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

Fundamentos de medicina basada en la evidencia

Comparación de dos medias. Pruebas de la t de Student

Molina Arias M[1], Ochoa Sangrador C[2], Ortega Páez E[3]
[1]Servicio de Gastroenterología. Hospital Universitario La Paz. Madrid. España.
[2]Servicio de Pediatría. Hospital Virgen de la Concha. Zamora. España.
[3]UGC de Maracena. Distrito Granada-Metropolitano. Granada. España.

Correspondencia: Manuel Molina Arias, mma1961@gmail.com

Palabras clave en español: estadística; inferencia estadística; comparación de medias.
Palabras clave en inglés: statistics; statistical inference; comparing means.

Fecha de recepción: 2 de diciembre de 2020 • **Fecha de aceptación:** 11 de diciembre de 2020
Fecha de publicación del artículo: 16 de diciembre de 2020

Evid Pediatr. 2020;16:51.

CÓMO CITAR ESTE ARTÍCULO

Molina Arias M, Ochoa Sangrador C, Ortega Páez E. Comparación de dos medias. Pruebas de la t de Student. Evid Pediatr. 2020;16:51.

Este artículo está disponible en: http://www.evidenciasenpediatria.es/EnlaceArticulo?ref=2020;16:51.

Comparación de dos medias. Pruebas de la t de Student

Molina Arias M[1], Ochoa Sangrador C[2], Ortega Páez E[3]
[1]*Servicio de Gastroenterología. Hospital Universitario La Paz. Madrid. España.*
[2]*Servicio de Pediatría. Hospital Virgen de la Concha. Zamora. España.*
[3]*UGC de Maracena. Distrito Granada-Metropolitano. Granada. España.*

Correspondencia: Manuel Molina Arias, mma1961@gmail.com

En un artículo previo de esta serie establecimos los fundamentos del contraste de hipótesis, que serán los que utilizaremos para comparar parámetros entre dos poblaciones o grupos.

De manera general, en el contraste de hipótesis se parte de un supuesto de igualdad de los dos parámetros que queremos comparar, supuesto al que denominamos hipótesis nula. Una vez obtenidos los parámetros, calcularemos la probabilidad de que, bajo el supuesto de la hipótesis nula, la diferencia que observemos entre ellos sea debida a error aleatorio o de muestreo. Si esta probabilidad es inferior a un determinado valor que, por convenio, suele situarse en 0,05, asumiremos que la probabilidad de que la diferencia se deba al azar es lo suficientemente baja como para rechazar la hipótesis nula y considerar como cierta la hipótesis alternativa de desigualdad (contraste bilateral) o de superioridad o inferioridad de uno de los parámetros (contraste unilateral).

Cuando queremos comparar dos parámetros de dos grupos diferentes, solemos utilizar un estadístico que se relacione con el parámetro y cuyos valores sigan una distribución de probabilidad conocida. Así, los pasos para realizar la comparación serán siempre los mismos: establecer la hipótesis nula de igualdad de los parámetros, seleccionar el estadístico adecuado para cada situación, utilizar la distribución de probabilidad correspondiente para calcular la probabilidad de ese valor del estadístico que hemos empleado y, según este valor de probabilidad, decidirnos en favor de la hipótesis nula o de la alternativa que hayamos elegido, unilateral o bilateral.

En este capítulo trataremos sobre la comparación de dos medias.

COMPARACIÓN DE UNA MEDIA CON UN VALOR DE REFERENCIA

El problema más sencillo es el de comparar la media de un parámetro obtenido en una muestra con un valor de referencia de una población conocida. Esto nos permitirá, por un lado, decidir si es razonable concluir que la muestra puede pertenecer a la población o bien, por otro lado, contrastar hipótesis sobre la media poblacional a partir de la obtenida en la muestra.

Aunque es posible comparar dos medias utilizando la distribución normal, para ello necesitamos conocer la desviación estándar poblacional del parámetro, que suele ser desconocida. En estos casos utilizamos, como aproximación a la desviación estándar poblacional (σ), la de la muestra (s), remplazando las probabilidades de la distribución normal por las de la t de Student, que varían en función del tamaño muestral (los grados de libertad). En cualquier caso, cuando el tamaño muestral es grande, el valor de probabilidad obtenido mediante la t de Student se aproxima al obtenido con la distribución normal.

En la práctica, calculamos el intervalo de confianza de la media poblacional, ya sea utilizando la distribución normal o la t de Student con n-1 grados de libertad (siendo n el tamaño de la muestra), y comprobamos si el intervalo incluye el valor de referencia. La fórmula para el cálculo del intervalo de confianza, utilizando la normal o la t de Student, respectivamente, sería:

$$\mu = z_{\alpha/2} \pm \frac{\bar{x}}{\sqrt{\frac{\sigma}{n}}}$$

$$\mu = t_{n-1;\alpha/2} \pm \frac{\bar{x}}{\sqrt{\frac{\sigma}{n}}}$$

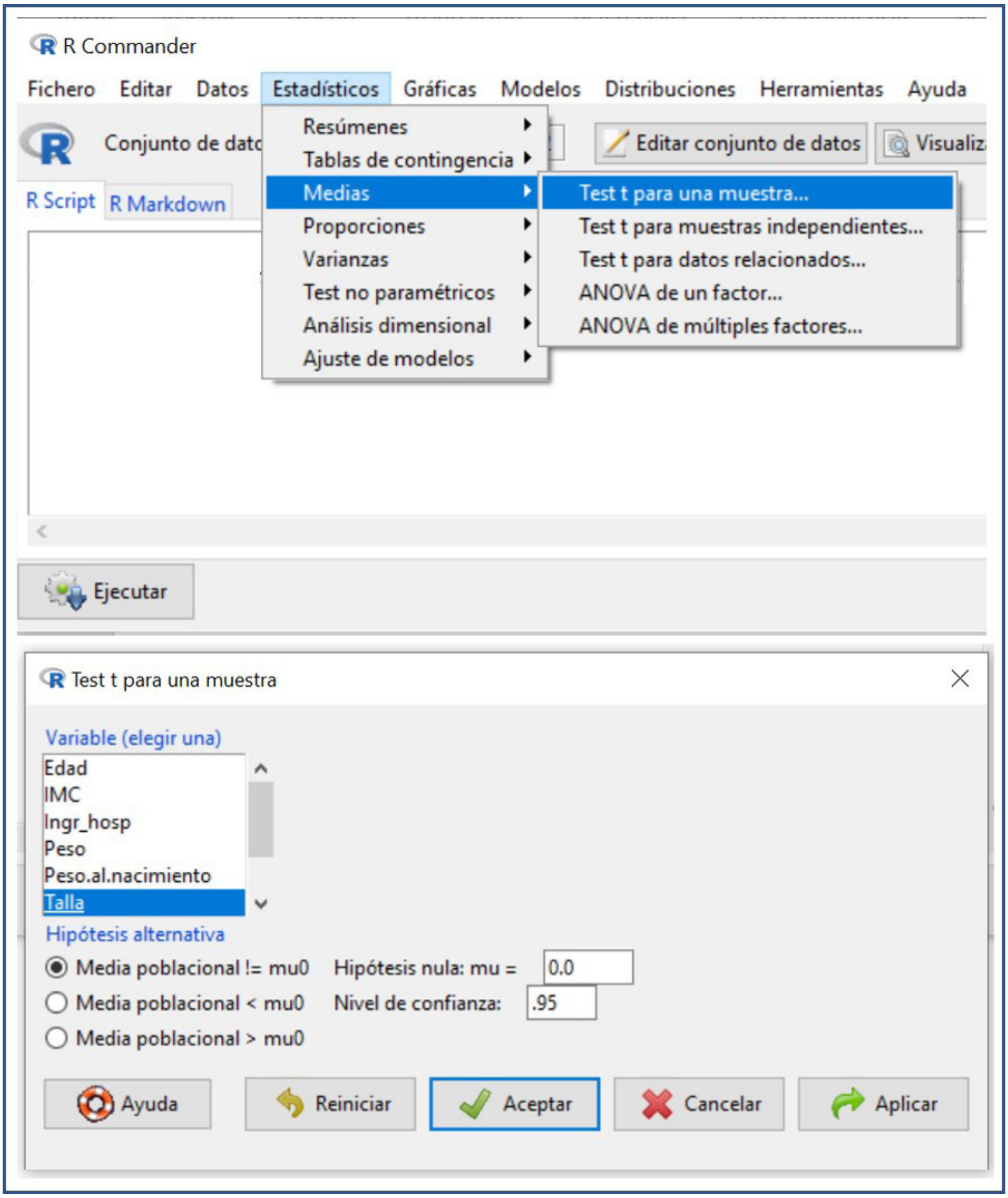
Aunque el cálculo es relativamente sencillo, aconsejamos utilizar un programa estadístico. Veamos un ejemplo utilizando un programa de acceso libre, el software estadístico R (www.r-project.org/) con el plugin RCommander (www.rcommander.com/) y la base de datos Fundamentos_graficos.RData, disponible en la web de *Evidencias en Pediatría*. 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).

En esta base de datos tenemos los registros de una serie de pacientes asmáticos. Queremos saber si la talla de estos pacientes es similar a la talla media de la población de la misma edad, que sabemos que es de 150 cm. Para ello, una vez abierto RCommander y cargado el conjunto de datos activos, seleccionamos en el menú las opciones Estadísticos/Medias/Test t para una muestra… (figura 1). En la nueva ventana debemos seleccionar la variable de la que queremos obtener el intervalo, la opción que elegimos para la hipótesis alternativa (en este caso, de igualdad de medias, contraste bilateral) y el nivel de significación estadística (lo dejamos en 0,05).

R nos presenta la media de talla de nuestra muestra, 122,21 cm, y su intervalo de confianza del 95%, de 114,85 cm a 129,58 cm. Podemos concluir que la talla de los niños asmáticos es inferior a la talla media conocida de la población general, ya que el intervalo no incluye el valor 150 cm.

El programa nos muestra también el valor del estadístico t, sus grados de libertad (n-1 en este caso) y su significación (p $<2{,}2 \times 10^{-16}$). Entenderemos mejor estos parámetros al tratar el siguiente punto.

FIGURA 1. COMPARACIÓN DE UNA MEDIA CON UN VALOR DE REFERENCIA MEDIANTE LA PRUEBA DE LA T DE STUDENT PARA UNA MUESTRA



COMPARACIÓN DE DOS MEDIAS INDEPENDIENTES

El supuesto más habitual es el de contrastar si hay una diferencia significativa en la media de una variable de resultado entre dos poblaciones diferentes e independientes. En estos casos, lo habitual es utilizar la prueba de la t de Student para dos muestras independientes.

Esta prueba compara las dos medias de una variable de resultado cuantitativo continuo obtenidas en dos categorías definidas por una variable cualitativa. Se basa en el cálculo del estadístico t, que tiene en cuenta la diferencia de medias a comparar y su error estándar, según la siguiente fórmula:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Siendo $\bar{x}_1, s_1^2$ y $\bar{x}_2, s_1^2$ las medias y las varianzas de las dos muestras respectivamente.

Bajo el supuesto de la hipótesis nula, la diferencia de medias es igual a cero, con lo que el valor de t será también igual a cero. Cuanto más se aleje t de ese valor, menos probable será que la diferencia observada se deba al azar.

Para poder aplicar esta prueba, debemos verificar previamente que se cumplen tres condiciones:

1. Los dos grupos deben ser independientes. Esto quiere decir que cada participante debe pertenecer a solo uno de los dos grupos y no tiene relación con los participantes del otro grupo.
2. La variable de resultado debe ser continua y seguir una distribución normal en los dos grupos.
3. Debe cumplirse el supuesto de homocedasticidad, esto es, igualdad de varianzas en los dos grupos.

El supuesto de normalidad de la variable en los dos grupos puede verificarse mediante la prueba de Shapiro-Wilk, más adecuada para muestras pequeñas, menores de 50, o la de Kolmogorov-Smirnov (con la modificación de Lilliefors en el supuesto habitual de desconocer la media y desviación estándar poblacional). Estas dos pruebas tienen el inconveniente de que asumen una hipótesis nula de normalidad. Si el resultado es significativo, podremos descartar la hipótesis nula y concluir que la variable no sigue una distribución normal. Sin embargo, un resultado no significativo no nos permite asegurar que la hipótesis alternativa de normalidad sea cierta (solo que no podemos rechazar la hipótesis nula). Además, ambas son poco potentes cuando el tamaño de la muestra no es grande, precisamente el supuesto en el que la t de Student es más sensible a la ausencia de normalidad. Por estos motivos, se recomienda completar la prueba de contraste con un método gráfico (como el histograma o el gráfico de cuantiles teóricos) y tener en cuenta el tamaño de la muestra.

En el caso de no seguir una distribución normal, se nos plantearán tres posibilidades. La primera, hacer el contraste de hipótesis con una prueba no paramétrica que, en este caso, sería la de la U de Mann-Whitney. La segunda, podemos intentar alguna transformación y comprobar si la variable transformada se distribuye de forma normal. La tercera, realizar una t de Student a pesar de no cumplirse la condición de normalidad, aplicando una corrección. Esto solo será aconsejable si hay ligeras desviaciones de la normalidad y el tamaño muestral es grande (mínimo de 30 a 50 participantes por grupo).

Una vez comprobada la normalidad, determinaremos que las dos varianzas son iguales y se cumple el supuesto de homocedasticidad. Para ello, puede calcularse una F de Snedecor con el cociente de las dos varianzas, colocándose en el numerador la varianza mayor de los dos grupos y en el denominador, la menor. Los grados de libertad son n-1 de cada grupo, siendo n el tamaño de cada grupo.

Bajo el supuesto de igualdad de varianzas, F valdrá 1. Cuanto mayor sea el valor de F, menos probable será que la diferencia observada entre las varianzas se deba al azar.

La prueba de la F de Snedecor es muy sensible a la falta de normalidad, por lo que en estos casos será recomendable recurrir a la prueba de Levene.

Una vez comprobado estos dos supuestos, procederemos a realizar el contraste mediante la prueba de la t de Student, aplicando la corrección de Welch si no existe homocedasticidad.

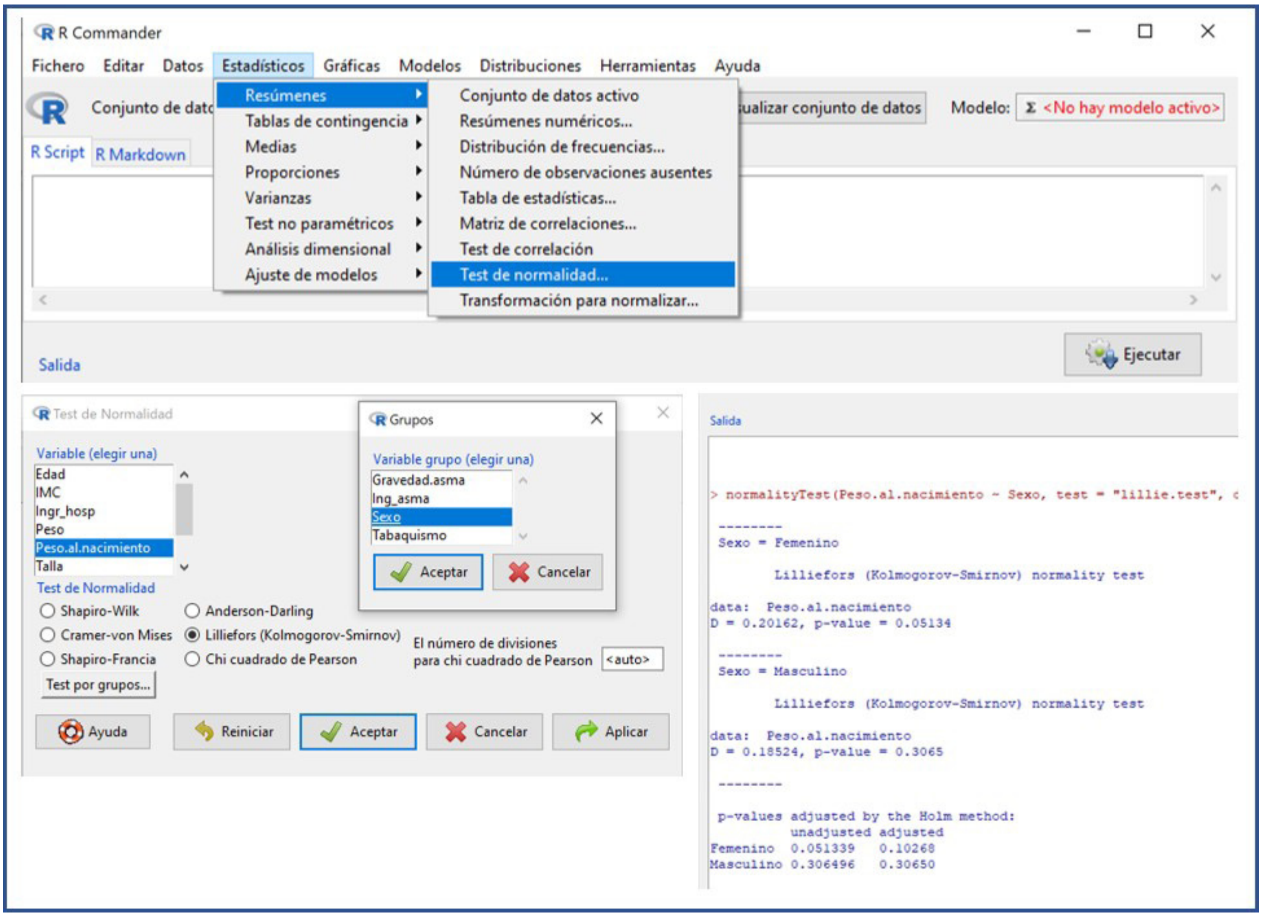
Vamos a ver un ejemplo práctico con la misma base de datos, comparando el peso al nacimiento entre niños y niñas.

Primero, calculamos la media para los dos valores. Con RCommander, seleccionamos el menú Estadísticos->Resúmenes->Tabla de estadísticas…. En la ventana emergente señalamos "Peso.al.nacimiento" como variable y "Sexo" como factor que diferencia los grupos a comparar. Como estadístico, señalaremos la media.

Obtenemos un peso al nacimiento medio de 2458 g en niñas y 2737 g en niños.

Comprobemos primero el supuesto de normalidad. Para ello, decidimos hacer una prueba de Kolmogorov-Smirnof. En el menú de RCommander seleccionamos las opciones Estadísticos/Resúmenes/Test de normalidad… y, en la ventana emergente, marcamos la variable "Peso.al.nacimiento", elegimos la prueba y pulsamos el botón "Test por grupos…" para seleccionar la variable "Sexo" (figura 2). En la ventana de salida podemos ver el resultado, siendo la p >0,05 para las dos categorías de la variable "Sexo", por lo que no rechazamos la hipótesis nula de normalidad.

FIGURA 2. COMPROBACIÓN DEL SUPUESTO DE NORMALIDAD MEDIANTE LA PRUEBA DE KOLMOGOROV-SMIRNOV



Comprobemos ahora el supuesto de homocedasticidad. Seleccionamos las opciones Estadísticos/Varianzas/Test F para dos varianzas… En la ventana emergente seleccionamos "Peso.al.nacimiento" como variable explicada y la agrupamos por "Sexo" (figura 3). Si marcamos la pestaña de "Opciones", podemos seleccionar el tipo de contraste y el nivel de significación. Dejamos la opción por defecto, que es un contraste bilateral con un nivel de significación de 0,05.

Vemos que el programa nos da el valor de F (1,2052), los grados de libertad para el numerador y el denominador y la probabilidad de encontrar ese valor de F por azar (valor de *p* = 0,7695). Por tanto, no rechazamos la hipótesis nula de igualdad de varianzas. Vemos que R nos proporciona también el intervalo de confianza del 95% del valor de F (de 0,36 a 3,45), que incluye el valor 1 que corresponde a varianzas iguales.

Una vez comprobadas normalidad y homocedasticidad, ya podemos realizar la prueba de la t de Student. En este caso, al haber homocedasticidad, no será necesario aplicar la corrección de Welch. Seleccionamos Estadísticos/Medias/Test t para muestras independientes… En la ventana emergente marcamos "Peso.al.nacimiento" como variable explicada y la agrupamos por "Sexo" (figura 4). A continuación, pulsamos la pestaña opciones, donde podemos cambiar el tipo de contraste y la significación (bilateral y 0,05, por defecto) y donde tenemos que señalar si las varianzas son o no iguales. En el caso de que marquemos "No", R hará la prueba aplicando la corrección de Welch. En este caso, marcamos "Sí".

El programa nos ofrece el valor del estadístico t (-2,36), sus grados de libertad (n-2, 28) y su valor de *p* (0,02). Por si tenemos alguna duda en la dirección del contraste, también nos dice cuál es la hipótesis alternativa: que la verdadera diferencia entre las medias de los dos grupos es distinta de cero. Al ser el valor de *p* <0,05, rechazamos la hipótesis nula de igualdad de medias y concluimos que sí existe una diferencia entre los dos grupos en el peso al nacimiento.

Merece la pena comentar que R nos ofrece también las medias de los dos grupos y el intervalo de confianza de la diferencia de medias. En este caso es de -520,88 a -37,17. Viendo que el intervalo no incluye el 0 (valor nulo para la diferencia de medias), podemos también rechazar la hipótesis nula de igualdad sin necesidad de recurrir al valor de *p*.

FIGURA 3. COMPROBACIÓN DEL SUPUESTO DE HOMOCEDASTICIDAD MEDIANTE LA PRUEBA DE LA F DE SNEDECOR

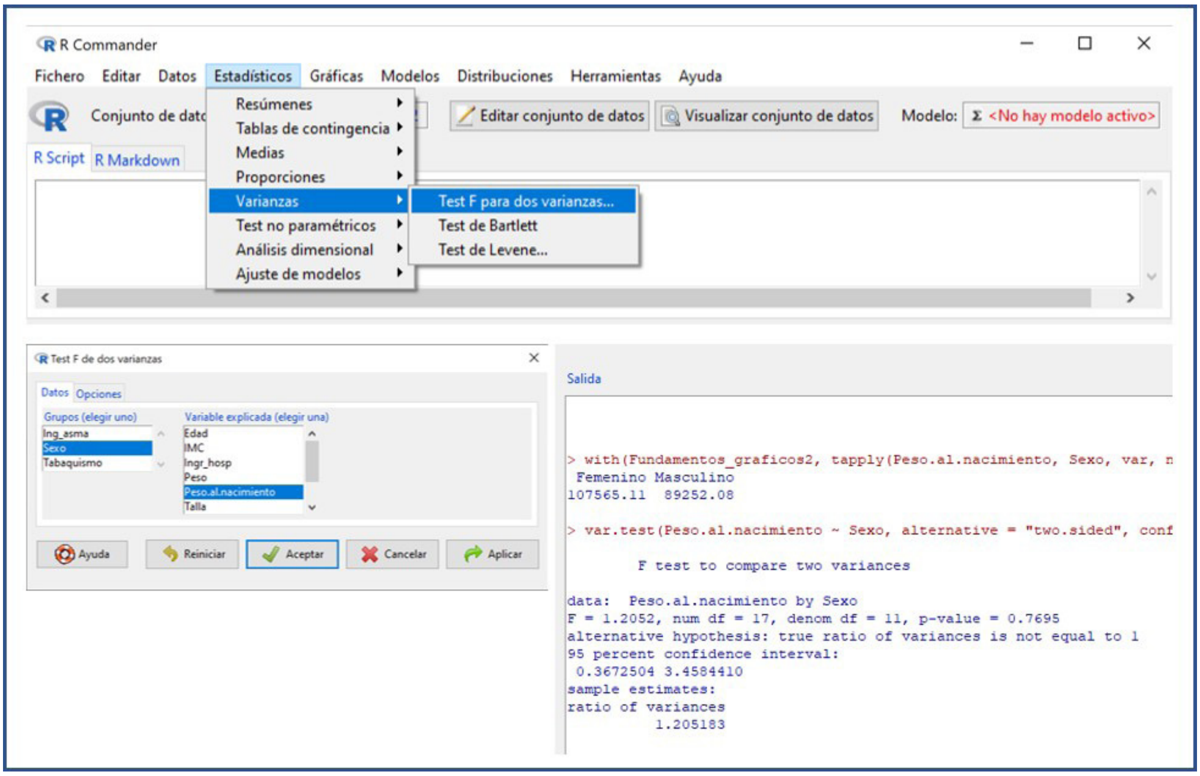
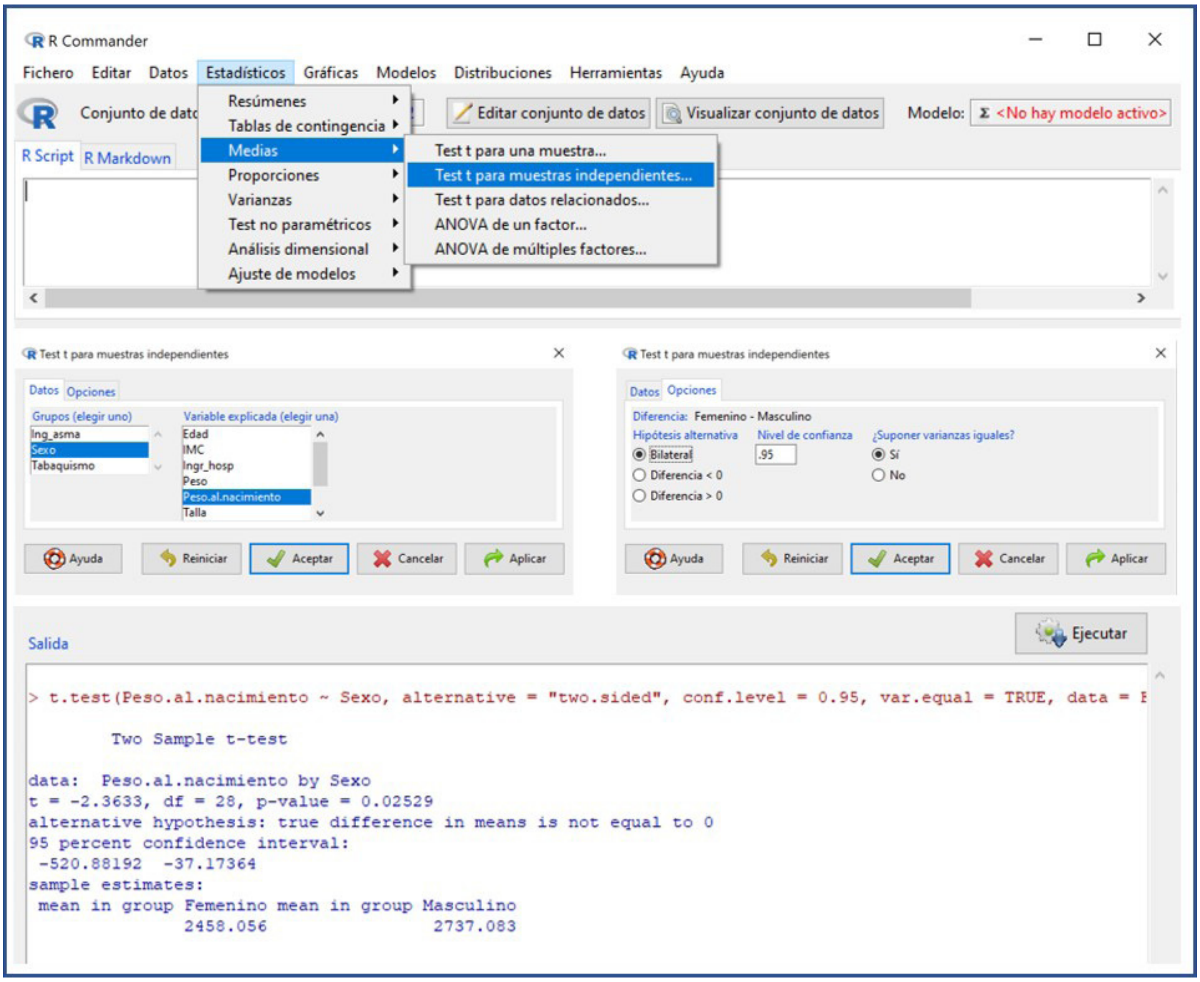


FIGURA 4. COMPARACIÓN DE DOS MEDIAS INDEPENDIENTES MEDIANTE LA PRUEBA DE LA T DE STUDENT PARA MUESTRAS INDEPENDIENTES



COMPARACIÓN DE DOS MEDIAS DEPENDIENTES O RELACIONADAS

En ocasiones se plantea el problema de comparar las medias de dos grupos que están relacionados, como puede ser el caso de medidas obtenidas del mismo participante en diferentes momentos, de diferentes localizaciones de la misma persona (por ejemplo, presión intraocular de ojo derecho e izquierdo) o cuando se comparan los datos de cada caso con su correspondiente control emparejado. Esto es muy típico de los estudios longitudinales, los estudios de antes-después de una intervención y los estudios de casos y controles.

En estos casos no existe una variable que defina los grupos, sino que la variable de resultado que se valorará será las diferencias entre los dos resultados de cada pareja, suponiendo la hipótesis nula que la media de estas diferencias es igual a cero. Así, en este tipo de análisis el interés no se centra en las diferencias entre individuos, sino en las que puede haber en el mismo individuo en dos momentos diferentes o entre las observaciones de los individuos relacionados.

La prueba que empleamos en estos casos es la de la t de Student para medidas repetidas (datos apareados o relacionados). Para poder aplicarla, debe cumplirse que la variable de interés sea cuantitativa continua, que la muestra de pares de datos haya sido obtenida al azar de la población y que la diferencia entre las parejas se distribuya de forma normal. Lógicamente, en este caso no tiene sentido plantear si hay igualdad de varianzas, ya que se trata de los mismos participantes en los dos grupos.

El planteamiento de la prueba es similar al de la t de Student para medias independientes, solo que en este caso se genera una nueva variable a partir de las dos medidas a comparar:

$$di = x_{i1} - x_{i2}$$

donde d_i es la diferencia de resultado de cada pareja en dos instantes diferentes, x_1 y x_2.

En este análisis, el estadístico t se obtiene con la media y la desviación estándar de esta variable, según la siguiente ecuación:

$$t = \frac{\bar{d}}{s_{d/\sqrt{n}}}$$

El contraste bilateral plantea la hipótesis nula de que la diferencia es igual a 0. En este caso, los grados de libertad son n-1, siendo n el tamaño de la muestra. Como en el caso de muestras independientes, si la hipótesis nula es cierta el valor de la diferencia será cero, por lo que t valdrá cero. Cuanto mayor sea el valor de t, menos probable será que la diferencia observada se deba al azar.

Desde un punto de vista práctico, el procedimiento es similar al que hemos descrito previamente, solo que esta vez seleccionaremos la opción Estadístico/Medias/Test t para datos relacionados… de RCommander. La lectura de los resultados será similar, proporcionándonos el programa el valor del estadístico t, su significación y el intervalo de confianza de la diferencia de medias. Animamos al lector a probar estas opciones.

BIBLIOGRAFÍA

- Estimating the difference between two population parameters. En: Bowers D (ed.). Medical statistics from scratch. An introduction for health professionals. 2.ª edición. West Sussex, Reino Unido: John Wiley & Sons Ltd.; 2018. pp. 119-31.
- Ochoa Sangrador C, Molina Arias M, Ortega Páez E. Inferencia estadística: contraste de hipótesis. Evid Pediatr. 2020;16:11.
- Ochoa Sangrador C, Molina Arias M, Ortega Páez E. Inferencia estadística: estimación del tamaño muestral. Evid Pediatr. 2020;16:24.
- Watt WW. Continuous variables: comparing two independent samples. En: Peat J, Barton B (eds.). Medical statistics. A guide to data analysis and critical appraisal. Massachusetts, EE. UU.: Blackwell Publishing Ltd.; 2005. pp. 51-85.
- Watt WW. Continuous variables: paired and one-sample t-tests. En: Peat J, Barton B (eds.). Medical statistics. A guide to data analysis and critical appraisal. Massachusetts, EE. UU.: Blackwell Publishing Ltd.; 2005. pp. 86-107.