Modelo de diseños factoriales y diseños 2^k

Introducción

En el tema anterior se analizaron la posible influencia de un factor sobre la variable respuesta, aleatorizando las observaciones para eliminar el efecto de otros factores. En este capítulo analizaremos modelos en los cuales dos o más factores pueden influir en la variable respuesta. Se emplea la siguiente metodología:

1. Identificar los factores que pueden influir en la variable respuesta y proponer un modelo
2. Realizar el experimento, tomando las observaciones necesarias
3. Estimar los parámetros del modelo
4. Contrastar si los factores influyen en la respuesta
5. Si los factores influyen en la respuesta, detectar dónde radican las diferencias
6. Si algún factor no influye, simplificar el modelo y repetir los pasos anteriores
7. Realizar la diagnosis del modelo mediante el análisis de los residuos

Un diseño factorial es aquél en el que se investigan todas las posibles combinaciones de los niveles de los factores en cada ensayo completo. En este caso se dicen que están cruzados, apareciendo el concepto de interacción.

Se supone la existencia de repeticiones del experimento en cada una de las posibles combinaciones de los niveles del factor correspondiente.

Concepto de interacción

Para ilustrar de forma intuitiva lo que es la interacción vamos a tomar dos conjuntos de datos. Consideramos dos factores: α (niveles α_1 y α_2) y β (niveles β_1 y β_2).

Primer caso: dos factores sin interacción. Los datos son:

$\alpha \backslash \beta$	β_1	β_2
α_1	10	20
α_2	30	40

El efecto principal del factor α es la diferencia entre la respuesta promedio de α_1 y α_2 :

$$E_\alpha = \frac{10+20}{2} - \frac{30+40}{2} = -20$$

y el efecto principal del factor β es:

$$E_\beta = \frac{10+30}{2} - \frac{20+40}{2} = -10$$

Ahora bien, para el nivel β_1, el efecto del factor α es:

$$E_\alpha | \beta_1 = 10 - 30 = -20$$

y para el nivel β_2 es:

$$E_\alpha | \beta_2 = 20 - 40 = -20$$

De forma similar, los efectos del factor β para los niveles α_1 y α_2 son, respectivamente:

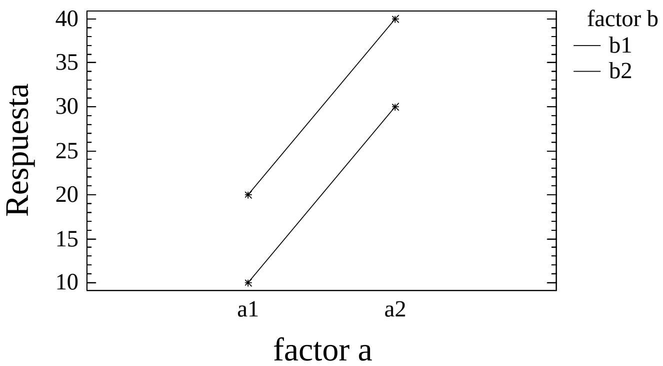
$$E_\beta | \alpha_1 = 10 - 20 = -10$$

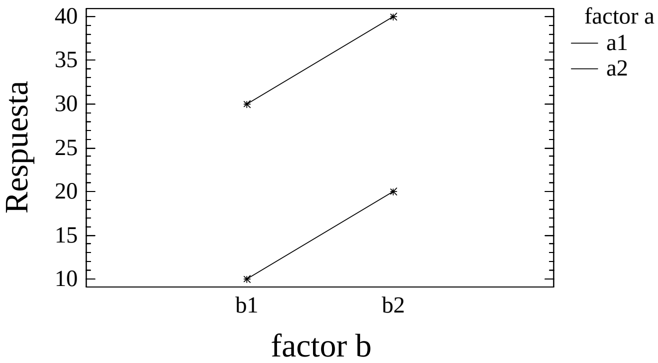
$$E_\beta|\alpha_2 = 30 - 40 = -10$$

Entonces, el efecto de uno de los factores no depende de los niveles del otro factor, lo cual indica que no hay interacción entre los factores. Cuando ambos factores tienen dos niveles, el efecto de la interacción es la diferencia entre los promedios de las diagonales, que es en este caso:

$$E_{\alpha\beta} = \frac{10 + 40}{2} - \frac{30 + 20}{2} = 0$$

lo que indica que no hay interacción. Los siguientes gráficos de perfil muestran la falta de interacción ya que las rectas que aparecen son paralelas.





Segundo caso: dos factores con interacción. Los datos son:

$\alpha\backslash\beta$	β_1	β_2
α_1	10	20
α_2	30	0

El efecto principal del factor α es

$$E_\alpha = \frac{10+20}{2} - \frac{30+0}{2} = 0$$

lo que indicaría que el factor α no tendría ningún efecto en la respuesta. Sin embargo, para el nivel β_1, el efecto del factor α es:

$$E_\alpha|\beta_1 = 10 - 30 = -20$$

y para el nivel β_2 es:

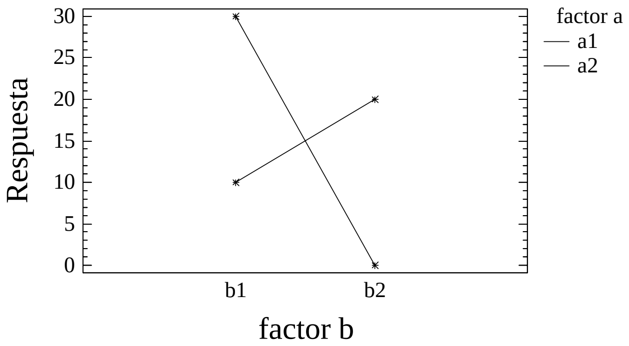
$$E_\alpha|\beta_2 = 20 - 0 = 20$$

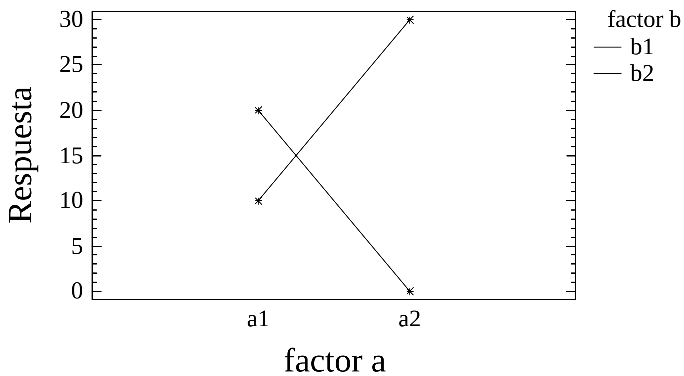
Entonces, aunque el efecto principal indique que el factor α no influye en la respuesta, el efecto que produce α depende del nivel seleccionado del factor β y se concluye que hay interacción entre α y β.

El efecto de la interacción es en este caso:

$$E_{\alpha\beta} = \frac{10+0}{2} - \frac{30+20}{2} = -20$$

lo que indica que hay interacción. Los siguientes gráficos de perfil muestran la existencia de interacción ya que las rectas que aparecen se cruzan entre sí.





En este caso, la variable respuesta Y puede depender también de dos factores α y β, pero éstos a su vez pueden potenciarse o interactuar.

Para comprobar la existencia de interacción, se puede considerar el gráfico de residuos frente a valores previstos de un modelo sin interacción. La idea de este resultado es que los residuos contendrán la influencia de todos aquellos efectos no considerados de forma explícita en el modelo. Por lo tanto, si la interacción es significativa y no ha sido incluida, su efecto se verá en los residuos. La forma de ver la interacción en los residuos es a través de cierta curvatura en la nube de puntos del gráfico de residuos frente a valores previstos. El motivo es que la interacción implica una relación no lineal entre los factores y la variable respuesta; por tanto si hay interacción y el modelo no la incluye, los residuos tendrán una estructura no lineal no incluida en el modelo.

Diseño bifactorial con replicaciones

Se consideran dos factores A y B con a y b niveles respectivamente. Se tienen $a \cdot b$ combinaciones o posibles tratamientos y n observaciones para cada tratamiento, esto es, un diseño balanceado.

El modelo es

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

para $i = 1, \ldots, a$ $j = 1, \ldots, b$, $k = 1, \ldots, n$ donde:

- μ es el efecto medio global.
- α_i es el efecto incremental sobre la media causado por el nivel i del factor A.
- β_j el efecto incremental sobre la media causado por el nivel j del factor B.
- $(\alpha\beta)_{ij}$ el efecto incremental sobre la media causado por la interacción del nivel i del factor A y el nivel j del factor B.
- ε_{ijk} el término de error

Supondremos además que:

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

Se obtiene una tabla con esta forma

$$\begin{array}{c|ccc} & A_1 & \cdots & A_a \\ \hline B_1 & y_{11k} & \cdots & y_{1ak} \\ \vdots & \vdots & \ddots & \vdots \\ B_b & y_{b1k} & \cdots & y_{bak} \end{array}$$

donde $k = 1, \ldots, n$.

Estimación de los parámetros

Se calcula

$$\min_{\mu,\alpha_i,\beta_j,(\alpha\beta)_{ij}} \phi = \min \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(y_{ijk} - \mu - \alpha_i - \beta_j - (\alpha\beta)_{ij}\right)^2$$

sujeto a

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

Se tiene

$$\frac{\partial\phi}{\partial\mu} = -2\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(y_{ijk} - \mu - \alpha_i - \beta_j - (\alpha\beta)_{ij}\right) = 0 \Longrightarrow$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk} - abn\mu = 0 \Longrightarrow$$

$$\hat{\mu} = \frac{1}{abn}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk} = \bar{y}_{\cdots}$$

Para i fijado

$$\frac{\partial\phi}{\partial\alpha_i} = -2\sum_{j=1}^{b}\sum_{k=1}^{n}\left(y_{ijk} - \mu - \alpha_i - \beta_j - (\alpha\beta)_{ij}\right) = 0 \Longrightarrow$$

$$\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk} - bn\mu - bn\alpha_i = 0 \Longrightarrow$$

$$\hat{\alpha}_i = \frac{1}{bn}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk} - \bar{y}_{\cdots} \Longrightarrow$$

$$\hat{\alpha}_i = \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}$$

Análogamente, para j fijado,

$$\hat{\beta}_j = \bar{y}_{\cdot j \cdot} - \bar{y}_{\cdots}$$

Para (i, j) fijado, se deriva respecto de $(\alpha\beta)_{ij}$

$$\begin{aligned}
&\frac{\partial \phi}{\partial (\alpha\beta)_{ij}} = -2 \sum_{k=1}^{n} \left(y_{ijk} - \mu - \alpha_i - \beta_j - (\alpha\beta)_{ij} \right) = 0 \Longrightarrow \\
&\sum_{k=1}^{n} y_{ijk} - n\bar{y}_{\cdots} - n\left(\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}\right) - n\left(\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots}\right) - n\left(\widehat{\alpha\beta}\right)_{ij} = 0 \Longrightarrow \\
&\left(\widehat{\alpha\beta}\right)_{ij} = \bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots}
\end{aligned}$$

Así

$$\hat{y}_{ijk} = \bar{y}_{\cdots} + \left(\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}\right) + \left(\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots}\right) + \left(\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots}\right) = \bar{y}_{ij\cdot}$$

es decir,

$$\hat{y}_{ijk} = \bar{y}_{ij\cdot}$$

El número de parámetros a estimar en total es

$$1 + (a-1) + (b-1) + (a-1)(b-1)$$

ya que la suma de las estimas de las interacciones por filas es igual a 0, con lo cual hay $(b-1)$ términos. Del mismo modo sucede para las columnas, y se obtienen $(a-1)$ términos. En total hay $(a-1)(b-1)$ términos.

Como el número total de observaciones es $N = abn$, entonces el número de grados de libertad es

$$abn - 1 - (a-1) - (b-1) - (a-1)(b-1) = abn - ab = ab(n-1).$$

De este modo, como

$$SCE = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \left(y_{ijk} - \hat{y}_{ijk}\right)^2 = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \left(y_{ijk} - \bar{y}_{ij\cdot}\right)^2$$

entonces la estima de la varianza total es

$$\hat{\sigma}^2 = \frac{1}{ab(n-1)} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \left(y_{ijk} - \bar{y}_{ij\cdot}\right)^2$$

Descomposición de la varianza

La técnica de análisis de la varianza se utilizará en los siguientes contrastes de hipótesis:

$$\begin{cases} H_0 : \alpha_1 = \cdots = \alpha_a \ = 0 \text{ (el factor } A \text{ no influye)} \\ H_1 : \text{algún } \alpha_i \neq 0 \ \text{ (el factor } A \text{ influye)} \end{cases}$$

$$\begin{cases} H_0 : \beta_1 = \cdots = \beta_b = 0 \ \text{ (el factor } B \text{ no influye)} \\ H_1 : \text{algún } \beta_i \neq 0 \ \text{ (el factor } B \text{ influye)} \end{cases}$$

$$\begin{cases} H_0 : (\alpha\beta)_{11} = \cdots = (\alpha\beta)_{ab} = 0 \text{ (no hay interacción)} \\ H_1 : \text{algún } (\alpha\beta)_{ij} \neq 0 \text{ (hay interacción)} \end{cases}$$

Para contrastar estas hipótesis, descomponemos la suma de cuadrados total en la siguiente suma de cuadrados:

$$\begin{aligned} SCT &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{\cdots})^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - N \cdot \bar{y}_{\cdots}^2 = \\ &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - \frac{1}{abn} y_{\cdots}^2 = \\ &= SC_A + SC_B + SC_{AB} + SCE \end{aligned}$$

donde $N = abn$ y

$$\begin{aligned} SC_A &= bn \sum_{i=1}^{a} (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots})^2 = bn \sum_{i=1}^{a} \bar{y}_{i\cdot\cdot}^2 - N \cdot \bar{y}_{\cdots}^2 = \\ &= \frac{1}{bn} \sum_{i=1}^{a} y_{i\cdot\cdot}^2 - \frac{1}{abn} y_{\cdots}^2 \\ &\equiv \text{“Suma de cuadrados explicada debido al factor } A\text{”} \end{aligned}$$

$$\begin{aligned} SC_B &= an \sum_{j=1}^{b} (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots})^2 = an \sum_{j=1}^{b} \bar{y}_{\cdot j\cdot}^2 - N \cdot \bar{y}_{\cdots}^2 = \\ &= \frac{1}{an} \sum_{j=1}^{b} y_{\cdot j\cdot}^2 - \frac{1}{abn} y_{\cdots}^2 \\ &\equiv \text{“Suma de cuadrados explicada debido al factor } B\text{”} \end{aligned}$$

$$
\begin{aligned}
SC_{AB} = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots})^2 = \\
= \quad n\sum_{i=1}^{a}\sum_{j=1}^{b}\left[(\bar{y}_{ij\cdot} - \bar{y}_{\cdots}) - (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}) - (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots})\right]^2 = \\
= \quad n\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{y}_{ij\cdot}^2 - N\cdot\bar{y}_{\cdots}^2 - SC_A - SC_B = \\
\frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b}y_{ij\cdot}^2 - \frac{1}{bn}\sum_{i=1}^{a}y_{i\cdot\cdot}^2 - \frac{1}{an}\sum_{j=1}^{b}y_{\cdot j\cdot}^2 + \frac{1}{abn}y_{\cdots}^2 \\
\equiv \text{“Suma de cuadrados explicada debido a la interacción”}
\end{aligned}
$$

$$
\begin{aligned}
SCE \quad &= \quad \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{ij\cdot})^2 = ab(n-1)\hat{\sigma}^2 \\
&\equiv \text{“Suma de cuadrados residual”}
\end{aligned}
$$

La tabla de análisis de la varianza es:

F. V.	**S. C.**	**G. L.**	**M. C.**	**F**
Factor A	SC_A	$a-1$	$MC_A = \frac{SC_A}{a-1}$	$F_A = \frac{MC_A}{MC_E}$
Factor B	SC_B	$b-1$	$MC_B = \frac{SC_B}{b-1}$	$F_B = \frac{MC_B}{MC_E}$
Interacción	SC_{AB}	$(a-1)(b-1)$	$MC_{AB} = \frac{SC_{AB}}{(a-1)(b-1)}$	$F_{AB} = \frac{MC_{AB}}{MC_E}$
Residual	SCE	$ab(n-1)$	$MC_E = \frac{SCE}{ab(n-1)}$	
Total	SCT	$abn-1$		

Entonces:

1. Rechazamos $H_0 : \alpha_1 = \cdots = \alpha_a = 0$ al nivel α cuando

$$F_A > F_{a-1,ab(n-1);\alpha}$$

2. Rechazamos $H_0 : \beta_1 = \cdots = \beta_b = 0$ al nivel α cuando

$$F_B > F_{b-1,ab(n-1);\alpha}$$

3. Rechazamos $H_0 : (\alpha\beta)_{11} = \cdots = (\alpha\beta)_{ab} = 0$ al nivel α cuando

$$F_{AB} > F_{(a-1)(b-1),ab(n-1);\alpha}$$

Observación: Siempre trataremos de buscar el modelo más sencillo que explique bien la variable respuesta. Por ejemplo, si aceptamos $H_0 : (\alpha\beta)_{11} = \cdots = (\alpha\beta)_{ab} = 0$, concluimos que la interacción no influye de manera apreciable en la respuesta, y pasaríamos a considerar un modelo con dos factores sin interacción y calcularíamos de nuevo la tabla ANOVA. Si además, aceptamos que uno de los factores no influye en la respuesta, lo eliminaríamos del modelo y trabajaríamos con un modelo de un factor.

Ejemplo. Se aplican pinturas tapaporos para aeronaves en superficies de aluminio, con dos métodos: inmersión y rociado. La finalidad del tapaporos es mejorar la adhesión de la pintura, y puede aplicarse en algunas partes utilizando cualquier método. El grupo de ingeniería de procesos responsable de esta operación está interesado en saber si existen diferencias entre tres tapaporos diferentes en cuanto a sus propiedades de adhesión.

Para investigar el efecto que tienen el tipo de pintura tapaporos y el método de aplicación sobre la adhesión de la pintura, se realiza un diseño factorial. Para ello, se pintan tres muestras con cada tapaporo utilizando cada método de aplicación, después se aplica una capa final de pintura y a continuación se mide la fuerza de adhesión. Los datos son los siguientes:

Tapaporos	**Inmersión**	**Rociado**
1	4 4.5 4.3	5.4 4.9 5.6
2	5.6 4.9 5.4	5.8 6.1 6.3
3	3.8 3.7 4	5.5 5 5

Entonces, $a = 3$, $b = 2$, $n = 3$, $N = 18$.

Las medias de las observaciones son:

Tapaporos	**Inmersión**	**Rociado**	$\bar{y}_{i\cdot\cdot}$
1	$\bar{y}_{11\cdot} = 4{,}267$	$\bar{y}_{12\cdot} = 5{,}3$	$28{,}7/6 = 4{,}783$
2	$\bar{y}_{21\cdot} = 5{,}3$	$\bar{y}_{22\cdot} = 6{,}067$	$34{,}1/6 = 5{,}683$
3	$\bar{y}_{31\cdot} = 3{,}833$	$\bar{y}_{32\cdot} = 5{,}167$	$27/6 = 4{,}5$
$\bar{y}_{\cdot j\cdot}$	$40{,}2/9 = 4{,}467$	$49{,}6/9 = 5{,}511$	$\bar{y}_{\cdots} = 89{,}8/18 = 4{,}989$

Las sumas de cuadrados son:

$$\begin{aligned} SCT &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{\cdots})^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - N \cdot \bar{y}_{\cdots}^2 = 4^2 + 4{,}5^2 + \cdots + 5^2 + 5^2 - \\ &\quad 18 \times 4{,}989^2 = 10{,}72 \end{aligned}$$

$$SC_A = bn\sum_{i=1}^{a}\bar{y}_{i\cdot\cdot}^2 - N\cdot\bar{y}_{\cdots}^2 = 6(4{,}783^2 + 5{,}683^2 + 4{,}5^2) - 18 \times 4{,}989^2 = 4{,}58$$

$$SC_B = an\sum_{j=1}^{b}\bar{y}_{\cdot j\cdot}^2 - N\cdot\bar{y}_{\cdots}^2 = 9(4{,}467^2 + 5{,}511^2) - 18 \times 4{,}989^2 = 4{,}91$$

$$\begin{aligned} SC_{AB} &= n\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{y}_{ij\cdot}^2 - N\cdot\bar{y}_{\cdots}^2 - SC_A - SC_B = 3(4{,}267^2 + 5{,}3^2 + 5{,}3^2 + 6{,}067^2 + \\ &\quad 3{,}833^2 + 5{,}167^2) - 18 \times 4{,}989^2 - 4{,}58 - 4{,}91 = 0{,}24 \end{aligned}$$

$$SCE = SCT - SC_A - SC_B - SC_{AB} = 10{,}72 - 4{,}58 - 4{,}91 - 0{,}24 = 0{,}99$$

La tabla ANOVA es:

F.V.	**S.C.**	**G.L.**	**M.C.**	**F**
Tapaporo (A)	4.58	2	2.29	27.7576
Método (B)	4.91	1	4.91	59.5152
Interacción	0.24	2	0.12	1.4545
Error	0.99	12	0.0825	
Total	10.72	17		

$$\begin{aligned} F_{2,12;0,05} &= 3{,}8853 \\ F_{1,12;0,05} &= 4{,}7472 \end{aligned}$$

Por tanto, no hay evidencia de la existencia de interacción entre los factores. Los efectos del tipo de tapaporos y del método de aplicación empleado afectan a la fuerza de adhesión. En este caso, debemos simplificar el modelo, considerando un modelo sin interacción (juntando las sumas de cuadrados de la interacción a las del error), donde la tabla ANOVA sería:

F.V.	**S.C.**	**G.L.**	**M.C.**	**F**
Tapaporo (A)	4.58	2	2.29	26.082
Método (B)	4.91	1	4.91	55.9225
Error	0.99+0.24=1.23	12+2=14	0.0878	
Total	10.72	17		

$$\begin{aligned} F_{2,14;0,05} &= 3{,}7389 \\ F_{1,14;0,05} &= 4{,}6001 \end{aligned}$$

Concluimos que los efectos del tipo de tapaporos y del método de aplicación empleado afectan a la fuerza de adhesión.

Si las medias de los tratamientos son diferentes entre sí se pueden considerar los tests de comparaciones múltiples y de rangos estudentizados, que se vieron para el modelo unifactorial general. Se ha de reemplazar el número de réplicas por nivel del factor (n), por el correspondiente número de elementos de cada casilla. A su vez, los grados de libertad del error $(N-a)$ han de cambiarse en el caso general por $ab(n-1)$.

Utilizaremos el método de Bonferroni para observar diferencias entre el nivel medio de los tres tipos de tapaporos. Como $a=3$, realizaremos $m=\binom{3}{2}=3$ comparaciones. Si utilizamos $\alpha_T=0{,}06$, entonces $\alpha=\frac{0{,}06}{3}=0{,}02$.

$$t_{ab(n-1);\alpha/2} \cdot \hat{\sigma}\sqrt{\frac{1}{bn}+\frac{1}{bn}} = t_{14;0,01}\sqrt{0{,}0878}\sqrt{\frac{1}{6}+\frac{1}{6}} =$$
$$= \quad 2{,}624\sqrt{0{,}0878}\sqrt{\frac{1}{3}} = 0{,}4489.$$

Las diferencias de medias en valor absoluto son:

	$\bar{y}_{2\cdot\cdot} = 5{,}683$	$\bar{y}_{3\cdot\cdot} = 4{,}5$
$\bar{y}_{1\cdot\cdot} = 4{,}783$	0.9	**0.283**
$\bar{y}_{2\cdot\cdot} = 5{,}683$	–	1.183

En negrita aparece el caso en el cual no rechazamos la hipótesis de igualdad de medias. Por tanto, concluimos que el tapaporos 2 es el más eficaz y los tapaporos 1 y 3 son iguales en cuanto a eficacia.

Ejemplo. Supongamos que un ingeniero diseña una batería para su uso en un dispositivo que será sometido a ciertas variaciones extremas de temperatura. El único parámetro de diseño que se puede seleccionar es el material de la cubierta de la batería, y tiene tres alternativas. Cuando el dispositivo se manufactura y se envía al campo, el ingeniero no tiene control sobre los extremos de temperatura a que será expuesto el dispositivo, y sabe por experiencia que es probable que la temperatura influya en la duración efectiva de la batería. Sin embargo, sí es posible controlar la temperatura en el laboratorio de desarrollo de productos para los fines del ensayo.

El ingeniero decide probar los tres materiales de la cubierta a tres niveles de temperatura (15, 70 y 125° F) consistentes en el entorno de uso final del producto. Se prueban cuatro baterías con cada combinación de material de la cubierta y temperatura, y las 36 pruebas se ejecutan al azar. Los datos son los siguientes:

Material	**15° F**		**70° F**		**125° F**	
1	130	155	34	40	20	70
	74	180	80	75	82	58
2	150	188	136	122	25	70
	159	126	106	115	58	45
3	138	110	174	120	96	104
	168	160	150	139	82	60

En este ejemplo $a = 3$, $b = 3$, $n = 4$, $N = 36$. Las medias de las observaciones son:

Material	**15º F**	**70 F**	**125º F**	$\bar{y}_{i\cdot\cdot}$
1	134.75	57.25	57.5	83.17
2	155.75	119.75	49.5	108.33
3	144	145.75	85.5	125.083
$\bar{y}_{\cdot j\cdot}$	144.83	107.583	64.17	$\bar{y}_{\cdots}$=105.53

Las sumas de cuadrados son:

$$SCT = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{\cdots})^2 = (130 - 105{,}53)^2 + (155 - 105{,}53)^2 + \cdots$$
$$+(82 - 105{,}53)^2 + (60 - 105{,}53)^2 = 77646{,}972$$
$$SC_A = bn\sum_{i=1}^{a}\bar{y}_{i\cdot\cdot}^2 - N\cdot\bar{y}_{\cdots}^2 = 12(83{,}17^2 + 108{,}33^2 + 125{,}083^2) -$$
$$-36 \times 105{,}53^2 = 10683{,}722$$
$$SC_B = an\sum_{j=1}^{b}\bar{y}_{\cdot j\cdot}^2 - N\cdot\bar{y}_{\cdots}^2 = 12(144{,}83^2 + 107{,}583^2 + 192{,}5^2) -$$
$$-36 \times 105{,}53^2 = 39118{,}722$$
$$SC_{AB} = n\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{y}_{ij\cdot}^2 - N\cdot\bar{y}_{\cdots}^2 - SC_A - SC_B = 4(134{,}75^2 + 57{,}25^2 + \cdots$$
$$+107{,}583^2 + 192{,}5^2) - 36 \times 105{,}53^2 - 10633{,}167 - 39083{,}167 = 9613{,}778$$
$$SCE = SCT - SC_A - SC_B - SC_{AB} = 18230{,}75$$

La tabla ANOVA es:

F.V.	**S.C.**	**G.L.**	**M.C.**	**F**
Material (A)	10683,722	2	5341.861	7,91
Temperatura (B)	39118,722	2	19559.361	28.97
Interacción	9613,778	4	2403.444	3.56
Error	18230,75	27	675.213	
Total	77646,972	35		

$$F_{2,27;0,05} = 3{,}3541$$
$$F_{4,27;0,05} = 2{,}7278$$

Por tanto, existe una interacción significativa entre los factores. Los efectos del tipo de material y de la temperatura son significativos.

Utilizaremos el método de Bonferroni para detectar diferencias en el nivel medio de los tres tipos de material. Como la interacción es significativa, las comparaciones se realizan en un sólo nivel de temperatura. Vamos a tomar, por ejemplo, el nivel 2 (70º F).

Como $I = 3$, realizaremos $m = \binom{3}{2} = 3$ comparaciones. Si utilizamos $\alpha_T = 0{,}06$, entonces $\alpha = \frac{0{,}06}{3} = 0{,}02$.

$$t_{ab(n-1);\alpha/2} \cdot \hat{\sigma}\sqrt{\frac{1}{n} + \frac{1}{n}} = t_{27;0{,}01}\sqrt{665{,}954}\sqrt{\frac{1}{4} + \frac{1}{4}} = 2{,}473\sqrt{665{,}954}\sqrt{\frac{1}{2}} = 45{,}126.$$

Las diferencias de medias en valor absoluto son:

	$\bar{y}_{22\cdot} = 119{,}75$	$\bar{y}_{32\cdot} = 145{,}75$
$\bar{y}_{12\cdot} = 57{,}25$	62.5	88.5
$\bar{y}_{22\cdot} = 119{,}75$		**26**

En negrita aparece el caso en el cual no rechazamos la hipótesis de igualdad de medias. Concluimos que para el nivel de temperatura 70º F, el voltaje medio producido por los materiales 2 y 3 es el mismo y el producido por el 1 es significativamente menor.

Estudio de las medias individuales

Si se rechaza la igualdad entre los efectos del factor A ó B se pueden considerar tests de recorrido studentizado (para el factor correspondiente), pero no son recomendables cuando aparece interacción significativa. Se puede hacer fijando un nivel concreto de uno de los factores.

Si se rechaza la igualdad entre las interacciones, se pueden contrastar las medias que aparecen $\bar{y}_{ij\cdot}$ en todos los posibles tratamientos.

Diseño bifactorial sin replicaciones

Se puede considerar un diseño en el que se presentas dos factores y sólo se realiza una observación por cada tratamiento.

El modelo es

$$y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ij}$$

para $i = 1, \ldots, a$ $j = 1, \ldots, b$, donde:

- μ es el efecto medio global.

- α_i es el efecto incremental sobre la media causado por el nivel i del factor A.

– β_j el efecto incremental sobre la media causado por el nivel j del factor B.

– $(\alpha\beta)_{ij}$ el efecto incremental sobre la media causado por la interacción del nivel i del factor A y el nivel j del factor B.

– ε_{ij} el término de error

Supondremos además que:

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

Se obtiene una tabla con esta forma

$$\begin{array}{c|ccc}
 & A_1 & \cdots & A_a \\
\hline
B_1 & y_{11} & \cdots & y_{1a} \\
\vdots & \vdots & \ddots & \vdots \\
B_b & y_{b1} & \cdots & y_{ba}
\end{array}$$

En este caso, el número de parámetros a estimar es igual que en el caso del diseño bifactorial replicado:

$$1 + (a-1) + (b-1) + (a-1)(b-1) = ab$$

y como el número de observaciones es ab, entonces no hay grados de libertad suficientes para estimar $\hat{\sigma}^2 = \dfrac{MCE}{ab-ab}$.

Una posible solución es considerar que la interacción es nula

$$(\alpha\beta)_{ij} = 0$$

donde $i = 1, \ldots, a$ $j = 1, \ldots, b$, obteniéndose la siguiente tabla ANOVA

F. V.	**S. C.**	**G. L.**	**F**
Factor A	$SC_A = \frac{1}{b}\sum_{i=1}^{a} y_{i\cdot}^2 - \frac{1}{ab}y_{\cdot\cdot}^2$	$a-1$	$F_A = \frac{\frac{SC_A}{a-1}}{\frac{SCE}{(a-1)(b-1)}}$
Factor B	$SC_B = \frac{1}{a}\sum_{j=1}^{b} y_{\cdot j}^2 - \frac{1}{ab}y_{\cdot\cdot}^2$	$b-1$	$F_B = \frac{\frac{SC_B}{a-1}}{\frac{SCE}{(a-1)(b-1)}}$
Error	$SCE = SCT - SCA - SCB$	$(a-1)(b-1)$	
Total	$SCT = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{1}{ab}y_{\cdot\cdot}^2$	$ab-1$	

Se observa que al suponer interacción nula, el efecto de la interacción y el error experimental se juntan.

Otra alternativa es suponer que el efecto de la interacción es de la forma

$$(\alpha\beta)_{ij} = k\alpha_i\beta_j$$

donde k es una constante desconocida que se determina mediante Regresión.

Se descompone la SCE en dos componentes:

(i) Una componente para la interacción con 1 grado de libertad, de modo que la suma de cuadrados correspondiente es

$$SC_N = \frac{1}{a \cdot b \cdot SC_A \cdot SC_B} \left[\sum_{i=1}^{a} \sum_{j=1}^{b} y_{ij} y_{i\cdot} y_{\cdot j} - y_{\cdot\cdot} \left(SC_A + SC_B + \frac{y_{\cdot\cdot}^2}{ab} \right) \right]^2$$

(ii) Una componente para el error

$$SCE^* = SCE - SC_N$$

con $(a-1)(b-1)-1$ grados de libertad

Se determina

$$F_0 = \frac{SC_N}{\frac{SCE^*}{(a-1)(b-1)-1}}$$

Si $F_0 > F_{1,(a-1)(b-1)-1,\alpha}$ la hipótesis nula de no interacción se rechaza.

Modelo de efectos aleatorios

Supongamos que los a niveles del factor A y los b niveles del factor B son una muestra aleatoria de *poblaciones* de niveles. De este modo, la inferencia se puede extender a todos los posibles niveles mediante el cálculo sobre la muestra anterior.

El modelo es

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

para $i = 1, \ldots, a$ $j = 1, \ldots, b$, $k = 1, \ldots, n$ donde los parámetros del modelo son variables aleatorias independientes entre sí, tales que

$$\begin{aligned}
\alpha_i &\sim N(0, \sigma_\alpha) \\
\beta_j &\sim N(0, \sigma_\beta) \\
(\alpha\beta)_{ij} &\sim N(0, \sigma_{\alpha\beta}) \\
\varepsilon_{ijk} &\sim N(0, \sigma)
\end{aligned}$$

De este modo la varianza de cada observación es

$$Var(y_{ijk}) = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_{\alpha\beta}^2 + \sigma^2$$

que son los componentes de la varianza.

Se puede demostrar que

$$\begin{aligned}
E(MC_A) &= E\left(\frac{SC_A}{a-1}\right) = \sigma^2 + n\sigma_{\alpha\beta}^2 + bn\sigma_\alpha^2 \\
E(MC_B) &= E\left(\frac{SC_B}{b-1}\right) = \sigma^2 + n\sigma_{\alpha\beta}^2 + an\sigma_\beta^2 \\
E(MC_{AB}) &= E\left(\frac{SC_{AB}}{(a-1)(b-1)}\right) = \sigma^2 + n\sigma_{\alpha\beta}^2 \\
E(MCE) &= E\left(\frac{SCE}{ab(n-1)}\right) = \sigma^2
\end{aligned}$$

Utilizando las expresiones anteriores, igualando las medias de cuadrados, se obtiene

$$\begin{aligned}
\hat{\sigma}^2 &= MCE \\
\hat{\sigma}_{\alpha\beta}^2 &= \frac{MC_{AB} - MCE}{n} \\
\hat{\sigma}_\alpha^2 &= \frac{MC_A - MC_{AB}}{bn} \\
\hat{\sigma}_\beta^2 &= \frac{MC_B - MC_{AB}}{an}
\end{aligned}$$

Las hipótesis que se tienen que contrastar son

$$\begin{array}{l} H_0 \equiv \hat{\sigma}_\alpha^2 = 0 \\ H_1 \equiv \hat{\sigma}_\alpha^2 > 0 \end{array}$$

$$\begin{array}{l} H_0 \equiv \hat{\sigma}_\beta^2 = 0 \\ H_1 \equiv \hat{\sigma}_\beta^2 > 0 \end{array}$$

$$\begin{array}{l} H_0 \equiv \hat{\sigma}_{\alpha\beta}^2 = 0 \\ H_1 \equiv \hat{\sigma}_{\alpha\beta}^2 > 0 \end{array}$$

Las sumas de cuadrados se calculan igual que en el modelo de efectos fijos, pero al calcular las medias de cuadrados esperadas, se observa que para contrastar $H_0 \equiv \hat{\sigma}_{\alpha\beta}^2 = 0$ el estadístico de contraste que ha de usarse es

$$F_0 = \frac{MC_{AB}}{MCE}$$

porque si H_0 es verdadera tanto el numerador como el denominador de F_0 tienen como valor esperado σ^2 y si es falsa entonces,

$$F_0 > F_{(a-1)(b-1),ab(n-1),\alpha}$$

rechazándose H_0 a nivel α.

Para contrastar $H_0 \equiv \hat{\sigma}_\alpha^2 = 0$ se usa el contraste

$$F_0 = \frac{MC_A}{MC_{AB}}$$

porque si H_0 es verdadera tanto el numerador como el denominador de F_0 tienen como valor esperado $\sigma^2 + n\sigma_{\alpha\beta}^2$ y si es falsa entonces,

$$F_0 > F_{(a-1),(a-1)(b-1),\alpha}$$

rechazándose H_0 a nivel α.

De modo similar, para contrastar $H_0 \equiv \hat{\sigma}_\beta^2 = 0$ se usa el contraste

$$F_0 = \frac{MC_B}{MC_{AB}}$$

porque si H_0 es verdadera tanto el numerador como el denominador de F_0 tienen como valor esperado $\sigma^2 + n\sigma^2_{\alpha\beta}$ y si es falsa entonces,

$$F_0 > F_{(b-1),(a-1)(b-1),\alpha}$$

rechazándose H_0 a nivel α.

Se observa que el método para contrastar efectos no es el mismo en el caso de que los factores A y B sean fijos.

En cualquier caso, los valores esperados de las medias de cuadrados se deben usar siempre como guía para construir los contrastes.

Ejemplo. En el proceso de fabricación de una pieza de un vehículo se puede elegir un gran número posible de materiales y también un gran número posible de temperaturas a las que puede estar sometida la pieza.

Como diseño del experimento se seleccionan 3 posibles tipos de material al azar y tres posibles temperaturas. Se obtiene la siguiente tabla de tiempos de duración:

Material	**-10º C**		**15º C**		**32º C**	
1	130	155	34	40	20	70
	74	180	80	75	82	58
2	150	188	136	122	25	70
	159	126	106	115	58	45
3	138	110	174	120	96	104
	168	160	150	139	82	60

Las sumas de cuadrados son:

$$SCT = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - \frac{1}{abn}y_{\cdots}^2 = (130^2 + 155^2 + \cdots + 60^2) - \frac{3799^2}{36} = 77646{,}972$$

$$SC_A = \frac{1}{bn}\sum_{i=1}^{a} y_{i\cdot\cdot}^2 - \frac{1}{abn}y_{\cdots}^2 = \frac{1}{12}(998^2 + 1300^2 + 1501^2) - \frac{3799^2}{36} = 10683{,}722$$

$$SC_B = \frac{1}{an}\sum_{j=1}^{b} y_{\cdot j\cdot}^2 - \frac{1}{abn}y_{\cdots}^2 = \frac{1}{12}(1738^2 + 1291^2 + 770^2) - \frac{3799^2}{36} = 39118{,}722$$

$$SC_{AB} = \tfrac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij\cdot}^2 - \tfrac{1}{abn}y_{\cdots}^2 - SC_A - SC_B = \tfrac{1}{4}(539^2 + 229^2 + \cdots + 342^2) -$$
$$- \tfrac{3799^2}{36} - 10633{,}17 - 39083{,}17 = 9613{,}778$$

$$SCE = SCT - SC_A - SC_B - SC_{AB} = 18230{,}75$$

Se obtiene la siguiente tabla ANOVA

F.V.	**S.C.**	**G.L.**	**M.C.**	**F**
Material (A)	10683.722	2	5341.861	$\frac{5341{,}861}{2403{,}444} = 2{,}223$
Temperatura (B)	39118.722	2	19559.361	$\frac{19559{,}361}{2403{,}444} = 8{,}138$
Interacción	9613.778	4	2403.444	$\frac{2403{,}444}{675{,}213} = 3{,}560$
Error	18230.75	27	675.213	
Total	77646.972	35		

Como

$$F_{(a-1)(b-1),ab(n-1),\alpha} = F_{4,27,0'05} = 2{,}73.$$

la interacción entre A y B resulta significativamente distinta de 0: $\sigma^2_{\alpha\beta} > 0$

Por otro lado, como

$$F_{(a-1),(a-1)(b-1),\alpha} = F_{(b-1),(a-1)(b-1),\alpha} = F_{2,4,0'05} = 6{,}94.$$

El efecto del factor A no resulta significativo: se acepta que $\sigma^2_\alpha = 0$

El efecto de B sí resulta significativamente distinto de 0: $\sigma^2_\beta > 0$.

Los componentes de la varianza estimados son

$$\begin{aligned}
\hat{\sigma}^2 &= MCE = 675{,}213 \\
\hat{\sigma}^2_{\alpha\beta} &= \frac{MC_{AB} - MCE}{n} = \frac{2403{,}444 - 675{,}213}{4} = 432{,}06 \\
\hat{\sigma}^2_{\alpha} &= \frac{MC_A - MC_{AB}}{bn} = \frac{5341{,}861 - 2403{,}444}{3 \cdot 4} = 244{,}87 \\
\hat{\sigma}^2_{\beta} &= \frac{MC_B - MC_{AB}}{an} = \frac{19559{,}361 - 2403{,}444}{3 \cdot 4} = 1429{,}7
\end{aligned}$$

Modelo bifactorial mixto

Se puede considerar un modelo de efectos mixtos en el que uno de los factores es fijo y el otro aleatorio:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

donde α_i (para $i = 1, \ldots, a$) corresponde al efecto fijo del factor A, de modo que $\sum_{i=1}^{a} \alpha_i = 0$ y el resto de los parámetros del modelo son variables aleatorias independientes entre sí, tales que

$$\begin{aligned} \beta_j &\sim N(0, \sigma_\beta) \\ (\alpha\beta)_{ij} &\sim N(0, \sigma_{\alpha\beta}) \\ \varepsilon_{ijk} &\sim N(0, \sigma) \end{aligned}$$

para $i = 1, \ldots, a$ $j = 1, \ldots, b$, $k = 1, \ldots, n$

Se puede demostrar que

$$E\left(\frac{SC_A}{a-1}\right) = \sigma^2 + n\sigma^2_{\alpha\beta} + bn\frac{\sum_{i=1}^{a} \alpha_i^2}{a-1}$$
$$E\left(\frac{SC_B}{b-1}\right) = \sigma^2 + n\sigma^2_{\alpha\beta} + an\sigma^2_{\beta}$$
$$E\left(\frac{SC_{AB}}{(a-1)(b-1)}\right) = \sigma^2 + n\sigma^2_{\alpha\beta}$$
$$E\left(\frac{SCE}{ab(n-1)}\right) = \sigma^2$$

Así, para contrastar

$$\begin{array}{l} H_0 \equiv \alpha_i = 0, \qquad (i = 1, \ldots, a) \\ H_1 \equiv \alpha_i \neq 0 \end{array} \Longrightarrow F_0 = \frac{MC_A}{MC_{AB}}$$

$$\begin{array}{l} H_0 \equiv \hat{\sigma}^2_\beta = 0 \\ H_1 \equiv \hat{\sigma}^2_\beta > 0 \end{array} \Longrightarrow F_0 = \frac{MC_B}{MC_{AB}}$$

$$\begin{array}{l} H_0 \equiv \hat{\sigma}^2_{\alpha\beta} = 0 \\ H_1 \equiv \hat{\sigma}^2_{\alpha\beta} > 0 \end{array} \Longrightarrow F_0 = \frac{MC_{AB}}{MCE}$$

y las estimas de efectos y los componentes de la varianza son

$$\begin{aligned}
\hat{\mu} &= \bar{y}_{\cdots} \\
\hat{\alpha}_i &= \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots} \qquad (i = 1, \ldots, a) \\
\hat{\sigma}^2_{\beta} &= \frac{MC_B - MC_{AB}}{an} \\
\hat{\sigma}^2_{\alpha\beta} &= \frac{MC_{AB} - MCE}{n} \\
\hat{\sigma}^2 &= MCE
\end{aligned}$$

Diseños factoriales 2^k

En diseños industriales es frecuente considerar dos niveles para cada uno de los factores que pueden intervenir en el diseño experimental. Un diseño con k factores que tienen dos niveles requiere un número de replicaciones igual a 2^k observaciones. En este tipo de modelos se asume que los efectos son fijos y la aleatorización completa y se consideran las mismas restricciones que en el caso de los diseños factoriales típicos.

El diseño 2^2

Se consideran dos factores A y B con dos niveles:

BAJO: 0

ALTO: 1

Los niveles altos de los factores se representan mediante las letras a y b respectivamente y los niveles bajos se representan por la ausencia de dichas letras. Si ambos niveles son bajos se considera un valor igual a (1).

$$\begin{array}{ccc}
(0,0) & \Longrightarrow & (1) \\
(1,0) & \Longrightarrow & a \\
(0,1) & \Longrightarrow & b \\
(1,1) & \Longrightarrow & ab
\end{array}$$

(1), a, b y ab son las respuestas para las n réplicas. Los efectos medios de A y B son

$$\begin{aligned} A &= \frac{1}{2n}(ab + a - b - (1)) \\ B &= \frac{1}{2n}(ab + b - a - (1)) \end{aligned}$$

Estos valores se obtienen considerando que, por ejemplo, el efecto de A se obtiene como la diferencia entre el nivel alto del factor menos el nivel bajo (en cada caso en relación a los niveles del otro factor): El efecto of A en el nivel bajo de B es $(a - (1))/n$ y el efecto of A en el nivel alto de B es $(ab - b)/n$.

Así, el efecto medio de A es

$$A = \frac{ab + a}{2n} - \frac{b + (1)}{2n} = \frac{1}{2n}(ab + a - b - (1))$$

El efecto de la interacción AB se define como la diferencia media entre el efecto de A al nivel alto de B, y el efecto de A al nivel bajo de B :

$$\begin{aligned} AB &= \frac{1}{2n}\left[(ab - b) - (a - (1))\right] = \\ & \frac{1}{2n}\left[ab + (1) - a - b\right]. \end{aligned}$$

Del mismo modo se puede definir BA, obteniéndose que $AB = BA$.

En general, se trata de medir la importancia y el efecto de los factores que intervienen, en términos de la magnitud y del signo de los efectos anteriores.

La sumas de cuadrados se pueden definir en términos, también, de las estimas anteriores:

$$\text{Suma de Cuadrados del Factor} = \frac{1}{n\sum_{i=1}^{a} c_i^2}\left[\sum_{i=1}^{a} c_i y_{i\cdot}\right]^2.$$

Así, en este caso,

$$\begin{aligned} SCA &= \frac{\left[ab + a - b - (1)\right]^2}{4n}, \\ SCB &= \frac{\left[ab + b - a - (1)\right]^2}{4n}, \\ SCAB &= \frac{\left[ab + (1) - a - b\right]^2}{4n}, \end{aligned}$$

La suma de cuadrados total es, como habitualmente,

$$SCT = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{\dots}^2}{4n}$$

y tiene $(2 \cdot 2 \cdot n) - 1$ grados de libertad.

La suma de cuadrados del error es

$$SCE = SCT - SCA - SCB - SCAB.$$

y SCE tiene $4(n-1)$ grados de libertad.

La tabla de análisis de la varianza es, entonces,

F. V.	**S. C.**	**G. L.**	**M. C.**	**F**
Factor A	SC_A	1		$F_A = \frac{SC_A}{MC_E}$
Factor B	SC_B	1		$F_B = \frac{SC_B}{MC_E}$
Interacción	SC_{AB}	1		$F_{AB} = \frac{SC_{AB}}{MC_E}$
Residual	SCE	$4(n-1)$	$MC_E = \frac{SCE}{4(n-1)}$	
Total	SCT	$4n-1$		

Ejemplo

Se trata de estudiar la influencia de los factores:

Temperatura (**1**: alta ó **0**: baja) y

Catalizador (**1**: se usa ó **0**: no se usa)

en la variable respuesta: *dureza* de un material cerámico. Los datos son:

	Replicación			
Combinación	**1**	**2**	**Respuesta total**	**Codificación**
$(0,0)$	86	92	178	(1)
$(0,1)$	47	39	86	a
$(1,0)$	104	114	218	b
$(1,1)$	141	153	294	ab
			$y_{\dots} = 776$	

Los efectos medios y las medias de cuadrados son

$$\begin{aligned} A &= \frac{294 + 86 - 218 - 178}{4} = -4 \\ B &= \frac{294 + 218 - 86 - 178}{4} = 62 \\ AB &= \frac{294 + 178 - 86 - 218}{4} = 42. \end{aligned}$$

$$
\begin{aligned}
SCA &= \frac{(4A)^2}{4\cdot 2} = 32 \\
SCB &= \frac{(4B)^2}{4\cdot 2} = 7688 \\
SCAB &= \frac{(4AB)^2}{4\cdot 2} = 3528
\end{aligned}
$$

$$
\begin{aligned}
SCT &= (86^2 + \cdots + 153^2) - \frac{776^2}{8} = 11420 \\
SCE &= 172.
\end{aligned}
$$

La tabla de análisis de la varianza es

F. V.	**S. C.**	**G. L.**	**M. C.**	**F**
Factor A	32	1	32	$F_A = 0{,}74$
Factor B	7688	1	7688	$F_B = 178{,}79$
Interacción	3528	1	3528	$F_{AB} = 82{,}05$
Residual	172	4	43	
Total	11420	7		

De modo que el factor B y la interacción entre A y B son significativos al nivel 0,05, ya que $F_{1,4;0'05} = 7{,}71$.

El diseño 2^3

Se introduce un breve resumen de este modelo. Supongamos que se tienen tres factores binarios A, B y C. El número de posibles combinaciones es 8, y con n replicaciones se tiene un total de $8n$ observaciones.

Para calcular los efectos se puede usar la siguiente tabla o *matriz de diseño*:

Efecto	**Combinación de Factores**							
Factorial	(1)	a	b	ab	c	ac	bc	abc
I	+	+	+	+	+	+	+	+
A	−	+	−	+	−	+	−	+
B	−	−	+	+	−	−	+	+
AB	+	−	−	+	+	−	−	+
C	−	−	−	−	+	+	+	+
AC	+	−	+	−	−	+	−	+
BC	+	+	−	−	−	−	+	+
ABC	−	+	+	-	+	−	−	+

La primera fila es la identidad y cualquier fila multiplicada por ella permanece invariante. El resto de filas tiene el mismo número de signos $+$ y signos $-$. Se pueden obtener los contrastes y los efectos sustituyendo los signos $+$ por 1 y los $-$ por -1, como se ve a continuación.

Por otro lado se pueden obtener las distintas filas a partir del producto entre ellas, por ejemplo:

$$\begin{aligned} A \cdot B &= AB, \\ (AB) \cdot (B) &= A \cdot B^2 = A, \\ (AC) \cdot (BC) &= A \cdot C^2 \cdot B = AB \end{aligned}$$

Estimación de los efectos:

Los efectos medios se calculan a partir de los contrastes indicados en la tabla anterior partidos entre $4n$:

$$\begin{aligned} A &= \frac{1}{4n}\left[a-(1)+ab-b+ac-c+abc-bc\right] \\ B &= \frac{1}{4n}\left[b+ab+bc+abc-(1)-a-c-ac\right] \\ C &= \frac{1}{4n}\left[c+ac+bc+abc-(1)-a-b-ab\right] \\ AB &= \frac{1}{4n}\left[(1)+ab+c+abc-a-b-ac-bc\right] \\ AC &= \frac{1}{4n}\left[(1)+b+ac+abc-a-ab-c-bc\right] \\ BC &= \frac{1}{4n}\left[(1)+a+bc+abc-b-ab-c-ac\right] \\ ABC &= \frac{1}{4n}\left[abc+a+b+c-ab-ac-bc-(1)\right] \end{aligned}$$

Las sumas de los cuadrados son, en cada caso, de manera semejante al diseño 2^2,

$$SC_{Efec} = \frac{Contraste^2}{8n}$$

Ejemplo

Supongamos la siguiente tabla con $n = 2$ réplicas

	Factor B			
	0		**1**	
Factor A	**Factor C**		**Factor C**	
	0	**1**	**0**	**1**
0	4	7	20	10
	5	9	14	6
1	4	2	4	14
	11	7	6	16

Se tiene que

$$\begin{array}{llll} (1) = 9 & c = 16 & b = 34 & bc = 16 \\ a = 15 & ac = 9 & ab = 10 & abc = 30 \end{array}$$

$$\begin{aligned}
A &= \frac{1}{8}[15 - 9 + 10 - 34 + 9 - 16 + 30 - 16] = -\frac{11}{8} = -1{,}375 \\
B &= \frac{1}{8}[34 + 10 + 16 + 30 - (9 + 15 + 16 + 9)] = \frac{41}{8} = 5{,}125 \\
C &= \frac{1}{8}[16 + 9 + 16 + 30 - (9 + 15 + 34 + 10)] = \frac{3}{8} = 0{,}375 \\
AB &= \frac{1}{8}[9 + 10 + 16 + 30 - (15 + 34 + 9 + 16)] = -\frac{9}{8} = -1{,}125 \\
AC &= \frac{1}{8}[9 + 34 + 9 + 30 - (15 + 10 + 16 + 16)] = \frac{25}{8} = 3{,}125 \\
BC &= \frac{1}{8}[9 + 15 + 16 + 30 - (34 + 10 + 16 + 9)] = \frac{1}{8} = 0{,}125 \\
ABC &= \frac{1}{8}[30 + 15 + 34 + 16 - (10 + 9 + 16 + 9)] = \frac{51}{8} = 6{,}375
\end{aligned}$$

Como, en cada caso,

$$SC_{Efec} = \frac{Contraste^2}{8n}$$

se tiene que

$$\begin{aligned}
SC_A &= \frac{1}{16}11^2 = 7{,}56 \\
SC_B &= \frac{1}{16}41^2 = 105{,}06 \\
SC_C &= \frac{1}{16}3^2 = 0{,}56 \\
SC_{AB} &= \frac{1}{16}9^2 = 5{,}06 \\
SC_{AC} &= \frac{1}{16}25^2 = 39{,}06 \\
SC_{BC} &= \frac{1}{16}1^2 = 0{,}06 \\
SC_{ABC} &= \frac{1}{16}51^2 = 162{,}56
\end{aligned}$$

y

$$\begin{aligned}
SCT &= (4^2 + 5^2 + \cdots + 14^2 + 16^2) - \frac{1}{16}139^2 = 389{,}44 \\
SCE &= 69{,}52
\end{aligned}$$

La tabla de análisis de la varianza es

F. V.	**S. C.**	**G. L.**	**M. C.**	**F**
Factor A	7,56	1	7,56	0,87
Factor B	105,06	1	105,06	12,09*
Interacción AB	5,06	1	5,06	0,58
Factor C	0,56	1	0,56	0,06
Interacción AC	39,06	1	39,06	4,49
Interacción BC	0,06	1	0,06	0,01
Interacción ABC	162,56	1	162,56	18,71*
Residual	69,52	8	8,62	
Total	389,44	15		

Como el valor de la F de Snedecor $F_{1,8,0,05} = 5{,}32$, entonces los valores marcados con $*$ son significativos a nivel 0,05.

Aplicación con R

Se puede usar la librería **Rcmdr** de **R**:

```
library(Rcmdr)
Commander()

Datos <- read.table("C:/CursoCIII/Disenno/Practicas06/dat2Fac.txt", header=TRUE,
sep="", na.strings="NA", dec=".", strip.white=TRUE)

Datos$Mat <- as.factor(Datos$Mat)
Datos$Temp <- factor(Datos$Temp, labels=c('-10','15','32'))

Anova(lm(Resist ~ Mat*Temp, data=Datos))

tapply(Datos$Resist, list(Mat=Datos$Mat, Temp=Datos$Temp), mean, na.rm=TRUE) #
means

tapply(Datos$Resist, list(Mat=Datos$Mat, Temp=Datos$Temp), sd, na.rm=TRUE) #
std. deviations

tapply(Datos$Resist, list(Mat=Datos$Mat, Temp=Datos$Temp), function(x)
sum(!is.na(x))) # counts

plotMeans(Datos$Resist, Datos$Mat, Datos$Temp, error.bars="none")

#...................................................................................

cosa <- read.table("C:/... /dat2Fac.txt", header=TRUE)
attach(cosa)

Mat <- as.factor(Mat)
Temp <- as.factor(Temp)

par(pty="s")
par(mfrow=c(2,2))
plot.design(Resist ~ Mat*Temp)
plot.design(Resist ~ Mat*Temp, fun=median)
plot(Resist ~ Mat*Temp)

modelo <- aov(Resist ~ Mat*Temp)
summary(modelo)
par(pty="s")
par(mfrow=c(2,2))
plot(modelo)

interaction.plot(Mat,Temp,Resist,type="l",xlab="Material",trace.label="temperatu
ra",col=1:3)

# Modelo de efectos aleatorios
summary(aov(Resist ~ Mat+Temp + Error(Mat:Temp)))
summary(aov(Resist ~ Mat:Temp + Error(Mat+Temp)))
```

Aplicación con SAS

```
options ps=66 ls=80 nodate;
title 'ANOVA bifactorial de efectos fijos';
data ano;
input catilaz presion precip;
datalines;
1 1  11
1 1   2
1 1   9
1 2   8
1 2  10
1 2  10
1 3  12
1 3  10
1 3  13
1 4   9
1 4  11
1 4  10
2 1  13
2 1  11
2 1  14
2 2  14
2 2  10
2 2  10
2 3   8
2 3  12
2 3  10
2 4   9
2 4   9
2 4   8
3 1   9
3 1   9
3 1   9
3 2  10
3 2   8
3 2  11
3 3  11
3 3  11
3 3   9
3 4   7
3 4  11
3 4   6
;
proc glm;
class catilaz presion;
model precip=catilaz presion catilaz*presion;
run;
quit;

proc means noprint;
var precip;
by catilaz presion;
output out=outmean mean=mn;
run;
quit;

symbol i=join;
proc gplot;
plot mn*presion=catilaz;
```

```
run;
quit;

/* ------------------------------------------------------- */

proc glm;
class catilaz presion;
model precip=catilaz presion catilaz*presion;
lsmeans catilaz | presion /tdiff adjust=tukey;
run;
quit;

/* ------------------------------------------------------- */

proc glm;
class catilaz presion;
model precip=catilaz presion catilaz*presion;
output out=diag r=res p=pred;
run;
quit;

symbol v=circle;
proc univariate noprint;
qqplot res / normal (l=1 mu=0 sigma=est);
hist res / normal (l=1 mu=0 sigma=est);
run;
quit;

proc gplot;
plot res*catilaz;
plot res*presion;
plot res*pred;
run;
quit;
```

Dependent Variable: precip

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	77.6666667	7.0606061	1.71	0.1325
Error	24	99.3333333	4.1388889		
Corrected Total	35	177.0000000			

R-Square	Coeff Var	Root MSE	precip Mean
0.438795	20.68908	2.034426	9.833333

Source	DF	Type I SS	Mean Square	F Value	Pr > F
catilaz	2	13.16666667	6.58333333	1.59	0.2245
presion	3	15.22222222	5.07407407	1.23	0.3219
catilaz*presion	6	49.27777778	8.21296296	1.98	0.1078

Source	DF	Type III SS	Mean Square	F Value	Pr > F
catilaz	2	13.16666667	6.58333333	1.59	0.2245
presion	3	15.22222222	5.07407407	1.23	0.3219
catilaz*presion	6	49.27777778	8.21296296	1.98	0.1078

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey

catilaz	precip LSMEAN	LSMEAN Number
1	9.5833333	1
2	10.6666667	2
3	9.2500000	3

Least Squares Means for Effect catilaz
t for H0: LSMean(i)=LSMean(j) / Pr > |t|

Dependent Variable: precip

i/j	1	2	3
1		-1.30436	0.40134
		0.4065	0.9154
2	1.304355		1.705695
	0.4065		0.2237
3	-0.40134	-1.7057	
	0.9154	0.2237	

Least Squares Means
Adjustment for Multiple Comparisons: Tukey

presion	precip LSMEAN	LSMEAN Number
1	9.6666667	1
2	10.1111111	2
3	10.6666667	3
4	8.8888889	4

Least Squares Means for Effect presion
t for H0: LSMean(i)=LSMean(j) / Pr > |t|

Dependent Variable: precip

i/j	1	2	3	4
1		-0.46343	-1.04271	0.810998
		0.9663	0.7265	0.8486
2	0.463428		-0.57928	1.274426
	0.9663		0.9373	0.5874
3	1.042712	0.579284		1.85371
	0.7265	0.9373		0.2741
4	-0.811	-1.27443	-1.85371	
	0.8486	0.5874	0.2741	

Least Squares Means
Adjustment for Multiple Comparisons: Tukey

catilaz	presion	precip LSMEAN	LSMEAN Number
1	1	7.3333333	1
1	2	9.3333333	2
1	3	11.6666667	3
1	4	10.0000000	4
2	1	12.6666667	5
2	2	11.3333333	6
2	3	10.0000000	7
2	4	8.6666667	8
3	1	9.0000000	9
3	2	9.6666667	10
3	3	10.3333333	11
3	4	8.0000000	12

Least Squares Means for Effect catilaz*presion
t for H0: LSMean(i)=LSMean(j) / Pr > |t|

Dependent Variable: precip

i/j	1	2	3	4	5	6
1		-1.20402	-2.60871	-1.60536	-3.21072	-2.40804
		0.9834	0.3277	0.8908	0.1131	0.4372
2	1.20402		-1.40469	-0.40134	-2.0067	-1.20402
	0.9834		0.9514	1.0000	0.6852	0.9834
3	2.60871	1.40469		1.00335	-0.60201	0.20067
	0.3277	0.9514		0.9961	1.0000	1.0000
4	1.60536	0.40134	-1.00335		-1.60536	-0.80268
	0.8908	1.0000	0.9961		0.8908	0.9995
5	3.21072	2.0067	0.60201	1.60536		0.80268
	0.1131	0.6852	1.0000	0.8908		0.9995
6	2.40804	1.20402	-0.20067	0.80268	-0.80268	
	0.4372	0.9834	1.0000	0.9995	0.9995	

Least Squares Means for Effect catilaz*presion
t for H0: LSMean(i)=LSMean(j) / Pr > |t|

Dependent Variable: precip

i/j	7	8	9	10	11	12
1	-1.60536	-0.80268	-1.00335	-1.40469	-1.80603	-0.40134
	0.8908	0.9995	0.9961	0.9514	0.7999	1.0000
2	-0.40134	0.40134	0.20067	-0.20067	-0.60201	0.80268
	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995
3	1.00335	1.80603	1.60536	1.20402	0.80268	2.20737
	0.9961	0.7999	0.8908	0.9834	0.9995	0.5598
4	0	0.80268	0.60201	0.20067	-0.20067	1.20402
	1.0000	0.9995	1.0000	1.0000	1.0000	0.9834
5	1.60536	2.40804	2.20737	1.80603	1.40469	2.80938
	0.8908	0.4372	0.5598	0.7999	0.9514	0.2368
6	0.80268	1.60536	1.40469	1.00335	0.60201	2.0067
	0.9995	0.8908	0.9514	0.9961	1.0000	0.6852

Least Squares Means
Adjustment for Multiple Comparisons: Tukey

Least Squares Means for Effect catilaz*presion
t for H0: LSMean(i)=LSMean(j) / Pr > |t|

Dependent Variable: precip

i/j	1	2	3	4	5	6
7	1.60536	0.40134	-1.00335	0	-1.60536	-0.80268
	0.8908	1.0000	0.9961	1.0000	0.8908	0.9995
8	0.80268	-0.40134	-1.80603	-0.80268	-2.40804	-1.60536
	0.9995	1.0000	0.7999	0.9995	0.4372	0.8908
9	1.00335	-0.20067	-1.60536	-0.60201	-2.20737	-1.40469
	0.9961	1.0000	0.8908	1.0000	0.5598	0.9514
10	1.40469	0.20067	-1.20402	-0.20067	-1.80603	-1.00335
	0.9514	1.0000	0.9834	1.0000	0.7999	0.9961
11	1.80603	0.60201	-0.80268	0.20067	-1.40469	-0.60201
	0.7999	1.0000	0.9995	1.0000	0.9514	1.0000
12	0.40134	-0.80268	-2.20737	-1.20402	-2.80938	-2.0067
	1.0000	0.9995	0.5598	0.9834	0.2368	0.6852

Least Squares Means for Effect catilaz*presion
t for H0: LSMean(i)=LSMean(j) / Pr > |t|

Dependent Variable: precip

i/j	7	8	9	10	11	12
7		0.80268	0.60201	0.20067	-0.20067	1.20402
		0.9995	1.0000	1.0000	1.0000	0.9834
8	-0.80268		-0.20067	-0.60201	-1.00335	0.40134
	0.9995		1.0000	1.0000	0.9961	1.0000
9	-0.60201	0.20067		-0.40134	-0.80268	0.60201
	1.0000	1.0000		1.0000	0.9995	1.0000
10	-0.20067	0.60201	0.40134		-0.40134	1.00335
	1.0000	1.0000	1.0000		1.0000	0.9961
11	0.20067	1.00335	0.80268	0.40134		1.40469
	1.0000	0.9961	0.9995	1.0000		0.9514
12	-1.20402	-0.40134	-0.60201	-1.00335	-1.40469	
	0.9834	1.0000	1.0000	0.9961	0.9514	

The UNIVARIATE Procedure
Fitted Distribution for res

Parameters for Normal Distribution

Parameter	Symbol	Estimate
Mean	Mu	0
Std Dev	Sigma	1.661102

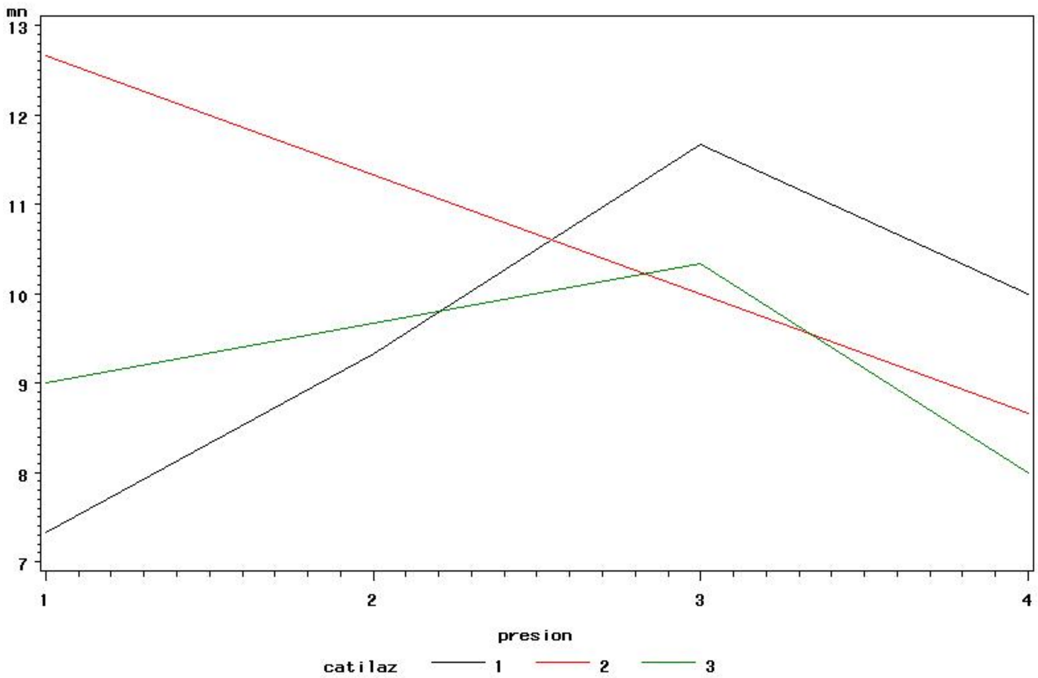
Goodness-of-Fit Tests for Normal Distribution

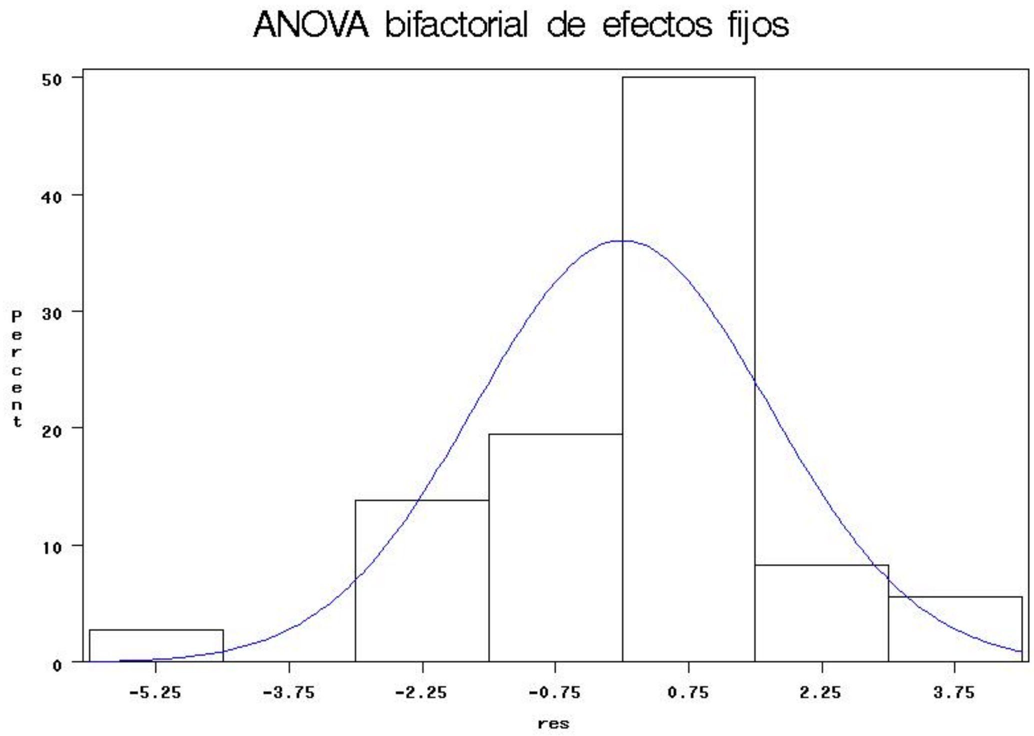
Test	---Statistic----		-----p Value-----	
Cramer-von Mises	W-Sq	0.07858188	Pr > W-Sq	>0.250
Anderson-Darling	A-Sq	0.51155988	Pr > A-Sq	>0.250

Quantiles for Normal Distribution

	------Quantile------	
Percent	Observed	Estimated
1.0	-5.33333	-3.864301
5.0	-2.00000	-2.732269
10.0	-1.66667	-2.128788
25.0	-1.33333	-1.120396
50.0	0.16667	-0.000000
75.0	0.83333	1.120396
90.0	2.00000	2.128788
95.0	3.00000	2.732269
99.0	3.66667	3.864301

ANOVA bifactorial de efectos fijos




ANOVA bifactorial de efectos fijos
percent
res
50
40
30
20
10
0
-5.25
-3.75
-2.25
-0.75
0.75
2.25
3.75

```
options ls=72 nodate nonumber;
title 'ANOVA BIFACTORIAL DE EFECTOS ALEATORIOS';
data ano;
input zona $ tienda ventas;
cards;
A 1  59
A 1  61
A 1  61
A 1  59
A 2  60
A 2  63
A 2  55
A 2  57
A 3  63
A 3  67
A 3  65
A 3  60
A 4  66
A 4  64
A 4  61
A 4  68
B 1  61
B 1  64
B 1  67
B 1  62
B 2  69
B 2  64
B 2  62
B 2  69
B 3  66
B 3  71
B 3  68
B 3  68
B 4  62
B 4  69
B 4  68
B 4  71
C 1  71
C 1  70
C 1  68
C 1  74
C 2  66
C 2  72
C 2  67
C 2  72
C 3  75
C 3  69
C 3  76
C 3  68
C 4  69
C 4  70
C 4  69
C 4  77
;
proc glm;
class zona tienda ;
model ventas= zona|tienda  ;
random  zona tienda zona*tienda/test;
run;

proc varcomp method=type1;
class zona tienda ;
```

```
model ventas= zona|tienda  ;
run;
```

Dependent Variable: ventas

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	817.062500	74.278409	7.53	<.0001
Error	36	355.250000	9.868056		
Corrected Total	47	1172.312500			

R-Square	Coeff Var	Root MSE	ventas Mean
0.696966	4.737186	3.141346	66.31250

Source	DF	Type I SS	Mean Square	F Value	Pr > F
zona	2	648.0000000	324.0000000	32.83	<.0001
tienda	3	123.7291667	41.2430556	4.18	0.0123
zona*tienda	6	45.3333333	7.5555556	0.77	0.6017

Source	DF	Type III SS	Mean Square	F Value	Pr > F
zona	2	648.0000000	324.0000000	32.83	<.0001
tienda	3	123.7291667	41.2430556	4.18	0.0123
zona*tienda	6	45.3333333	7.5555556	0.77	0.6017

Source	Type III Expected Mean Square
zona	Var(Error) + 4 Var(zona*tienda) + 16 Var(zona)
tienda	Var(Error) + 4 Var(zona*tienda) + 12 Var(tienda)
zona*tienda	Var(Error) + 4 Var(zona*tienda)

Type 1 Estimates

Variance Component	Estimate
Var(zona)	19.77778
Var(tienda)	2.80729
Var(zona*tienda)	-0.57813
Var(Error)	9.86806

```
options ls=80 ps=66 nodate;
data ano;
input  materi tempera durac;
title 'ANOVA bifactorial efectos mixtos';
datalines;
1 1  130
1 1  155
1 1   74
1 1  180
1 2   34
1 2   40
1 2   80
1 2   75
1 3   20
1 3   70
1 3   82
1 3   58
2 1  150
2 1  188
2 1  159
2 1  126
2 2  136
2 2  122
2 2  106
2 2  115
2 3   25
2 3   70
2 3   58
2 3   45
3 1  138
3 1  110
3 1  168
3 1  160
3 2  174
3 2  120
3 2  150
3 2  139
3 3   96
3 3  104
3 3   82
3 3   60
;
proc mixed covtest;
class materi tempera;
model durac=materi;
random tempera materi*tempera;
run;
quit;
```

ANOVA bifactorial efectos mixtos

The Mixed Procedure

Model Information

Data Set	WORK.ANO
Dependent Variable	durac
Covariance Structure	Variance Components
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Class Level Information

Class	Levels	Values
materi	3	1 2 3
tempera	3	1 2 3

Dimensions

Covariance Parameters	3
Columns in X	4
Columns in Z	12
Subjects	1
Max Obs Per Subject	36
Observations Used	36
Observations Not Used	0
Total Observations	36

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	352.41258855	
1	1	327.91147422	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error	Z Value	Pr Z
tempera	1429.66	1636.09	0.87	0.1911
materi*tempera	432.06	427.35	1.01	0.1560
Residual	675.21	183.77	3.67	0.0001

```
                  Fit Statistics

     -2 Res Log Likelihood           327.9
     AIC (smaller is better)         333.9
     AICC (smaller is better)        334.7
     BIC (smaller is better)         331.2


         Type 3 Tests of Fixed Effects

                 Num     Den
 Effect           DF      DF    F Value    Pr > F
 materi            2       4       2.22    0.2243
```

OTRA OPCION:

```
options ls=76 nodate nonumber;
title 'ANOVA bifactorial efectos mixtos';
data ano;
input  materi tempera durac;
cards;
1 1  130
1 1  155
. . .
3 3   82
3 3   60
proc anova;
class materi tempera;
model durac=materi tempera materi|tempera ;
test h=tempera e=materi*tempera;
test h=materi e=materi*tempera;
run;
```

```
                                    Sum of
 Source                    DF       Squares   Mean Square  F Value  Pr > F
 Model                      8   59416.22222    7427.02778    11.00  <.0001
 Error                     27   18230.75000     675.21296
 Corrected Total           35   77646.97222

             R-Square     Coeff Var      Root MSE    durac Mean
             0.765210      24.62372      25.98486      105.5278

 Source                    DF      Anova SS   Mean Square  F Value  Pr > F
materi*tempera             4    9613.77778    2403.44444     3.56  0.0186


                  Tests of Hypotheses Using the Anova MS
                    for materi*tempera as an Error Term

 Source                    DF      Anova SS   Mean Square  F Value  Pr > F
 tempera                    2   39118.72222   19559.36111     8.14  0.0389
 materi                     2   10683.72222    5341.86111     2.22  0.2243
```