verificar este supuesto como la prueba de aditividad de Tukey (1949).

La tabla de análisis de la varianza para este experimento, suponiendo un modelo aditivo, se muestra en la siguiente tabla.

Tabla 3: Cuadro de análisis de la varianza para de un experimento bifactorial para evaluar el efecto de distintos potenciales agua del substrato obtenidos por el agregado al medio de dos osmolitos: polietilenglicol(PEG) o cloruro de sodio (ClNa) sobre el porcentaje de germinación en A. cordobensis.

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F
Osmolito	50.0	1	50.0	5.6
Potencial Agua	6118.5	3	2172.8	241.4
Dentro	27.0	3	9.0	
Total	6195.5	7		

Consultando los valores críticos de una F con 1 y 3 grados de libertad para el factor osmolito y con 3 y 3 grados de libertad para potencial agua, se puede apreciar que ambos factores afectan significativamente el porcentaje de germinación.

Modelos con interacción

Si el experimentador supone o sospecha que la respuesta a dos o más factores no se puede explicar como la suma de sus efectos individuales entonces el modelo para el experimento factorial deberá incluir términos de interacción que den cuenta de este hecho. La inclusión de términos de interacción en el modelo conlleva la necesidad de tener repeticiones para cada tratamiento porque de otra forma no es posible estimar los parámetros adicionales. Aunque no se profundizará más en este tema, cuando el experimento tiene dos factores, existen solo interacciones de primer orden, cuando tiene tres factores, existen interacciones de primer y de segundo orden y así sucesivamente para factoriales de mayor orden.

A continuación se examinará con algún nivel de detalle un experimento bifactorial con interacción y se presentará un ejemplo. El modelo para un experimento bifactorial con interacciones es una ampliación del modelo para el experimento bifactorial descripto anteriormente, excepto que incluye un conjunto adicional de parámetros, conocidos como de interacción.

$Y_{ijk} = \mu + \alpha_i + \beta_j + \delta_{ij} + \varepsilon_{ijk}$ con $i=1,2$; $j=1,2$; $k=1,..,n_{ij}$

En este modelo Y_{ijk} representa la respuesta de la k-ésima repetición en el i-ésimo nivel del factor A y j-ésimo nivel de factor B, μ representa una media general, α_i el efecto que produce el i-ésimo nivel del factor A, β_j corresponde al efecto del j-ésimo nivel del factor B, δ_{ij} los efectos adicionales (interacciones) para cada combinación de los niveles de los factores y ε_{ijk} es el error asociado a la observación ijk-ésima que como siempre se supone normal e independiente con esperanza cero y varianza común σ^2.

Debe notarse que el subíndice k se mueve entre 1 y n_{ij}, es decir, el número de repeticiones para el tratamiento puede ser distinto.

El cuadro de Análisis de la Varianza para este diseño se calcula según las expresiones provistas en la Tabla 4.

Tabla 4: Expresiones para el cálculo del cuadro de análisis de la varianza de un experimento bifactorial con interacción en un diseño completamente aleatorizado.

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Promedio	F_o
A tratamientos	SS_A	$a-1$	$MS_A = \frac{SS_A}{a-1}$	$F_o = \frac{MS_A}{MS_E}$
B tratamientos	SS_B	$b-1$	$MS_B = \frac{SS_B}{b-1}$	$F_o = \frac{MS_B}{MS_E}$
Interacción	SS_{AB}	$(a-1)(b-1)$	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$F_o = \frac{MS_{AB}}{MS_E}$
Error	SS_E	$ab(n-1)$	$MS_E = \frac{SS_E}{ab(n-1)}$	
Total	SS_T	$abn-1$		

Los términos de la suma de los cuadrados se calculan como se muestra a continuación:

Suma de cuadrados totales:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{\bullet\bullet\bullet}^2}{abn}$$

Suma de los cuadrados de los efectos son:

$$SS_A = \frac{1}{bn}\sum_{i=1}^{a} y_{i\bullet\bullet}^2 - \frac{y_{\bullet\bullet\bullet}^2}{abn} \quad \text{y} \quad SS_B = \frac{1}{an}\sum_{j=1}^{b} y_{\bullet j\bullet}^2 - \frac{y_{\bullet\bullet\bullet}^2}{abn}$$

Es conveniente obtener la suma de los cuadrados de la interacción en dos fases. Primero, se calcula la suma de cuadrados entre los totales de las celdas ab que se conoce como la suma de cuadrados debido a "subtotales":

$$SS_{Subtotals} = \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij\bullet}^2 - \frac{y_{\bullet\bullet\bullet}^2}{abn}$$

Esta suma de cuadrados también contiene SS_A y SS_B . Por lo tanto, el segundo paso es calcular la suma de cuadrados de la interacción como sigue:

$$SS_{AB} = SS_{Subtotals} - SS_A - SS_B$$

Ahora por substracción podemos calcular la suma de cuadrados del error como sigue:

$$SS_E = SS_T - SS_{AB} - SS_A - SS_B \qquad \text{ó} \qquad SS_E = SS_T - SS_{Subtotals}$$

Ejemplo 2

En un estudio sobre la potencialidad forrajera de Atriplex cordobensis, un arbusto que crece en depresiones del chaco árido argentino, se evaluó la concentración de proteínas en hojas cosechadas en invierno y verano sobre plantas masculinas y femeninas. Para cada combinación de sexo y estación, se obtuvieron tres determinaciones del contenido proteico medido como porcentaje del peso seco. Los resultados se presentan en la siguiente tabla.

Tabla 5: Concentración proteica (% del peso seco) en hojas de Atriplex cordobensis cosechadas en invierno y verano de plantas masculinas y femeninas.

Femeninas		Masculinas	
Invierno	Verano	Invierno	Verano
24	17	17	24
28	18	18	25
26	16	16	23

La tabla que presenta los resultados del análisis de la varianza se muestra a continuación. Como puede observarse, ninguno de los factores ensayados muestra por si mismo un efecto significativo sobre la concentración de proteínas pero el término de interacción es altamente significativo, indicando que los factores estudiados efectivamente intervienen en la expresión final de la concentración de proteínas pero que sus efectos no son independientes del nivel del otro factor. La

Figura 2 presenta una representación gráfica de los valores medios en los cuatro tratamientos que permite interpretar fácilmente el resultado mostrado en el cuadro de análisis de la varianza.

Tabla 6: Cuadro de análisis de la varianza para el efecto del sexo y la época de cosecha sobre la concentración de proteínas en hojas de Atriplex cordobensis.

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F
Factor Sexo	3	1	3.00	1.71
Factor Época de cosecha	3	1	3.00	1.71
Interacción Época-Sexo	192	1	192.00	109.71
Dentro	14	8	1.75	
Total	212	11		

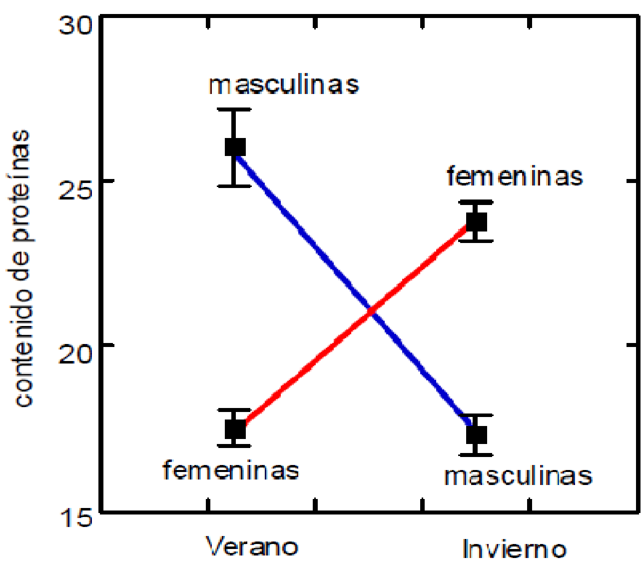

Figura 2: Media ± error estándar de la concentración de proteínas en hojas de Atriplex cordobensis por efecto del sexo y la época de cosecha.

Los modelos con interacción no siempre muestran comportamientos tan extremos como el del ejemplo anterior. De hecho, en muchas situaciones los perfiles de respuesta no se cruzan aunque exista interacción significativa. Un ejemplo se puede apreciar en la Figura 3 que también corresponde al trabajo sobre potencialidad forrajera de A. cordobensis, pero en este caso la variable estudiada es la proporción de fibras insolubles. Las plantas masculinas siempre presentaron mayor contenido de fibras insolubles que las femeninas (efecto principal del factor sexo) pero la diferencia entre femeninas y masculinas fue mayor en el invierno que en el verano (interacción).

De igual modo se puede interpretar el efecto de la época de cosecha, diciendo que en verano el contenido de fibras insolubles fue siempre mayor que en invierno (efecto principal del factor época de cosecha) pero que esta diferencia es más marcada en las plantas femeninas.

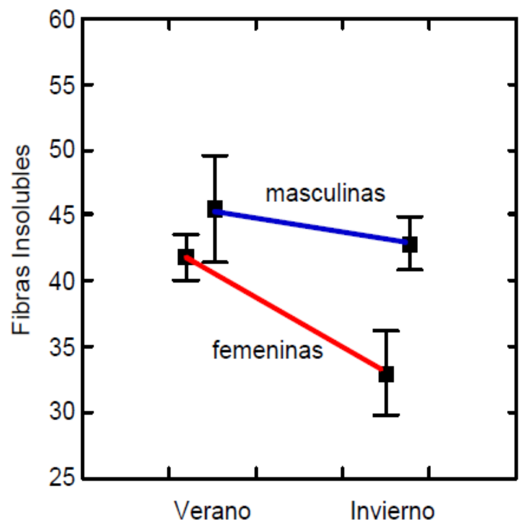

Figura 3: Media ± error estándar de la concentración de fibras insolubles en hojas de Atriplex cordobensis por efecto del sexo y la época de cosecha.

Aunque en los ejemplos anteriores se han presentado experimentos con estructura factorial de tratamientos sólo en diseños completos al azar, la combinación de estructuras factoriales y estructuras de parcela da lugar a una amplia variedad de arreglos experimentales. Así, un experimento como aquel en que se evaluaba el efecto de dos osmolitos y cuatros potenciales agua sobre la germinación se podría haber diseñado con repeticiones. Por diversas razones, quizás no puedan asegurarse las mismas condiciones experimentales de repetición a repetición, porque, por ejemplo, no siempre las cámaras de cultivo regulan de manera similar la temperatura del ensayo, y por lo tanto se tiene una fuente potencial de variación conocida que no es de interés por si misma pero que sí debe incorporarse al diseño y al modelo para eliminarla del error experimental. De este modo, cada repetición podría considerarse un bloque y el experimento completo sería un experimento con estructura bifactorial de tratamiento y estructura de parcelas en bloques completos al azar.

Como se anticipó al comienzo del capítulo, los temas de diseño experimental no se agotan en esta presentación. Queda una importante variedad de tópicos relativos a jerarquías en la estructura de parcelas y en la estructura de tratamientos, métodos de partición de sumas de cuadrados, análisis de la interacción en los modelos no aditivos, análisis de la covarianza, modelos con factores de efectos aleatorios, modelos con mezcla de factores con efectos aleatorios y fijos, diseño del número de repeticiones para alcanzar una potencia deseada, etc, etc, que el lector interesado deberá consultar en una obra más completa de Diseño de Experimentos (Montgomery, 1991; Milliken and Johnson, 1995; Steel y Torrie, 1985).

Ejercicios

Ejercicio 1

El siguiente conjunto de datos corresponde a proteína bruta en leche obtenida con dos suplementos (A y B) en dos dosis (1 y 2). Cada observación corresponde al contenido de proteína bruta en leche de una muestra obtenida de una muestra amalgamada por tambo.

Tambo	Control	A1	A2	B1	B2
I	3.19	3.03	3.06	3.22	3.33
II	3.16	3.07	3.08	3.28	3.20
III	3.25	3.23	3.24	3.45	3.45
IV	3.48	3.30	3.33	3.44	3.39
V	3.25	3.25	3.24	3.35	3.54
VI	3.10	3.05	2.93	3.28	3.35

a) Calcular la estadística descriptiva básica.

b) Identificar el modelo lineal para los datos anteriores.

c) Calcular la tabla de análisis de la varianza y, si corresponde, utilizar alguna técnica de comparaciones múltiples.

d) ¿Qué suplementación se recomendaría si el objetivo es maximizar la concentración de proteína bruta en la leche?

Ejercicio 2

En la siguiente tabla se muestran los resultados de un experimento montado según un diseño completamente aleatorizado con cuatro repeticiones, en el que nemátodos de género Pratylenchus fueron criados en cuatro condiciones de temperatura y discriminados según sexo para evaluar el efecto del sexo y la temperatura sobre la expresión fenotípica de diversos caracteres morfométricos. Los resultados presentados corresponden al largo promedio de la cola en unidades experimentales conformadas por 5 individuos.

	Hembras				**Machos**			
Temp. (ºC)	Rep 1	Rep 2	Rep 3	Rep 4	Rep 1	Rep 2	Rep 3	Rep 4
16	29.2	32.5	34.6	32.6	27.2	24.7	27.3	26.2
21	30.1	30.4	31.4	35.8	26.7	26.5	27.2	27.2
25	31.6	30.2	29.5	30.0	26.2	26.3	28.2	26.2
28	29.6	28.4	28.4	28.1	24.8	25.4	25.6	26.2

a) Identificar el modelo lineal para este experimento.

b) Representar gráficamente los valores medios según sexo y temperatura.

c) Construir la tabla de análisis de la varianza correspondiente.

d) Concluir sobre el efecto de la temperatura y el sexo sobre la expresión del largo de la cola y relacione sus conclusiones con la representación gráfica obtenida en b).

APLICACIONES EN INFOSTAT

Ejemplo 3: En un ensayo comparativo del efecto del estrés hídrico y salino sobre la germinación de Atriplex cordobensis, se sometieron lotes de semillas a cuatro niveles de potencial agua: 0, -0.5, -1.0 y -1.5 Mpa obtenidos mediante la aplicación al medio de dos osmolitos: polietilenglicol (PEG) y cloruro de sodio (ClNa). El experimento se condujo bajo un diseño completamente aleatorizado sin repeticiones. Los resultados se presentan en el archivo Factorial1.

El archivo de datos contiene tres columnas, una identificando al factor tratamiento A (estrés hídrico–potencial agua), otra al factor tratamiento B (estrés salino-osmolitos) y otra a la respuesta observada (germinación). Para este análisis elegir Menú – ESTADÍSTICAS - ANÁLISIS DE LA VARIANZA. Si en la ventana del selector de variables del Análisis de la varianza se declara “Factor A” y “Factor B” como Variable de clasificación y “Germinación” como Variable dependiente, la siguiente ventana Análisis de Varianza señalará que las variables “Factor A” y “Factor B” han sido seleccionadas como variables de clasificación y aparecerán en la subventana Especificación de los términos del modelo. Como existe más de un factor de clasificación aparecerá automáticamente el botón Agregar interacciones. En este caso como no existen repeticiones la interacción no puede ser evaluada y por lo tanto no se debe activar este botón. Al Aceptar (sin agregar interacciones), se abrirá una ventana de Resultados conteniendo la siguiente información:

Tabla 7: Cuadro de análisis de la varianza para un experimento bifactorial. Archivo Factorial1.

Análisis de la varianza

Variable	N	R^2	R^2 Aj	CV
Germinación	8	1.00	0.99	5.43

Cuadro de Análisis de la Varianza (SC Tipo III)

F.V.	SC	gl	CM	F	Valor p
Modelo	6568.50	4	1642.13	182.46	0.0007
Factor A	6518.50	3	2172.83	241.43	0.0004
Factor B	50.00	1	50.00	5.56	0.0997
Error	27.00	3	9.00		
Total	6595.50	7			

El valor p<0.0004, menor al nivel de significación nominal de la prueba (□=0.05), para el

efecto del factor A implica que, en el dominio estudiado, este factor tiene efecto

estadísticamente distinto de cero sobre la germinación promedio. No sucede lo mismo para

el factor B, ya que p=0.0997 es mayor al nivel de significación elegido. La verificación de las suposiciones realizadas sobre el término de error y la comparación de medias de tratamientos generalmente acompañan este tipo de salida. Son necesarias pruebas de comparaciones múltiples de medias para el factor A, ya que el valor p=0.0004 solo rechaza la hipótesis de igualdad de medias entre los 4 niveles de potencial agua, pero no se conoce cual o cuales son diferentes (ver Supuestos de ANAVA, Comparaciones múltiples y Contrastes).

Ejemplo 4: En un estudio sobre la potencialidad forrajera de Atriplex cordobensis, unarbusto que crece en depresiones del chaco árido argentino, se evaluó la concentración de proteínas en hojas cosechadas en invierno y verano sobre plantas masculinas y femeninas. Para cada combinación de sexo y estación, se obtuvieron tres determinaciones del contenido proteico medido como porcentaje del peso seco. Los resultados se encuentran en el archivo Factorial2.

El archivo de datos contiene tres columnas, una identificando al Factor A (sexo), otra al Factor B (estación) y otra a la respuesta observada (concentración de proteínas). Para este análisis elegir Menú – ESTADÍSTICAS - ANÁLISIS DE LA VARIANZA. Si en la ventana del selector de variables del Análisis de varianza se declara "Factor A" y "Factor B" como Variable de clasificación y "Conc. Prot." como Variable dependiente, la siguiente ventana del selector de variables Análisis de Varianza señalará que las variables "Factor A" y "Factor B" han sido seleccionadas como Variables de clasificación y aparecerán en la subventana Especificación de los términos del modelo. Para incluir la interacción entre el Factor A y el Factor B, se deberá activar Agregar interacciones. Al Aceptar se abrirá una ventana de Resultados conteniendo la información provista en la siguiente tabla.

Nota: al activar Agregar interacciones se adicionarán automáticamente, como términos del modelo, todas las posibles interacciones entre las variables de clasificación. Para eliminar cualquier término que no se desea explicitar en la ecuación del modelo y que figura en dicha lista, seleccionar los mismos y presionar la tecla Suprimir. Para no agregar al modelo términos de interacción no deseados, en la subventana Variables de clasificación se pueden seleccionar aquellos términos cuyas interacciones se desean agregar al modelo y luego activar el botón Agregar interacciones.

Tabla 8: Cuadro de análisis de la varianza para un diseño bifactorial. Archivo Factorial2.

Análisis de la varianza

Variable	N	R²	R² Aj	CV
Conc.Prot.	12	0.93	0.91	6.30

Cuadro de Análisis de la Varianza (SC Tipo III)

F.V.	SC	gl	CM	F	Valor p
Modelo	198.00	3	66.00	37.71	<0.0001
Factor A	3.00	1	3.00	1.71	0.2268
Factor B	3.00	1	3.00	1.71	0.2268
Factor A*Factor B	192.00	1	192.00	109.71	<0.0001
Error	14.00	8	1.75		
Total	212.00	11			

Como puede observarse el valor p asociado a la interacción es altamente significativo, indicando que los factores estudiados no actúan independientemente. Por este motivo no se establecerán conclusiones sobre los efectos principales a partir de esta tabla ya que la presencia de interacción podría estar afectando las diferencias promedios. En este caso se deberán comparar las medias de los niveles del factor A dentro de los tratamientos que reciben el mismo nivel del factor B o viceversa (Comparar los niveles de B para cada nivel de A por separado).

Para obtener conclusiones acerca del efecto principal de un factor, en presencia de una interacción significativa, el experimentador debe examinar los niveles de dicho factor, manteniendo fijos los niveles de los otros. La siguiente figura muestra una gráfica, obtenida en InfoStat, de los valores medios en los cuatro tratamientos que permite interpretar fácilmente el resultado mostrado en el cuadro de análisis de la varianza.

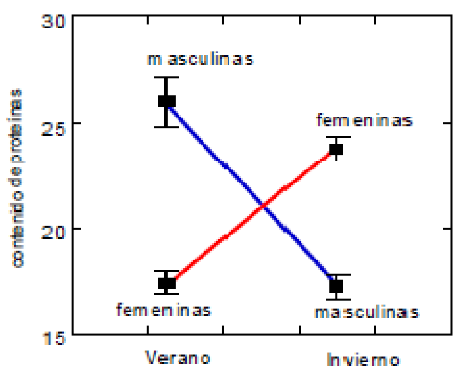

Figura 4: Media±error estándar de la concentración de proteínas en hojas Atriplex cordobensis para cada combinación de niveles de los factores sexo y época de cosecha.

Si alguna combinación de factores está ausente (celdas faltantes), InfoStat presentará automáticamente la suma de cuadrados de tipo I (secuenciales) e informará en la ventana de resultados que se trata de un diseño desbalanceado en celdas. Estas sumas son obtenidas para cada término del modelo secuencial si existe un subconjunto de datos que permite la estimación del contraste de interés (efectos principales e interacción) al menos para los datos existentes.