

Análisis de Datos con jamovi

© 2023, Josep Maria Losilla Vidal y Jaume Vives Brosa
Universitat Autònoma de Barcelona (UAB)

Citar como:

Losilla, J.M. y Vives, J. (2023). *Análisis de Datos con jamovi*. [URL de descarga]

Esta obra está licenciada bajo Licencia Creative Commons
Atribución – NoComercial – SinDerivadas 4.0 Internacional (CC BY-NC-ND 4.0)



Índice

0. Breve introducción a jamovi para R.....	3
0.1. Instalación de jamovi y de módulos adicionales desde su librería.....	3
0.2. Elementos principales de la interfaz de usuario de jamovi.....	4
0.3. Salir de jamovi guardando los datos y/o los resultados.....	5
1. Definición de la matriz de datos.....	5
1.1. Propiedades de las variables en jamovi.....	5
1.2. Ejercicio: definir la matriz de datos del <i>Cuestionario de hábitos tóxicos</i>	9
1.3. Importar una matriz de datos.....	11
1.4. Definir las propiedades de las variables.....	12
1.5. Guardar la estructura de la matriz de datos y de los resultados como plantilla (<i>template</i>).....	15
1.6. Introducir manualmente nuevos casos a la matriz de datos.....	15
1.7. Solución del ejercicio.....	16
2. Transformación y cálculo de variables.....	16
2.1. Recodificar variables.....	16
2.2. Calcular variables mediante expresiones aritméticas y estadísticas.....	19
2.3. Calcular variables mediante expresiones lógicas.....	24
2.4. Calcular recuentos.....	28
2.5. Calcular tiempos transcurridos entre fechas y horas.....	28
2.6. Ejercicio.....	30
3. Análisis estadísticos descriptivos.....	31
3.1. Descripción estadística de una variable categórica.....	31
3.2. Descripción estadística de una variable cuantitativa.....	36
3.3. Descripción estadística de la relación lineal entre dos variables cuantitativas.....	39
3.4. Descripción estadística de la relación entre dos variables categóricas.....	46
3.5. Descripción estadística de la relación entre una variable categórica y una variable cuantitativa.....	49
3.5.1. Relación entre una variable categórica binaria y una variable cuantitativa.....	49
3.5.2. Relación entre una variable categórica ordinal o politómica y una variable cuantitativa.....	53
3.6. Ejercicio.....	54
Parte 1. Descriptiva univariante.....	54
Parte 2. Descriptiva bivariante.....	56
3.7. Síntesis de menús y opciones jamovi para la descripción estadística uni y bivariante.....	58
A0. Filtro de datos.....	58
A1. Descripción estadística de una variable categórica.....	58
A2. Descripción estadística de una variable cuantitativa.....	58
A3. Relación lineal entre dos variables cuantitativas.....	59
A4. Relación entre dos variables categóricas.....	59
A5. Relación entre una variable categórica y una variable cuantitativa.....	59
4. Inferencia estadística y contraste de hipótesis.....	60
4.1. Fundamentos teóricos.....	60
4.1.1. Parámetros y estadísticos.....	60
4.1.2. Distribución muestral de un estadístico.....	60
4.1.3. Intervalo de confianza de un parámetro poblacional.....	64
4.1.4. Contraste de hipótesis mediante intervalos de confianza.....	66
4.1.5. Contraste de hipótesis mediante pruebas de significación.....	68
4.1.6. Significación estadística y relevancia práctica: valoración del tamaño del efecto.....	69
4.1.7. Significación estadística y relación causal.....	70

4.2.	Estimación por intervalo y contraste de hipótesis con una variable categórica	71
4.3.	Estimación por intervalo y contraste de hipótesis con una variable cuantitativa	75
4.4.	Contraste de hipótesis acerca de la asociación entre dos variables cuantitativas: coeficiente de correlación lineal de Pearson.....	77
4.5.	Contraste de hipótesis acerca de la asociación entre dos variables categóricas: comparación de proporciones.....	79
4.6.	Contraste de hipótesis acerca de la asociación entre una variable categórica y una variable cuantitativa: comparación de medias y análisis de la variancia	82
4.6.1.	Relación entre una variable categórica binaria y una variable cuantitativa continua	83
4.6.2.	Relación entre una variable categórica binaria y una variable cuantitativa de recuento	86
4.6.3.	Relación entre una variable categórica ordinal o politómica y una variable cuantitativa continua	87
4.6.4.	Relación entre una variable categórica ordinal o politómica y una variable cuantitativa de recuento.....	89
4.7.	Ejercicio	91
	Parte 1. Preparación de los datos (generación de variables)	91
	Parte 2. Inferencia estadística	91
	Parte 3. Contraste de hipótesis bivalente	92
4.8.	Síntesis de menús y comandos jamovi para la descripción estadística, la estimación de intervalos de confianza y el contraste de hipótesis univariante y bivalente	94
	A0. Filtro de datos	94
	A1. Intervalos de confianza y pruebas de conformidad con una variable categórica	94
	A2. Intervalos de confianza y pruebas de conformidad con una variable cuantitativa	94
	A3. Relación lineal entre dos variables cuantitativas (correlación lineal)	95
	A4. Relación entre dos variables categóricas (comparación de proporciones).....	95
	A5. Relación entre una variable categórica y una variable cuantitativa (comparación de medias).....	95
5.	Ejercicios basados en estudios reales.....	97
5.1.	Lenguaje en adultos con síndrome de Down	97
5.2.	Evocación del léxico nominal por niños/as con implante coclear en una tarea de denominación.....	99
5.3.	Percepción del profesorado tutor del alumnado sordo sobre la cultura y las prácticas educativas inclusivas	102
5.4.	Desarrollo evolutivo de la conciencia fonológica	104

0. Breve introducción a jamovi para R

R es un entorno completo para el análisis estadístico basado en un lenguaje del mismo nombre y que incluye un gran número de funciones estadísticas agrupadas en librerías (*packages*). Además de por tratarse de un entorno interactivo abierto y de libre distribución, **R** también es conocido por la alta calidad de sus gráficos y la extensa lista de funciones estadísticas que incorpora, fruto del trabajo colaborativo de centenares de investigadores en estadística. En este sentido, **R** cuenta con una de las comunidades más dinámicas en el ámbito del desarrollo de procedimientos estadísticos, y actualmente es una plataforma de referencia en la que los autores implementan sus funciones, ejemplos de sus cursos de estadística, etc. Las librerías de funciones estadísticas de **R** están en continua revisión y ampliación, y pueden descargarse gratuitamente (<http://cran.r-project.org>).

0.1. Instalación de jamovi y de módulos adicionales desde su librería

jamovi es un entorno de menús (GUI, *Graphical User Interface*) para **R** de uso muy sencillo, que incluye todos los índices y gráficos estadísticos que se exponen en este curso. **jamovi** se puede descargar gratuitamente desde su página web (<https://www.jamovi.org/download.html>). Recomendamos instalar la versión más actual (*current version*).

¡Atención! **jamovi** solo funciona en ordenadores con procesador de 64 bits.

Al instalar **jamovi** también instala automáticamente una versión interna del paquete estadístico **R**. No es necesario, por tanto, descargar e instalar previamente **R** para trabajar con **jamovi**. Al ejecutar las opciones de los menús y paneles de opciones de **jamovi**, éste traduce y ejecuta código **R** de forma oculta para el usuario, y presenta los resultados en su panel gráfico.

Una característica que otorga a **jamovi** una gran potencia y extensibilidad es la posibilidad de instalar **módulos** adicionales desde su librería. Para instalar estos módulos basta con clicar en el botón  **Modules** que se encuentra en la parte superior derecha de la ventana de **jamovi**. En el menú que aparece hay que clicar **jamovi-library** y desplazarse por la lista para elegir el módulo que se desea instalar pulsando el botón **INSTALL**.

Para llevar a cabo los análisis estadísticos numéricos y gráficos que se presentarán en los siguientes apartados es necesario que estén instalados los siguientes módulos:

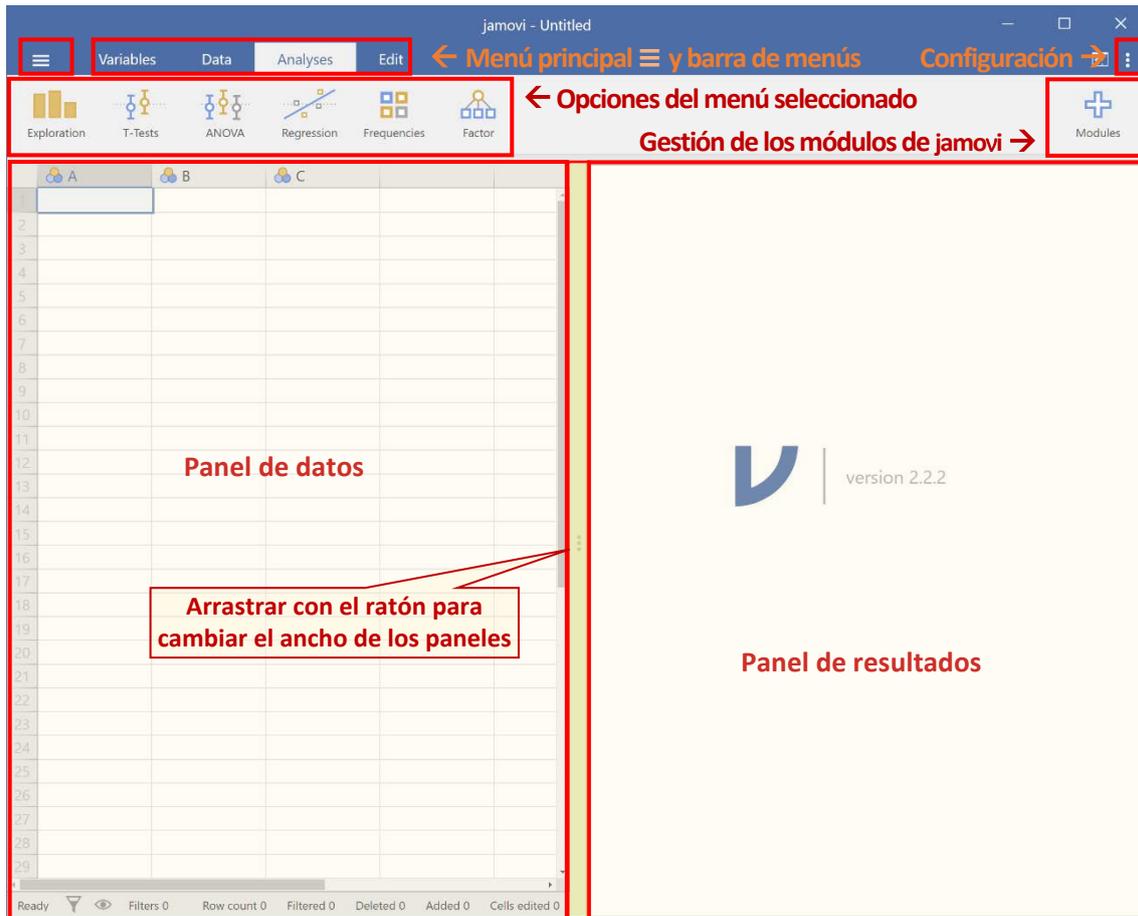
- *esci* – *Effect Sizes and Confidence Intervals for R and jamovi*
- *gamlj* – *General Analyses for Lineal Models in jamovi*
- *jpowers* – *Power Analysis for Common Research Designs*
- *Rj* – *Editor to run R code inside jamovi*
- *surveymv* – *Survey Plots*
- *scatr*

Los módulos adicionales que se instalan se muestran como un nuevo icono en el menú **Analyses**, o bien no muestran un nuevo icono, sino que añaden nuevas opciones a los menús de análisis existentes.

¡Atención! Cuando se instala una nueva versión de **jamovi** es necesario ejecutar el botón  **Modules** para actualizar manualmente aquellas librerías que lo requieran.

0.2. Elementos principales de la interfaz de usuario de jamovi

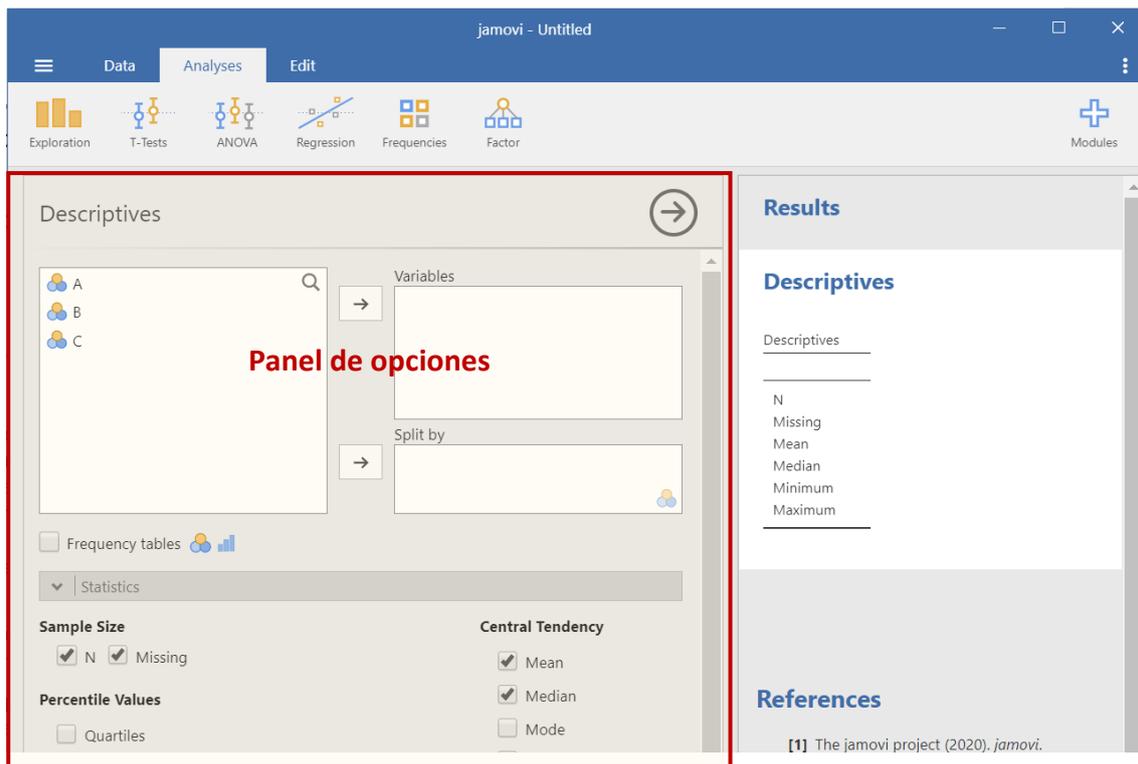
La siguiente figura muestra los cinco elementos principales de la interfaz de usuario de jamovi.



El panel izquierdo (**panel de datos**) funciona como una hoja de cálculo en la que se visualizará y se podrá editar el contenido de la matriz de datos que se esté analizando. En este panel se puede escribir directamente como lo harías en un programa de hoja de cálculo, o bien importar los datos desde un fichero guardado en cualquiera de los formatos habituales de los programas de análisis estadístico (SPSS, SAS, Stata, R, JASP, Excel, CSV, etc.).

En el panel derecho (**panel de resultados**) es donde se irán mostrando y actualizando en tiempo real los resultados de los análisis estadísticos que se ejecuten. Haciendo clic con el botón derecho del ratón en los resultados de cada análisis individual numérico y/o gráfico se pueden copiar al portapapeles para luego pegarlos en un documento (**Analysis | Copy**), guardarlos en ficheros independientes en diferentes formatos (PDF, HTML, etc.) (**Analysis | Export**), o bien borrarlos del panel (**Analysis | Remove**). Por último, la barra que separa el panel de datos y el de resultados se puede arrastrar hacia la izquierda o hacia la derecha para cambiar sus tamaños.

Al clicar cualquier menú se abre un **panel de opciones** en el que se pueden seleccionar e indicar las opciones particulares para llevar a cabo el análisis elegido. Una vez establecidas las opciones, el análisis se ejecuta haciendo clic en la flecha situada en la parte superior derecha del panel de opciones. Para cambiar posteriormente cualquiera de las opciones, basta con clicar en sus resultados.



0.3. Salir de jamovi guardando los datos y/o los resultados

Para salir de jamovi basta con cerrar su ventana, pero antes es recomendable guardar los cambios realizados durante la sesión de trabajo. jamovi permite guardar la matriz de datos junto con los resultados de los análisis realizados en un único fichero con un formato propio de jamovi y con la extensión **.omv**. Para guardar este fichero se debe ejecutar la opción **Save** del menú principal  (situado en la parte superior izquierda de la ventana), o con la opción **Save** si ya se había guardado previamente. También se puede guardar por separado la matriz de datos en diferentes formatos mediante la opción **Export** del menú principal de jamovi.

1. Definición de la matriz de datos

1.1. Propiedades de las variables en jamovi

Los análisis estadísticos se llevan a cabo sobre un conjunto de datos, que denominamos genéricamente **matriz de datos**, que presenta una estructura rectangular en la cual en las filas se encuentran los **casos** y en las columnas las **variables**. En este apartado nos centraremos en el diseño de la matriz de datos, es decir, en la creación de las variables y la definición de sus propiedades.

Estructura de la matriz de datos: **casos (filas)** y **variables (columnas)**

Como ya hemos dicho, la matriz de datos es una estructura bidimensional en la cual las filas (que denominaremos **casos**) contienen la información correspondiente a cada entidad que se desea almacenar y analizar (por ejemplo, la entidad "persona"). La información de cada **caso** se divide en columnas (que denominaremos **variables**), cada una de las cuales contiene un único valor o **dato** de una determinada característica o atributo. Por ejemplo, para la entidad "persona" puede ser de interés almacenar y analizar las variables NIF, género, fecha de nacimiento y de respuesta, peso, resultado de una prueba diagnóstica, etc.

	NIF	FNAC	FRESP	GENERO	PESO	TALLA	DIAG
CASO 1 →	77832898R	07-04-1970	25-01-1990	1	64.7	169	0112
CASO 2 →	39876564A	04-11-1969	24-01-1990	2	101.5	170	0109

CASO 20 →	38356912J	05-08-1969	24-01-1990	1	70.9	167	0117
	↑ VAR 1	↑ VAR 2	↑ VAR 3	↑ VAR 4	↑ VAR 5	↑ VAR 6	↑ VAR 7

Propiedades de las variables

jamovi permite indicar las siguientes propiedades para cada variable de la matriz de datos:

- **Nombre (Name):** debe ser único, empezar con una letra y puede contener a continuación cualquier combinación de caracteres alfanuméricos (sin signos de puntuación) y el símbolo especial de guion bajo (_ subrayado). Aunque la longitud del nombre puede ser larga, es preferible utilizar nombres cortos, de un máximo de 8 caracteres. Recuerde que jamovi presentará el nombre respetando las mayúsculas y minúsculas, y que **distingue entre mayúsculas y minúsculas en el momento de escribir los nombres de las variables**. Es recomendable utilizar la denominada “**notación húngara**”, que consiste en escribir en mayúsculas las iniciales de las palabras que compongan las etiquetas compuestas por más de un término (e.g., FResp, NumCig).
- **Etiqueta (Description):** texto descriptivo del contenido de la variable. **Es recomendable indicar explícitamente la unidad de medida de la variable en esta etiqueta**. Nota: **es preferible no incluir acentos ni símbolos especiales (e.g., la ñ) en la etiqueta** porque en algunos casos no se muestran correctamente.
- **Escala de medida (Measure type):** hace referencia a la escala de medida de la variable. jamovi permite distinguir entre **cuatro subtipos: Nominal** (escala categórica), **Ordinal** (escala categórica), **Continuous** (escala cuantitativa) e **ID** (identificador de caso):

- **Tipo de datos (Data type):** hace referencia a la codificación de los datos que almacena la variable. jamovi permite distinguir entre **dos subtipos numéricos (entero -integer- y decimal) y uno textual:**

Una limitación importante de jamovi respecto a otros programas de análisis estadístico, es el hecho de que no incluye las **fechas y horas** como tipo de dato. La solución más sencilla para almacenar fechas y horas en una matriz de datos jamovi es hacerlo como texto (con escala nominal y tipo de datos textual) y utilizar posteriormente las funciones `SPLIT()` e `INT()`, que se revisaran en el apartado 2.2 al tratar la generación de nuevas variables, para extraer en nuevas variables numéricas enteras los tres elementos que componen una fecha (día, mes y año) o una hora (hora, minutos, segundos). Otra opción es almacenarlas como variables numéricas separadas con los contenidos, escala de medida y tipo de datos mencionados.

- **Valores faltantes (*Missing values*):** la ausencia de valor en una variable de tipo cuantitativo, que se visualiza como un valor vacío en la matriz de datos jamovi, se considera automáticamente un valor faltante (*missing*). Si se desea considerar valores concretos como faltantes, por ejemplo, para distinguir entre los motivos que han dado lugar a la falta de valor (e.g., utilizar los valores -9998 y -9999 en una variable cuantitativa para indicar que la falta de respuesta es debida a que el sujeto *desconoce la respuesta* o a que *no desea responder*, respectivamente), jamovi también permite indicar dichos valores para que luego se pueda decidir si se incluyen o no en los análisis estadísticos. Para ello, jamovi presenta el siguiente panel de opciones en el que se pueden especificar como faltantes valores individuales mediante el operador de comparación `==`, o bien rangos de valores mediante los operadores de comparación `>`, `>=`, `<` y `<=`, tal y como se ilustra en la siguiente captura de pantalla:

En las variables con escala de medida continua jamovi excluye de forma automática los valores faltantes a la hora de realizar los cálculos estadísticos. Los valores declarados como faltantes se visualizan atenuados en gris en la matriz de datos.

- **Etiquetas de valores (*Levels*):** en las variables con escala de medida nominal u ordinal se pueden indicar etiquetas para cada valor. También se pueden asignar etiquetas a valores concretos de variables con escala continua (e.g., los que definen los valores faltantes descritos en el punto anterior). En jamovi se pueden asignar estas etiquetas únicamente para los valores existentes en la variable; por tanto, la asignación de etiquetas se realiza una vez la variable ya contiene datos. Una vez se han definido las etiquetas, en la matriz de datos se muestra la etiqueta en lugar del valor. Por último, hay que señalar que la opción `Retain unused levels` que aparece en el panel de opciones de definición de etiquetas permite indicar si se desea que en los análisis de frecuencias

se muestren como categorías con 0 casos los valores para los cuales se ha definido una etiqueta pero que no existen entre los datos de la variable (ningún caso tiene esos valores).

Levels	
Hombre	0
Mujer	1

Consideraciones adicionales sobre la definición de las variables

Al definir la estructura de una matriz de datos, es decir, las variables que la componen, hay que tener presente las siguientes consideraciones previas:

- Es recomendable que la matriz incluya al menos una variable cuyo contenido permita identificar de forma única cada caso (fila). A este tipo de variables se les asigna generalmente la escala de medida ID en jamovi.
- La estructura inicial de la matriz de datos no debe contener variables redundantes, es decir, no debe incluir variables cuyos datos se puedan calcular o derivar lógicamente a partir de los valores de otras variables. Así, por ejemplo, no hay que definir la variable *edad* si se han definido las variables *fecha de nacimiento* y *fecha de respuesta*, o la variable *índice de masa corporal* si ya se han definido las variables *peso* y *talla* que definen su cálculo aritmético. jamovi permite ampliar posteriormente la matriz de datos mediante la inclusión de nuevas variables calculadas (este tipo de operación se revisará en detalle en el apartado 2). De este modo se evita cometer los errores que habitualmente se producen cuando se introduce información redundante.
- Es preferible definir variables que contengan los valores de las fechas y horas que variables que contengan tiempos transcurridos (e.g., edades). jamovi permitirá calcular exactamente el tiempo transcurrido entre dos fechas y/u horas, pudiendo escoger la unidad de medida del tiempo transcurrido resultante (días, meses, años, horas, minutos o segundos).
- Es preferible codificar las variables de tipo categórico de forma numérica. Es muy aconsejable codificar numéricamente las categorías de las variables categóricas en lugar de hacerlo mediante códigos de texto o alfanuméricos. Como en otros programas estadísticos, en jamovi algunos procedimientos estadísticos sólo se pueden aplicar sobre variables numéricas. Por ejemplo, la variable *Genero* es preferible definirla como numérica y codificar sus categorías, por ejemplo, con los valores 1 y 2 en lugar de H y M. La elección del código numérico para cada categoría es discrecional, pero se acostumbra a atender a los siguientes criterios:
 - Variables **categóricas binarias**: cuando la variable es de tipo No/Sí, Falso/Verdadero, Ausente/Presente, se acostumbra a asignar el código 0 a la categoría que representa la negación, falsedad o ausencia, y el código 1 a la otra categoría.
 - Variables **categóricas ordinales**: se asignan códigos numéricos enteros consecutivos siguiendo el orden de las categorías.
 - La **categoría de referencia**: cuando en una variable categórica existe una categoría que se puede considerar “de referencia”, es decir, que se prevé realizar los análisis comparando la/las otra/s categoría/s respecto a ella, es preferible asignar el código 0 a dicha categoría. Esto evitará tener que recodificar la variable previamente a la realización de algunos análisis estadísticos.

1.2. Ejercicio: definir la matriz de datos del *Cuestionario de hábitos tóxicos*

Como ejercicio para trabajar las propiedades de las variables expuestas en el apartado anterior, vamos a definir la estructura de la matriz de datos adecuada para registrar las respuestas a la siguiente encuesta sobre hábitos tóxicos:

CUESTIONARIO SOBRE HáBITOS TÓXICOS			
		CUESTIONARIO Nº: _____	
		PSEUDÓNIMO: _____	
FECHA:	de respuesta al cuestionario: ____/____/____		
	de nacimiento: ____/____/____	EDAD:	_____
GÉNERO:	() Masculino () Femenino		
PESO:	_____ kg.		
TALLA:	_____ cm.		
TABACO:	Si fumas:		
	¿Cuántos cigarrillos fumas, en promedio, cada día? _____		
	¿Qué marca de tabaco fumas habitualmente? _____		

ALCOHOL:	Si bebes alcohol:		
	¿Cuántos vasos de vino bebes, en promedio, cada semana? _____		
	¿Cuántas cervezas bebes, en promedio, cada semana? _____		
	¿Cuántas copas (coñac, ginebra, ron, etc.) y combinados (cuba-libre, etc.) bebes, en promedio, cada semana? _____		
DROGAS:	Si tomas algún tipo de droga:		
	<u>Tipo de droga</u>	<u>No tomo</u>	<u>Ocasionalmente</u> <u>Habitualmente</u>
	Marihuana	()	() ()
	Alucinógenos	()	() ()
	Opiáceos	()	() ()

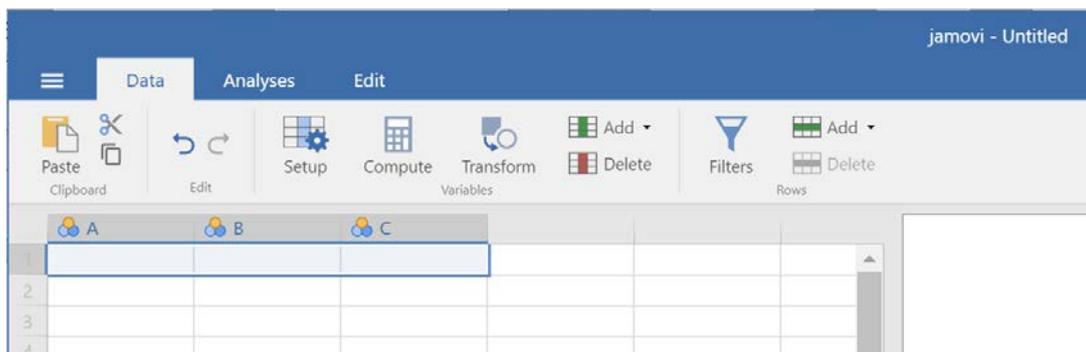
Nota: En el caso de la pregunta “marca de tabaco que fuma habitualmente”, a pesar de que las respuestas a este tipo de preguntas es recomendable recogerlas de forma textual en los cuestionarios en papel, durante la introducción de los datos se recomienda asignar un código numérico a cada respuesta posible. Por tanto, en la matriz de datos este tipo de variable se define como *numérica* en lugar de *cadena*.

1.3. Importar una matriz de datos

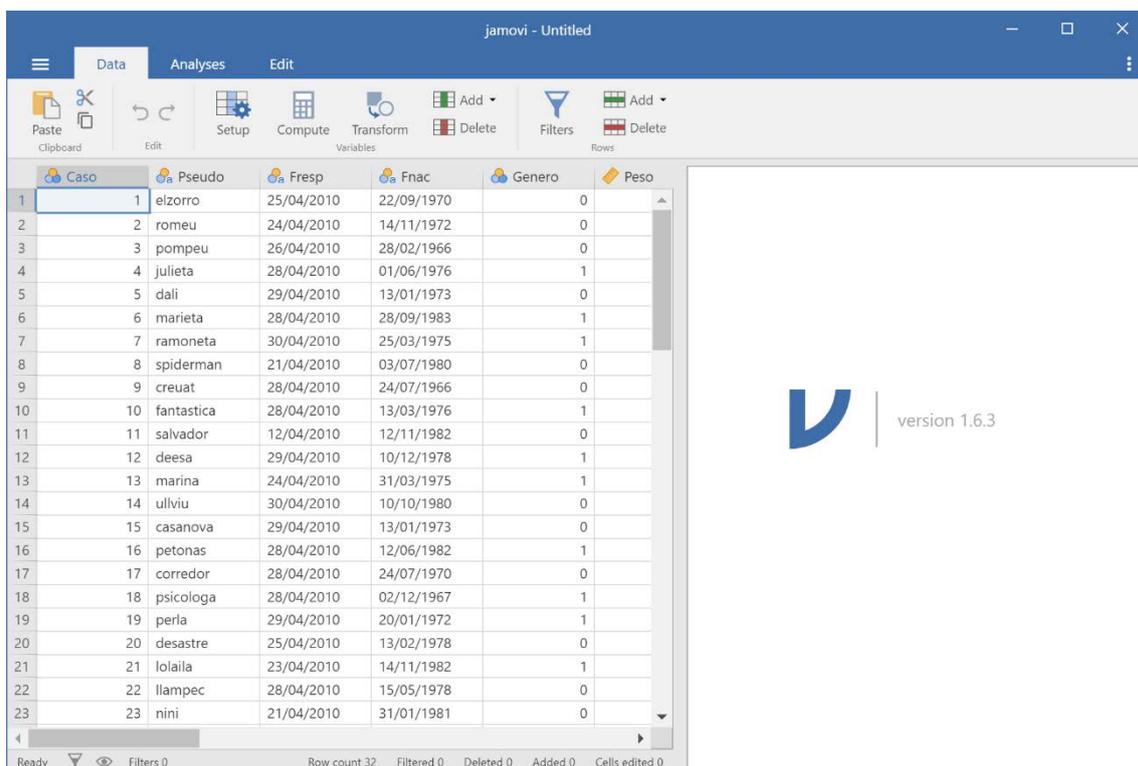
A pesar de que se pueden introducir los valores de la matriz de datos directamente en el panel de datos de jamovi, para facilitar esta tarea y para garantizar al máximo su calidad, es recomendable definir la matriz de datos e introducir la información utilizando, por ejemplo, un **gestor de bases de datos** o una **hoja de cálculo** para, posteriormente, importar ese fichero de datos desde jamovi.

Vamos a ilustrar el proceso de importación utilizando el fichero de hoja de cálculo MS-Excel *HABITOX.xlsx* que almacena los datos de las respuestas al *Cuestionario sobre hábitos tóxicos* de una muestra ficticia de 32 participantes de ese estudio:

1. Abre jamovi y selecciona el menú **Data**.
2. En el panel de datos aparecen predefinidas tres variables con los nombres A, B y C. Vamos a eliminarlas para dejar la matriz vacía antes de realizar la importación. Arrastra el cursor del ratón a lo largo de la primera fila para seleccionar las celdas de las tres variables.



3. A continuación, haz clic en **Delete** y confirma la eliminación de las tres columnas de datos.
4. Ejecuta la opción **Import** del menú principal  de jamovi y, a continuación, haz clic en el botón **Browse** que aparece en la parte superior del panel de opciones para localizar la carpeta en la que se encuentra el fichero *HABITOX.xlsx*. Haz clic sobre el nombre de este fichero y pulsa el botón **Abrir** para importar los datos que contiene:



jamovi mantiene el orden de las variables que tienen en la hoja de cálculo o fichero de datos exteno que se importa.

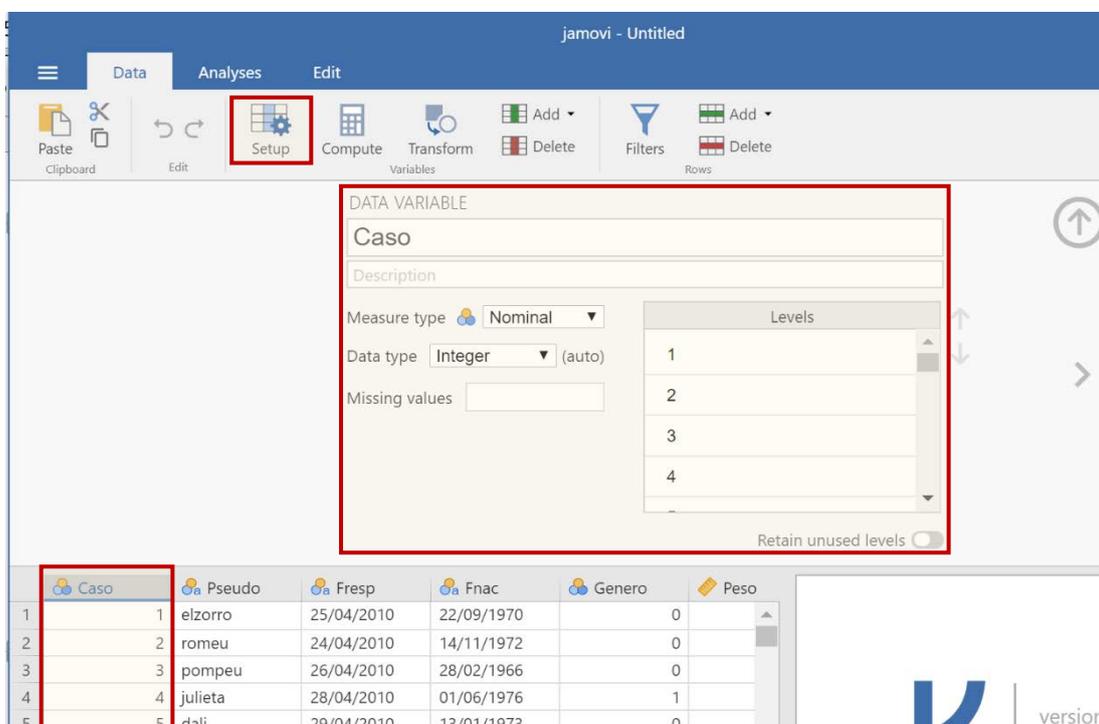
jamovi no incorpora todavía la opción para reordenar columnas de datos, pero se puede trasladar una variable a otra posición 1) insertando una nueva variable en esa posición, 2) copiando en el portapapeles el contenido de la variable que se quiere mover, 3) pegando el contenido copiado en la nueva variable insertada, 4) borrando la variable original, y 5) cambiando el nombre de la nueva variable insertada por el que tenía la variable original.

jamovi permite importar múltiples ficheros de datos de una vez, añadiendo secuencialmente sus filas.

1.4. Definir las propiedades de las variables

Una vez importada la matriz de datos desde el fichero de hoja de cálculo, vamos a modificar las propiedades de las variables de acuerdo el análisis de la estructura de la matriz de datos que hemos llevado a cabo con anterioridad a partir del *Cuestionario sobre hábitos tóxicos*.

1. Haz clic sobre la barra de título de la primera columna de la matriz (variable *Caso*) para seleccionarla y, a continuación, pulsa la opción de menú **Setup** para abrir el panel de opciones que permite definir las propiedades de esta variable:



jamovi ha predefinido esta variable como **nominal** respecto a su escala de medida y con valores numéricos **enteros** respecto al tipo de datos que almacena. Como irás comprobando, esta preasignación es la que jamovi realiza para todas las variables cuyo contenido sea de tipo numérico, excepto en aquellas que contengan decimales, que jamovi las predefine como variables con escala cuantitativa **continua** con tipo de datos **decimal**, y aquellas que contengan información alfanumérica, que predefine como **nominal** con datos de tipo **textual**.

2. Escribe la **Description** de la variable **Nº de cuestionario** y elige **ID** en **Measure type** para indicar que se trata de una variable de tipo **identificador** cuyo contenido se utilizará únicamente para identificar los casos de la matriz de datos.

DATA VARIABLE

Caso

Nº de cuestionario

Measure type ID

Data type Integer (auto)

Missing values

Levels

Retain unused levels

3. A continuación, haz clic sobre la barra de título de la segunda columna del panel de datos, o bien haz clic en el símbolo \triangleright que se encuentra a la derecha del panel de opciones, para pasar a definir las propiedades de la variable *Pseudo*. Escribe Pseudonimo en su Description y escoge ID en Measure type por tratarse también de una variable de tipo **identificador**.
4. Siguiendo el mismo procedimiento, abre el panel de opciones de la siguiente variable (*FResp*). En este caso se trata de una variable que almacena la Fecha de respuesta al cuestionario de cada participante del estudio. Tal y como ya se ha comentado en un apartado anterior, jamovi no implementa el tipo de dato **fecha**, por lo que considera esta variable como nominal textual.

Esto es así porque las celdas de esta variable en la hoja de cálculo que se ha importado tenían un formato *General* o bien *Texto*. Si en la hoja de cálculo estas celdas hubieran tenido un formato *Fecha* la importación las hubiera transformado en sendas variables cuantitativas discretas con el número de días transcurrido desde la fecha 01/01/1960 que jamovi, al igual que otros programas de análisis estadístico, considera como fecha inicial o “fecha 0”.

Dejaremos el Measure type y el Data type de esta variable tal y como están, y en el apartado 2.5 veremos cómo se puede extraer de esta variable los valores correspondientes al año, mes y día para utilizarlos en los cálculos estadísticos. Por tanto, para esta variable sólo hay que escribir su Description.

5. Avanza a la siguiente variable (*FNac*) y escribe Fecha de nacimiento en su Description.
6. Avanza a la siguiente variable (*Genero*). En este caso se trata de una variable **categoría nominal** que almacena los códigos numéricos correspondientes a sus dos categorías. Escribe Genero en su Description. En este caso Measure type y Data type ya están correctamente asignados con las opciones Nominal e Integer, respectivamente. Al tratarse de una variable categórica nominal debemos definir las etiquetas para sus posibles valores (categorías). Para ello, haz clic sobre el valor 0 que aparece en la lista Levels y escribe la etiqueta Masculino. A continuación, haz clic sobre el valor 1 de la lista Levels y escribe la etiqueta Femenino.

DATA VARIABLE

Genero

Género

Measure type Nominal

Data type Integer

Missing values

Levels

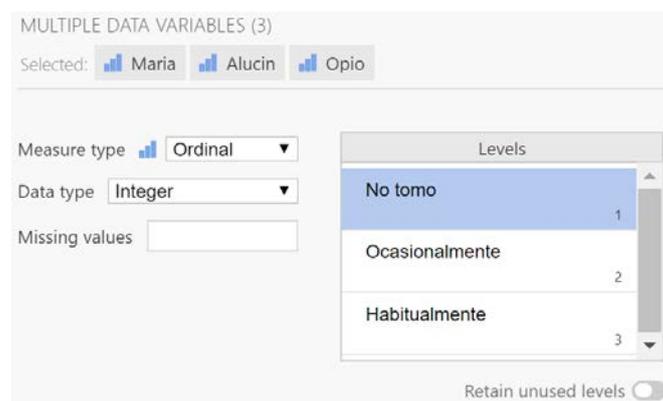
Masculino	0
Femenino	1

Retain unused levels

Estas flechas permiten cambiar el orden de las categorías. jamovi presentará los resultados de los análisis en el orden aquí establecido.

Este será el procedimiento que deberás seguir para asignar las etiquetas a los varlores de las variables de tipo categórico nominal u ordinal.

7. Avanza a la siguiente variable (*Peso*) y cambia sus propiedades siguiendo el procedimiento descrito. En este caso, al tratarse de una variable **cuantitativa continua** con valores decimales las propiedades **Measure Type** y **Data type** preasignadas por jamovi también son correctas. Recuerda que en las variables con escala cuantitativa es recomendable incluir la unidad de medida **Peso (Kg.)** en su **Description**. **Data type** ya está correctamente identificado como **Decimal**.
8. Pasa a la siguiente variable (*Talla*), escribe **Talla (cm.)** en su **Description** y cambia **Measure type** por **Continuous** al tratarse de una variable **cuantitativa discreta**. **Data type** ya está correctamente identificado como **Integer**.
9. Selecciona la siguiente variable (*NumCig*), escribe **Numero de cigarrillos (diario)** en su **Description**, y cambia **Measure type** por **Continuous** al tratarse de una variable **cuantitativa discreta**. **Data type** ya está correctamente identificado como **Integer**.
10. Sigue avanzando por el resto de las variables y cambiando sus propiedades de acuerdo con lo indicado en la solución del ejercicio (apartado 1.7) hasta que llegues a la variable *Maria*. Esta variable y las dos que le siguen (*Alucin* y *Opio*) son de tipo **categórico ordinal**. Escribe en cada una de ellas su **Description** y cambia su **Measure type** por **Ordinal**. Puesto que estas tres variables ordinales comparten las mismas etiquetas para sus posibles valores se puede realizar el proceso de etiquetación simultáneamente para todas ellas. Para ello, basta con seleccionar con el ratón las celdas de la primera fila de las tres variables y, a continuación, escribir las etiquetas de cada categoría en la lista **Levels**. jamovi titula el panel de opciones como **MULTIPLE DATA VARIABLES (3)** para indicar que se están definiendo las propiedades de tres variables a la vez.



11. Una vez definidas las propiedades de todas las variables de la matriz de datos, el siguiente paso es guardar el trabajo realizado. Tal y como ya se ha comentado, se puede guardar o bien la matriz de datos (en diversos formatos) con la opción **Export** del menú principal ☰, o bien la matriz y el contenido del panel de resultados en un fichero **.omv** con un formato particular de jamovi. Esta última opción es la recomendable cuando se desea finalizar la sesión de trabajo para continuarla en otro momento. Para crear este fichero elige la opción **Save As** del menú principal de jamovi.
12. En el panel de opciones haz clic en el botón **Browse** y, a continuación, selecciona la carpeta en la que desees almacenar el nuevo fichero.
13. Escribe el nombre del fichero **HABITOX_P1User** y pulsa **Guardar**. Puedes comparar el contenido y las propiedades de las variables del fichero **HABITOX_P1User.omv** que acabas de crear con la solución que se presenta en el fichero **HABITOX_P1Sol.omv**.

1.5. Guardar la estructura de la matriz de datos y de los resultados como plantilla (*template*)

jamovi incorpora también la posibilidad de guardar plantillas (*templates*). Una plantilla es un fichero particular de jamovi (con la extensión *.omt.*) que almacena las propiedades de las variables originales (*data variables*), de las nuevas variables de transformación (*transformed variables*) y calculadas (*computed variables*) contenidas en el panel de datos, junto con las órdenes que se hayan ejecutado para llevar a cabo los análisis estadísticos contenidos en el panel de resultados. A diferencia de los ficheros *.omv*, las plantillas no guardan los datos ni los resultados.

Cuando se abre una plantilla, el panel de datos muestra los nombres de las variables (originales y generadas) en los títulos de las columnas y las filas vacías, y el panel de resultados contiene sólo la estructura de los formatos de resultados de los análisis. Al importar una matriz de datos con la misma estructura (i.e., nombres de las variables originales) que tenía el panel de datos cuando se guardó la plantilla, jamovi genera automáticamente todas las variables de transformación y cálculo y ejecuta de nuevo todos los análisis con esos nuevos datos. El uso de plantillas es muy útil cuando se desea definir la matriz de datos y planificar los análisis que se llevarán a cabo antes de disponer de los datos finales. También son muy útiles cuando se prevé llevar a cabo un mismo bloque de análisis para diferentes matrices de datos que comparten una misma estructura.

Para almacenar la plantilla de la matriz de datos del *Cuestionario sobre hábitos tóxicos* que hemos definido en el apartado anterior debes seguir los siguientes pasos:

1. Elige la opción **Export** del menú principal  de jamovi.
2. Haz clic en **Browse** y, a continuación, selecciona la carpeta en la que deseas almacenar la plantilla.
3. Escribe el nombre del fichero **HABITOX_P1User**, elige **jamovi template (.omt)** en la lista desplegable y, finalmente, pulsa **Guardar**. Se guardará el fichero de plantilla jamovi *HABITOX_P1User.omt*.

1.6. Introducir manualmente nuevos casos a la matriz de datos

Se puede añadir manualmente más filas a la matriz de datos. Para ello basta con situar el cursor en una nueva fila del panel de datos e ir introduciendo los valores en cada celda.

Consideraciones durante la introducción de datos

- Para la introducción de los valores de las **variables categóricas** se debe escribir directamente el código numérico de la categoría y jamovi mostrará automáticamente la etiqueta correspondiente. Si se introduce un valor para el cual no se ha definido una etiqueta ésta no se mostrará y ello le ayudará a identificar posibles errores durante la entrada de datos.
- Las **fechas** y las **horas** hay que introducirlos como texto manteniendo constantes los delimitadores entre sus valores: / o – (barra o guion) entre día y mes y entre mes y año; : (dos puntos) entre hora y minutos y entre minutos y segundos.
- En los valores numéricos **decimales** debe utilizarse el punto (.) como símbolo de separación entre la parte entera y la parte decimal.
- Cuando **no se disponga de algún valor** simplemente hay que dejar la celda vacía.

1.7. Solución del ejercicio

Nombre	Etiqueta	Escala de medida	Tipo de datos	Valores faltantes	Etiquetas de valores
Caso	Nº de cuestionario	ID	Integer		
Pseudo	Pseudonimo	ID	Text		
FResp	Fecha de respuesta	Nominal	Text		
FNac	Fecha de nacimiento	Nominal	Text		
Genero	Genero	Nominal	Integer		0: Masculino 1: Femenino
Peso	Peso (Kg.)	Continuous	Decimal		
Talla	Talla (cm.)	Continuous	Integer		
NumCig	Numero de cigarrillos (diario)	Continuous	Integer		
Tabaco	Marca de tabaco	Nominal	Integer		1: Fortuna 2: Ducados negro 3: Ducados rubio etc.
Vino	Numero de vasos de vino (semanal)	Continuous	Integer		
Cerve	Numero de cervezas (semanal)	Continuous	Integer		
Copas	Numero de copas (semanal)	Continuous	Integer		
Maria	Marihuana (nivel de consumo)	Ordinal	Integer		1: No tomo 2: Ocasionalmente 3: Habitualmente
Alucin	Alucinogenos (nivel de consumo)	Ordinal	Integer		1: No tomo 2: Ocasionalmente 3: Habitualmente
Opio	Opiaceos (nivel de consumo)	Ordinal	Integer		1: No tomo 2: Ocasionalmente 3: Habitualmente

2. Transformación y cálculo de variables

Habitualmente, antes de iniciar el proceso de análisis estadístico es necesario crear nuevas variables a partir de las existentes en la matriz de datos. La creación de estas nuevas variables puede implicar desde operaciones de transformación de una variable cuantitativa en una variable ordinal agrupando rangos de valores, o realizar cálculos sencillos de diferencia entre dos fechas para obtener un tiempo transcurrido (como, por ejemplo, la edad en años cumplidos a partir de la fecha de nacimiento y la fecha de respuesta a un cuestionario), hasta complejas operaciones aritméticas o lógicas a partir de un gran número de variables originales (por ejemplo, para definir un diagnóstico DSM o CIE a partir de un conjunto de síntomas e indicadores).

jamovi incluye herramientas para la creación de nuevas variables en su menú **Data**. Como veremos, en la mayor parte de casos estas operaciones se podrán realizar mediante paneles de opciones que evitan tener que aprender sintaxis complejas. En otros casos aprender dicha sintaxis será inevitable para asegurar que la creación de la nueva variable se realiza correctamente, teniendo en cuenta contingencias como la presencia de valores faltantes (*missing*) o expresiones lógicas complejas.

2.1. Recodificar variables

Una de las operaciones habituales de transformación de datos consiste en cambiar la escala de medida de una variable cuantitativa continua o discreta para crear una nueva variable categórica ordinal, o bien asignar códigos numéricos a los valores de una variable alfanumérica (de tipo textual).

Para ilustrar el procedimiento de recodificación de variables con jamovi, vamos a crear la nueva variable nivel de consumo de tabaco (variable *Fumador*) a partir del número de cigarrillos que los sujetos fuman en promedio cada día (*NumCig*). Para ello establecemos previamente los siguientes criterios:

<i>NumCig</i>	<i>Fumador</i>	
Valor	Valor	Etiqueta
0	0	<i>No fumador</i>
1 a 4	1	<i>Menos de 5 cigarrillos/día</i>
5 a 9	2	<i>Menos de 10 cigarrillos/día</i>
Mayor o igual a 10	3	<i>10 o más cigarrillos/día</i>

1. Abre el fichero *HABITOX_P1Sol.omv* que contiene la matriz de datos resultante de la ejecución del apartado 1.4, con la opción *Open* del menú principal  de jamovi, o bien, haz doble clic sobre este fichero en el explorador de Windows (esta acción abre jamovi y carga el fichero de forma automática).
2. Selecciona la columna de la variable *NumCig* que vamos a recodificar y ejecuta *Data | Transform*.

Recuerda que en jamovi se pueden seleccionar un grupo consecutivo de columnas en el panel de datos haciendo clic sobre la primera de ellas y luego clic en la última mientras se mantiene pulsada la tecla Mayús, o bien un grupo de columnas no consecutivas haciendo clic sobre ellas mientras se mantiene pulsada la tecla Ctrl.

3. Escribe *Fumador* como nombre para la nueva variable recodificada y *Nivel de consumo de cigarrillos (diario)* como *Description*.
4. Despliega la lista *using transform* y elige la opción *Create New Transform...* Se abre un nuevo panel de opciones en el que se pueden introducir las reglas de transformación y también definir la escala de medida de la nueva variable, tal y como muestra la siguiente imagen. Escribe *Recodificación NumCig* como nombre para la nueva transformación que definiremos a continuación, *Nivel de consumo de cigarrillos/pipas* como descripción y, finalmente, elige *Ordinal* en la lista *Measure Type*.

Cuando la transformación se realiza simultáneamente sobre un conjunto de variables, el campo *Variable suffix* del panel de opciones permite asignar nombres automáticos a las nuevas variables generadas. Los siguientes ejemplos ilustran cómo funcionan los sufijos que se indiquen en este campo:

- *_R* → *nombre_variable_R*
- *R_...* → *R_nombre_variable*

Los tres puntos (...) se reemplazarán por el nombre de cada una de las variables originales para crear automáticamente los nombres de las nuevas variables transformadas.

5. A continuación, introduce las reglas de recodificación que se muestran en la siguiente captura de pantalla para definir las categorías de la nueva variable *Fumador* a partir del contenido de la variable *NumCig*. Hay que ir haciendo clic en *+ Add recode condition* para añadir nuevas filas para definir las reglas de transformación. Nota: actualmente jamovi permite asignar únicamente los códigos, o bien las etiquetas, de las categorías de la nueva variable. En este ejercicio asignaremos directamente las etiquetas, que deben escribirse entre comillas (").



Las reglas (condiciones) que se indican en el panel de opciones de transformación se ejecutan para cada fila de datos en el orden en que se indican. Esto quiere decir que, para cada fila de datos, jamovi deja de evaluar las restantes reglas cuando la información contenida en las variables originales cumple la regla que está evaluando en ese momento.

Cuando se está recodificando una variable **cuantitativa** hay que asegurarse que los intervalos que definen las reglas recogen todos los valores posibles de la variable original, es decir, son **exhaustivas**. En el ejemplo anterior, esto se logra por el hecho de indicar ≥ 10 en la última regla, en lugar de indicar > 10 , que dejaría fuera de la recodificación el valor 10. Si la variable *NumCig* fuera **cuantitativa continua** en lugar de discreta, un patrón general para realizar la transformación asegurando su exhaustividad sería el siguiente (nótese que se va indicando el límite superior de cada intervalo usando el operador de comparación \leq y se utiliza el operador $>$ para definir el límite superior del último intervalo):

- If $\$source == 0$ use "No fuma"
- If $\$source \leq 4$ use "1 a 4 cigarrillos diarios"
- If $\$source \leq 9$ use "5 a 9 cigarrillos diarios"
- If $\$source > 9$ use "10 o mas cigarrillos diarios"
- else use NA

Por último, señalar que en la mayoría de casos, es recomendable dejar vacía la última regla de transformación (else), o bien indicar else use NA para indicar que se desea que la nueva variable transformada contenga un valor faltante (*missing*) para todos aquellos casos cuyos valores en la variable original no cumplan ninguna de las reglas definidas. Los valores faltantes en las variables originales quedan automáticamente asignados también como valores faltantes (NA) en las nuevas variables transformadas, a no ser que se incluya una regla de transformación específica en la que se asignen los NA a un valor determinado (e.g., If $\$source == NA$ use "Sin respuesta").

Cómo comprobar si las reglas de transformación funcionan correctamente

A medida que se escriben las reglas de transformación se puede (de hecho, se debe) ir comprobando que funcionan correctamente observando en el panel de datos cómo jamovi las va aplicando:

Numcig	Fumador
8	5 a 9 cig/d
5	5 a 9 cig/d
1	1 a 4 cig/d
20	10 o más cig/d
8	5 a 9 cig/d
11	10 o más cig/d
8	5 a 9 cig/d
15	10 o más cig/d

6. Cierra el panel de opciones de recodificación pulsando las flechas \downarrow y \uparrow que se encuentran en la parte superior derecha del panel.
7. Finalmente, guarda los cambios realizados en la matriz de datos ejecutando la opción **Save** del menú principal \equiv de jamovi.

2.2. Calcular variables mediante expresiones aritméticas y estadísticas

Otra operación habitual de preparación de los datos consiste en crear nuevas variables como resultado de cálculos aritméticos y estadísticos a partir de los valores de variables existentes en la matriz de datos original. El menú **Data** | **Transform** de jamovi permite realizar esta operación.

Elementos de una expresión jamovi

Para facilitar la escritura de las expresiones, jamovi incluye en los paneles de opciones de transformación un conjunto de botones y de listas de funciones que al pulsarlos insertan en la expresión el elemento correspondiente:

- **Valores constantes:**
 - Números: se escriben directamente (ejemplo, 34).
 - Alfanuméricos: se escriben entre comillas ("valor").
 - Valores faltantes: se escribe NA.
- **Nombres de variables:** jamovi distingue entre mayúsculas y minúsculas (ejemplo, PESO, Peso).
- **Operadores aritméticos** (de mayor a menor orden de precedencia, es decir, de ejecución):
 - ^ (potencia)
 - / (división)
 - * (multiplicación)
 - - (resta)
 - + (suma)
- **Operadores de comparación (relacionales):**
 - == (igual que)
 - != (distinto de)
 - > (mayor que)
 - < (menor que)
 - <= (menor o igual que)
 - >= (mayor o igual que)
- **Operadores (nexos) lógicos** (de mayor a menor orden de precedencia):
 - NOT () (negación lógica; implementada como una función en jamovi)
 - and (y lógico; debe escribirse en minúsculas)
 - or (o lógico; debe escribirse en minúsculas)
- **Funciones.** jamovi incluye funciones para realizar diversos tipos de operaciones matemáticas, estadísticas, lógicas, de texto, de conversión de tipos de datos, etc. En el caso concreto de las funciones estadísticas, éstas se pueden realizar sobre una fila de datos, o bien, sobre los valores contenidos en una columna (en este último caso las funciones añaden una V inicial a su nombre). Se puede consultar la lista completa de funciones, así como de su sintaxis, haciendo clic en el botón f_x del panel de opciones de transformación y generación de variables. La siguiente tabla muestra algunos ejemplos de funciones matemáticas:

	function	example input	(answer)
square root	SQRT(x)	SQRT(25)	5
absolute value	ABS(x)	ABS(-23)	23
logarithm (base 10)	LOG10(x)	LOG10(1000)	3
logarithm (base e)	LN(x)	LN(1000)	6.908
exponentiation	EXP(x)	EXP(6.908)	1000.245
box-cox	BOXCOX(x, lamda)	BOXCOX(6.908, 3)	109.551

Orden de ejecución de los elementos de una expresión: uso de paréntesis

Como en cualquier otro programa estadístico o matemático, cuando se incluyen paréntesis en una expresión, jamovi evalúa en primer lugar la parte de la expresión que se encuentra entre los paréntesis. De esta manera se puede romper el orden de ejecución (de precedencia) de los operadores, orden que por defecto es el siguiente:

- (1) Funciones.
- (2) Operadores aritméticos.
- (3) Operadores de comparación (relacionales).
- (4) Operadores (nexos) lógicos.
- (5) Variables y valores constantes.

Comportamiento de los valores faltantes en las operaciones aritméticas y estadísticas

Cuando una expresión de cálculo utiliza operadores aritméticos, si alguno de los argumentos contiene un valor faltante (*missing*) el resultado será también un valor faltante.

Por el contrario, las funciones estadísticas `SUM()`, `MEAN()`, `STDEV()` y `VAR()` incluyen el argumento `ignore_missing=1` que permite indicar que retornen el resultado aunque los argumentos contengan valores faltantes. Además, la función `SUM()` tiene la particularidad de que retorna como resultado el valor 0 en el caso que todas las variables que se sumen contengan valores faltantes. Como se verá en el siguiente ejercicio, este comportamiento en general no es adecuado, por lo que se requerirá incluir también en la sintaxis de esta función el argumento `min_valid=` para indicar el número mínimo de valores válidos necesarios para que se lleve a cabo la suma; en caso contrario se retornará el valor faltante.

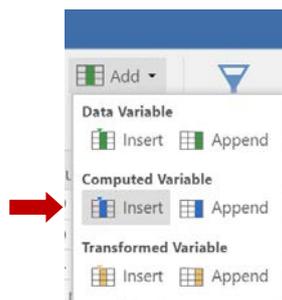
Vamos a crear la **variable calculada** *IMC* (índice de masa corporal) para ilustrar el proceso de creación con jamovi de una variable calculada (**computed variable**) con base en una expresión aritmética:

1. Selecciona la columna de la variable *NumCig* para insertar a su izquierda una nueva columna para la variable calculada *IMC*. De este modo la nueva variable *IMC* se situará en el panel de datos tras la variable *Talla*. Esta ubicación permitirá comprobar visualmente si se realiza correctamente el cálculo del *IMC* como resultado de dividir el *Peso* en kg. de los participantes por su *Talla* en metros al cuadrado. Hay que tener en cuenta que la *Talla* está registrada en cm., por lo que hay que pasarla a metros dividiendo su valor por 100. La expresión de cálculo del *IMC* sería la siguiente:

$$IMC = \text{Peso} / ((Talla / 100) ^ 2)$$

En esta expresión se puede observar también la importancia del uso de paréntesis para que jamovi realice correctamente la operación de cálculo. En este caso, el paréntesis $(Talla / 100)$ es necesario y el paréntesis exterior $((Talla / 100) ^ 2)$, aunque no es necesario, es recomendable incluirlo para hacer más explícito el orden en que se ejecutarán las operaciones incluidas en la expresión:

1. $(Talla / 100)$ para pasar la *Talla* de centímetros a metros.
 2. $((Talla / 100) ^ 2)$ para elevar al cuadrado la *Talla* en metros.
 3. $\text{Peso} / ((Talla / 100) ^ 2)$ para dividir el *Peso* (Kg.) por la *Talla* (m.) y obtener el *IMC*.
2. A continuación, ejecuta **Data | Add | Computed Variable | Insert** para insertar la nueva columna y abrir el panel de opciones para introducir las propiedades de la nueva variable *IMC*.



3. Escribe **IMC** como nombre para la nueva variable y **Índice de Masa Corporal** como **Description**. A continuación, escribe la expresión de cálculo del *IMC* tal y como se muestra en la siguiente captura de pantalla y pulsa **Entrar**.

Una vez escrita o modificada una expresión de cálculo, para que jamovi la ejecute hay que pulsar la tecla **Entrar**, o bien hacer clic fuera del recuadro de la expresión del panel de opciones. De este modo se puede comprobar en el panel de datos si el resultado de la expresión es correcto.

Jamovi opera y almacena los resultados numéricos con la máxima precisión (16 decimales), pero muestra por pantalla sólo el valor redondeado a tres cifras decimales.

Peso	Talla	IMC
76.4	178	24.113
85.2	179	26.591
78.6		
80.6	168	28.557
76.3	175	24.914
71.5	165	26.263

4. Verifica en algunas de las filas del panel de datos que la nueva variable contiene efectivamente el resultado correcto del cálculo del índice de masa corporal, con valores que oscilaran alrededor de los 20 puntos en los casos con una adecuada relación entre el peso y la talla, y con valores superiores a 25 en los casos con un peso superior al que correspondería a su talla. Observa que, si las variables *Peso* y/o *Talla* presentaban datos faltantes, la nueva variable *IMC* aparece como valor faltante, dado que en jamovi las expresiones de cálculo aritmético devuelven como resultado el valor faltante cuando no se dispone de valor en alguna de las variables con las que se opera.
5. Cierra el panel de opciones de creación de variables calculadas haciendo clic en la flecha **↑** que se encuentra en la parte superior derecha del panel.
6. Ejecuta la opción **Save** del menú principal **≡** de jamovi para guardar los cambios realizados.

Vamos a crear ahora la **variable calculada** *NumBebidas* que contenga el número de unidades de bebidas alcohólicas que consumen semanalmente los participantes del estudio. La expresión de cálculo de esta nueva variable permitirá ilustrar la diferencia entre el uso de **operadores aritméticos** y **funciones** en jamovi, en especial por lo que respecta al diferente manejo que realizan de los **datos faltantes**.

7. Selecciona la columna de la variable *Maria* para insertar a su izquierda una nueva columna para la variable calculada *NumBebidas*. De este modo la nueva variable se situará en el panel de datos tras la variable *Copas*. Esta ubicación permitirá comprobar visualmente si se realiza correctamente el cálculo de *NumBebidas* como resultado de sumar el consumo de *Vino*, *Cerve* y *Copas*.
8. A continuación, ejecuta **Data | Add | Computed Variable | Insert** para insertar la nueva columna y abrir el panel de opciones que permite introducir las propiedades de la nueva variable *NumBebidas*.
9. En el panel de opciones que se abre, escribe *NumBebidas* como nombre para la nueva variable calculada, **Numero de unidades de bebidas alcoholicas consumidas semanalmente** como

Description, y la expresión de cálculo $\text{Vino} + \text{Cerve} + \text{Copas}$ en el recuadro Formula. Una vez escrita la expresión, pulsa Entrar para que jamovi ejecute la expresión y poder comprobar sus resultados en la columna *NumBebidas* del panel de datos.

Observa que algunas de las filas contienen el valor faltante en la nueva variable *NumBebidas*. Se trata de las filas que contienen un valor faltante en al menos una de las tres variables incluidas en la expresión de cálculo. Esto es así porque, como ya se ha indicado en el cuadro de información de este apartado, los **operadores aritméticos** devuelven como resultado el valor faltante en aquellas filas con valor/es faltantes en que alguna/s de las variables incluidas en la expresión.

COMPUTED VARIABLE

NumBebidas

Número de unidades de bebidas alcohólicas consumidas semanalmente

Formula

$f_x = \text{Vino} + \text{Cerve} + \text{Copas}$

Retain unused

Vino	Cerve	Copas	NumBebidas
2	6	1	9
2	5	0	7
4	16	3	23
3	7	4	14
6	5	3	14
5	12	0	17
8	5	4	17
1	10	2	13
0	0	4	4
5	8	0	13
0	3	5	8
7	0	0	7
2	5	10	17
0	12	3	15
-	-	0	0

10. Ejecuta la opción Save del menú principal de jamovi para guardar los cambios realizados.
11. Para obtener el valor de la expresión de cálculo anterior aún en el caso que alguna de las variables *Vino*, *Cerve* o *Copas* contuviera valores faltantes, en lugar de utilizar operadores aritméticos (en este caso la suma +) en la expresión de cálculo, se deben utilizar las **funciones estadísticas**. Vamos a ilustrar el uso de este tipo de funciones para crear la variable *NumBebidas2*. Selecciona la columna de la variable *Maria* para insertar a su izquierda una nueva columna.
12. A continuación, ejecuta Data | Add | Computed Variable | Insert para insertar la nueva columna y abrir el panel de opciones que permite introducir las propiedades de la nueva variable *NumBebidas2*.
13. En el panel de opciones que se abre, escribe *NumBebidas2* como nombre para la nueva variable calculada y *Numero de unidades de bebidas alcohólicas consumidas semanalmente* que han sido informadas por los participantes como Description.
14. En este caso, para introducir la expresión de cálculo en el recuadro Formula vamos a utilizar el panel de opciones extendido, que nos dará acceso a la lista de funciones matemáticas y estadísticas que incluye jamovi. Haz clic en el botón f_x y, a continuación, localiza en la lista Functions la función SUM() del bloque **Statistical** y haz doble clic sobre ella para que jamovi la escriba en el recuadro de la

expresión de cálculo. Observa en la parte inferior del panel de opciones que la sintaxis de la función SUM() incluye el argumento ignore_missing= para indicar si se desea que la suma se lleve a cabo (ignore_missing=1) o no (ignore_missing=0) en el caso que las variables que se suman contengan algún valor faltante. Escribe la expresión SUM(Vino, Cerve, Copas, ignore_missing = 1) y pulsa Entrar para que jamovi ejecute la expresión y poder comprobar sus resultados en la columna NumBebidas2 del panel de datos.

The screenshot shows the 'COMPUTED VARIABLE' panel in Jamovi. The variable name is 'NumBebidas2' and its description is 'Número de unidades de bebidas alcohólicas consumidas semanalmente que han sido informadas por los participantes'. The formula entered is '= SUM(Vino, Cerve, Copas, ignore_missing = 1)'. Below the formula, there are lists for 'Functions' and 'Variables'. The 'Functions' list includes MEAN, MIN, SCALE, STDEV, SUM, and VMAX. The 'Variables' list includes Caso, Pseudo, Fresp, Fnac, Genero, and Peso. The data table below shows columns for 'Vino', 'Cerve', 'Copas', 'NumBebidas', and 'NumBebidas2'. Two rows are highlighted with red dashed boxes and arrows, showing that the result in 'NumBebidas2' is 0 for those rows, despite missing values in the input variables.

Vino	Cerve	Copas	NumBebidas	NumBebidas2
2	5	0	7	7
4	16	3	23	23
3	7	4	14	14
6	5	3	14	14
5	12	0	17	17
8	5	4	17	17
1	10	2	13	13
0	0	4	4	4
5	8	0	13	13
0	3	5	8	8
7	0	0	7	7
2	5	10	17	17
0	12	3	15	15
7	3	0	10	10

Observa que las dos filas señaladas con una flecha contienen el valor 0 para NumBebidas2 en lugar del valor faltante que se obtuvo cuando se realizó la suma de las tres variables mediante el operador aritmético +. Es decir, la función SUM() devuelve como resultado la suma de los valores de las variables Vino, Cerve y Copas aunque éstas contengan valores faltantes.

15. Observa que el resultado también es 0 en la primera fila señalada en que todas las variables que se suman contienen un valor faltante. Generalmente es más correcto obtener un valor faltante (NA) en este caso. Para conseguir este resultado hay que modificar la expresión de cálculo utilizada en el paso anterior, añadiendo el argumento min_valid= para indicar que la suma se lleve a cabo sólo en el caso que el número mínimo de valores válidos en las variables que se suman sea igual al que se indique en este argumento. Así, por ejemplo, si se desea que la suma sólo se realice en el caso que al menos una de las variables que se suman contenga un valor válido (i.e., distinto de faltante) habría que indicar min_valid=1 como argumento final en la función SUM().

Cambia la expresión de cálculo de NumBebidas2 por la siguiente, y comprueba en el panel de datos que la primera fila señalada con la flecha en la figura anterior contiene ahora un valor faltante, mientras que la segunda fila señalada continúa manteniendo el valor 0 por disponerse de un valor válido (0 en este caso) en la variable Copas, a pesar de que Vino y Cerve contengan valores faltantes:

SUM(Vino, Cerve, Copas, ignore_missing = 1, min_valid=1)

16. Ejecuta la opción Save del menú principal ≡ de jamovi para guardar los cambios realizados.

2.3. Calcular variables mediante expresiones lógicas

Otro tipo de variable calculada que habitualmente es necesario definir es aquella cuyo resultado no se obtiene a partir de un cálculo aritmético o estadístico, sino de una expresión lógica. Este tipo de variables también se pueden crear en jamovi utilizando la función `IF()`.

Un ejemplo típico de este tipo de variable calculada lo constituyen los **diagnósticos** (como, por ejemplo, la hipertensión o la obesidad según los criterios de la OMS, el trastorno bipolar según el DSM-IV, etc.) o, desde un punto de vista más general, los sistemas de clasificación.

Estructura de una expresión lógica simple

Una expresión lógica simple está compuesta por tres elementos:

A `operador_de_comparación` B

Siendo A y B valores, variables o expresiones numéricas. Ejemplo:

TALLA > 150

Resultado de una expresión lógica simple

Una expresión lógica devuelve como resultado:

- El valor 1 (*verdadero*) si la expresión es cierta.
- El valor 0 (*falso*) si la expresión es falsa.
- El valor NA (*faltante -missing-*) si la expresión no se puede evaluar por falta de información.

Estructura de una expresión lógica compleja

Una expresión lógica compleja está compuesta por dos o más expresiones lógicas simples concatenadas mediante operadores (nexos) lógicos:

(A `operador_de_comparación` B) `nexo_lógico` (C `operador_de_comparación` D)

Siendo A, B, C y D valores, nombres de variables o expresiones de cálculo numérico o estadístico. ¡Atención! Los nexos lógicos tienen un orden de precedencia (de ejecución), de modo que en una expresión que contenga varios nexos lógicos primero se ejecutan los *Y lógicos* (*and*) y después los *O lógicos* (*or*). El **uso de los paréntesis** es, por tanto, imprescindible para variar el orden de ejecución de los nexos lógicos que incluye la expresión. jamovi implementa el nexo lógico *NO* como una función `NOT()`, por lo que se evalúa siempre en primer lugar. Ejemplos:

Vino > 0 **and** Cerve > 0 **and** Copas > 0
Genero == 0 **and** IMC >= 25 **or** Genero == 1 **and** IMC >= 24

Resultado de una expresión lógica compleja

El nexo lógico **and** devuelve *verdadero* cuando se cumplen las dos condiciones lógicas simples que concatena y *falso* si al menos una de ellas es falsa. En el caso de valores faltantes (NA):

- (NA **and** NA) → NA
- (NA **and** *verdadero*), o bien, (*verdadero* **and** NA) → NA
- (NA **and** *falso*), o bien, (*falso* **and** NA) → *falso*

El nexo lógico **or** devuelve *verdadero* cuando se cumple al menos una de las dos condiciones lógicas simples que concatena. En el caso de valores faltantes (NA):

- (NA **or** NA) → NA
- (NA **or** *verdadero*), o bien, (*verdadero* **or** NA) → *verdadero*
- (NA **or** *falso*), o bien, (*falso* **or** NA) → NA

Construcción de un sistema de clasificación

Cuando se desea crear una variable que es el resultado de un sistema de clasificación (por ejemplo, un diagnóstico), es útil construir una tabla como la que se muestra a continuación, que contenga en columnas los antecedentes y una columna final con el valor consecuente:

Antecedentes		Consecuente	
V1	V2	V3	
valor1 →	valor3 →	1	Categoría 1
	valor4 →	2	Categoría 2
	NA →	NA	
valor2 →	valor5 →	1	Categoría 1
	valor6 →	2	Categoría 2
	NA →	NA	
NA →	valor7 →	1	Categoría 1
	valor8 →	2	Categoría 2
	NA →	NA	

Siendo *valorX* un valor, una lista o rango de valores, o bien, una expresión de cálculo numérico.

Como se puede observar, se trata de un diagrama en árbol en forma de tabla que permite indicar los *valores* (o *expresiones*) de cada variable (*antecedente*) que definen cada posible valor de la nueva variable (*consecuente*). Es importante incluir siempre en esta tabla de análisis la posibilidad de que tanto los antecedentes como el consecuente puedan contener **valores faltantes**.

Para transformar esta tabla de análisis en la expresión lógica que crea el sistema de clasificación sólo hay que aplicar la siguiente regla:

- Los *valores* que se encuentran en una misma fila (o rama) se concatenan con el nexo lógico **and**.
- Las filas (o ramas) que definen un mismo valor en el *consecuente* (variable que contendrá la clasificación) se concatenan con el nexo lógico **or**.

Así, para la tabla de análisis de antecedentes y consecuente genérica presentada antes, las expresiones lógicas necesarias para crear la nueva variable de clasificación serían (los paréntesis no serían necesarios, pero se incluyen para hacer más clara la expresión e identificar las expresiones resultantes de cada fila de la tabla de análisis anterior):

$(V1==valor1 \text{ and } V2==valor3) \text{ or } (V1==valor2 \text{ and } V2==valor5) \text{ or } (V2==valor7) \rightarrow V3=1$

$(V1==valor1 \text{ and } V2==valor4) \text{ or } (V1==valor2 \text{ and } V2==valor6) \text{ or } (V2==valor8) \rightarrow V3=2$

Por tanto, **para crear un sistema de clasificación hay que indicar tantas expresiones lógicas como categorías tenga dicho sistema.**

Vamos a crear la variable *Fuma* con los valores "fuma" o "no fuma" a partir del consumo diario de cigarrillos contenido en la variable *NumCig*. La variable *Fuma* se podría crear también siguiendo el procedimiento de recodificación descrito en el apartado 1, pero nos servirá aquí para ilustrar con un ejemplo muy sencillo la creación de variables calculadas con base en una expresión lógica.

1. Selecciona la columna de la variable *Tabaco* para insertar a su izquierda una nueva variable.
2. A continuación, ejecuta **Data | Add | Computed Variable | Insert** para insertar la columna y abrir el panel de opciones que permite introducir las propiedades de la nueva variable *Fuma*.
3. En el panel de opciones que se abre, escribe *Fuma* como nombre para la nueva variable calculada, *Fumador (Si/No)* como *Description*, y la siguiente expresión en el recuadro *Formula*:

`IF(NumCig > 0, "Si", "No")`

- Una vez escrita la expresión, pulsa **Entrar** para que jamovi ejecute la expresión y poder comprobar sus resultados en la columna *Fuma* del panel de datos. Observa que algunas de las filas contienen el valor faltante en la nueva variable *Fuma*. Se trata de las filas que contienen un valor faltante en la variable antecedente *NumCig*. Esto es así porque, como ya se ha indicado en el cuadro de información de este apartado, una **expresión lógica** devuelve como resultado el valor faltante cuando no se dispone de información suficiente para evaluarla.

COMPUTED VARIABLE

Fuma

Fumador (Si/No)

Formula f_x = IF(NumCig > 0, "Sí", "No")

Antecedente Consecuente Retain unused levels

NumCig	Fumador	Fuma	Tabaco	Vino
0	No fumador	No		
4	1 a 4 cig/d	Sí	West	7
0	No fumador	No		0
10	10 o más cig/d	Sí	Marlboro	4
10	10 o más cig/d	Sí	Winston	2
0	No fumador	No		0
0	No fumador	No		0
0	No fumador	No		5
15	10 o más cig/d	Sí	Ducados negro	8

- Ejecuta la opción **Save** del menú principal  de jamovi para guardar los cambios realizados.

Vamos a crear ahora una nueva variable que contenga el diagnóstico de obesidad a partir de la siguiente tabla de antecedentes (*Genero* e *IMC*) y consecuente (*Obesidad*), según la Sociedad Española para los Estudios de la Obesidad (SEEDO):

Antecedentes		Consecuente	
Genero	IMC	Obesidad	
Valor	Valor	Valor	Etiqueta
0 (masculino)	< 25	1	Normopeso
	>= 25 and < 30	2	Sobrepeso
	>= 30	3	Obesidad
	NA	NA	
1 (femenino)	< 24	1	Normopeso
	>= 24 and < 29	2	Sobrepeso
	>= 29	3	Obesidad
	NA	NA	
NA	< 24	1	Normopeso
	>= 25 and <29	2	Sobrepeso
	>= 30	3	Obesidad
	(>= 24 Y < 25) or (>= 29 Y < 30) or NA	NA	

2.4. Calcular recuentos

Otro tipo de variables calculadas que habituales son las que se definen a partir del recuento del número de sucesos o eventos que se producen en un conjunto de variables. Un suceso puede ser un valor, una lista y/o un rango de valores.

Vamos a crear una nueva variable, de nombre *NumDrogas*, que contendrá el número de drogas que consumen (ya sea habitual o bien ocasionalmente) los participantes que han respondido al *Cuestionario sobre hábitos tóxicos*.

1. Ejecuta **Data | Add | Computed Variable | Append** para añadir al final de la matriz de datos la nueva columna y abrir el panel de opciones que permite introducir las propiedades de la nueva variable *NumDrogas*.
2. En el panel de opciones que se abre, escribe **NumDrogas** como nombre para la nueva variable calculada, **Numero de drogas que consume (0 a 3)** como **Description**, y la siguiente expresión en el recuadro **Formula**:

$$\text{IF}(\text{Maria} > 1, 1, 0) + \text{IF}(\text{Alucin} > 1, 1, 0) + \text{IF}(\text{Opio} > 1, 1, 0)$$

Observa que la expresión es simplemente la suma de los resultados de la evaluación mediante la función **IF()** de las respuestas sobre la frecuencia de consumo de cada una de las tres drogas (*Maria*, *Alucin* y *Opio*) sobre las cuales se interroga en el cuestionario. Cada condición **IF()** retorna el valor **1** si la respuesta sobre el consumo de la droga ha sido “Ocasionalmente” (codificada con el valor **2**) o “Habitualmente” (codificada con el valor **3**), o bien retorna el valor **0** en caso contrario (i.e., cuando la respuesta ha sido “No tomo”, codificada con el valor **1**).

3. Ejecuta la opción **Save** del menú principal **≡** de jamovi para guardar los cambios realizados.

2.5. Calcular tiempos transcurridos entre fechas y horas

Como ya se comentó en el apartado 1.1, jamovi no incluye los tipos de datos **fecha** y **hora**. Las variables *FResp* y *FNac* se importaron en el procedimiento 1 como variables de tipo alfanumérico desde la hoja de cálculo *HABITOX.xlsx* porque las celdas de estas variables tenían un formato *General* o bien *Texto*. Si las celdas de estas variables hubieran tenido el formato *Fecha* en la hoja de cálculo *HABITOX.xlsx*, entonces la importación las hubiera transformado en sendas variables cuantitativas discretas con el número de días transcurridos desde la fecha que jamovi considera como fecha de referencia o “fecha 0”, el 01/01/1960. En este último caso, para calcular el tiempo transcurrido entre *FResp* y *FNac* bastaría con realizar una simple resta entre las mismas, y dividir el resultado por la cantidad correspondiente para transformar el número de días en la unidad temporal deseada (365.25 para años, 30 para meses o 7 para semanas):

- $\text{Int}((\text{FResp} - \text{FNac}) / 365.25)$ → Edad en años
- $\text{Int}((\text{FResp} - \text{FNac}) / 30)$ → Edad en meses
- $\text{Int}((\text{FResp} - \text{FNac}) / 7)$ → Edad en semanas
- $(\text{FResp} - \text{FNac})$ → Edad en días

Observe el uso de la función **INT()**, que retorna la parte entera de un valor numérico decimal.

A menudo los usuarios prefieren llevar a cabo las operaciones con fechas y horas en los programas de hoja de cálculo o de gestión de bases de datos que utilizan para introducirlos. En estos programas habitualmente la diferencia entre dos fechas retorna el número de días transcurridos entre ambas y la diferencia entre dos horas retorna el número de minutos transcurridos entre ellas.

Puesto que en nuestra matriz de datos *HABITOX* las variables *FResp* y *FNac* son de tipo texto, antes de operar numéricamente con sus valores deberemos crear a partir de cada una de ellas tres nuevas variables con los valores numéricos de día, mes y año que contienen. A continuación, veremos que esta operación se realiza de forma muy sencilla con las funciones **SPLIT()** y **VALUE()** de jamovi.

1. Selecciona la columna de la variable *Genero* para insertar a su izquierda una nueva variable.

2. A continuación, ejecuta `Data | Add | Computed Variable | Insert` para insertar la columna y abrir el panel de opciones que permite introducir las propiedades de la nueva variable que contendrá el valor del día de la variable `FResp`.
3. En el panel de opciones que se abre, escribe `FRespDia` como nombre para la nueva variable calculada, `Día de respuesta al cuestionario` como `Description`, y la siguiente expresión en el recuadro `Formula`:

$$\text{VALUE}(\text{SPLIT}(\text{FResp}, "/", 1))$$

La función `SPLIT()` divide la cadena alfanumérica contenida en la variable `FResp` en todas las partes delimitadas por el carácter `" / "` indicado como segundo argumento y, a continuación, retorna la parte de texto correspondiente a la posición indicada en el tercer argumento (la 1 en este caso, para obtener el valor del día situado en `FResp` antes del primer `" / "`). Por último, la función `VALUE()` transforma el resultado alfanumérico de `SPLIT()` en un valor numérico.

4. Realiza los pasos anteriores para crear las variables `FRespMes`, `FRespAnyo`, `FNacDia`, `FNacMes` y `FNacAnyo`. Sitúa estas variables secuencialmente tras la variable `FRespDia`.

Una vez creadas las seis variables resultantes de la extracción de los componentes numéricos de la fecha de respuesta al cuestionario `FResp` y la fecha de nacimiento `FNac`, vamos a crear ahora la variable `Edad` calculando el tiempo en años cumplidos transcurrido entre ambas fechas.

Consideraciones sobre la precisión de los tiempos transcurridos

Habitualmente se acostumbra a truncar (i.e., eliminar la parte decimal) los **tiempos transcurridos** entre dos momentos temporales (dos fechas o dos horas). Pero para que los cálculos estadísticos con este tipo de variables sean precisos este truncamiento se debe evitar. Por este motivo, para preservar dicha precisión de los cálculos es preferible, siempre que sea posible, obtener y registrar la información de **momentos temporales** (fechas y/o horas) en lugar de tiempos transcurridos.

En el caso que no se disponga de los valores de los momentos temporales necesarios para realizar el cálculo exacto del tiempo transcurrido entre ellos, sino que se dispone directamente de una medida completa de dicho tiempo transcurrido (por ejemplo, una edad en años cumplidos), antes de proceder a los análisis estadísticos hay que realizar una corrección sumando el valor 0.5 . Esto es así, porque los tiempos transcurridos completos en realidad expresan el límite inferior del intervalo temporal entre dicho valor y el instante anterior al valor consecutivo superior (e.g., una edad de 43 años cumplidos en realidad indica que la persona tiene una edad comprendida entre 43 y el instante anterior a cumplir los 44 años).

5. Selecciona la columna de la variable `Genero` para insertar a su izquierda una nueva variable.
6. A continuación, ejecuta `Data | Add | Computed Variable | Insert` para insertar la columna y abrir el panel de opciones que permite introducir las propiedades de la nueva variable que contendrá la nueva variable `Edad`.
7. En el panel de opciones que se abre, escribe `Edad` como nombre para la nueva variable calculada, `Edad en años cumplidos al responder al cuestionario` como `Description`, y la siguiente expresión en el recuadro `Formula`:

$$(\text{FRespAnyo} - \text{FNacAnyo}) - \text{IF}(\text{FRespMes} == \text{FNacMes} \text{ and } \text{FRespDia} < \text{FNacDia} \text{ or } \text{FRespMes} < \text{FNacMes}, 1, 0)$$

La siguiente expresión calcula la `Edad` en años cumplidos a partir de las variables alfanuméricas `FResp` y `FNac`, sin crear previamente las variables con los valores del día, mes y año de estas fechas:

$$\begin{aligned} & \text{VALUE}(\text{SPLIT}(\text{FResp}, "/", 3)) - \text{VALUE}(\text{SPLIT}(\text{FNac}, "/", 3)) - \\ & \text{IF}(\text{VALUE}(\text{SPLIT}(\text{FResp}, "/", 2)) == \text{VALUE}(\text{SPLIT}(\text{FNac}, "/", 2)) \text{ and } \\ & \text{VALUE}(\text{SPLIT}(\text{FResp}, "/", 1)) < \text{VALUE}(\text{SPLIT}(\text{FNac}, "/", 1)) \text{ or } \\ & \text{VALUE}(\text{SPLIT}(\text{FResp}, "/", 2)) < \text{VALUE}(\text{SPLIT}(\text{FNac}, "/", 2)), 1, 0) \end{aligned}$$

8. Ejecuta la opción `Save` del menú principal  de jamovi para guardar los cambios realizados.

2.6. Ejercicio

Abre el fichero jamovi *HABITOX_P2.omv* y añade las variables transformadas y calculadas que se indican a continuación. En el caso de variables calculadas con base en expresiones lógicas presta especial atención al análisis previo de la **tabla de antecedentes y consecuente** y también a la consideración en dicho análisis de la posible presencia de **valores faltantes** –el resultado de las nuevas variables deberá ser un valor faltante cuando no se disponga de información suficiente en las variables antecedentes.

- **NivDrogas**: grado de adicción a las drogas, expresado como la suma de los valores de las tres variables que recogen su frecuencia de consumo, siempre que se disponga de la respuesta de los sujetos en todas ellas. Además, *NivDrogas* debe devolver el valor 0 cuando no se consume ninguna droga y el valor 6 cuando las tres drogas se consumen de forma habitual.
- **Bebedor1**: grado de consumo de alcohol, a partir de la información de la variable *NumBebidas2* que contiene la suma de las variables *Vino*, *Cerve* y *Copas* (ignorando los valores faltantes), y con base en la siguiente clasificación:

<i>Vino + Cerve + Copas</i>	<i>Bebedor</i>	
Valor	Valor	Etiqueta
0	0	<i>No bebedor</i>
1 a 7	1	<i>Menos de 8 vasos o copas/semana</i>
8 a 14	2	<i>Menos de 15 vasos o copas/semana</i>
Mayor o igual a 15	3	<i>15 o más vasos o copas/semana</i>

- **Bebedor2**: grado de consumo de alcohol con base en la misma clasificación anterior, pero sólo para los sujetos de los cuales se disponga de respuesta al menos en dos de las tres variables que recogen las respuestas sobre el consumo de los diferentes tipos de bebidas alcohólicas.
- **Alcohol**: número de tipos distintos de bebidas alcohólicas que toman los sujetos, siempre que se disponga de respuesta en las tres variables que recogen el consumo de los diferentes tipos de bebidas alcohólicas.
- **Habito**: debe contener el valor “No” si el sujeto no fuma ni bebe ni toma habitualmente ningún tipo de droga, el valor “Sí” si tiene alguno de estos hábitos, o el valor faltante si no se dispone de información suficiente para tomar esta decisión.

Una vez hayas finalizado este ejercicio, guarda los resultados en un nuevo fichero jamovi con el nombre *HABITOX_P2User.omv*. Puedes comparar tus resultados con la solución que se presenta en el fichero *HABITOX_P2Sol.omv*.

Ejercicio opcional avanzado

Crea la variable **IMC_Edad_Sexo** ("Normopeso según la edad") que contenga el valor 1 si el sujeto tiene un IMC en el rango de la normalidad, y el valor 0 en caso contrario. En la siguiente tabla se muestran los valores de referencia de los *IMC* para ambos sexos en distintas franjas de edad. Al crear la nueva variable *IMC_Edad_Sexo* hay que tener en cuenta la posibilidad que no se conozca el *IMC*, el sexo o la edad, y también que la edad del sujeto esté por debajo del valor mínimo baremado.

Edad	Rango de normalidad del IMC	
	Hombres	Mujeres
19-24	19-24	18-23
> 24-34	20-25	19-24
> 34-44	21-26	20-25
> 44-54	22-27	21-26
> 54-65	23-28	22-27
> 65	24-29	23-28

3. Análisis estadísticos descriptivos

En este apartado se revisan los principales procedimientos estadísticos que incorpora jamovi para realizar análisis estadísticos descriptivos, tanto de una variable categórica o cuantitativa, como de la asociación entre dos variables. En estos análisis se introduce también el potente procedimiento de generación de gráficos que incorpora jamovi para complementar visualmente la descripción de variables y de sus asociaciones basada en índices estadísticos.

3.1. Descripción estadística de una variable categórica

En este apartado revisaremos cómo obtener los índices y gráficos estadísticos básicos para describir una variable categórica. En primer lugar, utilizaremos como ejemplo la variable *Obesidad* (diagnóstico de sobrepeso u obesidad en base al índice de masa corporal -IMC-) generada en el apartado 2.3.

Índices estadísticos descriptivos para variables categóricas

Los índices de resumen estadístico más adecuados para variables categóricas son las **frecuencias** y **los porcentajes**, que se basan en el recuento del número de casos para cada una de las categorías de la variable.

Representaciones gráficas de variables categóricas

Los gráficos más habituales para representar dichas frecuencias o porcentajes son los **gráficos de barras** y los **gráficos de sectores**. Cuando la variable es nominal también se puede construir el **diagrama de Pareto**. Para variables categóricas ordinales también se pueden utilizar **diagramas de líneas**.



1. Abre el fichero de datos *HABITOX_P2Sol.omv* ejecutando la opción **Open** del menú principal de jamovi, o bien haz doble clic sobre este fichero en el explorador de Windows.
2. Ejecuta el menú **Analyses | Exploration | Descriptives**.
3. En el panel de opciones selecciona la **variable categórica ordinal** *Obesidad* en la lista de variables y pásala a la lista **Variables** pulsando .
4. Marca la casilla **Frequency tables** del panel de opciones, despliega el bloque **Statistics** y desmarca las casillas **Mean**, **Median** y **Std. deviation** por no ser adecuadas para las variables categóricas, así como las opciones **Minimum** y **Maximum**. A continuación, despliega el bloque **Plots** para marcar la opción **Bar plot**, que es el tipo de gráfico estadístico más adecuado de entre los disponibles en este bloque de opciones para una variable categórica.

Como puedes observar en la tabla **Descriptives** del panel de resultados, la muestra de este estudio incluye 32 personas, 30 con valores válidos (*N*) y 2 con valores faltantes (*Missing*) en la variable *Obesidad*. La tabla **Frequencies** muestra las frecuencias absolutas (*Counts*), las frecuencias relativas (*% of Total*) y las frecuencias acumuladas (*Cumulative %*) de los 30 participantes con información válida en *Obesidad*: 14 (46,7%) de los participantes tienen un diagnóstico de “Normopeso”, 13 (43.3%) un diagnóstico de “Sobrepeso”, y 3 (10.0%) un diagnóstico de “Obesidad”. Por último, el gráfico (**Plot**) muestra el diagrama de barras de la distribución del número de casos en las tres categorías de la variable *Obesidad*.

Todos los resultados que produce jamovi son **objetos** interactivos, de modo que al hacer clic sobre ellos se abre automáticamente el panel de opciones que los produjo. Pulsando el botón derecho del ratón sobre los resultados se pueden **copiar** al portapapeles (en formato textual o gráfico, según el caso) para pegarlos posteriormente en otro documento, **duplicar** para facilitar la repetición de los análisis con otras

variables u opciones, **exportar** en un fichero externo (en formato PDF o HTML), o bien **eliminar** del panel de resultados. Otra opción muy útil que se incluye en el menú contextual de objetos de resultados es la que posibilita añadir **notas textuales** propias. Estas acciones se pueden realizar sobre la totalidad de los objetos de resultados, sobre uno de ellos, o incluso en ocasiones sobre una parte de un resultado.

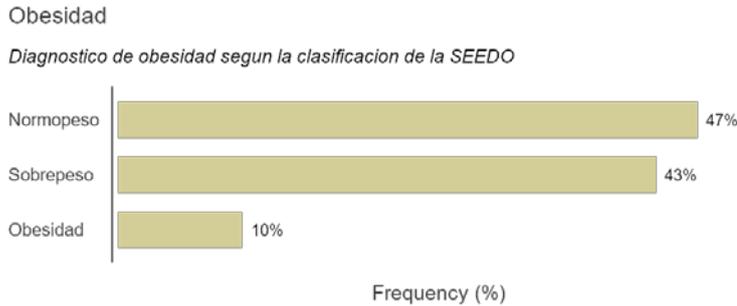
Por último, es destacable también el hecho que jamovi muestra los resultados formateados según la propuesta de reporte de análisis estadísticos y gráficos de la *American Psychological Association (APA)*.

The screenshot shows the Jamovi interface. On the left, the 'Descriptives' panel is active, showing a list of variables on the left and 'Obesidad' selected in the 'Variables' box. A red callout box points to a button in the top right of this panel labeled 'Cerrar panel de opciones'. On the right, the 'Results' section displays 'Descriptives' for 'Obesidad' with a table showing N=30 and Missing=2. Below this is a 'Frequencies' table for 'Obesidad' with three categories: Normopeso (14, 46.7%), Sobrepeso (13, 43.3%), and Obesidad (3, 10.0%). A bar chart titled 'Obesidad' shows the counts for each category. A red callout box points to a context menu over the bar chart, labeled 'Menú contextual', which contains options: All, Analysis, Group, Image, Copy, Export..., and Add Note.

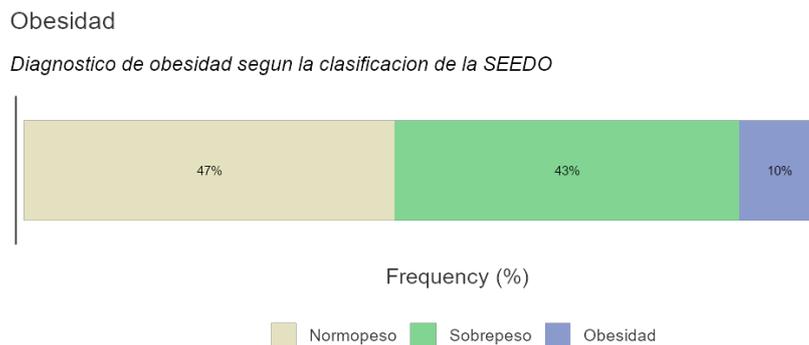
Inclusión o exclusión en los análisis de los casos con valores faltantes (*missing values*)

En la mayoría de los procedimientos, jamovi lleva a cabo los análisis estadísticos excluyendo los casos con información faltante en las variables analizadas (exclusión *by analysis* o bien *listwise*). Algunos procedimientos permiten indicar que se desea incluir los casos con valores faltantes en los análisis.

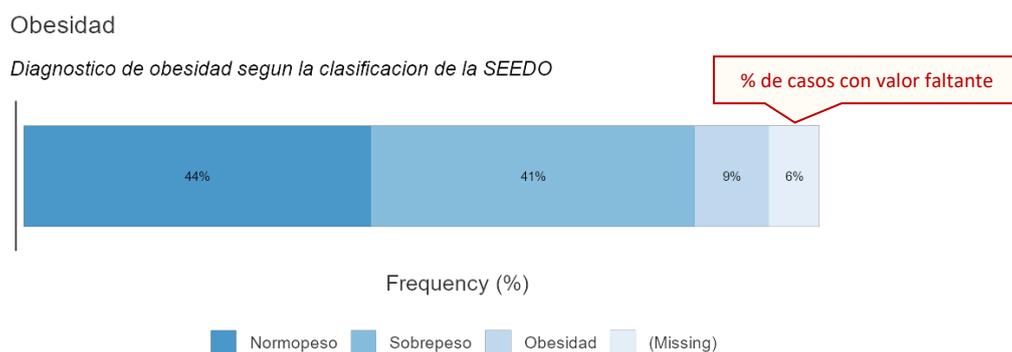
5. Cierra el panel de opciones pulsando el botón → que se encuentra en su esquina superior derecha.
6. Las barras del gráfico anterior muestran los recuentos de casos en cada categoría. Para obtener el gráfico de barras con los porcentajes de cada categoría de *Obesidad* se puede ejecutar el menú **Analyses | Exploration | Survey Plots**. Nota: este menú sólo está disponible si se ha instalado previamente el módulo *surveymv – Survey Plots* siguiendo los pasos descritos en el apartado 1.1. Ejecuta este menú pasando la variable *Obesidad* a **Variables**. A continuación, despliega el bloque **Nominal / Ordinal Plots** y marca **Grouped bar** en **Plot Type** y **Percentages** en **Frequency Type**.



7. Como irás viendo a lo largo de este apartado, jamovi ofrece una gran variedad de opciones para modificar el aspecto de los gráficos estadísticos. Así, por ejemplo, se puede cambiar el gráfico anterior de modo que muestre las barras apiladas (*Stacked bar*) en lugar de agrupadas (*Grouped bar*). Marca la opción **Stacked bar** en **Plot Type**:



8. También se pueden modificar los datos que se desean mostrar en el gráfico. Una opción de uso frecuente es la inclusión de los valores faltantes en el gráfico. Desmarca la casilla **Hide missing values** para que el gráfico anterior incluya también el porcentaje de participantes del estudio para los cuales no se pudo obtener el diagnóstico de sobrepeso u obesidad (casos con valores faltantes en la variable *Obesidad*). Para mostrar la facilidad con la que jamovi permite modificar el aspecto de los gráficos haz clic en el botón de los tres puntos (⋮) que se encuentra en la parte superior derecha de la ventana de la aplicación (bajo el botón ✕ de cierre de la ventana). En el panel de opciones que se despliega, elige **Sequential | Blues** en **Plots | Color Palette** y, a continuación, cierra el panel de opciones de jamovi y el panel de opciones de **Survey Plots** pulsando → en ambos casos:



Vamos ahora a obtener la distribución de frecuencias de la **variable categórica nominal** *Tabaco*, que contiene los códigos de la marca de tabaco que fuman los sujetos. Incluir en este análisis a los participantes que hayan respondido que no son fumadores sería incorrecto, ya que la pregunta sobre la marca de tabaco que se fuma **sólo es aplicable** a aquellos participantes que hayan manifestado que su consumo de cigarrillos al día es superior a 0 (variable *NumCig*).

9. Antes de llevar a cabo la descripción estadística de la variable *Tabaco* vamos a establecer un filtro en la matriz de datos, de modo que los análisis se realicen únicamente para los sujetos fumadores. Para ello, ejecuta el menú **Data | Filters**.

Establecimiento de filtros de casos en los análisis estadísticos

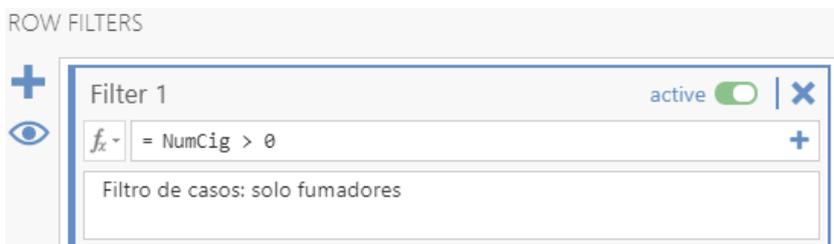
Aunque la mayor parte de análisis estadísticos se realizan sobre la totalidad de los casos contenidos en la matriz de datos, en ocasiones se desea realizar dichos análisis para un subconjunto de ellos. La aplicación de estos filtros de casos puede estar motivada por diferentes razones:

- Se desea realizar análisis sólo para un subconjunto de la muestra de casos que tienen alguna o algunas características en común, y no se desea comparar estadísticamente los resultados de dichos análisis con los de otros grupos de sujetos. Por ejemplo, se desea realizar análisis sólo para los sujetos de una determinada localización geográfica, o sólo para los sujetos de un sexo determinado, etc.
- Se desea excluir de los análisis los sujetos con valores faltantes debidos a que las variables eran de tipo **no aplicable**.
- Se desea incluir sólo a una muestra aleatoria del total de casos para llevar a cabo algún tipo de validación de la información, o bien, de los resultados de los análisis estadísticos.

Análisis estadísticos con valores faltantes (*missing*)

Al concluir sobre los valores faltantes, es importante distinguir los debidos a que la pregunta es **no aplicable** (en cuyo caso se dejan vacíos durante la introducción de datos) de los debidos a otras causas, como la **no respuesta deliberada** (por desconocimiento o bien por negativa a responder), o por ser **valores incorrectos** declarados como valores faltantes durante la fase de depuración de datos. En este último caso se pueden introducir en la matriz de datos con códigos numéricos fuera del rango de valores válidos de la variable –por ejemplo, -999, -998, etc.- para poder identificarlos, etiquetarlos y analizarlos posteriormente de forma separada. Recuerda que jamovi permite indicar los valores faltantes para una variable en el panel de opciones de definición de sus propiedades (véase el apartado 1.1 de definición de variables).

10. En el panel de opciones escribe la condición lógica que deben cumplir los casos que se incluirán en los análisis, tal y como muestra la siguiente captura de pantalla:



11. Cierra el panel de opciones pulsando  y observa que en el panel de datos se ha añadido una primera columna *Filter 1* que indica qué casos se incluirán y cuáles se excluirán de los análisis a partir de este momento. Los casos excluidos se muestran sombreados:

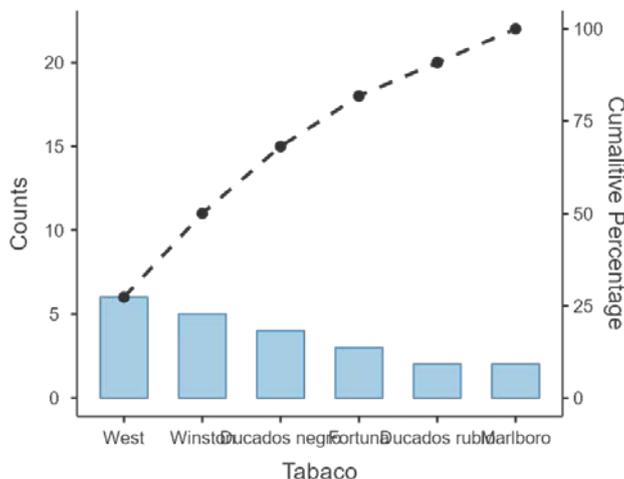
	Filter 1	Caso	Pseudo	Fresp	Fnac
1	✓	1	elzorro	25/04/2010	22/09/1970
2	✓	2	romeu	24/04/2010	14/11/1972
3	✓	3	pompeu	26/04/2010	28/02/1966
4	✓	4	julieta	28/04/2010	01/06/1976
5	✓	5	dali	29/04/2010	13/01/1973
6	✓	6	marieta	28/04/2010	28/09/1983
7	✓	7	ramoneta	30/04/2010	25/03/1975
8	✓	8	spiderman	21/04/2010	03/07/1980
9	✓	9	creuat	28/04/2010	24/07/1966
10	✓	10	fantastica	28/04/2010	13/03/1976
11	✗	11	salvador	12/04/2010	12/11/1982
12	✓	12	deesa	29/04/2010	10/12/1978
13	✗	13	marina	24/04/2010	31/03/1975
14	✓	14	ullviu	30/04/2010	10/10/1980

- Ejecuta el menú **Analyses | Exploration | Descriptives** y pasa la variable *Tabaco* a la lista **VARIABLES** pulsando **→**.
- Marca la casilla **Frequency tables** en el panel de opciones, despliega el bloque **Statistics** y desmarca las casillas **Mean, Median, Std. Deviation, Minimum y Maximum**:

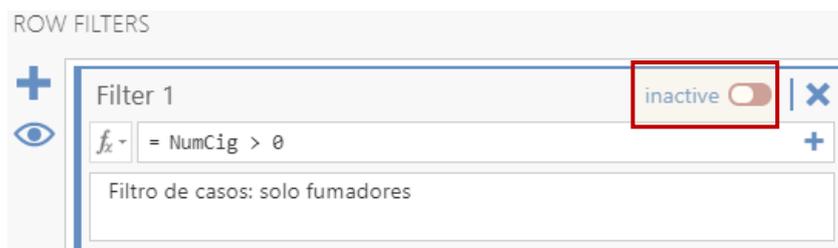
		Frequencies of Tabaco			
		Levels	Counts	% of Total	Cumulative %
Descriptives	Tabaco	Fortuna	3	13.6 %	13.6 %
		Ducados negro	4	18.2 %	31.8 %
		Ducados rubio	2	9.1 %	40.9 %
		West	6	27.3 %	68.2 %
		Winston	5	22.7 %	90.9 %
		Marlboro	2	9.1 %	100.0 %
	N	22			
	Missing	1			

Como puedes observar, el número de casos válidos (*N*), es decir, que han respondido la pregunta sobre la marca de tabaco que se fuma, es de 22, y que 1 caso no ha respondido a esta pregunta del cuestionario (*Missing*), aun siendo fumador. La tabla de frecuencias muestra las frecuencias absolutas (*Counts*), las frecuencias relativas (*% of Total*) y las frecuencias acumuladas (*Cumulative %*) sólo para la muestra de participantes que han respondido a la pregunta *Tabaco* (*N=22*).

- Vamos ahora a solicitar un **diagrama de Pareto** (*Pareto chart*), que es un tipo de diagrama de barras muy informativo para las **variables categóricas nominales**. Este gráfico presenta las barras de las categorías ordenadas en función de la frecuencia de casos de cada categoría (de mayor a menor frecuencia), solapando un diagrama de línea con las frecuencias acumuladas. Ejecuta el menú **Analyses | Exploration | Pareto Chart** y traslada la variable *Tabaco* a **X-Axis**:



- Finalmente, ejecuta de nuevo **Data | Filters** para desactivar el filtro de casos:



- Si observas el panel de resultados, los descriptivos de la variable *Tabaco* muestran que el número de casos (*N*) incluido en el análisis es ahora de 32, de los cuales 22 son casos válidos, es decir, que contienen la respuesta a la pregunta sobre la marca de tabaco que se fuma, y 10 casos son *Missing*, es decir, que no contienen respuesta a dicha pregunta. Fíjate que el número de valores faltantes ha

pasado de 1 a 10 respecto al análisis realizado con el filtro que incluía sólo a los sujetos que han respondido que fuman más de 0 cigarrillos al día. Estos 9 casos adicionales son los sujetos que no han respondido a la pregunta *Marca de tabaco* por ser no fumadores, mientras que el caso faltante restante es de naturaleza diferente, ya que se trata de un sujeto fumador que no ha indicado la marca de tabaco que fuma. Por tanto, hay 9 casos sin valor en *Tabaco* para los que en realidad la pregunta *Marca de tabaco* es **no aplicable** y sólo 1 caso con un valor faltante debido a **falta de respuesta** por su parte. En conclusión, el número de participantes que no han respondido la marca de tabaco que fuman es sólo de 1 sobre 23 (un 4,35%) –análisis con filtro- y no de 10 sobre 32 (un 31,25%) -análisis sin filtro-.

Descriptives	
Tabaco	
N	22
Missing	10

3.2. Descripción estadística de una variable cuantitativa

En este apartado revisaremos cómo obtener los índices y gráficos estadísticos básicos para describir una variable cuantitativa. Para ello utilizaremos como ejemplo la **variable cuantitativa discreta** *NumCig*, que registra el número de cigarrillos que fuman en promedio cada día los participantes.

Sobre la adecuación de los índices estadísticos para describir variables cuantitativas

Para describir variables cuantitativas se utilizan índices estadísticos basados en momentos, como la **media aritmética** o la **desviación estándar**, así como índices estadísticos basados en ordenaciones, como la **mediana** o la **amplitud intercuartil**.

Los índices estadísticos basados en momentos, como la media aritmética, no son aconsejables cuando la distribución de la variable cuantitativa es marcadamente asimétrica, ni tampoco cuando contienen valores extremos o atípicos, debido a su falta de resistencia. En estos casos es aconsejable utilizar los índices basados en ordenaciones, como la mediana. En el caso particular que la variable cuantitativa sea de naturaleza discreta con un rango reducido de valores, la descripción estadística más apropiada consiste en obtener la moda (el valor más frecuente) y la amplitud o rango (diferencia entre el valor mínimo y el valor máximo).

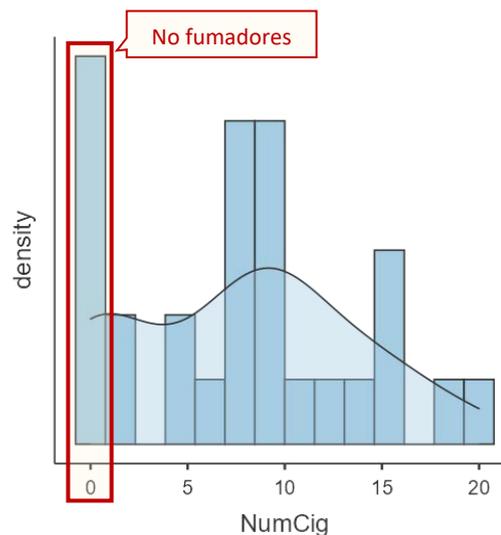
Representación gráfica de variables cuantitativas

Por otro lado, los gráficos que se emplean habitualmente para este tipo de variables son el **histograma** y el **polígono de frecuencias** (para variables cuantitativas continuas), el **diagrama de barras** y el **diagrama de líneas** (para variables cuantitativas discretas), y el **diagrama de tallo y hojas (stem & leaf)**, el **diagrama de violín** y el **diagrama de caja (boxplot)** para todos los tipos de variables cuantitativas.

1. Ejecuta el menú **Analyses | Exploration | Descriptives** y pasa la variable *NumCig* a la lista **Variabls** pulsando **→**.
2. Despliega el bloque **Statistics** y marca las casillas **Percentiles** y **Mode**. A continuación, despliega el bloque **Plots** y marca las casillas **Histogram** y **Density** para obtener el histograma del consumo de cigarrillos diario y solapar encima de él la función de densidad de probabilidad que ayudará a valorar la forma de la distribución de la variable.

Como puedes observar, se trata de una distribución bimodal, resultado de la mezcla de dos distribuciones distintas: la distribución de los participantes no fumadores (con valor 0) y la distribución de los participantes fumadores (con valores superiores a 0). No es correcto describir esta variable para la totalidad de los casos, ya que los descriptivos que obtenemos mezclan la información de estas dos muestras de sujetos. Así, por ejemplo, la media aritmética no indicará el número de cigarrillos que fuman diariamente los participantes fumadores, que es el valor que puede resultar de interés, sino un promedio menor por el hecho de incluir en su cálculo también a los participantes que no fuman.

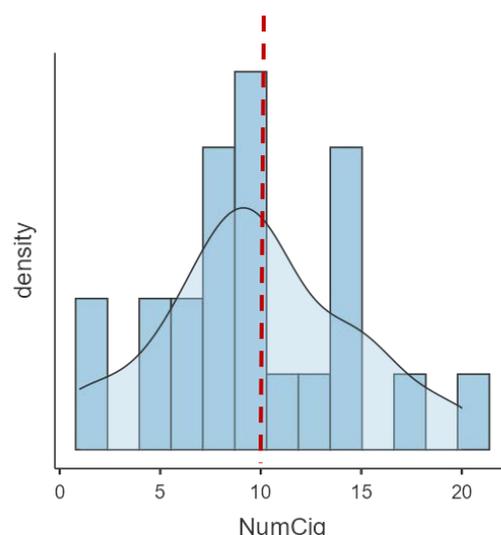
Descriptives	
	Numcig
N	29
Missing	3
Mean	7.79
Median	8
Mode	0.00
Standard deviation	5.92
Minimum	0
Maximum	20
25th percentile	1.00
50th percentile	8.00
75th percentile	11.0



3. Para describir adecuadamente la variable *NumCig* se debe realizar previamente un filtro de los datos, seleccionando únicamente los participantes fumadores. Para ello, ejecuta el menú **Data | Filters** y activa de nuevo el **Filter 1** definido anteriormente ($\text{NumCig} > 0$).
4. A continuación, para obtener de nuevo la descripción estadística de la variable *NumCig* puedes volver a ejecutar **Analyses | Exploration | Descriptives**. Otra forma de repetir un análisis consiste en señalarlo con el ratón en el panel de resultados y ejecutar el menú contextual **Analysis | Duplicate**. De este modo se puede duplicar cualquier análisis estadístico numérico y/o gráfico para cambiar alguna de sus propiedades (e.g., para realizarlo sobre otra/s variable/s). Duplica el análisis descriptivo para la variable *NumCig* y haz clic sobre la copia del resultado que jamovi inserta a continuación en el panel de resultados.

En este caso no sería necesario repetir el análisis porque jamovi actualiza automáticamente todos los resultados cuando se aplica un filtro de datos. Lo hacemos así aquí para destacar la facilidad con la que se pueden llevar a cabo nuevos análisis a partir de los ya realizados, y también para ilustrar el uso de anotaciones. En este caso sería recomendable incluir una nota en los resultados del nuevo análisis indicando que se debe activar el filtro de datos para obtenerlos.

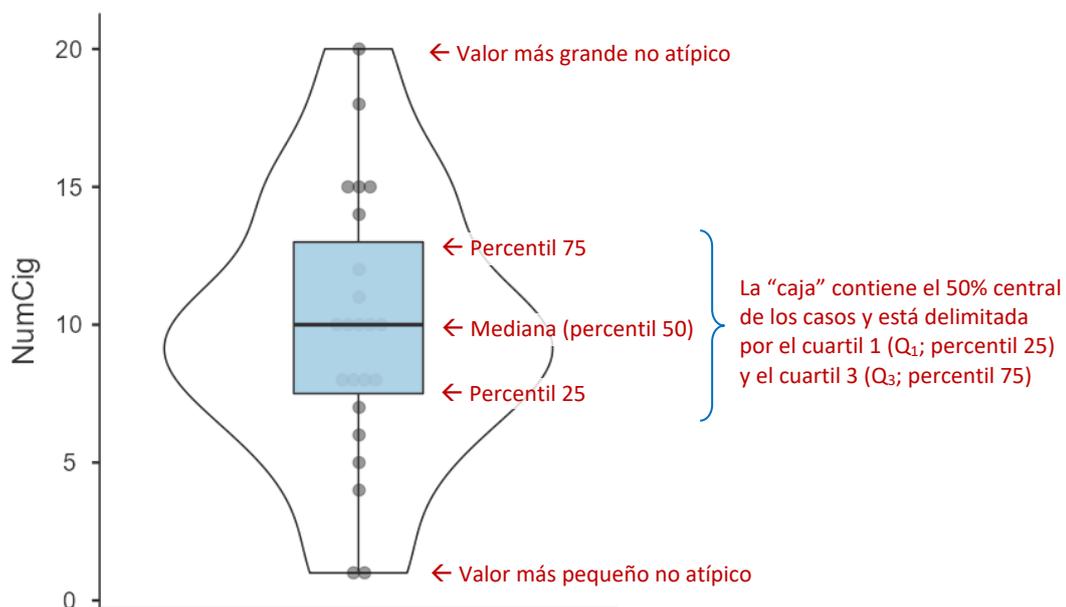
Descriptives	
	Numcig
N	23
Missing	0
Mean	9.83
Median	10
Mode	10.0
Standard deviation	4.88
Minimum	1
Maximum	20
25th percentile	7.50
50th percentile	10.0
75th percentile	13.0



El *histograma* muestra un patrón bastante simétrico, centrado aproximadamente alrededor de los 10 cigarrillos diarios, con frecuencias menores a medida que el consumo se aleja de este valor. En el eje horizontal del histograma se muestran los valores de la variable (*NumCig*), y el número de barras se corresponde con los intervalos en que jamovi ha dividido automáticamente el rango de valores observados, entre el valor mínimo (1) y el máximo (20). La altura de cada barra refleja la frecuencia de casos (*density*) con valores en su intervalo; la barra no se muestra para los intervalos sin casos.

Los estadísticos descriptivos obtenidos muestran la siguiente información:

- La matriz de datos contiene 23 participantes fumadores (N).
 - Los sujetos fuman en promedio (media aritmética -*Mean*-) 9.83 cigarrillos al día.
 - La mediana (*Median* o percentil 50 -*50th percentile*-) indica que la mitad de los sujetos fuman diariamente un máximo de 10 cigarrillos y, por tanto, el 50% restante tienen un consumo superior o igual a este valor. Nótese que el valor de la mediana es muy próximo a la media aritmética, lo cual indica que la distribución de esta variable es bastante simétrica alrededor de estos dos valores centrales, tal y como se puede comprobar visualmente en el histograma.
 - La desviación estándar (*Standard deviation*) es de 4.877 e indica la variabilidad o dispersión que presenta la distribución del consumo diario de cigarrillos. La amplitud intercuartil (*Interquartile range*, IQR) es otra medida de dispersión de uso habitual que indica la amplitud del intervalo que contiene al 50% central de los individuos (se obtiene como diferencia entre el percentil 25 y el percentil 75 de la distribución), en este caso 5.5 cigarrillos diarios.
 - Los valores del consumo de cigarrillos están comprendidos entre 1 (*Minimum*) y 20 (*Maximum*) y, por tanto, entre el valor máximo y el mínimo hay una diferencia de 19 puntos.
 - Los percentiles 25 y 75 indican que el 25% de los participantes de este estudio fuman menos de 7.5 cigarrillos y el 75% menos de 13 cigarrillos, respectivamente.
5. A continuación, veremos dos tipos de gráficos que permiten representar una variable cuantitativa, ya sea discreta o continua: el diagrama de caja (*Box plot*) y el diagrama de violín. Para obtener estos dos gráficos combinados basta con marcar las casillas **Box plot** y **Violin** del bloque **Plots** del menú **Analyses | Exploration | Descriptives**. Es recomendable marcar también en **Box Plots** la opción **Data** y seleccionar **Stacked** en la lista para que se muestren los datos en forma de puntos sobreimpresos en el diagrama; de este modo se puede visualizar en qué valores se encuentran los casos (a mayor número de puntos, mayor número de casos con un determinado valor):



El diagrama de violín (las dos líneas curvas) se puede interpretar de un modo parecido a un histograma -basta con apoyar la cabeza sobre el hombro 😊-.

El diagrama de caja o *boxplot* (en azul) permite observar rápidamente la mediana, los percentiles 25 y 75, la dispersión, la asimetría de la distribución y los valores atípicos (cuando existen). Se consideran valores atípicos aquéllos inferiores a $Q_1 - 1.5 \times IQR$ o superiores a $Q_3 + 1.5 \times IQR$, donde, como ya hemos señalado antes, IQR es la amplitud intercuartil ($Q_3 - Q_1$). Cuando la distribución es totalmente simétrica se puede doblar el gráfico por la línea de la mediana y todas las líneas coincidirían, lo que casi sucede en este caso.

3.3. Descripción estadística de la relación lineal entre dos variables cuantitativas

Tras haber revisado los principales índices estadísticos y gráficos para resumir la información contenida en una variable, en los siguientes apartados vamos a presentar las medidas y representaciones adecuadas para la descripción de la relación entre dos o más variables registradas en una matriz de datos.

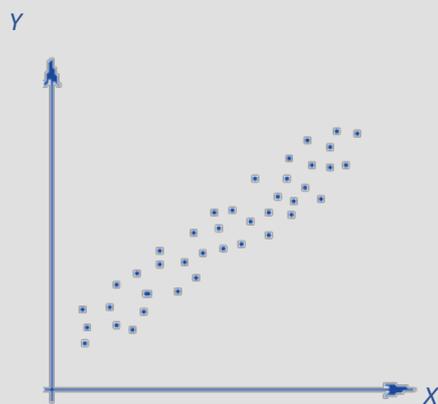
Concepto de asociación estadística entre dos variables

La asociación estadística entre dos variables se manifiesta a) a través de algún tipo de patrón de emparejamiento o covariación entre los valores de ambas variables, o b) en el hecho de que la distribución de valores de una variable difiere en función de los valores que toma la otra. Por tanto, podemos referirnos al concepto de asociación estadística en términos de **relación** entre los valores de ambas variables, o bien en términos de **diferencia** entre los valores de los estadísticos resumen de una de las variables obtenidos en cada grupo de observaciones definido a partir de los valores de la otra. Por el contrario, nos referimos a la ausencia de asociación estadística entre dos variables en términos de **independencia** entre ellas.

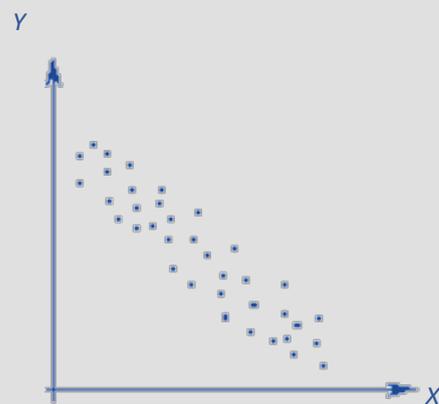
En la práctica, para afrontar el análisis estadístico descriptivo de la relación entre dos variables se parte, en primer lugar, de la consideración de sus **escalas de medida**, para decidir a partir de ellas los índices y gráficos adecuados para investigar dicha relación.

Interpretación del diagrama de dispersión para valorar la relación lineal entre dos variables cuantitativas

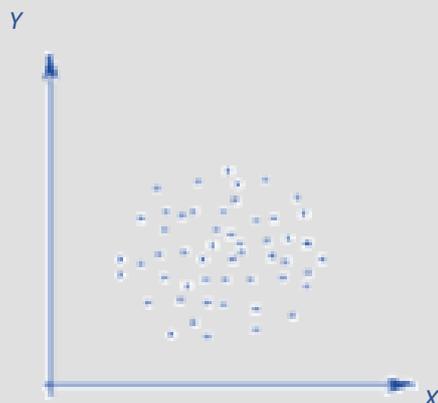
La relación entre dos variables cuantitativas se puede representar gráficamente con un **diagrama de dispersión (Scatterplot)**, que presenta en una nube de puntos la puntuación de cada participante en las dos variables estudiadas:



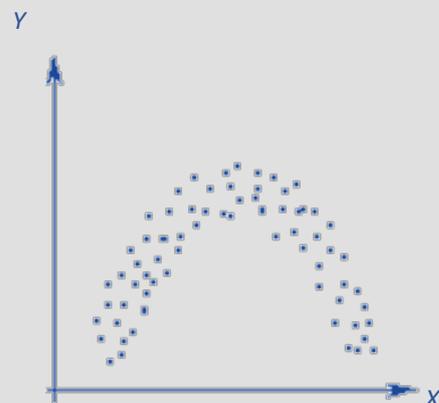
Relación lineal creciente o "directa"



Relación lineal decreciente o "inversa"



Ausencia de relación



Relación no lineal en "U" invertida

- Una nube de puntos descendente indica que existe una marcada relación negativa, es decir, que a medida que aumentan o disminuyen los valores de una variable disminuyen o aumentan en sentido contrario, en promedio, los valores de la otra variable.
- Una nube de puntos ascendente indica que hay una fuerte relación positiva, es decir, que a medida que aumentan o disminuyen los valores de una variable aumentan o disminuyen en el mismo sentido también, en promedio, los valores de la otra variable.
- Una nube de puntos horizontal o esférica indica ausencia de relación entre las dos variables.
- Una nube de puntos con un patrón en forma de “U” o “U invertida”, sinusoidal, etc. indica la presencia de una relación no lineal.

Disposición de las variables en los ejes X e Y

Es importante señalar que la disposición de cada variable en los ejes horizontal o de abscisas (X) y vertical o de ordenadas (Y) habitualmente no es discrecional. Si se tiene un interés por analizar la relación en un sentido concreto, por ejemplo, cómo el número de horas de trabajo semanal afecta al grado de exigencias psicológicas, entonces la primera variable se debe situar en el eje X y la segunda en el eje Y . Se habla en este caso de **relación asimétrica** entre las dos variables. En los estudios con diseños experimentales se utiliza la terminología *variable independiente* y *variable dependiente* para referir esta diferencia en el rol de las variables X e Y , respectivamente. Esta nomenclatura cambia cuando los diseños son de tipo no experimental, utilizándose los términos *variable de exposición* (también *variable predictiva* o *variable explicativa*) y *variable de respuesta* (también *variable criterio* o *variable de resultado*) para denominar las variables X e Y , respectivamente. Cuando entre las dos variables existe una **relación simétrica** respecto a su rol, su disposición en los ejes del diagrama sí es discrecional.

Asociación y causalidad

La presencia de asociación estadística es uno de los requisitos para poder establecer una relación de causalidad entre dos variables, pero no es suficiente para dar soporte a este tipo de interpretación; es otras palabras, **la presencia de asociación estadística es una condición necesaria pero no suficiente para establecer un vínculo causal entre dos variables.**

Por otro lado, en el ámbito de las ciencias sociales y de la salud se acepta que las relaciones de causa-efecto se dan en términos de relaciones funcionales probabilísticas y no de relaciones deterministas, como se puede establecer en cambio en otras disciplinas científicas como la física. Así, al estudiar el comportamiento humano, se asume que los valores de una variable causal no determinan los valores específicos de la variable que mide los resultados o efectos, sino que inciden únicamente sobre las distribuciones de probabilidad de estos resultados. Por este motivo, es preferible utilizar los términos *exposición*, *factor de riesgo*, etc., para referir las variables causales en psicología. Esto es así por la etiología multifactorial de los fenómenos psicológicos, la imposibilidad de manipular todas las posibles causas de un mismo fenómeno, la asunción de que la realidad psicológica es demasiado compleja para poder conocerla en su totalidad, la aceptación de que la medida en psicología siempre comporta un determinado grado de error, etc.

En cualquier caso, dejando de lado las limitaciones inherentes al establecimiento de relaciones causales, para poder establecer la existencia de una relación de este tipo entre dos variables se requiere el cumplimiento de tres condiciones básicas:

1. **Precedencia temporal de la exposición:** la variable de exposición debe preceder temporalmente a la variable que recoge su efecto. Esta condición comporta la necesidad de garantizar mediante el diseño de la investigación, o bien en base a una teoría psicológica o evidencia científica bien establecida, que el fenómeno causal es anterior al inicio del efecto. Así, por ejemplo, si se desea estudiar la relación entre la ansiedad ante una prueba de conocimientos de estadística y el rendimiento en dicha prueba en alumnos de primer curso de psicología, se podría pensar que un alto grado de ansiedad determina peores resultados en dicha prueba, y que un bajo nivel determina un alto rendimiento en la misma. Pero también se podría pensar que es la situación de la propia prueba de conocimientos de estadística la que provoca una reacción de ansiedad, de tal

modo que los alumnos poco preparados reaccionan con un aumento de la ansiedad mayor a la de los mejor preparados. En este ejemplo, por tanto, no se puede asegurar la precedencia temporal de ninguna de las dos variables.

2. **Presencia de asociación estadísticamente significativa:** debe existir una asociación o covariación significativa, es decir, con una probabilidad muy baja de que sea debida al azar, entre la variable de exposición y la variable de respuesta o efecto. Más adelante, en el capítulo dedicado a la inferencia estadística se abordará en detalle el concepto de significación estadística.
3. **Ausencia de relación espuria:** no deben existir otras variables que sean la causa de la relación observada entre la exposición y el efecto estudiados. Esta condición conlleva un sólido estudio teórico previo que permita tomar en consideración otras variables que mantengan una relación con la exposición y con la respuesta investigadas. Así, en el ejemplo enunciado sobre el estudio de la relación entre el grado de ansiedad y el rendimiento en la prueba de conocimientos de estadística, está claro que ambas variables pueden depender del nivel de preparación previo de los estudiantes, nivel éste que en última instancia es el causante tanto del grado de ansiedad como del resultado de la prueba de conocimientos. Otro ejemplo de relación espuria sería la asociación positiva que se da entre la talla y el salario en la población trabajadora. Esta relación depende en realidad de una tercera variable, el sexo, en el sentido que las mujeres tienen en promedio, todavía en 2012 ☹, sueldos más bajos que los hombres, al tiempo que su talla también es en promedio más baja.

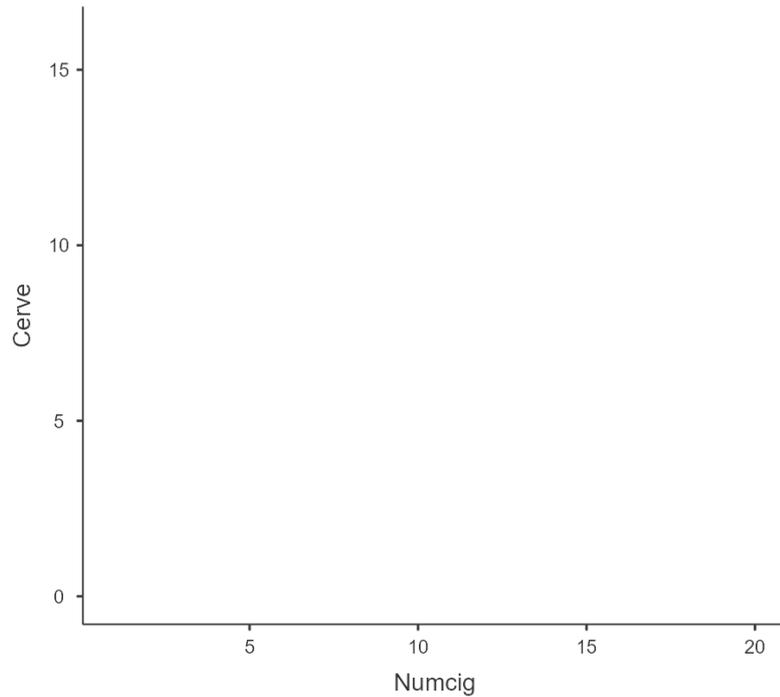
De lo expuesto hasta ahora se deducen tres conclusiones fundamentales a la hora de estudiar relaciones causales:

1. Las hipótesis sobre relaciones causales se deben basar en teorías, modelos o hechos sólidamente establecidos (denominados también evidencias científicas), que permitan justificar previamente la precedencia temporal entre la variable de exposición y la respuesta, así como la propia relevancia y coherencia del estudio de esta relación.
2. El diseño de la investigación debe ofrecer el máximo nivel de garantías posible sobre la ausencia de una relación espuria entre la exposición y la respuesta. En los diseños experimentales, cuando la muestra de estudio es suficientemente grande, la aleatorización permite el control de las variables confundidas, puesto que al asignar al azar a los sujetos a los diferentes grupos o momentos de intervención se consigue equilibrar entre dichos grupos o momentos la posible influencia de las variables confundidas, aumentando así su comparabilidad. Sin embargo, en los diseños no experimentales la ausencia de aleatorización desplaza el peso del control de las variables confundidas desde el diseño hacia estrategias de análisis estadístico más sofisticadas. Así, en este tipo de diseños, el procedimiento para descartar posibles relaciones espurias recae fundamentalmente en el registro de las potenciales variables de confusión para, posteriormente, aislar su efecto sobre la relación estudiada durante la fase de análisis estadístico. El control estadístico de la confusión se lleva a cabo mediante estrategias de análisis estratificados, que estudian la relación de interés de forma separada para diferentes grupos de sujetos definidos en base a los niveles de las potenciales variables confundidas, o bien incluyéndolas en análisis multivariantes que permiten obtener medidas “ajustadas” de los efectos de las exposiciones, “descontando” el efecto sobre éstas y sobre la respuesta de las potenciales variables de confusión. La exposición de este tipo de técnicas estadísticas más avanzadas escapa a los objetivos del presente texto.
3. El proceso de documentación científica es una tarea previa fundamental para la planificación de un estudio que investigue posibles relaciones causales. Este proceso de documentación es el que permitirá conocer las teorías, modelos y hechos establecidos sobre el fenómeno de interés que fundamentan, en última instancia, la pertinencia y la coherencia del estudio de la relación entre las variables de interés. Por otro lado, las investigaciones previas son la fuente de conocimiento principal sobre las variables de control que se deben registrar e incluir posteriormente en los análisis estadísticos. Por último, junto con las variables de control, el conocimiento sobre el tipo de diseños de investigación utilizados por otros investigadores en el ámbito de estudio de interés es siempre una guía extremadamente útil de cara a la planificación de una nueva investigación.

Vamos a revisar ahora los gráficos e índices estadísticos de uso más habitual para estudiar la relación lineal entre dos variables cuantitativas. Para ello estudiaremos la relación entre el consumo de tabaco (*NumCig*) y de cervezas (*Cerve*) de los participantes.

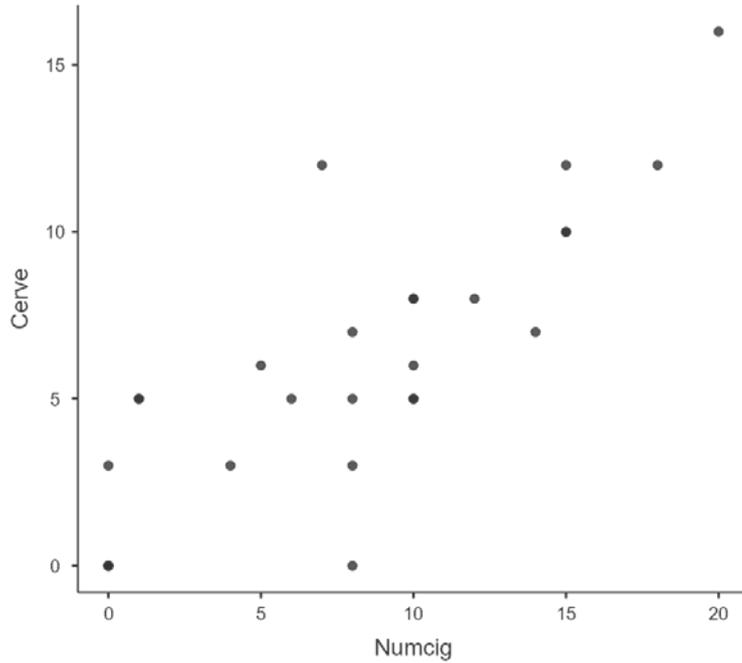
1. Realizaremos primero la representación gráfica de forma manual, para familiarizarnos con el procedimiento de construcción de un diagrama de dispersión. El siguiente listado muestra los valores de las variables *NumCig* y *Cerve* para los participantes con valores válidos en ambas variables. Se trata de que marques con un punto la intersección entre las líneas perpendiculares y horizontales correspondientes a los valores en *NumCig* y en *Cerve*, respectivamente, de cada uno de los 27 participantes:

Caso	Numcig	Cerve
1	8	3
2	5	6
3	1	5
4	20	16
5	8	7
7	8	5
8	15	12
9	1	5
10	15	10
12	12	8
14	8	0
15	6	5
16	18	12
18	4	3
19	0	0
20	10	8
21	10	5
22	0	3
23	0	0
24	0	0
26	15	10
27	7	12
28	10	5
29	0	0
30	14	7
31	10	8
32	10	6
N	27	27

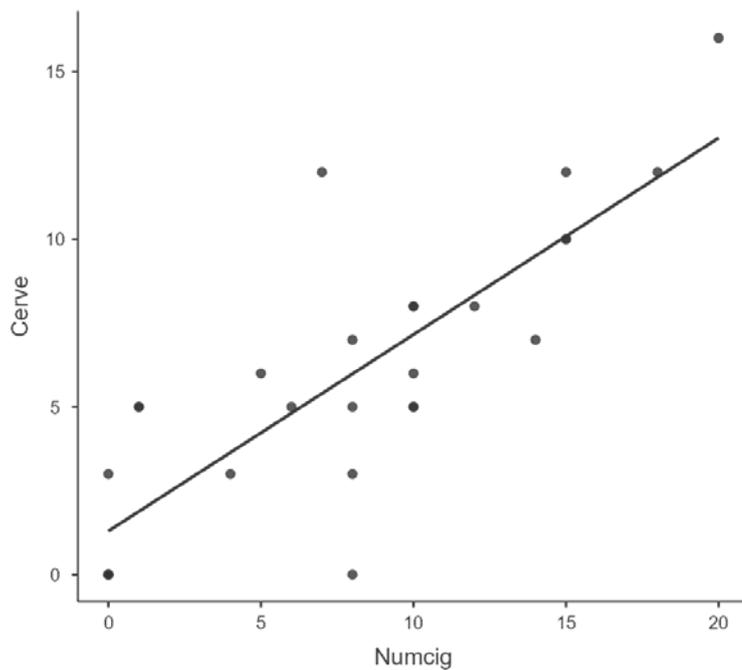


Como puedes ver, el diagrama de dispersión muestra un patrón claro de relación lineal directa entre las dos variables, dado que la nube de puntos tiene forma de elipse ascendente. Así pues, los valores altos en el consumo de tabaco están relacionados con valores también altos en el consumo de cervezas.

2. Veamos cómo se obtiene el gráfico de dispersión con jamovi. La disposición de cada variable en los ejes X e Y es en este caso discrecional. Si se tuviera un interés por analizar la relación en un sentido concreto, por ejemplo, cómo el número de cigarrillos afecta al consumo de cervezas, entonces la variable que se debe situar en el eje X es el número de cigarrillos y la variable del eje Y el consumo de cervezas. Ejecuta el menú **Analyses | Exploration | Scatterplot**. En el panel de opciones pasa la variable *NumCig* a X-Axis y la variable *Cerve* a Y-Axis. Comprueba si coincide con el que has realizado manualmente en el paso 2 y observa que jamovi ha excluido automáticamente los casos con valores faltantes en alguna de las dos variables:



3. Para facilitar la valoración de la relación lineal entre las dos variables se puede añadir una recta de ajuste superpuesta al gráfico anterior. Esta recta minimiza las distancias de todos los puntos a la recta. Haz clic sobre el diagrama de dispersión en el panel de resultados y marca la opción **Linear Regression Line**.



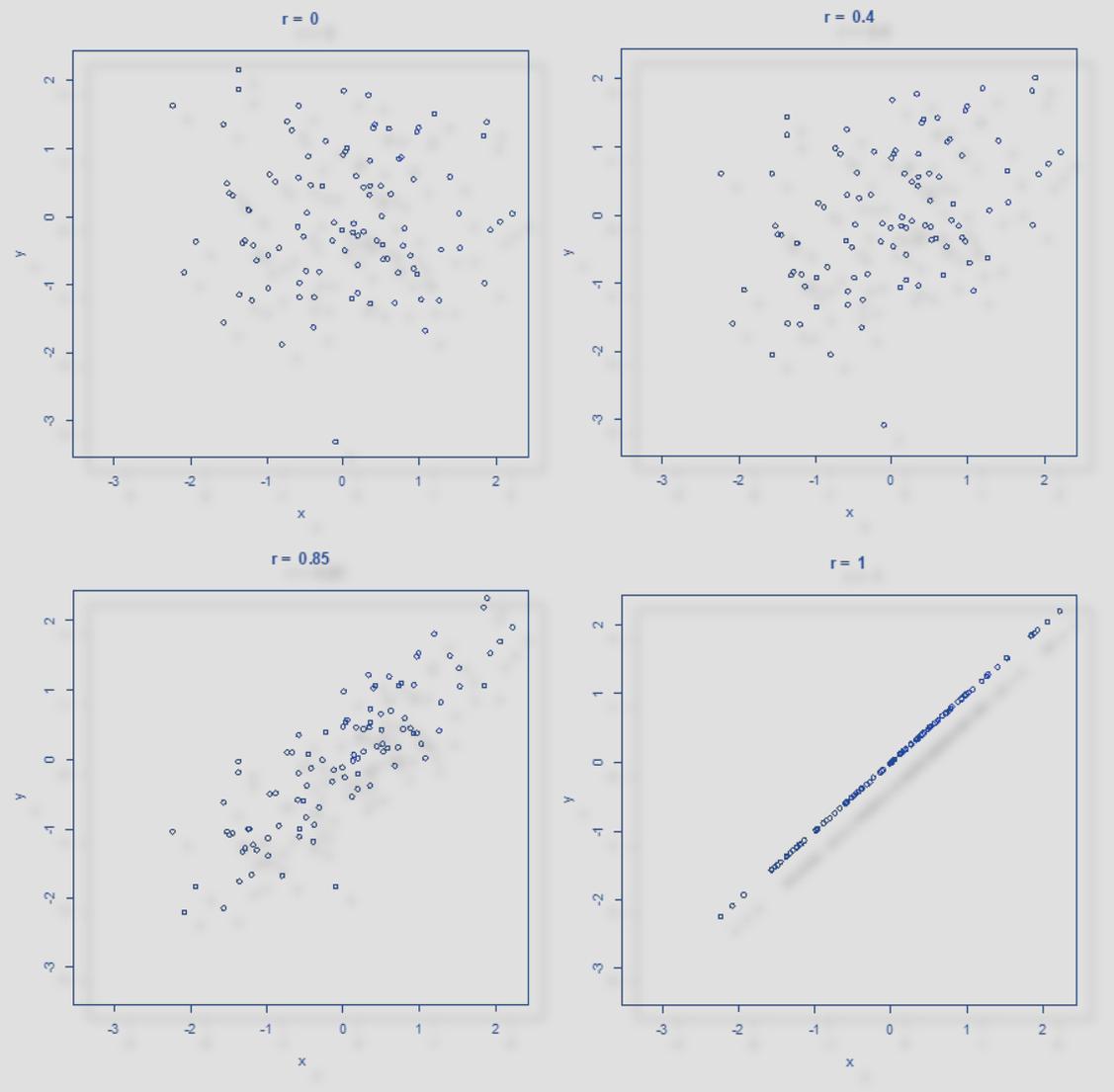
La recta de ajuste muestra una pendiente positiva, con la mayoría de los puntos relativamente cercanos a dicha recta, lo que evidencia una clara relación lineal directa entre ambas variables y con muy pocos casos que se alejan de este patrón.

El índice de correlación r de Pearson para evaluar la relación lineal entre dos variables cuantitativas

La relación lineal entre dos variables cuantitativas se describe a través del índice estadístico de correlación lineal de Pearson (también conocido como r de Pearson). El valor de este índice puede oscilar entre -1 y $+1$:

- Valores próximos a -1 indican que existe una marcada relación lineal negativa, es decir, que a medida que aumentan o disminuyen los valores de una variable disminuyen o aumentan en sentido contrario, en promedio, los valores de la otra variable.
- Valores próximos a $+1$ indican que hay una fuerte relación lineal positiva, es decir, que a medida que aumentan o disminuyen los valores de una variable aumentan o disminuyen en el mismo sentido también, en promedio, los valores de la otra variable.
- Valores próximos a 0 indican la ausencia de relación lineal entre las dos variables.

Los siguientes diagramas de dispersión ilustran, además de la interpretación comentada, el hecho de que a medida que el valor del estadístico r de Pearson se aproxima a 1 (o a -1) también disminuye la dispersión de la nube de puntos alrededor de una hipotética línea recta cuya pendiente refleja el sentido y la intensidad de la relación lineal, reflejando con más claridad la asociación entre las dos variables.



4. Para obtener el índice de correlación de Pearson ejecuta el menú **Analyses | Regression | Correlation Matrix**.

5. Selecciona las variables *NumCig* y *Cerve*, y en **Additional Options** marca la opción **N** y desmarca **Report significance**.

Correlation Matrix

		Numcig	Cerve
Numcig	Pearson's r	—	
	N	—	
Cerve	Pearson's r	0.819	—
	N	27	—

La *correlación de Pearson* entre el consumo semanal de tabaco y de cervezas es de 0.819, valor muy próximo a 1 que indica, por tanto, que existe una relación lineal directa entre ambos consumos, de tal forma que cuanto más se fuma más cervezas se beben (y viceversa).

Medidas de tamaño del efecto (*Effect size, ES*)

Tras obtener e interpretar los valores de los índices estadísticos de relación entre variables, el siguiente consiste en valorar la **relevancia práctica** de los resultados obtenidos. La forma más habitual de abordar esta cuestión consiste en obtener e interpretar un índice estadístico denominado genéricamente "**tamaño del efecto**".

Los tamaños del efecto son índices estadísticos estandarizados, es decir, libres de unidades de medida, aplicables, entre otras, en las siguientes situaciones:

- Para valorar la **intensidad de la asociación entre dos variables categóricas** mediante *odds ratios*.
- Para valorar la **magnitud de la diferencia entre dos medias** (o medianas) mediante el índice *d* de Cohen (o cualquiera de sus variantes, como la *g* de Hedges o la Δ -Delta- de Glass), al evaluar la **relación entre una variable categórica y una variable cuantitativa**.
- Para valorar la **intensidad de la asociación entre dos variables cuantitativas** mediante un coeficiente de correlación *r*.
- Para valorar la **magnitud del efecto de una variable cuantitativa o categórica binaria en un análisis de regresión múltiple** (o de análisis de la variancia).
- Para valorar la **cantidad de variabilidad de una variable cuantitativa atribuible a una (o varias) variable/s categórica/s** mediante índices η^2 (o cualquiera de sus variantes, como la ω^2).

Interpretación de los índices de tamaño del efecto

Cohen (1988, 1992) y Rosenthal (1996) proponen los siguientes criterios para interpretar los valores de las medidas de tamaño del efecto, que constituyen la referencia más extendida actualmente en el ámbito de las ciencias de la salud:

Effect size measure	Small effect size	Medium effect size	Large effect size	Very large effect size
Odds ratio	1.5	2.5	4	10
Cohen's <i>d</i> (or one of its variants)	0.20	0.50	0.80	1.30
<i>r</i>	0.10	0.30	0.50	0.70
Cohen's <i>f</i>	0.10	0.25	0.40	—
Eta-squared	0.01	0.06	0.14	—

^aCohen, 1992, 1988; Rosenthal, 1996.

Obtención de los índices de tamaño del efecto en jamovi

El módulo base de jamovi incluye en los paneles de opciones el cálculo del tamaño del efecto para algunos de los análisis estadísticos. Se pueden extender las opciones para obtener los índices estadísticos de tamaño del efecto, así como representaciones gráficas específicas que facilitan su interpretación, mediante la instalación del **módulo esci (Effect Sizes and Confidence Intervals for R and jamovi)**, propuesta en el apartado 0.1 como una de las librerías básicas para trabajar con jamovi.

6. La valoración de la intensidad (tamaño del efecto) de la relación entre *NumCig* y *Cerve* es sencilla, por el hecho que el índice de correlación lineal r de Pearson es un estadístico estandarizado interpretable directamente como tamaño del efecto. Siguiendo la propuesta presentada en el cuadro de información anterior para los índices estadísticos de la familia r de correlaciones, la magnitud de la relación entre *NumCig* y *Cerve* es muy grande ($r = .82$, valor que supera el criterio $.70$ para un *very large effect size*).

3.4. Descripción estadística de la relación entre dos variables categóricas

Vamos a revisar ahora los gráficos e índices estadísticos de uso más habitual para estudiar la relación lineal entre dos variables categóricas. Para ello, estudiaremos la relación entre las variables categóricas *Genero* y *Obesidad*.

Recomendaciones sobre la construcción de tablas de contingencia

La relación entre dos variables categóricas se analiza a través de tablas de doble entrada denominadas **tablas de contingencia**. Cada casilla de la tabla de contingencia contiene el número de participantes que forma parte de una combinación de categorías de las dos variables.

Cuando se construye la tabla, se aconseja seguir el mismo procedimiento para ubicar las variables que se aplica cuando éstas son cuantitativas: se situarán en columnas las categorías de la *variable independiente, exposición o pronóstica* de la relación, y en filas las categorías de la *variable dependiente o de respuesta*. Con esta disposición de las variables en base a su rol, la interpretación de los resultados requiere calcular los porcentajes respecto al total de cada columna. Por último, si las categorías tienen un orden hay que situarlas en sentido creciente, a partir del vértice inferior izquierdo de la tabla, tal y como muestra a continuación:

Variable dependiente o de respuesta	Variable independiente, de exposición o pronóstica		
	Categoría 1	Categoría 2	Categoría 3
Categoría 1	10 (20%)	20 (40%)	40 (80%)
Categoría 0	40 (80%)	30 (60%)	10 (20%)
TOTAL	50 (100%)	50 (100%)	50 (100%)

Construiremos primero la tabla de contingencia manualmente, para familiarizarnos con ella. Para ello, en las siguientes figuras se muestra el número de casos con cada nivel diagnóstico de *Obesidad* separadamente para cada *Genero*. A la izquierda se muestran los resultados para la submuestra del género masculino y a la derecha para la del género femenino:

Obesidad (diagnóstico)	Freq.		Obesidad (diagnóstico)	Freq.	
Normopeso	9	*****	Normopeso	5	*****
Sobrepeso	5	*****	Sobrepeso	8	*****
Obesidad	2	**	Obesidad	1	*
Total	16		Total	14	

1. Completa en la siguiente tabla el número de participantes en el espacio previsto con ____ y añade el resultado del cálculo de los porcentajes por columnas entre paréntesis (____%):

	Masculino	Femenino
Normopeso	____ (____ %)	____ (____ %)
Sobrepeso	____ (____ %)	____ (____ %)
Obesidad	____ (____ %)	____ (____ %)
Total	16 (100%)	14 (100 %)

2. Veamos cómo se obtiene la tabla de contingencia con jamovi. Ejecuta el menú **Analyses | Frecuencias | Contingency Tables | Independent Samples**. En el panel de opciones, selecciona la variable *Obesidad* en filas (Rows) y la variable *Genero* en columnas (Columns). Despliega, a continuación, el bloque **Cells** para marcar **Column** en **Percentages**, y desmarca las opciones χ^2 y **Confidence intervals** en el bloque **Statistics**. Comprueba si los valores que has escrito en la tabla anterior son correctos.

Contingency Tables

Obesidad		Genero		Total
		Masculino	Femenino	
Normopeso	Observed	9	5	14
	% within column	56.3 %	35.7 %	46.7 %
Sobrepeso	Observed	5	8	13
	% within column	31.3 %	57.1 %	43.3 %
Obesidad	Observed	2	1	3
	% within column	12.5 %	7.1 %	10.0 %
Total	Observed	16	14	30
	% within column	100.0 %	100.0 %	100.0 %

En la tabla de contingencia los porcentajes respecto al total de columnas indican que existe una relación estadística entre las dos variables analizadas, ya que se observa una diferente distribución del diagnóstico de la obesidad entre ambos géneros: en los hombres el porcentaje de participantes con peso normal (56,3%) es superior al de las mujeres (35,7%); así mismo, el porcentaje de casos con sobrepeso es superior en las mujeres (57,1%) que en los hombres (31,3%); por último, hay un porcentaje superior de obesos entre los hombres (12,5%) que entre las mujeres (7,1%).

Evaluación del tamaño del efecto de la asociación entre dos variables categóricas

La evaluación del tamaño del efecto al estudiar la relación entre dos variables categóricas se puede realizar mediante un conjunto de indicadores, como la **diferencia de proporciones (o porcentajes, DP)** (también denominados diferencia -o reducción- absoluta de riesgo, exceso de riesgo o riesgo atribuible), la **razón de proporciones o riesgo relativo (RR)**, la **razón de ventajas (odds ratio, OR)**, el **número necesario a tratar (number needed to treat, NNT)**, el **número necesario de cambios en no expuestos (number needed to be exposed, NNE)** o el **número necesario de cambios en expuestos (exposure impact number, EIN)**.

La interpretación de estos índices es en muchos casos complementaria, es decir, expresan un mismo efecto desde diferentes puntos de vista. Algunos de ellos son específicos para determinados diseños de investigación (e.g., RR para diseños de cohortes o transversales, OR para todo tipo de diseños -incluso de casos y controles-, NNT para diseños experimentales -aleatorizados-, NNE y EIN para diseños no experimentales longitudinales).

Los tamaños del efecto adecuados en este caso para valorar la magnitud de la asociación entre *Genero* y *Obesidad* ofrecen los siguientes resultados:

- La diferencia de porcentajes en el grupo de participantes diagnosticados de “obesidad” muestra un 5.4% ($DP = 12.5\% - 7.1\% = 5.4\%$) más de hombres que mujeres con este diagnóstico, y un 25.8% ($DP = 31.3\% - 57.1\% = -25.8\%$) menos de hombres que mujeres con “sobrepeso”.
- Las comparaciones anteriores ofrecerían valores de riesgo relativo de 1.76 ($RR = 12.5 / 7.1 = 1.76$) y 0.55 ($RR = 31.3 / 57.1 = 0.55$), respectivamente. Estos valores indican que la proporción de hombres con diagnóstico de “obesidad” es 1.76 veces mayor que la de mujeres, es decir, un 76% ($[1.76 - 1] * 100 = 76\%$) superior; contrariamente, la proporción de hombres con diagnóstico de “sobrepeso” es aproximadamente la mitad que la de mujeres, es decir, un 45% ($[0.55 - 1] * 100 = -45\%$) inferior.

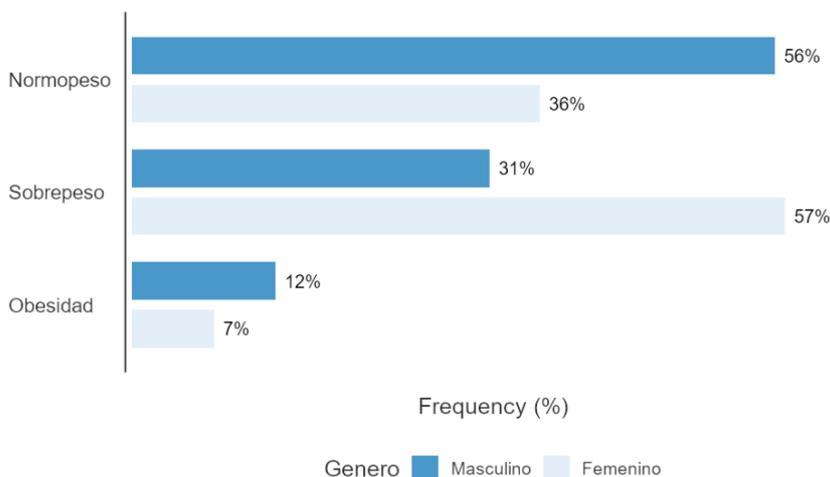
Representación gráfica de la asociación entre dos variables categóricas

Para describir gráficamente la relación entre dos variables categóricas se suele utilizar el **diagrama de barras agrupadas** y el **diagrama de barras apiladas (al 100% o no)**. Cuando las variables son ordinales también se pueden utilizar **diagramas de líneas múltiples** y **diagramas de áreas apiladas**.

3. Para obtener el gráfico de barras agrupadas, ejecuta el menú **Analyses | Exploration | Survey Plots**. Pasa la variable *Obesidad* a **Variables** y la variable *Genero* a **Grouping Variable**. A continuación, despliega **Nominal / Ordinal Plots** y marca **Percentages** en **Frequency Type**:

Obesidad

Diagnostico de obesidad segun la clasificacion de la SEEDO

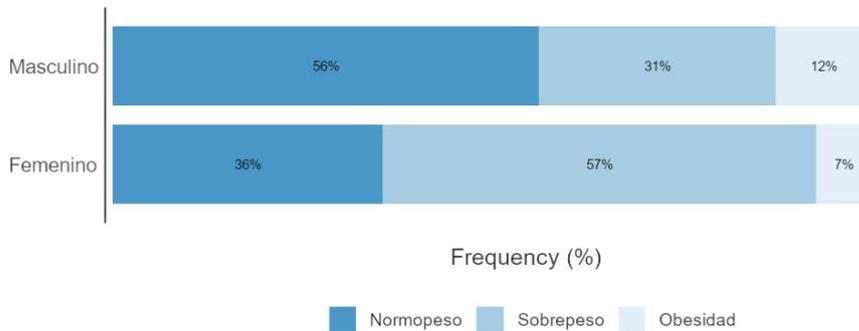


Este diagrama permite ver estas diferencias de forma rápida observando las alturas de las barras de hombres y mujeres en cada categoría diagnóstica de la obesidad. Puedes comprobar que los porcentajes de las tres barras de cada género suman el 100%.

- Otra forma de representar el gráfico de barras es hacerlo mediante barras apiladas (*stacked bars*). En el panel de opciones, marca **Stacked bar** en **Plot Type** para cambiar el tipo de gráfico:

Obesidad

Diagnostico de obesidad segun la clasificacion de la SEEDO



3.5. Descripción estadística de la relación entre una variable categórica y una variable cuantitativa

Por último, en este apartado se analiza gráfica y estadísticamente la relación entre una variable categórica y una variable cuantitativa. Para ello, utilizaremos los ejemplos del análisis de la relación entre la variable categórica binaria *Genero* y el número de cigarrillos que fuman diariamente los participantes (*NumCig*), y entre la variable categórica ordinal *Obesidad* y el número de unidades de bebidas alcohólicas que toman semanalmente (*NumBebidas2*).

Índices estadísticos para estudiar la relación entre una variable categórica y una cuantitativa

La estrategia más habitual para estudiar la relación entre una variable categórica y una variable cuantitativa consiste en comparar el índice de tendencia central (e.g., la media aritmética o la mediana) de la variable cuantitativa en cada grupo de casos definidos por las categorías de la variable categórica.

Representación gráfica de la relación entre una variable categórica y una variable cuantitativa

Los tipos de gráficos que permiten visualizar este tipo de relación cuando la variable cuantitativa es discreta son los **diagramas de barras agrupadas** y los **diagramas de líneas múltiples** (también los **diagramas de barras y de áreas apiladas**); cuando la variable cuantitativa es continua se utiliza habitualmente el **diagrama de barras de error**. En todos los casos es aconsejable obtener el **diagrama de caja múltiple**.

3.5.1. Relación entre una variable categórica binaria y una variable cuantitativa

Para familiarizarnos con el procedimiento, antes de obtener los resultados con jamovi vamos a realizar manualmente los cálculos de los promedios del número de cigarrillos que fuman diariamente los participantes en las submuestras de hombres y mujeres. En el siguiente listado se muestran, para cada género y en orden ascendente, los valores válidos de la variable *NumCig* de los participantes fumadores. Calcula y anota la media y la mediana de consumo de tabaco de las personas fumadoras de este estudio:

-> Genero = Masculino

Caso	Numcig
9	1
3	1
2	5
15	6
27	7
5	8
1	8
14	8
20	10
32	10
26	15
8	15
N	12

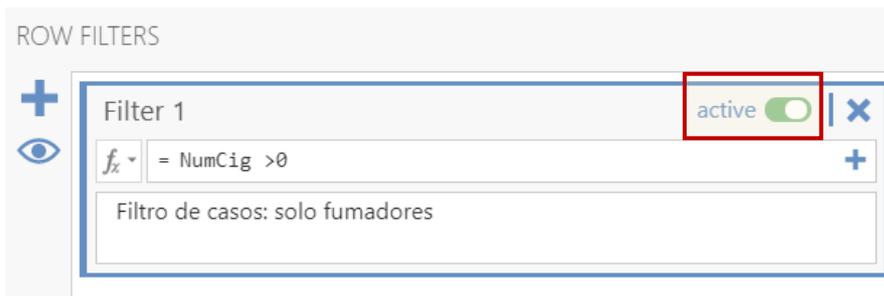
-> Genero = Femenino

Caso	Numcig
18	4
7	8
28	10
31	10
21	10
6	11
12	12
30	14
10	15
16	18
4	20
N	11

Media aritmética: _____

Mediana: _____

1. Para describir adecuadamente las diferencias entre el consumo de tabaco de hombres y mujeres, se debe aplicar el mismo filtro de datos que aplicamos para describir individualmente la variable *NumCig*, para que sólo se incluyan los participantes fumadores con información válida en *NumCig*. Ejecuta el menú **Data | Filters** y activa de nuevo el **Filter 1** definido anteriormente:



2. A continuación, ejecuta el menú **Analyses | T-test | Independent Samples T-Test**. En el panel de opciones, pasa la variable *Genero* a **Grouping Variable** y la variable *NumCig* a **Dependent Variable**, y marca las casillas **Mean difference**, **Effect size**, **Descriptives** y **Descriptives plots** en el bloque **Additional Statistics**:

Independent Samples T-Test

		Statistic	df	p	Mean difference	SE difference	Effect Size
NumCig	Student's t	-2.22	21.0	0.037	-4.17	1.87	Cohen's d -0.928

Diferencia entre los promedios de los dos grupos

Medida del tamaño del efecto

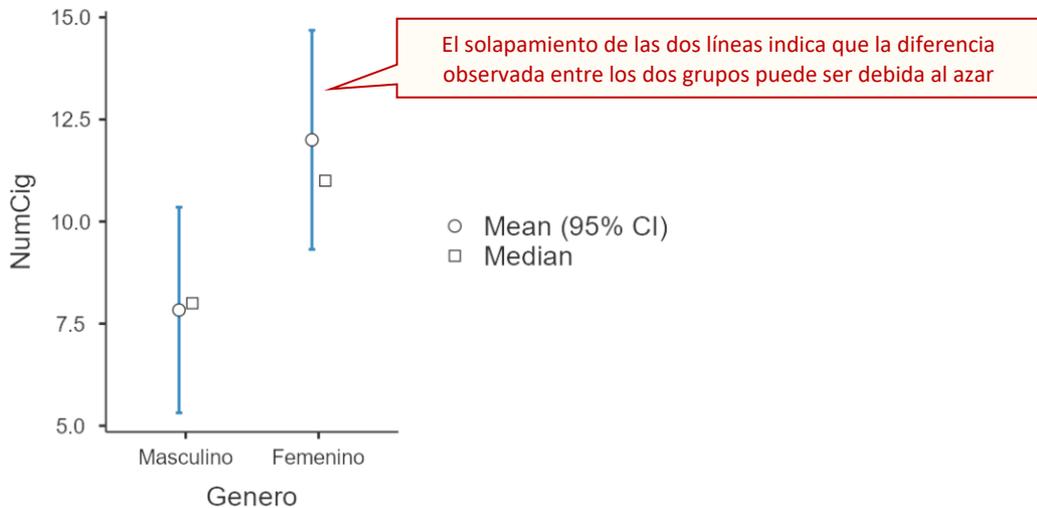
Group Descriptives

		Group	N	Mean	Median	SD	SE
NumCig	Masculino		12	7.83	8.00	4.45	1.28
	Femenino		11	12.0	11.0	4.54	1.37

Medidas de tendencia central de los dos grupos

Plots

NumCig



La media aritmética del consumo de tabaco es superior en las mujeres fumadoras (12 cig/día) que en los hombres fumadores (7.83 cig/día), lo que supone una diferencia entre los promedios de hombres y mujeres fumadores de 4.17 cigarrillos diarios. Por tanto, podemos concluir que existe una relación entre la variable *Genero* y la variable *NumCig*, en el sentido que las mujeres fumadoras fuman diariamente una mayor cantidad de cigarrillos que los hombres.

El valor -0.928 del índice de tamaño del efecto *d* de Cohen permite interpretar dicha diferencia como de magnitud grande por ser superior al criterio *large* propuesto por Cohen (1988) (ver el siguiente cuadro de información).

Evaluación del tamaño del efecto en la relación entre una variable categórica y una cuantitativa

Generalmente, la evaluación del tamaño del efecto al estudiar la relación entre una variable categórica y una variable cuantitativa se lleva a cabo mediante el cálculo e interpretación del índice estandarizado ***d* de Cohen** (o alguna de sus variantes, como las Δ -*Delta*- de Glass para estudios con diseños experimentales, o la más general *g* de Hedges, especialmente útil en meta-análisis).

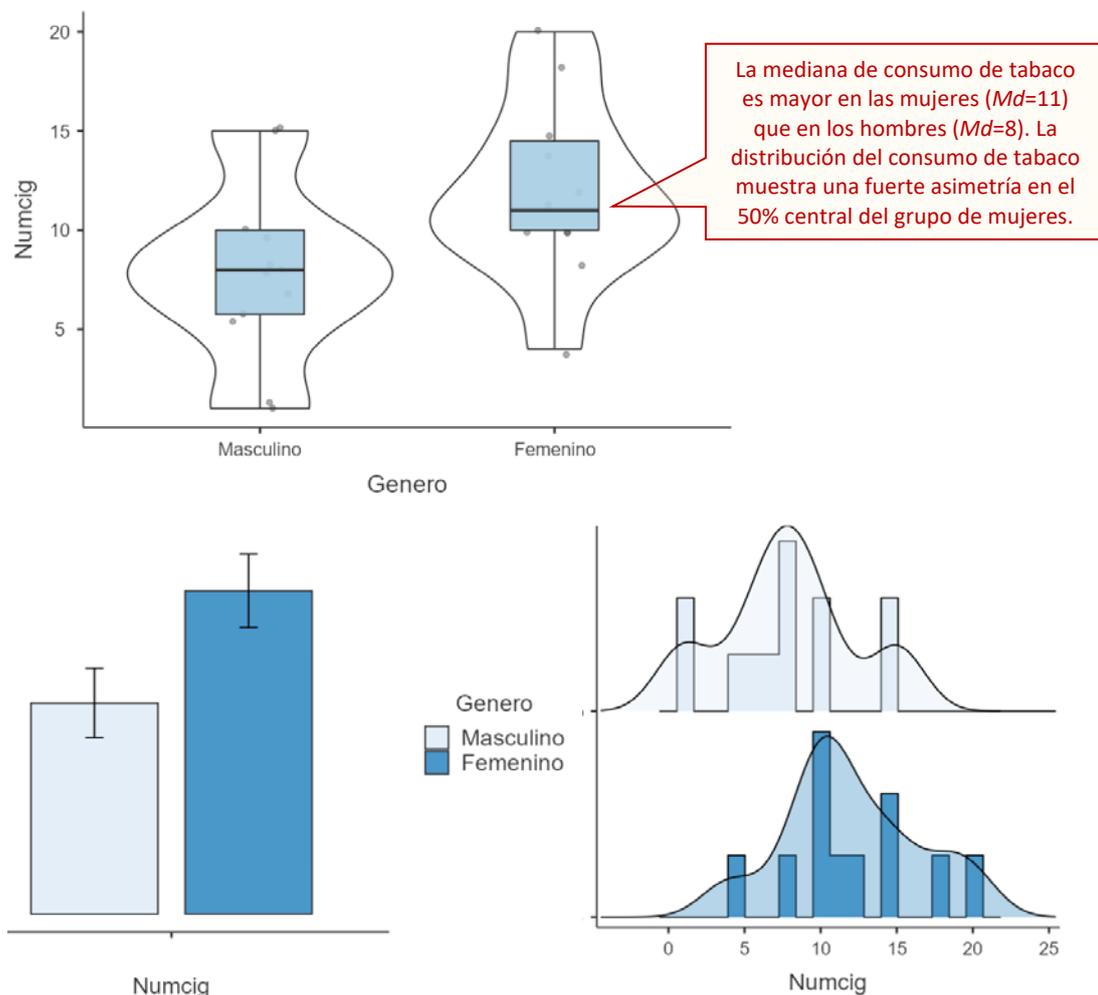
Se trata de índices que permiten valorar la magnitud del efecto de la variable categórica sobre la variable cuantitativa, a través de la estandarización de las diferencias entre los índices de tendencia central obtenidos en los grupos definidos por la variable categórica. Su interpretación, como ya se ha comentado, se realiza habitualmente comparando dichos índices con los valores criterio propuestos por Cohen (1988) y Sawilowsky (2009):

Effect size	Very small	Small	Medium	Large	Very large	Huge
d	0.01	0.20	0.50	0.80	1.20	2.0
η^2		0.01	0.06	0.14		

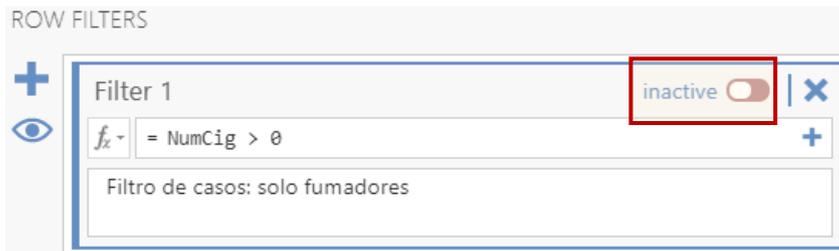
En palabras de Cohen (1992):

The ES index for the test of the difference between independent means is d , the difference expressed in units of (i.e., divided by) the within-population standard deviation. For this test, the $d = 0$ indicates that the difference equals zero; and the small, medium, and large ESs are $d = .20$, $.50$, and $.80$. Thus, an operationally defined medium difference between means is half a standard deviation; concretely, for IQ scores in which the population standard deviation is 15, a medium difference between means is 7.5 IQ points.

3. Vamos a obtener ahora otras representaciones gráficas para visualizar las diferencias entre hombres y mujeres en cuanto a su consumo diario de cigarrillos. Ejecuta el menú **Analyses | Exploration | Descriptives**. En el panel de opciones pasa la variable *NumCig* a **Variab**les y la variable *Genero* a **Split by**. A continuación, marca todas las opciones excepto Q-Q en **Plots**:



- Finalmente, ejecuta de nuevo **Data | Filters** para desactivar el filtro de casos:



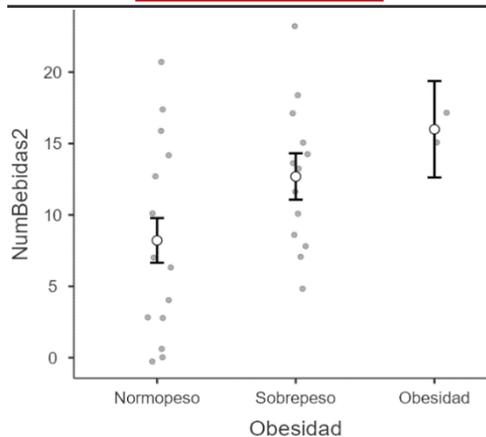
3.5.2. Relación entre una variable categórica ordinal o politómica y una variable cuantitativa

Por último, vamos a analizar la relación entre una variable cuantitativa y una variable categórica con más de dos categorías. En concreto, vamos a estudiar si existen diferencias entre los tres grupos de participantes definidos a partir de la variable diagnóstico de *Obesidad* y el número de unidades de bebidas alcohólicas que toman en promedio semanalmente (variable *NumBebidas2*).

- Ejecuta el menú **Analyses | ANOVA | ANOVA** y pasa la variable *NumBebidas2* a **Dependent Variables** y la variable *Obesidad* a **Fixed Factors**. Despliega el bloque **Estimated Marginal Means** para que jamovi muestre el gráfico de medias de los grupos con su barra de error y la nube de puntos, así como la tabla de descriptivos de *NumBebidas2* para los grupos definidos por las categorías de *Obesidad*. Pasa la variable categórica *Obesidad* como **Term 1** de **Marginal Means** y, a continuación, selecciona las opciones de **Output Marginal means plots** y **Marginal means tables**, y las opciones **Standard Error** en la lista **Error Bars** y **Observed scores** del bloque **Plot**.

Estimated Marginal Means - Obesidad

Obesidad	95% Confidence Interval			
	Mean	SE	Lower	Upper
Normopeso	8.21	1.56	5.00	11.4
Sobrepeso	12.69	1.62	9.36	16.0
Obesidad	16.00	3.38	9.07	22.9



Las medias aritméticas del número de vasos de vino, cervezas y copas que toman semanalmente los participantes del estudio es superior en los diagnosticados de obesidad (16 vasos/copas), seguidos de los diagnosticados de sobrepeso (12.69 vasos/copas). El grupo de participantes con diagnóstico de normopeso presenta el menor consumo promedio (8.21 vasos/copas).

El gráfico de medias marginales muestra claramente las diferencias comentadas entre las medias de los tres grupos de participantes. Observa, empero, que la dispersión que refleja la nube de puntos es bastante grande en los dos primeros grupos, y pequeña en el grupo de participantes con diagnóstico de obesidad, aunque en el grupo de diagnosticados de obesidad este aspecto no se puede valorar adecuadamente por el hecho de que sólo contiene los datos del consumo de unidades de bebidas alcohólicas de tres sujetos.

- En el panel principal de opciones marca la opción η^2 en **Effect Size**. A continuación, despliega el bloque de opciones **Post Hoc Tests**, pasa la variable *Obesidad* a la derecha, desmarca en **Correction** la opción **Tukey** y marca en **Effect Size** la opción **Cohen's d**.

ANOVA - NumBebidas2

	Sum of Squares	df	Mean Square	F	p	η^2
Obesidad	221	2	110.4	3.22	0.056	0.193
Residuals	925	27	34.3			

Post Hoc Comparisons - Obesidad

Comparison		Mean Difference	SE	df	t	Cohen's d
Obesidad	Obesidad					
Normopeso	- Sobrepeso	-4.48	2.25	27.0	-1.986	-0.765
	- Obesidad	-7.79	3.72	27.0	-2.091	-1.330
Sobrepeso	- Obesidad	-3.31	3.75	27.0	-0.882	-0.565

Note. Comparisons are based on estimated marginal means

En la primera tabla de resultados aparece el valor del índice de magnitud del efecto $\eta^2 = 0.193$, que indica que, según los criterios propuestos por Cohen (1988), existen diferencias de gran magnitud entre al menos dos de los grupos definidos por el diagnóstico de obesidad respecto al número de unidades de bebidas alcohólicas que consumen semanalmente los participantes.

Según esos mismos criterios, los valores del índice de tamaño del efecto d de Cohen que se muestran en la segunda tabla de resultados indican que la diferencia entre el grupo con normopeso y el grupo con diagnóstico de sobrepeso es de magnitud moderada ($d = -0.765$), y de magnitud muy grande ($d = -1.33$) respecto al grupo diagnosticado de obesidad. Entre el grupo con sobrepeso y el grupo con diagnóstico de obesidad también existe una diferencia de magnitud moderada ($d = -0.565$).

- Guarda todos los análisis realizados en este apartado en el fichero *HABITOX_P3User.omv*.

3.6. Ejercicio

Abre el fichero de datos *HABITOX_P2Sol.omv* para resolver las cuestiones planteadas en este ejercicio. Cuando acabes puedes consultar las soluciones a las cuestiones planteadas abriendo el fichero *HABITOX_P3Sol.omv*.

Parte 1. Descriptiva univariante

- ¿Los participantes que han respondido al cuestionario sobre hábitos tóxicos toman con la misma frecuencia los tres tipos de droga (marihuana, alucinógenos y opiáceos) sobre los que se interroga en el estudio? Anota tu respuesta:

- Entre los participantes que han respondido a la pregunta sobre consumo de opio, ¿cuál es el porcentaje de participantes que toman esta droga ocasional o habitualmente? Anota tu respuesta:

3. Construye con jamovi los gráficos más adecuados para representar la distribución de consumo de opio de los participantes que han respondido a esta pregunta. Anota tu respuesta:

4. Construye con jamovi los gráficos estadísticos más adecuados para representar la distribución de la edad de los participantes de este estudio. Anota tu respuesta:

5. ¿Cuáles son los índices de tendencia central y de dispersión más adecuados para describir la edad de los participantes de este estudio? Anota tu respuesta:

6. ¿Cuál es el valor mínimo del IMC del 25% de participantes con IMC más alto? ¿Y el valor máximo del 25% con IMC más bajo? Anota tu respuesta:

7. ¿Cuáles son los índices estadísticos y gráficos más adecuados para describir el número de vasos de vino que los participantes beben en promedio semanalmente? Anota tu respuesta:

8. ¿Cuántos participantes del estudio beben 2 o más vasos de vino en promedio a la semana? ¿Qué porcentaje representan respecto al total de participantes? Anota tu respuesta:

9. ¿Cuáles son los índices estadísticos y gráficos más adecuados para describir la variable *NumDrogas*?
Anota tu respuesta:

Parte 2. Descriptiva bivariante

10. ¿Existen diferencias entre los hombres y mujeres del estudio en cuanto a su hábito de fumar (variable binaria *Fuma*)? ¿Cuál sería el gráfico más adecuado para analizar esta relación? Anota tu respuesta:

11. ¿Hombres y mujeres consumen alucinógenos con la misma frecuencia? Anota tu respuesta:

12. ¿Los fumadores consumen marihuana con mayor frecuencia que los no fumadores? Anota tu respuesta:

13. ¿Existen diferencias entre los hombres y las mujeres de la muestra de estudio en cuanto a su edad? ¿Cuál sería el gráfico más adecuado para analizar la relación entre estas dos variables? Anota tu respuesta:

14. Entre los participantes que beben cervezas, ¿beben la misma cantidad semanalmente los hombres que las mujeres? Anota tu respuesta:

15. Analiza estadísticamente (numérica y gráficamente) la relación entre el consumo de cervezas y el IMC. ¿Qué conclusión puedes extraer? Anota tu respuesta:

16. ¿Con qué se asocia más el consumo de marihuana, con el consumo de cigarrillos o con el de copas? Anota tu respuesta:

17. ¿Existe relación entre el consumo de los tres tipos de bebidas alcohólicas sobre los que se interroga en este estudio? Anota tu respuesta:

3.7. Síntesis de menús y opciones jamovi para la descripción estadística uni y bivalente

A0. Filtro de datos

1. Seleccionar filas de datos para el análisis: Data | Filters [pág. 33]

- ✓ Row Filters: expresión lógica de filtro
- ✓ Active / Inactive

A1. Descripción estadística de una variable categórica

1. Gráfico de barras: Analyses | Exploration | Descriptives [pág. 31]

- ✓ Variables: variable categórica
- ✓ Plots | Bar Plots: Bar plot

2. Gráfico de barras agrupadas o apiladas: Analyses | Exploration | Survey Plots [pág. 32]

- ✓ Variables: variable categórica
- ✓ Nominal / Ordinal Plots | Plot Type: Grouped bar o Stacked bar
- ✓ Nominal / Ordinal Plots | Frequency Type: Percentages
- ✓ Nominal / Ordinal Plots | Additional Options: (opcional: Hide missing values)

3. Diagrama de Pareto: Analyses | Exploration | Pareto Chart [pág. 35]

- ✓ X-Axis: variable categórica

4. Índices estadísticos: Analyses | Exploration | Descriptives [pág. 35]

- ✓ Variables: variable categórica
- ✓ Frequency tables
- ✓ Statistics | Sample size: N, Missing
- ✓ Statistics | Central Tendency: desmarcar Mean, Median
- ✓ Statistics | Dispersion: desmarcar Std. deviation, Minimum, Maximum

A2. Descripción estadística de una variable cuantitativa

1. Histograma: Analyses | Exploration | Descriptives [pág. 36]

- ✓ Variables: variable cuantitativa
- ✓ Plots | Histograms: Histogram (opcional: Density)

2. Diagrama de caja (*boxplot*): Analyses | Exploration | Descriptives [pág. 38]

- ✓ Variables: variable cuantitativa
- ✓ Plots | Box Plots: Box Plot (opcional: Violin y/o Data)

3. Índices estadísticos: Analyses | Exploration | Descriptives [pág. 36]

- ✓ Variables: variable cuantitativa
- ✓ Sample size: N, Missing
- ✓ Statistics | Central Tendency: Mean, Median, Mode
- ✓ Statistics | Percentile Values: Percentiles
- ✓ Statistics | Dispersion: Std. deviation, Minimum, Maximum

A3. Relación lineal entre dos variables cuantitativas

1. Diagrama de dispersión (<i>scatter-plot</i>): Analyses Exploration Scatterplot [pág. iError! Marcador no definido.]
✓ X-Axis: variable X (independiente o exposición) ✓ Y-Axis: variable Y (dependiente o respuesta) ✓ Regression Line: Linear
2. Índice de correlación lineal de Pearson: Analyses Regression Correlation Matrix [pág. 44]
✓ Additional Options: marcar N y desmarcar Report significance ⇒ Tamaño del efecto: small: $r \geq 0.10$; medium: $r \geq 0.30$; large: $r \geq 0.50$; very large: $r \geq 0.70$

A4. Relación entre dos variables categóricas

1. Gráfico de barras agrupadas o apiladas: Analyses Exploration Survey Plots [pág. 48]
✓ Variables: variable Y (dependiente o respuesta) ✓ Grouping variable: variable X (independiente o exposición) ✓ Nominal / Ordinal Plots Plot Type: Grouped bar o Stacked bar
2. Tablas de contingencia y comparación de proporciones: Analyses Frequencies Contingency Tables Independent Samples [pág. 47]
✓ Rows: variable Y (dependiente o respuesta) ✓ Columns: variable X (independiente o exposición) ✓ Statistics Tests: desmarcar χ^2 ✓ Statistics Comparative Measures (2x2 only): desmarcar Confidence intervals ✓ Cells Percentages: Column (para obtener los porcentajes por columnas) ⇒ Tamaño del efecto: diferencia (DP) o razón (RR) entre porcentajes de cada fila

A5. Relación entre una variable categórica y una variable cuantitativa

1. Histograma: Analyses Exploration Descriptives [pág. 52]
✓ Variables: variable cuantitativa ✓ Split by: variable categórica ✓ Plots Histograms: Histogram (opcional: Density)
2. Gráfico de barras: Analyses Exploration Descriptives [pág. 52]
✓ Variables: variable cuantitativa ✓ Split by: variable categórica ✓ Plots Bar Plots: Bar Plot
3. Diagrama de caja múltiple (<i>multiple boxplot</i>): Analyses Exploration Descriptives [pág. 52]
✓ Variables: variable cuantitativa ✓ Split by: variable categórica ✓ Plots Box Plots: Box Plot (opcional: Violin y/o Data)
4. Comparación de medias (variable categórica binaria): Analyses T-test Independent Samples T-Test [pág. 50]
✓ Grouping Variable: variable categórica ✓ Dependent Variable: variable cuantitativa ✓ Additional Statistics: Mean difference, Effect size, Descriptives y Descriptives plots ⇒ Tamaño del efecto: small: $d \geq 0.20$; medium: $d \geq 0.50$; large: $d \geq 0.80$; very large: $d \geq 1.20$; huge: $d \geq 2.0$
5. Comparación de medias (variable categórica ordinal o politómica): Analyses ANOVA ANOVA [pág. 53]
✓ Fixed Factors: variable categórica ✓ Dependent Variables: variable cuantitativa ✓ Effect size: η^2 ⇒ Tamaño del efecto η^2 : small: $\eta^2 \geq 0.01$; medium: $\eta^2 \geq 0.06$; large: $\eta^2 \geq 0.14$ ✓ Post Hoc Tests Effect Size: Cohen's d ⇒ Tamaño del efecto d: small: $d \geq 0.20$; medium: $d \geq 0.50$; large: $d \geq 0.80$; very large: $d \geq 1.20$; huge: $d \geq 2.0$ ✓ Estimated Marginal Means Marginal Means: Term 1: variable X (independiente o exposición) ✓ Estimated Marginal Means Output: Marginal means plots, Marginal means tables ✓ Estimated Marginal Means Plot: (Error bars) Standard Error, Observed scores

4. Inferencia estadística y contraste de hipótesis

4.1. Fundamentos teóricos

La estadística inferencial permite dar respuesta, a partir del valor de estadísticos calculados con la información de muestras representativas de las poblaciones, tanto a preguntas acerca de cuáles son los valores de parámetros poblacionales de interés, como a preguntas que comparan dichas estimaciones con valores hipotéticos planteados por el investigador. En el primer caso se habla de «**estimación de parámetros**», mientras que en el segundo se habla de «**contraste de hipótesis**», aunque, como veremos, en este último caso la estimación también juega un papel central.

4.1.1. Parámetros y estadísticos

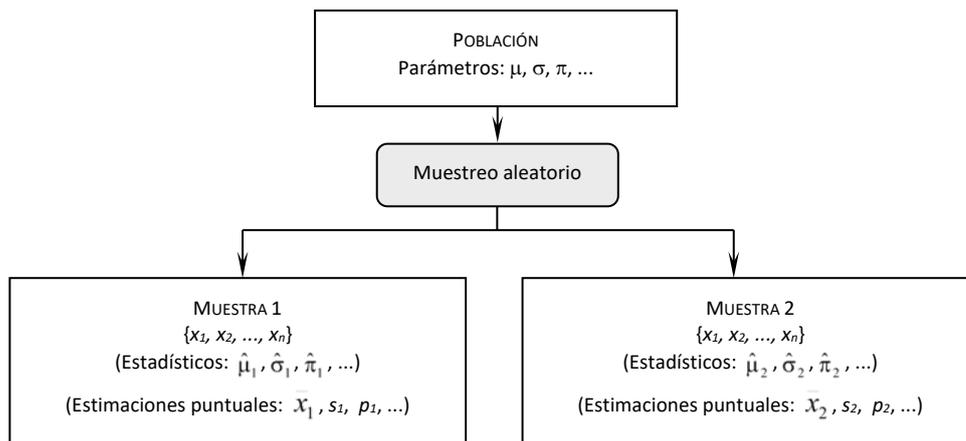
Se denominan «**parámetros**», y se representan con letras del alfabeto griego, a los valores que resumen estadísticamente (mediante una media μ , una proporción π , una desviación estándar σ , etc.) una determinada característica de los sujetos de una población.

Alfabeto griego.

α	β	γ	δ	ϵ	ζ	η	θ	ι	κ	λ	μ
Alfa	beta	gamma	delta	Épsilon	zeta	eta	theta	iota	Kappa	lambda	my
ν	ξ	\omicron	π	ρ	σ	τ	υ	ϕ, ϕ	χ	ψ	ω
Ny	xi	omicron	pi	Rho	sigma	tau	ípsilon	phi	Ji	psi	Omega

Habitualmente el valor de un parámetro poblacional es desconocido, debiendo ser «estimado» a partir del cálculo de una función denominada «estimador» (o también «**estadístico**») que se aplica a los datos observados en una muestra extraída de forma aleatoria de la población, es decir, mediante un procedimiento que garantice que cualquier elemento de la población tiene la misma probabilidad de ser elegido para formar parte de la muestra. Las estimaciones puntuales de los parámetros poblacionales obtenidas a partir de los datos de una muestra se acostumbra a representar mediante otros símbolos como, por ejemplo, \bar{x} o m para la media, s^2 para la variancia, p para la proporción, etc.

La siguiente figura esquematiza el proceso de muestreo aleatorio a partir de una población y la notación descrita.

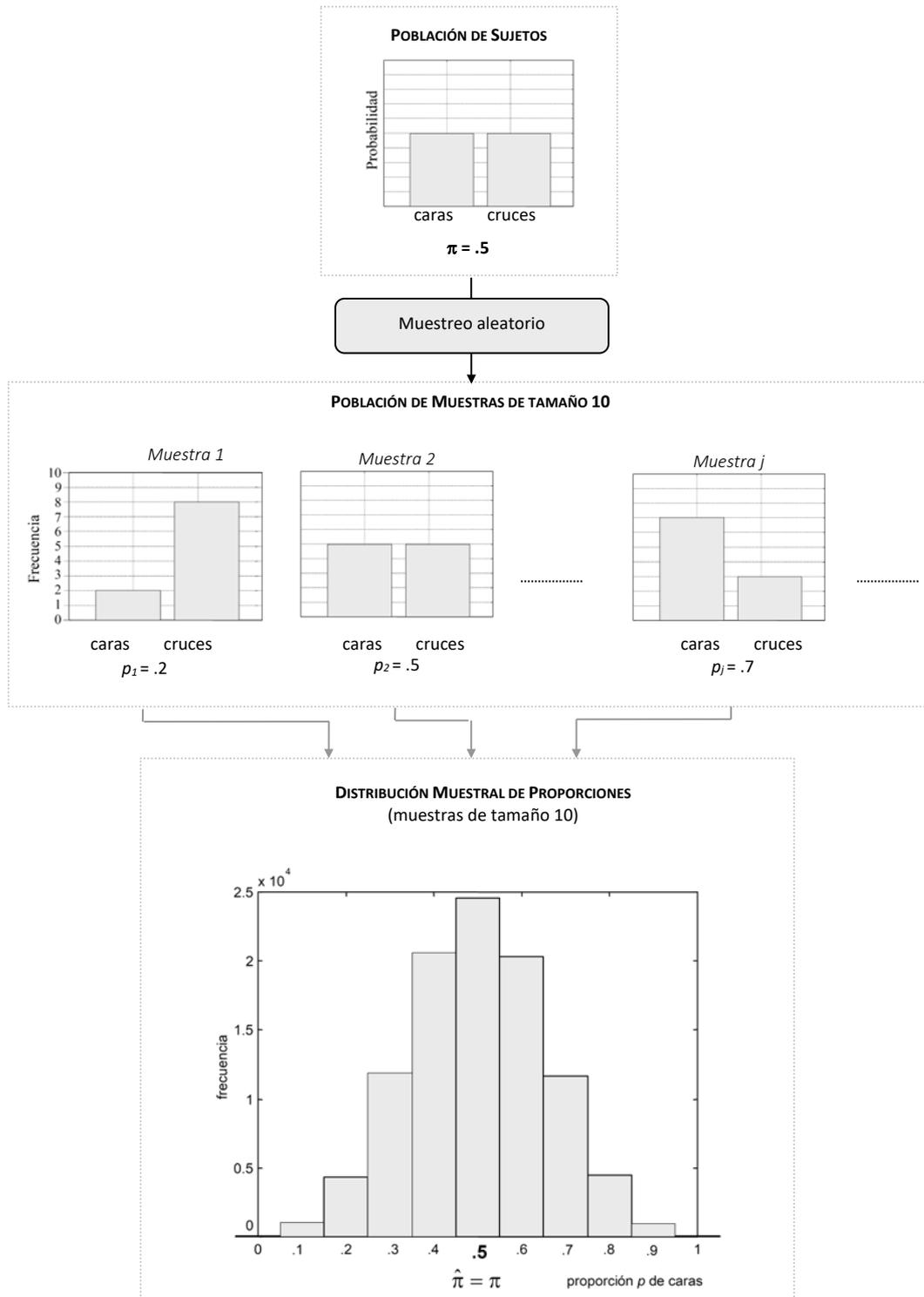


Parámetros, estadísticos y estimaciones puntuales.

4.1.2. Distribución muestral de un estadístico

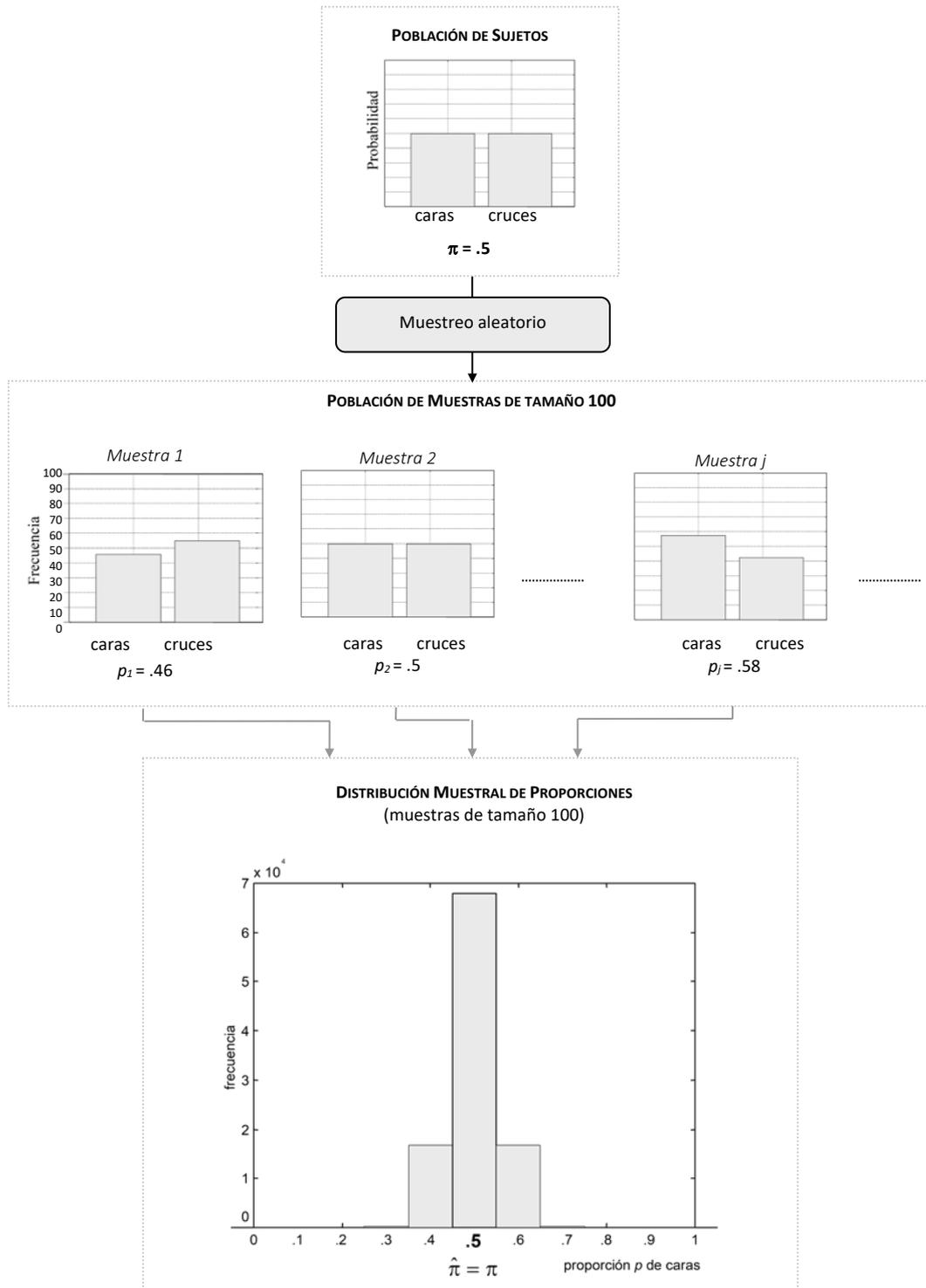
Una estimación puntual será más precisa, es decir, más fiable o informativa, cuanto mayor sea el tamaño de la muestra en relación con el tamaño de la población origen. El sencillo experimento que se expone a continuación pone de manifiesto la necesidad apuntada antes de acompañar las estimaciones puntuales con una medida de su precisión, al tiempo que permite aproximar empíricamente un concepto teórico central en el proceso de inferencia estadística: la «**distribución muestral**» de un estadístico.

Así, por ejemplo, si realizaras (extrajeras) 100 mil series (muestras) de 10 lanzamientos (casos) de una moneda y construyeras el gráfico de frecuencias del estadístico proporción p de *caras* en cada serie, obtendrías la «distribución muestral empírica de la p de *caras*», en la cual podrías observar que, a pesar de que en cada serie de 10 lanzamientos de una moneda se espera obtener 5 *caras* y 5 *cruces*, el error aleatorio asociado (azar) al muestreo da lugar a series de lanzamientos en las que la proporción p de *caras* es distinta a la que cabría esperar tomando en consideración el parámetro $\pi = 0.5$. Fíjate, empero, que la probabilidad de obtener p distintas de 0.5 es menor a medida que se alejan de π .



Distribución muestral (empírica) de la proporción de *caras* obtenidas en 100.000 series de 10 lanzamientos de una moneda (inspirada en la figura presentada por Pardo y San Martín, 1991, p. 62).

Por su parte, la siguiente figura presenta la distribución muestral empírica de la proporción p de *caras* en 100 mil series (muestras) de 100 lanzamientos de la misma moneda. Observando la dispersión de esta distribución puedes constatar que, comparada con la presentada en la figura anterior, al aumentar el tamaño muestral la fluctuación de las estimaciones puntuales es menor y, por tanto, el error máximo asociado a una estimación puntual, considerado como la diferencia entre la proporción p de *caras* observada en una muestra y la proporción teórica $\pi = 0.5$, también es menor.



Distribución muestral (empírica) de la proporción de *caras* en 100.000 series de 100 lanzamientos de una moneda (inspirada en la figura presentada por Pardo y San Martín, 1991, p. 62).

Las estimaciones puntuales son poco útiles en sí mismas, ya que, por el hecho de que varían de una muestra a otra, pueden coincidir o no con el verdadero valor del parámetro poblacional. La existencia de esta fluctuación debida al muestreo hace que sea imposible realizar una estimación puntual «sin error» del valor del parámetro poblacional a partir de los datos de una muestra determinada. Frente a esta situación, la única forma de realizar una estimación sin error sería enunciando que «*el valor del parámetro de interés oscilará entre el valor mínimo posible y el valor máximo posible*». Por ejemplo, si quisieras estimar la proporción de sujetos con una determinada característica en una muestra de $n = 10$ sujetos, sólo el intervalo (0 a 1) tendría probabilidad 1, es decir, la «certeza absoluta» de contener la proporción teórica π , intervalo éste que, no obstante, como puedes ver no aporta ninguna información.

Fíjate que la limitación enunciada se da por el hecho de querer obtener una estimación «sin error» en una situación de «incertidumbre». En estos casos, y ésta es una solución que los humanos vamos asumiendo durante nuestro proceso de aprendizaje, el problema se solventa rebajando nuestras pretensiones y aceptando que la estimación que realizaremos deberá contener algún grado de error, por pequeño que sea. Lo importante aquí, y éste es un aspecto crucial en el proceso de inferencia estadística, es conocer el modo de cuantificar el error que se comete, puesto que si se conoce la magnitud de dicho error se podrán tomar decisiones aun sabiendo que estas decisiones son «arriesgadas» en el sentido de que en algunos casos serán incorrectas.

En síntesis, podemos decir que para obtener información útil en una situación de incertidumbre habrá que asumir siempre un determinado grado de error, error éste que se trasladará a las decisiones que se basen en dicha información. Una vez realizada esta reflexión, la solución del problema se desplaza pues hacia la medida del error de estimación, que viene dada por la cuantía de la variabilidad de la distribución muestral del estadístico, que recibe el nombre de «**error estándar**» («**EE**») para diferenciarla de la desviación estándar de la distribución de valores individuales en una población o en una muestra.

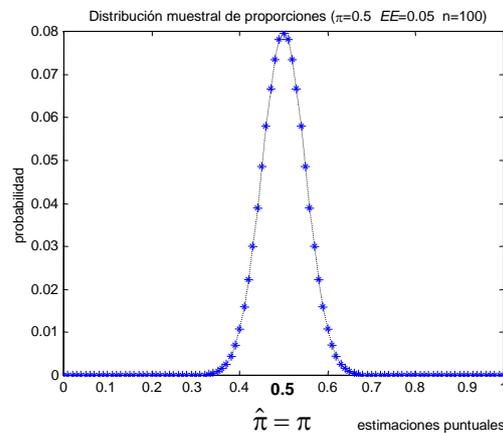
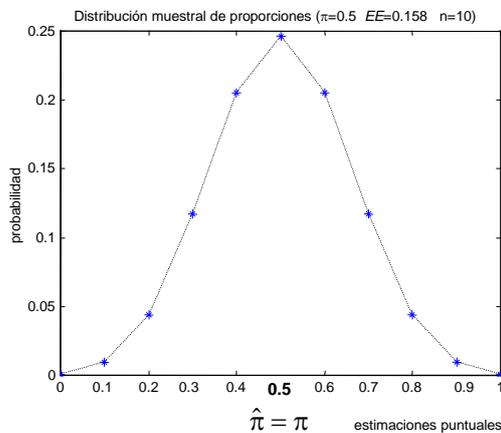
La formulación genérica del error estándar (EE) viene dada por la siguiente expresión que refleja la relación entre la dispersión σ y el tamaño n de las muestras que se extraen aleatoriamente de la población para construir dicha distribución muestral:

$$EE(\hat{\theta}) = \sigma_{\hat{\theta}} = \frac{\sigma}{\sqrt{n}}$$

Para ilustrar este concepto, la siguiente figura muestra las distribuciones muestrales teóricas (es decir, para *todas* las posibles muestras) de la proporción p de caras obtenidas en series de 10 y de 100 lanzamientos de una moneda (compárese con las correspondientes distribuciones empíricas presentadas en las dos figuras anteriores, respectivamente). En este caso, dado que la variancia σ^2 de un carácter binario es igual a $\pi \times (1 - \pi)$, el error estándar de las dos distribuciones muestrales de las estimaciones p es:

$$n = 10 \rightarrow EE(p) = \sigma_{\hat{\pi}} = \sqrt{\frac{\pi \times (1 - \pi)}{n}} = \sqrt{\frac{0.5 \times 0.5}{10}} = 0.1581$$

$$n = 100 \rightarrow EE(p) = \sigma_{\hat{\pi}} = \sqrt{\frac{\pi \times (1 - \pi)}{n}} = \sqrt{\frac{0.5 \times 0.5}{100}} = 0.05$$



Distribuciones muestrales teóricas de la proporción de caras en series de 10 y de 100 lanzamientos de una moneda.

4.1.3. Intervalo de confianza de un parámetro poblacional

Una vez comprendido que las estimaciones puntuales de características poblacionales son insuficientes para concluir sobre éstas, y una vez asumida también la existencia del error debido al muestreo, cuantificado en el cálculo del error estándar EE de una distribución muestral, se deduce que será necesario realizar estimaciones por intervalo, denominadas «intervalos de confianza» («IC»), que tomen en consideración dicho error de muestreo al tiempo que permitan controlar el «riesgo de error» de que el intervalo construido no contenga el verdadero valor del parámetro poblacional.

El riesgo de error asumido se simboliza habitualmente con la letra griega α , y se acostumbra a fijar en la práctica en los valores $\alpha=0.01$, $\alpha=0.05$ ó $\alpha=0.10$, correspondiendo a errores del 1%, 5% y 10%, respectivamente. El valor complementario $1-\alpha$ del riesgo de error fijado expresa el denominado «nivel de confianza» de la estimación.

Riesgo de error α y nivel de confianza ($1-\alpha$)

α	0.10	0.05	0.01	0.001
$(1-\alpha)$	90%	95%	99%	99.9%

Siguiendo a Wonnacott y Wonnacott (1991, p. 125-131), la figura que se presenta en la siguiente página contiene todos los elementos necesarios para la comprensión del mecanismo de construcción del intervalo de confianza de un parámetro θ , siguiendo la distribución muestral del estadístico $\hat{\theta}$ una ley Normal, y asumiendo un riesgo de error del 5% (IC95%). En la parte central de la figura aparece la distribución muestral del estadístico $\hat{\theta}$ que contiene todos los posibles valores que éste puede tomar en muestras extraídas aleatoriamente de la población. Siguiendo la distribución muestral, en este caso, una distribución Normal, los valores que encierran el 95% central de las observaciones se obtienen a partir de la siguiente expresión en la que interviene el valor ± 1.96 correspondiente a los percentiles 2.5 y 97.5 (ambos percentiles dejan, por tanto, una proporción de valores $\alpha/2 = 0.025$ por debajo y por encima, respectivamente) en dicha distribución Normal:

$$\theta \pm 1.96 \times EE$$

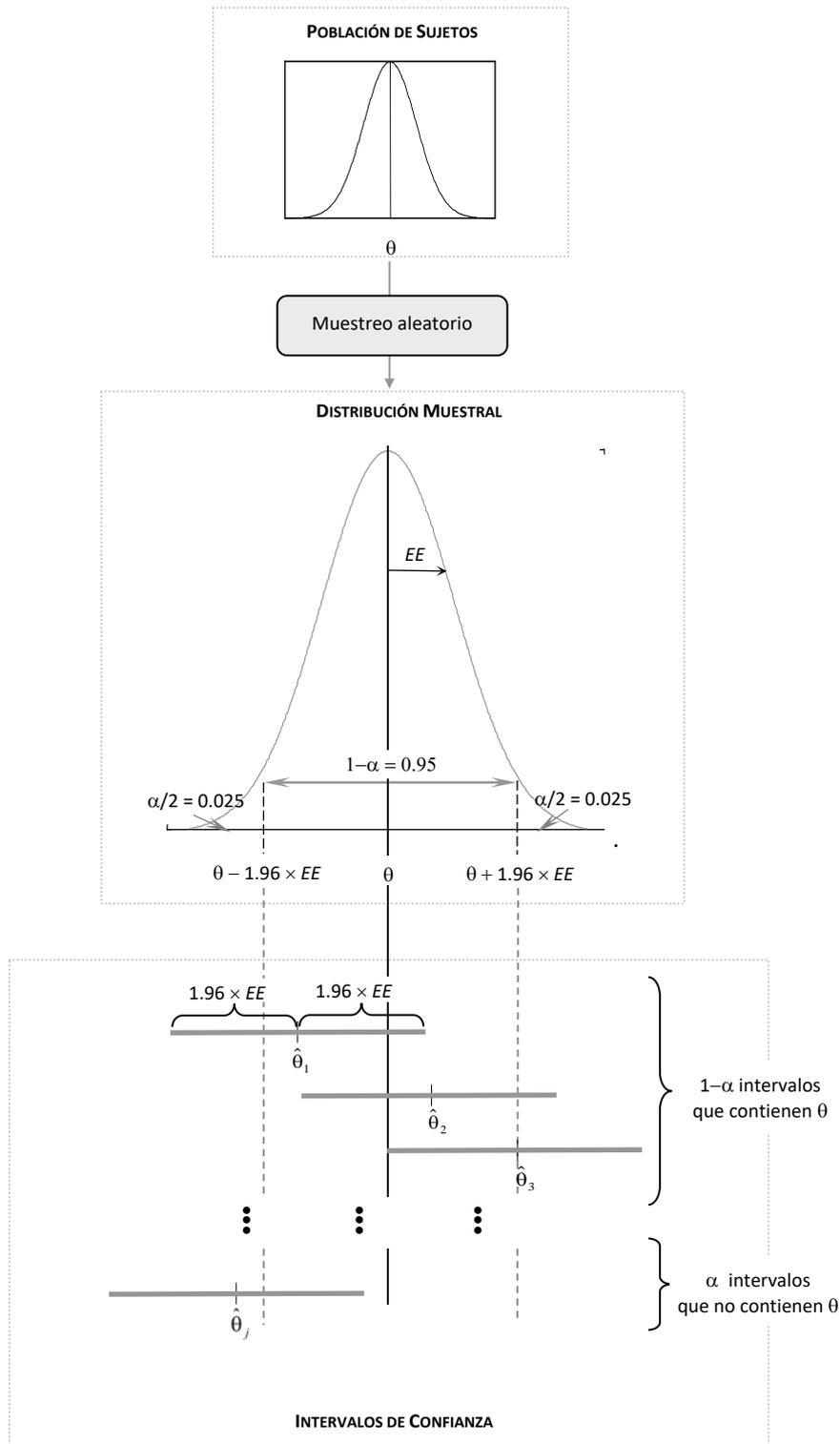
En la parte inferior de la figura se presentan los intervalos de confianza que se podrían construir utilizando esta misma expresión alrededor de cada uno de los valores de los estadísticos $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_j$ obtenidos en las muestras:

$$\hat{\theta} \pm 1.96 \times EE$$

Puesto que la cantidad que se suma y se resta es la misma en ambos casos, los intervalos construidos alrededor del 95% central de los posibles valores de $\hat{\theta}$ contendrán el valor del parámetro θ , mientras que el 5% restante no contendrán dicho parámetro. Como señalan Wonnacott y Wonnacott (1991, p. 128), es importante tener claro que, en la práctica, cuando se calcula el intervalo de confianza a partir del valor del estadístico obtenido en una muestra, *«este intervalo abarca a θ o no lo abarca. Pero lo importante es reconocer que se está aplicando un método con una posibilidad del 95% de éxito; a la larga, el 95% de los intervalos que construya de este modo encerrarán a θ »*.

Fíjate que no es un intervalo concreto el que tiene la probabilidad $1-\alpha$ de contener θ , sino el procedimiento de cálculo en el que se basa. Así pues, al intervalo que se obtiene se le atribuye una «confianza» de magnitud $1-\alpha$, y será incorrecto, por tanto, interpretar $1-\alpha$ como la probabilidad de que ese intervalo contenga el valor del parámetro poblacional. De hecho, a nivel interpretativo hay que considerar que *cuquiera de los valores contenidos en el intervalo de confianza tiene la misma probabilidad de ser el verdadero valor del parámetro poblacional θ* . En la práctica, aunque la expresión general para el cálculo de los límites de un intervalo de confianza es siempre la misma, la fórmula para obtener el valor del error estándar de la distribución muestral variará en función del estadístico a partir del cual se desea realizar la estimación.

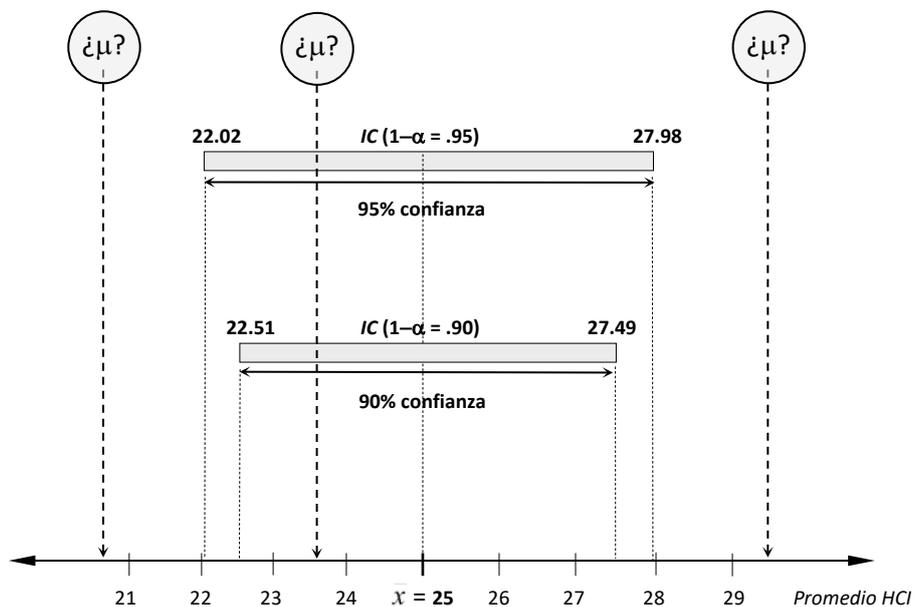
Por último, es importante observar que *la precisión (amplitud) del intervalo de confianza obtenido dependerá de dos factores: el nivel $1-\alpha$ que se elija y el tamaño n de la muestra*. Así, cuanto menor sea el riesgo de error α asumido mayor será la amplitud del intervalo y, por tanto, menor será la precisión de la estimación. Respecto al tamaño muestral, a medida que éste aumenta disminuye el error estándar de la distribución muestral del estadístico, motivo por el cual aumenta la precisión del intervalo de confianza.



Construcción de intervalos de confianza de un parámetro en base a la distribución muestral Normal (adaptada de Wonnacott y Wonnacott, 1991, p. 128).

Vamos a ilustrar estos conceptos con un ejemplo. Supón que se desea conocer la media del número de horas mensuales de conexión a Internet (*HCI*) de los estudiantes de la Universitat Autònoma de Barcelona (UAB). Se dispone de los datos observados en una muestra aleatoria de 100 estudiantes de esta universidad que presentan una media de $HCI = 25h$ y una desviación estándar de 15h. Para responder al problema planteado es necesario calcular el intervalo de confianza alrededor de la media observada $\bar{x} = 25h$. La siguiente figura ilustra los resultados que se obtendrían en este caso para los intervalos de

confianza del 95% y del 90%, así como los aspectos más relevantes de su interpretación: la media poblacional μ puede encontrarse, con la confianza correspondiente, en cualquier lugar del recorrido del intervalo, o incluso fuera de él.



Intervalos de confianza del 95% y del 90% de las medias de horas de conexión mensual a Internet.

Estos resultados indicarían que, **con una confianza del 95%**, en la población de estudiantes de la UAB, la media del número de horas mensuales de conexión a Internet se encuentra aproximadamente entre 22 y 28 horas. Si se desea asumir un riesgo de error mayor, por ejemplo, del 10% (y por tanto se fija un nivel de confianza $1-\alpha$ del 90%), entonces el IC90% resultante es más preciso, entre 22.51 y 27.49 horas, aunque con menor confianza de contener el valor del parámetro poblacional. Observa que la conclusión está formulada con base en los límites del intervalo de confianza, sin mencionar el valor promedio observado porque este valor es anecdótico desde el punto de vista estadístico (i.e, como estimación puntual, tiene un riesgo de error desconocido).

4.1.4. Contraste de hipótesis mediante intervalos de confianza

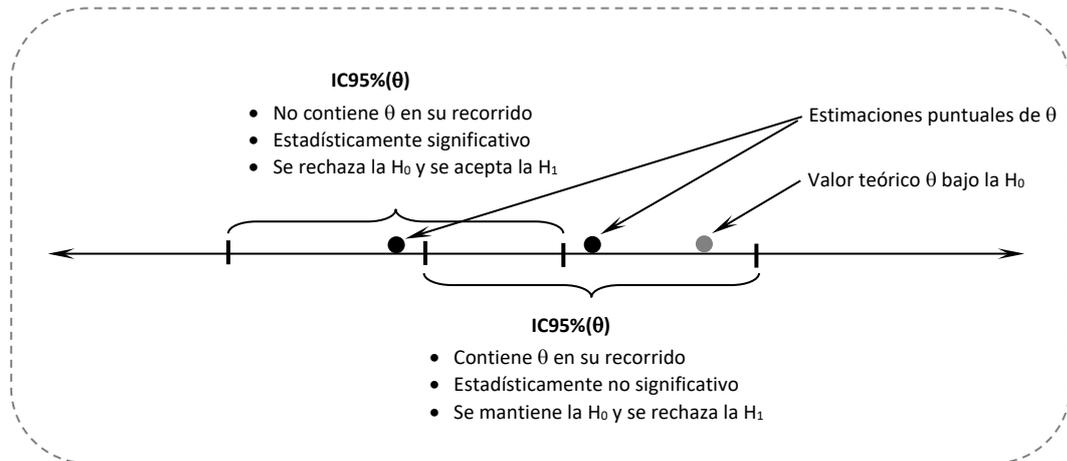
Por último, supón que estudios previos han establecido que los estudiantes de la Universitat de València (UV) tienen en promedio un valor de HCl = 20h. La pregunta que podríamos plantear es si esta aparente diferencia en las horas de conexión a internet de los estudiantes de la UV (20h) y de la UAB (IC95%: 22h a 28h) es debida al azar, es decir, al error de muestreo en la extracción de la muestra de la UAB, o bien se debe a que en realidad los estudiantes de ambas universidades tienen un comportamiento distinto respecto a las HCl.

Para responder a esta cuestión podemos utilizar dos estrategias alternativas que permitirán alcanzar la misma conclusión. La primera estrategia se basa en la consideración ya expuesta de que cualquiera de los valores contenidos en un intervalo de confianza tiene la misma probabilidad de ser el verdadero valor del parámetro poblacional. Esta estrategia consiste en comprobar si el valor estimado del valor de las HCl de los estudiantes de la UV se encuentra dentro o fuera del rango de valores definido por los límites del intervalo de confianza obtenido a partir de la muestra de estudiantes de la UAB: si se encuentra dentro podemos concluir que se trata de poblaciones con el mismo comportamiento respecto a las HCl; por el contrario, si se encuentra fuera, podremos afirmar, con un nivel de confianza $1-\alpha$ (i.e., con un error igual o inferior a α), que ambas poblaciones de estudiantes presentan un comportamiento estadísticamente **distinto** respecto a las HCl.

La primera de las conclusiones, planteada en términos de igualdad, contrasta la denominada «hipótesis nula», que se simboliza en este caso como $H_0: \hat{\theta} = \theta$ para expresar que $\hat{\theta}$ (el valor del estadístico obtenido en la muestra) proviene de una población caracterizada por el valor paramétrico θ . La segunda conclusión, planteada en términos de desigualdad, contrasta la hipótesis complementaria a H_0 , denominada

«hipótesis alternativa» y que se simboliza como $H_1: \hat{\theta} \neq \theta$ para expresar que $\hat{\theta}$ no proviene de una población caracterizada por el valor paramétrico θ .

La «regla de decisión estadística» acerca de si se debe mantener o bien si se debe rechazar la H_0 , consiste en observar si el valor del estadístico $\hat{\theta}$ obtenido en la muestra es un valor plausible para el parámetro poblacional θ , determinando si se halla dentro o fuera del intervalo, tal y como se ilustra en la siguiente figura:



Contraste de hipótesis en base al intervalo de confianza del parámetro poblacional.

- Si el intervalo de confianza *contiene* en su recorrido el valor teórico bajo la H_0 , se «acepta» dicha hipótesis y, por tanto, se «rechaza» la H_1 . En esta situación, se concluye que la diferencia entre el valor del estadístico $\hat{\theta}$ obtenido en la muestra y el valor θ del parámetro poblacional es «estadísticamente no significativa».
- Si el intervalo de confianza *no contiene* en su recorrido el valor teórico bajo la H_0 , se «rechaza» dicha hipótesis y, en consecuencia, se «acepta» la H_e . En esta situación, se concluye que la diferencia entre el valor del estadístico $\hat{\theta}$ obtenido en la muestra y el valor θ del parámetro poblacional es «estadísticamente significativa».

La hipótesis alternativa H_1 se puede plantear de forma equivalente en términos de la diferencia δ entre el valor del estadístico $\hat{\theta}$ estimado a partir de los datos muestrales y el valor θ del parámetro poblacional:

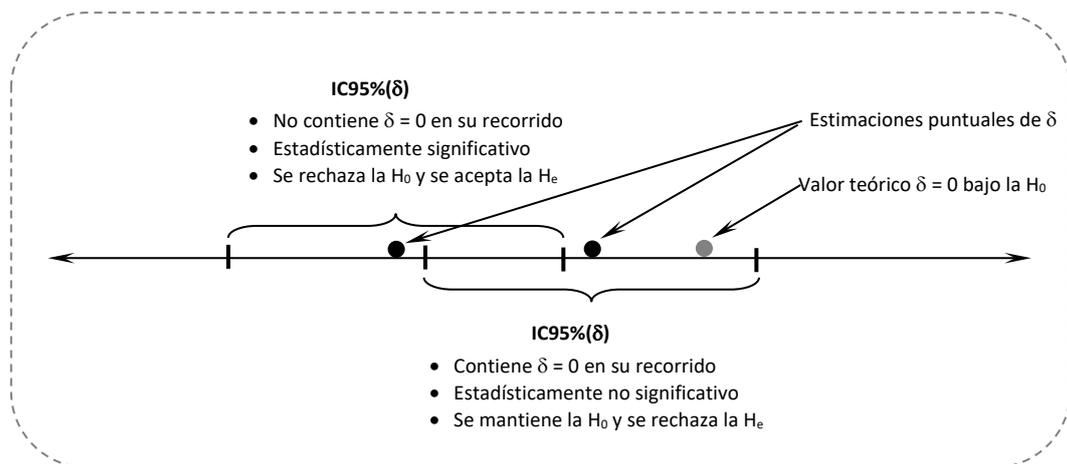
$$H_1: \delta = \hat{\theta} - \theta \neq 0$$

$$H_0: \delta = \hat{\theta} - \theta = 0$$

Bajo este planteamiento, la hipótesis de estudio se contrasta a partir de los límites del intervalo de confianza de la diferencia δ (que se puede obtener directamente restando el valor θ del parámetro poblacional a cada uno de los límites del intervalo de confianza), y el criterio de decisión acerca de la significación estadística de la diferencia observada se basa en valorar si el valor 0 que expresa la diferencia bajo la H_0 se encuentra dentro o fuera de dichos límites, tal como ilustra la siguiente figura.

Fíjate que el aspecto clave en este procedimiento de contraste es la consideración de los posibles valores del parámetro poblacional estimados mediante el cálculo del intervalo de confianza, que incluye en su cálculo el valor del error estándar de la distribución muestral que refleja la fluctuación debida al muestreo, fluctuación ésta que podría ocasionar que la comparación directa entre el valor de una estimación puntual y el valor del parámetro poblacional conllevara una conclusión incorrecta. En síntesis:

«Un intervalo de confianza se puede considerar simplemente como el conjunto de las hipótesis aceptables» (Wonnacott y Wonnacott, 1991, p. 159)



Contraste de hipótesis en base al intervalo de confianza de la diferencia entre el valor observado y el teórico.

4.1.5. Contraste de hipótesis mediante pruebas de significación

Pero hemos dicho que hay dos estrategias para abordar el problema planteado acerca de si los estudiantes de la UAB y la UV proceden de una misma población respecto a las *HCI*. La segunda estrategia consiste en llevar a cabo un procedimiento de «**contraste de hipótesis**». Este procedimiento consiste en obtener el valor del denominado «**grado de significación**» («**p-value**»), simbolizado habitualmente como P , que permite valorar la verosimilitud de la hipótesis nula $H_0: \hat{\theta} = \theta$ que el valor del estadístico $\hat{\theta}$ obtenido en la muestra pueda provenir de una población caracterizada por un determinado valor θ . La regla de interpretación del grado de significación P es la siguiente:

- $P < .05$ indica que el resultado obtenido es «estadísticamente significativo», es decir, con un error $\alpha < 0.05$, se rechaza la hipótesis H_0 y, por tanto, se acepta la H_1 alternativa. En otros términos, con un error inferior al 5%, se considera que **no es verosímil que el resultado $\hat{\theta}$ obtenido en la muestra de estudio pueda provenir de una población con el valor θ del parámetro poblacional.**
- $P \geq .05$ indica que el resultado obtenido es «estadísticamente no significativo», es decir, **nada se opone a aceptar que la hipótesis hipótesis H_0 es cierta.** En otros términos, con una confianza superior o igual al 95%, se mantiene la hipótesis de igualdad y se asume, por tanto, que **la diferencia entre el resultado $\hat{\theta}$ obtenido en la muestra de estudio y el valor θ del parámetro poblacional es debida al azar, es decir, al error o fluctuación debida al muestreo.**

Una última anotación. La $H_1: \hat{\theta} \neq \theta$ planteada antes expresa que el valor del estadístico de interés es diferente de un determinado valor del parámetro poblacional, sin especificar si se espera que sea superior o inferior a éste. En este caso, se dice que la H_1 se formula de forma «**bilateral**». En ocasiones, empero, la H_1 se establece en el sentido que el parámetro de interés es, o bien superior, o bien inferior al valor del parámetro poblacional, en cuyo caso se dice que dichas H_1 se formulan de forma «**unilateral por la derecha**» o de forma «**unilateral por la izquierda**», respectivamente. El tipo de formulación de la hipótesis afecta al cálculo del grado de significación, ya que el riesgo de error α se acumula en una de las colas de la distribución muestral del estadístico en el caso de las hipótesis unilaterales, incrementando así las posibilidades de hallar un resultado estadísticamente significativo, al contrario de lo que sucede en el caso que la hipótesis sea bilateral, que divide el error $\alpha/2$ en cada cola. Es importante tener en cuenta que la decisión acerca de plantear hipótesis unilaterales se debe fundamentar en el conocimiento teórico del ámbito sustantivo de la investigación, no a partir de los datos obtenidos en la misma.

Formulación bilateral o unilateral de hipótesis de contraste sobre un parámetro poblacional.

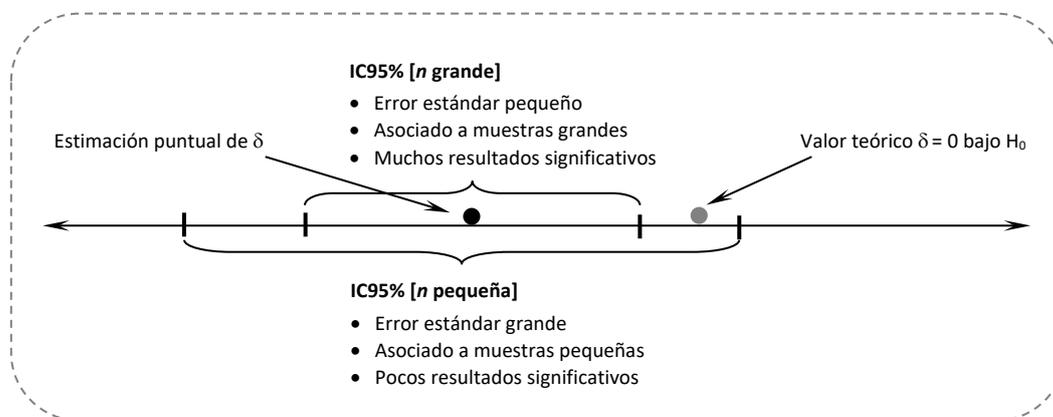
HIPÓTESIS	H_e	H_0
Bilateral	$\hat{\theta} \neq \theta$ —ó— $\delta = (\hat{\theta} - \theta) \neq 0$	$\hat{\theta} = \theta$ —ó— $\delta = (\hat{\theta} - \theta) = 0$
Unilateral por la derecha	$\hat{\theta} > \theta$ —ó— $\delta = (\hat{\theta} - \theta) > 0$	$\theta \leq \theta$ —ó— $\delta = (\hat{\theta} - \theta) \leq 0$
Unilateral por la izquierda	$\hat{\theta} < \theta$ —ó— $\delta = (\hat{\theta} - \theta) < 0$	$\theta \geq \theta$ —ó— $\delta = (\hat{\theta} - \theta) \geq 0$

La tabla anterior resume las posibles formulaciones tanto bilaterales como unilaterales de las H_1 y de las H_0 complementarias, simbolizando como θ el valor del parámetro poblacional con el que se desea

comparar el valor del estadístico $\hat{\theta}$ obtenido en una muestra extraída aleatoriamente de dicha población. Como se puede observar, al formular la H_1 se incluirá el signo $>$ o $<$ cuando se trate de una hipótesis unilateral por la derecha o por la izquierda, respectivamente, mientras que su formulación contendrá el signo de desigualdad \neq si se trata de una hipótesis bilateral.

4.1.6. Significación estadística y relevancia práctica: valoración del tamaño del efecto

Cabe notar también, que una conclusión en términos de «significación estadística» es un tipo de aseveración que hace referencia a la plausibilidad de una hipótesis, pero no a la relevancia práctica de los resultados obtenidos. Lógicamente, cuanto mayor sea la magnitud de la diferencia observada, mayor será la probabilidad de que la decisión estadística se decante hacia la significación, por hallarse el valor criterio de referencia bajo la H_0 fuera de los límites del intervalo de confianza. Pero puede no ser así cuando el tamaño de la muestra de estudio sea pequeño, porque el error estándar será grande y, en consecuencia, el intervalo de confianza será también más amplio, aumentando la probabilidad de que el valor criterio se halle dentro del intervalo. A la inversa, cuando el tamaño muestral es muy grande, el error estándar es más pequeño, el intervalo de confianza más estrecho y aumenta la probabilidad de rechazar la H_0 , incluso ante un resultado de nula relevancia práctica. Por tanto, ante una misma estimación puntual se pueden alcanzar conclusiones estadísticas distintas según sea el tamaño muestral, como ilustra la siguiente figura.



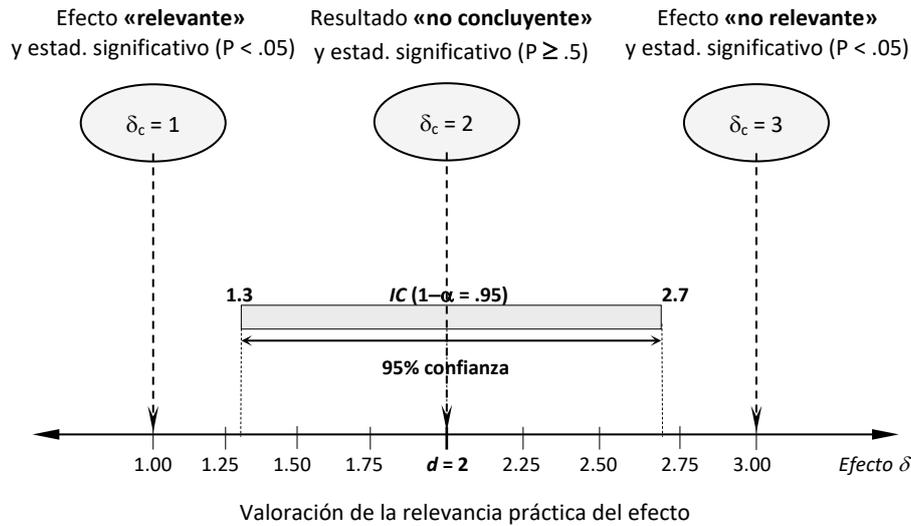
Relación entre la significación estadística, la magnitud de la diferencia observada y el tamaño muestral.

Una interpretación conjunta de la significación estadística y de la magnitud de la diferencia observada como guía para la «significación práctica» posibilita una evaluación más correcta de sus consecuencias teóricas o aplicadas. En este sentido, la estimación por intervalo de los estadísticos descriptivos, de contraste de hipótesis o de tamaño del efecto permiten integrar la información que por separado proporcionan la significación estadística y las estimaciones puntuales obtenidas en las muestras de datos.

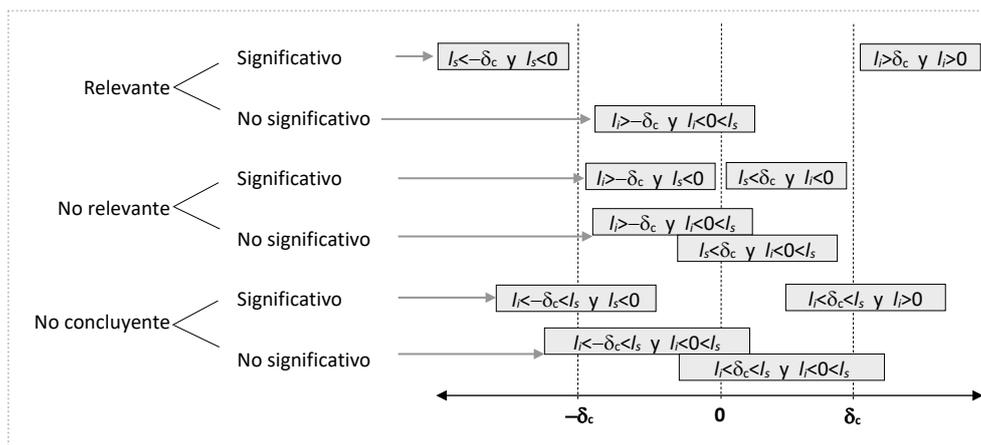
En general, para evaluar la relevancia práctica de los resultados a partir del intervalo de confianza del estadístico descriptivo, de contraste de hipótesis o de tamaño del efecto de interés δ , es necesario disponer de un criterio externo δ_c que, a modo de umbral, exprese la cantidad mínima que se puede considerar relevante en la práctica. Al considerar la posición del valor δ_c respecto a los límites del intervalo de confianza se pueden dar tres situaciones: que el efecto observado sea relevante, que sea no relevante, o bien que el intervalo no permita concluir en términos de relevancia del efecto.

La siguiente figura ilustra estas tres situaciones con un ejemplo hipotético en el que se ha obtenido el intervalo de confianza para una diferencia entre la media m observada en una muestra y una hipotética media poblacional μ (IC95% δ : 1.3 a 2.7 puntos). Se deduce por los valores del IC95% que la prueba de significación de la hipótesis $H_1: \delta \neq 0$ ha resultado estadísticamente significativa, puesto que el IC95% no incluye el valor $\delta = 0$ en su recorrido. Supón también que se consideran tres criterios de relevancia práctica: $\delta_c = 1$, $\delta_c = 2$ y $\delta_c = 3$. Para el criterio $\delta_c = 1$, puesto que δ_c se encuentra por debajo del límite inferior del IC95%, la diferencia estimada se consideraría relevante; es decir, todas las posibles diferencias que expresa el recorrido del IC95% son mayores que δ_c . Con el criterio $\delta_c = 2$ consideraríamos el resultado como no concluyente, puesto que δ_c se encuentra dentro del IC95%; es decir, el resultado se podría considerar relevante por ser δ_c superior al valor del límite inferior del intervalo, pero al mismo tiempo se podría considerar irrelevante por ser δ_c inferior al valor del límite superior. Por último, aplicando el criterio

$\delta_c = 3$ el resultado estimado se consideraría irrelevante por ser el límite superior del IC95% inferior a δ_c ; es decir, todas las posibles diferencias que expresa el recorrido del IC95% son menores que δ_c .



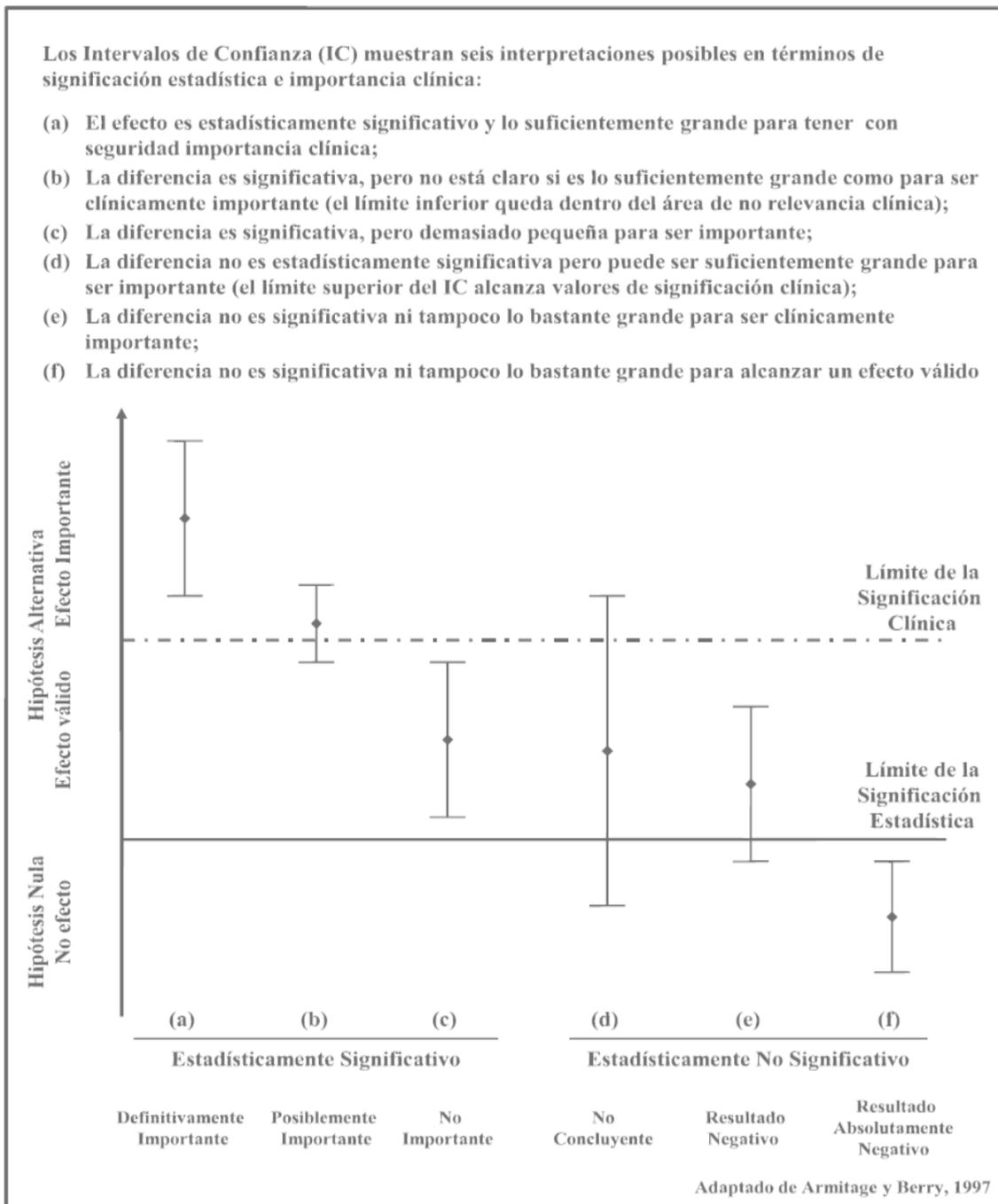
La siguiente figura y la de la siguiente página ilustran las combinaciones entre las tres conclusiones posibles acerca de la relevancia práctica de los resultados expresados con un intervalo de confianza y las dos posibles decisiones acerca de la significación estadística de un contraste de hipótesis bilateral.



Interpretaciones de un intervalo de confianza en función de la relevancia práctica y la significación estadística en el contraste de una hipótesis bilateral.

4.1.7. Significación estadística y relación causal

Por último, otro aspecto fundamental que se debe tener presente al aplicar una prueba de contraste de hipótesis es que la obtención de un resultado estadísticamente significativo no implica la existencia de una relación causal entre las variables estudiadas. Un resultado estadísticamente significativo sólo permite concluir que las diferencias observadas no son atribuibles al azar inherente al proceso de muestreo, de manera que son generalizables a la población. Para poder concluir, además, que las diferencias son debidas a la variable explicativa, se debe tomar en consideración el diseño del estudio. Así, si se ha llevado a cabo un diseño experimental, asignando aleatoriamente los sujetos a diferentes grupos, se podrá atribuir a esa variable *Grupo* el efecto observado en la respuesta, lo que apoyará una conclusión en términos de causalidad. Por contra, si el diseño utilizado ofrece pocas garantías sobre la comparabilidad de los grupos, no se podrán emitir juicios de causalidad sólidos, ya que el efecto observado puede ser debido a otros factores no controlados en el diseño de la investigación. Procedimientos estadísticos como el ajuste de modelos de regresión múltiple ofrecen herramientas que permiten ajustar los resultados frente al efecto de potenciales factores de confusión no controlados por el diseño del estudio.



Interpretación conjunta de los intervalos de confianza, de la significación estadística y de la relevancia práctica («significación clínica»).

En los siguientes apartados ilustraremos mediante un ejemplo práctico cómo se pueden obtener con jamovi los intervalos de confianza para estimar valores de parámetros poblacionales (e.g., proporciones, promedios, etc.) y los grados de significación para contrastar hipótesis acerca dichos parámetros (e.g., diferencias entre medias, asociación entre variables, etc.).

4.2. Estimación por intervalo y contraste de hipótesis con una variable categórica

En este apartado revisaremos cómo obtener con jamovi los valores de los intervalos de confianza para medidas categóricas y como llevar a cabo contrastes de hipótesis acerca de valores (parámetros) poblacionales. Para ello utilizaremos la matriz de datos de un estudio sobre el lenguaje en niños con síndrome de Down en el que se registraron las siguientes variables:

Variable	Etiqueta	Categorías
Sexo	Sexo	0: Masculino; 1: Femenino
Edad	Edad en años cumplidos	
Lectura	Lectura	0: No; 1: Sí
GD	Grado de discapacidad (escala 0 a 100)	
QI	Cociente intelectual (escala 0 a 200)	
ErrEF	Número de errores en la prueba de evaluación fonológica (EF)	
ExpoEF	Número de términos evaluados en la prueba de EF	

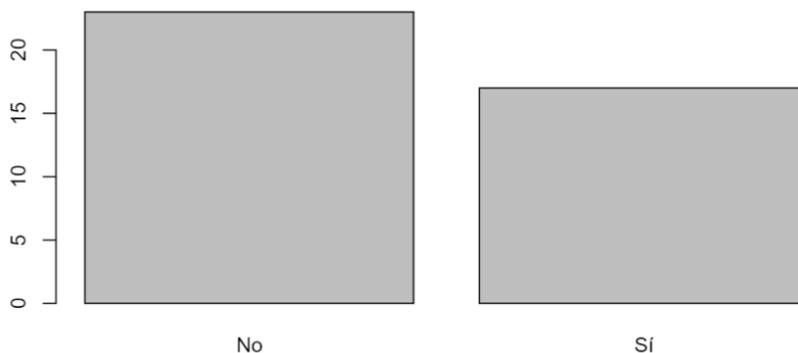
1. Abre el fichero de datos *DOWN.omv* ejecutando la opción *Open* del menú principal  de jamovi, o bien haz doble clic sobre este fichero en el explorador de Windows.
2. Vamos a llevar a cabo una estimación por intervalo a partir de las respuestas obtenidas para una **variable categórica binaria**, cuyo índice estadístico resumen más apropiado es la frecuencia relativa o proporción y cuya distribución muestral sigue una ley de probabilidad Binomial. Para ello, ejecuta el menú *Analyses | esci | Estimate Proportion*.
3. En el panel de opciones selecciona la variable categórica *Lectura* en la lista de variables y pásala a *Measure* pulsando .

Estimated Proportion

Group	Cases	Total N	P	95 % CI	
				Lower	Upper
No	23	40	0.575	0.422	0.715
Sí	17	40	0.425	0.285	0.578

Note. There were no missing responses.

Frequency Plot



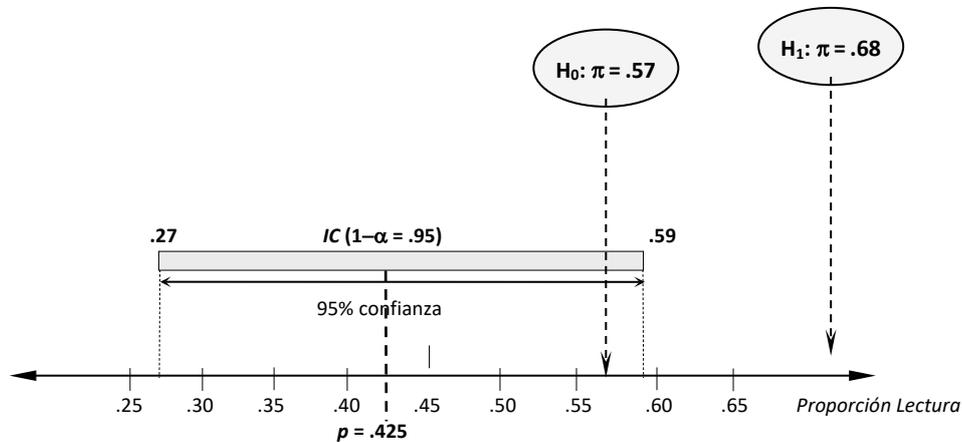
Estos resultados indican que, **con una confianza del 95%, el porcentaje de jóvenes y adultos con síndrome de Down en la población origen de la muestra que saben leer se encuentra entre el 28.5% y el 57.8%.**

Supón que se conoce que en una población distinta de la que se obtuvo la muestra del presente estudio, el porcentaje de jóvenes y adultos con síndrome de Down que saben leer fuera del:

- 68% ($H_0: \pi = .68$): deberíamos concluir que, **con un riesgo de error del 5%, ambas poblaciones son estadísticamente distintas por lo que respecta a esta característica**, puesto que el valor de referencia de la otra población (H_1 en la siguiente figura) se encuentra fuera del recorrido del intervalo de confianza del 95% obtenido en el presente estudio ($p = .425$; IC95%: .27 – .59).

- 57% ($H_0: \pi = .57$): nada se opone a afirmar que los resultados del presente estudio son compatibles con el valor de referencia de la otra población, puesto que dicho valor de referencia (H_0 en la siguiente figura) se encuentra dentro del recorrido del intervalo de confianza del 95% obtenido en el presente estudio ($p = .425$; IC95%: .27 – .59).

La siguiente figura ilustra estas conclusiones. Observa dónde se encuentran los dos hipotéticos valores de referencia poblacionales ($\pi = .68$ y $\pi = .57$) respecto a los límites del intervalo de confianza:



4. Como se ha comentado en el apartado anterior, para responder a las preguntas anteriores también podemos contrastar las hipótesis alternativas ($H_1: \pi \neq .68$) y ($H_1: \pi \neq .57$) obteniendo el grado de significación mediante una **prueba de conformidad** basada en una distribución muestral binomial (**Binomial Test**), adecuada para variables categóricas binarias. Para ello, ejecuta el menú **Analyses | Frecuencias | 2 Outcomes**. Selecciona la variable *Lectura* y escribe el valor $\pi = .68$ del parámetro poblacional en **Test value** y selecciona **Confidence intervals** en **Additional Statistics**:

Binomial Test						95% Confidence Interval	
	Level	Count	Total	Proportion	p	Lower	Upper
Lectura	No	23	40	0.575	0.175	0.409	0.730
	Sí	17	40	0.425	0.001	0.270	0.591

Note. H_a is proportion \neq 0.68

Como puedes observar, el valor del grado de significación $p = 0.001$ para la categoría *Sí* de la variable *Lectura* indica que la diferencia entre la proporción observada ($p = .425$) y la teórica ($\pi = .68$) es estadísticamente significativa ($P < .05$). La conclusión es, por tanto, la misma que alcanzamos en el paso 3 al comparar el valor poblacional teórico $\pi = 0.68$ con los límites del intervalo de confianza.

Para completar la interpretación de este resultado, podemos valorar también la magnitud de la diferencia entre la proporción observada y la teórica poblacional interpretando el IC95%: 0.27 a 0.59 de dicha diferencia. Como ya se expuso en el apartado 3, las medidas de tamaño del efecto adecuadas en este caso pasarían por el cálculo de la diferencia de proporciones (DP) entre los valores de los límites del IC95% obtenidos con los datos de la muestra del estudio y la proporción teórica: (DP = .27 – .68 = –.41; –.41 * 100 = –41%) y (DP = .591 – .68 = –.09; –.09 * 100 = –9%). Estas DP indican que el 42.5% de lectores observado en la muestra jóvenes y adultos con síndrome de Down es, con una confianza del 95%, entre un 9% y un 41% inferior al 68% poblacional de jóvenes y adultos con síndrome de Down que leen.

También podemos calcular el riesgo relativo (RR) como la razón entre los límites del IC95% y la proporción teórica: (RR = .27 / .68 = 0.397; [0.397 – 1] * 100 = –60.3%) y (RR = .591 / .68 = 0.869; [0.869 – 1] * 100 = –13.1%). Estos RR indican que el 42.5% de lectores observado en la muestra de jóvenes y adultos con síndrome de Down supone, con una confianza del 95%, una reducción relativa de entre el 13.1% (RR = 0.869) y el 60% (RR = 0.397) respecto al 60% de jóvenes y adultos con síndrome de Down que leen en la población.

5. Cambia el valor de **Test value** por **.57** para contrastar la segunda hipótesis (**H₁: $\hat{\pi} \neq .57$**) planteada:

Binomial Test

	Level	Count	Total	Proportion	p	95% Confidence Interval	
						Lower	Upper
Lectura	No	23	40	0.575	1.000	0.409	0.730
	Sí	17	40	0.425	0.078	0.270	0.591

Note: H_a is proportion \neq 0.57

En este caso, el valor del grado de significación $P = .078$ para la categoría **Sí** de la variable **Lectura** indica que la diferencia entre la proporción observada ($p = .425$) y la teórica ($\pi = .57$) no es estadísticamente significativa ($P \geq .05$). De nuevo, la conclusión es la misma que alcanzamos en el paso 3 al comparar el valor poblacional teórico $\pi = .57$ con los límites del intervalo de confianza.

La interpretación de las medidas del tamaño del efecto refleja el motivo por el cual la diferencia ha resultado estadísticamente no significativa. La diferencia de proporciones (DP) entre los valores de los límites del IC95% obtenidos con los datos de la muestra del estudio y la proporción teórica: (DP = $.27 - .57 = -.30$; $-.30 * 100 = -30\%$) y (DP = $.591 - .57 = .021$; $.021 * 100 = 2.1\%$). Estas DP indican que el **42.5% de lectores observado en la muestra de jóvenes y adultos con síndrome de Down es, con una confianza del 95%, entre un 30% inferior y un 2.1% superior al 57% poblacional de jóvenes y adultos con síndrome de Down que leen**. El hecho de que el intervalo de las DP incluya el valor 0% indica que el 42.5% observado en la muestra del estudio es, con una confianza del 95%, compatible (i.e., la diferencia es estadísticamente no significativa) con una población en la que el porcentaje es del 57%. Dicho de otro modo, la diferencia observada puede ser debida a la variabilidad previsible al extraer muestras aleatorias de esa población (i.e., al denominado «**error de muestreo**»).

Esta misma conclusión la podemos extraer del cálculo del riesgo relativo (RR) como la razón entre los límites del IC95% y la proporción teórica: (RR = $.27 / .57 = 0.474$; $[0.474 - 1] * 100 = -52.6\%$) y (RR = $.591 / .57 = 1.03$; $[1.03 - 1] * 100 = 3\%$). Estos RR indican que el **42.5% de lectores observado en la muestra de jóvenes y adultos con síndrome de Down supone, con una confianza del 95%, entre una reducción relativa del 52.6% (RR = 0.474) y un aumento relativo del 3% (RR = 1.03) respecto al 57% de jóvenes y adultos con síndrome de Down que leen en la población**. El hecho de que el intervalo de los RR incluya el valor 1 (IC95% RR: 0.474 a 1.03) indica que el 42.5% observado en la muestra del estudio puede suponer tanto una reducción como un incremento respecto al 57% existente en la población, motivo por el cual concluimos que los resultados obtenidos en la muestra son, con una confianza del 95%, estadísticamente compatibles con una población en la que el porcentaje es del 57%.

Las dos hipótesis **H₁** que hemos contrastado, ($H_1: \pi \neq .68$) y (**H₁: $\pi \neq .57$**), se han formulado de forma bilateral. En el bloque **Hypothesis** del panel de opciones de **Analyses | Frequencies | 2 Outcomes** se puede seleccionar las opciones **> Test value** o **< Test value** si se desea contrastar **hipótesis H1 unilaterales** por la derecha ($H_1: \hat{\pi} > \pi$) o por la izquierda ($H_1: \hat{\pi} < \pi$), respectivamente.

Por último, hay que señalar que el menú **Analyses | Frequencies | N Outcomes** permite llevar a cabo una **prueba de conformidad de una distribución observada con más de dos categorías** (para una variable categórica politómica u ordinal) respecto a una distribución teórica, seleccionando en el panel de opciones la **Variable** categórica, y desplegando el bloque **Expected Proportions** para indicar en **Ratio** la distribución de proporciones esperadas.

4.3. Estimación por intervalo y contraste de hipótesis con una variable cuantitativa

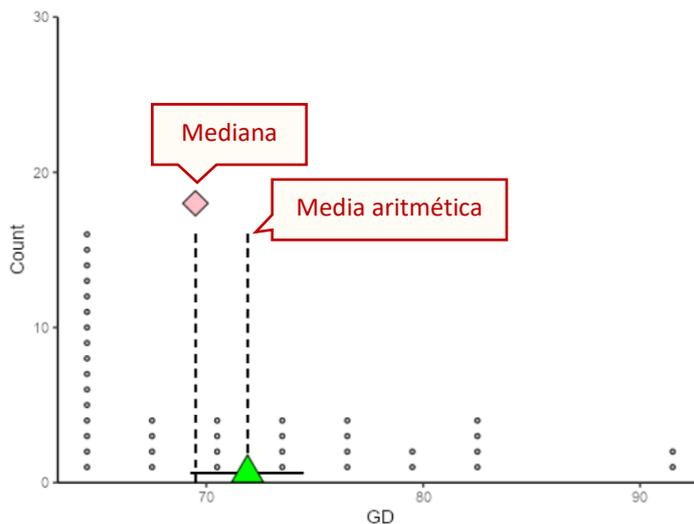
En este apartado, vamos a ejecutar el procedimiento de estimación y contraste univariante presentado en el apartado anterior para el caso de una variable cuantitativa.

1. Ejecuta el menú **Analyses | esci | Estimate Mean**. En el panel de opciones selecciona la **variable cuantitativa continua GD** en la lista de variables y pásala a **Measure**. A continuación, despliega el bloque **Distribution-Graph Options** y marca las casillas **Mean marker** y **Median marker**.

Descriptive Statistics

	95 % CI			s	N	Missing	Median	Quartiles	
	M	Lower	Upper					25%	75%
GD	71.9	69.3	74.5	8.08	40	0	69.5	65.0	78.0

Distribution



A partir de estos resultados, concluiríamos, con una confianza del 95%, que en la población origen de la muestra de estudio el promedio de la puntuación en la escala de grado discapacidad (*GD*) de los jóvenes y adultos con síndrome de Down se encuentra entre 69.3 y 74.5 puntos.

2. Supongamos ahora que se desea comparar la media observada en la muestra $m = 71.9$ con una media poblacional teórica $\mu = 70$. Para contrastar la hipótesis alternativa ($H_1: \mu \neq 70$) debemos ejecutar una **prueba de conformidad** basada en una distribución muestral *t* de Student (**T-test**). Para llevar a cabo este contraste de hipótesis ejecuta el menú **Analyses | T-test | One Sample T-Test**.
3. En el panel de opciones, pasa la variable *GD* a **Dependent Variables** y escribe el valor del parámetro poblacional 70 en **Test value** del bloque **Hypothesis**. Marca también en el bloque **Additional Statistics** las opciones **Mean difference** y **Effect size** y sus correspondientes **Confidence Interval 95%**:

p	Mean difference	95% Confidence Interval		Cohen's d	Effect Size	95% Confidence Interval	
		Lower	Upper			Lower	Upper
0.145	1.90	-0.685	4.48	0.235	-0.0806	0.548	

El valor del grado de significación $P = .145$ indica que la diferencia 1.90 entre la media observada ($m=71.9$) y la teórica ($\mu=70$) no es estadísticamente significativa ($P \geq .05$). Esta conclusión se desprende también del IC95%: -0.685 a 4.48 de esta diferencia, que incluye el valor 0 que expresa la igualdad.

De nuevo, se trata de la misma conclusión que alcanzaríamos al observar que el valor poblacional $\mu=70$ se encuentra entre los límites del IC95%: 65.0 a 78.0 de la media observada obtenidos en el paso anterior.

Sobre la interpretación de resultados estadísticamente no significativos, pero no concluyentes

Ya hemos expuesto en el apartado 0.6 que la interpretación de la magnitud de la diferencia entre el valor del estadístico observado en la muestra y el valor del parámetro poblacional teórico conduce a la conclusión de que dicha diferencia es atribuible al error de muestreo cuando el resultado del contraste de hipótesis es no significativo. Pero **aun cuando la prueba de contraste de hipótesis resulta estadísticamente no significativa, puede ser interesante obtener índices de tamaño del efecto para interpretar adecuadamente los resultados obtenidos.**

Así, frente a un resultado estadísticamente no significativo, por ejemplo en una **prueba de conformidad** para comparar el valor de un estadístico con el de un parámetro poblacional, la interpretación de las medidas de tamaño del efecto pueden resultar de interés cuando el objetivo que se desea contrastar no es valorar si la muestra puede o no proceder de una población con un determinado valor paramétrico (como en el caso analizado en el paso 5 del apartado anterior o en el caso que nos ocupa en este apartado), sino que es, por ejemplo, decidir si una intervención realizada con los participantes de un estudio es más eficaz que otra ya existente cuya eficacia es conocida, o bien comparar la eficacia observada en una muestra de estudio con un criterio de eficacia preestablecido (e.g., una reducción de una determinada proporción o porcentaje en el caso de una medida de eficacia categórica, o de una determinada cantidad en el caso de una medida de eficacia cuantitativa).

Pero ¿por qué en estos casos resulta interesante interpretar la magnitud del efecto si la prueba de contraste de hipótesis ha resultado estadísticamente no significativa? A pesar de que la prueba de significación arroje un resultado no significativo ($P \geq .05$) que no permite concluir que la $H_1: \hat{\mu} \neq \mu$ es cierta, los límites del IC95% aportan una información adicional de interés práctico cuando los límites del IC95% del índice de tamaño del efecto ofrecen resultados contradictorios: el valor de uno de los límites indica la existencia de un efecto relevante en la práctica en un determinado sentido, mientras que el valor del otro límite va en el sentido contrario o directamente indica que el efecto es irrelevante. En esta situación, puesto que todos los valores del intervalo de confianza tienen, con el margen de error α , la misma verosimilitud de ser el verdadero valor poblacional, el resultado obtenido es **«no concluyente»** desde un punto de vista aplicado. Esta conclusión tiene un significado y unas implicaciones desde el punto de vista de la investigación aplicada distintas de las que se derivan de un contraste con resultado **«estadísticamente no significativo»**. Un resultado no concluyente, pero con indicios de un posible efecto relevante, no niega de forma categórica la existencia de dicho efecto, sino que expresa la imposibilidad de concluir en un sentido concreto como consecuencia de una falta de precisión en la estimación estadística (i.e., la excesiva amplitud del intervalo de confianza). Esta falta de precisión, como se ha visto en el apartado 4.1.2, puede ser debida a diferentes causas, frecuentemente por una alta variabilidad de la medida analizada o por contar con una muestra reducida de casos. Otros estudios pueden obtener resultados más precisos y concluyentes en un sentido u otro, por ejemplo, reduciendo la variabilidad mediante el uso de medidas de mayor calidad, o trabajando con muestras con un mayor número de casos. Puesto que la amplitud del IC95% se relaciona inversamente con el tamaño muestral, a mayor tamaño muestral, menor será la amplitud (i.e., mayor precisión) del IC95%, disminuyendo así las posibilidades de obtener un resultado no concluyente. En la práctica, para asegurar que el resultado final del análisis sea concluyente (con el tamaño/relevancia que sea), se debe calcular previamente el tamaño muestral necesario para garantizar que se obtendrán IC95% con una determinada amplitud (precisión).

Para ilustrar el cálculo e interpretación del tamaño del efecto en una prueba de conformidad de comparación de una media observada con una media poblacional, vamos a analizar el resultado del índice de tamaño del efecto adecuado en este caso, la d de Cohen. Se trata de una medida estandarizada y, por tanto, libre de unidades de medida, de la diferencia entre dos medias (puedes revisar el apartado 3.3 para saber más sobre la d de Cohen). **Los límites del IC95%: -0.08 a 0.548 de la d de Cohen** (valores de **Effect Size**) **obtenidos estandarizando la diferencia $d = 1.90$ entre la media $m = 71.9$ observada y la media teórica $\mu = 70$ indican que, con una confianza del 95%, esta diferencia se puede considerar**, según los criterios de interpretación propuestos por Cohen, como **pequeña ($d = -0.08$)**, o bien **grande ($d = 0.55$)** ($d < 0.2$ para valores *very small* y $d > 0.5$ para efectos de magnitud *medium*). Como ya hemos comentado, a pesar de lo paradójico de este resultado, es compatible con la previsible fluctuación debida al error de muestreo y, por tanto, no contradice la hipótesis nula $H_0: \hat{\mu} = 70$ de que la muestra puede provenir de una población con este valor paramétrico.

4. En este estudio se ha recogido también la puntuación en una prueba de evaluación fonológica, medida como el número de errores (variable *ErrEF*) cometidos al evaluar un conjunto de palabras. Puesto que se trata de una **variable [cuantitativa discreta] de recuento**, el intervalo de confianza debe calcularse a partir del promedio de errores observado en la muestra de estudio con base en una distribución muestral que sigue una ley de probabilidad de Poisson, en lugar de una ley Normal (o *t* de Student), como sería el caso para las variables cuantitativas continuas. jamovi no incluye (todavía) la posibilidad directa de obtener los intervalos de confianza y el grado de significación de una prueba de conformidad entre una media observada y una media teórica con base a la distribución de Poisson. Se pueden obtener, sin embargo, los valores de la estimación por intervalo con base en la verdadera distribución muestral de Poisson ejecutando el menú **Analyses | Linear Models | Generalized Linear Models** (requiere la instalación del módulo *gamlj – General Analyses for Lineal Models in jamovi*). Ejecuta este menú y en el panel principal de opciones marca **Poisson** en **Frequencies**, pasa la variable *ErrEF* a **Dependent Variable** y marca **Odds Ratios (expB)** en **Effect Size** y **For exp(B)** en **Confidence Intervals**:

Parameter Estimates

Names	Estimate	SE	95% Exp(B) Confidence Interval			z	p
			exp(B)	Lower	Upper		
(Intercept)	2.58	0.0435	13.2	12.1	14.4	59.3	< .001

El intervalo de confianza obtenido indica que, con una confianza del 95%, el número de errores que en promedio cometen los jóvenes y adultos con síndrome de Down en la población origen de la muestra de este estudio se encuentra entre los 12.1 y los 14.4 errores.

En este estudio el recuento de errores cometidos en la prueba de evaluación fonológica no se llevó a cabo sobre el mismo número de palabras, sino que cada participante del estudio evaluó un número de palabras distinto (*ExpoEF*) durante el tiempo establecido para la prueba, tiempo éste que sí se mantuvo constante. En esta situación los cálculos requieren transformar los recuentos en **tasas (rates)** tomando en consideración la diferente “exposición” de cada participante. En los programas de análisis estadísticos este tipo de ajuste se realiza indicando como *exposure variable* aquella que recoge el número de palabras a las que estuvo “expuesto” cada participante e indicando *Poisson (rates)* como tipo de intervalo de confianza. Desafortunadamente, jamovi no incluye, todavía, la posibilidad de incluir la variable de exposición, ni la prueba de conformidad, en los análisis con variables de recuento.

4.4. Contraste de hipótesis acerca de la asociación entre dos variables cuantitativas: coeficiente de correlación lineal de Pearson

En este apartado vamos a revisar cómo llevar a cabo el contraste de hipótesis acerca de la relación lineal entre dos variables cuantitativas. Para ello estudiaremos la relación entre el grado de discapacidad (*GD*) y el número de errores cometido por los participantes en la prueba de evaluación fonológica (*ErrEF*).

El siguiente cuadro de información resume los índices estadísticos y gráficos que se deben obtener para contrastar una hipótesis de asociación entre las dos variables.

Procedimiento general de contraste de hipótesis bivalente

En general, la estrategia para llevar a cabo el contraste de hipótesis bivalente incluye la obtención de los siguientes resultados estadísticos:

- La **descripción univariante** de las variables cuya relación se desea analizar y el **análisis gráfico de la relación** entre las dos variables.
- La obtención del **índice o índices estadísticos de contraste** adecuados en función del tipo de hipótesis alternativa H_1 planteada (diferencias, asociación, etc.) y la escala de medida de las dos variables, su **intervalo de confianza** y el **grado de significación** asociado a H_1 .
- La obtención de la/s **medidas del tamaño del efecto** adecuadas en función de la escala de medida de las dos variables y su intervalo de confianza.

Interpretación del diagrama de dispersión para valorar la relación entre dos variables cuantitativas

La relación entre dos variables cuantitativas se puede representar gráficamente con un **diagrama de dispersión** (*scatterplot*), que presenta en una nube de puntos la puntuación de cada participante en las dos variables estudiadas:

- Una nube de puntos descendente indica que existe una marcada relación negativa, es decir, que a medida que aumentan o disminuyen los valores de una variable disminuyen o aumentan en sentido contrario, en promedio, los valores de la otra variable.
- Una nube de puntos ascendente indica que hay una fuerte relación positiva, es decir, que a medida que aumentan o disminuyen los valores de una variable aumentan o disminuyen en el mismo sentido también, en promedio, los valores de la otra variable.
- Una nube de puntos horizontal o esférica indica ausencia de relación entre las dos variables.

Disposición de las variables en los ejes del diagrama de dispersión

La disposición de las variables en los dos ejes del diagrama de dispersión no es discrecional cuando la interpretación de la asociación entre las dos variables sólo tiene sentido en una dirección concreta. Así, por ejemplo, si se tiene interés por analizar la relación en el sentido de cómo el grado de discapacidad afecta al rendimiento en la prueba de evaluación fonológica, siendo inverosímil plantear la relación en el sentido contrario, entonces la variable que se debe situar en el eje X del diagrama de dispersión es el grado de discapacidad y la que se debe situar en el eje Y es el número de errores cometido en la prueba de evaluación fonológica. Esta disposición facilita la interpretación correcta de la nube de puntos, cuya lectura se acostumbra a realizar recorriendo los valores del eje horizontal X y mientras se observa cómo se sitúan los puntos respecto a las coordenadas del eje vertical Y.

En los estudios con diseños experimentales se habla de “**variables independientes**” y de “**variable dependiente**”; esta terminología cambia cuando los diseños son de tipo no experimental, hablándose, por ejemplo, de “**variables de exposición**” o “**variables pronósticas**” y de “**variable de respuesta**”, respectivamente.

Interpretación del grado de significación P (p -value)

El grado de significación permite valorar la verosimilitud de que la asociación obtenida en la muestra de estudio pueda provenir de una población en la que no existe dicha asociación. La regla de interpretación del grado de significación P es la siguiente:

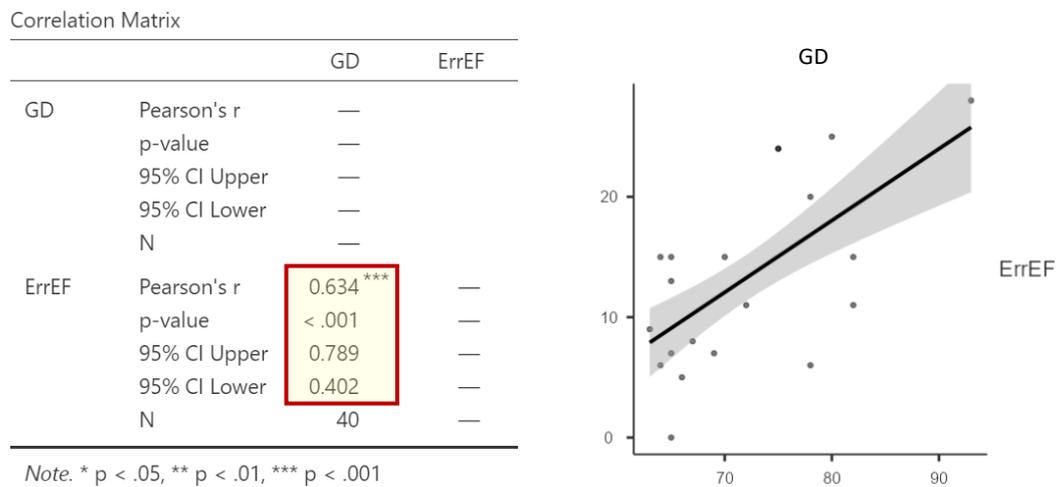
- $P < .05$ indica que el resultado obtenido es «**estadísticamente significativo**». Es decir, con un error α inferior al 5% se rechaza la hipótesis de ausencia de asociación y, por tanto, no es verosímil que el resultado obtenido en la muestra de estudio pueda provenir de una población en la que no existe asociación entre las variables analizadas.
- $P \geq .05$ indica que el resultado obtenido es «**estadísticamente no significativo**». Es decir, nada se opone a aceptar que la asociación obtenida en la muestra de estudio es debida al azar (error o fluctuación debida al muestreo), y que la muestra de estudio puede provenir, por tanto, de una población en la que no existe asociación entre las variables analizadas.

Interpretación del tamaño del efecto r

Tal y como se expuso en el apartado 3.3, Cohen (1988, 1992), propone los siguientes criterios para interpretar los valores de las medidas de tamaño del efecto basadas en índices de correlación:

<i>Effect size</i>	Small	Medium	Large	Very large
r	0.10	0.30	0.50	0.70

1. Para obtener el índice de correlación de Pearson y el diagrama de dispersión ejecuta el menú **Analyses | Regression | Correlation Matrix**.
2. Selecciona, en este orden, las variables *GD* y *ErrEF* y marca las opciones **Flag significant correlations, N y Confidence intervals** en **Additional Options**.



La recta de ajuste del diagrama de dispersión muestra una pendiente positiva, con la mayor parte de los puntos relativamente cercanos a dicha recta, lo que evidencia una relación lineal clara y directa entre ambas variables, con muy pocos casos que se alejan de este patrón.

El valor del índice de correlación r de Pearson entre el grado de discapacidad y el número de errores cometido en la prueba de evaluación fonológica es de $r = .634$, valor positivo, de magnitud “grande” ($r > .50$) y estadísticamente significativo ($P < .001$).

Por tanto, la interpretación global de los resultados obtenidos podría expresarse del siguiente modo: se puede afirmar que, **con un error inferior al 1%, los participantes con síndrome de Down presentan una relación lineal directa ($r = .634$; IC95%: .402 a .79) y de magnitud grande entre el grado de discapacidad y el número de errores en la prueba de evaluación fonológica, en el sentido que cuanto mayor es el grado de discapacidad mayor es el número de errores que se cometen.**

4.5. Contraste de hipótesis acerca de la asociación entre dos variables categóricas: comparación de proporciones

En este apartado revisaremos cómo llevar a cabo un contraste de hipótesis acerca de la relación (asociación) estadística entre dos variables categóricas. Para ello utilizaremos como ejemplo las variables categóricas binarias *Sexo* (0: Masculino y 1: Femenino) y *Lectura* (0: No; 1: Sí).

1. Ejecuta el menú **Analyses | Frequencies | Contingency Tables | Independent Samples**. En el panel de opciones, selecciona la variable *Sexo* en filas (Rows) y la variable *Lectura* en columnas (Columns).

A pesar de que en el apartado 3.3 hemos recomendado seguir la misma lógica para ubicar las variables en la tabla de contingencia que se aplica cuando éstas son cuantitativas, es decir, situar en columnas las categorías de la **variable independiente, de exposición o pronóstica**, y en filas las categorías de la **variable dependiente o de respuesta**, en jamovi debemos disponerlas al revés para que los cálculos de la medida del tamaño del efecto se realicen de forma correcta. En consecuencia, **los porcentajes que se interpretarán serán los calculados sobre el total de las filas (i.e., dentro de cada categoría de la variable independiente, de exposición o pronóstica) en lugar de hacerlo sobre los totales de las columnas (que en este caso contendrán las categorías de la variable dependiente o de respuesta).**

2. Despliega el bloque **Statistics** y marca la opción χ^2 en **Tests** y las opciones **Relative Risk, Difference in proportions y Confidence intervals** en **Comparative Measures (2x2 only)**.
3. Despliega el bloque **Cells** para marcar **Row** en **Percentages**.

Contingency Tables

Sexo		Lectura		Total
		No	Sí	
Masculino	Observed	16	4	20
	% within row	80.0 %	20.0 %	100.0 %
Femenino	Observed	7	13	20
	% within row	35.0 %	65.0 %	100.0 %
Total	Observed	23	17	40
	% within row	57.5 %	42.5 %	100.0 %

 χ^2 Tests

	Value	df	p
χ^2	8.29	1	0.004
N	40		

Comparative Measures

	Value	95% Confidence Intervals	
		Lower	Upper
Difference in 2 proportions	0.450*	0.177	0.723
Relative risk	2.29*	1.21	4.32

* rows compared

Como puedes observar, la distribución de frecuencias de *Lectura* difiere en cada grupo definido por *Sexo* (80% y 20% en los hombres; 35% y 65% en las mujeres). O, dicho de otro modo, estos resultados indican que en la muestra de estudio existe una relación entre *Sexo* y *Lectura*, en el sentido que hay más participantes que leen en el grupo de mujeres adultas con síndrome de Down que en el grupo de hombres (y al contrario respecto a la frecuencia de participantes que no leen).

Tal y como se expuso en el procedimiento 3, para evaluar la relevancia práctica de estos resultados se puede calcular, entre otros, los siguientes índices de tamaño del efecto:

- Diferencia de proporciones (o de porcentajes) (DP): $80\% - 35\% = 45\%$
- Razón de proporciones (o de porcentajes) (o riesgo relativo, RR): $80\% / 35\% = 2.29$

El IC95% de la diferencia de proporciones (DP) indica que, en la población origen de la muestra de estudio, la proporción de no lectores es entre un 17.7% y un 72.3% mayor en los hombres que en las mujeres; y el IC95% del riesgo relativo (RR) indica que la proporción de no lectores es en la población entre un $([1.21 - 1] * 100 =)$ 21% y 4.32 veces mayor en los hombres que en las mujeres. Ambas medidas (DP y RR) se pueden comparar con algún criterio de referencia para valorar su relevancia práctica (e.g., que la DP sea superior al 10%, o que el RR sea superior a 1.1, es decir, al 10%).

Por último, tras la tabla de contingencia, jamovi presenta el valor de la prueba de contraste χ^2 de **Pearson**, que evalúa la asociación estadística entre dos variables categóricas, así como su grado de significación estadística ($P = .004$). El valor del estadístico χ^2 no tiene una interpretación práctica, pero sí su grado de significación. A continuación, se presentan cuatro formas equivalentes y complementarias para expresar los resultados obtenidos. Siguiendo el patrón de interpretación expuesto en el apartado anterior, podemos afirmar que, con un error inferior al 5% (en este caso del 4 por mil):

- [interpretación de la DP]: ... existe una **diferencia** entre las distribuciones de frecuencias de lectura entre hombres y mujeres, en el sentido que en la población origen de la muestra el porcentaje de hombres no lectores es entre un 17.7% y un 72.3% superior al de mujeres no lectoras.

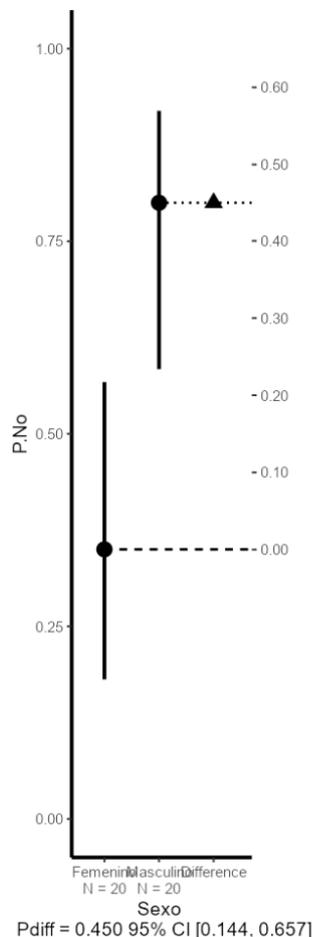
- [interpretación del RR]: ... existe una **diferencia** entre las distribuciones de frecuencias de lectura entre hombres y mujeres, en el sentido que en la población origen de la muestra el número de hombres con síndrome de Down que no leen es entre un 21% y 4.32 veces más que el de mujeres.
 - [interpretación en términos de asociación]: ... la frecuencia de lectura **está asociada** con el sexo de los participantes, en el sentido que entre los hombres hay un porcentaje de no lectores entre un 17.7% y un 72.3% superior al de mujeres no lectoras.
 - [interpretación en términos de inferencia poblacional]: ... la **población origen de la muestra de estudio presenta distintos porcentajes de no lectores** en los grupos de hombres (alrededor del 80%) y de mujeres (alrededor del 35%).
4. Esta última interpretación se puede precisar incluyendo los resultados de la estimación por intervalo de las proporciones de lectores en el grupo de hombres y de mujeres en la población origen de la muestra de estudio. Para ello hay que ejecutar el menú **Analyses | esci | Estimate Proportion Difference**. En el panel de opciones selecciona la variable *Lectura* como **Outcome variable** y la variable *Sexo* como **Category variable**:

Estimated Difference in Proportion

	No	Sí	Total N	P	95 % CI	
					Lower	Upper
Masculino	16	4	20	0.800	0.584	0.919
Femenino	7	13	20	0.350	0.181	0.567
Difference	.	.	.	0.450	0.144	0.657

Note. There were no missing responses.

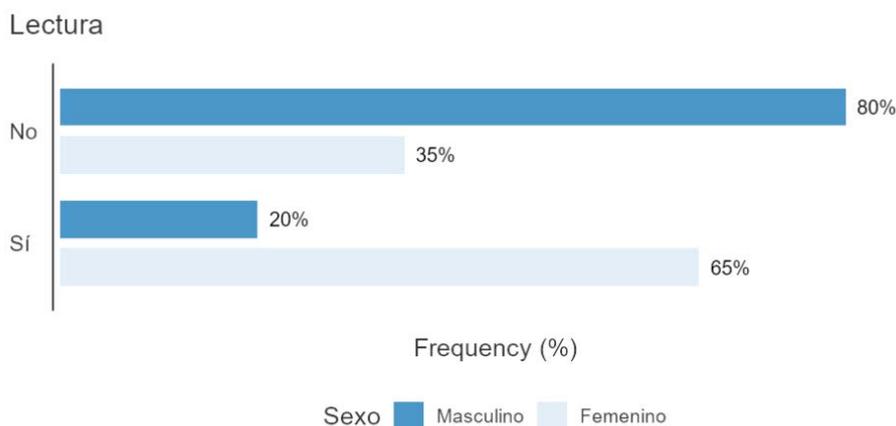
Proportion Plot



Ahora podemos realizar una interpretación más precisa acerca de los porcentajes estimados de lectores en la población origen de la muestra: **con una confianza del 95%, en la población origen de la muestra se estima que el porcentaje de hombres con síndrome de Down no lectores oscila entre el 58.4% y el 91.9%, mientras que el porcentaje de mujeres no lectoras se encuentra entre el 18.1% y el 56.7%, oscilando la diferencia porcentual entre ambos sexos entre el 14.4% y el 65.7%.**

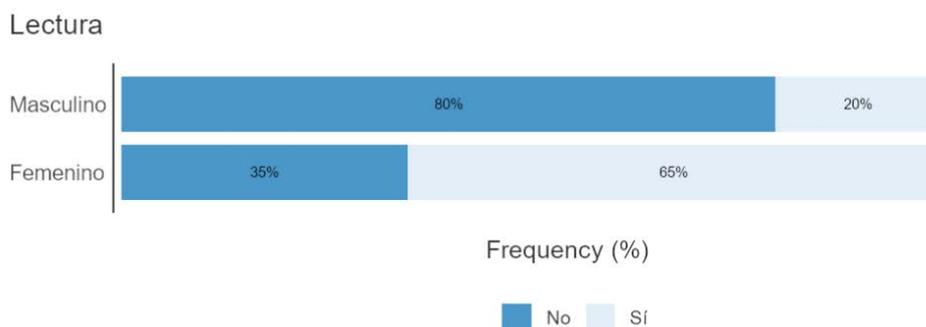
Observa que los intervalos de confianza de los dos grupos de participantes (lectores y no lectores) no se solapan, lo cual indica que el resultado de la diferencia de proporciones es estadísticamente significativo, es decir, que **hombres y mujeres provienen de poblaciones de jóvenes y adultos con síndrome de Down con distintos porcentajes de no lectores.**

5. Por último, vamos a presentar también este resultado mediante el correspondiente diagrama de barras agrupadas. Ejecuta el menú **Analyses | Exploration | Survey Plots**. Pasa la variable **Lectura** a **Variables** y la variable **Sexo** a **Grouping Variable**.



Este diagrama de barras permite ver las diferencias de forma rápida observando las alturas de las barras de las frecuencias de las categorías “No” y “Sí” en cada submuestra definida por **Sexo**. Puedes comprobar que las sumas de los porcentajes de las barras dentro de cada sexo suman el 100%.

6. Otra forma de representar el gráfico de barras es hacerlo mediante barras apiladas (*stacked bars*). Marca la opción **Stacked bar** en el panel de opciones para cambiar el tipo de gráfico:



4.6. Contraste de hipótesis acerca de la asociación entre una variable categórica y una variable cuantitativa: comparación de medias y análisis de la variancia

Por último, en este apartado se analiza gráfica y estadísticamente la relación entre una variable categórica y una variable cuantitativa. Para ello, utilizaremos cuatro ejemplos: el análisis de la relación entre la **variable categórica binaria Lectura** y las variables **QI** (cociente intelectual; **variable cuantitativa continua**) y **ErrEF** (número de errores que los participantes cometen en la prueba de evaluación fonológica (variable **cuantitativa discreta de recuento**), y entre la **variable categórica ordinal** nivel de discapacidad (**ND**) y las variables **QI** y **ErrEF**.

Estrategia de análisis de la relación entre una variable categórica y una variable cuantitativa

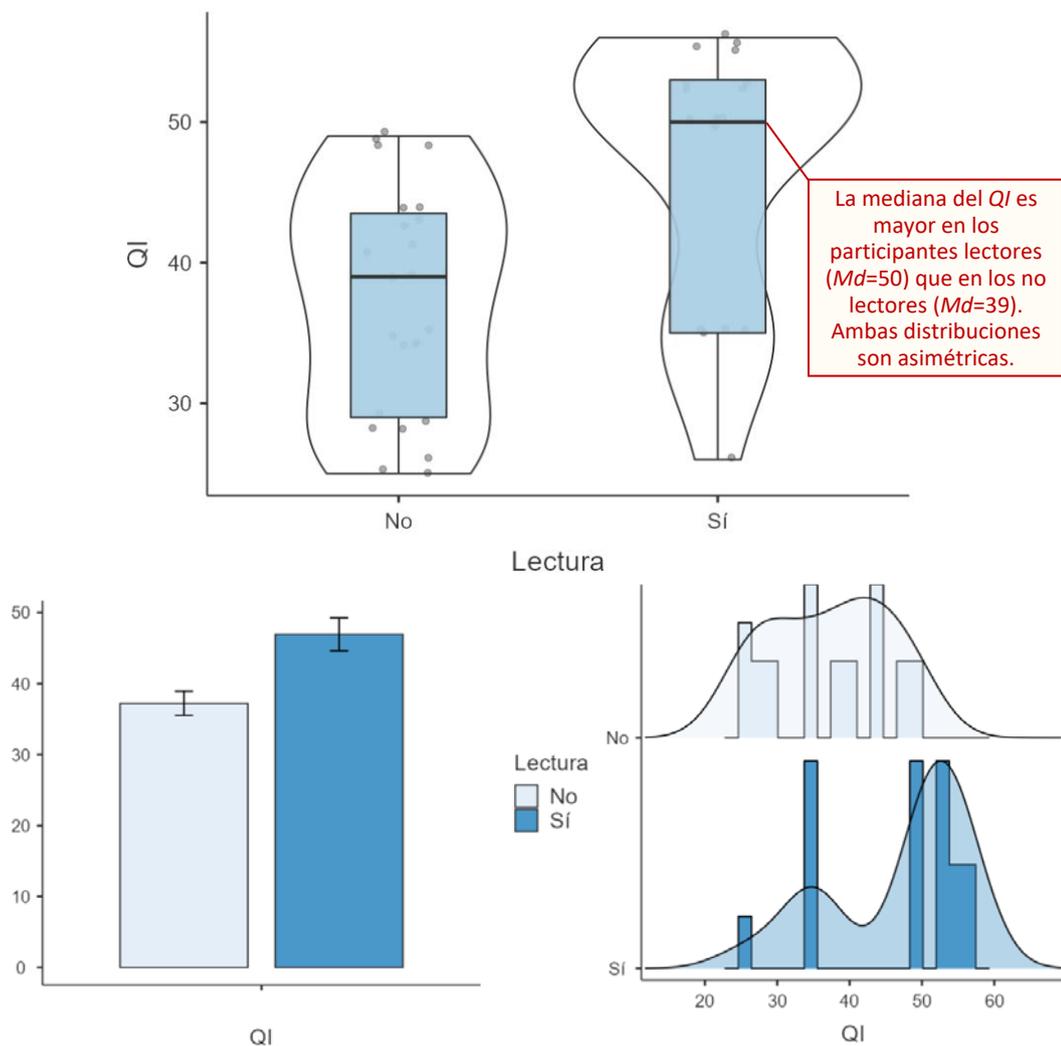
Para estudiar la **relación entre una variable categórica y una variable cuantitativa**, el procedimiento habitual es comparar el índice de tendencia central (media aritmética o mediana) de la variable cuantitativa en cada categoría de la otra variable.

Los tipos de gráficos que permiten visualizar este tipo de relación cuando la variable cuantitativa es discreta son los **diagramas de barras agrupadas** y los **diagramas de líneas múltiples** (también los **diagramas de barras y de áreas apiladas**); cuando la variable cuantitativa es continua se utiliza habitualmente el **diagrama de barras de error**. En todos los casos es aconsejable obtener el **diagrama de caja múltiple**.

4.6.1. Relación entre una variable categórica binaria y una variable cuantitativa continua

Iniciamos este bloque de análisis de la asociación estadística entre una variable categórica y una variable cuantitativa, con el estudio de la relación entre la **variable categórica binaria Lectura** y la **variable cuantitativa continua QI**.

1. En primer lugar, vamos a analizar gráficamente si existen diferencias en el cociente de inteligencia (*QI*) entre lectores y no lectores (*Lectura*). Para ello vamos a construir un diagrama de caja múltiple (*multiple boxplot*), un diagrama de barras y un histograma de la variable *QI* para cada grupo definido por la variable *Lectura*. Ejecuta **Analyses | Exploration | Descriptives** y en el panel de opciones pasa la variable *QI* a **Variabls** y la variable *Lectura* a **Split by**. A continuación, despliega el bloque **Plots** y marca las tres opciones de **Box Plots** (**Box plot**, **Violin** y **Data**), las dos de **Histograms** (**Histogram** y **Density**) y **Bar plot** en **Bar Plots**:



Observando el diagrama de caja se puede comprobar que la mediana del cociente de inteligencia de los participantes lectores es superior a la del grupo de no lectores, aunque las cajas centrales, que encierran el 50% central de los participantes, muestran una asimetría en ambas distribuciones, mostrando un desplazamiento de la mediana (línea interior de la caja) hacia la parte superior de la caja. Esta asimetría también se puede constatar en los histogramas.

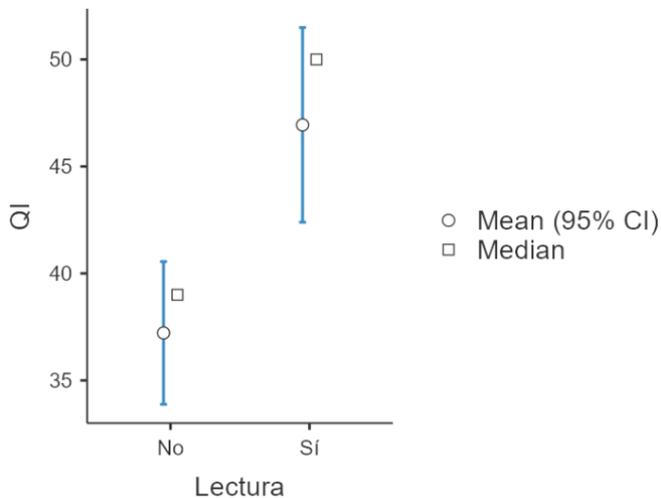
2. Vamos ahora a comprobar si la diferencia observada en el gráfico de caja es suficientemente grande como para poder concluir que existen diferencias estadísticamente significativas entre lectores y no lectores. Para ello, ejecuta el menú **Analyses | T-test | Independent Samples T-Test**. En el panel de opciones, pasa la variable *Lectura* a **Grouping Variable** y la variable *QI* a **Dependent Variable**, y marca en el bloque **Additional Statistics** las casillas **Mean difference**, **Effect size**, **Descriptives**, **Descriptives plots** y los dos **Confidence Interval**, para la diferencia de medias y para el tamaño del efecto:

Independent Samples T-Test												
	Statistic	df	p	Mean difference	SE difference	95% Confidence Interval		Cohen's d	95% Confidence Interval			
						Lower	Upper		Effect Size	Lower	Upper	
QI	Student's t	-3.46	38.0	0.001	-9.72	2.81	-15.4	-4.04		-1.11	-1.80	-0.391

Group Descriptives						
	Group	N	Mean	Median	SD	SE
QI	No	23	37.2	39.0	8.16	1.70
	Sí	17	46.9	50.0	9.57	2.32

Plots

QI



jamovi calcula las medias aritméticas de *QI* para cada uno de los dos grupos de comparación (lectores y no lectores), así como sus intervalos de confianza IC95%. Como puedes observar, el grupo de no lectores presenta un promedio de 37.2 puntos en su coeficiente de inteligencia, mientras que el grupo de lectores tiene en promedio 46.9 puntos. Observa también en el gráfico que los límites de los IC95% de ambas medias no se solapan (el límite inferior del IC95% del grupo de lectores es mayor que el límite superior del IC95% del grupo de no lectores), y que la diferencia de -9.72 puntos entre los promedios de ambos grupos, con un IC95%: -15.4 a -4.04, no incluye en su recorrido el valor 0. Por este motivo, el grado de significación correspondiente a la hipótesis de diferencias entre los dos grupos $H_1: \hat{\mu}_{\text{no lectores}} \neq \hat{\mu}_{\text{lectores}}$ ofrece un valor estadísticamente significativo $P = .001$.

Podemos concluir, por tanto, que existe una diferencia estadísticamente significativa ($P = .001$) entre el cociente de inteligencia (QI) de los participantes lectores y los no lectores, en el sentido que los no lectores tienen en promedio 9.72 (IC95%: -15.4 a -4.04) puntos menos que los lectores en la valoración de su QI. O, dicho de otro modo, existe una relación estadísticamente significativa ($P = .001$) entre leer o no leer y el cociente intelectual, en el sentido que los jóvenes y adultos con síndrome de Down no lectores tienen una puntuación de QI inferior a la de los lectores.

Evaluación del tamaño del efecto en la relación entre una variable categórica y una cuantitativa

Generalmente, la evaluación del tamaño del efecto al estudiar la relación entre una variable categórica y una variable cuantitativa se lleva a cabo mediante el cálculo e interpretación del índice estandarizado *d* de Cohen (o alguna de sus variantes, como las Δ -Delta- de Glass para estudios con diseños experimentales, o la más general *g* de Hedges, especialmente útil en meta-análisis).

Se trata de índices que permiten valorar la magnitud del efecto de la variable categórica sobre la variable cuantitativa, a través de la estandarización de las diferencias entre los índices de tendencia central obtenidos en los grupos definidos por la variable categórica. Su interpretación, como ya se ha comentado, se realiza habitualmente comparando dichos índices con los valores criterio propuestos por Cohen (1988) y Sawilowsky (2009):

Effect size	Very small	Small	Medium	Large	Very large	Huge
<i>d</i>	0.01	0.20	0.50	0.80	1.20	2.0
η^2		0.01	0.06	0.14		

En palabras de Cohen (1992):

*The ES index for the test of the difference between independent means is *d*, the difference expressed in units of (i.e., divided by) the within-population standard deviation. For this test, the $d = 0$ indicates that the difference equals zero; and the small, medium, and large ESs are $d = .20$, $.50$, and $.80$. Thus, an operationally defined medium difference between means is half a standard deviation; concretely, for IQ scores in which the population standard deviation is 15, a medium difference between means is 7.5 IQ points.*

Pero como ya hemos visto, la interpretación en términos exclusivamente de significación estadística no es suficiente. En la práctica, hay que valorar también si se trata de una diferencia relevante. A falta de un criterio externo para valorar la magnitud de la diferencia observada, podemos recurrir de nuevo al criterio propuesto por Cohen para valorar el tamaño del efecto estandarizado de *Lectura* sobre la puntuación de *QI*. La estimación puntual del tamaño del efecto mediante el cálculo de la *d* de Cohen ofrece un resultado de -1.11 , que en valor absoluto se podría calificar como un tamaño del efecto de magnitud “grande” según el criterio de Cohen. Sin embargo, como ya hemos justificado anteriormente, la interpretación correcta debe hacerse a partir de los valores del IC95%: -1.8 a -0.39 de la *d* de Cohen. Los valores de este IC95% indican que estamos frente a una diferencia entre el QI de lectores y no lectores que resulta muy relevante si interpretamos en valor absoluto el límite inferior del IC95% ($|d = -1.8| > 1.20$, criterio mínimo para considerar que un tamaño del efecto es “very large”) o, por el contrario, que se trata de una diferencia de magnitud “pequeña” si interpretamos el valor del límite superior del IC95% ($|d = -0.39| > 0.20$ según criterio de Cohen).

Una última reflexión. La hipótesis que hemos contrastado $H_1: \hat{\mu}_{\text{no lectores}} \neq \hat{\mu}_{\text{lectores}}$ ha valorado si la diferencia en la puntuación del QI es significativamente *distinta* entre ambos grupos. Pero en realidad, la diferencia que deseáramos contrastar en este caso probablemente no sería esta, sino la $H_1: \hat{\mu}_{\text{no lectores}} < \hat{\mu}_{\text{lectores}}$, es decir, que los jóvenes y adultos con síndrome de Down no lectores tendrán una puntuación de QI *menor* que los lectores. La primera hipótesis está formulada como «hipótesis bilateral», mientras que la segunda en términos de «hipótesis unilateral». Como ya se expuso en el apartado 0.5, el contraste de una hipótesis unilateral conlleva un cálculo distinto del grado de significación, acumulando el riesgo de error α en una de las colas de la distribución muestral del estadístico de contraste.

3. Para contrastar la hipótesis unilateral $H_1: \hat{\mu}_{\text{no lectores}} > \hat{\mu}_{\text{lectores}}$ basta con seleccionar **Group 1 < Group 2** en el bloque **Hypothesis** del panel de opciones del procedimiento **Independent Samples T-Test** que estamos ejecutando. Al hacer este cambio, jamovi actualiza automáticamente el resultado del cálculo del grado de significación P y de los límites del IC95% de la diferencia observada (en este caso, un intervalo abierto por la izquierda por el hecho de tratarse de una hipótesis unilateral):

Independent Samples T-Test											
QI	Statistic	df	p	Mean difference	SE difference	95% Confidence Interval		Cohen's d	Effect Size	95% Confidence Interval	
						Lower	Upper			Lower	Upper
Student's t	-3.46	38.0	< .001	-9.72	2.81	-Inf	-4.99	-1.11	-1.80	-0.391	

Note. H_0 : No < Sí

Se mantiene la interpretación realizada en el paso, aunque ahora con un resultado estadísticamente más significativo: $P < .001$ frente al resultado $P = .001$ obtenido con el contraste bilateral.

4.6.2. Relación entre una variable categórica binaria y una variable cuantitativa de recuento

Vamos ahora a analizar la relación entre la **variable categórica binaria** *Lectura* y la **variable cuantitativa [discreta] de recuento** *ErrEF* (número de errores en la prueba de evaluación fonológica). Tal y como ya se ha expuesto al llevar a cabo la estimación por intervalo del promedio de *ErrEF* (ver el paso 4 del apartado 2), la distribución muestral de las medias (y de la diferencia entre medias) de variables de recuento no se ajusta a una ley Normal (o t de Student), como sería el caso para las variables cuantitativas continuas, sino a una ley de probabilidad de Poisson.

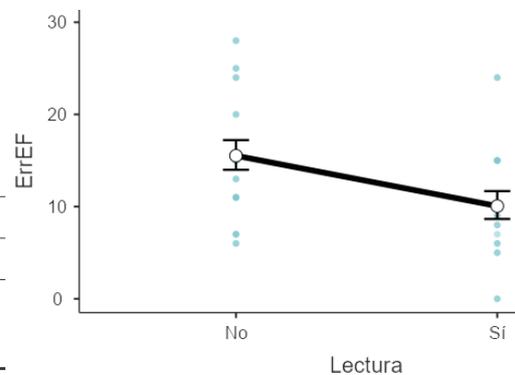
1. jamovi no incluye (todavía) la posibilidad directa de llevar a cabo el contraste de hipótesis bivalente para variables de recuento, pero se puede llevar a cabo este análisis desde el menú **Analyses | Linear Models | Generalized Linear Models** (requiere la instalación del módulo *gamlj – General Analyses for Linear Models in jamovi*). Ejecuta este menú y en el panel principal de opciones marca **Poisson** en **Frequencies**, pasa la variable *ErrEF* a **Dependent Variable** y la variable *Lectura* a **Factors**. A continuación, marca **Odds Ratios (expB)** en **Effect Size** y **For exp(B)** en **Confidence Intervals**. Despliega el bloque de opciones **Factors Coding** y elige la opción simple para *Lectura*. A continuación, despliega el bloque **Plots** y pasa la variable *Lectura* a **Horizontal axis** para obtener también el gráfico de medias con sus IC95% (marca la opción **Confidence intervals** en **Display** y la opción **Observed scores** en **Plot**). Por último, despliega el bloque **Estimated Marginal Means** y marca la opción **Estimated Marginal Means**:

Parameter Estimates

Names	Effect	Estimate	SE	exp(B)	95% Exp(B) Confidence Interval		z	p
					Lower	Upper		
(Intercept)	(Intercept)	2.525	0.0465	12.495	11.389	13.667	54.31	< .001
Lectura1	Sí - No	-0.434	0.0930	0.648	0.539	0.776	-4.66	< .001

Lectura

Lectura	Mean Count	SE	95% Confidence Interval	
			Lower	Upper
No	15.5	0.821	13.99	17.2
Sí	10.1	0.769	8.66	11.7



Los jóvenes y adultos con síndrome de Down de la muestra de este estudio que no leen cometen en promedio 15.5 errores en la prueba de evaluación fonológica, frente a los 10.1 errores que cometen los que sí leen, siendo esta diferencia de 5.4 puntos estadísticamente significativa ($P < .001$).

El IC95% del $\exp(B)$ del término *Lectura1* expresa el intervalo de la *odds ratio* (OR). Las OR se interpretan de modo análogo a como lo hemos hecho con el riesgo relativo (RR). En este caso, dado que $OR < 1$, se interpretaría como la reducción relativa en el número de errores que supone que los participantes sean lectores respecto a no que no lo sean: $[p(\text{Lectura=No}) = 15.5] * [\exp(B)=0.648] = 10.04 \approx [p(\text{Lectura=Sí})=10.1]$. Por tanto, podemos concluir que **los participantes lectores cometen entre un $([OR_{ls}=0.776-1]*100=)$ 22.4% y un $([OR_{li}=0.539-1]*100=)$ 46.1% menos errores en la prueba de evaluación fonológica que los participantes no lectores.**

En este estudio el recuento de errores cometidos en la prueba de evaluación fonológica no se llevó a cabo sobre el mismo número de palabras, sino que cada participante del estudio evaluó un número de palabras distinto (*ExpoEF*) durante el tiempo establecido para la prueba, tiempo éste que sí se mantuvo constante. En esta situación los cálculos requieren transformar los recuentos en **tasas (rates)** tomando en consideración la diferente “exposición” de cada participante. En los programas de análisis estadístico, este tipo de ajuste se realiza indicando como *exposure variable* aquella que recoge el número de palabras a las que estuvo “expuesto” cada participante e indicando *Poisson (rates)* como tipo de intervalo de confianza. Desafortunadamente, jamovi no incluye, todavía, la posibilidad de incluir la variable de exposición en los análisis con variables de recuento.

4.6.3. Relación entre una variable categórica ordinal o politómica y una variable cuantitativa continua

Para ilustrar el procedimiento de análisis y la interpretación de los resultados de la relación entre una **variable categórica ordinal o politómica** y una **variable cuantitativa**, vamos, en primer lugar, a generar la variable ordinal *ND* (nivel de discapacidad), para analizar después su relación con el coeficiente de inteligencia (*QI*) de los participantes del estudio, es decir, si existen diferencias en el promedio de puntuación del *QI* entre los participantes según sea su nivel de discapacidad.

1. Para generar la variable *ND*, selecciona la columna de la variable *GD* que vamos a recodificar y pulsa **Data | Transform**.
2. Escribe *ND* como nombre para la nueva variable recodificada y **Nivel de discapacidad** como **Description**. A continuación, despliega la lista **using transform**, elige la opción **Create New Transform...** y haz clic en **+ Add recode condition**.
3. Se abre un nuevo panel de opciones en el que se pueden introducir las reglas de transformación y también definir la escala de medida de la nueva variable. Elige **Ordinal** en la lista **Measure Type**.
4. A continuación, introduce las reglas de recodificación que se muestran en la siguiente captura de pantalla para definir las categorías de la nueva variable *ND* a partir del contenido de la variable cuantitativa *GD*. Hay que ir haciendo clic en **+ Add recode condition** para añadir nuevas filas para definir las reglas de transformación. Las etiquetas de las categorías deben escribirse entre comillas.

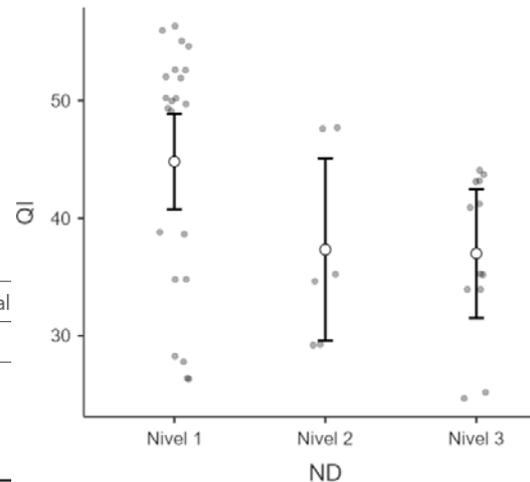


5. Cierra el panel de opciones de recodificación pulsando las flechas **↓** y **↑** que se encuentran en la parte superior derecha del panel.
6. Guarda los cambios realizados en la matriz de datos ejecutando **Save** del menú principal **≡** de jamovi.
7. Ejecuta el menú **Analyses | ANOVA | ANOVA** y pasa la variable *QI* a **Dependent Variables** y la variable *ND* a **Fixed Factors**. En el panel principal de opciones marca la opción η^2 en **Effect Size**.

Despliega el bloque **Estimated Marginal Means** para que jamovi incluya también en los resultados el gráfico de medias de los grupos con su intervalo de confianza y la nube de puntos, así como la tabla de descriptivos de *QI* para los grupos definidos por los niveles de discapacidad *ND*. Pasa la variable categórica *ND* como **Term 1** de **Marginal Means** y, a continuación, selecciona las opciones de **Output Marginal means plots** y **Marginal means tables**, y las opciones **Confidence interval** en la lista **Error Bars** y **Observed scores** del bloque **Plot**.

Estimated Marginal Means - ND

ND	Mean	SE	95% Confidence Interval	
			Lower	Upper
Nivel 1	44.8	2.00	40.8	48.9
Nivel 2	37.3	3.83	29.6	45.1
Nivel 3	37.0	2.71	31.5	42.5



Las medias aritméticas de la puntuación del *QI* de los participantes del estudio es superior en los participantes con un nivel 1 de discapacidad (44.8 puntos), seguidos de los otros dos grupos, que tienen puntuaciones del *QI* similares (37.3 y 37 puntos).

El gráfico de medias marginales muestra claramente las diferencias comentadas entre la media del primer grupo de participantes y las de los otros dos. Observa, empero, que la dispersión que refleja la nube de puntos es bastante grande en el primer grupo.

- Despliega el bloque de opciones **Post Hoc Tests**, pasa la variable *ND* a la lista, y marca **Tukey** en **Correction** y **Cohen's d** y **Confidence Interval** en **Effect Size**.

ANOVA - QI

	Sum of Squares	df	Mean Square	F	p	η^2
ND	588	2	294.2	3.33	0.047	0.153
Residuals	3265	37	88.2			

Post Hoc Comparisons - ND

Comparison	ND	ND	Mean Difference	SE	df	t	Ptukey	95% Confidence Interval		
								Cohen's d	Lower	Upper
Nivel 1 - Nivel 2	-	Nivel 2	7.485	4.33	37.0	1.7301	0.208	0.7968	-0.1550	1.75
			7.818	3.37	37.0	2.3193	0.065	0.8323	0.0792	1.59
Nivel 2 - Nivel 3	-	Nivel 3	0.333	4.70	37.0	0.0710	0.997	0.0355	-0.9776	1.05

Note. Comparisons are based on estimated marginal means

En la primera tabla de resultados aparece el valor del índice de magnitud del efecto $\eta^2 = 0.153$, que indica que, según los criterios propuestos por Cohen (1988), existen diferencias de una magnitud grande entre al menos dos de los grupos de participantes definidos por su nivel de discapacidad.

Según esos mismos criterios, los valores del índice de tamaño del efecto *d* de Cohen indican que la diferencia entre el grupo con nivel 1 de discapacidad y los otros dos grupos es de magnitud moderada ($d = 0.797$ y $d = 0.832$). La diferencia en el promedio de la puntuación del *QI* entre los grupos con niveles 2 y 3 de discapacidad es de magnitud muy pequeña ($d = 0.036$).

4.6.4. Relación entre una variable categórica ordinal o politómica y una variable cuantitativa de recuento

Por último, vamos a analizar la relación entre la **variable categórica ordinal** *ND* (nivel de discapacidad) y la **variable cuantitativa [discreta] de recuento** *ErrEF* (número de errores en la prueba de evaluación fonológica). Al igual que hemos hecho en el apartado 5.2, vamos a tomar en consideración el hecho de que la distribución muestral de las medias (y de la diferencia entre medias) de variables de recuento no se ajusta a una ley Normal (o *t* de Student) sino a una ley de probabilidad de Poisson.

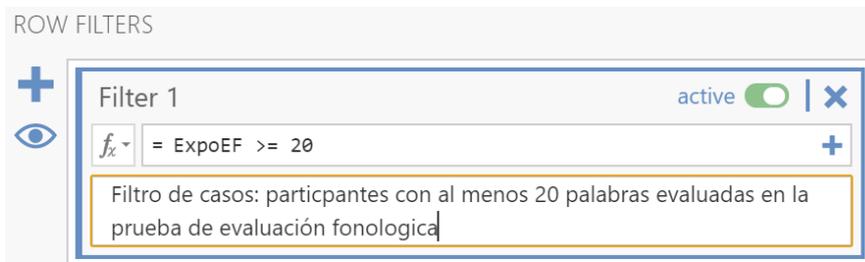
Como ya se ha comentado, en el estudio que estamos analizando el recuento de errores cometidos en la prueba de evaluación fonológica no se llevó a cabo sobre el mismo número de palabras, sino que cada participante del estudio evaluó un número de palabras distinto (*ExpoEF*) durante el tiempo establecido para la prueba, tiempo éste que sí se mantuvo constante. Para analizar la relación entre *ND* y *ErrEF*, vamos a seleccionar previamente sólo los participantes que evaluaron como mínimo 20 palabras durante la prueba de evaluación fonológica.

Establecimiento de filtros de casos en los análisis estadísticos

Aunque la mayor parte de análisis estadísticos se realizan sobre la totalidad de los casos contenidos en la matriz de datos, en ocasiones se desea realizar dichos análisis para un subconjunto de ellos. La aplicación de estos filtros de casos puede estar motivada por diferentes razones:

- Se desea realizar análisis sólo para un subconjunto de la muestra de casos que tienen alguna o algunas características en común, y no se desea comparar estadísticamente los resultados de dichos análisis con los de otros grupos de sujetos. Por ejemplo, se desea realizar análisis sólo para los sujetos de una determinada localización geográfica, o sólo para los sujetos de un sexo determinado, etc.
- Se desea excluir de los análisis los sujetos con valores faltantes debidos a que las variables eran de tipo **no aplicable**.
- Se desea incluir sólo a una muestra aleatoria del total de casos para llevar a cabo algún tipo de validación de la información, o bien, de los resultados de los análisis estadísticos.

1. Ejecuta el menú **Data | Filters** para establecer el filtro mencionado en la matriz de datos.
2. En el panel de opciones escribe la condición lógica que deben cumplir los casos que se incluirán en los análisis, tal y como muestra la siguiente captura de pantalla:



3. Cierra el panel de opciones pulsando **↑** y observa que en el panel de datos se ha añadido una primera columna *Filter 1* que indica qué casos se incluirán y cuáles se excluirán de los análisis a partir de este momento. Los casos excluidos se muestran sombreados:

	Filter 1	Id	Sexo	Edad	Lectura	GD	ND	QI	ErrEF	ExpoEF	
17	✓	17	Femenino	22	No		75 Nivel 2		29	24	49
18	✓	18	Masculino	23	Sí		75 Nivel 2		35	24	46
19	✓	19	Masculino	21	No		80 Nivel 3		25	25	47
20	✓	20	Masculino	19	No		93 Nivel 3		34	28	50
21	✗	21	Femenino	44	Sí		65 Nivel 1		55	0	15
22	✗	22	Femenino	43	Sí		66 Nivel 1		53	5	17
23	✗	23	Masculino	40	No		78 Nivel 3		43	6	17
24	✓	24	Femenino	36	Sí		64 Nivel 1		56	6	21
25	✓	25	Femenino	36	Sí		69 Nivel 1		26	7	26
26	✓	26	Masculino	20	No		65 Nivel 1		40	7	27

4. Ejecuta el menú **Analyses | Linear Models | Generalized Linear Models**. Ejecuta este menú y en el panel principal de opciones marca **Poisson** en **Frequencies**, pasa la variable *ErrEF* a **Dependent**

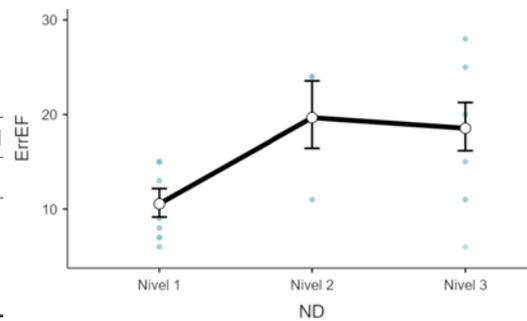
Variable y la variable *ND* a **Factors**. A continuación, marca **Odds Ratios (expB)** en **Effect Size** y **For exp(B)** en **Confidence Intervals**. Despliega el bloque de opciones **Factors Coding** y elige la opción **simple** para *ND*. A continuación, despliega el bloque **Plots** y pasa la variable *ND* a **Horizontal axis** para obtener también el gráfico de medias con sus IC95% (marca la opción **Confidence intervals** en **Display** y la opción **Observed scores** en **Plot**). Despliega también el bloque **Estimated Marginal Means** y marca la opción **Estimated Marginal Means**. Por último, despliega el bloque **Post Hoc Tests** y pasa la variable *ND* a la lista de la derecha.

Parameter Estimates

Names	Effect	Estimate	SE	exp(B)	95% Exp(B) Confidence Interval		z	p
					Lower	Upper		
(Intercept)	(Intercept)	2.752	0.0455	15.67	14.31	17.11	60.47	< .001
ND1	Nivel 2 - Nivel 1	0.622	0.1172	1.86	1.48	2.34	5.31	< .001
ND2	Nivel 3 - Nivel 1	0.564	0.1008	1.76	1.44	2.14	5.59	< .001

ND

ND	Mean Count	SE	95% Confidence Interval	
			Lower	Upper
Nivel 1	10.6	0.766	9.16	12.2
Nivel 2	19.7	1.810	16.42	23.6
Nivel 3	18.5	1.298	16.17	21.3



Post Hoc Comparisons - ND

Comparison		exp(B)	SE	z	Pbonferroni
ND	ND				
Nivel 2	- Nivel 3	1.060	0.1226	0.508	1.000
Nivel 1	- Nivel 2	0.537	0.0629	-5.309	< .001
Nivel 1	- Nivel 3	0.569	0.0574	-5.590	< .001

Los jóvenes y adultos con síndrome de Down de la muestra de este estudio con un nivel 1 de discapacidad cometen en promedio 10.6 errores en la prueba de evaluación fonológica, frente a los 19.7 que cometen los que tienen un nivel 2 de discapacidad y los 18.5 que cometen los que tienen un nivel 3, siendo la diferencia entre el primer grupo y los otros dos estadísticamente significativa ($P < .001$), y no significativa la diferencia entre los grupos con nivel 2 y 3 de discapacidad.

Los IC95% de los *exp(B)* de los términos *ND1* y *ND2* del modelo expresan los intervalos de las *odds ratios* (OR). En este caso, puesto que las $OR > 1$, se interpretarían como el incremento relativo en la puntuación del QI de los grupos con niveles 2 y 3 de discapacidad respecto al grupo con nivel 1: el grupo con nivel 2 entre 1.48 y 2.34 veces más, y entre 1.44 y 2.14 veces más el grupo con nivel 3. Por tanto, podemos concluir que los participantes con niveles de discapacidad 2 o 3 presentan, aproximadamente, entre 1.5 ($OR_{\text{Nivel}_2 - \text{Nivel}_1} = 1.48$ y $OR_{\text{Nivel}_3 - \text{Nivel}_1} = 1.44$) y un poco más del doble ($OR_{\text{Nivel}_2 - \text{Nivel}_1} = 2.34$ y $OR_{\text{Nivel}_3 - \text{Nivel}_1} = 2.14$) de errores en la prueba de evaluación fonológica que los participantes con nivel 1 de discapacidad.

Como ya hemos comentado anteriormente, este análisis requeriría incluir en el modelo como "exposure variable" la variable *ExpoEF* que recoge el número de palabras a las que estuvo "expuesto" cada participante. Desafortunadamente, jamovi no incluye todavía esta posibilidad.

5. Guarda todos los análisis realizados en este apartado en el fichero *HABITOX_P4User.omv*.

4.7. Ejercicio

Parte 1. Preparación de los datos (generación de variables)

1. Abre el fichero de datos *DOWN.omv*.
2. A partir de la *Edad* registrada, crea la variable *Edad2* que contenga las siguientes dos categorías:

<i>Edad</i>	<i>Edad2</i>	
Valor	Valor	Etiqueta
19 a 26	0	<i>Joven</i>
27 a 48	1	<i>Adulto/a</i>

3. Si todavía no lo has hecho, a partir del grado de discapacidad registrado (*GD*), crea la variable nivel de discapacidad (*ND*) con las tres categorías indicadas en el apartado 5.3.

Parte 2. Inferencia estadística

4. Estudios previos han establecido que el grado de discapacidad promedio de los jóvenes con síndrome de Down se sitúa alrededor de los 75 puntos (escala de 0 a 100). ¿La muestra del presente estudio es representativa de la población con respecto de esta característica?

5. ¿Cuál es el promedio estimado del número de errores que cometen en la prueba de evaluación fonológica las mujeres jóvenes con síndrome de Down de la población origen de la muestra de estudio?

6. ¿Y de la puntuación del cociente de inteligencia (*QI*)?

7. ¿El coeficiente de inteligencia (*QI*) promedio de los participantes de la muestra de estudio es compatible con una población caracterizada por un *QI* promedio de 50 puntos (escala 0 a 100)? ¿Y con una población con un promedio de 40 puntos?

8. En la población origen de la muestra de estudio, ¿cuál es la proporción de jóvenes y adultos con síndrome de Down que tienen un nivel 3 de discapacidad (*ND*)?

9. El resultado obtenido en la pregunta anterior es compatible con una población caracterizada por una prevalencia del 25% de jóvenes y adultos con síndrome de Down con un nivel 3 de discapacidad. ¿Y con una población con una prevalencia del 40%? Anota tu respuesta:

Parte 3. Contraste de hipótesis bivalente

10. A partir de los datos obtenidos en el presente estudio, ¿se confirma la hipótesis de que las mujeres con síndrome de Down obtienen mejores resultados en la prueba de evaluación fonológica (*ErrEF*) que los hombres (*Sexo*)? ¿De qué magnitud es esta diferencia?

11. A partir de los datos obtenidos en el presente estudio, ¿se confirma la hipótesis de que existe una asociación negativa tanto entre los resultados en la prueba de evaluación fonológica (*ErrEF*) y la *Edad*, como entre dichos resultados (*ErrEF*) y el cociente intelectual (*QI*)?

12. A partir de los datos de la muestra de estudio, ¿se puede afirmar que existe una asociación entre el cociente de inteligencia (*QI*) y el grado discapacidad (*GD*)? ¿En qué sentido y magnitud? Anota tu respuesta:

13. En la población origen de la muestra de estudio, ¿la distribución del nivel de discapacidad (ND) es la misma en hombres y mujeres (*Sexo*)? Anota tu respuesta:

14. En la población origen de la muestra de estudio, ¿hombres y mujeres (*Sexo*) difieren en cuanto a su rendimiento en la prueba de evaluación fonológica (*ErrEF*)? ¿Dicha diferencia, en caso de que sea estadísticamente significativa, es también relevante? Anota tu respuesta:

15. ¿Y en cuanto a su cociente intelectual (*QI*)? Anota tu respuesta:

16. En la población origen de la muestra de estudio, ¿existe relación entre la *Edad* y el nivel de discapacidad (*ND*)? Anota tu respuesta:

17. ¿Y entre ser jóvenes o adultos (*Edad2*) y el nivel de discapacidad (*ND*)? Anota tu respuesta:

18. Guarda los resultados de los análisis en el fichero jamovi *DOWN_P4User.omv*.

19. Compara los análisis realizados y tus resultados con los del fichero jamovi *DOWN_P4Sol.omv*.

4.8. Síntesis de menús y comandos jamovi para la descripción estadística, la estimación de intervalos de confianza y el contraste de hipótesis univariante y bivalente

A0. Filtro de datos

2. Seleccionar filas de datos para el análisis: Data | Filters [pág. 89]
 - ✓ Row Filters: expresión lógica de filtro
 - ✓ Active / Inactive

A1. Intervalos de confianza y pruebas de conformidad con una variable categórica

1. Gráfico de barras agrupadas o apiladas: Analyses | Exploration | Survey Plots [pág. 71]
 - ✓ Variables: variable categórica
 - ✓ Nominal / Ordinal Plots | Plot Type: Grouped bar o Stacked bar
 - ✓ Nominal / Ordinal Plots | Frequency Type: Percentages
 - ✓ Nominal / Ordinal Plots | Additional Options: (opcional: Hide missing values)
2. Índices estadísticos: Analyses | Exploration | Descriptives [pág. 71]
 - ✓ Variables: variable categórica
 - ✓ Frequency tables
 - ✓ Statistics | Sample size: N, Missing
 - ✓ Statistics | Central Tendency: desmarcar Mean, Median
 - ✓ Statistics | Dispersion: desmarcar Minimum, Maximum
3. Intervalo de confianza: Analyses | esci | Estimate Proportion [pág. 71]
4. Prueba de conformidad: Analyses | Frecuencias | 2 Outcomes [pág. 71]
 - ✓ Hypothesis: Test value (valor teórico)
 - ✓ Hypothesis: \neq Test value o $>$ Test value o $<$ Test value

A2. Intervalos de confianza y pruebas de conformidad con una variable cuantitativa

1. Histograma: Analyses | Exploration | Descriptives [pág. 75]
 - ✓ Variables: variable cuantitativa
 - ✓ Plots | Histograms: Histogram (opcional: Density)
2. Diagrama de caja (boxplot): Analyses | Exploration | Descriptives [pág. 75]
 - ✓ Variables: variable cuantitativa
 - ✓ Plots | Box Plots: Box Plot (opcional: Violin y/o Data)
3. Índices estadísticos: Analyses | Exploration | Descriptives [pág. 75]
 - ✓ Variables: variable cuantitativa
 - ✓ Statistics | Sample size: N, Missing
 - ✓ Statistics | Central Tendency: Mean, Median, Mode
 - ✓ Statistics | Percentile Values: Quartiles (opcional: Cut points for ___ equal groups)
 - ✓ Statistics | Dispersion: Std. deviation, Minimum, Maximum
4. Intervalo de confianza (X continua): Analyses | esci | Estimate Mean [pág. 75]
5. Intervalo de confianza (X recuento): Analyses | Linear Models | Generalized Linear Models [pág. 77]
 - ✓ Frecuencias: Poisson
 - ✓ Effect Size: Odds Ratios (expB)
 - ✓ Confidence Intervals: For exp(B)
6. Prueba de conformidad (X continua): Analyses | T-test | One Sample T-Test [pág. 75]
 - ✓ Hypothesis: Test value (valor teórico)
 - ✓ Hypothesis: \neq Test value o $>$ Test value o $<$ Test value
 - ✓ Additional Statistics: Mean difference, Confidence interval, Effect size, Confidence interval
 - ✓ Percentile Values: Quartiles (opcional: Cut points for ___ equal groups)

A3. Relación lineal entre dos variables cuantitativas (correlación lineal)

1. **Diagrama de dispersión (*scatter-plot*):** Analyses | Exploration | Scatterplot [pág. 79]
 - ✓ X-Axis: variable X (independiente o exposición)
 - ✓ Y-Axis: variable Y (dependiente o respuesta)
 - ✓ Regression Line: Linear
2. **Índice de correlación de Pearson:** Analyses | Regression | Correlation Matrix [pág. 79]
 - ✓ Additional Options: Report significance, Flag significant correlations, Ny Confidence intervals
 - ✓ Tamaño del efecto: small: $r \geq 0.10$; medium: $r \geq 0.30$; large: $r \geq 0.50$; very large: $r \geq 0.70$

A4. Relación entre dos variables categóricas (comparación de proporciones)

1. **Gráfico de barras agrupadas o apiladas:** Analyses | Exploration | Survey Plots [pág. 82]
 - ✓ Variables: variable Y (dependiente o respuesta)
 - ✓ Grouping variable: variable X (independiente o exposición)
 - ✓ Nominal / Ordinal Plots | Plot Type: Grouped bar o Stacked bar
2. **Intervalo de confianza:** Analyses | esci | Estimate Proportion Difference [pág. 79]
 - ✓ Outcome variable: variable Y (dependiente o respuesta)
 - ✓ Category variable: variable X (independiente o exposición)
3. **Prueba de significación (tabla de contingencia):** Analyses | Frequencies | Contingency Tables | Independent Samples [pág. 79]
 - ✓ Rows: variable Y (dependiente o respuesta)
 - ✓ Columns: variable X (independiente o exposición)
 - ✓ Statistics | Tests: χ^2
 - ✓ Statistics | Comparative Measures (2x2 only): Relative Risk, Difference in proportions, Confidence intervals
 - ✓ Cells | Percentages: Row (para obtener los porcentajes por filas)
⇒ Tamaño del efecto: diferencia (DP) o razón (RR) entre porcentajes de cada fila

A5. Relación entre una variable categórica y una variable cuantitativa (comparación de medias)

1. **Histograma:** Analyses | Exploration | Descriptives [pág. 83]
 - ✓ Variables: variable cuantitativa
 - ✓ Split by: variable categórica
 - ✓ Plots | Histograms: Histogram (opcional: Density)
2. **Gráfico de barras:** Analyses | Exploration | Descriptives [pág. 83]
 - ✓ Variables: variable cuantitativa
 - ✓ Split by: variable categórica
 - ✓ Plots | Bar Plots: Bar Plot
3. **Diagrama de caja (*boxplot*):** Analyses | Exploration | Descriptives [pág. 83]
 - ✓ Variables: variable cuantitativa
 - ✓ Split by: variable categórica
 - ✓ Plots | Box Plots: Box Plot (opcional: Violin y/o Data)
4. **Intervalo de confianza y prueba de significación (X categórica binaria - Y continua):** Analyses | T-test | Independent Samples T-Test [pág. 84]
 - ✓ Grouping Variable: variable categórica X
 - ✓ Dependent Variable: variable cuantitativa continua Y
 - ✓ Additional Statistics: Mean difference, Confidence Interval, Effect size, Confidence Interval, Descriptives, Descriptives plots
 - ✓ Hypothesis: Group 1 \neq Group 2 o Group 1 $>$ Group 2 o Group 1 $<$ Group 2
⇒ Tamaño del efecto: small: $d \geq 0.20$; medium: $d \geq 0.50$; large: $d \geq 0.80$; very large: $d \geq 1.20$; huge: $d \geq 2.0$

5. Intervalo de confianza y prueba de significación (X categórica binaria - Y recuento):

Analyses | Linear Models | Generalized Linear Models

[pág.86]

- ✓ Frequencies: Poisson
- ✓ Factors: variable categórica X
- ✓ Dependent Variable: variable cuantitativa de recuento Y
- ✓ Effect Size: Odds Ratios (expB)
- ✓ Confidence Intervals: For exp(B)
- ✓ Factors Coding: simple
- ✓ Plots: Horizontal axis (variable X categórica)
- ✓ Plots | Display: Confidence intervals
- ✓ Plots | Plot: Observed scores
- ✓ Estimated Marginal Means: Estimated Marginal Means
⇒ Tamaño del efecto: small: $d \geq 0.20$; medium: $d \geq 0.50$; large: $d \geq 0.80$; very large: $d \geq 1.20$; huge: $d \geq 2.0$

6. Intervalo de confianza y prueba de significación (X categórica ordinal o politómica - Y continua):

Analyses | ANOVA | ANOVA

[pág.87]

- ✓ Fixed Factors: variable categórica X
- ✓ Dependent Variables: variable cuantitativa continua Y
- ✓ Effect size: η^2
⇒ Tamaño del efecto η^2 : small: $\eta^2 \geq 0.01$; medium: $\eta^2 \geq 0.06$; large: $\eta^2 \geq 0.14$
- ✓ Contrasts: none o deviation o simple o difference o helmert o repeated o polynomial
- ✓ Post Hoc Tests | Correction: No correction y/o Tukey y/o Scheffe y/o Bonferroni y/o Holm
- ✓ Post Hoc Tests | Effect Size: Cohen's d, Confidence Interval
⇒ Tamaño del efecto: small: $d \geq 0.20$; medium: $d \geq 0.50$; large: $d \geq 0.80$; very large: $d \geq 1.20$; huge: $d \geq 2.0$
- ✓ Estimated Marginal Means | Marginal Means: Term 1 (variable X independiente o exposición)
- ✓ Estimated Marginal Means | Output: Marginal means plots, Marginal means tables
- ✓ Estimated Marginal Means | Plot | Error bars: Confidence interval
- ✓ Estimated Marginal Means | Plot: Observed scores

7. Intervalo de confianza y prueba de significación (X categórica ordinal o politómica - Y recuento):

Analyses | Linear Models | Generalized Linear Models

[pág.89]

- ✓ Frequencies: Poisson
- ✓ Factors: variable categórica X
- ✓ Dependent Variable: variable cuantitativa de recuento Y
- ✓ Effect Size: Odds Ratios (expB)
- ✓ Confidence Intervals: For exp(B)
- ✓ Factors Coding: simple
- ✓ Plots: Horizontal axis (variable X categórica)
- ✓ Plots | Display: Confidence intervals
- ✓ Plots | Plot: Observed scores
- ✓ Estimated Marginal Means: Estimated Marginal Means
- ✓ Post Hoc Tests: pasa la variable X (independiente o exposición) a la lista
⇒ Tamaño del efecto: small: $d \geq 0.20$; medium: $d \geq 0.50$; large: $d \geq 0.80$; very large: $d \geq 1.20$; huge: $d \geq 2.0$

5. Ejercicios basados en estudios reales

En los siguientes ejercicios se revisan los principales procedimientos estadísticos que incorpora jamovi para realizar análisis estadísticos descriptivos, tanto de una variable categórica o cuantitativa, como de la relación entre dos variables. En estos análisis se introduce también el potente procedimiento de generación de gráficos que incorpora jamovi para complementar visualmente la descripción de variables y relaciones basada en índices estadísticos.

5.1. Lenguaje en adultos con síndrome de Down

El ejemplo que se propone en este ejercicio se plantea a partir de un estudio llevado a cabo por estudiantes del grado en Logopedia de la Facultat de Psicologia de la UAB. La muestra está formada por 40 personas adultas con edades comprendidas entre los 19 y los 48 años con Síndrome de Down, 20 de ellos hombres y 20 mujeres.

Las **variables registradas**, todas ellas numéricas, son las siguientes:

Nombre	Etiqueta de la variable	Etiqueta de los códigos
Id	Identificador (se puede obviar la etiqueta)	
Sexo	Sexo	0: Masculino; 1: Femenino
Edad	Edad en años cumplidos	
Lectura	Lectura	0: No; 1: Sí
GD	Grado de discapacidad (escala 0 a 100)	
QI	Cociente intelectual (escala 0 a 200)	
ErrEF	Número de errores en la prueba de evaluación fonológica (EF)	
ExpoEF	Número de términos evaluados en la prueba de EF	

	Id	Sexo	Edad	Lectura	GD	QI	ErrEF	ExpoEF
1	1	Femenino	43	Sí	65	55	0	16
2	2	Femenino	48	Sí	66	53	5	17
3	3	Femenino	39	No	78	43	6	20
4	4	Femenino	35	Sí	64	56	6	22
5	5	Femenino	37	No	69	26	7	27
6	6	Masculino	29	No	65	49	7	29
7	7	Femenino	28	Sí	67	52	8	29
8	8	Femenino	25	Sí	63	50	9	30
9	9	Femenino	30	No	82	44	11	35
10	10	Masculino	23	No	72	48	11	34
11	11	Masculino	24	No	65	39	13	41
12	12	Masculino	28	No	65	28	15	43
13	13	Femenino	26	Sí	70	35	15	42
14	14	Masculino	25	No	82	35	15	45
15	15	Masculino	27	Sí	64	50	15	43
16	16	Masculino	24	No	78	41	20	48
17	17	Femenino	22	No	75	29	24	49
18	18	Masculino	23	Sí	75	35	24	46
19	19	Masculino	21	No	80	25	25	47
20	20	Masculino	19	No	93	34	28	50
21	21	Femenino	44	Sí	65	55	0	15
22	22	Femenino	43	Sí	66	53	5	17
23	23	Masculino	40	No	78	43	6	17
24	24	Femenino	36	Sí	64	56	6	21
25	25	Femenino	36	Sí	64	56	6	21
26	26	Femenino	36	Sí	64	56	6	21
27	27	Femenino	36	Sí	64	56	6	21
28	28	Femenino	36	Sí	64	56	6	21
29	29	Femenino	36	Sí	64	56	6	21
30	30	Femenino	36	Sí	64	56	6	21
31	31	Femenino	36	Sí	64	56	6	21
32	32	Femenino	36	Sí	64	56	6	21
33	33	Femenino	36	Sí	64	56	6	21
34	34	Femenino	36	Sí	64	56	6	21
35	35	Femenino	36	Sí	64	56	6	21
36	36	Femenino	36	Sí	64	56	6	21
37	37	Femenino	36	Sí	64	56	6	21
38	38	Femenino	36	Sí	64	56	6	21
39	39	Femenino	36	Sí	64	56	6	21
40	40	Femenino	36	Sí	64	56	6	21

La **variable de respuesta objeto de estudio (ErrEF)** informa del número de errores en la prueba de evaluación fonológica, de manera que una puntuación más alta es indicativa de un peor resultado.

En este estudio se desea contrastar las siguientes **hipótesis**:

- **H1:** Habrá una asociación positiva entre los resultados en la prueba de evaluación fonológica (*ErrEF*) y el grado de discapacidad (*GD*), mientras que la relación con el cociente intelectual (*QI*) será negativa.
- **H2:** Se obtendrán mejores resultados en la prueba de evaluación fonológica (*ErrEF*) en las personas que saben leer que en las que no saben leer (*Lectura*).

Ejercicio:

1. Los datos se proporcionan en el archivo *DOWN.xlsx*. Abre jamovi e importa el fichero de datos siguiendo los pasos indicados en el procedimiento P1J.
2. Etiqueta las variables y los códigos de las variables categóricas. Guarda el fichero de datos jamovi con el nombre *DOWN.omv*.
3. A partir de la *Edad* registrada, crea la variable *Edad2* que contenga las siguientes dos categorías de edad:

<i>Edad</i>	<i>Edad2</i>	
Valor	Valor	Etiqueta
19 a 26	0	Joven
27 a 48	1	Adulto

4. Realiza una descripción estadística gráfica y numérica de las respuestas obtenidas en las variables categóricas *Sexo*, *Edad2* y *Lectura* y en las variables cuantitativas *Edad*, *GD*, *QI* y *ErrEF*.
5. Evalúa numérica y gráficamente si hay diferencias en la *Lectura* en función del *Sexo* y la edad (*Edad* y *Edad2*).
6. Contrasta estadísticamente la hipótesis H1:
 - ¿Hay relación entre el número de errores en la prueba de evaluación fonológica (*ErrEF*) y las medidas de discapacidad (*GD*) e inteligencia (*QI*)?
 - En caso de que exista relación entre estas variables, ¿dicha relación se da en el sentido planteado en la hipótesis?
 - ¿Qué puedes decir acerca de la magnitud de la asociación entre estas variables?
7. Contrasta estadísticamente la hipótesis H2:
 - ¿Los resultados obtenidos en la prueba de evaluación fonológica varían en función de si los participantes saben o no leer?
 - En caso de que existan, ¿dichas diferencias se dan en el sentido planteado en la hipótesis?
 - ¿Qué puedes decir acerca de la magnitud de la diferencia observada entre lectores y no lectores en la prueba de evaluación fonológica?
8. Guarda los resultados de los análisis en el fichero jamovi *DOWN_User.omv*.
9. Compara los análisis realizados y tus resultados con los del fichero jamovi *DOWN_Sol.omv*.

5.2. Evocación del léxico nominal por niños/as con implante coclear en una tarea de denominación

En este estudio se analiza si los/las niños/as con implante coclear construyen el léxico nominal de la misma manera como lo hacen los/las niños/as con audición típica y, por lo tanto, si su evocación presenta las mismas características o, por el contrario, presentan dificultades lexicales que se manifiestan en la evocación cuando deben utilizar el léxico nominal en una actividad de denominación.

La muestra de estudio está formada por 40 niños sordos (21 niñas y 19 niños), de edades comprendidas entre los 5 y los 7 años, escolarizados en centros públicos en los niveles educativos de educación infantil o primer curso de educación primaria y con el catalán como lengua de aprendizaje. A cada niño/a sordo/a se les buscó una pareja de control con audición típica, del mismo grupo de clase y misma edad y sexo. La muestra total está formada, por tanto, por 80 niños y niñas.

El material utilizado para la actividad de evocación (denominación) fueron fotos a color basadas en las imágenes de la prueba AREHA (Aguilar y Serra, 2003), que tiene por objetivo evaluar los procesos fonológicos desde los 3 a los 6 años.

Las **variables registradas** fueron las siguientes:

Nombre	Etiqueta de la variable	Etiqueta de los códigos
Caso	Identificador del participante	
Genero	Género	0: Femenino; 1: Masculino
Edad	Edad en años cumplidos	
Ciclo	Ciclo formativo	0: Primaria; 1: Infantil
Idioma	Idioma materno y paterno	1: Catalán y castellano; 2: Castellano; 3: Catalán; 4: Otros
Audicion	Audición	0: Oyente; 1: Sordo/a
EdadImp	Edad en meses cuando se realizó el implante coclear	
EdadAud	Edad auditiva (número de meses que lleva el implante)	
Tipolmp	Tipo de implante coclear	0: Unilateral; 1: Bilateral
NP	Número de preguntas realizadas por la logopeda	
NoP	Número de preguntas no realizadas por la logopeda	
NC	Número de preguntas que no contesta	
NS	Número de preguntas que responde "no sé"	
NPResp	Número de preguntas respondidas	
Conoce	Número de nombres que evoca correctamente	
Evoca	Número de nombres que evoca correctamente sin latencia	
TiempoTot	Tiempo total en segundos que tarda en evocar todos los nombres	
DifSL	Número de nombres con dificultades semánticas-lexicales (solo de significado expandido)	
DifAC	Número de nombres evocados con dificultad de acceso	
DifE	Número de nombres con errores de evocación	

Caso	Genero	Edad	Ciclo	Idioma	Audicion	EdadImp	EdadAud	TipImp	NP	Nop	NC	NS	NPResp	Conoce	Evoca	Tiempotot	DI/SL	DI/AC	DI/E
30	7944 Masculino	5	Infantil (P5)	Catalan y cast...	Oyente				37	4	1	0	36	32	32	4	0	1	4
31	17934 Masculino	5	Infantil (P5)	Catalan y cast...	Sordo/a	15	49	Unilateral	37	4	3	0	34	24	23	4	0	1	11
32	7934 Masculino	5	Infantil (P5)	Oyente	Oyente				38	3	1	2	35	32	31	4	0	2	3
33	18924 Femenino	6	Infantil (P5)	Otros	Sordo/a	12	62	Unilateral	39	2	5	0	34	28	22	45	1	8	6
34	8924 Femenino	6	Infantil (P5)	Oyente	Oyente				38	3	1	1	36	30	29	18	0	4	5
35	18914 Femenino	6	Infantil (P5)	Castellano	Sordo/a	8	60	Bilateral	40	1	3	3	34	26	21	13	1	5	5
36	8914 Femenino	6	Infantil (P5)	Oyente	Oyente				40	1	1	0	39	35	31	12	0	5	6
37	18904 Femenino	7	Primaria (EP1)	Catalan y cast...	Sordo/a	48	42	Unilateral	38	3	12	1	25	23	18	14	0	7	2
38	8904 Femenino	7	Primaria (EP1)	Oyente	Oyente				38	3	1	4	33	26	25	5	0	2	7
39	18894 Femenino	6	Infantil (P5)	Catalan	Sordo/a	24	46	Bilateral	37	4	4	1	32	28	26	16	0	4	9
40	8894 Femenino	6	Infantil (P5)	Oyente	Sordo/a				39	2	0	3	36	29	26	3	1	3	6
41	19069 Femenino	6	Infantil (P5)	Catalan y cast...	Sordo/a	39	29	Unilateral	33	8	5	1	27	18	18	0	0	8	8
42	9069 Femenino	6	Infantil (P5)	Oyente	Oyente				36	5	2	2	32	25	25	0	1	0	6
43	19089 Femenino	6	Primaria (EP1)	Catalan	Sordo/a	24	51	Bilateral	35	6	4	0	31	24	24	0	0	0	6
44	9089 Femenino	6	Primaria (EP1)	Oyente	Oyente				33	8	0	1	32	26	24	12	1	2	5
45	19079 Femenino	7	Primaria (EP1)	Castellano	Sordo/a	18	65	Unilateral	33	8	1	0	32	26	25	3	0	1	6
46	9079 Femenino	7	Primaria (EP1)	Oyente	Oyente				33	8	0	3	30	26	25	4	0	1	4
47	17791 Masculino	7	Primaria (EP1)	Catalan y cast...	Sordo/a	23	58	Unilateral	32	9	2	1	29	24	23	27	1	3	3
48	7791 Masculino	7	Primaria (EP1)	Oyente	Oyente				37	4	3	1	33	26	26	0	1	0	3
49	18801 Femenino	7	Primaria (EP1)	Castellano	Sordo/a	52	26	Unilateral	34	7	3	2	29	19	18	28	3	4	8
50	8801 Femenino	7	Primaria (EP1)	Oyente	Oyente				35	6	6	0	29	24	20	25	3	4	8
51	19028 Femenino	5	Infantil (P5)	Catalan y cast...	Sordo/a	20	49	Unilateral	39	2	2	0	37	36	31	6	0	5	1
52	9028 Femenino	5	Infantil (P5)	Oyente	Oyente				39	2	0	0	39	39	35	8	0	4	0
53	19018 Femenino	6	Primaria (EP1)	Catalan	Sordo/a	51	28	Unilateral	40	1	0	2	38	37	26	23	0	11	1
54	9018 Femenino	6	Primaria (EP1)	Oyente	Oyente				41	0	0	0	41	38	36	6	0	2	3
55	18008 Masculino	6	Primaria (EP1)	Catalan	Sordo/a	7	70	Bilateral	39	2	0	1	38	38	33	8	0	5	0
56	8008 Masculino	6	Primaria (EP1)	Oyente	Oyente				40	1	0	0	40	40	37	4	0	0	3
57	17965 Masculino	7	Primaria (EP1)	Catalan	Sordo/a	18	62	Bilateral	40	1	1	0	39	38	0	0	0	0	1
58	7965 Masculino	7	Primaria (EP1)	Oyente	Oyente				40	1	1	0	39	37	36	4	0	2	2
59	17822 Masculino	5	Infantil (P5)	Castellano	Sordo/a	43	25	Unilateral	36	5	13	0	23	16	12	14	0	8	7
60	7822 Masculino	5	Infantil (P5)	Oyente	Oyente				38	3	0	1	37	29	25	6	0	0	8
61	18812 Femenino	7	Primaria (EP1)	Castellano	Sordo/a	43	31	Unilateral	40	1	12	0	28	25	19	8	0	4	3
62	8812 Femenino	7	Primaria (EP1)	Oyente	Oyente				41	0	2	0	39	38	32	17	0	6	1
63	17977 Masculino	5	Infantil (P5)	Catalan	Sordo/a	16	54	Unilateral	39	2	0	1	38	33	32	4	0	2	5
64	7977 Masculino	5	Infantil (P5)	Oyente	Oyente				40	1	2	0	38	33	30	7	1	3	3
65	18884 Femenino	6	Primaria (EP1)	Catalan	Sordo/a	35	41	Unilateral	35	6	3	0	32	25	24	2	0	1	6
66	8884 Femenino	6	Primaria (EP1)	Oyente	Oyente				37	4	0	0	37	28	24	10	0	4	9
67	17874 Masculino	6	Primaria (EP1)	Catalan y cast...	Sordo/a	25	49	Bilateral	37	4	3	3	31	20	14	15	0	6	11
68	8874 Masculino	6	Primaria (EP1)	Oyente	Oyente				38	3	1	0	37	27	24	6	0	3	10
69	17864 Masculino	5	Infantil (P5)	Catalan y cast...	Sordo/a	37	30	Unilateral	38	3	3	0	35	23	23	1	0	1	12
70	7864 Masculino	5	Infantil (P5)	Oyente	Oyente				37	4	0	0	37	35	35	1	0	1	2
71	17854 Masculino	6	Primaria (EP1)	Catalan y cast...	Sordo/a	17	62	Unilateral	37	4	3	2	32	18	18	5	0	2	13

Las hipótesis planteadas en este estudio son las siguientes:

Respecto a la comparación entre el grupo de niños/as con implante coclear y el grupo de niños/as con audición típica (*Audición*):

- **H1:** El grupo de niños/as con implante coclear producirá un mayor número de no respuestas (*NC* y *NS*) que el grupo de niños/as con audición típica.
- **H2:** El grupo de niños/as con implante coclear evocará una menor cantidad de nombres (*Conoce* y *Evoca*) que el grupo de niños/as con audición típica.
- **H3:** El grupo de niños/as con implante coclear y el grupo de niños/as con audición típica presentarán la misma cantidad de dificultades de evocación del tipo semántico-lexical (*DifSL*) y de acceso (*DifAC*), pero con más errores de evocación (*DifE*) en el grupo de niños/as con implante coclear.

En el grupo de niños/as con implante coclear:

- **H4:** La implantación precoz (*EdadImp*) y la mayor edad auditiva (*EdadAud*) permitirán evocar un mayor número de nombres (*Conoce* y *Evoca*), mientras que no se espera que el tipo de implante coclear (*Tipolmp*) sea una variable relevante en este sentido.
- **H5:** Se espera encontrar una menor cantidad de dificultades (*DifSL*, *DifAC* y *DifE*) en los niños a los que se les realizó el implante más precozmente (*EdadImp*) y en los de mayor edad auditiva (*EdadAud*), pero no se espera que el tipo de implante (*Tipolmp*) sea una variable relevante en este sentido.

Ejercicio:

2. Abre el archivo de datos jamovi *NAMING.omv*.
3. Contrasta estadísticamente la hipótesis H1:
 - ¿Hay diferencias entre los dos grupos de niños/as (*Audición*) respecto al número de no respuestas (*NC* y *NS*)?
 - En caso de que existan diferencias entre los dos grupos en estas variables, ¿dicha diferencia se da en el sentido planteado en la hipótesis?
 - ¿Qué puedes decir acerca de la magnitud de dichas diferencias?
4. Contrasta estadísticamente la hipótesis H2 siguiendo el mismo patrón de respuesta de la pregunta anterior.
5. Analiza las diferencias respecto a los tipos de dificultades de evocación planteados en la H3 y razona la respuesta en los términos indicados en las dos preguntas anteriores.
6. Antes de contrastar las hipótesis H4 y H5 relativas a los/las niños/as con implante coclear, comprueba que los grupos con implante unilateral y bilateral (*Tipolmp*) son comparables respecto al idioma materno y paterno (*Idioma*) y el género (*Genero*).
7. Comprueba también que no existen diferencias en el número de respuestas respondidas (*NPResp*) ni tampoco en la cantidad de nombres que evocan (*Conoce* y *Evoca*) en función del idioma materno y paterno (*Idioma*).
8. ¿Los datos recogidos en el grupo de niños/as con implante coclear permiten concluir en el sentido planteado por la H4? ¿Qué puedes decir acerca de la magnitud de las asociaciones analizadas?
9. Contrasta la H5 y concluye en términos de si existe asociación estadística entre las variables implicadas y, en el caso de que exista, si el sentido y la magnitud de dichas asociaciones son los planteados en la hipótesis.
10. Guarda los resultados de los análisis en el fichero jamovi *NAMING_User.omv*.
11. Compara los análisis que has realizado con los contenidos en el fichero jamovi *NAMING_Sol.omv*.

5.3. Percepción del profesorado tutor del alumnado sordo sobre la cultura y las prácticas educativas inclusivas

En este estudio se analiza la percepción del profesorado tutor del alumnado sordo sobre la realidad educativa en relación con la cultura y las prácticas inclusivas. Se utilizan los *índices de Ainscow* como marco general de consenso para la valoración de los criterios de inclusión educativa. La muestra está formada por 60 tutores/as de 53 centros educativos catalanes de educación infantil y primaria.

Las **variables registradas** son las siguientes:

Nombre	Etiqueta de la variable	Etiqueta de los códigos
Centro	Identificador del centro de enseñanza	
Ciclo	Ciclo formativo	0: Infantil; 1: Primaria
RE	Rendimiento auditivo (en dB) del alumnado sordo	
NivComp	Nivel de comprensión oral (Reynell; 0 a 100) alum. sordo	
NivExpre	Nivel de expresión oral (Reynell; 0 a 100) alum. sordo	
CursoAD	Ha recibido un curso sobre atención a la diversidad	0: No; 1: Sí
CursoTIC	Ha recibido un curso sobre el uso de las TIC	0: No; 1: Sí
NumCF	Número de cursos de formación recibidos	
AEFamO	Coordinación agente educativo: Familia (Oyentes)	0: Bajo; 1: Medio; 2: Alto
AEFamS	Coordinación agente educativo: Familia (Sordos/as)	0: Bajo; 1: Medio; 2: Alto
AEProfS	Coordinación agente educativo: Profesionales (Sordos/as)	0: Bajo; 1: Medio; 2: Alto
ExpecO	Expectativas respecto al alumnado Oyente	0: Bajo; 1: Medio; 2: Alto
ExpecS	Expectativas respecto al alumnado Sordo	0: Bajo; 1: Medio; 2: Alto
EDO	Estrategias docentes con el alumnado Oyente	0: Bajo; 1: Medio; 2: Alto
EDS	Estrategias docentes con el alumnado Sordo	0: Bajo; 1: Medio; 2: Alto
ASDS	Adaptaciones socioeducativas y didácticas con sordos/as	0: Bajo; 1: Medio; 2: Alto
CI	Percepción sobre el nivel de Cultura inclusiva del centro	1: Medio; 2: Alto
PI	Percepción sobre el nivel de Práctica inclusiva del centro	1: Medio; 2: Alto

	Centro	Ciclo	RE	NivComp	NivExpre	CursoAD	CursoTIC	NumCF	AEFamO	AEFamS	AEProfS	ExpecO	ExpecS	EDO	EDS	ASDS	CI	PI
1	1	Infantil	27	56	50	No	No	2	Alto	Alto	Bajo	Alto	Medio	Alto	Alto	Bajo	Alto	Alto
2	2	Primaria	18	50	65	No	Si	2	Alto	Alto	Medio	Alto	Medio	Medio	Alto	Alto	Alto	Alto
3	24	Primaria	31	60	39	Si	No	7	Medio	Alto	Medio	Alto	Medio	Alto	Medio	Alto	Medio	Alto
4	23	Primaria	26	67	59	Si	Si	11	Medio	Alto	Bajo	Medio	Medio	Medio	Alto	Bajo	Medio	Medio
5	13	Infantil	18	60	76	No	Si	8	Medio	Alto	Medio	Alto	Alto	Medio	Medio	Bajo	Alto	Medio
6	9	Infantil	38	35	52	Si	No	7	Alto	Alto	Bajo	Alto	Medio	Medio	Medio	Alto	Alto	Alto
7	10	Infantil	62	4	14	Si	Si	7	Medio	Alto	Bajo	Alto	Medio	Medio	Medio	Alto	Medio	Medio
8	7	Infantil	25	0	0	No	No	1	Medio	Alto	Bajo	Medio	Medio	Medio	Medio	Bajo	Medio	Medio
9	11	Infantil	33	21	28	Si	No	4	Medio	Alto	Bajo	Alto	Alto	Medio	Medio	Medio	Alto	Medio
10	6	Primaria	35	30	33	No	No	0	Alto	Alto	Medio	Alto	Alto	Medio	Alto	Alto	Alto	Alto
11	3	Infantil	20	43	59	No	No	6	Medio	Alto	Bajo	Medio	Medio	Alto	Alto	Alto	Medio	Alto
12	19	Infantil	22	45	45	No	Si	8	Medio	Medio	Alto	Alto	Alto	Medio	Medio	Medio	Alto	Medio
13	21	Primaria	42	24	27	No	No	6	Alto	Alto	Medio	Alto	Medio	Medio	Alto	Medio	Alto	Medio
14	18		20	66	59	No	Si	5	Alto	Alto	Medio	Alto	Medio	Alto	Alto	Bajo	Alto	Alto
15	22	Primaria	16	50	55	Si	Si	3	Alto	Alto	Alto	Alto	Alto	Alto	Alto	Alto	Alto	Alto
16	25	Infantil	28	0	16	Si	No	4	Alto	Alto	Alto	Alto	Medio	Alto	Alto	Alto	Alto	Alto
17	35	Primaria	30	69	63	No	No	2	Medio	Medio	Medio	Medio	Medio	Medio	Medio	Medio	Medio	Medio
18	33	Infantil	26	3	17	Si	Si	6	Alto	Alto	Bajo	Alto	Alto	Medio	Alto	Alto	Alto	Alto
19	30	Infantil	28	0	1	No	No	1	Alto	Alto	Medio	Alto	Alto	Alto	Alto	Alto	Alto	Alto
20	17	Infantil	31	29	20	No	Si	5	Alto	Medio	Alto	Alto	Medio	Alto	Medio	Alto	Alto	Alto
21	28	Primaria	26	55	63	Si	No	6	Alto	Alto	Alto	Alto	Alto	Alto	Alto	Alto	Alto	Alto
22	26	Infantil	31	57	6	Si	Si	3	Alto	Alto	Medio	Medio	Medio	Medio	Medio	Bajo	Medio	Medio

Ejercicio:

1. Abre el archivo de datos jamovi *AINSCOW.omv*.
2. ¿La coordinación de los/las tutores/as con las familias del alumnado sordo es la misma que con las familias del alumnado oyente?
3. ¿La coordinación de los/las tutores/as con las familias del alumnado sordo y con los/las profesionales es independiente del ciclo formativo?
4. Analiza las expectativas de los/las tutores/as respecto al alumnado oyente en comparación con las que tienen respecto al alumnado sordo. ¿Qué puedes concluir?
5. ¿Las estrategias docentes que aplican los/las tutores/as al alumnado oyente son las mismas que aplica al alumnado sordo?
6. ¿Las estrategias docentes que aplican los/las tutores/as al alumnado sordo son las mismas en los dos ciclos formativos?
7. ¿Qué puedes decir sobre las adaptaciones socioeducativas y didácticas que realizan los/las tutores/as por el hecho de tener alumnado con sordera en el aula? ¿Estas adaptaciones son las mismas en los dos ciclos formativos?
8. ¿La percepción de los/las tutores/as sobre la cultura y las prácticas inclusivas es la misma en los dos ciclos formativos?
9. ¿El rendimiento auditivo influye en la percepción de los/las tutores/as en relación con la cultura y las prácticas inclusivas?
10. ¿Y los niveles de comprensión oral y de expresión oral?
11. ¿El número de cursos de formación realizados guarda relación con la percepción de los/las tutores/as en relación con la cultura y las prácticas inclusivas?
12. ¿Y el hecho de haber recibido un curso sobre atención a la diversidad o sobre el uso de las TIC?
13. Guarda los resultados de los análisis en el fichero jamovi *AINSCOW_User.omv*.
14. Compara los análisis que has realizado con los del fichero jamovi *AINSCOW_Sol.omv*.

5.4. Desarrollo evolutivo de la conciencia fonológica

Este ejercicio se basa en una matriz de datos simulada a partir de los resultados del siguiente artículo:

de la Calle, A.M., Aguilar, M. y Navarro, J.I. (2016) Desarrollo evolutivo de la conciencia fonológica: ¿Cómo se relaciona con la competencia lectora posterior? *Revista de Investigación en Logopedia*, 1, 22-41. <https://doi.org/10.5209/rlog.58553>

Este estudio pretende comprender la secuencia normal de desarrollo en el mecanismo de adquisición lectora averiguando el grado de dificultad de distintas tareas de conciencia fonológica y su relación con la competencia lectora posterior. La muestra de este estudio está formada por 124 escolares de educación infantil de 4 (n=64) y 5 (n=60) años.

Las **variables registradas**, todas ellas numéricas, son las siguientes:

Nombre	Etiqueta de la variable	Etiqueta de los códigos
Caso	Identificador (se puede obviar la etiqueta)	
Sexo	Sexo	0: Niño; 1: Niña
NivEst	Nivel de estudios máximo de los padres	1: Sin estudios; 2: Primaria; 3: Secundaria; 4: Bachillerato; 5: Universitario (diploma/grado); 6: Universitario (máster); 7: Universitario (doctorado)
NivSE	Nivel socioeconómico de la familia	0: Bajo; 1: Medio
TipoCole	Tipo de colegio	0: Municipal; 1: Concertado
TipProg	Tipo de programa escolar	0: No Integración; 1: Integración
Grupo	Grupo de edad	0: 4 años; 1: 5 años
CF	Conciencia fonológica (puntuación)	
AS	Aislar sílabas (recuento)	
R1	Rima 1 (recuento)	
R2	Rima 2 (recuento)	
CS	Contar sílabas (recuento)	
OS	Omisión de sílabas (recuento)	
CL	Competencia lectora (puntuación)	
CLAlta	Competencia lectora por encima de la mediana del grupo de edad del/de la niño/a	0: No; 1: Sí

	Caso	Sexo	NivEst	NivSE	TipoCole	TipProg	Grupo	CF	AS	R1	R2	CS	OS	CL
1	1	Niña	Sin estudios	Medio	Concertado	No integracion	4 años	40	7	9	7	15	2	39
2	2	Niño	Sin estudios	Medio	Concertado	Integracion	4 años	46	7	5	8	15	5	33
3	3	Niña	Universitari...	Medio	Concertado	No integracion	4 años	46	9	7	10	15	2	42
4	4	Niña	Universitari...	Medio	Municipal	No integracion	4 años	27	5	2	5	10	1	24
5	5	Niño	Secundaria	Medio	Concertado	Integracion	4 años	33	5	3	6	14	0	32
6	6	Niño	Secundaria	Medio	Concertado	No integracion	4 años	25	3	7	8	4	1	21
7	7	Niño	Primaria	Bajo	Concertado	Integracion	4 años	24	3	6	6	10	1	20
8	8	Niña	Universitari...	Medio	Concertado	No integracion	4 años	44	8	5	7	18	4	32
9	9	Niña	Secundaria	Bajo	Municipal	No integracion	4 años	46	10	0	7	14	3	37
10	10	Niño	Secundaria	Bajo	Municipal	No integracion	4 años	25	4	4	6	10	1	23
11	11	Niña	Primaria	Medio	Concertado	No integracion	4 años	19	3	4	5	5	1	32
12	12	Niña	Universitari...	Medio	Concertado	No integracion	4 años	30	6	10	9	7	1	24
13	13	Niña	Secundaria	Bajo	Concertado	Integracion	4 años	27	10	0	3	8	0	33
14	14	Niño	Sin estudios	Medio	Concertado	No integracion	4 años	30	3	7	7	13	3	22
15	15	Niña	Sin estudios	Medio	Concertado	No integracion	4 años	24	4	3	4	12	1	25

Ejercicio:

1. Abre el fichero de datos jamovi *CFCL.omv*.
2. Crea la variable binaria *CF30* a partir de la variable *CF* de modo que la nueva variable contenga 2 categorías de acuerdo con los criterios siguientes:

<i>CF</i>	<i>CF30</i>	
Valor	Valor	Etiqueta
0 a 30	1	$CF \leq 30$
Mayor a 30	0	$CF > 30$

3. Crea la variable *Rimas* que contenga la suma de las puntuaciones de Rima 1 (*R1*) y Rima 2 (*R2*) para los/las niños/as que tengan una puntuación válida en ambas variables.
4. Realiza la descripción de las siguientes variables registradas en el estudio utilizando los índices y gráficos estadísticos más adecuados en cada caso:
 - a. Grupo de edad (*Grupo*).
 - b. Sexo (*Sexo*).
 - c. Nivel socioeconómico (*NivSE*).
 - d. Conciencia fonológica (*CF*).
 - e. Competencia lectora (*CL*).
 - f. *Rimas*.
 - g. Contar sílabas (*CS*).
5. Responde a las siguientes cuestiones acerca de la relación entre variables registradas en el estudio:
 - a. ¿Los grupos de edad (*Grupo*) están equilibrados en cuanto al número de niños y niñas (*Sexo*)?
 - b. ¿Existen diferencias entre los dos grupos de edad (*GRUPO*) en cuanto a la distribución del nivel socioeconómico (*NivSE*)?
 - c. ¿Existen diferencias en la conciencia fonológica (*CF*) entre los dos grupos de edad (*Grupo*)?
 - d. ¿Y en la competencia lectora (*CL*)?
 - e. ¿Los dos grupos definidos con base en el sexo (*Sexo*) presentan la misma proporción de niños/as con valores por encima de la mediana en competencia lectora (*CLAlta*)?
 - f. ¿Los resultados de la prueba de conciencia fonológica (*CF*) están asociados a los obtenidos en las tareas de aislar sílabas (*AS*), rima (*R1* y *R2*), contar sílabas (*CS*) y omisión de sílabas (*OS*)?
 - g. ¿Los resultados obtenidos en competencia lectora (*CL*) se asocian con los obtenidos en esas mismas tareas?
 - h. Analiza las diferencias en competencia lectora (*CL*) entre los grupos definidos con base en el nivel de estudios (*NivEst*).
6. Reproduce todos los resultados posibles que aparecen en las tablas 1 y 4 del artículo ejecutando un solo menú jamovi.
7. Guarda los resultados de los análisis en el fichero jamovi *CFCL_User.omv*.
8. Compara los análisis realizados y tus resultados con los del fichero jamovi *CFCL_Sol.omv*.