

Capítulo 4: Diferencias que presenta una variable numérica entre dos grupos

En este capítulo veremos procedimientos para contrastar si las diferencias numéricas obtenidas al comparar dos tratamientos (o dos poblaciones) son lo suficientemente grandes como para que su única causa sea atribuible al azar.

Este tipo de pruebas se suelen usar cuando se elige una muestra de individuos que han seguido cierto tratamiento (por ejemplo, un placebo), y otra muestra que ha recibido otro tratamiento (por ejemplo, un fármaco de pruebas).

Dependiendo de cómo se construyan ambas muestras, clasificamos el experimento en dos clases:

- Muestras apareadas: Cuando para cada observación en un grupo tenemos una observación asociada en el segundo grupo.
- Muestras independientes: Cuando puede suponerse que los individuos de un grupo de tratamiento han sido extraídos independientemente de los del otro.

4.1 Muestras apareadas o relacionadas

Aparecen por ejemplo en los estudios de pacientes “antes y después”, es decir, cuando en un grupo tenemos las observaciones de unos pacientes antes del tratamiento, y en el segundo tenemos las observaciones tras el tratamiento. En este caso también decimos que cada individuo es su propio control.

Se deben realizar de modo apareados los estudios en que al medir una variable sospechamos de más fuentes de variabilidad además del tratamiento. Por ejemplo, si creemos que el sexo de los individuos puede influir notablemente en la variable que medimos y queremos anular el efecto tenemos varias posibilidades:

- Hacer un estudio con todos los individuos del mismo sexo en cada muestra, y otro con todos los individuos del otro sexo. A esto se le denomina estratificar por sexos. En este caso lo que tenemos son dos estudios con muestras independientes.
- Asociar a cada individuo del primer grupo, un individuo “que se parezca” del segundo. Por ejemplo, si pensamos que conjuntamente el sexo y la edad pueden influir en el resultado, podemos hacer que por cada individuo del primer grupo, se elija un individuo del segundo, del mismo sexo y de edad similar. A pesar de las dificultades que comporta hacer un estudio con esas características, habremos reducido la posibilidad de que las variables sexo y edad hayan influido mucho en las diferencias observadas entre tratamientos.

En los contrastes con muestras apareadas, la hipótesis nula es que el tratamiento no presenta efecto, frente a la hipótesis alternativa de que sí hay diferencia. Como en todo contraste de hipótesis, se declara que el efecto es estadísticamente significativo si la significación calculada es inferior a cierta cantidad pequeña (5% o 1% típicamente).

Los contrastes se realizan calculando las diferencias existentes entre cada observación de un grupo y la observación asociada en el segundo. Si las mencionadas diferencias tienen una distribución aproximadamente normal o bien la muestra es grande, la prueba

que se suele usar es la **t-student para muestras apareadas**. Esta prueba funciona del siguiente modo. Si se dan las condiciones de validez (diferencias normales), las diferencias deberían ser aproximadamente normales de media cero. Si al calcular la media de las diferencias, el valor obtenido en la muestra no es consistente con una posible media de cero, se rechaza la hipótesis nula. Es decir, si la diferencia entre lo observado y la hipótesis nula no es atribuible al puro azar, aceptamos que hay diferencias entre los grupos.

Una prueba menos exigente es el **contraste no paramétrico de Wilcoxon**. Este último considera las diferencias entre cada observación y su control. Si la hipótesis nula fuese cierta, las diferencias negativas serían similares en cantidad y tamaño a las diferencias positivas. La prueba de Wilcoxon examina la discrepancia existente entre los resultados observados y la predicción de la hipótesis nula.

Ejemplo: Se pretendió comparar la producción de maíz con semillas normales (tipo I) y la de otras semillas que habían recibido un tratamiento especial (tipo II). En un principio se pensó en cultivar dos extensiones de terrenos similares en tamaño y comparar las producciones, pero tras una reflexión se concluyó que desde luego factores como las características del terreno, la humedad, el sol, etc., podían influir en el resultado. Para anular la influencia de estos se decidió realizar el experimento eligiendo múltiples parcelas de terreno, de pequeña extensión, y en cada una de ellas se plantó en una mitad las semillas de tipo I y en otro la semilla de tipo II. Ahora estaba claro que sea cuales sean las posibles causas de variabilidad, los dos tipos de semilla estaban afectados de la misma manera. Por tanto tenemos un diseño apareado.

Prueba de muestras relacionadas

	Diferencias relacionadas				Sig. (bilateral)
	Media	Error típ. de la media	95% Intervalo de confianza para la diferencia		
			Inferior	Superior	
Semilla tipo I - Semilla tipo II	-33,7273	19,95135	-78,1816	10,7271	,122

Un contraste de diferencia de medias usando la prueba t-student para muestras apareadas (relacionadas en SPSS)

muestra por resultado la tabla que se muestra al lado. Lo que leemos en ella se traduce del siguiente modo: La media de las diferencias fueron de 33.7 kilogramos a favor de la semilla de tipo II. Un intervalo de confianza al 95% para la diferencia de medias da un rango de entre 78.1 kg. a favor de la semilla de tipo II hasta 10 kg. para la semilla de tipo I. Esto es compatible con que realmente no haya diferencia entre ambas semillas, con un nivel de significación del 5%. Por último vemos que la significación del contraste es de 0.122=12,2%. Declaramos las diferencias no significativas, para un nivel de significación del 5%.

Estadísticos de contraste^b

	Semilla tipo II - Semilla tipo I
Z	-1,600 ^a
Sig. asintót. (bilateral)	,110

a. Basado en los rangos negativos.

b. Prueba de los rangos con signo de Wilcoxon

Otra posibilidad hubiese sido elegir una prueba no paramétrica para muestras apareadas (relacionadas en SPSS), para contrastar la diferencia entre las dos semillas. En este caso la significación del contraste es bastante similar a la prueba paramétrica (0.11=11%). No puede rechazarse para un nivel

de significación del 5% que entre las dos semillas haya diferencia de producción. Es decir, las diferencias de producción observadas son achacables al azar.

4.2 Muestras independientes

Sirven para tratar problemas como el de si el nivel de hierro es similar en los individuos que padecen determinada enfermedad frente a los individuos sanos. Para ello podríamos elegir una muestra de individuos enfermos y otra de sanos, y comparar si los valores de una muestra tienen tendencia a ser mayores que los de la otra.

En los contrastes con muestras independientes, la hipótesis nula es que los valores obtenidos en una y otra muestra son similares, frente a la hipótesis alternativa de que son diferentes. El valor obtenido en la significación nos permite decidir si se rechaza o no la hipótesis nula.

Hay varias maneras de realizar este tipo de contrastes:

- Prueba t-student para dos medias: Se basa en contrastar si las medias de cada grupo son similares. Este tipo de contrastes es válido cuando se da alguna de las siguientes condiciones:
 - Las varianzas son similares y las observaciones de cada muestra son normales.
 - Las varianzas son similares y los tamaños muestrales son grandes.
 - Hay diferencia notable entre las varianzas de cada grupo, pero los tamaños de cada muestra son similares y además las muestras son grandes o aproximadamente normales.
- Pruebas no paramétricas para muestras independientes: No requieren ningún tipo de suposición sobre la distribución de las muestras. Esto permite que se puedan usar con variables discretas u ordinales. Pruebas populares son las de Mann-Whitney, Wilcoxon o la prueba de Kolmogorov-Smirnov de dos muestras.

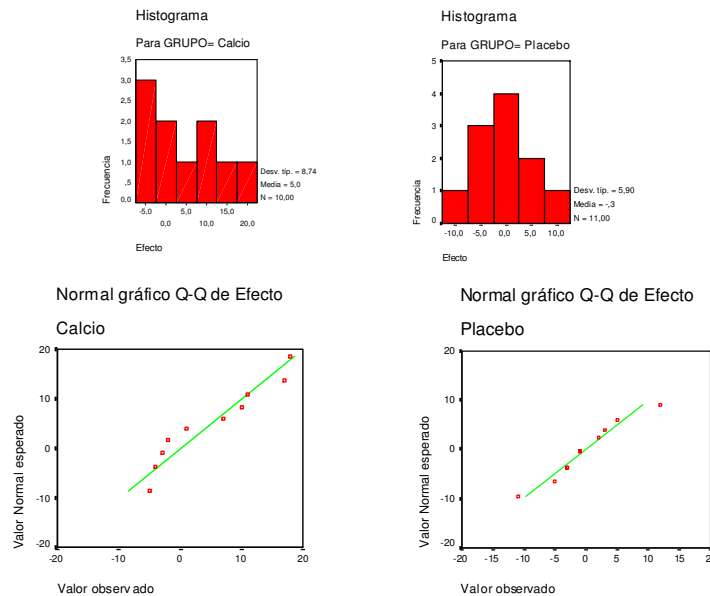
En el caso de que queramos aplicar la prueba t-student, debemos tener en cuenta que esta se ve muy afectada en caso de que las muestras no estén similarmente dispersas, y es necesario hacer una corrección. Debemos comprobar que no existan observaciones anómalas. Disponemos del contraste de Levene para la igualdad de varianzas. De hecho, SPSS lo realiza automáticamente siempre que se desea realizar la prueba t-student, y muestra los resultados para la prueba t-student, tanto en el caso de que el usuario desee admitir la igualdad de varianzas, como rechazarla. Se recomienda al usuario elegir un nivel de significación alto para esta prueba (15% o 20 %). Es tan sensible la prueba t-student sin corrección para desigualdad de varianzas cuando las muestras no son equilibradas, que a la menor indicación debemos considerar que son diferentes.

Descriptivos

Efecto		Estadístico	Error típ.
Placebo	Media	-,27	1,78
	Desv. típ.	5,90	
	Asimetría	,39	,66
	Curtosis	1,50	1,28
Calcio	Media	5,00	2,76
	Desv. típ.	8,74	
	Asimetría	,34	,69
	Curtosis	-1,56	1,33

Ejemplo: Se cree que la ingesta de calcio reduce la presión sanguínea. Para contrastarlo se decidió elegir a 21 individuos de características similares para participar en un estudio. A 10 de ellos elegidos al azar, se les asignó un tratamiento consistente en un suplemento de calcio durante 3 meses y se observó la

diferencia producida en la presión arterial (la que había “antes” menos la que había “después”). A los 11 individuos restantes se les suministró un placebo y se midió también la diferencia. Una exploración descriptiva parece indicar que en los individuos que siguieron el tratamiento con calcio se consiguió una cierta reducción media, mientras que los individuos que tomaron calcio se quedaron prácticamente igual. Usaremos un contraste de hipótesis para ver si esas diferencias pueden ser explicadas o no por el azar. En primer lugar observamos cierta disparidad entre las dispersiones. También se aprecia que la muestra de los individuos que tomaron calcio se aleja ligeramente de la normalidad, aunque la del grupo placebo no.



Prueba de muestras independientes

		Prueba de Levene	Prueba T para la igualdad de medias		
			Sig.	Sig. (bilateral)	95% Intervalo de confianza para la diferencia
					Inferior
Efecto	Se han asumido varianzas iguales	.051	.119	-12,02622	1,48077
	No se han asumido varianzas iguales			.129	-12,25749

En estas condiciones no sería muy adecuado realizar una prueba t-student. Mostramos al lado como ilustración el resultado de la prueba t-student para dos grupos independientes. Se observa que la prueba de Levene de igualdad de varianzas muestra

que estas son bastante diferentes (recordar que el nivel de significación para esta prueba se elige muy superior al 5%). Por tanto, debemos considerar simplemente la significación de la prueba t-student cuando no se asumen varianzas iguales. La significación es de 0.129=12.9%, con lo que las diferencias observadas a favor del tratamiento con calcio podrían ser atribuibles al azar. Seguramente, si se hubiesen escogidos muestras más grandes habríamos obtenido evidencia a favor del tratamiento con calcio, pero como es de esperar, es difícil obtener resultados estadísticamente significativos con muestras pequeñas.

Estadísticos de contraste

	Efecto
U de Mann-Whitney	40,500
Sig. asintót. (bilateral)	.306

Si en su lugar usamos la prueba no paramétrica de Mann-Whitney, que no requiere ninguna condición especial sobre las muestras obtenemos una significación de 0.306=30.6%. Concluimos lo mismo que con la prueba t-

student para un nivel de significación del 5%.