ESTIMACION DE LOS COMPONENTES DE VARIANZA PARA UN DISEÑO (p × q) FACTORIAL CUANDO LOS EFECTOS SON FIJOS O ALEATORIOS

CARLOS G. VÁSQUEZ PELÁEZ [1]

Resumen

En este trabajo se presentan, en forma de ejemplo, los métodos para estimar los componentes de varianza para el diseño factorial desbalanceado (p × q) para los modelos fijos o aleatorios. Se discute la importancia de la prueba F.

Introducción

En los diseños de experimentos es importante conocer cuáles son los orígenes de la variación, los componentes de varianza y las hipótesis relacionadas con ellos. En la mayoría de los experimentos, algunas de las causas de la variación que componen el modelo son impuestas por el investigador, ya sea en forma de diferentes tratamientos, variedades o prácticas. El total de la variación en un experimento puede ser debido a efectos fijos, aleatorios o a una combinación de ambos. Sin embargo, éstos en muchos casos dependen de la inferencia en que el investigador esté interesado; esto es, si todos los niveles de interés para el investigador están incluidos en el experimento, el factor es fijo y la inferencia de los resultados será válida solamente para los niveles usados en el experimento. El objetivo primordial del investigador en este caso es el análisis de las medias de todos los niveles del factor involucrado. Sin embargo, si el investigador utiliza una muestra al azar del total de niveles pertenecientes al factor, es denominado "aleatorio" y la inferencia relacionada con este factor será válida para todos los posibles niveles de la población en ese factor.

Considérese el caso en el cual existen Q niveles de un factor en la población y solamente q niveles son seleccionados al azar para representar en el experimento los Q niveles de la población, esto es $q \leq Q$. Es decir, existe un coeficiente de *corrección* $(1\text{-}q/Q)$ que se toma en cuenta para la estimación de los componentes de varianza. Se puede observar que si $q = Q$ el coeficiente de corrección es igual a cero, como es el caso de los efectos fijos y se aproxima a uno cuando Q se aproxima a infinito, como sucede en los efectos aleatorios. En general se supone que $Q = q$ y por lo tanto el factor de corrección es cero para los efectos fijos y Q es infinito, por lo que el coeficiente de corrección es uno para los efectos al azar. Bennet y Franklin (1954) presentan algoritmos par obtener esperanza de cuadrados medios suponiendo que Q es menor que infinito.

La estimación de los componentes de varianza para diseños balanceados (igual número de observaciones en cada subclase) con uno o más factores ya sean fijos, aleatorios o una combinación de éstos, ha sido presentada por el método de "Algoritmos de la esperanza de los cuadrados medios" (ECM), el cual ha sido descrito en varios libros de estadística; tal es el caso de Anderson y McLean (1974), Steel y Torrie (1980), etc. Sin embargo, en los modelos desbalanceados (desigual número de observaciones en una o más subclases) que aunque siguen las mismas reglas de los algoritmos de ECM, la estimación de los coeficientes K no es tan directa como en el caso de los diseños balanceados.

Tres métodos para estimar los componentes de varianza para diseños desbalanceados han sido presentados por Henderson (1953)

Recibido para su publicación el 13 de diciembre de 1982.

[1] Departamento de Genética Animal, Instituto Nacional de Investigaciones Pecuarias, SARH, Km 15.5 Carretera México-Toluca, México 10, D.F., C.P. 05110.

y revisados por Searle (1968). Los tres métodos de Henderson incluyen: estimación de los cuadrados medios, estimación de sus esperanzas y la solución de las ecuaciones lineales derivadas de la relación de los cuadrados medios estimados con sus esperanzas. Los tres métodos presentados por Henderson son equivalentes cuando se utilizan en diseños balanceados.

La teoría involucrada en la estimación de los componentes de varianza para diseños desbalanceados ha sido discutida por Searle (1971), Searle y Henderson (1961), y Anderson (1975), por lo que no será revisada en este trabajo.

También están disponibles paquetes estadísticos para la estimación de los componentes de varianza; tal es el caso de: Mínimos cuadrados (Harvey, 1975) y SAS (1979) en los procedimientos GLM y VAR-COM.

El propósito de este trabajo es presentar en forma de ejemplo el método para la estimación de los coeficientes K y la estimación de los componentes de varianza para el diseño desbalanceado: $p \times q$ factorial para los efectos fijos ($q = Q$) y aleatorios ($q < Q$) basado en el método 3 de Henderson.

Modelo:

Los orígenes de la variación en un diseño $p \times q$ factorial cuando la interacción está presente son como sigue:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{(ij)k}$$

Donde Y_{ijk} es la respuesta de la k-ésima observación ($k = 1, 2, \ldots n_{ij}$) en el i-ésimo nivel del factor α ya sea fijo ($q_\alpha = Q_\alpha$) o aleatorio ($q_\alpha < Q_\alpha$) y el j-ésimo nivel del factor β, ya sea fijo o aleatorio; Y_{ij} es el efecto de la interacción entre los niveles de los factores α y β, ya sea fija (si α y β son fijos) o aleatoria (si α o β o ambas son aleatorias); μ es la media general de la población y $\varepsilon_{(ij)k}$ es el error aleatorio DNI $(0,\sigma^2)$. Para efectos fijos de α el componente de varianza es 0 (α) una función de las α y para los efectos aleatorios de α el componente de varianza es σ^2_α, la varianza de la población de efectos α. De igual modo para los otros factores.

El análisis de varianza con la esperanza de los cuadrados medios (ECM) se presenta en el Cuadro 1. Como se puede observar, la estimación de los componentes de varianza es función de los coeficientes K en la columna de la ECM; asimismo esta columna muestra la forma apropiada en que la prueba F debe ser llevada a cabo Es decir, en los modelos fijos, el denominador en todos los casos es el error, mientras que en los modelos aleatorios el deno-

CUADRO 1

Análisis de varianza con la esperanza de los cuadrados medios para el diseño factorial ($p \times q$) para los modelos fijo y aleatorio con la interacción presente

Origen de la variación	Grados de libertad	Cuadrados medios	Esperanza de los cuadrados: Modelo fijo	Esperanza de los cuadrados: Modelo aleatorio
Debido al factor:				
α	$(A-1)$	$CM\alpha$	$\sigma^2 + K_5 \phi \alpha$	$\sigma^2 + K_4 \sigma^2 \gamma + K_5 \sigma^2 \alpha$
β	$(B-1)$	$CM\beta$	$\sigma^2 + K_3 \phi \beta$	$\sigma^2 + K_2 \sigma^2 \gamma + K_3 \sigma^2 \beta$
γ	$(r-a-b+1)$	$CM\gamma$	$\alpha^2 + K_1 \phi \gamma$	$\sigma^2 + K_1 \sigma^2 \gamma$
Error	$(n-r)$	$CM\epsilon$	σ^2	σ^2

A y B son los números de niveles de los factores α, β respectivamente γ es el número de combinaciones de niveles de α y β estudiados y n es el total de observaciones; K_i son los i-ésimos coeficientes de las esperanzas de los cuadrados medios para el modelo fijo y aleatorio.

minador para los factores α y β sería la interacción (γ) finalmente, la interacción será probada contra el error.

Cuando los modelos son balanceados, los coeficientes K_1, K_2 y K_4 son iguales $K_3 = K_5$ para los efectos aleatorios y en modelos fijos $K_1 = K_3 = K_5$ mientras que en los modelos desbalanceados estos coeficientes son diferentes. Debido a esto, es importante estimar los coeficientes K_i con la finalidad de:

i) Estimar los componentes de varianza.

ii) Efectuar la prueba de F con una mejor aproximación.

Estimación de los componentes de varianza

La representación de un diseño factorial ($p = 3 \times q = 2$) con diferente número de observaciones por celda, se muestra en el Cuadro 2, el cual será utilizado para obtener los componentes de varianza, basado en el método 3 de Henderson, así como las pruebas de F, para los efectos fijos o aleatorios.

CUADRO 2

Representación de un diseño (p×q) factorial

Factor α (i)

	1	2	3	n.j
Factor β (j) 1	X_{111}	X_{211}	X_{311} X_{312}	7
	(n_{ijk})	X_{212}	X_{313} X_{314}	
2	X_{121} X_{122}	X_{221}	X_{321} X_{322}	
	X_{123}			6
ni.	4	3	6	13 (n..)

La estimación de los coeficientes k con el método 3 de Henderson, está basada en la solución de las ecuaciones de los cuadrados mínimos.

Como es sabido, un diseño factorial es una extensión del análisis de regresión, en el cual el mismo modelo presentado anteriormente puede ser presentado como $Y = X\beta$ o $X'Y = X'X\beta$ donde Y' tiene dimensiones ($n \times 1$), X' dimensiones ($p \times n$), X'X dimensiones ($p \times p$) y β *dimensiones* ($p \times 1$), donde n es el número de observaciones y p el número de parámetros desconocidos en el modelo; como se puede observar, la matriz de diseño (X'X) está sobreparametrizada, esto es, hay más ecuaciones que incógnitas, por lo que es necesario reparametrizarla; para esto es necesario suponer que: $\Sigma\alpha_i = \Sigma\beta_j = \Sigma\gamma_{ij} = 0$. Al imponer esta condición, la matriz de diseño tendrá dimensiones iguales al número de grados de libertad del modelo, que en el ejemplo. Estas dimensiones serán X'X no reducida (12×12) y X'X reducida (6×6) que corresponden a 1, 2, 1, 2, grados de libertad para μ, α, β y γ respectivamente.

La solución de las ecuaciones simultáneas es $\beta = (X'X)^{-1} X'X$ donde $(X'X)^{-1} = Z$ y los coeficientes k_1, k_3 y k_5 son obtenidos por medio de la inversa de las matrices parciales de Z.

Así los valores K son obtenidos como sigue

$$K_i = \frac{1}{m}\left(\sum Z^{-}_{ii} - \frac{1}{gl}\sum_i \sum_j Z^{-}_{i \neq j}\right)$$

Donde m es el número de clases o niveles en el factor en cuestión, gl los grados de libertad para el mismo factor, y Z^{-} son los elementos ij para la inversa de la submatriz de la parte correspondiente al factor en cuestión en la matriz $(X'X)^{-1}$ reducida.

Para la estimación de los valores K_2 y K_4 en el modelo aleatorio solamente, ya que en el modelo fijo no existe, es necesario obtener la siguiente información: R de las matrices (X'X) reducida y N de la (X'X) no reducida:

$$\begin{array}{cc} & \begin{array}{ccc} \mu & \alpha \text{ o } \beta & \gamma \end{array} \\ R = & \begin{bmatrix} n_{..} & n_{i. \text{ o } .j} & n_{ij} \\ & n_{i. \text{ o } .j} & n_{ij} \\ & & n_{ij} \end{bmatrix} \end{array} \quad \begin{array}{cc} & \mu \\ N = & \begin{bmatrix} \alpha \text{ o } \beta \\ \gamma_{ij} \\ \ \end{bmatrix} \end{array} \begin{bmatrix} \gamma_{ij} \\ \ \\ \ \end{bmatrix}^{1}$$

Los coeficientes K_i se obtienen como sigue:

$$K_i = \frac{1}{gl}\left(n_{..} - \sum_i \sum_j R_{\bar{ij}}\, N_{ij}\right)$$

Donde $N_{ij} = N'N$ no reducida.

Basado en el mismo ejemplo (Cuadro 2) los valores de K son obtenidos como sigue:

La matriz de diseño (X'X) no reducida es:

$$\begin{array}{cc} & \begin{array}{cccccccccccc} \mu & \alpha_1 & \alpha_2 & \alpha_3 & \beta_1 & \beta_2 & \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{21} & \gamma_{22} & \gamma_{23} \end{array} \\ \begin{array}{c} (X'X) = \\ (12 \times 12) \end{array} & \begin{bmatrix} 13 & 4 & 3 & 6 & 7 & 6 & 1 & 2 & 4 & 3 & 1 & 2 \\ & 4 & 0 & 0 & 1 & 3 & 1 & 0 & 0 & 3 & 0 & 0 \\ & & 3 & 0 & 2 & 1 & 0 & 2 & 0 & 0 & 1 & 0 \\ & & & 6 & 4 & 2 & 0 & 0 & 4 & 0 & 0 & 2 \\ & & & & 7 & 0 & 1 & 2 & 4 & 0 & 0 & 0 \\ & & & & & 6 & 0 & 0 & 0 & 3 & 1 & 2 \\ & & & & & & 1 & 0 & 0 & 0 & 0 & 0 \\ & & & & & & & 2 & 0 & 0 & 0 & 0 \\ & & & & & & & & 4 & 0 & 0 & 0 \\ & & & & & & & & & 3 & 0 & 0 \\ & & & & & & & & & & 1 & 0 \\ & & & & & & & & & & & 2 \end{bmatrix} \end{array}$$

Con la restricción impuesta previamente ($\epsilon\, \alpha_i = \epsilon\, \beta_j = \epsilon\, \gamma_{ij} = 0$) en la columna y los renglones de esta matriz X'X de diseño

$$\alpha_1 = -\alpha_3$$
$$\alpha_2 = -\alpha_3$$
$$\beta_1 = -\beta_2$$
$$\gamma_{11} = \gamma_{11} - \gamma_{13} - \gamma_{21} + \gamma_{23}$$
$$\gamma_{12} = \gamma_{12} - \gamma_{13} - \gamma_{22} + \gamma_{23}$$

Así, la matriz reducida (X'X) queda reparametrizada en μ; α_1 y α_2 β_1; γ_{11} y γ_{12}; pudiendo obtener las otras constantes como sigue:

$$\alpha_3 = -\alpha_1 - \alpha_2$$
$$\beta_2 = -\beta_1$$
$$\gamma_{13} = -\gamma_{11} - \gamma_{12}$$
$$\gamma_{21} = -\gamma_{11}$$
$$\gamma_{22} = -\gamma_{12}$$
$$\gamma_{23} = -\gamma_{21} - \gamma_{22}$$

$$\begin{array}{cc} & \begin{array}{cccccc} \mu & \alpha_1 & \alpha_2 & \beta_1 & \gamma_{(11)} & \gamma_{(12)} \end{array} \\ \begin{array}{c} (x'x) = \\ \text{Reducida} \\ (6 \times 6) \end{array} & \begin{bmatrix} 13 & -2 & -3 & 1 & -4 & -1 \\ & 10 & 6 & -4 & 0 & 2 \\ & & 9 & -1 & 2 & 3 \\ \text{Simétrica} & & & 13 & -2 & -3 \\ & & & & 10 & 6 \\ & & & & & 9 \end{bmatrix} \end{array}$$

$$\begin{array}{cc} & \begin{array}{cccccc} \mu & \alpha_1 & \alpha_2 & \beta_1 & \gamma_{(11)} & \gamma_{(12)} \end{array} \\ (X'X)^{-} = Z = & \begin{bmatrix} .0995 & .0116 & .0255 & -.0023 & .0579 & -.0294 \\ & .2106 & -.1366 & .0579 & .0532 & -.0162 \\ & z_\alpha & .2245 & .-0394 & -.0162 & .0439 \\ & & z_\beta & .0995 & .0116 & .0255 \\ & & & & .2106 & -.1366 \\ & & & & z_\gamma & .2245 \end{bmatrix} \end{array}$$

$$z_\alpha^- = z_\gamma^- = \begin{bmatrix} .2106 & -.1366 \\ -.1366 & .2245 \end{bmatrix}^{-1} = \begin{bmatrix} 7.83 & 4.77 \\ 4.77 & 7.35 \end{bmatrix}$$

$$Z^- = (0.0995)^{-1} = 10.04$$

Usando estos resultados.

Los coeficientes K_1, K_3 y K_5 son obtenidos de la siguiente manera:

$$K_1 = \frac{1}{6}\left(7.83 + 7.35 - \frac{1}{2}\, 2\,(4.77)\right) = 1.73$$

$$K_3 = \frac{1}{2}\,(10.04) = 5.02$$

$$K_5 = \frac{1}{3}\left(7.83 + 7.35 - \frac{1}{2}\, 2\,(4.77)\right) = 3.47$$

Para la estimación del coeficiente K_2, en la ECM del factor β, la matriz R identificada como R_2 se obtiene de la matriz $(X'X)$ reducida y N_2 de la matriz X'X no reducida, así:

$$R_2 = \begin{array}{c} \begin{array}{ccccc} \mu & \alpha_1 & \alpha_2 & \gamma_{(11)} & \gamma_{(12)} \end{array} \\ \begin{bmatrix} 13 & -2 & -3 & -4 & -1 \\ & 10 & 6 & 0 & 2 \\ & & 9 & 2 & 3 \\ & & & 10 & 6 \\ & & & & 9 \end{bmatrix} \end{array}$$

$$N_2 = \begin{array}{c|c} & \begin{array}{cccccc} \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{21} & \gamma_{22} & \gamma_{23} \end{array} \\ \begin{array}{c} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{13} \\ \gamma_{21} \\ \gamma_{22} \\ \gamma_{23} \end{array} & \begin{bmatrix} 1 & 2 & 4 & 3 & 1 & 2 \\ 1 & 0 & 0 & 3 & 0 & 0 \\ 0 & 2 & 0 & 0 & 1 & 0 \\ 0 & 0 & 4 & 0 & 0 & 2 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix} \end{array}$$

Siendo la matriz NN' no reducida de dimensiones (10×10).

$$N N' = \begin{array}{c} \begin{array}{cccccccccc} \mu & \alpha_1 & \alpha_2 & \alpha_3 & \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{21} & \gamma_{22} & \gamma_{23} \end{array} \\ \begin{bmatrix} 35 & 10 & 5 & 20 & 1 & 4 & 16 & 9 & 1 & 4 \\ & 10 & 0 & 0 & 1 & 0 & 0 & 9 & 0 & 0 \\ & & 5 & 0 & 0 & 4 & 0 & 0 & 1 & 0 \\ & & & 20 & 0 & 0 & 16 & 0 & 0 & 4 \\ & & & & 1 & 0 & 0 & 0 & 0 & 0 \\ & \text{SIMETRICA} & & & & 4 & 0 & 0 & 0 & 0 \\ & & & & & & 16 & 0 & 0 & 0 \\ & & & & & & & 9 & 0 & 0 \\ & & & & & & & & 1 & 0 \\ & & & & & & & & & 4 \end{bmatrix} \end{array}$$

Con las mismas restricciones para columnas y renglones de $\sum \alpha_i = \sum \gamma_{ij} = 0$ esto es:

$$\alpha_1 - \alpha_3;\ \alpha_2 - \alpha_3$$

$$\gamma_{11} - \gamma_{13} - \gamma_{21} + \gamma_{23}$$

$$\gamma_{12} - \gamma_{22} - \gamma_{13} + \gamma_{23}$$

Obtenemos la matriz NN' reducida con dimensiones (5×5)

$$(N_{ij}) = NN' \text{ reducida} = \begin{bmatrix} 35 & -10 & 15 & -20 & -9 \\ & 30 & 20 & 4 & 12 \\ & & 25 & 12 & 15 \\ & & & 30 & 20 \\ & & & & 25 \end{bmatrix}$$

$$R_2^{-1} = \begin{bmatrix} .099 & .0129 & .0245 & .0581 & -.0387 \\ & .177 & -.1137 & .0465 & -.031 \\ & & .209 & -.0116 & -.0339 \\ & & & .2093 & -.139 \\ & & & & .218 \end{bmatrix}$$

$$K_2 = \frac{1}{1}\ 13\text{–}35\ (.099) - (-10)\ (.0129) -$$

$$- \ldots - 25\ (.218) = 1.67.$$

De la misma manera, para obtener K_4

$$R_4 = \begin{array}{c} \mu \quad \beta_1 \quad \gamma_{(11)} \quad \gamma_{(12)} \\ \begin{bmatrix} 13 & 1 & -4 & -1 \\ & 13 & -2 & -3 \\ & & 10 & 6 \\ & & & 9 \end{bmatrix} \end{array}$$

$$\begin{array}{l}(N_{ij}) = NN \\ \text{reducida}\end{array} \begin{bmatrix} 35 & 7 & -20 & -9 \\ & 35 & -10 & -15 \\ & & 30 & 20 \\ & & & 25 \end{bmatrix}$$

$$K_4 = \frac{1}{2}\ 13\text{–}35\ (.09) - 7\ (-.005) - \ldots -$$

$$- 25\ (.201) = .82$$

Hay que hacer notar que cuando se obtuvo el valor de K_2 la matriz R_2 y N_2 no contiene el efecto de β. De la misma manera cuando se obtuvo K las matrices R_4 y N_4 no muestran los valores de α.

Por lo tanto, la esperanza (E) de los cuadrados medios es igual a los estimadores de los componentes de varianza, pudiendo ser obtenidos bajo el siguiente sistema de ecuaciones lineales.

$$E \begin{bmatrix} CM\alpha \\ CM\beta \\ CM\gamma \\ CM\epsilon \end{bmatrix} = \begin{bmatrix} 1 & K_4 & 0 & K_5 \\ 1 & K_2 & K_3 & 0 \\ 1 & K_1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{\sigma}^2 \\ \hat{\sigma}^2_\gamma \\ \hat{\sigma}^2_\beta \\ \hat{\sigma}^2_\alpha \end{bmatrix}$$

Para el modelo aleatorio y para el modelo fijo, sólo que en este último los valores de K_4 y K_2 son iguales a 0.

Los valores obtenidos por el método 1 (Vásquez, 1983) y 3 de Henderson nos llevan a soluciones diferentes como puede verse en el Cuadro 3.

Cuando los modelos son desbalanceados, la varianza de los estimadores de los componentes es desconocida. Sin embargo, Harvey (1975) sugiere que el uso del método 3 de Henderson es más eficiente aunque más laborioso, sin embargo hay que hacer

CUADRO 3

Análisis de varianza con la esperanza de los cuadrados medios para un diseño factorial (3×2) basado en el método 1 y 3 de Henderson para los modelos fijos y aleatorios

		Esperanza de los cuadrados	
Origen de la variación	Grados de libertad	Medios Método 1	Modelo aleatorio Método 3
α	2	$\sigma^2 + 2.41\,\sigma^2\gamma + 4.16\,\sigma^2\alpha$	$\sigma^2 + 0.82\,\sigma^2\gamma + 3.47\,\sigma^2\alpha$
β	1	$\sigma^2 + 1.42\,\sigma^2\gamma + 6.46\,\sigma^2\beta$	$\sigma^2 + 1.67\,\sigma^2\gamma + 5.02\,\sigma^2\beta$
γ	2	$\sigma^2 + 1.43\,\sigma^2\gamma$	$\sigma^2 + 1.73\,\sigma^2\gamma$
Error	7	σ^2	σ^2
		Modelo fijo	
		Método 1	Método 3
		$\sigma^2 + 4.16\,\phi\alpha$	$\sigma^2 + 3.47\,\phi\alpha$
		$\sigma^2 + 6.46\,\phi\beta$	$\sigma^2 + 5.02\,\phi\beta$
		$\sigma^2 + 1.43\,\phi\gamma$	$\sigma^2 + 1.73\,\phi\gamma$
		σ^2	σ^2

notar que cualesquiera de estos métodos son aproximados y los dos son igualmente válidos. Con esto se hace notar que con el uso de paquetes estadísticos en donde puede obtenerse la estimación de los componentes de varianza éstos pueden mostrar valores diferentes aún con los mismos datos, debido al procedimiento de estimación.

Importancia de los Coeficientes K en la Prueba de F.

Por medio de la columna de la Esperanza de los Cuadrados Medios (ECM) se puede reconocer cuáles son los errores que probarán si el(los) efecto(s) de interés, muestran o no diferencias estadísticas significativas, es decir, la hipótesis de interés es que el componente de varianza con efectos aleatorios debida al factor de interés es igual a una constante, esto es $H_0{:}\sigma_i^2 = m$. generalmente $m = 0$. En los diseños balanceados o desbalanceados, esta prueba se realiza comparando el valor del cuadrado medio del factor de interés con su respectivo error, o lo que es lo mismo, comparando las esperanzas de los cuadrados medios involucrados en la prueba. Como se observa en el ejemplo, la prueba de F sería: $CM\alpha/CM\gamma$ y $CM\beta/CM\gamma$ para los factores α y β respectivamente en el modelo aleatorio balanceado. La prueba de F es exacta si: $K_1 = K_4$ para el factor α y $K_1 = K_2$ para el factor β. Sin embargo en los modelos desbalanceados donde $K_1 \neq K_2 \neq K_4$ como se mostró en el ejemplo, las pruebas de F no son exactas. Por ejemplo, tómese el factor β: su error es la interacción γ y la prueba de F será, pues:

$$F = \frac{CM_\beta}{CM_\gamma}$$

donde la esperanza

$$E\,(CM_\beta) = \sigma^2 + k_2\,\sigma_\gamma^2 + K_3\,\sigma_\beta^2$$

$$E\,(CM_\gamma) = \sigma^2 + K_1\,\sigma_\gamma^2$$

Si la hipótesis es cierta, es decir: $\sigma_\beta^2 = 0$

$$\frac{E\,(CM_\beta)}{E\,(CM_\gamma)} = \frac{\sigma^2 + K_2\,\sigma_\gamma^2}{\sigma^2 + K_1\,\sigma_\gamma^2}$$

Es obvio que si $K_2 = K_1$ y la hipótesis es cierta, la F es central con valores cercanos a uno como sucede en los diseños balanceados. Sin embargo, en los diseños desbalanceados donde $K_2 \neq K_1$ el cociente F puede no tener distribución central y mostrar tendencia a valores mayores de uno si $K_2 > K_1$ y menor de uno si $K_2 < K_1$. Estas diferencias entre los valores de K son de importancia cuando los diseños presentan un gran desbalance en las observaciones, debido a que en este caso, las diferencias estadísticas podrían estar confundidas con los valores de K y no *a* la varianza debida al factor de interés. Cuando el diseño es de esta naturaleza, es importante conocer los valores de K para ajustar las pruebas de F utilizando como denominador una combinación entre el cuadrado medio del error y la interacción, así la prueba F sería:

$$F = \frac{CM_\beta}{\dfrac{k_2}{k_2}\,CM) + (1 - \dfrac{k_2)}{k_1}\,CM_{error}}$$

con grados de libertad (β-1) para el numerador y usando la aproximación presentada por Satterthwaite (1946) para el denominador.

$$gl = \frac{M^2}{\sum_{i=1}^{k} \dfrac{(a_i\,M_i)^2}{f_i}}$$

Donde

$$M = \sum_{i=1}^{k} a_i\,M_i$$

M_i es el i-ésimo mínimo cuadrado del denominador; a_i es el coeficiente del i-ésimo mínimo cuadrado; f_i son los grados de libertad (gl) del i-ésimo cuadrado medio.

O bien la utilización de otro análisis, tal es el caso del análisis de mínimos cuadrados ponderados o análisis proporcional.

Summary

This paper presents, in form of example, the methods for estimating the variance components for a two-way clasification (p × q factorial design) for fixed and random unbalanced models. The F test is discussed.

Literatura citada

ANDERSON, L.R., 1975, Design and Estimators for Variance Components. J.N. Srivastava, Ed. A Survey of Statistical Designs and Linear Models. *North-Holland Publishing Co.*

ANDERSON, V.L. and R.A. MCLEAN, 1974, Design of Experiments. A Realistics Approach. *Marcel Dekker, Inc.*, N.Y.

BENNETT, C.A. and FRANKLIN, N.L., 1954, Statistical Analysis in Chemistry and the Chemical Industry. *Wiley, N.Y.*

HARVEY, R.W. 1975, Least-squares. Analysis of Data with Unequal Subclass Numbers. AHS H-4 Agriculture Research Service. *U.S. Department of Agriculture.*

HENDERSON, C.R., 1953, Estimation of Variance and Covariance Components. *Biometrics, 9:*226-252.

SAS Institute, 1979, *SAS User's Guide.* North Carolina.

SATTERTHWAITE, F.E., 1946, An approximate distribution of estimates of variance components. *Biometrics Bulletin, 2:*110-114.

SEARLE, S.R., 1956, Matrix Methods in Variance and Covariance Components Analysis. *Ann. Math. Statist., 27:*737-748.

SEARLE, S.R., 1968, Another look at Henderson's Methods of Estimating Variance Components. *Biometrics, 24:*749-488.

SEARLE, S.R., 1971, Topics in Variance Components Estimation. *Biometrics, 27:*1-76.

SEARLE, S.R. and C.R. HENDERSON, 1961, Computing Procedures for Estimating Components of Variance in the Two-way Classification, Mixed Model. *Biometrics, 17:*6-70616.

STEELE, G.D.R. and J.H. TORRIE, 1980, Principles and Procedures of Statistics. A Biometrical Approach. 2nd Ed. *McGraw-Hill Book Co.*