



EXOPOL S.L.

Circulares Exopol: www.exopol.com

Alberto Morillo Alujas, Veterinario
Daniel Villalba Mata, Doctor Ingeniero Agrónomo
Jaume Feixa Cercós, Ingeniero Técnico Agrícola

Tests & Trials, S.L.
San Isidro 8
22540 Altorricón. Huesca
España

ANÁLISIS DE VARIANZA

Introducción

En el anterior capítulo, utilizamos los contrastes de significación para determinar si la media muestral se correspondía con la media de la población, y también los utilizamos para comparar dos medias. En los contrastes para comparar dos medias, las técnicas usadas permitían determinar el efecto de una variable sobre la media de otra variable, por ejemplo, el efecto de incluir harina de pescado en la dieta de los lechones sobre la ganancia media diaria de estos lechones.

Tendríamos que aclarar que los estudios estadísticos a menudo comparan más de dos poblaciones. Por tanto, es necesario un método para comparar cualquier número de medias.

Las comparaciones múltiples

Siguiendo con el anterior ejemplo, nos podríamos preguntar, por ejemplo, qué efecto tiene sobre la ganancia media diaria de los lechones incluir en la dieta harina de pescado en diferentes cantidades. Planteemos el ejemplo. Se ha distribuido tres dietas a 20 lechones en la etapa de prestarter. La dieta 1 contenía un 0% de harina de pescado, la dieta 2 contenía un 4% de harina de pescado, mientras que la dieta 3 contenía 8% de harina de pescado. Los resultados en cuanto a ganancia media diaria fueron los siguientes.

G.M.D. Dieta 0%	G.M.D. Dieta 4%	G.M.D. Dieta 8%
0,224	0,248	0,265
0,211	0,240	0,187
0,227	0,228	0,289
0,228	0,266	0,250
0,211	0,215	0,286
0,213	0,304	0,198
0,231	0,210	0,199
0,202	0,215	0,275
0,204	0,200	0,294
0,195	0,258	0,277
0,210	0,238	0,284
0,248	0,247	0,200
0,230	0,270	0,264
0,239	0,290	0,235
0,259	0,240	0,319
0,231	0,262	0,273
0,200	0,224	0,298
0,243	0,257	0,299
0,250	0,210	0,268
0,235	0,265	0,221



EXOPOL S.L.

Circulares Exopol: www.exopol.com

Calculemos, como hacíamos en el primer capítulo, Algunos de los estadísticos básicos, para empezar a analizar el problema. Siendo los resultados los siguientes:

	A	B	C	D	E	F	G	H
1	<i>Dieta 0%</i>			<i>Dieta 4%</i>			<i>Dieta 8%</i>	
2								
3	Media	0,22459727		Media	0,24440951		Media	0,25907864
4	Error típico	0,00407625		Error típico	0,00621399		Error típico	0,00879124
5	Mediana	0,2276926		Mediana	0,24363611		Mediana	0,27086051
6	Desviación estándar	0,01822955		Desviación estándar	0,02778981		Desviación estándar	0,03931562
7	Varianza de la muestra	0,00033232		Varianza de la muestra	0,00077227		Varianza de la muestra	0,00154572
8	Curtosis	-0,94340244		Curtosis	-0,35237601		Curtosis	-0,80461513
9	Coefficiente de asimetría	0,07965605		Coefficiente de asimetría	0,31926906		Coefficiente de asimetría	-0,6235822
10	Rango	0,06391303		Rango	0,10436774		Rango	0,13180626
11	Mínimo	0,19464249		Mínimo	0,19956873		Mínimo	0,186997
12	Máximo	0,25855552		Máximo	0,30393647		Máximo	0,31880326
13	Suma	4,49194545		Suma	4,88819015		Suma	5,18157281
14	Cuenta	20		Cuenta	20		Cuenta	20
15								

Como podemos comprobar la media de la ganancia media diaria aumenta a medida que aumenta el contenido de harina de pescado de la dieta de los lechones, ya que la media de ganancia media diaria para la dieta con 0% de harina de pescado es de 0,224, para la dieta de 4% es de 0,244, mientras que para la dieta de 8% es de 0,259. Ahora, la pregunta que nos deberíamos hacer es la siguiente, ¿son estadísticamente significativas estas diferencias?

La problemática de las comparaciones múltiples

Antes de nada, empecemos dando nombre a las medias de las ganancias medias diarias de los lechones. Para la dieta con 0% de harina de pescado la media será μ_1 , para la dieta con un 4% de harina de pescado la media la llamaremos μ_2 , y para la dieta con el 8% de harina de pescado la media será μ_3 .

Si pretendemos comparar estas tres medias podríamos usar varias veces la prueba t de dos muestras que vimos en capítulo anterior, planteando las siguientes hipótesis:

- Contrastando, $H_0: \mu_1 = \mu_2$
 $H_a: \mu_1 \neq \mu_2$

obtendríamos,

Prueba t

	<i>Dieta 0%</i>	<i>Dieta 4%</i>
Media	0,224597273	0,244409508
Varianza	0,000332316	0,000772273
Observaciones	20	20
Diferencia hipotética de las medias	0	
Grados de libertad	33	
Estadístico t	-2,665925185	
P(T<=t) una cola	0,005897692	
Valor crítico de t (una cola)	1,692360456	
P(T<=t) dos colas	0,011795385	
Valor crítico de t (dos colas)	2,03451691	



EXOPOL S.L.

Circulares Exopol: www.exopol.com

- Contrastando, $H_0: \mu_1 = \mu_3$
 $H_a: \mu_1 \neq \mu_3$
obtendríamos,

Prueba t

	<i>Dieta 0%</i>	<i>Dieta 8%</i>
Media	0,224597273	0,259078641
Varianza	0,000332316	0,001545718
Observaciones	20	20
Diferencia hipotética de las medias	0	
Grados de libertad	27	
Estadístico t	-3,558342376	
P(T<=t) una cola	0,000702746	
Valor crítico de t (una cola)	1,703288035	
P(T<=t) dos colas	0,001405491	
Valor crítico de t (dos colas)	2,051829142	

- Contrastando, $H_0: \mu_2 = \mu_3$
 $H_a: \mu_2 \neq \mu_3$
obtendríamos,

Prueba t

	<i>Dieta 4%</i>	<i>Dieta 8%</i>
Media	0,244409508	0,259078641
Varianza	0,000772273	0,001545718
Observaciones	20	20
Diferencia hipotética de las medias	0	
Grados de libertad	34	
Estadístico t	-1,362584789	
P(T<=t) una cola	0,090985713	
Valor crítico de t (una cola)	1,690923455	
P(T<=t) dos colas	0,181971426	
Valor crítico de t (dos colas)	2,032243174	

El inconveniente que nos encontramos al hacer estas tres pruebas, es que obtenemos tres valores P (ver capítulo anterior), uno para cada prueba. Esto no nos dice cuál es la probabilidad que otras tres medias muestrales cualquiera estén tan separadas como éstas. Es decir, puede ser que μ_1 y μ_3 sean estadísticamente distintas si nos fijamos en la comparación de estas dos medias, pero podría ser que no fueran estadísticamente distintas si sabemos que son la media menor y mayor de los tres grupos. En conclusión, no podemos comparar, de una manera categórica, muchas medias de dos en dos, mediante pruebas de significación.

El problema de comparar varias medias conjuntamente, es un problema típico de la estadística. Lo resolveremos mediante las pruebas más sencillas del Análisis de varianza, cuya dificultad es mínima aunque



EXOPOL S.L.

Circulares Exopol: www.exopol.com

todas las medias conjuntamente. Nos permitirán discernir si hay diferencias significativas entre todas las medias. Una vez detectadas diferencias entre ellas, en un análisis posterior, que en este artículo no nos ocupa, podremos determinar las diferencias entre cada una de las medias por separado. Este último análisis ya se realiza con programas estadísticos especializados, mediante diferentes técnicas, como podrían ser los llamados tests de Duncan, tests de la Mínima Diferencia Significativa, etc.

Análisis de Varianza de un factor. Prueba F del Análisis de varianza

Una de las singularidades de la estadística es que uno de los métodos para poder comparar medias se denomine Análisis de la Varianza. El motivo es que esta prueba se lleva a cabo comparando dos tipos de variación. El Análisis de Varianza es un método general para estudiar las fuentes de variación en las respuestas. La comparación de varias medias se llama Análisis de Varianza de un factor. Es de un factor debido a que la variable respuesta, en nuestro ejemplo Ganancia Media Diaria de los lechones, sólo se ve influida por otra variable, en nuestro ejemplo, el contenido de harina de pescado en la dieta. La prueba más sencilla del Análisis de Varianza para comparar varias medias se denomina *prueba F*.

El estadístico *F* utilizado para comparar varias medias tiene la siguiente forma:

$$F = \text{variación entre medias muestrales} / \text{variación entre individuos de la misma muestra}$$

El estadístico *F* sólo puede tomar valores positivos o cero. Será cero cuando las medias sean todas iguales. De hecho, el efecto del azar crea algunas diferencias entre las medias muestrales, aún cuando las medias de las poblaciones son iguales. Así, cuando la hipótesis nula es cierta, esperamos que *F* tome valores próximos a uno. A medida que las medias muestrales están más separadas entre sí, el valor de *F* se hace mayor. Los valores grandes de *F* constituyen una prueba fehaciente en contra de la hipótesis nula, inclinándonos a pensar que la hipótesis correcta es la alternativa, la que contempla que alguna de las medias muestrales no es igual a las demás. No hay una mejor explicación que la que se pueda ver en un ejemplo. De esta manera, sigamos el hilo del ejemplo que planteamos en la introducción. Queremos contrastar la hipótesis nula de que no existen diferencias entre la Ganancia Media Diaria de los lechones para las tres dietas con diferente contenido en harina de pescado. Es decir:

- $H_0: \mu_1 = \mu_2 = \mu_3$

La hipótesis alternativa plantea que sí existe alguna diferencia entre las medias, es decir, no todas las medias son iguales:

- $H_a: \mu_1 \neq \mu_2 \neq \mu_3$

La H_a contempla el caso que $\mu_1 = \mu_2$, pero que μ_3 tenga un valor diferente, o cualquier otra combinación. El contraste de H_0 contra H_a se denomina prueba *F* del Análisis de varianza. El Análisis de varianza se abrevia comúnmente en ANOVA (Analysis of Variance).

Análisis de varianza en Excel

Para proseguir con el Análisis de varianza del ejemplo debemos utilizar el programa Excel. Una vez introducidos los datos del ejemplo, entraremos en la opción *Análisis de datos*, que estará debidamente activada (en caso contrario, ver capítulos anteriores), del menú Herramientas. A continuación, seleccionaremos la opción *Análisis de varianza de un factor* como se muestra en la figura 1.



EXOPOL S.L.

Circulares Exopol: www.exopol.com

	A	B	C	D	E	F	G	H
	G.M.D.	G.M.D.	G.M.D.					
1	Dieta 0%	Dieta 4%	Dieta 8%					
2	0,224	0,248	0,265					
3	0,211	0,240	0,187					
4	0,227	0,228	0,289					
5	0,228	0,266	0,250					
6	0,211	0,215	0,286					
7	0,213	0,304	0,198					
8	0,231	0,210	0,199					
9	0,202	0,215	0,275					
10	0,204	0,200	0,294					
11	0,195	0,258	0,277					
12	0,210	0,238	0,284					
13	0,248	0,247	0,200					
14	0,230	0,270	0,264					
15	0,239	0,290	0,235					
16	0,259	0,240	0,319					
17	0,231	0,262	0,273					
18	0,200	0,224	0,298					
19	0,243	0,257	0,299					
20	0,250	0,210	0,268					
21	0,235	0,265	0,221					

Análisis de varianza de un factor [?] [X]

Entrada

Rango de entrada: [X]

Agrupado por: Columnas Filas

Rótulos en la primera fila

Alfa:

Opciones de salida

Rango de salida:

En una hoja nueva:

En un libro nuevo

Aceptar Cancelar Ayuda

Figura 1. Análisis de varianza de un factor
El resultado de realizar esta operación es el siguiente:

	A	B	C	D	E	F	G
1	Análisis de varianza de un factor						
2							
3	RESUMEN						
4	Grupos	Cuenta	Suma	Promedio	Varianza		
5	G.M.D. Dieta 0%	20	4,491945454	0,224597273	0,000332316		
6	G.M.D. Dieta 4%	20	4,888190153	0,244409508	0,000772273		
7	G.M.D. Dieta 8%	20	5,181572813	0,259078641	0,001545718		
8							
9							
10	ANÁLISIS DE VARIANZA						
11	Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados	F	Probabilidad	Valor crítico para F
12	Entre grupos	0,011977819	2	0,00598891	6,779110435	0,002284684	3,15884563
13	Dentro de los grupos	0,050355846	57	0,000883436			
14							
15	Total	0,062333665	59				
16							
17							
18							

Fijémonos principalmente en el estadístico de contraste F, $F = 6,779$. como dijimos anteriormente a mayor valor de F nos daba una idea que la hipótesis nula no era la correcta. Ahora observemos su probabilidad, que es de 0,0022. Mientras el valor de la probabilidad se sitúe por debajo de 0,05 (el nivel de significación tradicionalmente aceptado), podremos afirmar que existe la evidencia de que los tres tipos de dieta no son iguales, o alguna de las tres dietas es diferente a las demás, es decir, afectan de forma diferente a la Ganancia Media Diaria de los lechones. Estamos en disposición de rechazar la hipótesis y aceptar como cierta la hipótesis alternativa.

La prueba F del ANOVA valora la evidencia de que existe alguna diferencia entre las medias poblacionales. En la mayoría de los casos siempre esperamos que esta prueba sea significativa. No emprenderíamos el estudio posterior, comparar una a una cada media, sino encontraríamos diferencias. Si bien parece contradictorio a lo que se ha dicho en el apartado La problemática de las comparaciones múltiples y, debido a que los programas estadísticos no están al alcance de todos, no sería descabellado que, una vez realizada la prueba F, separáramos



EXOPOL S.L.

Circulares Exopol: www.exopol.com

prueba posterior que realizamos tenemos una probabilidad de error del 5% (si comparamos 7 medias dos a dos en al menos una cometeremos el error de aceptar la Hipótesis alternativa cuando no es cierta) .La prueba F nos sirve como precaución para evitar ser engañados por las variaciones debidas al azar.

Análisis de Varianza de varios factores

Como hemos estado viendo en el ejemplo, el único factor que podía perturbar a nuestra variable respuesta, la Ganancia Media Diaria, era el contenido de harina de pescado de la dieta. Este es el caso más sencillo que nos podemos encontrar cuando analicemos unos datos y deseemos comparar medias. Que hubiese sucedido si hubiéramos separado a los lechones en dos mitades y les hubiéramos suministrado dos dietas diferentes en energía, a parte del contenido de harina de pescado. La variable Ganancia Media Diaria ya no sólo estaría influenciada por el factor contenido en harina de pescado, sino que también lo estaría o podría estarlo por la cantidad de energía que aporta la dieta. De hecho, el nivel de factores puede ser muy amplio pero, no se nos escapa, el análisis se complicaría de forma exagerada.

El programa Excel permite hacer un Análisis de Varianza hasta dos factores. En la opción Análisis de datos del menú Herramientas, debemos seleccionar la opción Análisis de Varianza de dos factores con varias muestras por grupo.

La intención de los autores en esta serie de artículos es proporcionar una serie de herramientas estadísticas básicas, que pueden ser llevadas a cabo, con más o menos dificultad, mediante el programa Excel, que está al alcance de todos. De esta forma, en la página www.testsandtrials.com los lectores encontrarán un ejemplo resuelto y debidamente explicado de un Análisis de Varianza para dos factores.