

Análisis de la Varianza (ANOVA) de un factor y test a posteriori.

Ejercicios Temas 8 y 9 (Resuelto)

1. Problema 5

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, obteniendo los resultados siguientes:

Control	50	65	72	46	38	29	70	85	72	40	57	59
25mg	49	47	30	602	62	60	19	28	56	62	55	40
50mg	20	59	64	61	28	47	29	41	60	57	61	38
100mg	20	23	38	31	27	16	27	18	22	12	24	11
125mg	18	30	22	26	31	11	15	12	31	36	16	13

Contrastar si el medicamento es efectivo contra los parásitos y si existen diferencias según la dosis aplicada.

1.1. Resolución

1.1.1. Planteamiento del contraste

En este caso se quiere contrastar si el medicamento es efectivo y en qué dosis respecto al control observado.

Por tanto, y dado que se dispone de datos para diferentes tratamientos (en este caso 2 o más), es factible utilizar un *análisis de la varianza de un factor*, contraste para la media de una variable continua normal en dos o más grupos, donde el factor es la dosis, que contiene cinco clases o grupos o niveles, y la variable numérica a analizar es el número de parásitos contabilizados en cada uno de los 60 peces estudiados, agrupados en los diferentes tratamientos y el control.

1.1.2. Planteamiento de la hipótesis

Del planteamiento del contraste anterior se deduce que las hipótesis a contrastar para analizar si existen diferencias entre los k grupos son:

$$\left. \begin{array}{l} H_0 : \text{Todas las medias poblaciones de los } k \text{ grupos son iguales} \\ H_1 : \text{Al menos una media poblacional difiere} \end{array} \right\} \begin{array}{l} \rightarrow \beta \\ \rightarrow \alpha = 0.05 \end{array}$$

$$\left. \begin{array}{l} H_0 : \mu_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 \\ H_1 : \text{Al menos una igualdad no es cierta} \end{array} \right\} \begin{array}{l} \rightarrow \beta \\ \rightarrow \alpha = 0.05 \end{array}$$

1.1.3. Estadístico de contraste y gráfico

El estadístico experimental se basará en una F de de Snedecor con $k - 1$ y $n - k$ grados de libertad para el numerador y denominador respectivamente. Por tanto se debe comparar dicho estadístico F_{expt} con el correspondiente punto crítico F_{teo} . Éstos se definen como:

$$F_{expt} = \frac{Q_E/k - 1}{Q_D/n - k}$$

$$P_c = F_{teo} = F_{gl_{num}, gl_{denom}, \alpha} = F_{k-1, n-k, \alpha}$$

La representación gráfica puede verse en el Apartado b) del desarrollo del contraste.

1.1.4. Requisitos del contraste y datos necesarios para su desarrollo

a) Datos necesarios

El resumen de los datos observados se presentan en la siguiente tabla, agrupados para cada una de las clases para poder observar las posibles diferencias entre las medias y también analizar las varianzas:

Clase	n_i	\bar{X}_i	S_i^2	S_i
Control	12	56.917	279.538	16.719
25mg	12	47.500	224.818	14.994
50mg	12	47.083	238.629	15.448
100mg	12	22.417	60.629	7.786
125mg	12	21.750	76.386	8.740

Y la información referente al conjunto de los datos:

$$\begin{aligned}
 k &= 5 \\
 n &= n_1 + n_1 + n_2 + n_3 + n_4 + n_5 = 60 \\
 \bar{X}_{..} &= \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij} = 39.133
 \end{aligned}$$

b) Requisitos

Las hipótesis necesarias para poder llevar a cabo un análisis ANOVA son:

1. *Independencia de los valores observados*, hipótesis necesaria que se comprueba mediante la inspección del diseño experimental. Si las unidades experimentales (en nuestro caso los peces) han sido seleccionadas al azar se asume como verdadero (por lo que se comenta en el enunciado).
2. *Normalidad de los datos* en cada una de las clases, se verifica mediante un contraste de bondad de ajuste a una variable normal. En este caso se debería utilizar un $K - S$ de Kolmogorov-Smirnov a partir de la hipótesis $H_0 =$ Los datos del nivel i se ajustan a la distribución normal para cada clase al no aparecer implícito en el enunciado (en caso de resolución para el examen se asume por simplicidad en la resolución y por la escasez de tiempo).
3. *Homogeneidad de varianzas*, debemos analizar si las varianzas son homogéneas (parecidas) y esto lo hacemos a partir de un contraste de homogeneidad de varianzas como la M de Bartlett para datos no balanceados o la G de Cochran para datos balanceados.

Como en este caso los datos son balanceados se desarrollará la G de Cochran en su versión resumida. En este contraste las hipótesis se definen como:

$$\left. \begin{aligned}
 H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 \\
 H_1 : \text{Al menos una igualdad no es cierta}
 \end{aligned} \right\} \begin{aligned}
 &\rightarrow \beta \\
 &\rightarrow \alpha = 0.05
 \end{aligned}$$

Para este contraste el estadístico y el punto crítico quedan definidos como:

$$\begin{aligned}
 G_{expt} &= \frac{\max(S_i^2)}{\sum_{i=1}^k S_i^2} \\
 P_c &= G_{teo} = G_{n,k,\alpha}^1
 \end{aligned}$$

Desarrollando y comparando ambos se obtiene que:

¹En este caso n es el N° de observaciones para una de las clases cualquiera, al estar los datos balanceados

$$G_{expt} = \frac{\max(S_i^2)}{\sum_{i=1}^k S_i^2} > G_{n,k,\alpha} = G_{teo} \Rightarrow \text{SE RECHAZA } H_0$$

$$G_{expt} = \frac{279.538}{880.000} = 0.318 < 0.4118 = G_{11,5,0.05}^2 \Rightarrow \text{NO SE RECHAZA } H_0$$

Por tanto no se puede demostrar que exista heterogeneidad de varianzas al no rechazar H_0 . (En este caso se cumple el requisito de Homogeneidad de varianzas).

1.1.5. Desarrollo del contraste

a) ANOVA

Para desarrollar el contraste se debe completar la tabla de ANOVA:

[0.4mm] F. de variación	Suma de cuadr.	g.l.	Cuadr. medios	F	Pr(> F)
[0.4mm] ENTRE	$Q_E = \sum_{i=1}^k n_i (\bar{X}_{i.} - \bar{X}_{..})^2$	$k - 1$	$S_E^2 = Q_E / k - 1$	$F = \frac{Q_E / k - 1}{Q_D / n - k}$	
DENTRO	$Q_D = \sum_{i=1}^k (n_i - 1) S_i^2$	$n - k$	$S_D^2 = Q_D / n - k$		
TOTAL	$Q = Q_E + Q_D$	$n - 1$	$S^2 = Q / n - 1$		

Desarrollando las ecuaciones obtenemos como resultado la tabla de ANOVA del ejercicio:

[0.4mm] F. de variación	Suma de cuadr.	g.l.	Cuadr. medios	F	Pr(> F)
[0.4mm] ENTRE	12372.878	4	3093.22	17.575	$p\text{valor} < 0.01$ ⁴
DENTRO	9680	55	176		
TOTAL	22052.878	59	373.778		

El contraste resulta ser significativo, dado que:

$$F_{expt} = 17.575 < 2.56 = F_{4,50,0.05} \approx F_{4,55,0.05}^5 \Rightarrow \text{SE RECHAZA } H_0$$

$$P(F_{4,55} > F_{expt}) = p - \text{valor} < 0.01 < 0.05 = \alpha \Rightarrow \text{SE RECHAZA } H_0$$

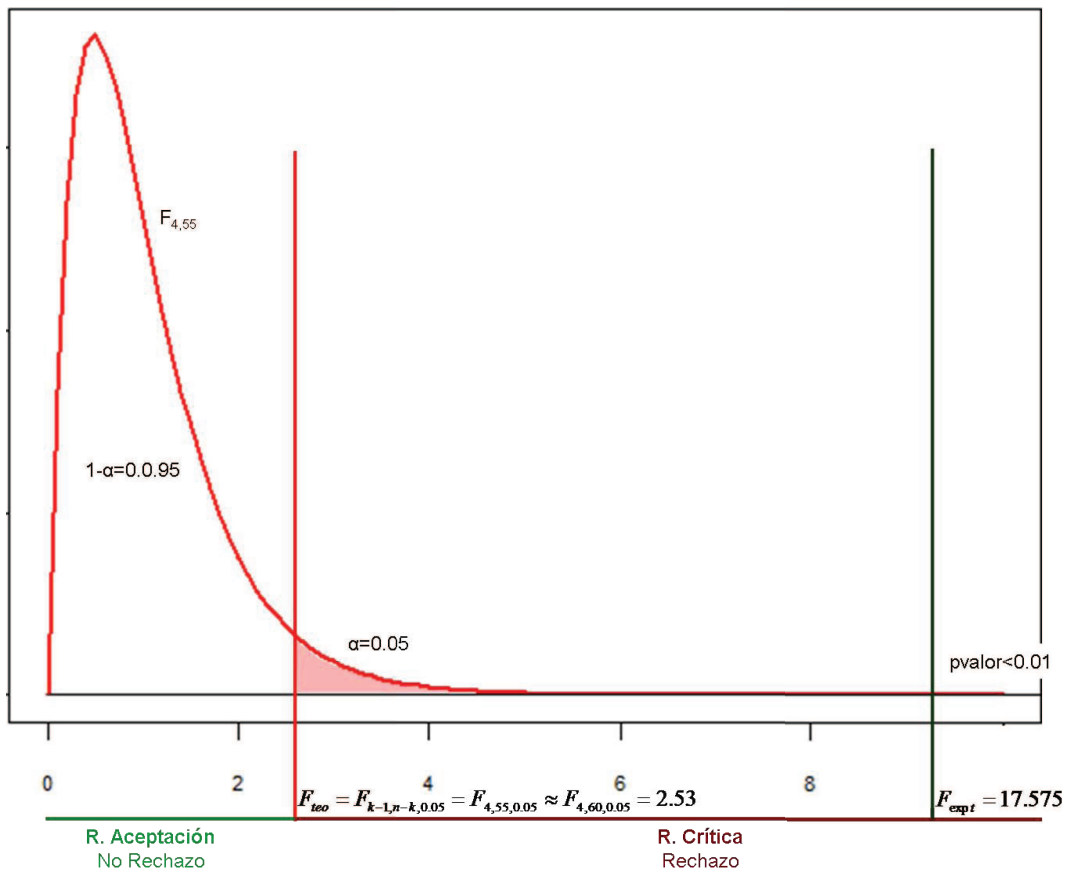
Conclusión: Se rechaza H_0 y por lo tanto al menos alguna de las clases son diferentes entre si (existen diferencias en los resultados para las dosis aplicadas). Es decir, existe *al menos* una dosis que tiene resultados diferentes a las demás. Como lo interesante una vez comprobado H_0 es averiguar cuál o cuáles son diferentes y en que sentido para detectar si alguna de las dosis es más efectiva se debe desarrollar un contraste a posteriori con objeto de identificar las diferencias ya detectas (pero no identificadas) en el análisis de ANOVA.

²En realidad sería $G_{12,5,0.05} = 0.4012$ pero como no está definido para $n = 12$, cogemos el menor más próximo a 12 en las tablas de la G

⁴A partir de la tabla de F, el p-valor obtenido a partir del programa R es $2.409e - 09$

⁵Como se puede comprobar, al no estar definido este punto en la tabla de la F, lo que se hace es redondear al siguiente inferior, en este caso $F_{4,50,0.05} = 2.56$

b) Representación gráfica del contraste ANOVA



c) Contraste a posteriori

Para identificar cuáles son las medias que pueden ser diferentes entre si se ha optado por desarrollar el contraste de comparaciones múltiples a posteriori de Bonferroni (no existen demasiadas clases y por tanto es factible su uso), basado en comparaciones por pares de todos con todos (los niveles) a partir de la t de Student para dos poblaciones con un parámetro penalizador (dado que se realizan comparaciones simultáneas).

Las hipótesis en este caso para la comparación de cada pareja de clases serán:

$$\left. \begin{array}{l} H_0 : \mu_i = \mu_j \rightarrow H_0 : \mu_i - \mu_j = 0 \\ H_1 : \mu_i \neq \mu_j \rightarrow H_1 : \mu_i - \mu_j \neq 0 \end{array} \right\} \begin{array}{l} \rightarrow \beta \\ \rightarrow \alpha/K' \end{array}$$

Y el estadístico experimental y el punto crítico para su comparación se definen como:

$$t_{expt_{ij}} = \frac{|\bar{X}_i - \bar{X}_j|}{S_d \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}}$$

$$P_c = t_{teo} = t_{n-k, \alpha/K'}$$

Para poder desarrollar el contraste se necesita conocer K' , S_d y t_{teo} :

$$K' = \frac{k!}{(k-2)!2!} = \frac{5!}{3!2!} = 10$$

$$S_d = \sqrt{\frac{Q_D}{n-k}} = \sqrt{S_D^2} = 13.266$$

$$t_{teo} = t_{n-k, \alpha/K'} = t_{60-5, 0.05/10} = t_{55, 0.005} = 2.6682$$

Desarrollando cada contraste para todas las posibles combinaciones de clases se obtiene que:

Muestras	$\bar{X}_i - \bar{X}_j$	t_{ij}	$t_{55,0.005}$	$P(t_{55} > t_{ij})$
Control-25mg	9.417	1.739	NoSignif.	$0.025 < p < 0.05$
Control-50mg	9.834	1.816	NoSignif.	$0.025 < p < 0.05$
Control-100mg	34.500	6.370	Signif.	$p < 0.005$
Control-125mg	35.167	6.493	Signif.	$p < 0.005$
25mg-50mg	0.417	0.077	NoSignif.	$p > 0.25$
25mg-100mg	25.083	4.631	Signif.	$p < 0.005$
25mg-125mg	25.750	4.755	Signif.	$p < 0.005$
50mg-100mg	24.666	4.554	Signif.	$p < 0.005$
50mg-125mg	25.333	4.678	Signif.	$p < 0.005$
100mg-125mg	0.667	0.123	NoSignif.	$p > 0.25$

Al ser diez contrastes mediante la t de Student, la representación gráfica se basaría en el gráfico de la t de Student para dos poblaciones normales de temas anteriores, uno para cada uno de los pares de comparaciones.

Analizando los resultados de los contrastes, podemos identificar dos subconjuntos homogéneos:

$$\underline{125mg = 100mg} < \underline{50mg = 25mg = control}$$

1.1.6. Conclusión biológica

El número de parásitos observados usando dosis de 25 y 50 mg no es significativamente distinto a los obtenidos en los peces control, por lo que estas dosis no son efectivas. Por el contrario, a pesar de que aplicando dosis de 125 y 100 mg el número de parásitos es significativamente menor a la observada en los controles, aplicando una dosis de 125 mg no se observa una diferencia significativa en el número de parásitos a la obtenida aplicando una dosis de 100 mg.

El análisis de los resultados nos permite concluir que el medicamento sí es efectivo contra los parásitos, pero logra reducir el número de parásitos de forma significativa a partir de una dosis de 100 mg.