Formas de Razonamiento Covariacional Informal alrededor de la recta de mejor ajuste en un ambiente computacional en estudiantes de octavo grado

Ana Patricia García Amado

Trabajo de Grado para optar por el título de Magister en Educación Matemática

Director

Gabriel Yáñez Canal

Doctor en Ciencias Especialidad Matemática Educativa

Universidad Industrial de Santander

Facultad de Ciencias

Escuela de Matemáticas

Maestría en Educación Matemática

Bucaramanga

2018

Agradecimientos

En primer lugar, quiero agradecer a la Energía Creadora Divina por permitirme vivir esta experiencia que me ha dejado más que aprendizaje, una semilla qué cultivar para seguir haciendo investigación. Esta experiencia no solo me ha permitido crecer como persona y profesional sino que también me ha dado la oportunidad de conocer a muchas personas que, indudablemente, me han aportado mucho y estoy muy agradecida con ellos, en especial con mis compañeros Laura, Silvia, Édinson y Mónica y con mis profesores que hicieron parte de esta formación: Sandra Parada, Solange Roa y Jorge Fiallo.

También quiero agradecer a mis padres por todo el apoyo y confianza que me han podido dar y a Bayo por su disposición, buena vibra y apoyo incondicional.

Mil gracias a la Escuela de matemáticas y a mi alma máter, UIS, por permitirme, con su apoyo, dedicar cien por ciento de mi tiempo a esta investigación. También a la Institución Educativa San Francisco y a Johana Quintero y María del Carmen López.

Finalmente, un especial agradecimiento a la infinita impaciencia del profesor Gabo, a su apoyo constante y a sus sugerencias y correcciones siempre pertinentes. ¡Eureka!, Gabo.

Tabla de contenido

Lista de Figuras

Lista de Tablas

Lista de Apéndices

RESUMEN

TÍTULO: FORMAS DE RAZONAMIENTO COVARIACIONAL INFORMAL ALREDEDOR DE LA RECTA DE MEJOR AJUSTE EN UN AMBIENTE COMPUTACIONAL EN ESTUDIANTES DE OCTAVO GRADO.

AUTOR: ANA PATRICIA GARCÍA AMADO**

PALABRAS CLAVES: RAZONAMIENTO COVARIACIONAL INFORMAL, RECTA DE MEJOR AJUSTE, TRAYECTORIAS HIPOTÉTICAS DE APRENDIZAJE.

DESCRIPCIÓN:

En investigaciones acerca del razonamiento de los estudiantes surge la necesidad de explorar y aprovechar sus conocimientos previos, ya sean formales o informales, así como sus creencias, experiencias, intuiciones e inclusive errores para diseñar el camino que, teniéndolos en cuenta, los conduzca hacia la comprensión de conceptos matemáticos. Por otra parte, recientemente existe una tendencia en la investigación estadística que se caracteriza por el uso de programas computacionales como una respuesta a la necesidad de generar diversas muestras, u obtener diferentes tipos de representación, que permitan a los estudiantes una mejor comprensión de los fenómenos estocásticos.

En este documento presentamos los resultados de una investigación que tuvo como objetivo analizar las formas de razonamiento covariacional informal de estudiantes de octavo grado, alrededor de la recta de mejor ajuste a partir del diseño, implementación y análisis de una serie de actividades. Como parte de los resultados se sugiere una Trayectoria Hipotética de Aprendizaje (THA) (Simon, 1995) basada en actividades computacionales y que va desde un análisis univariado con la media hasta un análisis bivariado con la ubicación de la recta de mejor ajuste.

En esta investigación se ratifican algunos de los resultados descritos en los antecedentes, pero también se evidencia cierto control sobre concepciones y estrategias erróneas ya reportadas en cuanto a la ubicación de la recta de mejor ajuste, lo anterior debido a ideas desarrolladas en la parte univariada que permitieron una extensión al caso bivariado.

Finalmente, la THA propuesta, alrededor de la recta de mejor ajuste, con un enfoque informal, es solo una de las muchas que se podrían dar en una clase normal. Por el momento, esta THA provee a los profesores en ejercicio de actividades para que lleven a sus estudiantes a un primer acercamiento previo al futuro abordaje formal de la recta de mejor ajuste.

---

[*] Trabajo de grado

[**] Facultad de Ciencias Básicas y Humanas. Escuela de Matemáticas. Director: Gabriel Yáñez Canal, Doctor en Ciencias de Especialidad en Matemáticas.

ABSTRACT

TITLE: FORMS OF INFORMAL COVARIATION REASONING AROUND THE LINE OF BEST FIT IN A COMPUTATIONAL ENVIRONMENT FOR STUDENTS FROM EIGHT GRADE. *

AUTHOR: ANA PATRICIA GARCÍA AMADO**

KEYWORDS: INFORMAL COVARIATIONAL REASONING, LINE OF BEST FIT, HYPOTHETICAL LEARNING TRAJECTORIES.

DESCRIPTION:

In research about the students' reasoning, there is a need to explore and take advantage of their previous knowledge, whether formal or informal, as well as their beliefs, experiences, intuitions and even their errors to design the path that, keeping them into consideration, leads them towards understanding of mathematical concepts. On the other hand, recently there is a trend in the statistical research that is characterized by the use of computer programs as a response to the need to generate various samples, or obtain different types of representation, which allow students a better understanding of the stochastic phenomena.

In this document we present the results of a research that aimed to analyze the forms of informal covariational reasoning from students of eight grade, around the line of best fit from the design, implementation and analysis of a series of activities. As part of the results a Hypothetical Learning Trajectory (HLT) is suggested (Simon, 1995) based on computational activities and that goes from a univariate analysis with the mean, to a bivariate analysis with the location of the line of best fit.

This research confirms some of the results described in the background, but also shows some control over conceptions and erroneous strategies already reported regarding the location of the line of best fit, the previous due to ideas developed in the univariate part that they allowed an extension to the bivariate case.

Finally, the proposed HLT, around the line of best fit, with an informal approach, is just one of many that it could be given in a normal class. At the moment, this HLT provides teachers in practice with activities to take their students to a first approach prior to the future formal approach of the line of best fit.

---

* Degree work

** Faculty of Basic and Human Sciences. School of Mathematics Director: Gabriel Yáñez Canal, PhD in Mathematics Specialty Sciences.

Introducción

Es inevitable la inmersión de la estadística en nuestro diario vivir, tanto en la vida profesional como personal, debido a las situaciones de incertidumbre que enfrentamos día a día. Encontrar una manera de controlar estas situaciones ha sido un tema de gran interés desde tiempos remotos. Para mencionar un ejemplo de ello, extraído del MEN (2004), podríamos hablar de la civilización egipcia. En ese momento, la situación de incertidumbre era saber cuándo el río Nilo crecería e inundaría sus cosechas. Para hacerle frente a esta situación, los egipcios observaron los cambios de posición de una estrella, llamada Sirio, en relación al Sol, cuándo salía y cuándo se ocultaba; y estimaron así, de manera empírica la duración de un año. De igual manera, elaboraron un calendario que les permitió controlar este tipo de situaciones.

Estas situaciones de incertidumbre se deben a que nada permanece igual, todo cambia, vivimos rodeados de fenómenos cambiantes y en su predicción y control juegan un papel importante la unión de las matemáticas con la estadística.

Lo anterior nos lleva a cuestionarnos sobre las destrezas y conocimientos estadísticos que debe tener un individuo para enfrentar situaciones de incertidumbre, así como para desarrollar un pensamiento interpretativo y crítico sobre la información que le proporcionan los medios. De igual manera surgen cuestiones sobre cómo lograr que la comprensión de conceptos estadísticos, tan sofisticadamente presentados hoy en día, sea factible y útil para los estudiantes desde sus primeros grados escolares. ¿Acaso todo debe reducirse a un banal recetario de fórmulas? Organismos nacionales como el Ministerio de Educación Nacional (MEN), e internacionales como el *National Council of Teachers of Mathematics* (NCTM, por sus siglas en inglés) y el Proyecto *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) dan orientaciones respecto a las cuestiones anteriores.

En cuanto a la formalización de objetos estadísticos señalan que:

> ...hoy día ya no es tan importante para los estudiantes el recuerdo de las fórmulas y la habilidad para calcular sus valores, como sí lo es el desarrollo del pensamiento aleatorio… con el fin de intentar predecir dentro de ciertos rangos el curso de los acontecimientos respectivos y de tomar decisiones lo más razonables posibles ante la imposibilidad de saber con certeza lo que va a pasar (MEN, 2004, pp. 65 - 66).

Por otra parte, entre los procesos generales presentes en la actividad matemática que contemplan los Lineamientos Curriculares de Matemáticas se encuentra el de *razonar* qué subyace a la memorización de reglas y fórmulas al darles sentido y lógica a su aplicación. Dentro de las actividades que permiten el desarrollo de un razonamiento crítico se encuentran: predecir, conjeturar, justificar o refutar conjeturas, explicar coherentemente, responder con argumentos y razones (MEN, 2004).

Desde una perspectiva internacional, los Principios y Estándares para la Educación Matemática (NTCM, 2000), en cuanto al *Análisis de Datos y Probabilidad,* hacen énfasis en que "los estudiantes necesitan saber hacer análisis de datos y otros aspectos relativos a la probabilidad para poder razonar estadísticamente" (NCTM, 2000, p.51). Para ello, necesitan capacitarse en:

- Formular preguntas que los conduzcan a recoger datos, organizarlos, representarlos. Ya a partir del nivel medio, los estudiantes deberían trabajar con datos generados por simulaciones o extraídos de otras fuentes.

- Analizar datos con métodos estadísticos apropiados. En cuanto a los datos bivariantes, a partir de los niveles medios los estudiantes deberían investigar asociaciones y tendencias.

- Hacer inferencias y predicciones basándose en los datos. Los estudiantes deberían desprenderse de sus juicios y creencias y observar lo que informan los datos, al igual que debería surgirles la necesidad de modelar situaciones reales y hacer simulaciones para obtener un conjunto mayor de datos que resultaría costoso hacerlo en la vida real.

- Manejar conceptos básicos de probabilidad y comprenderlos. En cuanto a la simulación, a partir de los niveles medios, los estudiantes deberían hacer simulaciones que les permitan cuantificar la probabilidad de ocurrencia de un evento, pasar de las simulaciones físicas a las computacionales.

Como complemento a lo propuesto por el NCTM (2000) respecto al estándar Análisis de datos y Probabilidad, nace el proyecto GAISE elaborado por Franklin et al (2007). Este proyecto proporciona un marco para la educación estadística el cual utiliza tres niveles de desarrollo de la alfabetización estadística: A, B y C. Señalan, para cada nivel, que:

- Nivel A: Los estudiantes deberían ser capaces de observar una posible asociación entre dos variables, una numérica y una categórica, por ejemplo, al observar un diagrama de puntos. O la de dos variables numéricas, en un diagrama de dispersión para buscar tendencias y patrones. O también observar el cambio de una variable numérica con el tiempo.

- Nivel B: Para este nivel los estudiantes deberían usar representaciones más sofisticadas que impliquen la relación entre dos variables numéricas, así como la intensidad de la asociación entre dos variables, posteriormente la cuantificación de las tendencias y patrones observados en el nivel y hasta modelos sencillos que den cuenta de la relación entre dos variables, por ejemplo una línea recta.

- Nivel C: Los estudiantes deberían reconocer cuándo la relación entre dos variables numéricas es lineal, comprender que el coeficiente de correlación de Pearson mide la fuerza de la relación lineal entre dos variables numéricas. También deberían contemplar que la asociación a veces se debe simplemente al azar. Deberían interpretar resúmenes estadísticos, como el coeficiente de correlación asociados a los datos y en el contexto de los mismos (Franklin et al, 2007).

Lo anterior justifica en parte las investigaciones que se han realizado alrededor de la enseñanza y el aprendizaje de la asociación estadística.

Recientemente, algunas investigaciones se han encaminado hacia la exploración del razonamiento informal de los estudiantes antes de la presentación formal de los conceptos estadísticos. Con base en los resultados de estas investigaciones planteamos nuestro problema de investigación que se describe a continuación:

Problema de investigación y justificación

La necesidad por buscar explicaciones de lo que pasa a nuestro alrededor nos lleva a relacionar unos eventos con otros. Sin embargo algunas explicaciones podrían ser equívocas debido a interpretaciones incorrectas o ideas falsas alrededor de la asociación estadística en general, inclusive entre aquellos que están muy familiarizados con el tema (Borovcnik, 2012).

Borovcnik (2012) señala que, para aquellos que buscan interrelacionar variables, es importante comprender, dar uso, evaluar e interpretar resultados de asociación estadística.

Señala, además, que no es sencillo hacerlo ya que detrás de conceptos como los de covariación y regresión se esconden algunos errores y trampas que impiden una interpretación acertada. Respecto a la justificación matemática, dice que su comprensión debe ir más allá del hecho de aplicar o explicar un algoritmo matemático y que ésta comprensión debe reflejarse en el contexto al que pertenecen los datos.

Estepa, Gea, Cañadas y Contreras (2012, p. 1) también hacen énfasis en que "entre las nociones estadísticas fundamentales, cuya enseñanza debe optimizarse, se encuentran las de correlación y regresión" ya que "conocer si los sucesos se relacionan y, con qué intensidad lo hacen, facilita a las personas explicar el pasado, controlar el presente y predecir el futuro" (Estepa et al., 2012, p. 1. Tomado de Crocker, 1981, p.272).

La investigación que aquí se desarrolla tiene como tema de interés el Razonamiento Covariacional Informal, en particular sobre la recta de mejor ajuste, ya que "explicar, controlar y predecir los sucesos que se presentan en nuestro día a día, depende de habilidades para detectar covariaciones" (Gea, 2012, p.17. Tomado de Alloy y Tabachnik, 1984) que incluye encontrar tendencias o patrones, así como el modelo de ese patrón, por ejemplo la recta de mejor ajuste.

Por otra parte, recientemente existe una tendencia en la investigación estadística que se caracteriza por el uso de programas computacionales como una respuesta a la necesidad de generar diversas muestras que permitan a los estudiantes una mejor comprensión de los fenómenos estocásticos.

Dada la dificultad de las estrategias formales en la enseñanza de la estadística, en los años recientes se ha venido investigando alrededor del razonamiento informal que utilizan los estudiantes, con el ánimo de impulsar su desarrollo asumiendo que es el paso previo para la comprensión posterior de los modelos formales.

De manera que esta investigación buscó responder las siguientes preguntas:

• *¿Cuáles son las formas de razonamiento covariacional informal de los estudiantes alrededor de la recta de mejor ajuste como producto de una secuencia de actividades en un ambiente computacional?*

• *¿Cómo evoluciona el razonamiento informal de los estudiantes, alrededor de la recta de mejor ajuste, a lo largo de la implementación y discusión de las actividades?*

• *¿Cuál podría constituir una THA para la enseñanza informal de la Recta de mejor ajuste?*

Adoptamos por razonamiento covariacional el asociado a la relación entre dos variables continuas y que, además del coeficiente de correlación de Pearson, se expresa a través modelos de regresión. En particular, la investigación hará énfasis en el análisis de los diagramas de dispersión y la recta de mejor ajuste.

En síntesis, en este trabajo nos propusimos investigar acerca del desarrollo del razonamiento covariacional informal en estudiantes de octavo grado cuando se enfrentan a tareas soportadas en programas computacionales. Para el desarrollo de esas tareas, los estudiantes disponían solo del conocimiento informal y del razonamiento informal que de alguna manera habían adquirido en experiencias previas dentro y fuera del aula de clase. Lo anterior porque hemos conjeturado que es posible, favorable y recomendable acercarse de manera informal a conceptos estadísticos como la regresión, antes de una presentación formal que poco sentido tiene para el estudiante. Además, como señala Casey (2015), "el aprendizaje informal de la regresión, sienta las bases para futuros estudiantes sobre el estudio de regresión formal (por ejemplo, la recta de regresión de mínimos cuadrados)" (p.2). Por otro lado, un aprovechamiento de programas computacionales, como Fathom o Geogebra, constituye una herramienta potente para ese primer acercamiento al que nos referimos.

Consideramos que los resultados de esta investigación redundan no solamente en mayor conocimiento de una enseñanza basada en actividades computacionales, sino que provee a los profesores en ejercicio de actividades debidamente elaboradas para su ejercicio profesional.

La presentación de este documento se ha organizado en cinco apartados que se describen brevemente a continuación:

1. *Antecedentes.* Debido a que existe una vasta gama de investigaciones enfocadas hacia el razonamiento covariacional, en esta sección se muestra una síntesis de la revisión bibliográfica realizada. Esta síntesis a su vez está sujeta a las diferentes disciplinas desde donde se ha estudiado el razonamiento covariacional: sicología, matemáticas, educación matemática y educación estadística y formación de profesores.

2. *Referentes conceptuales.* Como se mencionó anteriormente, hay una tendencia en la investigación en educación estadística hacia el estudio del razonamiento informal y hacia el uso de la simulación computacional como apoyo para la enseñanza y el aprendizaje de temas estadísticos, razón por la cual, en esta sección se caracterizan estos conceptos, específicamente, lo que se entiende en esta investigación por razonamiento covariacional informal. También se expone lo que se entiende en esta investigación por Recta informal de mejor ajuste.

3. *Metodología.* Esta sección se divide a su vez en cuatro apartados: i*) Marco metodológico*, basado en elementos del ciclo de enseñanza de las matemáticas que propone Simon (1995**)**, principalmente en una parte de este ciclo que corresponde a las Trayectorias Hipotéticas de Aprendizaje (THA); ii) Descripción de la *Población y muestra* en donde se llevó a cabo la investigación; iii) Descripción de las *Herramientas computacionales* y; iv) Descripción de las *Fases de la investigación.*

4. *Análisis de resultados.* En esa sección se describen las actividades y pruebas, y el análisis detallado de las mismas. Estos análisis incluyen transcripciones del material (hojas de trabajo, videograbaciones, entrevistas, vídeos de pantalla) recolectado durante la implementación, desarrollo y discusión de las actividades, y que además consideramos claves ya que evidencian algunas formas y/o cambios de razonamientos en los estudiantes.

5. *Conclusiones.* Para finalizar, exponemos en esta última sección las conclusiones generadas de los análisis de resultados. Esta sección se divide a su vez en dos apartados: i) Una THA del concepto de Recta de mejor ajuste y ii) Formas de razonamiento informal alrededor de la recta de mejor ajuste en estudiantes de octavo grado al finalizar la implementación de la THA. También incluimos en esta sección algunas sugerencias para posibles futuras investigaciones dirigidas en este mismo sentido, la enseñanza informal de la recta de mejor ajuste.

1. Antecedentes

Aunque nuestro tema de interés es la recta de mejor ajuste, consideramos pertinente resaltar algunas investigaciones que se han hecho alrededor de lo que podríamos llamar como un todo, *asociación estadística* o *covariación*, que incluye diagramas de dispersión, correlación y recta de mejor ajuste.

Como señalan Zieffler (2009) y Garfield y Ben-Zvi (2008) la covariación ha sido un tema de gran interés desde diferentes disciplinas como la psicología, las matemáticas, la estadística, la educación matemática, la educación estadística, y en general, para aquellos que hacen investigaciones empíricas para buscar establecer, justificar y describir relaciones entre dos o más variables (Borovcnik, 2012). Como resultado de este tipo de investigaciones se han identificado sesgos, concepciones, dificultades, errores e incluso trampas alrededor de la covariación (Borovcnik, 2012). Así mismo, se han identificado formas de razonamiento covariacional que, posteriormente, han generado instrumentos y trayectorias de aprendizaje.

Por lo anterior, esta sección estará dividida en tres partes: i) Concepciones de los estudiantes sobre la asociación estadística, en particular una subsección para la recta de mejor ajuste; ii) Razonamiento covariacional de los estudiantes y; iii) Diseños de instrucción para la enseñanza y el aprendizaje de la covariación.

1.1 Concepciones de los estudiantes sobre la asociación estadística

En cuanto a asociación estadística en general, Estepa y Batanero (1995) clasificaron las concepciones previas que tienen los estudiantes en cuatro grupos: *causalista, determinista, unidireccional y localista.*

- *Causalista:* Cuando se considera que existe dependencia entre las variables solo cuando se identifica una relación de tipo causal entre ellas.

- *Determinista:* Cuando se espera una relación de tipo funcional entre las variables, es decir, para un valor de la variable independiente se espera un solo valor de la variable dependiente.

- *Unidireccional*: Cuando se considera un solo sentido de asociación entre las variables, generalmente no se admite el de tipo dependencia inversa, inclusive llegándose a considerar como independencia en esos casos.

- *Localista:* Cuando los juicios de asociación se justifican en datos aislados.

Concepciones como las anteriores también se evidencian en estudiantes que ya han estudiado el tema, en particular la correlación. Sánchez, Estepa y Batanero (2000) realizan un estudio experimental acerca de la estimación de la correlación a partir de diferentes representaciones. Dentro de sus resultados está que no todos los estudiantes actúan de manera similar frente a problemas relacionados con el concepto de correlación, pero que muestran una buena capacidad para estimar la correlación y que ésta es más precisa cuando se tiene un diagrama de dispersión. Las mayores dificultades se presentan en la tarea de construir una nube de puntos a partir de una descripción verbal, así como estimar el coeficiente a partir de una tabla numérica. Por otra parte, hay mayor precisión en la estimación de la correlación cuando es más intensa, aunque se evidencian problemas de comprensión en el concepto de correlación así como en el significado del coeficiente de correlación. Respecto a esto último, el significado del coeficiente de correlación, Truran (1995) muestra algunas interpretaciones que los estudiantes hacen:

- Algunos manifestaron que el valor positivo en r significaba que un aumento de X implicaba una caída en Y.

- Otros decían que r estaba entre 0 y 1.

- No existe un acuerdo entre los estudiantes sobre lo que se entiende por una asociación fuerte o significativa.

Por otro lado, la dependencia estadística de dos variables no es tan familiar como sí lo es la dependencia funcional o determinista; inclusive entre los profesores, ya que muchos de ellos tienen un pensamiento determinista sobre datos bivariados. De manera que, para una comprensión adecuada de la correlación y la regresión se necesita una comprensión no solo acerca de las funciones sino también de la variación (Engel y Sedlmeier, 2011). En ese sentido, aunque algunas personas creen que un coeficiente de correlación de ± 1 supone una relación funcional entre dos variables X e Y, por otro lado consideran que solo hay una dependencia determinista entre estas variables si el coeficiente de correlación es diferente de cero, cuando puede haber dependencia determinista en relaciones no lineales con un coeficiente de correlación igual a cero (Tomado de Estepa y Sánchez, 2001).

1.1.1 Concepciones de los estudiantes sobre la recta de mejor ajuste. Casey (2015) investiga las concepciones que tienen los estudiantes sobre la recta de mejor ajuste antes de una introducción formal del concepto. El estudio, bajo un enfoque informal, se realizó con 33 estudiantes de octavo grado quienes tenían que resolver tareas en contextos familiares que implicaban colocar la recta de mejor ajuste en un diagrama de dispersión, siempre explicando el razonamiento que hacían. Investigaciones de este tipo son determinantes ya que informan a los profesores de estadística y a los escritores del plan de estudios sobre cómo los estudiantes piensan de manera natural sobre temas específicos, como la recta de mejor ajuste, y de esta forma anticiparse a las respuestas de los estudiantes, lo cual constituye un elemento clave en la práctica docente (Casey, 2015). Para el análisis de datos usan el método de la teoría fundamentada (Glaser y Strauss, 1967) y los resultados se sintetizan en cuatro aspectos, todos respecto a la recta de mejor ajuste: i) Significados; ii) criterios y métodos para la posición; iii) precisión en la posición y; iv) la interpretación.

Significados. Se identifican cuatro categorías: i) representación donde se espera ver la relación entre las variables; ii) muestra la apariencia de los datos; iii) algo que se usa para acercarse a predicciones y; iv) el promedio de los puntos.

Criterios y métodos para la posición. Se identifican siete métodos, ordenados de mayor a menor frecuencia: i) a través de tantos puntos como sea posible; ii) igual número de puntos a ambos lados; iii) cerca de todos los puntos como sea posible; iv) reflejar la relación que tienen las variables con base al conocimiento del contexto; v) a mitad de camino entre los puntos mínimo y máximo; vi) a través de los puntos primero y último y; vii) comenzando desde el primer punto y luego trazar maximizando el número de puntos.

Precisión en la posición. Existe una considerable variabilidad entre los estudiantes para ubicar la recta. Algunos la ubican muy cerca y paralela a la recta de regresión. Otros utilizan el criterio de que pase por la mayoría de puntos, inclusive un estudiante llama la atención de que la recta debe pasar por el origen. Algunos, reflejando un razonamiento univariado, trazan rectas horizontales para dejar igual número de puntos a cada lado, es decir buscando una aplicación para la media o mediana. En general, la precisión va muy de la mano con los criterios descritos anteriormente.

Interpretación. Se evidenció que 25 de 33 estudiantes interpretaron ampliamente la recta de mejor ajuste al presentar argumentos como “Muestra que cuanto mayor es la altura de la cual se deja caer la pelota, más alto rebota” o “A medida que el precio subió se podía ver que la cantidad de personas bajó”. Otras interpretaciones fueron como “no muestra nada” sobre la relación, y otras como una evidencia de una relación causal.

1.2 Razonamiento Covariacional de los estudiantes

Un razonamiento covariacional implica la interpretación y juicio de una relación entre dos variables y puede darse de tres maneras: i) Matemática: a través de una función lineal; ii) Estadística: por medio de un diagrama de dispersión o iii) Cualitativa: predicciones causales sobre eventos con base en asociaciones observadas Zieffler & Garfield (2009).

En un sentido matemático o funcional se destaca el trabajo de Carlson et al., (2002) quienes entienden por *comprensión de la covariación* como “mantener en la mente, de manera simultánea, una imagen sostenida de dos valores de cantidades (magnitudes)” y agregan que “esta actividad mental involucra la coordinación de las dos cantidades, es decir, hacer seguimiento al valor de cada cantidad y darse cuenta de que la otra cantidad también tiene un valor en cada instante”. Así mismo, definen razonamiento covariacional “como las actividades cognitivas implicadas en la coordinación de dos cantidades que varían mientras se atiende a las formas en que cada una de ellas cambia con respecto a la otra” (p.354) y muestran que el desarrollo de este tipo de razonamiento es fundamental para la comprensión de algunos conceptos en el cálculo, por ejemplo la derivada.

En un sentido estadístico o aleatorio, Moritz (2004) señala que

> El razonamiento sobre la covariación comprende los procesos de traducción entre datos numéricos crudos, representaciones gráficas, y las declaraciones verbales sobre covariación estadística y asociación causal. Otros procesos pueden incluir el cálculo y la interpretación

> de las pruebas estadísticas de asociación, el modelado matemático para ajustar los datos a una ecuación funcional específica, y la traducción desde y hacia las expresiones simbólicas de funciones algebraicas (p.228).

Por lo que considera como habilidades de razonamiento covariacional las asociadas a: i) Generar datos de forma especulativa, es decir, traducir de una representación verbal a una gráfica; ii) Interpretar verbalmente un gráfico, es decir, traducir de una representación gráfica a una verbal e; iii) Interpretar numéricamente un gráfico, esto es, leer valores e interpolar. En ese sentido, realiza un estudio del razonamiento covariacional en estudiantes de dos escuelas privadas en los grados tercero, quinto, séptimo y noveno. Este estudio tuvo como objetivo explorar esas tres habilidades.

Entre los resultados se muestran las asociaciones entre las tres habilidades bajo estudio concluyendo que la interpretación numérica de un gráfico está altamente correlacionada con la interpretación verbal y con la generación de datos de forma especulativa, mientras que la correlación entre la interpretación verbal de un gráfico con la generación de datos de forma especulativa es baja. Esta investigación muestra que los estudiantes pueden llegar a diferentes representaciones tanto gráficas como verbales de la covarianza antes de su normalización convencional; lo anterior a partir de contextos familiarizados.

Por otro lado, cuestiones pertinentes que abordan Zieffler y Garfield (2009) tienen que ver con la naturaleza, o el patrón de cambio, en el desarrollo del razonamiento covariacional de los estudiantes a lo largo de un curso; también sobre si la secuencia de conocimientos alrededor de datos bivariados se asocia al patrón de cambio en el razonamiento o si los cambios en el razonamiento en conceptos fundamentales, como en la distribución de una variable, se asocia con las diferencias en el patrón de cambio del razonamiento covariacional. Como respuesta a estas cuestiones, evidencian que los estudiantes tienen cierta capacidad de razonamiento covariacional antes de cualquier enseñanza formal sobre datos bivariados. Así mismo, concluyen que el desarrollo del razonamiento distributivo univariante fue mayor al del covariacional, mostrando así la dificultad que tienen los estudiantes para razonar sobre datos bivariados, aunque existe una relación positiva entre el razonamiento distributivo y el covariacional, de manera que si los estudiantes muestran mayores cambios en su razonamiento distributivo univariado tenderán a tener mayores cambios en su razonamiento sobre datos bivariados; así que vale la pena invertir suficiente tiempo y atención para potenciar el razonamiento de los estudiantes sobre la distribución de una variable. En cuanto a la secuenciación de conocimientos, no se vio una influencia sobre estos tipos de razonamiento.

1.3 Diseños de instrucción para la enseñanza y el aprendizaje de la covariación

Lavalle, Micheli y Rubio (2006), hacen énfasis en que el conocimiento y la identificación de los errores y dificultades tanto en la enseñanza como en el aprendizaje, así como el análisis de contenido antes de abordar un tema específico, permite el diseño de actividades didácticas para superar o prevenir esas dificultades. En ese sentido describen los conocimientos previos que se deben tener antes de abordar la correlación y la regresión:

- *Conocimientos matemáticos:* Puntos en el plano, variable, función, función lineal, ecuación de la recta, pendiente, ordenada al origen, representación gráfica de la recta, ecuación de la recta que pasa por un punto con pendiente conocida.

- *Conocimientos estadísticos:* Datos bivariados, media, desviación estándar y variabilidad.

De igual manera, para favorecer la comprensión de la correlación y regresión lineal en un análisis de contenido consideran tratar los siguientes conceptos: variable, datos, gráfico de dispersión, recta de regresión, estimación de valores de la variable respuesta y coeficiente de correlación lineal. A lo anterior se suma Gea (2013) quien sugiere hacer una presentación más equilibrada de situaciones que involucren el sentido, la intensidad y los tipos de dependencia, así como el uso de datos reales, recursos interactivos (applets) y el planteamiento de proyectos completos donde se parta de una pregunta que le permita al estudiante recorrer uno a uno los pasos del proceso de pensamiento estadístico: transnumeración, consideración de la variación, razonamiento con modelos estadísticos e integración de la estadística y el contexto (Wild y Pfannkuch, 2009).

1.3.1 Trayectorias Hipotéticas de Aprendizaje para desarrollar el razonamiento covariacional. Como se mencionó anteriormente, la identificación de errores, concepciones y formas de razonamiento, así como el análisis de contenido alrededor de la covariación, ha tenido implicaciones en la enseñanza de este tema. Dentro de las implicaciones están las investigaciones que tienen como resultado final el diseño de una trayectoria de aprendizaje o de una secuencia de actividades que permiten desarrollar el razonamiento covariacional de los estudiantes. Ejemplos de este tipo de investigación son los trabajos de Cobb, McClain y Gravemeijer (2003), Garfield y Ben-Zvi (2008) y Casey (2014, 2015).

Las trayectorias hipotéticas de aprendizaje (THA) (Simon, 1995) constituyen un constructo donde se consideran, de manera simultánea, la meta de aprendizaje, las formas de pensamiento y razonamiento de los estudiantes, los modelos de pensamiento de los estudiantes, profesores e investigadores, una secuencia de instrucciones y la interacción de todo lo anterior para un análisis detallado de los procesos (Clements y Sarama, 2004). A continuación presentamos una síntesis de estas investigaciones, centrándonos en la trayectoria que cada uno de estos estudios propone.

Cobb, McClain y Gravemeijer (2003) proponen una THA alrededor de las formas de razonamiento de los estudiantes sobre datos bivariados, a partir de la lectura de diagramas de dispersión, con el fin de que los estudiantes desarrollen una visión global de los datos y así evitar una concepción de tipo *localista.* La THA que proponen tiene las siguientes características.

- El punto de partida, para una secuencia de instrucción que apunte hacia la distribución de datos bivariados, es la distribución de conjuntos de datos univariados y sus formas de inscripción: en grupos de un tamaño específico, en intervalos de igual anchura, en dos grupos iguales o en cuatro grupos iguales. La inscripción inicial presentada se muestra en la Figura 1.

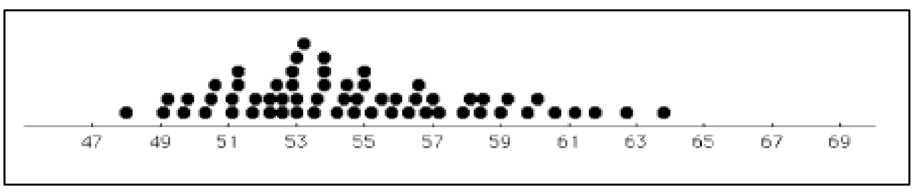
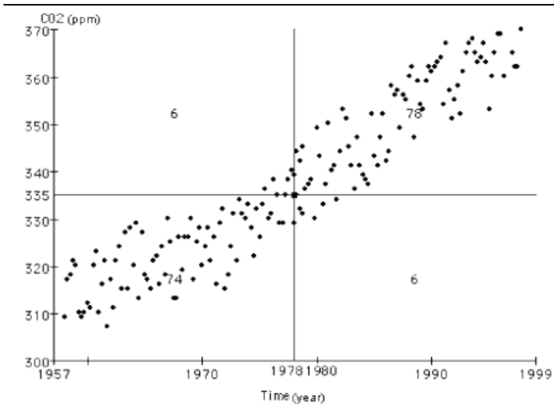


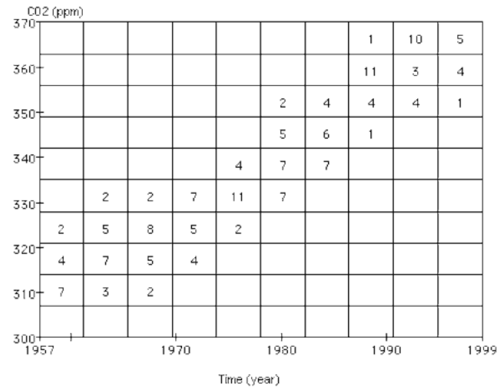
Figura 1. Distribución de datos univariados (Tomado de Cobb, McClain y Gravemeijer, 2003)

- Seguidamente, enfocarse en el desarrollo de las formas de inscripción de datos bivariados, las cuales constituyen una extensión de las formas de inscripción para datos univariados, como se observa en la Tabla 1.

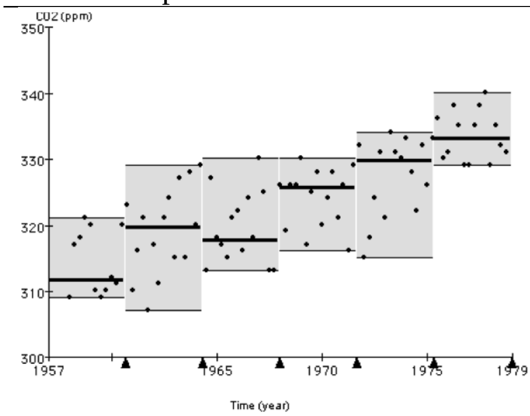
Tabla 1.
Formas de inscripción para datos bivariados. Cobb, McClain y Gravemeijer (2003)



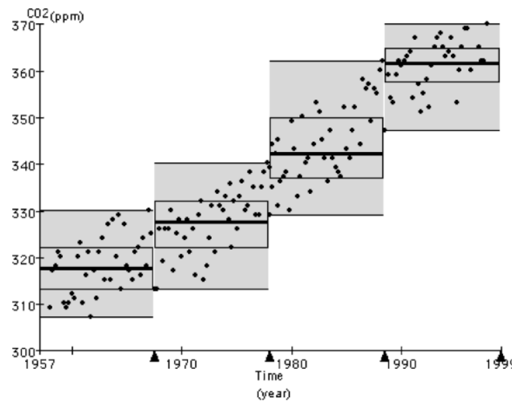
Cross. Una vez ubicado en algún lugar de la pantalla, esta opción divide los datos en cuatro celdas y muestra a la vez el número de puntos en cada cuadrante.



Grids. Permite particionar los datos en rejillas desde 4 x 4 hasta 10 x 10, mostrándose el número de datos que cae en cada celda.



Dos grupos con el mismo número de datos. Esta opción permite particionar los datos en columnas verticales, desde 4 hasta 10; a su vez, divide los datos de cada partición en dos grupos iguales, es decir que muestra la mediana para cada rebanada.



Cuatro grupos iguales. Como en Dos grupos con el mismo número de datos, con la diferencia de que los datos de cada partición se dividen en cuatro grupos iguales, es decir que muestra los cuartiles para cada rebanada.

- Luego, usar los datos apilados como distribuciones de datos bivariados (ver Figura 2).

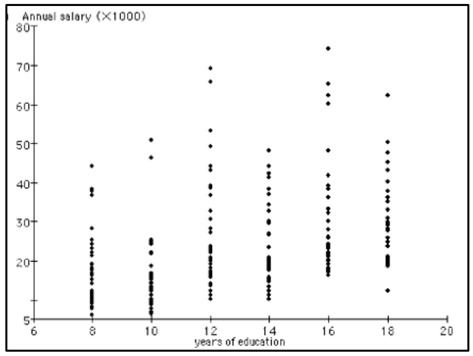


Figura 2. Ejemplo de datos apilados. Cobb, McClain y Gravemeijer (2003)

- Finalmente hacer la transición a los diagramas de dispersión como una distribución de datos bivariados.

Garfield y Ben-Zvi (2008) también proponen una secuencia de ideas y conceptos para desarrollar el razonamiento covariacional considerando las seis características, que proponen Watkins et al., (2004), que pueden guiar a los estudiantes para que aprendan a razonar sobre datos bivariados y que además pueden ser desarrolladas de manera informal: i) Las variables individuales y su variabilidad; ii) La forma de un patrón revelado en un gráfico de dos variables en términos de linealidad, agrupaciones, y valores atípicos; iii) La tendencia si hay una (positiva o negativa); iv) La fuerza de la tendencia: fuerte o débil, variable o constante; v) Si el patrón puede generalizar a otras situaciones (se puede probar con los métodos de inferencia, y también se basa en la forma en que se obtuvo la muestra) y; vi) Si hay explicaciones plausibles para el patrón.

Lo anterior apoyándose en la tecnología ya que las investigaciones revelan que los diferentes tipos y usos de la tecnología ayudan a que los estudiantes hagan juicios más precisos de covariación (Garfield y Ben-Zvi, 2008).

Por otra parte en Casey (2014) se sintetiza una THA para la enseñanza y el aprendizaje de la regresión lineal en secundaria y en el nivel medio como resultado de tres investigaciones de las cuales dos, se enfocaron en cómo los estudiantes de octavo grado progresan en su aprendizaje de la regresión lineal a partir de sus concepciones iniciales hasta la compresión que se espera que tengan en el grado doce. La tercera investigación, con estudiantes de noveno grado, tenía como objetivo describir el conocimiento didáctico del contenido que necesitan los profesores para la enseñanza informal de la recta de mejor ajuste. “Una THA proporciona un modelo para el sucesivo y gradual pensamiento por el que el estudiante debe pasar para la comprensión de un tema” (p.1).

De manera que existen dos componentes necesarios que hacen parte del conocimiento para la enseñanza estadística: i) las concepciones iniciales que tienen los estudiantes y ii) cómo estas concepciones pueden convertirse en una sólida comprensión del tema con un diseño y una implementación de aprendizaje adecuado (Casey, 2014). Estos dos componentes constituyen una base para el diseño de la THA, la cual, estructuralmente corresponde a un bucle que progresa a través de cinco bucles donde cada bucle tiene como base el aprendizaje que se ha producido en el bucle anterior. Para intereses propios, haremos alusión solo al primer bucle: recta de mejor ajuste de manera informal. En la Tabla 2, la THA que propone para este primer bucle.

Tabla 2.
Primeros dos bucles para la enseñanza y el aprendizaje de la regresión lineal (Tomado y adaptado de Bargagliotti et al. 2012)

	Formular la pregunta	**Recolectar datos**	**Analizar datos**	**Interpretar resultados**	**Clave del desarrollo de entendimientos**
Bucle 1	¿Cómo podemos mostrar una	a. Determinar un método	a. Representar gráficamente los datos en un	a. Interpretar la pendiente y la intersección con y	Los datos bivariados pueden ser visualizados en

	distribución cuantitativa de dos variables y ampliar los conceptos de forma, centro y la variabilidad para resumirlo?	para obtener las mediciones / conteos para cada variable de cada miembro del grupo.	diagrama de dispersión. b. Discutir las razones de por qué hay "dispersión" en el gráfico. c. Determinar visualmente y describir la tendencia del gráfico. d. Analizar maneras de modelar los datos. e. Examinar los diagramas de dispersión y hacer intentos para sacar un modelo del gráfico (por ejemplo, esto podría hacerse mediante la elaboración de una recta o una curva en el diagrama de dispersión). f. Para esos datos donde un modelo lineal es apropiado, discutir donde la "recta de mejor ajuste" se podría ubicar.	en el contexto de la pregunta planteada para la recta colocada en el diagrama de dispersión. b. De manera informal describir la correlación mediante el examen de qué tan cerca están los puntos a la recta. c. Predecir valores de y usando valores de la variable independiente dentro del rango de los valores del conjunto de datos de la muestra (interpolación). d. Analizar la limitación de predicciones usando los valores de la variable independiente fuera de los datos de la muestra (extrapolación). e. Analizar la viabilidad de examinar los diagramas de dispersión en el caso de un gran conjunto de datos.	un diagrama de dispersión. Una relación se dice que es "lineal" si los puntos se agrupan en una nube elíptica. Cuando la relación es lineal, el centro puede ser modelado por una recta, la cual puede ser utilizada para la predicción. El error en la predicción se puede estimar a partir del tamaño de los residuos, los cuales se miden en sentido vertical. La variabilidad de la recta se mide por la correlación.

1.4 Resumen de la revisión de la literatura

Las investigaciones descritas dan cuenta de las dificultades, errores y sesgos que tienen no solo los estudiantes, sino las personas que, en general, buscan relacionar dos o más variables, es decir, alrededor de la covarianza. En este sentido, para hacer frente a esas dificultades otras investigaciones, tanto en Educación Matemática como en Educación Estadística, han tenido como foco el estudio del razonamiento covariacional antes (razonamiento informal) y después de una instrucción formal. Identificadas las concepciones y formas de razonamiento covariacional de los estudiantes, otras investigaciones proponen trayectorias de aprendizaje, que incluyen el uso de herramientas tecnológicas y que tienen como fin la comprensión gradual, algunas partiendo desde lo informal, de estos conceptos: covariación, correlación y regresión, en particular la recta de mejor ajuste.

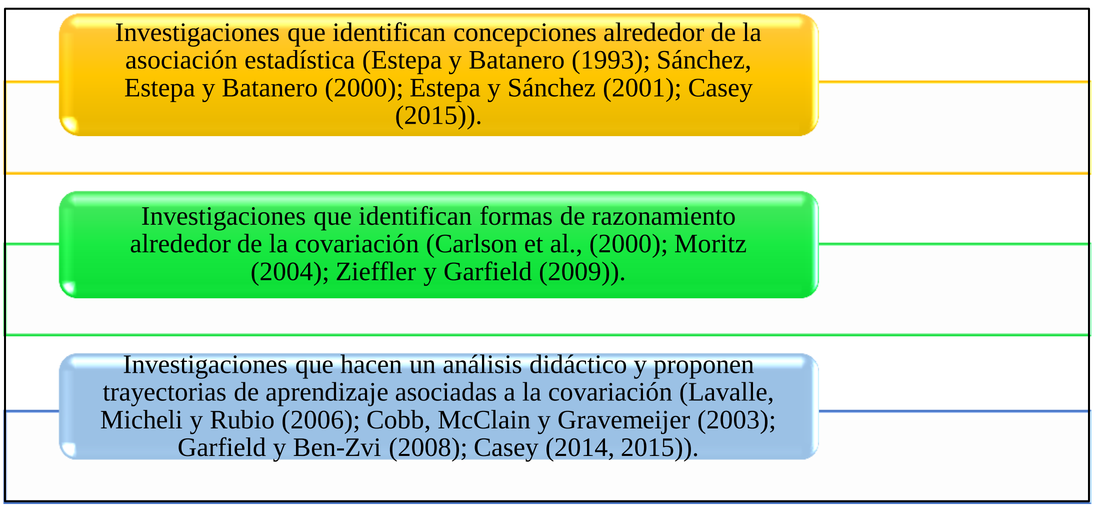


Figura 3. Investigaciones alrededor de la asociación estadística

A manera de resumen, presentamos algunas conclusiones obtenidas de la revisión de la literatura:

- La covariación, su comprensión e interpretación, ha sido el foco de estudio de investigaciones que van desde la sicología y la educación hasta investigaciones empíricas.

- Las concepciones previas (causalista, determinista, unidireccional y localista) que tienen los estudiantes alrededor de la asociación estadística influyen en la estimación y estrategias para establecer una relación entre dos variables.

- Los estudiantes presentan dificultades en la interpretación del significado de la correlación y del coeficiente de correlación aun cuando logran calcularlo matemáticamente.

- La estimación de la correlación es más precisa cuando ésta es más fuerte, positiva y además se tiene un diagrama de dispersión, aunque llegar a esa representación, a partir de una expresión verbal, presenta dificultades en los estudiantes

- Las creencias previas pueden llevar a emitir una correlación donde no la hay o a que las personas entren en conflicto con lo que muestran los datos.

- Existe una dificultad para identificar la variable explicativa y la variable respuesta.

- Dentro de las habilidades que identifican un razonamiento covariacional correcto está la traducción de una representación a otra, así como la interpolación y la extrapolación. Estas habilidades pueden ser desarrolladas usando datos reales que tengan sentido para los estudiantes, y usando tecnología.

- Los estudiantes tienen cierta capacidad de razonamiento covariacional antes de cualquier enseñanza formal sobre datos bivariados.

- El desarrollo adecuado de un razonamiento sobre datos univariados favorece la transición hacia el razonamiento sobre datos bivariados.

- Conocer las concepciones iniciales que tienen los estudiantes alrededor de un tema enriquece el diseño y la implementación de trayectorias hipotéticas de aprendizaje que conduzcan a los estudiantes hacia una sólida comprensión del tema.

- En el caso del estudio de la regresión lineal, los estudiantes deberían desarrollar ideas intuitivas acerca de datos bivariados usando datos apilados, pendiente, ordenada en el origen y ecuación de una recta.

Teniendo en cuenta el panorama descrito, así como las reflexiones finales resultado del estudio de la literatura, el objetivo principal de esta investigación fue: Analizar las formas de razonamiento informal de estudiantes, de octavo grado, alrededor de la recta de mejor ajuste mediante la implementación de una THA basada en actividades computacionales.

2. Referentes conceptuales

Investigaciones alrededor del razonamiento de los estudiantes surgen de cuestiones sobre cómo explorar y aprovechar sus conocimientos previos, ya sean formales o informales, así como sus creencias, experiencias, intuiciones e inclusive errores para facilitar el camino hacia la comprensión de conceptos estadísticos. Lo anterior antes de hacer un primer acercamiento formal.

Para nuestra investigación tomaremos algunos de los conceptos que Zieffler (2006), Zieffler Garfield, delMas y Reading (2008), Carlson et al. (2002), Moritz (2004) y Zieffler y Garfield (2009) proponen para definir el Razonamiento Covariacional Informal (RCI). Para llegar a esta definición antes debemos definir o hacer distinciones entre otros términos como Razonamiento Estadístico y Pensamiento Estadístico, Conocimiento Informal y Razonamiento Informal, Recta de mejor ajuste y Recta Informal de mejor ajuste.

2.1 Razonamiento Estadístico y Pensamiento Estadístico

Conocer cómo los estudiantes razonan y piensan estadísticamente ha sido el foco de muchos investigadores en educación estadística; lo anterior, argumentándose en que la enseñanza tradicional de la estadística se reduce a procedimientos y cálculos en donde los estudiantes no desarrollan estos procesos: pensar y razonar estadísticamente (Ben-Zvi y Garfield, 2004). Sin embargo no hay consenso, entre los investigadores en este campo, de lo que significa el Razonamiento Estadístico y el Pensamiento Estadístico, algunos inclusive hacen referencia a estos términos de manera indistinta (delMas, 2004).

De manera que el interés por definir o caracterizar estos términos se ha evidenciado en investigadores como Ben-Zvi y Garfield (2004), Pfannkuch y Wild (1999, 2004), delMas (2004), entre otros. Lo anterior porque

> las similitudes y diferencias entre estos procesos son importantes al considerar formular metas de aprendizaje para los estudiantes, diseñar actividades de instrucción y evaluar el aprendizaje usando instrumentos de evaluación apropiados (Ben-Zvi y Garfield, 2004, p. 6).

Aunque no se ha llegado a un acuerdo sobre las definiciones y distinciones entre estos dos procesos, Garfiel y Ben-Zvi (2004, p. 7) lo resumen de la siguiente manera:

• *El razonamiento estadístico* puede definirse como la forma en que las personas razonan con ideas estadísticas y le dan sentido a la información estadística. Lo que implica realizar interpretaciones basadas en conjuntos de datos, representaciones de datos o resúmenes estadísticos de datos. Implica también la conexión de un concepto con otro (por ejemplo el centro y la variabilidad), o combinar ideas sobre datos y el azar. Razonar significa entender y ser capaz de explicar los procesos estadísticos y ser capaz de interpretar completamente los resultados estadísticos.

- *El pensamiento estadístico* implica comprender por qué y cómo se realizan las investigaciones estadísticas y las "grandes ideas" que subyacen a las investigaciones estadísticas. Estas ideas incluyen la naturaleza omnipresente de la variación y cuándo y cómo usar métodos apropiados de análisis de datos tales como resúmenes numéricos. Incluye una comprensión de la naturaleza del muestreo, de cómo se usan los modelos para simular fenómenos aleatorios. Incluye poder entender y utilizar el contexto de un problema para formar investigaciones, sacar conclusiones, reconocer y comprender todo el proceso, es decir desde la presentación de preguntas hasta la recopilación de datos, la elección del método de análisis para probar suposiciones, etc.

A lo anterior su suman Wild y Pfannkuch (1999, 2004) quienes identifican cinco tipos de pensamiento fundamentales para la estadística: i) Reconocimiento de la necesidad de los datos; ii) Transnumeración; iii) Consideración de la variación; iv) Razonamiento con modelos estadísticos e; v) integración de la información estadística y contextual.

Finalmente delMas (2004) argumenta que, tanto el pensamiento estadístico como el razonamiento estadístico pueden involucrarse en una misma tarea, pero que no necesariamente se pueden distinguir de acuerdo al contenido de un problema. ¿Cómo podrían distinguirse entonces? Según delMas (2004), es posible distinguirlos por la naturaleza de la tarea y ejemplifica: Una persona manifiesta un pensamiento estadístico si sabe cuándo y cómo aplicar el conocimiento y los procedimientos estadísticos; mientras que manifiesta que está razonando estadísticamente cuando es capaz de explicar y argumentar el por qué se produjeron los resultados o por qué se justifica cierta conclusión, también cuando se justifica un modelo seleccionado para ver si representa un ajuste razonable a un contexto especificado.

2.2 Conocimiento Informal y Razonamiento Informal

Zieffler et al. (2008) desarrollan un marco para investigar el Razonamiento Inferencial Informal (RII) de los estudiantes y consecuentemente hacen sugerencias sobre los tipos de tareas que se deben usar para este tipo de exploración. Otros autores como Garfield y Ben-Zvi (2008) identifican dos temas principales que caracterizan la estadística inferencial: i) Generalizar a una población mayor a partir de una pequeña muestra y ii) Determinar si a un patrón encontrado en los datos se le puede atribuir un efecto real. Señalan además que son muchas las investigaciones que se han llevado a cabo para la comprensión de estos temas por parte de los estudiantes, así como las dificultades que estos presentan. Investigaciones que a su vez han dado paso a dos líneas de investigación en educación estocástica: La Inferencia Estadística Informal (IEI) y el RII. Dado que estas líneas de investigación son relativamente recientes, su significado no era del todo claro, constituyéndose así el principal objetivo de Zieffler et al. (2008): Proporcionar una definición así como un marco para el diseño de tareas para estudiar el razonamiento de los estudiantes respecto a la inferencia estadística. Para el cumplimiento de este objetivo, los autores consideran pertinente definir, a su vez, dos términos: *Conocimiento Informal* y *Razonamiento Informal.*

- El *Conocimiento Informal,* hace referencia, por una parte, a ese conocimiento cotidiano y experiencias que se adquieren fuera del salón de clases, y, por otra parte, a ese conocimiento no tan formalizado que resulta luego de una instrucción formal previa. Los autores hacen hincapié en la importancia y uso que se le debe dar a este tipo de conocimiento para el estudio posterior formal de un tema. Tal como ellos señalan "el conocimiento informal es un punto de partida para el desarrollo del conocimiento formal" (Zieffler et al., 2008, p.3).

- El *Razonamiento Informal,* se emplea en situaciones no deductivas como tomar una decisión así como en argumentación de la misma. Los autores señalan que la mejora de este tipo de razonamiento no necesariamente ocurre con el aumento del conocimiento de contenidos, tampoco con el nivel de maduración del individuo, la experiencia o el grado de escolaridad, pero sí a través de instrucciones. Aunque no hay consenso en la definición de este tipo de razonamiento, se está de acuerdo en que la argumentación constituye una forma de razonamiento informal, en ese sentido Sanz de Acedo (2001) señala que

> …una cualidad, quizá la más importante del razonamiento informal es la capacidad que tiene para elaborar argumentos intencionados, defenderlos frente a otros posibles y evaluarlos en relación con la información disponible para llegar a una conclusión (Sanz de Acedo, 2001, p. 356).

De manera que el RII es definido "como la forma en la que los estudiantes usan su conocimiento estadístico informal para elaborar argumentos que apoyen inferencias sobre poblaciones desconocidas a partir de muestras observadas" (Zieffler, et al. 2008, p. 44).

Zieffler, et al. (2008) señalan que son tres los componentes que distinguen un RII:

- Hacer juicios o predicciones sobre las poblaciones a partir de muestras, pero sin usar procedimientos o métodos estadísticos formales.

- Hacer uso e integrar los conocimientos previos en la medida de que estos estén disponibles.

- Articular argumentos basados en la evidencia para hacer y defender juicios o predicciones sobre las poblaciones basadas en muestras.

Por otra parte, si se quiere estudiar el RII así como su desarrollo, es importante tener en cuenta los siguientes dos enfoques que orientan el diseño de tareas o actividades:

- Centrarse en la naturaleza de este razonamiento o en los métodos ingenuos de razonamiento. En otras palabras, caracterizar este tipo de pensamiento asociado a la temática en estudio y su estado actual de los individuos que participan en el estudio.

- Centrarse en los cambios o desarrollos que puedan ocurrir en la naturaleza de este razonamiento cuando los individuos son sometidos a procesos instructivos y con cierto tipo de recursos o herramientas. En nuestro caso se dispondrá de la simulación computacional.

2.3 Asociación estadística. La recta de mejor ajuste

Dado que nuestra investigación se centra en la covariación o asociación estadística alrededor de la recta de mejor ajuste, en esta sección se abordará, grosso modo, teóricamente.

Hablar, en un sentido estadístico, de la relación entre dos o más variables continuas, nos lleva a hablar de técnicas estadísticas como las de correlación y regresión. Y en un sentido general, es decir la relación entre dos o más variables cualesquiera, cuantitativas o categóricas, nos conduce a hablar de asociación estadística. De manera que la asociación estadística extiende la noción de correlación (Estepa, 1993; Estepa y Batanero, 1995). El siguiente mapa conceptual (ver Figura 4) de la noción de asociación estadística, corresponde a una parte del que propone Casey (2008, p. 94) ya que nuestra investigación se limita a datos cuantitativos, además continuos, y hacia la comprensión de la recta de mejor ajuste de manera informal usando simulación computacional.

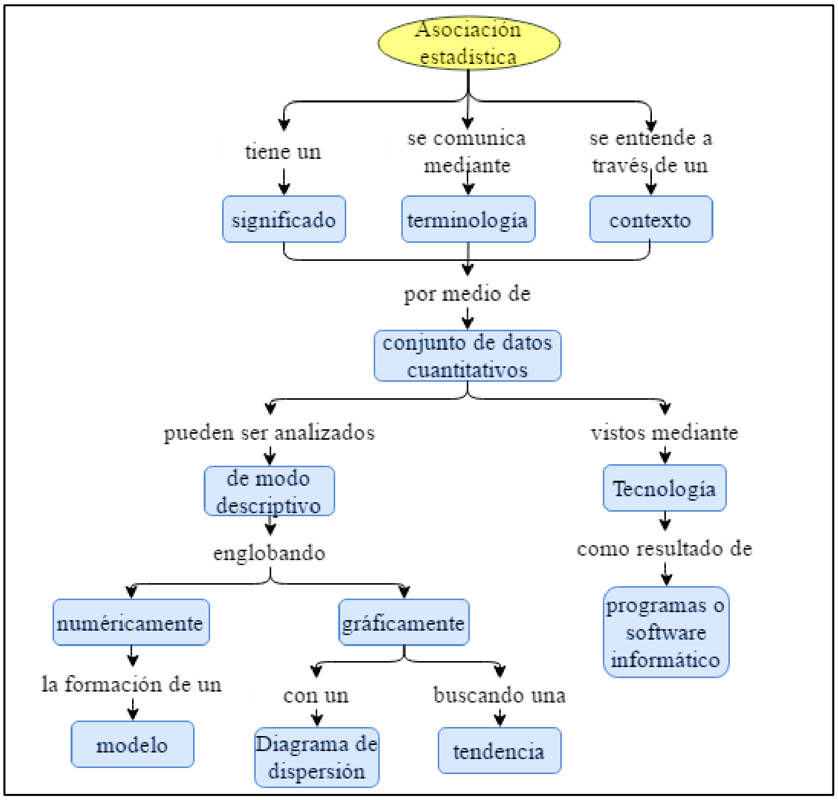


Figura 4. Mapa conceptual de la noción de asociación estadística (Tomado y adaptado de Casey, 2008, p.94).

Diagrama de dispersión

Existen fenómenos en los que no es posible encontrar una expresión que relacione, de manera funcional, todas las parejas (x_i, y_i) de las variables que determinan dicho fenómeno.

Si dichos pares de valores los representamos en un sistema cartesiano, los puntos, en general, no se ajustan de un modo preciso a una curva plana, sino que se obtiene un conjunto de puntos más o menos dispersos. Una representación de ese tipo recibe el nombre de *nube de puntos* o *diagrama de dispersión* (Batanero y Godino, 2001, p. 60).

La Figura 5 proporciona un ejemplo de un diagrama de dispersión.

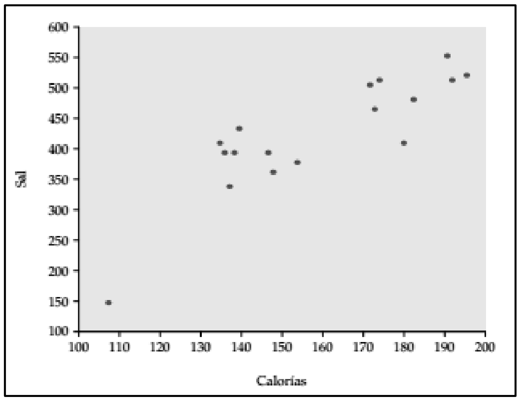


Figura 5. Diagrama de dispersión que relaciona las calorías y el contenido de sal de 17 marcas de salchichas (Tomado de Moore 2005, p. 115)

Recta de mejor ajuste formal – Método de regresión por mínimos cuadrados

Observando el diagrama de dispersión anterior, una pregunta natural es ¿de qué manera se relaciona las calorías con el contenido de sal? “Las relaciones lineales son especialmente importantes, ya que una recta es una figura sencilla bastante común” (Moore, 2005, p. 120). Ahora, si los datos manifiestan una relación de tipo lineal, ¿cómo encontrar la recta que mejor se ajusta? ¿Y para qué encontrarla? Estepa (1993, p.36) proporciona dos razones para hacer un análisis de regresión: i) se desea obtener una descripción de la relación entre las variables, como una indicación de una posible causalidad y ii) se puede desear obtener un predictor de la variable dependiente, a partir de los valores de las variables independientes. Ahora, ¿con qué intensidad se da esa relación lineal? La respuesta a esta pregunta corresponde ya a un problema de correlación.

Método de regresión por mínimos cuadrados

El método de regresión por mínimos cuadrados nos permite encontrar la recta de regresión la cual "es una recta que describe cómo cambia una variable respuesta y a medida que cambia una variable explicativa x" (Moore, 2005, p. 132). Por este método, "la recta de regresión mínimo-cuadrática de y con relación a x es la recta que hace que la suma de los cuadrados de las distancias verticales de los puntos observados a la recta sea lo más pequeña posible" (Moore, p. 135). Ahora, para encontrar, matemáticamente, la ecuación de la recta de regresión por mínimos cuadrados que relacione la variable explicativa x con la variable respuesta y, teniendo n puntos, se calcula: i) $\bar{x}$ e $\bar{y}$, que corresponden al valor promedio de cada variable; ii) s_x y s_y, es decir las desviaciones típicas para cada variable y; iii) la correlación, r, entre x e y. De manera que la recta de regresión por mínimos cuadrados está dada por:

$$\hat{y} = a + bx,$$

donde b, la pendiente, es $b = r\frac{s_y}{s_x}$ y a, la ordenada en el origen, es $a = \bar{y} - b\bar{x}$ (Moore, 2005, p. 135).

Calcular manualmente lo anterior puede resultar ser una tarea bastante tediosa, como señala Moore (2005) cualquier programa estadístico resuelve esa tarea, lo importante es comprender la recta de regresión.

Recta informal de mejor ajuste

Finalmente, expondremos brevemente a lo que hacemos referencia con *Recta informal de mejor ajuste.*

Para Casey (2015), ajustar de manera informal la recta, significa que, dados unos datos representados en un diagrama de dispersión, los estudiantes ajusten a ojo o intuitivamente, sin cálculos ni tecnología, la recta que mejor se ajusta para ellos. Concordamos con lo anterior, a diferencia de que el estudiante sí puede usar tecnología pero sin ir directamente a la herramienta "regresión".

De manera que nos tomamos la licencia de definir Recta informal de mejor ajuste, como lo llamaremos de aquí en adelante, a la recta que el estudiante ubica como producto de un razonamiento informal.

2.4 Razonamiento Covariacional

Para Zieffler (2006), después de examinar las definiciones que se le han dado al Razonamiento Covariacional desde los diferentes campos en que se ha estudiado (sicología, enseñanza de las ciencias, educación matemática y educación estadística), "el razonamiento covariacional se refiere a la forma en que la gente piensa acerca de, o razona acerca de la relación entre dos o más variables" (p. 6).

Para Moritz (2004) el "razonamiento sobre covariación general comprende los procesos de traslación entre los datos numéricos crudos, representaciones gráficas, y las declaraciones verbales sobre covariación estadística y asociación causal (Moritz 2004, pág. 165).

Ziffler (2006) afirma que las investigaciones en educación estadística sobre el razonamiento covariacional se interesan en cómo los estudiantes razonan a partir de la lectura de un diagrama de dispersión, o en la interpretación de las correlaciones y en otras tareas que conlleven al análisis de dos variables así como la interpretación de los resultados arrojados en ese análisis.

Los sicólogos definen la covariación como una expresión informal de la correlación. Estos examinan, haciendo uso de tablas de contingencia y gráficos, el juicio covariacional de la gente, su detección y evaluación del grado de asociación entre dos o más variables. En cuanto a los educadores de ciencias, estos lo hacen en el mismo sentido en que lo hacen los sicólogos solo que difiere el contexto en que lo hacen, es decir, se interesan en cómo los estudiantes usan su razonamiento covariacional en el momento de realizar un experimento científico, por ejemplo (Zieffler, 2006).

Por otra parte los educadores en matemáticas han estudiado el razonamiento covariacional desde un punto de vista determinista o funcional. En este sentido, Carlson et al. (2002) definen el razonamiento covariacional “como las actividades cognitivas implicadas en la coordinación de dos cantidades que varían mientras se atiende a las formas en que cada una de ellas cambia con respecto a la otra” (p.354).

En acuerdo con Ziffler (2006) debido a que la investigación en la enseñanza y el aprendizaje de la educación estadística es relativamente reciente en comparación con otras áreas, entonces tanto las investigaciones como la publicación de sus resultados sobre el razonamiento covariacional se hacían desde otras disciplinas como las mencionadas anteriormente, excepto desde la educación estadística. Es así, como nuestra investigación constituye un aporte para enriquecer la investigación en educación estadística, en particular en el tema de Razonamiento Covariacional, de manera aleatoria, el cual definimos en el sentido en que lo hace Ziffler: “El Razonamiento Covariacional se refiere a la forma en que la gente piensa acerca de, o razona acerca de la relación entre dos o más variables” (Zieffler, 2006, p.6). Esta definición abarca el saber cómo juzgar e interpretar esa relación (Zieffler y Garfield, 2009).

2.5 Razonamiento Covariacional Informal

En un intento por aunar las definiciones que Zieffler et al. (2008) hacen sobre Conocimiento Informal y Razonamiento Informal con la definición de Razonamiento Covariacional adoptamos la definición de Razonamiento Covariacional Informal (RCI) *como la forma en que la gente razona y argumenta acerca de la relación entre dos o más variables haciendo uso solo de su conocimiento cotidiano, experiencias previas fuera de clase y el conocimiento formal, o no tan formal, que ya ha adquirido en una instrucción previa*. Y en el mismo sentido en el que lo hacen Zieffler et al. (2008) con el RII, identificamos los siguientes componentes que señalan que un individuo está haciendo uso de su RCI cuando:

- Infiere sobre las relaciones entre dos variables.
- Infiere sobre si una variable explica o causa los cambios de la otra (variables explicativas y variables respuesta).
- Infiere si la relación tiene un patrón o comportamiento.
- Infiere sobre la medición (fuerza y sentido) de la relación entre dos variables.

2.6 Herramientas computacionales

Uno de los principios para las matemáticas escolares que señala el NCTM (2000) es *El principio tecnológico.* El uso de calculadoras, computadores, software y applets facilitan y permiten una mejor visualización de propiedades en objetos matemáticos que, en algunos casos, con mucho esfuerzo se verían en papel y lápiz. Cuando se dispone y se usan herramientas tecnológicas “los alumnos pueden centrar su atención en tomar decisiones, reflexionar, razonar y resolver problemas… formular y explorar conjeturas” (NCTM, 2000, p. 26), más que en la tediosa tarea de hacer cálculos o cuentas, imaginar movimientos o representaciones. Por otra parte, dentro de los Estándares de contenidos para las matemáticas escolares que propone el NCTM (2000) está *el Análisis de Datos y Probabilidad*.

Este Estándar recomienda que los alumnos formulen preguntas que puedan contestarse mediante datos y que afronten lo que esto requiere: la recogida de los datos y su acertado uso. Deberían aprender a recoger datos, organizar los propios y los ajenos, y representarlos en gráficos y diagramas que resulten útiles para responder a las preguntas. Incluye también el aprendizaje de algunos métodos para analizar los datos y algunas formas de hacer inferencias y obtener conclusiones a partir de ellos (p. 58).

En estadística para enriquecer y hacer frente a lo que se propone en el párrafo anterior, se disponen de mediadores semióticos (Moreno, 2014) como: SPSS, STATGRAPHICS, Fathom, EXCEL, Geogebra entre otros, los cuales permiten analizar conjuntos de datos grandes, simularlos, representarlos, graficarlos; en otras palabras como la esencia misma de la estadística, ver más allá de los datos y así razonar estadísticamente.

Como ya se ha mencionado, en esta investigación se desarrollaron actividades basadas en programas computacionales como Fathom y Geogebra para analizar las formas del razonamiento covariacional informal de los estudiantes.

3. Metodología

En esta sección presentamos la metodología que llevó a cabo esta investigación. Está dividida en cuatro apartados: i*) Marco metodológico*; ii) Descripción de la *población y muestra*; iii) *Herramientas de simulación* y; iv) Descripción de las *Fases de la investigación.*

3.1 Marco metodológico

El marco metodológico para esta investigación se basa en uno de los elementos del *Ciclo de enseñanza de las matemáticas* que propone Simon (1995). Ese elemento corresponde a las Trayectorias Hipotéticas de Aprendizaje (THA).

Las THA fueron introducidas por Martin A. Simon en el año 1995 como una complementación de su "*Ciclo de enseñanza de las matemáticas"* el cual constituye un modelo esquemático que relaciona de manera cíclica los siguientes elementos: El conocimiento del profesor de matemáticas; el conocimiento de las representaciones matemáticas, materiales y actividades; la evaluación de los conocimientos de los estudiantes; las hipótesis del profesor acerca del conocimiento de los estudiantes; las teorías del profesor sobre la enseñanza y el aprendizaje de las matemáticas; el conocimiento del profesor sobre el aprendizaje de los estudiantes de un contenido particular y; las THA.

Tal como lo señala Simón (1995), un principio en la enseñanza, y no solo de las matemáticas, es que "la única cosa predecible en la enseñanza es que las actividades de clase no irán como se predicen" (p.133), es decir, se pueden tener bien definido dos cosas: Primero;

el objetivo de enseñanza (el concepto como tal, en nuestro caso covariación) y segundo; el plan de aula (o de instrucción) para alcanzar dicho objetivo. Pero tanto el primero como el segundo se verán modificados, y no solo una vez sino de manera constante cuando los estudiantes se involucren en la ejecución del plan, es decir; esa modificación se da como producto de la interacción entre el profesor y los estudiantes y en general, de la interacción entre todos los elementos del *Ciclo de enseñanza de las matemáticas* previamente mencionado. La Figura 6 muestra cómo se da la relación cíclica entre esos elementos.

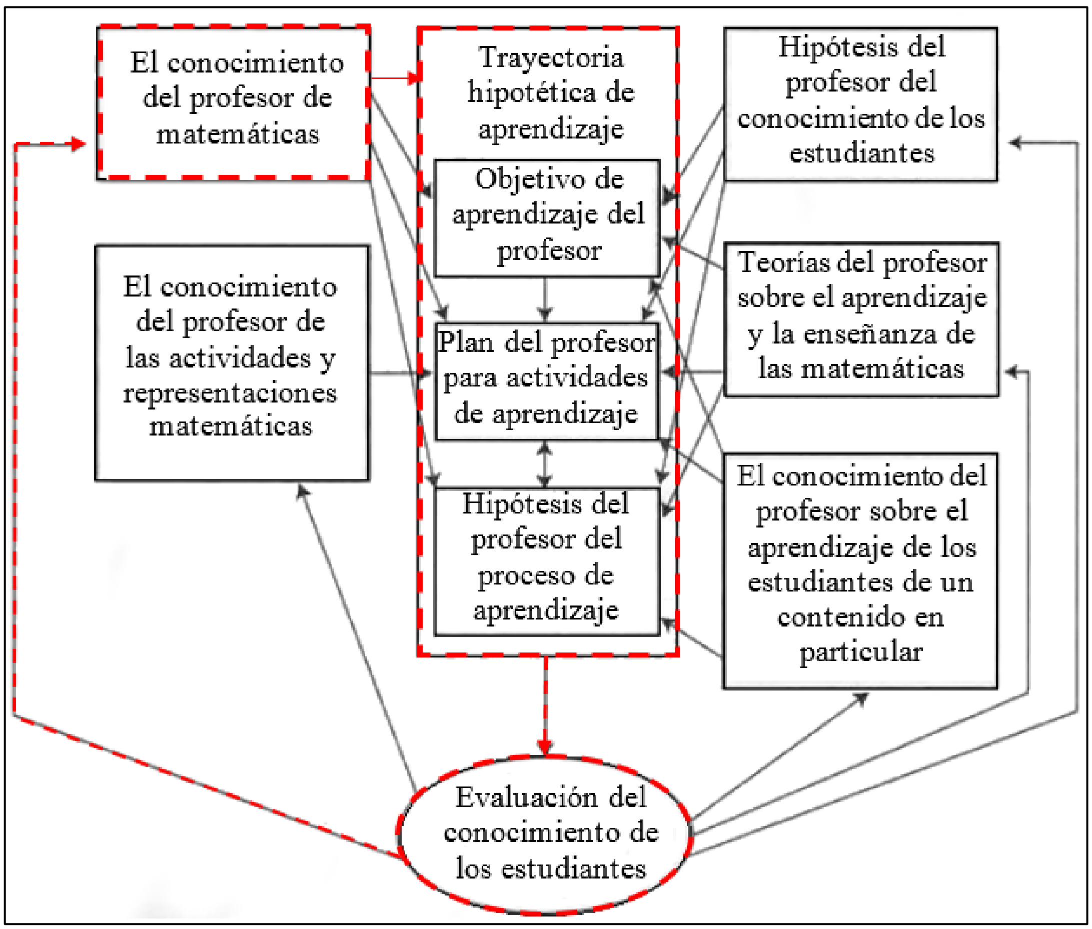


Figura 6. Ciclo de enseñanza de las matemáticas (Simon, 1995, p. 137) (Contorno punteado agregado “constitución interactiva de las actividades de clase”, el cual corresponde a la versión abreviada del ciclo).

Para nuestro interés, solo entraremos en detalle en la versión abreviada del ciclo, donde la THA constituye una parte central del ciclo de enseñanza de las matemáticas pero ¿qué se entiende por trayectoria hipotética? Para darle un sentido figurativo y un mejor entendimiento vamos a tomar la analogía que hace Simón (1995, p. 136):

> Imagina que has decidido navegar alrededor del mundo con el fin de visitar lugares que nunca has visto. Lo anterior no se hace al azar (por ejemplo, ir a Francia, luego a Hawai, a continuación a Inglaterra), pero tampoco hay un itinerario fijo a seguir. De manera que hay que adquirir el mayor conocimiento relevante como sea posible para la planificación de tu viaje. Dado lo anterior, haces un plan. Es posible planificar inicialmente el viaje, o solo parte de él. Propones la navegación de acuerdo a tu plan. Sin embargo, se debe ajustar constantemente debido a las condiciones que encuentres. Luego se continúa para adquirir conocimientos sobre la navegación, sobre las condiciones actuales, y sobre las áreas que desees visitar. Cambias de planes con respecto al orden de los destinos. Modificas la duración y la naturaleza de tus visitas como resultado de la interacción con la gente en el camino. Agregas destinos que eran desconocidos antes del viaje para ti. El camino que viajas es tu "trayectoria"; el camino que se anticipa en cualquier punto en el tiempo es tu "trayectoria hipotética".

De manera que la THA se refiere al "camino" que podría dar paso al aprendizaje, también describe cómo los estudiantes entienden un concepto en particular y cómo este debe progresar (Bargagliotti, 2014); y es hipotética porque se desconoce la trayectoria real de aprendizaje. Tres elementos la componen:

i. La consideración del objetivo de aprendizaje; el cual direcciona la THA.

ii. Las actividades de aprendizaje y;

iii. El aprendizaje y el pensamiento en el que los estudiantes pueden participar (hipótesis del proceso de aprendizaje)

A partir del objetivo se decide, diseña y rediseña la THA basándose en lo que acontece en el momento de la implementación de la misma y en las nuevas hipótesis de aprendizaje o razonamiento que van surgiendo. Por hipótesis nos referimos a cómo se espera que tanto el razonamiento como la comprensión de los estudiantes mejoren cuando se vean enfrentados a las actividades de aprendizaje diseñadas bajo simulación computacional. Notemos, además, en la Figura 6 el sentido con el que se relacionan los últimos dos elementos de la THA; Simon (1995) llama a esta relación una *relación simbiótica*, es decir, una depende de la otra.

Por otra parte, Simon y Tzur (2004, p.93) consideran los siguientes supuestos para la construcción de una THA:

1. La generación de una THA se basa en la comprensión de los conocimientos actuales de los estudiantes involucrados.

2. Una THA es un vehículo para el aprendizaje de la planificación de determinados conceptos.

3. Las tareas proporcionan herramientas para promover el aprendizaje de determinados conceptos y son, por lo tanto, una parte clave del proceso de instrucción.

4. Debido a la naturaleza hipotética e inherentemente incierta de este proceso, el profesor (investigador en nuestro caso) participa regularmente en la modificación de todos los aspectos de la THA.

Tal como lo hicieron Cobb, McClain y Gravemeijer (2003), tomaremos como unidad de análisis la *microcultura,* establecida por la comunidad de clase. Lo anterior para documentar la evolución y formas del razonamiento alrededor de la recta de mejor ajuste que sean comunes en el aula. Por *microcultura* los autores hacen referencia, y adoptamos, al contexto social inmediato donde acontece el aprendizaje de los estudiantes. Este tipo de enfoque analítico diferencia tres tipos de normas en el aula:

1. *Las normas sociales en el aula*, que sirven para documentar la estructura de la participación en clase. Estas normas incluyen que el estudiante justifique y explique sus respuestas, también incluye que los estudiantes le den o le encuentren sentido a las explicaciones de sus compañeros manifestando así su comprensión o no compresión de la explicación del otro. Dentro de estas normas también está la de hacer preguntas aclaratorias y desafiar los diferentes razonamientos que se evidencien en los estudiantes.

2. *Las normas sociomatemáticas,* por ejemplo, las que incluyen los criterios que se establecen para lo que se considera una solución matemática diferente, sofisticada o eficiente, así como lo que se considera como una explicación o razonamiento matemático coherente. Se centran en las acciones e interacciones específicas de las matemáticas.

3. *Los significados matemáticos normativos*, los cuales son más específicos ya que se centran en las ideas matemáticas que conducen a la aparición de contenidos matemáticos bajo estudio. Al igual que lo hacen Cobb, McClain y Gravemeijer (2003) nos enfocaremos en

las formas de hablar y razonar de los estudiantes acerca de datos bivariados. Se toma como *normativo* los significados matemáticos que se constituyen como legítimos en el aula, es decir que son comunes o aceptables en la comunidad de la clase y que no genera discusión o justificación sobre esa forma de razonar en particular.

3.2 Población y muestra

La población sobre la cual se realizó el estudio fueron estudiantes de bachillerato, específicamente de octavo grado; la selección de este grado se debió a que, para este nivel, los estudiantes no han recibido ningún tipo de enseñanza formal en los temas de covariación, correlación y regresión, inclusive, matemáticamente hablando, sobre rectas. La muestra constituyó un grupo de estos estudiantes de la Institución Educativa San Francisco (Piedecuesta).

3.3 Herramientas computacionales

Para la construcción de las actividades involucradas en la THA se utilizaron los programas Geogebra y Fathom.

3.4 Fases de la investigación

Las fases de esta investigación son cuatro: i) Implementación y análisis de dos pruebas diagnósticas, ii) Elaboración de las actividades de la THA, iii) Implementación de la THA y, vi) Análisis de resultados.

3.4.1 Implementación y análisis de dos pruebas diagnósticas. En primer momento, se elaboró y aplicó una prueba inicial de concepciones informales con el fin de diagnosticar y caracterizar el razonamiento covariacional informal de los estudiantes, los estudiantes disponían de dos horas para responderlo. También con este test se pretendió evaluar el conocimiento actual, formal o informal, que tenían los estudiantes. El análisis posterior de las respuestas a este test nos llevó al diseño y aplicación de un segundo test para ratificar algunas conjeturas acerca del razonamiento de los estudiantes en tareas de predicción y ubicación de la recta de mejor ajuste. Este segundo test se aplicó en menos de un mes, respecto al primero, a una muestra de estudiantes con las mismas características de la que se aplicó el primer test, con una disponibilidad también de dos horas para responderlo. De manera que los resultados del análisis de las dos pruebas iniciales, junto con el desarrollo de la fase 1, nos permitió generar y validar hipótesis sobre el conocimiento de los estudiantes alrededor de los objetos estadísticos que se abordaron.

3.4.2 Elaboración de las actividades de la THA. Con base en los resultados obtenidos en la fases I y la revisión bibliográfica sobre concepciones y/o errores de los estudiantes alrededor de la recta de mejor ajuste, se construyó un plan de actividades así como el planteamiento de hipótesis y/o conjeturas alrededor de la forma como los estudiantes razonan de manera informal sobre variables que están conjuntamente relacionadas y cómo este razonamiento se desarrollará a lo largo de la implementación de las mismas. Esta fase no fue otra cosa más que el diseño de la THA que estaba direccionada de acuerdo al objetivo del aprendizaje: la recta informal de mejor ajuste. Esta etapa duró aproximadamente cuatro meses, sin embargo las actividades fueron continuamente rediseñadas y/o ajustadas de acuerdo a las formas de razonamiento de los estudiantes y a las discusiones que se hacían con ellos a lo largo de la implementación de las mismas.

3.4.3 Implementación de la THA. En esta fase se llevó a cabo la implementación de la THA trazada en la fase anterior. Fue en esta etapa donde se interactuó con los estudiantes y, resultado de esta interacción, en algunos momentos surgió la necesidad de rediseñar o ajustar la THA para la siguiente sesión de trabajo. Esta implementación duró, aproximadamente, tres meses.

3.4.4 Análisis de resultados. El diseño y posterior ajuste (para cada sesión) de la THA, basada en actividades computacionales, permitió un análisis retrospectivo que finalmente nos llevó a describir las formas de razonamiento covariacional informal de los estudiantes así como su desarrollo desde la prueba diagnóstica hasta el final de la implementación de la THA.

La recolección de los datos analizados se hizo de diferentes maneras: En primer lugar, contábamos con las respuestas escritas por los estudiantes, adicional a ello, debido a que las actividades tenían gran parte computacional, en todo momento se grabó cada una de las pantallas que incluía, además, video cámara y voz para observar los ademanes que hacían los estudiantes o, si hacían algún comentario. Por otra parte, contábamos con otra cámara que grababa, en general, el discurso que se tenía en clase con los estudiantes. Al final, contamos también con dos entrevistas: La primera se hizo de manera grupal y la segunda de manera individual, alrededor de la prueba final, con el fin de obtener más conocimiento del razonamiento informal actual de los estudiantes, alrededor de la recta de mejor ajuste, finalizada la implementación y discusión de las actividades; estas entrevistas también fueron video grabadas.

El análisis fue de manera retrospectiva en el sentido de que para el diseño de una actividad específica, se tenían una serie de hipótesis y el posible discurso que se daría en clase. Ahora, el análisis de los discursos y de las respuestas los estudiantes, validaría o refutaría las hipótesis que se tenían llevándonos a rediseñar o ajustar la actividad que venía y a plantear nuevas hipótesis. La modificación de la THA es inherente dado lo anterior. En el análisis se buscaban formas comunes de razonamiento alrededor de nuestros ejes temáticos principales como eran variabilidad y predicción tanto en el caso univariado con la media, como en el caso bivariado con la recta de mejor ajuste. Lo anterior con el fin de llevar un seguimiento de estas formas y cómo estas iban cambiando en los estudiantes a lo largo de la implementación de las actividades. Estas mismas formas, de alguna manera, direccionaban, validaban o cambiaban el rumbo de la THA.

4. Análisis de Resultados

4.1 Análisis de la prueba diagnóstica 1: Interpretando gráficos

A continuación el análisis de la prueba diagnóstica que se aplicó a 39 estudiantes de octavo grado quienes no habían visto previamente conceptos de estadística como diagramas de dispersión y recta de mejor ajuste. La prueba constaba de doce ítems, de opción múltiple, relacionados con cinco diagramas de dispersión.

Los cinco diagramas de dispersión presentaban características diferentes, como se describirá más adelante, y lo que se pretendía evaluar era: i) si los estudiantes establecían o no una relación entre las dos variables; ii) conocer sus concepciones sobre la variabilidad; iii) los criterios que utilizan para predecir el valor de la variable respuesta dado un valor para la variable explicativa y; iv) conocer los criterios con que trazan la recta que mejor se ajusta a los datos y con qué precisión lo hacen. En pocas palabras, su nivel de razonamiento covariacional informal actual.

Ítems 1 a 3

Para los ítems 1 a 3 se mostraba un diagrama que relacionaba las variables estatura y peso de 30 niños entre los 12 y 13 años (Figura 7) las cuales tenían una fuerte asociación lineal positiva (r=0,98). Aquí se pretendía evaluar las ideas intuitivas y las estrategias de los estudiantes para describir un gráfico de dispersión con este tipo de asociación, en tareas que involucran asociación y la predicción.

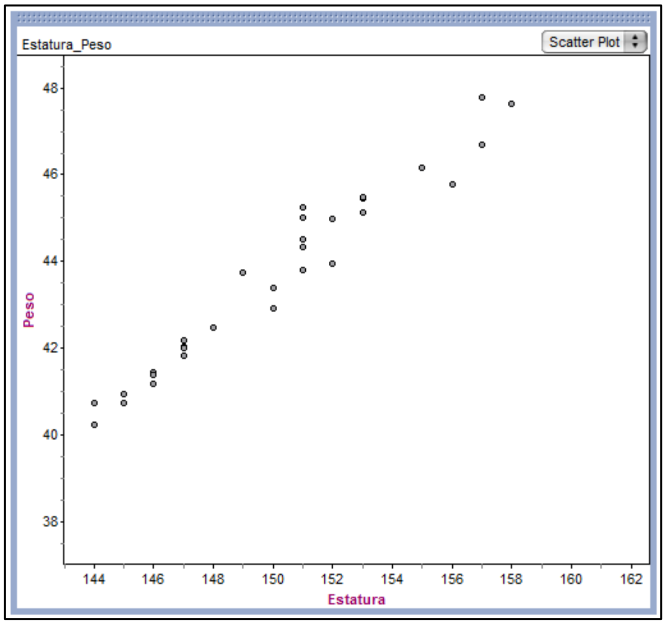


Figura 7. Estatura (en cm) y peso (en kg) de 30 niños de un curso de octavo grado entre 12 y 13 años.

Tabla 3.
Ítem 1 Prueba diagnóstica 1

1. Respecto a la relación entre la estatura y el peso de un niño entre 12 y 13 años se puede concluir que

a. los estudiantes con menor estatura tendrán un peso mayor.
b. los estudiantes con mayor estatura tendrán un peso mayor.
c. un estudiante con determinada estatura puede tener más de un peso.
d. no es posible establecer una relación entre la estatura de los estudiantes y el peso.
e. Otra opción ¿cuál?

Para el ítem 1, 21 (54%) estudiantes aciertan en la respuesta (b); sin embargo 19 estudiantes basan sus argumentos en sus creencias previas, en su contexto, en decir, tienen una *concepción causalista* (Estepa & Batanero, 1995): *"Porque a medida que van creciendo se van alimentando más y van aumentando su peso, energía"* o *"Porque un estudiante cuando es más alto pesa más, porque sus huesos son más largos y más pesados que los de una persona pequeña y tienen más masa muscular"*. Solo 2 estudiantes se apoyan en el gráfico para dar solución a la pregunta, sin embargo en sus argumentos se evidencia una *concepción local* (Estepa & Batanero, 1995) ya que sus juicios se justifican en algunos datos del gráfico, por ejemplo*: "Porque en la gráfica se ve que entre los que miden 1.52 a 1.58 tienen un peso entre 44 a 48 kg el mayor peso entre estudiantes de 8"* o *"Porque el estudiante que tiene de estatura 158 tiene 48 de peso en cambio el estudiante que tiene de estatura 144 pesa entre los 40 a los 41 o sea entre menos estatura menor peso y entre más estatura mayor peso"*.

La segunda opción con mayor escogencia fue la c. El siguiente argumento refleja un sentido de la variabilidad *"La c porque un estudiante que por ejemplo la estatura sea de 151 cm puede pesar más de un peso. Lo cual quiere decir que el peso de un estudiante puede ser variado según la estatura"*.

Tabla 4.
Ítem 2 Prueba diagnóstica 1

2. Para un estudiante que mide 154 cm, se puede predecir que su peso, aproximadamente,

a. sería de 44 kg
b. sería de 46 kg
c. estaría entre 44 kg y 46 kg
d. El gráfico no proporciona información para predecir ese valor.
e. Otra opción ¿cuál?

En el ítem 2, se pedía predecir el peso de un estudiante que mide 154 cm. En un sentido formal, evaluando 154 cm en la ecuación de la recta que mejor se ajusta, el peso para el estudiante sería de aproximadamente 45,6 kg aproximadamente. En ese sentido 27 (70%) estudiantes se acercaron muy bien a este valor al escoger la opción a, b o c. Sin embargo sus argumentos se basan en sus creencias o en su contexto más cercano, por ejemplo la estatura y el peso de ellos mismos: *"¿Cuánto puede pesar un estudiante de estatura de 154? Yo creo que 45 kg porque yo mido 150 y peso 40 kg".* Nuevamente se evidencia una concepción causal cuando se presentan argumentos como: *"Porque estos pesos pueden ser su peso correspondiente para un niño de esta edad".* Algunos estudiantes manifiestan que aunque ese valor no se muestre en el gráfico sí es posible predecir el peso: *"En el gráfico no muestran el peso, pero se puede decir que está entre ese peso 44 y 46 kg"* lo anterior lo hacen ubicando 154 en el eje horizontal y observando los valores que podría tomar en el eje vertical. Otros estudiantes basan sus argumentos en lo que está pasando en el gráfico para valores cercanos a la estatura dada: *"Porque no está definido cuánto pesa según los otros estudiantes que miden casi lo mismo entre 44 kg y 46 kg".*

Dentro de los argumentos de los estudiantes que escogieron la opción d (26%) consideran que la única manera de predecir este valor es ver al estudiante cuya estatura es de 154 cm, por ejemplo: *"Porque no me concluye bien la altura para sacar el promedio porque necesito saber cómo es el chico"* o que necesariamente ese valor debería estar en la gráfica para dar una respuesta acertada: *"Ninguna porque en la gráfica no me está dando algunos de los valores que pueda pesar"* o *"Porque en el gráfico no hay una información, de lo que aquí se pregunta para nosotros poder [dar] una respuesta clara".*

Tabla 5.
Ítem 3 Prueba diagnóstica 1

3. Si alguien te pregunta "*¿cuánto podría pesar un estudiante que mide 151 cm?*" Y tuvieras que dar un solo valor ¿cuál de las siguientes opciones escogerías? (Nota: Para 151 cm el gráfico proporciona los siguientes cinco valores (en kg), aproximadamente: 45,24; 45,02; 44,50; 44,34 y 43,80).

a. 45,24 kg porque es el mayor valor.
b. 43,80 kg porque es el menor valor.
c. 44,50 kg porque está "en el medio" de los cinco valores.
d. 44,58 kg porque es el valor promedio de los cinco valores.
e. Otro valor ¿cuál y por qué?

Para el ítem 3, como se observa en la Figura 7, hay 5 estudiantes que miden 151 cm y tienen diferente peso. Dentro de las estrategias que los estudiantes usan para emitir un solo valor están:

Tabla 6.
Estrategias que los estudiantes usan para emitir un solo valor en el ítem 3 de la Prueba diagnóstica 1

Porcentaje (%)	Criterio
26	El mayor valor
15	El menor valor
33	La mediana
21	El promedio

Sin embargo los argumentos no se basan en lo que muestra el gráfico sino en su contexto y creencias previas: *"Escogí la respuesta d porque un estudiante que pesa 151 debe tener un promedio de peso de 44 a 58 balanceado a su estatura".* Algunos estudiantes que escogieron la opción c sí basan sus argumentos en la gráfica y en particular en la mediana al emitir juicios como: *"44.50 Porque está en medio de los dos valores: Está en medio de los cuatro valores en la gráfica se ve que está entre 44 y 45 y el punto está en la mitad y como en los resultados estaba en las respuestas 44.50 era la más acertada".* Otros manifiestan una *concepción determinista* (Estepa & Batanero, 1995) al manifestar que debería existir para un valor de la variable independiente, un solo valor para la variable dependiente: *"Porque en la gráfica el 51 [151] está en la mitad y se supone que de acuerdo a la estatura hay un determinado peso…".*

Ítems 4 y 5

Para los ítems 4 y 5 se mostraba un diagrama que relacionaba las variables año y tiempo (en segundos) de los medallistas olímpicos (hombres) en atletismo en los 100 metros planos (Figura 8), con una fuerte asociación lineal negativa (r=-0,89). Al igual que en los ítems 1 a 3, aquí se pretendía evaluar las ideas intuitivas y las estrategias de los estudiantes para describir un gráfico de dispersión con este tipo de asociación, asociación negativa fuerte, en tareas que involucran asociación y predicción.

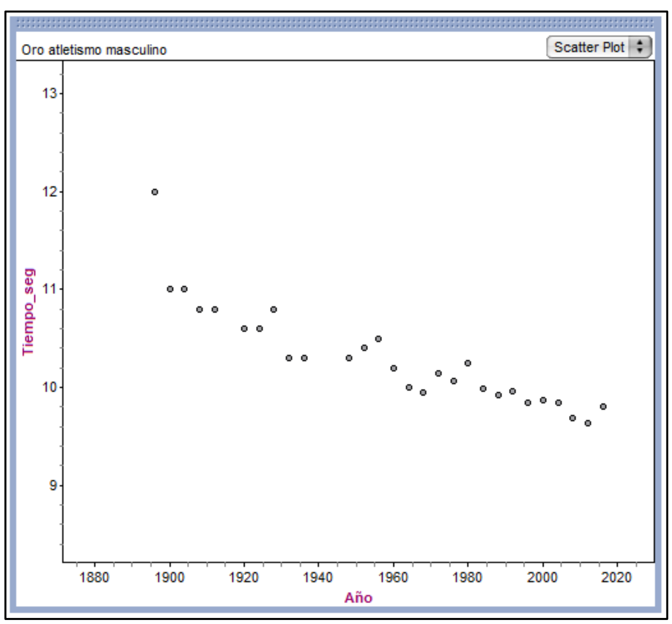


Figura 8. Año y tiempo (en segundos) de los medallistas olímpicos (hombres) en atletismo en los 100 metros planos

Tabla 7.
Ítem 4 Prueba diagnóstica 1

4. Respecto a la relación entre el año y el tiempo en que tardan en correr 100 metros planos los medallistas olímpicos, se puede concluir que

a. el tiempo en el que los medallistas olímpicos corren 100 metros planos ha disminuido a lo largo del tiempo.
b. el tiempo en el que los medallistas olímpicos corren 100 metros planos ha aumentado a lo largo del tiempo.
c. el tiempo en el que los medallistas olímpicos corren 100 metros planos no ha cambiado a lo largo del tiempo.
d. No es posible establecer una relación entre el año de los juegos olímpicos y el tiempo en que tardan los medallistas olímpicos en correr 100 metros planos.
e. Otra opción ¿cuál y por qué?

En el ítem 4, 67% (26) de los estudiantes responden acertadamente, opción a. Al igual que en el ítem 1, en un intento por justificar lo que observan en el diagrama, algunos basan sus argumentos en sus creencias e intentan, así, explicar la razón del por qué se da esa relación, por ejemplo: "*La a porque a través de los años van progresando más y tardan menos tiempo para recorrer 100 metros y esto se logra por medio de entrenamientos*" o "*Porque se han ejercitado y cada año disminuyen el tiempo*". A otros les resulta más que suficiente lo que el gráfico evidencia y tienen además una visión global de los datos, por ejemplo: "*El tiempo en el que corren los medallistas ha disminuido porque en la gráfica se ve que a los largo del tiempo corren más rápido y tardan menos en los 100 metros planos*".

Un estudiante habló en términos de promedio cuando en realidad cada punto representaba un deportista: "*Sí porque antes en un promedio de 12,5 segundos corrían 100 metros ahora en un promedio de 11,5 segundos corren los 100 metros*". El siguiente argumento se centró en el que resultaba ser un dato atípico, (1896; 12): *"Porque el primer punto está más arriba que los demás puntos"*; aunque su mirada es global, lo hace en el sentido de mirar lo extraño, o lo más grande respecto a lo demás.

Tabla 8.
Ítem 5 Prueba diagnóstica 1

5. Debido a la primera y segunda Guerra Mundial, los juegos olímpicos no se celebraron en los años 1916, 1940 y 1944. De acuerdo a lo que se observa en el Gráfico B, si se hubieran realizado los juegos olímpicos para el año 1944, se puede predecir aproximadamente, que el tiempo que hubiera tardado el medallista olímpico de ese año en correr los 100 metros planos

a. es de 10,47 segundos.
b. es de 10,30 segundos.
c. está comprendido entre 9,63 y 12 segundos.
d. El gráfico no proporciona información para predecir ese valor.
e. Otra opción ¿cuál y por qué?

En el ítem 5, el 79% (31) de los estudiantes escogen entre la opción a, b y c las cuales se consideran acertadas ya que en, un sentido formal, al evaluar este valor en la ecuación de la recta que mejor se ajusta se tiene que para el año 1944 el medallista olímpico tardaría aproximadamente 10,47 segundos en correr los 100 metros planos. Dentro de las estrategias para dar un valor aproximado está observar lo que pasa alrededor del valor que se pedía: "*Porque en los años anteriores el promedio de los corredores era de más o menos de 10,52, 10,27 y esa era la capacidad que tenían los corredores en el año 1944"* o "*Es de 10,30 porque entre 1940 y 1960 hay un punto que está cerca de los 10 segundos y muy cerca del 1940 deduje que los 10,30 es la respuesta correcta*". Esta estrategia de observar lo que pasa alrededor del valor para el cual se quiere predecir, considera, de alguna manera, rectas ajustadas parciales.

Ítems 6 a 9

Los ítems 6 a 9 se correspondían con un diagrama que relacionaba las variables minutos y pulsaciones de 23 sesiones de natación del profesor Moore (Figura 9). Las variables tenían un nivel de asociación negativa (r=-0,75). Además de evaluar lo descrito en los ítems anteriores, en particular con el ítem 8 se pretendía que los estudiantes reflexionaran sobre cómo se podría predecir el valor de la variable respuesta para cualquier valor dado en la variable explicativa; es decir, con este ítem se quería sembrar la necesidad de la construcción de una recta para señalar la tendencia lineal que mostraba el diagrama.

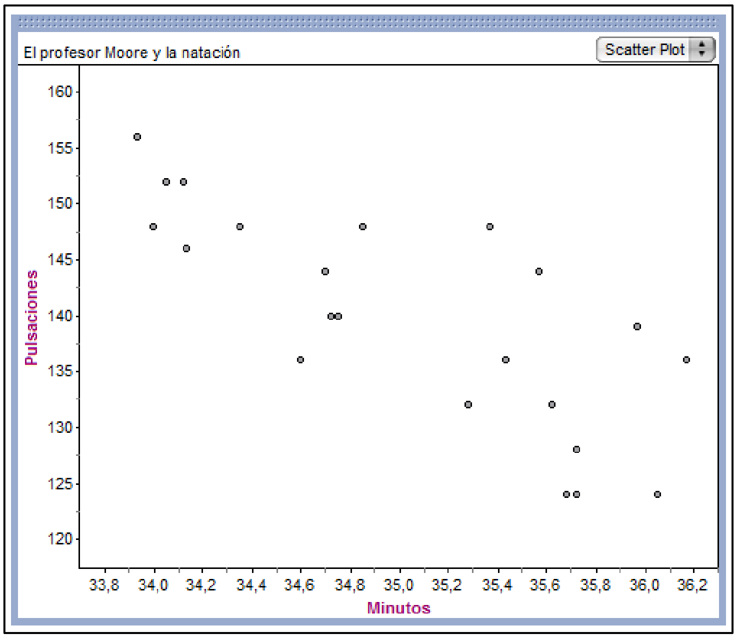


Figura 9. Tiempos (en minutos) que tarda el profesor Moore en nadar 1800 metros y pulsaciones por minuto después de bracear en 23 sesiones de natación

Tabla 9.
Ítem 6 Prueba diagnóstica 1

6. Respecto a la relación entre el tiempo que tarda el profesor Moore en nadar los 1800 metros y su ritmo cardíaco, se puede concluir que

a. un número mayor de pulsaciones indica que el profesor Moore tardó más tiempo en nadar los 1800m.
b. entre más tiempo tarde el profesor Moore en nadar los 1800m su rimo cardíaco será mayor.
c. un número menor de pulsaciones indica que el profesor Moore tardó menos tiempo en nadar los 1800m.
d. entre menos tiempo tarde el profesor Moore en nadar los 1800m su ritmo cardíaco será mayor.

e. Otra opción. ¿cuál y por qué?

Para el ítem 6, el 46% (18) de los estudiantes escogen la opción acertada (d) y sus argumentos son del mismo tipo como en los ítems anteriores donde se pedía la relación entre las dos variables. Puede pensarse que entre los estudiantes que escogieron la opción a, b y/o c, 20 de 39, comprendieron de otra manera el enunciado del problema o que no tuvieron en cuenta toda la información, por ejemplo: "*b. Porque si el profesor nada más sus pulsaciones cardíacas serán más y más. c. Porque si nada menos pues sus pulsaciones serían menores a las que sucederían si nadara más",* es decir, no tomaron en cuenta que era el tiempo en nadar los 1800 metros.

Tabla 10.
Ítem 7 Prueba diagnóstica 1

7. Juan, un alumno del profesor Moore, asegura que si para la siguiente sesión el profesor Moore tarda 34,30 minutos en nadar los 1800m, entonces su ritmo cardíaco será de 152 pulsaciones por minuto, aproximadamente. Respecto a la afirmación que hace Juan, tú estarías

a. de acuerdo ya que para 34,30 minutos le corresponde exactamente 152 pulsaciones.
b. en desacuerdo, ya que el gráfico no me proporciona información para ese valor.
c. de acuerdo, ya que el gráfico me proporciona información para predecir ese valor.
d. en desacuerdo, ya que la única manera de saber esa información es que el profesor tarde exactamente 34,30 minutos en nadar los 1800m.
e. Otra opción ¿cuál?

El ítem 7 tenía que ver con predicción, 5 estudiantes (13%) responden acertadamente (c) y, a diferencia de los ítems anteriores, pareciera que abandonan la estrategia de mirar lo que ocurre con los puntos "vecinos" al que se está pidiendo: "*Si porque el cuadro me ayuda a saber si es verdad*", "*El gráfico nos proporciona más de los valores que se encuentran en el* *gráfico*”. Un poco más de la mitad de los estudiantes estuvieron en desacuerdo (52%) considerando que la única manera de saber esa información era que el profesor tardara exactamente ese tiempo y medir sus pulsaciones, o que ese dato no estaba en el gráfico, es decir que no es posible interpolar. Lo anterior pudo darse tal vez por el grado de dispersión del diagrama mostrado en la Figura 9.

Tabla 11.
Ítem 8 Prueba diagnóstica 1

8. María, una alumna del profesor Moore, asegura que es posible predecir el ritmo cardíaco del profesor Moore para cualquier tiempo, comprendido entre 33,8 y 36,2 minutos, que tarde en nadar los 1800m. Respecto a la afirmación que hace María, tú estarías

a. de acuerdo, ya que para cualquier tiempo se puede tomar un valor cercano a éste y observar el correspondiente número de pulsaciones, el cual será muy similar al que se busca.
b. en desacuerdo, porque para algunos valores de tiempo el gráfico no muestra el correspondiente número de pulsaciones.
c. de acuerdo, ya que se puede establecer el aspecto general de la relación entre las dos variables.
d. en desacuerdo, ya que no se puede establecer el aspecto general de la relación entre las dos variables.
e. Otra opción ¿cuál y por qué?

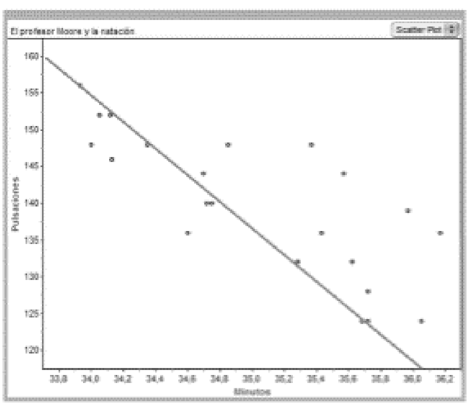
Aquí, ítem 8, 49% de los estudiantes estuvieron de acuerdo (opciones a y c), aunque reflejan una concepción local en los argumentos que emiten, al escoger la opción a: *“Se puede tomar un valor cercano a lo que se da es más los números eran 33,8 y 36,2 y ya los había hecho el profesor Moore”* o *“Escogí la a porque se puede acercar a el valor más cercano que se desea simular”* o *“De acuerdo porque podemos tomar un valor cercano a este y comprobar aproximadamente las pulsaciones y ver si es similar al que necesitamos encontrar* *para poder estar seguros".* Dentro de los que escogieron la opción c, 8 estudiantes, ninguno evidencia cómo sería el aspecto general de esa relación, es decir una recta, es más, ninguno de ellos emite un argumento coherente, puesto que no es una tarea sencilla de hacer.

El ítem 9 se diseñó con base en las concepciones encontradas por Casey (2015) sobre los criterios que usan los estudiantes para ubicar la recta que mejor se ajusta a los datos. Con este ítem se quería indagar sobre cómo nuestros estudiantes ubicaban la recta que mejor se ajusta, es decir, si están de acuerdo con algunos de los criterios encontrados por Casey (2015) o si proponen algún otro criterio como se considera en la última opción de respuesta.

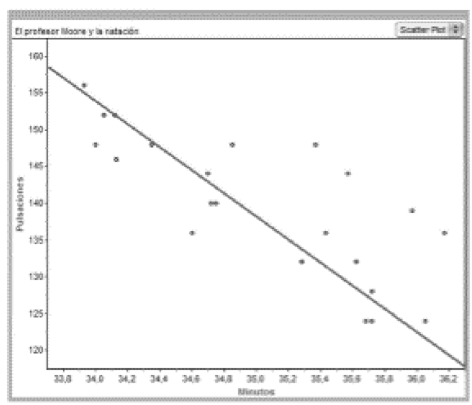
Tabla 12.
Ítem 9 Prueba diagnóstica 1

Se les pidió a cuatro estudiantes que trazaran una recta sobre el Gráfico C teniendo en cuenta que la recta se ajustara de la mejor manera a los datos del gráfico. Las siguientes gráficas muestran la respuesta de los cuatro estudiantes.

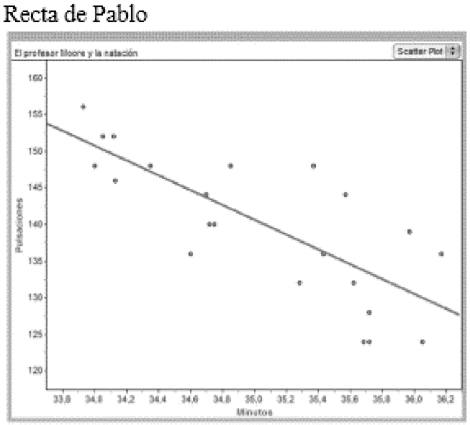
Recta de Susana



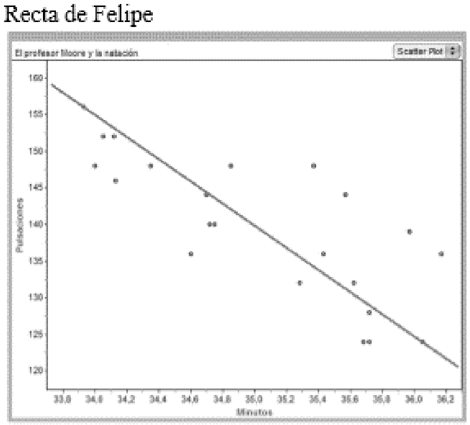
Recta de Julieta



Recta de Pablo

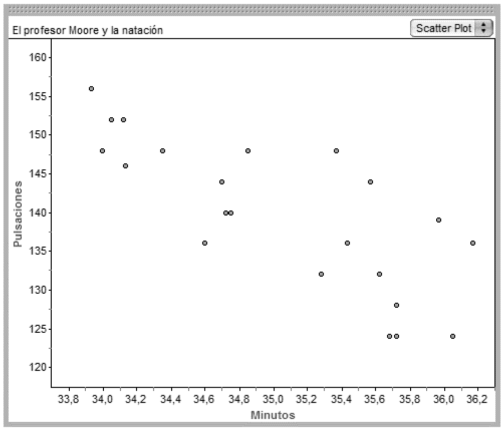


Recta de Felipe



9. Respecto a lo solicitado en la tarea, tú estarías de acuerdo con la recta trazada por [Justifica tu respuesta]

a. Susana
b. Julieta
c. Pablo
d. Felipe
e. Otra opción (traza tu recta y justifica)



Para el ítem 9, 11 (29%) de 38 estudiantes estuvieron de acuerdo con Susana (a) que, entre las opciones dadas, correspondía a la recta que más puntos tocaba. Entre los argumentos están: "*Porque tuvo más en cuenta los puntos sobre la gráfica*" o "*Susana, porque es la que mejor se adapta a el gráfico y utiliza mayor cantidad de puntos en el gráfico que se le dio*".

4 (11%) estudiantes escogieron la recta dada por Julieta (b), la cual dejaba igual número de puntos a lado y lado de la recta.

4 (11%) estudiantes acertaron al escoger la recta de Pablo (c), que era la recta que estaba más cerca, de manera simultánea, a todos y cada uno de los puntos. Un estudiante cree que es esta porque es la única que no toca los puntos: *"Pablo porque no pasa por encima de los puntos"*. Solo un estudiante se expresa en términos de cercanía: "*Porque sus datos son los más cercanos a la recta*".

10 (26%) escogieron la opción d (Felipe), la cual se trazó a través del punto mínimo y máximo respecto a la variable respuesta y 9 (24%) estudiantes decidieron trazar su propia recta pero sin justificación alguna.

Ítem 10

El ítem 10, tomado y adaptado de Moritz (2004), se relaciona con un diagrama (Figura 10) en donde las variables Número de personas en el salón de clase y Nivel de ruido tenían una asociación positiva (r=0,89). En una tarea similar a la que hace Casey (2015), aquí se pretendió indagar sobre qué tanto influye el contexto y las creencias previas que tienen los estudiantes para dar respuesta a la pregunta, ya que en el gráfico se muestra una asociación entre las variables contraria a lo que realmente se esperaría.

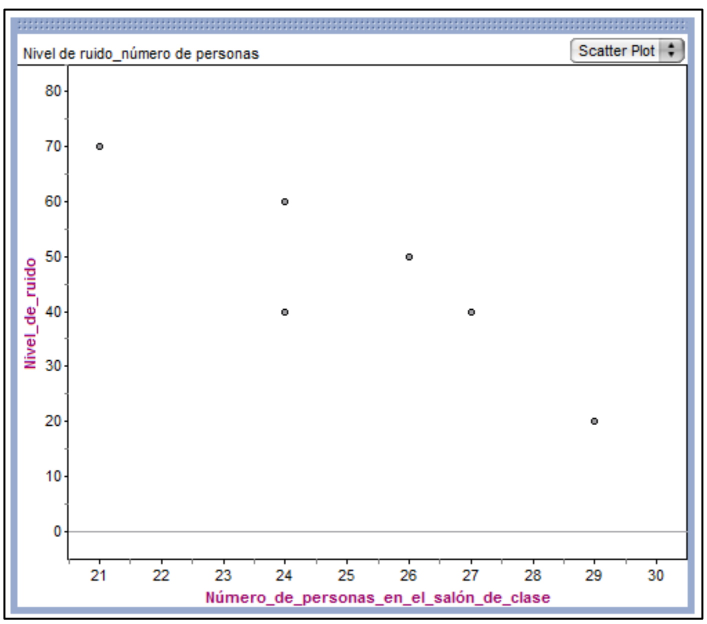


Figura 10. Número de personas en el salón de clase y nivel de ruido en la clase

Tabla 13.
Ítem 10 Prueba diagnóstica 1

10. Escoja la opción que mejor describe el gráfico anterior y justifique su elección.

a. Entre más personas haya en un salón de clases el ruido será mayor.
b. Entre más personas haya en un salón de clases el ruido será menor.
c. El gráfico muestra que el nivel del ruido no está relacionado con el número de personas en la clase.
d. El gráfico no proporciona información para establecer si existe una relación o no entre el número de personas en un salón de clases y el nivel de ruido.
e. Otra opción ¿cuál y por qué?

Para el ítem 10, el 32% (12) de los estudiantes escogieron la opción a, la cual no es cierta según lo que se evidencia en el gráfico, aunque sí es la más natural dentro del contexto de los estudiantes y eso se refleja en sus argumentos: *"Sí porque más niños hacen más ruido"* o *"Entre más personas el ruido será mayor porque cada persona hace mucho ruido y si hablan todos a la vez se hace mucho más ruido".* Aunque el 45% (17) de los estudiantes escogen la opción b sus argumentos se limitan al contexto y a sus creencias, buscando dar una explicación del hecho observado: *"Donde hay más alumnos el ruido será menor porque todos están más cerca y no tienen que alzar mucho la voz"*. Otros estudiantes sí basan sus argumentos en lo que la gráfica les informa, por ejemplo, se enfocan en un par de puntos: *"La b porque en la gráfica se ve que hay 21 personas que hacen un ruido de 70 y en otro salón de más personas hay 30 y el ruido es mucho menor por esto escogí la b"*. Solo un estudiante se expresa de manera global aunque sin dejar a un lado el contexto: *"Porque en el gráfico muestran que entre más personas menos ruido y entre menos personas más ruido porque algunos grados los estudiantes son más hiperactivos".*

Ítems 11 y 12

Los ítems 11 y 12 se relacionaban con un diagrama (Figura 11) en donde las variables IRM y CI tenían una baja asociación positiva (r= 0,33). Con el ítem 11 queríamos determinar cómo los estudiantes describen un gráfico en donde no se percibe una asociación fuerte entre las variables.

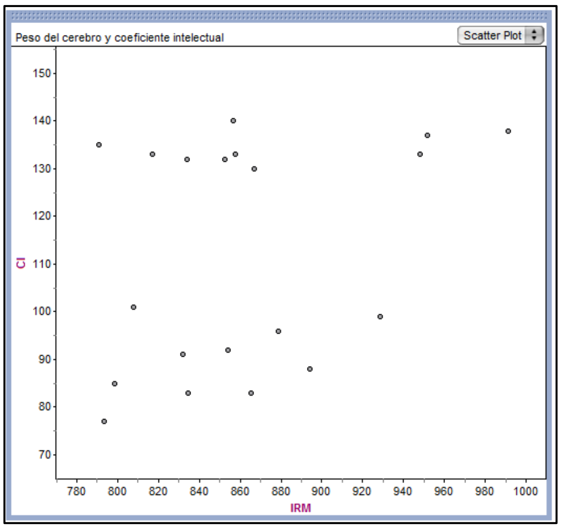


Figura 11. Peso del cerebro y coeficiente intelectual (CI) de 20 mujeres. El peso del cerebro se determinó mediante una imagen obtenida por resonancia magnética (IRM)

Tabla 14.
Ítem 11 Prueba diagnóstica 1

11. Respecto a la relación entre el peso del cerebro de una mujer y su coeficiente, se puede concluir que

a. las mujeres que tienen un cerebro con mayor peso tienen un coeficiente intelectual mayor.
b. las mujeres que tienen un cerebro con mayor peso tienen un coeficiente intelectual menor.
c. no existe una relación entre el peso del cerebro de una mujer y su coeficiente intelectual.
d. El gráfico no proporciona información para establecer si existe una relación o no.
e. Otra opción ¿cuál y por qué?

Respecto al ítem 11, el 29% (10) de los estudiantes escogieron la opción a y el 26% (9) la opción b, pero también aquí sus argumentos se basan en sus creencias, por ejemplo: *"Sí porque se le graban más las cosas"*, *"Pues creo que entre más grande debe ser más inteligente"* o *"Me parece que piensan más que los otros"*.

Algunos además de tener una concepción local, confunden la variable dependiente con la variable independiente: *"Porque en la gráfica se muestra que hay un peso del cerebro que es de 140 y tiene un coeficiente intelectual de 860 por eso estoy de acuerdo con la opción b"*. Otros incluyen otra variable como la edad para argumentar: *"Las mujeres con mayor edad tienen menos coeficiente intelectual"*.

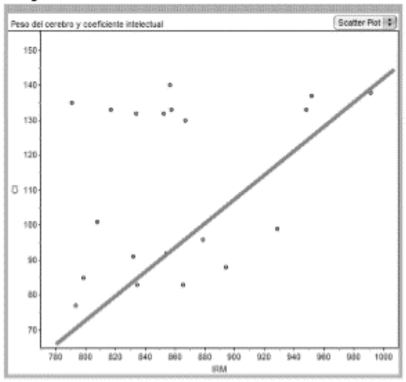
El 29% (10) escogió la opción c que era la correcta. Un argumento basado en sus creencias previas es, por ejemplo: *"Porque no existe la relación entre el peso y el coeficiente intelectual porque no importa el peso o el coeficiente intelectual, eso depende de la persona porque si pone cuidado entiende más"*. El siguiente argumento refleja un sentido de la variabilidad: "*Porque hay cerebros del mismo peso pero con diferente inteligencia"*.

Al igual que en la pregunta 9, con la pregunta 12 queríamos indagar sobre cómo nuestros estudiantes ubican la recta que mejor se ajusta, es decir, si está de acuerdo con algunos de los criterios encontrados por Casey (2015) o si propone algún otro criterio como se considera en la última opción de respuesta.

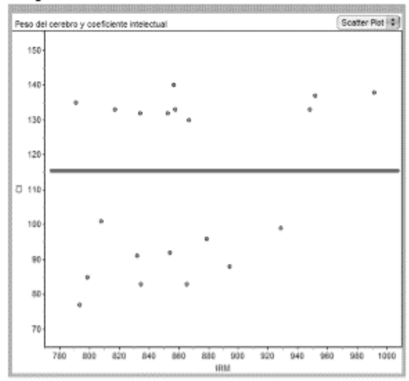
Tabla 15.
Ítem 12 Prueba diagnóstica 1

Se les pidió a cuatro estudiantes que trazaran una recta sobre el Gráfico E teniendo en cuenta que la recta se ajustara de la mejor manera a los datos del gráfico. Las siguientes gráficas muestran la respuesta de los cuatro estudiantes.

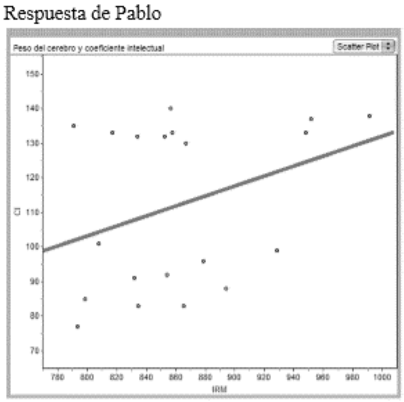
Respuesta de Susana



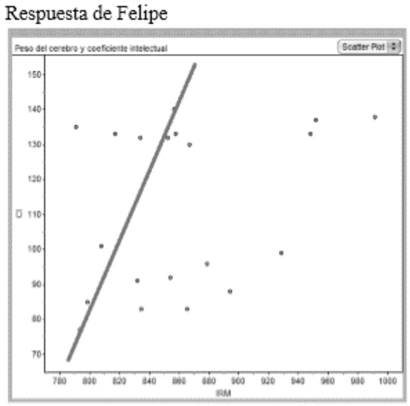
Respuesta de Julieta



Respuesta de Pablo

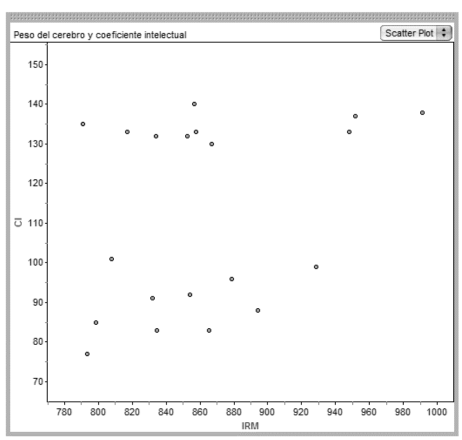


Respuesta de Felipe



12. Respecto a lo solicitado en la tarea, tú estarías de acuerdo con la recta trazada por [Justifica tu respuesta]

a. Susana
b. Julieta
c. Pablo
d. Felipe
e. Otra opción (traza tu recta y justifica)



Para el ítem 12, 11 (31%) de 36 estudiantes estuvieron de acuerdo con Susana (a) que, entre las opciones dadas, correspondía a la recta que más puntos tocaba.

11 (31%) estudiantes escogieron la recta dada por Julieta (b), que no está del todo incorrecta, la cual dejaba igual número de puntos a lado y lado de la recta. Aunque la recta se construyó con ese criterio, los argumentos de los estudiantes evidencias otros: "*Porque sus datos están lejos de la recta*" o "*La opción correcta es Julieta por separar las respuestas mayores de las inferiores".* Otros lo relacionan con que tiene que ser completamente horizontal, o con otro concepto como lo es la recta numérica: "Porque Julieta tiene su línea recta…" o "*Estoy de acuerdo con Julieta la línea está derecha y si esa es la imaginación de ella. Con esa se puede hacer una recta numérica*". Una argumentación muy interesante es la siguiente: "*La b porque la tarea de Julieta es la más cercana a la solución ya que no existe una relación*", ya que sabemos que la recta tenderá a ser horizontal cuando el nivel de asociación es muy cercano a cero.

Solo 2 (6%) estudiantes acertaron al escoger la recta de Pablo (c), que era la recta que estaba más cerca, de manera simultánea, a todos y cada uno de los puntos. Solo hubo una argumentación: "Porque está más cerca".

7 (19%) escogieron la opción d (Felipe), la cual se trazó a través del punto mínimo y máximo respecto a la variable respuesta, sin embargo surgieron otros argumentos como: *"Porque me está indicando que entre menos peso más coeficiente"* o "*Porque las mujeres con el CI más alto tienen menor IRM como lo muestra Felipe",* estos dos últimos argumentos reflejan lo que muestra la recta, así no se corresponda con los puntos, lo que da indicios a

pensar que sí entienden cómo sería la tendencia de los datos si existiera la relación que ellos dan.

Finalmente, 5 (14%) estudiantes decidieron trazar su propia recta con criterios como que no toque ningún punto, por ejemplo. Otra estrategia fue unir el valor mayor del eje vertical con el menor del eje horizontal argumentando *"Ninguna porque si la respuesta es que las mujeres tienen el cerebro más pesado menor es su coeficiente y ellos muestras es otras cosas. Yo la trazo así porque si dice la información de la gráfica que si tiene el cerebro más pesado menos es su coeficiente".*

4.2 Análisis de la prueba diagnóstica 2 Predicción y tendencia

La prueba que se analiza a continuación, fue aplicada a 40 estudiantes de octavo grado quienes no tenían una formación previa en conceptos estadísticos ni en diagramas de dispersión y recta de mejor ajuste, inclusive, sus antecedentes sobre el concepto como tal de recta era nulo, en términos formales. Esta prueba surgió del análisis de una primera prueba diagnóstica, de la cual resultaron algunas hipótesis, constaba de cinco ítems y se describe a continuación.

Descripción de la prueba

De la primera prueba diagnóstica surgieron algunas hipótesis que se irán mencionando junto con la descripción de cada ítem de esta segunda prueba diagnóstica.

1. En el momento de interpretar diagramas de dispersión, específicamente respecto a si los estudiantes establecían o no una relación o asociación entre las dos variables, se observó que los estudiantes, al parecer, pueden establecer dicha relación sin la necesidad de observar el gráfico, es decir, el contexto bajo el cual estaban las variables era muy cercano a ellos. De manera que, tomaremos como primera hipótesis el hecho de que *usar variables sin contexto, o por lo menos un contexto no cercano o no comprensible para los estudiantes los obliga a usar un razonamiento "puro" en sus interpretaciones.*

Por lo anterior, en el ítem 1 de la siguiente prueba las variables estaban descontextualizadas, la variable independiente se correspondía con X y la variable dependiente con Y. Queríamos conocer si los estudiantes establecían, o no, una relación entre las variables X e Y. Lo anterior ubicando un punto en el diagrama de dispersión dado para un valor específico en X (encerrados en rojo en la Figura 12), esperábamos que el estudiante identificara el "comportamiento" de los puntos en el diagrama y como consecuencia ubicara el punto que se le pedía para el correspondiente valor de X siguiendo ese comportamiento previamente identificado.

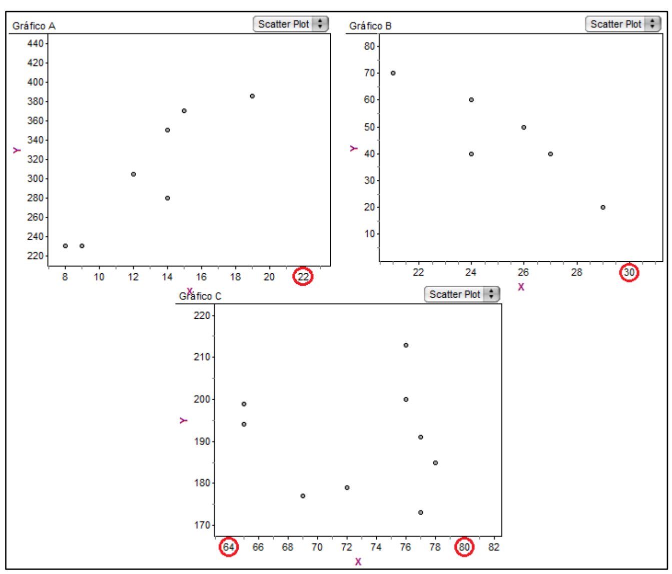


Figura 12. Diagramas de dispersión relacionados con el ítem 1 de la segunda prueba diagnóstica

Como se observa en la Figura 12, el primer ítem constaba de tres diagramas, cada uno con diferente sentido de asociación: el Gráfico A con una asociación positiva, el Gráfico B con una asociación negativa y el Gráfico C sin asociación.

Para el Gráfico A, donde se esperaba que ubicaran el punto arriba, 27 de los 40 estudiantes lo hicieron. Argumentos como *"Porque conforme asciende el valor de x el valor de y también"* o *"El valor de y que coloqué es de 440 porque creo que todo tiene un orden, entre más grande el número en x más grande el número y",* evidencian que reconocen claramente la existencia de una asociación entre las dos variables, mientras que otros no necesariamente, por ejemplo: *"Porque me guié por el más alto para colocarlo más alto que ese"* o *"Yo pienso que debe ir allí porque hace conexión con el otro punto que está a su misma altura".* Respecto al Gráfico B, 29 de 40 estudiantes ubicaron el punto en la parte de abajo, lo cual es correcto, con algunos argumentos muy similares a los presentados en el gráfico A, esta vez usando términos como "descendente", "disminuye", "de mayor a menor", es decir, pueden reconocer la asociación entre dos variables cuando la hay pero ¿qué pasa cuando no existe tal asociación como en el Gráfico C?

Para el Gráfico C solo 2, de 40 estudiantes, explicitan en sus argumentos la no asociación entre las dos variables: *"Pues porque no veo que tenga un orden respectivo", "Lo mismo no está aumentando ni disminuyendo solo se puede colocar donde sea".* El esfuerzo de los demás radica en encontrar un patrón, una "forma" en la nube de puntos, y justificar de alguna manera el valor de *y* que asociaron para el *x* pedido, por ejemplo: *"La puse arriba porque creo que esta lleva una secuencia de u".* Otros agruparon puntos del gráfico, por ejemplo para *x=80 "Y=215 del 78 hasta el 80 va de manera ascendente"* y para *x=64 "Y=205 creo que hasta el 77 va de manera descendente".*

En general, 21 de 40 estudiantes ubicaron correctamente los puntos sobre el gráfico A y B, y solo 2 de esos 21, ubicaron y argumentaron correctamente los puntos sobre el gráfico C. Por lo que la mayoría de los estudiantes, sí reconocen de alguna manera la asociación entre dos variables y cómo varían conjuntamente al observar un diagrama de dispersión, mientras que no reconocen fácilmente la no existencia de una asociación, es decir, buscan sí o sí relacionar de alguna manera las dos variables.

2. Respecto a la fuerza de la asociación entre las dos variables, se estableció la siguiente hipótesis: *los estudiantes asumen la intensidad de asociación de las dos variables de acuerdo al engrosamiento de la nube de puntos.* Por lo anterior, se presenta el ítem 2 de la siguiente (Figura 13) prueba el cual constaba de cuatro diagramas de dispersión cada uno con diferente coeficiente de correlación. Los diagramas se generaron con ayuda de un applet que se puede encontrar en http://www.rossmanchance.com/applets/GuessCorrelation.html. Aquí se pedía ordenar de mayor a menor los diagramas de acuerdo a la intensidad de asociación entre las dos variables las cuales, como en el ítem 1, no estaban contextualizadas.

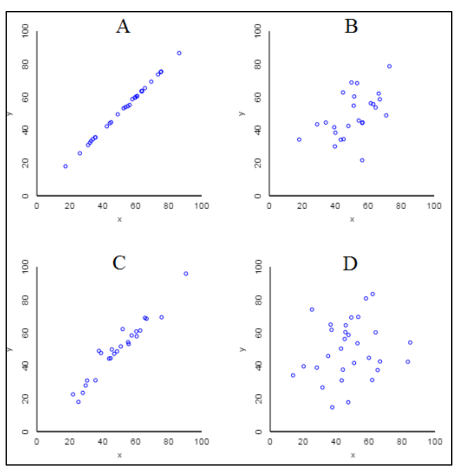


Figura 13. Diagramas de dispersión relacionados con el ítem 2 de la segunda prueba diagnóstica

Relacionando el "engrosamiento" de la nube de puntos con la intensidad o fuerza de asociación de las variables *x* e *y*, en el sentido en que a menor engrosamiento mayor es la intensidad, se puede concluir que, para los cuatro diagramas mostrados en la Figura 13, A tiene mayor intensidad, seguido de C, luego B y por último D. De los 40 estudiantes, 19 establecieron ese orden, ACBD. Entre los criterios, y con mayor frecuencia, para establecer el anterior orden está el de *dispersión o proximidad de los puntos*: *"Yo creo que el mayor es el que los punticos están más juntos y el menor pues el que tiene los puntos dispersos por todo el gráfico"*; el *ordenamiento* de los puntos, por ejemplo *"Porque el D y el B no tienen un orden respectivo mientras que el C ya se va formando y el A ya tiene un orden"*, el *nivel de rectitud: "Creo que porque la que tenga un nivel recto esta tiene un mayor nivel de intensidad";* como se observa en estos dos últimos criterios los estudiantes hablan implícita y explícitamente de recta, es decir, que la nube de puntos que se asemeje más a una recta es la que tiene mayor intensidad de asociación, lineal, entre las dos variables.

El segundo orden con más frecuencia fue DBCA, 8 estudiantes. Si nos fijamos, este orden es el "opuesto" al que observamos anteriormente, es decir, ese sería el orden correcto para establecer una intensidad de asociación de menor a mayor. El criterio más usado para esta elección coincide con uno ya mencionado: *dispersión o proximidad de los puntos* cuando argumentan "*Algunos están más amontonados y por eso me guié para dar esa respuesta",* también se evidencian los demás criterios pero con una interpretación opuesta a la que esperábamos: *"Porque D tiene más cantidad de puntos esparcidos, B porque tiene la 2 [segunda] más puntos esparcidos por el recuadro, C porque es el 2 [segundo] que tiene menos puntos esparcidos que son más unidos, A porque es más ordenado de todas y tiene una línea recta que conforman una recta.*

Otro criterio fue el *número de puntos* en el gráfico. De manera que el gráfico donde hubiera mayor cantidad de puntos, las variables tendrían una asociación más intensa: *"D porque tiene 30 puntos, B porque tiene 24 puntos, A porque tiene 23 puntos y C porque tiene 21 puntos".* En realidad los gráficos A y D tenían 30 puntos y B y C 25 puntos.

Independientemente del sentido de la interpretación, 27 de 40 estudiantes establecen un orden en cuanto al grado de dispersión o proximidad de los puntos. Entre esos, 4 estudiantes relacionan esa proximidad con asemejarse a una recta.

3. Otra hipótesis que establecemos es que, dado que nuestro interés es sembrar la necesidad de trazar la recta de mejor ajuste, *el hecho de observar "vecinos cercanos" para hacer una predicción puntual, lo cual manifiesta una concepción local, nos permite inferir que los estudiantes asumen de alguna manera la recta, por lo menos en esa localidad.* Lo anterior se refleja en acciones como observar el antes y el después del valor que se le pide o tomar la media de los "vecinos cercanos". Para ratificar lo anterior y explorar otras maneras de usar esta estrategia, se plantea el ítem 3 de la siguiente prueba (Figura 14).

Con el ítem 3, donde las variables sí estaban contextualizadas, queríamos ratificar las estrategias de los estudiantes para predecir el valor de la variable respuesta dado un valor para la variable explicativa (3a y 3b). Para el 3a, los "vecinos cercanos" son dos, uno a cada lado, mientras que para el 3b, los "vecinos cercanos" son tres, uno a un lado, y dos al otro lado.

Salón	Número de personas	Nivel de ruido
A	24	50
B	26	50
C	21	20
D	27	60
E	24	30
F	29	70

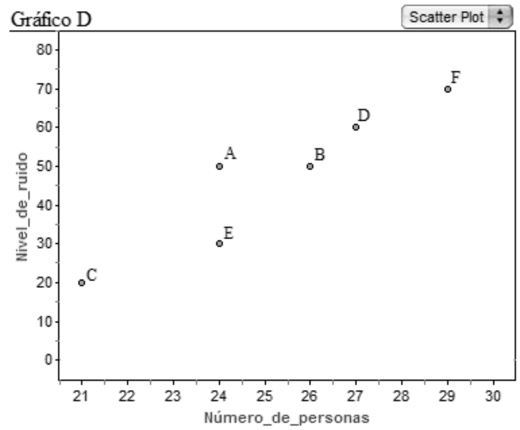


Figura 14. Tabla y diagrama de dispersión relacionado con el ítem 3 de la segunda prueba diagnóstica

En pocas palabras, se pedía predecir el nivel de ruido para un salón con 28 personas y para uno con 25 personas.

Para el de 28 personas, 30 de los 40 estudiantes eligieron un nivel de ruido entre 50 y 80 que podría considerarse como acertada. Dentro de esos 30, 8 estudiantes tomaron como estrategia la media de los “vecinos cercanos”, D y F, por ejemplo: *“El nivel de ruido es 65 el valor lo encontré mirando el D y F y en la mitad de D y F encontré el valor de 65”* o de otros vecinos: *“60 porque 28 está después de 26 y 26 es de 50 y está antes de 29 es de 70 entonces la mitad de 70 y 50 es 60”.* Otros aunque no usaron la media de los vecinos cercanos, sí se guiaron por los mismos para emitir un valor: *“Nivel de ruido sería 61 ya que me guie por el que tenía 27 personas pero porque tenía uno más le subió más la altura”* o *“60 Porque el salón D se acerca a los 28 estudiantes”.*

Un estudiante puso dos puntos, entre 50 y 65, manifestando un sentido de la variabilidad: *“Pues entre más personas hay más ruido, pero puede que hayan muchas personas y no tanto ruido, varía la respuesta”.*

Para el de 25 personas se tenían tres vecinos con ciertas características, es decir, 24 tenía dos imágenes y, a su vez, una imagen de 24 coincidía con la imagen de 26 (Figura 14). El criterio de mayor frecuencia para predecir el nivel de ruido, donde 32 estudiantes lo estimaron entre 31 y 55, fue también el de observar los vecinos cercanos, sin embargo muchos ignoraron el hecho de que 24 tenía dos imágenes, solo tomaron una de ellas para emitir un valor para 25, o un solo vecino: *"Me guie por el que tenía 24 personas y le puse como ruido 31"* o *"El de 25 personas es de 50 porque el de 24 y 26 también lo son";* implícitamente lo ubicaron en la media de los vecinos cercanos.

4. Siguiendo en línea con la hipótesis 3, nuestra cuarta hipótesis es que *la extensión de una concepción local podría conducirnos, en un principio, al ajuste de una recta poligonal, o ajuste parcial, que bien podría "enderezarse" y convertirse en la recta que queremos encontrar, la que mejor se ajusta.*

Por lo anterior, en el ítem 4 (Figura 15), donde las variables también estaban contextualizadas, partimos preguntando sobre la recta que mejor se ajusta a dos puntos (4a) ¿será "trivial" para los estudiantes que esta recta corresponde a la recta que pasa por los dos puntos? Enseguida queremos indagar sobre cómo trazan la recta que mejor se ajusta para tres puntos no colineales (4b y 4c) y finalmente se les presenta todo el conjunto de puntos y se pide la misma tarea, trazar la recta que mejor se ajusta. En últimas lo que queremos conocer con este ítem es qué entienden por "la recta que mejor se ajusta".

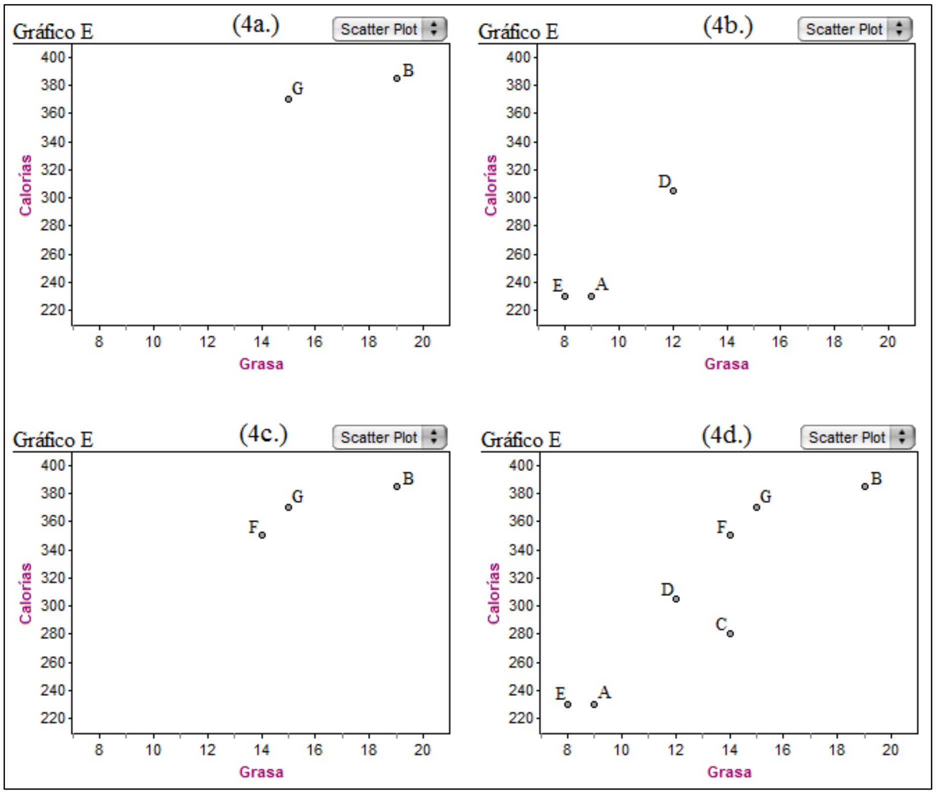


Figura 15. Diagramas de dispersión relacionados con el ítem 4 de la segunda prueba diagnóstica

Entre las estrategias para responder 4a la más utilizada fue trazar el segmento que unía los dos puntos, algunos trazaron dos segmentos, uno partiendo del origen y luego el que unía los dos puntos del gráfico, otros trazaron líneas verticales y horizontales hasta cada punto. En total, 7 estudiantes trazaron una recta, donde 3 de ellos trazaron la recta que unía los dos puntos con argumentos como *"… se ajusta porque solo hay dos puntos y es la única recta que se puede trazar".* Otra estrategia fue trazar la recta de manera que pasara por un punto pero muy cerca al otro. 3 estudiantes trazaron la recta entre los dos puntos y partiendo del origen, argumentando de la siguiente manera: *"la pasé sin que tocara G o B".*

Para 4b, 4c y 4d la estrategia con mayor frecuencia fue la de unir los puntos con segmentos. De los 7 estudiantes que trazaron una recta en el inciso 4a, 2 trazaron varias rectas uniendo dos puntos, mientras que los otros 5 trazaron una sola recta. Con estos 5 estudiantes se identifican algunas de las estrategias reportadas ya en la literatura (Casey, 2015) por ejemplo: la que toque la mayor cantidad de puntos; la que pasa por el "medio" dejando igual número de puntos a cada lado de la recta: *"porque así reparto mejor los puntos".* Respecto al inciso 4e, donde se les preguntaba que por qué creían que la recta trazada en 4d era la que mejor se ajustaba, solo 2 estudiantes argumentaron en términos de "cercanía": *"porque los puntos quedan más cerca a la línea"* o *"porque es como la que más se acerca a los datos".* En general, tal como se esperaba, la mayoría de los estudiantes trazaron "rectas poligonales" al unir los puntos por segmentos.

5. Un criterio muy usado es que la recta que mejor se ajusta es la que pasa por más puntos. Lo anterior se evidenció en una tarea propuesta por Casey (2015) en la que se daba un diagrama de puntos y los estudiantes ubicaban la recta, representada por un alambre, que mejor se ajustaba para ellos. Con el ítem 5 (Figura 16) de esta prueba nos preguntamos qué pasa con "el regreso", es decir suponemos una recta que se ajusta a seis puntos e indagamos sobre cómo deberían ir los puntos. La hipótesis 5 que establecemos acá es que el criterio mencionado anteriormente no va a cambiar, aunque expresado de otra manera: *la nube de puntos que se corresponde con la recta que mejor se ajusta, es la que más puntos toca la recta*.

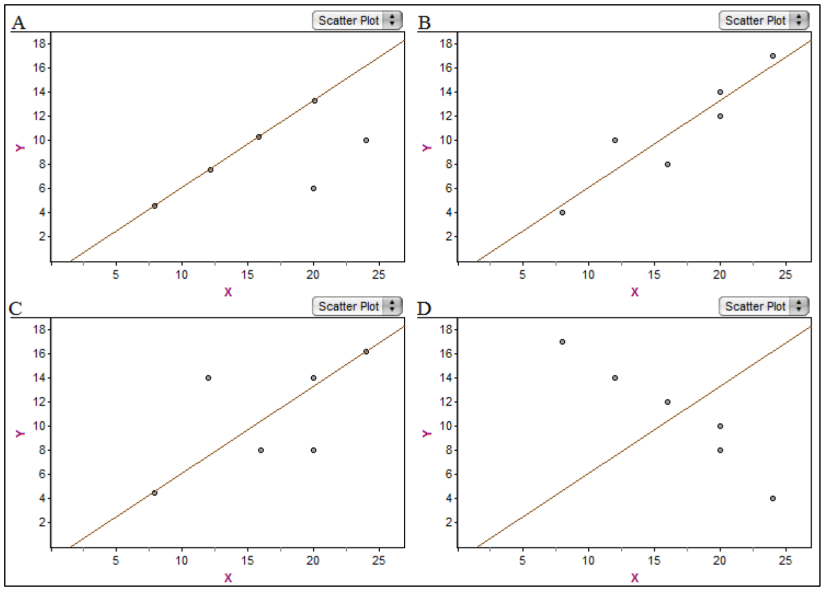


Figura 16. Diagramas de dispersión relacionados con el ítem 5 de la segunda prueba diagnóstica

De manera que en el ítem 5 se mostraban cuatro diagramas con la misma recta pero seis puntos distribuidos de manera diferente en cada gráfico (Figura 16). Se pedía ordenar, de mayor a menor, los gráficos de acuerdo al nivel de ajuste de la recta a los seis puntos. Para elaborar este ítem usamos la herramiena Fathom que nos permite ver los cuadrados y la suma de las áreas de los mismos que se forman de cada punto a la recta (Figura 17); sabemos que si esta suma es mínima, tendremos, en este caso, la nube de puntos para la cual la recta dada es la que mejor se ajusta.

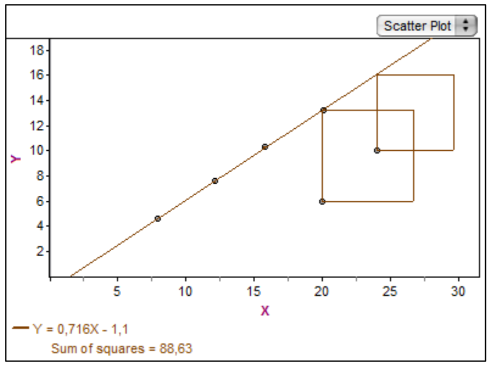


Figura 17. Suma de cuadrados realizada con Fathom para la opción A del ítem 5

Así, la suma de las áreas de los cuadrados para las opciones A, B, C y D eran, aproximadamente, 88, 15, 75 y 381, respectivamente, por lo que la respuesta correcta era BCAD, solo un estudiante respondió correctamente, y su estrategia fue unir los puntos por segmentos y observando que: *"BCAD porque el B y C son los que están más cerrados mientras que A y D están abiertos".*

ACBD fue la respuesta con mayor frecuencia, 19 estudiantes, que si observamos la Figura 16 ese sería el orden en cuanto al número de puntos que toca la recta o que están muy próximos a tocarla. Sin embargo de esos 19, solo 11 hablaron en esos términos y 8 hablaron en términos de "alineamiento" u "organización": *"Porque desde A los puntos están más alineados y desde C B y D empiezan a desalinearse".*

11 estudiantes que aunque no eligieron la respuesta correcta, basaron sus argumentos en términos de "cercanía" pero al parecer entendido como el número de puntos más cercanos a la recta, por ejemplo: *"Empezamos con los seis puntos cerca a la línea luego se alejan dos, luego se alejan 3 y luego se alejan 4 quedando B, A, C, D".* Otro argumentos no dejan muy en claro qué es lo que ellos entienden por cercanía, por ejemplo: *"Yo escogí ese orden [ACBD] porque miré que los puntos de A están más cerca a la recta que los puntos de el D".*

En general, con este ítem se ratifica nuestra hipótesis 5 pero además se abre la reflexión sobre qué entienden o cómo conciben los estudiantes la “cercanía” de los puntos a una recta, ya que fue uno de los argumentos con mayor frecuencia, pero estableciéndose un orden diferente.

4.2.1 Sobre el análisis de las dos pruebas diagnósticas. El análisis de los resultados de estas pruebas nos proporcionó información sobre el Razonamiento Covariacional Informal (RCI) actual de los estudiantes. Se ratificaron algunas estrategias y concepciones ya reportadas en investigaciones anteriores y algunas novedades que nos orientaron para establecer más hipótesis de aprendizaje y, consecuentemente, diseñar el plan para el diseño de actividades computacionales para el aprendizaje de la recta informal de mejor ajuste:

- Construir actividades sin contexto, o por lo menos no comprensible al estudiante, podría evitar que el estudiante tenga una *concepción causal*. En su lugar, se obligaría a tener un razonamiento “puro” en tareas de asociación y predicción.
- La mayoría de los estudiantes, sí reconocen de alguna manera la asociación entre dos variables y cómo varían conjuntamente al observar un diagrama de dispersión, mientras que no reconocen fácilmente la no existencia de una asociación, es decir, buscan sí o sí relacionar de alguna manera las dos variables.
- La mayoría de los estudiantes sí establecen un nivel de asociación en cuanto al grado de dispersión o proximidad de los puntos pero pocos, además de eso, en cuanto a la forma.
- La estrategia de los “vecinos cercanos”, manifiesta una *concepción local*, que si se extiende conduce al ajuste de una recta poligonal que bien podría “enderezarse” y convertirse en la recta de mejor ajuste global.

4.3 Actividad 0. Obteniendo datos del curso al que pertenezco

Con el fin de que los estudiantes se familiarizaran y se motivaran con el manejo del software Fathom, la Actividad 0 consistió en generar la necesidad de tomar datos, datos reales y cercanos a los estudiantes como, por ejemplo: edad, género, peso y estatura de sus compañeros de curso. Lo anterior motivados por preguntas como ¿Cuántos niños y cuántas niñas hay en el curso de octavo grado? ¿Qué edades tienen los estudiantes? ¿Qué peso y estatura? ¿Son más altos los niños o las niñas?

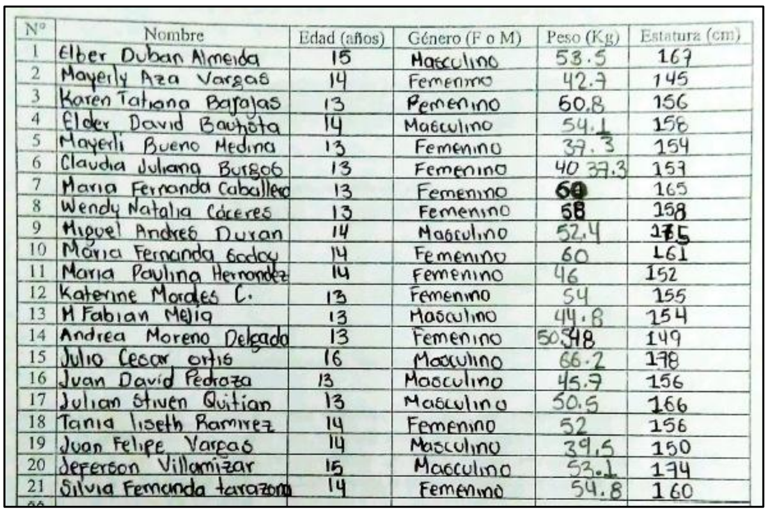
Obtenidos los datos, con el ánimo de facilitar la respuesta a las anteriores preguntas, se procedió a registrarlos en una tabla en Fathom así como visualizarlos en gráficos.

Los datos fueron tomados en grupos con el propósito de que para los atributos “peso” y “estatura”, cada grupo obtuviera mediciones diferentes. Lo anterior dado que cada grupo contó con sus propios instrumentos de medición (báscula y metro), como se muestra en la Figura 18. Esto nos permitiría más adelante generar una primera discusión alrededor de la variabilidad y sobre la decisión de quedarnos con una sola medida, un solo valor.

Figura 18. Estudiantes tomando datos para la Actividad 0

Análisis

El desarrollo de esta primera actividad giró alrededor de una tabla como la que se muestra en la Figura 19, producto de la experiencia que ellos vivieron recolectando y registrando datos de cada uno de sus compañeros de curso. Hecho lo anterior se generó una primera discusión sobre los datos en cuestiones como cuántos estudiantes aparecen en los datos; cuántas variables y cómo esos datos describen a su curso.



N°	Nombre	Edad (años)	Género (F o M)	Peso (Kg)	Estatura (cm)
1	Elber Duban Almeida	15	Masculino	53.5	167
2	Mayerly Aza Vargas	14	Femenino	42.7	145
3	Karen Tatiana Barajas	13	Femenino	50.8	156
4	Elder David Bautista	14	Masculino	54.1	158
5	Mayerli Bueno Medina	13	Femenino	37.3	154
6	Claudia Juliana Burgos	13	Femenino	~~40~~ 37.3	157
7	Maria Fernanda Caballero	13	Femenino	60	165
8	Wendy Natalia Cáceres	13	Femenino	58	158
9	Miguel Andres Duran	14	Masculino	52.4	175
10	Maria Fernanda Godoy	14	Femenino	60	161
11	Maria Paulina Hernandez	14	Femenino	46	152
12	Katerine Morales C.	13	Femenino	54	155
13	M Fabian Mejia	13	Masculino	44.8	154
14	Andrea Moreno Delgado	13	Femenino	50 48	149
15	Julio Cesar ortis	16	Masculino	66.2	178
16	Juan David Pedraza	13	Masculino	45.7	156
17	Julian Stiven Quitian	13	Masculino	50.5	166
18	Tania liseth Ramirez	14	Femenino	52	156
19	Juan Felipe Vargas	14	Masculino	39.5	150
20	Jeferson Villamizar	15	Masculino	53.1	174
21	Silvia Fernanda tarazona	14	Femenino	54.8	160

Figura 19. Datos registrados por uno de los estudiantes para la Actividad 0

Respecto a las cuestiones anteriores, los siete estudiantes estuvieron de acuerdo en cuanto al número de estudiantes y seis de ellos en que las variables que se midieron fueron cinco.

En cuanto a la descripción de los datos, es decir, qué información se podía extraer de la tabla que caracterizara el curso al que pertenecían, la mayoría de los estudiantes describieron, o bien entre qué valores se encontraba alguna variable, o cuál era el valor máximo que alcanzaba, por ejemplo: “Que en el grado octavo hay una cantidad de 21 estudiantes donde la mayoría son niñas, que la edad promedio es de 16 a 13 años, que el peso está entre 67 y 39 kg y la estatura ronda 182 cm y 145 cm” o “Que hay más niñas que niños, que la moda en edad es de 13 años, el peso de algunos estudiantes no sobrepasa de 62 kg”

Por otra parte, cinco de ellos hablaron implícita o explícitamente de la moda de alguna de las variables al usar expresiones como “hay más”, “la mayoría” o, directamente, “la moda”, por ejemplo: “Que hay más niñas que niños, que la moda en edad es de 13 años, el peso de algunos estudiantes no sobrepasa de 62 kg”. Cabe destacar que los estudiantes ya tenían algunos conocimientos previos sobre medidas de tendencia central.

Finalmente, solo dos estudiantes hablaron en términos de promedio, pero, según como lo argumentan, entendido como un intervalo y no como un solo valor: “…que la edad promedio es de 16 a 13 años…” o “Que los estudiantes del grado octavo están en un promedio de estatura de 1,45 m y 1,82m”.

Otro tema para abordar, que justificó esta actividad, es la variabilidad, la cual se hizo muy natural por el hecho de que ellos mismos recolectaron los datos. De manera que a la pregunta ¿Por qué tenemos diferentes mediciones, en peso y estatura, por ejemplo, para un mismo estudiante? Sus respuestas fueron inmediatas y muy válidas, por ejemplo: “En el peso porque no se paraban derechos, o los compañeros metían la mano en el peso y esto alteraba el peso y en la estatura porque tal vez se sumó mal” o “Tenemos diferentes medidas porque en algunos metros se pudo haber colocado recto y en otros encorvado. En el peso de pronto por las pilas o algunos mirando hacia el suelo”.

Ahora, una vez hecha la discusión sobre por qué un mismo estudiante había tenido diferentes mediciones, se tentó a los estudiantes a quedarse con un solo valor, uno para peso y otro para estatura. Por ejemplo, la Tabla 16 muestra las medidas que tuvo Julio, un compañero de ellos, en cada uno de los grupos (G) donde fue medido.

Tabla 16.
Medidas de un estudiante del curso en cada uno de los grupos para la Actividad 0

JULIO	**G1**	**G2**	**G3**
Peso (kg)	67	66,2	66
Estatura (cm)	182	178	181

Entre las estrategias para quedarse con un solo valor están las siguientes:

- *El promedio:* Tres estudiantes argumentaron en términos de promedio, "Me quedaría el peso 66 kg. Estatura 181 porque es el promedio entre los tres valores".
- *Un valor entre el mínimo y el máximo:* "Yo escogería el peso (kg) 67 por su forma física. Y la estatura 181 porque está en un promedio de 178 cm y 182 cm".
- *Creencias:* Algunos basaron sus argumentos en sus creencias de lo que pudo haber pasado en el momento de la medición: "Me quedaría con el peso del grupo N°1 porque creo que este grupo pudo medir la estatura de Julio mejor, tal vez porque era el primer grupo y los alumnos debían pasar primero por este. Y creo que mi grupo y el segundo grupo íbamos un poco apresurados porque los estudiantes no pasaban en orden y debíamos llamarlos".
- *El que más se repite:* Aunque se pedía por estatura y peso, un estudiante argumentó en términos de edad: "Con la edad de 13 años porque es la que más se repite y es de 13,0, es decir hay 10 niños de 13 años".

Para abordar un poco el tema de predicción, se propuso predecir la estatura para un nuevo estudiante, hombre, de 13 años que llega al curso. Todos estuvieron de acuerdo en que sí era posible predecir la estatura, aunque difirieron en sus estrategias de la siguiente manera:

- Cuatro estudiantes decidieron extraer de los datos a los niños hombres de 13 años, sin embargo, tres de ellos no especifican cómo usarían esos datos para hacer la predicción: “Sí, porque tendría en cuenta la estatura de mis otros compañeros (hombres) de 13 años para poder darle una estatura”. Solo uno de ellos toma el promedio de la estatura de esta extracción: “Total sí podría porque tomaría la estatura de los compañeros de 13 años y sacaría su promedio”.
- Los otros tres estudiantes usan como estrategia o bien medirlo o preguntarle directamente por su estatura o bien compararlo con alguno de sus compañeros del que ya conocen ese dato: “Le preguntaría o si hubiese posibilidad de medirlo lo haría, o lo compararía con alguno de mis compañeros, puede ser que él y mi compañero tengan la misma estatura y si se cuánto mide mi compañero tal vez pueda saber cuánto mide el estudiante nuevo dependiendo si es más alto o más bajo o tal vez tenga la misma estatura de mi compañero”.

Las anteriores estrategias fueron discutidas entre ellos y el hecho de que algunos se enfocaran solo en los estudiantes hombres de 13 años nos dio paso para visualizar, con ayuda de Fathom, esos casos específicos y, teniendo en cuenta eso, contestar nuevamente la situación de predicción.

Con lo anterior queríamos ver si después de la discusión y con esa nueva visualización ratificaban, cambiaban, o no, su estrategia de predicción. Dos de los tres estudiantes que tenían como estrategia medirlo o preguntarle directamente al estudiante, cambiaron su estrategia para tomar el promedio de las mediciones. Lo anterior posiblemente porque, en la discusión, el estudiante que propuso sacar el promedio refutaba la estrategia de los demás al decir que ellos no conocían al estudiante nuevo y que no era posible medirlo o compararlo directamente con otros.

Por otra parte, en el caso de los cuatro estudiantes que tomaron en cuenta a los hombres de 13 años, se evidenció mejor cómo usarían esos datos para hacer la predicción: Tres, sacando el promedio y uno, describiendo un intervalo de los valores que toma la estatura de los hombres de 13 años. De manera que, al final de esta actividad, cuatro de los siete estudiantes asumen la media como estrategia de predicción.

Como se observa, esta primera actividad fue más allá de una exploración de los estudiantes con el manejo de Fathom ya que a su vez nos permitió, a nosotros como investigadores, tener un primer acercamiento a las intuiciones, concepciones e ideas de los estudiantes en aspectos como la variabilidad y la predicción en el caso unidimensional. Creemos que el hecho de que los datos fueran reales y además tomados por ellos mismos, les permitió asumir de manera natural el tema de la variabilidad ya que fueron testigos de las razones por las que pudo ocurrir.

También nos permitió ver sus conocimientos previos en medidas de tendencia central en donde algunos asumen la media, o el promedio, como un intervalo de valores. En cuanto al tema de predicción, en primera instancia, solo un estudiante usó como estrategia la media; los demás basaron sus juicios en sus creencias o en sus “vecinos” más cercanos, en el sentido de que tomaban a otro de sus compañeros para compararlos y, así, hacer una predicción.

Por otra parte, vimos también cómo después de la discusión entre ellos sobre sus estrategias de predicción, algunos cambian o se sienten correspondidos o convencidos con la estrategia de alguno de sus compañeros, en donde finalmente cuatro concuerdan en que tomar la media es una buena estrategia.

4.4 Actividad 1 (Parte 1). El Machín Machón

Como se observó, en la Actividad 0, algunos estudiantes hablaron de “promedio” (o media, como se hace referencia en este texto). ¿Pero qué significa ese valor para ellos? ¿Cómo se calcula? ¿Por qué es el mejor representante en un conjunto de datos univariados? ¿Cómo está ubicado respecto a los demás valores? ¿Qué propiedades cumple? Esta actividad pretende abordar las cuestiones anteriores.

El objetivo es esta actividad fue que los estudiantes establecieran un criterio para, dado un conjunto de datos univariados, encontrar el mejor representante que finalmente será la media; la media como un valor justo, como veremos en un primer momento llamado “Repartiendo caramelos”, o un punto de “equilibrio” o “balanceador”, como se apreciará en el segundo momento llamado “Balanceando el Machín Machón”.

Análisis

Situación A: Repartiendo Caramelos

Como ya se dijo, en primera instancia quisimos ver la media como un valor justo, en un sentido equitativo. Es por lo anterior, que la primera situación de esta actividad se llamó “Repartiendo caramelos”. Treinta y cinco caramelos fueron distribuidos entre los siete estudiantes como se muestra en la Tabla 17.

Tabla 17.
Tabla asociada a la Situación A de la Actividad 1.1

Nombre	# de caramelos
Andrea	6

Miguel	4
Wendy	7
Silvia	6
Dubán	7
Elder	3
María Fernanda	2

Cuando se indagó sobre si consideraban, o no, justa la manera en que se habían distribuido los caramelos, naturalmente, todos la consideraron injusta argumentando ya fuera en términos de comparación como: "Es injusta porque unos tienen menos caramelos que otros" o dando indicios de cómo hubiera sido justa: "Es injusta porque a todos no les dieron la misma cantidad de caramelos".

El hecho de que todos estuvieran de acuerdo en lo injusto de la distribución de caramelos nos dio paso a que dieran o recibieran caramelos de sus compañeros hasta que consideraran que cada uno de ellos tenía la cantidad justa. Por ejemplo, Wendy, quién tenía siete caramelos en un principio, procedió de la siguiente manera: "Le di a María Fernanda Godoy dos caramelos porque [ella] tenía tres caramelos". Lo que da a entender que Wendy buscaba que tanto ella como María Fernanda tuvieran cinco caramelos. De manera similar procedieron los demás hasta lograr la cantidad justa para todos y cada uno de ellos, tal como lo argumenta un estudiante: "Ahora todos los compañeros tienen cantidad igual porque dividimos 35 caramelos entre 7 y así todos quedaron con 5 caramelos que es la cantidad justa por estudiante".

Como se observa en esta situación, el proceso matemático de sumar la cantidad de caramelos que le correspondió a cada estudiante y dividirlo entre la cantidad de estudiantes

$$\frac{6+4+7+6+7+3+2}{7}=5$$

corresponde al cálculo de la media, solo que vista como un valor justo y además en un contexto muy familiar para ellos, tan natural como cuando promedian la nota definitiva de una materia, por lo menos cuando cada nota tiene el mismo peso. Es como si el cálculo estuviera implícito en ellos, pero hasta el momento no sabemos si sus propiedades también; es por esto que se plantea la siguiente situación: Balanceando el Machín Machón.

Situación B: Balanceando el Machín Machón

Para esta situación, en primera instancia se hizo una discusión sobre la dinámica del juego Machín Machón. Aquí, todos manifestaron que ya en alguna ocasión lo habían jugado. De manera que se inició con un ejercicio de lápiz y papel, planteado de la siguiente manera:

Dos niños, Édinson y Gabriel, del mismo peso, se encuentran sentados a una distancia x_1, x_2 metros, respectivamente, del extremo izquierdo de un Machín Machón en el cual desean jugar. El Machín Machón tiene una longitud de 10 metros, bastante grande.

El primer caso que se propuso fue muy natural para ellos, $x_1 = 0$ y $x_2 = 10$, es decir, cada uno de los niños ubicados en un extremo del Machín Machón, como normalmente se acostumbra jugar. De manera que la respuesta a ¿Dónde ubicarías el pivote para que el Machín Machón quede balanceado? fue unánime, "a cinco metros", es decir, en la mitad del Machín Machón. Argumentando que "… porque es la mitad de 10 metros" o tomando la mitad del extremo derecho: "La ubicaría en la mitad porque si Edinson está ubicado en el extremo x_1 que es igual a 0 y Gabriel está ubicado en el extremo x_2 es igual a 10 entonces la mitad de 10 es 5 entonces el pivote lo ubicaría a 5 metros". ¿Pero dónde ubican el pivote en los demás casos es que los dos niños, no necesariamente, ocupan los extremos del Machín Machón? Y, en general, ¿dónde ubican el pivote para cualquier valor de x_1 y x_2?

Pues bien, tal como se esperaba, en los demás casos las estrategias, y por tanto las respuestas, variaron. A continuación presentamos las tres estrategias a las que los estudiantes llegaron y que fueron discutidas para llegar a una expresión algebraica:

• *En la mitad de los dos niños:* Tres estudiantes propusieron esta estrategia. Por ejemplo, en el caso en que $x_1 = 3$ y $x_2 = 9$, "A seis metros porque Gabriel está a seis metros de Édinson porque 9-3=6 / 2 = 3 metros más 3 = 6 mt". Es decir, se toma la semidistancia entre los dos niños y al resultado se le suma x_1. Tal como lo generaliza un estudiante: "Se restan los dos valores y el resultado se divide en dos y se suma la distancia de x_1" y lo expresa algebraicamente como:

$$\frac{\{x_2 - x_1\}}{2} + x_1$$

Después de que la estudiante simplificara frente a sus compañeros la expresión anterior, el resultado fue el siguiente:

$$\frac{x_1 + x_2}{2}$$

Dos de los tres estudiantes, aunque no mostraron una expresión algebraica, se limitaron a decir que, en general, ubicarían el pivote en el "centro" de los dos niños.

• *El extremo derecho entre el extremo izquierdo:* Tres estudiantes optaron por esta estrategia, es decir, dividir x_2, puesto que x_2 siempre estuvo a la derecha de x_1, entre x_1. Algebraicamente,

$$\frac{x_2}{x_1}$$

Uno de ellos argumentó de la siguiente manera: Dividir el resultado de la derecha entre la izquierda y el valor dado es la ubicación para cada pivote.

- *El extremo derecho entre el extremo izquierdo, más el extremo izquierdo.* Un estudiante argumentó así: "Yo tendría en cuenta siempre una división porque así podría encontrar la distancia para que el Machín Machón se pueda mantener balanceado. La división sería del punto x_2 entre x_1 por ejemplo $x_1 = 4$ y $x_2 = 12$, entonces la distancia del Machín Machón sería 3". Lo que nos podría llevar a pensar en que consiste en la estrategia inmediatamente anterior: "El extremos derecho entre el extremo izquierdo", pero, tal como ubica el pivote en el caso en que $x_1 = 2$ y $x_2 = 8$, se observa que difiere en el sentido de que el resultado de ese cociente lo cuenta a partir de la ubicación del niño de la izquierda, como se muestra en la Figura 20. Ubicación del pivote de uno de los estudiantes cuando se tienen dos niños.

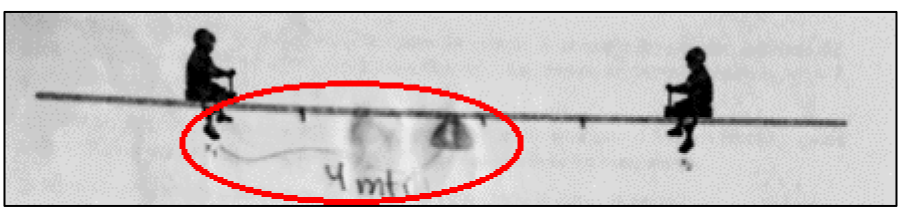


Figura 20. Ubicación del pivote de uno de los estudiantes cuando se tienen dos niños

De manera que una vez discutido su estrategia, se llegó a la siguiente expresión algebraica,

$$\frac{x_2}{x_1} + x_1$$

En la Tabla 18, se resumen las tres expresiones algebraicas que proponen los estudiantes para cualquier valor de x_1 y x_2.

Tabla 18.
Resumen de las expresiones algebraicas que proponen los estudiantes para la ubicación del pivote dado dos valores

$\frac{x_1 + x_2}{2}$	$\frac{x_2}{x_1}$	$\frac{x_2}{x_1} + x_1$

Simulando el Machín Machón con Geogebra

Como se observó, los estudiantes llegaron a tres conjeturas para la ubicación del pivote en donde se balancea el Machín Machón, de manera que lo que seguía era verificar la que cada uno de ellos había propuesto. Para lo anterior contaban con un archivo[1] de Geogebra (ver Figura 21) que podían manipular sin necesidad de estar familiarizados con este software, de hecho ninguno de ellos lo estaba.

[1] El archivo de Geogebra puede ser descargado en: https://www.dropbox.com/s/yiw60l39axf9doh/Mach%C3%ADn%20Mach%C3%B3n.ggb?dl=0

Figura 21. Visualización Actividad Machín Machón en Geogebra

Para cada uno de los casos previos realizados en lápiz y papel, debían verificarlo con la simulación. Como era de esperarse, en el caso a, donde $x_1 = 0$ y $x_2 = 10$, todos lo lograron, es decir, movieron el pivote hasta 5,0 y, dado que efectivamente en esa posición el Machín Machón se balancearía, aparecía la expresión ¡Eureka! como se muestra en la Figura 22.

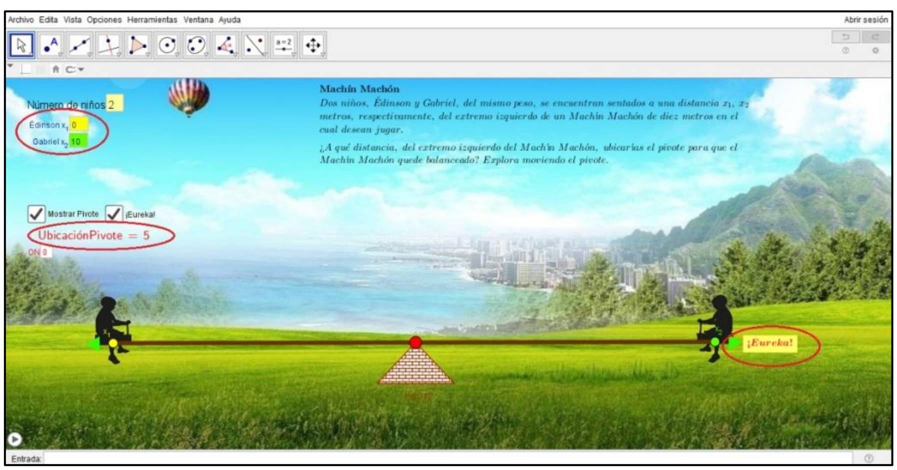


Figura 22. Visualización Actividad Machín Machón en Geogebra (UbicaciónPivote y Eureka)

Lo anterior se repitió en cada uno de los casos previos y haciendo uso de su expresión algebraica conjeturada. Ya en la discusión se había acordado que si su conjetura no funcionaba intentara con las de sus compañeros o planteara una nueva hasta encontrar la forma general de ubicar el pivote para cualquier valor de x_1 y x_2, como se muestra en la siguiente discusión.

I: Vamos a poner a x_2 en 0 y a x_1 vamos a ponerlo en 4… Entonces Miguel, dónde cree usted que va el pivote.

M: Dos metros.

I: ¿En dos? … Silvia…

S: En tres

I: Elder…

E: En dos

I: Elder en dos y Elber…

D: En tres.

Después cada quien ubicó el pivote donde creía que se nivelaría el Machín Machón.

I: ¿Dónde se niveló?

A: En dos

I: ¿De qué nos estamos dando cuenta?

A: De que teníamos que sumarlos y dividirlos en dos.

I: Sumarlo y dividirlo en dos… Andrea dice que sumarlo y dividirlo en dos, Mafer. Mafer, deme un número de cero a diez.

M: Ocho

I: Buen entonces vamos a ubicar x_1 en ocho. Silvia deme un número de cero a diez…

S: Seis

I: Seis, x_1 en 8 y x_2 en 6. Entonces según lo que dice Andrea, Mafer, dónde iría el pivote…

I: Mafer, dónde iría el pivote… Andrea que es lo que usted dice…

A: Sumar y dividirlo en dos.

I: Sumar qué.

A: Los dos valores… o sea 8 + 6…

E: En siete

I: Entonces en siete, muévalo hasta siete. ¿Sí se balancea?

W: Sí

Así mismo se verificó en otros casos y todos estuvieron de acuerdo en que la forma general para encontrar la ubicación del pivote correspondía a la expresión

$$\frac{x_1 + x_2}{2}$$

Quedando así descartadas las otras dos expresiones algebraicas propuestas.

Descubriendo una propiedad de la media.

Es claro que hasta el momento ninguno de los estudiantes, dio cuenta de que ese cálculo correspondía al cálculo de la media entre x_1 y x_2, tampoco se habló en términos de la misma puesto que más adelante queríamos ver cómo encontraban la ubicación del pivote en el caso de tres niños o más. Acaso ¿sumando los tres y dividiéndolo en tres? No obstante, queríamos indagar, o tener un primer acercamiento a una de las propiedades que tiene la media: “La suma de las desviaciones sobre la media es cero” o “la suma de las distancias a la derecha sobre la media, es igual a la suma de las distancias a la izquierda sobre la media”.

Para lo anterior, en el simulador del Machín Machón, una vez que se encontraba balanceado, es decir, cuando aparecía a la expresión ¡Eureka!, se activaba, a su vez, una casilla nombrada “ON” que, por defecto, tenía el valor de 0. Con los estudiantes se discutió sobre lo que ocurría en la simulación teniendo balanceado el Machín Machón para el caso c, $x_1 = 3$ y y $x_2 = 9$.

I: Entonces Wendy, dónde iría el pivote.

W: En seis.

I: ¿Están todos de acuerdo?

Todos: Sí

I: ¿Si ven que cuando ustedes lo balancean aparece una palabra que dice “ON”?

Todos: Sí

I: ¿Y dice cero?

Todos: Sí

I: Entonces vamos a cambiar cero por uno. ¿Qué salió? Elder, qué salió cuando cambió cero por uno.

E: Pues tres y tres y las flechitas

Lo anterior se muestra en la Figura 23.

Figura 23. Visualización Actividad Machín Machón en Geogebra asociado a una conversación sobre las distancias del pivote a dos niños

I: Qué creen que sean esas flechitas.

A: La distancia que hay desde el niño al pivote.

I: La distancia que hay del niño al pivote ¿cierto? ¿y cómo son esas distancias?

M: Iguales.

Lo anterior se verificó con otros valores para x_1 y x_2.

Hasta el momento, de manera intuitiva e implícita, vimos cómo se calcula la media para dos valores y qué propiedad cumple. Nuestro objetivo final era extender lo anterior, para el caso de tres niños o más, por lo que el objetivo siguiente era encontrar una forma general de balancear el Machín Machón en donde se encuentran jugando tres niños, como se muestra en la Figura 24. En primera instancia tenían que x_1 =1, x_2 =3 y $x_3 = 8$ y debían encontrar la ubicación del pivote antes de usar el simulador.

Figura 24. Visualización Actividad Machín Machón en Geogebra con tres niños

Cinco estudiantes acertaron en que la posición del pivote era a cuatro metros del extremo izquierdo, sin embargo, difirieron en sus estrategias. Por ejemplo, uno de ellos dijo lo siguiente: “Ubicaría el pivote a cuatros metros porque 1 mt más 3 mt = cuatro metros. Y Cristian está a ocho metros entonces 8m – 4m = 4 metros”, que algebraicamente es

$$x_3 - (x_1 + x_2)$$

Otros tres coincidieron en que debían sumar x_1 y x_2, luego dividir x_3 en la suma anterior y al resultado sumarle lo que sobraba a la derecha del Machín Machón, en decir 2.

$$\frac{x_3}{x_1 + x_2} + 2$$

El razonamiento de otro se enfocó en la lógica del juego, argumentando que debía estar cerca a x_1 y a x_2 para poder “levantar” a x_3, de manera que sumó x_1 y x_2, obteniendo así cuatro.

Otro estudiante dijo que a seis metros argumentando así: “Calculo las dos distancias y después las sumo y las divido por 2 y así me da el resultado donde colocaría el pivote” con las “dos distancias” se refería a la distancia de x_3 a cada uno de los otros dos niños. Algebraicamente es

$$\frac{(x_3 - x_1) + (x_3 - x_2)}{2}$$

Finalmente, otra opción fue a cinco metros procediendo de la siguiente manera: “Ubicaría el pivote a 5mt. Sumando las tres cantidades que me dieron que daría un total de 12, esta cantidad lo dividiría en la cantidad de niños que sería en 3 y a este resultado le sumaría x_1 y el resultado que me de ubicaría el pivote que sería a 5mt” que algebraicamente es:

$$\frac{x_1 + x_2 + x_3}{3} + x_1$$

Dado que la respuesta correcta era cuatro, entonces, en común acuerdo, las estrategias de los estudiantes que dijeron cinco y seis fueron descartadas, ahora ellos debían probar las estrategias de sus compañeros o encontrar su propia estrategia.

Como la idea era generalizar, cada uno de los estudiantes manipuló la simulación en diferentes casos para x_1, x_2, y x_3, hasta que, por ensayo y error, llegaron a un procedimiento que funcionaba. De los siete estudiantes, seis concluyeron que la ubicación del pivote se calculaba procediendo de la siguiente manera:

$$\frac{x_1 + x_2 + x_3}{3}$$

El otro no llegó a una generalización puesto que solo se limitó a un caso. Entre los argumentos de los que acertaron están: “El procedimiento: primero se tiene en cuenta x_1, x_2, x_3 y después los resultados se suman y se divide por la cantidad de niños que haya en el Machín Machón” o “en general: se puede utilizar mi procedimiento porque funciona en todos los ensayos que hice, en general el procedimiento sería: Sumar las cantidades dadas y dividirlas en la cantidad de niños (3) y el resultado de la división es la ubicación del pivote”.

Al igual que como se hizo en el caso de dos niños, los estudiantes observaron las distancias de los niños al pivote, tanto a la derecha como a la izquierda, acordando en que la suma de las distancias a la derecha del pivote era la misma que a la izquierda del pivote.

Por ejemplo, en el caso en que $x_1 = 1$, $x_2 = 3$ y $x_3 = 8$, como se observa en la Figura 25, vemos que “al sumar la distancia de un lado da 4 porque 3 + 1 = 4 y el otro da 4”, como lo expresa uno de ellos. En general todos dieron cuenta de ello con otros ejemplos, sin embargo, para ninguno le fue posible expresarlo algebraicamente, por lo que fue necesario intervenir y discutir este hecho entre todos.

Figura 25. Visualización Actividad Machín Machón en Geogebra asociado a las distancias del pivote a tres niños

De manera que si P es la ubicación del pivote, se tiene que:

$$(P - x_1) + (P - x_2) = x_3 - P$$

$$P - x_1 + P - x_2 = x_3 - P$$

$$P + P + P = x_3 + x_1 + x_2$$

$$3P = x_1 + x_2 + x_3$$

$$P = \frac{x_1 + x_2 + x_3}{3}$$

Es decir, acá se pretendió hacer énfasis en que, partiendo de la propiedad que cumple la media se encuentra la expresión que la define.

Después de lo anterior, se hizo una retroalimentación de lo que habíamos estado haciendo desde la actividad 0 hasta el momento y hacer mención, por primera vez del término “media”.

I: Cuando veíamos el caso de Luis (en la Actividad 0) y que teníamos tres mediciones nosotros qué hacíamos, para quedarnos con una sola.

A: Las sumaba y las dividía en tres.

I: Cuanto teníamos el caso de los caramelos, que yo les daba a ustedes cierta cantidad de caramelos, diferentes a cada uno, qué hacíamos al final.

M: Encontrar una cantidad justa.

I: Cómo calculábamos esa cantidad justa, Miguel.

M: Los sumábamos todos y los dividíamos en siete.

I: Ahora acá qué hacemos, tenemos tres individuos en un Machín Machón a cierta distancia del extremo izquierdo y cómo encontramos el pivote.

M: Sumando los tres y dividiéndolo entre tres.

I: Entonces qué tiene en común lo que hicimos con Luis, con las medidas de Luis, lo que hicimos con los caramelos, y lo que hicimos acá [con el Machín Machón]…

M: Que se suma y se divide

I: Que se suma todo y se divide…

A: Se divide por la cantidad

I: Bueno, y eso cómo se llama… es decir, en el caso de la actividad 0, ese valor nos ayudó a predecir, en el caso de los caramelos, ese valor fue un valor qué…

MF: Justo

I: Y en el caso del Machín Machón… qué pasa cuando yo encuentro ese valor, qué pasa con el Machín Machón

M: Se equilibra

I: O sea que también me sirve como un valor para equilibrar. Pero al final estamos haciendo la misma operación ¿cierto?

M: Sí

I: Bueno pero ustedes… ¿Ustedes han visto Estadística? ¿Qué han visto?

(En algún momento hicieron alusión a la media)

I: La media, qué era eso de la media.

A: Sumar todos los datos y dividirlo… igual que esto.

I: Igual que esto ¿cierto? O sea, todo lo que hemos estado haciendo es calcular la… la media.

Como se observa en la anterior discusión, ellos hacen alusión a la media como un cálculo, pero en el desarrollo de estas actividades pudieron explorar o experimentar, que la media va más allá de un algoritmo y que cumple ciertas características: como un valor justo, en el caso de los caramelos; como un valor que nos ayuda a predecir, en el caso de Julio; y como un punto de equilibrio, en el caso del Machín Machón y, de igual manera, que su ubicación entre los datos no es al azar, como la propiedad que se evidenció en la actividad del Machín Machón al observar la suma de las distancias a lado y lado de la media, que también fue vista en términos de desviaciones o diferencias, donde la suma es cero.

De manera que, en esta actividad, se acordó que la media nos indica dónde debería ir el pivote, el siguiente paso fue trabajar con un archivo en Fathom, manteniendo el contexto del Machín Machón, para descubrir una segunda propiedad: "La suma de los cuadrados de las distancias de los valores observados a la media, es la más pequeña", tal como se desarrolló en la siguiente actividad, 1.2.

4.5 Actividad 1 (Parte 2). El Machín Machón

La Actividad 1.1 concluyó con el descubrimiento de una propiedad de la media: "la suma de las desviaciones sobre la media es cero" o "la suma de las distancias a la derecha sobre la media, es igual a la suma de las distancias a la izquierda sobre la media". De manera que el principal objetivo en esta segunda parte fue el descubrimiento de una segunda propiedad: "La suma de los cuadrados de las distancias de los valores observados a la media es la más pequeña" es decir, dicho de otra manera, es la media el valor más cercano a todos y cada uno de los demás valores.

Para facilitar una mejor visualización de esta propiedad, la herramienta computacional usada en esta parte fue Fathom.

Análisis

Puesto que en la Actividad 0, los estudiantes habían tenido un primer acercamiento con Fathom, esta actividad se inició creando una colección, llamada Machín Machón, una tabla y un gráfico como se muestra en la Figura 26

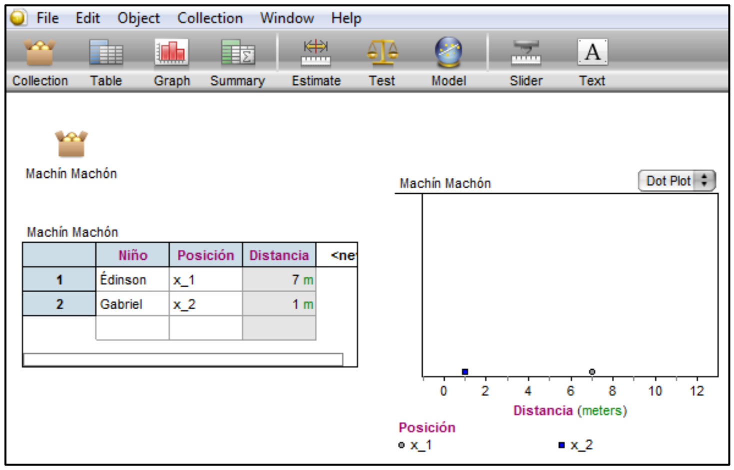


Figura 26. Pantalla de Fathom asociada a la Actividad 1.2

La columna “Distancia” estaba bajo la función “randomInteger(10)” de manera que cada vez que ellos oprimieran “Ctrl+y”, los valores cambiaban aleatoriamente. Lo anterior fue realizado por ellos bajo indicaciones del taller.

En primera instancia, debían calcular la ubicación del pivote cada vez que oprimieran “Ctrl+y”. Como ya se había discutido en la Actividad 1.1, todos coincidieron en que se sumaban los dos valores y se dividía en dos, es decir, la media, por lo que el siguiente paso era visualizar este valor en la gráfica.

Después de ubicar la media (línea azul en la Figura 27) y de hacer una retroalimentación sobre las distancias de los valores a la media, lo siguiente fue ubicar una línea movible (línea café en la Figura 27) y asociar cada distancia con un cuadrado como se observa en la Figura 27.

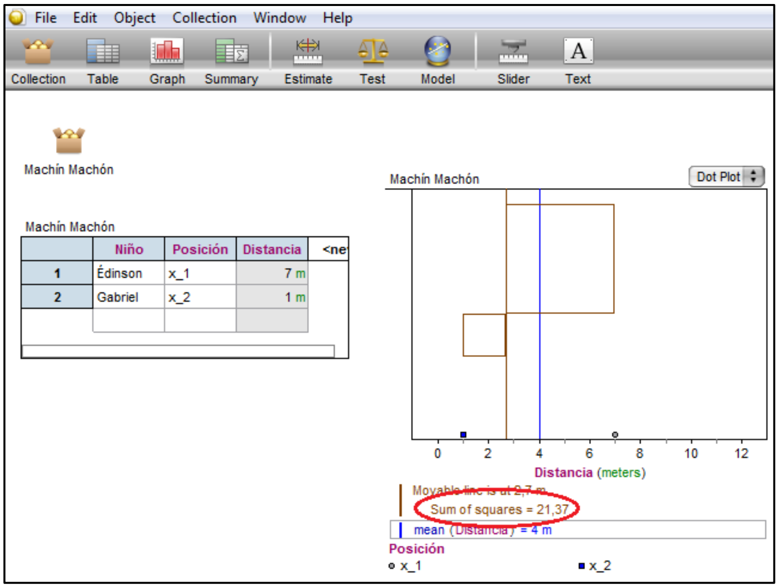


Figura 27. Pantalla de Fathom asociada a la Actividad 1.2 (media y suma de cuadrados).

Las preguntas asociadas a la Figura 27 fueron las siguientes:

1. ¿Cómo se forma cada cuadrado?
2. ¿Cómo se calcula el área de cada cuadrado?
3. ¿Qué significa la suma de las áreas de todos los cuadrados?
4. ¿Para qué posición de la línea café esa suma es mínima?

Las tres primeras preguntas anteriores, sirvieron de preámbulo para acercarnos a la propiedad que queríamos evidenciar. Estas fueron discutidas entre todos, acordando que la suma de las áreas de todos los cuadrados equivalía a la suma de los cuadrados de las distancias de cada punto a la línea movible.

Para la cuarta pregunta bastaba con que prestaran atención a lo que pasaba con el valor de "Sum of squares" (ver Figura 27) cuando movían la línea café a lado y lado de la línea azul, la media, especialmente cuando estas dos líneas coincidían. De los siete estudiantes, solo dos afirmaron que cuando las dos líneas coincidían, la suma era mínima: "Cuando la línea café llega a la línea azul que es la media el resultado de la suma es mínima” o “En la misma posición de la azul porque si la movemos a la izquierda o la derecha es mayor que 18”. Los demás decían, por ejemplo, que a la izquierda aumentaba y a la derecha disminuía, o que cuando se acercaba a uno de los valores de x_1 o x_2 era mínima, o comparaban cómo era la suma cuando estaban sobre x_1 o x_2 y escogían la mínima entre las dos.

Las anteriores respuestas fueron discutidas en grupo con el fin de observar cómo éstas cambiaban para cuando se repitiera la cuarta pregunta en el caso de tres valores, x_1, x_2 y x_3. Los dos estudiantes que decían que la suma era mínima cuando las dos líneas coincidían, ratificaron su respuesta para este caso, como es natural. Sin embargo, entre los demás estudiantes solo uno sumó a este hecho argumentando que “cuando coinciden el valor de las sumas es el mínimo” cuando antes de la discusión decía que “Para el lado derecho de la línea azul”.

Lo anterior se volvió a discutir entre todos con el ánimo de convencer a los demás de que, efectivamente, la suma era mínima cuando las dos líneas coincidían. Lo anterior no tendría por qué ser difícil puesto que, como se dijo anteriormente, bastaba con ver el valor de “Sum of squares” que Fathom nos proporcionaba. Al final, después de relacionar las respuestas a las cuatro preguntas previas, cuando se les pidió describir la propiedad de la media que acabábamos de observar solo cuatro estudiantes la enunciaron correctamente: “La suma de los cuadrados de las distancias a la media es el valor mínimo”. A continuación, un fragmento de la discusión que se hizo para llegar a esa conclusión.

I: ¿Qué se puede decir del valor de la suma de los cuadrados antes y después de la… es decir, si usted aleja la línea café de la media, qué pasa con la suma de los cuadrados?

W: Va aumentando

I: Va aumentando ¿cierto? Y cuando se acerca a la línea azul, qué pasa con la suma de los cuadrados.

M: Disminuye

I: Disminuye ¿cierto? ¿Qué pasa exactamente cuando la línea azul y la línea café coinciden?

A: Es el valor mínimo.

I: La suma es el valor mínimo… Entonces, en la actividad pasada nosotros habíamos visto una propiedad de la media, ¿se acuerdan cuál era? Que la suma de las qué… en el Machín Machón…

A: De las distancias…

I: Pero luego vimos otra, que si nos daba negativo cómo se llamaba eso.

A: De las… ¿desviaciones?

I: Desviaciones… Que la suma de las desviaciones a la media daba cuánto.

Todos: Cero

I: Cero ¿sí? ¿Qué podemos decir acá? Qué propiedad podemos encontrar acá… Muchachos a ver, tratemos de observar la propiedad que estamos viendo ahí. Estamos viendo que la línea café es movible ¿sí?

E: Sí

I: Y nos está dando unos cuadrados y la suma de las áreas de esos cuadrados, que viene siendo la suma de los cuadrados de las distancias. Pero qué pasa cuando esa línea viene siendo la media… Qué pasa con esa suma.

A: Es el valor más pequeño.

I: Entonces Andrea qué podríamos decir… ¿qué propiedades podríamos decir de la media?

A: (risas)

I: En la anterior… la suma de las desviaciones a la media es…

Todos: Cero

I: Entonces qué podemos decir acá… la suma de qué, Miguel…

A: La suma de los cuadrados es mínima cuando está… ¿en la media?

I: Algo así, la suma de los cuadrados es mínima cuando está en la media, es decir, la suma de los cuadrados de las distancias a la media es…

Todos: Mínima.

¿Pero qué significado tiene esa propiedad, en términos de la ubicación de la media respecto a los demás valores? ¿Acaso es "natural" para los estudiantes que este hecho implica que la media es el valor más cercano, en términos de distancias al cuadrado, a todos y cada uno de los demás valores? Pues no, no lo fue, por lo menos en nuestros estudiantes. Cabe destacar que era importante para nosotros que los estudiantes dieran cuenta de este último hecho puesto que, en el caso bivariado, es análogo cuando se habla de la recta de mejor ajuste. Lo anterior se evidenció en la siguiente situación que surgió en clase y que buscaba, en últimas, hablar de la media en términos de cercanía.

Situación: ¿Dónde ubicar la tienda?

A continuación, un fragmento de la discusión alrededor de la situación problema.

I: Imaginemos la siguiente situación: Cinco casitas que están sobre una carretera, que es recta, vamos a imaginarla así ¿listo? (Ver Figura 28)

E: Listo

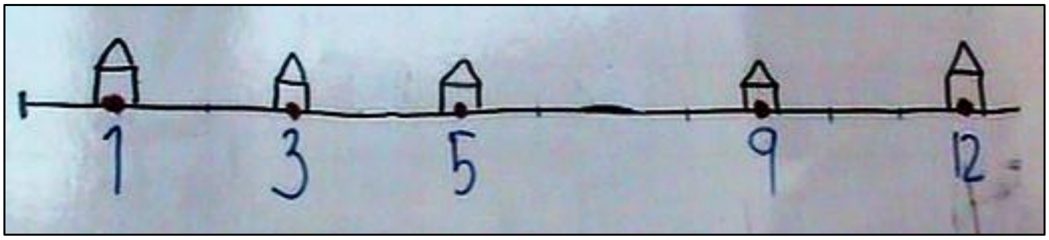


Figura 28. Figura asociada a la situación ¿Dónde ubicar la tienda?

I: Tenemos que ubicar una tienda, ustedes ya más o menos saben que, por su experiencia, ir a una tienda es lejos ¿sí? Puede que quede lejos, ¿sí o no? ¿A quién les queda cerca una tienda desde su casa?

E: A mí

I: ¿Usted tiene la tienda?

Todos: (Risas)

I: ¿Los demás? Lejos… ¿Es complicado, cierto? Lo digo porque yo también viví como año y medio en zona rural, trabajaba en un colegio, e ir a una tienda era complicado... Entonces, están las cinco casitas… Vamos a ubicar una tienda de manera que les quede más cerca a todos. ¿Dónde ustedes ubicarían la tienda?

I: A algunos les va quedar lejos, a otros les va quedar cerca, pero… denme valores y vamos mirando.

A, M, W: Seis

I: ¿Todos dicen que seis?

Todos: Sí

I: ¿por qué?

M: La suma de todos…

A: Daría treinta, dividido entre seis… cinco.

I; Ahora convénzanme, por qué. ¿Acá iría la tienda? (Ver Figura 29, tienda en rojo)

E: No

M: Sí

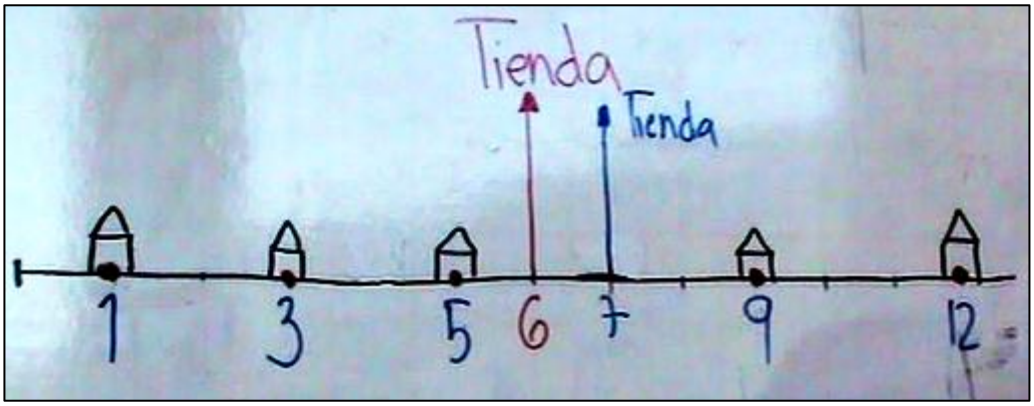


Figura 29. Ubicaciones propuestas por los estudiantes en la situación ¿Dónde ubicar la tienda?

I: Bueno, Elder dice que no ¿por qué no?

E: Porque no va en la mitad.

I: ¿Usted dice que en la mitad? ¿En la mitad de qué?

E: En la mitad de nueve y cinco.

I: ¿En toda la mitad de nueve y cinco? O sea, en dónde... ¿En siete? De cinco a nueve hay cuatro, dividido en dos, dos. Entonces según Elder, la tienda quedaría acá ¿listo? (Ver Figura 29, tienda en azul).

E: No sería justo que los otros caminaran más…

I: Unos van a quedar más cerca y otros van a quedar más lejos, eso sí. Pero entonces, vamos a mirar que los que queden más lejos…

E: Se ubiquen en… que quede más cerca.

I: Que los que queden más lejos, les convenga la más cerca… de alguna manera. Entonces vamos a mirar las distancias para la tienda roja y para la tienda azul.

A continuación, para facilitar la escritura, se denominó cada casa y cada tienda, siendo la tienda roja T_1 y la tienda azul T_2, y las casas C_1, C_2, C_3, C_4 y C_5, de izquierda a derecha; y se calcularon las respectivas distancias de cada tienda a cada casa (ver Figura 30).

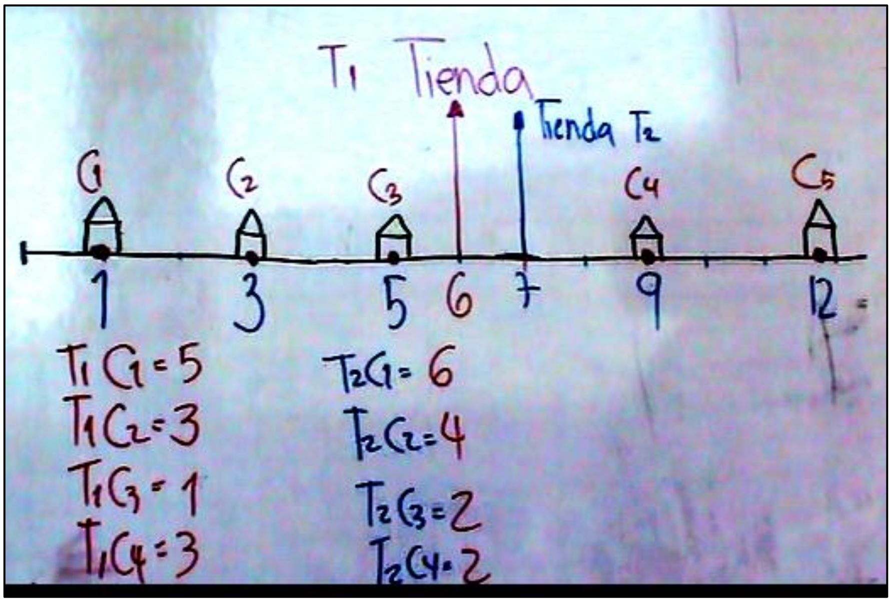


Figura 30. Figura asociada a la situación ¿Dónde ubicar la tienda?

El estudiante que escogió ubicar la tienda en 7, justificó su razonamiento de que ahí porque estaba en la mitad de cinco y nueve, y cuando se le indagó de que por qué no en la mitad de nueve y doce, dijo que porque ahí le quedaba muy lejos a las casas C_1, C_2 y C_3, y de forma análoga si la ubicábamos en la mitad de tres y cinco, quedaría muy lejos de 9 y 12. De manera que este estudiante estaba, de alguna manera, justificando el porqué de su elección, pero los otros seis estudiantes, no, además del cálculo de la media.

I: Miguel, por qué seis… además de que sumo y divido entre cinco y me da seis. Además de eso por qué seis. ¿Qué le garantiza a usted que esa posición es la más cercana a todos?

W: Pues porque está en la mitad de 1 y 12.

I: ¿Y los demás? La idea es que el recorrido que haga cada persona hasta la tienda sea qué…

Algunos: Igual

I: ¿Igual?

Otros: Justo

E: Y que no les quede tan lejos a los que viven…

J: Pero si fuera justo, significaría que… en este caso de justo, todos tendrían que recorrer lo mismo… ¿Pero será que eso ocurre?

E: No, porque unos viven más cerquita.

I: Pero que sí recorra lo mínimo ¿cierto? Que cada uno recorra lo mínimo. Unos van a tener que recorrer más que otros, nada qué hacer ¿cierto?

M: Sí

I: Pero hay una manera de que cada uno recorra lo mínimo… la mínima distancia.

Una vez calculadas las respectivas distancias, se generó la duda sobre qué hacer con esos valores y algunos de los estudiantes comenzaron a hacer sus cálculos.

I: Qué operación está haciendo, Elder.

E: Yo intento sumar el número de la distancia de cada casa hasta su tienda y dividirla por el número de casas que hay.

I: O sea está sumando acá esto (las distancias a T_1) Y cuánto suma eso…

E: Dieciocho

I: Suma dieciocho, y lo va a dividir en cuánto…

E: En cinco.

I: ¿En cinco? O sea, eso sería más o menos… ¿tres coma algo? ¿Y acá? (para T_2)

M: Diecinueve, dividido en cinco…

W: Tres coma…

I: ¿Daría más que acá (en T_1) o menos?

A: Más

I: Más… ¿Y de acuerdo a eso, Elber, entonces…? Ya le ayudamos a hacer la operación que usted estaba haciendo, que es sumando y dividiéndolo en cinco. Y acá (en T_2) dio más que acá (en T_1) ¿y entonces?

E: Entonces mi tienda fracasó (risas)

I: O sea que con hacer esa operación… ¿pero por qué se le ocurre hacer esa operación?

E: Porque así puedo sacar las distancias… de cada casa… de cuánto es de cada casa a la tienda.

I: Pero aquí ya sacamos cuánto es la distancia a cada casa.

Cabe resaltar que el hecho de considerar las distancias de cada casa a la tienda es muy válido, es decir, considerar las distancias al punto que se crea que va a producir el valor más pequeño. El punto buscado es la mediana. El elevar al cuadrado las desviaciones es para evitar el asunto de los negativos, el problema es que ya cuando se comparan las distancias al cuadrado el mejor punto se desplaza de la mediana a la media. Hecho que se recordó de lo que hacíamos en la actividad pasada, al considerar la suma de los cuadrados de las distancias a la media.

I: ¿Qué hacíamos nosotros ahorita en el taller que pasó? Qué calculábamos…

A: La media.

I: Sí, la media que fue lo que usted calculó acá (seis). Pero qué mirábamos luego.

A: La distancia de cada punto hasta la línea. De un punto hasta la media.

I: Bueno, hasta la línea movible que, en este caso, viene siendo la media. En este caso, esta tienda roja (T_1) vendría siendo qué.

E: La media…y la otra la línea movible.

I: Y la azul como la línea movible. Porque acá no hay manera de que haya otra tienda porque ya sabemos que la media es un solo valor ¿cierto? Porque ustedes al principio decían "No, el promedio está entre tanto y tanto" ¿Eso está bien decirlo?

E: No… porque toca sacar la operación primero.

I: Porque la media viene siendo un solo valor.

I: Entonces acá esta roja (T_1) no hay manera, o sea… está ahí, no se puede quitar porque es el promedio de los cinco. Pero la azul sí se podría cambiar porque dice "no, está en la mitad de cinco y nueve" pero, usando el mismo criterio, yo podría ubicarla en la mitad de nueve y doce, la podría mover allá, o en la mitad entre uno y cinco… la puedo cambiar. Pero si alguien me dice "es sumar y dividir en cinco" es seis y ya, no puedo hacer nada más. Entonces vamos a tomar esta (T_1) como la media y esta (T_2) como la línea movible. Entonces nosotros mirábamos las distancias… ¿y qué pasaba justo en la media?

E: El mínimo

I: El valor mínimo de qué…

E: De las distancias… de los cuadrados

I: La suma de los cuadrados de las distancias… Es decir, yo puedo tomar acá una distancia, entonces ahora vamos a tomar los cuadrados.

De manera que se tomaron los cuadrados de las distancias que ya se habían calculado y se sumaron para cada una de las dos tiendas propuestas, dando 85 para T_2 y 80 para T_1 (ver Figura 31).

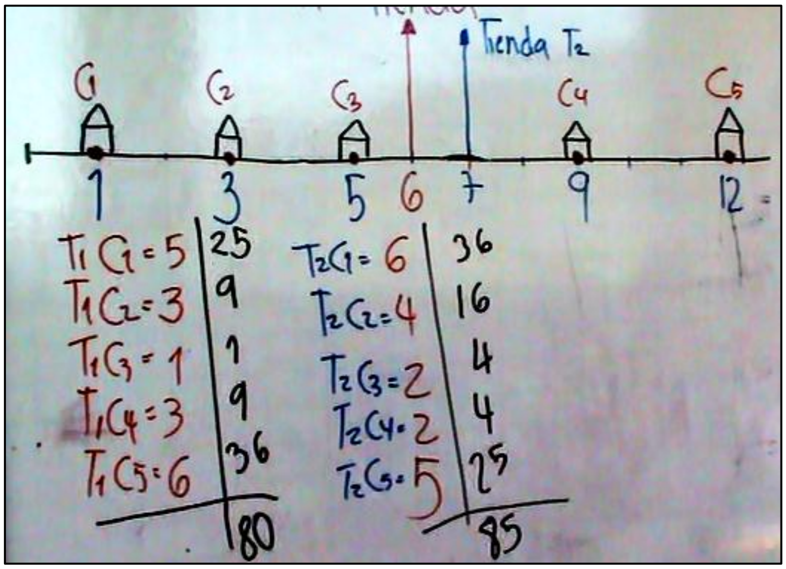


Figura 31. Figura asociada a la situación ¿Dónde ubicar la tienda?

I: Entonces dónde la suma de los cuadrados de las distancias dio mínima.

A y M: En la de la tienda uno.

I: ¿Entonces dónde ubicamos la tienda?

A: En seis.

Con esta actividad se dio por cerrado el caso univariado, es decir, hasta el momento nuestra intención fue darle un significado a la media más allá de su cálculo, concluyendo que, en primera instancia, es un representante del conjunto de datos, en el sentido de que nos permite *resumir* el conjunto de datos a un solo valor que, a su vez, nos ayuda a *predecir*. Además, es un valor que se encuentra *entre* los demás valores y que tiene algunas propiedades que la hace, como se vio en el caso de la tienda, la más *cercana* a todos y cada uno de los demás valores, cuando se considera los cuadrados de las distancias.

Lo que se resalta en cursiva, son ideas que pretendimos extender para el caso bivariado, como se mostrará en los análisis de las siguientes actividades.

4.6 Actividad 2. Actividades del profesor Gabo

Esta actividad, llamada "Actividades del profesor Gabo", fue nuestro primer acercamiento al caso bivariado. Como se dijo anteriormente, de aquí en adelante nuestro objetivo fue extender las ideas a las que habíamos llegado con datos univariados, a dos variables que, de alguna manera, se relacionaban.

Esta actividad se dividió en dos partes: Con la primera parte, que se llamó "De la casa a la universidad", y que seguía siendo un caso univariado, quisimos ratificar nuestra hipótesis de que, con lo visto en las actividades anteriores, los estudiantes asumirían la media como un buen predictor. Además de lo anterior, quisimos retomar el tema de la variabilidad. De manera que el primer ejercicio fue uno de predicción como se enuncia a en la Tabla 19

De la casa a la universidad

Tabla 19.

Tabla asociada a la situación "De la casa a la universidad" de la Actividad 2

El profesor Gabo, quien vive a unos kilómetros del campus universitario de la Universidad Industrial de Santander, ha registrado durante 10 días consecutivos, excepto uno, el tiempo en minutos que tarda conduciendo desde su casa hasta la universidad. En la siguiente tabla se muestran los registros del profesor Gabo.

Día	Tiempo (minutos)
1	15.0
2	10.0
3	8.5
4	
5	7.6
6	8.7
7	7.8
8	10.4
9	8.9
10	9.5

Como primer ejercicio, los estudiantes dieron razones del por qué el profesor Gabo no siempre demora la misma cantidad de tiempo, es decir, justificaron la variabilidad de los datos. Por ejemplo: “Que unos días se gasta menos tiempo en llegar y otros más. No se demora siempre la misma cantidad de tiempo porque puede haber congestión vehicular o a veces sale temprano o tarde” o “Que unos días tarda más que otros porque tal vez a veces hay algún accidente en la vía en que transita o porque se olvidó algo y se regresó a la mitad de camino, etc”.

A continuación, un fragmento de lo que se discutió en cuanto al reconocimiento de los datos y la variabilidad de los mismos.

I: Silvia, ¿cuántos registros vemos?

S: Diez

I: ¿Qué se está registrando? Elber.

D: El tiempo que gasta en ir a la universidad.

I: Pero hubo un día que no lo registró.

W: El cuatro.

I: El cuatro, ¿por qué pudo pasar esto, Elder?

E: Porque de pronto no fue, estaba enfermo.

A: Se le olvidó

I: Miguel, el profesor Gabo, no se ha cambiado de casa, tiene el mismo carro, se va por la misma ruta… ¿por qué cree que hay tiempos diferentes?

M: Porque pudo haber un trancón.

MF: Porque a veces se va tarde, y a veces se va temprano.

I: Por muchas razones hay esa variabilidad.

Otra discusión que se hizo fue sobre los valores “extraños” o atípicos, por ejemplo, en el día uno (ver Tabla 19).

I: ¿Qué opinan del día uno? El día uno cuánto tardó?

A: 15 minutos

I: ¿Es un dato muy similar a los demás o es un dato muy raro?

A: Mayor

I: ¿Muy mayor?

A: Por cinco minutos.

I: Por cinco minutos, pero los otros, más o menos… ¿están entre cuánto y cuánto?

A: Entre 7 y 10…

I: Más o menos ¿sí? Pero ahí (día 1) decía que tardo 15 minutos ¿es muy similar a los otros?

A, S y D: (niegan)

I: Es algo raro ¿sí? Tarda mucho más de lo normal, pudo haber pasado algo ahí.

Hecho lo anterior, el siguiente ejercicio fue uno de predicción. Se pedía predecir el tiempo que pudo haber tardado el profesor Gabo para el día 4 que, como se observa en la Tabla 19, no fue registrado y el tiempo que podría tardar el día 11.

El procedimiento para hacer la predicción tanto para el día 4 como para el día 11 fue unánime, la media. Sin embargo, para el día 4, un estudiante dividió la suma de todos los datos en 10 y no en 9, es decir incluye el valor que se está prediciendo: “Se sumaría, el resultado lo dividiría entre 10, si hay más de dos decimales tomaría el primero” mientras que para el día 11 sí dividió en 10.

Para el día 11 bastaba con incluir el valor predicho para el día 4, sumar todos los valores y dividirlos en 10, tal como lo hicieron cinco estudiantes: “Se tiene en cuenta el día cuatro y se suman todos los valores y se dividen en 10, en el día 11 tardaría 9,6”. Los otros dos estudiantes difirieron de los demás al dividir entre 11: “Él tardaría 8,7 minutos. Yo sumé todos los datos agregándole 9,6 y dividiéndolo en 11 que es el día que me piden porque es el procedimiento que vengo haciendo desde un principio”.

Lo anterior fue discutido, trayendo siempre a colación lo visto en actividades anteriores, hasta acordar que lo correcto es dividir entre el número de datos que se tenga: para el día 4 en 9 y, para el día 11, en 10, incluyendo el del día 4.

Ahora, ¿por qué proceder de esa manera, es decir calculando la media, y no de otra? La siguiente discusión pretendía indagar un poco más sobre sus razones al confrontarlos con otras estrategias que el investigador proponía.

I: Bueno, ¿pero por qué hacer ese procedimiento, Andrea? Por qué no escoger… no sé, si en el día 10 gastó 9,5 pues en el día 11 gastará… no sé, 9,6. ¿Por qué no hacer algo así? O bueno… si más o menos está entre 7 y 10 pues… no sé, la mitad, 8,5. Por qué ese es mejor del que yo estoy diciendo, Andrea?

A: Porque con la media se puede reducir muchos valores a uno solo… eh… es más fácil… (risas)

I: Ayúdenle a Andrea, porque todos hicieron eso. Y yo estoy diciendo "¡No! Si para 10 fue 9,5 para 11 es 9,6". O "¡No!, cojo el de la mitad, 8,5". Y ustedes están diciendo eso, entonces defiendan su argumento, ¿por qué ese valor?

Todos: (Silencio)

I: Ya Elber dijo algo… ¿Cuál es la diferencia entre lo que ustedes dicen y lo que yo estoy diciendo? ¿Yo estoy tomando cuántos valores?

M: Uno solo.

I: Uno solo ¿cierto? Mientras que ustedes están tomando…

M y A: Todos los valores.

I: Ustedes están teniendo en cuenta todos los… valores ¿cierto?... Inclusive, ¿la media tiene que ser un valor de los que me muestran en la tabla? Mafer…

MF: No

I: ¿Cuánto le dio a usted para el día 4?

MF: 9,6

I: ¿Ese valor estaba en la tabla?

MF: (niega)

I: Es decir, la media no necesariamente…

M: Tiene que estar en la tabla.

Hasta el momento, el hecho de que siguiendo el mismo procedimiento, el valor de la predicción hubiera dado el mismo en los dos días, no les inquietó a los estudiantes y, por nuestra parte, tampoco se indagó ya que, algebraicamente, se puede mostrar por qué pasa eso. Sin embargo, no fue hasta la Actividad 4 donde surge la necesidad de hacerlo, como se mostrará más adelante.

La segunda parte de esta actividad se llamó "Practicando natación" y se basó en la siguiente información:

Practicando natación

Tabla 20.
Tomado y adaptado de Moore (1995 p. 117-118)

El profesor Gabo, quien en su tiempo libre practica natación, ha registrado durante 23 sesiones consecutivas, el tiempo, en minutos, que tarda en nadar 1800 metros y su ritmo cardíaco, en pulsaciones por minuto, una vez ha completado los 1800 metros. Los tiempos y las pulsaciones se registran en la siguiente tabla.

	Minutos	Pulsaciones por minuto
1	34	152
2	36	124
3	35	140
4	34	152
5	34	155
6	36	128
7	36	128
8	35	144
9	35	148
10	36	130
11	35	136
12	36	124
13	35	148
14	35	144
15	35	140
16	34	156
17	35	136
18	34	148
19	34	148
20	36	132
21	36	124
22	35	132
23	36	130

Como se observa en la Tabla 20, se tienen dos variables. El primer ejercicio fue uno de predicción, queríamos ver cómo lo hacían cuando tenían dos variables, de manera que se les pidió predecir el tiempo y las pulsaciones del profesor Gabo para la siguiente sesión.

Dado que para este momento aún no se había hecho la discusión mencionada en el ejercicio anterior, es natural que los mismos estudiantes que calcularon correctamente la media arriba lo hicieran también ahora, es decir, sumar y dividir en 23: “El tiempo sería 35 porque se suman los tiempos y se divide por la cantidad, 23. En la siguiente sesión se suman todas las pulsaciones que da como resultados 3199 y se divide en 23 que son la cantidad de pulsaciones y da como resultado 139”.

Otro estudiante siguió con su proceso de dividir entre el dato siguiente, que en este caso sería 24: “El tiempo sería 35 minutos y sus pulsaciones serían 138 porque sumé y dividí sus minutos y pulsaciones y lo dividí por 24 que es el número de la sesión que sigue” Y, finalmente, una estudiante mostró otra estrategia al escoger el máximo valor en cada una de las dos variables: “Tendría en cuenta la máxima pulsación y el máximo de minutos o sea 156 y minutos 36”.

Como se muestra, los estudiantes que tomaron como estrategia sacar la media, trataron las dos variables de manera aislada. Es decir, el hecho de que en minutos les hubiera dado 35 no los detuvo para prestar más atención en los casos en que el profesor Gabo tardó 35 minutos, para observar cómo variaban las pulsaciones en ese caso en particular. Cabe resaltar que no esperábamos que pasara lo anterior, sino que asumieran la media como un buen predictor; pero independientemente de lo anterior ¿Qué ventajas tiene para ellos asumir la media? A continuación, algunas de sus respuestas:

- Es útil y fácil de calcular:

“Que siempre me va a dar un resultado correcto al sumar los valores y dividir entre los valores para predecir el valor oculto o desconocido. Este procedimiento es útil y fácil de usar. Reduce los valores a uno solo”. Al señalar que es correcto, está obviando la variabilidad.

“Que nos ayuda a encontrar un valor más rápido y sin tanto problema para hacer las operaciones y nos ayuda de tantos valores a predecir solo uno. Y nos ayuda a predecir muy bien porque si uno hace bien la operación le queda bien y además de eso nos ayuda a distribuir cosas de la vida cotidiana o diaria porque podemos hacer la operación y quedar con una misma cantidad para todos, por igual. Y para mí esta es mejor que todas para calcular”.

- Resume los datos a uno solo:

“Que podemos saber de muchos resultados un valor predictor. Sumando y dividiendo los valores dados”.

- Es un valor preciso y seguro:

“La media es un procedimiento el cual nos puede ayudar a predecir algo que necesitamos, nos ayuda a dar un resultado seguro y preciso el que nos ayuda también a dar una afirmación para cualquier operación o cualquier caso en la vida real para predecir algo que necesitamos saber”.

Como se observa en los anteriores argumentos, los estudiantes le adjudican a la media mucho más que un cálculo. Se refieren a ella en términos de un solo valor y no como un intervalo, como lo consideraban algunos en un principio. Un valor que toma en cuenta los demás valores, por la manera como se calcula. Un valor útil en situaciones cotidianas, como de repartición. Y un valor preciso y seguro, aunque este último argumento está obviando la variabilidad. Esto último también fue discutido como se muestra a continuación:

I: ¿El mejor predictor para la siguiente sesión podría ser la media?

Todos: (Asienten)

I: Ahora… ¿será que va a ocurrir eso necesariamente?

Todos: (Silencio)

I: Estamos prediciendo… ¿será que ese valor que yo predigo… necesariamente va a ocurrir al siguiente día?

MF: Puede ser

I: Puede ser… podría ser, pero no necesariamente. Pero si no ocurrió ese valor… ¿podría ser muy parecido o muy diferente?

M, A, D y E: Parecido.

Continuando con el tema de predicción, quisimos "sembrar" la primera semilla sobre predecir un valor de la variable respuesta (pulsaciones) para un valor de la variable explicativa (minutos). Por lo que se les pidió predecir las pulsaciones si un día, el profesor Gabo, completa los 1800 metros en 34 minutos. Ya sabemos que un experto en el tema, procedería con el método de mínimos cuadrados, calcularía la recta de regresión y remplazaría $x = 34$ y encuentra el valor de y. ¿Pero cómo lo hacen los estudiantes a este nivel?

Bueno, cabe destacar que el hecho de que se les haya pedido específicamente para 34 minutos "encendió" la alarma de los estudiantes para buscar otra estrategia diferente a sacar la media de los 23 datos. Se destacan tres estrategias:

- Cinco estudiantes tomaron como estrategia extraer de los datos los casos en que el profesor Gabo tardó 34 minutos y sacar la media de esos datos: "Busco los días que gasta 34 minutos y los divido entre el número de valores que es 6 y me da como resultado 147 entonces las pulsaciones si un día el profesor Gabo completa los 1800 metros en 34 minutos sería 147 pulsaciones por minuto” o “Sería de 151 pulsaciones porque si sumamos las pulsaciones de los días que el profesor hizo 34 minutos que fueron 5 días entonces sumé las pulsaciones de esos cinco días y los dividiría por 5 y el resultado sería 151. Pues eso sería la media la cual me ayudaría a dar esa afirmación”.

- Un estudiante dividió los metros recorridos por el profesor Gabo entre el tiempo que tardó, es decir, que en realidad calculó los metros por minutos, sin embargo, dijo que 152 cuando esa operación da 52, aproximadamente: “La pulsación del profesor Gabo es 152 saqué ese resultado colocando 1800 dividido en 34 y me dio 152, porque 1800 son los metros y 34 minutos”.

El hecho de que haya modificado su respuesta de 52 a 152 nos permite inferir que este estudiante comprende que su predicción no debería ser un valor “extraño” frente a los demás valores, como lo sería 52. O, por lo menos, que su predicción debería estar entre los valores que se tiene, como ya se había discutido previamente.

- Finalmente, una estudiante dio un intervalo tomando, de la tabla, el mínimo y el máximo valor en pulsaciones: “Las pulsaciones fueron 156, con respecto a la tabla y a las pulsaciones hechas en ese tiempo pudieron ser de 124 a 148 o 156, como lo dije anteriormente”.

Como se observó en el desarrollo de esta actividad, el eje temático fue Predicción, en primera instancia con una variable, en seguida con dos, y finalmente, ya de manera más específica, predecir el valor de la variable respuesta para un valor de la variable explicativa dado. Tanto en el caso univariado como en el caso bivariado se ratificó, en la mayoría de los casos, la estrategia de la media.

Por otra parte, si viéramos los datos de la Tabla 20 en un gráfico de dispersión se observarían datos apilados, de manera que, por lo que observamos hasta este momento, podríamos inferir que, en un gráfico de dispersión, donde los datos se encuentran apilados, los estudiantes los resumirían tomando la media de cada pila, siendo, así, viable nuestra THA.

4.7 Actividad 3. El profesor Gabo y la natación

Esta actividad tuvo como principal objetivo explorar la relación entre dos variables cuantitativas, en particular, las de la actividad prevista: "Practicando natación". Es decir, la relación entre las variables *Minutos* y *Pulsaciones por minuto.*

Debido a que hasta el momento solo habíamos graficado para una sola variable, se les pidió a los estudiantes que hicieran, con ayuda de Fathom, un gráfico para cada una de las variables (ver Figura 32). Así, una manera de observar la relación entre las dos variables fue seleccionando una parte de los datos en el gráfico de *Minutos* y observar lo que pasaba en el de *Pulsaciones por minuto.*

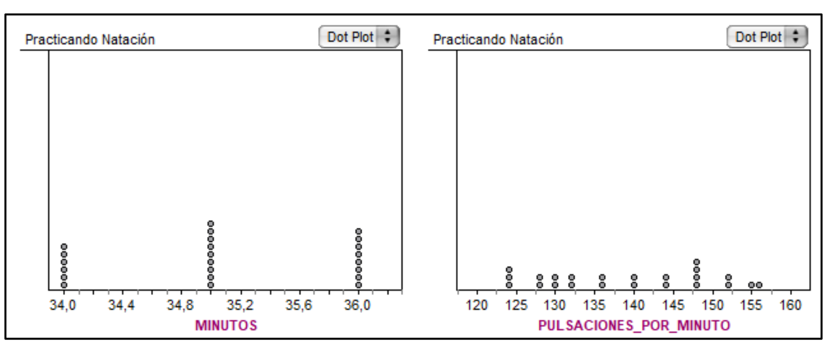


Figura 32. Gráficos asociados a la Actividad 3

Por ejemplo, se les indicó que seleccionaran los valores más pequeños en *Minutos*, es decir, los casos en que vale 34, y que indagaran sobre los datos que se seleccionaban en *Pulsaciones por Minuto* (ver Figura 33)*:* ¿Cuáles y cuántos datos se seleccionaron? ¿Cómo eran esos datos en *Pulsaciones por minuto* respecto a los demás? Una ventaja de Fathom es que, al seleccionar los valores en 34, automáticamente se seleccionan también en la tabla, permitiéndoles a los estudiantes dar cuenta de que los valores que se seleccionan en *Pulsaciones por minuto* se corresponden con los seleccionados en *Minutos.*

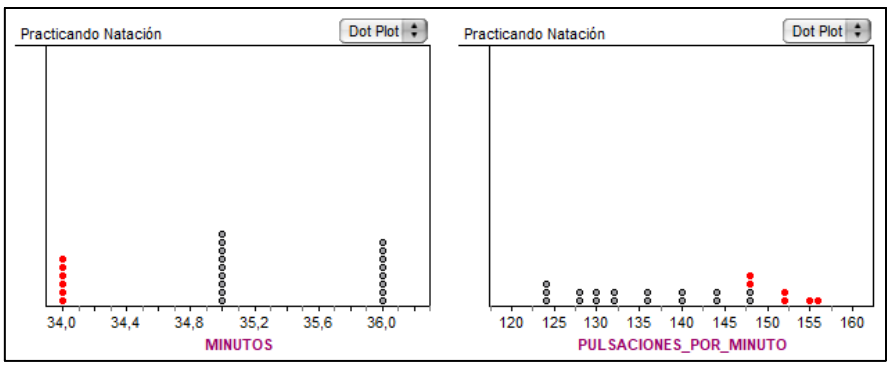


Figura 33. Relación entre Minutos y Pulsaciones por minuto desde diagramas separados. Acitivdad 3

Ahora, frente a la acción de seleccionar los valores más pequeños en *Minutos* están las siguientes explicaciones por parte de los estudiantes:

- Aquellos que se limitan a la acción que ocurrió automáticamente en la pantalla como: “Se seleccionan automáticamente y se colocan en rojo”.
- Los que dan cuenta de que corresponden a las pulsaciones en los casos de 34 minutos: “En la gráfica de pulsaciones, también se marcan las pulsaciones que tuvo el profesor Gabo por las veces que utilizó 34 minutos de natación” u “Observo que en el gráfico de pulsaciones

aparecen las pulsaciones hechas en los seis valores seleccionados en el primer gráfico de minutos…”.

- Aquellos que además señalan entre cuánto y cuánto van esos valores seleccionados: “Ocurre que se seleccionan los datos automáticamente en pulsaciones, hay 3 valores de 148, 2 de 152 y 1 de 155 y otro de 156 ubicados en el gráfico de pulsaciones” o “…Entonces en los valores más pequeños en minutos hay desde 148 pulsaciones a 156 pulsaciones”.
- Los que, además, dan un primer indicio de la relación existente entre las dos variables: “Al seleccionar los cinco valores en el gráfico de minutos; en el gráfico de pulsaciones se muestran las pulsaciones correspondientes a los minutos pequeños que son 34 minutos. Los valores en pulsaciones son los más grandes, y en minutos los más pequeños”.

Puesto que el objetivo principal era que evidenciaran la relación que había entre las dos variables, entonces, se les pidió específicamente que compararan los valores seleccionados en *Pulsaciones por minuto* con los demás valores, cuando seleccionaban los más pequeños en *Minutos*: “Son los más mayores porque en el gráfico muestra 148, 148, 152, 152, 155, 156 y son las pulsaciones mayores”. Un estudiante los compara directamente con los casos en que el profesor Gabo gasta 35 y 36 minutos: “Algunos valores son mayores y otros menores: cuando seleccionamos los valores más pequeños en minutos se seleccionan los valores más grandes en pulsaciones. Los valores seleccionados son mayores que las pulsaciones hechas en los minutos 35 y 36”.

Lo anterior se repitió seleccionando ahora los casos en los que *Minutos* eran mayores y, a manera de conclusión, se les preguntó si existía algún patrón o relación entre las dos variables, o cómo variaban las pulsaciones a medida que variaban los minutos.

Por ejemplo, al seleccionar los minutos mayores un estudiante señala que: “Ahora en la gráfica muestra las pulsaciones menores, no es nada igual a lo anterior porque ahora es lo contrario las pulsaciones son menores, 124, 124, 124, 128, 128,130, 130, 132 las pulsaciones menores” y concluye que: “A medida que los minutos van avanzando las pulsaciones van disminuyendo”. Por otra parte, otro estudiante, al no encontrar un patrón claro entre las dos variables, no aceptaba el hecho de que estuvieran relacionadas. Lo anterior se discutió entre todos como se muestra a continuación, hasta llegar a un consenso de que sí existe tal relación:

I: ¿Existe algún patrón o relación entre las dos variables?

E: Pues no porque si no entonces… porque entonces serían… siempre números pares. Digamos, en 36 tendría su número par, más grande.

I: ¿Cómo así?

E: Pues digamos, si existiera algún patrón o relación… Entonces en 36, como es en los minutos, entonces tendría que ser el doble de 36. Que sería el par… entonces no puede haber esto…

I: Por ejemplo, en 36, el doble sería 72… 72 en qué…

E: En pulsaciones… Pero no creo que haya algún patrón.

I: ¿Los demás? Elber… ¿Existe alguna relación o algún patrón?

D: No

I: ¿Por qué?

E: Porque entre menos minutos, mayor pulsación y entre mayor pulsación menos minutos.

I: ¿Y eso que está diciendo no me está mostrando una relación?

A: Sí

I: ¿Sí o no?

M: Sí

I: Eso que está diciendo ahí, precisamente me está mostrando la relación entre las dos variables.

I: ¿En general cómo se relaciones las variables minutos y pulsaciones sin tener que decir números?

M: Están variando.

I: Sí, están variando, pero de qué manera.

M: De manera que cuando hay mayores minutos hay menos pulsaciones.

Siguiendo la conversación anterior, también se evidencia que los estudiantes siguen asumiendo la media como una manera de resumir datos que se encuentran de manera apilada:

I: Pero pues acá no es tan fácil verlo… es decir, encontramos la relación "que a menos minutos más pulsaciones o que a más minutos menos pulsaciones" pero, encontrar el patrón pues no es tan sencillo ¿sí? No es tan sencillo porque, empezando para 34 tenemos muchos valores… ¿cuántos valores hay para 34, cuántas pulsaciones hay?

A: Seis.

I: Entonces en 34 qué valores había

M: 148, 152, 155, 156, 148 y 152.

I: Porque hay variabilidad… Entonces Elder… Usted estaba intentando encontrar algún patrón. Y, para empezar, 34 tiene muchos valores acá. Si nos queremos quedar con uno solo… ¿usted qué haría?

E: ¿los sumo y los divido?

I: Ah listo, suma y divide entre cuánto…

A: Entre seis.

I: Cuánto da eso…

M: [después de hacer el cálculo] 151.

I: 151 ¿ven que 151 no es ninguno de los que están acá, ¿no? Porque no necesariamente la media tiene que ser alguno de eso, estamos recordando. Pero sí está…

A: Ente los valores.

I: Entonces podríamos quedarnos con que para 34, más o menos es 151.

De la misma manera se procedió para 35 y para 36. Siendo para 35 la media 141, y para 36, 127.

I: Acá se ve la relación, Elder. Que al aumentar minutos, qué pasa con pulsaciones…

E: Disminuye.

I: Pero el patrón…

E: No lo encontramos…

I: No tanto… Porque para 34, 151: para 35, 141 ¿cuánto disminuyó?

M: Diez.

I: Para 36, 127… cuánto disminuyó.

E: 14

I: Sí va disminuyendo, pero no sabemos exactamente cuánto disminuye, sí una aproximación… es una aproximación. Pero la relación sí se puede ver.

Ahora, una vez identificada esa relación entre las variables, se les pidió que graficaran en un solo gráfico las dos variables, siendo así, nuestro primer acercamiento al gráfico de dispersión.

Curioseando un poco, los siete estudiantes lograron elaborar un solo gráfico. Tal como lo esperábamos, los gráficos se dividieron en dos grupos como se muestran en la Figura 34, donde tres estudiantes le apostaron a la forma A y cuatro a la forma B.

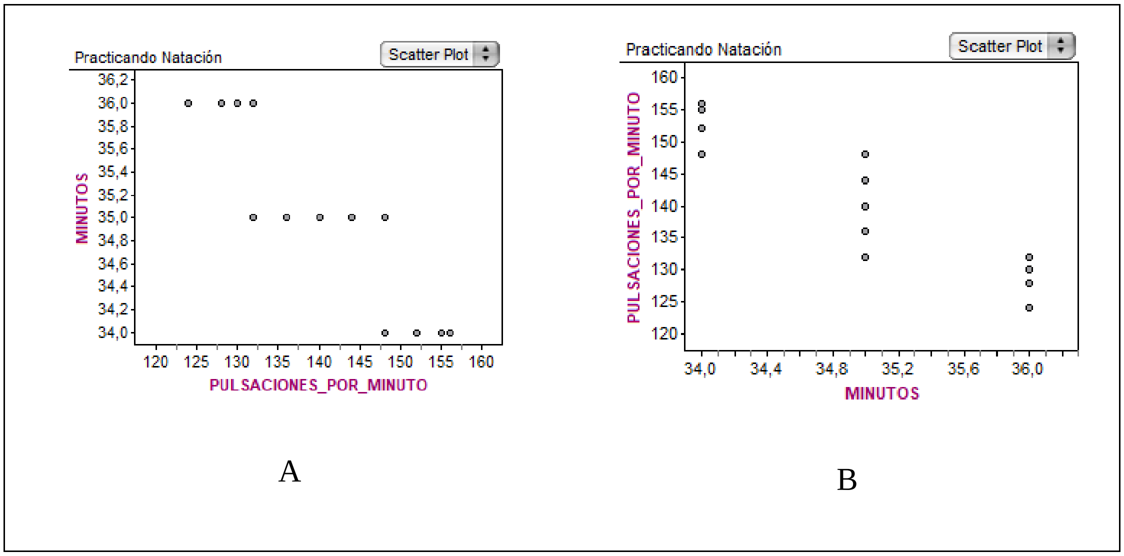


Figura 34. Formas en que graficaron los estudiantes dos variables en un solo gráfico

Una vez generada la discusión sobre qué significa cada punto en los gráficos se les preguntó lo siguiente: ¿Cómo se refleja, en esos dos gráficos, la relación o la variación entre las dos variables? En realidad, ante esta pregunta los estudiantes no respondieron otra cosa diferente a lo que ya habían dicho anteriormente respecto a cómo variaban las pulsaciones a medida que variaban los minutos, como es natural.

Ahora, al observar la Figura 34A, lo más correcto sería hablar en términos de cómo varían los minutos a medida que varían las pulsaciones y viceversa para la Figura 34B, cómo varían las pulsaciones a medida que varían los minutos. Y, yendo un poco más allá, lo más cómodo sería graficar como en la Figura 34B identificando la variable explicativa y la variable respuesta.

Pero hasta el momento pues no sabíamos nada de lo anterior, por lo que era natural que los estudiantes graficaran de esa manera, como en la Figura 34A; y también el hecho de que, un estudiante que graficó como en la Figura 34B, argumentara de esta forma: “Que en cada vez que las pulsaciones son mayores los minutos descienden” o que combinaran los dos gráficos: “Entonces mientras las pulsaciones aumentan, los minutos disminuyen y así también pasa cuando los minutos aumentan las pulsaciones disminuyen”.

Por otra parte, lo anterior no era relevante, puesto que, de una manera u otra, todos los estudiantes dieron cuenta de la relación que había entre las dos variables. Sin embargo, ninguno habló sobre cómo se reflejaba esa relación en el gráfico, o por lo menos no como nosotros lo esperábamos. Por lo que debimos indagar más en términos de forma, por ejemplo: cómo van los puntos, o hacia dónde van a medida que avanzamos hacia la derecha, etc… Como se discute a continuación:

I: Cómo se refleja el hecho de decir: “Al aumentar los minutos disminuyen las pulsaciones” ¿De qué manera van los puntos en este gráfico (como en la Figura 34B)?

W: Porque entre menos minutos hay más pulsaciones y por eso los puntos están más arriba.

I: Ajá, y acá… entre más minutos…

M: Menos pulsaciones, estarán más abajo.

I: Y en este (como en la Figura 34A), entre menos pulsaciones…

M: Mayores minutos…

I: Eso cómo se refleja ahí.

A: Están más arriba los puntos.

I: Y acá, entre mayores pulsaciones…

A: Menos minutos…

I: Entonces los puntos…

M: Están más abajo.

I: Entonces, los puntos, en cualquiera de los dos gráficos… ¿están ascendiendo o descendiendo?

A: Descendiendo.

Aprovechando lo anterior, se indagó sobre cómo sería el gráfico en el caso en que las dos variables tienen una asociación positiva. Y, puesto que hasta el momento habíamos visto datos apilados, esperábamos que graficaran de igual manera como, finalmente, lo hizo una estudiante:

I: Cómo sería el gráfico si las variables tuvieran una "relación ascendente". Por ejemplo, tengo dos variables x e y… ¿qué significa que esas dos variables tengan una relación ascendente?

A: Que las dos van a subir.

I: Entonces cómo sería el gráfico, Wendy.

En un principio marca un solo punto (ver Figura 35), como si lo comparara con los ejes x y y, es decir, a un valor grande en x, le corresponde un valor grande en y.

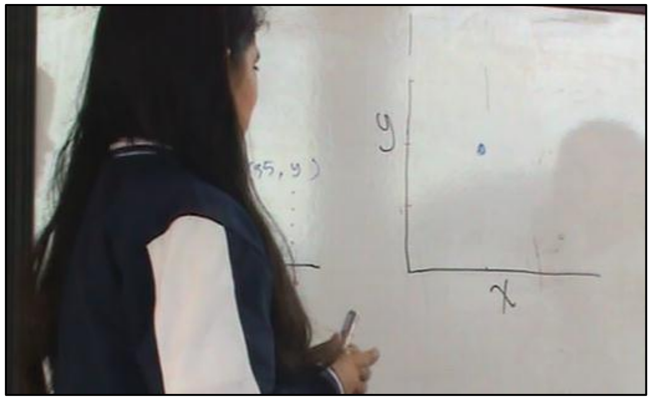


Figura 35. Forma en la que una estudiante grafica dos variables que tienen una asociación positiva

I: Pero ese es un solo punto… ¿ese solo punto me está diciendo algo?

M: No

I: Tengo que compararlo con otro.

De manera que se le pidió que marcara siete puntos (ver Figura 36) que, como se observa, los ubica de manera apilada.

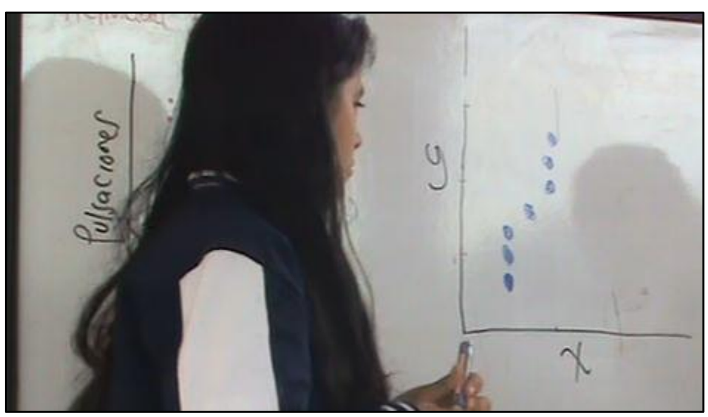


Figura 36. Forma en la que una estudiante grafica dos variables que tienen una asociación positiva (segunda vez)

I: Acá, a medida que aumenta x se ve que va aumentando y.

Finalmente, se hizo un paréntesis entre lo que habíamos estado haciendo en actividades anteriores respecto a la media y lo que estábamos haciendo ahora, que era relacionar dos variables y que una manera de ver esa relación era visualizando los datos en un gráfico de dispersión:

I: Entonces muchachos, cuál es la diferencia entre los que estamos viendo acá, y lo que vimos con la media…

A: Cómo así.

I: O sea, acá estamos relacionando cuántas variables…

A: Dos.

I: Estamos relacionando dos variables… Entonces, una manera de mirar la relación entre dos variables es graficarlo en un plano cartesiano… Esto se llama un gráfico de dispersión. Entonces, si los puntos van así [como en la Figura 34B], como hacia abajo… significa que las dos variables tienen una relación…

A: Descendente.

I: Descendente ¿sí? O una relación inversa. Que al aumentar una, disminuye…

M: La otra.

I: Ahora, si tienen una relación ascendente, significa qué…

A: Que van subiendo.

I: Que al aumentar una… qué pasa con la otra.

M: Va disminuyendo…

A: Va aumentando.

I: También aumenta…

M: ¡Ay sí!

I: Tienen una relación directa, una relación ascendente…

Ahora, después de discutir la forma como van los puntos cuando tienen una relación "ascendente" o "descendente" sin temor alguno, se indagó sobre cómo irían los puntos si las dos variables no se relacionan.

Es decir, grosso modo, si las dos variables se relacionan, esa relación se refleja de alguna *forma* en el gráfico, por lo que era natural que los estudiantes pensaran que, si las dos variables no se relacionaban, los puntos iban a estar "desordenados" como lo muestra una estudiante:

I: ¿Cómo creen ustedes que sería el gráfico si las dos variables no tienen ninguna relación?

A: Sería…

I: Andrea, sí… Vamos a pensar en dos variables que no tienen ninguna relación. Cuáles dos… porque por ejemplo la estatura y el peso… ¿tienen alguna relación esas dos variables?

E: (asiente)

I: Cómo, Elder.

E: Cuando uno va al médico siempre le dicen eso… que tiene que… que depende de la estatura eso es lo del peso, lo que debe tener.

I: Y qué relación tiene, ascendente o descendente… ¿directa o inversa?

E: (silencio)

I: Por ejemplo, Miguel y Elber… quién es más alto.

A: Miguel.

I: Quién cree que pese más, ¿Elber o Miguel?

E: Elber.

A: Nooooo, Miguel.

I: ¿Por qué?

A: Porque depende de la estatura pues va aumentando.

I: Ojo a lo que dice Andrea acá… si la persona es más alta, qué se espera del peso.

W: Que sea más.

I: Que sea mayor, es decir, hay una relación entre el peso y la estatura. Por ejemplo, si acá fuera la estatura [eje x] y acá fuera el peso [eje y], cómo iría… si yo le tomo la estatura y el peso a muchas personas, no solamente a estudiantes de octavo, sino… no sé… desde Kinder hasta once… cómo iría el gráfico, cómo irían los puntos, de qué forma.

A: [señala con su mano una recta diagonal con pendiente positiva]

I: Ajá, como así [imitando su seña] Ahora, yo les digo… hay dos variables que no se relacionan… cómo creen que irían los puntos. No sé… voy a inventarme dos variables: La edad con el número de mascotas.

I: Es decir, si dos variables se relacionan de manera directa, los puntos tienen una forma, tienden hacia dónde…

A: A ascender.

I: Ahora, si dos variables se relacionan de manera descendente, sí… que al aumentar una disminuye la otra… cómo se ve esto en el gráfico… [los estudiantes señalan como una recta con pendiente negativa] Se ven como de esta forma [imitando sus señas] ¿sí? Como hacia abajo, descendente, empiezan como a bajar… Ahora, si dos variables no se relacionan…

A: Pues estarían como en desorden…

I: Cómo… cómo estarían si dos variables no se relacionan.

A continuación, se le pidió a la estudiante que graficara cómo irían los puntos si las dos variables no se relacionan (ver Figura 37).

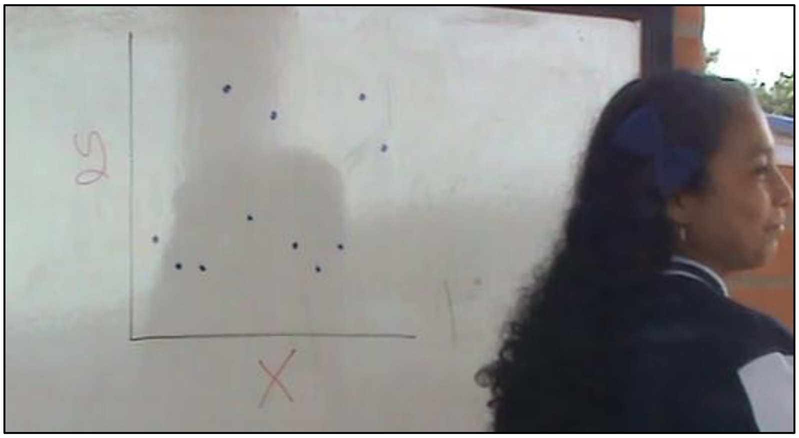


Figura 37. Forma en la que un estudiante grafica dos variables que no tienen ninguna asociación

I: Miren acá… estos puntos acá no tienen ninguna forma. ¿No podemos decir que al aumentar x aumenta y?

A: No

I: ¿o que al disminuir x aumenta y?

A: No

I: Entonces, esas dos variables no tienen una…

A y M: Relación.

Si observamos la Figura 37, podemos darnos cuenta que la estudiante abandona los datos apilados, puesto que pudo haberlos puesto de esa manera y, aun así, no reflejar ninguna relación. El hecho de que los haya ubicado así podría llevarnos a pensar que se está olvidando del tema de la variabilidad o que está reflejando un pensamiento determinista.

Como pudimos darnos cuenta, esta actividad fue nuestra llave para comenzar a trabajar con gráficos de dispersión. Se evidenció que los estudiantes resumen los datos apilados a un solo valor calculando la media de los mismos, por lo que nuestra THA era viable en cuanto a seguir trabajando con datos presentados de esa forma.

Respecto a la relación, de aquí en adelante los estudiantes hablaron en términos de “relación ascendente” o “relación descendente” cuando las dos variables tenían una asociación positiva o negativa, respectivamente. Un primer indicio de que tal relación puede ser resumida con una recta, fue el hecho de que los estudiantes señalaran con sus manos la *forma* de esa relación, como una recta con pendiente positiva o negativa.

Como vimos en actividades anteriores, los estudiantes asumen la media como un predictor en el caso univariado, pero ¿cómo predicen en un diagrama de dispersión, es decir, en el caso bivariado? Las Actividades 4 y 5 buscan dar respuesta a esa pregunta, para datos apilados y no apilados, respectivamente.

4.8 Actividad 4. Nivel de ruido

La predicción es uno de los ejes temáticos principales en el diseño de las actividades. Ahora, en un diagrama de dispersión donde se observa una relación lineal entre dos variables, predecir un valor de la variable respuesta para cualquier valor de la variable explicativa se resume en encontrar la mejor recta.

Para el caso bivariado, quisimos comenzar con la predicción entre dos puntos, luego dados tres, y aumentando así la cantidad de puntos hasta finalmente obtener varias pilas de datos que, de manera global, reflejaban ya fuera una asociación positiva o negativa. ¿Cómo predicen los estudiantes en un diagrama de dispersión, entre dos o tres puntos y entre datos apilados? A continuación, unas primeras formas de predicción.

Esta actividad que llamamos “Nivel de ruido”, fue tomada y adaptada de Moritz (2004, p.238). En un principio los estudiantes debían predecir, con base a unos datos mostrados en una tabla y, en seguida, validar sus respuestas con un archivo en Geogebra en donde ahora los datos se mostraban en un diagrama de dispersión, de manera que adicionaban sus predicciones al diagrama y decidían si movían o no sus valores predichos.

Análisis

Predicción dados dos puntos

Como se dijo anteriormente, un ejercicio inicial fue el de predicción dados dos puntos como se enuncia en la Tabla 21.

Tabla 21.
Tabla asociada a la Actividad 4 (predicción dados dos puntos)

<table>
<tr>
<td>Gabo, Édinson y Cristian hacen parte de un proyecto sobre el ruido. El proyecto consiste en visitar salones, contar el número de personas y medir el nivel de ruido en la clase con un medidor de sonido.</td>
<td>
<table>
<tr><td>Salón</td><td># de Personas</td><td>Nivel de ruido</td><td rowspan="5"></td></tr>
<tr><td>I</td><td>21</td><td>23</td></tr>
<tr><td>F</td><td>29</td><td>75</td></tr>
<tr><td>W</td><td>25</td><td></td></tr>
<tr><td>Z</td><td>28</td><td></td></tr>
</table>
</td>
</tr>
</table>

Se pedía predecir el nivel de ruido para un valor que estaba justo en la mitad de 21 y 29, es decir para el salón W y luego se pedía predecir para el salón Z, que también estaba entre los valores dados, pero no en la mitad de ellos.

Para este punto de la THA ya el tema de la media era tan familiar para los estudiantes que todos, de manera unánime, procedieron de la siguiente manera: Para el salón W, con 25 personas, sacaron la media de los dos valores mostrados, 23 y 75, y para el salón Z, con 28 personas, sacaron la media de los tres valores, es decir, adicionando ahora el hallado para el salón W. Algunos argumentan de la siguiente manera: “Para predecir el valor del salón W debo que tener en cuenta los datos I, F, y sumo el nivel del ruido y me da 98 y ese resultado lo divido por 2, esta forma porque sería algo fácil para predecir” o “Para predecir el valor del salón Z, hice el mismo procedimiento, sumar todos los valores sobre el nivel del ruido y dividirlos entre 3 y obtuve un valor de 49. Porque creo que esta operación puede ayudarme a calcular el valor y así poder predecir”.

El hecho de que les haya dado el mismo valor tanto para W como para Z no fue extraño para ellos, es decir, ignoraron completamente la variable *# de personas* e inclusive se salieron del contexto, despreciando la relación que había entre las dos variables.

Puesto que habíamos previsto que algo así ocurriría, se les pidió a los estudiantes que observaran el diagrama de dispersión generado por los datos de la Tabla 21, incluyendo sus valores predichos. Y en seguida se les indagó sobre si consideraban cambiar los valores que habían predicho para los salones W y Z. Lo que todos observaron se muestra en la Figura 38, siendo los puntos en rojo, las predicciones de los estudiantes.

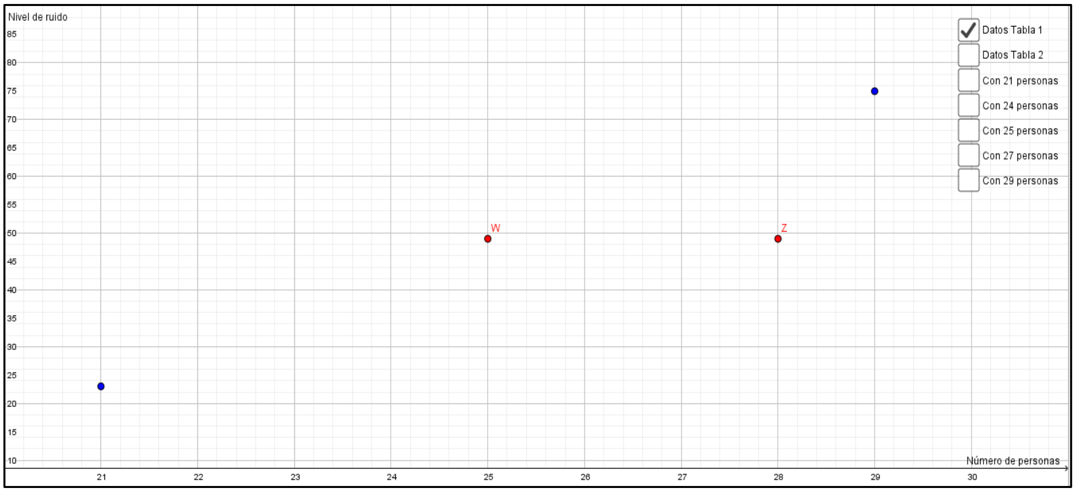


Figura 38. Gráfico de dispersión asociados a los datos y predicciones de los estudiantes de la Tabla 21

Aun así, visualizando los datos de esta manera, todos los estudiantes, a excepción de uno, consideraron que no deberían mover los puntos. Algunos, basando sus argumentos en la operación que hicieron: “Creo que no porque esa es la predicción de acuerdo a la operación de suma y dividido por la cantidad de los valores sumados de acuerdo a la media” y otros de acuerdo a la forma en que se encontraban los puntos: “No, porque quedaría como forma de escala y esa para mí quedaría también como una forma organizada”. El estudiante que consideró mover sus puntos argumentó de la siguiente manera: “Sí, porque me hubiera gustado que Z tuviera mayor nivel de ruido” donde, este último, implícitamente, está considerando la relación entre las dos variables donde al aumentar el número de personas, se espera que aumente el nivel de ruido.

Dado lo anterior, se hizo necesario la intervención del investigador incitándolos a predecir para otras cantidades de personas. Con estos nuevos valores los estudiantes continuaron haciendo la operación de sacar la media, obteniendo siempre el mismo nivel de ruido. En la siguiente discusión se llega a un acuerdo de que lo anterior no parece muy lógico.

I: En el salón… es decir, para el salón W, ustedes sumaron los dos anteriores, I y F y dividieron en dos.

M: Sí.

I: ¿Listo? Para el salón Z, sumaron los tres y dividieron en tres. Para un salón X, sumaron los cuatro y dividieron en…

A: Cuatro

I: Cuatro y va dar… va seguir dando qué…

A: Cuarenta y nueve.

I: ¿Y eso está bien? ¿Suena bien?

A y E: No (Risas)

I: Es decir, si seguimos haciendo ese procedimiento y llegamos no sé… a la letra "XXF" no sé… que tiene 50 personas… ¿Entonces sumo los 49 anteriores y divido entre 49?

A: No

I: ¿Siempre va dar 49 cierto? Es decir ¿qué no estoy teniendo en cuenta ahí?

D: ¿El número de personas?

I: El número de personas, ¿sí? El número de personas no lo estoy teniendo en cuenta. Es decir, ustedes están aplicando… ¿Qué están sacando en realidad ahí?

A: La media.

I: La media ¿cierto, Andrea? Pero acá la media como que no me está dando mucha información. Entonces si yo les pregunto a ustedes para uno que esté acá (señalando en un gráfico) en 28, en 25, en 24… siempre me va dar 49 ¿cierto? ¿Y ese gráfico les suena o no les suena?

A, W y MF: (niegan)

I: Aaah, entonces en el 2a, ya todos dijeron que en… “¿Consideran cambiar los puntos?” dijeron… qué dijeron…

Todos: No

I: Y de acuerdo a lo que discutimos ahorita ¿ustedes qué creen?

Todos: Que sí.

I: Es decir, un salón con 50 personas…

A: No va hacer el mismo ruido que uno por ahí de 23.

I: Uno qué esperaría

M: Que hiciera mayor ruido

Otra discusión que se hizo alrededor de lo anterior fue sobre *¿por qué está pasando eso?* Es decir, por qué al operar como ellos lo estaban haciendo, siempre daba el mismo valor, cuarenta y nueve.

I: ¿No sabemos por qué está pasando eso?

E: ¿No será porque la media sirve para algunas cosas y para otras no?

I: Es decir ¿si todos los salones tuvieran 25 personas serviría ese procedimiento?

A y M: Sí

I: ¿Por qué?

M: Hay el mismo número de personas.

I: Todos tienen en común la cantidad de personas. Entonces vamos a mirar por qué está pasando eso.

La profesora-investigadora, demostró algebraicamente la razón de la igualdad de las predicciones cuando se usan los resultados previamente calculados.

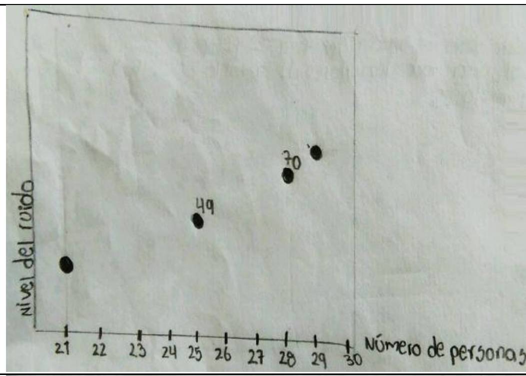
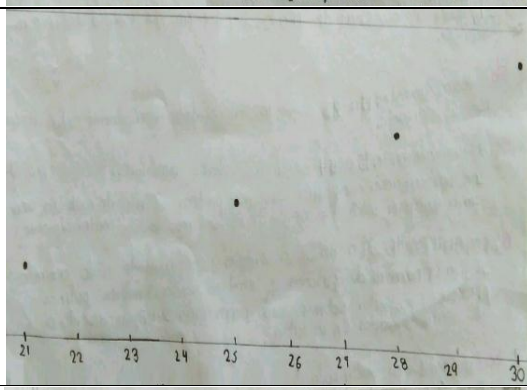
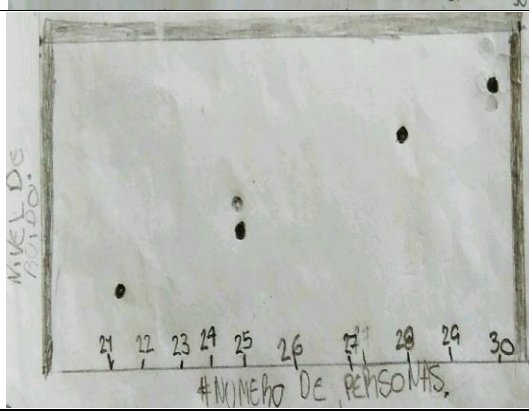
Se observa en las respuestas de los estudiantes como han asumido la media como la respuesta automática a toda pregunta que se relacione con predicción como bien lo hicieron en el caso univariado, sin hacer ninguna referencia a la variable número de personas. Logran relacionar las dos variables cuando la profesora les hace caer en cuenta de lo absurdo que sería que el ruido fuera igual en grupos de diferente tamaño. En este caso, el contexto ayudó claramente a asumir un punto de vista bivariado en el que, claramente, las dos variables están relacionadas.

Finalmente, los estudiantes aceptaron el hecho de que este procedimiento no les permitiría predecir y que sí debían mover los puntos, por ejemplo: "Sí, porque si en unos salones hay más personas pues es mayor cantidad de ruido y en donde hay menos personas el ruido es menor" o "Sí, porque como vamos a esperar que todo lo sigamos dividiendo y sumando y nos daría el mismo resultado y eso no sería lógico".

¿Pero y entonces cómo? Claro, lo que resultaba ser una buena predicción en el caso univariado, no lo era acá, en el caso bivariado. De manera que se les pidió que buscaran otra manera para mejorar sus predicciones y que movieran los puntos hasta donde consideraran que estaban bien ubicados. Los siete estudiantes reflejaron en sus gráficos una asociación positiva entre las variables, como se observan algunos en la Tabla 22. Según los argumentos, la estrategia de ellos fue mover libremente los puntos para que al aumentar el número de personas el nivel de ruido también aumentara. Observemos también que, en cuanto a forma, casi que buscan que los puntos sean colineales.

Tabla 22.
Gráficos de dispersión asociados a los datos y nuevas predicciones de los estudiantes de la Tabla 21

Nuevo diagrama de dispersión	Argumento

	Porque considero que a medida que hay más personas el nivel de ruido sube.
	Estoy de acuerdo porque: Los valores obtenidos son diferentes y en el salón con 28 personas hay más ruido que en el salón con 25 personas. El valor con 28 personas tendría 42 (valor de ruido) y para las 25 personas 32. Estoy de acuerdo porque cada valor varía de acuerdo al número de personas.
	Porque creo que es así por su orden.

Las predicciones para un salón con 25 y otro con 28 personas se discutieron entre todos acordando que para 25 estaba bien pero que para 28 sí deberíamos moverlo. En la Figura 39 se observan las predicciones de todos los estudiantes para 28, en las que algunos coincidieron, y también la predicción que propone el investigador para 28 que es 10, el punto más bajo.

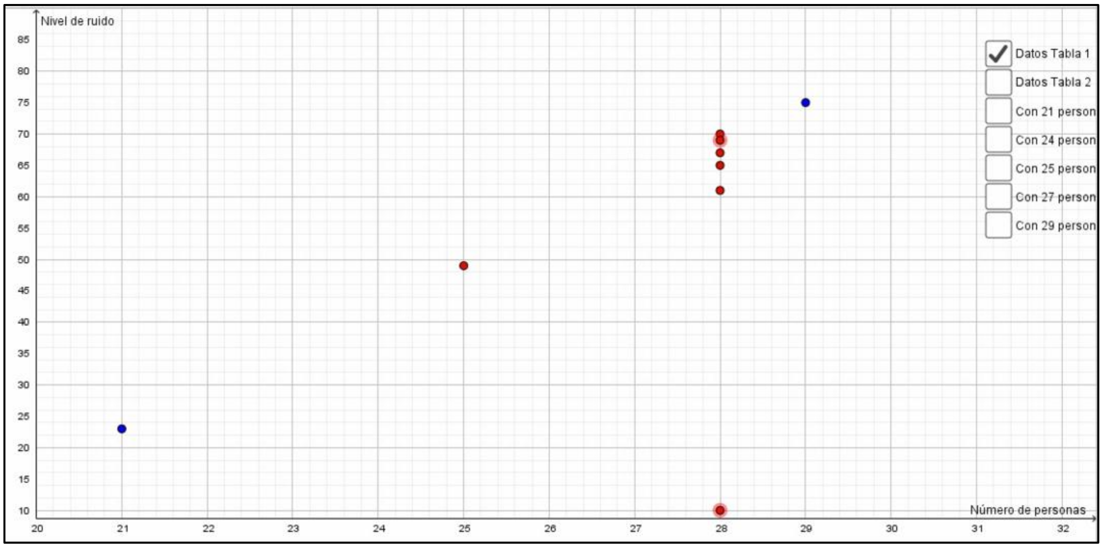


Figura 39. Predicciones de los estudiantes para un salón 28 personas. Actividad 4

La siguiente discusión se hace alrededor de la Figura 39 e incluye una reflexión sobre datos atípicos como el que se observa para 28. También con qué valor, entre los que ellos predicen para 28, nos quedamos.

I: Entonces ¿todos esos valores [para 28 en la Figura 39] que están ahí son posibles?

A: No

I: ¿Cuál no es posible, Andrea?

A: El que usted colocó

E: El 10.

I: Pero si yo soy la profe… ¿por qué no?

M: Porque no se puede

E: Porque usted se puede equivocar

W: Entonces cómo va a ser eso.

I: Ajá… ¿Y todos esos valores que ustedes dijeron esos sí son posibles?

E, M y A: Sí.

I: ¿Por qué?

A: Porque está entre el valor de 25 y el de 29.

I: Definitivamente este valor que está acá [el punto más bajo en la Figura 39] es un valor raro. En estadística se llama dato atípico, es un dato raro, extraño. Que puede pasar debido a algunas razones, podría pasar porque… Esto podría pasar, es raro pero podría pasar.

I: ¿Qué podemos hacer con esos datos raros? Si yo ya sé que la relación que hay entre las dos variables es "ascendente" Qué puedo hacer yo con este dato…

Todos: [Silencio]

W: Descartarlo

I: Descartarlo, es decir, podría ignorar ese dato porque ese dato es raro. Es decir, cuando se hizo la medición qué pudo haber pasado.

E: Midieron mal

I: Midieron mal, el aparato no tenía pilas

M: Estaban muy callados

I: De pronto estaban regañados los estudiantes de ese salón.

I: Entonces, tengo cinco valores [en 28], quiero solamente un valor… ¿Cuál es el mejor de todos esos?

E: Profesora se puede sacar entonces la media ahí.

Predicción dados tres puntos

Lo anterior se repitió para el caso de tres puntos, en donde, como se observa en la Tabla 23, para un valor de la variable explicativa, *# de personas*, se tenían dos valores en la variable respuesta, *Nivel de ruido*. Nuevamente se pidió predecir el nivel de ruido para los salones W y Z.

Tabla 23.
Tabla asociada a la Actividad 4 (predicción dados tres puntos)

Salón	# de Personas	Nivel de ruido
N	21	25
J	29	81
M	29	85
W	25	
Z	28	

Entre las formas de predicción de los estudiantes se rescatan las siguientes:

Calculando: Cuatro estudiantes hicieron cálculos diferentes para predecir. Por ejemplo, una estudiante procedió de la siguiente manera:

- Para el salón W: "Sumé los valores anteriores del número de las personas del salón N, J, M y el nivel del ruido del salón N, J, M y los dividí en 6 y me da como resultado 45 que es un valor coherente porque hay 25 personas y el ruido puede ser 45. W=45
- Para el salón Z: "Sumar los valores anteriores pero esta vez incluyendo los valores del salón W y dividir en 8 y da 42. Z=42"

Ya, en el ejercicio anterior, se había discutido la relación entre las dos variables, es decir cómo varían de manera conjunta y que era necesario tener en cuenta las dos variables para hacer una predicción. Por lo que, como se puede apreciar en sus argumentos, su predicción no obedece a esa relación, Es decir, su esfuerzo por tener en cuenta las dos variables, radicó en calcular con todos los datos de una u otra manera. Y no es hasta que observa los datos en un diagrama de dispersión cuando acepta el hecho de que sus predicciones no son las mejores.

Sin duda la representación geométrica juega el papel de juez del procedimiento algebraico utilizado gracias a que permite una percepción conjunta de los valores en juego, algo que es muy difícil con el solo trabajo algebraico. Se nota que los estudiantes, si bien tienen en claro que la predicción que se haga debe tener en cuenta las dos variables no logran concretarla en una operatividad adecuada.

Por lo que a la pregunta de si considera o no cambiar de posición sus predicciones contesta que: "Sí, porque el nivel del ruido para 28 personas no puede ser menor que el nivel de ruido para 25 personas, debe ser un poco mayor", es decir, esa forma de representación tabular no le permitió apreciar que los puntos siguieran una "forma", ascendente en este caso, como sí lo hizo el diagrama de dispersión. En la Figura 40 se observa cómo esta estudiante finalmente ubica los puntos, donde deja igual para 25 y cambia su predicción para 28.

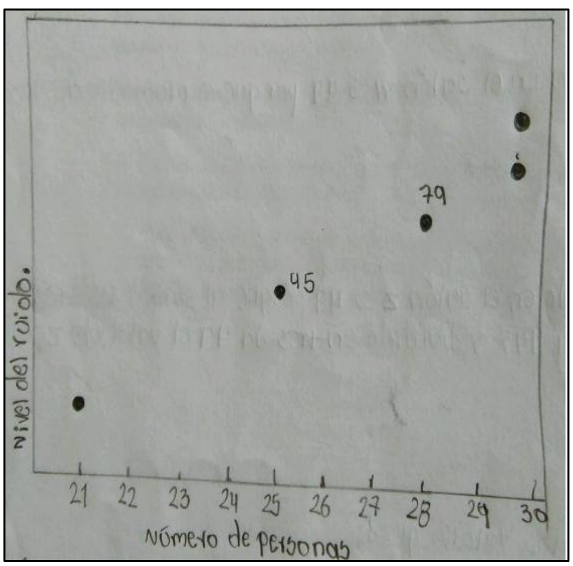


Figura 40. Predicciones finales de una estudiante asociadas a los datos de la Tabla 23

Argumentando que ahora sí está bien: "Considero que está bien porque está de manera ascendente". De igual manera, otra estudiante también movió los puntos al observar que, en su gráfico de dispersión, W estaba más arriba que Z.

De los otros dos estudiantes, que también hicieron algún tipo de cálculo, uno consideró no cambiar de ubicación sus predicciones puesto que en el gráfico de dispersión conservaban una forma ascendente. Mientras que el otro, a pesar de que en su gráfico (ver Figura 41) se observa que sus predicciones (las dos del centro), junto con los otros dos puntos, no respetan la relación entre las dos variables considera que están bien: "No, porque considero que, así como está, [está] bien… así que no veo la necesidad de cambiarlo".

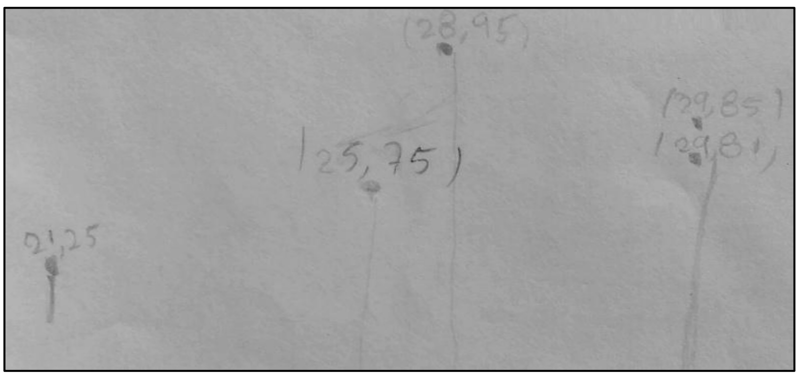


Figura 41. Predicciones de un estudiante asociadas a los datos de la Tabla 24

Aproximaciones: La otra manera de proceder de los otros tres estudiantes, se puede resumir en que hicieron aproximaciones teniendo en cuenta la relación entre las dos variables y los datos mostrados en la tabla. A continuación, el argumento de uno de ellos: "Mis valores dados son una aproximación a los valores que me están mostrando en la tabla, con respecto a los demás salones: Para el salón W aproximadamente 30, para el salón Z aproximadamente 75. Con respecto a la cantidad de personas y el ruido de los demás salones" y decide no cambiarlos “porque los valores acerca del ruido ascienden de acuerdo con el número de personas”.

Es decir, independientemente del procedimiento que hayan utilizado los siete estudiantes, todos, a excepción de uno, validan sus predicciones al observar que, en el gráfico de dispersión, reflejen la relación existente entre las dos variables, de no ser así, mueven los puntos.

Predicción con pilas de datos

Ahora se les presentaron a los estudiantes datos repetidos, esto es, pilas de datos para cada valor de la variable explicativa. Inicialmente dos pilas, luego tres pilas. Se pedía predecir para valores entre las pilas e inclusive predecir para un valor donde se tenía una pila de datos, ¿seguirían asumiendo la media y, de ser así, de qué manera?

En el caso de dos pilas de datos, los datos se presentaron organizados en tablas como en la Tabla 24, pero también se podían visualizar en un diagrama de dispersión en un archivo de Geogebra. Ya para el caso de tres pilas, y en adelante, solo se contaba con el archivo de Geogebra. Se pedía predecir el nivel de ruido para un salón con 25 personas y para uno con 28.

Tabla 24.
Tabla asociada a la Actividad 4 (predicción con dos pilas de datos)

Salón	# de personas	Nivel de ruido
A	21	11
E	21	19
F	29	75
I	21	23
J	29	81
M	29	85
N	21	25

¿Hacia dónde íbamos? Pues el objetivo final era que los estudiantes encontraran una estrategia para predecir el nivel de ruido para cualquier salón que tuviera entre 21 y 29 personas. A continuación, vamos a mostrar el proceso llevado por una estudiante (señalado por ✓) junto con su estrategia y predicciones. También mostraremos, en línea, "*Otras formas"* de predicción (señalado por •) destacadas en los otros estudiantes.

Predicción para un salón con 25 personas

✓ Con los datos de la Tabla 24, para predecir el nivel de ruido de un salón con 25 personas, procedió de la siguiente manera: "El nivel del ruido para un salón que tiene 25 personas sería 35 porque al sumar todos los valores de la tabla da 490 y este valor dividido en 14 da 35". Es decir, sacó *la media de todos los datos* como si correspondieran a una sola variable.

Otras formas:

- *Calculando la media de la variable "Nivel de ruido":* "El nivel de ruido para un salón que tiene 25 personas es 45, calculé este resultado sumando el nivel de ruido de la tabla anterior y después el resultado que me dio lo dividí por 7 y me dio 45".
- *Un valor mayor al número de personas dado:* "El nivel del ruido para el salón que tiene 25 personas sería 38 el nivel del ruido, ya que sería algo coherente para predecir para un salón de 25 personas y sería coherente ya que sería un valor más alto que 25, es decir más mayor".
- *Observando los vecinos y aproximando:* "Tendría en cuenta la cantidad de personas que hay en los otros salones y el nivel de ruido y teniendo en cuenta eso yo ubicaría el punto en

Z=(25,56). Lógicamente teniendo en cuenta las posiciones de los puntos con 21 y 29 personas".

Como se observa, se identifican tres formas: la media, ya sea del conjunto de todos los datos, o del conjunto de datos de la variable dependiente; un valor mayor al valor de la variable dependiente y, por último, observando lo que pasa alrededor del valor que se quiere predecir. Esta última forma refleja, de manera implícita, una consideración de la relación entre las dos variables, al observar de qué manera se están comportando los valores vecinos.

Predicción para un salón con 28 personas

✓ Con base en los mismos datos de la Tabla 24, para predecir el nivel de ruido en un salón con 28 personas cambió su estrategia: "El nivel del ruido para 28 personas sería 78 porque debe ser un valor menor que el nivel del ruido en promedio de los tres valores de 29", es decir, tomó el vecino más cercano, que correspondía a la pila de 29 y promedió. Y, dado que conocía la relación entre las dos variables, entonces su predicción fue *un valor menor al promedio que le dio en 29*, puesto que el número de personas era menor.

Otras formas:

- *Sumando las dos columnas y dividiendo en el próximo dato, ocho:* "El nivel de ruido para un salón que tiene 28 personas es 57, calculé ese resultado sumando el número de personas que me da como resultado 117, después sumé el nivel de ruido que me dio 343 y esos dos resultados los sumé, después los dividí por 8 y me dio 57". El estudiante sabe que debería tener en cuenta de alguna manera las dos variables, por lo que involucra en su cálculo todos los datos.

▪ *Sumar los datos de la pila más cercana y dividirlo entre el número de personas del que tenía que predecir:* "Sumar los datos sobre el ruido de los salones con 29 personas y dividirlos en 28. Cuando lo hice en el gráfico me di cuenta de que estaba mal el resultado, no es coherente". Como se lee en este argumento, la estudiante valida su predicción observándolo en el gráfico, pareciéndole poco coherente.

▪ *Observando los vecinos y aproximando:* "Tendría en cuenta la ubicación del nivel de ruido de los salones con 21, 25 y 29 y podría dar un valor que sería 69 el nivel de ruido W = (28,69). Teniendo en cuenta la cantidad de personas que hay en el salón".

Es decir, hasta el momento, de los siete estudiantes, solo dos abandonan los cálculos para utilizar como estrategia la aproximación de acuerdo a lo que observan en los datos y a la relación entre las dos variables. Aunque una de ellas, como hemos visto, calcula la media de las pilas, finalmente su predicción es una aproximación teniendo en cuenta la media (o medias) calculada.

Predicción sobre una pila de datos

✓ En seguida, se adicionaba una pila de datos, la de 25 personas (ver Figura 42). Esas tres pilas se observaban en un diagrama de dispersión. En primera instancia, se pedía predecir el nivel de ruido para un salón con 25 personas: "El nivel del ruido para un salón que tiene 25 personas sería 55 porque este valor está entre los valores anteriores. 64+58+52+47=221/4=55".

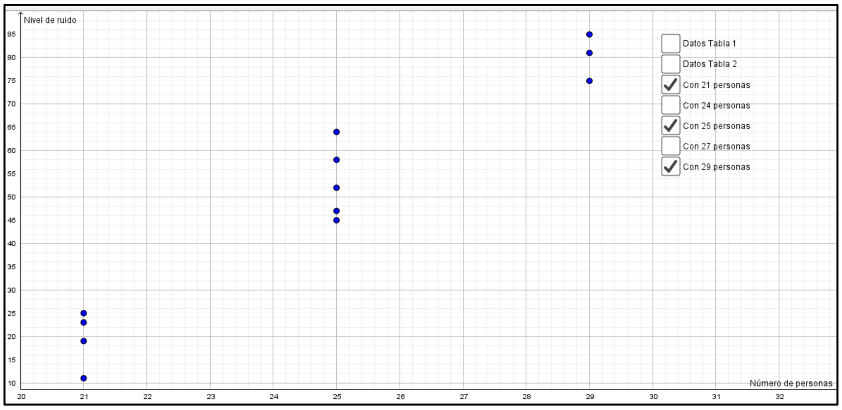


Figura 42. Gráfico asociado a la Actividad 4 (predicción dados tres pilas de datos)

Es decir, para una pila de datos, sacó *la media de los datos* y esta fue su predicción, tal como lo habíamos predicho de acuerdo a discusiones anteriores en otras actividades; de la misma manera procedieron tres estudiantes más. Más adelante veremos cómo lo haría entre pilas de datos.

Otras formas:

- *El promedio de todos los datos en Nivel de ruido:* “El nivel de ruido para un salón que tiene 25 personas es 48. Calculé este resultado sumando el nivel de ruido de los salones 21, 25 y 29 después los dividí por 12 el número de salones y me dio 48. Es coherente ese resultado teniendo en cuenta los anteriores niveles de ruido. C=(25,45), D=(25,47), O=(25,52), T=(25,58), L=(25,64)”.
- *Intervalo aproximado tomando como referencia a un vecino:* “Aproximadamente: 15 a 38 de nivel de ruido. De acuerdo a los valores de 21, donde algunos están de 11 a 25. Porque creo que estaría entre esos valores”.
- *Una aproximación de acuerdo a los valores de la misma pila de datos:* “Tendría en cuenta los valores de los otros salones con 25 personas y pondría un valor que se encuentre

entre 64 y 46 en nivel de ruido, pondría en 61 para el nivel de ruido del otro salón con 25 personas. Z=(25,61)”.

El ejercicio de predecir el nivel de ruido para un valor en donde se tenía una pila de datos (25 personas) se reducía a un análisis univariado. Como vimos, la estudiante que seguimos dio cuenta de ello al fijar su atención en los salones que solo tenían 25 personas, siendo la media su predicción. Otros estudiantes también fijaron su atención en la pila de 25 personas aunque difiriendo en su estrategia: Aproximaciones e intervalos. Respecto a dar un intervalo como una predicción, podríamos pensar que existe cierta reserva para dar una predicción cuando se percibe la aleatoriedad que existe en el ruido para la misma cantidad de personas. Esta actitud de los estudiantes evidencia la necesidad de una forma de control de la variabilidad como es esa: dar un intervalo de posibles valores y no un solo valor.

Predicción entre pilas de datos

Se pedía predecir el nivel de ruido para un salón con 23 y 27 personas, que se ubican en la mitad de 21 y 25, y 25 y 29, respectivamente.

- ✓ Para 23: “El nivel del ruido para un salón que tiene 23 personas sería 49 porque si en un salón con 25 personas el nivel del ruido es 55 para el salón de 23 personas sería un valor menor que 55 entonces 49 es un valor coherente”. Es decir, procedió de la misma manera que cuando predijo el nivel de ruido para un salón con 28 personas, *observando la media de un vecino cercano*, para 25 personas, *y teniendo en cuenta la relación* entre las dos variables.
- ✓ Para 27: “El nivel del ruido para un salón que tiene 27 personas sería 70 porque si en un salón con 29 personas el nivel del ruido es 79 y en un salón que tiene 25 personas el

nivel del ruido es 55, entonces 70 sería un valor que está entre 29 y 79. En este caso, la estudiante observó que 27 se encontraba entre dos pilas de datos, de manera que su predicción fue *un valor entre la media de cada pila de datos*, la de 25 y la de 29, pero no necesariamente la semisuma de las medias.

Otras formas:

- *Sacando la media de los valores de Nivel de ruido de las pilas entre las que se encontraba 23 o 27:* “El nivel de ruido para un salón que tiene 27 personas es 63. Este resultado lo encontré teniendo en cuenta el nivel de ruido de los salones 25 y 29. Sumé el nivel de ruido de los salones 25 y 29. Sumé el nivel de ruido de los dos salones y ese resultado lo dividí en 8 y me dio 63. No utilicé el salón de número de estudiantes 21 personas porque no era coherente al resultado que me dio”.
- *Intervalo aproximado tomando como referencia a un vecino:* Para 23, por ejemplo, “Aproximadamente: 15 a 45 según los valores dados anteriormente sobre un salón con 25 personas”.
- *Un valor aproximado de acuerdo a la media de una de las pilas de datos cercana:* Para 23, por ejemplo, “11+19+23+25=78/4=19 observé en la vista algebraica la lista con 21 observé los valores los sumé y dividí por la cantidad que es 4, lo dividí y ese resultado sería aproximadamente el valor del salón. Que el punto no sea menor que 21 y que no sea mayor que 25 porque entre más personas mayor ruido, entre menos personas es menor el ruido lo ubiqué en (23, 31) de acuerdo a lo anterior”.

- *Teniendo en cuenta los datos de las tres pilas:* Para 27, por ejemplo, "Como he estado haciendo en los casos anteriores tendría en cuenta la ubicación de los valores de los otros salones y lo pondría en 67 para el nivel de ruido para un salón con 27 personas. W= (27,67)".

Como se observa, a lo largo de las predicciones, desde los datos mostrados en tablas, hasta los datos observados en diagramas de dispersión, los razonamientos de algunos de los estudiantes fueron evolucionando. En un principio los estudiantes no son muy razonables en los cálculos que realizan, por ejemplo cuando calculan la media de todo el conjunto de datos, reflejando así que algunas ideas, como la relación entre las dos variables, no han terminado de estructurarse adecuadamente. Sin embargo después evalúan la razonabilidad que tiene el valor encontrado frente a los valores que ya se poseen, con expresiones como: "No es coherente el resultado que me dio".

Una predicción de forma general

¿Cómo sería una forma general para predecir el nivel de ruido para cualquier salón en donde hay entre 21 y 29 personas?

- ✓ Pues bien, para esta estudiante una forma general sería como la que enuncia: "El nivel del ruido para 21 sería 19 y para 29 sería 79 porque es un valor coherente. En donde elegí el nivel del ruido de los salones con 21 personas y saqué la media y con el salón de 29 personas hice lo mismo. El nivel del ruido para los demás salones entre 21 y 29 personas sería un valor que está entre 19 y 79 el nivel del ruido" Es decir, toma la media de las dos pilas y sus predicciones estarán entre esos dos valores, pero siempre teniendo en cuenta la relación entre las dos variables como se observa en la Tabla 25 que corresponde a sus predicciones. De igual manera procede otro estudiante.

Tabla 25.

Predicciones, de la estudiante, del nivel de ruido para salones entre 21 y 29 personas

Número de personas	Nivel de ruido	Argumento
21	19	Al completar la tabla tuve en cuenta la media del nivel de los salones que tienen 21 personas y luego lo mismo con los salones que tienen 29 personas y busqué valores que estuvieran entre 19 y 80.
22	25	
23	49	
24	52	
25	55	
26	67	
27	70	
28	75	
29	80	

Otras formas:

• *Para un valor entre dos pilas, calculando la media de las dos pilas y luego la semisuma de las dos medias:* "Lo que quería en este punto era sumar y dividir por dos, hallar la media de todos los salones y los dividía en dos, así me pareció como una forma para hallar el término que me preguntaban".

Por ejemplo, si se tenía una pila de datos para 21 y otra para 24, entonces saca la media para cada una de las pilas, suma y divide en dos y esa sería su predicción para 22, luego, tomando en cuenta este último valor, lo suma con la media que le dio en 24, divide en dos y ese nuevo valor sería su predicción para 23 y así sucesivamente sigue operando para predecir el nivel de ruido en los otros salones.

• *Teniendo en cuenta la relación entre las dos variables y aproximando:* "Se tendría en cuenta el número de personas que hay en cada salón y así podría predecir el valor para el nivel de ruido de ese salón. Porque entre más personas haya mayor es el nivel de ruido y entre menos personas, menor es el nivel de ruido. Por qué creo que son buenas, porque las variables tienen una relación ascendente y esta nos puede afirmar que si una de las variables sube la otra también tiende a subir porque existe variabilidad”.

Este estudiante asume que la causa de la tendencia de la curva es la existencia de variabilidad, como una forma de explicitar el hecho de que la tendencia no es constante, es decir, una forma de decir que la variabilidad en los valores de la variable respuesta se podrían explicar por cambios en la variable explicativa, en otras palabras, existe asociación entre las dos variables.

Independientemente de la forma general que los siete estudiantes propusieron, al momento de llenar la tabla todos tuvieron en cuenta la relación que había entre las dos variables (ver Tabla 26), una “relación ascendente”.

Tabla 26.
Predicciones de los otros seis estudiantes del nivel de ruido para salones entre 21 y 29 personas

		Predicciones de los otros seis estudiantes					
		D	**E**	**MF**	**M**	**S**	**W**
Número de personas	**21**	23	19,5	22	19	29	25
	22	38	30,625	27	31	30	31
	23	42	36,35	34	40	39	34
	24	53	41,75	38	42	41	52
	25	61	53,2	42	53	45	61
	26	72	59,85	48	54	50	64
	27	85	66,5	56	66	55	67
	28	92	73,415	68	72	60	69
	29	97	80,33	74	80	66	75

Hechas sus predicciones, se les pedía que las observaran en un gráfico de dispersión y que decidieran si cambiaban o no sus predicciones. El gráfico de dispersión de la estudiante que seguimos se observa en la Figura 43. Los de los otros seis estudiantes se resumen en la Figura 44.

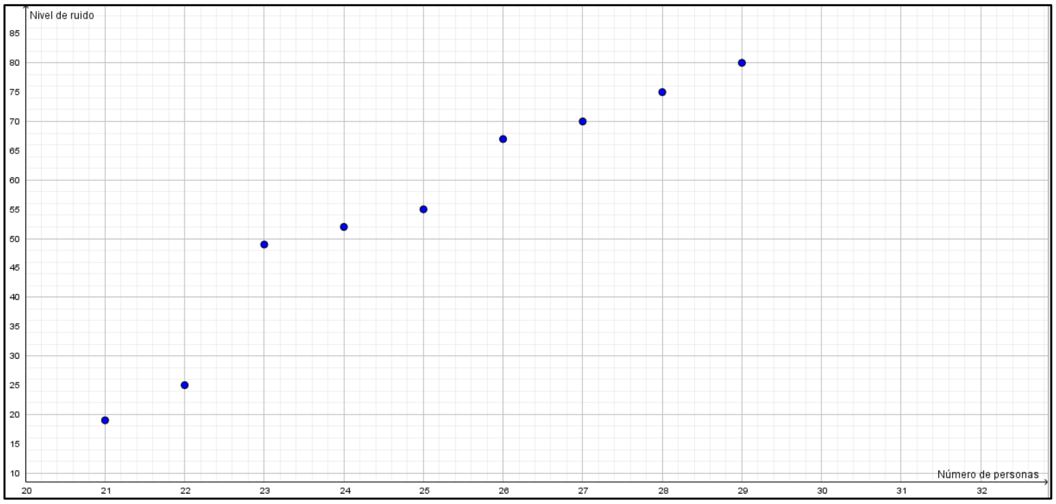


Figura 43. Diagrama de dispersión asociado a la Tabla 25

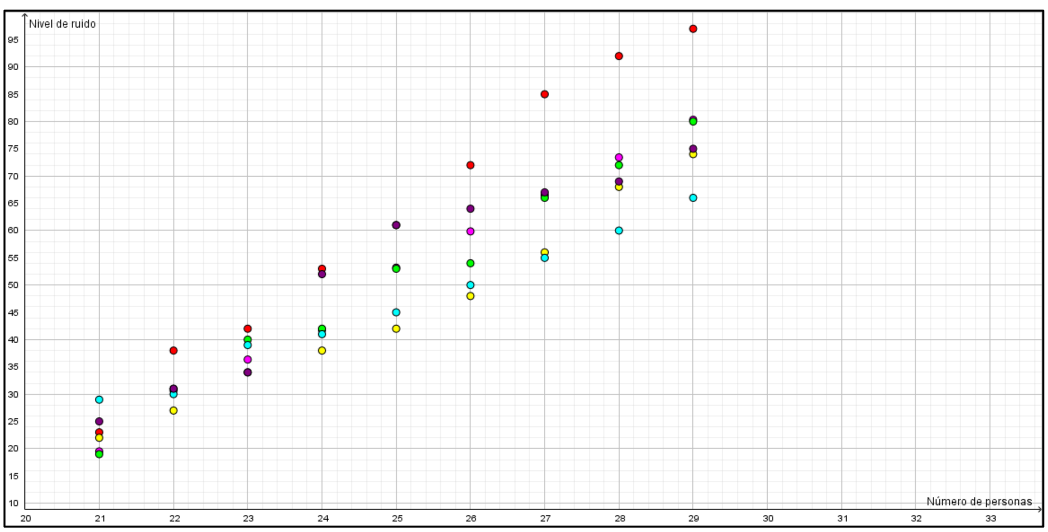


Figura 44. Diagrama de dispersión asociado a la Tabla 26

Como se aprecia en la Figura 43 y en la Figura 44 (los colores se relacionan con la Tabla 26), las predicciones de los estudiantes respetan la relación entre las dos variables, tal como uno de ellos lo justifica para decidir no cambiarlos de posición: “No, porque son valores que cumplen una relación porque entre más personas haya el nivel del ruido es mayor, se cumple una relación ascendente”.

Finalmente, para que los estudiantes validaran sus respuestas, se les pidió exponer las ventajas y desventajas de la forma general que habían propuesto, es decir, por qué lo que ellos proponían era válido: “Porque se tiene en cuenta la relación en donde el número de personas se relaciona con el nivel del ruido y es fácil por[que] se buscan valores coherentes que estén entre los valores. Este procedimiento tiene algunas desventajas. Las desventajas es que en algunos casos no podrían estar bien las predicciones porque podría ser que los valores que busco son varios y debo elegir uno solo que se aproxime o esté en los valores y en algunos casos el valor tal vez no sea el apropiado”.

Es decir, una de las ventajas es que su forma general conserva la relación entre las dos variables y entre las desventajas está su preocupación al tener que escoger un solo valor para cada número de personas, y que éste no sea el mejor, manifestando así un sentido de variabilidad.

A continuación, en la Tabla 27, se muestran las ventajas y desventajas de las otras dos formas generales expuestas anteriormente.

Tabla 27.

Ventajas y desventajas que los estudiantes reconocen en sus formas generales de predicción con datos apilados

Forma general	Ventajas	Desventajas
Para un valor entre dos pilas, calculando la media de las dos pilas y luego la semisuma de las dos medias	La ventaja es que para un caso de que faltan valores se podrá hacer la misma operación.	Pero las desventajas es que para algunos casos sirve y para otros no, no siempre va a dar un valor preciso ni un valor exacto pero siempre va haber variabilidad eso es una desventaja.
Teniendo en cuenta la relación entre las dos variables y aproximando:	Podemos dar una respuesta a lo que se nos pregunta.	No podemos dar un valor preciso. Puede casi siempre no funcionar.

El ejercicio de predecir el nivel de ruido para salones entre 21 y 29 personas se discutió entre todos. Como se mostró en una discusión anterior alrededor de la Figura 39, se acordó que, entre las predicciones que ellos habían hecho para un salón con 28 personas, tomaríamos la media para así quedarnos con uno (ver Figura 45, punto amarillo).

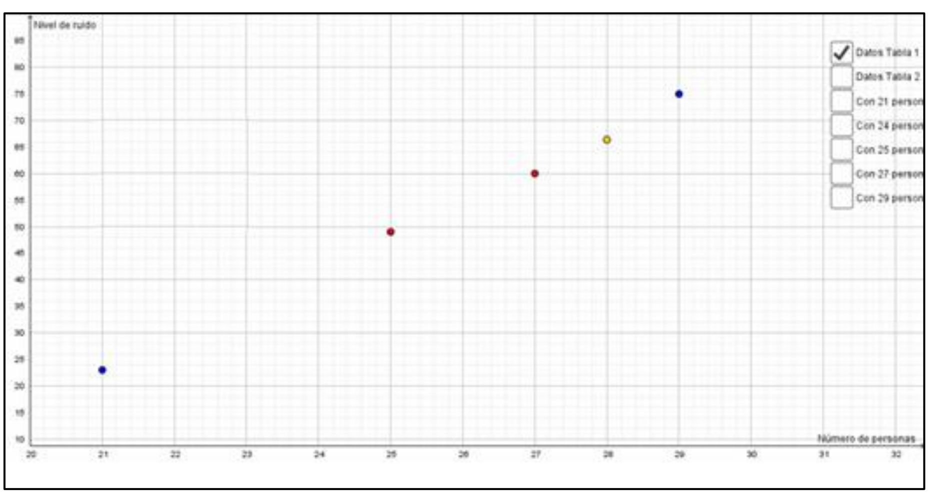


Figura 45. Diagrama de dispersión asociado a la Figura 39

De ahí en adelante cada uno de ellos fue prediciendo para cada uno de los salones. En la siguiente discusión se muestra cómo fuimos más allá, en cuanto a forma, por ejemplo.

I: Ahora sí ¿cómo se ve? [Figura 45] ¿Sí les suena o no les suena?

Todos: Síiii

A: Síi, tan bonito.

I: ¿Por qué se ve bonito?

M: ¡Cuál bonito, hala! Se ve ascendente.

Todos: (Risas)

I: Se ve ascendente y qué más.

E: Que se ve alineado profesora

M: Que está en recta.

I: ¿Cómo?

M: Que está en recta.

I: ¿Está casi en una recta? Elder dice que están alineados, ¿se ven alineados? ¿más o menos alineados?

E: Más o menos sí.

I: Yo puedo trazar una recta para ver si están alineados (ver Figura 46). No se ven alineados pero están casi alineados ¿cierto?

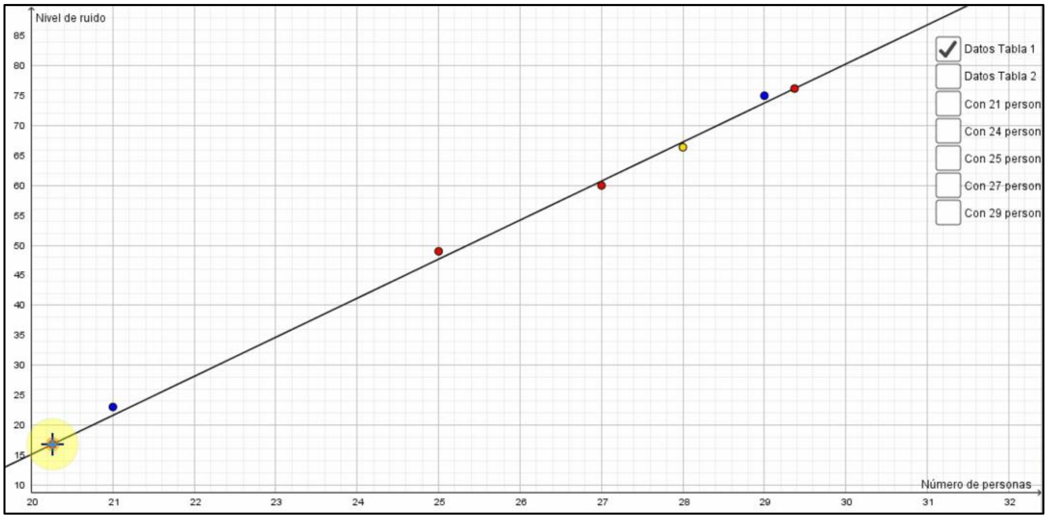


Figura 46. Recta asociada al diagrama de dispersión de la Figura 45

I: Es decir [quitando la recta] si yo les pido a ustedes para 22, para 23, para 28… ¿ustedes cómo esperan que se vea ese gráfico?

Como los estudiantes estaban visualizando en un televisor lo que el investigador hacía desde su computador, entonces entre todos ubicaron los puntos restantes en el diagrama de dispersión (ver Figura 47).

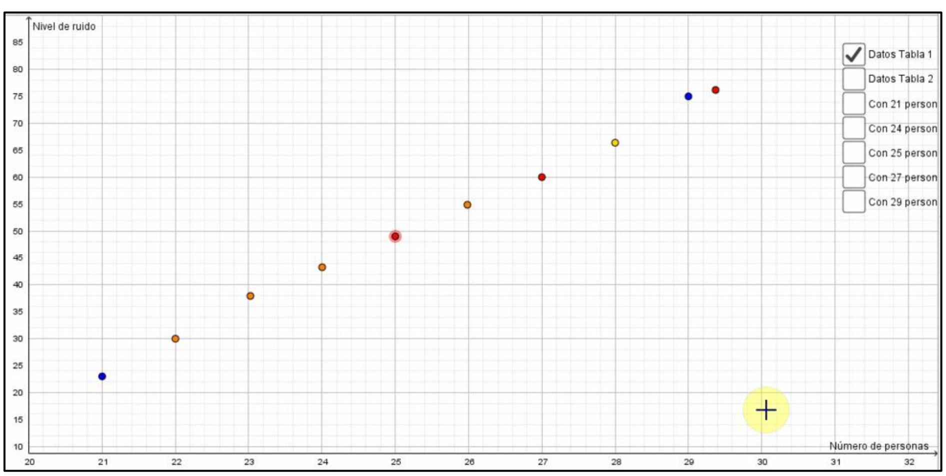


Figura 47. Predicciones de los estudiantes de manera grupal

¿Pero por qué ubicarlos así? ¿Qué buscaban o tenían en cuenta para ubicarlos de esa manera? A continuación una discusión alrededor de la Figura 47.

I: Yo sé que en un salón no pueden haber veintinueve personas y media, pero entonces ya no pensemos en número de personas, sino en cualquier variable x ¿listo? Es decir, ese valor x puede tomar cualquier número real. Miguel, entonces si x vale 28,5 más o menos cuánto valdría y siguiendo lo que estamos viendo.

I: [Entonces se empieza a mover el punto y él indica si más arriba o más abajo… inclusive proponen mover el último punto un poco más arriba, como se observa en la Figura 48]

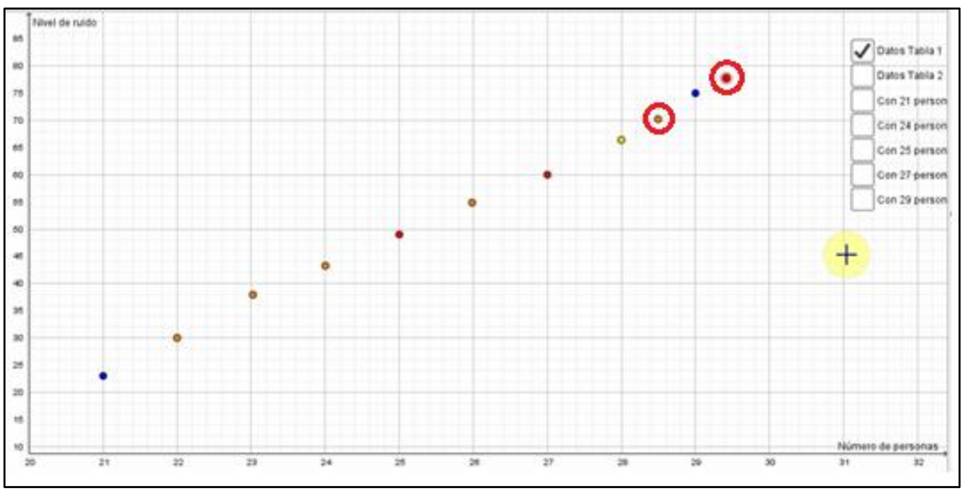


Figura 48. Gráfico de dispersión asociado al de la Figura 47

I: ¿Qué están buscando ustedes?

M y E: La recta (Risas)

I: ¿Que se vea una qué?

M: Una recta

I: ¿Una recta? [Y se dibuja una recta como en la Figura 49]

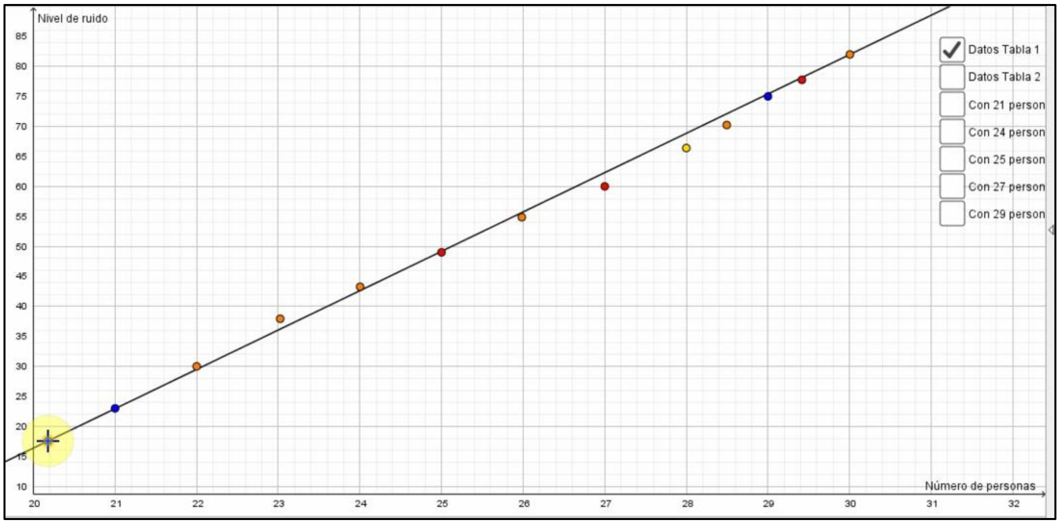


Figura 49. Recta asociada al gráfico de dispersión de la Figura 48

I: ¿Y ustedes por qué están buscando la recta?

A: Porque es más bonito

I: Es decir ¿podría ser una recta perfecta?

A: (Niega)

I: ¿Por qué?, Andrea.

A: Porque por ejemplo… el 23 no queda bien ahí ¿no? [Ver Figura 49] Entonces podría ser porque…

I: Porque qué… Ayúdenle a Andrea, ¿por qué no una recta perfecta?

A: Porque puede variar el nivel de ruido en 23 personas.

I: Es decir, siempre va haber…

A: Variabilidad

I: Siempre va haber variabilidad, no estamos en un mundo perfecto ¿no? No podemos decir que si para 21 es 23 y para 22 es 30… aumentó cuánto.

M: Siete

I: Entonces no podemos decir que para 23 va tener cuánto… si de 21 a 23 aumentó 7 ¿entonces para 23 será 37?

M y A: No

I: Hay variabilidad. No podemos decir que exactamente 27 pero sí algo parecido, miren, 37,93. ¿Cierto?

A: Por eso es que una desventaja puede ser que no siempre uno va a dar un valor perfecto porque siempre hay variabilidad.

I: Ajá

Hasta el momento, aunque surgió el tema de la recta, al quitarla e indagar o hablar en términos de la "forma" que reflejaban los puntos, no asumieron la recta como esa "forma" sino que pensaron en una figura geométrica. Es decir, aunque una estudiante sí asume la recta como una "forma" su idea pasa desapercibida al no ser aceptada por los demás estudiantes, como se muestra a continuación.

I: ¿Y acá [mostrando solo los puntos, sin la recta] se refleja la relación que hay entre las dos variables?

A y M: Sí

I: Y en cuanto a forma cómo se refleja esa relación ascendente. En cuanto a forma a qué se parece eso.

M: [Susurra que a un triángulo]

M: ¿Si hacemos una forma?

I: Sí, si hacen una forma cuál sería.

A: Una recta.

M: No pero una forma…

I: ¿Una recta podría ser?

M: No porque va…

A: ¿pero si uno traza así también se puede tener en cuenta esto así? [abandona la idea de recta y ahora se refiere también a los ejes]

I: Cómo así

M: O sea, por ejemplo profesora trazando así [una recta a través de los puntos] se puede tener en cuenta así y así [rectas paralelas a los ejes].

I: Aaaah… para que se parezca a qué…

A: Un triángulo

I: Es decir, algo así… [Ver Figura 50]

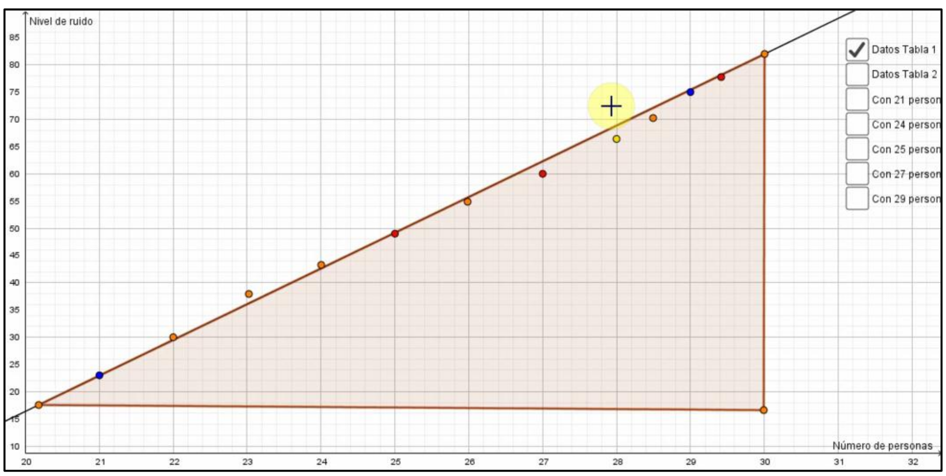


Figura 50. Forma asociada al gráfico de dispersión de la Figura 48

M, A y E: Sí

I: Pero bueno, yo acá [en los catetos del triángulo] no tengo nada. Solo esto [la hipotenusa] acá qué parece, me interesa donde están los puntos. Qué parece, qué forma tiene.

A: Una recta [retomando su idea inicial]

M: No…

I: ¿Por qué no, Miguel?

M: Porque no se ven así, que todos vayan por la línea.

I: Pues no, pero es decir esto acá a qué se parece [dibujando un cuadrado no tan cuadrado]

W: Un cuadrado

I: Un cuadrado ¿pero es un cuadrado perfecto?

Todos: No

I: No, así yo lo haga con mucho pulso [dibujando otro mejor]… ¿es un cuadrado perfecto?

M y E: No (Risas)

I: O sea, la perfección solamente está acá [señalando mi sien] Yo sé que un cuadrado es perfecto si qué…

W: Si tiene todos sus lados iguales…

I: Ajá y si tiene todos sus ángulos…

W: Iguales

I: Iguales y además son qué…

W: Rectos.

I: Eso es lo que yo tengo acá [señalando mi sien] pero por más que yo lo haga con regla y compás, siempre va haber…

M: Un defecto.

I: Entonces solo los puntos qué forma tienen… entonces cuando ustedes estaban ubicando los puntos, que me decían “más arriba” “más abajo” ustedes qué estaban buscando.

M: Una recta

I: ¿Todos estaban buscando una recta?

A y M: Síi

Y, asumiendo la recta, se indagó sobre por qué la recta de esa manera o en esa dirección y no en otra, esta vez volviendo a datos apilados.

I: ¿Cómo me refleja esa recta la relación que hay entre las dos variables? Mafer…

MF: Se ve bien

I: ¿Cómo es que se ve bien? Porque es que miren [mostrando en la pantalla otra recta con una relación opuesta a la que tenían los puntos, ver Figura 51], esto también es una recta…

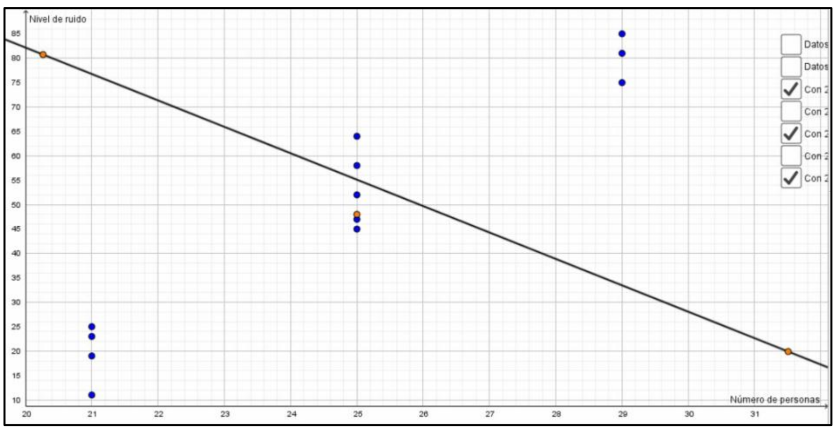


Figura 51. Recta propuesta en una discusión para un diagrama con datos apilados

A: Aaaaaah, porque la recta está ascendente.

I: Aaaaaah… Bueno y si las dos variables tuvieran una relación descendente, Silvia… ¿Cómo creería usted que iría eso, usted qué buscaría? [Se mostraron varias en la pantalla hasta que acertaron con una]

I: ¿Y cómo sería una recta que no es ascendente ni descendente?

MF: Horizontal

De manera que se mostró cómo irían las rectas en variables que tuvieran una relación ascendente (pendiente positiva), descendente (pendiente negativa) o que no tuvieran una relación (pendiente cero). Cabe recordar que estos estudiantes tenían la noción de recta pero, hasta este momento de su formación, no habían visto nada como pendiente o ecuación de una recta, sin embargo lo anterior no fue un impedimento para hablar de la misma y de su relación con los puntos en un diagrama de dispersión.

Claro, no negamos el hecho de que nuestra THA tomaría otro rumbo si elementos como pendiente y ecuación de una recta fueran conocidos por los estudiantes. Sin embargo lo que hicimos acá podría constituir un trabajo previo a lo que ellos verán más adelante, en su formación académica, cuando vean puntos completamente alineados y que pueden ser expresados, matemática y perfectamente, por una ecuación. Lo anterior lo discute Casey (2015, p.24) al dejar una reflexión abierta sobre cómo las rectas de mejor ajuste deberían ser integradas con las rectas que se ven en matemáticas, además de que cuál sería la mejor secuencia de enseñanza, es decir ¿qué enseñar primero, rectas de mejor ajuste, en un sentido estadístico o si rectas en un sentido matemático, determinista?

Como se observó, el desarrollo y discusión de esta actividad superó nuestras expectativas ya que nos permitió ir más allá y nos abrió las puertas para empezar a hablar en términos de rectas. Es decir debemos reconocer que en realidad, luego de esto, nuestra THA estaba encaminada hacia las nubes de puntos, ya no con datos apilados, y hablar en términos del engrosamiento de la nube de puntos para concluir que la nube más delgada, con una correlación muy cerca de 1 o -1, se asemejaría a una recta, pero como se observó nuestra THA tomó otra dirección para pasar a hablar directamente de recta, de la mejor, como se verá en las siguientes actividades.

4.9 Actividad 5

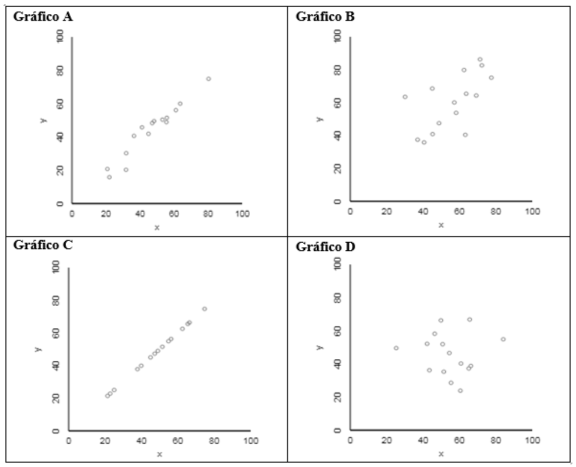
Continuando con nuestro eje temático, Predicción, en esta actividad, a diferencia de la anterior, las predicciones se hicieron inmediatamente sobre diagramas de dispersión y no sobre tablas. Lo anterior debido a que queríamos que los estudiantes centraran su atención en la ubicación de los puntos y que abandonaran la estrategia de hacer cálculos que, como vimos en algunos casos, resultaban pocos razonables. Adicional a lo anterior, los diagramas no estaban contextualizados para evitarnos concepciones de tipo Causal (Estepa y Batanero, 1995), por ejemplo. Sumado todo lo anterior, esta actividad nos permitiría observar sus formas de razonamiento de una manera más limpia o pura.

Por otro lado, esta actividad no se soportó en ningún software y el objetivo final era que, nuevamente, pero ahora sobre un diagrama de dispersión, encontraran una forma general que les permitiera predecir un valor de y para cualquier valor que tomara x. Es decir, como ejercicios final, se les motivó a que, tal como lo hacíamos en el caso univariado donde podíamos reducir un conjunto de datos a un solo valor, la media, con el cual podíamos predecir, encontraran una *forma* para reducir una nube de puntos y que esa *forma* diera cuenta de dos cosas: i) la relación que se observaba entre las dos variables y ii) que les facilitara el ejercicio de predicción.

Análisis

Ya en la actividad pasada vimos cómo los estudiantes muestran una coordinación entre la forma como van los puntos, si van “subiendo” o “bajando”, con la relación entre las dos variables, “ascendente” o “descendente”, respectivamente”. ¿Pero cómo reconocen que esa relación se refleja más en unos gráficos que en otros? Pues bien, ese fue el primer ejercicio en esta actividad, como se enuncia en la Tabla 28.

Tabla 28.
Tabla asociada al ítem 1 de la Actividad 5



Lee la siguiente afirmación y responde: *"En los gráficos A, B, C y D, se podría decir que si la variable x aumenta la variable y aumenta"* ¿En cuál de los cuatro gráficos podrías dar esa afirmación con mayor seguridad y en cuál con menor seguridad? Explica tu respuesta.

Cinco estudiantes escogieron el gráfico C, uno el gráfico A y otro los gráficos A y C. Estos dos últimos estudiantes eligieron A en primera instancia, pero luego se inclinaron también por el gráfico C, sin embargo mostramos sus criterios de elección iniciales. Por otra parte, de manera unánime, todos concluyeron que con el gráfico D esa afirmación se podía dar con menor seguridad. A continuación se destacan algunos criterios de elección.

- *Los puntos se asemejan a una recta o se ven rectos:* Dos estudiantes argumentaron en términos de recta, por ejemplo: "En el cuadro que se refleja más es el C y en el que menos se refleja es el D porque nos están diciendo que X y Y son una pareja entonces si uno de los dos aumenta el compañero también debe aumentar y yo digo que es C porque es el cuadro que está más reflejada una recta que el en gráfico D porque ahí están como si los puntos los hubieran ubicado al azar sin tener en cuenta la relación entre las dos variables".

- *Los puntos se ven ordenados:* "En el gráfico C porque a medida que va aumentando los valores de X en el gráfico se va representando mejor en forma ascendente. Y en el gráfico D es el que menor seguridad podemos tener porque no tiene como un orden comparándolo con el gráfico C. Comparando al gráfico C al aumentar X va aumentando Y y en el gráfico D no se muestra igual".

- *Variabilidad:* Un estudiante, en primer momento, se inclinó por el gráfico A argumentando de la siguiente manera: "Con mayor seguridad el gráfico "A" porque es como la afirmación más segura que me ha parecido, con menor seguridad es el gráfico D ya que el que no cumple con las condiciones dadas, es de mayor seguridad es el gráfico A porque tiene variabilidad y tiene que ver que si x aumenta "Y" entonces cumpliría el requisito que me piden, el de menos seguridad es el gráfico D no me pareció porque si "X" aumenta la variable "Y" debería que aumentar y hay unos puntos que no cumple la condición por iniciar muy alto y no debería iniciar tan alto, debería que iniciar el punto "O" [se refiere al origen] la coordenada muy baja para que fuera subiendo los puntos a lo que se alejan las variables de 0".

Es decir, este estudiante considera que el gráfico C es demasiado "perfecto" como para que se dé y que, debido a la variabilidad, siempre presente, algo como C no debería ocurrir. Sin embargo, luego cambió su respuesta por C, al enfrentarse con el siguiente ejercicio que ya veremos en qué consistía. Es como si el tema de la variabilidad se hubiera convertido en un obstáculo para él.

• *De acuerdo al grado de dispersión:* “En el gráfico C porque se ve más claramente el valor que va aumentando de (X) y (Y). Con menor seguridad sería el gráfico con la menor es: El gráfico D, porque hay valores con la variable (X) aumentada pero con la variable (Y) disminuye al valor superior a (X) y están más dispersos en la gráfica de acuerdo a la afirmación”.

Como se observa, entre los criterios para decidir la intensidad de relación entre las dos variables, el de mayor frecuencia es que los datos reflejen un orden, como en el gráfico C en donde los puntos seguían un orden semejándose a una recta. Mientras que para decidir el de menor grado de intensidad de la relación se inclinan por el grado de dispersión que, comparándolos con el gráfico C, no siguen una relación: “como si los puntos los hubieran ubicado al azar”. Se observa también un hecho interesante cuando un estudiante deja de lado el gráfico C para escoger el gráfico A en donde se refleja cierto grado de variabilidad pero, aun así, se evidencia la relación entre las dos variables. Recordemos que, en actividades anteriores, cuando se pedía predecir sobre una pila de datos un criterio fue tomar un intervalo, de alguna manera acá está pasando lo mismo, pues el estudiante siente la necesidad de controlar la variabilidad al inclinarse por el gráfico de A en lugar de C.

Prediciendo sobre los cuatro diagramas de dispersión de la Tabla 28

En seguida se les pidió predecir, en cada uno de los cuatro gráficos, un valor de y si x era 72 y que indicaran cuál gráfico les permitía hacer una mejor predicción para cualquier valor de x.

En los cuatro gráficos las predicciones de los siete estudiantes estuvieron bastante razonables en el sentido en que, por lo menos en los gráficos A, B y C donde se refleja una asociación positiva, respetaron la relación que había entre las dos variables. Y bueno, en el gráfico D, donde la correlación era casi 0, la ubicación del punto era al azar. Solo una estudiante hizo predicciones de otro valor al que se le había pedido, sin embargo tuvo en cuenta la relación que observó entre las dos variables (ver Figura 52).

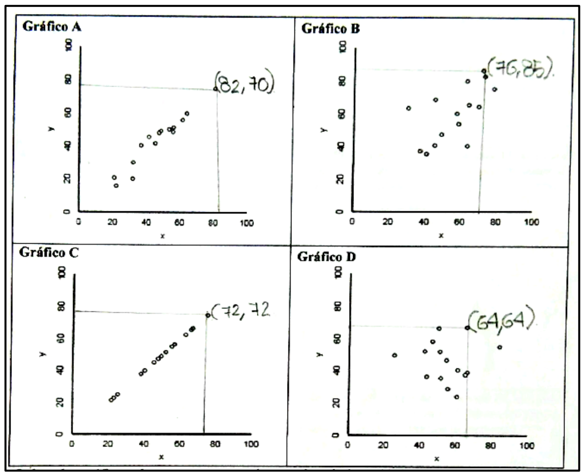


Figura 52. Predicciones de una estudiante sobre los diagramas de la Tabla 28

Haciendo énfasis en la predicción sobre el gráfico C donde la correlación era 1, es decir que, como los puntos se mostraban colineales, esperábamos que la predicción estuviera "en línea" con los demás puntos y, esto es, $y = 72$. Claro, como los estudiantes no contaban con tablas de datos, no era tan fácil hacer una predicción exacta, de manera que, aunque variando sus valores entre 70 y 73, cinco, de los siete estudiantes, ubicaron el punto como lo esperábamos, en línea con los demás puntos, donde dos, de esos cinco, predijeron que y era 72. En la Figura 53 se observa la predicción en el gráfico C de uno de ellos.

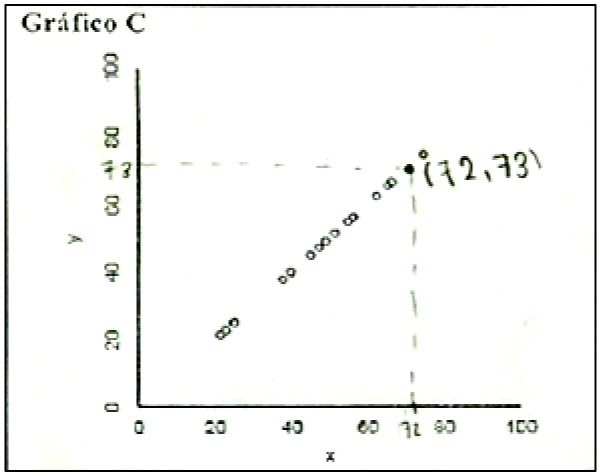


Figura 53. Predicción de uno de los estudiantes sobre el Gráfico C de la Figura 52

Las predicciones de los otros dos estudiantes fueron 75 y 79 y, como se observa en el gráfico de uno de ellos (ver Figura 54), no estaban preocupados por conservar esa “colinealidad”, pero sí en que se respetara la relación creciente entre las dos variables. Cabe destacar que estos dos estudiantes siempre manifestaron su afán por la variabilidad, es decir que algo tan perfecto, como poner el punto en línea con los demás, no podría darse, pero sí muy cerca.

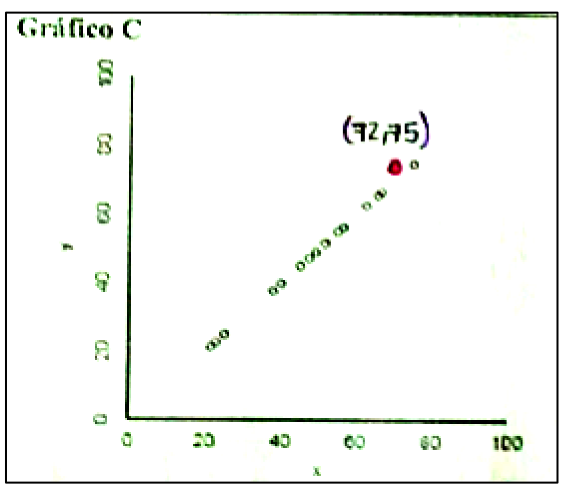


Figura 54. Predicción de uno de los estudiantes sobre el Gráfico C de la Figura 52

Entre los criterios para ubicar la predicción en los cuatro gráficos se destacan dos:

- *Orden u organización:* Cinco, de los siete estudiantes, argumentaron en términos de orden u organización, por ejemplo: "Para ubicar los puntos tuve en cuenta cómo iban los puntos o cómo iba su organización, es decir como seguía el orden de los puntos, y decidí colocar los valores que fueran coherentes es decir, según como estaban ordenados los puntos y que no fueran valores tan bajos".

Inclusive una estudiante, se animó a dar un orden en cuanto a la dificultad de hacer una predicción en los cuatro gráficos: "Para el gráfico A=(72, 66), B=(72, 66), C=(72, 73), D=(72, 77). Decidí hacerlo así porque a medida que aumenta X aumentaría Y. A diferencia de que el valor que me dan es el mismo para cada uno, entonces la variable Y aumenta a medida que la variable X aumenta. Pero solo está claro en el gráfico A y C. A diferencia que en el gráfico B y D los puntos están un poco desordenados… Los ordenaría así: C, A, B y D porque el gráfico D está más desordenado que el gráfico B".

- *La relación:* "Predije un valor de Y que fue 70 porque de acuerdo a la relación entre las dos variables 70 cumple una relación ascendente de acuerdo a los otros puntos".

Criterios de elección entre los cuatro diagramas de dispersión de la Tabla 28 para hacer una mejor predicción

Por otra parte, respecto a qué gráfico les permitiría hacer una mejor predicción de un valor de y para cualquier valor de x, en este caso entre 0 y 100, seis estudiantes acordaron con C y uno con A, teniendo en cuenta los criterios mencionados anteriormente.

- *Orden u organización:* Cinco, de los siete estudiantes, argumentaron en términos de orden u organización, por ejemplo: "En el gráfico C porque es el más ordenado y me ayuda a predecir mejor" o "El gráfico C ya que aquí se muestra que los valores están organizados de una manera ascendente y por eso de da que si un valor aumenta el otro también debe aumentar. Organizado en el sentido de que en los otros gráficos no se ve tanto una forma ascendente".

- *La relación:* "En el gráfico C porque se ve mejor la relación esto hace que permita una mejor predicción para cualquier valor y se dificultó en el gráfico D porque no [hay] una relación y es un poco difícil ubicar cualquier valor".

Podemos apreciar que dentro de los criterios de los estudiantes para decidir cuál de los gráficos les permite hacer una mejor predicción, uno de ellos se corresponde con los criterios para seleccionar el gráfico en donde se refleja mayor intensidad de relación entre las variables: *Orden u organización*, y en mayor frecuencia de elección.

A continuación mostramos una discusión que se hizo alrededor de estas primeras tareas de la Actividad 5:

I: Entonces Miguel, si tuviera que ordenar los cuatro gráficos [ver Tabla 28], de mayor a menor, es decir, donde se refleja más la relación a donde se refleja menos la relación, ¿cómo los ordenaría, Miguel?

M: C, A, B y D.

I: Los demás qué dicen, ¿están de acuerdo?

Todos: Sí

I: Si yo tuviera que predecir un valor de y para un valor de x cualquiera, Elder… ¿dónde sería más fácil predecir?

E: En el gráfico… en el C.

I: ¿En el C, sería más fácil? ¿Por qué?

E: Porque se ven los puntos ordenados.

I: ¿Y si tuviéramos que ordenar, en la dificultad de predicción, entonces cómo sería del más fácil al más difícil?

E: Lo mismo profesora, C, A, B, D.

I: Es decir, donde se refleje mejor la relación que hay entre las dos variables ¿va ser más fácil predecir?

E: Sí

I: Qué dicen los demás

Todos: [Asienten]

De manera que los estudiantes relacionan el grado “del orden” de los puntos, con la dificultad de predicción, o sea, que el hecho de que en un gráfico los puntos se ven más “ordenados”, facilita la predicción. Es decir, a mayor correlación entre las variables mejor la predicción.

También se hizo una pequeña discusión frente al gráfico D en la Tabla 28, donde no se percibía ninguna relación entre las dos variables, quisimos indagar sobre cómo predecir en ese caso. Veamos que asumen cualquier valor.

I: Elder, ¿en el gráfico D las dos variables están relacionadas o no?

E: Mmm… no.

I: No se refleja una relación, entonces si x vale 64, qué valor pudo haber tomado y.

E: 80

M: Bueno sí, porque dice que no hay variabilidad.

A: Sí hay variabilidad pero no hay una relación.

M: Ah sí.

I: Entonces si no hay una relación y puede tomar…

A: Cualquier valor.

Por lo que en un gráfico donde las variables no manifiesten ningún tipo de relación, se asume que, para un valor de la variable explicativa, la variable respuesta puede tomar cualquier valor. Es decir, en el caso de correlación cercana a 0, la predicción no se puede garantizar, en principio, cualquier resultado es posible.

Predicciones sobre diagramas de dispersión sin contexto

El siguiente ejercicio consistió en predecir sobre diagramas de dispersión, sin contexto, que iban aumentando su cantidad de puntos, no apilados, y variando la asociación entre las dos variables, a veces positiva y a veces negativa. Como ya lo dijimos al inicio del análisis de esta actividad, esta fue muy similar a la anterior difiriendo en la representación de los datos y en la ausencia de un contexto. Pero el objetivo final era el mismo: Encontrar una forma general que les permitiera predecir un valor de y para cualquier valor que tomara x.

Predicción con dos puntos

En la Figura 55, se pedía predecir un valor de y si x era 50. Observe que 50 no estaba justo en la mitad de 10 y 80.

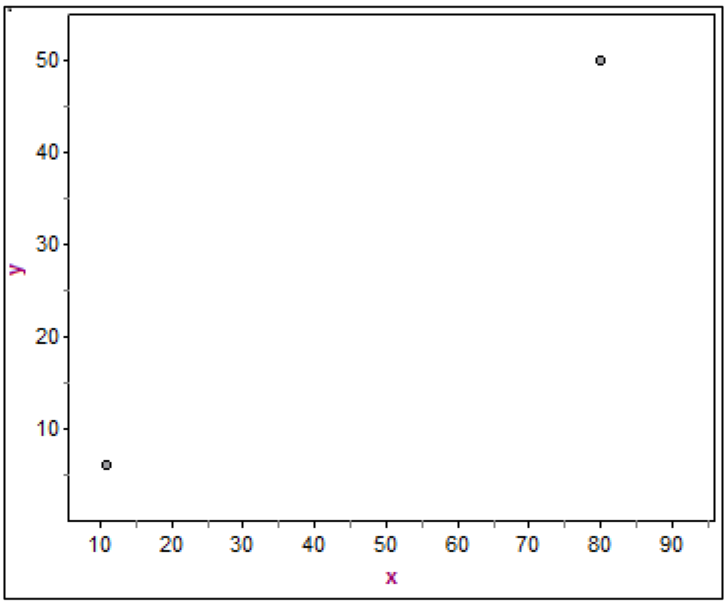


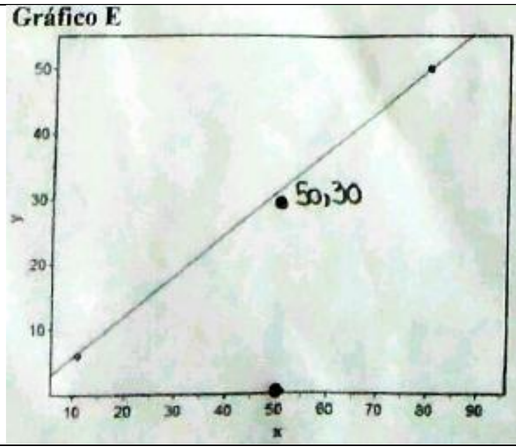
Figura 55. Diagrama asociado al ítem 2 de la Actividad 5. Predicción dados dos puntos

A continuación, algunas formas de razonamiento para predecir dados dos puntos.

- *En la mitad de los dos puntos* (ver Tabla 29, omita la recta):

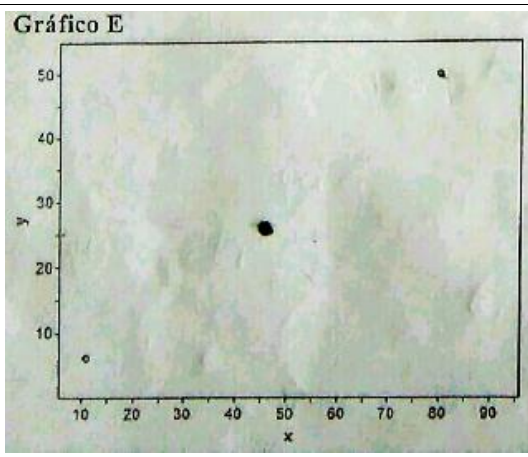
Tabla 29.
Tabla asociada a la forma de razonamiento de predicción "En la mitad de los dos puntos"

Predicción	**Argumento**

	Marqué el punto en el valor de 30 para Y por de acuerdo a los dos puntos que están en el gráfico E 30 sería como tomar un valor central entre los dos puntos.

- *Conservando la forma* (ver Tabla 30):

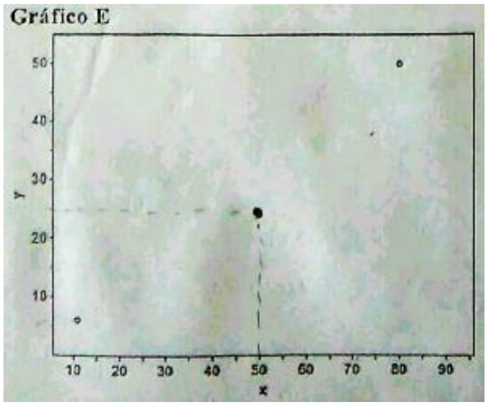
Tabla 30.
Tabla asociada a la forma de razonamiento de predicción "Conservando la forma"

Predicción	Argumento
	El valor de Y que predecí es 25. Coloqué ese valor para que así con los dos puntos que tengo ahí me ayude a hacer una forma ascendente.

- *Siendo y la mitad de x* (ver Tabla 31):

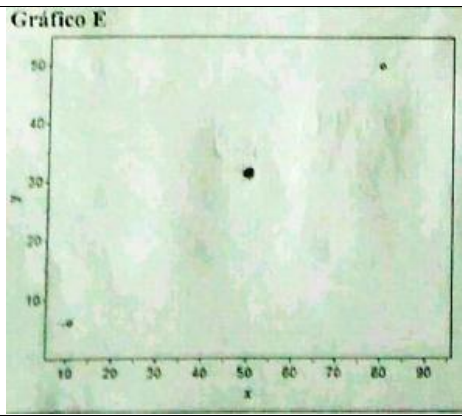
Tabla 31.
Tabla asociada a la forma de razonamiento de predicción "Siendo y la mitad de x"

Predicción	Argumento

	Para el gráfico C aproximadamente 25 porque tomé la mitad de 50 y ubiqué el punto.

- *Entre el mínimo valor y el máximo valor de y* (verTabla 32):

Tabla 32.
Tabla asociada a la forma de razonamiento de predicción "Entre el mínimo valor y el máximo valor de y"

Predicción	**Argumento**
	Predije este valor teniendo en cuenta que si uno de los valores está en (80, 50) y el otro está en (11, 12), entonces el valor que debo ubicar para (50, Y) es (50, 33) porque están entre los valores. Hay variabilidad. (50, 33)

Enfocándonos en la ubicación del punto de la Tabla 29 a la Tabla 32, independientemente de la forma de razonamiento que cada uno de ellos expresó haber utilizado, seis de los siete estudiantes lo ubican casi de manera colineal con los otros dos. Solo una estudiante no lo hizo así (ver Figura 56), aunque su argumento no se refleja del todo en su predicción: "(Y) Valdría 50 igual que (X)=50 porque tengo en cuenta que la variable (X) aumenta, y la variable (Y) también aumenta". Es como si hubiera olvidado el último punto cuyo valor x es 80.

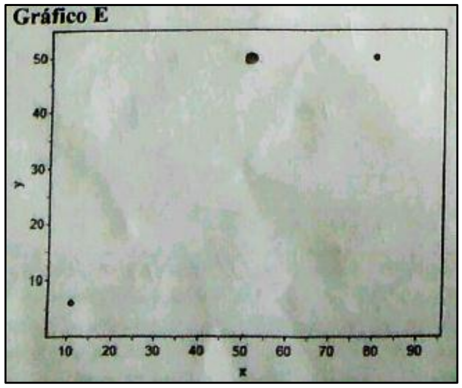


Figura 56. Predicción de una estudiante en un diagrama de dispersión con dos puntos

Predicción con tres puntos

En el siguiente diagrama de dispersión se tenían tres puntos que reflejaban una asociación negativa (ver Figura 57), y aunque ya se había discutido sobre la *forma* de los puntos cuando las variables tenían una "relación descendente", era la primera vez que predecían sobre un gráfico con estas características. En la Figura 57 se pedía predecir para 24 y para 28.

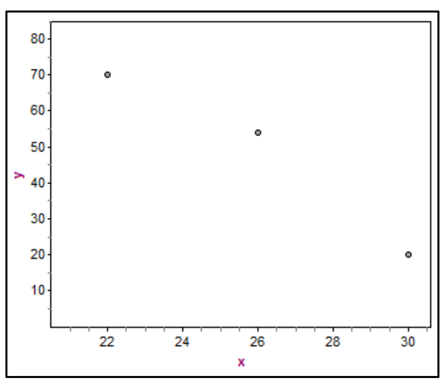


Figura 57. Diagrama asociado al ítem 2 de la Actividad 5. Predicción dados tres puntos y con una asociación negativa

Cuatro de los siete estudiantes ubicaron los puntos de manera que, globalmente, los puntos conservaran una asociación negativa, como en la Figura 58.

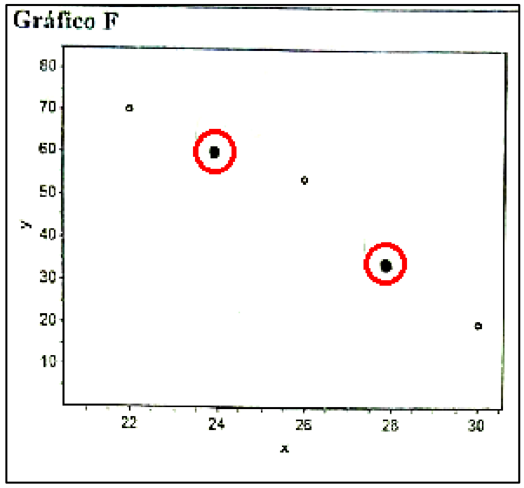


Figura 58. Predicción de un estudiante sobre el diagrama de dispersión asociado al de la Figura 57

Los otros tres estudiantes ubicaron los dos puntos donde se aprecia una asociación positiva, es decir, que localmente esos dos puntos evidencian una "relación ascendente" como en la Figura 59. Se podría inferir que este último hecho ocurrió debido a lo que dijimos anteriormente, era la primera vez que se enfrentaban con este tipo de predicción, no podemos negar del todo que reconocieran la asociación de los tres puntos dados inicialmente. Se podría pensar que hicieron caso omiso de la relación existente entre los puntos iniciales que estaban en el gráfico asumiendo que debían colocar los dos puntos en forma ascendente.

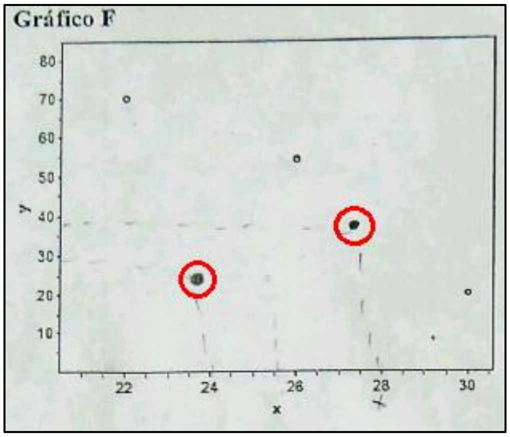
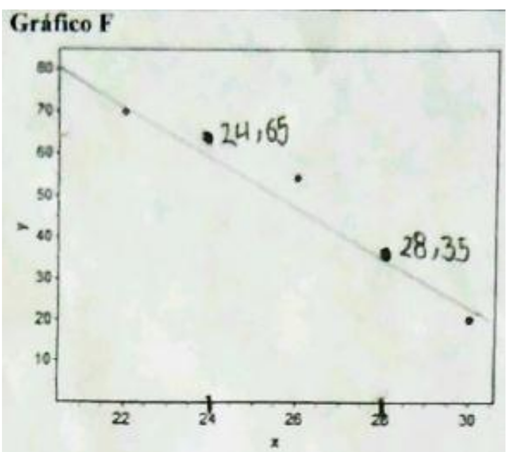


Figura 59. Predicción de un estudiante sobre el diagrama de dispersión asociado al de la Figura 57

Entre las formas de razonamiento del grupo de estudiantes que predijeron como en la Figura 58 se encuentran las siguientes:

- *En la mitad de los dos puntos* (ver Tabla 33, omitir la recta):

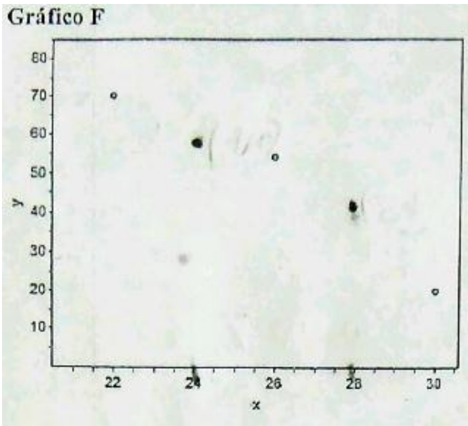
Tabla 33.
Tabla asociada a la forma de razonamiento de predicción "En la mitad de los dos puntos" dados tres puntos

Predicción	**Argumento**
	El valor de Y si X vale 24 aproximadamente sería 65 porque este punto está en medio de los dos puntos y seguiría la relación de una forma ascendente. Y el valor de Y si X vale 28 aproximadamente sería 35 porque se sigue cumpliendo la relación.

- *Conservando la forma* (ver Tabla 34):

Tabla 34.

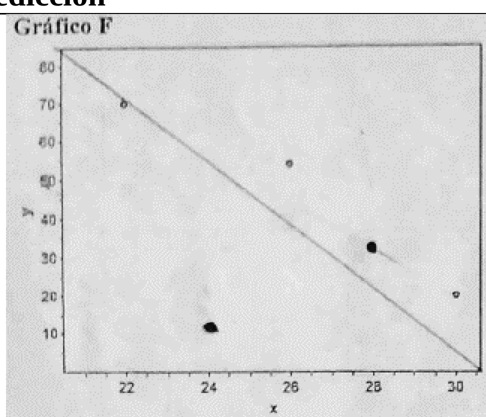
Tabla asociada a la forma de razonamiento de predicción "Conservando la forma" dados tres puntos

Predicción	Argumento
	El valor de Y para los dos puntos son 60, 40. Coloqué esos dos puntos ahí para tener como una forma o una recta descendente.

• *Observando uno de los vecinos* (ver Tabla 35, omitir la recta):

Tabla 35.

Tabla asociada a la forma de razonamiento de predicción "Observando uno de los vecinos" dados tres puntos

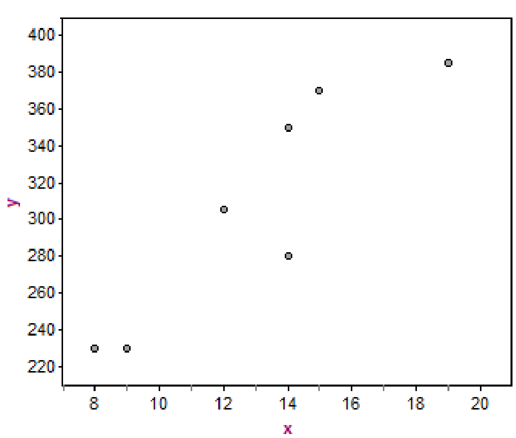
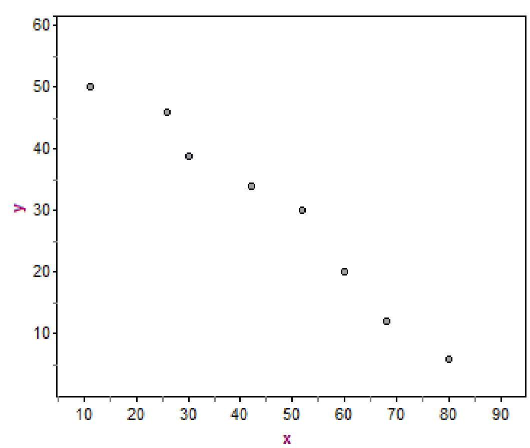
Predicción	Argumento
	El valor que yo le daría a 24 sería de 10 porque tuve en cuenta el valor de 30 y el de 28 sería de 35 porque tuve en cuenta el valor de 26.

Observemos que entre la predicción dado dos puntos y la de tres puntos algunas formas de razonamiento fueron abandonadas ("*Siendo y la mitad de x", "Entre el mínimo valor y el* *máximo valor de y"),* lo anterior debido a que el resultado de esas estrategias no satisfacían, por ejemplo, la relación de las variables que evidenciaban los tres puntos dados, obligando así a los estudiantes a abandonar su estrategia para buscar otra. Por otra parte vemos que otra forma de razonamiento resurgió: "*Observando uno de los vecinos"* que ya había sido reconocida en actividades anteriores.

Predicción con más de tres puntos:

Como ya se dijo, la cantidad de puntos en los diagramas de dispersión fue aumentando así como la cantidad de predicciones que se pedía. Lo anterior con el fin de encontrar una *forma* general, es decir, una que funcionara para cualquier nube de puntos que reflejara ya fuera una asociación positiva o negativa. Los siguientes diagramas en los que se pedía predecir se muestran en la Tabla 36, el primero con una asociación positiva y el segundo con una negativa.

Tabla 36.
Diagramas asociados a la Actividad 5 en los que se pedía predecir

	
Se pedía predecir un valor de y si x vale 11, 14 y 17.	- Se pedía predecir un valor de y si x vale 20, 35, 45 y 75.

Algunas formas de razonamiento se muestran a continuación:

• *En la mitad de dos puntos:* Cabe destacar que la estudiante que utilizó esta forma, la conservó hasta el último diagrama donde se le pedía predecir. Esta estudiante es la misma que, por ejemplo para predecir entre pilas de datos, sacaba la media de cada pila y luego la semisuma entre las medias, siendo este resultado su predicción. De manera que su estrategia es una extensión a datos no apilados. En la Figura 60 (omita la recta), se muestran sus predicciones en uno de los diagramas.

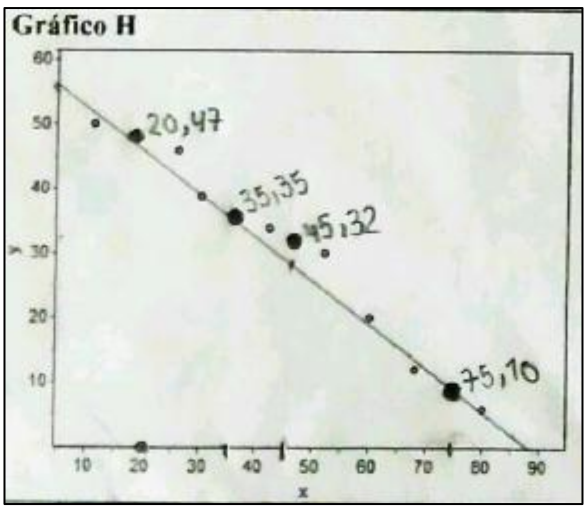
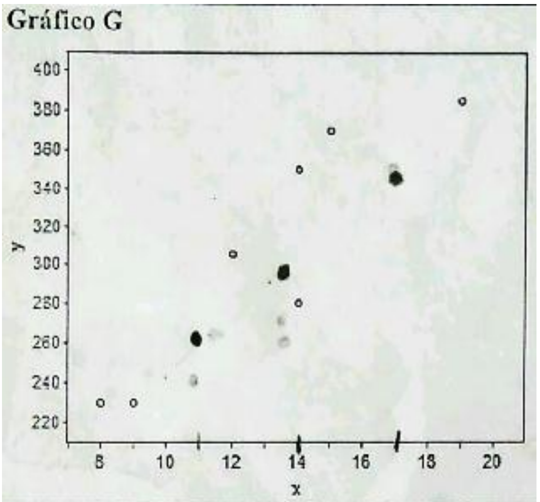


Figura 60. Figura asociada a la forma de razonamiento de predicción "En la mitad de dos puntos" con más tres puntos

- *Entre el mínimo valor y el máximo valor de y* (ver Tabla 37):

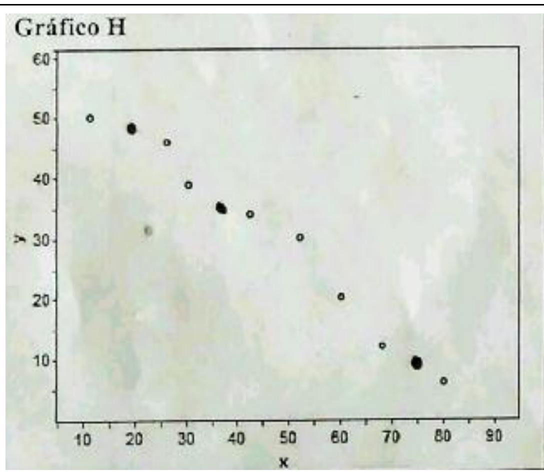
Tabla 37.
Figura asociada a la forma de razonamiento de predicción "Entre el mínimo valor y el máximo valor de y" con más tres puntos

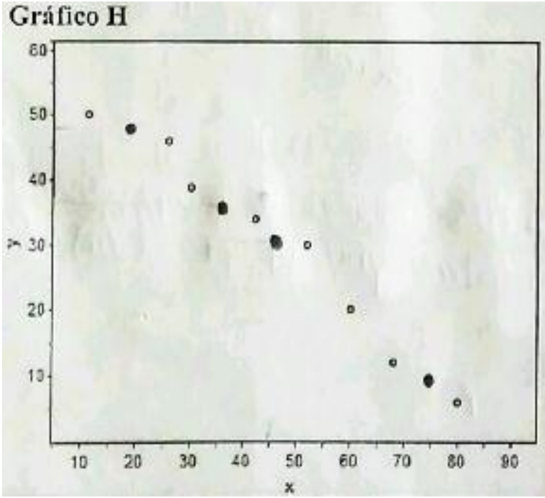
Predicción	Argumento

	El valor de Y para los tres puntos son 260, 300, 340. En este gráfico tuve en cuenta los puntos 230 y 380 para con los dos puntos me ayude mejor a predecir teniendo en cuenta que me dé una forma ascendente.

- *Entre los valores y/o de acuerdo a la relación* (verTabla 38):

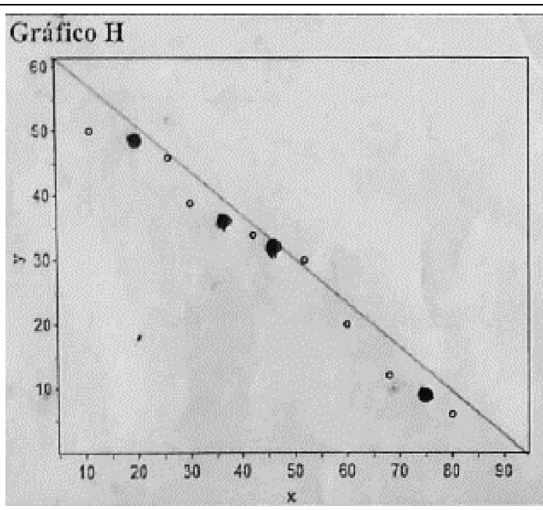
Tabla 38.
Figura asociada a la forma de razonamiento de predicción "Entre los valores y/o de acuerdo a la relación" con más tres puntos

Predicción	Argumento
	De acuerdo a que los puntos desciendan como se muestra en el gráfico y que mis valores dados se ubiquen entre los valores.

	Lo ubiqué de manera que si uno de los valores X aumenta el Y disminuye y si X disminuye el Y aumenta. Según estén ordenados de ascendente o descendente. Hay variabilidad. (20, 48) (35, 45) (45, 30) (75, 10)

- *Observando uno de los vecinos* (ver Tabla 39, omitir la recta):

Tabla 39.
Figura asociada a la forma de razonamiento de predicción "Observando uno de los vecinos" con más tres puntos

Predicción	**Argumento**
	Para 20 sería 48 porque tuve en cuenta el valor más alto, para 35 sería 37 porque tuve en cuenta el valor anterior, para 45 sería de 30 porque tuve en cuenta el valor que estaba delante de ese y para 75 sería de 10 porque tuve en cuenta el valor más pequeño.

Observemos que independientemente de las cuatros formas mostradas anteriormente, la estrategia para cada estudiante resulta ser coherente para ellos, puesto que la ubicación de sus puntos conserva la forma que siguen los puntos dados, ya fuera ascendente o descendente. Es decir, que conservan la relación de las variables que percibieron de manera global. Inclusive si observamos con más detalle de la Tabla 37 a la Tabla 39, se refleja que los estudiantes buscan que sus predicciones sean colineales, entre ellas.

A continuación mostraremos parte de una discusión que se hizo sobre el ejercicio de predicción inmediatamente anterior. Veremos la opinión de algunos de los estudiantes frente a las predicciones que hizo uno de sus compañeros.

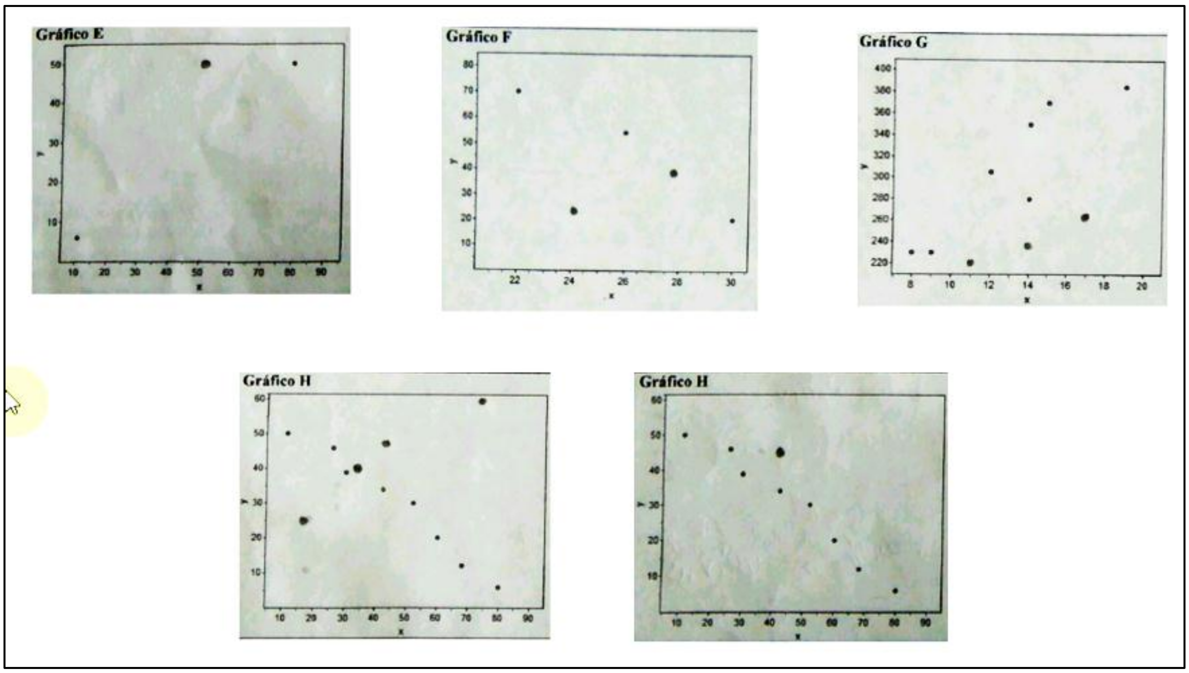


Figura 61. Predicciones de un estudiante en diferentes diagramas de dispersión

I: Miguel, qué podríamos decir ahí (ver Figura 61).

M: Decir sobre qué.

I: Sobre la ubicación de los puntos que esa persona hizo.

M: Que no tuvo en cuenta la relación.

I: ¿En cuál por ejemplo no está teniendo en cuenta la relación?

A: En el E, F y H.

I: ¿Por qué en el E no? ¿Usted dónde esperaría que estuviera ese punto, Andrea?

A: Por ahí en 30.

I: Acá en el F, por ejemplo para 28 ¿esa es una buena predicción?

Todos: Sí

M: Pero para el otro no.

I: En 24 dónde debería ir.

D: Más arriba.

I: Elder, qué pasa en el gráfico H (el de la izquierda).

E: Quedó mal.

I: ¿Por qué quedó mal?

E: Porque los puntos que él colocó van en sentido contrario de los que se plantean.

I: Qué estaría buscando…

M: ¿Esa persona?

W: Pues estaría buscando que estuvieran en una forma ascendente.

I: Ajá.

Finalmente, se les pedía que encontraran una forma general para predecir y que la mostraran, primeramente, sobre el gráfico de la Figura 62 y, después, en cada uno de los gráficos en los que habían predicho antes.

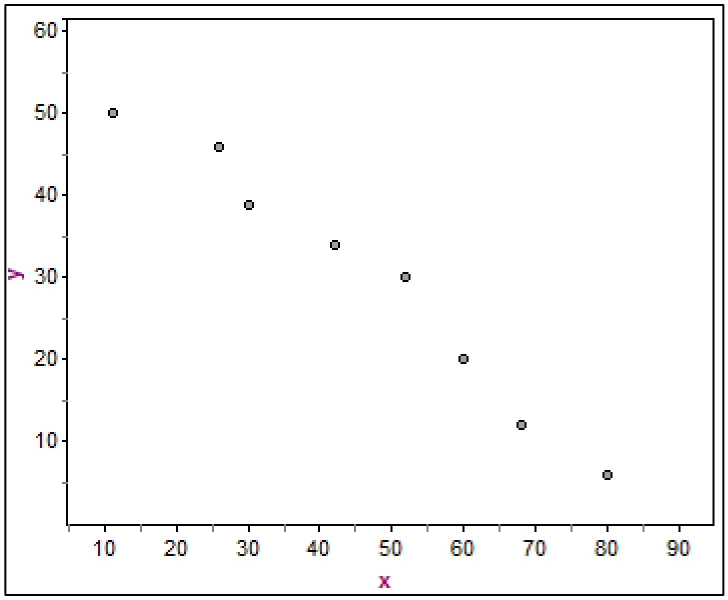


Figura 62. Diagrama de dispersión donde debían mostrar una forma general para predecir

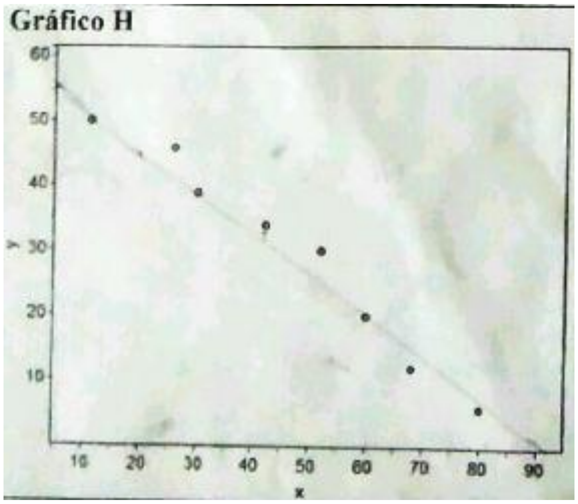
Puesto que el ejercicio podría resultar algo confuso, se introdujo haciendo alusión al caso univariado de la siguiente manera:

> *"En el caso de una variable vimos cómo un conjunto de datos podía reducirse a un solo valor, la media, el cual nos ayudaba a predecir. Para cada uno de los gráficos anteriores (A-H) traza la forma en que podría reducirse la nube de puntos teniendo en cuenta lo siguiente:*
>
> *i) Que refleje la relación que tienen las dos variables.*
>
> *ii) Que te permita hacer una buena predicción del valor de y para cualquier valor que tome x, según se muestra en cada gráfico. Explica tu respuesta".*

A continuación mostramos las formas generales que ellos proponen:

- *Trazar una recta:* Tres estudiantes trazaron una recta (ver Tabla 40 a Tabla 42).

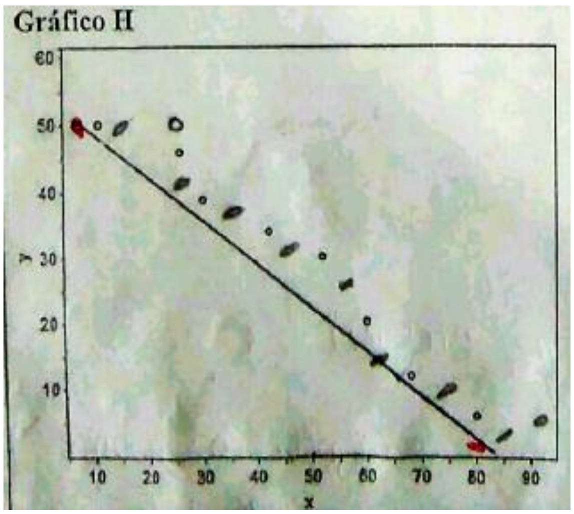
Tabla 40.
Tabla asociada a la forma general de predicción "Trazar una recta" sobre un diagrama de dispersión

Forma	**Argumento**
	Tuve en cuenta que los puntos tengan una relación ascendente o descendente, de acuerdo a los puntos la recta que tracé en cada caso es una aproximación que está o pasa cerca de cada uno de los puntos… Esta recta me está permitiendo predecir para los valores faltantes de y.

Como podemos observar, esta estudiante centra su atención en que su recta o bien pase cerca de cada uno de los puntos o sobre los puntos. Es decir, implícitamente está haciendo alusión al método formal que se acostumbraría hacer. Adicional a ellos, admite que la recta le permitiría predecir, aunque no enuncia de qué manera.

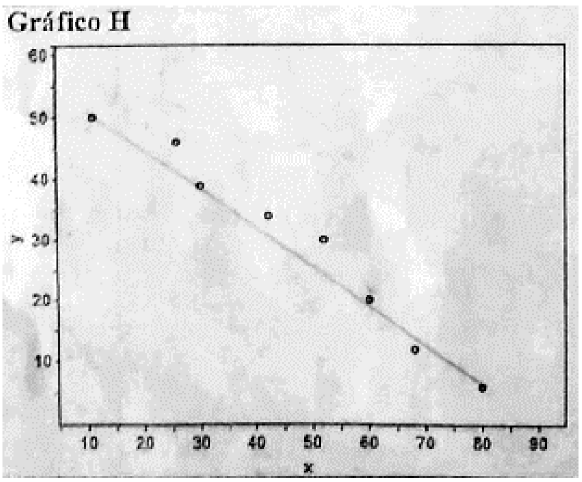
Tabla 41.
Tabla asociada a la forma general de predicción "Trazar una recta" sobre un diagrama de dispersión

Forma	**Argumento**

	Mis pasos para poder desarrollar gráficos: 1. Se mira y se detalla la gráfica, es muy importante tener todo en cuenta. 2. Se observan los puntos dados a ver cómo están ordenados si en forma ascendente o descendente. 3. Se tiene en cuenta el paso dos y respecto como están ordenados los puntos usted ordena los puntos como sigue el gráfico o como están. El por qué así, ya que así sería una forma más fácil, solo tendría en cuenta cómo están ordenados los puntos dados si de manera ascendente o descendente. Para poder hacer el tercer punto tuve en cuenta si los puntos iban de forma ascendente o descendente y decidí hacer de forma una línea ya que con eso podría tener más claro si me había quedado bien mi hipótesis entonces con esa línea podría saber más cosas del gráfico sin complicarme.

Este estudiante tomaba como referencia dos puntos (puntos en rojo), que no hacían parte de los datos, y trazaba la recta. Vemos cómo crea una secuencia de pasos, en primera instancia para predecir. Al referirse que “es muy importante tener todo en cuenta” se evidencia una visión global del gráfico, en cuanto a su forma si es ascendente o descendente. Hecho lo anterior, procede a ubicar sus predicciones respetando el “orden” o la “forma” que llevan los puntos. De la misma manera procedió para trazar la recta, asumiéndola, a la vez, como una manera de ratificar su forma de razonamiento, en cuanto a predicción. Es decir, su recta se ajusta tanto a los datos dados, como a los valores predichos.

Tabla 42.
Tabla asociada a la forma general de predicción "Trazar una recta" sobre un diagrama de dispersión

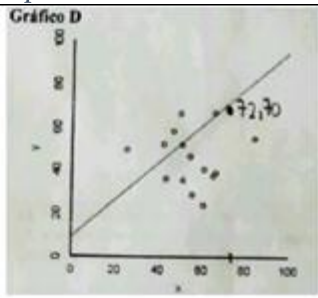
Forma	**Argumento**
	La forma general para mí sería tratar de buscar una recta entre cualquier valor que tenga X o Y.

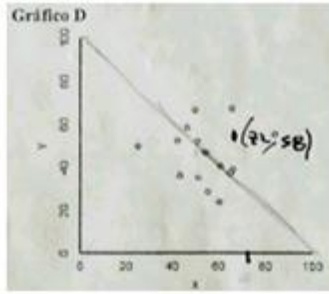
Como se lee en el argumento de este estudiante, hay una extensión de lo que se observó en el caso univariado ahora para el caso bivariado, al usar la expresión "entre". Es decir, así como la media es un valor que se encuentra entre el máximo y el mínimo valor de los datos, la recta también está entre los datos.

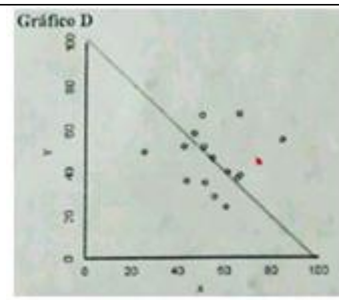
Ninguno de los tres estudiantes señaló cómo o de qué manera la recta que habían trazado les permitiría predecir un valor de y para un valor de x dado. Estos tres estudiantes trazaron una recta con pendiente positiva o negativa, en donde las variables tenían una asociación positiva o negativa, respectivamente. Es decir, sus rectas reflejaban la relación entre las dos variables (Tabla 40 - Tabla 42) tal cual como lo pedía el ejercicio, a excepción del gráfico en donde las variables no tenían ningún tipo de relación (ver Tabla 43).

Tabla 43.

Tabla asociada a la forma general de predicción "Trazar una recta" sobre un diagrama de dispersión en donde las variables no están asociadas



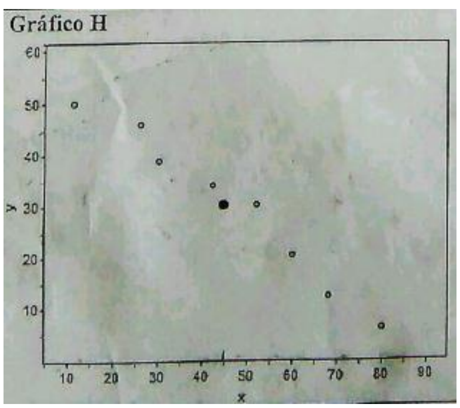




- *Un solo punto* (ver Tabla 44)*:* Dos estudiantes procedieron de esa manera

Tabla 44.

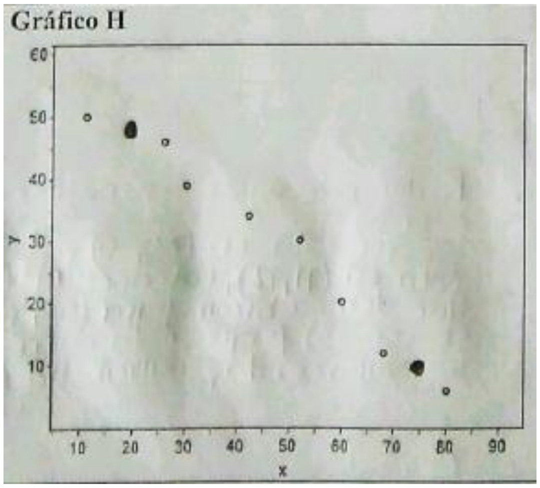
Tabla asociada a la forma general de predicción "Un solo punto" sobre un diagrama de dispersión

Forma	**Argumento**
	La forma en que podría reducirse la nube de puntos es coger un valor que a medida que aumente X aumente [disminuya] Y así tener como un mejor orden sin menos puntos.

Es decir, que al igual que en el caso univariado, donde se reducía el conjunto de datos a un solo valor, este estudiante lo redujo a un punto, donde se conserve la relación entre las dos variables. Lo anterior podría asumirse como que está buscando el punto medio para ambas variables, esto es, el conformado por la media de X y la media de Y.

- *A dos puntos, casi de extremo a extremo* (Ver Tabla 45)*:*

Tabla 45.
Tabla asociada a la forma general de predicción "A dos puntos, casi de extremo a extremo" sobre un diagrama de dispersión

Forma	Argumento
	(20, 49) (75, 15) Lo ubicaría de manera en que estén ordenados; depende si está de manera ascendente o descendente. Porque hay variabilidad.

Observemos que esta estudiante ubicó los dos puntos pero no en la parte más alta (mayor al y máximo) ni en la parte más baja (menor al y mínimo). Se podría inferir que, de alguna manera, está extendiendo, al caso bivariado, una de las características que cumple la media en el caso univariado, la cual dicta que es un valor que se encuentra entre los valores. En este caso serán dos puntos que se encuentren entre los demás puntos y que además conservan la relación entre las dos variables.

- *Ubicando el punto de manera aproximada teniendo en cuenta la relación:* "Mi forma general sería aproximándome a los valores de acuerdo a que los puntos en el gráfico asciendan o desciendan pero que concuerden con el gráfico. En cada gráfico se busca reflejar la relación de las variables ya sea que los valores de esta asciendan o desciendan según el gráfico. Los valores dados deben encontrarse entre los valores del gráfico para que esa relación sobresalga".

En realidad esta estudiante no está generalizando, sino que está emitiendo una forma para predecir cada valor que se le pida, considerando características como la relación entre las dos variables y que sus predicciones estén entre los valores dados.

A continuación mostraremos algunas de las discusiones que se dieron alrededor de esta última tarea que consistía en describir y trazar esa forma general:

Alrededor de la segunda y la tercera forma: *A un solo punto* y *A dos puntos.*

I: Entonces acá en este (Tabla 44, gráfico H) me marcaron un punto. ¿Qué opinan ustedes?

A: No.

I: Por qué.

W: ¿Debería haber al menos dos valores?

I: ¿Dos valores?

W: ¿O sea al menos dos puntos para poder predecir?

I: ¿Cómo dos puntos me ayudan a predecir para cualquier valor de x?

W: Pues porque se ubicaría entre esos puntos.

I: Es decir, si yo pusiera un punto y otro punto acá… [dos puntos sobre el gráfico H] podría predecir entre esos dos puntos ¿pero y los demás, fuera de esos dos puntos?

I: Acá tengo dos puntos (Tabla 45, gráfico H), Wendy ¿esos dos puntos me podrían ayudar a predecir?

W: Sí

I: Para cuáles valores de x ¿para cualquier valor de x?

A: No

I: Wendy, qué haría con esos dos puntos.

W: No sé…

I: Usted dijo "si hubieran dos puntos, mejor" Listo, ya tengo los dos puntos…

W: Pues porque se podría mirar para los valores que hay entre el punto de arriba y el punto de abajo… dar un valor porque van [señalando una recta con pendiente negativa con su mano derecha]… están como en línea recta, lineales. Y pues para los valores más pequeños pues se tiene en cuenta que el punto de abajo, tiene menor respecto al eje y y mayor en el eje x, entonces se pondría menos.

Es decir, tal como lo dijimos anteriormente, para esta estudiante no es suficiente reducir el conjunto de datos a uno solo como se hacía en la caso univariado, pero sí lo es, hacerlo con dos puntos que conserven la relación entre las dos variables. De manera que para predecir, basta con ubicar puntos entre esos dos puntos. Y para los valores que están por fuera, seguir el orden o la relación. Lo anterior no es otra cosa que una descripción de la forma como trazaría la recta: toma dos puntos extremos para cobijar más valores de los dados y después los une.

Alrededor de la primera forma: *Trazar una recta.*

I: Acá (Tabla 42, Gráfico H), "la forma general para mí sería tratar de buscar una recta entre cualquier valor que tenga x o y. Según mi forma de trazar la recta sí me sirve para la primera condición…" ¿Esa recta refleja la relación que hay entre las dos variables?

Todos: Asienten

I: Elder

E: Sí

I: ¿Sí, por qué?

A: Porque va de una manera…

I: Elder

E: Ascendente

A, M y D: ¡Descendenteee!

I: A medida que aumenta x, y va bajando. ¿Qué opinan ustedes de esa forma muchachos? Él dice "la forma general para mí sería tratar de buscar una recta entre cualquier valor que tenga x o y…" ¿cómo así que entre cualquier valor que tenga x o y?

A: Que esté entre x o y.

Una pregunta natural y, por lo que veremos, no tan trivial, que surge acá es ¿dónde ubicar la predicción una vez se asume la recta como esa forma general? Veamos:

I: Ahora, cumple con la primera condición [refleja la relación entre las dos variables, Tabla 42, Gráfico H] ¿cumple con la segunda? Es decir ¿me permite predecir para cualquier valor de x, el valor de y?

A: Sí

I: Elber

D: Sí

I: ¿cómo? Si yo a usted le doy ese cuadro tal como está (Tabla 42, Gráfico H) y le digo "prediga un valor de y cuando x vale 20". Con esa forma usted qué haría, ¿cómo lo ubicaría? No me diga un valor, sino dónde lo ubicaría.

D: Al pie de la recta

I: ¿al pie de la recta? Los demás que dicen ¿dónde lo ubicarían?

A: Lo ubicaría de manera que, o sea se va en 20, cerca de la recta o sobre la recta.

I: Es decir, según esta forma, para predecir y simplemente me ubico en x…

A: Y subo…

I: Y pongo el punto en dónde…

A y M: En la recta.

I: ¿Y me funciona para cualquier x?

A y W: Sí

La mayoría de los estudiantes asumen que la recta es una buena forma, sin embargo, como pudimos observar, sus predicciones serían sobre la recta o muy cerca de la recta, no es trivial para ellos ubicar su predicción directamente sobre la recta debido a la variabilidad, es como si asumieran que sobre la recta sería el valor perfecto, pero, puesto que hay variabilidad, podría estar muy cerca de la recta.

Enseguida se discutió sobre cuándo sería apropiado, o no, trazar una recta, como lo hicieron tres estudiantes.

I: Es decir, esos tres estudiantes redujeron esos puntos a una…

A: Recta

I: A una cosa

A: Que cumple la relación entre esas dos variables

I: Cuando yo veo una nube de puntos, ¿qué me da derecho a trazar una recta?

A: Que haya una relación

I: Primero que se refleje la relación… ¿o sea que la nube de puntos tenga como forma de qué, Mafer?

MF: De recta.

I: De recta. No tiene que ser perfecta, pero sí tiene que ser muy…

A: Parecida.

I: Bueno, estos tres estudiantes redujeron la nube de puntos a una recta y yo les decía que trazaran esa forma en todos los gráficos. Entonces ahí ustedes van a mirar cómo la trazaron (se observó que en los gráficos E a H, los tres estudiantes trazaron la recta respetando la relación entre las dos variables).

Recordemos que uno de los gráficos, el D (Figura 63), no reflejaba ninguna relación. A continuación, una discusión que se hizo alrededor de este gráfico y de cómo trazaron la recta en ese caso.

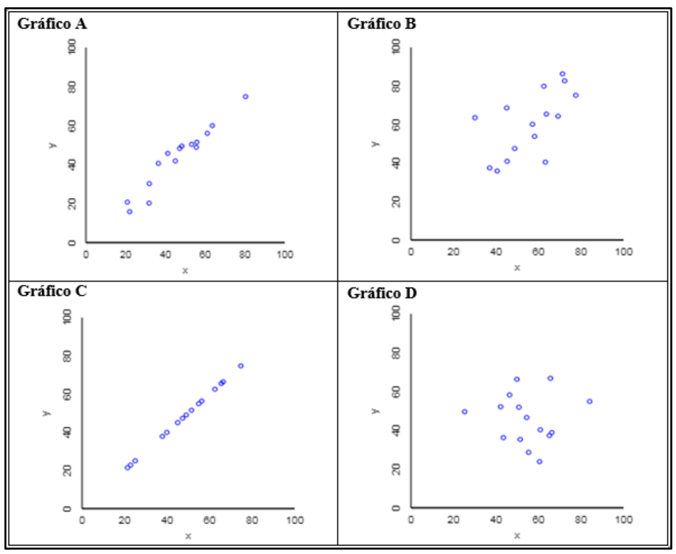


Figura 63. Diagramas donde se debía trazar una forma general de predicción

I: Pero veamos ahora este (gráficos de A a D, Tabla 43)… Ya nosotros decíamos que en el gráfico A hay una relación qué…

Todos: Ascendente.

I: Ascendente, igual en el gráfico…

Todos: C

I: En el gráfico B más o menos… ¿y en el gráfico D?

M: Nada

I: Definitivamente no se ve. Entonces… cuando esos tres estudiantes trazaron la recta, mire cómo la trazaron (como en la Tabla 43, dos con pendiente negativa y uno con pendiente positiva).

A: Yo no hallaba cómo trazar la del último [D]

I: Eso… a ustedes tres, ¿dónde se les dificultó trazar la recta?

M: En el D

A: Y en el B.

I: ¿Cómo sería la recta en un gráfico donde no hay relación? ¿Ustedes qué creen? Es decir, miren que esta recta acá [Figura 63, gráfico D], me está diciendo que al aumentar x aumenta y ¿Y eso se ve en los puntos?

Todos: No

I: Eso no se refleja en los puntos… ¿entonces cómo sería?

[Silencio]

Es decir, aunque ya se había discutido en actividades anteriores, y ya habían dado cuenta de que una recta horizontal refleja la nula relación entre las dos variables, vemos ahora que ese hecho no fue aceptado en su totalidad.

Los resultados muestran cómo los estudiantes fueron asumiendo poco a poco a la recta como una buena forma que cumplía con las dos condiciones fijadas: Que reflejara la relación y que permitiera predecir de la mejor manera y de forma global para cualquier valor X.

Ahora, como tres estudiantes trazaron una recta, quisimos indagar con sus compañeros con cuál de las tres rectas se quedarían, es decir, cuál consideraban mejor. La discusión se hizo alrededor de la Figura 64.

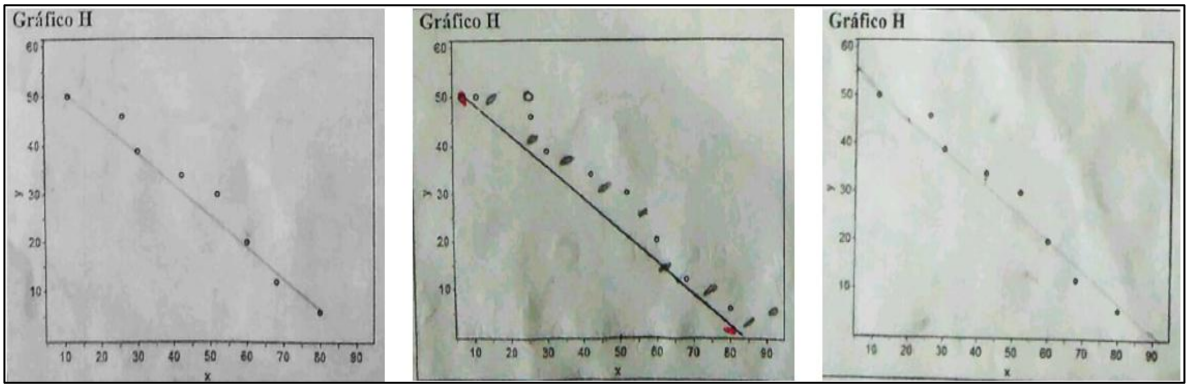


Figura 64. Rectas trazadas por tres estudiantes sobre un diagrama de dispersión

I: Pero bueno pregunta… (después de observar las rectas que trazaron los tres estudiantes sobre el gráfico H) ¿Con cuál de esas tres me quedo? ¿Cuál es la mejor recta?

E: La de Andrea (en la Figura 64, la tercera de izquierda a derecha)

M: Con las tres

A: Con la recta que esté más cerca de los puntos. Porque… por ejemplo mire la de Elder (en la Figura 64, la del medio), ninguna está sobre los puntos.

I: ¿Qué dicen los demás? Es decir, Wendy, si usted tuviera que escoger… ¿con cuál se queda? Vamos a poner las tres en un solo gráfico (como en la Figura 64)… Vamos hacer una votación acá… Wendy, con cuál se queda… con la primera, con la segunda o con la tercera.

[Cinco estudiantes votaron por la tercera y uno por la segunda]

I: ¿Por qué con la tercera, por qué no con las otras?

[Silencio]

I: ¿Porque es Andrea?

Todos: Nooo

I: ¿Entonces?

A: Porque es la que está más cerca a los puntos.

I: Eso es lo que dice Andrea ¿y los demás?

E: Se refleja una relación

I: En todos se refleja la relación descendente. Pero por qué Mafer, por qué la tercera.

MF: Yo creo que está más clara la relación.

I: ¿Está más clara la relación en la tercera?

MF: Porque están sobre algunos puntos y otros están más cerca… en cambio en la una y en la dos… mmm…

Aunque no todos argumentaron el porqué de su elección, la autora de la tercera recta en la Figura 64 la defendió argumentando que era la que estaba más cerca de todos puntos. Otros argumentos fueron respecto a la relación y a la cantidad de puntos que tocaba la recta.

Como pudimos observar, a este punto de la trayectoria, la mayoría de los estudiantes asumen la recta como esa forma de reducir una nube de puntos en donde las variables reflejan una relación lineal, evidenciándose las primeras estrategias para trazarla.

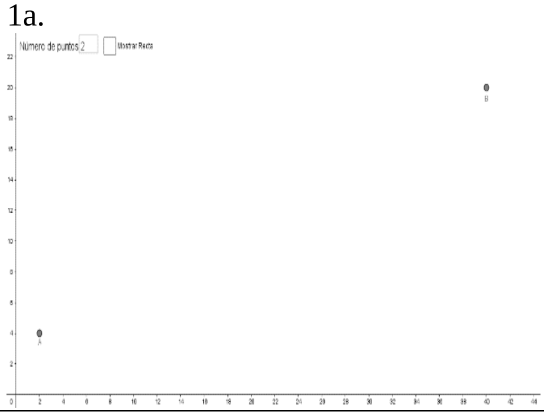
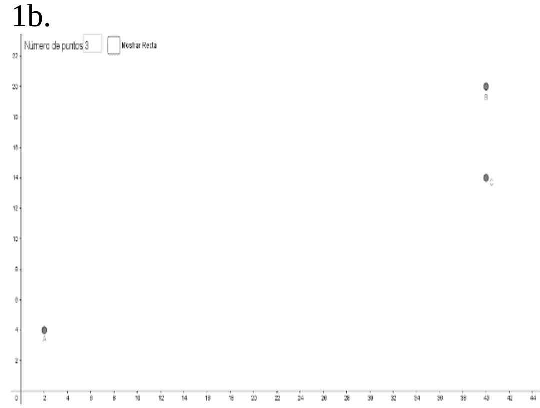
4.10 Actividad 6. Juego: Encontrando la mejor recta

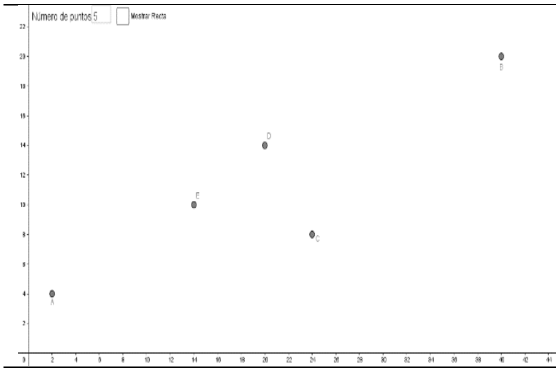
En la actividad anterior vimos cómo tres de los siete estudiantes proponían, como una forma de reducir una nube de puntos en donde las dos variables tienen una asociación lineal, trazar una recta. Vimos también cómo seis de los siete estudiantes, en el momento de quedarse con una de las tres rectas, se inclinaron por la recta que fue trazada bajo el criterio de que estuviera más cerca de todos los puntos.

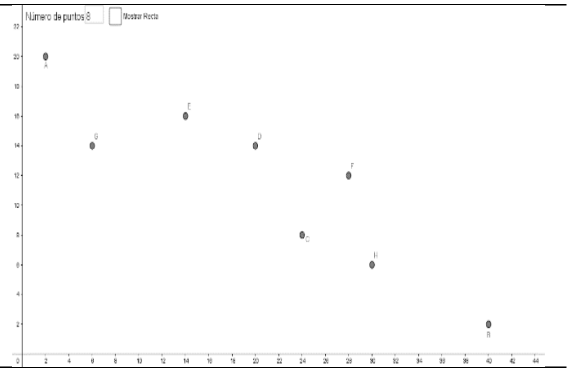
La sexta actividad que relatamos ahora, se dividió en dos partes, 6.1 y 6.2. La primera parte consistía en un juego llamado *Encontrando la mejor recta*, el cual fue simulado en Geogebra y cuyo objetivo final era que los estudiantes encontraran una estrategia para, como su nombre lo indica, encontrar la mejor recta.

4.10.1 Actividad 6 (Parte 1). Juego: Encontrando la mejor recta. La primera parte de la Actividad 6 consistió en un juego, “Encontrando la mejor recta”, en donde los estudiantes debían trazar rectas en varios diagramas de dispersión, sin contexto y con diferentes características en cuanto a la cantidad de puntos, el sentido de asociación, y la distribución de los puntos (apilados, no apilados, variables relacionadas y no relacionadas) (ver Tabla 46). Mostraremos la estrategia que proponen para trazar la mejor recta en cualquier diagrama de dispersión y sus opiniones sobre lo que les permitiría hacer esa recta ¿Acaso predecir?

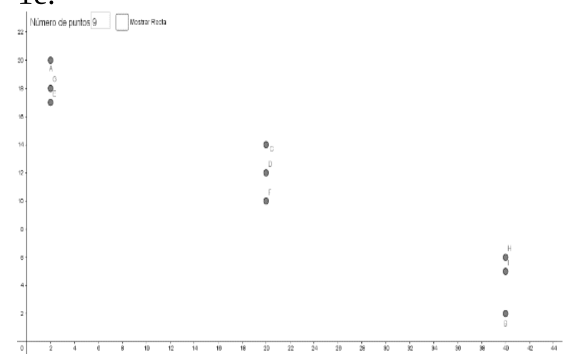
Tabla 46.
Diagramas de dispersión asociados a la Actividad 6

1a. 	1b. 
1c.	1d.

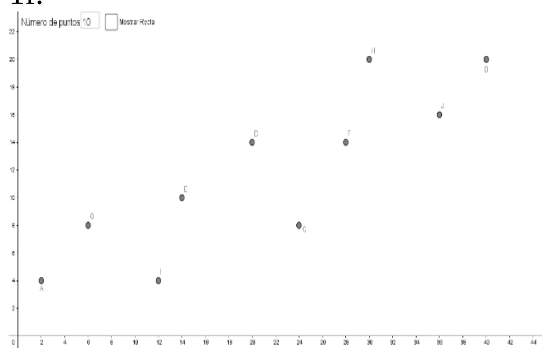




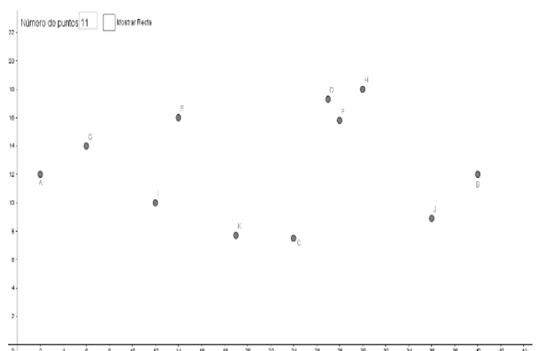
1e.



1f.



1g.



En un principio los estudiantes se encontraban con una pantalla como la que se muestra en la Figura 65.

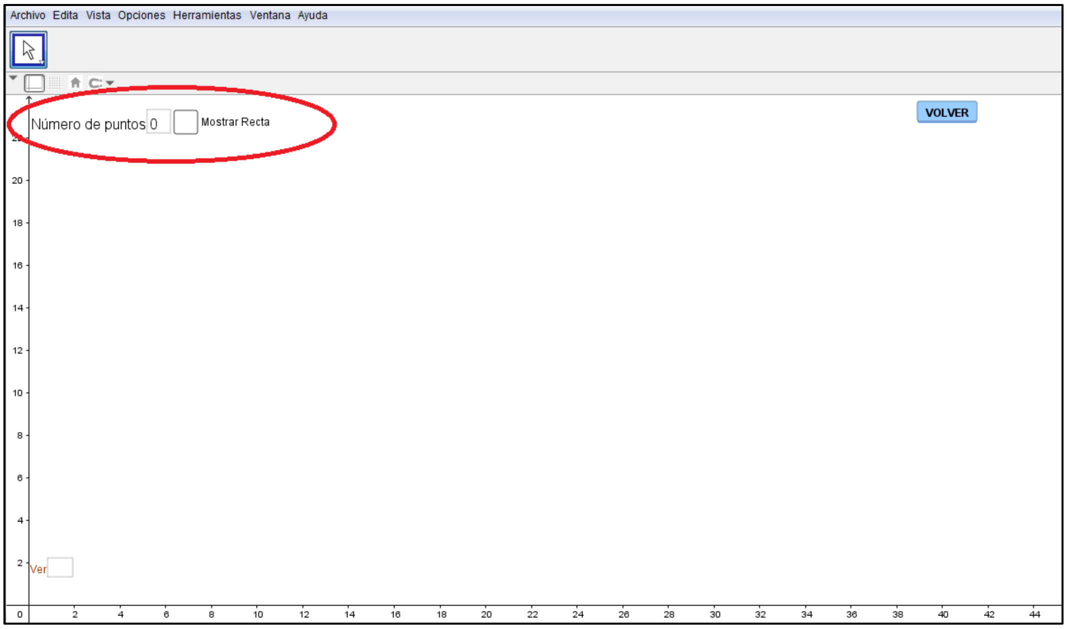


Figura 65. Pantalla de Geogebra asociada a la Actividad 6

En la Figura 65 podemos observar dos casillas: “Número de puntos”, que iba de 2 a 11, y “Mostrar recta”. Para empezar, debían trazar la recta cuando se tenían dos puntos, de manera que en “Número de puntos” escribían 2 y, en seguida, seleccionaban la casilla “Mostrar recta” (ver Figura 66). Una vez que trazaban la recta, el investigador pasaba por cada uno de los puestos revisando la recta e indicaba quién fue el feliz ganador, es decir, el que más se aproximara a la mejor recta, éste recibiría los puntos acordados en el principio del juego. Lo anterior se repitió hasta llegar a los once puntos en el diagrama de dispersión; en algunas ocasiones los estudiantes tuvieron más de una oportunidad para acercarse a la mejor recta. Al final se sumaban los puntos y se hacía la premiación.

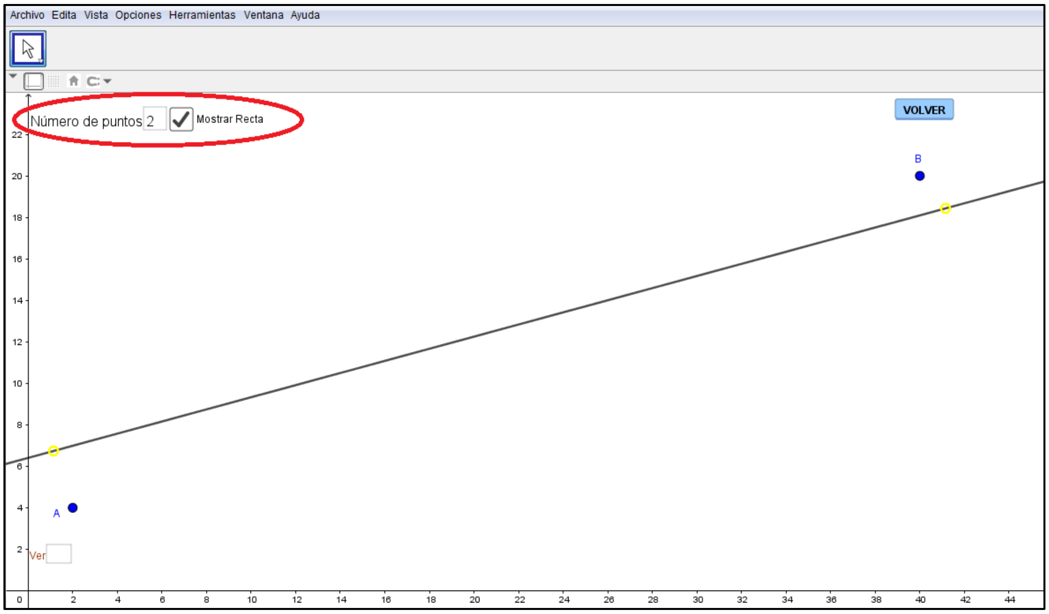


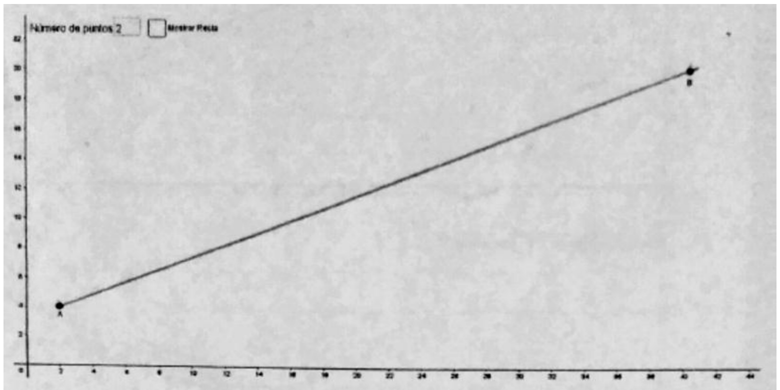
Figura 66. Pantalla de Geogebra asociada a la Actividad 6.1

Como lo mencionamos anteriormente, en primera instancia quien decidía cuál estudiante se había acercado a la mejor recta era el investigador. Para lo anterior, el criterio de decisión era la suma de diferencias entre los puntos y la recta; quien más se aproximara a cero, en esa suma, era el ganador. Claro, lo anterior no lo sabían los estudiantes.

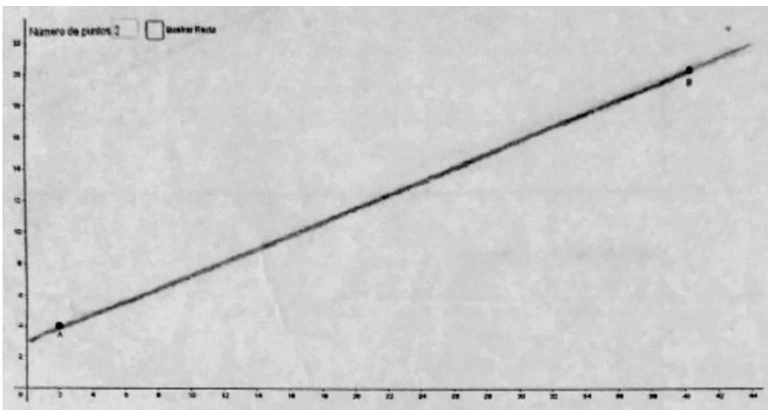
En cuanto al diagrama de dispersión con dos puntos (ver Tabla 46, 1a), los siete estudiantes trazaron una línea recta que pasaba por los dos puntos de tres maneras (ver Tabla 47): i) Cuatro estudiantes trazaron un segmento, uniendo los dos puntos (a); ii) Dos estudiantes trazaron una recta (b) y; un estudiante trazó una semirecta (c).

Tabla 47.
Formas en que los estudiantes trazan la recta dados dos puntos

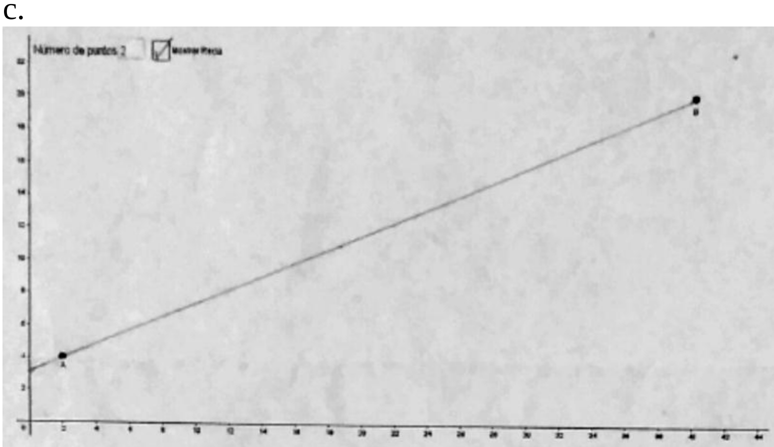
a.



b.



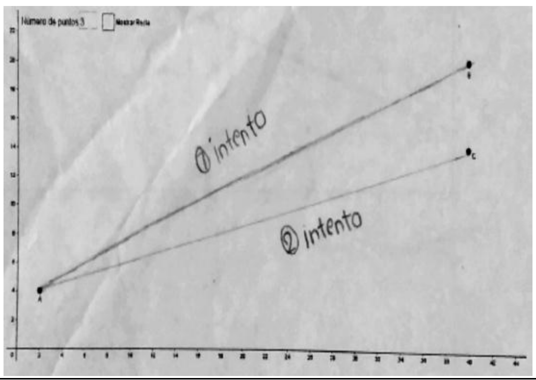
c.



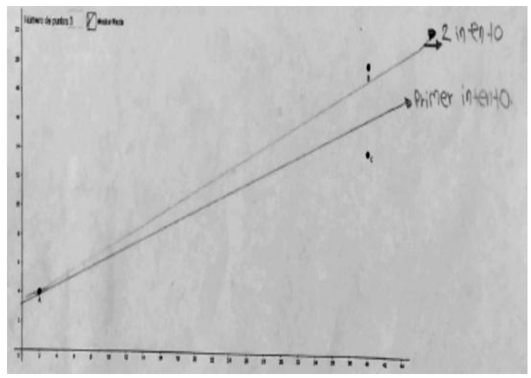
Si para el caso de dos puntos bastó con unir, o que pasara por los dos puntos, era natural que para el caso de tres puntos (ver Tabla 46 b) buscaran lo mismo. De manera que para el diagrama de dispersión con tres puntos, los siete estudiantes trazaron una línea recta en donde, en todos los casos, pasaba por A. En un primer intento, cuatro estudiantes trazaron la línea recta de manera que pasara por dos puntos (ver Tabla 48 a). Mientras que los otros tres estudiantes que pasara por A y entre B y C (ver Tabla 48 b). Estos últimos tres estudiantes, al pasar la recta entre B y C reflejan una extensión del trabajo realizando en actividades anteriores cuando se pedía predecir para datos apilados.

Tabla 48.
Formas en que los estudiantes trazan la recta dados tres puntos

a.



b.



Para el segundo intento de trazar la recta sobre el diagrama de dispersión con tres puntos, cinco estudiantes trazaron la recta por A y entre B y C y dos estudiantes continuaron con la estrategia de unir dos puntos.

La idea de que pasara por un punto se extendió al siguiente diagrama de dispersión que tenía cinco puntos (ver Tabla 46 c), es decir, los siete estudiantes pasaron su recta por A que era el punto que se encontraba más a la izquierda. Para este caso, se pueden identificar dos estrategias para trazar la recta:

i) *Uniendo dos puntos:* En un primer intento, cuatro estudiantes unieron los dos puntos más extremos tanto a la derecha como a la izquierda, es decir, el primer y último punto. Y en un segundo intento unieron el primer punto y otro (ver Figura 67).

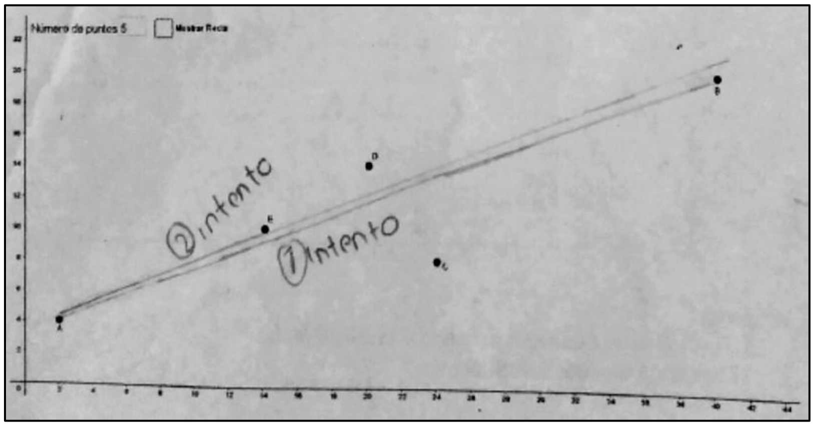


Figura 67. Forma "Uniendo dos puntos" en que los estudiantes trazan la recta dados cinco puntos

ii) *Partiendo del primer punto y pasando entre los demás puntos:* Tres estudiantes partieron del primer punto y trazaron la recta entre los demás puntos, sin tocarlos (ver Figura 68).

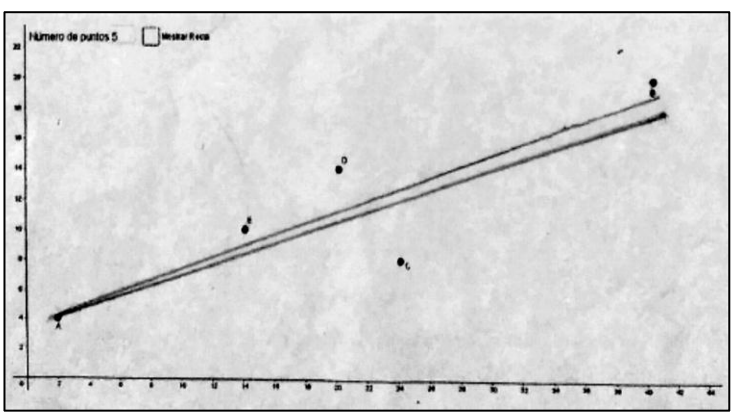


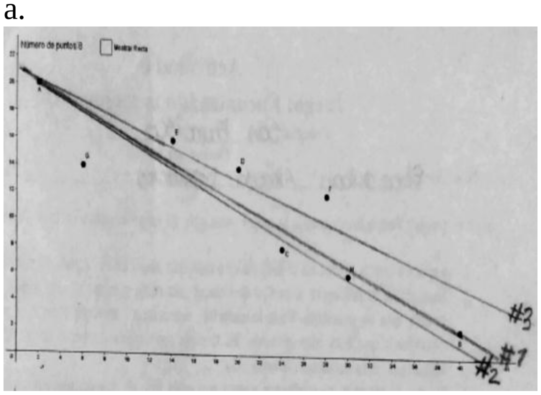
Figura 68. Forma "Partiendo del primer punto y pasando entre los demás puntos" en que los estudiantes trazan la recta dados cinco puntos

Las anteriores estrategias se fueron repitiendo en el siguiente diagrama de dispersión que tenía ocho puntos (ver Tabla 46 d). A diferencia de los anteriores diagramas, la asociación de las variables en éste era negativa. Cabe destacar que en todos los intentos para este diagrama, los siete estudiantes reflejaron esa asociación con la recta que trazaron, es decir, con una pendiente negativa.

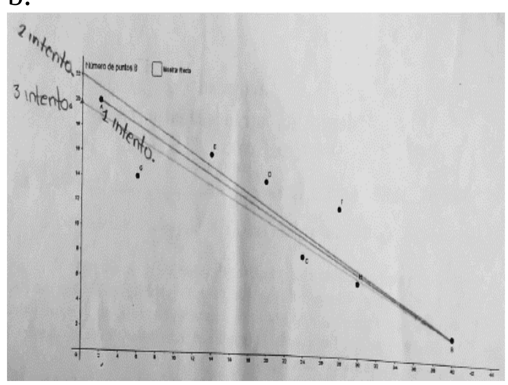
Nuevamente, como primer intento cinco estudiantes trazaron la recta por el primer y último punto y para los siguientes intentos partían del primer o último punto y trazaban la recta entre los demás puntos (ver Tabla 49).

Tabla 49.
Forma "Partiendo del primer punto, o último punto, y pasando entre los demás puntos" en que los estudiantes trazan la recta dados ocho puntos

a.



b.



Para los siguientes diagramas se trabajó en equipo, se formaron dos equipos con dos estudiantes y uno con tres. La agrupación se hizo de manera que por lo menos uno de los integrantes del equipo se estuviera aproximando bastante a la mejor recta en los diagramas anteriores. Lo anterior se hizo con la intención de que entre ellos compartieran sus estrategias para finalmente quedarse con la mejor.

En el siguiente diagrama de dispersión, con nueve puntos, los datos estaban apilados (ver Tabla 46 e). A diferencia de los intentos anteriores por encontrar la mejor recta, a este punto del juego algunos estudiantes empiezan a abandonar la estrategia de que necesariamente la recta tiene que pasar por algún punto, o dos. Por ejemplo, de la anterior manera, hubo un equipo, de dos estudiantes, que en su primer intento se acercaron bastante a la mejor recta, por lo que decidieron no cambiarla para los demás intentos (ver Figura 69).

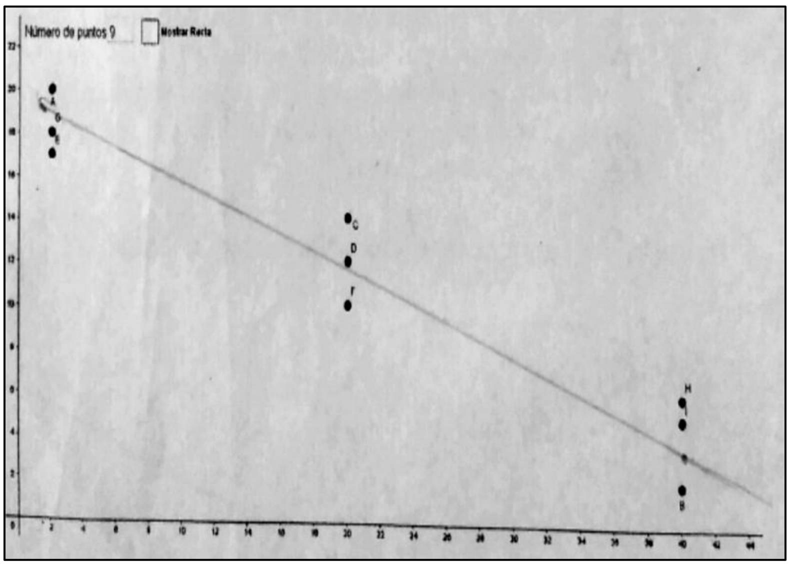
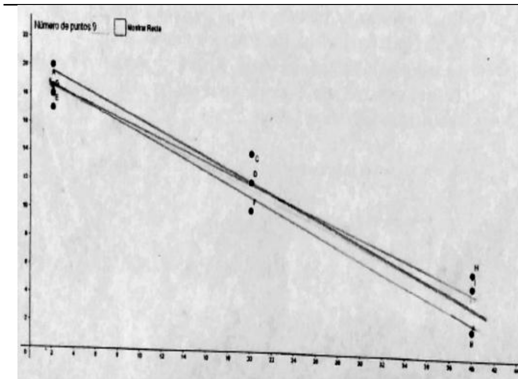
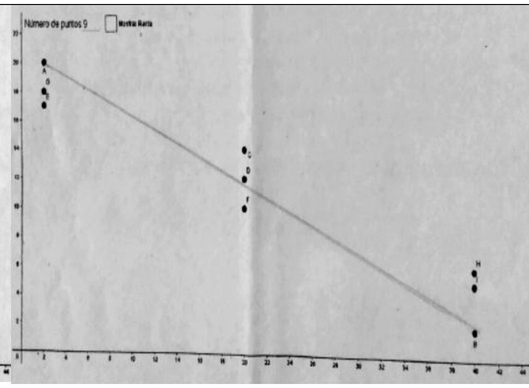


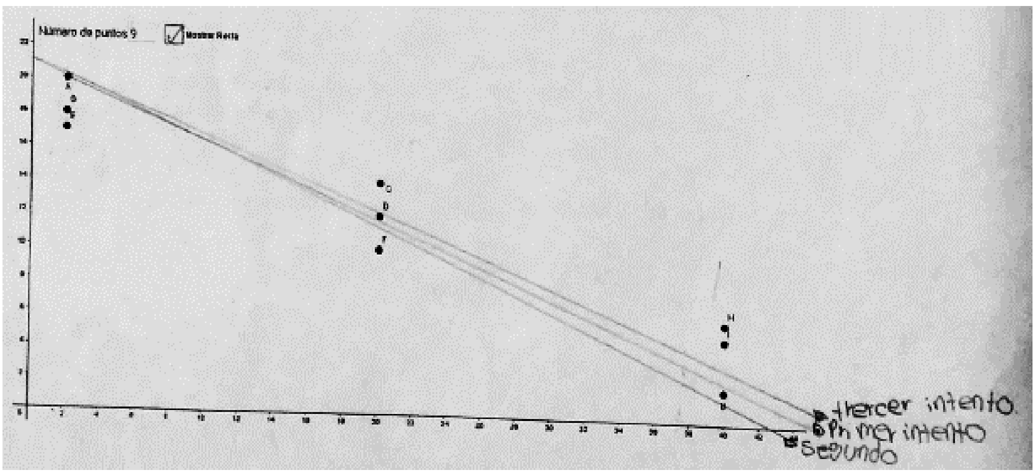
Figura 69. Forma en que dos estudiantes trazan la recta dados nueve puntos apilados

Los demás estudiantes seguían aferrados a la idea de que debía pasar por uno o dos puntos, o en unir el primer y el último punto (ver Tabla 50).

Tabla 50.
Forma en que los estudiantes trazan la recta dados nueve puntos apilados





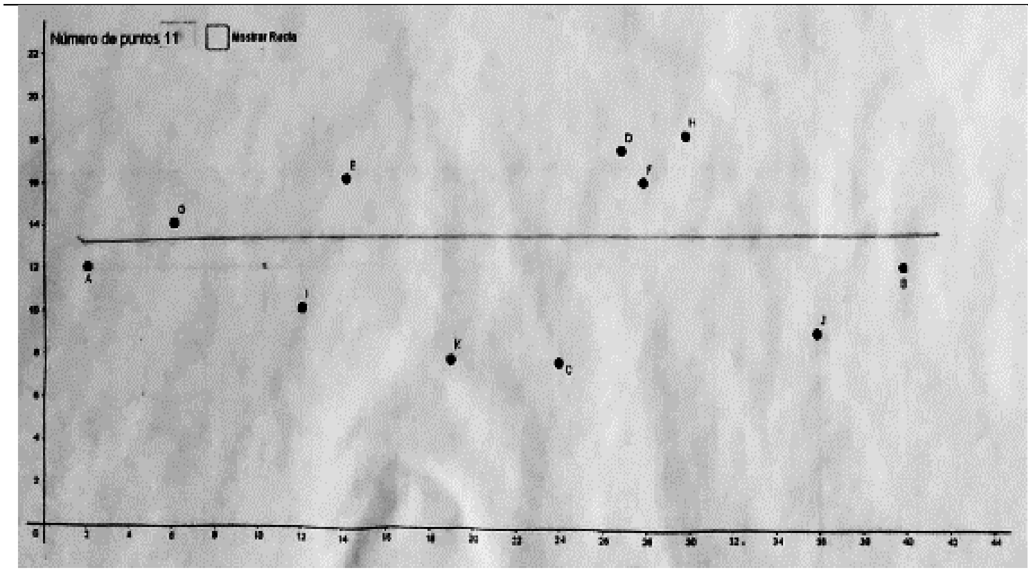


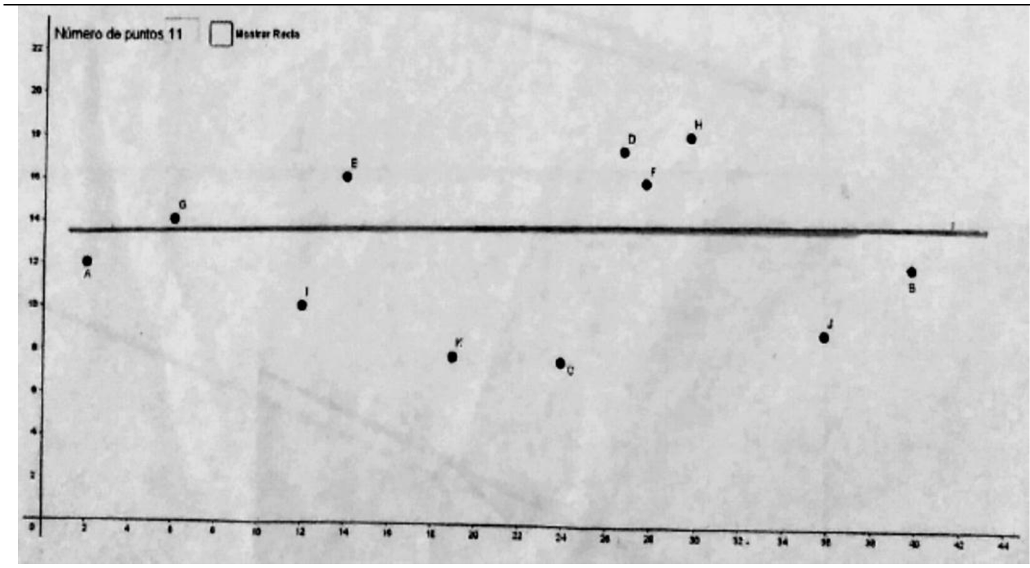
Puesto que las estrategias se siguieron repitiendo en el siguiente diagrama de dispersión con diez puntos (Tabla 46 f), pasaremos directamente al último diagrama en donde los datos no reflejaban ningún tipo de asociación (ver Tabla 46 g).

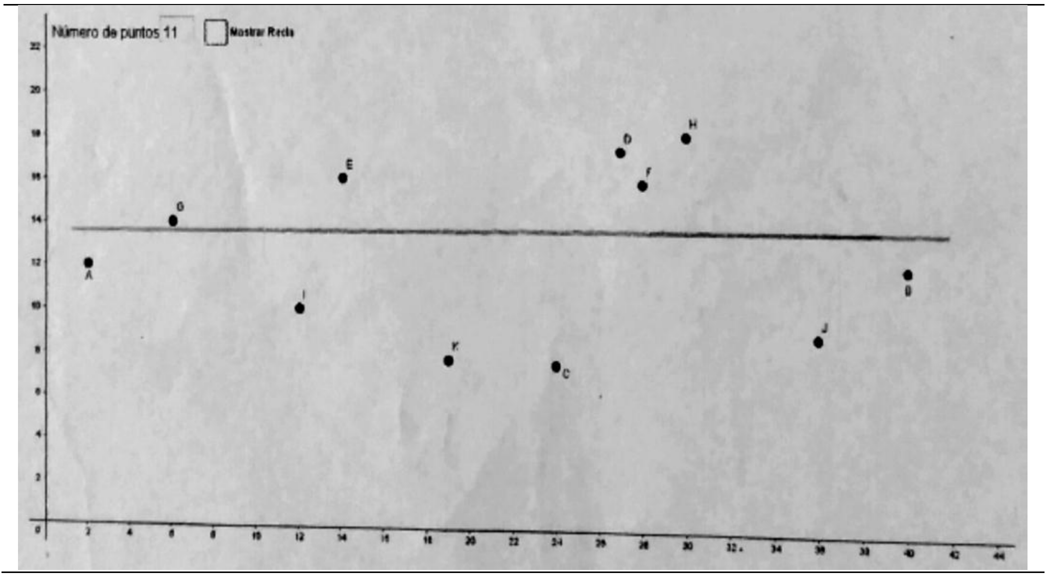
Cabe recordar que en actividades previas se había discutido sobre cómo sería la forma de la recta para un diagrama con estas características, es decir, una recta que no reflejara ningún tipo de relación entre las variables. Sin embargo, las formas de trazar la recta para este diagrama, difirió inclusive entre los mismos integrantes de algunos grupos. Solo un grupo conformado por dos, acordó con que la recta trazada tenía que ser horizontal.

i) *Horizontal y entre los puntos:* Tres estudiantes trazaron una recta de esta manera. Se podría pensar que los estudiantes identifican la tendencia de los puntos o que una recta así no refleja una relación, ascendente o descendente, entre las variables (ver Tabla 51).

Tabla 51.
Forma "Horizontal y entre los puntos" en que los estudiantes trazan la recta donde no hay asociación







ii) *Entre los puntos con una leve inclinación* (ver Figura 70):

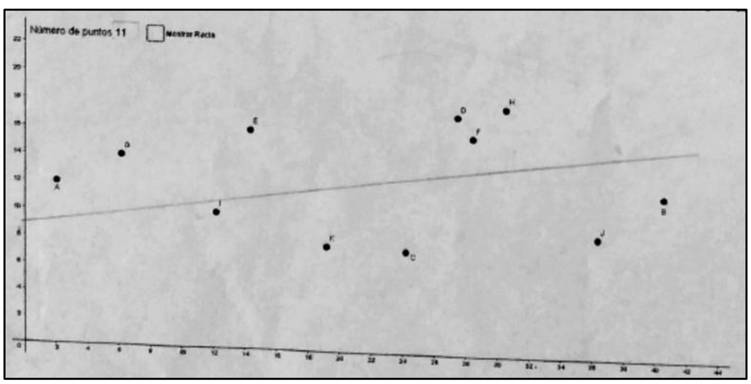
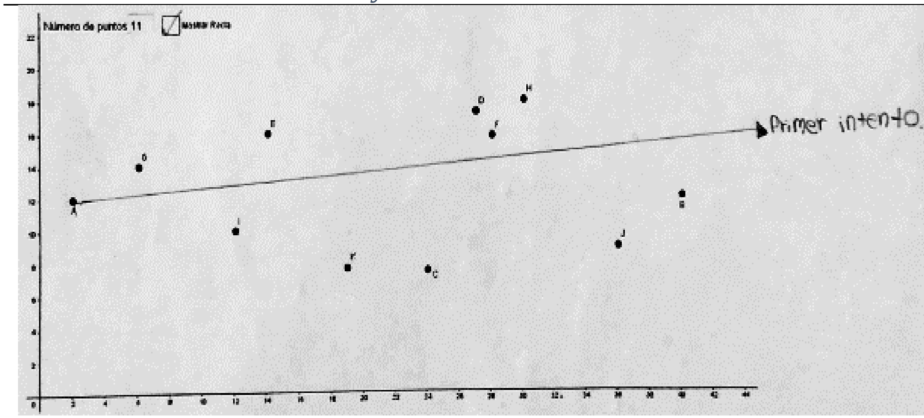
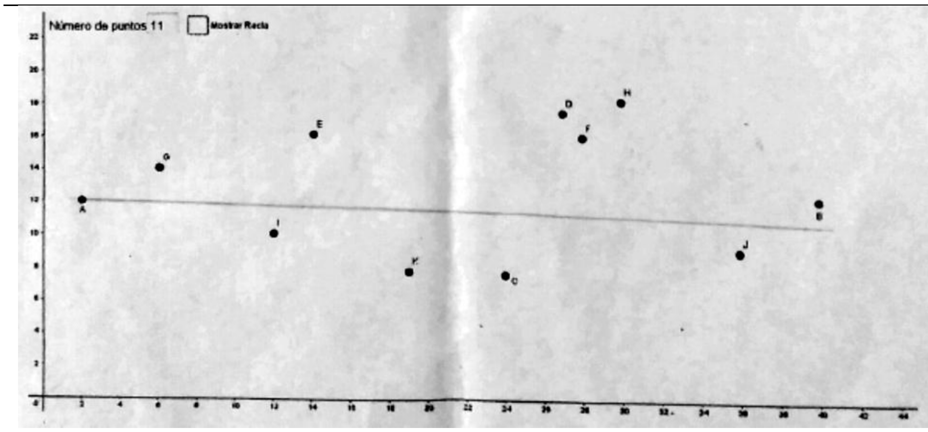


Figura 70. Forma "Entre los puntos con una leve inclinación" en que los estudiantes trazan la recta donde no hay asociación

iii) *Partiendo del primer punto y pasando entre los demás puntos* (ver Tabla 52): Dos estudiantes trazaron la recta de esta manera, uno mostrando una leve asociación positiva entre las variables y el otro una asociación negativa.

Tabla 52.
Forma "Partiendo del primer punto y pasando entre los demás puntos" en que los estudiantes trazan la recta donde no hay asociación





iv) *Partiendo del origen y entre los puntos* (ver Figura 71): Una estudiante trazó la recta como si las variables tuvieran una asociación positiva. La manera en que van los puntos no le está dando mucha información para trazar la recta, puesto que hasta este punto ya se había discutido que una recta de esa manera, nos indicaba que al aumentar una variable, la otra también aumentaba, lo que no se ve, de manera global, en el diagrama.

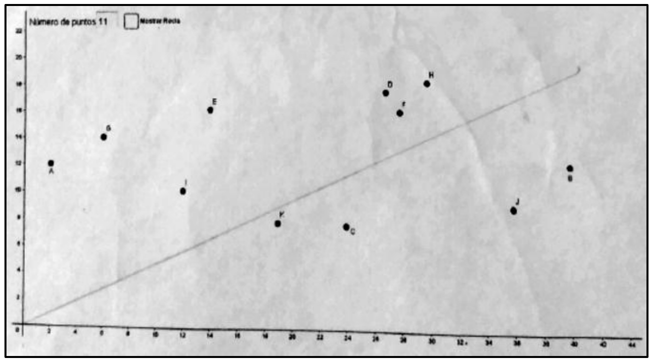


Figura 71. Forma "Partiendo del origen y entre los puntos" en que los estudiantes trazan la recta donde no hay asociación

Hasta el momento, las formas de trazar la recta por los estudiantes en diferentes diagramas de dispersión en donde las variables reflejaban una asociación lineal, ya fuera positiva o negativa, o ninguna asociación, son interpretaciones nuestras. Es decir, lo que nos permitió ver los trazos que ellos hicieron. A continuación veremos la estrategia que finalmente ellos proponen para trazar la recta y lo que les permitiría hacer esa recta.

Finalizado el juego, esta primera parte de la actividad terminaba con dos preguntas:

1. ¿Cuál es tu estrategia para trazar la mejor recta?
2. ¿Qué te permite hace esa recta?

Independientemente de la estrategia, la forma con que trazaron las rectas respetó la asociación, positiva o negativa, que tenían las variables. Como ya se mostró, hubo menor acierto, en trazar la recta, en el diagrama donde las variables tenían una correlación casi nula.

A continuación resumimos las estrategias que, finalmente, los estudiantes proponen y describen para trazar la mejor recta:

• *Cerca de todos los puntos:* Cinco estudiantes hablaron en términos de cercanía. Sin embargo, se diferencian, a su vez, cuatro estrategias:

i) Partiendo del primer punto:

“Colocar un punto en la coordenada A y trazarla por donde pasen cerca o entre los puntos”.

Para este estudiante, una condición necesaria es que la recta parta del primer punto.

ii) Al lado de los puntos o en los puntos:

“Mi estrategia para trazar la mejor recta fue tener en cuenta la agrupación de los puntos y también que la línea que trace pase al lado de los puntos o en los puntos”.

Lo anterior supone que si la recta se pudiera trazar por la mayor cantidad de puntos posibles, sería mejor, algo así como una “recta modal”. De lo contrario, lo más cercano de todos los puntos o al lado, como el estudiante lo expresa.

iii) Cerca de los puntos:

“Mi estrategia fue que los puntos quedaran muy cerca o sobre la recta, que no quedaran alejados de la recta. Para que la relación se viera a simple vista, que la mayoría de puntos quedaran más cerca a la recta”.

“Cuando trazo la recta lo primero que hago es mirar cuál es la relación entre las dos variables, luego trato de que la recta que trace vaya acorde con los puntos, es decir, que no esté fuera de lo normal o que esté muy cerca a los puntos de manera que cuando me den un valor yo lo pueda ubicar cerca o en la recta”.

El anterior argumento además de exponer las consideraciones a tener en cuenta, como la relación entre las variables y la tendencia de los puntos, explicita dos formas para predecir, una vez se tenga la recta: ubicar el punto sobre la recta o muy cerca de la recta. Esta última forma de predecir se justifica con la variabilidad, es decir, es como si el valor real de la variable respuesta, para un valor de la variable explicativa, fuera sobre la recta, pero, debido a la variabilidad, es muy difícil que ocurra exactamente ese valor, podría ser ese valor o uno muy cercano a él.

iv) Uniendo el primer y el último punto:

“Que la recta empiece desde el primer punto al último y que los puntos estén en la recta o cerca a ella de forma ascendente o descendente”.

En las cuatro estrategias mencionadas anteriormente, el hecho de hablar en términos de "cercanía" involucra, de alguna manera u otra, el método de mínimos cuadrados, pues están pensando en minimizar las distancias entre los puntos y la recta.

Las siguientes dos estrategias, toma consideraciones del caso univariado, como las medidas de tendencia central.

- *Dejando igual número de puntos a lado y lado de la recta:*

"La estrategia era unir los puntos que quedara una estrategia para poder unir los puntos, uno los dos puntos con una línea y que debajo y arriba sea la misma cantidad de puntos, es decir, la línea separa la mitad de los puntos, y tiene que ser una línea trazada pero que tenga relación".

Lo anterior podría denominarse como "la recta mediana" haciendo una extensión del caso univariado al caso bivariado, donde se deja igual número de puntos a lado y lado de la recta.

- *Entre los puntos*

"La estrategia que utilicé fue la misma que utilizamos en la actividad de ubicar la tienda entre las casas, la estrategia es que la recta debe estar ubicada entre los valores".

Como se observa en el anterior argumento, esta estudiante extiende, de forma explícita al traer a colación un ejemplo visto en el caso de una variable, una característica propia de la media, que consiste en ubicarla entre los valores.

Como se observa, las anteriores estrategias se incluyen dentro de los criterios que utilizan los estudiantes para la ubicación de la recta de mejor ajuste, en los estudios realizados por Sorto et al. (2011) y Casey (2015), descrito en los antecedentes.

En cuanto a la segunda pregunta sobre qué les permitiría hacer la recta, en realidad debimos preguntar directamente sobre si la recta les permitiría, o no, predecir y cómo, las respuestas de los estudiantes se relacionaron con:

- *Reflejar la relación entre las variables*: Cuatro estudiantes argumentaron en términos de que la recta permite ver, con mayor claridad, la relación que tienen la variables, por ejemplo:

“Lo que me permite hacer la recta es que haya una relación ascendente o descendente, porque el orden de los puntos me permiten ubicar mejor la recta y si no hay una relación entre las dos variables sería un poco más complicado la ubicación de la recta”.

- *Para predecir:* Solo un estudiante dejó explícito que con la recta se podría predecir un valor de la variable respuesta, para cualquier valor de la variable explicativa. Adicional a eso, supone cómo debería ir la nube de puntos dada la recta.

“Para predecir y encontrar cualquier valor y lograr formar una nube de puntos con los valores que me dan”.

Las dos anteriores respuestas se asocian con lo que Casey (2015) hace referencia al significado que los estudiantes le dan a la recta de mejor ajuste.

Dos estudiantes justificaron sus respuestas más en la ubicación de la recta que en lo que les permitiría hacer la recta.

• *Estar cerca de los puntos:* “Que pueda pasar o estar cerca de los puntos así sea de forma descendente o ascendente”.

• *Entre los puntos:* “Lo que me permite hacer esa recta es tener en cuenta que debe estar ubicada entre los valores”.

Como vimos, el objetivo del desarrollo de la primera parte de esta actividad, fue conocer las estrategias de los estudiantes para trazar la recta de mejor ajuste así como sus intuiciones sobre lo que ésta les permitiría hacer u observar. Vimos que estas estrategias y significados alrededor de la recta de mejor ajuste se corresponden con resultados de estudios previos en otras investigaciones. De manera que se identificaron tres estrategias: Cerca de todos los punto, Dejando igual número de puntos a lado y lado de la recta y, Entre los puntos. Respecto a lo que les permitiría la recta, la mayoría de estudiantes la relaciona con el tipo de asociación que tienen las variables y solo uno dice que le permite predecir.

4.10.2 Actividad 6 (Parte 2). Juego: Encontrando la mejor recta. Para la segunda parte de la actividad seis, los estudiantes debían ubicar nuevamente la recta con la estrategia que ya habían establecido, en cada uno de los diagramas mostrados en la primera parte. Seguidamente, en cada diagrama, debían verificar que la ubicación de la recta. Para lo anterior, una vez ubicada la recta, en la parte inferior de la pantalla en “Ver” debían escribir “EK” en donde, con esta acción, se visualizaba una casilla llamada “¡Eureka!”, seleccionaban la casilla y si la posición de la recta estaba bien, aparecía la palabra “¡Eureka!” (ver Figura 72). De lo contrario debían mover la recta hasta encontrar la mejor posición. Esta actividad se torna válida en la medida en que los estudiantes sean capaces de describir las razones que identifican la recta correcta.

Figura 72. Pantalla de Geogebra asociada a la Actividad 6.2

Hecho lo anterior en cada uno de los diagramas, debían cuestionarse sobre la ubicación de la recta de mejor ajuste. A continuación, veremos que algunos de los estudiantes cambian su estrategia mientras que otros se mantienen firmes en la que ya habían propuesto:

• *Cerca de los puntos:* Cuatro estudiantes hablaron en términos de cercanía, bajo ciertas condiciones:

i) *Tomando en cuenta el primer y último punto:* "Que al ubicar la recta trato de que siempre los dos últimos puntos estén en o cerca de la recta mirando la relación y si no hay relación la recta no es ni de forma ascendente ni descendente". Los dos últimos puntos a los que se refiere la estudiante, en realidad corresponden al primer y último punto en el diagrama de dispersión.

ii) *Que toque siempre un punto:* "Mi estrategia final para trazar la recta es tener en cuenta que los puntos están agrupados, que la recta está cerca y que pase por un punto siempre…".

iii) *Por la mitad de los puntos:* “La estrategia para la recta sería que esta recta pasara por la mitad de los puntos y muy cerca de ellos”.

iv) *Sobre algunos puntos:* “Mi estrategia de la recta es que la gráfica de en la que están ubicados los puntos, la recta esté cerca a los puntos según que en algunos casos pase sobre el punto o en el punto”.

- *Igual número de puntos a lado y lado de la recta:* Aunque algunos diagramas tenían un número impar de datos, un estudiante justificó que, para un diagrama con un número par de datos, la recta debía dejar a lado y lado, de la misma, igual número de puntos. Mientras que para un diagrama con un número impar de datos no estableció ninguna estrategia: “Entonces, yo consideraría que debería ser: en los números pares debería ser la línea iría por la mitad, entonces es decir la línea iría en la mitad, entonces los puntos de arriba de la línea son la misma cantidad de los puntos de abajo de la línea”. En la siguiente discusión se le sugirió que su estrategia debería funcionar para cualquier cantidad de puntos, sugerencia que no logró disuadir al estudiante de cambiar su estrategia.

E: Profesora ya tengo mi estrategia, ahora sí la tengo… ¿usted qué ve ahí profesora?

I: A ver, cuénteme.

E: Usted desde su punto de vista qué ve ahí…

I: No, dígame usted… (risas) A ver, cuál es su estrategia.

E: Mire… hay un, dos, tres, cuatro… (contando los puntos por debajo de la recta) Entonces, como tratar de encontrar la variabilidad entre los puntos de abajo y los puntos de arriba.

I: ¿Encontrar la variabilidad? ¿Cómo así?

E: Que quede como una balanza, como nivelado.

I: Y qué significa nivelado, nivelado en qué sentido.

E: El Eureka es como si fuera la mitad de los puntos, la línea es como si dividiera la mitad de los puntos.

I: Pero ahí tengo nueve puntos

E: Por eso, ahí no se puede.

I: La idea es encontrar una sola estrategia.

- *Fijar un punto:* "Mi estrategia general sería colocar un punto fijo y mover el otro punto hasta que me salga la mejor recta" Este estudiante desde la primera parte de esta actividad, se mantuvo en que necesariamente la recta debería pasar por un punto, específicamente el primero. Infortunadamente, debido a la sensibilidad del ajuste de la mejor recta en la simulación, esta condición que él propone se ajustaba a la mejor recta en varios diagramas, pero no en el diagrama con once puntos, por ejemplo, como se muestra en la siguiente conversación:

I: ¿Listo, ya con 11?

D: No… No la encuentro

I: Pues intente mover los dos puntos… No necesariamente tiene que mover uno solo.

D: No porque entonces ahí sí se iría "a bote" mis anteriores…

I: ¿Su estrategia?

D: Sí

I: Es que la estrategia tiene que funcionar para todos

D: Para los otros sí me funcionó este… entonces cómo lo voy a cambiar…

Este estudiante no cambió de estrategia.

- *Entre los puntos:* “La mejor recta se ubicaría entre los valores, no siempre debe pasar sobre un punto. En estos casos la recta se debería ubicar entre los puntos y no necesariamente debe pasar sobre un punto”. Como se observa, esta estudiante extiende ideas desarrolladas en el caso univariado con la media, que se ubica entre los valores y que, no necesariamente, corresponde a uno de los datos.

I: ¿Cómo está ubicando la recta?

W: Que quede entre los valores, como lo que hicimos en lo de las tiendas y eso…

I: Que quede entre los… ¿estos son valores? (señalando los puntos)

W: O sea entre los…

I: Estos son puntos ¿sí? ¿Y necesariamente tiene que pasar por algún punto?

W: Pues no necesariamente tiene que pasar por algún punto, en algunos casos.

Como observamos, más de la mitad de los estudiantes hacen referencia a la cercanía ¿pero qué significa ser la más cercana a todos y cada uno de los puntos? ¿O bajo qué condiciones se da dicha cercanía? No es tan natural e inmediata la respuesta. De manera que el propósito de esta actividad era extender las ideas que se habían visto en el caso univariado al caso bivariado. Una de ellas era observar la suma de las diferencias de los puntos a la recta y el cuadrado de esa suma. Por lo que, en primera instancia, con ayuda de la simulación, los estudiantes pudieron darse cuenta, en cada uno de los diagramas, de que las diferencias por encima de la recta eran positivas, por debajo negativas y de que, efectivamente, la suma de las diferencias de los puntos a la mejor recta era cero, como se hizo por ejemplo con el Machín Machón.

Por ejemplo, consideremos el caso en que “Número de puntos” es 3. Los estudiantes ubicaban la mejor recta y, en la parte inferior de la pantalla en “Ver” escribían MD, la anterior acción mostraba una casilla llamada “Mostrar diferencias” que, al ser seleccionada, mostraba las diferencias de los puntos a la recta (ver Figura 73).

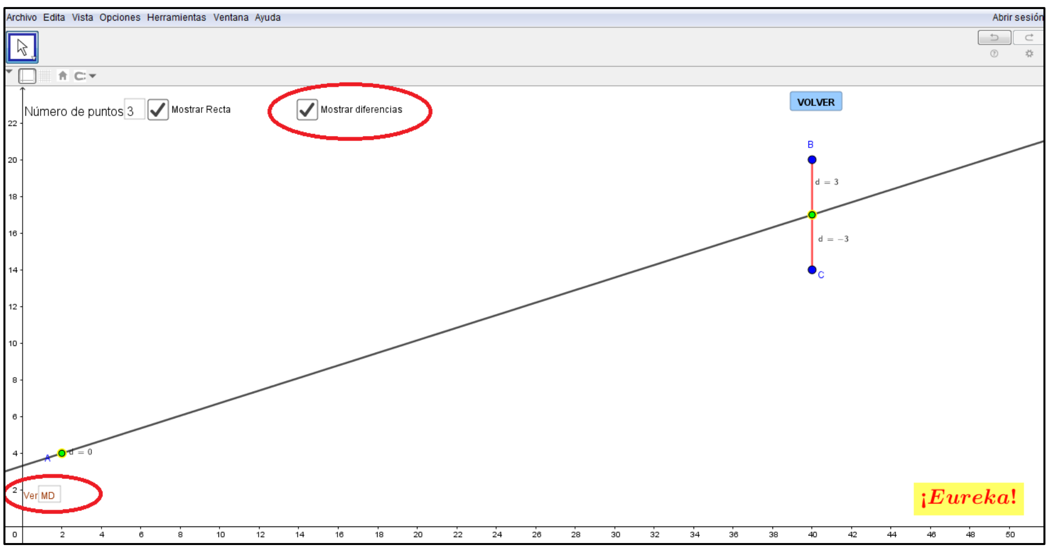


Figura 73. Pantalla de Geogebra asociada a la Actividad 6.1 donde se muestran las diferencias de los puntos a la recta

Aunque en la guía de trabajo se hablaba en términos de diferencias, cuatro estudiantes asumieron esas diferencias como distancias, por ejemplo: *“Los valores son la distancia que hay de la recta a cada punto, cuando los puntos están debajo de la recta son negativos y cuando están arriba de la recta los valores son positivos y si están sobre la recta es valor es 0, es neutro”* o *“Estos valores se encuentran cuando se mide las distancias de los puntos a la recta. Son negativos cuando se ubican a la izquierda y como se muestra en la pantalla en la* *parte inferior y positivos cuando se encuentran en la parte superior de la recta. Y las distancias de la parte negativa debe de ser la misma que la suma de las distancias a las derechas"*, observemos que en este último argumento la estudiante relaciona negativo con izquierda y con parte inferior, análogamente positivo con derecha y con parte superior, lo que nos da a pensar que la estudiante está pensando que lo que está ocurriendo acá, en el caso bivariado, es semejante a lo que ocurría en el caso univariado en las primeras actividades.

Dos estudiantes asumieron las diferencias como "el signo" de los valores de los puntos: "Los valores serían que los puntos que están por encima de la recta son positivos y los puntos que están por debajo de la recta son negativos y la recta es neutra" o "Creo que los valores de los puntos en su diferencia es de acuerdo a la recta de donde empieza y donde termina contando hacia atrás en lo último, según. Los puntos ubicados arriba sobre la recta son positivos, lo que están debajo de la recta son negativos y los puntos sobre la recta son neutros = (0)".

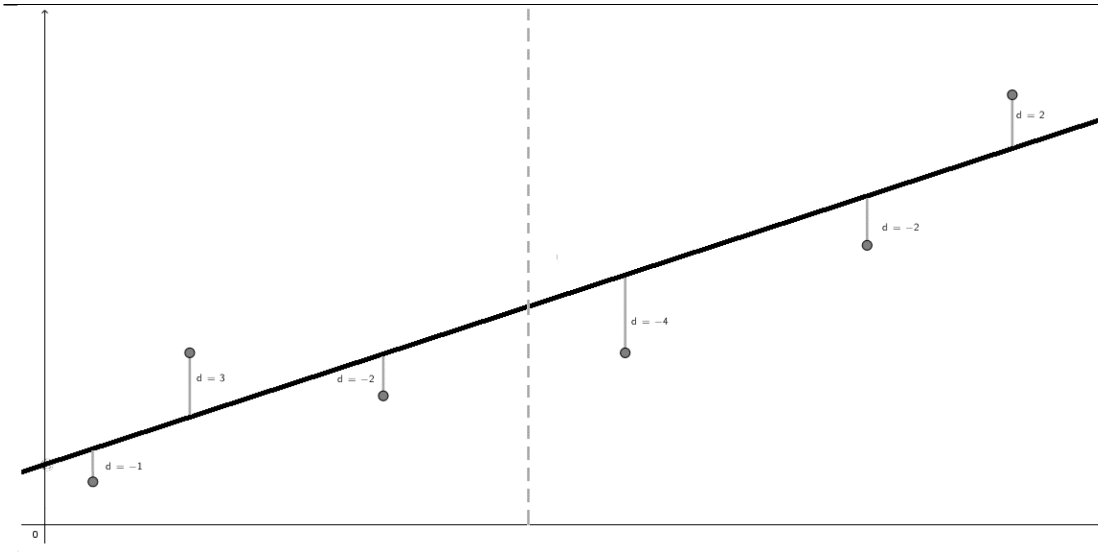
A otro estudiante le es indiferente los términos distancia y diferencia: "Creo que los valores que están por debajo de la recta son negativos y los que están por debajo (encima) son positivos. Creo que los valores pueden ser diferencias o distancias".

En cuanto a la suma de las diferencias, o desviaciones como uno de ellos hace referencia, todos estuvieron de acuerdo en que era cero: "La suma de las desviaciones es cero. Porque arriba es positivo y abajo es negativo y si son los mismos valores su suma es cero".

El siguiente ejercicio consistía en, dada la mejor recta para cierta cantidad de puntos en donde faltaba un punto, observar las diferencias y ubicar el punto faltante. Si la exploración anterior les había dado alguna información, bastaba con sumar las diferencias por encima y por debajo de la recta, comparar y hallar el valor faltante para lograr que la suma diera cero y así ubicar el punto (ver Tabla 53).

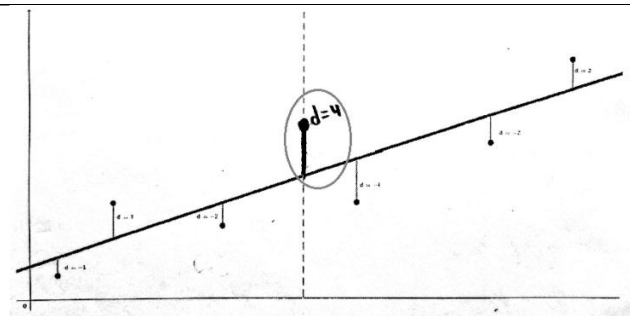
Tabla 53.
Ítem 3d de la Actividad 6.2

Observa el siguiente gráfico en donde se muestra la mejor recta para siete puntos. Como te darás cuenta falta un punto. Ubica sobre la línea punteada el punto faltante y, además, escribe el valor de la diferencia "d" para ese punto. Explica cómo lo encontraste.



Seis estudiantes ubicaron el punto por encima de la recta a una distancia de cuatro unidades, lo que era correcto. En la Tabla 54 se observa cómo procedió uno de ellos; los otros cinco procedieron de manera similar.

Tabla 54.
Respuesta de uno de los estudiantes al ítem 3d de la Actividad 6.2



Argumento: *"El punto faltante es 4 porque al sumar (-1) + (-2) + (-4) + (-2) = -9 y arriba 3+2 = 5 y faltan 4 para llegar a nueve y al sumar -9 + 9 = 0".*

Uno de esos estudiantes continuó hablando en términos de "derecha" e "izquierda", ratificando que considera una relación entre lo que se observaba en el caso univariado y lo que observa ahora en el caso bivariado: *"Lo encontré teniendo en cuenta que las sumas de las desviaciones de la derecha, tanto como en la izquierda debe ser cero (0), el punto lo ubiqué de manera que los puntos de que están ubicados a la derecha debe ser la misma suma de distancias tanto de la izquierda, sino que en la de la izquierda deben ser negativos y de manera que al sumar las desviaciones debe de ser cero (0)"*.

Solo una de los siete estudiantes ubica el punto en otra posición, sobre la recta (ver Figura 74), argumentando que *"Mi ubicación para ese punto es en cero. Su valor es neutro porque quedaría en la mitad de los seis puntos, entre tres y tres y el punto ubicado en cero es el siete (el séptimo punto)".* En la discusión alrededor del por qué procedió de esa manera, ella menciona que no tuvo en cuenta las diferencias, aunque hizo una visualización global de los puntos, su estrategia fue dejar igual número de puntos a lado y lado del nuevo punto, algo así como que este correspondiera a la mediana.

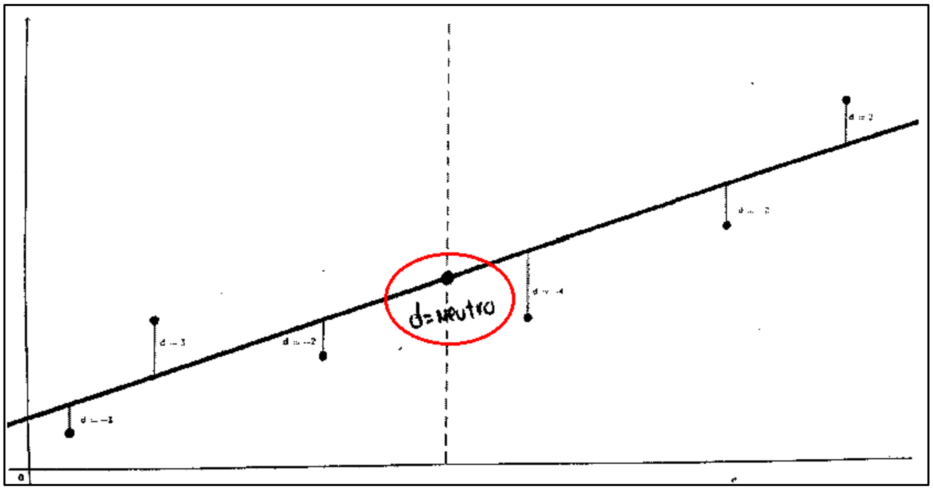


Figura 74. Respuesta de uno de los estudiantes al ítem 3d de la Actividad 6.2

Se podría inferir que hay cierto grado de aceptación con una de las propiedades que debe cumplir la mejor recta, en donde la suma de los residuos es cero, sin embargo sabemos que no es condición suficiente para ser la mejor. En la Figura 75 se observa un contraejemplo.

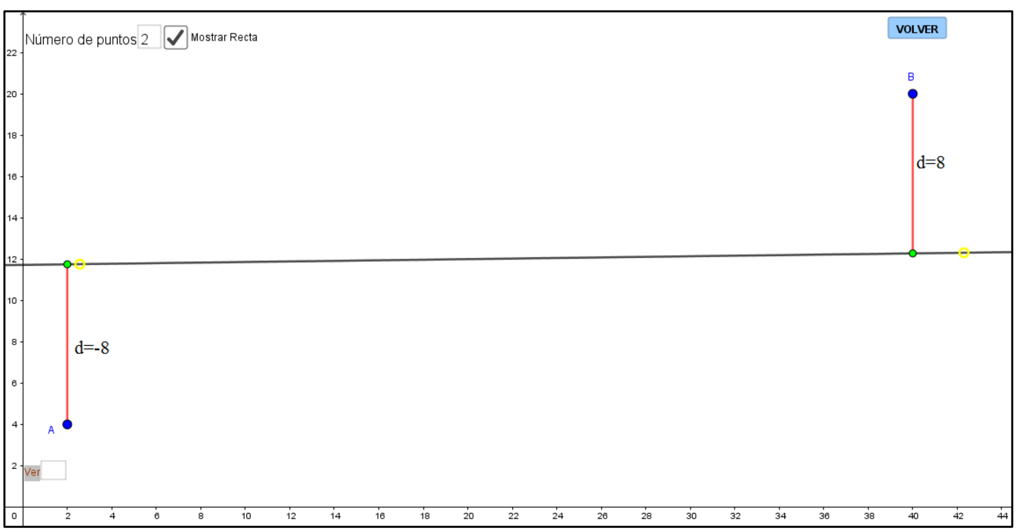


Figura 75. Recta para dos puntos donde la suma de los residuos es cero

Toda la exploración anterior, así como el último ejercicio que acabamos de presentar pretendía darles a entender a los estudiantes de que la mejor recta debe cumplir con ciertas propiedades, como la que acabamos de mencionar, es decir, es como irnos hacia atrás: si la recta cumple con que la suma de las desviaciones es cero, es señal de que vamos por buen camino. Aprovechando el diagrama de la Tabla 54, discutimos que, por ejemplo, la mejor recta en este caso no encajaba dentro de algunas estrategias que ellos habían propuesto como que pasara por el primer punto, o por el primero y el último, o que dejara igual número de puntos a lado y lado, pues, para empezar, tenía un número de puntos impar, pero que sí debía darse que la suma de las diferencias fuera cero.

Continuando con la exploración, hacia la caracterización de la mejor recta, nuestro siguiente objetivo era dar cuenta de una segunda propiedad: “La suma de los cuadrados de las diferencias es la más pequeña”. Para lo anterior, teníamos dos rectas, una movible (de color negro) y una fija (de color verde) que correspondía a la mejor recta. Las diferencias “d” que vimos en la exploración anterior las asociamos con un cuadrado y el objetivo acá era observar cómo variaba el valor de “SumaÁreasCuadrados” cuando la recta movible se movía alrededor de la mejor recta y cómo era ese valor cuando las dos rectas coincidían (ver Figura 76). Recordemos que una exploración muy similar se hizo en el caso univariado en las primeras actividades.

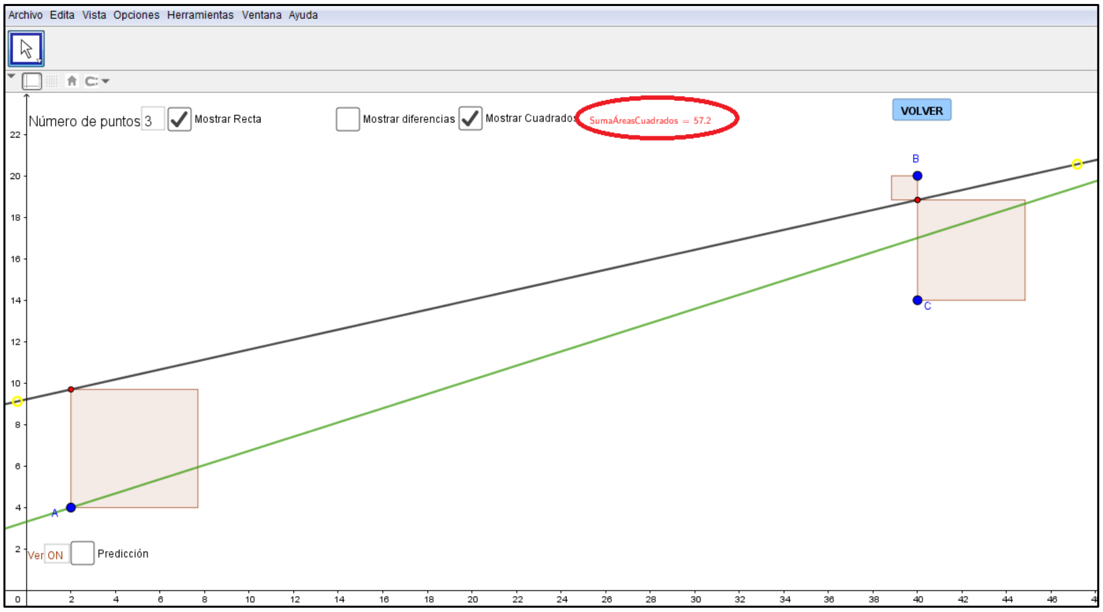


Figura 76. Pantalla de Geogebra asociada a la Actividad 6.1 donde se muestran los cuadrados y la suma de los cuadrados de las distancias de los puntos a la recta

La mayoría de los estudiantes dieron cuenta de que al alejar la recta movible de la recta de mejor ajuste, la suma de las áreas de los cuadrados aumentaba y que cuando coincidían el valor era el más pequeño comparado con otras posiciones, por ejemplo: *"Lo que pasa es lo ste [siguiente]. Si la muevo para arriba el valor de la suma de los valores del área de los cuadrados cambia y lo mismo pasa cuando la muevo hacia abajo. Cuando se alinea con la recta, o coincide el valor cuando coincide con la recta, es el valor más pequeño"* o *"Cuando muevo la recta hacia arriba la suma de los cuadrados aumenta y hacia abajo también aumenta y cuando la recta está sobre la mejor recta el valor es el más pequeño"*.

Después de la exploración anterior alrededor de la recta de mejor ajuste, se les preguntó a los estudiantes sobre qué es lo que hace la mejor recta, la mejor, es decir, ¿qué condición debería cumplir una recta para catalogarla como la mejor recta, o sea, para garantizar que nos encontramos ante la mejor recta? Tres de los estudiantes estuvieron de acuerdo en que la suma de los cuadrados de las diferencias, o que la suma de las áreas de los cuadrados, como aparecía en el archivo de Geogebra, tiene que ser la más pequeña, además dos de ellos agregan cómo debería estar ubicada la recta para darse lo anterior: *"Que la suma de las diferencias al cuadrado es la más pequeña y que la mejor recta tiene el valor más pequeño, está entre los valores"* o *"Que la propiedad que debe cumplir es que en la recta la suma del área de cuadrados sea la más pequeña, que no esté tan alejada la recta de todos los puntos".* Dos se refirieron a la suma de las desviaciones, la cual debe ser o bien cero, o la más pequeña: "Que en la mejor recta la suma de las desviaciones es la más pequeña" o "La mejor recta: Que esté entre los valores, que esté cerca de los puntos y que la suma de los valores sea cero", pero lo anterior podría darse no solo en la mejor recta.

Aunque hasta el momento los estudiantes pudieron evidenciar dos propiedades que cumple la mejor recta, no es natural que los estudiantes piensen que, por ese hecho, nos encontramos frente a la recta más cercana a todos los puntos. Ya vimos que varios hablan en términos de cercanía ¿pero qué significado le dan los estudiantes al término "cercanía"? Por ejemplo, si se tiene un diagrama de dispersión, en donde se ha trazado la mejor recta ¿Qué tendría en cuenta para catalogarla como tal? A continuación una discusión al respecto:

I: Si alguien tiene una nube de puntos así, y le dice "para mí la mejor recta en esa nube de puntos, es esta" Entonces Silvia, ¿usted cómo diría que es la mejor recta?

S: Porque está entre los puntos.

I: Está entre los puntos, qué más. Porque si ese fuera el único criterio, yo puedo trazar otra… Esa sigue estando entre los puntos ¿o no?

A: Sí

I: Entonces tengo que ser más detallista, qué más puedo mirar.

A: Que esté más cerca a todos los puntos.

I: ¿Y cómo mido yo que esté más cerca a todos los puntos?

MF: Que los que estén alejados sean mínimos.

I: Cómo así, Mafer.

MF: Que estén lejos… o sea que esté más cerca de los puntos. Por lo menos, yo creo que ahí no porque los tres puntos están más abajo, o sea solo está cerca de dos puntos.

I: Entonces dónde

MF: Yo creo que más bajo…

I: ¿Usted está mirando ahí que haya una cantidad de puntos cerca de la recta?

MF: Sí… O sea, que queden pocos puntos alejados.

I: Por ejemplo, ahí tengo tres cerca y dos alejados…

MF: Y está entre los puntos.

I: Wendy usted qué diría… ¿está de acuerdo con la recta?

W: [Niega]

I: Por qué

W: Pues porque la suma de las distancias debe que ser… de la parte de debajo, de la izquierda, son… deben que ser iguales a las de arriba.

En la anterior discusión se muestra, como lo manifestaron dos estudiantes, que, en primer momento, un criterio para que sea la mejor recta es que ésta se encuentre entre los puntos, y, directamente como lo dice otra estudiante, que esté más cerca de todos los puntos. Posterior a eso, se evidencian dos formas de asumir esa cercanía: observando que las distancias por encima y por debajo de la recta sean las mismas, o que la cantidad de puntos alejados de la recta sea mínima. Evidentemente la primera forma, además de que evidencia una mirada global sobre todos los puntos, produce un valor que es fácilmente verificable; es decir, este es un procedimiento algorítmico que conlleva una respuesta sin dudas. La segunda forma, sigue siendo intuitiva y poco práctica porque no es para nada claro el significado de “puntos alejados” ni que la cantidad de ellos sea mínima.

¿Cómo predecir teniendo la mejor recta?

Finalmente, retomamos la cuestión sobre cómo predecir teniendo la mejor recta. Seguramente en un curso con conocimientos previos sobre la recta y la ecuación de la misma, esta tarea podría no resultar difícil, inclusive podría ser muy obvia para algunos; pero como ya lo hemos mencionado, nuestros estudiantes no tenían conocimientos previos al respecto. Veremos que ninguno de los estudiantes, asume, como única forma, ubicar el punto sobre la recta:

Cuatro estudiantes señalan las siguientes formas: i) En la recta o cerca de la recta; ii) El valor de la variable respuesta es el mismo para el valor de la variable explicativa; iii) Nunca sobre la recta y iv) Cerca de la recta. Los otros tres estudiantes, hacen referencia a la relación entre las variables y a las propiedades de la recta: *“Me ayuda a encontrar la relación entre las dos variables ya sea ascendente o descendente, pero que esta relación cumpla con las propiedades para encontrar la mejor recta”* o *“Me ayuda porque gracias a esto puedo tener en cuenta que la suma de las diferencias es cero y que la suma de las diferencias al cuadrado es la más pequeña”.*

Los estudiantes generaron argumentos encontrados entre su concepción de la mejor recta que no tiene por qué pasar forzosamente por alguno de los puntos y la idea de cercanía: es como si el valor predicho también se considerara como uno de los puntos iniciales, lo que lo obliga a no afectar la “suma cero” pero al mismo tiempo sin el ideal de “cercanía cero”.

Las anteriores formas se reflejaron al momento de predecir sobre un diagrama de dispersión con diez puntos. Por ejemplo, se les pedía predecir un valor de y para $x = 20$, en la Figura 77 se resumen las predicciones donde dos estudiantes dijeron que 11, dos 10, y los demás, 4, 14 y 12. El valor sobre la recta era 11,6. La estudiante que más se aproximó dijo que 12 y, según argumenta, lo ubicó tan cerca como pudo de manera que no afectara tanto en que la suma de las desviaciones fuera cero: *“El valor de Y=12 de manera que la suma de las diferencias sea cero (0). Para que no afecte la suma de las diferencias sea cero”*. De manera que ningún estudiante lo ubicó sobre la recta, sus estrategias se enfocaron más en conservar la relación entre las variables y ubicarlo cerca de la recta, a excepción de la que dijo 4 aunque así fuera su intención: *“Mi valor dado es 4 (20,4), busco que mi valor quede cerca a la recta”*.

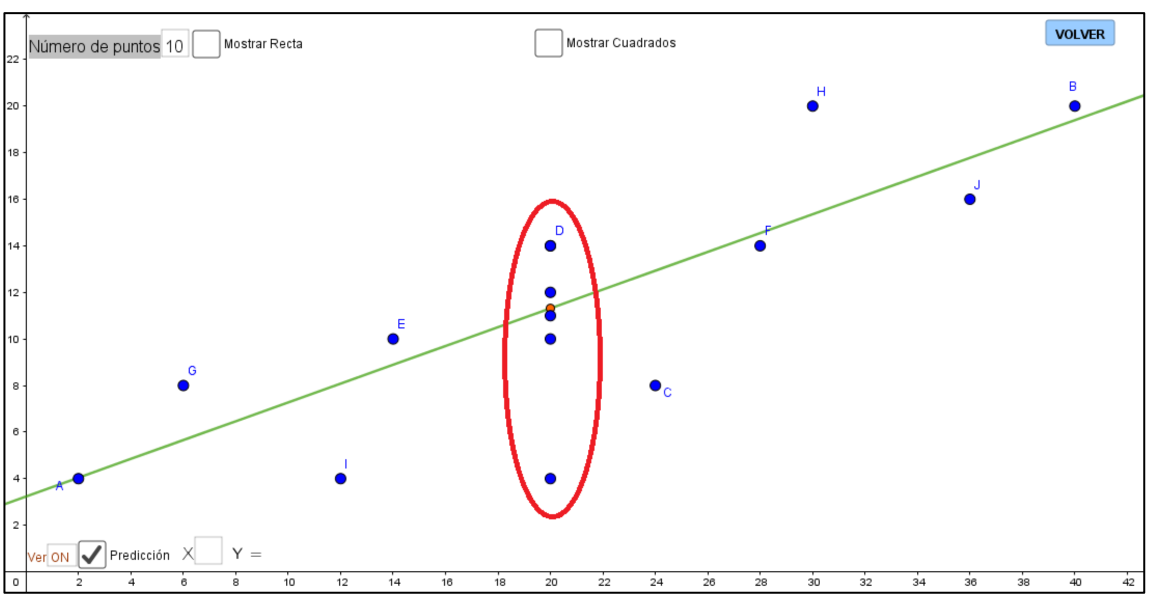


Figura 77. Predicciones de los estudiantes dada la recta de mejor ajuste

4.11 Análisis prueba final

A continuación se presenta el análisis de la prueba final aplicada a los siete estudiantes de octavo grado que participaron en el desarrollo de una THA alrededor del análisis de datos bivariados, hacia la recta de mejor ajuste. La prueba constaba de siete ítems, los primeros seis de opción múltiple y el último abierto. Los primeros tres ítems se relacionaban con datos univariados y los cuatro últimos con datos bivariados. La prueba se realizó en una sala de cómputo en donde cada estudiante tenía a su disposición un computador y acceso a programas como Excel, Fathom, Geogebra y la calculadora, si lo requerían o deseaban usar. El tiempo estimado para contestar la prueba fue dos horas.

Ítem 1.

Para el primer ítem se daba una serie de mediciones y se pedía predecir la siguiente medición (ver Tabla 55).

Tabla 55.
Ítem 1 asociado a la prueba final

1. La concentración de determinadas sustancias en la sangre influye en la salud de las personas. He aquí las mediciones del nivel de fosfatos en la sangre de un paciente que realizó seis visitas consecutivas a una clínica, expresadas en miligramos de fosfato por decilitro de sangre.

5.7	5.2	4.6	4.9	5.7	6.4

Predice el nivel de fosfatos en la sangre, del mismo paciente, para su próxima visita.

a. 6.4
b. 5.4
c. 4.6
d. 5.7

De los siete estudiantes, seis responden b y uno c. De manera que seis estudiantes asumen la media como un buen predictor, reflejando así una concepción global de los datos: *"El nivel de fosfato lo predije fue de 5.4 porque hallé la media para dar un valor acertado sobre cuál sería el nivel de fosfato para ese mismo paciente, en una próxima consulta"*, al justificar "un valor acertado" podría pensarse que no está teniendo en cuanta la variabilidad puesto que no se puede asegurar 100% que así ocurrirá. Otra respuesta fue la siguiente: *"Para predecir el nivel de fosfato del próximo paciente, lo que realicé fue, tomar los 6 valores que me dieron los sumé y los dividí en 6 para poder sacar un solo valor y tener un cálculo para el próximo paciente. Le media me ayudó a predecir el valor del próximo paciente ya que es un solo valor y la media sirve solo para eso".* Aquí vemos que el estudiante reconoce que, al proceder de esa manera, no solo está teniendo en cuenta todos los valores sino que los está resumiendo a un solo valor que lo ayuda a predecir.

La estudiante que contestó c también considera la media como una forma para predecir, sin embargo su cálculo no fue el correcto puesto que dividió en siete, en lugar de seis, es decir, dividió en la próxima medición: *"El nivel de fosfatos en la sangre para su próxima visita será 4.6 porque los anteriores resultados sumados y divididos en siete da 4.6 y me permite predecir".*

El segundo y tercer ítem requería la comprensión de características y propiedades de la media (ver Tabla 56).

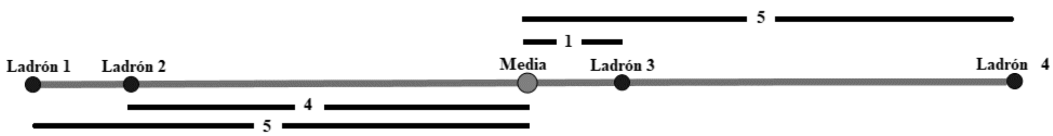
Tabla 56.
Ítems 2 y 3 asociados a la prueba final

Una banda de ladrones entre los 25 y los 30 años, de edad, ha sido capturada por la policía. El equipo de investigación sabe que la banda está integrada por cinco ladrones, incluyendo al jefe y que el nombre de la banda hace honor al valor de la media de la edad actual de sus integrantes.

2. ¿Cuál de las siguientes opciones NO podría ser el nombre de la banda?

a. “Los veintiséis”
b. “Los veintiocho”
c. “Los treinta”
d. “Los treinta y dos”

3. Como broma, la policía ha ubicado a cada ladrón en una celda a cierta distancia de otra celda que tiene como número la media de la edad actual de sus integrantes, incluyendo al jefe. En la siguiente imagen se observa las celdas de los cuatro ladrones menos la del jefe.



¿Dónde está ubicada la celda del jefe?

a. Cuatro unidades a la derecha de la media
b. Una unidad a la izquierda de la media
c. Tres unidades a la derecha de la media
d. Tres unidades a la izquierda de la media

Ítem 2

Para el ítem 2, todos acertaron en la respuesta, que es la d: “Los treinta y dos”. Entre los argumentos, se identifica claramente que dos estudiantes comprenden una característica de la media, la cual es que es un valor que se encuentra entre el valor más pequeño y el más grande de los dados: *“No podría ser el nombre de la banda "los treinta y dos" porque la media debe estar entre los valores y 32 no está entre 25 y 30 pero 26, 28 y 30 sí está entre 25 y 30 entonces 32 no puede ser el nombre”* o *“¿Por qué? Porque una característica de la media es que está entre los valores, en este caso de 25 a 30. Por lo tanto el nombre "los treinta y dos" no sería apropiado para este grupo de ladrones, porque no está entre los valores”.*

Cuatro estudiantes basan sus argumentos en que 32 no está entre 25 y 30. Uno de ellos argumenta de la siguiente manera: *"El nombre de la banda no puede ser treinta y dos porque dice que la edad de los integrantes está entre los 25 y los 30, así que el nombre de la banda no debe ser mayor a 30 ni menor que 25"*. Uno de ellos considera que 32 es valor extraño, un valor atípico y por esa razón descarta que ese pueda ser el nombre de la banda: *"Sobre el nombre de la banda no podría ser "los treinta y dos" ya que ese valor no se encuentra entre los 25 y los 30 años y por eso ese valor es un valor atípico, es decir muy extraño"*.

De igual manera una estudiante reconoce que 32 es un valor muy grande para la media, específicamente, la que ella calculó, que consistió en promediar los dos valores 25 y 30. Lo anterior pudo darse por una mala interpretación del enunciado o una incomprensión de la media: *"Para mí sería la d, porque ya sería un valor muy grande para la media que hallé la cual fue 27, que fue hallar la media entre 25 y 30 y luego multiplicarlo por 5 la cual es la cantidad de ladrones y luego dividirlo en 5"*. También da a entender que la media que ella calculó la corresponde con la edad de todos los integrantes, aproximadamente 27.

Ítem 3

En cuanto al ítem tres, se requería comprender una propiedad de la media: La suma de las desviaciones a la media es cero o, dicho de otra manera, la suma de las distancias a la media a la derecha, es la misma a la izquierda, de manera que la respuesta correcta era la c. Cuatro estudiantes respondieron acertadamente, de los cuales tres muestran en sus argumentos la comprensión de la propiedad mencionada, por ejemplo: *"A tres unidades a la derecha de la media porque las distancias a la derecha y a la izquierda deber ser iguales al sumar cinco más cuatro da nueve y al sumar cinco más uno da seis entonces para que las distancias sean* *iguales a la derecha y a la izquierda faltan tres unidades a la derecha de la media para que se cumpla la propiedad”.*

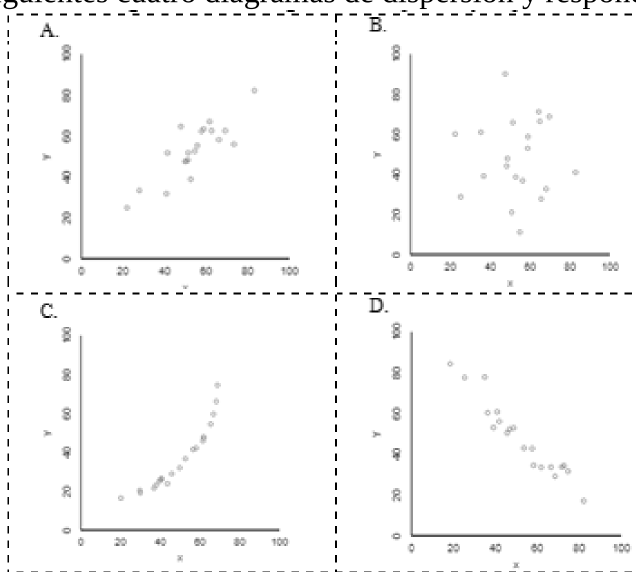
El cuarto y quinto ítem requerían la interpretación de dos conjuntos de diagramas de dispersión, sin contexto.

Ítem 4

El cuarto ítem estaba enfocado hacia la intensidad de relación entre las dos variables (ver Tabla 57).

Tabla 57. *Ítem 4 asociado a la prueba final*

Observa los siguientes cuatro diagramas de dispersión y responde



4. Lee la siguiente afirmación y responde: *“En los gráficos A, B, C y D, se puede decir que las variables x e y se relacionan de alguna manera”* Ordena los gráficos de mayor a menor, de acuerdo al nivel de seguridad con que se puede dar esa afirmación.

a. A-C-B-D
b. A-D-C-B
c. C-A-D-B
d. C-D-A-B

En la Tabla 57 se puede apreciar que el orden de los gráficos, de mayor a menor intensidad, con que se relacionan las dos variables es C-D-A-B que, como se observa, no todos son de forma lineal, como C y B. Durante el desarrollo de la THA, específicamente en la parte bivariada, los gráficos sobre los que se trabajaron o bien mostraban una relación lineal, positiva o negativa, o no se relacionaban, solo, en una ocasión, se permitió una discusión sobre un gráfico como C, de manera que quisimos ver cómo asumían este tipo de relación, poniendo, de manera intencional, al gráfico C con la mayor intensidad de relación.

Cuatro de los siete estudiantes escogen la opción correcta, la d. Por sus argumentos se puede concluir que, para tal elección, toman dos consideraciones: i) Que se refleje la forma de la relación, ascendente o descendente y, ii) La claridad del "orden" de los puntos. Resaltamos los siguientes dos argumentos: *"La "D" porque se puede ver con claridad el orden de los puntos más ordenados sea de forma ascendente o descendente"* y *"Ordenados de mayor a menor de acuerdo al nivel de seguridad quedan primero C porque si hay una relación ascendente que está clara, segundo D porque hay una relación descendente que también está clara, tercero A porque hay una relación ascendente que también se ve un poco reflejada y por último el gráfico B porque no hay alguna relación"*.

La siguiente opción con mayor frecuencia fue la b, A-D-C-B. Según los argumentos, se podría pensar que estos estudiantes estaban buscando una relación de tipo lineal, como se refleja en los gráficos A y D, siendo el de mayor intensidad el A. Uno de ellos no asume que el C tenga alguna relación pero que, en definitiva, el de menor relación es el B: *"El orden sería A, D, C y B porque en los gráficos A y D se puede ver con más claridad su afirmación ya sea de forma ascendente y descendente y en los otros dos gráficos no se puede ver la relación entre las dos variables"*. Mientras que otro señala que en el C se intenta dar una relación ascendente: *"De esta forma se ve la relación que hay entre X y Y en el A se ve la relación que se quiere mostrar la cual es ascendente, en la D. la forma descendente, la C. porque ahí se trata de dar una forma ascendente y en la B ya se pierde esta relación"*.

Por último, un estudiante escoge la opción C, C-A-D-B, por lo que, como se observa, le está dando prioridad a los gráficos que manifiestan una relación ascendente y, poca aceptación a los que presentan una asociación negativa, como el gráfico D.

En general, si omitimos por un momento el gráfico C en la Tabla 57, se pueden destacar dos cosas: La primera es que todos los estudiantes comprenden cómo debería ser un gráfico de dispersión para dos variables que no se relacionan y la segunda es que todos, a excepción de uno, aceptan el tipo de asociación negativa que, como hemos visto en la revisión bibliográfica, Estepa y Batanero (1995), este tipo de asociación difícilmente la asumen los estudiantes.

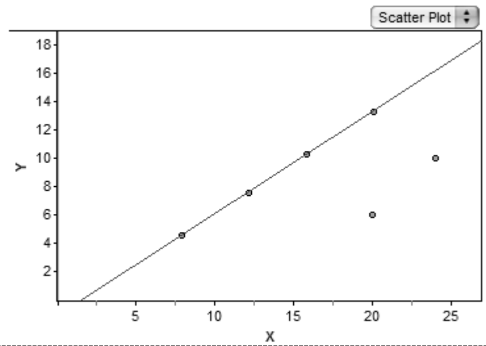
Ítem 5

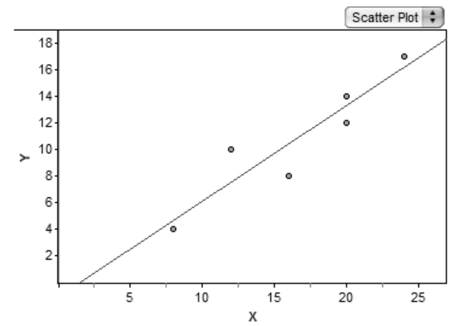
El quinto estaba enfocado hacia la recta de mejor ajuste (ver Tabla 58), en el siguiente sentido: dada la recta de mejor ajuste, asignar la nube de puntos.

Tabla 58.
Ítem 5 asociado a la prueba final

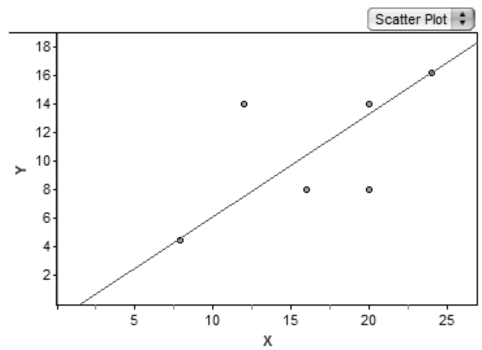
5. En los siguientes gráficos se muestra un conjunto de seis puntos y una recta que es LA MISMA para todos los gráficos, la cual refleja el comportamiento de los seis puntos. Ordena los gráficos de acuerdo al nivel de ajuste de la recta a los seis puntos, es decir, el primer gráfico que escojas será el gráfico en el cual la recta se ajusta mejor a la nube de puntos dada y el último que escojas, la que menos se ajusta.

A.	B.

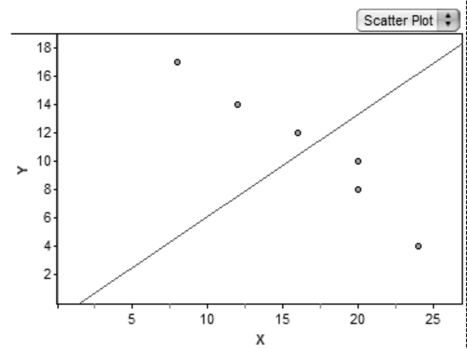




C.



D.



a. B-D-C-A
b. B-C-D-A
c. A-C-B-D
d. B-C-A-D

Los gráficos de la Tabla 58, fueron elaborados de acuerdo a las estrategias que ya los estudiantes habían manifestado en actividades anteriores, específicamente, en la de “Encontrando la mejor recta”, por ejemplo: Que pase por el primer y último punto, que deje igual número de puntos a lado y lado de la recta, más cerca de todos los puntos y, que pase por el primer punto. Con este ítem quisimos ver cómo esas concepciones habían cambiado después de la socialización con sus compañeros.

Cuatro de los estudiantes acertaron en la respuesta, d. B-C-A-D, dentro de los criterios para hacer su elección están los siguientes:

- Por los puntos o cerca de ellos: *“Porque tuve en cuenta que pasara por los puntos o cerca a ellos pero que no estuviera tan lejos como en el caso del gráfico D*”. Como se observa este estudiante considera dos cosas para ajustar la nube de puntos: Cantidad de puntos sobre la recta, y cercanía de los puntos. Pero es claro que su prioridad es la segunda consideración, la cercanía.

- Que refleje la relación que muestra la recta: *“La d. porque esta forma la nube de puntos en cada caso hay una relación con la recta en el B se ve una mejor relación de los puntos con la recta. En la C. también se ve esa relación de la recta con los puntos, en la A se alcanza a ver un poco la relación con la recta pero en la D no se ve la relación que estoy tratando de buscar como en los demás casos”*. Es decir, los puntos se deben corresponder con la recta, en el sentido de que deben encaminarse hacia la relación que señala la recta.

- Que esté entre los puntos, que haya una cercanía y que se cumpla la relación: Esta estudiante utiliza más de un criterio para hacer su elección.

1. En primera instancia la recta debería estar entre los puntos.
2. Como segundo criterio, para decidir el orden entre A y C, está la cercanía y,
3. Finalmente, para justificar por qué dejar al final a D, utiliza el criterio de relación, es decir, tanto la nube de puntos como la recta deberían reflejar la misma relación.

Tal como lo argumenta en el siguiente párrafo:

“En el gráfico que más se ajusta la recta es en el gráfico B porque es una recta que está entre los puntos y además cumple la relación ascendente, el segundo gráfico sería C porque está entre los puntos cumple la relación pero es el segundo porque los puntos no están tan cerca a la recta como en el gráfico B, luego sigue en el gráfico A porque cumple la relación ascendente pero no está entre los puntos y además se aleja de dos puntos y en

el gráfico que menos se ajusta la recta es en el b porque en primer lugar no está cumpliendo la relación de los puntos la cual es descendente y la recta está ascendente".

• Otra estudiante hace alusión, explícitamente, a propiedades y características de la media con el ánimo de extenderlas a la recta de mejor ajuste, usando así también varios criterios: *"Porque de esta manera la nube de puntos se ajusta mejor a la recta. El "B" es el primer gráfico porque cumple con las propiedades de la media, es decir, la recta está entre los valores, está cerca de los puntos etc. La D es la última que está dentro del orden porque me parece que no se ajusta y no está tan cerca de los puntos, no se ve una relación clara*".

Como pudimos ver, el ítem 5 era el camino de vuelta a las actividades que habíamos hecho hasta el momento respecto a ubicar la recta de mejor ajuste. En dichas actividades los estudiantes de alguna u otra manera caracterizaron la recta de mejor ajuste y, valiéndose de esa caracterización, asignaron la nube de puntos a la que sería la mejor recta. Dentro de esas características están: Que los puntos estén cerca de la recta o sobre la recta; que los puntos reflejen la relación que muestra la recta, o que se ubiquen de tal manera que la recta esté entre los puntos.

Los ítems 6 y 7 se relacionaban con predicción en modelos bivariados. Para los ítems seis y siete se mostraba una situación y una tabla de datos con dos variables, altura de lanzamiento y altura del primer rebote. (ver Tabla 59).

Tabla 59.

Ítems 6 y 7 asociados a la prueba final

Se deja caer una pelota de goma desde diferentes alturas y se mide la altura que alcanza en el primer rebote. Las mediciones se muestran en la siguiente tabla.

Altura de lanzamiento [cm]	Altura del primer rebote [cm]
12	5
26	12
32	22
43	23
50	26
61	37
66	48

6. Juan, Camilo, Pedro y Pablo desean lanzar nuevamente la pelota desde una altura de 45 cm y cada uno de ellos ha dado una predicción de la altura del primer rebote. Juan dice que 27 cm, Camilo 24 cm, Pedro 5 cm y Pablo 48 cm. ¿Con cuál de los cuatro estás más de acuerdo?

a. Juan
b. Camilo
c. Pedro
d. Pablo

7. Pregunta abierta. Predice la altura del primer rebote si la misma pelota se deja caer desde las siguientes alturas: a) 20 cm; b) 35 cm; c) 55 cm y d) 60 cm.

Quisimos representar los datos de manera tabular, con el ánimo de que ellos, sin indicarles, hicieran un diagrama de dispersión, sin embargo, ningún estudiante lo realizó. De manera que sus argumentos se basaron netamente en la tabla.

Ítem 6

Por otra parte, si los datos del ítem 6 en la Tabla 59 se muestran en un diagrama de dispersión, como en la Figura 78 (puntos en gris), con la respectiva recta de mejor ajuste y las opciones de respuesta (puntos en azul), vemos que la opción correcta es a. Juan y que una buena aproximación sería la b. Camilo.

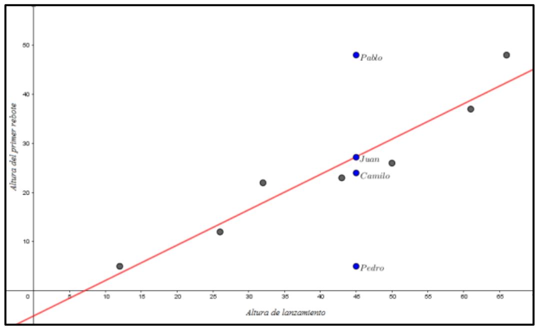


Figura 78. Diagrama de dispersión asociado a los datos del ítem 6 de la Tabla 59

Solo una estudiante acertó con la respuesta a. Juan, argumentando de la siguiente manera: *"Sumé 12+26+32+43+50+61+66=296/7=41 y lo dividí por el número de valores dados lo que me dio =41 Por lo cual escogí a Juan que le falta solo 14 a diferencia de Camilo que le falta 17 veces, Pedro que le falta 36 veces y Pablo que se pasa del valor 7 veces lo cual se acerca más Juan aunque le falte su teoría es mejor que la de los demás"*. Como se observa, la buena respuesta fue una coincidencia feliz. Es decir, la estudiante promedió valores, hecho que refleja una especie de fijación sobre la media como estrategia de solución salvadora. El hecho de promediar los valores de la variable explicativa, Atura de lanzamiento, es decir, de fijar su atención en una sola variable, refleja la ausencia de coordinación entre las dos variables, que caracterizaría un razonamiento covariacional.

La respuesta con mayor frecuencia fue la b. Camilo, dada por cinco estudiantes. Como se observa en la Figura 78 es la más cercana a la recta, de manera que es una buena aproximación en comparación con las otras dos. Veamos cómo lo justifican estos cinco estudiantes.

Dos estudiantes toman en cuenta las dos variables y su estrategia se basa en mirar los valores vecinos, de dos maneras diferentes:

i) El promedio de los vecinos: *“La b. porque observando la tabla en la parte de altura y lanzamiento y miré dos valores que estuvieran cerca de (45cm) y los cuales escogí 43 cm y 50 cm. Luego miré las alturas del primer rebote en esos dos casos hallé la media entre 26 y 23 y el resultado me dio 24.5 entonces la respuesta que más se acerca a esto es la de Camilo con la predicción de 24 cm”.*

ii) Un valor entre los vecinos: *“Estoy de acuerdo con "Camilo" porque según la tabla un valor para el rebote de una altura de 45 cm de la pelota sería entre 23 y 26 cm y Camilo dice 24 cm y es una predicción que se ajusta porque está entre los valores y Juan, Pedro y Pablo no cumple la propiedad de que esté entre los valores”.*

La estrategia de observar los vecinos cercanos estuvo siempre presente desde la aplicación de las pruebas diagnósticas y aunque esta estrategia se conservó durante el desarrollo de algunas actividades, cabe resaltar que esta estrategia contemplaba otros aspectos cuando se trataba de predecir sobre un diagrama de dispersión, por ejemplo, la relación o forma que seguían todos los puntos. Sin embargo, ante la falta del gráfico los estudiantes perdieron la idea de colectivo que la tabla en sí misma posee. Volviendo así a la antigua estrategia de considerar solamente los “vecinos más cercanos” para hacer la predicción. Muy probablemente este hecho hubiera cambiado si se les hubiera condicionado a graficar los datos en un diagrama de dispersión.

Por otra parte, los otros tres estudiantes promedian los valores de la variable respuesta, Altura del primer rebote, reflejándose, una vez más, la no coordinación entre las dos variables, por ejemplo: *"Para escoger la mejor predicción tuve que realizar o encontrar la media es decir, escogí todos los siete valores de la altura del primer rebote los sumé todos y los dividí en siete y me dio como resultado final 24 cm, por eso he escogido la respuesta de Camilo".* Otra vez la media pero con la variable dependiente.

Como se vivió en algún momento, durante el desarrollo de las actividades, la estrategia de promediar los valores de una sola variable, para hacer una predicción, se previó. Por tal razón, en el último ítem, se pedía predecir para cuatro valores más:

"Pregunta abierta. Predice la altura del primer rebote si la misma pelota se deja caer desde las siguientes alturas: a) 20 cm; b) 35 cm; c) 55 cm y d) 60 cm".

Como sabemos, si continuaban con la misma estrategia de sumar y dividir entre el número de datos, la predicción siempre iba a ser la misma, lo que no tendría mucho sentido si se comprende la relación que hay entre las dos variables: A mayor altura de lanzamiento, mayor altura del primer rebote. Por otra parte, para este ítem, todos los estudiantes, nuevamente, solo trabajaron sobre la representación tabular de los datos.

A continuación presentamos solo las estrategias de los estudiantes que tuvieron en cuenta, ya fuera implícita o explícitamente, la coordinación entre las dos variables para hacer sus predicciones, es decir, que en sus predicciones se reflejara la relación entre las dos variables. Estos estudiantes fueron tres y sus estrategias son las siguientes:

i) La Media entre los dos vecinos más cercanos: Dos estudiantes tomaron los valores más cercanos a lado y lado del que iban a predecir y calcularon la media. De manera que sus predicciones para 20, 35 y 55 fueron las mismas:

"a) 20 cm. Primero tendría en cuenta las alturas de lanzamiento y escogería las dos que estén más cerca a 20 cm en este caso 12 y 26cm. Luego hallaría la media de esos dos valores y sería 8.5 este sería el valor para calcular la altura para ese primer rebote de 20 cm de altura…".

En la Figura 79 se muestran sus predicciones junto con la recta de mejor ajuste para los datos del ítem 6 de la Tabla 59.

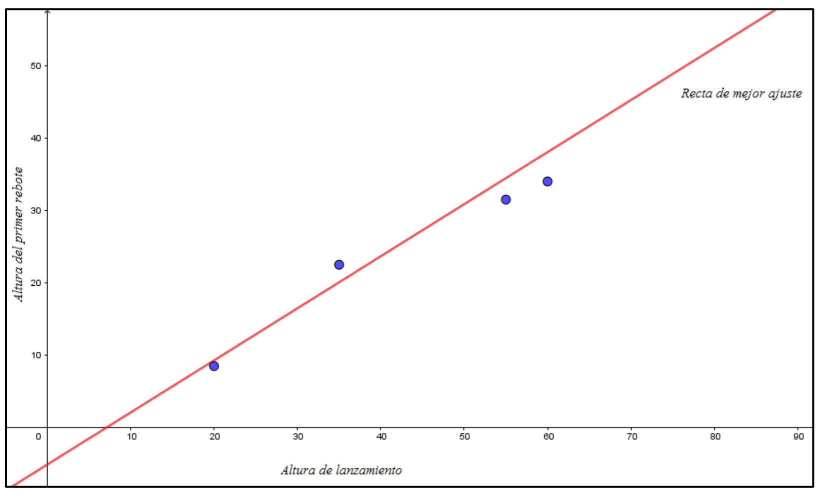


Figura 79. Predicciones de un estudiante bajo la forma "La Media entre los dos vecinos más cercanos" y recta de mejor ajuste

"a) 20 cm: El rebote sería 8.5 cm porque sería una predicción adecuada porque, primero la segunda variable que es el rebote aumenta a media que la primera variable aumenta y además, sumé 12 más a 5 y me da 17 dividido en 2 da 8,5 y 12 y 5 son los valores de la altura del lanzamiento desde 12 y 26 cm, escogí estas dos alturas porque 20 cm está entre 12 cm y 26 cm…".

En la Figura 80 se muestran sus predicciones junto con la recta de mejor ajuste para los datos del ítem 6 de la Tabla 59.

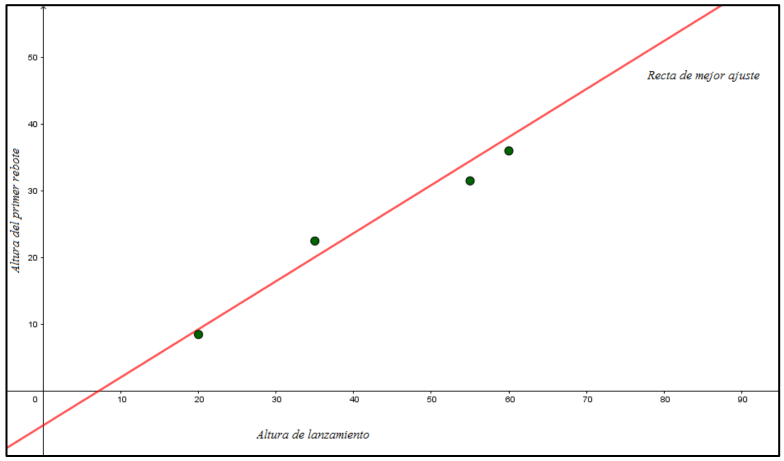


Figura 80. Predicciones de un estudiante bajo la forma "La Media entre los dos vecinos más cercanos" y recta de mejor ajuste

Como se observa, las dos estudiantes están teniendo en cuenta las dos variables. Inclusive en el último argumento, la estudiante enuncia la relación entre las dos variables: "*la segunda variable que es el rebote aumenta a media que la primera variable aumenta*".

Por otra parte, sus predicciones para 60 variaron, ya que una tomó la predicción que hizo para 55 y la altura del primer rebote de 61, en la tabla, y promedió:

"d) 60 cm. Primero escojo los valores que estén más cerca de 60 cm. En este caso podría escoger el valor que predije para 55 cm el cual está cerca de 60 cm y el de 61 cm luego hallaría la media que sería 34 este sería el valor para la altura del primer rebote de 60 cm de altura" (ver Figura 79).

Mientras que la otra estudiante dio un valor teniendo en cuenta la relación:

"d) 60 cm: El rebote sería 36 cm porque es una predicción que cumple la relación de la segunda variable aumenta cuando la primera variable aumenta" (ver Figura 80).

ii) Aproximaciones: Otra estudiante hizo aproximaciones, teniendo en cuenta que sus predicciones estuvieran entre el menor y mayor valor de la tabla y que además sus predicciones respetaran una relación, aunque no lo hace explícito:

"Mis procedimientos serían aproximaciones, pero que mis valores estén entre los valores de acuerdo a la tabla del punto 6 sobre la altura del primer rebote que está en unos valores de 5 a 48. Para: a) 20 cm: Aproximadamente 16 cm. Para: b) 35 cm: Aproximadamente 24 cm. Para: c) 55 cm: Aproximadamente 32 cm. Para: d) 60 cm: Aproximadamente 41 cm".

En la Figura 81 se muestran sus predicciones junto con la recta de mejor ajuste para los datos del ítem 6 de la Tabla 59.

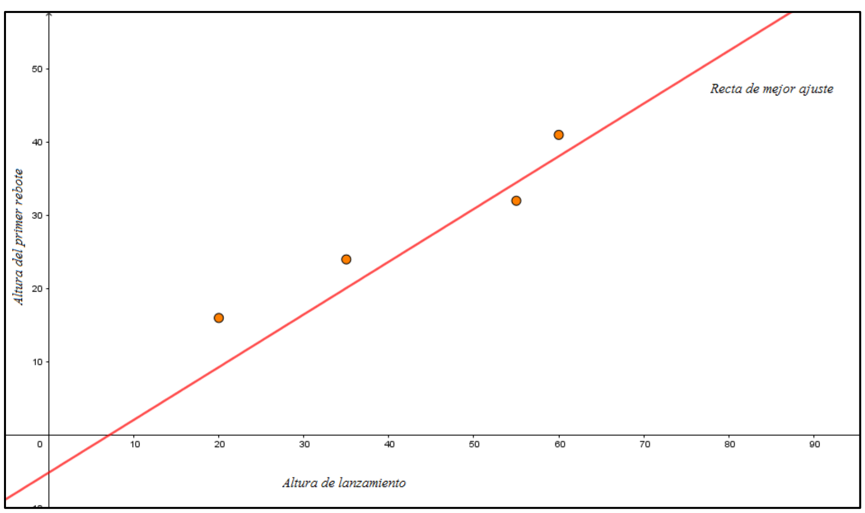


Figura 81. Predicciones de un estudiante bajo la forma "Aproximaciones" y recta de mejor ajuste

Como es claro, las predicciones de los estudiantes se basaron en la representación tabular de los datos y se limitaron, básicamente, a la consideración de los vecinos cercanos. Para indagar cómo predicen cuando la representación de los datos se les presenta en un diagrama de dispersión (ver Figura 82 puntos en azul), se realizaron entrevistas con cada uno de ellos.

Por ejemplo, la siguiente conversación se hace sobre la Figura 82 con uno de los estudiantes que fijó su atención en una sola variable, “Altura del primer rebote”, para hacer sus predicciones (puntos rojos Figura 82).

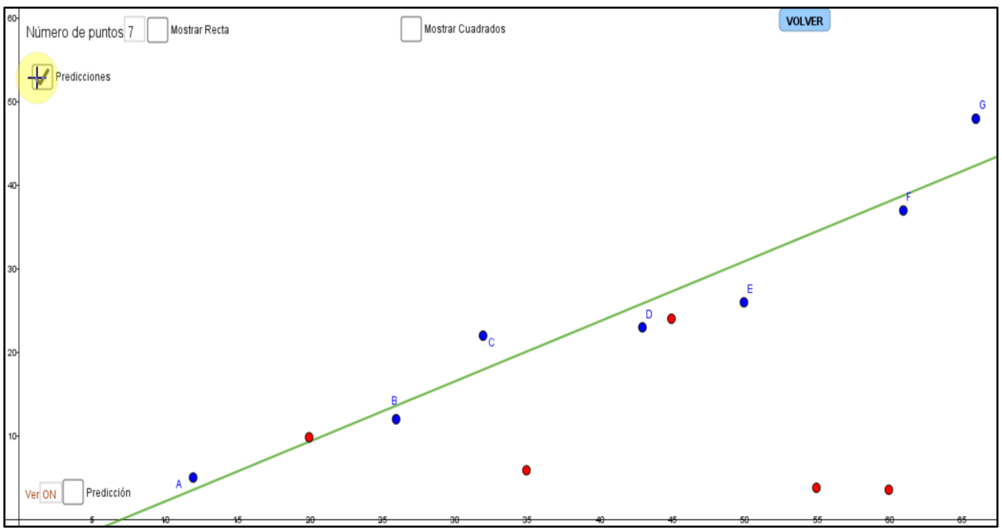


Figura 82. Predicciones de un estudiante del ítem 8 de la prueba final, sobre un diagrama de dispersión.

I: Ah, listo… Entonces mire, resulta que esto que usted ve acá, esos puntos azules que usted ve ahí… son los de la prueba.

E: Ujum

I: Es decir, estos puntos que usted ve acá son los de los datos del punto seis de la prueba final… estos (señalando la tabla de la prueba final)

E: Ujum

I: Y usted hizo unas predicciones… pero sus predicciones las hizo fue sobre la tabla

E: Sí

I: Y sus predicciones ya ubicándolas sobre el diagrama son estas (Figura 82 puntos en rojo)

I: Cómo las ve

E: (Risa) Yo qué culpa, profesora. Quedó mal…

I: Por qué quedó mal, a ver.

E: Lo uno, hay muchos datos atípicos… (señalando los tres de abajo en la Figura 82)

I: Ajá

E: Hay tres datos atípicos

I: Por qué son atípicos

E: Porque, porque están por fuera de lo cercano que es la línea, prácticamente

I: Y si tuviera la oportunidad de moverlos… si quiere muévalos hasta donde sea la mejor predicción (ver Figura 83 puntos en rojo).

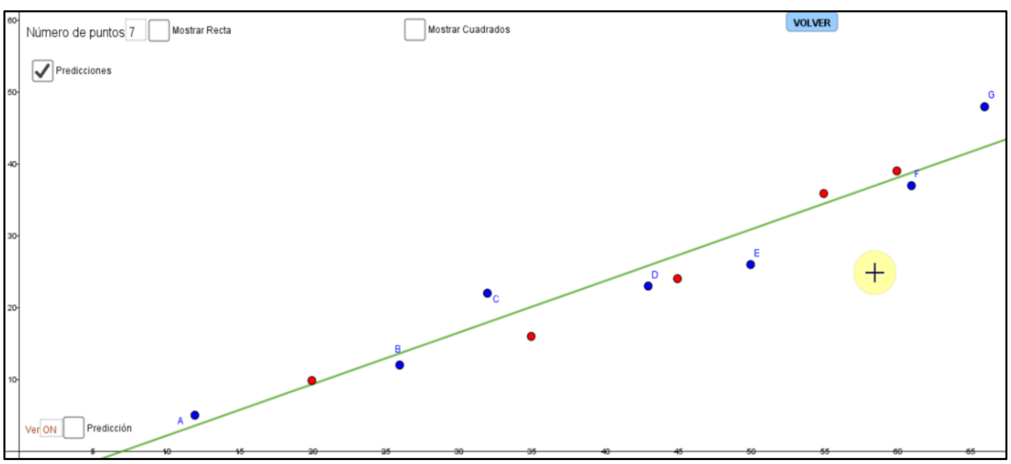
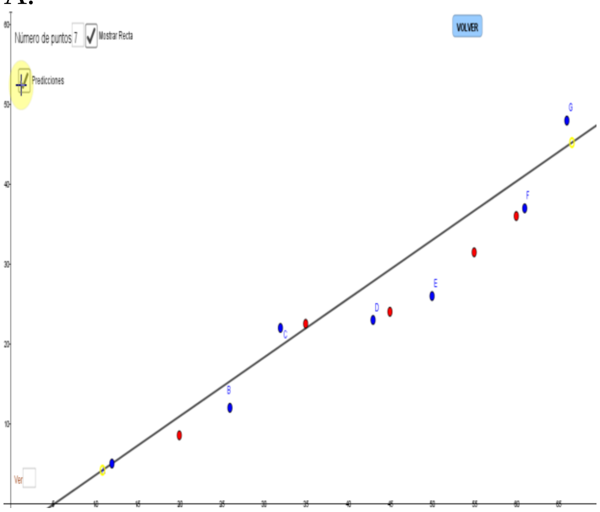
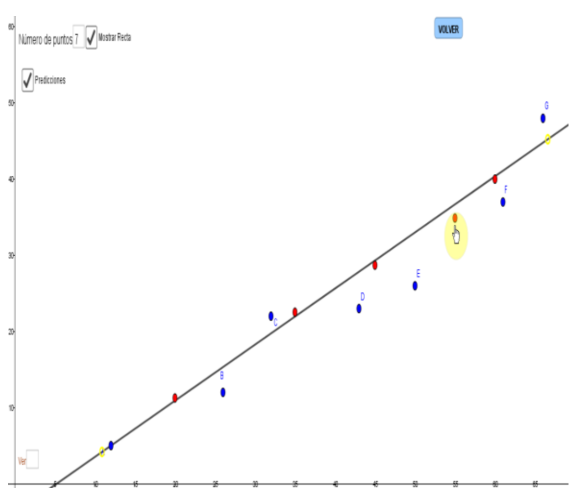


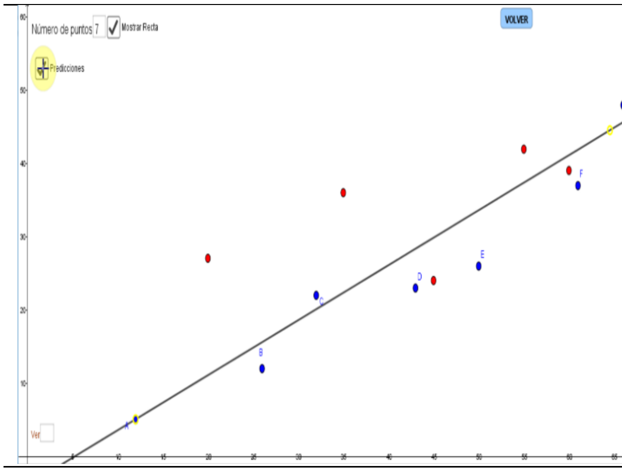
Figura 83. Nuevas predicciones de un estudiante del ítem 8 de la prueba final, después de observar los datos en un diagrama de dispersión

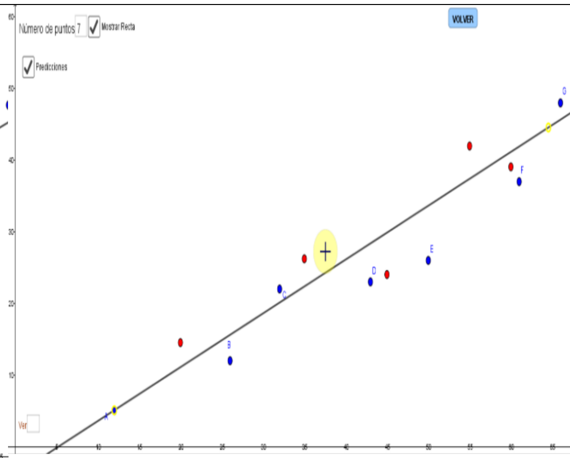
Como se observa en la Figura 83, teniendo la mejor recta, las predicciones de este estudiante están muy cerca de la recta. Cuando se le preguntó que si nunca sobre la recta, dijo que no porque *"cuando está sobre la recta entonces no se le podría sacar medición… de cuánto la distancia está… del punto a la recta"*, es decir, como si estuviera buscando que los puntos que corresponden a las predicciones, junto con los datos, conserven que la suma de las diferencias a la recta sea cero, sin caer en cuenta que si los pone justo en la recta, también se da.

El anterior ejercicio de ver las predicciones que hicieron sobre la tabla, ahora sobre un diagrama de dispersión, juzgar sus predicciones y moverlas hasta donde creían convenientes, también se hizo con los otros seis estudiantes durante la entrevista individual. En la Tabla 60 se observan, al lado derecho, las predicciones basadas en la tabla, ahora sobre un diagrama de dispersión y, al lado derecho, las predicciones cambiadas.

Tabla 60.
Al lado derecho, las predicciones de los estudiantes basadas en datos mostrados en una tabla y, a la derecha, las predicciones modificadas después de ver los datos en un diagrama de dispersión

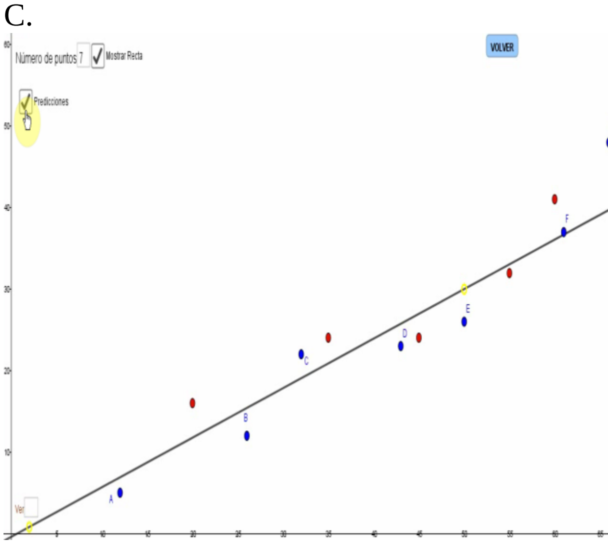
A.	A. Modificado
	
B.	B. Modificado

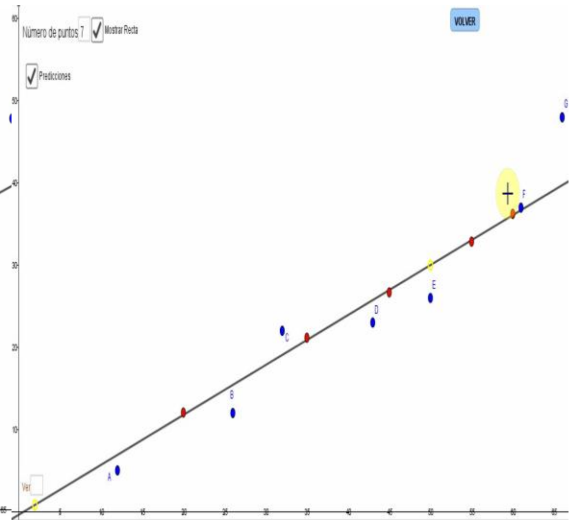




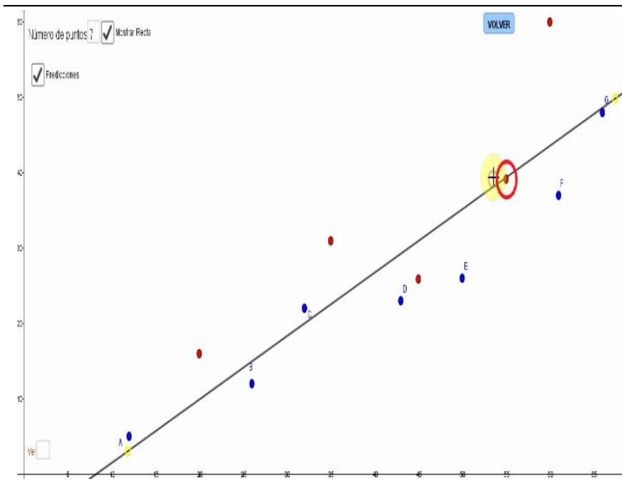
C.

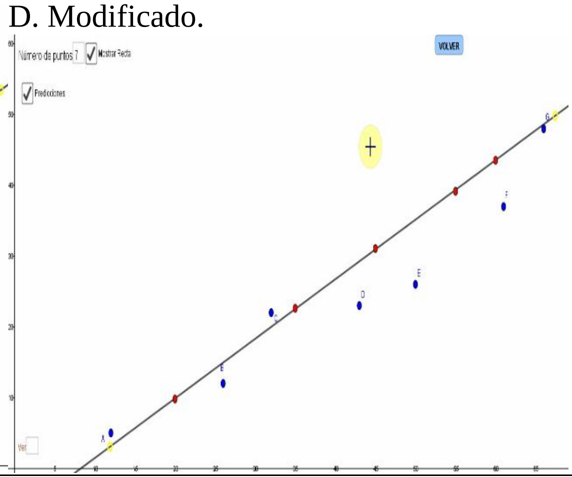
C. Modificado





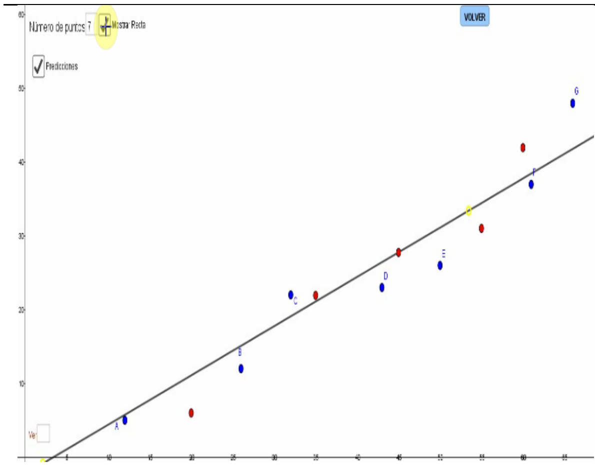
D. Modificado.

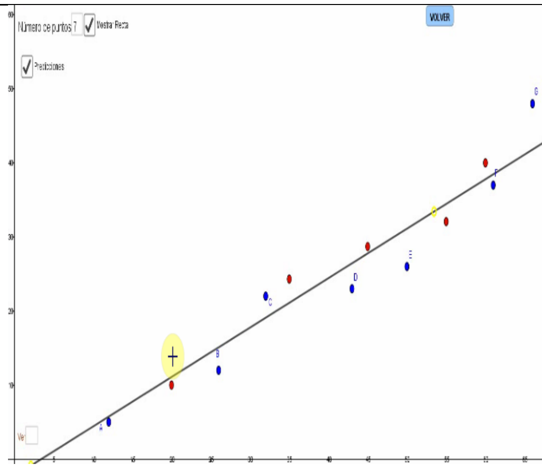




E.

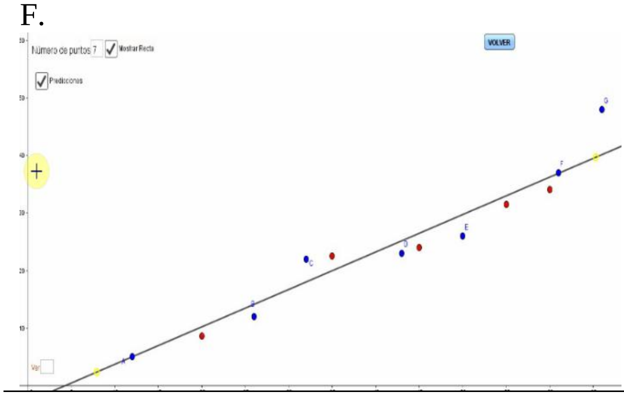
E. Modificado.



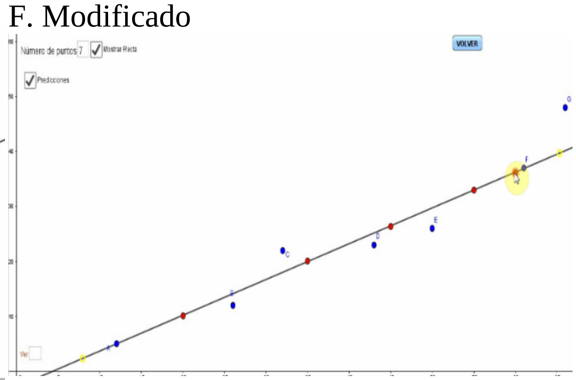


I: Qué está pensando cuando los ubica, Silvia.
S: Que los puntos de acuerdo a… acá… (C, por ejemplo) estén más arriba que los menores.
I: Ah bueno listo… ¿y en ese caso? (el que está cerca a F) Ahí, cómo iría… el más cercano está más abajo.
S: …
I: Ahí con cuál lo está comparando (para 20)
S: Entre A y B.
I: Listo ¿ahí sería? (los puntos)
S: Creo que sí

F.



F. Modificado



De manera que, observando la Tabla 60, junto con la Figura 83, en total tres estudiantes mueven sus predicciones sobre su recta de mejor ajuste y cuatro muy cerca de la recta. Uno de los que ubicaron los puntos cerca de la recta (Figura 83 A. Modificado) comentaba lo siguiente:

I: O sea que si usted tiene la recta, no en todos los casos va a poner el punto sobre la recta.

A: No

I: ¿Por qué, si esa es la recta que mejor se ajusta?

A: Porque los valores que ya están, no todos van sobre la recta…

I: Cuáles, no todos van sobre la recta ¿ah, los azules?

A: Sí

Es decir, esta estudiante asume que como los datos dados no necesariamente van todos sobre la recta, pues sus predicciones tampoco deberían ir todos sobre la recta pero sí muy cercanos a la recta.

Por otra parte, esta misma estudiante manifestó que la tarea de predecir es más sencilla cuando los datos se visualizan sobre un diagrama de dispersión que sobre una tabla:

I: Finalmente, ¿cómo le ayuda la recta de mejor ajuste a predecir un valor de Y para cualquier valor de X?

A: Porque ya estando la recta entonces uno solo se ubica en el punto y sube hasta la recta o cerca de la recta. En cambio como estaba… en el último punto (prueba final) era más difícil predecir un valor casi exacto

I: Por qué

A: Porque como no tenía la recta entonces podría ser cualquier… entre los valores… porque yo cogía los dos valores que estuvieran (más cerca) al punto que usted me daba y buscaba como el centro de esos dos.

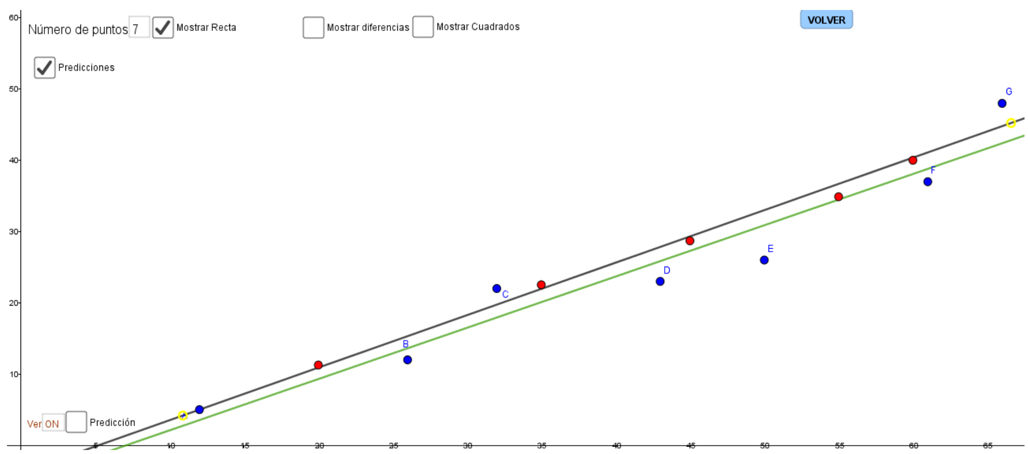
I: ¿O sea que si yo le hubiera dado el diagrama de dispersión hubiera sido más fácil?

A: Sí (risa)

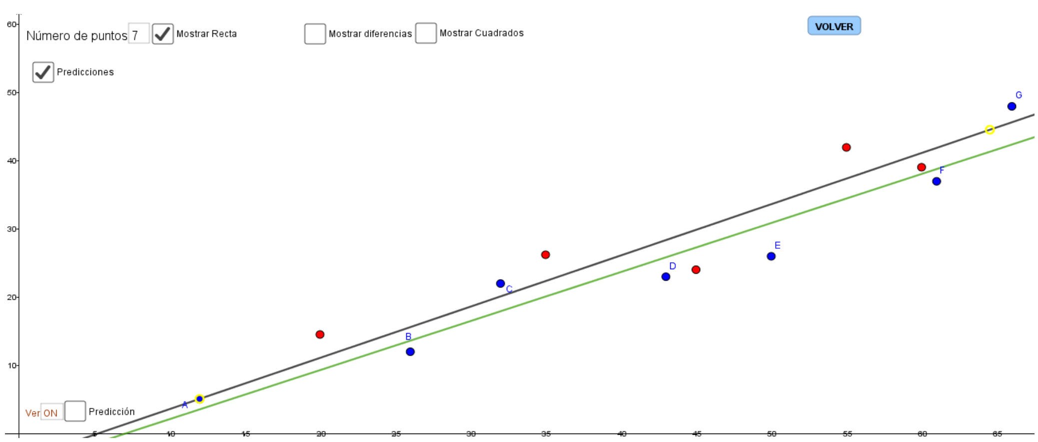
Finalmente, en la Tabla 61, se pueden ver las dos rectas, la que cada estudiante ajustó (color negro), durante la entrevista individual, a la nube de puntos correspondiente a los datos del último ítem de la prueba final y la recta de mejor ajuste por el método de mínimos cuadrados (color verde). Como se observa, todas las rectas reflejan la relación entre las dos variables y se podría decir que son bastante acertadas.

Tabla 61.
En color negro, recta de mejor ajuste que cada estudiante propone durante la entrevista y, en verde, recta de mejor ajuste por el método de mínimos cuadrados

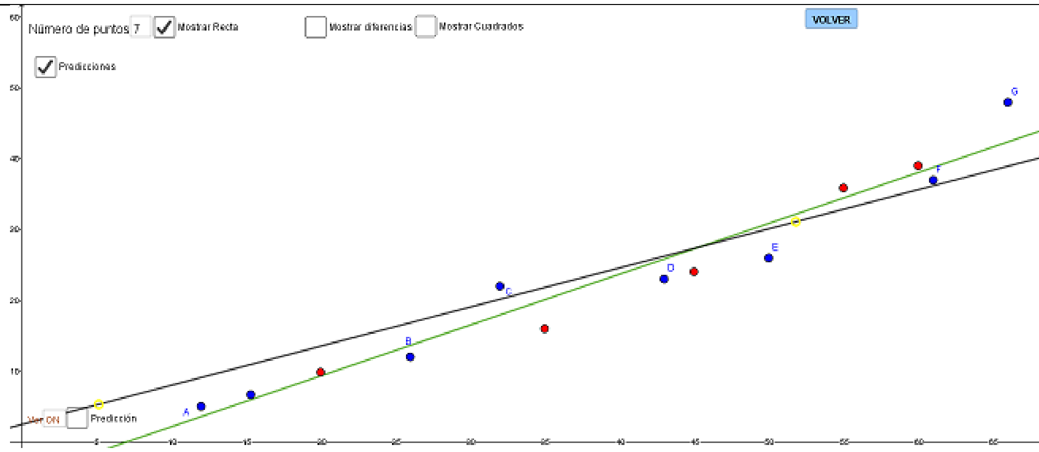
A.



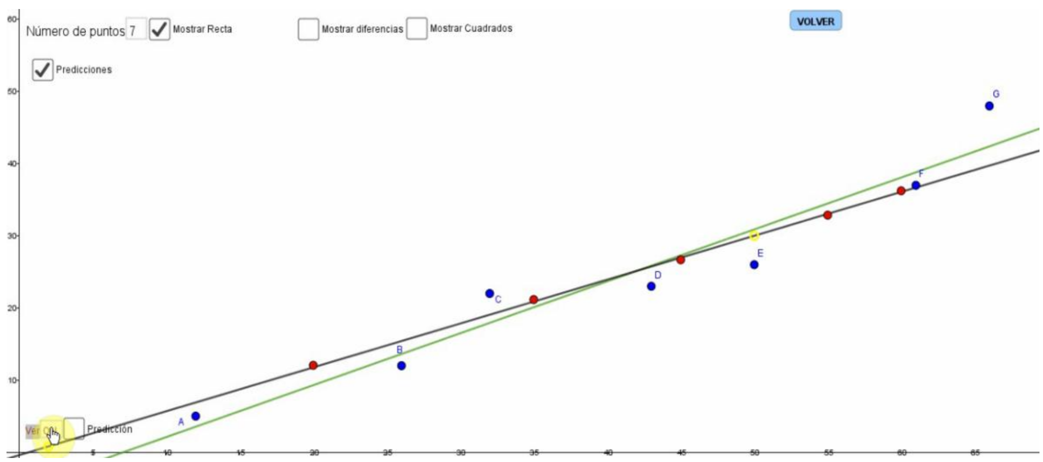
B.



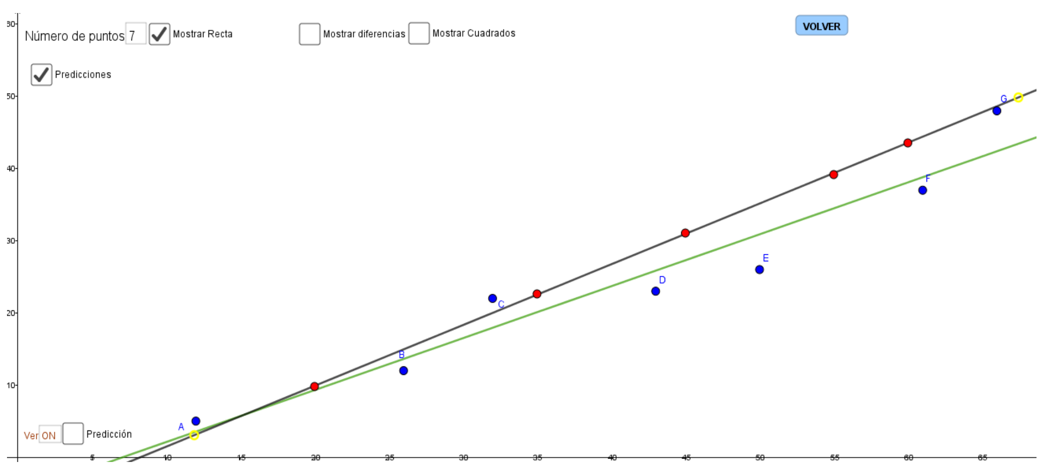
C.


Número de puntos 7
Mostrar Recta
Mostrar diferencias
Mostrar Cuadrados
VOLVER
Predicciones
Ver ON
Predicción

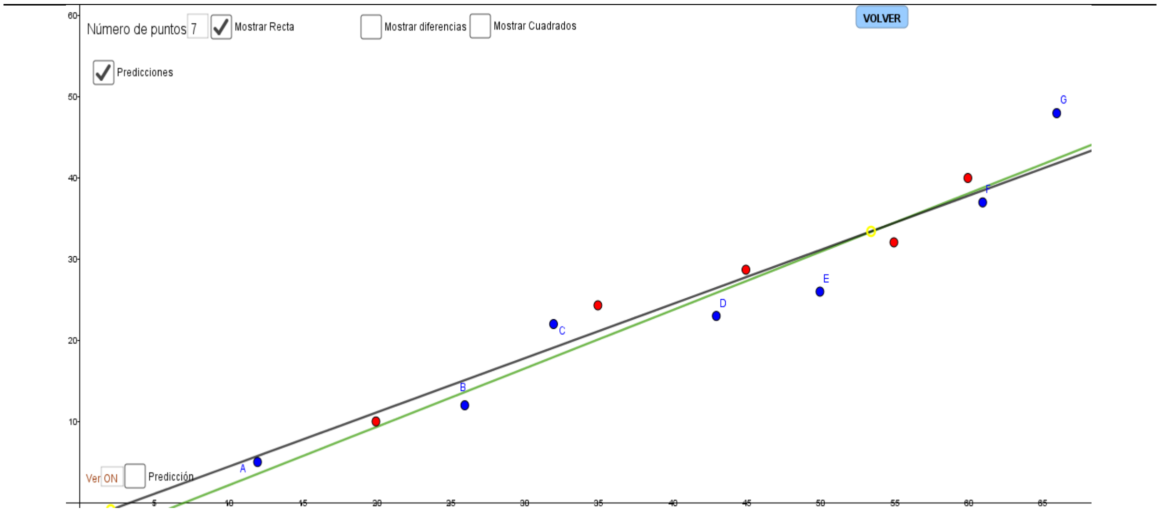
D-


Número de puntos 7
Mostrar Recta
Mostrar diferencias
Mostrar Cuadrados
VOLVER
Predicciones
Predicción

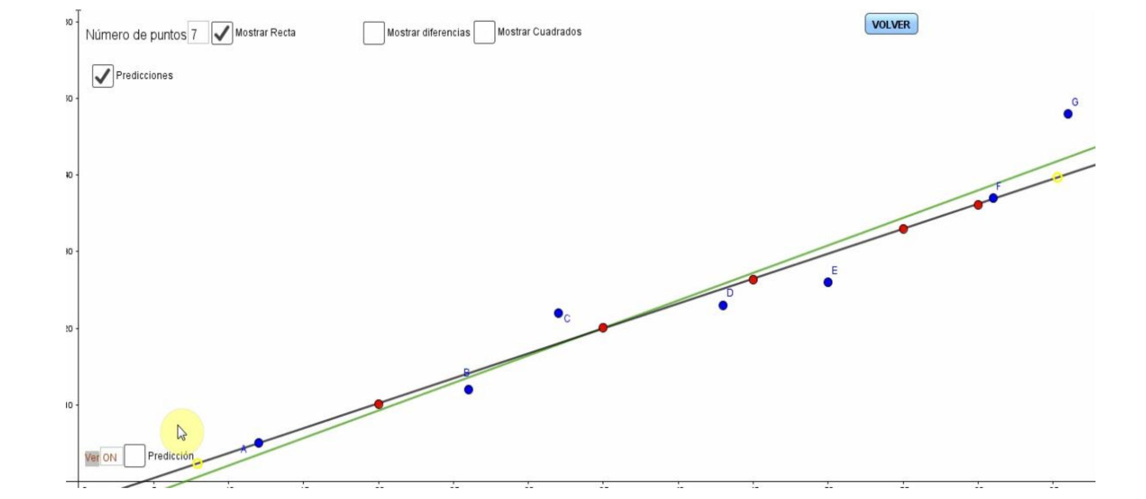
E.


Número de puntos 7
Mostrar Recta
Mostrar diferencias
Mostrar Cuadrados
VOLVER
Predicciones
Ver ON
Predicción

F.



G.



En la entrevista, durante la ubicación de la recta de mejor ajuste, se les cuestionaba sobre qué estaban pensando o teniendo en cuenta para trazar la mejor recta. A continuación, algunas de las conversaciones:

CONVERSACIÓN I CON LA ESTUDIANTE A.

I: Qué tuvo en cuenta para trazar esa recta

A: Que estuviera entre los puntos

I: Ahí está entre los puntos ¿y algo más?

A: Sí, y que esté cerca a los puntos

CONVERSACIÓN II CON EL ESTUDIANTE D.

I: O sea, cuál es su estrategia, Duván.

D: Yo tuve en cuenta el primer punto que es A y que los otros pasaran o estuvieran cerca a la recta.

I: Cerca a la recta, y cómo medimos esa cercanía.

D: Ehh… pues…

I: O sea, cómo yo mido… o sea si nuestro amigo quiere trazar la mejor recta y usted le está diciendo que esté muy cerca, entonces usted cómo podría medir esa cercanía para garantizar que está trazando la mejor recta.

D: Eh… ¿Qué trace la mejor recta?

I: Qué hacíamos nosotros luego de esto

D: Ahí después de hallar la línea pues mirábamos a ver que los puntos estuvieran cerca o en la recta

I: Y cómo mirábamos que estuvieran cerca

D: Pues… cómo así eh… cómo así no le entiendo. Ahí, tuve en cuenta que un punto pasara… que el primer punto pasara por la recta y los otros, cerca.

CONVERSACIÓN III CON LA ESTUDIANTE W.

I: Bueno, si un amigo le pidiera ayuda para trazar la recta de mejor ajuste en un diagrama de dispersión o una nube de puntos, qué indicaciones le daría usted para que él encontrara la mejor recta.

W: Pues que primero se debe que ubicar entre los puntos, que la distancia que hay de los puntos a la recta tanto a la derecha…

I: ¿A la derecha?

W: No, o sea… si la recta…

I: Por arriba

W: Sí, por arriba, la suma de las distancias de los puntos por arriba de la recta, debe ser la misma que la distancia de los puntos por abajo.

I: Ajá, eso en cuando a distancias, y algo más para ser… bueno listo, está entre los puntos, tiene que cumplir esa propiedad de las distancias y ¿algo más para ubicar la mejor… la mejor recta?

W: No… yo lo hacía así.

Sus respuestas reflejan el abandono de algunas estrategias adoptadas en el desarrollo de la actividad "Encontrando la mejor recta". Aunque no se muestran todas las conversaciones, la mayoría de los estudiantes se inclinaron porque la recta esté entre los puntos o cerca de los puntos, como en la conversación I, reflejando una visión global de los datos. En la conversación II se refleja que el tema de medir esa cercanía de la que ellos hablan no es tan sencillo de entender, de manera que bien vale la pena profundizar más en el tema de lo que los estudiantes entienden por cercanía. Una estudiante va más allá de que la recta esté entre los puntos al tener en cuenta que la suma de las diferencias, de los puntos a la recta, sea cero, como en la conversación III.

4.12 Análisis entrevista grupal final

Una última toma de datos correspondió a una entrevista final, con los siete estudiantes involucrados en este proceso. La entrevista tenía como objetivo conocer la comprensión de los estudiantes de los temas abordados durante el desarrollo de la THA como media, variabilidad y predicción. De manera tal que iremos destacando algunos fragmentos de la entrevista donde pensamos que se evidencia el desarrollo de algunas ideas alrededor de los temas mencionados.

En el siguiente fragmento vemos cómo los estudiantes asumen la media más allá de un cálculo, así como su papel a la hora de hacer una predicción. No faltan las menciones a la variabilidad propia de los datos y de la media misma.

I: Wendy, usted qué aprendió en todo este período de tiempo. ¿Cómo le contaría a usted a su abuelita, lo que aprendió?

W: Pues que aprendí que la media nos puede ayudar para muchos casos, cuando necesitamos hacer algún cálculo, para dar un valor preciso. Que un solo valor lo podemos… o sea, de varios valores podemos dar uno solo si necesitamos dar una afirmación para predecir algo, ah…

Claramente Wendy expone la utilidad de la media: para “hacer algún cálculo”, tal vez pensando en la media como medida de resumen y “para predecir algo”.

I: ¿Ese solo? Y ese solo cómo es que se calcula

W: Eeeh, pues sumando todos los valores y dividiéndolo en la cantidad de valores

I: Aaah y ese valor, María Fernanda, ¿qué nos garantiza?

MF: Nos garantiza que se puede reducir todos los valores a uno solo. O sea, para predecir…

I: Para predecir… ¿será que cuando yo mido algo siempre obtengo el mismo resultado?

Todos: No

I: No, no. Dubán, explique eso.

D: Eh, no porque esto según lo que usted vaya a medir, no todos son esto iguales… eh, como dice mi compañera para… ahí se pueden dar unas medidas pero que no son iguales pero si necesito saber esto… el promedio en qué está o la media, entonces ahí sí sumaría y dividiría y ahí obtendría un resultado.

En el próximo apartado se observa como los estudiantes comprendieron que la media es un representante de todos los datos y que, por ende, es la mejor alternativa para una predicción. Esta idea que contiene elementos de variabilidad fue una constante durante toda la experiencia realizada y que, como se ha descrito, previamente no les resultó nada fácil de controlar.

El diálogo se inicia cuando el investigador sugiere utilizar el máximo de los valores como un mejor predictor.

I: ¿Y no será mejor tomar el máximo valor?

D: Eh…

I: ¿Tomar el más grande?

D: No

I: ¿No? ¿Por qué?

D: Porque esto… eh…

I: ¿Por qué, el promedio, o media, da más tranquilidad respecto a la medición que estoy haciendo, por qué no tomo el máximo de los valores?

W: ¿Pues porque con la media podía dar un valor preciso?

I: ¿Preciso? O sea, yo doy ese valor, después vuelvo y mido ¿y me da exactamente la media?

W: No, pero podría dar un valor que más se acercara a ese

I: ¿Pero el valor máximo no me sirve, Miguel?

M: También podría servirnos pero así como dice Wendy si uno por ejemplo suma todos los valores y los divide por la cantidad ese es el número que nos está dando según todos, por ejemplo ¿cómo uno puede hacer otra medición? Puede estar cerca a ese valor, o puede que esté alejado, pero siempre puede…

Los estudiantes no poseen argumentos sólidos en defensa del uso de la media para predecir un futuro valor de la variable en estudio.

La discusión se dirigió ahora hacia las propiedades de la media con el ánimo de ayudar a los estudiantes a justificar su uso. En la ejecución de la THA, los estudiantes tuvieron la oportunidad, a través de actividades como la del Machín Machón y la del reparto de los caramelos, de descubrir propiedades y características que cumple la media.

I: Pero la media… ¿una media puede ser más grande que el máximo de los valores?

Todos: No

I: Yo tengo una serie de valores…

A: No, porque la media está entre los valores.

I: ¿Está entre los valores? ¿O sea, no es más grande que el máximo?

W: No

I: ¿Y por qué no?

W: Porque debe estar entre los valores.

I: Debe estar, pero… es decir, “debe estar” suena feo, nada en la vida debe, esas cosas se deducen por alguna razón. Yo creo que la media puede ser más pequeña que el mínimo o más grande que el máximo. Quién me da una razón para convencerme de que no puede ser así.

(Silencio)

I: Jaja, quién me da una razón. Andrea, deme una razón, convénzame. Piénsenlo en un ejemplo… piensen en la situación de los caramelos… vamos a hacer una colecta de caramelos para regalárselos a los niños de primero o de segundo que por alguna razón tienen un festeje y nos pidieron una colaboración. Sabemos que son siete niños entonces vamos a juntar caramelos. ¿Cierto? Y yo quiero darle a cada niño la misma cantidad de caramelos, ¿cómo calculamos esa cantidad igual?

W: Pues la media nos puede ayudar, nos da las herramientas necesarias para poder dar…

I: Aquí, por ejemplo, somos ocho personas… ¿usted cuántos caramelos tiene?

MF: Mmmm treinta

I: Cuántos, usted Miguel…

M: siete

E: Quince

D: Diez

S: Veinte

W: Veintisiete

A: Seis

I: Y yo siempre pongo diecisiete. Cuánto da.

…

I: 132. Sumados da 132 divididos por ocho. ¿Cuánto es?

D: Dieciséis

I: ¿16 exactos?

D: No, 16.5

I: Entonces 16.5, cómo vamos hacer. En principio, cuántos caramelos damos. Miguel, usted cómo hace. Da 16.5 pero nosotros no tenemos sino caramelos enteros, usted se imagina el lío… una pepa dura de esas partirla por la mitad, ahí nos quedamos…

A: Entonces buscar un valor cercano.

I: Como cuál

A: Dieciséis

I: Usted pone dieciséis, y por qué no diecisiete.

MF: ¿Porque sumaría más?

I: Claro, no tendríamos caramelos… no tendríamos ¿no? Porque cuánto a 17 por 8… bueno estamos suponiendo que vamos a entregarle a ocho personas. Entonces vamos a entregarle a ocho personas, entonces cuánto da 17 por ocho…

W: 136

I: Y cuántos caramelos teníamos

Todos: 132

I: Entonces sencillamente 17 no, porque no nos alcanza. ¿Y dieciséis?

W: 128

I: Entonces qué pasó

A: Sobraron

I: Entonces claro, un problema práctico ahí… por encima no porque es más de los que tenemos, por debajo sí pero nos van a sobrar. Pero bueno, volvemos al cuento, quién entregó la mayor cantidad de caramelos ¿parece que fue María Fernanda con 30? 30, ¿la media cuánto nos dio?

Todos: 16.5

I: Menor que el máximo

Todos: Sí

I: Y cuál fue el mínimo de caramelos… siete quién dio

Todos: No, Andrea

I: Andrea dio seis

Todos: Sí

I: La media es mayor que el menor valor. Es un ejemplo concreto. En general, ¿la media es superior al mínimo?

Todos: Sí

Inclusive se les cuestionó sobre algún caso donde la media correspondería al mínimo o al máximo de los datos. Y al final se evidencia la comprensión de cómo la media se vería afectada si todos los datos son iguales a excepción de uno.

I: ¿Siempre? ¿Hay alguna situación donde la media sea exactamente igual al mínimo?

A: No

I: ¿No? ¿Nunca? ¿No? ¿Seguro? Una situación como los caramelos, a ver… una situación donde la media sea igual al mínimo… si todos hubieran dado diez caramelos, todos diez… ¿cuál es el mínimo?

Todos: Diez

I: Y cuál es la media

Todos: Diez

I: Son iguales. Entonces sí hay un caso. Cuál es el caso donde el mínimo es igual a la media.

A: Cuando todos los valores son iguales

I: Son iguales. Y si hay uno diferente… qué pasa con la media

A: Sube otro poquito

I: ¿Y con el máximo… la media puede ser mayor que el máximo?

A: Sí

I: Cuándo, Andrea

A: Cuando todos los valores son iguales

I: Cuando otra vez los valores son iguales. Cuál es el máximo es todos esos diez, diez, diez… cuál es el máximo

A: Pues diez, que es el único caso es que mínimo y máximo son…

Todos: Iguales

I: Iguales. Entonces en ese caso, la media también es igual al máximo. Pero si hay uno diferente, más chico ese chico qué hace, que la media sea…

A y W: Menor que el máximo

I: Más chica, como la media los considera a todos, entonces el más pequeñito que el máximo pues hace que la media también sea más pequeñita que el máximo.

En la siguiente conversación se muestra el razonamiento de los estudiantes cuando se les presenta una situación en la que se les da tres datos y la media de cuatro datos, y se les pide encontrar el dato faltante. Veremos cómo algunos proceden de manera correcta al comprender que la media debe relacionarse con los cuatro datos, llevándolos a encontrar el dato que falta, mientras que una de ellos promedia los tres datos que le dan. El argumento de esta última estudiante se ve confrontado con el de otra hasta llegar a un acuerdo.

I: Digamos que tenemos cuatro valores, cuatro… yo tengo cuatro datos pero voy a esconder uno. Como yo tengo los cuatro, voy a calcularle la media a los cuatro y voy a esconder uno. Les voy a dar tres datos y el valor de la media. Ustedes me adivinan el cuarto, el que falta. Digamos que los números son tres, cinco y ocho: la media es seis, ¿cuál es el cuarto valor?

…

I: ¿Listo? Elder ¿listo?

E: Más o menos… yo sumé y esto… a 16 le agregué 6 para que me diera 24…

I: No, 16 y 6 da 22.

E: Bueno, le agregué 8 que diga… le agregué 8 para que me diera 24

I: Y por qué 24

E: Y 24 lo dividí entre…

I: Pero por qué 24… de dónde salió 24. Ese número 24 yo no lo dicté… Elder, cuénteme el 24 de dónde salió

E: Porque usted dijo que la media era seis ¿cierto?

I: Sí

E: Entonces lo único que hice fue multiplicar seis por cuatro y para que me diera 24 le sumaba ocho, porque había 16…

I: Entonces

E: Entonces al sumarle ocho me daba 24.

I: Entonces el número que falta cuál es…

E: El ocho.

I: ¿Ese es el mismo razonamiento suyo, Wendy?

E: Sí señor.

I: ¿A todos les dio ocho? ¿Silvia cuánto le dio a usted, realmente?

S: Eh, yo primero sumé los tres valores y los dividí en tres

I: ¿En tres?

S: Y me dio cinco…

I: Cinco algo

S: Sí, cinco coma…

I: Tres…

S: Después esto lo sumé con los anteriores valores y me dio 21 y lo dividí en cuatro y me dio cinco coma veinticinco

I: ¿Qué opinan de lo que ella hizo? Andrea...

A: Que no porque, si ya tengo la media porque tengo que sacar otra vez la media de los tres valores

I: ¿Si ve Silvia, no? Yo ya le di la media de los cuatro, o es que hay alguna relación entre la media entre tres y la media entre cuatro… es como si ella estuviera diciendo, pareciera que si yo hago, cojo tres y hago la media, y después cojo los cuatro y hago la media… ¿hay alguna relación entre esas dos medias?

A y M: No

Pasando al caso de dos dimensiones, en primera instancia pudimos ver cómo los estudiantes relacionan una asociación negativa y una asociación positiva en el sentido de que al aumentar una variable disminuye la otra y, al aumentar una variable, aumenta la otra, respectivamente. De igual manera la forma en que deberían ir los puntos en cada caso, veamos:

I: Bueno muy bien, entonces vámonos para dos dimensiones… cómo hago yo predicciones cuando tengo dos variables. Por ejemplo, ¿un ejemplo con dos variables cuál sería? Cuando ya los valores de una dependen de la otra, ya es un caso más interesante, un poquito más complicado pero más interesante ¿no? Entonces qué iba a decir, Andrea… cómo hace uno ahí una predicción.

A: Miro los anteriores y miro qué relación tienen las dos variables. Por ejemplo en un caso nos daban digamos 35 minutos y 36 minutos de natación, cuántas pulsaciones en 35 minutos se hacen, entonces por ejemplo en 35 minutos digamos 128 pulsaciones y en 36… 138…

I: Cuando aumentan los minutos… qué pasaba con las pulsaciones

A: Disminuyen

I: Entendiendo…Ustedes están viendo el eje ¿no? Subir es hacer así (dibujando una recta con pendiente positiva) Qué quiere decir subir, que si aumento en X…

M: Y también

I: Y también. Entonces en ese ejemplo que mencionó Andrea, ¿vamos subiendo o vamos bajando? Lo acabó de decir, Miguel… ¿bajamos o subimos?

A: Bajamos

I: Bajamos. Porque a medida que avanza, o aumenta el tiempo en que se demora recorriendo la misma distancia, las pulsaciones

De igual manera, como se muestra a continuación, los estudiantes asumen la intensidad de relación entre dos variables cuando los puntos se asemejan a una recta, o cuando la nube de puntos es más "flaca":

I: Y esa calidad de esa asociación cuando yo tengo una serie de puntos en el gráfico, cómo digo "huy esa asociación es buena o esa asociación es mala", qué me indica cuando yo veo esa nube de puntos, cuándo estoy más seguro de que la asociación es buena.

A: Cuando tienen forma de recta los puntos

I: Cuando tiene forma de recta, o sea que hablando en términos de gorda o flaca, la nube… qué es mejor, ¿flaca o gorda? La nube…

M y A: Flaca

En cuanto a predicción, vemos cómo la estrategia de observar los vecinos prevalece, ya sea ubicándolo en la mitad, o solo entre los valores que ya nos han dado. Vemos también que esas predicciones no las asumen con precisión debido a que son conscientes de la existencia de la variabilidad siempre.

I: Bueno, muy bien y ahí cómo hacemos entonces predicción. Nos dieron 30 minutos, 31, 32, de 32 brincó a 35, 38… y me dice, y si se hubiera demorado 33 minutos, que ese dato no me lo dieron, ¿cuánto cree usted que podrían ser las pulsaciones?

W: Pues…

I: Cómo lo hace, fíjese que tiene una serie de puntos, tiene la nube

W: Pues tendría en cuenta que… 33 está entre 32 y 34 y entre esos valores estaría…

I: O sea, tengo 33 que es el que quiero predecir, tengo 32 y tengo 34. Entonces en 32 tengo un valor y en 34 tengo otro valor, más abajo porque dijimos que está decreciendo y entonces dónde pongo 33

A: En la mitad… en la mitad entre 32 y 34

I: En la mitad de qué

A: De los dos valores

I: 33 está en la mitad, pero yo quiero… esto es tiempo, yo quiero en términos de pulsaciones, y tengo las pulsaciones en 32, supongamos que hayan sido 120, y tengo las pulsaciones en 34, supongamos que hayan sido 100, entonces en 33 en dónde la pone

A: ¿En 110?

I: ¿Lo pongo exactamente en la mitad?

A: No porque hay variabilidad

Como se pudo observar, en el caso univariado los estudiantes asumen la media como un buen candidato para hacer una predicción, mientras que en el caso bivariado prevalece la idea de observar lo que pasa alrededor y hacer la predicción ubicando el valor entre los valores dados, siempre y cuando se conserve la relación que reflejan, de manera global, los datos. Vimos también que hay un grado de consciencia frente al tema de la variabilidad tanto en el caso univariado como en el caso bivariado, y que ese aspecto jamás les permitirá ser precisos pero sí obtener una buena predicción.

En cuanto a la media, los estudiantes la identifican como un valor que cumple ciertas características y propiedades como que debe estar ubicada entre los valores. En el caso bivariado, se muestra que los estudiantes asumen el tipo de asociación entre dos variables de acuerdo con la forma que muestran los puntos y que el nivel de asociación podría juzgarse de acuerdo al grosor de la nube de puntos o a que estos se asemejen más a una línea recta.

5. Conclusiones

La investigación desarrollada muestra las formas de razonamiento informal de estudiantes de octavo grado alrededor de la recta de mejor ajuste, como resultado del diseño, ejecución y análisis de una THA soportada en actividades computacionales utilizando Geogebra y Fathom. Los estudiantes no habían estudiado la recta en el colegio desde el punto de vista algebraico.

Los ejes temáticos principales abordados en esta THA fueron variabilidad y predicción partiendo desde un análisis unidimensional, enfocado en la media, hasta uno bidimensional, en la recta de mejor ajuste.

En esta investigación se ratifican algunos de los resultados descritos en los antecedentes, pero también se evidencia cierto control sobre concepciones y estrategias erróneas ya reportadas en cuanto a la ubicación de la recta de mejor ajuste, lo anterior debido a ideas desarrolladas en la parte univariada que permitieron una extensión al caso bivariado.

En concordancia con lo que señalan Simon y Tzur (2004, p.93), para el diseño de la THA que sugerimos, se siguieron los supuestos que ellos consideran en la forma que a continuación se describe:

1. La generación de una THA se basa en la comprensión de los conocimientos actuales de los estudiantes involucrados. Lo anterior nos los proporcionó, en primer momento,

la revisión bibliográfica en cuanto a concepciones y estrategias alrededor de la recta de mejor ajuste y, en segundo momento, el análisis de dos pruebas diagnósticas.

2. Una THA es un vehículo para el aprendizaje de la planificación de determinados conceptos matemáticos. Aunque nuestro objetivo final de aprendizaje era la ubicación de la recta de mejor ajuste, durante la planeación surgió la necesidad de abordar antes el concepto de media.

3. Las tareas de matemáticas proporcionan herramientas para promover el aprendizaje de determinados conceptos matemáticos y son, por lo tanto, una parte clave del proceso de instrucción. Herramientas computacionales como Geogebra y Fathom nos permitieron diseñar actividades como las del Machín Machón y Encontrando La Mejor Recta, entre otras, las cuales facilitaron la exploración de propiedades que presentan conceptos como la media y la recta de mejor ajuste.

4. Debido a la naturaleza hipotética e inherentemente incierta de este proceso, el profesor participa regularmente en la modificación de todos los aspectos de la THA. De acuerdo a las ideas discutidas y desarrolladas en actividades previas, constantemente se hacía necesario un rediseño o ajuste de las actividades que venían y en el discurso de las mismas. Pero siempre enfocados hacia el aprendizaje del concepto abordado.

De manera que en este apartado mostramos las conclusiones de esta investigación que incluye, como sugerencia, nuestra THA del tema que abordamos. Por lo que este apartado se divide en dos partes: i) Una THA del concepto de Recta informal de mejor ajuste y ii) Formas de razonamiento informal alrededor de la recta de mejor ajuste en estudiantes de octavo grado al finalizar la implementación de la THA.

5.1. Una Trayectoria Hipotética de Aprendizaje del concepto de Recta de Mejor Ajuste Informal.

Convenimos en que considerar la recta como un modelo para la asociación lineal entre dos variables, debería ser el resultado de un razonamiento previo en tareas que involucren:

i. Variabilidad en un conjunto de datos.

ii. Predicción de un nuevo valor en un conjunto de datos.

iii. Análisis univariado alrededor de la media.

iv. El reconocimiento de la existencia o no de una asociación entre dos variables cuantitativas.

v. La predicción local del valor de la variable respuesta.

vi. La predicción global que conlleve a la forma que "resume" la asociación entre las dos variables, en este caso, la recta.

vii. La exploración de la recta que mejor se ajusta a la nube de puntos.

Para el desarrollo de la lista anterior, la THA que se propone consta de dos fases principales como se muestra en la Figura 84:

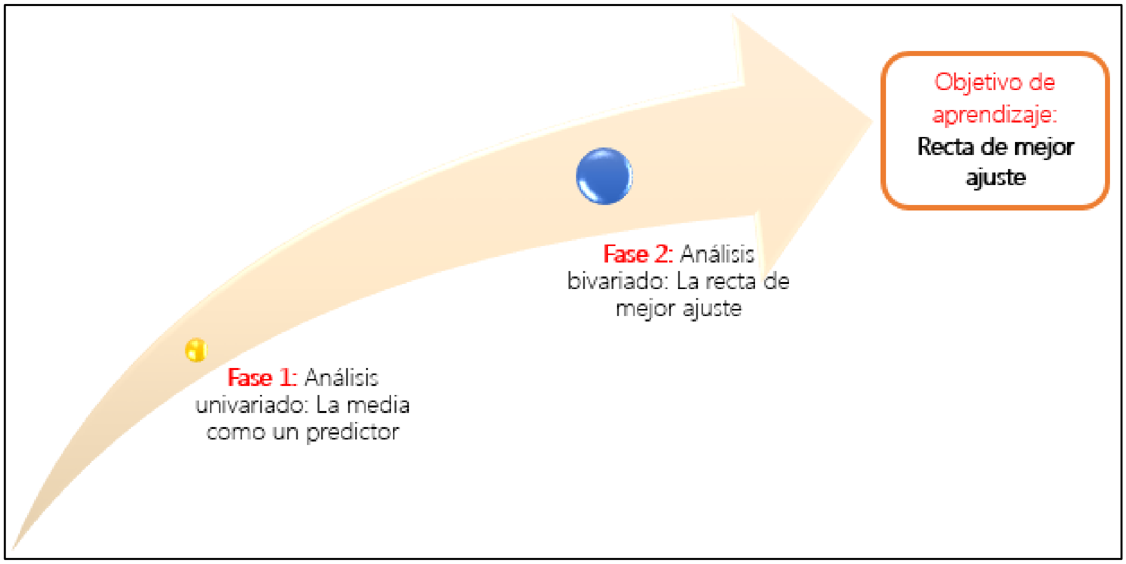


Figura 84. Fases de la THA propuesta del concepto de Recta Informal de Mejor Ajuste

A continuación, la justificación y descripción de las fases. Al finalizar cada fase, en una tabla resumen, se definen unos objetivos específicos, análisis claves, conceptos claves, hipótesis de razonamiento, actividades y herramientas de apoyo (applet, software,…) para el desarrollo de las mismas.

5.1.1 Fase 1: Análisis de conjuntos de datos univariados alrededor de la media. ¿Por qué partir del análisis de un conjunto de datos univariados con un enfoque hacia la media?

En un problema bivariado, en donde se refleja gráficamente una tendencia lineal, una tarea interesante sería resumir ese conjunto de datos en una recta, la mejor ¿pero cómo encontrar esa recta y qué características o propiedades tiene que la hacen única?

La idea de encontrar un modelo en el plano que resuma un conjunto de datos, representados gráficamente, guarda una similitud con el análisis que se podría hacer en el caso univariado.

De manera que responder a la pregunta inicialmente planteada es equivalente a responder ¿Qué se podría extraer de un análisis univariado que facilite la extensión a un análisis bivariado? A continuación el desarrollo de los ítems de la lista anterior asociados a la Fase 1.

i. *Variabilidad en un conjunto de datos*. El hecho de observar datos dispersos en un gráfico bivariado se debe a la variabilidad, es por eso que en un gráfico de dispersión a un valor de la variable explicativa le pueden corresponder diferentes valores en la variable respuesta; es decir, el determinismo matemático no existe. Por lo anterior se tiene una primera respuesta: *la variabilidad,* ¿por qué una medición en x puede tener muchas mediciones en y? Lo anterior puede abordarse desde el caso univariado, por ejemplo, cuando se mide una variable en una población que presentan características comunes ¿por qué el valor de la variable no es la misma en todos los individuos si se miden bajo las mismas condiciones, inclusive en el mismo individuo? Un primer acercamiento, que consideramos genera una mayor aceptación sobre la variabilidad, puede aprovecharse con la Actividad 0 “Obteniendo datos del curso al que pertenezco”, puesto que los estudiantes no solo abordarán datos reales, sino tomados por ellos mismos.

ii. *Predicción de un nuevo valor en un conjunto de datos*. En el caso bivariado se sabe que la mejor recta nos permite encontrar las mejores predicciones y, en el univariado, un buen predictor es la media. Por lo que la segunda respuesta es *resumir.* Un mayor aprovechamiento de la Actividad 0 puede darse en este momento al pedirle a los estudiantes hacer sus primeras predicciones. Lo anterior permitirá conocer sus intuiciones, concepciones y estrategias para el desarrollo de esta tarea, inclusive sus conocimientos previos sobre medidas de tendencia central, es decir, si asumen la media, mediana o moda como una predicción.

iii. *Análisis univariado alrededor de la media*. Respecto a *resumir*, muchas rectas podrían resumir un conjunto de datos bivariados, pero solo existe una con características o propiedades que la hacen la mejor, como las que se mencionan a continuación:

- La suma de las desviaciones de los puntos a la recta es cero.
- La suma de los cuadrados de las distancias verticales de los puntos observados a la recta es la más pequeña. Es decir, la recta es la más cercana a todos (en términos de la distancia vertical al cuadrado) y cada uno de los puntos.
- Permite encontrar la mejor predicción.

De lo anterior, si se hiciera una analogía con el análisis univariado, la recta de mejor ajuste también podría llamarse de alguna manera como *"la recta media"*. Veamos por qué.

Imaginemos un conjunto de datos univariados, es decir, varias medidas de una sola variable, por ejemplo la estatura de los estudiantes, hombres, de un curso de octavo grado. Imaginemos también que queremos resumir ese conjunto de datos en un solo valor, a uno muy bueno que permita predecir la estatura del estudiante nuevo que está próximo a ingresar al curso, sin ser deterministas claro, es decir, considerando también la variabilidad alrededor de ese valor. Un buen predictor, en el análisis univariado, es la media[2], que cumple con algunas propiedades, por ejemplo:

---

[2] En algunos casos un mejor valor corresponde a la Mediana debido a que es más robusta frente a valores atípicos. Para intereses propios, como lo que queremos considerar es la Media, tendremos que "negociar" con los estudiantes qué hacer con esos valores atípicos.

- Es un valor que se encuentra entre los valores extremos del conjunto de datos.
- La suma de las desviaciones a la media es cero.
- La suma de los cuadrados de las distancias de los valores observados a la media es más pequeña que para cualquier otro valor.

Actividades como la 1.1 y 1.2, “El Machín Machón”, permiten abordar estas propiedades y algunas características de la media. Por ejemplo, la Actividad 1.1, en la Situación A: Repartiendo Caramelos, permite acercarse a la media desde un punto de vista justo o equitativo; mientras que la Situación B: Balanceando el Machín Machón, como un punto de equilibrio o balanceador. Esta última situación, permite además explorar la propiedad de media respecto a la suma de las desviaciones. Por otra parte, la Actividad 1.2 permite la exploración de la propiedad de la media respecto a la suma de los cuadrados de las distancias. Para el desarrollo de estas actividades se necesita de programas como Geogebra y Fathom, respectivamente.

En la Tabla 62 se define el objetivo de la Fase 1 así como su posible desarrollo en cuanto a la secuencia de instrucción.

Tabla 62.
Fase 1 de la THA propuesta para la Recta Informal de Mejor Ajuste

Fase	Objetivo(s) específico(s)	Análisis de datos (claves)	Interpretaciones y conceptos claves para desarrollar	Hipótesis de razonamientos de los estudiantes	Actividad(es)	Herramienta (s) de apoyo
1 Análisis de conjuntos de datos univariados alrededor de la media	•Caracterizar a la media como el mejor representante en un conjunto de datos •Explorar las propiedades y características de la media	•Proceso de generación de datos •Examinar la variabilidad •Predicción de un nuevo valor •Significado de la media como un valor justo y como un punto de equilibrio. •Propiedades y Características de la media • Resumir el conjunto de datos a uno solo (la media) •Predicción de un nuevo valor (se espera uno cercano a la media o inclusive la misma media)	• La forma en que se distribuyen los datos serán similares si el proceso de generación de datos se repitiera, no exactos debido a la variabilidad •Un buen predictor en un conjunto de datos univariados es la media • La suma de las desviaciones sobre la media es cero •La suma de los cuadrados de las distancias de los valores observados a la media es la más pequeña	•Un dato representativo podría ser: El que más se repite, el mayor, el menor, el que está en medio del mayor y el menor, la media, la mediana •A menos que se haga la medida, no es posible hacer una predicción	Actividad 0. Actividad 1.1 Actividad 1.2	Geogebra Fathom

5.1.2 Fase 2: Análisis de conjuntos de datos bivariados alrededor de la recta informal de mejor ajuste. En un análisis unidimensional, la media resulta ser la más cercana a todos los valores. y es precisamente esta idea clave la que se pretende extender a un análisis bivariado en la THA que acá se propone, pues de alguna manera este hecho está asociado al criterio de mínimos cuadrados (ver Figura 85) que permite encontrar la recta que mejor se ajusta, es decir, la más cercana simultáneamente a todos los puntos.

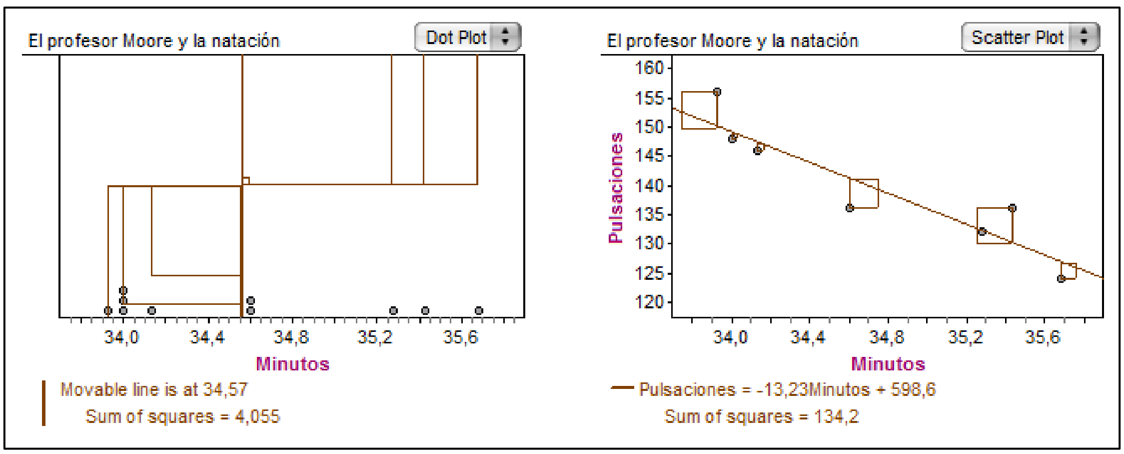


Figura 85. Representación criterio de mínimos cuadrados para el caso univariado y bivariado

Las investigaciones muestran que esa recta no es tan intuitiva para los estudiantes (por ejemplo, Casey, 2015) pero sí podría serlo el valor de la media como un dato representante o un punto de equilibrio o central, en el caso univariado. Por ejemplo, en los resultados de las pruebas diagnósticas que hicimos, en la tarea de predecir el valor de la variable respuesta fue natural para los estudiantes tomar la media de los "vecinos cercanos", es decir, un problema bidimensional quedó reducido a uno unidimensional. De manera que la extensión del papel de la media y sus propiedades, en el caso univariado, al caso bivariado es viable en la THA que acá se sugiere.

Continuando con la THA, es momento de pasar al caso bivariado. Ahora los estudiantes se verán enfrentados a dos variables en donde, en la mayoría de los casos, una se relaciona con la otra. Es importante que las ideas desarrolladas en la Fase 1 como el proceso de generación de datos, datos extraños, la variabilidad y la distribución de los mismos, se continúen discutiendo también en esta fase. Una discusión alrededor de lo anterior se permite en la Actividad 2, "Actividades del profesor Gabo", y en primera instancia con la situación "De la casa a la universidad".

A continuación el desarrollo de los ítems de la lista asociados a la Fase 2.

iv. *El reconocimiento de la existencia o no de una asociación entre dos variables cuantitativas*. Al igual que en la Fase 1, se hace necesario conocer las intuiciones, concepciones o estrategias de los estudiantes en tareas de predicción, ahora con dos variables. De acuerdo a lo desarrollado en la Fase 1, es natural que, en el caso bivariado, los estudiantes adopten, como solución inmediata, la media como una forma de predecir. Una exploración alrededor de lo que mencionamos es posible en la siguiente situación de la Actividad 2, "Practicando natación". ¿Continuarán asumiendo la media para cualquier número de predicciones? Si no se es consciente de que existe una relación entre las dos variables, la respuesta es sí y es por eso que se propone la Actividad 3 que tiene como objetivo explorar la relación entre dos variables cuantitativas, puesto que si siguiéramos en el mismo proceso de sacar la media, obtendríamos el mismo valor siempre, lo que no resulta muy coherente si existe tal relación.

Por otra parte, hasta el momento los datos han sido visualizados en tablas, pero la Actividad 3 nos permite ir un poco más allá para acercarnos a otra forma de representación como lo es el diagrama de dispersión, ya que se les incita a los estudiantes a graficar las dos variables en un solo gráfico con ayuda de Fathom. Dado lo anterior se podría discutir sobre cómo se refleja en este gráfico la relación entre las dos variables y qué tipo de relación tienen.

Gráficos de dispersión (Datos apilados verticalmente)

v. *La predicción local del valor de la variable respuesta*. Si la THA que proponemos es viable hasta el momento, los estudiantes toman la media como una forma de predicción también en el caso bivariado. Para ratificar lo anterior se propone la Actividad 4, "Nivel de ruido", en donde los estudiantes tendrán acceso a los datos bajo dos representaciones, en tablas y en diagramas de dispersión. Así mismo se continuará con la tarea de predicción, primero con los datos representados en tablas y en seguida sobre diferentes diagramas de dispersión para validar sus predicciones, es decir, si bajo esta representación sus predicciones son coherentes respecto a los datos dados.

Los diagramas de dispersión en la Actividad 4 parten con dos puntos y van aumentando paulatinamente pero siempre conservando una característica, que son datos apilados verticalmente. Lo anterior para tener un mayor aprovechamiento de la media como un predictor, es decir, se podría prever que al momento de reducir un diagrama de puntos apilados de esa manera, los estudiantes le den un tratamiento a cada pila de datos como lo hacían en el caso univariado, reduciendo cada pila a un solo valor, la media. Si lo anterior es viable, nos encontramos ahora frente a un nuevo gráfico de dispersión de puntos no apilados que, a su vez, permite generar las primeras discusiones sobre la forma que siguen esos puntos y cómo predecir de manera general, tal como hacia dónde va encaminada la Actividad 5.

Gráficos de dispersión (Datos no apilados)

vi. *La predicción global que conlleve a la forma que "resume" la asociación entre las dos variables, en este caso, la recta*. En adelante, las actividades se basan sobre gráficos de dispersión de datos no apilados con características diferentes: asociación positiva, negativa, ninguna asociación, inclusive sin ningún tipo de contexto con el fin de evitar concepciones de tipo Causal (Estepa y Batanero, 1995). Por ejemplo, en la Actividad 5 queremos que los estudiantes hablen en términos de "orden", "alineados", "recta" como formas que siguen los puntos en un diagrama de dispersión donde se refleja una asociación y además, cómo esa forma les permite predecir un valor de la variable respuesta para cualquier valor de la variable explicativa. De manera que al igual que en la Actividad 4, en la Actividad 5 también se presentan diferentes diagramas de dispersión que van aumentando el número de datos, esta vez no apilados, hasta asemejarse a una recta, que es lo que buscamos. En últimas lo que pretendemos acá es extender lo que hacíamos en el caso univariado: Resumir un conjunto de datos a "una sola cosa" que nos permita predecir, que en ese caso esa "cosa" correspondía a la media.

vii. *La exploración de la recta que mejor se ajusta a la nube de puntos*. Bajo la hipótesis de que los estudiantes asumen la recta como una forma reducida de un conjunto de datos representados en un diagrama de dispersión y que dicha recta refleja la relación entre las dos variables, el objetivo próximo sería caracterizar la mejor recta. La Actividad 6, "Encontrando la mejor recta", se divide en dos partes. Con la primera parte, a través de un juego, lo que se pretende es que los estudiantes empiecen a trazar sus primeras rectas. Aquí se evidenciarán concepciones ya reportadas en investigaciones anteriores pero que se espera que los estudiantes vayan abandonando a medida que se desarrolla el juego y se discute con cada uno de ellos. Al finalizar esta primera parte los estudiantes deberán emitir una forma para trazar la recta que mejor se ajusta dado cualquier diagrama de dispersión.

La forma que ellos propongan se pondrá a prueba con la segunda parte de la Actividad 6, puesto que nuevamente trazarán la recta bajo su estrategia y ésta será validada de acuerdo a las propiedades que cumple la mejor recta que, en un principio, están implícitas en el diseño de la misma actividad. Aquí se espera que los estudiantes ratifiquen o mejoren su estrategia.

En la Tabla 63 se define el objetivo de la Fase 2 así como su posible desarrollo en cuanto a la secuencia de instrucción.

Tabla 63.
Fase 2 de la THA propuesta para la Recta Informal de Mejor Ajuste

Fase	Objetivo(s) específicos(s)	Análisis de datos (claves)	Interpretaciones y conceptos claves para desarrollar	Hipótesis de razonamientos de los estudiantes	Actividad(es)	Herramienta(s) de apoyo
2. Análisis de conjuntos de datos bivariados alrededor de la recta informal de mejor ajuste	Extender las ideas desarrolladas en la Fase 1 al caso bivariado para encontrar la recta informal de mejor ajuste	•Proceso de generación de datos •Examinar la variabilidad •Reconocimiento de la asociación, o no, entre dos variables cuantitativas y el tipo de asociación •Representación de los datos en un gráfico de dispersión •Discutir el significado de cada punto en un gráfico de dispersión •Predicción local del valor de la variable respuesta sobre gráficos con datos apilados verticalmente •Resumir el gráfico de dispersión de datos apilados (es posible que encuentren la media para cada pila de datos o un valor entre los datos de cada pila) •Discutir sobre la forma o tendencia que siguen los puntos en un gráfico de	•Los datos bivariados pueden ser visualizados en un diagrama de dispersión •Los datos en un diagrama de dispersión consisten en las medidas de dos atributos de un número de casos •Cada punto en un diagrama de dispersión significa un solo caso cuyas medidas se indican por su ubicación con respecto a los ejes •El hecho de que una medida en la variable explicativa tenga varias medidas en la variable respuesta se debe a la variabilidad •La forma, tendencia o patrón de los conjuntos de datos apilados, o no apilados, describen una asociación y el nivel de asociación, entre los dos	•Asumir la media de una de las variables o de todo el conjunto de datos como una forma de predecir. •Negación de una asociación entre dos variables cuando esta es inversa. •Concepciones de tipo causal y local para predecir •La recta que mejor se ajusta a una nube de puntos es: la que más puntos toca, la que deja igual número de puntos a lado y lado de la recta, la más cercana a los puntos, la que une el primer y último punto... •Dada la recta que mejor se ajusta a una nube de puntos, la predicción se ubica	Actividad 2 Actividad 3 Actividad 4 Actividad 5 Actividad 6	Geogebra

		dispersión. •Discutir sobre una predicción general: una forma que resuma el conjunto de datos (se espera las primeras discusiones alrededor de la recta) •Caracterizar la recta de mejor ajuste •Discutir sobre cómo usar la recta para predecir un valor para cualquier valor de la variable explicativa	variables. •Si la relación es de tipo lineal, el conjunto de datos se puede resumir a una recta •La recta de mejor ajuste se caracteriza por cumplir ciertas propiedades •La recta puede ser utilizada para hacer predicción.	sobre la recta o muy cerca de la recta.

5.2. Formas de razonamiento informal alrededor de la recta de mejor ajuste en estudiantes de octavo grado al finalizar la implementación de la THA.

Se podría decir que todos los estudiantes terminaron por asumir, en el caso univariado, a la media como un buen candidato para hacer predicciones pero siempre de la mano de la variabilidad, es decir, siendo un valor bastante preciso pero, no necesariamente, exacto. Por otra parte, la mayoría de los estudiantes, a través de las actividades, caracterizaron la media como un punto de equilibrio o balanceador, también como un valor justo y como un valor muy particular, al presentar ciertas propiedades.

En cuanto al caso bivariado, la mayoría de los estudiantes reflejan una coordinación entre las dos variables al asociar expresiones como “al aumentar una variable aumenta la otra” con una “relación ascendente” y, de igual manera, “al aumentar una variable disminuye la otra” con una “relación descendente”. Así mismo, el nivel de asociación lo relacionan con la forma que siguen los puntos, siguiendo un “orden” o reflejando, de manera global, una línea recta. En este mismo aspecto, en la entrevista final, se evidenció un buen grado de aceptación en cuanto al grosor de la nube de puntos, es decir, si esta es más “flaca” el nivel de asociación es más fuerte. Respecto a predicción, a menos que se les sugiera trazar la recta, los estudiantes reflejan una concepción local al predecir observando los valores de los vecinos cercanos pero, a su vez, reflejan una visión global de los datos para que sus predicciones sean coherentes con el tipo de relación que tienen las variables.

Por lo que corresponde a la ubicación de la recta de mejor ajuste, al finalizar la trayectoria, la mayoría de los estudiantes le atribuyen a esta recta características como: se ubica entre los puntos, refleja la relación entre las dos variables y es la más cercana a los puntos. Aunque se cuestionó sobre qué entendían por cercanía o a qué se referían con esa expresión, sus respuestas fueron prácticamente nulas o evasivas, es decir, aunque, tanto en el caso univariado como en el caso bivariado, los estudiantes evidenciaron que las sumas de los cuadrados de las distancias a la media y a la recta, respectivamente, es mínima, ese hecho no les dio luz de que nos encontrábamos ante el valor y la recta más cercana a todos y cada uno de los datos. Por lo que bien vale la pena seguir investigando en este aspecto: ¿Qué comprenden o a qué le atribuyen los estudiantes el término "cercanía"? ¿Cómo miden esa cercanía?

Finalmente, un componente que siempre estuvo presente a lo largo de la THA fue la variabilidad. Lo anterior se evidenció cuando, dada la recta de mejor ajuste, se les pedía predecir. La mayoría de los estudiantes emitían sus predicciones muy cerca de la recta y los que ubicaban sus predicciones sobre la recta, de igual manera, adicionaban "o muy cerca de la recta". La justificación fue la variabilidad, como si asumieran que sobre la recta sería el valor exacto pero que, debido a la variabilidad, no necesariamente tenía que ser así. Es decir, en el momento de predecir, la confrontación: cercanía vs variabilidad, aparecía como una manera de controlar la variabilidad como "no podemos ser tan exactos pero sí muy cercanos".

5.3. Limitaciones y sugerencias para futuras investigaciones

Como ya lo hemos mencionado, los estudiantes que participaron en el desarrollo de esta investigación fueron estudiantes que previamente no habían recibido ningún tipo de formación en cuanto a recta y ecuación de la misma. Lo anterior lo asumimos desde un principio cuando elegimos nuestra muestra para llevar a cabo esta investigación, puesto que queríamos ver esas formas de razonamiento en ausencia de métodos matemáticos como lo es la recta y su ecuación. Sin embargo, al igual que lo señala Casey (2015) sería interesante mirar y confrontar esas formas de razonamiento covariacional cuando esos métodos matemáticos ya hacen parte del conocimiento informal de los estudiantes.

Por otra parte, como vimos, tanto en el caso univariado como en el caso bivariado, se exploró una de las propiedades de la media y de la recta, respectivamente, que consistió en la suma de los cuadrados de las diferencias y se exploró cuándo esta suma era mínima. Los estudiantes a través de las simulaciones dieron cuenta que esta suma era mínima cuando se hacía sobre la media y sobre la recta de mejor ajuste, sin embargo no es fácil relacionar este hecho con el término de cercanía. Por lo que sería pertinente abordar más sobre qué entienden, cómo asumen, o cómo miden esa cercanía.

Consideramos que el hecho de empezar con un análisis univariado alrededor de la media y sus propiedades nos permitió extender características como "cercanía" y "entre" al caso bivariado. Sin embargo en algunos casos, cuando los datos eran tabulados, el cálculo de la media se convirtió en una salida para hacer predicción sin contar con la relación implícita de las dos variables medidas. Por lo que se hace necesario trabajar más en problemas donde los estudiantes tengan que pasar de una representación a otra para no perder el hilo conductor cuando se vean enfrentados a datos representados en forma tabular como ocurrió en un ítem de la prueba final.

Otro aspecto en el que se hace importante profundizar es la confrontación que se presentó en nuestros estudiantes entre cercanía vs variabilidad, en el caso univariado esta confrontación se evidenció en estudiantes que asumían un intervalo como un predictor, y en el caso bivariado cuando, teniendo la recta de mejor ajuste, la ubicaban tímidamente tan cerca de la recta como pudieran, como si ese nuevo dato se obligara a pertenecer al conjunto de datos inicial. De manera que sería bastante interesante profundizar en este aspecto.

Finalmente la THA, alrededor de la recta informal de mejor ajuste, que aquí se propone es solo una de las muchas que se podrían dar en una clase normal. Por el momento, esta THA provee a los profesores en ejercicio de actividades para que lleven a sus estudiantes a un primer acercamiento previo al futuro abordaje formal de la recta de mejor ajuste.

Referencias bibliográficas

Alloy, L. B. & Tabachnik, N. (1984). Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review*, 91(1), 112-149.

Bargagliotti, A., Anderson, C., Casey, S., Everson, M., Franklin, C., Gould, R., Groth, R., Haddock, J. & Watkins, A. (2014). Project-SET Materials for the teaching and learning of Sampling Variability and Regression. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute.

Bargagliotti, A., Anderson, C., Casey, S., Everson, M., Franklin, C., Gould, R., Groth, R., Haddock, J., & Watkins, A. (2012). *Project-SET linear regression learning trajectory*. http://project-set.com/presentations/121712-regressionlp-final-released/

Batanero, C. & Godino, J. (2001) Análisis de Datos y su Didáctica. Departamento de Didáctica de la Matemática Universidad de Granada.

Ben-Zvi, D., & Gardfiel, J. (2004). Statistical Literacy, Reasoning, and Thinking: Goals, Definitions, and Challenges. En D. Ben-Zvi, & J. Gardfiel (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 3-15). Dordrecht, The Netherlands: Kluber Academy Publishers.

Borovcnik, M. (2012). Correlation and Regression – Errors and Pitfalls. *Selçuk Journal of Special*, special issue, 51-68.

Carlson, M., Jacobs, S., Coe, E., Larsen, S., & Hsu, E. (2002). Applying covariational reasoning while modeling dynamic events: A framework and a study. *Journal for Research in Mathematics Education*, *33*(5), 352-378.

Casey, S.A. (2008). *Subject matter knowledge for teaching statistical association*. Tesis doctoral. Universidad de Illinois.

Casey, S. (2014). Teachers' knowledge of students's conceptions and their development when learning linear regression. En K. Makar, B. de Sousa, & R. Gould (Eds.), Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA. Voorburg, The Netherlands: International Statistical Institute.

Casey, S. (2015). Examining Student Conceptions of Covariación: A Focus on the Line of Best Fit. *Journal of Statistics Education* 23 (1), 1-33.

Clements, D. & Sarama, J. (2004). Learning trajectories in mathematics education. *Mathematical Thinking and Learning* 6 (2), 81-89.

Cobb, P., McClain, K. & Gravemeijer, K. (2003). Learning About Statistical Covariation, *Cognition and Instruction*, 21(1), pp. 1-78.

Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin 90* (2), 272-292.

delMas, R. (2004). A comparison of Mathematical and Statistical Reasoning. En D. Ben-Zvi, & J. Gardfiel (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 79-95). Dordrecht, The Netherlands: Kluber Academy Publishers.

Engel, J. & Sedlmeier, P. (2011). Correlation and regression in the training of teachers. En C. Batanero, G. Burrill y C. Reading (Eds.), *Teaching statistics in school mathematics challenges for teaching and teacher education: A Joint ICMI/IASE study* (pp,247-258). New York: Springer.

Estepa, A. (1993). *Concepciones iniciales sobre la asociación estadística y su evolución como consecuencia de una enseñanza basada en el uso de ordenadores*. Tesis doctoral. Universidad de Granada.

Estepa, A., Gea, M., Cañadas, G. & Contreras, J. (2012). Algunas notas históricas sobre la correlación y regresión y su uso en el aula. *Revista de Didáctica de las Matemáticas Números,* 81, 5-14.

Estepa, A. & Batanero, C. (1995). Concepciones iniciales sobre la asociación estadística. *Revista Enseñanza de las ciencias*, 13(2), 155-170.

Estepa, A., & Sanchez Cobo, F. (2001). Empirical research on the understanding of association and implications for the training of researchers. In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 37–51). Granada, Spain: International Association for Statistical Education.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., y Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A preK-12 curriculum framework.* Alexandria, VA: American Statistical Association.

Garfield, J. & Ben-Zvi, D. (2008). Learning to Reason about Covaration. In *Developing Students' Statistical Reasoning*, eds. D. Ben-Zvi y J. Garfield, Springer Sciende+Business Media B.V. pp. 289-308.

Glaser, B. & Strauss, A. (1967). *The Discovery of Grounded Theory*, Chicago, IL: Aldine.

Gea, M. (2013). *La correlación y regresión en bachillerato: Análisis de libros de texto y del conocimiento de los futuros profesores.* Tesis doctoral. Universidad de Granada, España.

Lavalle, A., Micheli, E. & Rubio, N. (2006). Análisis didáctico de regresión y correlación para la enseñanza media. *Revista Relime*, 9(3), 383-406.

Ministerio de Educación Nacional. (2004). Pensamiento variacional y tecnologías computacionales. Proyecto: Incorporación de nuevas tecnologías al currículo de matemáticas de la educación básica secundaria y media de Colombia.

Moore, D. (1995). *Estadística aplicada básica.* Barcelona: Antoni Bosch.

Moreno, L. (2014). *Educación Matemática: Del signo al píxel.* Bucaramanga, Colombia: Universidad Industrial de Santander.

Moritz, J. (2004). Reasoning about Covariation. En D. Ben-Zvi y J. Garfield (eds), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 227-255). The Netherlands: Kluver Academic Publishers.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston: VA.

Pfannkuch, M. & Wild, C. (2004). Towards an understanding of Statistical Thinking. En D. Ben-Zvi, & J. Gardfiel (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 17-46). Dordrecht, The Netherlands: Kluber Academy Publishers.

Sánchez, F., Estepa, A. & Batanero, C. (2000). Un estudio experimental de la estimación de la correlación a partir de diferentes representaciones. *Revista Enseñanza de las ciencias*, 18(2), 297-310.

Sanz de Acedo, M. (2001). La argumentación: Una forma de razonamiento informal. *Revista de Psicología General y Aplicaciones,* 54 (3), 355-370.

Simon, M. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. Journal for Research in Mathematics Education, 26(2), 114-145.

Simon, M. A. y Tzur, R. (2004). Explicating the role of mathematical tasks in conceptual learning: an elaboration of the hypothetical learning trajectory. *Mathematical Thinking and Learning*, 6(2), 91-104.

Truran, J. (1995). Some undergraduates' understanding of the meaning of a correlation coefficient, en B. Atweh, y S. Clavel (eds.), *Proceedings of the Eigteenth Annual Conference of the Mathematics Education Research Group of Australasia (MERGA),* (pp. 524-529). Darwin, Australia: Northern Territory University.

Watkins, A. E., Scheaffer, R. L., & Cobb, G. W. (2004). *Statistics in action: Understanding a world of data*. Emeryville, CA: Key Curriculum Press.

Wild, C. J., y Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67* (3), 223-265.

Zieffler, A. S. (2006). *A longitudinal investigation of the development of college students' reasoning about bivariate data during an introductory statistics course*. Tesis doctoral. Universidad de Minnesota.

Zieffler, A, & Garfield, J. (2009). Modeling the growth of students' covariational reasoning during an introductory statistics course. *Statistics Education Research Journal*, 8 (1), 7-31.

Zieffler, A., Garfield, J., Delmas, R. & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 5-19.

Apéndices

Apéndice A: Prueba Diagnóstica 1. Interpretando gráficos

Prueba Diagnóstica 1

Institución Educativa San Francisco

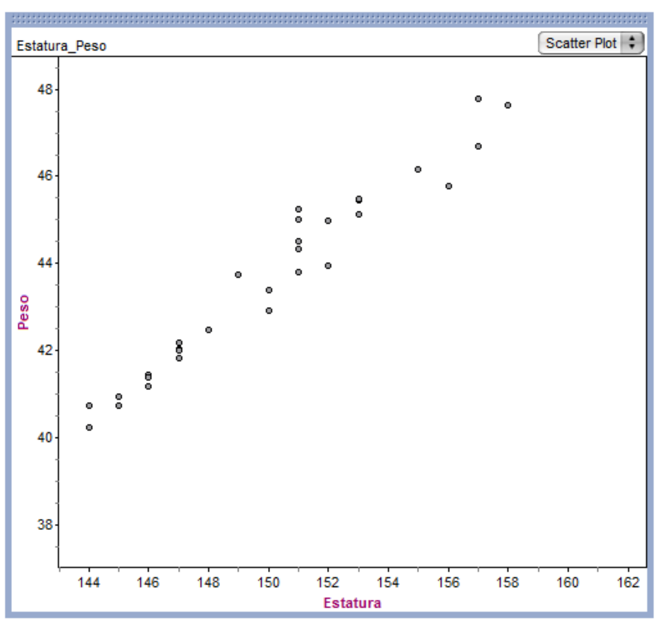
Octavo grado

Nombre: ____Código: _________

Interpretando gráficos

Para cada una de las siguientes preguntas escoge la opción que consideres que mejor describe lo que se muestra en el respectivo gráfico. **Justifica la opción que escojas.**

Gráfico A. Estatura (en cm) y peso (en kg) de 30 niños de un curso de octavo grado entre 12 y 13 años. Cada punto del plano se corresponde con la altura y el peso de un estudiante del grupo.



1. Respecto a la relación entre la estatura y el peso de un niño entre 12 y 13 años se puede concluir que

a. los estudiantes con menor estatura tendrán un peso mayor.
b. los estudiantes con mayor estatura tendrán un peso mayor.
c. un estudiante con determinada estatura puede tener más de un peso.
d. no es posible establecer una relación entre la estatura de los estudiantes y el peso.
e. Otra opción ¿cuál y por qué?

2. Para un estudiante que mide 154 cm, se puede predecir que su peso, aproximadamente,

a. es de 44 kg
b. es de 46 kg
c. está entre 44 kg y 46 kg
d. El gráfico no proporciona información para predecir ese valor.
e. Otra opción ¿cuál y por qué?

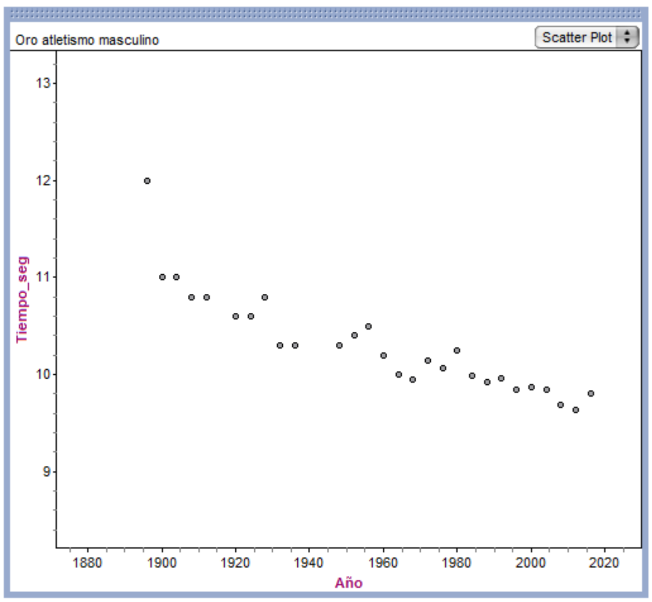
3. Si alguien te pregunta "*¿cuánto podría pesar un estudiante que mide 151 cm?*" Y tuvieras que dar un solo valor ¿cuál de las siguientes opciones escogerías? (Nota: Para 151 cm el gráfico proporciona los siguientes cinco valores (en kg), aproximadamente: 45,24; 45,02; 44,50; 44,34 y 43,80).

f. 45,24 kg porque es el mayor valor.
a. 43,80 kg porque es el menor valor.
b. 44,50 kg porque está "en el medio" de los cinco valores.

c. 44,58 kg porque es el valor promedio de los cinco valores.

d. Otra opción ¿cuál y por qué?

Grafico B. Año y tiempo (en segundos) de los medallistas olímpicos (hombres) en atletismo en los 100 metros planos. Cada punto del plano se corresponde con el año y el tiempo en correr 100 metros planos.



4. Respecto a la relación entre el año y el tiempo en que tardan en correr 100 metros planos los medallistas olímpicos, se puede concluir que

a. el tiempo en el que los medallistas olímpicos corren 100 metros planos ha disminuido a lo largo del tiempo.

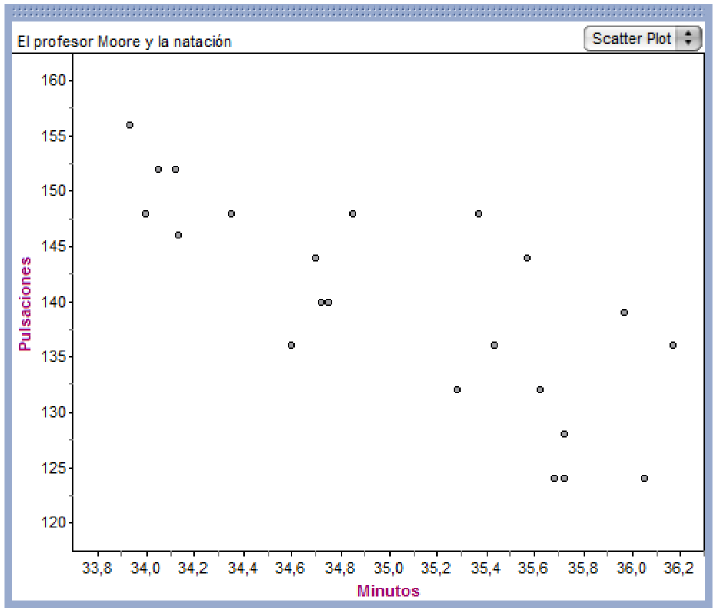
b. el tiempo en el que los medallistas olímpicos corren 100 metros planos ha aumentado a lo largo del tiempo.

c. el tiempo en el que los medallistas olímpicos corren 100 metros planos no ha cambiado a lo largo del tiempo.
d. No es posible establecer una relación entre el año de los juegos olímpicos y el tiempo en que tardan los medallistas olímpicos en correr 100 metros planos.
e. Otra opción ¿cuál y por qué?

5. Debido a la primera y segunda Guerra Mundial, los juegos olímpicos no se celebraron en los años 1916, 1940 y 1944. De acuerdo a lo que se observa en el Gráfico B, si se hubieran realizado los juegos olímpicos para el año 1944, se puede predecir aproximadamente, que el tiempo que hubiera tardado el medallista olímpico de ese año en correr los 100 metros planos

a. es de 10,47 segundos.
b. es de 10,30 segundos.
c. está comprendido entre 9,63 y 12 segundos.
d. El gráfico no proporciona información para predecir ese valor.
e. Otra opción ¿cuál y por qué?

Gráfico C. El profesor Moore y la natación. He aquí los tiempos (en minutos) que tarda el profesor Moore en nadar 1800 metros y su ritmo cardíaco después de bracear (en pulsaciones por minuto) en 23 sesiones de natación. Cada punto del plano se corresponde con el tiempo en nadar 1800 metros y el ritmo cardíaco.



6. Respecto a la relación entre el tiempo que tarda el profesor Moore en nadar los 1800 metros y su ritmo cardíaco, se puede concluir que

a. un número mayor de pulsaciones indica que el profesor Moore tardó más tiempo en nadar los 1800m.

b. entre más tiempo tarde el profesor Moore en nadar los 1800m su rimo cardíaco será mayor.

c. un número menor de pulsaciones indica que el profesor Moore tardó menos tiempo en nadar los 1800m.

d. entre menos tiempo tarde el profesor Moore en nadar los 1800m su ritmo cardíaco será mayor.

e. Otra opción. ¿cuál y por qué?

7. Juan, un alumno del profesor Moore, asegura que si para la siguiente sesión el profesor Moore tarda 34,30 minutos en nadar los 1800m, entonces su ritmo cardíaco será de 152 pulsaciones por minuto, aproximadamente. Respecto a la afirmación que hace Juan, tú estarías

a. de acuerdo ya que para 34,30 minutos le corresponde exactamente 152 pulsaciones.
b. en desacuerdo, ya que el gráfico no me proporciona información para ese valor.
c. de acuerdo, ya que el gráfico me proporciona información para predecir ese valor.
d. en desacuerdo, ya que la única manera de saber esa información es que el profesor tarde exactamente 34,30 minutos en nadar los 1800m.
e. Otra opción ¿cuál y por qué?

8. María, una alumna del profesor Moore, asegura que es posible predecir el ritmo cardíaco del profesor Moore para cualquier tiempo, comprendido entre 33,8 y 36,2 minutos, que tarde en nadar los 1800m. Respecto a la afirmación que hace María, tú estarías

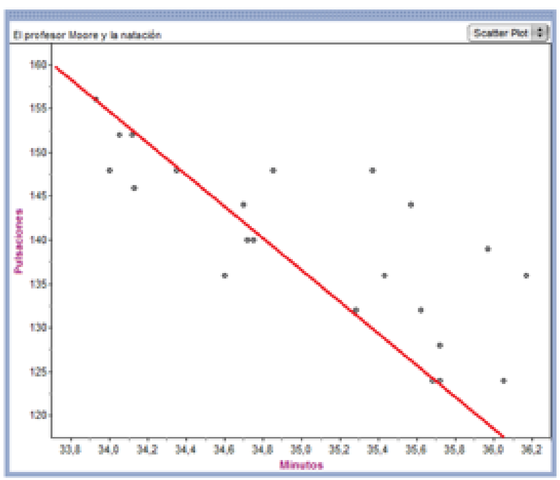
a. de acuerdo, ya que para cualquier tiempo se puede tomar un valor cercano a éste y observar el correspondiente número de pulsaciones, el cual será muy similar al que se busca.
b. en desacuerdo, porque para algunos valores de tiempo el gráfico no muestra el correspondiente número de pulsaciones.
c. de acuerdo, ya que se puede establecer el aspecto general de la relación entre las dos variables.

d. en desacuerdo, ya que no se puede establecer el aspecto general de la relación entre las dos variables.

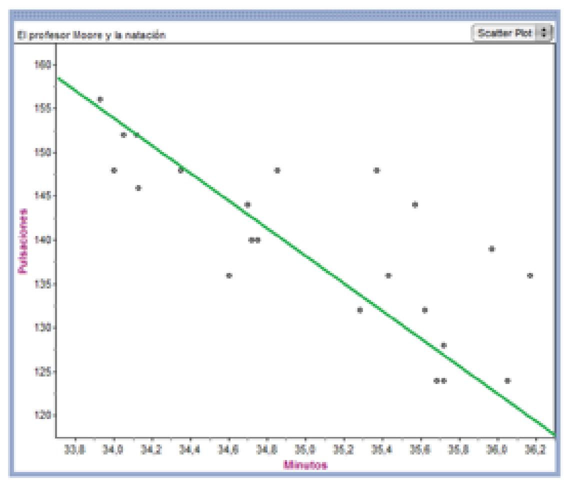
e. Otra opción ¿cuál y por qué?

Se les pidió a cuatro estudiantes que trazaran una recta sobre el Gráfico C teniendo en cuenta que la recta se ajustara de la mejor manera a los datos del gráfico. Las siguientes gráficas muestran la respuesta de los cuatro estudiantes.

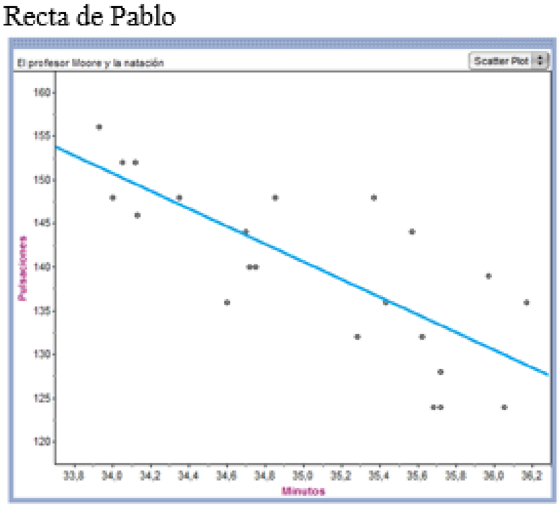
Recta de Susana



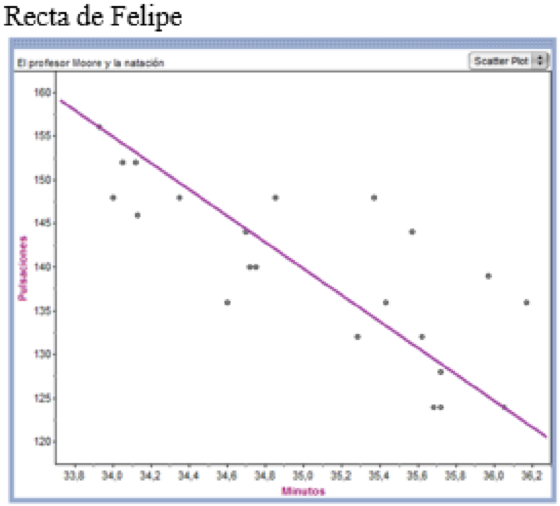
Recta de Julieta



Recta de Pablo



Recta de Felipe



9. Respecto a lo solicitado en la tarea, tú estarías de acuerdo con la recta trazada por [Justifica tu respuesta]

a. Susana

b. Julieta

c. Pablo

d. Felipe

e. Otra opción (traza tu recta y justifica)

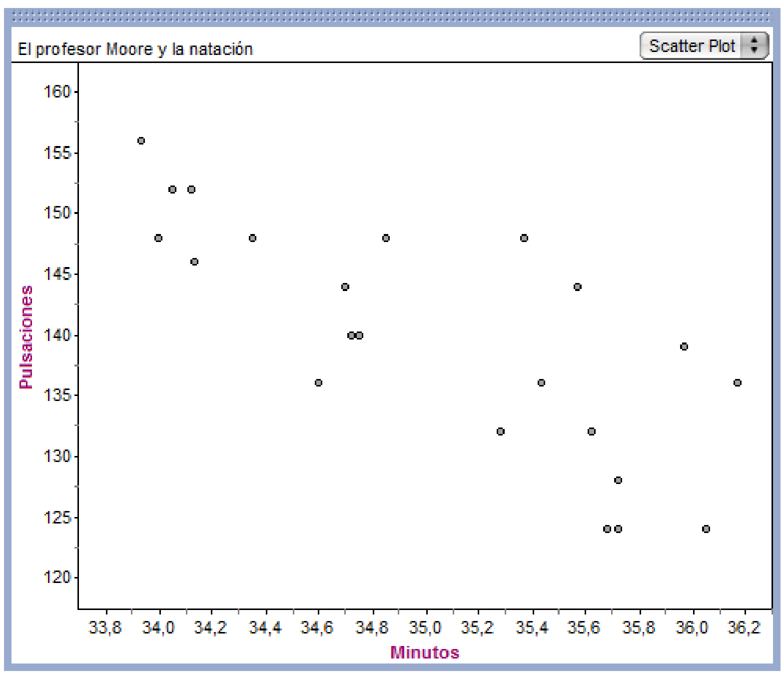
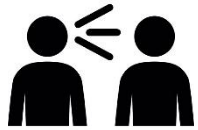
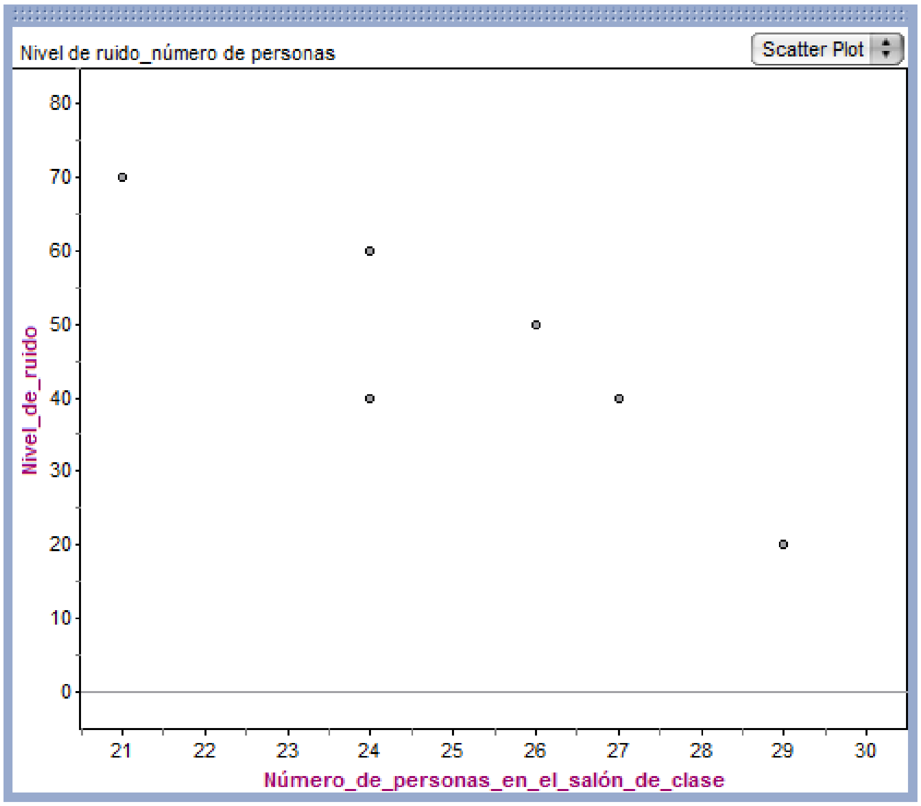


Gráfico D. Algunos estudiantes estaban haciendo un proyecto sobre el ruido, de manera que visitaron 6 salones de clase diferentes y midieron el nivel de ruido en la clase con un medidor de sonido. Además contaron el número de personas que había en cada salón de clase en el momento de la medición.





10. Escoja la opción que mejor describe el gráfico anterior y justifique su elección.

a. Entre más personas haya en un salón de clases el ruido será mayor.

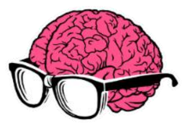
b. Entre más personas haya en un salón de clases el ruido será menor.

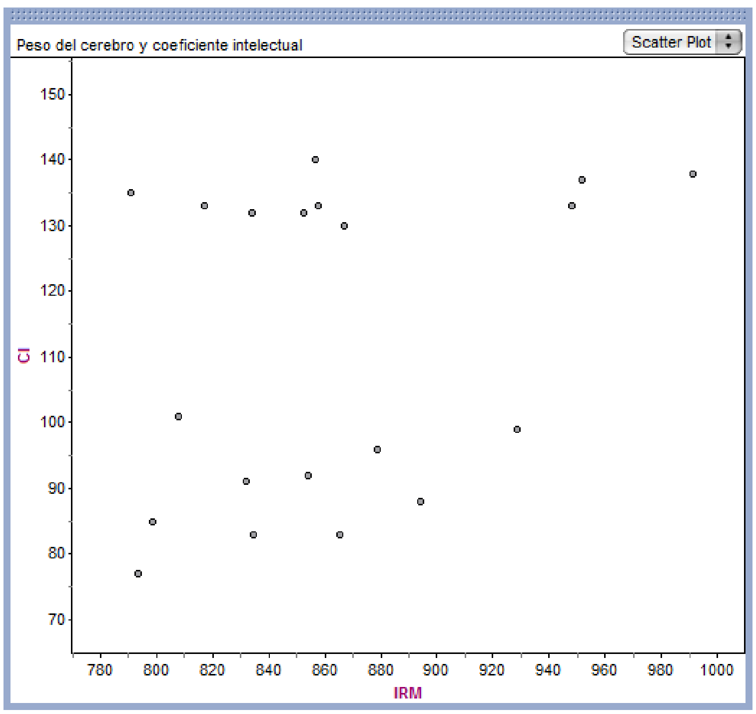
c. El gráfico muestra que el nivel del ruido no está relacionado con el número de personas en la clase.

d. El gráfico no proporciona información para establecer si existe una relación o no entre el número de personas en un salón de clases y el nivel de ruido.

e. Otra opción ¿cuál y por qué?

Gráfico E. Peso del cerebro y el coeficiente intelectual (CI) de 20 mujeres. El peso del cerebro se determinó mediante una imagen obtenida por resonancia magnética (IRM). Cada punto del plano se corresponde con el peso del cerebro y el coeficiente intelectual.





11. Respecto a la relación entre el peso del cerebro de una mujer y su coeficiente, se puede concluir que

a. las mujeres que tienen un cerebro con mayor peso tienen un coeficiente intelectual mayor.

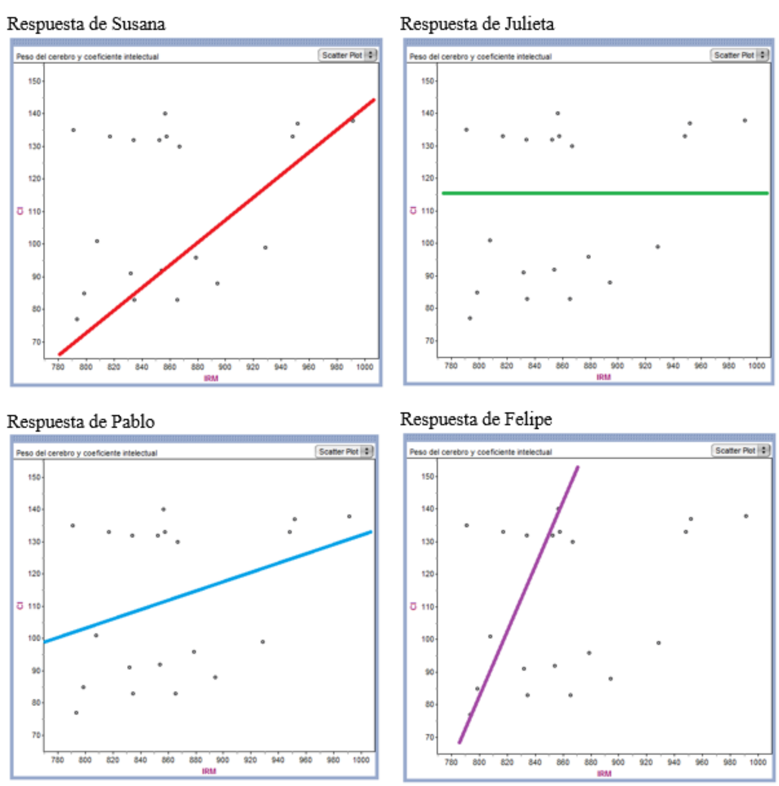
b. las mujeres que tienen un cerebro con mayor peso tienen un coeficiente intelectual menor.

c. no existe una relación entre el peso del cerebro de una mujer y su coeficiente intelectual.

d. El gráfico no proporciona información para establecer si existe una relación o no.

e. Otra opción ¿cuál y por qué?

Se les pidió a cuatro estudiantes que trazaran una recta sobre el Gráfico E teniendo en cuenta que la recta se ajustara de la mejor manera a los datos del gráfico. Las siguientes gráficas muestran la respuesta de los cuatro estudiantes.



12. Respecto a lo solicitado en la tarea, tú estarías de acuerdo con la recta trazada por [Justifica tu respuesta]

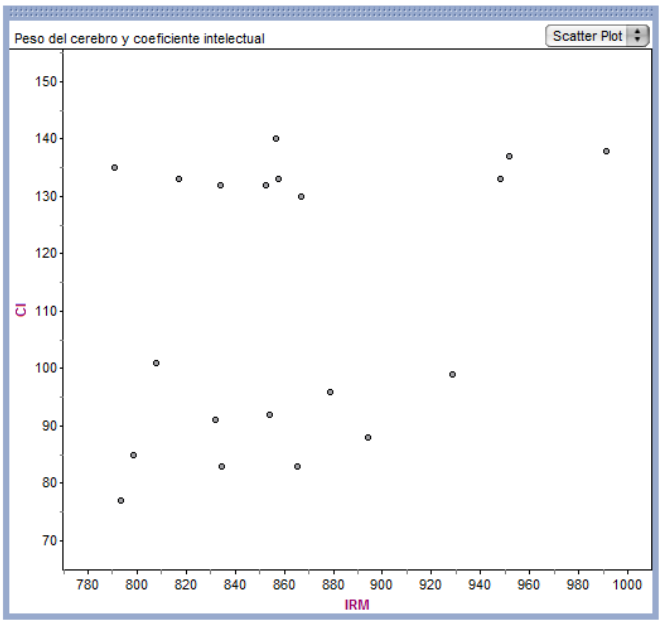
a. Susana

b. Julieta

c. Pablo

d. Felipe

e. Otra opción (traza tu recta y justifica)



Apéndice B: Prueba Diagnóstica 2. Predicción y tendencia

Prueba Diagnóstica 2

Colegio: ________________________________

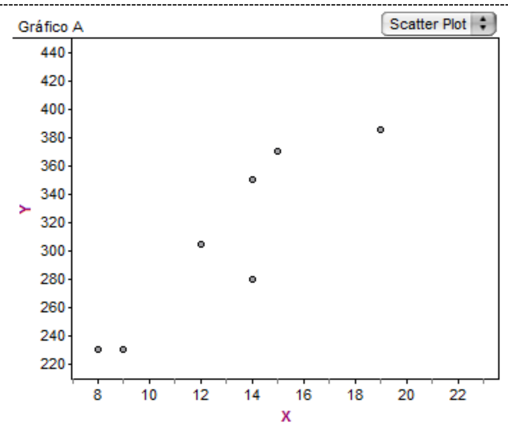
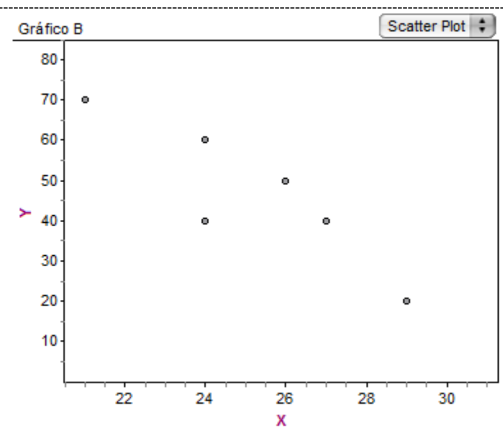
Octavo grado

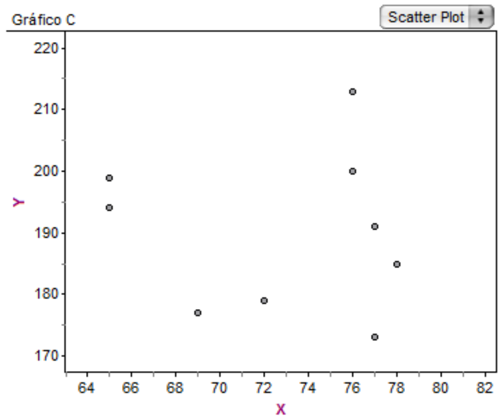
Nombre: ____ Código: _________

Predicción y tendencia

Responde cada una de las siguientes preguntas CON BASE EN LO QUE SE MUESTRA EN EL RESPECTIVO GRÁFICO.

1. Observa los gráficos A, B, y C y ubica sobre el respectivo gráfico el valor de Y que corresponde al valor de X que se te da. Justifica tu respuesta.

<table>
<tr><td>
</td><td>X=22 Justificación:</td></tr>
<tr><td>
</td><td>X=30 Justificación:</td></tr>
</table>

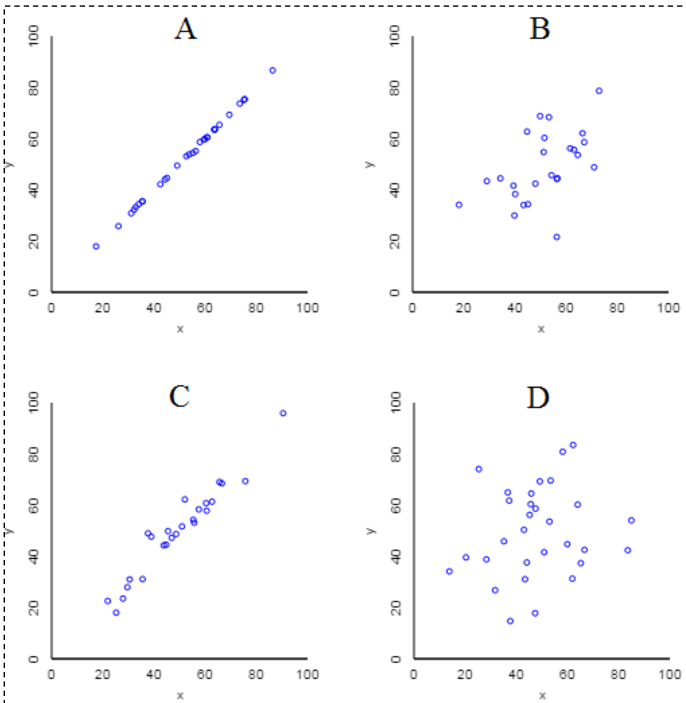


X=80 Justificación:

X=64 Justificación:

2. Observa los siguientes diagramas (A, B, C y D) y ordénalos de mayor a menor de acuerdo a la intensidad de la asociación entre las variables x e y. Pon en las siguientes líneas el orden que tú consideres.

__, __, __, __ ¿Por qué?



3. Algunos estudiantes estaban haciendo un proyecto sobre el ruido, de manera que visitaron 6 salones de clase diferentes y midieron el nivel de ruido en la clase con un medidor de sonido. Además contaron el número de personas que había en cada salón de clase en el momento de la medición. Los datos se muestran en la siguiente tabla y se visualizan en el Gráfico D, donde en el eje x está el número de personas y en el eje y el nivel de ruido.

Salón	Número de personas	Nivel de ruido
A	24	50
B	26	50
C	21	20
D	27	60
E	24	30
F	29	70

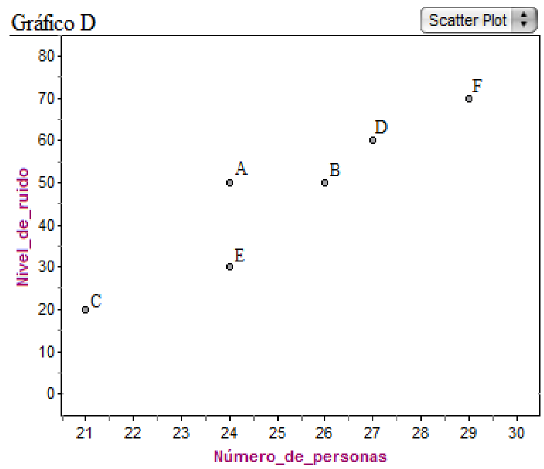


Tabla y Gráfico D para el problema 3.

3a. A Juan le correspondía medir el nivel de ruido para un salón con 28 personas y por alguna razón no pudo hacerlo. Con base en el Gráfico D ¿Cuál crees que sea el nivel de ruido para el salón que le correspondió a Juan? Explícale a Juan cómo encontraste ese valor.

3b. Te han invitado a participar en este proyecto para el cual se te ha pedido predecir el nivel de ruido para un salón con 25 personas. Observa el Gráfico D y predice el valor que se te ha pedido. Explica tu procedimiento.

4. Pizzas Floridablanca ofrece a sus clientes 9 variedades de pizza. Para cada clase de pizza se ha medido su contenido de grasa (en gramos) y calorías. Los datos se muestran en la siguiente tabla y se visualizan en el Gráfico E, donde en el eje x está la cantidad de grasa y en el eje y las calorías.

Clase de pizza	Grasa (gr)	Calorías
A	9	230
B	19	385
C	14	280
D	12	305
E	8	230
F	14	350
G	15	370

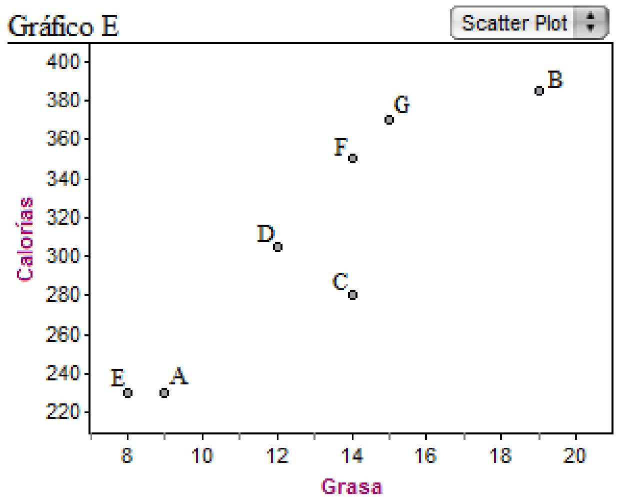
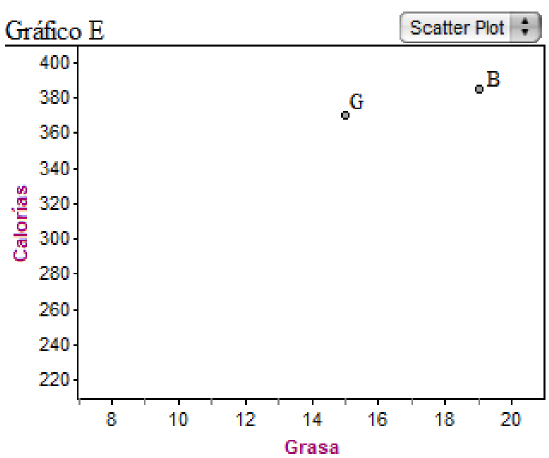
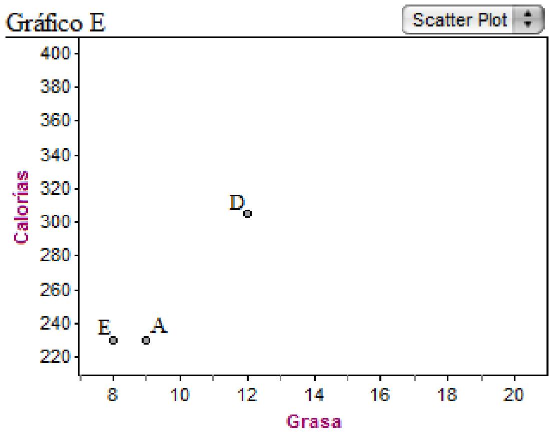


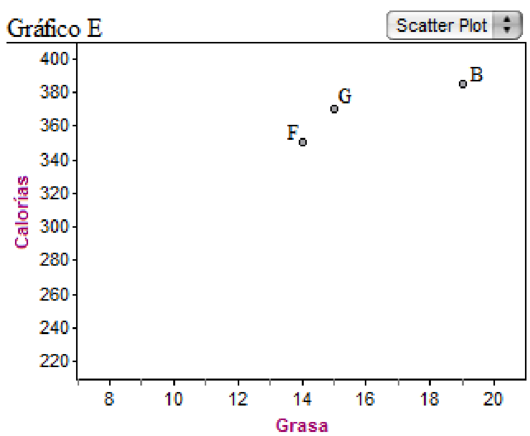
Tabla y Gráfico E para el problema 4.



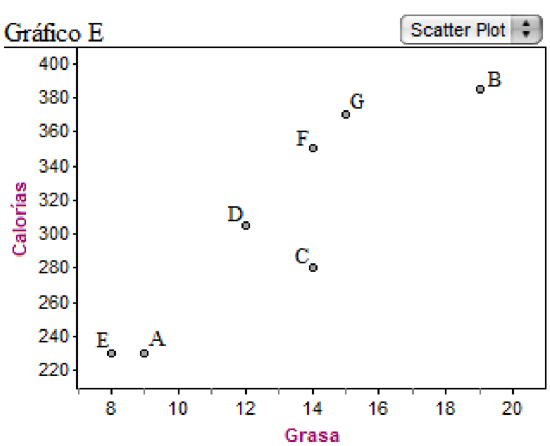
4a. Traza la recta que mejor se ajusta a los puntos G y B. Explica cómo la trazaste.



4b. Traza la recta que mejor se ajusta a los puntos E, A y D. Explica cómo la trazaste.



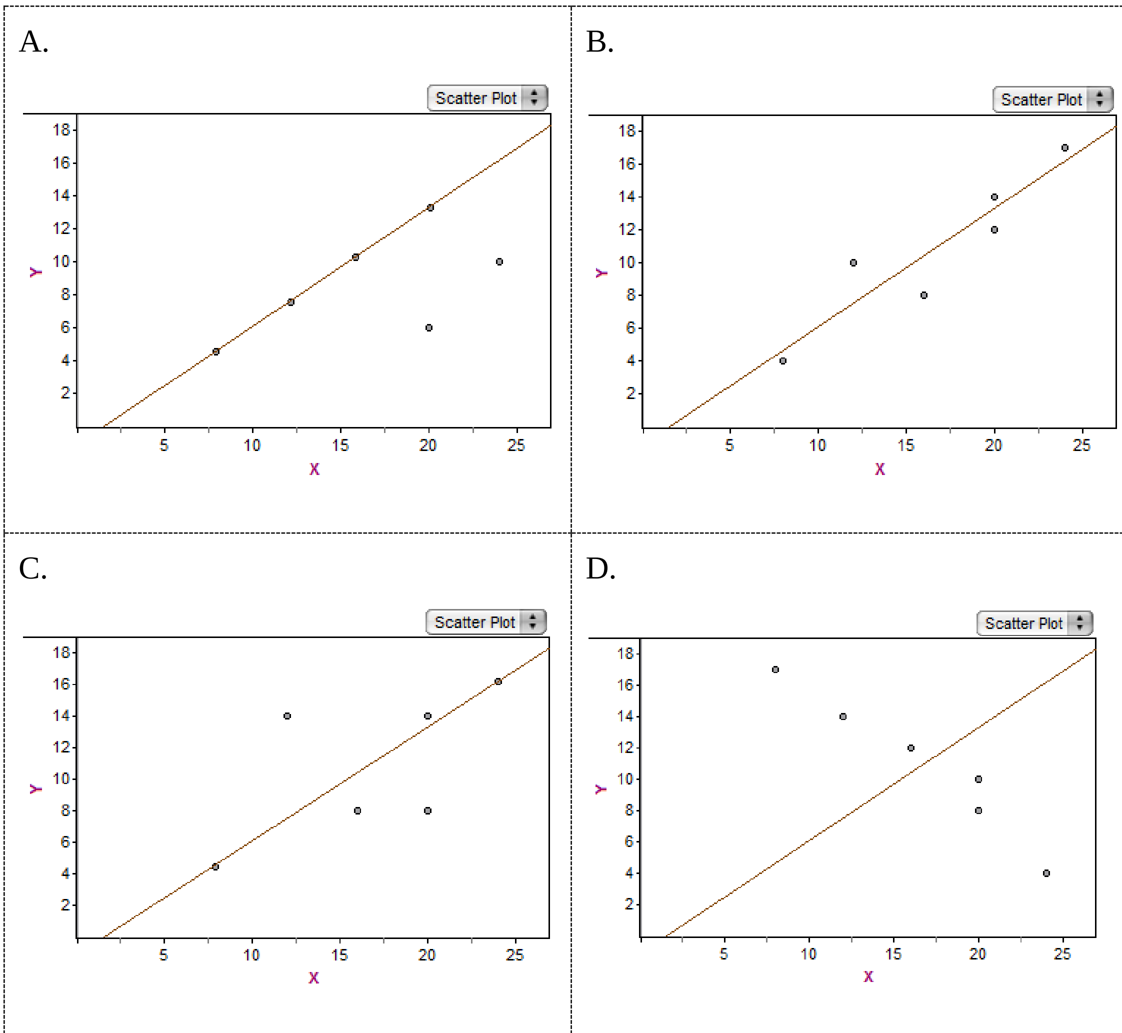
4c. Traza la recta que mejor se ajusta a los puntos F, G y B. Explica cómo la trazaste.



4d. Traza la recta que mejor se ajusta a todos los puntos. Explica cómo la trazaste.

4e. ¿Por qué crees que la recta que trazaste en 3d es la que mejor se ajusta a los datos?

5. En los siguientes gráficos se muestra un conjunto de seis puntos y una recta que es la misma para todos los gráficos y se asume que esa recta refleja el comportamiento de los seis puntos. Ordena los gráficos de acuerdo al nivel de ajuste de la recta a los seis puntos, es decir, el primer gráfico que escojas será el gráfico en el cual la recta se ajusta mejor a la nube de puntos dada y el último que escojas, la que menos se ajusta. Explica la razón del orden dado.



Pon en las siguientes líneas el orden que tú consideres.

__, __, __, __ Explicación:

Apéndice C: Actividad 0. Obteniendo datos del curso al que pertenezco

Actividad 0. Obteniendo datos del curso al que pertenezco

Colegio: ________________________________

Octavo grado

Nombre: ____ Código: _________

Tomando y registrando datos

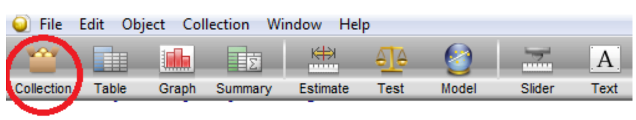
Dirígete con tu grupo de trabajo a cada uno de tus compañeros de curso y pregúntale o toma la medida de cada uno de estos atributos: Nombre, Edad, Género, Peso y Estatura. Registra los datos, inclusive los tuyos, en la siguiente tabla.

N°	Nombre	Edad (años)	Género (F o M)	Peso (Kg)	Estatura (cm)
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					

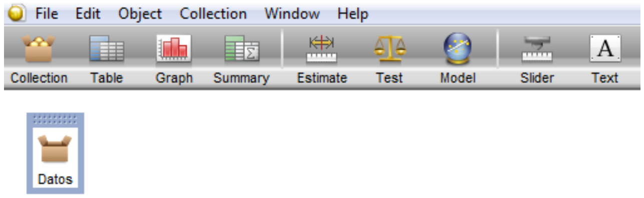
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					
31					
32					
33					
34					
35					
36					
37					
38					
39					
40					

Registro de datos en una tabla en Fathom

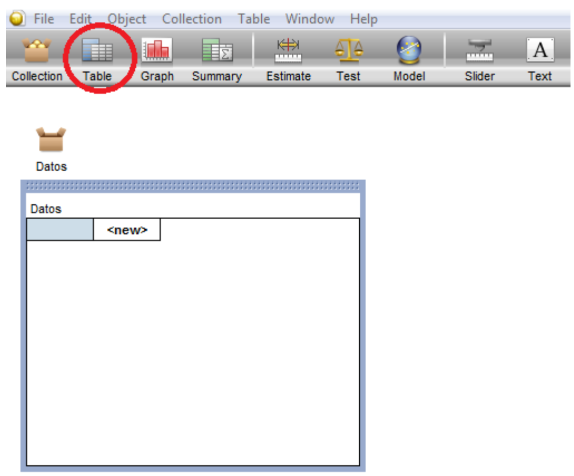
Abre el software Fathom y selecciona “Collection”, aparecerá una mano como si estuviera “agarrando” ese objeto. Mueve la mano que ves hasta un espacio en blanco y da clic.



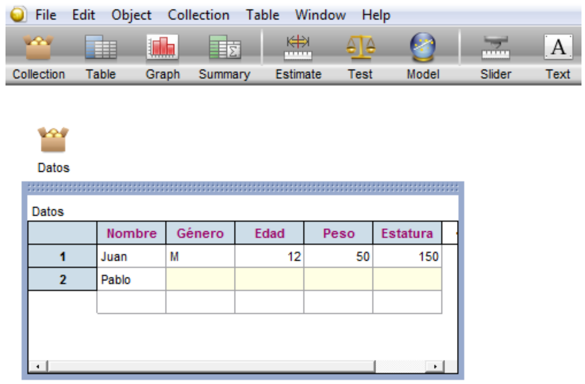
Aparecerá una “caja” que, por defecto, se llama “Collection 1”. Da doble clic sobre la caja y cambia el nombre Collection 1 por Datos.



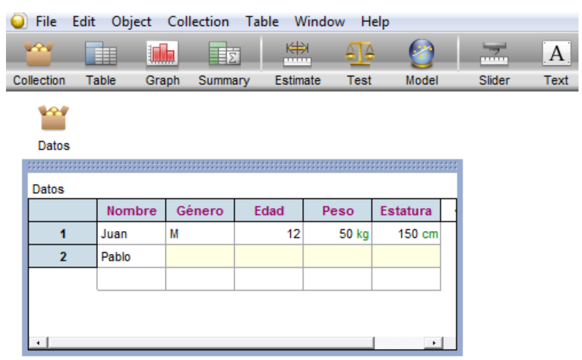
A continuación, vamos a registrar los datos que tenemos en una tabla. Da clic en la caja "Datos", para asegurarnos de que está seleccionada (dentro de un marco azul), y enseguida selecciona en la parte superior "Table", y mueve la mano hasta un espacio en blanco de tu pantalla y da clic. Enseguida aparecerá una tabla llamada "Datos". Si en lugar de "Datos" se llama "no data" significa que no seleccionaste previamente la colección "Datos".



En la tabla en "<new>" escribe el primer atributo que tomaste, por ejemplo "Nombre", y da "Enter". Repite lo anterior en cada <new> que va apareciendo hasta terminar con todos los atributos que tomaste: Nombre, Género, Edad, Peso y Estatura. Si deseas modificar algún nombre, solo da doble clic y cámbialo. Enseguida registra debajo de cada atributo los datos que tomaste; extiende la tabla si deseas visualizar todos los datos.



Podemos también ponerle unidades a Peso y Estatura, por ejemplo, kg y cm, respectivamente. Para lo anterior, sobre la tabla da clic izquierdo y selecciona “Show Units”. Enseguida aparecerá una fila en blanco llamada “units”. Debajo de peso escribe “kg” y debajo de estatura “cm”. A continuación se mostrará las unidades en todos los datos para estos dos atributos. Si no deseas ver la fila de “units”, da clic izquierdo sobre la tabla y selecciona “Hide Units”.

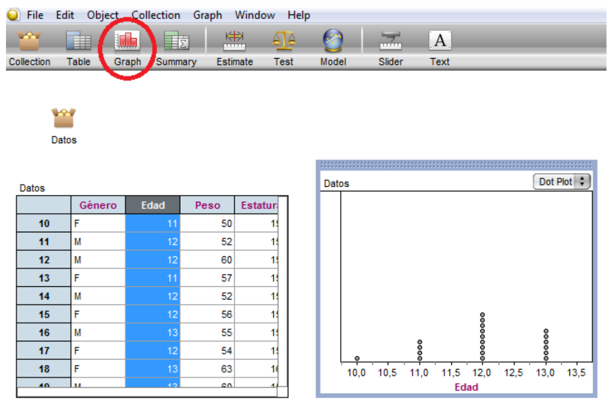


Observa los datos y reflexiona sobre las siguientes preguntas:

1. ¿Cuántos estudiantes aparecen en los datos? ¿Cuántas variables?

2. ¿Qué nos dicen los anteriores datos sobre ese curso de octavo grado? Es decir, si tuvieras que describir el curso a un amigo, quien no conoce el curso, ¿qué le podrías decir?

Otra manera de visualizar los datos es a través de un gráfico que nos ayudaría a responder preguntas como ¿Qué edad tienen la mayoría de los estudiantes del curso? Para lo anterior, selecciona en la parte superior “Graph”, muévelo hasta un espacio en blanco de la pantalla y da clic. Aparecerá un recuadro en blanco. Enseguida selecciona “Edad” en la tabla y, manteniendo la selección, dirígete hasta el nuevo recuadro donde dice “Drop an attribute here”.



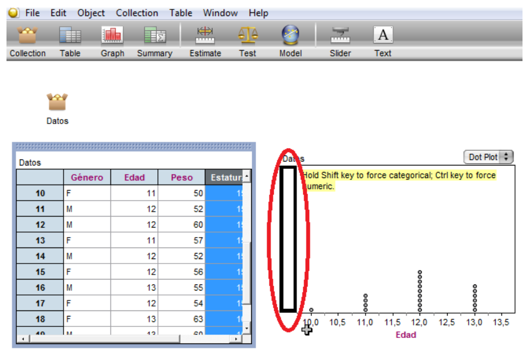
Dado que cada grupo tomó las medidas de manera independiente, reflexiona sobre la siguiente pregunta con el fin de llegar a un acuerdo y manejar los mismos datos de ahora en adelante.

3. ¿Por qué tenemos diferentes mediciones, en peso y estatura por ejemplo, para un mismo estudiante? Y si tenemos que quedarnos con un solo valor ¿con cuál valor nos quedamos? ¿por qué? Discútelo con los demás grupos.

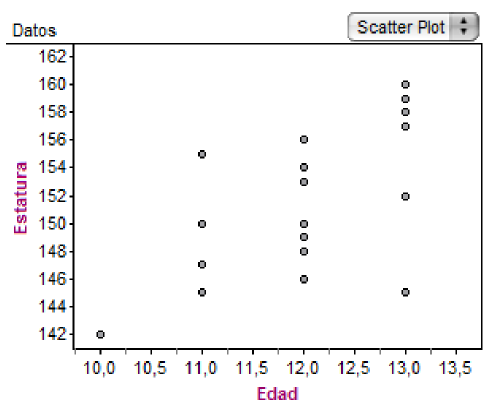
4. Imagina la siguiente situación. Un nuevo estudiante (hombre) de 12 años llega a este curso ¿podrías predecir su estatura? ¿Cómo lo harías?

5. Para fijar nuestra atención en solo los estudiantes (hombres) de 12 años, podríamos visualizar los datos de la siguiente manera:

Crea un gráfico y ubica Edad tal como lo hiciste en el punto f. A continuación, selecciona Estatura en la tabla y ubícalo en el gráfico en el espacio donde se muestra a continuación.

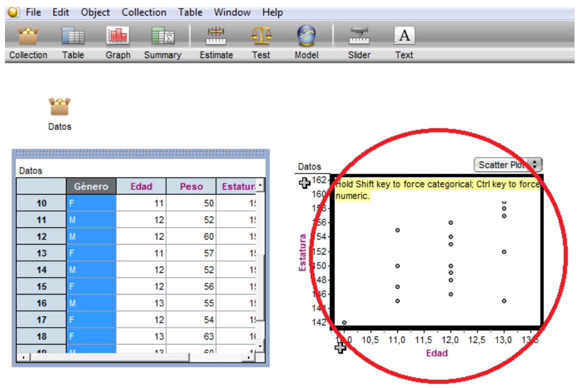


Obtendrás el siguiente gráfico

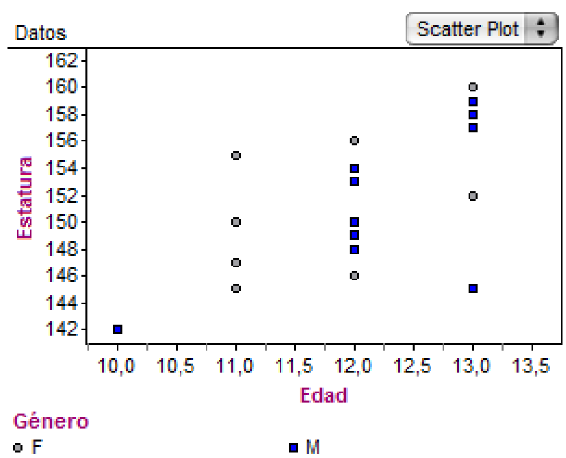


¿Cuántos puntos hay en el gráfico y qué significa cada punto en el gráfico? Explica y discútelo con tus compañeros.

6. Como observarás, el nuevo gráfico no nos dice cuáles medidas corresponden a los hombres y cuáles a las mujeres. Para saber cuáles pertenecen a hombres y cuáles a mujeres, una manera de proceder es la siguiente: En tu tabla selecciona “Género” y “suéltalo” en el gráfico previamente generado, tal como se observa en la siguiente imagen, es decir, en el plano mismo, en ninguno de los ejes.

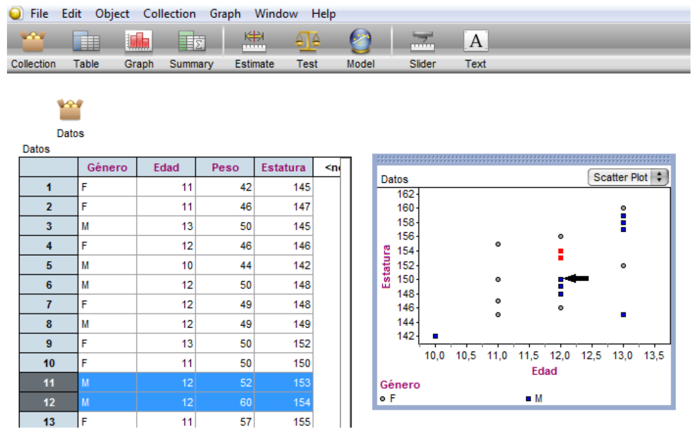


Obtendrás el siguiente gráfico



Describe lo que observas en el nuevo gráfico.

7. Si te ubicas sobre uno de los puntos, observarás que aparece una flecha que lo señala, además si das clic verás que el punto se pone en rojo y además se resalta en azul en la tabla. Si quieres seleccionar varios puntos al tiempo, usa la tecla “Shift” y, manteniéndola oprimida, selecciona los puntos los cuales se irán resaltando en tu tabla; asegúrate de extender la tabla hasta visualizar todos los datos.



Tomando en cuenta todo lo anterior, responde nuevamente la situación planteada en el inciso 4: “Un nuevo estudiante (hombre) de 12 años llega a este curso ¿podrías predecir su estatura? ¿Cómo lo harías?”. Discute tu respuesta con tus compañeros y explica cómo la obtuviste.

Apéndice D: Actividad 1 (Parte 1). El Machín Machón

Actividad 1 (Parte 1). El Machín Machón

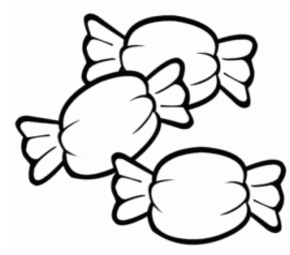
Colegio: ________________________________

Octavo grado

Nombre: ____ Código: _________

Situación A: Repartiendo caramelos

Cierta cantidad de caramelos serán distribuidos entre ustedes. Registra en la siguiente tabla el nombre, de acuerdo a la cantidad de caramelos que te correspondió.

Nombre	# de caramelos	
	6	
	4	
	7	
	6	
	7	
	3	
	2	

1. ¿Estás de acuerdo en que la distribución de los caramelos es injusta? Explica por qué.

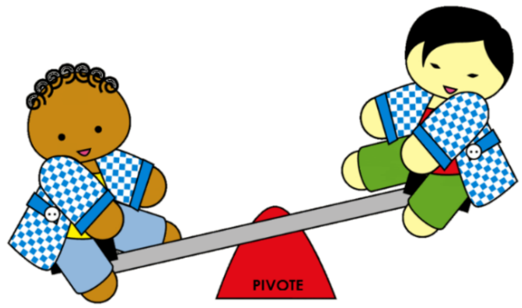
2. Considerando que la distribución es injusta, da o recibe caramelos de tus compañeros hasta que consideres que todos tienen la cantidad justa de caramelos. Describe todas las acciones que hiciste hasta lograr lo anterior.

3. Resuelto el ítem 2, ¿por qué crees que ahora todos tienen la cantidad justa de caramelos?

4. Cómo calcularías la cantidad justa de caramelos a partir de los datos presentados en la tabla. Descríbelo.

Situación B: Balanceando el Machín Machón

Dos niños, Édinson y Gabriel, del mismo peso, se encuentran sentados a una distancia x_1, x_2 metros, respectivamente, del extremo izquierdo de un Machín Machón en el cual desean jugar. El Machín Machón tiene una longitud de 10 metros, bastante grande.



1. Para cada una de los siguientes casos indique a qué distancia, del extremo izquierdo, ubicarías el pivote para que el Machín Machón quede balanceado. Explica tu respuesta a tus compañeros

Caso a. $x_1 = 0$ y $x_2 = 10$

Caso b. $x_1 = 2$ y $x_2 = 8$

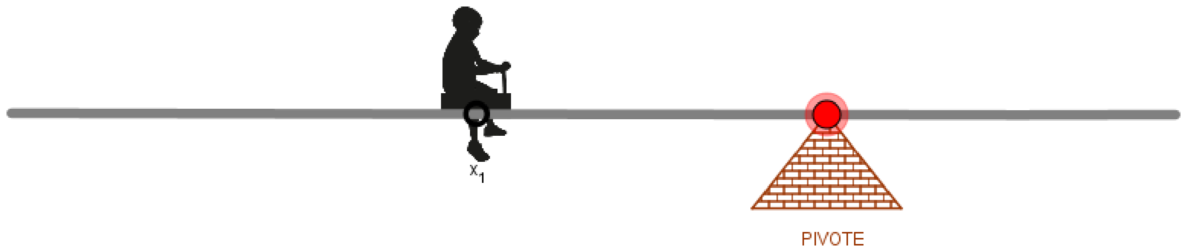
Caso c. $x_1 = 3$ y $x_2 = 9$

Caso d. $x_1 = 2$ y $x_2 = 10$

Caso e. ¿En general, qué tendrías en cuenta para ubicar el pivote, dado cualquier valor de x_1 y x_2? Explica tu respuesta.

2. En cada uno de los casos anteriores, ¿qué tan lejos está cada niño al pivote que ubicaste?

3. Ahora considera la siguiente situación: Édinson se encuentra sentando a una distancia $x_1 = 4$ metros del extremo izquierdo del Machín Machón y el pivote a 7 metros. ¿A qué distancia del pivote, ubicarías a Gabriel para que el Machín Machón se mantenga balanceado? Explica tu respuesta.



4. En general, ¿a qué distancia, del extremo izquierdo, ubicarías el pivote para que el Machín Machón quede balanceado, dado cualquier valor de x_1 y x_2? Explica tu respuesta.

5. Ahora imagina la siguiente situación:

Tres niños, Édinson, Gabriel y Cristian, del mismo peso, se encuentran sentados a una distancia x_1, x_2, x_3 metros, respectivamente, del extremo izquierdo de un Machín Machón de diez metros en el cual desean jugar, como se muestra en la siguiente imagen.

Suponga que el primer niño, Édinson, se encuentra sentado a un metro del extremo izquierdo del Machín Machón, es decir, $x_1 = 1$, el segundo, Gabriel, a tres metros del extremo izquierdo, $x_2 = 3$ y el tercero, a ocho metros, $x_3 = 8$.

¿A qué distancia, del extremo izquierdo, ubicarías el pivote para que el Machín Machón quede balanceado? Explica tu respuesta a tus compañeros y apunta la de tus compañeros.

6. Abre el archivo MachinMachon.ggb en donde visualizarás la siguiente pantalla

Da clic en el recuadro blanco "Mostrar Pivote" para visualizar el pivote. Selecciona el punto rojo y, manteniendo la selección (u oprimiendo la tecla →), mueve el punto rojo y observa el valor que va tomando "UbicaciónPivote". Detente cuando ese valor llegue hasta el valor que diste en la pregunta anterior (5).

¿Quedó balanceado el Machín Machón en el valor que diste? De no ser así, explora con los otros valores que dieron tus compañeros.

Si con tu valor, o los de tus compañeros, no has logrado balancear el Machín Machón, selecciona el recuadro blanco ¡Eureka! y mueve libremente el pivote hasta que el Machín Machón esté balanceado. Sabrás que está balanceado cuando aparezca un símbolo verde en los extremos del Machín Machón y la palabra "¡Eureka!".

Observa el valor que tomó "UbicaciónPivote" y reflexiona sobre cómo encontrar ese valor.

7. Cambia los valores de las casillas en blanco "Edinson x_1, Gabriel x_2 y Cristian x_3" los cuales no deben exceder a 10 pues esa es la longitud del Machín Machón y repite lo que hiciste en el punto anterior.

Observa el valor que tomó "UbicaciónPivote" y reflexiona sobre cómo encontrar ese valor.

8. Repite muchas veces lo anterior hasta que encuentres un criterio, que funcione, para encontrar la ubicación del pivote dada cualquier posición de los tres niños en el Machín Machón. Discute tu criterio con tus compañeros y muéstrales que sí funciona.

9a. Volvamos al caso en que Édinson se encuentra sentado a un metro del extremo izquierdo del Machín Machón, es decir, $x_1 = 1$, Gabriel, a tres metros del extremo izquierdo, $x_2 = 3$ y Cristian a ocho metros, $x_3 = 8$. ¿Qué tan lejos está cada niño del pivote en situación de equilibrio?

Para visualizar claramente las distancias, balancea el Machín Machón y observa que sale una casilla llamada “ON”, a continuación cambia “0” por “1” y responde qué tan lejos está cada niño del pivote.

9b. ¿Qué notas en particular de las distancias de los niños al pivote? Escribe una expresión algebraica. Discútelo con tus compañeros.

10. Ahora considera la siguiente situación:

Cambia los valores x_1, x_2, x_3 por 2, 4, 1, respectivamente, y mueve el pivote hasta el valor de 5. Observarás que el Machín Machón no quedó balanceado. Sin mover el pivote, cambia el valor de SOLO UNA de las casillas, x_1, x_2 o x_3, da Enter. Lo anterior hasta lograr que el Machín Machón quede balanceado.

Nota: Asegúrate que el recuadro “Eureka” esté chuleado.

¿Qué puedes concluir de lo anterior?

Apéndice E: Actividad 1 (Parte 2). Machín Machón

Actividad 1 (Parte 2). Machín Machón

Colegio: ________________________________

Octavo grado

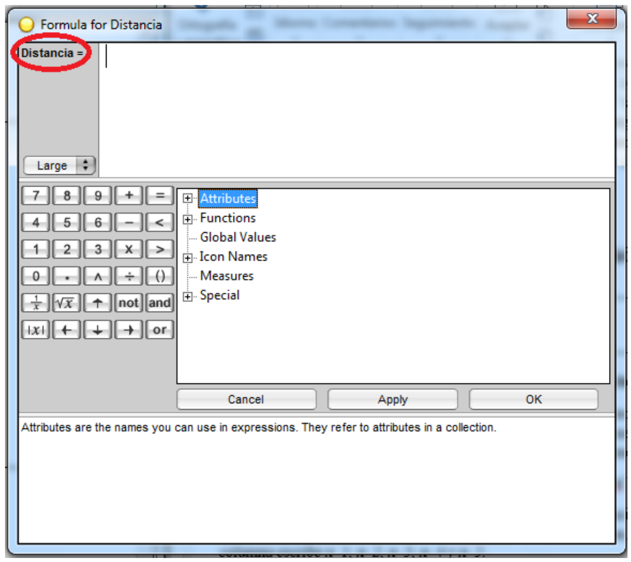
Nombre: ____ Código: _________

Balanceando el Machín Machón

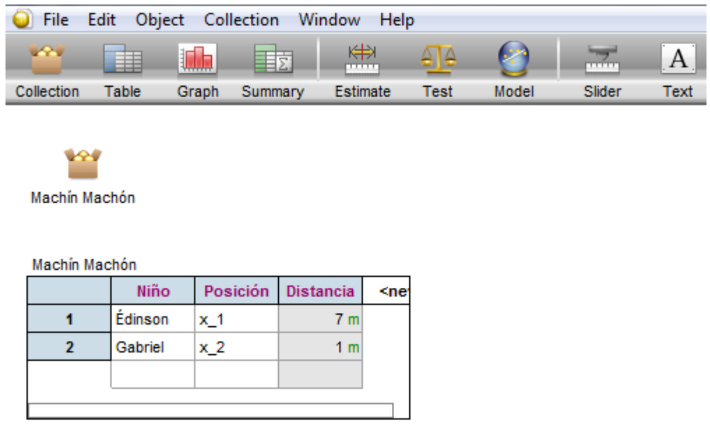
Dos niños, Édinson y Gabriel, del mismo peso, se encuentran sentados a una distancia x_1, x_2 metros, respectivamente, del extremo izquierdo de un Machín Machón en el cual desean jugar. El Machín Machón tiene una longitud de 10 metros.

- Abre un nuevo documento en Fathom y crea una colección llamada “Machín Machón”.

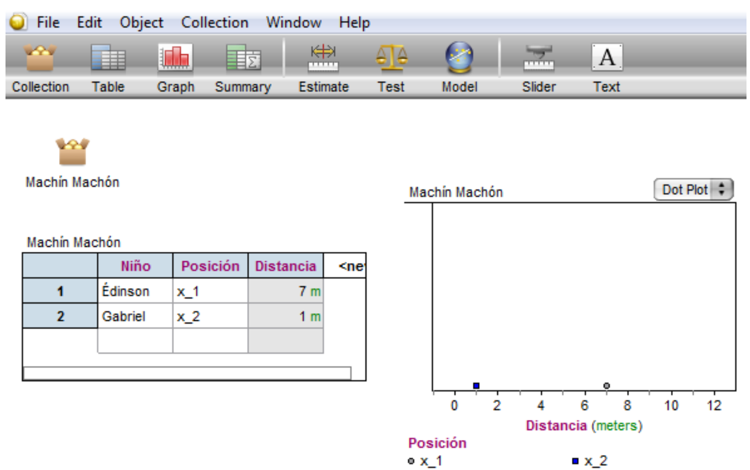
- Elabora una tabla con tres columnas: Niño, Posición, Distancia. En la columna “Niño” escribe en cada celda los respectivos nombres: Édinson y Gabriel. En la segunda columna escribe x_1, para Édinson y x_2 para Gabriel.
- En la tercera columna vas a adicionar una fórmula de manera que cada vez que oprimas “Ctrl+y” los valores de las distancias cambien entre 0 y 10. Para lo anterior, da clic derecho sobre “Distancia” y selecciona “Edit Formula”; se abrirá la siguiente ventana:



- En el espacio superior en blanco escribe randomInteger(10), luego da clic en “Apply” y en seguida en “OK”. Verifica cómo cambian los valores en la tabla cada vez que das “Ctrl+y”. Puedes adicionarle las unidades a Distancia, en *meters*.

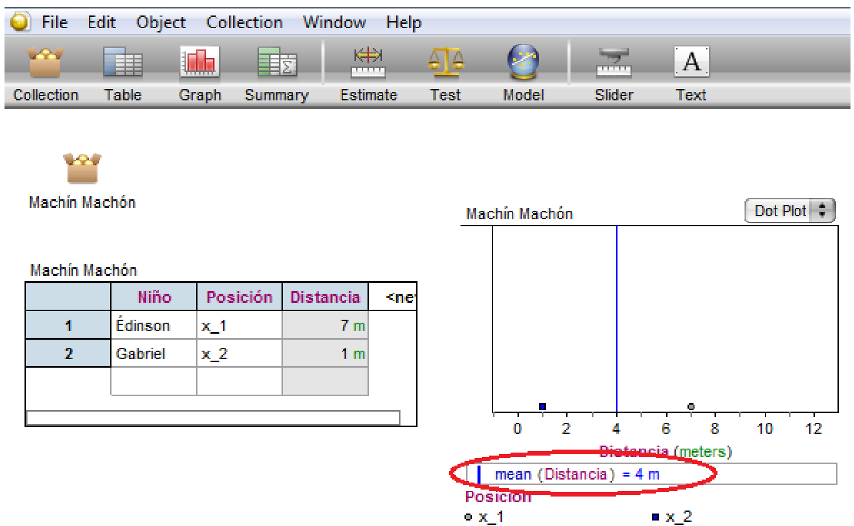


- Visualiza los datos en una gráfica ubicando "Distancia" en el eje horizontal. Observarás que los puntos son indistinguibles; para evitar lo anterior, selecciona "Posición" y llévalo hasta un espacio en blanco de la gráfica; verás cómo cada punto toma diferente forma, como se muestra en la siguiente imagen.

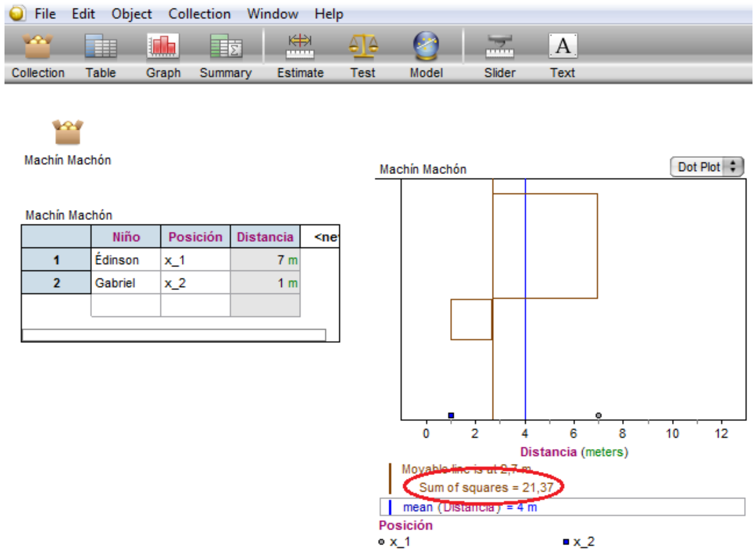


1. Encuentra qué valor debería tomar el pivote, es decir la media, según las posiciones que tienes, para que el Machín Machón quede balanceado. Repite lo anterior cada vez que des "Ctrl+y".

- Ubica el valor de la media sobre el gráfico. Para lo anterior, da clic derecho sobre el gráfico y selecciona “Plot Value”. En el espacio en blanco escribe mean(), en el paréntesis selecciona en la parte de abajo el “+” junto a “Attributes” y selecciona con doble clic “Distancia”, dale “Apply” y en seguida “OK”. Se visualizará una línea azul. Verifica que el valor donde se ubica la línea azul corresponde a la media que tú ya habías calculado previamente.



Descubriendo otra propiedad. Vamos a asociar cada distancia, del punto al valor donde se ubica la media, con un cuadrado. Para lo anterior da clic derecho sobre el gráfico y selecciona "Add Movable Line", observarás que sale una línea color café. Hecho lo anterior, da clic nuevamente sobre el gráfico y selecciona “Show Squares”; extiende la gráfica hasta visualizar totalmente los cuadrados. Observa que, a diferencia de la línea azul, la café la puedes mover.



2a. Describe cómo se forma cada cuadrado

2b. Describe la forma de calcular el área de cada cuadrado

2c. Describe qué significaría la suma de las áreas de todos los cuadrados, “Sum of squares”.

- Mueve la línea café y observa cómo cambia el valor de "Sum of squares" antes y después de la media (línea azul).

2d. ¿Qué se puede decir del valor de "Sum of squares" cuando las dos líneas (la café y la azul) coinciden? Plantea una conclusión de acuerdo a lo que observas en el gráfico.

3. Adiciona en la tabla una tercera fila con Cristian y x_3. Responde nuevamente el inciso 2.

4. Describe la propiedad de la media que acabas de observar.

Apéndice F: Actividad 2. Actividades del profesor Gabo

Actividad 2. Actividades del profesor Gabo

Colegio: ________________________________

Octavo grado

Nombre: ____ Código: _________

- **De la casa a la universidad**

El profesor Gabo, quien vive a unos kilómetros del campus universitario de la Universidad Industrial de Santander, ha registrado durante 10 días consecutivos, excepto uno, el tiempo en minutos, que tarda conduciendo desde su casa hasta la universidad. En la siguiente tabla se muestran los registros del profesor Gabo.

	Tiempo (minutos)
1	15.0
2	10.0
3	8.5
4	
5	7.6
6	8.7
7	7.8
8	10.4
9	8.9
10	9.5

1a. Describe la distribución de los tiempos que tarda el profesor Gabo conduciendo desde su casa a la universidad. Explica por qué no se demora siempre la misma cantidad de tiempo.

1b. Como se observa, para el día 4 no hay registro puesto que el profesor Gabo olvidó hacerlo. Ten en cuenta los valores de la tabla y predice el tiempo que tardó en llegar a la universidad ese día. Explica cómo encontraste o calculaste ese valor.

1c. Predice el tiempo que tardará el profesor Gabo para el día 11, teniendo en cuenta el que pusiste en el inciso b para el día 4. Explica cómo encontraste o calculaste ese valor.

- **Practicando natación**

El profesor Gabo, quien en su tiempo libre practica natación, ha registrado durante 23 sesiones consecutivas, el tiempo, en minutos, que tarda en nadar 1800 metros y su ritmo cardíaco, en pulsaciones por minuto, una vez ha completado los 1800 metros. Los tiempos y las pulsaciones se registran en la siguiente tabla.

	Minutos	Pulsaciones por minuto
1	34	152
2	36	124
3	35	140
4	34	152
5	34	155
6	36	128
7	36	128
8	35	144
9	35	148
10	36	130
11	35	136
12	36	124
13	35	148
14	35	144
15	35	140
16	34	156
17	35	136
18	34	148
19	34	148
20	36	132
21	36	124
22	35	132
23	36	130

2a. Predice el tiempo y las pulsaciones del profesor Gabo para la siguiente sesión. Explica tu respuesta.

- Abre un nuevo documento en Fathom y crea una colección llamada Practicando Natación. Elabora una tabla con los datos dados y, en seguida, dos gráficos, uno para minutos y otro para pulsaciones.

2b. Con base en los gráficos describe los tiempos y las pulsaciones del profesor cuando nada los 1800 metros; hazlo de tal forma como si le estuvieras contando a un amigo, sin mostrarle los datos, acerca de la actividad de natación del profesor Gabo.

2c. Para cada uno de los gráficos, ubica el valor de la media y compáralo con los que encontraste en el inciso 2a. ¿Qué tan cercanos son tus valores al de la media?

2d. ¿Qué ventajas tiene asumir la media como predictor?

2e. Predice las pulsaciones si un día, el profesor Gabo, completa los 1800 metros en 34 minutos. Explica tu respuesta.

- Elabora una tabla para los casos en que se demora 34 minutos. Para lo anterior, selecciona en el gráfico de minutos los casos en que son 34; verás que en la tabla también se seleccionan. Da clic derecho sobre uno de los casos que esté seleccionado, en la primera columna de la tabla, y selecciona “Copy Cases”. En seguida, da clic en algún espacio en blanco de la pantalla y crea una nueva tabla. En “New” escribe “Minutos” (tal cual como lo tienes escrito en la tabla) y da Enter, finalmente, da clic derecho y selecciona “Paste Text”.

2f. Con los datos de esa nueva tabla, elabora un gráfico para pulsaciones. Ubica el valor de la media y compáralo con el que diste en el inciso 2e. ¿Qué tan cerca estuvo tu valor al de la media?

Apéndice G: Actividad 3. El profesor Gabo y la natación

Actividad 3. El profesor Gabo y la natación

Colegio: ________________________________

Octavo grado

Nombre: ____ Código: _________

El profesor Gabo, quien en su tiempo libre practica natación, ha registrado durante 23 sesiones consecutivas, el tiempo, en minutos, que tarda en nadar 1800 metros y su ritmo cardíaco, en pulsaciones por minuto, una vez ha completado los 1800 metros. Los tiempos y las pulsaciones se registran en la siguiente tabla.

	Minutos	Pulsaciones por minuto
1	34	152
2	36	124
3	35	140
4	34	152
5	34	155
6	36	128
7	36	128
8	35	144
9	35	148
10	36	130
11	35	136
12	36	124
13	35	148
14	35	144
15	35	140

16	34	156
17	35	136
18	34	148
19	34	148
20	36	132
21	36	124
22	35	132
23	36	130

- Abre un nuevo documento en Fathom y crea una colección llamada Practicando Natación. Elabora una tabla con los datos dados y, en seguida, dos gráficos, uno para minutos y otro para pulsaciones (o puedes usar el archivo de la Actividad 2).

1a. Selecciona los cinco valores más pequeños en "Minutos". ¿Qué ocurre en el gráfico de "Pulsaciones"?

1b. ¿Los valores que se ponen en rojo en el gráfico de "Pulsaciones" son mayores, menores o iguales, comparados con los que seleccionaste en "Minutos"?

1c. Repite lo anterior con otros valores. Por ejemplo, ahora selecciona los valores mayores en el gráfico "Minutos". ¿Qué puedes concluir de lo anterior? ¿Existe algún patrón o relación entre las dos variables?

1d. ¿Cómo varían las pulsaciones a medida que varían los minutos?

2. Grafica las variables Minutos y Pulsaciones que se muestran en la tabla en un gráfico y describe el gráfico; hazlo de tal manera como si le estuvieras contando a un amigo quien no puede ver el gráfico.

3. ¿Cómo se refleja lo que respondiste en 1c y 1d en este nuevo gráfico?

Apéndice H: Actividad 4. Nivel de ruido

Actividad 4. Nivel de ruido

Colegio: ________________________________

Octavo grado

Nombre: ____ Código: _________

Gabo, Édinson y Cristian hacen parte de un proyecto sobre el ruido. El proyecto consiste en visitar salones, contar el número de personas y medir el nivel de ruido en la clase con un medidor de sonido.

En la Tabla 1 se registran los datos para cuatro salones.

Salón	# de Personas	Nivel de ruido	
I	21	23	
F	29	75	
W	25		
Z	28		

Tabla 1

1. Como observas, no registraron el nivel de ruido para los salones W y Z.

1a. Predice el nivel de ruido para el salón W con base en los datos obtenidos en los salones I y F. Explica cómo encontraste o calculaste ese valor.

1b. Predice el nivel de ruido para el salón Z.

2. Abre el documento Nivel de ruido-1.ggb y selecciona la casilla “Datos Tabla 1”; aparecerá los puntos correspondientes a los salones I y F. Ubica también los puntos que predijiste para los salones W y Z. Para lo anterior, en la barra de entrada escribe (25, el valor que encontraste) y da Enter. Repite lo mismo para el salón Z.

2a. Observa el gráfico. ¿Consideras que deberías cambiar los valores que predijiste para los salones W y Z? Sí__ No__ ¿por qué?

2b. Si tu respuesta en el ítem anterior fue "Sí" mueve los puntos que corresponden a los salones W y Z hasta donde consideres que estarían bien ubicados. Explica por qué consideras que están bien ahora.

3. Repite los ítems 1-2 para la siguiente tabla.

Salón	# de Personas	Nivel de ruido
N	21	25
J	29	81
M	29	85
W	25	
Z	28	

Tabla 2

Nota: Para 2a, selecciona "Datos Tabla 2".

4. En el archivo Nivel de ruido-1.ggb selecciona las casillas "Con 21 personas" y "Con 29 personas". Observarás los datos que se muestran en la siguiente tabla.

Salón	# de personas	Nivel de ruido
A	21	11
E	21	19
F	29	75
I	21	23
J	29	81
M	29	85
N	21	25

Tabla 3

4a. Con base en los datos de la tabla, predice el nivel de ruido para un salón que tiene 25 personas. Explica cómo encontraste o calculaste ese valor.

4b. Con base en los datos de la tabla, predice el nivel de ruido para un salón que tiene 28 personas. Explica cómo encontraste o calculaste ese valor.

5. En el archivo Nivel de ruido-1.ggb selecciona las casillas “Con 21 personas”, “Con 29 personas” y “Con 25 personas”

5a. Con base en los datos, predice el nivel de ruido para un nuevo salón que tiene 25 personas. Explica cómo encontraste o calculaste ese valor.

5b. Con base en los datos, predice el nivel de ruido para un salón que tiene 23 personas. Explica cómo encontraste o calculaste ese valor.

5c. Con base en los datos, predice el nivel de ruido para un salón que tiene 27 personas. Explica cómo encontraste o calculaste ese valor.

6. Escribe una forma general, sin tener que hacer cálculos, para predecir el nivel de ruido para cualquier salón en donde hay entre 21 y 29 personas. Explica tu respuesta.

7. Abre el archivo Nivel de ruido-2.ggb, selecciona las casillas “Con 21 personas” y “Con 29 personas” y, de acuerdo a la forma que propones en el ítem 6, completa la siguiente tabla para el “Nivel de ruido. Explica tu respuesta.

Número de personas	Nivel de ruido
21	
22	
23	
24	

25	
26	
27	
28	
29	

8. En el archivo Nivel de ruido-2.ggb, quita la selección y ubica los puntos que dijiste en el ítem 7.

8a. Observa el gráfico. ¿Consideras que deberías cambiar los valores que predijiste en el ítem 7? Sí__ No__ ¿por qué?

8b. Si tu respuesta en el ítem anterior fue "Sí" mueve los puntos hasta donde consideres que estarían bien ubicados. Explica por qué consideras que están bien ahora.

9. Por qué lo que tú propones es válido, es decir, cuáles son las ventajas y desventajas de la forma general que propones en el ítem 6.

Apéndice I: Actividad 5

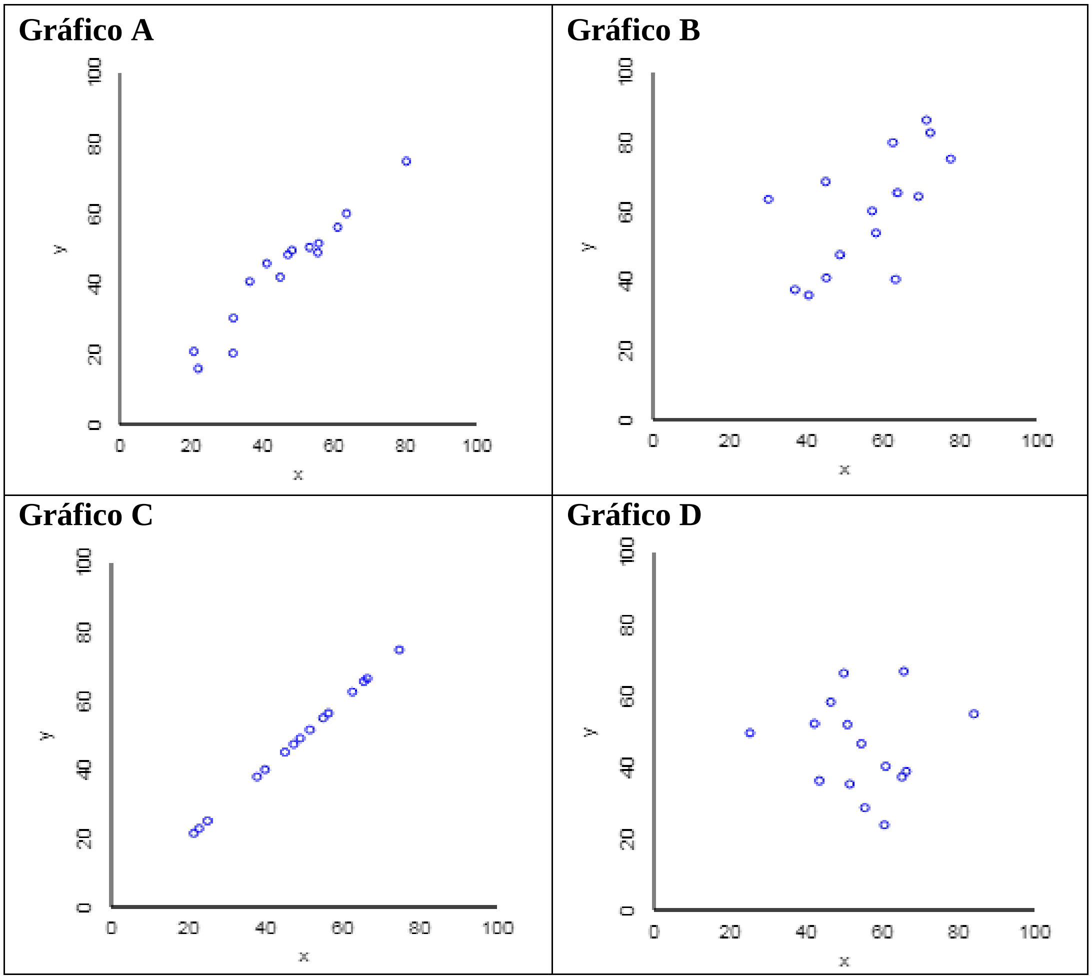
Actividad 5

Colegio: ________________________________

Octavo grado

Nombre: ____ Código: _________

1. Observa los siguientes diagramas de dispersión, que corresponden a las medidas de dos variables x e y, y responde.

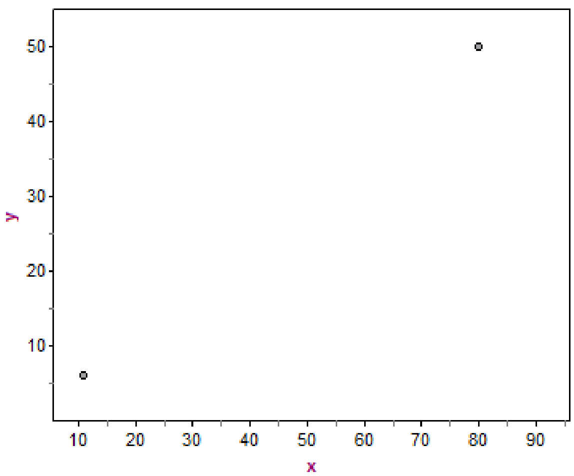
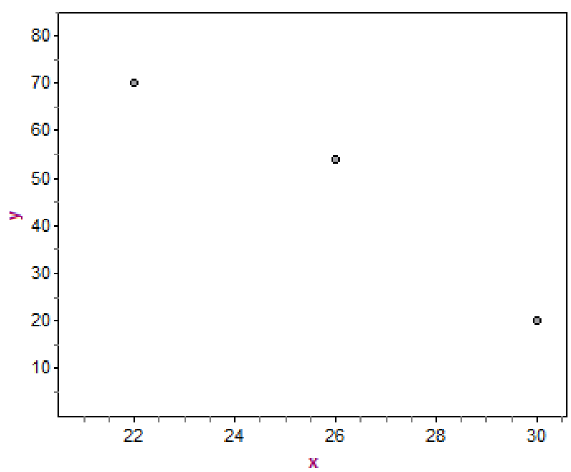


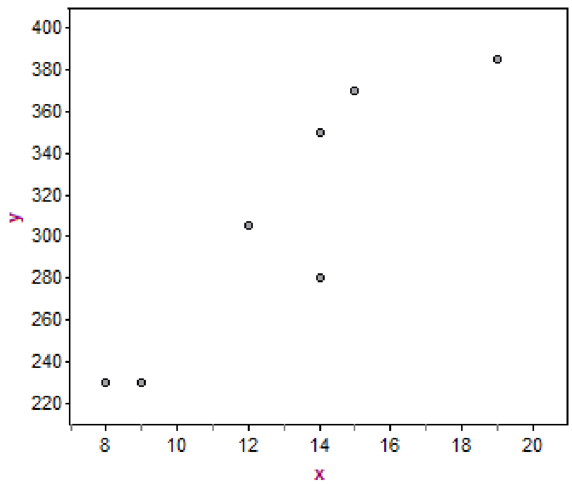
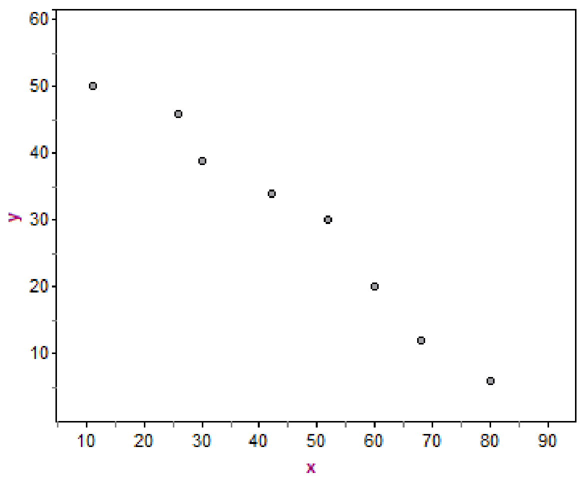
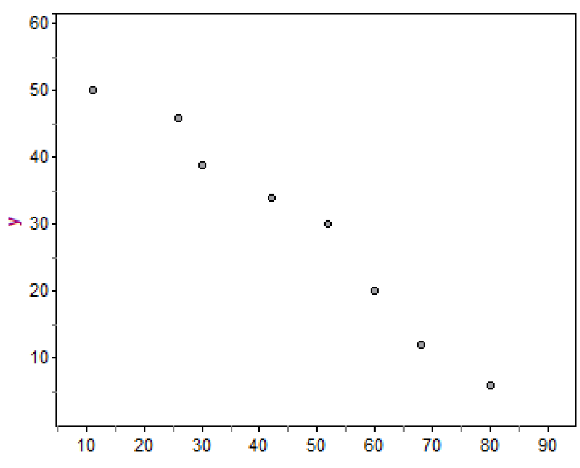
1a. Lee la siguiente afirmación y responde: *"En los gráficos A, B, C y D, se podría decir que si la variable x aumenta la variable y aumenta"* ¿En cuál de los cuatro gráficos podrías dar esa afirmación con mayor seguridad y en cuál con menor seguridad? Explica tu respuesta.

1b. ¿Cuál de los gráficos te permite hacer una mejor predicción del valor de y para un valor de x entre 0 y 100? ¿Por qué?

1c. En cada uno de los gráficos, A, B, C y D, predice un valor de y si x vale 72, aproximadamente. Explica tu respuesta.

2. Observa los siguientes diagramas de dispersión, que corresponden a las medidas de dos variables x e y, y responde.

<table>
<tr>
<td>Gráfico E

</td>
<td>2a. Predice un valor de y si x vale 50. Marca el punto sobre el gráfico y explica tu respuesta.</td>
</tr>
<tr>
<td>Gráfico F

</td>
<td>2b. Predice un valor de y si x vale 24, y si x vale 28. Marca los puntos sobre el gráfico y explica tu respuesta.</td>
</tr>
<tr>
<td>Gráfico G</td>
<td>2c. Predice un valor de y si x vale 11, si x vale 14 y si x vale 17. Marca los puntos sobre el gráfico y explica tu respuesta.</td>
</tr>
</table>

<table>
<tr><td>
</td><td></td></tr>
<tr><td>Gráfico H

</td><td>2d. Predice un valor de y si x vale 20, si x vale 35, si x vale 45 y si x vale 75. Marca los puntos sobre el gráfico y explica tu respuesta.</td></tr>
<tr><td>Gráfico H

</td><td>2e. Encuentra una forma general que te permita predecir un valor de y para cualquier valor que tome x entre 10 y 80. Muestra esa forma sobre el gráfico y explica tu respuesta.</td></tr>
</table>

3. En el caso de una variable vimos cómo un conjunto de datos podía reducirse a un solo valor, la media, el cual nos ayudaba a predecir. Para cada uno de los gráficos anteriores (A-H) traza la forma en que podría reducirse la nube de puntos teniendo en cuenta lo siguiente: i) Que refleje la relación que tienen las dos variables. ii) Que te permita hacer una buena predicción del valor de y para cualquier valor que tome x, según se muestra en cada gráfico. Explica tu respuesta.

Apéndice J: Actividad 6.1. Juego: Encontrando la mejor recta

Actividad 6.1

Juego: Encontrando la mejor recta

Colegio: ________________________________

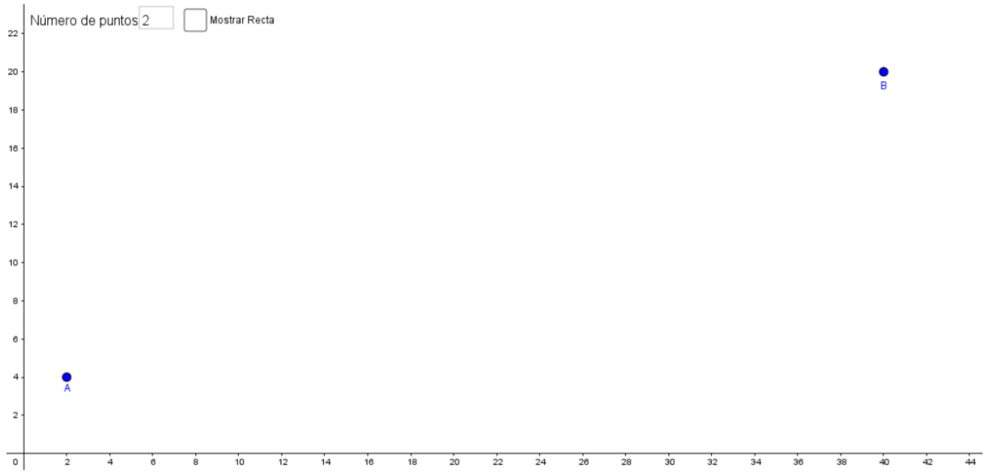
Octavo grado

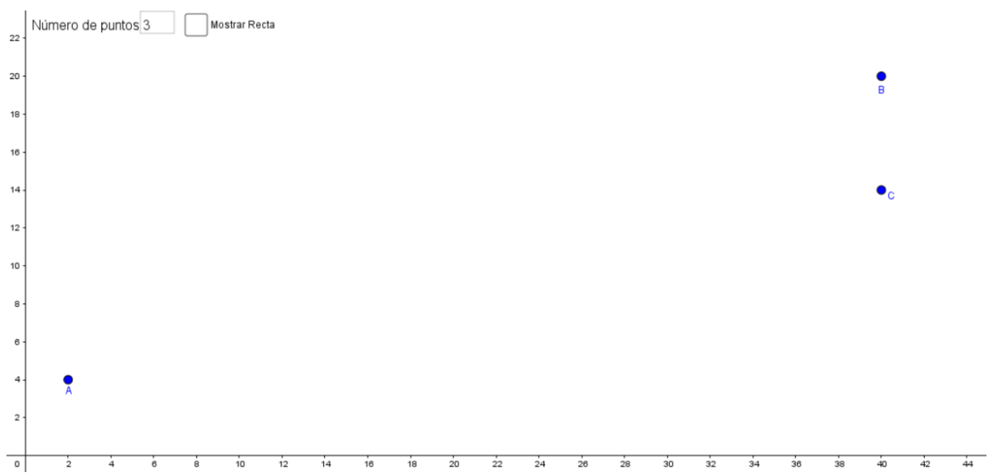
Nombre: ____ Código: _________

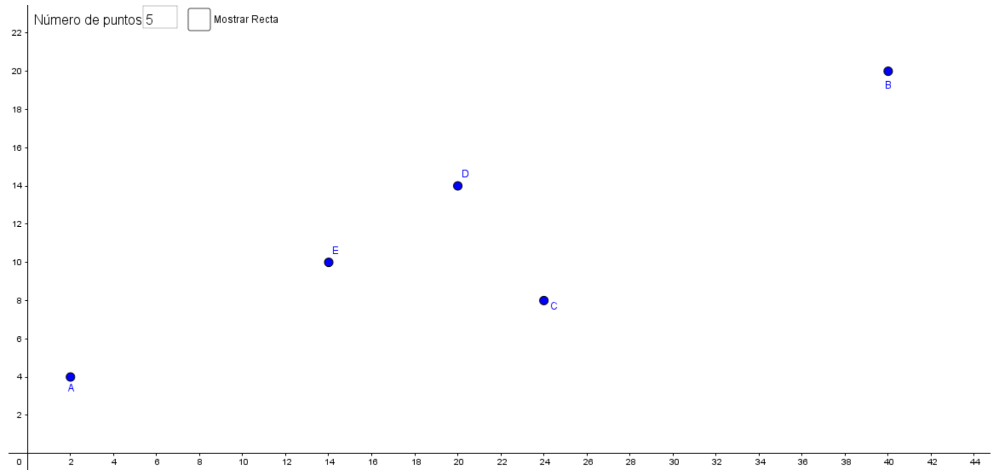
Abre el archivo Juego Encontrando la mejor recta.ggb. El juego consiste en lo siguiente:

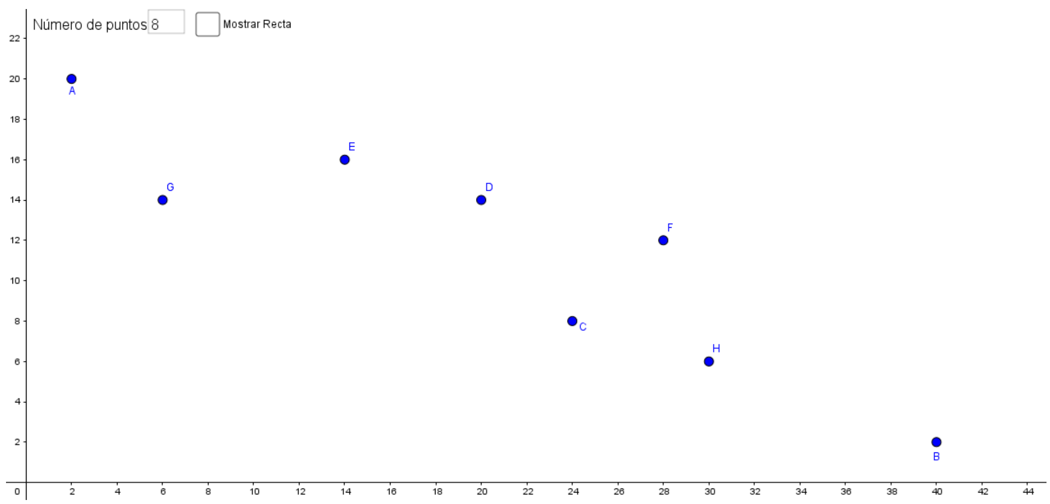
1. Espera a que tu profesora te indique qué cantidad poner en “Número de puntos”.
2. Enseguida, tu profesora te pedirá que traces una recta que, de alguna forma, se relacione con los puntos que se muestran. Para lo anterior, selecciona “Mostrar Recta” y muévela hasta donde consideres que está bien ubicada. El tiempo para trazar la recta es de 15 segundos, una vez ubicada la recta no debes moverla más.
3. Hecho lo anterior, tu profesora pasará por cada uno de los computadores revisando las rectas e indicará quién fue el feliz ganador, el cual será premiado.
4. Lo anterior se irá repitiendo aumentando el número de puntos en el diagrama de dispersión.
5. El objetivo final del juego es que encuentres una estrategia para encontrar la mejor recta, la cual deberás contársela a tus compañeros.

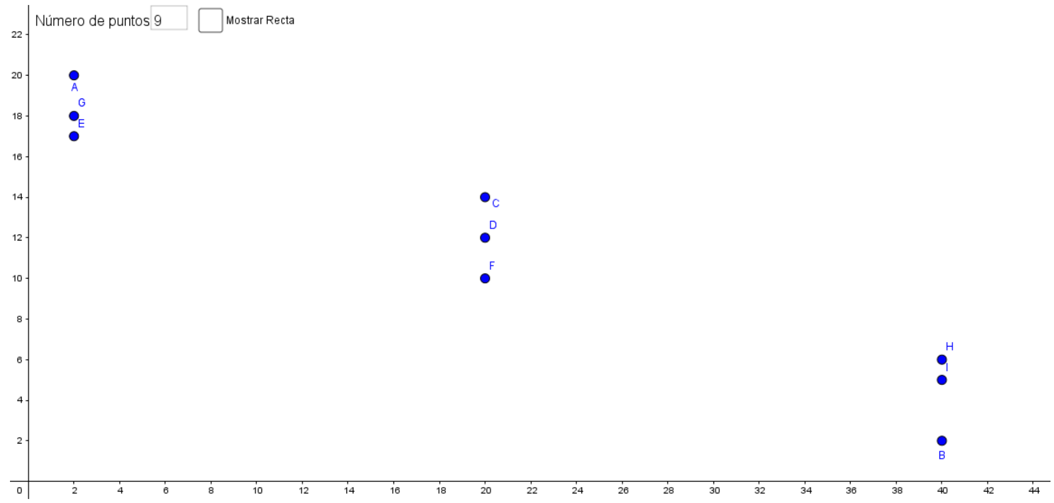
Traza sobre los siguientes diagramas la recta que encontraste en el juego.

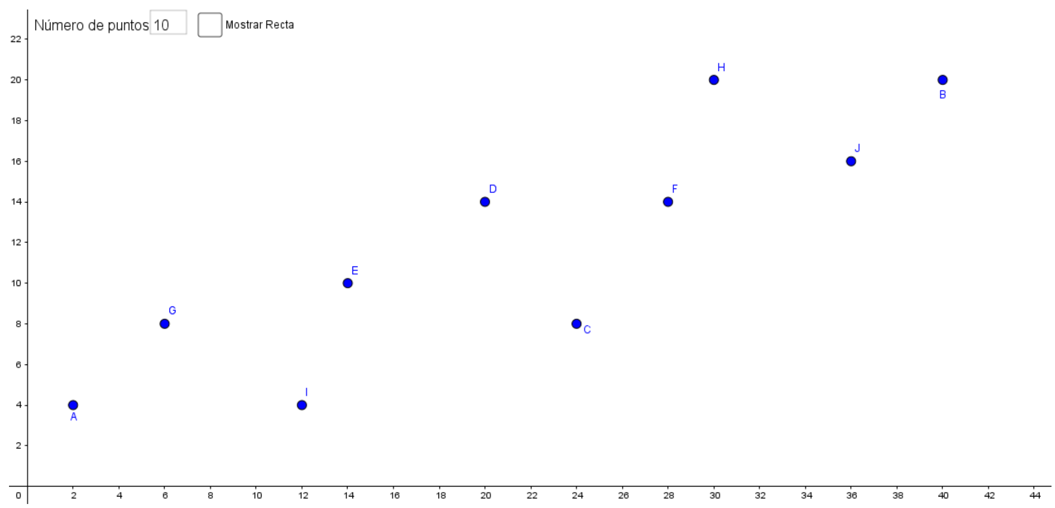


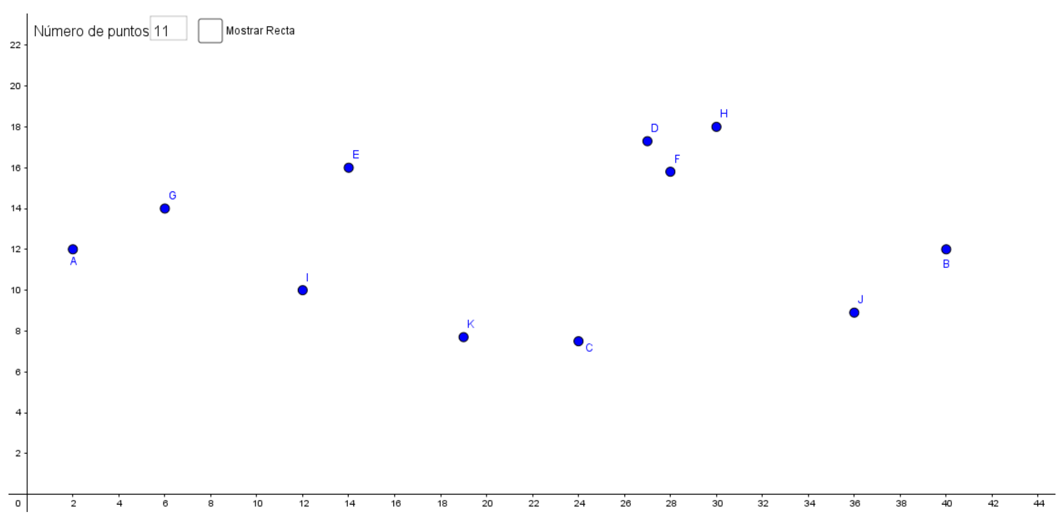



Número de puntos 5
Mostrar Recta
A
B
C
D
E


Número de puntos 8
Mostrar Recta
A
B
C
D
E
F
G
H


Número de puntos 9
Mostrar Recta
A
G
E
C
D
F
H
I
B


Número de puntos 10
Mostrar Recta
H
B
J
D
F
E
G
C
I
A


Número de puntos 11
Mostrar Recta
H
D
F
E
G
A
B
I
J
K
C

Responde:

1. ¿Cuál es tu estrategia para trazar la mejor recta?
2. ¿Qué te permite hacer esa recta?

Apéndice K: Actividad 6.2. Juego: Encontrando la mejor recta (Parte 2)

Actividad 6 (Parte 2). Juego: Encontrando la mejor recta

Colegio: ________________________________

Octavo grado

Nombre: ____ Código: _________

Abre el archivo Juego Encontrando la mejor recta.ggb.

1. En la actividad 6 (parte 1) jugaste a encontrar la mejor recta para diferentes nubes de puntos. Para cada uno de los casos vistos (Número de puntos= 2, 3, 5, 6, 8, 9, 10 y 11) vuelve a ubicar la recta que mejor consideres que se ajusta a la nube de puntos. En la parte inferior de la pantalla en “Ver” escribe “EK” y selecciona el recuadro blanco ¡Eureka!

Si tu recta corresponde a la mejor, en la parte inferior derecha de la pantalla aparecerá un recuadro amarillo con la palabra ¡Eureka! De lo contrario, mueve libremente la recta hasta encontrar la mejor. Reflexiona sobre cómo ubicar la mejor recta y descríbelo.

2. Considera el caso en que “Número de puntos” es 3 y encuentra la mejor recta, es decir, hasta que aparezca “¡Eureka!”. Hecho lo anterior, en la parte inferior de la pantalla en “Ver” escribe “MD” y selecciona la casilla “Mostrar diferencias”. Observa y reflexiona sobre cómo se obtienen esos valores; cuándo los valores son positivos y cuándo negativos. Descríbelo

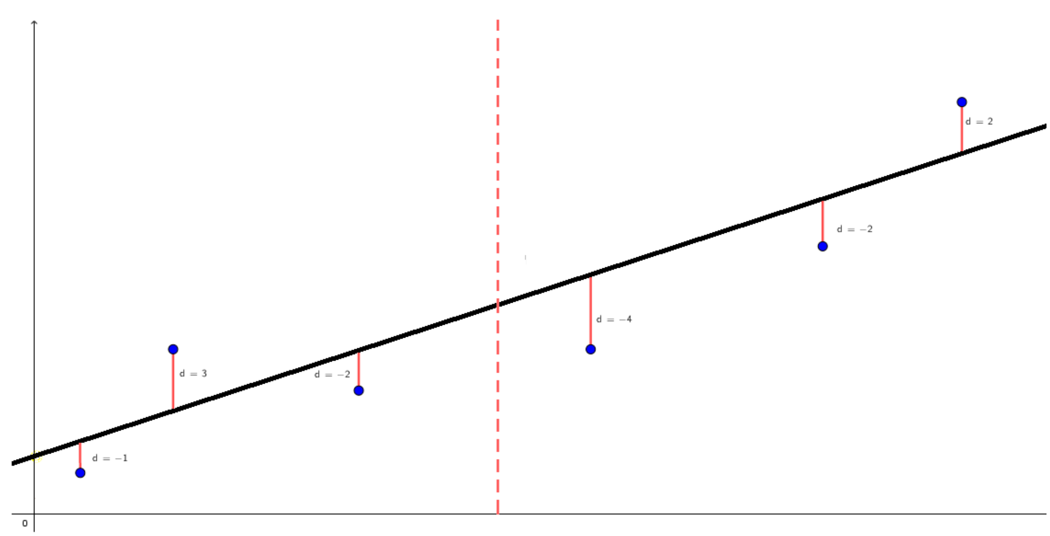
3. Oprime “VOLVER” y repite el anterior ítem para cada uno de los casos y contesta, en general, las siguientes preguntas.

3a. ¿Qué signo tienen las diferencias por encima de la recta?

3b. ¿Qué signo tienen las diferencias por debajo de la recta?

3c. En el caso de la mejor recta, ¿cuánto suma el total de las diferencias? (Para ver la suma de las diferencias, en la parte inferior de la pantalla en “Ver” escribe SD, la suma aparecerá en rojo debajo de “Mostrar diferencias”. Selecciona nuevamente “Mostrar diferencias”).

3d. Observa el siguiente gráfico en donde se muestra la mejor recta para siete puntos. Como te darás cuenta falta un punto. Ubica sobre la línea punteada el punto faltante y, además, escribe el valor de la diferencia “d” para ese punto. Explica cómo lo encontraste.



4. Descubriendo otra propiedad. Da clic sobre el botón “VOLVER” y en “Número de puntos” escribe 3.

- Ubica la mejor recta para esos puntos. Para lo anterior, en la parte inferior de la pantalla en “Ver” escribe “ON”. La mejor recta es la que aparece de color verde y verifica que no es movible.
- Selecciona “Mostrar recta”, y vamos a asociar cada diferencia “d” con un cuadrado. Para lo anterior, selecciona “Mostrar Cuadrados”.
- Mueve la recta de color negro y observa cómo cambia el valor de "SumaÁreasCuadrados” antes y después de la mejor recta (recta verde).

Contesta: ¿Qué se puede decir del valor de "SumaÁreasCuadrados" cuando las dos rectas (la negra y la verde) coinciden? Plantea una conclusión de acuerdo a lo que observas.

5. Oprime “VOLVER” y explora lo anterior para otra cantidad de “Número de puntos”. Describe la propiedad de la mejor recta que acabas de observar.

6. ¿Cómo te ayuda la mejor recta a predecir un valor de Y para un valor de X?

7. Oprime “VOLVER” y en “Número de puntos” pon 10. Muestra la mejor recta escribiendo “ON” en “Ver” y predice un valor de Y para cuando X vale 15.

8. En la parte inferior de la pantalla selecciona “Predicción”; observarás que sale un valor de Y para cuando X es 15. Compáralo con el valor que predijiste en el punto anterior y describe dónde se ubica ese nuevo punto (de color naranja) en el gráfico. Repite lo anterior para diferentes valores de X.

Contesta: En general, si tienes la mejor recta para una nube de puntos ¿cómo predices un valor de Y para un valor de X dado? Explica tu respuesta.

Apéndice L: Prueba final

Prueba final

Colegio: ________________________________

Octavo grado

Nombre: ____ Código: _________

Las siguientes preguntas son de opción múltiple con única respuesta. Escoge solo una de las opciones y JUSTIFICA tu elección.

1. La concentración de determinadas sustancias en la sangre influye en la salud de las personas. He aquí las mediciones del nivel de fosfatos en la sangre de un paciente que realizó seis visitas consecutivas a una clínica, expresadas en miligramos de fosfato por decilitro de sangre.

5.7	5.2	4.6	4.9	5.7	6.4

Predice el nivel de fosfatos en la sangre, del mismo paciente, para su próxima visita.

a. 6.4

b. 5.4

c. 4.6

d. 5.7

Contesta las preguntas 2 y 3 de acuerdo a la siguiente información:

Una banda de ladrones entre los 25 y los 30 años, de edad, ha sido capturada por la policía. El equipo de investigación sabe que la banda está integrada por cinco ladrones, incluyendo al jefe, y que el nombre de la banda hace honor al valor de la media de la edad actual de sus integrantes.

2. ¿Cuál de las siguientes opciones NO podría ser el nombre de la banda?

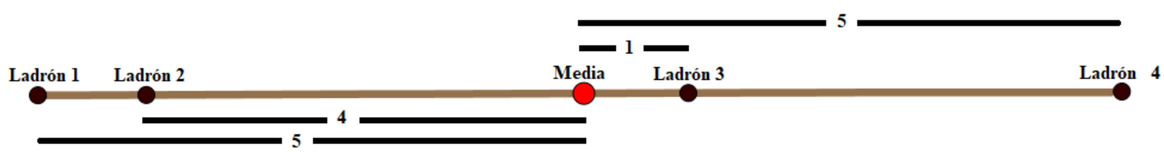
a. “Los veintiséis”

b. “Los veintiocho”

c. “Los treinta”

d. “Los treinta y dos”

3. Como broma, la policía ha ubicado a cada ladrón en una celda a cierta distancia de otra celda que tiene como número la media de la edad actual de sus integrantes, incluyendo al jefe. En la siguiente imagen se observa las celdas de los cuatro ladrones menos la del jefe.



¿Dónde está ubicada la celda del jefe?

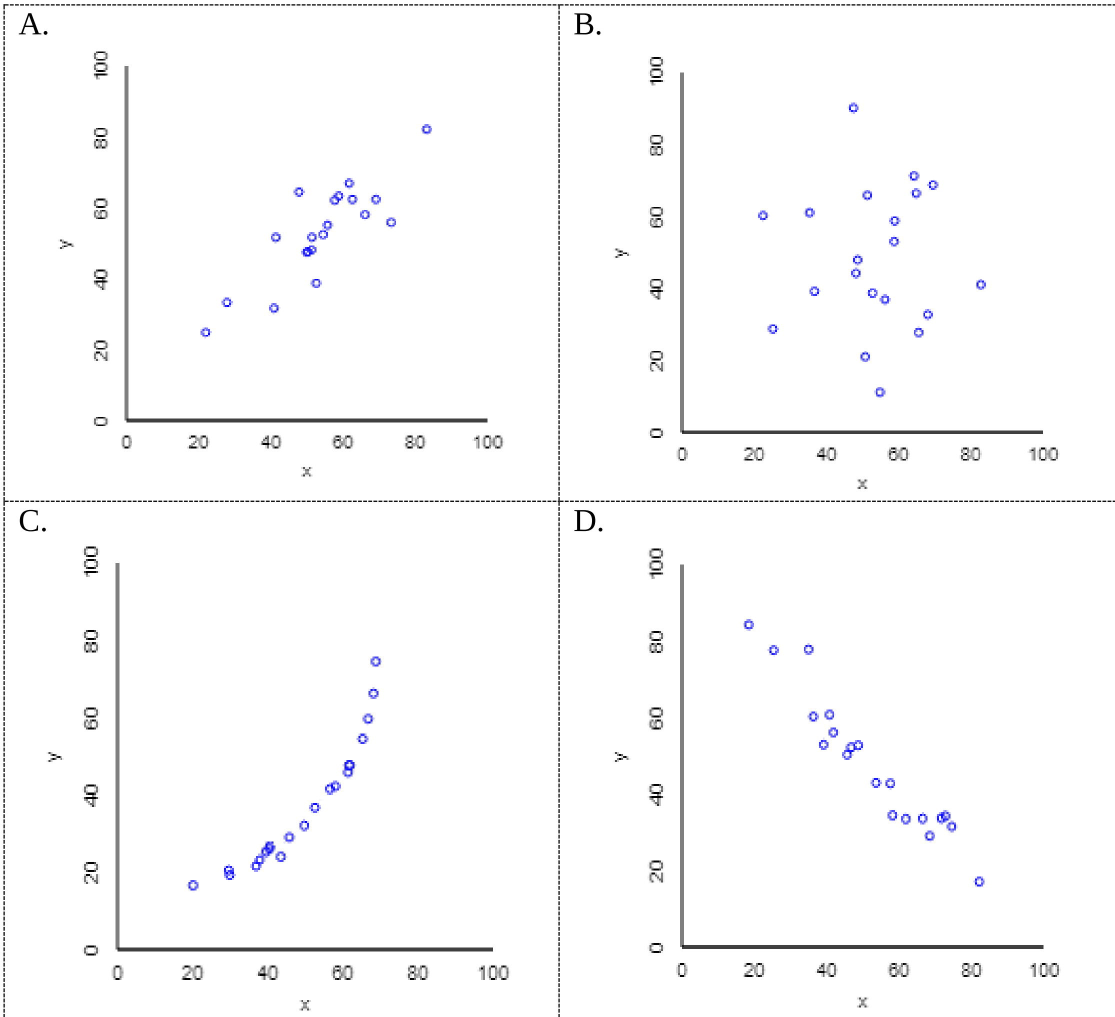
a. Cuatro unidades a la derecha de la media

b. Una unidad a la izquierda de la media

c. Tres unidades a la derecha de la media

d. Tres unidades a la izquierda de la media

Observa los siguientes cuatro diagramas de dispersión y responde



4. Lee la siguiente afirmación y responde: *"En los gráficos A, B, C y D, se puede decir que las variables x e y se relacionan de alguna manera"* Ordena los gráficos de mayor a menor, de acuerdo al nivel de seguridad con que se puede dar esa afirmación.

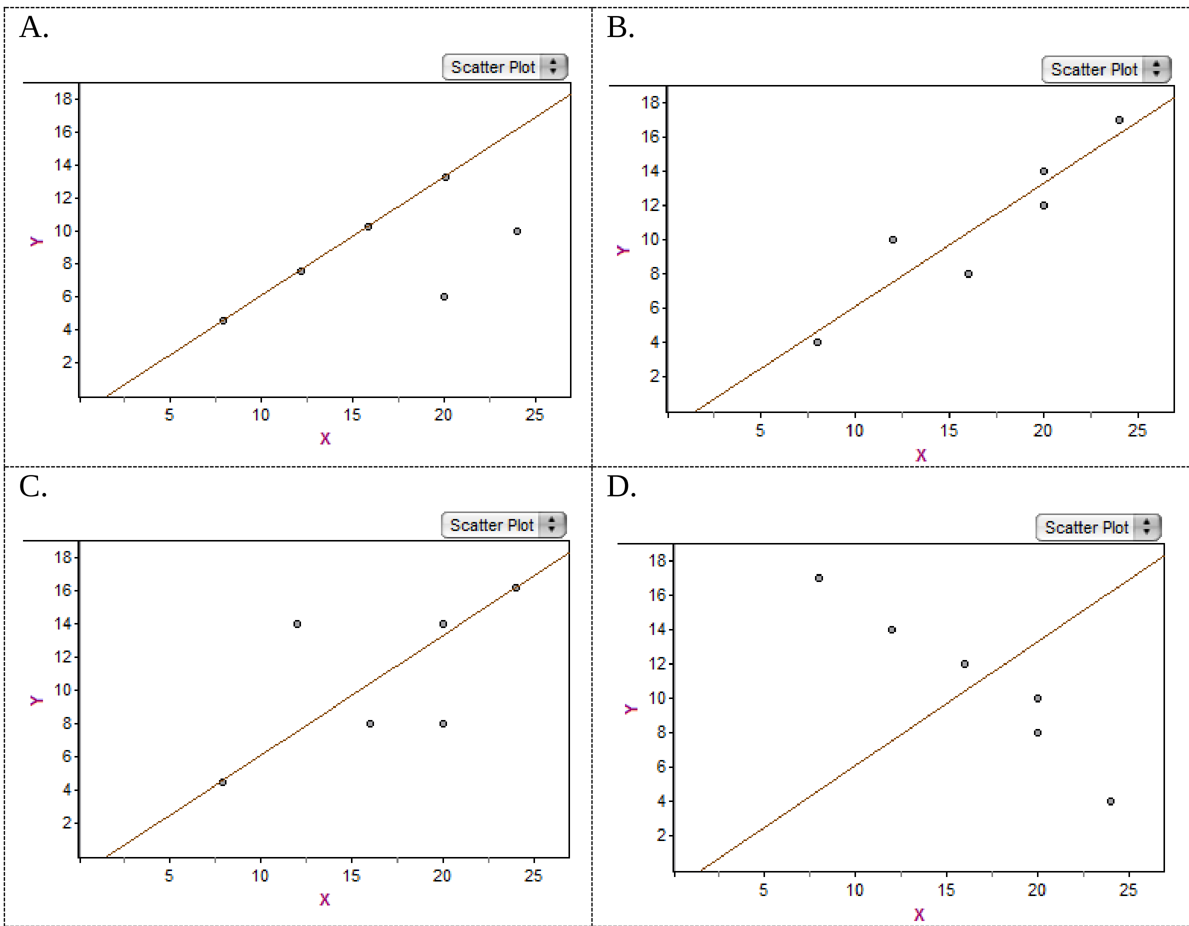
a. A-C-B-D

b. A-D-C-B

c. C-A-D-B

d. C-D-A-B

5. En los siguientes gráficos se muestra un conjunto de seis puntos y una recta que es LA MISMA para todos los gráficos, la cual refleja el comportamiento de los seis puntos. Ordena los gráficos de acuerdo al nivel de ajuste de la recta a los seis puntos, es decir, el primer gráfico que escojas será el gráfico en el cual la recta se ajusta mejor a la nube de puntos dada y el último que escojas, la que menos se ajusta.



a. B-D-C-A

b. B-C-D-A

c. A-C-B-D

d. B-C-A-D

Contesta las preguntas 6 y 7 de acuerdo a la siguiente información:

Se deja caer una pelota de goma desde diferentes alturas y se mide la altura que ésta alcanza en el primer rebote. Las mediciones se muestran en la siguiente tabla.

Altura de lanzamiento [cm]	Altura del primer rebote [cm]
12	5
26	12
32	22
43	23
50	26
61	37
66	48

6. Juan, Camilo, Pedro y Pablo desean lanzar nuevamente la pelota desde una altura de 45 cm y cada uno de ellos ha dado una predicción de la altura del primer rebote.

Juan dice que 27 cm, Camilo 24 cm, Pedro 5 cm y Pablo 48 cm. ¿Con cuál de los cuatro estás más de acuerdo?

a. Juan

b. Camilo

c. Pedro

d. Pablo

7. Pregunta abierta. Predice la altura del primer rebote si la misma pelota se deja caer desde las siguientes alturas: a) 20 cm; b) 35 cm; c) 55 cm y d) 60 cm.