



UN MODELO PARA CARACTERIZAR EL RAZONAMIENTO ESTADÍSTICO DE ESTUDIANTES DE SECUNDARIA AL COMPARAR CONJUNTOS DE DATOS REPRESENTADOS MEDIANTE GRÁFICOS DE CAJA

Jorge Zambrano y Rosa María Palacios
Secretaría de Educación Distrital (Colombia)
willzam@hotmail.com, rosamariwill@hotmail.com

En este trabajo se presenta un modelo para caracterizar el razonamiento estadístico de los estudiantes al interpretar la información que es representada por el gráfico de cajas. El origen de dicho modelo se motiva en una experiencia de aula que considera y aplica los resultados obtenidos en una investigación realizada como trabajo de grado de la Maestría en Docencia de las Matemáticas y adscrita a la línea de investigación en Educación Estadística de la Universidad Pedagógica Nacional en el año 2009. Esta investigación pretende categorizar el razonamiento estadístico de un grupo de estudiantes de secundaria en un colegio público de la ciudad de Bogotá. Para obtener dicha categorización se propuso comparar conjuntos de datos representados mediante gráficos de caja, y se empleó la teoría de clasificación conocida como taxonomía SOLO, la cual a su vez fue articulada con siete elementos de razonamiento sugeridos por los autores del presente trabajo.

PALABRAS CLAVE

Razonamiento estadístico, gráfico de cajas, taxonomía SOLO.

INTRODUCCIÓN

Los contenidos y la manera en la cual se debe enseñar la estadística en la escuela, sigue siendo un tema poco trabajado en nuestro país a pesar del notable auge y crecimiento del interés por parte de la comunidad académica sobre la Educación Estadística.

En cuanto a la calidad y cantidad de trabajos sobre Educación Estadística en el contexto educativo colombiano, la situación merece que se le preste más atención. Basta con revisar las investigaciones que se han realizado al respecto en las comunidades de educadores matemáticos del país, para darse cuenta que a pesar de que existe interés en el tema, las referencias no son suficientes para establecer un precedente claro sobre la Educación Estadística en Colombia. A propósito, Rocha (2007) plantea que existen algunos problemas estructurales en la educación estadística en el país como las escasas investigaciones en torno a la enseñanza y aprendizaje de la probabilidad y estadística en el país, la ausencia de una comunidad académica que oriente y oxigene los diferentes procesos a desarrollar dentro de la enseñanza



y el aprendizaje de la estadística, la carencia de medios de comunicación entre los profesores de estadística y el débil vínculo entre la investigación y la instrucción. (p. 24)

Esto tiene como consecuencia que la situación en el aula no sea la esperada por quienes se interesan por mejorar el aprendizaje de la estadística escolar y que su enseñanza sea una disculpa para usar otros temas matemáticos y ejercitar la capacidad del cálculo o la representación gráfica sin tener en cuenta la actividad con datos reales y las orientaciones hacia el razonamiento estadístico. Trabajos, como el de Fernández, Sarmiento y Soler (2008), caracterizan los conocimientos estadísticos y probabilísticos de profesores de educación básica y media de colegios oficiales de Bogotá, y dan cuenta de las actitudes hacia la enseñanza, el aprendizaje y la utilidad de la estadística en el contexto escolar, visto desde la perspectiva del razonamiento estadístico. Por otro lado, es necesario reconocer que la enseñanza de la estadística se ha limitado al enfoque descriptivo, cuyo propósito en palabras de Inzunza (2006):

[...] se centra en la descripción de los datos, utilizando para ello procedimientos de agrupamiento a través de tablas o distribuciones de frecuencia y el cálculo de medidas descriptivas, como la tendencia central y la dispersión. Si bien, en estadística descriptiva se utilizan representaciones gráficas, [...], el análisis no enfatiza en el uso de ellas en forma sistemática, sino más bien como recurso adicional al cálculo de estadísticos que permiten dar una idea global del comportamiento de los datos. (p. 1)

Esto dificulta que los estudiantes estén en capacidad de desarrollar habilidades básicas que según Gal (2002) se establecen para la alfabetización estadística en términos de su capacidad para interpretar y evaluar críticamente la información estadística y para discutir y comunicar opiniones respecto a estas informaciones. Una de las razones fundamentales es que, como ya se mencionó, la enseñanza de la estadística descriptiva se centra principalmente en el trabajo sobre algoritmos y cálculo de estadísticos numéricos, en detrimento de establecer relaciones o generar comprensiones al comparar muestras.

Afortunadamente con la aparición de los lineamientos curriculares se integró la enseñanza de la estadística y la probabilidad al pensamiento aleatorio y en los estándares básicos de competencias matemáticas se establece que los conceptos de estadística y probabilidad deben enseñarse en la educación colombiana.

También se presenta el enfoque del Análisis exploratorio de datos como la propuesta curricular sugerida, lo cual da gran importancia al estudio del gráfico de cajas. En consecuencia, en el documento titulado 'pensamiento estadístico y tecnologías computacionales' (MEN, 2004), han propuesto la introducción del AED para la enseñanza de la estadística, ya que esta nueva filosofía en el tratamiento de la información concibe: los propósitos de la formación estadística menos desde el uso fundamentado en las reglas del cálculo, y más desde el uso fundamentado en la comprensión de los datos y sus características. Esta concepción nos permite sostener que un alumno que es capaz de hacer inferencias, extrapolar en los datos o los gráficos la información que está más allá de la percepción, adquiere una mejor formación que le permita tomar decisiones en presencia de la incertidumbre con soporte en sus exploraciones sobre los datos. (p. 15)



De esta manera, el interés principal de la presente investigación es el de reportar una caracterización sobre el razonamiento estadístico de los estudiantes cuando utilizan el gráfico de cajas para comparar dos distribuciones de datos. Se espera que, en consonancia con lo sugerido por el MEN (2004), esta caracterización aporte elementos para el diseño y aplicación de propuestas curriculares que contribuyan en la enseñanza y aprendizaje de la estadística. Por ello, consideramos importante dar respuesta a la siguiente pregunta de investigación:

¿Cómo puede ser caracterizado el razonamiento estadístico de los estudiantes de grado undécimo (15 a 17 años) cuando interpretan y comparan distribuciones de datos representados por gráficos de caja?

MARCO DE REFERENCIA CONCEPTUAL

El marco teórico en el cual se fundamentó esta investigación está dividido en tres dimensiones fundamentales. La primera, está encaminada a presentar, de una manera breve, la filosofía del Análisis Exploratorio de Datos (AED) como una de las corrientes actuales del análisis de datos, además de las características educativas que lo han llevado a proponerse como enfoque de reforma curricular para posteriormente, presentar el gráfico de cajas el cual fue objeto de estudio de esta investigación, abordando elementos importantes como su origen, construcción y estructura.

Una segunda sección está destinada a la presentación de la metodología de clasificación, es decir, la taxonomía SOLO; la cual es un intento por hacer una categorización por niveles de complejidad de las estrategias que los estudiantes presentan al aprender un objeto estadístico, aunque también tiene aplicaciones a otras áreas del conocimiento. Finalmente, se presenta el componente del conocimiento estadístico, el cual se centra en la caracterización de los términos alfabetización, pensamiento y razonamiento estadístico, desarrollando especialmente lo relacionado al razonamiento, por ser la perspectiva teórica del aprendizaje estadístico que nos interesó observar.

Fundamentos del análisis exploratorio de datos

El AED es una filosofía relativamente nueva en el tratamiento de la información estadística, la cual emplea una variedad de técnicas gráficas para explorar los datos. Este grupo de técnicas gráficas y analíticas creado por el estadístico norteamericano Jhon Tukey (1977), permite conseguir un conocimiento previo de los datos a analizar siempre desde una perspectiva exploratoria, considerándose como una nueva rama de la estadística. Sin embargo, algunos autores como Bakker (2004) sostienen que “a diferencia de la estadística descriptiva tradicional el objetivo en el AED no es probar una hipótesis preestablecida con datos de un muestreo aleatorio, sino más bien éste se enfoca en la dirección de patrones y tendencias no anticipadas de todo tipo de datos, ya sea una muestra aleatoria o no” (p. 2).

Las técnicas estadísticas empleadas por el AED, buscan favorecer el estudio de estadísticos robustos que permitan observar la estructura de un conjunto de datos sin que esta se vea demasiado afectada por la presencia de datos atípicos. En ocasiones se selecciona una muestra donde se nota la presencia de alguno de estos datos inusitados y es necesario establecer si tal observación hace parte importante de la muestra y si se

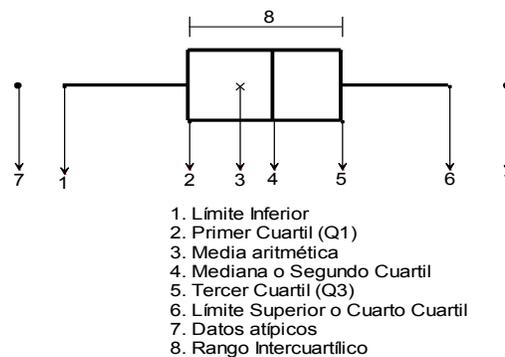


debe retener o rechazar, pues la presencia de este dato hace que las medidas descriptivas que generalmente se calculan, pueden verse afectadas y por ejemplo, la media aritmética ya no sea una medida representativa del conjunto de datos.

El gráfico de cajas

Este gráfico cuyo nombre original es 'Box and whisker plot', es decir, gráfico de caja y bigotes fue introducido por Tukey (1977) en el segundo capítulo de su libro *Exploratory Data Analysis*. Esta representación gráfica pretende proporcionar los valores de los datos pero sin que estos aparezcan con todo detalle, a fin de no perder su configuración espacial; para ello mantiene únicamente sus valores indicativos principales, que pueden resumirse en: localizaciones y agrupaciones significativas de valores, zonas en las que predomine la dispersión, referencia visual de la simetría central y de los extremos, referencia visual de la curtosis, relacionando la longitud de la caja y de los bigotes, rango valores atípicos, anomalías o valores significativamente alejados.

Es un gráfico representativo de las distribuciones de un conjunto de datos en cuya construcción se usan cinco medidas descriptivas de los mismos, a saber: mediana, primer cuartil, tercer cuartil y dos valores extremos: el valor máximo y el valor mínimo; además presenta al mismo tiempo, información sobre la tendencia central, la dispersión y la simetría de un conjunto de datos permitiendo identificar con claridad y de forma individual observaciones que se alejan de manera poco usual del resto de los datos; a estas observaciones se les llama valores atípicos. Admite igualmente comparar a la vez varios grupos de datos sin perder información.

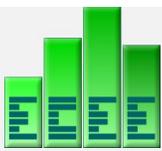


Gráfica 1. Elementos de un diagrama de cajas o box plot

Fundamentos y niveles de la taxonomía SOLO

Biggs y Collis (1982) basaron su modelo sobre la noción de que en cualquier situación de aprendizaje, los resultados tanto cuantitativos, ¿cuánto aprendió?, como cualitativos, ¿qué tan bien y de qué manera se aprendió?, están determinados por una compleja interacción entre las técnicas de enseñanza usadas y las características de los estudiantes, ya mencionadas.

En cuanto a la perspectiva cuantitativa, Biggs y Collis, (1982) conciben el aprendizaje “como la agregación de contenido; al ser un aprendizaje en el que más se aprende, el más exitoso” (p. 2), mientras que desde la perspectiva cualitativa “se asume que los



estudiantes aprenden acumulativamente, al interpretar y combinar el nuevo material con el que ellos ya conocen, de manera que al cambiar su comprensión progresivamente van aprendiendo” (p. 3). Así, las construcciones de los estudiantes vienen a ser los resultados de aprendizaje, por lo que el objetivo de la evaluación es determinar en qué punto están actualmente los estudiantes acerca de su comprensión del concepto o habilidad evaluada. De esta manera, en la evaluación cualitativa, en la que tiene sentido la taxonomía SOLO, es necesario determinar cómo sería la evolución de un concepto para definir el desarrollo que presentarían los estudiantes acerca de su aprendizaje. A continuación se presentan los niveles que la conforman.

En el nivel idiosincrásico o preestructural, las respuestas de los estudiantes en este nivel se caracterizan por no demostrar ninguna aproximación a la posibilidad de hacer conexiones entre los componentes de la información que se presenta. En el nivel Uniestructural, las respuestas de los estudiantes se enfocan sobre sólo un aspecto del concepto evaluado, sin hacer ningún tipo de conexión con otras ideas, pero es capaz de razonar sobre algún contenido significativo. Para el nivel multiestructural, las respuestas de los estudiantes identifican más de un aspecto del concepto evaluado, pero no determinan el significado del todo pues las ideas involucradas están desordenadas y sin relación.

En el nivel relacional, los estudiantes a este nivel pueden indicar conexiones entre los hechos y la teoría, las acciones y los propósitos. Las respuestas muestran cierta comprensión de algunos componentes, los cuales son integrados conceptualmente, demostrando cómo las partes contribuyen al todo para darle estructura y significado. Por último el nivel abstracto extendido se caracteriza porque las respuestas a este nivel consideran toda la información relevante y relaciona sus componentes bajo una estructura hipotética abstracta que hace posible realizar deducciones acerca de información que no fue incluida originalmente.

Razonamiento estadístico

Al referirnos al razonamiento estadístico es posible caracterizarlo en términos de habilidades específicas que emergen principalmente en la actividad estadística de resolución de problemas y que lo diferencian en varios aspectos del tratamiento de problemas matemáticos.

En trabajos no tan recientes como los de Chervaney, Collier, Fienberg, Jonson y Neter (1977, citados en Garfield, 2002), lo definían como lo que una persona es capaz de hacer con el contenido estadístico, como recordar, reconocer y discriminar entre conceptos estadísticos, y las habilidades que demuestra tener en la utilización de conceptos estadísticos para la resolución de problemas; en este sentido se le caracteriza como un proceso de tres etapas: comprensión del problema, en el cual se identifica la situación dada con una clase de problemas; planificación y ejecución, que implica la aplicación de métodos apropiados para resolver el problema y su ejecución; y finalmente la evaluación e interpretación que exige la explicación de los resultados en relación con el problema original.



ASPECTOS METODOLÓGICOS

El estudio se puede caracterizar como descriptivo e interpretativo, en el que se plantea un nuevo modelo que permite categorizar los razonamientos estadísticos de los estudiantes cuando comparan dos conjuntos de datos representados mediante gráficos de caja. La primera fase de la investigación fue el diseño y aplicación de actividades que aproximaron a los estudiantes a la construcción, interpretación y uso del gráfico de caja, la segunda fase tuvo que ver con el diseño del cuestionario piloto que generó el cuestionario de estudio, el cual proporcionó información para el diseño de las categorías de análisis sobre el razonamiento estadístico siendo ésta la última fase. Las categorías de análisis se diseñaron sobre una tabla de doble entrada que relaciona siete elementos de razonamiento que los autores consideran están presentes en cualquier lectura del gráfico de caja con cuatro niveles de razonamiento inspirados en la taxonomía SOLO.

Estos son: tamaño de la muestra, escala, localización, dispersión, intervalos de concentración, datos adyacentes y datos atípicos. Dichos elementos de análisis fueron determinados por los autores, considerando la investigación de Pfannkuch (2006) y algunos otros que se han considerado importantes. La construcción del modelo de razonamiento se ha diseñado a partir del análisis de los resultados evidenciados en la aplicación del cuestionario aplicado a los estudiantes de la muestra.

PROPUESTA

Para presentar las categorías del razonamiento estadístico manifestado por los estudiantes de la muestra de estudio al comparar dos muestras representadas por gráficos de cajas, se ha diseñado una tabla de contingencia en la que están relacionados los diferentes elementos de razonamiento planteados por los autores y cuatro niveles de razonamiento inspirados en la taxonomía SOLO.

De esta manera, los elementos de razonamiento están ubicados en las filas de la Tabla 1 y para cada uno de estos se establecen los cuatro niveles organizados progresivamente, empezando con el nivel más bajo y terminando con el de mayor complejidad. Las celdas de dicha tabla contienen los descriptores del elemento de razonamiento para cada nivel y las frecuencias absolutas de estudiantes pertenecientes a la muestra de estudio que se pueden ubicar en dicha categoría.



NIVEL ELEMENTO	NIVEL 1 IDIOSINCRÁSICO Prestructural	NIVEL 2 TRANSICIONAL Uniestructural	NIVEL 3 CUANTITATIVO Multiestructural	NIVEL 4 ANALÍTICO Relacional
TAMAÑO DE LA MUESTRA	Afirma que el tamaño de la muestra está relacionado con la amplitud y forma del gráfico.	Independiza el tamaño de la muestra de la interpretación inicial del gráfico, asumiendo un solo gráfico de caja.	Caracteriza el efecto del muestreo en dos gráficos de caja, pero no establece relaciones entre ellos.	Reconoce que no se puede identificar el tamaño de la muestra en una lectura del gráfico, pero relaciona el tamaño de los gráficos para obtener conclusiones.
ESCALA	No relaciona la escala de la recta numérica con la representación gráfica de cada caja.	Identifica los valores de los cinco resúmenes en un solo gráfico de cajas	Reconoce e interpreta separadamente para cada gráfico cada uno de las medidas de posición, pero no compara dichos resúmenes entre los dos gráficos.	Relaciona los valores de las medidas de resumen de cada gráfico y establece conclusiones acerca de las características de ambas distribuciones de datos.
LOCALIZACIÓN	No identifica la mediana, en el gráfico de caja, como centro porcentual de un conjunto de datos.	Reconoce la mediana como una medida de tendencia central representativa de todo el conjunto de datos en un solo conjunto de datos.	Describe las medianas de dos conjuntos de datos, pero no establece relaciones entre ellos.	Establece relaciones entre dos gráficos de caja al comparar sus dos medianas.
DISPERSIÓN	No identifica el rango intercuartílico en un gráfico de caja.	Identifica el rango intercuartílico como una medida de dispersión en un solo gráfico de caja.	Reconoce los rangos intercuartílicos en dos gráficos de caja, pero no establece relaciones entre estas distribuciones.	Identifica la dispersión como una característica de la forma del conjunto de datos relacionando los rangos intercuartílicos de los gráficos de caja.



INTERVALOS DE CONCENTRACIÓN	No reconoce las agrupaciones de los datos respecto al recorrido de los valores de la variable.	Describe el comportamiento de un conjunto de datos en particiones específicas de un solo gráfico de caja.	Identifica el comportamiento de la agrupación de los datos en dos gráficos de caja, pero sin establecer relaciones entre ellos.	Compara la concentración de los datos en zonas específicas de dos conjuntos de datos y establece conclusiones acerca de la relación entre ambas distribuciones de datos.
DATOS ADYACENTES	No identifica los valores que son límites de las cercas.	Identifica el mayor y el menor dato considerados en los segmentos exteriores de un gráfico de caja.	Describe los valores considerados como límites de los bigotes en dos conjuntos de datos por separado, sin establecer relaciones entre los conjuntos de datos.	Establece conclusiones acerca de la distribución de dos conjuntos de datos al comparar los datos adyacentes de ambas distribuciones.
DATOS ATÍPICOS	No tiene en cuenta la presencia de datos atípicos en la distribución de datos.	Considera y caracteriza los datos atípicos en un solo gráfico de caja.	Determina los datos atípicos en los gráficos de caja para ambas distribuciones de datos pero sin establecer relaciones.	Caracteriza el efecto que tiene la inclusión de los valores atípicos en la distribución del conjunto de datos y establece conclusiones al comparar los cambios en las distribuciones por esta inclusión.

Tabla 1. Categorías de razonamiento estadístico al comparar dos conjuntos de datos representados por gráficos de caja

CONCLUSIONES

Este trabajo puede ser de utilidad para aquellas personas interesadas en investigar acerca de la introducción del AED en el currículo oficial colombiano, puesto que consideramos importante que antes de insertar las herramientas propias del AED se debe hacer un análisis de la manera que los estudiantes interpretan la información presentada a través de estas técnicas.

El modelo jerárquico propuesto se puede adaptar a diferentes grados cuando se pretenda que los estudiantes interpreten y comparen gráficos de caja. Además, también se puede utilizar para analizar el razonamiento estadístico cuando se interpreten otros gráficos o cualquier objeto de estudio estadístico. Para esto es necesario identificar los elementos propios de cada una de las representaciones gráficas más usualmente empleadas en la enseñanza de la estadística.



La lista de elementos de razonamiento y los niveles propuestos se pueden considerar como fundamento para el diseño de unidades didácticas para el aula y en la construcción de instrumentos de evaluación del razonamiento estadístico sobre análisis de datos.

REFERENCIAS

- Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, 3 (2), 64-83. <http://www.stat.auckland.ac.nz/serj>.
- Biggs, J. y Collis, K. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.
- Fernández, F., Sarmiento, B. y Soler, N. (2008). Conocimiento estadístico y probabilístico de profesores de educación básica y media. Bogotá: Universidad Pedagógica Nacional.
- Gal, I. (2002). Adult's statistical literacy. Meanings, components, responsibilities. *International Statistical Review*, 70 (1), 1-25.
- Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education*, 10 (3).
- Inzunza, S. (2006). ¿Qué enseñar, estadística descriptiva o análisis exploratorio de datos? Trabajo presentado en II Reunión Nacional de Análisis. La actividad experimental en la enseñanza de las Ciencias Naturales y Exactas. Culiacán Sinaloa: Centro de Ciencias de Sinaloa.
- MEN (2004). *Pensamiento estadístico y tecnologías computacionales*. Bogotá: MEN.
- Pfannkuch, M. (2006). Comparing box plot distributions: teacher's reasoning. *Statistics Education Research Journal*, 5 (2), 27-45.
- Rocha, P. (2007). Epistemología del razonamiento estadístico y aleatorio y su desarrollo a partir de proyectos de trabajo estadístico como innovación en la enseñanza de los objetos de estudio estocásticos. En Memorias del ENAES. Bogotá.
- Tukey, J. (1977). *Exploratory data analysis*. New York. Editorial Addison Wesley.