<u>Comparación de dos proporciones realizadas con datos apareados.</u>

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

<u>España</u>

Vimos en el anterior capítulo de esta serie cómo se analizan las proporciones con datos independientes. En este capítulo vamos a conocer cómo se analizan cuando tenemos datos que provienen de datos apareados y posteriormente estudiaremos la concordancia entre resultados u observaciones.

En primer lugar definamos qué son datos apareados. Datos apareados son aquellos datos que provienen de la medición de una variable en el mismo individuo tras la aplicación de un proceso. Los datos apareados más habituales en nuestro trabajo son los datos que provienen de la aplicación de un tratamiento, por ejemplo la medición de la temperatura en un animal antes o después de la aplicación de un antitérmico. Estos datos deben ser procesados de diferente forma a los obtenidos, por ejemplo, en el problema del número anterior ya que las variables son distintas y por tanto la lógica matemática para su análisis debe ser también distinta.

Iniciemos el tema con un ejemplo. Supongamos que queremos verificar la eficacia de un antitérmico en el tratamiento de la fiebre en un lote de terneros de carne. Los resultados tras la aplicación del antitérmico son los que aparecen en la siguiente tabla.

Tabla 1. Temperatura rectal en un lote de terneros frisones tras la aplicación de un antitérmico en el tratamiento de la fiebre (consideramos en este ejemplo que un animal tiene fiebre si su temperatura rectal es superior a 39,5ºC. Este punto de corte se considera sólo a efectos de este análisis estadístico, lo que no implica que sea real)

Crotal	Temperatura día 1	Temperatura día 3
2352	40.2	39.0

2374	39.2	37.8
2384	39.0	39.3
2415	38.4	39.8
2469	38.5	39.3
2470	39.5	39.5
2472	38.5	38.6
2500	39.9	38.8
2514	38.7	37.8
2551	38.8	38.3
2568	39.5	38.5
2602	39.3	39.1
2620	38.7	37.8
2652	39.8	39.5
2658	39.3	39.2
2667	39.7	38.5
2701	40.0	37.8
2724	39.1	39.7
2725	40.1	39.8
2746	40.4	39.7
2760	39.8	39.5
2777	38.4	39.7
2795	38.3	39.8
2796	39.7	38.7
2797	39.7	38.2
2813	40.3	38.6
2816	39.9	39.9
2857	38.2	38.7
2896	39.5	38.2
2903	39.3	38.4
2920	38.9	39.0
2942	39.2	38.7
2946	40.2	39.5
2978	39.3	39.7
2997	38.4	39.1
3016	39.2	39.8
3027	40.1	38.5
3040	38.9	38.2
3052	39.1	38.9
3057	40.4	38.7
3070	38.6	38.4
3096	39.6	38.5
3101	38.6	38.5
3121	40.3	38.9
3150	39.0	39.5
3156	39.3	39.9

3180	38.3	38.6
3224	39.8	39.4
3263	39.8	38.8
3274	38.6	38.3
3282	38.4	38.7
3283	40.1	38.0
3294	39.8	39.0
3331	38.6	39.5
3333	39.7	38.9
3373	40.2	39.2
3393	40.0	38.2
3435	40.5	37.9
3439	39.1	39.7
3452	39.5	39.5

Esta tabla 1 también podemos verla de otra forma (tabla 2):

Tabla 2. Evolución de la fiebre en un lote de terneros frisones tras la aplicación de un antitérmico (consideramos en este ejemplo que un animal tiene fiebre si su temperatura rectal es superior a 39,5ºC. Este punto de corte se considera sólo a efectos de este análisis estadístico, lo que no implica que sea real)

		Fiebre el día 1		
		No	**Sí**	**Total**
Fiebre el día 3	**Sí**	8	3	11
	No	27	22	49
	Total	35	25	60

Para analizar estos datos vamos a seguir la siguiente lógica matemática:

1. Prescindimos de los sujetos que en ambos casos dan resultados concordantes. En nuestro caso los que tenían fiebre el día 1 y el día 3 (3 sujetos) y los que no tenían fiebre ni el día 1 ni el día 3 (27 sujetos).
2. La hipótesis nula que vamos a confrontar es que el tratamiento no mejora la fiebre. La prueba estadística que vamos a aplicar se conoce con el nombre de prueba de McNemar y su cálculo se realiza a partir de la prueba z.

Prueba z de comparación de dos proporciones con datos apareados:

$$z = \frac{d - a}{\sqrt{d + a}} = \frac{22 - 8}{\sqrt{22 + 8}} = 2.556$$

Para una descripción de la nomenclatura de una tabla de contingencia ver el artículo anterior.

Si aceptamos un riesgo _ del 5 % (_ = 0.05) debemos ver si la diferencia de proporciones es significativa. Como nuestro valor z es mayor que el valor $z_{0.025}$ (1.96) podemos concluir con que la diferencia es significativa y nuestro tratamiento es eficaz.

Hay que tener en cuenta que la prueba de McNemar tiene una condición de aplicación que es que los casos discordantes deben sumar más de 10; (a + d) > 10. Si esta condición no se cumple debemos calcular el grado de significación con la prueba P de la Ley Binomial que escapa al ámbito de este artículo.

También estamos interesados en conocer el intervalo de confianza de la diferencia entre las proporciones obtenidas en nuestro estudio de datos apareados.

La estimación de la diferencia y de su intervalo de confianza se obtiene con los siguientes cálculos:

$$p_a = \frac{b + d}{n} = \frac{3 + 22}{60} = \frac{25}{60}$$

$$p_b = \frac{a + c}{n} = \frac{8 + 27}{60} = \frac{35}{60}$$

$$p_a - p_b = \frac{d - a}{n} = \frac{22 - 8}{60} = \frac{14}{60} = 0.2333$$

La estimación del error estándar de esta diferencia;

$$\hat{EE} = \frac{\sqrt{d + a - \frac{(d - a)^2}{n}}}{n} = \frac{\sqrt{22 + 8 - \frac{(22 - 8)^2}{60}}}{60} = 0.0862$$

Si la muestra es grande y los valores a y d no son extremadamente diferentes, los límites del intervalo de confianza se pueden calcular con la Ley Normal. En caso contrario debe aplicarse el cálculo exacto del intervalo de confianza. (En el caso de que algún lector esté interesado en el cálculo por el método exacto de este intervalo de confianza, le rogamos que se ponga en contacto con nosotros a través del e-mail tyt@testsandtrials.com).

En nuestro caso,

$$p_a - p_b = (p_a - p_b) \pm \hat{E}E * z_{á/2}$$

$$p_a - p_b = 0.2333 \pm 0.0862 * 1.96 = (0.0644;0.4023)$$

Este intervalo, lo podemos expresar en porcentajes y decir que el resultado de aplicar el antitérmico problema mejora la fiebre en un 23.33 % de animales con un intervalo de confianza al 95 % de entre 6.44 % al 40.23 %.

Y como no podía ser menos os animamos a realizar el siguiente ejercicio. Tras aplicar a un lote de 60 animales un tratamiento durante 5 días, 10 animales no variaron su condición de enfermos de un total de 20 enfermos el primer día, mientras que 30 animales que no estaban enfermos el primer día permanecieron sin enfermar durante periodo del tratamiento. ¿Es eficaz el tratamiento?. ¿Cuan eficaz es?.