



Exploración de la noción de *distribución* desde la variabilidad

- Exploration of the Notion of Distribution from Variability
- Exploração da noção de distribuição da variabilidade

Resumen

Este artículo presenta los resultados de una investigación dirigida a comprobar si el trabajo desde la variabilidad contribuye al desarrollo de la noción de *distribución estadística*. Para ello se opta por ejecutar un experimento de enseñanza en el que se propone una secuencia de instrucción orientada a que los candidatos a profesores de matemáticas lleguen a una idea más amplia de este concepto. A partir de la exploración intuitiva de colecciones de datos y del tránsito por varios registros de representación con la ayuda de la tecnología, se encuentra que los estudiantes en formación efectivamente notan y consideran diversos aspectos relacionados con la variabilidad que a su vez están ligados a la noción de distribución. En otras palabras, al indagar de manera informal sobre la variabilidad de colecciones de datos, en realidad los estudiantes están develando características de la distribución.

Palabras clave

distribución; variabilidad; experimento de enseñanza; dispersión; densidad; centro; simetría

Abstract

This article presents the results of a teaching experiment aimed at verifying whether the work from variability contributes to the development of the notion of statistical distribution. For this, it is decided to run a teaching experiment in which a sequence of instruction is proposed so that the candidates for mathematics teachers achieve a broader idea of this concept. From the intuitive exploration of the data collections and the transit through several representation registers with the help of technology, it is found that the students notice and identify diverse aspects related to the variability that in turn are linked to the notion of distribution. In other words, by informally inquiring about the variability of data collections, students are actually uncovering characteristics of the distribution

Keywords

distribution; variability; teaching experiment; spread; clustering; center; symmetry

Luisa Andrade Escobar*
Felipe Jorge Fernández Hernández**
Maritza Méndez-Reina***

* Matemática, Universidad Pontificia Javeriana. Profesora investigadora Departamento de Matemáticas, Universidad Pedagógica Nacional, Bogotá, Colombia.
Correo electrónico: landrade@pedagogica.edu.co
Código Orcid: <https://orcid.org/0000-0002-1168-1468>

** Magíster en Scientiae-Estadística, Universidad Nacional de Colombia. Profesor Departamento de Matemáticas, Universidad Pedagógica Nacional, Bogotá, Colombia.
Correo electrónico: ffernandez@pedagogica.edu.co
Código Orcid: <https://orcid.org/0000-0003-3775-4757>

*** Magíster en Docencia de la Matemática, Universidad Pedagógica Nacional. Profesora Departamento de Matemáticas, Universidad Pedagógica Nacional, Bogotá, Colombia.
Correo electrónico: mmendezr@pedagogica.edu.co
Código Orcid: <https://orcid.org/0000-0003-0830-0924>



Resumo

Este artigo apresenta os resultados de uma investigação que procurou verificar se o trabalho de variabilidade contribui para raciocinar sobre a noção de distribuição estatística. Assim propomos uma sequência de instrução com estudantes candidatos para professores de matemática que visa alcançar uma ideia de distribuição mais ampla. A partir da exploração intuitiva das coleções de dados e do trânsito através de diversos registros de representação com a ajuda da tecnologia, constata-se que os alunos efetivamente percebem e consideram vários aspectos relacionados à variabilidade que por sua vez estão ligados à noção de distribuição. Em outras palavras, é claro que, quando os estudantes examinam informalmente a variabilidade das coletas de dados, na realidade, estão desenvolvendo características da distribuição.

Palavras-chave

distribuição; variabilidade; experimento de ensino; dispersão; aglomeração; centro; simetria

Introducción

La noción de *distribución estadística* es básica para la comprensión y, sin embargo, es usual que se aborde en la enseñanza sin ahondar en los elementos que la componen. Se plantea así un proyecto¹ que pretende contribuir a razonar sobre dicha noción desde la exploración de la variabilidad, en estudiantes de nivel universitario. El proyecto se concreta en un experimento de enseñanza que se desarrolla a lo largo de un semestre de clases, y propone un trabajo en varios registros de representación y con tecnología. Se presenta, entonces, una sistematización del trabajo de los estudiantes con una caracterización de la variabilidad ligada a la noción de distribución que emerge de la revisión de ideas de la literatura a la luz del trabajo de los estudiantes. Para finalizar, se exponen unas reflexiones acerca del proceso de aproximación de los estudiantes a la noción de *distribución estadística* durante el desarrollo de la secuencia de instrucción propuesta.

Aspectos metodológicos

La metodología de los experimentos de enseñanza forma parte del enfoque conocido como *investigación de diseño*. Este se enmarca en la perspectiva de investigación que se centra en la relación reflexiva entre la teoría y la práctica, donde, según Cobb (2000), y Steffe, Thompson y Von Glasersfeld (2000), la teoría emerge de la práctica y la realimenta para guiarla. Es usual que la investigación de diseño también involucre el diseño de materiales de instrucción y de análisis retrospectivos.

El experimento de enseñanza desarrollado en el proyecto está orientado a observar y analizar el aprendizaje de los estudiantes sobre la noción de *distribución estadística* a partir del trabajo desde la variabilidad, con la hipótesis de que este trabajo posibilita la evolución del pensamiento de los estudiantes hacia una idea de distribución más complejizada.

El escenario del experimento de enseñanza es el espacio académico de Estadística de la formación de estudiantes para profesores de matemáticas de la Universidad Pedagógica Nacional (UPN), donde el profesor del curso es miembro del equipo de investigación. La secuencia de instrucción diseñada, que busca que los estudiantes sean aprendices activos, propone el contexto de la alimentación sana. Así, desde la intuición y la observación sobre la variabilidad de colecciones de datos numéricos, en conjunto con el tránsito por diferentes registros de representación mediante tecnología, se pretende generar progreso en la formalización de las apreciaciones e interpretaciones.

Después de cada clase, se realiza un análisis de lo ocurrido allí, que guía la planeación de las clases siguientes y las modificaciones a la secuencia de instrucción. Así mismo, al finalizar el experimento de enseñanza, se lleva a cabo un análisis retrospectivo de todo lo sucedido, con énfasis en la mirada del aprendizaje de los estudiantes.

Se forman 13 grupos de 3 estudiantes cada uno, nombrados más adelante como G_1, G_2, \dots, G_{13} , en el interior de los cuales se trabaja en el desarrollo de las tareas propuestas en la secuencia de instrucción. Se realizan con frecuencia puestas en común y socializaciones guiadas por el profesor, que permiten el trabajo colaborativo en todo el curso. Los estudiantes utilizan el computador como herramienta y registran el trabajo elaborado

¹ Proyecto de investigación desarrollado durante 2017 por la línea de investigación en Educación Estadística, del Departamento de Matemáticas de la Universidad Pedagógica Nacional. Financiado por el Centro de Investigaciones (CIUP) de la misma Universidad.

por escrito. Adicionalmente, las interacciones de los grupos y las socializaciones globales se graban en audio y video.

La sistematización de la información surge de la selección de apartes de las transcripciones, que son revisados y discutidos hasta llegar a una interpretación consensuada por los investigadores a la luz del marco conceptual. Las respuestas y expresiones formuladas por los grupos constituyen la información que se organiza según elementos comunes que se referencian en el trabajo de los estudiantes, los cuales van conformando las categorías relacionadas con características de la variabilidad de una colección de datos. En definitiva, estas características se ven como elementos constitutivos de la noción de *distribución estadística* y componen la caracterización de la variabilidad que emerge en este proyecto.

Marco conceptual

Según Wild (2006), un posible origen de los problemas que enfrenta la educación en estadística puede ser el hecho de que algunos aspectos de la estadística son tan elementales, que a pesar de que las personas los usan al razonar, no los nombran o exploran. Los conceptos *variabilidad*² y *distribución* son unas de esas nociones que no son enunciadas, aunque sí son usadas de manera habitual dentro del campo de la estadística y la probabilidad; sin embargo, pero en su estudio no hay claridad sobre sus características, ni tampoco hay suficiente indagación, como lo sugieren Andrade et ál. (2017), aspectos ligados a la variabilidad como objeto o tema de estudio, pues en disertaciones doctorales han tenido una importancia relativamente menor (apenas 12 de 129 trabajos doctorales la han considerado).

Shaughnessy y Ciancetta (2001, citados por Canada, 2004) muestran que hay un acuerdo en que “la variabilidad es el fundamento del pensamiento estadístico, la razón más profunda de la existencia de la disciplina”. Autores como Moore (1998) y Slauson (2008) reafirman la importancia de que los estudiantes aprecien y comprendan la variabilidad como componente central del razonamiento estadístico, y de manera similar, Hassad (2007), Meletiou (2000), Porter (2001) y Zieffler (2006) alegan que la clave para el razonamiento estadístico es que los estudiantes reconozcan y aprecien la omnipresencia de la variabilidad en contextos diversos, y entiendan cómo es dicha variabilidad, cómo se cuantifica, y se explica con las herramientas estadísticas que permiten darle sentido y control.

La definición de Moore (1990) para la naturaleza del pensamiento estadístico incluye cinco elementos: la omnipresencia de la variabilidad en los procesos, la necesidad de datos sobre los procesos, el diseño de producción de datos con la variabilidad en mente, la cuantificación de la variabilidad, y la explicación de

2 En este artículo nos referimos siempre a *variabilidad* para aludir a las características de una entidad que cambian y son observables, y no necesariamente a la medición del cambio, siguiendo a Reading y Shaughnessy (2004), quienes distinguen así *variabilidad* de *variación*.

la variabilidad, que son corroborados por Pfannkuch y Wild (2004) al ubicar la variabilidad en el centro de los modelos de pensamiento estadístico de los profesionales que involucran notar, entender, evaluar críticamente y distinguir los diferentes tipos de variabilidad. La omnipresencia de la variabilidad es ratificada por la posibilidad de observarla y detectarla en todos los sistemas y entidades de la realidad (Wild y Pfannkuch, 1999).

En la mayoría de formas estadísticas de razonamiento sobre la variabilidad está presente la noción de *distribución*, la cual subyace a las grandes ideas de la estadística como la variabilidad y el muestreo (Bakker y Gravemeijer, 2004). Según Reading y Shaughnessy (2004), “sin variación no hay distribución”. Al respecto, Wild (2006) subraya que la comprensión de la noción de *distribución* inicia por la percepción de la naturaleza omnipresente de la variabilidad. Para este autor, la noción de distribución es, en su nivel más básico, “el patrón de variación en una variable; esta noción es el fundamento del trabajo estadístico que investiga desentraña y modela patrones de variación, con el fin de aprender de ellos y hacerlos predecibles” (p. 11). De esta manera, Wild (2006) establece la distribución como un lente para mirar la variabilidad; es decir, existe una estrecha relación entre la variabilidad de una colección de datos que constituye una variable y la distribución correspondiente³. Además, apunta este investigador que mirar la variación en los datos a través de la distribución conduce a la exploración de la naturaleza de la variación explicada e inexplicada.

3 Mellissinos et ál. (1997, citados en Canada, 2004), precisan la diferencia entre una colección de datos y una distribución indicando que “una colección de datos es una colección de medidas de una o más características de objetos o personas y una distribución es un atributo del conjunto de datos que comunica cómo las medidas en el conjunto de datos están distribuidas a lo largo de un rango de valores”.

Es usual que los estudiantes vean los datos como valores de cada caso, es decir, consideren las frecuencias de valores particulares, y así, la percepción se centre en las características de los casos de valores individuales. Agregan Konold et ál. (2014) que cuando la atención se centra en la visualización de valores similares a los de una unidad o categoría, los datos se pueden ver como clasificadores, donde usualmente se contempla solo la frecuencia de cada unidad o categoría. Bakker y Gravemeijer (2004) señalan la necesidad de que los estudiantes vean los datos como valores de una variable y no como una característica personal de alguien; es decir, que desconecten las medidas o datos de los sujetos u objetos y consideren los datos comparados con un contexto global de medidas, para así desarrollar una noción de distribución como una estructura organizada o entidad conceptual. Parafraseando a Wild (2006), se da un gran paso si se llega a pensar en los datos como un agregado, pues se pueden aprender lecciones más ampliamente aplicables como ver los patrones que se disciernen en el todo. En la misma dirección, Konold et ál. (2014) indican que, al ver los datos como un agregado, la unidad perceptiva es todo el lote de datos y se razona sobre las características de la colección de datos como un todo; la distribución es entendida como una entidad con características de forma, centro y dispersión, entre otras. De acuerdo con Porter (1986, citado en Konold et ál., 1997), enfocarse en dónde se concentran los datos de una colección y en cómo estos están distribuidos, además de permitir pronósticos individuales, tiene la ventaja de posibilitar el aventurar tendencias y propensiones del todo. Para Konold et ál. (1997), algunas de las dificultades que los estudiantes presentan al formular e interpretar argumentos estadísticos pueden originarse en no haber adoptado esa perspectiva estadística.

También Garfield y Ben-Zvi (2008) consideran necesaria la idea de centro para procurar un significado a los datos analizados en el estudio de la variabilidad. Bakker y Gravemeijer (2004) concretan esta idea y otros aspectos comunes de las colecciones de datos y de las distribuciones en cuatro grandes propiedades: centro, dispersión, densidad y asimetría, estas se presentan en la tabla 1.

Tabla 1. Entre los datos y la distribución

Distribución (entidad conceptual)			
Centro	Dispersión	Densidad	Asimetría
promedio, mediana, rango medio, ...	rango, distancias, desviación estándar, rango intercuartil, ...	frecuencia absoluta, frecuencia relativa, mayoría, cuartiles	Posición de la mayoría de datos
Datos (valores individuales)			

Fuente: Bakker y Gravemeijer (2004).

Estos autores anotan que en este esquema la perspectiva hacia arriba es usual en los estudiantes principiantes, quienes calculan, por ejemplo, la media, la mediana, el rango o los cuartiles, sin que reconozcan necesariamente tales medidas como de centro o representativas de un grupo. De hecho, Zawojewski y Shaughnessy (2000, citados por Bakker y Gravemeijer, 2004) afirman que los estudiantes precisan desarrollar una noción de *distribución* antes de que puedan realizar una selección de la medida apropiada. Por consiguiente, es necesario que desarrollen también la perspectiva hacia abajo y conciben la centralidad, la dispersión y la asimetría como características de una distribución, mirando los datos como una entidad conceptual que se organiza de manera estructurada.

Secuencia de instrucción propuesta

En la tabla 2 se exponen algunas de las tareas desarrolladas que fueron propuestas en la secuencia de instrucción, las cuales ilustran el tipo de trabajo planteado a los estudiantes para profesor de matemáticas. En esta secuencia se trabaja de manera informal la noción de *distribución* en coherencia con la idea de variabilidad mediante hojas de cálculo en Excel.

Tabla 2. Tareas secuencia de instrucción propuesta

Sesión 1
<p>En la actualidad la divulgación de algunas pautas de alimentación para mantenerse en forma, conocidas con el nombre de FIT, su alimentación se basa en la disminución de carbohidratos y el aumento de la ingesta de proteínas y grasas saludables. Si se desea hacer un estudio sobre las pautas de alimentación que siguen los estudiantes del DMA, mencione al menos tres variables relevantes para el estudio.</p> <p>Discutir las variables propuestas hasta acordar las variables relevantes: porciones de distintos alimentos, proteínas, carbohidratos, grasas que contiene cada porción.</p>

Sesión 2

Generar una encuesta y con ella recolectar información de 30 estudiantes, una vez se hayan recopilado los datos generar tablas de Excel con la información recogida. Considerar dos distribuciones (porciones de lácteos y porciones de verduras y frutas).

En Excel, considerar la cantidad de porciones de lácteos consumidas. Ordenar los datos.

Describir todo lo que observa sobre la variabilidad de los datos.

Señalar cómo se pueden agrupar los datos.

Compartir y discutir las características de la variabilidad observadas y las formas de agrupar los datos.

Sesión 3

En Excel, generar un gráfico de columnas para las distribuciones de porciones de lácteos y porciones de verduras y frutas. Con base en dicha gráfica, describir qué más observa sobre la variabilidad de los datos.

Discutir otras características de la variabilidad que se han observado, se destaca cómo están presentes las frecuencias.

Sesión 4

En Excel, convertir los datos de cantidades de porciones de lácteos de cada persona a datos de proteínas en gramos.

En Excel, sumar los datos de las proteínas que consume cada persona, y generar una columna para la variable *proteínas* con el total los datos. Ordenar los datos.

En Excel, generar la tabla de frecuencias y el "Histograma".

Describir todo lo que observa sobre la variabilidad de los datos, teniendo en cuenta lo que se ha discutido hasta el momento.

Discutir otras características de la variabilidad que se han observado.

Sesión 5

En Excel, convertir los datos de cantidades de porciones de lácteos de cada persona, a datos de carbohidratos en gramos.

En Excel, sumar los datos de carbohidratos de cada persona, y generar una columna para la variable *carbohidratos* con el total los datos. Ordenar los datos.

En Excel, generar la tabla de frecuencias y el "Histograma".

Describir todo lo que observa sobre la variabilidad de los datos en la gráfica.

Describir las similitudes y diferencias entre las gráficas de las distribuciones de las variables *proteínas* y *carbohidratos*.

Fuente: elaboración propia.

Resultados

Se presentan aquí, en primer lugar, las respuestas y expresiones formuladas por los estudiantes en los grupos de trabajo durante el desarrollo de las tareas propuestas, agrupadas según la característica de la variabilidad a la que aluden. En segundo lugar, se muestra la caracterización de la variabilidad ligada a la noción de *distribución* que resulta del trabajo en el proyecto.

Respuestas de los estudiantes en el desarrollo de la secuencia de instrucción

Surge en los estudiantes, en principio y ante la presentación de datos numéricos en una lista, la necesidad de organizar u ordenar los datos numéricos, como acción indispensable para observar la variabilidad en unos datos. La aproximación intuitiva a la variabilidad se concreta en la identificación de los valores

máximos y mínimos, del rango y de valores atípicos, según se ejemplifica en las declaraciones a continuación, en las que se identifican con “G” las intervenciones por grupos de estudiantes:

G2: El mayor valor de esta lista es 1713,5 gramos de proteínas y el menor es 73,5 gramos de proteínas.

G3: El número de porciones de consumo de carnes y huevos a la semana de las personas encuestadas oscilan entre 0 y 30 [...] dato menor 170. dato mayor 3995.

G9: Variabilidad de las proteínas: mínimo 73,5; máximo 1713,5 [...] se puede observar que existe un dato atípico que se encuentra alejado de los demás datos de la frecuencia, el dato atípico es 1713,5.

G12: Rango de la variable que va de 0 porciones a 30 porciones [...] el rango que presentan los datos $(3995-170) = 3825$.

Es de anotar, que a lo largo del trabajo el reconocimiento de los valores máximos, mínimos, del rango y de los valores atípicos, se manifiesta también a partir del manejo de otros registros de representación como el tabular y el gráfico. Así se evidencia en las descripciones siguientes:

G6: Es posible observar... el valor atípico (el menos repetido)... la mayor cantidad de proteínas que consume una persona es 1713,5 y la menor cantidad es 73,5, el dato atípico es 1713,5 gramos [...] no se observa un valor atípico. [...] la gráfica de proteínas representa un valor atípico, mientras que la de carbohidratos no.

G13: Se puede realizar una comparación entre la cantidad de proteínas que consume siendo: el más bajo 73,5; el más alto 1713,5 semanal [...] encontramos un valor máximo en cada una de ellas [gráficas] y valores mínimos, el valor máximo siempre corresponde al valor correspondiente o el rango donde aparece la persona que consume 30 porciones [...] se mantienen máximos y mínimos en las dos gráficas, que varían.

Cuando los datos se presentan como una lista, los estudiantes suelen identificar el valor de la variable que más se repite; en los otros registros de representación también es común que destaquen este valor, es decir, el valor correspondiente a la mayor frecuencia; algunos se refieren a dicho valor con el término *moda*. En algunos grupos, los estudiantes mencionan además al promedio y lo calculan en las diferentes representaciones, a veces de manera no convencional. Así se percibe en las afirmaciones expuestas enseguida:

G1: El dato que más se repite es la cantidad (de proteínas) 7.

G4: 7 es el valor de mayor tendencia.

G7: la moda de los datos es 7 [...]. La moda para estos valores fue 361,8 [...] el promedio fue 398,41 [...]. El promedio para la tabla 1 tuvo 26,3 mientras que para la segunda tabla fue de 13,8.

G9: Se observa que la mayor cantidad de personas consumen (6-8) porciones de alimentos [...]. Promedio: 398, 414 → 100 % media: 363 moda: 361,8.

En pocos grupos los estudiantes dan cuenta de la manera en que cambian los valores de forma cualitativa o cuantitativa, lo cual podría asociarse con percibir distancias entre ellos, según se lee de las siguientes expresiones:

G10: Observamos que la mayoría de valores cambian en décimas.

G12: El rango que presenta los datos (3995-170) = 3825 y como son 263 datos implica que hay bastante variabilidad de los datos.

Casi todos los grupos registran la variabilidad de las frecuencias tanto en forma cualitativa como cuantitativa, y pareciera que se refieren a la frecuencia como la variable de interés para el estudio, ya que sugieren agrupaciones sobre los datos por los valores de las frecuencias, como lo indican sus respuestas:

G1: Se pueden agrupar las variables que sólo tienen una repetición, los que tienen de 2 a 15 repeticiones, de 16 a 35 repeticiones [...] "cantidad de repeticiones" [determina grupos así] 1-14, 2-15, 16-35, 1 y 0 [...] los datos "cantidad de repetición" varían de 1 a 114.

G2: Se pueden agrupar por la cantidad de personas que consumen ciertas porciones [...] La variable es "cant-per" ya que lo que varía es la cantidad de personas que consumen en un intervalo determinado, porciones.

G4: Agrupar por cantidad de personas a la semana [0 -5), [5 - 10), [10 - 15), ...

G5: Las clases 170 ... , 3755, 9375 con frecuencia igual a 1. Las clases 409,0652, ..., 3038,75 y mayor, con frecuencia igual a 3.

G10: Reconocemos que la variable es número de personas contra cantidad de porciones.

En el trabajo con tablas de frecuencia, los estudiantes dan cuenta de la variabilidad relacionada con los intervalos o clases; se refieren a la cantidad de intervalos, a la longitud de ellos, a la igualdad o diferencia de tamaño entre intervalos, al orden de los intervalos, tal y como se aprecia en estos enunciados:

G6: Tiene intervalos de 100 en 100, [...] en la columna denominada "clase" (gramos por proteína) se generan intervalos de grupos de 55 gramos [...] en el punto 31 tiene más clases que las del punto 33 [...] son agrupaciones al azar a excepción del gráfico 1, que hace intervalos de 8 en 88.

G9: Cuando se hacen las gráficas con intervalos iguales y con intervalos diferentes podemos observar que las frecuencias se distribuyen de la misma manera [...] en las proteínas los intervalos son un poco más cortos. Podemos observar que si los intervalos son más pequeños la distribución se logra observar con más claridad.

G10: 102,5 es la diferencia entre cada dato, cada "clase" [...] los datos se dividieron en intervalos de 239,0625.

G13: Y la amplitud de los intervalos es de dos, pero en el último se toma de 21 a 30 [...] con la herramienta histograma genera un rango automático de alrededor de 100 por intervalo [...] 200 - (menos que 200), 300 - (200-300), 400 - (300-400).

En las referencias de los estudiantes a las frecuencias y a sus valores correspondientes, e incluso a las frecuencias como porcentajes, es posible ver el tratamiento de los datos como clasificadores en el sentido expuesto en el marco conceptual, como se muestra a continuación:

G2: El 43,3 % consume 7 porciones [...] el 3,8 % consume 10 porciones.

G11: La variable en cantidad de porciones (0) tiene una cantidad significativa de proporción [...] la mayoría de personas consumen 7 porciones de carnes y huevos por semana [...] la mayoría de personas está en el rango de (1-7) porciones de carne y huevos [...] un gran porcentaje de los encuestados está en un rango de 381 proteínas consumidas a la semana [...] la mayor parte de los encuestados está alrededor de 648,125 carbohidratos a 1365,3125 carbohidratos.

G12: Una persona consume 30 porciones de carne y huevo a la semana [...] la clase numerada con el 2 presenta la mayor frecuencia absoluta (173 personas que consumen entre 5 y 10 porciones) [...] las clases numeradas con 4, 5 y 6 presentan los menores consumos de porciones [...] el consumo mayor de proteínas se puede observar en la frecuencia mayor de su gráfica la cual corresponde a 109 personas y en cambio en los carbohidratos su bajo consumo se refleja en la frecuencia de 69 personas.

En el trabajo con la representación gráfica los estudiantes visualizan y expresan otras características de la variabilidad, que son más notorias a través de este registro. Así, establecen comparaciones entre las barras, describen la forma de las gráficas, se refieren a la variabilidad de los datos en relación con las frecuencias, es decir, a la variación conjunta de la variable y la frecuencia; al respecto, indican si está inclinada o no, si asciende o desciende, algunos utilizando términos del álgebra y cálculo como *crecimiento* y *decrecimiento*, según se percibe en estas anotaciones:

G2: Se puede observar el crecimiento y decrecimiento de las columnas de la gráfica [...] desde la clase 381 hasta 1611 la frecuencia decrece [...] desde la clase 73,5 hasta 381 la frecuencia crece [...] en la gráfica de carbohidratos se observan 5 intervalos de crecimiento y 4 de decrecimiento [...] en la gráfica de proteínas se observan 2 intervalos de crecimiento y 1 de decrecimiento.

G3: A mayor cantidad de clases mayor variación.

G8: Cuando tienen muchas clases se pueden ver los valores más bajos que tienen

G9: Podemos observar que, si los intervalos son más pequeños, la distribución se logra observar con más claridad; si los intervalos son muy pequeños se pierde la efectividad de la gráfica de tenerla condensada mientras que si [...] son todos intervalos muy grandes no se ve con claridad la variabilidad de los datos [...] en la gráfica de proteínas existe una curva más alta; también podemos observar que el decrecimiento es menos exponencial [...] cuando el último intervalo se acerca al último dato, la gráfica se comporta como una curva en la que se detalla más la curva y se ve el decrecimiento y crecimiento de los datos

En la referencia a la forma de la gráfica, también los estudiantes en varios grupos señalan los llanos, no necesariamente nombrados con este término, identificación que se percibe así mismo en el trabajo con las tablas, donde es común que ellos aludan a las frecuencias nulas o que son cero. Las descripciones enseñada ilustran esto:

G2: En la gráfica de proteínas se observan seis clases nulas de 17 clases en cambio en la de carbohidratos se ve una sola clase nula.

G3: Hay intervalos en los que la cantidad de personas que consumen ciertas cantidades de carne son cero [...] de 791 en adelante la frecuencia tiende a ser llana [...] desde 1956,9 en adelante hay una frecuencia "llana".

G5: Se ve un "vacío" provocado porque no hay personas con esa cantidad de proteínas.

G8: Cuando tiene muchas clases se pueden ver los valores más bajos que tienen.

G10: Desde 996 en adelante podemos notar que están los llanos en contraste con la otra que la podemos encontrar 3038,75 y 3277,8 [...] se observa una diferencia en cuanto a los "llanos".

Hay igualmente, indicios de reconocimiento de aglomeraciones en las gráficas, es decir, de cantidades numerosas de datos en valores cercanos de la variable, por parte de la mayoría de los grupos de estudiantes, como se puede ver en las afirmaciones a continuación:

G3: la mayoría de los datos están aglomerados en un rango de (176, 791) [...] los datos están aglomerados entre (596, 1955) [...] la mayoría de las personas tienen un consumo de carbohidrato entre 596,56 y 1729,37.

G4: La mayor cantidad de personas consumen de 5 a 10 porciones de carne y huevos a la semana.

G9: Se presenta una curva en el intervalo de la clase [405 a 1849] lo que permite decir que hay una alta concentración en estos intervalos.

De manera similar los estudiantes perciben en las gráficas, apuntamientos o picos asociados a los intervalos o clases correspondientes, según lo señalan en sus apreciaciones.

G1: El pico que se observa en esta gráfica está alrededor del cuarto intervalo [...] tiene más apuntamientos (intervalos muy altos) la gráfica de carbohidratos que la gráfica de proteínas [...] hay un pico en la clase (1049,6, 1276,2) [...] similitudes [de las gráficas] ambas tienen un pico, [...], antes del pico de las proteínas hay tres datos, mientras en la gráfica de carbohidratos hay cuatro datos antes del pico

G4: El histograma muestra y presenta la forma ascendente de los valores hasta el pico o la moda.

G5: Picos: Proteínas NO, Carbohidratos sí [...] en las tres gráficas de proteínas ninguna presenta picos.

Se evidencian algunas referencias a los datos como un todo o como un agregado, cuando en un grupo los estudiantes comparan partes con el todo:

G2: El 85,5 % de las personas consumen menos de 10 porciones.

G12: la mayoría de los datos de consumo de proteína están agrupados en porcentaje entre 60 % y 90 % [...] se observa que 114 personas consumen 7 porciones de carne y huevo a la semana, es decir el 43,4 % del total [...] la clase número 2 representa el 65 % de las personas.

Solo un grupo de estudiantes menciona medidas de dispersión específicas:

G10: Observamos que en torno a la varianza podemos decir que entre más intervalos se puede deducir más cómodamente la variabilidad de los datos.

Se perciben referencias a la simetría de la gráfica según se observa en las respuestas de más adelante. Algunos grupos de estudiantes emplean los términos formales, posiblemente porque ya han tomado cursos de probabilidad.

G1: [la gráfica] Es asimétrica.

G3: Hay mayor frecuencia en los intervalos anteriores al intervalo de mayor frecuencia (datos acumulados a la izquierda) derecha, hay mayor frecuencia en los intervalos siguientes al intervalo de mayor frecuencia (datos acumulados a la derecha).

G7: Encontramos que la gráfica es asimétrica.

G12: la gráfica de carbohidratos no es simétrica (asimétrica) / la gráfica de proteínas es asimétrica.

Hay, quizás, por parte de estudiantes de cursos de probabilidad previos, alusión a la asociación de la forma de la gráfica con el modelo teórico de la distribución normal:

G3: Tienen una distribución normal [...] también se puede ver que los datos se pueden agrupar como una distribución normal.

G5: Hay una distribución en forma de campana, se evidencia en la gráfica de proteínas, consideramos que esta distribución es normal.

Caracterización de la variabilidad ligada a la distribución

En la sistematización de las respuestas que se expone, se percibe que a pesar del lenguaje informal utilizado y de que el término *distribución* no aparece, los estudiantes notan y reafirman diversas características de la variabilidad, que a su vez son elementos constitutivos de la noción de *distribución*, lo que va consolidando y formalizando una comprensión de la variabilidad más completa y simultáneamente, una concepción de distribución más amplia.

Las características apuntadas por los estudiantes dan lugar a una caracterización de la variabilidad ligada a la noción de *distribución* que amplía la conceptualización propuesta por Bakker y Gravemeijer (2004). En la tabla 3 se presenta esta caracterización no solo con los aspectos indicados por estos autores, sino con otros que el trabajo en el proyecto ha permitido concretar para complementar el espectro.

En particular, la ampliación de la caracterización aporta una comprensión que trasciende de la mera identificación de propiedades o conceptos asociados a la distribución, a la de explicitar procesos asociados a estas; por ejemplo, no solo se identifica el rango o los valores atípicos como componentes de la dispersión, sino que se precisan acciones particulares como “identificar valores máximos y

mínimos” y “apreciar de manera general la variabilidad”, entre otras. Además, se ha incorporado la categoría de modelos teóricos, en donde se sugiere la conexión que se puede hacer entre una distribución empírica y una forma teórica como la asociada a una distribución normal.

Tabla 3. Características de la variabilidad asociadas a la noción de distribución

Dispersión
Ver la necesidad de organizar u ordenar los datos.
Identificar valores máximo y mínimo.
Determinar el rango.
Establecer relaciones cualitativas entre los datos, o relaciones cuantitativas como distancias.
Apreciar de manera general la variabilidad.
Reconocer valores atípicos.
Estimar varianza, desviación estándar.
Densidad
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.
Centro
Reconocer el valor de la frecuencia más alta, moda.
Calcular mediana, media (promedio).
Simetría y asimetría
Reconocer simetría o asimetría.
Modelos teóricos
Relacionar con modelos teóricos.

Fuente: elaboración propia.

Reflexiones finales

En el trabajo inicial de exploración intuitiva de colecciones de datos cuantitativos enteros, presentadas en listas, los estudiantes advierten con facilidad mínimos, máximos y rango; las modas son también objeto de su reconocimiento, especialmente cuando agrupan los datos en tablas. En este registro tabular, algunos hacen referencia a características de centro al referirse a promedios, no siempre calculados ortodoxamente. Aparte de esto, de que el lenguaje usado es en general de carácter informal sin necesariamente nombrar los términos convencionales (por ejemplo: “inicia en”, “finaliza en”, “el valor más alto”, “el valor de mayor tendencia”), y de que los valores de la frecuencia sean reiterativamente considerados por los estudiantes como una variable de interés para la situación, es decir, las agrupaciones sean elaboradas a partir del número de repeticiones de los datos (frecuencias), se aprecia aquí un primer acercamiento a razonar sobre las características de una distribución.

Cabe anotar que el trabajo siguiente con el computador, entre otras cosas, dirige a los estudiantes a dejar de lado las frecuencias como variables y a contemplar las colecciones de datos como las variables relevantes de estudio. Perciben entonces otras características de la variabilidad primordialmente asociadas de manera natural con la representación tabular y gráfica, que además son componentes esenciales de la noción de distribución, mostrando avances en su razonamiento. En el trabajo de generación de tablas, los estudiantes aprecian el sentido de las frecuencias y la conveniencia de la construcción de clases; aluden así mismo, a valores atípicos y a valores faltantes determinados por las frecuencias nulas, o valores de la variable que no existen; también identifican frecuencias relativas y porcentajes de valores particulares y de intervalos, que los llevan a hacer comparaciones de partes de los datos con el todo, lo que de alguna manera sugiere que ven la colección de datos como un agregado, e incluso a veces visualizan percentiles.

Sin embargo, en el trabajo de los estudiantes es más frecuente ver el tratamiento de los datos como clasificadores que como un agregado, en especial cuando se refieren a las frecuencias o porcentajes y a los datos correspondientes; usualmente se centran en un valor o intervalo de valores de la variable y en su frecuencia correspondiente, de manera individual, y no en los datos como medidas comparables con la globalidad de los datos. De acuerdo con Bakker y Gravemeijer (2004), esta percepción está más relacionada con el trabajo de la distribución enfocado en los datos y sus frecuencias, es decir en la distribución de frecuencias, mientras que tratar los datos como un todo o agregado se acerca más a abordar las distribuciones de probabilidad con sus modelos.

En la representación gráfica con la ayuda de la tecnología, las descripciones de los estudiantes dan muestras de razonar sobre características de la variabilidad aún más relevantes para una concepción de distribución, como referir ideas en

torno al crecimiento y decrecimiento de los valores de la variable, a aglomeraciones, a apuntamientos, a llanos, a la forma de la curva, a colas, a la simetría y, en ocasiones, a la asimetría; también dan cuenta en este registro de aspectos evidenciados en los otros registros de representación, como los valores extremos y atípicos, el rango, los llanos. De nuevo el lenguaje de los estudiantes no necesariamente incluye esos términos, y es muchas veces informal (por ejemplo: “no todas pero tampoco tantas”, “el valor más alto”, “el valor de mayor tendencia”, “valor muy alejado”, “la barra que se encuentra más alta o más baja”, “la barra más grande”, “la curva más alta”, “cambios bruscos”, “los picos son irregulares”, “hay un salto”, “cambio drástico de rango”, “se ve un vacío”, “la gráfica se condensa”, “tiene forma de campana”).

En general, el tránsito por diversos sistemas de representación permite que se caractericen aspectos de la variabilidad asociados a la noción de distribución, por cuanto cada representación (listas de datos, tablas y gráficos) revela diversos atributos de las distribuciones tratadas. En efecto, los procesos de transnumeración implicados en cambios de listas de datos a tablas hacen emerger de manera natural el máximo y el mínimo de una distribución, así como los valores de las variables con mayor frecuencia, es decir, la(s) moda(s); mientras que la transnumeración de lo tabular a lo gráfico hace visibilizar las tendencias de crecimiento o decrecimiento de una distribución y la posible existencia de simetría o sesgos.

Conclusiones

El objetivo principal de este proyecto fue conocer y promover el razonamiento en torno al concepto de *distribución estadística* para así avanzar en la conceptualización de esta noción. Al evaluar la secuencia de instrucción

propuesta en este experimento de enseñanza, se evidencia que, con los resultados, el trabajo de exploración informal desde lo intuitivo en situaciones de la vida cotidiana, en particular el contexto de la alimentación FIT y a través de tres registros de representación, lista de datos, tabular y gráfico, de los conjuntos de datos resultantes, contribuye a que los estudiantes razonen sobre las características de la variabilidad asociadas a la noción de distribución a medida que avanzan en las tareas.

En general, teniendo en cuenta la conceptualización de distribución ligada a la caracterización de la variabilidad resultante, se concluye que los estudiantes para profesor de matemáticas señalan elementos que tienen que ver primordialmente con la dispersión y la densidad, mientras que los elementos relativos al centro y a la asimetría son menos nombrados. La referencia a la moda, común entre los estudiantes, se percibe más cercana a los aspectos de la densidad que a los de centro, pues es motivada por la observación de las frecuencias más altas y no es usada en relación con delineaciones de la posible centralidad de la colección de datos.

El trabajo en este proyecto muestra que crear oportunidades de exploración en situaciones de la vida real, donde los estudiantes aportan sus ideas intuitivas de forma colaborativa, enriquece el proceso de aprendizaje. De manera similar, fue evidente que la posibilidad de usar su propio lenguaje informal no impide el razonamiento sobre el objeto estadístico en la transición hacia una conceptualización más formalizada. Es claro que conceptos de centralidad, como *media* y *mediana*; de dispersión, como *varianza* y *desviación*, y de posición, como *cuartiles* y en general *percentiles*, que apenas son vislumbrados por algunos estudiantes, requieren un tratamiento posterior al desarrollo intuitivo de una noción de *distribución*.

Considerar, por parte de los estudiantes, la frecuencia como variable de interés para la situación es un hecho detectado con regularidad por profesores de estadística en sus clases, que podría usarse como motivo de un primer acercamiento a razonar sobre las características de una distribución.

La caracterización de la variabilidad resultante recrea, amplia y concreta aspectos específicos que se pueden considerar en el trabajo de clase, para organizar de una manera más enfocada las ideas de la variabilidad discutidas y recomendadas en la comunidad de *educación estadística*. Esta caracterización puede aportar no solo al conocimiento del profesor, sino que puede constituirse en una herramienta para la enseñanza y el aprendizaje, en aras de impulsar una mejor conceptualización de la noción de *distribución*.

Referencias

- Andrade, L., Fernández, F. y Álvarez, I. (2017). Panorama de la investigación en educación estadística desde tesis doctorales 2000-2014. *Revista Tecné, Episteme y Didaxis: TED*, 41, 87-107.
- 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). Kluwer Academic Publishers.
- Canada, D. (2004). *Pre-service elementary teachers' conceptions of variability* [tesis de doctorado]. Portland State University, Portland, Oregón, Estados Unidos.
- Cobb, P. (2000). Conducting teaching experiments in collaboration with teachers. En A. Kelly y R. Lesh (eds.), *Handbook of research design in mathematics and science education* (pp. 307-333). Lawrence Erlbaum.
- Garfield, J. y Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Springer Science & Business Media.
- Hassad, R. (2007). *Development and validation of a scale for measuring instructors' attitudes toward concept-based or reform-oriented teaching of introductory statistics in the health and behavioral sciences*. [Tesis de doctorado, Universidad internacional de Touro].
- Konold, C., Higgins, T., Russell, S. y Khalil, K. (2014). Data seen through different lenses. *Educational Studies in Mathematics*, 88, 305-325. 10.1007/s10649-013-9529-8.
- Konold, C., Pollatsek, A., Well, A. y Gagnon, A. (1997). *Students analyzing data: research of critical barriers. Research on the role of technology in teaching and learning statistics*. International Statistical Institute Voorburg. <https://goo.gl/u80vZR>
- Meletiou, M. (2000). *Student understanding of variation. An untapped well in statistical reasoning*. [Tesis de doctorado, Universidad de Texas].

- Moore, D. (1990). Uncertainty. En L. A. Steen (ed.), *On the shoulders on giants: A new approach to numeracy* (pp. 95-137). National Academy of Sciences.
- Moore, D. (1998). Statistics among the liberal arts. *Journal of the American Statistical Association*, 93(444), 1253-1259.
- Pfannkuch, M. y Wild, C. (2004). Towards an understanding of statistical thinking. En J. Garfield y D. Ben-Zvi (eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 17-46). Kluwer Academic Publishers.
- Porter, A. (2001). *Improving statistical education through the experience of reflective practice*. [Tesis de doctorado, Universidad de Wollongong].
- Reading, C. y Shaughnessy, J.M. (2004). Reasoning about variation. En J. Garfield y D. BenZvi (eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 201-226). Kluwer Academic Publishers.
- Slauson, L. (2008). *Students' conceptual understanding of variability*. [Tesis de doctorado, Universidad de Ohio].
- Steffe, L.P., Thompson, P.W. y Von Glasersfeld, E. (2000). Teaching experiment methodology: Underlying principles and essential elements. En A. Kelly y R. Lesh (eds.), *Handbook of research design in mathematics and science education* (pp. 267-306). Routledge.
- Wild, C. J. (2006). The concept of distribution. *Statistics Education Research Journal*, 5(2), 10-26.
- Wild, C. J. y Pfannkuch, M. (1999). Statistical thinking in empirical enquire. *International Statistics Review*, 67, 23-265.
- Zieffler, A. (2006). *A longitudinal investigation of the development of college students' reasoning about bivariate data during an introductory statistics course*. [Tesis de doctorado, Universidad de Minnesota].

Para citar este artículo

Méndez Reina, M., Fernández Hernández, F. J., & Andrade Escobar, L. (2021). Exploración de la noción de distribución desde la variabilidad. *Tecné, Episteme y Didaxis: TED*, (49), 73-90. <https://doi.org/10.17227/ted.num49-9233>

