NÚMEROS
Revista de Didáctica de las Matemáticas
http://www.sinewton.org/numeros
ISSN: 1887-1984
Volumen 112, noviembre de 2022, páginas 65-82

Razonamientos informales de estudiantes de bachillerato al enfrentar una tarea para introducir la recta de mejor ajuste

Cindy Nathalia Morgado Hernández
Ernesto Alonso Sánchez Sánchez
(Centro de Investigación y de Estudios Avanzados – CINVESTAV. México)

Fecha de recepción: 02 de diciembre de 2021
Fecha de aceptación: 09 de marzo de 2022

Resumen

Se reportan resultados de una investigación sobre el razonamiento estadístico covariacional informal realizada con 22 estudiantes entre 16 y 18 años, en la cual se diseñó e implementó una tarea utilizando el software *CODAP* y un applet de GeoGebra para introducir la recta de mejor ajuste. El diseño de la tarea se caracteriza por poseer elementos que prevén las malas concepciones reportadas en la literatura, y por enfocarse en cuatro ideas estadísticas que se consideran son centrales para el desarrollo del razonamiento informal acerca de la recta de mejor ajuste. Después de haber utilizado *CODAP* y el applet de GeoGebra, los estudiantes transitaron de ver a una nube de puntos como puntos individuales, o fragmentada en subconjuntos a verla como un agregado a partir de un mecanismo matemático que los une con la noción de distancia de una nube a una recta.

Palabras clave Razonamiento estadístico informal, recta de mejor ajuste, diseño de experimento, tecnología

Title **Informal reasoning of high school students when faced with a task to introduce the line of best fit**

Abstract

The results of an investigation on informal covariational statistical reasoning carried out with 22 students between 16 and 18 years old are reported, in which a task was designed and implemented in a digital technology environment to introduce the best line fit. The task design is characterized by having elements that anticipate misconceptions reported in the literature and by focusing on four statistical ideas that are central to the development of informal reasoning about the line of best fit. After using the digital technology environment, the students move from seeing a point cloud as individual points or fragmented into subsets to seeing it as an aggregate from a mathematical mechanism that unites them with the notion of distance from a cloud to a line.

Keywords Informal statistical reasoning, line of best fit, design of experiment, technology

1. Introducción

En la ciencia, la tecnología, y en la vida diaria y profesional es indispensable conocer las relaciones entre diferentes variables; por ejemplo, el volumen de un gas a temperatura constante está relacionado con la presión; el costo total de la factura de agua en relación con los litros consumidos por mes en cada vivienda; la distancia que recorre un objeto cuando cae depende del tiempo transcurrido, etc. Pero muchas relaciones entre variables no son determinísticas en el sentido de que el valor de una de ellas conduzca a un valor preciso de la otra. Por ejemplo, un estudio determinó que las mujeres de estatura baja son más propensas a sufrir un infarto cardiaco que las mujeres altas, pero no toda mujer de estatura baja sufrirá un infarto (Moore, 2005); un recién nacido con peso por debajo del promedio es propenso a ser enfermizo durante su crecimiento, pero no todos los nacidos con peso por debajo del promedio serán enfermizos (Wild y Seber, 2000). Así, debido a que contienen elementos de incertidumbre, las relaciones entre variables estadísticas tienen una mayor complejidad que las relaciones entre variables determinísticas. ¿Cómo representar, medir y utilizar las relaciones entre dos variables estadísticas? Bajo la expresión regresión y correlación se agrupa un sistema de conceptos y procedimientos que abordan y responden a dicha pregunta.

Diversas investigaciones didácticas muestran que los estudiantes tienen dificultades para aprender el tema de correlación, por ejemplo, las ideas preconcebidas que tienen los inhibe de utilizar los datos adecuadamente para detectar la relación entre dos variables cuando es diferente a la que esperan; suelen subestimar o sobrestimar la correlación entre dos variables dependiendo de la parte de datos en la que enfocan su atención; también tienen dificultades para entender la relación de la correlación con la presencia o ausencia de relaciones causales y, a veces, identifican correlación con causalidad (Estepa y Sánchez-Cobo, 2003; Zieffler y Garfield, 2009). Con respecto a problemas de regresión, los estudiantes tienen dificultades en distinguir entre la variable explicativa y la variable de respuesta en problemas ubicados en contextos no-matemáticos; les cuesta trabajo aplicar de manera pertinente sus conocimientos sobre funciones, frecuentemente unen los puntos de una nube creyendo que se revelará una función; cuando se les pide que tracen la recta de mejor ajuste a los puntos de un diagrama de dispersión exhiben estrategias de pertenencia de los puntos a la recta (que pasa por dos puntos extremos, o la que más puntos atrape) o de partición de la nube en dos conjuntos con el mismo número de puntos (Casey, 2014, 2015); en general, tienen dificultades para determinar la tendencia lineal de una nube de puntos y la fuerza de dicha tendencia.

Varios investigadores han comentado que detrás de algunas de estas dificultades se encuentra una tendencia de los estudiantes a concebir a un conjunto de datos bivariados como la suma de puntos aislados con sus propiedades particulares y no como una entidad con propiedades globales. El presente trabajo presenta una situación-problema que con la ayuda de la tecnología (CODAP y GeoGebra) ofrece una manera de avanzar en la superación de concebir un conjunto bivariado como puntos aislados (Bakker y Gravemeijer, 2004; Casey, 2015; Hancock et al., 1992; Konold y Higgins, 2002). Más adelante veremos con más detalle lo que significa ver un conjunto de puntos como una entidad versus como la reunión de datos aislados.

El objetivo del estudio se centró en analizar los razonamientos informales de 22 estudiantes entre 16 y 18 años al enfrentar una tarea utilizando tecnología digital (hacemos referencia al uso del software *CODAP* y un applet de GeoGebra) para introducir la recta de mejor ajuste.

2. Antecedentes

Los órganos humanos se pueden pensar de dos maneras, una como órganos aislados (oídos, cerebro, estomago, etc.), otra, como parte de la entidad “cuerpo humano”; pensarlo de esta manera permite percibir las funciones de cada órgano en términos de propiedades del cuerpo humano. De manera análoga, un conjunto de datos se puede pensar como datos aislados o como parte de una entidad que los trasciende. En el análisis estadístico los datos deben ser concebidos como una entidad o parte de una entidad que tiene propiedades globales que suelen no percibirse en cada dato particular. En inglés se utiliza el término "agregado" para referirse a tales entidades y corresponde a la definición en castellano de: "Conjunto de cosas homogéneas que se consideran formando un cuerpo" (Real Academia Española, definición 6). Varios investigadores en educación estadística 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.

Para Hancock et al. (1992), el razonar acerca del agregado se define como un razonamiento basado en construir valores representativos para grupo, y se ocupa de los patrones y relaciones en el conjunto de datos como un todo (Pfannkuch y Wild, 2004), se diferencia del razonamiento basado en datos individuales con poco intento de relacionarlos con el conjunto de datos más amplio. Para estos autores no poseer un razonamiento basado en agregados es un serio problema para la modelización de datos porque el concepto de valores representativos es un vínculo crítico en la lógica de la mayoría de las investigaciones basadas en datos, donde se establece un diálogo entre los datos y los modelos estadísticos.

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).

Bakker (2004) y Bakker y Gravemeijer (2004) hacen referencia a la vista agregada de datos cuando se perciben como un todo que tiene características de una entidad que no son visibles en ninguno de los casos individuales. 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 quedará 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).

La definición más reciente sobre la noción de agregado desde una perspectiva estadística es dada por Stigler (2016) quien establece que dicha noción es uno de los siete pilares de la sabiduría estadística. Para Stigler (2016) un agregado es un objeto nuevo derivado de un conjunto de datos que subsume o representa a los datos individuales contenidos en la totalidad.

3. Marco Conceptual

En las siguientes cuatro subsecciones se exponen algunos conceptos que son importantes para entender la investigación de la que ha surgido el presente informe. La noción de marco conceptual que aquí se maneja es entendido de una manera similar como lo definen Miles y Huberman:

> Un marco conceptual explica, ya sea gráficamente o en forma narrativa, las principales cosas que van a ser estudiadas – los factores clave, las variables o los constructos – y las relaciones que se presume hay entre ellas. Un marco puede ser simple o elaborado, del sentido común o dirigido por una teoría, descriptivo o causal (Miles y Huberman, 1994, p.18)

Se propone cuatro conceptos: agregado, agregación, razonamientos estadístico informal, razonamiento informal sobre covariación y la influencia de la tecnología digital en el razonamiento covariacional; estos conceptos interconectados estructuran esta investigación.

3.1. Definición de agregado y agregación en estadística

Tanto la correlación como la regresión son objetos estadísticos que expresan propiedades globales de un conjunto de datos, propiedades que no pertenecen a puntos particulares sino a todos a la vez. Un dato individual, aislado, no contiene las propiedades que emergerán cuando se asocia con otros datos, es decir, cuando se trata un conjunto de datos como una entidad. Esta caracterización trata de sintetizar los comentarios de Hancock et al. (1992), Konold y Higgins (2002), Bakker y Gravemeijer (2004) referente al agregado.

Una condición para que la media de un conjunto de datos o la recta de mejor ajuste de una nube de puntos se vean como representantes de sus respectivos conjuntos de datos es que estos se deben concebir como agregados. Desde una perspectiva educacional surge la pregunta ¿Cómo llega un estudiante a concebir que un conjunto de datos es un agregado? Nosotros sostenemos la hipótesis de que hay dos posibles fuentes; a saber, el contexto y la matemática. Respecto a la primero, se puede percibir que todos los datos pertenecen a un fenómeno o que se asocian con una relación causal, en este caso se ven los datos dentro de un contexto que los unifica. Con relación a la perspectiva matemática, se puede percibir a un dato como parte de un agregado si pertenece a un conjunto que está unido por una propiedad matemática, v, gr., números pares en un conjunto de números, o estar alineados en un conjunto de datos bivariados, etc.

3.2. El razonamiento estadístico informal

Se entiende por razonamiento a los procesos para obtener y verificar proposiciones (conclusiones) sobre la base de la evidencia (datos) o de conocimientos establecidos o suposiciones. El razonamiento puede tomar muchas formas que van desde argumentaciones informales hasta demostraciones deductivas (Martin et al., 2009). El razonamiento estadístico informal es el razonamiento cuyo contenido se relaciona con datos, muestras, azar, inferencia y relaciones entre variables estadísticas. En consecuencia, el razonamiento estadístico informal sobre covariación está relacionado con conjuntos de datos bivariados y relaciones entre variables estadísticas. En la presente investigación el propósito es desarrollar el razonamiento estadístico informal de los estudiantes sobre la recta de mejor ajuste. Para esto se definieron nociones informales de distancia lineal (en lugar de cuadrática) y de línea de mejor ajuste, que no coinciden con los conceptos formales, pero no son inconsistentes con estos y tienen la ventaja de estar más cerca de las intuiciones de los estudiantes.

3.3. El razonamiento informal sobre covariación

Un enfoque informal para el desarrollo del razonamiento estadístico plantea formas alternativas de abordar el tema que, por un lado, permitan evitar los altos niveles de sofisticación matemática que conlleva el acercamiento formal, por el otro, se apoye de manera conveniente en las intuiciones y conocimientos que ya tienen los estudiantes, teniendo en cuenta su propensión a ciertas falsas concepciones. El uso de la tecnología ayuda a evitar la complejidad matemática, mientras que la literatura didáctica sobre el tema proporciona información acerca de algunas intuiciones y falsas concepciones que frecuentemente tienen los estudiantes. En seguida, destacaremos algunos conocimientos didácticos sobre correlación y recta de mejor ajuste reportados en la literatura. Uno es la tendencia de los estudiantes a juzgar sobre la relación entre dos variables basados más en sus creencias personales acerca de la relación que sobre los datos disponibles (Moritz, 2004). También los conceptos previos de función lineal en matemáticas suelen influir de manera importante, pero no siempre conveniente, en la tarea de ajustar una recta (Casey y Nagle, 2016). Estepa y Batanero (1996) descubrieron lo que llamaron una concepción determinística de la correlación; ésta se presenta cuando la relación entre las variables se considera solo desde un punto de vista funcional. También mencionan una concepción local de la correlación, esta se presenta cuando se usa sólo una parte de los datos y se generalizan las conclusiones a todo el conjunto de datos. Casey (2014, 2015) describe las siguientes estrategias observadas en las respuestas de los estudiantes a la tarea de trazar la recta de mejor ajuste: trazar una especie de mediana, en el sentido de proponer una recta que divide los puntos de datos, la mitad de los puntos a un lado de la recta y la otra mitad en el otro lado de la recta. Otra estrategia es trazar la recta a través de puntos que son puntos medios de diferentes grupos de puntos. Además, la concepción “cercanía” entre la línea y la nube de puntos puede surgir (vagamente) en las respuestas de los estudiantes, pero falta mayor precisión, por ejemplo, considerando los residuos o la suma de los cuadrados de las distancias verticales entres los valores observados y los valores predichos.

3.4. La influencia de la tecnología digital en el razonamiento sobre covariación estadística

La investigación sugiere que la tecnología puede ayudar a los estudiantes a hacer juicios de covariación más precisos (Batanero et al., 1998; Cobb et al., 2003; Inzunza, 2016); de hecho, la tecnología posee un rol muy importante en la estadística, dado que produce representaciones visuales, interactivas y dinámicas de sus conceptos. El uso de tales representaciones permite un enfoque centrado más en los conceptos en lugar de los algoritmos y cálculos, donde la interactividad y la calidad de uso de las gráficas permiten llevar a cabo experimentaciones con los datos. Todo lo anterior permite enganchar a los estudiantes en actividades productivas (Biehler et al., 2013). Específicamente, para las temáticas de regresión y correlación la tecnología ofrece los siguientes rasgos relevantes para su enseñanza:

1. La posibilidad para formar gráficos, en especial diagramas de dispersión, y ajustar visualmente una línea de regresión mostrando visualmente las desviaciones cuadráticas cambiantes de la línea a medida que esta se ajusta a los datos.
2. Obtener el valor numérico del coeficiente de correlación y determinar la expresión algebraica de la recta de regresión.
3. Arrastrar puntos del diagrama de dispersión y observar en tiempo real el efecto de su ubicación dentro de la nube sobre la fuerza (coeficiente de correlación) y dirección (línea de regresión) de la relación, lo cual permite ver las interacciones entre los elementos que se arrastran y las medidas estadísticas.
4. La vinculación de múltiples representaciones que permite descubrir y observar patrones y tendencias en los datos simultáneamente desde diferentes perspectivas (la gráfica, las medidas resumen, la recta de regresión) En el presente estudio mostramos cómo con la ayuda de la tecnología los estudiantes pueden concebir a la recta de mejor ajuste como resultado de un proceso de agregación.

4. Marco Metodológico

A continuación, se detallan cada uno de los elementos metodológicos que guiaron la presente investigación.

4.1. Tipo de investigación

Corresponde a un experimento de diseño basado en los cinco aspectos del entorno del aula de los "principios del diseño instruccional" de Cobb y McClain (2004): El enfoque sobre ideas estadísticas centrales, las tareas instruccionales, la estructura de la tarea en el salón de clase, el uso de herramientas computacionales y el discurso en la clase.

1. *Ideas estadísticas centrales:* Con relación al contenido, el diseño se debe enfocar en el desarrollo de ideas estadísticas centrales y no, o no solo, en el dominio de los algoritmos y procedimientos que instrumentalizan dichas ideas. Una idea central se concibe como un objetivo de aprendizaje al que se quiere llegar desarrollando varias ideas y recursos. Las técnicas, procedimientos y subconceptos constituyen una idea central. Por lo tanto, una idea central se podría componer de una o más ideas. La idea central en esta investigación es la recta de mejor ajuste que comprende los siguientes elementos:

- La recta de mejor ajuste depende de todos los puntos del diagrama de dispersión, esto significa que en el proceso de su construcción deben utilizarse todos los datos dados
- Dada una nube de puntos y una recta, se define el error (llamado residual) de un punto como el valor absoluto de la diferencia entre la ordenada del punto y la ordenada de la proyección del punto en la recta de ajuste, es decir:
$$e = |y_i - \hat{y}_i\,|$$
- La medida (residuos) de cercanía o lejanía entre la nube de puntos y la recta de ajuste se calcula sumando todos los residuales, mediante:
$$\sum_{i=1}^{n} |y_i - \hat{y}_i\,|$$

- Utilizamos una noción de distancia lineal y no de distancia cuadrática porque aquella resulta más intuitiva para los estudiantes, no obstante, nos parece claro que, entendiendo la idea con distancia lineal, se puede generalizar fácilmente considerando distancia cuadrática, argumentando la razón de las ventajas de esta.

- La recta que mejor se ajusta a los datos es aquella donde la suma de residuales es mínima, lo que implica determinar:

$$min\left\{\sum_{i=1}^{n} |y_i - \hat{y}_i\,|\right\}$$

  Este concepto es difícil porque se ha definido una función distancia cuya variable independiente son las rectas en el plano. Es decir, para cada recta en el plano se determina el conjunto: $\{\hat{y}_1, \hat{y}_2 \ldots \hat{y}_n\}$ y es sobre todos estos conjuntos que se determina el mínimo. Conviene subrayar que cada conjunto está determinado por los parámetros m y b de una recta de la forma $y = mx + b$.

2. *Uso de tecnología*: Se utiliza la plataforma online gratuita CODAP (https://codap.concord.org/) dado que es de interacción para realizar gráficos de dispersión y calcular coeficientes de

correlación, además diseña un applet con el software GeoGebra, pues sus características permiten el diseño de los elementos que consideramos relevantes para que el estudiante observe e interactúe. Para la línea de mejor ajuste dos rasgos son relevantes: 1) La posibilidad para formar gráficos de dispersión y 2) La de trazar, manipular y ocultar rectas, distancias y expresiones algebraicas. Asumimos la perspectiva de Biehler et al. (2013) de utilizar la tecnología digital para hacer estadísticas visuales, interactivas y dinámicas, enfocándose más en los conceptos en lugar de los algoritmos y cálculos, donde la interactividad y la calidad de uso de las gráficas permite experimentar con los datos y enganchar a los estudiantes. La tecnología digital influiría en la exploración de datos y en la preparación para la inferencia estadística; con el fin de ser utilizada para promover razonamientos estadísticos más allá de ser una caja que contiene recetas algorítmicas.

3. Para el *discurso en el aula* se sugiere que el profesor sea el encargado de monitorear, coordinar y dar sentido a las interacciones propias de los estudiantes, así como entre los estudiantes y la herramienta tecnológica. El profesor debe abandonar la instrucción formal del tema. Además, debe intervenir para que los estudiantes hagan uso de la terminología que va apareciendo y del lenguaje que se quiere que vayan adquiriendo durante el desarrollo de la tarea.

4. *La estructura de la tarea en el salón de clase:* debe organizarse la administración de la tarea y problemas de modo que los estudiantes tengan la oportunidad de explorar soluciones, de compararlas con las de sus compañeros, y aclararlas en reunión grupal. Debe definirse: qué tareas se llevan a cabo individualmente; cuáles en equipos; cuáles se harán en el salón de clase sin disponibilidad de equipo de cómputo y cuáles en la sala de cómputo. Estos aspectos hacen parte del método, y se puede resumir en determinar los participantes, instrumentos y procedimientos, lo cuales se describen en el siguiente apartado junto con la tarea. El quinto aspecto hace referencia a la tarea instruccional que es explicada en el siguiente apartado.

4.2. Método

4.2.1. Sujetos

En el estudio participaron 22 estudiantes de 4º semestre (16-18 años) de un colegio ubicado en el norte de la Ciudad de México. Los estudiantes pertenecían a un grupo que cursaba la asignatura Matemáticas IV (4º semestre) cuyo contenido es funciones (polinomiales, racionales, radicales, trigonométricas, exponenciales y logarítmicas). Los contenidos de las asignaturas de los semestres previos abarcan aritmética, álgebra, geometría y geometría analítica, pero no incluyen probabilidad y estadística, ésta la estudiarán hasta el 5º semestre. No obstante, en el ciclo escolar previo (nivel medio básico) se estudian los temas de lectura y elaboración de gráficas, medidas de tendencia central e introducción a la probabilidad.

Se eligió realizar la experimentación con estudiantes de 4º semestre pues en este grado se asegura que los estudiantes han visto plano cartesiano, funciones lineales y funciones cuadráticas que son antecedentes útiles para estudiar desde un acercamiento informal el tema de regresión y correlación. Aunque desde el diseño se valoró que convenía trabajar con estudiantes de 4º semestre, el grupo con el que se llevó a cabo este estudio fue elegido por conveniencia, ya que la profesora titular de dicho grupo es una compañera que tiene estudios de doctorado en Educación Matemática y que ha participado, con los autores de este artículo, en seminarios sobre temas de esta disciplina. Se le solicitó que nos permitiera insertar en su curso las actividades de regresión y correlación diseñadas para el estudio. Cabe mencionar que el sistema de bachillerato del colegio público en Ciudad de México se caracteriza por dar una gran libertad a los profesores para diseñar sus cursos sin dar directrices que fuercen a seguir un programa rígido, por lo que fue posible incluir en el curso las actividades de regresión y recta de mejor ajuste con ayuda de tecnología.

El curso fue mixto (hombres y mujeres) y de jornada matutina, la aplicación del instrumento se realizó en una clase de dos horas en enero del 2020.

4.2.2. Instrumento: La tarea

Para el instrumento de recolección de datos se tomó un problema del libro de texto "The practice of statistics" de Starnes et al. (2010), el cual se utilizó para realizar el diseño de la tarea. A continuación, se muestra la tarea que comprende un archivo digital de GeoGebra junto con enunciados instruccionales para los estudiantes.

Tarea: ¿Estar inquieto te mantiene delgado? 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 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: inquietud, vida diaria y cosas por el estilo.

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

Tabla 1. Medidas de cambio en ANE y aumento de grasa en los 12 adultos jóvenes.
Datos tomados de Starnes et al. (2010)

A lápiz y papel.

Discute con tu compañero y responde.

a) ¿Cuál consideras es la variable respuesta y la variable explicativa? Describe la relación entre el cambio ANE y el aumento de grasa.

Utilizando tecnología digital.

b) Realiza el diagrama de dispersión en CODAP (https://codap.concord.org/) y describe el comportamiento de la nube de puntos (intensidad y dirección).
c) Abre el archivo de GeoGebra llamado ANE_Grasas.ggb (ver figura 1, esta figura no es mostrada en la hoja de trabajo del estudiante). El archivo de GeoGebra muestra el diagrama de dispersión de los datos (puntos azules) de la tabla 1 junto con los siguientes elementos:

- Una línea amarilla la cual puedes mover con los símbolos en rojo ◆ .
- El segmento punteado de color rojo muestra la distancia que va desde el punto azul a la recta amarilla, su valor corresponde a la diferencia entre la ordenada del punto azul (y) y la ordenada del punto que pertenece a la recta amarilla ($\hat{y}$), a este valor lo llamaremos residual. Es decir,

e =valor y (punto de la nube) – valor $\hat{y}$ (punto posible de la línea amarilla).

- La palabra residuos muestra la suma de todos los residuales.

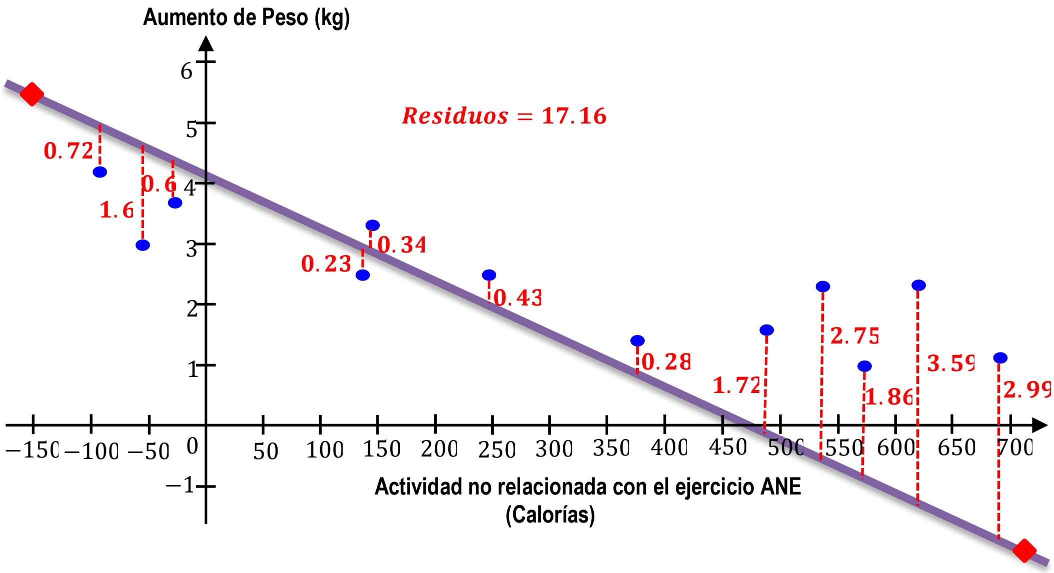


Figura 1. Captura de pantalla del archivo de GeoGebra ANE_Grasas.

d) Mueve la línea amarilla, explora lo que ocurre con el valor de cada residual y responde las siguientes preguntas:

1. ¿Qué ocurre con el valor de cada residual si alejas la línea amarilla de la nube de puntos azules? ¿Qué ocurre con el valor de los residuos?
2. ¿Qué ocurre con el valor de cada residual si acercas la línea amarilla a la nube de puntos azules? ¿Qué ocurre con el valor de los residuos?
3. ¿Dónde consideras debes ubicar la línea amarilla para que el valor de los residuos sea el mínimo? ¿Puedes asegurar que la ubicación que consideras es única? ¿sí? ¿no? ¿por qué?
4. Sigue las indicaciones del profesor y determina la línea de mejor ajuste (deberá aparecer de color rojo). ¿La manera como ubicaste la línea amarilla es igual a la línea roja arrojada por GeoGebra? ¿Sí? ¿No? ¿En qué se diferencian?
5. ¿Cuál crees es el criterio que utiliza GeoGebra para determinar la línea de mejor ajuste?

Los elementos que comprenden la tarea fueron establecidos con base en los “principios del diseño instruccional” de Cobb y McClain (2004) expuestos en la subsección 4.1 y los conceptos que abarcan el marco conceptual descritos en la sección 3. El análisis previo se describe a continuación:

La vista gráfica de GeoGebra muestra una recta amarilla que es movible sosteniendo el clic en cualquier lugar sobre ella, esta recta se mueve de diferentes maneras como se desee, variando su pendiente o variando su punto de corte con el eje y. La característica intrínseca de la línea amarilla de estar conectada a los puntos azules (datos de la tabla 1) se enfoca en la idea de concebir a una nube de puntos como un agregado, es decir, todos los puntos de la nube influyen en la determinación de la línea de mejor ajuste con el fin de prever la concepción local de la asociación (Estepa y Batanero, 1996) donde los estudiantes conciben la nube de puntos como dos grupos, los que pertenecen a un modelo lineal y los que no pertenecen, trazando la línea teniendo en cuenta algunos puntos de la nube, ejemplo, por el primer punto y el último punto (Casey, 2015).

Los interrogantes 1 y 2 se encaminan en la idea estadística del error como la distancia entre un punto (posible) en el modelo y un punto de la nube (ver figura 2), es decir:

$$d = \text{valor } y \text{ (punto de la nube)} - \text{valor } \hat{y} \text{ (punto posible).}$$

La anterior distancia es llamada residual.

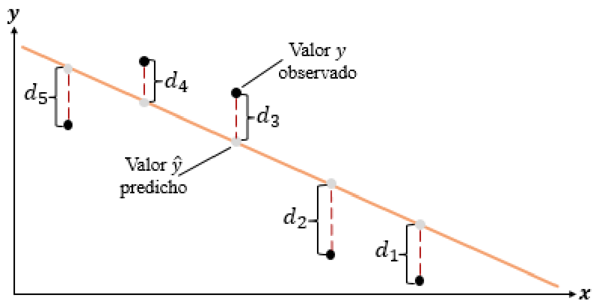


Figura 2. Algunos elementos que caracterizan la línea de mejor ajuste.

La pregunta 3 pretende llevar a definir una medida de la cercanía (residuos) entre la recta y la nube de puntos, donde se tiene que la suma de todas las distancias (residuales) representa que tan cerca o lejos se encuentra la recta de la nube de puntos.

Con las preguntas 4 y 5 se busca definir la línea que mejor se ajusta a los datos como aquella donde los errores (residuales) o las distancias verticales son las más pequeñas posibles en algún sentido promedio en general.

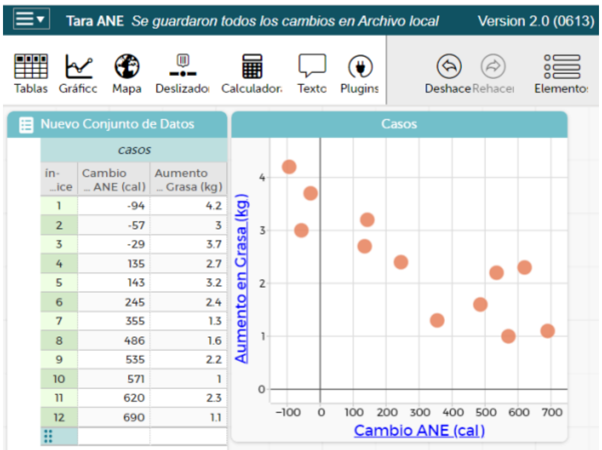
Es importante aclarar que en bachillerato se enseña la métrica de mínimos cuadrados y que la recta que arroja GeoGebra es la recta de mínimos cuadrados en la cual se minimiza la suma de los cuadrados de las distancias. Sin embargo, el objetivo de la tarea no es llegar a la formalización del método de mínimos cuadrados, sino generar razonamientos entorno a la nube de puntos y la posible recta de mejor ajuste. Consideramos que, entendiendo la idea con distancia lineal, se puede generalizar fácilmente considerando distancia cuadrática, argumentando la razón de las ventajas de esta. La tarea se podría extender en proponer preguntas para que los estudiantes razonen acerca del por qué es más adecuado asumir la suma del cuadrado de los residuales en lugar de la suma de su valor absoluto.

Destacamos que, a pesar de la aparente idoneidad de la tecnología digital en la clase de estadística, es necesario documentar mediante investigaciones empíricas el uso específico de herramientas tecnológicas digitales dentro de las actividades de instrucción, las posibilidades y limitaciones de las conceptualizaciones que pueden llevar a cabo los estudiantes.

4.2.3. Recolección de datos

La obtención de datos fue por medio de hojas de trabajo para desarrollar en parejas. Lo importante fue adquirir evidencia donde los estudiantes expresaran su razonamiento. La autora del presente estudio realizó la aplicación del instrumento. El proceso para recopilar los datos se describe a continuación:

i) El tiempo de aplicación del instrumento fue de 2 horas. La primera parte a lápiz y papel tuvo una duración de 30 minutos, la segunda parte utilizando CODAP y el archivo de GeoGebra tuvo una duración de una hora y 30 minutos.
ii) La aplicación del instrumento se realizó en una sala de ordenadores donde los estudiantes se ubicaron en parejas para desarrollar la tarea.
iii) Primero, el profesor entregó la situación problema de la tarea y orientó a los estudiantes a juzgar la relación entre las variables a lápiz y papel (sin uso de tecnología digital).
iv) Seguido, el profesor explicó cómo realizar el diagrama de dispersión en CODAP para pedirles que describan el comportamiento de la nube de puntos. En la figura 3 se muestra el diagrama de dispersión realizado en CODAP (esta figura no se muestra en la hoja de trabajo del estudiante).
v) Finalmente, les indica abrir el archivo de GeoGebra y seguir las instrucciones de la hoja de trabajo. Como se describe en la subsección 4.1 el profesor es el encargado de monitorear, coordinar y dar sentido a las interacciones propias de los estudiantes, así como entre los estudiantes y la herramienta tecnológica y abandona la instrucción formal del tema.



Índice	Cambio ANE (cal)	Aumento Grasa (kg)
1	-94	4.2
2	-57	3
3	-29	3.7
4	135	2.7
5	143	3.2
6	245	2.4
7	355	1.3
8	486	1.6
9	535	2.2
10	571	1
11	620	2.3
12	690	1.1

Figura 3. Captura de pantalla del diagrama de dispersión realizado en CODAP

Se digitalizaron las respuestas de los estudiantes presentes en sus hojas de trabajo y empleando algunos métodos y procesos de la teoría fundamentada (Birks y Mills, 2015; Glaser y Strauss, 2008) se codificaron las respuestas para generar categorías y relaciones entre categorías. Las dos personas que estuvieron involucradas en el proceso de codificación fueron los autores del presente trabajo. Se desarrollaron los códigos de manera consensual e independiente.

Utilizando el método analítico de comparación constante, se compara entre sí datos (incidentes, frases, respuestas, etc.) para encontrar rasgos comunes y diferentes entre ellos que permita formar agrupaciones e identificar patrones. En la codificación inicial se generaron muchos códigos, donde se utilizaron palabras o grupos de palabras que se identificaron con un patrón o característica común en los datos. Los nombres de los códigos se derivan de diferentes maneras, algunos de las frases directamente o con un poco de variación, otros son nombrados por el investigador o están relacionados con los antecedentes teóricos. Con base en los primeros códigos que se continuó con el análisis comparativo constante entre las respuestas de los estudiantes, se anexaron nuevos incidentes, algunos de los cuales se etiquetan con los códigos que ya se tienen establecidos, y otros incidentes generaron nuevos códigos. Seguido, los códigos se comienzan a transformar en conceptos más abstractos y se van identificando las categorías, los códigos se van conglomerando de tal manera que tienen propiedades similares y producen una categoría. En esta etapa una categoría central comienza a hacerse evidente a medida que las categorías desarrolladas se forman alrededor de un concepto central (Chun Tie et al., 2019).

En la codificación final, los conceptos que llegan a formar las categorías y serán abstractos, representan historias de muchos. Es importante resaltar que las categorías no se desarrollan para un determinado estudiante o pareja, se desarrollan respecto a varias respuestas de los estudiantes (Birks y Mills, 2015).

5. Resultados

En seguida, exponemos algunas observaciones sobre los códigos y categorías que emergieron de las respuestas de los estudiantes al juzgar la relación entre las variables mostradas en la situación problema junto con la tabla de valores a lápiz y papel. Del análisis de las respuestas a la única pregunta sin uso de tecnología identificamos dos categorías que hemos llamado c*ovariación funcional* y *aleatoriedad*.

La estrategia de *covariación funcional* consiste en buscar y aislar los datos bivariados que se apeguen a un modelo matemático, es decir, ponen su atención hacia los puntos que logran ubicar sobre una recta. En general, los estudiantes cuyas soluciones entran en esta categoría dividen el conjunto de datos en dos partes, los que, para ellos, corresponden a un modelo lineal y los que no lo hacen. Entonces su descripción de los datos se refiere sólo al subconjunto de datos que corresponde al modelo e ignoran aquellos que quedan fuera; y con base en los datos que seleccionaron, describen la tendencia general. Así, sus afirmaciones, muchas veces coinciden con la tendencia de toda la nube, afirman, por ejemplo: "entre más grasa consumas más calorías aumentas", "entre menos cambio en las calorías mediante los movimientos, menos la grasa disminuirá", pero su sustento es sobre una parte de los datos Por otro lado, hay quienes involucran en sus descripciones características del *contexto del problema*: "la grasa aumenta porque se queda encapsulada", "entre menos actividad realices más grasa", "las grasas se quedan en el cuerpo". Es posible que la elección que hacen del conjunto de datos de la nube de puntos esté influenciada por sus creencias al respecto, aparte de seguir el criterio de que estén en una recta.

La estrategia de *aleatoriedad* se presenta cuando los estudiantes enfocan su atención en lo que ocurre en el paso de un punto al que le sigue dentro de la tabla de valores. Al darle seguimiento a las diferencias entre puntos sucesivos, en la nube de puntos, conforme repasan los puntos de izquierda a derecha, notan que no hay patrones previsibles, a veces la diferencia es positiva a veces negativa y el tamaño varía. Así, concluyen: "la grasa aumenta o disminuye dependiendo de cuántas calorías queman en ANE". "si las calorías disminuyen o aumentan dependiendo del ANE, la grasa aumentará o disminuirá también", "las calorías que se queman aumentarán los kg o las calorías que se quemen disminuirán los kg". Para los estudiantes que siguen esta estrategia es evidente que no hay una correlación, pues no se puede saber si de dato a dato subirán o bajarán los niveles de grasa. Ambas estrategias se caracterizan en que se mira a los datos poniendo atención en sus características individuales o parciales, y no como una entidad.

Así, el análisis de acuerdo con la primera estrategia consiste en separar los datos en dos, en la segunda, en repasarlos uno a uno y ver si aumenta o disminuye y qué tanto.

Cuando utilizan la plataforma CODAP para digitar la tabla de valores y realizar el diagrama de dispersión y juzgar el comportamiento de las variables, surge la *covariación inversa*, el cual describe de manera global el comportamiento de la nube, por ejemplo: “al aumentar las calorías hay más quema de grasa”, “entre más cambio en ANE, menos aumento de grasa presenta”, “cuando el cambio en ANE es mayor, el aumento de grasa es menor”, “más incremento en ANE más pérdida de grasa hay. Entre menos incremento en ANE menos pérdida de grasa”, también influye el *contexto del problema*, por ejemplo: “la grasa no aumenta porque, aunque no se hace ejercicio si se mantiene en movimiento”, “el aumento de grasa depende de la cantidad de ANE que se realiza diariamente”. De modo que sin utilizar tecnología los estudiantes enfocan su atención en la variación de punto a punto (*aleatoriedad*: la grasa a veces aumenta y a veces disminuye), mientras que el uso de la tecnología influye para que hagan una descripción de la tendencia general de la nube, ignorando las fluctuaciones particulares (*covariación inversa:* los valores de la grasa tienden a disminuir conforme los valores de ANE tienden aumentar). La representación gráfica de los datos permite su captación de un solo vistazo y, probablemente esto, propicia que se comiencen a considerar como una totalidad.

La siguiente tabla resume la frecuencia obtenida para cada código respecto a las preguntas que involucran la descripción de la relación entre las variables, donde los códigos con mayor frecuencia fueron el contexto, aleatoriedad y covariación inversa.

Pregunta	Código	Frecuencia (Parejas)
A lápiz y papel describir la relación entre las variables	Covariación funcional	3
	Aleatoriedad	5
	Contexto	3
Utilizando CODAP realizar el diagrama de dispersión, determinar la correlación y describir la relación	Covariación Inversa	8
	Contexto	3

Tabla 2. Frecuencia para cada código respecto a las preguntas de juzgar la relación entre las variables

Vemos que tratar la situación con ayuda de la tecnología y el uso de algunas de sus posibilidades puede influir en los estudiantes para abandonar su tendencia a destacar en su descripción de la nube a un conjunto de puntos individuales que comparan entre sí, para destacar una propiedad global que se hace evidente cuando se posee una vista agregada del conjunto de puntos. Además, con tecnología y sin tecnología la descripción de la relación entre las variables se ve influenciado por el contenido contextual del problema, los estudiantes no ven los datos simplemente como números, si no números con un contexto, lo cual para Moore (1990) es la definición de dato estadístico.

Respecto al análisis de las respuestas de los estudiantes cuando se les pide establecer la recta de mejor ajuste a los datos utilizando el applet de GeoGebra surgieron razonamientos que resumimos en los códigos relacionados con cercanía, distancia y residuos:

En muchas respuestas de los estudiantes cuando utilizan tecnología se revela la intuición de *cercanía* del conjunto de puntos a una recta; por ejemplo: “que los puntos azules estuvieron más cercanos entre sí a la recta”, “creo que es la recta que pasa cerca de todos los puntos y no los une, así cada uno tiene cierta distancia”. Esta idea intuitiva se vuelve operativa con una noción de *distancia* de una recta a un conjunto de puntos. Como se ha visto, el software permite calcular los residuos, es decir, las diferencias de las ordenadas de cada pareja de puntos con la misma abscisa, uno perteneciente a la nube y otro sobre la recta. Cada residuo puede verse como una distancia del punto de la nube a la recta (para el propósito actual no afecta el hecho de que estrictamente la distancia de un punto a una recta se defina como la distancia del segmento que pasa por el punto perpendicular a la recta) y la suma de residuos como la distancia total de la nube a la recta. Como dicha suma se puede ver en el software y se actualiza en tiempo real conforme se mueve la recta, los estudiantes de hecho, manipulan y ven una función real determinada por la nube de puntos y cuya variable independiente son las rectas en el plano; entonces se comienzan a dar cuenta que la recta que minimiza la función es la recta más cercana a los puntos, es decir, la de mejor ajuste. Las siguientes expresiones de los estudiantes indican algunas de sus ideas al respecto: “entre menos distancia haya entre los puntos y la línea, el residuo va a cambiar”, “entre más alejada esté la recta de ajuste de los puntos de la tabla en la gráfica es mayor el valor de los residuos ya que hay una mayor distancia” “cuando la recta está mejor centrada, el nivel de los residuos es menor, es decir, pasa centrándose en medio de los puntos”, “por el centro, ya que pasa cerca de la mayoría de los puntos y ahí el residuo es menor entre los puntos y la recta”.

A continuación, la tabla resume la frecuencia obtenida para cada código respecto a las preguntas que implicaban el uso del archivo de GeoGebra al determinar la recta de mejor ajuste, donde los códigos con mayor frecuencia en diferentes preguntas fueron distancia y cercanía.

Pregunta	**Código**	**Frecuencia (Parejas)**
¿Qué ocurre con el valor de cada residual si alejas la línea amarilla de la nube de puntos azules? ¿Qué ocurre con el valor de los residuos? ¿Qué ocurre con el valor de cada residual si acercas la línea amarilla a	Distancia (nube, recta)	4
	Distancia (nube y la recta) y residuos	5
	Recta centrada implica menor residuos	2
¿Dónde consideras debes ubicar la línea amarrilla para que el valor de los residuos sea el mínimo? ¿Puedes asegurar que la ubicación que consideras es única? ¿sí? ¿no? ¿por qué? ¿Cuál crees es el criterio que utiliza GeoGebra para determinar la línea de mejor ajuste?	Cercanía	4
	Distancia	4
	Ninguna Respuesta	3
¿La manera como ubicaste la Recta amarilla es igual a	Cercanía o lejanía	9
	Menor Residuos	2

la Recta de regresión arrojada por GeoGebra? ¿Sí? ¿No? ¿En qué se diferencian?		

Tabla 3. Frecuencia para cada código respecto a las preguntas para determinar la recta de mejor ajuste

Una condición para ver a un conjunto de datos como un agregado es imaginar un mecanismo racional que los una a todos en una entidad que los represente, aunque se pierda información individual. Cuando los estudiantes eligen entre la nube de puntos un subconjunto en una recta, consideran al subconjunto como agregado, pero ignoran y descartan muchos datos, no ven a toda la nube como un agregado. En cambio, el proceso de buscar la recta más cercana a la nube de puntos con una noción de distancia pone en evidencia que todos y cada uno de los puntos contribuyen para determinar dicha recta. Este hecho resulta significativo debido a que en la literatura se ha destacado que uno de los problemas para que los estudiantes entiendan las nociones estadísticas de centros, variación, comparación de datos y distribución de datos, es que vean al conjunto de datos en ciernes como un agregado (Bakker y Gravemeijer, 2004; Konold y Higgins, 2002; Roseth et al., 2008). Por otro lado, también el contexto es un factor importante para ver a un conjunto de datos como sistema, pues en este caso es la existencia de un mecanismo causal (el término es de Zimmerman, 2007) que sugiere que todos los datos están relacionados, en nuestro caso, que el ANE y la acumulación de grasa son parte de un proceso causal y que se manifiesta mediante la tendencia de la nube de puntos. Sin embargo, como advierten Konold et al. (1997) el contexto indujo al error a los estudiantes, pues magnificaron sus experiencias personales al producir la respuesta en lugar de confiar en la evidencia basada en datos “agregado” (entidad); por esto, la concepción de sistema de datos es adquirida mediante un balance de consideraciones matemáticas y consideraciones de contexto.

6. Conclusión y comentarios finales.

Concebir a un conjunto de datos bivariados como un agregado es una condición necesaria para entender los temas de regresión y correlación. Sin embargo, los estudiantes no alcanzan dicha concepción de manera espontánea (Mokros y Russell, 1995). En las exposiciones clásicas del tema no se define o discute la noción de agregado (Gonzalvo, 1978; Moore, 2005), pues este no es un concepto estadístico, sino más bien un término que han elaborado los educadores estadísticos para describir un fenómeno relacionado con el aprendizaje de la estadística.

Esta investigación sobre el razonamiento estadístico covariacional informal realizada con un grupo de estudiantes entre 16 y 18 años, en la cual se diseñó e implementó una tarea utilizando tecnología (uso de la plataforma CODAP y el applet de GeoGebra) para introducir la recta de mejor ajuste, muestra que hay dos planos en los que conviene analizar la aparición y desarrollo de un conjunto de puntos como un agregado: el plano matemático y el plano del contexto. En el primero, los recursos tecnológicos permiten a los estudiantes transitar de ver a una nube de puntos como puntos individuales, o fragmentado en subconjuntos, a verlo como un agregado a partir de un mecanismo matemático que los une con la noción de distancia de una nube a una recta. En el plano del contexto, en los casos en que subyace en la correlación alguna(s) relación(es) causal(es), es la existencia de un mecanismo causal que ayuda a imaginar una unidad en el conjunto de datos, en estos casos restaría trabajar sobre las fuentes de la variabilidad para que los estudiantes se expliquen por qué los datos difieren del modelo causal probable.

Este trabajo muestra que el diseño de la tarea tuvo un impacto positivo en llevar a los estudiantes a concebir un conjunto de datos bivariados como un agregado, dado que trataron la nube de puntos no como puntos particulares sino a todos a la vez, es decir, concibieron un conjunto de datos como una entidad. La disponibilidad de recursos tecnológicos y la elección de un problema en un contexto accesible hizo posible diseñar una actividad que, sin distraer la atención de los estudiantes en los aspectos técnicos del cálculo, los llevara a volver operativas sus intuiciones de residuos y cercanía mediante el concepto de distancia de la nube a una recta. Con esta herramienta, los estudiantes pueden volver a considerar su interpretación del problema y los datos en contexto y ajustar sus creencias iniciales para conseguir una interpretación coordinada con una apreciación global de los datos.

Creemos que el diseño, los resultados y el análisis de la experiencia que exponemos en este trabajo puede ofrecer una modesta orientación para los profesores de estadística no sólo para realizar actividades con sus estudiantes sobre el tema de correlación y regresión, sino también para diseñar actividades sobre aquellos conceptos en los que la concepción de los datos como agregado es importante (v.gr. medidas de tendencia central, distribución, muestreo). En efecto, el supuesto subyacente es que en cada problema de estos temas se presentan los aspectos contextual y matemático. Para el tratamiento de este, la tecnología puede ofrecer una herramienta útil y, con una elección adecuada de problemas, conducir a los estudiantes a que puedan integrar ambos aspectos.

Finalmente, identificamos ciertas limitaciones en nuestro estudio que podrían ser mejoradas en un estudio posterior. Por ejemplo, considerar las distancias cuadráticas en el diseño de la tarea en lugar de las distancias lineales que se utilizaron; enriquecer los problemas que den lugar a nube de puntos que tengan diferentes tendencias, incluyendo la independencia.

Bibliografía

Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, *3*(2), 64–83.

Bakker, A., y Gravemeijer, K. (2004). Learning to reason about distribution. In D. Ben-Zvi y J. Garfield (Eds.), *The Challenge of developing statistical literacy, reasoning and thinking* (pp. 147–168). Dordrecht, The Netherlands: Kluwer. https://doi.org/10.1007/1-4020-2278-6_7

Batanero, C., Godino, J. D., y Estepa, A. (1998). Building the meaning of statistical association through data analysis activities. In A. Olivier y K. Newstead (Eds.), *Proceedings of the 22 Conference of the International Group for the Psychology of Mathematics Education* (pp. 221–236). South Africa: University of Stellembosch.

Biehler, R., Ben-Zvi, D., Bakker, A., y Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. In M. A. Clement, J. Bishop, C. Keitel, J. Kilpatrick, y A. Y. L. Leung (Eds.), *Third International Handbook of Mathematics Education* (pp. 643–689). New York, NY, United States: Springer. https://doi.org/10.1007/978-1-4614-4684-2_21

Birks, M., y Mills, J. (2015). *Grounded theory: A practical guide (2nd ed)*. Los Ángeles: CA: Sage. https://doi.org/10.13140/2.1.2982.0484

Casey, S. A. (2014). Teacher's knowledge of students' conceptions and their development when learning linear regression. In 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). Voorburg, The Netherlands: 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

Casey, S. A., y Nagle, C. (2016). Students' use of slope conceptualizations when reasoning about the line of best fit. *Educational Studies in Mathematics*, *92*(2), 163–177. https://doi.org/10.1007/s10649-

015-9679-y

Chun Tie, Y., Birks, M., y Francis, K. (2019). Grounded theory research: A design framework for novice researchers. *SAGE Open Medicine*, *7*, 1–8. https://doi.org/10.1177/2050312118822927

Cobb, P., y McClain, K. (2004). Principles of Instructional Design for Supporting the Development of Students' Statistical Reasoning. In Dani Ben-Zvi y J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 375–395). Dordrecht: Springer Netherlands. https://doi.org/10.1007/1-4020-2278-6_16

Cobb, P., Mcclain, K., y Gravemeijer, K. (2003). Learning About Statistical Covariation. *Cognition and Instruction*, *21*, 1–78.

Estepa, A., y Batanero, C. (1996). Judgments of correlation in scatterplots: Students' intuitive strategies and preconceptions. *Hiroshima Journal of Mathematics Education*, *4*, 24–41.

Estepa, A., y Sánchez-Cobo, F. T. (2003). Evaluación de la comprensión de la correlación y regresión a partir de la resolución de problemas [Evaluation of the understanding of correlation and regression through problem solving]. *Statistics Education Research Journal*, *2*(1), 54–68. Retrieved from http://fehps.une.edu.au/serj

Glaser, B. G., y Strauss, A. L. (2008). *The discovery of Grounded Theory: Strategies for qualitative research*. New Brunswick, USA: Aldine Transaction.

Gonzalvo, G. (1978). *Diccionario de metodología estadística*. Madrid: Ediciones Morata.

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

Inzunza, S. (2016). Análisis de datos bivariados en un ambiente basado en applets y software dinámico. *Revista Educación Matemática*, *28*(3), 61–89.

Konold, C., y Higgins, T. (2002). Highlights of related research. In y V. B. S.J. Russell, D. Schifter (Ed.), *Developing mathematical ideas: Working with data* (pp. 165–201). Parsippany, NJ: Dale Seymour Publications. Retrieved from https://srri.umass.edu/publications/konold-2002hrr/

Konold, C., Pollatsek, A., Well, A., y Gagnon, A. (1997). Students analyzing data: Research of critical barriers. In J. Garfield y G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics* (pp. 151–167). Voorburg, The Netherlands: International Statistical Institute.

Martin, G., Carter, J., Forster, S., Howe, R., Kader, G., Kepner, H., … Valdez, P. (2009). *Focus in High School Mathematics : Reasoning and Sense Making in Statistics and Probability*. Reston, Va: National Council of Teachers of Mathematics.

Medina, M., Sanchez, E., y Silvestre, E. (2019). Covariational reasoning patterns exhibited by high school students during the implementation of a hypothetical learning trajectory. In 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)*. Utrecht, Netherlands. Retrieved from https://hal.archives-ouvertes.fr/hal-02412811

Miles, M. B., y Huberman, A. M. (1994). Qualitative data analysis: An expanded sourcebook, 2nd ed. In *Qualitative data analysis: An expanded sourcebook, 2nd ed.* Thousand Oaks, CA, US: Sage Publications, Inc.

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. (1990). Uncertainty. In *On the Shoulders of Giants: New Approaches to Numeracy* (pp. 95–137). USA: National Academy Press.

Moore, D. S. (2005). *Estadística aplicada básica*. Barcelona: Bosh, Antoni.

Moritz, J. (2004). Reasoning about Covariation. In D. Ben-Zvi y J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 227–255). Dordrecht: Springer Netherlands. https://doi.org/10.1007/1-4020-2278-6_10

Pfannkuch, M., y Wild, C. (2004). Towards an understanding of statistical thinking. In D. Ben-Zvi y J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 17–46). Dordrecht: Springer Netherlands. https://doi.org/10.1007/1-4020-2278-6_2

Roseth, C. J., Garfield, J. B., y Ben-Zvi, D. (2008). Collaboration in learning and teaching statistics. *Journal of Statistics Education*, *16*(1). https://doi.org/10.1080/10691898.2008.11889557

Starnes, D. S., Yates, D., y Moore, D. S. (2010). *The practice of statistics* (4th ed.). New York: W.H. Freeman and Company/BFW.

Stigler, S. M. (2016). *The Seven Pillars of Statistical Wisdom.* Cambridge, Massachusett: Harvard University Press.

Wild, C. J., y Seber, G. A. F. (2000). *Chance encounters: A first course in data analysis and inference.* New York: John Wiley.

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.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27*(2), 172–223. https://doi.org/https://doi.org/10.1016/j.dr.2006.12.001

Cindy Nathalia Morgado Hernández. Estudiante de doctorado en Matemática Educativa del Centro de Investigación y de Estudios Avanzados del IPN (CINVESTAV), Ciudad de México. Magister en Educación Matemáticas de la Universidad Industrial de Santander (2013, Colombia), y Licenciada en Matemáticas de la Universidad Industrial de Santander (2011, Colombia).
Email: cindy.morgado@cinvestav.mx.

Ernesto Alonso Sánchez Sánchez. Investigador titular del Departamento de Matemática Educativa del Centro de Investigación y de Estudios Avanzados del IPN (CINVESTAV), Ciudad de México. Doctor en Ciencias, Magister en Ciencias y Licenciado en Matemáticas. Su investigación se centra en la enseñanza y aprendizaje de la Probabilidad y Estadísticas y Nuevas Tecnologías en la Enseñanza de la Matemáticas.
Email: esanchez@cinvestav.mx.