JASP

Mark A Goss-Sampson

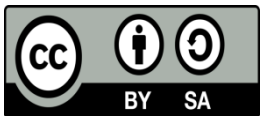
Análisis estadístico con JASP: una guía para estudiantes

PID_00265008

Traducción revisada por:

Julio Meneses

Traducción revisada por Julio Meneses, profesor agregado de los Estudios de Psicología y Ciencias de la Educación de la Universitat Oberta de Catalunya (UOC).
Av. Tibidabo, 39-43, 08035 Barcelona
Realización editorial: FUOC



Contenidos

PREFACIO

El acrónimo JASP tiene su origen en la expresión inglesa **Jeffrey's Amazing Statistics Program**, en reconocimiento al pionero de la inferencia bayesiana Sir Harold Jeffreys. Se trata de un paquete estadístico de código abierto multiplataforma, desarrollado y actualizado ininterrumpidamente (en su versión 0.9.2 a diciembre de 2018) por un grupo de investigadores de la Universidad de Amsterdam. Su objetivo era desarrollar un programa libre y de código abierto que incluyera tanto los estándares como las técnicas estadísticas más avanzadas, poniendo especial énfasis en lograr una interfaz de usuario simple e intuitiva.

En contraste con muchos otros paquetes de estadística, JASP facilita una interfaz simple de arrastrar y soltar, menús de fácil acceso, análisis intuitivo con computación a tiempo real y visualización de todos los resultados. Todas las tablas y los gráficos están presentados en formato APA y pueden ser copiados directamente y/o independientemente. Las tablas también pueden exportarse desde JASP a formato LaTeX.

JASP puede ser descargado desde el sitio web https://jasp-stats.org/ y está disponible para Windows, Mac OS X y Linux. También se puede descargar una versión preinstalada para Windows que funcionará directamente desde una unidad USB o un disco duro externo, sin necesidad de instalarlo localmente. El instalador WIX para Windows permite elegir una ruta para la instalación de JASP –no obstante, esta opción puede estar bloqueada en algunas instituciones debido a normas administrativas locales–.

El programa también incluye una librería de datos con una colección inicial con más de 50 conjuntos de datos procedentes del libro de Andy Field, *Discovering Statistics using IBM SPSS statistics*,[1] y de *The Introduction to the Practice of Statistics*,[2] de Moore, McCabe y Craig.

Desde mayo de 2018, JASP también puede ejecutarse directamente desde el navegador vía rollApp™ sin necesidad de instalar nada en el ordenador (https://www.rollapp.com/app/jasp). No obstante, podría no tratarse de la versión más reciente de JASP.

¡¡Es importante prestar atención a las actualizaciones regulares de JASP, y a los vídeos y los posts de ayuda de su blog!!

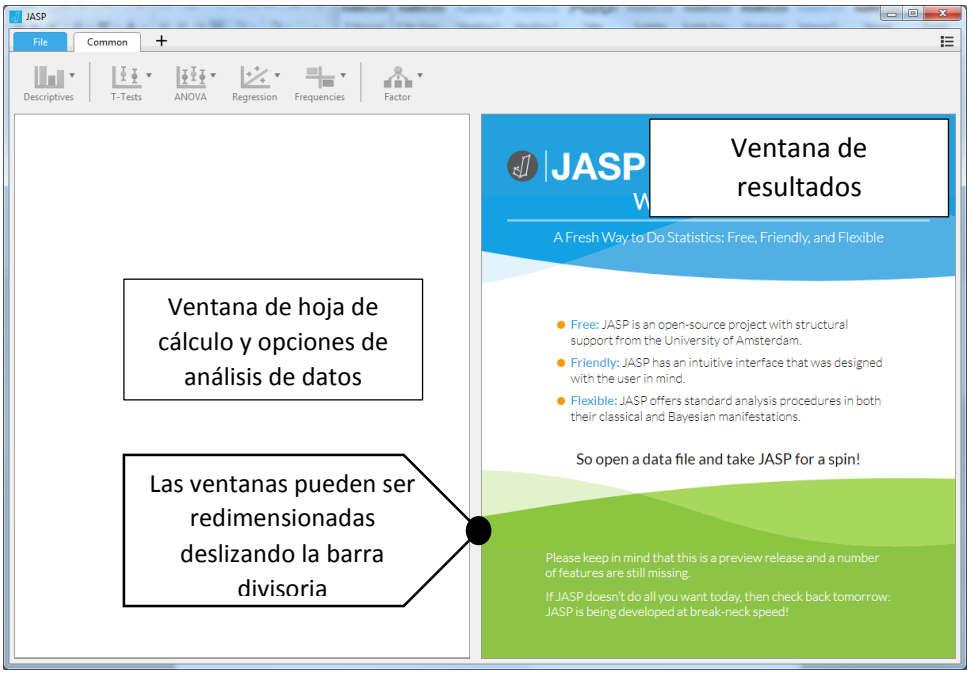
Este documento es una colección de capítulos independientes que cubren los análisis estadísticos más habituales (basados en el modelo frecuentista) utilizados por los estudiantes de ciencias biológicas. Los conjuntos de datos utilizados en este documento están disponibles para su descarga en http://bit.ly/2wlbMvf.

Dr. Mark Goss-Sampson
Centro para la Ciencia y la Medicina en el Deporte
Universidad de Greenwich
2018

[1] A Field. (2017) Discovering Statistics Using IBM SPSS Statistics (5th Ed.) SAGE Publications.
[2] D Moore, G McCabe, B Craig. (2011) Introduction to the Practice of Statistics (7th Ed.) W H Freeman.

USO DE LA INTERFAZ DE JASP

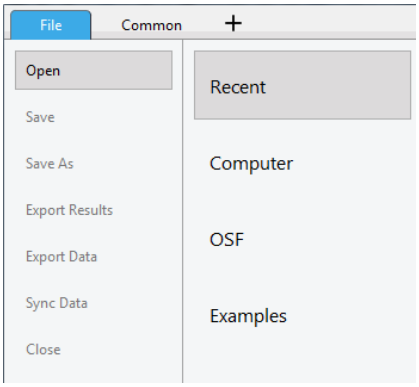
Abra JASP:

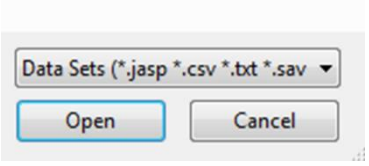


JASP tiene su propio formato .**jasp** pero acepta una gran variedad de formatos de conjuntos de datos, como:

- **.csv** (*comma separated values*, valores separados por comas), normalmente guardados en Excel
- **.txt** (texto plano) también puede ser guardado en Excel
- .**sav** (archivo de datos IBM SPSS)
- .**ods** (*open document spreadsheet*, hoja de cálculo de código abierto)

Haciendo clic en la pestaña «File» o en «So open a data file and take JASP for a spin» de la pantalla de inicio se pueden abrir los archivos recientes, buscar entre las carpetas del equipo y acceder al Open Science Framework (OSF), o a un amplio abanico de ejemplos incluidos en JASP.





Todos los archivos deben incluir una etiqueta de encabezado en la primera fila. Una vez cargado, el conjunto de datos aparece en la ventana izquierda:

	Game	Country code	Number of England Injuries
1	1	France	7
2	2	Tonga	4
3	3	New Zealand	2
4	4	France	5
5	5	Tonga	1
6	6	Wales	2
7	7	Wales	5
8	8	New Zealand	4
9	9	Wales	4
10	10	Tonga	3
11	11	Wales	5
12	12	Wales	3
13	13	France	6

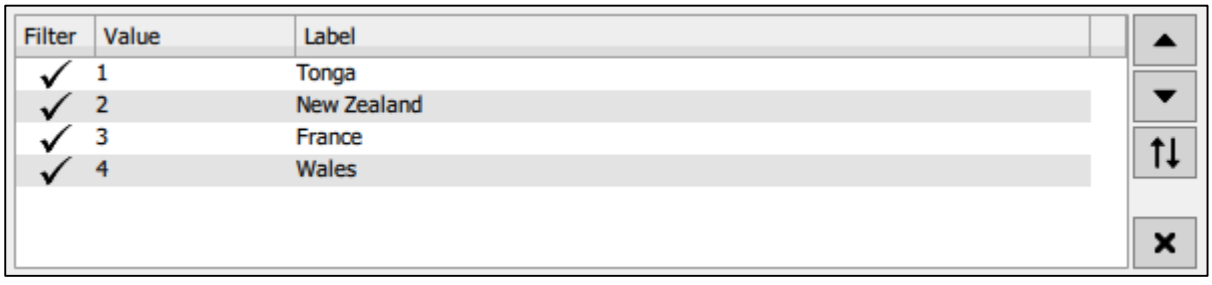
En conjuntos de datos grandes, el icono de la mano permite desplazarse fácilmente por las mismas.

Al importar, JASP trata de asignar de manera automática los datos a los diferentes tipos de variables:

Nominal  **Ordinal**  **Continua** 

Si JASP ha identificado incorrectamente el tipo de dato, solo hay que hacer clic sobre el icono apropiado en el título de columna para cambiarlo al formato correcto.

Si se han codificado los datos, se puede clicar sobre el nombre de variable para abrir la ventana siguiente que permite etiquetar cada código. Estas etiquetas reemplazan los códigos en la visualización de la hoja de cálculo. Si se guarda este documento como archivo .**jasp**, estos códigos, así como todos los análisis y las notas, se guardarán automáticamente. Esto permite que el análisis de datos sea totalmente reproducible.

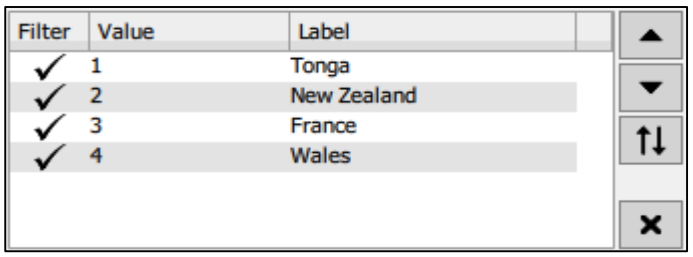


En esta ventana también se puede llevar a cabo un filtrado simple de datos; por ejemplo, si se deselecciona la etiqueta «Wales», no se usará en los análisis subsiguientes.

 Clicando en este icono de la ventana de la hoja de cálculo se abre un conjunto de opciones de filtrado de datos mucho más completo:

El uso de esta opción no se describe en este documento. Para información detallada sobre el uso de filtros más complejos, consulte el siguiente enlace: https://jasp-stats.org/2018/06/27/how-to-filter-your-data-in-jasp/

Por defecto, JASP grafica los datos según el valor (p. ej., 1-4). El orden puede cambiarse seleccionando la etiqueta y moviéndola arriba o abajo usando los cursores pertinentes:

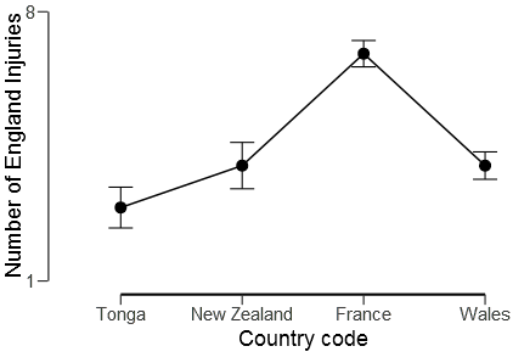


Mover arriba

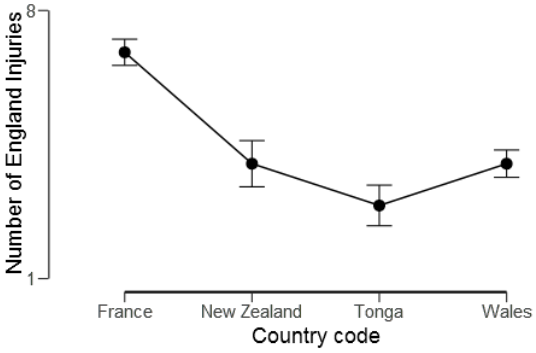
Mover abajo

Invertir el orden

Cerrar

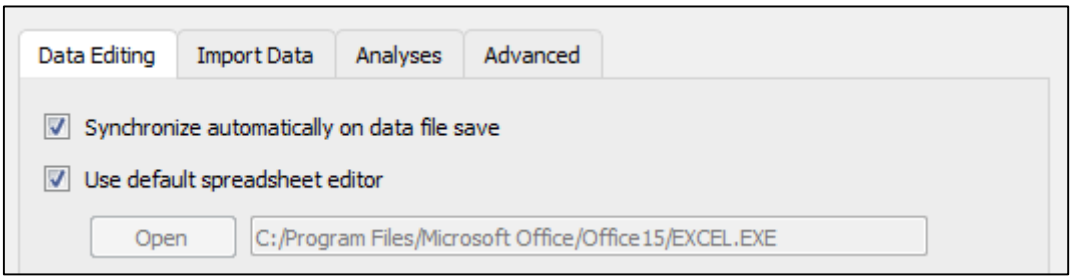


Filter	Value	Label
✓	1	Tonga
✓	2	New Zealand
✓	3	France
✓	4	Wales



Filter	Value	Label
✓	3	France
✓	2	New Zealand
✓	1	Tonga
✓	4	Wales

Si se precisa editar los datos en la hoja de cálculo, basta con hacer doble clic sobre la celda y el dato se abrirá en la hoja de cálculo original, p. ej., en Excel. Se puede cambiar la opción del editor de hojas de cálculo que se utiliza clicando sobre el icono  en la esquina superior derecha de la ventana de JASP y seleccionando «**Preferences**».



En esta ventana se puede cambiar la opción de la hoja de cálculo a SPSS, ODS, etc. Volveremos sobre las preferencias más adelante.

Una vez editados los datos y guardada la hoja de cálculo original, JASP se actualizará automáticamente para reflejar los cambios que se hayan realizado, siempre que no se haya modificado el nombre del archivo.

MENÚ DE ANÁLISIS DE JASP

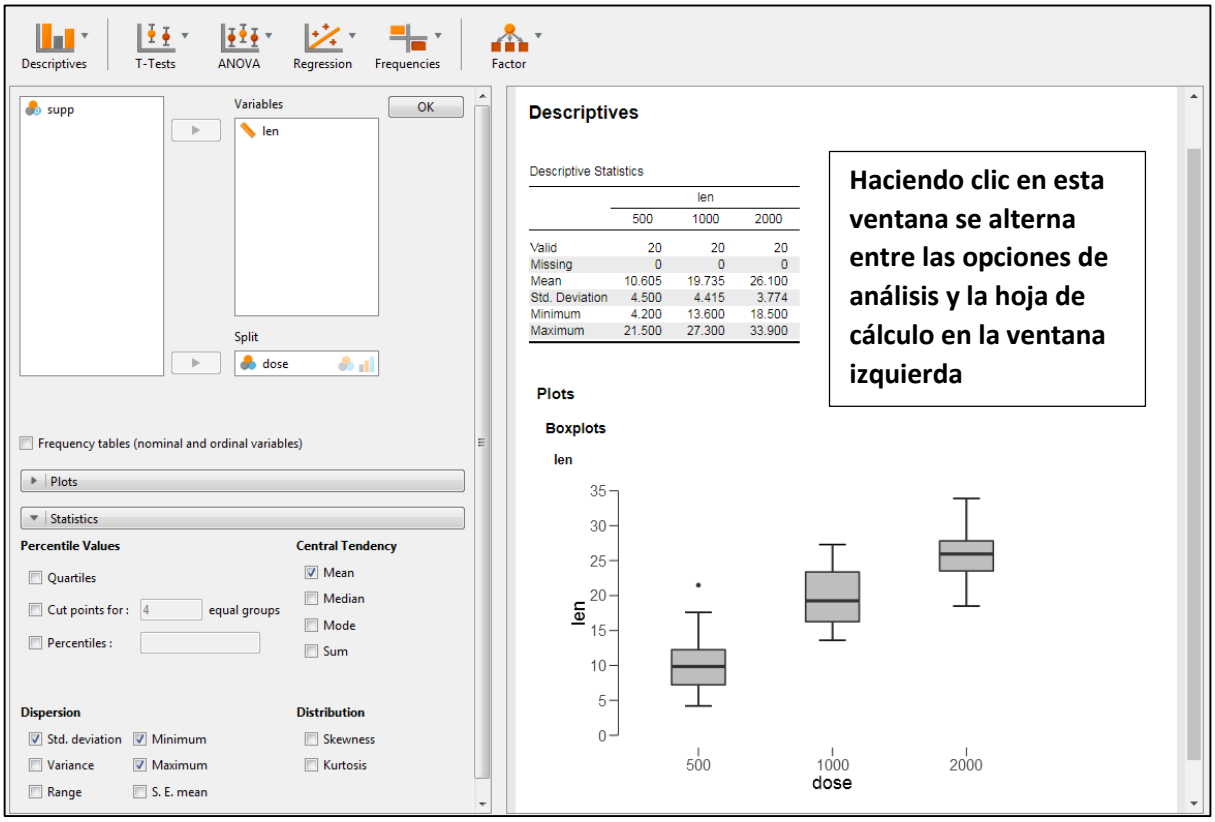
Se puede acceder a las opciones de análisis más **comunes** desde la barra de herramientas principal. Actualmente (v0.9.0.1), ofrece las siguientes pruebas basadas en el modelo frecuentista (estadística más habitual) y las alternativas bayesianas siguientes:

Descriptivas • Estadística descriptiva • Análisis de fiabilidad*	**Regresión** • Correlación • Regresión lineal • Regresión logística
Pruebas t • Para dos muestras independientes • Para dos muestras apareadas • Para una muestra única	**Frecuencias** • Test binomial • Test multinomial • Tablas de contingencia • Regresión log-lineal*
ANOVA • Medidas independientes • Medidas repetidas • ANCOVA*	**Análisis Factorial** • Análisis de componentes principales (ACP, PCA en inglés)* • Análisis factorial exploratorio (AFE, EFA en inglés)*

* No se trata en el presente documento

Clicando sobre el icono + del menú superior se puede acceder a las opciones avanzadas, incluyendo análisis de redes, metaanálisis, modelos de ecuaciones estructurales y estadística bayesiana.

Tras seleccionar el análisis requerido, todas las opciones estadísticas posibles aparecen en la ventana izquierda y los resultados se muestran en la ventana derecha.



Si se sitúa el cursor encima de «Results», aparece el icono ▼ y, clicando, se puede acceder a varias opciones que incluyen:

- **Remove all.** Elimina todos los análisis de la ventana de resultados.
- **Remove.** Elimina los análisis seleccionados.
- **Collapse.** Oculta el resultado.
- **Add notes.** Añade notas a cada resultado.
- **Copy**. Copiar.
- **Copy special (LaTeX code).** Copiado especial (código LaTeX).
- **Save image as.** Guardar imagen como.

La opción «Add notes» permite añadir fácilmente anotaciones a los resultados y exportarlos a un archivo HTML seleccionando «File» ⟶ «Export results».

ANOVA - Number of England Injuries

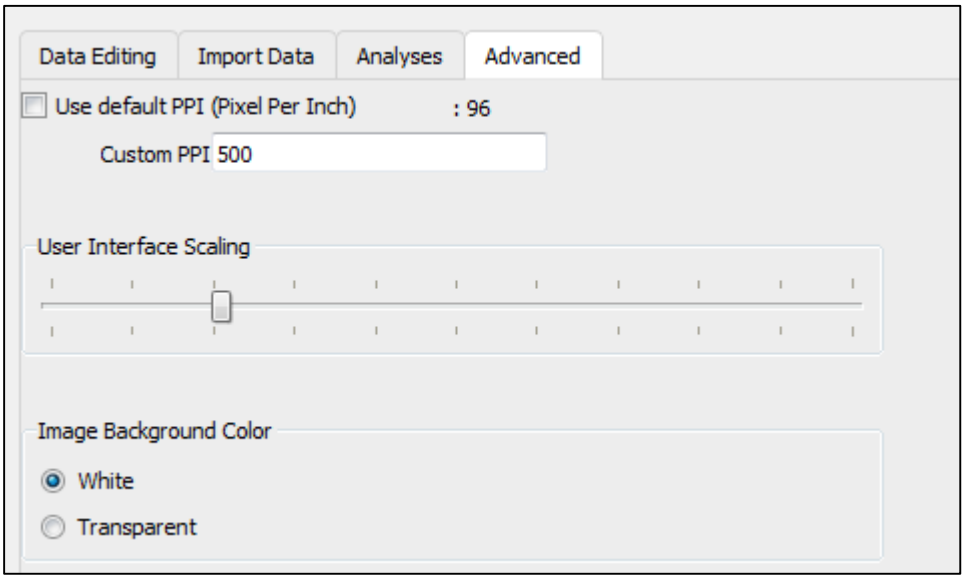
Cases	Sum of Squares	df	Mean Square	F	p
Country code	97.09	3	32.364	13.23	< .001
Residual	97.82	40	2.445		

Note. Type III Sum of Squares

One way ANOVA of injuries received by England rugby players against Tonga, New Zealand, France and Wales

Se puede cambiar el tamaño de todas las tablas y los gráficos usando ctrl+ (aumentar) ctrl- (reducir) ctrl= (volver al tamaño por defecto). Los gráficos también pueden ser redimensionados arrastrando la esquina inferior derecha del gráfico.

Como se ha mencionado anteriormente, todas las tablas y figuras cumplen con el estándar APA y pueden copiarse directamente en cualquier otro documento. Desde la v0.9.2, todas las imágenes pueden ser copiadas o guardadas con fondo blanco o transparente. Esto se puede seleccionar en «Preferences» ⟶ «Advanced»:



En la misma ventana también se puede cambiar el tamaño de la fuente de la interfaz de la hoja de cálculo mediante el controlador de escala de interfaz de usuario («User Interface Scaling»).

	Trial 1	Trial 2	Trial 3
1	12	12	12
2	36	26	16
3	27	27	27

	Trial 1	Trial 2	Trial 3
1	12	12	12
2	36	26	16
3	27	27	27

Un último consejo en relación con las preferencias («Preferences»): para que las tablas estén menos saturadas se puede ajustar el número de decimales que se muestran, así como mostrar los valores p exactos; por ejemplo, de p < 0,001 a p < 0,00084.

Hay muchos más recursos sobre el uso de JASP en el sitio web https://jasp-stats.org/

ESTADÍSTICA DESCRIPTIVA

Es muy difícil para el lector visualizar o hacer inferencias a partir de una presentación de los datos brutos. La estadística descriptiva y los gráficos relacionados son un modo conciso de describir y resumir los datos, pero no prueban ninguna hipótesis. Hay distintos tipos de estadísticos que se pueden usar para describir los datos:

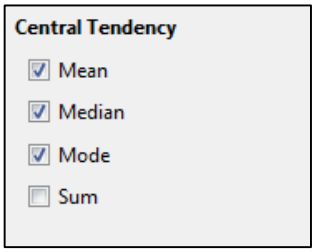
- Medidas de tendencia central.
- Medidas de dispersión.
- Percentiles.
- Medidas de distribución.
- Gráficos descriptivos.

Para estudiar estas medidas, cargue **Descriptive data.csv** en JASP. Vaya a «Descriptives» ⟶ «Descriptive statistics» y traslade los datos variables a la caja «Variables» de la derecha.

TENDENCIA CENTRAL

Puede ser definida como la tendencia de las variables a agruparse alrededor de un valor central. Las tres formas de describir este valor central son la media, la mediana o la moda. Si se considera el total de la población, se utiliza el término media, mediana o moda poblacionales. Si se analiza una muestra / subconjunto de población, se utiliza el término media, mediana o moda muestrales. Las medidas de tendencia central se mueven hacia un valor constante cuando el tamaño de la muestra es suficiente para ser representativa de la población.

En las opciones estadísticas, hay que asegurarse de que todo está deseleccionado excepto la media, la mediana y la moda.



Descriptive Statistics

	Variable
Valid	810
Missing	0
Mean	17.71
Median	17.90
Mode	20.00

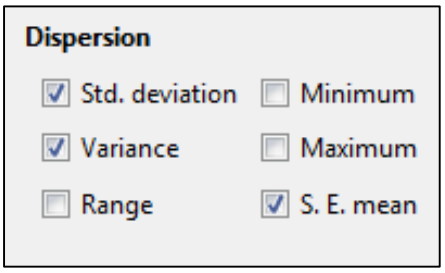
La ***media***, **M o $\bar{x}$** (17,71), es igual a la suma de todos los valores dividida por el número de valores de la tabla. Es decir, el promedio de los valores. Se usa para describir datos continuos. Proporciona un modelo estadístico simple del centro de la distribución de los valores y es una estimación teórica del "valor típico". Sin embargo, puede quedar fuertemente influenciada por valores "extremos".

La ***mediana***, **Mdn** (17,9) es el valor central en un conjunto de datos que ha sido ordenado del valor más pequeño al más grande, y es la medida tradicional utilizada para datos continuos ordinales o continuos no paramétricos. Es menos sensible a los valores atípicos y a las distribuciones asimétricas.

La ***moda*** (20,0) es el valor más frecuente en el conjunto de datos y normalmente la barra más alta en un histograma de una distribución.

DISPERSIÓN

En las opciones estadísticas, asegúrese de que todo está deseleccionado menos la desviación estándar («Std. deviation»), la varianza («Variance») y el error estándar de la media («S. E. mean»).



Descriptive Statistics

	Variable
Valid	810
Missing	0
Std. Error of Mean	0.24
Std. Deviation	6.94
Variance	48.10

La ***desviación estándar (Standard deviation), S o SD*** (6,94) se usa para cuantificar el grado de dispersión de los datos respecto a la media. Una desviación estándar baja indica que los valores están cerca de la media, mientras que una desviación estándar alta indica que el rango de dispersión de los valores es más amplio.

La ***varianza (Variance)*** (S^2 = 48,1) es otra estimación de hasta qué punto los datos se separan de la media. También es el cuadrado de la desviación estándar.

El ***error estándar de la media (The standard error of the mean), SE*** (0,24) es una medida que expresa hasta qué punto se espera que la media obtenida a partir de una muestra difiera de la media real de la población. A medida que aumenta el tamaño de la muestra, el SE disminuye en comparación con la S y la verdadera media de la población se conoce con mayor especificidad.

Los **intervalos de confianza (CI)**, aunque no se muestren en los resultados de la estadística descriptiva, se usan en muchos otros test estadísticos. Cuando se toma una muestra de la población para obtener una estimación de la media, los intervalos de confianza representan un rango de valores dentro del cual se está n% seguro de que se incluye la verdadera media. Un CI del 95% es, por lo tanto, un rango de valores del que uno puede estar un 95% seguro de que contiene la verdadera media de la población. Esto **no** es lo mismo que un rango que contenga el 95% de **todos** los valores.

Por ejemplo, en una distribución normal, se espera que el 95% de los datos tenga una SD de ± 1,96 respecto a la media, y el 99% una SD de ± 2,576.

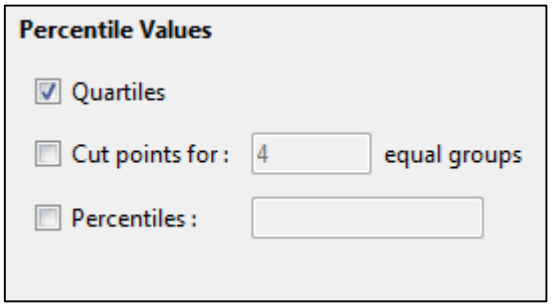
95% CI = M ± 1,96 * el error estándar de la media.

Basándonos en los datos a los que nos hemos referido hasta ahora, M = 17,71; SE = 0,24; esto será 17,71 ± (1,96 * 0,24) o 17,71 ± 0,47.

Por tanto, el CI del 95% para este conjunto de datos es 17,24-18,18 y sugiere que la media real se halla dentro de este rango en un 95% de las ocasiones.

CUARTILES

En las opciones estadísticas, asegúrese de que todo está deseleccionado excepto los cuartiles.



Descriptive Statistics

	Variable
Valid	810
Missing	0
25th percentile	13.05
50th percentile	17.90
75th percentile	22.30

Los cuartiles son los puntos en los cuales los conjuntos de datos se dividen en 4 partes iguales, a partir de los valores de las medianas una vez ordenados los datos. Por ejemplo, para este conjunto de datos:

1	1	2	2	3	3	4	4	4	4	5	5	5	6	7	8	8	9	10	10	10
				25%						50%					75%						

El valor de la mediana que divide los datos por el 50% = percentil 50 = 5.

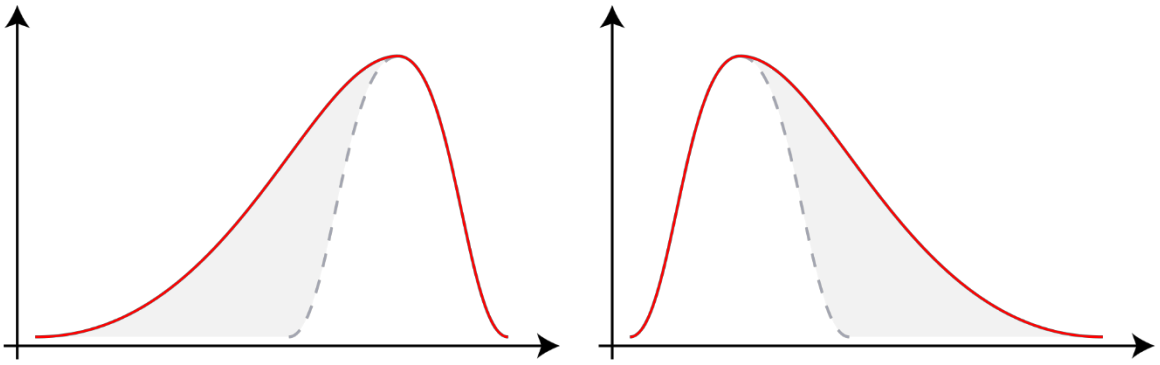
El valor de la mediana del lado izquierdo = percentil 25 = 3.

El valor de la mediana del lado derecho = percentil 75 = 8.

A partir de esto, se puede calcular el rango intercuartil (IQR), esto es, la diferencia entre los percentiles 75 y 25, es decir, 5. Estos valores se utilizan para construir, más adelante, los gráficos de caja descriptivos.

DISTRIBUCIÓN

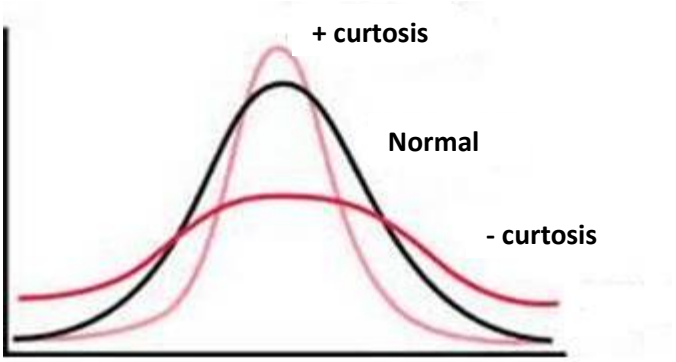
La asimetría describe el desplazamiento de la distribución respecto a una distribución normal. Una asimetría negativa muestra que la moda se mueve hacia la derecha dando como resultado una cola izquierda dominante. Una asimetría positiva muestra que la moda se desplaza hacia la izquierda resultando en una cola derecha dominante.



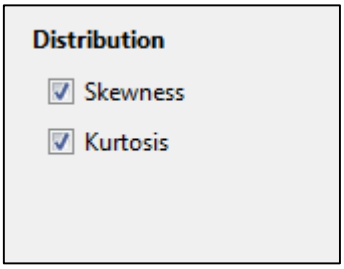
Asimetría negativa

Asimetría positiva

La curtosis describe cuán pronunciadas o suaves son las colas. Una curtosis positiva da como resultado un "vértice" de la distribución más agudo, con colas más pronunciadas (largas); en cambio, una curtosis negativa muestra una distribución mucho más uniforme o aplanada, con colas suaves (cortas).



En las opciones estadísticas, asegúrese de que todo está deseleccionado excepto la asimetría (*skewness*) y la curtosis (*kurtosis*).



Descriptive Statistics

	Variable
Valid	810
Missing	0
Skewness	−0.004
Std. Error of Skewness	0.086
Kurtosis	−0.410
Std. Error of Kurtosis	0.172

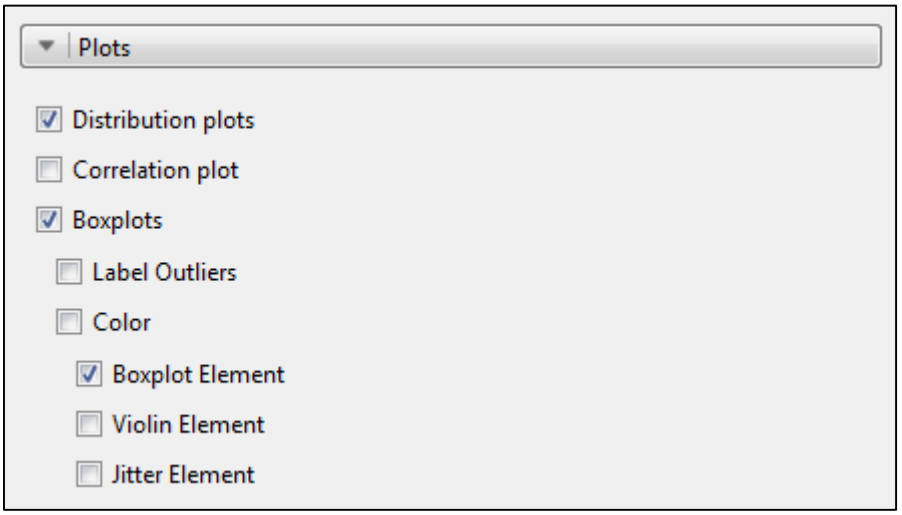
Podemos usar los resultados descriptivos para calcular las asimetrías y las curtosis. Para una distribución normal, ambos valores deberían ser cercanos a cero (ver "Exploración de la integridad de los datos en JASP" para más detalles).

GRÁFICOS DESCRIPTIVOS EN JASP

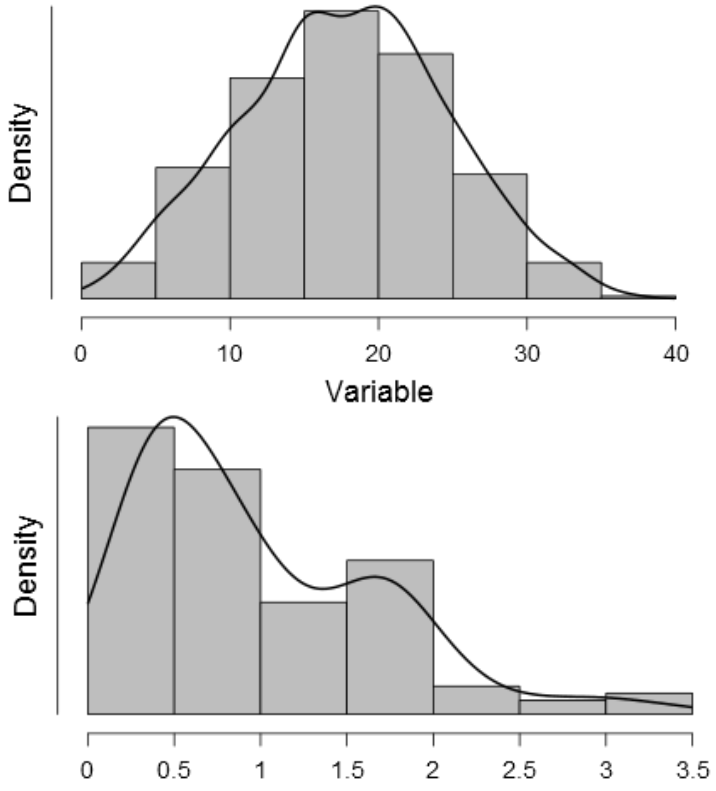
Actualmente, JASP produce tres tipos de gráficos descriptivos:

- Gráficos de distribución («Distribution plots»).
- Gráficos de correlación («Correlation plot»).
- Gráficos de caja, con 3 opciones («Boxplots»):
  - Elemento gráfico de caja («Boxplot Element»).
  - Elemento violín («Violin Element»).
  - Elemento jitter («Jitter Element»).

De nuevo, usando **Descriptive data.csv**, una vez introducidas las variables en la caja «Variables», vaya a las opciones estadísticas y debajo de «Plots», seleccione «Distribution plots» y «Boxplots» – «Boxplot Element».

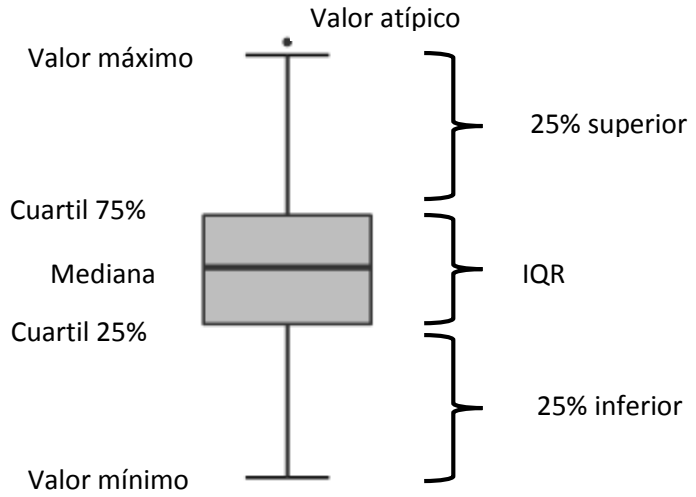


El gráfico de distribución («Distribution plots») está basado en una división de los datos en intervalos de frecuencia, que se superpone a la curva de distribución. Como se ha dicho anteriormente, la barra más alta es la moda (el valor más frecuente en el conjunto de datos). En este caso, la curva parece casi simétrica, lo que sugiere que los datos se distribuyen de un modo aproximadamente normal. El segundo gráfico de distribución es de otro conjunto de datos, que muestran una asimetría positiva.



Los gráficos de caja muestran los estadísticos descritos anteriormente en un gráfico:

- Mediana.
- Cuartiles del 25% y el 75%.
- Rango intercuartil (IQR), o sea valores del cuartil de 75%-25%.
- Valores máximos y mínimos representados una vez excluidos los valores atípicos.
- Si se solicita, también se muestran los valores atípicos.



Vuelva a las opciones estadísticas. En «Descriptive plots», marque «Boxplot Element» y «Violin Element», y vea cómo ha cambiado el gráfico. Tras ello, seleccione los elementos «Boxplot Element», «Violin Element» y «Jitter Element». El gráfico de violín ha adoptado la curva suavizada del gráfico de distribución, girándola 90° y superponiéndola en el gráfico de caja. El gráfico *jitter* ha agregado, además, todos los puntos de los datos.

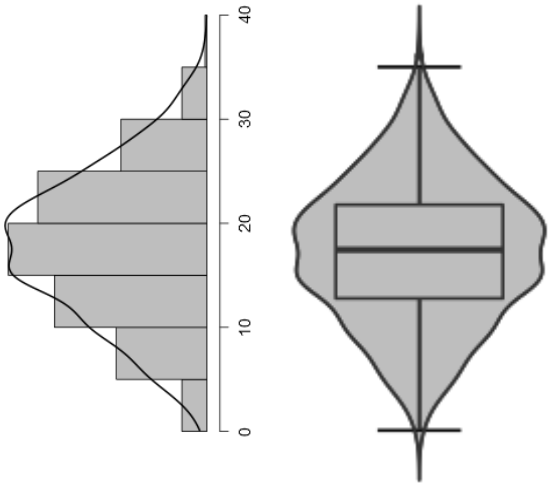


Gráfico de caja + gráfico de violín

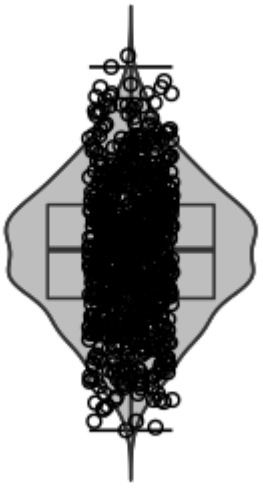


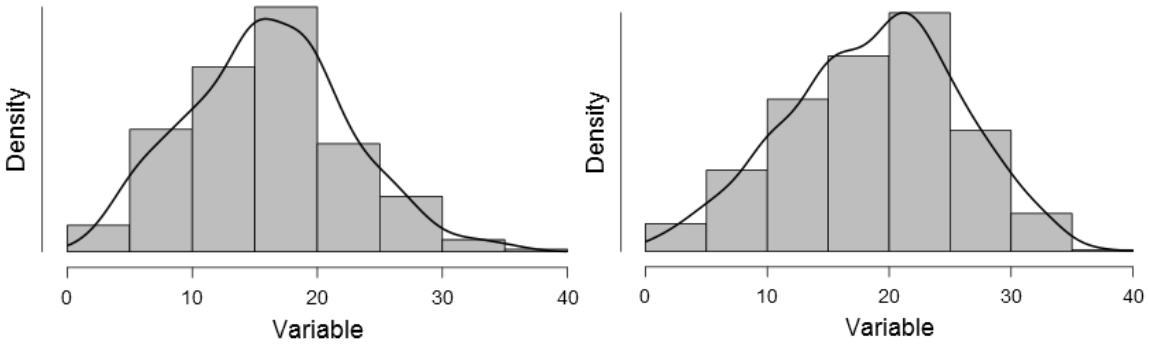
Gráfico de caja + gráfico de violín + *jitter*

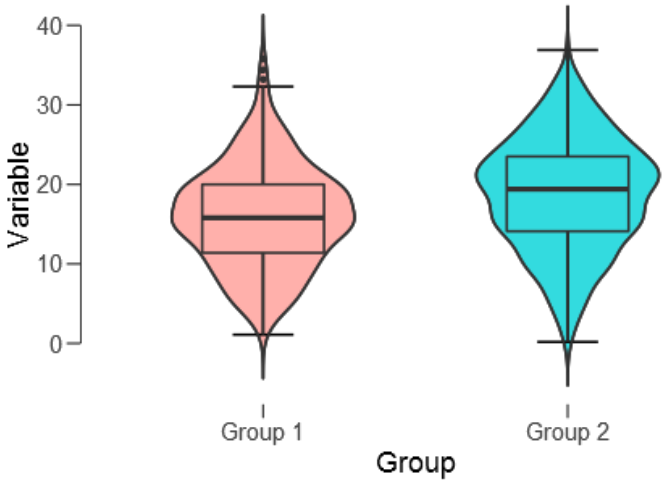
DIVISIÓN DE LOS ARCHIVOS DE DATOS

Si existe una variable de agrupación (categórica u ordinal), se pueden elaborar gráficos y estadísticos descriptivos para cada grupo. Usando **Descriptive data.csv** con las variables en la caja «Variables», añada una variable de agrupación a la caja «Split». El resultado se mostrará como sigue:

Descriptive Statistics ▼

	Variable	
	Group 1	Group 2
Valid	315	495
Missing	0	0
Mean	16.021	18.787
Median	15.800	19.400
Mode	20.000	20.200
Std. Deviation	6.424	7.040
Variance	41.269	49.556
Skewness	0.200	−0.176
Std. Error of Skewness	0.137	0.110
Kurtosis	−0.101	−0.397
Std. Error of Kurtosis	0.274	0.219
Minimum	1.100	0.200
Maximum	35.800	36.900





EXPLORACIÓN DE LA INTEGRIDAD DE LOS DATOS

Los datos obtenidos a partir de una muestra se utilizan para estimar los parámetros de la población, teniendo en cuenta que un parámetro es una característica medible de una población, como la media, la desviación estándar, el error estándar o los intervalos de confianza, etc.

¿Cuál es la diferencia entre un estadístico y un parámetro? Supongamos que realizamos una encuesta sobre la calidad del bar estudiantil a un grupo de estudiantes seleccionados aleatoriamente, y que el 75% de los mismos se muestra satisfecho. Esto es un **estadístico** muestral ya que solo se encuestaría a una muestra de la población. Se calcularía lo que la población probablemente haría en base a la muestra. Si se preguntara a **todos** los estudiantes de la universidad y un 90% se declarase satisfecho se obtendría un **parámetro**, ya que se habría encuestado al total de la población universitaria.

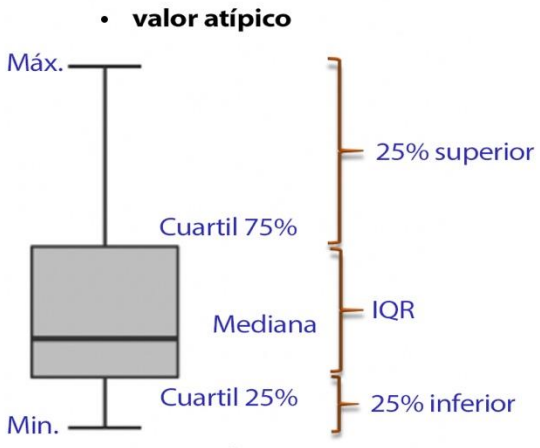
El sesgo puede ser definido como la tendencia de una medición a sobreestimar –o subestimar– el valor de un parámetro de una población. Hay muchos tipos de sesgo que pueden aparecer en el diseño de la investigación y la recogida de datos, entre ellos:

- Sesgo en la selección de participantes –algunos son más propensos que otros a ser seleccionados para el estudio–.
- Sesgo en la exclusión de participantes –por la exclusión sistemática de ciertos individuos–.
- Sesgo analítico –debido al modo en el que se evalúan los resultados en el estudio–.

Sin embargo, el sesgo estadístico puede afectar: a) a la estimación de los parámetros; b) a los errores estándar y los intervalos de confianza; o c) a los test estadísticos y los valores *p*. Entonces, ¿cómo podemos comprobar si hay sesgo?

¿SON SUS DATOS CORRECTOS?

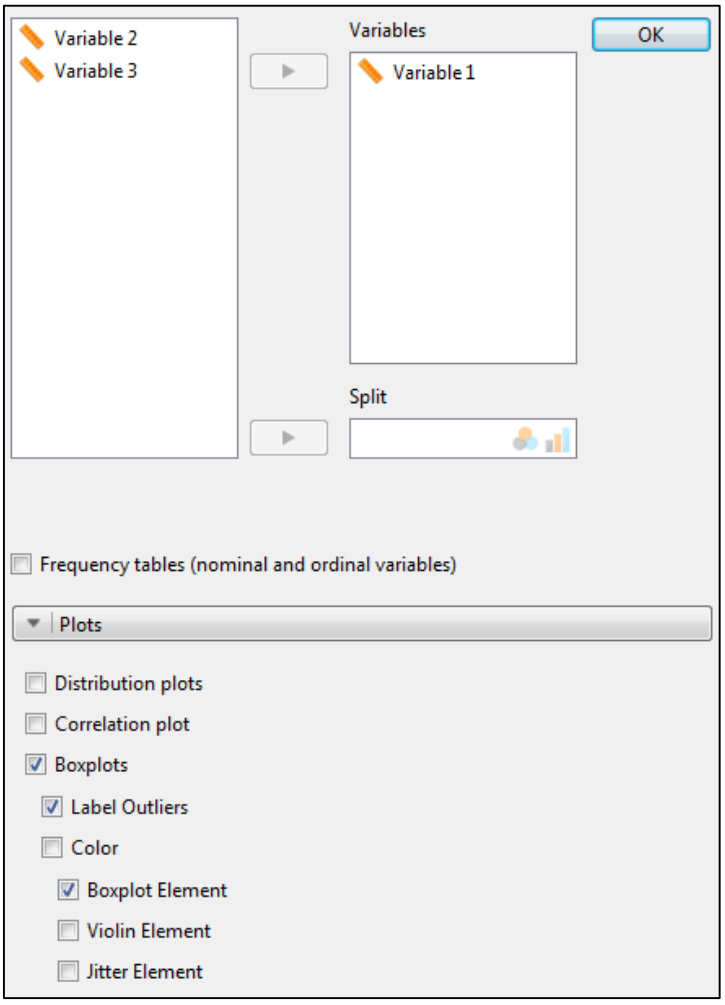
Los valores atípicos son puntos de los datos que se encuentran anormalmente lejos de otros puntos. Un valor atípico puede deberse a distintos motivos, como errores en la introducción de datos o errores analíticos cometidos en el momento de la recogida de los datos. Los gráficos de caja permiten visualizar fácilmente estos puntos de datos en los que los valores atípicos están fuera del límite del cuartil superior (75% + 1,5 * IQR) o inferior (25% - 1,5 * IQR).



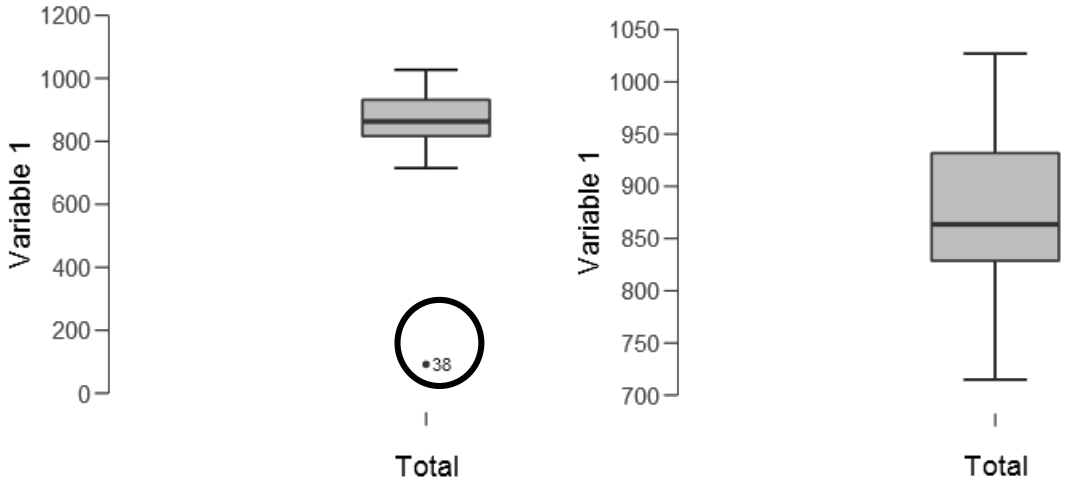
El gráfico de caja muestra:

- Mediana.
- Cuartiles del 25% y 75%.
- IQR –Rango intercuartil–.
- Valores máximos y mínimos representados una vez excluidos los valores atípicos.
- Si se solicita, también se muestran los valores atípicos.

Cargue **Exploring Data.csv** en JASP. En «Descriptives» ⟶ «Descriptive statistics», añada la variable 1 a la caja «Variables». En gráficos («Plots»), seleccione gráficos de caja («Boxplots»), etiquetar valores atípicos («Label Outliers») y elemento gráfico de caja («Boxplot Element»).



El gráfico de caja resultante que se muestra hacia la izquierda se ve muy comprimido y se puede observar un valor atípico evidente en la fila 38 del conjunto de datos. Esto se puede deber a un error en la introducción de los datos, al introducir 91,7 en lugar de 917. El gráfico de caja de la derecha muestra los datos "limpios".



Cómo se maneje un valor atípico dependerá de su causa. La mayoría de las pruebas paramétricas son muy sensibles a los valores atípicos, mientras que las no paramétricas generalmente no lo son.

¿Corregirlo? – Comprobamos los datos originales para asegurar que no se trate de un error de introducción de los datos; si es así, lo corregimos y ejecutamos el análisis de nuevo.

¿Mantenerlo? – Incluso en conjuntos de datos con distribución normal se pueden esperar datos atípicos para muestras grandes y no deben descartarse automáticamente si se da el caso.

¿Eliminarlo? – Es una práctica controvertida en conjuntos de datos pequeños en los que no se puede asumir una distribución normal. Pueden excluirse los valores atípicos debidos a un error de lectura en el instrumento, pero primero deben verificarse.

¿Reemplazarlo? – También conocida como *winsorización*. Esta técnica reemplaza los valores atípicos por los valores máximos y/o mínimos relevantes, hallados tras excluir el valor atípico.

Cualquier método que se utilice debe estar justificado por la metodología estadística adoptada y los análisis subsiguientes.

HACEMOS MUCHAS SUPOSICIONES SOBRE NUESTROS DATOS

Cuando usamos pruebas paramétricas, partimos de una serie de suposiciones sobre nuestros datos y si se violan estos supuestos se producirá un sesgo, en particular:

- Normalidad.
- Homogeneidad de la varianza u homocedasticidad.

Muchas pruebas estadísticas son en realidad un conjunto de pruebas "ómnibus", algunas de las cuales verifican estos supuestos.

PRUEBA DEL SUPUESTO DE NORMALIDAD

La normalidad no significa necesariamente que los datos estén normalmente distribuidos *per se*, sino si el conjunto de datos puede estar bien modelado por una distribución normal. La normalidad puede explorarse por distintas vías:

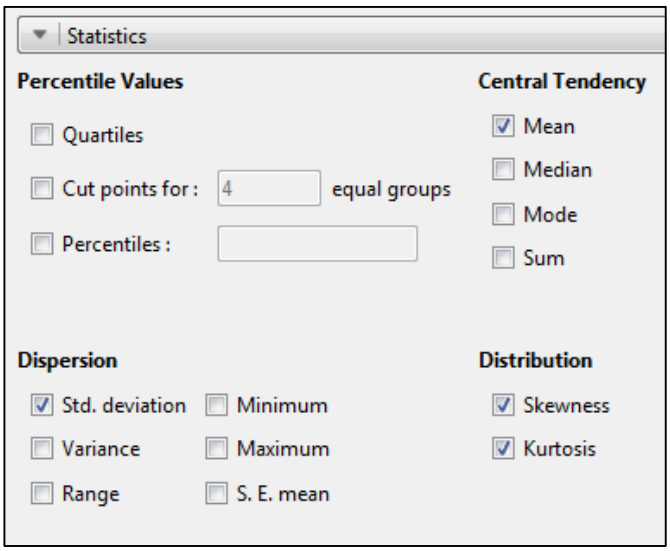
- Numéricamente.
- Visualmente / gráficamente.
- Estadísticamente.

Numéricamente, podemos usar los resultados descriptivos para calcular la asimetría y la curtosis. En una distribución normal, ambos valores deberían ser cercanos a cero. Para determinar la significación de la asimetría o la curtosis, calculamos las puntuaciones z (*z-scores*) dividiéndolas por sus errores estándar respectivos:

$$\text{Asimetría}_{Z} = \frac{\text{asimetría}}{\text{error estándar de la asimetría}} \qquad \text{Curtosis}_{Z} = \frac{\text{curtosis}}{\text{error estándar de la curtosis}}$$

Significación de la puntuación z: $p < 0{,}05$ si $z > 1{,}96$ $p < 0{,}01$ si $z > 2{,}58$ $p < 0{,}001$ si $z > 3{,}29$

Usando **Exploring data.csv**, vaya a «Descriptives» ⟶ «Descriptive statistics» y mueva la variable 3 a la caja «Variables»; en el menú desplegable de «Statistics», seleccione «Mean», «Std. Deviation», «Skewness» y «Kurtosis» tal como se muestra a continuación en la correspondiente tabla de resultados.



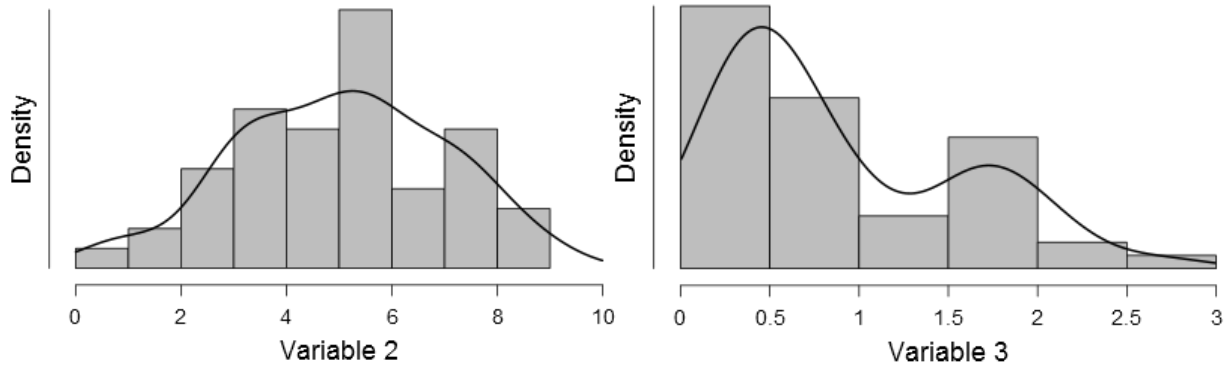
	Variable 3
Valid	50
Missing	0
Mean	0,893
Std. Deviation	0,673
Skewness	0,839
Std. Error of Skewness	0,337
Kurtosis	-0,407
Std. Error of Kurtosis	0,662

Se puede ver que la asimetría y la curtosis no son cercanas a 0. La asimetría positiva sugiere que los datos están más distribuidos hacia la izquierda (ver los gráficos a continuación), mientras que la curtosis negativa sugiere una distribución plana. Al calcular sus puntuaciones Z, se puede ver que los datos son asimétricos ($p < 0,05$).

$$\text{Asimetría}_{Z} = \frac{0,839}{0,337} = 2,49 \qquad \text{Curtosis}_{Z} = \frac{-0,407}{0,662} = 0,614$$

[Nótese, como advertencia, que la asimetría y la curtosis se muestran significativas en grandes conjuntos de datos, aunque la distribución sea normal.]

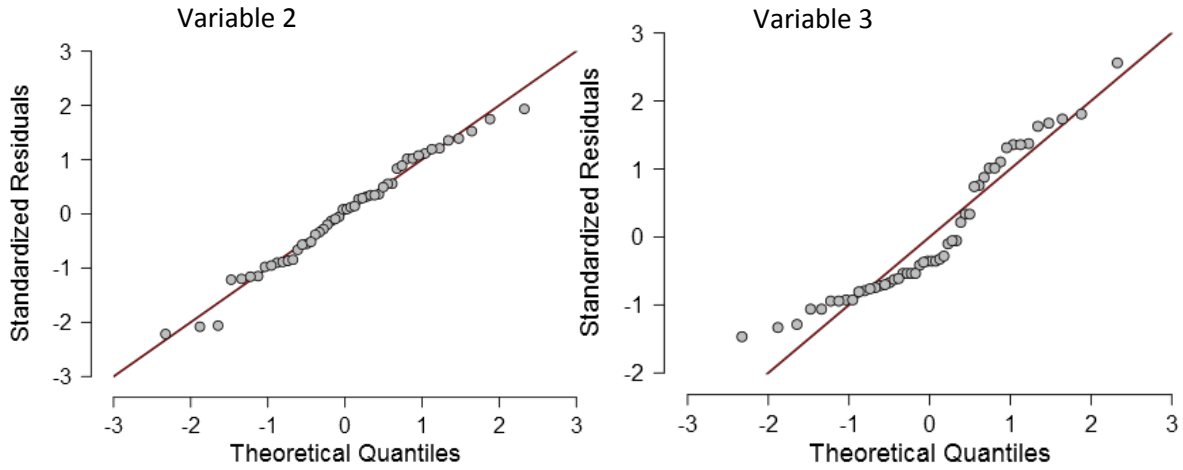
Ahora, añada la variable 2 a la caja «Variables» y, en «Plots», seleccione «Distribution plots». Esto proporcionará los dos gráficos siguientes:



Resulta fácil ver que la variable 2 tiene una distribución simétrica. La variable 3 presenta una asimetría hacia la izquierda, como confirma la puntuación Z de la asimetría.

Otro modo de comprobar gráficamente la normalidad es mediante un gráfico Q-Q. Este procedimiento forma parte de la comprobación de los supuestos de la **regresión** y el **ANOVA**. Los gráficos Q-Q muestran los cuantiles de los datos reales frente a los esperados para una distribución normal.

Si los datos se distribuyen normalmente, todos los puntos estarán cerca de la línea diagonal de referencia. Si los puntos "caen" por encima o por debajo de la línea, hay un problema con la curtosis. Si los puntos serpentean alrededor de la línea, entonces el problema es la asimetría. A continuación, se muestran los gráficos Q-Q para las variables 2 y 3. Compárense con los gráficos de distribución y las puntuaciones Z de asimetría / curtosis anteriores.



La prueba de Shapiro-Wilk es una forma estadística utilizada por JASP para verificar el supuesto de normalidad. Se utiliza en las pruebas t para dos muestras independientes (distribución de los dos grupos) y apareadas (distribución de diferencias entre pares). El test proporciona un valor de W, donde los valores pequeños indican que la muestra no está distribuida normalmente (la hipótesis nula de que la población está distribuida normalmente si sus valores están por debajo de un cierto umbral puede, por lo tanto, ser rechazada). La siguiente tabla es un ejemplo de la tabla de resultados de Shapiro-Wilk que no muestra ninguna desviación significativa de la normalidad en los 2 grupos.

Test de normalidad (Shapiro-Wilk)

		W	p
Variable 2	Control	0,971	0,691
	Test	0,961	0,408

Nota. Los resultados significativos sugieren una desviación de la normalidad.

La limitación más importante es que la prueba puede estar sesgada por el tamaño de la muestra. Cuanto mayor sea la muestra, mayor será la probabilidad de obtener un resultado estadísticamente significativo.

Probando el supuesto de normalidad. ¡Nota de advertencia!

Para que la mayoría de los test paramétricos sean fiables, uno de los supuestos es que los datos se distribuyen de manera **aproximadamente** normal. Una distribución normal alcanza su punto máximo

en el medio y es simétrica respecto a la media. No obstante, los datos no tienen que estar distribuidos de manera perfectamente normal para que los test sean fiables.

Entonces, ¿era necesario extendernos tanto sobre los test de normalidad?

El teorema del límite central establece que, a medida que el tamaño de la muestra aumenta –es decir, > 30 puntos de datos– la distribución de las medias muestrales se aproxima a una distribución normal. Por lo tanto, cuantos más puntos de datos se tengan, más normal parecerá la distribución y más se acercará la media de la muestra a la media de la población.

Los conjuntos de datos grandes pueden dar como resultado pruebas significativas de normalidad; es decir, mostrar Shapiro-Wilk o puntuaciones Z de asimetría y curtosis significativas cuando los gráficos de distribución parecen bastante normales. Y, al contrario, los conjuntos de datos pequeños reducirán la potencia estadística para detectar la no normalidad.

Sin embargo, los datos que definitivamente no cumplen con el supuesto de normalidad ofrecerán resultados deficientes en ciertos tipos de test (en concreto, aquellos que asumen que se debe cumplir con este supuesto). ¿Hasta qué punto deben ajustarse sus datos a una distribución normal? Para tomar una decisión en relación con este supuesto, es mejor observar los datos.

¿QUÉ HAGO SI MIS DATOS NO SE DISTRIBUYEN NORMALMENTE?

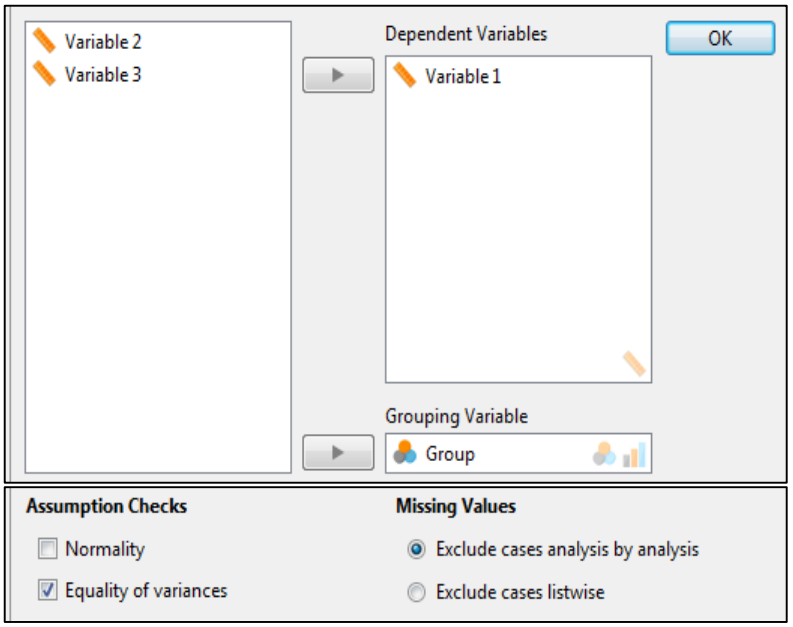
Se deben transformar los datos y realizar nuevamente comprobaciones de normalidad para los datos transformados. Las transformaciones comunes incluyen calcular el logaritmo o la raíz cuadrada de los datos.

Es mejor usar test no paramétricos, dado que se trata de pruebas de distribución libre y se pueden usar en lugar de su equivalente paramétrico.

PRUEBAS DE HOMOGENEIDAD DE LA VARIANZA

El test de Levene se usa frecuentemente para probar la hipótesis nula de que las varianzas en los diferentes grupos son iguales. El resultado del test (F) se reporta como valor de p; si no es significativo, se puede asumir que la hipótesis nula debe ser mantenida (que las varianzas son iguales); si el valor de p es significativo, entonces la implicación es que las varianzas son desiguales. El test de Levene se incluye en la **prueba t independiente** y el **ANOVA**, en JASP, como parte de la comprobación de los supuestos.

Usando **Exploring data.csv**, vaya a «T-Tests» ⟶ «Independent samples t-test», traslade la variable 1 a la caja «Variables», la variable Group a la caja «Grouping Variable» y marque «Assumption Checks» ⟶ «Equality of variances».

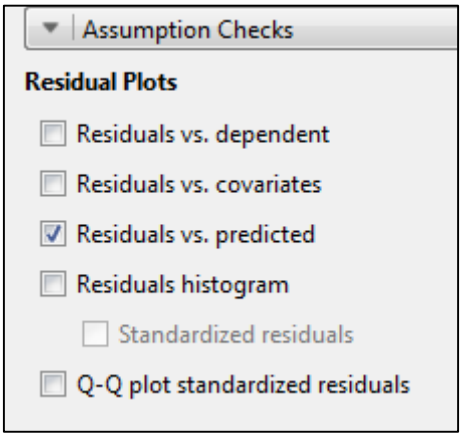


Test de igualdad de varianzas (test de Levene)

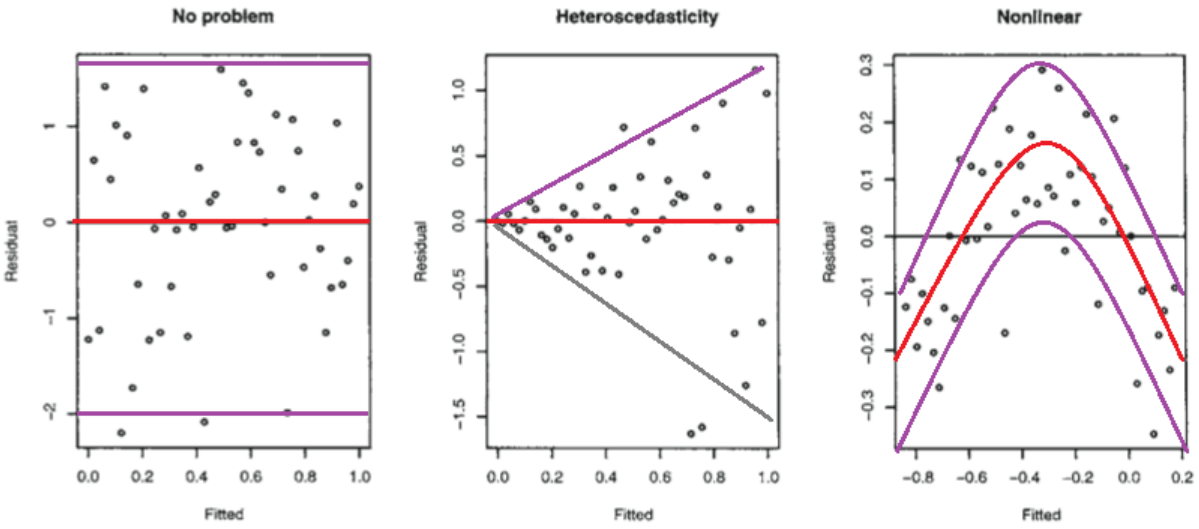
	F	df	p
Variable 1	0,218	1	0,643

En este caso, no hay diferencias significativas en la varianza entre los dos grupos: F (1) = 0,218, p = 0,643.

El supuesto de homocedasticidad (igualdad de varianza) es importante en los modelos **de regresión lineal**, como lo es la linealidad. Esta prueba asume que la varianza de los datos alrededor de la línea de regresión es la misma para todos los puntos de datos de las variables predictoras. La heterocedasticidad (la violación de la homocedasticidad) se presenta cuando la varianza difiere en los valores de una variable independiente. Esto se puede evaluar visualmente en una regresión lineal representando los residuos obtenidos en relación con los residuos predichos por el modelo.

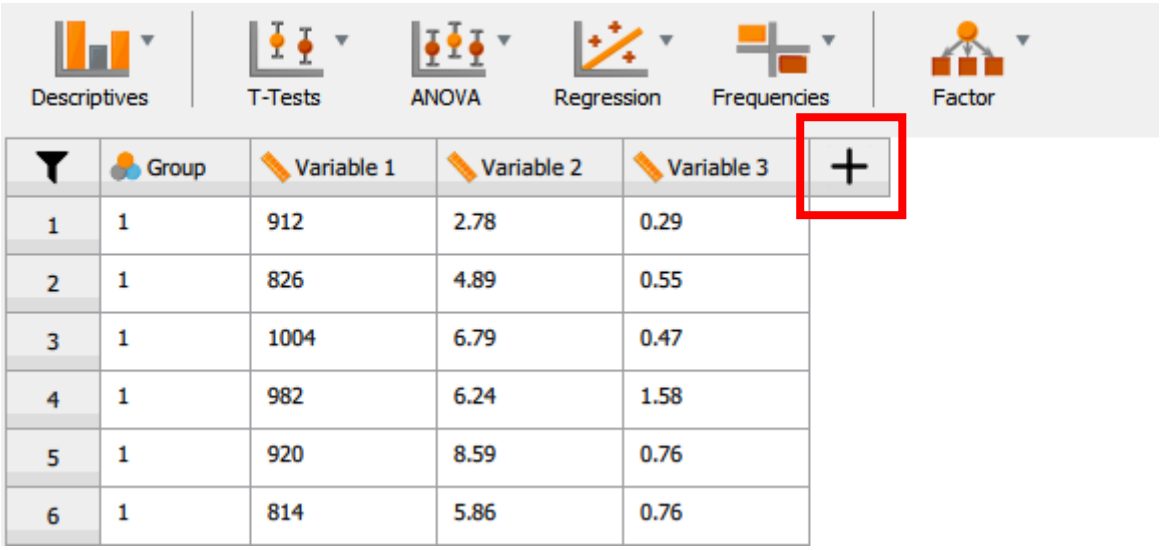


Si no se violan la homocedasticidad y la linealidad, no debería haber una relación entre lo que el modelo predice y sus errores, como muestra el gráfico de la izquierda. Cualquier tipo de canalización (gráfico del medio) sugiere que se ha violado la homocedasticidad y cualquier curva (gráfico de la derecha) sugiere que no se ha cumplido con el supuesto de linealidad.



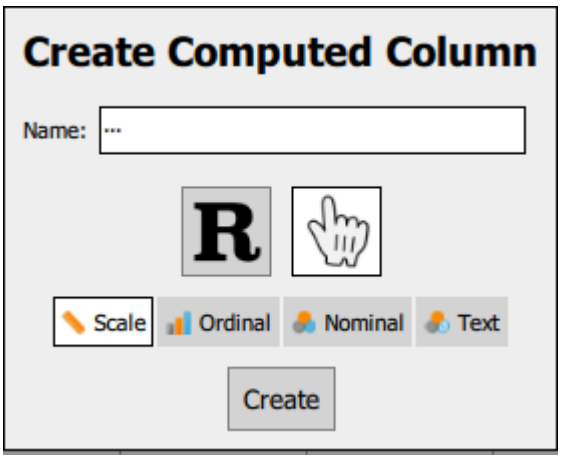
TRANSFORMACIÓN DE LOS DATOS

La capacidad para calcular nuevas variables o transformar datos fue introducida en la versión 0.9.1. En algunos casos, puede ser útil calcular las diferencias entre medidas repetidas o, para que un conjunto de datos esté distribuido de un modo más normal, aplicar una transformación logarítmica, por ejemplo. Cuando un conjunto de datos esté cargado, habrá un signo más (**+**) al final de las columnas.



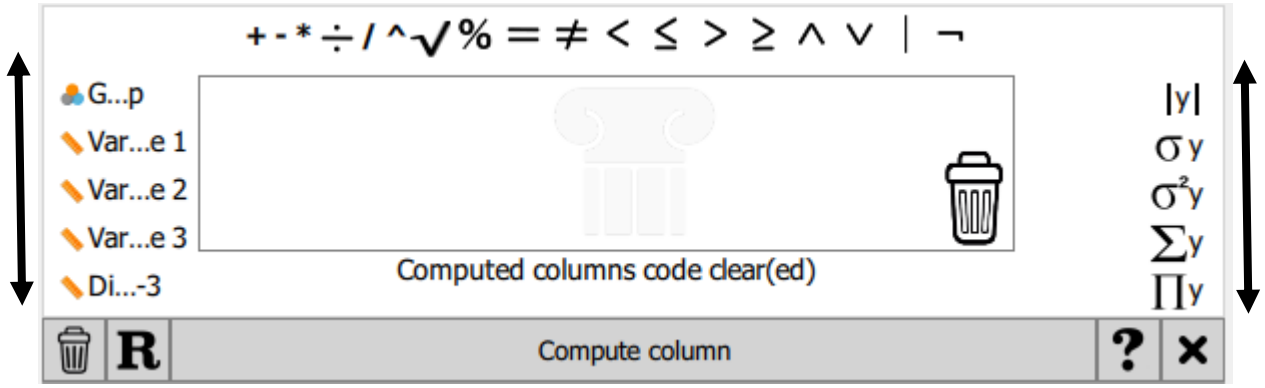
Haciendo clic en **+** se abre un pequeño cuadro de diálogo en el que se puede:

- Introducir el nombre de una nueva variable o de la variable transformada.
- Seleccionar si se introduce el código R directamente o se usan los comandos integrados en JASP.
- Seleccionar qué tipo de dato se requiere.

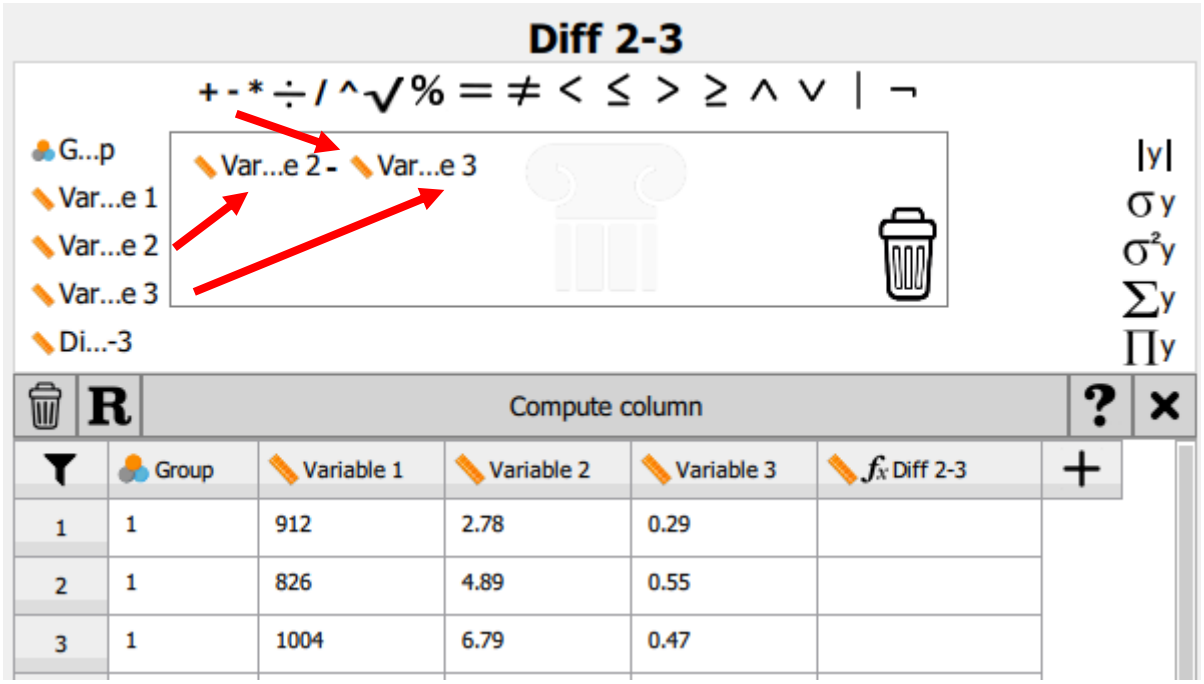


Una vez nombrada la nueva variable y elegidas las demás opciones, clique «Create».

Si se elige la opción manual en lugar del código R, se abrirán todas las opciones integradas para crear y transformar. A pesar de no ser muy intuitivo, se puede navegar por las opciones que hay a mano izquierda y a mano derecha para encontrar más variables y otros operadores, respectivamente.

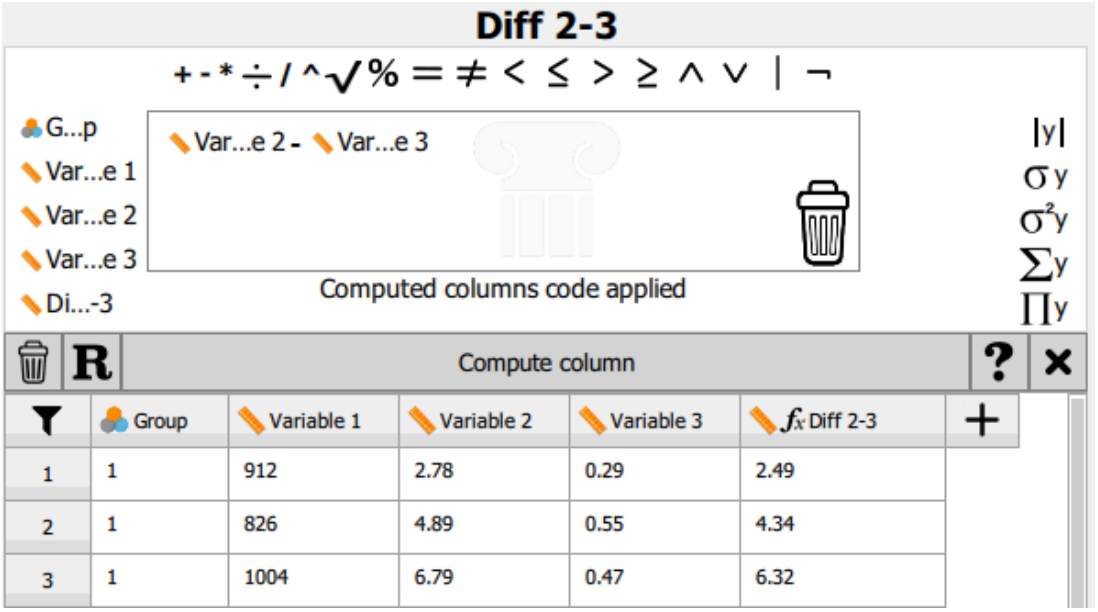


Por ejemplo, queremos crear una columna de datos que muestre la diferencia entre la variable 2 y la variable 3. Una vez introducido el nombre de la columna en el cuadro de diálogo «Create computed column», este aparecerá en la ventana de la hoja de cálculo. Ahora será necesario definir las operaciones matemáticas. En este caso, arrastre la variable 2 hasta la caja de ecuaciones, haga lo mismo con el signo “menos” y finalmente arrastre la variable 3.



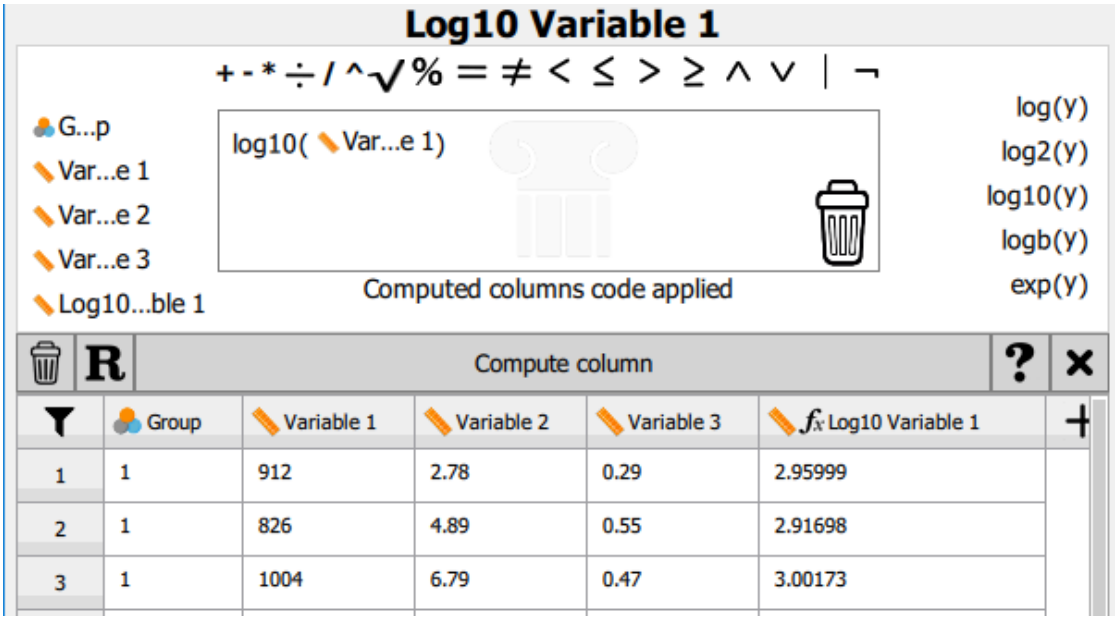
Si ha cometido algún error, por ejemplo, si ha usado una variable o un operador erróneos, elimínelo arrastrando el ítem a la papelera que se encuentra en la esquina inferior derecha.

Cuando esté conforme con la ecuación / operación, clique en «Compute column» y el dato quedará incorporado.

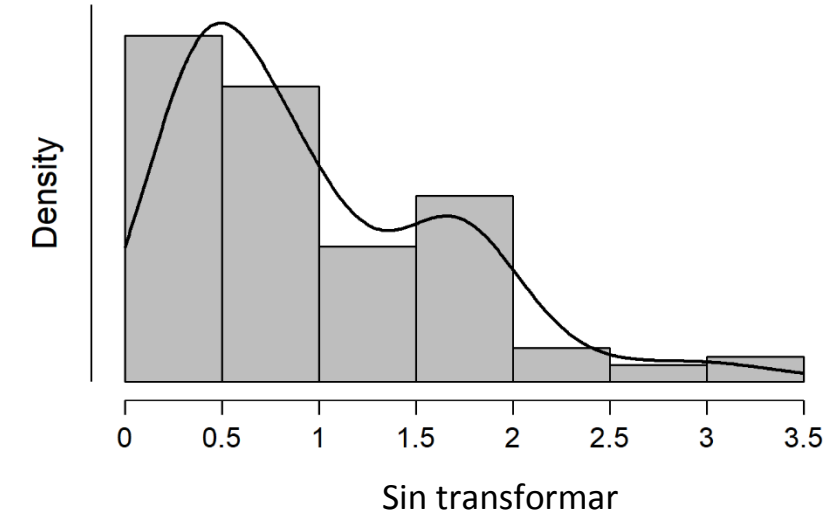


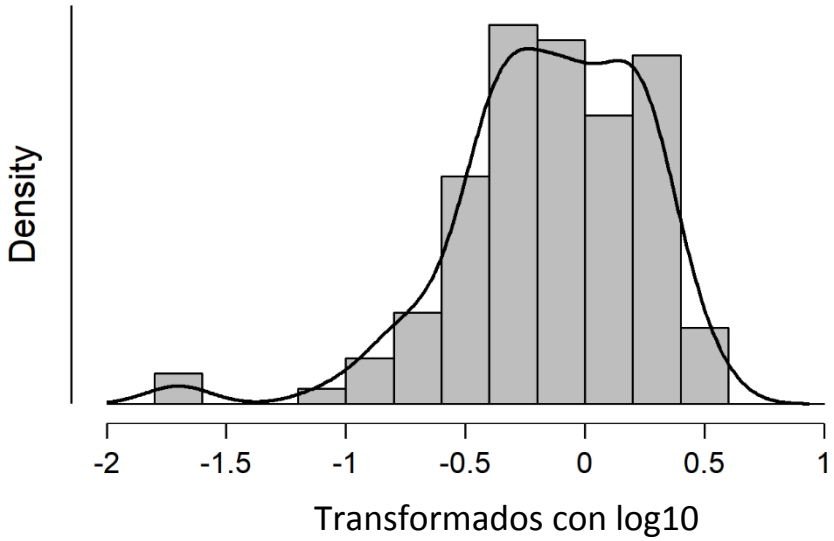
Si se decide no conservar los datos derivados, se puede eliminar la columna clicando en el otro icono de la papelera situado al lado de **R**.

Otro ejemplo sería realizar una transformación logarítmica de los datos. En el caso siguiente, la variable 1 ha sido transformada desplazándose por los operadores de la izquierda y seleccionando la opción «log10(y)». Reemplace la *y* con la variable que desea transformar y luego clique en «Compute column». Al terminar, haga clic en **X** para cerrar el diálogo.



Los dos gráficos siguientes muestran los datos sin transformar y los transformados con log10. Los datos claramente asimétricos han sido transformados en un perfil con una distribución más normal.





La función «Export» también exportará todas las nuevas variables que hayan sido creadas.

PRUEBA T PARA UNA MUESTRA ÚNICA

La investigación se lleva a cabo, normalmente, con muestras obtenidas de una población, pero ¿cuán cerca está la muestra de reflejar el conjunto de la población? La prueba t paramétrica para una muestra única determina si la media de la muestra es estadísticamente diferente de la media conocida o hipotética de la población.

La hipótesis nula (H_o) que se pone a prueba es que la media de la muestra es igual a la media de la población.

SUPUESTOS

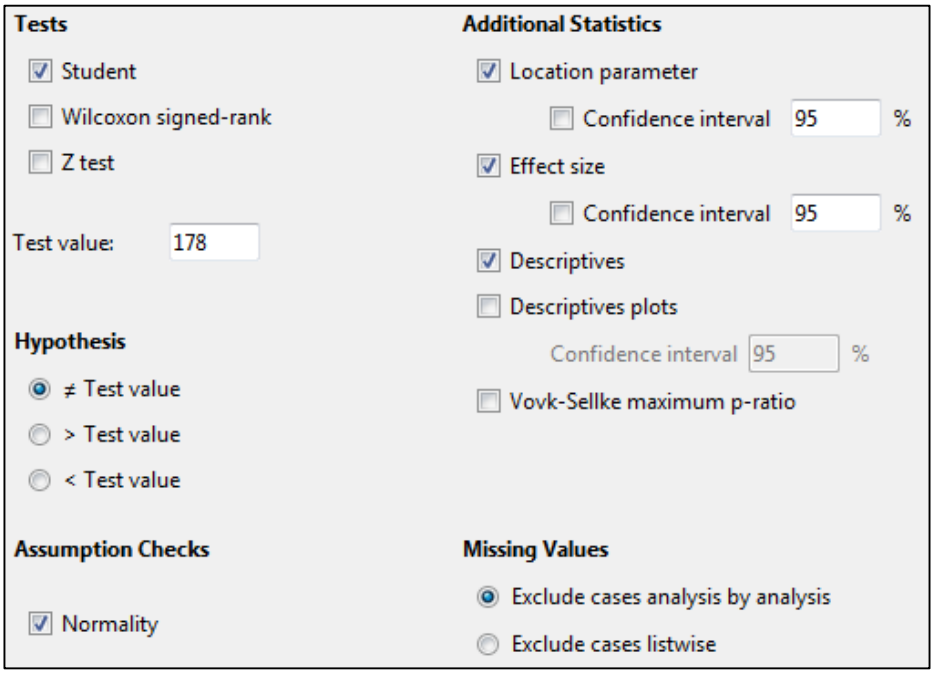
Se requieren tres supuestos para obtener un resultado válido en la prueba t para una muestra única:

- La variable de la prueba debe medirse en una escala **continua**.
- Los datos de la variable de la prueba deben ser **independientes**, es decir, sin relación entre ninguno de los puntos de datos.
- Los datos deben seguir una **distribución aproximadamente normal**.
- No debe haber **valores atípicos** significativos.

EJECUTANDO LA PRUEBA T PARA UNA MUESTRA ÚNICA

Abra **one sample t-test.csv**. Este archivo contiene dos columnas de datos que representan la altura (cm) y las masas corporales (kg) de una muestra de hombres usada en un estudio. En 2017, las medias de la población adulta masculina en el Reino Unido eran **178 cm** de altura y **83,6 kg** de masa corporal.

Vaya a «T-Tests» → «One sample t-test» y añada, en primera instancia, la altura a la caja de análisis de la derecha. Tras ello, seleccione las opciones siguientes y añada **178** como valor de prueba:



ENTENDIENDO EL RESULTADO

El resultado debe contener tres tablas.

Test of Normality (Shapiro-Wilk)

	W	p
height	0.969	0.507

Note. Significant results suggest a deviation from normality.

La comprobación del supuesto de normalidad (Shapiro-Wilk) no es significativa, lo que sugiere que las alturas están distribuidas normalmente; por lo tanto, este supuesto no es violado. Si el análisis mostrase una diferencia significativa, debería repetirse usando el equivalente no paramétrico, **la prueba de los rangos con signo de Wilcoxon** (*Wilcoxon's signed rank test*), probada sobre la mediana de altura de la población.

One Sample T-Test ▼

	t	df	p	Mean Difference	Cohen's d
height	−0.382	22	0.706	−0.391	−0.080

Note. Student's t-test.
Note. For the Student t-test, location parameter is given by mean difference *d*.
Note. For the Student t-test, effect size is given by Cohen's *d*.
Note. For all tests, the alternative hypothesis specifies that the population mean is different from 178.

Esta tabla muestra que no existen diferencias significativas entre las medias: p = 0,706.

Descriptives

	N	Mean	SD	SE
height	23.000	177.609	4.915	1.025

Los datos descriptivos muestran que la altura media de la muestra era de 177,6 cm comparada con el promedio de 178 cm de los hombres británicos.

Repita el procedimiento reemplazando altura por masa y cambiando el valor de prueba a 83,6.

Test of Normality (Shapiro-Wilk)

	W	p
mass	0.941	0.185

Note. Significant results suggest a deviation from normality.

La comprobación del supuesto de normalidad (Shapiro-Wilk) no es significativa, lo que sugiere que las masas están distribuidas normalmente.

One Sample T-Test

	t	df	p	Mean Difference	Cohen's d
mass	−7.159	22	< .001	−10.487	−1.493

Note. Student's t-test.
Note. For the Student t-test, location parameter is given by mean difference *d*.
Note. For the Student t-test, effect size is given by Cohen's *d*.
Note. For all tests, the alternative hypothesis specifies that the population mean is different from 83.4.

Esta tabla muestra una diferencia significativa entre la media de la muestra (72,9 kg) y la masa corporal de la población (83,6 kg): $p < 0,001$.

Descriptives

	N	Mean	SD	SE
mass	23.000	72.913	7.025	1.465

REPORTANDO LOS RESULTADOS

Una prueba t para una muestra única no exhibió diferencias significativas en la altura en comparación con la media de la población: t (22) = -0,382, p = 0,706. No obstante, los participantes eran significativamente más delgados (menor masa corporal) que el promedio de la población masculina del Reino Unido: t (22) = -7,159, p < 0,001.

TEST BINOMIAL

El test binomial es una versión no paramétrica de la prueba t para una muestra única destinado a usarse con conjuntos de datos categóricos dicotómicos (es decir, sí / no). Esta prueba sirve para determinar si la frecuencia de la muestra es estadísticamente diferente de la frecuencia poblacional conocida o hipotética.

La hipótesis nula (H_o) que se pone a prueba es que la frecuencia de la muestra es igual a la frecuencia poblacional esperada.

SUPUESTOS

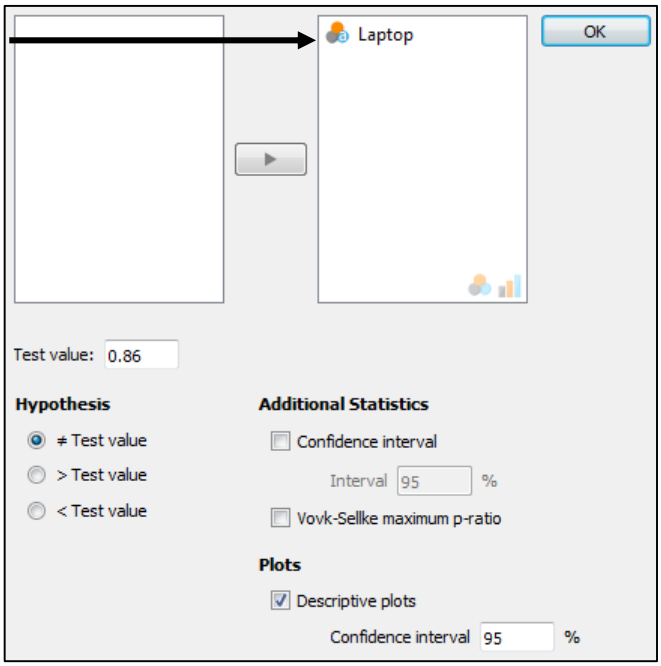
Se requieren tres supuestos para que un test binomial ofrezca un resultado válido:

- La variable del test debe tener una escala dicotómica (como sí/no, masculino/femenino, etc.).
- Las respuestas de la muestra deben ser independientes.
- El tamaño de la muestra es más pequeño, pero sigue siendo representativa de la población.

EJECUTANDO EL TEST BINOMIAL

Abra **binomial.csv**. Este archivo contiene una columna de datos que muestra el número de estudiantes que usan o bien un portátil Windows, o bien un MacBook en la universidad. En enero de 2018, comparando estos dos sistemas operativos, la cuota de mercado de Windows en el Reino Unido era del 86%, y la de Mac IOS del 14%.[3]

Vaya a «Frequencies» ⟶ «Binomial test». Traslade la variable Laptop a la ventana de datos e indique el valor de prueba en 0,86 (86%). Seleccione, también, «Descriptive plots».



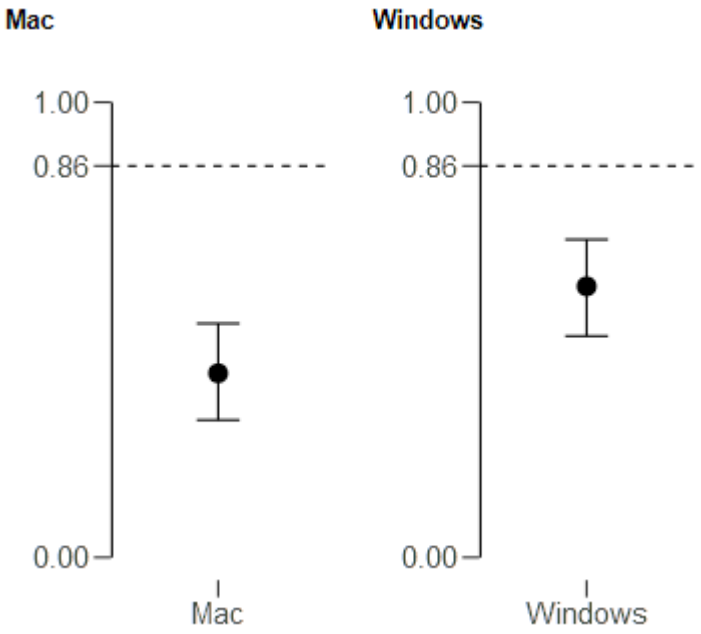
[3] https://www.statista.com/statistics/268237/global-market-share-held-by-operating-systems-since-2009/

La tabla y el gráfico siguientes muestran que las frecuencias de ambos portátiles son significativamente inferiores al 86%. En particular, estos estudiantes están usando portátiles Windows de un modo significativamente inferior a lo esperado, comparado con la cuota de mercado en el Reino Unido.

Binomial Test

	Level	Counts	Total	Proportion	p
Laptop	Mac	36	89	0.404	< .001
	Windows	53	89	0.596	< .001

Note. Proportions tested against value: 0.86.

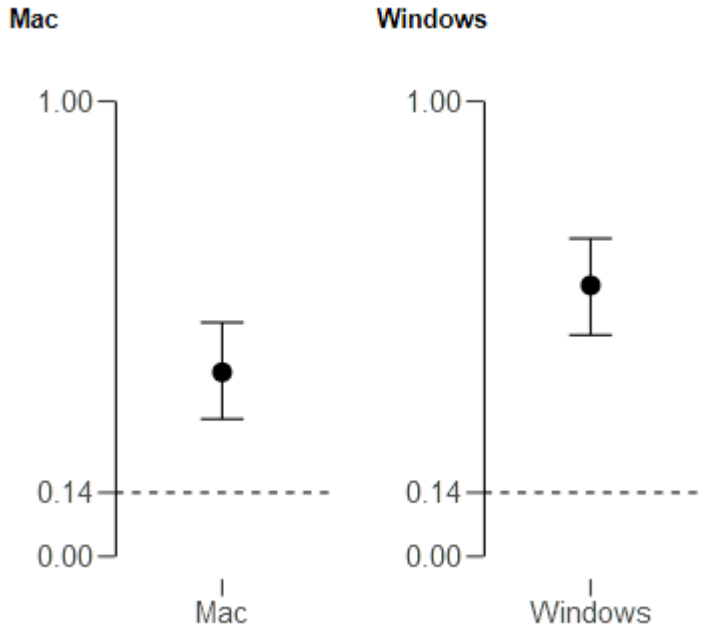


¿Sucede lo mismo con los usuarios de MacBook? Vuelva a la ventana de opciones y cambie el valor de prueba por 0,14 (14%). Esta vez, la frecuencia es significativamente superior al 14%. Esto muestra que los estudiantes usan MacBooks de un modo significativamente superior a lo esperado, comparado con la cuota de mercado en el Reino Unido.

Binomial Test

	Level	Counts	Total	Proportion	p
Laptop	Mac	36	89	0.404	< .001
	Windows	53	89	0.596	< .001

Note. Proportions tested against value: 0.14.



REPORTANDO LOS RESULTADOS

La proporción reportada de usuarios británicos de Windows y MacBook fue, respectivamente, del 86% y del 14%. En una cohorte de estudiantes universitarios (N = 90), un test binomial reveló que la proporción de estudiantes usuarios de portátiles Windows era significativamente inferior (59,6%, $p < 0,001$) y los que utilizaban MacBooks lo hacían de forma significativamente superior (40,4%, $p < 0,001$) a lo esperado.

TEST MULTINOMIAL

El test multinomial es una versión extendida del test binomial, destinado a usarse con conjuntos de datos categóricos que contengan tres o más factores. Esta prueba sirve para determinar si la frecuencia de la muestra es o no es estadísticamente diferente de una frecuencia poblacional hipotética (test multinomial) o conocida (test de "bondad de ajuste" chi cuadrado).

La hipótesis nula (H_o) que se pone a prueba es que la frecuencia de la muestra es igual a la frecuencia poblacional esperada.

SUPUESTOS

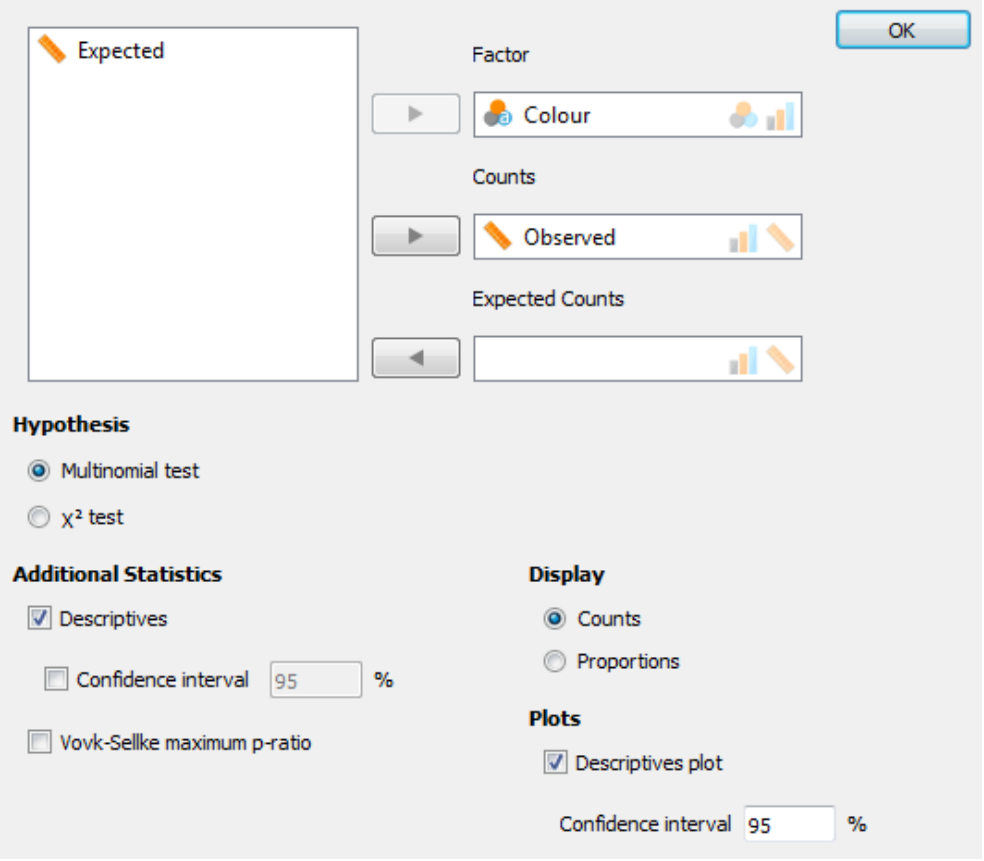
Se requieren tres supuestos para que un test multinomial proporcione un resultado válido:

- La variable del test debe tener una escala categórica con 3 o más factores.
- Las respuestas de la muestra deben ser independientes.
- El tamaño de la muestra es más pequeño, pero sigue siendo representativa de la población.

EJECUTANDO EL TEST MULTINOMIAL

Abra **multinomial.csv**. Este archivo contiene tres columnas de datos que muestran el número de M&M de diferentes colores repartidos en cinco bolsas. Sin ningún conocimiento previo, se podría suponer que los M&M de diferentes colores se distribuyen por igual.

Vaya a «Frequencies» ⟶ «Multinomial test». Traslade el color del M&M a «Factor» y el número observado de M&M a «Counts». Seleccione «Descriptives» y «Descriptives plot».



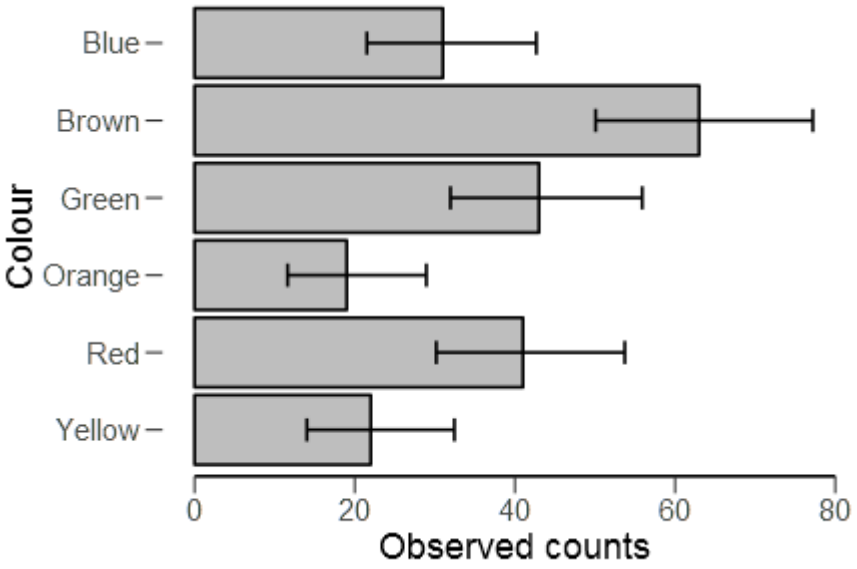
Como se puede ver en la tabla de descriptivas, el test asume una misma expectativa para las proporciones de M&M de colores (36 de cada color). Los resultados del test multinomial muestran que la distribución observada es significativamente diferente ($p < 0,001$) a una distribución equitativa.

Multinomial Test

	χ^2	df	p
Multinomial	35.932	5	< .001

Descriptives

Colour	Observed	Expected: Multinomial
Blue	31	36
Brown	63	36
Green	43	36
Orange	19	36
Red	41	36
Yellow	22	36



TEST DE "BONDAD DE AJUSTE" CHI CUADRADO

Sin embargo, investigaciones adicionales muestran que los fabricantes producen M&M de colores en diferentes proporciones:

Color	Azul	Marrón	Verde	Naranja	Rojo	Amarillo
Proporción	24	13	16	20	13	14

Ahora, estos valores pueden ser usados como recuentos estimados, por tanto, mueva la variable Expected a la caja «Expected Counts». Esto ejecuta automáticamente el test de "bondad de ajuste" χ^2 dejando en gris las opciones de hipótesis.

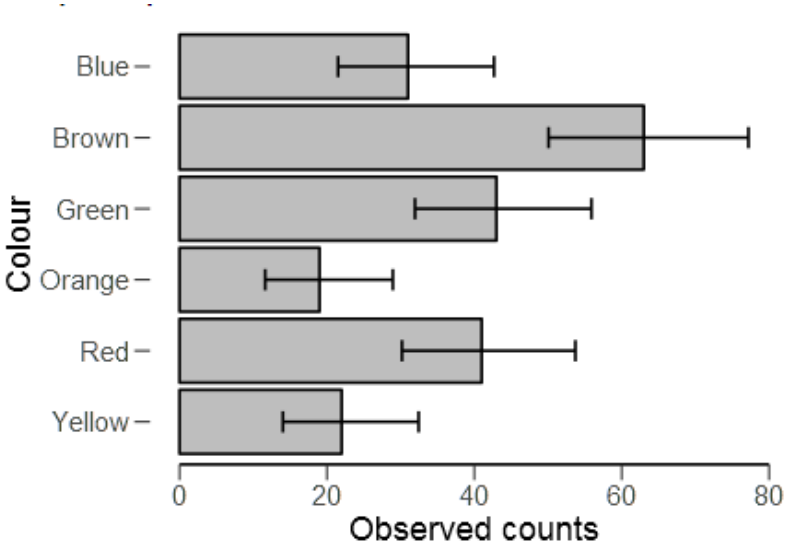
Como puede verse en la tabla de descriptivas, JASP ha calculado el número esperado de M&M de cada color en base a la ratio de producción reportada por los fabricantes. Los resultados del test muestran que las proporciones observadas para los M&M de distintos colores son significativamente diferentes (χ^2 = 74,5, p < 0,001) de las proporciones declaradas por el fabricante.

Multinomial Test

	χ^2	df	p
Expected	74.535	5	< .001

Descriptives

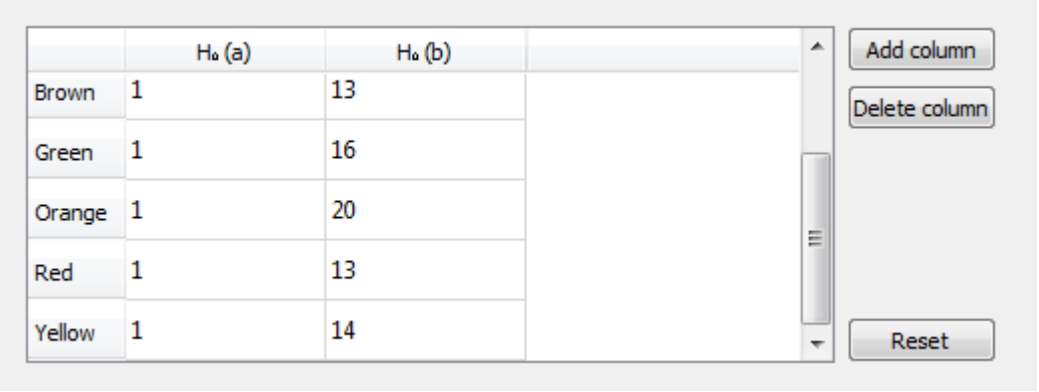
Colour	Observed	Expected: Expected
Blue	31	52
Brown	63	28
Green	43	35
Orange	19	43
Red	41	28
Yellow	22	30



TEST MULTINOMIAL Y DE "BONDAD DE AJUSTE" X^2

JASP también proporciona otra opción mediante la cual ambas pruebas se pueden ejecutar al mismo tiempo. Regrese a la ventana de opciones y agregue la variable Colour a la caja «Factor» y Observed a la caja «Counts»; elimine Expected de la caja «Expected Counts» si la variable aún se encuentra ahí. En «Hypothesis», marque el test χ^2. Esto abrirá una pequeña ventana de hoja de cálculo que mostrará el color y H_o (a) con un 1 en cada celda. Esto implica que las proporciones de cada color son las mismas (test multinomial).

En esta ventana, añada otra columna que se etiquetará automáticamente como H_o (b). Ahora se pueden introducir las proporciones estimadas para cada color.



Ahora, una vez ejecutado el análisis, se muestran los resultados de las pruebas para las dos hipótesis. H_o (a) comprueba la hipótesis nula de que las proporciones de cada color están distribuidas por igual, mientras que H_o (b) comprueba la hipótesis nula de que las proporciones son las mismas que las esperadas. Como se puede observar, ambas hipótesis son rechazadas. En concreto, la evidencia indica que los colores de los M&M no coinciden con las proporciones publicadas por los fabricantes.

Multinomial Test

	χ^2	df	p
Hₒ (a)	35.932	5	< .001
Hₒ (b)	74.535	5	< .001

Descriptives

		Expected	
Colour	Observed	Hₒ (a)	Hₒ (b)
Blue	31	36	52
Brown	63	36	28
Green	43	36	35
Orange	19	36	43
Red	41	36	28
Yellow	22	36	30

COMPARACIÓN DE DOS GRUPOS INDEPENDIENTES

PRUEBA T PARA DOS MUESTRAS INDEPENDIENTES

La prueba t paramétrica para dos muestras independientes, también conocida como prueba t de Student (*Student's t-test*), se usa para determinar si existe diferencia estadística entre las medias de dos grupos independientes. La prueba requiere una variable dependiente continua (p. ej., masa corporal) y una variable independiente que contenga dos grupos (p. ej., hombres y mujeres).

Con esta prueba se obtiene una puntuación t (*t-score*) que es el cociente de las diferencias entre los dos grupos y las diferencias dentro de los dos grupos:

$$t = \frac{\textbf{media grupo 1} - \textbf{media grupo 2}}{\textbf{error estándar de las medias}}$$

$$t = \frac{(X_1 - X_2)}{\sqrt{\frac{(S_1)^2}{n_1} + \frac{(S_2)^2}{n_2}}}$$

X = media

S = desviación estándar

n = número de puntos de datos

Una puntuación t alta indica que existe una gran diferencia entre los grupos. Cuanto más baja sea la puntuación t, mayor será la similitud entre los grupos. Una puntuación t de 5 indica que los grupos son cinco veces más diferentes entre ellos de lo que lo son dentro de cada uno de ellos.

La hipótesis nula (H_o) que se pone a prueba es que las medias poblacionales de los dos grupos no relacionados son iguales.

SUPUESTOS DE LA PRUEBA T PARAMÉTRICA PARA DOS MUESTRAS INDEPENDIENTES

Independencia del grupo:

Ambos grupos deben ser independientes entre sí. Cada participante solo proporcionará un punto de datos para un solo grupo. Por ejemplo, el participante 1 solo puede estar en un grupo, masculino o femenino, pero no en ambos. Las medidas repetidas se evalúan con la **prueba t para dos muestras apareadas** (*paired t-test*).

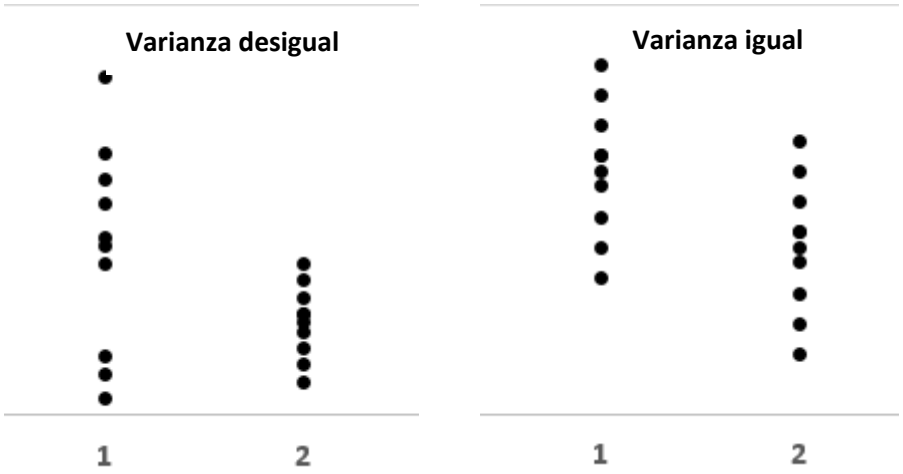
Normalidad de la variable dependiente:

La variable dependiente también debe medirse en una escala continua y debe tener una distribución aproximadamente normal, sin valores atípicos significativos. Esto se puede comprobar mediante el test Shapiro-Wilk. La prueba t es bastante robusta, por lo que pueden aceptarse pequeñas desviaciones de la normalidad. Sin embargo, esto no es así en el caso de grupos con tamaños muy diferentes. Como regla general, la ratio entre los tamaños de grupo debe ser < 1,5 (p. ej., grupo A = 12 participantes y grupo B = > 8 participantes).

Si la normalidad ha sido violada, puede intentar transformar los datos (p. ej., transformaciones logarítmicas o raíz cuadrada) o, si los tamaños de grupo son muy diferentes, usar el test **U de Mann-Whitney**, el equivalente no paramétrico que no requiere el supuesto de normalidad (ver más adelante).

Homogeneidad de la varianza:

Las varianzas de la variable dependiente deben ser iguales en cada grupo. Esto se puede comprobar con el test de igualdad de varianzas de Levene.

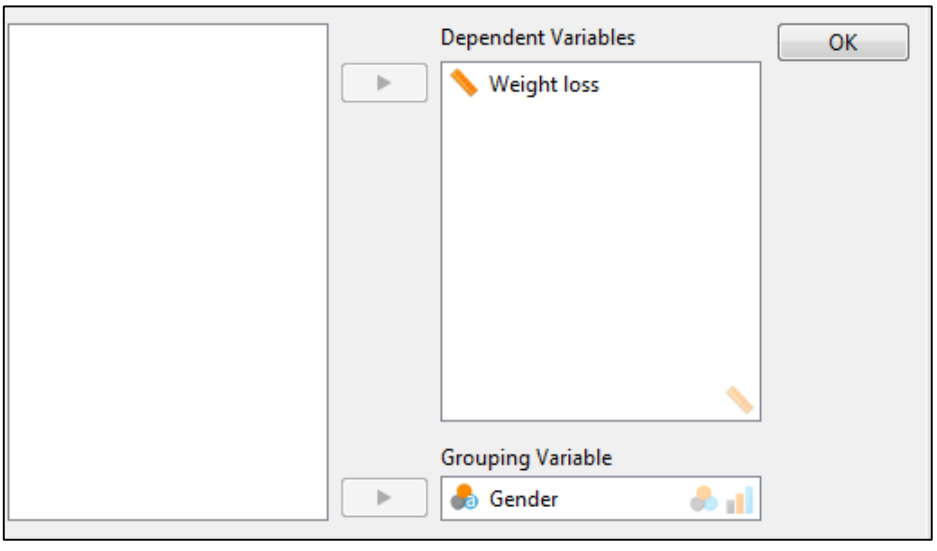


Si el test de Levene es estadísticamente significativo, indicando que las varianzas de los grupos son desiguales, se puede corregir esta violación usando una prueba t ajustada según el método de **Welch**.

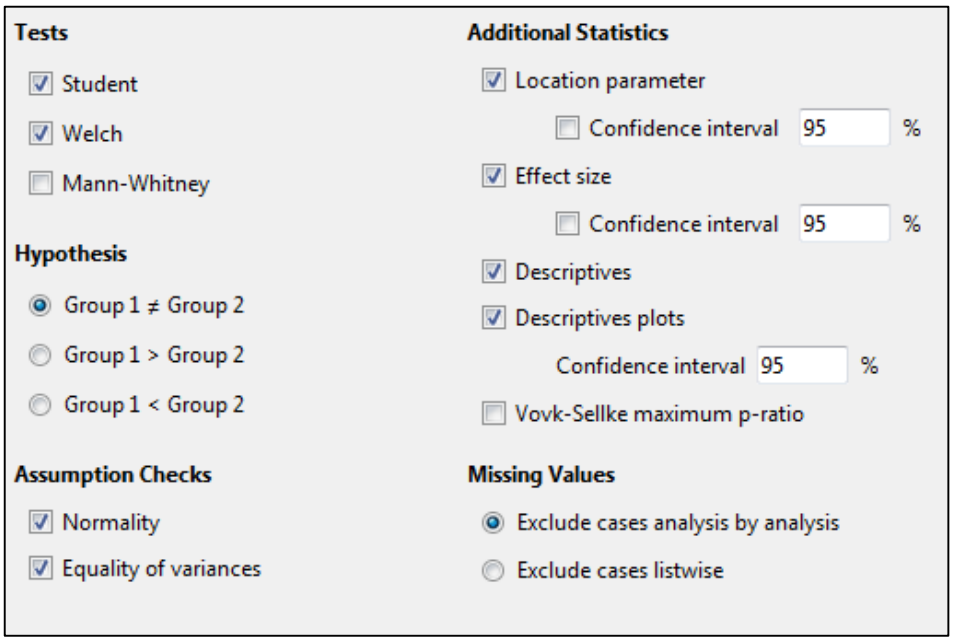
EJECUTANDO LA PRUEBA T PARA DOS MUESTRAS INDEPENDIENTES

Abra **Independent t-test.csv**. Este archivo contiene la pérdida de peso con una dieta autocontrolada de 10 semanas entre hombres y mujeres. Es una buena práctica comprobar la distribución y los gráficos de caja en «Descriptives», para verificar visualmente la distribución y los valores atípicos.

Vaya a «T-Tests» → «Independent samples t-test», e introduzca la pérdida de peso en la caja «Dependent Variables» y el género (variable independiente) en la caja «Grouping Variable».



En la ventana de análisis, seleccione las opciones siguientes:



ENTENDIENDO LOS RESULTADOS

El resultado debe contener cuatro tablas y un gráfico. En primer lugar, hace falta comprobar que no se violan los supuestos paramétricos requeridos.

Test of Normality (Shapiro-Wilk)

		W	p
Weight loss	Females	0.968	0.282
	Males	0.971	0.310

Note. Significant results suggest a deviation from normality.

El test Shapiro-Wilk muestra que ambos grupos tienen datos distribuidos normalmente, por lo que no se viola el supuesto de normalidad. Si uno o ambos fuesen significativos, habría que considerar el uso del test equivalente no paramétrico de **Mann-Whitney**.

Test of Equality of Variances (Levene's)

	F	df	p
Weight loss	2.278	1	0.135

La prueba de Levene muestra que no hay diferencia en la varianza, por lo tanto, no se viola el supuesto de homogeneidad de la varianza. Si la prueba de Levene fuese significativa, se debería reportar la prueba t con la corrección de Welch, los grados de libertad y los valores de p.

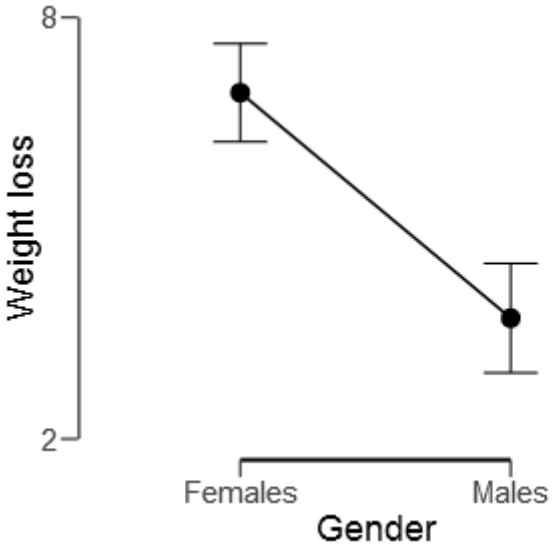
Independent Samples T-Test

	Test	Statistic	df	p	Mean Difference	SE Difference	Cohen's d
Weight loss	Student	6.160	85.000	< .001	3.209	0.521	1.322
	Welch	6.191	84.544	< .001	3.209	0.518	1.325

Esta tabla muestra el cálculo de las dos pruebas t (**Student y Welch**). Debemos recordar que el estadístico t se obtiene dividiendo la diferencia de medias por el error estándar de la diferencia. Ambos muestran que hay una diferencia estadística significativa entre los dos grupos ($p < 0,001$), y la d de Cohen sugiere que se trata de un efecto importante.

Group Descriptives

	Group	N	Mean	SD	SE
Weight loss	Females	42	6.929	2.242	0.346
	Males	45	3.720	2.588	0.386



A partir de los datos descriptivos, se puede ver que las mujeres tuvieron una pérdida de peso mayor que los hombres.

REPORTANDO LOS RESULTADOS

Una prueba t para dos muestras independientes mostró que las mujeres han perdido significativamente más peso tras 10 semanas de dieta que los hombres: **t(85) = 6,16, p < 0,001**. La d de Cohen (1,322) sugiere que se trata de un efecto importante.

PRUEBA U DE MANN-WITNEY

Si se da el caso de que los datos no están normalmente distribuidos (resultado significativo del test de Shapiro-Wilk) o si la distribución es ordinal, la prueba no paramétrica para dos muestras independientes equivalente es la **prueba U de Mann-Whitney**.

Abra **Mann-Whitney pain.csv**. Este archivo contiene puntuaciones de dolor subjetivo (0-10) con y sin tratamiento con hielo. **Nota**: compruebe que el tratamiento sea categórico y que la puntuación del dolor sea ordinal. Vaya a «T-test» ⟶ «Independent t-test» y añada la puntuación del dolor en la caja «Dependent Variables», usando el tratamiento como variable de agrupación.

En las opciones de análisis, seleccione solo:

- ✓ Mann-Whitney.
- ✓ Parámetro de localización (*Location parameter*).
- ✓ Tamaño del efecto (*Effect size*).

No hay ninguna razón para comprobar los supuestos, ya que Mann-Whitney no asume el supuesto de normalidad ni el de homogeneidad de la varianza requeridos por las pruebas paramétricas.

ENTENDIENDO EL RESULTADO

Esta vez solo se obtiene una tabla:

Independent Samples T-Test

	W	p	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Pain score	207.000	< .001	3.000	0.840

Note. Mann-Whitney U test.

El test estadístico U de Mann-Whitney (JASP la reporta como *W*, ya que se trata de una adaptación del test de los rangos con signo de Wilcoxon) es altamente significativo: **U = 207, p < 0,001**.

El parámetro de localización, la estimación Hodges-Lehmann, es la diferencia **mediana** entre los dos grupos. La correlación de rango biserial (*Rank-Biserial Correlation*, r_B) puede ser considerada como tamaño del efecto e interpretada del mismo modo que la r de Pearson, por lo que 0,84 es un tamaño del efecto importante.

Para datos no paramétricos, se deben reportar valores **medianos** como estadística descriptiva y usar gráficos de caja en lugar de gráficos de líneas e intervalos de confianza, barras SD / SE. Vaya a «Descriptive statistics», introduzca la puntuación de dolor en la caja «Variables» y el tratamiento a la caja «Split».

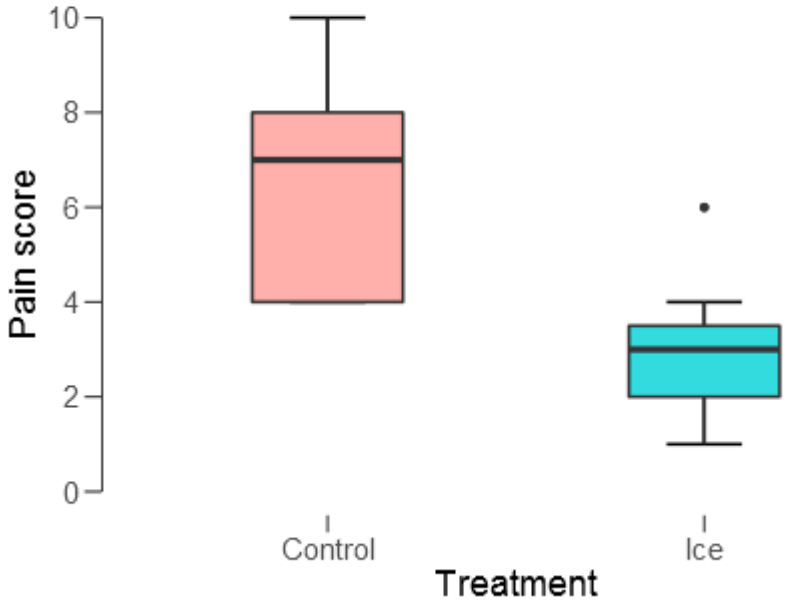
Descriptive Statistics

	Pain score	
	Control	Ice
Valid	15	15
Missing	0	0
Median	7.000	3.000
Minimum	4.000	1.000
Maximum	10.000	6.000

Plots

Boxplots

Pain score



REPORTANDO LOS RESULTADOS

El test de Mann-Whitney mostró que el tratamiento con hielo reduce significativamente las puntuaciones de dolor (Mdn = 3), en comparación con el grupo de control (Mdn = 7), U = 207, $p < 0,001$.

COMPARACIÓN DE DOS GRUPOS RELACIONADOS

PRUEBA T PARA DOS MUESTRAS APAREADAS

Como sucede con la prueba t para dos muestras independientes, JASP ofrece ambas opciones: la paramétrica y la no paramétrica. La prueba t paramétrica para dos muestras apareadas (también conocida como prueba t para muestras dependientes o prueba t para medidas repetidas) compara las medias entre dos grupos relacionados en la misma variable continua dependiente. Por ejemplo, observando la pérdida de peso antes y después de las 10 semanas de dieta.

$$\text{Estadístico } t \text{ apareado} = \frac{\text{media de las diferencias entre las parejas de los grupos}}{\text{error estándar de las diferencias de las medias}}$$

Con la prueba t para dos muestras apareadas, la hipótesis nula (H_o) que se pone a prueba es que la diferencia entre las parejas de los dos grupos es cero.

SUPUESTOS DE LA PRUEBA T PARAMÉTRICA PARA DOS MUESTRAS APAREADAS

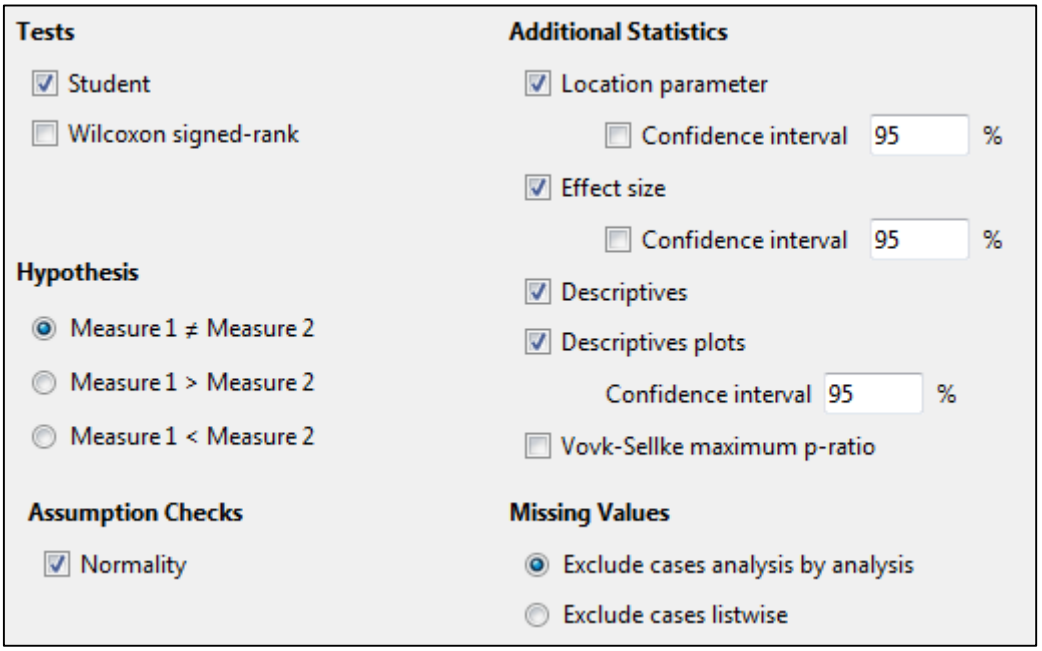
Para que la prueba t paramétrica proporcione un resultado válido, se requieren cuatro supuestos:

- La **variable dependiente** debe ser medida en una escala continua.
- La **variable independiente** debe contar con 2 grupos categóricos relacionados / emparejados, es decir, que cada participante aparece en ambos grupos.
- Las diferencias entre las parejas deben estar aproximadamente **distribuidas normalmente**.
- No debe haber **valores atípicos** significativos en las diferencias entre los 2 grupos.

EJECUTANDO LA PRUEBA T PARA MUESTRAS APAREADAS

Abra **Paired t-test.csv** en JASP. Este archivo contiene dos columnas de datos apareados: masa corporal anterior a la dieta y tras 4 semanas haciendo dieta. Vaya a «T-test» ⟶ «Paired samples t-test». Haga clic sobre ambas variables manteniendo la tecla Ctrl presionada y añádalas a la caja de análisis de la derecha.

En las opciones de análisis, marque lo siguiente:



ENTENDIENDO EL RESULTADO

El resultado debe incluir tres tablas y un gráfico.

Test of Normality (Shapiro-Wilk)

			W	p
Pre diet body mass	-	Post 4 weeks diet	0.975	0.124

Note. Significant results suggest a deviation from normality.

La comprobación del supuesto de normalidad (Shapiro-Wilk) no es significativa, sugiriendo que las diferencias apareadas están distribuidas normalmente, de forma que se cumple el supuesto. Si mostrase una diferencia significativa, el análisis debería repetirse usando el equivalente no paramétrico, la **prueba de rangos con signo de Wilcoxon.**

Paired Samples T-Test

			t	df	p	Mean Difference	SE Difference	Cohen's d
Pre diet body mass	-	Post 4 weeks diet	13.039	77	< .001	3.782	0.290	1.476

Note. Student's t-test.

Esto muestra que hay una diferencia significativa de masa corporal entre las condiciones previas y las posteriores a la dieta, con una diferencia media (parámetro de localización) de 3,783 kg. La d de Cohen establece que se trata de un efecto importante.

El gráfico y la estadística descriptiva muestran que hubo una reducción de masa corporal tras seguir la dieta durante 4 semanas.

Descriptives

	N	Mean	SD	SE
Pre diet body mass	78	72.526	8.723	0.988
Post 4 weeks diet	78	68.744	9.009	1.020

REPORTANDO LOS RESULTADOS

Los participantes perdieron, de promedio, 3,78 kg (SE: 0,29 kg) de masa corporal siguiendo un programa de dieta de 4 semanas. La prueba t para muestras apareadas mostró que esta disminución es significativa (**t(77) = 13,039, p < 0,001**). La d de Cohen sugiere que se trata de un efecto importante.

EJECUTANDO LA PRUEBA NO PARAMÉTRICA PARA MUESTRAS APAREADAS

PRUEBA DE RANGOS CON SIGNO DE WILCOXON

Si se observa que los datos no están normalmente distribuidos (resultado significativo del test Shapiro-Wilk) o si la distribución es ordinal, la prueba no paramétrica equivalente es la prueba de rangos con signo de Wilcoxon. Abra **Wilcoxon's rank.csv**. Este archivo contiene dos columnas: una con las puntuaciones de ansiedad antes del tratamiento y otra con las puntuaciones después de un tratamiento con hipnoterapia (de 0 a 50). Al mostrarse el conjunto de datos, asegurarse de que ambas variables están asignadas como variables ordinales.

Vaya a «T-test» → «Paired samples t-test» y siga las instrucciones explicadas anteriormente, pero esta vez seleccione, únicamente, las opciones siguientes:

- ✓ Rango con signo de Wilcoxon (*Wilcoxon signed rank*).
- ✓ Parámetro de localización (*Location parameter*).
- ✓ Tamaño del efecto (*Effect size*).

El resultado se mostrará en una única tabla:

Paired Samples T-Test

			W	p	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Pre-anxiety	-	Post-anxiety	322.000	< .001	8.000	0.480

Note. Wilcoxon signed-rank test.

El estadístico W de Wilcoxon es altamente significativo, $p < 0{,}001$.

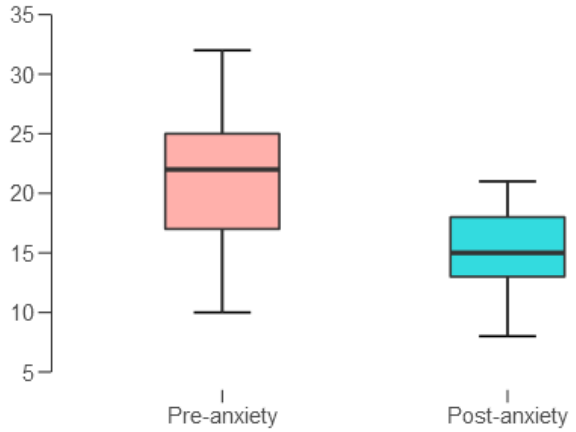
El parámetro de localización, la estimación Hodges-Lehmann, es la diferencia mediana entre los dos grupos. La correlación de rango biserial (*Rank-Biserial Correlation*, r_B) puede ser considerada como un tamaño del efecto y se interpreta como la r de Pearson, por lo que 0,48 es un tamaño del efecto entre medio y grande.

Tamaño del efecto	Irrelevante	Pequeño	Medio	Grande
Rango biserial (r_B)	< 0,1	0,1	0,3	0,5

Para datos no paramétricos, se deben reportar los valores medianos como estadística descriptiva y usar gráficos de caja en lugar de gráficos de línea e intervalos de confianza, barras SD / SE.

Descriptive Statistics

	Pre-anxiety	Post-anxiety
Valid	29	29
Missing	0	0
Median	22.0	15.0
Minimum	10.0	8.0
Maximum	32.0	21.0



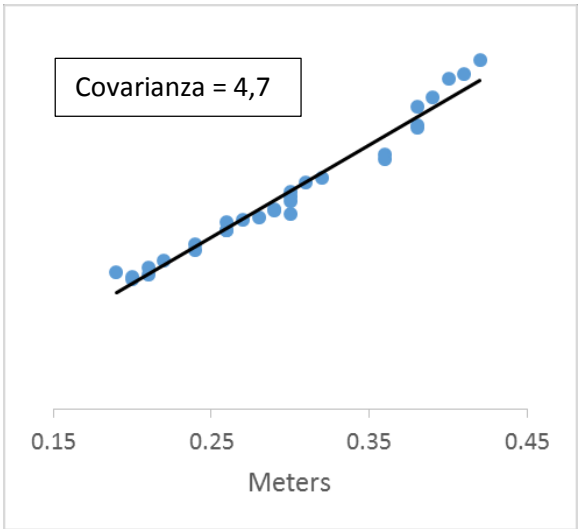
REPORTANDO LOS RESULTADOS

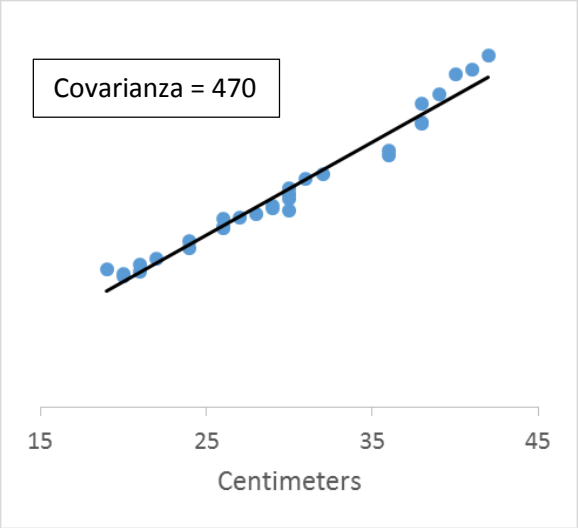
La prueba de rangos con signo de Wilcoxon mostró que la hipnoterapia reduce significativamente las puntuaciones de ansiedad (Mdn = 15), en comparación con las puntuaciones de ansiedad anteriores al tratamiento (Mdn = 22), W = 322, p < 0,001.

ANÁLISIS DE CORRELACIÓN

La correlación es una técnica estadística que se puede usar para determinar si hay pares de variables relacionados y con qué fuerza lo están. La correlación solo es apropiada para datos cuantificables que tengan significado, como datos continuos u ordinales. No puede usarse para datos puramente categóricos; para estos, lo indicado es el análisis de tabla de contingencia (ver Análisis chi cuadrado en JASP).

En esencia, ¿diferentes variables covarían? Es decir, ¿se dan cambios en una variable que tengan su reflejo en cambios similares en otra variable? Si una variable se desvía de su media, ¿la otra variable se desvía de su media en la misma dirección o en la opuesta? Esto se puede evaluar midiendo la covarianza, aunque no es un método estandarizado. Por ejemplo, se puede medir la covarianza de dos variables medidas en metros. Sin embargo, si transformamos los valores a centímetros, obtenemos la misma relación, aunque con un valor de la covarianza completamente distinto.





Para superar esta situación, se usa una covarianza estandarizada, conocida como el **coeficiente de correlación de Pearson** (*Pearson's correlation coefficient*, o r). Adopta un valor en el intervalo entre -1,0 y +1,0. Cuanto más cerca está r de +1 o -1, más estrechamente relacionadas entre sí están las dos variables. Si r es cercano a 0, no hay relación. Si r es (+), cuando los valores de una variable son más altos, los de la otra también lo son. Si r es negativo (-), cuando los valores de una variable son más altos, los de la otra son más bajos (llamada a veces correlación "inversa").

No se debe confundir el coeficiente de correlación (r) con R^2 –coeficiente de determinación (*coefficient of determination*)–, ni con R –coeficiente de correlación múltiple (*multiple correlation coefficient*), tal como se usa en la regresión–.

El supuesto principal en este análisis es que los datos tienen una distribución normal y son lineales. Este análisis no funcionará bien con relaciones curvilíneas.

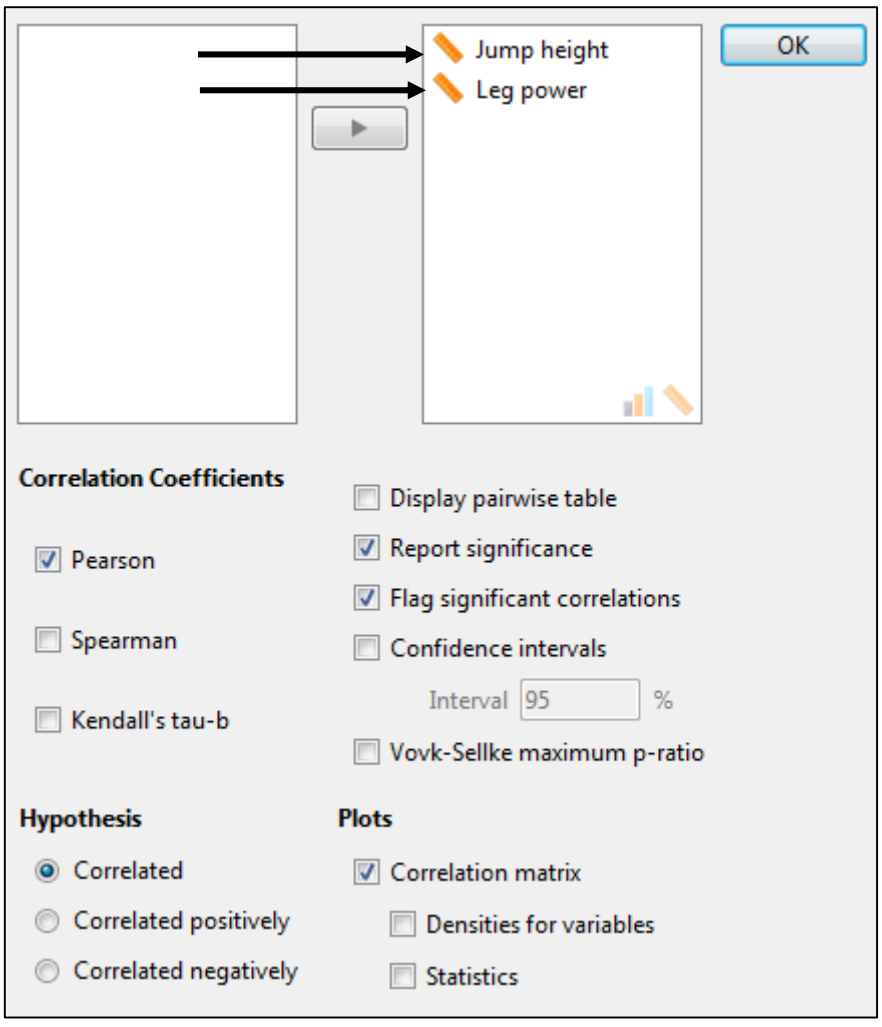
EJECUTANDO LA CORRELACIÓN

El análisis pone a prueba la hipótesis nula (H_0) de que no hay relación entre dos variables.

De los datos de ejemplo, abra **Jump height correlation.csv**. Este archivo contiene 2 columnas de datos, Jump height (m) y Leg power (W). En primer lugar, vaya a «Descriptive statistics» y compruebe los gráficos de caja por si hubiera valores atípicos.

Para ejecutar el análisis de correlación, vaya a «Regression» ⟶ «Correlation matrix». Traslade las 2 variables a la caja de análisis de la derecha. Marque:

- ✓ Pearson.
- ✓ Reportar significación («Report significance»).
- ✓ Marcar correlaciones significativas («Flag significant correlations»).
- ✓ Matriz de correlación («Correlation matrix») (en «Plots»).



ENTENDIENDO EL RESULTADO

La primera tabla muestra la matriz de correlación con los valores de la r de Pearson y sus p. Se observa una correlación altamente significativa (p < 0,001), con un valor de r cercano a 1 (r = 0,984), que nos permite rechazar la hipótesis nula.

Pearson Correlations

		Jump height	Leg power
Jump height	Pearson's r	—	
	p-value	—	
Leg power	Pearson's r	0.984***	—
	p-value	< .001	—

* p < .05, ** p < .01, *** p < .001

Para correlaciones simples como esta, resulta más sencillo observar la tabla de valores por parejas. Vuelva al análisis y seleccione la opción «Display pairwise table». Esto sustituye la matriz de correlación en los resultados y puede facilitar su lectura.

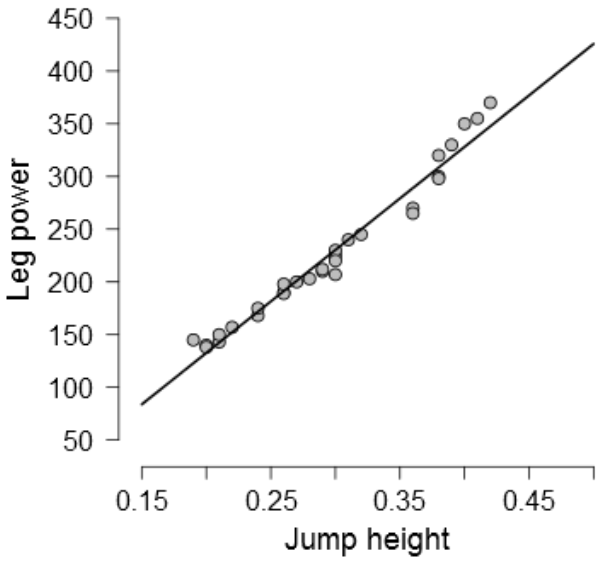
Pearson Correlations

			Pearson's r	p
Jump height	-	Leg power	0.984***	< .001

* p < .05, ** p < .01, *** p < .001

En realidad, el valor r de Pearson muestra un tamaño del efecto donde < 0,1 es irrelevante, de 0,1 a 0,3 es un efecto pequeño, de 0,3 a 0,5 es un efecto moderado y > 0,5 es un efecto grande.

El gráfico permite visualizar de una forma simple esta fuerte correlación positiva (r = 0,984, p < 0,001).



YENDO UN PASO MÁS ALLÁ

Si se toma el coeficiente de correlación r y se eleva al cuadrado, se obtiene el coeficiente de determinación (R^2). Es una medición estadística de la proporción de la varianza de una variable que se explica por la otra variable. O:

R^2 = Varianza explicada / Varianza total.

R^2 produce siempre un valor entre 0 y 100% en el que:

- un 0% indica que el modelo no explica nada sobre la variabilidad de los datos en torno a su media, y
- un 100% indica que el modelo explica toda la variabilidad de los datos en torno a su media.

En el ejemplo anterior, r = 0,984, por lo que R^2 = 0,968. Esto sugiere que la altura de salto representa un 96,8% de la varianza en la potencia de pierna.

REPORTANDO LOS RESULTADOS

La correlación de Pearson mostró una correlación significativa entre la altura de salto y la potencia de pierna (r = 0,984, p < 0,001), representando la altura de salto un 96,8% de la varianza en la potencia de pierna.

EJECUTANDO LA CORRELACIÓN NO PARAMÉTRICA: LA TAU DE KENDALL Y LA RHO DE SPEARMAN

Si los datos son ordinales o si son datos continuos que han violado los supuestos requeridos para el uso de la estadística paramétrica (normalidad y/o varianza), debería usar alternativas no paramétricas al coeficiente de correlación de Pearson.

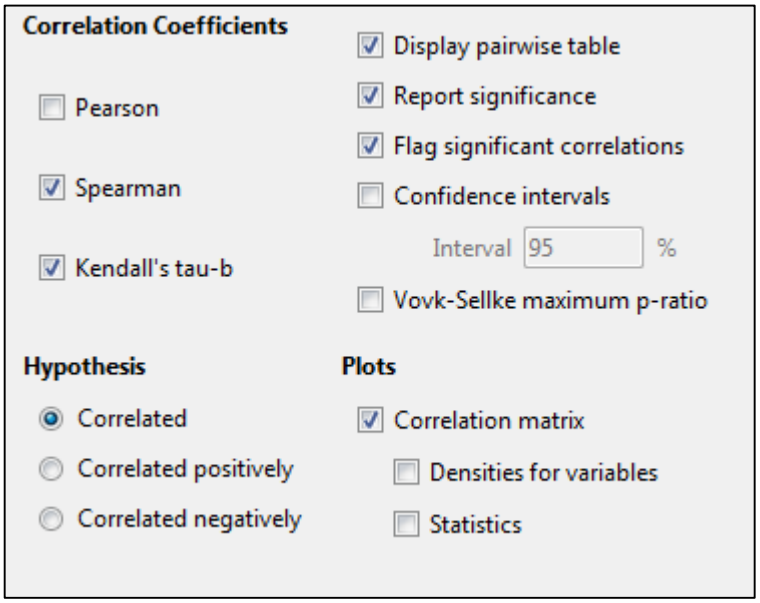
Las alternativas son los coeficientes de correlación de Spearman (rho) o Kendall (tau). Ambos están basados en datos de clasificación (ordenados de mayor a menor), y no están afectados por la presencia de valores atípicos o violaciones de la varianza / normalidad.

La rho de Spearman se usa habitualmente para datos de escala ordinal y la tau de Kendall se usa en muestras pequeñas o cuando hay muchos valores con la misma puntuación (empates). En la mayoría de los casos, la tau de Kendall y el coeficiente de correlación de Spearman son muy similares y, por lo tanto, conducen invariablemente a las mismas inferencias.

Los tamaños del efecto son los mismos que la r de Pearson. La principal diferencia es que se puede usar rho^2 como una aproximación no paramétrica al coeficiente de determinación, cosa que no sucede en el caso de la tau de Kendall.

De los datos de ejemplo, abra **Non-parametric correlation.csv**. Este archivo contiene 2 columnas de datos: una con puntuaciones de creatividad y otra con las posiciones en la competición de “El mayor mentiroso del mundo” (*World’s biggest liar*; gracias a Andy Field).

Ejecute el análisis como en el caso anterior, pero esta vez usando los coeficientes de Sperman y tau-b de Kendall en lugar del de Pearson.

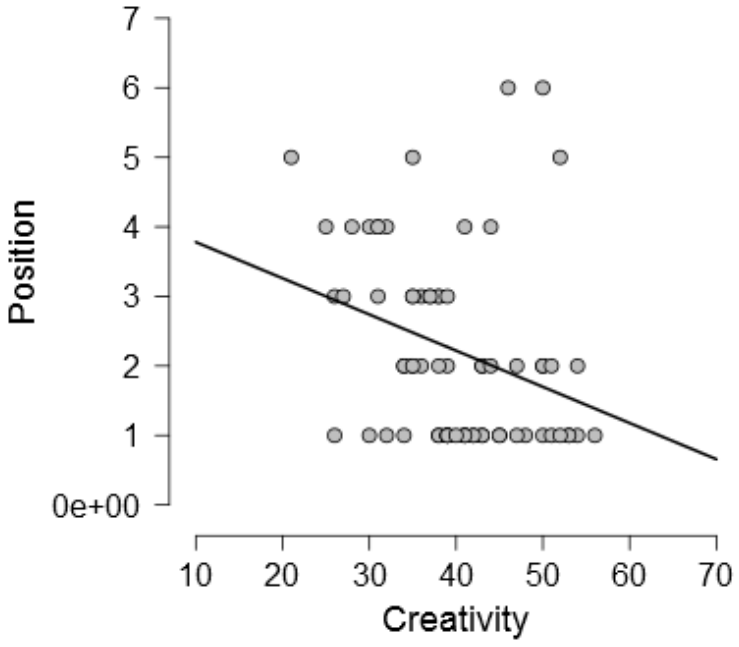


Correlation Table

			Spearman		Kendall	
			rho	p	tau B	p
Creativity	-	Position	−0.373**	0.002	−0.300**	0.001

* p < .05, ** p < .01, *** p < .001

Como puede verse, hay una correlación significativa entre las puntuaciones de creatividad y la posición final en la competición *World's biggest liar*: cuanto mayor es la puntuación, mejor es la posición final en la competición. Sin embargo, el tamaño del efecto es moderado.



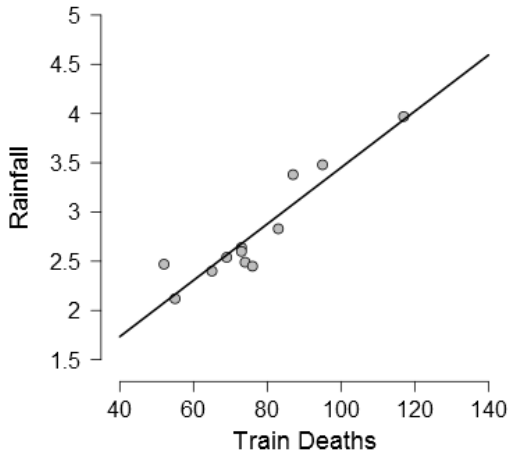
NOTA DE ADVERTENCIA

En realidad, la correlación solo ofrece información sobre la fortaleza de la asociación. No informa sobre la dirección, es decir, sobre qué variable hace que la otra cambie. Por ello, no puede ser usada para afirmar que una cosa es causa de otra. A menudo, una correlación significativa no quiere decir absolutamente nada y es puramente casual, en especial si se correlacionan miles de variables. Esto puede verse en correlaciones extrañas como las siguientes:

El número de peatones muertos en un atropello de tren correlaciona con la lluvia en Missouri

Pearson Correlations

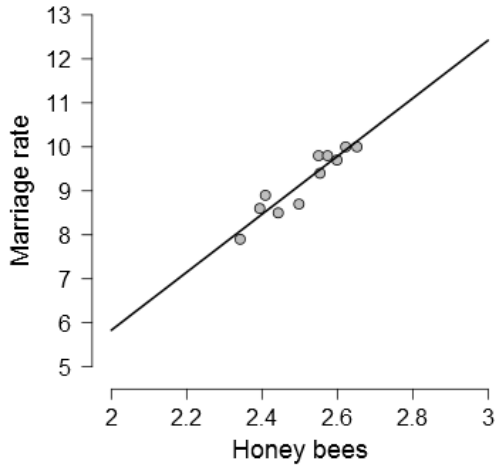
			Pearson's r	p
Train Deaths	-	Rainfall	0.928	< .001



El número de colonias de abejas productoras de miel (por 1.000) correlaciona fuertemente con la tasa de matrimonios en Carolina del Sur (por 1.000 matrimonios)

Pearson Correlations

			Pearson's r	p
Honey bees	-	Marriage rate	0.938	< .001



REGRESIÓN

Mientras que las pruebas de correlación se usan para las asociaciones entre variables, la regresión es el paso siguiente usado habitualmente para los análisis predictivos, es decir, para predecir una variable de resultado dependiente a partir de una (regresión simple) o más (regresión múltiple) variables predictivas independientes.

La regresión resulta en un modelo hipotético de relación entre la variable resultado y una o más variables predictivas. El modelo usado es lineal, definido por la fórmula:

$$y = c + b*x + \varepsilon$$

- y = puntuación de la variable de resultado dependiente estimada
- c = constante
- b = coeficiente de regresión
- x = puntuación de la variable independiente predictiva
- ε = componente de error aleatorio (basado en los residuos)

La regresión lineal proporciona tanto la constante como el o los coeficientes de regresión.

La regresión lineal asume los siguientes supuestos:

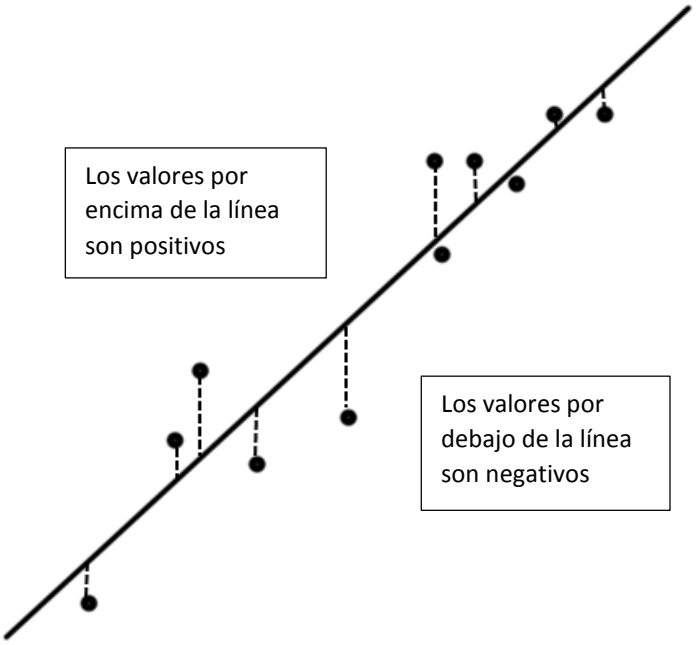
1. **Relación lineal**: es importante revisar los valores atípicos, ya que la regresión lineal es sensible a sus efectos.
2. **Independencia** de las variables.
3. **Normalidad multivariante**: requiere que todas las variables estén distribuidas normalmente.
4. **Homocedasticidad**: homogeneidad de la varianza de los residuos.
5. **Multicolinealidad / autocorrelación mínima**: cuando las variables independientes / los residuos están muy correlacionados entre sí.

Respecto a los tamaños de las muestras, hay mucha literatura sobre distintas reglas generales que van desde los 10-15 puntos de datos por predictor incluido en el modelo (es decir, 4 variables predictivas requerirán entre 40 y 60 puntos de datos) a 50 puntos + (8*número de predictores). Así, 4 variables requerirían 82 puntos de datos (50 + 8 * 4 = 50 + 32 = 82). En cualquier caso, cuanto mayor sea el tamaño de la muestra, mejor será el modelo.

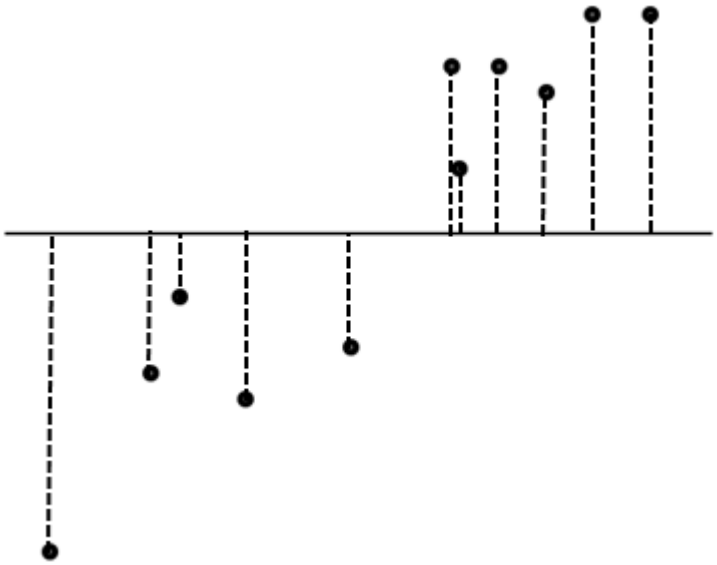
SUMAS DE CUADRADOS (Aburrido, pero básico para la evaluación del modelo de regresión)

La mayoría de los análisis de regresión producirán el mejor modelo posible, pero este modelo, ¿cuán bueno es en realidad y cuánto error se comete con él?

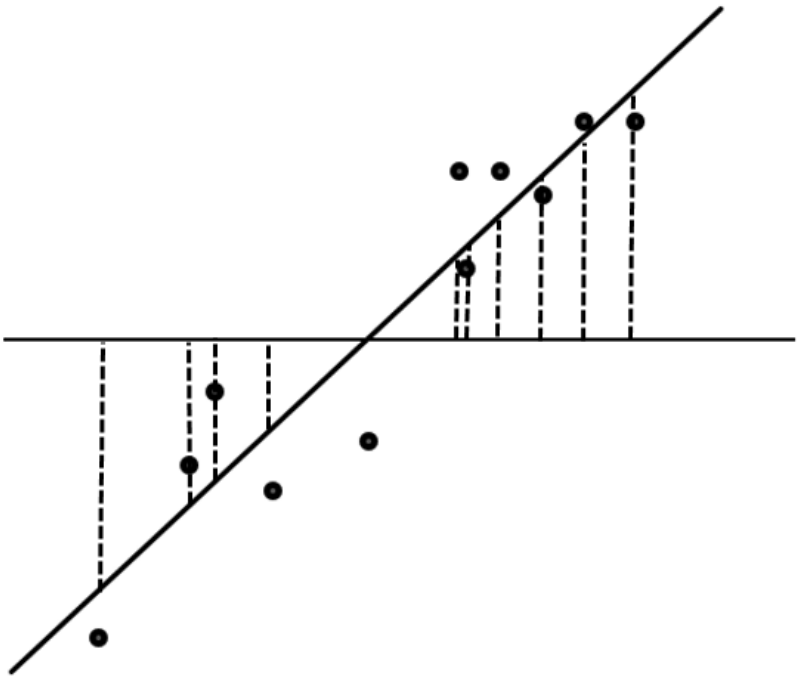
Esto se puede determinar comprobando la “bondad de ajuste” basada en las sumas de cuadrados. Se trata de una medida para determinar cuán cerca están los puntos de datos reales de la línea de regresión modelada.



La diferencia vertical entre los puntos de datos y la línea de regresión predicha se conocen por el nombre de **residuos**. Estos valores se elevan al cuadrado para eliminar los números negativos y luego se suman para obtener $\mathbf{SS_R}$ (SC_R, suma de cuadrados de los residuos, en castellano). Este es, efectivamente, el error del modelo o "**bondad de ajuste**"; de modo que cuanto más pequeño sea el valor, menor error habrá en el modelo.



Se puede calcular la diferencia vertical entre los puntos de datos y la media de la variable resultado. Estos valores se elevan al cuadrado para eliminar los números negativos y luego se suman para obtener la suma **total** de cuadrados $\mathbf{SS_T}$ (SC_T, suma de cuadrados total en castellano). Esto muestra cuán bueno es el valor medio como modelo de las puntuaciones de la variable resultado.



Ahora, podemos determinar la diferencia vertical entre la media de la variable resultado y la línea de regresión predicha. De nuevo, estos valores se elevan al cuadrado para eliminar los números negativos y luego se suman para obtener la suma de cuadrados del modelo **SS_M** (SC_M, suma de cuadrados del modelo en castellano). Esto indica cómo de bueno es el modelo comparado con el uso únicamente de la media de la variable resultado.

Por lo tanto, cuanto mayor sea **SC_M** mejor será el modelo para predecir el resultado comparado con el valor medio por sí solo. Si viene acompañado de un pequeño **SC_R** el modelo también tendrá un error pequeño.

R^2 es similar al coeficiente de determinación en la correlación, en tanto que muestra hasta qué punto la variación en la variable resultado puede ser predicha por la(s) variable(s) predictiva(s).

$$R^2 = \frac{SC_M}{SC_R}$$

En la regresión, el modelo se evalúa mediante el estadístico F que se basa en la mejora de la predicción del modelo (**SC_M**) y el error (**SC_R**). Cuanto mayor sea el valor de F, mejor será el modelo.

$$F = \frac{\text{Media } SC_M}{\text{Media } SC_R}$$

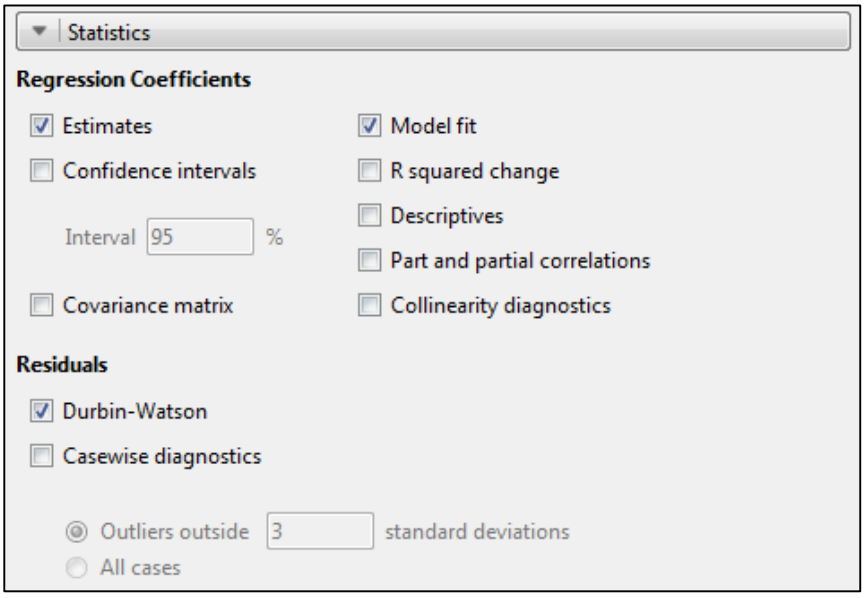
REGRESIÓN SIMPLE

La regresión pone a prueba la hipótesis nula (H_o) de que la(s) variable(s) predictiva(s) no predecirá(n) significativamente la variable dependiente (resultado).

Abra **Rugby kick regression.csv**. Este archivo contiene datos sobre pateos en el rugby, incluyendo la distancia recorrida, la fuerza y la flexibilidad de la pierna derecha / izquierda, y la fuerza de pierna bilateral.

Primero, vaya a «Descriptives» ⟶ «Descriptive statistics» y compruebe los gráficos de caja por si hubiera valores atípicos. En este caso no debería haber ninguno, pero la comprobación es una buena práctica.

Para esta regresión simple, vaya a «Regression» ⟶ «Linear regression» e introduzca la distancia en «Dependent Variable» (outcome), y R_Strength en la caja «Covariates» (Predictor). Marque las siguientes opciones en «Statistics»:



ENTENDIENDO EL RESULTADO

Ahora obtendrá los siguientes resultados:

Model Summary

Model	R	R^2	Adjusted R^2	RMSE	Durbin-Watson
1	0.784	0.614	0.579	55.285	1.524

Aquí se puede ver que la correlación (R) entre las dos variables es alta (0,784). El valor R^2 de 0,614 nos dice que la fuerza de la pierna derecha representa el 61,4% de la varianza en la distancia de pateo. Durbin-Watson comprueba las correlaciones entre los residuos, lo que podría invalidar el test. Debería estar por encima de 1 y por debajo de 3, idealmente cerca de 2.

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
1	Regression	53589.863	1	53589.863	17.533	0.002
	Residual	33621.061	11	3056.460		
	Total	87210.923	12			

La tabla de ANOVA muestra todas las sumas de los cuadrados antes mencionados. “Regression” es el modelo y “Residual” el error. El estadístico F es significativo p = 0,002. Esto nos dice que el modelo es un predictor de la distancia de pateo significativamente mejor que la distancia media.
Reporte de la siguiente manera: **F (1, 11) = 17,53, p < 0,001.**

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p
1	(Intercept)	57.105	103.588		0.551	0.592
	R_Strength	6.425	1.534	0.784	4.187	0.002

Esta tabla proporciona los coeficientes no estandarizados (“Unstandardized”) que pueden introducirse en la ecuación lineal.

y = c + b*x

y = puntuación estimada de la variable dependiente resultado.

c = constante (“**(Intercept)**”).

b = coeficiente de regresión (“**R_Strength**”).

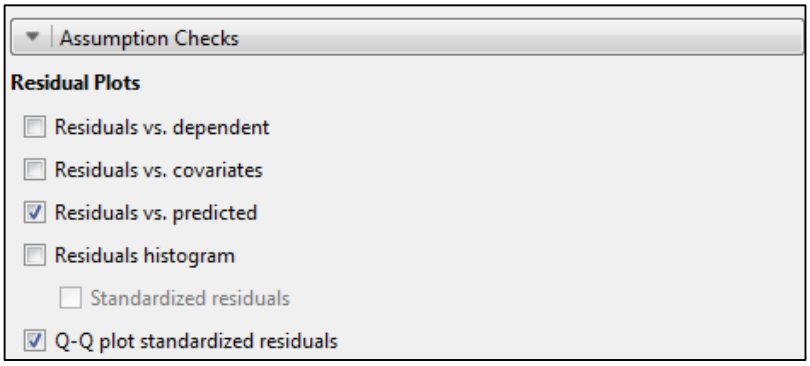
x = puntuación en la variable predictiva independiente.

Por ejemplo, para una fuerza de pierna de 60 kg, la distancia de pateo se puede predecir con la fórmula siguiente:

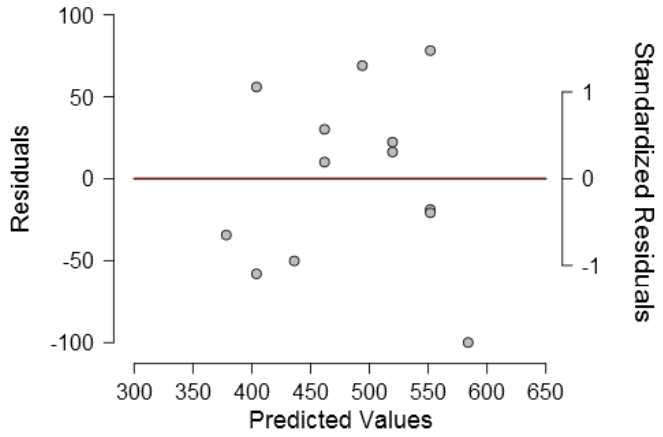
Distancia = 57,105 + (6,452 * 60) = 454,6 m

COMPROBACIONES ADICIONALES

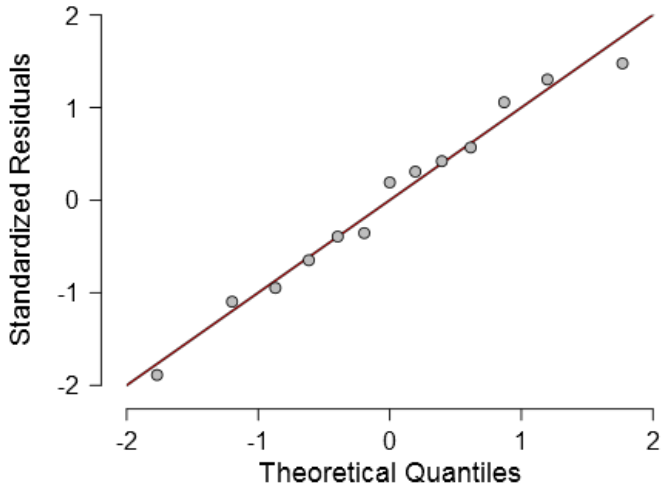
En «Assumption Checks», seleccione las dos opciones siguientes:



A partir de esto se obtendrán dos gráficos:



Este gráfico muestra una distribución aleatoria equilibrada de los residuos alrededor de la línea de base, sugiriendo que el supuesto de homocedasticidad no ha sido violado. Ver "Exploración de la integridad de los datos en JASP" para más detalles.



El gráfico Q-Q muestra que los residuos estandarizados coinciden con la diagonal, sugiriendo que los dos supuestos de normalidad y linealidad no han sido violados.

REPORTANDO LOS RESULTADOS

La regresión lineal muestra que la fuerza de la pierna derecha puede predecir significativamente la distancia de pateo F (1,11) = 17,53, p < 0,001 usando la siguiente ecuación de regresión:

Distancia = 57,105 + (6,452 * fuerza de la pierna derecha)

REGRESIÓN MÚLTIPLE

El modelo usado sigue siendo lineal, definido por la fórmula:

$$y = c + b*x + \varepsilon$$

- y = puntuación estimada de la variable dependiente resultado.
- c = constante.
- b = coeficiente de regresión.
- x = puntuación en la variable predictiva independiente.
- ε = componente de error aleatorio (basado en los residuos).

No obstante, ahora tenemos más de 1 coeficiente de regresión para la puntuación en cada variable predictiva. Es decir:

$$y = c + b_1*x_1 + b_2*x_2 + b_3*x_3 \ldots b_n*x_n$$

Métodos de entrada de datos

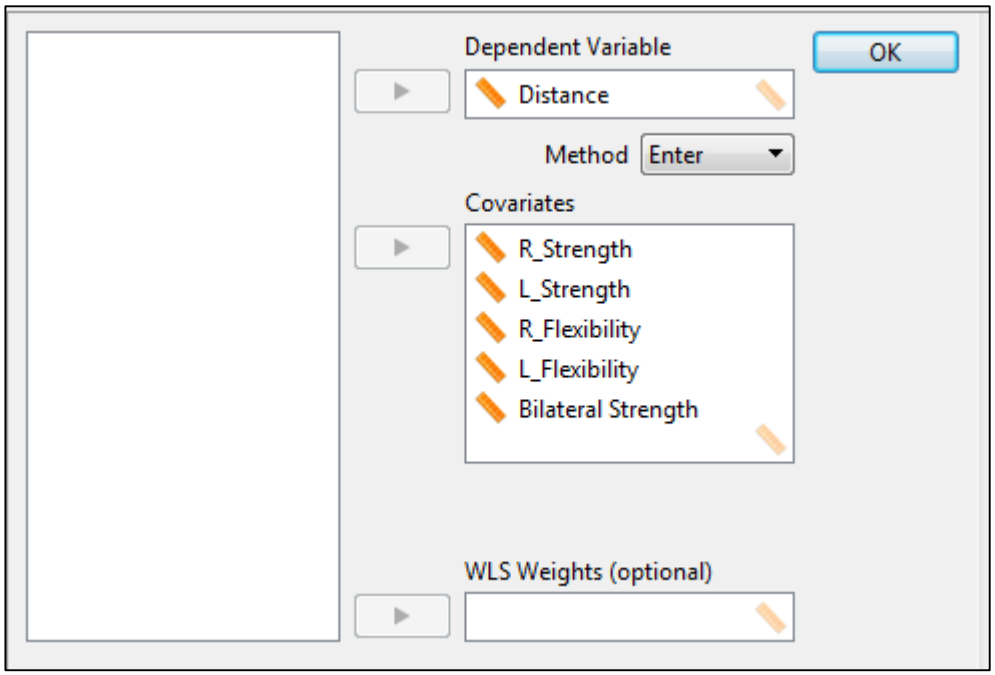
Si las variables predictivas no están correlacionadas, su orden de entrada no tiene importancia para el modelo. En la mayoría de los casos, las variables predictivas están en alguna medida correlacionadas y, por ello, el orden en el que se introduzcan puede tener consecuencias. Los distintos métodos disponibles han sido objeto de un gran debate.

a) Entrada **forzada** («**Enter**»): este es el **método por defecto** en el que se fuerza la entrada de las variables predictoras en el orden en que aparecen en la caja de covariables («Covariates»). Se considera el mejor método.
b) Entrada **por bloques** (*hierarchical entry*): el investigador, normalmente basado en conocimientos y estudios previos, decide en primer lugar el orden en el que se introducen las variables predictoras, en función de su importancia en la predicción de la variable resultado. En pasos posteriores se introducen predictoras adicionales.
c) Entrada **por pasos hacia atrás** («**Backward**»): todas las variables predictoras se introducen inicialmente en el modelo y se calcula la contribución de cada una de ellas. Se eliminan las predictoras con un nivel de contribución inferior al nivel establecido ($p < 0,1$). Se repite el proceso hasta que todas las variables predictoras que se conservan en el modelo son estadísticamente significativas.
d) Entrada **por pasos hacia adelante** («**Forward**»): se introduce, en primer lugar, la variable predictora con la correlación simple más alta respecto a la variable resultado. Las predictoras subsiguientes se eligen en función del tamaño de su correlación semiparcial respecto a la variable resultado. Este proceso se repite hasta que han quedado incluidas todas las predictoras que contribuyen con una variación única significativa al modelo.
e) Entrada **por pasos** («**Stepwise**»): similar al método de entrada hacia adelante («Forward»), excepto que cada vez que se añade una variable predictora al modelo, se realiza un test para eliminar la predictora menos útil. El modelo se revisa constantemente para comprobar si las predictoras redundantes pueden ser eliminadas.

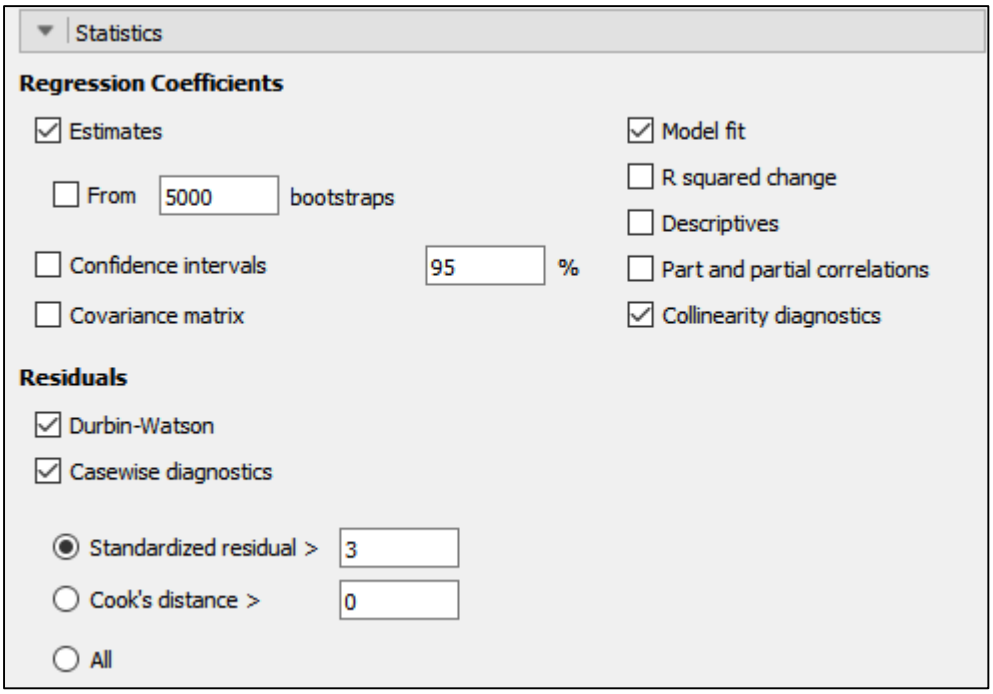
Se han descrito muchos inconvenientes relacionados con el uso de métodos de entrada por pasos. Sin embargo, el método **hacia atrás** puede ser útil para explorar variables predictoras no utilizadas previamente o para afinar el modelo con el fin de seleccionar las mejores de entre las disponibles.

EJECUTANDO LA REGRESIÓN MÚLTIPLE

Abra **Rugby kick regression.csv**, que también hemos usado para la regresión simple. Vaya a «Regression» ⟶ «Linear regression», introduzca la distancia en la caja «Dependent Variable» (resultado) y el resto de variables en la caja «Covariates» (predictoras).



En «Method» deje el método de entrada forzada («**Enter**») que aparece por defecto. Marque las siguientes opciones en «Statistics options»: «Estimates», «Model fit», «Collinearity diagnostics» y marque «Durbin-Watson».



ENTENDIENDO EL RESULTADO

Ahora, obtendrá los siguientes resultados:

Model Summary

Model	R	R^2	Adjusted R^2	RMSE	Durbin-Watson
1	0.902	0.814	0.681	48.132	1.328

El R^2 ajustado (usado para múltiples predictoras) muestra que se puede predecir un 68,1% de la varianza de la variable resultado. Durbin-Watson comprueba que las correlaciones entre los residuos se encuentran entre 1 y 3, como se requiere.

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
1	Regression	70994.078	5	14198.816	6.129	0.017
	Residual	16216.845	7	2316.692		
	Total	87210.923	12			

La tabla de ANOVA muestra que el estadístico F es significativo p = 0,017, sugiriendo que el modelo predice significativamente mejor la distancia de pateo que la distancia media.

Coefficients

							Collinearity Statistics	
Model		Unstandardized	Standard Error	Standardized	t	p	Tolerance	VIF
1	(Intercept)	−92.367	218.389		−0.423	0.685		
	R_Strength	1.747	3.321	0.213	0.526	0.615	0.162	6.180
	L_Strength	0.703	3.590	0.086	0.196	0.850	0.138	7.231
	R_Flexibility	4.078	4.759	0.373	0.857	0.420	0.140	7.125
	L_Flexibility	−1.339	2.447	−0.135	−0.547	0.601	0.438	2.281
	Bilateral Strength	1.665	0.946	0.423	1.759	0.122	0.458	2.181

Esta tabla muestra un modelo y la constante ("(Intercept)"), y los coeficientes de regresión ("Unstandardized") para todos los predictores forzados en el modelo. Aunque la tabla de ANOVA muestre que el modelo es significativo, ¡ninguno de los coeficientes de regresión predictivos lo es!

Los estadísticos de colinealidad, tolerancia y VIF (siglas en inglés de *Variance Inflation Factor*, o factor de inflación de la varianza) comprueban el supuesto de multicolinealidad. Como regla general, si el VIF > 10 y la tolerancia < 0,1, el supuesto ha sido ampliamente violado. Si el **promedio** de los valores del VIF > 1 y la tolerancia < 0,2, el modelo podría estar sesgado. En este caso, el promedio del VIF es bastante grande (alrededor de 5).

¡La tabla de diagnóstico por casos ("Casewise Diagnostics") está vacía! Son buenas noticias. Esta tabla muestra los casos (filas) con residuos que se encuentren a 3 o más desviaciones estándar respecto a la media. Estos casos con los errores más grandes podrían ser valores atípicos. La presencia de demasiados valores atípicos tendrá un impacto sobre el modelo y deberían tratarse del modo habitual (ver "Exploración de la integridad de los datos").

Casewise Diagnostics

Case Number	Std. Residual	Distance	Predicted Value	Residual	Cook's Distance

Como comparación, vuelva a ejecutar los análisis, pero esta vez eligiendo «Backward» (hacia atrás) como método de entrada.

Los resultados son los siguientes:

Model Summary

Model	R	R^2	Adjusted R^2	RMSE	Durbin-Watson
1	0.902	0.814	0.681	48.132	
2	0.902	0.813	0.720	45.146	
3	0.897	0.805	0.740	43.505	
4	0.884	0.782	0.738	43.618	1.676

JASP ahora ha calculado 4 modelos de regresión potenciales. Se puede ver que cada modelo consecutivo incrementa el R^2 ajustado, donde el modelo 4 explica el 73,5% de la variable resultado. La puntuación Durbin-Watson también es más alta que con el método de entrada forzada (*Enter*).

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
1	Regression	70994.078	5	14198.816	6.129	0.017
	Residual	16216.845	7	2316.692		
	Total	87210.923	12			
2	Regression	70905.329	4	17726.332	8.697	0.005
	Residual	16305.594	8	2038.199		
	Total	87210.923	12			
3	Regression	70176.855	3	23392.285	12.359	0.002
	Residual	17034.068	9	1892.674		
	Total	87210.923	12			
4	Regression	68185.712	2	34092.856	17.920	< .001
	Residual	19025.211	10	1902.521		
	Total	87210.923	12			

La tabla de ANOVA indica que cada modelo sucesivo es mejor, tal como muestra el aumento del valor de la F y la mejora del valor de p.

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p	Collinearity Statistics	
							Tolerance	VIF
1	(Intercept)	−92.367	218.389		−0.423	0.685		
	R_Strength	1.747	3.321	0.213	0.526	0.615	0.162	6.180
	L_Strength	0.703	3.590	0.086	0.196	0.850	0.138	7.231
	R_Flexibility	4.078	4.759	0.373	0.857	0.420	0.140	7.125
	L_Flexibility	−1.339	2.447	−0.135	−0.547	0.601	0.438	2.281
	Bilateral Strength	1.665	0.946	0.423	1.759	0.122	0.458	2.181
2	(Intercept)	−110.347	185.840		−0.594	0.569		
	R_Strength	2.218	2.148	0.271	1.033	0.332	0.340	2.938
	R_Flexibility	4.501	3.978	0.411	1.131	0.291	0.177	5.658
	L_Flexibility	−1.370	2.291	−0.138	−0.598	0.566	0.440	2.272
	Bilateral Strength	1.605	0.840	0.408	1.910	0.092	0.512	1.954
3	(Intercept)	−116.892	178.772		−0.654	0.530		
	R_Strength	2.710	1.911	0.331	1.418	0.190	0.399	2.505
	R_Flexibility	2.886	2.814	0.264	1.026	0.332	0.328	3.048
	Bilateral Strength	1.642	0.807	0.418	2.033	0.073	0.515	1.944
4	(Intercept)	46.251	81.820		0.565	0.584		
	R_Strength	3.914	1.512	0.478	2.588	0.027	0.641	1.561
	Bilateral Strength	2.009	0.725	0.511	2.770	0.020	0.641	1.561

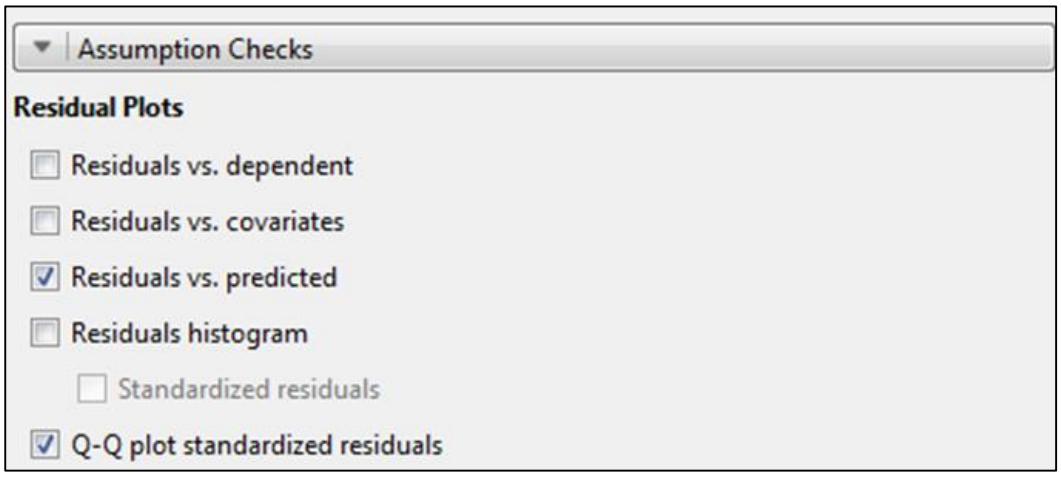
El modelo 1 es el mismo que el método de entrada forzada (*Enter*) usado en primer lugar. La tabla muestra que a medida en que se eliminan de modo secuencial las predictoras con una contribución significativamente menor, acabamos obteniendo un modelo con dos coeficientes de regresión predictivos significativos: la fuerza de la pierna derecha (R_Strength) y la fuerza de pierna bilateral (Bilateral Strength). Tanto la tolerancia como el VIF son aceptables.

Ahora podemos reportar que la entrada de las variables predictoras por pasos hacia atrás (*Backward*) resulta en un modelo altamente significativo: **F (2, 10) = 17,92, p < 0,001**, y una ecuación de regresión como la que sigue:

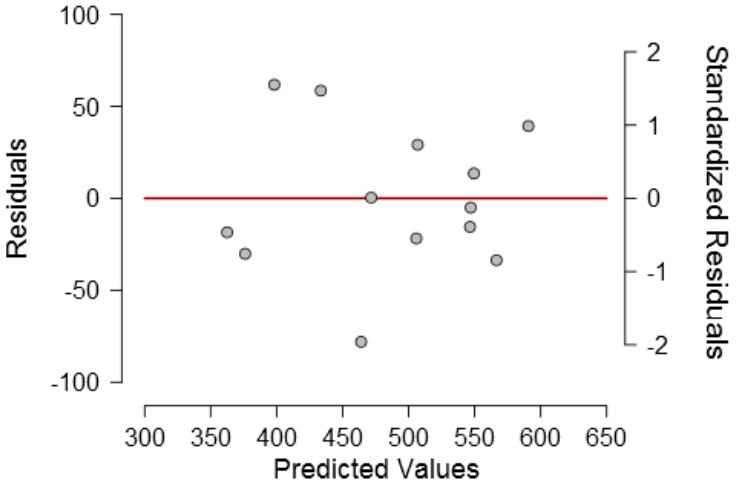
Distancia = 46,251 + (3,914 * R_Strength) + (2,009 * Bilateral Strength)

COMPROBACIÓN DE SUPUESTOS ADICIONALES

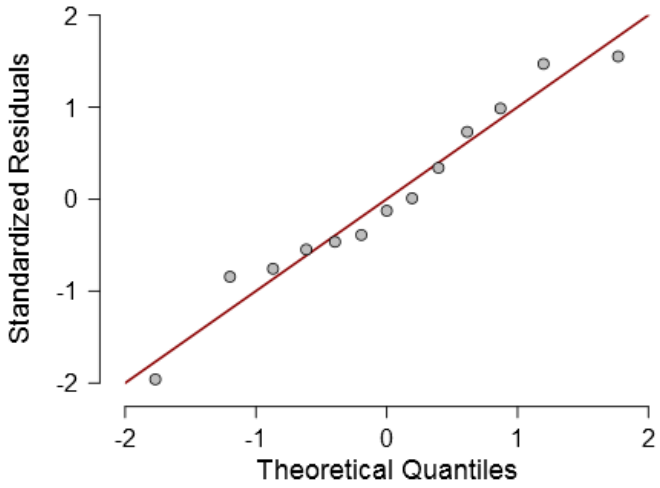
Como en el ejemplo de regresión lineal simple, seleccione las opciones siguientes.



Residuals vs. Predicted



Q-Q Plot Standardized Residuals ▼



La distribución equilibrada de los residuos alrededor de la línea base sugiere que el supuesto de homocedasticidad no ha sido violado.

El gráfico Q-Q muestra que los residuos estandarizados se ajustan a lo largo de la diagonal, lo que sugiere que ambos supuestos de normalidad y linealidad tampoco han sido violados.

REPORTANDO LOS RESULTADOS

La regresión lineal múltiple basada en el método de entrada por pasos hacia atrás muestra que la fuerza de pierna derecha y la fuerza bilateral pueden predecir significativamente la distancia de pateo **F (2,10) = 17,92, p < 0,001** usando la ecuación de regresión:

Distancia = 57,105 + (3,914 * R_Strength) + (2,009 * Bilateral Strength)

EN RESUMEN

R^2 proporciona información sobre cuánta varianza puede ser explicada utilizando las variables predictoras introducidas en el modelo.

El estadístico F proporciona información sobre cómo de bueno es el modelo.

El valor de los coeficientes no estandarizados proporciona una constante que refleja la fuerza de la relación entre cada una de las variables predictoras y la variable resultado.

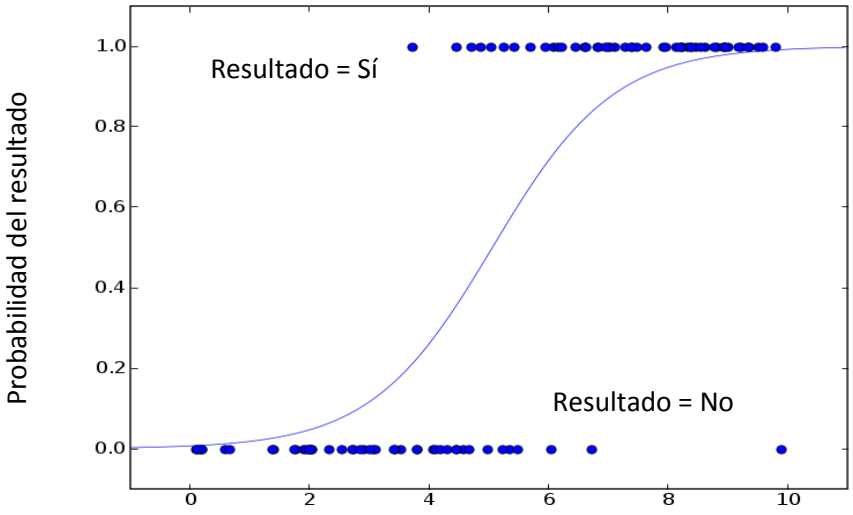
La violación de los supuestos puede ser comprobada usando el valor de Durbin-Watson, los valores de tolerancia / VIF y los gráficos de residuos *vs.* predichos y Q-Q.

REGRESIÓN LOGÍSTICA

En la regresión lineal simple y múltiple, la variable resultado y las variables predictoras eran continuas. ¿Pero qué sucedería si la variable resultado fuese una medida binaria / categórica? ¿Puede, por ejemplo, predecirse una variable resultado de sí o no, a partir de otras variables continuas o categóricas? La respuesta es que sí, si se utiliza una regresión logística binaria. Este método se usa para predecir la probabilidad de una variable resultado binaria de sí o no.

La hipótesis nula que se pone a prueba es que no existe relación entre las variables resultado y las predictoras.

Como se puede ver en el gráfico siguiente, una línea de regresión lineal entre las respuestas de sí y no tendría poco sentido como modelo predictivo. En su lugar, se ajusta una curva de regresión logística sigmoide con un mínimo en 0 y un máximo en 1. Puede verse que algunos valores de la variable predictora se superponen entre el sí y el no. Por ejemplo, un valor de 5 tendría una probabilidad del 50% de resultar en un sí o un no. Por lo tanto, se calcula un umbral para determinar si el valor en una variable predictora se clasificará como un sí o como un no en la variable resultado.



SUPUESTOS DE LA REGRESIÓN LOGÍSTICA BINARIA

- La variable dependiente debe ser binaria, es decir, sí o no, hombre o mujer, bueno o malo.
- Una o más (variables predictivas) independientes que pueden ser variables categóricas o continuas.
- Una relación lineal entre las variables independientes continuas y la transformación *logit* (logaritmo natural de la probabilidad de que la variable resultado equivalga a una de las categorías) de la variable dependiente.

MÉTRICAS DE LA REGRESIÓN LOGÍSTICA

AIC (por las siglas en inglés de *Akaike Information Criteria,* o Criterio de Información Akaike) y **BIC** (por *Bayesian Information Criteria*, o Criterio de Información Bayesiano) son medidas de ajuste para el modelo; el mejor modelo tendrá los valores AIC y BIC más bajos.

En JASP se calculan tres valores **pseudo R^2**: McFadden, Nagelkerke y Tjur. Estos son análogos al R^2 en la regresión lineal y todos proporcionan valores diferentes. Lo que constituye un buen valor de R^2 varía, pero son útiles cuando se comparan diferentes modelos con los mismos datos. Se considera que el mejor modelo es el que posea un valor en el estadístico R^2 más alto.

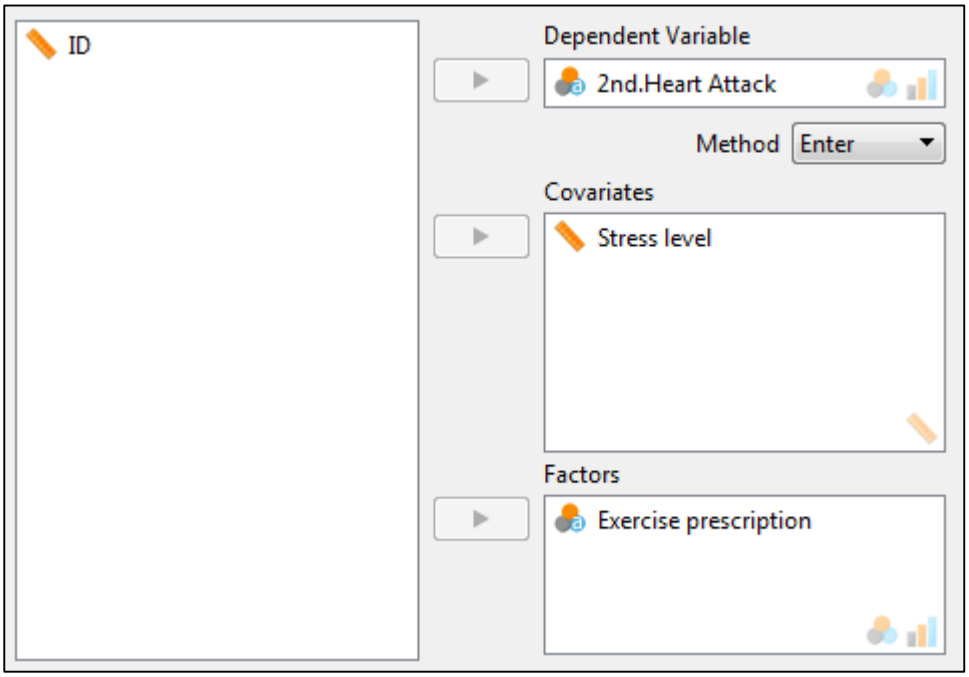
La **matriz de confusión** (***confusion matrix***) es una tabla que muestra los resultados reales *vs*. los predichos, y puede ser utilizada para determinar la precisión del modelo. A partir de ella, pueden derivarse la sensibilidad y la especificidad.

La **sensibilidad** (***sensitivity***) es el porcentaje de casos en los que el resultado observado fue predicho correctamente por el modelo (es decir, verdaderos positivos).

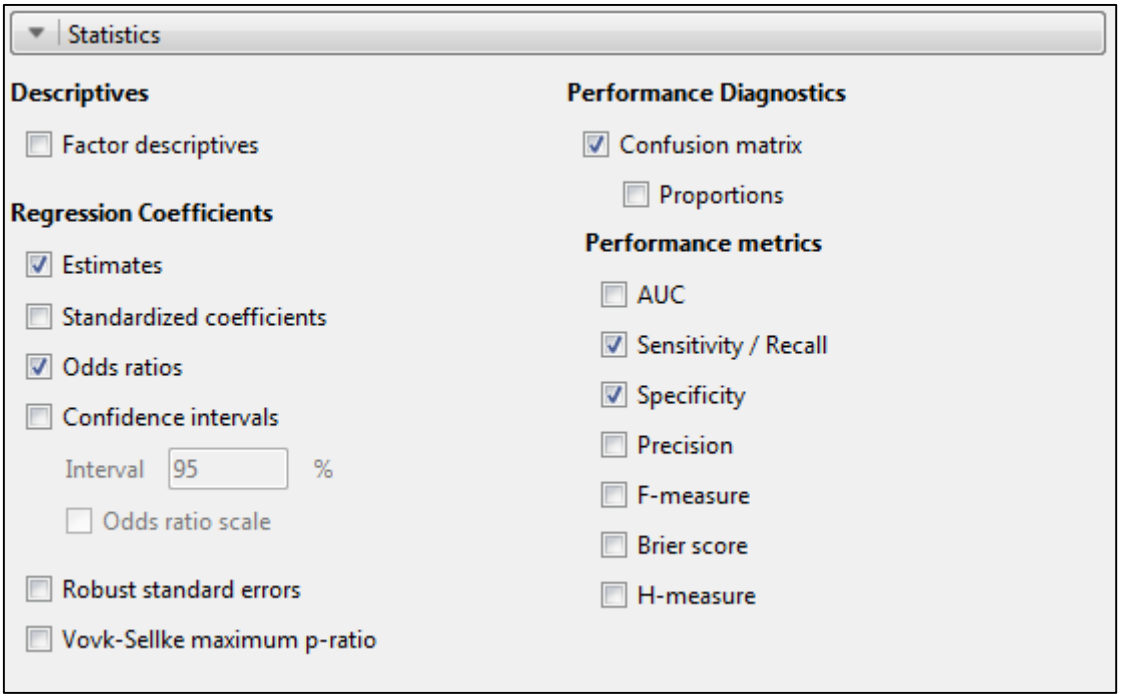
La **especificidad** (***specificity***) es el porcentaje de observaciones que también se predijeron correctamente como aquellas que no tenían los resultados observados (es decir, verdaderos negativos).

EJECUTANDO LA REGRESIÓN LOGÍSTICA

Abra **Heart attack.csv** en JASP. Este archivo contiene 4 columnas de datos: la ID de paciente (Patient ID), si sufrieron un segundo ataque al corazón (sí / no), si se les prescribió ejercicio (sí / no) y sus niveles de estrés (valor alto = estrés alto).
Ponga la variable de resultado (2nd.Heart.Attack) en la caja «Dependent Variable», añada los niveles de estrés a la caja «Covariates» y la prescripción de ejercicio en la caja «Factors». Deje el método de entrada como forzada («Enter»).



En «Statistics», marque «Estimates», «Odds ratios», «Confusion matrix», «Sensitivity» y «Specificity».



ENTENDIENDO EL RESULTADO

El resultado inicial consiste en 4 tablas.
El resumen del modelo muestra que H1 (con las puntuaciones de AIC y BIC más bajas) sugiere una relación significativa ($X^2(37) = 21{,}257$, $p < 0{,}001$) entre la variable resultado (Segundo ataque al corazón) y las variables predictoras (prescripción de ejercicio y niveles de estrés).

Model summary

Model	Deviance	AIC	BIC	df	X^2	p	McFadden R^2	Nagelkerke R^2	Tjur R^2
H_0	55.452	57.452	59.141	39					
H_1	34.195	40.195	45.261	37	21.257	< .001	0.383	0.550	0.126

El R^2 de McFadden = 0,383. Se suele aceptar que un valor en un rango entre 0,2 y 0,4 indica un buen ajuste del modelo.

Coefficients

	Estimate	Standard Error	Odds Ratio	z	p
(Intercept)	−4.368	2.550	0.013	−1.713	0.087
Stress level	0.089	0.041	1.093	2.159	0.031
Exercise prescription (Yes)	−2.043	0.890	0.130	−2.295	0.022

Note. 2nd.Heart Attack level 'Yes' coded as class 1.

Tanto el nivel de estrés como la prescripción de ejercicio son variables predictoras significativas (p = 0,031 y 0,022, respectivamente). Los valores más importantes en la tabla de coeficientes son las razones de probabilidades ("Odds Ratio"). Para una variable predictora continua, una razón de probabilidades mayor que 1 sugiere una relación positiva mientras que < 1 implica una relación negativa. Esto sugiere que altos niveles de estrés están significativamente relacionados con una mayor probabilidad de tener un segundo ataque al corazón. La razón de probabilidades de 0,13 se puede interpretar como la existencia de solo un 13% de probabilidad de sufrir un segundo ataque cardíaco cuando se realiza ejercicio físico.*

Confusion matrix

	Predicted	
Observed	No	Yes
No	15.000	5.000
Yes	5.000	15.000

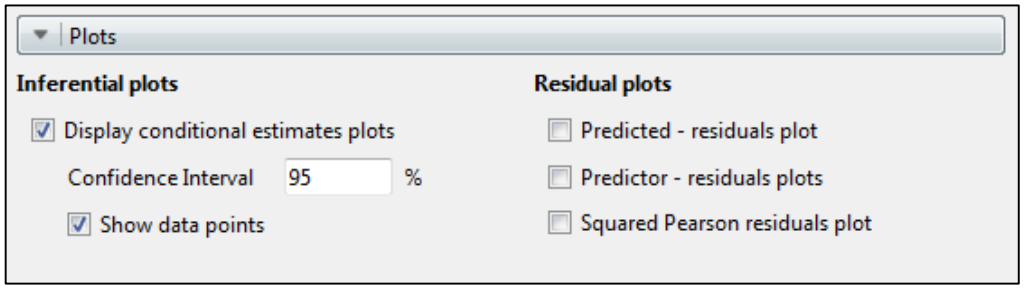
Performance metrics

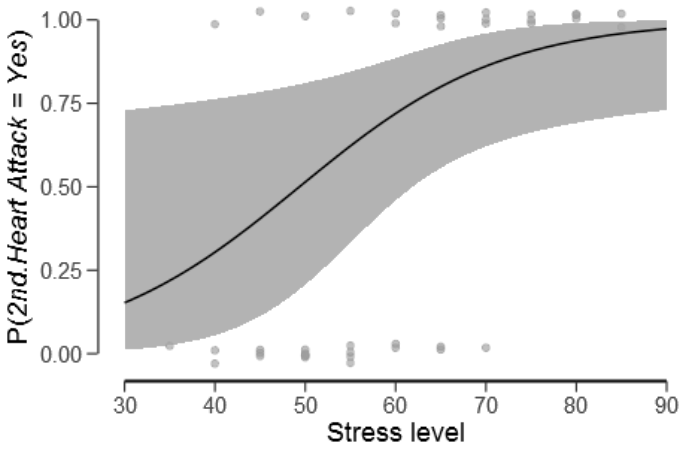
	Value
Sensitivity	0.750
Specificity	0.750

La matriz de confusión muestra que el modelo ha predicho correctamente 15 casos de verdaderos negativos y otros 15 más de verdaderos positivos, mientras que ha cometido un error en 5 casos de falsos negativos y otros 5 de falsos positivos. Estos resultados se reflejan en las métricas de rendimiento ("Performance metrics"), donde la sensibilidad (% de los casos con el resultado correctamente predicho) y la especificidad (% de casos correctamente predichos como aquellos que no tienen resultado; es decir, verdaderos negativos) son ambas del 75%.

GRÁFICOS

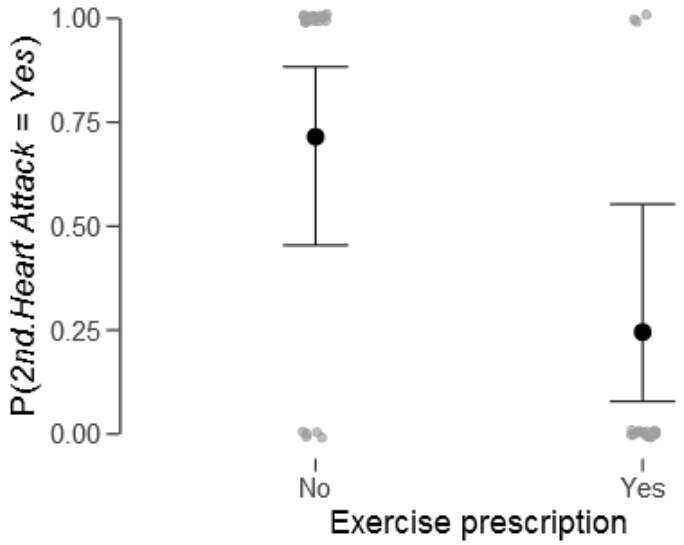
Estos resultados se pueden visualizar fácilmente a través de los gráficos inferenciales («Inferential plots»).





*O lo que es lo mismo, el ejercicio físico está asociado con una reducción del 87% (0,13 - 1 = -0,87) de las probabilidades de sufrir un segundo ataque al corazón en comparación con aquellos que no lo hicieron. (Nota del revisor.)

Si se incrementa el nivel de estrés, aumenta la probabilidad de sufrir un segundo ataque al corazón.



Si no se realiza ejercicio, aumenta la probabilidad de sufrir un segundo ataque al corazón, mientras que se reduce si ha prescrito.

REPORTANDO LOS RESULTADOS

Se realizó una regresión logística para determinar los efectos del estrés y la prescripción de ejercicio físico sobre la probabilidad de que los participantes sufrieran un segundo ataque al corazón. El modelo de regresión logística fue estadísticamente significativo, χ^2 (37) = 21,257, p < 0,001. El modelo clasificó correctamente el 75,0% de los casos. El incremento del estrés se asoció con un aumento de la probabilidad de sufrir un segundo ataque cardíaco, mientras que una reducción del estrés se asoció con una disminución de esta probabilidad. La prescripción de un programa de ejercicio redujo al 13% la probabilidad de un segundo ataque al corazón.

COMPARACIÓN DE MÁS DE DOS GRUPOS INDEPENDIENTES

ANOVA

Mientras que las pruebas t comparan las medias de dos grupos / condiciones, el análisis de varianza (**ANOVA,** *ANalysis Of VAriance*) de un solo factor (o unifactorial) compara las medias de 3 o más grupos / condiciones. En JASP están disponibles los dos tipos de ANOVA, de medidas independientes (o muestras independientes) y de medidas repetidas (o muestras relacionadas). El ANOVA ha sido descrito como una "prueba ómnibus" (global) que proporciona un estadístico F que compara si la varianza explicada es significativamente mayor que la varianza no explicada. **La hipótesis nula que se pone a prueba es que no hay diferencia significativa entre las medias de todos los grupos**. Si la hipótesis nula es rechazada, el ANOVA simplemente afirma que hay una diferencia significativa entre los grupos, pero no dónde se hallan estas diferencias. Para determinar en qué grupos se encuentran las diferencias, a continuación se deben llevar a cabo pruebas post hoc (del latín *post hoc*, que significa "después de esto").

¿Por qué no realizar simplemente múltiples comparaciones entre pares? Si hay 4 grupos (A, B, C, D), por ejemplo, y las diferencias fueran comparadas usando múltiples pruebas t:

- A *vs*. B P < 0,05 95% sin error Tipo I
- A *vs*. C P < 0,05 95% sin error Tipo I
- A *vs*. D P < 0,05 95% sin error Tipo I
- B *vs*. C P < 0,05 95% sin error Tipo I
- B *vs*. D P < 0,05 95% sin error Tipo I
- C *vs*. D P < 0,05 95% sin error Tipo I

Asumiendo que cada prueba fuera independiente, la probabilidad global sería entonces:

$$0{,}95 * 0{,}95 * 0{,}95 * 0{,}95 * 0{,}95 * 0{,}95 = 0{,}735$$

Esto se conoce como *familywise error* o error de Tipo I acumulativo, y en este caso resulta en solo un 73,5% de probabilidad de que no cometamos un error de Tipo I, por lo que la hipótesis nula podría rechazarse cuando en realidad es verdadera. Esto se evita con las pruebas post hoc, que realizan comparaciones múltiples por parejas con criterios de aceptación más estrictos y permiten así prevenir este tipo de error.

SUPUESTOS

El ANOVA de medidas independientes tiene los mismos supuestos que la mayoría de los demás test paramétricos.

- La variable independiente debe ser categórica y la variable dependiente debe ser continua.
- Los grupos deben ser independientes entre sí.
- La variable dependiente debe tener una distribución aproximadamente normal.
- No debería haber valores atípicos significativos.
- Debería haber homogeneidad de varianza entre los grupos; de otro modo, el valor de p para el estadístico F podría no ser fiable.

Normalmente, los 2 primeros supuestos se controlan con un diseño de investigación adecuado.

Si los tres últimos supuestos son violados, entonces debería considerarse la prueba de Kruskal-Wallis, su equivalente no paramétrico.

PRUEBAS POST HOC

JASP proporciona 4 alternativas para llevar a cabo con la prueba de ANOVA de medidas independientes:

a) ***Bonferroni*** – puede ser muy conservador, pero ofrece garantías de control del error de Tipo I a riesgo de reducir la potencia estadística.
b) ***Holm*** – el test Holm-Bonferroni, un método Bonferroni secuencial menos conservador que el test Bonferroni original.
c) ***Tukey*** – uno de los test más frecuentemente usados que proporciona un error de Tipo I controlado para grupos con el mismo tamaño de muestra y la misma varianza.
d) ***Scheffe*** – controla el nivel global de confianza si los grupos tienen diferentes tamaños de muestra.

JASP también proporciona 4 tipos:

a) ***Standar*** – tal como se describen arriba les cuatro alternativas.
b) ***Games-Howell*** – se utiliza cuando no tenemos seguridad sobre la igualdad de las varianzas de los grupos.
c) **Dunnett's** – se usa cuando queremos comparar todos los grupos con uno solo, es decir, el grupo de control.
d) ***Dunn*** – un test post hoc no paramétrico que se usa para poner a prueba pequeños subgrupos de pares.

TAMAÑO DEL EFECTO

JASP proporciona 3 cálculos alternativos del tamaño del efecto para usar con la prueba de ANOVA de medidas independientes:

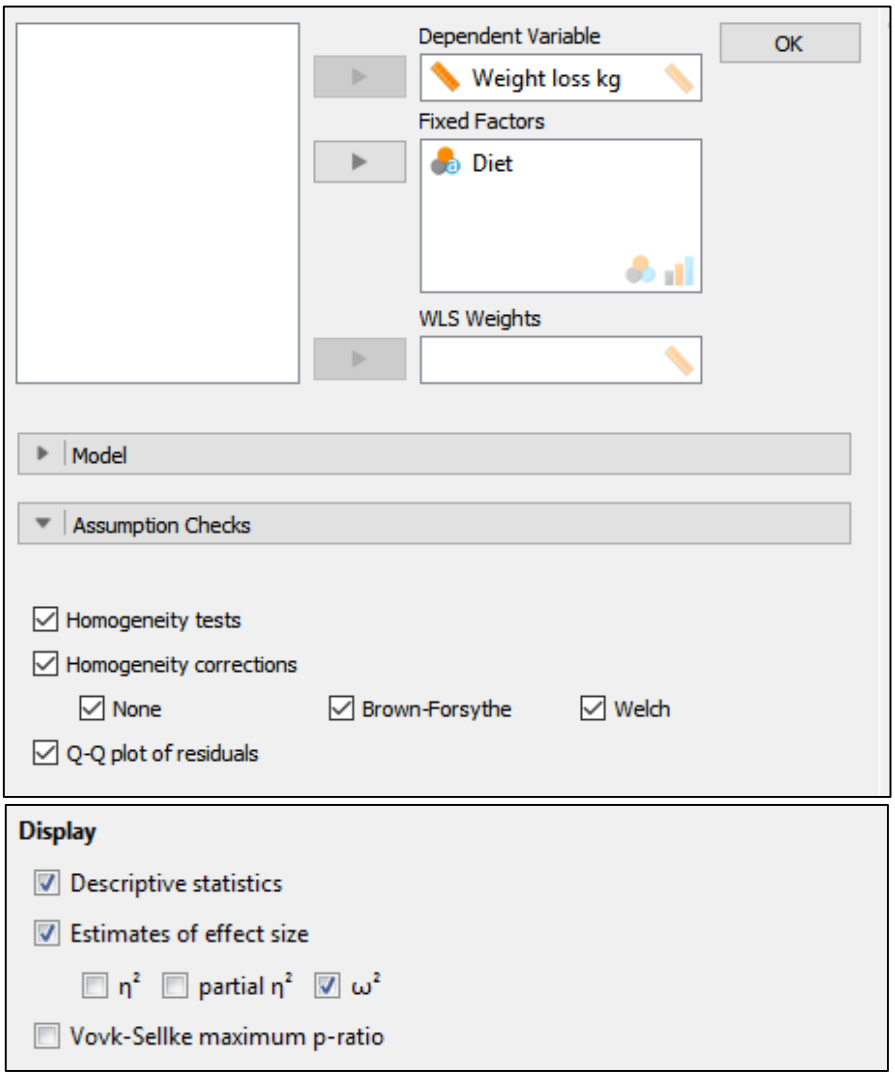
a) ***Eta cuadrado (η^2)*** – preciso para estimar la varianza explicada en la muestra pero sobreestima la varianza en la población. Dificulta la comparación del efecto de la misma variable en distintos estudios.
b) ***Eta cuadrado parcial (η_p^2)*** – resuelve el problema de la sobreestimación de la varianza en la población permitiendo la comparación del efecto de la misma variable en distintos estudios.
c) ***Omega cuadrado (ω^2)*** – normalmente, el sesgo estadístico deviene muy pequeño a medida que se incrementa el tamaño de muestra, pero para cuando tenemos muestras pequeñas ($n < 30$) ω^2 proporciona una medida no sesgada del tamaño del efecto.

Test	Medida	Irrelevante	Pequeño	Medio	Grande
ANOVA	**Eta**	**< 0,1**	**0,1**	**0,25**	**0,37**
	Eta parcial	**< 0,01**	**0,01**	**0,06**	**0,14**
	Omega cuadrado	**< 0,01**	**0,01**	**0,06**	**0,14**

EJECUTANDO EL ANOVA DE MEDIDAS INDEPENDIENTES

Cargue **Independent ANOVA diet.csv**. Este archivo contiene una columna A con 3 dietas usadas (A, B y C) y otra columna con la cantidad total de peso perdido tras 8 semanas siguiendo una de las 3 dietas diferentes. Es una buena práctica comprobar la estadística descriptiva y los gráficos de caja por si hubiera valores atípicos extremos.

Vaya a «ANOVA» ⟶ «ANOVA», introduzca la pérdida de peso en la caja «Dependent Variable» y las agrupaciones según la dieta en la caja de factores fijos («Fixed Factors»). En primera instancia, seleccione las comprobaciones de supuestos («Assumption Checks») y, en las opciones adicionales («Additional Options»), marque «Descriptive statistics» y ω^2 como tamaño del efecto:



Esto dará un resultado en 3 tablas y un gráfico Q-Q.

ENTENDIENDO EL RESULTADO

ANOVA - Weight loss kg

Cases	Homogeneity Correction	Sum of Squares	df	Mean Square	F	p
Diet	None	92.37	2.000	46.184	10.83	< .001
Diet	Brown-Forsythe	92.37	2.000	46.184	10.83	< .001
Diet	Welch	92.37	2.000	46.184	11.45	< .001
Residual	None	294.37	69.000	4.266		
Residual	Brown-Forsythe	294.37	64.352	4.574		
Residual	Welch	294.37	44.987	6.544		

Note. Type III Sum of Squares

La tabla de ANOVA principal muestra que el estadístico F es significativo ($p < 0,001$) y que hay un tamaño del efecto grande. Por lo tanto, hay una diferencia significativa entre las medias de los 3 grupos de dietas.

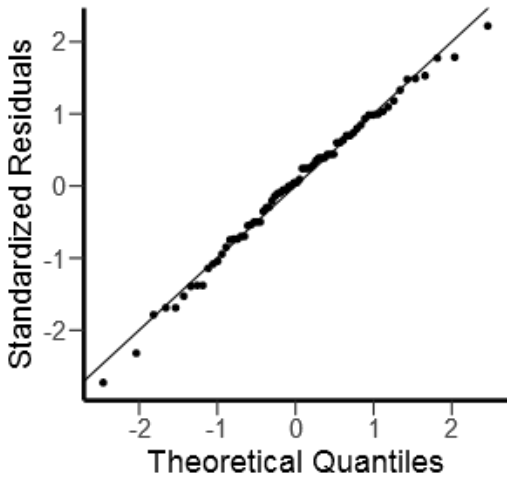
COMPROBACIÓN DE LOS SUPUESTOS

Antes de dar por buenos estos resultados, se debe comprobar que no se violan los supuestos requeridos por la prueba de ANOVA.

Test for Equality of Variances (Levene's)

F	df1	df2	p
1.298	2.000	69.000	0.280

El test de Levene muestra que la homogeneidad de la varianza no es significativa. Sin embargo, si la prueba de Levene muestra una diferencia significativa en la varianza, debería reportarse la corrección Brown-Forsythe o la de Welch.



El gráfico Q-Q muestra que los datos parecen tener una distribución normal y que son lineales.

Descriptives - Weight loss kg

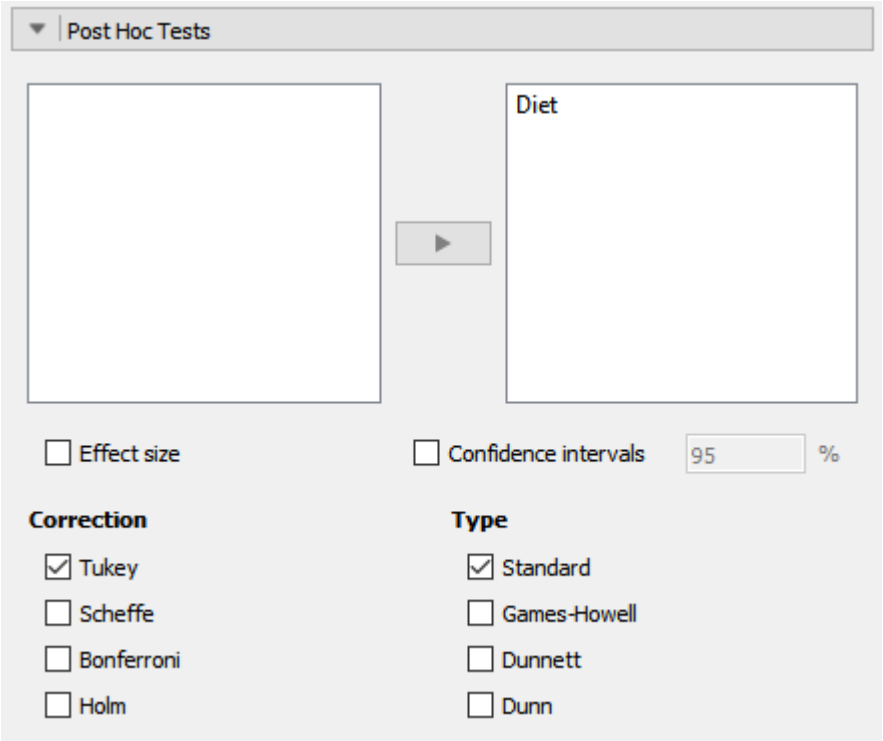
Diet	Mean	SD	N
Diet A	3.008	1.668	24.000
Diet B	3.413	2.361	24.000
Diet C	5.588	2.108	24.000

La estadística descriptiva sugiere que la dieta 3 consigue la mayor pérdida de peso tras 8 semanas.

Si el ANOVA no reporta una diferencia significativa, no debe proseguir con el análisis.

PRUEBAS POST HOC

Si el ANOVA es significativo, ahora se pueden llevar a cabo el análisis post hoc. En «Post Hoc Tests», añada Diet a la caja de análisis de la derecha, seleccione «Effect size» y, en este caso, use «Tukey» para la corrección post hoc.



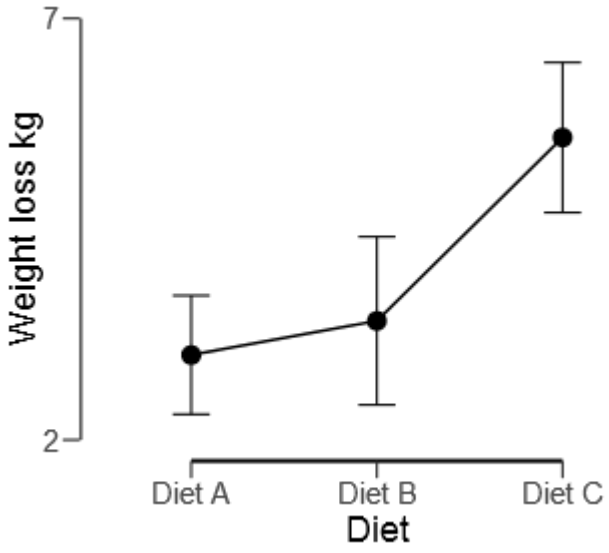
Añada también, en «Descriptive plots», el factor Diet al eje horizontal y elija «Display error bars».

Post Hoc Comparisons - Diet

		Mean Difference	SE	t	Cohen's d	p_{tukey}
Diet A	Diet B	−0.404	0.596	−0.678	−0.198	0.777
	Diet C	−2.579	0.596	−4.326	−1.357	< .001
Diet B	Diet C	−2.175	0.596	−3.648	−0.972	0.001

Note. Cohen's d does not correct for multiple comparisons.

El análisis post hoc muestra que no hay diferencia significativa en la pérdida de peso entre las dietas A y B. No obstante, es significativamente superior en la dieta C comparada con la dieta A ($p < 0,001$) y la dieta B ($p = 0,001$). La d de Cohen muestra que estas diferencias se corresponden con un tamaño del efecto grande.



REPORTANDO LOS RESULTADOS

El ANOVA unifactorial mostró un efecto significativo del tipo de dieta sobre la pérdida de peso tras 8 semanas **($F (2, 69) = 46,184$, $p < 0,001$, $\omega^2 = 0,214$)**

El análisis post hoc mediante la corrección de Tukey reveló que la dieta C consiguió una pérdida de peso significativamente superior que la dieta A ($p < 0,001$) o la dieta B ($p = 0,001$). No hubo diferencias significativas de pérdida de peso entre las dietas A y B ($p = 0,777$).

KRUSKAL-WALLIS: EL ANOVA NO PARAMÉTRICO

Si los datos no cumplen con los supuestos paramétricos o son de naturaleza nominal, la prueba *H* de Kruskal-Wallis es un equivalente no paramétrico al ANOVA para medidas o muestras independientes. Se puede usar para comparar dos o más grupos independientes con un tamaño de muestra igual o diferente. Como las pruebas de Mann-Whitney y Wilcoxon, es un test basado en rangos.

Como el ANOVA, la prueba *H* de Kruskal-Wallis (también conocida como ANOVA unifactorial por rangos) es una prueba global que no especifica qué grupos de la variable independiente son significativamente diferentes entre sí. Para poder hacerlo, JASP proporciona la opción de ejecutar la prueba post hoc de Dunn. Esta prueba de comparaciones múltiples puede ser muy conservadora, especialmente cuando se realizan un gran número de comparaciones.

Cargue el conjunto de datos **Kruskal-Wallis ANOVA.csv** en JASP. Este conjunto de datos contiene puntuaciones de dolor subjetivo en participantes que no reciben tratamiento (control), que reciben crioterapia o que reciben una combinación de crioterapia con compresión para tratar el dolor muscular de aparición tardía tras el ejercicio.

EJECUTANDO LA PRUEBA DE KRUSKAL-WALLIS

Vaya a «ANOVA» ⟶ «ANOVA». En la ventana de análisis, añada la puntuación de dolor a la caja de variable dependiente («Dependent Variable») y el tratamiento a la caja de factores fijos («Fixed Factors»). Compruebe que la puntuación de dolor está asignada como variable ordinal. Esto ejecutará, automáticamente, el ANOVA de medidas independientes convencional. En «Assumption Checks», seleccione «Homogeneity tests» y «Q-Q plot of residuals».

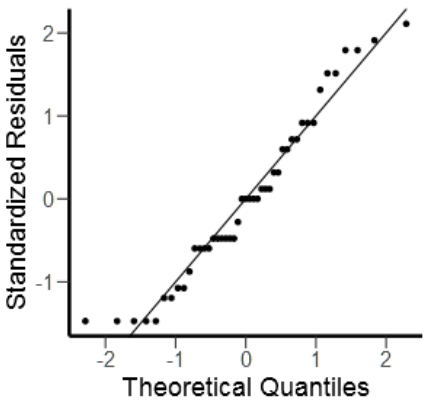
ANOVA - Pain Score

Cases	Sum of Squares	df	Mean Square	F	p
Treatment	98.844	2.000	49.422	16.457	< .001
Residual	126.133	42.000	3.003		

Note. Type III Sum of Squares

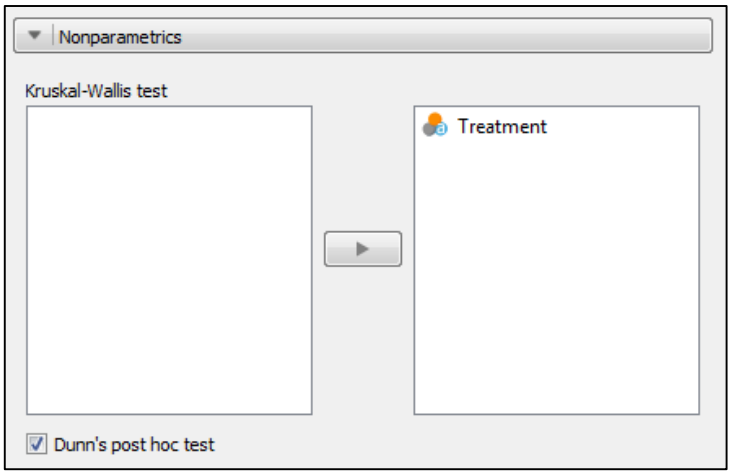
Test for Equality of Variances (Levene's)

F	df1	df2	p
3.832	2.000	42.000	0.030



Aunque el ANOVA muestra un resultado significativo, los datos no cumplen con los supuestos de homogeneidad de la varianza tal como se puede observar por el test significativo de Levene, y solo muestran linealidad en el centro del gráfico Q-Q que se curva en los extremos, indicando la presencia de valores extremos. Esto, añadido al hecho de que la variable dependiente está basada en puntuaciones de dolor subjetivo, sugiere el uso de una alternativa no paramétrica.

Vuelva a las opciones estadísticas y abra la opción «Nonparametrics», situada al final. Para obtener el test de Kruskal-Wallis, mueva la variable Treatment a la caja de la derecha y seleccione «Dunn's post hoc test».



ENTENDIENDO EL RESULTADO

El resultado muestra dos tablas. La prueba de Kruskal-Wallis muestra que hay una diferencia significativa entre las tres modalidades de tratamiento.

Kruskal-Wallis Test

Factor	Statistic	df	p
Treatment	19.693	2	< .001

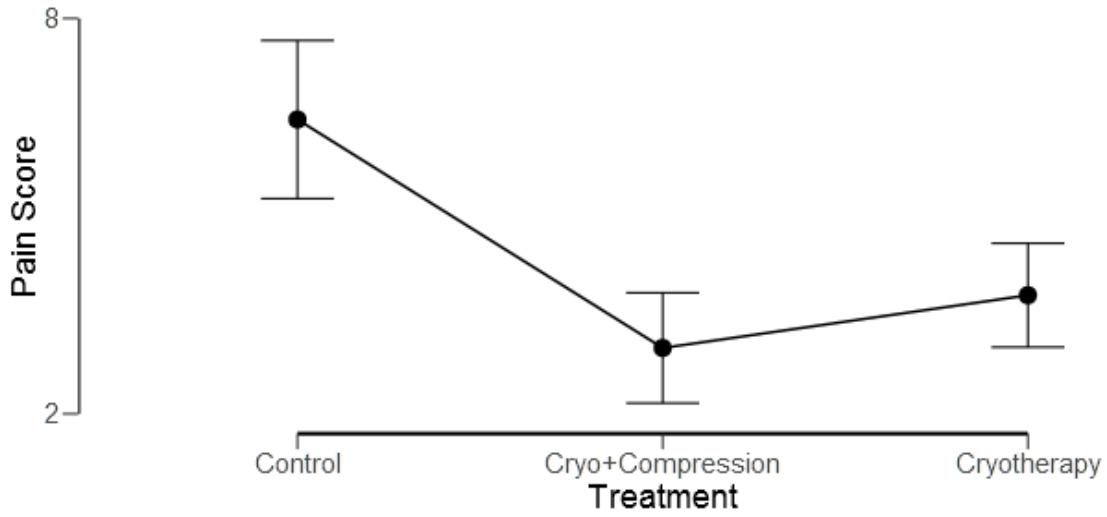
Dunn's Post Hoc Comparisons - Treatment

		z	W_i	W_j	p	p_{bonf}	p_{holm}
Control	Cryo+Compression	4.317	34.600	14.200	< .001	< .001	< .001
	Cryotherapy	3.048	34.600	20.200	0.001	0.003	0.002
Cryo+Compression	Cryotherapy	−1.270	14.200	20.200	0.102	0.306	0.102

El test post hoc de Dunn facilita su propio valor de p, así como los de Bonferroni y la corrección de Holm. Como se puede ver, ambas condiciones de tratamiento son significativamente diferentes respecto al grupo control, pero no entre sí.

REPORTANDO LOS RESULTADOS

Las puntuaciones de dolor estaban afectadas significativamente por la modalidad de tratamiento **H (2) = 19,693, p < 0,001**. Las comparaciones dos a dos mostraron que, tanto la crioterapia como la crioterapia con compresión, reducen significativamente las puntuaciones de dolor (p = 0,001 y p < 0,001, respectivamente) en comparación con el grupo de control. No hubo diferencias significativas entre la crioterapia y la crioterapia con compresión (p = 0,102).



COMPARACIÓN DE MÁS DE DOS GRUPOS RELACIONADOS

ANOVA MR

El ANOVA de un solo factor de medidas repetidas (**ANOVA MR**) se usa para evaluar si existen diferencias en las medidas entre 3 o más grupos (donde los participantes son los mismos en cada grupo) que hayan sido tratados en varias ocasiones o bajo diferentes condiciones. Como diseño de la investigación, por ejemplo, los mismos participantes podrían ser tratados tomando una medida de resultado en 1, 2 y 3 semanas o que la medida de resultado fuera tomada bajo las condiciones 1, 2 y 3.

La hipótesis nula que se pone a prueba es que no hay diferencia significativa entre las medias de las diferencias entre todos los grupos.

La variable independiente debería ser categórica y la variable dependiente tiene que ser una medida continua. En este análisis las categorías de la variable independiente son **niveles** designados, es decir, son los grupos relacionados. Por lo tanto, en el caso en el que una variable resultado fuese medida en 1, 2 y 3 semanas, los 3 niveles serían semana 1, semana 2 y semana 3.

El **estadístico F** se calcula dividiendo los cuadrados medios de la variable (varianza explicada por el modelo) por los cuadrados medios de su error (varianza no explicada). Cuanto mayor sea el estadístico *F*, más probable será que la variable independiente haya tenido un efecto significativo sobre la variable dependiente.

SUPUESTOS

El ANOVA MR tiene los mismos supuestos que la mayoría de los test paramétricos.

- La variable dependiente debería tener una distribución aproximadamente normal.
- No debería haber valores atípicos significativos.
- Esfericidad, que tiene que ver con la igualdad de las varianzas de las diferencias entre los niveles del factor de medidas repetidas.

Si los supuestos han sido violados, entonces debería considerarse el **test de Friedman**, su equivalente no paramétrico descrito más adelante en esta sección.

ESFERICIDAD

Si un estudio tiene 3 niveles (A, B y C), la esfericidad asume lo siguiente:

$$\text{Varianza (A-B)} \approx \text{Varianza (A-C)} \approx \text{Varianza (B-C)}$$

El ANOVA MR comprueba el supuesto de esfericidad usando el test de esfericidad de Mauchly (pronunciado como “Mockley”). Este test pone a prueba **la hipótesis nula de que las varianzas de las diferencias son iguales**. En muchos casos, las medidas repetidas violan el supuesto de esfericidad, lo que puede conducir a un error de Tipo I. Si este es el caso, se pueden aplicar correcciones al estadístico *F*.

JASP ofrece dos métodos de corrección del estadístico *F*: las correcciones épsilon (ε) de **Greenhouse-Geisser** y de **Huynh-Feldt**. Como recomendación general, si los valores ε son < 0,75 se debe usar la corrección de Greenhouse-Geisser y, si son > 0,75, la corrección de Huynh-Feldt.

PRUEBAS POST HOC

Aunque el análisis post hoc es limitado en el caso del ANOVA MR, JASP proporciona dos alternativas:

a) ***Bonferroni*** – puede ser muy conservador, pero garantiza el control del error de Tipo I a riesgo de reducir la potencia estadística.
b) ***Holm*** – el test Holm-Bonferroni es un método secuencial Bonferroni menos conservador que el test Bonferroni original.

Si se solicitan las correcciones post hoc de Tukey o de Scheffe, JASP reportará un error NaN (por el inglés *not a number*, o "no es un número").

TAMAÑO DEL EFECTO

JASP proporciona las mismas alternativas al cálculo del tamaño del efecto que las usadas en la prueba de ANOVA de medidas independientes:

a) ***Eta cuadrado (η^2)*** – preciso para estimar la varianza explicada en la muestra pero sobreestima la varianza en la población. Dificulta la comparación del efecto de la misma variable en distintos estudios.
b) ***Eta cuadrado parcial (η_p^2)*** – resuelve el problema de la sobreestimación de la varianza en la población permitiendo la comparación del efecto de la misma variable en distintos estudios. Esta es la forma más habitual de reportar el tamaño del efecto en el ANOVA de medidas repetidas.
c) ***Omega cuadrado (ω^2)*** – normalmente, el sesgo estadístico deviene muy pequeño a medida que se incrementa el tamaño de muestra, pero para cuando tenemos muestras pequeñas ($n < 30$) ω^2 proporciona una medida no sesgada del tamaño del efecto.

Niveles de tamaño del efecto:

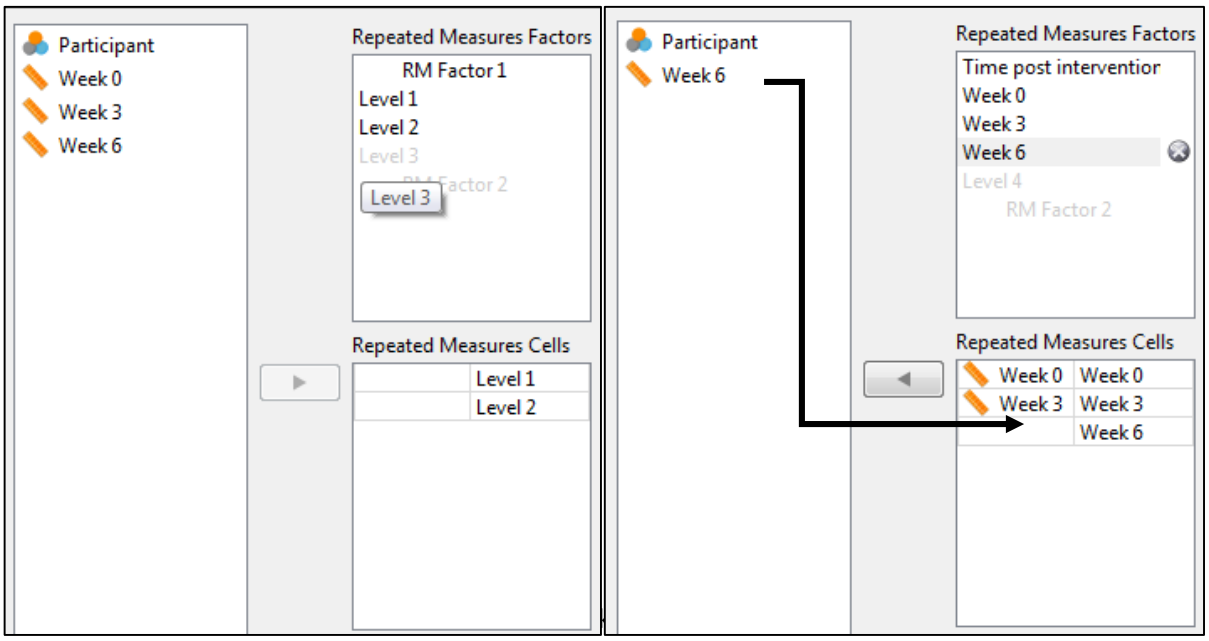
Test	Medida	Irrelevante	Pequeño	Medio	Grande
ANOVA	**Eta**	**< 0,1**	**0,1**	**0,25**	**0,37**
	Eta parcial	**< 0,01**	**0,01**	**0,06**	**0,14**
	Omega cuadrado	**< 0,01**	**0,01**	**0,06**	**0,14**

EJECUTANDO EL ANOVA DE MEDIDAS REPETIDAS

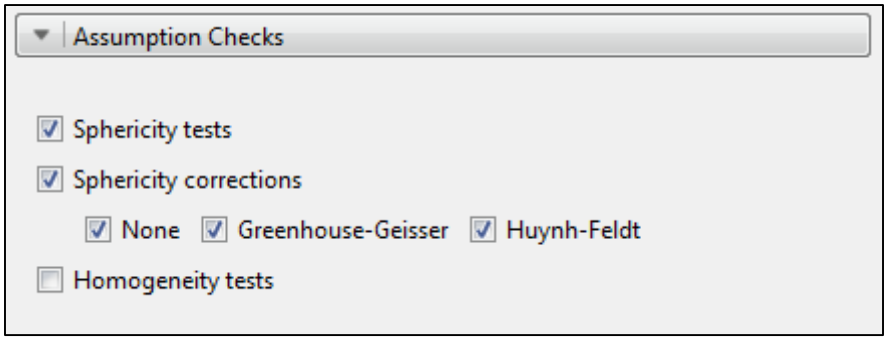
Cargue **Repeated ANOVA cholesterol.csv**. Este archivo contiene una columna con las ID de los participantes y otras 3 columnas más, una para cada medida repetida del colesterol en sangre tras una intervención. Es una buena práctica revisar la estadística descriptiva y los gráficos de caja por si hubiera algún valor atípico extremo.

Vaya a «ANOVA» ⟶ «Repeated measures ANOVA». Como se ha comentado antes, la variable independiente (el factor de medidas repetidas) tiene niveles, en este caso 3. Cambie el nombre de la variable RM Factor 1 a Time post intervention y haga lo mismo con los 3 niveles poniendo Week 0, Week 3 y Week 6, respectivamente.

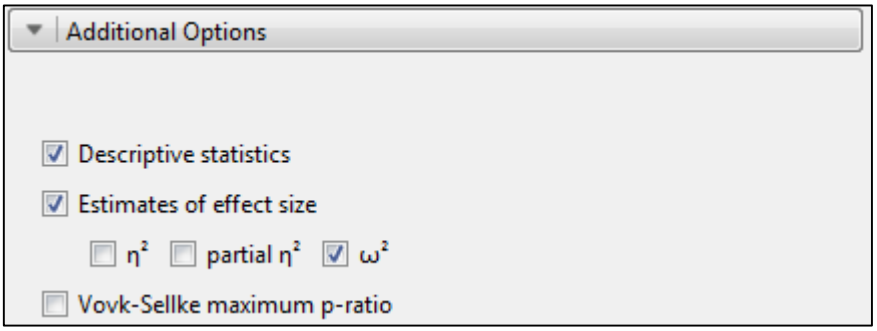
Una vez hecho, aparecerán en la caja «Repeated Measures Cells». Ahora, añada los datos apropiados al nivel apropiado.



En «Assumption Checks», seleccione «Sphericity tests» y todas las opciones incluidas en «Sphericity corrections».



En «Additional Options», marque «Descriptive statistics», «Estimates of effect size» y «ω^2».



El resultado debería incluir 4 tablas. La tercera tabla, la que corresponde a los efectos inter-sujetos, puede ignorarse en este análisis.

ENTENDIENDO EL RESULTADO

Within Subjects Effects

	Sphericity Correction	Sum of Squares	df	Mean Square	F	p	ω²
Time post intervention	None	4.320[a]	2.000[a]	2.160[a]	212.321[a]	< .001[a]	0.058
	Greenhouse-Geisser	4.320[a]	1.235[a]	3.497[a]	212.321[a]	< .001[a]	0.058
	Huynh-Feldt	4.320[a]	1.284[a]	3.365[a]	212.321[a]	< .001[a]	0.058
Residual	None	0.346	34.000	0.010			
	Greenhouse-Geisser	0.346	21.001	0.016			
	Huynh-Feldt	0.346	21.822	0.016			

Note. Type III Sum of Squares
[a] Mauchly's test of sphericity indicates that the assumption of sphericity is violated (p < .05).

La tabla de efectos intra-sujetos ("Within Subjects Effects") muestra un estadístico F grande, altamente significativo (p < 0,001) y con un tamaño del efecto entre pequeño y medio (0,058). Esta tabla presenta los estadísticos que asumen la esfericidad ("None") y los dos métodos de corrección. Las principales diferencias están en los grados de libertad ("df", por el inglés *degrees of freedom*) y el valor de los cuadrados medios. Bajo la tabla se indica que el supuesto de esfericidad ha sido violado.

La tabla siguiente ofrece los resultados del test de esfericidad de Mauchly. Como puede verse, hay una diferencia significativa (p < 0,001) en las varianzas de las diferencias entre los grupos. Los valores épsilon (ε) de Greenhouse-Geisser y de Huynh-Feldt están por debajo de 0,75. Por lo tanto, el resultado del ANOVA debería basarse en la corrección de Greenhouse-Geisser:

Test of Sphericity

	Mauchly's W	p	Greenhouse-Geisser ε	Huynh-Feldt ε
Time post intervention	0.381	< .001	0.618	0.642

Para obtener una tabla más clara, vuelva a «Assumption Checks» y seleccione únicamente «Greenhouse-Geisser» como corrección de la esfericidad.

Within Subjects Effects

	Sphericity Correction	Sum of Squares	df	Mean Square	F	p	ω²
Time post intervention	Greenhouse-Geisser	4.320[a]	1.235[a]	3.497[a]	212.321[a]	< .001[a]	0.058
Residual	Greenhouse-Geisser	0.346	21.001	0.016			

Note. Type III Sum of Squares
[a] Mauchly's test of sphericity indicates that the assumption of sphericity is violated (p < .05).

Hay una diferencia significativa entre las medias de las diferencias entre todos los grupos: **F (1,235, 21,0) = 212,3, p < 0,001, ω^2 = 0,058.**

Descriptives

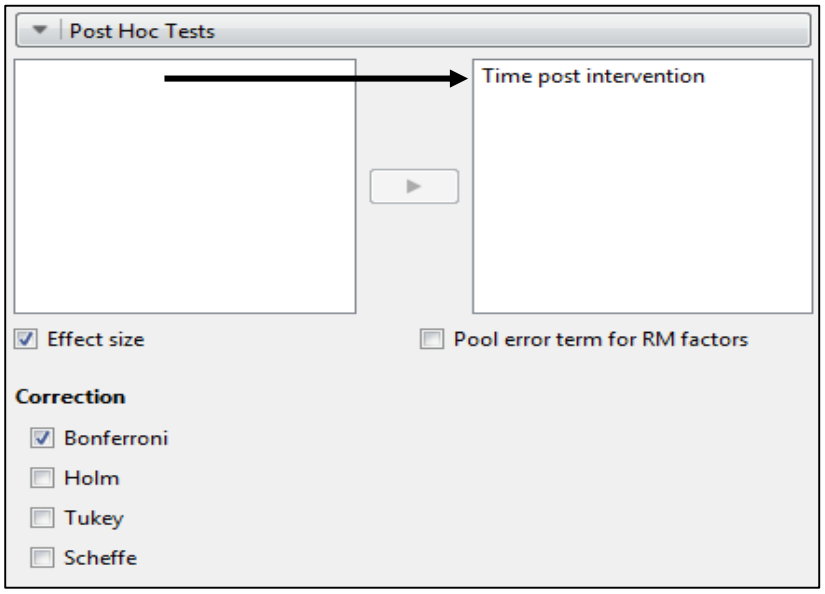
Time post intervention	Mean	SD	N
Week 0	6.408	1.191	18.000
Week 3	5.842	1.123	18.000
Week 6	5.779	1.102	18.000

El análisis descriptivo sugiere que los niveles de colesterol en sangre fueron más altos en la semana 0, comparados con los de las semanas 3 y 6.

Sin embargo, si el ANOVA no reporta diferencias significativas, no se puede proseguir con el análisis.

PRUEBAS POST HOC

Si el ANOVA es significativo, se puede llevar a cabo el análisis post hoc. En «Post Hoc Tests» añada Time post intervention a la caja de análisis de la derecha, seleccione «Effect size» y, en este caso, use Bonferroni para la corrección post hoc.



Añada también, en «Descriptive plots», el factor Time post intervention a la caja «Horizontal axis» y marque «Display error bars».

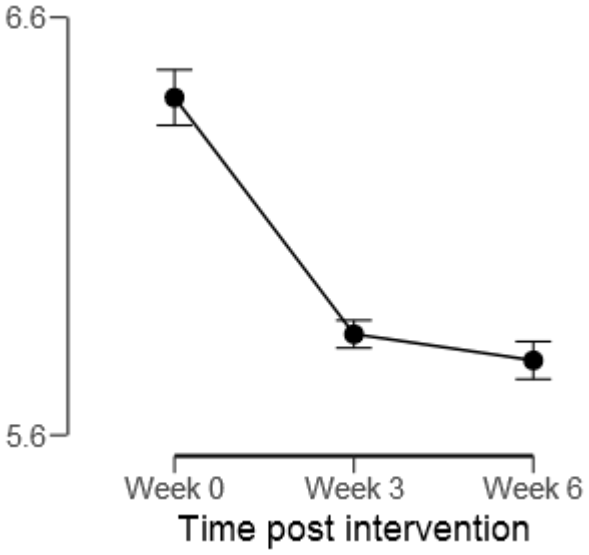
Post Hoc Comparisons - Time post intervention

		Mean Difference	SE	t	Cohen's d	p_{bonf}
Week 0	Week 3	0.566	0.037	15.439	3.639	< .001
	Week 6	0.629	0.042	14.946	3.523	< .001
Week 3	Week 6	0.063	0.017	3.781	0.891	0.004

Note. Cohen's d does not correct for multiple comparisons.

El análisis post hoc muestra que hay diferencias significativas en los niveles de colesterol en sangre entre todas las combinaciones de valores de tiempo y que están asociadas con tamaños del efecto grandes.

REPORTANDO LOS RESULTADOS



Se usó la corrección de Greenhouse-Geisser, dado que el test de esfericidad de Mauchly fue significativo. El análisis mostró que los niveles de colesterol difirieron significativamente: F (1,235, 21,0) = 212,3, p < 0,001, ω^2 = 0,058.

El análisis post hoc usando la corrección de Bonferroni reveló que los niveles de colesterol disminuyeron significativamente a medida que pasó el tiempo entre las semanas 0 y 3 (diferencia de las medias = 0,566 unidades, p < 0,001) y entre las semanas 3 y 6 (diferencia de las medias = 0,063 unidades, p = 0,004).

ANOVA DE MEDIDAS REPETIDAS DE FRIEDMAN

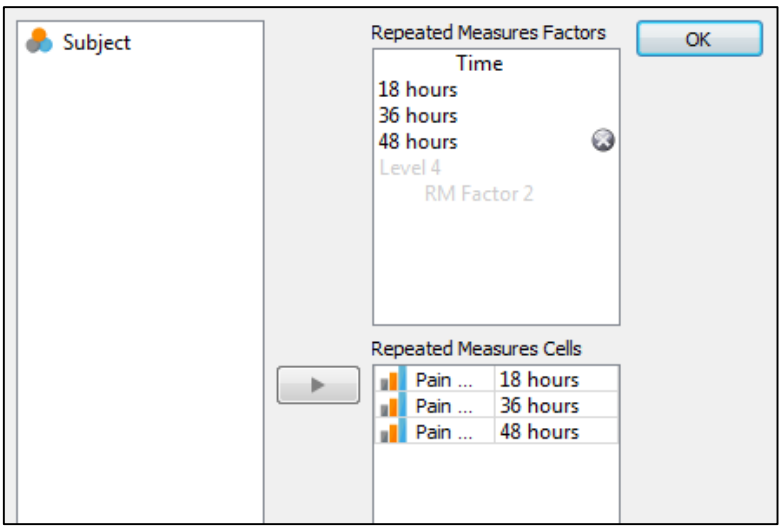
Si se violan los supuestos paramétricos o los datos son ordinales, debería considerarse el uso de la alternativa no paramétrica, el test de Friedman. Como el test de Kruskal-Wallis, la prueba de Friedman se usa para el análisis de la varianza de medidas repetidas de un solo factor por rangos, y no supone que los datos provengan de una distribución en particular. Se trata de otra "prueba ómnibus" (global) que no especifica qué grupos de la variable independiente son significativamente diferentes entre sí. Si el test de Friedman es significativo, JASP proporciona la opción de ejecutar la prueba post hoc de Conover.

Cargue **Friedman RMANOVA.csv** en JASP. Este archivo contiene 3 columnas con las puntuaciones de dolor subjetivo registradas a las 18, 36 y 48 horas tras haber realizado ejercicio. Compruebe que las puntuaciones de dolor están asignadas como variables ordinales.

EJECUTANDO EL TEST DE FRIEDMAN

Vaya a «ANOVA» ⟶ «Repeated Measures ANOVA». La variable independiente (factor de medidas repetidas) tiene 3 niveles. Cambie el nombre de la variable RM Factor 1 a Time y haga lo mismo con los 3 niveles poniendo 18 hours, 36 hours y 48 hours, respectivamente.

Tras hacer esto, aparecerán en la caja «Repeated Measures Cells». Añada ahora los datos apropiados al nivel que corresponda.



De esta forma, obtendrá la tabla de ANOVA estándar de medidas repetidas intra-sujetos. Para ejecutar el test de Friedman, expanda la pestaña «Nonparametrics», mueva Time a la caja «RM Factor» y marque «Connover's post hoc tests».

ENTENDIENDO EL RESULTADO

Deberían haberse obtenido dos tablas.

Friedman Test

Factor	Chi-Squared	df	p	Kendall's W
Time	26.772	2	< .001	0.764

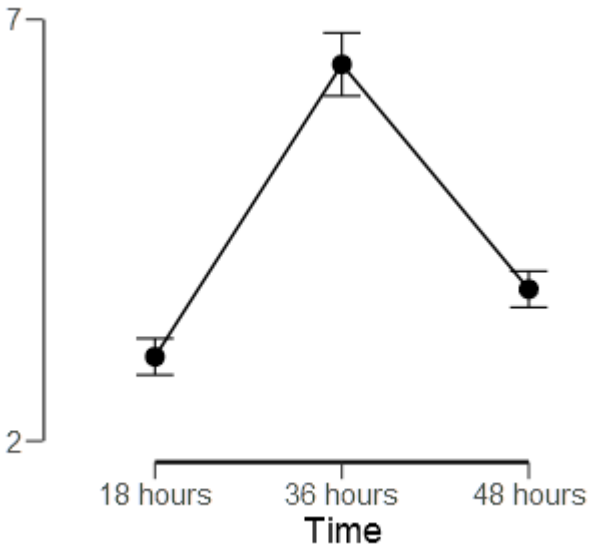
Connover's Post Hoc Comparisons - Time

		T-Stat	df	W_i	W_j	p	p_{bonf}	p_{holm}
18 hours	36 hours	15.171	28	17.000	44.500	< .001	< .001	< .001
	48 hours	6.344	28	17.000	28.500	< .001	< .001	< .001
36 hours	48 hours	8.827	28	44.500	28.500	< .001	< .001	< .001

La prueba de Friedman muestra que el tiempo tiene un efecto significativo sobre la percepción del dolor. Las comparaciones post hoc dos a dos de Connor muestran que las percepciones de dolor son significativamente diferentes en cada momento.

REPORTANDO LOS RESULTADOS

El tiempo tiene un efecto significativo en las puntuaciones de dolor subjetivo χ^2 (2) = 26,77, p < 0,001. Las comparaciones dos a dos muestran que la percepción del dolor es significativamente diferente en cada momento (todos los valores de p < 0,001).



ANOVA DE MEDIDAS INDEPENDIENTES DE DOS FACTORES

Si el ANOVA de un solo factor evalúa situaciones en las que solo se manipula una variable independiente, el ANOVA de dos factores se usa cuando se ha manipulado más de 1 variable independiente. En este caso, las variables independientes se conocen como factores.

Factor 1	Factor 2	
Condición 1	Grupo 1	Variable dependiente
	Grupo 2	Variable dependiente
Condición 2	Grupo 1	Variable dependiente
	Grupo 2	Variable dependiente
Condición 3	Grupo 1	Variable dependiente
	Grupo 2	Variable dependiente

Los factores están divididos en niveles, de modo que, en este caso, el factor 1 tiene 3 niveles y hay 2 niveles para el factor 2.

El "efecto principal" (*Main effect*) es el efecto de una de las variables independientes sobre la variable dependiente, ignorando los efectos de cualquier otra variable independiente. Hay 2 efectos principales que se ponen a prueba, ambos "inter-sujetos" (*Between-subjects*): en este caso, comparando las diferencias en el factor 1 (es decir, la condición), y las diferencias en el factor 2 (los grupos). Se produce una **interacción** cuando un factor influye en el otro.

El ANOVA de dos factores de medidas independientes es otra "prueba ómnibus" (global) que se usa para probar 2 hipótesis nulas:

1. **No hay ningún efecto significativo inter-sujetos, es decir, no existen diferencias significativas entre las medias de los grupos en cualquiera de los factores.**
2. **No hay ningún efecto de interacción significativo, es decir, no existen diferencias de grupo significativas entre las condiciones.**

SUPUESTOS

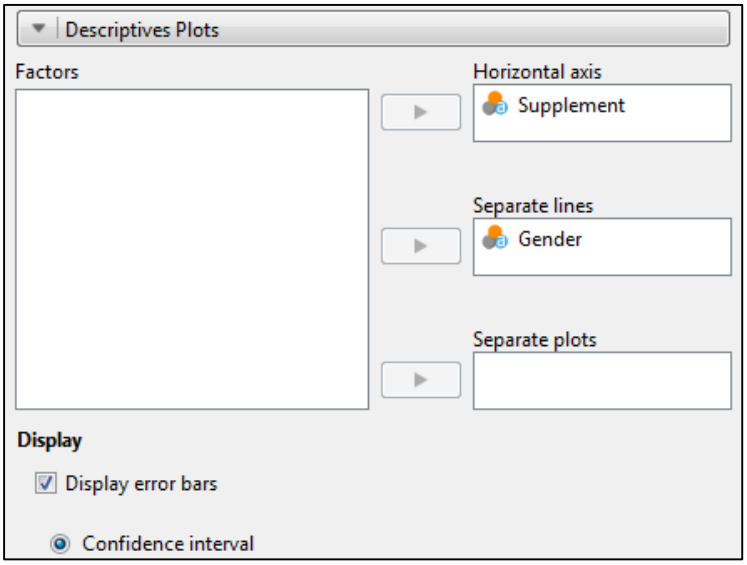
Como las demás pruebas paramétricas, el ANOVA de dos factores de medidas independientes realiza una serie de asunciones que deberían abordarse en el diseño de la investigación o que deberían ser comprobadas.

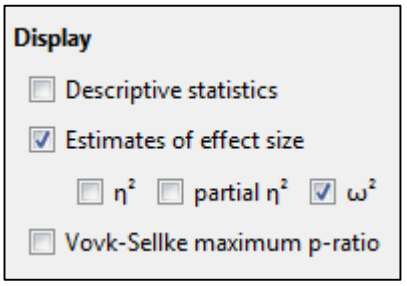
- Las variables independientes (factores) deberían tener al menos dos grupos independientes categóricos (niveles).
- La variable dependiente debería ser continua y mostrar una distribución aproximadamente normal a lo largo de todas las combinaciones de los factores.
- Debería haber homogeneidad de la varianza en cada una de las combinaciones de los factores.
- No debería haber valores atípicos significativos.

EJECUTANDO EL ANOVA DE DOS FACTORES DE MEDIDAS INDEPENDIENTES

Abra **2-way independent ANOVA.csv** en JASP. Este archivo incluye 3 columnas de datos: Factor 1 – Gender, con 2 niveles (hombre y mujer); Factor 2 – Supplement, con 3 niveles (control, carbohidratos CHO y proteínas) y la variable dependiente (Jump power). En «Descriptive statistics», compruebe los datos por si hubiera valores atípicos significativos. Vaya a «ANOVA» ⟶ «ANOVA», añada Jump power a la caja «Dependent Variable», y Gender y Supplement a la caja «Fixed Factors».

En «Descriptives Plots», añada Supplement a «Horizontal axis» y Gender a «Separate lines». En «Additional Options», marque «Descriptive statistics», «Estimates of effect size» y «ω^2».





ENTENDIENDO EL RESULTADO

El resultado debería contener 2 tablas y un gráfico.

ANOVA - Jump power

Cases	Sum of Squares	df	Mean Square	F	p	ω²
Gender	119108.037	1.000	119108.037	9.589	0.003	0.058
Supplement	896116.137	2.000	448058.068	36.071	< .001	0.477
Gender ✻ Supplement	275806.438	2.000	137903.219	11.102	< .001	0.138
Residual	521712.054	42.000	12421.716			

Note. Type III Sum of Squares

La tabla de ANOVA muestra que hay efectos principales significativos para Gender y Supplement ($p = 0{,}003$ y $p < 0{,}001$, respectivamente), con tamaños del efecto medio y grande, respectivamente. Esto sugiere que hay una diferencia significativa entre la potencia de salto de cada género, con independencia del suplemento, y diferencias significativas entre suplementos, independientemente del género.

También hay una interacción significativa entre Gender y Supplement ($p < 0{,}001$), que también tiene un tamaño del efecto entre medio y grande (0,138). Esto sugiere que las diferencias en la potencia de salto entre los géneros están afectadas de algún modo por el tipo de suplemento utilizado.

La estadística descriptiva y el gráfico sugieren que las diferencias principales se dan entre géneros cuando se usa un suplemento proteínico.

Descriptives - Jump power ▼

Gender	Supplement	Mean	SD	N
Female	Control	877.500	134.563	8.000
	CHO	789.286	102.283	7.000
	Protein	986.667	91.924	9.000
Male	Control	788.125	64.417	8.000
	CHO	901.875	117.502	8.000
	Protein	1263.125	140.863	8.000

COMPROBACIÓN DE SUPUESTOS

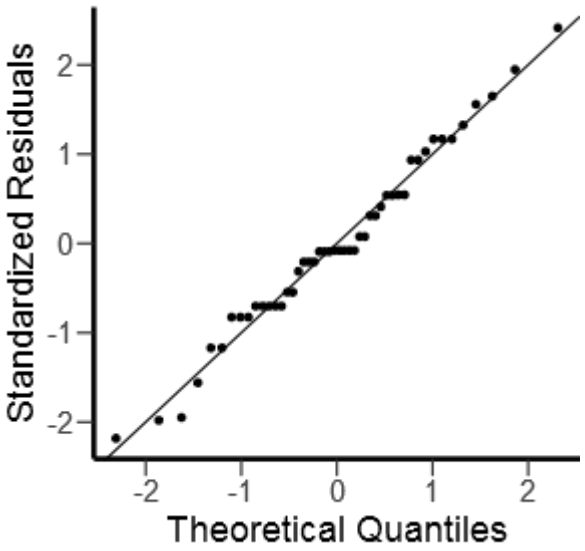
En «Assumption Checks», marque «Homogeneity tests» y «Q-Q plot of residuals».

Assumption Checks

Test for Equality of Variances (Levene's)

F	df1	df2	p
1.100	5.000	42.000	0.375

El test de Levene no muestra diferencias significativas de la varianza, por lo que la homogeneidad de la varianza no ha sido violada.



El gráfico Q-Q muestra que los datos parecen estar distribuidos normalmente y ser lineales. Se puede, por tanto, aceptar el resultado del ANOVA ya que ninguno de estos supuestos ha sido violado.

No obstante, si el ANOVA no muestra ninguna diferencia significativa, no se puede proseguir con el análisis.

PRUEBAS POST HOC

Si el ANOVA es significativo, se puede realizar un análisis post hoc. En «Post Hoc Tests» añada Supplement en la caja de análisis de la derecha, marque «Effect size» y, en este caso, marque «Tukey» para la corrección post hoc.

El test post hoc no se realiza para Gender porque solo hay 2 niveles.

Post Hoc Comparisons - Supplement

		Mean Difference	SE	t	Cohen's d	p_{tukey}
Control	CHO	−12.768	40.102	−0.318	−0.109	0.946
	Protein	−292.083	38.853	−7.518	−1.919	< .001
CHO	Protein	−279.315	39.561	−7.060	−1.782	< .001

Note. Cohen's d does not correct for multiple comparisons.

El test post hoc no muestra diferencias significativas entre el grupo de control y el de suplemento CHO, independientemente del género, pero sí muestra diferencias significativas entre el grupo de control y el de proteínas ($p < 0{,}001$) y entre el de CHO y el de proteínas ($p < 0{,}001$).

Vaya ahora a las opciones «Simple Main Effects». Aquí, añada Gender a la caja «Simple effect factor» y Supplement a la caja «Moderator factor 1». Efectivamente, los efectos principales simples son comparaciones dos a dos.

Simple Main Effects - Gender

Level of Supplement	Sum of Squares	df	Mean Square	F	p
Control	31951.563	1	31951.563	2.572	0.116
CHO	47325.030	1	47325.030	3.810	0.058
Protein	323700.184	1	323700.184	26.059	< .001

Esta tabla muestra que no hay diferencias de género en la potencia de salto entre los grupos control y CHO (p = 0,116 y p = 0,058, respectivamente). No obstante, hay una diferencia significativa (p < 0,001) de potencia de salto entre géneros en el grupo de suplemento proteínico.

REPORTANDO LOS RESULTADOS

Se usó un ANOVA de dos factores para examinar el efecto del género y el tipo de suplemento sobre la potencia de salto. Se hallaron efectos principales para los dos géneros (F (1, 42) = 9,59, p = 0,003, ω^2 = 0,058) y el suplemento (F (2, 42) = 30,07, p < 0,001, ω^2 = 0,477). Hubo una interacción estadísticamente significativa entre los efectos del género y el suplemento en la potencia de salto (F (2, 42) = 11,1, p < 0,001, ω^2 = 0,138).

La corrección post hoc de Tukey mostró que la potencia de salto fue significativamente superior en el grupo de proteínas comparado con los grupos control y CHO (t = -1,919, p < 0,001 y t = -1,782, p < 0,001, respectivamente).

Los efectos principales simples mostraron que la potencia de salto fue significativamente mayor entre los hombres que entre las mujeres en el grupo de los que usaron un suplemento de proteínas (F (1) = 28,06, p < 0,001).

ANOVA MIXTO CON JASP

El ANOVA mixto (otro ANOVA de dos factores) es una combinación del ANOVA de medidas repetidas y el de medidas independientes, en el que se hallan involucradas más de 1 variable independiente (conocidas como factores).

Variable independiente (factor 2)	***Variable independiente (factor 1) = tiempo o condición***		
	Tiempo/condición 1	**Tiempo/condición 2**	**Tiempo/condición 3**
Grupo 1	Variable dependiente	Variable dependiente	Variable dependiente
Grupo 2	Variable dependiente	Variable dependiente	Variable dependiente

Los factores están divididos en niveles, en este caso, el factor 1 tiene 3 niveles y el factor 2 posee 2 niveles. Esto resulta en 6 combinaciones posibles.

Un "efecto principal" es el efecto de una de las variables independientes sobre la variable dependiente, ignorando los efectos de cualquier otra variable independiente. Se ponen a prueba 2 efectos principales: en este caso, la comparación de los datos a lo largo del factor 1 (es decir, el tiempo) se conoce como factor "**intra-sujetos**", mientras que la comparación de las diferencias en el factor 2 (es decir, los grupos) se denomina factor "**inter-sujetos**". Se da una interacción cuando un factor influye sobre el otro factor.

El efecto principal del tiempo o la condición (factor intra-sujetos) pone a prueba, con independencia del grupo a que pertenezcan los datos:

Variable independiente (factor 2)	***Variable independiente (factor 1) = tiempo o condición***		
	Tiempo/condición 1	**Tiempo/condición 2**	**Tiempo/condición 3**
Grupo 1	Todos los datos	Todos los datos	Todos los datos
Grupo 2			

El efecto principal del grupo (factor inter-sujetos) pone a prueba que, con independencia de la condición a que pertenezcan los datos:

Variable independiente (factor 2)	***Variable independiente (factor 1) = tiempo o condición***		
	Tiempo/condición 1	**Tiempo/condición 2**	**Tiempo/condición 3**
Grupo 1	Todos los datos		
Grupo 2	Todos los datos		

Los efectos principales simples son, efectivamente, comparaciones dos a dos:

Variable independiente (factor 2)	***Variable independiente (factor 1) = tiempo o condición***		
	Tiempo/condición 1	**Tiempo/condición 2**	**Tiempo/condición 3**
Grupo 1	Datos	Datos	Datos
Grupo 2	Datos	Datos	Datos

Un ANOVA mixto es otra "prueba ómnibus" (global) usada para poner a prueba 3 hipótesis nulas:

1. **No hay efecto significativo intra-sujetos, es decir, no hay diferencias significativas entre las medias de las diferencias entre todas las condiciones / los tiempos**.
2. **No hay efecto significativo inter-sujetos, es decir, no hay diferencias significativas entre las medias de los grupos**.
3. **No hay efecto de interacción significativo, es decir, no hay diferencias significativas de los grupos a través de las condiciones / el tiempo**.

SUPUESTOS

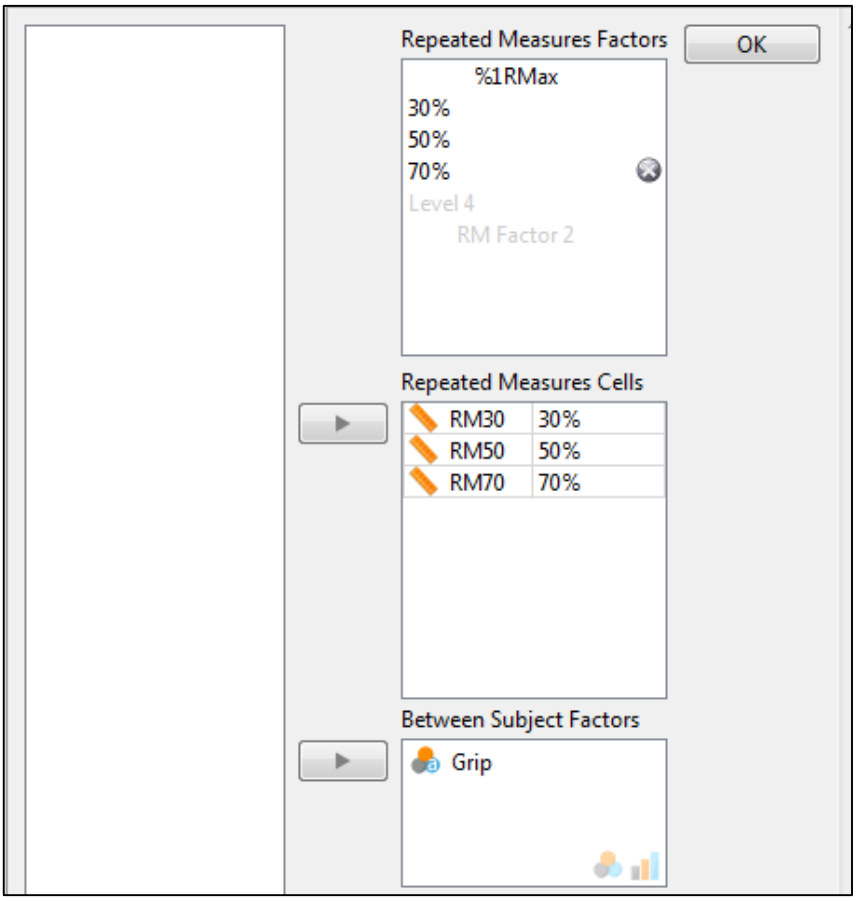
Como las demás pruebas paramétricas, el ANOVA mixto realiza una serie de supuestos que deberían tenerse en cuenta en el diseño de la investigación o que podrían probarse.

- El factor "**intra-sujetos**" debería contener al menos dos grupos categóricos (niveles) relacionados (medidas repetidas).
- El factor "**inter-sujetos**" debería contener al menos dos grupos categóricos (niveles) no relacionados (medidas independientes).
- La variable independiente debería ser continua y tener una distribución aproximadamente normal para todas las combinaciones de factores.
- Debería haber homogeneidad de varianza para cada uno de los grupos y, si hubiera más de 2 niveles, esfericidad entre los grupos relacionados.
- No debería haber valores atípicos significativos.

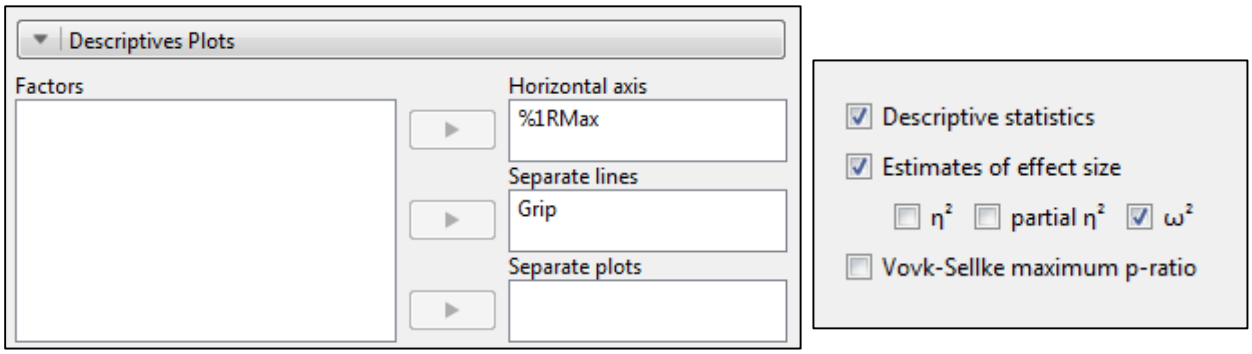
EJECUTANDO EL ANOVA MIXTO

Abra **2-way Mixed ANOVA.csv** en JASP. Este archivo contiene 4 columnas de datos relativos a las empuñaduras en el levantamiento de pesas y a la velocidad del levantamiento con 3 cargas de peso distintas (%1RM). La columna 1 contiene el tipo de agarre, las columnas 2-4 contienen las 3 medidas repetidas (30%, 50% y 70%). Compruebe si existen valores atípicos significativos mediante los gráficos de caja y vaya a «ANOVA» ⟶ «Repeated measures ANOVA».

Defina el factor de medidas repetidas introduciendo %1RM en la caja «Repeated Measures Factor» y añada 3 niveles (30%, 50% y 70%). Añada la variable apropiada en la caja «Repeated Measures Cells» y añada Grip a la caja «Between-Subjects Factors»:



En «Descriptive Plots», mueva %1RM al eje horizontal y Grip a líneas separadas. En «Additional Options», marque «Descriptive statistics», «Estimates of effect size» y «ω^2».



ENTENDIENDO EL RESULTADO

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	ω²
%1RM	5.605[a]	2[a]	2.803[a]	115.450[a]	< .001[a]	0.744
%1RM ✻ Grip	0.583[a]	2[a]	0.291[a]	12.003[a]	< .001[a]	0.218
Residual	0.874	36	0.024			

Note. Type III Sum of Squares

[a] Mauchly's test of sphericity indicates that the assumption of sphericity is violated (p < .05).

El resultado debería incluir 3 tablas y un gráfico.

Para el efecto principal respecto a %1RM, la tabla de efectos intra-sujetos ("Within Subjects Effects") reporta un estadístico F grande, que es altamente significativo (p < 0,001) y además muestra un tamaño del efecto grande (0,744). Así, independientemente del tipo de agarre, hay una diferencia significativa entre las tres cargas.

Finalmente, existe una interacción significativa entre %1RM y Grip (p < 0,001), que también muestra un tamaño del efecto grande (0,218). Esto sugiere que las diferencias entre las cargas están de algún modo afectadas por el tipo de agarre empleado.

No obstante, JASP informa, bajo la tabla, de que el supuesto de esfericidad ha sido violado. Trataremos el asunto en la siguiente sección.

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p	ω²
Grip	1.095	1	1.095	20.925	< .001	0.499
Residual	0.942	18	0.052			

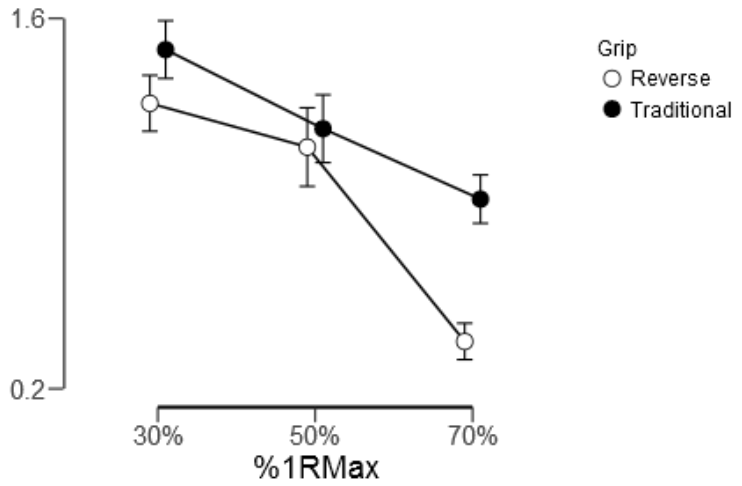
Note. Type III Sum of Squares

Para el efecto principal respecto a Grip, la tabla de efectos inter-sujetos ("Between Subjects Effects") muestra una diferencia significativa entre los diferentes tipos de agarres (p < 0,001), con independencia de las cargas.

A partir de la estadística descriptiva y del gráfico, parece que hay una diferencia mayor entre los dos tipos de agarre con la carga de peso más alta del 70%.

Descriptives

%1RMax	Grip	Mean	SD	N
30%	Reverse	1.279	0.178	10.000
	Traditional	1.482	0.217	10.000
50%	Reverse	1.114	0.198	10.000
	Traditional	1.183	0.256	10.000
70%	Reverse	0.379	0.105	10.000
	Traditional	0.917	0.086	10.000



COMPROBACIÓN DE SUPUESTOS

En «Assumptions Checks», marque «Sphericity tests», «Sphericity corrections» y «Homogeneity tests».

Test of Sphericity

	Mauchly's W	p	Greenhouse-Geisser ε	Huynh-Feldt ε
%1RMax	0.649	0.025	0.740	0.791

La prueba de esfericidad de Mauchly es significativa, por lo que el supuesto ha sido violado. Por tanto, debería usarse la corrección de Greenhouse-Geisser ya que épsilon es < 0,75. Vuelva a «Assumption Checks» y, en «Sphericity corrections», deje marcado únicamente «Greenhouse-Geisser». Esto dará como resultado una tabla actualizada de efectos intra-sujetos ("Within Subjects Effects"):

Within Subjects Effects ▼

	Sphericity Correction	Sum of Squares	df	Mean Square	F	p	ω²
%1RM	Greenhouse-Geisser	5.605[a]	1.480[a]	3.787[a]	115.450[a]	< .001[a]	0.744
%1RM ✻ Grip	Greenhouse-Geisser	0.583[a]	1.480[a]	0.394[a]	12.003[a]	< .001[a]	0.218
Residual	Greenhouse-Geisser	0.874	26.639	0.033			

Note. Type III Sum of Squares

[a] Mauchly's test of sphericity indicates that the assumption of sphericity is violated (p < .05).

Test for Equality of Variances (Levene's)

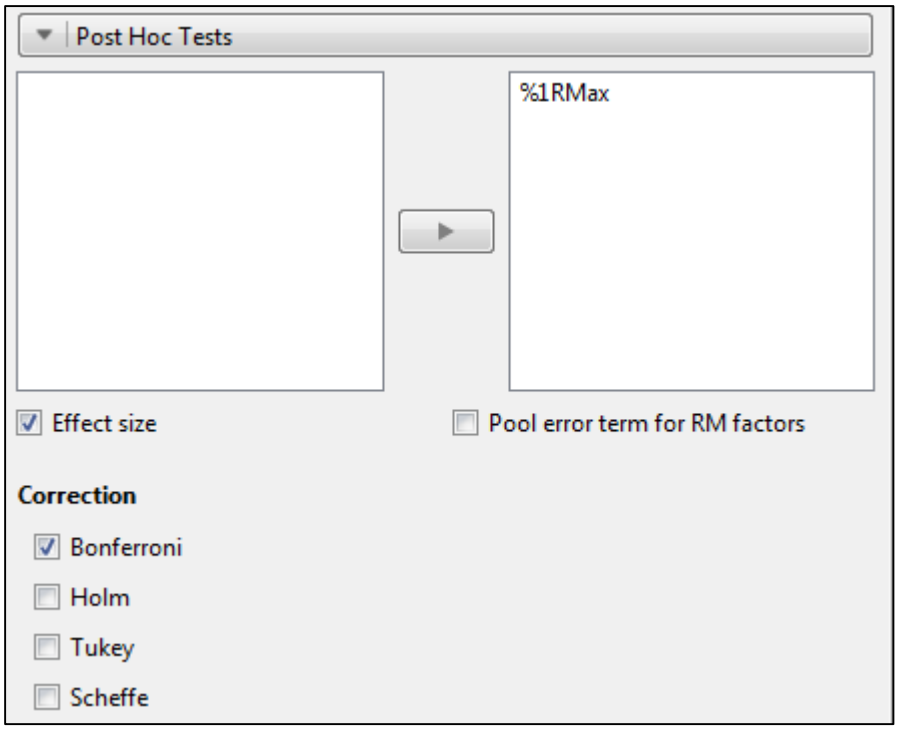
	F	df1	df2	p
RM30	0.523	1.000	18.000	0.479
RM50	0.346	1.000	18.000	0.564
RM70	0.183	1.000	18.000	0.674

La prueba de Levene muestra que no hay diferencia significativa en la varianza de la variable dependiente a través de los dos tipos de agarre.

Sin embargo, si el ANOVA no reporta diferencias significativas, no puede ir más allá con el análisis.

PRUEBAS POST HOC

Si el ANOVA es significativo, puede llevarse a cabo el análisis post hoc. En «Post Hoc Tests», añada %1RM a la caja de análisis de la derecha, marque «Effect size» y, en este caso, utilice Bonferroni para la corrección post hoc. En el análisis de medidas repetidas solo están disponibles las correcciones de Bonferroni y de Holm.

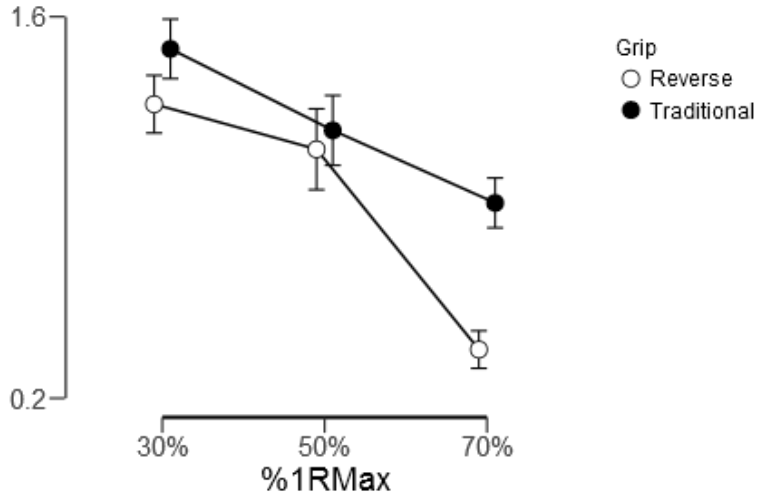


Post Hoc Comparisons - %1RMax

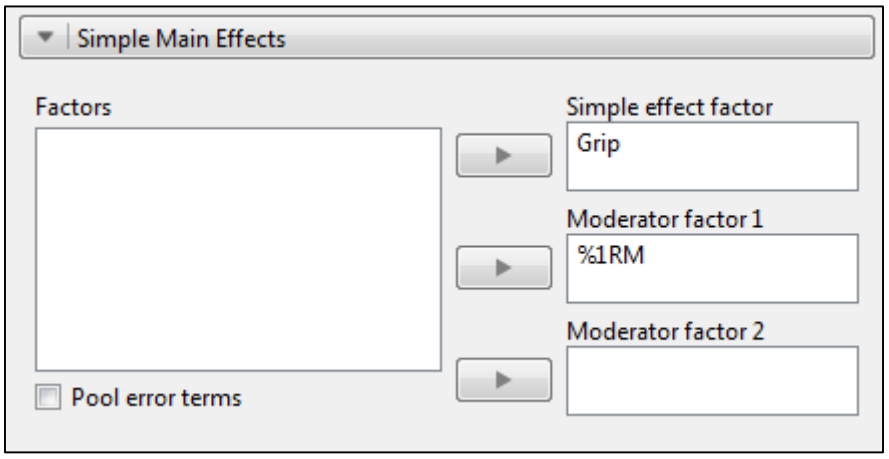
		Mean Difference	SE	t	Cohen's d	p_{bonf}
30%	50%	0.232	0.060	3.856	0.862	0.003
	70%	0.733	0.050	14.583	3.261	< .001
50%	70%	0.500	0.073	6.839	1.529	< .001

Note. Cohen's d does not correct for multiple comparisons.

El análisis post hoc muestra que, con independencia del tipo de agarre utilizado, cada carga de peso es significativamente diferente del resto y, como se ve en el gráfico, la velocidad de levantamiento decrece a medida que aumenta el peso.



Finalmente, en «Simple Main Effects», añada Grip a la caja «Simple effect factor» y %1RM a la caja «Moderator factor 1».



Simple Main Effects - Grip

Level of %1RM	Sum of Squares	df	Mean Square	F	p
30%	0.206	1	0.206	5.229	0.035
50%	0.024	1	0.024	0.461	0.506
70%	1.447	1	1.447	157.212	< .001

Los resultados muestran que hay una diferencia significativa en la velocidad de levantamiento a lo largo de los dos tipos de agarre en la carga inferior del 30%, así como en la carga mayor del 70% ($p = 0,035$ y $p < 0,001$, respectivamente).

REPORTANDO LOS RESULTADOS

Usando la corrección de Greenhouse-Geisser, hubo un efecto principal significativo de la carga (F= (1,48, 26,64) = 115,45, p < 0,001). El análisis post hoc con la corrección de Bonferroni mostró una disminución secuencial significativa en la velocidad de levantamiento entre el 30% y el 50% de carga (p = 0,035) y entre el 50% y el 70% de carga (p < 0,001).

Hubo un efecto principal significativo para el tipo de agarre (F (1, 18) = 20,925, p < 0,001) mostrando una velocidad global mayor de levantamiento con el agarre tradicional que con el reversible.

Usando la corrección de Greenhouse-Geisser, hubo una interacción significativa entre la carga y el tipo de agarre (F (1,48, 26,64) = 12,00, p < 0,001), mostrando que el tipo de agarre afectó a la velocidad de levantamiento a través de les diferentes cargas.

PRUEBA DE CHI CUADRADO PARA LA ASOCIACIÓN

La prueba de chi cuadrado (χ^2) de independencia (también conocida como prueba χ^2 de Pearson o prueba χ^2 de asociación) puede usarse para determinar si existe relación entre dos o más variables categóricas. El test produce una tabla de contingencia, o tabla de doble entrada, que muestra las agrupaciones cruzadas de las variables categóricas.

El test χ^2 pone a prueba la hipótesis nula de que no hay asociación entre dos variables categóricas. Compara las frecuencias observadas de los datos con las frecuencias que deberían esperarse si no hubiera relación entre ambas variables.

El análisis requiere cumplir con dos supuestos:

1. Las dos variables deben ser categóricas (nominales u ordinales).
2. Cada variable debería comprender dos o más grupos categóricos independientes.

La mayoría de los test estadísticos ajustan un modelo a los datos observados asumiendo la hipótesis nula de que no hay diferencia entre los datos observados y los modelados (esperados). El error o la desviación del modelo se calcula como:

$$\text{Desviación} = \sum (observado - modelo)^2$$

La mayoría de los modelos paramétricos se basan en medias y desviaciones estándar poblacionales. El modelo χ^2, en cambio, se basa en frecuencias esperadas.

¿Cómo se calculan las frecuencias esperadas? Por ejemplo, hemos categorizado a 100 personas entre hombres y mujeres y entre personas altas y bajas. Si existiera una distribución homogénea entre las 4 categorías, la frecuencia esperada = 100/4 o 25%. No obstante, los datos reales observados no presentan una distribución de la frecuencia homogénea.

Distribución homogénea	Hombres	Mujeres	Total fila
Alto/a	25	25	50
Bajo/a	25	25	50
Total columna	50	50	

Distribución observada	Hombres	Mujeres	Total fila
Alto/a	57	24	**81**
Bajo/a	14	5	**19**
Total columna	**71**	**29**	

El modelo basado en los valores esperados se puede calcular del siguiente modo:

Modelo (valores esperados) = (*total de fila x total de columna*) / 100

Modelo – hombre alto = (81 x 71) / 100 = 57,5
Modelo – mujer alta = (81 x 29) / 100 = 23,5
Modelo – hombre bajo = (19 x 71) / 100 = 13,5
Modelo – mujer baja = (19 x 29) / 100 = 5,5

Estos valores se pueden añadir a la tabla de contingencia:

	Hombre (H)	Mujer (M)	Total fila
Alto/a (A)	57	24	81
Esperado	**57,5**	**23,5**	
Bajo/a (B)	14	5	19
Esperado	**13,5**	**5,5**	
Total columna	71	29	

El estadístico χ^2 se deriva de $\sum \frac{(\textbf{observado}-\textbf{esperado})^2}{\textbf{esperado}}$

Validez

La prueba de χ^2 solo es válida cuando se dispone de un tamaño de muestra razonable, es decir, menos del 20% de las celdas tienen un valor esperado inferior a 5 y ninguna de ellas inferior a 1.

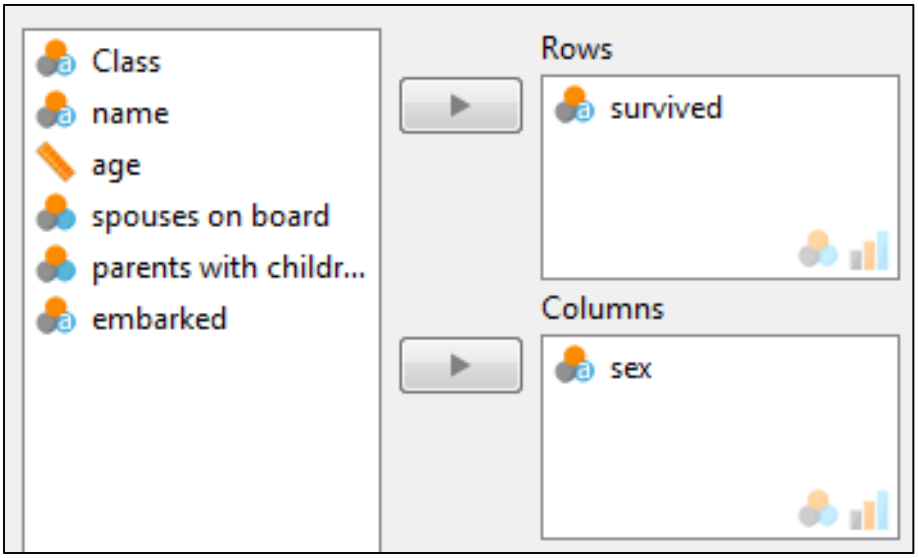
EJECUTANDO EL ANÁLISIS

El conjunto de datos **Titanic survival** es un conjunto clásico de datos usado en *machine learning* que contiene datos sobre 1.309 pasajeros y la tripulación que estaban a bordo del Titanic cuando se hundió en 1912. Podemos usarlo para ver las relaciones entre su supervivencia y otros factores. La variable dependiente es Survived y las variables independientes posibles son el resto de variables disponibles.

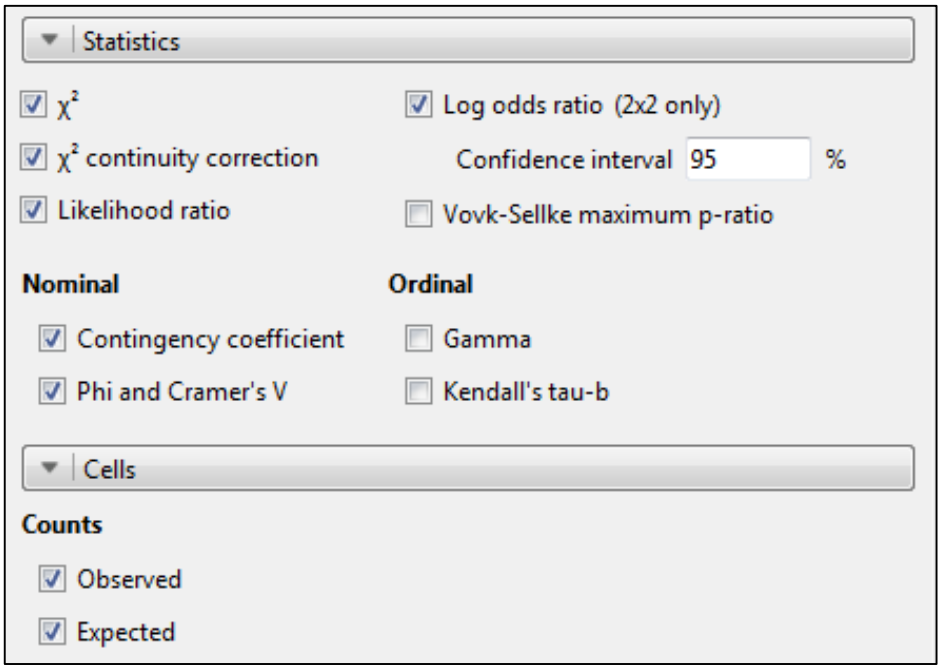
Class	survived	name	sex	age
Third	No	Abbing, Mr. Anthony	male	42
Third	No	Abbott, Master. Eugene Joseph	male	13
Third	No	Abbott, Mr. Rossmore Edward	male	16
Third	Yes	Abbott, Mrs. Stanton (Rosa Hunt)	female	35
Third	Yes	Abelseth, Miss. Karen Marie	female	16
Third	Yes	Abelseth, Mr. Olaus Jorgensen	male	25
Second	No	Abelson, Mr. Samuel	male	30
Second	Yes	Abelson, Mrs. Samuel (Hannah Wizosky)	female	28
Third	Yes	Abrahamsson, Mr. Abraham August Johannes	male	20
Third	Yes	Abrahim, Mrs. Joseph (Sophie Halaut Easu)	female	18
Third	No	Adahl, Mr. Mauritz Nils Martin	male	30
Third	No	Adams, Mr. John	male	26

Por convención,* la variable independiente se suele ubicar en las columnas de la tabla de contingencia y la variable dependiente en las filas.

Abra **Titanic survival.csv** en JASP, añada Survived en la caja «Rows» (filas) como variable dependiente y Sex en la caja «Columns» (columnas) como variable independiente.



Tras ello, marque las siguientes opciones:



* En realidad, el análisis produce el mismo resultado independientemente de la convención utilizada. Algunos autores, de hecho, recomiendan utilizar la contraria: las variables independientes en las filas y las variables dependientes en las columnas. En este texto seguimos la convención utilizada por el autor. (Nota del revisor.)

ENTENDIENDO EL RESULTADO

En primer lugar, eche un vistazo a la tabla de contingencia generada.

Contingency Tables

survived		sex: female	sex: male	Total
No	Count	127.0	682.0	809.0
	Expected count	288.0	521.0	809.0
	% within row	15.7 %	84.3 %	100.0 %
	% within column	27.3 %	80.9 %	61.8 %
	% of Total	9.7 %	52.1 %	61.8 %
Yes	Count	339.0	161.0	500.0
	Expected count	178.0	322.0	500.0
	% within row	67.8 %	32.2 %	100.0 %
	% within column	72.7 %	19.1 %	38.2 %
	% of Total	25.9 %	12.3 %	38.2 %
Total	Count	466.0	843.0	1309.0
	Expected count	466.0	843.0	1309.0
	% within row	35.6 %	64.4 %	100.0 %
	% within column	100.0 %	100.0 %	100.0 %
	% of Total	35.6 %	64.4 %	100.0 %

Recuerde que la prueba de χ^2 solo es válida cuando disponemos de un tamaño de muestra razonable, es decir, menos del 20% de las celdas con un valor esperado inferior a 5 y ninguna de ellas inferior a 1.

En la tabla, fijándonos en el % de fila ("% within row"), podemos observar que en el Titanic murieron más hombres que mujeres, y que sobrevivieron más mujeres que hombres. Sin embargo, ¿existe una relación significativa entre el género y la supervivencia?

Los resultados se muestran aquí:

Chi-Squared Tests

	Value	df	p
X^2	365.9	1	< .001
X^2 continuity correction	363.6	1	< .001
Likelihood ratio	372.9	1	< .001
N	1309		

El estadístico χ^2 (χ^2 (1) = 365,9, p < 0,001) sugiere que existe una relación significativa entre el género y la supervivencia.

La corrección por continuidad de χ^2 ("χ^2 continuity correction") puede usarse para prevenir una sobreestimación de la significación estadística en el caso de disponer de conjuntos de datos pequeños. Principalmente se usa cuando al menos una celda de la tabla tiene un valor esperado inferior a 5.

Como precaución, tenga en cuenta que esta corrección puede sobrecorregir el resultado del análisis y resultar demasiado conservadora, hasta el punto de que puede no rechazar la hipótesis nula cuando debería hacerlo (un error Tipo II).

La razón de verosimilitud ("Likelihood ratio") es una alternativa al chi cuadrado de Pearson. Se basa en la teoría de la máxima verosimilitud. Para muestras grandes, produce el mismo resultado que el χ^2 de Pearson. Se recomienda especialmente para muestras de tamaño pequeño, es decir, < 30.

En el caso de las variables nominales, el coeficiente Phi ("Phi-coefficient"; solo para tablas de contingencia de 2 x 2) y la V de Cramér ("Cramér's V"; la más popular), son pruebas de la magnitud de la asociación (es decir, tamaños del efecto). Ambos valores se encuentran en un rango de entre 0 (no hay relación) y 1 (relación perfecta). Puede verse que la magnitud de la relación entre las variables muestra un tamaño del efecto grande.

Nominal

	Value
Contingency coefficient	0.5
Phi-coefficient	0.5
Cramer's V	0.5

El coeficiente de contingencia ("Contingency coefficient") produce un valor ajustado de Phi y solo se recomienda en el caso de disponer de tablas de contingencia de gran tamaño, como las tablas de 5 x 5 o superiores.

Tamaño del efecto[4]	**df**	**Pequeño**	**Moderado**	**Grande**
Phi y V de Cramér (solo 2 x 2)	1	0,1	0,3	0,5
V de Cramér	2	0,07	0,21	0,35
V de Cramér	3	0,06	0,17	0,29
V de Cramér	4	0,05	0,15	0,25
V de Cramér	5	0,04	0,13	0,22

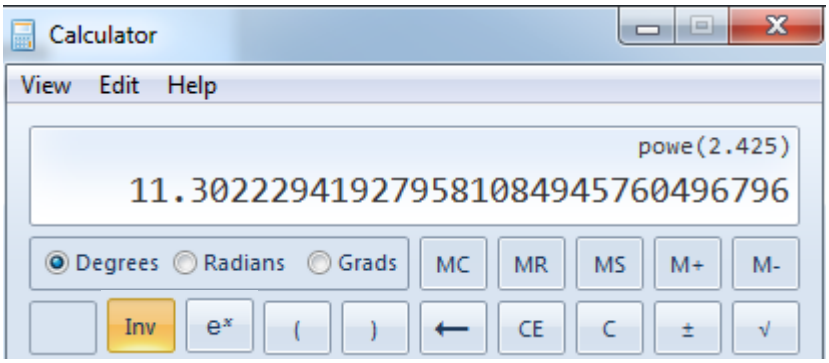
JASP también proporciona la razón de probabilidades (OR, del inglés *odds ratio*), usada para comparar la probabilidad relativa de ocurrencia del resultado de interés (supervivencia), dada la exposición a la variable de interés (en este caso, el género).

[4] Kim HY. Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. Restor. Dent. Endod. 2017; 42(2):152-155.

Log Odds Ratio ▼

		95% Confidence Intervals	
	Log Odds Ratio	Lower	Upper
Odds ratio	−2.425	−2.692	−2.159
Fisher's exact test	−2.423	−2.701	−2.150

Por alguna razón, JASP calcula las OR como un logaritmo natural. Para convertir estos valores, calcule el antilogaritmo natural (p. ej., utilizando la calculadora de Windows: introduzca el valor y después clique en **Inv** seguido de **e^x**), que en este caso es 11,3. Esto sugiere que los hombres tuvieron 11,3 veces más probabilidades de morir que las mujeres.



¿Cómo se calcula? Se deben usar los valores de la tabla de contingencia en las fórmulas siguientes:

Probabilidad[hombres] = Murieron / Sobrevivieron = 682 / 162 = 4,209.
Probabilidad[mujeres] = Murieron / Sobrevivieron = 127 / 339 = 0,374.

OR = Probabilidad[hombres] / Probabilidad[mujeres] = 11,3.

YENDO UN PASO MÁS ALLÁ

También se puede descomponer aún más la tabla de contingencia a modo de análisis post hoc, convirtiendo los recuentos y los recuentos esperados de cada celda en un residuo estandarizado. Esto puede revelar si las frecuencias observadas y las frecuencias esperadas son significativamente diferentes en cada celda.

El residuo estandarizado para cada celda de una tabla es una versión de la puntuación z estandarizada, calculada como:

$$z = \frac{\text{observado} - \text{esperado}}{\sqrt{\text{esperado}}}$$

En el caso especial en que df = 1, el cálculo del residuo estandarizado incluye un factor de corrección:

$$z = \frac{|\text{observado} - \text{esperado}| - 0{,}5}{\sqrt{\text{esperado}}}$$

El valor resultante de la z tiene un signo positivo si observado > estimado, y uno negativo si observado < estimado. Las significaciones de las puntuaciones z se muestran a continuación.

Puntuación z	**Valor p**
< -1,96 o > 1,96	< 0,05
< -2,58 o > 2,58	< 0,01
< -3,29 o > 3,29	< 0,001

Contingency Tables

survived		sex: female	sex: male	Total
No	Count	127.0	682.0	809.0
	Expected count	288.0	521.0	809.0
	% within row	15.7 %	84.3 %	100.0 %
	% within column	27.3 %	80.9 %	61.8 %
	% of Total	9.7 %	52.1 %	61.8 %
Yes	Count	339.0	161.0	500.0
	Expected count	178.0	322.0	500.0
	% within row	67.8 %	32.2 %	100.0 %
	% within column	72.7 %	19.1 %	38.2 %
	% of Total	25.9 %	12.3 %	38.2 %
Total	Count	466.0	843.0	1309.0
	Expected count	466.0	843.0	1309.0
	% within row	35.6 %	64.4 %	100.0 %
	% within column	100.0 %	100.0 %	100.0 %
	% of Total	35.6 %	64.4 %	100.0 %

Mujeres no z= -9,5	Hombres no z = 7,0
Mujeres sí z = 12,0	Hombres sí z = -8,9

Cuando calculamos las puntuaciones z para cada celda de la tabla de contingencia, se puede observar que murieron significativamente menos mujeres y más hombres de lo esperado (p < 0,001).

DISEÑO EXPERIMENTAL Y ORGANIZACIÓN DE LOS DATOS EN EXCEL PARA IMPORTAR A JASP

Prueba t para dos muestras independientes

Ejemplo de diseño:

Variable independiente	**Grupo 1**	**Grupo 2**
Variable dependiente	Dato	Dato

Variable independiente **Variable dependiente**

Categórica **Continua**

	A	B
1	Group	Data
2	1	0
3	1	0
4	1	3.8
5	1	6
6	1	0.7
7	1	2.9
8	1	2.8
9	1	2
10	1	2
11	1	8.5
12	1	1.9
13	1	3.1
14	1	1.5
15	1	3
16	1	3.6
17	1	0.9
18	1	-2.1
19	2	2
20	2	1.7
21	2	4.3
22	2	7
23	2	0.6
24	2	2.7
25	2	3.6

Si se requiere, se pueden añadir más variables dependientes.

Prueba t para dos muestras apareadas

Ejemplo de diseño:

Variable independiente	Pretest	Postest
Participante	**Variable dependiente**	
1	Dato	Dato
2	Dato	Dato
3	Dato	Dato
...n	Dato	Dato

	A	B
1	Pre-test	Post-test
2	60	60
3	103	103
4	58	54
5	60	54
6	64	63
7	64	61
8	65	62
9	66	64
10	67	65
11	69	61
12	70	68
13	70	67
14	72	71
15	72	69
16	72	68
17	82	81
18	58	60
19	58	56
20	59	57
21	61	57
22	62	55
23	63	62
24	63	60
25	63	59

Correlación

Ejemplo de diseño:

Correlación simple

Participante	Variable 1	Variable 2	Variable 3	Variable 4	Variable ...n
1	Dato	Dato	Dato	Dato	Dato
2	Dato	Dato	Dato	Dato	Dato
3	Dato	Dato	Dato	Dato	Dato
…n	Dato	Dato	Dato	Dato	Dato

Correlación múltiple

	A	B	C	D	E	F
1	Participant	Variable 1	Variable 2	Variable 3	Variable 4	Variable 5
2	1	533	77	77	106	106
3	2	472	63	59	92	93
4	3	484	82	77	93	78
5	4	536	72	72	103	93
6	5	630	77	68	104	93
7	6	563	68	68	101	87
8	7	531	77	82	108	106
9	8	344	50	50	86	92
10	9	346	54	50	90	86
11	10	386	59	54	85	80
12	11	460	54	63	89	83
13	12	492	63	59	92	94

Regresión

Ejemplo de diseño:

Regresión simple

Participante	Resultado	Predictor 1	Predictor 2	Predictor 3	Predictor ...n
1	Dato	Dato	Dato	Dato	Dato
2	Dato	Dato	Dato	Dato	Dato
3	Dato	Dato	Dato	Dato	Dato
…n	Dato	Dato	Dato	Dato	Dato

Regresión múltiple

	A	B	C	D	E	F
1	Participant	Outcome	Predictor 1	Predictor 2	Predictor 3	Predictor 4
2	1	533	77	77	106	106
3	2	472	63	59	92	93
4	3	484	82	77	93	78
5	4	536	72	72	103	93
6	5	630	77	68	104	93
7	6	563	68	68	101	87
8	7	531	77	82	108	106
9	8	344	50	50	86	92
10	9	346	54	50	90	86
11	10	386	59	54	85	80
12	11	460	54	63	89	83
13	12	492	63	59	92	94

Regresión logística

Ejemplo de diseño:

	Variable dependiente (categórica)	Factor (categórico)	Covariable (continua)
Participante	**Resultado**	**Predictor 1**	**Predictor 2**
1	Dato	Dato	Dato
2	Dato	Dato	Dato
3	Dato	Dato	Dato
…n	Dato	Dato	Dato

	A	B	C	D
1	ID	Outcome	Factor	Covariate
2	1	Yes	Yes	70
3	2	Yes	No	80
4	3	Yes	Yes	50
5	4	Yes	No	60
6	5	Yes	No	40
7	6	Yes	No	65
8	7	Yes	No	75
9	8	Yes	No	80
10	9	Yes	No	70
11	10	Yes	No	60
12	11	No	Yes	65
13	12	No	Yes	50
14	13	No	Yes	45
15	14	No	Yes	35
16	15	No	Yes	40
17	16	No	Yes	50
18	17	No	No	55
19	17	Yes	No	65
20	18	No	Yes	45

Si se requiere, se pueden añadir más factores y covariables.

ANOVA de medidas independientes de un factor

Ejemplo de diseño:

Variable independiente	**Grupo 1**	**Grupo 2**	**Grupo 3**	**Grupo…n**
Variable dependiente	Dato	Dato	Dato	Dato

Variable independiente

(Categórica)

Variable dependiente

(Continua)

	A	B
1	Group	Dependent variable
2	Group 1	3.8
3	Group 1	6
4	Group 1	0.7
5	Group 1	2.9
6	Group 1	2.8
7	Group 1	2
8	Group 1	2
9	Group 1	3.5
10	Group 2	1.9
11	Group 2	3.1
12	Group 2	1.5
13	Group 2	3
14	Group 2	3.6
15	Group 2	0.9
16	Group 2	-0.6
17	Group 3	1.1
18	Group 3	4.5
19	Group 3	6.1
20	Group 3	5
21	Group 3	2.4
22	Group 3	3.9
23	Group 3	3.5
24	Group 3	5.1
25	Group 3	3.5

Si se requiere, se pueden añadir más variables dependientes.

ANOVA de medidas repetidas de un factor

Ejemplo de diseño:

	Variable independiente (factor)			
Participante	**Nivel 1**	**Nivel 2**	**Nivel 3**	**Nivel ...n**
1	Dato	Dato	Dato	Dato
2	Dato	Dato	Dato	Dato
3	Dato	Dato	Dato	Dato
4	Dato	Dato	Dato	Dato
…n	Dato	Dato	Dato	Dato

Factor (tiempo)

	A	B	C	D
1	Participant	Week 0	Week 3	Week 6
2	1	6.42	5.83	5.75
3	2	6.76	6.2	6.13
4	3	6.56	5.83	5.71
5	4	4.8	4.27	4.15
6	5	8.43	7.71	7.67
7	6	7.49	7.12	7.05
8	7	8.05	7.25	7.1
9	8	5.05	4.63	4.67
10	9	5.77	5.31	5.33
11	10	3.91	3.7	3.66
12	11	6.77	6.15	5.96
13	12	6.44	5.59	5.64
14	13	6.17	5.56	5.51
15	14	7.67	7.11	6.96
16	15	7.34	6.84	6.82
17	16	6.85	6.4	6.29
18	17	5.13	4.52	4.45
19	18	5.73	5.13	5.17

Niveles

(Grupos relacionados)

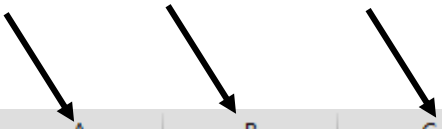
Si se requiere, se pueden añadir más niveles.

ANOVA de medidas independientes de dos factores

Ejemplo de diseño:

Factor 1	**Suplemento 1**			**Suplemento 2**		
Factor 2	**Dosis 1**	**Dosis 2**	**Dosis 3**	**Dosis 1**	**Dosis 2**	**Dosis 3**
Variable dependiente	Dato	Dato	Dato	Dato	Dato	Dato

Factor 1 **Factor 2** **Variable dependiente**



	A	B	C
1	supp	dose	len
2	OJ	1000	19.7
3	OJ	1000	23.3
4	OJ	1000	23.6
5	OJ	1000	26.4
6	OJ	1000	20
7	OJ	1000	25.2
8	OJ	1000	25.8
9	OJ	1000	21.2
10	OJ	1000	14.5
11	OJ	1000	27.3
12	OJ	2000	25.5
13	OJ	2000	26.4
14	OJ	2000	22.4
15	OJ	2000	24.5
16	OJ	2000	24.8
17	OJ	2000	30.9
18	OJ	2000	26.4
19	OJ	2000	27.3
20	OJ	2000	29.4
21	OJ	2000	23
22	VC	1000	16.5
23	VC	1000	16.5
24	VC	1000	15.2
25	VC	1000	17.3

Si se requiere, se pueden añadir más factores y variables dependientes.

ANOVA mixto

Ejemplo de diseño:

Factor 1 (Inter-sujetos)	**Grupo 1**			**Grupo 2**		
Niveles del factor 2 (Medidas repetidas)	**Prueba 1**	**Prueba 2**	**Prueba 3**	**Prueba 1**	**Prueba 2**	**Prueba 3**
1	Dato	Dato	Dato	Dato	Dato	Dato
2	Dato	Dato	Dato	Dato	Dato	Dato
3	Dato	Dato	Dato	Dato	Dato	Dato
...n	Dato	Dato	Dato	Dato	Dato	Dato

Factor 1

(Categórico)

Niveles del factor 2

(Continuo)

	A	B	C	D
1	Group	Level 1	Level 2	Level 3
2	Group 1	1.31	0.9	0.9
3	Group 1	1.29	0.89	0.72
4	Group 1	1.8	0.9	0.96
5	Group 1	1.4	1.26	0.97
6	Group 1	1.49	1.18	0.88
7	Group 1	1.35	1.15	0.92
8	Group 1	1.45	1.19	1
9	Group 1	1.21	1.2	0.85
10	Group 1	1.79	1.48	0.99
11	Group 1	1.73	1.68	0.98
12	Group 2	1.55	0.9	0.55
13	Group 2	1.27	0.95	0.41
14	Group 2	1.53	0.87	0.42
15	Group 2	1.26	1.15	0.44
16	Group 2	1.14	1.12	0.38
17	Group 2	1.11	1.08	0.34
18	Group 2	1.1	1.0758	0.18
19	Group 2	1.08	1.18	0.24
20	Group 2	1.3	1.26	0.39
21	Group 2	1.45	1.55	0.44

Chi cuadrado: tablas de contingencia

Ejemplo de diseño:

Participante	Respuesta 1	Respuesta 2	Respuesta 3	Respuesta …n
1	Dato	Dato	Dato	Dato
2	Dato	Dato	Dato	Dato
3	Dato	Dato	Dato	Dato
…n	Dato	Dato	Dato	Dato

Todos los datos deberían ser categóricos.

	A	B	C	D	E
1	Respondant	Response 1	Response 2	Response 3	Response 4
2	1	Female	clay	Morning	yes
3	2	Male	astro	Morning	No
4	3	Female	grass	Evening	No
5	4	Male	clay	Afternoon	No
6	5	Male	clay	Morning	No
7	6	Male	grass	Evening	No
8	7	Female	grass	Evening	yes
9	8	Male	clay	Morning	yes
10	9	Female	grass	Morning	No
11	10	Male	clay	Afternoon	No
12	11	Female	clay	Afternoon	No
13	12	Male	astro	Afternoon	No
14	13	Male	astro	Afternoon	No
15	14	Male	astro	Afternoon	yes
16	15	Female	clay	Morning	No
17	16	Male	astro	Afternoon	yes
18	17	Female	astro	Afternoon	yes
19	18	Male	grass	Morning	No
20	19	Male	clay	Afternoon	No

ALGUNOS CONCEPTOS EN ESTADÍSTICA FRECUENTISTA

La aproximación frecuentista es la metodología estadística más comúnmente enseñada y utilizada. Describe los resultados obtenidos a partir de una muestra basados en la frecuencia o proporción de los datos a partir de estudios repetidos mediante los cuales se define la probabilidad de los sucesos.

La estadística frecuentista utiliza marcos de referencia rígidos que incluyen la prueba de hipótesis, los valores de p, los intervalos de confianza, etc.

Prueba de hipótesis

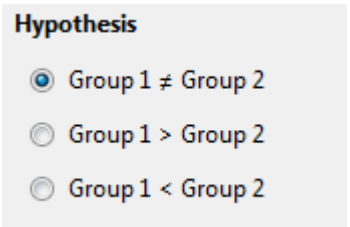
Una hipótesis puede ser definida como "**una explicación tentativa basada en evidencias limitadas como punto de partida para investigaciones adicionales**".

Hay dos tipos básicos de hipótesis, una **hipótesis nula** (H_0) y una **hipótesis alternativa o experimental** (H_1). La **hipótesis nula** es la posición por defecto para la mayoría de los análisis estadísticos en los cuales se ha establecido que no existe relación ni dependencia entre grupos. La **hipótesis alternativa** establece que existe una relación o una diferencia entre los grupos y la dirección de esta diferencia o relación. Por ejemplo, si se lleva a cabo un estudio para observar los efectos de un suplemento sobre el tiempo de esprint en un grupo de participantes comparado con un grupo placebo:

1) H_0 = **no** hay diferencias en los tiempos de esprint entre los dos grupos.
2) H_1 = hay diferencias en los tiempos de esprint entre los dos grupos.
3) H_2 = el grupo 1 es mejor que el grupo 2.
4) H_3 = el grupo 1 es peor que el grupo 2.

La prueba de hipótesis se refiere a los procedimientos estrictamente predefinidos que se utilizan para aceptar o rechazar las hipótesis y la probabilidad de que pudiera ser el resultado de la mera casualidad. La confianza con la que se acepta o rechaza una hipótesis nula se denomina nivel de significación. El nivel de significación se denota por α, normalmente 0,05 (5%). Esta es la probabilidad de aceptar un efecto como verdadero (95%) y que solamente haya un 5% de probabilidad que el resultado se dé por mera casualidad.

En JASP se pueden seleccionar fácilmente distintos tipos de hipótesis. Sin embargo, la hipótesis nula siempre aparece marcada por defecto.



Errores de Tipo I y Tipo II

La probabilidad de rechazar la hipótesis nula cuando en realidad es verdadera se llama error de Tipo I, mientras que la probabilidad de aceptar la hipótesis nula cuando no es verdadera se conoce como error de Tipo II.

		La verdad	
		No culpable (H_0)	**Culpable (H_1)**
El veredicto	**Culpable (H_1)**	**Error de Tipo I** Un inocente es encarcelado	Decisión correcta
	No culpable (H_0)	Decisión correcta	**Error de Tipo II** Un culpable queda libre

El error de Tipo I se considera el peor error que se puede cometer en el análisis estadístico.

La potencia de una prueba estadística se define como la probabilidad de que el test rechace la hipótesis nula cuando la hipótesis alternativa es verdadera. Para un determinado nivel de significación, si el tamaño de la muestra aumenta, la probabilidad de cometer errores de Tipo II disminuye, por lo que se incrementa la potencia estadística.

Prueba de hipótesis

La esencia de la prueba de hipótesis es definir en primer lugar la **hipótesis nula (o la alternativa)**, establecer el nivel de **α**, normalmente 0,05 (5%), y recopilar y analizar datos de una muestra. Utilizamos un **estadístico** para determinar a qué distancia (o el número de desviaciones estándar) está la media observada en la muestra en relación con la media de la población establecida en la hipótesis nula. El valor del estadístico se compara con un valor crítico. Este es un valor de corte que define el límite en el que se pueden obtener menos del 5% de las medidas de diferentes muestras si la hipótesis nula es verdadera.

Si la probabilidad de obtener por casualidad una diferencia entre las medias es inferior al 5% cuando se define la hipótesis nula, se puede rechazar la hipótesis nula y aceptar la hipótesis alternativa.

El **valor de p** es la probabilidad de obtener un resultado en una muestra, suponiendo que el valor definido en la hipótesis nula es verdadero. Si el valor de p es inferior al 5% ($p < 0,05$), se rechaza la hipótesis nula. Cuando el valor de p es superior al 5% ($p > 0,05$), aceptamos la hipótesis nula.

Tamaño del efecto

El tamaño del efecto es una medida estándar que puede calcularse en muchos tipos de análisis estadístico. Si la hipótesis nula es rechazada, el resultado es significativo. Esta significación solo evalúa la probabilidad de obtener el resultado en la muestra por casualidad, pero no indica cómo de grande es la diferencia (significación práctica), ni se puede utilizar para comparar entre diferentes estudios.

El tamaño del efecto indica la magnitud de la diferencia entre los grupos. Así, por ejemplo, si se diera una disminución significativa de los tiempos en el esprint en distancias de 100 m en un grupo que toma suplementos alimenticios en comparación con otro grupo placebo, el tamaño del efecto indicaría cuánto más efectiva fue la intervención con estos suplementos. A continuación, se muestran algunos de los tamaños del efecto más comunes.

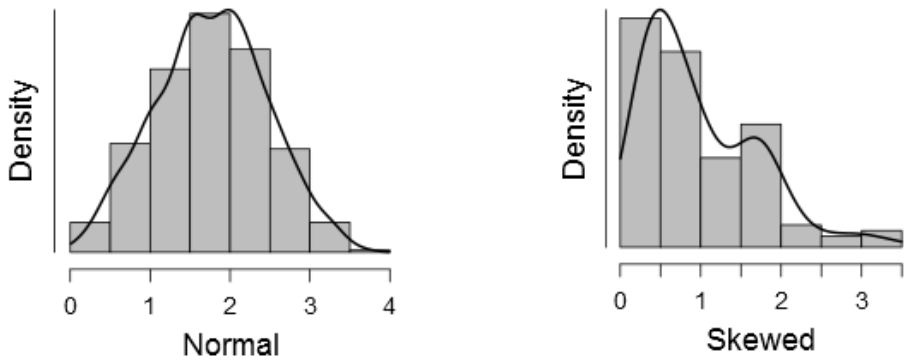
Test	Medida	Irrelevante	Pequeño	Medio	Grande
Entre medias	d de Cohen	< 0,2	0,2	0,5	0,8
Correlación	Coeficiente de correlación (r) Rango biserial (r_B) Rho de Spearman	< 0,1 < 0,1 < 0,1	0,1 0,1 0,1	0,3 0,3 0,3	0,5 0,5 0,5
Regresión múltiple	Coeficiente de correlación múltiple (R)	< 0,10	0,1	0,3	0,5
ANOVA	Eta Eta parcial Omega cuadrado	< 0,1 < 0,01 < 0,01	0,1 0,01 0,01	0,25 0,06 0,06	0,37 0,14 0,14
Chi cuadrado	Phi (solo en tablas 2x2) V de Cramér Razón de probabilidades (solo en tablas 2x2)	< 0,1 < 0,1 < 1,5	0,1 0,1 1,5	0,3 0,3 3,5	0,5 0,5 9,0

En conjuntos de datos pequeños, puede haber un tamaño del efecto de moderado a grande pero no haber diferencias significativas. Esto puede sugerir que el análisis no tuvo suficiente potencia estadística y que el aumento en el número de puntos de datos podría mostrar un resultado significativo. Por el contrario, cuando se usan conjuntos de datos grandes, las pruebas significativas pueden ser engañosas, ya que efectos pequeños o irrelevantes pueden producir resultados estadísticamente significativos.

PRUEBA PARAMÉTRICA *vs*. PRUEBA NO PARAMÉTRICA

La mayoría de las investigaciones recogen información a partir de una muestra de la población de interés ya que normalmente resulta imposible recopilar datos de toda la población. Sin embargo, queremos saber hasta qué punto los datos recogidos reflejan adecuadamente la media, la desviación estándar, la proporción, etc., de la población basándonos en la distribución paramétrica de estas funciones. Estas medidas son los **parámetros poblacionales**. Las estimaciones de estos parámetros en la muestra son los estadísticos. La estadística paramétrica requiere que se establezcan supuestos sobre los datos que incluyen la distribución normal y la homogeneidad de la varianza.

En algunos casos se pueden violar estos supuestos, en el sentido en que los datos pueden ser notablemente asimétricos:

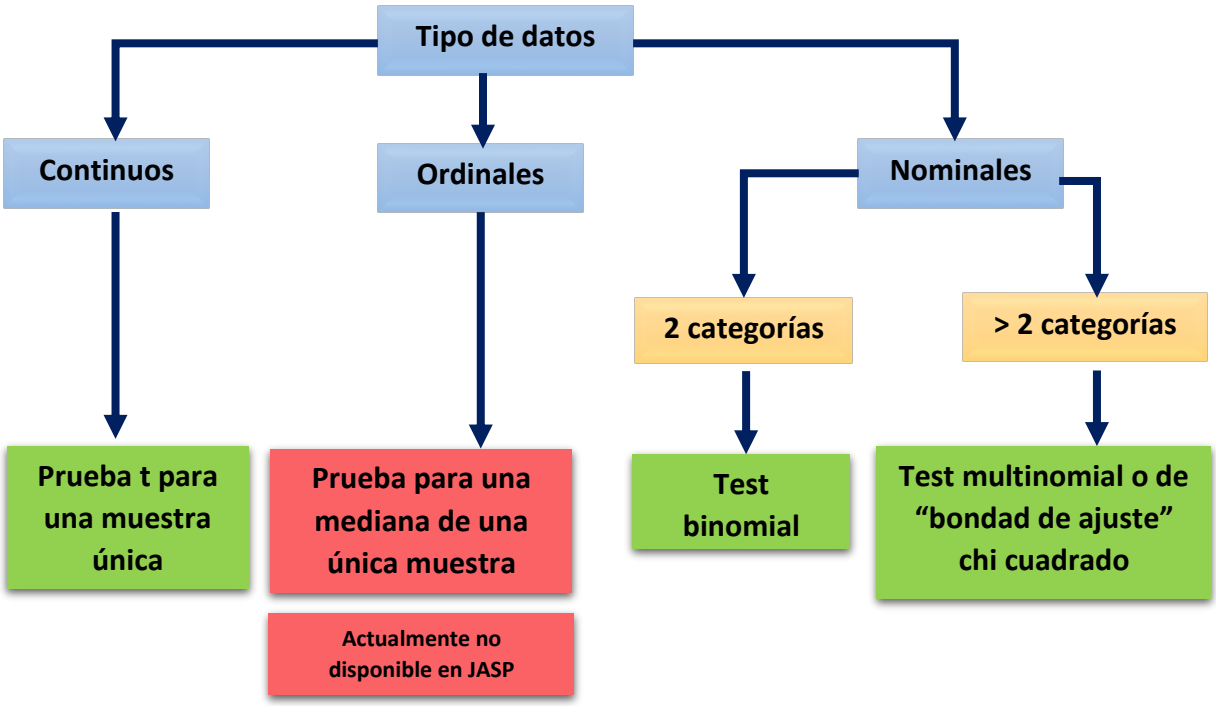


A veces, transformar los datos sirve para rectificar esta situación, pero no siempre es así. También es común recoger datos ordinales (p. ej., puntuaciones en escalas Likert) para los cuales algunos términos como media y desviación estándar no tienen sentido. Como tal, no hay parámetros asociados con datos ordinales (**no paramétricos**). Las alternativas no paramétricas incluyen, entre otras, la mediana y los cuartiles.

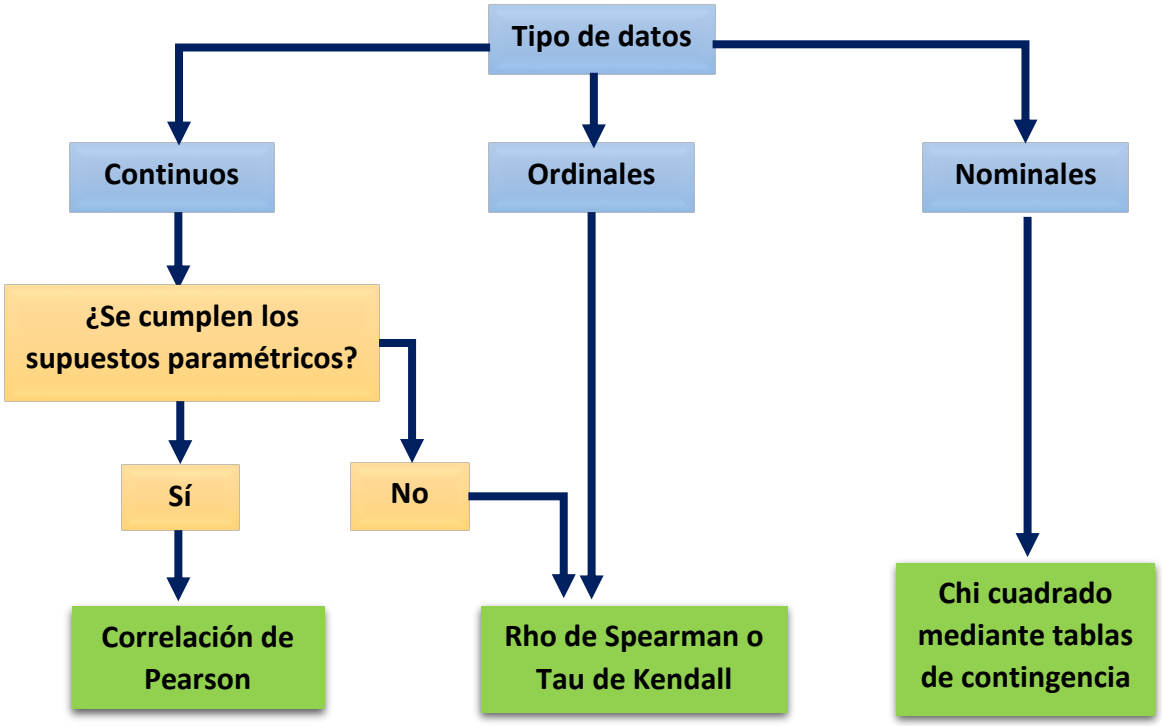
Para estos dos casos disponemos de pruebas estadísticas no paramétricas. Existen equivalentes para la mayoría de las pruebas paramétricas clásicas más comunes. Estas pruebas no asumen una distribución normal de los datos o la existencia de parámetros en la población, y se basan en la ordenación de los datos en rangos, de los valores más bajos a los más altos. Todos los cálculos posteriores se realizan con estos rangos en lugar de hacerlo con los valores de los datos reales.

¿QUÉ PRUEBA DEBERÍA USAR?

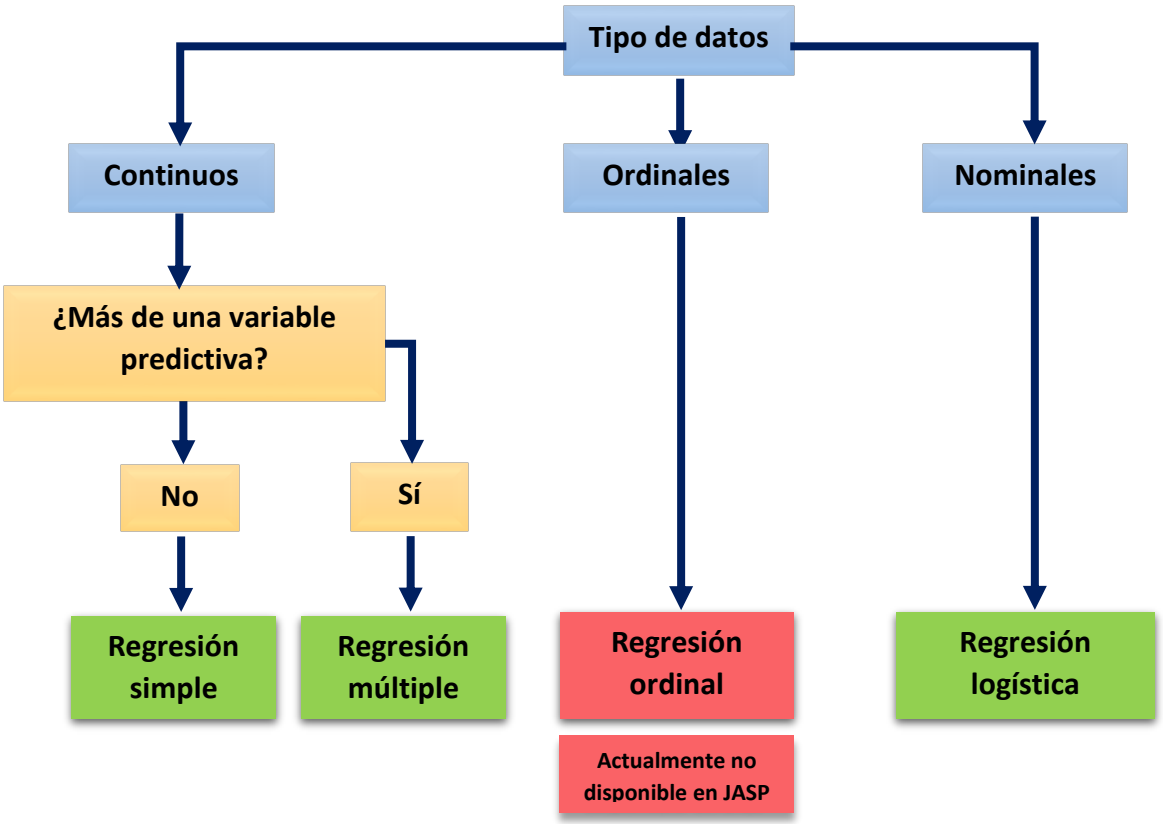
Comparación de una media muestral con la media conocida o hipotética poblacional



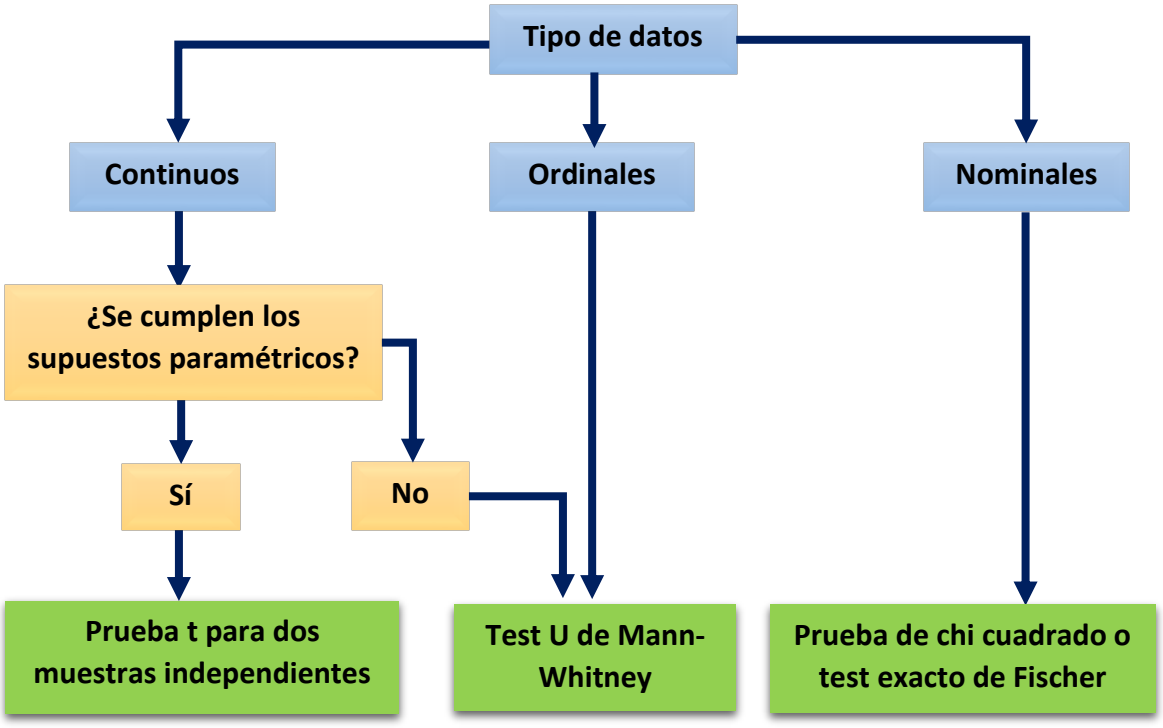
Prueba para la relación entre dos o más variables



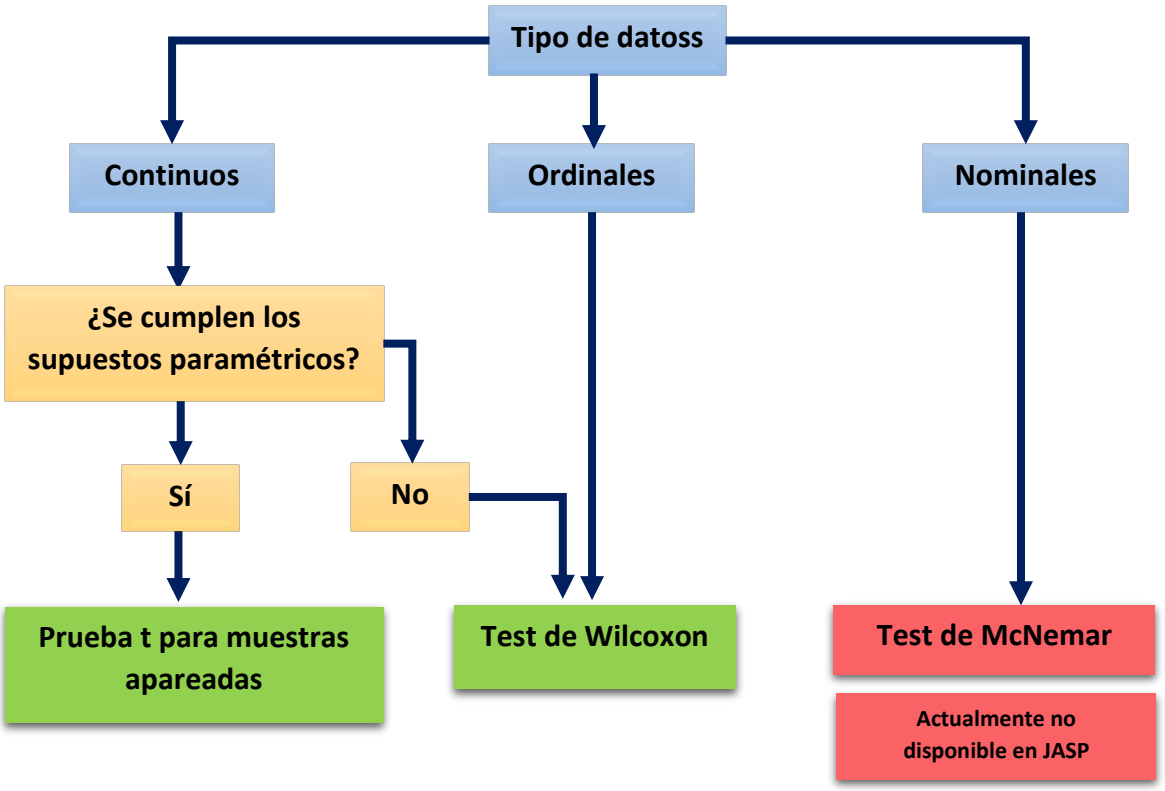
Predicción de resultados



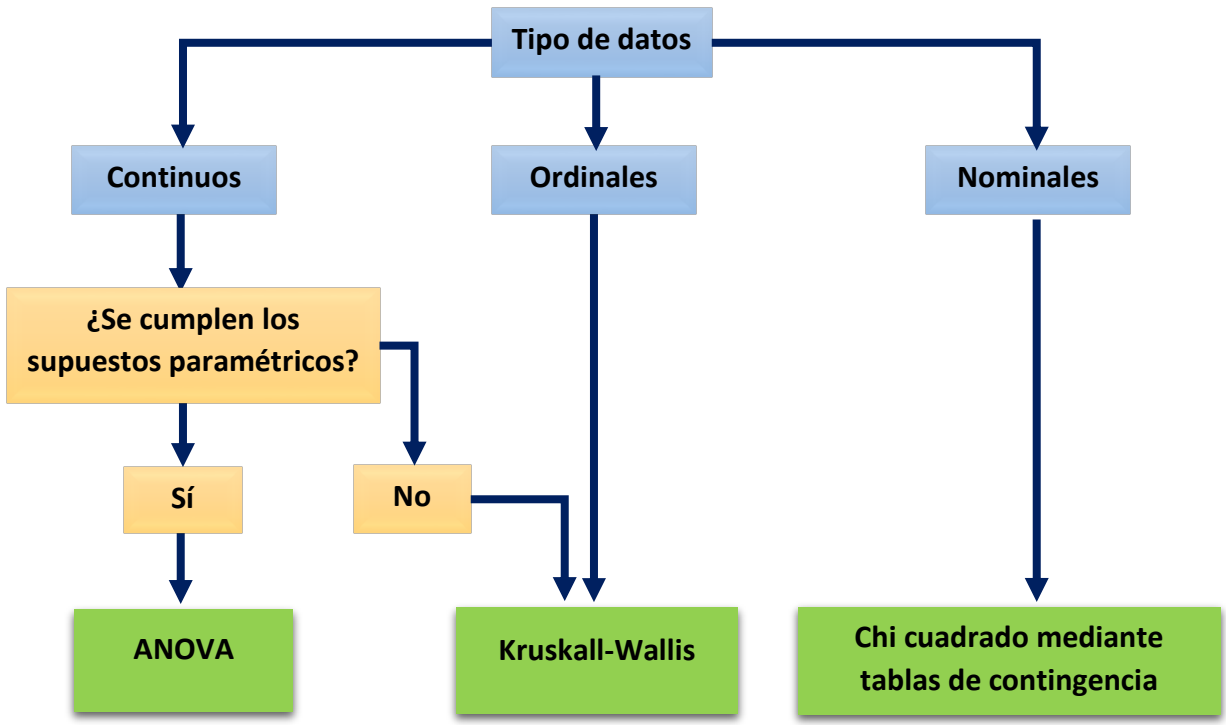
Prueba para las diferencias entre dos grupos independientes



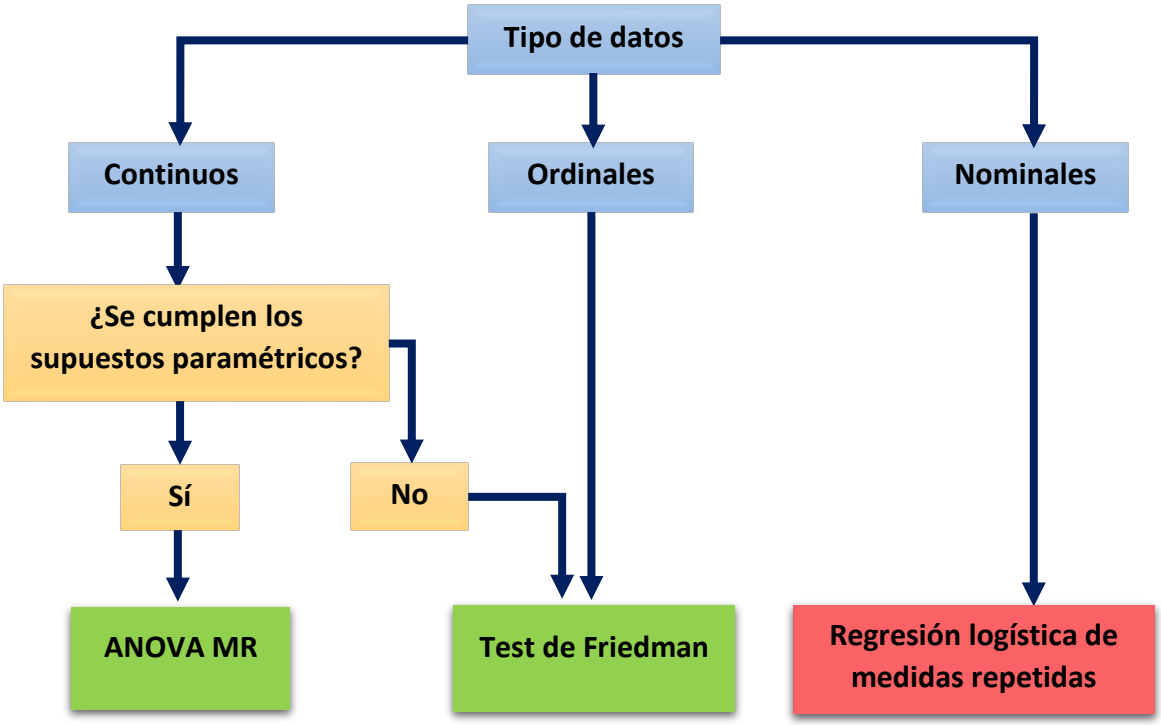
Prueba para dos grupos relacionados



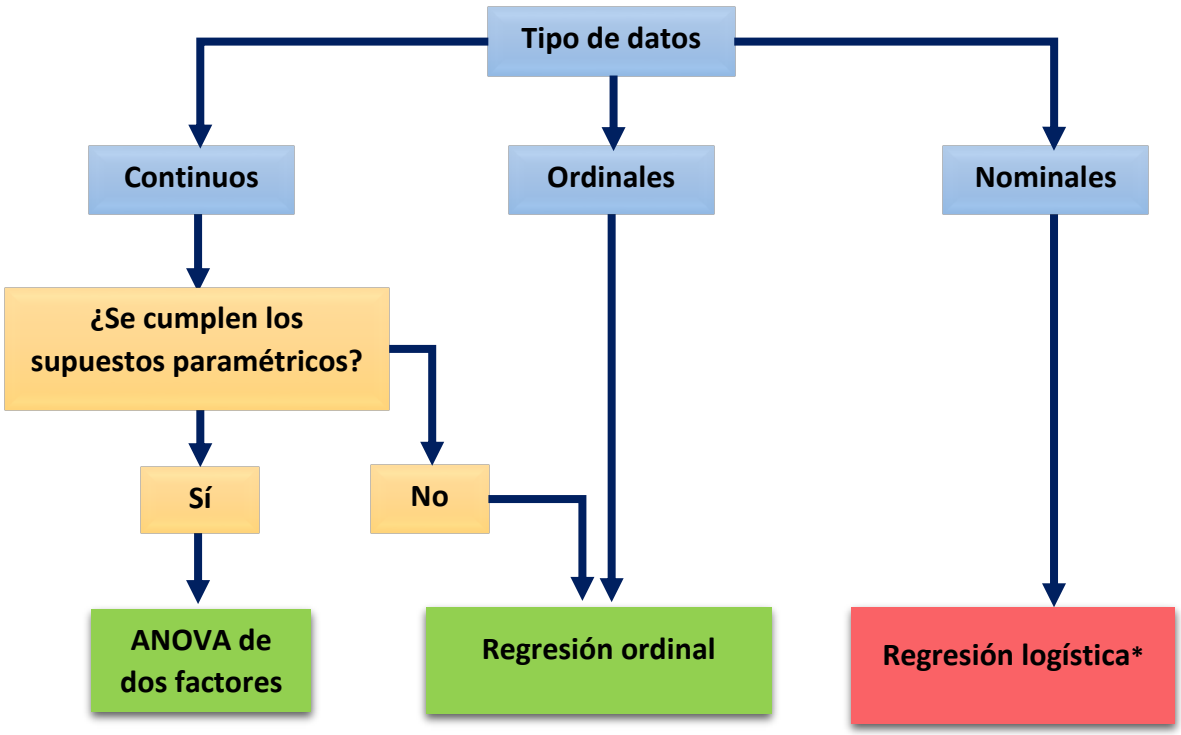
Prueba para las diferencias entre tres o más grupos independientes



Prueba para las diferencias entre tres o más grupos relacionados



Prueba para interacciones entre dos o más variables independientes



* Aunque aparece como no disponible en este diagrama, la regresión logística es un procedimiento disponible en la versión 0.9.2 de JASP. (Nota del revisor.)