

PAULA ELOSUA
MARTIN EGAÑA

PSICOMETRÍA APLICADA

GUÍA PARA EL ANÁLISIS DE DATOS Y
ESCALAS CON JAMOVI



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

Psicometría aplicada

Guía para el análisis de datos y escalas con *jamovi*

CIP. Biblioteca Universitaria

Elosua Oñiden, Paula

Psicometría aplicada [Recurso electrónico]: guía para el análisis de datos y escalas con *jamovi* / Paula Elosua Oñiden. – Datos. – Bilbao : Universidad del País Vasco / Euskal Herriko Unibertsitatea, Argitalpen Zerbitzua = Servicio Editorial, [2020]. – 1 recurso en línea : PDF (124 p.)

Modo de acceso: World Wide Web

ISBN: 978-84-1319-150-8.

1. Psicometría. 2. Tests psicológicos. 3. Jamovi (Programa de ordenador)

(0.034)159.938.3

(0.034)159.9.072

Agradecimientos: Trabajo parcialmente financiado por el Ministerio de Ciencia, Innovación y Universidades; PID2019-103859RB

© Servicio de Publicaciones de la Universidad del País Vasco
Euskal Herriko Unibertsitateko Argitalpen Zerbitzua

ISBN: 978-84-1319-150-8

Índice

Índice de imágenes	5
Índice de tablas	8
1. Test: uso	9
2. jamovi	12
2.1. Entorno R y licencias públicas	12
2.2. <i>jamovi</i> . Sencillez y efectividad	13
2.3. Obtener e instalar <i>jamovi</i>	14
2.4. Pantalla principal.	16
2.5. Copiar y pegar	18
3. Presentación del ejemplo	19
4. Gestión de archivos	21
4.1. Abrir ficheros	22
4.2. Guardar ficheros	24
5. Preparar los datos	25
5.1. Configurar las variables	26
5.2. Recodificar ítems.	29
6. Describir la muestra	34
6.1. Describir una variable	34
6.2. Describir dos variables conjuntas	38
7. Análisis de ítems	41
7.1. Índices basados en la distribución del ítem.	41
7.2. Índice de discriminación	48
7.3. Matriz de correlación entre ítems	51
8. Fiabilidad de la escala	54
8.1. Coeficientes alfa y omega	55
8.2. Error típico de medida.	58
8.2.1. Calcular la puntuación observada	58
8.2.2. Desviación típica de las puntuaciones	60
8.2.3. Estimar el error típico de medida.	62
8.2.4. Intervalo de confianza en torno a la puntuación verdadera	62

9. Dimensionalidad. Análisis factorial exploratorio	65
9.1. Supuestos	68
9.2. Extracción factorial	70
9.3. Rotación factorial	70
9.4. Definición del número de factores	71
9.5. Formato de salida de resultados	72
9.6. Ejemplo práctico	73
9.6.1. Solución unidimensional	73
9.6.2. Solución bifactorial	76
9.7. Análisis de contenido. Definición sustantiva	78
10. Validez externa. Comparación de medias	79
10.1. Comparar dos medias	80
10.2. Comparar más de dos medias	86
11. Validez externa. Relación test-criterio	95
11.1. Correlación. Coeficiente de validez	95
11.2. Predicción: Regresión lineal	98
11.2.1. Diagrama de dispersión. Instalar módulos	101
12. Interpretación normativa	105
12.1. Transformaciones lineales	106
12.2. Transformaciones no lineales	111
12.2.1. Centiles por grupo	114
13. Informe	117
14. Referencias	119

Índice de imágenes

1.1. imagen. Componentes de un test	11
2.1. imagen. <i>jamovi</i> web	14
2.2. imagen. Descargar <i>jamovi</i>	15
2.3. imagen. <i>jamovi</i> ventana principal	16
2.4. imagen. Menús de segundo nivel	17
2.5. imagen. Copiar tablas y gráficos.	18
4.1. imagen. Gestión de archivos	21
4.2. imagen. Abrir ficheros	23
4.3. imagen. Hoja de datos.	23
5. 1. imagen. Características de las variables	26
5. 2. imagen. Definir una variable.	26
5. 3. imagen. Variable edad características (1)	28
5. 4. imagen. Variable edad características (2)	28
5. 5. imagen. Ítems inversos	29
5. 6. imagen. Recodificar ítems (1)	30
5. 7. imagen. Recodificar ítems (2)	31
5. 8. imagen. Recodificar ítems (3)	31
5. 9. imagen. Recodificar ítems (4)	32
5.10. imagen. Recodificar ítems (5)	32
5.11. imagen. Recodificar ítems (6)	33
6.1. imagen. Descriptivos (1)	34
6.2. imagen. Descriptivos (2)	35
6.3. imagen. Descriptivos (3)	35
6.4. imagen. Descriptivos (4)	36
6.5. imagen. Descriptivos (5)	37
6.6. imagen. Descriptivos en función del sexo (1)	38
6.7. imagen. Descriptivos en función del sexo (2)	39
6.8. imagen. Descriptivos en función del sexo (3)	40
7. 1. imagen. Análisis de ítems (1)	42
7. 2. imagen. Análisis de ítems (2)	42
7. 3. imagen. Análisis de ítems (3)	43
7. 4. imagen. Análisis de ítems (4)	44
7. 5. imagen. Análisis de ítems (5)	46
7. 6. imagen. Índice de discriminación (1)	48
7. 7. imagen. Índice de discriminación (2)	49

7. 8. imagen. Índice de discriminación (3)	49
7. 9. imagen. Índice de discriminación (4)	50
7.10. imagen. Matriz de correlaciones inter-ítem (1).	51
7.11. imagen. Matriz de correlaciones inter-ítem (2).	52
7.12. imagen. Matriz de correlaciones inter-ítem (3).	52
8. 1. imagen. Coeficiente de fiabilidad (1)	55
8. 2. imagen. Coeficiente de fiabilidad (2)	56
8. 3. imagen. Coeficiente de fiabilidad (3)	56
8. 4. imagen. Puntuación observada (1)	58
8. 5. imagen. Puntuación observada (2)	59
8. 6. imagen. Puntuación observada (3)	59
8. 7. imagen. Puntuación observada (4)	60
8. 8. imagen. Desviación típica de la puntuación observada (1).	60
8. 9. imagen. Desviación típica de la puntuación observada (2).	61
8.10. imagen. Desviación típica de la puntuación observada (3).	61
8.11. imagen. Intervalo de confianza	63
8.12. imagen. Intervalo de confianza puntuación verdadera (1)	63
8.13. imagen. Intervalo de confianza puntuación verdadera (2)	64
8.14. imagen. Intervalo de confianza puntuación verdadera (3)	64
9. 1. imagen. Análisis factorial (1)	66
9. 2. imagen. Análisis factorial (2)	66
9. 3. imagen. Análisis factorial (3)	67
9. 4. imagen. Análisis factorial (4)	68
9. 5. imagen. Análisis factorial (5)	69
9. 6. imagen. Análisis factorial (6)	70
9. 7. imagen. Análisis factorial (7)	71
9. 8. imagen. Análisis factorial (8)	72
9. 9. imagen. Análisis factorial (9)	73
9.10. imagen. Análisis factorial (10)	74
9.11. imagen. Análisis factorial (11)	76
9.12. imagen. Estructuras factoriales	78
10. 1. imagen. Satisfacción corporal en función del sexo.	81
10. 2. imagen. t.test (1)	82
10. 3. imagen. t-test (2)	82
10. 4. imagen. t-test (3)	83
10. 5. imagen. t-test (3)	84
10. 6. imagen. t-test (4)	85
10. 7. imagen. Diferentes variabilidades intragrupo.	87
10. 8. imagen. ANOVA (1).	88
10. 9. imagen. ANOVA (2).	88
10.10. imagen. ANOVA (3).	89
10.11. imagen. ANOVA (4).	89
10.12. imagen. ANOVA (5).	90
10.13. imagen. ANOVA (6).	91
10.14. imagen. ANOVA (7).	92
10.15. imagen. Modelo completo ANOVA	93
11. 1. imagen. Matriz de correlaciones (1)	96
11. 2. imagen. Matriz de correlaciones (2)	97
11. 3. imagen. Regresión lineal (1).	99
11. 4. imagen. Regresión lineal (2).	100
11. 5. imagen. Instalar módulo (1)	101

ÍNDICE DE IMÁGENES

11. 6. imagen. Instalar módulo (2)	101
11. 7. imagen. Instalar módulo (3)	102
11. 8. imagen. Instalar módulo (4)	102
11. 9. imagen. Gráfico de dispersión	103
11.10. imagen. Gráfico de dispersión por grupo	104
12. 1. imagen. Distribución normal	107
12. 2. imagen. Puntuaciones típicas (1)	107
12. 3. imagen. Puntuaciones típicas (2)	108
12. 4. imagen. Puntuaciones típicas (3)	108
12. 5. imagen. Puntuaciones típicas (4)	109
12. 6. imagen. Puntuaciones T (1)	109
12. 7. imagen. Puntuaciones T (2)	110
12. 8. imagen. Puntuaciones T (3)	110
12. 9. imagen. Estatinos	112
12.10. imagen. Puntuaciones centiles (1).	112
12.11. imagen. Puntuaciones centiles (2).	113
12.12. imagen. Puntuaciones centiles (3).	113
12.13. imagen. Puntuaciones centiles (4).	114
12.14. imagen. Informe. Ejemplo	115

Índice de tablas

2.1. tabla. Menú principal	17
4.1. tabla. Formatos de archivo reconocidos por <i>jamovi</i>	22
4.2. tabla. Exportar ficheros	24
5.1. tabla. Variables: naturaleza	25
5.2. tabla. Variables: <i>jamovi</i>	27
5.3. tabla. Variables: tipo	27
5.4. tabla. Reglas de recodificación	30
7.1. tabla. Niveles de discriminación	48
9.1. tabla. Métodos de extracción factorial	70
9.2. tabla. Rotaciones factoriales	70
9.3. tabla. Análisis factorial. Información adicional	72
10.1. tabla. Pruebas estadísticas comparación de medias	84
10.2. tabla. Interpretación de la <i>d</i> de Cohen	86
11.1. tabla. Coeficientes de correlación	97
12.1. tabla. Escalas transformadas	106
12.2. tabla. Percentiles	111

1

Test: uso

Desde la aparición de la escala de Binet-Simon (1905) para la medición de aptitudes intelectuales superiores, la disponibilidad y el uso de los test se han extendido a todos los ámbitos de actuación de la psicología (sanitario, educativo, social, forense, organizacional, deportivo, clínico, investigador...). Se utilizan como herramientas de apoyo en el quehacer diario, y asisten al profesional en la toma de decisiones; decisiones que afectan en mayor o menor grado a personas o a instituciones. Un diagnóstico clínico, una evaluación educativa, el diseño y evaluación de un programa de intervención, la selección de personal o un peritaje judicial entre otros, demandan la recogida de información objetiva. La utilización de test de probada calidad proporciona un modo fiable, válido y eficiente para esta tarea.

Un test es básicamente un instrumento de medición estandarizado que sirve para recoger información sobre una muestra de conducta (American Educational Research Association, American Psychological Association, y National Council on Measurement in Education, 2014). Los desarrollos técnicos y psicométricos de las últimas décadas en el área de los test hacen difícil una clasificación exhaustiva de los tipos de test disponibles (Olea, Abad y Barrada, 2010); entre los criterios de clasificación podrían utilizarse : (a) campo de estudio (personalidad, inteligencia, neuropsicología, intereses vocacionales, rendimiento, competencias, aptitudes, actitudes...); (b) forma de administración (individual/ colectivo, secuencial/adaptativa, online/offline); (c) soporte (papel, ordenador, móvil, internet); (d) información requerida (autoinformes, ejecución, observación); (e) tipo de interpretación (normativa, criterial, ipsativa), y/o (f) tipo de respuesta (ejecución típica, ejecución máxima).

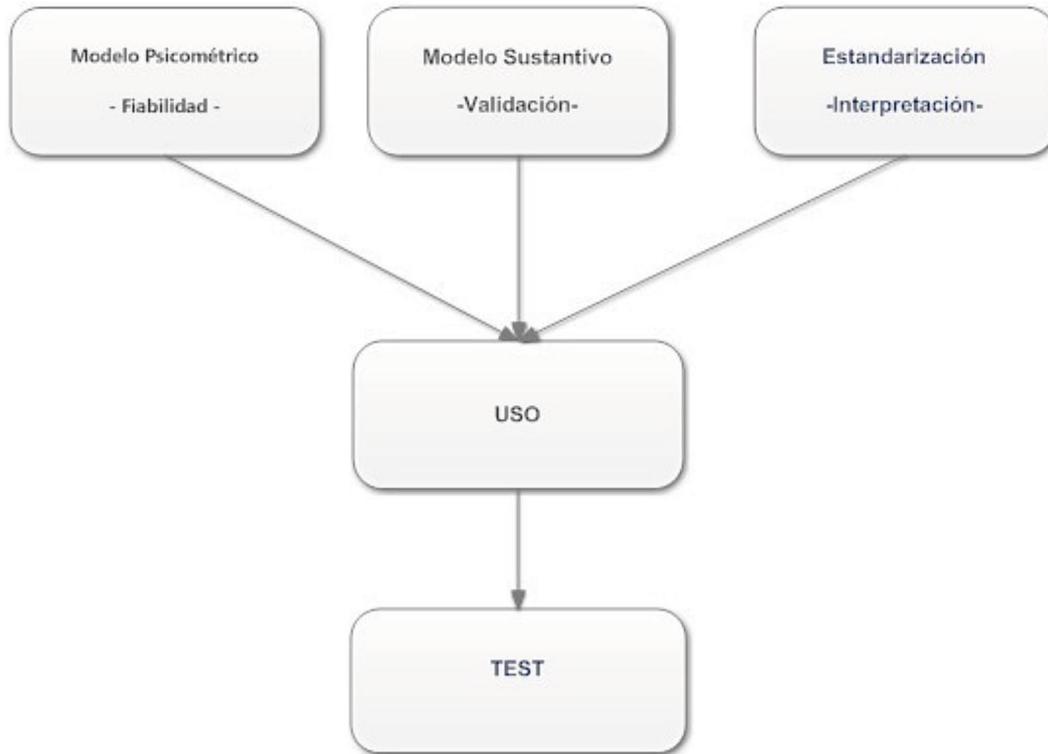
El objetivo de los test es obtener información métrica que junto con el apoyo de otros recursos puede utilizarse en una situación de evaluación. Constituyen el aspecto más conocido y de mayor impacto social relacionado con la investigación psicométrica, y psicológica. Sean conocidos como cuestionarios, exámenes, escalas, instrumentos, autoinformes o medidas, los test comparten exigencias formales cuyo cumplimiento los facultará para su uso en la práctica profesional. Alcanzar el propósito con el que fueron diseñados depende de que se cumplan dos requisitos: (a) que su construcción siga principios psicométricos que garanticen su calidad técnica (Irwin, 2018; Muñiz y Fonseca-Pedrero, 2019; Wilson, 2005), y (b) que sean utilizados de acuerdo a criterios que permitan salvaguardar ésta. Sólo un uso apropiado garantizará la validez de las inferencias derivadas de las puntuaciones.

Tras la administración de un test se obtienen una o varias puntuaciones que se consideran representaciones observadas de la variable o variables latentes que el psicólogo está interesado en medir. El resultado en forma numérica, y conocido como puntuación observada o puntuación empírica posee características distintivas de la medición psicológica: (a) no es una medida directa del rasgo evaluado; (b) no es un valor numérico interpretable en una escala de razón como lo sería la medida del peso de una persona, y (c) su interpretación y valoración están sujetas a los cuatro principios básicos por los que se rige la medición en psicología: fiabilidad, validez, estandarización y uso.

La inferencia psicológica, es decir la interpretación de las puntuaciones o su uso con fines métricos y/o evaluativos se apoya en la consideración y conjunción de estos *cuatro componentes*:

- El componente de *fiabilidad* se relaciona directamente con el error aleatorio de medida. Ninguna medida, en ningún área de conocimiento científico, está libre de error, y la estimación de su efecto sobre la puntuación es imprescindible para su correcta interpretación. El modelo psicométrico más utilizado para la estimación de la fiabilidad es la Teoría Clásica de Test según el cual la puntuación observada X es una combinación lineal de la puntuación verdadera (V) y el error aleatorio de medida (E ; $X = V + E$) (ver Abad, Olea, Ponsoda y García, 2011; Elosua, 2011; Muñiz, 2018). Junto a ella, la Teoría de Respuesta al Ítem (TRI) postula que la probabilidad de respuesta correcta a un ítem es una función de las características del ítem y del nivel de la persona en la variable medida (de Ayala, 2009; Hambleton y Swaminathan, 1985; Lord, 1980).
- El segundo componente a considerar en la interpretación de una puntuación es su *validez*. La validación, se relaciona con el modelo sustantivo, teoría psicológica o base racional utilizada en la construcción del test. Un test es algo más que un procedimiento estandarizado para la obtención de valores escalares; es un instrumento de medida, y como tal tiene que justificarse argumentando que la puntuación es un indicador del constructo o variable no observada que se desea medir.
- El tercer componente de un test hace referencia a su carácter de medida estandarizada. La *estandarización* asegura que tanto las instrucciones, como la administración, corrección e interpretación se realizan siguiendo pautas de actuación normalizadas.
- Sin embargo, estos tres componentes, fiabilidad, validez y estandarización, carecerían de sentido si el uso del test no respetara los objetivos, la finalidad, el contexto de aplicación o las características definitorias de la población a la que este va dirigido. Únicamente un uso correcto (International Test Commission, 2001) puede salvaguardar sus propiedades psicométricas. Hoy somos más conscientes que nunca de las consecuencias que emanan del uso incorrecto de los test (Messick, 1995).

Todo test se construye sobre esos cuatro pilares. Los modelos psicométricos se encargarán del análisis formal de las puntuaciones. Los estudios de validación tienen como cometido su análisis sustantivo. El proceso de estandarización garantizará en la medida de lo posible, que las diferencias encontradas no son debidas a una aplicación/corrección incorrecta del test, y finalmente un uso correcto preservará las propiedades psicométricas del test.



1.1. IMAGEN

Componentes de un test

En este marco definido por los requisitos de calidad psicométrica y uso correcto, el objetivo de esta guía formativa es ofrecer al estudiante y profesional herramientas que le permitan desarrollar competencias para el análisis psicométrico de datos. El propósito son la autonomía y la solvencia en el análisis; aspectos ambos que incidirán directamente en un mejor uso de los test. Se trata de un documento reducido y eminentemente práctico que viene acompañado de un fichero de datos, que permitirá al lector replicar cada uno de los procedimientos descritos.

2

jamovi

En el ámbito de las ciencias sociales se dispone de una variada gama de programas para el análisis de datos, que bien bajo licencia privativa o bajo licencia libre cumplen las exigencias del usuario, independientemente del tipo de explotación o de la sencillez o complejidad de los modelos que quiera estimar. Entre los primeros, se podrían citar programas tan conocidos como SPSS, SAS, STATISTICA, Systat, Stata o GenStat. Desde sus primeras versiones hasta las actuales, estos programas han ido incorporando modelos de análisis cada vez más complejos a la par que han ido mejorando sus interfaces gráficas (GUI, *Graphical User Interfaces*). El desarrollo de los menús contextuales basados en la gestión de ventanas los han convertido en programas de fácil manejo, y al mismo tiempo ha permitido la socialización del análisis de datos con relación al uso de modelos de análisis que ahora se hacen accesibles al usuario no iniciado.

Sin embargo, el carácter comercial y la licencia privativa de todos ellos son una limitación a su uso. Desde el punto de vista del profesor que asume la responsabilidad de ofrecer y facilitar al alumno las bases para una formación continua que no finaliza con el abandono de la universidad, la accesibilidad es importante. La disponibilidad de herramientas fiables y sólidas a coste 0 ofrecen al alumno y al profesional una autonomía de trabajo que repercute en el buen desarrollo profesional.

2.1. Entorno R y licencias públicas

R (Ihaka y Gentleman, 1996; R Core Team, 2018) es un entorno de programación y análisis estadístico y gráfico, derivado del lenguaje de programación S (Becker, Chambers y Wilks, 1988). Fue desarrollado por Ross Ihaka y Robert Gentleman del Departamento de Estadística de la Universidad de Auckland (Nueva Zelanda). El nombre R proviene de las iniciales de sus autores (R), y de la fonética de R «our» —nuestro— que lo enlaza con la filosofía del software libre. La primera versión de R se difundió rápidamente y su expansión es hoy imparable (Elosua, 2009). R crece con las aportaciones no-lucrativas de investigadores provenientes de prácticamente todas las ramas del conocimiento. Las actualizaciones del entorno corren a cargo del R Development Core Team, grupo constituido en el año 1997.

En R confluyen características que lo hacen único; es libre, de código abierto, dispone de versiones para distintas plataformas (Microsoft Windows, Linux/UNIX o Macintosh) y está a la van-

guardia de los más avanzados modelos estadísticos. R posee una estructura versátil, adaptable a las necesidades del usuario básico, mediano o avanzado. Su capacidad para el análisis y la construcción de gráficos son excelentes.

Parte de la vasta información disponible sobre R es accesible a través de la web CRAN (*Comprehensive R Archive Network*; <http://cran.r-project.org/>), sitio oficial de R. Es la página base del proyecto R, desde la cual se puede descargar la última versión del programa, consultar manuales, obtener ayuda, y, en definitiva, estar al corriente de los movimientos en este entorno de trabajo.

R, en tanto en cuanto software libre, se inscribe dentro del proyecto GNU *General Public License* (Licencia Pública General, GNU). Se trata de una licencia creada por *Free Software Foundation* (Fundación para el software libre) organización fundada por Richard Matthew Stallman (rms) en el año 1985. El principal propósito de la licencia GNU es declarar la libertad del uso, modificación y distribución del software, y protegerlo de intentos de privatización que puedan restringir su uso (el contenido de la licencia puede consultarse en el sitio <http://www.gnu.org/copyleft/gpl.html>). Dentro de esta licencia se distribuyen programas como la suite ofimática *OpenOffice*, el navegador Mozilla, los artículos de wikipedia, el sistema operativo GNU/Linux, o el editor de textos Emacs.

A pesar de las bondades de R, tal vez su mayor desventaja es la curva de aprendizaje asociada al trabajo con códigos. Para solventar esta situación se han construido varias GUIs —interfaces gráficas—, que permiten el uso y acceso a este entorno por medio de diálogos y menús. Entre las GUIs más utilizadas destacaríamos BlueSky Statistics, RKWard (Rödiger, Friedrichsmeier, Kapat, y Michalke, 2012), R Commander (Elosua, 2011; Fox, 2005), Deducer (Fellows, 2012), JASP (2019) o *jamovi* (2019).

Entre todas ellas dedicamos esta guía práctica a *jamovi*. Es una de las últimas en aparecer en el panorama del análisis de datos; la facilidad de su uso, ligado a una estética atractiva, buen rendimiento y amplia cobertura de los modelos estadísticos más utilizados en psicología y en las ciencias sociales, la convierten en una alternativa potente para el profesional y estudiante que desea analizar datos sin dedicar tiempo al aprendizaje de operativas más o menos complejas.

2.2. *jamovi*. Sencillez y efectividad

jamovi es una herramienta de reciente aparición en el panorama del análisis de datos; la primera versión se lanzó en 2017. Con una interfaz elegante, sencilla e intuitiva el usuario no se percata de que está trabajando sobre R. Los análisis se ejecutan en tiempo real, y cualquier cambio en los datos tiene un efecto inmediato sobre los resultados. Las salidas se ofrecen en el formato APA, y estas pueden trasladarse a cualquier editor con la conocida fórmula de copiar + pegar (CtrlC + CtrlV). Una de las características distintivas de *jamovi* es que la opción de guardar genera un único fichero que contiene los datos y los análisis efectuados sobre ellos; al abrirlo se recuperan las operaciones y la secuencia de análisis efectuadas.

jamovi incorpora los modelos incluidos en cualquier curso de pregrado de análisis de datos, y los refuerza con análisis de mediación/moderación, modelos mixtos y análisis bayesianos. Se presenta como un software impulsado por la comunidad científica (*community driven*), en la cual se fo-

menta el desarrollo de funciones que se pondrán a disposición del usuario. Estas funciones se añaden como módulos (ver 11.2.1).

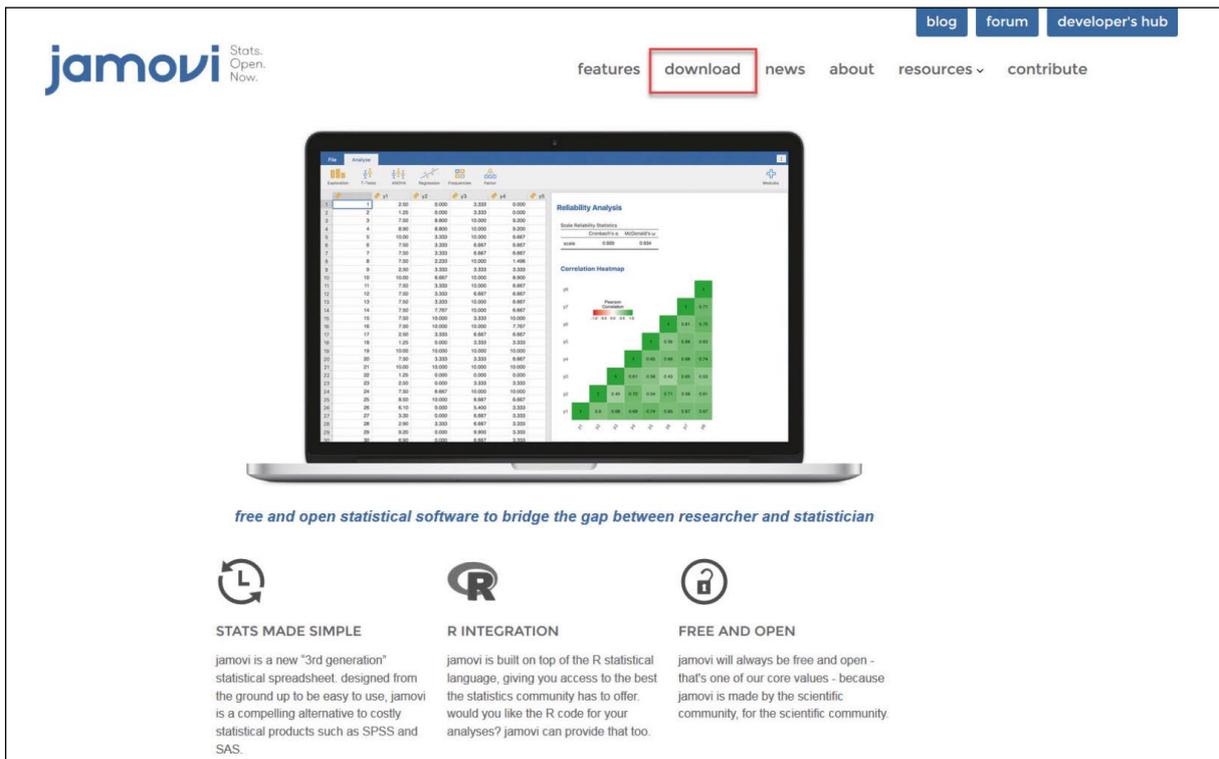
La gestión de ficheros reconoce un amplio espectro de extensiones entre las que se encuentran SPSS, Stata, SAS, R, Excel o Libre Office. La manipulación de datos incluye posibilidades para añadir/modificar/borrar variables o establecer filtros (la versión actual no permite ordenar o transponer los datos).

Para las situaciones en que *jamovi* no responde a las necesidades del investigador se puede recurrir a la utilización de códigos a través de la instalación de módulo *jmv*.

2.3. Obtener e instalar *jamovi*

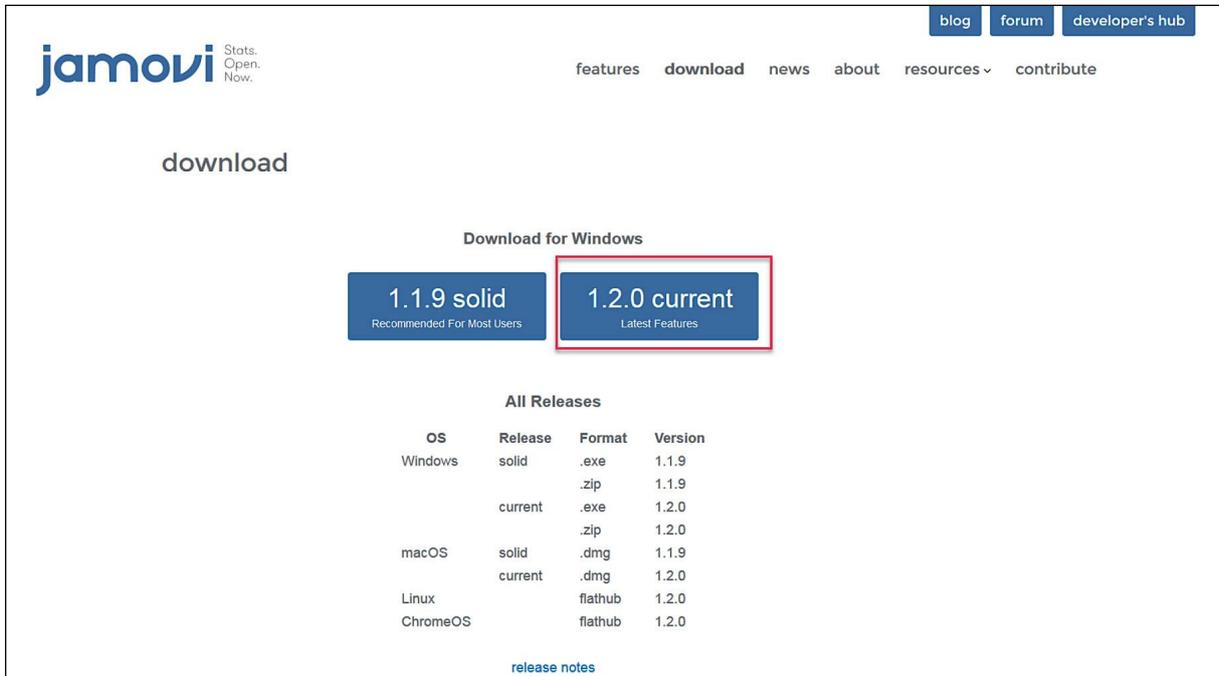
La instalación de *jamovi* requiere una única fase. Para obtener e instalar este software seguiremos las siguientes pautas:

1. Acceder a la web oficial de *jamovi* (<https://www.jamovi.org/>).



2.1. IMAGEN
jamovi web

2. Seleccionar la plataforma adecuada (Linux, MacOSX o Windows) y descargar el programa. El proceso de instalación es automático; a diferencia de otros programas, únicamente requiere un paso, y una vez ejecutado el archivo de instalación, *jamovi* junto a todas sus dependencias quedarán instaladas en nuestro ordenador.

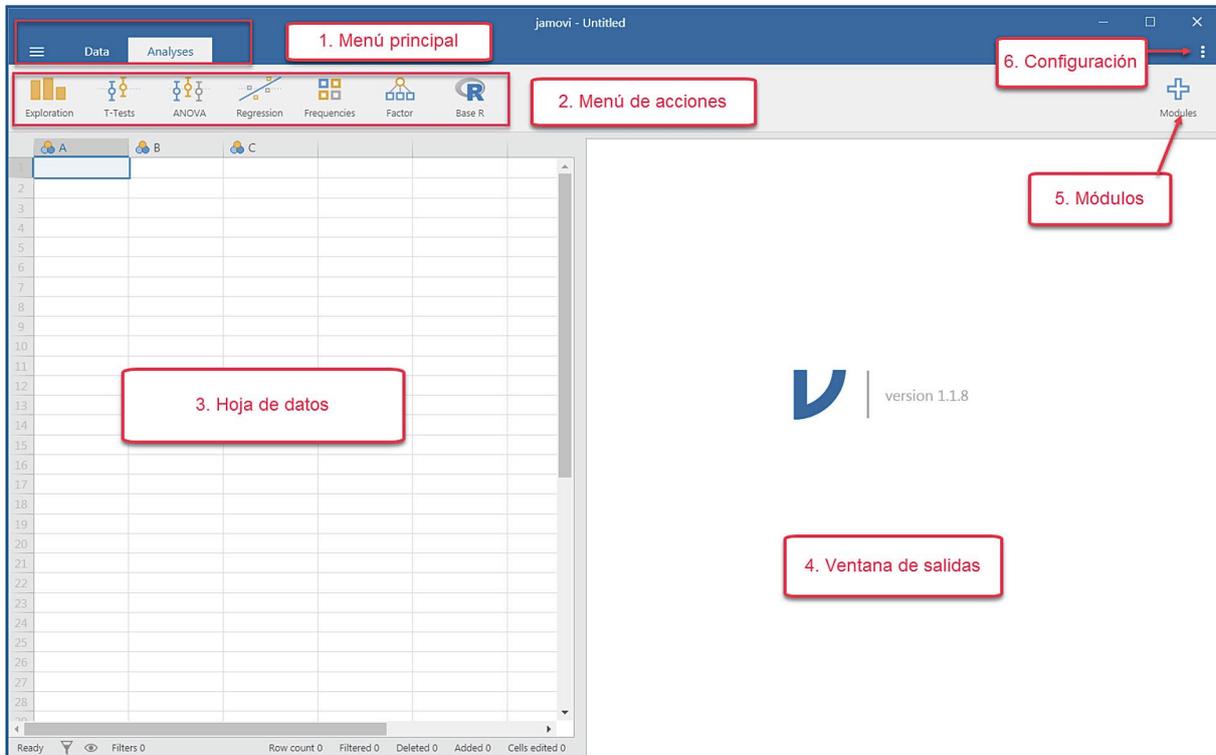


2.2. IMAGEN

Descargar *jamovi*

2.4. Pantalla principal

Cuando se ejecuta *jamovi* el usuario se encuentra con esta pantalla, a la que se denomina interfaz de usuario. En ella podemos distinguir varias secciones:



2.3. IMAGEN

jamovi ventana principal

1. *Menú-principal* Al igual que cualquier software diseñado para el análisis de datos, en *jamovi* están claramente diferenciadas tres secciones: ficheros, datos y análisis.

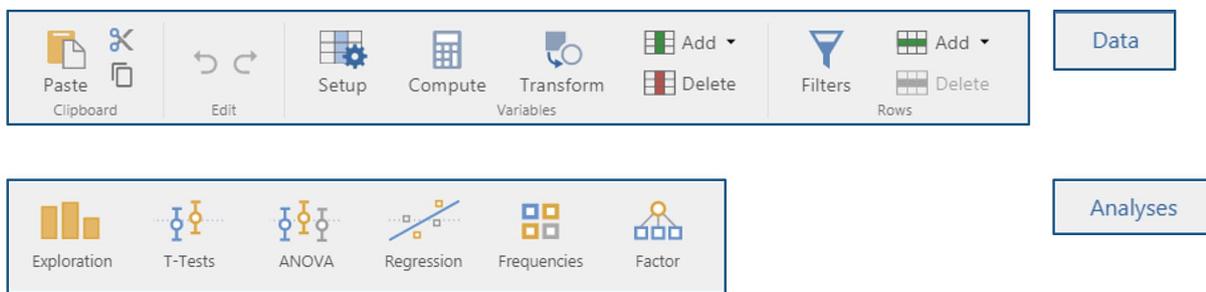
- Ficheros: permite la gestión de ficheros con acciones destinadas a abrir, importar, guardar y exportar. Todas ellas están incluidas dentro del icono hamburguesa (\equiv).
- Datos: la opción *Data* incluye acciones para el manejo de datos, como crear, borrar o modificar variables, añadir y borrar casos, y definir filtros.
- Análisis: esta opción incluye los modelos para el análisis de datos.

2.1. tabla

Menú principal

Menú principal	
	Icono hamburguesa. Permite la gestión de ficheros: abrir, importar, guardar, exportar
	Gestión de datos: añadir variables, modificar variables, calcular variables, borrar variables.
	Modelos estadísticos y de análisis.

2. *Menú de acciones.* Para manipular los datos la opción es *Data*. Si se desea analizarlos habrá que seleccionar *Analyses*.



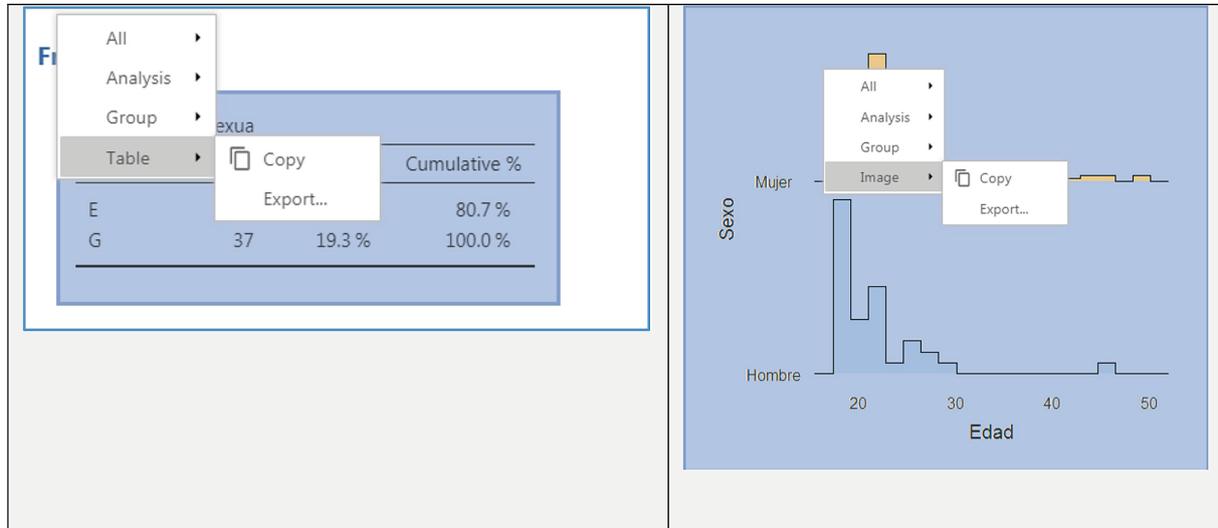
2.4. IMAGEN

Menús de segundo nivel

3. *Hoja de datos.* *jamovi* tiene la apariencia de una hoja de datos, en la cual las filas están ocupadas por casos (sujetos) y las columnas contienen información sobre las variables.
4. *Ventana de resultados o salidas.* En ella irán apareciendo los resultados de los análisis efectuados.
Entre el área de la hoja de datos y la ventana de salidas existe una barra que el usuario puede mover para modificar el área de ambas secciones.
5. *Módulos.* Las opciones gráficas y de análisis incluidas en *jamovi* pueden ampliarse con la instalación de funciones integradas en módulos específicos, que se gestionan través de *Modules* (ver 11.2.1).
6. *Opciones de configuración.* *jamovi* ofrece varias opciones para modificar los colores de los gráficos, el tamaño de la fuente o el número de dígitos; todas ellas se encuentran en la elipse vertical (:).

2.5. Copiar y pegar

Las tablas y gráficos generados por *jamovi* pueden copiarse y pegarse en editores de texto o gráficos, utilizando las opciones de copiar o exportar, accesibles —como en cualquier otro software— con la presión del botón derecho del ratón.



2.5. IMAGEN

Copiar tablas y gráficos

3

Presentación del ejemplo

La utilización de *jamovi* en el estudio de las características de un inventario, test o escala se presenta con el apoyo de los análisis ejecutados sobre un fichero de datos que contiene las respuestas a una adaptación de la escala de satisfacción corporal perteneciente a *Eating Disorder Inventory-3* (Garner, 2004; Elosua, López-Jáuregui y Sánchez-Sánchez, 2010). La escala satisfacción corporal está compuesta por 10 ítems de respuesta graduada con 6 categorías de respuesta (Siempre, Casi siempre, A menudo, A veces, Pocas veces, Nunca) que se codifican respectivamente con valores 1-6. La dirección de la escala está en línea con el constructo satisfacción corporal; a mayor puntuación mayor satisfacción.

El contenido de los ítems ficticios de la escala satisfacción corporal es el siguiente:

1. Pienso que mis manos son demasiado grandes.
2. Creo que mi nariz es demasiado larga.
3. Creo que el tamaño de mis manos está bien.
4. Me siento satisfecho con mi imagen.
5. Me gusta la forma de mi cara.
6. Creo que mis orejas son demasiado anchas.
7. Me siento hinchado después de comer.
8. Creo que mi nariz tiene el tamaño adecuado.
9. Creo que mi cara es demasiado ancha.
10. Creo que mis orejas tienen el tamaño adecuado.

El fichero incluye variables que añaden información sobre cada participante en la investigación:

1. Edad: variable numérica que representa la edad en años.
2. Categorías: grupos en función de la edad. Las categorías adolescente, joven y adulto representan grupos de participantes con edades entre los 18 y los 21 años, 22 y 26 años, y mayores de 26 años.
3. Sexo: variable con dos niveles Hombre/Mujer.
4. Peso: peso en kilogramos de cada participante.
5. Altura: altura en centímetros de cada participante.
6. Ítems1-Ítem10: ítems pertenecientes a la escala satisfacción corporal (SaCor).
7. ST: puntuación obtenida en un test de siluetas (variable criterio).

El análisis de datos sigue este esquema general, que se presentará desde una perspectiva práctica (si bien es posible y aconsejable en algunas situaciones adaptar esta secuencia). Las pautas que ofrecemos son habituales en un análisis de escalas, y las intentaremos exponer desde una perspectiva práctica.

1. Lectura del fichero de datos.
2. Preparación de las variables.
3. Descripción de la muestra.
4. Análisis de ítems.
5. Fiabilidad.
6. Estructura interna. Análisis factorial exploratorio.
7. Validez externa. Comparación de grupos, predicción de criterios.
8. Normas de interpretación.

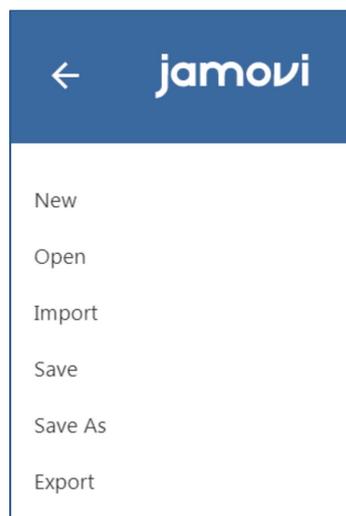
4

Gestión de archivos

La gestión de archivos se ocupa de las acciones sobre los ficheros de datos que son fundamentales para iniciar el análisis, gestionar las salidas, guardar las pautas de análisis, y en definitiva mejorar la replicabilidad de la investigación:

- Lectura o importación de los datos a analizar.
- Grabación o exportación de los datos una vez éstos han sido manipulados.
- Grabación de las salidas de resultados.
- Grabación de las secuencias de análisis efectuados.

Todas las acciones relacionadas con la gestión de ficheros —abrir, importar, guardar— se esconden en *jamovi* dentro del icono hamburguesa (≡).



4.1. IMAGEN

Gestión de archivos

- *New*: nuevo. Crear un nuevo fichero de datos.
- *Open*: abrir. Abrir un fichero de datos existente.
- *Import*: importar. Abrir un fichero de datos sobre otro previamente abierto.
- *Save*: guardar. Guardar un fichero de datos, análisis y salidas.
- *Save As*. Guardar un fichero con un nombre diferente al original.
- *Export*: exportar. Guardar un fichero en un formato diferente al utilizado por *jamovi*.

Cada software utiliza un formato específico en la gestión de archivos; *jamovi* puede trabajar con los siguientes formatos:

4.1. tabla

Formatos de archivo reconocidos por *jamovi*

Origen	Extensión
<i>jamovi</i> (datos y análisis)	.omv
<i>jamovi</i> (plantilla)	.omt
Archivo separado por comas	.csv .txt
SPSS	.sav .zsav .por
R	.RData .RDS
Stata	.dta
SAS	.xpt .sas7bdat
JASP	.jasp
Excel	.xlsl
Libre Office	.ods

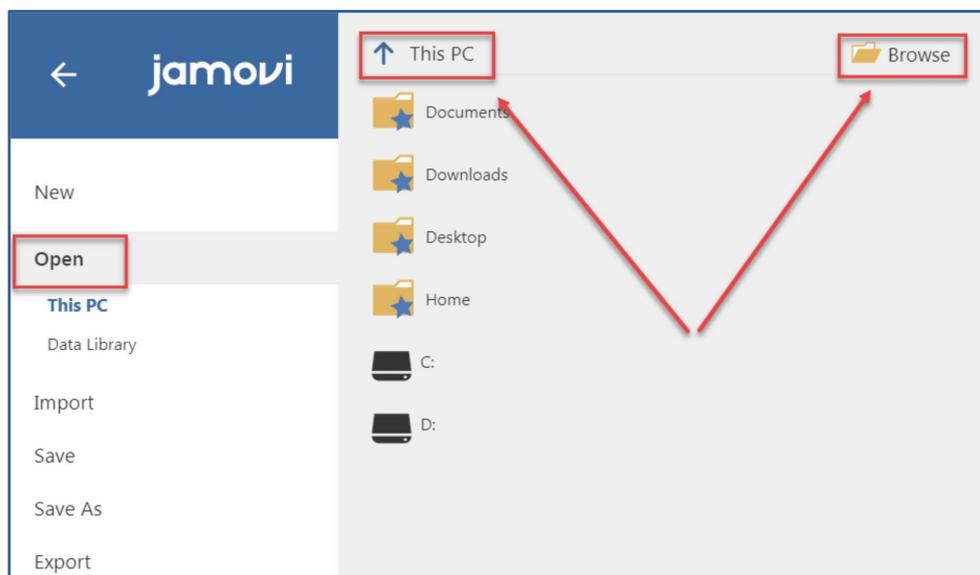
4.1. Abrir ficheros

jamovi ofrece dos opciones para la lectura de datos: abrir e importar. Para ambas acciones reconoce el mismo tipo de formatos, pero entre ambas existen diferencias importantes. Para *jamovi* la opción de importar datos significa trasladar esos datos sobre un archivo que está ya abierto. Esa acción origina lo siguiente:

- Se borran las filas del archivo que estaba abierto.
- Se trasladan los nuevos valores sobre las columnas previamente definidas.
- Si existen variables diferentes a las del archivo abierto se añaden columnas a la hoja de datos.
- Las variables del archivo original que no tienen correspondencia en el nuevo archivo quedan en blanco.

Si la hoja de trabajo está vacía y se opta por la opción *import*, *jamovi* añade tres columnas a la hoja de datos porque por defecto *jamovi* contiene 3 columnas que se mantendrán tras la importación de nuevos datos.

Las opciones de importar y abrir dan acceso a una ventana en la cual se podrá seleccionar el archivo a leer.



4.2. IMAGEN

Abrir ficheros

El archivo abierto tiene la apariencia de una hoja de datos donde cada variable es una columna.

	Edad	Categorías	Sexo	Peso	Altura	item1
1	29	Adulto	G	84.0	188	5
2	21	Adolescente	G	117.0	182	1
3	46	Adulto	G	80.0	173	4
4	24	Joven	G	77.0	186	2
5	19	Adolescente	G	80.0	187	6
6	18	Adolescente	G	74.0	185	2
7	21	Adolescente	G	78.0	183	6
8	25	Joven	G	82.6	174	1
9	20	Adolescente	G	70.0	182	4
10	18	Adolescente	G	85.0	181	6
11	21	Adolescente	G	76.0	181	6
12	19	Adolescente	E	82.0	166	3
13	19	Adolescente	G	75.0	170	6
14	18	Adolescente	G	65.0	180	6
15	20	Adolescente	G	75.0	180	2

4.3. IMAGEN

Hoja de datos

4.2. Guardar ficheros

Las modificaciones realizadas sobre la hoja de datos pueden guardarse en *jamovi* con las opciones *Save*, *Save As* y *Export*, accesibles a través del icono hamburguesa.

Las opciones *Save* y *Save as*, permiten almacenar tanto las modificaciones sobre la hoja de datos, como los análisis efectuados sobre los datos en el formato específico de *jamovi*, *omv*. La apertura de un fichero *.omv* ocasiona la recuperación automática de las operaciones realizadas sobre los datos, de la secuencia de modelos estimados, y de los resultados obtenidos. *Save* almacena el fichero con el mismo nombre con el que se abrió. La opción *Save as* permite modificar el nombre del fichero, y si se desea guardar el fichero en algún otro formato la opción a utilizar es *Export*.

NOTA. Para recuperar las modificaciones realizadas sobre los datos, así como los análisis efectuados la extensión a utilizar es *.omv*.

4.2. tabla

Exportar ficheros

Exportar	Extensión
Portable Document Format	.pdf
Web	.html .html
<i>jamovi</i> patrón	.omt
Archivos separados por comas	.csv .txt
SPSS	.sav .por
R	.RData .RDS
Stata	.dta
SAS	.xpt .sas7bdat

5

Preparar los datos

La hoja de trabajo de *jamovi* está organizada en columnas, cada una de las cuales se refiere a una variable. El tipo de análisis a realizar está ligado al tipo de variables. En general estas se dividen en variables cualitativas y variables cuantitativas.

— Variables cualitativas: son variables no numéricas.

- Variables nominales: aquellas entre las que únicamente se pueden establecer relaciones de identidad ($a = b$; $a \neq b$).
Por ejemplo: religión, idioma, origen...
- Variables ordinales: además de las relaciones de igualdad/desigualdad, es posible valorar su orden ($a \leq b$; $a \geq b$).
Por ejemplo: insuficiente < suficiente < bien < muy bien; pequeño < mediano < grande; nunca < a veces < periódicamente < siempre.

— Variables cuantitativas: son variables numéricas.

- Variables discretas: aquellas que pueden tomar algunos valores dentro de un mínimo conjunto numerable.
Por ejemplo: número de hijos, número de dedos, número de árboles en un parque...
- Variables continuas: pueden tomar un valor fijo dentro de un intervalo determinado; una variable continua toma valores a lo largo de un continuo, esto es, en todo un intervalo de valores.
Por ejemplo: medidas como la altura, peso, distancia (1; 1,1;1,11;1,111).

5.1. tabla

Variables: naturaleza

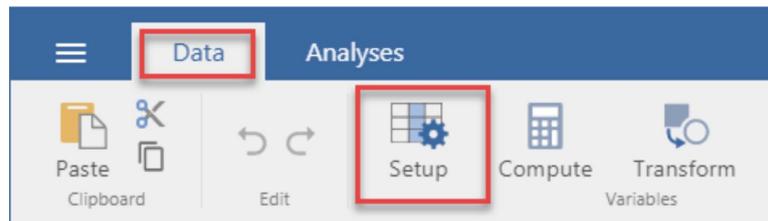
Naturaleza	Tipo	Ejemplo
Cualitativas	Nominales	Origen
	Ordinales	Pequeño < mediano < grande
Cuantitativas	Discretas	Número de árboles
	Continuas	Altura

5.1. Configurar las variables

Antes de iniciar el proceso de análisis de datos es fundamental definir correctamente las variables. *jamovi* asigna a cada variable leída o importada características referidas a su nombre, naturaleza o categorías (niveles), que hay que comprobar y en su caso, mantener o modificar.

Para acceder a las características de una variable *jamovi* ofrece dos opciones:

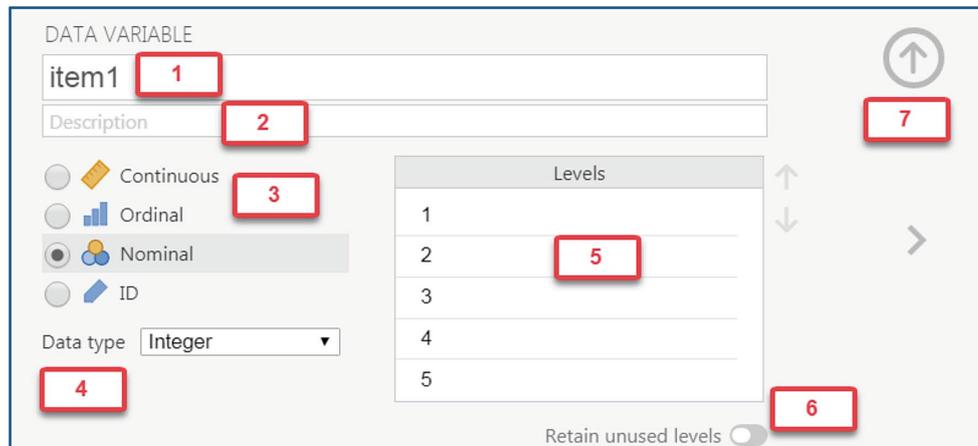
1. Situarse en la cabecera de la variable de interés y presionar dos veces el botón izquierdo del ratón.
2. Situarse en cualquier fila de la variable, y presionar *Setup* dentro de la opción *Data* del menú principal.



5.1. IMAGEN

Características de las variables

Por ejemplo para la variable *item1*, *jamovi* muestra la siguiente pantalla:



5.2. IMAGEN

Definir una variable

1. *Nombre de la variable*. Es posible mantener el nombre original o modificarlo. Para ello bastará con escribir el nuevo nombre en el espacio habilitado para ello.
2. *Descripción*. Se puede añadir un contenido descriptivo de la variable, que si bien no es necesario, puede ser aconsejable.
3. *Naturaleza de la variable*. En la hoja de datos de *jamovi* en la primera fila de cada columna aparece un símbolo que define la naturaleza de la variable. *jamovi* reconoce cuatro tipos de variables:

5.2. tablaVariables: *jamovi*

Naturaleza de las variables	
Continua (<i>Continuous</i>)	<input checked="" type="radio"/>  Continuous
Ordinal (<i>Ordinal</i>)	<input type="radio"/>  Ordinal
Nominal (<i>Nominal</i>)	<input type="radio"/>  Nominal
ID	<input type="radio"/>  ID

— *Continuous*. Variables de naturaleza numérica.

— *Nominal* y *ordinal*. Variables cualitativas.

— ID es un tipo de variables utilizado por *jamovi* como variable indicador; este tipo de variable no se analiza.

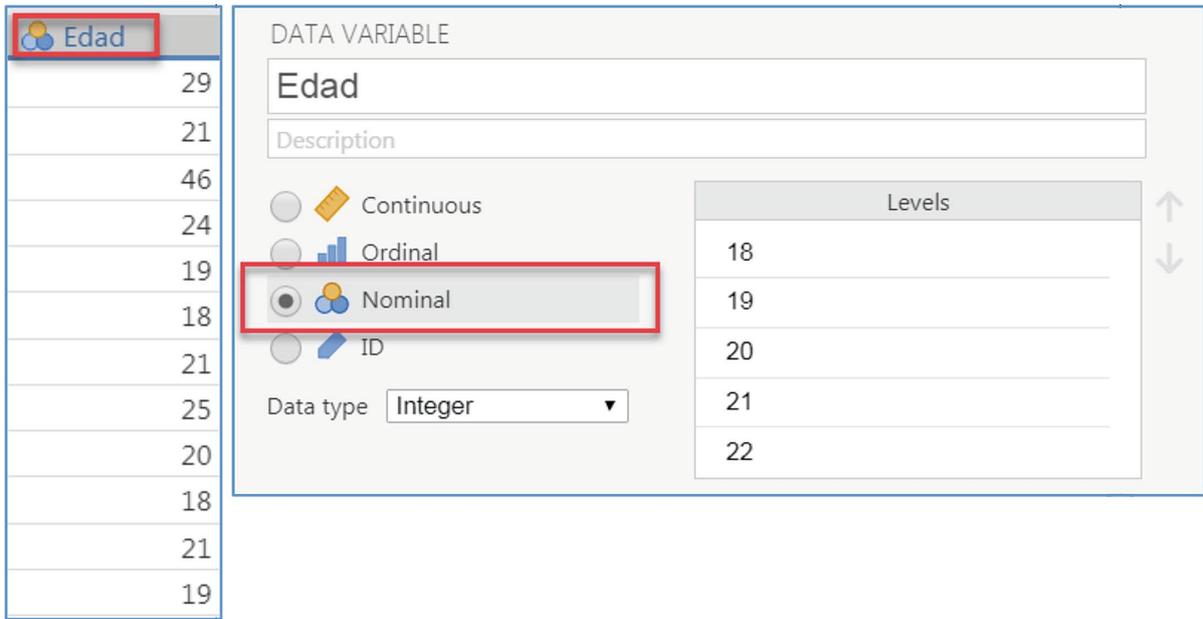
4. *Tipo*. Las variables pueden ser de tres tipos: decimales, enteras y variables de texto. Las variables continuas pueden ser decimales o enteras. Las variables ordinales pueden ser variables enteras o de texto.

5.3. tablaVariables: *tipo*

Tipo de variables	
Números enteros	Integer
Números decimales	Decimal
Nominales	Text

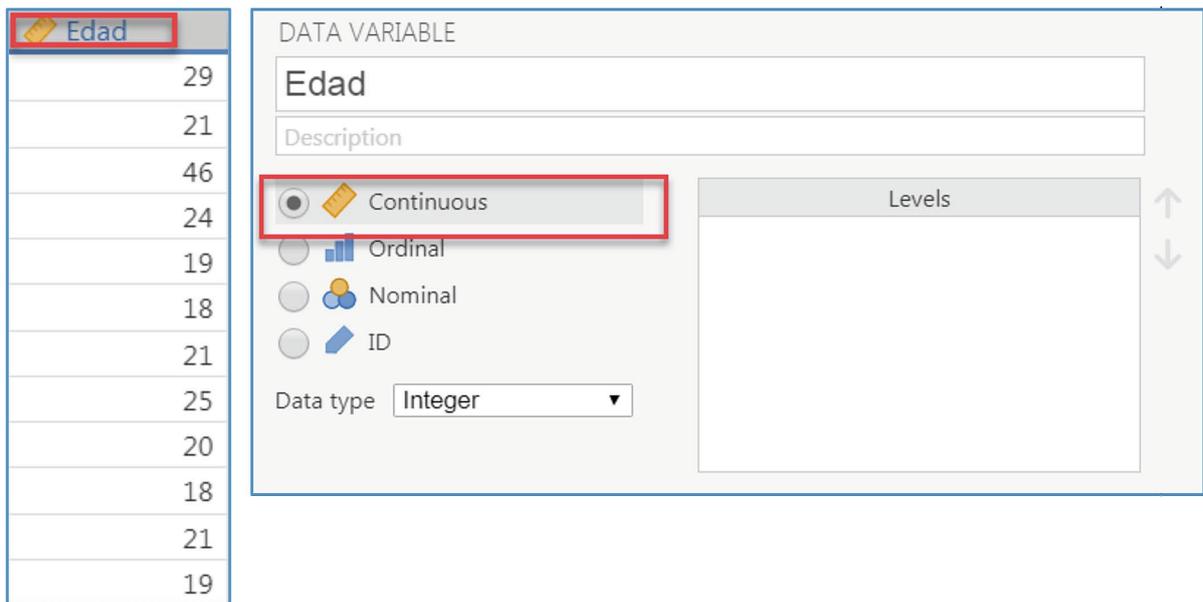
5. Las variables ordinales y las variables nominales tienen categorías o niveles. En esta tabla aparecen los niveles de la variable. Las flechas al lado del recuadro sirven para definir o modificar el orden de los niveles de una variable. El orden de los niveles es importante en los gráficos y en las tablas.
6. Puede ocurrir que en un archivo de datos las variables no contengan todos los niveles posibles. Si se desea mantener los niveles originales habrá que marcar la opción *Retain unused levels*.
7. Esta flecha sirve para indicar a *jamovi* que vuelva a la hoja de datos porque se ha finalizado con la definición de la variable.

En este ejemplo, *jamovi* ha considerado que la variable edad es una variable nominal; para modificar esa consideración, y definir la variable edad como variable continua, entraremos en la pantalla de especificaciones de la variable y marcaremos la casilla correspondiente.



5.3. IMAGEN

Variable edad características (1)



5.4. IMAGEN

Variable edad características (2)

Con esta acción *jamovi* modificará la naturaleza de la variable. Puede comprobarse el cambio en el icono que acompaña al nombre de la variable en la primera fila de la hoja de datos.

5.2. Recodificar ítems

Antes de iniciar el análisis estadístico, es importante determinar el sentido de la medida o la valencia de los ítems; hay que comprobar la dirección en que los ítems reflejan el constructo a medir, redireccionando los ítems inversos.

1. Creo que mi estómago es demasiado grande
 1. Siempre
 2. Casi siempre
 3. Muchas veces
 4. Algunas veces
 5. Casi nunca
 6. Nunca
2. Creo que mi estómago tiene el tamaño adecuado
 1. Siempre
 2. Casi siempre
 3. Muchas veces
 4. Algunas veces
 5. Casi nunca
 6. Nunca

Ambos ítems se relacionan con el mismo constructo, satisfacción corporal, pero en sentido contrario. La inferencia en cuanto a la satisfacción corporal de una persona que selecciona la opción «Siempre» en el primer ítem es opuesta a las conclusiones sobre una persona que elige la opción «Siempre» en el segundo ítem. En el primer caso la puntuación asignada sería 1 y reflejaría un nivel de satisfacción corporal menor que el de aquella persona con una puntuación de 6. Sin embargo, aquél que responde «Siempre» al segundo ítem (1) poseerá un nivel de satisfacción corporal mayor que la persona que responde al mismo ítem con la opción «Casi nunca» (5). Dado que la dirección de la escala indica que a mayor puntuación mayor satisfacción corporal, los ítems que representan el constructo en la dirección opuesta son ítems inversos.



5.5. IMAGEN
Ítems inversos

Antes de proceder al análisis sobre datos que contienen ítems inversos, es necesario redireccionarlos situando todos los ítems de la escala en el mismo sentido. En general, se aplicará la siguiente regla de reconversión:

$$X' = (X_{Ma} + X_{mi}) - X$$

donde X' es la puntuación recodificada,

X_{Ma} y X_{mi} , los valores teóricos máximo y mínimo de la variable a recodificar, y X , la opción de respuesta seleccionada.

De este modo los valores 1, 2, 3, 4, 5 y 6 se reconvierten en 6, 5, 4, 3, 2 y 1.

$$\begin{aligned} X'_1 &= (6 + 1) - 1 = 6 & X'_2 &= (6 + 1) - 2 = 5 \\ X'_3 &= (6 + 1) - 3 = 4 & X'_4 &= (6 + 1) - 4 = 3 \\ X'_5 &= (6 + 1) - 5 = 2 & X'_6 &= (6 + 1) - 6 = 1 \end{aligned}$$

La utilización de ítems inversos en la construcción de escalas pretende controlar el sesgo relacionado con el estilo de respuesta. Se utilizan ítems formulados positivamente junto con ítems inversos para el control de la deseabilidad social o tendencia a responder en función de lo socialmente aceptado, y la aquiescencia o tendencia a manifestar un acuerdo sistemático con el enunciado del ítem independientemente de su contenido. Es una práctica frecuente que sin embargo, está siendo cuestionada (Suárez-Álvarez *et al.*, 2018).

Para el ejemplo que nos ocupa las reglas de recodificación son:

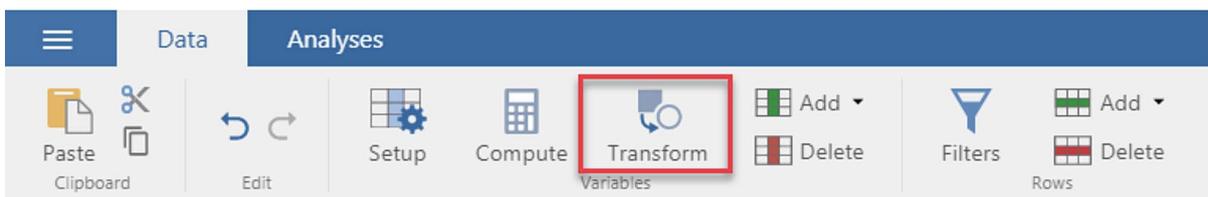
5.4. tabla

Reglas de recodificación

Ítem	Sentido
1, 2, 6, 7, 9	Directo
3, 4, 5, 8, 10	Inverso

Aunque es posible realizar análisis de fiabilidad en *jamovi* sin recodificar los ítems, en este ejemplo y dado que necesitamos calcular puntuaciones totales, y obtener información detallada sobre cada ítem, es aconsejable situar todos los ítems en la misma dirección sintomática. Para ello, seguiremos la siguientes pautas:

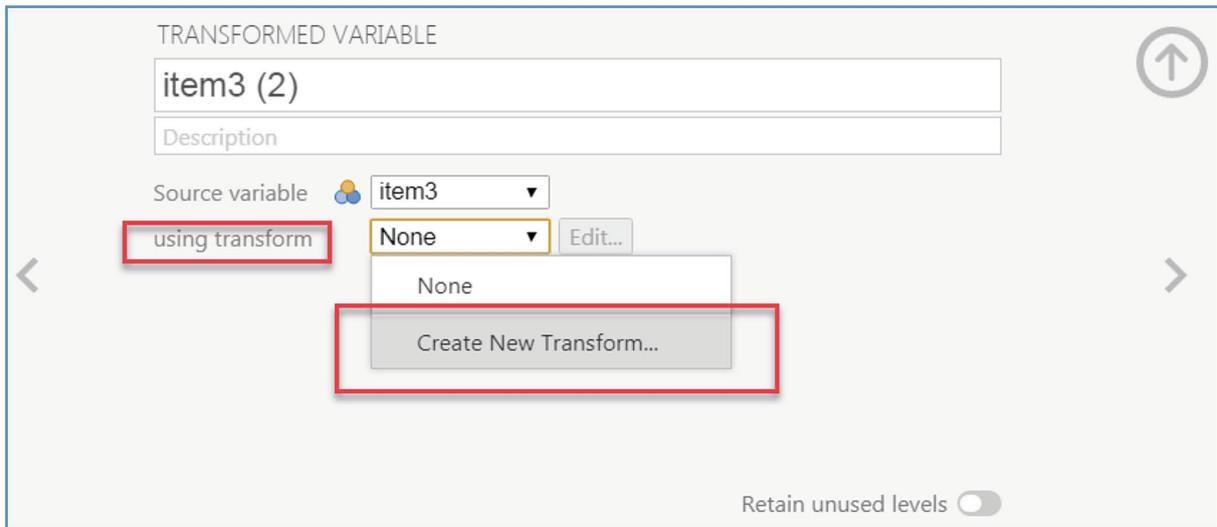
1. Situándonos en cualquier punto de la variable a recodificar presionar la opción *Transform*.



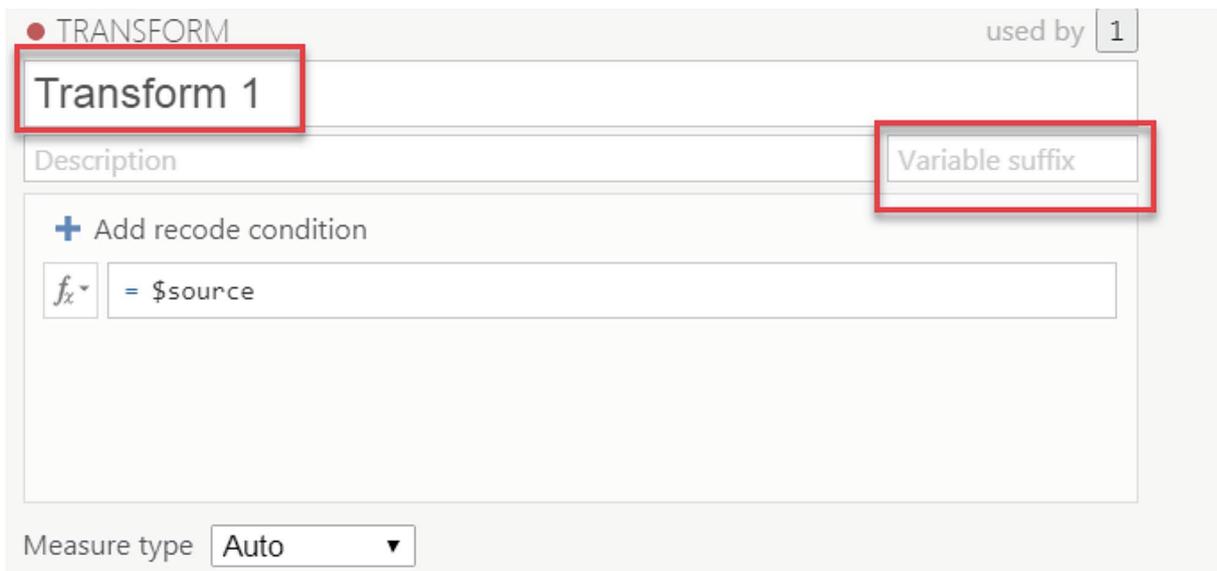
5.6. IMAGEN

Recodificar ítems (1)

2. En esta pantalla seleccionaremos la opción *using transform* y dentro de ella *Create New Transform*.



5.7. IMAGEN

Recodificar ítems (2)

5.8. IMAGEN

Recodificar ítems (3)

3. Es posible identificar la transformación con un nombre específico, por ejemplo, *item_rec*, y especificar un sufijo *Variable suffix* que será añadido al nombre de la variable recodificada; por ejemplo *_rec*. Junto a ello definiremos la regla de recodificación:

$$X' = (6 + 1) - X$$

Valor máximo (6) más valor mínimo (1) menos la opción de respuesta, que es añadida por *jamovi* utilizando la sintaxis $\$source$.

TRANSFORM used by 1

item_rec

Description: _rec

+ Add recode condition

$f_x = (6+1) - \$source$

Measure type: Auto

5.9. IMAGEN

Recodificar ítems (4)

Estas especificaciones generan una nueva variable, *item3_rec*:

item3	item3_rec
2	5
6	1
5	2
1	6
2	5
3	4
1	6
5	2
2	5
6	1
1	6
4	3

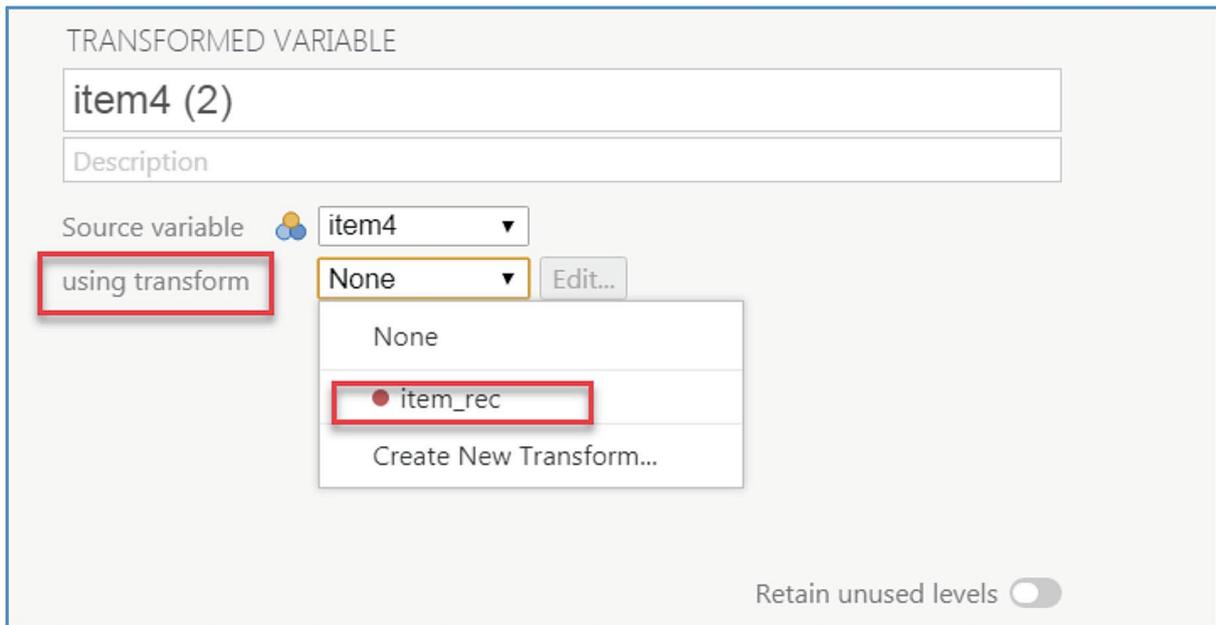
5.10. IMAGEN

Recodificar ítems (5)

Al lado del nombre de la nueva variable, *item3_rec*, *jamovi* añade un punto de color, que indica que esta variable no existía en el fichero original; se trata de una variable generada durante el proceso de análisis.

Una vez recodificado un ítem, se puede aplicar la misma regla de recodificación a otras variables (ítems). Para ello se seguirán las siguientes pautas:

1. Posicionados en el ítem a recodificar seleccionar la opción *Transform* presionando el botón derecho del ratón (del mismo modo que en el ejemplo anterior).
2. En esta ventana y dentro de la opción *using transformation* seleccionar la transformación definida en el ejemplo anterior, item-rec:



5.11. IMAGEN

Recodificar ítems (6)

Con esos dos pasos *jamovi* generará la versión recodificada del ítem correspondiente.

NOTA. Una vez definida una regla de transformación es posible aplicarla a otras variables.

6

Describir la muestra

La descripción de los participantes es esencial dentro de cualquier investigación. Exige una presentación lo más clara posible de: 1) la población destinataria del test; 2) el tipo de muestreo utilizado y su justificación; 3) el número de personas que componen la muestra, y 4) la descripción de las características de los participantes y de grupos específicos o distintivos, si es que los hubiere (edad, sexo, nivel de estudios, nivel socioeconómico...).

Para describir la muestra pueden utilizarse análisis de frecuencias, estadísticos descriptivos (media, desviación típica), tablas de contingencia e incluso representaciones gráficas (histogramas, diagramas de cajas, gráficos de densidad...). Todas ellas está disponibles en la opción de exploración de datos (*Exploration*) del menú de análisis (*Analyses*).

6.1. Describir una variable

Por ejemplo, para conocer la distribución de los participantes en esta investigación en función del sexo, podemos solicitar la tabla de frecuencias de esta variable. Para ello seguiremos los siguientes pasos:

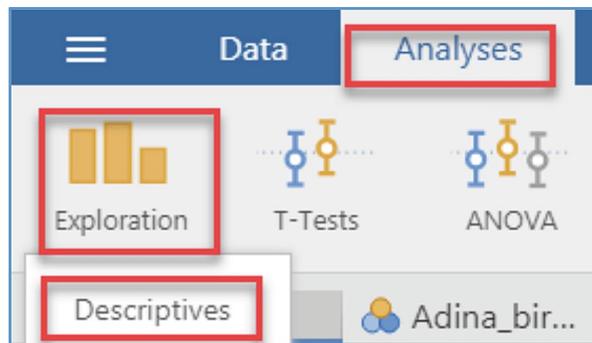
1. *Seleccionar la opción adecuada.* Dentro del menú principal de *jamovi* seleccionar la opción *Analyses*.



6.1. IMAGEN

Descriptivos (1)

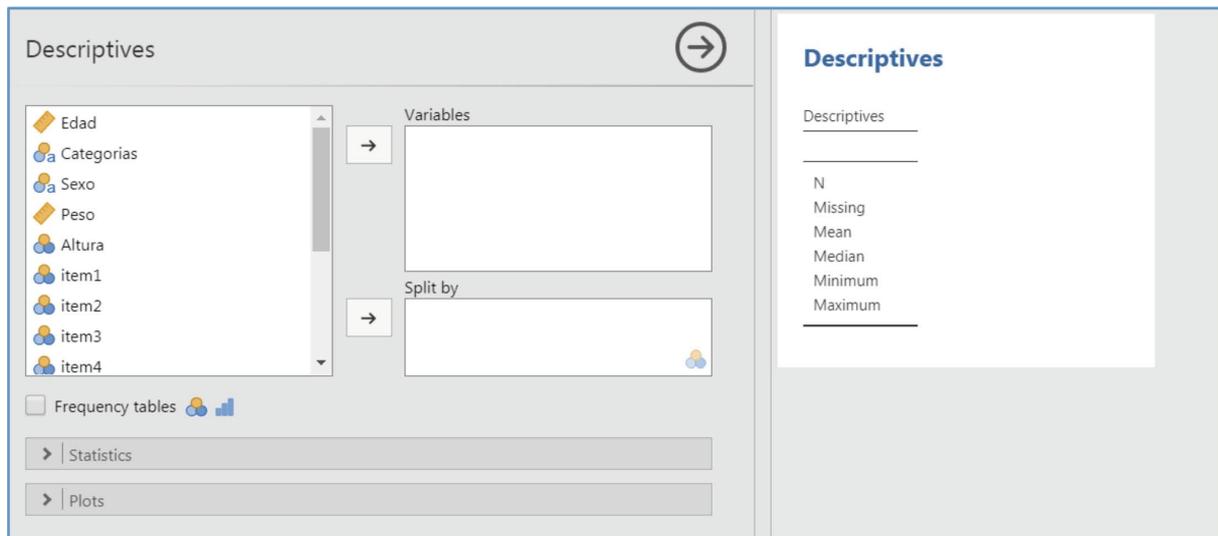
Presionamos la opción de exploración de datos, *Exploration* y dentro de ella seleccionamos *Descriptives*.



6.2. IMAGEN

Descriptivos (2)

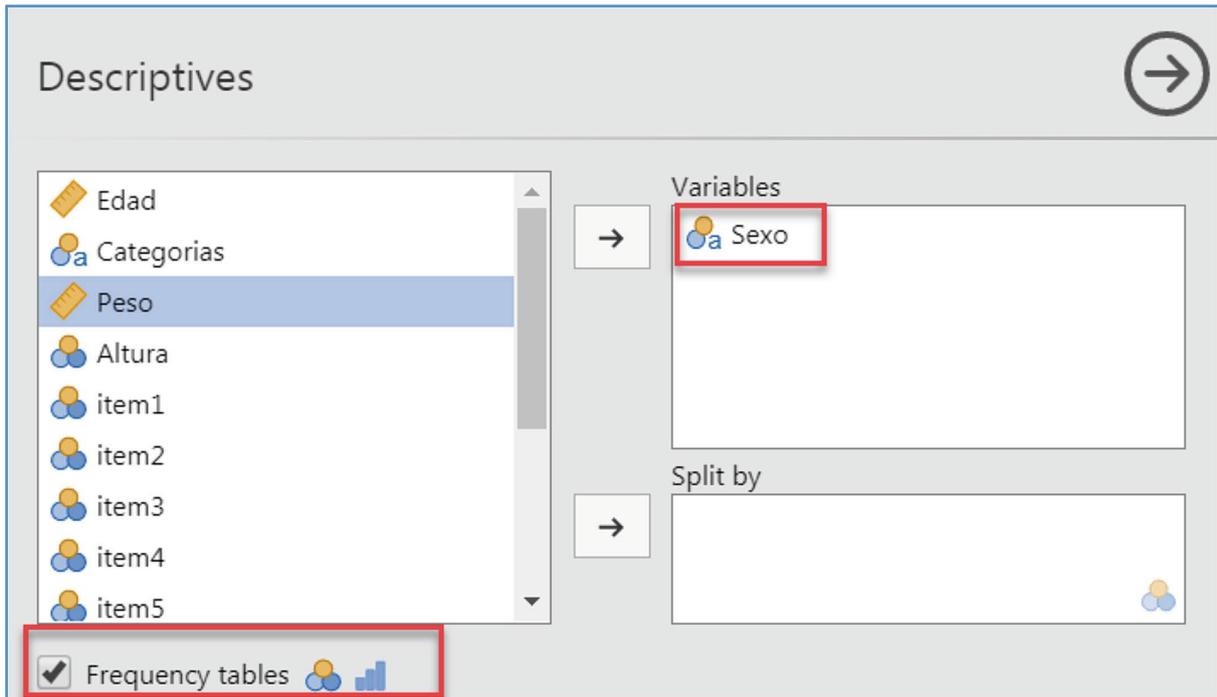
Esa acción nos lleva a la siguiente pantalla, donde podremos seleccionar las variables a analizar, así como definir los estadísticos descriptivos y gráficos deseados:



6.3. IMAGEN

Descriptivos (3)

2. *Seleccionar variables.* En esta pantalla se seleccionan las variables que se quieren analizar, y se trasladan a la ventana derecha. Para ello basta moverlas con el ratón de una ventana a otra, o una vez marcada la variable presionar la flecha que separa ambas ventanas. También es posible seleccionar más de una variable al mismo tiempo.



6.4. IMAGEN

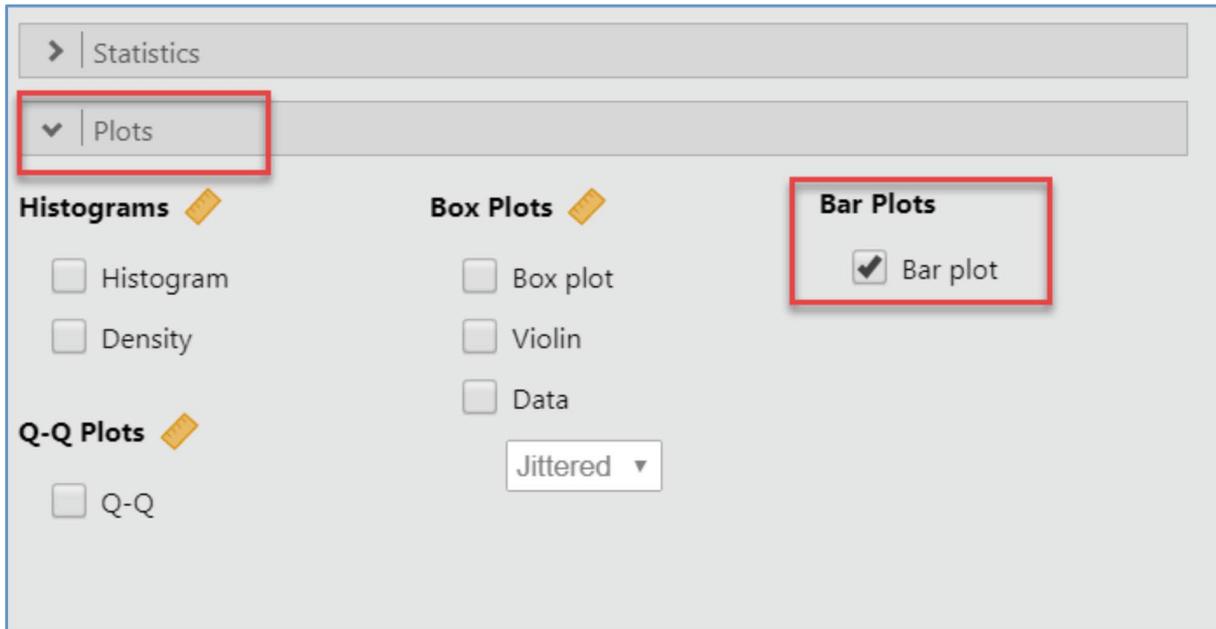
Descriptivos (4)

3. *Seleccionar el estadístico apropiado.* Para la descripción de la variable sexo, que es una variable categórica, solicitamos a *jamovi* la tabla de frecuencias. Como resultado obtenemos la siguiente salida:

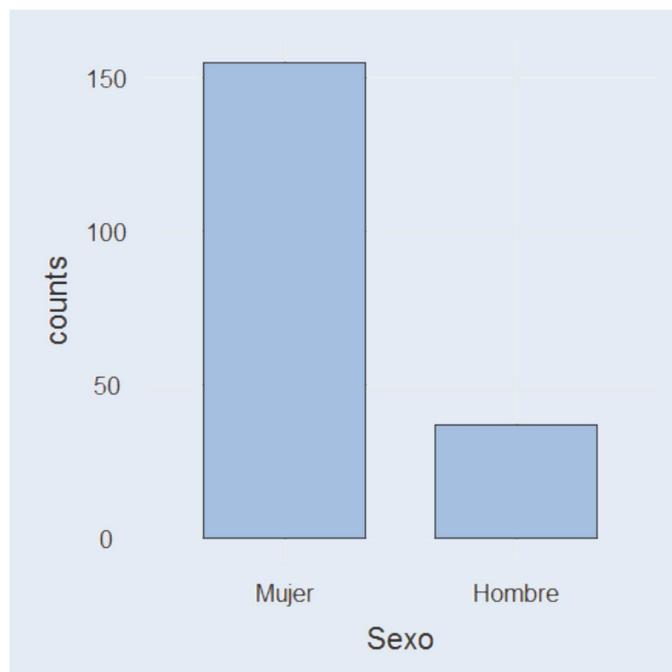
Descriptives			
Sexo			
N	192		
Frequencies			
Frequencies of Sexo			
Levels	Counts	% of Total	Cumulative %
Mujer	155	80.73 %	80.73 %
Hombre	37	19.27 %	100.00 %

De los resultados se desprende que en este estudio participaron 192 personas, de las cuales 155 (80,73%) son mujeres y 37 (19,27%) son varones.

4. *Seleccionar opciones gráficas.* Para solicitar un diagrama de barras que recoja esta información abriremos la opción de gráficas (*plots*), y marcaremos la opción correspondiente:



6.5. IMAGEN

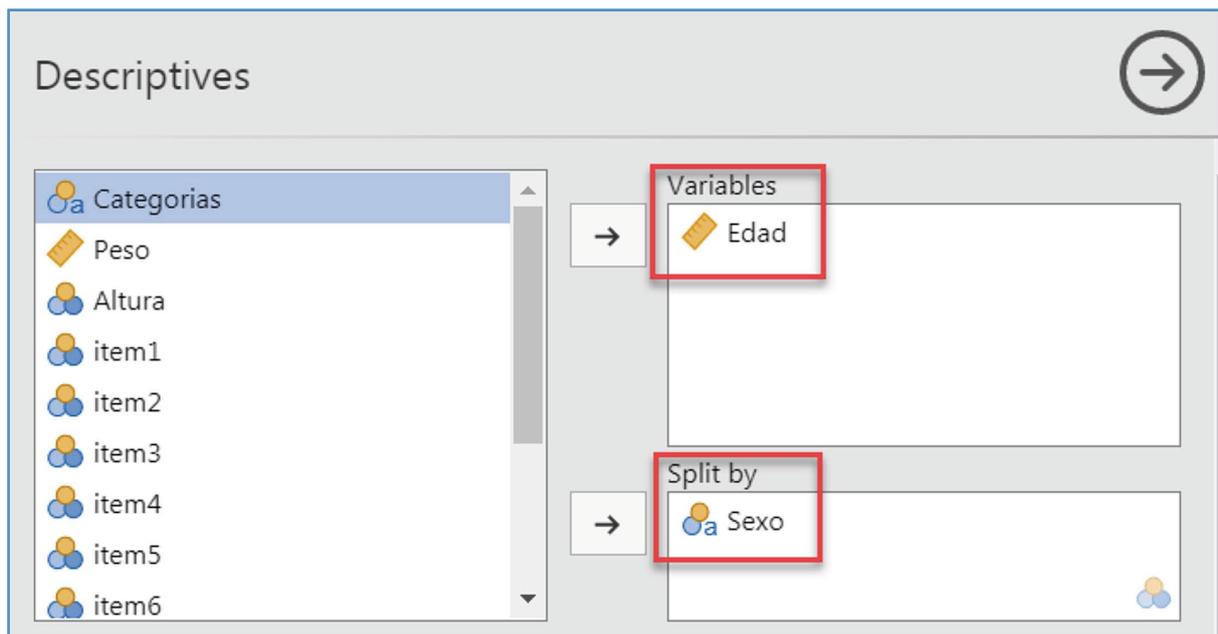
Descriptivos (5)

6.2. Describir dos variables conjuntas

Puede interesarnos conocer la edad media de los participantes en función del sexo; es decir, la media de edad de los varones, y la media de edad de las mujeres.

Para ello dentro de la opción *Analyses y Exploration*:

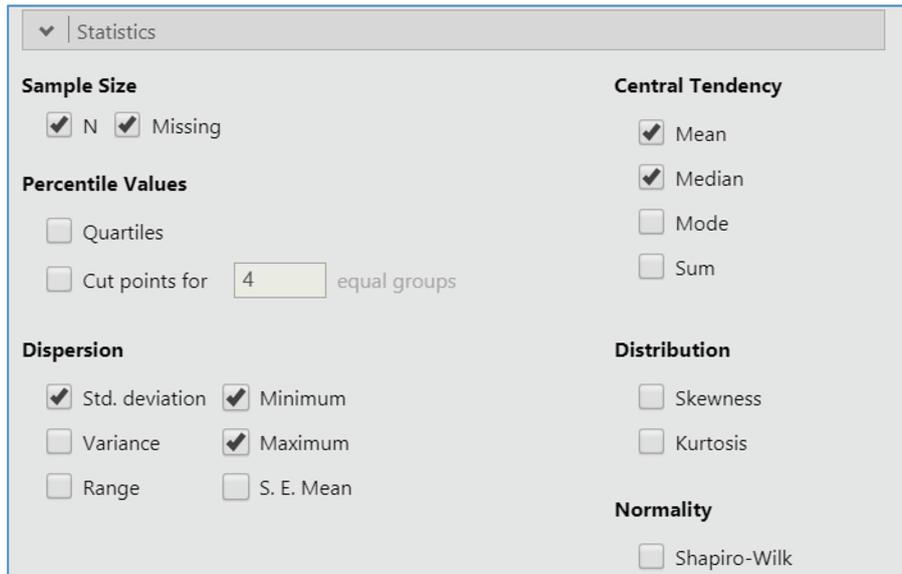
1. *Seleccionar variables*. Marcamos la variable Edad como variable a analizar y la movemos a la ventana derecha. Para indicar que queremos conocer la edad en función del sexo, movemos esta última a la ventana *Split by*. Esta acción, genera dos sub-muestras, una compuesta por mujeres y la segunda formada por hombres.



6.6. IMAGEN

Descriptivos en función del sexo (1)

2. *Seleccionar el estadístico apropiado*: En la ventana de estadísticos descriptivos (*Statistics*) podemos marcar entre otras, las opciones correspondientes a:
 - *N*: número de participantes.
 - *Mean*: media aritmética.
 - *Median*: mediana.
 - *Std. Deviation*: desviación estándar.
 - *Minimum*: mínimo.
 - *Maximum*: máximo.
 - *Variance*: varianza.
 - *Range*: rango.
 - *Mode*: moda.
 - *Sum*: suma.
 - *Skewness*: asimetría.
 - *Kurtosis*: curtosis.



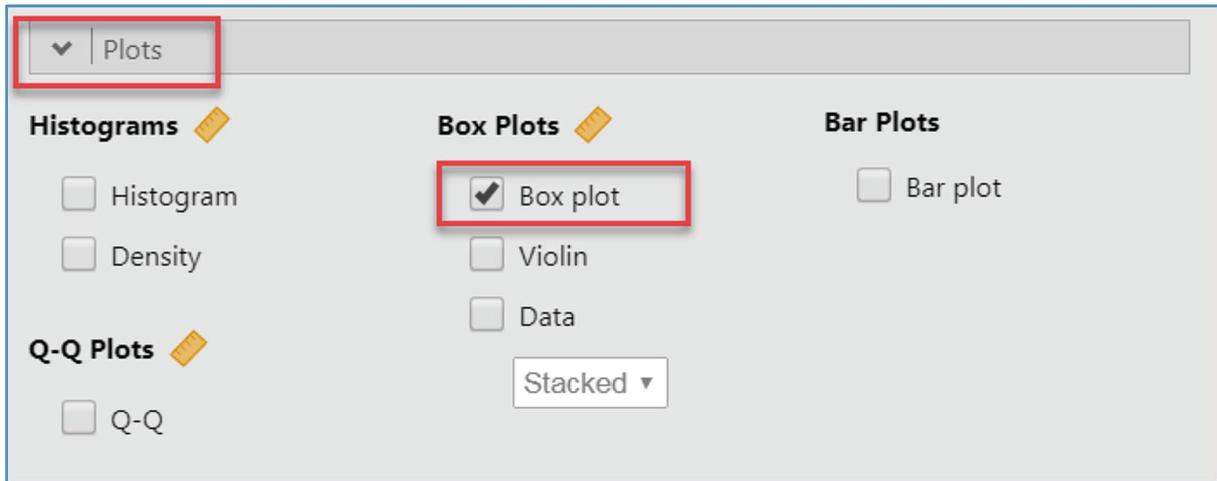
6.7. IMAGEN

Descriptivos en función del sexo (2)

Esta acción genera en la ventana de salidas la siguiente tabla. En ella puede leerse que el grupo de mujeres tiene una edad media de 22,43 años (desviación típica = 6,05 años) y que el grupo de varones tiene una edad media de 21,59 años (DT = 5,06). Los más jóvenes, tanto en el grupo de hombres como en el grupo de mujeres, tienen 18 años; y los mayores tienen 49 y 46 años. Los valores de la mediana para ambas submuestras son 21 y 20.

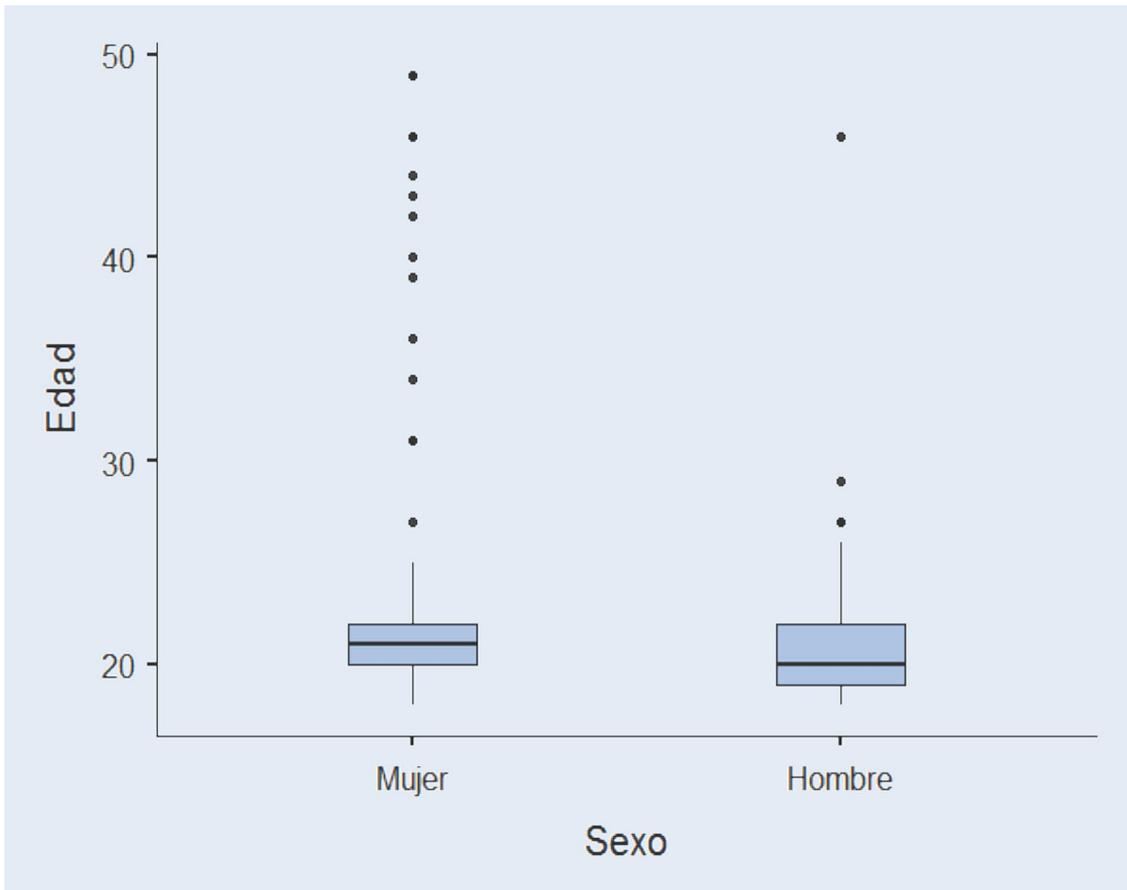
Descriptives		
	Sexo	Edad
N	Mujer	155
	Hombre	37
Mean	Mujer	22.43
	Hombre	21.59
Median	Mujer	21
	Hombre	20
Standard deviation	Mujer	6.055
	Hombre	5.069
Minimum	Mujer	18
	Hombre	18
Maximum	Mujer	49
	Hombre	46

3. *Seleccionar opciones gráficas.* Si se desea aportar información gráfica, dentro del menú *Plots*, podríamos marcar la opción para el diagrama de cajas (*Box plot*).



6.8. IMAGEN

Descriptivos en función del sexo (3)



7

Análisis de ítems

El análisis de ítems estudia las características sustantivas y formales de los ítems de una escala. El análisis ha de ser cualitativo y cuantitativo, y siempre ligado a los objetivos del test. El análisis cualitativo de ítems estudia aspectos lingüísticos y de contenido con el fin de evaluar la adecuación lingüística, cultural y sustantiva con relación al objeto medido y a la población diana (Elosua, 2003). Sólo tras una primera fase, ineludible, en la que un grupo de expertos evalúa (*raters mediated assessment*) cada uno de los ítems, se dará inicio a una fase operacional en la que cobra relevancia el aspecto cuantitativo.

Los índices numéricos más habituales en el análisis de ítems incluyen los índices relacionados con el modelo formal y los índices conectados con la validez de las puntuaciones. Dentro de la teoría clásica de test pertenecen al primer grupo el índice de dificultad y el índice de discriminación. El primero se apoya en la distribución univariada de las respuestas, y el segundo se construye sobre las relaciones entre ítems. Entre los índices relacionados con la validez el índice de validez y el funcionamiento diferencial del ítem son los más destacados. El primero estudia la relación del ítem con un criterio externo, y el segundo es una herramienta útil en el estudio de la estructura interna del test.

Excepto el funcionamiento diferencial del ítem, el resto son índices clásicos de la literatura psicométrica, y son tratados prácticamente en todos los manuales de psicometría. El estudio del funcionamiento diferencial del ítem es algo más reciente respecto a los demás índices; sin embargo, ha de reconocerse que tiene ya una amplia trayectoria dentro de la psicometría y que existen procedimientos de estimación sencilla accesibles al investigador novel.

7.1. Índices basados en la distribución del ítem

En los ítems de ejecución máxima (por ejemplo ítems de aptitudes, rendimiento...) la media aritmética del ítem informa sobre el nivel de dificultad del elemento; de hecho, en condiciones de dicotomía (correcto/incorrecto) la media aritmética del ítem se conoce como índice de dificultad del ítem (p_i), y su valor es igual a la proporción de sujetos que responden correctamente al ítem.

En el caso de ítems de ejecución típica (por ejemplo los ítems utilizados en inventarios o cuestionarios de personalidad, intereses, actitudes...) el concepto de dificultad carece de sentido, y en su lugar habría de hablarse de la media aritmética del ítem.

En cualquiera de los dos casos es importante complementar la información aportada por la media aritmética con estadísticos que informen sobre la distribución de las respuestas; tales como, número o porcentaje de personas por opción, desviación típica de la distribución, y valores de asimetría y curtosis. Todos ellos permitirán estudiar la normalidad de la distribución, estimar los efectos techo y suelo, y en definitiva ahondar en las características formales del ítem (Ho y Carol, 2015). Por ejemplo, la literatura considera los índices de asimetría y curtosis como una aproximación a la evaluación de la normalidad de las variables. Se consideran valores aceptables aquellos situados en el rango $[-1, + 1]$ (Ferrando y Anguiano-Carrasco, 2010), $[-1.5, + 1.5]$ (Forero, Maydeu-Olivares, y Gallardo-Pujol, 2009), o incluso entre valores de $[-2, + 2]$ (Bandalos y Finney, 2010).

Para obtener información sobre los índices basados en la distribución utilizaremos las opciones incluidas en el menú *Analyses*. Los pasos a seguir serían:

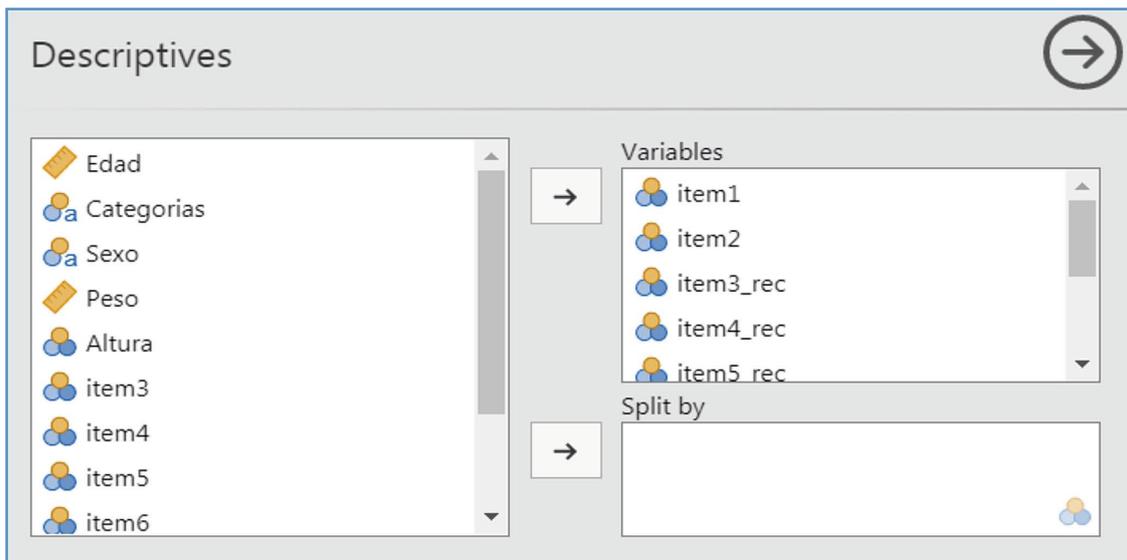
1. *Seleccionar la opción adecuada.* Dentro del menú *Analyses* elegir la opción *Exploration* y dentro de ella *Descriptives*:



7.1. IMAGEN

Análisis de ítems (1)

2. *Seleccionar variables.* En la ventana que ofrece *jamovi*, seleccionar los ítems a analizar.



7.2. IMAGEN

Análisis de ítems (2)

NOTA. Se seleccionarán los ítems recodificados en los casos en los que existieren.

3. *Seleccionar el estadístico apropiado.* En la ventana *Statistics* se marcarán aquellos estadísticos que se deseen obtener.

The screenshot shows the 'Statistics' dialog box with the following options selected:

- Sample Size:** N, Missing
- Percentile Values:** Quartiles, Cut points for 4 equal groups
- Dispersion:** Std. deviation, Minimum, Variance, Maximum, Range, S. E. Mean
- Central Tendency:** Mean, Median, Mode, Sum
- Distribution:** Skewness, Kurtosis

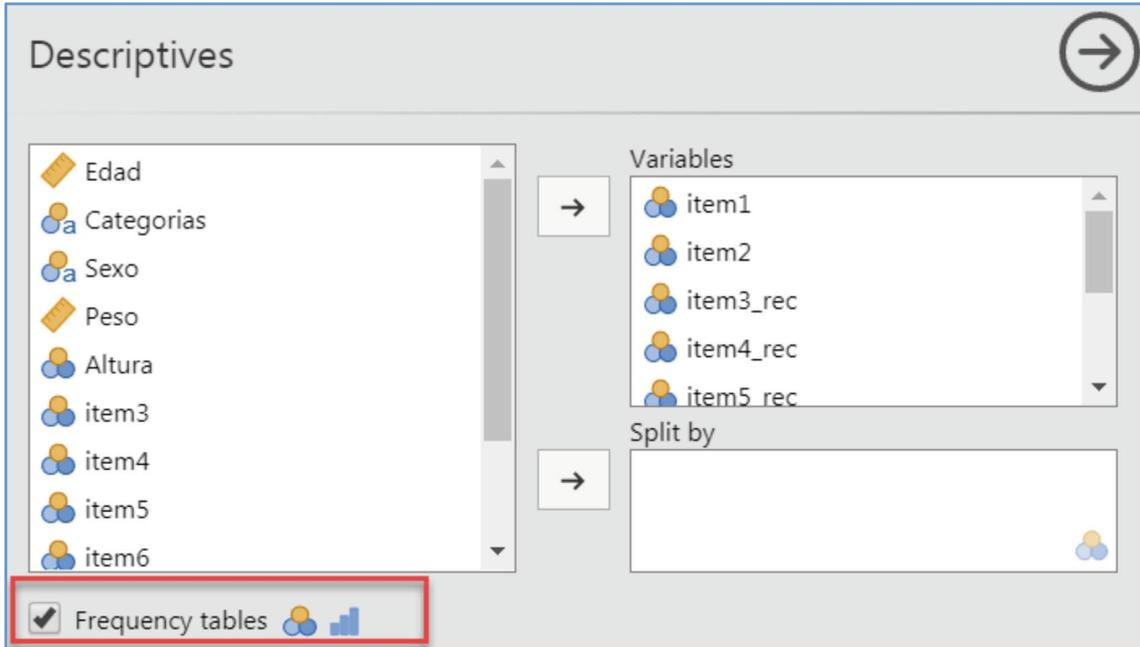
7.3. IMAGEN

Análisis de ítems (3)

Con esas pautas *jamovi* genera la siguiente salida:

Descriptives										
	item1	item2	item3_rec	item4_rec	item5_rec	item6	item7	item8_rec	item9	item10_rec
N	192	192	192	192	192	192	192	192	192	192
Missing	0	0	0	0	0	0	0	0	0	0
Mean	4.34	3.95	4.10	3.56	3.45	4.46	4.40	3.38	4.46	3.60
Median	5.00	4.00	4.50	3.00	3.00	5.00	5.00	3.00	5.00	4.00
Standard deviation	1.74	1.77	1.68	1.37	1.68	1.77	1.24	1.69	1.66	1.70
Minimum	1	1	1	1	1	1	1	1	1	1
Maximum	6	6	6	6	6	6	6	6	6	6
Skewness	-0.711	-0.385	-0.433	0.0103	0.0452	-0.893	-0.752	0.109	-0.868	-0.0744
Std. error skewness	0.175	0.175	0.175	0.175	0.175	0.175	0.175	0.175	0.175	0.175
Kurtosis	-0.840	-1.12	-1.12	-0.995	-1.30	-0.596	0.111	-1.26	-0.517	-1.28
Std. error kurtosis	0.349	0.349	0.349	0.349	0.349	0.349	0.349	0.349	0.349	0.349

4. *Solicitar tabla de frecuencias.* La tabla de frecuencias de cada ítem permite analizar la atracción que cada opción de respuesta genera en la muestra.



7.4. IMAGEN

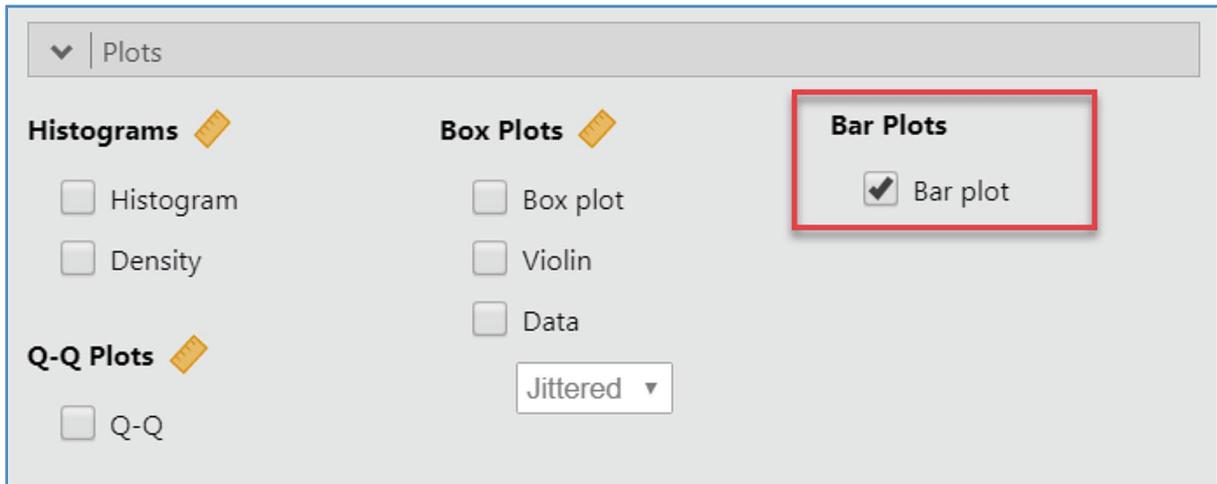
Análisis de ítems (4)

Frequencies of item1				Frequencies of item2			
Levels	Counts	% of Total	Cumulative %	Levels	Counts	% of Total	Cumulative %
1	22	11 %	11 %	1	30	16 %	16 %
2	14	7 %	19 %	2	15	8 %	23 %
3	23	12 %	31 %	3	25	13 %	36 %
4	23	12 %	43 %	4	41	21 %	58 %
5	38	20 %	63 %	5	27	14 %	72 %
6	72	38 %	100 %	6	54	28 %	100 %
Frequencies of item3_rec				Frequencies of item4_rec			
Levels	Counts	% of Total	Cumulative %	Levels	Counts	% of Total	Cumulative %
1	17	9 %	9 %	1	10	5 %	5 %
2	26	14 %	22 %	2	39	20 %	26 %
3	27	14 %	36 %	3	49	26 %	51 %
4	26	14 %	50 %	4	34	18 %	69 %
5	42	22 %	72 %	5	47	24 %	93 %

7. ANÁLISIS DE ÍTEMS

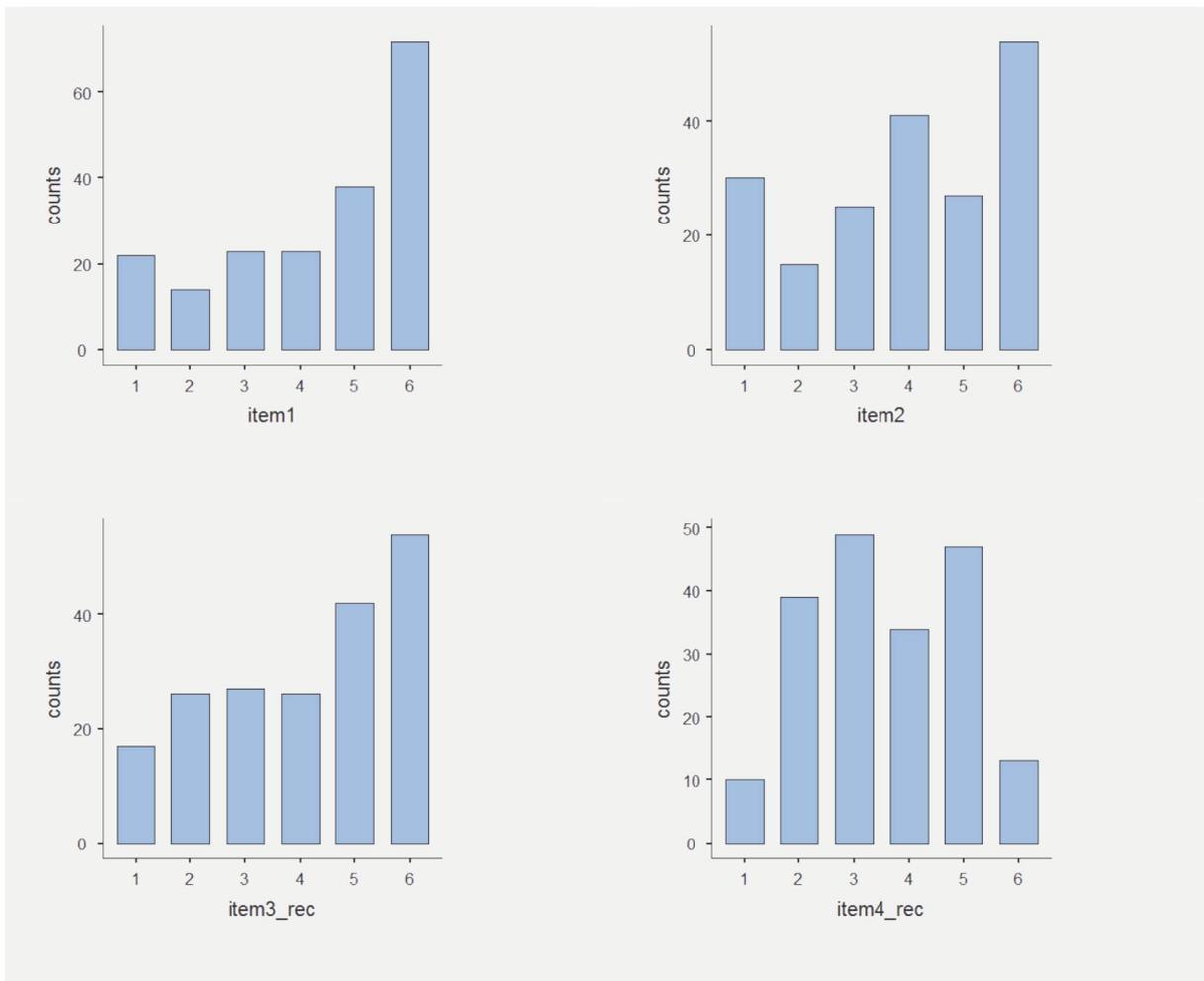
6	54	28 %	100 %	6	13	7 %	100 %
Frequencies of item5_rec				Frequencies of item6			
Levels	Counts	% of Total	Cumulative %	Levels	Counts	% of Total	Cumulative %
1	29	15 %	15 %	1	24	13 %	13 %
2	38	20 %	35 %	2	13	7 %	19 %
3	36	19 %	54 %	3	12	6 %	26 %
4	21	11 %	65 %	4	24	13 %	38 %
5	42	22 %	86 %	5	40	21 %	59 %
6	26	14 %	100 %	6	79	41 %	100 %
Frequencies of item7				Frequencies of item8_rec			
Levels	Counts	% of Total	Cumulative %	Levels	Counts	% of Total	Cumulative %
1	4	2 %	2 %	1	33	17 %	17 %
2	16	8 %	10 %	2	35	18 %	35 %
3	16	8 %	19 %	3	40	21 %	56 %
4	54	28 %	47 %	4	21	11 %	67 %
5	67	35 %	82 %	5	37	19 %	86 %
6	35	18 %	100 %	6	26	14 %	100 %
Frequencies of item9				Frequencies of item10_rec			
Levels	Counts	% of Total	Cumulative %	Levels	Counts	% of Total	Cumulative %
1	17	9 %	9 %	1	28	15 %	15 %
2	18	9 %	18 %	2	31	16 %	31 %
3	13	7 %	25 %	3	36	19 %	49 %
4	26	14 %	39 %	4	24	13 %	62 %
5	48	25 %	64 %	5	41	21 %	83 %
6	70	36 %	100 %	6	32	17 %	100 %

5. *Seleccionar opciones gráficas.* La información puede completarse con gráficos que se generarán marcando la opción correspondiente al diagrama de barras (*bar-plot*) del menú de gráficos (*Plots*).

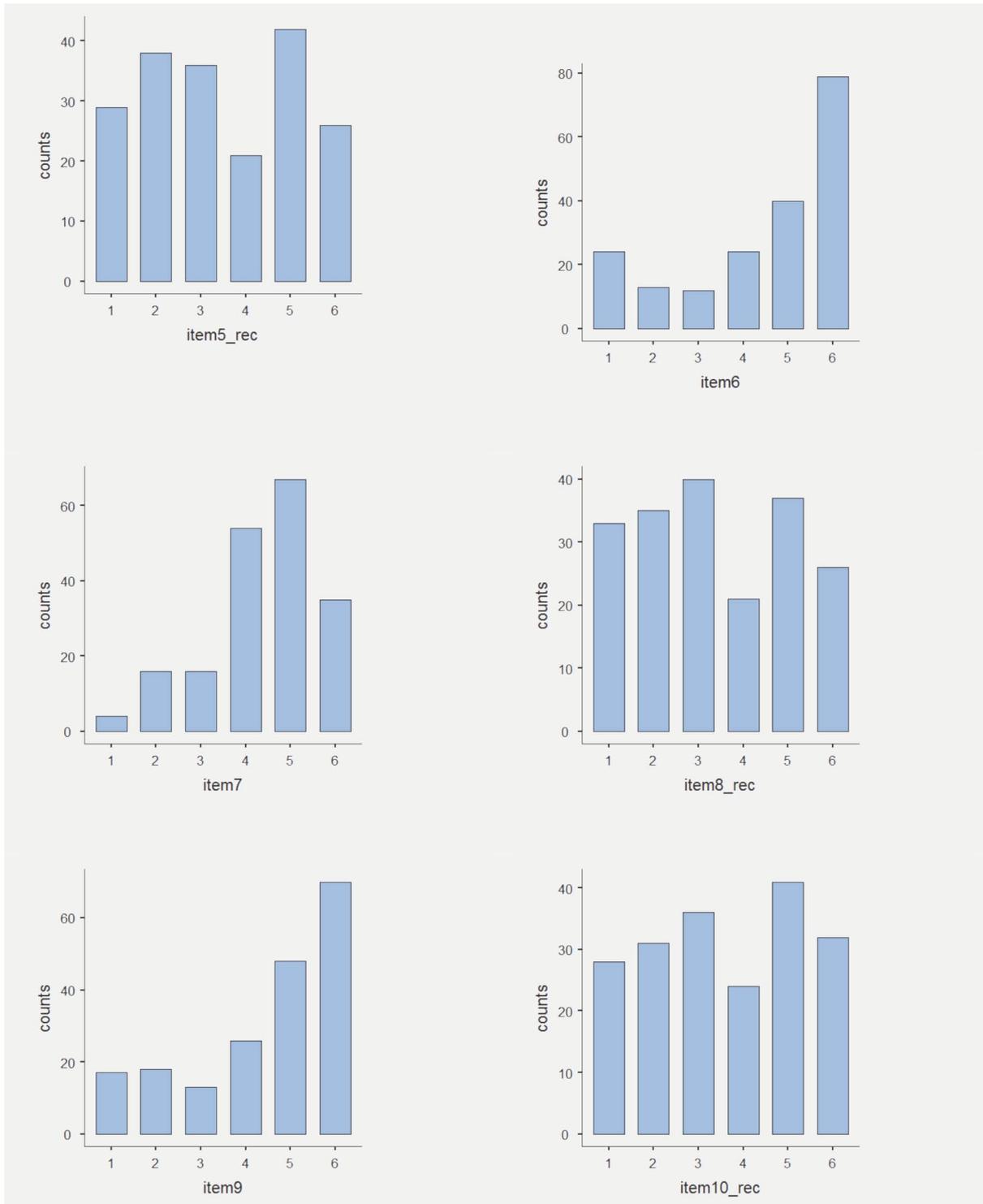


7.5. IMAGEN

Análisis de ítems (5)



7. ANÁLISIS DE ÍTEMS



7.2. Índice de discriminación

La relación entre cada uno de los ítems que componen una escala y la escala misma se analiza por medio del índice de discriminación (r_{iX}); la información aportada por este índice está ligada a la consistencia interna del test. El índice se obtiene por medio del coeficiente de correlación entre el ítem objeto de estudio y la puntuación total. Aunque teóricamente y al tratarse de una correlación los valores del índice de discriminación varían en un rango de $[-1, +1]$, en el contexto del análisis de ítems no tendrían sentido los índices de discriminación negativos.

A pesar de que la fijación de puntos de corte referidos a la calidad de un índice es siempre interpretable en función del contexto, entre la comunidad psicométrica se acepta la siguiente clasificación referida al índice de discriminación (Ebel, 1965).

7.1. tabla

Niveles de discriminación

r_{iX}	Nivel de discriminación
$0,40 < r_{iX}$	Muy bueno
$0,30 < r_{iX} < 0,39$	Bueno
$0,20 < r_{iX} < 0,29$	Insuficiente
$r_{iX} < 0,19$	Malo

Para obtener el índice de discriminación de un ítem se seguirán las siguientes pautas:

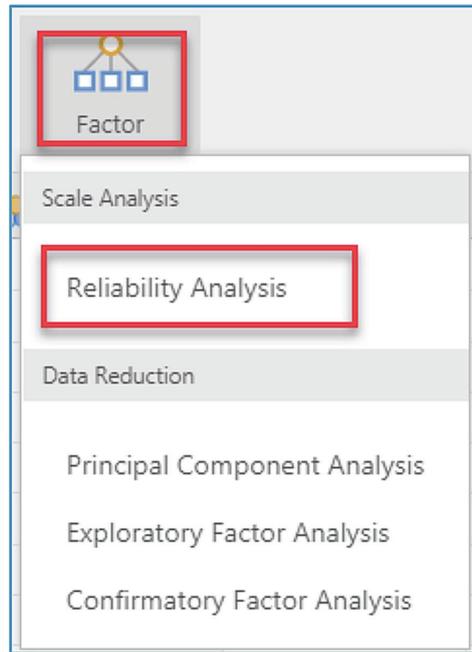
1. *Seleccionar la opción adecuada.* El índice de discriminación se encuentra dentro del menú *Analyses* bajo las opciones que encierra *Factor*.



7.6. IMAGEN

Índice de discriminación (1)

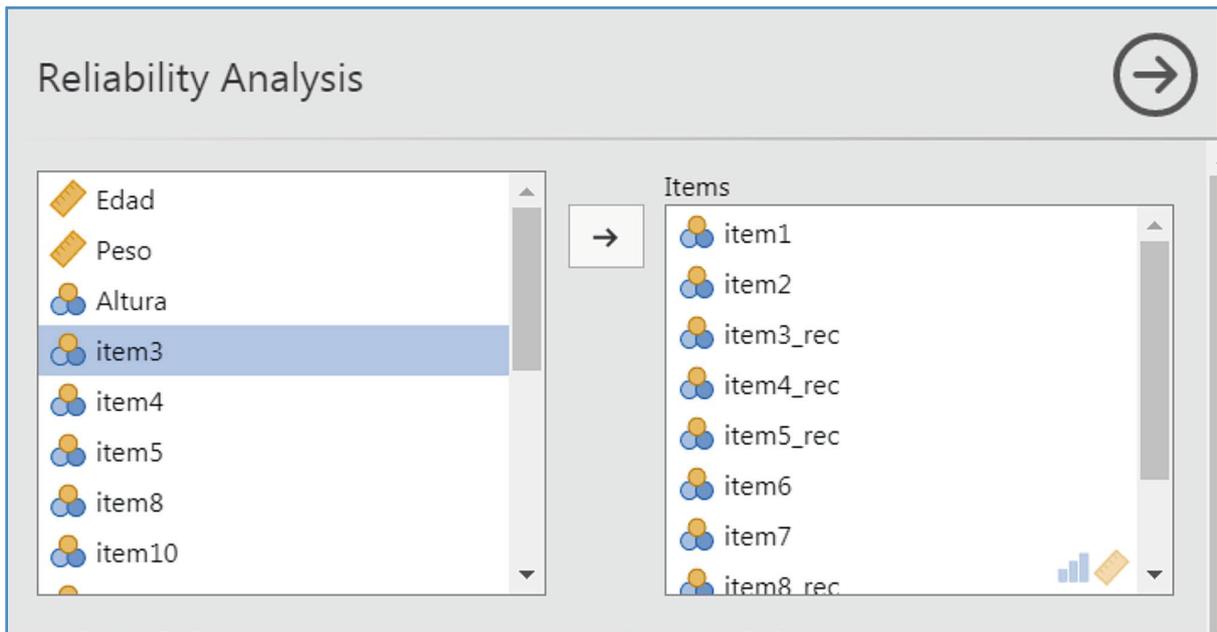
Dentro de ella marcaremos *Reliability Analysis*.



7.7. IMAGEN

Índice de discriminación (2)

2. *Seleccionar variables.* En la ventana *Reliability Analysis* se seleccionan los ítems que se quieren analizar.



7.8. IMAGEN

Índice de discriminación (3)

3. *Seleccionar el estadístico apropiado.* Dentro de la opción *Statistics* marcar la correspondiente al índice de discriminación, *item-test correlation*.

The image shows a screenshot of the SPSS Statistics dialog box. It is divided into two main sections: 'Scale Statistics' and 'Item Statistics'. Under 'Scale Statistics', there are four unchecked checkboxes: Cronbach's α , McDonald's ω , Mean, and Standard deviation. Under 'Item Statistics', there are five checkboxes: Cronbach's α (if item is dropped), McDonald's ω (if item is dropped), Mean, Standard deviation, and Item-rest correlation. The 'Item-rest correlation' checkbox is checked and highlighted with a red rectangular box. Below these sections is an 'Additional Options' section with one unchecked checkbox: Correlation heatmap.

7.9. IMAGEN

Índice de discriminación (4)

Esta acción genera una tabla con los índices de discriminación de cada ítem. El estudio de estos valores permite adoptar decisiones sobre la pertinencia de modificar o eliminar alguno/s ítem/s de la escala.

Item Reliability Statistics	
	item-rest correlation
item1	0.5171
item2	0.7686
item3_rec	0.6156
item4_rec	0.7209
item5_rec	0.6434
item6	0.6717
item7	0.2060
item8_rec	0.7356
item9	0.7028
item10_rec	0.7352

Los valores obtenidos muestran índices de discriminación aceptables excepto para el ítem 7 que con un valor de 0,2060 es marcadamente inferior al resto de los ítems que componen la escala.

NOTA. Si durante el proceso de análisis se opta por eliminar alguno de los ítem, es preceptivo reanalizar los datos.

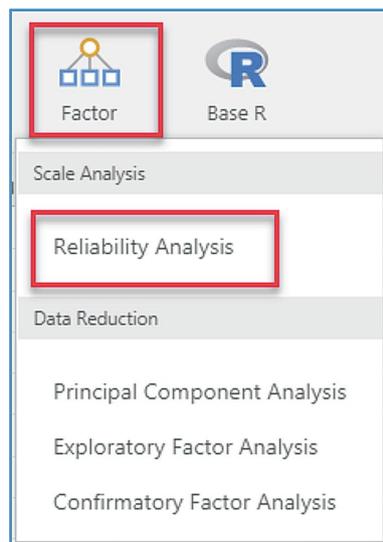
7.3. Matriz de correlación entre ítems

El estudio de las relaciones entre ítems ocupa un lugar destacado en el análisis de ítems; su estudio es abordado por los índices de relación, que básicamente analizan las correlaciones entre ítems. El procedimiento de estimación de la correlación depende de la naturaleza (discreta/continua) de las variables. Es posible estimar una correlación biserial, policórica (tetracórica en el caso de ítems dicotómicos) o producto-momento de Pearson.

Los valores esperables en la matriz de correlaciones entre ítems son por lo general, positivos. La aparición de valores negativos se debe con frecuencia a la presencia de ítems inversos no recodificados. En el caso de que se analicen ítems pertenecientes a distintas escalas, las correlaciones entre ítems pertenecientes a la misma escala serán mayores que las estimadas entre ítems pertenecientes a distintas escalas. La inspección de la matriz de correlaciones ofrece información relevante sobre la estructura interna del test. Sin embargo, a medida que aumenta al número de ítems es difícil extraer información significativa y clara; habría que aplicar sobre ella procedimientos y modelos más adecuados entre los que destaca el modelo de análisis factorial.

Para obtener la matriz de correlación entre ítems se seguirán las siguientes pautas:

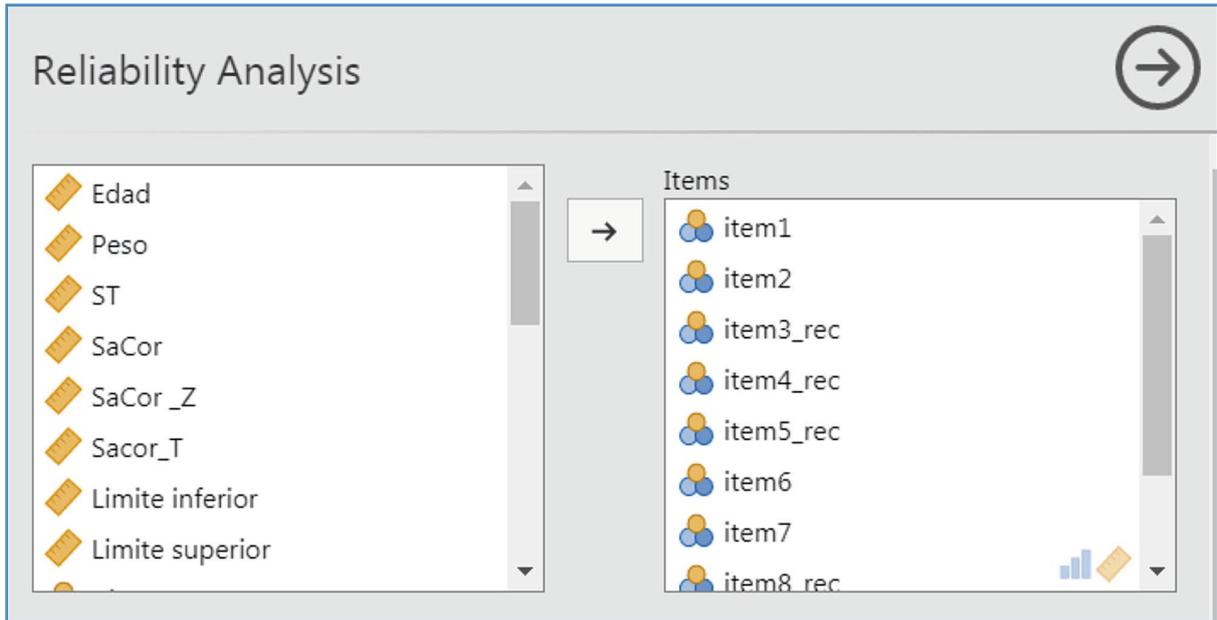
1. *Seleccionar la opción adecuada.* *jamovi* ofrece varias opciones para la obtención de la matriz de correlaciones, entre ellas, para esta ocasión utilizaremos la incluida dentro de la opción *Factor* y *Reliability Analysis* por ser especialmente atractiva.



7.10. IMAGEN

Matriz de correlaciones inter-ítem (1)

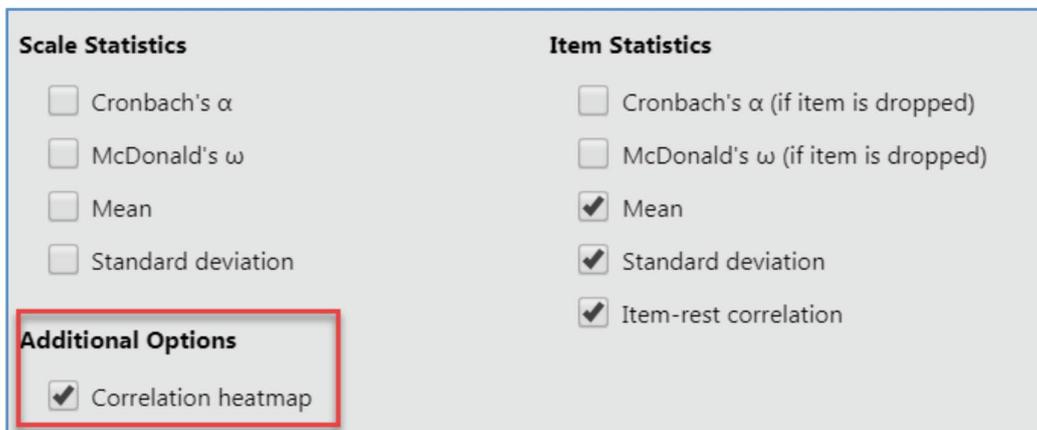
2. *Seleccionar variables.* Seleccionar los ítems cuya matriz de correlaciones se desea obtener.



7.11. IMAGEN

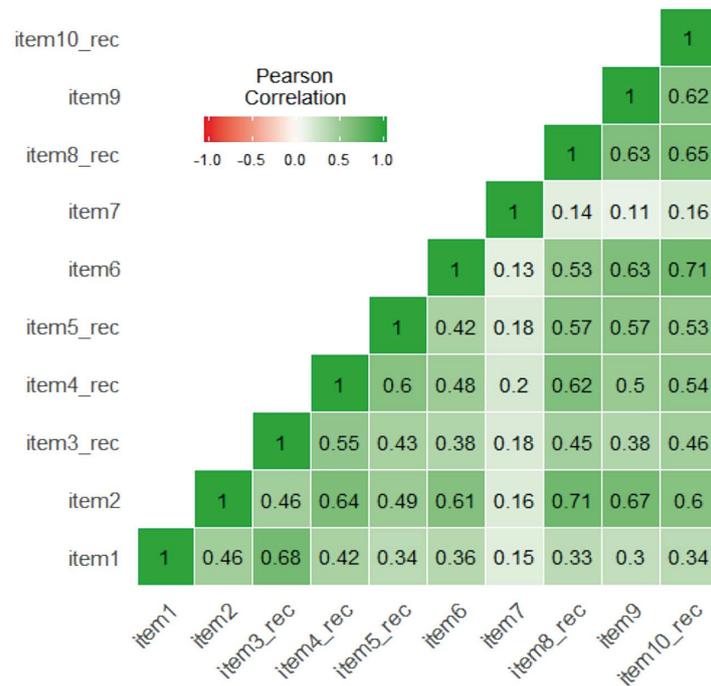
Matriz de correlaciones inter-ítem (2)

3. *Seleccionar el estadístico adecuado.* Se marcará la opción *Correlation heatmap* o mapa de calor.



7.12. IMAGEN

Matriz de correlaciones inter-ítem (3)



El mapa de calor (*heatmap*) ofrece además de los valores numéricos correspondientes a los coeficientes de correlación, información visual sobre la intensidad de la relación. Los colores más intensos indican correlaciones elevadas. A medida que la intensidad de la correlación se atenúa, los colores del mapa se van pastelizando.

8

Fiabilidad de la escala

En todo proceso de medida por el cual se asignan valores numéricos a ciertas propiedades de los objetos, interfieren los errores de medida. Los errores de medida afectan a todas las medidas; físicas, médicas, químicas o las medidas utilizadas en el marco de las ciencias sociales. Asumir la existencia de errores de medida es básico, y estimar su efecto sobre las puntuaciones es uno de los objetivos que persiguen los modelos psicométricos. Dentro de la teoría clásica de test el término que hace referencia a la precisión de la medida es la fiabilidad (De Boeck y Elosua, 2016).

La fiabilidad se refiere a la precisión de las puntuaciones, y su objetivo es estimar hasta qué punto los errores afectan al proceso de medida. Estos, son por definición aleatorios, no controlados y de su estimación se deriva el grado de fiabilidad de las puntuaciones del test.

Existen varios diseños y procedimientos para la estimación del coeficiente de fiabilidad. Entre ellos destacan tres; formas paralelas, test-retest y método de las mitades. Cada uno de ellos tiene ventajas e inconvenientes que es preciso conocer para una correcta interpretación del coeficiente de fiabilidad.

- *Formas paralelas*. La estimación del coeficiente de fiabilidad descansa en la construcción de dos formas paralelas de un mismo test. Una vez aplicadas a la misma muestra se estimará la correlación entre las puntuaciones. El coeficiente de fiabilidad así estimado indica el grado de equivalencia entre las formas paralelas.
- *Test-retest*. Asumiendo que un test es paralelo a sí mismo, la obtención del coeficiente de fiabilidad exige la administración en dos ocasiones diferentes del mismo test a la misma muestra. Una vez obtenidos los datos, la fiabilidad del test se obtendría por medio de una correlación entre puntuaciones. Este coeficiente de fiabilidad sería interpretado en el sentido de estabilidad temporal de las puntuaciones.
- *Método de las mitades*. Es el diseño más simple. El coeficiente de fiabilidad se estima a partir de los datos obtenidos en una única administración del test, y se interpreta en términos de consistencia interna.

Los valores del coeficiente de fiabilidad están comprendidos entre 0 y 1; en general se consideran aceptables valores superiores a .8 en contextos de investigación, y mayores a .9 cuando las puntuaciones son la base de inferencias o decisiones sobre personas individuales (Nunnally, 1978).

8.1. Coeficientes alfa y omega

Entre todos los procedimientos disponibles para la estimación del coeficiente de fiabilidad basados en una única aplicación del test se ha impuesto el coeficiente alfa de Cronbach (1951). Es un indicador del grado de covariación de los ítems de una escala o correlación intra-clase, y se interpreta como coeficiente de consistencia interna. Su extendido uso ha sido el origen de un debate que hoy sigue abierto (Viladrich, Angulo-Brunet y Doval, 2017).

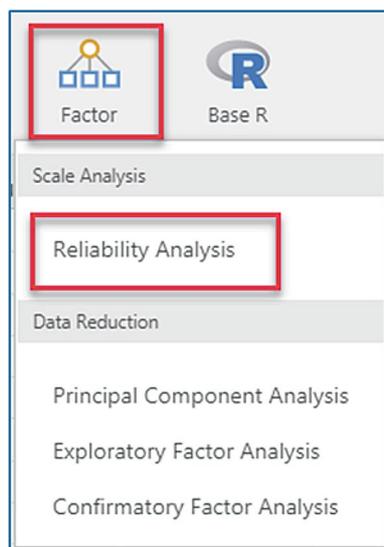
jamovi ofrece, además, el coeficiente omega de McDonald (1999). Este indicador se estima a partir de una solución factorial jerárquica (Trizano-Hermosilla y Alvarado, 2016; Revelle y Zinbarg, 2009) que *jamovi* implementa sobre un modelo factorial exploratorio unidimensional. En situaciones en que los datos se ajustan a un modelo unidimensional, las cargas factoriales de los ítems son iguales, y no existen errores correlacionados, es decir cuando las medidas son tau-equivalentes, ambos índices son similares.

Una correcta interpretación de estos coeficientes habría de considerar que:

- La estimación de alfa y omega no aporta información sobre la estabilidad de la medida. Son indicadores de consistencia interna, no de estabilidad temporal.
- Alfa y omega se definen formalmente como indicadores de consistencia para variables continuas. Existen versiones de estos índices para variables ordinales (Elosua y Zumbo, 2008). En la práctica podría asumirse que las escalas ordinales (Likert) sean tratadas como continuas a partir de 5 categorías de respuesta y distribuciones aproximadamente normales (Rhemtulla, Brosseau-Liard y Savalei, 2012).

Para obtener los indicadores de fiabilidad de una escala se seguirán las siguientes pautas:

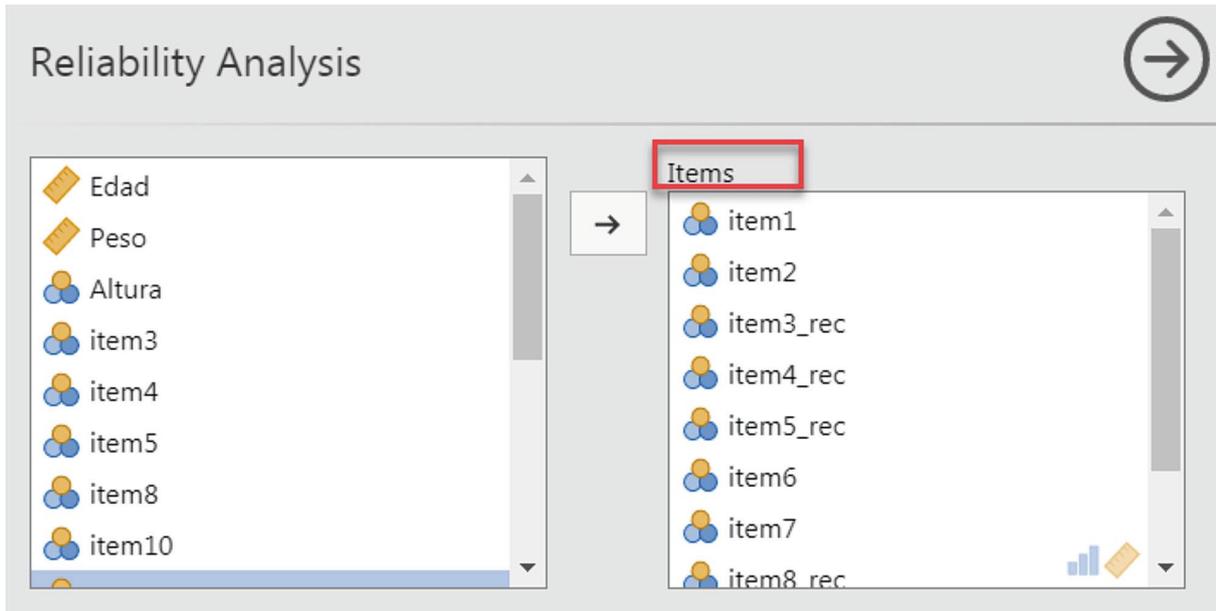
1. *Seleccionar la opción adecuada.* Continuamos con las opciones que encierra el menú *Factor y Reliability Analysis*.



8.1. IMAGEN

Coeficiente de fiabilidad (1)

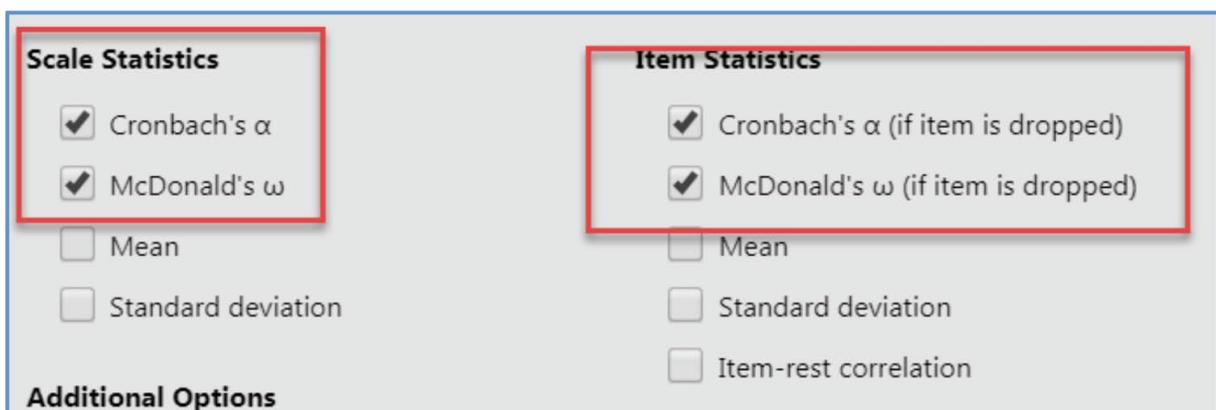
2. *Seleccionar variables.* Mover a la ventana *Items* los ítems que constituyen la escala que se desea analizar.



8.2. IMAGEN

Coefficiente de fiabilidad (2)

3. *Seleccionar el estadístico apropiado.* Para estimar los coeficientes de fiabilidad por los métodos alfa de Cronbach y omega de McDonald marcaremos las opciones correspondientes, tanto para la escala global (*Cronbach's α* y *C McDonald's ω*) como para cada uno de los ítems; α y (*Cronbach's α (if item is dropped)* y *McDonald's ω (if item is dropped)*).



8.3. IMAGEN

Coefficiente de fiabilidad (3)

Esas opciones generarán la siguiente salida:

Scale Reliability Statistics		
	Cronbach's α	McDonald's ω
scale	0.8929	0.8958

El valor del coeficiente alfa de Cronbach (Cronbach's α) es 0,89; valor que en este caso coincide con el coeficiente ω de *McDonald*. Además de estos índices generales referidos al conjunto de la escala, es interesante valorar la aportación de cada uno de los ítems a la fiabilidad total. Para ello *jamovi* ofrece la siguiente información:

Item Reliability Statistics		
	if item dropped	
	Cronbach's α	McDonald's ω
item1	0.8909	0.8939
item2	0.8724	0.8760
item3_rec	0.8837	0.8882
item4_rec	0.8781	0.8803
item5_rec	0.8818	0.8853
item6	0.8798	0.8834
item7	0.9049	0.9079
item8_rec	0.8752	0.8782
item9	0.8776	0.8811
item10_rec	0.8752	0.8788

En esta tabla puede comprobarse que aquellos ítems con valores más altos están inversamente relacionados con la fiabilidad de la escala. Si se eliminara el ítem 7 de la escala, el valor del coeficiente alfa de Cronbach se incrementaría de 0,8929 a 0,9049, y el valor de ω pasaría de 0,8958 a 0,9079.

8.2. Error típico de medida

El error típico de medida está estrechamente ligado a la fiabilidad del test, y es fundamental para la estimación del intervalo de confianza en torno a la puntuación verdadera. Se calcula mediante la fórmula:

$$S_e = S_X \sqrt{1 - r_{XX'}}$$

Donde, S_e es el error típico de medida,

S_X es la desviación típica de las puntuaciones observadas, y
 $r_{XX'}$ es el coeficiente de fiabilidad del test.

Esta ecuación se corresponde con la desviación típica de los errores de medida, y podría interpretarse como la desviación estándar de las puntuaciones empíricas de un sujeto en aplicaciones reiteradas de un test, es decir, su variación. A medida que aumenta la distribución de las puntuaciones empíricas disminuirá el coeficiente de fiabilidad del test.

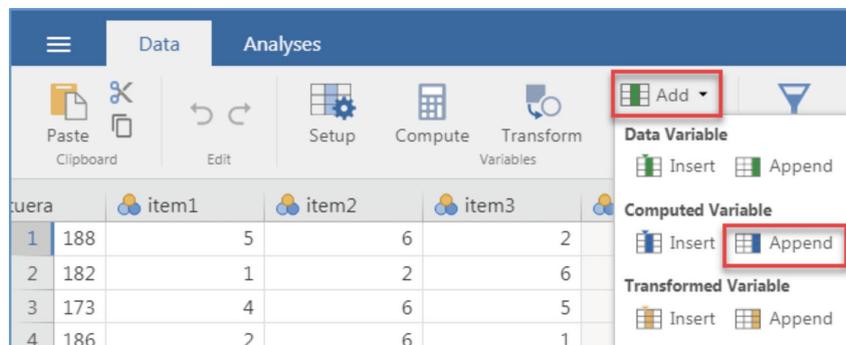
Para su estimación seguiremos la siguiente pauta:

1. Obtener la puntuación observada (puntuación total).
2. Calcular la desviación típica de la puntuación observada.
3. Estimar el error típico de medida.

8.2.1. Calcular la puntuación observada

La puntuación total de una persona es la suma de las puntuaciones en cada uno de los ítems que componen la escala. Desde un punto de vista operativo, su cálculo supone añadir una nueva variable a la hoja de datos que recoja este nuevo valor. En este ejemplo llamaremos, SaCor (satisfacción corporal) a esa nueva variable. Para calcularla seguiremos las siguientes pautas:

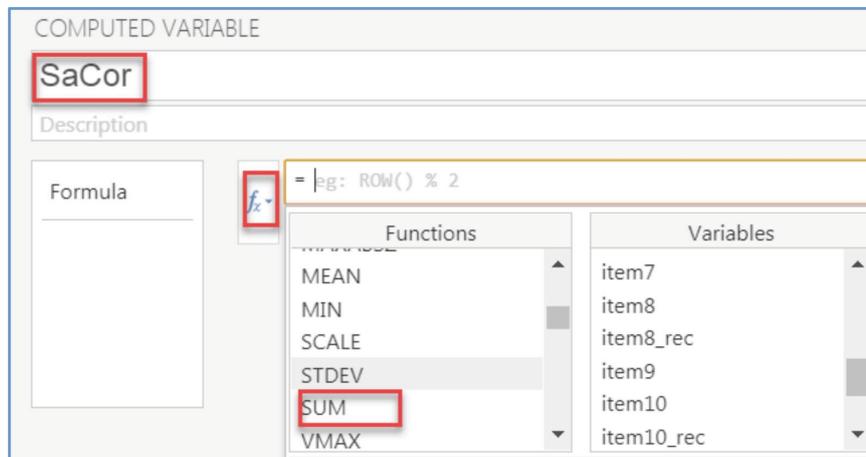
1. *Seleccionar la opción adecuada.* En el menú *Data* seleccionamos la opción *Add* y dentro de esta indicamos a *Jamovi* que queremos crear una nueva variable *Computed Variable*, y que la queremos añadir al final de la hoja de datos, *Append*. Esta elección crea una nueva variable en la hoja de datos que se añade al final de esta.



8.4. IMAGEN

Puntuación observada (1)

- Una vez creada la nueva variable a la que *jamovi* asignará un nombre, pulsamos sobre ella con el botón izquierdo del ratón para definir sus características.

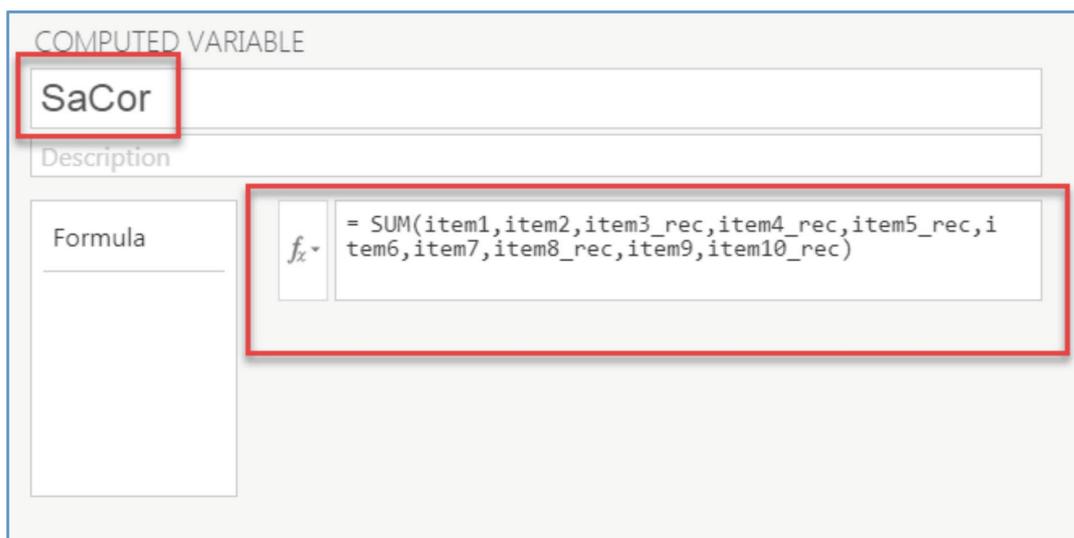


8.5. IMAGEN

Puntuación observada (2)

En esta ventana asignaremos un nombre (SaCor) a la variable e indicaremos a *jamovi* a través del menú que se abre tras presionar *fx* que queremos crearla por medio de una suma, SUM().

- Tras seleccionar la opción *SUM()*, señalaremos entre paréntesis y separados entre comas, los ítems cuyo valor queremos sumar para calcular la puntuación total. Para elegir un ítem basta marcarlo con el ratón presionando con el botón izquierdo sobre él (es posible introducirlo manualmente). En cualquier de los dos modos, es importante introducir en la fórmula los ítems recodificados.



8.6. IMAGEN

Puntuación observada (3)

Una vez especificada la formula y presionando la flecha con la que indicamos que hemos terminado con esta pantalla, aparecerá la nueva variable, SaCor, en la hoja de datos.

	item10_rec	ST	SaCor
1	5	508	52
2	1	480	17
3	6	504	51
4	4	494	51
5	3	510	47
6	6	515	50
7	6	506	52
8	4	493	39
9	4	511	47
10	4	505	48
11	6	506	55

8.7. IMAGEN

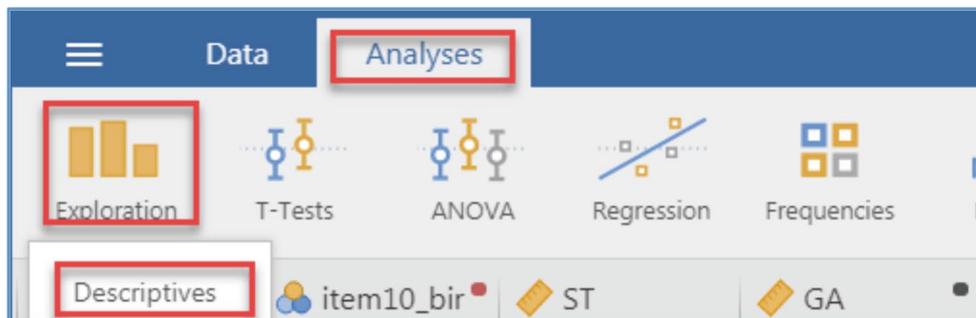
Puntuación observada (4)

NOTA. Incluir los ítems recodificados en caso de que existieran.

8.2.2. Desviación típica de las puntuaciones

Una vez calculada la variable que recoge la puntuación observada es necesario obtener su desviación típica. Para ello seguiremos las siguiente pautas:

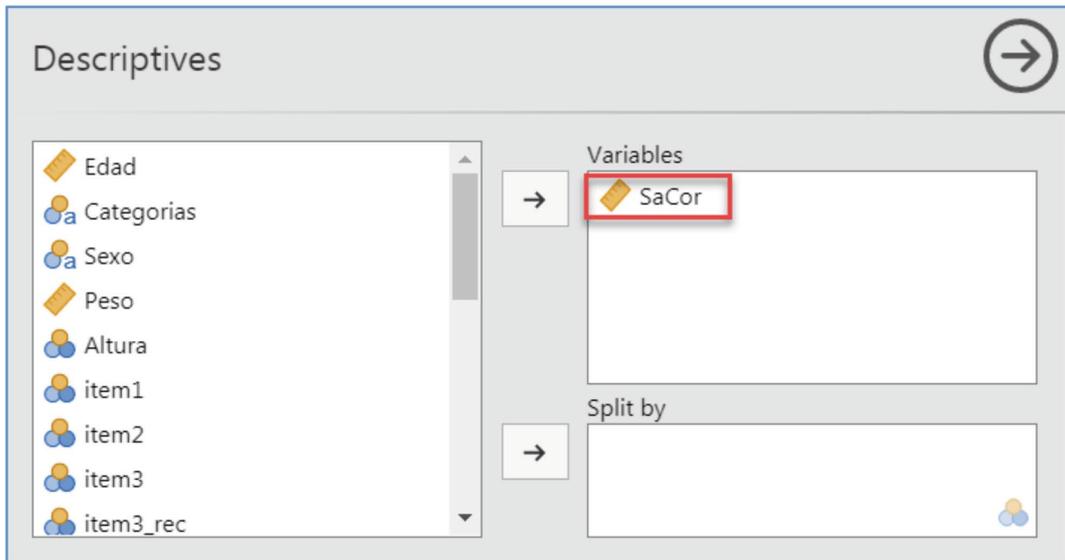
1. *Seleccionar la opción adecuada.* Dentro del menú *Analyses* seleccionar *Exploration* y *Descriptives*.



8.8. IMAGEN

Desviación típica de la puntuación observada (1)

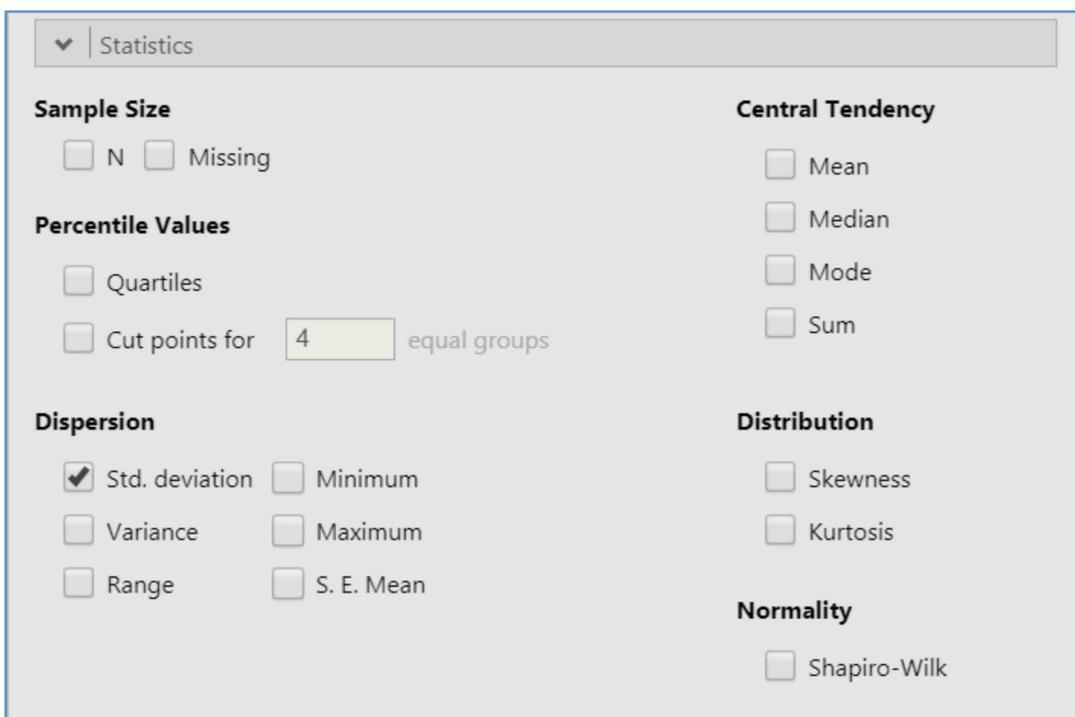
2. *Seleccionar variables.* En esta ventana seleccionaremos la variable cuya desviación queremos conocer, SaCor.



8.9. IMAGEN

Desviación típica de la puntuación observada (2)

3. *Seleccionar el estadístico apropiado.* Seleccionar la opción *Std. Deviation*.



8.10. IMAGEN

Desviación típica de la puntuación observada (3)

En la ventana de resultados podremos recoger el valor de la desviación típica de la variable SaCor; 11,68.

Descriptives	
	SaCor
Standard deviation	11.68

8.2.3. Estimar el error típico de medida

Para estimar el valor del error típico de medida, aplicaremos la fórmula anterior:

$$S_e = S_X \sqrt{1 - r_{XX'}} = 11,68 \sqrt{1 - 0,89} = 3,87$$

La desviación típica de la escala satisfacción corporal es 11,68, y el valor del coeficiente de fiabilidad estimado por medio del alfa de Cronbach es 0,89; el error de medida de la escala es 3,87.

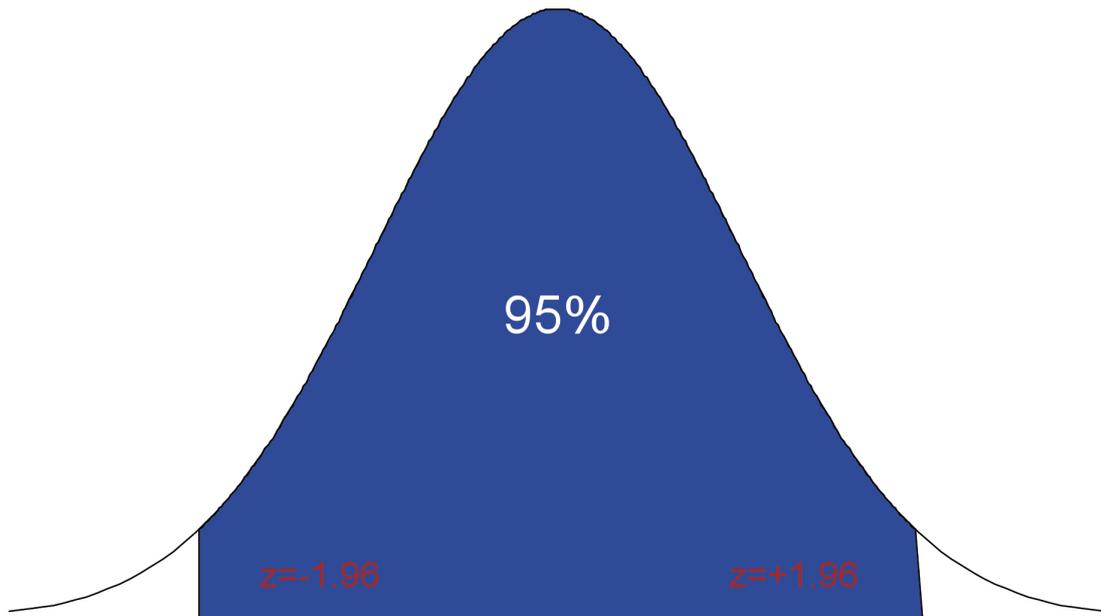
8.2.4. Intervalo de confianza en torno a la puntuación verdadera

Conocido el valor del error típico de medida se puede estimar con un grado de confiabilidad determinado el intervalo de confianza para la puntuación verdadera de los sujetos a partir de cualquier puntuación observada.

$$\text{Intervalo de confianza} = X_i \pm z_{\alpha/2} S_e$$

$$P\left(X_i - z_{\alpha/2} S_e \leq V \leq X_i + z_{\alpha/2} S_e\right) = 1 - \alpha$$

Suponiendo que el nivel de confianza es del 95% ($z_{\alpha/2} = \pm 1,96$), podemos estimar los límites inferior y superior del intervalo en torno a la puntuación verdadera para cada puntuación observada.

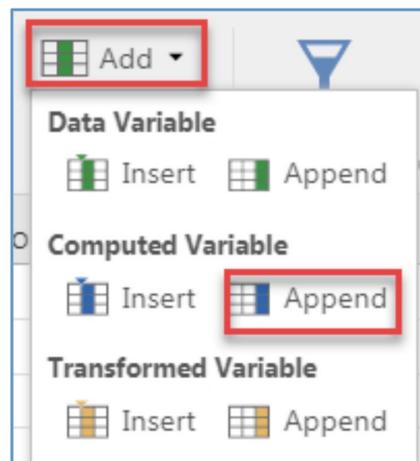


8.11. IMAGEN

Intervalo de confianza

Para estimar el límite superior del intervalo de confianza referido a la puntuación verdadera en la escala de satisfacción corporal se seguirán las siguientes pautas:

1. *Seleccionar la opción adecuada.* Indicar a *jamovi* que se desea generar una nueva variable.

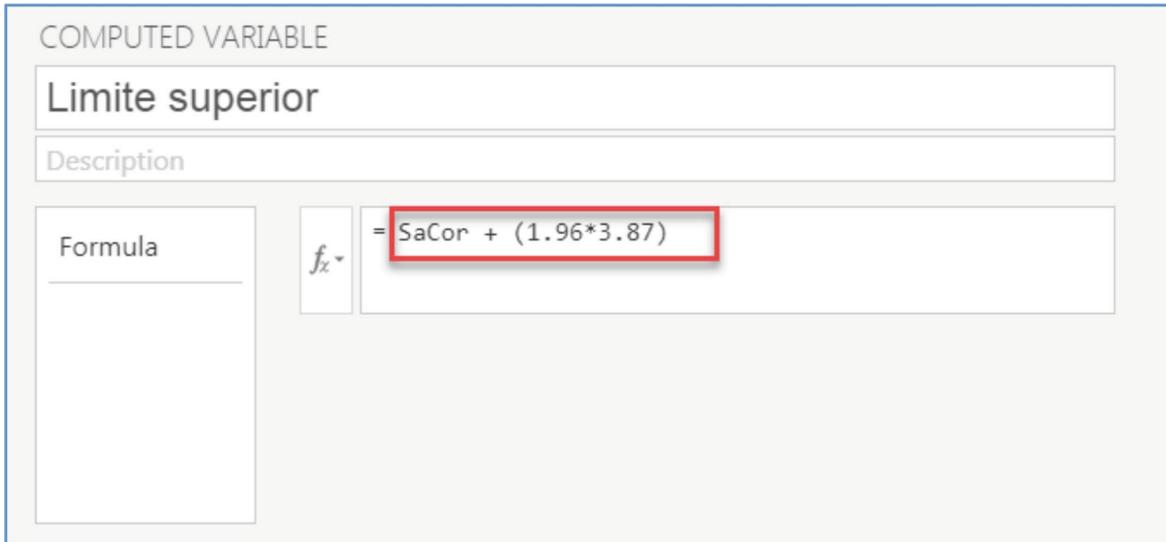


8.12. IMAGEN

Intervalo de confianza puntuación verdadera (1)

2. Especificar las características de la nueva variable indicando el nombre y la función utilizada para generarla. En este ejemplo la variable generada se llama límite superior, y en el espacio destinado a especificar la función que la generará indicaremos lo siguiente:

$$\text{Límite superior} = SaCor + (1,96 \times 3,87)$$



8.13. IMAGEN

Intervalo de confianza puntuación verdadera (2)

Se seguirá el mismo procedimiento para estimar el límite inferior. Para ello la función a utilizar sería:

$$\text{Límite inferior} = SaCor - (1,96 \times 3,87)$$

Estas pautas añadirán dos nuevas variables a la hoja de cálculo, que representan los límites inferior y superior del intervalo de confianza de la puntuación verdadera.

Limite inf...	Limite su...
44.415	59.585
9.415	24.585
43.415	58.585
43.415	58.585
39.415	54.585
42.415	57.585
44.415	59.585
31.415	46.585
39.415	54.585
40.415	55.585
47.415	62.585
31.415	46.585

8.14. IMAGEN

Intervalo de confianza puntuación verdadera (3)

9

Dimensionalidad. Análisis factorial exploratorio

El estudio de la evidencia de validez centrada en el análisis de la estructura interna evalúa el grado en que las relaciones entre los ítems y los componentes del test conforman el constructo que se quiere medir y sobre el que se basarán las interpretaciones. Según los últimos estándares de la APA, se centra en el estudio de la dimensionalidad de la prueba, y del funcionamiento diferencial de los ítems.

El estudio de la dimensionalidad cubre básicamente dos aspectos: 1) la verificación o refutación de la hipótesis de unidimensionalidad, y 2) la evaluación de la estructura multidimensional de los datos. Ambos objetivos implican la utilización de una metodología encaminada a determinar qué y cuántos factores explican la matriz de correlaciones, o determinan la respuesta de un sujeto.

Generalmente, el objetivo es especificar y demostrar la presencia de una «estructura simple» (Thurstone, 1947), o «simplicidad factorial» (Kaiser, 1974); definida como aquella que resulta de la agrupación de ítems en núcleos dimensionalmente homogéneos entre ellos, y dimensionalmente distintos entre sí, en la cual se espera que cada variable muestre una única saturación diferente de cero en un único factor. De entre todos los procedimientos utilizables para ello, el modelo de análisis factorial goza de mayor popularidad.

El análisis factorial engloba un conjunto de técnicas de análisis multivariadas (McDonald, 1985) cuyo objetivo es resumir la información contenida en un conjunto de variables observadas por medio de un número reducido de variables hipotéticas, conocidas normalmente como factores.

Es un procedimiento que puede utilizarse como técnica de análisis exploratoria (Análisis Factorial Exploratorio), o confirmatoria, para contrastar la hipotética estructura subyacente a un conjunto de datos (Análisis Factorial Confirmatorio). La tendencia reciente en el estudio del análisis factorial considera que el estudio de la dimensionalidad de los ítems puede situarse entre dos perspectivas que se sitúan en los polos de un continuo en función de las restricciones impuestas a los datos; exploratoria y confirmatoria. En la primera se obtendrá una descripción de la estructura interna de los datos, en la segunda se imponen restricciones al modelo para contrastar estadísticamente hipótesis referidas a la dimensionalidad. El contexto de investigación delimitará el tipo de acercamiento más apropiado; entre ambos extremos existe una tercera vía conocida como ESEM (*Exploratory Structural Equation Modeling*) (Ferrando y Lorenzo-Seva, 2014; Morin, Marsh y Nagengast, 2013).

En este ejemplo analizaremos la estructura interna de los ítems que componen la escala de satisfacción corporal. Para ello se seguirán las siguiente pautas:

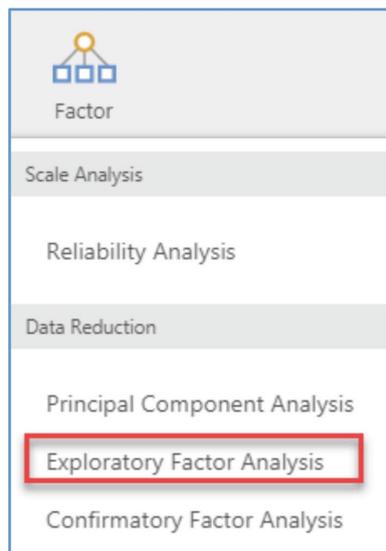
1. *Seleccionar la opción adecuada.* El análisis factorial se encuentra dentro del menú general de *Analyses* en la opción *Factor*.



9.1. IMAGEN

Análisis factorial (1)

jamovi ofrece tres opciones generales para estudiar la dimensionalidad de los datos: Análisis de componentes principales (*Principal Component Analysis*), análisis factorial exploratorio (*Exploratory Factor Analysis*) y análisis factorial confirmatorio (*Confirmatory Factor Analysis*). De entre ellas este curso trabajaremos con en el análisis factorial exploratorio.

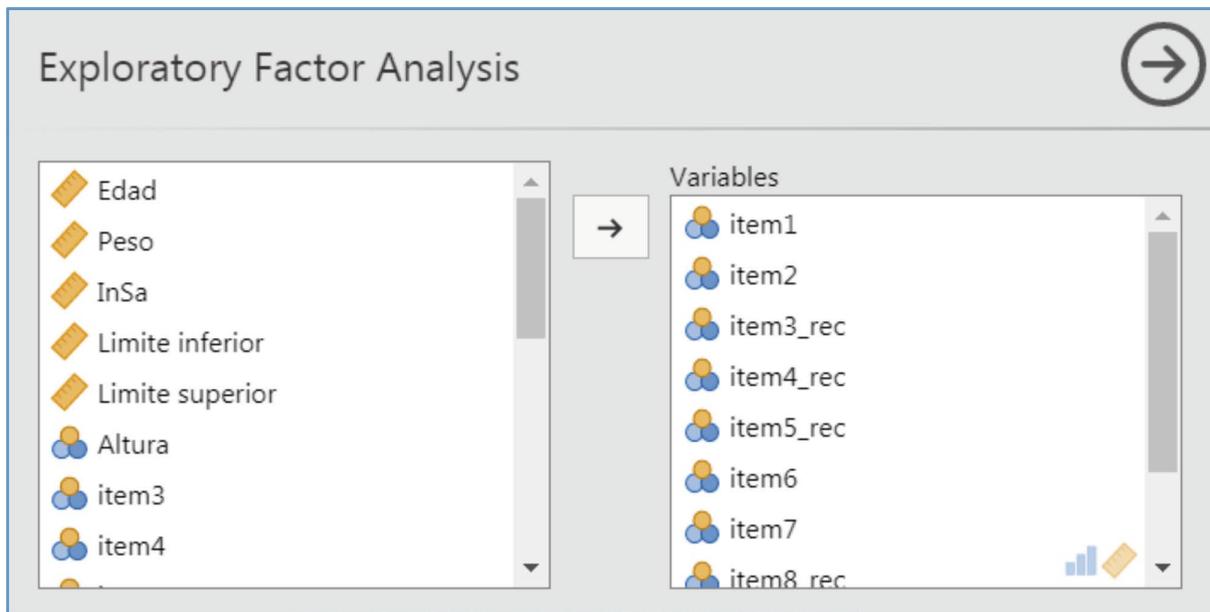


9.2. IMAGEN

Análisis factorial (2)

NOTA. Aunque, desde un punto de vista formal, el modelo de componentes principales no es un modelo de análisis factorial, su uso como tal es común en la investigación aplicada. De hecho, programas con licencias privativas incluyen el modelo de componentes principales como un método de extracción factorial. El objetivo del modelo de componentes principales es reproducir la varianza total mientras que el análisis factorial está interesado en la varianza común. Esa diferencia en los objetivos marca la diferencia entre los modelos. Es cierto que en la investigación aplicada se ha utilizado sobremana el modelo de componentes principales como técnica de evaluación de la dimensionalidad de los datos, sin embargo, es una utilización inapropiada del modelo.

2. *Selección de variables.* En la ventana correspondiente se seleccionarán las variables (ítems) objeto de análisis.



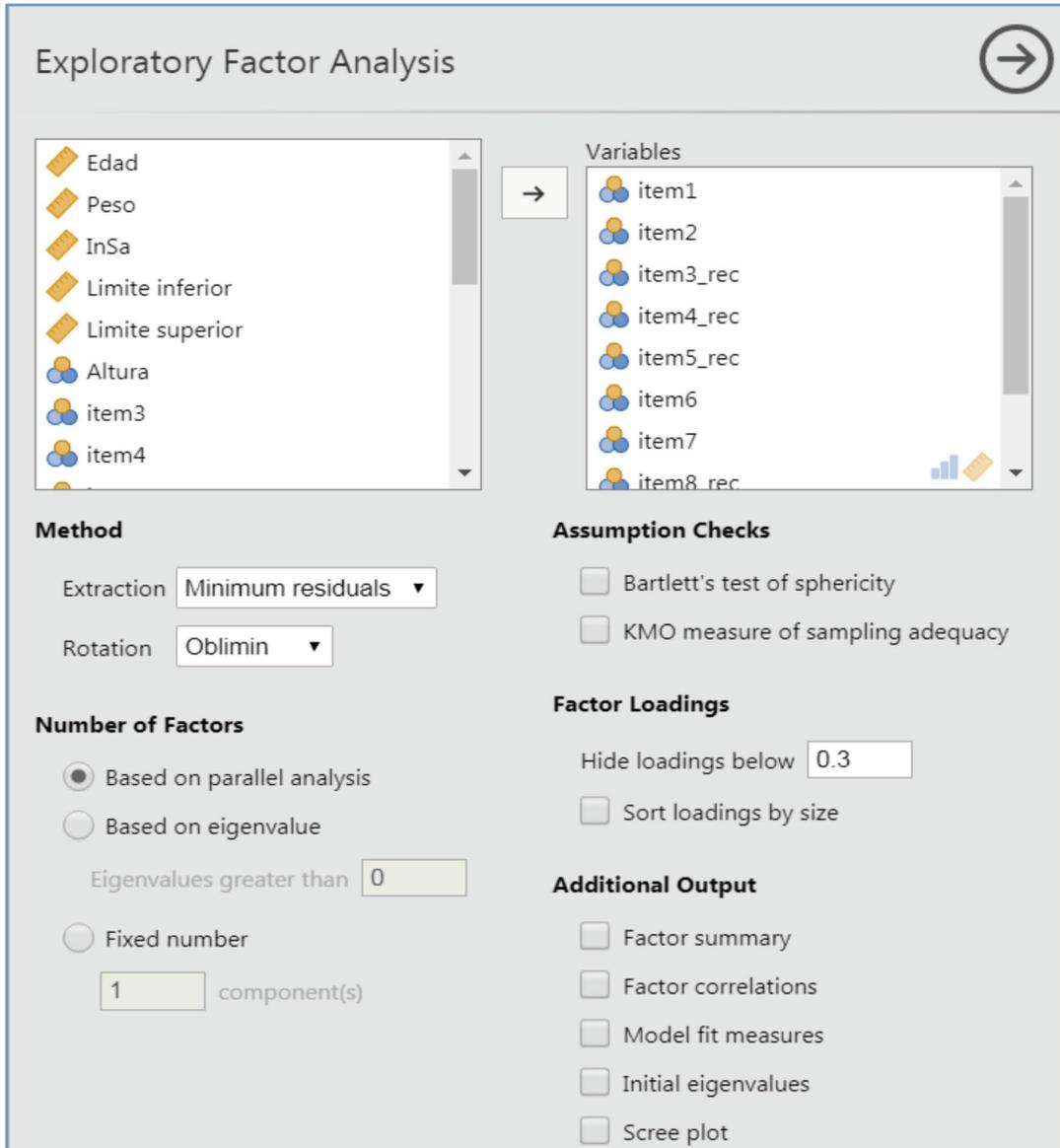
9.3. IMAGEN

Análisis factorial (3)

El proceso de análisis factorial consta de varias etapas que es conveniente conocer para optar en cada situación por la alternativa más adecuada.

1. *Supuestos.* Antes de la estimación de cualquier modelo estadístico es importante comprobar que los datos a analizar cumplen los supuestos exigidos por el modelo.
2. *Hipótesis.* El análisis factorial ha de ir guiado por un conocimiento de los datos a analizar que permita formular hipótesis sobre su estructura interna.
3. *Extracción factorial.* Viene determinada por la elección de un método de extracción que se fijará en función de las características de los datos.
4. *Rotación factorial.* En muchas ocasiones la primera solución factorial es difícil de interpretar y se opta por buscar una solución formalmente equivalente pero cuya interpretación sea más clara desde el punto de vista sustantivo.
5. *Adecuación de la solución.* La adecuación de la solución factorial dependerá en parte del grado en que los datos a analizar cumplen los supuestos del modelo factorial. A medida que aumenta la distancia entre los supuestos del modelo y las características de los datos la solución obtenida será más débil.
4. *Interpretación.* La interpretación de la solución factorial ha de aunar los aspectos formal y sustantivo del análisis. Las decisiones que se tomen relativas al número de factores a retener, su interpretación psicológica o la adecuación del modelo no pueden guiarse nunca por criterios exclusivamente numéricos; el análisis está siempre supeditado a la teoría.

En línea con estas etapas, *jamovi* ofrece las siguientes opciones:



9.4. IMAGEN

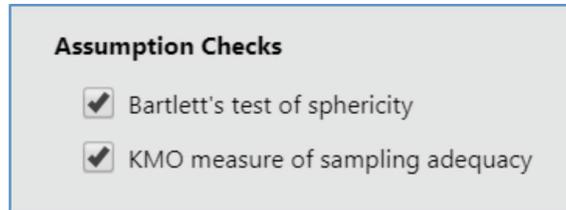
Análisis factorial (4)

9.1. Supuestos

Básicamente se utilizan dos indicadores para evaluar la matriz de correlaciones inter-ítem y obtener información sobre la pertinencia de someter los datos a un análisis factorial.

- *Bartlett's test of sphericity*. Prueba de esfericidad de Bartlett (1950): Contrasta la hipótesis nula de que la matriz de correlaciones es una matriz identidad. Si la prueba no rechaza la hipótesis nula, se cuestionaría la viabilidad del análisis factorial.

– *KMO measure of sampling adequacy*. Índice de Kaiser-Meyer-Olkin: Este índice contrasta los coeficientes de correlación parcial con los coeficientes de correlación observados. Valores bajos desaconsejarían la ejecución del análisis factorial. En la investigación aplicada se aceptan valores por encima de 0,70.



9.5. IMAGEN

Análisis factorial (5)

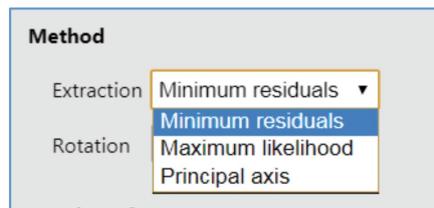
Los resultados del examen de los supuestos, muestran un índice KMO de 0,873, y se rechaza la hipótesis de que la matriz de correlaciones es una matriz identidad ($\chi^2(45) = 1924$; $p < .001$).

```
## Bartlett's Test of Sphericity
# -----
#  χ²   df  p
# -----
# 1024  45  < .001
# -----
```

```
# KMO Measure of Sampling Adequacy
# -----
#           MSA
# -----
# Overall      0.873
# item1        0.769
# item2        0.873
# item3_rec    0.814
# item4_rec    0.900
# item5_rec    0.893
# item6        0.868
# item7        0.939
# item8_rec    0.905
# item9        0.898
# item10_rec   0.887
# -----
```

9.2. Extracción factorial

jamovi ofrece tres procedimientos de extracción factorial: máxima verosimilitud, residuales mínimos, y ejes principales. El método de máxima verosimilitud es un procedimiento inferencial, adecuado y aconsejado para las situaciones en las que las variables se distribuyen de acuerdo a una distribución normal multivariada; si bien existe abundante literatura que muestra la robustez del modelo cuando las variables tiene una distribución univariada normal (Forero, Maydeu-Olivares y Gallardo-Pujol, 2009). Los procedimientos de ejes principales y residuales mínimos se englobarían dentro de los procedimientos conocidos como mínimos cuadrados ordinarios. Estos procedimientos generan soluciones similares (Ferrando y Anguano-Carraso, 2010), y se aconseja su uso para las situaciones de no-normalidad.



9.6. IMAGEN

Análisis factorial (6)

9.1. tabla

Métodos de extracción factorial

Procedimientos de extracción	
<i>Minimum residuals</i>	Residuales mínimos
<i>Maximum likelihood</i>	Máxima verosimilitud
<i>Principal axes</i>	Ejes principales

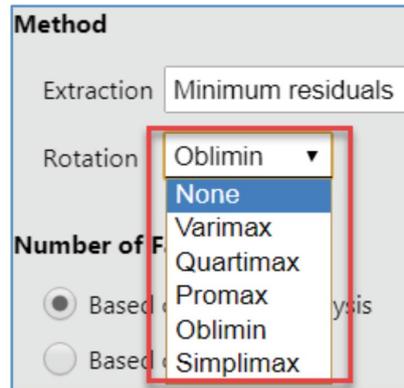
9.3. Rotación factorial

Una vez extraídos los factores, estos puede rotarse aproximando la solución a la estructura simple. *jamovi* dispone de varios procedimientos de rotación, tanto ortogonal como oblicua. El primer grupo genera factores independientes y el segundo asume la existencia de factores correlacionados. *Varimax* y *Quartimax* son rotaciones ortogonales, y el resto son rotaciones oblicuas.

9.2. tabla

Rotaciones factoriales

Rotación factorial	
Varimax Quartimax	Ortogonal
Promax Oblimin Sinplimax	Oblicua



9.7. IMAGEN

Análisis factorial (7)

- *None*. Ninguno: no se aplica ningún método de rotación.
- *Varimax*: formulada por Kaiser (1958), maximiza la varianza factorial minimizando el número de variables con altas saturaciones.
- *Quartimax*: simplifica la explicación de las variables minimizando el número de factores comunes asociados a ellas. Es recomendable cuando el número de variables observadas es elevado.
- *Oblimin*, *Promax* y *Simplimax*: son métodos de rotación oblicua de entre los cuales *simplimax* se aprecia como el más efectivo (Kiers, 1994; Ferrando y Anguiano-Carrasco, 2010).

9.4. Definición del número de factores

La determinación del número de factores a extraer puede fijarse por medio de tres opciones diferentes:

- *Based on parallel analyses*. Análisis paralelo. Selecciona los factores comunes que presentan valores propios mayores que los que se obtendrían por azar (Horn, 1965).
- *Based on eigenvalue*. Valores propios mayores de la unidad o Criterio de Kaiser-Guttman (Guttman, 1954; Kaiser, 1960). Estos autores equiparan el número de dimensiones con el número de factores que poseen valores propios mayores a la unidad. A pesar de ser un criterio muy utilizado en muchas ocasiones sobreestima el número de dimensiones.
- *Fixed number*. Determinación por parte del investigador. El investigador fija el número de factores a extraer.

9.8. IMAGEN

Análisis factorial (8)

9.5. Formato de salida de resultados

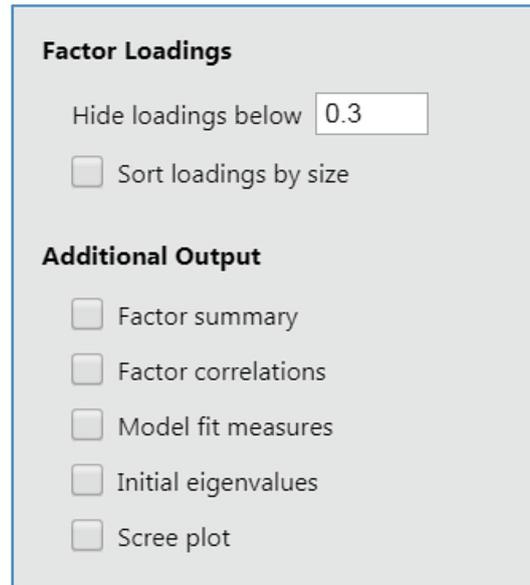
Además de especificar las características del proceso de extracción y rotación *jamovi* ofrece la posibilidad de fijar el formato de salida de resultados a través de las siguientes opciones.

- *Factor Loadings* permite definir el aspecto de la matriz factorial. Si no se desean mostrar las saturaciones menores a un valor fijo, habrá que indicarlo en el campo *Hide loadings below*.
- Si se desea que las saturaciones aparezcan ordenadas, se marcará la casilla *Sort loadings by size*.
- Las casillas de *Additional Output* permiten especificar:

9.3. tabla

Análisis factorial. Información adicional

Salidas	
<i>Factor summary</i>	Información sobre los factores
<i>Factor correlations</i>	Correlaciones entre factores
<i>Model fit measures</i>	Índices de ajuste del modelo
<i>Initial eigenvalues</i>	Valores propios iniciales
<i>Scree plot</i>	Gráfico de sedimentación de Catell



The image shows a settings panel for factor analysis. It is divided into two sections: 'Factor Loadings' and 'Additional Output'. In the 'Factor Loadings' section, there is a text input field labeled 'Hide loadings below' with the value '0.3' entered. Below it is a checkbox labeled 'Sort loadings by size'. The 'Additional Output' section contains five checkboxes, all of which are currently unchecked: 'Factor summary', 'Factor correlations', 'Model fit measures', 'Initial eigenvalues', and 'Scree plot'.

9.9. IMAGEN

Análisis factorial (9)**9.6. Ejemplo práctico**

En este ejemplo se analizarán dos soluciones factoriales diferentes; unidimensional y bidimensional.

9.6.1. Solución unidimensional

Indicaremos a *jamovi* lo siguiente:

- Número de factores a extraer: uno (no hay posibilidad de rotación).
- Procedimiento de estimación: ejes principales.
- Mostrar saturaciones mayores a 0,20.
- Orden de los ítems original.
- Información sobre el factor extraído.
- Gráfico de sedimentación o *Scree-plot*.

La salida muestra que el factor extraído explica un %47,8% de la varianza observada. Este valor se utiliza en la literatura psicométrica para apoyar la idea de existencia de un factor dominante. Reckase definió para ello un punto de corte del 20% (Hattie, 1985).

<p>Method</p> <p>Extraction <input type="text" value="Principal axis"/></p> <p>Rotation <input type="text" value="None"/></p> <p>Number of Factors</p> <p><input type="radio"/> Based on parallel analysis</p> <p><input type="radio"/> Based on eigenvalue</p> <p>Eigenvalues greater than <input type="text" value="0"/></p> <p><input checked="" type="radio"/> Fixed number</p> <p><input type="text" value="1"/> component(s)</p>	<p>Assumption Checks</p> <p><input checked="" type="checkbox"/> Bartlett's test of sphericity</p> <p><input checked="" type="checkbox"/> KMO measure of sampling adequacy</p> <p>Factor Loadings</p> <p>Hide loadings below <input type="text" value="0.2"/></p> <p><input type="checkbox"/> Sort loadings by size</p> <p>Additional Output</p> <p><input checked="" type="checkbox"/> Factor summary</p> <p><input type="checkbox"/> Factor correlations</p> <p><input type="checkbox"/> Model fit measures</p> <p><input type="checkbox"/> Initial eigenvalues</p> <p><input checked="" type="checkbox"/> Scree plot</p>
--	---

9.10. IMAGEN

Análisis factorial (10)

```
#
# Summary
# -----
# Factor  SS Loadings  % of Variance  Cumulative %
# -----
# 1      4.78         47.8          47.8
# -----
```

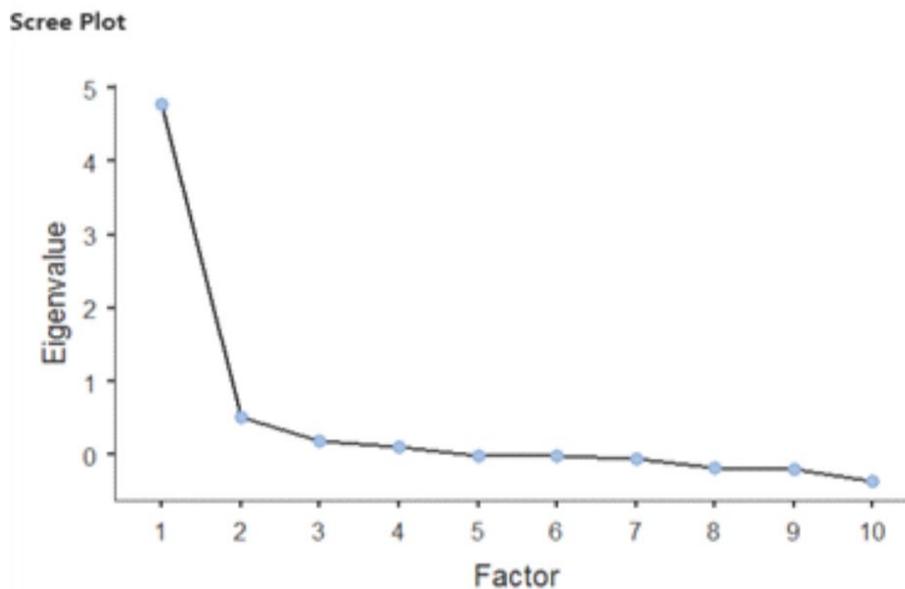
En la matriz factorial, las saturaciones factoriales representan la correlación entre cada ítem y el factor. Desde la perspectiva de la regresión lineal se trataría de coeficientes de regresión. La salida muestra que el séptimo ítem de la escala presenta un valor bajo (0,214), y por tanto una estimación de la unicidad alta, *Uniqueness* (0,954). El análisis de fiabilidad de la escala también reveló valores del índice de discriminación insuficientes para este ítem ($r_{iX} = 0,21$).

```

# Factor Loadings
# -----
#           1      Uniqueness
# -----
# item1      0.532    0.717
# item2      0.823    0.322
# item3_rec  0.629    0.604
# item4_rec  0.759    0.424
# item5_rec  0.685    0.531
# item6      0.718    0.484
# item7     0.214   0.954
# item8_rec  0.795    0.369
# item9      0.756    0.429
# item10_rec 0.784    0.386
# -----
# Note. 'Principal axis factoring' extraction method
# was used in combination with a 'none' rotation

```

El *screeplot* (gráfico de sedimentación) de Cattell (1966) es uno de los procedimientos gráficos más utilizados como apoyo en la determinación del número de factores a retener en un análisis de dimensionalidad. El gráfico ofrece una representación de la magnitud de los autovalores extraídos de la matriz de correlaciones. El punto de inflexión en la línea descendente se utiliza como regla para la determinación del número óptimo de factores a retener. La simplicidad y efectividad del procedimiento hace de él el método de decisión más utilizado en la investigación aplicada, si bien no es el más adecuado.



9.6.2. Solución bifactorial

En un estudio de dimensionalidad se exploran soluciones ofrecidas por diferentes configuraciones factoriales. En este ejemplo se ha estimado una solución bifactorial que se ha especificado del siguiente modo:

- Número de factores a extraer: dos.
- Procedimiento de estimación: ejes principales.
- Rotación: oblicua por el método *promax*.
- No ocultación de ninguna saturación (*Hide loadings below 0*).
- Información sobre los factores extraídos.
- Correlación entre los factores.

9.11. IMAGEN

Análisis factorial (11)

#	Factor	SS Loadings	% of Variance	Cumulative %
#	1	3.92	39.2	39.2
#	2	1.68	16.8	56.0

El primer factor explica un 39,2% de la varianza, y el segundo un 16,8%. Entre ambos explican una 56% de la variabilidad observada.

```
#
# Factor Loadings
# -----
#           1           2           Uniqueness
# -----
# item1    - 0.0980    0.81562    0.433
# item2     0.7553    0.09584    0.323
# item3_rec 0.0736    0.92643    0.228
# item4_rec 0.5182    0.30372    0.428
# item5_rec 0.5851    0.13064    0.538
# item6     0.7683   -0.04345    0.453
# item7     0.0859    0.15930    0.949
# item8_rec 0.8123   -0.00438    0.345
# item9     0.9004   -0.14766    0.346
# item10_rec 0.8202   -0.02518    0.354
# -----
# Note. 'Principal axis factoring' extraction method was used in combination with a
# 'promax' rotation
```

Con respecto a los pesos factoriales la matriz factorial revela que el primer ítem y el tercer ítem presentan cargas elevadas en el segundo factor; los ítems 2,6, 8, 9 y 10 definen claramente un primer factor. Es reseñable que el ítem 7 presenta correlaciones bajas con los dos factores extraídos. Los ítems 4 y 5 presentan saturaciones aceptables en el primer factor, si bien se podría revisar su contenido; «me siento», «me gusta».

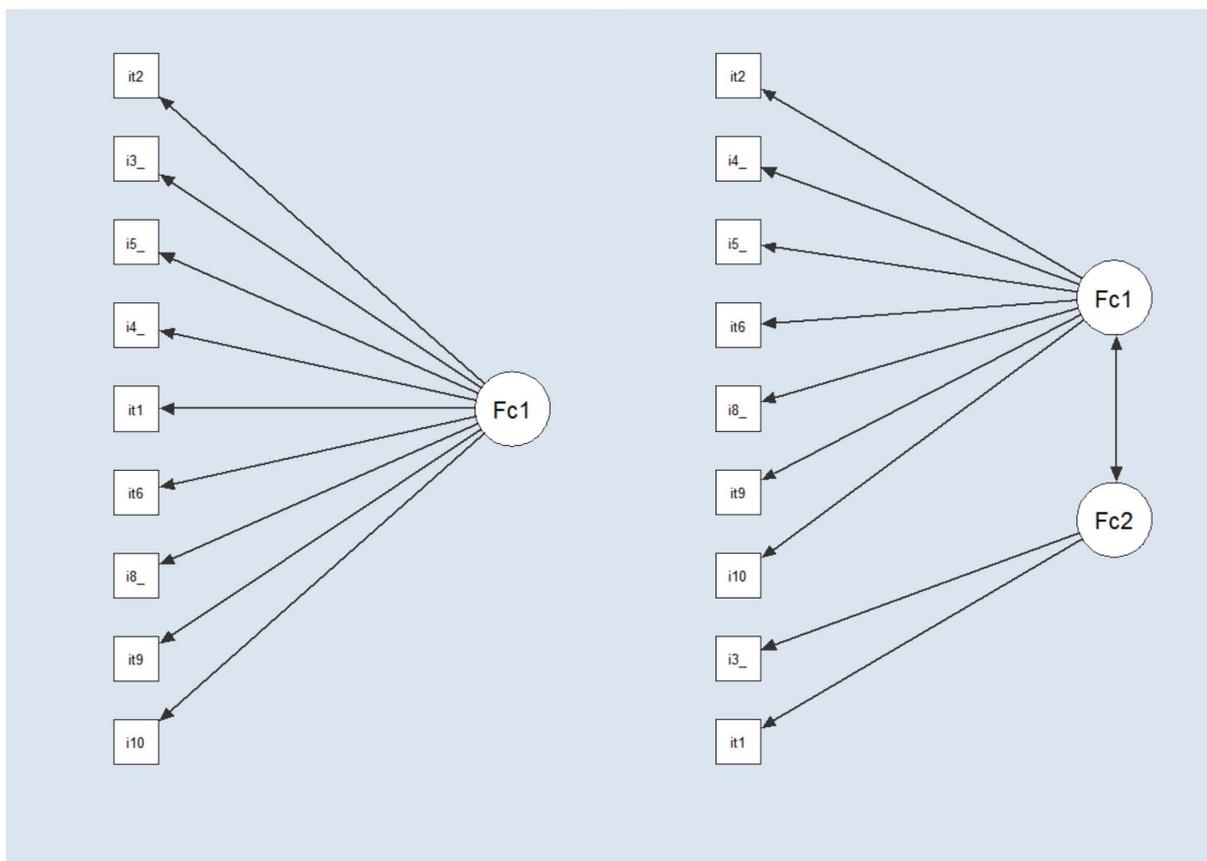
Finalmente *jamovi* ofrece información sobre la correlación entre los factores extraídos; es una correlación alta, 0,736.

```
#
# Correlation Matrix
# -----
#           1           2
# -----
# 1  —           0.736
# 2           —
# -----
```

9.7. Análisis de contenido. Definición sustantiva

Todo análisis factorial ha de complementarse con un análisis de contenido con relación al modelo teórico. En el ejemplo que nos ocupa es claro que el ítem 7 es el único ítem de la escala que no se refiere a una parte del cuerpo, separándose del contenido representado por el resto de ítems.

Las conclusiones a partir de las soluciones factoriales se acompañan de una análisis teórico, que puede complementarse con métodos factoriales confirmatorios que ayuden al investigador en la tarea de delimitar la dimensionalidad subyacente a los datos. Todas las consideraciones han de estar guiadas por un buen conocimiento teórico de las variables estudiadas; nunca se ha de anteponer la solución estadística a la solución teórica.



9.12. IMAGEN

Estructuras factoriales

10

Validez externa. Comparación de medias

El estudio de las relaciones entre las puntuaciones obtenidas tras la administración del test y otras variables externas, conocida como el aspecto externo de la validez por Loevinger (1957) y Messick (1989), o como amplitud nomológica por Embretson (1983, 1985), tal vez sea la fuente de información más utilizada en el proceso de validación de las puntuaciones. Su defensa como fuente de validez por el enfoque funcionalista (Prieto y Delgado, 1999) en la construcción de test ha avalado su uso desde los primeros estándares de la APA.

Esta fuente de información en el proceso de validación se centra en la búsqueda de pruebas que relacionen la puntuación con algún criterio que se espera prediga el test, o con otros test que hipotéticamente midan constructos similares, constructos relacionados o constructos diferentes (AERA, APA y NMCE, 2014). Los resultados de estos análisis permitirían evaluar el grado en que las relaciones hipotetizadas son consistentes con la interpretación propuesta.

Relaciones test-criterio. El objetivo es constatar la relación entre las puntuaciones obtenidas por el test y un criterio. Un criterio es un indicador de un constructo, de un diagnóstico, o de una conducta que el test intenta predecir. La bondad de los estudios centrados en las relaciones test/criterio dependerá de la calidad del criterio seleccionado y de la calidad de la medida que de él se haya tomado. El usuario habrá de valorar la calidad del criterio con referencia al propósito del test. La información reportada en el manual a modo de coeficiente de validez, es la correlación obtenida en una muestra representativa entre las puntuaciones en el test y la medida del criterio. A mayor coeficiente de validez mayor será la relación entre ambas medidas.

Validez predictiva. En determinados ámbitos del uso de test, como en el organizacional, en el que un test es usado para selección de personal, la información sobre la relación entre el test y un criterio ha de ampliarse con información referida a la precisión de la predicción. Para ello se tiene en cuenta el número de errores tanto por falsos positivos como por falsos negativos así como la razón de selección o porcentaje de candidatos que serán seleccionados. La cantidad de información aportada por una prueba más allá de la tasa base puede determinarse utilizando las tablas de Taylor-Russell. Las tablas muestran el porcentaje de solicitantes seleccionados que se espera tengan «éxito» en la situación de selección en función del coeficiente de validez, de la tasa base y de la razón de selección. El uso de las tablas supone una definición dicotómica del «éxito» en una situación de selección; se han elaborado enfoques similares para criterios continuos basados en la teoría de la decisión y en la teoría de la utilidad (Raju, Burke y Normand, 1990).

Validez discriminante. Clasificación. En la práctica clínica la información test/criterio tampoco es suficiente, y es necesario aportar información sobre la capacidad discriminatoria del instrumento, es decir, sobre su sensibilidad y especificidad. La sensibilidad se refiere a la proporción de casos diagnosticados como afirmativos, a partir del criterio o regla de decisión. La especificidad es la proporción de casos diagnosticados como negativos, a partir de la regla de decisión. La especificidad y sensibilidad para las distintas puntuaciones obtenidas por un cuestionario se pueden valorar a través de curvas ROC (*Receiver Operating Characteristic*, o Característica Operativa del Receptor). La curva ROC resume en un único gráfico sensibilidad y especificidad para todos los puntos de corte en una prueba, es decir, la capacidad predictiva para la detección de casos verdaderamente positivos.

Evidencia convergente y discriminante. Una de las formas más utilizadas en la recogida de información durante el proceso de validación de las puntuaciones consiste en evaluar su relación (correlación estadística) con medidas similares obtenidas por medio de otros test (validación convergente), o con medidas diferentes con las que teóricamente el test no presenta relación (validación discriminante). Las correlaciones convergentes habrían de ser altas, y las correlaciones discriminantes bajas.

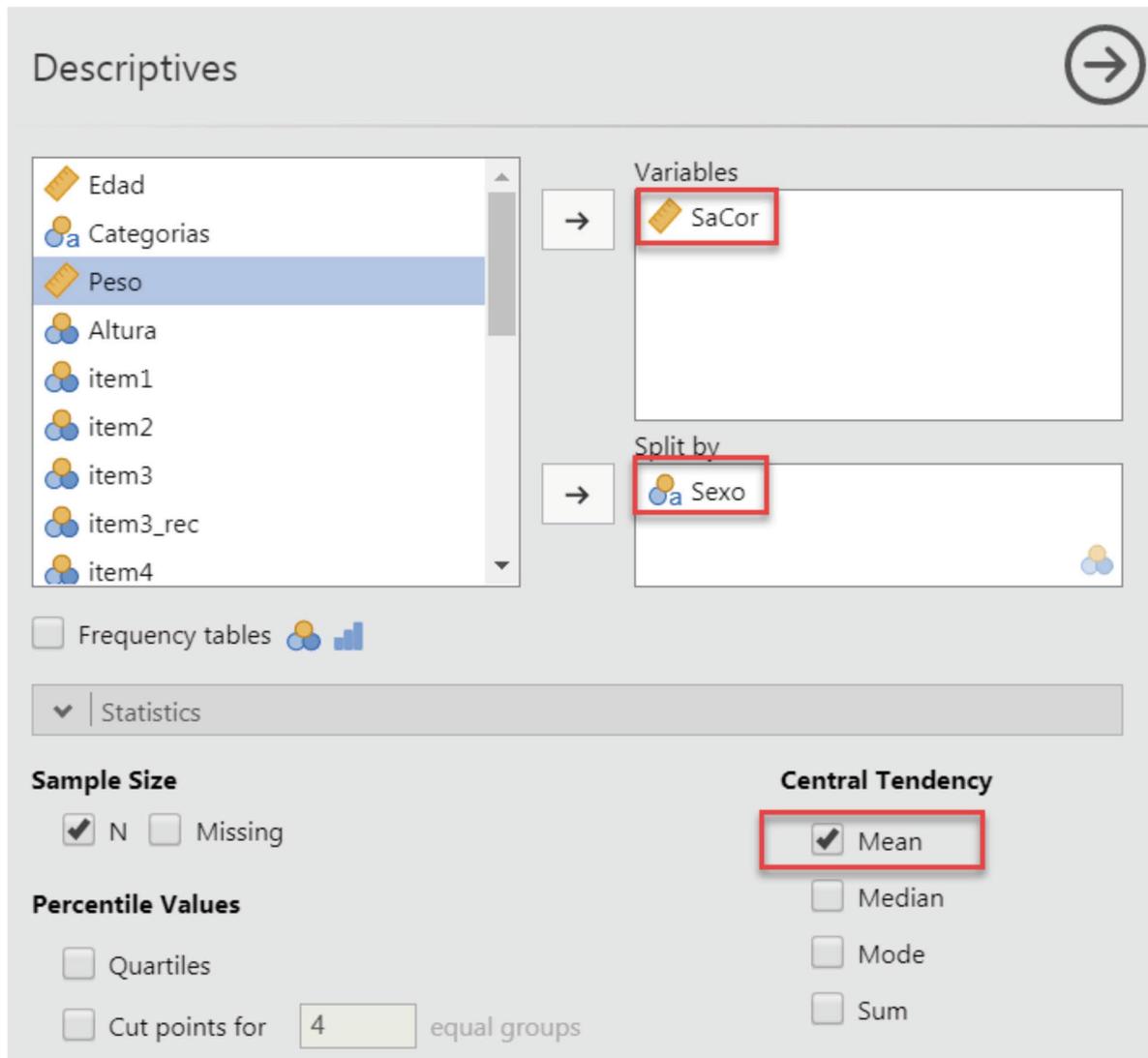
10.1. Comparar dos medias

Una de las fuentes de información más utilizada en el proceso de validación de puntuaciones es la comparación entre grupos. En ella el investigador no está únicamente interesado en describir sus datos (muestra); el objetivo es la generalización de resultados (inferencia) más allá de la muestra de estudio; busca extender las conclusiones derivadas de su estudio a la población de la que la muestra no es más que una representación. El «salto» de la muestra a la población se apoya en la estadística inferencial, que ha desarrollado multitud de técnicas con ese fin. La utilización de una u otra dependerá del diseño de investigación y de la naturaleza de los datos a analizar. La selección de la prueba más adecuada a cada caso es una decisión del investigador.

Descriptives		
	Sexo	SaCor
N	Mujer	155
	Hombre	37
Mean	Mujer	38.03
	Hombre	46.70
Standard deviation	Mujer	11.57
	Hombre	9.434

Para contestar a la pregunta de si existen diferencias significativas en el grado de satisfacción corporal en función del sexo, o más específicamente, si este es superior en las mujeres que en los

varones, partimos de las medias en la escala de satisfacción corporal en el grupo de hombres y en el grupo de mujeres: 38,03 y 46,70. La diferencia es de 8,67 puntos a favor de los varones (*Analyses > Exploration > Descriptives*). Ahora bien, el objeto de nuestra pesquisa no es la diferencia entre los hombres y las mujeres en nuestra muestra concreta; queremos saber si existen diferencias entre las poblaciones de mujeres y hombres, y por ello necesitaremos aplicar una prueba de significación estadística.



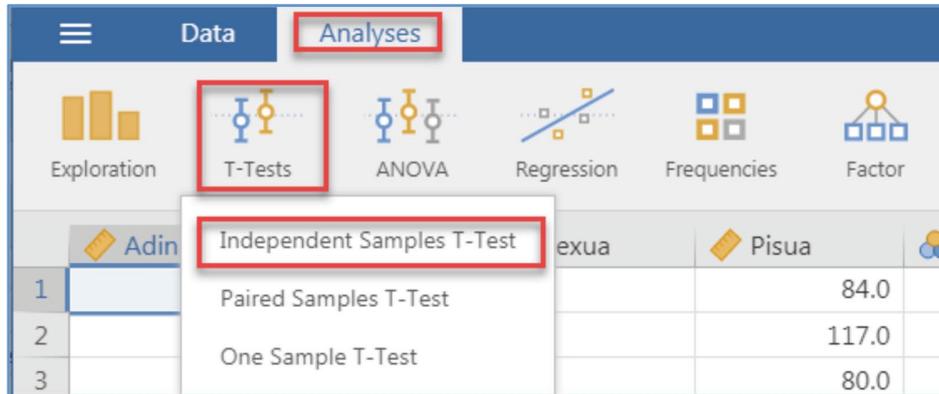
10.1. IMAGEN

Satisfacción corporal en función del sexo

El test clásico para la comparación entre dos medias es la prueba t de Student. A partir del valor de t se obtendrá el «grado de significación», es decir la probabilidad asociada con el valor t. El grado de significación (p) evalúa la probabilidad de la diferencia observada en los datos en el caso de que la hipótesis nula fuese cierta, es decir, si efectivamente las dos poblaciones de hombres y mujeres tuvieran medias idénticas en la escala de satisfacción corporal.

Los pasos a seguir para llevar a cabo una comparación de medias serían:

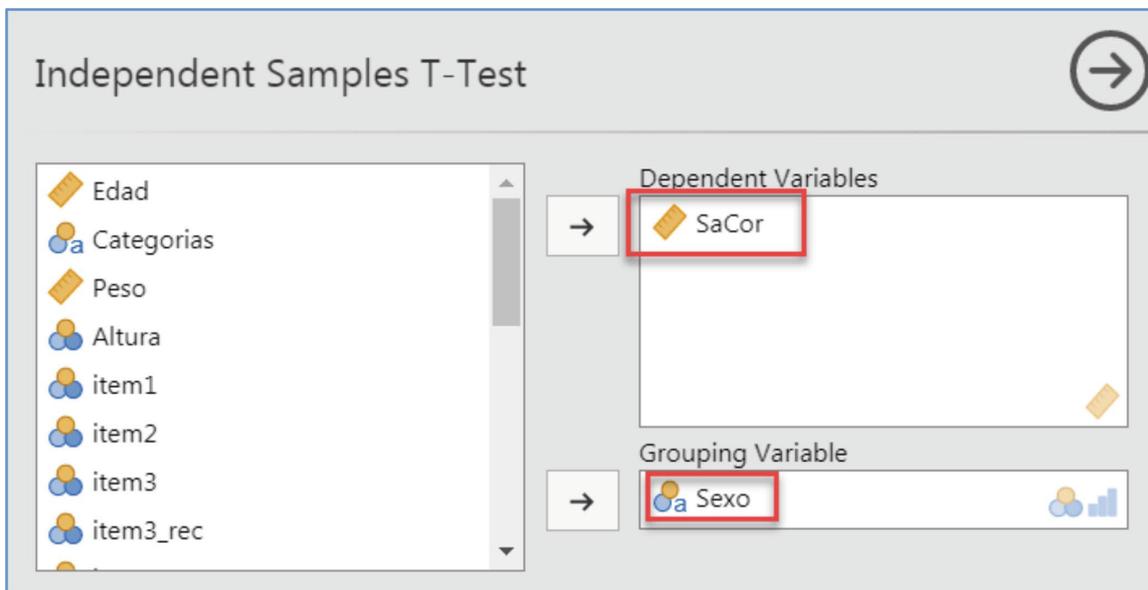
1. *Seleccionar el procedimiento adecuado.* La comparación de medias se encuentra dentro del menú *Analyses*, *T-Test* y dentro de esta en la opción *Independent Samples T-test*.



10.2. IMAGEN

t.test (1)

2. *Seleccionar variables.* En esta ventana se seleccionará la variable que representa la satisfacción corporal (SaCor) y se moverá a la ventana *Dependent Variables*. En la ventana titulada *Grouping Variable* se colocará la variable Sexo.



10.3. IMAGEN

t.test (2)

3. *Supuestos*. *jamovi* ofrece varias opciones para llevar a cabo una comparación de medias, para seleccionar la óptima es importante evaluar el cumplimiento de los supuestos de modelo (*Assumption Checks*); marcando las opciones *Normality* y *Equality of Variances* obtenemos lo siguiente:

Assumption Checks

Normality (Shapiro-Wilk)

Normality (Q-Q plot)

Equality of variances

10.4. IMAGEN

t-test (3)

Assumptions				
Test of Normality (Shapiro-Wilk)				
		W		p
SaCor		0.979		0.005
Note. A low p-value suggests a violation of the assumption of normality				
Test of Equality of Variances (Levene's)				
	F	df	df2	p
SaCor	5.07	1	190	0.025
Note. A low p-value suggests a violation of the assumption of equal variances				

La variable SaCor no se distribuye de acuerdo a la ley normal ($W = 0.979$; $p = .005$), y las varianzas en las muestras de mujeres y de hombres no pueden considerarse equivalentes ($F(1,190) = 5.07$; $p = 0.025$).

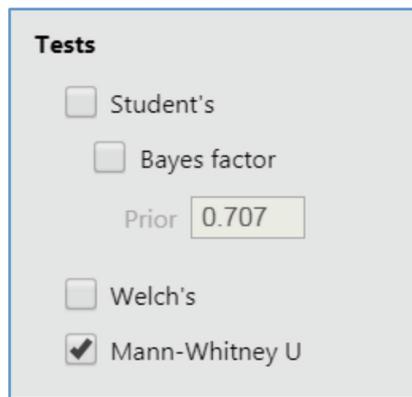
4. *Seleccionar el estadístico apropiado.* jamovi ofrece tres pruebas diferentes: t de Student, prueba de Welch y Mann-Whitney U. La tabla siguiente muestra las diferencias entre ellas:

10.1. tabla

Pruebas estadísticas comparación de medias

Prueba	Tipo	Supuestos
Mann Whitney	No paramétrica	Sin asunciones
Welch	Paramétrica	Distribución normal Varianzas diferentes
Student	Paramétrica	Distribución normal Varianzas iguales

Dado que los datos muestrales no cumplen los supuestos indicados, la prueba más adecuada para este caso es la U de Mann-Whitney.



10.5. IMAGEN

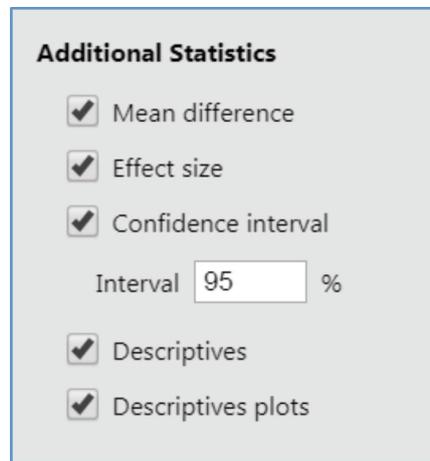
t-test (3)

Independent Samples T-Test			
		statistic	p
SaCor	Mann-Whitney U	1559	< .001

Según los resultados los valores de satisfacción corporal son estadísticamente diferentes para hombres y para mujeres ($U = 1559$; $p < .001$).

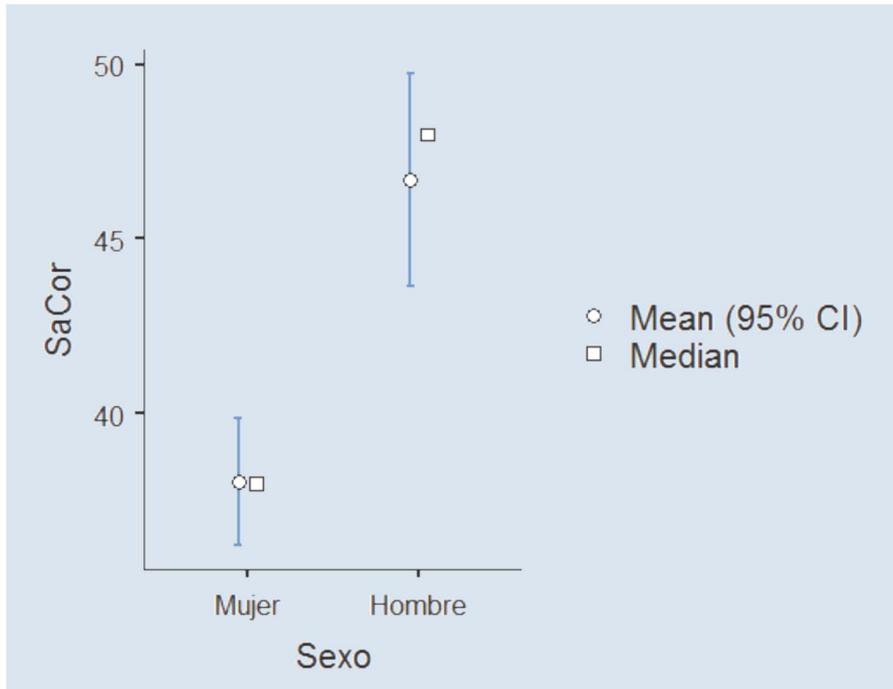
5. *Información adicional.* En la ventana correspondiente a la prueba *T-test*-marcando las opciones *Descriptives* y *Descriptive plots*, *jamovi* mostrará las medias aritméticas de esta variable para cada grupo, su representación gráfica, el intervalo de confianza en torno a la diferencia, y una estimación del tamaño del efecto que es menester incluir junto a toda prueba estadística. Al respecto citamos la siguiente recomendación de la APA que indica que las pruebas de significación sobre la hipótesis nula no son más que el punto inicial en la comunicación de resultados:

«Para interpretar el alcance de los resultados es necesario ofrecer información sobre los tamaños del efecto, los intervalos de confianza ... Uno de los requerimientos mínimos de las publicaciones periódicas de la APA es la presentación completa de las hipótesis examinadas, y de las estimaciones de tamaños del efecto e intervalos de confianza adecuados (APA, 2010, p. 33).»



10.6. IMAGEN
t-test (4)

		statistic	p	Mean difference	SE difference	95% Confidence Interval		Cohen's d
						Lower	Upper	
SaCor	Mann-Whitney U	1559	< .001	-9.00		-13.0	-5.00	-0.775
Group Descriptives								
	Group	N	Mean	Median	SD	SE		
SaCor	Mujer	155	38.0	38.0	11.6	0.929		
	Hombre	37	46.7	48.0	9.43	1.55		



Para la interpretación del tamaño del efecto es frecuente utilizar los valores que aparecen en la siguiente tabla (Cohen, 1988; Sawilowsky, 2009), si bien para su correcta interpretación es importante considerar el contexto de investigación:

10.2. tabla

Interpretación de la d de Cohen

Tamaño	d
Muy pequeño	0,01
Pequeño	0,20
Medio	0,50
Grande	0,80
Muy grande	1,20

En este ejemplo el valor de la d de Cohen es $-0,775$; un valor mediano.

10.2. Comparar más de dos medias

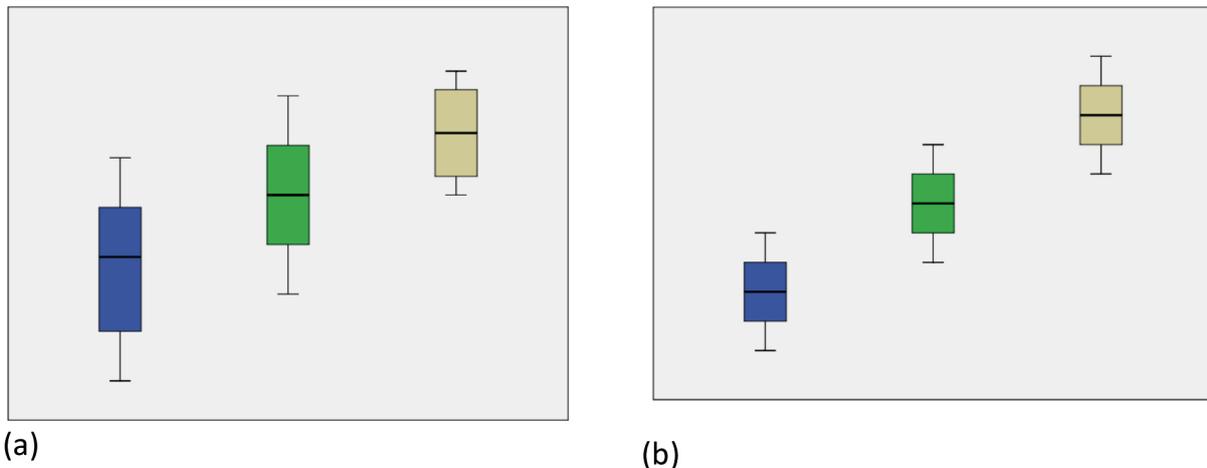
Cuando se precisa la comparación entre más de dos grupos se podrían tomar todos los pares de comparaciones y aplicar a cada una de ellas el test *t*. Sin embargo, este proceder conduciría a un incremento grave del error Tipo I. El análisis de varianza (ANOVA o AVAR) permitirá probar

la significatividad de las diferencias entre varias medias sin que se incremente el error de tipo I. El método ANOVA básicamente consiste en dividir la varianza de la variable dependiente en dos o más componentes, cada uno de los cuales puede ser atribuido a una fuente identificable.

Supongamos que nos interesa averiguar si los tres diferentes grupos de edad (adolescentes, jóvenes, adultos) obtienen medias similares en la variable satisfacción corporal, o si por el contrario las medias son diferentes. A la variable independiente (Categorías) en el ANOVA se denomina factor (¡no confundirlo con el «factor» del análisis factorial!) y en este ejemplo tiene tres categorías o niveles. El factor puede ser una variable propiamente categórica, o puede haber sido categorizada; como en este caso. La variable dependiente, satisfacción corporal (SaCor) es continua.

Aún cuando la hipótesis nula sea cierta y todas las medias poblacionales sea cual sea la edad fuesen iguales, esperaríamos una pequeña fluctuación de las medias muestrales de satisfacción corporal. ¿Cuánto se espera que varíen estas medias? Pues bien, a partir de la variación de los datos, podemos tener una idea de la variación entre las medias. Si las medias poblacionales fuesen en realidad diferentes, esperaríamos que esta variación fuese aún mayor. Por ello necesitamos un test de hipótesis con el que se pueda verificar si la variación entre las medias es mayor que la esperada en el caso de meras fluctuaciones aleatorias. Se precisa un nuevo modelo de distribución muestral, denominada distribución F.

Para aclarar como funciona este test examinemos la siguiente imagen que contiene dos conjuntos de diagramas de cajas. Se trata de decidir si las medias muestrales son suficientemente dispares como para rechazar la hipótesis nula.



10.7. IMAGEN

Diferentes variabilidades intragrupo

En el gráfico de la derecha (b), es difícil admitir que las diferencias entre las muestras se deban únicamente a la variabilidad muestral. En cambio en (a) es más fácil deducir que estas observaciones podrían haberse dado a partir de grupos con las mismas medias. Los dos conjuntos (a) y (b) tienen la misma diferencia entre los centros de los grupos. Sin embargo, lo que los diferencia es la variación dentro de cada grupo. En el segundo es tan pequeña que las diferencias entre los grupos se evidencian. Pues bien, esta es la idea central del test F. Cuando las diferencias entre las medias son grandes comparadas con la variación dentro de los grupos, rechazaremos la hipótesis

nula y concluiremos que las medias no son iguales. Sin embargo, en el gráfico (a) las diferencias entre las medias hacen pensar que bien podrían haberse obtenido por azar a partir de grupos con medias iguales, por ello no hay evidencia suficiente como para rechazar la hipótesis nula (H_0).

Ahora bien, se necesitará una prueba estadística que permita hacer esta comparación de modo más preciso. Para comparar las medias de más de dos grupos se seguirán las siguientes pautas:

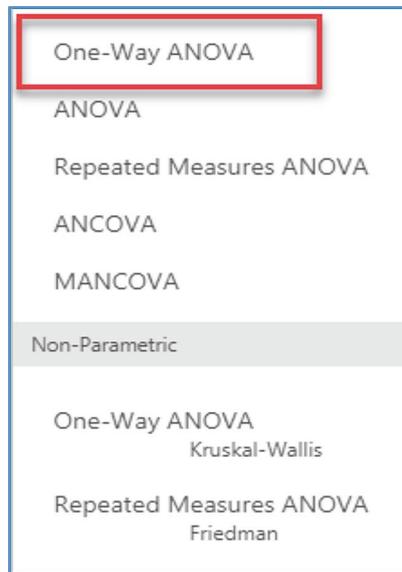
1. *Seleccionar el procedimiento adecuado.* ANOVA se encuentra dentro del menú *Analyses*.



10.8. IMAGEN

ANOVA (1)

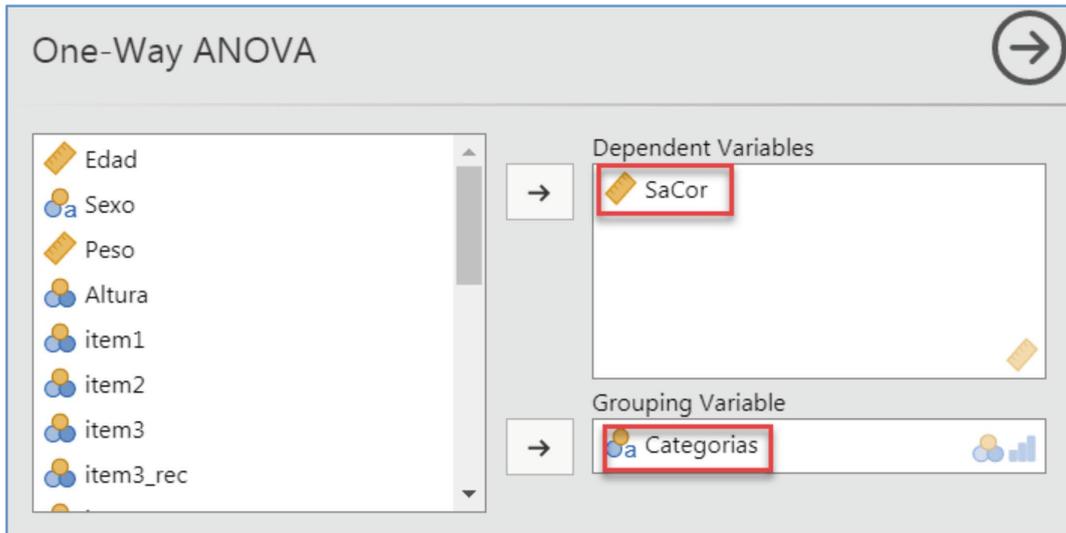
Entre las opciones ofrecidas por *jamovi* seleccionaremos One-Way ANOVA, que ofrece una versión simplificada de la prueba ANOVA.



10.9. IMAGEN

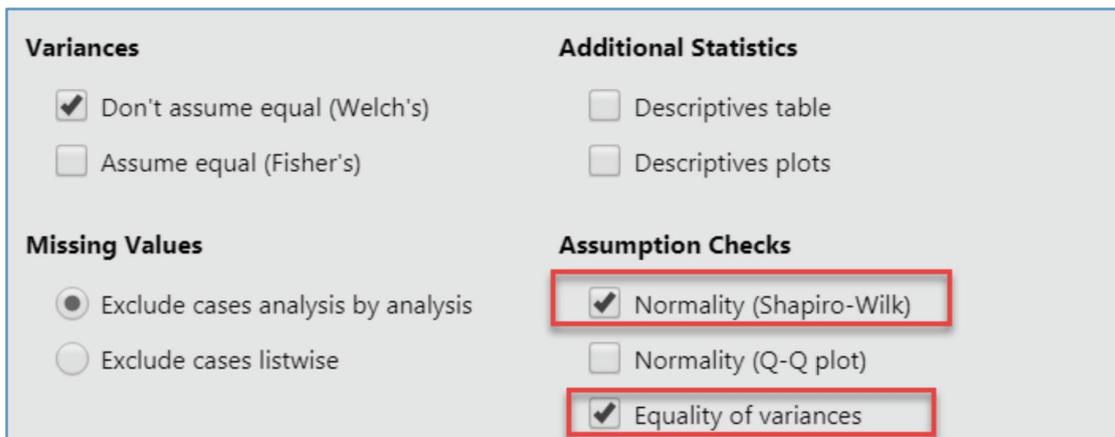
ANOVA (2)

2. *Seleccionar variables.* Queremos comprobar si la variable satisfacción corporal cambia con la edad. La variable dependiente (*Dependent Variables*) es SaCor, y la variable categórica que definirá los grupos es Categorías (*Grouping Variable*).



10.10. IMAGEN
ANOVA (3)

3. *Supuestos*. Para la selección de la prueba estadística más adecuada es conveniente analizar si las varianzas de los grupos son iguales, y las distribuciones son normales.



10.11. IMAGEN
ANOVA(4)

Los resultados muestran que las varianzas pueden considerarse equivalentes ($F(2,289) = 0,761; p = 0,469$).

# Test for Equality of Variances (Levene's)				
#	F	df1	df2	p
# SaCor	0.761	2	189	0.469

Sin embargo, la distribución de la variable satisfacción corporal no cumple el criterio de normalidad.

Test of Normality (Shapiro-Wilk)		
	W	p
SaCor	0.967	0.00015

Note. A low p-value suggests a violation of the assumption of normality

4. *Seleccionar el estadístico apropiado.* De acuerdo a los resultados anteriores se deduce que la prueba más adecuada en la prueba de Welch, que no asume varianzas iguales.

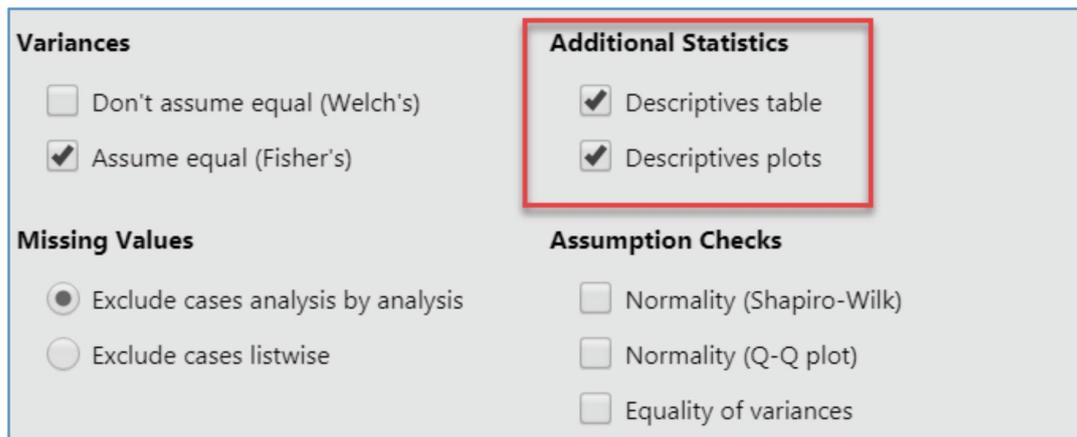
<p>Variaciones</p> <p><input checked="" type="checkbox"/> Don't assume equal (Welch's)</p> <p><input type="checkbox"/> Assume equal (Fisher's)</p>	<p>Additional Statistics</p> <p><input type="checkbox"/> Descriptives table</p> <p><input type="checkbox"/> Descriptives plots</p>
<p>Missing Values</p> <p><input checked="" type="radio"/> Exclude cases analysis by analysis</p> <p><input type="radio"/> Exclude cases listwise</p>	<p>Assumption Checks</p> <p><input checked="" type="checkbox"/> Normality (Shapiro-Wilk)</p> <p><input type="checkbox"/> Normality (Q-Q plot)</p> <p><input checked="" type="checkbox"/> Equality of variances</p>

10.12. IMAGEN
ANOVA (5)

Los resultados de la prueba ANOVA muestran que no existen diferencias en la variable SaCor en función de la edad ($F(2, 39,5) = 0,785; p = 0,46309$).

One-Way ANOVA (Welch's)				
	F	df1	df2	p
SaCor	0.785	2	39.5	0.46309

5. *Información adicional.* Se puede solicitar a *jamovi* información sobre los estadísticos descriptivos por grupo y la representación gráfica de los resultados.



Variances

Don't assume equal (Welch's)

Assume equal (Fisher's)

Missing Values

Exclude cases analysis by analysis

Exclude cases listwise

Additional Statistics

Descriptives table

Descriptives plots

Assumption Checks

Normality (Shapiro-Wilk)

Normality (Q-Q plot)

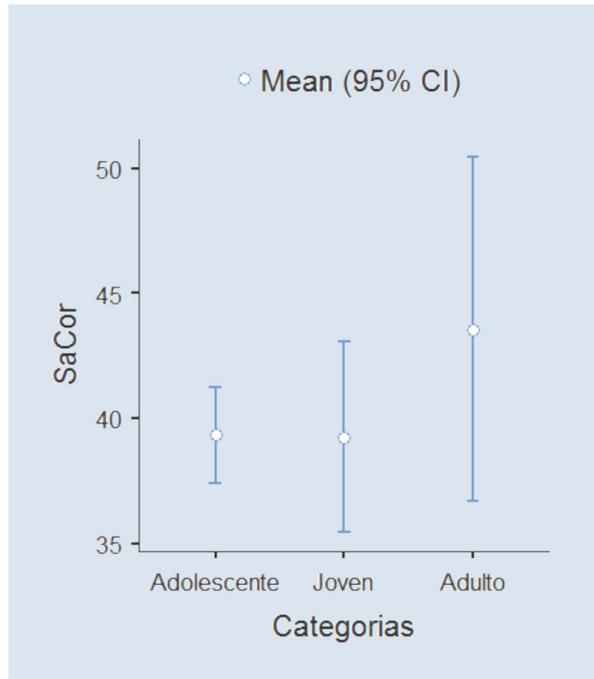
Equality of variances

10.13. IMAGEN

ANOVA (6)

Group Descriptives					
	Categorías	N	Mean	SD	SE
SaCor	Adolescente	132	39.32	11.19	0.9738
	Joven	42	39.24	12.24	1.8879
	Adulto	18	43.56	13.81	3.2552

Los datos muestran que la media en la variable satisfacción corporal en el grupo de 132 adolescentes es 39,32; esta media se mantiene estable en la submuestra de jóvenes (39,24), y es más elevada en el grupo de adultos (43,56).



6. *Diferencias post-hoc*. El test F del ANOVA proporciona una respuesta global a la pregunta ¿son las diferencias entre las medias de los grupos significativas?; un valor p pequeño únicamente indica que las medias no son iguales pero no dice específicamente cuáles son las medias que difieren. Aunque una inspección de los gráficos de las medias o de sus valores da alguna pista de donde se producen estas diferencias, es deseable complementarla con otros test estadísticos suplementarios al ANOVA. Las comparaciones múltiples son las que se llevan a cabo tras un ANOVA y en ellas se comparan sistemáticamente todos los pares de medias para encontrar los grupos que han producido esta significación, en caso de que la haya. Las pruebas diseñadas para llevarlas a cabo mantienen el riesgo alfa en cierto nivel para el conjunto de comparaciones que se efectúen. *jamovi* ofrece dos opciones diferentes para analizar las diferencias por pares dependiendo de que se cumpla o no el supuesto de igualdad de las varianzas, la prueba de Tukey (*Honestly Significantly Differences*) y la prueba de Games-Howell.

▼ Post-Hoc Tests

Post-Hoc Test	Statistics
<input type="radio"/> None	<input checked="" type="checkbox"/> Mean difference
<input checked="" type="radio"/> Games-Howell (unequal variances)	<input checked="" type="checkbox"/> Report significance
<input type="radio"/> Tukey (equal variances)	<input type="checkbox"/> Test results (t and df)
	<input type="checkbox"/> Flag significant comparisons

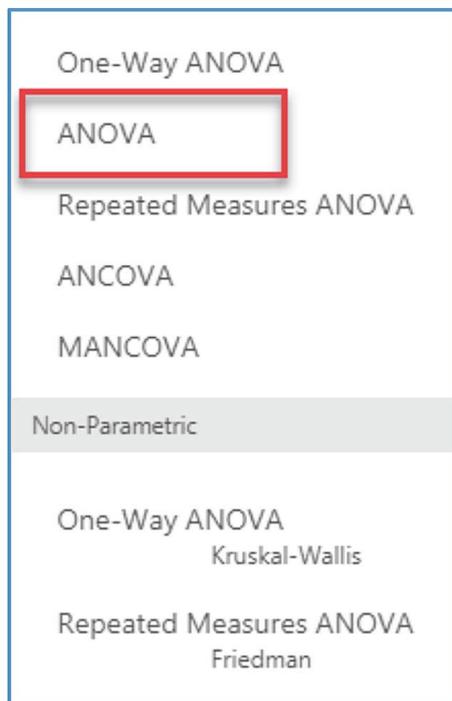
10.14. IMAGEN
ANOVA (7)

En este ejemplo si bien no se han apreciado diferencias significativas entre grupos, queremos ejemplificar la utilización de las pruebas *post-hoc*. La salida siguiente muestra las diferencias dos a dos entre los grupos de edad en la variable satisfacción corporal, ninguna de las cuales es significativa ($p > 0,05$)

Games-Howell Post-Hoc Test – SaCor

		Adolescente	Joven	Adulto
Adolescente	Mean difference	—	0.0801	-4.24
	p-value	—	0.99922	0.44026
Joven	Mean difference		—	-4.32
	p-value		—	0.49349
Adulto	Mean difference			—
	p-value			—

En este ejemplo se ha ejecutado el modelo ANOVA simplificado; puede obtenerse información más completa a través de la opción ANOVA del menú de *jamovi*:



10.15. IMAGEN

Modelo completo ANOVA

Esta opción ofrece información sobre el tamaño del efecto a través de tres índices: eta cuadrado (η^2), eta cuadrado parcial (η_p^2) y omega cuadrado (ω^2), cuyos valores para este ejemplo, como era esperable, son próximos a cero.

ANOVA								
	Sum of Squares	df	Mean Square	F	p	η^2	η^2_p	ω^2
Categorías	295.8	2	147.9	1.084	0.34024	0.0113	0.0113	0.0009
Residuals	25778.7	189	136.4					

11

Validez externa. Relación test-criterio

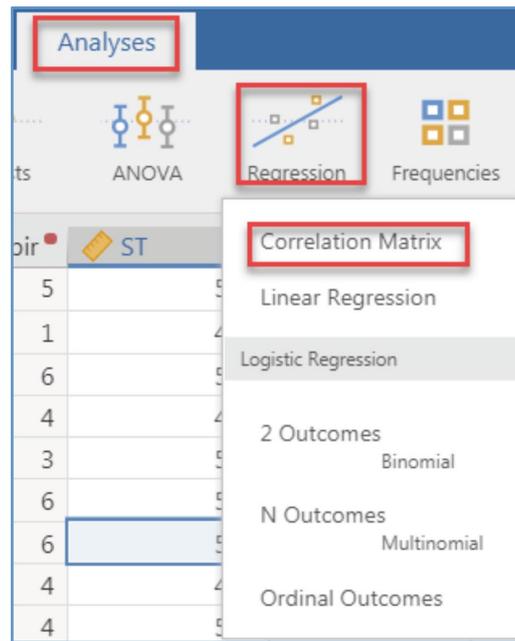
El estudio de las relaciones test-criterio adquiere una gran relevancia en contextos de utilidad donde es fundamental la precisión con que se efectúa una predicción. Su estudio incluye además del grado de precisión, un análisis de los factores que inciden directamente sobre ella, como las características propias del instrumento evaluado, la relevancia, fiabilidad y validez del criterio, y en general todos aquellos que afecten a la fiabilidad de las variables como el tamaño muestral o la restricción del rango. Los diseños utilizados para la obtención de índices de este tipo dependen del tiempo transcurrido entre la recogida de datos en el test y en el criterio, siendo conocidos como el predictivo, concurrente o retrospectivo. En función tanto del número de variables empleadas como de su carácter sería posible la utilización de diversas técnicas de análisis multivariado. Entre ellas, regresión y correlación simple (un test / un criterio), regresión múltiple, regresión logística, análisis discriminante (varios predictores / un sólo criterio) o la correlación canónica y el análisis de regresión multivariante para el caso de varias variables predictoras y varios criterios.

En el estudio de las relaciones entre el test y criterios externos es usual estimar la correlación entre las puntuaciones obtenidas por el test y las puntuaciones obtenidas por otros procedimientos de medida que pueden ser test que miden constructos similares. La correlación entre ambas medidas se conoce como coeficiente de validez del test.

11.1. Correlación. Coeficiente de validez

En este ejemplo utilizamos como variable criterio las puntuaciones obtenidas en un test de Siluetas como indicador de satisfacción corporal. El coeficiente de validez de la escala satisfacción corporal podrá definirse como la correlación entre las puntuaciones obtenidas por ambos test. Para obtener la correlación entre dos o más variables se seguirán las siguientes pautas:

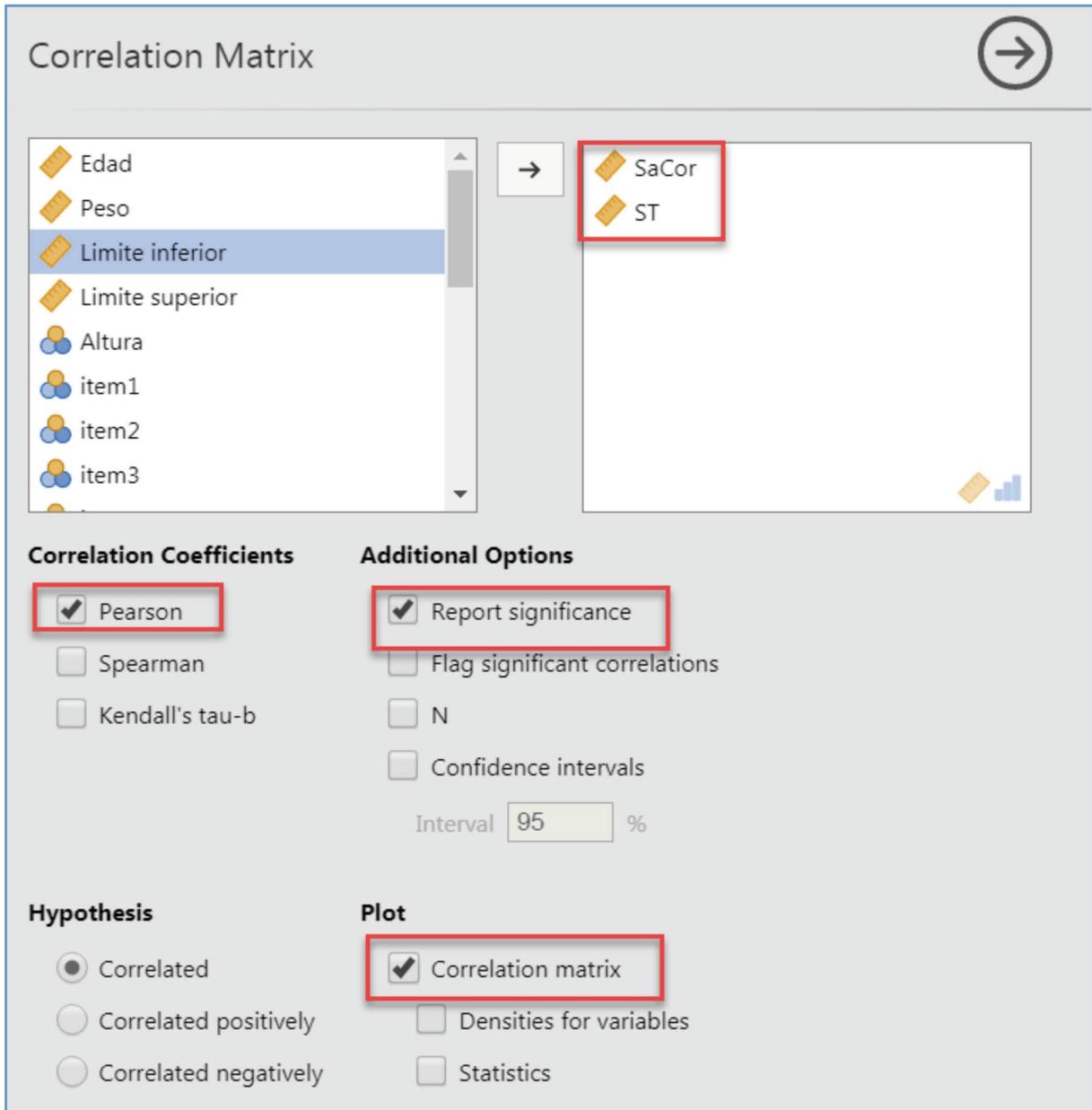
1. *Seleccionar del procedimiento adecuado.* Dentro de la opción general *Analyses* seleccionar *Regression*, y entre las alternativas que nos ofrece *jamovi* optar por *Correlation Matrix*.



11.1. IMAGEN

Matriz de correlaciones (1)

2. *Seleccionar variables.* En la ventana que ofrece *jamovi* seleccionar las variables de interés: SaCor y ST. *jamovi* ofrece varias opciones complementarias: tipo de coeficiente de correlación a estimar, significación estadística, y representación gráfica. En este ejemplo al tratarse de variables continuas optamos por la correlación de Pearson, y solicitamos a *jamovi* que evalúe su significación estadística (*Report significance*).



11.2. IMAGEN

Matriz de correlaciones (2)

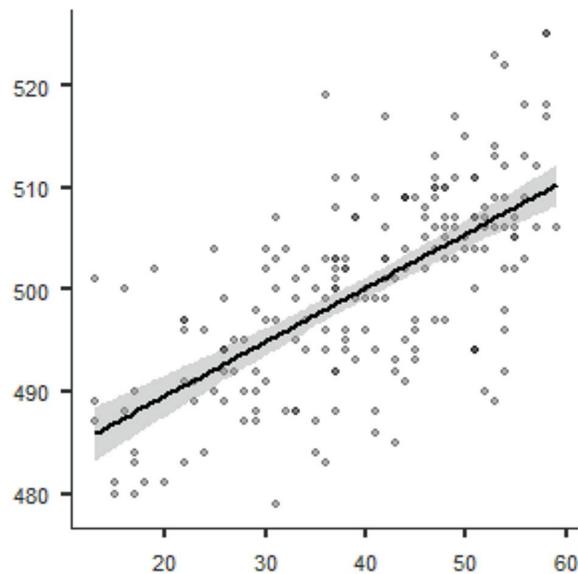
11.1. tabla

Coeficientes de correlación

Correlación	Variables
Pearson	Cuantitativas
Spearman	Cuantitativas (orden)
Kendall	Ordinales

```
# Correlation Matrix
# -----
#           ST           SaCor
# -----
# ST  Pearson's r    —
#     p-value        —
#
# SaCor Pearson's r   0.649    —
#       p-value       <.001    —
# -----
```

El coeficiente de correlación de Pearson entre las puntuaciones obtenidas a través del test de siluetas y la escala de satisfacción corporal es el coeficiente de validez del test de satisfacción corporal. El coeficiente de validez es 0,649; un valor estadísticamente significativo ($p < 0,001$). A través del diagrama de dispersión (*scatter plot*) se puede observar la relación positiva existente entre ambas. El eje X representa la escala de satisfacción corporal y el eje Y el test de siluetas.



11.2. Predicción: Regresión lineal

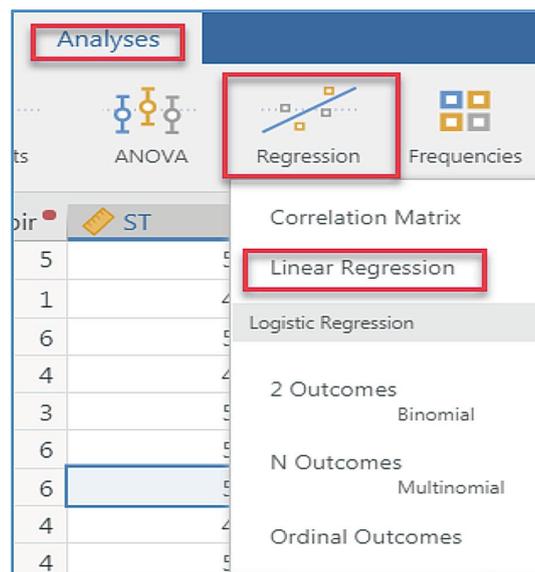
El fin último de numerosos test o cuestionarios es la predicción de un criterio. Cuando se quiere pronosticar el rendimiento académico, el éxito laboral u otro tipo de conducta se utilizan medidas simples o combinadas que permitan una predicción precisa. Aunque los contextos de predicción son muchos y variados, en la sustitución de un criterio por una variable (o compuesto de variables) predictora se persiguen simplicidad, sencillez y ahorro. Si, para pronosticar una conducta futura o una conducta presente, utilizamos test o cuestionarios es necesario garantizar la relación estrecha entre las inferencias derivadas de estos y el criterio a predecir.

El modelo de regresión lineal ofrece un marco teórico que posibilita el estudio entre las relaciones test/criterio; su aplicabilidad dependerá en cada caso del cumplimiento de los supuestos del modelo por los datos. En el caso más simple el modelo de regresión lineal explicaría las relaciones entre una variable predictora y un criterio. En el contexto de la teoría de test, la variable predictora es la puntuación obtenida en el test y la variable criterio es aquella variable que se desea predecir. En psicología y educación las predicciones se basan con frecuencia en modelos más complejos que utilizan como predictores un conjunto de puntuaciones obtenidas tras la administración de varios test o de la utilización de métodos basados en la observación, entrevistas... En este guía únicamente trabajaremos con el modelo de regresión simple.

Como venimos indicando, la utilización y aplicación de todo modelo matemático implica la asunción de varios presupuestos, que en el caso del modelo de regresión simple se podrían resumir en la relación lineal entre la variable predictora y el criterio. Formalmente, es posible utilizar la ecuación de una recta para explicar los efectos que producen sobre la variable dependiente (Y) los cambios en la variable predictora (X).

Para estimar un modelo de regresión simple se seguirán las siguientes pautas:

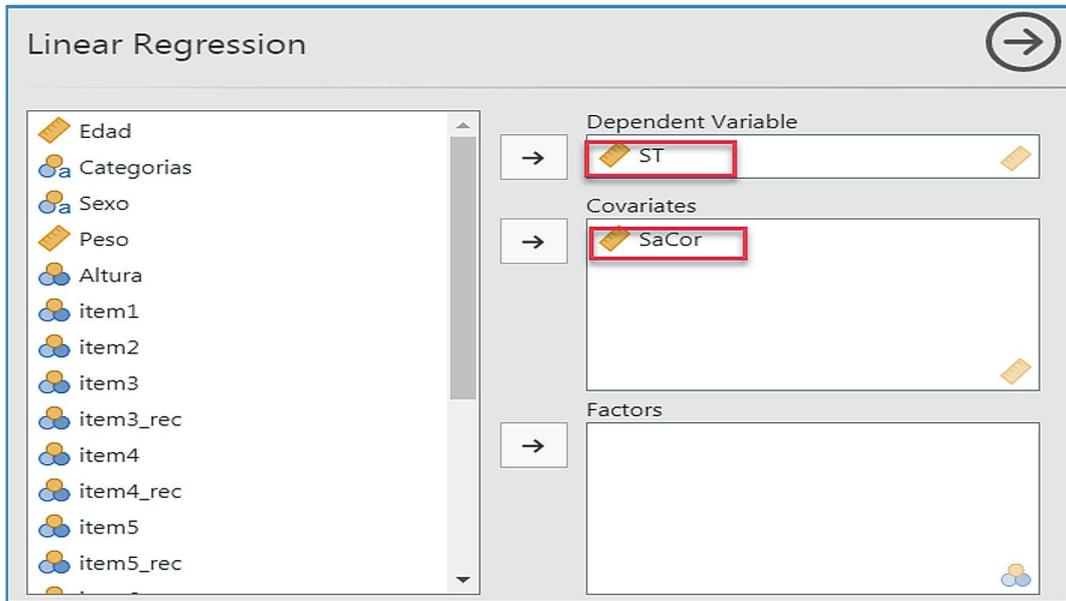
1. *Seleccionar la opción adecuada.* Dentro de las opciones que ofrece *Analyses* seleccionar *Regression*, y dentro de ella *Linear Regression*.



11.3. IMAGEN

Regresión lineal (1)

2. *Seleccionar variables.* En la ventana que ofrece *jamovi* definir las siguientes variables:
 - a) *Dependent Variable.* Variable dependiente. En este caso la puntuación del test de siluetas, ST.
 - b) *Covariates.* Covariables. Variables predictoras continuas. En este ejemplo únicamente disponemos de una, la puntuación en la escala de satisfacción corporal, SaCor.
 - c) *Factors.* Factores. Variables categóricas que pueden incluirse en el modelo predictivo. En este ejemplo no vamos a utilizar factores.



11.4. IMAGEN

Regresión lineal (2)

Esas indicaciones son suficientes para obtener la siguiente salida:

Linear Regression

Model Fit Measures

Model	R	R ²
1	0.6491	0.4214

Model Coefficients - ST

Predictor	Estimate	SE	t	p
Intercept	479.0086	1.86117	257.37	< .00001
SaCor	0.5292	0.04499	11.76	< .00001

La salida tiene dos partes; los parámetros del modelo e información sobre su ajuste. Los valores de los parámetros se pueden leer en la tabla encabezada como *Model Coefficients-ST*. El valor interceptal (*Intercept*) es 479,01 y la pendiente de la recta de regresión (*SaCor*) es 0,53. En función de esos valores la recta de regresión quedaría definida del siguiente modo:

$$ST = 479,01 + 0,53 \times SaCor$$

Es decir, una puntuación de 27 en la escala de satisfacción corporal, se correspondería con una puntuación de 493,32 en el test de siluetas.

$$ST = 479,01 + 0,53 \times 27$$

$$ST = 493,32$$

Junto a cada parámetro *jamovi* ofrece el valor del error típico de estimación (SE), el valor *t* correspondiente al parámetro estimado, y su significación ($Pr(> |t|)$). Esa información puede complementarse con más pruebas estadísticas, pero para este ejemplo es suficiente la información que *jamovi* aporta por defecto.

3. *Evaluación del ajuste*. La evaluación de la bondad de ajuste del modelo se encuentra bajo el epígrafe *Model fit measures*. En esta tabla pueden leerse el valor del coeficiente de correlación de Pearson (0,659), y el valor del coeficiente de determinación (R^2), 0,42. Según estos valores, la escala satisfacción corporal explica el 42% de la varianza observada en el test de siluetas.

11.2.1. Diagrama de dispersión. Instalar módulos

Cuando se estudia la relación lineal entre dos variables, el gráfico de dispersión ofrece información relevante que es interesante conocer. Si bien en el ejemplo anterior hemos ofrecido un gráfico de dispersión, *jamovi* dispone de un módulo específico para la generación de gráficos bivariados, *scatr*.

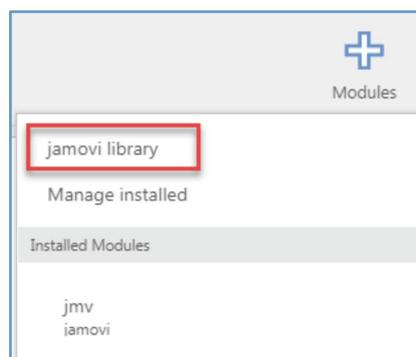
Para instalar un módulo se seguirán las siguientes pautas:

1. Acceder a la opción de gestión de módulos que es visible en el menú de *Analyses*, y dentro de ella seleccionar la opción *jamovi library*



11.5. IMAGEN

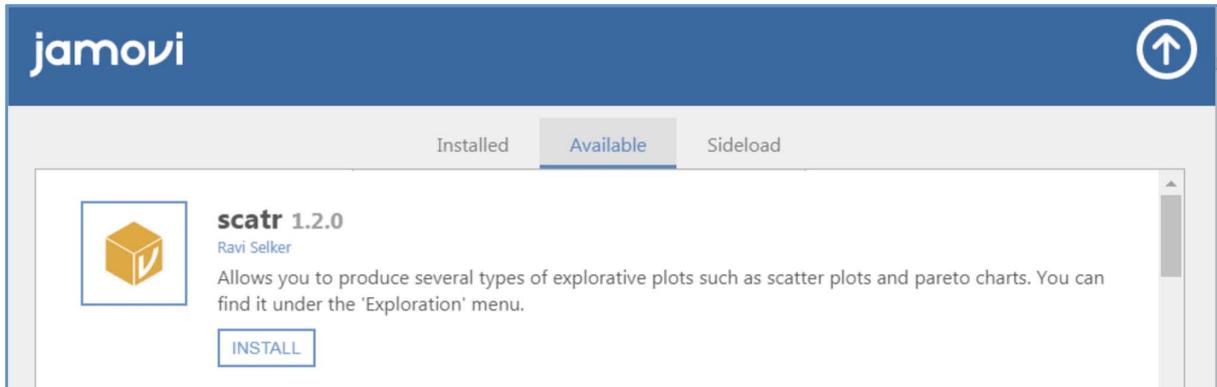
Instalar módulo (1)



11.6. IMAGEN

Instalar módulo (2)

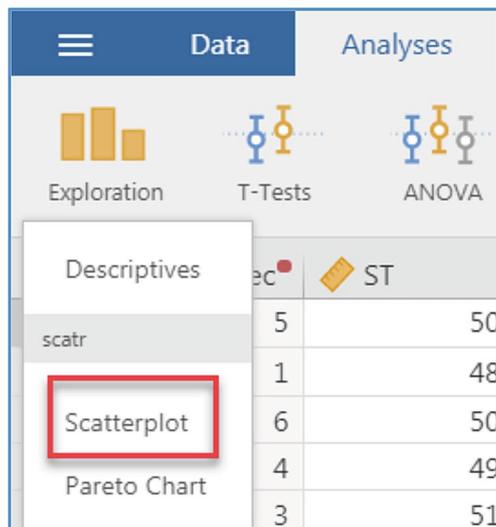
2. A través de esta opción se accede a un listado de los módulos disponibles en *jamovi*. Seleccionar el modulo *scatr*.



11.7. IMAGEN

Instalar módulo (3)

3. Una vez instalado el modulo, dentro de la opción *Exploration* incluida en *Analyses* aparecerán nuevas funciones; de entre ellas *Scatterplot* genera diagramas de dispersión.

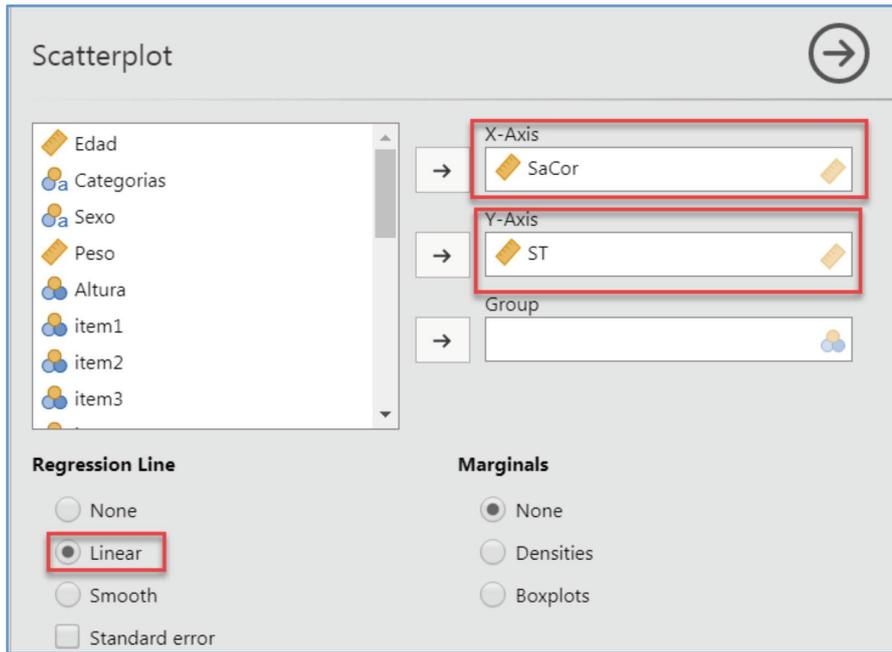


11.8. IMAGEN

Instalar módulo (4)

Para generar la distribución bivariada de las variables satisfacción corporal y ST seguiremos las siguientes pautas:

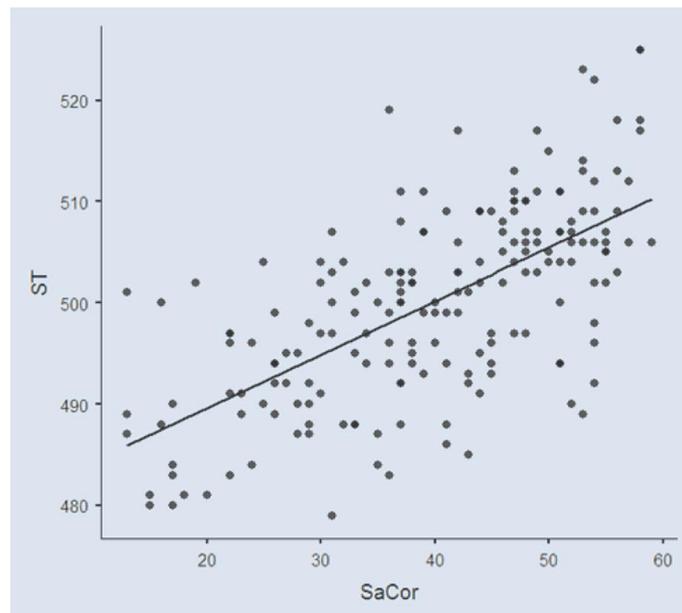
1. *Seleccionar la opción adecuada.* En este caso será *Scatterplot*.
2. *Seleccionar variables.* Moveremos la variable indicadora de satisfacción corporal al eje X, y la variable dependiente, ST, al eje Y.



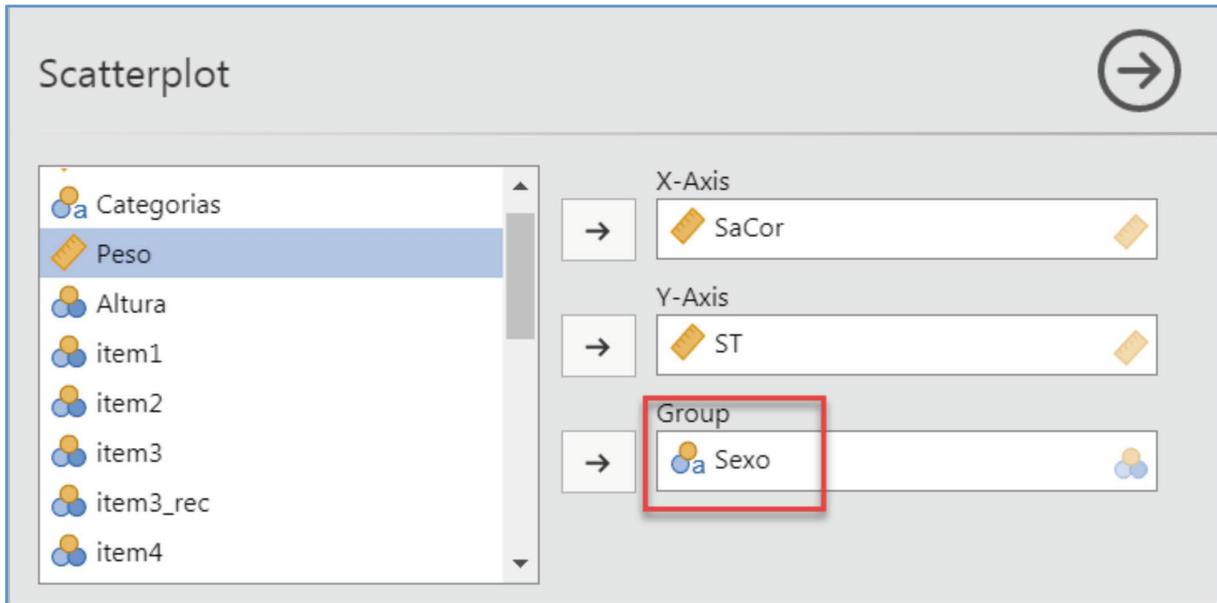
11.9. IMAGEN

Gráfico de dispersión

3. *Seleccionar estadístico apropiado.* Indicaremos a *jamovi* que trace la recta de regresión lineal sobre los datos.

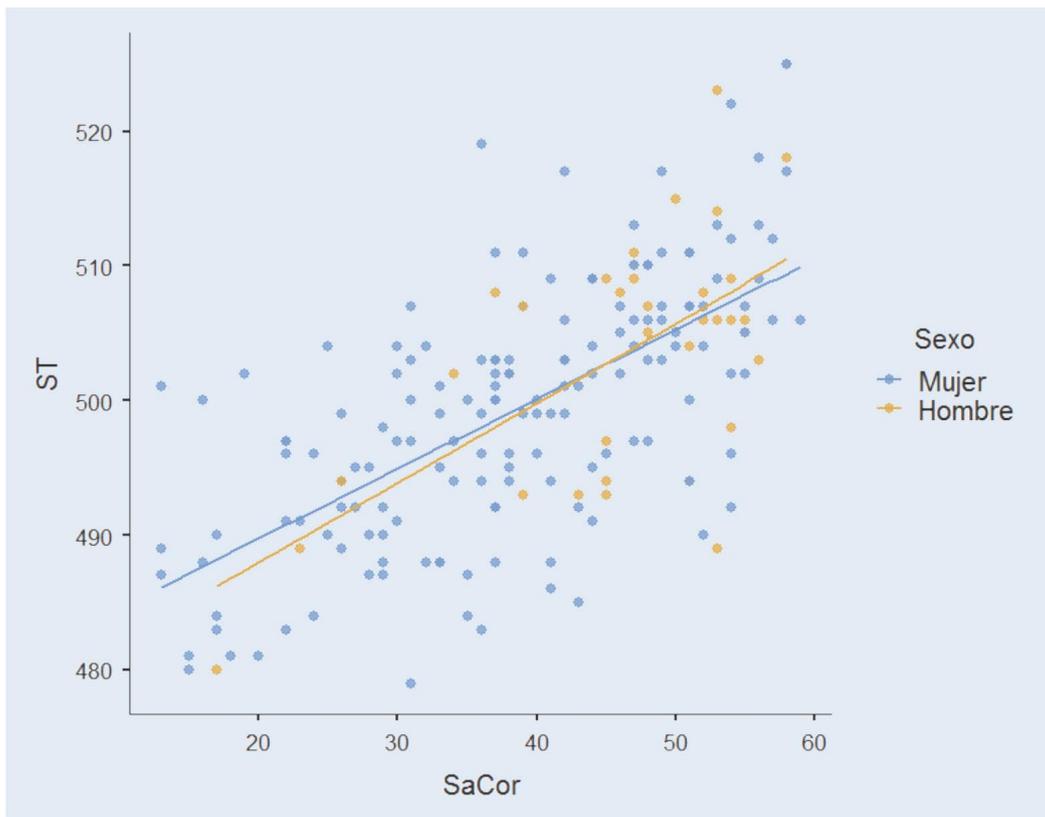


4. *Rectas por grupo.* Si se deseara diferenciar en el diagrama los valores para hombres y mujeres, bastaría con mover la variable sexo al recuadro destinado a la definición de grupos, *Group*.



11.10. IMAGEN

Gráfico de dispersión por grupo



12

Interpretación normativa

El valor empírico o resultado numérico obtenido tras la administración de un test no puede considerarse en términos absolutos, en tanto en cuanto las escalas de puntuación no son escalas de razón; carecen de un 0 absoluto que ejerza de referencia para interpretar un valor numérico. ¿Qué se puede inferir de la información contenida en el siguiente enunciado? «Javier obtuvo 16 puntos en una prueba de razonamiento lógico». Nada, si no se dispone de más información sobre la escala utilizada. La interpretación de la mayor parte de los test psicológicos se lleva a cabo con referencia a la norma (test normativos) o con referencia al criterio (test criterios); si bien existen también test ipsativos —todavía con poca tradición— en los que la interpretación es intrasujeto (Brown y Maydeu-Olivares, 2018).

Para el caso anterior, la puntuación 16 sólo tiene sentido, referida al test a partir del cual se obtuvo, y con relación a la muestra (grupo normativo) en la que se estandarizó el test. Sólo con ambas referencias (test y grupo) se podrá concluir si la puntuación 16 se sitúa por encima o por debajo de la media aritmética del grupo, o se podrá conocer cuántas desviaciones estándar por encima o por debajo de la media se sitúa el valor 16; de hecho, la media aritmética y la desviación estándar se consideran el punto 0 y la unidad de medida de la escala para los test de norma de grupo.

La muestra de estandarización necesaria para la interpretación de las puntuaciones es aquella muestra representativa de la población a la que el test va dirigido. El manual del test ha de incluir información relevante sobre ella, y sobre su distribución en función de las variables relacionadas con el constructo o muestra de conducta evaluado. El contenido de esa información es crucial para la valoración de la aplicabilidad del test. Si la muestra de normalización no es equiparable a las personas/grupo a la que se va a administrar el test, las propiedades psicométricas del test podrán verse comprometidas. Un test puede incluir más de una muestra de normalización. Este sería el caso de aquellas variables cuya distribución varía en función de la edad, grado, sexo, nivel socioeconómico... o cualquier otra que tenga relación con los resultados del test.

En los test de norma de grupo, las puntuaciones empíricas observadas se transforman a puntuaciones de localización con relación al grupo, con el fin de aportar un contexto para su interpretación sin referencia a los parámetros de escala. La transformación de puntuaciones engloba el conjunto de procedimientos diseñados para este cometido (Angoff, 1984; Petersen, Kolen y Hoover, 1989). Básicamente, los tipos de transformación son dos; transformaciones lineales y transformaciones no-lineales; ambas generan tablas interpretativas discretas. Ante las deficien-

cias asociadas con la naturaleza discreta de las normas, la psicometría está trabajando en el desarrollo de procedimientos conocidos como *continuous norming* que con el uso de muestras más reducidas modulan el cambio en las distribuciones de la puntuación observada como una función de la edad u otra variable de interés (Lenhard, Lenhard, Suggate, y Segerer, 2016).

12.1. Transformaciones lineales

Las transformaciones lineales están sujetas a la siguiente ecuación:

$$Y_i = a + bX_i$$

Donde Y_i es el valor de escala transformado,
 a, la media aritmética de la escala destino, y
 b, la desviación estándar de la escala destino.

Como consecuencia de una transformación lineal la forma de la distribución original no sufre ninguna alteración; si fuera normal en origen seguiría siendo normal tras la transformación, y si estuviera sesgada la distribución transformada también lo estaría.

Escala típica: a partir de los parámetros de transformación, $a = 1/S_x$ y $b = \bar{X}/S_x$, se obtienen las puntuaciones z o puntuaciones típicas; tienen una media aritmética de 0 y una desviación estándar de 1, y poseen una relevancia especial en las ciencias sociales.

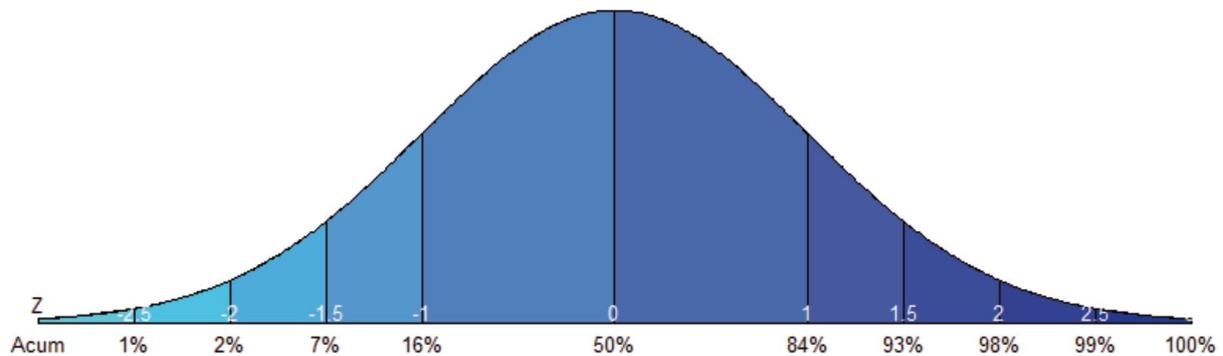
Escala típica derivada: los valores negativos de las escalas típicas (dado que el punto medio de la escala es 0, la mitad de los valores son negativos) no son sencillos de interpretar. Para solventar este inconveniente es frecuente utilizar escalas derivadas con medias aritméticas y desviaciones estándar prefijadas que se obtienen a partir de las puntuaciones típicas (Z). Existen varias escalas que gozan de tradición en la práctica psicológica y educativa; entre ellas, las escalas T, Wechsler, Stanford-Binet, MMPI o Delta:

12.1. tabla

Escalas transformadas

Escala	Media aritmética	Desviación típica
Z	0	1
T	50	10
Wechsler	100	15
Standford-Binet	100	16
MMPI	50	10
Delta	4	13
Estanino o Eneatipo	5	2
Decatipos	5,5	2

La interpretación correcta de las puntuaciones típicas ha de tener en cuenta si la distribución de referencia se ajusta o no a la distribución normal. En el primer caso, podrían hacerse afirmaciones del tipo, «el 95,44% de la población se encuentra en el intervalo definido por la media \pm dos desviaciones típicas». Sin embargo, a medida que la distribución original se aleja de la distribución normal, sentencias como la anterior serían incorrectas.

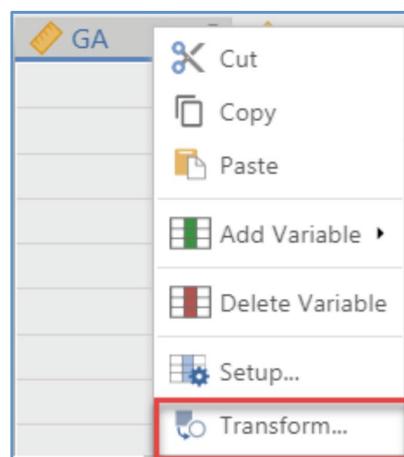


12.1. IMAGEN

Distribución normal

Para una transformación lineal de las puntuaciones se seguirán las siguientes pautas:

1. *Situarse sobre la variable (puntuación) a transformar* —en este ejemplo SaCor— y presionando sobre el botón derecho del ratón seleccionar la opción *Transform*.



12.2. IMAGEN

Puntuaciones típicas (1)

2. En la ventana que abre *jamovi* asignar un nombre a la nueva variable —en este ejemplo se ha asignado el nombre SaCor_Z—, e indicar que se quiere definir una nueva transformación, *Create New Transform* que se aplicará sobre la variable origen SaCor (*Source variable*).

TRANSFORMED VARIABLE

SaCor_Z

Description

Source variable SaCor

using transform None Edit...

None

item_rec

Create New Transform...

12.3. IMAGEN

Puntuaciones típicas (2)

3. En esta nueva pantalla se *definirá la regla de transformación*. Se asigna un nombre a la regla —lo cual permitirá poder reutilizarla—, y se define la función de transformación, $SCALE(\$source)$ en el recuadro correspondiente:

● TRANSFORM used by 1

Escala Z

Description Variable suffix

+ Add recode condition

f_x = SCALE(\$source)

12.4. IMAGEN

Puntuaciones típicas (3)

Como resultado aparecerá la nueva variable en la hoja de datos, SaCor_Z. Puede comprobarse que la media aritmética y la desviación estándar de esta nueva variable son 0 y 1.

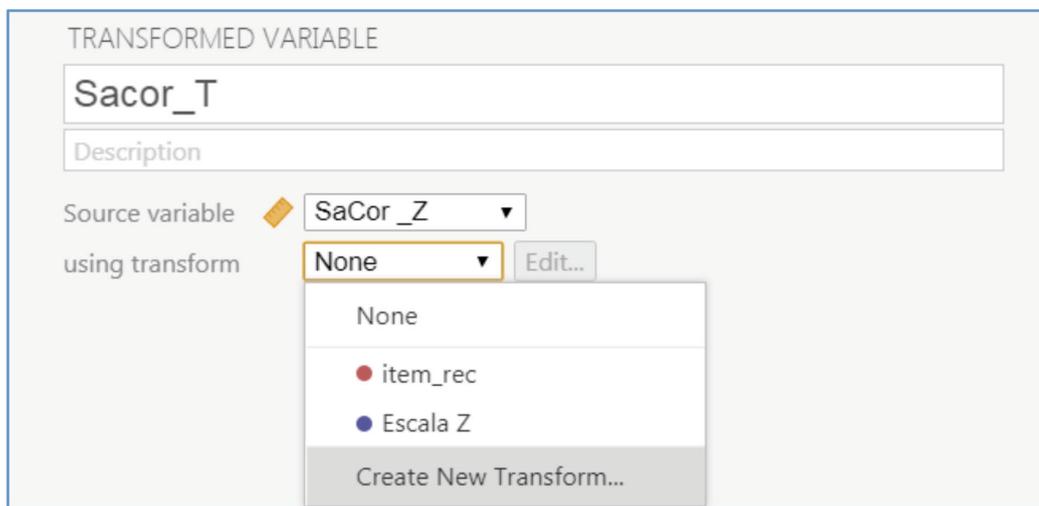
SaCor	SaCor_Z
52	1.053
17	-1.943
51	0.967
51	0.967
47	0.625
50	0.882
52	1.053
39	-0.060
47	0.625
48	0.711
55	1.310

12.5. IMAGEN

Puntuaciones típicas (4)

Para transformar las puntuaciones típicas a cualquiera otra escala se seguirá el mismo procedimiento que el utilizado para obtener los valores Z. En este ejemplo crearemos valores T sobre la variable origen SaCor_Z.

1. *Situarse sobre la variable a transformar* —en este ejemplo SaCor_Z— y presionar el botón derecho del ratón para seleccionar la opción *Transform*.
2. Indicar que se quiere definir una nueva regla de transformación, y asignar un nuevo nombre a la variable a crear, SaCor_T.

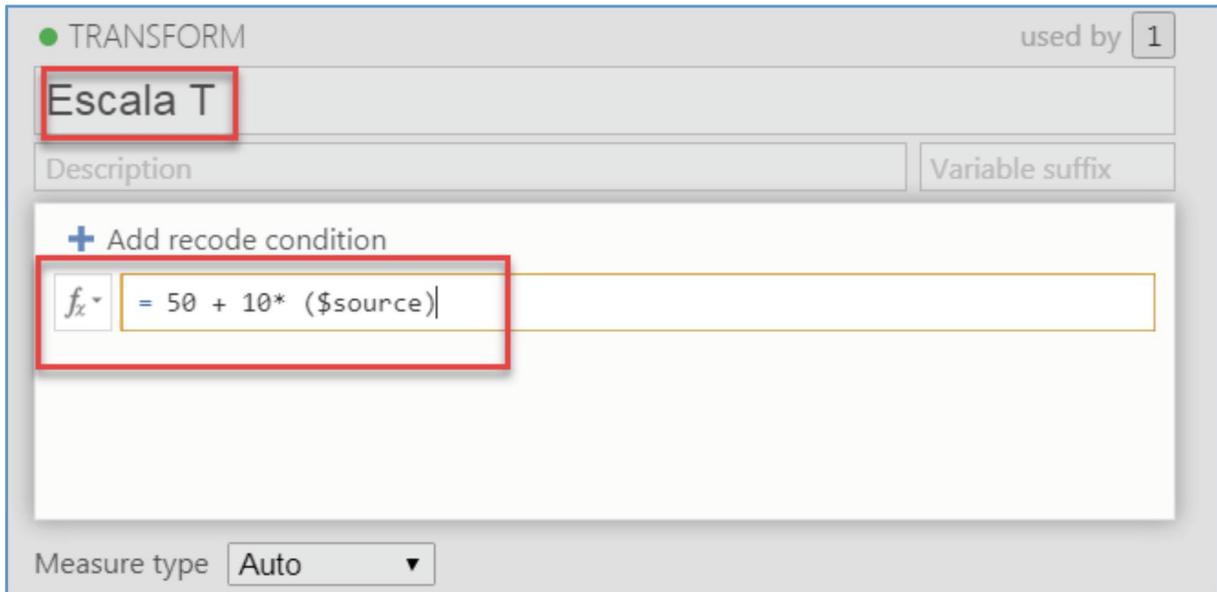


12.6. IMAGEN

Puntuaciones T (1)

3. *Definir la regla de transformación.* En este ejemplo generamos la escala T para la variable SaCor. Para ello aplicamos la siguiente fórmula:

$$T = 50 + \text{SaCor_Z} \times 10$$



12.7. IMAGEN

Puntuaciones T (2)

Como resultado aparecerá una nueva variable en la hoja de datos; Sacor_T. Su media aritmética y desviación típica son respectivamente 50 y 10.

SaCor	SaCor_Z	Sacor_T
17	-1.943	30.573
51	0.967	59.673
51	0.967	59.673
47	0.625	56.250
50	0.882	58.817
52	1.053	60.529
39	-0.060	49.403
47	0.625	56.250
48	0.711	57.106
55	1.310	63.097
39	-0.060	49.403
56	1.395	63.952
39	-0.060	49.403
45	0.454	54.538

12.8. IMAGEN

Puntuaciones T (3)

12.2. Transformaciones no lineales

Además de las transformaciones lineales, existen transformaciones que siendo monótonas crecientes sí alteran la distribución original de las puntuaciones; entre ellas, las más utilizadas son las puntuaciones centiles o percentiles y las puntuaciones estándar normalizadas.

Centiles o percentiles. Los percentiles son las escalas transformadas más comunes; son especialmente relevantes cuando las distribuciones no se ajustan a la distribución normal. La lectura de los percentiles es sencilla tanto para el profesional como para el evaluado. Un percentil indica el porcentaje en la muestra de baremación que se sitúa por debajo de un determinado valor. Sus valores no definen escalas de intervalo, sino de orden, y las diferencias observadas entre las puntuaciones originales no son mantenidas por las escalas percentiles. Es decir, en la lectura de los percentiles es importante tener presente que los percentiles: (a) no son una transformación lineal de la puntuación; (b) maximizan las diferencias en la zona centro de la distribución, y (c) minimizan las diferencias entre puntuaciones brutas en las colas de la distribución. Esto significa que diferencias menores en el centro de la distribución (valores cercanos a la media aritmética) se corresponden con diferencias mayores en los valores percentiles; sin embargo, como en las colas de la distribución hay un número reducido de personas, diferencias pronunciadas en los valores brutos pueden corresponderse con diferencias mínimas en los valores percentiles asociados.

Por ejemplo, la tabla siguiente muestra una transformación en percentiles de una escala que tiene un valor mínimo de 25 y un máximo de 81. La media de la distribución es 50 y su desviación típica es 10. Puede comprobarse que en los valores próximos a la media un incremento de un punto (por ejemplo, puntuaciones 51 y 52) supone el cambio del percentil 50 al percentil 56 (P50-P56); pero en la cola derecha de la distribución el incremento de 5 puntos en la puntuación empírica (p. ej. 70-75) se asocia con un cambio de un punto en la escala de percentiles (P98-P99).

12.2. tabla

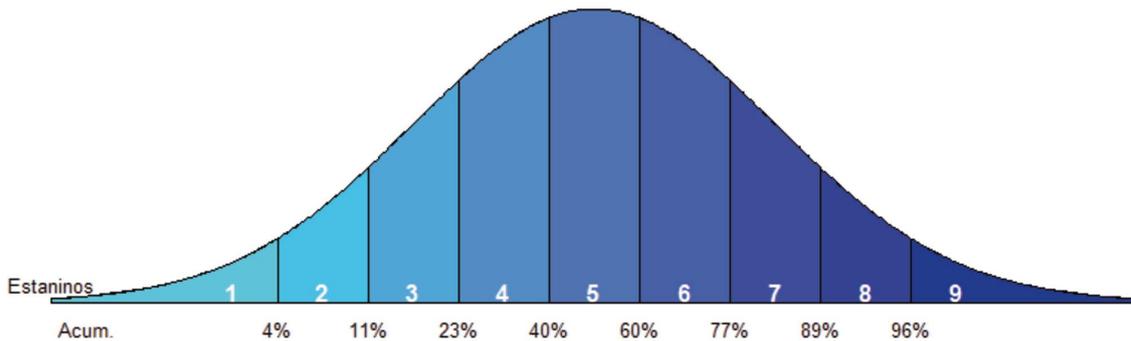
Percentiles

Centil	X	Centil	X	Centil	X	Centil	X	Centil	X
0	25	10	37	31	46	65	55	89	64
1	26	11	38	34	47	68	56	92	65
2	28	13	39	38	48	71	57	94	66
3	30	15	40	41	49	75	58	95	67
4	32	17	41	45	50	77	59	96	69
5	33	20	42	50	51	81	60	98	70
6	34	23	43	56	52	83	61	99	75
7	35	26	44	59	53	86	62	100	81
8	36	28	45	63	54	87	63		

Puntuaciones típicas normalizadas. A partir de los valores centiles es posible obtener puntuaciones típicas normalizadas. Basta con relacionar las puntuaciones centiles con los valores z correspondiente al área de la curva normal que limitan aquellas. Esta transformación normaliza la

distribución original de las puntuaciones, por ello antes de llevarla a cabo es conveniente analizar la condición de normalidad de los datos. Si existen diferencias significativas entre ambas (original-normal) esta transformación no sería aconsejable (Angoff, 1984).

Puntuaciones normalizadas derivadas. Una vez normalizadas las puntuaciones pueden aplicárseles transformaciones lineales que las conviertan a cualquier otra escala. En la tradición psicométrica la escala de estatinos o eneatipos (media = 5; DT = 2), y decativos (media = 55; DT = 2) son de las más utilizadas. Estas escalas segmentan el rango de puntuaciones en 9 y 10 intervalos de igual amplitud (no frecuencia). Los límites de los intervalos se fijan de acuerdo a los valores de la distribución normal



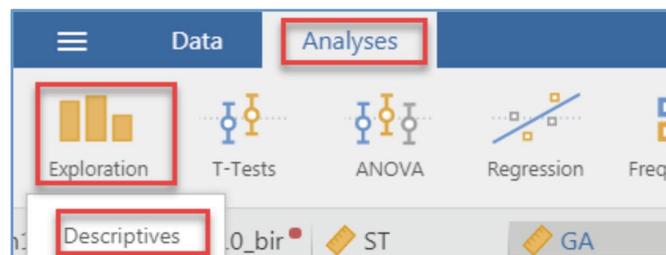
12.9. IMAGEN

Estatinos

En la lectura y valoración de las tablas de baremación incluidas en los test es importante recordar que las poblaciones cambian a lo largo del tiempo, y como consecuencia las tablas interpretativas pueden quedar obsoletas. Al respecto cabe apuntar las directrices de la Comisión Internacional de Test sobre la revisión de test, test obsoletos y la retirada de test (International Test Commission, 2015) que señalan que un test queda obsoleto cuando la teoría subyacente, el contenido de los ítems, las normas sobre las que se construye, o su adecuación técnica ya no satisfacen las necesidades que requiere su uso, o cuando éste puede generar decisiones o diagnósticos inapropiados o imprecisos.

Para obtener los valores percentiles se seguirán las siguientes pautas.

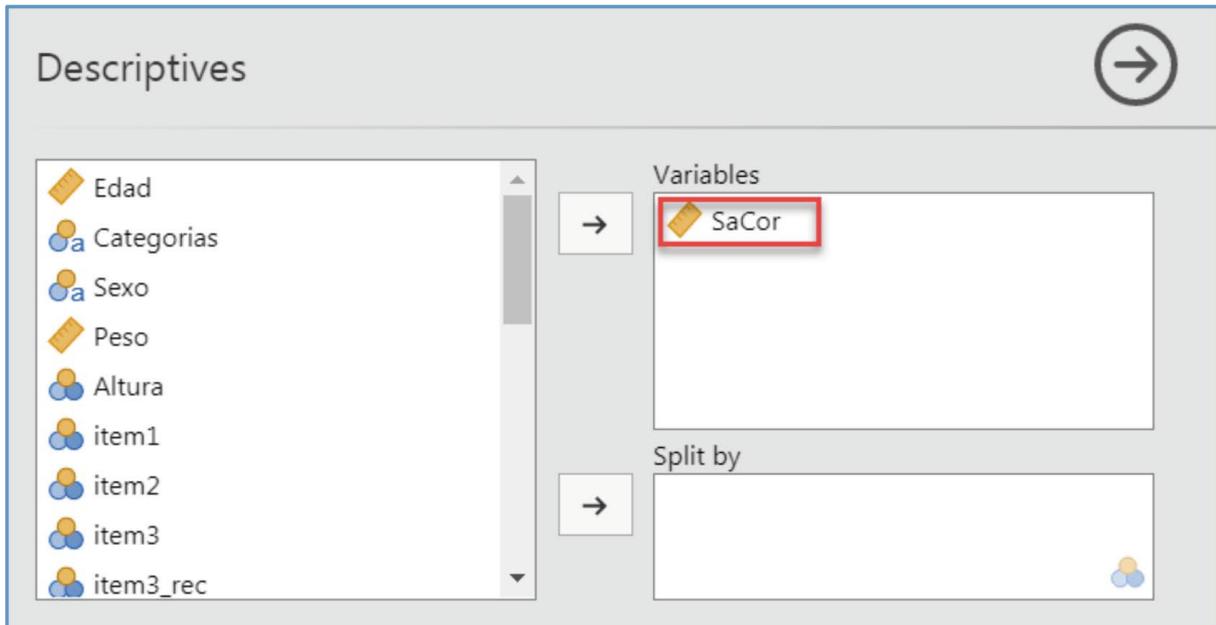
1. *Seleccionar el procedimiento adecuado.* Dentro del menú *Analyses* y *Exploration* seleccionar *Descriptives*.



12.10. IMAGEN

Puntuaciones centiles (1)

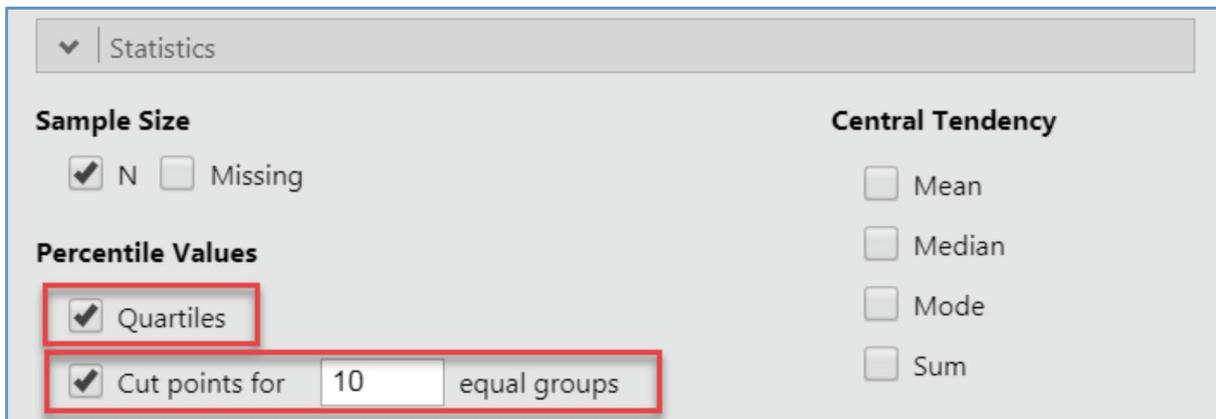
2. *Seleccionar variables.* En esta ventana seleccionar la variable satisfacción corporal (SaCor).



12.11. IMAGEN

Puntuaciones centiles (2)

3. *Seleccionar el estadístico adecuado.* En el recuadro destinado a los valores percentiles *Percentile Values* marcar las opciones cuartiles y percentiles. En este ejemplo pedimos a *jamovi* los cuartiles de la distribución, y que divida esta en 10 segmentos con el mismo número de personas en cada uno de ellos.



12.12. IMAGEN

Puntuaciones centiles (3)

Con esta instrucción se obtienen, a) los valores que dividen la distribución en 4 grupos —cuartiles—, y b) los valores percentiles. En la salida se lee que las puntuaciones 31, 41 y 49 dejan bajo ellas el 25%, 50% y 75% de la muestra. Con referencia a los valores percentiles puede

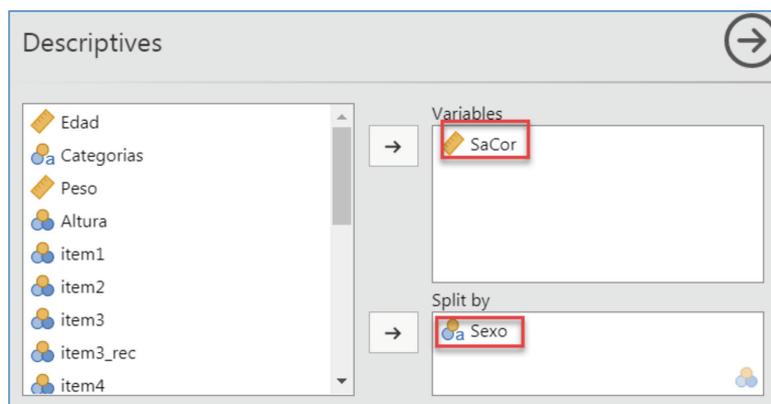
comprobarse que el percentil 20 se corresponde con la puntuación 29, y que la puntuación que deja por debajo un 60% de la muestra es 44.

	SaCor
25th percentile	31.00
50th percentile	41.00
75th percentile	49.00
10th percentile	23.00
20th percentile	29.00
30th percentile	34.00
40th percentile	37.00
50th percentile	41.00
60th percentile	44.00
70th percentile	48.00
80th percentile	51.00
90th percentile	54.00

12.2.1. Centiles por grupo

Si una submuestra de baremación es sensible a la variable medida (sexo, edad, nivel socio-económico...), es decir si se aprecian diferencias significativas entre grupos podría ser interesante construir baremos específicos para ella.

En este ejemplo creamos tablas de baremación diferenciadas en función del sexo. Para ello una vez seleccionado el procedimiento adecuado en la ventana de selección de variables situamos la variable satisfacción corporal (SaCor), y en el recuadro destinado a segmentar la muestra (*Split by*) ubicamos la variable sexo.



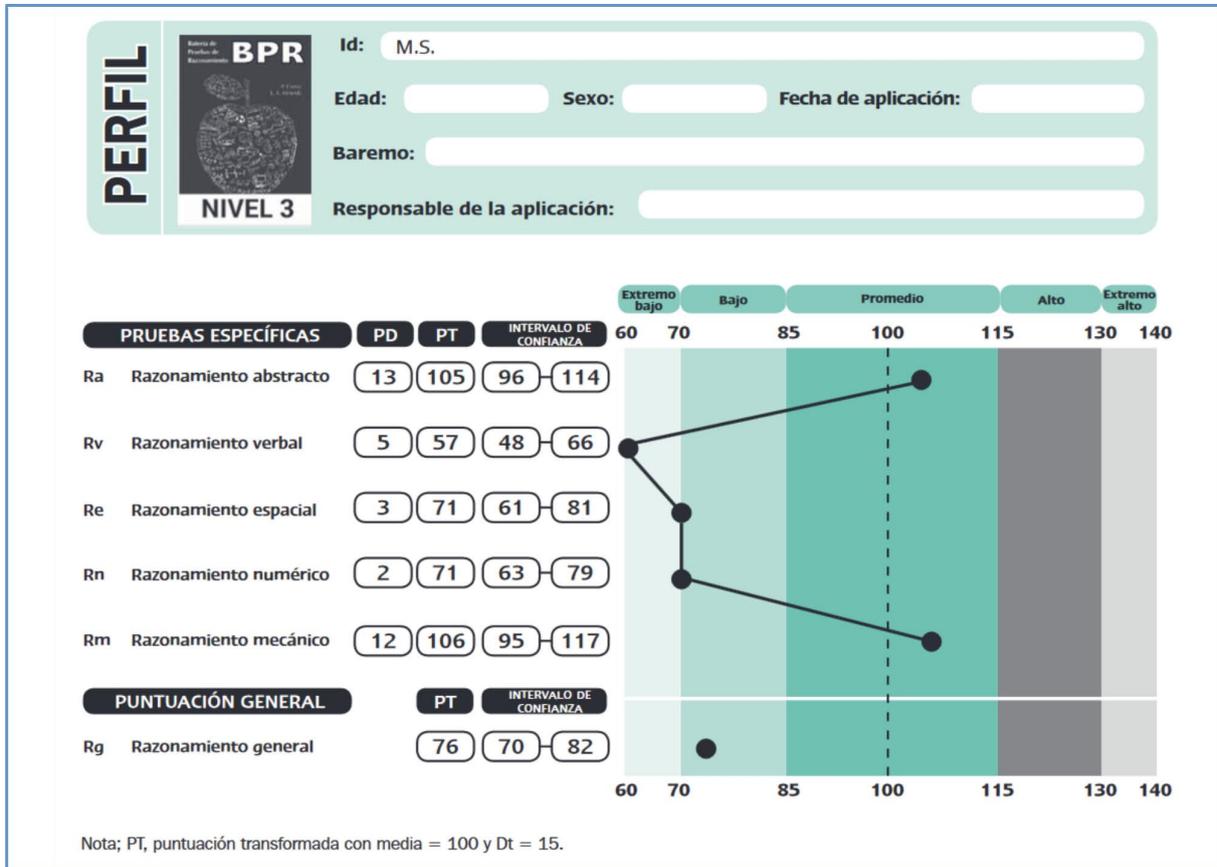
12.13. IMAGEN

Puntuaciones centiles (4)

Con esa instrucción obtendremos los valores percentiles para las mujeres y para los hombres.

Descriptives		
	Sexo	SaCor
N	Mujer	155
	Hombre	37
25th percentile	Mujer	30.0
	Hombre	45.0
50th percentile	Mujer	38.0
	Hombre	48.0
75th percentile	Mujer	47.5
	Hombre	53.0
10th percentile	Mujer	22.0
	Hombre	35.8
20th percentile	Mujer	28.0
	Hombre	43.2
30th percentile	Mujer	31.2
	Hombre	45.0
40th percentile	Mujer	36.0
	Hombre	47.0
50th percentile	Mujer	38.0
	Hombre	48.0
60th percentile	Mujer	42.0
	Hombre	51.0
70th percentile	Mujer	45.8
	Hombre	53.0
80th percentile	Mujer	49.0
	Hombre	53.8
90th percentile	Mujer	53.6
	Hombre	55.0

Una vez finalizado el proceso de transformación de puntuaciones, se ofrece al usuario una tabla de lectura sencilla o un informe gráfico que recoja la información más relevante. Por ejemplo para la Batería de Pruebas de Razonamiento (Elosua y Almeida, 2016) se ofrece este gráfico como parte del informe final:



12.14. IMAGEN
Informe. Ejemplo

13

Informe

Finalizados los análisis, se debe elaborar un informe en el que se describa el proceso seguido, y se comenten razonadamente los resultados obtenidos. El informe incluirá los siguientes apartados:

1. Descripción de la muestra.
Descripción lo más clara posible de: 1) el número de personas que componen la muestra; 2) características de los participantes (edad, sexo..) y, de los 3) grupos diferenciados si es que los hay.
2. Análisis de ítems.
Se debe informar e interpretar la media, desviación típica, atracción de las opciones, efectos techo/suelo e índice de discriminación de los ítems, con el fin de responder a la pregunta ¿Son adecuados los valores? Hay que explicar detalladamente el proceso seguido para determinar cuántos ítems y porqué han de eliminarse de la escala inicial.
3. Fiabilidad de la escala.
Adjuntar e interpretar el coeficiente de fiabilidad obtenido por el método alfa de Cronbach, respondiendo a las cuestiones: ¿Cuál es el coeficiente de consistencia interna de la escala? ¿Cuál es el error estándar de medida de la escala? ¿Cual es la aportación de cada ítem a la fiabilidad de la escala?
Presentar un ejemplo de estimación de los intervalos de confianza en torno a la puntuación verdadera, y explicarlo.
4. Validez interna.
Informar sobre el proceso y resultados del estudio de la estructura interna. Analizar y valorar: 1) cumplimiento de los supuestos del modelo: 2) factores extraídos; cuantos, procedimiento de extracción y rotación; 3) solución factorial: saturaciones, varianza explicada por los factores retenidos y correlaciones entre factores, y 4) ajuste del modelo e importancia de cada factor.
5. Validez externa.
Mostrar e interpretar los resultados obtenidos en las comparaciones entre grupos: 1) cumplimiento de los supuestos del modelo; 2) prueba estadística utilizada; 3) descripción de las diferencias, y 4) tamaño del efecto. ¿Puede asumirse que existen diferencias entre subgrupos?

Relaciones test/criterio: 1) cumplimiento de los supuestos del modelo; 2) prueba estadística utilizada; 3) descripción de la relación; coeficiente de validez y ecuación de regresión, y 4) evaluación del ajuste del modelo. ¿Es adecuado utilizar la escala para predecir el criterio? ¿Cuál es el coeficiente de determinación?

6. Baremos.

Incluir la/s tabla/s que muestre/n los baremos por centiles y puntuaciones T. Explicar el uso de la tabla y la interpretación de sus valores aportando algún ejemplo concreto con una o dos puntuaciones observadas.

7. Valoración general de la calidad de la escala. Puntos fuertes, criterios para mejorarla; necesidades y deficiencias más notorias.

14

Referencias

- Abad, F. J., Olea, J., Ponsoda, V., y García, C. (2011). *Medición en ciencias sociales y de la salud*. Madrid: Síntesis.
- American Educational Research Association, American Psychological Association, y National Council on Measurement in Education. (1999). *Standards for European Psychologist educational and psychological testing*. Washington, DC: American Psychological Association.
- American Psychological Association. (2010). *Publication manual of the American Psychological Association (6.a ed.)*. Washington, DC: Autor.
- Angoff, W. H. (1984). *Scales, Norms and Equivalent Scores*. Princeton, New Jersey: Educational Testing Service.
- Bandalos, D. L. y Finney, S. J. (2010). Factor Analysis: Exploratory and Confirmatory. En G. R. Hancock y R. O. Mueller (Eds.), *Reviewer's guide to quantitative methods*. Routledge: New York.
- Bartlett, M. S. (1950). Test of significance in factor analysis. *British Journal of Psychology*, 3, 77-85.
- Becker, R. A., Chambers, J. M., y Wilks, A. R. (1988). *The new S language: A programming environment for data analysis and graphics*. PacificGrove, CA: Wadsworth.
- Binet, A., y Simon, Th. A. (1905). Méthode nouvelle pour le diagnostic du niveau intellectuel des anormaux. *L'Année Psychologique*, 11, 191-244.
- Blue Sky statistics [Computer software]. (2019). Retrieved from <https://www.blueskystatistics.com/Default.asp>.
- Brown, A., y Maydeu-Olivares, A. (2018). Modeling forced-choice response format. In Irwing, P., Booth, T., Hughes, D., D. (eds.), *The Wiley handbook of psychometric testing* (pp. 523-569). London, England: Wiley.
- Cattell, R. B. (1966). *Handbook of multivariate experimental psychology*. Chicago: Rand McNally.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York: Routledge.
- Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, 16, 297-334.
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: The Guilford Press
- de Boeck, P., y Elosua, P. (2016). Reliability and Validity: History, Notions, Methods, Discussion. En F.T. L. Leong, D. Bartram, F. Cheung, K. F. Geisinger y D. Iliescu (Eds.), *The ITC International Handbook of Testing and Assessment* (pp. 408-421). New York: Oxford University Press.
- Ebel, R. L. (1965). *Measuring Educational Achievement*. Englewood, N.J.: Prentice-Hall.

- Elosua, P. (2003). Sobre la validez de los test. *Psicothema*, 15(2), 315-321.
- Elosua, P. (2009). ¿Existe vida más allá de SPSS? Descubre R. *Psicothema*, 21(4), 652-655.
- Elosua, P. (2011). *Psicometría. Conceptos básicos y aplicaciones prácticas con Rcommander*. Leioa, UPV/EHU. <https://addi.ehu.es/bitstream/handle/10810/15581/UWLGPS5661.pdf?sequence=1&isAllowed=y>
- Elosua, P. y Almeida, L. (2016). *BPR. Batería de Pruebas de Razonamiento*. Madrid, TEA ediciones.
- Elosua, P., López-Jáuregui, A., y Sánchez-Sánchez, F. (2010). *Manual técnico con la adaptación al español del Eating Disorder Inventory-3*. Madrid: TEA Ediciones.
- Embretson, S. E. (1983). Construct validity: Construct Representation versus Nomothetic Span. *Psychological Bulletin*, 93, 179-197.
- Embretson, S. E. (1985). *Test design: Developments in psychology and psychometrics*. New York: Academic Press, Inc.
- Epskamp, S. (2017). semPlot: Path Diagrams and Visual Analysis of Various SEM Packages' Output. [R package].
- Fellows, I. (2012). Deducer: A Data Analysis GUI for R. *Journal of Statistical Software*, 49(8), 1-15. <http://www.jstatsoft.org/v49/i08/>.
- Ferrando, P. J. y Anguiano-Carrasco, C. (2010). El análisis factorial como técnica de investigación en psicología. *Papeles del Psicólogo*, 31, 18-33.
- Ferrando, P. J. y Lorenzo-Seva, U. (2014). El análisis exploratorio de los ítems: algunas consideraciones adicionales. *Anales de psicología*, 1170-1175.
- Forero, C. G, Maydeu-Olivares, A. y Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling*, 16, 625-641.
- Fox, J. (2005). The R Commander: A Basic Statistics Graphical User Interface to R. *Journal of Statistical Software*, 14(9), 1-42. <http://www.jstatsoft.org/v14/i09>
- Garner, D. M. (2004). *Eating Disorder Inventory-3*. Lutz, FL: Psychological Assessment Resources, Inc.
- Guttman, L. (1954). Some Necessary Conditions for Common Factor Analysis. *Psychometrika*, 19, 149-162.
- Hambleton, R. K. y Swaminathan, H. (1985). *Item response theory. Principles and applications*. Boston, Kluwer: Nijhoff Publishing.
- Hattie, J. (1985). Methodology review: assessing unidimensionality of test and items. *Applied psychological measurement*, 9(2), 139-164.
- Ho, A. D. y Carol, C. Y. (2015). Descriptive statistics for modern test score distributions skewness, kurtosis, discreteness, and ceiling effects. *Educational and Psychological Measurement*, 75, 365-388.
- Horn, J. L. (1965). A rationale and test for the number of factors in a factoranalysis. *Psychometrika*, 30, 179-185.
- Ihaka, R. y Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, 5, 299-314.
- International Test Commission (2001). International Guidelines for Test Use, *International Journal of Testing*, 1:2, 93-114. [Traducción de la Comisión de Test del Colegio Oficial de Psicólogos de España]. <https://www.cop.es/index.php?page=directrices-internacionales>
- International Test Commission (2015). *International Guidelines for Practitioner Use of Test Revisions, Obsolete Tests, and Test Disposal* [www.Intestcom.org].

- Irwin, P. (Ed.) (2018). *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development*. UK: John Wiley & Sons Ltd.
- Izquierdo, I., Olea, J., y Abad, F. (2014). Exploratory factor analysis in validation studies: Uses and recommendations, *Psicothema*, 26, 395-400.
- JASP Team (2019). JASP [Computer software].
- Kaiser, H. F. (1958). The Varimax Criterion for Analytic Rotation in Factor Analysis. *Psychometrika*, 23, 187-200.
- Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*, 20, 141-151.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39, 31-36.
- Kiers, H. A. L. (1994). Simplimax: oblique rotation to an optimal target with simple structure, *Psychometrika*, 59, 567-579.
- Lenhard, A., Lenhard, W., Suggate, S., y Segerer, R. (2016). A continuous solution to the norming problem. *Assessment*, 25, 112-25.
- Loevinger, J. (1957). Objective test as instruments of psychological theory. *Psychological Reports* (Monograph Supp. 9), 3, 635-694.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lloret-Segura, S., Ferreres-Taver, A., Hernández-Baeza, A., y Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *Anales de Psicología*, 30, 1151-1169.
- McDonald, R. P. (1985). *Factor Analysis and Related Methods*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- McDonald, R. P. (1999). *Test theory. A Unified Treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Messick, S. (1989). Validity. En R. L. Linn (Eds.), *Educational Measurement* (Third Edition, 13-104 pág.). New York: American Council on Education Macmillan Publishing Company.
- Messick, S. (1995). Validity of psychological assessment. *American Psychologist*, 50, 741-749.
- Morin, A. J. S., Marsh, H. W., y Nagengast, B. (2013). Exploratory structural equation modeling. En G. R. Hancock, y R. O. Mueller (eds.), *Structural equation modeling: A second course* (pp. 395-440). Charlotte, NC: Information Age Publishing, Inc.
- Muñiz, J. (2018). *Introducción a la psicometría*. Madrid: Pirámide.
- Muñiz, J. y Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, 31, 7-16.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Olea, J., Abad, F., y Barrada, J. R. (2010). Tests informatizados y otros nuevos tipos de tests. *Papeles del Psicólogo*, 31, 94-107.
- R Core Team (2018). R: A Language and environment for statistical computing. [Computer software].
- Raju, N. S., Burke, M. J., y Norrmand, J. (1990). A new approach to utility analysis. *Journal of Applied Psychology*, 75: 3-12.
- Petersen, N. S., Kolen, M. J., y Hoover, H. D. (1989). Scaling, norming, and equating. In R. L. Linn (ed.), *The American Council on Education/Macmillan series on higher education. Educational measurement* (p. 221-262). Macmillan Publishing Co, Inc; American Council on Education.
- Polit, D. F. y Yang, F.M. (2016). *Measurement and the Measurement of Change*. Hong Kong: Wolters Kluwer.

- Prieto, G. y Delgado, A. (1999). Medición cognitiva de las aptitudes. En J.Olea, V. Ponsoda y G. Prieto (eds.), *Tests informatizados. Fundamentos y aplicaciones* (207-226 pág.). Madrid: Pirámide.
- Rhemtulla, M., Brosseau-Liard, P. É., y Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17, 354-373.
- Revelle, W. (2019). *psych: Procedures for Psychological, Psychometric, and Personality Research*. [R package].
- Revelle, W., y Zinbarg, R. (2009). Coefficients alpha, beta, omega, and the glb: comments on Sijtsma. *Psychometrika* 74, 145-154.
- Rödiger, S., Friedrichsmeier, T., Kapat, P., y Michalke, M. (2012). RKWard: A Comprehensive Graphical User Interface and Integrated Development Environment for Statistical Analysis with R. *Journal of Statistical Software*, 49, 9
- Rosseel, Y. et al. (2018). *lavaan: Latent Variable Analysis*. [R package].
- Sawilowsky, S (2009). New effect size rules of thumb, *Journal of Modern Applied Statistical Methods*, 8, 467-474.
- Suárez-Alvarez, J., Pedrosa, I., Lozano, L. M., García-Cueto, E., Cuesta, M., y Muñoz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema*, 30, 149-158.
- The jamovi project (2019). *jamovi*. (Version 1.0) [Computer Software].
- Trizano-Hermosilla, I. y Alvarado, J. M. (2016). Best AI-ternatives to Cronbach's Alpha Reliability in Realistic Conditions: Congeneric and Asymmetrical Measurements. *Frontiers in Psychology*, 7, 1-8.
- Viladrich, C., Angulo-Brunet, A., y Doval, E. (2017). Un viaje alrededor de alfa y omega para estimar la fiabilidad de consistencia interna. *Anales de psicología*, 33, 755-782.
- Wilson, M. (2005). *Constructing Measures. An Item Response Modeling Approach*. Mahwah, NJ: Lawrence Erlbaum.

UNIBERTSITATEKO ESKULIBURUAK
MANUALES UNIVERSITARIOS

INFORMAZIOA ETA ESKARIAR • INFORMACIÓN Y PEDIDOS

UPV/EHUko Argitalpen Zerbitzua • Servicio Editorial de la UPV/EHU
argitaletxea@ehu.eus • editorial@ehu.eus
1397 Posta Kutxatila - 48080 Bilbo • Apartado 1397 - 48080 Bilbao
Tfn.: 94 601 2227 • www.ehu.eus/argitalpenak

eman ta zabal zazu



Universidad del País Vasco Euskal Herriko Unibertsitatea