

**UNIVERSIDAD DE EL SALVADOR
FACULTAD MULTIDISCIPLINARIA DE OCCIDENTE
DEPARTAMENTO DE MATEMÁTICA**



Licenciatura en Estadística

Control Estadístico del Paquete R

**”UNIDAD SEIS”
Práctica 25 - Diseños por bloques**

**Alumna:
Martha Yoana Medina Sánchez**

**Fecha de elaboración
Santa Ana - 27 de noviembre de 2015**

Una variable o factor cuyo efecto sobre la variable respuesta no es directamente de interés, pero que se introduce en el experimento para obtener comparaciones más homogéneas, se denomina una variable bloque. La diferencia principal entre un factor cualquiera y una variable bloque es que, en general, se supone que no hay interacción entre la variable bloque y la variable factor. En resumen, la variable bloque se introduce para eliminar de manera sistemática las comparaciones estadísticas entre los tratamientos (la variable bloque se introduce con el fin de reducir la variabilidad experimental).

Supondremos que tenemos una variable factor con k niveles, o mejor dicho tenemos k tratamientos; mientras que tenemos una variable bloque con n niveles, o si lo prefiere n bloques. Supondremos que tomamos una observación para cada combinación de tratamiento-bloques (se supone que los tratamientos son asignados de manera aleatoria dentro de cada uno de los bloques).

El modelo (basado en los resultados para un único factor) que genera los datos es el siguiente:

$$y_{ij} = \mu + t_i + B_j + e_{ij}$$

Donde:

- y_{ij} : Representa la observación en el j -ésimo bloque del i -ésimo tratamiento.
- μ : Representa un promedio o efecto global.
- t_i : Representa el efecto del i -ésimo tratamiento. Debe cumplirse sumatoria de $t_i = 0$.
- B_j : Representa el efecto del j -ésimo bloque. Deben cumplir sumatoria de $B_j = 0$.
- e_{ij} : Representa un componente de error aleatorio, llamado perturbaciones, que incorpora todas las demás fuentes de variabilidad del experimento (no incluidas ni en los tratamientos ni en los bloques).

Las cuatro hipótesis básicas del modelo se resumen en $u_{ij} = NIID N(0; \sigma^2)$; para todo i, j .

La hipótesis a probar es como siempre:

$$H_o: \mu_1 = \mu_2 = \mu_3 \dots \mu_k$$

$$H_1: \mu_i \text{ distinto } \mu_j; \text{ para al menos un par } i \text{ distinto de } j$$

Que en términos de efectos de grupos son:

$$H_o: t_1 = t_2 \dots t_k = 0$$

$$H_1: t_i \text{ distinto } 0; \text{ para al menos un } i$$

No resulta difícil verificar utilizando el método de máxima verosimilitud que el modelo estimado para una muestra aleatoria de tamaño $N = kn$ es:

$$\hat{y}_{ij} = \hat{\mu} + \hat{t}_i + \hat{B}_j$$

Y por consiguiente:

$$y_{ij} = \hat{\mu} + \hat{t}_i + \hat{B}_j + \hat{u}_{ij}$$

Donde:

- $\hat{\mu} = \tilde{y} ..$
- $\hat{t}_i = \tilde{y} i. - \tilde{y} ..$
- $\hat{B}_j = \tilde{y} .j - \tilde{y} ..$
- $\hat{u}_{ij} = y_{ij} - \tilde{y} .j - \tilde{y} i. + \tilde{y} ..$

Y se tendrán las siguientes medidas de interés:

- $\tilde{y} i.$ es el promedio para el i-ésimo tratamiento.
- $\tilde{y} .j$ es el promedio para el j-ésimo bloque.
- $\tilde{y} ..$ es la media general de la característica de interés.

El Análisis de Varianza establece que se debe cumplir la siguiente relación (al ser cada uno de las fuentes ortogonales entre sí):

$$VT = VE(t) + VE(B) + VNE$$

Donde:

- VT es la variabilidad total del experimento.
- $VE(t)$ es la variabilidad explicada por los tratamientos.
- $VE(B)$ es la variabilidad explicada por los bloques.
- VNE es la variabilidad no explicada o residual.

Para poder contrastar simultáneamente la igualdad de las k medias, se hace uso de lo siguiente:

Tratamientos:

- Sumas de Cuadrados: $VE(t)$.
- Grados de Libertad: $K - 1$
- Medias de Cuadrados: $MCE(t) = VE(t) / K - 1$
- F_o : $F(t) = MCE(t)/MCNE$

Bloques:

- Sumas de Cuadrados: $VE(B)$.
- Grados de Libertad: $n - 1$
- Medias de Cuadrados: $MCNE(B) = VE(B) / n - 1$

- $F_o: F(B) = MCE(B)/MCNE$

Error:

- Sumas de Cuadrados: VNE .
- Grados de Libertad: $(K - 1)(n - 1)$
- Medias de Cuadrados: $MCNE = VNE / (K - 1)(n - 1)$

Total:

- Sumas de Cuadrados: $VT = VE(t) + VE(B) + VE$
- rados de Libertad: $N - 1$

De tal modo que la hipótesis nula se rechaza (a un nivel de confianza del $100(1 - \alpha)\%$ si

$$F_o > F_{\alpha, (K-1), (K-1)(N-1)}$$

Por otra parte el contraste de que los bloques no influyen, se realiza con las siguientes hipótesis:

$$H_o: B_1 = B_2 = B_3 \dots B_k$$

$$H_1: B_j \text{ distinto } B_j; \text{ para al menos un par } j \text{ distinto de } j$$

De tal modo que la hipótesis nula se rechaza (a un nivel de confianza del $100(1 - \alpha)\%$ si

$$F_B > F_{\alpha, (K-1), (K-1)(n-1)}$$

Y es un contraste independiente del anterior.

EJEMPLO 1.

Se probaran 5 raciones respecto a sus diferencias en el engorde de novillos. Se dispone de 20 novillos para el experimento, que se distribuyen en 4 bloques (5 novillos por bloque) con base a sus pesos, al iniciar la prueba de engorde, los novillos más pesados se agruparon en un bloque, en otro se agruparon los 5 siguientes más pesados y así sucesivamente. Los 5 tratamientos (raciones) se asignaron al azar dentro de cada bloque. Se obtuvieron los siguientes datos:

```
# Definiendo el vector que contendr\ 'a el bloque al cual pertenecen los novillos. # El p
bloques <- gl(n=4, k=1, length=20);bloques
## [1] 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
## Levels: 1 2 3 4
# k=1 especifica que no se repite el mismo factor m\ 'as de una vez consecutiva
# ;mientras que n=4 indica que el factor contendr\ 'a 4 niveles.
```

```

# Se crea el vector que contendr\ 'a los tratamientos de los novillos (raciones de
# alimento) los primeros cuatros se les asigna el tratamiento 1, los siguientes
# cuatro el 2, y as\ sucesivamente.

tratamientos <- gl(n=5, k=4);tratamientos

## [1] 1 1 1 1 2 2 2 2 3 3 3 3 4 4 4 4 5 5 5 5
## Levels: 1 2 3 4 5

# k=4 especifica que los primeros cuatro elementos ser\ 'an asignados al primer
# factor; los siguientes cuatro al segundo, y as\ 'i suceamente; mientras que n=5
# indica que el factor contendr\ 'a 5 niveles.

peso <- c(0.9, 1.4, 1.4, 2.3, 3.6, 3.2, 4.5, 4.1, 0.5, 0.9, 0.5, 0.9, 3.6, 3.6, 3.2, 3.6, 1.8, 1.8, 0.9, 1.4)
peso

## [1] 0.9 1.4 1.4 2.3 3.6 3.2 4.5 4.1 0.5 0.9 0.5 0.9 3.6 3.6 3.2 3.6 1.8
## [18] 1.8 0.9 1.4

datos2 <- data.frame(bloques = bloques, tratamientos = tratamientos, peso = peso)
datos2

##      bloques tratamientos peso
## 1          1             1 0.9
## 2          2             1 1.4
## 3          3             1 1.4
## 4          4             1 2.3
## 5          1             2 3.6
## 6          2             2 3.2
## 7          3             2 4.5
## 8          4             2 4.1
## 9          1             3 0.5
## 10         2             3 0.9
## 11         3             3 0.5
## 12         4             3 0.9
## 13         1             4 3.6
## 14         2             4 3.6
## 15         3             4 3.2
## 16         4             4 3.6
## 17         1             5 1.8
## 18         2             5 1.8
## 19         3             5 0.9
## 20         4             5 1.4

```

- Las hipótesis son las siguientes:

$$H_o: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

H_1 : μ_1 distinto μ_2 distinto μ_3 distinto μ_4 distinto μ_5

- Ejecutar el script `.anova2.R`

```
# Definiendo el vector que contendr\ 'a el bloque al cual pertenecen los
# novillos. El primer novillo es asignado al bloque 1, el siguiente al 2, el
# tercero al 3 el cuarto al 4, y se inicia el ciclo.

bloques <- gl(n=4, k=1, length=20);bloques

## [1] 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
## Levels: 1 2 3 4

# k=1 especifica que no se repite el mismo factor m\ 'as de una vez consecutiva
# ;mientras que n=4 indica que el factor contendr\ 'a 4 niveles.

# Se crea el vector que contendr\ 'a los tratamientos de los novillos (raciones
# de alimento) los primeros cuatros se les asigna el tratamiento 1, los
# siguientes cuatro el 2, y as\ 'i sucesivamente.

tratamientos <- gl(n=5, k=4);tratamientos

## [1] 1 1 1 1 2 2 2 2 3 3 3 3 4 4 4 4 5 5 5 5
## Levels: 1 2 3 4 5

# k=4 especifica que los primeros cuatro elementos ser\ 'an asignados al primer
# factor; los siguientes cuatro al segundo, y as\ 'i suceamente; mientras que
# n=5 indica que el factor contendr\ 'a 5 niveles.

peso <- c(0.9, 1.4, 1.4, 2.3, 3.6, 3.2, 4.5, 4.1, 0.5, 0.9, 0.5, 0.9, 3.6, 3.6,
          3.2, 3.6, 1.8, 1.8, 0.9, 1.4 )
peso

## [1] 0.9 1.4 1.4 2.3 3.6 3.2 4.5 4.1 0.5 0.9 0.5 0.9 3.6 3.6 3.2 3.6 1.8
## [18] 1.8 0.9 1.4

datos2 <- data.frame(bloques = bloques, tratamientos = tratamientos,
                     peso = peso)
datos2

##      bloques tratamientos peso
## 1          1             1 0.9
## 2          2             1 1.4
## 3          3             1 1.4
```

```
## 4      4      1  2.3
## 5      1      2  3.6
## 6      2      2  3.2
## 7      3      2  4.5
## 8      4      2  4.1
## 9      1      3  0.5
## 10     2      3  0.9
## 11     3      3  0.5
## 12     4      3  0.9
## 13     1      4  3.6
## 14     2      4  3.6
## 15     3      4  3.2
## 16     4      4  3.6
## 17     1      5  1.8
## 18     2      5  1.8
## 19     3      5  0.9
## 20     4      5  1.4

# Se aplica el an\`alisis de varianza

mod2 <- aov(peso ~ tratamientos + bloques, data = datos2)

# Observe con el signo + se indican c\`omo se descompone la varianza
# (tratamientos y bloques no hay interacci\`on entre ellos, es decir,
# son independientes).

# Se muestra la tabla ANOVA del experimento

summary(mod2)

##              Df Sum Sq Mean Sq F value    Pr(>F)
## tratamientos  4 30.712    7.678   39.107 8.6e-07 ***
## bloques        3  0.462    0.154    0.784  0.526
## Residuals     12  2.356    0.196
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Note que seg\`un los resultados, concluimos que si existen diferencia
# en el efecto de las raciones de comida en el engorde de los novillos.
# No as\`i con el efecto de los bloques.
```

COMENTARIOS FINALES

La eficacia del diseño por bloques depende de los efectos de los bloques; si éstos son muy pequeños, habremos ganado muy poco y, en el límite, si los bloques no influyesen, el contraste sería

menos eficaz que el diseño unifactorial, ya que la variabilidad no explicada tendrá menos grados de libertad $((k - 1)(n - 1))$ en lugar de los $k(n - 1)$ del diseño unifactorial). Sin embargo, cuando los bloques realmente influyen mucho, este diseño es enormemente superior al diseño unifactorial, pues será más sensible a percibir diferencias entre tratamientos. No podemos perder precisión si las variables no influyen, y podemos ganar mucho cuando si lo hacen.