Capítulo 4

Modelos unifactoriales de efectos aleatorizados

En el modelo de efectos aleatorios, los niveles del factor son una muestra aleatoria de una población de niveles. Este modelo surge ante la necesidad de estudiar un factor que presenta un número elevado de posibles niveles, que en algunas ocasiones puede ser infinito. En este modelo las conclusiones obtenidas se generalizan a toda la población de niveles del factor, ya que los niveles empleados en el experimento fueron seleccionados al azar.

Formalmente la expresión del modelo es la misma que en el modelo de efectos fijos dado por la ecuación (**??**), es decir:

$$y_{ij} = \mu + \tau_i + u_{ij} \quad , \tag{4.1}$$

con la diferencia de que ahora τ_i son variables aleatorias. Además este modelo requiere que las variables τ_i y u_{ij} sean independientes y sigan distribuciones normales con media 0 y varianza constante σ_τ^2 y σ^2, respectivamente. Así, por la independencia entre las variables τ_i y u_{ij}, la varianza de cualquier observación de la muestra, es decir, la varianza total, denotada por σ_T^2, vale

$$\sigma_T^2 = \sigma_\tau^2 + \sigma^2 \quad .$$

La mecánica del Análisis de la Varianza es la misma que en el modelo de efectos fijos, excepto en el cálculo de las esperanzas de los cuadrados medios, ya que mientras en el modelo de efectos fijos, los efectos τ_i eran constantes que cumplían la condición $\sum_{i=1}^{I} n_i\tau_i = 0$, ahora son variables aleatorias $N(0, \sigma_\tau)$, no sometidas a ninguna restricción[1]. En este modelo, carece de sentido probar la hipótesis que se refiere a los efectos de tratamiento individuales y, en su lugar, debe realizarse el contraste:

$$H_0: \quad \sigma_\tau^2 = 0$$

$$H_1: \quad \sigma_\tau^2 > 0$$

Si se acepta H_0, $\sigma_\tau^2 = 0$, significa que todos los tratamientos son idénticos; en cambio, si se acepta H_1, $\sigma_\tau^2 > 0$, significa que existe variabilidad entre los tratamientos.

Al igual que en el modelo de efectos fijos, se distingue entre el caso equilibrado y el caso no-equilibrado.

4.1. Modelo de efectos aleatorios no-equilibrado

Como en el modelo de efectos fijos, la variabilidad total de los datos se puede expresar como la suma de la variabilidad entre los tratamientos y la variabilidad dentro de los mismos. A partir de estas variabilidades se definen las correspondientes varianzas muestrales.

Para establecer el procedimiento del contraste de hipótesis, para el modelo de efectos aleatorios, calculemos en primer lugar los valores esperados de la varianza entre tratamientos y de la varianza residual.

Como en el modelo de efectos fijos, consideremos las expresiones de $y_{i.}$, $\bar{y}_{i.}$, $y_{..}$e $\bar{y}_{..}$, en función de los parámetros del modelo, con objeto de poder hallar las esperanzas de las varianzas muestrales.

$$y_{i.} = n_i\mu + n_i\tau_i + u_{i.} \quad ; \qquad \bar{y}_{i.} = \mu + \tau_i + \bar{u}_{i.}$$

$$y_{..} = N\mu + \sum_{i=1}^{I} n_i\tau_i + u_{..} \quad ; \quad \bar{y}_{..} = \mu + \frac{1}{N}\sum_{i=1}^{I} n_i\tau_i + \bar{u}_{..}$$

[1] Al ser los efectos τ_i variables aleatorias independientes entonces la condición $\sum_{i=1}^{I} n_i\tau_i = 0$, del modelo de efectos fijos, no puede imponerse ya que conduciría a dependencia entre los τ_i.

1º) La varianza entre tratamientos se puede expresar como

$$\begin{aligned}
\widehat{S}^2_{Tr} = {} & \frac{1}{I-1}\left[\sum_{i=1}^{I} n_i\left(\bar{y}_{i.}-\bar{y}_{..}\right)^2\right] = \\
& \frac{1}{I-1}\sum_{i=1}^{I} n_i\left[\mu+\tau_i+\bar{u}_{i.}-\mu-\frac{1}{N}\sum_{k=1}^{I} n_k\tau_k-\bar{u}_{..}\right]^2 = \\
& \frac{1}{I-1}\sum_{i=1}^{I} n_i\left[\left(\tau_i-\frac{1}{N}\sum_{k=1}^{I} n_k\tau_k\right)+\left(\bar{u}_{i.}-\bar{u}_{..}\right)\right]^2 = \\
& \frac{1}{I-1}\left[\sum_{i=1}^{I} n_i\left(\tau_i-\frac{1}{N}\sum_{k=1}^{I} n_k\tau_k\right)^2+\sum_{i=1}^{I} n_i\left(\bar{u}_{i.}-\bar{u}_{..}\right)^2+\right. \\
& \left. 2\sum_{i=1}^{I} n_i\left(\tau_i-\frac{1}{N}\sum_{k=1}^{I} n_k\tau_k\right)\left(\bar{u}_{i.}-\bar{u}_{..}\right)\right] \quad ,
\end{aligned}$$

y su esperanza matemática será la suma de las esperanzas matemáticas de cada sumando, es decir

$$\begin{aligned}
\mathrm{E}[\widehat{S}^2_{Tr}] = {} & \frac{1}{I-1}\left\{\mathrm{E}\left[\sum_{i=1}^{I} n_i\left(\tau_i-\frac{1}{N}\sum_{k=1}^{I} n_k\tau_k\right)^2\right]+\mathrm{E}\left[\sum_{i=1}^{I} n_i\left(\bar{u}_{i.}-\bar{u}_{..}\right)^2\right]+\right. \\
& \left. 2\,\mathrm{E}\left[\sum_{i=1}^{I} n_i\left(\tau_i-\frac{1}{N}\sum_{k=1}^{I} n_k\tau_k\right)\left(\bar{u}_{i.}-\bar{u}_{..}\right)\right]\right\} \quad .
\end{aligned} \tag{4.2}$$

Ahora bien, puesto que:

a) Las variables aleatorias τ_i son independientes con media 0, entonces

$$\mathrm{E}\left(\sum_{i\neq k} n_i n_k \tau_i \tau_k\right) = 0$$

y además, como $\mathrm{E}(\tau_i^2) = \sigma_\tau^2$, también se verifica que

$$\mathrm{E}\left[\tau_i \sum_{k=1}^{I} n_k \tau_k\right] = n_i \sigma_\tau^2 \quad .$$

Por lo tanto,

$$\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^{I} n_i \left(\tau_i - \sum_{k=1}^{I} \frac{n_k \tau_k}{N}\right)^2\right] = & \sum_{i=1}^{I} n_i \,\mathrm{E}\left(\tau_i^2 + \left(\sum_{k=1}^{I} \frac{n_k \tau_k}{N}\right)^2 - \right. \\
& \left. 2\tau_i \sum_{k=1}^{I} \frac{n_k \tau_k}{N}\right) = \\
& \sum_{i=1}^{I} n_i \,\mathrm{E}\left(\tau_i^2 + \frac{1}{N^2} \sum_{k=1}^{I} n_k^2 \tau_k^2 + \right. \\
& \left. \frac{1}{N^2} \sum_{k \neq r}^{I} n_k n_r \tau_k \tau_r - \frac{2}{N} \tau_i \sum_{k=1}^{I} n_k \tau_k\right) = \\
& \sum_{i=1}^{I} n_i \left(\sigma_\tau^2 + \frac{1}{N^2} \sum_{k=1}^{I} n_k^2 \sigma_\tau^2 - \frac{2}{N} n_i \sigma_\tau^2\right) = \\
& N\sigma_\tau^2 - \frac{1}{N} \sum_{i=1}^{I} n_i^2 \sigma_\tau^2 \quad .
\end{aligned} \tag{4.3}$$

b) $\mathrm{E}\left[\sum_{i,j}(\bar{u}_{i.}-\bar{u}_{..})^2\right]=I\sigma^2-\sigma^2$, ya que

$$\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^{I} n_i\,(\bar{u}_{i.}-\bar{u}_{..})^2\right] &= \sum_{i=1}^{I} n_i\,\mathrm{E}\left(\frac{1}{n_i}\sum_{j=1}^{n_i}u_{ij}-\frac{1}{N}\sum_{k=1}^{I}\sum_{h=1}^{n_k}u_{kh}\right)^2 = \\
&\sum_i n_i\,\mathrm{E}\left[\frac{1}{n_i^2}\left(\sum_j u_{ij}\right)^2+\frac{1}{N^2}\left(\sum_{k,h}u_{kh}\right)^2-\right. \\
&\left. 2\frac{1}{Nn_i}\sum_j u_{ij}\sum_{k,h}u_{kh}\right] = \\
&\sum_i n_i\left[\frac{1}{n_i^2}\sum_j\sigma^2+\frac{1}{N^2}\sum_{k,h}\sigma^2-2\frac{1}{Nn_i}\sum_j\sigma^2\right] = \\
&\sum_i n_i\left[\frac{\sigma^2}{n_i}-\frac{\sigma^2}{N}\right]=I\sigma^2-\sigma^2=\sigma^2(I-1)
\end{aligned} \tag{4.4}$$

c) Puesto que las variables aleatorias τ_i y u_{ij} son independientes entre sí y

$$\mathrm{E}(\tau_i)=\mathrm{E}(u_{ij})=0 \quad ,$$

entonces

$$\mathrm{E}\left[\sum_i n_i\left(\tau_i-\frac{1}{N}\sum_k n_k\tau_k\right)(\bar{u}_{i.}-\bar{u}_{..})\right]=0 \tag{4.5}$$

Por lo tanto, sustituyendo las expresiones (4.3), (4.4) y (4.5) en (4.2), tenemos

$$\mathrm{E}(\widehat{S}^2_{T_r})=\frac{1}{I-1}\left(N\sigma_\tau^2-\frac{1}{N}\sum_i n_i^2\sigma_\tau^2+I\sigma^2-\sigma^2\right)=\frac{N^2-\sum_i n_i^2}{N(I-1)}\sigma_\tau^2+\sigma^2 \quad . \tag{4.6}$$

2º) La esperanza matemática de la varianza residual es la varianza poblacional, en efecto

$$\mathrm{E}(\widehat{S}_R^2) = \mathrm{E}\left[\frac{1}{N-I}\left(\sum_{i,j}(y_{ij}-\bar{y}_{i.})^2\right)\right] = \frac{1}{N-I}\sum_{i,j}\mathrm{E}\left[u_{ij}-\frac{1}{n_i}\sum_j u_{ij}\right]^2 =$$

$$\frac{1}{N-I}\sum_{i,j}\mathrm{E}\left[u_{ij}^2+\frac{1}{n_i^2}\left(\sum_j u_{ij}\right)^2-\frac{2}{n_i}u_{ij}\sum_j u_{ij}\right] =$$

$$\frac{1}{N-I}\sum_{i,j}\left(\sigma^2+\frac{\sigma^2}{n_i}-\frac{2}{n_i}\sigma^2\right) = \frac{1}{N-I}\left(N\sigma^2-I\sigma^2\right) = \sigma^2 \quad ,$$

por tanto, $\widehat{S}_R^2$ es un estimador insesgado del parámetro σ^2.

De la ecuación (4.6) se deduce que si no existe variabilidad entre los tratamientos, es decir, $\sigma_\tau^2 = 0$, entonces $\widehat{S}_{Tr}^2$ es también un estimador insesgado de σ^2.

De modo similar que en el modelo de efectos fijos se puede demostrar, bajo la hipótesis nula correspondiente a este modelo, que $SCTr/\sigma^2$ y SCR/σ^2 siguen distribuciones χ^2 independientes con $I-1$ y $N-I$ grados de libertad, por tanto el cociente

$$F = \frac{\dfrac{SCTr/\sigma^2}{I-1}}{\dfrac{SCR/\sigma^2}{N-I}} = \frac{\widehat{S}_{Tr}^2}{\widehat{S}_R^2} \tag{4.7}$$

sigue una distribución F de Snedecor con $I-1$ y $N-I$ grados de libertad y es el estadístico de contraste utilizado para probar la hipótesis de interés en el modelo de efectos aleatorios.

Bajo la hipótesis alternativa, es decir $\sigma_\tau^2 \neq 0$, el valor esperado del numerador del estadístico de contraste es mayor que σ^2 y por tanto rechazaremos la hipótesis nula para valores significativamente grandes de F. En otras palabras, en el modelo de efectos aleatorios para contrastar la hipótesis $H_0 : \sigma_\tau^2 = 0$ frente a $H_1 : \sigma_\tau^2 > 0$, se actúa de la siguiente forma

$$Si \;\; F_{exp} > F_{\alpha;I-1,N-I} \;, \quad \text{se rechaza } H_0$$

$$Si \;\; F_{exp} \leq F_{\alpha;I-1,N-I} \;, \quad \text{se acepta } H_0$$

La tabla del análisis de la varianza para el modelo de efectos aleatorios es idéntica a la del modelo de efectos fijos, salvo con la diferencia de que las conclusiones obtenidas se generalizan a toda la población de niveles del factor.

Además de efectuar el contraste de hipótesis, en el modelo de efectos aleatorios interesa estimar los valores σ_τ^2. El procedimiento utilizado para ello se denomina "*método de componentes de la varianza*". Dicho procedimiento consiste en igualar las esperanzas de las varianzas entre tratamientos y residual con sus correspondientes valores muestrales y resolver el sistema resultante:

$$\left.\begin{array}{rl} \widehat{S}^2_{Tr} = & \dfrac{N^2 - \sum\limits_i n_i^2}{N(I-1)}\sigma_\tau^2 + \sigma^2 \\ & \\ \widehat{S}^2_R = & \sigma^2 \end{array}\right\} .$$

Así, se obtienen los siguientes estimadores de las varianzas:

$$\widehat{\sigma}^2 = \widehat{S}^2_R \tag{4.8}$$

$$\widehat{\sigma}^2_\tau = \frac{N(I-1)}{N^2 - \sum\limits_i n_i^2}\left[\widehat{S}^2_{Tr} - \widehat{S}^2_R\right] . \tag{4.9}$$

Puede comprobarse que si la hipótesis alternativa es cierta, entonces $\widehat{\sigma}^2_\tau$ es un estimador insesgado de σ^2_τ.

En el caso de que la varianza residual sea mayor que la varianza entre tratamientos, el método de componentes de la varianza conduce a un estimación negativa de σ^2_τ. Evidentemente ésto carece de sentido, al tratarse de un parámetro no negativo. Cuando ésto ocurre se puede adoptar alguna de las siguientes alternativas:

- Rechazar por inadecuado el modelo propuesto y replantear el problema.
- Reinterpretar la estimación de σ^2_τ como evidencia de que su verdadero valor es cero; es decir, admitir que $\sigma^2_\tau = 0$. Esto puede ocasionar que las propiedades estadísticas de los restantes estimadores se vean perturbadas por esta decisión.
- Utilizar otro método de estimación que no conduzca a estimaciones negativas.

La selección de una opción concreta dependerá del experimento analizado o del criterio del investigador.

Para ilustrar el análisis de la varianza unifactorial de efectos aleatorios (caso no-equilibrado), vamos a resolver el Ejemplo 1-3.

Ejemplo 4.1

En una forja se utilizan varios hornos para calentar muestras de metal. Se supone que todos los hornos operan a la misma temperatura, aunque se sospecha que quizás ésto probablemente no sea cierto. Se seleccionan aleatoriamente tres hornos y se anotan sus temperaturas en sucesivos calentamientos, obteniéndose las observaciones que se muestran en la Tabla 1-14.

Tabla 1-14*. Datos del Ejemplo 1-3*

Horno	*Temperatura*					
1	*91.50*	*98.30*	*98.10*	*93.50*	*93.60*	
2	*88.50*	*84.65*	*79.90*	*77.35*		
3	*90.10*	*84.80*	*88.25*	*73.00*	*71.85*	*78.65*

A partir de estos datos, se desea saber si existe variación significativa en la temperatura de los hornos al nivel de significación del 5%.

El análisis de la varianza resultante se presenta en la siguiente tabla.

Tabla 1-15. Análisis de la varianza para los datos del Ejemplo 1-3

Fuentes de variación	Suma de cuadrados	Grados de libertad	Cuadrados medios	F_{exp}
Entre grupos	594.53	2	297.26	8.62
Dentro de grupos	413.81	12	34.48	
TOTAL	1008.34	14		

Como el valor del estadístico de contraste, 8.62, es mayor que $F_{0{,}05,2,12} = 3{,}89$, se rechaza la hipótesis de que todos los hornos operan a la misma temperatura.

A continuación, aplicando el método de las componentes de la varianza, obtenemos que la estimación de la varianza del error es

$$\widehat{\sigma}^2 = \widehat{S}_R^2 = 34{,}48$$

y la estimación de la varianza de la temperatura es

$$\widehat{\sigma}_\tau^2 = \frac{N(I-1)}{N^2 - \sum_i n_i^2}\left[\widehat{S}_{Tr}^2 - \widehat{S}_R^2\right] = \frac{15 \times 2}{15^2 - 77}\,(297{,}26 - 34{,}48) = 53{,}26 \quad ,$$

de modo que la estimación de la varianza total es igual a

$$\widehat{\sigma}_T^2 = \widehat{\sigma}_\tau^2 + \widehat{\sigma}^2 = 53{,}26 + 34{,}48 = 87{,}74 \quad .$$

Por tanto, la varianza total (87.74) se descompone en una parte atribuible a la diferencia entre hornos (53.26) y otra procedente de la variabilidad existente dentro de ellos (34.48). Comprobamos que en dicha varianza tiene mayor peso la variación entre hornos, en porcentaje un 60.70 % frente a la variación dentro de los hornos, que representa el 39.30 % del total.

4.2. Modelo de efectos aleatorios equilibrado

En el modelo equilibrado o balanceado, el número de observaciones en cada nivel del factor es el mismo. Fijando dicho valor en n, el número total de observaciones es $N = I \times n$. En este caso se obtiene la siguiente expresión para la esperanza de la varianza entre tratamientos

$$E(\widehat{S}_{Tr}^2) = n\sigma_\tau^2 + \sigma^2$$

y los siguientes estimadores para las componentes de la varianza

$$\widehat{\sigma}^2 = \widehat{S}_R^2 \tag{4.10}$$

$$\widehat{\sigma}_\tau^2 = \frac{\widehat{S}_{Tr}^2 - \widehat{S}_R^2}{n} \tag{4.11}$$

Naturalmente, todo lo dicho sobre la estimación de $\widehat{\sigma}_\tau^2$ en el modelo no-equilibrado, es aplicable a este caso.

Desde el punto de vista práctico, el modelo equilibrado presenta la ventaja de que se pueden obtener con Statgraphics las componentes de la varianza, como veremos en la subsección **??** con el siguiente ejemplo

Ejemplo 4.2

Una fábrica de textiles dispone de un gran número de telares. En principio, se supone que cada uno de ellos debe producir la misma cantidad de tela por unidad de tiempo. Para investigar esta suposición se seleccionan al azar cinco telares, y se mide la cantidad de *tela producida en cinco ocasiones diferentes. Se obtienen los datos de la tabla adjunta.*

Tabla 1-16. *Datos de la producción de tela*

Telares	*Producción*				
1	*14.0*	*14.1*	*14.2*	*14.0*	*14.1*
2	*13.9*	*13.8*	*13.9*	*14.0*	*14.0*
3	*14.1*	*14.2*	*14.1*	*14.0*	*13.9*
4	*13.6*	*13.8*	*14.0*	*13.9*	*13.7*
5	*13.8*	*13.6*	*13.9*	*13.8*	*14.0*

¿Del estudio se concluye que todos los telares tienen el mismo rendimiento?

El análisis de la varianza resultante se presenta en la siguiente tabla.

Tabla 1-17. Análisis de la varianza para los datos del Ejemplo 1-4

Fuentes de variación	Suma de cuadrados	Grados de libertad	Cuadrados medios	F_{exp}
Entre grupos	0.3416	4	0.0854	5.77
Dentro de grupos	0.2960	20	0.0148	
TOTAL	0.6376	24		

Si realizamos el contraste al nivel de significación del 5 % y comparamos el valor de la $F_{exp} = 5{,}77$, con el valor de la F teórica ($F_{0{,}05,4,20} = 2{,}87$), se concluye que se rechaza la hipótesis de que todos los telares tienen el mismo rendimiento.

A continuación, aplicamos el método de las componentes de la varianza y obtenemos

$$\widehat{\sigma}^2 = \widehat{S}_R^2 = 0{,}0148$$

$$\widehat{\sigma}_\tau^2 = \frac{\widehat{S}_{Tr}^2 - \widehat{S}_R^2}{n} = \frac{0{,}0854 - 0{,}0148}{5} = 0{,}014 \quad ,$$

por tanto la estimación de la varianza total es igual a

$$\widehat{\sigma}_T^2 = \widehat{\sigma}_\tau^2 + \widehat{\sigma}^2 = 0{,}014 + 0{,}0148 = 0{,}0289 \quad .$$

La varianza total (0.029) se descompone en una parte atribuible a la diferencia entre los telares (0.014) y otra procedente de la variabilidad existente dentro de ellos (0.015).

Comprobamos como en la varianza total tiene más peso la variación dentro de los telares, 51.18 %, que la variación entre los telares, que representa el 48.82 % del total.

Bibliografía utilizada

* **García Leal, J. & Lara Porras, A.M.** (1998). *"Diseño Estadístico de Experimentos. Análisis de la Varianza."* Grupo Editorial Universitario.

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