Capítulo 1

Diseños en bloques aleatorizados: Comprobación de la idoneidad del modelo

La comprobación de la idoneidad del modelo requiere contrastar las hipótesis de normalidad, homocedasticidad e independencia: Los contrastes son análogos a los realizados en el modelo completamente aleatorizado, con la diferencia de que en el modelo en bloques aleatorizados también hay que estudiar la variabilidad del error por bloques.

En este modelo hemos supuesto otra hipótesis adicional, la aditividad de los efectos de tratamiento y bloque, es decir, hemos supuesto un modelo en el que el efecto de un factor no depende del nivel del otro factor. A pesar de que este modelo aditivo es a menudo útil, existen situaciones en las que resulta inadecuado, por ejemplo, las diferencias entre dos fertilizantes pueden ser mayores cuando se aplican en el bloque 1 que cuando se aplican en el bloque 3. Cuando esto ocurre se dice que existe interacción entre los tratamientos y los bloques, y entonces el modelo en bloques propuesto no es adecuado. Por lo tanto, la hipótesis de aditividad también debe ser contrastada.

1.0.1. Test de Interacción de Tukey

Cuando existe interacción entre tratamiento y bloque, el modelo tendrá en general la siguiente ecuación

$$y_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + u_{ij} \qquad i = 1, 2, \cdots, I \quad j = 1, 2, \cdots, J \quad , \tag{1.1}$$

donde el término $(\tau\beta)_{ij}$ representa la interacción. Se suponen las siguientes restricciones para los parámetros

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

por lo tanto el número de parámetros independientes de este modelo sería

$$1 + (I-1) + (J-1) + (I-1)(J-1) = IJ \quad ,$$

que, puesto que hay una sola observación por celdilla, coincide con el número de observaciones, no habiendo suficientes grados de libertad para estimar la varianza residual. Un procedimiento para solventar este problema consiste en tomar más de una observación por celdilla.

Tukey (1949) desarrolló un método para determinar si existe interacción entre tratamientos y bloques, cuando sólo hay una observación por celdilla, conocido como el *test de interacción de un grado de libertad de Tukey.* Este procedimiento supone que la forma de la interacción es particularmente simple, es decir

$$(\tau\beta)_{ij} = \gamma\tau_i\beta_j \quad ,$$

en donde γ es una constante desconocida. De esta forma las interacciones añaden únicamente un parámetro, y el modelo resultante es:

$$y_{ij} = \mu + \tau_i + \beta_j + \gamma\tau_i\beta_j + u_{ij} \qquad i = 1, 2, \cdots, I \quad ; \quad j = 1, 2, \cdots, J \quad , \tag{1.2}$$

siendo las restricciones para este modelo

$$\sum_i \tau_i = \sum_j \beta_j = \sum_i \gamma\tau_i\beta_j = \sum_j \gamma\tau_i\beta_j = 0 \quad .$$

En primer lugar abordaremos la estimación de los parámetros del modelo. Para ello, se construye la función de verosimilitud asociada a la muestra $\mathbf{y} = (y_{11}, \cdots, y_{1J}, \cdots, y_{I1}, \cdots, y_{IJ})$:

$$\mathbb{L}(\mu, \tau_i, \beta_j, \gamma, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{I} \sum_{j=1}^{J} \left[y_{ij} - \mu - \tau_i - \beta_j - \gamma\tau_i\beta_j \right]^2 \right) \quad , \tag{1.3}$$

se determina el logaritmo de dicha función

$$\ln(\mathbb{L}(\mu,\tau_i,\sigma^2)) = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{I}\sum_{j=1}^{J}\left[y_{ij}-\mu-\tau_i-\beta_j-\gamma\tau_i\beta_j\right]^2 \quad , \tag{1.4}$$

y se hallan las primeras derivadas parciales respecto de los parámetros del modelo, obteniéndose las mismas expresiones que en (**??**)-(**??**) para los estimadores de los parámetros μ, τ_i y β_j.

El estimador máximo verosimil de γ se obtiene realizando la correspondiente derivada parcial, es decir

$$\frac{\partial \ln \mathbb{L}}{\partial \gamma} = \frac{1}{\sigma^2}\sum_{i=1}^{I}\sum_{j=1}^{J}\left[y_{ij}-\mu-\tau_i-\beta_j-\gamma\tau_i\beta_j\right]\tau_i\beta_j = 0 \tag{1.5}$$

de donde

$$\sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij}\tau_i\beta_j - \mu\sum_{i=1}^{I}\tau_i\sum_{j=1}^{J}\beta_j - \sum_{i=1}^{I}\tau_i^2\sum_{j=1}^{J}\beta_j - \sum_{j=1}^{J}\beta_j^2\sum_{i=1}^{I}\tau_i - \gamma\sum_{i=1}^{I}\tau_i^2\sum_{j=1}^{J}\beta_j^2 = 0 \quad .$$

Por lo tanto

$$\widehat{\gamma} = \frac{\displaystyle\sum_{i=1}^{I}\sum_{j=1}^{J}\widehat{\tau}_i\widehat{\beta}_j y_{ij}}{\displaystyle\sum_{i=1}^{I}\widehat{\tau}_i^2\sum_{j=1}^{J}\widehat{\beta}_j^2} \quad , \tag{1.6}$$

y reemplazando $\widehat{\tau}_i$ y $\widehat{\beta}_j$ por sus respectivos estimadores muestrales, tenemos

$$\widehat{\gamma} = \frac{\displaystyle\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{y}_{i.}-\bar{y}_{..})(\bar{y}_{.j}-\bar{y}_{..})y_{ij}}{\displaystyle\sum_{i=1}^{I}(\bar{y}_{i.}-\bar{y}_{..})^2\sum_{j=1}^{J}(\bar{y}_{.j}-\bar{y}_{..})^2} \quad . \tag{1.7}$$

Como hemos supuesto la existencia de interacción entre los factores, hay que introducir en este modelo una suma de cuadrados que represente dicha interacción, esta suma de cuadrados se denota por $SCIT$ y viene dada por la expresión

$$SCIT = \sum_i\sum_j \widehat{\gamma}^2\widehat{\tau}_i^2\widehat{\beta}_j^2 \quad , \tag{1.8}$$

que recibe el nombre de suma de cuadrados debida a la interacción o no-aditividad del modelo, que también se puede expresar, sustituyendo los correspondientes estimadores muestrales, de la siguiente forma

$$SCIT = \sum_{i=1}^{I}\sum_{j=1}^{J}\widehat{\gamma}^2(\bar{y}_{i.}-\bar{y}_{..})^2(\bar{y}_{.j}-\bar{y}_{..})^2 = \frac{\left[\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{y}_{i.}-\bar{y}_{..})(\bar{y}_{.j}-\bar{y}_{..})y_{ij}\right]^2}{\sum_{i=1}^{I}(\bar{y}_{i.}-\bar{y}_{..})^2\sum_{j=1}^{J}(\bar{y}_{.j}-\bar{y}_{..})^2} \quad . \tag{1.9}$$

Por lo tanto, la ecuación básica del análisis de la varianza para el modelo con interacción (1.2) debe tener un término más que la citada ecuación del modelo en bloques aleatorizados. Dicha ecuación se expresa simbólicamente en la forma

$$SCT = SCTr + SCBl + SCIT + SCR \quad , \tag{1.10}$$

donde

- $SCTr$ es la suma de cuadrados entre tratamientos
- $SCBl$ es la suma de cuadrados entre bloques
- $SCIT$ es la suma de cuadrados debida a la interacción por Tukey
- SCR es la suma de cuadrados residual que se obtiene como

$$SCR = SCT - SCTr - SCBl - SCIT \quad .$$

Se demuestra que si $\gamma = 0$, esto es, si no hay interacción del tipo $\gamma\tau_i\beta_j$, entonces $SCIT/\sigma^2$ y SCR/σ^2 son variables aleatorias independientes distribuidas según una χ^2 con 1 y $IJ - I - J$ grados de libertad, respectivamente. Por tanto, si $\gamma = 0$, el estadístico de contraste

$$F = \frac{SCIT/1}{SCR/(IJ - I - J)} \quad , \tag{1.11}$$

se distribuye según una F de Snedecor con $1, IJ - I - J$ grados de libertad. Así, para contrastar

$$\begin{array}{ll} H_0: & \gamma = 0 \ \text{(no hay interacción)} \\ H_1: & \gamma \neq 0 \ \text{(hay interacción del tipo } \gamma\tau_i\beta_j) \quad , \end{array} \tag{1.12}$$

puesto que, F se distribuye como una $F_{1,IJ-I-J}$ cuando H_0 es cierta y puesto que los valores grandes de F_{exp} conducen a la conclusión H_1, la decisión apropiada para controlar un riesgo de error de Tipo I igual a α, es:

$$\begin{aligned} &Si\ \ F_{exp} \leq F_{\alpha,1,IJ-I-J}\ , \quad \text{se acepta } H_0 \\ &Si\ \ F_{exp} > F_{\alpha,1,IJ-I-J}\ , \quad \text{se rechaza } H_0 \quad , \end{aligned} \tag{1.13}$$

donde $F_{\alpha,1,IJ-I-J}$ es el punto crítico superior de la distribución F con 1 y $IJ - I - J$ grados de libertad.