

## Intervalos de confianza asimétricos para el índice la validez de contenido: Un programa Visual Basic para la V de Aiken

César Merino Soto<sup>1\*</sup> y José Livia Segovia<sup>2</sup>

<sup>1</sup> Universidad de San Martín de Porres (Perú)

<sup>2</sup> Universidad Nacional Federico Villarreal (Perú)

**Resumen:** El presente artículo presenta un programa informático en el lenguaje Visual Basic para el cálculo del coeficiente V de Aiken, usado para cuantificar la validez de contenido, y su intervalo de confianza en niveles de tasa de error tipo I previamente seleccionados. La construcción de los intervalos de confianza se hizo por la derivación del método score y publicado en Penfield y Giacobbi (2004). El creciente uso de este método y del enfoque de intervalos de confianza se ajusta a las demandas actuales mejoramiento de los resultados estadísticos inferenciales en la investigación psicológica. Se discute su uso en contextos aplicados.

**Palabras clave:** Validez de contenido; V de Aiken; intervalos de confianza; psicometría; validez.

**Title:** Confidence intervals for the content validity: A Visual Basic computer program for the Aiken's V.

**Abstract:** The present document reporting a Visual Basic computer program for compute the Aiken's V coefficient, this used for quantification of the content validity, and their confidence interval in specified levels of previously Type I error rate. The construction of confidence intervals is performed using the *score* method, and it was published in Penfield and Giacobbi (2004). The growing use of this method and of the confidence interval approach is fitted to the request of improving the results in inferential statistic procedures for the psychological research. It is discussed their use in applied context.

**Key words:** Content validity; Aiken's V; confidence interval; psychometry; validity.

Particularmente, los intervalos de confianza es forma más informativa y útil de expresar el grado de imprecisión o incertidumbre asociada con los resultados cuantitativos de una investigación o algún cálculo en general; eso sería una consecuencia apropiada de la situación muy común de utilizar una muestra de limitado tamaño. Actualmente, la construcción de intervalos de confianza para los puntajes obtenidos es una práctica recomendada y a veces es referida como un elemento esencial para la interpretación del reporte de resultados de un sujeto (Charter, 2003).

En el presente artículo presentamos un programa informático para calcular intervalos de confianza para el índice de validez de contenido,  $V$ , propuesto por Aiken (1980, 1985). El procedimiento para obtener el intervalo de confianza es por medio del método *score*, que recientemente fue derivado en Penfield y Giacobbi (2004) para este coeficiente. Hay pocos precedentes sobre el uso del coeficiente  $V$  de Aiken en publicaciones científicas hispanas, excepto la sencilla presentación descriptiva por Ecurra (1989) que no se incluía un programa informático para su cálculo. El desarrollo de un programa que calcule este coeficiente podría ser de menor utilidad dado que el cálculo no presenta complejidades aún para el usuario sin conocimientos estadísticos, pero el programa presentado aquí incluye cálculos que van más allá del planteamiento original de Aiken, quien consideraba la tradicional prueba de hipótesis del coeficiente V, particularmente con su contraste de hipótesis nula fijada en  $V = 0.50$  para establecer los valores críticos. Esta característica nueva resuelve los cálculos para obtener intervalos de confianza para  $V$  usando el método *score*, pero ya que calcularlos puede llevar a errores debido a sus ecuaciones (el lector los verá más adelante en el presente artículo).

El método de los intervalos de confianza para la  $V$  de Aiken representa los límites que tienen una alta probabilidad de que ocurra V en la población; y esta probabilidad es de una extensión del 95% u otro límite crítico que el examinador elija. La justificación de utilizar intervalos de confianza para la cuantificación de la validez de contenido va de acuerdo con el actual énfasis de su uso para reportar hallazgos de investigación psicológica (Wilkinson y APA Task Force on Statistical Inference, 1999; American Psychological Association, 2001; Fidler, 2002).

### Coefficiente de validez de contenido V de Aiken

En la literatura metodológica se han descrito algunos enfoques de análisis cuantitativos para la validez de contenido que parecen promisorios (por ejemplo, Anderson y Gerbing, 1991; Schriesheim, Powers, Scandura, Gardiner y Lankau, 1993; Sireci, 1998a; Hinkin y Tracey, 1999), pero requieren de mayor examen metodológico para obtener de ellas resultados confiables en investigaciones aplicadas. Un método sencillo es el cálculo del coeficiente  $V$  de Aiken (Aiken, 1980; 1985), y se aplica en un método lógico de validez: la opinión de expertos sobre la validez de un material evaluativo. Este coeficiente es una de las técnicas para cuantificar de validez de contenido o relevancia del ítem respecto a un dominio de contenido en  $N$  jueces, cuya magnitud va desde 0.00 hasta 1.00; el valor 1.00 es la mayor magnitud posible que indica un perfecto acuerdo entre los jueces respecto a la mayor puntuación de validez de los contenidos evaluados. La interpretación del coeficiente usa la magnitud hallada y la determinación de la significancia estadística mediante las tablas de valores críticos que se pueden hallar en Aiken (1985). La ecuación, algebraicamente modificada por Penfield y Giacobbi (2004), es:

\* Dirección para correspondencia [Correspondence address]: César Merino Soto. Calle Enrique Palacios 430, Lima 9 (Perú).  
E-mail: [sikavax@yahoo.com.ar](mailto:sikavax@yahoo.com.ar)

$$V = \frac{\bar{X} - l}{k} \quad (1)$$

$\bar{X}$  es la media de las calificaciones de los jueces en la muestra,  $l$  es la calificación más baja posible, y  $k$  es el rango de los valores posibles de la escala Likert utilizada. Por ejemplo, si  $l = 1$  y  $k = 5$ , entonces  $k = 5 - 1 = 4$ . La estimación de los intervalos de confianza para el coeficiente *V de Aiken* usará el método *score* (Wilson, 1927; Penfield y Giocobbi, 2004), que a continuación describimos brevemente.

### Intervalos de confianza para V de Aiken por el método score

Para la construcción de intervalos de confianza se usa tradicionalmente el método de Wald, que asume la distribución normal asintótica de la variable en estudio; pero particularmente su aplicación al coeficiente *V* es inapropiada (Penfield y Giocobbi, 2004). Uno de los métodos para afrontar las limitaciones de este método tradicional es el método de Wilson (1927), conocido como *método score*, que tiene muy buenas propiedades para el análisis debido que no depende de la distribución normal de la variable, es asimétrica respecto a la variable y es altamente exacto. Una presentación técnica del método *score* aplicado a las proporciones y sus diferencias está en Newcombe (1998a, 1998b, 1998c), y otra menos técnica en Newcombe y Merino (2006). La derivación de la formulación original del método *score* para su uso con *V* de Aiken aparece en el apéndice de Penfield y Giocobbi (2004), quienes trataron la *V de Aiken* como una proporción para obtener la construcción del intervalo en un nivel de confianza determinado. La ecuación para el límite inferior del intervalo es:

$$L = \frac{2nkV + z^2 - z\sqrt{4nkV(1-V) + z^2}}{2(nk + z^2)} \quad (2)$$

Y para el límite superior del intervalo:

$$U = \frac{2nkV + z^2 + z\sqrt{4nkV(1-V) + z^2}}{2(nk + z^2)} \quad (3)$$

- L límite inferior del intervalo
- U límite superior del intervalo
- Z valor en distribución normal estándar
- V *V* de Aiken, calculado por la fórmula 1.
- n número de jueces

La complejidad computacional de este procedimiento puede hacer que el usuario sea proclive a errores, y por tal motivo facilitamos al lector un programa informático de libre distribución.

### El programa y su disponibilidad

El programa está escrito y compilado en Visual Basic 6.0, y corre bajo las versiones actuales de Windows para PC, específicamente Windows 95 o superior. Para ejecutar el programa, se presenta una interfaz gráfica fácil de manejar. La activación del programa requiere hacer doble clic en su icono respectivo.

El usuario ingresará los siguientes datos: a) la calificación promedio o rating obtenido previamente para el ítem seleccionado para el análisis, b) la calificación mínima y máxima posibles en la escala, y c) el número de jueces o expertos.

Para pasar de un cuadro de texto al siguiente, el usuario puede usar el ratón, presionar la tecla *enter* o *tab* luego del ingreso de cada dato. Con estos datos, el programa calculará automáticamente el rango de valores de las calificaciones (calificación máxima - calificación mínima), el índice *V* de Aiken, y los intervalos de confianza en los niveles del 90%, 95% y 99%, que son los niveles más usuales en las estimaciones de intervalos de confianza. El programa, de uso sencillo, está disponible sin costo escribiendo al autor principal en las direcciones de contacto.

### Usos

El intervalo de confianza para la *V* de Aiken permite al usuario probar si la magnitud obtenida del coeficiente es superior a una que es establecida como mínimamente aceptable para concluir sobre la validez de contenido de los ítems. Por ejemplo, de desea probar si un coeficiente *V* es significativamente diferente del mínimo nivel de validez según los estándares de los expertos; este estándar puede estar en un nivel liberal (Cicchetti, 1994) de  $V_o = 0.50$ , o en un nivel más conservador, como  $V_o = 0.70$  o más (Charter, 2003). Luego de recolectar el juicio de 5 jueces, en una escala de 1 al 5, el rating promedio es 4.2, y al 95% de confianza el intervalo del coeficiente obtenido será [0.62, 0.90]. Con este dato, no podemos aceptar el ítem a la luz del estándar mínimo más exigente (0.70), ya que el límite inferior del intervalo está por debajo de tal punto. En las fases iniciales de la construcción de ítems, se puede elegir un criterio más liberal ( $V_o = 0.50$ ) o un nivel de confianza igual a 90%, especialmente si el número de jueces es pequeño (Penfield y Giacobbi, 2004). También, para fines inferenciales, el enfoque de intervalos de confianza permite probar la hipótesis nula de igualdad mediante coeficientes *V*, ya que la información contenida en el intervalo de confianza conduce a juzgar el grado en que los coeficientes *V* comparados se traslapan o se mantienen lo suficientemente alejados como para concluir que son diferentes.

La significancia estadística para el coeficiente V de Aiken tiene por hipótesis nula que un valor del contraste de 0.50, que representa la variación aleatoria; pero este nivel es inaceptable para fines prácticos y está por debajo de las recomendaciones en la construcción de pruebas para cualificar la magnitud de un coeficiente de validez (Cicchetti y Sparrow, 1981; Cicchetti, 1994; Nunnally y Bernstein, 1995; Anastasi y Urbina, 1997; Prieto y Muñiz, 2000; Charter, 2003; Lindley, Bartram y Kennedy, 2005). Al investigador aplicado y al usuario además de interesarle si el valor estimado del coeficiente de validez es diferente de lo que ocurriría en relaciones aleatorias, observará si tal magnitud es lo suficientemente alto como para tomar decisiones acertadas sobre la aceptabilidad de los ítems para capturar el contenido del constructo. Contrastando con la original propuesta de prueba de hipótesis de Aiken, se puede probar cualquier valor declara-

do como valor nulo o  $V_0$  y no únicamente  $V_0 = 0.50$ . Esto da al usuario flexibilidad sobre los valores críticos apropiados al contexto de estudio y a la fase de construcción de escalas en que se halla.

Finalmente, el programa permite interactivamente hacer una estimación del número de jueces necesarios para lograr un nivel deseado intervalo de confianza, y obtener un grado de precisión del coeficiente *V de Aiken*. Bajo el mismo contexto del ejemplo anterior, se necesitaría un mínimo de 14 jueces y un nivel de confianza de 90% para lograr un intervalo que incluya al estándar establecido, ya que el intervalo del coeficiente *V de Aiken* se modificaría a [0.70, 0.87]. Es fácil concluir que a medida que el tamaño muestral se incrementa, la amplitud del intervalo será menor, y por lo tanto, la precisión de la estimación del coeficiente *V* será mejor.

## Referencias

- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45, 131-142.
- Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, 40, 955-959.
- Aiken, L. R. (1996). *Tests psicológicos y evaluación*. México, D. F.: Prentice Hall.
- American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.
- Anastasi, A. y Urbina, S. (1997). *Psychological testing* (7th ed). New York: McMillan.
- Anderson, J. C., y Gerbing, D. W. (1991). Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities. *Journal of Applied Psychology*, 76, 732-740.
- Charter, R. A. (2003). A breakdown of reliability coefficients by test type and reliability method, and the clinical implications of low reliability. *Journal of General Psychology*, 130(3), 290-304.
- Cicchetti D.V. y Sparrow, S.S. (1981) Developing criteria for establishing the interrater reliability of specific items in a given inventory. *American Journal of Mental Deficiency*, 86, 127-137.
- Cicchetti, D. V. (1994) Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessments*, 6, 284-290.
- Escurra, L. (1989) Cuantificación de la Validez de Contenido por Criterio de Jueces. *Revista de Psicología - PUCP*, 6, 103-111.
- Fidler, F. (2002). The 5th edition of the APA Publication Manual: Why its statistics recommendations are so controversial. *Educational and Psychological Measurement*, 62, 749-770.
- Hinkin, T. R. y Tracey, J. B. (1999) An analysis of variance approach to content validation. *Organizational Research Methods*, 2(2), 175-186.
- Lindley, P., Bartram, D., y Kennedy, N. (2005) *EFPA review for the description and evaluation of psychological test: Test review form and notes for reviewers (version 3.41)*. Report of the Standing Committee on Test and Testing to General Assembly. Granada: EFPA.
- May, R. B., Masson, M. J., y Hunter, M. A. (1990). *Application of statistics in behavioral research*. New York: Harper & Row.
- Miller, J. M., y Penfield, R. D. (2006). Using the Score method to construct asymmetric confidence intervals: An SAS program for content validation in scale development. *Behavior Research, Methods, Instruments, and Computers*, 37, 450-452.
- Newcombe, R. G. y Merino, C. (2006) Intervalos de confianza para las estimaciones de proporciones y sus diferencias entre ellas. *Interdisciplinaria*, 23(2), 141-154.
- Newcombe, R.G. (1998a). Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Statistics in Medicine*, 17, 857-872.
- Newcombe, R.G. (1998b). Interval estimation for the difference between independent proportions: comparison of eleven methods. *Statistics in Medicine*, 17, 873-890.
- Newcombe, R.G. (1998c). Improved confidence intervals for the difference between binomial proportions based on paired data. *Statistics in Medicine*, 17, 2635-2650.
- Nunnally, J. C., y Bernstein, I. J. (1995). *Teoría psicométrica* (3ra ed.). México, D. F.: McGraw-Hill
- Penfield, R. D. y Giacobbi, P. R., Jr. (2004) Applying a score confidence interval to Aiken's item content-relevance index. *Measurement in Physical Education and Exercise Science*, 8(4), 213-225.
- Penfield, R. D. y Miller, J. M. (2004) Improving content validation studies using an asymmetric confidence interval for the mean of expert ratings. *Applied Measurement in Education*, 77(4), 359-370.
- Prieto, G. y Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 77, 65-71.
- Schriesheim, C. A., Powers, K. J., Scandura, T. A., Gardiner, C. C., y Lankau, M. J. (1993). Improving construct measurement in management research: Comments and a quantitative approach for assessing the theoretical adequacy of paper-and-pencil and survey-type instruments. *Journal of Management*, 19, 385-417.
- Sireci, S. G. (1998a) Gathering and analyzing content validity data. *Educational Assessment*, 5(4), 299-321.
- Sireci, S. G. (1998b) The construct of content validity. *Social Indicators Research*, 45, 83-117.
- Wilkinson, L., y APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594-604.
- Wilson, E.B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22, 209-212.

(Artículo recibido: 6-6-2008; aceptado: 21-10-2008)