

Conceptualización de la recta de mejor ajuste desde un enfoque informal por estudiantes de bachillerato

The conceptualization of the best fit line from an informal approach by high school students

Cindy Morgado,¹ Ernesto Sánchez²

Resumen. Se investiga cómo los estudiantes de bachillerato conceptualizan la recta de mejor ajuste utilizando tecnología. El marco teórico incluye la noción de agregado, un enfoque informal hacia la recta de ajuste y el uso de herramientas tecnológicas. La investigación involucró a 22 estudiantes, trabajando en parejas, en un taller compuesto por tres actividades enfocadas en la relación entre dos variables. Las tareas incluyeron la construcción y análisis de diagramas de dispersión, la interpolación y ajuste de una recta y, la manipulación de una recta para minimizar las distancias a los puntos. Los resultados se agrupan según similitudes en las respuestas, destacando cómo los estudiantes desarrollaron una comprensión agregada de los datos. Se identifican tres fases clave en la conceptualización de la recta de mejor ajuste: 1) uso del diagrama de dispersión, 2) interpolación de una recta en la nube de puntos y, 3) comprensión de residuos y la minimización de la distancia entre la recta y los puntos. Se revela un entendimiento de que la recta de mejor ajuste es aquella que minimiza su distancia al conjunto de puntos.

Fecha de recepción: 9 de noviembre de 2023. **Fecha de aceptación:** 19 de marzo de 2024.

¹ Universidad Autónoma de Bucaramanga (UNAB), cmorgado@unab.edu.co, <https://orcid.org/0000-0002-6085-7458>.

² Departamento de Matemática Educativa, Cinvestav, esanchez@cinvestav.mx, <https://orcid.org/0000-0002-8995-7962>.

Palabras clave: *Recta de mejor ajuste, agregado, enfoque informal, tecnología*

Abstract. This study investigates how high school students conceptualize the best fit line using technology. The theoretical framework includes the notion of aggregate, an informal approach to the fitting line, and the use of technological tools. The research involved twenty-two high school students working in pairs in a workshop comprising three activities focused on the relationship between two variables. Tasks included the construction and analysis of scatter plots, interpolation and adjustment of a line, and manipulation of a line to minimize distances to the points. Results are grouped according to similarities in responses, highlighting how students developed an aggregate understanding of the data. Three key phases in the conceptualization of the best fit line are identified: Use of the scatter plot, interpolation of a line in the point cloud, and understanding of residuals and minimizing the distance between the line and the points. An understanding is revealed that the best fit line is one that minimizes its distance to the set of points.

Keywords: *Best fit line, aggregate, informal approach, technology.*

INTRODUCCIÓN

En la sociedad actual, donde la tecnología de la información inunda a los ciudadanos con datos, la enseñanza y el aprendizaje de la estadística cobran una relevancia especial, pues puede permitir que los estudiantes se conviertan en ciudadanos capaces de procesarlos y analizarlos favoreciendo su comprensión del mundo y la toma de decisiones y, con ello, su desarrollo profesional, laboral y personal (Gravemeijer, 2013). Una de las áreas de la estadística más interesantes en este sentido es la asociación y modelación de las relaciones entre dos variables estadísticas (y el razonamiento covariacional asociado), tema considerado por Burril y Biehler (2011) como una de las ideas fundamentales de la estadística. Los conceptos de correlación y regresión son creaciones matemáticas de gran complejidad conceptual que históricamente emergieron asociados a profundas discusiones filosóficas y estadísticas acerca de la relación entre matemáticas y realidad (Desrosières, 2000). Una de las dificultades del tratamiento

escolar de este tópico es el problema de encontrar un enfoque de su enseñanza que atienda aspectos de ambos dominios.

El tema de la correlación y regresión es parte de los cursos universitarios de Estadística (Starnes, *et al.* 2010), pero también se prescriben subtemas como: la relación entre dos variables estadísticas, modelamiento de la relación, intensidad en ellas, diagrama de dispersión, estimaciones y predicciones, entre otros, en los programas de estudio de secundaria y bachillerato en varios países (NCTM, 2000, CCH-UNAM, s/f.). El enfoque tradicional de la enseñanza, en el nivel universitario, se carga hacia aspectos de la técnica matemática o algorítmicos, mientras que un enfoque más actual en los niveles básico y medio, pone más énfasis en la búsqueda de significado en el contexto del problema. Por un lado, un aprendizaje matemático, es decir, algorítmico y procedimental mediante las fórmulas del 'coeficiente de correlación' y 'recta de regresión' no es suficiente para comprender su potencial en la modelación de situaciones con fuertes dosis de incertidumbre. Por otro lado, esta comprensión no surge de manera natural mediante el tratamiento descriptivo de situaciones en contextos ricos y cercanos a los estudiantes en las que intervienen dos variables estadísticas. Se suele caracterizar el primer acercamiento como formal y el segundo intuitivo. En el caso de la inferencia, para establecer un puente entre ambos se ha sugerido explorar acercamientos informales (Zieffler *et al.* 2008; Makar y Rubin, 2009); ahora bien, los mismos argumentos también sugieren explorar acercamientos informales con relación a la regresión y correlación. En este artículo se informa sobre una experiencia en la que se introdujo desde un enfoque informal la recta de mejor ajuste a estudiantes de bachillerato.

Un concepto que se ha utilizado en el ámbito de la educación estadística y que conviene tener en cuenta en los enfoques informales para la enseñanza de conceptos estadísticos, incluyendo la recta de mejor ajuste, es el de agregado. En efecto, en las investigaciones sobre cómo los estudiantes piensan y analizan, en general, conjuntos de datos, frecuentemente se ha encontrado que la mayoría, sin importar el nivel escolar, tiende a poner atención en las características de datos particulares en detrimento de las propiedades globales del conjunto de datos; se juzga entonces que los estudiantes no ven a los datos como un agregado (Bakker y Gravemeijer, 2004; Konold y Higgins, 2002). También se ha detectado la tendencia de los estudiantes a poner atención a datos particulares en los estudios relacionados con trazar la recta de mejor ajuste, es decir, en el contexto de este problema tampoco suelen considerar a los datos bivariados como un agregado (Sorto

et al., 2017). Sin embargo, en nuestra opinión, por un lado, hace falta clarificar mejor el significado de agregado y, por otro, proporcionar sugerencias de cómo lograr que los estudiantes puedan llegar a ese nivel de pensamiento en relación con la recta de mejor ajuste. El presente estudio intenta hacer algunas precisiones sobre el concepto de agregado y sugiere que las actividades propuestas contribuyen a que los estudiantes lleguen a comprender a la recta de mejor ajuste como un agregado.

Una componente que apuntala los enfoques informales es el uso de tecnología (Rubin *et al.*, 2006) para desarrollar acercamientos a los conceptos y procesos estadísticos, permitiendo que los estudiantes tengan contacto con las ideas centrales sin necesidad de que enfrenten las complejas técnicas matemáticas que subyacen en tales conceptos y procesos. En este sentido, se han hecho propuestas de software didáctico para apoyar el aprendizaje de los estudiantes en estos temas, no obstante, aún hace falta documentar sobre cómo ciertos dispositivos les permite conceptualizar a la recta de mejor ajuste, teniendo en cuenta que parte de dicha conceptualización implica la noción de agregado.

En el presente artículo, se proponen actividades que incluyen el uso de la plataforma CODAP y una aplicación en GeoGebra con el fin de responder la siguiente pregunta de investigación: ¿Cómo conceptualizan la recta de mejor ajuste los estudiantes de bachillerato con ayuda de la tecnología?

ANTECEDENTES

Esta sección se divide en dos apartados: primero, se abordan las investigaciones sobre el aprendizaje y enseñanza de la correlación y regresión lineal por niveles escolares (universitario, bachillerato y secundaria). Segundo, los estudios que han señalado que una dificultad para entender el significado de conceptos estadísticos proviene de una inhabilidad de los estudiantes para ver los datos como un agregado.

INVESTIGACIONES SOBRE EL APRENDIZAJE Y ENSEÑANZA DE LA CORRELACIÓN Y REGRESIÓN LINEAL

A nivel universitario, los primeros estudios sobre la correlación y regresión fueron de naturaleza cognitiva, como el de Truran (1995) quien observa que estos frecuentemente interpretan el coeficiente de correlación con base en concepciones

erróneas, asumiendo la correlación como medida de la causa en lugar de asociación, o una medida que explica la variación. Sánchez-Cobo *et al.* (2000) muestran que sus estudiantes universitarios tienen capacidad para estimar el coeficiente de correlación, pero la precisión depende del tipo de representación a partir del cual lo estiman; tienen más éxito con el diagrama de dispersión que con las expresiones verbales o las tablas de datos. Sorto *et al.* (2011) solicitan a los estudiantes que dibujen la línea que mejor se ajuste a los datos representados en un diagrama de dispersión; emergen tres estrategias para trazar la recta: una que divide los puntos de datos por la mitad, una que pasa por puntos medios y, otra que une el primer y el último punto. Los autores concluyen que los estudiantes tienen dificultades para articular un criterio para el ajuste que pueda ser operacionalizado. Inzunza (2016), Inzunza y Ward (2015) exploran el razonamiento covariacional de estudiantes universitarios mediante el uso de la tecnología computacional y describen los resultados en una jerarquía SOLO, destacando que la aparente sencillez de un diagrama de dispersión queda entredicha en las respuestas de los estudiantes, dado que solo uno de ellos logró ubicarse en el nivel relacional del modelo SOLO.

En los niveles de secundaria y bachillerato, los estudios se han enfocado en dos aspectos: la representación gráfica de la covariación y el método para determinar la recta de mejor ajuste. Hourigan y Leavy (2021) hallaron que estudiantes de sexto grado a menudo tienen una comprensión local de la asociación estadística y enfrentan dificultades al crear gráficos de covariación. Moritz (2004) y Watson y Moritz (2007) identificaron cuatro niveles de razonamiento en estudiantes de grado 5, 6, 7, 8, grado 9 y 10 de secundaria y bachillerato, al interpretar gráficos de covariación: no estadístico (nivel 0), aspecto único (nivel 1), covariación inadecuada (nivel 2), y covariación apropiada (nivel 3).

Estepa y Batanero (1996), descubrieron concepciones variadas de correlación en estudiantes de último grado de bachillerato, incluyendo visiones deterministas, locales y causales. Por su parte, Casey (2014, 2015) y Nagle *et al.* (2017) investigaron las concepciones de los estudiantes sobre la línea de mejor ajuste y observaron estrategias similares a las de Sorto *et al.* (2011); también identificaron influencias de conceptos previos en estas tareas, como el de función lineal, este puede interferir en la capacidad de dar sentido a la tarea de ajustar una línea que no va necesariamente a través de todos los puntos, también el concepto de pendiente en la función lineal en matemáticas genera conflictos cognitivos con la configuración estadística de la recta de ajuste, debido a la variación aleatoria.

Entre los estudios de enseñanza recientes a nivel de bachillerato que incorporan tecnología, se encuentra el de Medina *et al.* (2019) quienes utilizaron

Fathom para enseñar sobre diagramas de dispersión y la línea de mejor ajuste, encontrando que los estudiantes poseen razonamientos como dividir los datos en dos partes, los que se comportan linealmente y los que no. Gil y Gibbs (2017) incorporaron tecnología para enseñar razonamiento covariacional en grandes datos, mostrando que la noción de tendencia surgió en los razonamientos de los estudiantes y les permitió describir y explicar el comportamiento de variables en sistemas del mundo, y Dierdorp *et al.* (2011) se centraron en visualización y tendencia de datos en contextos auténticos, lo cual hace que sean tareas realistas para que los estudiantes puedan experimentar autenticidad y se sintieran comprometidos para encontrar una solución al problema contextual.

En resumen, los estudios de aprendizaje dan cuenta de la dificultad que tienen los estudiantes para superar sus concepciones y creencias previas al evaluar e interpretar, el coeficiente de correlación, los diagramas de dispersión y la recta de regresión. Por otro lado, las investigaciones que han realizado algún tipo de diseño de enseñanza con el objetivo de promover el desarrollo del razonamiento covariacional, han utilizado tecnología digital. A nivel bachillerato se han ubicado tres estudios cubriendo cada artículo un tema: covariación en contextos de grandes datos (Gil y Gibbs, 2017), diagramas de dispersión y la línea de mejor ajuste (Medina *et al.*, 2019), visualización y tendencia en los datos (Dierdorp *et al.*, 2011).

INVESTIGACIONES QUE INVOLUCRAN LA NOCIÓN DE AGREGADO

Autores como Mokros y Russell (1995) evidenciaron que la mayoría de los estudiantes de tercer grado describieron sus datos enumerando observaciones sobre valores individuales para describir frecuencias, lo cual mostró que la mayoría de ellos no trataron un conjunto de datos como una entidad. Por el contrario, los estudiantes dan un salto conceptual cuando pasan de ver los datos como una amalgama de individuos únicos a verlos como un agregado con propiedades emergentes que no son evidentes en ninguno de los elementos individuales (Konold y Higgins, 2002).

Para Casey (2015) algunos criterios que establecen los estudiantes respecto a la recta de mejor ajuste dan evidencia de que poseen una vista agregada de los datos. Por ejemplo, colocar la línea de mejor ajuste de tal manera que quedara un número igual de puntos a cada lado de esta o aquellos que consideran que la línea debe estar lo más cercana posible a todos los puntos. En contraste, los estudiantes que no conciben la nube de puntos como un

agregado, establecen la línea a partir de los puntos que, según su criterio, corresponden a un modelo lineal dejando de lado aquellos que consideran no pertenecen a dicho modelo (Medina *et al.*, 2019).

MARCO CONCEPTUAL

TERMINOLOGÍA SOBRE COVARIACIÓN

Para Zieffler y Garfield (2009) el razonamiento covariacional es el razonamiento sobre la relación entre dos variables estadísticas, por lo que incluye el razonamiento sobre varios conceptos asociados como recta de mejor ajuste, datos bivariados, diagramas de dispersión y correlación. Este estudio se enfoca en la recta de mejor ajuste, la cual se relaciona con otros conceptos que mencionaremos enseguida. Los datos bivariados numéricos son parejas formadas por observaciones o medidas correspondientes a dos variables numéricas diferentes; las dos observaciones o medidas que constituyen un dato bivariado se obtienen de una misma unidad de análisis; por ejemplo, el peso y la altura de una misma persona se representan en un dato bivariado. Un diagrama de dispersión es la representación en un plano cartesiano de un conjunto de datos bivariados; al conjunto de puntos resultante se le llama nube de puntos. La covariación estadística es la manera en que varían conjuntamente dos variables estadísticas; según Moritz (2004) “se refiere a la correspondencia de variación de dos variables estadísticas que varían sobre escalas numéricas” (p. 228). Un diagrama de dispersión permite tener una imagen general de la covariación. La correlación se refiere al grado en que dos variables se relacionan linealmente y el coeficiente de correlación es una medida que ofrece información sobre la tendencia de la nube y sobre la cercanía de sus puntos entre sí y con una línea (fuerza de la correlación). Una recta de regresión es una recta que describe cómo cambia una variable de respuesta a medida que cambia una variable explicativa. La recta de regresión mínimo-cuadrática de un conjunto de datos bivariados es la recta que minimiza la distancia cuadrática de los puntos a la recta (Moore, 2000).

UN ENFOQUE INFORMAL DEL RAZONAMIENTO COVARIACIONAL

Es bien conocido y ampliamente documentado que los estudiantes tienen grandes dificultades para entender las ideas fundamentales de la estadística (Garfield y Alghren, 1988; Shaughnessy, 1992; Castro-Sotos *et al.*, 2007) incluyendo la regresión y la correlación (Engel y Sedlmeier, 2011). En el área del aprendizaje y la enseñanza de la inferencia estadística se ha propuesto la exploración de acercamientos informales como una estrategia para hacerla más accesible a los estudiantes e incluso introducirla desde niveles preuniversitarios (Zieffler, *et al.* 2008; Makar y Rubin, 2009, Tobías-Lara y Gómez-Blancarte, 2019). De manera análoga, la investigación explora acercamientos informales para temas de regresión y correlación (Casey, 2015; Biehler *et al.*, 2018), que implican también cierto tipo de inferencias.

Makar y Rubin (2009) propusieron un marco para la inferencia estadística informal (IEI) basado en tres categorías generales: datos, generalización e incertidumbre. Una afirmación o razonamiento pertenece a la IEI cuando utiliza los datos como evidencia, generaliza a una población más amplia y, además, se considera que dicha generalización mantiene cierto grado de incertidumbre. Tales categorías también pueden orientar una caracterización del razonamiento covariacional informal. Así, proponemos que un razonamiento covariacional informal consiste en juicios, razonamientos y/o procedimientos para descubrir y determinar la relación entre dos variables que no tienen el nivel de elaboración simbólica y matemática de los procedimientos formales, pero que tienen al menos tres propiedades análogas a la IEI, que son las siguientes: un juicio o razonamiento informal sobre la relación entre dos variables: 1) se basa en los datos bivariados disponibles (datos), 2) va más allá de los datos permitiendo estimar o predecir el valor de la variable resultado a partir de un valor de la variable explicativa (generalización), 3) se reconoce que tal estimación o predicción representa un valor probable pero no seguro (incertidumbre). Esta caracterización nos permite evaluar en qué medida las actividades propuestas fomentan el desarrollo de un razonamiento covariacional informal.

Hacemos la precisión de que hay una diferencia entre un acercamiento intuitivo y un acercamiento informal (Rossman, 2008); el primero, consiste en los procedimientos que espontánea y automáticamente ponen en juego los estudiantes para resolver un problema; en cambio, los procedimientos informales, al igual que los formales, son aprendidos mediante procesos de enseñanza dirigida y son aplicados con parsimonia, no sin creatividad, en la solución de problemas.

AGREGACIÓN, AGREGADO Y VISTA AGREGADA

De acuerdo con Stigler (2016), la agregación es el proceso de combinar múltiples observaciones o medidas individuales para formar un objeto estadístico que los asimila y engloba; el objeto así obtenido es un agregado, por ejemplo, los promedios, diagramas, gráficas, distribuciones o curvas de ajuste son ejemplos de agregados. Estos tienen propiedades del conjunto de datos que no están presentes en cada dato particular. La agregación busca ganar información más allá de lo que los valores individuales pueden proporcionar, creando un nuevo objeto que es más significativo que la mera colección de los datos individuales. Un proceso de agregación implica una simplificación y generalización de los datos. A pesar de la pérdida de detalles individuales, la agregación es fundamental para revelar tendencias y patrones generales en los datos. Un aspecto interesante es que con el resultado de una agregación se pueden hacer inferencias, es decir, ir más allá de los datos y obtener conclusiones probabilísticas. Un agregado permite hacer afirmaciones simples de variables complejas que pueden ser parte de un argumento. Algunos autores sugieren que los estudiantes deben desarrollar una “vista agregada” de un conjunto de datos, con lo que se quiere indicar que deben entender y analizar los datos como un todo colectivo, en lugar de considerar las observaciones individuales por separado (Bakker y Gravemeijer, 2004; Casey, 2015; Hancock *et al.*, 1992; Konold y Higgins, 2002). Pero con base en la afirmación de Hancock *et al.* (1992, p. 355) de que “para que un estudiante pueda pensar en el agregado, el agregado debe estar ‘construido’”, conviene remarcar que una “vista agregada” depende también de un agregado específico y no solo del conjunto de datos; por ejemplo, una visión agregada de un conjunto de datos que surge de la media es diferente de la que surge de su desviación estándar. La razón es que las propiedades globales de un conjunto de datos se revelan a través de los agregados.

LA TECNOLOGÍA Y EL RAZONAMIENTO ESTADÍSTICO COVARIACIONAL

La revolución tecnológica de las últimas décadas ha impactado en la investigación en educación estadística, surgen cada vez más aplicaciones de estadística y procesamiento de datos debido a los logros tecnológicos y la integración de estos avances en nuestra sociedad, lo que constituye un desafío serio para la educación de las futuras generaciones (Andre y Lavicza, 2019, p. 254); se han

producido y explorado diversos recursos para apoyar el razonamiento de los alumnos para analizar datos (Biehler, *et al.* 2013).

Biehler *et al.* (2018) proponen varias dimensiones para diseñar entornos de aprendizaje estadístico, 1) el enfoque en el desarrollo de ideas centrales más que en procedimientos, 2) el uso de tareas bien diseñadas, 3) conjuntos de datos reales y motivadores y 4) integración de herramientas tecnológicas. El presente estudio sigue estas recomendaciones poniendo atención en el papel del software para apoyar el desarrollo de la idea de que la recta de mejor ajuste es la que mejor aproxima a los datos. La facilidad para representar datos bivariados en diagramas de dispersión, y realizar con ellos transformaciones diversas de manera dinámica, ha creado condiciones muy favorables para el desarrollo del razonamiento covariacional informal en las aulas de estadística. En efecto, la exploración de la relación entre la forma general de una nube de puntos y su coeficiente de correlación lineal era prácticamente imposible sin la tecnología por el trabajo técnico que implicaba. Con esta, ahora se pueden analizar clases de nubes de puntos que comparten el mismo coeficiente de correlación, así como también investigar el efecto de variar un punto y analizar su efecto en el coeficiente de correlación lineal de todo el conjunto. La tecnología permite introducir conceptos importantes con un sentido claro para los estudiantes como el concepto de residuo, la distancia entre una nube de puntos y una recta, y la minimización de esta distancia. Debe tenerse en cuenta que el manejo formal de la noción de distancia a la nube de puntos es una función real de dos variables. En realidad, el dominio son rectas del plano cartesiano, pero cada recta se puede caracterizar con dos variables. La minimización formal de dicha distancia requiere de cálculo avanzado. Como hemos hecho en la presente investigación, con ayuda de recursos tecnológicos se pueden crear dispositivos que propicien un nivel de comprensión razonable por parte de los estudiantes de las ideas de residuos, distancia nube-recta y minimización sin que requieran las herramientas matemáticas que implica su acercamiento formal. Para la actividad de esta investigación se utilizó la plataforma CODAP y el software GeoGebra.

METODOLOGÍA

CONTEXTO DE LA INVESTIGACIÓN

En el presente estudio se exploran las posibilidades de una forma de introducir a los estudiantes desde un acercamiento informal al razonamiento sobre la relación entre dos variables estadísticas y la recta de mejor ajuste, por lo que se clasifica como una investigación cualitativa y exploratoria que se caracteriza por promover la actividad de resolución de problemas en parejas, con uso de tecnología e intervenciones del profesor.

El análisis de las relaciones entre dos variables con base en un conjunto de datos bivariados es un problema genuino para los estudiantes porque no cuentan con un procedimiento previamente aprendido para hacerlo. Elegimos datos bivariados en un contexto de nutrición, el cual puede ser de interés debido a que se relaciona con la salud y las costumbres alimenticias. La utilización de datos de variables de la vida real del estudiante puede ser un factor importante tanto cognitivo como motivacional para propiciar el aprendizaje de la estadística (Garfield y Ben-Zvi, 2008; Neumann *et al.*, 2013).

Algunos investigadores sostienen que la actividad de resolución de problemas se vuelve más potente si se realiza de forma colaborativa en pequeños grupos y no solo individualmente. El trabajo en parejas permite que uno de ellos exprese sus ideas y las contraste con las del otro, que negocien la interpretación del problema y las estrategias de solución, así como que encuentren una forma conjunta de expresarla por escrito (Yackel *et al.*, 1991; Webb y Mastergeorge, 2003). En consecuencia, esperamos que las respuestas dadas por las parejas representen patrones de razonamiento que reflejen la micro cultura del aula del salón de clase (Cobb *et al.*, 2003).

La disponibilidad de recursos tecnológicos ha hecho posible diseñar situaciones de aprendizaje que sin tecnología son impensables (Biehler, *et al.* 2013). Por ejemplo, en CODAP se puede arrastrar directamente un punto de un diagrama de dispersión y ver en tiempo real cómo cambian la correlación, la recta de mejor ajuste y su ecuación, o bien, en una aplicación de GeoGebra es posible representar un conjunto de datos bivariados, construir una recta movable, así como definir y calcular una distancia de la recta al conjunto de puntos. La minimización de esta distancia aproxima a la recta de mejor ajuste. Estos dispositivos pueden ayudar a los estudiantes a entender la lógica subyacente a la construcción de la recta de mejor ajuste. CODAP (<https://codap.concord.org/>) es un software educativo para el análisis de datos, diseñado como una plataforma para desarrolladores y como

una aplicación para estudiantes de 6º a 14º grado. Como complemento se utilizó GeoGebra, ya que en este se pueden elaborar programas que combinan álgebra y geometría, lo que permitió elaborar un programa para que los estudiantes se acercaran a la recta de mejor ajuste. Ambos softwares son gratuitos y dinámicos y fueron diseñados para la enseñanza.

En el diseño de actividades, se asegura que las parejas de estudiantes dejen rastros de sus razonamientos acerca de la correlación y recta de mejor ajuste mediante sus respuestas a las preguntas que se les formulan. Las intervenciones del profesor se centraron en explicar el funcionamiento del software y en aclarar la intención y sentido de la actividad, pero se redujeron al mínimo en cuanto a la información del contenido.

PARTICIPANTES

La investigación contó con la participación de 22 estudiantes de secundaria, cuyas edades iban desde los 16 hasta los 18 años. La autora de este artículo dirigió las sesiones con el apoyo de la profesora del grupo. Durante los semestres anteriores, los estudiantes han estudiado temas como aritmética, álgebra, geometría y geometría analítica, pero no han abordado la probabilidad y la estadística hasta el quinto semestre. Sin embargo, en el ciclo escolar anterior, se han familiarizado con la lectura y elaboración de gráficas, medidas de tendencia central e introducción a la probabilidad. Aunque los estudiantes no tenían experiencia previa con conceptos como correlación y regresión lineal, ya estaban familiarizados con el plano cartesiano, las funciones lineales y las funciones cuadráticas, lo cual les sirvió como base para comprender de manera informal el tema de la regresión y la correlación.

INSTRUMENTOS Y RECOLECCIÓN DE DATOS

En la investigación de la cual se deriva el presente artículo (Morgado, 2023) se desarrollaron tres actividades, pero solo se presentan dos de ellas, porque la intermedia fue una actividad reportada en otros estudios y los resultados que obtuvimos son similares (Morgado *et al.*, 2022; Casey, 2015; Medina *et al.*, 2022). La fuente de datos son las hojas de trabajo que llenaron los estudiantes trabajando en parejas al mismo tiempo, se formaron 11 parejas con los 22 estudiantes. El diseño e implementación se describe a continuación. La tabla 1 resume la duración y estructura de las dos actividades.

Tabla 1. Duración y estructura de las actividades

Actividad	Duración	Estructura
Actividad 1 (ver tabla 2)	2 horas	1ª parte: Describir la relación entre variables sin uso de tecnología. 2ª parte: Crear el diagrama de dispersión con CODAP y describir la relación entre las variables.
Actividad 2	2 horas	1ª parte: Describir la relación entre las variables sin uso de tecnología, y luego crear el diagrama con CODAP y juzgar la relación a partir de este (ver tabla 3). 2ª parte: Establecer el criterio para la línea de mejor ajuste (ver tabla 4).

Para la primera actividad se utilizó la información nutricional de grasa y calorías de 6 productos de McDonalds y 6 productos Burger King. Estos datos fueron descargados de las páginas oficiales de las marcas (<https://www.mcdonalds.com>, <https://www.bk.com/menu/search-by-nutrition>).

Para la segunda actividad se tomó como referencia una situación del libro de texto *The practices of Statistics* de Starnes *et al.* (2010) y se diseñó un applet digital de GeoGebra junto con instrucciones y preguntas.

PRESENTACIÓN Y ANÁLISIS DE LA ACTIVIDAD 1

En la tabla 2 se observa la situación y preguntas que se administraron a los estudiantes para explorar su razonamiento acerca de la relación entre variables. Los rasgos que conviene destacar es que la situación está en un contexto familiar, las preguntas son abiertas y no presuponen conocimientos específicos, además se promueve el uso de tecnología. Los estudiantes se enfrentan a dos aspectos que están en tensión, por un lado, el contexto propicia la recuperación de creencias sobre las relaciones entre las variables involucradas, por otro, el diagrama de dispersión informa sobre tales relaciones, pero no necesariamente confirma las creencias de los estudiantes. Para los estudiantes, es difícil ignorar sus creencias previas y observar los datos en términos puramente cuantitativos (Moritz, 2004; Confrey, 1990). Se busca conocer los juicios que establecen sobre la relación entre dos variables en un contexto definido.

La herramienta CODAP permite la creación y visualización de un conjunto de datos bivariados, con la situación de la actividad 2 se busca que los estudiantes establezcan que una relación entre dos variables se considera lineal cuando los datos en la nube de puntos tienden a agruparse formando una línea recta en el plano cartesiano.

Tabla 2. Situación y problema para explorar el razonamiento sobre correlación

Actividad 1.

Situación: Quizás, oyes o creas que las grasas y las calorías son dañinas para tu salud. Es verdad, que algunas personas ingieren más grasas y más calorías de las que necesitan, pero sin ellas no tendrían suficiente energía para realizar sus actividades y crecer. ¿Qué has escuchado sobre las grasas y las calorías que contiene un alimento? Muchos alimentos contienen grasas y calorías, mucha gente cree que la grasa influye notablemente en la cantidad de calorías totales que contiene ¿Tú qué opinas? ¿Cómo podrías saber si la cantidad de grasa influye en la cantidad de calorías? A continuación, los datos de la tabla A muestran la cantidad de grasa en gramos y el número de calorías en algunos productos de McDonald's y Burger King.

Tabla A. Grasa y Calorías de algunos productos de McDonalds y Burger King

	Producto	Grasa(g)	Calorías (Kcal)
McDonald's	Big Mac	27	540
	Cheeseburger	12	300
	Quarter Pounder With Cheese	26	510
	Mcdouble	18	390
	Double Quarter Pounder with Cheese	45	780
	Bacon Ranch Grilled Chicken Salad	13	300
Burger King	Hamburger	10	240
	Cheeseburger	16	320
	Bacon Y Cheese Whopper	51	790
	Double Whopper	58	900
	Double Pretzel Bacon King	60	920
	Chicken Club Salad-Grilled	41	610

1ª parte

a) Observa los datos de la tabla A ¿Cuál crees que es la relación entre los gramos de grasa y las calorías? Explica tu respuesta.

2ª parte

Utilizando tecnología digital. Abrir la plataforma digital CODAP y crear el diagrama de dispersión de los datos bivariados (grasa, calorías) presentados en la tabla A. Sigue las indicaciones del profesor.

b) ¿Qué características se podrían tener en cuenta al interpretar un diagrama de dispersión? (esta pregunta resultó irrelevante, pues los estudiantes se concentraron en responder la siguiente pregunta e ignoraron esta).

c) ¿Qué puedes decir de la relación entre los gramos de grasa y el número de calorías?

Tabla 3. Situación y preguntas para explorar el razonamiento sobre recta de mejor ajuste (1ª parte)

Actividad 2.												
Situación. Un estudio investigaba por qué algunas personas no aumentan de peso incluso cuando comen en exceso. Tal vez por estar inquieto, por ejemplo, al hacer pequeños movimientos, especialmente con las manos y pies, por nerviosismo o impaciencia y otras “actividades no relacionadas con el ejercicio (ANE)” se gasta energía y quemas calorías. Los investigadores se preguntaron lo siguiente: ¿los cambios en la inquietud y otras actividades no relacionadas con el ejercicio explican el aumento o disminución de peso en las personas que comen en exceso? Para contestar la pregunta, deliberadamente sobrealimentaron a 12 adultos jóvenes sanos durante 8 semanas. Midieron el aumento de grasa (en kilogramos) y el cambio en el uso de energía (en calorías) a partir de otra actividad no relacionada con el ejercicio ya sea inquietud, vida diaria y cosas por el estilo (tabla B).												
Tabla B. Medidas de Cambio en ANE y Aumento de Grasa en los 12 Adultos Jóvenes												
Cambio ANE (cal)	-94	-57	-29	135	143	245	355	486	535	571	620	690
Aumento de Grasa (kg)	4.2	3	3.7	2.7	3.2	2.4	1.3	1.6	2.2	1	2.3	1.1
a) Discute con tu compañero y describe la relación entre el cambio ANE y el aumento de grasa. b) Realiza el diagrama de dispersión en CODAP y describe el comportamiento de la nube de puntos.												

PRESENTACIÓN Y ANÁLISIS DE LA ACTIVIDAD 2

En las tablas 3 y 4 se presentan la situación y preguntas que se administraron a los estudiantes para explorar su razonamiento sobre la recta de mejor ajuste de un conjunto de datos bivariados. Al igual que la actividad 1, se formula dentro de un contexto de nutrición. La primera parte (tabla 3) se aborda inicialmente la discusión acerca de la relación entre las variables cuantitativas sin uso de tecnología, seguido se realiza el diagrama de dispersión en CODAP para describir el comportamiento de la nube de puntos.

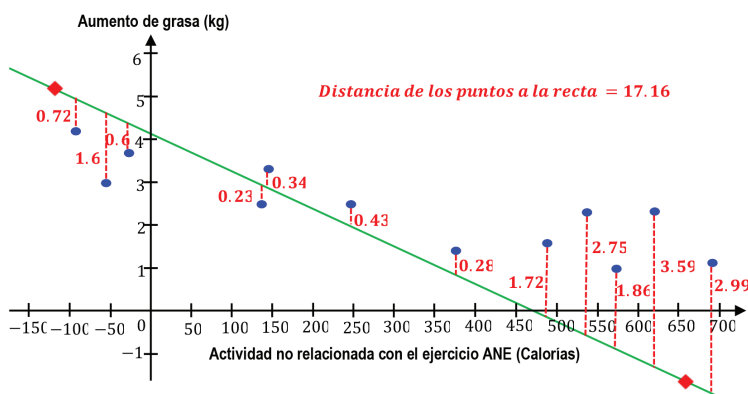
Con ayuda de un applet de GeoGebra (2ª parte-tabla 4) se busca que los estudiantes entiendan las ideas subyacentes en la técnica de regresión lineal, específicamente, el proceso de optimización de distancias. Se ha decidido no emplear el concepto de distancia cuadrática con el fin de transparentar la idea de construir una medida de distancia de un conjunto de datos a una recta; consideramos que el tomar distancias cuadráticas puede oscurecer la idea simple de una distancia

global de la nube de puntos a una recta es la suma de las distancias verticales de la recta a cada punto. La vista gráfica de GeoGebra muestra una recta que es movable utilizando el ratón (o mouse). La recta está ligada a la función que calcula automáticamente la distancia de una nube de puntos a la recta y la exhibe en una ventana de la pantalla. Para describir dicha función se define el error o residual de un punto como la distancia vertical entre el punto y la recta, siendo la distancia de la nube a la recta la suma de los residuales.

Tabla 4. Situación y preguntas para explorar el razonamiento sobre recta de mejor ajuste (2ª parte)

Abre el archivo de GeoGebra (el estudiante visualiza la figura A). Lee cuidadosamente lo siguiente: El archivo muestra el diagrama de dispersión de los datos (puntos azules) de la tabla B junto con una línea verde la cual puedes mover con los símbolos en rojo $\blacklozenge$. El segmento punteado de color rojo muestra la distancia que va desde el punto azul a la recta verde, su valor corresponde a la diferencia entre la ordenada del punto azul (y) y la ordenada del punto que pertenece a la recta verde ($\hat{y}$), a este valor lo llamaremos residual, y la "Distancia de los puntos a la recta" es la suma de todos los residuales.

Figura A. Captura de pantalla de la vista gráfica de la tarea en GeoGebra



Mueve la línea verde, explora lo que ocurre con el valor de cada residual y responde:

- ¿Dónde colocarías una recta para que corresponda a la recta que mejor se ajuste a los puntos? Explica el criterio que utilizaste.
- ¿Qué ocurre con el valor de cada residual si mueves la recta cerca o lejos de la nube de puntos?
- ¿La manera como ubicaste la recta movable es igual a la recta arrojada por GeoGebra? ¿Sí? ¿No? ¿En qué se diferencian? ¿Cuál crees es el criterio que utiliza GeoGebra para determinar la recta de mejor ajuste?

PROCEDIMIENTO DE ANÁLISIS

Las respuestas de los estudiantes fueron digitalizadas y se compararon entre sí para determinar patrones de respuesta, es decir, los rasgos comunes a los procedimientos o razonamientos similares presentes en dos o más respuestas a una pregunta. La codificación de tales patrones de respuesta permite reducir y esquematizar las respuestas para facilitar la descripción y análisis. La intención es desarrollar e identificar características en común, en los datos que den evidencia de las conceptualizaciones de los estudiantes al juzgar la relación entre dos variables estadísticas y al establecer el criterio para la recta de mejor ajuste.

RESULTADOS

En esta sección se expone cómo los estudiantes respondieron a las preguntas de las actividades, destacando patrones identificados al comparar sus respuestas. Estos patrones, denominados con términos específicos como *global-cualitativo*, *linealidad* o *distancia*, reflejan aspectos comunes en dos o más respuestas. Este enfoque brinda una visión general de las respuestas del grupo y esboza sus procesos de razonamiento, complementándose con comentarios analíticos para su conceptualización.

RESULTADOS DE LA ACTIVIDAD 1

Frente a la pregunta sobre la relación entre gramos de grasa y calorías (1ª parte de la actividad 1), emergen dos enfoques: uno global-cualitativo y otro particular-cuantitativo. El primero evalúa la tendencia general de los datos y su descripción cualitativa; mientras que el segundo busca expresar los cambios específicos en cada dato a través de secuencias numéricas o funciones, aunque sin éxito. Ocho parejas ofrecieron una valoración cualitativa, observando una tendencia positiva entre grasa y calorías, ejemplificada en la figura 1, la respuesta es evidencia de la tendencia positiva al describir que ambas variables aumentan. Tres parejas intentaron un análisis cuantitativo utilizando proporciones para establecer una relación matemática, como se muestra en la figura 2.

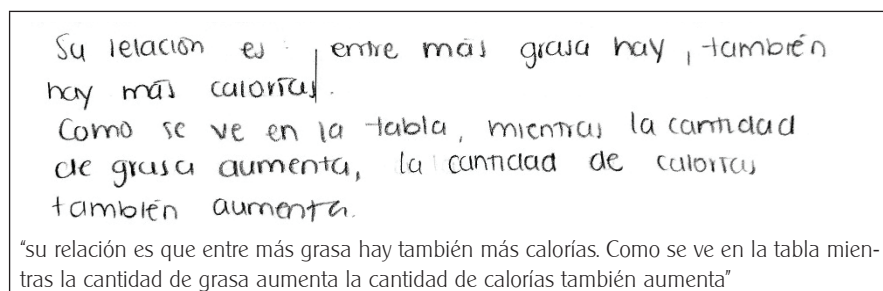


Figura 1. Respuesta que involucra juzgar la relación entre las dos variables en la actividad 1.

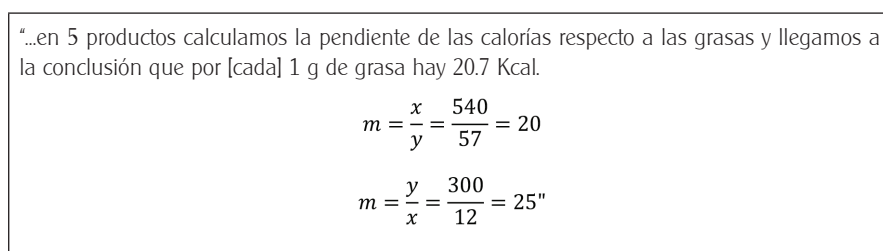


Figura 2. Respuesta que involucra juzgar la relación entre las dos variables.

En la segunda parte de la actividad 1, donde se utilizó la plataforma CODAP para representar los datos en un diagrama de dispersión, las respuestas se clasifican en dos categorías: casi-linealidad e irregularidad, con un caso atípico. Cinco parejas identificaron una casi-linealidad en la distribución de los datos, como se ejemplifica en la figura 3. Las otras cinco parejas notaron irregularidad en la distribución de puntos, aunque reconocieron una tendencia positiva general, ilustrada en la figura 4. Una pareja no establece ninguna descripción.

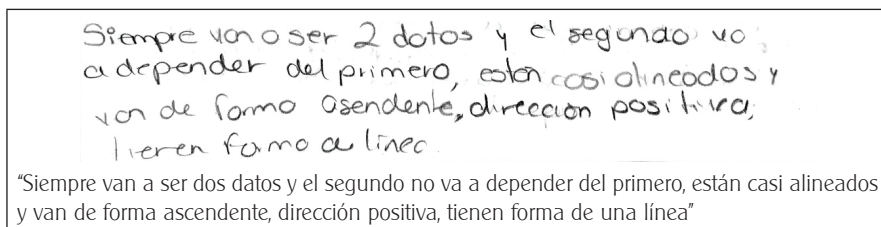


Figura 3. Ejemplo de respuesta al analizar la relación en el diagrama de dispersión en la actividad 1.

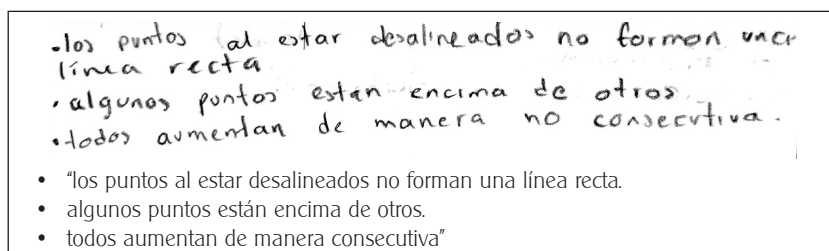


Figura 4. Ejemplo de respuesta al analizar la relación en el diagrama de dispersión en la actividad 1.

No es conveniente juzgar que las ocho parejas que utilizan un enfoque global tienen una visión agregada de la correlación simplemente porque observan la propiedad general y no se enfocan en puntos particulares. Lo que hace falta para desarrollarla es concebir el agregado correspondiente y la manera en que los puntos se combinan para construirlo. Tampoco es conveniente juzgar del todo inapropiada la estrategia de las tres parejas que buscan una regla para describir los puntos, ya que están buscando cómo combinarlos para obtener un objeto que describa la relación. De hecho, es lo que se debe hacer, solo que las herramientas con las que cuentan los estudiantes (secuencias aritméticas y proporcionalidad) no son las apropiadas. Juzgamos entonces que no son concepciones falsas sino embrionarias.

El efecto de utilizar el diagrama de dispersión, facilitado con el software, es significativo, porque con la visualización de los puntos emergen respuestas que muestran que perciben una tendencia (la categoría: casi-linealidad), pero también la irregularidad. Se siguen presentando en esta actividad respuestas que se enfocan en lo global, pero menciona, lo particular (figura 3) o que se enfocan

en lo particular, pero en la forma de irregularidad y no en la búsqueda de una regla. En estas respuestas se presenta una prefiguración del agregado, pues hay mención indirecta, pero inclusiva, de una recta: “están casi alineados” (figura 3) o la mencionan directamente, aunque de manera excluyente: “no forman una línea recta” (figura 4). El software y la posibilidad de trazar el diagrama de dispersión ha permitido que los estudiantes tengan una percepción más clara del conjunto de datos considerando a la vez aspectos globales y puntuales, es decir, una tendencia y la irregularidad.

Aunque los estudiantes prefiguran el agregado, todavía no saben exactamente cómo combinar los datos para construirlo. Antes de pasar a la siguiente actividad, conviene recordar que los estudiantes llevaron a cabo una actividad intermedia cuyos resultados decidimos no exponer aquí con detalle por coincidir en gran medida con otros informes (Sorto *et al.*, 2011; Casey, 2015). Consistió en presentar un diagrama de dispersión y pedirles que trazaran una recta que se ajustara a los datos. Las estrategias que desarrollaron los estudiantes fueron similares a las reportadas por los autores mencionados. Las rectas que proponen los estudiantes son aún inapropiadas o imprecisas: la recta que pasa por más puntos del conjunto, la recta que pasa por el primero y el último, la recta que deja la mitad de los puntos de un lado y la otra por el otro, lo que a simple vista parece la más cercana a la recta. La función de esta tarea fue motivar el proceso de agregación, es decir, formular el problema de buscar una manera de combinar los datos para dar lugar al agregado (la recta de mejor ajuste). Varios estudiantes se acercan a la noción de que la recta de mejor ajuste es la más cercana a todos los puntos, pero bajo una idea intuitiva de “cercanía” que es necesario operacionalizar. La actividad 2, descrita previamente, tiene ese propósito.

En resumen, el diagrama de dispersión y la tarea de que ajusten una recta constituye una estrategia que permite a los estudiantes pensar en combinaciones de los puntos para construir una recta hipotética. Proponen combinaciones para construir el agregado, pero les hace falta un método más preciso que concrete las ideas centrales de distancia del conjunto de datos a una recta y, sobre todo, la minimización de las distancias. Para que esta opción sea viable para ellos, hemos propuesto la actividad con el programa de GeoGebra que se ha descrito previamente.

RESULTADOS DE LA ACTIVIDAD 2 (SIN PROGRAMA DE GEOGEBRA)

En la Actividad 2, previo al uso del applet en GeoGebra, se repitieron preguntas similares a la Actividad 1. En la 1ª parte, pregunta (a), se solicitó describir las relaciones entre ANE y grasa. Cinco parejas indicaron una tendencia global, por ejemplo: “menos actividad implica más grasa”. Tres parejas reconocieron una dependencia entre ANE y grasa, sin establecer una regla definida, por ejemplo: “Si las calorías varían con el ANE, la grasa fluctúa” (ver figura 5). Las tres restantes basaron sus respuestas en un modelo contextual, por ejemplo: “El aumento en el consumo de calorías incrementa la grasa” (ver figura 6).

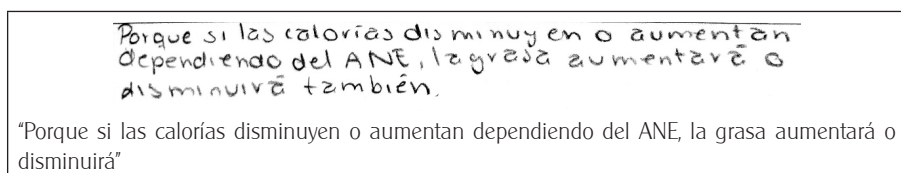


Figura 5. Ejemplo de respuesta al juzgar la relación entre las variables en la actividad 2.

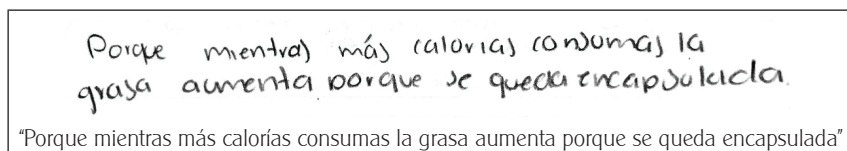


Figura 6. Ejemplo de respuesta al juzgar la relación entre las variables en la actividad 2.

La pregunta (b) requería un diagrama de dispersión en CODAP y la descripción de la relación observada. Ocho parejas identificaron una relación inversa, con la idea de que “A mayor ANE, menor aumento de grasa” (ver figura 7). Tres parejas se enfocaron en el contexto, sin datos concretos, por ejemplo: “El ANE no relacionado con ejercicio no incrementa la grasa si hay movimiento” (ver figura 8).

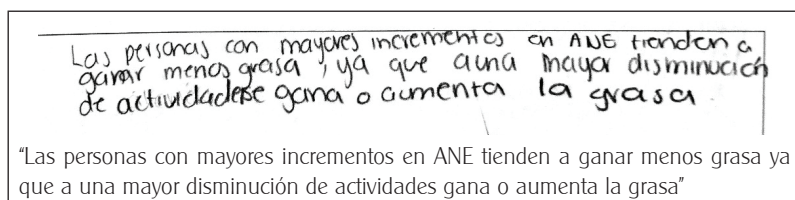


Figura 7. Ejemplo de respuesta al analizar la relación en el diagrama de dispersión en la actividad 2.

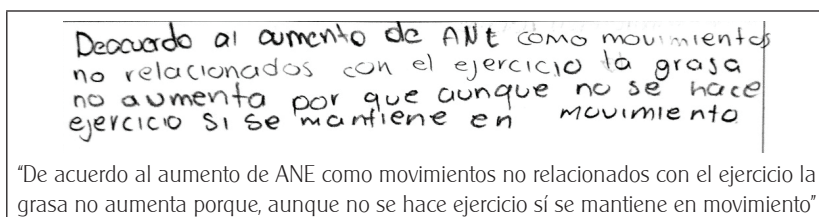


Figura 8. Ejemplo de respuesta al analizar la relación en el diagrama de dispersión en la actividad 2.

La diferencia con la actividad 1, aparte del contexto, es que ahora la tendencia es negativa, es decir, mientras una variable aumenta la otra disminuye. La variable ANE es menos familiar para los estudiantes, aunque la asimilan a la idea de movimiento no relacionado con el ejercicio. Probablemente por la falta de familiaridad de la variable ANE, los estudiantes no notaron que lo que se mide es el cambio de ANE y no la actividad misma; de ahí que se tengan valores negativos, que significan pasar de una actividad intensa a una actividad más leve. Probablemente por esto los resultados no son mejores a los obtenidos en la actividad 1, debido a que se presentaron respuestas basadas en creencias previas sobre el contexto sin atender a la información que proporcionan los datos. De cualquier manera, el objetivo de esta primera parte es introducir a los estudiantes al contexto de la situación del problema, y proponerles utilizar el programa en GeoGebra para centrar su atención en la idea de distancia de un conjunto de puntos a una recta y la de minimizar la distancia.

RESULTADOS DE LA ACTIVIDAD 2 (CON PROGRAMA DE GEOGEBRA)

En esta parte de la actividad los estudiantes utilizan el programa GeoGebra. A través de él, se representan los datos del problema y una recta movable en el plano cartesiano, en la que se pueden observar los residuos correspondientes y en un pequeño cuadro la suma de todos los residuos. En seguida reproducimos cada pregunta y las clases de respuesta que se obtuvieron.

La pregunta a) "¿Dónde colocarías una recta para que corresponda a la recta que mejor se ajuste a los puntos? Explica el criterio que utilizaste", es la misma que se formuló en la actividad anterior, pero ahora los estudiantes tienen la posibilidad de mover una recta preconstruida en la pantalla del software y de observar tanto la distancia vertical de cada punto a la recta como la suma de tales distancias.

Clasificamos las respuestas de los estudiantes en tres categorías: Puntos en la recta (1 de 11 parejas), Cercanía (4 parejas), Minimización de residuales (4 parejas), Sin respuesta (2 parejas). Pertenecen a la primera categoría la respuesta de una pareja que asocia la recta de mejor ajuste a la minimización de distancia de la recta a los puntos, pero considera que "es mínima cuando pasa por casi todos los puntos". Las respuestas de la categoría "Cercanía" sugieren que la recta de mejor ajuste debe estar cerca de los puntos, pero la forma de medir la cercanía es simplemente como la perciben, es decir, de manera subjetiva (figura 9) y sin precisarlas con ayuda del software, pero perciben que les da información de los residuos.

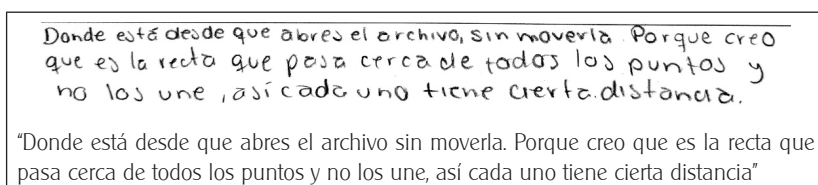


Figura 9. Ejemplo de respuesta del criterio para la recta de mejor ajuste en la actividad 2.

Las cuatro respuestas de la tercera categoría corresponden a quienes manipulan los residuos tratando de minimizarlos, e infieren que así se minimiza la distancia de los puntos a la recta (figura 10). La estrategia se basa ya en los conceptos pertinentes de distancia y minimización, pero a diferencia de la noción vaga de "cercanía", esta se asocia a las medidas precisas que proporciona el software.

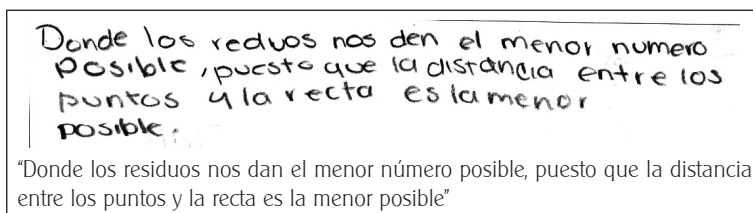


Figura 10. Ejemplo de respuesta del criterio para la recta de mejor ajuste en la actividad 2.

Un detalle que revela la dificultad para construir el agregado es que, al parecer, los estudiantes buscan minimizar los residuos individualmente e infieren que así se obtiene la distancia mínima de los puntos a la recta, pero no les parece natural minimizar la distancia global e inferir que al hacerlo se consigue una distribución de residuos de minimización óptima. Es decir, piensan que el control de los residuos particulares les permite controlar la distancia global, pero no a la inversa. En las respuestas a la pregunta que siguió, los estudiantes evidenciaron esta sutil dificultad.

La pregunta b) "¿Qué ocurre con el valor de cada residual si mueves la recta cerca o lejos de la nube de puntos?", busca que los estudiantes reflexionen sobre la relación entre la distancia de la recta a los puntos y el valor de cada residuo. La pregunta sugiere una intuición sobre la proximidad (cercanía) de la línea a la nube de puntos, relacionada con la suma de residuos. Si esta suma se minimiza, se infiere que se ha logrado una distribución de residuos con suma mínima.

En esta pregunta encontramos tres grupos de respuesta: descriptivo (5 de 11 parejas), enfoque en lo particular (4 de 11 parejas) y la recta centrada (2 de 11 parejas). El primer grupo solo constata que al cambiar los residuos cambia la distancia, pero no se refieren a la minimización; por ejemplo: "Que la distancia que hay entre cada uno de los puntos y la recta sumadas es el resultado de los residuos". El segundo grupo resalta que al minimizar los residuos se minimiza la distancia de la nube a la recta; por ejemplo: "Entre menores sean los residuos, menor será la distancia de los puntos a la recta". Finalmente, en dos casos, sugieren que de una posición centrada de la recta se obtiene la minimización de los residuos, por ejemplo: "Cuando la recta está mejor centrada, el nivel de residuos es menor, es decir, pasa centrándose en medio de los puntos".

Una idea implícita en la pregunta y que pasa desapercibida (o es eludida) por los estudiantes, es que el proceso de ajuste de cada residuo no del conjunto total de residuos, es decir, no es posible minimizar un residuo, luego minimizar

un segundo residuo y así sucesivamente, para alcanzar la recta de mejor ajuste; por el contrario, se debe producir una minimización coordinada que solo es posible controlando la suma de los residuos, sin poner atención en lo que pasa con los residuos particulares. Percibir este detalle contribuiría a consolidar una vista agregada de la recta de mejor ajuste. De cualquier manera, consideramos que el diseño del applet y su uso permite que los estudiantes asocien los conceptos de distancia de la nube de puntos a la recta y su minimización a la determinación de la recta de mejor ajuste.

Para la pregunta c) "¿La manera como ubicaste la recta movable es igual a la recta arrojada por GeoGebra? ¿Sí? ¿No? ¿En qué se diferencian? ¿Cuál crees que es el criterio que utiliza GeoGebra para determinar la recta de mejor ajuste?" Cuatro respuestas se centran en la aproximación, por ejemplo: "Son casi iguales a la amarilla está un poco arriba de la recta roja, pero están pegaditas y casi en el mismo lugar", y siete se centran en la diferencia, por ejemplo: "No se encuentran en el mismo sitio". Otra distinción en las respuestas es que nueve se enfocan en la posición relativa de las rectas, por ejemplo: "La recta de regresión aumenta la separación en comparación de la amarilla", mientras que en dos se comparan la suma de los residuos de cada recta, por ejemplo: "Nuestra recta tenía una distancia de 10 y la de regresión tenía una de 5". No hubo ninguna respuesta para la pregunta sobre el criterio que utiliza GeoGebra para determinar la recta.

En estas repuestas se manifiesta una de las propiedades de los agregados, a saber, que se pueden comparar entre sí. Es decir, la recta que construyeron los estudiantes es comparable con otro agregado (la recta de regresión). Los criterios como: comparar aproximación de las rectas (que no coinciden, la distancia de cada una de ellas es mayor o menor a los puntos, y las rectas son paralelas), se basan en propiedades generales de las rectas y no se refieren a puntos particulares. En este sentido, la actividad ha propiciado que los estudiantes hayan avanzado en tener una vista agregada de la recta de mejor ajuste.

CONCLUSIONES

La pregunta que motivó esta investigación fue ¿cómo conceptualizan la recta de mejor ajuste los estudiantes de bachillerato con ayuda de la tecnología? Con base en los resultados mostrados, podemos mencionar que el proceso de conceptualización es gradual a través de tres momentos propiciados por diferentes tareas apoyadas con recursos tecnológicos. Conviene notar que las tareas se

organizan para que los estudiantes comprendan la recta de mejor ajuste como un agregado. A continuación, se describen estos momentos.

Momento 1: uso de diagrama de dispersión. A diferencia de la tarea en la que solo se les presentaba la tabla de datos bivariados, la construcción de un diagrama de dispersión propició que los estudiantes percibieran tanto una tendencia general de los datos como la irregularidad de las diferencias entre los puntos. En la primera tarea, en la que no hicieron un diagrama de dispersión, algunos estudiantes se fijaban solo en la tendencia y otros solo buscaban relaciones punto a punto, pero con el diagrama todos mencionaron un aspecto global y otro individual. Medina *et al.* (2019) observó que, ante la tarea de analizar la relación entre dos variables a partir de un conjunto de datos bivariados dados en tablas, los estudiantes no suelen trazar el diagrama de dispersión de manera espontánea y tienden a buscar relaciones aritméticas entre ello, de ahí que la contribución de un software como CODAP, es significativa al propiciar una perspectiva geométrica.

Momento 2: interpolación de una recta en la nube de puntos. La tarea de interpolar una recta enfrenta a los estudiantes a utilizar los datos para fijar una recta representativa. Generan entonces estrategias diferentes, como trazar la recta por los puntos extremos a la nube, que pase por en medio de los puntos o ajustarla a simple vista para que esté más cerca de todos los puntos. Estas estrategias han sido documentadas consistentemente en diferentes estudios (Sorto *et al.* 2011, Casey, 2015, Medina, *et al.* 2019). Son aún estrategias intuitivas que no alcanzan el nivel de informales puesto que son solo visuales, sin traducirse en operaciones con los valores de las entradas de los datos. No obstante, con tales estrategias los estudiantes buscan una manera de combinar los datos para dar lugar a la recta; esto los prepara para la siguiente fase, en la que se les propone un programa que opera con los valores específicos de los datos. Por esto, interpretamos el inicio de la construcción de la recta de mejor ajuste como un proceso de agregación.

Momento 3: residuos y distancia de una recta a la nube de puntos. Entre las estrategias propuestas por los estudiantes en la tarea previa se destaca la que utiliza el criterio de la cercanía de los puntos a la recta; esta se operacionaliza con el uso del programa de GeoGebra que describimos previamente. En efecto, el programa permite introducir el concepto de residuo junto con la manera de medir los residuos con relación a cualquier recta dada; entonces se tiene una manera operativa de definir la distancia de la nube de puntos a una recta y decidir que la que mejor ajusta a los puntos minimiza dicha distancia.

En cada una de estas fases los estudiantes progresan en adquirir elementos para alcanzar una visión agregada del conjunto de datos bivariados; en particular, llegan a entender cómo se combinan los datos para dar lugar a la recta de mejor ajuste. En efecto, captan la idea central que está en la base de la construcción de esta: “minimizar la ‘distancia’ total entre la línea y los puntos observados”. Los conceptos de residuo y la suma de todos los residuos, disponibles a través del software, fueron cruciales para operacionalizar su idea intuitiva de cercanía de una recta a todos los puntos.

Un aspecto importante que conviene aclarar es que el programa de GeoGebra que hemos utilizado se basa en el concepto de residuo y en la consideración de la distancia lineal del conjunto de puntos a la recta (es decir, como la suma de los valores absolutos de los residuos). Sin embargo, la recta de regresión utiliza el concepto de distancia cuadrática. Por tanto, falta propiciar que los estudiantes comprendan que es mejor la distancia cuadrática que la lineal. Una discusión sobre esta ventaja se puede encontrar en Lesser (1999a) y en Lesser (1999b) de cómo utilizar la tecnología para motivarla.

En nuestra opinión, la aportación del presente trabajo a la educación matemática consiste en mostrar una posible trayectoria para ayudar a los estudiantes a comprender algunos conceptos esenciales para construir la recta de mejor ajuste, especialmente la idea de distancia de una recta a la nube de puntos con base en el concepto de residuo. Esta trayectoria propicia que los estudiantes adquieran los elementos para tener una vista agregada del conjunto de datos en relación con la recta de mejor ajuste. No obstante, reconocemos que solo se avanzó en que los estudiantes entiendan cómo se combinan los datos para generar la recta de mejor ajuste, quedando pendiente un mayor avance en las componentes de variabilidad e inferencia. Este y la transición a distancias cuadráticas son nuestros desafíos para investigaciones futuras.

REFERENCIAS

- Andre, M., y Lavicza, Z. (2019). Technology changing statistics education: Defining possibilities, opportunities and obligations. *The Electronic Journal of Mathematics and Technology*, 13, 253-264.
- Bakker, A., y Gravemeijer, K. (2004). Learning to reason about distribution. En D. Ben-Zvi y J. Garfield (Eds.), *The Challenge of developing statistical literacy, reasoning and thinking* (pp. 147-168). Kluwer. https://doi.org/10.1007/1-4020-2278-6_7

- Biehler, R., Ben-Zvi, D., Bakker, A., Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. En M. A. Clements *et al.* (Eds.). *Third International Handbook of Education* (pp. 643-689). Springer.
- Biehler, R., Frischmeier, D., Reading, C. y Shaughnessy, J. M. (2018). Reasoning about data. En D. Ben-Zvi, K. Makar, J. Garfield (Eds.). *International Handbook of Research in Statistics Education*. Springer.
- Burrill, G. y Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training 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. 57-69). Springer.
- Casey, S. A. (2014). Teacher's knowledge of students' conceptions and their development when learning linear regression. En K. Makar, B. Sousa, y R. Gould (Eds.), *Sustainability in Statistics Education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July 2014)* (pp. 1-6). International Statistical Institute.
- Casey, S. A. (2015). Examining student conceptions of covariation: A focus on the line of best fit. *Journal of Statistics Education*, 23(1), 1-33. <https://doi.org/10.1080/10691898.2015.11889722>
- Castro-Sotos, A. E., Vanhoof, S., Van den Noortgate, W., y Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, 2, 98-113.
- CCH-UNAM (s/f). Programa de Estudio de Estadística y Probabilidad I y II. https://www.cch.unam.mx/sites/default/files/plan_estudio/mapa_estadistica.pdf
- Cobb, P., McClain, K., Gravemeijer, K. (2003). Learning about statistical covariation. *Cognition and Instruction*, 21(1), 1-78.
- Confrey, J. (1990). A Review of the Research on Student Conceptions in Mathematics, Science, and Programming. En *Review of Research in Education* (vol. 16, pp. 3-56). Sage Publications, Inc., American Educational Research Association. <https://doi.org/10.2307/1167350>
- Desrosières, A. (2000). *La política de los Grandes Números. Historia de la razón estadística*. Editorial Melusina.
- Dierdorff, A., Bakker, A., Eijkelhof, H., & van Maanen, J. (2011). Authentic Practices as Contexts for Learning to Draw Inferences Beyond Correlated Data. *Mathematical Thinking and Learning*, 13(1-2), 132-151. <https://doi.org/10.1080/10986065.2011.538294>
- Estepa, A., y Batanero, C. (1996). Judgments of correlation in scatterplots: Students' intuitive strategies and preconceptions. *Hiroshima Journal of Mathematics Education*, 4, 24-41.

- Engel, J., y Sedlmeier, P. (2011). Correlation and regression in the training of teachers. En C. Batanero, G. Burril, C. Reading (Eds.), *Teaching Statistics in School Mathematics-Challenges for teaching and Teacher Education*. ICMI, Springer.
- Garfield, J., y Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for Research. *Journal for Research in Mathematics Education*, 19(1), 44-63.
- Garfield, J.B. y Ben-Zvi, D. (2008). *Developing Students' Statistical Reasoning. Connecting Research and Teaching Practice*. Springer.
- Gil, E., y Gibbs, A. L. (2017). Promoting modelling and covariational reasoning among secondary school students in the context of Big Data. *Statistics Education Research Journal*, 16(2), 163-190.
- Gravemeijer, K. (2013). Mathematics Education and the Information Society. En A. Damlamian, J. F. Rodrigues, R. Sträßer (Eds.) *Educational Interfaces between Mathematics and Industry* (pp.279-286). ICMI, Springer.
- Hancock, C., Kaput, J. J., y Goldsmith, L. T. (1992). Authentic Inquiry With Data: Critical Barriers to Classroom Implementation. *Educational Psychologist*, 27(3), 337-364. https://doi.org/10.1207/s15326985ep2703_5
- Hourigan, M., Leavy, A. M. (2021). Interrogating a measurement conjecture to introduce the concept of statistical association in upper elementary education. *Teaching Statistics*, 43(2), 62-71.
- Inzunza, S. (2016). Análisis de datos bivariados en un ambiente basado en applets y software dinámico. *Revista Educación Matemática*, 28(3), 61-89.
- Inzunza, S., y Ward, S. (2015). Explorando el razonamiento covariacional mediante un ambiente computacional en un curso introductorio de estadística. En C. Fernández, M. Molina, y N. Planas (Eds.), *Investigación en Educación Matemática XIX* (pp. 317-325). SEIEM.
- Konold, C., y Higgins, T. (2002). Highlights of related research. En y V. B. S.J. Russell, D. Schifter (Ed.), *Developing mathematical ideas: Working with data* (pp. 165-201). Dale Seymour Publications. <https://srri.umass.edu/publications/konold-2002hrr/>
- Lesser L. (1999a). The 'Ys' and 'why nots' of line of best fit. *Teaching Statistics*, 21(2), 54-55.
- Lesser, L. (1999b). Making the Black Box Transparent. *The Mathematics Teacher* 92(9),
- Makar, K. y Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, 8(1), 82-105.
- Medina, M., Sanchez, E., y Silvestre, E. (2019). Covariational reasoning patterns exhibited by high school students during the implementation of a hypothetical learning trajectory. En Uffe Thomas Jankvist and Marja van den Heuvel-Panhuizen and Michiel Veldhuis (Ed.), *Proceedings of the Eleventh Congress of the European Society for Research in Mathematics Education (CERME11)*. <https://hal.archives-ouvertes.fr/hal-02412811>

- Mokros, J., y Russell, S. J. (1995). Children's Concepts of Average and Representativeness. *Journal for Research in Mathematics Education*, 26(1), 20–39. <https://doi.org/10.2307/749226>
- Moore, D. S. (2000). *Estadística Aplicada Básica*. Antoni Bosch editor.
- Morgado, C. (2023). *Experimento de diseño con soporte tecnológico para el desarrollo del razonamiento estadístico covariacional de estudiantes de bachillerato*. Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, México.
- Morgado, N., y Sánchez, E. (2022). Razonamientos informales de estudiantes de bachillerato al enfrentar una tarea para introducir la recta de mejor ajuste. *Números*, 112, 65–82. <http://sinewton.es/publicacion-numeros/articulo-4-112/>
- Moritz, J. (2004). Reasoning about covariation. En D. Ben-Zvi y J. Garfield (Eds.). *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*. Kluwer Academic Publishers.
- Nagle, C., Casey, S., y Moore-Russo, D. (2017). Slope and line of best fit: A transfer of knowledge case study. *School Science and Mathematics*, 117(1–2), 13–26. <https://doi.org/10.1111/ssm.12203>
- NCTM (2000). *Principles and Standards for School Mathematics*. The National Council of Teachers of Mathematics, Inc.
- Neumann, D. L., Hood, M., Neumann, M. M. (2013). Using real-life data when teaching statistics: Students perceptions of this strategy in an introductory statistics course. *Statistics Education Research Journal*, 12(2), 59–70.
- Rossman, A. J. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal*, 7(2), 5–16.
- Rubin, A., Hammerman, J., y Konold, C. (2006). Exploring informal inference with interactive visualization software. En A. Rossman and B. Chance (Eds.), *Working cooperatively in statistics education: Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil. [CDROM]. International Statistical Institute.
- Sánchez-Cobo, F., Estepa, A., Batanero, C., Castro, E., y Bernabeu, B. (2000). Un estudio experimental de la estimación de la correlación a partir de diferentes representaciones. *Enseñanza de Las Ciencias*, 18(2), 297–310.
- Shaughnessy J. M., (1992). Research in probability and statistics: Reflections and directions. En D. A. Grouws (Ed), *Handbook of Research on Mathematics Teaching and Learning* (pp. 465-494). Macmillan Publishing Company.
- Sorto, M. A., White, A., y Lesser, L. M. (2011). Understanding student attempts to find a line of fit. *Teaching Statistics*, 33(2), 49–52. <https://doi.org/10.1111/j.1467-9639.2010.00458.x>
- Starnes, D. S., Yates, D., y Moore, D. S. (2010). *The practice of statistics* (4th ed.). W.H. Freeman and Company/BFW.

- Stigler, S. M. (2016). *The Seven Pillars of Statistical Wisdom*. Harvard University Press.
- Tobías-Lara, M. G. y Gomez-Blancarte, A. L. (2019). Assessment of informal and formal inferential reasoning: a critical research review. *Statistics Education Research Journal*, 18(1), 8–25.
- Turan, J. (1995). Some Undergraduates' Understanding of the Meaning of a Correlation Coefficient. En B. Atweh y S. Flavel (Ed.), *Annual conference; 18th, Mathematics Education Research Group of Australasia (MERGA)* (pp. 524–529). Northern Territory University.
- Watson, J., y Moritz, J. (2007). Developing aspects of distribution in response to a media-based statistical literacy task. International Statistical Institute 56th Session.
- Webb, N. M. y Mastergeorge, A. (2003). Promoting effective helping behavior in peer-directed groups. *International Journal of Educational Research* 39(2), 73–97. doi:10.1016/S0883-0355(03)00074-0
- Yackel, E., Cobb, P., y Wood, T. (1991). Small-group interactions as a source of learning opportunities in second-grade mathematics. *Journal for Research in Mathematics Education*, 22(5), 390–408.
- Zieffler, A. S., Garfield, J., DelMas, R., y Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 40–58.
- Zieffler, A. S., y Garfield, J. (2009). Modeling the growth of students' covariational reasoning during an introductory statistics course. *Statistics Education Research Journal*, 8(1), 7–31.

Autor de correspondencia

CINDY MORGADO

Dirección: Universidad Autónoma de Bucaramanga (UNAB), Bucaramanga, Colombia
Código Postal: 680003
cmorgado@unab.edu.co