Revista Colombiana de Estadística
Nº 9 - 1984

# DISENOS DE BLOQUES TOTALMENTE BALANCEADOS ANALISIS DE VARIANZA Y FACTOR DE EFICIENCIA

*José Alberto Vargas Navas*

**Profesor Asistente**
**Universidad Nacional**

## 1. Introducción.

En diversas situaciones experimentales, es frecuente utilizar los diseños de bloques; aunque el más común es el diseño de bloques completos al azar, éste no es aplicable en todos los casos, pues una de las primeras dificultades para probar el efecto de varios tratamientos, es conseguir bloques homogéneos; a medida que aumenta el tamaño del bloque, disminuye la homogeneidad de las unidades experimentales en el mismo. Se acostumbra entonces a construir bloques homogéneos mas pequeños, de tal forma que el número de tratamientos aplicados en cada bloque sea menor que el número total a probarse. En el presente artículo se derivan sistemáticamente los principales resultados del análisis de experimentos de bloques, usando notación matricial.

Además de presentar un método para el análisis de varianza de un diseño de bloques incompletos, se estudia el factor de eficiencia, en razón a que el experimentador puede tener a su disposición un gran número de diseños del mismo tamaño y por tanto necesitar una herramienta que le permita seleccionar un buen diseño; el factor de eficiencia es una de ellas.

## 2. Notación.

Se consideran $t$ tratamientos aplicados a $N$ unidades experimentales, distribuidas en $b$ bloques de tamaño $k_1, k_2, \ldots, k_b$; sea $K' = (k_1, k_2, \ldots, k_b)$ el vector de tamaños de bloque y $R' = (r_1, r_2, \ldots, r_t)$ el vector de replicaciones de tratamiento, donde $r_i$ es el número de unidades a las que se le ha aplicado el $i$-ésimo tratamiento; se nota además por $\lambda_{ii'}$ el número de bloques en que ocurren juntos el $i$-ésimo e $i'$-ésimo tratamientos ($i, i' = 1, 2, \ldots, t$; $i \neq i'$).
Sean $T' = (T_1, T_2, \ldots, T_t)$ y $B' = (B_1, B_2, \ldots, B_b)$

los vectores de totales de tratamiento y totales de bloque respectivamente; $n = (n_{ij})_{t\times b}$ la matriz de incidencia, donde $n_{ij}$ es el número de veces que el $i$-ésimo tratamiento ocurre en el $j$-ésimo bloque, si $n_{ij} = 1$ ó $n_{ij} = 0$, el diseño se llama binario, que son los que se consideran en el presente trabajo.

Se denota por $X^{a\delta}$ la matriz diagonal, cuyos elementos diagonales son los del vector $X$ elevados a la potencia $a$.

## 3. Modelo y estimación de los parámetros.

El modelo lineal general $Y = A\theta+\varepsilon$ de un diseño de bloques es

$$Y = [\mathbb{1},D',\Delta'][\alpha,\beta',\gamma']'+ \varepsilon$$

donde: $A \doteq [\mathbb{1},D',\Delta']$ es la matriz de diseño $N\times(1+b+t)$ de rango $1+b+t-2$; $\mathbb{1}$ es el vector de unos, $D'$ es la matriz $N\times b$ con una fila para cada unidad y una columna para cada bloque y $\Delta'$ es la matriz $N\times t$ con una fila para cada unidad y una columna para cada tratamiento.

$\theta = [\alpha,\beta',\gamma']'$, donde $\alpha$ es la media general, $\beta$ es el vector $b\times 1$ de parámetros de bloque y $\gamma$ es el vector $t\times 1$ de parámetros de tratamiento.

$\varepsilon$ es el vector $N\times 1$ de errores aleatorios, con una distribución especificada por $E(\varepsilon) = 0$ y $E(\varepsilon\varepsilon') = \sigma^2 I$.

Se cumplen las siguientes relaciones: $n = \Delta D'$, $K^\delta = DD'$, $R^\delta = \Delta\Delta'$, $B = DY$, $T = \Delta Y$ y el gran total $G = B'\mathbb{1} = T'\mathbb{1}$.

Las ecuaciones normales del modelo propuesto son:

$$A'A\hat{\theta} = A'Y$$

por tanto, una solución de las ecuaciones normales obtenida por el método de mínimos cuadrados es

$$\hat{\theta} = (A'A)^{-}A'Y$$

donde $(A'A)^{-}$ es una inversa generalizada de la matriz $A'A$; una de estas es la inversa de la matriz $A'A + C'C$ donde $C$ está dada por:

$$C = \frac{1}{\sqrt{N}}\begin{bmatrix} 0 & K' & 0' \\ 0 & 0' & R' \end{bmatrix}.$$

Los parámetros $\theta = [\alpha, \beta', \gamma']'$ son entonces estimables bajo las condiciones $C\theta = 0$ y así:

$$\hat{\theta} = (A'A + C'C)^{-1}A'Y$$

se obtiene

$$\hat{\theta} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} G/N \\ K^{-\delta}(B-(\frac{G}{N})K+n'\Omega nK^{-\delta}B-n'\Omega T) \\ \Omega Q \end{bmatrix}$$

donde $$\Omega^{-1} = R^{\delta} - nK^{-\delta}n' + \frac{RR'}{N}$$

y

$$Q = T - nK^{-\delta}B \ .$$

A fin de efectuar el análisis de varianza, se calculan las respectivas sumas de cuadrados como se muestra a continuación: La suma de cuadrados del error está dada por la siguiente expresión

$$\begin{aligned} SCE &= (Y-A\hat{\theta})'\ (Y-A\hat{\theta}) \\ &= Y'Y-2\hat{\theta}'A'Y+\hat{\theta}'A'A\hat{\theta} \\ &= Y'Y-B'K^{-\delta}B-Q'\Omega Q. \end{aligned}$$

Como la suma de cuadrados de bloques viene dada por $SCB = B'K^{-\delta}B- \frac{G^2}{N}$, la suma de cuadrados de tratamientos, obtenida por diferencia, es

$$SCT = Q'\Omega Q.$$

La tabla de análisis de varianza ya se puede construir.

*Tabla 1*

**Análisis de varianza de un diseño de bloques incompletos**

| F.de V. | gl | SC |
|---|---|---|
| Bloques | $b - 1$ | $B'K^{-\delta}B - G^2/N$ |
| Tratamientos | $t - 1$ | $Q'\Omega Q$ |
| Error | $N-b-t-1$ | $Y'Y - B'K^{-\delta}B - Q'\Omega Q$ |
| Total | $N - 1$ | $Y'Y - G^2/N$ |

## 4. Caracterización de los contrastes de tratamiento.

**4.1.** Descomposición de los contrastes de tratamiento en componentes inter e intrabloque.

**Definición 4.1.1.** Una función lineal $S'\Delta Y = S'T$ es llamada un contraste de tratamiento cuando $S'\Delta\mathbf{1} = S'R = 0$, se nota por $S$; y, una función lineal $U'DY = U'B$ es llamada un contraste de bloque cuando $U'D\mathbf{1} = U'K = 0$, se nota por $U$.

**Definición 4.1.2.** Dos contrastes $P_1$ y $P_2$ son ortogonales si $P_1'P_2 = 0$.

**Definición 4.1.3.** Un constraste que sea ortogonal a todos los contrastes de bloque se denomina contraste intrabloque y un contraste de blo que se distingue de los contrastes intrabloques describiéndolos como interbloques.

Dado un contraste de tratamiento $S$ se puede encontrar $U$ tal que $S'\Delta Y = U'DY+(S'\Delta-U'D)Y$, donde $\Delta'S-D'U$ es un contraste intrabloque; la condición para que se cumpla este requisito es $D(\Delta'S-D'U) = 0$, requisito que es equivalente a $DD'U = D\Delta'S$, o sea $K^{\delta}U = n'S$, determinando $U = K^{-\delta}n'S$ de manera única.

Se tiene entonces que $S'T = S'nK^{-\delta}B+S'Q$, es decir, que un contraste de tratamiento es resoluble de manera única en componentes inter e intrabloque.

**4.2.** Condición para que las componentes inter e intrabloque estimen el mismo contraste de tratamiento.

La descomposición en 2 componentes puede ser aplicada a cualquier contraste de tratamiento, pero es útil, solo si las componentes mismas estiman el mismo contraste.

Sea $\tau$ el vector $t\times 1$ cuyos elementos son los valores esperados de una sola observación en las diferentes clases de tratamientos, es decir

$$E(Y) = \Delta'\tau$$

así,

$$E(S'\Delta Y) = S'\Delta\Delta'\tau = S'R^{\delta}\tau$$

y

$$E(U'\mathcal{D}Y) = U'\mathcal{D}\Delta'\tau = U'n'\tau \ .$$

Si los dos estimativos tienen la misma esperanza, aparte de un factor constante, para todo $\tau$ se tiene:

$$U'n' = \mu S'R^{\delta}$$

ó

$$S'nK^{-\delta}n' = \mu S'R^{\delta}$$

donde $\mu$ es un escalar que depende de $S$, $\mathcal{D}$ y $\Delta$.

Si se escribe

$$M = R^{-\delta}nK^{-\delta}n'$$

la ecuación

$$S'nK^{-\delta}n' = \mu S'R^{\delta}$$

resulta ser

$$MS = \mu S.$$

Luego una condición necesaria y suficiente para que un contraste de tratamiento $S$ sea resoluble en componentes inter e intrabloque, cada una estimado el mismo contraste de tratamiento, es que debe ser un vector propio de $M$.

Sea
$$M_o = M - \frac{\mathbf{1}R'}{N} \ .$$

Post-multiplicando por $S$ se obtiene:

$$M_o S = MS - \frac{\mathbf{1}R'}{N} S$$

$$= \mu S - \frac{\mathbf{1}R'}{N} S = \mu(I - \frac{\mathbf{1}R'}{N}) S$$

o sea $M_o S = \mu S$ ya que $R'S = 0$.

$\mu$ es el único valor propio diferente de cero de $M_o$, en razón a que la ecuación característica es:

$$|M_o - \alpha I| = (-1)^t \alpha(\alpha-\mu)^{t-1}$$

y por tanto la multiplicación de $\mu$ es $t-1$.

Como $M_o \mathbf{1} = 0$, el valor propio cero corresponde el vector $\mathbf{1}$. Luego si $M_o$ es de la forma $M_o = \mu(I - \frac{\mathbf{1}R'}{N})$, cualquier contraste de tratamiento es un vector propio de $M_o$ y también de $M$.

## 5. Conceptos de balance y eficiencia.

Las siguientes definiciones fueron dadas por Jones (1959) y Calinski (1971):

**Definición 5.1.** Un diseño de bloques es balanceado para un contraste $S$, cuando $S$ es un vector propio de $M_o$. Es decir $M_o S = \mu S$; el correspondiente valor propio de $M_o$, $\mu$, se llama el factor de pérdida y $(1-\mu)$ el factor de eficiencia del diseño con respecto a $S$.

**Definición 5.2.** Si un diseño de bloques es balanceado para todos los contrastres de parámetros de tratamiendo con exactamente el mismo factor de pérdida, este será llamado un diseño de bloques totalmente balanceado.

Se ha demostrado por lo tanto, que la condición necesaria y suficiente para que un diseño de bloques sea un diseño de bloques totalmente balanceado es que $M_o$ sea de la forma:

$$M_o = \mu\left(I - \frac{\mathbf{1}R'}{N}\right)$$

donde $\mu$ es el único valor propio diferente de cero de la matriz $M_o$ con multiplicidad $t-1$.

Es conveniente aquí, explicar en mas detalle el significado que tiene el factor de eficiencia en vista de su importancia para la elección de un diseño: $\mu$ es llamado el factor de pérdida ya que es una medida de la pérdida relativa de información debido a la confusión parcial de los contrastes de tratamiento con el efecto de bloques. En un diseño totalmente balanceado, esta pérdida es igual para todos los contrastes de tratamiento y, por tanto, la eficiencia del diseño, definida en términos de la información efectiva obtenida sobre cualquier contraste de tratamiento es $1-\mu$.

## 6. Un diseño totalmente balanceado especial.

Una clase particular de los diseños totalmente balanceados son los diseños de bloques incompletos balanceados (BIB). En estos diseños se consideran las unidades arregladas en $b$ bloques de $k$ unidades cada uno, y $t$ tratamientos, cada uno replicado $r$ veces; además, dos tratamientos cualesquiera ocurren juntos en el mismo bloque $\lambda$ veces.

Entonces

$$M = R^{-\delta} n K^{-\delta} n' = (\frac{1}{r} I) n (\frac{1}{k} I) n' = \frac{1}{rk} nn'$$

$$M = \frac{1}{rk} \begin{bmatrix} r & \lambda & \lambda & \dots & \lambda \\ \lambda & r & \lambda & \dots & \lambda \\ \vdots & & & & \\ \lambda & \lambda & \lambda & \dots & r \end{bmatrix}$$

y como la suma de las filas debe ser 1, pues $M\mathbb{1} = \mathbb{1}$.

$$\frac{r+(t-1)\lambda}{rk} = 1$$

luego,

$$\lambda = \frac{r(k-1)}{t-1}$$

Ahora bien,

$$M_o = M - \frac{\mathbb{1} R'}{N} ,$$

es decir:

$$M_o = \begin{bmatrix} \frac{t-k}{kt} & \frac{\lambda}{rk} - \frac{1}{t} & \cdots & \frac{\lambda}{rk} - \frac{1}{t} \\ \frac{\lambda}{rk} - \frac{1}{t} & \frac{t-k}{kt} & \cdots & \frac{\lambda}{rk} - \frac{1}{t} \\ \vdots & & & \\ \frac{\lambda}{rk} - \frac{1}{t} & \frac{\lambda}{rk} - \frac{1}{t} & \cdots & \frac{t-k}{kt} \end{bmatrix}$$

por lo tanto, los diseños BIB satisfacen

$$M_o = \mu(I - \frac{\mathbb{1}R'}{N})$$

con $\mu = \frac{t-k}{k(t-1)}$ y el factor de eficiencia estará dado por

$$1 - \mu = 1 - \frac{t-k}{k(t-1)} = \frac{t(k-1)}{t(k-1)} \; .$$

## 7. Conclusiones.

**7.1.** Se han deducido y presentado matricialmente las propiedades de los diseños de bloques incompletos. La notación matricial, a la vez que facilita la manipulación de los datos cuando se usa un diseño de bloques incompletos totalmente balanceado, permite visualizar y comprender las propiedades de estos diseños; es así como las propiedades de un diseño de bloques incompletos balanceado, se deducen fácilmente.

**7.2.** El cálculo y propiedades de las matrices $\Omega$ y $M_o$ es importante en el análisis de los datos de un experimento que ha sido diseñado en bloques incompletos. $\Omega$ por ser la matriz de covarianzas del vector de medias de tratamiento ajustadas y $M_o$ en razón a que sus propiedades relacionadas con los contrastes de tratamiento determinan el factor de eficiencia del diseño. Se ha mostrado, que cuando el diseño de bloques es totalmente balanceado, el cálculo de estas matrices no presenta problema.

**7.3.** El análisis de varianza que se ha presentado aquí es el intrabloque; por tanto, el método desarrollado es útil, solo cuando la información interbloque se puede considerar despreciable.

**7.4.** El adjetivo de balanceado asegura que todos los pares de tratamiento se comparen con la misma precisión, aunque existen grandes diferencias entre los bloques.

**7.5.** Aunque se dispone de un gran número de diseños, es posible que no haya un diseño adecuado para el número exacto de tratamientos proyectados inicialmente para comparar. Es importante entonces, el trabajo conjunto entre el experimentador y el estadístico, a fin de intercambiar ideas sobre los recursos disponibles para el tra bajo, de tal forma que el número de tratamientos en ciertos casos pudiera aumentarse, ó, en otros, disminuirse ligeramente, con el objeto de ajustarlos a un diseño disponible.

* *

# BIBLIOGRAFIA

Calinski, T., On some desirable patterns in block designs. Biometrics. Nº 27, p.275-292, 1971.

Ceranka, B., Affine resovable incomplete block designs. Applicationes Mathematicae. V. 14, Nº 4, p.565-572, 1975.

Jones, R.M., On a property of incomplete block. J.R. Statist. Soc. Serie B. Nº 27, p.172-179, 1959.

Pearce, S.C., The efficiency of block design in general. Biometrika V. 57, Nº 2, p.339-346, 1970.

Raghavarao, D., Constructions and combinatorial problem in designs of experiments. John Wiley & Sons, Inc. 1971.

Rees, D.H., The analysis of variance of designs with many non-orthogonal classifications. J.R. Statist. Soc. Serie B, Nº 28, p.110-

117, 1966.

Tocher, K.D., The design and analysis of block experiments. J.R. Statist. SOC. Serie B, Nº 14, p. 45-91, 1952.

* *