

Apuntes de estimación de la fiabilidad de consistencia interna de los ítems de un instrumento de medida

Dolores Frías-Navarro

Universidad de Valencia

Citar como:

Frías-Navarro, D. (2025). *Apuntes de estimación de la fiabilidad de consistencia interna de los ítems de un instrumento de medida*. Universidad de Valencia. España. Disponible en: <https://www.uv.es/friasan/AlfaCronbach.pdf>

(revisado 19 noviembre 2025)

Este material de apuntes está siendo revisado y ampliado para ser un capítulo del libro en:

Frías-Navarro, D. y Pascual-Soler, M. (2023). *Escritura científica y lectura crítica. Recomendaciones para redactar el informe de investigación*. Universidad de Valencia. España. Proyecto: "Research design, analysis and writing of results". <https://doi.org/10.17605/osf.io/kngtp>.

Otros materiales relacionados con la metodología de investigación:

-Frías-Navarro, D. (2025). *Esquemas*. Recomendaciones para elaborar el informe de investigación: lectura y valoración crítica. Universidad de Valencia. España. https://www.uv.es/friasan/Recomendaciones_ElaborarInforme_2025.pdf Este material se actualiza periódicamente. Más información en: <https://www.uv.es/friasan/TFG>

-Frías-Navarro, D. (2025). *Protocolo de lectura crítica y análisis metodológico (LECR)*. Universidad de Valencia.

-Frías-Navarro, D. y Pascual-Soler, M. (Eds.). (2022). *Diseño de la investigación, análisis y redacción de los resultados*. Valencia: Palmero Ediciones. [Se trata de una edición revisada y ampliada con nueva información y capítulos, 750 páginas. En este documento se detalla el índice del libro. Se ofrece información teórica e información para llevar a cabo el análisis de las hipótesis y la redacción de los resultados. Se explican los análisis con SPSS (principalmente), JASP y JAMOVI. *Está disponible en copias y revelados* (Palmero Ediciones), calle Menéndez Pelayo, 29, 46010 Valencia y podéis escribir a jpablo@copiasyreveados.com].

-Frías-Navarro, D. (2022). *Anotaciones* sobre el Manual "Diseño de la investigación, análisis y redacción de los resultados". *Recomendaciones para la redacción de los resultados*. Universidad de Valencia. España. [Este material es una ayuda para encontrar en el Manual los aspectos clave para llevar a cabo la redacción de los resultados: ¿cómo se redactan los resultados? Se detallan las principales ideas y las páginas del **Manual de referencia** (Frías-Navarro, D. y Pascual-Soler, M., 2022) donde se desarrolla cómo redactar los resultados de los diseños de investigación].

<https://www.uv.es/friasan>

Índice

Análisis de la consistencia interna de las puntuaciones de un instrumento de medida.....	4
Validez y fiabilidad.....	5
Alfa de Cronbach: coeficiente de fiabilidad de consistencia interna (<i>Internal consistency reliability</i>).....	7
Omega de MacDonald: coeficiente de fiabilidad de consistencia interna	13
Cálculo del alfa de Cronbach.....	14
Interpretación de los valores de consistencia interna.....	15
Tamaño de muestra necesario.....	17
SPSS: análisis de la consistencia interna	18
Ejercicio 1.....	18
Ventana de Resultados del SPSS	21
Interpretación de los resultados que ofrece el SPSS del alfa de Cronbach	23
Correlación elemento-total corregida: índice de homogeneidad corregido	24
Intervalo de confianza del valor de la consistencia interna.....	25
Ejercicio 2	26
SPSS.....	27
JASP.....	29
JAMOVI	29
Estimar el intervalo de confianza cuando se conoce alfa y el número de observaciones (N)	29
Referencias (en proceso).....	31
Análisis factorial de los ítems de una escala.....	31

Análisis de la consistencia interna de las puntuaciones de un instrumento de medida

En muchas ocasiones en las Ciencias Sociales o en las Ciencias de la Salud es necesario medir y cuantificar la expresión de un constructo o variable no observable directamente (por ejemplo, prejuicio, racismo, xenofobia, homofobia internalizada, ansiedad, ira, inteligencia, empatía, razones para emigrar, percepción de beneficios y riesgos de la inteligencia artificial...) ya que es imposible medirlo de forma explícita o directa. Para realizar su medición se utiliza un conjunto de cuestiones, preguntas o ítems que describen de forma operativa u observable al constructo y, posteriormente, se combinan las respuestas de los sujetos sumando sus valores y obteniendo una puntuación total en el instrumento. Pero cuando los ítems son sumados para formar una escala es necesario demostrar previamente que existe consistencia interna entre dichos ítems y, por lo tanto, es correcto sumarlos para definirlos como puntuación total de la medida de un constructo común. Con ello, se podrá conocer si las puntuaciones de la muestra evaluada en el estudio son fiables desde el punto de vista de la consistencia interna de sus respuestas, aportando evidencia a la calidad de medición. Es decir, todos los ítems deben medir un mismo constructo (un mismo dominio psicológico) y por lo tanto estarán correlacionados entre sí. Se trata de que exista consistencia interna u homogeneidad entre los ítems que forman la dimensión o escala.

La fiabilidad de una escala o instrumento de medida se refiere al grado en que los resultados obtenidos con la prueba son consistentes y reproducibles. Es decir, la fiabilidad hace referencia a cuán estables o consistentes serían los resultados si la prueba se repitiera bajo las mismas condiciones, en el mismo individuo, por diferentes personas/evaluadores/clínicos o cuando es completada por el sujeto en más de una ocasión. Se considera fiable si se obtienen resultados similares cada vez que se aplica al mismo sujeto en varias ocasiones. Un instrumento de medida puede ser fiable y no ser válido, pues mide un constructo de forma consistente, pero mide de forma consistente el constructo equivocado. La fiabilidad no garantiza la validez del instrumento.

Existen tres tipos principales de fiabilidad para cualquier instrumento. El primero se denomina fiabilidad test-retest: evalúa la consistencia de los resultados de la prueba a lo largo del tiempo. El segundo, es la fiabilidad interevaluador o fiabilidad interjueces: un buen instrumento también debería medir un constructo de forma consistente independientemente de quién lo administre. En tercer lugar, es muy importante que los ítems que forman la escala tengan consistencia interna: el instrumento debe medir la variable o constructo de manera consistente si los ítems que forman esa

escala miden el mismo constructo o característica subyacente; es decir, los ítems están interrelacionados

Este material está dedicado al estudio de la fiabilidad como consistencia interna de los ítems que forman un constructo. Una alta consistencia interna es deseable para cualquier instrumento ya que implica que todos los ítems de la prueba están correlacionados entre sí y que todos ellos miden el mismo constructo. Por el contrario, una baja consistencia interna indica que los ítems están poco relacionados entre sí y, por lo tanto, probablemente están midiendo constructos diferentes.

Validez y fiabilidad

Los instrumentos de medida de los constructos tienen que demostrar que sus puntuaciones en la muestra con la que se está trabajando poseen las propiedades psicométricas de validez y fiabilidad. Y sus propiedades deben calcularse con cada una de las muestras en las que se aplica, pues sus cualidades psicométricas no son una propiedad inherente del instrumento y podría ser fiable y/o válido con una muestra de participantes, pero no serlo con otra muestra diferente. Por ello, la fiabilidad o la validez no es del test sino de las puntuaciones obtenidas con el instrumento de medida en una muestra concreta de participantes y no es una propiedad inherente (inmutable) del test o instrumento de medida. En este sentido, **Sánchez-Meca y López-Pina (2008)** señalan que utilizar una expresión en un informe como “la fiabilidad de la escala es .80” es totalmente incorrecto. La redacción debe señalar, por ejemplo, “la fiabilidad de las puntuaciones de la escala en la muestra del estudio es .80”.

“The reliability of a scale indicates whether it is measuring an attribute in a way that is reproducible and consistent”

La validez de una escala se refiere al grado en que esa medida evalúa con precisión el concepto, rasgo o constructo específico que afirma estar evaluando. Hace referencia a la veracidad de la medida llevada a cabo con un instrumento. Es decir, se refiere al grado en que el instrumento mide aquello que pretende medir o en qué grado las puntuaciones obtenidas representan a la variable objeto de medición. En otras palabras, los ítems que forman la escala miden aquello para lo que el instrumento fue desarrollado.

La fiabilidad de la escala se refiere a la consistencia, estabilidad y repetición de una medición. Analiza cuánto error aleatorio puede estar distorsionando las puntuaciones o introduciendo “ruido” no deseado en los datos. La fiabilidad de las puntuaciones está relacionada con la posibilidad de reproducir los resultados en repetidas ocasiones con el mismo instrumento, es decir, se trata de la estabilidad de la medición (reproducible y consistente). Por ejemplo, un cuestionario para medir la

depresión puede incluir varias preguntas sobre aspectos como la tristeza, cambios en el sueño y el apetito, fatiga y pérdida de interés. Se asume que las respuestas de las personas a estos diferentes ítems de síntomas deberían ser bastante consistentes.

En resumen, la validez se refiere a qué se está midiendo, mientras que la fiabilidad se refiere a con cuánta consistencia se está midiendo.

Una medida no fiable no puede ser verdaderamente válida, porque si una medida produce puntuaciones inconsistentes e impredecibles, claramente no está midiendo el rasgo o la cualidad que pretende medir de un modo veraz y sistemático. Establecer la fiabilidad proporciona la base para poder determinar la validez de la medida. Pero conviene tener en cuenta que la fiabilidad es una condición necesaria, pero no suficiente, para la validez. Es decir, una prueba puede ser fiable, produciendo resultados de forma consistente, sin ser válida, es decir, sin medir con precisión el atributo o constructo que se pretende evaluar. Sin embargo, una prueba válida (realmente mide aquello que afirma medir) debe ser fiable. En la búsqueda de una investigación psicológica rigurosa, tanto la validez como la fiabilidad son imprescindibles: la validez para asegurar que se está midiendo el constructo correcto y la fiabilidad para garantizar que se está midiendo de manera exacta y precisa.

Nunnally (1967) definía la fiabilidad como el grado en que las medidas de los constructos son repetibles y cualquier influencia aleatoria que pudiese hacer las medidas diferentes de una medición a otra es una fuente de error de medida. Por lo tanto, la fiabilidad indica una consistencia de la medida del constructo. En este sentido, es deseable que se obtengan puntuaciones similares (consistentes) cuando el instrumento de medida se administra repetidamente a los sujetos (diferentes muestras o la misma muestra en diferentes contextos o tiempos) bajo las mismas condiciones. Se trataría de un instrumento con una alta fiabilidad, es decir libre de errores (pocos errores). Si el valor de fiabilidad es bajo indicará que hay errores de medición y, por lo tanto, los resultados de los análisis estadísticos estarían sesgados al carecer de fiabilidad o consistencia interna. Una definición formal de la fiabilidad desde el modelo de la teoría clásica de test define a la fiabilidad como la razón entre la varianza de la puntuación verdadera y la varianza de la puntuación total (**Revelle y Zinbarg, 2009**).

Por ejemplo, un instrumento de medida es fiable si proporciona aproximadamente el mismo tipo de respuestas para diferentes grupos de sujetos o en repetidas aplicaciones a un mismo grupo de sujetos. Además, será válido si correlaciona con las puntuaciones de otros instrumentos que miden constructos donde es esperable una convergencia (validez convergente). En cambio, la

validez discriminante está presente cuando se utilizan instrumentos para medir diferentes constructos y sus puntuaciones están débilmente correlacionadas.

Desde el punto de vista de la teoría clásica de tests (*Classical Test Theory*, CTT; ver, por ejemplo, [Lord y Novick, 1968](#)), el cálculo del estadístico de consistencia interna considera que una parte de la variabilidad de las respuestas de los participantes se debe a una verdadera diferencia subyacente (varianza verdadera, T) en el constructo que está siendo medido (X). El resto de la variabilidad se considera que es error aleatorio de medida (E). Por lo tanto, $X = T + E$.

La fiabilidad se define como la razón entre la variabilidad de T (σ^2_T) y la variabilidad de X (σ^2_X). Es decir, es la proporción de la variabilidad en la puntuación del instrumento (por ejemplo, puntuación en la variable de homofobia) que se atribuye al rasgo que está siendo medido (homofobia) respecto a la variabilidad total que contienen las respuestas de los sujetos (el rasgo más el error de medida). Por ello, la fiabilidad es la proporción de la varianza total que se atribuye a la varianza verdadera y así, cuanto mayor sea la proporción de la varianza total atribuida a la varianza verdadera mayor será la fiabilidad de la medida. Por lo tanto, se trata de computar la σ^2_T / σ^2_X , donde $\sigma^2_X = \sigma^2_T + \sigma^2_E$.

En definitiva, la teoría clásica de test plantea que cualquier cuestión que influya en el error cambiará de forma inevitable la fiabilidad y por ello la fiabilidad no es una propiedad del test o prueba, sino que es una propiedad de una escala aplicada en un determinado contexto con una muestra particular ([Miller, 1995](#); [Thompson & Vacha-Haase, 2000](#)).

Existen diferentes tipos de análisis de fiabilidad, pero la fiabilidad basada en la consistencia interna de los ítems es uno de los más populares. Además, existen diferentes métricas para proporcionar evidencia de la fiabilidad basada en la consistencia interna, siendo el alfa de Cronbach (α) el coeficiente más popular. El coeficiente omega de McDonald (ω) es cada vez más utilizado. [Doval y cols. \(2023\)](#) señalan que tanto α como ω son adecuados para ítems que se distribuyen de forma aproximadamente normal y medidas aproximadamente unidimensionales y congenéricas sin cargas factoriales extremas. Sin embargo, cuando los ítems tienen otra distribución, un fuerte componente específico o sus errores están correlacionados resultan más adecuadas variantes de ω .

Alfa de Cronbach: coeficiente de fiabilidad de consistencia interna (*Internal consistency reliability*)

La fiabilidad de la consistencia interna de las puntuaciones del instrumento se refiere al grado en que los distintos ítems de una prueba o cuestionario, que están diseñados para medir el mismo

constructo, producen puntuaciones similares. Por ejemplo, un cuestionario que mide la sintomatología de depresión puede incluir varias preguntas sobre aspectos como la tristeza, los cambios en el sueño y el apetito, la fatiga y la pérdida de interés. En ese caso, se asume que las respuestas de las personas a estos distintos ítems de síntomas de depresión deberían ser bastante consistentes entre sí. Se supone que todos los ítems reflejan el mismo constructo subyacente, por lo que las puntuaciones de las personas en esos ítems deberían estar correlacionadas entre sí. En la Escala de Autoestima de Rosenberg, por ejemplo, las personas que están de acuerdo con que son “una persona valiosa” deberían tender a estar de acuerdo también con que “tienen buenas cualidades”. Si las respuestas de las personas a los distintos ítems no estuvieran correlacionadas entre sí, entonces dejaría de tener sentido afirmar que todos ellos están midiendo el mismo constructo subyacente.

La fiabilidad de la consistencia interna de las puntuaciones del instrumento se puede estimar con el alfa de Cronbach (α) o también con el método omega de McDonald (ω), entre otros coeficientes. Actualmente, son los dos coeficientes que más se utilizan en las publicaciones de las Ciencias Sociales y de la Salud, especialmente el alfa de Cronbach como coeficiente tradicional, pero la omega de McDonald se aplica cada vez más para tratar de paliar el uso inadecuado del alfa de Cronbach cuando no se cumplen sus supuestos estadísticos. El grupo de trabajo de la American Psychological Association ([Wilkinson and APA Task Force on Statistical Inference, 1999](#)) enfatiza el uso correcto de la fiabilidad.

En resumen, la fiabilidad de consistencia interna se refiere a la consistencia de las respuestas de las personas a lo largo de los ítems de una medida con múltiples ítems. En general, todos los ítems de este tipo de instrumentos se supone que reflejan el mismo constructo subyacente, por lo que las puntuaciones de las personas en esos ítems deberían estar correlacionadas entre sí.

El alfa de Cronbach es un estadístico que permite cuantificar la fiabilidad de consistencia interna. Calcula la correlación media entre ítems del conjunto de ítems que forman la escala. Sus valores oscilan entre 0 y 1, donde los valores más altos indican una mayor consistencia interna. Como criterio general, el valor de alfa debería ser igual o superior a .70 para sugerir una fiabilidad adecuada. Por ejemplo, un alfa de .90 para un cuestionario de depresión significa que existe una alta correlación media entre las puntuaciones de las personas en los distintos ítems. Este resultado indica que todos los ítems están midiendo de forma consistente el mismo constructo subyacente (la depresión). Es decir, señala la unidimensionalidad de la escala, aportando evidencia de que está midiendo una sola cosa. Si algunos ítems no estuvieran relacionados con los demás, las correlaciones medias entre ítems serían más bajas, lo que daría lugar a un alfa más bajo. Esto

indicaría la presencia de múltiples dimensiones en la escala, en lugar de un único concepto unificado.

El coeficiente más conocido y aplicado en la literatura psicológica, hasta el momento, para medir la fiabilidad de la consistencia interna de los ítems que forman la puntuación total de una escala de medida es el denominado alfa de Cronbach (α , [Cronbach, 1951](#)). Es decir, el alfa de Cronbach estima cómo de fiables son las respuestas dadas a un conjunto de ítems señalando el grado de consistencia de las respuestas (estabilidad) respecto al dominio psicológico medido. Una alta fiabilidad de consistencia interna, mostrada por un alfa de Cronbach elevado, apoya la idea de que los distintos ítems de una escala realmente están captando la misma variable latente que se pretende medir. Por lo tanto, los ítems se articulan de forma coherente y significativa entre sí para medir de manera fiable dicho constructo.

En la **Escala de Autoestima de Rosenberg**, por ejemplo, las personas que están de acuerdo con que son “una persona con valía” deberían tender a estar de acuerdo también con que “tienen una serie de buenas cualidades”. Si las respuestas de las personas a los distintos ítems no estuvieran correlacionadas entre sí, entonces dejaría de tener sentido afirmar que todos ellos están midiendo el mismo constructo subyacente.

Esto es tan cierto para las medidas conductuales y fisiológicas como para las medidas de autoinforme. Por ejemplo, se podría pedir a las personas que realicen una serie de apuestas en un juego de ruleta simulado como medida de su nivel de **búsqueda de riesgo**. Esta medida tendría consistencia interna en la medida en que las apuestas de cada participante fueran de forma consistente altas o bajas a lo largo de los distintos ensayos.

Según el estudio de [Hogan y cols. \(2000\)](#), el 75% de los artículos publicados utilizan el alfa de Cronbach como coeficiente para estimar la fiabilidad, entendida como consistencia interna de los ítems que miden un constructo. Y, a pesar de las críticas que ha recibido ([Green y cols., 1977; Green y Yang, 2015; Green y Hershberger, 2000; Huysamen, 2006; Kelley y Cheng, 2012; Raykov, 1998; Sijtsma, 2009; Zimmerman y cols., 1993; Zinbarg y cols., 2005](#)) sigue siendo el estadístico de consistencia interna más utilizado en la actualidad.

El método de consistencia interna permite estimar la fiabilidad de un instrumento de medida a través de un conjunto de ítems que se espera que midan el mismo constructo o una única dimensión teórica de un constructo latente. Aporta información sobre la homogeneidad de los ítems que forman la escala. Es decir, informa en qué grado las puntuaciones de cada ítem se correlaciona con las puntuaciones del resto de ítems que forman la escala ([Cronbach y Shavelson, 2004](#)). [Streiner y Norman \(2003\)](#) señalan que la fiabilidad de la consistencia interna es el promedio de las

correlaciones entre todos los ítems de la escala. Sus valores pueden oscilar de 0 a 1 y cuanto mayor es su valor mayor es el grado de consistencia interna (homogeneidad) de los ítems que forman la escala. Cuando los datos tienen una estructura multidimensional el valor de consistencia interna será bajo. Es decir, no se observa una consistencia en las puntuaciones que forman el constructo teórico que se desea medir.

El coeficiente alfa de Cronbach y el omega de McDonald son modelos de consistencia interna, basados en el promedio de las correlaciones entre los ítems. En ambos casos se puede evaluar cuánto mejoraría (o empeoraría) la fiabilidad de la prueba si se excluyera un determinado ítem; procedimiento que posteriormente se analizará con más detalle y cuya principal utilidad radica en observar si eliminar un ítem mejoraría de forma sustancial el valor de la consistencia interna. Conviene tener en cuenta que se eliminará un ítem solamente cuando mejore de forma sustancial el valor de la consistencia interna o si se ha obtenido un valor de consistencia interna bajo y al eliminar un ítem se obtiene un valor que se considera aceptable. Es importante valorar con detalle si se elimina el ítem ya que si se hace entonces no se podrán llevar a cabo comparaciones directas con los datos de las puntuaciones medias de otros estudios ya que la amplitud de la escala en cada situación será diferente. Si un ítem no funciona adecuadamente lo primero es reflexionar sobre por qué ha ocurrido, qué ha podido pasar, qué tipo de muestra se utilizó, en qué contexto se llevó a cabo la medición y con ello valorar una posible explicación relacionada con el diseño del estudio y las condiciones de su ejecución. El coeficiente Kuder-Richardson 20 es un tipo especial de alfa de Cronbach para escalas con respuestas dicotómicas ([Kuder y Richardson, 1937](#)).

[Cronbach \(1951\)](#) propone el coeficiente alfa (α) que estima la proporción de varianza de un instrumento de medida debido al factor común entre los ítems. Y es conveniente que los investigadores e investigadoras tengan en cuenta la valoración del cumplimiento de sus supuestos básicos:

- 1) La denominada tau-equivalencia que consiste en que los ítems midan el mismo rasgo o la misma variable latente con el mismo o parecido grado de precisión ([Cho, 2016](#)).
- 2) La no correlación de los errores ya que se asume que son independientes (la puntuación de error de cualquier par de ítems no está correlacionada).
- 3) La unidimensionalidad, es decir, que todos los ítems o preguntas deben medir un solo rasgo latente
- 4) La medida del constructo debe de ser continua.

El supuesto de equivalencia tau asume la misma puntuación verdadera para todos los ítems de la escala o cargas factoriales de todos los ítems iguales en un modelo factorial y es un requisito para que α no esté sesgado (Cronbach, 1951). Si el supuesto de la tau-equivalencia no se cumple, el valor del alfa de Cronbach tiende a infraestimar (subestimar) la fiabilidad de la puntuación del test. Y, si el supuesto de ausencia de correlación entre los errores se viola entonces el valor del alfa de Cronbach tiende a sobreestimar el valor de la fiabilidad (Cho y Kim, 2015; Lucke, 2005; Raykov y Marcoulides, 2017).

Por lo tanto, si el modelo de medida tau-equivalente no se cumple entonces otros coeficientes pueden ser una alternativa mejor como el coeficiente omega. Cuando existe tau-equivalencia entonces los valores de alfa de Cronbach y de omega de MacDonald coinciden o son muy próximos (Trizano-Hermosilla y Alvarado, 2016).

Es necesario que el investigador o investigadora valore la naturaleza de sus datos antes de elegir el coeficiente de fiabilidad más adecuado, pues en muchas ocasiones son los usos inadecuados de los estadísticos y coeficientes y las interpretaciones incorrectas los problemas que rodean a la calidad de los resultados de la Ciencia (Frías-Navarro y cols., 2021; Monterde-i-Bort y cols., 2010).

El valor de alfa oscila de 0 a 1. Cuanto más cerca se encuentre el valor del alfa a 1 mayor es la consistencia interna de los ítems analizados, es decir, se asume que los ítems están midiendo una misma dimensión. Si los ítems están positivamente correlacionados entonces la varianza de la suma de los ítems se incrementa. Por ello, si las puntuaciones en todos los ítems fuesen idénticas y, por lo tanto, las puntuaciones estarían perfectamente correlacionadas, el valor de alfa sería igual a 1. En cambio, si los ítems fuesen totalmente independientes, no mostrando ningún tipo de relación entre ellos, el valor de alfa sería igual a 0.

Se recomienda informar en los artículos o informes de la estimación puntual del valor de consistencia interna junto con su intervalo de confianza para valorar la precisión de la estimación y permitir los análisis secundarios como los estudios de meta-análisis de fiabilidad de un instrumento determinado o de un constructo concreto (Dominguez & Merino, 2015; Sánchez-Meca & López-Pina, 2008). Además, si el valor del intervalo inferior es extremadamente bajo ($\alpha < .60$) se podría tomar la decisión de considerar que la fiabilidad de la escala es inadecuada ya que el error de medida sería grande (Dominguez-Lara, 2016). Posteriormente se detallará cómo estimar el intervalo de confianza.

Una vez comprobado que el valor de alfa es aceptable ya se puede sumar las puntuaciones de los ítems y formar una puntuación total de una escala. Por ello, siempre hay que comprobar la

consistencia interna de las puntuaciones de la muestra antes de computar puntuaciones totales o realizar cualquier otro análisis estadístico. Si la consistencia interna no es adecuada entonces será necesario estudiar si representan a dimensiones diferentes (por ejemplo, realizando un análisis factorial), pero nunca se utilizarán como elementos que miden un mismo constructo.

Los resultados del estudio de la consistencia interna mediante el alfa de Cronbach u otro coeficiente deben redactarse siempre en el apartado de instrumentos del informe de investigación o artículo, señalando su valor y su intervalo de confianza. Con ello se facilita que el lector o lectora valore la calidad de las puntuaciones que forman la dimensión o escala desde el punto de vista de la consistencia de los ítems que la forman. Y en ese apartado se debe describir de forma detallada cómo es la medición que se lleva a cabo con el instrumento: breve descripción del constructo que se mide, número de ítems que forman la escala, tipo de escala de respuesta, correlación media de los ítems con la puntuación total, valor del coeficiente de fiabilidad y su intervalo de confianza y el autor o autora del instrumento original que se ha utilizado cuya obra siempre debe ser referenciada o detallar si se ha elaborado un conjunto de ítems para la investigación.

En cambio, si el estudio que se ha realizado tiene como principal objetivo presentar la creación de un nuevo instrumento o se trata de una adaptación de un instrumento entonces se recomienda que la redacción detallada de las propiedades psicométricas se realice en el apartado de Resultados ya que es el objetivo de ese trabajo (necesidad de conocimiento, [Frías-Navarro & Pascual-Soler, 2022](#)).

Respecto a las limitaciones del coeficiente alfa de Cronbach, hay que tener en cuenta que le afecta el número de ítems, el número de alternativas de respuesta y la proporción de la varianza del test ([Domínguez-Lara & MerinoSoto, 2015](#)). Es recomendable tener al menos 3 ítems en la escala o instrumento de medida para poder estimar el valor del coeficiente alfa de Cronbach y cuanto mayor el número de ítems mayor será la fiabilidad de la escala. La medida de la consistencia interna mediante el alfa de Cronbach asume que los ítems (medidos en escala tipo *Likert*) miden un mismo constructo y que están altamente correlacionados entre sí ([Welch & Comer, 1988](#)). Por lo tanto, conviene tener en cuenta que la utilización del alfa de Cronbach podría no ser correcta cuando la naturaleza de la escala de respuesta es de tipo ordinal ya que uno de los supuestos del coeficiente alfa de Cronbach es la naturaleza continua de las variables donde tiene sentido aplicar escalas de respuesta tipo *Likert* ya que plantean intervalos de respuesta que son comparables entre sí ([Elosua & Zumbo, 2008](#)).

Todas las escalas de respuesta que utilizan la escala tipo *Likert* son polítómicas, es decir, tienen más de dos alternativas de respuesta. Se trata de escalas de respuesta ordenada o graduadas donde

la asignación de los números enteros a las distintas opciones de respuestas es una escala ordinal. A partir de las puntuaciones obtenidas por distintos sujetos en un mismo ítem se pueden realizar inferencias sobre quiénes tienen una actitud más o menos favorable, pero no se puede determinar cuál es la distancia entre los sujetos que responden con alternativas diferentes. Estrictamente hablando, en la mayoría de los instrumentos que miden constructos psicológicos, no se trata de escalas de respuesta de intervalo ya que realmente son escalas ordinales. De ahí que, en general, si las opciones de respuesta son de 6 o menos alternativas se recomienda optar por el estadístico de omega de McDonald. Sin embargo, habría que comprobar si se produce la tau-equivalencia de los ítems, en cualquier caso. El SPSS, hasta la versión 28 de 2024, no ofrece la estimación del valor de omega, siendo necesario recurrir al programa JASP o JAMOVI, por ejemplo.

Omega de MacDonald: coeficiente de fiabilidad de consistencia interna

El coeficiente Omega de McDonald (ω , [McDonald, 1999](#)) se plantea como un sustituto del alfa de Cronbach ya que controla algunos de sus sesgos como, por ejemplo, cuando no se mantiene el modelo de medida de la tau-equivalencia. Cuando dicho modelo se viola entonces el verdadero valor de fiabilidad se subestimará con el coeficiente alfa de Cronbach, dependiendo su valor de la gravedad de la violación. En ese caso, omega de McDonald corrige el sesgo ya que no exige el modelo de medida de la tau-equivalencia ni tampoco que los errores no estén correlacionados tal y como se especifica en el alfa de Cronbach ([Dunn y cols., 2014](#)). La omega de McDonald también se conoce como fiabilidad compuesta ('Composite Reliability' o como coeficiente Jöreskog, [Moral de la Rubia, 2019; Viladrich y cols., 2017](#)). Cuando los ítems mantienen la tau-equivalencia entonces los coeficientes de alfa y omega son iguales a la fiabilidad compuesta para el conjunto de ítems.

El coeficiente omega de McDonald se aplica cuando la escala de respuesta es ordinal y, en ocasiones se considera más adecuado, pues el alfa de Cronbach tiene como supuesto aplicarse cuando la escala de respuesta de las variables es de naturaleza continua en escala de intervalo tal y como ya se ha comentado. Además, el alfa de Cronbach podría subestimar la magnitud de la consistencia interna cuando la escala de respuesta tiene cinco alternativas de respuesta o menos, siendo preferible optar por la omega de McDonald cuando la escala de respuesta es binaria y ordinal ([Zumbo y cols., 2007](#)). Ante esta situación algunos autores recomiendan abandonar el uso del alfa de Cronbach y optar por la omega de McDonald dentro del ámbito de la Psicología ([Peters, 2014](#)). Por todo ello, se recomienda utilizar / construir instrumentos de medida con al menos 6 opciones de

respuesta ya que con este tipo de opciones se estabiliza el valor de alfa de Cronbach (Lozano y cols., 2008).

El coeficiente de consistencia interna ordinal omega es conceptualmente semejante al alfa de Cronbach y la diferencia principal se encuentra en que el valor de omega se basa en la matriz de correlación policórica entre los ítems (más adecuada para datos ordinales) en lugar de la matriz de covarianza (correlación) de Pearson (más adecuada para datos continuos) y por ello es más apropiado para estimar la consistencia interna cuando los datos son ordinales (Peters, 2014).

En definitiva, no se aconseja estimar el alfa de Cronbach cuando no se cumple el supuesto de tau-equivalencia de los ítems, los errores están correlacionados y la escala de medida no es de intervalo y tiene menos de 6 alternativas de respuesta. Requisitos, por otra parte, difíciles de cumplir en los instrumentos de medida que se aplican en general en las Ciencias Sociales y la Psicología (Dunn y cols., 2014; Graham, 2006) y, por lo tanto, se necesitará estimar otros coeficientes como el omega de McDonald que es más robusto, tal y como ya se ha comentado (Viladrich y cols., 2017).

Cálculo del alfa de Cronbach

El alfa de Cronbach puede ser calculado a partir de la correlación entre los ítems como una función del número de ítems del instrumento y la correlación media entre los ítems. La fórmula sería la siguiente:

$$\alpha = \frac{N \times \bar{r}}{1 + (N - 1) \times \bar{r}}$$

Donde N es igual al número de ítems y \bar{r} es la correlación media entre los ítems.

Como se puede observar en la fórmula, si se incrementa el número de ítems entonces se incrementa el valor del alfa de Cronbach. Además, si la correlación media entre los ítems es baja entonces el valor del alfa de Cronbach será bajo. En este sentido, a medida que la correlación media entre los ítems se incrementa también se incrementa el valor del alfa de Cronbach. Si la correlación entre los ítems es alta entonces habrá evidencia de que los ítems están midiendo un mismo constructo o dimensión latente.

El análisis de las correlaciones ítem-puntuación total permite valorar la relevancia y la utilidad de cada ítem dentro de la escala. Se recomienda que los autores o autoras de los manuscritos informen de la correlación media entre los ítems y describan la amplitud de los valores de los

coeficientes de correlación obtenidos con el total de los ítems. Así, los lectores y lectoras podrán llevar a cabo una lectura crítica o activa de la calidad de la consistencia interna de los ítems.

Cuando se obtiene un valor de alfa de Cronbach bajo entonces los ítems tienen una estructura multidimensional y es aconsejable realizar un análisis factorial de los ítems para observar cómo se agrupan los ítems en dimensiones latentes diferentes y entonces calcular el valor de alfa de Cronbach para cada dimensión por separado que ha detectado el análisis factorial.

Un valor aceptable de alfa de Cronbach señala que existe consistencia interna, es decir, los ítems están inter-relacionados, pero nunca señala el grado de unidimensionalidad de los ítems que miden el constructo. La consistencia interna es una condición necesaria pero no suficiente para señalar la unidimensionalidad. Un conjunto de ítems puede que estén inter-relacionados y, sin embargo, mostrar multidimensionalidad. Por ello, conviene tener en cuenta que obtener un valor de alfa de Cronbach alto no significa que la medida del constructo es unidimensional. Para comprobar que la escala o instrumento es unidimensional es necesario ejecutar otros análisis como el Análisis Factorial Exploratorio (EFA) o el Análisis Factorial Confirmatorio (AFC) y comprobar la dimensionalidad del conjunto de ítems (Frías-Navarro & Pascual-Soler, 2012). En el trabajo de Boateng (2018) se puede ampliar con detalle la información y los valores que se utilizan como referente para considerar el ajuste de las soluciones factoriales en el AFC.

Interpretación de los valores de consistencia interna

Como nota importante, destacar que los lectores y lectoras deben tener en cuenta que el valor de alfa de Cronbach de un instrumento es específico de la muestra del estudio. Por tanto, los autores deben informar ese valor para cada instrumento que utilicen en su investigación. Su valor no es del instrumento sino de las respuestas de la muestra de participantes. Por ello, es necesario que en la descripción del instrumento de medida se anote siempre el valor de consistencia obtenido y su intervalo de confianza (95% IC). Esta información es necesaria e imprescindible cuando se publica un trabajo de investigación ya que facilita la lectura crítica de los lectores y lectoras para valorar si sus valores podrían comprometer los resultados aportados con el uso de ese instrumento de medida.

Como criterio general, los autores sugieren las recomendaciones siguientes para evaluar los valores de los coeficientes de alfa de Cronbach (las mismas recomendaciones para el coeficiente omega de McDonald) (George & Mallery, 2003, p. 231; Menon y cols., 2025, p. 727):

- Coeficiente alfa >.95: es redundante.
- Coeficiente alfa >.90 a .95 es excelente

- Coeficiente alfa > .80 es bueno
- Coeficiente alfa > .70 es aceptable
- Coeficiente alfa > .60 es cuestionable
- Coeficiente alfa < .50 es inaceptable

Opiniones de otros investigadores o investigadoras:

- **Gliem y Gliem (2003):** un valor de alfa de .80 es probablemente una meta razonable.
- **Hair (1998):** el límite menor del alfa de Cronbach debe ser .70.
- **Huh, Delorme, y Reid (2006):** el valor de consistencia interna en investigación exploratoria debe ser igual o mayor a .60 y en estudios confirmatorios debe estar entre .70 y .80.
- **Kaplan y Saccuzzo (1982):** el valor de consistencia interna para la investigación básica entre .70 y .80 y en investigación aplicada sobre .95.
- **Loewenthal (1996):** sugiere que un valor de consistencia interna de .60 puede ser considerado aceptable para escalas con menos de 10 ítems.
- **Loo (2001):** el valor de consistencia que se considera adecuado es de .80 o más.
- **Nunnally (1978):** dentro de un análisis exploratorio estándar, el valor de consistencia interna en torno a .70 es adecuado y es el nivel mínimo aceptable (**Nunnally & Bernstein, 1994**). El autor reconoce que valores más bajos son utilizados a veces en la literatura. En las primeras fases de la investigación o estudios exploratorios un valor de consistencia interna de .60 puede ser suficiente. Con investigación básica se necesita, al menos, .80 y en investigación aplicada entre .90 y .95.
- **Shrestha (2021):** valor de consistencia interna de, al menos, .70.

Por lo tanto, valores de consistencia interna menores a .70 señalan una baja correlación entre los ítems. Valores menores .50 se considera totalmente inaceptables. Valores por encima de .95 son muy elevados y se considera que son indicadores de redundancia o duplicación de ítems (solapamiento excesivo y posible redundancia en algunos ítems de la escala), pues por lo menos un par de ítems miden exactamente el mismo aspecto de constructo y uno de ellos debería eliminarse (**Kline, 2015**).

En definitiva, el punto de corte del valor de .7 o más se considera óptimo para decidir si el conjunto de ítems es consistente (fiabilidad) en la medida del constructo.

Tal y como se ha anotado anteriormente, cuando se comprueba que hay ítems que tienen una escasa correlación entre el ítem y la puntuación total se puede incrementar el valor de la consistencia interna eliminando ese ítem de la escala ya que dicho ítem se considera de poca calidad metodológica como medio para explorar el constructo que es objeto de medición. Sin embargo, hay que tener muy en cuenta que dicha modificación en el número total de ítems que forman la escala se anotará siempre en el apartado de instrumentos cuando se describa al instrumento para alertar al lector o lectora de que las puntuaciones totales se han calculado con un ítem menos y por ello no se pueden comparar las puntuaciones medias de diferentes estudios que han aplicado el mismo instrumento de medida ya que su amplitud es diferente.

En resumen, la consistencia interna es una medida de la correlación que existe entre los ítems que forman el instrumento de medida. En general, los valores son aceptables cuando son iguales o superiores a .70 y menores o iguales a .95. Cuando se interpreta el valor del coeficiente de consistencia interna es conveniente tener en cuenta la longitud de la escala. Algunos autores consideran que las escalas con 10 ítems homogéneos es el mínimo recomendable para obtener un nivel de fiabilidad adecuado (Kline, 1993). Sin embargo, actualmente se han desarrollado muchas escalas con un número menor de ítems. En esos casos se puede considerar que el conjunto de los ítems tiene consistencia interna con valores inferiores a .7 (Todd y Bradley, 1994).

El análisis de la fiabilidad debe realizarse siempre que se aplica el instrumento, aunque se haya comprobado su consistencia interna con otras muestras ya que su valor varía según las características de los participantes. Por lo tanto, la fiabilidad no es una propiedad del test sino de las puntuaciones que se obtienen con el test en una determinada muestra.

Tamaño de muestra necesario

Respecto al tamaño de la muestra, generalmente, las muestras grandes reducen la posibilidad de error. Como señalan Campos-Arias y Oviedo (2008), desde un punto de vista conservador se recomienda utilizar muestras con al menos 300 (Nunnally y Bernstein, 1994) o 400 participantes (Sánchez & Gómez, 1998). Desde un punto de vista más flexible, para escalas de 20 ítems se acepta un tamaño de la muestra entre 5 y 20 participantes por cada ítem de la escala, es decir, entre 100 y 400 participantes (Cortina, 1993). Como regla general se considera que es necesario al menos 10 participantes para cada ítem de la escala (Nunnally, 1978). Rouquette y Falissard (2011) destacan que para determinar el tamaño de muestra es necesario tener en cuenta 3 criterios: el número de ítems de la escala, el número de factores o sub-escalas que tiene el instrumento y el tipo de análisis factorial que se desea aplicar: análisis factorial exploratorio (PCA) o análisis factorial confirmatorio

(EFA). Por ejemplo, en la tabla 6 de su artículo, detallan que para una escala de 10 ítems y dos factores se necesitarían 300 respuestas si se aplica PCA y 500 si se aplica EFA. Concluyen **Rouquette y Falissard (2011)** que, en términos generales, se necesita una muestra de 300 respuestas, pero sería necesario incrementarla en tres casos: cuando el número de factores o sub-escalas es alto, cuando se desea aplicar EFA como método de análisis y cuando el número de ítems es pequeño.

SPSS: análisis de la consistencia interna

Se puede consultar información relevante en la siguiente dirección:

<https://www.ibm.com/docs/es/spss-statistics/25.0.0?topic=features-reliability-analysis>

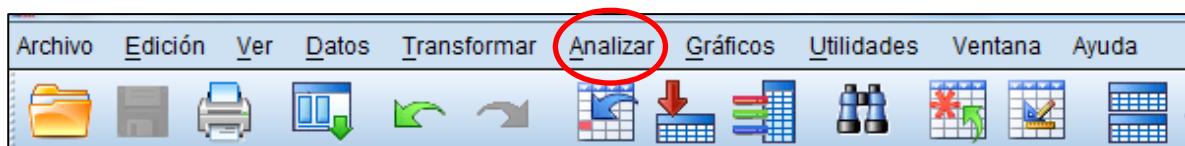
Ejercicio 1.

Realizar la estimación de la consistencia interna para la sub-escala de Prejuicio Manifiesto y la de Prejuicio Sutil utilizando el alfa de Cronbach.

Para computar la fiabilidad es necesario seguir la siguiente ruta con las ventanas del SPSS:

ANALIZAR----->**ESCALA**----->**ANÁLISIS DE FIABILIDAD**

En la siguiente imagen se puede observar cómo acceder a la ventana de “Analizar” y a partir de ahí seguir la ruta que se ha señalado en la imagen anterior.



Una vez se entra en la pantalla de “Análisis de fiabilidad”, se mueven a la derecha aquellos ítems de los que se desea analizar la fiabilidad o consistencia y que se supone forman una escala o dimensión que mide un determinado constructo.

Es muy importante que todos los ítems estén puntuados en el mismo sentido y nunca tener ítems en sentido directo e inverso en un mismo análisis de consistencia interna, pues daría valores de alfa de Cronbach bajos e incluso negativos. Si fuese necesario, para invertir los ítems hay que cambiar el formato de respuesta, por ejemplo, si la respuesta del ítem 1 es 1, 2, 3, 4, 5 y 6 hay que cambiar ese ítem invirtiendo la respuesta a 6, 5, 4, 3, 2 y 1. Es decir, cambiar el 1 por el 6, el 2 por el 5 y

así sucesivamente. Este proceso de inversión se puede realizar en la Ventana del SPSS de “Transforma y recodificar en la misma variable” (se sustituye a la variable original) o “Recodificar en distinta variable” (se crea una nueva variable con un nombre distinto que aparecerá al final de la base de datos). La opción de “Recodificar en distinta variable” es más recomendable para no machacar o eliminar la versión original. Es útil poner el mismo nombre del ítem que se invierte, pero con una R al final del nombre; por ejemplo, autoestima1R, autoestima5R para recordar que ese ítem se ha invertido.

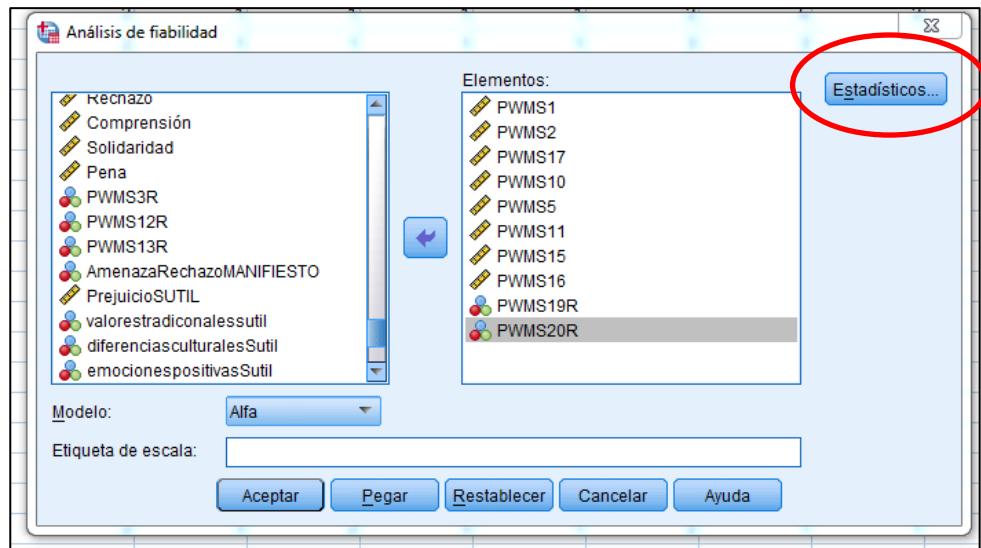
Se puede controlar que todos los ítems están en el mismo sentido observando en los resultados del SPSS la columna de las correlaciones del ítem con la puntuación total donde todas tienen que ser correlaciones positivas. Si hay algún valor de coeficiente de correlación negativo entonces ese ítem está en un sentido inverso al resto de ítems y habría que transformar su escala de respuesta antes de ejecutar el análisis de consistencia interna. Si la muestra es muy pequeña también podrían aparecer valores de alfa de Cronbach negativos.

Es muy importante tener en cuenta que nunca se deben sumar los ítems de una dimensión o escala que tiene ítems directos e ítems inversos, siempre deben estar medidos en el mismo sentido.

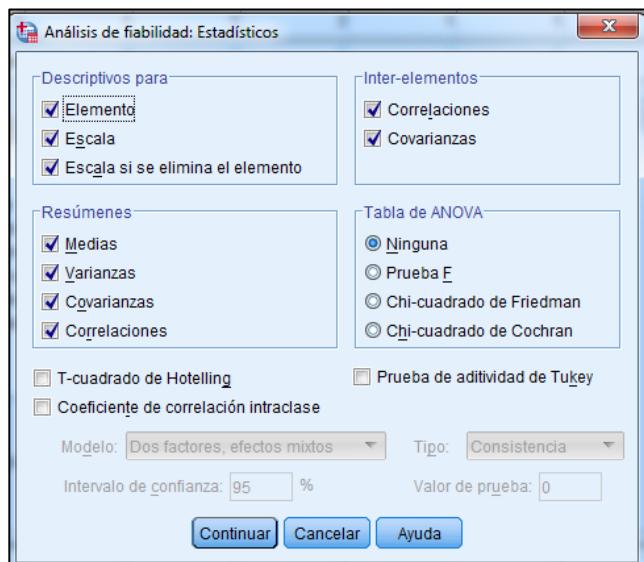
Una vez nos encontramos en la ventana de “Análisis de fiabilidad” existen diferentes modelos de fiabilidad y hay que seleccionar uno de ellos. Así, en el apartado de “Modelo” de fiabilidad se puede seleccionar un modelo concreto:

- *Alfa*: se trata del alfa de Cronbach y es un modelo de consistencia interna basado en la correlación promedio entre los elementos
- *Dos mitades*: este modelo actúa dividiendo la escala en dos partes y analizando la correlación entre dichas partes.
- *Guttman*: este modelo calcula los límites inferiores de Guttman para la fiabilidad verdadera.
- *Paralelo*: con este modelo se asume que todos los elementos que se van a analizar tienen varianzas iguales y varianzas de error también iguales a través de las réplicas.
- *Paralelo estricto*: este modelo asume los mismos supuestos del modelo paralelo, pero también asume que las medias son iguales en todos los elementos.

Antes de ejecutar el análisis de consistencia interna es necesario entrar en la opción de “Estadísticos” (ver la imagen siguiente) para que la salida de resultados del programa ofrezca el análisis de consistencia interna de los ítems.



En “Estadísticos” se seleccionan las opciones de “Descriptivos para” elemento, escala y escala si se elimina el elemento. Este apartado de descriptivos es esencial para analizar de forma apropiada la consistencia interna de alfa de Cronbach y el análisis de la calidad de los ítems que forman la escala (ver imagen siguiente). Y si se desea, también se puede ampliar la información con “Resúmenes” de medias, varianzas, covarianzas y correlaciones y en “Inter-elementos” se señalan las correlaciones y covarianzas. En la imagen siguiente faltaría seleccionar “Coeficiente de correlación intraclass” para estimar el intervalo de confianza del valor puntual del alfa de Cronbach, cuestión que se analizará posteriormente.



La **sintaxis** para obtener los resultados ejecutando los comandos en la ventana de Sintaxis del SPSS es la siguiente:

```

RELIABILITY
/VARIABLES=PWMS1 PWMS2 PWMS17 PWMS10 PWMS5 PWMS11 PWMS15 PWMS16 PWMS19R PWMS20R
/SCALE('ALL VARIABLES') ALL
/MODEL=ALPHA
/STATISTICS=DESCRIPTIVE SCALE CORR
/SUMMARY=TOTAL MEANS VARIANCE COV CORR.

```

Para dar solución al ejercicio propuesto: una vez se han movido los ítems a la derecha y situados en “Elementos”, seleccionado el modelo de Alfa y seleccionados en estadísticos los valores que se desean conocer, ya se puede clicar sobre “Aceptar”. A continuación, se analizan los resultados que aportan los programas estadísticos después de ejecutar el estudio de la fiabilidad de una escala.

Ventana de Resultados del SPSS

En la ventana de Resultados del SPSS se obtienen un conjunto de tablas relacionadas con toda la información solicitada en “Opciones”. Los resultados aparecen en la ventana de Resultados del SPSS (ver imagen siguiente). Se observa un valor de alfa de Cronbach = .74, siendo la puntuación media de los ítems de 3.14 (oscilando entre 1.69 a 3.71), todos los valores de correlación entre el elemento y la total corregida (eliminando el ítem, índice de homogeneidad corregido) superan el valor de .35 y no se mejora de forma sustancial el valor de alfa de Cronbach si se elimina algún ítem.

➔ Análisis de fiabilidad

[Conjunto_de_datos1] K:\mim2014\BaseMIMbruto[1]_1.sav

Escala: TODAS LAS VARIABLES

Resumen del procesamiento de los casos

	N	%
Casos Válidos	194	93,3
Excluidos ^a	14	6,7
Total	208	100,0

a. Eliminación por lista basada en todas las variables del procedimiento.

Estadísticos de fiabilidad

Alfa de Cronbach	Alfa de Cronbach basada en los elementos tipificados	N de elementos
,743	,748	10

Estadísticos de los elementos

	Media	Desviación típica	N
PWMS1	1,69	1,061	194
PWMS2	3,70	1,504	194
PWMS17	2,98	1,386	194
PWMS10	3,27	1,013	194
PWMS5	3,30	1,060	194
PWMS11	3,71	1,138	194
PWMS15	2,92	1,217	194
PWMS16	3,39	1,226	194
PWMS19R	3,52	1,565	194
PWMS20R	2,93	1,324	194

Matriz de correlaciones inter-elementos

	PWMS1	PWMS2	PWMS17	PWMS10	PWMS5	PWMS11	PWMS15	PWMS16	PWMS19R	PWMS20R
PWMS1	1,000	,393	,384	,222	,029	,175	,178	,112	,181	,022
PWMS2	,393	1,000	,485	,172	,242	,164	,106	,285	,317	,242
PWMS17	,384	,485	1,000	,165	,264	,201	,162	,278	,369	,333
PWMS10	,222	,172	,165	1,000	,377	,589	,412	,350	,082	,041
PWMS5	,029	,242	,264	,377	1,000	,331	,243	,332	,120	,052
PWMS11	,175	,164	,201	,589	,331	1,000	,467	,471	,069	-,078
PWMS15	,178	,106	,162	,412	,243	,467	1,000	,447	,103	-,097
PWMS16	,112	,285	,278	,350	,332	,471	,447	1,000	,133	-,080
PWMS19R	,181	,317	,369	,082	,120	,069	,103	,133	1,000	,460
PWMS20R	,022	,242	,333	,041	,052	-,078	-,097	-,080	,460	1,000

Matriz de covarianzas inter-elementos

	PWMS1	PWMS2	PWMS17	PWMS10	PWMS5	PWMS11	PWMS15	PWMS16	PWMS19R	PWMS20R
PWMS1	1,127	,627	,565	,239	,032	,211	,230	,146	,300	,031
PWMS2	,627	2,262	1,011	,262	,387	,281	,194	,526	,746	,482
PWMS17	,565	1,011	1,922	,232	,388	,317	,273	,472	,801	,610
PWMS10	,239	,262	,232	1,028	,405	,679	,508	,435	,131	,054
PWMS5	,032	,387	,388	,405	1,125	,399	,314	,431	,200	,072
PWMS11	,211	,281	,317	,679	,399	1,294	,646	,656	,124	-,118
PWMS15	,230	,194	,273	,508	,314	,646	1,481	,667	,196	-,155
PWMS16	,146	,526	,472	,435	,431	,656	,667	1,503	,256	-,129
PWMS19R	,300	,746	,801	,131	,200	,124	,196	,256	2,448	,952
PWMS20R	,031	,482	,610	,054	,072	-,118	-,155	-,129	,952	1,752

Estadísticos de resumen de los elementos

	Media	Mínimo	Máximo	Rango	Máximo/mínimo	Varianza	N de elementos
Medias de los elementos	3,142	1,691	3,711	2,021	2,195	,345	10
Varianzas de los elementos	1,594	1,026	2,448	1,422	2,385	,242	10
Covarianzas inter-elementos	,357	-,155	1,011	1,166	-,6,502	,072	10
Correlaciones inter-elementos	,229	-,097	,589	,685	-,6,102	,027	10

Estadísticos total-elemento

	Media de la escala si se elimina el elemento	Varianza de la escala si se elimina el elemento	Correlación elemento-total corregida	Correlación múltiple al cuadrado	Alfa de Cronbach si se elimina el elemento
PWMS1	29,73	42,220	,345	,287	,730
PWMS2	27,72	36,816	,495	,353	,707
PWMS17	28,43	36,848	,555	,402	,697
PWMS10	28,15	41,195	,453	,429	,718
PWMS5	28,11	41,728	,384	,247	,726
PWMS11	27,71	40,426	,442	,460	,717
PWMS15	28,49	40,883	,369	,324	,727
PWMS16	28,03	39,688	,448	,373	,716
PWMS19R	27,90	38,255	,383	,295	,728
PWMS20R	28,48	42,759	,208	,326	,752

Estadísticos de la escala

Media	Varianza	Desviación típica	N de elementos
31,42	48,110	6,936	10

Interpretación de los resultados que ofrece el SPSS del alfa de Cronbach

El procedimiento de análisis de la consistencia interna de alfa de Cronbach ofrece dos valores:

- 1) El valor de alfa de Cronbach.
- 2) La consistencia interna basada en valor de alfa de Cronbach que utiliza los elementos o ítems tipificados (estandarizados).

Cuando se utiliza un conjunto de ítems que tienen la misma unidad de medida (por ejemplo, escala Tipo *Likert*) la interpretación se realiza con la consistencia interna del valor del alfa de Cronbach que utiliza la correlación entre los ítems de tal manera que cuanto mayor sea la correlación entre los ítems más probable será que las puntuaciones de los ítems que forman el instrumento sean consistentes. La consistencia interna basada en los ítems estandarizados utiliza las covarianzas entre los ítems y se utiliza cuando los ítems que forman la escala usan diferentes unidades de medida (alfa de Cronbach basado en los elementos tipificados). Habitualmente suele trabajarse con el denominado valor de alfa de Cronbach.

Correlación elemento-total corregida: índice de homogeneidad corregido

El análisis de la relación que mantiene cada uno de los ítems que forman la escala y la puntuación total de la escala misma es muy importante para valorar la calidad de la medida. Se trata de comprobar si el ítem está midiendo lo mismo que la escala global o total, es decir, en qué grado el ítem contribuye a la homogeneidad o consistencia interna de la escala. Esta información se obtiene con el índice de homogeneidad corregido o índice de discriminación y ayuda a valorar la consistencia interna de los ítems que forman el instrumento de medida.

El índice de homogeneidad corregido se obtiene por medio del coeficiente de correlación entre la puntuación del ítem y la puntuación total de la escala (eliminado, por supuesto, ese ítem). Este índice se basa en la idea de que si un ítem es discriminativo entonces si se obtiene una puntuación alta en el ítem también ese sujeto tendrá una alta puntuación en la puntuación total de la escala que pretende medir el constructo y si puntúa bajo entonces el sujeto tenderá a puntuar bajo en la escala global.

Si el valor del índice de homogeneidad de un ítem es bajo entonces se puede inferir que ese ítem no mide lo mismo que el resto de ítems que forman la escala. Si su valor es negativo entonces hay que cuestionar cómo se ha llevado a cabo la medición en ese ítem, pues probablemente es un ítem inverso que no ha sido invertido antes de estimar la consistencia interna de la escala.

En resumen, para considerar que los ítems del instrumento son adecuados, los valores del índice de homogeneidad corregido de cada uno de sus ítems deben ser altos y positivos (con signo positivo) respecto a las puntuaciones de la escala (sumatorio del resto de ítems).

Por todo ello, conviene observar con detenimiento en los resultados aportados por el SPSS la información relacionada con el apartado de “Correlación elemento-total corregida” (correlación ítem-test). Esta correlación indica la correlación lineal entre el ítem y la puntuación total en la escala (sin considerar el ítem que se está evaluando), señalando la magnitud y la dirección de esta relación). Dicha correlación se conoce como el *índice de homogeneidad*. Si el ítem analizado mide lo mismo que el resto de ítems (es decir, que la prueba globalmente) el índice de homogeneidad será alto, de manera que los sujetos que puntúan alto en el ítem también tenderán a puntuar alto en la escala y los sujetos que puntúan bajo en el ítem tenderán a puntuar bajo en la escala. Si el ítem tiene un índice de homogeneidad bajo entonces se infiere que estará midiendo algo diferente a lo que la escala mide en su conjunto. Conviene tener en cuenta que el índice de discriminación u homogeneidad puede variar entre -1 y 1, pero un valor negativo no tendría sentido en este tipo de

análisis y, probablemente, estará indicando que el ítem está en sentido inverso respecto al resto de ítems que forman la puntuación total.

Los ítems cuyos coeficientes de correlación ítem-total (extrayendo el ítem analizado) arrojan valores menores a .2 deben ser desechados o reformulados ya que su nivel de homogeneidad es insuficiente. Se puede inferir que en ese caso los ítems de la escala no miden lo mismo que el resto de ítems y no tiene sentido combinarlos en una puntuación total. Una baja correlación entre el ítem y la puntuación total puede deberse a diversas causas, ya sea por una mala redacción del ítem o que el ítem no sirve para medir lo que se desea medir. Se recomienda utilizar los ítems que tienen al menos un valor de .3 como correlación del ítem con la puntuación total (índice de homogeneidad corregido). Y si superan los valores de .4 se considera que los niveles de homogeneidad son muy buenos (Ebel, 1965; Kline, 1999; Nunnally y Bernstein, 1994).

Si se aplica la fórmula del alfa de Cronbach a los datos del ejercicio se observa que su valor es de .75:

$$\alpha = \frac{N \times \bar{r}}{1 + (N - 1) \times \bar{r}}$$

$$\alpha = 10 \frac{10 \times 0.229}{1 + (10 - 1) \times 0.229} = 0.75$$

Intervalo de confianza del valor de la consistencia interna

Cuando se estima el valor del alfa de Cronbach o el omega de McDonald es muy conveniente añadir el intervalo de confianza de esa estimación ya que permite valorar la precisión de la estimación puntual observando su amplitud (a mayor amplitud menor precisión) y, a su vez, valorando si el límite inferior del intervalo de confianza tiene un valor que se considera adecuado.

A continuación, se detalla un ejemplo junto al desarrollo de la estimación del coeficiente alfa de Cronbach y su intervalo de confianza, así como el de omega de McDonald. En primer lugar, con el programa SPSS (solo calcula el alfa de Cronbach), en segundo lugar, con JAMOVI y en tercer lugar con JASP. Tanto JASP como JAMOVI calculan el intervalo de confianza de la estimación puntual del alfa de Cronbach.

Ejercicio 2

Supongamos que una investigadora ha realizado un estudio donde se ha utilizado la escala denominada “Escala de creencias sobre el ajuste infantil en familias con padres / madres del mismo sexo” (Frías-Navarro, 2009, 2012). El estudio consta de 634 observaciones. La escala consta de los siguientes 14 ítems:

1. En general, el desarrollo social de un niño es mejor cuando es educado por un padre y una madre heterosexuales y no por una pareja de padres gay o madres lesbianas.
2. En general, los niños criados con padres del mismo sexo tendrán más problemas que los que son criados por un padre y una madre heterosexuales.
3. Es más probable que un niño sufra aislamiento social si sus amigos saben que sus padres son gay o lesbianas.
4. Si los niños son criados por padres del mismo sexo tendrán más problemas de confusión con su propia identidad sexual que si son criados por un padre y una madre heterosexuales.
5. Si queremos defender los intereses del niño, sólo las parejas de heterosexuales deberían poder adoptar.
6. Un niño adoptado por una pareja de padres del mismo sexo será objeto de broma y rechazo por parte de sus compañeros.
7. Si un niño es adoptado por una pareja de gay o de lesbianas seguramente tendrá problemas psicológicos en el futuro.
8. Seguramente los compañeros rechazarán a un niño cuyo padre o madre es gay o lesbiana.
9. Un niño que es educado por una pareja de padres del mismo sexo sufrirá las burlas de sus compañeros.
10. Probablemente el niño criado por padres gay o madres lesbianas no será escogido como líder por sus compañeros de clase o amigos.
11. Cuando un niño manifiesta conductas homosexuales sería conveniente llevarlo al psicólogo.
12. Si los padres son gay o lesbianas será difícil que los hijos sean invitados a las fiestas de sus amigos.
13. Un niño educado por madres lesbianas será un niño afeminado.

14. Lo más común es que el niño oculte la orientación homosexual de sus padres a sus amigos por el temor al rechazo social.

Se ha demostrado que el instrumento tiene una estructural bifactorial: la dimensión de Oposición Individual (actitudes más personales, basadas en las propias opiniones, rechazo más agresivo o tradicional) y la dimensión de Oposición Normativa (actitudes más sutiles, basadas en argumentos externos las justifiquen, rechazo moderno adaptado a la realidad actual). Todos los ítems miden el constructo de forma directa y no es necesario invertir ningún ítem. La forma de corregir la escala para obtener las puntuaciones totales de cada dimensión es la siguiente:

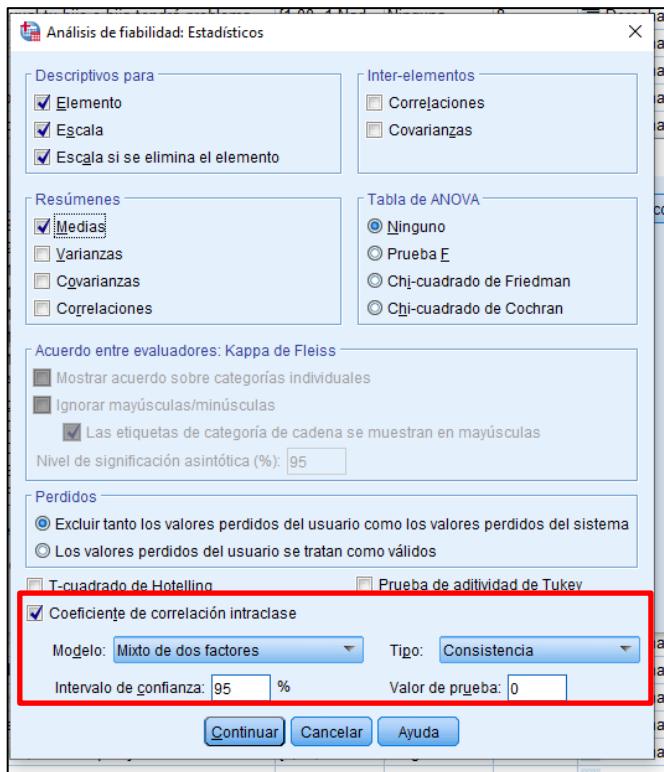
-*Oposición Individual*: ítems 1, 2, 4, 5, 7, 11, 13.

-*Oposición Normativa*: ítems 3, 6, 8, 9, 10, 12,14.

Antes de calcular la puntuación total de cada dimensión es necesario comprobar que la consistencia interna de los ítems que forman cada dimensión es adecuada y valorar su intervalo de confianza. Anteriormente se ha comprobado que su estructural factorial es de dos factores y los ítems que forman cada uno de ellos se ajustan a las indicaciones de la autora de la escala de Oposición Individual y Oposición Normativa.

SPSS

Se procede con el análisis de la sub-escala de Oposición Individual tal y como se ha detallado anteriormente, pero dentro de “Opciones” se activa también la opción de estimar el intervalo de confianza clicando sobre “Coeficiente de correlación intra-clase”. Se puede cambiar el nivel de confianza si se desea. Gracias a esta selección el programa SPSS ofrece información sobre el intervalo de confianza asociado a la estimación puntual del valor de alfa de Cronbach. El programa SPSS en su versión 26 no calcula el valor del coeficiente omega de McDonald.



Los resultados son los siguientes:

Estadísticas de fiabilidad		
Alfa de Cronbach	Alfa de Cronbach basada en elementos estandarizados	N de elementos
,842	,862	7

Coeficiente de correlación intraclass							
Correlación intraclass ^b	95% de intervalo de confianza		Prueba F con valor verdadero 0				
	Límite inferior	Límite superior	Valor	gl1	gl2	Sig	
Medidas únicas	,432 ^a	,398	,468	6,331	633	3798	,000
Medidas promedio	,842 ^c	,823	,860	6,331	633	3798	,000

Modelo de dos factores de efectos mixtos donde los efectos de personas son aleatorios y los efectos de medidas son fijos.

Por lo tanto, la redacción que se llevaría a cabo en el artículo podría ser, por ejemplo:

“El análisis de la consistencia interna de los ítems que forman la dimensión de Oposición Individual tiene un valor del alfa de Cronbach de .84 [95% IC .82, .86], siendo un valor adecuado”.

JASP

JAMOVI

Estimar el intervalo de confianza cuando se conoce alfa y el número de observaciones (N)

En ocasiones puede ocurrir que se dispone de la información del valor de alfa de Cronbach, pero no de su intervalo de confianza. Por ejemplo, realizando la lectura crítica o activa de un artículo se desea conocer el intervalo de confianza de los valores de alfa de Cronbach que sí se redactan en el texto, pero el autor o autora no redactó su intervalo. La aplicación que se encuentra en Internet y se ejecuta on-line denominada “Parametric confidence interval for Cronbach alpha” permite llevar a cabo un análisis para obtener el intervalo de confianza siempre que se disponga del valor de estimación puntual de alfa de Cronbach y del valor de N o número de observaciones con las que se estimó dicho valor de alfa. Por ello, se recomienda anotar el número de observaciones que se utilizó al estimar el valor de alfa de Cronbach en la redacción:

“El análisis de la consistencia interna de los ítems que forman la dimensión de Oposición Individual ($N = 634$) tiene un valor del alfa de Cronbach de .84 [95% IC .82, .86], siendo un valor adecuado”.

La aplicación se encuentra en la siguiente dirección y ha sido elaborado por <https://www.psyctc.org/psyctc/> donde se aporta más información sobre otras utilidades estadísticas:

<https://shiny.psyctc.org/apps/Cronbach1Feldt/>

Parametric confidence interval for Cronbach alpha

This shiny app is part of a number from my site at PSYCTC.org. There is a form there to contact me so do please use that if you think there is anything wrong here, or anything that could be improved.

Put your values in here, replacing the existing ones

Sample size (n), positive integer

Number of items (k), positive integer

Observed/reported Cronbach alpha value (<=1.0)

A referential alpha (if you need this)

Width of CI (usually .95, i.e. 95% CI, <=.99)

Number of decimal places

Utilizando los resultados del supuesto de investigación se rellena cada apartado y se debe tener en cuenta que se utiliza la coma para separar el valor decimal. La información que proporciona esta aplicación es la siguiente y como se observa se ha obtenido que el intervalo de confianza al 95% es de .82 a .86 tal y como se computó con los programas SPSS; JAMOVI y JASP.

Your input and results

Given:
Observed alpha = 0.84
n = 634
Number of items = 7
Gives two-tailed p = 0
against alternative alpha of 0
95% confidence interval from 0.82 to 0.86

App created by Chris Evans PSYCTC.org licenced under a Creative Commons, Attribution Licence-ShareAlike Please respect that and put an acknowledgement and link back to here if re-using anything from here.

Otras utilidades que se ofrecen en esa página Web: conocido el tamaño del efecto d de Cohen, obtener el valor g de Hedges (Getting Hedges's g from Cohen's d and n): https://shiny.psyctc.org/apps/g_from_d_and_n/; correlación media del ítem con la puntuación total: (Mean inter-item correlation for alpha and k): https://shiny.psyctc.org/apps/inter-item_mean_from_alpha/; intervalo para una correlación de Pearson: Confidence interval for a Pearson correlation (<https://shiny.psyctc.org/apps/CIcorrelation/>). Toda la información se detalla en <https://shiny.psyctc.org/>

Referencias (en proceso)

Análisis factorial de los ítems de una escala

Se recomienda la lectura y estudio de los siguientes trabajos:

Ferrando Piera, P. J., Lorenzo Seva, U., Hernández Dorado, A., & Muñiz Fernández, J. (2022). Decálogo para el análisis factorial de los ítems de un test. *Psicothema*, 34(1), 7-17. doi: 10.7334/psicothema2021.456

Frías-Navarro, D., & Pascual Soler, M. (2012). Prácticas del análisis factorial exploratorio (AFE) en la investigación sobre conducta del consumidor y marketing. *Suma Psicológica*, 19(1), 47-58.
