**Ejercicio Propuesto 5 (Resuelto)**

**Un investigador quiere evaluar la productividad de cuatro variedades de aguacates, 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,** **para controlar los efectos de la pendiente y la disponibilidad de Nitrógeno, utilizó un diseño de cuadrado latino, los datos corresponden a la producción en kg/parcela.**



**Responder a las siguientes cuestiones:**

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

**2. ¿Qué supuestos han de verificarse?**

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

**Solución:**

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

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, cada uno de ellos con 4 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***, ambos con 4 niveles y ambos de efectos fijos.
* **Tamaño del experimento**: Número total de observaciones (42).

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

productividad aguacate nitrogeno pendiente

785 D N1 P1

730 A N1 P2

700 C N1 P3

595 B N1 P4

855 A N2 P1

775 B N2 P2

760 D N2 P3

710 C N2 P4

950 C N3 P1

885 D N3 P2

795 B N3 P3

780 A N3 P4

945 B N4 P1

950 C N4 P2

880 A N4 P3

835 D N4 P4

Nos situamos en el directorio donde están los datos:

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

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.

> propuesto5<-read.table("propuesto5.txt", header = TRUE)

> propuesto5

 productividad aguacate nitrogeno pendiente

1 785 D N1 P1

2 730 A N1 P2

3 700 C N1 P3

4 595 B N1 P4

5 855 A N2 P1

6 775 B N2 P2

7 760 D N2 P3

8 710 C N2 P4

9 950 C N3 P1

10 885 D N3 P2

11 795 B N3 P3

12 780 A N3 P4

13 945 B N4 P1

14 950 C N4 P2

15 880 A N4 P3

16 835 D N4 P4

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.

> propuesto5$latina = factor(propuesto5$aguacate)

> propuesto5$latina

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

Levels: A B C D

> propuesto5$Bloque1 = factor(propuesto5$nitrogeno)

> propuesto5$Bloque1

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

Levels: N1 N2 N3 N4

> propuesto5$Bloque2 = factor(propuesto5$pendiente)

> propuesto5$Bloque2

 [1] P1 P2 P3 P4 P1 P2 P3 P4 P1 P2 P3 P4 P1 P2 P3 P4

Levels: P1 P2 P3 P4

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

> mod <- aov(productividad~ aguacate + nitrogeno + pendiente, data = propuesto5 )

donde:

* **productividad**: Nombre de la columna de las observaciones
* **aguacate** :  Nombre de la columna en la que están representados los tratamientos
* **nitrogeno** : Nombre de la columna en la que está representado el primer factor bloque
* **pendiente:** 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

> mod <- aov(productividad~ aguacate + nitrogeno + pendiente, data = propuesto5 )

> mod

Call:

 aov(formula = productividad ~ aguacate + nitrogeno + pendiente,

 data = propuesto5)

Terms:

 aguacate nitrogeno pendiente Residuals

Sum of Squares 5556.25 92518.75 52556.25 112.50

Deg. of Freedom 3 3 3 6

Residual standard error: 4.330127

Estimated effects may be unbalanced

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

> summary(mod)

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

aguacate 3 5556 1852 98.78 1.70e-05 \*\*\*

nitrogeno 3 92519 30840 1644.78 3.92e-09 \*\*\*

pendiente 3 52556 17519 934.33 2.13e-08 \*\*\*

Residuals 6 113 19

---

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

Observando los valores de los p-valores**,** 1.70e-05, 3.92e-09 **y** 2.13e-08**; menores respectivamente que el nivel de significación del 5%, deducimos que los tres efectos son significativos**. Tanto las variedades de aguacates utilizadas, como la pendiente del terreno y la disponibilidad de nitrógeno influyen en la productividad de los aguacates.

**2. ¿Qué supuestos han de verificarse?**

Los supuestos que han de verificarse en un diseño de cuadrados latinos son Normalidad, Homocedasticidad e Independencia además del supuesto de aditividad entre filas, columnas y tratamientos (es decir, que no haya interacciones entre los mismos).

**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(mod$residuals)**

 **Shapiro-Wilk normality test**

**data: mod$residuals**

**W = 0.89854, p-value = 0.07616**

Observamos el contraste de Shapiro-Wilk que es adecuado cuando las muestras son pequeñas (n<50) y es una alternativa más potente que el test de Kolmogorov-Smirnov. El p-valor (0.07616) 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(propuesto5$productividad, propuesto5$aguacate)

 Bartlett test of homogeneity of variances

data: propuesto5$productividad and propuesto5$aguacate

Bartlett's K-squared = 3.355, df = 3, p-value = 0.3401

bartlett.test(propuesto5$productividad, propuesto5$aguacate)

El p-valor es **0.3401** por lo tanto no se puede rechazar la hipótesis de homogeneidad de las varianzas y se concluye que la cuatro variedades tienen varianzas homogéneas.

> bartlett.test(propuesto5$productividad, propuesto5$nitrogeno)

 Bartlett test of homogeneity of variances

data: propuesto5$productividad and propuesto5$nitrogeno

Bartlett's K-squared = 0.56609, df = 3, p-value = 0.9041

> bartlett.test(propuesto5$productividad, propuesto5$pendiente)

 Bartlett test of homogeneity of variances

data: propuesto5$productividad and propuesto5$pendiente

Bartlett's K-squared = 0.4168, df = 3, p-value = 0.9368

Los p-valores son mayores que 0.05, por lo tanto no se puede rechazar la hipótesis de homogeneidad de las varianzas en ninguno de los factores 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(mod)



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.

**3. ¿Se obtiene la misma producción con las cuatro variedades de aguacate? En caso negativo, analizar mediante el procedimiento de Tukey y Newman-Keuls, con qué variedad de aguacate hay mayor producción.**

> TukeyHSD(mod, "aguacate", ordered = TRUE)

 Tukey multiple comparisons of means

 95% family-wise confidence level

 factor levels have been ordered

Fit: aov(formula = productividad ~ aguacate + nitrogeno + pendiente, data = propuesto5)

$aguacate

 diff lwr upr p adj

A-B 33.75 23.1507168 44.34928 0.0001393

D-B 38.75 28.1507168 49.34928 0.0000629

C-B 50.00 39.4007168 60.59928 0.0000134

D-A 5.00 -5.5992832 15.59928 0.4289199

C-A 16.25 5.6507168 26.84928 0.0072734

C-D 11.25 0.6507168 21.84928 0.0392225

La tabla de comparaciones múltiples muestra los intervalos simultáneos construidos por el método de Tukey para cada posible combinación de variedades de aguacates. En la tabla se muestra un resumen de las comparaciones de cada tratamiento con los restantes. Es decir, aparecen comparadas dos a dos las cuatro medias de los tratamientos.

En esta tabla, las columnas:

* **diff:** muestra las medias de cada par
* **p adj:** muestra los p-valores de los contrastes, que permiten conocer si la diferencia entre cada pareja de medias es significativa al nivel de significación considerado (en este caso 0.05)
* **lwr** y **upr:** proporcionan los intervalos de confianza al 95% para cada diferencia.

Así por ejemplo, si comparamos la concentración media correspondiente a las variedades A Y B, tenemos una diferencia entre ambas medias de **33.75**, un p-valor (Sig.) de **0.0001393** significativo. Por lo tanto, las producciones medias de las variedades A y B pueden considerarse distintas estadísticamente y un intervalo de confianza con un límite inferior 23.1507168 y un límite superior 44.34928 y por lo tanto no contiene al cero de lo que también deducimos que hay diferencias significativas entre estas dos variedades de aguacate.

Como se puede observar, todos los intervalos de confianza construidos para las diferencias entre las producciones medias de las variedades no contienen al 0, excepto el correspondiente a la pareja de variedades de aguacates A y D. Lo que significa que todas las producciones medias pueden considerarse distintas estadísticamente excepto las producciones medias correspondientes a las variedades **A y D**. Se deduce que únicamente no se observan diferencias significativas entre las producciones de las variedades de aguacates **A y D** (**P-valor = 0.4289199**).

> plot(TukeyHSD(mod, "aguacate"))



Esta gráfica muestra los intervalos para cada pareja de medias

> Newman\_Keuls <- SNK.test(mod,"aguacate", console=TRUE)

Study: mod ~ "aguacate"

Student Newman Keuls Test

for productividad

Mean Square Error: 18.75

aguacate, means

 productividad std r Min Max

A 811.25 68.84463 4 730 880

B 777.50 143.38177 4 595 945

C 827.50 141.50972 4 700 950

D 816.25 55.43389 4 760 885

Alpha: 0.05 ; DF Error: 6

Critical Range

 2 3 4

 7.492106 9.394633 10.599283

Means with the same letter are not significantly different.

 productividad groups

C 827.50 a

D 816.25 b

A 811.25 b

B 777.50 c

El contraste de Newman-keuls muestra una tabla de Subconjuntos homogéneos. En nuestro estudio sobre las producciones de aguacates se observan que hay tres subgrupos homogéneos, al primer subgrupo pertenece la **Variedad C**, con una producción de 827.50 media **kg/parcela**, el segundo las variedades **D y A**, con producciones medias de 816.25 y 811.25 Kg(parcela, respectivamente y el tercero la **Variedad B, con una producción media de 777.50 kg/parcela**. Por lo tanto, se observa que la producción media mayor se obtiene con la **Variedad C** (827.5 Kg/ parcela) y la menor con la **Variedad B** (777.50 Kg/parcela).

**Editor de R**

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

propuesto5<-read.table("propuesto5.txt", header = TRUE)

propuesto5

propuesto5$latina = factor(propuesto5$aguacate)

propuesto5$latina

propuesto5$Bloque1 = factor(propuesto5$nitrogeno)

propuesto5$Bloque1

propuesto5$Bloque2 = factor(propuesto5$pendiente)

propuesto5$Bloque2

mod <- aov(productividad~ aguacate + nitrogeno + pendiente, data = propuesto5 )

mod

summary(mod)

shapiro.test(mod$residuals)

bartlett.test(propuesto5$productividad, propuesto5$aguacate)

bartlett.test(propuesto5$productividad, propuesto5$nitrogeno)

bartlett.test(propuesto5$productividad, propuesto5$pendiente)

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

plot(mod)

TukeyHSD(mod, "aguacate", ordered = TRUE)

plot(TukeyHSD(mod, "aguacate"))

Newman\_Keuls <- SNK.test(mod,"aguacate", console=TRUE)