**Ejercicio Propuesto 3  (Resuelto)**

**Se quiere estudiar el efecto a distintas dosis de un medicamento que está en proceso de experimentación para combatir al parásito *Sparicotyle chrysophrii*, que comúnmente se encuentra en las branquias de las doradas (*Sparus aurata*) criadas en acuicultura. Para ello, se tomaron 30 doradas parasitadas al azar, y se dividieron en 3 grupos de 10 individuos cada uno. El primer grupo llamado Control no fue medicado, pero a los dos restantes se les suministró el medicamento en dosis distintas. Tras una semana de tratamiento, se contabilizó el número de parásitos existentes en cada individuo, obteniendo los resultados siguientes:**

****

Tabla 1: Datos del Ejercicio Propuesto 3

1. **¿Existen diferencias significativas en la cantidad de parásitos según la dosis del medicamento suministrada? Utilice un nivel de significación del 5%.**

**2. En caso afirmativo en el apartado anterior, ¿entre qué dosis? Utilice un nivel de significación del 5%.**

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

**Solución**

1. **¿Existen diferencias significativas en la cantidad de parásitos según la dosis del medicamento suministrada? Utilice un nivel de significación del 5%.**

Nos interesa saber si el control y las distintas dosis de medicamento influyen significativamente en el número de parásitos existentes en las branquias de las doradas

* El número de parásitos por individuo es una **variable numérica**
* La dosis de medicamento suministrada es una **variable categórica** que divide a los individuos en grupos

Nuestro contraste de hipótesis es:



Es decir,

$$\begin{matrix}H\_{0}:el número de parásitos en las branquias de las doradas es el mismo para los tres grupos\\H\_{1}:el número de parásitos en las branquias de las doradas es diferente según el grupo \end{matrix}$$

En primer lugar, cargamos el paquete BrailleR

> library(“BrailleR”)

almacenamos los datos de las dos variables en dos vectores:

> grupo = factor(c(rep("Control",10), rep("50mg",10), rep("125mg",10)),levels = c("Control", "50mg", "125mg"))

> parasitos=c(50, 65, 60, 46, 38, 29, 61, 85, 62, 40, 20, 59, 64, 61, 28, 47, 29, 41, 60, 57, 30, 45, 52, 46, 31, 21, 34, 32, 51, 36)

Agrupamos las 2 variables en un data frame, al que llamamos *medicamento*:

> medicamento = data.frame (grupo, parasitos)

Realizamos el ANOVA pedido mediante la función **OneFactor** sin que nos muestre el Test post-hoc.

> OneFactor("parasitos", "grupo", medicamento, HSD = FALSE)

[Muestra una salida en html:](file:///C%3A%5CUsers%5CUsuario%5CDesktop%5Cdatos%5CParasitos.Grupo-OneFactor.html)

# **Analysis of the Medicamento data, using Parasitos as the response variable and Grupo as the single grouping factor.**

#### **Prepared by BrailleR**

A continuación, se muestra y se detalla cada resultado:

1. **Una tabla resumen que incluye para cada grupo la media, la desviación típica, el número de datos y el error estándar**

****

Tabla 2: Resumen para cada grupo

1. **Diagramas de caja y bigotes comparativos para cada grupo**

****

Figura 1: Diagrama de caja y bigotes para cada grupo

* + En ninguno de los tres grupos se ve claramente la simetría de la distribución.
	+ No hay valores atípicos.
	+ El tamaño de las cajas es similar para los distintos niveles, por lo que no hay indicios de falta de homocedasticidad.
	+ A simple vista no parece que vaya a haber diferencia de medias.
1. **Gráfico de puntos comparativos para cada grupo**

****

Figura 2: Gráfico de puntos para cada grupo

Al igual que en el diagrama de caja y bigotes, se puede tener una idea intuitiva de si las medias difieren unas de otras o no. Obteniendo las mismas conclusiones que con el gráfico anterior.

1. **Análisis de la varianza de un factor**

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

grupo 2 1254 626.8 2.986 0.0674 .

Residuals 27 5668 209.9

---

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

Esta es la tabla ANOVA que nos proporciona un p-valor=0.0674 por lo que no tenemos evidencia muestral para rechazar la hipótesis nula, esto quiere decir que no hay diferencias significativas en la cantidad de parásitos según la dosis del medicamento suministrada.

1. **En caso afirmativo en el apartado anterior, ¿entre qué dosis? Utilice un nivel de significación del 5%.**

Puesto que en el apartado anterior no hemos obtenido diferencias significativas, no tiene sentido responder a esta pregunta.

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

Para comprobar la idoneidad del modelo propuesto, la salida de la función **OneFactor** muestra los gráficos:

1. **Valores ajustados frente a residuos**
2. **Gráfico Q-Q de normalidad**
3. **Valores ajustados frente a raíz cuadrada de los residuos estandarizados (en valor absoluto)**
4. **Residuos estandarizados frente a leverages**

****

Figura 3: Gráficos de comprobación de las hipótesis

Los gráficos 1 y 3 se utilizan para contrastar gráficamente si la media de los residuos es cero y la homocedasticidad de los residuos. La idea intuitiva para que se cumplan estas hipótesis es que no se observe ningún patrón ni forma de embudo. Además, en el gráfico 3 se usa como regla general que cualquier punto por encima de 2 en el eje Y sugeriría heterogeneidad de varianza. Por tanto, se confirma que los residuos tienen media cero y son homocedásticos.

Observando el gráfico Q-Q, se observa que los residuos estandarizados están próximos a la diagonal punteada que aparece en el gráfico, por tanto, los residuos siguen una distribución normal.

El gráfico de Leverage para cada grupo frente a los residuos estandarizados se utiliza para detectar valores atípicos. En caso de detectarse alguna observación fuera del rango [-2,2] debe estudiarse este punto de forma aislada para detectar, por ejemplo, si la elevada importancia de esa observación se debe a un error. No se detectan valores atípicos.

La hipótesis de independencia es intrínseca al diseño del experimento, por tanto, se supone que las observaciones son independientes ya que se están considerando distintas dosis de medicamento.

1. **Dos tests de homogeneidad de varianzas**

Vamos a confirmar la homocedasticidad mediante dos métodos analíticos.

* **El test de homogeneidad de varianzas de Bartlett**

 Bartlett test of homogeneity of variances

data: parasitos by grupo

Bartlett's K-squared = 2.168, df = 2, p-value = 0.3382

Considerando como nivel de significación 0.05, y observando el p-valor=0.3382 no tenemos evidencia muestral para rechazar la hipótesis nula, por lo que la varianza permanece constante entre grupos.

* **El test de homogeneidad de varianzas de Fligner-Killeen**

 Fligner-Killeen test of homogeneity of variances

data: parasitos by grupo

Fligner-Killeen:med chi-squared = 2.2805, df = 2, p-value = 0.3197

Considerando como nivel de significación 0.05, y observando el p-valor=0.3197 no tenemos evidencia muestral para rechazar la hipótesis nula, por lo que la varianza permanece constante entre grupos.

Considerando como nivel de significación 0.05, y observando el p-valor=0.3197 no tenemos evidencia muestral para rechazar la hipótesis nula, por lo que la varianza permanece constante entre grupos.