12

Comparación de medias

Irene Moral Peláez

12.1. Introducción

Cuando se desea comprobar si los valores de una característica que es posible cuantificar (como podría ser la edad o la cifra de tensión arterial, entre otras) difieren al agruparlas en dos o más grupos (por ejemplo según género, o por diagnóstico de hipertensión arterial) hablaremos de comparación de medias.

La comparación de medias en un sentido más general, abarca la comparación de los valores de una variable continua según los valores de una variable (o factor) que se puede resumir en dos o más categorías (como el ejemplo expuesto previamente) y que englobaríamos dentro de las pruebas para datos independientes, así como la comparación de los valores de una variable continua evaluada en dos o más momentos en el tiempo (por ejemplo comparar si hay diferencias entre la medición de la presión arterial realizada por la mañana o por la noche) y que englobaríamos dentro de las pruebas para datos apareados.

12.2. Comparación de medias: pruebas para datos independientes

Existen varias pruebas estadísticas que permiten comparar las medias de una variable continua entre dos o más grupos. Cada una de estas pruebas ha sido diseñada para poder ser aplicada cuando se cumplen una serie de supuestos necesarios, bajo diferentes condiciones de aplicación.

Prácticamente todas las hipótesis que podamos plantear (como comparar las medias de una característica entre dos grupos) se pueden analizar bajo una base paramétrica o una base no paramétrica. La decisión de cuándo aplicar una prueba correspondiente a cada uno de estos grupos, depende básicamente de las características inherentes a la variable que deseamos analizar. En general, como se ha indicado en capítulos anteriores, las pruebas paramétricas son más potentes que las pruebas no paramétricas, pero exigen que se cumplan una serie de supuestos como la normalidad en la distribución de la variable, la homocedasticidad (igualdad de varianzas) y la independencia de las observaciones, requiriendo algunas pruebas que se cumplan todo el conjunto de supuestos o alguno de ellos, dependiendo de la prueba a utilizar, sin los cuales, estas pruebas pierden todo su potencial y resulta imprescindible recurrir a sus homólogas no paramétricas. Exceptuando algunas pruebas que exigen más sofisticación y complejidad de las que las pruebas no paramétricas pueden ofrecer, casi todas las pruebas disponibles bajo una base paramétrica, disponen de una homóloga o equivalente bajo una base no paramétrica.

12.2.1. Pruebas paramétricas en la comparación de medias para datos independientes

Si los datos que deseamos analizar cumplen los supuestos necesarios establecidos para poder aplicar pruebas paramétricas, la prueba t de Student para datos independientes o el análisis de la varianza (ANOVA) son las más indicadas para comparar las medias de una característica entre dos o más grupos. La prueba t de Student para datos independientes se utiliza cuando deseamos comparar única y exclusivamente las medias entre dos grupos (por ejemplo deseamos saber si la altura varía según el género), mientras que el ANOVA resulta conveniente cuando deseamos comparar las medias entre más de dos grupos (por ejemplo deseamos saber si las cifras de tensión arterial sistólica difieren según la edad, considerando a los sujetos en 3 grupos de edad: menores de 35 años, entre 35 y 55 años y mayores de 55 años).

12.2.1.1. Prueba paramétrica para datos independientes en la comparación de dos grupos

Como sabemos, en toda prueba existe una hipótesis nula que es normalmente la igualdad de medias, frente a la hipótesis alternativa, que engloba la existencia de un rasgo diferencial entre las medias, es decir, no son iguales. En la prueba t de Student, el estadístico de contraste utilizado para probar la hipótesis nula planteada (las medias de los dos grupos son iguales) se construye en función de las diferencias registradas entre los valores de la variable de estudio evaluada en cada uno de los grupos a comparar. Para ello se utiliza la información procedente de las medias y desviaciones estándar (medidas resumen) de cada uno de los grupos de estudio. El estadístico que se calcula varía ligeramente en base a si las varianzas de los dos grupos en estudio son conocidas, desconocidas pero iguales o desconocidas y distintas. Obviamente, el primer problema a resolver es el de encontrar un método estadístico que nos permita decidir si la varianza en ambos grupos es o no la misma. El test de la razón de varianzas o test de Levene viene a resolver este problema. Bajo la suposición de que las dos poblaciones siguen una distribución normal y tienen igual varianza (H_0: $\sigma_1 = \sigma_2$) se espera que la razón de varianzas siga una distribución F de Snedecor con parámetros (n-1) y (m-1):

$$F = \frac{\frac{1}{n-1}\sum_{i-1}^{n}(X_i - \bar{X})^2}{\frac{1}{m-1}\sum_{i-1}^{m}(Y_i - \bar{Y})^2} = \frac{\hat{S}_1^2}{\hat{S}_2^2}$$

Esta prueba permitirá conocer si las varianzas de los dos grupos de observaciones son o no iguales. Si su p-valor es menor a 0,05, rechazaremos la hipótesis nula y supondremos que la variabilidad en ambos grupos es sustancialmente distinta (varianzas no homogéneas). Por lo tanto, el estadístico a calcular, variará ligeramente en función de las variabilidades muestrales. El estadístico a utilizar sería el siguiente:

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{(n-1)\hat{S}_1^2 + (m-1)\hat{S}_2^2}{n+m-2}}\sqrt{\frac{1}{n}+\frac{1}{m}}}$$

Otro modo de obtener esta misma información es mediante el cálculo de intervalos de confianza para la diferencia de la respuesta media en ambos grupos. El intervalo de confianza constituye una medida de la incertidumbre con la que se estima esa diferencia a partir de la muestra, permitiendo valorar tanto la significación estadística como la magnitud clínica de esa diferencia. En el caso de asumir la misma variabilidad en ambos grupos mediante el test de Levene ($p \geq 0,05$), el intervalo de confianza vendrá dado como:

$$(\bar{X} - \bar{Y}) \pm t^{n+m-2}_{0,975} \sqrt{\frac{(n-1)\hat{S}_1^2 + (m-1)\hat{S}_2^2}{n+m-2}} \sqrt{\frac{1}{n} + \frac{1}{m}}$$

$t^{n+m-2}_{0,975}$ donde denota el valor que según la distribución t de Student con n+m-2 grados de libertad deja a su derecha el 2,5% de los datos. En definitiva, el intervalo de confianza para la diferencia de medias expresará un rango de valores entre los que se puede encontrar el valor real de la diferencia entre los pesos de ambos grupos. Proporciona además la misma información que obteníamos del contraste estadístico. El hecho de que el valor cero pertenezca al intervalo indica que no se dispone de evidencia para concluir que la variable que pretendemos comparar sea distinta en ambos grupos. Por el contrario, si asumimos distintas varianzas en los grupos ($p<0.05$), el intervalo de confianza se expresará tal como se indica a continuación:

$$(\bar{X} - \bar{Y}) \pm t^{f}_{0,975} \sqrt{\frac{\hat{S}_1^2}{n} + \frac{\hat{S}_2^2}{m}}$$

Ejemplo 1. Comparación de dos medias independientes.

Se desea comprobar si la función renal en pacientes hipertensos realmente está más afectada en hombres que en mujeres. Por ello, se ha realizado un estudio en el que se han seleccionado para participar un total de 50 pacientes hipertensos, 24 hombres y 26 mujeres, y se les ha analizado los valores de creatinina sérica. Se han comparado los valores de creatinina sérica entre hombres y mujeres mediante la prueba t de Student, dado que se cumplen los requisitos mínimos básicos para su aplicación.

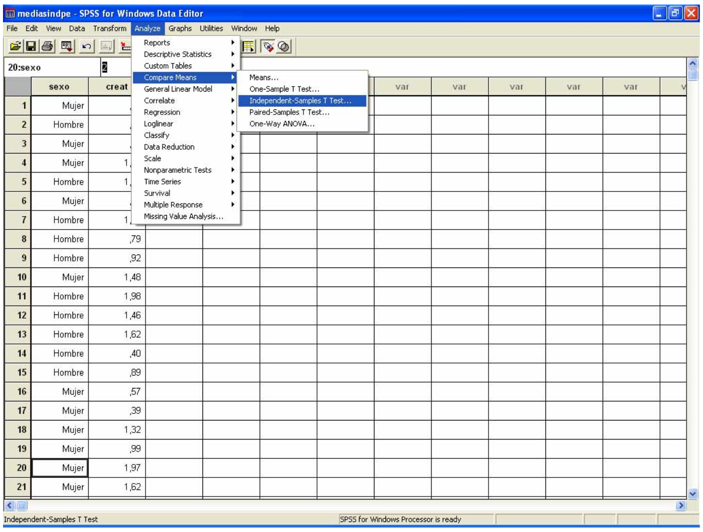
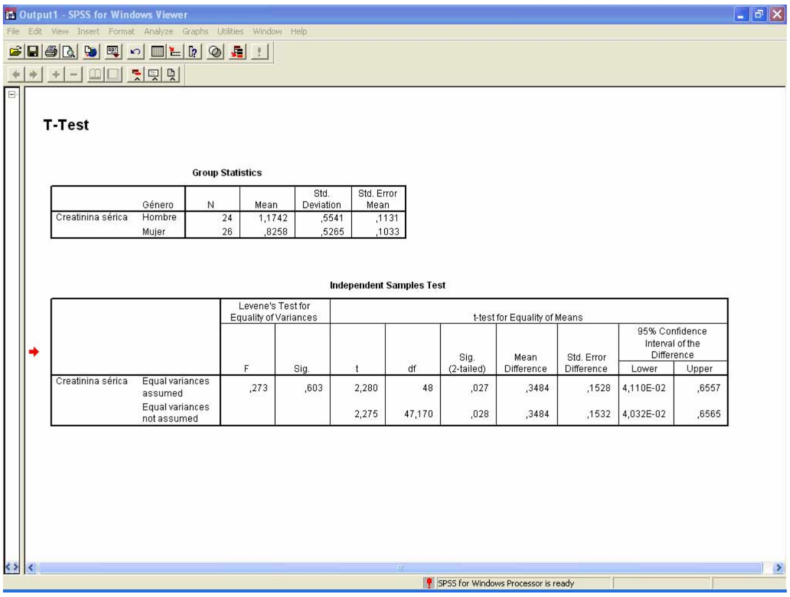


Figura 44. Comparación de dos medias en SPSS. T-Test para datos independientes

Los resultados obtenidos se describen a continuación (Figura 45). La primera tabla nos muestra un resumen descriptivo de la variable creatinina sérica estratificando por género. El valor medio (DE) registrado en hombres fue de 1,2 (0,55) mg/dl, mientras que el valor medio (DE) registrado en mujeres fue de 0,8 (0,53) mg/dl. Según la prueba de Levene, que se muestra en la segunda tabla de la figura, podemos asumir que la hipótesis de igualdad de varianzas es cierta, por lo que si miramos los resultados de la comparación de medias asumiendo igualdad de varianzas, obtenemos un p-valor (la significación estadística correspondiente al estadístico de contraste calculado) resultante de contrastar nuestra prueba de hipótesis de igualdad de medias es de 0,027. A partir de este valor podremos contestar a la siguiente pregunta: ¿Son las medias de creatinina sérica iguales entre ambos sexos? Si a priori definimos que α=0,05 (recordemos del capítulo 10 que corresponde a la región crítica para aceptar o rechazar hipótesis), el p-valor obtenido en la prueba es inferior a este valor, por lo que podemos rechazar la hipótesis nula de que no existen diferencias entre las medias de creatinina sérica según el género. Como se puede observar, en la segunda tabla se muestra adicionalmente, cual es la media muestral de las diferencias, el error estándar y el intervalo de confianza al 95% de la diferencia de medias.

Coincidiendo con el resultado del contraste de hipótesis obtenemos un intervalo de confianza al 95% que no incluye el 0 [0,04-0,66], por lo que aceptamos la existencia de diferencias en los niveles de creatinina sérica según el género de los pacientes.



T-Test

Group Statistics

	Género	N	Mean	Std. Deviation	Std. Error Mean
Creatinina sérica	Hombre	24	1,1742	,5541	,1131
	Mujer	26	,8258	,5265	,1033

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Creatinina sérica	Equal variances assumed	,273	,603	2,280	48	,027	,3484	,1528	4,110E-02	,6557
	Equal variances not assumed			2,275	47,170	,028	,3484	,1532	4,032E-02	,6565

Figura 45. Resultados obtenidos en la comparación de dos medias en SPSS. T-Test para datos independientes

12.2.1.2. Prueba paramétrica para datos independientes en la comparación de más de dos grupos

Por su parte, en el ANOVA, básicamente se considera que la variable analizada depende de un sólo factor, de manera que las causas de su variabilidad se engloban en una componente aleatoria que se denomina error experimental. Esta prueba podemos denominarla como la prueba de contraste de medias general para dos o más grupos, cuyo resultado para dos grupos coincide con la prueba de la t de Student. Para aplicar el ANOVA debemos asumir lo siguiente:

1) las observaciones proceden de poblaciones normales,
2) las muestras (los grupos a comparar) son aleatorias e independientes,
3) dentro de cada muestra (grupo) las observaciones son independientes y, finalmente,
4) homocedasticidad (igualdad de varianzas) de las observaciones.

Más explícitamente, supongamos que deseamos comparar las medias de k grupos entre sí (k>2) en una muestra de tamaño n, para ello asumimos que se trata de k muestras aleatorias, independientes entre sí y extraídas de la misma población normal. La varianza de la población se puede estimar a partir de la varianza dentro de los grupos (MCI), que se calcula como la suma de cuadrados del error (SCI) dividido por los grados de libertad (términos independientes de la suma de cuadrados) y a partir de la varianza entre grupos (MCE). En base al cociente de estos dos estimadores de la varianza de una población normal, que se distribuye según la ley de Fisher-Schnedecor con los grados de libertad del numerador y del denominador, se evalúa la hipótesis de que las k medias provengan de la misma población.

Los resultados de una prueba ANOVA (Analysis of Variance) se expresan como la descomposición de la varianza entre-grupos, intra-grupos y total y se resumen de la siguiente manera:

Fuente de variación	Grados de libertad	Suma de cuadrados		Media de cuadrados	F
Entre-grupos	k-1	SCE	$SCE = \Sigma\ (\breve{y}_i - \breve{y})^2$	MCE= SCE/(k-1)	MCE/MCI
Intra-grupos	(n-1) x k	SCI	$SCI = \Sigma\ (y_{ij} - \breve{y}_i)^2$	MCI= SCI/k x (n-1)	
Total	K x n-1	SCT	$SCT = \Sigma\ (y_{ij} - \breve{y})^2$		

Tabla 14. Análisis de varianza. (ANOVA)

Donde:

1) SCE corresponde a la suma de cuadrados entre grupos y se calcula como el sumatorio del cuadrado de la resta a la media observada en cada uno de los grupos de la media global.
2) SCI corresponde a la suma de cuadrados intra-grupos y se calcula como el sumatorio del cuadrado de la resta a cada observación de la media global observada en su grupo.
3) Y SCT corresponde a la suma de cuadrados total y se calcula como el sumatorio del cuadrado de la resta a cada observación de la media global.

Estas sumas de cuadrados se relacionan entre sí de forma que la suma de cuadrados total es igual a la suma de cuadrados entre-grupos más la suma de cuadrados intra-grupos: SCT = SCI + SCE.

A modo de un primer ejemplo, podríamos asegurar que la altura tiene una variabilidad dada por los distintos sujetos (SCT); sin embargo, al segmentar por género, descubrimos que además de obtener medias distintas en cada grupo, se ha disminuido su variabilidad (SCE), dado que en las mujeres se habrá "eliminado" la variabilidad que aportaban los hombres y viceversa. Por último, tendremos la fuente de variabilidad propia de los sujetos, que es lo que llamamos error aleatorio (SCI). Por tanto, para poder interpretar este capítulo, podemos decir que la variabilidad de la altura ha quedado explicada por el factor género.

Finalmente, comentar que el estadístico de contraste obtenido (F) se distribuye según la distribución de Fischer-Schnedecor con $\alpha_{(k-1\,,\,(n-1)^{*}k)}$ grados de libertad. Si el estadístico F es igual a 1 indica que la variabilidad entre grupos es igual a la variabilidad dentro de cada grupo, por lo que el factor analizado no tiene ninguna influencia en la variabilidad de la muestra. Si por el contrario, F es mayor a 1 con un p-valor asociado menor a 0,05, la variabilidad entre grupos será mayor a la aportada por todas las observaciones individuales, por lo que el factor analizado explicará parte de la variabilidad detectada entre los sujetos.

12.2.2. Pruebas no paramétricas en la comparación de medias para datos independientes

Las pruebas no paramétricas se basan mayoritariamente en la ordenación de las observaciones registradas en la muestra. Para poder cuantificar dicha ordenación se identifican los valores reales registrados en la variable de interés con unos valores llamados rangos. El valor del rango que se asigna a cada una de las observaciones de la muestra es un número natural que oscila entre 1 y n, donde n es el tamaño de la muestra, y que identifica el tipo de ordenación realizada por la prueba en cuestión. La manera en que se asignan los rangos (se realiza la ordenación) dependerá de la hipótesis planteada y por tanto de la prueba no paramétrica que sea necesario utilizar. Por ejemplo, la mecánica de realización de la prueba de Kruskall-Wallis consiste en ordenar todas las observaciones (independientemente del grupo al que pertenezcan) de mayor a menor y asignar los rangos consecutivamente. Posteriormente, se suman los rangos asignados a cada observación agrupándolos para cada uno de los grupos de estudio y se comparan las sumas obtenidas en cada uno de los grupos mediante un estadístico de contraste (fórmula no mostrada), evaluando su valor respecto a la ley de distribución de la Ji cuadrado con k – 1 grados de libertad, donde k indica el número de grupos que se están comparando.

Si los datos que deseamos comparar no cumplen los requisitos para poder aplicar las pruebas paramétricas correspondientes, existe la posibilidad de aplicar pruebas no paramétricas equivalentes que no requieren que los datos cumplan todos los supuestos establecidos para sus homólogas paramétricas. La prueba análoga a la prueba t de Student para datos independientes en su versión no paramétrica es la prueba U de Mann-Whitney, mientras que la prueba análoga al ANOVA en su versión no paramétrica es la prueba de Kruskall-Wallis.

La mecánica de funcionamiento de la prueba U de Mann-Whitney consiste, a grandes rasgos, en calcular un estadístico que se distribuye según una normal de media 0 y desviación 1. Para calcular dicho estadístico se compara cada una de las observaciones de uno de los grupos a comparar con todas las observaciones del otro grupo, asignando un valor de 1 en caso de que la observación del primer grupo sea superior a la observación del segundo grupo, un valor de 0.5 en caso de empate o un valor de 0 en otro caso. Tras la comparación de todas las observaciones se suman los valores obtenidos y se calcula dicho estadístico para poder probar la hipótesis planteada.

Ejemplo 2: Comparación de más de dos medias independientes (k-medias).

Se desea evaluar la influencia que el tipo de diálisis tiene en la concentración de eritropoyetina en suero en pacientes con IRC. Se han seleccionado pacientes con IRC que se sometían a hemodiálisis (n=30), diálisis peritoneal intermitente (n=35) o diálisis peritoneal ambulatoria continua (n=35) y se han registrado los valores de eritropoyetina en suero de la muestra. Para comparar la hipótesis de que no existen diferencias estadísticamente significativas entre los 3 tipos de técnicas de diálisis, necesitamos utilizar una prueba que permita comparar más de dos grupos. Puesto que no se cumplen los criterios para poder aplicar una prueba paramétrica (ANOVA), deberemos aplicar la prueba no paramétrica H de Kruskall-Wallis, que es la prueba análoga a la generalizada y que su realización con SPSS se muestra en la Figura 46.

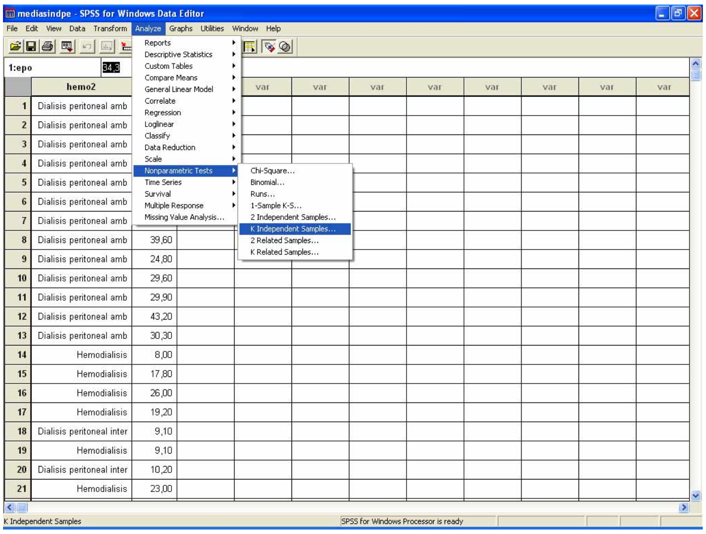


Figura 46. Comparación de k medias independientes en SPSS. Test de Kruskall-Wallis

Los resultados obtenidos se muestran en la Figura 47. En la primera tabla de resultados se muestra un análisis descriptivo de los datos que estamos analizando. Observamos que estamos comparando 3 grupos independientes entre sí (hemodiálisis, diálisis peritoneal intermitente y diálisis peritoneal ambulatoria continua), indicando el número de observaciones que forman cada grupo (n=30, n=35 y n=35, respectivamente) y de la media de los rangos asignados, según el funcionamiento de la prueba. La media de los rangos es superior en los pacientes que reciben diálisis peritoneal ambulatoria continua, seguida de los pacientes que reciben diálisis peritoneal intermitente y finalmente de los pacientes en hemodiálisis. Para comprobar si estos grupos presentan diferencias estadísticas tal como se nos ocurre a priori, en la segunda tabla se muestra el valor del estadístico de contraste, los grados de libertad de la distribución y la significación (el p-valor) que resulta en nuestro caso de 0,000, inferior a 0,05, valor umbral que hemos fijado como error tipo I (o alfa), por lo que podemos rechazar la hipótesis nula de que el valor de eritropoyetina en suero no difiere significativamente según el tipo de diálisis a la que se somete el paciente con enfermedad renal terminal. Así pues, podemos afirmar que se han encontrado diferencias estadísticamente significativas entre tipos de diálisis en los valores de eritropoyetina en sangre.

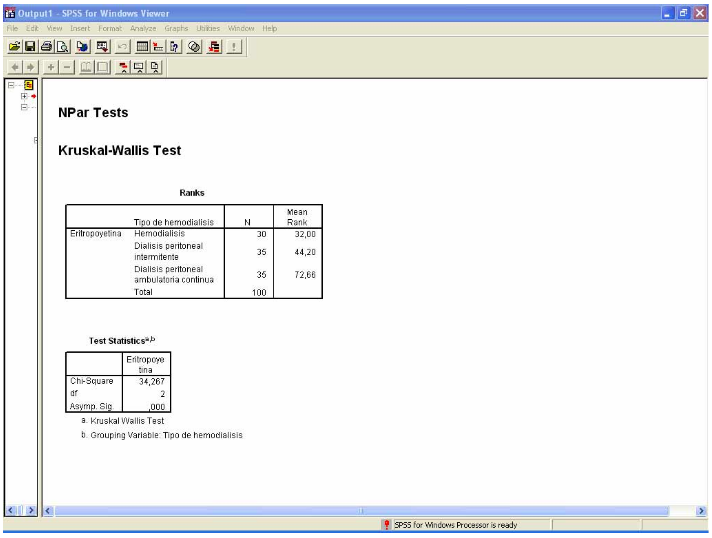


Figura 47. Resultados obtenidos en la comparación de k medias independientes en SPSS. Test de Kruskall-Wallis

12.3. Comparación de medias: pruebas para datos apareados

Imaginemos que deseamos comprobar si las cifras de tensión arterial sistólica varían significativamente si son registradas antes o después de comer, porque vamos a realizar un estudio y debemos determinar en qué momento recogeremos esta información. Para realizar la prueba de hipótesis escogeríamos una muestra de pacientes hipertensos a los cuales mediríamos la cifra de TAS antes de comer y les volveríamos a hacer idéntica medición a los mismos sujetos después de comer. Los valores de TAS estarían identificados para cada paciente, por lo que compararíamos si las cifras de TAS de cada individuo de la muestra difirieren significativamente entre los dos momentos del tiempo en los que fueron registradas, teniendo en cuenta que la información en ambas ocasiones procede del mismo individuo, es decir que existe una variabilidad intra-individual, inherente al propio sujeto, que hay que considerar al realizar las comparaciones, no pudiendo asumir la independencia de las observaciones. Estas pruebas generalmente son aplicables en los estudio "pre-post" tratamiento, en los que es necesario conocer la evolución clínica de parámetros que pueden verse alterados por la administración de tratamientos farmacológicos o distintas terapias.

Al igual que en el caso anterior, existen unas condiciones de aplicabilidad, bajo las cuales es adecuado realizar pruebas paramétricas, debiendo aplicar en caso contrario, pruebas análogas no paramétricas, que no necesiten cumplir dichos supuestos.

12.3.1. Pruebas paramétricas en la comparación de datos apareados

La base de las pruebas para la comparación de medias apareadas consiste en analizar las diferencias entre las observaciones de un mismo individuo. Suponiendo que la variable aleatoria que define la diferencia entre dos observaciones registradas en un mismo individuo (modelo antes-después) fuera una variable aleatoria que se distribuyera normalmente, y queremos contrastar la hipótesis de que se produjo un efecto entre ambas observaciones (cambio). En el caso de resultar cierta, el estadístico de contraste que utilizaríamos se distribuiría según la ley de probabilidad de la t de Student, por lo que la prueba que resultaría más adecuada sería la prueba paramétrica de la t de Student para datos apareados.

El estadístico de contraste desarrollado a partir del planteamiento de la hipótesis a contrastar es:

$$t = \frac{\bar{d}}{\hat{S}_d}\sqrt{n}$$

donde:

1) d: media muestral de la diferencia entre las observaciones "pre" y "post"
2) n: tamaño de la muestra
3) Sd: desviación estándar muestral de las diferencias
4) t_{n-1}: ley de probabilidad de la t de Student con n-1 grados de libertad

El cálculo del intervalo de la diferencia de medias al 95% de confianza, responde a la siguiente fórmula:

$$\left(\bar{d} \pm t_{0975}^{n-1} \frac{\hat{S}_d}{\sqrt{n}}\right)$$

En el caso que hemos mencionado en la introducción, la pregunta sería: ¿existen cambios en las cifras de TAS después de comer? Intuitivamente, podemos apreciar que lo realmente importante en este caso será analizar lo que incrementa o disminuye la cifra de TAS, independientemente del valor inicial o final. Por tanto, si no existen diferencias, el valor de antes y el de después serán iguales (o variarán muy poco) y la resta de sus valores será cercana a cero.

12.3.2. Pruebas no paramétricas en la comparación de datos apareados

A continuación se explicará cómo llevar a cabo el análisis de datos en el caso de comparar variables continuas entre 2 o más observaciones, teniendo en cuenta datos apareados, utilizando medidas no paramétricas. Se mostrarán dos ejemplos realizados mediante el paquete estadístico SPSS. El primer ejemplo corresponderá al caso de comparación de 2 observaciones con pruebas no paramétricas y el segundo, utilizando la comparación de k observaciones apareadas con una prueba no paramétrica.

12.3.2.1. Pruebas no paramétricas para datos apareados en la comparación de dos grupos

La prueba de contraste de hipótesis análoga, en su versión no paramétrica es la prueba de los rangos con signo de Wilcoxon. Básicamente, la prueba consiste en ordenar las diferencias de menor a mayor y obtener sus rangos respectivos. A continuación se suman los rangos correspondientes a las diferencias negativas y a las diferencias positivas, es decir cuando la primera observación es mayor que la segunda, y a la inversa, cuando la segunda observación es mayor a la primera. Una vez construido el estadístico de contraste se evalúa a partir de las tablas de Wilcoxon si se encuentra dentro de la región crítica, para decidir si se acepta la hipótesis nula (no hay diferencias en las observaciones apareadas) o se rechaza (si las hay).

Ejemplo 3: Comparación de dos observaciones apareadas.

Un individuo hipertenso, con función renal y proteinuria normal puede ser donante de riñón. Se desea evaluar si se registran cambios en la Tensión Arterial Sistólica (TAS) en los pacientes hipertensos, con función renal y proteinuria normales al año de haber donado un riñón. Para ello, se registraron los valores de TAS de estos pacientes antes de haber donado el riñón y al año de seguimiento. Tras comprobar que no se cumplían los supuestos necesarios para poder aplicar la prueba paramétrica de primera elección (prueba t de Student para datos apareados), se compararon ambos valores mediante la prueba de Wilcoxon para datos apareados. La Figura 48 muestra los pasos a seguir para realizar la prueba de Wilcoxon utilizando el paquete estadístico SPSS.

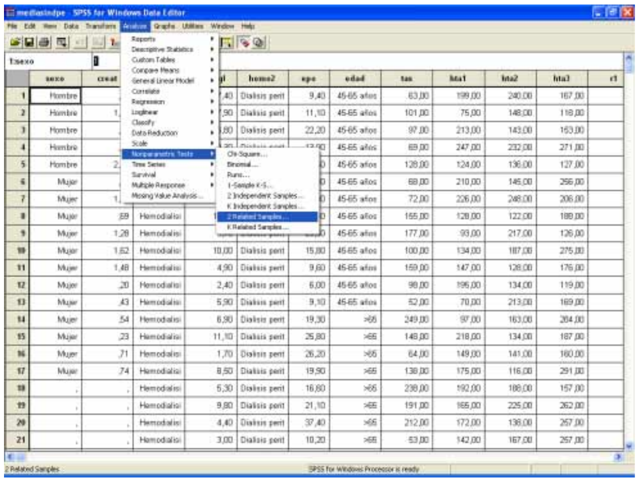


Figura 48. Comparación de 2 medias para datos apareados en SPSS. Test de Wilcoxon

Los resultados obtenidos tras la aplicación de la prueba correspondiente, se muestra en la Figura 49. En el estudio realizado no se observaron diferencias estadísticamente significativas respecto a las cifras de tensión arterial al año de haber donado un riñón en pacientes hipertensos, con función renal y proteinuria normales. El valor de la significación del estadístico Z es de 0,790 con lo que, al no ser superior a 0,05, no podemos rechazar la hipótesis nula y, por tanto, no podemos considerar que se hayan detectado diferencias entre los valores "pre" y "post".

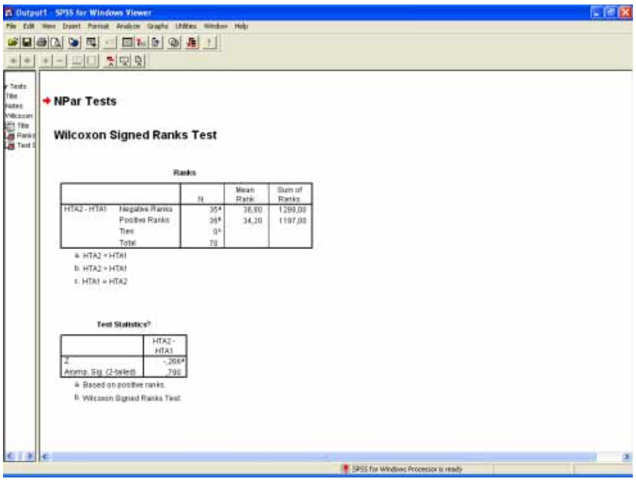


Figura 49. Resultados obtenidos en la comparación de 2 medias para datos apareados en SPSS. Test de Wilcoxon

12.3.2.2. Pruebas no paramétricas para datos apareados en la comparación de más de dos grupos

En caso de querer comparar los valores de una variable continua registrada en más de dos momentos de tiempo la prueba paramétrica idónea para realizar el contraste de hipótesis sería el modelo lineal general (GLM) de medidas repetidas, prueba basada en el análisis de la varianza que no abordaremos en este libro debido a su mayor complejidad, mientras que su análoga no paramétrica sería la prueba de Friedman.

La prueba de Friedman consiste en asignar rangos a las observaciones de un mismo individuo, es decir si existen 4 mediciones de una misma observación para un mismo individuo registradas en diferentes momentos del tiempo, se ordenarán las 4 observaciones para cada individuo y se les asignará un rango que oscilará entre 1 y 4. Bajo la hipótesis de que no hay variación entre las 4 mediciones, la suma de los rangos de todos los individuos entre las 4 mediciones no debería diferir.

Ejemplo 4. Comparación de más de dos observaciones apareadas.

Se ha comprobado que un individuo hipertenso, con función renal y proteinuria normal donante de riñón no presenta fluctuaciones estadísticamente significativas en los valores de TAS hasta un año después de haber donado el órgano, pero no se dispone de información a más largo plazo. Un estudio plantea la hipótesis de que a los 5 años de haber donado el órgano y debido al proceso de envejecimiento, entre otros factores, si es posible que los valores de la TAS se muestren significativamente alterados, respecto a los valores iniciales y a los valores registrados al año de la intervención. En un estudio realizado se registraron los valores de TAS en pacientes hipertensos, con función renal y proteinuria normales antes de donar un riñón, al año y a los 5 años de seguimiento y se compararon las observaciones registradas mediante la prueba de Friedman, dado que los supuestos necesarios para poder aplicar el modelo lineal general (GLM) de medidas repetidas no se cumplieron.

La Figura 50 nos muestra los pasos a seguir para realizar la prueba de Friedman en SPSS. Se puede apreciar que HTA1 es la variable que indica el valor de TAS previa a la donación, HTA2 al cabo de un año y HTA3, a los 5 años de seguimiento.

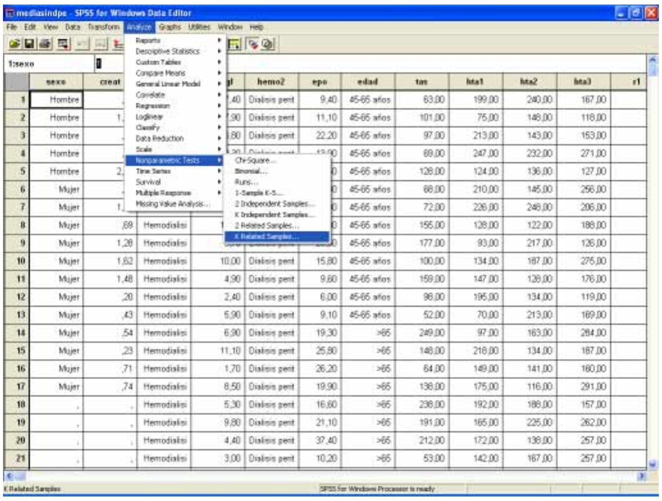


Figura 50. Comparación de k medias de datos apareados en SPSS. Test de Friedman

Los resultados obtenidos se describen en la Figura 51: Las cifras de tensión arterial en pacientes hipertensos, con función renal y proteinuria normales fueron significativamente diferentes entre el momento previo a la donación del riñón, al año de la intervención y a los 5 años. Si observamos la tabla resumen de la asignación de rangos, el rango medio para la variable HTA3 (que corresponde a los valores de tensión arterial a los 5 años de seguimiento) se distancia respecto a los rangos medios registrados en las variables HTA1 (valor de la tensión arterial en el momento previo a la intervención) y HTA2 (valor de la tensión arterial al año de la intervención), lo que nos hace suponer que a los 5 años de seguimiento las cifras de tensión arterial aumentan significativamente con una p=0,037, que es menor a nuestro alfa de 0,05.

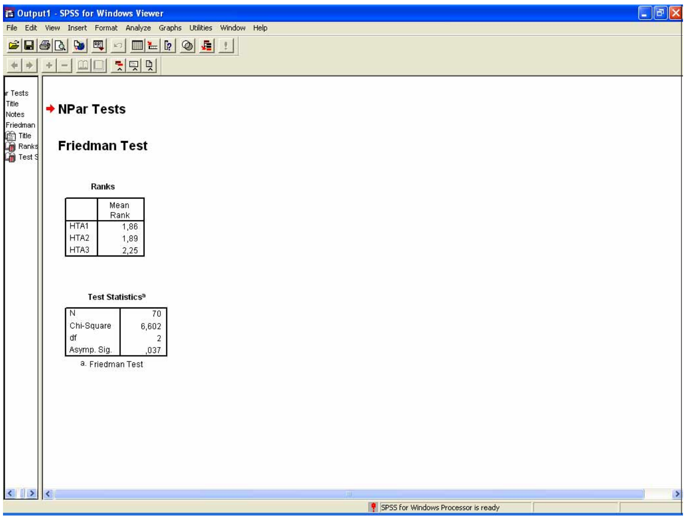


Figura 51. Resultados obtenidos en la comparación de 2 medias para datos apareados en SPSS. Test de Friedman

12.4. Pruebas post-hoc o a posteriori

Si en el ANOVA o en el procedimiento GLM un factor resulta significativo, puede interesar conocer a posteriori cuáles son los niveles que pueden considerarse iguales y cuáles distintos. En el caso de tener dos niveles no es importante, puesto que si el estadístico tiene una significación asociada menor a nuestro valor crítico, entonces consideraremos que existen diferencias entre las medias de los dos grupos. Sin embargo, cuando existen más de dos grupos, el ANOVA sólo nos indica que se han encontrado diferencias entre los grupos y que nuestra variable de agrupación o factor es significativa, con lo que sabemos que existen diferencias pero no sabemos dónde.

Existen varios tests que permiten conocer cuáles son los grupos que difieren entre sí, pero las técnicas de comparación son muy variadas. Lo más importante es conocer algunos de estos métodos que podemos encontrar en los paquetes estadísticos y la conveniencia de reajustar nuestro valor alfa. El hecho de realizar pruebas post-hoc indica que deberemos hacer múltiples comparaciones entre los niveles del factor a comparar, lo cual implica una multiplicidad de pruebas que puede hacer que aumentemos el riesgo de considerar dos niveles distintos cuando sean iguales (Error alfa). Es por eso que debe realizarse un ajuste de alfa a valores cercanos a 0,01, que nos dé más seguridad al considerar dos medias iguales.

Algunos de los estadísticos más conocidos son los siguientes: LSD (Least Significant Difference), Bonferroni, Scheffé, S-N-k (Student-Neuwman-Keuls), Tukey, Duncan, Dunnet, entre otros. Muchos paquetes estadísticos han incorporado estos contrastes post-hoc en el análisis de la varianza. En este manual no profundizaremos más en el tema, pero es importante saber que son procedimientos que existen y que son necesarios para conocer qué diferencias tenemos cuando existen más de dos grupos de una variable.

12.5. Consideraciones importantes

En muchos estudios y en la mayoría de los ensayos clínicos, es necesario comparar ciertas características en dos o más grupos de sujetos. La elección de un método de análisis apropiado en este caso dependerá de la naturaleza de los datos y la forma en la que estos hayan sido obtenidos, es decir, si se trata de un estudio transversal o longitudinal a lo largo del tiempo. Otro aspecto a tener en consideración será el tipo y distribución de los datos. Para grupos independientes, los métodos paramétricos requieren que las observaciones en cada grupo provengan de una distribución aproximadamente normal con una variabilidad semejante, de modo que si los datos disponibles (o alguna transformación simple de ellos) no verifican tales condiciones, se debería recurrir a la utilización de procedimientos no paramétricos.

Pruebas para muestras independientes. Estudios transversales.

Como se ha comentado, la aplicación de un contraste paramétrico requiere la normalidad de las observaciones para cada uno de los grupos, aunque un número suficiente de observaciones (digamos mayor de 30) justifica, no obstante, la utilización de un test paramétrico. Cabe especificar que, siempre que se pueda y se cumplan las condiciones, se deben aplicar pruebas paramétricas puesto que tienen asociados intervalos de confianza que permitirán conocer cuánto puede alejarse nuestra diferencia de medias muestrales de la diferencia de medias poblacionales.

En caso de tener dos muestras en las que se pueden aplicar pruebas paramétricas, se desarrollará la t de Student, con las distintas vertientes, sean las variabilidades de las muestras iguales o estadísticamente distintas. Sin embargo, en caso de tener más de dos grupos de estudio, el ANOVA será la prueba que deberemos aplicar.

Por otro lado, las pruebas no paramétricas que pueden utilizarse corresponderán a la U de Mann-Whitney en el caso de querer comparar dos grupos y la prueba H de Kruskall-Wallis en caso de tener más de dos grupos.

Pruebas para muestras dependientes. Estudios longitudinales.

Las series dependientes surgen normalmente cuando se evalúa un mismo dato más de una vez en cada sujeto de la muestra (estudios longitudinales). También se puede encontrar este tipo de observaciones en estudios de casos y controles donde cada caso se aparea individualmente con un control.

Las pruebas que se utilizan sobre la diferencia entre observaciones en caso de poder aplicar pruebas paramétricas son: la t de Student para datos apareados en caso de tener dos momentos, o bien, el procedimiento ANOVA para medidas repetidas (GLM) en caso de medir la variable en dos o más momentos en el tiempo. Análogamente, las pruebas no paramétricas para evaluar la media de las diferencias entre dos momentos corresponde a la prueba de rangos con signo de Wilcoxon y, en general, para más de dos momentos, la prueba de Friedman.

En la siguiente tabla se muestra esquemáticamente a qué prueba estadística podemos recurrir cuando deseamos comparar dos o más medias, evaluando la situación global de los datos a comparar.

	Comparación de 2 medias		Comparación de > 2 medias	
	Pruebas paramétricas	Pruebas no paramétricas	Pruebas paramétricas	Pruebas no paramétricas
Transversalmente (en un momento del tiempo) Datos independientes	T de Student para datos independientes	U de Mann -Whitney	ANOVA	H de Kruskall Wallis
Longitudinalmente (a lo largo del tiempo) Datos apareados	T de Student para datos apareados	Prueba de los rangos con signo de Wilcoxon	GLM para medidas repetidas	Prueba de Friedman

Tabla 15. Resumen de pruebas para la comparación de medias.