TEMA 2
Diseño de experimentos: modelos con varios factores

José R. Berrendero
Departamento de Matemáticas
Universidad Autónoma de Madrid

Análisis de Datos - Grado en Biología

Esquema del tema

- Modelo bifactorial aditivo.
- Modelo bifactorial con interacciones.
- Otros modelos y extensiones.

Ejemplo

El abeto posee una regeneración natural escasa por lo que puede resultar conveniente incrementar su producción de semillas mediante un tratamiento hormonal. Se desean comparar cuatro tipos de hormonas A, B, C y D. Las condiciones naturales de reproducción en árboles diferentes no son las mismas. Para controlar este efecto, se han seleccionado 10 árboles y en cada árbol 4 ramas similares. Cada rama recibe exactamente uno de los 4 tratamientos. Tras aplicar los tratamientos se mide el número de semillas producidas en cada rama.

Ejemplo

Árbol	A	B	C	D	Media
1	89	59	20	51	54.75
2	87	56	15	47	51.25
3	84	52	14	45	48.75
4	92	67	26	56	60.25
5	95	70	28	60	63.25
6	90	62	22	53	56.75
7	89	60	19	51	54.75
8	88	56	17	50	52.75
9	82	50	14	45	47.75
10	94	63	24	53	58.50
Media	89.0	59.5	19.9	51.1	54.875

- Parte de la variabilidad en la respuesta hay que asignarla a las diferencias entre los árboles.
- El árbol utilizado es un factor a incluir en el modelo.
- ¿Cómo habría que haber hecho el experimento para que sí fuese apropiado el modelo unifactorial?
- A veces las diferencias entre los niveles de cierto factor no son de interés directo, pero se diseña el experimento teniendo en cuenta este factor para reducir la variabilidad no explicada.
- En este caso al factor se le suele llamar **bloque**.

En el ejemplo:

- El **tratamiento hormonal** es un **factor** con cuatro niveles: A, B, C y D.
- El **árbol** es **otro factor (o bloque)** con 10 niveles: los 10 árboles utilizados.
- El **número de semillas** es la **variable respuesta**.
- Cada combinación lineal de niveles de los factores es un **tratamiento**. En el ejemplo hay 40 tratamientos (10 árboles $\times$ 4 tipos de hormonas).

Estructura de los datos

Y_{ij} es la respuesta correspondiente a la combinación del nivel i del factor α con el nivel j del factor β.

		Factor β				
		1	2	$\cdots$	J	Medias por filas
Factor α	1	Y_{11}	Y_{12}	$\ldots$	Y_{1J}	$\bar{Y}_{1\cdot}$
	2	Y_{21}	Y_{22}	$\ldots$	Y_{2J}	$\bar{Y}_{2\cdot}$
	$\vdots$	$\vdots$			$\vdots$	$\vdots$
	I	Y_{I1}	Y_{I2}	$\ldots$	Y_{IJ}	$\bar{Y}_{I\cdot}$
	Medias por columnas	$\bar{Y}_{\cdot 1}$	$\bar{Y}_{\cdot 2}$		$\bar{Y}_{\cdot J}$	$\bar{Y}_{\cdot\cdot}$

Modelo general

Supongamos que μ_{ij} es la respuesta esperada para los niveles i (factor α) y j (factor β)

El modelo bifactorial general es:

$$Y_{ij} = \mu_{ij} + u_{ij},$$

donde los errores aleatorios u_{ij} verifican las hipótesis habituales, es decir, distribución normal de media 0 y varianza σ^2 e independencia.

Con los datos de los abetos no es posible ajustar el modelo general porque el número de parámetros (41) es mayor que el número de datos (40).

Esto siempre sucede en un experimento con dos factores en el que haya una única respuesta para cada combinación de niveles: IJ datos y $IJ + 1$ parámetros.

Modelo bifactorial aditivo

Una solución es especificar con más detalle el valor de μ_{ij}.

- μ: Respuesta media global.
- α_i: Efecto adicional debido al nivel i del factor α.
- β_j: Efecto adicional debido al nivel j del factor β

Para $i = 1, \ldots, I$ y $j = 1, \ldots, J$,

$$Y_{ij} = \mu + \alpha_i + \beta_j + u_{ij},$$

donde los errores u_{ij} tienen distribución normal de media 0 y varianza σ^2 (la misma para cualquier valor de i y j) y son independientes. Además,

$$\sum_i \alpha_i = \sum_j \beta_j = 0,$$

ya que podemos interpretar α_i y β_j como desviaciones a la media debidas a los niveles de los factores.

Modelo bifactorial aditivo

El modelo se puede escribir:

$$Y_{ij} \equiv \mathrm{N}(\mu + \alpha_i + \beta_j; \sigma),$$

donde $\sum_i \alpha_i = \sum_j \beta_j = 0$ y las variables son independientes.

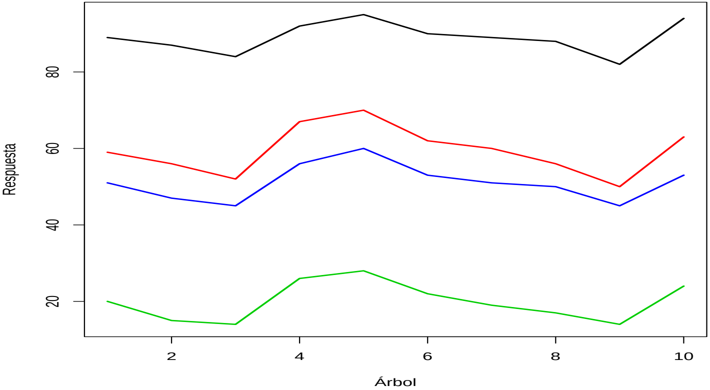
El modelo aditivo tiene $(I-1)+(J-1)+1+1 = I+J$ parámetros. Salvo en el caso $I = J = 2$, siempre $IJ > I + J$.

El modelo aditivo implica que el efecto en la respuesta del nivel de uno de los factores **no depende** del nivel fijado para el otro factor.

Se dice que no hay **interacciones** entre los dos factores.

¿Es razonable un modelo aditivo para los datos de abetos?

Representamos las respuestas como función del árbol (una curva de color diferente para cada tipo de hormona)



Si las curvas son aproximadamente paralelas, el modelo aditivo es razonable.

Estimadores de los parámetros

Parámetro	Estimador
μ	$\hat{\mu} = \bar{Y}_{..}$
α_i	$\hat{\alpha}_i = \bar{Y}_{i.} - \bar{Y}_{..}$
β_j	$\hat{\beta}_j = \bar{Y}_{.j} - \bar{Y}_{..}$

Calcula los estimadores de los parámetros μ y β_j para $j = 1, 2, 3, 4$ en el ejemplo de los abetos.

Los estimadores verifican las mismas restricciones que los parámetros, es decir,

$$\sum_i \hat{\alpha}_i = \sum_j \hat{\beta}_j = 0$$

Residuos y valores ajustados

- **Valores ajustados o pronosticados:** Es la parte de la respuesta que los factores incluidos en el modelo pueden explicar:

$$\hat{Y}_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j = \bar{Y}_{i.} + \bar{Y}_{.j} - \bar{Y}_{..}$$

- **Residuos:** Es la parte de la respuesta que los factores incluidos en el modelo no pueden explicar:

$$e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}$$

Calcula la descomposición $Y_{ij} = \hat{Y}_{ij} + e_{ij}$ para la primera fila de datos del ejemplo (es decir, cuando $i = 1$).

¿Cuánto vale la suma de los residuos de cada fila? ¿Y la suma de los residuos de cada columna?

Variabilidad no explicada y estimación de la varianza

La suma de los residuos al cuadrado mide la variabilidad no explicada o residual:

$$\mathrm{SCR} = \sum_i \sum_j e_{ij}^2 = \sum_i \sum_j (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2$$

Esta suma tiene $(I-1)(J-1)$ gl

Normalizando la suma de cuadrados por sus gl se obtiene un estimador insesgado de la varianza de los errores (la **varianza residual** en este modelo):

$$S_R^2 = \frac{\mathrm{SCR}}{(I-1)(J-1)} = \frac{\sum_i \sum_j (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2}{(I-1)(J-1)}$$

Puede demostrarse que $\mathrm{SCR}/\sigma^2 \equiv \chi^2_{(I-1)(J-1)}$.

Descomposición de la variabilidad

Se verifica:

$$Y_{ij} - \bar{Y}_{..} = \hat{\alpha}_i + \hat{\beta}_j + e_{ij}$$

Si elevamos al cuadrado y sumamos todos estos términos:

$$\sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2 = J \sum_i \hat{\alpha}_i^2 + I \sum_j \hat{\beta}_j^2 + \sum_i \sum_j e_{ij}^2$$

- **Suma de cuadrados total:**

$$\mathrm{SCT} = \sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2$$

- **Suma de cuadrados debida al factor α:**

$$\mathrm{SCE}(\alpha) = J \sum_i \hat{\alpha}_i^2 = J \sum_i (\bar{Y}_{i.} - \bar{Y}_{..})^2$$

- **Suma de cuadrados debida al factor β:**

$$\mathrm{SCE}(\beta) = I \sum_j \hat{\beta}_j^2 = I \sum_j (\bar{Y}_{.j} - \bar{Y}_{..})^2$$

Tabla ANOVA

De acuerdo con las definiciones anteriores, se cumple

$$\text{SCT} = \text{SCE}(\alpha) + \text{SCE}(\beta) + \text{SCR}$$

¿Cuántos gl tiene cada suma de cuadrados?

Tabla ANOVA para el modelo bifactorial aditivo

Fuente	SC	gl	cuadrados medios	F
Factor α	SCE(α)			$F(\alpha)$
Factor β	SCE(β)			$F(\beta)$
Residual	SCR	$(I-1)(J-1)$	SCR$/[(I-1)(J-1)]$	
Total	SCT			

Ejemplo

- Completa la siguiente tabla ANOVA correspondiente a un modelo bifactorial aditivo ajustado a los datos de los abetos:

Fuente	Suma de cuadrados	gl	cuadrados medios	estadístico F
Árbol	886.6			
Hormona			8078.0	
Residual	49.7			
Total				

- ¿Cuál es el estimador de la varianza σ^2 en este caso?
- Calcula un IC de nivel 95 % para σ^2.

Tabla ANOVA con SPSS

Variable dependiente:semillas

Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.
Modelo corregido	25120,700[a]	12	2093,392	1137,827	,000
Intersección	120450,625	1	120450,625	65468,885	,000
arbol	886,625	9	98,514	53,546	,000
hormona	24234,075	3	8078,025	4390,673	,000
Error	49,675	27	1,840		
Total	145621,000	40			
Total corregida	25170,375	39			

a. R cuadrado = ,998 (R cuadrado corregida = ,997)

Tabla ANOVA con SPSS (interpretación)

Modelo corregido	$\mathrm{SCE}(\alpha) + \mathrm{SCE}(\beta) = 25120{,}700$
Intersección	$n\bar{Y}_{..}^2 = 120450, 625$
arbol	$\mathrm{SCE}(\alpha) = 886{,}625$
hormona	$\mathrm{SCE}(\beta) = 24234{,}075$
Error	$\mathrm{SCR} = 49{,}675$
Total	$\sum_i \sum_j Y_{ij}^2 = 145621$
Total corregida	$\mathrm{SCT} = 25170, 375$

Coeficiente de determinación:

$$R^2 = \frac{\mathrm{SCE}(\alpha) + \mathrm{SCE}(\beta)}{\mathrm{SCT}} = \frac{886{,}625 + 24234{,}075}{25170{,}375} = 0{,}998$$

Contrastes sobre el efecto de los factores en la respuesta

¿Tiene el factor α un efecto significativo sobre la variable respuesta Y? Para responder planteamos el siguiente contraste de hipótesis:

$$H_0 : \alpha_1 = \cdots = \alpha_I = 0 \text{ frente a } H_1 : \alpha_i \neq \alpha_j \text{ para algún par } i, j$$

Parece razonable afirmar que el efecto del factor es significativo si la parte de la variabilidad que el factor explica,

$$\mathsf{SCE}(\alpha) = J(\hat{\alpha}_1^2 + \cdots + \hat{\alpha}_I^2),$$

es "suficientemente grande".

Cuando H_0 es cierta se verifica

$$\frac{\mathsf{SCE}(\alpha)}{\sigma^2} \equiv \chi^2_{I-1}.$$

Contrastes sobre el efecto de los factores en la respuesta

Como no conocemos σ^2, lo reemplazamos por su estimador.
Cuando H_0 es cierta,

$$F(\alpha) = \frac{\mathrm{SCE}(\alpha)/(I-1)}{\mathrm{SCR}/[(I-1)(J-1)]} \equiv F_{I-1,(I-1)(J-1)}$$

Un contraste de nivel α para contrastar $H_0 : \alpha_1 = \cdots = \alpha_I = 0$ viene dado por la región crítica:

$$R = \{F(\alpha) > F_{I-1,(I-1)(J-1);\alpha}\}$$

- Escribe la región crítica para contrastar si el factor β tiene un efecto significativo sobre la respuesta.
- Contrasta, para los datos de abetos, si el tipo de hormona tiene un efecto significativo sobre la respuesta.
- Contrasta, para los datos de abetos, si el árbol tiene un efecto significativo sobre la respuesta.

Comparaciones entre dos niveles de un factor

Un IC de nivel $1-\alpha$ para la diferencia $\alpha_i - \alpha_j$ es:

$$\left[(\bar{Y}_{i.} - \bar{Y}_{j.}) \mp t_{(I-1)(J-1),\alpha/2} S_R \sqrt{\frac{1}{J} + \frac{1}{J}}\right]$$

Para contrastar $H_0: \alpha_i = \alpha_j$ frente a $H_1: \alpha_i \neq \alpha_j$ a nivel α, una posibilidad es rechazar H_0 cuando el intervalo anterior no contiene al 0.

¿Cuál es la región crítica que debemos usar para contrastar $H_0: \alpha_i \leq \alpha_j$ frente a $H_1: \alpha_i > \alpha_j$?

Comparaciones múltiples

Para llevar a cabo m comparaciones con un nivel global α_T usamos el método de Bonferroni:

- Calcular $\alpha = \alpha_T / m$.
- Llevar a cabo cada contraste individual a nivel α (para que el nivel de significación global sea α_T).
- Calcular cada IC individual a nivel $1 - \alpha$ (para que el nivel de confianza global sea $1 - \alpha_T$.

Comparaciones múltiples

Comparaciones múltiples

semillas
Bonferroni

(I)hormona	(J)hormona	Diferencia de medias (I-J)	Error típ.	Sig.	Intervalo de confianza 95% Límite inferior	Límite superior
1	2	29,50*	,607	,000	27,77	31,23
	3	69,10*	,607	,000	67,37	70,83
	4	37,90*	,607	,000	36,17	39,63
2	1	-29,50*	,607	,000	-31,23	-27,77
	3	39,60*	,607	,000	37,87	41,33
	4	8,40*	,607	,000	6,67	10,13
3	1	-69,10*	,607	,000	-70,83	-67,37
	2	-39,60*	,607	,000	-41,33	-37,87
	4	-31,20*	,607	,000	-32,93	-29,47
4	1	-37,90*	,607	,000	-39,63	-36,17
	2	-8,40*	,607	,000	-10,13	-6,67
	3	31,20*	,607	,000	29,47	32,93

Basadas en las medias observadas.
El término de error es la media cuadrática(Error) = 1,840.

*. La diferencia de medias es significativa al nivel 0,05.

Modelo bifactorial con interacciones

Introducimos un nuevo modelo que presenta **dos diferencias** respecto al anterior:

- **Permitimos que haya interacciones** entre los dos factores: el efecto en la respuesta de cada nivel de un factor puede depender del nivel que fijamos para el otro factor.
- **El experimento se replica** un número K de veces: para cada combinación de niveles i, j se repite el experimento K veces. Se dispone de $n = IJK$ respuestas.

Para cada respuesta Y_{ijk}, donde $i = 1, \dots, I$, $j = 1, \dots, J$ y $k = 1, \dots, K$ tenemos:

$$Y_{ijk} = \mu_{ij} + u_{ijk},$$

donde las medias μ_{ij} no están restringidas y los errores verifican las hipótesis habituales.

Ejemplo

Se asignaron aleatoriamente 20 individuos con sobrepeso (10 hombres y 10 mujeres) a dos posibles dietas con el fin de estudiar su eficacia. Después de 10 semanas de seguimiento de la dieta, se midió la pérdida de peso (en libras) en cada individuo.

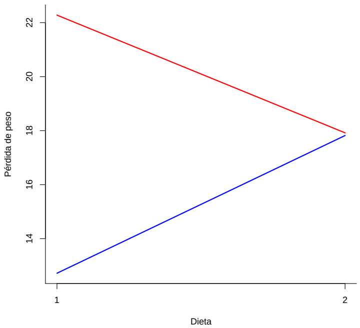
¿Cuánto valen I, J y K en este ejemplo?

	Dieta 1	Dieta 2
Mujeres	7.6	19.5
	8.8	17.6
	12.5	16.8
	16.1	13.7
	18.6	21.5
Hombres	22.2	30.1
	23.4	24.2
	24.2	9.5
	32.2	14.6
	9.4	11.2

Ejemplo

Medias para cada combinación de niveles ($\bar{Y}_{ij.}$):

	Dieta 1	Dieta 2
Mujeres	12.72	17.82
Hombres	22.28	17.92



Modelo bifactorial con interacciones

Vamos a escribir el modelo de manera que los parámetros reflejen los efectos principales de cada factor y las interacciones.

- Respuesta media global: $\mu = \bar{\mu}_{..}$
- Efecto adicional debido al nivel *i* del factor α:

$$\alpha_i = \bar{\mu}_{i.} - \bar{\mu}_{..}$$

- Efecto adicional debido al nivel *j* del factor β:

$$\beta_j = \bar{\mu}_{j.} - \bar{\mu}_{..}$$

- Efecto adicional debido a la interacción entre los niveles i de α y *j* de β:

$$(\alpha\beta)_{ij} = \mu_{ij} - \mu - \alpha_i - \beta_j$$

Descomposición de la media: $\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$.

Modelo bifactorial con interacciones

Para $i = 1, \dots, I$, $j = 1, \dots, J$ y $k = 1, \dots, K$

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + u_{ijk},$$

donde

$$\sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \beta_j = \sum_{i=1}^{I} (\alpha\beta)_{ij} = \sum_{j=1}^{J} (\alpha\beta)_{ij} = 0$$

y las variables u_{ijk} tienen distribución normal de media 0 y varianza σ^2, y son independientes.

- ¿Cuántos parámetros hay que estimar en este modelo?
- ¿Qué ocurre si no se replica el experimento ($K = 1$)?

Estimadores de los parámetros

Parámetro	Estimador
μ	$\hat{\mu} = \bar{Y}_{...}$
α_i	$\hat{\alpha}_i = \bar{Y}_{i..} - \bar{Y}_{...}$
β_j	$\hat{\beta}_j = \bar{Y}_{.j.} - \bar{Y}_{...}$
$(\alpha\beta)_{ij}$	$\widehat{(\alpha\beta)}_{ij} = \bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}$

Calcula todos los estimadores de los parámetros en el ejemplo de las dietas.

Los estimadores verifican las mismas restricciones que los parámetros.

Residuos y valores ajustados

- **Valores ajustados o pronosticados:** Es la parte de la respuesta que los factores incluidos en el modelo pueden explicar:

$$\hat{Y}_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \widehat{(\alpha\beta)}_{ij} = \bar{Y}_{ij.}$$

- **Residuos:** Es la parte de la respuesta que los factores incluidos en el modelo no pueden explicar:

$$e_{ijk} = Y_{ijk} - \hat{Y}_{ijk} = Y_{ijk} - \bar{Y}_{ij.}$$

Calcula la descomposición $Y_{ijk} = \hat{Y}_{ijk} + e_{ijk}$ para las 5 mujeres que siguen la dieta 1 en el ejemplo.

¿Qué restricciones verifican los residuos?

Variabilidad no explicada y estimación de la varianza

La suma de los residuos al cuadrado mide la variabilidad no explicada o residual:

$$\mathrm{SCR} = \sum_i \sum_j \sum_k e_{ijk}^2 = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij.})^2$$

Esta suma tiene $IJ(K-1)$ gl

Normalizando la suma de cuadrados por sus gl se obtiene un estimador insesgado de la varianza de los errores (la **varianza residual** en este modelo):

$$S_R^2 = \frac{\mathrm{SCR}}{IJ(K-1)} = \frac{\sum_i \sum_i \sum_k (Y_{ijk} - \bar{Y}_{ij.})^2}{IJ(K-1)}$$

Puede demostrarse que $\mathrm{SCR}/\sigma^2 \equiv \chi^2_{IJ(K-1)}$.

Descomposición de la variabilidad

Se verifica:

$$Y_{ijk} - \bar{Y}_{...} = \hat{\alpha}_i + \hat{\beta}_j + \widehat{(\alpha\beta)}_{ij} + e_{ijk}$$

Si elevamos al cuadrado y sumamos todos estos términos resulta

$$\text{SCT} = \text{SCE}(\alpha) + \text{SCE}(\beta) + \text{SCE}(\alpha\beta) + \text{SCR},$$

donde:

- **Suma de cuadrados total:**

$$\text{SCT} = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{...})^2$$

- **Suma de cuadrados debida al factor α:**

$$\text{SCE}(\alpha) = JK \sum_i \hat{\alpha}_i^2 = JK \sum_i (\bar{Y}_{i..} - \bar{Y}_{...})^2$$

- **Suma de cuadrados debida al factor β:**

$$\text{SCE}(\beta) = IK \sum_j \hat{\beta}_j^2 = IK \sum_j (\bar{Y}_{.j.} - \bar{Y}_{...})^2$$

Descomposición de la variabilidad

- **Suma de cuadrados debida a las interacciones:**

$$\text{SCE}(\alpha\beta) = K \sum_i \sum_j (\widehat{\alpha\beta})^2_{ij} = K \sum_i \sum_j (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$$

- **Suma de cuadrados residual:**

$$\text{SCR} = \sum_i \sum_j \sum_k e^2_{ijk} = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij.})^2$$

¿Cuántos gl tiene cada una de estas sumas de cuadrados?

Tabla ANOVA para el modelo bifactorial con interacciones

Fuente	SC	gl	cuadrados medios	F
Factor α	SCE(α)	$I-1$	SCE(α)/$(I-1)$	$F(\alpha)$
Factor β	SCE(β)	$J-1$	SCE(β)/$(J-1)$	$F(\beta)$
Interacciones	SCE($\alpha\beta$)	$(I-1)(J-1)$	SCE($\alpha\beta$)/$[(I-1)(J-1)]$	$F(\alpha\beta)$
Residual	SCR	$IJ(K-1)$	SCR/$[IJ(K-1)]$	
Total	SCT	$IJK-1$		

Coeficiente de determinación en este modelo:

$$R^2 = \frac{\text{SCE}(\alpha) + \text{SCE}(\beta) + \text{SCE}(\alpha\beta)}{\text{SCT}}$$

Tabla ANOVA con SPSS

Variable dependiente:Peso

Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.
Modelo corregido	229,194[a]	3	76,398	1,731	,201
Intersección	6255,185	1	6255,185	141,726	,000
Sexo	116,645	1	116,645	2,643	,124
Dieta	,685	1	,685	,016	,902
Sexo * Dieta	111,865	1	111,865	2,535	,131
Error	706,172	16	44,136		
Total	7190,550	20			
Total corregida	935,366	19			

a. R cuadrado = ,245 (R cuadrado corregida = ,103)

Tabla ANOVA con SPSS (interpretación)

Modelo corregido	$SCE(\alpha) + SCE(\beta) + SCE(\alpha\beta) = 229{,}194$
Intersección	$n\bar{Y}^2_{\cdots} = 6255{,}185$
sexo	$SCE(\alpha) = 116{,}645$
dieta	$SCE(\beta) = 0{,}685$
sexo * dieta	$SCE(\alpha\beta) = 111{,}865$
Error	$SCR = 706{,}172$
Total	$\sum_i \sum_j \sum_k Y^2_{ijk} = 7190{,}550$
Total corregida	$SCT = 935{,}366$

Coeficiente de determinación:

$$R^2 = \frac{SCE(\alpha) + SCE(\beta) + SCE(\alpha\beta)}{SCT} = \frac{229{,}194}{7190{,}550} = 0{,}245$$

Contrastes sobre los efectos de los factores en la respuesta

¿Tiene la dieta (factor β) un efecto significativo sobre la respuesta?

Contraste: $H_0 : \beta_1 = \cdots = \beta_J = 0$

La variabilidad debida al factor β es:

$$\mathrm{SCE}(\beta) = IK \sum_j \hat{\beta}_j^2 = IK \sum_j (\bar{Y}_{.j.} - \bar{Y}_{...})^2$$

Parece razonable rechazar H_0 cuando $\mathrm{SCE}(\beta)$ sea suficientemente grande.

Bajo H_0,

$$\frac{\mathrm{SCE}(\beta)}{\sigma^2} \equiv \chi^2_{J-1}.$$

Contrastes sobre los efectos de los factores en la respuesta

Como no conocemos σ^2 lo reemplazamos por su estimador insesgado.

Bajo H_0,

$$F(\beta) = \frac{\mathrm{SCE}(\beta)/(J-1)}{\mathrm{SCR}/[IJ(K-1)]} \equiv F_{J-1,IJ(K-1)}$$

Una región crítica de nivel α es:

$$R = \left\{F(\beta) > F_{J-1,IJ(K-1),\alpha}\right\}$$

El contraste para el otro factor, $H_0 : \alpha_1 = \cdots = \alpha_I = 0$, es totalmente análogo.

Contrastes sobre las interacciones

En este modelo, podemos también preguntarnos si hay interacciones significativas entre los dos factores:

$$H_0 : (\alpha\beta)_{ij} = 0, \quad \text{para todo } i = 1, \ldots, I, \; j = 1, \ldots, J$$

El razonamiento es totalmente análogo al de lo contrastes de los efectos principales. La región crítica de nivel α es:

$$R = \left\{ F(\alpha\beta) > F_{(I-1)(J-1), IJ(K-1), \alpha} \right\},$$

donde

$$F(\alpha\beta) = \frac{\text{SCE}(\alpha\beta)/[(I-1)(J-1)]}{\text{SCR}/[IJ(K-1)]}$$

tiene distribución $F_{(I-1)(J-1), IJ(K-1)}$ bajo H_0.

¿Cuáles son las conclusiones en el ejemplo de las dietas?

Comparaciones múltiples

Para responder a las cuestiones siguientes se razona de manera análoga a los modelos anteriores:

- Escribe la fórmula de un intervalo de confianza de nivel $1 - \alpha$ para la diferencia $\beta_i - \beta_j$ en este modelo.
- Escribe una región crítica para contrastar a nivel α la hipótesis $H_0 : \alpha_i = \alpha_j$.
- ¿Cómo habría que modificar los intervalos si llevamos a cabo m comparaciones y queremos que el nivel de confianza global sea $1 - \alpha_T$?

Reactivos y catalizadores

Se desea investigar el efecto de cuatro tipos de reactivo y de tres catalizadores en la cantidad de producto resultante en una reacción. Se repite la reacción dos veces para cada tipo de reactivo y cada tipo de catalizador y se obtienen las cantidades de producto que aparecen en la tabla siguiente:

	Cat 1	Cat 2	Cat 3
Reactivo A	4-6	11-7	5-9
Reactivo B	6-4	13-15	9-7
Reactivo C	13-15	15-9	13-13
Reactivo D	12-12	12-14	7-9

Tabla ANOVA del modelo bifactorial con interacciones:

Pruebas de los efectos inter-sujetos

Variable dependiente: producto

Fuente	Suma de cuadrados tipo III	gl	Media cuadrática	F	Significación
Modelo corregido	252,000[a]	11	22,909	5,727	,003
Intersección	2400,000	1	2400,000	600,000	,000
cataliza	48,000	2	24,000	6,000	,016
reactivo	120,000	3	40,000	10,000	,001
cataliza * reactivo	84,000	6	14,000	3,500	,031
Error	48,000	12	4,000		
Total	2700,000	24			
Total corregida	300,000	23			

a. R cuadrado = ,840 (R cuadrado corregida = ,693)

Modelo bifactorial con interacciones: algunas cuestiones

- Calcula un estimador del efecto principal del reactivo A.
- Calcula un estimador del efecto principal del catalizador 1.
- Calcula un estimador del parámetro que mide la interacción entre el reactivo A y el catalizador 1.
- Calcula el valor pronosticado para la cantidad de producto obtenida cuando se utiliza el reactivo A y el catalizador 1.
- Calcula los residuos para las observaciones correspondientes al tratamiento del apartado anterior.
- ¿Cuál es la tabla ANOVA correspondiente a un modelo bifactorial aditivo que no incluye interacciones? Escribe este modelo.
- ¿Hay alguna diferencia entre las conclusiones obtenidas usando el modelo aditivo y usando el modelo con interacciones?

Tabla ANOVA para el modelo aditivo

Pruebas de los efectos inter-sujetos

Variable dependiente: producto

Fuente	Suma de cuadrados tipo III	gl	Media cuadrática	F	Significación
Modelo corregido	168,000[a]	5	33,600	4,582	,007
Intersección	2400,000	1	2400,000	327,273	,000
cataliza	48,000	2	24,000	3,273	,061
reactivo	120,000	3	40,000	5,455	,008
Error	132,000	18	7,333		
Total	2700,000	24			
Total corregida	300,000	23			

a. R cuadrado = ,560 (R cuadrado corregida = ,438)

Modelos con tres o más factores

Un modelo general con tres factores, α, β y γ incluye:

- Media global: μ
- Efectos principales: α_i, β_j, γ_k
- Interacciones de segundo orden: $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{ik}$, $(\beta\gamma)_{jk}$
- Interacciones de tercer orden: $(\alpha\beta\gamma)_{ijk}$

SPSS permite seleccionar cuáles de estos parámetros queremos incluir en el modelo y nos da la correspondiente tabla ANOVA, cuyo análisis es análogo al de los modelos que ya hemos visto.

El número de observaciones requeridas para ajustar estos modelos crece muy rápido.

¿Cuántas observaciones son necesarias para ajustar un modelo completo con 3 factores y 4 niveles cada uno?

Diseño por cuadrados latinos

Es un diseño alternativo que requiere menos observaciones.

Se puede aplicar cuando:

- Hay tres factores.
- Todos los factores se presentan con el mismo número de niveles K.
- No hay interacciones entre los factores.

El diseño por cuadrados latinos requiere únicamente K^2 observaciones.

En este diseño cada nivel de un factor aparece una única vez con cada uno de los niveles de los otros factores.

Diseño por cuadrados latinos

4	2	5	7	1	6	8	3	9
3	6	8	5	4	9	1	7	2
1	9	7	3	8	2	6	5	4
8	4	9	1	6	7	5	2	3
5	3	1	2	9	4	7	6	8
2	7	6	8	5	3	4	9	1
9	1	2	6	7	8	3	4	5
6	8	3	4	2	5	9	1	7
7	5	4	9	3	1	2	8	6

Diseño por cuadrados latinos: ejemplo

En un estudio se mide la velocidad de lectura (número de palabras leídas por minuto) bajo distintas circunstancias:

- Tipo de papel: Satinado, Blanco o Color.
- Tamaño de letra: Grande, Normal o Pequeña.
- Iluminación: Natural (A), Muy fuerte (B), Escasa (C).

Diseño del experimento y respuestas obtenidas:

	Papel satinado	Papel blanco	Papel color
Letra grande	258 A	230 C	240 B
Letra normal	235 B	270 A	240 C
Letra pequeña	220 C	225 B	260 A

Diseño por cuadrados latinos: ejemplo

Pruebas de los efectos inter-sujetos

Variable dependiente:Velocidad

Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.
Modelo corregido	2329,333[a]	6	388,222	89,590	,011
Intersección	527076,000	1	527076,000	121632,923	,000
Iluminacion	1938,667	2	969,333	223,692	,004
Tamaño	268,667	2	134,333	31,000	,031
Papel	122,000	2	61,000	14,077	,066
Error	8,667	2	4,333		
Total	529414,000	9			
Total corregida	2338,000	8			

a. R cuadrado = ,996 (R cuadrado corregida = ,985)