**Ejercicio Propuesto 6 (Resuelto)**

**Consideremos de nuevo el ejercicio propuesto 5 del investigador que quiere evaluar la productividad de cuatro variedades de aguacate, A, B, C y D. Para ello, decide realizar el ensayo en un terreno que posee un gradiente de pendiente de oriente a occidente y además, diferencias en la disponibilidad de Nitrógeno de norte a sur. Se seleccionan cuatro disponibilidades de nitrógeno, pero sólo dispone de tres gradientes de pendiente. Para controlar estas posibles fuentes de variabilidad, el investigador decide utilizar un diseño en cuadrado de Youden con cuatro filas, las cuatro disponibilidades de Nitrógeno (NI, N2, N3, N4), tres columnas, los tres gradientes de pendientes (P1, P2, P3) y cuatro letras latinas, las variedades de aguacates (A, B, C, D). Los datos corresponden a la producción en kg/parcela.**

|  |  |
| --- | --- |
|  | **Pendiente** |
| **Nitrógeno** | **P1** | **P2** | **P3** |
| **N1** | **D****956** | **A****820** | **C****689** |
| **N2** | **A****867** | **B****975** | **D****680** |
| **N3** | **C****850** | **D****775** | **B****699** |
| **N4** | **B****950** | **C****870** | **A****980** |

**Responder a las siguientes cuestiones:**

1. **Estudiar cuál es el tipo de diseño adecuado a este experimento y escribir el modelo matemático asociado.**
2. **¿Se puede afirmar que la productividad media de las cuatro variedades de aguacate es la misma?**
3. **Comprobar la hipótesis de homocedasticidad**
4. **¿Se obtiene la misma producción con las cuatro variedades de aguacate? En caso negativo, analizar mediante el procedimiento de Duncan, con qué variedad de aguacate hay mayor producción.**

**Solución:**

**1. Estudiar cuál es el tipo de diseño adecuado a este experimento y escribir el modelo matemático asociado.**

El análisis de la productividad de las variedades de aguacate corresponde al análisis de un factor con 4 niveles. Dado que en el estudio intervienen dos fuentes de variación: la **Disponibilidad de Nitrógeno** y la **Pendiente**, se consideran dos factores de bloque, el primero con 4 niveles y el segundo con tres niveles.

Se pretende, entonces dar respuesta al contraste:

$$H\_{0}≡μ\_{A}=μ\_{B}=μ\_{C}=μ\_{D} $$

$$H\_{1}≡μ\_{i}\ne μ\_{j} para algún i \ne j$$

* **Variable respuesta**: ***Productividad.***
* **Factor**: ***Variedad de aguacate***. Es un factor de efectos fijos ya que desde el principio se establecen los niveles concretos que se van a analizar.
* **Bloques**: ***Disponibilidad de Nitrógeno*** y ***Pendiente***, con 4 y 3 niveles, respectivamente y ambos de efectos fijos.
* **Tamaño del experimento**: Número total de observaciones: **12**.

Este experimento se modeliza mediante un diseño en **Cuadrados de Jouden**.

Un cuadrado de Youden se puede considerar como un diseño en bloques incompletos balanceado y simétrico en el que las filas corresponden a los bloques. En efecto, si asignamos

* El factor principal a las letras latinas,
* Un factor secundario, el que tiene el mismo número de niveles que el factor principal, a las filas,
* un factor secundario, el que tiene menor número de niveles que el factor principal, a las columnas.

Entonces, un cuadrado de Youden es un diseño en bloques incompletos balanceado y simétrico en el que

* Cada tratamiento ocurre una vez en cada columna.
* La posición del tratamiento dentro de un bloque indica el nivel del factor secundario correspondiente a las columnas.
* El número de réplicas de un tratamiento dado es igual al número de tratamientos por bloque.

Recordamos que los parámetros que caracterizan este modelo son:

* *I, J* y *K* son el número de tratamientos, el número de bloques y el número de tratamientos por bloque, respectivamente.
* *R,* número de veces que cada tratamiento se presenta en el diseño, es decir el número de réplicas de un tratamiento dado.
* λ , número de bloques en los que un par de tratamientos ocurren juntos.
* *N*, número de observaciones.

Los valores de los parámetros del modelo en este ejemplo son:

 *N = I R = J K.* En efecto, ya que *N= 12; I = 4 = J; R = K = 3*

$$λ=R\frac{K-1}{I-1}=3\frac{3-1}{4-1}=2$$

El modelo matemático es:

$$y\_{ij k}=μ+τ\_{i }+β\_{j}+γ\_{k}+u\_{ijk} , i=1, …,4 ; j=1, ….,4; k=1,…, 3$$

**2. ¿Se puede afirmar que la productividad media de las cuatro variedades de aguacate es la misma?**

Productividad Nitrogeno Pendiente Variedad

756 N1 P1 D

720 N1 P2 A

689 N1 P3 C

596 N2 P1 A

855 N2 P2 B

780 N2 P3 D

750 N3 P1 C

975 N3 P2 D

899 N3 P3 B

950 N4 P1 B

870 N4 P2 C

880 N4 P3 A

Fichero: propuesto6.txt

> setwd("C:/Users/Usuario/Desktop/Datos")

> propuesto6<-read.table("propuesto6.txt", header = TRUE)

> propuesto6

 Productividad Nitrogeno Pendiente Variedad

1 756 N1 P1 D

2 720 N1 P2 A

3 689 N1 P3 C

4 596 N2 P1 A

5 855 N2 P2 B

6 780 N2 P3 D

7 750 N3 P1 C

8 975 N3 P2 D

9 899 N3 P3 B

10 950 N4 P1 B

11 870 N4 P2 C

12 880 N4 P3 A

A continuación debemos transformar tanto la columna de los tratamiento como la de los bloques en un factor para podemos realizar los cálculos posteriores adecuadamente.

> propuesto6$Nitrogeno <- factor(propuesto6$Nitrogeno)

> propuesto6$Nitrogeno

 [1] N1 N1 N1 N2 N2 N2 N3 N3 N3 N4 N4 N4

Levels: N1 N2 N3 N4

> propuesto6$Pendiente <- factor(propuesto6$Pendiente)

> propuesto6$Pendiente

 [1] P1 P2 P3 P1 P2 P3 P1 P2 P3 P1 P2 P3

Levels: P1 P2 P3

> propuesto6$Variedad <- factor(propuesto6$Variedad)

> propuesto6$Variedad

 [1] D A C A B D C D B B C A

Levels: A B C D

Para poder analizar los datos mediante un diseño BIB debemos instalar y cargar

los paquetes de R especializados en este tipo de diseños:

>library(daewr)

>library(AlgDesign)

Nota: Depende de la versión de R, también hay que cargar e instalar los paquetes

> library(colorspace)

> library(zoo)

La función “BIBsize(t , k)” de la librería daewr nos permite saber si el diseño puede realizarse. Calcula los parámetros del diseño donde

* t = número de niveles del factor tratamiento.
* k = número de tratamientos por bloque.

> BIBsize(t = 4 , k = 3)

Posible BIB design with b= 4 and r= 3 lambda= 2

**El análisis de este modelo lo podemos realizar en R de dos formas:**

**1. Realizaremos el análisis evaluando primero el efecto del tratamiento y después el de los bloques utilizando tres funciones**

Para cada factor realizamos una tabla ANOVA:

* **Factor principal: Variedad**

Para evaluar el efecto de los tratamientos, la suma de cuadrados de tratamientos debe ajustarse por bloques, por lo tanto primero se introducen los bloques y después los tratamientos.

Para calcular la tabla ANOVA hacemos uso de la función “**aov**” (asume suma de cuadrados tipo I) de la siguiente forma:

> mod1 <- aov(Productividad~ Pendiente + Nitrogeno + Variedad, data = propuesto6 )

donde:

* **Productividad**: Nombre de la columna de las observaciones.
* **Variedad:** Nombre de la columna en la que están representados los tratamientos.
* **Pendiente:** Nombre de la columna en la que está representado el primer factor bloque.
* **Nitrogeno:** Nombre de la columna en la que está representado el segundo factor bloque (letras latinas).
* **data** = data.frame en el que están guardados los datos.

> mod1

Call:

 aov(formula = Productividad ~ Pendiente + Nitrogeno + Variedad,

 data = propuesto6)

Terms:

 Pendiente Nitrogeno Variedad Residuals

Sum of Squares 16952.00 73454.00 47805.25 3012.75

Deg. of Freedom 2 3 3 3

Residual standard error: 31.6899

Estimated effects may be unbalanced

> summary(mod1)

 Df Sum Sq Mean Sq F value Pr(>F)

Pendiente 2 16952 8476 8.44 0.0586 .

Nitrogeno 3 73454 24485 24.38 0.0131 \*

Variedad 3 47805 15935 15.87 0.0241 \*

Residuals 3 3013 1004

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

El  p-valor, **0.0241,** es menor que el nivel de significación del 5%, deducimos que el **factor principal: Variedades del aguacate es significativo**.

* **Factor Bloque: Pendiente**

Para evaluar el efecto del primero de los bloques, la suma de cuadrados de bloques debe ajustarse por los tratamientos, por lo tanto primero se introducen los tratamientos y después los bloques:

> mod2 <- aov(Productividad~ Variedad + Nitrogeno + Pendiente, data = propuesto6 )

> mod2

Call:

 aov(formula = Productividad ~ Variedad + Nitrogeno + Pendiente,

 data = propuesto6)

Terms:

 Variedad Nitrogeno Pendiente Residuals

Sum of Squares 50344.67 70914.58 16952.00 3012.75

Deg. of Freedom 3 3 2 3

Residual standard error: 31.6899

Estimated effects may be unbalanced

> summary(mod2)

 Df Sum Sq Mean Sq F value Pr(>F)

Variedad 3 50345 16782 16.71 0.0224 \*

Nitrogeno 3 70915 23638 23.54 0.0138 \*

Pendiente 2 16952 8476 8.44 0.0586 .

Residuals 3 3013 1004

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

El  p-valor, **0.0586,** es mayor que el nivel de significación del 5%, deducimos que el **Factor Bloque: Pendiente no es significativo.**

* **Factor Bloque: Nitrogeno**

Para evaluar el efecto del segundo bloque, la suma de cuadrados de bloques debe ajustarse también por los tratamientos, por lo tanto primero se introducen los tratamientos y después los bloques:

> mod3 <- aov(Productividad~ Variedad + Pendiente + Nitrogeno, data = propuesto6 )

> mod3

Call:

 aov(formula = Productividad ~ Variedad + Pendiente + Nitrogeno,

 data = propuesto6)

Terms:

 Variedad Pendiente Nitrogeno Residuals

Sum of Squares 50344.67 16952.00 70914.58 3012.75

Deg. of Freedom 3 2 3 3

Residual standard error: 31.6899

Estimated effects may be unbalanced

> summary(mod3)

 Df Sum Sq Mean Sq F value Pr(>F)

Variedad 3 50345 16782 16.71 0.0224 \*

Pendiente 2 16952 8476 8.44 0.0586 .

Nitrogeno 3 70915 23638 23.54 0.0138 \*

Residuals 3 3013 1004

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

El p-valor es **0.0138**; menor que el nivel de significación del 5%, deducimos que **Factor Bloque: Nitrógeno es significativo**.

2. **Realizaremos el análisis evaluando tanto los tratamientos como los bloques ejecutando solo una** **función.**

Para ello necesitamos instalar y cargar el paquete “car”

> mod4 <- lm(Productividad~ Variedad + Nitrogeno + Pendiente, data = propuesto6 )

> mod4

Call:

lm(formula = Productividad ~ Variedad + Nitrogeno + Pendiente,

 data = propuesto6)

Coefficients:

(Intercept) VariedadB VariedadC VariedadD NitrogenoN2 NitrogenoN3

 634.38 133.13 -6.75 127.63 -24.63 108.62

NitrogenoN4 PendienteP2 PendienteP3

 176.50 92.00 49.00

> car::Anova(mod4, type="III")

Anova Table (Type III tests)

Response: Productividad

 Sum Sq Df F value Pr(>F)

(Intercept) 536576 1 534.3047 0.0001774 \*\*\*

Variedad 47805 3 15.8676 0.0240651 \*

Nitrogeno 70915 3 23.5382 0.0137944 \*

Pendiente 16952 2 8.4401 0.0586204 .

Residuals 3013 3

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Los resultados obtenidos coinciden con los realizados primero a los bloques y después al tratamiento.

**3. Comprobar la hipótesis de homocedasticidad**

> bartlett.test(propuesto6$Productividad, propuesto6$Variedad)

 Bartlett test of homogeneity of variances

data: propuesto6$Productividad and propuesto6$Variedad

Bartlett's K-squared = 1.8021, df = 3, p-value = 0.6145

> bartlett.test(propuesto6$Productividad, propuesto6$Pendiente)

 Bartlett test of homogeneity of variances

data: propuesto6$Productividad and propuesto6$Pendiente

Bartlett's K-squared = 0.49855, df = 2, p-value = 0.7794

> bartlett.test(propuesto6$Productividad, propuesto6$Nitrogeno)

 Bartlett test of homogeneity of variances

data: propuesto6$Productividad and propuesto6$Nitrogeno

Bartlett's K-squared = 3.8699, df = 3, p-value = 0.2759

Los p-valores del factor tratamiento, Variedad de aguacate (0.6145), del factor bloque Pendiente (0.779) y del factor bloque Nitrógeno (0.2759) son mayores que 0.05, por lo tanto no se puede rechazar la hipótesis de homogeneidad de la varianza del tratamiento y de los bloques.

**4. ¿Se obtiene la misma producción con las cuatro variedades de aguacate? En caso negativo, analizar mediante el procedimiento de Duncan, con qué variedad de aguacate hay mayor producción.**

Para realizar el contraste de comparaciones múltiples hay que cargar e instalar el paquete **agricolae**

> library(agricolae)

> (duncan=duncan.test(mod4, "Variedad" , group = T))

$statistics

 MSerror Df Mean CV

 1004.25 3 810 3.912334

$parameters

 test name.t ntr alpha

 Duncan Variedad 4 0.05

$duncan

 Table CriticalRange

2 4.500659 82.34484

3 4.515652 82.61915

4 4.472854 81.83611

$means

 Productividad std r Min Max Q25 Q50 Q75

A 732.0000 142.37977 3 596 880 658.0 720 800.0

B 901.3333 47.54296 3 855 950 877.0 899 924.5

C 769.6667 92.08873 3 689 870 719.5 750 810.0

D 837.0000 120.11245 3 756 975 768.0 780 877.5

$comparison

NULL

$groups

 Productividad groups

B 901.3333 a

D 837.0000 ab

C 769.6667 bc

A 732.0000 c

attr(,"class")

[1] "group"

En la tabla se muestran los subgrupos formados de medias iguales al utilizar el método de Duncan.  Hay tres subconjuntos que se diferencian entre sí. Por una parte el formado por la variedad de aguacate B y D, el subgrupo formado por D y C y el formado por A y C. También se observa que la mayor productividad de aguacate es la del tipo B, con una producción de 901.3333 Kg por parcela y la menor el tipo A, 732.0000 kg por parcela

**Editor de R**

library(colorspace)

library(zoo)

library(daewr)

library(AlgDesign)

BIBsize(t = 4 , k = 3)

mod1 <- aov(Productividad~ Pendiente + Nitrogeno + Variedad, data = propuesto6 )

mod1

summary(mod1)

mod2 <- aov(Productividad~ Variedad + Nitrogeno + Pendiente, data = propuesto6 )

mod2

summary(mod2)

mod3 <- aov(Productividad~ Variedad + Pendiente + Nitrogeno, data = propuesto6 )

mod3

summary(mod3)

mod4 <- lm(Productividad~ Variedad + Nitrogeno + Pendiente, data = propuesto6 )

mod4

car::Anova(mod4, type="III")

bartlett.test(propuesto6$Productividad, propuesto6$Variedad)

bartlett.test(propuesto6$Productividad, propuesto6$Pendiente)

bartlett.test(propuesto6$Productividad, propuesto6$Nitrogeno)

library(agricolae)

(duncan=duncan.test(mod4, "Variedad" , group = T))