

EVIDENCIAS EN PEDIATRÍA

Toma de decisiones clínicas basadas en las mejores pruebas científicas
www.evidenciasenpediatria.es

Fundamentos de medicina basada en la evidencia

Pruebas para muestras relacionadas. Variables cuantitativas

Ortega Páez E¹, Ochoa Sangrador C², Molina Arias M³

¹UGC Góngora. Distrito Granada-Metropolitano. Granada. España.

²Servicio de Pediatría. Hospital Virgen de la Concha. Complejo Asistencial de Zamora. Zamora. España.

³Servicio de Gastroenterología. Hospital Universitario Infantil La Paz. Madrid. España.

Correspondencia: Eduardo Ortega Páez, edortegap@gmail.com

Palabras clave en español: estadística; inferencia estadística; muestras relacionadas.

Palabras clave en inglés: statistics; statistical inference; paired samples.

Fecha de recepción: 14 de noviembre de 2022 • **Fecha de aceptación:** 24 de noviembre de 2022

Fecha de publicación del artículo: 7 de diciembre de 2022

Evid Pediatr. 2022;18:41.

CÓMO CITAR ESTE ARTÍCULO

Ortega Páez E, Ochoa Sangrador C, Molina Arias M. Pruebas para muestras relacionadas. Variables cuantitativas. Evid Pediatr. 2022;18:41.

Para recibir Evidencias en Pediatría en su correo electrónico debe darse de alta en nuestro boletín de novedades en <http://www.evidenciasenpediatria.es>

Este artículo está disponible en: <http://www.evidenciasenpediatria.es/EnlaceArticulo?ref=2022;18:41>.

©2005-22 • ISSN: 1885-7388

Pruebas para muestras relacionadas. Variables cuantitativas

Ortega Páez E¹, Ochoa Sangrador C², Molina Arias M³

¹UGC Góngora. Distrito Granada-Metropolitano. Granada. España.

²Servicio de Pediatría. Hospital Virgen de la Concha. Complejo Asistencial de Zamora. Zamora. España.

³Servicio de Gastroenterología. Hospital Universitario Infantil La Paz. Madrid. España.

Correspondencia: Eduardo Ortega Páez, edortegap@gmail.com

Hasta ahora hemos visto las pruebas estadísticas que relacionaban variables de dos muestras independientes. Es decir, que cada sujeto se observaba una sola vez y cada observación en un grupo no guardaba relación con ninguna observación del otro grupo. En el presente capítulo abordaremos el estudio de la comparación entre variables con muestras relacionadas, también llamadas dependientes, apareadas o emparejadas. Cuando el emparejamiento se realiza por medidas repetidas de un mismo individuo, hablamos de autoemparejamiento, medidas repetidas o de cambio. Si el emparejamiento se produce de forma natural, por ejemplo, una comparación entre gemelos, tenemos el emparejamiento natural. Si el emparejamiento se realiza mediante el diseño epidemiológico, asignando controles con las mismas características (sexo, edad) que los casos, hablamos de emparejamiento artificial, propio de los estudios de casos y controles. Así mismo, podemos clasificar estas pruebas en dos grandes bloques: por el tipo de variable, en cuantitativas o categóricas; y según su aproximación a la distribución normal, en paramétricas y no paramétricas (**Figura 1**). Son más eficientes que las pruebas con muestras independientes, ya que con menor tamaño muestral se obtiene mayor potencia, porque tienen menor variabilidad al ser comparaciones entre los mismos individuos.

En este capítulo veremos la comparación entre muestras relacionadas con variables cuantitativas, realizadas en varias mediciones en distintos momentos (medidas repetidas), también llamados diseños intrasujeto.

PRUEBAS PARAMÉTRICAS PARA MUESTRAS RELACIONADAS

Comparación de dos medias para muestras relacionadas. Pruebas T para muestras relacionadas

Estas pruebas ya se comentaron brevemente en el capítulo de comparación de medias, pruebas T de Student. Recordemos que tenemos que comparar dos medias en dos muestras que se han realizado en los mismos sujetos en momentos diferentes. Las variables que valoramos son las diferencias entre las parejas de datos en los dos momentos de la medición, bajo la hipótesis nula de que la diferencia es igual a 0. Los supuestos de la prueba son que las variables sean cuantitativas continuas, que las muestras de pares de datos se hayan obtenido al azar y que la diferencia entre las parejas se distribuya de forma normal.

Para ilustrar la explicación, veamos un ejemplo utilizando un programa de acceso libre, el software estadístico R (<https://www.r-project.org/>) con el plugin RCommander (<https://estadistica-dma.ulpgc.es/cursosR4ULPGC/12-Rcommander.html>) y la base de datos ficticia fund_med_rep, disponible en la web de *Evidencias en Pediatría* (https://evidenciasenpediatria.es/files/43-227-RUTA/fund_med_rep.RData). Si necesita saber cómo instalar RCommander, puede consultar el siguiente tutorial en línea (http://sct.uab.cat/estadistica/sites/sct.uab.cat.estadistica/files/instalacion_r_commander_0.pdf). El ejercicio se ha realizado con la versión de RStudio Versión 1.1.463 para mac y RCommander v 2.7.1 en inglés.

Figura 1. Pruebas estadísticas para muestras apareadas

	Variables cuantitativas		Variables categóricas	
Pruebas paramétricas	Dos muestras	Pruebas T para muestras dependientes	Dos muestras	Prueba de McNemar
	Más de dos muestras	ANOVA para medidas repetidas		
Pruebas no paramétricas	Dos muestras	Prueba de Wilcoxon	Más de dos muestras	Prueba de la Q de Cochran
	Más de dos muestras	Prueba de Friedman		

Una vez cargado RCommander, para cargar la base de datos basta con ir a la pestaña de *Data* (Datos) → *Load data set* (Cargar base de datos). Se abre una caja, buscamos la base de datos en nuestro ordenador y la cargamos. Seguidamente, en la pestaña *No active data set* (base de datos no activas), la buscamos y la cargamos.

En la base de datos se recogen los registros de la calprotectina fecal de diez pacientes afectos de alergia a las proteínas de leche de vaca no mediada por IgE (APLV) al inicio del diagnóstico, al mes y a los tres meses desde la exclusión de las proteínas de leche de vaca (PLV).

En primer lugar, comprobamos los supuestos de aplicación de la prueba. Los dos primeros se cumplen claramente porque son variables cuantitativas continuas y las muestras son recogidas al azar. Para ello hemos creado la variable *Calprob_1*, que es el resultado de las diferencias entre la calprotectina fecal basal y al mes del inicio del tratamiento. La comprobación del supuesto de la normalidad de la distribución de las diferencias de las medias se hace, como ya se ha visto con anterioridad, de forma gráfica con el gráfico de caja, el de comparación de cuantiles y por el test de normalidad, en este caso, el Shapiro-Wilk, que es más robusto cuando las muestras son pequeñas (**Figura 2**). Podemos observar, en primer lugar, en el gráfico de cajas, que la variable es bastante simétrica y no existen grandes valores atípicos; en la gráfica de cuantiles las diferencias de las medias de ambas variables es-

tán muy próximas a la diagonal que representaría la diferencia de medias igual a 0; y, por último, el test de normalidad de Shapiro-Wilk nos muestra una $p > 0,5$ ($p = 0,62$), no pudiendo rechazar el supuesto de normalidad.

El segundo paso es realizar la prueba. Desplegamos la pestaña *Statistics* (estadísticos) → *Means* (medias) → *Paired t-test* (prueba t para muestras apareadas) (**Figura 3**). Después, elegimos las variables relacionadas a contrastar, en este caso *Calpro_basal* y *Calpro_1* mes. Nuestra hipótesis es que el tratamiento con dieta de exclusión de PLV ha disminuido los valores de calprotectina fecal; en este supuesto podemos elegir una hipótesis alternativa unilateral cuya diferencia de medias sea mayor de 0. En la pestaña *Options* (opciones) elegimos *Difference > 0* y clicamos en OK. En la ventana de resultados obtenemos el valor de $t = 7,65$ (un valor muy alto, lo que quiere decir que se aleja mucho de la hipótesis nula), con 9 grados de libertad ($df = 9$) y un valor de $p = 0,00001576$, rechazando por tanto la hipótesis nula de igualdad de medias. R también nos recuerda la hipótesis alternativa unilateral, que el verdadero valor de la diferencia de medias es > 0 , con una diferencia media de 275,1 y un intervalo de confianza al 95% de 209,1 a infinito, puesto que no incluye el 0 sería una diferencia significativa. La interpretación sería que, con una probabilidad muy pequeña de equivocarnos, los niveles de calprotectina fecal disminuyen tras un mes de dieta de exclusión (**Figura 3**). Es preciso advertir que el intervalo de confianza es muy amplio, debido al reducido tamaño muestral.

Figura 2. Comprobación de los supuestos de normalidad de la diferencia entre la calprotectina basal y al mes

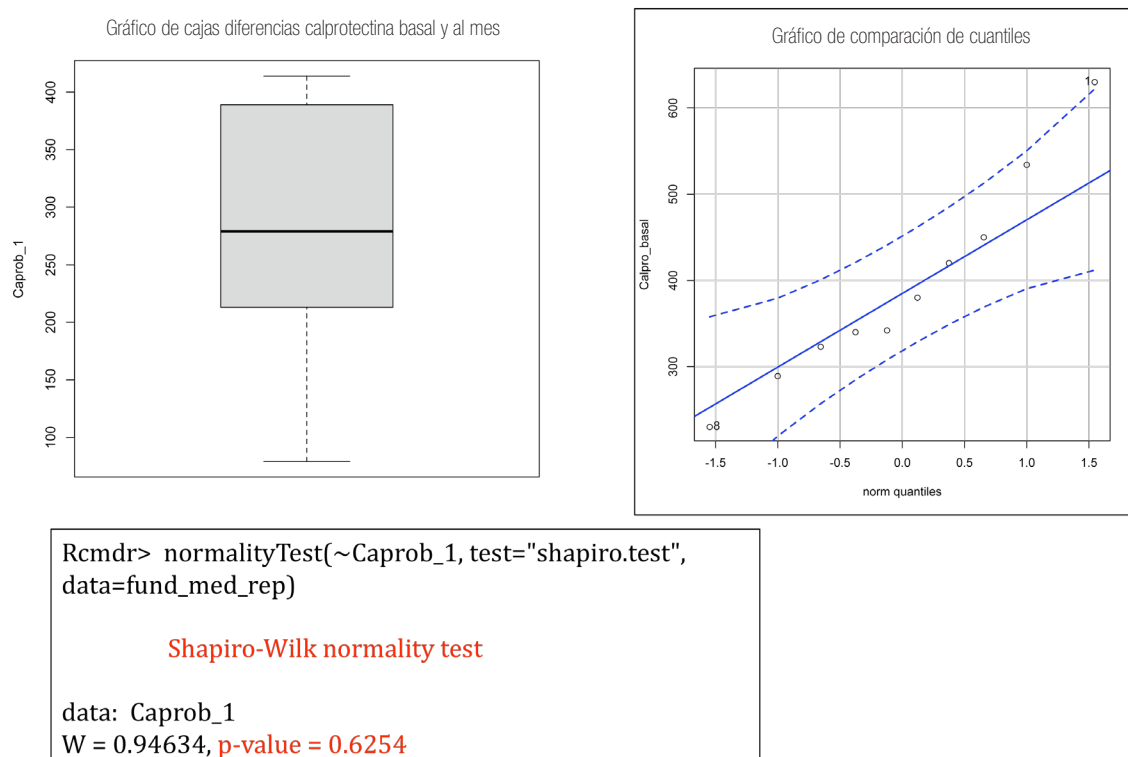
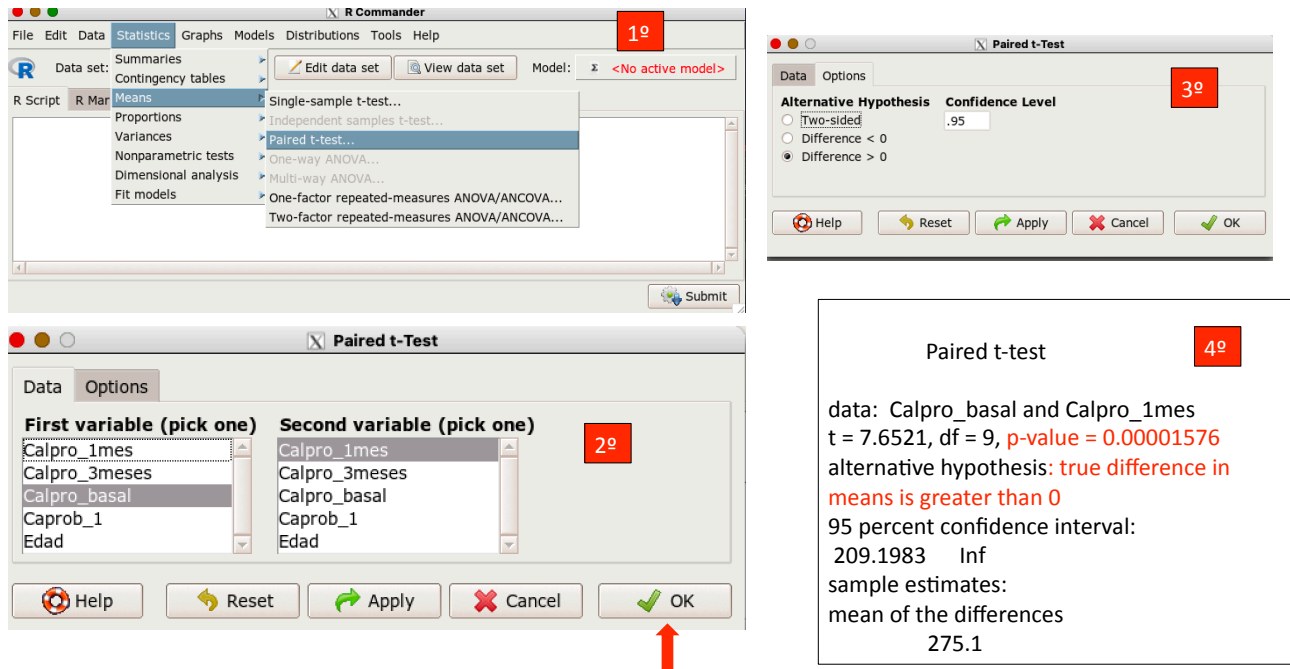


Figura 3. Prueba T para muestras apareadas



Comparación de más de dos medias para muestras relacionadas. Análisis de la varianza (ANOVA) para medidas repetidas

Se utiliza cuando tenemos más de dos medidas intrasujeto en tiempos diferentes. La hipótesis nula es que las medias de la variable de desenlace no cambian a lo largo del tiempo, siendo la hipótesis alternativa que al menos una de ellas sí cambia. Supongamos que tenemos tres medidas de una variable en tres tiempos diferentes: A, B y C.

H_0 : media (A) = media (B) = media (C)
 H_1 : media A \neq media B; media B \neq media C; media A \neq media C.
 Cualquiera de las tres es válida.

El fundamento es similar al visto con anterioridad en el capítulo de ANOVA, salvo que para el cálculo del estadístico de contraste (F) la variabilidad residual (varianza) está compuesta por la variabilidad total menos la variabilidad intrasujeto (varianza entre la repetición de cada medida) y la variabilidad entre los sujetos.

$$F = \frac{\text{Variabilidad intrasujeto}}{\text{Variabilidad residual} = \text{Variabilidad total} - (\text{variabilidad intrasujeto} + \text{variabilidad entresujetos})}$$

Condiciones para la aplicación del ANOVA de medidas repetidas

1. Las observaciones deben ser independientes.
2. Las variables de medidas repetidas deben seguir una distribución normal.
3. Debe existir esfericidad. Supone que las varianzas de la variable desenlace sean iguales en cada una de las medidas

repetidas. Sería el análogo a la homogeneidad de las varianzas en el ANOVA.

El estudio de la esfericidad se realiza mediante la prueba de Mauchly, bajo la hipótesis nula de que existe esfericidad. Se expresa mediante el coeficiente ϵ ; un valor de 1 significa esfericidad completa, cuanto menor sea el valor menor probabilidad habrá de que no se cumpla la condición hasta un nivel crítico de rechazo. En las situaciones en que este supuesto no se cumple, podemos aplicar las estimaciones de la ϵ mediante las correcciones de Greenhouse-Geisser (GG) y la de Huynh-Feldt (HF), o bien aplicar la prueba no paramétrica de Friedman que veremos posteriormente. La corrección de GG es más conservadora (mayor p) y es la recomendada.

Las comparaciones *a posteriori* (*post hoc*) es posible hacerlas comparando dos a dos las diferentes medias por medio de la prueba T para medidas apareadas, con corrección posterior de la probabilidad mediante la prueba de Bonferroni o de Holm.

Ejemplo de ANOVA de medidas repetidas

Volvemos a la base de datos del ejemplo anterior (fund_med_rep). Nos interesa conocer si los niveles de calprotectina fecal descienden de manera significativa tras el tratamiento con exclusión de PLV en el primer mes y el tercer mes.

Para la realización del ejercicio se ha utilizado el *plugin* RcmdrPlugin.EZR, disponible solo en inglés, ya que RCommander no tiene implementado en sus rutinas el análisis de las comparaciones *post hoc*. Para ello, una vez que hemos abierto el

programa R, debemos instalar el paquete correspondiente al *plugin* (si no se ha instalado y usado antes), para lo cual el método más rápido es teclear el comando `install.packages(RcmdrPlugin.EZR)`. Si este método falla, puede seleccionarse en R el menú Paquetes → Instalar paquetes(s)... y seleccionar un CRAN de la lista que ofrece la ventana emergente. Una vez hecho, veremos la lista de todos los paquetes de R ordenados alfabéticamente. Buscamos y marcamos el *plugin* RcmdrPlugin.EZR y pulsamos OK.

Empezaremos comprobando los supuestos de la prueba. En la **Figura 4** podemos ver cómo la distribución de las tres medidas de calprotectina fecal (Calpro_basal, Calpro_1mes, Calpro_3meses) siguen una distribución normal, puesto que la prueba de Shapiro-Wilk toma un valor $p > 0,05$ y la distribución de los datos en el gráfico de cuantiles se distribuyen alrededor de la diagonal.

Para la comprobación de la esfericidad, debemos realizar la prueba estadística en RCommander. Para ello, seleccionamos la pestaña *Statistical análisis* (análisis estadísticos) → *Continuos variables* (variables continuas) → *Repeated measures ANOVA*. En la ventana emergente en la caja de *Repeatedly measured data* (datos para las medidas repetidas) seleccionamos las tres

medidas. La prueba de esfericidad de Mauchly y las correcciones de GG y HF están implementadas por defecto. Seleccionamos *Lyne discrimination* (color) para conocer el gráfico de los valores de las medias y *Pairwise comparison Bonferroni* (prueba de comparación de medias con corrección de Bonferroni) y *Pairwise comparison Holm* (prueba de comparación de medias con corrección de Holm) para las comparaciones *post hoc* y clicamos en OK (**Figura 5**).

En la **Figura 6**, mostramos la ventana de resultados. En la tabla de ANOVA, además de mostrarnos la suma de cuadrados, los grados de libertad (df) y la F, vemos que por defecto R ha asignado el término *time* a las medidas repetidas, al que le corresponde un valor de $p < 0,05$ ($p = 0,0000000073$), lo que significa que los valores de calprotectina fecal en las tres medidas no permanece constante a lo largo del tiempo, que al menos una de las medias es diferente a las demás. A continuación, la comprobación de la esfericidad con la prueba de Mauchly, con $p > 0,05$ ($p = 0,17$), que nos confirma que existe esfericidad. Mas abajo, las correcciones de la esfericidad de GG and HF. Aunque en esta ocasión no las necesitamos, podemos observar resultados muy parecidos en la prueba HF ($p = 0,00000007696261$) y un poco más conservador (mayor p) en la prueba de GG ($0,0000004662$). Por último, tenemos las

Figura 4. Comprobación de supuestos de normalidad de Calprotectina basal, al mes y a los tres meses, previo a la realización de la ANOVA con medidas repetidas

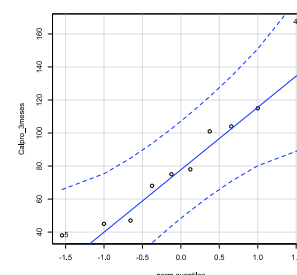
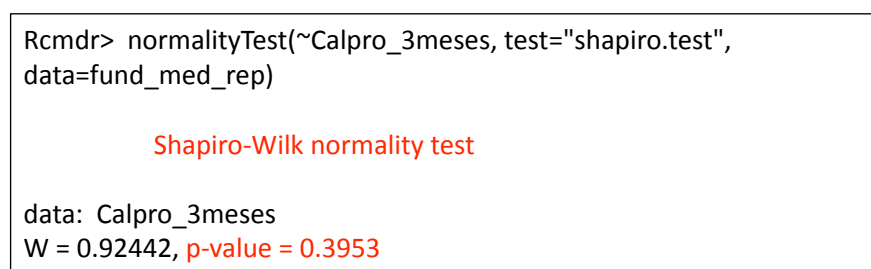
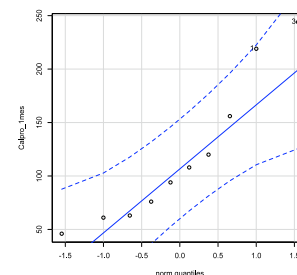
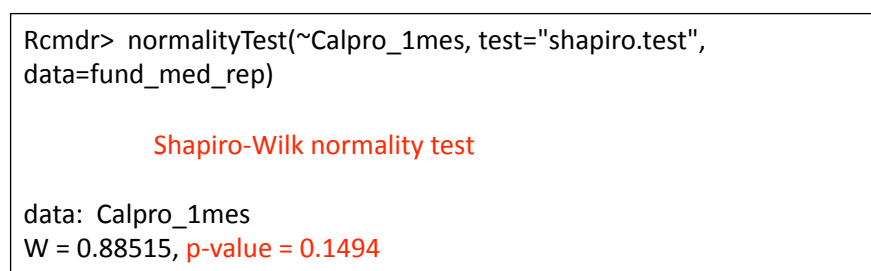
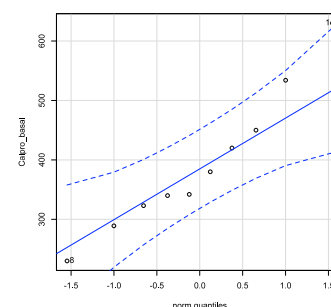
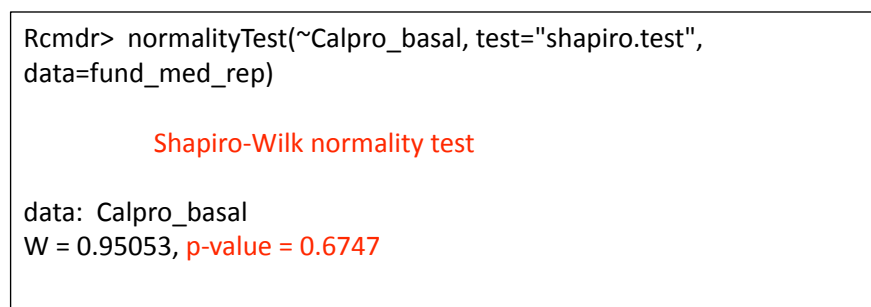
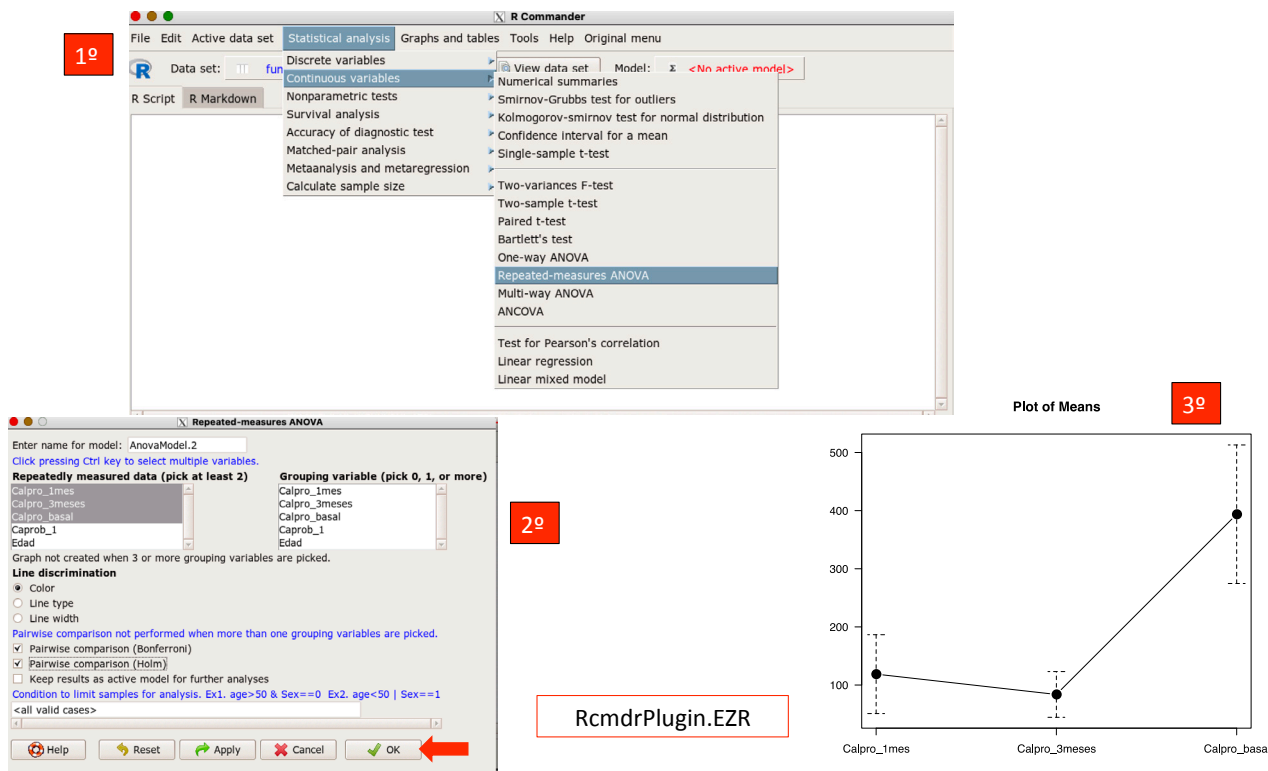


Figura 5. Realización de la ANOVA para medidas repetidas

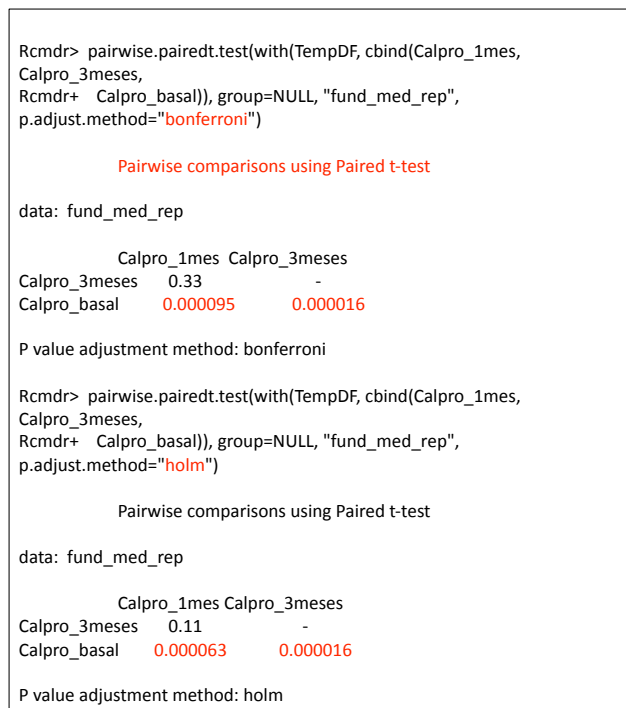
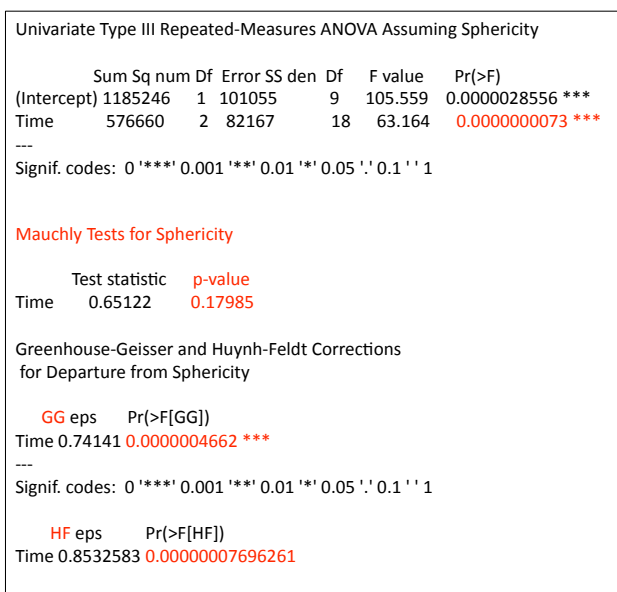


comparaciones *post hoc* realizadas dos a dos con la prueba de Wilcoxon apareadas con corrección de Bonferroni y Holm. Ambas ofrecen resultados muy parecidos, mostrando diferencias estadísticamente significativas de la disminución de la media de calprotectina fecal basal, al mes y a los tres meses, no así entre el mes y a los tres meses. Esta disminución tam-

bién es posible sospecharla en el gráfico de medias (*plot of means*) de la **Figura 5**.

En resumen, podemos decir que las medias de calprotectina fecal disminuyen de manera significativa después de la exclusión de PLV en los pacientes afectados de APLV y esta disminu-

Figura 6. Resultados ANOVA para medidas repetidas



ción se produce al mes y a los tres meses respecto al valor basal de manera significativa, no ocurriendo entre el mes y los tres meses.

PRUEBAS NO PARAMÉTRICAS PARA MUESTRAS RELACIONADAS

Comparación para dos muestras relacionadas. Prueba de Wilcoxon para muestras relacionadas

Se utiliza cuando los supuestos de normalidad no se cumplen o bien los datos a comparar son ordinales. El fundamento de la prueba es parecido al de la prueba de la U de Mann-Whitney vista con anterioridad en el capítulo de pruebas no paramétricas. Se calculan las diferencias entre cada pareja de individuos, luego se ordenan (rangos) las diferencias, las positivas y negativas por separado, y se suman. La hipótesis nula (H_0) sería que las diferencias entre ambos rangos (df) sean igual a 0; la alternativa (H_1) es que sean distintas de 0.

$$H_0: df = 0 \quad H_1: df \neq 0$$

Veámoslo con un ejemplo. En la base de datos fund_med_rep2, disponible en la web de *Evidencias en Pediatría* (https://evidenciasenpediatria.es/files/43-227-RUTA/fund_med_rep2.RData), con las mismas variables que la base de datos anterior, pero con datos distintos.

En primer lugar, comprobamos los supuestos de la prueba, que la distribución de los datos de al menos una muestra no sigue una distribución normal. Para ello, realizamos los mismos pasos que en el apartado anterior con la variable Calpro_basal y Calpro_1mes y obtenemos que en los gráficos de caja son bastante asimétricos, así como la existencia de varios valores alejados de la diagonal en el gráfico de cuantiles. Aplicando el test de normalidad de Shapiro-Wilk tenemos que, mientras Calpro_basal no sigue una distribución normal ($p = 0,03897$), Calpro_1mes sí la sigue ($p = 0,10699$). Por curiosidad, el lector puede también comprobar que las diferencias de las medias (Calpro_b1) tampoco se distribuyen normalmente ($p = 0,02452$) (**Figura 7**).

Ya estamos en disposición de realizar la prueba. En RCommander, vamos a la pestaña *Statistics* (estadísticos) → *Nonparametrics test* (pruebas no paramétricas) → *Paired samples Wilcoxon test* (prueba pareada de Wilcoxon). En la pantalla siguiente elegimos la Calpro_basal como primera variable (*first variable*) y Calpro_1mes como segunda variable (*second variable*). Al igual que en el caso de la prueba T apareada nos

interesa la prueba unilateral: *difference > 0* (la diferencia de las medianas es mayor de 0) y clicamos en OK. En la ventana de resultados obtenemos una $p = 0,000976$, lo que significa que existe una disminución estadísticamente significativa en los valores de calprotectina fecal al mes de comenzar con exclusión de las PLV ($p = 0,0009766$) (**Figura 8**). En los diagramas de caja de las dos mediciones de la **Figura 7** podemos ver la magnitud de la diferencia; hay que tener en cuenta que los diagramas tienen escalas diferentes en sus ejes verticales.

Comparación de más de dos muestras relacionadas. Prueba de Friedman

En un diseño intrasujetos, se utiliza cuando las medidas relacionadas a comparar no siguen una distribución normal, hay ausencia de esfericidad en la prueba de ANOVA repetidas, no es posible dar un resultado satisfactorio con las correcciones de GG y HF y, por tanto, no es posible utilizar el ANOVA de medidas repetidas. Se considera una extensión de la prueba de Wilcoxon para medidas apareadas, bajo la hipótesis nula de que no existen cambios en las medidas repetidas.

Retomemos la base de datos fund_med_rep2. Al igual que en el ANOVA para medidas repetidas, queremos conocer si los valores de calprotectina fecal varían después de la exclusión de PLV.

Solo la medida de Calpro_basal no sigue una distribución normal ($p = 0,03897$), las otras dos medidas tienen valores de $p > 0,05$ (**Figura 9**).

Para realizar la prueba, retomamos el *plugin RcmdrPlugin.EZR*, ya instalado anteriormente, accedemos a la pestaña *Statistics* (estadísticos) → *Nonparametrics test* (pruebas no paramétricas) → *Friedman test* (prueba de Friedman) y en la ventana emergente seleccionamos las tres medidas en la caja *Repeated measures variables* (variables para medidas repetidas), seleccionamos *Pairwise comparison Bonferroni* (comparación por pares corrección de Bonferroni) y clicamos OK. En la ventana de resultados obtenemos la p de la prueba ($p = 0,00184$); como es menor de 0,5 quiere decir que los valores de calprotectina fecal no son iguales en las tres medidas. Para conocer qué medida es la que contribuye a la desigualdad, vemos las comparaciones por pares usando la prueba de Wilcoxon (*Pairwise comparison using Wilcoxon signed rand test*), donde podemos comprobar que es la calprotectina al mes respecto a la basal la que es significativa ($p = 0,0058$) y el resto de las comparaciones no lo son (**Figura 10**).

Figura 7. Comprobación de la normalidad de Calprotectina basal, al mes previo a la realización de la prueba de Wilcoxon

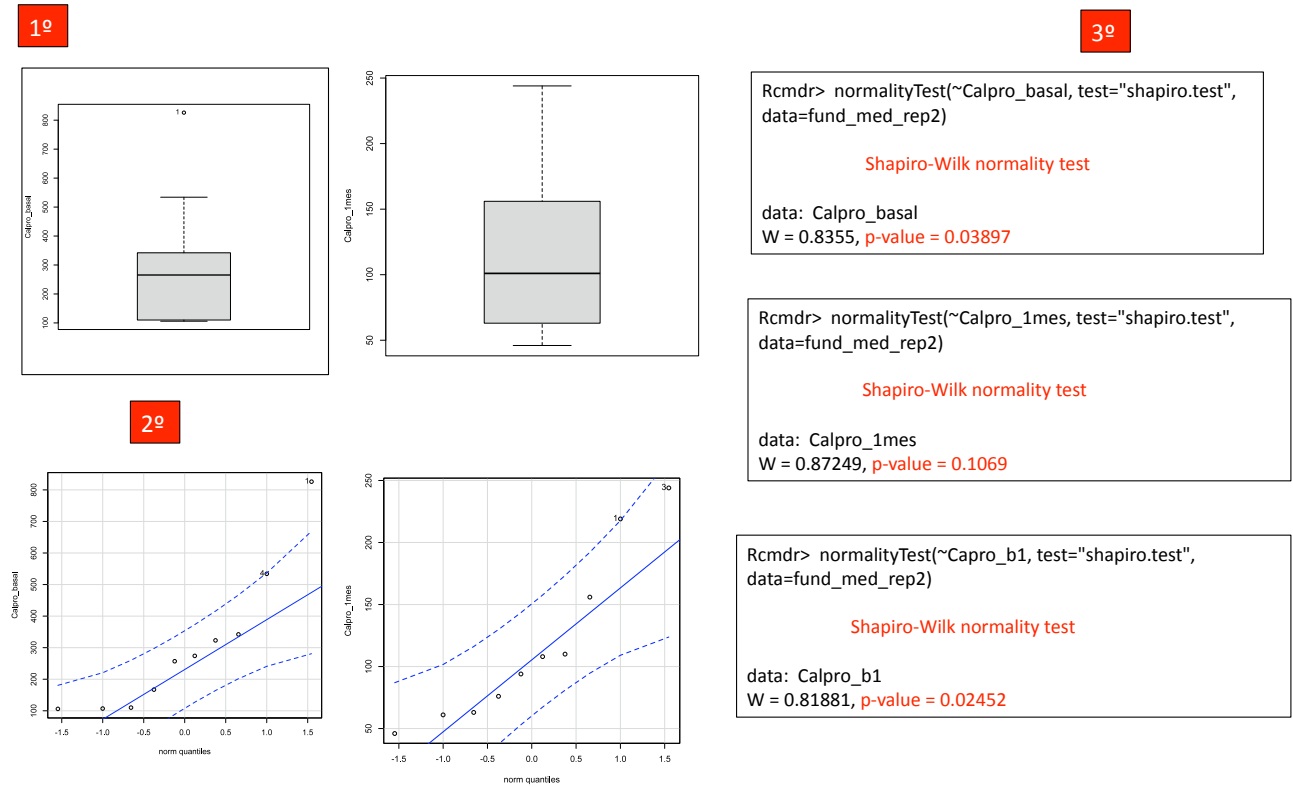


Figura 8. Prueba de Wilcoxon para muestras apareadas

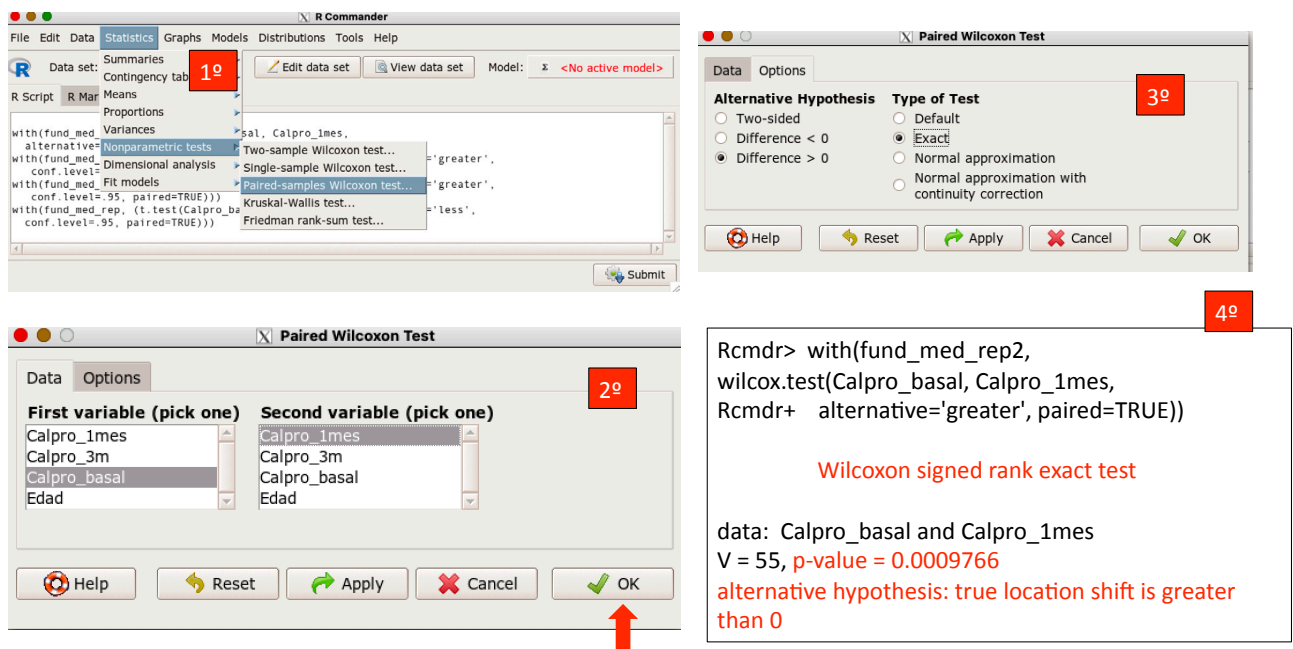


Figura 9. Pruebas de normalidad de Calprotectina basal, al mes y a los tres meses, previo a la prueba de Friedman

```
Rcmdr> normalityTest(~Calpro_basal, test="shapiro.test",
data=fund_med_rep2)
```

Shapiro-Wilk normality test

data: Calpro_basal
W = 0.8355, p-value = 0.03897

```
Rcmdr> normalityTest(~Calpro_1mes, test="shapiro.test",
data=fund_med_rep2)
```

Shapiro-Wilk normality test

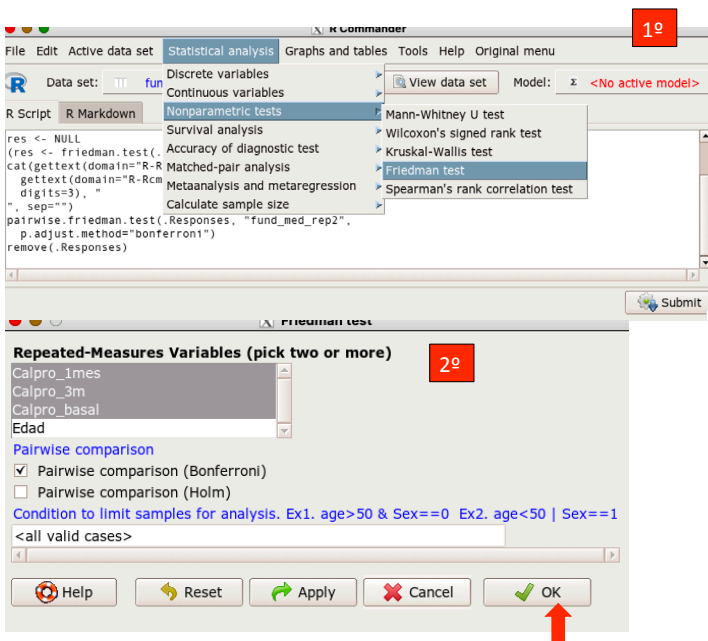
data: Calpro_1mes
W = 0.87249, p-value = 0.1069

```
Rcmdr> normalityTest(~Calpro_3mes, test="shapiro.test",
data=fund_med_rep2)
```

Shapiro-Wilk normality test

data: Calpro_3mes
W = 0.87249, p-value = 0.1069

Figura 10. Prueba de Friedman



1º

3º

```
Friedman test p.value = 0.00184
```

```
Rcmdr> pairwise.friedman.test(.Responses,
"fund_med_rep2",
Rcmdr+ p.adjust.method="bonferroni")
```

Pairwise comparisons using Wilcoxon signed rank test

data: fund_med_rep2

	Calpro_1mes	Calpro_3m
Calpro_3m	0.9960	-
Calpro_basal	0.0058	0.0585

P value adjustment method: bonferroni
RcmdrMsg: [4] WARNING: Warning in wilcox.test.default(pairwise.response1, RcmdrMsg+ pairwise.response2, alternative = "two.sided", : RcmdrMsg+ cannot compute exact p-value with ties

RcmdrPlugin.EZR

BIBLIOGRAFÍA

- Amat Rodrigo J. ANOVA análisis de varianza para comparar múltiples medias [en línea] [consultado el 05/12/2022]. Disponible en www.cienciadedatos.net/documentos/19_anova.html
- Amat Rodrigo J. Test de Friedman [en línea] [consultado el 05/12/2022]. Disponible en www.cienciadedatos.net/documentos/21_friedman_test
- Domenech Massons JM. Métodos estadísticos en ciencias de la salud. Unidad didáctica 15: Medias de cambio: Análisis de diseños con datos apareados. Editor: Domenech JM. Barcelona. Signo. 1998.
- Gonzalo M. Manual RCommander ANOVA de un factor para muestras relacionadas. Psicometría. Universidad Complutense de Madrid (UCM) [en línea]. Disponible en www.doccity.com/es/anova-de-dos-factores-de-medida-repetidas-con-r-commander/4156928/
- Molina Arias M, Ochoa Sangrador C, Ortega Páez E. Comparación de dos medias. Pruebas de la t de Student. *Evid Pediatr.* 2020;16:51.
- Molina Arias M, Ochoa Sangrador C, Ortega Páez E. Comparación de dos medias. Análisis de la varianza. *Evid Pediatr.* 2021;17:11.
- Ortega Páez E, Ochoa Sangrador C, Molina Arias M. Pruebas no paramétricas. *Evid Pediatr.* 2021;17:36.
- Sayón Orea C, Toledo E, Núñez Córdoba JM. ANOVA factorial, modelos lineales generalizados y ANCOVA. En: Martínez MA, Sánchez Villegas A, Toledo EA, Faulin A (eds.). *Bioestadística amigable* (3.ª ed.). Barcelona: Elsevier España; 2014. p. 241-268.