

EDUCACIÓN

Medición en Investigación Educativa con Apoyo del SPSS y el AMOS

Angel Alberto Valdés Cuervo
Fernanda Inéz García Vázquez
Gisela Margarita Torres Acuña
Maricela Urías Murrieta
Christian Samhir Grijalva Quiñonez



Medición en investigación educativa con apoyo del SPSS y el AMOS

Angel Alberto Valdés Cuervo
Fernanda Inéz García Vázquez
Gisela Margarita Torres Acuña
Maricela Urías Murrieta
Christian Samhir Grijalva Quiñonez

Departamento de Educación, Instituto Tecnológico de Sonora



CONACYT
Registro Nacional de Instituciones
y Empresas Científicas y Tecnológicas
Registro: 2016/17732

Medición en investigación educativa con apoyo del SPSS y el AMOS

- © Angel Alberto Valdés Cuervo
- © Fernanda Inéz García Vázquez
- © Gisela Margarita Torres Acuña
- © Maricela Urias Murrieta
- © Christian Samhir Grijalva Quiñonez

D. R. © 2019, Departamento de Educación, Instituto Tecnológico de Sonora

Dirección del proyecto

Carlos Herver Díaz
Esther Castillo Aguilar
José Eduardo Salinas de la Luz

Arte

Livia M. Rocco Sarmina
Paulina Cordero Mote
Laura Isabel Soler Navarro
Vanessa Alejandra Vázquez Fuentes

Preprensa

José Luis de la Rosa Meléndez

Diseño y formación de interiores

Livia M. Rocco Sarmina

1a. edición

© 2019 Fernando de Haro y Omar Fuentes

ISBN: 978-607-437-503-9

CLAVE EDITORIAL

Paseo de Tamarindos 400B, Suite 109.
Col. Bosques de las Lomas. C.P. 05120, Ciudad de México, México
Tel. 52 (55) 5258 0279/80/81
ame@ameditores.mx www.ameditores.com
ecastillo@ameditores.mx

Ninguna parte de este libro puede ser reproducida, archivada o transmitida en forma alguna o mediante algún sistema, ya sea electrónico, mecánico o de fotorreproducción, sin la previa autorización de los editores.

Elaborado en México.

Índice

Prólogo	09
Capítulo 1. Medición. Aspectos generales.....	13
¿Por qué es importante medir para un investigador?	
Constructos e indicadores	
Escala de medición	
Papel de la teoría y la psicometría en la medición	
Retos de la medición psicoeducativa	
Conclusiones	
Bibliografía sugerida	
Capítulo 2. Conceptos básicos en medición	21
Escala de medida	
Diferencias individuales	
Medidas de normalidad	
Procedimiento en el SPSS para calcular asimetría y curtosis	
Pruebas de normalidad	
Procedimiento en el SPSS para calcular pruebas de normalidad	
Medidas de tendencia central y dispersión	
Teoría clásica de medición	
Conclusiones	
Bibliografía sugerida	
Capítulo 3. Dimensionalidad de las escalas de medición.....	37
Escala unidimensional	
Escala multidimensional con dimensiones correlacionadas	
Escala multidimensional con dimensiones no correlacionadas	
Procedimientos para determinar la dimensionalidad	
Conclusiones	
Bibliografía sugerida	
Capítulo 4. Fiabilidad	43
Definición de fiabilidad	
Varianza explicada	
Correlación entre los puntajes de los ítems	
Factores que afectan la fiabilidad	
Costos de una pobre fiabilidad	
Medidas de fiabilidad	
Procedimiento para el análisis de fiabilidad de los ítems en SPSS	
Procedimiento para el análisis de fiabilidad con el Alfa de Cronbach en	

SPSS	
Bibliografía sugerida	
Capítulo 5. Validez	51
Definición de validez	
Validez de contenido	
Validez de constructo (estructura interna)	
Validez de constructo (asociación con otras variables)	
Conclusiones	
Bibliografía sugerida	
Capítulo 6. Etapas para el desarrollo de escalas de medida	55
Etapa 1. Delimitación de lo que se pretende medir	
Etapa 2. Generación de un banco de ítems	
Etapa 3. Determinación del formato de medición de las respuestas	
Etapa 4. Revisión por expertos (validez de contenido)	
Etapa 5. Validación de la escala	
Etapa 6. Propuesta de la escala final	
Bibliografía sugerida	
Capítulo 7. Análisis Factorial Exploratorio (AFE) con apoyo del SPSS	59
Verificación de los supuestos	
Extracción de los factores	
Rotación de los factores	
Interpretación de los resultados	
Delimitación teórica de los factores	
Procedimiento para el cálculo del AFE en el SPSS	
Consideraciones para el reporte de resultados	
Ejemplo del reporte de un AFE	
Conclusiones	
Bibliografía sugerida	
Capítulo 8. Análisis Factorial Confirmatorio (AFC) con apoyo de AMOS	73
Verificación de los supuestos	
Especificación del modelo	
Identificación estadística del modelo	
Selección del método de estimación	
Especificación del bootstrap	
Definición de parámetros a calcular	
Análisis de los índices de ajuste globales	
Análisis de los índices de ajuste específicos	
Procedimiento para el cálculo del AFC en AMOS	
Ejemplo del reporte de un AFC	
Conclusiones	
Bibliografía sugerida	
Referencias	96
Índice de tablas	99
Índice de figuras	100

Prólogo

Un prólogo es la puerta de entrada al universo conceptual de aquel que construye a través de un imaginario alimentado por las rutinas y sorpresas de la vida cotidiana, una ruta, una historia o un formulario. El prólogo pues se constituye en un elemento fundamental de la obra y por ello la responsabilidad y compromiso de aquellos que se encargan de esta tarea es cada vez más apremiante y a su vez provocadora, estimulante y un reto para el ingenio y la producción de ideas.

Frente a nosotros tenemos una obra denominada “Medición en investigación educativa con apoyo de SPSS y AMOS”. ¿Por qué es importante que el lector se tome su tiempo para escudriñar estas páginas? Primeramente, sería un motivo de elección, el gusto de la persona por la investigación educativa dado que todos los lectores tienen sus preferencias y el acceso al material escrito depende en gran parte de sus necesidades, sus gustos y el disfrute personal hacia la temática y la forma en que está presentada. En un segundo momento una vez que el investigador educativo o en general de las ciencias sociales decide que es de su interés y de su agrado introducirse en la medición, él espera encontrar algunas sorpresas, elementos que en la comprensión del texto le generen interpretaciones diversas y al vincularlos con la lógica propia de su pensamiento le provoquen alternativas innovadoras en su quehacer. En un tercer plano el lector va intentar a través del estudio de la medición en investigación educativa de utilizar lo aprendido para integrarlo todo a su bagaje conceptual y empírico, logrando por medio de la creación de nuevas ideas e imágenes, un mundo icónico más comprensivo, con tareas por emprender y generando innovaciones.

En el primer plano es importante señalar que se trata de un material educativo para los educadores y en general los investigadores en ciencias sociales. Nos educa, porque nos indica no sólo la importancia que tienen los procesos de medición sino adicionalmente nos podemos percatar de la dificultad que asume un investigador de las ciencias sociales al estar tratando con constructos complejos, sumamente variados y en su mayoría de carácter multidisciplinario que en algunas ocasiones son asumidos de manera ligera en la práctica profesional, lo cual trae consigo serios problemas éticos en la práctica y problemas de interpretación que pueden afectar la vida y el destino de las personas que están siendo diagnosticadas, así como los grupos que están siendo tamizados o evaluados.

Así pues, resultará una sorpresa el recorrido a través de los ocho capítulos e ir dando cuenta de las diferentes decisiones, interrogantes y posiciones que debe tomar un sujeto pensante toda vez que pretende describir, comparar, explicar o predecir un concepto en educación o ciencias sociales. El camino como verán no es fácil pero los autores se esmeran en lograr que haya una continuidad en la consecución de ideas y una profundidad mínima considerable para que los alumnos de educación de pregrado o posgrado puedan acceder a este conocimiento como elemento preliminar para despertar el interés la conciencia y el gusto y posteriormente se introduzcan a obras citadas en el texto en donde encontrarán con mayor grado de dificultad y complejidad descritos los momentos y conceptos de la medición en educación.

Con el fin de mantener el interés del lector a través de cada uno de los componentes que constituyen momentos en la toma de decisiones para medir algún concepto relacionado con la educación, los autores elaboran una estrategia que combina la didáctica basada en el ejemplo y en el seguimiento de ciertas reglas preliminares para ir logrando algunos objetivos, se colocan ejemplos en cada uno de los capítulos útiles para ir dilucidando acerca de la toma de decisiones sobre un proceso que finalmente desembocará en análisis de estadísticos de validez y confiabilidad.

Para lograr mantener la motivación y el esfuerzo del lector a través del tiempo se diseñaron los capítulos de manera consecutiva con una lógica adaptada fundamentalmente para personas que tienen un conocimiento elemental o aquellos que tienen un conocimiento promedio de la medición. Lo anterior promueve la permanencia del lector sobre cada uno de estos capítulos y puede ir a través de ellos entendiendo y resolviendo sus dudas, además algunos elementos de procedimiento que le serán útiles si pretende llevar a cabo o tiene la necesidad de medir.

La lógica didáctica del documento les permite también a aquellos profesionales que se encuentran resolviendo problemas en el día a día tanto de educación como de psicología estar en condiciones de poder hacer elecciones de los instrumentos de medida que tengan aquellos indicadores, parámetros o evidencias de validez y fiabilidad que cumplan con los criterios de ajuste para poderse considerar

una prueba fiable y válida. Esto es que el trabajo aquí presentado no solamente va dirigido a quienes buscan medir sino también a todos aquellos terapeutas y educadores que estando en la situación de cambio en el contexto de la profesionalización cotidiana, requieren elegir buenos instrumentos de diagnóstico y evaluación para sus clientes, empresas o escuelas.

Por la diversidad de ámbitos en los que esta obra puede ser útil se evitó colocar ejercicios personalizados a un área de la educación o un escenario específico educativo, con el objeto que el docente de pregrado o posgrado pueda llevar a cabo ejercicios vinculados con el contexto propio de sus alumnos y darle así una utilidad práctica a la obra. Porque tal como lo podemos observar en las asignaturas de psicometría, o se les capacita a los alumnos para aplicar instrumentos de medida sin conocer la forma en que están contruidos, su validez y confiabilidad o su pertenencia a un constructo teórico. Dependiendo del tipo de curso o licenciatura requerirán ejemplos organizacionales, clínicos, escolares sociales demográficos, etc. Entonces el escenario se transforma, pero la teoría de la medida sigue siendo útil en todos y cada uno de estos escenarios.

Es fundamental no solo entender cada uno de los capítulos sino mantener la relación entre ellos porque la medición sigue esa secuencia lógica de las escalas de medición a los estadígrafos de normalidad y de allí a las dimensiones que se proponen para el constructo para continuar a los niveles de fiabilidad y validez y finalmente desarrollar los parámetros de ajuste para el análisis factorial exploratorio y confirmatorio.

En general, tenemos una obra didáctica que será de gran utilidad no solo en las materias de estadística o psicometría en las licenciaturas de ciencias sociales incluyendo la administración, la sociología, la psicología y la educación, sino que además la obra permite una reflexión profunda y analítica de este proceso importante también para los que toman decisiones sobre diagnóstico y evaluación.

Hermosillo, Sonora Octubre del 2019
José Ángel Vera Noriega
Consejo Nacional de Ciencia y Tecnología
Academia Mexicana de la Ciencia

Dictaminadores

Dra. Deneb Elí Magaña Medina
Universidad Juárez Autónoma de Tabasco

Dr. Pedro Antonio Sánchez Escobedo
Universidad Autónoma de Yucatán

Dr. Jesús Tánori Quintana
Instituto Tecnológico de Sonora

Dr. José Ángel Vera Noriega
Centro de Investigación en Alimentación y Desarrollo

Dr. Ernesto Alonso Carlos Martínez
Instituto Tecnológico Superior de Cajeme

Capítulo 1

Medición. Aspectos generales

Gran parte de los individuos en países desarrollados tienen contacto con instrumentos de medición educativa o psicológica. Estas evaluaciones influyen en aspectos relacionados con su educación, salud y oportunidades laborales. En el contexto educativo los resultados de las mediciones se utilizan como base para el diseño de políticas educativas, intervenciones y acciones de capacitación. Por ejemplo, pueden llegar a brindar información para la toma de decisiones acerca del ingreso de estudiantes a las instituciones educativas. En este sentido es frecuente que los posgrados reconocidos en el Padrón Nacional de Posgrados de Calidad (PNPC) en México consideren como uno de sus requisitos de ingreso la obtención de 1000 puntos en el Examen Nacional de Ingreso (EXANI III).

Es indiscutible que las pruebas educativas y psicológicas afectan la vida de muchas personas. Sin embargo, en algunos casos las decisiones que se toman con los resultados de estas pruebas no están fundamentadas en la investigación empírica y la teoría psicométrica. Por ejemplo, no se justifica rechazar a un estudiante en un posgrado por obtener 995 puntos en el EXANI, ya que hasta donde nosotros conocemos no existen estudios sólidos que señalen que los 1000 puntos es lo mínimo necesario para obtener un buen desempeño académico en este nivel educativo.

¿Por qué es importante medir para un investigador?

La investigación educativa se enfoca en la descripción de fenómenos complejos, para lo cual se requiere la medición de constructos que no son directamente observables, tales como la autoeficacia, la orientación al logro y la identidad moral, por mencionar algunos. Para lo cual resulta necesario trabajar en la definición teórica de los constructos, lo que involucra su conceptualización a partir de un marco teórico específico. Así como la definición operacional de los mismos, que implica la identificación de indicadores observables que den cuenta de su existencia en un contexto o situación determinada, y que aporten evidencia incontrovertible de su manifestación.

La definición imprecisa del constructo y de sus indicadores observables lleva al investigador a elaborar conclusiones erróneas acerca de las variables y sus relaciones. Por lo que los investigadores deben ser cuidadosos en la revisión de la definición teórica y operacional de los constructos, ya que en ocasiones son definidos y operacionalizados de forma diferente. Recientemente, en un estudio acerca de tareas escolares enfrentamos un problema de este tipo. Revisando la literatura encontramos resultados contradictorios acerca de los efectos del apoyo parental en las tareas escolares en el desempeño académico (Doctoroff & Arnold, 2017; Núñez *et al.*, 2015; Silinskas & Kikas, 2017; Valle *et al.*, 2016). Al menos en parte, estos hallazgos inconsistentes se originaban por las diferencias en la definición del constructo, ya que mientras algunos autores consideran que el apoyo parental en las tareas involucra acciones que implican el fomento a la autonomía y la auto-determinación del hijo (Tian, Chen, & Huebner, 2014), otros incluyen en este apoyo actividades de los padres enfocadas en el dominio del contenido de las materias (Dumont, Trautwein, Nagy, & Nagengast, 2014).

Constructos e indicadores

Los constructos (o variables latentes) no son directamente observables (ej., *'la inteligencia'*, *'la autoeficacia académica'* y *'la empatía'*), mientras que los indicadores (o variables observables) constituyen conductas específicas con evidencia directa por observación o autoreporte (ej., *'el estudiante pregunta dudas en clase'* o *'me percató cuando algún compañero se siente mal'*).

Como se ha señalado antes, la medición de los constructos implica tanto su definición teórica como operacional. La definición teórica es importante, ya que existen constructos con denominaciones similares que en realidad representan fenómenos distintos (véase comentarios respecto al apoyo en las tareas escolares). Además, permite al investigador evaluar la pertinencia de los indicadores observables seleccionados para dar cuenta de la manifestación del constructo.

El análisis de una escala implica por ende la revisión de la definición teórica

del constructo y los indicadores observables utilizados para medirlo. Ambos aspectos, deben ser reportados en estudios que aborden el desarrollo de escalas de medición. Por ejemplo, García, Valdés, Carlos y Alcántar (2019) diseñaron una escala para medir desconexión moral en niños, para esto desde el punto de vista teórico definieron la desconexión moral como aquellos mecanismos cognitivos que disminuyen el malestar emocional causado por conductas que transgreden normas morales (Bandura, Barbaranelli, Caprara, & Pastorelli, 1996). Con base en esta definición se desarrollaron indicadores observables para medir el grado de expresión del constructo (ver Figura 1).

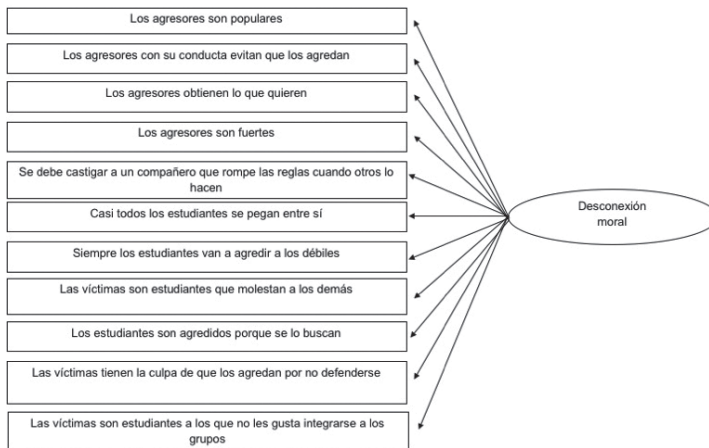


Figura 1. Modelo para medir Desconexión Moral en Niños.

Nota. Tomado de “Propiedades psicométricas de una escala para medir desconexión moral en niños mexicanos” por F. I. García, A. A. Valdés, E. A. Carlos y C. Alcántar, 2019, *Acta Colombiana de Psicología*, 22, p. 110.

La complejidad de los constructos hace necesario utilizar varios indicadores observables para captar la expresión del mismo. Aunque no existe total acuerdo, algunos autores recomiendan que se empleen por menos tres indicadores observables para medir el constructo (DeVellis, 2012; Furr & Bacharach, 2014). La combinación de la respuesta a los indicadores (mínimo tres) puede captar mejor el rasgo (mayor validez de contenido y menor error de medición) que cuando se utilizan uno o dos indicadores. En la Figura 2 se aprecia que los investigadores midieron tres roles de espectadores en el bullying: pro-acoso, pro-social y no comprometido (constructos) mediante 11 indicadores observables que ilustran conductas típicas de cada rol (variables manifiestas u observables).

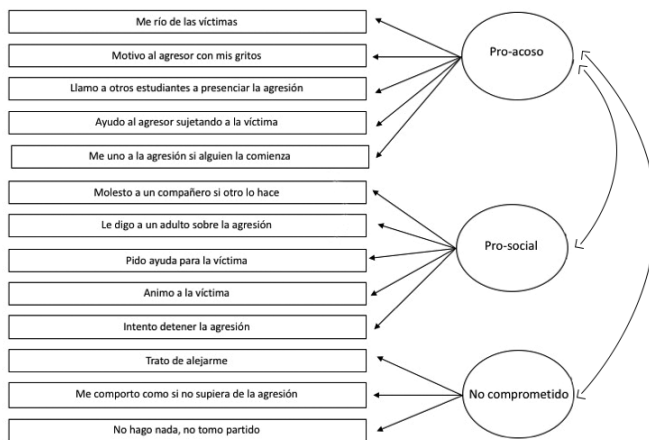


Figura 2. Roles de Espectadores en el Bullying en Niños.

Nota. Adaptado de “Propiedades psicométricas de la adaptación al español de la *Participant role approach*” por C. Alcántar, A. A. Valdés, E. A. Carlos, B. Martínez y F. García, 2018, *Revista Colombiana de Psicología*, 27, p. 182.

Finalmente, es necesario puntualizar que habitualmente los investigadores se interesan más en los constructos que en los indicadores observables; lo cual se debe a que estos tienen menor error de medición. Además, permiten el desarrollo de teorías para explicar los diferentes fenómenos educativos y psicológicos.

Escalas de medición

Las escalas de medición involucran procesos sistemáticos para obtener muestras de conducta con el fin de determinar la expresión de un constructo en un individuo (Anastasi & Urbina, 1997; Cronbach, 1960). Estos permiten identificar diferencias entre los individuos (interindividuales) y en un individuo (intraindividuales) en diferentes circunstancias y puntos de tiempo. Aunque existen muchas clasificaciones de los instrumentos de medición, consideramos relevante para los investigadores educativos distinguir entre dos tipos: de desempeño y de auto-reporte.

Los instrumentos de desempeño miden tareas realizadas por los individuos y suponen son indicadores observables del constructo medido, mientras que los de auto-reporte solicitan al individuo su opinión con respecto a la frecuencia con que realiza cierta conducta o el grado en que concuerda con algunas afirmaciones.

Un constructo puede ser medido con un instrumento de desempeño o de auto-reporte. Cuando un investigador mide el constructo Competencias para la Investigación en Universitarios, con una escala de auto-reporte cuestiona a

los estudiantes acerca del grado en que consideran que poseen determinadas habilidades (ej., *‘puedes calcular un modelo de regresión lineal múltiple’*). Este mismo investigador puede medir el constructo con un instrumento de desempeño cuando le pide a los estudiantes que resuelvan actividades relacionadas con la investigación (ej., *‘calcular un modelo de regresión lineal múltiple’*).

Para los investigadores también es útil diferenciar entre instrumentos referidos a criterios o normas. En los instrumentos referidos a criterios, el investigador basándose en la teoría o la evidencia empírica selecciona un punto de corte para diferenciar a individuos con alto o bajo nivel del rasgo. En contraste, en los referidos a normas los individuos son comparados con muestras normativas, es decir, sus puntajes se comparan con puntajes obtenidos de muestras poblacionales. En este último, es importante señalar que dicha comparación tiene valor cuando la muestra representa a la población de la que proviene el individuo.

Retomando el caso de la medición de las competencias científicas en estudiantes universitarios, el investigador utiliza un instrumento referido a criterio cuando basándose en la norma aprobatoria de las universidades considera que un desempeño mínimamente aceptable son 70 de 100 puntos en la medición; o en base a estadísticos muestrales, por ejemplo, los cuartiles clasifican a los estudiantes como bajo (<25), mediano (>25 y <75) y alto desempeño (>75). Por otra parte, en un instrumento referido a normas el investigador necesita contar normas poblacionales derivadas de la misma población de la que provienen los participantes en el estudio. Si se conoce, por ejemplo, que la norma ubica los puntajes en una media de 80 puntos con una desviación estándar de 5 puntos, una calificación de 93 puntos permite afirmar que los estudiantes de esta universidad se encuentran a una desviación estándar por encima de la media.

Papel de la teoría y la psicometría en la medición

Frecuentemente la medición de variables observables (ej., el *‘sexo’*, la *‘edad’*, la *‘escolaridad’*) no requiere de teoría. Sin embargo, la medición de los constructos debe realizarse desde una perspectiva teórica explícita. Sea o no consciente el investigador, toda medición de un constructo asume cierta perspectiva teórica acerca del mismo.

Es recomendable asumir de forma explícita una postura teórica para desarrollar una escala de medición o estar consciente de cuál fue adoptada para realizar su medición. Por ejemplo, cuando se asume la postura procesal de la inteligencia, los investigadores deben utilizar un instrumento como la Escala de Inteligencia de Weschler (WISC-V) que asume esta teoría y no versiones anteriores de esta escala (WISC-RM) que parten de la teoría factorial de la inteligencia.

Además de la sustentabilidad teórica, una escala debe contar con evidencias empíricas que constaten la calidad de sus mediciones. La psicometría es la

rama de la ciencia donde se generan los procedimientos para evaluar los atributos métricos de las escalas de medición. Al menos dos atributos son esenciales en investigación: (a) la fiabilidad de los puntajes y (b) la validez de los datos obtenidos.

Retos de la medición psicoeducativa

En ocasiones los investigadores minimizan la importancia de la correcta medición de los constructos. Sin embargo, es indudable que una correcta medición incrementa la validez de los hallazgos del estudio. En este sentido, la complejidad de las variables educativas y psicológicas conduce a retos particulares en la medición que afectan en ocasiones la confianza en nuestras interpretaciones (Furr & Bacharach, 2014; Nunnally & Bernstein, 2010).

Un primer reto se refiere a la *reactividad de los participantes*, la cual se deriva del conocimiento de los individuos de los objetivos de la medición. Esta provoca entre otros efectos indeseados la deseabilidad social en las respuestas, la cual implica que el individuo conteste de acuerdo con lo que considera que la sociedad o el investigador espera de él. Por ejemplo, en un cuestionario que mide actitudes hacia las personas homosexuales, las respuestas del individuo pueden ser influenciadas hacia aquellas que le eviten ser catalogado como homofóbico, más que representativas de sus actitudes reales hacia los homosexuales.

Un segundo reto involucra los efectos del sesgo. Este se presenta cuando los resultados de las mediciones son afectados por variables distintas al constructo que se quiere medir. Varios estudios reportan que los resultados en los test estandarizados de inteligencia son más bajos en niños de bajo nivel socio-económico con relación a los de alto nivel socio-económico (Von Stumm & Plomin, 2015); sin embargo, se ha constatado que estos resultados son más afectados por el nivel socio-económico (sesgo) que por el nivel real de habilidades intelectuales del niño (Croizet & Dutrévis, 2004). También se han reportado sesgos de medición asociados al sexo, lo que implica que los resultados son diferentes en hombres y mujeres, independientemente del nivel real con que se presenta el rasgo (Rouquette *et al.*, 2018; Valdés, Carlos, & Torres, 2018).

Un tercer reto se relaciona con la *necesidad de utilizar puntajes compuestos* para obtener información acerca de los constructos. Esta forma de medición presenta debilidades. Por ejemplo, cuando se conforma un puntaje mediante varios ítems, estos tienen igual peso en la expresión del constructo, sin embargo, ¿en realidad todos los ítems informan de manera similar acerca de un constructo?

Conclusiones

En este capítulo se enfatizó la importancia de una adecuada medición para fortalecer la validez de los resultados de investigación. También, se mostró que la

medición se enfoca en constructos que poseen mayor valor teórico y menor error de medida. La calidad de las escalas se vincula con la teoría que los sustenta y con sus propiedades psicométricas. La medición en el campo de la educación tiene retos importantes que se relacionan con la reactividad de los individuos, el sesgo y la necesidad de puntajes compuestos para informar acerca de los constructos.

Bibliografía sugerida

- Beretvas, S. N., Meyers, J. L., & Leite, W. L. (2002). A reliability generalization study of the Marlowe-Crowne Social Desirability Scale. *Educational and Psychological Measurement*, 62(4), 570-589. doi: 10.1177/0013164402062004003
- Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53, 605-634. doi: 10.1146/annurev.psych.53.100901.135 239
- Furr, R. M., & Bacharach, V. R. (2014). *Psychometrics. An introduction* (2nd ed.). Thousand, Oaks, California: SAGE.
- Nunnally, J. C., & Bernstein, I. H. (2010). *Psychometric theory* (3rd. ed.). New York: McGraw-Hill.

Capítulo 2

Conceptos básicos en medición

Escalas de medida

La medición involucra la asignación de números a variables bajo ciertas reglas (Stevens, 1946), con el fin de representar numéricamente la cantidad de los atributos o definir los objetos que pertenecen a ciertas categorías por poseer dichos atributos (Nunnally & Bernstein, 2010). Las escalas de medida constituyen formas específicas de relacionar los números con variables observables para crear una medición (Crocker & Algina, 1986). Stevens (1946) identifica cuatro niveles de medición: nominal, ordinal, intervalo y razón. Recientemente, otros autores agregan un nuevo nivel de medida que denominan como sumativo (Meyers, Gamst, & Guarino, 2013).

Las escalas de medida representan ciertos atributos educativos o psicológicos, que se asocian de forma diferente con los principios numéricos (Furr & Bacharach, 2014). En la Tabla 1 se describe la relación de cada escala de medida con los principios numéricos. Como puede apreciarse la propiedad de cantidad se representa únicamente en las escalas sumativas, de intervalo y razón.

Tabla 1. Relación entre los principios numéricos y el nivel de medición.
 Nota. Adaptado de “*Psychometrics. An introduction*”
 por Furr y Bacharach, 2014, p. 30.

Principios	Nivel de medida				
	Nominal	Ordinal	Sumativa	Intervalo	Razón
Identidad	X	X	X	X	X
Orden		X	X	X	X
Cantidad			X	X	X
Cero absoluto					X
Ejemplo	Estado civil	Nivel de escolaridad	Autoestima	Promedio académico	Edad

Nominal. Es el nivel más básico de medida, los números se utilizan para identificar categorías con base en ciertos atributos. Es decir, identifican grupos de objetos, personas o situaciones que poseen atributos comunes que no se encuentran en otros grupos. Por ejemplo, el estado civil se puede medir agrupando a los individuos en dos categorías: solteros y casados, representados como grupo 0 y grupo 1, respectivamente. Considere que estos grupos también pueden ser denominados como 1 y 2 respectivamente, sin afectar la medición pues como se comentó en este nivel de medición los números no representan cantidades. Asimismo es necesario puntualizar que la utilización de los números para identificar a individuos (número de pasaporte) u objetos (número de vivienda) no constituyen mediciones nominales, ya que no conforman grupos con base en ciertas propiedades que los diferencian de otros.

Ordinal. En las escalas ordinales los números tienen la propiedad del orden. Los objetos, situaciones o personas se organizan en un rango de acuerdo al grado en que presentan el atributo medido. Por ejemplo, en un estudio el nivel socio-económico de las familias es rankeado a juicio de los padres como bajo (1), medio (2) o alto (3) según los recursos económicos disponibles.

Como se comentó anteriormente, los números en esta escala no representan cantidades solo informan acerca de la posición relativa con base al grado que se manifiesta un atributo. Estos números no permiten inferencias precisas acerca del cambio en los atributos. Los números informan únicamente que un atributo se presenta más en un objeto, situación o persona que en otro similar.

Si bien este nivel de medición aporta más información que el nominal, es

también limitado. Una limitación es, por ejemplo, que aun cuando dos familias se clasifiquen en bajo nivel socioeconómico, éstas pueden diferir sustancialmente en sus recursos económicos cuando se realiza la investigación en Suecia o Haití. Como es visible en este nivel de medición es poco factible realizar comparaciones entre los hallazgos de estudios realizados en diferentes contextos, para llevar a cabo estas comparaciones se requiere que los números posean la propiedad de la cantidad.

Sumativa. Esta escala parte del supuesto de que los puntajes compuestos de ítems ordinales se comportan como escalas de intervalo, cuando tienen cuatro o más opciones de respuesta (Furr & Bacharach, 2014; Meyers *et al.*, 2013). En ésta los individuos asignan valores a los constructos basados en un continuo definido por el rango de la escala de respuesta. Los números son organizados habitualmente en orden creciente de respuesta con cinco o más opciones, para que reflejen el nivel del atributo. Por ejemplo, la sumatoria de los puntajes de los 10 ítems de un cuestionario que mide autoestima (0 = nada de acuerdo hasta 4 = totalmente de acuerdo). Se asume que mayores valores representan mayor grado de expresión del constructo, en este caso la autoestima.

Estas escalas se denominan como sumativas debido a que se promedian las respuestas a los ítems para obtener un puntaje global. Aunque estrictamente hablando este tipo de escala no cumple con los criterios necesarios para ser considerada una escala de intervalo, ya que no existen iguales distancias entre números adyacentes, se ha constado que el promedio de los ítems tiene una interpretación significativa cuando es tratado como de intervalo (Darlington & Hayes, 2017; Heck, Thomas, & Tabata, 2014; Meyer *et al.*, 2013). Esto implica asumir que los números representan cantidades del atributo medido.

Intervalo. En esta escala los números poseen la propiedad de la cantidad y los espacios entre ellos son constantes. Sin embargo, el cero es arbitrario, lo que implica que no se puede hablar de proporciones, es decir, aunque permiten afirmar que un estudiante que obtiene 80 puntos en prueba de habilidades matemáticas tiene 40 puntos más que otro que alcanzó 40 puntos, no se puede considerar que el primero posee el doble de habilidades que el segundo. Estrictamente hablando existen pocas mediciones en escalas de intervalo en la investigación psicoeducativa. Aunque por lo general se asume que las pruebas que miden desempeño académico y habilidades intelectuales se encuentran en este nivel de medida.

Razón. En esta escala, la más poderosa, el número tiene la propiedad de la cantidad y el cero no es arbitrario. Algunas medidas como la velocidad de res-

puesta, el ingreso económico y el número de hijos son ejemplos de este nivel de medición. Algo importante de señalar es que aunque el cero absoluto existe en algunas variables su presencia no es factible en la medición de atributos concretos, por ejemplo, aunque el tiempo de respuesta tiene un cero absoluto posible en realidad no es factible que un individuo no utilice tiempo en una tarea.

Diferencias individuales

Los constructos representan rasgos que varían entre las personas (ej., '*Inteligencia*', '*Motivación*', '*Autoestima*'), lo cual implica que en la población existen diferentes niveles de expresión de los mismos. Si se mide inteligencia, implica considerar que existen individuos con poca inteligencia, con una inteligencia mediana y con una elevada inteligencia. Una buena escala de medición debe ser capaz de captar la mayor variabilidad posible en la expresión del constructo.

De esta postura se deduce que para medir Habilidades Matemáticas donde la media de calificación de los estudiantes fue de 97 puntos ($DE = 3$; máximo 100 puntos), capta menos variabilidad que otro donde la media fue de 84 puntos ($DE = 4$). En el primer instrumento, la mayor parte de los estudiantes (68%) obtienen calificaciones entre 94 y 100 puntos, lo cual dificulta identificar a aquellos con bajo y alto nivel de habilidades matemáticas, ya que por ejemplo, cualquier calificación menor a 94 implicaría, un pobre nivel de habilidades matemáticas ¿se puede considerar un pobre nivel de habilidades matemáticas obtener 93 puntos?, por otra parte, resulta imposible seleccionar a estudiantes con altas habilidades matemáticas, debido a que el rango de la calificación promedio incluye el valor máximo de la escala.

En una buena medición, sobre todo si la muestra es grande, las variables se deben distribuir de forma normal (ver Figura 3). Asumiendo normalidad, se puede decir que la media representa el puntaje del 50% de los individuos. También se conoce que el 68% se encuentran entre -1 y +1 desviación estándar de la media. Puntajes mayores a +1 desviación estándar implica que el individuo tiene una elevada expresión de rasgo medido, por ejemplo, si medimos inteligencia, implica que se sitúa entre el 16% de las personas más inteligentes.

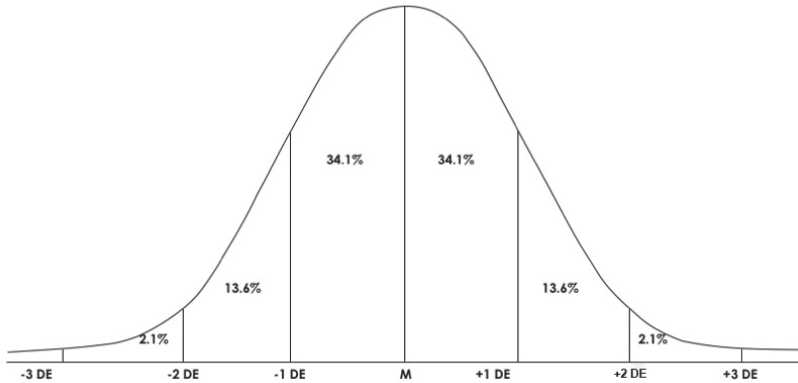


Figura 3. Distribución normal.

Medidas de normalidad

Las aseveraciones anteriores, acerca de la distribución de los individuos, se pueden realizar cuando los datos se distribuyen normalmente. La normalidad es además un presupuesto de varias técnicas estadísticas bivariadas y multivariadas utilizadas para determinar las propiedades psicométricas de las escalas. Dada la importancia de la normalidad en la medición se han desarrollado diversas medidas para evaluarla.

Asimetría y curtosis. La asimetría es una medida de normalidad de la distribución de los datos. En una distribución simétrica, existe el mismo número de valores a la derecha que a la izquierda de la media, por lo tanto, son iguales los valores con signo positivo que negativo. Existen dos tipos de asimetría: negativa, cuando la cola de la simetría se alarga para valores inferiores a la media y positiva, cuando la cola se alarga hacia valores superiores a la media (ver Figura 4).

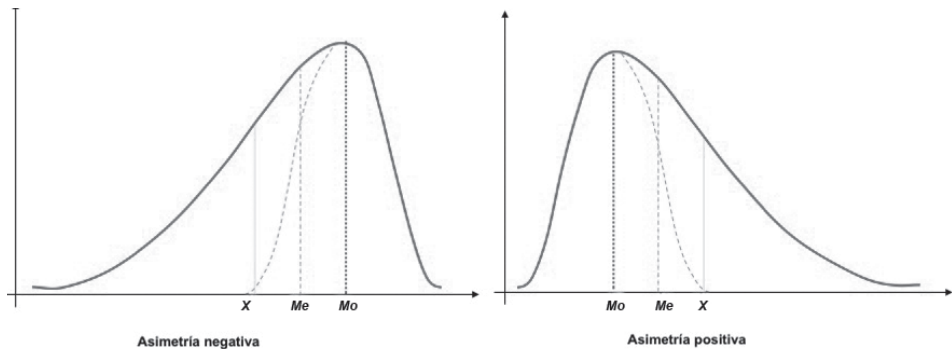


Figura 4. Tipos de asimetría.

Algunos autores, consideran que valores de asimetría entre -1 y +1 son indicadores de normalidad (Darlington & Hayes, 2017; Heck *et al.*, 2014). Sin embargo, existen mediciones más precisas, como el coeficiente de asimetría de Pearson y la medida de Yule Bowley o medida cuartilica.

$$\text{Coeficiente de asimetría de Pearson } As = 3 \frac{(M - Md)}{s}$$

Donde M = media aritmética; Md = mediana; s = desviación típica.

El coeficiente varía entre -3 y +3

$As < 0$ la distribución es asimétrica negativa

$As = 0$ la distribución es simétrica

$As > 0$ la distribución es simétrica positiva

$$\text{Medida de Yule Bowley o medida cuartilica } As = \frac{Q1 + Q2 - 2 Q2}{Q3 - Q1}$$

Donde $Q1$ = cuartil uno, $Q2$ = cuartil 2, $Q3$ = cuartil 3

La medida varía entre -1 y 1

$As < 0$ la distribución es asimétrica negativa

$As = 0$ la distribución es simétrica

$As > 0$ la distribución es simétrica positiva

Por su parte, la curtosis mide cuán apuntada o achatada es la distribución de los puntajes alrededor de la media; mientras más datos existan alrededor de la media más apuntada es la curva. Existen tres tipos de curtosis: leptocúrtica, los datos están concentrados en la media, siendo una curva muy apuntada; mesocúrtica, igual a la normal; y platicúrtica, poca concentración de los datos alrededor de la media, lo que ocasiona una forma achatada de la curva (ver Figura 5).

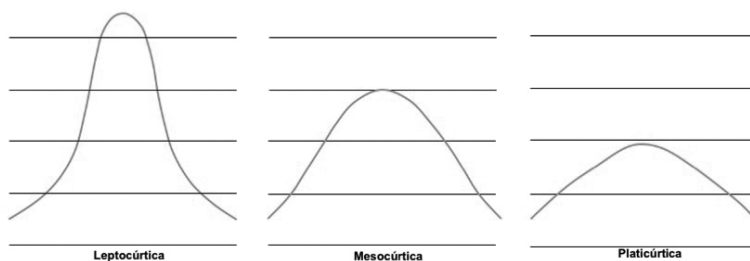


Figura 5. Tipos de curtosis.

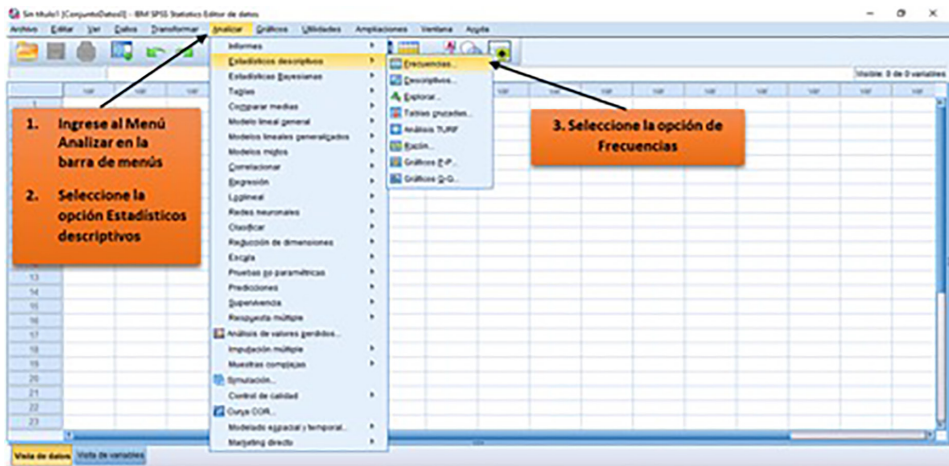
En general, los autores coinciden que valores de curtosis en el rango entre -1 y 1 sugieren una distribución cercana a la normal. Existen medidas más precisas, aunque complicadas en su cálculo, como el coeficiente de curtosis de Fischer.

Coeficiente de curtosis de Fisher

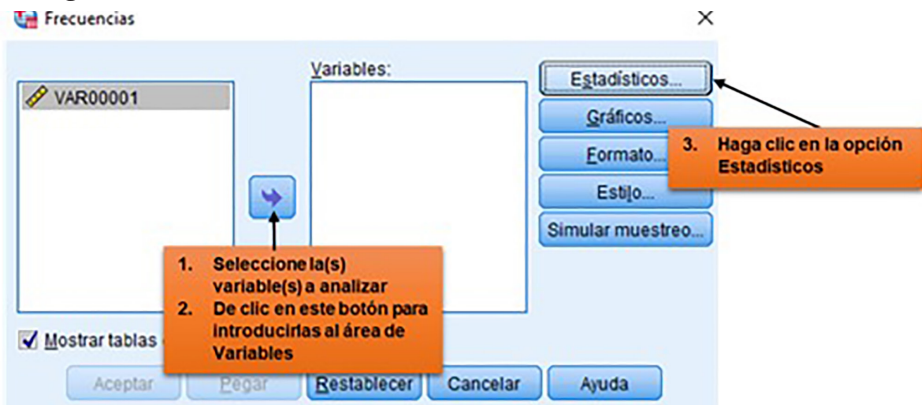
$$g^2 = \frac{(1/n)\sum (X_i - M)^4 n - 3}{(1/n)\sum (X_i - M)^2}$$

Procedimiento en el SPSS para calcular asimetría y curtosis

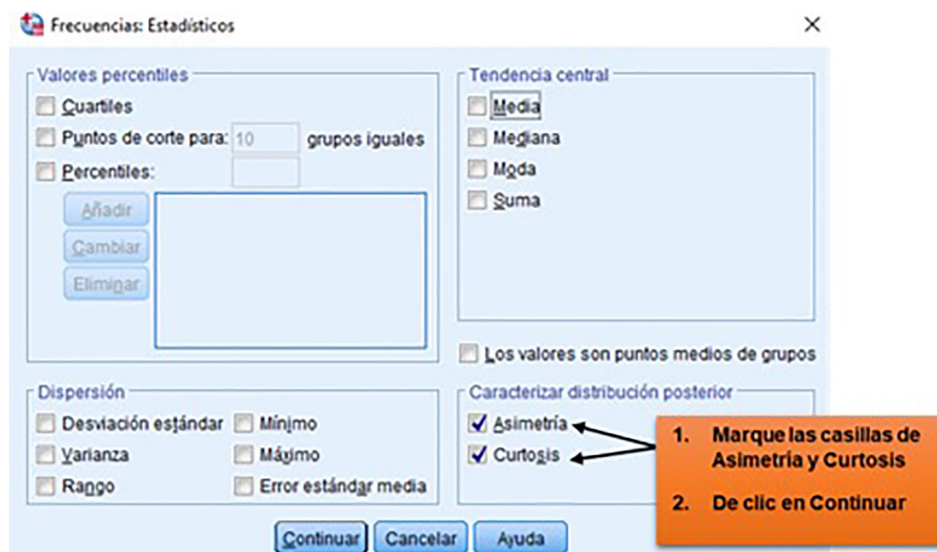
Como se muestra en la imagen, siga la ruta iniciando en la Barra de Menús: *Analizar* → *Estadísticos descriptivos* → *Frecuencias*



Dentro de la ventana de *Frecuencias*, seleccione la variable a analizar y pásela al área de *Variables* dando clic en la flecha como se muestra en la imagen. Posteriormente, ingrese a *Estadísticos* en el menú de la derecha.



Dentro de la ventana de *Estadísticos*, ubíquese en el área de *Caracterizar distribución posterior* y elija las opciones de *Asimetría* y *Curtosis*, luego de clic en *Continuar*. En la siguiente ventana solo de clic en *Aceptar*.



Pruebas de normalidad

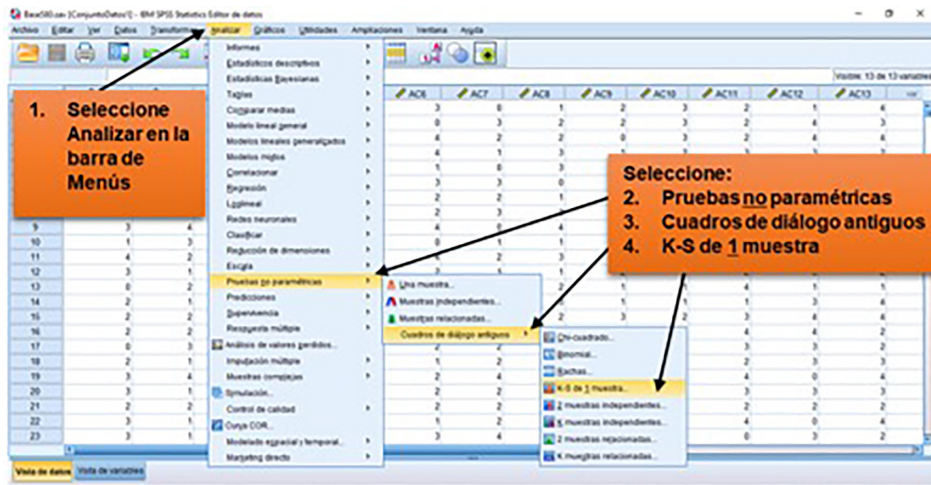
Las pruebas de Kolmogorov-Smirnov (KS), Kolmogorov-Smirnov con la corrección de Lilliefors (KSL) y Shapiro-Wilks (SW) se utilizan para contrastar la distribución de un conjunto de datos con la distribución normal. La prueba de Shapiro-Wilks se utiliza cuando la muestra es menor a 50 casos, el test de Kolmogorov-Smirnov es recomendable utilizarlo con más de 50 observaciones. Para atenuar problemas identificados en la prueba de KS se desarrolla la corrección de Lilliefors (KSL), la cual se recomienda utilizar en los análisis de normalidad. Estas son pruebas de hipótesis, donde valores de $p > 0.05$ sugieren que los datos proceden de una distribución normal (se acepta la hipótesis nula).

H0: los datos proceden de una distribución normal.

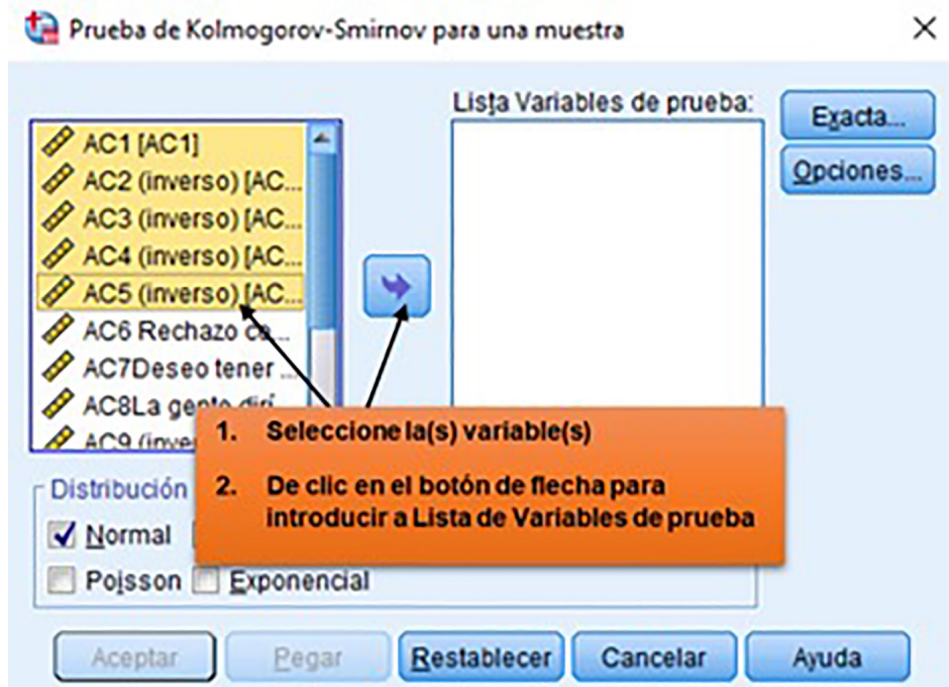
H1: los datos no proceden de una distribución normal.

Procedimiento en el SPSS para calcular pruebas de normalidad

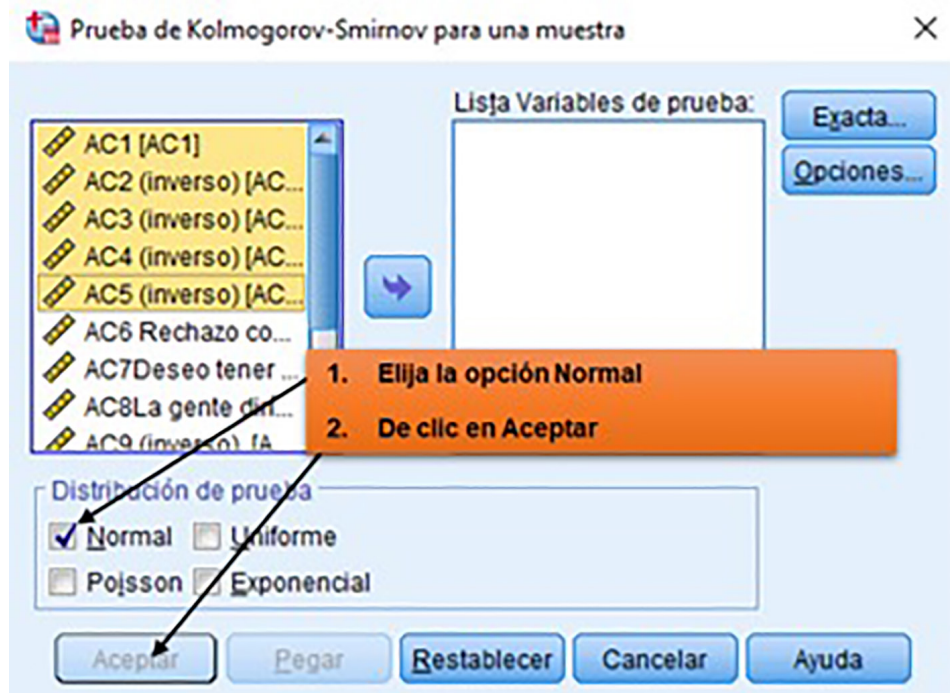
Prueba de Kolmogorov-Smirnov (KS). Como se muestra en la imagen, siga la ruta iniciando en la Barra de Menús: *Analizar* → *Pruebas no paramétricas* → *Cuadros de diálogo antiguos* → K-S de 1 muestra.



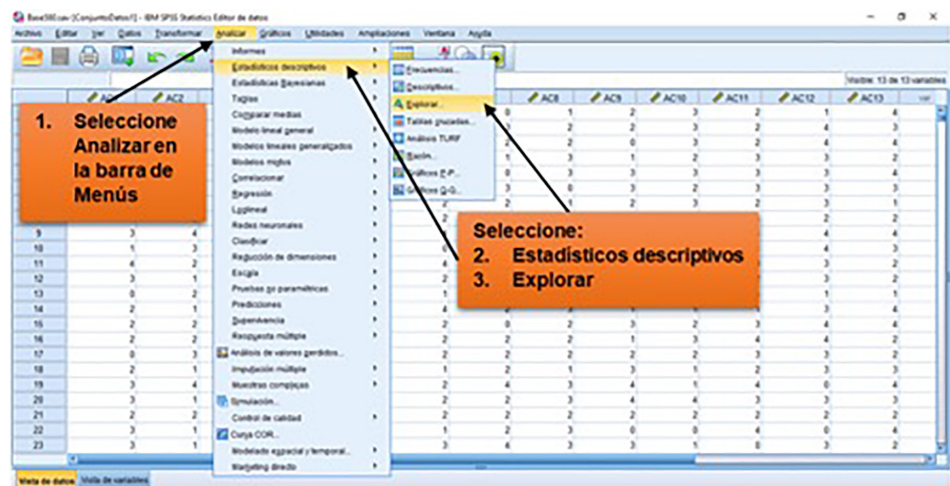
Dentro de la ventana de K-S de 1 muestra (Prueba de Kolmogorov-Smirnov para una muestra), seleccione la(s) variable(s) a analizar y pásela(s) al área de *Lista de Variables de Prueba* dando clic en la flecha como se muestra en la imagen.



En esta misma ventana, ubíquese en el área de *Distribución de prueba* y seleccione la opción *Normal*, luego de clic en *Aceptar*.



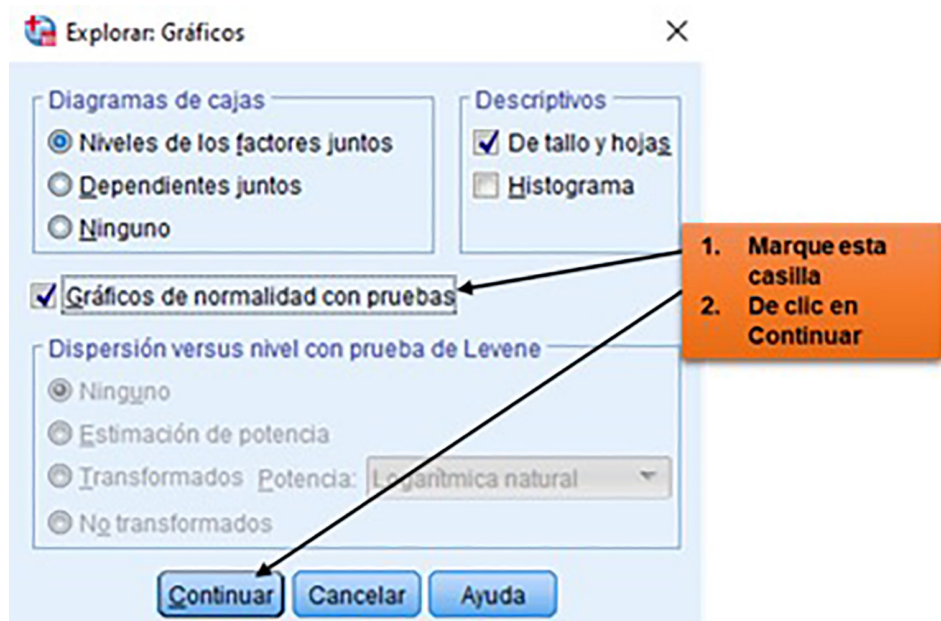
Para calcular la Prueba de Kolmogorov-Smirnov con la corrección de Lilliefors (KSL), siga la ruta iniciando en la Barra de Menús: *Analizar* → *Estadísticos descriptivos* → *Explorar*.



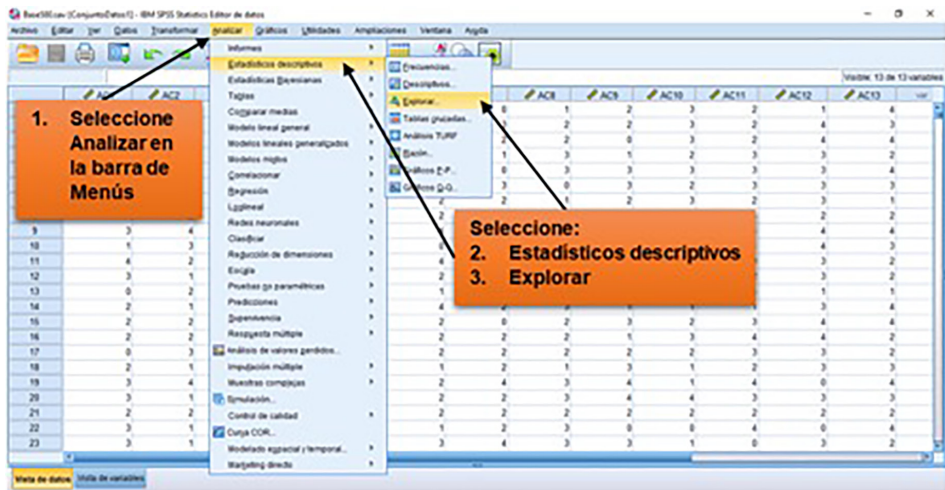
Dentro de la ventana de *Explorar*, seleccione la(s) variable(s) a analizar y pásela(s) al área de *Lista de dependientes* dando clic en la flecha como se muestra en la imagen. Posteriormente, ingrese a *Gráficos* en el menú de la derecha.



Dentro de la ventana de *Gráficos*, seleccione la opción de *Gráficos de normalidad con pruebas* y luego de clic en *Continuar*. En la siguiente ventana solo de clic en *Aceptar*.



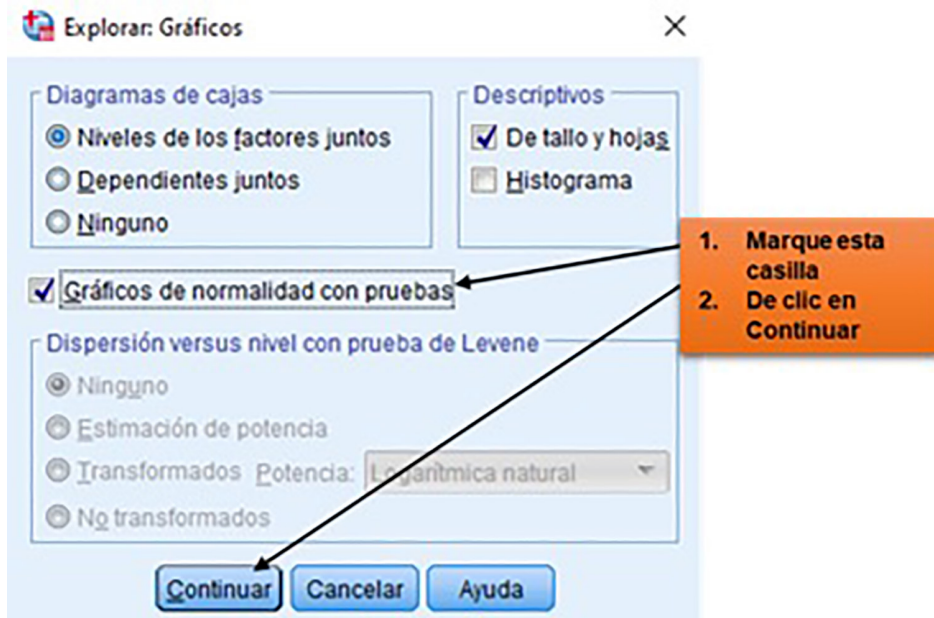
Para calcular la Prueba Shapiro-Wilks (SW), siga la ruta iniciando en la Barra de Menús: *Analizar* → *Estadísticos descriptivos* → *Explorar*.



Dentro de la ventana de *Explorar*, seleccione la(s) variable(s) a analizar y pásela(s) al área de Lista de dependientes dando clic en la flecha como se muestra en la imagen. Posteriormente, ingrese a *Gráficos* en el menú de la derecha.



Dentro de la ventana de *Gráficos*, seleccione la opción de *Gráficos de normalidad con pruebas* y luego de clic en *Continuar*. En la siguiente pantalla solo de clic en *Aceptar*.



Nota: Se observa que el procedimiento para llegar al cálculo de las Pruebas Kolmogorov-Smirnov con la corrección de Lilliefors (KSL) y Shapiro-Wilks (SW) es el mismo, sin embargo, en la pantalla de *Resultados* se muestra información de ambas pruebas, como se puede ver en la Tabla 2.

Tabla 2. Resultado de pruebas de normalidad.

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Medios_Consola_12_e	.472	916	.000	.531	916	.000

a. Corrección de significación de Lilliefors

Medidas de tendencia central y dispersión

Las medidas de tendencia central y dispersión brindan información acerca de la distribución de los puntajes en el constructo medido (ver Tabla 3). Los estadísticos de tendencia central (media, mediana y moda) informan acerca de lo

que ocurre en el centro de la distribución normal, indican el valor promedio que toma la distribución del rasgo. La media es el estadístico de tendencia central más ampliamente utilizado. Sin embargo, cuando existen problemas de normalidad, la mediana es un estadístico más preciso.

Las medidas de dispersión (varianza, desviación típica, rango y cuartiles) indican cómo se alejan los datos respecto de la media. La desviación típica es particularmente importante cuando la distribución de los datos es normal, ya que permite conocer la proporción de los individuos en cada desviación estándar.

Tabla 3. Estadísticos de tendencia central y dispersión.

		Definición	Símbolo
Tendencia central	Media	Promedio o la media aritmética del grupo.	M
	Mediana	Valor medio cuando los datos se organiza de menor a mayor.	Md
	Moda	Valor que se repite con mayor frecuencia.	Mo
Dispersión	Varianza	Variabilidad de los puntajes con relación a la media	s^2
	Desviación estándar	Es la raíz cuadra de la varianza.	de
	Rango	Es el intervalo entre el valor mínimo y máximo en una distribución de puntajes.	R
	Cuantil	Espacios en que se divide la distribución de los puntajes.	Q

Teoría clásica de medición

Los procedimientos descritos en este texto para determinar las propiedades psicométricas de las escalas se enmarcan dentro de la teoría clásica de medición. En esta teoría se presupone que el puntaje observable de un constructo es resultado de la puntuación verdadera más el error de medida (ver Figura 6). En otras palabras, se sostiene que el puntaje de los ítems es resultado de la variable latente y el error de medida.

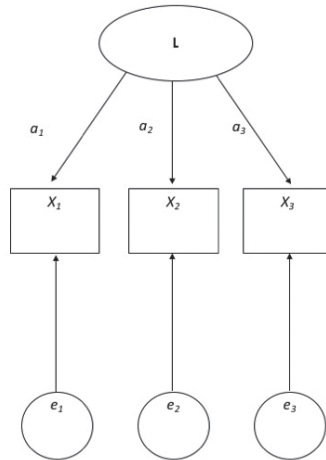


Figura 6. Modelo clásico de medida.

En sus inicios, los modelos desarrollados dentro de esta teoría suponían que la cantidad de error de los ítems individuales es aleatoria y que los términos de error entre los ítems no se correlacionan. Recientemente, han surgido otros modelos, como el congénico (Jöreskog, 1971) que asume que todos los ítems no tienen la misma relación con la variable latente, el error de medición no es aleatorio y que los errores de los ítems pueden estar correlacionados. Este modelo es ampliamente utilizado en la actualidad, por ejemplo, es la base de la técnica de Análisis Factorial Confirmatorio.

Conclusiones

En este capítulo se definió la medición como un proceso donde, bajo ciertas reglas, se asignan números a los atributos. Con relación a sus propiedades numéricas las escalas de medida se clasifican como nominales, ordinales, sumativas, de intervalo y de razón; y que únicamente en las escalas sumativa, de intervalo y razón el número posee la propiedad de la cantidad.

También, se analizó el papel de las diferencias individuales en la medición, afirmando que una buena escala de medida debe captar la variabilidad del constructo entre las personas o en una misma persona a través del tiempo. En general, se considera que la variabilidad en los constructos sigue una distribución normal. La normalidad, un supuesto de muchas técnicas estadísticas, se puede calcular por procedimientos, tales como asimetría y curtosis, o pruebas de hipótesis (KS, KSL o SW).

Finalmente, se presentaron estadísticos que permiten conocer la distribución de los puntajes. Éstos incluyen medidas de tendencia central (media, mediana y moda) que informan acerca de lo que sucede en el centro de la distribución; y de dispersión (varianza, desviación estándar, rango y cuantil), que describen como varían los puntajes del promedio.

Bibliografía sugerida

- DeVellis, R. F. (2012). *Scale development* (3rd ed.). Thousand, Oaks, California: SAGE.
- Harwell, M. R., & Gatti, G. G. (2001). Rescaling ordinal data to interval data in educational research. *Review of Educational Research*, 71(1), 105-131. doi: 10.3102/00346543071001105
- Meyers, L. S., Gamst, G., & Guarino, A. J. (2013). *Applied multivariate research. Design and interpretation* (2nd ed.). Thousand, Oaks, California: SAGE.
- Norman, G. (2010). Likert scales, level of measurement and the “laws” of statistics. *Advances in Health Sciences Education*, 15(5), 625-632. doi: 10.1007/s10459-010-9222-y

Capítulo 3

Dimensionalidad de las escalas de medición

Imagine que un grupo académicos quiere utilizar un cuestionario para medir el liderazgo en directores escolares que incluye tres estilos: transformacional, transaccional y *laissez-faire*. Un grupo de investigadores sostiene que las respuestas a los ítems se deben combinar para dar información acerca del estilo de liderazgo. Sin embargo, otro grupo que disiente de esta propuesta, sostiene que cada estilo conforma una dimensión independiente, lo que implica que cada puntuación se debe tratar por separado.

Esta situación ilustra el tema de la dimensionalidad de las mediciones, la cual es esencial en el desarrollo e interpretación de los resultados de las escalas. Los investigadores en el estudio acerca del liderazgo antes de tomar una decisión deben responder varias preguntas, por ejemplo, ¿existe una característica general que subyace a los estilos directivos medidos?, ¿cuál es la correlación entre los puntajes de los diferentes estilos?

Con respecto a la dimensionalidad existen tres temas básicos. Un primer tema, involucra cuestionarse acerca de ¿cuántas dimensiones reflejan los ítems de la escala? Esta pregunta es importante, ya que cada dimensión requiere un análisis psicométrico por separado. Un segundo tema se relaciona con indagar los casos en que existe más de una dimensión ¿en qué grado dos dimensiones se correlacionan? La respuesta a esta pregunta permite determinar cuándo el

puntaje global de la escala tiene significado. Asimismo como tercer tema, es necesario preguntarse cuándo existe multidimensionalidad, ¿cuántas dimensiones posee la escala?

Escalas unidimensionales

Las escalas unidimensionales se conforman por ítems homogéneos, lo que implica que la respuesta a los ítems se relaciona únicamente con el constructo medido (ver Figura 7). En estas escalas las propiedades psicométricas (validez y fiabilidad) se analizan basándose en un puntaje único obtenido de los ítems. Además, los investigadores deben interpretar únicamente el puntaje global de la escala. Un dato importante es que aun cuando desarrolladores de escalas propongan estructuras multidimensionales para medir el constructo, la existencia de correlaciones extremadamente altas entre las dimensiones sugiere unidimensionalidad.

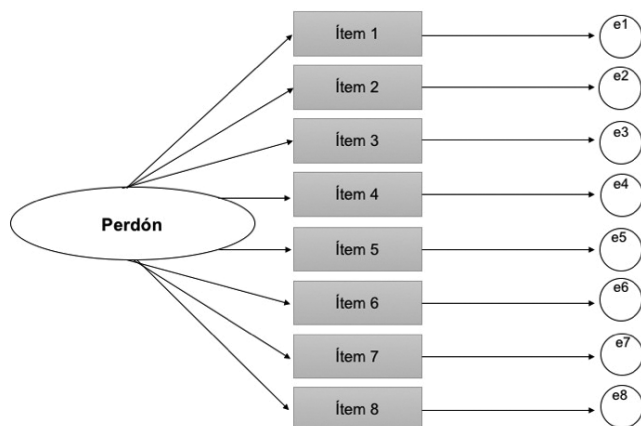


Figura 7. Modelo unidimensional de medición del Perdón en adolescentes.

Escalas multidimensionales con dimensiones correlacionadas

Como se comentó antes, cuando la escala es multidimensional es necesario determinar el grado de correlación entre sus dimensiones. Cuando las dimensiones se encuentran moderadamente correlacionadas ($\geq .40$ y $\leq .60$), estamos frente a una escala multidimensional con dimensiones correlacionadas (ver Figura 8).

En este tipo de escala se considera la existencia de una dimensión de orden superior que explica los puntajes de las dimensiones particulares. Por ejemplo, cuando los autores encuentran correlaciones entre los puntajes de los tipos de Ci-

beragresión (Denigración y Usurpación de identidad) pueden inferir la existencia de una dimensión de orden superior (Ciberagresión).

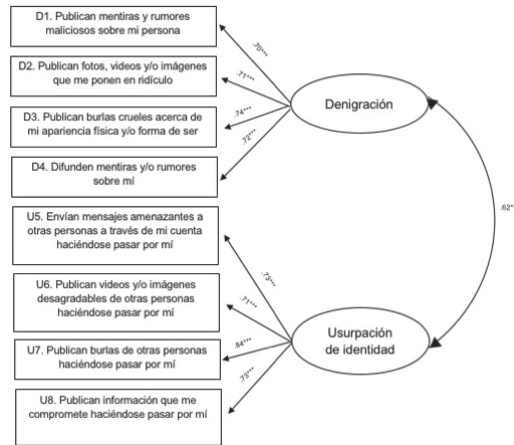


Figura 8. Escala de Ciberagresión con dos dimensiones correlacionadas.

Nota. Tomado de “Propiedades psicométricas de una escala para medir cibervictimización en universitarios” por A. A. Valdés, E. A. Carlos y G. M. Torres, 2018, *Revista Electrónica de Investigación Educativa*, 20, p. 39.

Las escalas multidimensionales con dimensiones correlacionadas producen diversos puntajes, habitualmente se analizan los puntajes de cada dimensión por separado, ya que en principio cada dimensión es unidimensional al comprender ítems homogéneos. Además, es posible formar un puntaje global a partir de las dimensiones individuales. En el caso anterior, los investigadores pueden considerar tanto los puntajes de cada tipo de ciberagresión por separado como el puntaje global de Ciberagresión. En este tipo de escalas las propiedades psicométricas se determinan para cada dimensión, ya que es posible encontrar propiedades psicométricas adecuadas para una dimensión y deficientes para otra. Además, es necesario analizar la calidad de las propiedades psicométricas del puntaje global.

Escalas multidimensionales con dimensiones no correlacionadas

Cuando una escala es multidimensional no tiene correlaciones entre sus dimensiones o son bajas ($< .20$), entonces se puede decir que es una escala multidimensional con dimensiones no correlacionadas (ver Figura 9). Aquí, cada dimensión es medida por un conjunto homogéneo de ítems, sin embargo, su escasa correlación sugiere que no existe un factor de orden superior que explique las dimensiones.

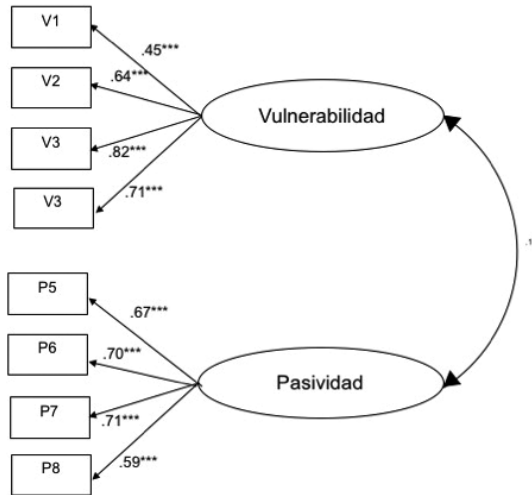


Figura 9. Escala de Dificultades en Habilidades Sociales con dimensiones no correlacionadas.

Nota. Adaptado de “Propiedades psicométricas de una escala para medir dificultades en habilidades sociales relacionadas con la victimización” por A. A. Valdés, E. J. Madrid, E. A. Carlos y B. Martínez, 2016, *Pensamiento Psicológico*, 14, p. 82.

Con respecto a su uso, estas escalas son similares a los multidimensionales con dimensiones correlacionadas, con la excepción de que los puntajes de cada dimensión no pueden combinarse para formar un puntaje global. La calidad de las propiedades psicométricas se determina para cada escala por separado.

Procedimientos para determinar la dimensionalidad

Aunque existen varios procedimientos para determinar la dimensionalidad de las escalas (por ejemplo, escalamiento múltiple y análisis de conglomerados), el análisis factorial es el más ampliamente utilizado. Existen dos grandes tipos de análisis: Análisis Factorial Exploratorio (AFE) y Análisis Factorial Confirmatorio (AFC). Para el AFE están disponibles software como el SPSS y el SAS, y para el AFC software como AMOS, Mplus y Lisrel.

Conclusiones

En este capítulo se analizó la importancia de determinar la dimensionalidad de las escalas de medición. En este sentido, existen escalas unidimensionales, multi-

dimensionales con dimensiones correlacionadas y multidimensionales con dimensiones no correlacionadas. Se describió cómo se analizan las propiedades psicométricas y se interpretan los puntajes de los diferentes tipos de escalas. Finalmente, se refieren que tanto el Análisis Factorial Exploratorio (AFE) y el Análisis Factorial Confirmatorio (AFC) son métodos para el análisis de su dimensionalidad.

Bibliografía sugerida

DeVellis, R. F. (2012). *Scale development* (3rd ed.). Thousand, Oaks, California: SAGE.

Furr, R. M., & Bacharach, V. R. (2014). *Psychometrics. An introduction* (2nd ed.). Thousand, Oaks, California: SAGE.

Martínez, M. R., Hernández, M. J., & Hernández, M. V. (2006). *Psicometría*. Madrid: Alianza Editorial.

Capítulo 4

Fiabilidad

Definición de fiabilidad

Las escalas de medida deben ser capaces de captar la variabilidad en la expresión de los constructos. Los investigadores utilizan los resultados de las mediciones para determinar el nivel del rasgo en los individuos y establecer relaciones entre constructos. Para ambos objetivos es esencial que las escalas reflejen de forma real los constructos medidos, lo cual es precisamente el tema fundamental de la fiabilidad.

La fiabilidad en la teoría clásica de medida refleja el grado en que las diferencias entre los individuos en las mediciones son resultado del puntaje verdadero y no de características desconocidas (error de medida). Por ende, la fiabilidad es influida por el grado en que las respuestas a la escala son resultado de diferencias interindividuales o intraindividuales reales en los individuos y no son función de un error de medida.

Los investigadores deben considerar que la fiabilidad no es un asunto de todo o nada, es decir, una escala no es totalmente fiable o no fiable. Esta cualidad de la medición es un continuo, lo que implica definir las mediciones como más o menos fiables. En general, la fiabilidad se determina por métodos estadísticos relacionados con la varianza y la correlación. Existen dos procedimientos esta-

dísticos para determinar la fiabilidad: 1) la varianza explicada y 2) la correlación entre los ítems de la escala.

Varianza explicada

Este procedimiento implica definir la fiabilidad como la proporción de la varianza que es explicada por la puntuación verdadera ($R = S_v/S_e$). Cuando tenemos un $R = .60$, entonces se puede afirmar que el 60% de las diferencias entre los individuos se debe a la puntuación verdadera. R es un coeficiente que oscila entre 0 y 1, donde 0 indica la ausencia de fiabilidad y 1 una fiabilidad perfecta. Ambos extremos son prácticamente inalcanzables, por lo tanto, se ha establecido que un porcentaje de varianza atribuida al puntaje verdadero superior a .50 es aceptable, siendo deseable un porcentaje igual o superior a .70 (Cea, 2004; Furr & Bacharach, 2014; Nunnally & Bernstein, 2010).

Correlación entre los puntajes de los ítems

Consistencia interna. Se relaciona con la correlación entre los puntajes observados y los verdaderos. Existen diversos estadísticos para determinar este tipo de fiabilidad (Alfa de Cronbach, Omega de McDonald, Theta Ordinal y Fiabilidad Compuesta). Los valores de estos coeficientes oscilan entre 0 y 1, donde 0 indica ausencia de fiabilidad y 1 una fiabilidad perfecta, donde valores de .70 o mayores son aceptables (George & Mallery, 2003; Martínez, Hernández, & Fernández, 2006).

Estabilidad temporal. La correlación entre los puntajes de los ítems también puede evaluarse mediante la estabilidad temporal de los mismos. Éste parte del supuesto de que cuando se mide adecuadamente el constructo debe mantenerse relativamente constante los puntajes a través del tiempo. Algunos autores señalan que existen problemas en la mediación de la fiabilidad temporal. Es difícil determinar cuándo el cambio ocurre en realidad por la modificación del constructo a través del tiempo, se debe al método utilizado en la evaluación o a la falta de fiabilidad temporal de la prueba (DeVellis, 2012).

Factores que afectan la fiabilidad

Como se ha visto, la fiabilidad es una propiedad importante de las escalas de medida, ya que informa la precisión de la medida del constructo. Existen dos factores que afectan la fiabilidad de las escalas, el primero se relaciona con la consistencia interna entre los ítems de la escala, una mayor consistencia conduce al incremento de la fiabilidad; y el segundo con la longitud del test, se afirma que siendo los demás factores similares mientras más largo es un test mayor será su fiabilidad.

Esto implica que un investigador puede mejorar la fiabilidad buscando ítems más homogéneos que favorezcan la consistencia interna de los puntajes o en su caso agregando más ítems a la escala original. Nosotros recomendamos la primera opción, ya que en la investigación es conveniente poseer escalas fiables que midan de forma parsimoniosa los constructos.

Costos de una pobre fiabilidad

Una baja fiabilidad conduce a una medición poco precisa del constructo, lo que implica que gran parte del puntaje obtenido de la misma es error. Esto por supuesto afecta la calidad de la investigación, ya que genera dudas acerca de que los resultados informen acerca de la variable medida. Por otra parte, una baja fiabilidad debilita las asociaciones entre las variables estudiadas. El grado de atenuación depende de la fiabilidad, a menor fiabilidad menor es la relación encontrada entre las variables. Como resultado, las pruebas de significatividad y el tamaño del efecto se vinculan con la fiabilidad. Los efectos de la fiabilidad en la significatividad y el tamaño del efecto hacen que esta sea un elemento que afecte los resultados de los estudios. Una pobre fiabilidad puede sesgar el estudio conduciendo al investigador a interpretaciones erróneas de sus resultados.

Medidas de fiabilidad

Existen muchas medidas de fiabilidad, los investigadores deben utilizar la que mejor se vincule con el tipo de estudio que realizan. Los métodos más frecuentes son aquellos basados en la consistencia interna. Algunos de los estadísticos utilizados con frecuencia son el Coeficiente Alfa de Cronbach, el Coeficiente Omega de McDonald, la Varianza Media Extractada y la Fiabilidad compuesta.

En SPSS únicamente está disponible el cálculo del Coeficiente Alfa de Cronbach. Para otros estadísticos como el Omega de McDonald (más efectivo en escalas ordinales como las de tipo Likert) existen otros softwares como el JASP. Finalmente, el cálculo de Fiabilidad compuesta y la Varianza extractada se realiza de forma manual o con apoyo de plantillas de Excel.

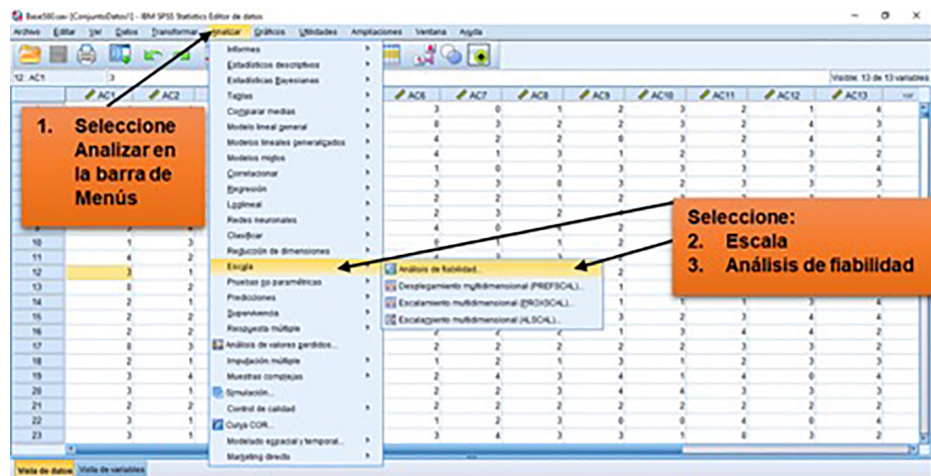
Coeficiente Alfa de Cronbach. En primera instancia, se sugiere realizar un análisis de la fiabilidad de los ítems, indagando la correlación entre cada ítem con el puntaje verdadero de la escala. Esto permite conocer si existen ítems que afectan la fiabilidad global, en general se recomienda no incluir en el análisis a ítems con fiabilidad menor a .20 (Nunnally & Bernstein, 2010; Martínez *et al.*, 2006) y a aquellos que tengan una relación negativa con el puntaje verdadero de la escala (los ítems inversos deben ser recodificados antes de los análisis).

Posteriormente, se debe realizar el análisis de fiabilidad. Cuando una escala

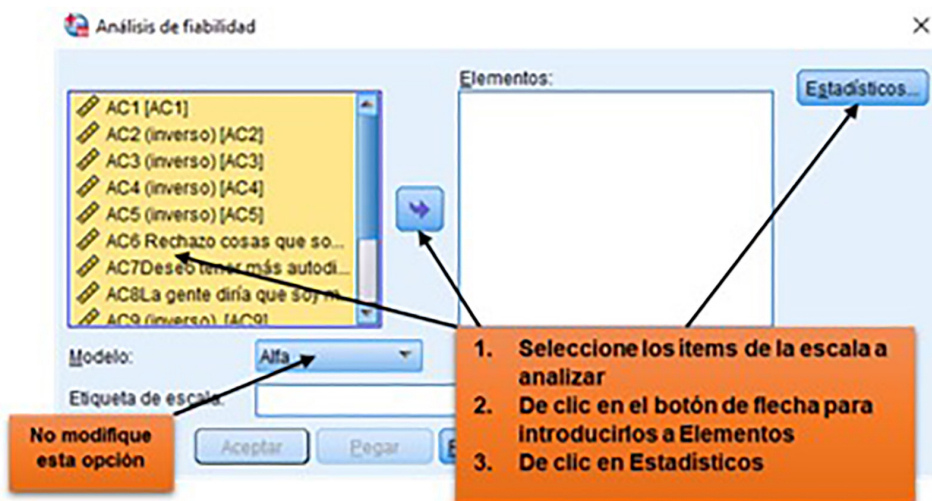
es multidimensional con dimensiones correlacionadas, el análisis de fiabilidad se reporta para cada dimensión y de forma global; mientras que cuando la escala es multidimensional con dimensiones no correlacionadas, la fiabilidad se reporta para cada dimensión, pero no de forma global; y, por último, cuando la escala es unidimensional solo se reporta la fiabilidad global de la escala. Recuerde que se consideran aceptables valores de fiabilidad iguales o superiores a .70 (George & Mallery, 2003; Martínez *et al.*, 2006).

Procedimiento para el análisis de fiabilidad de los ítems en SPSS

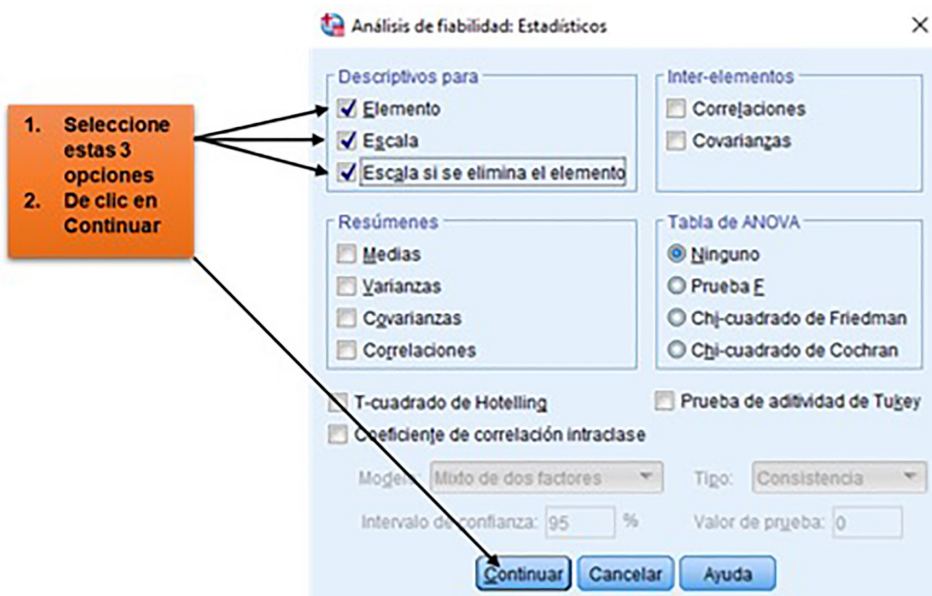
Como se muestra en la imagen, siga la ruta iniciando en la Barra de Menús: *Analizar* → *Escala* → *Análisis de fiabilidad*.



Dentro de la ventana de *Análisis de fiabilidad*, seleccione los ítems de la escala a analizar y páselos al área de *Elementos* dando clic en la flecha como se muestra en la imagen. Considere no modificar el *Modelo: Alfa* de la parte inferior. Posteriormente, de clic en *Estadísticos* en el menú de la derecha.

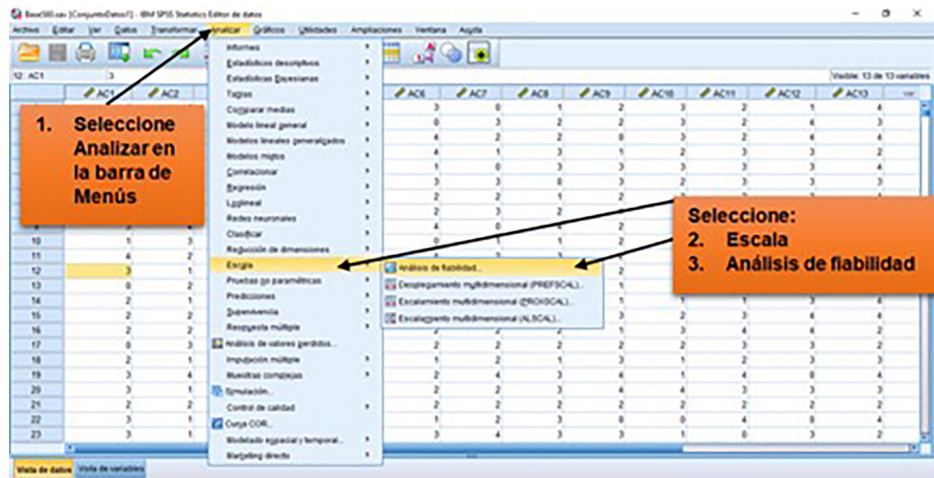


Dentro de la ventana de *Estadísticos*, ubíquese en el área de *Descriptivos para*, y seleccione las opciones de *Elemento*, *Escala* y *Escala si se elimina el elemento*, luego de clic en *Continuar*. En la siguiente pantalla solo de clic en *Aceptar*.

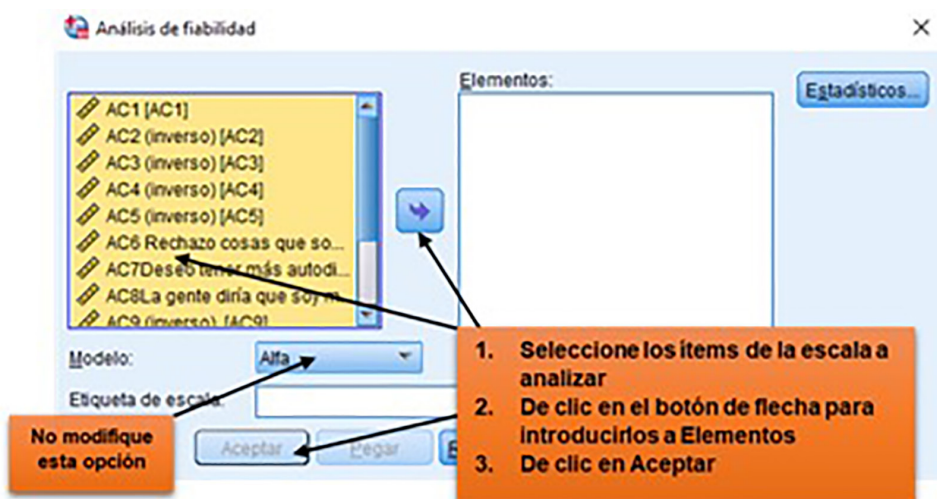


Procedimiento para el análisis de fiabilidad con el Alfa de Cronbach en SPSS

Como se muestra en la imagen, siga la ruta iniciando en la Barra de Menús: *Analizar* → *Escala* → *Análisis de fiabilidad*.



Dentro de la ventana de *Análisis de fiabilidad*, seleccione los ítems de la escala a analizar y páselos al área de *Elementos* dando clic en la flecha como se muestra en la imagen. Considere no modificar el Modelo: *Alfa* de la parte inferior. Posterior, de clic en *Aceptar*.



Bibliografía sugerida

- Cortina, J. M. (1993). What is coefficient Alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98-104. doi: 10.1037/0021-9010.78.1.98
- Furr, R. M., & Bacharach, V. R. (2014). *Psychometrics. An introduction* (2nd ed.). Thousand, Oaks, California: SAGE.
- Nunnally, J. C., & Bernstein, I. H. (2010). *Psychometric theory* (3rd. ed.). New York: McGraw-Hill.
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's Alpha. *Psychometrika*, 74(1), 107-120. doi: 10.1007/s11336-008.9101-0

Capítulo 5

Validez

Definición de validez

De forma sencilla la validez implica que la escala mida lo que dice medir. En este mismo sentido la validez se define como el grado en que la evidencia y la teoría sostiene la interpretación de los puntajes obtenidos (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999).

Los investigadores deben considerar que la validez no se refiere al instrumento sino a la interpretación de los puntajes de los mismos. Esto implica que la validez de la interpretación de los puntajes de una escala en una población específica, no se puede generalizar cuando ésta se utiliza en otro contexto. También, es necesario tener en cuenta que la validez no es asunto de todo o nada, sino de más o menos evidencias. Aunque existen múltiples evidencias de validez, en este texto nos referiremos a las más utilizadas en la investigación psicoeducativa: contenido, estructura interna y de criterio.

Validez de contenido

En este tipo de validez se busca precisar un acuerdo entre el contenido del test y el constructo que se mide (Furr & Bacharach, 2014). La validez de contenido

evalúa el grado en que los ítems son adecuados para medir el constructo (Polit & Beck, 2004, p. 423). Esta evidencia de validez es afectada por la incorporación de ítems irrelevantes. También, es limitada por la baja representatividad del constructo, lo que implica que si bien los ítems son pertinentes para su medición no resultan suficientes para medir todo el espectro del mismo.

Análisis de la validez de contenido. El juicio de expertos es el método utilizado por los investigadores para analizar la validez de contenido de la escala. En éste se le pide a un grupo de individuos con experiencia práctica o de investigación en la temática que valoren de forma cualitativa o cuantitativa el contenido de las definiciones y la pertinencia de los ítems para medir el constructo en un contexto determinado (DeVellis, 2012; Escobar-Pérez & Cuervo-Martínez, 2008).

Las valoraciones de los expertos se pueden realizar cualitativamente o cuantitativamente. En la valoración cualitativa los expertos emiten opiniones acerca de aspectos tales como la pertinencia contenido, la adecuación cultural, la claridad y la suficiencia de los ítems. De forma cuantitativa se les pide a los expertos que valoren los ítems como más o menos pertinentes, más o menos claros, para posteriormente calcular la concordancia entre sus valoraciones mediante el Coeficiente Kappa (respuestas dicotómicas), el Coeficiente de concordancia W de Kendall (respuestas ordinales) o el Índice de Validez de Contenido (ICV).

Validez de constructo (estructura interna)

Este tipo de validez se refiere a la pertinencia de la estructura interna (dimensionalidad) de la escala. El análisis de este tipo de validez permite aclarar el número de factores dentro de los ítems, determinar la asociación entre los factores en los test multidimensionales y precisar la relación de los ítems con los factores. El Análisis Factorial Exploratorio (AFE) y el Análisis Factorial Confirmatorio (AFC) son dos técnicas estadísticas utilizadas para determinar la estructura interna de las escalas. Dado que ambos tipos de análisis serán abordados en capítulos posteriores, aquí solo queremos mencionar que el AFC tiene mayor sustento en la teoría que el AFE.

Validez de constructo (asociación con otras variables)

Este tipo de evidencia de validez se basa en determinar si los resultados de la escala tienen las relaciones esperadas con otras variables. Existen varias formas de evaluar esta forma de validez: convergente, discriminante, concurrente y predictiva.

Validez convergente. La validez convergente es el grado en que los puntajes de la escala se relacionan con los puntajes de otras escalas que miden constructos similares. Se pueden relacionar los puntajes con las escalas que miden el mismo constructo, por ejemplo, se relacionan los puntajes de una escala que fue desarrollada para medir Desconexión Moral en Niños con otras previas reportadas en la literatura.

También, el investigador puede relacionar los puntajes de la escala con otras que miden constructos asociados. Así, por ejemplo, con base en la teoría un investigador puede esperar que la Autoeficacia Académica se relacione con la Autorregulación del aprendizaje y la Orientación al logro. Esto le permite evaluar la validez convergente de la escala relacionándolo con puntajes de otras que miden estos constructos.

Validez discriminante. Este tipo de evidencia de validez se obtiene cuando los resultados no se correlacionan con otros constructos con los cuales, de acuerdo con la teoría, no debe correlacionarse. Por ejemplo, un investigador cuando desarrolla una escala para medir Inteligencia Emocional espera basándose en la teoría que sus puntajes tengan escasa correlación con los resultados de un Test de Habilidades Cognitivas. Cuando ocurre lo contrario, se puede decir que no posee evidencias de validez discriminante, ya que también está midiendo habilidades intelectuales.

Validez concurrente. La validez concurrente es el grado en el cual los resultados de la escala correlacionan con otra variable que es medida en el mismo tiempo. Por ejemplo, cuando un investigador quiere evaluar la validez concurrente de una escala para medir Pensamiento Crítico en estudiantes universitarios puede relacionarlo con el desempeño académico actual de los estudiantes, ya que con base en la teoría se esperarían relaciones positivas entre ambas mediciones.

Validez predictiva. Esta validez se evalúa mediante el grado de relación de los puntajes de la escala con una variable que es medida en un tiempo posterior. Por ejemplo, se espera que los puntajes de una escala que mide Habilidades Intelectuales utilizado antes del ingreso a la universidad se relacione con el desempeño de los estudiantes en el primer semestre de la licenciatura.

Conclusiones

En este capítulo se define la validez como el grado en que la evidencia empírica y la teoría sustenta las interpretaciones de los puntajes de los instrumentos de medición. Se enfatiza que la validez es una propiedad de los puntajes y

que no es asunto de todo o nada, sino de más o menos evidencias. Se analizan algunas de las evidencias más utilizadas de validez en la investigación psicoeducativa: contenido, estructura interna y relaciones con otras variables.

Bibliografía sugerida

- Borsboom, D., Mellenbergh, G. J., & van Herdeem, J. (2004). The concept of validity. *Psychological Review*, 111(4), 1061-1071. doi: 10.1037/0033-295X.111.4.1061
- Escobar-Pérez, J., & Cuervo-Martínez, A. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances en Medicina*, 6, 27-36. Recuperado de humanas.unal.edu.co/psicometria/files/71113/8574/5708/Articulo3_Juicio-de-expertos_27-36-pdf
- Furr, R. M., & Bacharach, V. R. (2014). *Psychometrics. An introduction* (2nd ed.). Thousand, Oaks, California: SAGE.
- Nunnally, J. C., & Bernstein, I. H. (2010). *Psychometric theory* (3rd. ed.). New York: McGraw-Hill.

Capítulo 6

Etapas para el desarrollo de escalas de medida

Antes de comenzar a describir las etapas para el desarrollo de escalas de medición psicoeducativas creemos conveniente puntualizar que optar por diseñar un nuevo instrumento debe ser la última opción del investigador. En la mayor parte de los casos con una buena revisión de literatura se identifican escalas previamente desarrolladas para medir el constructo, en este caso los investigadores pueden optar por utilizar la versión original o realizar adaptaciones del mismo.

El diseño de nuevas escalas es necesario cuando no se encuentran opciones publicadas en la literatura para medir el constructo, se identifican debilidades teóricas en el contenido de los mismos, los existentes no reportan adecuadas propiedades psicométricas o se pretende medir desde una teoría que no es utilizada en otras propuestas. En dichos casos, es necesario para el investigador desarrollar escalas originales para el estudio, para lo cual DeVellis (2012) sugiere que se sigan varias etapas:

Etapas 1. Delimitación de lo que se pretende medir

En esta etapa el investigador debe precisar con claridad el constructo que desea medir, los objetivos de la escala y la población en la cual se realizará

la medición. Un aspecto conveniente en este momento es la selección de la teoría desde la cual se medirá el constructo. Si no existe o el investigador no pretende utilizar una teoría previa entonces debe especificar su propio marco teórico de referencia.

Posteriormente, cada constructo y dimensiones de la escala deben ser conceptualmente definidos con base en el marco teórico utilizado para la medición. Es necesario que estas definiciones sean claras y precisas antes de operacionalizar los constructos. Resulta importante, que la definición del constructo permita distinguirlos de otros similares.

Etapa 2. Generación de un banco de ítems

Cuando se desarrolla una nueva escala es más conveniente contar con múltiples ítems que poseer un número limitado de los mismos. Es importante que los ítems abarquen todo el aspecto de manifestación del constructo, incluso alguna redundancia puede ser deseable siempre y cuando no sea resultado de aspectos gramaticales o de vocabulario. No existe un acuerdo en cuanto al número de ítems a incluir en esta fase de desarrollo, nuestra sugerencia es incluir todos aquellos que se consideren relevantes para medir el constructo desde el marco teórico que se utiliza. Una recomendación práctica es incluir un mínimo de 6 a 8 ítems para medir cada constructo.

Se debe procurar que los ítems posean: (a) un contenido relevante para medir el constructo, (b) brevedad, (c) claridad, (d) una sola idea y (d) pertinencia para la cultura, la edad y el nivel educativo. Un aspecto discutido es la conveniencia de incluir ítems inversos en la medición, es decir ítems cuya respuesta indica un bajo grado de expresión del constructo, por ejemplo, si se mide conducta prosocial un ítem que diga 'le pego a mis compañeros' es inverso. Aunque muchos autores sugieren su inclusión como una forma de evitar la deseabilidad social en las respuestas, nosotros coincidimos con DeVellis (2012) en que no es recomendable su inclusión dado que pueden confundir a los participantes.

Etapa 3. Determinación del formato de medición de las respuestas

Aunque existen muchos formatos de respuestas (Thurstone, Guttman, Likert and Diferencial Semántico) nosotros nos referiremos únicamente a los formatos Likert por ser los más utilizados en la investigación psicoeducativa. Las escalas tipo Likert son ampliamente discutidas en la literatura, en este formato de respuesta los ítems son presentados como una sentencia seguidos por opciones de respuesta que indican cierto grado de acuerdo o desacuerdo con la misma o la frecuencia con que realizan cierto tipo de acciones.

Es conveniente que el investigador utilice opciones que tengan similares intervalos entre ellas, por ejemplo, 0 (*Fuertemente en desacuerdo*), 1 (*Moderadamente en desacuerdo*), 2 (*Ligeramente en desacuerdo*), 3 (*Ligeramente en acuerdo*), 4 (*moderadamente de acuerdo*) y 5 (*Fuertemente de acuerdo*). Es necesario puntualizar que no existe previo acuerdo con respecto al número de opciones que se deben utilizar en este tipo de escalas, ya que este parece relacionarse con el constructo medido y el contexto en que es utilizado el instrumento (Chang, 1994; Nunnally & Bernstein, 2010; Wu & Leung, 2017). Sin embargo, nosotros recomendamos que se utilicen más de cuatro opciones para facilitar los análisis estadísticos.

Etapas 4. Revisión por expertos (validez de contenido)

En esta etapa los investigadores, a través del juicio de experto (ver capítulo de validez), evalúan la validez de contenido de la escala. Es necesario que el instrumento posea este tipo de evidencia de validez antes de administrarlo.

Etapas 5. Validación de la escala

En este momento el investigador analiza las propiedades psicométricas de la escala (ver capítulos 4 y 5). Habitualmente es necesario presentar evidencias de fiabilidad y validez (estructura interna y relación con otras variables).

Etapas 6. Propuesta de la escala final

Una vez establecidas las evidencias de validez y fiabilidad el investigador presenta la versión final de la escala. Es necesario que se reporte: (a) el objetivo de la escala, (b) las características de la población en la que se utilizó, (c) la definición de constructos y dimensiones, (d) número de ítems que miden cada constructo y (e) las distintas evidencias de fiabilidad y validez.

Bibliografía sugerida

- DeVellis, R. F. (2012). Scale development. *Theory and applications* (3rd ed.). Thousand, Oaks, California: SAGE.
- Friborg, O., Martinussen, M., & Rosenvinge, J. H. (2006). Likert-based vs. semantic differential-based scoring of positive psychological constructs: A psychometric comparison of two versions of a scale measuring resilience. *Personality and Individual Differences*, 40, 873-884. doi: 10.1016/j.paid.2005.08.015

- Nunnally, J. C., & Bernstein, I. H. (2010). *Psychometric theory* (3rd ed.). New York: McGraw-Hill/Pearson.
- Olaogun, M. O. B., Adedoyin, R. A., Ikem, I. C., & Anifaloba, O. R. (2004). Reliability of rating low back pain with a visual analogue scale and a semantic differential scale. *Physiotherapy Theory and Practice*, 20, 135-142. doi: 10.1080/0959398490453048
- Wu, H., & Leung, S-O. (2017). Can Likert scales be treated as interval scales? - A simulation studies. *Journal of Social Service Research*, 43(4), 527-532. doi: 10.1080/01488376.2017.1329775

Capítulo 7

Análisis Factorial Exploratorio (AFE) con apoyo del SPSS

Cuando se analiza una escala de medida es importante determinar su dimensionalidad o el número de factores que posee, ya que es poco factible medir adecuadamente un constructo sin conocer sus dimensiones. El análisis factorial es una técnica que permite al investigador poner a prueba la dimensionalidad de un instrumento. En esta técnica se distinguen dos aproximaciones: el Análisis Factorial Exploratorio (AFE) y el Análisis Factorial Confirmatorio (AFC).

El AFE tiene como finalidad la búsqueda de dimensiones por medio de las correlaciones entre los ítems o variables observadas (Martínez *et al.*, 2006). Este análisis no parte de un número previo de factores, sino que se realiza con el fin de conocer los factores que surgen de un grupo de variables observables (Cea, 2004). El AFE comprende las siguientes fases:

Verificación de los supuestos

En la primera fase se eligen las variables que se analizarán y se procura que la muestra posea un tamaño adecuado. Debido a que este análisis es una técnica multivariable de interdependencia se requiere que el tamaño de la muestra sea elevado. El mínimo necesario ronda entre los 200 y 300 casos, aunque para

lograr una mayor precisión debe tenerse en cuenta el número de variables a analizar (Cea, 2004). Es además deseable que las variables cumplan con los supuestos de normalidad multivariada y linealidad con el fin de que el análisis sea más preciso, de no ser así se espera que los ítems se correlacionen moderadamente entre sí (Méndez & Rondón, 2012).

Extracción de los factores

En la segunda fase se determina el número de factores a extraer, mediante la regla del autovalor o el gráfico de sedimentación (Ho, 2006), lo cual se explica más adelante en este capítulo. En el análisis se puede dejar libre el número de factores o seleccionar un número fijo (Field, 2009). Es importante establecer el método de extracción de los factores, dos de los métodos más comunes son: (a) ejes principales, cuyo propósito es exponer la varianza común o compartida entre las variables y (b) máxima verosimilitud, que requiere que las variables cumplan con el supuesto de normalidad y a diferencia de los métodos anteriores, sus resultados pueden generalizarse. Este método permite corroborar la bondad de ajuste de un modelo factorial específico al de una matriz muestral (Cea, 2004).

Aunque es una opción en la mayoría de los programas, como por ejemplo el SPSS, el Análisis de *Componentes principales* no es propiamente un método de Análisis Factorial (Martínez *et al.*, 2006), pero es una técnica que se puede emplear para la reducción de ítems. El método de *Componentes principales*, parte de la idea de que no es necesario que los factores a extraer tengan validez teórica y se emplea sobre todo cuando se requiere reducir el número de ítems para que representen los factores mínimos necesarios (Ho, 2006).

Rotación de los factores

En la tercera fase del análisis se selecciona la técnica de rotación de los factores, esto es imprescindible para obtener una solución factorial más clara. La decisión depende de si se espera que los factores estén relacionados o no. En caso de que no exista fundamentación teórica que sustente la existencia de relación de los factores se debe seleccionar la rotación ortogonal, cuando se suponga que los factores están relacionados se emplea la rotación oblicua.

Rotación ortogonal. El SPSS presenta tres métodos de rotación ortogonal: varimax, quartimax y equamax (Field, 2009). El procedimiento más empleado es el varimax que tiene como objetivo maximizar la varianza de los coeficientes cuadrados para que su interpretación sea más sencilla (Cea, 2004).

Rotación oblicua. Mientras que el enfoque ortogonal tiene varias opciones dentro del programa SPSS, el enfoque oblicuo se limita al método Oblimin, el cual parte de que los factores se correlacionan (Ho, 2006).

Interpretación de los resultados

Índices de ajuste global. Es conveniente realizar un análisis preliminar de la matriz de correlaciones, lo que permitirá ayudar a identificar si es posible realizar el AFE. En caso de que las correlaciones sean muy bajas, es muy poco posible que existan factores comunes. Se deben verificar los valores del índice *KMO* (Kaiser-Meyer-Olkin) y del test de esfericidad de Bartlett. El índice *KMO* tiene valores entre 0 y 1, cuanto más próximo esté a 1 mejor es para la factorización (Field, 2009). Por su parte, el test de esfericidad de Bartlett realiza una estimación de la significación de la matriz de correlaciones. Se espera un *p*-valor menor que 0,001. Si obtiene este valor la matriz es adecuada y puede procederse a la extracción de factores (Martínez *et al.*, 2006).

Análisis de las communalidades. La comunalidad se define como la proporción de varianza explicada por los factores comunes (Ho, 2006). Es importante que después de la extracción las communalidades continúen presentando valores altos (Pérez, 2004). Cuanto más cercanas estén las communalidades a 1 mejor serán los factores para explicar los datos, en general se aceptan comunalidad mayores a .30 (Field, 2009).

Número de factores. El número de factores se determina siguiendo dos reglas principales. La primera indica que sólo los factores con autovalores con valor de 1 o mayores son significativos, los que posean valores menores son descartados. La segunda regla emplea la representación gráfica de los autovalores, llamada gráfico de sedimentación, se emplea para identificar la cantidad de factores que se pueden extraer. Se considera que el punto en el que la curva comienza a enderezarse indica el número máximo de factores a extraer (Ho, 2006).

Fiabilidad de la medición. Para esto se utiliza la varianza total explicada que expresa el porcentaje de la varianza acumulado de factores sucesivos, es decir el porcentaje de varianza que se explica por este factor y los anteriores a él. En ciencias sociales se espera un porcentaje de varianza acumulado de 60% e incluso valores menores en ciertos casos (Cea, 2004).

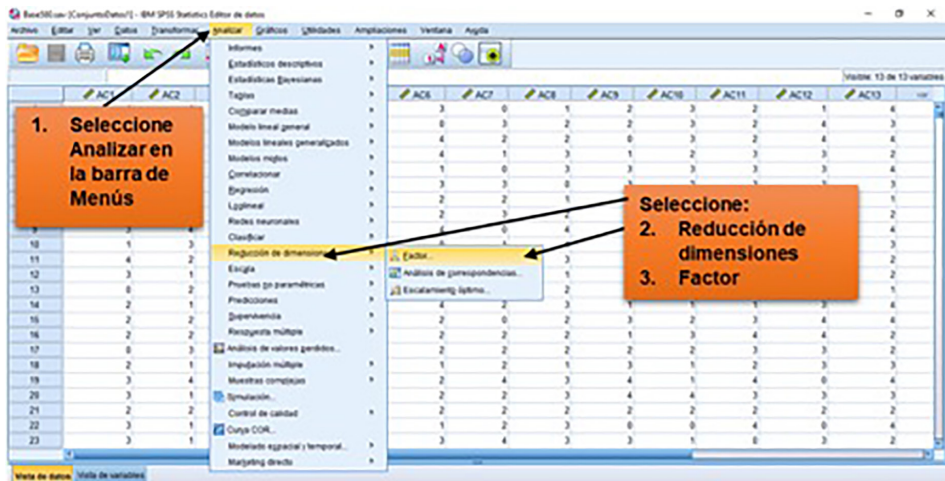
Análisis de la matriz factorial. La matriz factorial contiene los factores y los ítems que los componen, los primeros se sitúan en la columna de la matriz y los segundos en las filas (Cea, 2004). La tabla incluye los pesos factoriales, los cuales indican que tan relacionadas están las variables con cada factor (Ho, 2006). Algunos autores afirman que cargas mayores a 0.6 asocian a la variable con el factor (Pérez, 2004), mientras que otros aseveran que mayores 0.4 son suficientes (Stevens, 2002). Es necesario verificar que los ítems posean cargas factoriales iguales o mayores a .40 en un solo factor, ya que un ítem que posee cargas altas en varios factores afecta los futuros análisis estadístico. Éste, por lo general, es un ítem confuso teóricamente.

Delimitación teórica de los factores

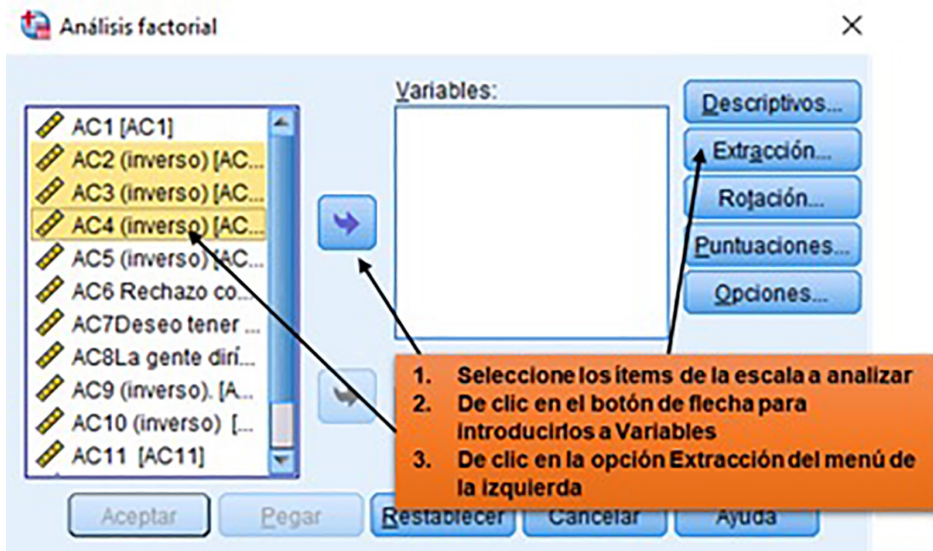
La última fase consiste en la interpretación teórica de los factores, la cual se realiza tomando en cuenta la teoría existente acerca de los constructos evaluados. En esta fase, se deben conceptualizar los factores desde el punto de vista teórico.

Procedimiento para el cálculo del AFE en el SPSS

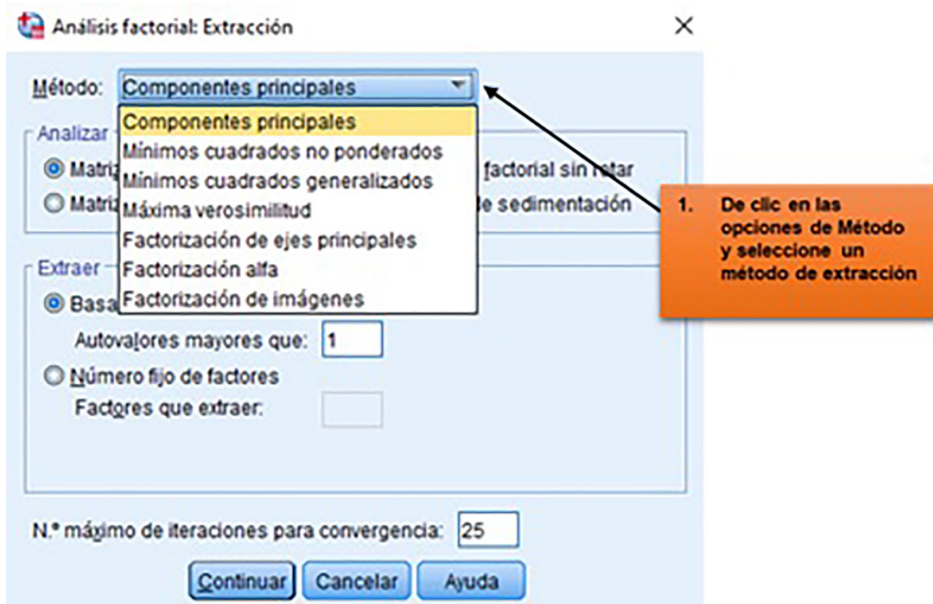
Como se muestra en la imagen, siga la ruta iniciando en la Barra de Menús: *Analizar* → *Reducción de dimensiones* → *Factor*.



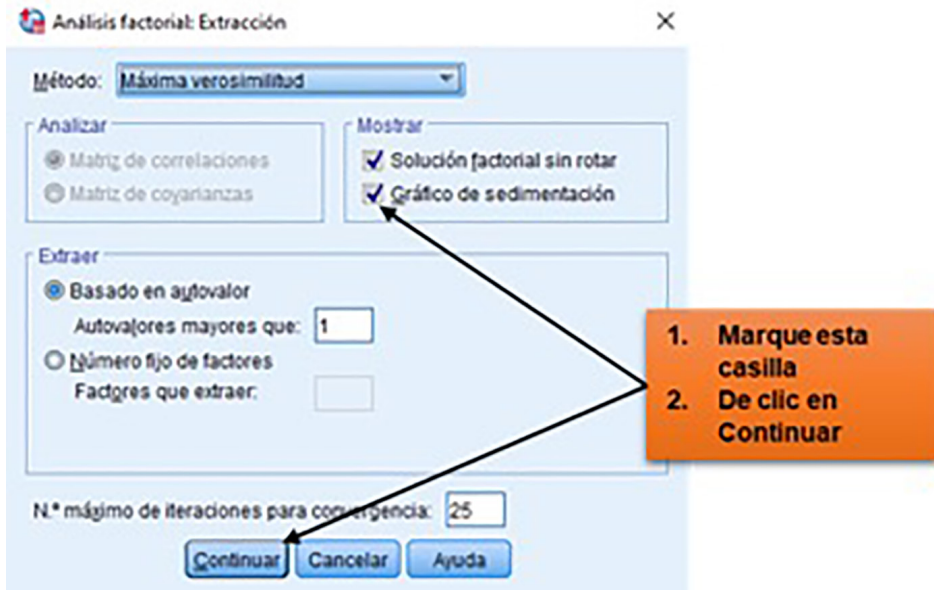
Dentro de la ventana de *Análisis factorial*, seleccione los ítems de la escala a analizar y páselos al área de *Variables* dando clic en la flecha como se muestra en la imagen. Posterior, de clic en *Extracción* en el menú de la derecha.



Dentro de la ventana de *Extracción* seleccione el Método de extracción de los que ofrece SPSS: *Máxima verosimilitud*, *Componentes principales*, *Mínimos cuadrados no ponderados*, *Mínimos cuadrados generalizados*, *Factorización de ejes principales*, *Factorización alfa* y *Factorización de imágenes*.



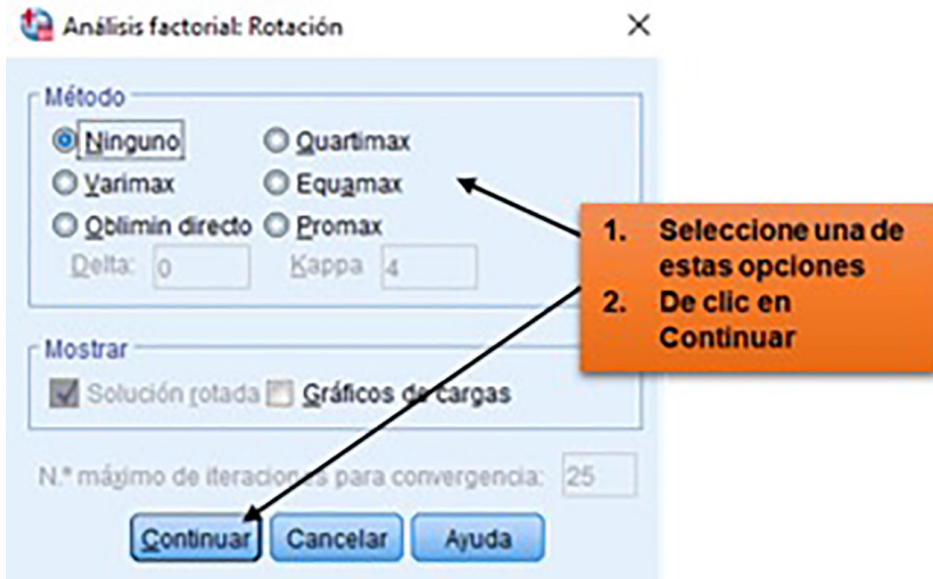
Una vez seleccionado el método de extracción, ubíquese en el área de *Mostrar* y seleccione la opción de *Gráfico de sedimentación*, luego de clic en *Continuar*.



Volviendo a la ventana de *Análisis factorial*, de clic en *Rotación* en el menú de la derecha.



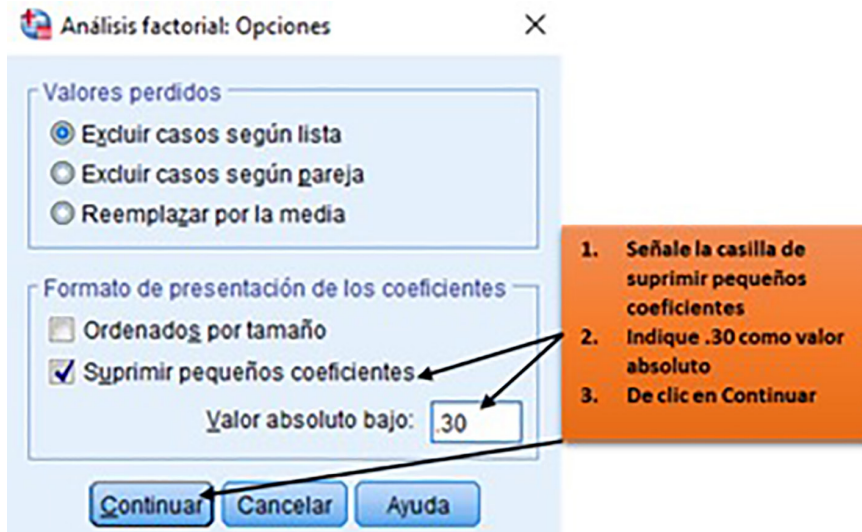
Dentro de la ventana de *Rotación*, seleccione el Método de rotación de los que ofrece SPSS, y que pueden ser: *Varimax*, *Oblimin directo*, *Quartimax*, *Equamax* y *Promax*. Posterior, de clic en *Continuar*.



Volviendo a la ventana de *Análisis factorial*, de clic en *Opciones* en el menú de la derecha.



Dentro de la ventana de *Opciones*, ubíquese en el área de *Formato de presentación de los coeficientes* y seleccione las opciones de *Suprimir coeficientes pequeños*, también ubíquese en el área de *Valor absoluto bajo* e indique *.30*. Posterior, de clic en *Continuar*. En la siguiente pantalla solo de clic en *Aceptar*.



En la Tabla 4, se presenta una salida del programa SPSS correspondiente a un AFE realizado a la escala de Empatía (Vossen, Piotrowski, & Valkenburg, 2015) que consta de dos dimensiones o factores: Empatía cognitiva y Empatía afectiva. El AFE se realiza con el método de Máxima Verosimilitud y Rotación Oblimin (se supone relación entre los factores).

En primer lugar, se sugiere revisar que los valores de estadístico *KMO* y la prueba de esfericidad de Bartlett sean adecuados. Como se muestra en la Tabla 4, el valor de *KMO* es cercano a 1 y el valor de probabilidad de la prueba de esfericidad de Bartlett es *.000*.

Tabla 4. Prueba de *KMO* y Bartlett de la escala de Empatía.

Medida Kaiser-Meyer-Olkin de adecuación de muestreo		.826
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado	2458.664
	gl	28
	Sig.	.000

En segundo lugar, se analiza que los valores de las comunalidades (columna de Extracción) sean superiores a .30, en el caso de la escala de Empatía todos los valores son superiores (ver Tabla 5).

Tabla 5. Comunalidades de los ítems de la escala de Empatía.

	Inicial	Extracción
89. Puedo entender como otro compañero se siente cuando es agredido incluso antes que me diga.	.330	.377
90. Me doy cuenta cuando un compañero está enfadado incluso si trata de esconderlo.	.479	.606
91. Me doy cuenta cuando un compañero actúa como si estuviera feliz cuando realmente no lo está.	.501	.630
92. Puedo darme cuenta fácilmente de los sentimientos de mis compañeros.	.487	.573
93. Cuando un compañero siente miedo me siento asustado.	.482	.583
94. Cuando un compañero está triste me siento triste también.	.490	.601
95. Cuando un compañero está enojado me siento enojado también.	.451	.548
96. Cuando un compañero está nervioso me siento nervioso también.	.411	.491

Método de extracción: máxima verosimilitud.

En tercer lugar, se establece el número de factores a considerar en la escala. En la Tabla 6, la sección de *Autovalores iniciales* (columna Total) existen dos valores superiores a uno (3.52 y 1.76), lo que indica que la escala tiene dos factores.

Tabla 6. Autovalores iniciales de la escala de Empatía.

Factor	Autovalores iniciales			Sumas de cargas al cuadrado de la extracción			Sumas de cargas al cuadrado de la rotación ^a
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado	Total
1	3.524	44.047	44.047	3.084	38.550	38.550	2.577
2	1.769	22.116	66.163	1.324	16.548	55.099	2.481
3	.648	8.104	74.268				
4	.517	6.463	80.731				
5	.429	5.362	86.093				
6	.394	4.921	91.014				
7	.379	4.736	95.750				
8	.340	4.250	100.000				

Método de extracción: máxima verosimilitud.

a. Cuando los factores están correlacionados, las sumas de las cargas al cuadrado no se pueden añadir para obtener una varianza total.

El cuarto paso consiste en comprobar el porcentaje de varianza atribuida a cada factor y el acumulado. En la última columna de la Tabla 7, se reporta el porcentaje de varianza acumulada que es del 55%, el cual sin ser alto es un valor aceptable.

Tabla 7. Varianza total explicada de la escala de Empatía.

Factor	Autovalores iniciales			Sumas de cargas al cuadrado de la extracción			Sumas de cargas al cuadrado de la rotación		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	3.524	44.047	44.047	3.084	38.550	38.550	2.255	28.188	28.188
2	1.769	22.116	66.163	1.324	16.548	55.099	2.153	26.911	55.099
3	.648	8.104	74.268						
4	.517	6.463	80.731						
5	.429	5.362	86.093						
6	.394	4.921	91.014						
7	.379	4.736	95.750						
8	.340	4.250	100.000						

Método de extracción: máxima verosimilitud.

Finalmente, la Tabla 8 se denomina Matriz de patrón, puesto a que se emplea el método de extracción de máxima verosimilitud con rotación Oblimin.

Tabla 8. Matriz de patrón de la escala de Empatía.

	Factor	
	1	2
1. Puedo entender como otro compañero se siente cuando es agredido incluso antes que me diga.	-.026	.623
2. Me doy cuenta cuando un compañero está enfadado incluso si trata de esconderlo.	-.077	.803
3. Me doy cuenta cuando un compañero actúa como si estuviera feliz cuando realmente no lo está.	.011	.789
4. Puedo darme cuenta fácilmente de los sentimientos de mis compañeros.	.175	.675
5. Cuando un compañero siente miedo me siento asustado.	.735	-.069
6. Cuando un compañero está triste me siento triste también.	.760	-.040
7. Cuando un compañero está enojado me siento enojado también.	.752	.035
8. Cuando un compañero está nervioso me siento nervioso también.	.715	.042
Método de extracción: máxima verosimilitud.		
Método de rotación: Oblimin con normalización Kaiser. ^a		
a. La rotación ha convergido en 6 iteraciones.		

Consideraciones para el reporte de resultados

A continuación, se presentan una serie de consideraciones para guiar el reporte del AFE:

1. Se recomienda especificar que se realizó este análisis.
2. Describir el método de extracción (ej. máxima verosimilitud o com-

- ponentes principales, etc.) y de rotación (ej. Varimax, Oblimin, etc.).
3. Colocar los valores de las pruebas del test de Bartlett y del Índice *KMO* y su interpretación.
 4. Determinar el número de factores que posee la escala y nombrarlos especificando el número de ítems por cada factor.
 5. Mencionar el total de varianza explicada.
 6. Colocar la tabla con la matriz factorial y los pesos factoriales (de acuerdo con APA).

Ejemplo del reporte de un AFE

Se realizó un análisis factorial exploratorio con el método de extracción de Máxima Verosimilitud y rotación Oblimin. Los datos demostraron un buen ajuste para este tipo de modelo, lo cual se evidenció en los resultados de la prueba de esfericidad de Bartlett ($X^2 = 2458.66, p < .000$) y el valor de Kaiser-Meyer-Olkin (*KMO*) de .82 (Cea, 2004; Martínez *et al.*, 2006). Como criterio para la inclusión de los ítems se consideraron pesos factoriales de 0.30 o mayores en sólo uno de los factores, lo cual refleja la solidez teórica del reactivo (Hair, Anderson, Tatham, & Black, 1999). Los ocho reactivos de la escala se agruparon en dos factores que explicaron de manera conjunta el 55.1% de la varianza de los puntajes de la escala. El primer factor, integrado por cuatro ítems, explicó un 38.5% de la varianza, y el segundo, compuesto por el mismo número de reactivos, el 16.5% de la misma (ver Tabla 9).

Tabla 9. Resultados del análisis factorial exploratorio de la escala Empatía.

Ítems	Carga factorial		Comunalidad
	F1	F2	
1. Puedo entender como otro compañero se siente cuando es agredido incluso antes que me diga	.62	-.03	.37
2. Me doy cuenta cuando un compañero está enfadado incluso si trata de esconderlo	.80	-.07	.60
3. Me doy cuenta cuando un compañero actúa como si estuviera feliz cuando realmente no lo está	.78	.01	.63
4. Puedo darme cuenta fácilmente de los sentimientos de mis compañeros	.67	.17	.57
5. Cuando un compañero siente miedo me siento asustado	-.07	.73	.58
6. Cuando un compañero está triste me siento triste también	-.04	.76	.60
7. Cuando un compañero está enojado me siento enojado también	.03	.75	.54
8. Cuando un compañero está nervioso me siento nervioso también	.04	.71	.49
Correlaciones factoriales			
Factor 1		-	
Factor 2		.37	-

Conclusiones

En este capítulo se describieron aspectos teóricos y prácticos para la realización de un AFE con apoyo del SPSS. Se enfatiza que el AFE es una técnica estadística multivariable de interdependencia que permite indagar acerca de la dimensionalidad de las escalas de medición. Aquí, es necesario que el investigador decida el método de rotación y extracción que utilizará en sus análisis, esta decisión se relaciona con los supuestos acerca de la relación entre los factores. También, debe optar por dejar libre el número de factores o precisar la cantidad que quiere obtener, lo cual se vincula con el conocimiento de la teoría. Finalmente, se presentan los pasos a seguir para el análisis en el SPSS y la interpretación de los resultados.

Bibliografía sugerida

- Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research & Evaluation*, 18(6), 1-13. Recuperado de http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=C41CA_67730_57760ED95CBFA6EB-74D379?doi=10.1.1.500.1714&rep=rep1&type=pdf
- Cea, M. A. (2004). *Análisis multivariable. Teoría y práctica en la investigación social*. Madrid: Síntesis.
- Ho, R. (2006). *Handbook of univariate and multivariate data analysis and interpretation with SPSS*. Boca Raton, Florida: CRC.
- Reio, T. G., & Schuck, B. (2014). Exploratory factor analysis. *Advances in Developing Human Resources*, 17(1), 12-25. doi: 10.1177/1523422314559804

Capítulo 8

Análisis Factorial Confirmatorio (AFC) con apoyo de AMOS

El AFC aborda la dimensionalidad de los modelos de medida que incluyen la relación entre una serie de indicadores observables (ítems) y variables latentes (constructos). Esto permite tomar decisiones acerca de la composición de los puntajes, es decir, permite determinar la pertinencia del uso de puntajes por escala y de manera global.

A diferencia del AFE en el AFC se debe especificar todos los aspectos del modelo de medición. Aunque no serán tema de este texto, es conveniente señalar que el AFC tiene otras posibilidades que no están disponibles en el AFE, tales como el análisis de la invarianza en el modelo a través del tiempo e informantes, la comparación entre modelos de medida y la evaluación de los efectos del método. El AFC comprende las siguientes fases:

Verificación de los supuestos

Es importante que antes de realizar el cálculo del modelo factorial se verifiquen los supuestos para evitar sesgos en su estimación:

- Linealidad. El modelo presupone relaciones lineales entre las variables.

- Se puede explorar con los gráficos de dispersión.
- Tamaño muestral elevado. Se sugieren muestras mayores a 200 casos, una regla práctica es 20 casos por ítem.
 - Normalidad multivariante. Esta se puede calcular mediante el estadístico de Mardía. El AMOS tiene la opción del bootstrap para procurar que las estimaciones no se afecten por las violaciones de este supuesto.
 - Relación entre el constructo y los ítems.

Especificación del modelo

El investigador debe especificar, con base en la teoría y la evidencia empírica, uno o varios modelos de medición teóricos donde se defina la relación entre los constructos (en caso de ser multidimensional) y sus indicadores (ver Figura 10).



Figura 10. Modelo teórico de medición de la Cibervictimización en estudiantes universitarios.

Nota. Tomado de “Propiedades psicométricas de una escala para medir cibervictimización en universitarios” por A. A. Valdés, E. A. Carlos y G. M. Torres, 2018, *Revista Electrónica de Investigación Educativa*, 20, p. 39.

Identificación estadística del modelo

El AFC puede ser estimado solamente cuando el número de parámetros libremente estimados no exceden el número de piezas de información en la matriz de varianzas y covarianzas. Con base en lo anterior, los modelos se consideran no identificados, identificados o sobre identificados.

Un modelo se encuentra no identificado cuando el número de parámetros desconocidos (libremente estimados) exceden el número de parámetros conocidos, lo que deriva en grados de libertad negativos. Este modelo no puede ser resuelto dado que tiene un infinito número de soluciones (ver Figura 11a). Un modelo identificado es aquel donde el número de parámetros desconocidos es similar al número de parámetros conocidos, éste posee 0 grados de libertad. Su cálculo ofrece una solución perfecta, por lo que carece de interés al no poderse generalizar (ver Figura 11b). Por último, un modelo sobreidentificado es aquel donde el número de parámetros conocidos excede el número de parámetros desconocidos, lo que implica que sus grados de libertad son positivos (ver Figura 11c).

Para lograr una correcta identificación del modelo, el investigador debe: (a) asignar una escala a la variable latente, lo cual se puede hacer asignando un ítem como indicador o fijando la varianza en el factor (habitualmente a 1). El AMOS de forma automática fija como indicador (asignándole el valor 1) al primer ítem de cada factor. El investigador puede decidir mantener o utilizar otro ítem como marcador. Una regla práctica es utilizar el ítem con mayor carga como marcador.

Es necesario señalar que también existen dificultades en la identificación empírica de los modelos de medición relacionada con la no correlación de uno o varios ítems con el factor al cual fue asignado.

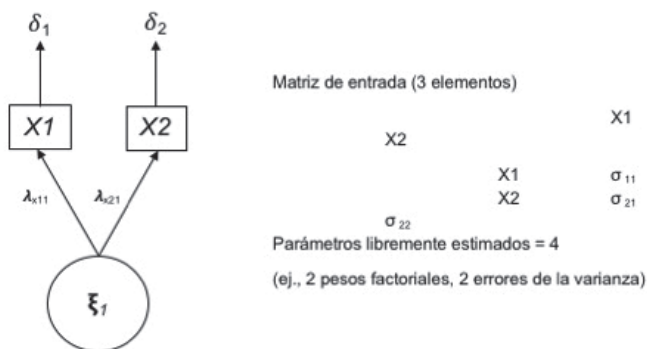


Figura 11a. Modelo no identificado ($gl = -1$).

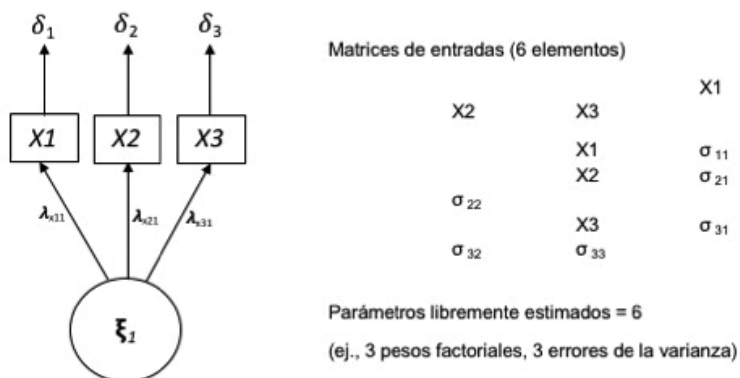


Figura 11b. Modelo identificado ($gl = 0$).

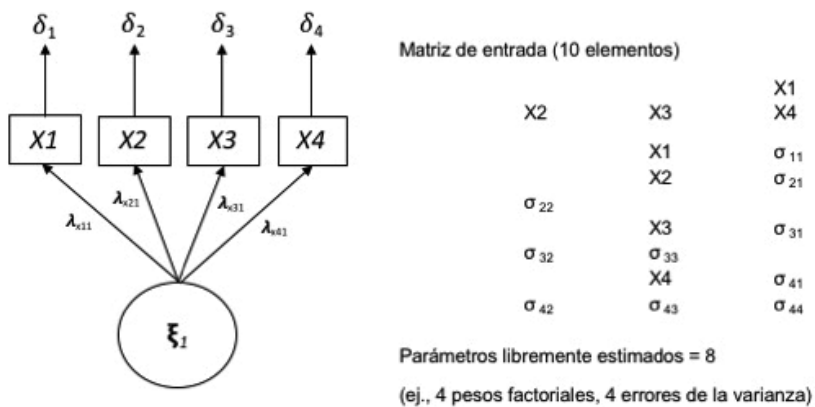


Figura 11c. Modelo sobreidentificado ($gl = 2$).

Selección del método de estimación

El modelo es estimado para minimizar las diferencias entre la matriz de covarianza empíricas y la propuesta en el modelo teórico de medición. El AMOS ofrece como métodos de estimación:

- Máxima verosimilitud (ML). Este método requiere de normalidad multivariada de las variables observables y de muestras grandes.

b) Mínimos cuadrados no ponderados (WLS). Este método no asume supuestos acerca de la distribución de las variables. Sin embargo, no arroja los valores de X^2 y de otros índices de ajuste.

c) Mínimos cuadrados generalizados (ULS). Se basa en el mismo principio del método de mínimos cuadrados de la regresión. Reporta los mismos estadísticos que el método de ML.

d) Método de mínimos cuadrados libre de escala (SLS). Se basa en el análisis de la matriz de correlación. Realiza una transformación lineal de las variables manifiestas para su análisis.

e) Estimación asintótica libre de distribución (ADF). No asume normalidad multivariada, sin embargo, requiere muestras muy grandes de más de 1000 participantes.

El método de Máxima Verosimilitud (ML) es el más utilizado para realizar las estimaciones en el AMOS. Sin embargo, dado que requiere de normalidad multivariada de las variables observables, un supuesto que no se cumple frecuentemente, se sugiere utilizarlo asociado con el método de bootstrap para evitar que las estimaciones se encuentren sesgadas.

Especificación del bootstrap

Es un método de re-muestreo que combina los casos en diferentes grupos de muestras para obtener precisión estadística. El bootstrap simula un muestreo aleatorio con reemplazo. Se recomienda utilizar el método con 500 repeticiones y el método de percentiles con un intervalo de confianza del 95%. Es necesario revisar los intervalos de confianza de los parámetros estimados, éstos no deben pasar por cero no ser demasiado grandes.

Definición de parámetros a calcular

El AMOS ofrece varios estadísticos, nosotros en el caso de los modelos de medición sugerimos solicitar:

a) Estimados estandarizados. Ofrece los pesos factoriales estandarizados de la relación entre la variable latente y los ítems. Aunque son aceptables valores entre .40 y .60, lo ideal es que sean mayores que .60.

b) Correlación múltiple al cuadrado. Permite conocer la varianza explicada en cada ítem por la variable latente.

c) Índices de modificación. Valores mayores que cinco sugieren cambios que mejoran el ajuste del modelo. Es importante señalar que el investigador es quien decide con base en la teoría las modificaciones que incluye en el modelo.

d) Pesos factoriales. Permiten conocer la varianza explicada por las variables latentes.

e) Covarianza y correlaciones. Como su nombre lo indica brinda información acerca la relación entre los parámetros estimados. Es importante considerar las correlaciones entre los factores para analizar la dimensionalidad de la escala.

f) Pruebas de normalidad y casos atípicos. La salida del AMOS ofrece valores de asimetría, curtosis, el estadístico de Mardia para evaluar normalidad multivariada (valores menores a cinco son aceptables), la distancia la Mahalanobis y estadísticos de probabilidad asociados para identificar los casos atípicos.

Análisis de los índices de ajuste globales

Los índices de ajuste globales informan acerca del ajuste promedio del modelo de medición (Kline, 2015). Estos índices no informan acerca de la dimensionalidad del modelo, no excluyen que existan partes del modelo que no ajusten o que tengan un sentido contrario al esperado. Se han desarrollado diversidad de índices de ajuste globales, nosotros nos referiremos a aquellos que varios autores sugieren reportar (Blunch, 2013; Brown, 2015; Byrne, 2010; Kline, 2015).

Índices de ajuste absolutos. Evalúan el ajuste del modelo sin considerar otros aspectos (parsimonia, restricciones), se basan en la comparación de las matrices de varianza y covarianzas del modelo propuesto y el generado de los datos. Dentro de este grupo se reportan con mayor frecuencia: la Chi cuadrado con probabilidad asociada (X^2, p), el Índice de Bondad de Ajuste (GFI) y la Raíz Cuadrada de la Media de Residuos Cuadrados (RMSR).

a) *Chi cuadrado con probabilidad asociada (X^2, p).* El estadístico X^2 es el tradicional índice de ajuste dentro del modelo de AFC, en éste se establece una hipótesis nula donde se espera que no existen diferencias entre el modelo teórico de medición y el generado de los datos, y una hipótesis alterna que sostiene que existen diferencias entre el modelo teórico y el empírico. La obtención de valores de probabilidad mayores a .05 implica que el modelo teórico se ajusta a los datos ($p > .05$). Este estadístico por lo general no se utiliza solo como índice de ajuste, ya que es afectado por la normalidad de los datos y es inflado por el tamaño de muestra.

b) *Índice de Bondad de Ajuste (GFI).* Mide la cantidad relativa de varianza y covarianzas que se explica por el modelo propuesto. Su valor oscila entre 0 (mal ajuste) a 1 (ajuste perfecto), un valor $> .90$ es considerado necesario para un ajuste aceptable del modelo.

c) *Raíz Cuadrada de la Media de Residuos Cuadrados (RMSR).* Se basa en el análisis de los residuos, mientras menos residuos mejor será el ajuste del modelo. En general, valores $< .05$ se consideran indicadores de un buen ajuste del modelo.

Índices de ajuste incremental. Estos índices comparan el modelo propuesto con el modelo nulo que se caracteriza porque todos sus parámetros estructurales han sido fijados a cero, lo cual implica ninguna relación entre las variables.

a) *Índice de Tucker-Lewis (TLI).* Este índice cuantifica el grado en que un modelo particular es una mejora sobre un modelo nulo, especialmente cuando se estima por el método de máxima verosimilitud. El rango de valores va de 0 a 1, valores $> .90$ se consideran aceptables.

b) *Índice de Ajuste Comparativo (CFI).* Este también compara el modelo especificado con el nulo. El rango de valores va de 0 a 1, valores $> .95$ se consideran aceptables.

Índices de ajuste de parsimonia. Estos índices consideran la parsimonia del modelo, es decir, obtener modelos con pocos parámetros y muchos grados de libertad. Estos permiten comparar modelos con diferentes números de parámetros estimados.

a) *Índice de Bondad de Ajuste Ajustado (AGFI).* Es una extensión del GFI que considera los grados de libertad, de esta manera favorece a los modelos con más grados de libertad. El rango de los valores oscila entre 0 (bajo ajuste) a 1 (ajuste perfecto), a diferencia de otros índices puede tomar valores negativos que no son deseables, se consideran valores $> .90$ como indicadores de un buen ajuste del modelo.

b) *Error de la Raíz Cuadrada de la Media de Aproximación (RMSEA).* Compara el ajuste del modelo propuesto con el nulo, cuando más pequeño es su valor mejor es el ajuste del modelo. Valores $< .05$ indican un buen ajuste, mientras que valores $> .05$ y $< .08$ se consideran aceptables.

Índices de ajuste teórico. Se desarrollaron para comparar modelos con diferentes números de parámetros, lo cual facilita comparar diferentes modelos teóricos de medición del constructo. En general, los modelos con pocos parámetros estimados obtienen menores valores (indican mejor ajuste) por ser más fácilmente replicables.

a) *Criterio de Información de Akaike (AIC).* Cuando más pequeño sea el valor del AIC, mejor es el ajuste del modelo por ser más parsimonioso. Estos valores se alcanzan con modelos que tengan pocos parámetros estimados y bajos valores de X^2 .

b) *Criterio de Información de Bayes (BIC).* Valores menores sugieren que el modelo presenta un mejor ajuste, a diferencia del AIC este considera el tamaño de la muestra.

Análisis de los índices de ajuste específicos

Los índices de ajuste específicos evalúan el ajuste de relaciones específicas propuestas, es conveniente considerar que un modelo que presenta un ajuste global aceptable puede tener aspectos específicos del mismo donde el ajuste no se obtenga.

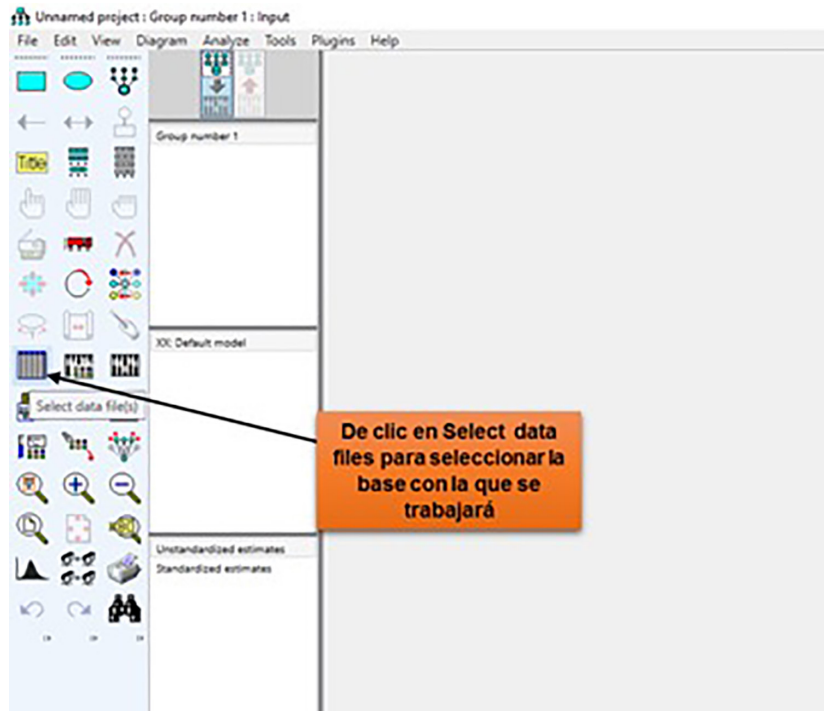
Coefficientes de regresión estandarizados. Indican la cantidad de varianza del ítem que es explicada por el constructo, lo recomendable es que tengan un valor $> .40$ y de preferencia $> .60$. Valores menores indican que la varianza en el ítem es poco explicada por el constructo, lo que sugiere que no es un adecuado indicador del mismo.

Covarianza y correlación. Indican la relación entre las dimensiones en los modelos multidimensionales. Un valor de correlación $< .40$ y $> .70$ sugiere relación sin colinealidad entre los constructos (no existe una excesiva correlación entre las dimensiones). Por su parte, un valor $< .40$ indica escasa relación entre las dimensiones lo que implica que no existe un constructo de orden mayor que los agrupe. Finalmente, un valor $> .70$ es una señal de colinealidad que hace dudar sobre la multidimensionalidad de la escala.

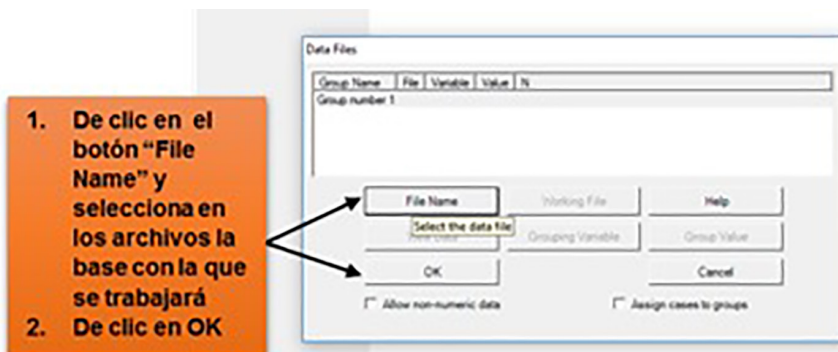
El impacto de los valores de correlación en el ajuste depende de la propuesta hecha por los autores. Por ejemplo, cuando se propone un modelo conformado por tres dimensiones relacionados, lo ideal sería obtener valores de correlación $< .40$ y $> .70$. Por otra parte, si el investigador espera que los factores no se encuentren correlacionados entonces los valores esperados serían $< .40$.

Procedimiento para el cálculo del AFC en AMOS

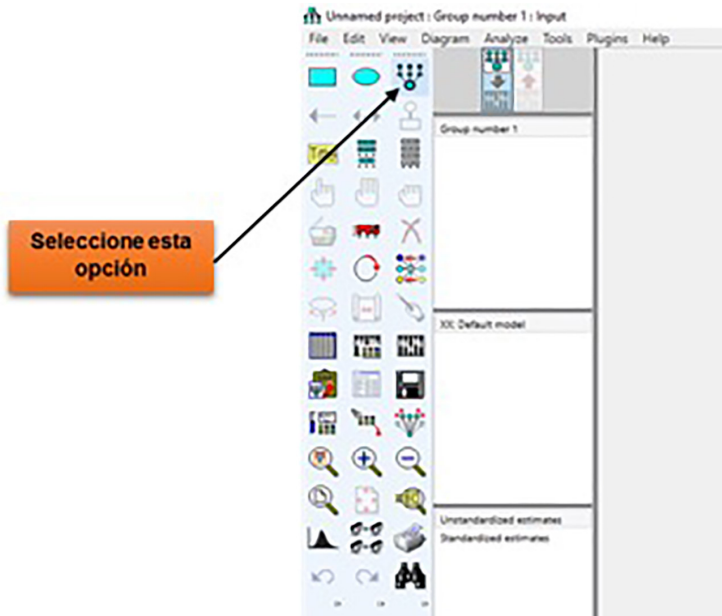
Como se muestra en la imagen, de clic en la opción *Select data files* para seleccionar el archivo de la base de datos de SPSS en donde tiene capturado los datos a analizar.



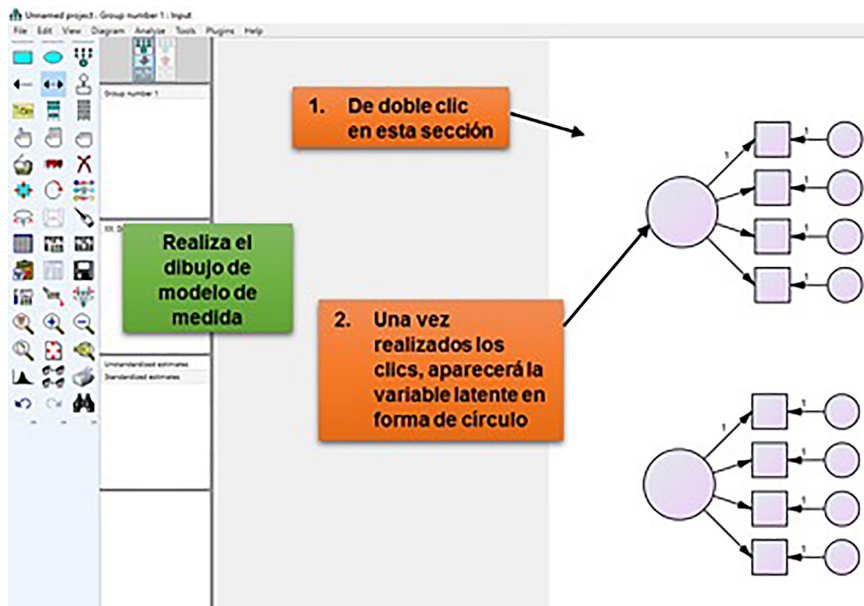
Posteriormente, en la pantalla de *Data files* de clic en la opción *File Name* y seleccione los archivos de SPSS con que se trabajará. Luego de clic en el botón *OK*.



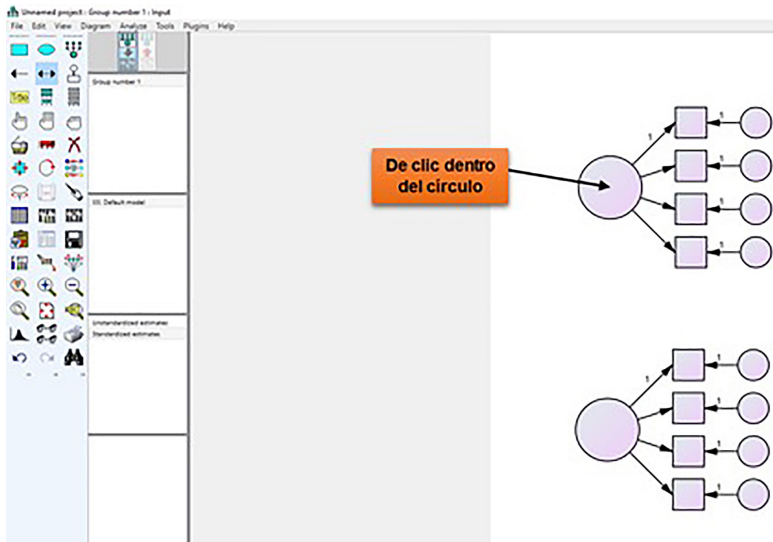
Una vez hecho lo anterior, de clic en la opción *Draw a latent variable* para dibujar una variable latente y sus indicadores.



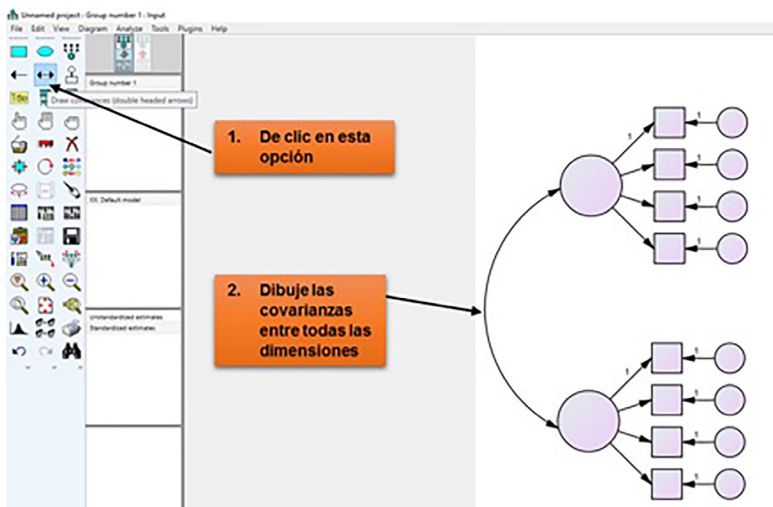
Para dibujar el modelo de medida a probar, de clic en el recuadro en blanco en la parte derecha de la pantalla una vez para que aparezca la variable latente (círculo) como se muestra en la imagen.



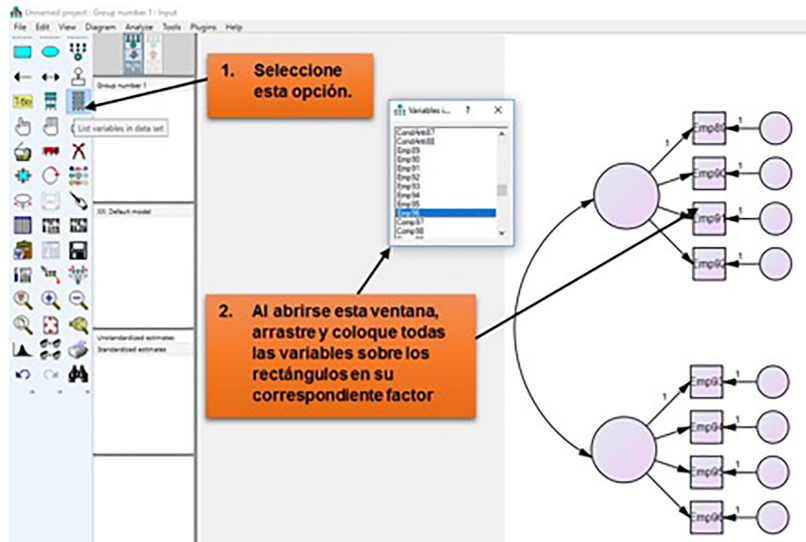
Para continuar, de clic dentro del círculo para que aparezcan sus variables manifiestas (rectángulos). El número de clics que dé corresponde al número de ítems que desea considerar en el constructo a evaluar o cada dimensión que lo integra (ej. Si la dimensión tiene cuatro ítems debe dar cuatro clics).



Repita el paso anterior si el constructo tiene varias dimensiones o factores. Es decir, si la escala consta de dos dimensiones, debe repetir el paso 5 y 6 dos veces. Para continuar, de clic en la opción *Draw covariances* del menú de la izquierda para dibujar las covarianzas entre las dimensiones dibujadas en el modelo. Inmediato, coloque el mouse sobre una variable latente y dibuje la relación con el resto de las variables de modelo. Repita este proceso hasta que se hayan establecido relaciones entre todas las variables latentes.



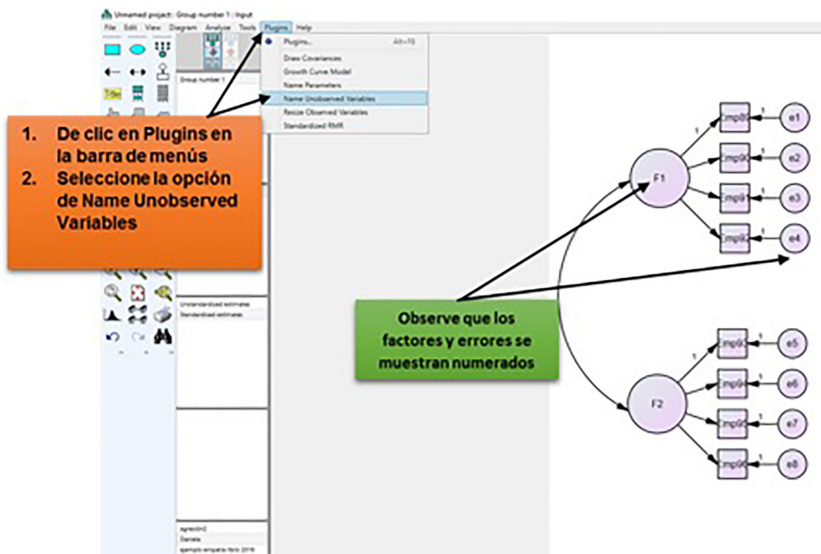
El siguiente paso requiere que de clic en la opción de *List variables in data set*, en el menú de la izquierda. Al hacerlo se abrirá la ventana de *Variables in dataset*. Arrastre cada uno de los ítems de su base de datos y colóquelos en las variables manifiestas (rectángulos) según corresponda su dimensión.



1. Seleccione esta opción.

2. Al abrirse esta ventana, arrastre y coloque todas las variables sobre los rectángulos en su correspondiente factor

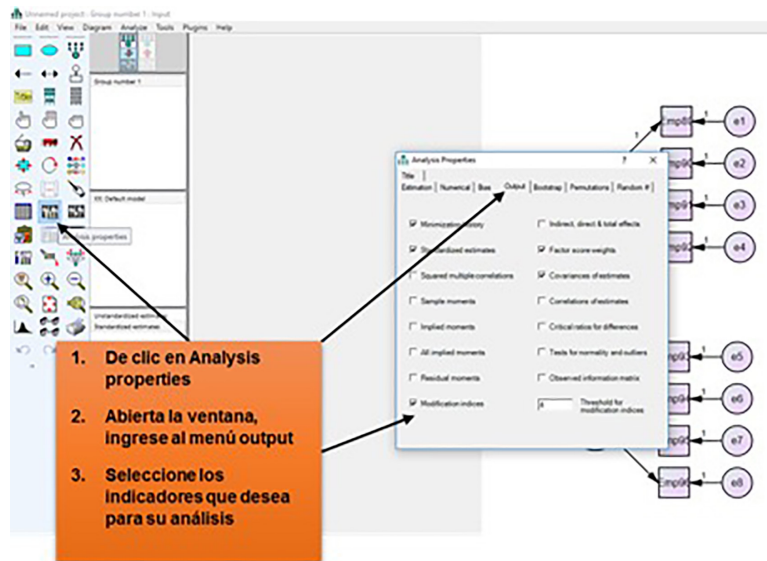
Una vez integradas las variables manifiestas (rectángulos) en el dibujo del modelo, siga la ruta iniciando en la Barra de Menús: *Plugins* → *Name Unobserved Variables*. Al hacerlo se observa que los factores y los errores aparecen numerados, como se ejemplifica en la imagen.



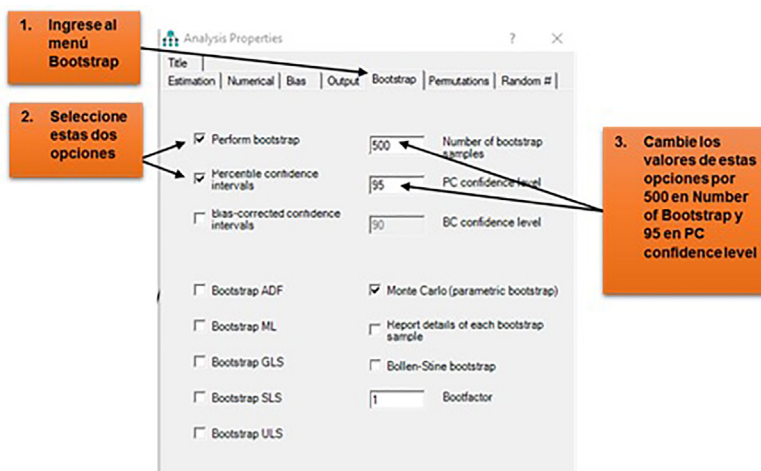
1. De clic en Plugins en la barra de menús
2. Seleccione la opción de Name Unobserved Variables

Observe que los factores y errores se muestran numerados

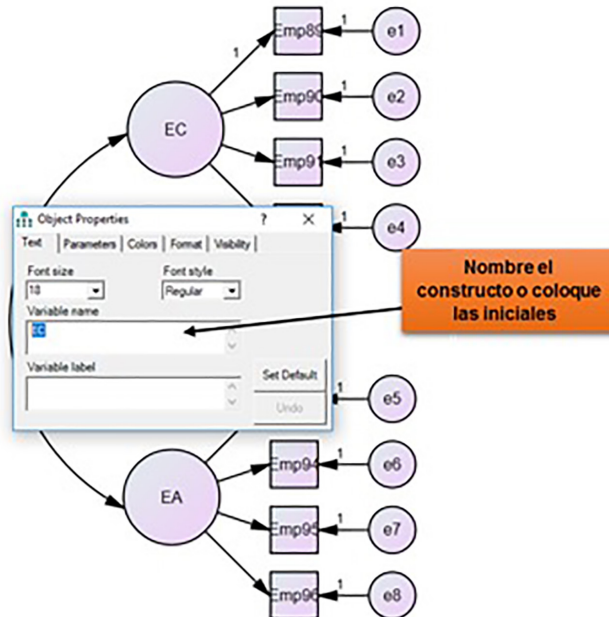
Una vez completo el dibujo del modelo, de clic en la opción de *Analysis properties*, en el menú de la izquierda para que se abra la ventana correspondiente. En dicha ventana ingrese al menú *Output* en donde deberá elegir los indicadores que requiere para su análisis. En la imagen se muestra como ejemplo la selección de *Minimization history*, *Standardized estimates*, *modification indices*, *factor score weights*, *covariances of estimates*. Es importante señalar que las casillas que se marquen dependen del análisis que se realiza y de la naturaleza del constructo.



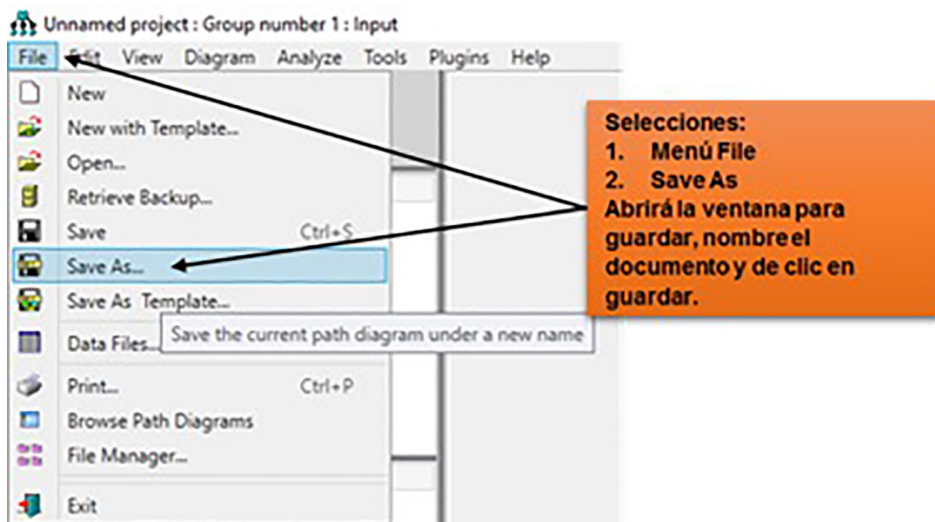
En la misma ventana de *Analysis properties*, ingrese al menú *Bootstrap*. En las opciones de la izquierda, señale la casilla *Perform bootstrap* e indique en el cuadro de la derecha *Number of bootstrap samples* el número 500. De igual modo, señale la casilla *Percentile confidence intervals* e indique en *Pc confidence level* el número 95. Para finalizar cierre la ventana dando clic en la X.



Para continuar, de doble clic el círculo que representa el factor que contiene las iniciales F1, se abrirá una ventana llamada *Object Properties*, ubíquese en el área de *Variable name* y escribe un nombre corto o las iniciales del constructo. Cierre la ventana para continuar, dando clic en la X.



Para iniciar el cálculo del modelo es necesario que guarde el trabajo realizado hasta el momento, para ello siga la ruta iniciando en la Barra de Menús. *File* → *Save As*. Hecho esto se abrirá la ventana para guardar el documento, nómbrelo y de clic en el botón de *Guardar*.



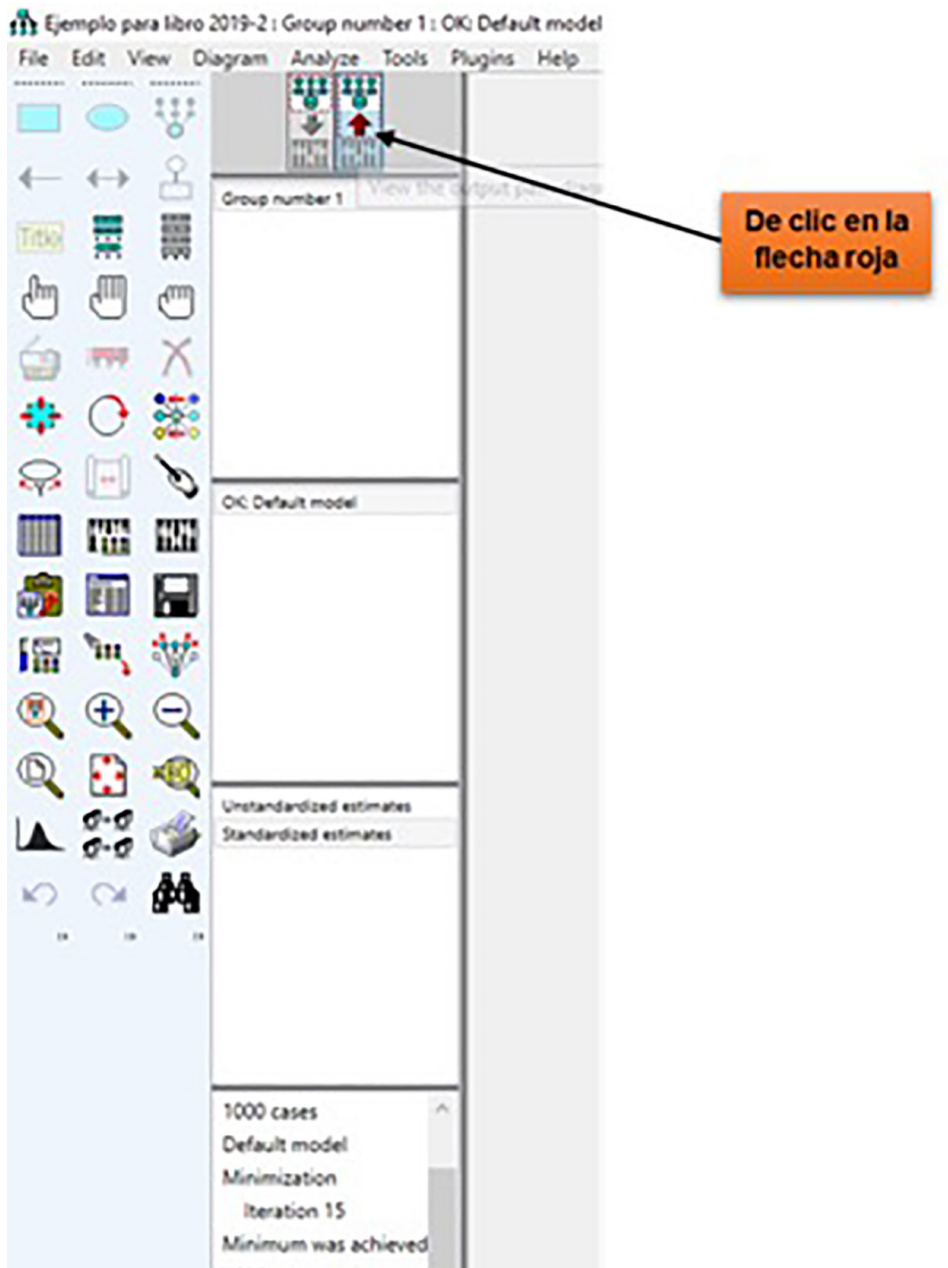
Para iniciar con el cálculo, de clic en la opción de *Calculate estimates* en el menú de la izquierda para que se muestren los resultados y luego de clic en *Standardized estimates*, tal como se muestra en la imagen.

The screenshot displays the AMOS software interface. The title bar reads "Ejemplo para libro 2019-2 : Group number 1 : Input". The menu bar includes "File", "Edit", "View", "Diagram", "Analyze", "Tools", "Plugins", and "Help". The left sidebar contains various tool icons. The main workspace shows a diagram area with "Group number 1" and a "Calculate estimates" button. Below this, there are options for "Unstandardized estimates" and "Standardized estimates". At the bottom, it indicates "1000 cases" and "Default model".

Two arrows point from an orange callout box to the "Calculate estimates" button and the "Standardized estimates" option. The callout box contains the following instructions:

1. De clic en esta opción para ver los resultados
2. De clic en Standardized estimates

Continúe dando clic en la flecha roja que apunta hacia arriba ubicada en la parte superior llamada *View the output path diagram*, como se muestra en la imagen.



Para visualizar resultados, elija la opción *View text* en el menú de la izquierda para desplegar la ventana llamada *Amos Output*. En dicha ventana verifique que los valores de *Chi-cuadrada*, los *grados de libertad* y *p* sean adecuados según la literatura especializada.

1. De clic en la opción View text para desplegar la ventana Amos output

2. Verificar la Chi-cuadrada, grados de libertad y el P valor

En la misma ventana de Amos Output, de clic en Estimates y verifique que los pesos factoriales sean significativos, esto en el área de *Regression Weights*. De igual modo, revise los pesos factoriales estandarizados en el área de *Standardized Regression Weights*, y las Correlaciones en el área de *Covariances*.

1. De clic en Estimates

2. Verifique los pesos factoriales en Regression Weights

3. Revise los Standardized Regression Weights

4. Revisar las Covariances

Estimates (Group number 1 - Default model)

Scalar Estimates (Group number 1 - Default model)

Maximum Likelihood Estimates

Regression Weights (Group number 1 - Default model)

	Estimate	S.E.	C.R.	P	Label
Emp09 <-- EC	1.000				
Emp00 <-- EC	1.163	.061	19.032	***	par_1
Emp01 <-- EC	1.291	.067	19.346	***	par_2
Emp02 <-- EC	1.159	.062	18.817	***	par_3
Emp03 <-- EA	1.000				
Emp04 <-- EA	1.049	.045	23.239	***	par_4
Emp05 <-- EA	.857	.038	22.682	***	par_5
Emp06 <-- EA	.888	.042	21.058	***	par_6

Standardized Regression Weights (Group number 1 - Default model)

	Estimate
Emp09 <-- EC	.625
Emp00 <-- EC	.786
Emp01 <-- EC	.810
Emp02 <-- EC	.771
Emp03 <-- EA	.772
Emp04 <-- EA	.790
Emp05 <-- EA	.758
Emp06 <-- EA	.702

Covariances (Group number 1 - Default model)

	Estimate	S.E.	C.R.	P	Label
EC <--> EA	.296	.031	9.529	***	par_7

Continuando en la ventana de *Amos Output*, de clic en *Model fit* para verificar los índices de ajuste del modelo, los cuales se revisan en el área que se despliega a la derecha llamada *Model Fit Summary*.

1. De clic en Model fit para verificar los índices de ajuste del modelo

2. En la sección de Model Fit Summary

Model Fit Summary

Model	NPAB	CMEN	DF	P	CMEN/DF
Default model	17	151.263	19	.000	7.961
Saturated model	36	.000	0		
Independence model	8	3350.667	28	.000	119.667

Model	ENR	GFI	AGFI	PGFI
Default model	.879	.962	.923	.908
Saturated model	.000	1.000		
Independence model	.492	.453	.297	.353

Model	NR1	RF1	RF1	TL1	CFI
	Delta1	obs1	Delta2	obs2	
Default model	.955	.953	.960	.941	.960
Saturated model	1.000	1.000	1.000	1.000	1.000
Independence model	.000	.000	.000	.000	.000

Model	PRATIO	PNFI	PCFI
Default model	4.79	4.63	4.52
Saturated model	.000	.000	.000
Independence model	1.000	.000	.000

Model	NCP	LO 90	HI 90
Default model	132.263	96.767	179.239
Saturated model	.000	.000	.000
Independence model	3322.667	3136.238	3514.607

Model	FMEN	FI	LO 90	HI 90
Default model	1.51	1.32	.897	1.79
Saturated model	.000	.000	.000	.000
Independence model	3.354	3.328	3.339	3.320

En caso de que los índices de ajuste, revisados en el paso anterior, no sean los adecuados se puede realizar lo siguiente para que el modelo ajuste: de clic en *Modification indices* o índices de modificación que se encuentra como parte del menú de la izquierda en la ventana de *Amos Output*. Detecte en el área de *Covariances* los errores que se pueden relacionar** o los ítems que se pueden eliminar***.

1. De clic en Modification Indices

2. En el área de Covariances detecte los ítems que se pueden relacionar o eliminar

Modification Indices (Group number 1 - Default model)

	M.I.	Par Change
e8 <-> EC	4.211	-.044
e7 <-> e8	28.620	-.109
e6 <-> e8	5.304	-.054
e5 <-> e8	4.111	-.047
e5 <-> e7	4.575	-.041
e5 <-> e6	8.878	.066
e4 <-> EA	38.512	.148
e4 <-> EC	6.686	-.051
e4 <-> e6	17.579	.096
e4 <-> e5	5.241	.051
e3 <-> e4	5.020	.047
e2 <-> EA	19.950	-.104
e2 <-> e5	4.769	-.047
e1 <-> e4	16.383	-.105
e1 <-> e2	38.951	.157

Covariances (Group number 1 - Default model)

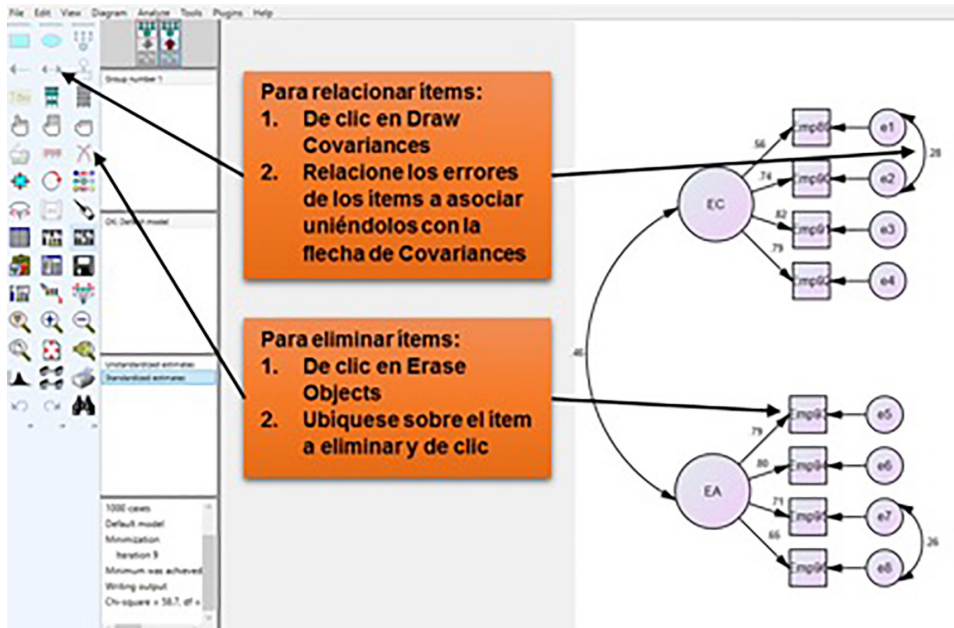
Variance (Group number 1 - Default model)

***Para relacionar ítems

Se debe volver a la ventana del modelo, como se muestra en la siguiente imagen, y elegir la opción de *Draw Covariances* en el menú de la izquierda, luego ubique el mouse sobre el error que corresponde al ítem a asociar y únalo con el otro error del ítem a relacionar. **No olvide que para relacionar dos errores estos deben pertenecer al mismo factor.**

***Para eliminar un ítem

En la ventana del modelo de clic en el botón *Erase Objects* del menú de la izquierda para retomar la cruz de color rojo, la cual deberá ubicar sobre el ítem que desea eliminar y dar clic sobre él. **No olvide que también debe eliminar el error asociado al ítem eliminado.**

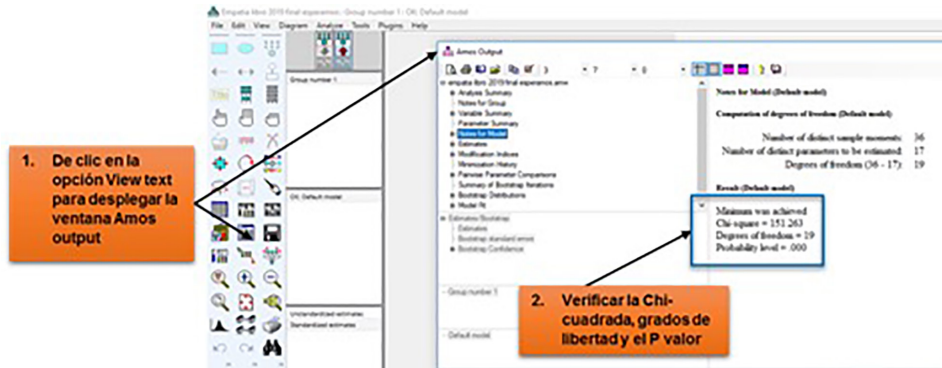


Una vez realizados los ajustes necesarios, de clic en *Calculate estimates* y luego en *Standardized estimates*.



1. De clic en esta opción para ver los resultados
2. De clic en Standardized estimates

Posteriormente, de clic en *View text* para abrir de nuevo la ventana *Amos Output* y revisar que los índices de ajuste sean adecuados. Para ello, de clic en la opción de *Model fit* en el menú de la izquierda y ubíquese en el área de *Notes for Model* del lado izquierdo.



Ejemplo del reporte de un AFC

Se realizó un análisis factorial confirmatorio utilizando el método de estimación de Máxima Verosimilitud (ML). Los resultados del AFC sugieren un buen ajuste global a los datos del modelo de medición bidimensional propuesto ($\chi^2 = 151.11$, $gl = 8$, $p = .057$; RMR = .04; AGFI = .93; TLI = .94; CFI = .96; RMSEA = .04 IC [.01, .05]). Los valores de los coeficientes de regresión sugieren que los factores explican una parte aceptable de la varianza de los ítems (ver Tabla 10). La correlación entre los factores fue de .43, de esto se sucede que son factores relacionados pero que no presentan problemas de colinealidad. Los valores del coeficiente Omega de McDonald ($\Omega = .81$) y de la Varianza Media Extractada (VME) de .62 en el primer y de .59 en el segundo factor sugieren una adecuada fiabilidad de la escala.

Tabla 10. Coeficientes de regresión estandarizados del Modelo Factorial Confirmatorio de la escala para medir Empatía.
 Nota. Los resultados del AFC se pueden reportar en forma de tabla o figura.

Ítems	Coeficientes de regresión	
	F1	F2
1. Puedo entender como otro compañero se siente cuando es agredido incluso antes que me diga	.72	
2. Me doy cuenta cuando un compañero está enfadado incluso si trata de esconderlo	.84	
3. Me doy cuenta cuando un compañero actúa como si estuviera feliz cuando realmente no lo está	.82	
4. Puedo darme cuenta fácilmente de los sentimientos de mis compañeros	.77	
5. Cuando un compañero siente miedo me siento asustado		.78
6. Cuando un compañero está triste me siento triste también		.79
7. Cuando un compañero está enojado me siento enojado también		.77
8. Cuando un compañero está nervioso me siento nervioso también		.74

Conclusiones

En este capítulo se presentaron elementos básicos de la técnica de AFC. Esta técnica tiene como propósito evaluar la estructura interna o dimensionalidad de las escalas de medición. Es un análisis guiado por la teoría, ya que el investigador intenta probar el ajuste a los datos de un modelo teórico de medida propuesto para el constructo. Sin embargo, también puede ser utilizado de forma exploratoria cuando a partir del análisis de los índices de ajuste parciales, los residuos y los índices de modificación el investigador reajuste la propuesta inicial de su modelo.

En primer lugar, es importante que los investigadores analicen en primera ins-

tancia los índices de ajuste global. El estadístico X^2 con su p asociada es el índice tradicionalmente utilizado, sin embargo, dada la vulnerabilidad de esta prueba al tamaño de muestra y los problemas de normalidad multivariada se han propuesto diversos índices de ajuste. Con base en varios autores nosotros proponemos utilizar el RMR, el AGFI, el TLI, el CFA y el RMSEA con sus intervalos de confianza asociados.

En segundo lugar, los investigadores deben evaluar los índices de ajuste parcial (particularmente los coeficientes de regresión estandarizados y los valores de correlación) teniendo en cuenta que un ajuste global aceptable no implica que todos los índices de ajuste parcial sean aceptables.

Es necesario considerar que este capítulo es únicamente una pequeña introducción a las posibilidades del AFC. Se sugiere que los investigadores consulten otras lecturas que fortalezcan su conocimiento teórico sobre el AFC, además que les brinden otras posibilidades como la exploración de la invarianza de medida.

Bibliografía sugerida

- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: The Guildford Press.
- Byrne, B. M. (2010). *Structural equation modeling with AMOS* (2nd ed.). New York, NY: Taylor and Francis.
- Jackson, D. L., Gillaspay, J. A., Jr., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, 14(1), 6-23. doi: 1037/a0014694
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99(6), 323-338. doi: 10.3200/JOER.99.6.323-338

Referencias

- Alcántar, C., Valdés, A. A., Carlos, E. A., Martínez, B., & García, F. I. (2018). Propiedades psicométricas de la adaptación al español de la Participant Role Approach (PRA). *Revista Colombiana de Psicología*, 27(2), 177-192. doi: 10.15446/rcp.v27n2.68721
- American Educational Research Association, American Psychological Association, and National Council of Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: Author.
- Anastasi, A., & Urbina, S. (1997). *Psychological test* (7th ed.). Upper Saddle River, NJ: Prentice Hall/Pearson Education.
- Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology*, 71, 364-374. doi: 10.1037/0022-3514.71.2.364
- Blunch, N. S. (2013). *Introduction structural equation modeling using IBM SPSS Statistic and AMOS* (2nd ed.). Thousand, Oaks, California: SAGE.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: The Guildford Press.
- Byrne, B. M. (2010). *Structural equation modeling with AMOS* (2nd ed.). New York, NY: Taylor and Francis.
- Cea, M. A. (2004). *Análisis multivariable. Teoría y práctica en la investigación social*. Madrid: Síntesis.
- Chang, L. (1994). A psychometric evaluation of 4-point Likert-type scales in relation to reliability and validity. *Applied Psychological Measurement*, 18(3), 205-215. doi: 10.1177/014662169401800302
- Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.
- Croizet, J.-C., & Dutrévis, M. (2004). Socioeconomic status and intelligence: Why test scores do not equal merit. *Journal of Poverty*, 8(3), 91-107. doi: 10.1300/j134v08n03_05
- Cronbach, L. J. (1960). *Essentials of psychological testing* (2nd ed.). New York, NY: Harper & Row.
- Darlington, R. B., & Hayes, A. F. (2017). *Regression analysis and linear models. Concepts, applications, and implementation*. New York, NY: The Guilford Press.
- DeVellis, R. F. (2012). *Scale development. Theory and applications* (3rd ed.). Thousand, Oaks, California: Sage.

- Doctoroff, G. L., & Arnold, D. H. (2017). Doing homework together: The relation between parenting strategies, child engagement, and achievement. *Journal of Applied Developmental Psychology*, 48, 103-113. doi: 10.1016/j.appdev.2017.01.001
- Dumont, H., Trautwein, U., Nagy, G., & Nagengast, B. (2014). Quality of parental homework involvement mediate the relationship between family background and educational outcomes? *Contemporary Educational Psychology*, 37, 55-69. doi: 10.1016/j.cedpsych.2011.09.004
- Escobar-Pérez, J., & Cuervo-Martínez, A. (2008). Validez de contenido y juicio de expertos: Una aproximación a su utilización. *Avances en Medición*, 6, 27-36. Recuperado de <http://servicio.bc.uc.edu.ve/educacion/revista/n33/art12.pdf>
- Field, A. (2009). *Discovering statistic using SPSS*. London: SAGE.
- Furr, R. M., & Bacharach, V. R. (2014). *Psychometrics. An introduction* (2nd ed.). Thousand, Oaks, California: SAGE.
- García, F. I., Valdés, A. A., Carlos, E. A., & Alcántar, C. (2019). Propiedades psicométricas de una escala para medir desconexión moral en niños mexicanos. *Acta Colombiana de Psicología*, 22, 118-128. doi: 10.14718/ACP.2019.22.1.6
- George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn & Bacon.
- Hair, J., Anderson, R., Tatham, R., & Black, W. (1999). *Análisis multivariante* (5^{ta} ed.). Madrid: Prentice Hall.
- Heck, R. H., Thomas, H. I., & Tabata, L. N. (2014). *Multilevel and longitudinal modeling with IBM SPSS* (2nd ed.). New York, NY: Routledge.
- Ho, R. (2006). *Handbook of univariate and multivariate data analysis and interpretation with SPSS*. Boca Raton, Florida: CRC.
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 109-134. doi: 10.1007/BF02291366
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: The Guilford Press.
- Martínez, M. R., Hernández, M. J., & Hernández, M. V. (2006). *Psicometría*. Madrid: Alianza Editorial.
- Méndez, C., & Rondón, M. (2012). Introducción al análisis factorial exploratorio. *Revista Colombiana de Psiquiatría*, 41(1), 197-207. doi: 10.1016/S0034-7450(14)60077-9
- Meyers, L. S., Gamst, G., & Guarino, A. J. (2013). *Applied multivariate research. Design and interpretation* (2nd ed.). Thousand, Oaks, California: SAGE.
- Nunnally, J. C., & Bernstein, I. H. (2010). *Psychometric theory* (3rd ed.). New York: McGraw-Hill/Pearson.
- Núñez, J., Suárez, N., Rosário, P., Vallejo, G., Valle, A., & Epstein, J. (2015).

- Relationships between perceived parental involvement in homework, student homework behaviors, and academic achievement: Differences among elementary, junior high, and high school students. *Metacognition and Learning*, 10, 375-406. doi: 10.1007/s11409-015-9135-5
- Pérez, C. (2004). *Técnicas de análisis multivariante de datos. Aplicaciones con SPSS*. Madrid: Pearson.
- Polit, D. F., & Beck, C. T. (2004). *Nursing research: Principles and methods* (7th ed.). Philadelphia: Lippincott, Williams, & Williams.
- Rouquette, A., Nadot, T., Labitrie, P., Van den Broucke, S., Mancini, J., Rigal, L., & Ringa, V. (2018). Validity and measurement invariance across sex, age, and education level of the French short versions of the European Health Literacy Survey Questionnaire. *Plos One*, 13(12). doi: 10.1371/journal.pone.0208091
- Silinskas, G., & Kikas, E. (2017). Parental involvement in math homework: Links to children performance and motivation. *Scandinavian Journal of Educational Research*, 63, 17-37. doi: 10.1080/00313831.2017.1324901
- Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Hillsdale, NJ: Erlbaum.
- Stevens, S. S. (1946). On the theory of scales measurement. *Science*, 1103, 677-680.
- Tian, L., Chen, H., & Huebner, E. S. (2014). The longitudinal relationships between basic psychological needs satisfaction at school and school-related subjective well-being in adolescent. *Social Indicators Research*, 119, 353-372. doi: 10.1007/s11205-013-0495-4
- Valdés, A., Carlos, E., & Torres, G. (2018). Propiedades psicométricas de una escala para medir cibervictimización en universitarios. *Revista Electrónica de Investigación Educativa*, 20(4), 36-48. doi: 10.24320/redie.2018.20.4.1841
- Valdés, A., Madrid, E., Carlos, E., & Martínez, B. (2016). Propiedades psicométricas de una escala para medir dificultades en habilidades sociales relacionadas con la victimización. *Pensamiento Psicológico*, 14(2), 77-88.
- Valle, A., Regueiro, B., Núñez, J. C., Suárez, N., Freire, C., & Ferrádas, M. (2016). Percepción de la implicación parental en los deberes escolares y rendimiento académico en estudiantes de secundaria. *Revista Española de Pedagogía*, 74(265), 481-498. Recuperado de https://revistapedagogia.org/wp-content/2016/12/percepcion_implicacion.pdf
- Von Stumm, S., & Plomin, R. (2015). Socioeconomic status and the growth of intelligence from infancy through adolescence. *Intelligence*, 48, 30-36. doi: 10.1016/j.intell.2014.10.002
- Vossen, H., Piotrowski, J., & Valkenburg, P. (2015). Development of the adolescent measure of empathy and sympathy (AMES). *Personality and Individual Differences*, 74, 66-71. doi: 10.1016/j.paid.2014.09.040
- Wu, H., & Leung, S-O. (2017). Can Likert scales be treated as interval scales? - A simulation studies. *Journal of Social Service Research*, 43(4), 527-532. doi: 10.1080/01488376.2017.1329775

Índice de tablas

Tabla 1. Relación entre los principios numéricos y el nivel de medición	22
Tabla 2. Resultado de pruebas de normalidad	33
Tabla 3. Estadísticos de tendencia central y dispersión	34
Tabla 4. Prueba de KMO y Bartlett de la escala de Empatía	67
Tabla 5. Comunalidades de los ítems de la escala de Empatía	67
Tabla 6. Autovalores iniciales de la escala de Empatía	68
Tabla 7. Varianza total explicada de la escala de Empatía	68
Tabla 8. Matriz de patrón de la escala de Empatía	69
Tabla 9. Resultados del análisis factorial exploratorio de la escala Empatía	71
Tabla 10. Coeficientes de regresión estandarizados del Modelo Factorial Confirmatorio de la escala para medir Empatía	94

Índice de figuras

Figura 1. Modelo para medir Desconexión Moral en Niños	15
Figura 2. Roles de Espectadores en el Bullying en Niños	16
Figura 3. Distribución normal	25
Figura 4. Tipos de asimetría	25
Figura 5. Tipos de curtosis	26
Figura 6. Modelo clásico de medida	35
Figura 7. Modelo unidimensional de medición del Perdón en adolescentes	38
Figura 8. Escala de Ciberagresión con dos dimensiones correlacionadas	39
Figura 9. Escala de Dificultades en Habilidades Sociales con dimensiones no correlacionadas	40
Figura 10. Modelo teórico de medición de la Cibervictimización en estudiantes universitarios	74
Figura 11a. Modelo no identificado ($g/l = -1$)	75
Figura 11b. Modelo identificado ($g/l = 0$)	76
Figura 11c. Modelo sobreidentificado ($g/l = 2$)	76

Acerca de los autores

Ángel Alberto Valdés Cuervo

Doctor en Ciencias por el Centro de Investigación en Alimentación y Desarrollo. Profesor Investigador Titular C del Departamento de Educación, Instituto Tecnológico de Sonora. Miembro del Cuerpo Académico 'Procesos Educativos'. Perfil PRODEP. Miembro del Sistema Nacional de Investigadores Nivel I. Líneas de Investigación: Familia-escuela, Violencia escolar.

Fernanda Inéz García Vázquez

Doctora en Ciencias Sociales por la Universidad de Sonora. Profesora Investigadora Asociada B del Departamento de Educación, Instituto Tecnológico de Sonora. Miembro del Cuerpo Académico 'Procesos Educativos'. Perfil PRODEP. Miembro del Sistema Nacional de Investigadores Nivel Candidata. Líneas de Investigación: Violencia escolar, Psicología Positiva.

Gisela Margarita Torres-Acuña

Doctora en Sistemas y Ambientes Educativos, Benemérita Universidad Autónoma de Puebla. Profesora Investigadora Titular B del Departamento de Educación, Instituto Tecnológico de Sonora. Miembro del Cuerpo Académico 'Procesos Educativos'. Perfil PRODEP. Miembro del Sistema Nacional de Investigadores Nivel Candidata. Líneas de Investigación: Violencia escolar, Ciberacoso.

Maricela Urías Murrieta

Doctora en Educación, Nova Southeastern University. Miembro del Cuerpo Académico 'Procesos Educativos'. Perfil PRODEP. Miembro del Sistema Nacional de Investigadores Nivel I. Líneas de Investigación: Familia-escuela.

Christian Samhir Grijalva Quiñonez

Maestro en Investigación Educativa. Asistente del Cuerpo Académico Procesos Educativos. Profesor de Asignatura, Dirección de Ciencias Sociales y Humanidades, Instituto Tecnológico de Sonora. Línea de investigación: Familia-escuela.

**MEDICIÓN EN INVESTIGACIÓN EDUCATIVA
CON APOYO DEL SPSS Y EL AMOS**

Diciembre, 2019

Esta obra pretende contribuir a la formación y el desarrollo de investigadores educativos. En particular, contribuir al desarrollo de conocimientos y habilidades para la medición psico-educativa. A diferencia de otros textos en este se muestra cómo utilizar dos softwares ampliamente utilizados por los investigadores (SPSS y AMOS) para realizar análisis de las propiedades psicométricas de los instrumentos. Se describe cómo realizar procedimientos y la interpretación de análisis de normalidad de los datos, fiabilidad y validez de las medidas (análisis factorial exploratorio y confirmatorio). Esperamos que esta obra contribuya al desarrollo de competencias para la medición educativa en los estudiantes e investigadores y a elevar la calidad de sus investigaciones.



ISBN 978-607-437-503-9



9 786074 375039