

Regresión y ANOVA: ANOVA de 1 factor

Cuervo Fernández, Esther García Prado, Sergio Martín Villares, Pablo

4 de octubre de 2017

1. Ejercicio [TODO elegir uno]

[TODO]

2. Ejercicio Acuicultura

El enunciado del ejercicio es el siguiente: “Se quiere estudiar el efecto de distintas dosis de un medicamento para combatir a los parásitos de peces criados en acuicultura. Para ello, se tomaron 60 peces al azar, y se dividieron en 5 grupos de 12 individuos cada uno. El primer grupo no fue medicado, pero a los restantes se les suministró el medicamento en dosis crecientes. Tras una semana de tratamiento, se contabilizaron los parásitos existentes en cada individuo”.

2.1. Representa gráficamente los datos. ¿Te parece que el medicamento es efectivo contra los parásitos? Realiza un contraste de hipótesis para verificarlo.

Lo primero que se ha hecho es realizar una representación gráfica para conocer mejor la relación entre las variables. Por tanto, se ha elegido representar los datos mediante un *Box Plot* que de la variable `niveles` particionada en los subgrupos determinados por el tipo de `grupo`.

A partir de dicho gráfico se pueden apreciar dos grupos de medias diferentes, el primero de ellos formado por los tratamientos de `Control`, `25mg` y `50mg`, y el siguiente constituido por los tratamientos de `100mg` y `125mg`.

Para poder asegurar que el tratamiento es efectivo contra los parásitos se ha realizado el siguiente contraste de hipótesis:

$$H_0 : \mu_{control} = \frac{\mu_{25mg} + \mu_{50mg} + \mu_{100mg} + \mu_{125mg}}{4} \quad (1)$$

$$H_1 : \mu_{control} \neq \frac{\mu_{25mg} + \mu_{50mg} + \mu_{100mg} + \mu_{125mg}}{4} \quad (2)$$

Si se acepta entonces podemos asegurar que el tratamiento no tiene efecto. Se ha escogido este contraste de hipótesis puesto que lo que se pretende comprobar es si el efecto del medicamento (en promedio) es diferente de los resultados que se obtienen sin el uso del mismo. Por tanto, se ha planteado como la igualdad de la media del valor de control y el promedio de las medias de cada cantidad de dosis del medicamento.

Figura 1: Box Plot de la variable `niveles` condicionada por la variable `grupo`

Dicho hipótesis se ha realizado con un nivel de confianza del 95 %, para la cual se ha obtenido un p-valor ≈ 0 . Por esta razón **tenemos que rechazar la hipótesis nula, lo que significa que el medicamento es efectivo**. Puesto que el estimador obtenido tiene signo negativo, es decir, que los distintos tratamientos disminuyen el nivel de parásitos con respecto del valor de control.

2.2. Obtén estimadores puntuales e intervalos de confianza para las medias de los tratamientos, utilizando dos métodos distintos, y explica las ventajas e inconvenientes de cada uno de ellos.

Para la obtención de los estimadores puntuales así como los intervalos de confianza se han utilizado los *procs* MEANS y UNIVARIATE obteniendo los siguientes resultados en cada uno de ellos:

[TABLAS de RESULTADOS]

Estos dos métodos ofrecen los mismos resultados, sin embargo lo que les diferencia es la cantidad de información que aportan cada uno de ellos. En el caso de MEANS la información es mucho más compacta y resumida, lo cual facilita su comprensión. Mientras que en el caso de UNIVARIATE se obtiene información mucho más extendida y ampliada, como los valores de la varianza e intervalos de confianza para los estimadores de la esperanza y la varianza. Para casos en que se requiere de un resumen sencillo del conjunto de datos como es en este caso, es más apropiado el uso del *proc* MEANS

2.3. Analiza las diferencias significativas entre pares de medias de los 5 grupos utilizando diferentes métodos, comentando y comparando los resultados. ¿Cambian las conclusiones si utilizas $\alpha = 0,1$?

Para la realización de los distintos tests se han utilizado los siguientes métodos: *Método de la T*, *Duncan*, *NewmanKeuls*, *Tukey*, *Bonferroni* y *Scheffé*. Estos se dividen en dos grupos, los que aseguran los resultados para experimentos múltiples con una confianza global de $1 - \alpha$ (*NewmanKeuls*, *Tukey*, *Bonferroni* y *Scheffé*) y los que no aseguran dicho nivel de confianza (*Método de la T*, *Duncan*).

Se ha probado con un valor $\alpha = 0,05$, para el cual se han obtenido los mismos resultados en todos los tests: Existen dos grupos de medias, el formado por **Control**, 25mg y 50mg, y el siguiente constituido por los tratamientos de 100mg y 125mg tal y como se podía intuir en el gráfico de la figura .

Después se ha realizado el mismo conjunto de tests con $\alpha = 0,1$ obteniendo los mismos resultados que para el caso anterior en los tests que aseguran un nivel de confianza global $1 - \alpha$, mientras que en el resto (*Método de la T*, *Duncan*), estos indican que la media del grupo **Control** es diferente del grupo formado por los tratamientos 25mg y 50mg, por lo que afirman la existencia de 3 grupos de medias diferentes.

2.4. Realiza el test de Dunnett para comparar los efectos de las distintas dosis con el grupo control y comenta el resultado.

Tras aplicar el test de *Dunnett* usando el tratamiento **control** para compararlo con el resto de tratamientos con un nivel de confianza de $\alpha = 0,05$ para todo el experimento, se ha concluido que las medias de los tratamientos 100mg y 125mg son diferentes de las de **control**. Sin embargo, las de 25mg y 50mg son tomadas como iguales.

- 2.5. Se desea comparar el efecto de las dosis bajas del medicamento, 25-50 mg con el de las dosis altas, 25-50 mg. Construye un intervalo de confianza para la diferencia entre ambos efectos e interpreta el resultado obtenido**

En este caso se ha planteado el contraste de hipótesis como la igualdad del promedio de medias de dosis altas con las dosis bajas, por tanto puede escribirse como:

$$H_0 : \frac{\mu_{25mg} + \mu_{50mg}}{2} = \frac{\mu_{100mg} + \mu_{125mg}}{2} \quad (3)$$

$$H_1 : \frac{\mu_{25mg} + \mu_{50mg}}{2} \neq \frac{\mu_{100mg} + \mu_{125mg}}{2} \quad (4)$$

Se ha realizado dicho contraste con un nivel de confianza $1 - \alpha$ con $\alpha = 0,05$, para el cual se obtiene un p-valor ≈ 0 por lo que se debe rechazar la hipótesis nula, es decir, no existe igualdad de medias entre los resultados obtenidos con dosis bajas y dosis altas.

El intervalo de confianza para el estimador de la diferencia de medias obtenidos es $[-32,88, -17,53]$. Esto puede interpretarse como que las dosis altas disminuyen el nivel de parásitos entre 17,53 y 32,88 respecto de las dosis bajas.

- 2.6. Verifica las hipótesis del modelo con ayuda de gráficos de residuos.**

[TODO]

- 2.7. Suponiendo que los grupos de datos correspondieran a cinco medicamentos elegidos aleatoriamente entre todos los posibles en el mercado, ¿cómo cambiaría el modelo adecuado? ¿Podría afirmarse que existe variación significativa entre los efectos de los distintos medicamentos? Estima las componentes de la varianza bajo dicho supuesto**

[TODO]