**Ejercicio Propuesto 4 (Resuelto)**

**Consideremos de nuevo el ejercicio propuesto 3 sobre un grupo de científicos que estudia la calidad de varios tipos de césped para implantarlo en invierno en los campos de golf. Para ello, miden la distancia recorrida por una pelota de golf, en el campo, después de bajar por una rampa (para proporcionar a la pelota una velocidad inicial constante). El terreno del que disponen tiene mayor pendiente en la dirección norte-sur, por lo que se aconseja dividir el terreno en seis bloques de manera que las pendientes de las parcelas individuales dentro de cada bloque sean las mismas. Se utilizó el mismo método para la siembra y las mismas cantidades de semilla. Las mediciones son las distancias desde la base de la rampa al punto donde se pararon las pelotas, y al realizar dichas mediciones no se han podido obtener una para cada combinación de tipo de césped y tipo de terreno, sino que sólo se han podido realizar con tres de las variedades del césped en cada uno de los bloques de terreno. Para controlar el efecto del tipo de terreno deciden utilizar un diseño en bloques incompletos. En el estudio se incluyeron las variedades: Agrostis Tenuis (Césped muy fino y denso, de hojas cortas y larga duración), Agrostis Canina (Hoja muy fina, estolonífera. Forma una cubierta muy tupida), Paspalum Notatum (Hojas gruesas, bastas y con rizomas. Forma una cubierta poco densa), Paspalum Vaginatum (Césped fino, perenne, con rizomas y estolones).**

|  |  |
| --- | --- |
| **Variedad de césped** | **Bloques** |
| **Bloque1** | **Bloque2** | **Bloque3** | **Bloque4** | **Bloque5** | **Bloque6** |
| **Agrosty Tennis** | **1.30** |  | **0.50** |  | **1.80** |  |
| **Agrosty Canina** | **2.20** |  |  | **1.50** |  | **1.80** |
| **Paspalum Notatum** |  | **2.40** | **2.00** |  |  | **1.60** |
| **Paspalum Vaginatum** |  | **4.40** |  | **4.10** | **3.40** |  |

**Se pide:**

1. **Identificar los elementos del estudio (factores, unidades experimentales, variable respuesta, etc.) y plantear detalladamente el modelo matemático utilizado en el experimento.**
2. **¿Son los bloques fuente de variación?**
3. **Existen diferencias reales entre las distancias medias recorridas por una pelota de golf en los distintos tipos de césped?**
4. **Comprobar que se cumplen las hipótesis del modelo.**
5. **Utilizando el método de Newman-Keuls, ¿qué tipo de cesped ofrece menor resistencia al recorrido de las pelotas?**

**Solución:**

**1. Identificar los elementos del estudio (factores, unidades experimentales, variable respuesta, etc.) y plantear detalladamente el modelo matemático utilizado en el experimento.**

* **Variable respuesta: *Distancia.***
* **Factor: *Tipo\_Cesped*** que tiene cuatro niveles. Es un factor de efectos fijos ya que viene decidido qué niveles concretos se van a utilizar.
* **Bloque:** ***Bloques*** que tiene seis niveles. Es un factor de efectos fijos ya que viene decidido qué niveles concretos se van a utilizar.
* **Modelo incompleto**: Todos los tratamientos no se prueban en cada bloque.
* **Tamaño del experimento:** Número total de observaciones (12).

Denotemos:

* *T, B* y *K* son el número de tratamientos, el número de bloques y el número de tratamientos por bloque, respectivamente. *T = 4; B = 6; K = 2*
* *R,* número de veces que cada tratamiento se presenta en el diseño, es decir el número de réplicas de un tratamiento dado. (R = 3)
* λ , el número de veces que cada par de tratamientos ocurre en el mismo bloque. (λ=1)
* *N*, número de observaciones. (N = 12)

En el diseño se verifica que:

* Cada tratamiento ocurre el mismo número de veces en el diseño (3 veces)
* Cada par de tratamientos ocurren juntos el mismo número de veces que cualquier otro par (1 vez)
* Estos parámetros deben verificar las siguientes relaciones:

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

donde ***B ≥ T***   y   ***N = T R = B K; 12=4X3=6X2***

Es un diseño de bloques incompletos balanceado *(BIB*

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

> propuesto4<-read.table("propuesto4.txt", header = TRUE)

> propuesto4

 distancia cesped bloque

1 1.3 C1 B1

2 2.2 C2 B1

3 2.4 C3 B2

4 4.4 C4 B2

5 0.5 C1 B3

6 2.0 C3 B3

7 1.5 C2 B4

8 4.1 C4 B4

9 1.8 C1 B5

10 3.4 C4 B5

11 1.8 C2 B6

12 1.6 C3 B6

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

> propuesto4$Tratamiento = factor(propuesto4$cesped)

> propuesto4$Tratamiento

 [1] C1 C2 C3 C4 C1 C3 C2 C4 C1 C4 C2 C3

Levels: C1 C2 C3 C4

>propuesto4$Bloque = factor(propuesto4$bloque)

>propuesto4$Bloque

 [1] B1 B1 B2 B2 B3 B3 B4 B4 B5 B5 B6 B6

Levels: B1 B2 B3 B4 B5 B6

**2. ¿Son los bloques fuente de variación?**

Para poder analizar los datos mediante un diseño BIB debemos instalar y cargar dos paquetes de *R* especializados en este tipo de diseños:

> library(daewr)

Al instalar el paquete daewr, puede dar el siguiente mensaje:

local({pkg <- select.list(sort(.packages(all.available = TRUE)),graphics=TRUE)

+ if(nchar(pkg)) library(pkg, character.only=TRUE)})
Error: package or namespace load failed for ‘daewr’ in
loadNamespace(i, c(lib.loc, .libPaths()), versionCheck = vI[[i]]):
 there is no package called ‘colorspace’

Por lo que tenemos también que cargar e instalar el paquete colorspace

> library(colorspace)

Error: package or namespace load failed for ‘daewr’ in loadNamespace(i, c(lib.loc, .libPaths()), versionCheck = vI[[i]]):

 there is no package called ‘zoo’

Cargamos e instalamos el paquete zoo

> library(zoo)

> library(daewr)

> library(AlgDesign)

 Hemos cargado e instalado cuatro paquetes:

> library(colorspace)

> library(zoo)

> library(daewr)

> library(AlgDesign)

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.

Ejecutamos:

> BIBsize(t = 4 , k = 2)

Posible BIB design with b= 6 and r= 3 lambda= 1.

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

**A) Realizaremos el análisis: evaluando primero el efecto de los bloques y después el de los tratamientos utilizando dos funciones**

* Para evaluar el efecto 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:
* Para calcular la tabla ANOVA hacemos uso de la función “**aov**” (**aov(y ~ A + B, data=mydataframe**) asume suma de cuadrados tipo I) de la siguiente forma:

> mod1 <- aov(distancia ~ cesped + bloque, data = propuesto4 )

> mod1

Call:

 aov(formula = distancia ~ cesped + bloque, data = propuesto4)

Terms:

 cesped bloque Residuals

Sum of Squares 12.856667 0.890833 1.062500

Deg. of Freedom 3 5 3

Residual standard error: 0.595119

Estimated effects may be unbalanced

> summary(mod1)

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

cesped 3 12.857 4.286 12.100 0.035 \*

bloque 5 0.891 0.178 0.503 0.766

Residuals 3 1.062 0.354

---

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

El p-valor 0.035 menor que el nivel de significación 0.05, por lo que se rechaza la hipótesis nula de igualdad de bloques y concluimos que los bloques son una fuente de variación.

**3. Existen diferencias reales entre las distancias medias recorridas por una pelota de golf en los distintos tipos de césped?**

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.

> mod2 <- aov(distancia ~ bloque + cesped, data = propuesto4 )

donde:

* distancia = nombre de la columna de las observaciones
* ceped = nombre de la columna en la que están representados los tratamientos
* bloque = nombre de la columna en la que están representados los bloques
* data = data.frame en el que están guardados los datos

> mod2

Call:

 aov(formula = distancia ~ bloque + cesped, data = propuesto4)

Terms:

 bloque cesped Residuals

Sum of Squares 6.6000 7.1475 1.0625

Deg. of Freedom 5 3 3

Residual standard error: 0.595119

Estimated effects may be unbalanced

> summary(mod2)

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

bloque 5 6.600 1.3200 3.727 0.1540

cesped 3 7.147 2.3825 6.727 0.0759 .

Residuals 3 1.062 0.3542

---

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

El valor del estadístico de contraste de igualdad de ***tipo de césped***, ***F* = 3.727,** deja a su derecha un **p-valor 0.1540**, mayor que el nivel de significación del 5%, por lo que no se rechaza la Hipótesis Nula de igualdad de tratamientos. Por lo tanto no hay diferencias reales entre las distancias medias recorridas por una pelota de golf en los distintos tipos de césped ya que el p-valor es mayor que 0.05.

###### B) **Realizaremos el análisis evaluando tanto los tratamientos como los bloques ejecutando solo una** **función.**

Para ello necesitamos instalar y cargar el paquete “car”:

Una vez  instalado cargado el paquete realizamos el ANOVA

> mod3 <- lm(distancia ~ cesped + bloque, data = propuesto4 )
> mod3

Call:

lm(formula = distancia ~ cesped + bloque, data = propuesto4)

Coefficients:

(Intercept) cespedC2 cespedC3 cespedC4 bloqueB2 bloqueB3

 1.4375 0.6250 0.8250 2.5500 0.2750 -0.6000

 bloqueB4 bloqueB5 bloqueB6

 -0.2250 -0.1125 -0.4625

> car::Anova(mod3, type="III")

Anova Table (Type III tests)

Response: distancia

 Sum Sq Df F value Pr(>F)

(Intercept) 2.7552 1 7.7794 0.06847 .

cesped 7.1475 3 6.7271 0.07589 .

bloque 0.8908 5 0.5031 0.76563

Residuals 1.0625 3

---

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

Los resultados obtenidos coinciden con los realizados primero a los bloque y después a los tratamientos.

**4. Comprobar que se cumplen las hipótesis del modelo.**

Los supuestos que han de verificarse son Normalidad, Homocedasticidad e Independencia además del supuesto de aditividad entre los tratamientos y los bloques

**Hipótesis de normalidad**

Comprobamos la hipótesis de normalidad mediante el análisis de la normalidad de los residuos. Para ello, hacemos uso del test de Shapiro-Wilks:

> shapiro.test(mod3$residuals)

 Shapiro-Wilk normality test

data: mod3$residuals

W = 0.94151, p-value = 0.5178

El p-valor, 0.5178, es mayor que el nivel de significación del 5%, aceptándose la hipótesis de normalidad.

**Homogeneidad de varianzas**

En este caso hacemos uso del Test de Barlett para contrastar la igualdad entre varianzas del factor.

> bartlett.test(propuesto4$distancia, propuesto4$cesped)

 Bartlett test of homogeneity of variances

data: propuesto4$distancia and propuesto4$cesped

Bartlett's K-squared = 0.76822, df = 3, p-value = 0.8571

El p-valor es **0.8571** por lo tanto no se puede rechazar la hipótesis de homogeneidad de las varianzas y se concluye que la cuatro variedades de césped tienen varianzas homogéneas.

> bartlett.test(propuesto4$distancia, propuesto4$bloque)

 Bartlett test of homogeneity of variances

data: propuesto4$distancia and propuesto4$bloque

Bartlett's K-squared = 3.1837, df = 5, p-value = 0.6717

El p-valor es mayor que 0.05, por lo tanto no se puede rechazar la hipótesis de homogeneidad de las varianzas los bloques.

**Independencia de los residuos**

Para ello tenemos que representar un gráfico de los residuos tipificados frente a los pronosticados. En R obtenemos varios gráficos a la vez que están incluidos en la estimación del modelo.

Para verlos de forma correcta hacemos uso de las siguientes órdenes:

> layout(matrix(c(1,2,3,4),2,2)) # para que salgan en la misma pantalla

> plot(mod3)



En esta salida se muestran cuatro gráficas, interpretamos el gráfico que se muestra en la Fila 1, Columna 1. Es decir, el gráfico el que se representan los residuos en el eje de ordenadas y los valores ajustados por el modelo en el eje de abscisas. Este gráfico no muestra ningún aspecto que haga sospechar de la hipótesis de independencia de los residuos.

**5. Utilizando el método de Newman-Keuls, ¿qué tipo de cesped ofrece menor resistencia al recorrido de las pelotas?**

Cargar e instalar el paquete agricolae

> library(agricolae)

> Newman\_Keuls <- SNK.test(mod3,"cesped", console=TRUE, main="

Contraste de Newman-Keuls para el factor tipo de cesped")

Study: Contraste de Newman-Keuls para el factor tipo de cesped

Student Newman Keuls Test

for distancia

Mean Square Error: 0.3541667

cesped, means

 distancia std r Min Max

C1 1.200000 0.6557439 3 0.5 1.8

C2 1.833333 0.3511885 3 1.5 2.2

C3 2.000000 0.4000000 3 1.6 2.4

C4 3.966667 0.5131601 3 3.4 4.4

Alpha: 0.05 ; DF Error: 3

Critical Range

 2 3 4

1.546391 2.030514 2.344854

Means with the same letter are not significantly different.

 distancia groups

C4 3.966667 a

C3 2.000000 b

C2 1.833333 b

C1 1.200000 b

En la  tabla  se muestran los subgrupos formados de medias iguales al utilizar el método de Newman-Keuls.  Hay dos subconjuntos que se diferencian  entre sí. Por una parte el formado por el tipo de césped C4 y por otra parte el subgrupo formado por los tipos de césped: C3, C2 y C1. También se observa que el tipo de césped donde la distancia recorrida es más grande es el C4 con una distancia de 3.96667 y la distancia más corta es de 1.2 en el césped C1

**Editor de R**

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

propuesto4<-read.table("propuesto4.txt", header = TRUE)

propuesto4

propuesto4$Tratamiento = factor(propuesto4$cesped)

propuesto4$Tratamiento

propuesto4$Bloque = factor(propuesto4$bloque)

propuesto4$Bloque

library(colorspace)

library(zoo)

library(daewr)

library(AlgDesign)

BIBsize(t = 4 , k = 2)

mod1 <- aov(distancia ~ cesped + bloque, data = propuesto4 )

mod1

summary(mod1)

mod2 <- aov(distancia ~ bloque + cesped, data = propuesto4 )

mod2

summary(mod2)

mod3 <- lm(distancia ~ cesped + bloque, data = propuesto4 )

mod3

car::Anova(mod3, type="III")

shapiro.test(mod3$residuals)

bartlett.test(propuesto4$distancia, propuesto4$cesped)

bartlett.test(propuesto4$distancia, propuesto4$bloque)

layout(matrix(c(1,2,3,4),2,2)) # para que salgan en la misma pantalla

plot(mod3)

library(agricolae)

Newman\_Keuls <- SNK.test(mod3,"cesped", console=TRUE, main=" Contraste de Newman-Keuls para el factor tipo de cesped")