**Ejercicio Propuesto 3 (Resuelto)**

**Debido a la proliferación de los campos de golf y a la gran cantidad de agua que necesitan, un grupo de científicos 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 cinco 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. 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** |
| **Bloque 1** | **Bloque 2** | **Bloque 3** | **Bloque 4** | **Bloque 5** |
| **Agrosty Tennis** | **1.30** | **1.60** | **0.50** | **1.20** | **1.10** |
| **Agrosty Canina** | **2.20** | **2.40** | **0.40** | **2.00** | **1.80** |
| **Paspalum Notatum** | **1.80** | **1.70** | **0.60** | **1.50** | **1.30** |
| **Paspalum Vaginatum** | **3.90** | **4.40** | **2.00** | **4.10** | **3.40** |

Figura 1: Tabla de datos del Ejercicio Propuesto 3.doc

**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. **Estudiar las interacciones de los factores.**
5. **Comprobar que se cumplen las hipótesis del modelo.**
6. **Utilizando el método de Duncan y Newman-Keuls, ¿qué tipo de césped 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\_Césped*** 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 cinco niveles. Es un factor de efectos fijos ya que viene decidido qué niveles concretos se van a utilizar
* **Modelo completo**: Los cuatro tratamientos se prueban en cada bloque exactamente una vez
* **Tamaño del experimento:** Número total de observaciones (20).

Este experimento se modeliza mediante un diseño en **Bloques completos al azar**. El modelo matemático es:

$$y\_{ij }=μ+τ\_{i }+β\_{j}+u\_{ij} , i=1, …,4 ; j=1, ….,5$$

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

Para realizar este supuesto en R debemos introducir primero los datos de forma correcta. Podemos introducir los datos directamente en R de forma manual o introducirlos previamente en un archivo de texto o Excel y leerlos en R.

En este caso lo hacemos en un archivo de texto:

distancia cesped bloque

1.30 C1 B1

1.60 C1 B2

0.50 C1 B3

1.20 C1 B4

1.10 C1 B5

2.20 C2 B1

2.40 C2 B2

0.40 C2 B3

2.00 C2 B4

1.80 C2 B5

1.80 C3 B1

1.70 C3 B2

0.60 C3 B3

1.50 C3 B4

1.30 C3 B5

3.90 C4 B1

4.40 C4 B2

2.00 C4 B3

4.10 C4 B4

3.40 C4 B5

**[Figura 2: Tabla de datos del Ejercicio Propuesto 3.](https://wpd.ugr.es/~bioestad/wp-content/uploads/propuesto3.txt)**[txt](https://wpd.ugr.es/~bioestad/wp-content/uploads/propuesto3.txt)

Para cargar los datos utilizamos la función **read.table** indicando el nombre del archivo (que debe de estar en el directorio de trabajo) e indicando además que tiene cabecera.

***Nota****: La ruta hasta llegar al fichero varía en función del ordenador. Utilizar la  orden****setwd()*** *para situarse en el directorio de trabajo*

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

> propuesto3<- read.table("propuesto3.txt", header = TRUE)

> propuesto3

 distancia cesped bloque

1 1.3 C1 B1

2 1.6 C1 B2

3 0.5 C1 B3

4 1.2 C1 B4

5 1.1 C1 B5

6 2.2 C2 B1

7 2.4 C2 B2

8 0.4 C2 B3

9 2.0 C2 B4

10 1.8 C2 B5

11 1.8 C3 B1

12 1.7 C3 B2

13 0.6 C3 B3

14 1.5 C3 B4

15 1.3 C3 B5

16 3.9 C4 B1

17 4.4 C4 B2

18 2.0 C4 B3

19 4.1 C4 B4

20 3.4 C4 B5

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.

> propuesto3$cesped= factor(propuesto3$cesped)

> propuesto3$cesped

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

Levels: c1 C1 C2 C3 C4

> propuesto3$bloque= factor(propuesto3$bloque)

> propuesto3$bloque

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

Levels: B1 B2 B3 B4 B5

Para calcular la tabla ANOVA primero hacemos uso de la función “**aov**” de la siguiente forma

> mod = aov(distancia ~ cesped + bloque, data = propuesto3)

donde:

* **distancia:**Nombre de la columna de las observaciones
* **cesped:** 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

Ejecutamos

> mod

Call:

 aov(formula = distancia ~ cesped + bloque, data = propuesto3)

Terms:

 cesped bloque Residuals

Sum of Squares 18.044 6.693 0.951

Deg. of Freedom 3 4 12

Residual standard error: 0.2815138

Estimated effects may be unbalanced

y a continuacion mostramos un resumen de los resultados con la función “**summary**” (verdadera tabla ANOVA):

> summary(mod)

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

cesped 3 18.044 6.015 75.89 4.52e-08 \*\*\*

bloque 4 6.693 1.673 21.11 2.32e-05 \*\*\*

Residuals 12 0.951 0.079

---

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

En la *Tabla ANOVA,* el valor del estadístico de contraste de igualdad de medias de tratamientos, **F = 75.89** deja a su derecha un p-valor igual a **4.52e-08**, menor que el nivel de significación del 5%, por lo que se rechaza la Hipótesis nula de igualdad de tratamientos. Así, los tipos de césped influyen en las distancias recorridas por las pelotas. Es decir, existen diferencias significativas en las distancias recorridas por las pelotas entre los cuatro tipos de césped..

En esta *Tabla ANOVA,* también se observa que el valor del estadístico de contraste de igualdad de medias de bloques, **F = 21.11** deja a su derecha un p-valor igual a **2.32e-05**, menor que el nivel de significación del 5%, por lo que se rechaza la Hipótesis nula de igualdad de bloques. La eficacia de este diseño depende de los efectos de los bloques. En este caso este diseño es más eficaz que el diseño completamente aleatorizado y el contraste principal de las medias de los tratamientos será más sensible a las diferencias entre tratamientos. Por lo tanto la inclusión del factor bloque en el modelo es acertada. Así, las distancias recorridas por las pelotas dependen del tipo de terreno.

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

Esta cuestión está contestada afirmativamente en el apartado anterior, en el que la tabla ANOVA nos muestra un valor de **F = 75.89** y un **Sig. menor que 0.00**1

4. **Estudiar las interacciones de los factores.**

La interacción entre el factor bloque y los tratamientos vamos a estudiarla analíticamente mediante el  Test de Interacción de un grado de [**Tukey**](https://es.wikipedia.org/wiki/John_W._Tukey)

Para realizar este test en ***R*** tenemos que utilizar la library “**daewr**” y dentro de ella la función “**Tukey1df**”. De la siguiente forma:

Primero hay que instalar el paquete **daewr**

Para ello, seleccionar **Paquetes/Instalar paquetes**y de la lista escoger **daewr**. O bien utilizar la siguiente orden

> utils:::menuInstallPkgs()

Para realizar este contraste hay que utilizar la libray ***daewr,*** para ello realizamos la siguiente orden

> library(daewr)

> Tukey1df(propuesto3)

Source df SS MS F Pr>F

A 3 18.044 6.0147

B 4 6.693 1.6732

Error 12 0.951 1.268

NonAdditivity 1 0.6155 0.6155 20.18 9e-04

Residual 11 0.3355 0.0305

Puesto que el p-valor (Pr > F) es 9e-04 rechazamos la hipótesis nula de no interacción.

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

#### **Hipótesis de Normalidad**

La normalidad las vamos a comprobar analíticamente y gráficamente.

Analíticamente mediante el  contraste de Shapiro-Wilk que es adecuado cuando las muestras son pequeñas (n<50)

> shapiro.test(mod$residuals)

 Shapiro-Wilk normality test

data: mod$residuals

W = 0.98077, p-value = 0.9437

Como podemos observar tenemos un p-valor de 0.9437 que aceptaría la hipótesis de normalidad por ser mayor al 5% (nivel de significación usual).

Gráficamente mediante el gráfico probabilístico normal. Para ello utilizamos la orden “*qqnorm*”

> qqnorm(mod$residuals)



Figura 1: Estudio gráfico de la hipótesis de normalidad mediante el gráfico Q-Q Plot

La figura 1 se representan cuantiles muestrales frente a los cuantiles teóricos. En esta gráfica observamos que podemos considerar que los puntos se encuentran sobre la diagonal principal por lo tanto podemos deducir gráficamente la normalidad de los datos.

#### **Hipótesis de Homogeneidad de Varianzas**

Para comprobar la hipótesis de homocedasticidad utilizamos el Test de Barlett distinguiendo entre la igualdad entre varianzas del factor principal y la igualdad de varianzas del factor bloque.

En nuestro ejemplo, el test para igualdad de varianzas del factor principal sería:

> bartlett.test(propuesto3$distancia, propuesto3$cesped)

Bartlett test of homogeneity of variances

data: propuesto3$distancia and propuesto3$cesped

Bartlett's K-squared = 3.3347, df = 3, p-value = 0.3428

El p-valor es del 0.3428 que al ser mayor del nivel significación usual del 5% no podemos rechazar la hipótesis de igualdad de varianzas en el factor principal.

De la misma manera procedemos para el factor bloque:

> bartlett.test(propuesto3$distancia, propuesto3$bloque)

 Bartlett test of homogeneity of variances

data: propuesto3$distancia and propuesto3$bloque

Bartlett's K-squared = 0.94442, df = 4, p-value = 0.9181

El p-valor es 0.9181, mayor que 0.05 por lo que no podemos rechazar la hipótesis de igualdad de varianzas en el factor bloque.

#### **Hipótesis de Independencia**

Comprobaremos si se satisface el supuesto de independencia entre 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))
> plot(mod)



Figura 2: Estudio gráfico de la hipótesis de independencia de los residuos

La figura 2 está formada por cuatro gráficos, nos fijamos en el primer gráfico que representa los  valores ajustados frente a los residuos y observamos que no hay ninguna tendencia sistemática. Concluimos que no hay sospechas para que se incumpla la hipótesis de independencia.

6. **Utilizando el método de Duncan y Newman-Keuls, ¿qué tipo de césped ofrece menor resistencia al recorrido de las pelotas?**

Hemos probado anteriormente que se rechaza la Hipótesis nula de igualdad de tratamientos. Así, los tratamientos influyen en el número de semillas. Es decir, existen diferencias significativas en el número de semillas entre los tres tratamientos. Para saber entre que parejas de días estas diferencias son significativas aplicamos una prueba **Post-hoc.**

El contraste de Comparaciones múltiples que vamos a utilizar es el Test de Duncan. Para poder hacer uso de él en R tenemos que instalar en primer lugar el paquete “**agricolae**” y dentro de él la función “duncan.test”.

> library(agricolae)

Destacar que este test hace las comparaciones especificándole si es para el factor principal o el factor bloque. En este caso nos piden que las realicemos para los tratamientos

> (duncan=duncan.test(mod, "cesped" , group = T))

$`statistics`

 MSerror Df Mean CV

 0.07925 12 1.96 14.36295

$parameters

 test name.t ntr alpha

 Duncan cesped 4 0.05

$duncan

 Table CriticalRange

2 3.081307 0.3879266

3 3.225244 0.4060478

4 3.312453 0.4170272

$means

 distancia std r Min Max Q25 Q50 Q75

C1 1.14 0.4037326 5 0.5 1.6 1.1 1.2 1.3

C2 1.76 0.7924645 5 0.4 2.4 1.8 2.0 2.2

C3 1.38 0.4764452 5 0.6 1.8 1.3 1.5 1.7

C4 3.56 0.9449868 5 2.0 4.4 3.4 3.9 4.1

$comparison

NULL

$groups

 distancia groups

C4 3.56 a

C2 1.76 b

C3 1.38 bc

C1 1.14 c

attr(,"class")

[1] "group"

En el apartado “$groups” concluimos que:

A partir de los resultados obtenidos, se deduce que las distancias medias recorridas por las pelotas son similares para los céspedes **Agrostis Tenuis (C1)** y **Paspalum Notatum (C3)** por una parte, también son similares en el **Paspalum Notatum (C3)** y **Agrostis Canina (C2)**, y en ambos grupos dichas distancias medias difieren significativamente de las recorridas en el césped **Paspalum Vaginatum (C4).** Por lo tanto, se pueden establecer tres agrupaciones con características similares para las distancias medias recorridas. El tipo de césped que ofrece menor resistencia al recorrido de las pelotas es el **Paspalum Vaginatum (C4)**, donde las pelotas tienen un recorrido medio de **3.56 u.d**.

**Método de Newman-Keuls**

SNK.test(mod, "cesped", console =TRUE)

Study: mod ~ "cesped"

Student Newman Keuls Test

for distancia

Mean Square Error: 0.07925

cesped, means

 distancia std r Min Max

C1 1.14 0.4037326 5 0.5 1.6

C2 1.76 0.7924645 5 0.4 2.4

C3 1.38 0.4764452 5 0.6 1.8

C4 3.56 0.9449868 5 2.0 4.4

Alpha: 0.05 ; DF Error: 12

Critical Range

 2 3 4

0.3879266 0.4749996 0.5285978

Means with the same letter are not significantly different.

 distancia groups

C4 3.56 a

C2 1.76 b

C3 1.38 bc

C1 1.14 c

propuesto3<- read.table("propuesto3.txt", header = TRUE)

propuesto3

propuesto3$cesped= factor(propuesto3$cesped)

propuesto3$bloque= factor(propuesto3$bloque)

mod = aov(distancia ~ cesped + bloque, data = propuesto3)

mod

summary(mod)

library(daewr)

Tukey1df(propuesto3)

shapiro.test(mod$residuals)

qqnorm(mod$residuals)

bartlett.test(propuesto3$distancia, propuesto3$cesped)

bartlett.test(propuesto3$distancia, propuesto3$bloque)

layout(matrix(c(1,2,3,4),2,2))

plot(mod)

library(agricolae)

(duncan=duncan.test(mod, "cesped" , group = T))

SNK.test(mod, "cesped", console =TRUE)