

INSTRUMENTO DE MEDIDA PARA EL RAZONAMIENTO ESTADÍSTICO DE ESTUDIANTES DE POSGRADO DE LAS CIENCIAS DE LA SALUD

Jaime Gaviria-Bedoya, Difariney González-Gómez, y Jhony Villa-Ochoa
 Universidad de Antioquia, Colombia
jaime.gaviria@udea.edu.co

Se presenta el diseño y validación de un instrumento de medida para el razonamiento estadístico de estudiantes de posgrado de Ciencias de la Salud, de una universidad pública colombiana. Los ítems fueron adaptados de instrumentos como el CAOS y el GOALS, y de investigaciones y artículos publicados en Ciencias de la Salud. La validez de contenido se realizó mediante juicio de expertos en educación estadística y/o estadística. Los jueces evaluaron cada ítem en una escala ordinal. Se calcularon estadísticos descriptivos e índices psicométricos de porcentaje de acuerdo y de validez de contenido. El proceso de revisión del instrumento dio como resultado una versión final de 41 ítems y los estadísticos obtenidos mostraron evidencia de un alto nivel de acuerdo entre los jueces expertos.

INTRODUCCIÓN

El razonamiento estadístico es considerado un objetivo de aprendizaje que es importante promover y desarrollar como parte de la formación de los estudiantes que toman cursos de estadística en diferentes niveles escolares (Andrade et al., 2017; Ben-Zvi et al., 2018). En particular, este razonamiento es esencial para los estudiantes de las Ciencias de la Salud, puesto que les permite comprender la información cuantitativa que encuentran en su contexto, como interpretar riesgos y probabilidades en estudios médicos, al igual que tomar decisiones en situaciones en las que la variabilidad y la incertidumbre estén presentes (Fioravanti et al., 2019). Pese a la reconocida importancia del razonamiento estadístico, en especial en la formación universitaria, investigadores sugieren realizar más estudios que aborden este razonamiento y su medición en estudiantes de las Ciencias de la Salud (Ferreras, 2014; Fioravanti et al., 2019). Además, distintas revisiones de literatura coinciden en la importancia de usar instrumentos validados para medir este constructo teórico (Langrall et al., 2017; Zieffler et al., 2008).

Aunque los estudiantes de posgrado tienen características diferentes a las de estudiantes de secundaria y universitarios, como la edad, la formación académica y profesional, y la experiencia en investigación, esta ha sido una población poco considerada en investigaciones en educación estadística (Andrade et al., 2017). Los estudiantes de posgrado de las Ciencias de la Salud deben utilizar la estadística como una herramienta metodológica para resolver problemas de su propio campo profesional, por lo que requieren de un razonamiento estadístico que les permita articular los métodos y técnicas de la estadística con las situaciones de su contexto profesional (Ocaña-Riola, 2016). Un objeto de estudio que permite articular el razonamiento estadístico con las situaciones de contexto de los estudiantes, es el razonamiento estadístico situado (Bakker et al., 2012). Dado que el razonamiento estadístico situado no ha sido suficientemente investigado en la población de estudiantes de posgrado de las Ciencias de la Salud (Cui et al., 2010; Fioravanti et al., 2019), el objetivo del presente estudio es describir el proceso de diseño y validación de un instrumento para su medición en estudiantes de posgrado de las Ciencias de la Salud de una universidad pública colombiana.

MÉTODO

Diseño del Instrumento

El diseño del instrumento se realizó siguiendo las recomendaciones de Los Estándares para Pruebas Educativas y Psicológicas (American Educational Research Association et al., 2014). Primero, se realizó una conceptualización teórica del razonamiento estadístico mediante una revisión de literatura sobre este constructo teórico (Gaviria-Bedoya et al., 2021) y se propuso una definición del constructo adaptado a las Ciencias de la Salud. Se implementó una revisión de literatura semi-sistématica sobre las comprensiones del razonamiento estadístico en algunos congresos en educación matemática mediante una síntesis cualitativa. La inclusión de los documentos en esta revisión de literatura comprendió cuatro etapas: identificación, tamización, elección e inclusión. De 84 documentos hallados, para la revisión se seleccionaron 19 ponencias publicadas en ocho congresos en educación matemática y estadística entre 2010 y 2020. Como resultado de la revisión, se definió el razonamiento estadístico como el conjunto de

procesos mentales que permiten realizar inferencias e integrar conceptos e ideas estadísticas (Garfield et al., 2008).

Luego, se definieron las dimensiones del razonamiento estadístico situado, en correspondencia con los procesos estadísticos fundamentados en los marcos conceptuales *Guidelines for Assessment and Instruction in Statistics Education II* (GAISE II) (Bargagliotti et al., 2020) y la práctica de la estadística (Watson et al., 2018) adaptados a las Ciencias de la Salud. Las dimensiones definidas fueron: (a) Problema y planificación estadística, (b) Análisis de datos estadísticos y epidemiológicos, y (c) Conclusiones e interpretación de resultados.

Se eligió un formato de prueba objetiva de opción múltiple con única respuesta y verdadero-falso, puesto que este tipo de ítems son adecuados para evaluar habilidades cognitivas, como las involucradas en el razonamiento estadístico sobre ideas abstractas y su relación. Además, los ítems objetivos se han utilizado comúnmente en investigación en educación estadística, porque están estructurados, son fáciles de calificar objetivamente y permiten su aplicación a grandes muestras de estudiantes. Luego, se construyó la prueba de especificaciones (blueprint), y un banco inicial de 55 ítems que, luego del proceso de revisión de expertos, se redujo a 41 ítems en la versión final.

Para seleccionar los ítems, se tomaron dos tipos de fuentes: (a) instrumentos para el razonamiento estadístico reconocidos, publicados y validados, como el Comprehensive Assessment of Outcomes in Statistics (CAOS) (delMas et al., 2007), el Goals and Outcomes Associated with Learning Statistics (GOALS) (Sabbag y Zieffler, 2015), y el Cuestionario de Evaluación de Razonamiento EStadístico (CERES) (Ferreras, 2014); y (b) investigaciones publicadas en las áreas de las Ciencias de la Salud que trataran sobre enfermedades infecciosas, como el SIDA, y no infecciosas, como el cáncer, además de investigaciones relacionadas con la pandemia, como la evaluación del desempeño del sistema de vigilancia en salud pública de Colombia durante los primeros 50 días de la pandemia de la COVID-19, o la relevancia del conocimiento estadístico para profesionales de la salud durante la pandemia. También se adaptaron ítems de publicaciones que trataran el tema del uso de la estadística en las Ciencias de la Salud. Las áreas de contenido a evaluar se corresponden con el programa de estudios de los cursos de bioestadística de primer año de los estudiantes de maestría y doctorado, e incluyen estadística descriptiva, probabilidad, inferencia estadística paramétrica y no paramétrica.

Validación del Instrumento

La validez de contenido se realizó mediante el juicio por parte de cuatro expertos en diferentes disciplinas como educación estadística, bioestadística, epidemiología o psicometría, quienes evaluaron cada ítem respecto a: claridad (se comprende fácilmente), coherencia (guarda relación lógica con su respectiva dimensión), relevancia (debe ser incluido), y suficiencia (los ítems de una dimensión son suficientes para obtener su medición) (Escobar-Pérez y Cuervo-Martínez, 2008). Para calificar el nivel de cumplimiento de criterio de los ítems en cada categoría, se empleó la escala: 1: no cumple, 2: nivel bajo, 3: nivel moderado, y 4: nivel alto.

Análisis de la Información

Con los resultados de las evaluaciones del juicio de expertos, se obtuvieron resúmenes estadísticos descriptivos por ítem y por dimensión, como media, mediana y desviación, así como el porcentaje de acuerdo entre los jueces con puntuaciones mayores o iguales a 3 (nivel moderado en cada categoría) (Sireci y Faulkner-Bond, 2014). Además, se obtuvieron algunos índices de congruencia ítem-objetivo, como el índice de validez de contenido V propuesto por Aiken (1980), que es la suma ponderada de las calificaciones asignadas a cada ítem, y está dado por la fórmula:

$$V = \sum_{i=1}^{c-1} \frac{i * n_i}{N(c - 1)}$$

donde N es el número de jueces, c el número de categorías a evaluar en cada ítem, n_0 el número de calificaciones iguales a 1, y n_i el número de calificaciones iguales a i-1, para $i = 1, 2, \dots, c - 1$ (Aiken, 1980). El índice de validez de contenido V oscila entre 0 y 1, y se interpreta como la proporción de jueces expertos que evaluaron el ítem por encima del punto medio de la escala de congruencia (Sireci y Faulkner-Bond, 2014).

RESULTADOS

Definición del Constructo Teórico y sus Dimensiones

El razonamiento estadístico situado se definió como el conjunto de procesos mentales que permiten la comprensión de los eventos de salud, a través de una práctica estadística que integra el conocimiento estadístico con los contextos específicos de las Ciencias de la Salud. Se identificaron tres dimensiones para este constructo: (a) Problema y planificación estadística: proceso mental en el que los estudiantes de posgrado deben evidenciar la capacidad de reconocer preguntas estadísticas en diferentes escenarios; (b) Análisis de datos estadísticos y epidemiológicos: esta dimensión es el núcleo del razonamiento estadístico; y (c) Conclusiones e interpretación de resultados: es la capacidad de interpretar los resultados estadísticos en el contexto de los datos de las Ciencias de la Salud y la capacidad para evaluar la evidencia de esos resultados en relación con el problema estadístico y la pregunta correspondiente de la primera dimensión.

Ejemplo de un Ítem del Instrumento para el Razonamiento Estadístico Situado

Los ítems 27 y 28 se refieren a la siguiente situación: el gráfico muestra la distribución de la edad en la que se produjo la menopausia en un grupo de 1000 mujeres de la consulta de ginecología de un hospital.

27. Refiriéndose al gráfico de la Figura 1 y sin hacer ningún cálculo, ¿qué medidas de resumen recomendaría informar y por qué? (elaboración propia adaptada del instrumento CAOS, delMas et al., 2007).

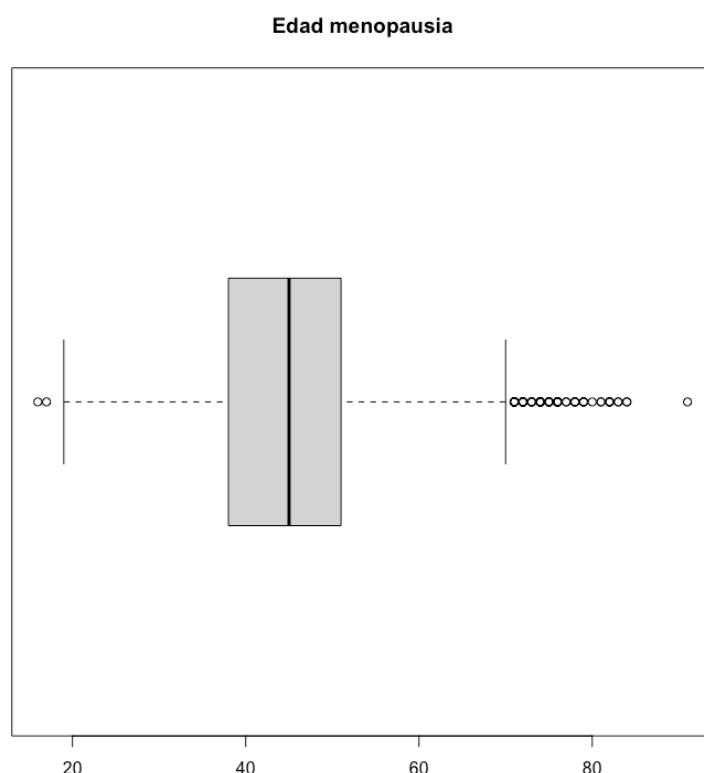


Figura 1. Gráfico de cajas y bigotes de la edad en la que ocurrió la menopausia en un grupo de 1000 mujeres (elaboración propia, datos simulados mediante el software R)

- Recomendaría informar de la media y la desviación estándar de la edad, ya que es con estas estadísticas con las que se construyen los intervalos de confianza.
- Recomendaría informar de la mediana y el rango Inter cuartil de la edad, ya que son medidas más robustas ante datos extremos.
- Recomendaría reportar la moda y el rango de la edad, ya que estos representan todos los datos.

Características del Ítem 27:

- Dimensión: 3. Conclusiones e interpretación de resultados (CIR).
- Contenido: Medidas descriptivas adecuadas según la presencia de valores atípicos.
- Tipo de razonamiento: Razonamiento estadístico sobre medidas descriptivas afectadas por valores atípicos.
- Indicador: CIR3-Utilizar pruebas estadísticas para responder a preguntas estadísticas y comunicar los resultados.

Resultados del Juicio de Expertos

El proceso de revisión del instrumento dio como resultado una versión final de 41 ítems, e implicó revisar la redacción de aquellos ítems calificados por debajo de tres en alguna de las categorías evaluadas. Dado el número de ítems del instrumento final y el límite de espacio, en la Tabla 1 se muestran los promedios de las calificaciones por dimensión en cada una de las cuatro categorías evaluadas por los jueces. Estos resultados muestran que, en general, la mediana de las calificaciones de los jueces es cercana a cuatro, nivel alto en cada categoría. Además, los porcentajes de acuerdo promedio entre los jueces y los índices de validez de contenido V de Aiken (1980) fueron cercanos a uno en las cuatro categorías evaluadas, lo que indica un alto nivel de acuerdo entre los jueces expertos, en las tres dimensiones del instrumento.

Tabla 1. Medidas descriptivas e índices psicométricos de validez de contenido para el instrumento de medida razonamiento estadístico situado

Categoría	Estadísticos	Dimensión		
		1	2	3
Claridad	Media	3.52	3.51	3.58
	Mediana	3.92	3.69	3.95
	Desviación	0.87	0.67	0.76
	% Acuerdo	0.83	0.85	0.80
	Índice V	0.84	0.84	0.84
Consistencia	Media	3.79	3.79	3.76
	Mediana	4.00	4.00	4.00
	Desviación	0.42	0.42	0.47
	% Acuerdo	0.90	0.92	0.90
	Índice V	0.93	0.93	0.90
Relevancia	Media	3.62	3.83	3.78
	Mediana	3.96	4.00	4.00
	Desviación	0.72	0.33	0.42
	% Acuerdo	0.85	0.93	0.90
	Índice V	0.87	0.94	0.91
Suficiencia*	Media	3.5	3.75	3.75
	Mediana	4	4	4
	Desviación	1	0.5	0.5
	% Acuerdo	0.75	1	1
	Índice V	0.83	0.92	0.92

*Nota: puesto que la suficiencia es un aspecto que se evalúa para cada dimensión

En la Tabla 2 se muestran los resultados del porcentaje de acuerdo y el índice de validez de contenido V de Aiken (1980) para las cuatro categorías consideradas en el juicio de expertos. Estos resultados brindan evidencia de la validez de contenido del instrumento, e indican que hubo un alto acuerdo intergrupal entre los expertos.

Tabla 2. Índices psicométricos de validez de contenido por categoría evaluada en el juicio de expertos

Categoría	% Acuerdo ≥ 3	Índice V de Aiken (1980)
Claridad	0,83	0,84
Consistencia	0,91	0,92
Relevancia	0,90	0,91
Suficiencia	0,92	0,89

CONCLUSIONES

El presente estudio busca hacer un aporte teórico al campo de la educación estadística en cuanto al conocimiento sobre el razonamiento estadístico de los estudiantes de posgrado de las Ciencias de la Salud. Los resultados del juicio de expertos proporcionaron evidencia de la validez de contenido del instrumento propuesto y mostraron que las calificaciones de los jueces, en general, fueron altas en cada categoría evaluada. Además, los índices psicométricos obtenidos por dimensión y por categoría evaluada fueron cercanos a uno, resultados que brindan evidencia sobre la validez de contenido del instrumento, e indican que hubo un alto acuerdo intergrupal entre los expertos. Se espera realizar la prueba piloto y la aplicación de la versión final en el segundo semestre de 2022 a una muestra de estudiantes de posgrado de una universidad en Colombia, con el fin de proveer evidencia acerca de la validez y fiabilidad del instrumento.

El diseño y validación de un instrumento de medida para el razonamiento estadístico situado, busca aportar a los profesores que enseñan bioestadística a nivel universitario en las facultades de las áreas de la salud, información útil sobre los procesos cognitivos involucrados en el razonamiento estadístico, que permita identificar aquellos aspectos más relevantes para el aprendizaje y la adquisición de las habilidades en la dimensión específica.

AGRADECIMIENTOS

Los autores agradecen a los grupos de investigación MATHEMA-FIEM, Demografía y Salud; y Aplicaciones Estadísticas y Salud Pública de la Universidad de Antioquia por el apoyo a este trabajo.

REFERENCIAS

Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, 40(4), 955–959. <https://doi.org/10.1177/001316448004000419>

American Educational Research Association, American Psychological Association, y Nacional Council on Measurement in Education. (2014). *The standards for educational and psychological testing*. American Educational Research Association.

Andrade, L., Fernández, F., y Álvarez, I. (2017). Panorama de la investigación en educación estadística desde tesis doctorales 2000–2014. *TED: Tecné, Episteme y Didaxis*, 41, 87–107. <https://doi.org/10.17227/01203916.6039>

Bakker, A., Mierlo, X., y Akkerman, S. (2012). Learning to integrate statistical and work-related reasoning. In D. Ben-Zvi y K. Makar (Eds.), *Teaching and learning statistics: Topic study group 12. 12th International Congress on Mathematical Education* (pp. 21–31).

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., y Spangler, D. A. (2020). *Pre-K–12 guidelines for assessment and instruction in statistics education II (GAISE II). A framework for statistics and data science education*. American Statistical Association; National Council of Teachers of Mathematics. https://www.amstat.org/asa/files/pdfs/GAISE/GAISEIIPreK-12_Full.pdf

Ben-Zvi, D., Makar, K., y Garfield, J. (Eds.). (2018). *International handbook of research in statistics education*. Springer. <https://doi.org/10.1007/978-3-319-66195-7>

Cui, Y., Roberts, M. R., y Gotzmann, A. (2010, May). *Evaluating statistical reasoning of college students in the social and health sciences with cognitive diagnostic assessment* [Paper presentation]. 2010 American Educational Research Association Annual Meeting. <http://www.statlit.org/pdf/2010-Cui-Roberts-Gotzmann-AERA.pdf>

delMas, R., Garfield, J., Ooms, A., y Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, 6(2), 28–58. <https://doi.org/https://doi.org/10.52041/serj.v6i2.483>

Escobar-Pérez, J., y Cuervo-Martínez, Á. (2008). Validez de contenido y juicio de expertos: Una aproximación a su utilización. *Avances En Medición*, 6(2), 27–36.

Ferreras, S. (2014). *Desarrollo del razonamiento estadístico en estudiantes de enfermería* [Tesis doctoral, Universidad Pontificia Comillas]. DSpace. <https://repositorio.comillas.edu/xmlui/handle/11531/1300>

Fioravanti, R., Greca, I., y Meneses, J. (2019). Caminos do ensino de estatística para a área da saúde. *Revista Latinoamericana de Investigación En Matemática Educativa*, 22(1), 67–96. <https://doi.org/10.12802/relime.19.2213>

Garfield, J., Ben-Zvi, D., Chance, B., Medina, E., Roseth, C., y Zieffler, A. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Springer. <https://doi.org/10.1007/978-1-4020-8383-9>

Gaviria-Bedoya, J., González-Gómez, D., y Villa-Ochoa, J. A. (2021). Una revisión de las comprensiones del razonamiento estadístico. In L. Tauber y J. Pinto (Eds.), *Tendencias y nuevos desafíos de la investigación en educación estadística en Latinoamérica: Libro de ponencias de las III Jornadas Latinoamericanas de investigación en educación estadística* (pp. 204–211). Universidad Nacional del Litoral. <https://www.fluc.unl.edu.ar/institucional/wp-content/uploads/sites/3/2018/08/Tendencias-y-nuevos-desafios-Educ-Estadistica.pdf>

Langrall, C., Makar, K., Shaughnessy, M., y Nilsson, P. (2017). Teaching and learning probability and statistics: An integrated perspective. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 490–525). National Council of Teachers of Mathematics.

Ocaña-Riola, R. (2016). The use of statistics in health sciences: Situation analysis and perspective. *Statistics in Biosciences*, 8(2), 204–219. <https://doi.org/10.1007/s12561-015-9138-4>

Sabbag, A., y Zieffler, A. (2015). Assessing learning outcomes: An analysis of the GOALS-2 instrument. *Statistics Education Research Journal*, 14(2), 93–116. <https://doi.org/https://doi.org/10.52041/serj.v14i2.263>

Sireci, S., y Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema*, 26(1), 100–107. <https://doi.org/10.7334/psicothema2013.256>

Watson, J., Fitzallen, N., Fielding-Wells, J., y Madden, S. (2018). The practice of statistics. In D. Ben-Zvi, K. Makar, y J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 105–137). Springer. https://doi.org/10.1007/978-3-319-66195-7_4

Zieffler, A., Garfield, J., Alt, S., Dupuis, D., Holleque, K., y Chang, B. (2008). What does research suggest about the teaching and learning of introductory statistics at the college level? A review of the literature. *Journal of Statistics Education*, 16(2). <https://doi.org/10.1080/10691898.2008.11889566>