### **Ejercicio Propuesto 1 (Resuelto)**

**La convección es una forma de transferencia de calor por los fluidos debido a sus variaciones de densidad por la temperatura; las partes calientes ascienden y las frías descienden formando las corrientes de convección que hacen uniforme la temperatura del fluido. Se ha realizado un experimento para determinar las modificaciones de la densidad de fluido al elevar la temperatura en una determinada zona. Los resultados obtenidos han sido los siguientes:**

|  |  |
| --- | --- |
| Temperatura | Densidad |
| 100 | 21.8 | 21.9 | 21.7 | 21.6 | 21.7 |
| 125 | 21.7 | 21.4 | 21.5 | 21.4 |  |
| 150 | 21.9 | 21.8 | 21.8 | 21.6 | 21.5 |
| 175 | 21.9 | 22.1 | 21.85 | 21.9 |  |

**Responder a las siguientes cuestiones:**

1. **¿Afecta la temperatura a la densidad del fluido?**
2. **Determinar qué temperaturas producen modificaciones significativas en la densidad media del fluido.**
3. **Estudiar las hipótesis del modelo: Homocedasticidad, independencia y normalidad.**
4. **Se puede afirmar que las temperaturas de 100 y 125 producen menos densidades de fluido en promedio que las temperaturas de 150 y 175.**

#### **Solución:**

El problema planteado se modeliza a través de un diseño **unifactorial totalmente aleatorizado de efectos fijos no-equilibrado.**

* **Variable respuesta:** *Densidad del fluido.*
* **Factor: *Temperatura*:** Es un factor de Efectos fijos.
* **Modelo no-equilibrado**: Los niveles de los factores tienen distinto número de elementos.

**1. ¿Afecta la temperatura a la densidad del fluido?**

Para responder a este apartado, se plantea el siguiente contraste de igualdad de medias:

$H\_{0}≡μ\_{1}+μ\_{2}+μ\_{3}+μ\_{4}$ vs $H\_{1}≡μ\_{i}\ne μ\_{j}$ para algún i $\ne j$

Para realizar este supuesto en **R** debemos introducir primero los datos de forma correcta. Podemos realizarlo 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:

Densidad Temperatura

21.80 100

21.90 100

21.70 100

21.60 100

21.70 100

21.70 125

21.40 125

21.50 125

21.40 125

21.90 150

21.80 150

21.80 150

21.60 150

21.50 150

21.90 175

22.10 175

21.85 175

21.90 175

* **Variable respuesta: *Concentración de CO.***
* **Factor**: ***Día de la semana*** que tiene cinco niveles. Es un factor de **efectos fijos** ya que viene decidido qué niveles concretos se van a utilizar (5 días de la semana).
* **Modelo equilibrado:** Los niveles de los factores tienen el mismo número de elementos (8 elementos).
* **Tamaño del experimento:** Número total de observaciones, en este caso 40 unidades experimentales.

El problema planteado se modeliza a través de un **diseño unifactorial totalmente aleatorizado de efectos fijos equilibrado.**

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")
> propuesto1 <- read.table("propuesto1.txt", header = TRUE)

> propuesto1

 Densidad Temperatura

1 21.80 100

2 21.90 100

3 21.70 100

4 21.60 100

5 21.70 100

6 21.70 125

7 21.40 125

8 21.50 125

9 21.40 125

10 21.90 150

11 21.80 150

12 21.80 150

13 21.60 150

14 21.50 150

15 21.90 175

16 22.10 175

17 21.85 175

18 21.90 175

Debemos transformar la variable referente a los niveles del factor fijo como factor para poder hacer los cálculos de forma adecuada

**> propuesto1$Temperatura <- factor(propuesto1$Temperatura)**

**> propuesto1$Temperatura**

 **[1] 100 100 100 100 100 125 125 125 125 150 150 150 150 150 175 175 175 175**

**Levels: 100 125 150 175**

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

**> mod <- aov(Densidad ~ Temperatura, data = propuesto1)**

donde:

* **Densidad**: Nombre de la columna de las observaciones.
* **Temperatura:** Nombre de la columna en la que están representados los tratamientos.
* data= data.frame en el que están guardados los datos.

**> mod**

**Call:**

 **aov(formula = Densidad ~ Temperatura, data = propuesto1)**

**Terms:**

 **Temperatura Residuals**

**Sum of Squares 0.384375 0.256875**

**Deg. of Freedom 3 14**

**Residual standard error: 0.1354556**

**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)**

**Temperatura 3 0.3844 0.12813 6.983 0.00419 \*\***

**Residuals 14 0.2569 0.01835**

**---**

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

El valor de F experimental 6.983, deja a su derecha un p-valor = 0.00419 inferior a 0.05, por lo que se rechaza la hipótesis nula de igualdad de medias. Concluyendo que existen diferencias significativas en la densidad del fluido en función de la modificación de la temperatura

**2. Determinar qué temperaturas producen modificaciones significativas en la densidad media del fluido.**

Se plantea entonces la pregunta de si la densidad media del fluido es significativamente diferente para las 4 temperaturas analizadas o sólo para alguna de ellas. Esta cuestión se resuelve mediante los contrastes de comparaciones múltiples. Vamos a utilizar la prueba de Tukey, obteniéndose los siguientes resultados:

**> mod.tukey <- TukeyHSD(mod, ordered = TRUE)**

**> mod.tukey**

 **Tukey multiple comparisons of means**

 **95% family-wise confidence level**

 **factor levels have been ordered**

**Fit: aov(formula = Densidad ~ Temperatura, data = propuesto1)**

**$`Temperatura`**

 **diff lwr upr p adj**

**150-125 0.2200 -0.04410917 0.4841092 0.1184044**

**100-125 0.2400 -0.02410917 0.5041092 0.0807121**

**175-125 0.4375 0.15910449 0.7158955 0.0021907**

**100-150 0.0200 -0.22900452 0.2690045 0.9953076**

**175-150 0.2175 -0.04660917 0.4816092 0.1240769**

**175-100 0.1975 -0.06660917 0.4616092 0.1785203**

La tabla de comparaciones múltiples muestra los intervalos simultáneos construidos por el método de Tukey para cada posible combinación de temperaturas.

A partir de la tabla de comparaciones múltiples se deduce que sólo se observan diferencias significativas entre las densidades de los fluidos cuando se ha modificado la temperatura a 125 y 175 grados (significación inferior a 0.05).

Como se puede observar todos los intervalos de confianza construidos para las diferencias entre las densidades medias contienen al 0, lo que significa que dichas densidades medias pueden considerarse iguales. Sin embargo, la significación asociada al contraste de las densidades medias correspondientes a las temperaturas de 125 y 175 grados es inferior a 0.05, lo que se traduce en que existe evidencia empírica de que ambas densidades medias son diferentes.

Aplicaremos también el contraste de comparaciones múltiples de Duncan

Primero instalaremos la librería “**agricolae”**

> library(agricolae)

> duncan <-duncan.test(mod, "Temperatura" ,   main= " Modificaciones de la densidad de fluido al elevar la temperatura ")
> duncan

$`statistics`

 MSerror Df Mean CV

 0.01834821 14 21.725 0.6235009

$parameters

 test name.t ntr alpha

 Duncan Temperatura 4 0.05

$duncan

NULL

$means

 Densidad std r Min Max Q25 Q50 Q75

100 21.7400 0.1140175 5 21.60 21.9 21.7000 21.70 21.80

125 21.5000 0.1414214 4 21.40 21.7 21.4000 21.45 21.55

150 21.7200 0.1643168 5 21.50 21.9 21.6000 21.80 21.80

175 21.9375 0.1108678 4 21.85 22.1 21.8875 21.90 21.95

$comparison

NULL

$groups

 Densidad groups

175 21.9375 a

100 21.7400 ab

150 21.7200 b

125 21.5000 c

attr(,"class")

[1] "group"

En la tabla Subconjuntos homogéneos asociada al contraste de Duncan se muestra por columnas los subgrupos de medias iguales. En nuestro estudio sobre las densidades de los fluidos se observan que la densidad media para temperaturas de 100, 125 y 150 se pueden considerar iguales y las densidades medias del fluido analizado pueden considerarse similares cuando las temperaturas son 100, 150 y 175 grados

El p-valor asociado al primer grupo de temperaturas (100, 125 y 150) es 0.105, mayor que 0.05 lo que significa que se puede aceptar la hipótesis de igualdad en la densidad media para este subgrupo. Análogamente ocurre con el otro subgrupo formado.

La densidad media de fluido es mayor a la temperatura de 175 grados (21.94) y menor a la temperatura de 125 grados (21.5).

**3. Estudiar las hipótesis del modelo: Homocedasticidad, independencia y normalidad**

Validar el modelo propuesto consiste en estudiar si las hipótesis básicas del modelo están o no en contradicción con los datos observados. Es decir, si se satisfacen los supuestos del modelo: Normalidad, Independencia y Homocedasticidad.

Hipótesis de **HOMOCEDASTICIDAD**

El primer aspecto que vamos a considerar es el de la homocedasticidad, la igualdad de varianzas, para ello, vamos a utilizar el test de Barlett

> bartlett.test(propuesto1$Densidad, propuesto1$Temperatura)

 Bartlett test of homogeneity of variances

data: propuesto1$Densidad and propuesto1$Temperatura

Bartlett's K-squared = 0.69212, df = 3, p-value = 0.8751

La salida muestra el resultado del contraste de **Barlett** de igualdad de varianzas en todos los grupos. El estadístico de contraste experimental, **B= 0.69212,** deja a la derecha un **p-valor = 0.8751,** que nos indica que no se debe rechazar la igualdad entre las varianzas entre las densidades registradas para las diferentes temperaturas.

Hipótesis de **INDEPENDENCIA**

**Hipótesis de Independencia:** Esta hipótesis la comprobaremos gráficamente mediante la representación de los residuos frente a los valores pronosticados por el modelo.

> layout(matrix(c(1,2,3,4),2,2))
> plot(mod)



Figura 1

En esta figura se muestran cuatro gráficos. Para comprobar que se satisface el supuesto de independencia entre los residuos, observamos el primer gráfico que se representan los residuos en el eje de ordenadas y los valores pronosticados en el eje de abscisas. La presencia de alguna tendencia en el gráfico puede indicar la alteración de dicha hipótesis. En este gráfico no se observa ninguna tendencia concreta lo que muestra la no existencia de relación de dependencia.

Hipótesis de **NORMALIDAD**

La hipótesis de Normalidad la comprobaremos gráficamente y analíticamente

Gráficamente comprobaremos la normalidad mediante el histograma y el gráfico **Q-Q plot**

En la figura 1, el gráfico **Q-Q plot** es el que se muestra en la Fila 2, Columna 1. En el que se representa los residuos frente a los cuantiles.

El gráfico Q-Q plot es el procedimiento gráfico más utilizado para comprobar la normalidad de un conjunto de datos. El Gráfico representa las funciones de distribución teórica y empírica (Cuantiles teóricos frente a cuantiles muestrales). En el eje de ordenadas se representa la función teórica bajo el supuesto de normalidad y en el eje de abscisas, la función empírica. Desviaciones de los puntos del gráfico respecto de la diagonal indican alteraciones de la normalidad. Observamos la ubicación de los puntos del gráfico en la figura 1, estos puntos se aproximan razonablemente bien a la diagonal lo que confirma la hipótesis de normalidad.

Para representar el histograma, en primer lugar calculemos los residuos del modelo

> g = mod$residuals

Calculamos la media de los residuos

> m <- mean(g)
> m
[1] -6.717912e-19

Calculamos la desviación típica

> std <- sqrt(var(g))
> std
[1] 0.1229239

Representamos el histograma

> hist(g, prob = TRUE, xlab ="Residuos", ylim = c(0, 5), main = "F. dens. Normal hist. de residuos")



Figura 2

En la figura 2 hemos representado el histograma de los residuos

Y para representar la curva Normal sobre el Histograma realizamos la siguiente orden

**> curve(dnorm(x, mean = m, sd = std), col ="darkblue", lwd =2, add = TRUE, yaxt ="n")**



Figura 3

En la Figura 3 hemos representado el histograma y la curva normal. Observamos que, como el gráfico Q-Q, no muestra desviaciones importantes a la normalidad

Analíticamente, vamos a comprobar la hipótesis de normalidad mediante el contraste de **Shapiro-Wilk**

> shapiro.test(mod$residuals)

 **Shapiro-Wilk normality test**

**data: mod$residuals**

**W = 0.94639, p-value = 0.3712**

El valor del p-valor es de **0.3712**, por lo tanto no podemos rechazar la hipótesis de normalidad de los residuos

**4. Se puede afirmar que las temperaturas de 150 y 125 producen menos densidades de fluido en promedio que las temperaturas de 100 y 175**

En el apartado 2, el test de Duncan nos muestra

**$groups**

 **Densidad groups**

**175 21.9375 a**

**100 21.7400 ab**

**150 21.7200 b**

**125 21.5000 c**

**Se muestra en la tabla que las temperaturas de 150 y 125 producen menos densidades de fluido en promedio que las temperaturas de 100 y 175**

**Editor de R**

setwd("C:/Users/Usuario/Desktop/Datos") \*Directorio donde estén los datos
propuesto1 <- read.table("propuesto1.txt", header = TRUE)

propuesto1$Temperatura <- factor(propuesto1$Temperatura)

propuesto1$Temperatura

mod <- aov(Densidad ~ Temperatura, data = propuesto1)

summary(mod) # Tabla ANOVA

mod.tukey <- TukeyHSD(mod, ordered = TRUE) # Comparaciones múltiples

mod.tukey

library(agricolae)

duncan <-duncan.test(mod, "Temperatura" ,   main= " Modificaciones de la densidad de fluido al elevar la temperatura ")
duncan

bartlett.test(propuesto1$Densidad, propuesto1$Temperatura) # Homocedasticidad

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

 plot(mod)

m <- mean(g) # Normalidad

 std <- sqrt(var(g))

 hist(g, prob = TRUE, xlab ="Residuos", ylim = c(0, 5), main = "F. dens. Normal hist. de residuos")

curve(dnorm(x, mean = m, sd = std), col ="darkblue", lwd =2, add = TRUE, yaxt ="n")

 shapiro.test(mod$residuals)