CONCEPTUALIZACIÓN DE MEDIDAS DE TENDENCIA CENTRAL Y DE DISPERSIÓN EN OCTAVO GRADO CON FATHOM

JUAN FRANCISCO MANTILLA GÓMEZ

UNIVERSIDAD INDUSTRIAL DE SANTANDER

FACULTAD DE CIENCIAS

ESCUELA DE MATEMÁTICAS

BUCARAMANGA

2014

CONCEPTUALIZACIÓN DE MEDIDAS DE TENDENCIA CENTRAL Y DE DISPERSIÓN EN OCTAVO GRADO CON FATHOM

JUAN FRANCISCO MANTILLA GÓMEZ

Trabajo de grado para optar al título de

Licenciado en Matemáticas

Director: GABRIEL YÁÑEZ CANAL

Doctor en Matemática Educativa

UNIVERSIDAD INDUSTRIAL DE SANTANDER

FACULTAD DE CIENCIAS

ESCUELA DE MATEMÁTICAS

BUCARAMANGA

2014

A Dios y a todas las personas que ha puesto en mi camino y que tanto quiero

Juan Francisco Mantilla Gómez

AGRADECIMIENTOS

El autor expresa su gratitud a:

Al profesor Gabriel Yáñez Canal, director del proyecto, por la orientación y gentileza brindada.

Al profesor Juan de Dios Urbina, docente de matemáticas de la Institución Educativa Las Américas de Bucaramanga, por su ayuda desinteresada.

A la Universidad Industrial de Santander, por los años de formación profesional dedicados.

A todas las personas que de alguna forma colaboraron con la elaboración del proyecto.

CONTENIDO

LISTA DE FIGURAS

RESUMEN

TÍTULO: CONCEPTUALIZACIÓN DE MEDIDAS DE TENDENCIA CENTRAL Y DE DISPERSIÓN EN OCTAVO GRADO CON FATHOM[*]

AUTOR: JUAN FRANCISCO MANTILLA GÓMEZ[**]

PALABRAS CLAVE: tendencia central, dispersión, software educativo, Fathom™.

DESCRIPCIÓN

Debido al creciente avance de las tecnologías y la comunicación, la estadística ha cobrado gran importancia dentro de las matemáticas y la sociedad, ya que permite el análisis de datos de distintas fuentes en diferentes ámbitos y para diversos fines. Los computadores y el desarrollo de software educativo han hecho que el acceso a la estadística sea cada vez más necesario y generalizado.

A la hora de describir, analizar o comparar conjuntos de datos se pueden llevar a cabo diferentes procedimientos tales como elaborar representaciones gráficas, encontrar el centro de los datos y medir qué tan dispersos están los datos respecto de ese centro. En un esfuerzo por relacionar los computadores y la enseñanza de la estadística se presentan en este trabajo algunas ideas sobre la forma como se puede usar el paquete estadístico Fathom para la enseñanza de conceptos básicos de la estadística como lo son las medidas de tendencia central y las medidas de dispersión, con el fin de generar una mayor comprensión de estos conceptos asociados a un conjunto de datos y de esta manera describir brevemente su distribución. Los estudiantes escogidos para esta investigación son de octavo grado de la Institución Educativa Las Américas de Bucaramanga.

[*] Trabajo de grado
[**] Facultad de Ciencias, Escuela de Matemáticas. Director: PhD. Gabriel Yáñez Canal

ABSTRACT

TITLE: CONCEPTUALIZATION OF MEASURES OF CENTRAL TENDENCY AND SPREAD IN EIGHTH GRADE WITH FATHOM*

AUTHOR: JUAN FRANCISCO MANTILLA GÓMEZ**

KEYWORDS: central tendency, spread, educational software, Fathom™.

DESCRIPTION

Due to the increasing advancement of technology and communication, statistics has become very important in mathematics and society, allowing data analysis in different areas and for different purposes. Computers and educational software development have made access to statistical increasingly necessary and widespread.

When describing, analyzing and comparing data sets can perform different procedures such as drawing graphs, try to find the center of the data and measure how scattered are data to that center. In an effort to connect computers and teaching statistics is proposed in this paper to give some ideas on how you can use statistical software Fathom to teach basic concepts of statistics such as measures of central tendency and spread, in order to generate a greater understanding of the concepts associated with a dataset with the aim to briefly describe their distribution. Students chosen for this research are eighth grade of school Las Américas of Bucaramanga.

* Graduate project

** Faculty of Sciences, School of Mathematics. Director: PhD. Gabriel Yáñez Canal

INTRODUCCIÓN

Aunque para los profesores parezca simple, conceptos elementales como las medidas de tendencia central y de dispersión en estadística resultan de gran complejidad para los estudiantes. Todo docente de matemáticas que trate de diseñar un proceso de enseñanza de estos conceptos debe ser consciente de las dificultades que pueden tener sus educandos para su comprensión; debe hacer un análisis epistemológico de su significado y planear situaciones que sugieran buscar un representante de un conjunto de datos y medir la dispersión de los datos con respecto a ese representante. La enseñanza de estos conceptos no puede reducirse solamente a procedimientos algorítmicos que están lejos de ser muestra de verdadera comprensión.

Para lograr que los estudiantes comprendan las medidas de tendencia central y de dispersión es necesario reflexionar sobre la complejidad de estas desde diferentes perspectivas y tener en cuenta que el aprendizaje de estos conceptos es un proceso que requiere tiempo y evoluciona lentamente. Como se dijo anteriormente, no basta con aprender procedimientos algorítmicos o definiciones, es necesario explorar problemas donde surgen estos conceptos y tener en cuenta los procedimientos que emplean los alumnos para resolverlos, aunque su forma de proceder no sea muy precisa; es muy importante que los alumnos trabajen con diferentes representaciones de estos conceptos y exploren y traten de argumentar el porqué de sus principales características.

Con el propósito de alcanzar todo lo mencionado anteriormente, la implementación de software educativo parece ser una buena alternativa, pues, además de que el desarrollo de los computadores facilita el acceso a la estadística, los contenidos de esta ciencia han tenido cambios importantes: la exploración, la interpretación y el análisis adquieren mayor protagonismo y pierden importancia los procedimientos meramente algorítmicos, al reducir el tiempo dedicado a cálculos tediosos. Fathom es un software diseñado para que los estudiantes exploren y analicen datos disponiendo de numerosas representaciones gráficas para aprender estadística, además de llevar a cabo cálculos en poco tiempo, cambiando cualitativamente la forma de enseñar los conceptos que son el objeto de estudio de esta investigación.

Este trabajo intenta dar respuesta a la pregunta de cómo utilizar software educativo para la enseñanza de conceptos básicos en estadística, específicamente analizar de qué manera se puede usar el paquete estadístico Fathom para lograr que los estudiantes de octavo grado comprendan y usen las medidas de tendencia central y de dispersión para describir e interpretar el comportamiento de conjuntos de datos.

Este proyecto se realizó con estudiantes de octavo grado del colegio Las Américas de Bucaramanga. Para analizar el impacto que tiene el uso de Fathom para conceptualizar las medidas de tendencia central y de dispersión, fue necesario en un primer momento elaborar una prueba diagnóstica para tener idea de sus conocimientos y concepciones previas; la segunda etapa del proceso fue el diseño y la aplicación de actividades para la conceptualización de estas medidas, llevando registro escrito de las respuestas de los estudiantes. Al terminar el proceso se aplicó una evaluación final para determinar el nivel de comprensión de los estudiantes para analizar cómo influyó el uso de Fathom en el proceso de enseñanza de los conceptos que son la materia de estudio de este trabajo.

JUSTIFICACIÓN

El creciente avance de las tecnologías y la comunicación desde el siglo pasado ha hecho indispensable su manejo en todos los ámbitos y en todas las profesiones. Debido a esta evolución todo individuo, independientemente del campo en que se desempeñe, debe estar manejar las tecnologías de la información y la comunicación (TIC); esto genera un gran desafío en la educación ya que hace necesario realizar cambios, integrar las TIC en los procesos de enseñanza y aprendizaje e inculcar en los alumnos el análisis de información.

La estadística ha cobrado importancia dentro de las matemáticas y la sociedad, ya que describe y predice características de datos para diversos fines. Los computadores y el desarrollo de software educativo han permitido el acceso a la estadística y se han usado con el objetivo de facilitar los procesos de enseñanza y aprendizaje de la misma. Un ejemplo de ello es el paquete didáctico Fathom, que proporciona un entorno dinámico y visualmente atractivo que permite a los estudiantes explorar, analizar y modelar datos.

Cuando se quiere describir, analizar o comparar conjuntos de datos se usan diferentes métodos como elaborar gráficas; otra forma de hacerlo es tratando de encontrar el centro de los datos y medir qué tan dispersos están los datos de ese centro. Para esto se usan las medidas de tendencia central y las medidas de dispersión, que son conceptos básicos dentro de la estadística.

A pesar de la aparente simplicidad de las medidas de tendencia central y las medidas de dispersión, son conceptos que resultan muy complejos para los estudiantes, en particular, cuando se les pide encontrar un representante en un conjunto de datos. Un error a la hora de enseñar estas medidas o cualquier concepto de esta ciencia es pensar que llevar a cabo procedimientos algorítmicos es muestra de comprensión de su significado por parte de los alumnos. Para este problema Fathom puede ser una herramienta muy poderosa pues con este es posible analizar conjuntos de datos con gran variedad de representaciones y permite trabajar con los datos de forma dinámica, siendo un medio para fomentar la intuición y la exploración, permitiendo que los estudiantes conjeturen propiedades de estas medidas en las representaciones gráficas. Además de esto realiza cálculos rápidamente, haciendo que no se dedique tanto tiempo en procedimientos algorítmicos.

En un intento por relacionar los computadores y la enseñanza de la estadística, se propone en este trabajo dar algunas respuestas a la siguiente pregunta:

¿De qué manera se puede usar el paquete estadístico Fathom para lograr que los estudiantes de octavo comprendan y usen las medidas de tendencia central y las medidas de dispersión para describir e interpretar el comportamiento de un conjunto de datos?

Se pretende en este trabajo dar algunas ideas de la forma como se podría utilizar el paquete Fathom para generar una mayor comprensión de las medidas de tendencia central y de dispersión asociadas a un conjunto de datos, con el ánimo de describir brevemente su distribución. Los estudiantes escogidos para esta investigación son de octavo grado de la Institución Educativa Las Américas.

1. ANTECEDENTES

Para lograr una verdadera comprensión de las medidas de tendencia central y de dispersión por parte de los alumnos se debe reflexionar sobre la complejidad de estos conceptos. Más importante que los estudiantes conozcan las definiciones de estas medidas, o sepan calcularlas, es que reconozcan situaciones relacionadas con ellas y que sepan interpretar sus valores.

Para la realización de este trabajo se tienen en cuenta algunas ideas de Batanero (2000, 2001) en sus textos "Significado y Comprensión de las Medidas de Posición Central" y "Didáctica de la Estadística". La autora afirma que el significado de la media y de las medidas de tendencia central tiene un carácter complejo e identifica en el segundo los siguientes tipos de elementos que deben tenerse en cuenta para realizar un análisis de los conceptos en estudio, incluso cuando se analizan las respuestas de los estudiantes (p.4):

> *Elementos extensivos*: campo de problemas donde surge el concepto.
> *Elementos actuativos*: las prácticas empleadas en la solución de problemas.
> *Elementos ostensivos*: notaciones, gráficos, palabras y en general todas las representaciones del objeto abstracto.
> *Elementos intensivos*: definiciones y propiedades características y sus relaciones con otros conceptos.
> *Elementos validativos*: demostraciones que se emplean para probar las propiedades del concepto y que llegan a formar parte de su significado y los argumentos que se emplean para mostrar a otras personas la solución de problemas.

La autora en el primer texto hace una síntesis de diferentes investigaciones en torno a la enseñanza de las medidas de tendencia central, analizando diferentes dificultades que se pueden encontrar en el aprendizaje de estas medidas. El aprendizaje de una persona es un proceso lento y va evolucionando de forma similar a la construcción de objetos en la ciencia.

Batanero (2001, p. 90) también identifica algunas dificultades en la comprensión de las medidas de tendencia central teniendo en cuenta los elementos anteriores y afirma:

> La comprensión de un concepto no puede reducirse a conocer las definiciones y propiedades (elementos intensivos), sino a reconocer los problemas donde debe emplearse el concepto (elementos extensivos), las notaciones y palabras con que lo denotamos y en general todas sus representaciones (elementos ostensivos), habilidad operatoria en los diferentes algoritmos y procedimientos relacionados con el concepto (elementos actuativos) y capacidad de argumentar y justificar propiedades, relaciones y soluciones de problemas (elementos validativos).

Entre las conclusiones que da la autora se destacan las siguientes (Batanero, 2000, p.10-11):

- Las medidas de tendencia central tienen un significado complejo y por lo tanto es necesario un periodo dilatado de enseñanza a lo largo de la educación primaria y secundaria para lograr el progresivo acoplamiento de los significados personales que construyen los alumnos a los significados institucionales que se pretende que adquieran.
- Tradicionalmente, en el aprendizaje de la estadística, se ha dado una gran importancia al cálculo y a los aspectos actuativos, que ahora pierden importancia, debido a las nuevas tecnologías. En lugar de tener que ejercitarse en la realización con lápiz y papel de cálculos y gráficos, el alumno debe aprender el uso de calculadoras gráficas y programas de computador, como la hoja de cálculo. Las nuevas tecnologías introducen también nuevos elementos actuativos y ostensivos, ya que el rango de representaciones es mucho mayor. Permiten también plantear situaciones

de aprendizaje en las que el alumno se enfrente a problemas más reales cuya solución requiera el uso y aprendizaje de conceptos estadísticos. Estas situaciones requieren también el trabajo cooperativo, motivan el interés del alumno y le permiten explorar tanto los datos, como los conceptos implicados, reforzando los elementos intensivos y validativos.

- La enseñanza de la estadística en la escuela usando computadores requiere una planificación cuidadosa. Frecuentemente los datos reales son demasiado complejos y es necesario tomar versiones simplificadas de los conjuntos de datos. Si se desea mostrar cierta propiedad, será preciso a veces manipular el conjunto de datos, para, por ejemplo, conseguir que el valor de la media, la mediana y la moda sean marcadamente diferentes.
- El trabajo con computador debe ser complementado con otras situaciones encaminadas a que el alumno se familiarice con los campos de los problemas, las representaciones, tipos de prácticas y propiedades de los promedios, y que ejercite su capacidad de argumentación.

En "Didáctica de la Estadística" la autora nombra tres tipos diferentes de capacidades, según Mokros y Russell (1991), necesarios para la comprensión de la idea de valor típico:

1. Dado un conjunto de datos, comprender la necesidad de emplear un valor central, y elegir el más adecuado.
2. Construir un conjunto de datos que tenga un promedio dado.
3. Comprender el efecto que, sobre los promedios (media, mediana o moda), tiene un cambio en todos los datos o parte de ellos.

Batanero habla sobre la importancia de los computadores en el desarrollo de la estadística, ya que ha permitido facilitar el acceso a esta, aumentando la demanda en formación básica en estadística; ha aumentado los contenidos a enseñar, incluyendo el uso adecuado de software y manejo de información. Además, los contenidos en estadística han tenido un gran cambio debido a la difusión de computadores y tecnologías, donde ahora se presta mayor importancia a los aspectos interpretativos y conceptuales que a los procedimentales y algoritmos. El uso del computador reduce el tiempo dedicado a cálculos, facilitando el estudio de conjuntos de gran cantidad de datos *“No tiene pues, sentido, hacer perder el tiempo a los alumnos ocupándoles en repetir una y otra vez cálculos tediosos para intentar aumentar sus destrezas en cálculo, sino que es preferible dedicar ese tiempo a actividades interpretativas y a la resolución de problemas”* (2001, p.148).

La autora distingue los siguientes tipos de software para la enseñanza de la estadística (2001, 138-139):

> *Paquetes estadísticos profesionales*, como por ejemplo *SPSS* que tienen una amplia disponibilidad de presentación gráfica y numérica.
> *Software didáctico*, como *Fathom* que es un medio de aprendizaje para análisis exploratorio de datos y se utiliza en secundaria y en cursos introductorios de estadística a nivel de bachillerato.
> *Software de uso general*, como las hojas de cálculo, por ejemplo *EXCEL*.
> *Tutoriales*, que son programas desarrollados para enseñar a los estudiantes sobre habilidades estadísticas específicas o evaluar su conocimiento.
> *Software en Internet*, material que puede obtenerse “on-line”.

Los computadores y los paquetes estadísticos como Fathom permiten que los estudiantes exploren y analicen datos de forma dinámica con gran cantidad de representaciones gráficas, además de realizar cálculos que serían muy tediosos para los estudiantes en poco tiempo.

A la hora de describir y analizar un conjunto de datos no basta solamente encontrar un centro o representante de esos datos. También es importante explorar y medir qué tan cerca o alejados se encuentran los datos de ese centro, medir su dispersión. Para explorar un conjunto de datos o compararlo con otros, medir la dispersión es tan importante como encontrar un representante de los datos.

Para la elaboración de secuencias didácticas para la conceptualización de medidas de tendencia central y de dispersión es necesario reconocer su complejidad; también es importante tener en cuenta su fundamento epistemológico, cómo surgieron y se desarrollaron, en qué situaciones están inmersos. La estadística es una ciencia que siempre está relacionada con un contexto, por ello los procesos de enseñanza de esta ciencia deben enfocarse en la resolución de problemas prácticos, en un ambiente de exploración donde se fomente la intuición y la experimentación.

2. OBJETIVOS

2.1 OBJETIVO GENERAL

Analizar el impacto que tiene el uso del paquete estadístico Fathom en la conceptualización de medidas de tendencia central y medidas de dispersión en estudiantes de octavo grado.

2.2 OBJETIVOS ESPECÍFICOS

- Elaborar una prueba diagnóstica a los estudiantes para tener idea sobre sus conocimientos, concepciones e intuiciones de las medidas de tendencia central y las medidas de dispersión.
- Diseñar las actividades en Fathom que se aplicarán a los estudiantes.
- Elaborar y aplicar una evaluación final que dé cuenta del grado de conceptualización logrado por los estudiantes de las medidas de tendencia central y de dispersión.

3. MARCO TEÓRICO

Cuando se va a analizar un conjunto de datos, un primer paso que es realizar un análisis descriptivo y elaborar representaciones gráficas; para ir más allá, se puede tratar de encontrar el centro de los datos y medir qué tan dispersos están los datos de ese centro, es decir, se usan las medidas de tendencia central y las medidas de dispersión. Las medidas de tendencia central son los valores alrededor de los cuales se agrupan los datos y las medidas de dispersión son una medida del alejamiento de los datos con respecto a las medidas de tendencia central.

3.1 MEDIDAS DE TENDENCIA CENTRAL

Cuando se pretende resumir datos numéricamente, la característica básica es el “centro” de los datos, que es una medida del “tamaño típico” de las observaciones. Existen varias formas de medir esto; las más usadas comúnmente son la *media*, la *mediana* y la *moda*. Estas medidas, que se conocen como medidas de tendencia central, son los valores alrededor de los cuales se agrupan los datos.

3.1.1 La media. La media de una muestra es el promedio aritmético de las observaciones.

$$Media = \frac{Suma\ de\ las\ observaciones}{Número\ de\ observaciones}$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Algunas propiedades de la media son:

- La media es un valor comprendido entre los extremos de la distribución.

$$x_{mín} \leq \bar{x} \leq x_{máx}$$

- La media del producto de una constante a por una variable X es igual al producto de la constante por la media de la variable dada. Es decir, si se efectúa un cambio de unidad de medida a los datos (por ejemplo de metros a centímetros), la media queda afectada por dicho cambio de escala.

$$\frac{\sum ax_i}{n} = \frac{a\sum x_i}{n} = a\bar{x}$$

- La media de la suma de una constante entera a con una variable X es igual a la suma de la constante con la media de la variable dada. O sea, al efectuar un cambio en el origen desde el que se han medido los datos, la media queda afectada por dicho cambio de origen.

$$\frac{\sum(a + x_i)}{n} = \frac{na + \sum x_i}{n} = a + \bar{x}$$

- La media no tiene por qué ser igual a uno de los valores de los datos, ni siquiera de su misma naturaleza: datos enteros pueden tener una media decimal.
- La media es un representante de los datos a partir de los que ha sido calculada, es decir, es un número que distingue un grupo de datos de otros (aunque es importante tener en cuenta medidas de dispersión para diferenciar grupos de datos con la misma media).
- La media se ve influenciada por la presencia de valores extremos.

3.1.2 La mediana. Afonso (2000, p.47) define la mediana de la siguiente manera:

La mediana de una muestra es el valor que ocupa la posición central de la lista, cuando los valores están colocados en orden creciente o decreciente e incluyendo también los valores repetidos, individualmente, en lista ordenada. La mediana de la muestra divide al conjunto total en dos partes iguales, con la mitad de los valores sobre la mediana de la muestra y la otra mitad debajo de ella. La mediana de la muestra puede no pertenecer a un conjunto original de valores.

Cuando la cantidad de valores es impar, la mediana de la muestra es el valor que ocupa la posición central, la cual, es única.

Cuando la cantidad de valores es par, hay dos posiciones centrales en la lista ordenada; entonces, la mediana de la muestra es la media aritmética de esos dos valores que ocupan las posiciones centrales.

3.1.3 La moda. Cuando la variable es cualitativa no se puede calcular la media o la mediana. Para describir un grupo se puede entonces usar la moda, que es el valor de la variable que tiene mayor frecuencia. En una distribución puede haber más de una moda. También es posible calcular la moda de variables numéricas.

3.2 MEDIDAS DE DISPERSIÓN

Aparte de conocer los valores típicos de una distribución, es decir, los valores que caracterizan al conjunto de datos, también es importante conocer qué tan dispersos están los datos que se tienen para describir sus características más importantes. Las medidas de dispersión proporcionan una medida de la desviación de los datos alrededor de una medida de tendencia central.

Las medidas de tendencia central indican los valores alrededor de los cuales se distribuyen los datos. Las medidas de dispersión son estadísticos que proporcionan una medida del mayor o menor agrupamiento de los datos respecto a los valores de tendencia central. Todas ellas son valores mayores o iguales a cero, indicando un valor 0 la ausencia de dispersión.

3.2.1 El rango. El rango es la medida más simple de qué tan dispersos están los datos, siendo simplemente la diferencia entre el valor más grande y el valor más pequeño.

$$Rango = Observación\ mayor - Observación\ menor$$

3.2.2 La desviación media. Una primera medida de dispersión es la desviación media que puede calcularse con respecto a cada uno de los valores centrales (la media, la mediana o la moda). Se define como la media de las desviaciones respecto al valor central que se considere, tomadas en valor absoluto. Se calcula de la siguiente forma:

$$D_c = \frac{\sum |x_i - c|}{n}$$

Donde c es el valor central.

3.2.3 La varianza de la muestra. Wackerly, Mendenhall y Scheaffer (2010, p.10) definen la varianza de una muestra de mediciones $x_1, x_2, \ldots, x_n$ como la suma de los cuadrados de las diferencias entre las mediciones y su media, dividida entre $n - 1$.

$$var = sd^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

3.2.4 La desviación estándar de la muestra. La desviación estándar es probablemente la medida de dispersión más popular, es la raíz cuadrada positiva de la varianza. Aunque esta medida no es fácil de interpretar, juega un papel fundamental en inferencia estadística. La fórmula para la desviación estándar es:

$$sd = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n - 1}} = \sqrt{\frac{1}{n - 1} \sum (x_i - \bar{x})^2}$$

3.3 PERTINENCIA DEL ESTUDIO A PARTIR DE LOS ESTÁNDARES Y LINEAMIENTOS CURRICULARES EN MATEMÁTICAS

3.3.1 Pensamiento aleatorio. Este tipo de pensamiento ayuda a tomar decisiones en situaciones de incertidumbre, azar, riesgo o ambigüedad por falta de información confiable, donde no se puede predecir con certeza un resultado final. En los lineamientos curriculares del Ministerio de Educación Nacional (MEN) establecen que el desarrollo del pensamiento aleatorio, por medio de contenidos de la probabilidad y la estadística debe estar infundido de espíritu de exploración y de investigación tanto por parte de los estudiantes como de los docentes. Deben estar presentes la comparación y evaluación de diferentes formas de aproximación a los problemas con el objeto de hacer seguimiento posibles concepciones y representaciones erradas (MEN, 1998, p.47). El desarrollo del pensamiento aleatorio se da en la resolución de problemas.

La enseñanza de las matemáticas muchas veces se centra en buscar una respuesta correcta y única. La aparición de la estadística en la clase de matemáticas permite sobre un conjunto de datos proponer inferencias e interpretaciones. Para la enseñanza de conceptos, representaciones y métodos estadísticos como teoría matemática, tiene que tenerse mucho cuidado para que esto no cause la pérdida de su carácter aleatorio. En los cursos de educación básica las representaciones gráficas permiten captar la aleatoriedad y la incertidumbre en forma cuantitativa y cualitativa, sobre los cuales los estudiantes pueden hacer evaluaciones y tomar decisiones sin recurrir a un esquema único de cálculo que los llevaría a encontrar valores deterministas definidos (MEN, 1998, p.48).

El proyecto del Consejo Escolar de Educación Estadística (Holmes, 1980, p.11-16) presenta tres principios para tener en cuenta a la hora de introducir conceptos estadísticos:

- Los conceptos y las técnicas deben introducirse dentro de un contexto práctico.
- No es necesario desarrollar completamente las técnicas en el momento que se presentan por primera vez.
- No es necesario ni deseable una justificación teórica completa de todos los temas, algunos de ellos se tratarán dentro de un problema particular, otros se considerarán mediante experiencias y no se justificarán teóricamente.

Los docentes deben preparar y utilizar situaciones abiertas en el marco aleatorio y estadístico, susceptibles de cambios y resultados inesperados e imprevisibles.

Para la elaboración de la prueba diagnóstica se tuvieron en cuenta los estándares de pensamiento aleatorio dispuestos por el MEN de los grados sexto a séptimo y octavo a noveno.

3.3.2 Sexto a séptimo. Para el estudio que se llevó a cabo con los estudiantes se tuvieron en cuenta los siguientes estándares (MEN, 2006, p.85)

- Comparar e interpretar datos provenientes de diversas fuentes.
- Reconocer la relación entre un conjunto de datos y su representación.
- Usar medidas de tendencia central para interpretar comportamiento de un conjunto de datos.
- Resolver y formular problemas a partir de un conjunto de datos y su representación.

3.3.3 Octavo a noveno. Se tuvieron en cuenta para el estudio los siguientes estándares (MEN, 2006, p.87):

- Interpretar analítica y críticamente información estadística proveniente de diversas fuentes.
- Interpretar y utilizar conceptos de media, mediana y moda y explicito sus diferencias en distribuciones de distinta dispersión y asimetría.

- Resolver y formular problemas seleccionando información relevante en conjuntos de datos.

En los lineamientos y estándares del MEN se da importancia al uso de medidas de tendencia central y de dispersión para el manejo de información de diversas fuentes. En el estudio que se hizo con los estudiantes de octavo grado de la Institución Educativa Las Américas de Bucaramanga se pretende que ellos conozcan las principales características de las medidas de tendencia central y de dispersión de forma visual y física por medio de las representaciones que ofrece el paquete estadístico Fathom en situaciones prácticas. No se pretende llegar a justificaciones teóricas de las propiedades de estas medidas, sino que se verán dentro de problemas particulares.

4. METODOLOGÍA

Las actividades se desarrollaron con un grupo de estudiantes de octavo grado de la Institución Educativa Las Américas de Bucaramanga, en la jornada de la mañana, a cargo de un estudiante de la Licenciatura en Matemáticas de la Universidad Industrial de Santander. Se realizaron en la sala de matemáticas del colegio dentro del horario de clases. Los registros de cada actividad se hicieron de forma escrita.

4.1 ETAPAS DEL PROCESO

4.1.1 Diagnóstico. En esta etapa se aplicó un cuestionario a los estudiantes que permitiera tener idea de los conocimientos, concepciones e intuiciones que tienen sobre las medidas de tendencia central y las medidas de dispersión. Antes de diseñar y desarrollar las actividades para la conceptualización de los conceptos a tratar, es necesario saber en qué nivel se encontraban los estudiantes para resolver problemas donde se requiere elegir un representante de un grupo y medir la variabilidad de los datos.

4.1.2 Instrucción. En esta etapa se hizo el diseño y la aplicación de actividades a los estudiantes para la conceptualización de las medidas de tendencia central y de dispersión. Se diseñó una actividad por semana y se trazaron para cada una objetivos o metas que se esperaban de los alumnos en la clase. La realización de los talleres con los estudiantes se hizo en la sala de matemáticas de la institución educativa (que cuenta con computadores y un video-beam) usando Fathom, registrando las respuestas de los estudiantes de forma escrita.

4.1.3 Evaluación final. Al terminar la etapa de instrucción se hizo una valoración del grado de comprensión de los estudiantes sobre las medidas de tendencia central y de dispersión por medio de una evaluación, para analizar las ventajas y desventajas que ofrece el uso de Fathom en la enseñanza y el aprendizaje de estos conceptos en estudiantes de octavo grado.

El análisis y elaboración del informe final se hizo a partir de los resultados obtenidos en las etapas descritas anteriormente. Se llevó registro escrito de las respuestas y reacciones de los estudiantes en el diagnóstico y a la hora de tratar de resolver los problemas propuestos en los talleres usando algunas herramientas de Fathom. Se analizó cada una de las actividades y se comparó las respuestas de los estudiantes con los elementos tenidos en cuenta en los antecedentes para la comprensión de las medidas de tendencia central y de dispersión.

4.2 RECURSOS

4.2.1 Fathom. Fathom es un software diseñado para cursos de secundaria y universidad con el objetivo que los estudiantes reúnan, exploren y analicen datos con gran cantidad de representaciones gráficas para aprender estadística y algunos conceptos o procedimientos matemáticos en general.

El análisis de datos y la modelación son ahora componentes integrales de los cursos de secundaria. Fathom proporciona un entorno visualmente atractivo dinámico para que los estudiantes exploren, analicen y modelen datos. Fathom puede ser usado para representar rápidamente datos con variedad de gráficos (Key Curriculum Press).

Fathom fue elaborado por KPC technologies, una compañía de investigación de software afiliada a Key Curriculum Press cuyo objetivo principal es el de desarrollar aplicaciones de software para cambiar cualitativamente la dinámica de las matemáticas y las clases de ciencias mediante el apoyo de materiales de alta calidad que ayudan a mejorar el desarrollo conceptual (Alvarado & Sánchez, 2010).

5. DIAGNÓSTICO

Antes de diseñar y desarrollar actividades para la conceptualización de las medidas de tendencia central y las medidas de dispersión, es necesario saber en qué nivel se encuentran los alumnos, conocer los conceptos previos que tengan, sus intuiciones para resolver problemas donde se requiere elegir un representante de un grupo. Estas medidas estadísticas tienen muchas aplicaciones y son un concepto que está muy lejos de ser simple como parece.

Según el profesor de la clase de matemáticas del grupo 8-05 de la Institución Educativa Las Américas, los estudiantes hasta el momento no han trabajado medidas de tendencia central ni medidas de dispersión.

Para conocer las concepciones previas que puedan tener los alumnos se hizo un cuestionario con problemas diversos, con el fin de determinar cómo resuelven estos alumnos problemas de encontrar un representante y medir la variabilidad en un conjunto de datos. Gran parte de los problemas propuestos son adaptaciones de problemas que ya propuso Batanero (Universidad de Granada, 2000) en sus investigaciones.

A continuación se presenta los problemas del cuestionario con las conclusiones que se obtuvieron al analizar las respuestas de los alumnos.

Problema 5.1. *Un objeto se pesa con una misma balanza por ocho estudiantes, obteniéndose los siguientes valores en gramos:*

62 60 60 63 61 62,3 61,5 62

¿Cuál es el peso real del objeto?

Para resolver este problema la mayoría elegía un número que estuviera entre el valor más pequeño 60 y el valor más grande 63, unos pocos dijeron que el peso real del objeto está por fuera de ese intervalo.

La mayor parte de los estudiantes dijo que el peso real del objeto debía ser 60 o 62 pues estos valores se repiten, además algunos de ellos afirmaban que el peso real es 60 porque es este valor se repite en dos casos consecutivos.

“60 porque él es el que se repite consecutivamente.”

“60 porque es el número con más probabilidad porque hay dos números iguales. 60 y 60, sería lo más probable y además coinciden.”

Los estudiantes pensaron en la moda como mejor representante pero, a pesar de que hubiera dos modas, el hecho de que un mismo dato estuviera dos veces consecutivas les hizo pensar que ese dato fuera el representante del grupo.

Algunas de las respuestas son muy particulares. Hay estudiantes que conocen el procedimiento algebraico para hallar el promedio, y trataron de obtener una respuesta. Otros sin embargo señalan que el peso real del objeto debe ser 61,5 porque es el valor del conjunto de datos que más cerca está a la media.

“61,5 porque al sumar y dividir todo da 61,475 y 61,5 es el más cerca.”

Esta respuesta muestra que algunos alumnos tienden a pensar que el representante de los datos tiene que ser necesariamente uno de los del conjunto. Para este estudiante el resultado que obtuvo mediante su procedimiento no tiene sentido pues no es uno de los datos que se le dan y por eso elige el valor más cercano al valor que encontró.

La mayoría daba su respuesta afirmando que esta es el “promedio” de los datos, lo que muestra que en algún momento han oído hablar de ese término.

Problema 5.2. *La altura media de los alumnos de un colegio es 140cm. Si se toma una muestra aleatoria de 5 estudiantes y resulta que la altura de los 4 primeros es 138, 142, 160 y 140. ¿Cuál sería la altura más probable del quinto estudiante?*

A diferencia de Batanero en sus trabajos, quien afirma que generalmente los alumnos buscan un valor de la estatura del quinto sujeto tal que sumada a las cuatro alturas anteriores dé una media de 140, ninguno de los estudiantes del curso hizo este procedimiento. Muchos alumnos dijeron que la altura debe ser 145 porque es el promedio de las cuatro medidas dadas. De nuevo afirmaban que su respuesta es un “promedio”.

Algunos dieron como respuesta 144 o 162 pues dos de los datos, 140 y 142, son 2cm más grandes que el dato anterior al ordenarlos en orden ascendente.

“144 porque cada uno tiene 2cm de diferencia y después de 142 va a 144 y luego 160.”

“162 porque miro el intervalo en cual están las medidas que van de dos en dos...”

La forma en que está redactado el problema tal vez hace pensar a los estudiantes que los datos están dados en forma de sucesión cuando ordenan los datos, después tratan de completar el término que creen que hace falta en esa ‘secuencia’.

Algunos pocos alumnos conciben la media como un representante de una población.

“Es probable que sea 140 porque es una altura media y es respectiva al colegio.”

Pocos estudiantes encuentran esa media dada como el valor representativo de una población y como el valor más probable.

Problema 5.3. *La nota final de una materia es el promedio de la calificación de 4 evaluaciones. Las notas van de 0 a 5 y el mínimo para aprobar es 3,5. Si un estudiante obtiene en las primeras tres evaluaciones: 4,0; 2,8 y 3,9. ¿Cuál es la nota mínima que debe obtener en la última evaluación para aprobar la materia?*

Algunos alumnos dieron una respuesta cualquiera sin dar una justificación, o daban una nota alta diciendo que si obtenían esa nota aseguraban la nota y no corrían ningún riesgo de perderla.

Otros obtuvieron el promedio de las primeras tres notas y afirmaron que ese debía la última nota. Incluso algunos dijeron que el promedio de las tres primeras notas iba a ser la nota final, no respondieron a la pregunta que se planteó.

Muy pocos dieron la respuesta correcta, por lo que se puede pensar que unos pocos alumnos conocen realmente el procedimiento algebraico para hallar la media aritmética.

La mayoría de los alumnos respondió que se necesita 3,5 o 3,4. Esto lo dedujeron por ensayo y error. Al preguntarles a estos alumnos cómo hallaron sus resultados, algunas de las respuestas fueron:

"Yo lo que hice fue que dieron cuatro notas. Tocaría sumar las tres notas y las dividí en tres, me dio un número, 2,6, le sumé 3,4 y dividí eso en 2."

"Sumé las tres notas y las dividí en tres. Después sumé cuatro notas agregando 3,5 pues me di cuenta de que este número me sirve para pasar la materia."

"Sumé las cuatro notas y las dividí por cuatro. Probé con 3,5 y me sirvió para pasar la materia."

"Cogí las tres notas, después intenté con una nota diferente que fue 3,4 y con esta no me dio para aprobar, por eso tomé 3,5."

"Cogí las tres notas que daban y las dividí en tres, el resultado dio 3,5 y ese resultado lo sumé con las otras y lo dividí en cuatro."

“Sumé los términos, el resultado me dio 14,2 y lo dividí en cuatro y me dio 3,5.”

A partir de esas respuestas se puede ver que algunos alumnos piensan que la media es asociativa, pues primero hallan la media de un conjunto de datos y luego suman el nuevo dato que necesitan y dividen en dos. Otros estudiantes piensan que hay que sumar las tres notas que tenían dadas y dividir en cuatro para obtener la nota que necesita. Algunos otros tomaron diferentes números hasta obtener el mínimo que necesita para aprobar la materia. Los estudiantes de octavo ya se han enfrentado a este tipo de problema en el colegio, pues sus notas precisamente son entre 0 y 5 y el mínimo para aprobar es 3,5. Este factor puede ser determinante para que los alumnos dieran una respuesta.

Problema 5.4. *Hay 10 personas en un ascensor, 4 mujeres y 6 hombres. El peso medio de las mujeres es de 60 kilos y el de los hombres de 80. ¿Cuál es el peso medio de las 10 personas en el ascensor?*

Muchos estudiantes dejaron ese problema sin responder y otros daban respuestas sin ninguna justificación.

Sorprendentemente algunos estudiantes dieron la respuesta correcta, 72 kilos. Sin conocer aún el concepto, usaron intuitivamente la media ponderada para obtener el valor correcto, o tal vez la información del problema da a entender a los alumnos que deben hacer ese procedimiento.

Al entrevistar algunos de los estudiantes que dieron una respuesta correcta, dieron comentarios como:

“Primero, miré el número de mujeres que había en el ascensor y el de los hombres. Multipliqué el peso de las mujeres por el número de mujeres que había, me dio 240; lo mismo hice para los hombres, y me dio 480. Luego dividí 480 entre

todas las personas, que eran 10, me dio 48; después dividí 240 entre 10 y me dio 24. Sumé 48 y 24, me dio 72.”

“Se multiplican 60 kilos y 80 kilos, luego el resultado se divide en 10 y se suman los resultados.”

“Multipliqué 60 kilos por las 4 mujeres que había y luego 80 por los 6 hombres. Los dos resultados los sumé y la suma la dividí en 10.”

“Multiplico 60 kilos por 4 mujeres, que me da 240. Después 80 kilos por los 6 hombres, me da 480. Después sumé 240 y 480, que me dio 720, que lo dividí en 10 y me dio 72.”

“Las cuatro mujeres las multipliqué por el peso promedio y luego multipliqué el peso de los hombres por el número de hombres. Y luego sumé los resultados y me dio 720. Eso lo dividí entre las 10 personas del ascensor y me dio 72 kilos.”

Estos estudiantes tuvieron en cuenta que en el conjunto de datos hay dos subconjuntos, cada uno con un peso representativo: el peso medio de las mujeres y el de los hombres. También se dieron cuenta que hallar el promedio entre esos dos pesos promedio no sería un buen representante de ese grupo de personas, pues el número de mujeres que hay dentro del ascensor no es el mismo que el de los hombres.

Algunos estudiantes dijeron que el peso medio debe ser 70 kilos obteniendo la media simple de los pesos promedio, 60 y 80, resultado que también se dio en las investigaciones de Batanero.

Varios estudiantes obtuvieron una respuesta satisfactoria en este problema, tuvieron en cuenta toda la información que se da en el enunciado y usaron intuitivamente la media ponderada como una medida de tendencia central para un conjunto de datos con diferentes pesos relativos haciendo el procedimiento aritmético que ellos mismo dieron cuando se les entrevistó. Sin embargo, los alumnos en general no pudieron resolver el problema anterior, de las notas, en el que se esperaba que no tuvieran mayor dificultad. La dificultad del problema de la nota final está en encontrar una nota tal que al sumarla con las anteriores y dividirlas todas entre el número de evaluaciones, que eran cuatro, obtuviera 3,5 que es el mínimo necesario para aprobar, pero los estudiantes de este grado (octavo) no tienen los elementos de álgebra necesarios para hallar este valor desconocido y por eso lo hicieron por medio de ensayo y error.

Problema 5.5. *Se tiene dos conjuntos diferentes de bloques A y B. Las longitudes de los bloques del conjunto A son 10, 20, 30, 40, 50 y 60 cm y las longitudes de los bloques del conjunto B son 10, 10, 10, 60, 60 y 60 cm. ¿Cuál de los dos conjuntos presenta mayor variabilidad?*

Para muchos de los estudiantes los conjuntos eran “casi lo mismo” o “igual de variables”, pues la suma de las longitudes en los dos conjuntos era la misma y por tener la misma cantidad de elementos, el promedio de estas iba a ser el mismo. Tal vez se debió plantear el problema con conjuntos de distinto número de elementos para explorar mejor qué podrían interpretar los alumnos.

Estos estudiantes respondieron sobre la media acertadamente, pero se hizo necesario interrogarlos de nuevo sobre su concepción sobre el término variabilidad.

“Yo lo que me imaginé que había que hacer era una suma. Sumé los elementos del conjunto A y las longitudes del conjunto B. Los dos resultados eran 210. Entonces consideré los conjuntos como iguales.”

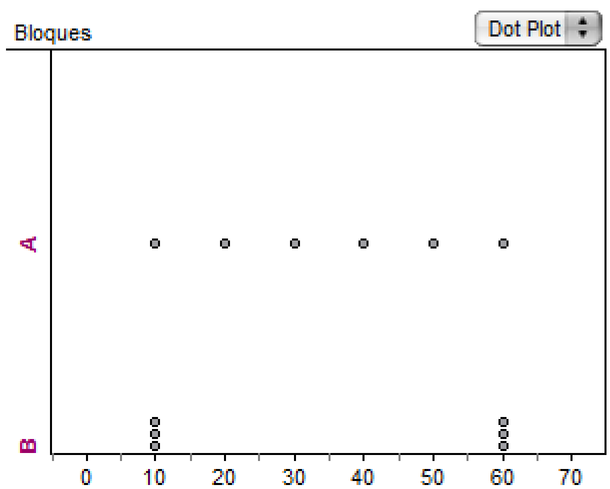
“Son iguales los conjuntos A y B porque al sumarlos dan 210.”

“El conjunto A y el conjunto B presentan la misma variabilidad, pero la diferencia es que los dos tienen los números invertidos.”

Algunos alumnos afirmaban que el conjunto A es el que presenta mayor variabilidad porque todas las longitudes son distintas, mientras que en el conjunto B hay dos valores que se repiten tres veces cada uno. Esto muestra que estos estudiantes entienden el término variabilidad asociado con el número de datos repetidos: máxima variabilidad si todos los datos son distintos y mínima variabilidad si todos los datos son iguales.

Si se hace una representación gráfica de los datos, por ejemplo un diagrama de puntos, se puede ver que en el conjunto A todos los puntos, que representan la longitud de cada bloque, están en un lugar distinto, mientras que en el B los puntos están amontonados en dos grupos.

Figura 1. Diagrama de puntos de los conjuntos de bloques



Viendo la representación gráfica de la información que da el problema los estudiantes podrían afirmar todavía más su forma de entender el término variabilidad, asumiendo máxima variabilidad al tener todos los puntos en diferentes posiciones en el diagrama de puntos y mínima variabilidad al tener los datos acumulados en igual posición. Se hace necesario trabajar de forma especial este término, pues habrá que considerar los valores centrales de un conjunto de datos para hablar de su dispersión teniendo estos valores como referencia y medir qué tan buenos representantes son.

“El conjunto que tiene mayor variabilidad es el A porque tenían más números diferentes, el B sólo tenía dos números que se repetían.”

De lo anterior se puede concluir que algunos de los estudiantes asumen el término variabilidad mirando qué tan distintos los datos, uno respecto a otros, pero no miden la dispersión de un conjunto respecto a un representante o valor central, resultado obtenido también por Batanero.

Problema 5.6. *¿Qué quiere decir que el número promedio de hijos por familia en una población es 2,3?*

Casi todos los estudiantes asumieron que la mayoría de personas en esa población debía tener entre 2 y 3 hijos, que es lo más común, o incluso que todas las familias tienen 2 o 3 hijos.

“La mayoría de las familias tiene 2 hijos.”

“Que a veces se tiene 3 y a veces 2 y casi nunca se tiene 1 hijo.”

“2,3 quiere decir que cada familia tiene entre 2 y 3 hijos.”

Al igual que en las investigaciones de Batanero, resulta difícil para los alumnos entender el número 2,3 como representante de una población. Es necesario prestar atención al significado del valor de la media de un conjunto de datos en un contexto dado.

A partir de las respuestas de los alumnos se puede concluir que cuando se les habla de un promedio de una población, ellos tienden a pensar que la mayoría de datos de esta, o incluso todos, asume ese valor.

Problema 5.7. *¿Qué quiere decir que el salario promedio de un empleado en una ciudad es de 900.000 pesos?*

Para este problema se tuvieron respuestas muy parecidas a las del anterior. Los estudiantes asumen que en esa población la mayoría de personas gana esa cantidad de dinero o un valor cercano.

Esto indica que los estudiantes presentan confusión en los conceptos de media y moda.

Analizando las respuestas de los alumnos a los problemas planteados se pueden ver diferentes conocimientos, concepciones, ideas e intuiciones sobre medidas de tendencia central y de dispersión. Todas estas concepciones previas podrán ser afinadas con el uso del paquete estadístico Fathom cuando se vean conjuntos de datos como un todo y se representen de forma gráfica.

6. PRIMERA ACTIVIDAD: *CALZADO DE 8-05*

6.1 OBJETIVOS

- Ver la necesidad de encontrar un representante en un conjunto de datos.
- Usar la mediana y la moda como representantes de un conjunto de datos.

6.2 DESARROLLO DE LA ACTIVIDAD

Para el desarrollo de esta actividad se dividieron los estudiantes en dos grupos: uno de 20 estudiantes y otro de 19. Por separado, los dos grupos trabajaron el mismo taller.

6.3 *¿QUÉ NÚMERO DE CALZADO TIENEN LOS ESTUDIANTES DE 8-05?*

Esta es la pregunta con que inició la actividad con el objetivo de dar una respuesta para el final de la clase: encontrar un número que represente al grupo de 8-05.

En un primer momento los estudiantes dijeron algunos números de calzado que ellos creían debía ser el número de calzado del grupo y argumentaban sus respuestas diciendo que el número que decían es el de su propio calzado, o que el número de calzado más común es 37 o 38 porque habían oído decir que esos números son los más frecuentes de calzado de las personas en general.

Se pidió entonces a los estudiantes de cada grupo decir cuál es su número de calzado para para, con todos los datos, describir el grupo de la clase. Los estudiantes dijeron su número de calzado y se pidió que registraran todos los datos en una tabla en Fathom. A continuación aparecen las tablas de datos de los dos grupos.

Figura 2. Tabla de datos del número de calzado de cada grupo de estudiantes

Collection 1

	Calzado1
1	37
2	37
3	36
4	38
5	39
6	37
7	36
8	36
9	36
10	36
11	41
12	37
13	38
14	39
15	36
16	40
17	38
18	39
19	38
20	38

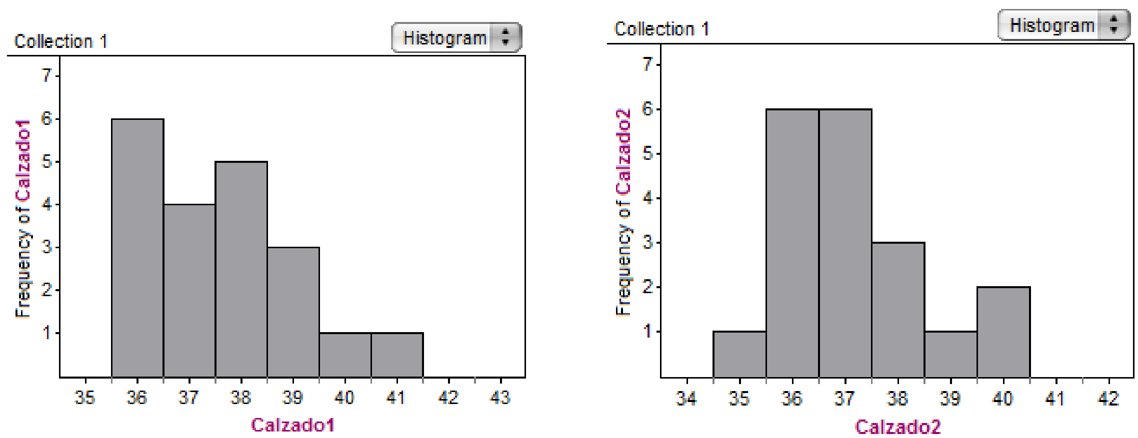
Collection 1

	Calzado2
1	36
2	36
3	39
4	36
5	36
6	35
7	36
8	37
9	40
10	37
11	40
12	38
13	38
14	37
15	37
16	38
17	37
18	37
19	36

6.3.1 La moda. Se planteó la pregunta de clase de nuevo. Para tratar de describir su grupo, los estudiantes comenzaron fijándose en el de mayor valor y el valor más pequeño entre los datos. En un intento por encontrar un dato que representara su grupo, los estudiantes del primer grupo dijeron que el calzado de ellos debe ser 36, pues era el valor que más se repetía (moda). Los del segundo dijeron que debía ser 36 o 37 porque eran los dos números de calzado que más se repetían.

Se preguntó a los estudiantes si tener la lista de datos y sin ordenar era una buena forma de ver el conjunto de datos que tenían. Los estudiantes dieron la idea de construir un gráfico de barras. Este tipo de gráficos ya lo habían visto previamente en sus clases. Los siguientes son los diagramas de barras de los dos grupos.

Figura 3. Gráfico de barras del número de calzado de cada grupo de estudiantes



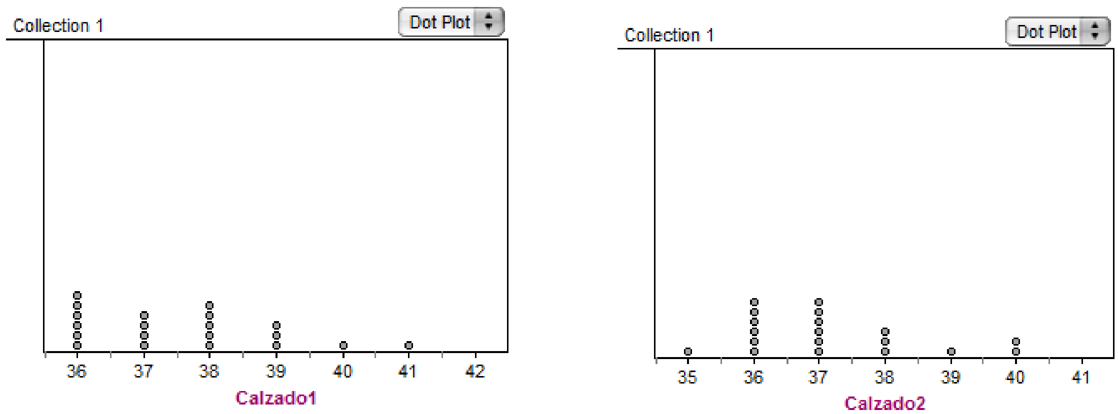
Al observar el gráfico de barras y teniendo en cuenta los valores mayor y menor del calzado de los estudiantes se vio que todos los datos estaban dentro de un intervalo, por ejemplo, no hay un estudiante que calce 32 o 45. Se llegó a una primera conclusión: no tiene sentido elegir como representante del grupo un número de calzado menor que el valor más pequeño (36 en el primer grupo, 35 en el segundo) o mayor que el valor más grande (41 en el primer grupo, 40 en el segundo).

Para encontrar un dato que representara el grupo los estudiantes del primer grupo tenían la idea de que la moda es el mejor representante del grupo; en el segundo, los estudiantes estaban con la idea de elegir un solo número que describiera el conjunto de datos, creían que el representante del grupo era 36 o 37, pero también pensaban que las dos modas eran representativas para el grupo.

6.3.2 La mediana. Una idea que surgió en el primer grupo fue pensar que, a partir de que se tienen 20 datos, el dato número 10, al ordenar los datos de menor a mayor (como está en el diagrama de barras) iba a describir a todo el grupo, pues según el estudiante que pensó esto, se iban a tener igual cantidad de datos debajo de este que por encima.

El investigador dio la idea de trabajar con un diagrama de puntos, pues con este se ve más fácilmente cada uno de los datos. A continuación aparecen los diagramas de puntos de los dos grupos.

Figura 4. Diagrama de puntos del número de calzado de cada grupo de estudiantes



Entonces el estudiante que dio la idea de seleccionar el dato número 10 de menor a mayor contó diez puntos de izquierda a derecha, *"entonces el representante del grupo puede ser 37"*.

El investigador, para poner en duda al estudiante, preguntó: *"pero yo veo que el décimo dato tiene nueve datos por debajo y diez por encima, ¿será realmente buen representante?"*.

El estudiante dijo entonces: *"pues si elijo el dato once voy a tener entonces diez datos por debajo y nueve por encima. Y el representante ahora sería 38"*.

A lo que el encargado de la actividad dijo: *"entonces, ¿cuál se puede elegir como representante?"*. Los estudiantes pensaron por un momento y otro estudiante contestó: *"pues debe ser la mitad entre 37 y 38, o sea 37 y medio"*.

Con el segundo grupo hubo mayor dificultad para surgir el concepto de mediana, pues al tener un número de datos impar, este número no es divisible por dos y a estos estudiantes se les presentó más difícil.

El investigador intervino e hizo una pregunta: *"¿Se puede encontrar en la gráfica un número de calzado tal que haya la misma cantidad de valores por debajo que por encima?"*. Los estudiantes estuvieron mirando en el gráfico y al ver que tenían 19 datos, se dieron cuenta que si elegían el dato número 10 de menor a mayor, iban a tener 9 números de calzado menores o iguales a este y 9 mayores o iguales. Entonces los estudiantes escogieron este décimo valor, 37 como posible representante. Estos alumnos al ver que este número coincidía con una de las modas fue el que eligieron como el representativo para su grupo.

El encargado de la actividad hizo la siguiente pregunta: "*¿Qué pasaría si le llega a usted un estudiante que calce 33? ¿O un estudiante que calce 43? ¿Es posible que ese estudiante sea del grupo 8-05?"*.

Los estudiantes inmediatamente respondieron que no. Se basaron en la gráfica que tenían y en los valores extremos de la distribución. Para describir un grupo no es suficiente tener un representante, es necesario tener en cuenta qué tan dispersos se encuentran los datos. Por ahora los estudiantes se fijaron en los extremos de la distribución y en que tener un calzado en el grupo de 33 o 43 sería "*muy extraño"*. Los estudiantes, en esta actividad, se enfrentaron con la dispersión solamente de forma gráfica.

6.4 EVALUACIÓN Y CONCLUSIONES DEL TALLER

A partir del desarrollo de este taller se puede concluir que los estudiantes intuitivamente eligen la moda como el representante de un conjunto de datos, esto se dio incluso sin tener todavía un gráfico de la distribución, y se afirmó todavía más al construirlo. Esta es la primera medida de tendencia central que apareció en el curso.

También se puede decir que el hecho de tener un número par de datos favorece el surgimiento del concepto de mediana, ya que hace más fácil pensar en un elemento que se encuentre en la mitad de los datos. Caso distinto pasó con un número impar de datos, pues hizo más difícil la tarea de encontrar un valor que tuviera la misma cantidad de valores menores o iguales que este y mayores o iguales, para lo cual el investigador tuvo que hacer una pregunta muy dirigida.

Para describir un conjunto de datos no es suficiente tener un representante. En un primer momento se puede explorar la dispersión de los datos de forma gráfica sin aún conocer las medidas de dispersión, fijándose en los extremos de la distribución y los representantes encontrados.

7. SEGUNDA ACTIVIDAD: *DESCRIBIENDO AL FC BARCELONA*

7.1 OBJETIVOS

- Ver las ventajas y limitaciones que pueden tener la mediana y la moda al tomarlas como representantes de un conjunto de datos.
- Ver la necesidad en algunos casos de encontrar un representante en un conjunto de datos que sea sensible a todos los datos.
- Comprender visual y físicamente el concepto de media.

7.2 DESARROLLO DE LA ACTIVIDAD

Para la realización de este taller, los estudiantes trabajaron con un archivo de Fathom con la lista de jugadores del FC Barcelona de la temporada 2013-2014[1] con los siguientes datos:

Nombre del jugador
Costo del traspaso al club (en millones de euros)
Salario anual (en millones de euros)
Cláusula de rescisión (en millones de euros)
Edad (años)

Se dejó como primera medida que los alumnos exploraran los datos que ellos quisieran, con la idea de tratar de encontrar un representante para cada conjunto de datos (costo, salario, cláusula, edad) y hacerlo en el orden que desearan los estudiantes. Se hizo en cada caso una serie de preguntas con el objetivo de que los estudiantes describieran cada conjunto y encontraran valor central.

La tabla de datos en el Fathom era como la que aparece a continuación:

[1] Fuente: http://paco11.blogspot.com/2013/05/barca-contratos-jugadores-salario.html

Figura 5. Tabla de datos de los jugadores del equipo de fútbol

FC BARCELONA

FC BARCELONA

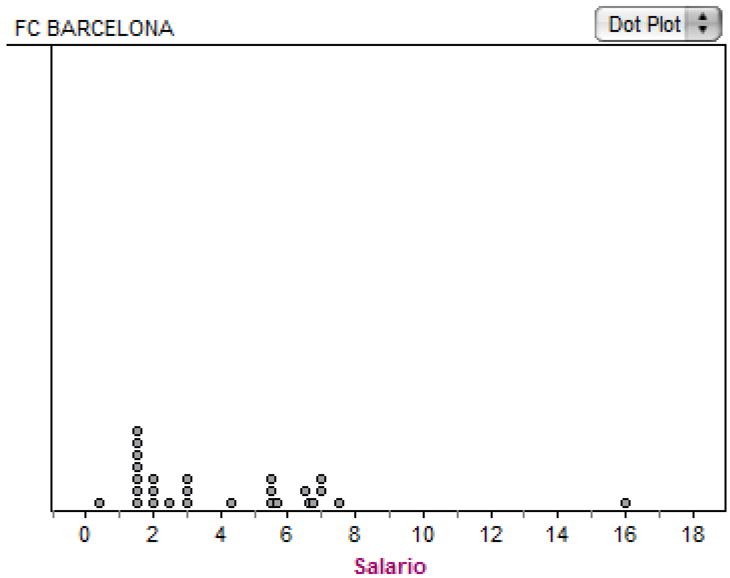
	Jugador	Costo	Salario	Cláusula	Edad	<new>
1	VALDÉS	0.0	6.60	150	31	
2	PINTO	0.5	1.50	5	37	
3	OIER	0.0	0.40	12	23	
4	BARTRA	0.0	1.50	30	22	
5	ALVES	35.5	6.72	125	30	
6	ALBA	14.0	3.00	90	24	
7	PUYOL	0.0	5.50	10	35	
8	ADRIANO	9.5	3.00	90	28	
9	MONTOYA	0.0	1.50	20	22	
10	PIQUÉ	5.0	5.50	200	26	
11	MASCHERANO	20.0	5.50	100	29	
12	SERGI ROBERTO	0.0	1.50	12	21	
13	SONG	19.0	2.50	80	25	
14	DOS SANTOS	0.0	1.50	30	23	
15	CESC FABREGAS	34.0	5.70	200	26	
16	BUSQUETS	0.0	6.50	150	25	
17	XAVI	0.0	7.50	80	33	
18	INIESTA	0.0	7.00	200	29	
19	TELLO	0.0	2.00	25	22	
20	PEDRO	0.0	3.00	150	26	
21	MESSI	0.0	16.00	250	26	
22	ALEXIS SÁNCHEZ	26.0	4.30	100	24	
23	NEYMAR JR	57.0	7.00	190	21	
24	AFELLAY	3.0	2.00	100	27	
25	CUENCA	0.0	1.50	20	22	
26	RAFINHA	0.0	1.50	30	20	
27	DELOFEU	0.0	2.00	35	19	

7.3 ¿CUÁNTO ES EL SALARIO ANUAL DE LOS JUGADORES DEL BARCELONA?

Los estudiantes quisieron trabajar en un inicio con el salario de los jugadores del equipo de fútbol, les causó curiosidad y lo eligieron para explorarlo de primero. Se planteó entonces la pregunta de cuánto gana un jugador del Barcelona al año.

En un primer momento los estudiantes se fijaron en valores particulares, por ejemplo en el salario del jugador que más gana en el equipo y el salario del jugador que menos cobra. Después construyeron un diagrama de puntos para ver mejor la distribución de los datos y tener ideas claras para responder a la pregunta.

Figura 6. Diagrama de puntos de los salarios de los jugadores



Algunos estudiantes propusieron la moda (1,5 millones de euros) como el salario anual que representa al equipo, pero esta respuesta no convenció a todos pues había varios datos que se repiten, como se puede ver en el diagrama de puntos, y porque los sueldos más altos del equipo se encuentran muy lejos de ese valor que más se repite.

Después los estudiantes recordaron lo que habían hecho en el taller anterior (donde trataban de encontrar el número del calzado del grupo), cuando trataban de encontrar un dato que tuviera igual número de datos por debajo de él que por encima, es decir, la mediana.

Un estudiante comentó: *"Hay 27 jugadores, entonces el sueldo que me representa a todos los jugadores es el del jugador 14, o sea el sueldo de Dos Santos que es 1.5 millones, pues va a tener 13 sueldos por encima que por debajo."*

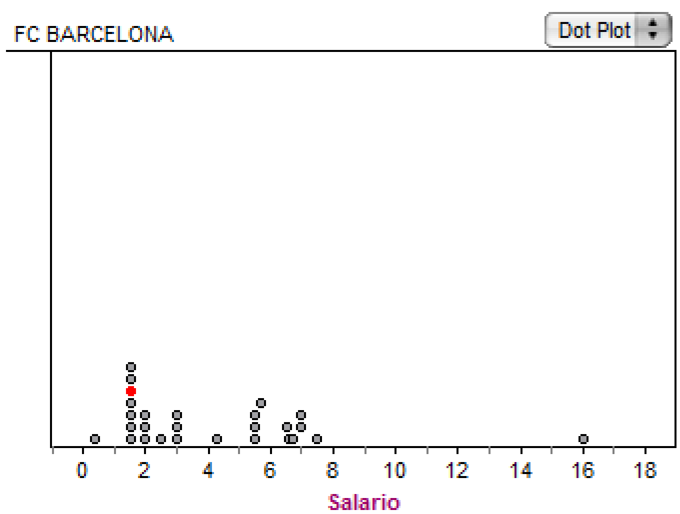
Este estudiante trató de encontrar la mediana del conjunto de datos a partir de la lista, pero no notó que en esta los valores no estaban ordenados y que para hallar un valor que tenga igual número de datos menores o iguales a él que mayores o iguales, es necesario ordenar esta lista.

Al escuchar esta respuesta, el investigador le pidió a ese alumno que mostrara en el proyector lo que llevaba hasta el momento en su computador, y señalara dando clic al dato 14 de la lista, el cual se señalaba automáticamente en el diagrama de puntos.

Figura 7. Jugador número 14 en la lista (sin ordenar)

FC BARCELONA

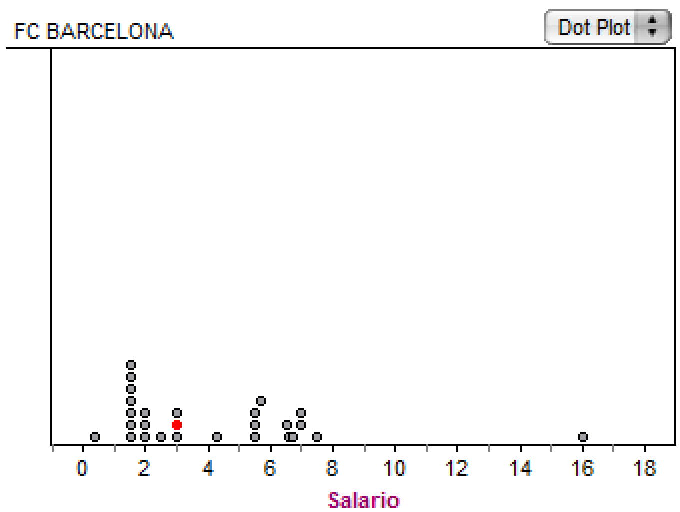
	Jugador	Costo	Salario	Cláusula	Edad	<new>
11	MASCHERANO	20.0	5.50	100	29	
12	SERGI ROBERTO	0.0	1.50	12	21	
13	SONG	19.0	2.50	80	25	
14	DOS SANTOS	0.0	1.50	30	23	
15	CESC FABREGAS	34.0	5.70	200	26	
16	BUSQUETS	0.0	6.50	150	25	
17	XAVI	0.0	7.50	80	33	



Al hacer esto el estudiante se dio cuenta que el valor 14 en la lista no era realmente la mediana, como se puede ver en el diagrama de puntos. Entonces los alumnos vieron que para hallar la mediana es necesario tener ordenados los datos de mayor a menor o de menor a mayor, lo cual está ya hecho en el gráfico que se hace en Fathom. Los estudiantes entonces contaron de menor a mayor (o viceversa) el valor número 14 y lo señalaron.

“Entonces el valor que representa los salarios de los jugadores es 3 millones de euros” dijo un estudiante.

Figura 8. Diagrama de puntos de los salarios con la mediana señalada por los estudiantes

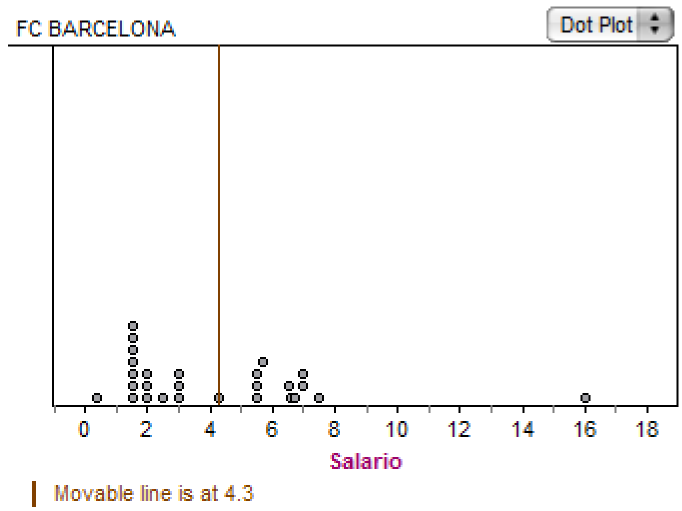


Algunos estudiantes consideraban que el representante de los datos era entonces la mediana, que era 3.5 millones, y otros consideraban que era 1.5 millones pues era el valor que más se repetía, la moda. También había un grupo de estudiantes que pensaba que el representante de los datos no podía ser ninguno de los valores que se habían considerado.

En ese momento una estudiante dijo que una forma en la que podrían encontrar el representante de los datos era sumar todos los sueldos y dividirlos entre el número de jugadores, que era 27. La estudiante dio esa respuesta justificando que ese tipo de procedimiento lo hacen los profesores con todas las notas que tienen en un periodo para dar una nota definitiva. El investigador por el momento le dijo que usara ese valor si quería.

Ante la diferencia de criterio entre los estudiantes, el investigador propuso a los alumnos que crearan dentro del diagrama de puntos una línea movible y la ubicaran donde ellos creyeran que debía ir el representante de los datos.

Figura 9. Diagrama de los salarios con una línea movible generada con Fathom



El investigador dijo a los estudiantes que imaginaran que la distribución de los datos en el diagrama de puntos estaba sobre una balanza y debían entonces encontrar un punto donde esta se equilibrara. Los alumnos siguieron esta indicación y cada uno indicó el que creía era el representante de los datos.

El investigador propuso entonces la siguiente situación:

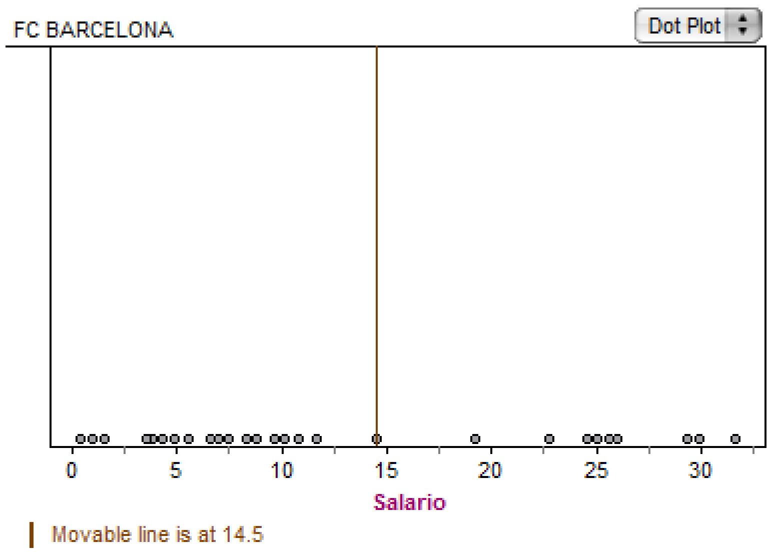
Suponga que cambian los sueldos de algunos jugadores, o todos, por diversos motivos (renovación de contrato, incremento o decrecimiento de su nivel de juego). ¿Qué debe pasar con el representante que escogió para el salario de los jugadores?

Los alumnos empezaron a mover los puntos de la gráfica y se dieron cuenta entonces que al cambiar los datos también debían modificar el valor del representante.

En un primer momento los estudiantes dedujeron que si aumentaba el sueldo de todos los jugadores, el representante también debía hacerlo; y si disminuía el sueldo de todos los jugadores, también debía hacerlo el representante.

Al mover los puntos, un estudiante obtuvo una distribución como la siguiente.

Figura 10. Diagrama de puntos con los salarios modificados



"¿Hay algún valor que se repita más que el resto?" preguntó el investigador.

"No. Yo acomodé los puntos para que me quedaran todos distintos." Respondió el estudiante.

El investigador dejó la pregunta: *"¿Entonces tiene sentido pensar en un valor que se repita más que el resto para tomarlo como representante?"*

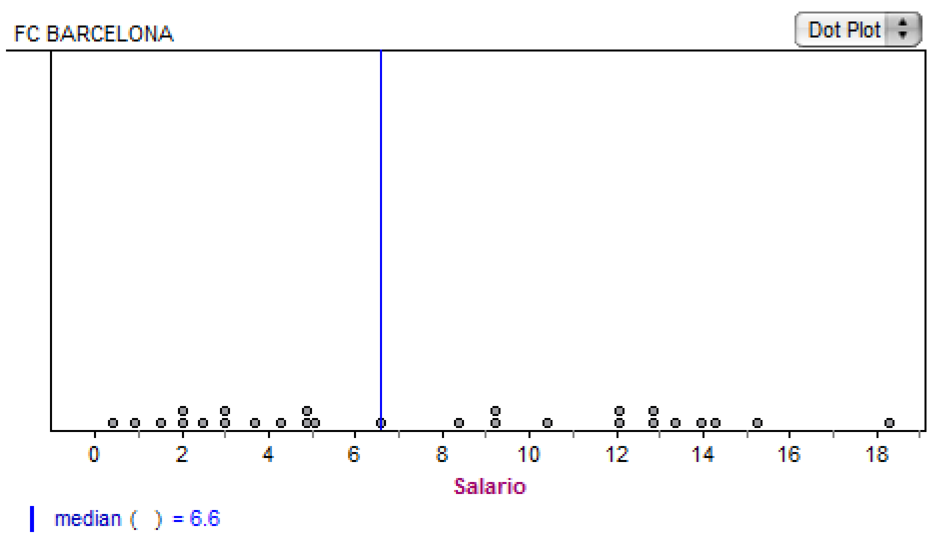
Los estudiantes se dieron cuenta que para ciertos conjuntos de datos no tiene sentido pensar en la moda, como por ejemplo el de la gráfica anterior donde todos los valores de la distribución son distintos.

7.3.1 Explorando la mediana. El investigador propuso a los estudiantes que con ayuda del programa construyeran dentro del diagrama de puntos la mediana.

¿La mediana cambia siempre que cambian los puntos de la gráfica?

Cada estudiante construyó la mediana en la distribución que tenía (pues cada alumno había movido los puntos y cambiado el sueldo de cada jugador) y volvieron a mover los puntos.

Figura 11. Diagrama de salarios modificados con la mediana

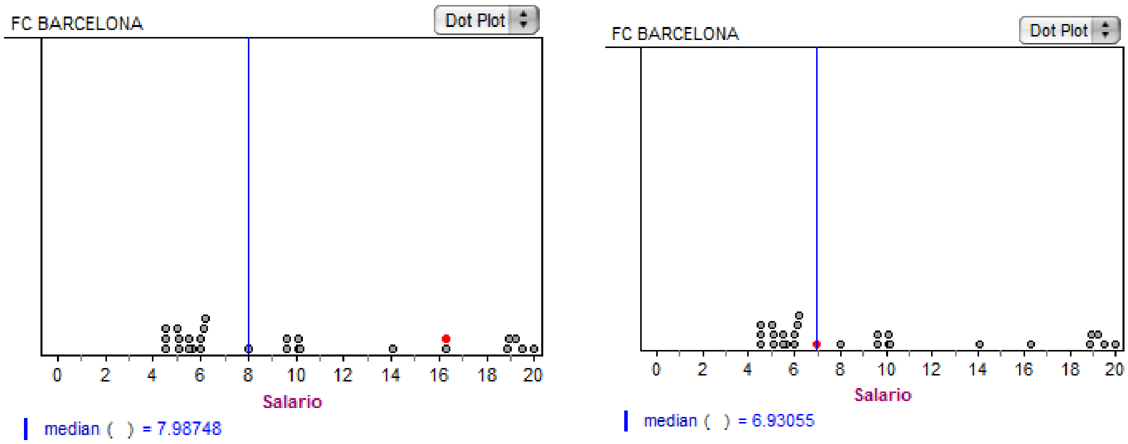


"No. Solamente se mueve con el punto que está en la mitad." Afirmó un alumno.

Este estudiante había movido algunos puntos de la distribución que tenía, mas no hacía que cambiara la mediana y solamente veía que la línea de la mediana permanecía con un mismo punto.

Los estudiantes siguieron cambiando la posición de los puntos y uno de ellos pidió que se mostrara en el proyector lo que él había encontrado:

Figura 12. Exploración del movimiento de la mediana

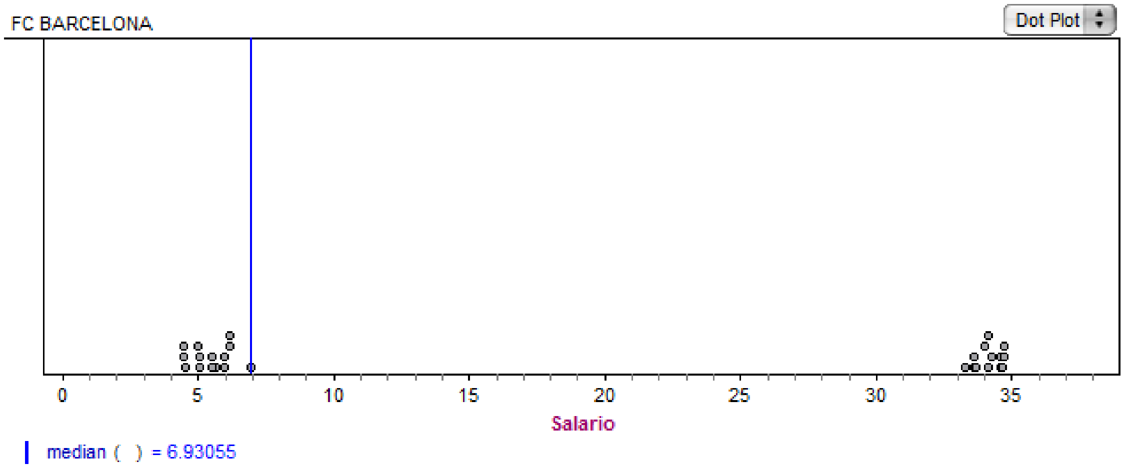


"Si yo muevo uno de los puntos que está a la derecha del valor de la mitad hacia la izquierda de la línea, esta va a estar ahora en ese punto, pero si aumento el valor de ese punto y lo sigo corriendo más a la derecha no me cambia la línea".

El investigador dijo: *"¿Y qué pasa si no mueve uno sino varios valores mayores que la mediana todavía más a la derecha?"*

El estudiante movió varios puntos y vio que la mediana no se movía.

Figura 13. Movimiento de puntos del diagrama donde la mediana no se movía



"¿Entonces para este caso la mediana es un buen representante?" intervino el investigador, a lo que los estudiantes respondieron que no.

¿Cuándo se mueve la mediana?

Los alumnos ya se habían dado cuenta que la mediana no es sensible a los extremos de la distribución, pues movieron a la derecha valores mayores que la mediana y valores menores a la izquierda y vieron que esta no se movía. También se dieron cuenta que como la mediana era uno de los datos, pues había un número impar de datos, cuando movían ese dato la mediana se movía con él hasta llegar al siguiente punto o al anterior. Y si movían uno de los puntos de la derecha de la mediana hacia su izquierda o uno de los de la izquierda hacia su derecha, esta iba a seguirlo hasta encontrar un nuevo punto.

"¿Por qué razón la mediana se mueve de esa forma?" preguntó el investigador.

Los estudiantes siguieron moviendo los puntos y pensando en una respuesta. A los estudiantes se les hizo difícil encontrar la razón para justificar por qué se movía la mediana de esa forma, hasta que uno de los alumnos dijo: *"para encontrar la mediana lo que hacíamos era mirar cuántos datos hay, que son 27 y buscar el dato del centro que es 14 porque tiene 13 datos por encima y 13 por debajo.*

Cuando movemos los puntos lo que pasa es que va cambiando ese dato 14 y por eso la línea se queda donde se encuentra ese dato."

Con este ejercicio los estudiantes exploraron visualmente la mediana y vieron que esta no es sensible a los extremos de la distribución.

7.3.2 La media. Después de descartar la moda y la mediana como representantes para ciertas distribuciones, los estudiantes vieron necesario buscar la forma de construir un representante que fuera sensible a todos los datos de la distribución que cada uno tenía, que se moviera cuando se mueve cada punto de la gráfica.

¿Cómo se puede construir un nuevo representante que se mueva con todos los puntos de la gráfica?

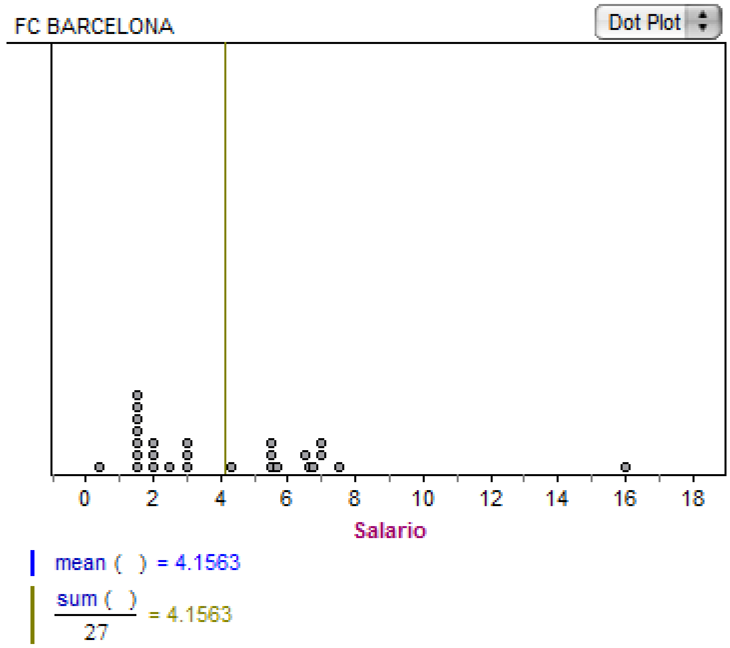
La estudiante que anteriormente propuso hallar la media de los datos preguntó si era posible construirla dentro del diagrama de puntos.

"¿Cómo se halla el promedio de los sueldos de los jugadores del equipo de fútbol?" preguntó el investigador.

"Sumando todos los sueldos y dividiéndolos en 27 que es el número de jugadores que hay" respondió la alumna.

Entonces el investigador indicó a los alumnos una forma de construir la media de los sueldos de los jugadores primero sumando todos los sueldos y después dividiendo esa suma entre el número de jugadores.

Figura 14. Diagrama de puntos de los salarios señalando la media con ayuda de Fathom



Una vez construyeron los estudiantes la media, el investigador preguntó:

¿Qué pasa con el valor de la media si aumenta alguno de los sueldos? ¿Y qué pasa con el valor de la media si disminuye alguno de los sueldos?

Los estudiantes empezaron a mover los puntos de la gráfica o a cambiar sus valores desde la tabla de datos y concluyeron que la media es sensible a todos los datos; si aumenta uno de los sueldos, la media aumenta y si disminuye, lo mismo hace el promedio. Además, al ver los valores que iba tomando la media se dieron cuenta que esta puede no ser igual a ninguno de los datos.

Después de esto, el investigador propuso la siguiente actividad a los estudiantes:

Muevan los puntos y acomódelos en la gráfica de tal forma que la media esté por encima o por debajo de todos.

Los estudiantes trataron de hacer lo que propuso el investigador. Lo intentaron por algún rato pero después se convencieron de que no es posible hacer lo que se había propuesto.

Se concluyó entonces que la media siempre se encuentra entre los datos, debe ser mayor o igual que el valor más pequeño y menor o igual que el valor más grande.

Tras completar esta parte del taller se obtuvieron las siguientes conclusiones:

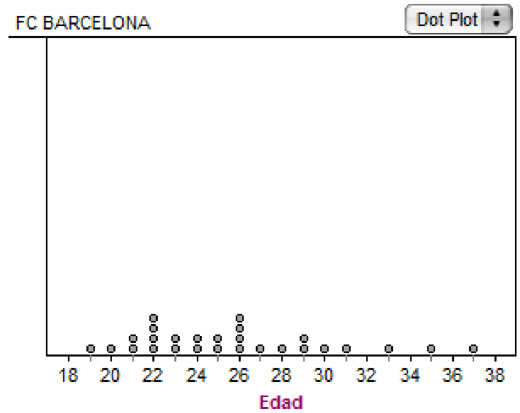
- La media es sensible a todos los datos.
- Si se mueve un punto del diagrama, la media se mueve en la dirección de ese punto.
- La media siempre está entre el valor más pequeño y el valor más grande.
- La media puede no ser igual a ninguno de los datos.

Con esta parte de la actividad los estudiantes vieron que en algunos casos no tiene sentido tomar la(s) moda(s) como valor que represente el conjunto de datos y que a veces la mediana no es tan buen representante, debido a que es poco sensible, y es necesario usar la media que está influenciada por todos los valores.

7.4 *¿CUÁL ES LA EDAD DE LOS JUGADORES DEL BARCELONA?*

Después de haber trabajado con el salario de los jugadores, los estudiantes se interesaron en averiguar la edad de los jugadores del equipo de fútbol. Al igual que en la parte anterior de este taller y la primera actividad, los alumnos primero se fijaban en valores particulares, por ejemplo se fijaban en cuál era el jugador de mayor edad y el jugador más joven. Luego construyeron un diagrama de puntos para ver mejor todos los datos de forma ordenada.

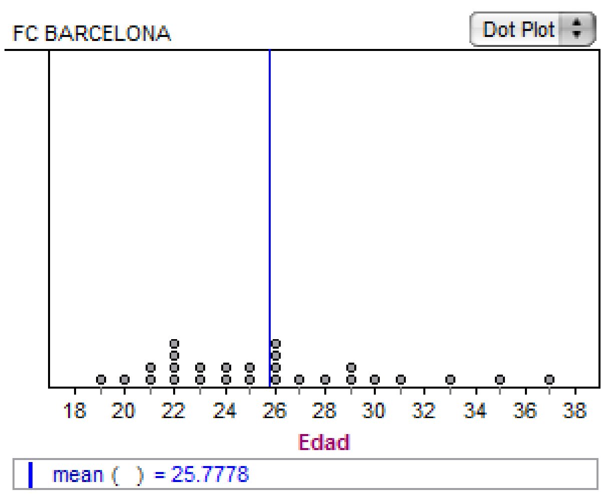
Figura 15. Diagrama de puntos de las edades de los jugadores

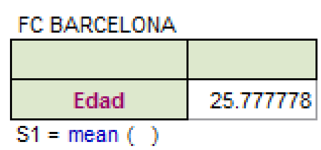


Se hizo entonces necesario encontrar un valor que represente a todas las edades de los jugadores. Como los estudiantes habían usado la media en la parte anterior y visto algunas de sus propiedades, decidieron usarla como representante.

Los estudiantes con la ayuda del software hallaron el promedio de las edades y lo construyeron dentro del diagrama de puntos.

Figura 16. Diagrama de puntos de las edades señalando la media con Fathom





"¿Cuál es la edad de los jugadores?" preguntó el investigador.

“25.7778 años” dijo la mayoría de alumnos. Pero un estudiante pidió la palabra y dijo: *“tenemos todas las edades en años: 19, 20, 25, 30 años y ese 25.7778 está cerquita de 26 años que es una de los dos valores que más se repite con el 24. ¿Por qué no lo tomamos 26 años como el representante?”*

Entonces los estudiantes estuvieron de acuerdo en decir que la edad de los jugadores del equipo de fútbol es de 26 años.

Al tener una respuesta los estudiantes, intervino el investigador y preguntó:

“Ustedes dicen que la edad de los jugadores del Barcelona es de 26 años. ¿Cuál será la edad de los jugadores dentro de tres años (contando con estos mismos jugadores aunque llegaran a cambiar de equipo)?”

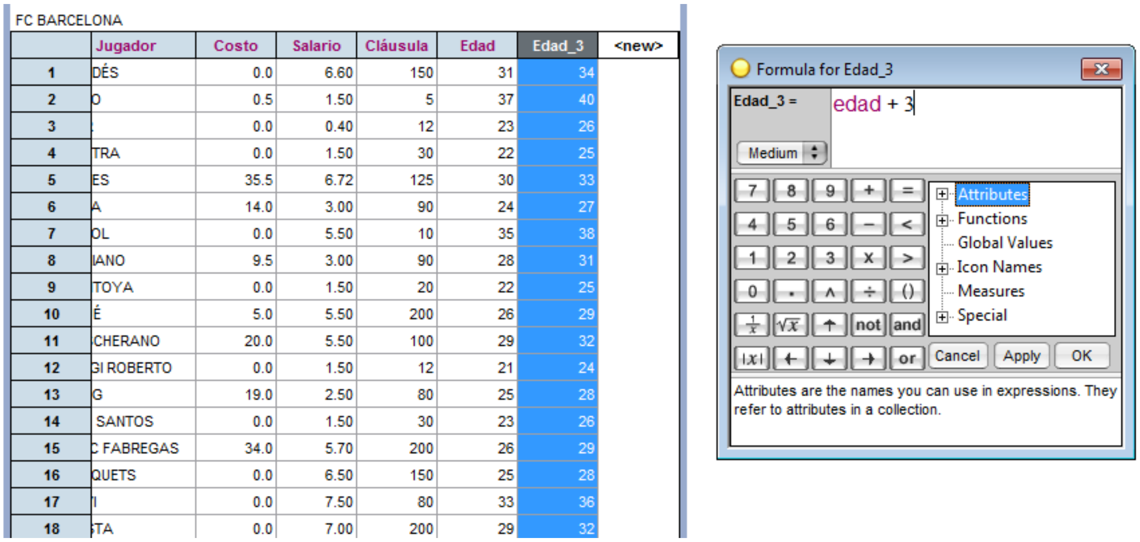
“Pues 29, fácil” respondió un alumno.

“¿Por qué 29?” preguntó el investigador. Casi todos los alumnos pensaban que si pasaban tres años, la media de los datos también tenía que aumentar esa cantidad de años.

El investigador volvió a intervenir: *“¿será que si se aumenta tres años la edad de los jugadores del equipo de fútbol aumenta necesariamente la media?”*

Los estudiantes estuvieron pensando por un momento en una razón por la que pudieran justificar lo que creían. Para terminar de resolver el problema, el investigador propuso a los estudiantes crear en la tabla de datos un nuevo atributo que sería la edad de los jugadores más tres años.

Figura 17. Creación de nuevo atributo: edad más tres años en Fathom

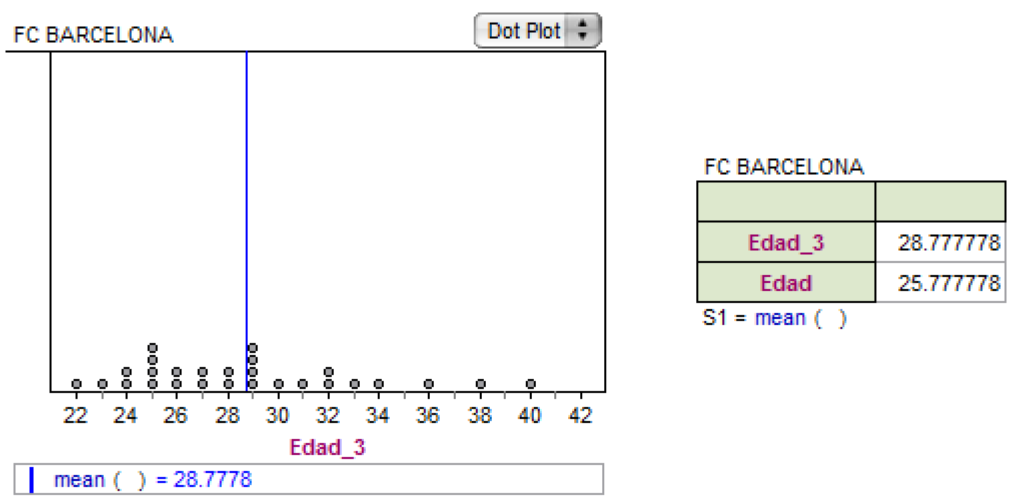


FC BARCELONA

	Jugador	Costo	Salario	Cláusula	Edad	Edad_3	<new>
1	DÉS	0.0	6.60	150	31	34	
2	O	0.5	1.50	5	37	40	
3		0.0	0.40	12	23	26	
4	TRA	0.0	1.50	30	22	25	
5	ES	35.5	6.72	125	30	33	
6	A	14.0	3.00	90	24	27	
7	OL	0.0	5.50	10	35	38	
8	IANO	9.5	3.00	90	28	31	
9	TOYA	0.0	1.50	20	22	25	
10	É	5.0	5.50	200	26	29	
11	CHERANO	20.0	5.50	100	29	32	
12	GI ROBERTO	0.0	1.50	12	21	24	
13	G	19.0	2.50	80	25	28	
14	SANTOS	0.0	1.50	30	23	26	
15	C FABREGAS	34.0	5.70	200	26	29	
16	QUETS	0.0	6.50	150	25	28	
17	I	0.0	7.50	80	33	36	
18	TA	0.0	7.00	200	29	32	

Enseguida los estudiantes hicieron un diagrama de puntos con la edad de los jugadores dentro de tres años y hallaron la media.

Figura 18. Diagrama de puntos de la edad de los jugadores más tres años



Los estudiantes vieron que al aumentar tres años a la edad de los jugadores, la media también aumentaría tres años.

"La edad que representa a los jugadores es casi 26 años. Y la edad dentro de tres años es casi 29 (28.7778 o sea 25.7778+3). Entonces si yo aumento tres años las edades, lo mismo le pasa al promedio." Dijo uno de los estudiantes.

Después el investigador dijo: *"¿Y cuál será la edad promedio de los jugadores en cinco años?"*

Los estudiantes respondieron 31 de una vez, aunque enseguida crearon un nuevo atributo con las edades más cinco años, construyeron un diagrama de puntos y hallaron la media.

"¿Cuál era la edad de los jugadores hace 2 años?" preguntó el investigador.

Los alumnos respondieron 24, y creando un nuevo atributo con la edad de los jugadores menos tres años, construyendo su respectivo diagrama de puntos y hallando la media de los datos rectificaron su respuesta.

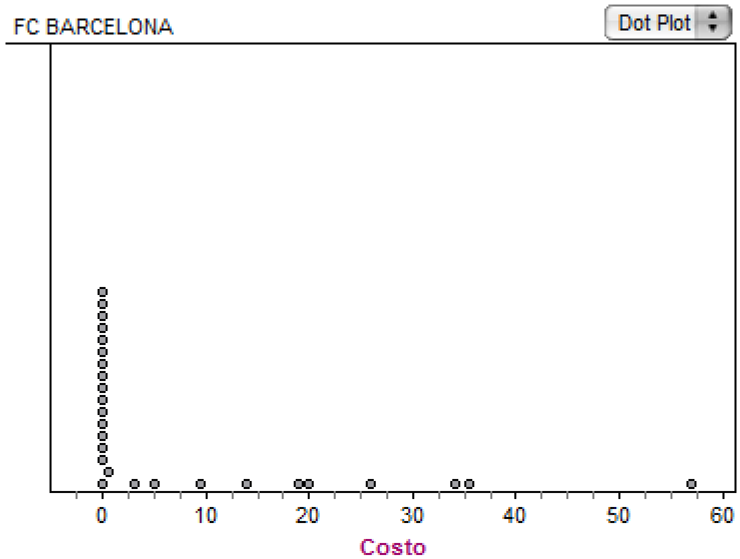
Con esta parte de la actividad se concluyó con los estudiantes que si a todos los valores de un conjunto de datos se les añade o se les sustrae un mismo valor, la media de esos datos será afectada por ese valor.

7.5 *¿CUÁNTO GASTA EL BARCELONA POR ADQUIRIR UN NUEVO JUGADOR?*

Para continuar con el taller, los estudiantes del curso escogieron explorar cuánto le costó al equipo adquirir a cada uno de sus jugadores. Una vez más el procedimiento que hicieron los estudiantes era mirar datos particulares, valores máximo y mínimo, construir un diagrama para ver mejor la distribución y encontrar un representante del grupo.

Los estudiantes desde un principio observaron que gran cantidad de jugadores tenían costo cero, aunque también hay jugadores que tuvieron un costo muy alto en comparación a los demás. A continuación aparece el diagrama de puntos del costo del traspaso al club de los futbolistas.

Figura 19. Diagrama de puntos del costo de los jugadores



Los estudiantes al ver el diagrama de puntos vieron que la mayoría de los jugadores no tuvieron costo, por lo tanto la moda y la mediana es la misma, cero. Por lo tanto ellos lo tomaron como el representante del grupo, no tuvieron en cuenta la media, que es un poco más de 8 millones de euros.

"¿Qué significa que tantos jugadores tengan costo cero?" preguntó el investigador.

Algunos de los estudiantes respondieron que habían oído hablar en la televisión y en la radio que la mayoría de los jugadores del Barcelona vienen desde sus inferiores y que este equipo muy rara vez compra futbolistas a alto costo. Por esta razón, y al ver la gráfica, los estudiantes se convencieron que el precio que paga el equipo de fútbol por sus jugadores es cero.

Al analizar la pregunta de cuánto paga el equipo por adquirir un jugador, los estudiantes se dieron cuenta de que a veces el hecho de que la media sea sensible a todos los datos puede ser desfavorable tomarla como representante de un conjunto de datos; en ocasiones gana más importancia el uso de la moda y de la mediana.

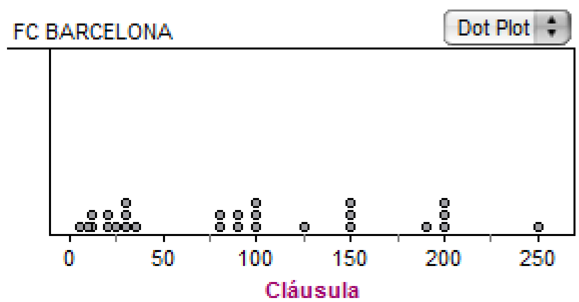
7.6 ¿CUÁNTO PUEDE ESPERAR GANAR EL CLUB POR LA VENTA DE UN JUGADOR?

Para el desarrollo de esta parte de la actividad se explicó a los estudiantes que la cláusula de rescisión es el precio que debería pagar en teoría al Barcelona otro club que quiera comprar uno de sus jugadores.

Una vez más los estudiantes primero se fijaron en valores particulares, por ejemplo que el jugador de menor costo saldría del club por 5 millones de euros y que el jugador más cotizado cuesta 250 millones de euros.

Los estudiantes construyeron el diagrama de puntos de la cláusula de los jugadores.

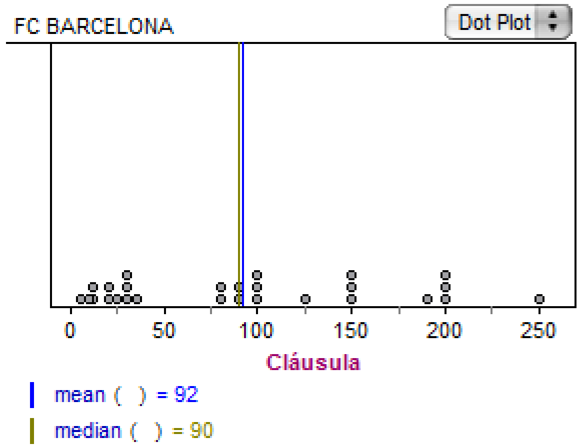
Figura 20. Diagrama de puntos de la cláusula de rescisión de los jugadores



Al ver la distribución, los estudiantes vieron que hay varias modas, por lo que sería difícil escoger una que representara el conjunto de datos.

Los estudiantes entonces hallaron la media y la mediana, cuyos valores fueron 92 y 90 millones de euros respectivamente.

Figura 21. Diagrama cláusula de rescisión, señalando la media y la mediana



Los estudiantes entonces dijeron que el club podría esperar ganar por la venta de un jugador entre 90 y 92 millones de euros.

7.7 EVALUACIÓN Y CONCLUSIONES DEL TALLER

Con el desarrollo de las preguntas de esta actividad los estudiantes lograron comprender visual y físicamente el concepto de media y algunas de sus características más importantes, con ayuda de las representaciones visuales y dinámicas que ofrece el paquete estadístico Fathom.

Los estudiantes vieron las ventajas y desventajas de elegir la media, la mediana o la moda como representantes de un conjunto de datos sin mayor intervención del investigador. Los mismos alumnos con observar la gráfica de los datos son capaces de elegir la mejor entre estas tres medidas de tendencia central dependiendo de la distribución sin necesidad de tener a alguien que les diga cuál deben escoger y en qué circunstancias.

Al dar respuesta a los problemas del taller, los alumnos reconocieron las ventajas y desventajas de la sensibilidad de la media y de escogerla como representante de un conjunto de datos.

8. TERCERA ACTIVIDAD: *COMPARANDO DOS EQUIPOS DE FÚTBOL*

8.1 OBJETIVOS

- Usar las medidas de tendencia central para comparar dos grupos de datos.
- Ver la necesidad de medir la dispersión de un conjunto de datos.

8.2 DESARROLLO DE LA ACTIVIDAD

Los estudiantes tenían en cada computador un archivo de Fathom con la lista de jugadores del Barcelona y del Real Madrid, temporada 2013-2014, con los siguientes datos:

Nombre de los jugadores

Salario (millones de euros por año)

Edad (años)

Figura 22. Tabla de datos de los dos equipos de fútbol

FC BARCELONA Y REAL MADRID CF

	Jugadores_FCB	Salario_FCB	Edad_FCB	_	Jugadores_RMCF	Edad_RMCF	Salario_RMCF
1	VALDÉS	6.60	31		CASILLAS	32	7.50
2	PINTO	1.50	37		D. LÓPEZ	32	2.00
3	OIER	0.40	23		JESÚS	25	0.60
4	BARTRA	1.50	22		ARBELOA	30	1.80
5	ALVES	6.72	30		F. COENTRAO	25	2.00
6	ALBA	3.00	24		SERGIO RAMOS	27	4.50
7	PUYOL	5.50	35		MARCELO	25	2.00
8	ADRIANO	3.00	28		PEPE	30	4.50
9	MONTOYA	1.50	22		CARVAJAL	21	0.80
10	PIQUÉ	5.50	26		NACHO	23	0.06
11	MASCHERANO	5.50	29		R. VARANE	20	1.20
12	SERGI ROBERTO	1.50	21		G. BALE	24	13.31
13	SONG	2.50	25		XABI ALONSO	32	4.50
14	DOS SANTOS	1.50	23		KHEDIRA	26	5.00
15	CESC FABREGAS	5.70	26		MODRIC	28	6.76
16	BUSQUETS	6.50	25		ISCO	21	2.00
17	XAVI	7.50	33		DI MARÍA	25	3.00
18	INIESTA	7.00	29		CASEMIRO	21	1.25
19	TELLO	2.00	22		ILLARRAMENDI	23	2.50
20	PEDRO	3.00	26		BENZEMA	26	5.00
21	MESSI	16.00	26		C. RONALDO	28	15.00
22	ALEXIS SÁNCHEZ	4.30	24		MORATA	21	1.00
23	NEYMAR JR	7.00	21		JESÉ	20	1.20
24	AFELLAY	2.00	27				
25	CUENCA	1.50	22				
26	RAFINHA	1.50	20				
27	DELOFEU	2.00	19				

La idea de este taller es la de comparar dos grupos de datos, primero gráficamente y después usando medidas de tendencia central y midiendo la dispersión de cada conjunto.

8.3 *¿QUIÉN GANA MÁS, UN JUGADOR DEL BARCELONA O UNO DEL REAL MADRID?*

Esta pregunta se hizo para que los estudiantes buscaran la forma de comparar dos grupos de datos: los salarios del Barcelona y los salarios del Real Madrid.

En un primer intento por resolver la pregunta los estudiantes compararon datos particulares como por ejemplo los jugadores mejor pagos de cada equipo y los jugadores de menor sueldo.

"El jugador que más gana en el Barcelona es Messi, con 16 millones de euros y el que más gana en el Real Madrid es Ronaldo con 15. Y el jugador que menos le pagan en el Barcelona es Oier con 0.4, mientras que el jugador que menos le pagan del Madrid es Nacho con 0.06 millones. Entonces los jugadores del Barcelona ganan más." Dijo uno de los alumnos. Este estudiante trató de responder la pregunta comparando los extremos de cada distribución, no se contentó con comparar los máximos, como suele suceder, sino que también comparó los mínimos. Como tuvo la fortuna de que en ambos casos el análisis conducía al mismo resultado, pudo concluir, sin contradicciones, el resultado final.

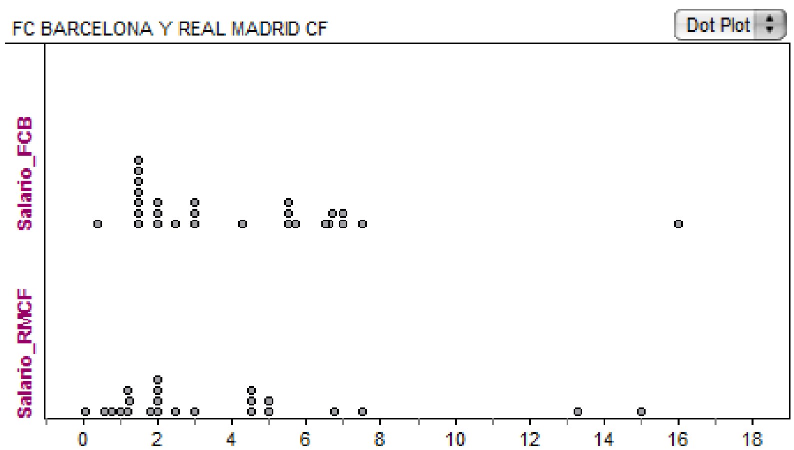
"¿Qué opinan los demás?" preguntó el investigador. La mayoría de alumnos estuvo de acuerdo con la intervención de su compañero.

"¿No ven algún problema a lo que dice su compañero?" preguntó de nuevo el investigador. Los alumnos se quedaron en silencio.

El investigador hizo entonces la aclaración de que este estudiante por el momento sólo se había fijado en unos datos particulares y que debía compararse los dos grupos y tener en cuenta todos los datos.

Los estudiantes entonces construyeron en Fathom un diagrama de puntos con el salario de los jugadores de cada equipo.

Figura 23. Diagrama de puntos de los salarios distinguiendo cada equipo



Al ver el diagrama de puntos, los estudiantes se fijaron más que todo en datos atípicos: dentro de los salarios del Real Madrid hay dos valores lejanos del resto del grupo, mientras que en el Barcelona sólo hay un jugador que tiene un sueldo muy alto en comparación a los demás. Esto hizo pensar a los alumnos que el salario de los jugadores del Real Madrid es más alto que el de los jugadores del Barcelona.

"En el Real Madrid hay dos jugadores que ganan mucho: Bale y Cristiano Ronaldo. Mientras que en el Barcelona Messi es el único que gana mucho más que el resto." Dijo un estudiante.

Después de ver datos particulares y el diagrama de puntos de los datos, los alumnos pensaron en encontrar un número que representara los salarios de cada equipo y después compararlos.

“¿Cómo podemos saber en cuál equipo gana más dinero un jugador?” Preguntó el investigador.

“Pues por ahora tenemos que Messi, que juega en el Barcelona, es el que más gana de todos los jugadores de los dos equipos. Pero Bale y Ronaldo, que son del Madrid también ganan sueldos muy altos.” Dijo un estudiante.

“¿Y cómo podríamos hacer para comparar los dos equipos teniendo en cuenta los salarios de todos los jugadores?” Volvió a preguntar el investigador.

“Podemos buscar un salario que represente a cada equipo.” Dijo uno de los estudiantes.

“¿Y para qué serviría tener un salario representante de cada equipo?” Preguntó una vez más el investigador.

“Para mirar cuál equipo tiene mayor representante y ese va a ser el que tenga sueldos más altos.” Intervino una estudiante.

Los estudiantes hallaron la media y la mediana de cada conjunto de datos y las construyeron dentro del diagrama de puntos.

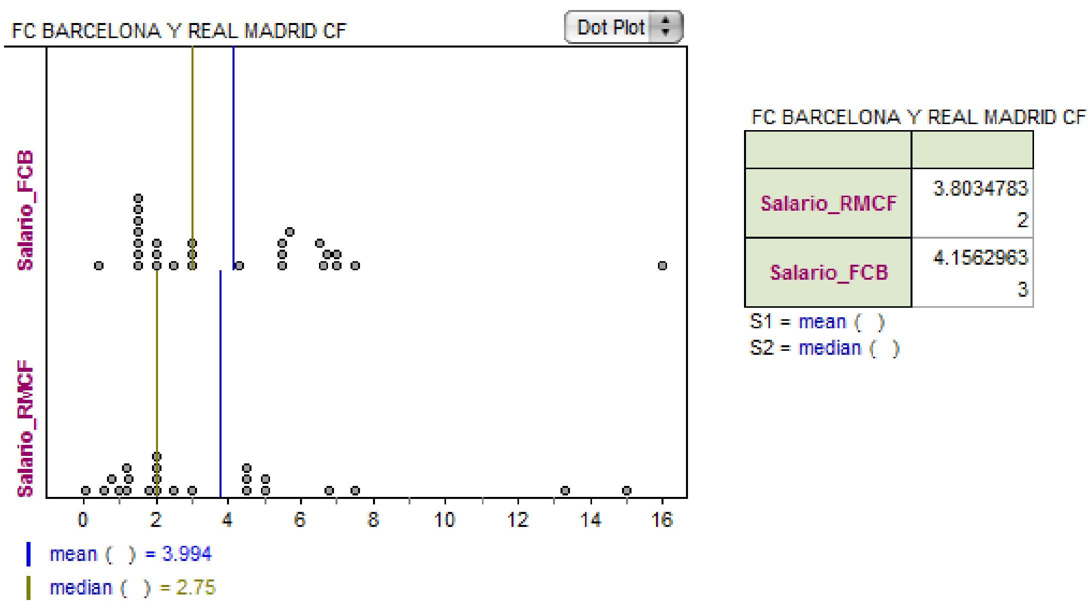
“¿Tiene sentido pensar en la moda como representante?” Preguntó el investigador.

“No. La vez pasada vimos que con el Barcelona no tenía sentido, además hay varios sueldos que se repiten en el Madrid, pero las veces que se repiten no son muchas.” Dijo uno de los alumnos.

De esta manera los estudiantes usaron la media y la mediana para tratar de responder a la pregunta.

Al ver las dos medidas dentro del diagrama de puntos, los estudiantes observaron que la media del salario de los jugadores del Barcelona es mayor que la de los jugadores del Real Madrid, lo mismo pasa con la mediana. Por lo tanto los alumnos concluyeron que los jugadores del primer equipo ganan más que los del segundo.

Figura 24. Diagrama salarios de cada equipo con la media y la mediana



“¿Qué valor tomamos como representante?” Preguntó el investigador.

“Pues cualquiera que tomemos, la respuesta va a ser la misma: los del Barcelona ganan más.” Dijo un estudiante.

“¿Pero cuál medida será mejor representante de los datos para cada equipo?” preguntó de nuevo el investigador.

“La clase pasada vimos que cuando uno movía los datos la media también se movía y no siempre lo hacía la mediana. Y si suponemos que cambian los salarios de los jugadores, como habíamos hecho la clase pasada, pues en ese caso va a ser mejor la media.” Dijo un estudiante.

El hecho de que este último estudiante hubiera hecho la suposición de que cambiaran los salarios de los jugadores, hizo que los alumnos entonces tomaran la media como representante de los salarios de cada equipo y así concluir que los jugadores del Barcelona ganan más.

Después de responder a la pregunta de quiénes ganan más dinero, se hizo en clase la siguiente pregunta.

8.4 *¿EN CUÁL DE LOS DOS EQUIPOS LOS SALARIOS DE LOS JUGADORES SON MÁS DESIGUALES?*

Esta pregunta se hizo con el propósito de que los alumnos midieran de alguna forma la dispersión de los datos de cada conjunto.

Los estudiantes se fijaron de nuevo en los extremos de cada distribución para responder esta pregunta. Ellos vieron en el diagrama de puntos que hay más distancia entre el sueldo más bajo y el más alto del Barcelona que entre el salario menor y el mayor del Real Madrid.

"¿Cómo hacemos para saber cuál equipo tiene salarios más desiguales? ¿O cómo hacemos para saber cuál tiene salarios más parecidos?" preguntó el investigador.

"Hay que mirar en los dos equipos si los sueldos de los jugadores que más ganan son más o menos parecidos, o muy diferentes de los que menos ganan." Dijo uno de los alumnos.

"¿Y cómo hacemos eso?" Preguntó el investigador.

"Podríamos mirar la distancia entre el sueldo más bajo y el sueldo más alto en cada equipo con la gráfica." Respondió el estudiante.

8.4.1 El rango. El investigador intervino y definió el rango como una forma de medir la variabilidad de un conjunto de datos de la siguiente manera:

$$rango = distancia\ entre\ el\ valor\ más\ pequeño\ y\ el\ valor\ más\ grande$$

“¿Y cómo se puede medir la distancia entre el valor más pequeño y el valor más grande?” preguntó el investigador. Los estudiantes pensaron un poco hasta que uno de ellos dijo *“restando el menor salario al mayor”*.

Entonces el rango quedó definido de la siguiente forma:

$$rango = valor\ más\ grande - valor\ más\ pequeño$$

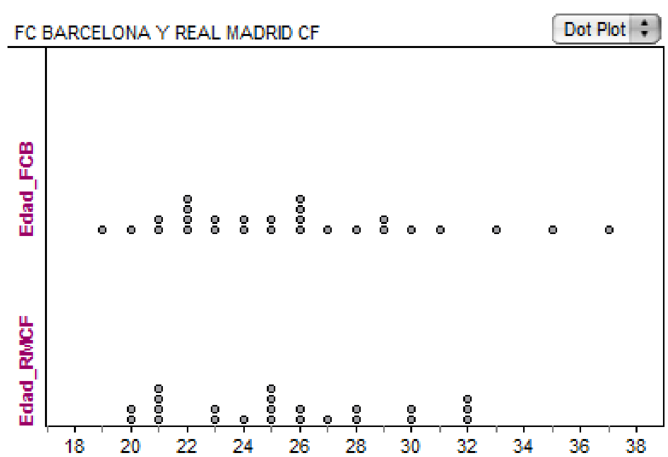
Los estudiantes entonces hallaron el rango de cada conjunto de datos y vieron que la distancia entre el menor sueldo y el mayor del Barcelona es de 14.94 millones de euros y la diferencia entre el sueldo más alto y el más bajo del Real Madrid es 15.6 millones.

Además de utilizar el rango, los estudiantes también se fijaron de nuevo en los valores atípicos de los salarios: dos en el Madrid y uno en el Barcelona.

8.5 *¿CUÁL EQUIPO TIENE MAYOR EDAD?*

Ahora se pretendía comparar las edades de los jugadores de los dos equipos de fútbol. Los estudiantes construyeron un diagrama de puntos y comenzaron comparando datos particulares, especialmente los extremos de cada conjunto de datos.

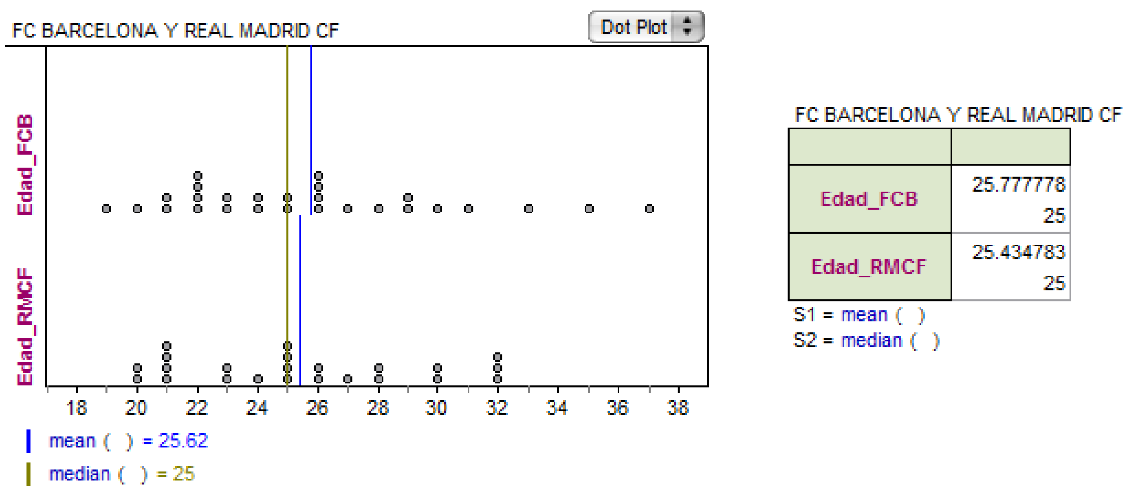
Figura 25. Diagrama de puntos de la edad de los jugadores distinguiendo cada equipo



Los alumnos se fijaron que el jugador más joven del Barcelona tiene 19 años, mientras que el menor del Real Madrid tiene 20. Esto en un principio podría ser una pista para pensar que los jugadores del primer equipo son menores que los del segundo; sin embargo, la edad máxima del equipo catalán es de 37 años, además los tres jugadores de más edad de este equipo tienen 33, 35 y 37 años, mientras que la edad de los tres jugadores más veteranos del club madridista es 32 años. Luego esto daría a pensar ahora que los jugadores del Barcelona son mayores que los del Madrid.

Después de fijarse en algunas particularidades de los datos, los estudiantes procedieron a hallar la media y la mediana de los dos conjuntos de datos y a construirlas dentro del diagrama.

Figura 26. Diagrama edad de jugadores de cada equipo señalando media y mediana



Los alumnos vieron que la media de la edad de los jugadores del Barcelona era muy parecida a la media de la edad de los jugadores del Real Madrid, además la mediana era la misma para los dos equipos. Esto hizo pensar a los alumnos las edades de los jugadores de los dos equipos son muy parecidas o casi iguales y concluir por el valor de la media que la edad del equipo catalán es un tanto mayor.

8.6 EVALUACIÓN Y CONCLUSIONES DEL TALLER

Para comparar grupos de datos el procedimiento que usan los estudiantes es: primero comparar datos particulares como los extremos de las distribuciones y valores atípicos; después construir un diagrama para ver las características más relevantes de cada distribución; posteriormente usar medidas de tendencia central para contrastar los conjuntos de datos que se tienen y por último medir la dispersión de los datos de cada grupo para encontrar diferencias que no pueden encontrarse por medio del representante.

El rango es una medida de dispersión que los mismos estudiantes lograron definir viendo la representación gráfica de conjuntos de datos, al comparar la distancia entre los extremos de cada distribución.

9. CUARTA ACTIVIDAD: *ELIMINATORIAS BRASIL 2014*

9.1 OBJETIVOS

- Explorar y medir la dispersión de un conjunto de datos.
- Usar medidas de tendencia central y de dispersión para comparar conjuntos de datos.

9.2 DESARROLLO DE LA ACTIVIDAD

Los estudiantes trabajaron con un archivo de Fathom en donde está la tabla de la eliminatoria al mundial del 2014 de Sudamérica y de Europa, con los siguientes datos:

País

Partidos jugados

Goles anotados

Puntos obtenidos

La tabla de posiciones final de las eliminatorias al mundial que se ve al iniciar Fathom está en la figura de la siguiente página.

Figura 27. Tabla Eliminatorias Mundial Brasil 2014

Grupo	EUROPA	PJ_E	GF_E	PTS_E
A	Bélgica	10	18	26
A	Croacia	10	12	17
A	Serbia	10	18	14
A	Escocia	10	8	11
A	Gales	10	9	10
A	Macedonia	10	7	7
B	Italia	10	19	22
B	Dinamarca	10	17	16
B	Rep. Checa	10	13	15
B	Bulgaria	10	14	13
B	Armenia	10	12	13
B	Malta	10	5	3
C	Alemania	10	36	28
C	Suecia	10	19	20
C	Austria	10	20	17
C	Rep. de Irlanda	10	16	14
C	Kazajstán	10	6	5
C	Islas Feroe	10	4	1
D	Holanda	10	34	28
D	Rumanía	10	19	19
D	Hungría	10	21	17
D	Turquía	10	16	16
D	Estonia	10	6	7
D	Andorra	10	0	0

SUDAMÉRICA	PJ_SA	GF_SA	PTS_SA
Argentina	16	35	32
Colombia	16	27	30
Chile	16	29	28
Ecuador	16	20	25
Uruguay	16	25	25
Venezuela	16	14	20
Perú	16	17	15
Bolivia	16	17	12
Paraguay	16	17	12

Grupo	EUROPA	PJ_E	GF_E	PTS_E
E	Suiza	10	17	24
E	Islandia	10	17	17
E	Eslovenia	10	14	15
E	Noruega	10	10	12
E	Albania	10	9	11
E	Chipre	10	4	5
F	Rusia	10	20	22
F	Portugal	10	20	21
F	Israel	10	19	14
F	Azerbaiyán	10	7	9
F	Irlanda del N...	10	9	7
F	Luxemburgo	10	7	6
G	Bosnia-Herz...	10	30	25
G	Grecia	10	12	25
G	Eslovaquia	10	11	13
G	Lituania	10	9	11
G	Letonia	10	10	8
G	Liechtenstein	10	4	2
H	Inglaterra	10	31	22
H	Ucrania	10	28	21
H	Montenegro	10	18	15
H	Polonia	10	18	13
H	Moldavia	10	12	11
H	San Marino	10	1	0
I	España	8	14	20
I	Francia	8	15	17
I	Finlandia	8	5	9
I	Georgia	8	3	5
I	Bielorrusia	8	7	4

La pregunta que se trató de responder en la clase es la siguiente:

9.3 *¿DÓNDE SON MÁS REÑIDAS LAS ELIMINATORIAS?*

En un principio se planteó esta pregunta sin todavía abrir el archivo de Fathom, para conocer qué opinaban los alumnos al respecto.

“Yo creo que las eliminatorias son más difíciles en Europa porque los equipos juegan menos partidos. Acá en América los equipos tienen más oportunidad de remontar porque juegan más partidos.”

“La eliminatoria es más difícil en Europa porque en algunos grupos hay veces que les toca jugar equipos muy difíciles.”

“Pues yo he oído decir a los comentaristas en televisión que las eliminatorias son más fáciles para los equipos poderosos como España o Alemania porque muchas veces quedan en un grupo donde les toca jugar contra equipos muy débiles como San Marino o Islas Feroe. Además también dicen en los partidos y en los noticieros deportivos que los equipos sudamericanos están en ascenso, como el caso de Colombia, y además en las ligas más poderosas del mundo hay muchos sudamericanos jugando en clubes europeos.”

“Pero en el caso de España, a ellos les tocó en el mismo grupo con Francia, entonces les tocó duro a los españoles para estar primeros en su grupo.”

“Pero de todos modos cuando en las noticias daban las tablas de las eliminatorias uno veía que todos los equipos acá en Sudamérica estaban con chances de clasificar casi hasta el final, mientras que en Europa hubo equipos que siempre perdían.”

Los alumnos tenían distintos puntos de vista en cuanto a la pregunta de dónde son más reñidas las eliminatorias al campeonato mundial de fútbol. Después de que los alumnos discutieran un poco y dijeran sus opiniones, el investigador propuso entonces trabajar con la tabla final de la eliminatoria que estaba en el archivo de Fathom. *“Vamos a mirar la tabla de posiciones de las eliminatorias tanto en Europa como en Sudamérica e intentaremos responder la pregunta basándonos en los datos que tenemos”*.

Los alumnos comenzaron mirando los datos de algunos países particulares y empezaron a sacar algunas conclusiones.

Un estudiante dijo: *“El último país de la tabla de la eliminatoria de Sudamérica es Paraguay que consiguió 12 puntos. Mientras que Islas Feroe solo obtuvo 1, Andorra y San Marino no sacaron ni un solo punto, ni siquiera empataron, perdieron todo. Entonces obviamente a los países sudamericanos les fue mejor.”*

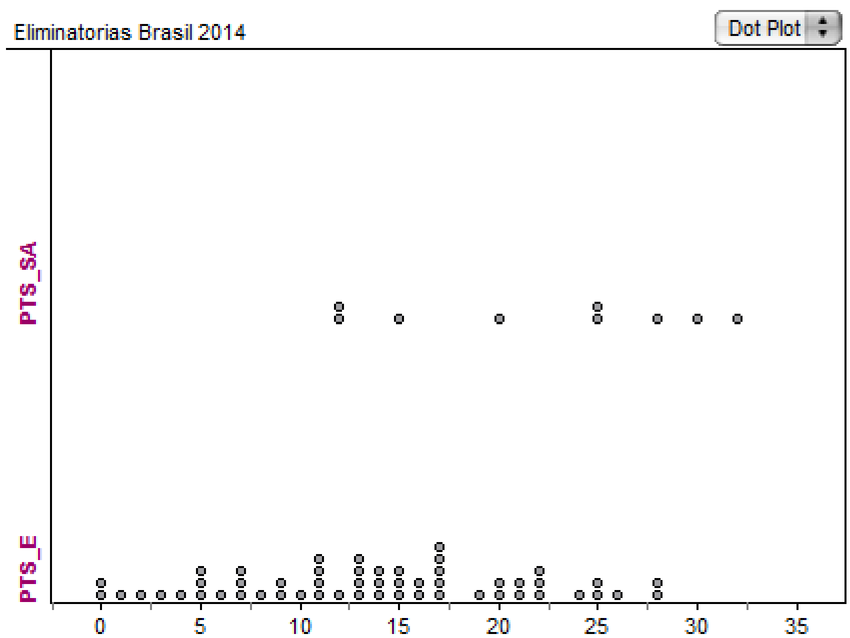
“¿Están todos de acuerdo? ¿Qué pueden decir los que pensaban que las eliminatorias europeas son más reñidas?” dijo el investigador.

“En el grupo A -dijo uno de los alumnos- *Bélgica, que fue el que quedó de primero, le sacó bastante ventaja a los demás, pero el segundo, que es Croacia no le sacó mucha ventaja a los demás.”*

“¿Qué podemos hacer entonces?” preguntó el investigador.

Los estudiantes construyeron un diagrama de puntos del puntaje de todos los equipos en las eliminatorias, solamente distinguiendo si eran equipos europeos o sudamericanos, como lo muestra la figura a continuación.

Figura 28. Diagrama puntaje de eliminatorias: Sudamérica y Europa

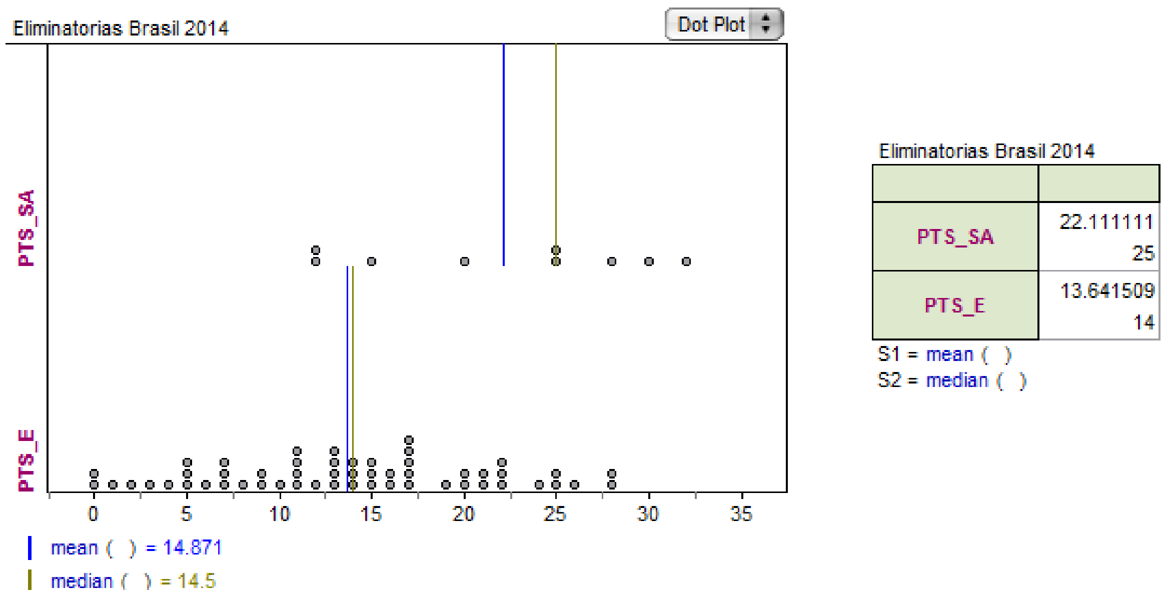


“¿Será que con el diagrama de puntos podemos decir cuál eliminatoria es más reñida?” preguntó el investigador.

“Yo hallé la mediana y la media en el gráfico -dijo una alumna- *pero no sé cuál elegir como representante del puntaje.”*

Los demás alumnos también hallaron estas medidas. *“Entonces ¿cómo elegimos un representante?* -preguntó el investigador- *será que nos sirve tener la mediana que no es sensible a los extremos o la media que es sensible a todos los datos?”*

Figura 29. Media y mediana puntajes eliminatorias



“Pues para Europa las dos líneas están cerquita, pero para Suramérica no es así. Pero de todos modos, habíamos dicho que a veces los equipos más poderosos agarraban mucha ventaja y había también equipos débiles que conseguían muy poquitos puntos. Entonces yo creo que es mejor elegir la mediana” dijo un estudiante.

“Bueno, ya tienen el representante, ¿pero eso de qué sirve para responder la pregunta?” dijo el investigador.

“Pues ahí vemos que los equipos de Sudamérica consiguen más puntos -dijo la alumna que propuso usar las medidas de tendencia central- *y por lo tanto podemos decir que los equipos suramericanos son mejores.”*

“¿Qué opinan los demás, están de acuerdo?” preguntó el encargado de la actividad.

Los alumnos estuvieron pensando un momento. *“No, porque en Sudamérica se juegan más partidos. El representante no nos dice dónde son más reñidas las eliminatorias”* respondió un estudiante.

“¿Qué podemos hacer entonces?” preguntó el investigador.

“Pues queremos saber cuál eliminatoria fue más reñida, más peleada, entonces lo que hay que mirar es en cuáles equipos los puntajes estuvieron más cercanos” dijo un alumno.

“¿Y cómo miramos eso?”. El estudiante dijo: *“en la clase pasada medíamos la distancia entre el dato más pequeño y el más grande”*. Entonces él halló el rango para la eliminatoria europea y para la eliminatoria sudamericana.

A partir del comentario del estudiante se trató de medir entonces la dispersión de los datos. Entre las respuestas de los estudiantes se pueden analizar diferentes formas de percibir y medir la dispersión en conjuntos de datos.

9.3.1 El rango. *“Los dos últimos de Sudamérica, que son Paraguay y Bolivia consiguieron 12 puntos, mientras que el primero que fue Argentina consiguió 32. Hay una diferencia de 20 puntos; ahora para Europa, los equipos que más mal les fueron son Andorra y San Marino que no sacaron ni un solo punto y los que más puntaje sacaron fueron Alemania y Holanda con 28 puntos. El rango es de 28 puntos. Entonces la eliminatoria sudamericana es más difícil porque su rango es más pequeño.”*

El estudiante halló el rango, que fue la primera medida de dispersión que usaron los alumnos intuitivamente desde el taller anterior. Para los alumnos es una primera forma de medir y resumir numéricamente la dispersión de un conjunto de datos.

Otro estudiante pidió la palabra y preguntó: *"¿pero será que sí nos sirve el rango? De todos modos los equipos de Sudamérica son menos que los de Europa y juegan más partidos."*

"¿Qué opina usted?" preguntó el investigador. El alumno dijo: *"pues estamos tratando de mirar qué tan dispersos están los datos, pero con el rango sólo sabemos la distancia entre el último y el primero."*

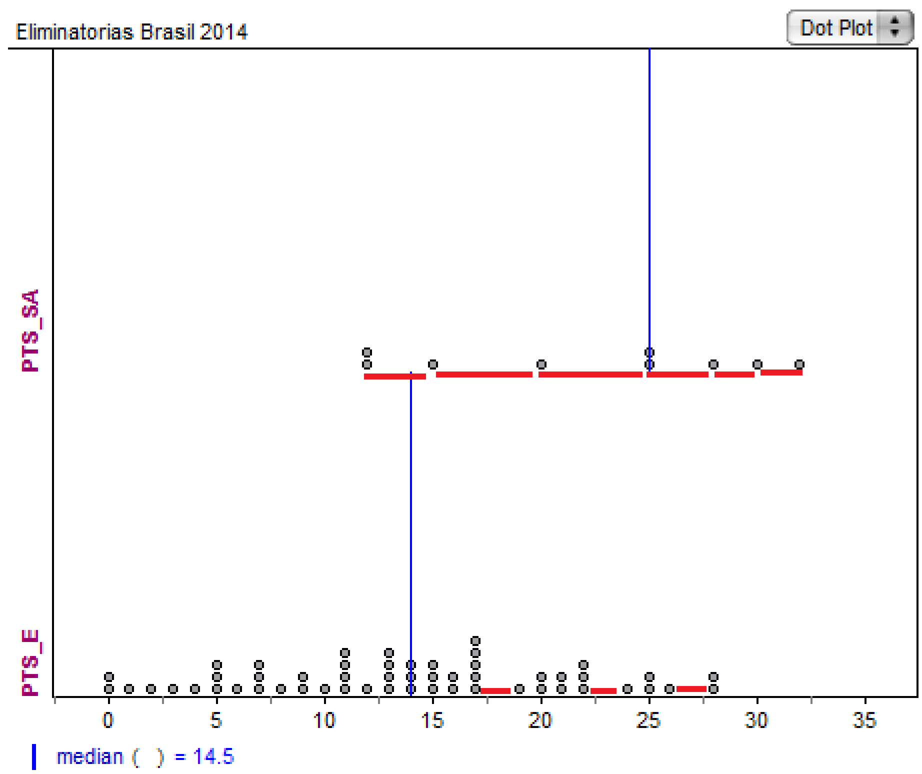
"¿Entonces qué se puede hacer?" dijo el investigador. *"Habría que medir qué tan distantes están los puntos, pero todos"* respondió el alumno. *"¿Y cómo hacemos eso?"* preguntó el encargado de la actividad.

9.3.2 Diversidad entre los datos. Aunque la mayoría de estudiantes tuvo en cuenta el rango para medir la dispersión en el puntaje de las selecciones nacionales en las eliminatorias para decidir en dónde son más reñidas, algunos alumnos no estaban convencidos de que esta medida fuera suficiente para responder a la pregunta, pues sólo mide la distancia de los extremos de cada distribución y no tiene en cuenta los demás datos. La pregunta que se hicieron estos alumnos fue buscar la forma de medir qué tan dispersos están todos los puntos en el diagrama, tener en cuenta todos los datos.

¿Cómo medir qué tan distantes están todos los puntos respecto a los demás?

Los estudiantes se fijaron en la gráfica para ver cómo podían responder esa pregunta. Uno de ellos pidió la palabra, pasó al tablero, proyectó la gráfica que tenía y señaló:

Figura 30. Diversidad de los datos



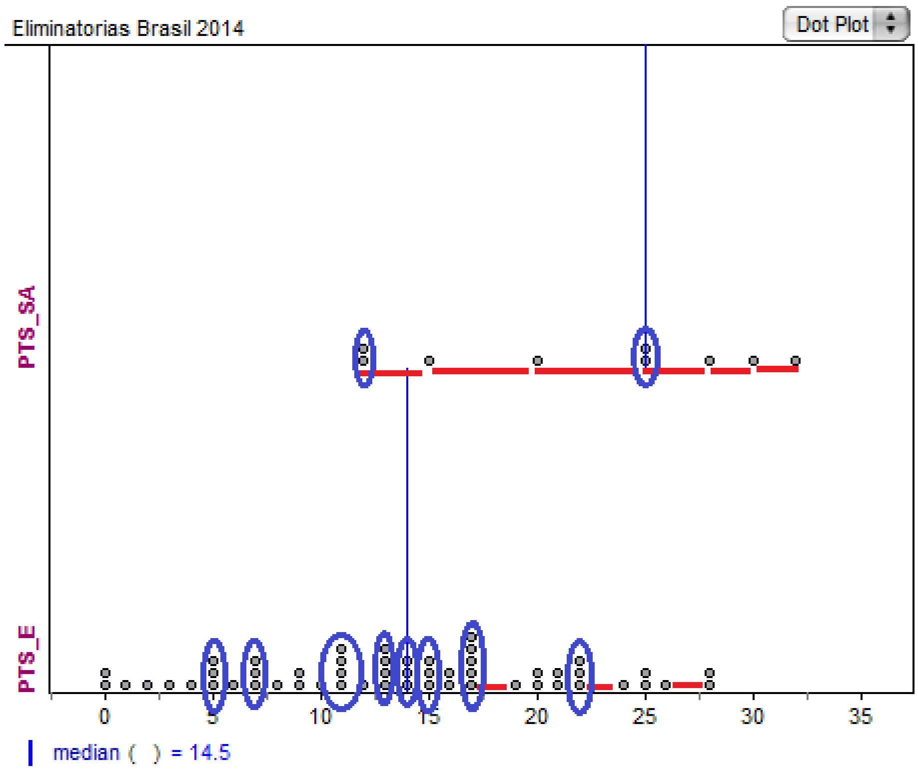
"Yo veo que en Sudamérica hay todos estos espacios (señalando con marcador las líneas rojas de la figura anterior) *y en Europa sólo tengo tres espacios pequeños. Entonces podría yo decir que son más parecidos los puntos en las eliminatorias europeas."*

"¿Qué opinan los demás?" preguntó el investigador.

"Pero de todos modos, en Europa hay 53 equipos y en Suramérica solo hay 9. Entonces por eso es que están esos huecos" dijo una alumna.

Otro compañero pidió pasar al tablero y señaló en la gráfica proyectada: *"pues yo estoy de acuerdo con que son más parecidos los puntos en la eliminatoria europea, porque si miro la gráfica veo que los puntos están más juntos. Además hay más datos que se repiten* (señalando los óvalos azules de la siguiente figura) *en Europa que en Suramérica. Además los puntos que están cerquita de la mediana están muy juntos."*

Figura 31. Datos repetidos



Este grupo de alumnos intentó medir la dispersión de los dos conjuntos de datos teniendo en cuenta: valores repetidos, datos repetidos es muestra de menor variabilidad; distancia entre los datos, menor distancia entre los datos implica menor variabilidad y presencia de huecos indica mayor variabilidad.

Al preguntarle a los estudiantes que primero afirmaron que las eliminatorias Sudamericanas son más reñidas porque su rango es menor si lo que dijeron sus compañeros al mirar la diversidad de los datos, estos dijeron que no cambiaban de opinión.

"Yo sigo creyendo que la eliminatoria es más reñida en Sudamérica porque si uno compara, los primeros no le sacan tanta ventaja a los últimos, todos hicieron más

o menos buen puntaje; pero hay equipos europeos que les fue demasiado mal" dijo uno de los alumnos que tuvo en cuenta el rango como la mejor forma de medir la dispersión.

Algunos de los estudiantes dijeron comentarios parecidos al anterior que hicieron dudar a los alumnos que trataron de medir la dispersión de forma distinta al rango.

"¿Qué pueden decir los que piensan que la eliminatoria europea es más reñida?" preguntó el investigador, pero estos estudiantes no respondieron. *"Bueno, ya que los que decían que la eliminatoria europea fue más reñida no se defienden entonces ¿aceptamos que la eliminatoria sudamericana es la más peleada?"* preguntó de nuevo.

La respuesta de casi todos fue positiva. Al hallar el rango, los alumnos se convencieron de que comparar las distancias entre el primer puesto con el último en cada eliminatoria sería la muestra adecuada para determinar cuál es la más reñida. Esto hizo que los alumnos que buscaron otra forma de ver y medir la dispersión de los datos no le dieran tanta importancia a su respuesta.

9.4 EVALUACIÓN Y CONCLUSIONES DEL TALLER

Los alumnos vieron que para comparar dos grupos de datos no es suficiente hallar un representante. En el caso de esta actividad, el hecho de que en las dos eliminatorias no se jugara la misma cantidad de partidos sirvió para que ellos entendieran esa idea. La pregunta de la clase motivó a los alumnos para buscar la forma de comparar dos grupos midiendo qué tan dispersos son los datos.

La representación de los datos en Fathom permitió que los alumnos exploraran formas distintas de ver y medir la dispersión de conjuntos de datos. Aunque la pregunta también hizo que los estudiantes se fijaran más en el rango que en otras medidas de dispersión, los estudiantes reconocieron intuitivamente otras formas de concebir el concepto de dispersión en un conjunto de datos que son muy interesantes. Sería entonces necesario trabajar más en el concepto de dispersión en las siguientes actividades.

10. QUINTA ACTIVIDAD: *AUTOS*

10.1 OBJETIVO

- Comparar y diferenciar grupos de datos a partir de su representación gráfica, sus valores representantes y su dispersión.

10.2 DESARROLLO DE LA ACTIVIDAD

Para el desarrollo de este taller, los alumnos trabajaron con un archivo de Fathom en el que tenían una tabla con algunos automóviles que se venden en el país. Cada modelo de automóvil contaba con los siguientes datos:

Marca
Modelo
Cilindrada
Precio (en millones de pesos)[2]
Rendimiento (kilómetros por litro)[3]

Debido a que ya los alumnos habían trabajado con conjuntos de datos de diversos tipos manejando Fathom, además de tratar de resolver problemas que sugieren el uso de medidas de tendencia central o dispersión para esta actividad se dejó más libertad a los alumnos para que, con los datos que tenían, plantearan ellos mismos preguntas que quisieran y pudieran responder con la información de la tabla.

A continuación está la tabla de datos de los automóviles que se tenía en el archivo de Fathom.

[2] Fuente: Revista Motor. Disponible en: http://www.motor.com.co/revista-motor/precios/590/ARCHIVO/ARCHIVO-13127676-0.pdf

[3] Fuente: http://www.consumovehicular.cl/?q=comparador

Figura 32. Tabla de datos de automóviles

	Marca	Modelo	Cilindrada	Precio	Rendimiento
1	Chevrolet	Spark GT	1.2	28.990000	15.90
2	Chevrolet	Aveo	1.4	25.990000	14.90
3	Chevrolet	Camaro SS	6.2	141.990000	7.00
4	Chevrolet	Cruze	1.8	45.990000	11.50
5	Chevrolet	Sail	1.4	28.690000	16.10
6	Chevrolet	Sonic	1.6	42.990000	13.30
7	Fiat	Uno Way	1.4	34.300000	13.70
8	Fiat	Palio	1.4	33.900000	12.50
9	Fiat	500	1.4	35.900000	13.20
10	Ford	Fiesta	1.6	37.900000	15.10
11	Ford	Focus	2.0	47.990000	14.70
12	Ford	Fusion	2.0	71.500000	11.05
13	Ford	Mustang	5.0	109.550000	9.10
14	Hyundai	i10	1.1	24.290000	19.60
15	Hyundai	Accent i25	1.6	37.490000	15.60
16	Hyundai	i30	1.6	52.190000	16.00
17	Hyundai	Elantra i35	1.6	45.990000	15.10
18	Hyundai	Veloster	1.6	58.790000	16.10
19	Hyundai	Génesis	2.0	77.590000	10.10
20	Hyundai	i40	2.0	67.990000	12.50
21	Kia	Rio	1.5	26.770000	15.80
22	Kia	Cerato	2.0	54.580000	14.20
23	Kia	Optima	2.0	62.820000	12.20
24	Kia	Cadenza	3.5	92.700000	10.30
25	Kia	Quoris	3.8	164.800000	9.20
26	Mazda	2	1.5	37.950000	14.70
27	Mazda	3	1.6	42.900000	12.90
28	Mazda	6	2.0	71.800000	14.40
29	Mazda	MX-5	2.0	79.500000	11.50
30	Nissan	Tiida	1.6	31.000000	14.70
31	Nissan	March	1.6	27.900000	15.50
32	Nissan	Versa	1.6	37.900000	15.00
33	Nissan	Note	1.6	39.900000	14.70
34	Nissan	Sentra	1.6	24.200000	11.70
35	Nissan	Altima	2.5	79.900000	13.70
36	Nissan	370Z	3.7	115.000000	9.10
37	Renault	Clio	1.6	22.990000	13.70
38	Renault	Twingo	1.2	35.690000	12.50
39	Renault	Sandero	1.6	26.370000	13.90
40	Renault	Stepway	1.6	36.570000	13.00
41	Renault	Logan	1.6	28.140000	13.60
42	Renault	Fluence	2.0	48.990000	13.00
43	Renault	Mégane	2.0	56.830000	12.40
44	Volkswagen	Gol	1.6	29.900000	13.70
45	Volkswagen	Voyage	1.6	29.900000	13.70
46	Volkswagen	Jetta	2.0	35.990000	12.20

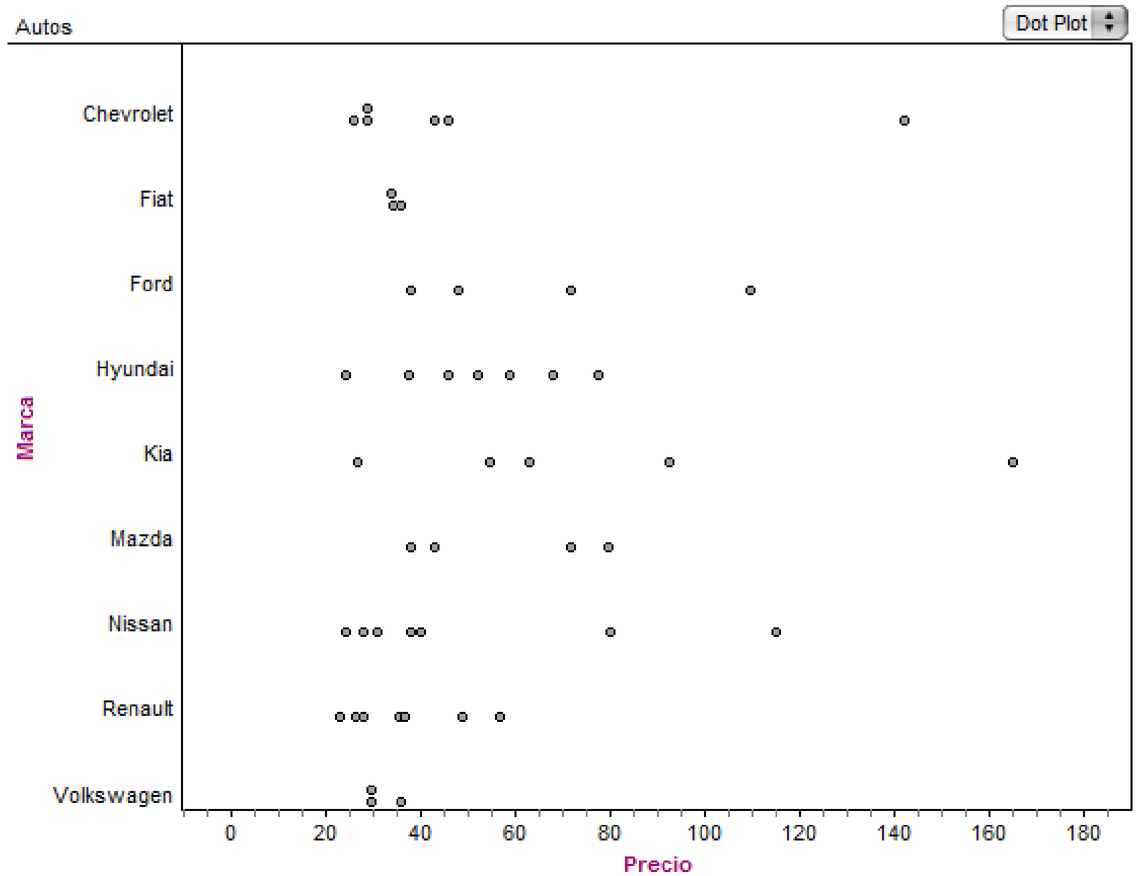
Al tratar de encontrar características de los datos y responder a diferentes inquietudes, los alumnos siguieron el mismo proceso que seguían en actividades anteriores, cuando el investigador era quien planteaba las preguntas: primero, miraban la tabla de datos y se preguntaban por cosas muy particulares, por ejemplo encontrar el automóvil más caro o cuál carro consume más gasolina; en un segundo momento construían gráficas de los datos; y por último, se plantearon preguntas que trataran describir datos como conjunto.

10.3 *¿QUÉ MARCA DE CARROS ES LA MÁS BARATA? ¿CUÁL ES LA MÁS CARA?*

Después de que algunos alumnos se preguntaran cuál es el carro más caro y cuál es el más barato, un estudiante propuso la pregunta de encontrar la marca de carros más económica y la más costosa. El investigador preguntó si es posible responder a la pregunta con los datos que se tenían.

El estudiante respondió: *"Por ahora yo construí una gráfica donde en el eje horizontal tengo los precios de los carros y en el eje vertical tengo la marca de los carros"*.

Figura 33. Diagrama de puntos precio distinguiendo marcas



Este estudiante construyó un diagrama de puntos con los precios de todos los automóviles agrupándolos por su respectiva marca. Al mostrarle su procedimiento hasta el momento en el tablero, los demás estudiantes construyeron en sus computadores la misma gráfica.

"Ya tenemos un diagrama de puntos con los precios de los carros y los tenemos clasificados por marcas, ahora ¿cómo podemos saber cuál es la marca más barata o la más cara?" preguntó el investigador.

Uno de los alumnos pidió la palabra y pasó al tablero. *"Kia tiene el carro más caro de todos* (señalando en la gráfica el punto del modelo con el mayor precio) *y Renault tiene el de menos precio* (señalando el punto del modelo menos costoso).*"*

“¿Y qué pasa con que Kia tenga el modelo más caro y Renault el más barato? ¿De qué nos puede servir?” preguntó el encargado de la actividad.

“Es que si el modelo más caro lo tiene Kia y el más barato lo tiene Renault hace pensar que lo más seguro es que esas sean la marca más cara y la más barata.” Respondió el alumno.

“¿Y todos piensan que la marca más barata es Renault y la más cara es Kia?” preguntó el investigador.

Los alumnos estuvieron mirando más detenidamente la gráfica para sacar conclusiones. Uno de ellos dijo: *“Yo no creo que la marca más barata sea Renault. Puede que tenga el modelo más barato de todos, que es un poco más de 20 millones, pero el carro más costoso de los Renault cuesta cerca de 60 millones. Y si uno mira por ejemplo Fiat, todos cuestan un poco más de 30 millones, entonces Renault no creo que sea la marca más barata.”*

Otro estudiante añadió: *“Sí. Además con los modelos de Volkswagen pasa más o menos lo mismo, pero los precios de los Fiat son más parecidos, los puntos están más pegados.”*

Con estas intervenciones de los alumnos se puede ver que en esta actividad, al igual que en las anteriores, el primer paso que usan para describir y comparar conjuntos de datos es el de graficar y mirar datos particulares, especialmente los extremos de cada distribución.

“¿Entonces cuál es la marca más barata, Renault, Fiat o Volkswagen? ¿Cómo podemos decidir?” preguntó el investigador.

“Podemos hallar un representante de cada marca y mirar cuál es el menor” dijo una estudiante.

Los alumnos entonces trataron de encontrar y añadir en la gráfica un precio que represente cada marca de automóviles con ayuda de Fathom.

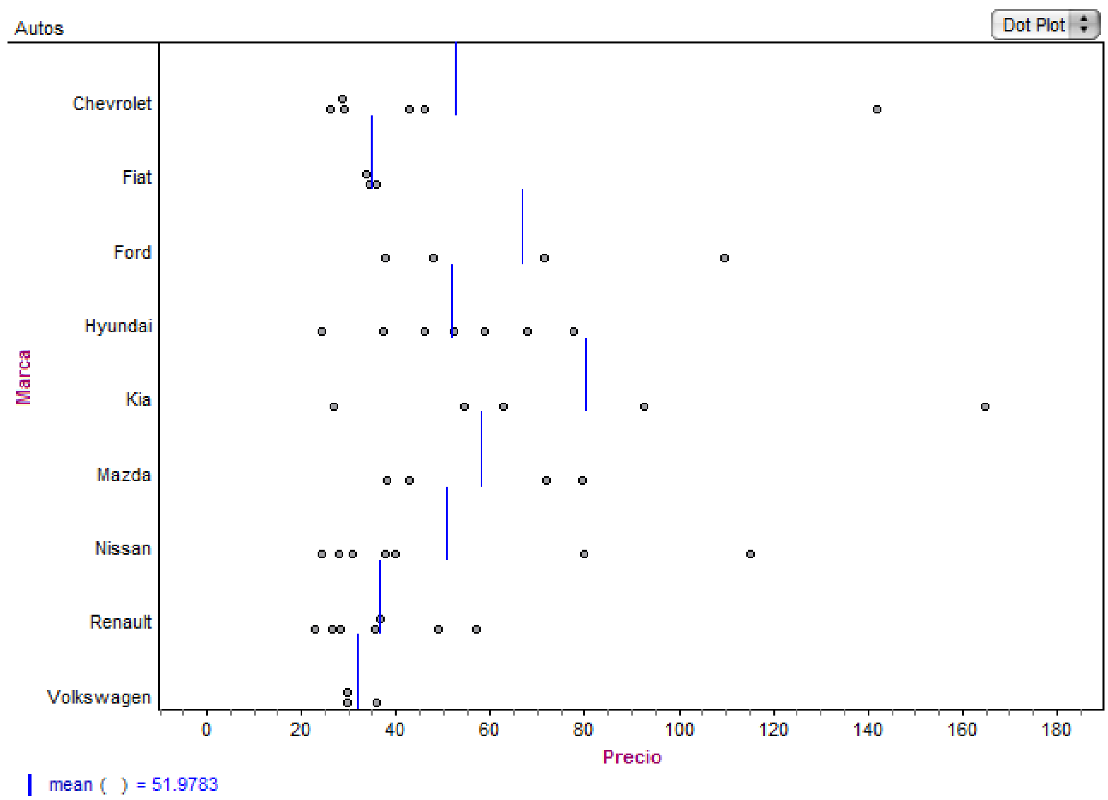
“¿Tiene sentido pensar en la moda como representante de los precios de cada marca?” Preguntó el investigador.

La respuesta de los estudiantes fue negativa.

“No porque no hay precios que se repitan muchas veces en los carros de una misma marca.”

Los estudiantes con ayuda de Fathom hallaron la media y la ubicaron dentro del diagrama de puntos.

Figura 34. Diagrama de precios y media de cada marca



Autos

	Marca									Row Summary
	Chevrolet	Fiat	Ford	Hyundai	Kia	Mazda	Nissan	Renault	Volkswagen	
Precio	52.44	34.7	66.735	52.047143	80.334	58.0375	50.828571	36.511429	31.93	51.978261

S1 = mean ( )

“¿Y por qué usan la media y no la mediana u otro representante?” preguntó el encargado de la actividad.

Los alumnos estuvieron pensando por un momento hasta que un estudiante dijo: *“pues yo creo que todos los precios tienen que influir, además si llegaran a cambiar los precios de los carros la media también cambiaría”*.

Este estudiante recordó la segunda actividad, en el momento donde se introdujo la media como medida de tendencia central[4], donde se miraban los sueldos de un equipo de fútbol y se suponía el cambio de los salarios para mirar el comportamiento de esta.

“¿Pero qué pasa por ejemplo con Chevrolet? ¿Será realmente buen representante la media?” preguntó el investigador, pero los alumnos insistieron que el precio más alto de esta marca también debería influir para encontrar un representante y además aludieron de nuevo a la suposición de que pudieran cambiar los precios de los carros.

Los estudiantes al tener la media como representante de los precios de cada marca respondieron la pregunta.

“La marca más barata es Volkswagen porque tiene la menor media y la más cara es Kia porque tiene la mayor. Además podríamos decir en orden las marcas de acuerdo al precio.” dijo uno de los estudiantes.

Los estudiantes terminaron esta parte de la actividad ordenando las marcas de automóviles por precio.

“¿Y qué pasó con Renault? ¿No tenía el carro más barato?” preguntó el investigador. Los estudiantes se dieron cuenta de que tener el modelo de menor precio no implica ser la marca más económica.

---

[4] Ver sección **La media** (p.63) en el capítulo **7**.

Ya respondido el interrogante de saber cuál es la marca más costosa y cuál la más económica, el investigador propuso la siguiente pregunta para medir la dispersión de los precios de los carros.

10.4 *¿CUÁL MARCA DE CARROS PRESENTA PRECIOS MÁS VARIABLES?*

Los alumnos primero observaron el diagrama de puntos para obtener algunas primeras conclusiones. Algunos estudiantes no entendían la pregunta y se les dificultó idear estrategias para responder. Para que el problema no resultara tan complicado, el investigador decidió plantear esta pregunta: *¿Cuál marca maneja precios más parecidos en sus modelos?*

Los estudiantes vieron el gráfico y respondieron.

"Fiat es la marca que tiene los precios más parecidos porque los precios de los tres carros están muy pegados, mientras que en el resto de marcas no tanto" dijo una alumna.

El investigador entonces comentó: *"bueno, ya sabemos cuál marca tiene precios más parecidos en sus carros, pero ahora miremos todo lo contrario, cuál es la marca que tiene precios más distintos."*

Los estudiantes de nuevo se fijaron en el diagrama de puntos. Uno de los alumnos pidió la palabra y dijo: *"Yo creo que la marca que tiene los precios más distintos es Kia porque hay mucha distancia entre el modelo más barato y el más caro."*

Otro estudiante añadió: *"podemos encontrar el rango y responder la pregunta."*

La mayoría de estudiantes procedió entonces a calcular el rango de los precios de los carros de cada marca.

Figura 35. Media y Rango

Autos

	Marca									Row Summary
	Chevrolet	Fiat	Ford	Hyundai	Kia	Mazda	Nissan	Renault	Volkswagen	
Precio	52.44	34.7	66.735	52.047143	80.334	58.0375	50.828571	36.511429	31.93	51.978261
	116	2	71.65	53.3	138.03	41.55	90.8	33.84	6.09	141.81

S1 = mean ( )
S2 = max ( ) − min ( )

“¿Entonces la marca que tiene precios más distintos es Kia?” Preguntó el investigador.

La mayoría de los estudiantes se contentaron hasta el momento con el rango para medir la dispersión de los precios, pero un alumno pidió pasar al tablero y dijo: *“pero hay una cosa, por ejemplo, yo veo que la distancia entre el más caro de los Kia y el anterior, no el más barato, sino el que va antes de él en precio, es grande, pero con el más caro de Chevrolet y el anterior la distancia todavía es más grande. Además los dos más caros de Nissan están lejos del resto.”*

Otro estudiante añadió: *“además por lo menos en Hyundai la distancia entre los precios de todos los carros es más o menos la misma.”*

En ese momento los estudiantes empezaron a dudar del rango como medida de dispersión. Empezaron a medir la distancia entre datos que no fueran los extremos de cada distribución.

“¿Y no hay una fórmula para medir la distancia entre todos los datos?” preguntó un estudiante.

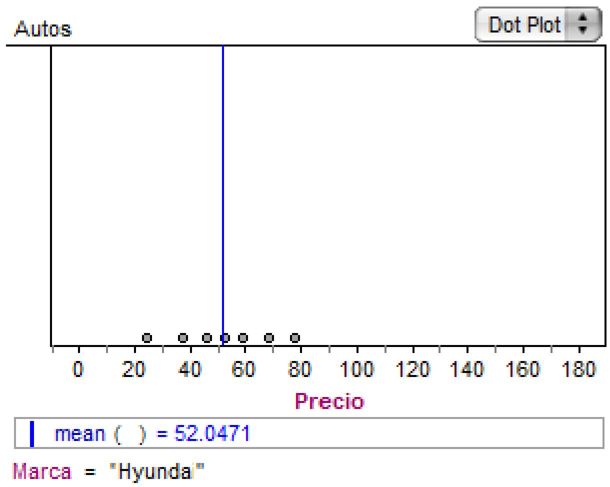
El investigador dijo: *“¿Cómo lo harían ustedes? ¿Cómo harían para tener en cuenta las distancias entre todos los datos?”*

Los estudiantes pensaron un poco pero se desanimaron rápidamente.

“Bueno... ¿Y qué tal si nos fijamos en la distancia entre los datos y el representante de cada marca?” propuso el investigador.

Se sugirió a los alumnos construir un diagrama de puntos de los precios de una sola marca y se señaló la media para tratar de medir la dispersión de los datos respecto a su representante.

Figura 36. Diagrama de puntos precios Hyundai



El encargado de la actividad hizo una serie de preguntas con el objetivo de definir con los alumnos la desviación media.

"¿Cómo podemos medir la distancia entre el precio más alto y la media?"

La mayoría de los alumnos no sabía qué responder. *"El precio más alto menos la media"* respondió uno de los estudiantes.

"¿Y cuál es la distancia entre el menor precio y la media?"

"La media menos el precio más pequeño" respondió el mismo estudiante.

"¿Y cuál es la distancia entre uno de los precios, cualquiera que escojamos, y la media?" se preguntó.

Ninguno de los alumnos se atrevió a responder.

"¿Cuál es la distancia entre este precio (señalando un punto a la derecha de la media en el diagrama de puntos) *y la media?"* dijo el investigador.

“El precio menos la media” participó uno de los estudiantes.

“¿Y cuál es la distancia entre este precio (señalando un punto a la izquierda de la media) *y la media?”* preguntó de nuevo el investigador.

“La media menos ese precio” contestó uno de los alumnos.

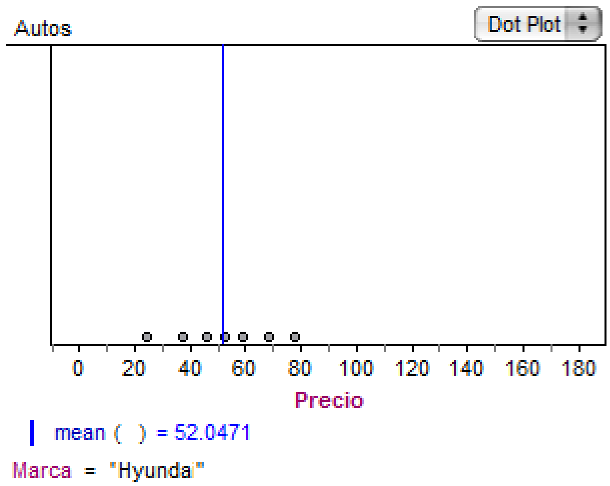
“Entonces, ¿cuál es la distancia entre uno de los precios, cualquiera que escojamos a la derecha o a la izquierda de la media, y la media?” volvió a preguntar el encargado de la actividad.

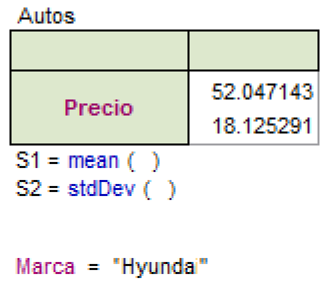
Los alumnos quedaron de nuevo confundidos.

Debido a no poder lograr el objetivo de definir la desviación media con los estudiantes, el investigador en ese momento comenzó a hacer de una forma más dirigida. Se dijo a los estudiantes que se iba a ver en el programa una herramienta para medir qué tan cercanos o qué tan lejanos están los puntos respecto a la media, mas no se trabajó en Fathom con la desviación media sino con la desviación estándar.

El investigador les indicó a los estudiantes cómo hallar con Fathom la desviación estándar, sin decirles la fórmula de esta medida de dispersión. Simplemente se dijo que se usaría este número para saber cuán cercanos o lejanos están los datos de la media.

Figura 37. Diagrama de puntos y desviación estándar



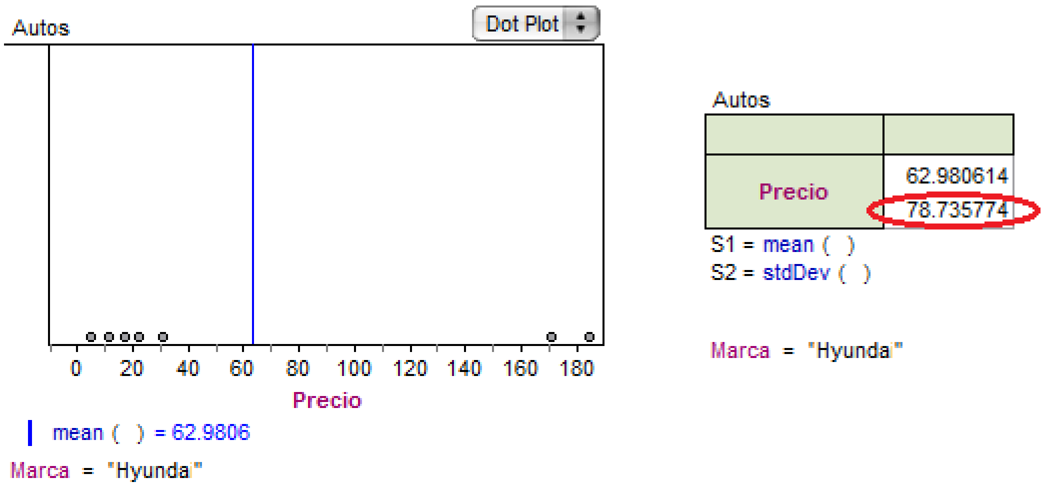


10.4.1 La desviación estándar. Para conocer algunas de las propiedades de esta medida de dispersión, el encargado de la actividad pidió a los estudiantes que movieran los puntos del diagrama de los precios de los autos de una de las marcas e hizo unas preguntas para que ellos obtuvieran conclusiones.

¿Qué ocurre cuando se alejan los datos de la media?

Los estudiantes empezaron a mover los puntos del diagrama y se dieron cuenta que entre más lejos estén los valores de la media mayor será el valor de la desviación estándar.

Figura 38. Aumento de valor de la desviación estándar

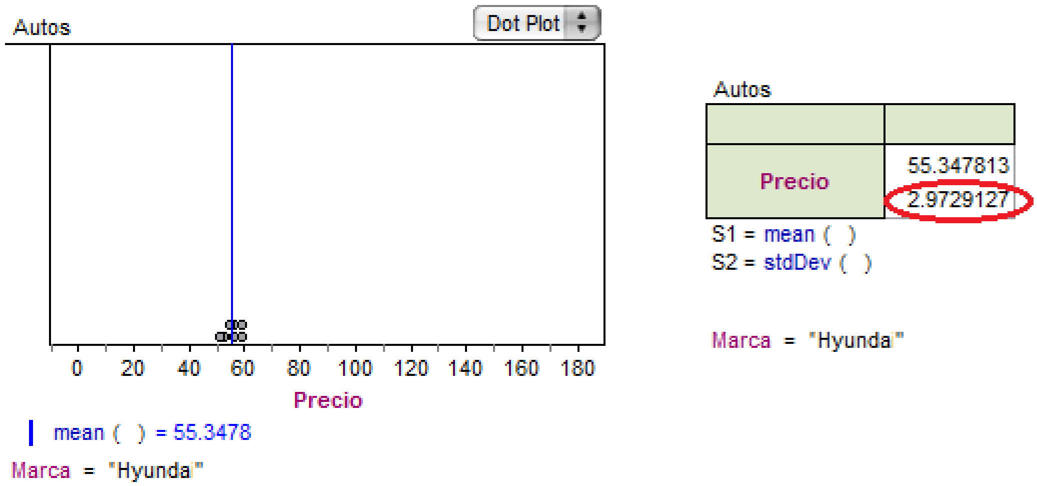


Los estudiantes vieron que al cambiar los precios, también lo hace la media (conclusión que ya se había obtenido en la segunda actividad). Pero si se mueven los puntos de manera que queden alejados de esta, el nuevo valor que tienen, la desviación estándar, va a aumentar su valor.

¿Qué ocurre cuando se acercan los datos a la media?

Los estudiantes volvieron a mover los puntos de manera que estuvieran lo más cerca que pudieran de la media. Ellos se dieron cuenta que entre más juntos estuvieran los puntos, menor sería el valor de la desviación estándar.

Figura 39. Disminución valor de la desviación estándar

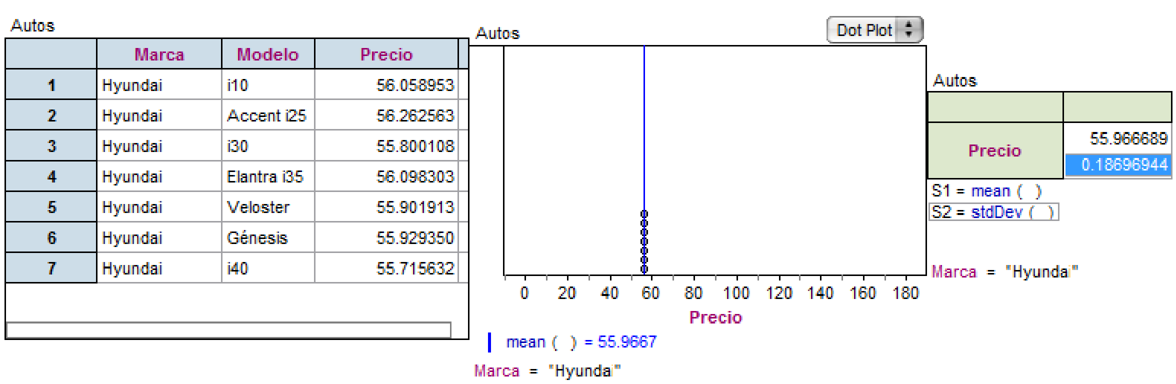


Los alumnos vieron que al cambiar los datos de forma que quedaran cada vez más cerca, haría que el valor de la desviación estándar fuera cada vez más pequeño.

¿Es posible que el valor de la desviación estándar sea cero?

Los alumnos de nuevo movieron los puntos del diagrama para intentar que el valor de la desviación estándar fuera cero. Como habían reconocido ya que entre más cercanos estuvieran los puntos, el valor de esta sería cada vez más pequeño, trataron de mover los puntos para que los precios fueran iguales.

Figura 40. Valor mínimo de la desviación estándar

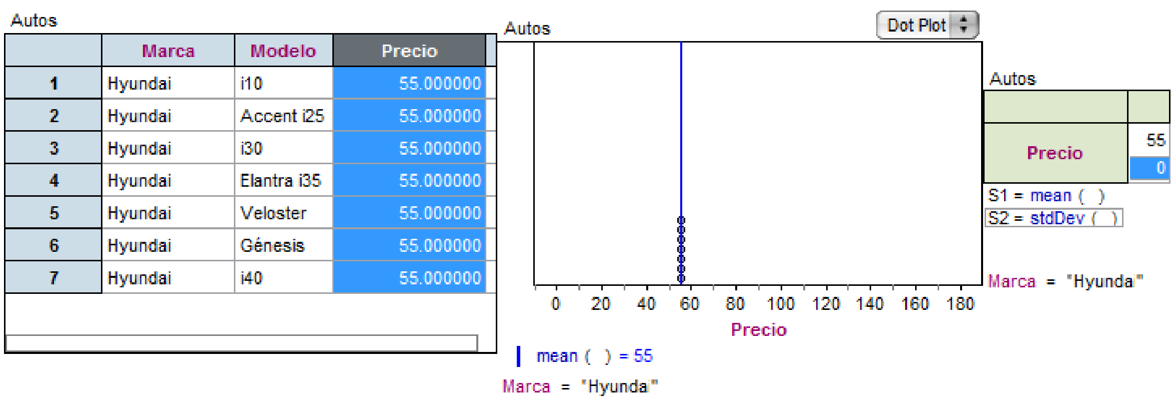


Autos

	Marca	Modelo	Precio
1	Hyundai	i10	56.058953
2	Hyundai	Accent i25	56.262563
3	Hyundai	i30	55.800108
4	Hyundai	Elantra i35	56.098303
5	Hyundai	Veloster	55.901913
6	Hyundai	Génesis	55.929350
7	Hyundai	i40	55.715632

Los alumnos, a pesar de mover los puntos de manera que estuvieran lo más cerca posible, no eran capaces de que el valor de la desviación estándar fuera exactamente cero. Esta situación se dio por un tiempo, hasta que uno de los estudiantes se dio cuenta, al mirar la tabla de datos, de que por más cercanos que tuvieran los puntos, no tenían el mismo valor. Por lo tanto él procedió a cambiar los valores desde la tabla de manera que todos los carros tuvieran el mismo precio.

Figura 41. Valor nulo de la desviación estándar


Autos

	Marca	Modelo	Precio
1	Hyundai	i10	55.000000
2	Hyundai	Accent i25	55.000000
3	Hyundai	i30	55.000000
4	Hyundai	Elantra i35	55.000000
5	Hyundai	Veloster	55.000000
6	Hyundai	Génesis	55.000000
7	Hyundai	i40	55.000000

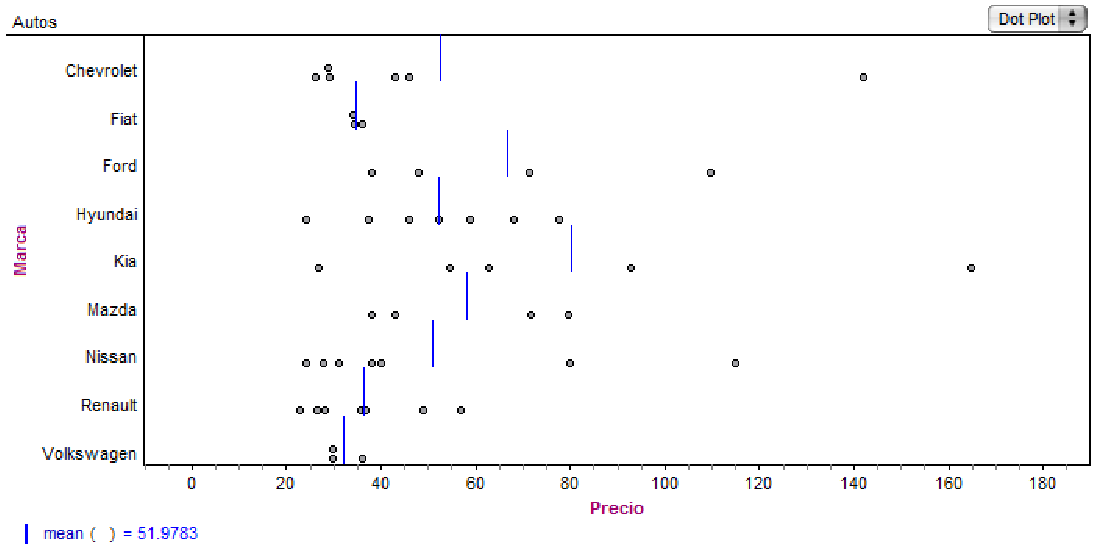
El alumno que cambió los valores directamente de la tabla pasó al tablero y describió a sus compañeros lo que había hecho. Los estudiantes entonces concluyeron que el valor de la desviación estándar es cero solamente cuando todos los precios son el mismo.

A partir de esta parte de la actividad los alumnos conocieron la desviación estándar como una medida que sirve para saber qué tan dispersos están los datos respecto a la media.

10.4.2 Dispersión de los precios. Al haber conocido la desviación estándar como una forma de medir la variabilidad de los datos distinta al rango, los alumnos tuvieron otro criterio para responder a la pregunta de cuál marca de carros presenta precios más distintos.

Se volvió a trabajar con los estudiantes el archivo original con todos los modelos de carros y procedieron a hallar la desviación estándar de los precios de cada marca.

Figura 42. Diagrama, media, rango y desviación estándar



Autos

	Marca									Row Summary
	Chevrolet	Fiat	Ford	Hyundai	Kia	Mazda	Nissan	Renault	Volkswagen	
Precio	52.44	34.7	66.735	52.047143	80.334	58.0375	50.828571	36.511429	31.93	51.978261
	116	2	71.65	53.3	138.03	41.55	90.8	33.84	6.09	141.81
	44.64006	1.0583005	31.825889	18.125291	52.741193	20.67766	33.856743	12.416358	3.5160631	31.282598

S1 = mean ( )
S2 = max ( ) − min ( )
S3 = stdDev ( )

Los alumnos notaron que la marca de carros que más presenta variabilidad en los precios con respecto a la media es Kia, que reforzó la primera respuesta que se había dado con el rango y se rechazó la idea de que la Chevrolet lo fuera, a pesar que su precio más alto estuviera más alejado del resto.

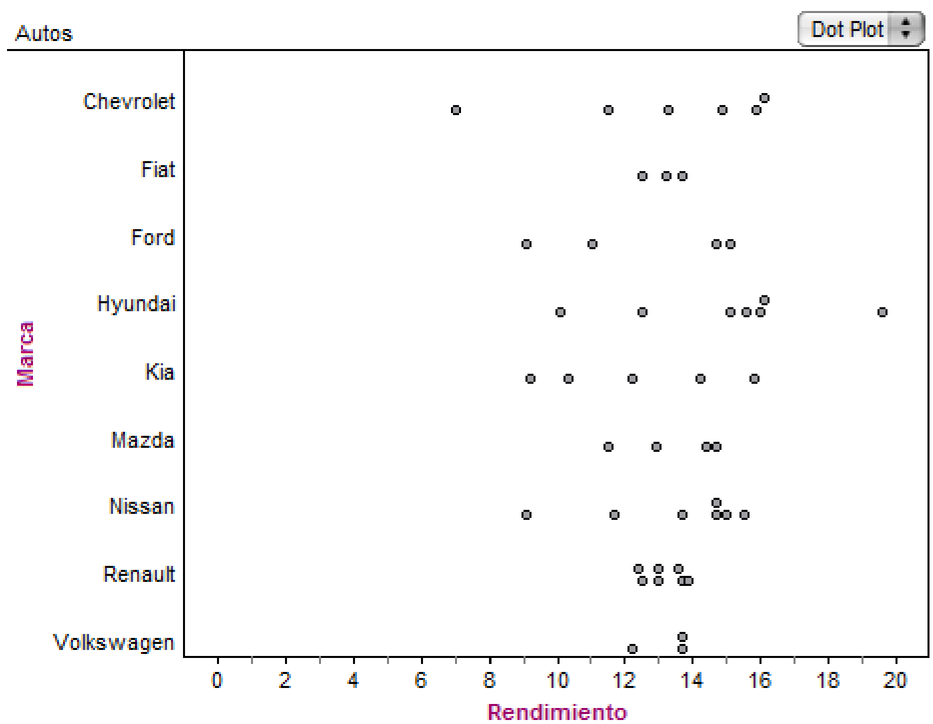
Con la pregunta de cuál es la marca que tiene precios más distintos en sus autos y cuál tiene costos más parecidos sirvió para que los alumnos midieran de alguna forma la dispersión de los datos y tuvieran en cuenta criterios distintos al rango, en este caso la desviación estándar.

Para continuar con el taller de nuevo se dejó que los estudiantes se preguntaran sobre características que les interesaran de los datos. Un aspecto que causó curiosidad fue el consumo de gasolina de los carros.

10.5 ¿CUÁL MARCA VENDE CARROS MÁS EFICIENTES EN EL CONSUMO DE GASOLINA?

En un comienzo los estudiantes construyeron un diagrama de puntos con el rendimiento de los carros y distinguiendo cada marca. Trataron mirar datos particulares, como cuál es el carro que más consume gasolina y cuál es el que consume menos, pero no tenían claridad sobre qué significa el rendimiento.

Figura 43. Diagrama de puntos rendimiento



El investigador explicó que esa medida es el promedio de cuánta distancia puede recorrer cada modelo con un litro de gasolina a un velocidad determinada. Así, tener un número alto en el rendimiento quiere decir consumo eficiente de gasolina y tener un número bajo significa alto consumo.

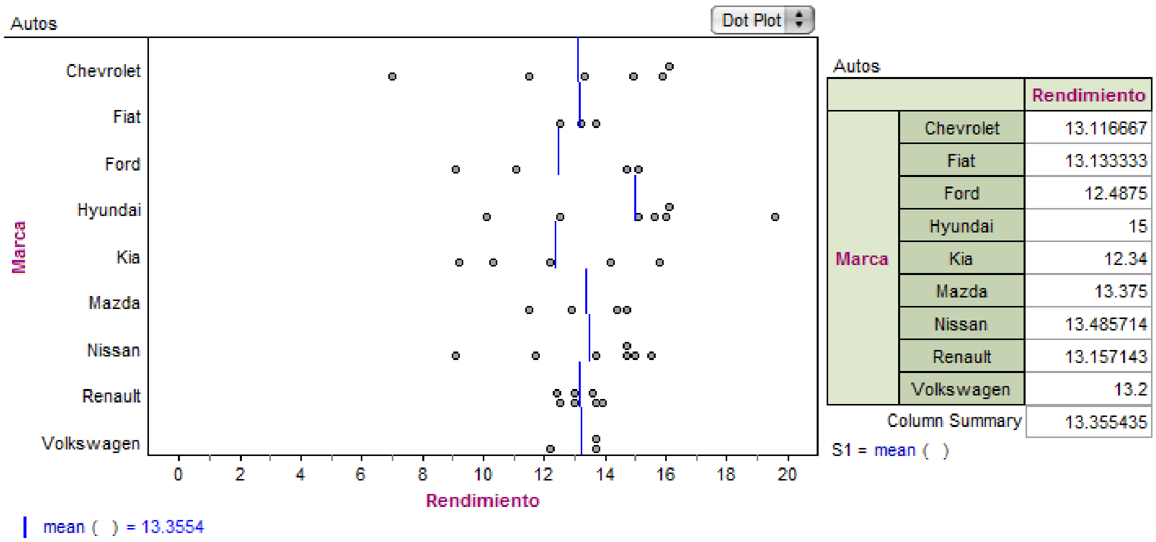
Los alumnos al comprender ya lo que significa el rendimiento, miraron en el diagrama de puntos y en la tabla de datos cuál modelo tiene el consumo de gasolina más eficiente y cuál es el que consume más gasolina. Después se interesaron en saber cuál es la marca que ofrece a los clientes autos más eficientes en el gasto del combustible.

"¿Podemos con los datos que tenemos responder la pregunta?" preguntó el investigador.

"Sí. Ya tenemos el rendimiento de cada carro y tenemos cada marca separada en la gráfica. Entonces tendríamos que sacar un representante de cada marca y mirar cuál es el más grande." Dijo uno de los estudiantes.

Debido a que los alumnos resolvieron el primer problema de saber cuál es la marca más cara y cuál la más barata, esta pregunta fue muy sencilla para ellos, pues sugería el mismo procedimiento. Los estudiantes hallaron y señalaron en el diagrama de puntos la media del rendimiento de los carros de cada marca y miraron cuál era la de mayor valor.

Figura 44. Rendimiento: la media en el diagrama de puntos



Al haber encontrado características de los datos que causaran curiosidad en los alumnos, el investigador propuso al final de la clase una pregunta con la intención de conocer, saliendo un poco del tema de investigación, algunas respuestas o reacciones de los estudiantes a la hora de relacionar dos variables cuantitativas.

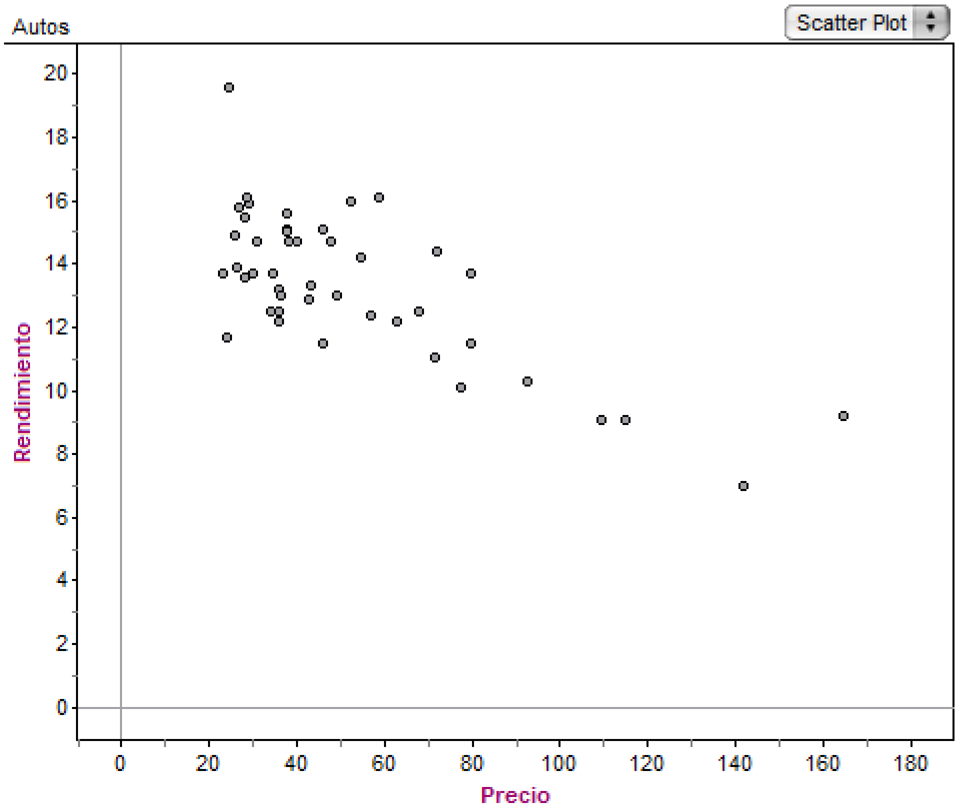
¿Qué carros tienen mejor precio, los de consumo eficiente de gasolina o los de alto consumo?

10.6 RENDIMIENTO Y PRECIO

La mayoría de estudiantes no sabía qué hacer para responder esa pregunta. Algunos miraban la tabla de datos, pero al tener los precios y el rendimiento sin orden no eran capaces de idear alguna estrategia. Otros construyeron diagramas de puntos de estas variables por aparte y hallaban la media de los datos, pero después se dieron cuenta de que esto no serviría para encontrar alguna relación.

Dos alumnos que estaban trabajando juntos pidieron al encargado de la actividad que se acercara para mostrarle lo que habían hecho. Habían construido un gráfico de dispersión relacionando el rendimiento con el precio de los carros.

Figura 45. Gráfico de dispersión: rendimiento y precio



El investigador pidió a los alumnos que pasaran al tablero a describirle a sus compañeros lo que habían hecho.

"Nosotros construimos un gráfico con el precio en el eje horizontal y el rendimiento en el eje vertical."

Al ver el gráfico de dispersión en el tablero los alumnos vieron que a mayor precio menor rendimiento.

"A medida que se hace mayor el precio, el rendimiento se hace cada vez menor."

"¿Y eso se cumple para todos los carros?" preguntó el investigador.

"No. Por ejemplo el carro más caro es de mayor rendimiento que el segundo más caro. Pero viendo la gráfica, uno espera que los carros más caros consuman más gasolina." Respondió un alumno.

Aunque la relación entre las dos variables no es fuerte, los estudiantes la notaron. Es importante tener en cuenta que los alumnos de ese grado no habían visto el concepto de función, ni sus representaciones. Este hecho puede ser fundamental para que los alumnos vean la relación, aunque el comportamiento de que mayor precio implique menor rendimiento o viceversa no se dé en todos los modelos. Tal vez si ya hubieran visto ese concepto, se les dificultaría ver esa relación.

10.7 EVALUACIÓN Y CONCLUSIONES DE LA ACTIVIDAD

Al tratar de encontrar características relevantes de los datos, los alumnos siguieron los mismos pasos que seguían en actividades anteriores: primero, mirar la tabla de datos y mirar casos particulares; en un segundo momento construir gráficas; y por último, plantear preguntas que sugieran describir los datos como conjunto.

Los estudiantes se fijaron en que para medir la dispersión de los datos, el rango puede no ser suficiente y en ocasiones es necesario tener en cuenta otros criterios para concebirla como la distancia entre datos diferentes a los extremos y qué tan cerca se encuentran del representante

El grupo conoció una nueva forma para medir la dispersión de los datos: la desviación estándar. Esta indica qué tan cercanos o qué tan distantes están los valores con respecto a la media. Entre más dispersos se encuentren los datos de la media, la desviación estándar será mayor; cuanto más cercanos se encuentren, esta será menor; y esta medida tiene valor cero solamente cuando todos los valores son iguales.

Los alumnos usaron medidas de tendencia central y de dispersión para comparar diferentes grupos de datos y encontrar características que les llamaran la atención. Fathom fue una herramienta de gran ayuda para que los estudiantes tuvieran una representación de los conceptos que estaban trabajando y así conocer propiedades importantes de las mismas. La información que se tenía y las preguntas que se plantearon durante la clase insinuaron la exploración y el uso de esas medidas.

11. SEXTA ACTIVIDAD: *¿CUÁNTO MIDE 8-05?*

11.1 OBJETIVOS

- Ver la necesidad de usar tanto las medidas de tendencia central como las medidas de dispersión para describir un conjunto de datos.
- Definir la desviación media.

11.2 DESARROLLO DE LA ACTIVIDAD

En un principio se planteó la pregunta a los estudiantes. El encargado de la actividad preguntó cómo se podría responder; *"hay que averiguar la estatura de todos"* dijo uno de los estudiantes. Se procedió entonces a pasar cada alumno al frente para medirlo y señalando su estatura en el tablero. Una vez hecho esto los alumnos registraron en una tabla de Fathom las alturas de todos los compañeros.

Figura 46. Tabla de datos estaturas de los estudiantes

Estatura 8-05

	Nombre	Altura	<new>
1	Camilo	1.72	
2	Danilo	1.70	
3	Cristian	1.65	
4	Sebastián	1.69	
5	Julián	1.64	
6	Luis	1.60	
7	Natalia	1.60	
8	Richard	1.61	
9	Catalina	1.55	
10	Natalia	1.55	
11	Alejandra	1.54	
12	Jhon	1.41	
13	Johana	1.49	

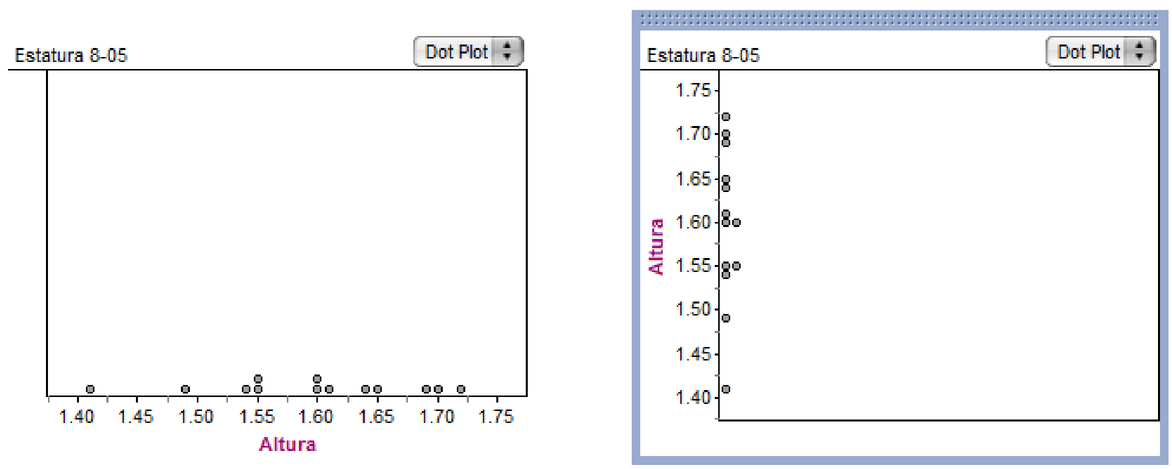
Después de haber registrado en Fathom la estatura de todos, los estudiantes procedieron a elaborar con el programa una representación que mostrara mejor la distribución.

"¿Es suficiente tener la lista de las alturas para responder la pregunta?" preguntó el investigador.

"No. Para ver mejor los datos es más fácil con una gráfica." Dijo una estudiante.

Los estudiantes construyeron un diagrama de puntos con las alturas.

Figura 47. Diagramas de puntos de estaturas horizontal y vertical



Los estudiantes construyeron estos dos diagramas, pero prefirieron después el de la derecha, pues las alturas están registradas en el eje vertical, además de que previamente en el tablero se había hecho una representación similar cuando se registró la estatura de cada uno.

"Entonces ¿cuál es la estatura del curso 8-05?" dijo el investigador.

"Podemos encontrar un representante." Respondió un estudiante.

"¿Qué criterios podemos tener en cuenta para escoger una altura que represente a todo el grupo?" intervino de nuevo el investigador.

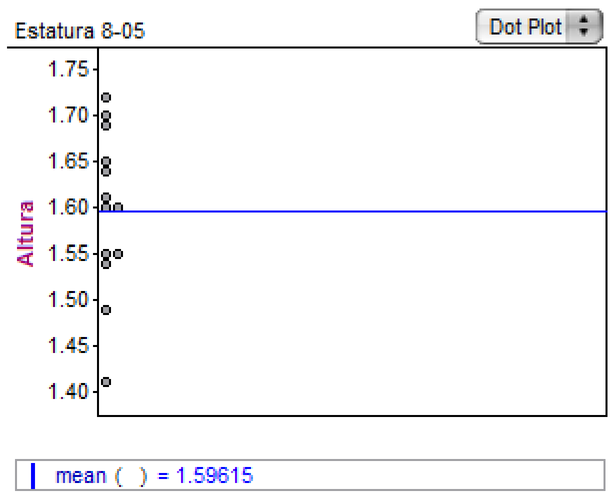
“Pues hemos visto la moda, la mediana y la media.” Dijo el estudiante.

“¿Y cuál elegimos?”

“Usar la moda no sirve. Porque las alturas repetidas solo se repiten dos veces.” Participó otro estudiante. Él vio que para encontrar un representante de las estaturas del grupo no tiene sentido pensar en las modas, pues solamente se repiten dos veces.

Algunos estudiantes hallaron la media de los datos.

Figura 48. Media



“El representante sería 1.59615 que es el promedio de las alturas.” Dijo un estudiante.

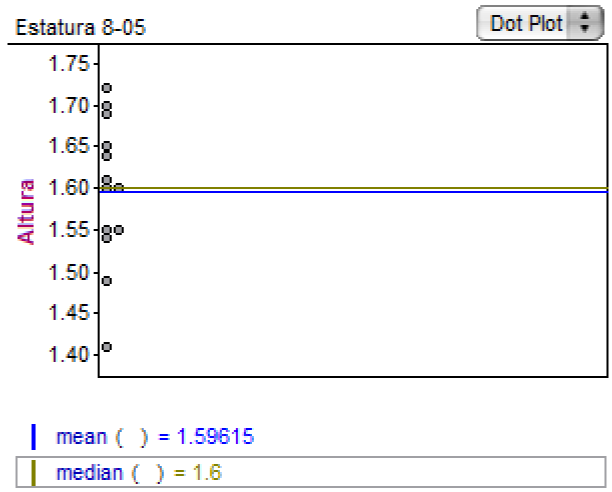
“¿Están todos de acuerdo?” preguntó el investigador.

Otro alumno intervino: *“yo pienso que el representante debe ser 1.59 porque cuando se mide a una persona se dice que mide 1.50, 1.58, 1.60,... pero nunca uno ve una estatura de 1.59615.”*

Una estudiante dijo: *“pues yo creo que el representante debe ser 1.60 porque la media está más cerca de 1.60 que de 1.59.”*

Uno de los estudiantes pidió la palabra y dijo: *"yo también creo que el representante es 1.60 y no 1.59 porque además 1.60 es la mediana."*

Figura 49. Media y mediana



La media y la mediana de las estaturas de los alumnos resultaron ser muy parecidas, por lo tanto estuvieron de acuerdo en pensar que el representante es 1.60, además coincidió con la estatura de dos estudiantes.

11.3 *¿QUÉ TAN BUENO ES EL REPRESENTANTE?*

"¿Quiénes son los que miden 1.60?" preguntó el investigador. Los dos estudiantes alzaron la mano y él le pidió a uno de los dos que pasara al tablero.

"Natalia mide 1.60, podemos decir que es la representante de la altura del grupo -dijo el investigador- *ahora vamos a ver qué tan buen representante es."*

"Necesito otro alumno que pase al tablero, puede ser el más alto." El investigador pasó al frente al alumno más alto del grupo.

"¿Cuánto mide Camilo?" preguntó el investigador. Los estudiantes contestaron: 1.72.

“Ahora, ¿qué distancia hay entre la altura de Camilo y la de Natalia?” Los estudiantes contestaron: *“12 centímetros”*.

“Necesito ahora que pasa al tablero el más bajito.” El investigador pasó al frente al estudiante de menor estatura.

El investigador preguntó *“¿Cuánto mide Jhon?”* Los alumnos respondieron: 1.42.

“¿Y cuánto le falta a Jhon para tener la misma altura de Natalia?” Los alumnos contestaron: *“18 centímetros”*.

“Bien. Ahora, ¿cuál es la distancia promedio de Camilo y de Jhon con respecto a Natalia?”

“La distancia promedio es 15.” Algunos estudiantes argumentaron esa respuesta diciendo que la distancia de 12 a 15 es la misma que la distancia de 15 a 18. Otros hicieron la suma de las dos distancias y dividieron en dos.

“Ahora, ¿qué distancia hay entre la altura de Natalia y el representante?” preguntó el investigador.

“¿Cómo así? -Dijo un estudiante- *si la altura de Natalia es el mismo representante”.* El investigador dijo: *“entonces, ¿cuál es la distancia?”* A lo que el alumno contestó: *“pues cero.”*

“Bien. Y ahora ¿cuál es la distancia promedio de Camilo, Jhon y Natalia con respecto a Natalia?”

Los alumnos hallaron el promedio de las tres distancias, algunos usando calculadora. *“10 centímetros”* contestaron.

“Necesito que otra persona más pase al tablero. ¿Quién quiere pasar?”

Una estudiante pasó. El investigador de nuevo preguntó: *“¿Cuánto mide Catalina?”* Los alumnos contestaron: 1.55.

“¿Cuál es la distancia entre la altura de Catalina y la de Natalia?” preguntó el investigador. Los alumnos respondieron: *“5 centímetros.”*

“¿Y cuál es la distancia promedio de Camilo, Jhon, Natalia y Catalina con respecto a Natalia?” los alumnos hallaron la media de las cuatro distancias: *“8.75 centímetros.”*

“Muy bien. Ahora ¿cómo hacemos para hallar la distancia promedio de todos con respecto a Natalia que es la representante?” Preguntó el investigador.

“Sumando la distancias de las alturas a la altura de Natalia y dividiendo entre 13.” Respondieron los estudiantes.

Los estudiantes se dieron cuenta de que, así como pueden usar la media de un conjunto de datos como valor representativo, también pueden hallar el promedio de la distancia de estos con respecto al representante.

Ahora la idea que trabajó el investigador con los estudiantes era construir la fórmula de la desviación media. *“¿Cómo podemos escribir el promedio de las distancias de las estaturas respecto al representante por medio de operaciones?”* preguntó.

“Hay que sumar cada distancia de cada altura al representante y el resultado dividirlo entre 13 que es el número de estudiantes que estamos” dijo un estudiante.

“Bueno entonces por ahora tenemos algo así” dijo el representante anotando en el tablero lo siguiente:

$$\frac{suma\ de\ distancia\ de\ cada\ dato\ al\ representante}{número\ de\ datos}$$

“¿Y cómo se calcula la distancia entre un dato y el representante?” preguntó.

“El dato menos el representante” dijo un estudiante.

“¿Seguro? ¿Qué pasa si el dato es menor que el representante?” dijo el investigador.

“Que la resta va a dar negativo y no hay distancias negativas” participó una alumna.

“¿Y cómo hacemos para que siempre tengamos un valor positivo?” preguntó de nuevo el investigador.

Los estudiantes no supieron responder en ese momento.

“¿No recuerdan ustedes lo que usan para tomar el valor de un número sin que importe el signo? Cuando le ponen una barrita a cada lado del número.” Dijo el encargado de la actividad.

“Ah. Ya. El valor absoluto.” Dijo uno de los estudiantes.

Así, la fórmula quedó de la siguiente fórmula

$$\frac{suma\ de\ |dato - representante|}{\#\ de\ datos}$$

“Al promedio de las distancias de los datos al representante se le llama desviación media. Ahora vamos a calcular la desviación media de las estaturas con Fathom. ¿Cómo escribiríamos la fórmula para las alturas?” Propuso el investigador.

Un estudiante pasó al tablero y escribió lo siguiente:

$$\frac{suma\ de\ |estatura - 1.60|}{13}$$

“... ¿le ven algún problema a la fórmula?” preguntó el investigador. Los alumnos no respondieron.

“¿Qué pasaría si cambio los datos?” preguntó de nuevo.

“Que el representante ya no sería 1.60 -dijo una estudiante- *hay que cambiar ese 1.60 por la media o por la mediana.”*

Así la forma de hallar la desviación media con Fathom sería de esta forma:

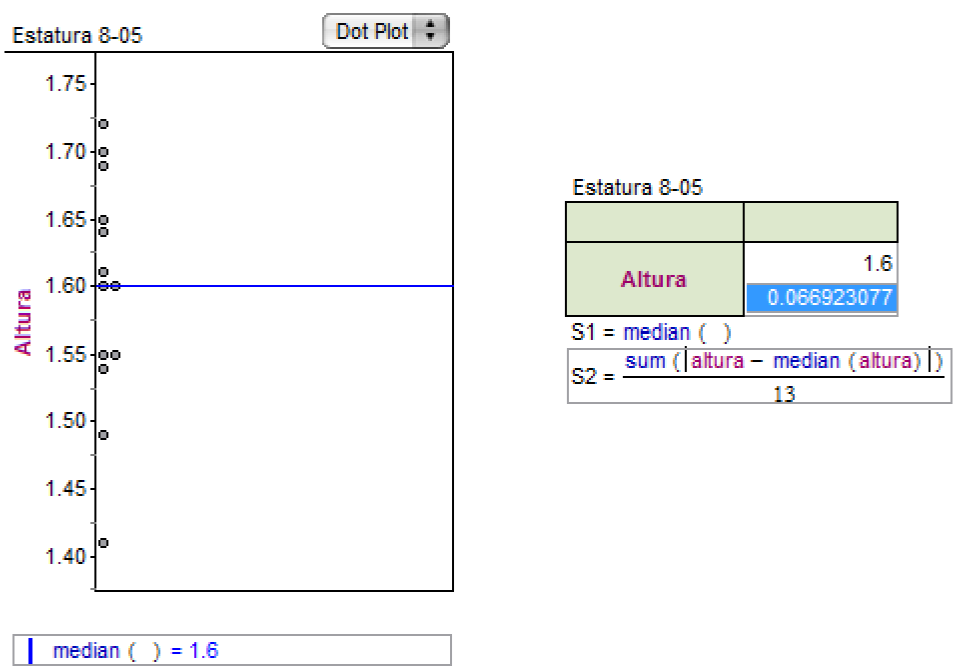
$$\frac{suma\ de\ |altura - media|}{13}$$

Ó

$$\frac{suma\ de\ |altura - mediana|}{13}$$

Los alumnos procedieron a hallar la desviación media de las estaturas.

Figura 50. Desviación media



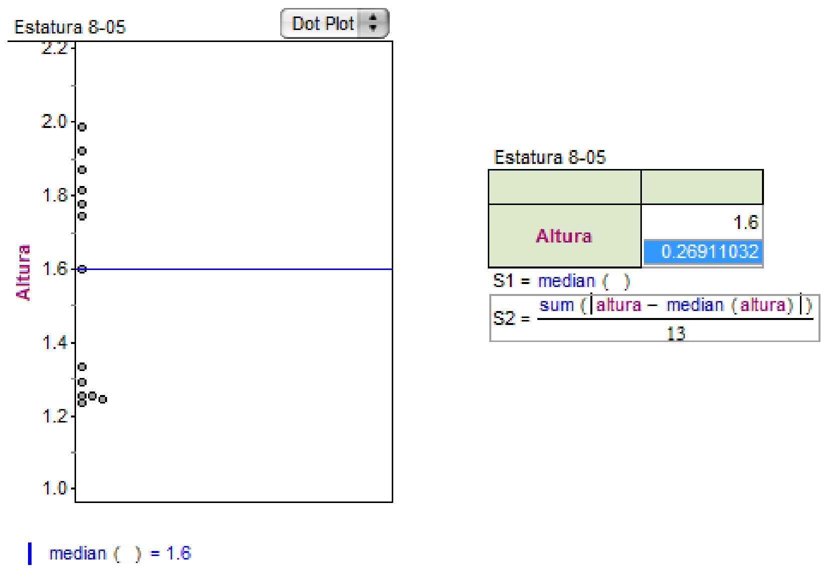
“¿Cuánto es la desviación media de la estatura entonces?” preguntó el investigador.

“0.0669 metros... casi 7 centímetros” respondió uno de los alumnos.

Enseguida el encargado de la actividad pidió a los estudiantes que cambiaran las estaturas moviendo los puntos del diagrama para ver el comportamiento de la desviación media.

“Traten de mover los puntos de manera que estén cada vez más lejos del representante. ¿Será ahora el valor de la línea mejor o peor representante que antes?”

Figura 51. Aumento de valor desviación media



“Peor, porque los puntos están lejos de él. El representante no se parece casi a los datos.” Respondió un estudiante.

“¿Qué pasa con el valor de la desviación media?”

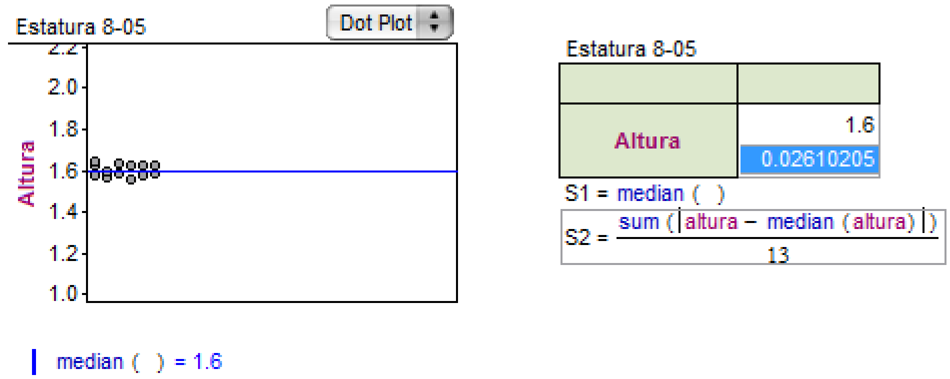
“El valor de la desviación aumenta” dijeron los estudiantes.

“Ahora traten de mover los puntos de manera que estén cerca del representante. ¿Ese valor representa mejor o peor al grupo?”

“Mejor, los puntos están cerca al representante y entonces se parecen más.”

“¿Y qué pasa con el valor de la desviación media cuando las estaturas son parecidas?”

Figura 52. Disminución valor desviación media

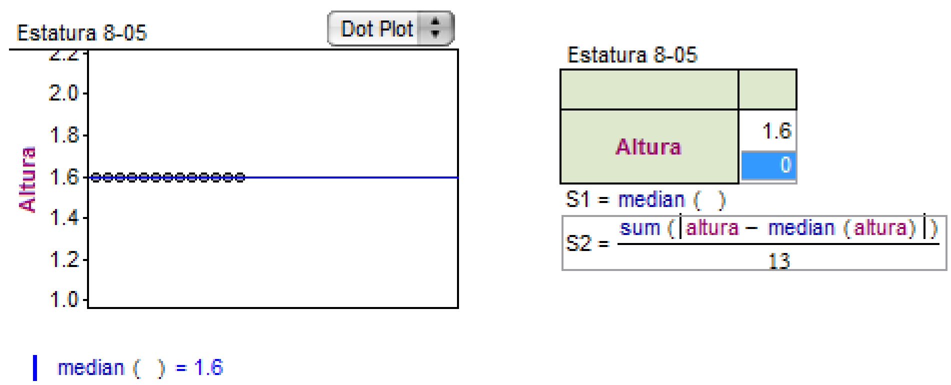


“Disminuye” respondieron.

“¿Es posible que el valor de la desviación media sea cero?”

“Todos las estaturas tendrían que ser iguales” respondieron los alumnos.

Figura 53. Valor nulo de la desviación media



Los alumnos se dieron cuenta que la desviación media mide qué tan parecidos son los datos al representante que eligieron y tiene notaron que su valor tiene un comportamiento similar al de la desviación estándar. Este sería un nuevo criterio para medir la dispersión de conjuntos de datos, diferentes de los que habían considerado antes (rango y diversidad de los valores).

11.4 EVALUACIÓN Y CONCLUSIONES DEL TALLER

Los estudiantes en la primera parte de la actividad fortalecieron la idea de buscar un valor central para describir un conjunto de datos.

En este último taller se reforzó con los alumnos la concepción de dispersión teniendo en cuenta la distancia de los datos al representante. Se valoraría después en la evaluación final cuál criterio tienen más en cuenta al medir la dispersión de conjuntos de datos.

Se definió con los estudiantes la desviación media y se vio cómo se comporta su valor.

12. EVALUACIÓN FINAL

Con el fin de analizar los conocimientos adquiridos por los estudiantes con el desarrollo de las actividades y medir el impacto del uso del paquete estadístico Fathom en el proceso, se diseñó una evaluación final que sugiere el uso de medidas de tendencia central y de dispersión.

La evaluación consta de dos problemas y se planteó para ser resuelta en 50 minutos. Debido a que durante todos los talleres se promovió el uso de Fathom, lógicamente el examen fue diseñado en este entorno para que los estudiantes utilizaran las herramientas del programa para su desarrollo. Debido a la poca disponibilidad de espacio y de tiempo en el salón de matemáticas, se hizo necesario que los alumnos trabajaran por parejas.

12.1 LOS PROBLEMAS

Problema 12.1.1 La siguiente tabla contiene la altura (en metros) de los jugadores de cinco equipos de la NBA.

¿Cuál es el equipo con mayor estatura?

¿Cuál es el equipo con menor estatura?

¿Cuál equipo tiene estaturas más variables o dispersas?

JUSTIFIQUE CADA UNA DE SUS RESPUESTAS

Figura 54. Estaturas NBA

NBA

	CELTICS	BULLS	HEAT	MEMPHIS	LAKERS	<new>
1	1.80	1.88	1.85	1.85	1.91	
2	1.85	1.90	1.88	1.91	1.91	
3	1.91	1.88	1.93	1.98	1.88	
4	1.93	1.91	1.96	1.96	1.93	
5	1.96	2.01	1.96	1.93	1.98	
6	1.96	2.01	2.03	2.03	1.98	
7	1.96	2.06	2.03	2.01	2.03	
8	2.01	2.06	2.03	2.06	2.01	
9	2.06	2.08	2.08	2.06	2.01	
10	2.03	2.08	2.08	2.08	2.03	
11	2.06	2.06	2.03	2.08	2.06	
12	2.06	2.08	2.08	2.13	2.11	
13	2.13	2.11	2.06	2.16	2.13	
14	2.10		2.08		2.08	
15			2.13		2.13	
16					2.13	

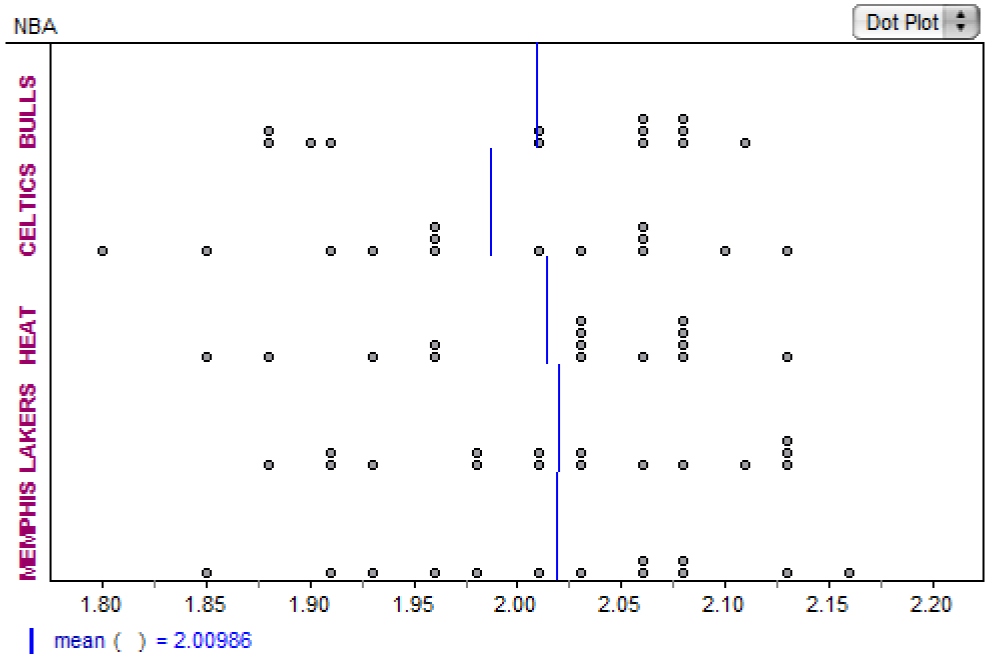
Problema 12.1.2 *Se tiene dos conjuntos diferentes de bloques A y B. Las longitudes de los bloques en el conjunto A fueron 10, 20, 30, 40, 50 y 60 cm y las longitudes de los bloques en el conjunto B fueron 10, 10, 10, 60, 60 y 60 cm. ¿Cuál de los dos conjuntos presenta mayor variabilidad? JUSTIFIQUE SU RESPUESTA*

12.2 RESPUESTAS DE LOS ESTUDIANTES Y ANÁLISIS

Para el desarrollo del primer problema los estudiantes trataron de encontrar un valor representativo para cada equipo de baloncesto y ver cuál es el mayor y el menor. Todos los alumnos construyeron una representación gráfica de los datos.

Algunos estudiantes utilizaron la media como valor representativo de la estatura de los jugadores de cada equipo de básquet.

Figura 55. Diagrama de puntos estaturas equipos y media



El equipo que presenta mayor estatura es el LAKERS porque su media nos muestra que su dato es 2.01935 y es mayor al de los otros.

El equipo que presenta menor estatura es el equipo CELTICS porque su media muestra que su dato es 1.987149 y eso lo hace menor que los otros.

El equipo con mayor estatura son los LAKERS porque el valor de la media es el más grande.

Una de las parejas respondió a la pregunta mirando cuál era el equipo que tenía mayor suma en las estaturas de sus jugadores.

LAKERS es el equipo con mayor estatura porque la suma de los metros de todos los jugadores es la mayor suma.

Algunos de los estudiantes se fijaron en las modas y valores repetidos de cada grupo.

HEAT es el equipo con mayor altura de los jugadores porque tiene dos columnas e cuatro jugadores que miden igual.

La respuesta anterior muestra que estos estudiantes pensaron en mirar cuál equipo presentaba la mayor cantidad de datos repetidos en la mayor altura posible.

Otros estudiantes lograron solamente fijarse en datos particulares, sobre todo en los extremos de todos los valores y mirar a cuál equipo pertenecen el jugador más alto y el de menor estatura.

El equipo el cual creemos que tiene mayor estatura es el MEMPHIS porque la mayor estatura que hay en ese equipo es 2.16 y es el más alto de todas las estaturas.

El equipo que nosotros creemos que tiene menor estatura es el CELTICS porque la medida que menor tiene es 1.80.

El equipo con menor estatura son los CELTICS porque empiezan con menor estatura y terminan con menor estatura.

A la hora de comparar grupos de datos la mayoría de estudiantes trató de encontrar un valor que representara a cada equipo -casi todos usaron la media- y comparar cada uno de los representantes de los equipos. Se logró que casi todos los alumnos comprendieran la idea de buscar un valor central para identificar y comparar grupos de datos. Unos pocos estudiantes se quedaron con la idea de buscar los extremos o datos particulares en la distribución.

Para responder la última pregunta del primer problema y el segundo problema los estudiantes debían medir de alguna forma la dispersión de los datos. A continuación se verá las consideraciones que tuvieron ellos para resolver estas preguntas.

En cuanto a saber cuál es el equipo con más variabilidad en sus estaturas, la mayoría de estudiantes usó el rango para responder.

El equipo que tiene más dispersión o variabilidad es el equipo de CELTICS porque nos muestra que su dato es 0.33 y tiene mayor rango a los otros equipos.

Estos dos estudiantes hallaron el rango de cada uno de los equipos y se fijaron en el equipo que mayor tuviera esta medida de dispersión.

Algunos estudiantes en el problema de saber cuál equipo presenta más dispersión en la estatura de sus jugadores fueron más allá y hallaron la desviación estándar para medir qué tan cercanos o lejanos están los datos respecto a la media.

Figura 56. Media, desviación estándar y rango

NBA

CELTICS	1.9871429 0.094740826 0.33
BULLS	2.0092308 0.085679785 0.23
HEAT	2.014 0.080958543 0.28
MEMPHIS	2.0184615 0.089707502 0.31
LAKERS	2.019375 0.083703345 0.25

S1 = mean ( )
S2 = stdDev ( )
S3 = max ( ) − min ( )

Viendo la tabla anterior que contiene la media, la desviación estándar y el rango de las estaturas de los equipos de básquet, estos estudiantes ratificaron con las dos últimas su respuesta del equipo con estaturas más variables, al ver el equipo que tiene el mayor rango tiene también la mayor desviación estándar.

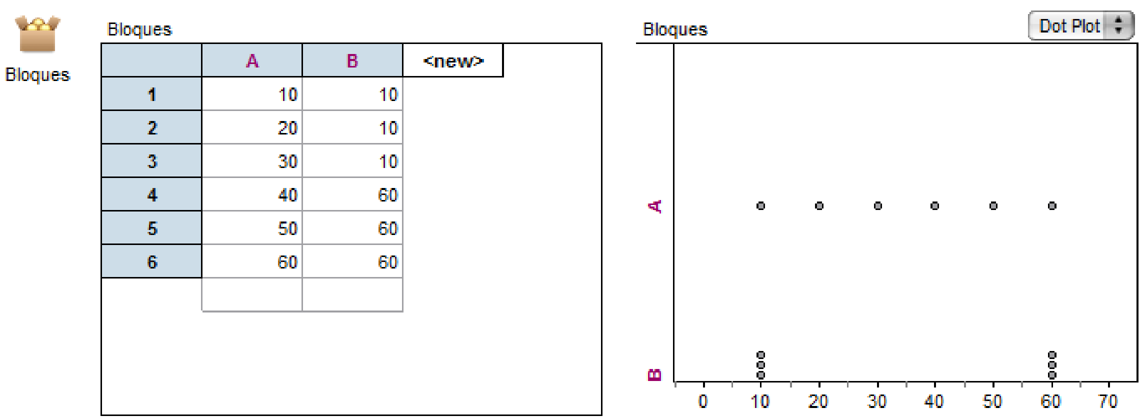
Otros estudiantes siguieron concibiendo la idea de dispersión como la ausencia o menor cantidad posible de valores repetidos.

MEMPHIS es el equipo de mayor variabilidad porque los jugadores están más dispersos.

El equipo con estaturas más variables es los MEMPHIS porque están más separados unos de otros.

En cuanto al segundo problema, es importante recordar que este era parte del diagnóstico, con el objetivo de reconocer si los estudiantes cambiaron o no su forma de percibir el concepto de variabilidad. Lo primero que hicieron los alumnos fue crear un diagrama de puntos de los dos conjuntos de bloques. A diferencia de la última parte del primer punto, el rango de los dos conjuntos de datos es el mismo.

Figura 57. Tabla de datos y diagrama de puntos Bloques



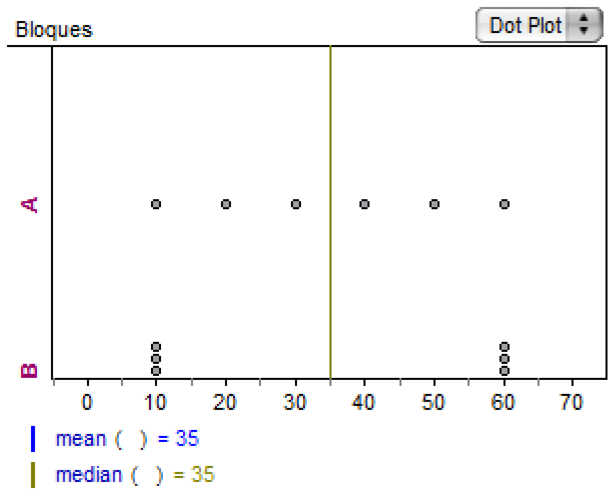
	A	B	<new>
1	10	10	
2	20	10	
3	30	10	
4	40	60	
5	50	60	
6	60	60	

Debido a las respuestas de algunos estudiantes se confirma que algunos siguen considerando el concepto de dispersión o variabilidad como la diversidad entre los datos.

El conjunto A porque están más variables el uno del otro, como lo muestra la gráfica.

Incluso se notó de nuevo que algunos alumnos consideraron los conjuntos como iguales, o igual de variables, debido a que la media y la mediana son la misma en los dos.

Figura 58. Media y mediana iguales en los dos conjuntos de bloques



Los dos, porque la variabilidad del bloque A es igual al bloque B, ya que cada suma de su longitud es 35.

Los dos presentan la misma variabilidad porque suman igual A y B.

Algunos estudiantes dieron su respuesta teniendo en cuenta la distancia entre los datos, unos de otros.

El conjunto B es el de mayor variabilidad porque de 10 a 60 hay mucha más diferencia que de 10 en 10 hasta 60.

La mitad de los estudiantes resolvió el problema teniendo en cuenta la distancia de los datos a su representante. Algunos tuvieron como referencia la desviación estándar.

Figura 59. Media, desviación estándar y rango Bloques

Bloques

A	35 18.708287 50
B	35 27.386128 50

S1 = mean ( )
S2 = stdDev ( )
S3 = max ( ) − min ( )

Los dos bloques tienen el mismo rango y la misma media pero la variabilidad es mayor en el bloque B.

La mayoría de los estudiantes que dieron una respuesta satisfactoria a este punto usaron la desviación estándar y no la desviación media, debido seguramente a que la primera es mucho más sencilla de escribir en una fórmula en Fathom.

Otros de los estudiantes que respondieron a la pregunta, tuvieron en cuenta la distancia de los datos respecto al representante viendo el diagrama de puntos sin usar ninguna fórmula para medir la dispersión.

Es B porque está más disperso de la mediana y tiene el mismo conjunto de lado a lado.

Esta respuesta es muy interesante. Estos estudiantes no usaron la desviación estándar o la desviación media para responder a la pregunta. Viendo solamente el diagrama, estos dos estudiantes notaron que el conjunto B presenta más dispersión de los datos respecto a la media que el conjunto A, sin necesidad de un cálculo numérico, lo que muestra una buena comprensión del concepto de variabilidad en cuanto a la distancia de los valores a su representante.

Con el desarrollo de la evaluación se pudo ver que los estudiantes a la hora de describir y comparar conjuntos de datos ven utilidad en encontrar un representante. Pocos siguieron con la idea de fijarse en datos particulares como los extremos de cada distribución. Se encontró dificultad al preguntárseles sobre el concepto de dispersión o variabilidad: la mitad del curso cambió su concepción respecto a este tema, viendo ventajoso medir qué tan cercanos o lejanos se encuentran los datos de un conjunto con respecto a su representante; la otra mitad siguió con la noción de asociar este concepto solamente con la diversidad de los datos o con el rango. Es importante tener en cuenta que para lograr una verdadera comprensión de las medidas de tendencia central y de dispersión es necesario dedicar tiempo y no cometer el error de ignorar la dispersión cuando se comparan grupos de datos, encontrar un valor representativo y medir la intensidad con los que los datos se desvían respecto a él siempre deben ir de la mano. El concepto de variabilidad es bastante complejo y es necesario prestarle mucha atención. Su significado puede ser claro para el profesor, pero no para el estudiante (Batanero & Godino, 2001, p.3-17) y da lugar a distintas interpretaciones como se pudo ver durante el desarrollo de las actividades y la evaluación.

13. CONCLUSIONES Y RECOMENDACIONES

Los estudiantes conocieron y exploraron diferentes problemas en los que se encuentran inmersas las medidas de tendencia central y las medidas de dispersión. Más allá de aprender definiciones, reconocieron situaciones en donde surgen y se emplean estos conceptos.

Durante el desarrollo de las actividades se pudo ver que para describir y comparar conjuntos de datos, los estudiantes en un primer momento se fijan en datos particulares, especialmente valores extremos; para ver mejor la distribución ven la necesidad de construir una representación gráfica y así percibir las características más relevantes de los datos; y, por último, usan las medidas de tendencia central y de dispersión para describir los datos como conjunto.

Con el uso de Fathom y sus representaciones dinámicas, los estudiantes pudieron ver diferentes características de las principales medidas de tendencia central: la moda, la mediana y la media aritmética. Con relación a la primera, que fue la que primero usaron intuitivamente como representante de un conjunto de datos, se dieron cuenta que en muchos casos no tiene sentido pensar en esta como valor representativo, especialmente cuando no se tienen datos repetidos; respecto a la mediana, vieron su utilidad como representante debido a no estar influenciada por valores extremos, sino por el orden de los datos. Respecto a la media, al manipular las representaciones de Fathom como diagramas de puntos y modificar los datos, reconocieron las siguientes propiedades:

- La media es sensible a todos los datos. Los alumnos se dieron cuenta que, al modificar cualquiera de los valores de un conjunto en la tabla de datos o moviendo uno de los puntos en su representación gráfica, la media también cambiará su valor y se moverá en la dirección que se mueva ese punto.

- La media es un valor que está comprendido entre los extremos de la distribución. En el ejercicio de cambiar los datos en las representaciones de Fathom, los alumnos se dieron cuenta que es imposible que el valor de la media sea menor que el mínimo de los datos o mayor que el máximo.
- La media no tiene por qué ser igual a alguno de los valores de los datos. Incluso, puede tomar un valor decimal cuando todos los datos son enteros.
- Si a todos los valores de un conjunto de datos se les añade o se les sustrae un mismo valor, la media de los datos será la misma sumando o restando respectivamente ese valor.

Utilizando Fathom, los estudiantes vieron algunas medidas de dispersión: el rango, la desviación estándar y la desviación media. La primera permite saber la distancia entre el menor y el mayor valor de los datos; con relación a la segunda y a la tercera, con el ejercicio de modificar valores en las tablas de datos y mover puntos en las representaciones gráficas, reconocieron las siguientes propiedades:

- Esta mide qué tan parecidos son los datos a la media,
- su valor aumenta cuando los puntos se alejan de la media,
- su valor disminuye cuando los puntos se acercan a la media,
- su valor es cero solamente cuando todos los datos son iguales, y por lo tanto
- entre menor sea su valor, la media será mejor representante de los datos.

Usando el paquete estadístico Fathom los estudiantes exploraron las medidas de tendencia central y de dispersión. Por medio de las diferentes representaciones que ofrece este software, reconocieron las propiedades más importantes de estas medidas de forma numérica, física y dinámica al ver cómo cambian sus valores al modificar los datos. Para este nivel no era necesario llegar a demostraciones formales de estas propiedades, se vieron visualmente en el programa.

Con el desarrollo de las actividades y la evaluación final, con ayuda de Fathom los estudiantes comprendieron y aplicaron propiedades de las medidas de tendencia central y de dispersión para describir y comparar conjuntos de datos en diferentes problemas.

Las medidas de tendencia central y de dispersión, a pesar de que parecen simples y claras para el profesor, resultan realmente complejas y difíciles de comprender para los estudiantes. Para lograr una buena comprensión de estas medidas se requiere un proceso largo, donde la intuición y las ideas que tengan los alumnos, aunque no sean muy precisas, juegan un papel fundamental en la solución de problemas que sugieren encontrar un representante en un conjunto de datos y medir la variabilidad de estos datos.

Las medidas de tendencia central fueron de utilidad para los estudiantes para describir brevemente la distribución de conjuntos de datos en diferentes contextos; sin embargo, el concepto de dispersión resulta más complicado de comprender para ellos. Es necesario prestar atención especial hacia cómo conciben los estudiantes el concepto de variabilidad y cómo cambiar esta concepción para ver la utilidad de usar las medidas de dispersión con el fin de determinar qué tan lejos o cerca se encuentran los datos respecto a un representante y por ende, evaluar la misma calidad de este representante.

BIBLIOGRAFÍA

AFONSO, P. (2000). *Probabilidad y Estadística, conceptos, modelos y aplicaciones en Excel.* Bogotá: Pearson.

ALVARADO, E., & SÁNCHEZ, M. (2010). *Programa Fathom 2.* [en línea] http://www.slideshare.net/Elykang13/programa-fathom-2 [citado en 13 de octubre de 2013]

Anónimo. (2013). *Amigos Blaugranas.* [en línea] http://paco11.blogspot.com/2013/05/barca-contratos-jugadores-salario.html [citado en 18 de septiembre de 2013]

BATANERO, C. (2000). *Significado y Comprensión de las Medidas de Posición Central.* Departamento de Didáctica de la Matemática, Universidad de Granada. Disponible en: URL: http://www.ugr.es/~batanero/ARTICULOS/isboa.pdf [citado en 24 de septiembre de 2013]

BATANERO, C. (2001). *Didáctica de la Estadística.* Grupo de Investigación en Educación Estadística, Departamento de Didáctica de la Matemática, Universidad de Granada. Disponible en: URL: http://www.uruguayeduca.edu.uy/Userfiles/P0001%5CFile%5C118didactica estadistica.pdf [citado en 20 de septiembre de 2013]

BATANERO, C., & GODINO, J. D. (2001). *Análisis de Datos y su Didáctica.* Granada, España: Departamento de Didáctica de la Matemática Universidad de Granada. Disponible en: URL: http://www.ugr.es/~batanero/ARTICULOS/Apuntes.pdf [citado en 26 de septiembre de 2013]

HOLMES, P. (1980) Teaching Statiatics, Slough, Foulshans Educational.

KEY CURRICULUM PRESS. (s.f.). *Fathom Dynamic Data.* [en línea] https://www.keycurriculum.com/products/fathom [citado en 14 de octubre de 2013]

MINISTERIO DE EDUCACIÓN NACIONAL REPÚBLICA DE COLOMBIA. (1998). *Serie Lineamientos Curriculares Matemáticas.* Bogotá.

MINISTERIO DE EDUCACIÓN NACIONAL REPÚBLICA DE COLOMBIA. (2006). *Estándares Básicos de Competencias en Matemáticas.*

MINISTERIO DE ENERGÍA GOBIERNO DE CHILE. (2012). *Etiqueta de Consumo Energético Para Vehículos Livianos: portal de indicadores de consumo energético y emisiones vehiculares.* Santiago de Chile. [en línea] http://www.consumovehicular.cl/?q=comparador [citado en 24 de octubre de 2013]

REVISTA MOTOR. (2013). *Nuevos.* [en línea] www.motor.com.co/revista-motor/precios/590/ARCHIVO/ARCHIVO-13127676-0.pdf [citado en 24 de octubre de 2013]

RUSELL, S. J., & MOKROS, J. R. (1991). *What's Typical? Children's Ideas about Average.* Voorburg, Holanda.

WACKERLY, D., MENDENHALL, W., & SCHEAFFER, R. (2008). *Estadística matemática con aplicaciones Séptima Edición.* Ixtapaluca: Cosegraf.

WILD, C. J., & SEBER, G. A. (1999). *Chance Encounters: A Firs Course in Data Analysis and Inference.* John Wiley & Sons, Inc.