

CARACTERÍSTICAS Y CONCEPCIONES ACERCA DE LA NOCIÓN DE DISTRIBUCIÓN EN EL RAZONAMIENTO SOBRE LA COMPARACION DE CONJUNTOS DE DATOS: EL CASO DE UN ESTUDIANTE

CHARACTERISTICS AND CONCEPTIONS ABOUT THE NOTION OF DISTRIBUTION IN THE REASON ON THE COMPARISON OF DATA SETS: THE CASE OF A STUDENT

Vivian Yulieth Marulanda Aroca, Maritza Méndez Reina
Universidad Pedagógica Nacional (Colombia).
vymarulandaa@upn.edu.co, mmendezr@pedagogica.edu.co

Resumen

Esta comunicación es producto de un trabajo de grado desarrollado en la Maestría en Docencia de la Matemática de la Universidad Pedagógica Nacional, en particular, expone el análisis de los razonamientos que un estudiante de grado séptimo exhibe cuando compara distribuciones de datos a partir del trabajo con una tarea mediada por tecnología digital, considerando elementos teóricos centrados en características y concepciones acerca de la noción de distribución. La estrategia investigativa empleada fue la entrevista basada en tareas, los resultados sugieren que el estudiante considera aspectos de la noción de distribución de manera informal cuando hace comparaciones de distribuciones.

Palabras clave: pensamiento relacionado con probabilidad, estadística, entrevista basada en tareas

Abstract

This report is the result of a thesis developed in the Mathematics Teaching Master's Degree Program at the National Teacher Training University. In particular, it presents the reasoning analysis that a seventh-grade student shows when comparing data distributions from the work with a digital technology-mediated-task; considering theoretical elements based on characteristics and conceptions concerning the distribution notion. The research strategy used was the task-based interview. The results suggest that the student considers distribution notion aspects informally when making distribution comparisons.

Key words: probability-related thinking, statistics, task-based interview

■ Introducción

La propuesta de trabajo de grado que se encuentra en curso, tiene como aspecto central el identificar los diversos razonamientos y concepciones que los estudiantes exhiben cuando comparan conjuntos de datos a partir del trabajo con una tarea mediada por tecnología digital, para dicho propósito la idea de distribución es medular. En este sentido, esta propuesta parte del supuesto que los estudiantes haciendo uso de sus conocimientos iniciales y de una herramienta digital diseñada en GeoGebra, son influenciados en sus razonamientos respecto a las características y concepciones de la distribución, lo cual contribuirá al desarrollo del pensamiento aleatorio y a la formación de ciudadanos estadísticamente cultos.

La cultura estadística, como lo propone Batanero, Diaz, Contreras y Roa (2013) involucra la interpretación de la información estadística que se puede encontrar en diferentes contextos, esta interpretación no puede darse correctamente si no se han desarrollado ideas estadísticas fundamentales, específicamente la de distribución de frecuencias. En este sentido, Ben-Zvi y Arcavi (2001) citados en Garfield y Ben-Zvi (2008) encontraron que las tareas cuidadosamente diseñadas (por ejemplo, comparar distribuciones, manejar valores atípicos), la orientación de los docentes y preguntas desafiantes, junto con conjuntos de datos motivadores y herramientas tecnológicas apropiadas ayudan a los estudiantes a hacer la transición del razonamiento local al global.

El conjunto de tareas que contempla la propuesta, se vale de preguntas orientadas a la comparación de conjuntos de datos para promover la noción de distribución, ya que:

1. Comparar dos o más conjuntos de datos puede estructurarse como una versión inicial e informal de inferencia estadística.
2. Los problemas que involucran comparaciones de conjuntos de datos son a menudo más interesantes que los que involucran a un solo conjunto.
3. Estudiantes de cualquier nivel requieren desarrollar estrategias para comparar conjuntos de datos.
4. En la comparación de conjuntos de datos es importante realizar representaciones gráficas y obtener resúmenes (centro y dispersión) de los datos. Sánchez y Orta (2013, p. 68)

Para efectos de esta comunicación se mostrará parte del marco referencial que orienta la propuesta, la metodología seguida, un episodio de la entrevista basada en la tarea propuesta, y su correspondiente análisis.

■ Marco referencial

Para Batanero y Godino (2001) la distribución proporciona información para analizar una variable, en particular aspectos como saber cuántos datos hay en total, con qué frecuencia aparecen los diferentes valores de la variable y el comportamiento de los datos, sin embargo, no señalan una definición clara sobre esta noción. Algunos autores se han interesado por investigar acerca de la noción de distribución, al respecto Wild (2006) indica que la distribución son los lentes a través de los cuales se observa como varían los datos, por lo que comprender la idea de variabilidad de los datos es un componente clave para comprender el concepto de distribución.

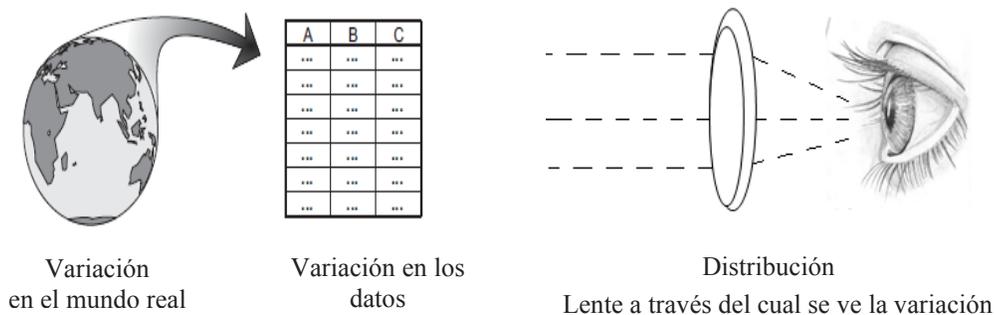


Figura 1. Distribución como lente.
Fuente: Wild (2006).

En línea con estas perspectivas, Wild (2006) hace una distinción entre dos tipos de distribuciones, las empíricas y las teóricas. Las distribuciones empíricas, son aquellas en las que se explica la variación de los datos mediante la observación directa de diferentes representaciones del conjunto de datos, dicha observación permite describir la variabilidad existente en los datos, por ello, este tipo de distribuciones no tienen un componente inferencial. Las distribuciones teóricas por su parte, hacen alusión a las distribuciones que explican la variación de los datos de variables aleatorias bajo modelos paramétricos como la distribución normal, binomial, etc., es decir, se refiere a modelos de probabilidad.

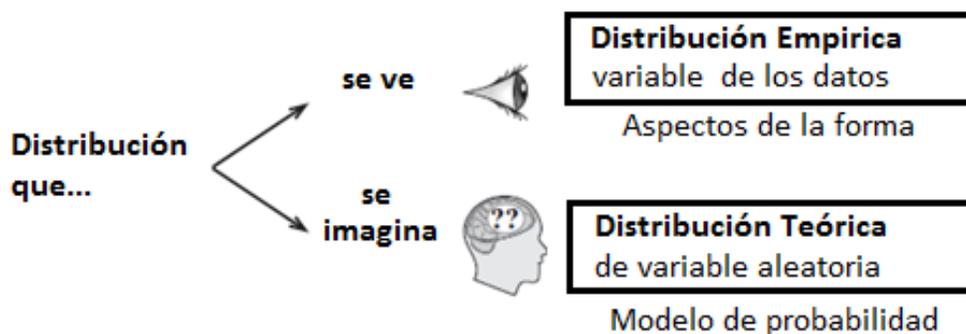


Figura 2. Distribución Empírica vs. Distribución Teórica.
Fuente: Wild (2006).

Las tareas planteadas en la propuesta, y en particular la mostrada en esta comunicación, centra su atención en las distribuciones de tipo empírico, dado que, lo que busca principalmente es favorecer la visualización gráfica, aprovechando las posibilidades que brindan los recursos tecnológicos. Así mismo se busca promover el aprendizaje de la estadística particularmente la distribución de frecuencias mediante la exploración dinámica. A continuación, se presentan las características de la distribución basadas en las investigaciones de algunos autores expertos en el tema.

Distribución y variabilidad

Antes de describir aspectos importantes acerca de la variabilidad, es necesario abordar la cuestión de la terminología que se usará a partir de este momento. Algunos autores usan el término “variación” y “variabilidad” para referirse al mismo concepto, sin embargo, hay quienes hacen distinción de los términos. Reading y Shaughnessy (2004;

citado en Garfield y Ben-Zvi 2008) hacen las siguientes distinciones entre “variación” y “variabilidad”: variación es el nombre que se usa para describir el acto de variar, mientras que variabilidad es el nombre del adjetivo “variable”, lo que significa que es probable que algo varíe o cambie. Teniendo en cuenta que esta distinción no ha sido acordada en la comunidad de la investigación en educación estadística, para este trabajo se tomara el argumento de los autores mencionados y se usara el término variabilidad para las ideas descritas y desarrolladas en este apartado.

Un requisito para comprender la distribución es la idea de variabilidad, que según Wild (2006) está siempre presente en la distribución de frecuencias. Por lo tanto, es importante que los estudiantes perciban la variabilidad en los conjuntos de datos.

Sobre la variabilidad

Según Garfield y Ben-Zvi (2008), es importante que los estudiantes perciban la presencia de la variabilidad en las distribuciones, aun cuando no se hable de ella, por lo que es necesario mostrarles que la variabilidad es una medida que puede ser explicada mediante una comprensión profunda que puede ser guiada como se muestra a continuación:

1. Desarrollando ideas intuitivas de variabilidad
2. Describiendo y representando la variabilidad con medidas numéricas
3. Usando la variabilidad para hacer comparaciones
4. Reconociendo la variabilidad en tipos especiales de distribuciones
5. Identificando patrones de variabilidad en los modelos de adaptación
6. Usando la variabilidad para predecir muestras aleatorias o resultados
7. Considerando la variabilidad como parte del pensamiento estadístico

Otro requisito para concebir la noción de distribución según Bakker y Gravemeijer (2004) es enfocarse en los aspectos de la forma. La forma de una distribución está influenciada por varios aspectos estadísticos como la centralidad, la dispersión de los datos, la densidad y la simetría.

Características de la distribución

Tabla 1. Características de la distribución de datos.

DISTRIBUCIÓN (Entidad conceptual)			
Centralidad	Dispersión	Densidad	Simetría
Media y Mediana	Desviación estándar y Rango	Frecuencia relativa y Cuartiles	Posición de la mayoría de los datos
Datos (Valores individuales)			

Fuente: Bakker y Gravemeijer (2004)

Esta estructura se puede leer de arriba hacia abajo o de abajo hacia arriba. La lectura en forma descendente según Bakker y Gravemeijer (2004) indica el centro, la dispersión, la densidad y la simetría como características de una distribución. En la lectura en ascendente según Mokros y Russell (1995) citados en Bakker y Gravemeijer (2004), se conciben los datos como valores individuales, que se pueden usar para calcular, por ejemplo, la media, la mediana, la desviación estándar y el rango. Esto no implica que se vea la media o la mediana como medida de centralidad o como representante de un conjunto de datos, no obstante, esta perspectiva puede conducir a la comprensión de la noción de distribución.

En Andrade, Fernández y Méndez (2018) se presenta una caracterización de la variabilidad ligada a la noción de distribución que amplía la conceptualización propuesta por Bakker y Gravemeijer (2004) (Tabla 2). Estas características se tienen en cuenta para analizar los razonamientos de los estudiantes y se adaptarán orientadas a la comparación de distribuciones de datos.

Tabla 2. Características de la variabilidad ligadas a la noción de distribución.

<i>Dispersión</i>
Ver la necesidad de organizar u ordenar los datos
Identificar valores máximo y mínimo
Determinar el rango
Establecer relaciones cualitativas entre los datos, y/o relaciones cuantitativas como distancias
Apreciar de manera general la variabilidad
Reconocer valores atípicos
Estimar varianza, desviación estándar
<i>Densidad</i>
Apreciar de forma cualitativa la variación de las frecuencias
Agrupar datos por valores de la frecuencia, es decir, determinar grupos para la frecuencia y asignar a estos los valores de la variable correspondientes
Agrupar datos por valores de la variable, es decir, determinar grupos para la variable y asignar a estos las frecuencias correspondientes
Referirse a la colección de datos por frecuencias relativas o porcentajes de valores particulares de la variable
Apreciar variación entre las longitudes de los intervalos o clases de una o más tablas de frecuencias
Apreciar de forma cualitativa la variación en la relación entre la variable y la frecuencia, determinar si la gráfica es ascendente, descendente, creciente, decreciente, describir curvas, inflexiones de las curvas, prolongaciones, colas
Apreciar de forma cuantitativa la variación en la relación entre la variable y la frecuencia
Describir relaciones de cercanía o lejanía entre las barras, reconocer condensaciones, aglomeraciones
Reconocer valores faltantes, frecuencias nulas, valles, llanos
Describir relaciones de altura entre las barras, reconocer picos, apuntamientos
Reconocer el valor de la variable que corresponde a la mayoría de los datos
Comparar partes con el todo
<i>Centro</i>
Reconocer el valor de la frecuencia más alta, moda
Calcular mediana, media (promedio)
<i>Simetría y asimetría</i>
Reconocer simetría o asimetría
<i>Modelos teóricos</i>
Relacionar con modelos teóricos

Fuente: Andrade, Fernández y Méndez (En prensa)

Concepciones acerca de la noción de distribución

El estudio acerca del razonamiento de los estudiantes sobre los datos de Konold, Higgins, Russell y Khalil (2003), sugiere un marco para describir los niveles en la forma en que las personas entienden los datos, centrándose en cómo las distintas representaciones y el uso de estas resaltan o le quitan énfasis a las características de los datos, se considera necesario describir a continuación estos niveles, ya que es una forma para clasificar la manera en que cada estudiante percibe un conjunto de datos:

1. *Datos como apuntadores*: Esta manera de ver los datos se percibe usualmente en estudiantes muy jóvenes que han recolectado datos por sí mismos. En estos casos los estudiantes tratan la representación de los datos como una imagen que le recuerda lo que sea sobresaliente del evento, por ejemplo: se realiza con los estudiantes una encuesta acerca del juego mecánico que más les gusta en una salida al parque mundo aventura de Bogotá, luego de la recolección de datos, la profesora les pide que den una conclusión acerca de los datos recolectados, a lo que Juan contesta que todos subieron a muchas atracciones mecánicas, lo que hace pensar que Juan ve la representación como una imagen para recordar la salida pedagógica a mundo aventura.
2. *Datos como valores de cada caso*: El estudiante proporciona información sobre el valor de algún atributo para cada caso individual, es decir el estudiante ve los datos como códigos de aspectos particulares de un evento. Este aspecto es a menudo evidente en niños pequeños que tienden a centrarse en la identidad de cada pieza individual de datos, especialmente los valores de datos que les pertenecen, por ejemplo, la profesora le preguntó a Jaime qué podía decir de los datos recolectados acerca de la salida pedagógica a mundo aventura, a lo que Jaime contesta que la atracción mecánica que más le gustó es el martillo, lo que muestra que el estudiante aunque reconoce que los datos pertenecen a un evento específico no es capaz de concluir acerca de los datos sino que lo hace desde su experiencia.
3. *Datos como clasificadores*: El estudiante da información teniendo en cuenta la frecuencia de los datos, es decir da conclusiones de los datos fijándose en el número de veces que aparece un dato. Siguiendo con el mismo ejemplo, la profesora le pregunta a Valentina qué puede decir acerca de los datos recolectados, a lo que Valentina, dice que cinco niños prefieren la montaña rusa, con lo que se observa que la niña dio este tipo de información remitiéndose a la frecuencia de los casos (atracción mecánica).
4. *La distribución como un agregado*: El estudiante emite conclusiones atendiendo a características que no son evidentes en ninguno de los valores de los datos individuales, es decir hace uso de aspectos como la centralidad (media, mediana y rango medio), densidad (frecuencia relativa, mayoría de datos y cuartiles) y asimetría (posición de la mayoría de los datos) para hablar acerca del conjunto de datos. Siguiendo con el ejemplo inicial, la profesora le pregunta a Milena qué puede decir acerca de los datos recolectados, a lo que contesta que la mayoría de los niños no prefieren la atracción mecánica de la araña.

Herramientas tecnológicas para desarrollar el concepto de distribución

Según Garfield y Ben-Zvi (2008) la tecnología puede desempeñar un papel importante en el desarrollo de la noción de distribución al proporcionar un acceso fácil a múltiples representaciones y oportunidades infinitas para manipular y comparar simultáneamente las representaciones del mismo conjunto de datos. Sin embargo, esta no es una tarea simple. Biehler (1997) informa que, a pesar de usar una innovadora herramienta de software para generar y moverse entre diferentes gráficos de datos, la interpretación y la descripción verbal de estos gráficos fueron profundamente difíciles para los estudiantes de secundaria y universitarios, a menos que tuvieran una comprensión conceptual de los conceptos fundamentales.

Garfield y Ben-Zvi (2008), la tecnología puede usarse para ayudar a los estudiantes a ver las conexiones entre las diferentes representaciones gráficas de datos, ayudando a los estudiantes a construir la idea de la distribución como una entidad. Tres usos importantes de la tecnología según los mismos autores son:

1. Permite visualizar la transición de gráficos de casos y argumentos a gráficos de puntos a histogramas, todos basados en el mismo conjunto de datos.
2. Permite ilustrar las formas en que diferentes gráficos de los mismos datos revelan diferentes aspectos de los datos, al tener de manera flexible múltiples representaciones en la pantalla al mismo tiempo, lo que permite a los estudiantes identificar dónde se encuentran uno o más casos en un gráfico.
3. Permite cambiar de manera flexible un gráfico (por ejemplo, hacer que los contenedores sean más anchos o estrechos para los histogramas) para que un patrón o forma sea más clara, o para agregar y eliminar valores para ver el efecto en el gráfico resultante.

■ Metodología

El enfoque que se sigue en este estudio es de tipo fenomenológico, pues se describen y comprenden los aspectos relevantes que se presentan a medida que se dan las diferentes interacciones entre los estudiantes, profesora y la tarea planteada. En consonancia, se opta por la estrategia investigativa entrevista basada en tareas con una aproximación interpretativa, dado que busca rastrear los diversos razonamientos que los estudiantes puedan construir cuando comparan distribuciones de datos a partir del trabajo con una tarea mediada con un applet de GeoGebra. Se parte de la hipótesis de que este estudio posibilita la evolución del pensamiento de los estudiantes hacia la idea de distribución más complejizada.

Para llevar a cabo el estudio, se realizó una entrevista basada en una tarea mediada con un applet de GeoGebra. A continuación, se describe el contexto de ésta y se da cuenta de las preguntas propuestas a los estudiantes:

- Los participantes de la entrevista fueron dos estudiantes de grado séptimo y la profesora de matemáticas quien asumió el rol de entrevistadora e investigadora. Ella diseñó con antelación la tarea y pregunta contingentes que fueron aplicadas en la entrevista.
- La elección de los estudiantes se hizo de forma aleatoria entre los estudiantes del curso 701 de la jornada mañana del colegio Kimy Pernía Domico IED que mostraron interés por participar de la entrevista. Se seleccionaron dos estudiantes, un niño y una niña que demuestran gusto por las clases donde se incluyen herramientas tecnológicas.
- Al observar el plan de estudios se presume que los estudiantes que participaron en la entrevista cuentan con un proceso de instrucción previo de estadística, pues en el grado anterior se dedica un periodo de 13 semanas al componente de formación estadística, en las que se señala de manera explícita el trabajo con tablas y gráficos de barras, y en torno a las medidas de tendencia central.
- La entrevista se aplicó en una sesión de 110 minutos, esta se registró con videos tanto de los estudiantes como de lo que acontecía en la pantalla del computador. Adicionalmente, la investigadora recolectó material escrito de los estudiantes, tomó capturas de pantalla y realizó la transcripción de la entrevista.
- Una vez recogida la información, se organizó en dos tipos de archivos: físico y digital. En el archivo físico se organizaron los escritos de los estudiantes. En el archivo digital se organizó el material registrado en capturas de pantalla, transcripciones y en grabaciones de video.
- Los archivos con la información en bruto fueron el punto de partida para la obtención de los datos investigativos. Se seleccionó la información que fue empleada en el análisis y se descartó la información irrelevante. Se generaron nuevos archivos con la información que se utilizó, especificando de manera escrita: el material registrado, el informante, la fecha, el tiempo empleado, el escenario donde se desarrolló la entrevista, la información que se registró, los dispositivos de registro que se emplearon y la ruta de ubicación del material.

■ Análisis de resultados

A continuación, se muestra un fragmento que constituye un dato representativo del trabajo de grado, este ilustra el dialogo entre la profesora y el estudiante en torno a una tarea que requería que el estudiante observara y manipulara gráficos de dos distribuciones (Figura. 3) presentados en un applet de GeoGebra para dar respuesta a la pregunta: ¿Qué puedes decir de los gráficos?

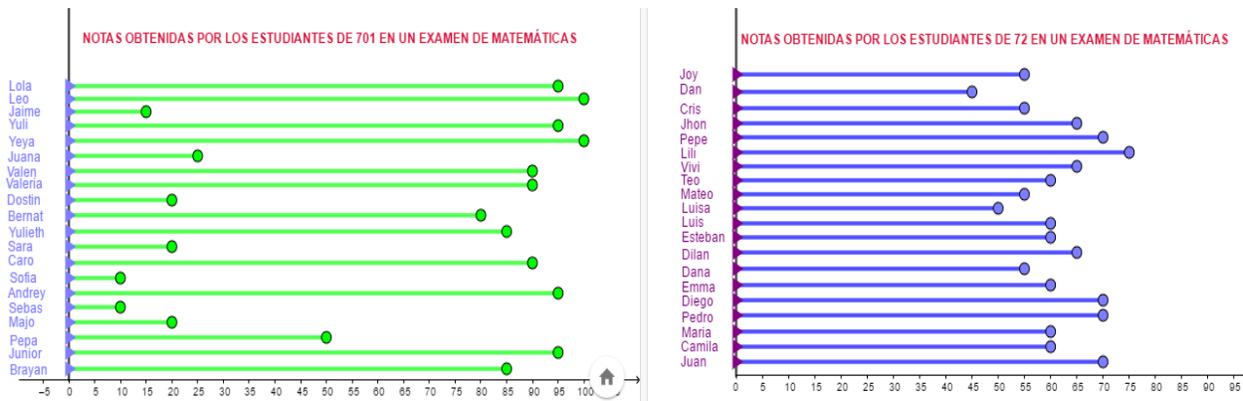


Figura 5. Comparación de distribuciones a partir de gráficos de barras de datos individuales propuesta en la Tarea 1.
Fuente: Elaboración propia

- 1 Richard: Lo que podemos decir de los gráficos de 702 y 701 es que hay diferencia entre los gráficos. Es decir, en 701 hay mayor puntaje, mayor elevación de las notas. Los de 701 tuvieron un puntaje muy elevado. [...]
- 2 Profesora: ¿Pueden organizar los datos de otra manera?
- 3 Richard: Vamos a mover las gráficas. Los que fueron de menor porcentaje [puntaje]. Toca mirar el menor. [...]
- 4 Profesora: ¿Por qué los organizaron así? [figura 5]
- 5 Richard: Por el porcentaje [puntaje] de los cursos en el examen de matemáticas. En 702 solo hay un disparejo que fue la estudiante Lili, que tuvo el mayor porcentaje [puntaje] de ese curso.
- 6 Profesora: Y ¿en 701?
- 7 Richard: En 701 también hubo estudiantes que sacaron desde el 95 al 100
- 8 Profesora: Listo ¿Qué más pueden decir de esos gráficos?... ¿Se ve mejor así [figura 6] o como estaba antes [figura 5]?
- 9 Richard: Se ve mejor así, porque se puede notar cuantos estudiantes sacaron el mismo porcentaje [puntaje] y no confundirse colocando notas a otros.



Figura 6. Comparación de distribuciones a partir de gráficos de barras de datos individuales organizado por los estudiantes. **Fuente:** Elaboración propia

El análisis de la intervención realizada por Richard en [1] permite evidenciar que su atención estaba centrada en describir relaciones de altura entre las barras; y en reconocer picos y apuntamientos. Esta manifestación permite afirmar que su razonamiento estaba centrado en características de la densidad de los datos. En [3], surge en el estudiante la necesidad de encontrar el valor mínimo para a partir de él, organizar los datos. También, de [5-7] gracias a la organización de los datos, el estudiante, además de ver los datos como valores individuales de cada caso, reconoce valores atípicos. Estos indicios muestran que el razonamiento del estudiante estaba centrado en las concepciones de la noción de distribución y en las características de la dispersión de los datos. La intervención del estudiante en [9] permite vislumbrar que prestó especial atención a la agrupación de los datos por valores de la variable, es decir, aunque no brinda información cuantitativa de la frecuencia de los datos, se evidencia que la considera importante para brindar información sobre una variable. Lo anterior, permite inferir que el razonamiento del estudiante estaba centrado en las características de la densidad de los datos.

■ Conclusiones

En el episodio acá mostrado, se aprecia que la secuencia instruccional que se propone en el trabajo contribuye a que el estudiante razone sobre la distribución de los datos a medida que se dan las distintas interacciones con la tarea mediada por tecnología digital. Es así como en el trabajo inicial de exploración de datos presentados en una hoja de GeoGebra de manera estática, el estudiante exhibió razonamientos relacionados con las formas de las distribuciones, en tanto comparó que en uno de los conjuntos de datos tienen valores superiores en comparación con la otra distribución, empleando las observaciones de altura de las barras individuales de cada caso.

Es necesario señalar que cuando el estudiante usó la herramienta de “arrastre” en GeoGebra para organizar los datos lo movilizó a usar otros elementos importantes para comparar distribuciones de datos como la necesidad de identificar el valor mínimo, el reconocimiento de datos atípicos y la agrupación de los datos por valores de la variable. En general, el estudiante señaló elementos que tienen que ver primordialmente con la densidad y la dispersión de los datos mientras que la simetría y la centralidad no fueron nombrados.

■ Referencias bibliográficas

Andrade, L.; Fernández, F. y Méndez, M. (En prensa). Exploración de la noción de distribución desde la variabilidad.

- Bakker, A. y Gravemeijer, K. (2004). Learning to reason about distributions. En D. Ben-Zvi y J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 147-168). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Batanero, C. y Godino, Juan. (2001). Análisis de datos y su didáctica.
- Batanero, C., Díaz, C., Contreras, J. y Roa, . (2013). El sentido estadístico y su desarrollo. *Números*. 83. 7-18.
- Biehler, R. (1997). Software for Learning and for Doing Statistics. *Internlt~onol Srottrst*. 65. 167-189. 10.2307/1403342.
- Garfield, J., y Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. New York: Springer Science & Business Media.
- Konold, C. Higgins, T. Russell, S. y Khalil, K. (2014). Data seen through different lenses. *Educational Studies in Mathematics*. 10.1007/s10649-013-9529-8.
- Sánchez, E. y Orta, J. (2013). Problemas de mediciones repetidas y de riesgo para desarrollar el razonamiento de estudiantes de secundaria en los temas de media y dispersión. *Números*. 83. 65-77.
- Wild, C.J. (2006). The concept of distribution. *Statistics Education Research Journal*, 5 (2), 10-26. <http://www.stat.auckland.ac.nz/serj>