Capítulo 6

Diseños factoriales con dos factores

En primer lugar vamos a estudiar los diseños más simples, es decir aquellos en los que intervienen sólo dos factores. Supongamos que hay a niveles para el factor A y b niveles del factor B, cada réplica del experimento contiene todas las posibles combinaciones de tratamientos, es decir contiene los ab tratamientos posibles.

6.1. El modelo sin replicación

El modelo estadístico para este diseño es:

$$y_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + u_{ij} \qquad i = 1, 2, \cdots, a \quad ; \quad j = 1, 2, \cdots, b \quad \text{, donde}$$

- y_{ij} : Representa la observación correspondiente al nivel (i) del factor A y al nivel (j) del factor B.
- μ: Efecto constante denominado media global.
- τ_i: Efecto producido por el nivel i-ésimo del factor A, $(\sum_i \tau_i = 0)$.
- β_j: Efecto producido por el nivel j-ésimo del factor B, $\left(\sum_j \beta_j = 0\right)$.
- $(\tau\beta)_{ij}$: Efecto producido por la interacción entre $A{\times}B$, $\left(\sum_i (\tau\beta)_{ij} = \sum_j (\tau\beta)_{ij} = 0\right)$.
- u_{ij} son vv aa. independientes con distribución$N(0, \sigma)$.

Supondremos que se toma una observación por cada combinación de factores, por tanto, hay un total de $N = ab$ observaciones.

Parámetros a estimar:

Parámetros	Número
μ	1
τ_i	$a-1$
β_j	$b-1$
$(\tau\beta)_{ij}$	$(a-1)(b-1)$
σ^2	1
Total	$ab+1$

A pesar de las restricciones impuestas al modelo,

$\sum_i \tau_i = \sum_j \beta_j = \sum_i (\tau\beta)_{ij} = \sum_j (\tau\beta)_{ij} = 0,$

el número de parámetros $(ab+1)$ supera al número de obsevaciones (ab). Por lo tanto, algún parámetro no será estimable.

6.1.1. Estimación de los parámetros del modelo

Los estimadores máximo verosímiles de los parámetros del modelo son

$$\widehat{\mu} = \bar{y}_{..} \quad , \quad \widehat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..} \quad , \quad \widehat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..} \quad \text{y} \quad \left(\widehat{\tau\beta}\right)_{ij} = y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}$$

Los residuos de este modelo son: $e_{ij} = y_{ij} - \widehat{y}_{ij} = y_{ij} - \widehat{\mu} - \widehat{\tau}_i - \widehat{\beta}_j - \left(\widehat{\tau\beta}\right)_{ij} = 0$.

Por lo tanto, al ser los residuos nulos no es posible estimar la varianza del modelo y no se pueden contrastar la significatividad de los efectos de los factores. Dichos contrates sólo pueden realizarse si:

a) Suponemos que la interacción entre $A \times B$ es cero. Entonces $e_{ij} = y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}$.

b) Replicamos el experimento (Tomamos varias observaciones por cada combinación de factores).

6.2. El modelo con replicación

El modelo estadístico para este diseño es:

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + u_{ijk} \;\; ; \;\; i = 1, 2, \cdots, a \;\; ; \;\; j = 1, 2, \cdots, b \;\; ; \;\; k = 1, 2, \cdots, r$$

donde r es el número de replicaciones y $n = abr$ es el número de observaciones.

El número de parámetros de este modelo es, como en el modelo de dos factores sin replicación, $ab+1$ pero en este caso el número de observaciones es abr.

6.2.1. Estimación de los parámetros del modelo

Los estimadores máximo verosímiles de los parámetros del modelo son
$\widehat{\mu} = \bar{y}_{...}$, $\widehat{\tau}_i = \bar{y}_{i..} - \bar{y}_{...}$, $\widehat{\beta}_j = \bar{y}_{.j.} - \bar{y}_{...}$ y $\left(\widehat{\tau\beta}\right)_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$
donde

*) $\bar{y}_{ij.}$ es la media de las r observaciones en la celdilla ij: $\bar{y}_{ij.} = \left(\sum_k y_{ijk}\right)/r$

*) $\bar{y}_{i..}$ es la media de las observaciones del nivel i del factor A :
$\bar{y}_{i..} = \left(\sum_{j,k} y_{ijk}\right)/(br) \quad ; \quad i = 1, \cdots, a$

*) $\bar{y}_{.j.}$ es la media de las observaciones del nivel j del factor B :
$\bar{y}_{.j.} = \left(\sum_{i,k} y_{ijk}\right)/(ar) \quad ; \quad j = 1, \cdots, b$

*) $\bar{y}_{...}$ es la media total de las observaciones : $\bar{y}_{...} = \left(\sum_{i,j,k} y_{ijk}\right)/r$.

Los residuos de este modelo son:

$$e_{ijk} = y_{ijk} - \widehat{y}_{ijk} = y_{ij} - \widehat{\mu} - \widehat{\tau}_i - \widehat{\beta}_j - \left(\widehat{\tau\beta}\right)_{ij} = y_{ijk} - \bar{y}_{ij.} \quad .$$

Se verifica que todos los residuos de una celdilla deben sumar cero es decir, en cada celdilla hay $r-1$ residuos independientes. Por lo tanto, en total habrá $ab(r-1)$ residuos independientes.

Se verifican las mismas propiedades para los estimadores máximo-verosímiles que en los modelos anteriores. La varianza residual tiene la siguiente expresión

$$\widehat{S}_R^2 = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r} e_{ijk}^2}{ab(r-1)}.$$

6.2.2. Descomposición de la variabilidad

La ecuación básica del análisis de la varianza es:

$$\begin{aligned}\sum_{i,j,k}(y_{ijk} - \bar{y}_{...})^2 \;=\;& br\sum_{i=1}^{a}(\bar{y}_{i..} - \bar{y}_{...})^2 + ar\sum_{j=1}^{b}(\bar{y}_{.j.} - \bar{y}_{...})^2 + \\ & r\sum_{i,j}(y_{ijs} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 + \sum_{i,j,k}(y_{ijk} - \bar{y}_{ij.})^2\end{aligned}$$

que simbólicamente podemos escribir: $SCT = SCA + SCB + SC(AB) + SCR$. Estas sumas de cuadrados también se pueden expresar como:

- $SCT = \sum_{i,j,k} y_{ijk}^2 - \left(y_{\cdots}^2\right)/r$: *Suma total de cuadrados.*
- $SCA = \left(\sum_i y_{i\cdot\cdot}^2\right)/(br) - \left(y_{\cdots}^2\right)/(abr)$: *S. C. entre niveles de A.*
- $SCB = \left(\sum_j y_{\cdot j\cdot}^2\right)/(ar) - \left(y_{\cdots}^2\right)/(abr)$: *S. C. entre niveles de B.*
- $SC(AB) = \left(\sum_{i,j} y_{ij\cdot}^2\right)/r - \left(y_{\cdots}^2\right)/(abr) - SCA - SCB$: *S. C. de la interacción $A \times B$*
- $SCR = SCT - SCA - SCB - SC(AB)$: *S. C. del error.*

A partir de la ecuación básica del ANOVA se pueden construir los cuadrados medios definidos como:

* Cuadrado medio total: $CMT = (SCT)/(n-1)$
* Cuadrado medio de A: $CMA = (SCA)/(a-1)$
* Cuadrado medio de B: $CMB = (SCB)/(b-1)$
* Cuadrado medio de la interacción $A \times B$: $CM(AB) = (SC(AB))/((a-1)(b-1))$
* Cuadrado medio residual: $CMR = (SCR)/(ab(r-1))$

6.2.3. Análisis estadístico

El objetivo del análisis es realizar los contrastes de hipótesis nula:

i) $H_{0A} \equiv \tau_1 = \cdots = \tau_a = 0$. Es decir, considerando la presencia de las interacciones con el factor B, contrastar si los efectos de los niveles del factor A son nulos. El estadístico de contraste es

$$F_A = \frac{(SCA)/(a-1)}{(SCR)/(ab(r-1))} = \frac{CMA}{CMR} \rightsquigarrow^{H_{0A}} F_{(a-1),ab(r-1)}$$

Se rechaza H_{0A} al nivel α si $F_{\alpha(\exp)} > F_{(a-1),ab(r-1)}$

ii) $H_{0B} \equiv \beta_1 = \cdots = \beta_b = 0$. Es decir, considerando la presencia de las interacciones con el factor A, contrastar si los efectos de los niveles del factor B son nulos. El estadístico de contraste es

$$F_B = \frac{(SCB)/(b-1)}{(SCR)/(ab(r-1))} = \frac{CMB}{CMR} \rightsquigarrow^{H_{0B}} F_{(b-1),ab(r-1)}$$

Se rechaza H_{0B} al nivel α si $F_{\alpha(\exp)} > F_{(b-1),ab(r-1)}$.

iii) $H_{0(AB)} \equiv (\tau\beta)_{ij} = 0 \quad$ para todo i, j. Es decir, contrastar si los efectos de las interacciones entre los factores A y B son nulos. El estadístico de contraste es

$$F_{(AB)} = \frac{(SC(AB)) \,/\, ((a-1)(b-1))}{(SCR) \,/\, (ab(r-1))} = \frac{CM(AB)}{CMR} \quad \rightsquigarrow^{H_{0(AB)}} F_{(a-1)(b-1),ab(r-1)}$$

Se rechaza $H_{0(AB)}$ al nivel α si $F_{\alpha(\exp)} > F_{(a-1)(b-1),ab(r-1)}$.

Tabla ANOVA para el modelo bifactorial con replicación

F. V.	S. C.	G. L.	C. M.	F_{exp}
Factor A	SCA	$a-1$	CMA	CMA/CMR
Factor B	SCB	$b-1$	CMB	CMB/CMR
Interacción	$SC(AB)$	$(a-1)(b-1)$	$CM(AB)$	$CM(AB)/CMR$
Residual	SCR	$ab(r-1)$	CMR	
TOTAL	SCT	$abr-1$	CMT	

La diagnosis y validación del modelo se realiza igual que en los modelos anteriores.

Ejemplo 6.1

En unos laboratorios se está estudiando los factores que influyen en la resistencia de un tipo particular de fibra. Se eligen al azar cuatro máquinas y tres operarios y se realiza un experimento factorial usando fibras de un mismo lote de producción. Los resultados obtenidos se muestran en la tabla adjunta. Analizar los resultados y obtener las conclusiones apropiadas.

	Tipos de máquinas			
Operario	*A*	*B*	*C*	*D*
1	*109* *110*	*110* *115*	*108* *109*	*110* *108*
2	*110* *112*	*110* *111*	*111* *109*	*114* *112*
3	*116* *114*	*112* *115*	*114* *119*	*120* *117*

Para realizar el análisis organizamos los datos en forma tabular de la manera siguiente:

	Tipos de máquinas				
Operario	A	B	C	D	$y_{i..}$
1	109 110 219	110 115 225	108 109 217	110 108 218	879
2	110 112 222	110 111 221	111 109 220	114 112 226	889
3	116 114 230	112 115 227	114 119 233	120 117 237	927
$y_{.j.}$	671	673	670	681	2695

Las *Sumas de Cuadrados* y la Tabla ANOVA se muestran a continuación

$$SCT = \left(109^2 + \cdots 117^2\right) - \frac{(2695)^2}{24} = 262{,}97$$

$$SCA = \frac{879^2 + 889^2 + 927^2}{3 \times 2} - \frac{(2695)^2}{24} = 12{,}46$$

$$SCB = \frac{671^2 + 673^2 + 670^2 + 681^2}{4 \times 2} - \frac{(2695)^2}{24} = 160{,}33$$

$$SC(AB) = \frac{219^2 + \cdots + 237^2}{2} - \frac{(2695)^2}{24} - 12{,}46 - 160{,}33 = 44{,}67 \;\text{ y }\; SCR = 45{,}5.$$

F. V.	S. C.	G. L.	C. M.	F_{exp}
Factor A	12,46	3	4,15	1,10
Factor B	160,34	2	80,17	21,14
Interacción	44,67	6	7,45	1,96
Residual	45,50	12	3,79	
TOTAL	262,97	23		

Realizando los contrastes al nivel de significación del 5 %, se concluye que es significativo el efecto principal del "operario" (factor B) ($F_{0,05,2,12} = 3{,}49$), pero no son significativos el efecto principal del tipo de máquina (factor A) ($F_{0,05,3,12} = 3{,}89$) y la interacción entre el tipo de máquina y operario (factor $A \times B$) ($F_{0,05,6,12} = 3{,}00$).

Bibliografía utilizada

∗ **Lara Porras, A.M.** (2000). *"Diseño Estadístico de Experimentos, Análisis de la Varianza y Temas Relacionados: Tratamiento Informático mediante SPSS."* Proyecto Sur de Ediciones.