UNIVERSIDAD DE EL SALVADOR
FACULTAD MULTIDISCIPLINARIA DE OCCIDENTE
DEPARTAMENTO DE MATEMÁTICA

Licenciatura en Estadística

Control Estadistico del Paquete R

”UNIDAD SEIS”
Práctica 26 - Diseños bifactoriales.

Alumna:
Martha Yoana Medina Sánchez

Fecha de elaboración
Santa Ana - 27 de noviembre de 2015

Los diseños bifactoriales se diferencian a los diseños por bloques en que ahora si nos interesa conocer el efecto entre este segundo factor(variable bloque) y nuestra variable dependiente; y además los dos factores ya no son independientes, por lo que es necesario conocer también el efecto de la interacción de ambos factores en nuestra variable dependiente.

Supondremos que tenemos un primer factor, A, con k niveles; mientras que tenemos un segundo factor, B, con n niveles. Supondremos que tomamos para cada combinación posible de los niveles de ambos factores un total de m observaciones. De lo contrario tendremos más parámetros a estimar en el modelo que el número de observaciones disponibles; y por consiguiente será imposible realizar el contraste.

El modelo que genera los datos es el siguiente:

$y_{\ ijl} = \mu + t_i + B_j +$ (tB)$_{ij}$ $+\ u_{\ ijl}$

Donde:

- $y_{\ ijl}$: Representa la l-ésima observación para la combinación ij-ésima de niveles de los factores A y B (celda ij-ésima de la matriz de datos).
- μ: Representa un promedio o efecto global.
- t_i: Representa el efecto del factor A cuando éste se encuentra en el i-ésimo nivel. Debe cumplirse que la sumatoria de $t_i = 0$.
- B_j: Representa el efecto del factor B cuando éste se encuentra en el j-ésimo nivel. Deben cumplir que la sumatoria de $B_j = 0$.
- (tB)$_{ij}$: Representa el efecto de la interacción de los factores A y B cuando el primero se encuentra en el nivel i y el segundo en el nivel j.
- $u_{\ ijl}$: Representa un componente de error aleatorio, llamado perturbaciones, que incorpora todas las demás fuentes de variabilidad del experimento.

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

El primer contraste de hipótesis a realizar es (efecto del factor A):

H_o: $t_1 = t_2$ ... $t_k = 0$

H_1: t_i distinto 0; para al menos un i

El segundo contraste de hipótesis a realizar es (efecto del factor B):

H_o: $B_1 = B_2$ ... $B_k = 0$

H_1: B_j distinto 0; para al menos un j

Mientras que el tercer contraste de hipótesis a realizar es (efecto de la interacción de los factores A y B):

H_o: $(\text{tB})_{ij} = 0$; para todo i,j

H_1: $(\text{tB})_{ij}$ disntinto 0 para al menos una combinación ij

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

$\hat{y}\ _{ij} = \hat{}\mu + \hat{t}_i + \hat{B}_j + \hat{(}\text{alfaB})\ _{ij}$
Y por consiguiente:

$y\ _{ijl} = \hat{}\mu + \hat{t}_i + \hat{B}_j + \hat{(}\text{alfaB})_{ij} + \hat{u}\ _{ijl}$

Donde:

- $\hat{}\mu = \tilde{y}_{..}$
- $\hat{t}_i = \tilde{y}_i..$ - $\tilde{y}_{..}$
- $\hat{B}_j = \tilde{y}_{.j.}$ - $\tilde{y}_{..}$
- $\hat{(}\text{alfaB})_{ij} = \tilde{y}_i j.$ - $\tilde{y}_i..$ - $\tilde{y}_{.j.}$ - $\tilde{y}_{..}$
- $\hat{u}\ _{ijl} = y\ _{ijl}$ - $\tilde{y}_i j.$

Y se tendrán las siguientes medidas de interés:

- $\tilde{y}_i..$ es el promedio para el i-ésimo nivel del factor A.
- $\tilde{y}_{.j.}$ es el promedio para el j-ésimo nivel del factor B.
- $\tilde{y}_i j.$ es el promedio para la combinación ij-ésima de los factores A y B.
- $\tilde{y}_{..}$ 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) + VE(tB) + VNE

Donde:

- VT es la variabilidad total del experimento.
- VE(t) es la variabilidad explicada por el factor A.
- VE(B) es la variabilidad explicada por el factor B.
- VE(tB) es la variabilidad explicada por la combinación de los factores A y B.

- VNE es la variabilidad no explicada o residual.

Para poder contrastar cada una de los diferentes contrastes anteriores, se utiliza la tabla ANOVA siguiente:

Factor A:

- Sumas de Cuadrados: VE(t).
- Grados de Libertad: $K - 1$
- Medias de Cuadrados: $MCE(\text{t}) = VE(\text{t})\ /\ K - 1$
- F_o: $F_{\ t} = MCE(\text{t})/MCNE$

Factor B:

- Sumas de Cuadrados: $VE(B)$.
- Grados de Libertad: $n - 1$
- Medias de Cuadrados: $MCNE(B) = VE(\text{t})\ /\ k - 1$
- F_o: $F_{\ B} = MCE(B)/MCNE$

Interacción:

- Sumas de Cuadrados: $VE(\text{tB})$.
- Grados de Libertad: $(K - 1)(n - 1)$
- Medias de Cuadrados: $MCNE = VNE(\text{tB})\ /\ (K - 1)(n - 1)$
- F_o: $F_{\ tB} = MCE(\text{tB})/MCNE$

Error:

- Sumas de Cuadrados: VNE.
- Grados de Libertad: $kn(m - 1)$
- Medias de Cuadrados: $MCNE = VNE\ /\ kn(m - 1)$

Total:

- Sumas de Cuadrados: $VT = VE(\text{t}) + VE(B) + VE(\text{tB}) + VNE$
- grados de Libertad: $N - 1$

De tal modo que la hipótesis nula de igualdad de los efectos del factor A se rechaza (a un nivel de confianza del 100(1 - $alfa$) % si

$F_t > F_{\ alfa,(K-1),kn(m-1)}$

De tal modo que la hipótesis nula de igualdad de los efectos del factor B se rechaza (a un nivel de confianza del 100(1 - $alfa$) % si

$F_B > F_{\ alfa,(n-1),kn(m-1)}$

De tal modo que la hipótesis nula de igualdad de los efectos del factor A y B se rechaza (a un nivel de confianza del 100(1 - $alfa$) % si

$F_{\ tB} > F_{\ alfa,(k-1)(n-1),kn(m-1)}$

EJEMPLO 1.

Se llevó a cabo un estudio del efecto de la temperatura sobre el porcentaje de encogimiento de telas teñidas, con dos réplicas para cada uno de cuatro tipos de tela en un diseño totalmente aleatorizado. Los datos son el porcentaje de encogimiento de dos réplicas de tela secadas a cuatro temperaturas; los cuales se muestran a continuación.

```
# Definiendo el vector que contendr\'a el factor A. 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.

FactorA <- gl(n=4, k=8, length=32);FactorA

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

# k=8 especifica que se introducir\'an las observaciones en orden de cada
# factor. Para mayor comodidad se introducir\'an los datos en orden de
# celda.

# 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.

FactorB<- gl(n=4, k=2,length=32);FactorB

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

# k=2 especifica que las primeras dos observaciones ser\'an correspondien
# tes al primer nivel, las dos siguientes al segundo, y as\'is
# sucesivamente. Cuando se finalice en el cuarto nivel se iniciar\'a
```

```r
# nuevamente en el nivel 1.

# Se digitan los pesos de los novillos

Porcentaje <- c(1.8, 2.1, 2.0, 2.1, 4.6, 5.0, 7.5, 7.9, 2.2, 2.4,
                4.2, 4.0, 5.4, 5.6, 9.8, 9.2, 2.8, 3.2, 4.4, 4.8,
                8.7, 8.4, 13.2, 13.0, 3.2, 3.6, 3.3, 3.5, 5.7, 5.8,
                10.9, 11.1)
Porcentaje

##  [1]  1.8  2.1  2.0  2.1  4.6  5.0  7.5  7.9  2.2  2.4  4.2  4.0  5.4  5.6
## [15]  9.8  9.2  2.8  3.2  4.4  4.8  8.7  8.4 13.2 13.0  3.2  3.6  3.3  3.5
## [29]  5.7  5.8 10.9 11.1

# Se registra en una hoja de datos los resultados del experimento

datos3 <- data.frame(FactorA = FactorA, FactorB = FactorB, Porcentaje=Porcentaje)
datos3

##    FactorA FactorB Porcentaje
## 1        1       1        1.8
## 2        1       1        2.1
## 3        1       2        2.0
## 4        1       2        2.1
## 5        1       3        4.6
## 6        1       3        5.0
## 7        1       4        7.5
## 8        1       4        7.9
## 9        2       1        2.2
## 10       2       1        2.4
## 11       2       2        4.2
## 12       2       2        4.0
## 13       2       3        5.4
## 14       2       3        5.6
## 15       2       4        9.8
## 16       2       4        9.2
## 17       3       1        2.8
## 18       3       1        3.2
## 19       3       2        4.4
## 20       3       2        4.8
## 21       3       3        8.7
## 22       3       3        8.4
## 23       3       4       13.2
## 24       3       4       13.0
## 25       4       1        3.2
## 26       4       1        3.6
## 27       4       2        3.3
```

```
## 28        4        2         3.5
## 29        4        3         5.7
## 30        4        3         5.8
## 31        4        4        10.9
## 32        4        4        11.1
```

Utilizando un nivel de significancia del 5 %, contraste el siguiente conjunto de hipótesis:

- Las hipótesis a contrastar para el factor A son:
  H_o: $t_1 = t_2 = t_3 = t_4$
  H_1: t_i distinto 0; para al menos un i
- Las hipótesis a contrastar para el factor B son:
  H_o: $B_1 = B_2 = B_3 = B_4$
  H_1: B_j distinto 0; para al menos un j
- Las hipótesis a contrastar para la interacción de los factores A y B son:
  H_o: (tB)$_{ij}$ = 0; para todo i,j
  H_1: (tB)$_{ij}$ disntinto 0 para al menos una combinación ij
- Ejecutar el script .ªnova3.R”

```
# Definiendo el vector que contendr\'a el factor A. 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.

FactorA <- gl(n=4, k=8, length=32);FactorA

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

# k=8 especifica que se introducir\'an las observaciones en orden de cada
# factor. Para mayor comodidad se introducir\'an los datos en orden de
# celda.

# 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.

FactorB<- gl(n=4, k=2,length=32);FactorB

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

# k=2 especifica que las primeras dos observaciones ser\'an correspondien
# tes al primer nivel, las dos siguientes al segundo, y as\'is
# sucesivamente. Cuando se finalice en el cuarto nivel se iniciar\'a
```

```
# nuevamente en el nivel 1.

# Se digitan los pesos de los novillos

Porcentaje <- c(1.8, 2.1, 2.0, 2.1, 4.6, 5.0, 7.5, 7.9, 2.2, 2.4,
                4.2, 4.0, 5.4, 5.6, 9.8, 9.2, 2.8, 3.2, 4.4, 4.8,
                8.7, 8.4, 13.2, 13.0, 3.2, 3.6, 3.3, 3.5, 5.7, 5.8,
                10.9, 11.1)
Porcentaje

##  [1]  1.8  2.1  2.0  2.1  4.6  5.0  7.5  7.9  2.2  2.4  4.2  4.0  5.4  5.6
## [15]  9.8  9.2  2.8  3.2  4.4  4.8  8.7  8.4 13.2 13.0  3.2  3.6  3.3  3.5
## [29]  5.7  5.8 10.9 11.1

# Se registra en una hoja de datos los resultados del experimento

datos3 <- data.frame(FactorA = FactorA, FactorB = FactorB, Porcentaje=Porcentaje)
datos3

##    FactorA FactorB Porcentaje
## 1        1       1        1.8
## 2        1       1        2.1
## 3        1       2        2.0
## 4        1       2        2.1
## 5        1       3        4.6
## 6        1       3        5.0
## 7        1       4        7.5
## 8        1       4        7.9
## 9        2       1        2.2
## 10       2       1        2.4
## 11       2       2        4.2
## 12       2       2        4.0
## 13       2       3        5.4
## 14       2       3        5.6
## 15       2       4        9.8
## 16       2       4        9.2
## 17       3       1        2.8
## 18       3       1        3.2
## 19       3       2        4.4
## 20       3       2        4.8
## 21       3       3        8.7
## 22       3       3        8.4
## 23       3       4       13.2
## 24       3       4       13.0
## 25       4       1        3.2
## 26       4       1        3.6
```

```
## 27       4       2        3.3
## 28       4       2        3.5
## 29       4       3        5.7
## 30       4       3        5.8
## 31       4       4       10.9
## 32       4       4       11.1
```

```
# Se aplica el anÃ¡lisis de varianza

mod3 <- aov(Porcentaje ~ FactorA * FactorB, data = datos3)

# Observe con el signo * se indican c\'omo se descompone la varianza; es
# decir, ser\'a el efecto del Factor A m\'as el efecto del Factor B m\'as
# el efecto de la interacci\'on de los factores A y B (* indica que tome
# en cuenta los efectos principales m\'as efecto de la interacci\'on).

# Se muestra la tabla ANOVA del experimento

summary(mod3)
```

```
##                 Df Sum Sq Mean Sq F value   Pr(>F)    
## FactorA          3  41.88   13.96  279.18 5.05e-14 ***
## FactorB          3 283.94   94.65 1892.91  < 2e-16 ***
## FactorA:FactorB  9  15.86    1.76   35.24 7.09e-09 ***
## Residuals       16   0.80    0.05                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Note que seg\'un los resultados, rechazamos cada una de las hip\'otesis.
```