ID: 10.33881/2027-1786.RIP.14101

ÏbērAM
Reflexión/Discusión | Reflection

Propuesta de estimador de la fiabilidad mediante Alfa-Game

La significancia estadística del coeficiente de fiabilidad

Proposal for a reliability estimator using Alfa-Game

José Alejandro **González Campos**
Juan Elías **Aspeé Chacón**

Rip 14[1]
Volumen 14 #1 ene-abr
14 Años

2021

Revista Iberoamericana de
Psicología
ISSN-l: 2027-1786 | e-ISSN: 2500-6517
Publicación Cuatrimestral

Revista Iberoamericana de
Psicología
ISSN-l: 2027-1786 | e-ISSN: 2500-6517
Publicación Cuatrimestral

ID: 10.33881/2027-1786.RIP.14101

Title: Proposal for a reliability estimator using Alfa-Game

Título: Propuesta de estimador de la fiabilidad mediante Alfa-Game

Subtítulo: La significancia estadística del coeficiente de fiabilidad

Alt Title / Título alternativo:

[en]: Proposal for a reliability estimator using Alfa-Game

Author (s) / Autor (es):

González Campos & Aspeé Chacón

Keywords / Palabras Clave:

[en]: Cronbach's alfa; reliability; significance; inference; distributional model.

[es]: Alfa de Cronbach; fiabilidad; significancia; inferencia; modelo distribucional.

Proyecto / Project:
Postdoctorado 2019-2020, del Instituto Interuniversitario de Investigación Educativa (IESED-CHILE).

Submited: 2020-05-25

Acepted: 2020-09-18

Resumen

Generalmente cuando el coeficiente alfa de Cronbach es superior a 0.7 se indica por consenso que una prueba cualquiera es fiable, sin embargo, dicho umbral es un criterio subjetivo que carece de respaldo inferencial, y en algunas ocasiones, se utiliza desconociendo las limitaciones de este coeficiente. Así, en el presente se propone de un soporte compacto para el estimador de la fiabilidad, de manera de superar las restricciones del efecto de las covarianzas negativas. De esta manera, se expone el uso del Coeficiente alfa Game y la definición de la estadística p-valor en este contexto, lo que permite disponer de una herramienta objetiva en la decisión en significancia para el estimador de fiabilidad, reemplazando así el consenso en la comunidad científica. Con ello, alfa Game hace posible obtener afirmaciones en significancia y superar situaciones adversas para el estimador de fiabilidad alfa de Cronbach que tienden a confundir a las y los investigadores no expertos en estadística. Palabras clave: alfa de Cronbach; fiabilidad; significancia; inferencia; modelo distribucional.

Abstract

Generally, when Cronbach's alpha coefficient is greater than 0.7, it is indicated by consensus that any test is reliable, however, this threshold is a subjective criterion that lacks inferential support, and on some occasions, it is used without knowing the limitations of this coefficient. Thus, at present compact support for the reliability estimator is proposed, in order to overcome the restrictions of the effect of negative covariances. In this way, the use of the alpha Game Coefficient and the definition of the p-value statistic in this context are exposed, which allows having an objective tool in the decision on significance for the reliability estimator, thus replacing the consensus in the scientific community. With this, the alpha Game makes it possible to obtain statements of significance and overcome adverse situations for Cronbach's alpha reliability estimator that tend to confuse non-statistical researchers.Keywords: Cronbach's alfa; reliability; significance; inference; distributional model.

Citar como:

González Campos, J. A., & Aspeé Chacón, J. E. (2021). Propuesta de estimador de la fiabilidad mediante Alfa-Game: La significancia estadística del coeficiente de fiabilidad. Revista Iberoamericana de Psicología , 14 (1), 1-10. Obtenido de: https://reviberopsicologia.ibero.edu.co/article/view/1860

Dr José Alejandro **González Campos**, Msc Psi
ORCID: https://orcid.org/0000-0003-4610-6874

Source | Filiacion:
Universidad de Playa Ancha de Ciencias de la Educación. Valparaíso, Chile

BIO:
Chileno. Académico e investigador de la Universidad de Playa Ancha de Ciencias de la Educación. Coordinador General de Investigación Facultad de Ciencias Naturales y Exactas, y Coordinador laboratorio de investigación de saberes matemáticos.
Especialista en estadística aplicada a la educación.

City | Ciudad:
Valpariaiso [cl]

Dr Juan Elías **Aspeé Chacón**, Msc Psi
Research ID: AAN-9224-2020
ORCID: http://orcid.org/0000-0003-3456-8414

Source | Filiacion:
Agencia de Estudios Sociales. Viña del Mar, Chile. Universidad Técnica Federico Sanata María. Valparaíso, Chile.

BIO:
Chileno. Director e investigador de la Agencia de Estudios Sociales (Viña del Mar, Chile).
Profesional de la Dirección de Relaciones Estudiantiles de la Universidad Técnica Federico Santa María (Valparaíso, Chile).
Especialista en políticas y gestión en educación superior, y teoría y metodología del trabajo social.

City | Ciudad:
Valpariaiso [cl]

Propuesta de estimador de la fiabilidad mediante Alfa-Game

La significancia estadística del coeficiente de fiabilidad

Proposal for a reliability estimator using Alfa-Game

José Alejandro **González Campos**
Juan Elías **Aspeé Chacón**

La medición determinística asociada a atributos físicos es un proceso directo y la precisión es un concepto de relativa facilidad de consenso. Sin embargo, el proceso de medición de constructos metacognitivos, como lo son la mayor parte de los fenómenos sociales, es un proceso sumamente complejo, para los cuales no existen metros o balanzas **(González, Viveros & Carvajal, 2016)**. Para este proceso de abstracción aparecen como técnica-herramienta de medición el uso de pruebas o cuestionarios. No obstante, la misma utilización de tales instrumentos debe ser reflexionada, pues su calidad y precisión deben ser ponderadas, y siempre deben ser susceptibles a la mejora. Así, y para optimizar las características métricas de un instrumento, aparecen dos conceptos fundamentales en medición, a saber: la fiabilidad y la validez.

En el presente trabajo, se desarrolla un mecanismo de medición de la fiabilidad utilizando como base el Alfa de Cronbach **(1951)**, mecanismo que complementa dicha prueba, especialmente en escenarios que son desfavorables para el uso e interpretación de esta, por ejemplo, en muestras pequeñas o en resultados con expresiones negativas. Aquí la fiabilidad es un concepto teórico sinónimo de consistencia interna del instrumento. Es decir, si cada partición de la prueba, llámese formas paralelas, establece el mismo valor métrico de la unidad a medir. Otra manera de entender este concepto es como estabilidad temporal, eso significa que, si aplicamos el instrumento en un tiempo t1 y replicamos en un tiempo t2, las mediciones realizadas deberían ser consistentes **(Prieto & Delgado, 2010)**. En torno a estas ideas teóricas surgen variadas alternativas de cuantificación de la fiabilidad, todas estructuradas sobre la misma noción basal a describir.

ISSN-L: 2027-1786

En general el coeficiente de fiabilidad es definido como la correlación entre las puntuaciones obtenidas por las personas en estudio en dos formas paralelas de una prueba, digamos y , dicha correlación la representaremos por:

$$Px_1x_2 \qquad (1)$$

Es importante entender el concepto de correlación, pues representa el grado de asociación lineal entre dos variables en estudio **(Casella & Berger, 2002)**. En este caso, las variables están caracterizadas por las puntuaciones obtenidas en las dos formas paralelas generadas a partir de una prueba. Ahora independiente si se usa la técnica de formas paralelas o como estabilidad temporal, y si no hubiese errores aleatorios de medida, la correlación debería ser perfecta:

$$Px_1x_2 = 1 \qquad (2)$$

Por lo tanto, el grado en que Px_1x_2 se aleja de 1, indicará en qué grado las mediciones están afectadas por errores aleatorios, siempre bajo el supuesto que las formas x_1 y x_2y realmente sean paralelas.

Con ello en mente, y para enriquecer las características métricas del coeficiente de cuantificación de la fiabilidad, en el año 1951 el profesor Lee J. Cronbach presenta uno de los trabajos de mayor importancia en este ámbito. La trascendencia del coeficiente Alfa de Cronbach se debe a que es una generalización de las fórmulas de KR-20 y KR-21 de consistencia interna y su condición dicotómica propuestas por **Kuder y Richardson en 1937** quehasta hoy son motivo de estudio.

En este artículo se retoma la problemática del uso y la interpretación del coeficiente de cuantificación de la fiabilidad mediante el alfa de Cronbach **(Cervantes, 2005; Contreras & Novoa-Muñoz, 2018; Domínguez-Lara, 2012; Maroco & Garcia-Marques, 2013, Gempp & Saiz, 2013; González & Pazmiño, 2015; Vaske, Beaman & Sponarski, 2017)**. Se pretende así, aportar criterios objetivos para su interpretación, de manera de superar los umbrales subjetivos generales que han primado en el uso de este coeficiente, expresados en el criterio que establece aceptable un alfa superior a 0.7 **(Campo-Arias & Oviedo, 2008; Henson, 2001; Oviedo & Campo-Arias, 2005)**, lo que es base de un sinfín de conclusiones en ciencias sociales, tales como las de **Cunha, Xu, Rosário y Núñez (2018), Figueroa, Rodriguez, Díaz y Zapata (2018), López-González et al. (2018), González e Ibáñez (2018), o Vilchez (2017)** entre otras.

A este tenor, **Taber (2018)** mostró que una parte importante de las investigaciones en el año 2015 referían valores del alfa de Cronbach con escasos comentarios y explicaciones sobre los mismos, indicando con un umbral arbitrario su aceptabilidad, con una diversidad alta en terminologías según cada investigación. Para superar ello, esta propuesta establece como mecanismo de interpretación la aleatorización del coeficiente alfa de Cronbach, estructurando procedimientos inferenciales para su exégesis. Pero, para avanzar, es necesario primero dilucidar qué implica la expresión diferencias significativas, efectos o valores significativos, puesto que se han convertido en una expresión frecuente en el ámbito de las ciencias sociales, a tal punto que los efectos de su utilización o los contextos en los que se resalta han perdido importancia.

En este sentido, hablar de nivel de significancia, generalmente indicado con la letra alfa (α), es indicar la cota superior para el error de tipo I en teoría de decisión, o particularmente planteamientos de hipótesis *(Error tipo I $\leq\alpha$)*, al mismo tiempo de ser el complemento del nivel de confianza. El nivel de significancia es prefijado por el investigador (generalmente 5%, 3% y 10%). El error tipo I dice relación con rechazar la hipótesis nula cuando esta es verdadera, dejando en evidencia que el concepto de significatividad es una expresión inferencial sometida a supuestos y restricciones. Asimismo, se entenderá como proceso inferencial al acto de realizar inferencia, esto es tomar como evidencia lo soportado en los datos muestrales, para extenderlo a la población de estudio. Las implicancias de utilizar erradamente esta expresión nos pueden llevar a asumir procedimientos o a implementar metodologías o técnicas que pueden tener repercusiones totalmente opuestas a las esperadas, lo que definitivamente en el ámbito educacional, psicológico, económico, social, etcétera, puede traducirse en conclusiones inexactas que tienen efectos en personas concretas. Una política educativa incorrecta, un instrumento psicométrico diseñado erradamente, una política económica basada en una evidencia trastocada, y así sucesivamente.

La pregunta que se debe plantear es ¿en qué contexto o situación podemos cuestionarnos la significancia de los efectos? Por ejemplo, en la Figura 1 se presenta a dos niños de diferentes alturas. En este contexto cabe preguntarse si José es significativamente más alto que Rafaela, y si es afirmativa la respuesta ¿cuál es el nivel de significación? Así, no es necesario usar instrumentos muy sofisticados para darse cuenta de que José es más alto que Rafaela. Es más, es posible afirmar que José es 30 centímetros más alto que Rafaela. Pero ¿es posible afirmar que José es significativamente más alto que Rafaela?, la respuesta es un rotundo no. Esto se debe a que hablar de diferencias significativas, implica conocer el nivel de significancia determinado, es decir, el umbral de decisión respecto de comparaciones **(Casella & Berger, 2002)**. Por tanto, es un concepto que adquiere su valor, siempre y cuando se cumplan los supuestos pertinentes entre los que se incluye el disponer de una muestra aleatoria y un tamaño muestral que permita usar alguna ley de los grandes números. Es decir, una referencia de contexto para comparar las alturas de José y de Rafaela.

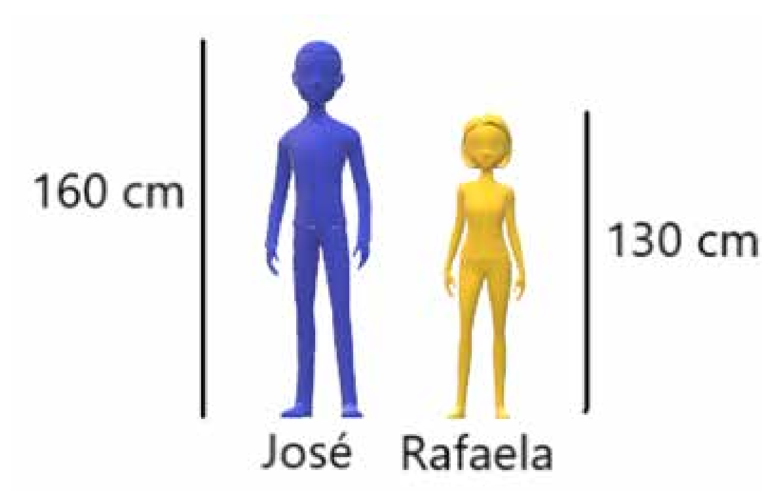

Figura 1. Alturas en centímetros entre José y Rafaela, sin contexto de comparación.

Ahora, en la Figura 2 se presenta un grupo de alumnos, entre ellos José y Rafaela, para los cuales es posible observar que Rafaela tiene una altura similar a la mayoría. Entonces en este contexto ¿Rafaela es significativamente de menor altura que José? Solo con el contexto de comparación es posible responder a dicha interrogante. Así, serán los datos los que determinarán si la diferencia de las alturas es significativa o no, y no la simple interpretación del observador **(Merino, Domínguez-Lara & Contreras, 2018; Ventura-León, 2017)**.

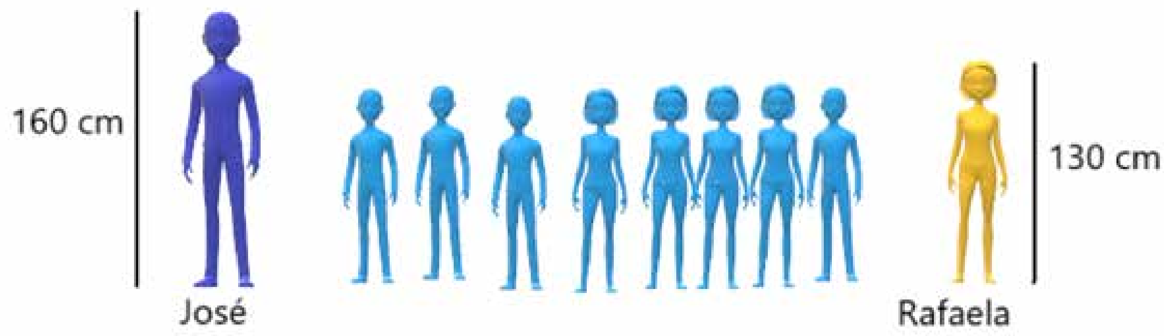

Figura 2. Alturas en centímetros entre José y Rafaela, con contexto de comparación.

El ejemplo, simple, por cierto, permite entender el concepto de diferencias significativas, pues es la muestra la que nos debe proveer evidencia a favor o en contra de la hipótesis que nos hemos planteado, en función de la superación o no del umbral de corte dado por el nivel de significancia. Es importante esta discusión, pues cuando estimamos la fiabilidad de una prueba basada en el coeficiente alfa de Cronbach no es suficiente para hacer afirmaciones respecto de la significatividad, pues es necesario para ello definir un proceso de muestreo, incorporar la aleatoriedad, proponer algún modelo distribucional coherente a la naturaleza de los valores que este genera, y finalmente, basado en una construcción teórica de un método inferencial, tomar una decisión. Solo en este momento se pueden hacer afirmaciones en significancia **(Casella & Berger, 2002)**.

Antecedentes y limitaciones del alfa de Cronbach α_C

La forma general de escribir el coeficiente fue presentada por el mismo Cronbach en 1951:}

$$\alpha_C = n/(n-1)\ (1-(\Sigma_i V_i)/V_t) \qquad (3)$$

Donde representa la varianza de las puntuaciones totales de la prueba en cada ítem V_t y es la varianza de las puntuaciones de cada ítem. Asimismo, la interpretación del coeficiente α_C se desprende de su formulación, esto es: La varianza de las puntuaciones totales a cada ítem, es decir V_t, puede ser expresada como:

$$V_t = \Sigma_i V_i + \Sigma_{i\neq j} Cov(I_i,I_j) \qquad (4)$$

Donde I_i y I_j representan las puntuaciones en pares de ítems diferentes y $Cov\ (I_i,I_j)$las covarianzas entre estas puntuaciones. Esta simple descomposición, nos permite reformular el coeficiente α_C, del siguiente modo:

$$\alpha_C = n/(n-1)\ (1-(\Sigma_i V_i)/(\Sigma_i V_i+\Sigma_{i\neq j} Cov(I_i,I_j))) \qquad (5)$$

Con esta reformulación se observa que la componente aditiva $\Sigma_{i\neq j} Cov(I_i,I_j)$ es la que permite alcanzar la interpretación adecuada. Por su parte, $Cov(I_i,I_j)$ viene a cuantificar la covarianza, es decir, la variabilidad conjunta o producto de la interacción. Por ello, informalmente podríamos decir que es la complicidad entre las puntuaciones de los ítems.

Cuando las puntuaciones de los ítems son independientes, implica que $Cov(I_i,I_j) = 0$ y si esta independencia es generalizada para todos los pares de ítems diferentes, entonces se tiene que $\Sigma_{i\neq j}\ Cov(I_i,I_j) = 0$. Este supuesto trae sobre el coeficiente α_c la siguiente consecuencia:

$$\alpha_C=n/(n-1)\ (1-(\Sigma_i V_i)/(\Sigma_i V_i+\Sigma_{i\neq j} Cov_{Ii,Ij}))=n/(n-1)\ (1-(\Sigma_i V_i)/(\Sigma_i V_i))=n/(n-1)\ (1-1)=0 \qquad (6)$$

Esto significa que la independencia entre las puntuaciones de los ítems no es favorable a la cuantificación de la fiabilidad de una prueba basada en el coeficiente . Por el contrario, si las covarianzas son altas y positivas, entonces el coeficiente α_c converge a 1, debido a $(\Sigma_i V_i)/(\Sigma_i V_i+\Sigma_{i\neq j} Cov(I_i,I_j))$ que converge a cero. En consecuencia, teóricamente lo que se busca es la consistencia entre los ítems, esto es que cada ítem sea una transformación de otro ítem. Es decir, si tenemos los ítems I_i y I_j, de una prueba con un alto coeficiente α_c, entonces I_i es una transformación de I_j. En términos de una asociación lineal, se dice que $I_i \approx a+bI_j$, para dos constantes reales .a y b

No obstante, cabe preguntarse por el comportamiento del coeficiente cuando las covarianzas son negativas y de valores absolutos altos. Este supuesto puede implicar que:

$$\Sigma_i V_i+\Sigma_{i\neq j} Cov(I_i,I_j)<\Sigma_i V_i \qquad (7)$$

Luego:

$$(\Sigma_i V_i)/(\Sigma_i V_i+\Sigma_{i\neq j} Cov(I_i,I_j))>1 \qquad (8)$$

Por tanto:

$$\alpha_c=n/(n-1)\ (1-(\Sigma_i V_i)/(\Sigma_i V_i+\Sigma_{i\neq j} Cov(I_i,I_j)))<0 \qquad (9)$$

La situación descrita en la tabla 1 corresponde a una simulación de 8 ítems aplicados a 30 personas (casos), donde sus respuestas correlacionan fuertemente con covarianzas negativas. Para esta situación se obtiene un coeficiente $\alpha_c=-2.7623$. Esto da origen a algunas inquietudes como: ¿Qué significado tiene esto? ¿Cómo se interpreta este tipo de resultados? ¿Tienen sentido estos valores? Este tipo de dificultades son las que complican las formulaciones estadísticas inferenciales, pues no se tiene control sobre el coeficiente , al posibilitar secciones de valores que carecen de interpretación. Aunque es válido precisar que este coeficiente no se creó como indicador inferencial, solo descriptivo como cuantificación de la consistencia interna de un instrumento de medición o test.

RIP 14 (1) pág. 1 -10

ISSN-L: 2027-1786

Tabla 1. Simulación de un instrumento de medición con covarianzas fuertes y negativas

Personas	Preguntas							
	1	2	3	4	5	6	7	8
1	4	1	1	5	5	2	2	5
2	2	3	4	1	0	5	5	0
3	4	2	3	2	1	2	3	3
4	1	5	5	1	0	5	5	2
5	2	3	3	2	1	4	4	1
6	1	4	4	1	0	3	4	1
7	2	4	4	1	1	3	4	2
8	1	4	4	2	1	4	5	1
9	1	4	5	1	0	4	4	3
10	4	2	2	3	3	3	3	4
11	3	2	3	2	2	4	5	1
12	2	3	3	2	2	2	2	3
13	2	3	4	2	2	3	4	3
14	1	5	5	1	0	5	5	1
15	3	2	2	3	3	4	5	1
16	3	3	4	1	1	3	4	3
17	1	4	5	0	1	3	3	4
18	5	1	1	5	5	1	1	5
19	1	5	6	0	1	4	4	2
20	4	2	3	2	2	4	5	0
21	1	5	6	1	1	4	4	2
22	3	3	4	1	1	2	2	4
23	4	2	3	2	2	3	3	4
24	3	3	4	1	0	4	4	3
25	1	5	6	1	1	5	5	2
26	2	4	4	1	1	4	4	3
27	3	2	3	3	3	4	4	2
28	2	3	3	2	1	3	4	3
29	2	4	4	1	0	4	4	1
30	5	1	1	4	4	3	4	3

Fuente: Elaboración propia.

A mayor abundamiento, los gráficos incorporados en la Figura 3 presentan la superficie generada por el coeficiente alfa de Cronbach, simulando diferentes varianzas totales y suma de varianzas de los ítems. En ella, se observa que cuando la varianza total es menor que la suma de las varianzas individuales de cada ítem, la superficie alcanza valores negativos. Para la generación del gráfico de la Figura 3, se consideró la fórmula de estimación de la fiabilidad, identificada con el alfa de Cronbach para un *n* (100) fijo, a partir de ello se definió funcionalmente, esto es:

$$f(x,y)=k(1-x/(x+y)),\ \text{donde}\ n/(n-1)=k,\ \Sigma_i V_i=x>0,\ \Sigma_{i\neq j} Cov(I_i,I_j)=y\in R \quad (10)$$

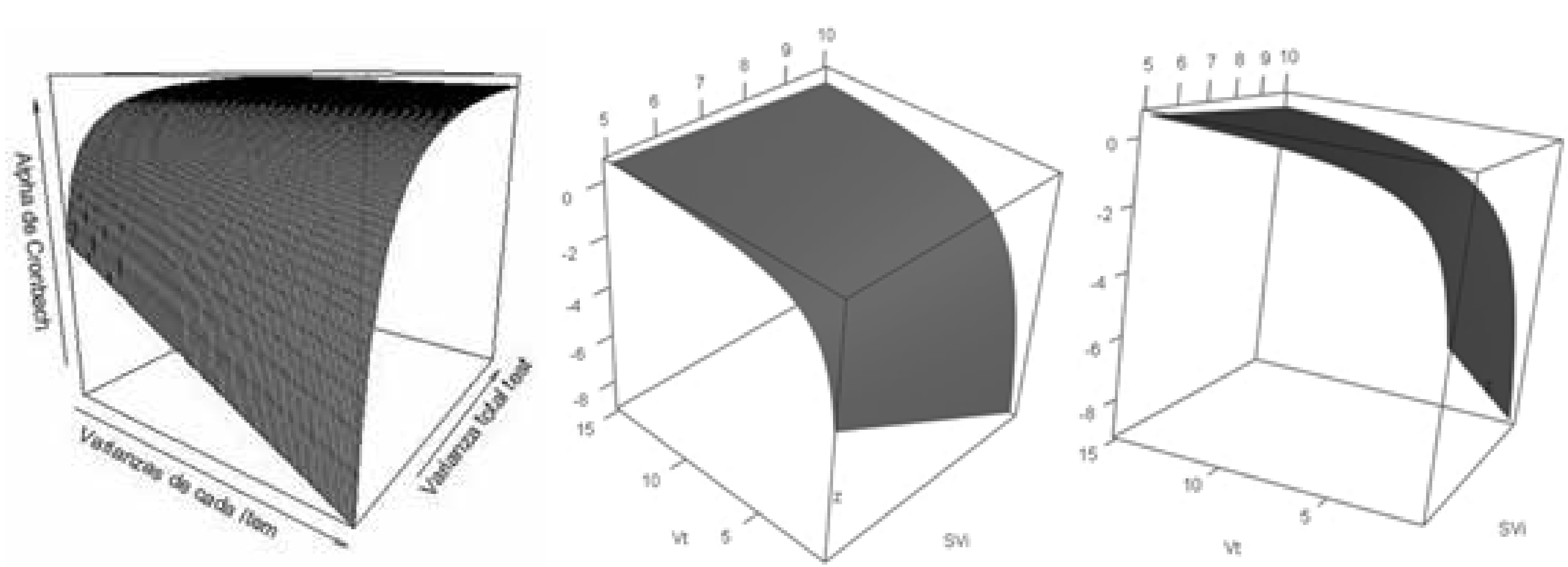

Figura 3. Superficie simulada para el alfa de Cronbach con valores negativos.

Los ejemplos simulados descritos precedentemente en tabla y gráficos expresan algunas de las falencias del alfa de Cronbach de las que pretende hacerse cargo el presente trabajo teórico-práctico. En este sentido, independiente de la naturaleza de la organización numérica que se suceda al aplicar una prueba, la formulación debe ser consistente y presentar un recorrido conocido. Solo de esa forma se permitirá la caracterización de un procedimiento inferencial para tomar decisiones sobre el coeficiente que puedan ser extendidas a la población y que den el soporte teórico estadístico a las conclusiones. Así las cosas, se presenta a continuación el desarrollo metodológico del denominado alfa-Game, que, con base al alfa de Cronbach, genera un procedimiento para tratar los escenarios adversos al mismo, permitiendo el uso de niveles de significancia. Dicho procedimiento se ejemplifica en un caso real, para luego terminar con discusiones y conclusiones pertinentes al caso.

Propuesta: Alfa-Game

Para la definición del coeficiente alfa-Game, se toma como base la definición del coeficiente alfa de Cronbach (α_c), previamente explicitada, cuya definición deja en evidencia la razón entre las covarianzas de cada par de ítem y las sumas de las varianzas, tal como se muestra en el desarrollo siguiente:

$$\alpha_c = n/(n-1)\ (1-(\Sigma_i V_i)/(\Sigma_i V_i+\Sigma_{i\neq j})\ Cov(I_i,I_j)))$$

$$= n/(n-1)\ (1-1/((\Sigma_i V_i+\Sigma_{(i\neq j)}\ Cov(I_i,I_j))/(\Sigma_i V_i)))$$

$$= n/(n-1)\ (1-1/((\Sigma_i V_i)/(\Sigma_i V_i)+(\Sigma_{(i\neq j)}\ Cov(I_i,I_j))/(\Sigma_i V_i)))$$

$$= n/(n-1)\ (1-1/(1+(\Sigma_{(i\neq j)}\ Cov(I_i,I_j))/(\Sigma_i V_i))) \qquad (11)$$

Por tanto, la limitación del coeficiente α_c (la posibilidad de obtención de covarianza negativas), puede ser resulta de forma simple, sustituyendo $\Sigma_{i\neq j}\ Cov(I_i,I_j)$ por $\Sigma_{i\neq j}\ |Cov(I_i,I_j)|$, permitiéndonos caracterizar el coeficiente α_c (alfa-Game) como una reformulación del coeficiente alfa de Cronbach, definido por:

$$\alpha_{\text{ə}} = n/(n-1)\ (1-1/(1+(\Sigma_{i\neq j}\ |Cov(I_i,I_j)|)/(\Sigma_i V_i))) \qquad (12)$$

Esta reformulación asegura un recorrido compacto dado por:

$$0 \leq \alpha_{\text{ə}} \leq 1 \qquad (13)$$

Además, implica que el coeficiente $\alpha_{\text{ə}}$ caracteriza una cota superior para el coeficiente α_c, esto es:

$$\alpha_c \leq \alpha_{\text{ə}} \qquad (14)$$

La igualdad es alcanzada cuando para todo par de ítems de la prueba se tiene que $Cov(I_i,I_j) \geq 0$.

El alfa-Game, dado su soporte compacto ($0 \leq \alpha_{\text{ə}} \leq 1$), puede ser definido como una variable aleatoria con distribución beta. Así, la expresión general para la función de densidad de probabilidad de una distribución beta está dada por **(González, Galvis & Hurtado, 2014)**:

$$f_x(x) = ((x-a)^{p-1}\ (b-x)^{q-1})/(B(p,q)\ (b-a)^{p+q+1}\ ),\ a \leq x \leq b; p,q > 0 \qquad (15)$$

Donde p y q son parámetros de forma, a y b son los límites superior e inferior respectivamente de la distribución, y B(p,q) es la función beta. La función beta se define por:

$$B(p,q) = \int_0^1 t^{\alpha-1}\ (1-t)^{\beta-1}\ dt \qquad (16)$$

El caso donde $a = 0$ y $b = 1$ se denomina distribución beta estándar, donde la fórmula general, según **Casella y Berger (2002)**, puede ser reducida a:

$$f_x(x) = (x^{p-1}\ (1-x)^{q-1})/B(p,q),\ 0 \leq x \leq 1\ ;\ p,q > 0 \qquad (17)$$

La función de distribución de probabilidades de una densidad beta está dada por:

$$F_x(x)=(\int_0^x t^{p-1}\ (1-t)^{q-1})/B(p,q)\ ,0 \leq x \leq 1;\ p,q > 0 \qquad (18)$$

Para este tipo de variables aleatorias se tienen las siguientes estadísticas (Tabla 2).

Tabla 2. Estadísticos para variables aleatorias beta

Media	Moda	Rango	Desviación estándar	Coeficiente de Variación	Asimetría
p/(p+q)	(p-1)/(p+q-2)	0 a 1	√(pq/((p+q)² (p+q+1)))	√(q/p(p+q+1))	(2(q-p) √(p+q+1))/((p+q+2) √pq)

Fuente: elaboración propia.

Para la estimación de los parámetros, bajo el método de momentos, se tiene que:

$$p = \underline{x}\ ((\underline{x}\ (1-\underline{x}))/s^2 -1) \qquad (19)$$

$$q = (1-\underline{x})((\underline{x}\ (1-\underline{x}))/s^2 -1) \qquad (20)$$

Donde $\underline{x}$ es la media muestral y s^2 la varianza muestral.

De esta manera, para poder introducir el concepto de significatividad en el coeficiente $\alpha_{\text{ə}}$, es necesario formalizar un proceso inferencial en torno a su definición. Por tanto, se dirá que $\alpha_{\text{ə}1},\alpha_{\text{ə}2},\ldots,\alpha_{\text{ə}n}$ es una muestra aleatoria de tamaño n de coeficientes $\alpha_{\text{ə}}$, desde una prueba de m ítems, para ello se utilizará un proceso de re-muestreo (bootstrapping) sobre los ítems de la prueba **(Ledesma, 2004)**. El procedimiento de re-muestreo consiste en generar $(m\ m-1\ )$ subprueba y para cada uno ellos seleccionar aleatoriamente n. Así, para cada subprueba es posible estimar el coeficiente $\alpha_{\text{ə}}$, que se han identificados como $\alpha_{\text{ə}1},\alpha_{\text{ə}2},\ldots,\alpha_{\text{ə}n}$.

En tal sentido, se impone un supuesto sumamente restrictivo como primera aproximación inferencial, indicando que cada $\alpha_{\text{ə}1},\alpha_{\text{ə}2},\ldots,\alpha_{\text{ə}n}$ es independiente e idénticamente distribuido y además que n es estadísticamente grande. Este supuesto permite utilizar el teorema central del límite **(Casella & Berger, 2002)**, que afirma que, ante una muestra aleatoria independiente e idénticamente distribuida como media conocida y varianza finita y positiva, se tiene que:

$$(\sqrt{n}\ (\ \underline{\alpha}_{\text{ə}} - E(\ \alpha_{\text{ə}}\ )))\ /\ Var(\ \alpha_{\text{ə}}\ ) \sim N(0.1) \qquad (21)$$

Donde $\underline{\alpha}_{\text{ə}}$ es la media de la muestra aleatoria de $\alpha_{\text{ə}1}$. Por otro lado, se indica que el valor esperado de cada variable aleatoria $\alpha_{\text{ə}}$ será μ, esto es $E(\ \alpha_{\text{ə}}\ ) = \mu$ y la varianza de cada variable aleatoria constituyente

RIP 14 (1) pág. 1-10
Revista Iberoamericana de Psicología
ISSN-L: 2027-1786

de la muestra aleatoria será σ^2, esto es $Var(\alpha_{ə})=\sigma^2$. Por ello, si se desconoce la varianza poblacional σ^2, un resultado equivalente es:

$$(\sqrt{n}\,(\alpha_{ə} - \mu))/s \sim t_{n-1;1-\alpha} \qquad (22)$$

Donde $\mu = E(\alpha_{ə})$, s^2 la varianza muestral y $t_{n-1;1-\alpha}$ es el cuantil $1 - \alpha$ de la distribución t - student con n - 1 grados de libertad.

En el primer pivote, se asume conocida la varianza mientras que, en el segundo pivote, es estimada por medio de la varianza muestral. Así, y en función de los pivotes anteriores, son presentados dos intervalos de confianza para μ.

Intervalo de confianza para μ, asumiendo conocida la varianza de $(1 - \alpha)\ 100\%$ de confianza, es:

$$IC_{1-\alpha}(\mu) = [\,\underline{\alpha}_{ə} - z_{1-\alpha/2}\ \sigma/\sqrt{n};\ \underline{\alpha}_{ə} + z_{1-\alpha/2}\ \sigma/\sqrt{n}\,] \qquad (23)$$

El intervalo de confianza para μ asumiendo desconocida la varianza, de $(1 - \alpha)\ 100\%$ de confianza, es:

$$IC_{1-\alpha}(\mu) = [\,\underline{\alpha}_{ə} - t_{1-\alpha/2;n-1}\ s/\sqrt{n};\ \alpha_{ə} + t_{1-\alpha/2;n-1}\ s/\sqrt{n}\,] \qquad (24)$$

Disponer de intervalos de confianza, es contar con herramientas de teoría de decisión en inferencia estadística y, por tanto, posibilita realizar pruebas de significancia sobre los parámetros de la variable aleatoria . Estas definiciones permiten generar un medio de decisión objetivo para el coeficiente . Cabe indicar que las restricciones inferenciales aquí también son necesarias, como, por ejemplo, el número de ítems que constituyen la prueba debe obedecer a las restricciones propias del teorema central del límite **(Casella & Berger, 2002)**.

La Figura 4 ejemplifica la densidad esperada, asumiendo un modelo beta, para los coeficientes

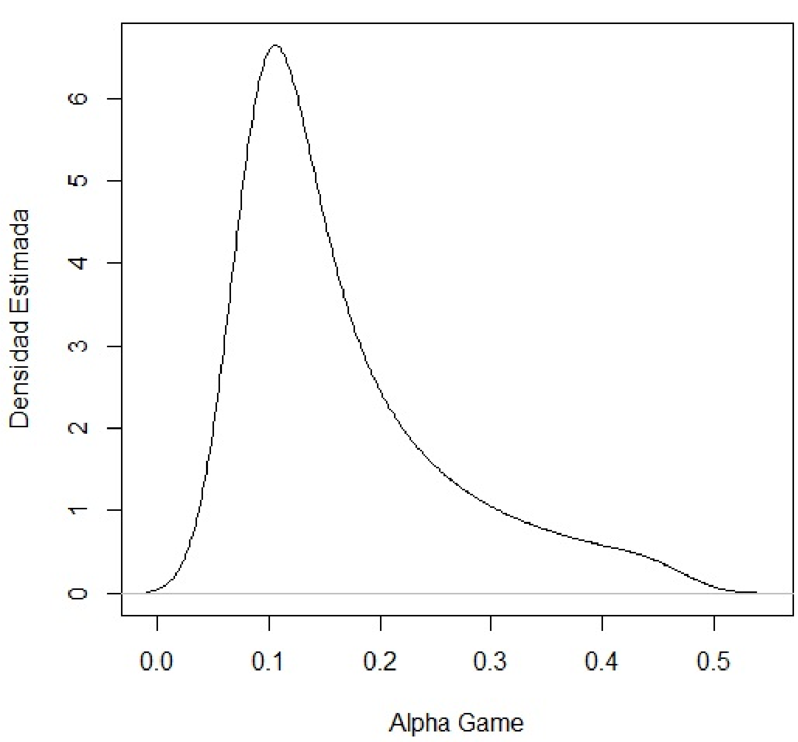

Figura 4. Simulación de la densidad de Alfa-Game

Para el caso simulado de la Figura 4, se observa que el valor medio para el coeficiente $\alpha_{ə}$ es bajo, por tanto, estaría caracterizando un instrumento con una baja fiabilidad. Sin embargo, para decidir si el valor obtenido para el coeficiente $\alpha_{ə}$ es alto, podría decidirse equivalentemente basado en intervalos de confianza sobre $\alpha_{ə}$, esto es si el intervalo de confianza contiene o no el valor 1. Otra forma, sería con base en la prueba de hipótesis, particularmente:

$$[H_0: \alpha_{ə} = 1\ \ H_1: \alpha_{ə} < 1\,] \qquad (25)$$

En esta situación se plantea el siguiente procedimiento:

- Escoja α (nivel de significancia). Límite máximo de probabilidad de cometer error tipo I.
- Seleccione la prueba a estudiar de m ítems y genere n subpruebas aleatorias de $m-1$ ítems y a partir de ellos genere una muestra aleatoria simple de $m-1$ coeficientes $\alpha_{ə}$.
- Determine el p - valor o fuerza de evidencia contenida en los datos.
- Si p-valor $\leq \alpha$, rechazar H_0

El p-valor (p-value) es dado por la siguiente expresión:

$$p\text{ - }value = P((T \leq (\underline{\alpha}_{ə} - \alpha_{ə0})/(s/\sqrt{n})) \qquad (26)$$

Donde la variable aleatoria T se distribuye según el modelo t - Student, lo que formalmente se escribe como:

$$T \sim t\text{-}student_{n-1;1-\alpha} \qquad (27)$$

Mientras que $P(k)$representa la probabilidad del evento k. Con base en el teorema central del límite, se puede obtener la siguiente aproximación:

$$p\text{ - }value = \varnothing((\underline{\alpha}_{ə} - \alpha_{ə0})/(s/\sqrt{n})) \qquad (28)$$

Donde $\varnothing(k) = P(Z \leq k)$ y Z es la variable aleatoria normal estándar, que formalmente se escribe $Z \sim N(0.1)$.

Para las hipótesis de interés planteadas, basta con considerar $\alpha_{ə0} = 1$, esto es:

$$p\text{ - }value = \varnothing(\alpha_{ə} - 1)/(s/\sqrt{n})) \qquad (29)$$

Definir la estadística p-valor, no es otra cosa que definir el criterio de decisión para las hipótesis, esto es, condicionado a un nivel de significancia (α), si $p\text{ - }valor < \alpha$ entonces la hipótesis nula es rechazada. Sin embargo, en el proceso de estimación de la fiabilidad de un test puede presentarse situaciones de homogeneidad en las varianzas, volviendo poco operativo el procedimiento al generar intervalos de confianza para el coeficiente $\alpha_{ə}$ poco conservativos. Si esto ocurre, se sugiere flexibilizar la metodología y se propone ponderar o corregir la varianza **(Correa & Sierra, 2003)**, del siguiente modo:

$$\sigma_*^2 = ((n+1)^4/4)\ \sigma^2 \qquad (30)$$

Donde σ_*^2 representa la varianza ponderada, equivalentemente para la varianza ponderada muestral:

$$s_*^2 = ((n+1)^4/4)\ s^2 \qquad (31)$$

Esta redefinición de la varianza genera intervalos de confianza mucho más conservativos y test más robustos.

Aplicación real

Para aplicar el alfa-Game propuesto, se recurrió a una muestra aleatoria de 50 estudiantes de postgrado de una universidad chilena, quienes fueron sometidos al Oxford Online Placement Test, que es una prueba para determinar el nivel de inglés. Esta prueba contiene 15 preguntas para medir la habilidad auditiva; y 15 preguntas para medir el uso de vocabulario, la gramática y la comprensión del significado en una conversación.

Aplicado el instrumento indicado, se determinó en primer lugar el coeficiente Alfa de Cronbach, que arrojó un valor de 0.617. Mientras aplicado el alfa–Game se obtuvo un valor muy similar de 0.67. Esta diferencia se debe a la existencia de 6 ítems cuya covarianza fue negativa. Por su parte, al tomar 7 submuestras o generar 7 subtest a partir del original, se obtiene el siguiente vector de coeficientes alfa-Game: 0.658, 0.672, 0.645, 0.657, 0.673, 0.666 y 0.662.

Considerando ello, se aplicó, previo análisis de normalidad (Shapiro-Wilk 0.943, p-valor 0.667; Kolmogorov-Smirnov 0.166, p-valor 0.973), la prueba paramétrica t-student cuyos resultados se detallan la Tabla 3. Para la obtención de estos resultados se aplicaron los softwares Jamovi 1.2.27 y R 3.5.1. Con ellos se indica que, la hipótesis alternativa que afirma que la media poblacional para el coeficiente alfa-Game es significativamente menor que 1 (a un nivel de significancia del), no es rechazada (p-valor <0.01), equivalentemente se rechaza la hipótesis nula. Es decir, el coeficiente alfa-Game es significativamente menor que 1. Por ende, la prueba Oxford Online Placement Test, no es fiable.

Tabla 3. Prueba t-student para Alfa-Game a un nivel de significancia del 5%

Estadística	df	p	Intervalo de confianza 95%	
			Lower	Upper
180	6.00	<.001	0.653	0.671

Nota: H_a media poblacional ≠ 1.

Fuente: elaboración propia.

El disponer del procedimiento descrito, permite realizar otro tipo de pruebas. Si bien se ha descartado estadísticamente que el coeficiente alfa-Game poblacional sea igual a uno en la situación anterior, cabe preguntarse, por ejemplo, ¿es significativamente menor que 0.7?, ello, considerando 0.7 un valor aceptable consuetudinariamente para la estimación de la fiabilidad. Así, y a partir de la Tabla 4, se concluye que los datos soportan evidencia a favor de la hipótesis alternativa, y por tanto, la fiabilidad del instrumento en cuestión es significativamente menor que 0.7. En consecuencia, se puede concluir que la prueba no es considerado fiable.

Tabla 4. Prueba t-student para Alfa-Game a un nivel de significancia del 5%

Estadística	df	p	Intervalo de Confianza 95%	
			Lower	Upper
-10.4	6.00	<.001	0.653	0.671

Nota. Considerando ecuación 25 para una región particular de H_a

H_a media poblacional < 0.7

Fuente: elaboración propia

Discusión y Conclusiones

Crear herramientas para que las mediciones sean más eficaces está en la esencia misma de las ciencias sociales, y de las ciencias en general. Pero, nunca se debe olvidar que la medición es una expresión de imperfección onto-epistemológica **(Covarrubias & Cruz, 2016)**. Por ende, una cualidad medida no sería más que una cualidad estimada, donde el resultado únicamente estará completo cuando se acompaña de la correspondiente incertidumbre **(García, Ruiz & Mesa, 2011)**.

En tal sentido, la utilización de una prueba objetiva, como lo es el alfa Cronbach para determinar la fiabilidad/confiabilidad de una prueba que mide conceptos abstractos, ayuda a dar mayor precisión a un acto impreciso por definición. Esto significó un gran aporte a la medición en ciencias sociales. No obstante, el uso acostumbrado de este coeficiente, así como de su determinación de efectividad con base a escalas subjetivas, han trivializado esta herramienta, desconociendo sus factores potenciadores, y más preocupante aún, sus limitantes. En tal sentido, esta propuesta coincide con **Cascaes et al. (2015)** respecto de estar conscientes de las limitaciones de esta potente herramienta, que, como toda obra humana, es perfectible. Por ello, es necesario ponderar las condiciones que hacen posible el uso del alfa de Cronbach, pues de lo contrario se generan conclusiones erróneas **(Contreras & Novoa-Muñoz, 2018)**.

Domínguez-Lara y Merino-Soto (2015), mencionan las limitaciones que posee el coeficiente alfa de Cronbach, resaltando la necesidad de los intervalos de confianza y sus reportes. Sin embargo, sus reflexiones no generan un método para robustecer el coeficiente alfa de Cronbach, cuya principal limitación para la definición de intervalos de confianza es el no disponer de un soporte acotado. En ello avanza la propuesta presentada en las líneas precedentes, posibilitando un recorrido compacto del alfa aun con covarianzas negativas. Asimismo, **Ledesma (2004)** sugería utilizar la técnica de bootstrapping como proceso de generación de la muestra aleatoria para iniciar el proceso inferencial, no obstante, la limitación del soporte compacto no fue superada. Por otro lado, **Koning y Franses (2006)** focalizaron su atención en la dependencia del número de ítems y el coeficiente alfa de Cronbach, pero no así en su definición. Por ello, los autores del presente texto consideran que la propuesta avanza de manera importante en la superación de las limitaciones de dicha prueba, al menos en cuanto a expresiones negativas y posibilidad de establecer un nivel de significancia para su interpretación.

Así las cosas, la presente propuesta teórico-aplicada pretende realizar un salto sustancial respecto al uso de este coeficiente. De esta manera, se propone la generación de un p-valor que indique que tan alejado (o no) está el coeficiente alfa Cronbach respecto de 1, reemplazando el consenso en la comunidad científica, con estándares surgidos de datos concretos y no en la interpretación arbitraria de los investigadores **(Taber, 2018)**. Con ello, es posible utilizar la expresión de significancia adicionada al coeficiente de alfa de Cronbach, que en la propuesta se denomina alfa-Game. Esta herramienta, resulta especialmente útil ante situaciones adversas para el alfa de Cronbach, como lo son test pequeños, tamaños de muestras reducidos o ante resultados con expresiones negativas, que tienden a confundir a las y los investigadores no expertos en estadística. Así, y como pudiera decir **Domínguez-Lara y Merino-Soto (2015)** esta propuesta va en la línea de que los instrumentos cuenten con evidencias satisfactorias de su confiabilidad, pero con conocimiento de sus restricciones.

Finalmente, para facilitar el proceso de determinación de las inferencias sobre alfa-Game, se puede utilizar el software Jamovi, con la aplicación Reliability analysis y Reverse Scaled Items (La descarga puede ser realizada de manera directa en https://www.jamovi.org/download.html), el cual detecta covarianzas negativas y permite su adecuación. Aquí lo importante es dejar la interpretación sin reportar los intervalos de confianza y poder redefinir la investigación en ciencias sociales con base a lo que los mismos datos indican. La propuesta queda a vuestra disposición.

Agradecimientos

El autor José González agradece a la Universidad de Playa Ancha y el apoyo del Ministerio de Educación de Chile, a través del Plan de Fortalecimiento de Universidades Estatales, UPA 1799.

Referencias

Campo-Arias, A., & Oviedo, H. C. (2008). Propiedades psicométricas de una escala; la consistencia interna. (5), 831-839. doi: https://doi.org/10.1590/S0124-00642008000500015

Casella, G., & Berger, R. L. (2002). . Pacific Grove, CA: Duxbury.

Cascaes, F., Gonçalves, E., Valdivia, B., Bento, G., Da Silva, T., Soleman, S., & Da Silva, R. (2015). Estimadores de consistencia interna en las investigaciones en salud: el uso del coeficiente alfa. (1), 129-138. doi: https://doi.org/10.17843/rpmesp.2015.321.1585

Cervantes, V. H. (2005). Interpretaciones del coeficiente de Cronbach. (1), 9-28. Recuperado de http://www.humanas.unal.edu.co/psicometria/files/1113/8574/8604/Articulo_1_Alfa_de_Cronbach_9-28_2.pdf

Contreras, S., & Novoa-Muñoz, F. (2018). Ventajas del alfa ordinal respecto al alfa de Cronbach ilustradas con la encuesta AUDIT-OMS. , 42: e65. doi: https://doi.org/10.26633/RPSP.2018.65

Correa, J. C., & Sierra, E. (2003). Intervalos de confianza para la comparación de dos proporciones. , 26(1), 61-75. Recuperado de https://revistas.unal.edu.co/index.php/estad/article/view/28677

Covarrubias, F., & Cruz, M. G. (2016). El estatuto epistemológico de los instrumentos científicos. , (12), 1-33. Recuperado de http://www.intersticiossociales.com/index.php/is/article/view/98

Cunha, J., Xu, J., Rosário, P., & Núñez, J. C. (2018). Validity and reliability of the Parental Homework Management Scale. , 30(3), 337-343. doi: https://doi.org/10.7334/psicothema2017.426

Cronbach, L. J. (1951). Coefficient and the internal structure of tests. (3), 297-334. doi: https://doi.org/10.1007/BF02310555

Domínguez-Lara, S., & Merino-Soto, C. (2015) ¿Por qué es importante reportar los intervalos de confianza del coeficiente alfa de Cronbach? 13, 2. Recuperado de http://revistaumanizales.cinde.org.co/rlcsnj/index.php/Revista-Latinoamericana/article/view/2030

Domínguez-Lara, S. (2012). Propuesta para el cálculo del Alfa Ordinal y Theta de Armor. (1), 213-217. doi: https://doi.org/10.15381/rinvp.v15i1.3684

Figueroa, T., Rodriguez, K., Díaz, D., & Zapata, A. (2018). Actitudes en torno a las brechas generacionales en el trabajo: desarrollo y validación de una escala. Revista Iberoamericana De Psicología, 11(2), 61-68. https://doi.org/10.33881/2027-1786.rip.11207

García, J. L., Ruiz, A. M., & Mesa, J. L (2011). Evaluación de la incertidumbre de medidas en ángulos, distancias y desniveles medidos con instrumentación topográfica (149), 6-27. Recuperado de http://www.mappinginteractivo.com/plantilla.asp?id_articulo=1835

Gempp, R., & Saiz, J. (2013). El coeficiente K2 de Livingston y la fiabilidad de una decisión dicotómica en un test psicológico. (1), 217-226. doi: http://doi.org//10.11144/Javeriana.UPSY13-1.eckl

González, M., & Ibáñez, I. (2018). Propiedades psicométricas de una versión española breve de 30 ítems del Cuestionario de Ansiedad y Depresión (MASQE30). (1), 1-10. doi: https://doi.org/10.11144/Javeriana.upsy17-1.ppve

González, J., & Pazmiño, M. (2015). Cálculo e interpretación del Alfa de Cronbach para el caso de validación de la consistencia interna de un cuestionario, con dos posibles escalas tipo Likert. (1), 62-67. Recuperado de https://nbn-resolving.org/urn:nbn:de:0168-ssoar-423821

González, J., Viveros, F., & Carvajal, C. (2016). Coeficientes edumétricos para la validez y dificultad de un test: Propuesta. (3), 467-481. doi: https://dx.doi.org/10.4067/S0718-07052016000400025

González, J., Galvis, D., & Hurtado, C. (2014). La distribución Beta Generalizada como un modelo de sobrevivencia para analizar la evasión universitaria. (3), 467-481. doi: https://dx.doi.org/10.4067/S0718-07052014000100008

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient . , (3), 177-189. doi: https://doi.org/10.1080/07481756.2002.12069034

Koning, A. J., & Franses, P. H. (2006). Confidence intervals for Cronbach's coefficient values. , 1(1), 1-43. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=423658

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. , 151-160. doi: https://doi.org/10.1007/BF02288391

Ledesma, R. (2004). CI: un programa de cálculo de intervalos de confianza para el coeficiente alfa de Cronbach. , 9(1), 31-37. doi: http://dx.doi.org/10.1590/S1413-82712004000100005

López-González, L., Amutio, A., Orio, X., Gázquez, J. J., Pérez-Fuentes, M. C., & Molero, M. M. (2018). Development and validation of the Relaxation-Mindfulness Scale for Adolescents (EREMIND-A). , 30(2), 224-231. doi: https://doi.org/10.7334/psicothema2017.225

Maroco, J., & Garcia-Marques, T. (2013). Qual a fiabilidade do alfa de Cronbach? Questões antigas e soluções modernas? (1), 65-90. doi: https://doi.org/10.14417/lp.763

Merino, C., Domínguez-Lara, S., & Contreras, A. (2018). Significancia práctica o significancia estadística: re-análisis de la magnitud de las diferencias en Contreras et al. , (4), 559-560. doi: http://dx.doi.org/10.4067/S0370-41062018005000506

Oviedo, H. C., & Campo-Arias, A. (2005). Aproximación al uso del coeficiente alfa de Cronbach. (4), 572-580. Recuperado de http://www.scielo.org.co/pdf/rcp/v34n4/v34n4a09.pdf

Prieto, G. & Delgado, A. (2010). Fiabilidad y validez. (1), 67-74. Recuperado de http://www.papelesdelpsicologo.es/pdf/1797.pdf

Taber, K. S. (2018). The Use of Cronbach's When Developing and Reporting Research Instruments in Science Education. 48, 1273–1296. doi: https://doi.org/10.1007/s11165-016-9602-2

Vaske, J. J., Beaman, J., & Sponarski, C. C. (2017). Rethinking internal consistency in Cronbach's (2), 163-173. doi: https://doi.org/10.1080/01490400.2015.1127189

Ventura-León, J. L. (2017). El significado de la significancia estadística: comentarios a Martínez-Ferrer y colaboradores. , 59(5), 499-500. doi: http://dx.doi.org/10.21149/8482

Vilchez, J. (2017). Agresividad en el aula: Influencia de factores sociopsciológicos [medida y fiabilidad]. (2), 93-98. *https://doi.org/10.33881/2027-1786.rip.10210*