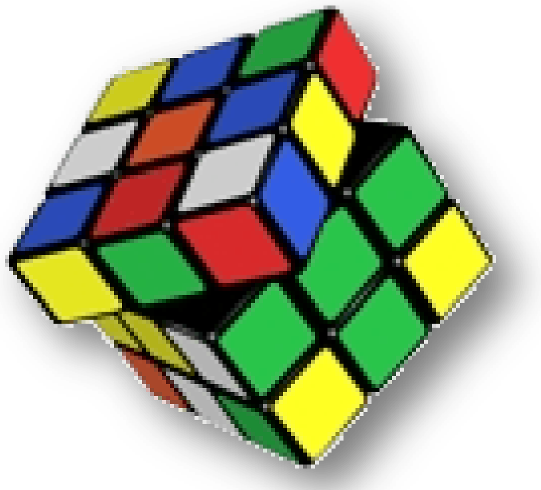


Diseños Factoriales

PROF. ZORITZA BRAVO

Introducción

- En los temas anteriores se analizaron las posibles influencias de un factor sobre la variable respuesta, aleatorizando las observaciones para eliminar el efecto de otros factores.

- En este caso analizaremos modelos en los cuales dos o más factores pueden influir en la variable respuesta.

Diseños factoriales a dos niveles

- Son diseños en los cuales las variables explicativas son factores (es decir, son categóricas o se han categorizado).

- Para cada factor se consideran solo dos niveles, genéricamente alto (+) y bajo (-).

- Todas las variables explicativas involucradas son tratamientos.

- El número de puntos diferentes para un diseño con k factores es $n = 2^k$.

Modelos asociados

- Por ser un conjunto de datos con tratamientos categóricas, el modelo lógico a utilizar es un modelo de análisis de varianza con *k* vías que incluya todas las interacciones entre factores.

- También puede utilizarse un modelo de regresión lineal con variables codificadas, el cual resulta equivalente al modelo *ANOVA*.

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

Definición de efecto

- En el ámbito de los diseños 2^k se denomina *efecto* de una variable (o de una interacción) a la diferencia entre la respuesta esperada que se obtiene en el nivel alto de la variable y la respuesta esperada que se obtiene en el nivel bajo de la misma.

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

Concepto de interacción

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

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

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

Concepto de interacción

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

$$\boxed{E_\alpha|\beta_1 = 10 - 30 = -20}$$

y para el nivel β_2 es:

$$\boxed{E_\alpha|\beta_2 = 20 - 40 = -20}$$

Concepto de interacción

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

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

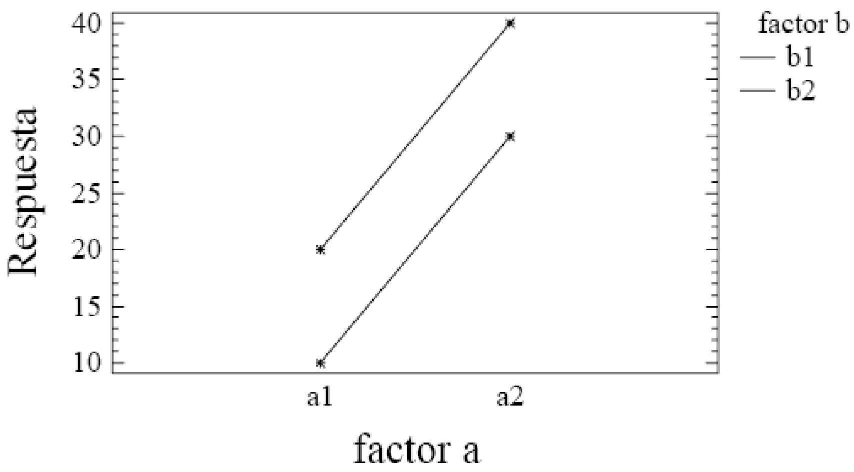
$$E_\beta|\alpha_2 = 30 - 40 = -10$$

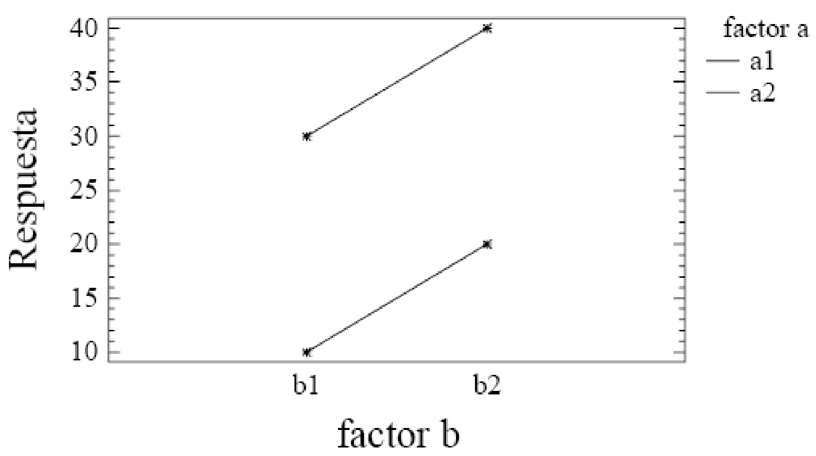
El efecto de uno de los factores no depende de los niveles el otro factor, lo cual indica que no hay interacción entre los factores

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

No hay interacción

Concepto de interacción





Concepto de interacción

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

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

Concepto de interacción

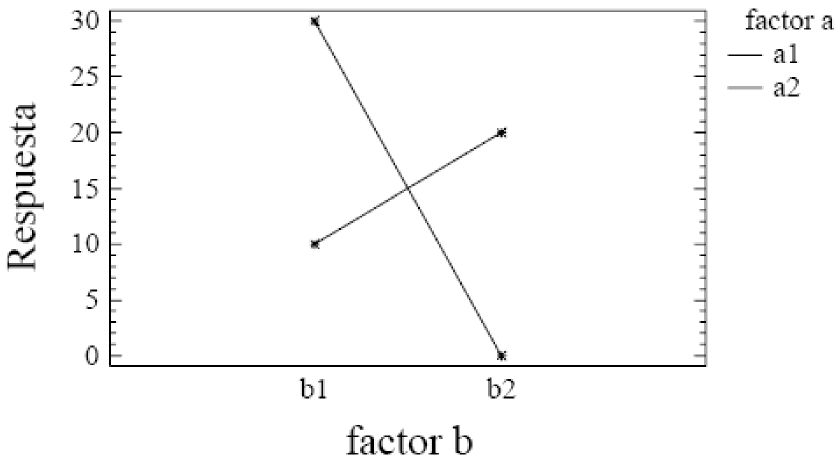
$$E_\alpha|\beta_1 = 10 - 30 = -20$$

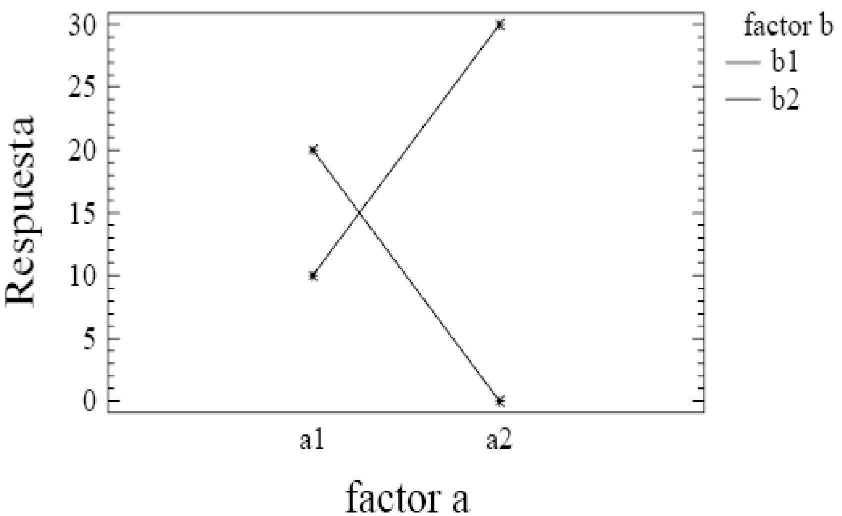
$$E_\alpha|\beta_2 = 20 - 0 = 20$$

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*

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

Concepto de interacción





Diseño factorial de dos factores

Se consideran dos factores *A* y *B* con *a* y *b* niveles respectivamente.

Se tienen *a·b* combinaciones o posibles tratamientos y *n (réplicas)* observaciones para cada tratamiento, esto es, un diseño balanceado.

		Factor *B*			
Factor *A*		*1*	*2*	*...*	*b*
	1	y_{111}, …,y_{112}, y_{11n}	.	.	y_{1b1}, …,y_{1b2}, y_{1bn}
	2	y_{211}, …,y_{212}, y_{21n}	.	.	.
		.	.	.	.
	a	y_{a11}, …,y_{a12}, y_{a1n}	.	.	y_{ab1}, …,y_{ab2}, y_{abn}

Modelo

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

Error aleatorio

Efecto medio global

Efecto sobre la media, causado por el nivel *i* del factor A

Efecto sobre la media, causado por el nivel *j* del factor B

Efecto sobre la media, causado por la interacción

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

Estimación de los parámetros

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

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

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

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

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

$$\left(\widehat{\alpha\beta}\right)_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$$

Descomposición de la suma de cuadrados total

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

$$\boxed{N = abn}$$

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

Descomposición de la suma de cuadrados total

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

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

Descomposición de la suma de cuadrados total

$$SC_{AB} = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots})^2 =$$

$$= \quad n\sum_{i=1}^{a}\sum_{j=1}^{b}\left[(\bar{y}_{ij\cdot} - \bar{y}_{\cdots}) - (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}) - (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots})\right]^2 =$$

$$= \quad n\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{y}_{ij\cdot}^2 - N\cdot\bar{y}_{\cdots}^2 - SC_A - SC_B =$$

$$\frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b}y_{ij\cdot}^2 - \frac{1}{bn}\sum_{i=1}^{a}y_{i\cdot\cdot}^2 - \frac{1}{an}\sum_{j=1}^{b}y_{\cdot j\cdot}^2 + \frac{1}{abn}y_{\cdots}^2$$

$$\equiv \text{“Suma de cuadrados explicada debido a la interacción”}$$

Estimador de la varianza

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

Efecto	Grados de libertad
A	a-1
B	b-1
Interacción AB	(a-1)(b-1)
Error	ab(n-1)
Total	abn-1

Hipótesis

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

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

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

Tabla ANOVA

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	$abn-1$		

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

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

Tabla ANOVA

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

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

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

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

Observaciones

Siempre trataremos de buscar el modelo más sencillo que explique bien la variable respuesta.

Por ejemplo, si aceptamos $H_0 : (\alpha\beta)_{11} = \cdot \cdot \cdot = (\alpha\beta)_{ab} = 0$, concluimos que la interacción no influye de manera apreciable en la respuesta, entonces 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

Un ingeniero está diseñando una batería que se usará en un dispositivo que se someterá a variaciones de temperaturas extremas. El único parámetro del diseño que puede seleccionar en este punto es el material de la placa o ánodo de la batería, y tiene tres elecciones posibles.
Cuando el dispositivo esté fabricado y se envíe al campo, el ingeniero no tendrá control sobre las temperaturas extremas en las que operará el dispositivo, pero sabe por experiencia que la temperatura probablemente afectará la vida efectiva de la batería. Sin embargo, la temperatura puede controlarse en el laboratorio donde se desarrolla el producto para fines de prueba.

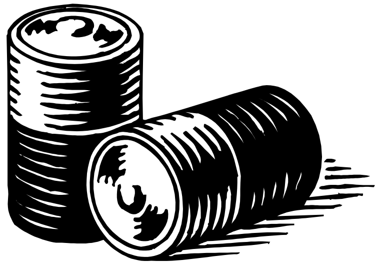
Ejemplo

El ingeniero decide probar los tres materiales de la placa con tres niveles de temperatura, por ser consistentes con el medio ambiente donde se usará finalmente el producto.
Se prueban 4 baterías con cada combinación del material de la placa y la temperatura, y las 36 pruebas se corren de manera aleatoria.

a. ¿Qué efectos tienen el tipo de material y la temperatura sobre la vida de la batería?
b. ¿Existe alguna elección del material que produzca de *manera regular una vida larga de la batería independientemente de la temperatura*?

Ejemplo

	Temperatura		
Tipo de material	**15**	**70**	**125**
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

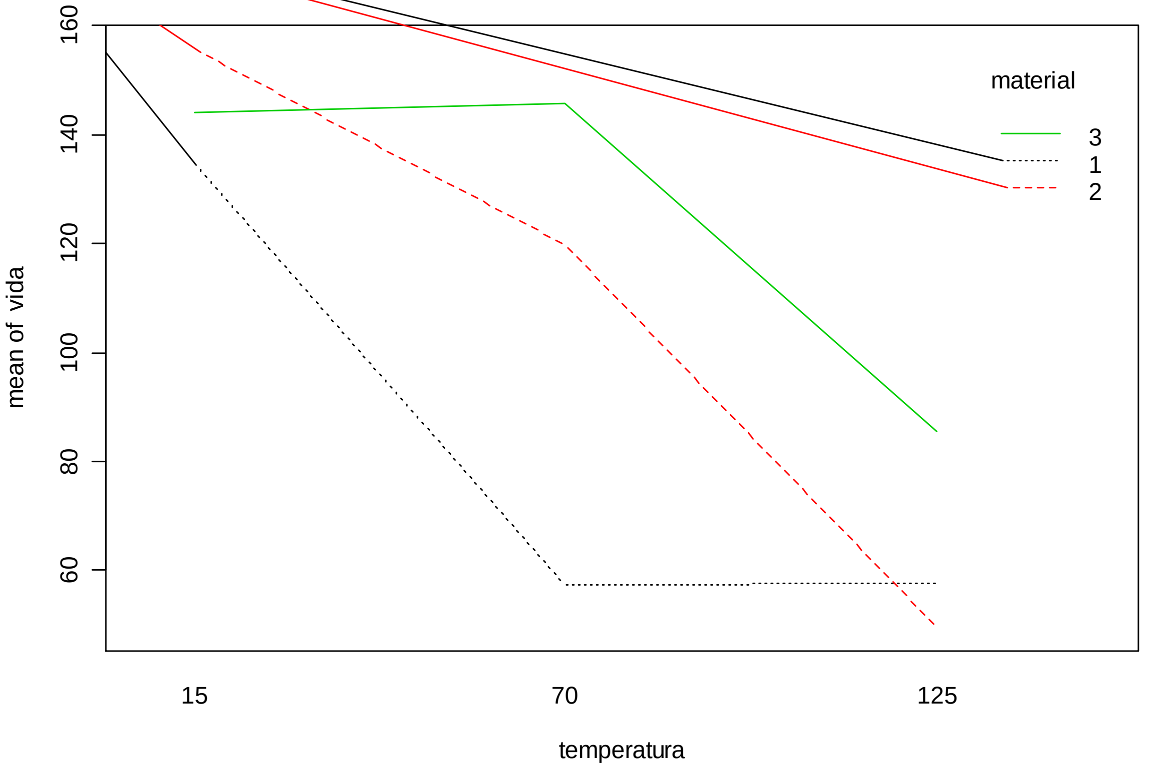


Factores: tipo de material y las temperaturas, con tres niveles cada una. ***Variable de interés:*** vida efectiva de las baterías (en horas). Entonces $a = 3$, $b = 3$ y $n = 4$

Análisis de varianza

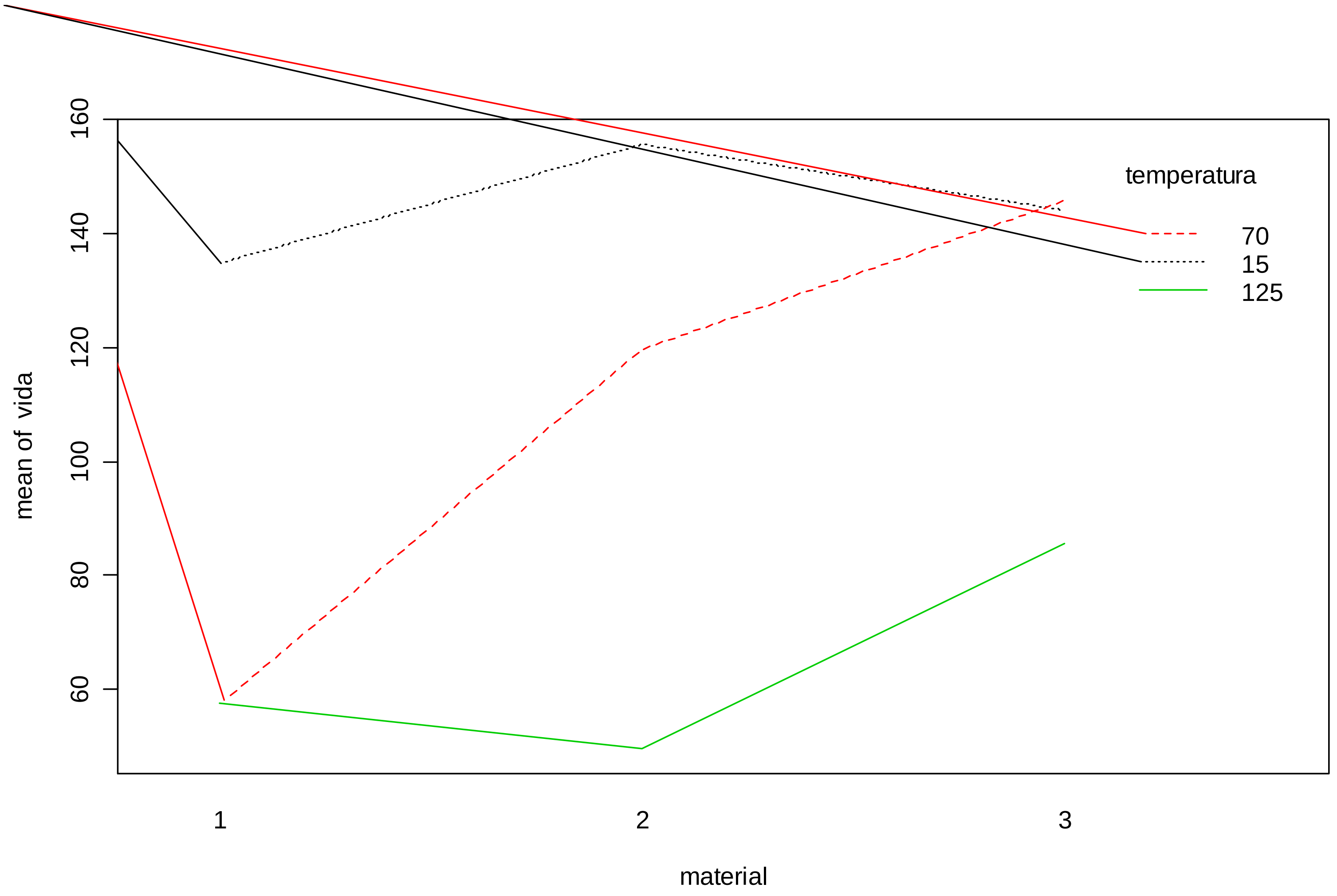
	Temperatura			
Tipo de material	**15**	**70**	**125**	$y_{i..}$
1	130 ,155	34, 40	20, 70	998
	74, 180	80, 75	82, 58	
2	150,188	136, 122	25, 70	1300
	159, 126	106,115	58, 45	
3	138, 110	174, 120	96, 104	1501
	168, 160	150, 139	82, 60	
$y_{.j.}$	1738	1291	770	$y_{...}$ **=3799**

Gráficos de interacción



En general se consigue una vida más larga con una temperatura baja, independientemente del tipo de material.
Al cambiar de una temperatura baja a una intermedia, la vida de la batería con el material tipo 3 tiene un incremento real, mientras que con los materiales 1 y 2 disminuye.
El material tipo 3 parece producir mejores si se quiere una pérdida menor de la vida cuando la temperatura cambia.

Gráficos de interacción



Análisis de varianza

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
material	2	10684	5342	7.9114	0.001976 **
temperatura	2	39119	19559	28.9677	1.909e-07 ***
material:temperatura	4	9614	2403	3.5595	0.018611 *
Residuals	27	18231	675		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

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

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

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

$F_{0.05,4,27} = 2.7277$

Hay una interacción significativa entre los tipos de material y la temperatura.
Los efectos principales del tipo de material y la temperatura también son significativos

Hipótesis del modelo

Normalidad: ε_{ij} sigue una distribución normal

$E(\varepsilon_{ij}) = 0$

Homocedasticidad: $Var(\varepsilon_{ij}) = \sigma^2$

Independencia: ε_{ij} son independientes entre sí

Método de Tuckey

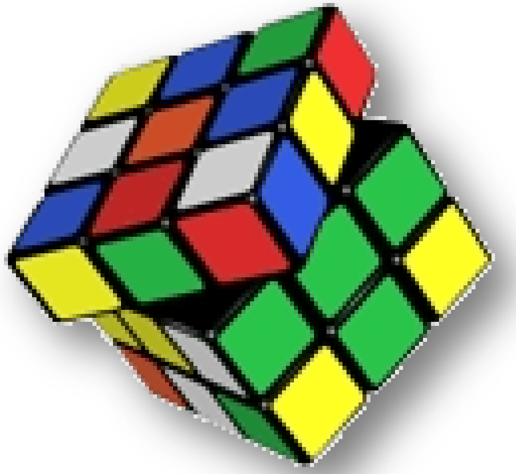
Se requiere que $n_i = n$, $i = 1, \ldots, a$.

Si esto no se cumple, entonces se toma $n = \min_i\{ni\}$

$$\text{Si } |\bar{y}_{i\cdot} - \bar{y}_{j\cdot}| > q_{a,N-a;\alpha}\hat{\sigma}\sqrt{\frac{1}{n}} \Longrightarrow \text{Se rechaza que } \mu_i = \mu_j \text{ a nivel } \alpha.$$

$$\text{Si } |\bar{y}_{i\cdot} - \bar{y}_{j\cdot}| < q_{a,N-a;\alpha}\hat{\sigma}\sqrt{\frac{1}{n}} \Longrightarrow \text{Se acepta que } \mu_i = \mu_j \text{ a nivel } \alpha.$$

```
TukeyHSD(aov(vida~material+temperatura+material*temperatura,
data=datos))
```



Diseños Factoriales (2)

PROF. ZORITZA BRAVO

Diseño sin réplicas

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

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

Error aleatorio

Efecto medio global

Efecto sobre la media, causado por el nivel *i* del factor A

Efecto sobre la media, causado por el nivel *j* del factor B

Efecto sobre la media, causado por la interacción AB

$$i = 1, \ldots, a \quad j = 1, \ldots, b$$

Diseño sin réplicas

En este caso, el número de parámetros a estimar es igual que en el caso previo:

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

y como el número de observaciones es ab, entonces no hay grados de libertad suficientes para estimar

$$\boxed{\hat{\sigma}^2 = \frac{MCE}{ab - ab}.}$$

Diseño sin réplicas

Una posible solución es considerar que la interacción es nula

$$\boxed{(\alpha\beta)_{ij} = 0} \qquad i = 1, \ldots, a \; j = 1, \ldots, b;$$

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

Diseño sin réplicas

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

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

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

Descomposición de la suma de cuadrados total

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

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

Una componente para el error

$$SCE^* = SCE - SC_N$$

con (a − 1)(b − 1) − 1 grados de libertad

Diseño sin réplicas

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