



**CENTRO DE INVESTIGACIÓN Y ESTUDIOS
AVANZADOS DEL INSTITUTO POLITÉCNICO
NACIONAL**

Unidad Zacatenco

Departamento de Matemática Educativa

**EXPERIMENTO DE DISEÑO CON SOPORTE TECNOLÓGICO
PARA EL DESARROLLO DEL RAZONAMIENTO ESTADÍSTICO
COVARIACIONAL DE ESTUDIANTES DE BACHILLERATO**

Tesis que presenta:

M. en C. Cindy Nathalia Morgado Hernández

Para optar al grado de Doctora en Ciencias en la Especialidad de Matemática Educativa

Director de tesis:

Dr. Ernesto Alonso Sánchez Sánchez

México, Ciudad de México.

Marzo, 2023

Especial agradecimiento al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico brindado para la realización de mis estudios de doctorado.

Cindy Nathalia Morgado Hernández

Becario No. 971033

DEDICATORIA

A tu memoria, Marujita

ÍNDICE GENERAL

RESUMEN	12
ABSTRACT	13
INTRODUCCIÓN	14
CAPÍTULO I. UNA REVISIÓN DE LAS INVESTIGACIONES SOBRE LA ENSEÑANZA Y APRENDIZAJE DE LA CORRELACIÓN Y REGRESIÓN LINEAL	18
1.1. Investigaciones acerca del aprendizaje de la correlación y regresión.	19
1.1.1. Concepciones de los estudiantes universitarios después de haber estudiado la correlación y regresión.	19
1.1.2. Concepciones de los estudiantes de secundaria, bachillerato, y de estudiantes universitarios antes del estudiar las temáticas de correlación y regresión.	24
1.1. Diseños de enseñanza y el avance en el aprendizaje de los estudiantes	32
1.1.1. Diseños de enseñanza a nivel universitario	34
1.1.2. Diseños de enseñanza a nivel bachillerato	41
1.1.3. Diseños de enseñanza a nivel primaria y secundaria	47
CAPÍTULO II. MARCO CONCEPTUAL	52
2.1. Contenido: Aspectos matemáticos y estadísticos	54
2.1.1. Aspectos Matemáticos	54
2.1.2. Aspectos Estadísticos	58
2.1.2.1. Definición de agregado en estadística	58
2.1.2.2. Modelo aleatorio lineal	59
2.1.2.3. Incertidumbre	60
2.2. Objeto de estudio: Razonamiento estadístico covariacional	61
2.3. Posición de la enseñanza	64
2.3.1. Uso de la tecnología	64
2.3.2. Resolución de problemas	65
2.3.3. El contexto	66
2.3.4. Trabajo colaborativo	66
CAPÍTULO III. MARCO METODOLÓGICO	68
3.1. Tipo de investigación: El experimento de diseño	69
3.1.1. Ideas estadísticas centrales	70
3.1.2. Uso de la tecnología	73
3.1.3. El Discurso en el aula	74
3.1.4. Las actividades y su estructura	75

3.2. Método de recolección de datos	75
3.2.1. Sujetos	75
3.2.2. Instrumentos y recolección de datos	76
3.2.2.1. Actividad 1.	78
3.2.2.2. Actividad 2.	83
3.2.2.3. Actividad 3.	87
3.3. Metodología para el análisis de resultados: Teoría Fundamentada	91
<i>CAPÍTULO IV. RESULTADOS</i>	95
4.1. Análisis de los datos obtenidos de las actividades	96
4.1.1. Codificación inicial.	96
4.1.2. Codificación intermedia.	99
4.1.2.1. Dimensión covariación.	100
4.1.2.2. Dimensión: Coeficiente de correlación junto con diagramas de dispersión	103
4.1.2.3. Dimensión: recta de mejor ajuste	105
4.2. Análisis de los datos obtenidos del cuestionario diagnóstico y el cuestionario evaluativo.	108
4.2.1. Cuestionario diagnóstico	109
4.2.2. Cuestionario Evaluativo	113
<i>CAPÍTULO V. DISCUSIÓN DE LOS RESULTADOS</i>	117
5.1. Discusión de los resultados en relación con la literatura	118
5.2. Modelos de desarrollo del razonamiento estadístico	124
5.3. Codificación avanzada: Modelos de desarrollo del razonamiento estadístico covariacional	129
5.3.1. Modelo para el razonamiento cuando se juzga la relación entre dos variables.	130
5.3.2. Modelo para el razonamiento cuando se analiza el coeficiente de correlación y el diagrama de dispersión.	131
5.3.3. Modelo para el razonamiento al establecer la recta de mejor ajuste.	132
5.3.4. Modelo para el razonamiento al estimar y predecir	134
<i>CAPÍTULO VI. CONCLUSIONES</i>	135
6.1. El papel de la tecnología en el desarrollo del razonamiento sobre la regresión y la correlación	136
6.2. La transición de ver los datos individuales a verlos como un agregado con la correlación y regresión lineal	138
6.3. De acercamientos determinísticos al reconocimiento de la incertidumbre	139
6.4. Respuesta a cada pregunta de investigación.	141

REFERENCIAS BIBLIOGRÁFICAS	146
Apéndice A: Método de búsqueda bibliográfica	153
Apéndice B: Actividad 1	158
Apéndice C: Actividad 2	164
Apéndice D: Actividad 3	168
Apéndice E: Cuestionario diagnóstico	173
Apéndice F: Cuestionario evaluativo	179
Apéndice G: Codificación inicial de las actividades	183
Codificación inicial de la actividad 1.	183
Codificación inicial de la actividad 2	190
Codificación inicial de actividad 3	203
Apéndice H: Codificación intermedia de las actividades	211
Apéndice I: Codificación del cuestionario diagnóstico y evaluativo	218
Codificación Inicial del cuestionario diagnóstico	218
Componente de álgebra básica	218
Componente estadística	221
Codificación Inicial del cuestionario evaluativo	229

LISTA DE TABLAS

Tabla 1.1. Estudios a nivel universitario sobre las concepciones después de la enseñanza entre el 1988 y 2019 según edad de los estudiantes y temática abordada.	20
Tabla 1.2. Estudios sobre las concepciones antes de la enseñanza del tema entre el 1988 y 2019 según el nivel escolar, edad de los estudiantes y temática abordada.....	24
Tabla 1.3. Estudios que realizaron intervención de enseñanza utilizando tecnología entre el 1988 y 2019 según el nivel escolar, edad de los estudiantes, temática y software utilizado	33
Tabla 1.4. Estudios de intervención de enseñanza entre el 1988 y 2019 a nivel universitario, según la metodología utilizada para el diseño, la teoría empleada para el análisis.....	34
Tabla 1.5. Niveles de modelado SOLO en la interpretación de diagrama de dispersión establecido por Inzunza (2016)	39
Tabla 1.6. Estudios de intervención de enseñanza entre el 1988 y 2019 a nivel bachillerato, según la metodología utilizada para el diseño, la teoría empleada para el análisis y software utilizado.	41
Tabla 1.7. THA diseñada e implementada por Gil y Gibbs (2017).....	44
Tabla 1.8. Resumen de la THA diseñada e implementada por Dierdorp et al., (2011).....	45
Tabla 1.9. Resume de las nociones estadísticas que abordó los estudios de enseñanza a nivel de bachillerato	47
Tabla 1.10. Estudios de intervención de enseñanza entre el 1988 y 2019 a nivel primaria, según la metodología utilizada para el diseño, la teoría empleada para el análisis y software utilizado	48
Tabla 3.1. Grasa y Calorías de algunos productos de McDonalds y Burger King	79
Tabla 3.2. Mediciones de grasa, carbohidratos y azúcar en los productos de McDonalds y Burger King.....	84

Tabla 3.3. Medidas de cambio en ANE y aumento de grasa en los 12 adultos jóvenes.....	89
Tabla 4.1. Ejemplo de palabras clave o frases para los códigos.....	98
Tabla 4.2. Ejemplo de tabla de frecuencia para cada código.	98
Tabla 4.3. Códigos y frecuencia para los razonamientos al juzgar la relación entre las variables.....	100
Tabla 4.4. Códigos y frecuencia para los razonamientos al analizar correlación junto con diagramas de dispersión	103
Tabla 4.5. Códigos y frecuencia para los razonamientos al establecer la recta de mejor ajuste	106
Tabla 4.6. Frecuencia para cada código del componente álgebra básica	109
Tabla 4.7. Frecuencia para cada código del componente evaluativo de correlación y regresión	110
Tabla 4.8. Frecuencia para cada código de la componente estadística del cuestionario evaluativo.	113
Tabla 5.1. Estudios previos de Modelo de desarrollo para el razonamiento covariacional	126
Tabla 5.2. Marco para analizar e interpretar datos a nivel de secundaria propuesto por Jones et al. (2000) y Mooney (2002).....	127
Tabla 5.3. Modelo para el razonamiento cuando se juzga la relación entre dos variables.	130
Tabla 5.4. Modelo para el razonamiento cuando se analiza la fuerza entre las variables.	132
Tabla 5.5. Modelo para el razonamiento al establecer la recta de mejor ajuste.	133
Tabla 5.6. Modelo para el razonamiento al estimar y predecir.	134

LISTA DE FIGURAS

Figura 1.1. Secuencias de enseñanza propuesta por Zieffler y Garfield (2009)	38
Figura 2.1. Marco conceptual de la investigación (elaboración propia)	53
Figura 2.2. Tipos de fuerza de la relación lineal	55
Figura 2.3. Tipos de dirección de la relación lineal.	55
Figura 2.4. Métrica de los mínimos cuadrados	56
Figura 2.5. Componentes de la covariación estadística.....	60
Figura 2.6. Modelo lineal simple (distribución de y para un valor dado de x)	60
Figura 3.1. Secuencia de los instrumentos y duración de la aplicación.	77
Figura 3.2. Ejemplo de recuadro con la definición de variable respuesta y variable explicativa.....	77
Figura 3.3. Recuadro variable respuesta y variable explicativa.	80
Figura 3.4. Diagrama de Dispersión de la Grasa y Calorías	81
Figura 3.5. Captura de pantalla del diagrama de dispersión y la recta de regresión realizado en CODAP	83
Figura 3.6. Recuadro definición de coeficiente de correlación	85
Figura 3.7. Captura de pantalla de los diagramas de dispersión realizados en CODAP....	86
Figura 3.8. Captura de pantalla para determinar la correlación numérica en CODAP	87
Figura 3.9. Captura de pantalla del archivo de GeoGebra ANE_Grasas.	90
Figura 3.10. Captura de pantalla del diagrama de dispersión realizado en CODAP	91

Figura 3.11. Marco de Procesos y Métodos de la Teoría Fundamentada (Adaptación de Chun Tie et al., 2019).....	92
Figura 4.1. Ejemplo de generación de la etiqueta distancia.	97
Figura 4.2. Dimensiones y categorías que emergieron de la codificación inicial e intermedia	100
Figura 6.1. Componentes de la covariación estadística.....	121

LISTA DE APÉNDICES

Apéndice A: Método de búsqueda bibliográfica	146
Apéndice B: Actividad 1	158
Apéndice C: Actividad 2	164
Apéndice D: Actividad 3	168
Apéndice E: Cuestionario diagnóstico	173
Apéndice F: Cuestionario evaluativo	179
Apéndice G: Codificación inicial de las actividades	183
Apéndice H: Codificación intermedia de las actividades	211
Apéndice I: Codificación del cuestionario diagnóstico y evaluativo	218

RESUMEN

En esta investigación nos preguntamos: ¿Cómo cambian los razonamientos que muestran estudiantes de bachillerato frente a tareas de correlación y recta de mejor ajuste antes y después de la implementación de tareas instruccionales apoyados con tecnología digital?, tal pregunta resulta conveniente ya que el razonamiento estadístico covariacional ha sido identificado como importante para los estudiantes de bachillerato, por lo tanto, son necesarios estudios educativos para documentar la forma en que los estudiantes progresan en el uso de herramientas estadísticas, apoyadas con tecnología digital, en el desarrollo del razonamiento covariacional, en particular en los conceptos de correlación y recta de mejor ajuste. El estudio que se presenta consiste en un experimento de diseño, cuyo rasgo importante es el diseño e implementación de tareas instruccionales basadas en la resolución de problemas utilizando tecnología (plataforma online CODAP y un *applet* de GeoGebra) con el propósito de promover el razonamiento covariacional en estudiantes de bachillerato. Para el experimento de diseño se considera el marco de los "principios del diseño instruccional" de Cobb y McClain (2004) que comprende: el enfoque sobre ideas estadísticas centrales, el discurso en la clase, el uso de herramientas computacionales, las actividades y su estructura en el salón de clase. La experimentación fue realizada con 22 estudiantes entre 16 y 18 años. Se emplearon los principios y métodos de la Teoría Fundamentada para el análisis de datos. Como resultado se reportan cuatro modelos de desarrollo del razonamiento estadístico covariacional (covariación, recta de mejor ajuste, coeficiente de correlación y diagramas de dispersión). Además, se identificaron los progresos y limitaciones de las conceptualizaciones que pueden llevar a cabo los estudiantes dentro de las actividades de instrucción en función de los elementos del experimento de diseño (el uso de tecnología y el tipo de problemas), estos resultados se exponen en formato de ensayo, y se titulan: i) El papel de la tecnología en el desarrollo del razonamiento sobre la regresión y la correlación. ii) La transición de ver un conjunto de datos singulares a verlo como un agregado. iii) De acercamientos determinísticos al reconocimiento de la incertidumbre.

Finalmente, identificamos ciertas limitaciones en nuestro estudio que podrían ser mejoradas en un estudio posterior. Así, como las implicaciones para la didáctica de la estadística y prospectiva de la investigación.

ABSTRACT

In this research we ask: how do high school students' reasoning change when faced with correlation and line of best fit tasks before and after the implementation of instructional tasks supported with digital technology? This question is convenient since covariate statistical reasoning has been identified as important for high school students, therefore, educational studies are needed to document how students' progress in the use of statistical tools, supported with digital technology, in the development of covariate reasoning, in the concepts of correlation and line of best fit. The presented study consists of a design experiment, whose important feature is the design and implementation of instructional tasks based on problem solving using technology (CODAP online platform and a GeoGebra applet) with the purpose of promoting covariate reasoning in high school students. For the design experiment, the framework of the "principles of instructional design" of Cobb and McClain (2004) is considered, which includes: the focus on central statistical ideas, the discourse in the classroom, the use of computational tools, the activities, and their structure in the classroom. The experimentation was conducted with 22 students between 16 and 18 years of age. Grounded Theory principles and methods were used for data analysis. As a result, four models of development of covariate statistical reasoning (covariation, line of best fit, correlation coefficient and dispersion diagrams) are reported. In addition, we identified the progress and limitations of the conceptualizations that students can carry out within the instructional activities as a function of the elements of the design experiment (the use of technology and the type of problems), these results are presented in essay format, and are titled: i) The role of technology in the development of reasoning about regression and correlation. ii) The transition from viewing a singular data set to viewing it as an aggregate. iii) From deterministic approaches to the recognition of uncertainty.

Finally, we identified certain limitations in our study that could be improved in a subsequent study. As well as implications for the didactics of statistics and prospective research.

INTRODUCCIÓN

Esta investigación se refiere a un experimento de diseño, cuyo rasgo importante es el diseño e implementación de tareas instruccionales basadas en la resolución de problemas utilizando tecnología con el propósito de promover el razonamiento acerca de la covariación estadística en estudiantes de bachillerato.

La *covariación estadística* es la correspondencia de variación conjunta de dos variables estadísticas que toman valores numéricos (Moritz, 2004). Los valores para cada variable se obtienen de la misma unidad de observación y son expresados como pares ordenados que componen un conjunto llamado datos bivariados. Las técnicas más utilizadas para investigar la covariación estadística son la correlación y la regresión. La correlación cuantifica la fuerza de la relación lineal entre un par de variables, mientras que la regresión expresa la relación en forma de un modelo matemático (ecuación lineal). El razonamiento sobre la covariación estadística es un rasgo fundamental del pensamiento científico, por lo que Zimmerman (2007) en su estudio sobre el desarrollo de las habilidades científicas de niños y adolescentes, destaca el razonamiento covariacional como un pilar para la formación científica.

El *razonamiento estadístico covariacional*, es el “razonamiento sobre datos bivariados [e] implica saber cómo juzgar e interpretar una relación entre dos variables” (Zieffler y Garfield, 2009, p. 7). En otras palabras, lo constituyen los procesos mediante los cuales se formulan y sustentan proposiciones referentes al comportamiento de dos variables que toman valores afectados por grados de incertidumbre, y a las relaciones entre ellas. El razonamiento covariacional involucra llevar a cabo juicios sobre correlación a partir de un conjunto de datos bivariados y un diagrama de dispersión para examinar el patrón de los puntos, observar su dirección y su forma, y si el patrón de puntos se acerca cada vez más a una línea recta, sugiere que la relación entre las variables se puede representar mediante un modelo matemático lineal. También incluye la determinación e interpretación numérica del coeficiente de correlación, establecer la recta de mejor ajuste a los datos y su expresión algebraica, así como la habilidad de utilizar la recta de regresión para hacer estimaciones y predicciones.

Algunos estudios de investigaciones didácticas sobre los temas de correlación y regresión han sido de naturaleza cognitiva asociadas al razonamiento. En efecto, los primeros estudios fueron acerca de las concepciones de los estudiantes, como el de Truran (1995) quien se interesó por la detección y caracterización de las interpretaciones de los estudiantes universitarios del coeficiente de correlación y el coeficiente de determinación. Sánchez-Cobo et al. (2000) estudiaron las representaciones verbales, gráficas y numéricas de la correlación. Sorto et al. (2011) estudiaron las concepciones de los estudiantes universitarios acerca de la línea de mejor ajuste cuando se les pide dibujarla en diagramas de dispersión. También se han generado estudios interesados por conocer las formas de razonamiento de los estudiantes de niveles escolares anteriores al universitario, como el estudio de Watson y Moritz (2007) y Moritz (2004) a nivel de secundaria, quienes analizaron el razonamiento de los estudiantes al realizar representaciones gráficas sobre la covariación, y el estudio de Casey (2014, 2015) sobre las concepciones que poseen los estudiantes acerca de lo que es la línea de mejor ajuste. Respecto a las concepciones de los estudiantes de bachillerato se determinó dos estudios relacionados con la representación gráfica de la covariación; el estudio de Watson y Moritz (1997), quienes analizaron la representación gráfica que establecían los estudiantes sobre la covariación en contextos no simbólicos, y el estudio de Estepa y Batanero (1996) quienes establecieron tres concepciones (determinística, local y causal) de la covariación en los estudiantes cuando juzgan la relación de dos variables con base en los diagramas de dispersión.

Una tendencia actual en educación estadística es realizar investigaciones en las que indague sobre las relaciones entre el diseño de la enseñanza y el avance en el aprendizaje de los estudiantes, se ha comenzado a llevar a cabo estudios en los que el diseño de la intervención en el aula es una componente importante. Por ello, surgieron los primeros experimentos de enseñanza (Batanero et al., 1998; Batanero et al., 1997) a nivel universitario y fueron orientados a la resolución de problemas de correlación y regresión lineal utilizando un *software* estadístico, e investigaciones de diseño (Inzunza y Ward, 2015; Inzunza, 2016) basadas en el análisis de datos reales multivariados mediante el uso de tecnología computacional para el análisis de correlación y regresión de datos bivariados. A *nivel bachillerato* se han ubicado 3 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).

Es de notar que todas investigaciones que han realizado intervención en el aula utilizaron en su diseño de enseñanza tecnología digital, sugiriendo que puede ayudar a los estudiantes a hacer juicios de covariación más precisos (Batanero et al., 1998; Cobb, McClain, et al., 2003; Inzunza, 2016), de hecho, la tecnología posee un papel muy importante en las estadísticas, dado que hace que sean visuales, interactivas y dinámicas, permitiendo 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 permite experimentar con los datos y enganchar a los estudiantes (Biehler et al., 2013).

En lo concerniente al tema de regresión y correlación sigue vigente el problema que ha formulado Shaughnessy (2007) acerca de la necesidad de documentar *experiencias de enseñanza* de estadística para mostrar lo que los estudiantes son capaces de hacer, y si es posible, cómo superar las concepciones erróneas y dificultades que enfrentan y que han sido caracterizadas por la investigación. Zimmerman (2007) destaca la importancia de enfrentar a los estudiantes a situaciones de covariación que conlleve utilizar razonamientos basados en la evidencia de los datos (Kuhn, 1993), donde las teorías que los estudiantes poseen se contradigan con lo mostrado en las situaciones que enfrentan. El llamado de Shaughnessy y la sugerencia de Zimmermann se vuelven más importantes si se tienen en cuenta las nuevas posibilidades que para la enseñanza de las matemáticas y la estadística ofrece la tecnología digital, en particular, para el tema de regresión y correlación (Biehler et al., 2013).

Específicamente, falta avanzar en entender cómo son los cambios en el razonamiento covariacional de estudiantes de bachillerato cuando se implementan en el aula tareas instruccionales y explotando las posibilidades que ofrece la tecnología.

En consecuencia, con base en las siguientes tres hipótesis:

- Es importante llevar a cabo experimentos de diseño con estudiantes de nivel bachillerato con el objetivo de entender mejor cómo y en qué medida se desarrolla su razonamiento covariacional en función de los elementos de diseño.

- En el diseño instruccional es necesario tener en cuenta los conocimientos establecidos sobre las concepciones y dificultades de los estudiantes, relacionados con el desarrollo del razonamiento covariacional.
- Un elemento de diseño se relaciona con tener en cuenta el papel que puede jugar el uso de recursos tecnológicos actuales que han sido elaborados expresamente para la enseñanza de la estadística y que, en particular, pueden ser útiles para el desarrollo del razonamiento covariacional.

Se establecieron las siguientes preguntas que guían nuestro estudio y son respondidas en las diferentes etapas de la investigación:

1. ¿Cómo cambian los razonamientos que muestran los estudiantes frente a tareas de correlación y recta de mejor ajuste antes y después de la implementación de las tareas instruccionales?
2. ¿Cómo se relacionan los cambios en los razonamientos de los estudiantes con los elementos del diseño, en especial, con las características de los problemas y la herramienta computacional?
3. ¿Qué concepciones y dificultades se revelan en las respuestas de los estudiantes cuando abordan las tareas instruccionales?

CAPÍTULO I. UNA REVISIÓN DE LAS INVESTIGACIONES SOBRE LA
ENSEÑANZA Y APRENDIZAJE DE LA CORRELACIÓN Y REGRESIÓN
LINEAL

En este capítulo se expone una revisión de los reportes de investigación empíricos sobre el razonamiento estadístico covariacional, publicados en revistas importantes de educación matemática y estadística en un período aproximado de hace 30 años a la fecha. La estrategia diseñada para la búsqueda bibliográfica se describe en el apéndice A. La intención es proporcionar un informe más detallado de los estudios que justifican el problema de investigación y brindar una visión actualizada de las investigaciones acerca de la enseñanza y aprendizaje de la correlación y regresión dentro del marco de la educación estadística.

Este capítulo contiene dos apartados: 1. los estudios sobre el aprendizaje de los estudiantes acerca de la correlación y regresión, y 2. las investigaciones que han realizado algún tipo de diseño de enseñanza con el objetivo de promover el desarrollo del razonamiento covariacional.

1.1. Investigaciones acerca del aprendizaje de la correlación y regresión.

Confrey (1990) menciona que las ideas sobre el mundo, las creencias, teorías, el desarrollo de significados, explicaciones y estrategias de los niños son la base del uso del término concepciones del estudiante, y cuando estas concepciones están en conflicto con los significados aceptados en ciencias o en las matemáticas se denominan conceptos erróneos. Las investigaciones sobre las concepciones de los estudiantes acerca de la correlación y regresión podrían tomar dos caminos: i) el cambio o resistencia de las concepciones después de la instrucción tradicional o con uso de herramientas tecnológicas a nivel universitario, ii) las concepciones acerca de la correlación y regresión en edades tempranas (secundaria y bachillerato) y antes del estudio formal de la correlación y regresión a nivel preuniversitario. Los anteriores aspectos comprenden el marco de análisis para la revisión de los estudios de aprendizaje. El análisis y discusión se realiza por niveles escolares y según la temática con el fin de proporcionar un marco organizativo que pueda ser útil para el diseño y desarrollo de esta investigación o para futuras investigaciones.

1.1.1. Concepciones de los estudiantes universitarios después de haber estudiado la correlación y regresión.

Cronológicamente, las primeras investigaciones se dieron a nivel universitario, su interés estaba en conocer cuáles eran las concepciones que tenían los estudiantes al finalizar un

curso de estadística, pues la correlación y regresión son temáticas que hacen parte del plan de estudios en los primeros niveles de las carreras universitarias, y son requisito previo para comprender conceptos y procedimientos estadísticos posteriores, como la regresión múltiple y diversos métodos multivariantes (Sánchez-Cobo et al., 2000). La tabla 1.1 muestra los estudios cuyo interés fueron las concepciones de los estudiantes después de finalizar su primer curso de estadística universitario. Las investigaciones están en orden cronológico. Se identificaron 5 estudios, de los cuales 2 se enfocaron en las concepciones de los estudiantes sobre el coeficiente de correlación, 1 estudio sobre la interpretación gráfica, verbal y numérica de la correlación, y 2 investigaciones enfocadas en las dos temáticas (correlación y regresión).

Tabla 1.1. *Estudios a nivel universitario sobre las concepciones después de la enseñanza entre el 1988 y 2019 según edad de los estudiantes y temática abordada.*

Autores	Nivel	Edad	Temática
Truran (1995)	Universitario		Coefficiente de correlación
Sánchez-Cobo, Estepa y Batanero (2000)	Universitario	20 años	Interpretación gráfica, verbal y numérica de la correlación
Estepa y Sánchez Cobo (2003)	Universitario	20 años	Correlación y regresión
Delgado y Ojeda Salazar (2007)	Universitario	-	Correlación y regresión
Castro-Sotos, Van Hoof, Van den Noortgate y Onghena (2009)	Universitario	-	Coefficiente de correlación

El *coeficiente de correlación* fue el parámetro estadístico que Truran (1995) escogió como objeto de estudio, su interés estaba en conocer cómo los estudiantes universitarios interpretan este parámetro estadístico, pues desde su experiencia argumentaba que las escuelas promovían un enfoque determinista en los estudiantes bajo una enseñanza enfocada en el cálculo relevante de parámetros.

Truran (1995) evidenció la creencia de los estudiantes de que la correlación mide la causa, en lugar de asociación. Además, describe las interpretaciones que los estudiantes dan a r (coeficiente de correlación de Pearson) y r^2 , algunas de estas interpretaciones se explican por una falta de comprensión de la idea subyacente y otras, que llamó aberrantes, por estar basadas en asociaciones a ideas extrañas.

- Interpretaciones aberrantes de r : equivalente a la pendiente de la recta de regresión, como la suma mínima de desviaciones al cuadrado de la línea de regresión, una alta correlación que indica una gran desviación de la línea de regresión. Las respuestas fueron bastante deterministas.
- Mal entendimiento de la relevancia del r^2 : Creen que fue una medida de cuánto podría explicarse el cambio en una variable por el cambio en la otra, creían que era una medida de cuánta variación podía explicarse o que era una medida de cuánto podría explicarse la cantidad de una variable por la cantidad de la otra y que creía que era una medida tanto de la varianza como de la variable.
- Interpretaciones aberrantes de r^2 : Como la pendiente de la recta de regresión o como una medida probabilística.

Truran (1995) sugiere que se debe enseñar explícitamente a los estudiantes que algunos parámetros son capaces de proporcionar una variedad de tipos de información, y qué cuestiones no se deben asumir respecto a estos parámetros. Por ejemplo, una propiedad que no se debe asumir para el coeficiente de correlación es la propiedad transitiva, la cual no se cumple y desde el punto de vista matemático y estadístico se encuentran explicaciones de por qué esta propiedad no se cumple. El estudio de Castro-Sotos et al. (2009) muestra que esta falsa concepción de asumir que si dos variables X y Z se correlacionan positivamente con una tercera variable Y , entonces X y Z tienen correlación positiva, lo cual no siempre es cierto, esta concepción se presentó frecuentemente en los estudiantes de diferentes edades, sexo y año escolar. Sin embargo, el contexto del problema provocaba el uso de la transitividad con menor frecuencia, por lo que afirman que el contexto parece ayudar a los estudiantes a evitar la idea errónea, pero sugieren que se necesita más investigación para explorar el efecto del contexto sobre la aparición de la concepción.

Interpretación verbal, gráfica y numérica de la correlación fue objeto de estudio por Sánchez-Cobo et al. (2000), quienes aplican un cuestionario con el fin de caracterizar el conocimiento de los estudiantes al finalizar la enseñanza sobre correlación en los primeros cursos universitarios. Las tareas del cuestionario son: dibujar diagramas de dispersión con determinadas características (ejemplo, el número de pintores pintando una habitación y tiempo en horas para acabar el trabajo), indicar el valor del coeficiente de correlación para distintas situaciones (ejemplo, el $r = ?$ para la ordenación por grado de timidez de los estudiantes de la Universidad de Jaén y ordenación por número de ciudades diferentes que han visitado), también se presenta varios diagramas de dispersión donde se debe estimar el valor de su coeficiente de correlación, y se debe describir dos variables razonables para determinados valores de coeficientes de correlación. Dentro de sus resultados describen las siguientes estrategias utilizadas por los estudiantes:

- Hacen referencia a alguna característica del diagrama de dispersión o lo dibujan (ejemplo, mencionan función decreciente, no es lineal).
- Se apoyan en algún hecho del marco numérico implícito (ejemplo, para un $r = 0,8$ se tiene relación una variable con otra, a medida que aumenta una variable, la otra aumenta con un incremento cada vez menor)
- Utilizan el marco numérico y gráfico en forma conjunta (ejemplo, afirman lineal negativa, relación indirecta decreciente, ofrecen información redundante).
- Sustentan sus argumentos en las teorías previas que poseen sobre el contexto del problema planteado (ejemplo, el tiempo depende de la memoria de cada persona y de la concentración que cada uno ponga)
- Usan otros conceptos como la proporcionalidad o la dependencia funcional. (ejemplo, estas variables tienen relación funcional, por tanto, $r = 1$).

Respecto a la estimación del coeficiente de correlación mencionan que es más precisa a partir de un diagrama de dispersión. Los errores son mayores al construir una nube de puntos a partir de una descripción verbal y estimar el coeficiente de correlación desde una tabla de valores numéricos.

Estepa y Sánchez-Cobo (2003) evalúan la **correlación y regresión** a través de resolución de problemas, y utilizando técnicas de análisis cualitativo recomendadas en Miles y Huberman (1994) con el propósito de obtener distintas categorías de las respuestas de los estudiantes y clasificarlas para identificar las estrategias. A continuación, se describe el tipo de problema que propusieron junto con las estrategias realizrealizadas.

- En el problema donde una recta de regresión tiene una pendiente de 16 y corta al eje de ordenadas en el punto $y = 4$. Si la media de la variable independiente es 8, ¿Cuál es la media de la variable dependiente?: Se evidenció la dificultad de discriminar entre la variable explicativa y la variable explicada. Además, los estudiantes utilizan otros parámetros como la media y la covarianza, siendo aplicados de manera inadecuada.
- En el problema donde se muestra una tabla de datos bivariados y se pregunta sobre el tipo de relación: Se evidenció que los estudiantes conocen el algoritmo del coeficiente de correlación, pero no lo realizan correctamente, tienen ciertos errores debidos al cálculo y son pocos los errores debidos a conocimiento incorrecto sobre la correlación.
- En el problema de estimación de un valor para la variable dependiente, y predicción para la variable independiente: utilizaron el coeficiente de correlación, la variación conjunta de las variables, uso de la covarianza, uso del coeficiente de determinación, utilizaron un intervalo de variación del coeficiente de correlación, usaron el diagrama de dispersión o compararon la asociación y la proporcionalidad.

Delgado y Ojeda (2007) realizan un estudio exploratorio en el cual consideran los registros semióticos de Duval (1998) y los tipos de conocimiento de Pollatsek et al. (1981) para analizar las respuestas a tres problemas. El primer problema correspondía a un conjunto de datos bivariados donde se pedía realizar la gráfica, explicar el tipo de relación (casual o no), realizar una estimación y una predicción: las respuestas de los estudiantes mostraron un conocimiento funcional acerca de la correlación y regresión. En el segundo problema se daba información en un gráfico de dispersión para contestar una serie de preguntas relacionadas con la relación entre las variables: las respuestas mostraron una conversión entre el registro

gráfico y la lengua natural. El tercer problema correspondía a un problema del valor esperado de la media de una variable utilizando tecnología para determinar la respuesta: se evidenció la comprensión del significado de la regresión a la media en relación con el contexto del problema.

1.1.2. Concepciones de los estudiantes de secundaria, bachillerato, y de estudiantes universitarios antes del estudiar las temáticas de correlación y regresión.

La tabla 1.2 muestra los estudios que se interesaron por las concepciones de los estudiantes a edades tempranas y a nivel preuniversitario. Las investigaciones están ordenadas por el nivel escolar de los estudiantes (secundaria, bachillerato y preuniversitario), seguido de la edad y el tema específico que fue objeto de estudio. Se identificó un estudio que comprendió estudiantes de secundaria y bachillerato, 7 estudios a nivel de secundaria, de los cuales 2 se enfocaron en las concepciones de los estudiantes cuando se les pide representar gráficamente dos variables cuantitativas, y 5 centradas en las concepciones respecto a la línea que mejor se ajuste a los datos. A nivel universitario, se determinaron 2 estudios, uno de ellos respecto a la línea de mejor ajuste y otro acerca de las concepciones al determinar la expresión algebraica de la ecuación que representa la regresión lineal.

Tabla 1.2. Estudios sobre las concepciones antes de la enseñanza del tema entre el 1988 y 2019 según el nivel escolar, edad de los estudiantes y temática abordada.

Autores	Nivel	Edad	Temática
Watson y Moritz (1997)	Secundaria y Bachillerato	-	Representación gráfica de la covariación
Moritz (2004)	Secundaria	-	Covariación
Watson y Moritz (2007)	Secundaria	-	Representación gráfica de la covariación
Casey (2014)	Secundaria	11 y 14 años	Línea de mejor ajuste
Casey (2015)	Secundaria	11 y 14 años	Línea de mejor ajuste
Casey y Nagle (2016)	Secundaria	11 y 14 años	Línea de mejor Ajuste
Nagle, Casey, Moore y Russo (2017)	Secundaria	11 y 14 años	Línea de mejor ajuste

García y Yáñez (2018)	Secundaria	12 y 13 años	Línea de mejor ajuste
Estepa y Batanero (1996)	Bachillerato	18 años	Diagramas de dispersión
Sorto, White y Lesser (2011)	Universitario	-	Línea de mejor ajuste
Yáñez, Rátiva y Morgado (2016)	Universitario	-	Ecuación de la regresión lineal

La **representación gráfica de dos variables cuantitativas** que realizan los estudiantes de grado 5, 6, 7, 8, grado 9 y 10 de secundaria y bachillerato fue el interés de Watson y Moritz (1997, 2007) y Moritz (2004); quienes querían conocer a partir de la descripción verbal de una covariación entre dos variables ¿cómo los estudiantes la representan gráficamente? ¿Cómo coordinan las variables involucradas? Además, querían conocer si se produce el desarrollo de la comprensión de los estudiantes durante la escolarización, es decir, estudiar el desarrollo o génesis de los conceptos a lo largo del tiempo, a esto Confrey (1990) le llama epistemología genética. Cabe mencionar que en los estudios de estos investigadores también se consideró la interpretación gráfica de datos cualitativos, pero esta revisión se centra en datos bivariados cuantitativos, por lo tanto, se discutirán sus hallazgos referentes a los ítems que involucraron variables cuantitativas.

Los cuestionarios aplicados a los estudiantes consistieron en fragmentos tomados de un periódico que los estudiantes debían representar gráficamente; algunos ítems fueron cambiando durante las investigaciones y otros se mantuvieron. En el primer estudio de Watson y Moritz (1997) un cuestionario fue aplicado a 1291 estudiantes; Moritz (2004) aplicó una encuesta a 167 estudiantes y Watson y Moritz (2007) aplicaron un cuestionario a 1285 estudiantes. En los primeros estudios las respuestas se analizaron mediante la técnica iterativa de Miles y Huberman (1994), refinando categorías y subcategorías por medio de la comparación y contraste de características de los gráficos en las respuestas escritas. Su último estudio actualizó las categorías determinadas en sus anteriores investigaciones bajo el marco de evaluación de Biggs y Collis (1982) Structure of Observed Learning Outcomes (SOLO) para el análisis de las respuestas. Este marco postula que el aprendizaje de las

personas ocurre en uno de cinco modos de funcionamiento: sensoriomotor, icónico, simbólico concreto, formal y postformal. Estos son similares a las etapas de desarrollo cognitivo de Piaget, pero hay algunas diferencias entre ambos. Dentro de cada modo las respuestas de los estudiantes se vuelven cada vez más complejas y esta sofisticación se describe en términos de niveles dentro de cada modo: Preestructural, Uniestructural, Multiestructural y Relacional. En los estudios realizados sobre datos bivariados y sus respectivos diagramas de dispersión, Watson y Moritz (2007) describieron estos niveles SOLO con los rasgos que se resumen a continuación:

- Preestructural: Realizan un dibujo del enunciado o realizan una gráfica, pero sin mostrar ningún valor, ni relación. Describen algún tipo de variación en el gráfico, pero no son capaces de identificarlo de manera significativa al mostrar una tendencia o un sentido de las variables involucradas.
- Uniestructural: Muestran variaciones para producir una tendencia, pero no hay indicación de las medidas que varían.
- Multiestructural: Consideran una variable respecto al tiempo o cada variable la relacionan con respecto al tiempo.
- Relacional: Realizan comparaciones relacionando las dos variables de varias maneras: a través de gráficos de barras o de líneas, las dos variables en el mismo gráfico con respecto al tiempo, gráficos bivariados con líneas de tendencia.

Watson y Moritz (1997, 2007) y Moritz (2004) mencionan en sus conclusiones dos aspectos muy importantes respecto al cambio o resistencia de las concepciones a través de la instrucción. El primer aspecto se refiere a que los estudiantes avanzados realizan gráficos que muestran variación apropiada asociada a los contextos, pero es poco probable que usen la palabra “variación”. El segundo aspecto, es la dependencia del contexto del problema, donde es difícil para los estudiantes separar sus creencias a priori temporalmente para observar los datos y evaluar la correlación; los autores sugieren que los estudiantes deben experimentar muchos contextos diferentes para explorar la variación.

La **interpretación de los diagramas de dispersión** realizada por estudiantes de bachillerato fue el interés de Estepa y Batanero (1996), quienes aplicaron un cuestionario a estudiantes de último grado de bachillerato antes de que iniciaran su primer curso de estadística y

probabilidad. El cuestionario constó de tres ítems donde se mostraba tres diagramas de dispersión diferentes, cada uno con un caso para las tres posibilidades del signo de la correlación (directa, inversa, e independencia) para la relación de dos variables, y tres valores diferentes para el coeficiente de correlación (0.55, -0.77 y 0.11). Los investigadores determinaron dos estrategias correctas en las respuestas de los estudiantes: la comparación global de la relación entre las dos variables y la estrategia de crecimiento que consiste en utilizar el incremento, decrecimiento o la forma constante del diagrama de dispersión para justificar el tipo de dependencia. Además, establecieron las siguientes concepciones:

- Concepción determinista de la correlación: Esperan una correspondencia donde a cada valor de la variable dependiente se le asigne otro valor de la variable independiente, cuando no es así, ellos consideran que no hay dependencia entre las variables.
- Concepción local de la correlación: Basan su juicio en solo una parte de los datos que provee el diagrama de dispersión.
- Concepción causal de la correlación: Cuando identifican correlación y causalidad.

La **noción de la línea de mejor ajuste** significa dibujar una línea que se acerque lo más posible a los datos bivariados que están representados en el diagrama de dispersión, esta estrategia se deriva del método por mínimos cuadrados, ya que el método determina la línea recta que hace que la suma de los cuadrados de las distancias verticales entre los valores observados y los valores predichos sea la más pequeña posible.

La investigación de Sorto et al. (2011) con 18 estudiantes universitarios fue uno de los primeros estudios sobre las concepciones de los estudiantes acerca de la interpretación de la línea de mejor ajuste cuando se les pide a lápiz y papel dibujarla en diferentes diagramas de dispersión. Los investigadores categorizaron las respuestas de los estudiantes y establecieron las siguientes categorías:

- i. Generalización del concepto de Mediana: Trazar una línea que divide los puntos de datos a la mitad en el sentido de que la mitad de los puntos están sobre la línea y la mitad están debajo de la línea.

- ii. Generalización del concepto de media: Trazar la línea a través de puntos que son puntos medios de diferentes grupos de puntos.
- iii. Ajustar la tendencia de los datos: Trazar la línea que une el primer y último punto y trazar "la línea está más estrechamente relacionada con todos los puntos y la tendencia de los datos".

Los estudios sobre la recta de mejor ajuste comenzaron a explorar a nivel medio básico (11-14 años), cuestionando a los estudiantes respecto a lo que ellos conciben puede ser la línea de mejor ajuste y el método que utilizaban para encontrar la línea de mejor ajuste. Las investigaciones de Casey (2015, 2014) refinan las categorías determinadas por Sorto et al. (2011) utilizando la teoría fundamentada de Glaser y Strauss (2008) mediante la cual desarrolló cuatro categorías principales para las concepciones de los estudiantes acerca de lo que conciben como línea de mejor ajuste es:

- Es una representación de una relación general entre las variables
- Muestra la apariencia de los datos.
- Es algo que uso para hacer predicciones.
- Es el promedio de los puntos.

Casey (2015) argumenta que las anteriores categorías no son independientes, estas pueden estar relacionadas. Por ejemplo, un estudiante para quién el significado de la línea de mejor ajuste es “algo que se utilice para obtener predicciones cercanas” o “muestra la apariencia de los datos”, constituye el criterio de escoger la línea de mejor ajuste como aquella que pase por el mayor número de puntos. Sin embargo, esta estrategia muestra que los estudiantes están preocupados con la línea que representa el conjunto de datos, pero a nivel de casos individuales, por lo tanto, se considera que hay una transición de una vista orientada de casos individuales a una visión total de datos. La concepción de la línea de mejor ajuste como el promedio de todos los puntos está presente tanto a nivel de bachillerato como universitario, lo que muestra que el estudio previo en matemáticas o estadística influye en la conceptualización de la línea de mejor ajuste. Casey (2015) menciona que algunos estudios previos en matemáticas, como el estudio de la función lineal o el aprendizaje sobre análisis estadísticos de datos univariados pueden influir de manera improductiva y productiva en la concepción de la línea de mejor ajuste, resaltando los siguientes aspectos:

- El aprendizaje previo de función lineal en matemáticas puede interferir en la capacidad de dar sentido a la tarea de ajustar una línea, como una línea que no va necesariamente a través de todos los puntos.
- La idea preconcebida que todas las líneas comienzan en el origen o la preocupación originada en matemáticas con la intercepción que se utiliza como punto de partida para la representación gráfica de la línea influye en el trazo de una línea que intercepte con el eje y o pase por el origen, en estadística es muy probable que la intersección con y de la línea de ajuste no tenga ningún significado en el contexto del problema.
- Aplicar la idea de promedio o media desde la perspectiva de datos univariados, por lo que colocan su línea de tal manera que tenga un número igual por encima y por debajo de la línea.

Para anexar a las concepciones a nivel medio básico descritas por Casey (2014, 2015), el estudio realizado por García y Yáñez (2018a) describe la concepción de “cercanía” de la recta a la nube de puntos, sin embargo, cuando los investigadores indagaron a qué se referían con esa expresión, sus respuestas fueron prácticamente nulas o evasivas. Por lo tanto, sugieren seguir investigando acerca de lo que los estudiantes comprenden o a qué le atribuyen el término cercanía y cómo miden esa cercanía.

Las concepciones de los estudiantes universitarios cuando se enfrentan a un problema de predicción fue el interés de Yáñez et al. (2016) quienes evidenciaron que algunos estudiantes utilizan el concepto matemático pendiente de una recta tomando dos pares de valores y estableciendo la expresión algebraica de la recta para resolver un problema de predicción, una idea intuitiva que puede ser usada por los profesores para aclarar la diferencia entre una relación funcional y una relación estadística tal y como lo menciona Casey (2014). Además, Yáñez et al. (2016) mencionan que algunas concepciones como la idea de promedio o media puede ser utilizada en la determinación de la recta de regresión, pues se puede generar una discusión con los estudiantes alrededor de la importancia de los promedios de las dos variables en la ecuación de la recta de mínimos cuadrados, ya que el punto cuyos valores son los promedios de las dos variables hacen parte de la recta de regresión.

La noción de pendiente en el estudio de la función lineal en matemáticas pareciera influir en la noción de la línea de mejor ajuste, por ello Casey y Nagle (2016) se interesaron por el rol de las concepciones de los estudiantes acerca de la pendiente en la conceptualización de la recta de mejor ajuste, evidenciando en su estudio que los estudiantes hacen referencia a múltiples conceptualizaciones de la pendiente cuando definen y ubican la línea de mejor ajuste informal, y son: (a) La conceptualización de la constante lineal de la pendiente desarrollada en matemáticas genera conflictos cognitivos con la configuración estadística donde los puntos no caen exactamente en la línea recta debido a la variación aleatoria. (b) La conceptualización del indicador de comportamiento se evidenció al evaluar si los puntos en el diagrama de dispersión exhibían un aumento o una tendencia decreciente, donde los estudiantes hicieron la evaluación desde una vista orientada a casos individuales y no ajustan la tendencia del conjunto de datos como un todo. (c) La conceptualización de situación del mundo real y la propiedad funcional, donde explicaron que la línea muestra covarianza entre las variables del mundo real lo cual apoyó la conceptualización de la línea de mejor ajuste correctamente y se colocó con precisión en el diagrama de dispersión. Más adelante, por esta misma línea acerca del conocimiento que tienen los estudiantes de la pendiente y se transfiere a la determinación del lugar de la línea de mejor ajuste Nagle et al. (2017) se enfocan en el caso de dos estudiantes de octavo para investigar qué factores podrían estar influyendo en el cambio de razonamiento. Los investigadores determinaron dos factores: (i) el contexto de la tarea influye en el conocimiento previo que los estudiantes transferirán a la nueva tarea, y ii) el razonamiento covariacional no se transfiere, pues los investigadores conjeturan que los estudiantes carecen de una comprensión conceptual de la covarianza desde el punto de vista funcional. En conclusión, estas investigaciones afirman que una comprensión conceptual de la pendiente no garantiza que se transferirá a un nuevo contexto problemático, y un contexto del mundo real no es solo una característica superficial de una tarea porque puede llegar a obstaculizar la transferencia de conocimientos.

Resumen. A continuación, se resume las concepciones en los estudiantes universitarios acerca del razonamiento estadístico covariacional:

- El parámetro coeficiente de correlación puede presentar la interpretación errónea de asumirse como una medida de la causa en lugar de una medida de asociación. Quizás,

la mal interpretación se debe a que proporciona una variedad de información como la dirección y fuerza de la relación entre las variables (ejemplo, su signo indica la dirección negativa o positiva) y es una medida entre -1 y 1, entre otras propiedades que posee.

- La concepción errónea de la transitividad se da cuando un estudiante considera que, sí dos variables X y Z se correlacionan positivamente con una tercera variable Y, entonces X y Z tienen correlación positiva, lo cual no siempre es cierto.
- Los estudiantes estiman con mayor precisión el coeficiente de correlación a partir de un diagrama de dispersión, los errores son mayores a partir de una tabla de valores, pues conocen el algoritmo, pero no lo realizan correctamente. Así mismo, es mayor el error al construir un diagrama de dispersión a partir de una descripción verbal.
- En problemas de regresión se presenta la dificultad de identificar la variable explicativa y la variable respuesta.

Respecto a las concepciones de los estudiantes acerca del razonamiento estadístico covariacional a nivel de secundaria, bachillerato y universitario, se tiene que:

- Los estudiantes pueden tener distintas estrategias al momento de representar gráficamente dos variables cuantitativas, algunos pueden mostrar gráficos más elaborados que otros y que muestren variación apropiada, pero es difícil que incluyan en su vocabulario la palabra “variación”.
- Es difícil para los estudiantes separar sus creencias previas para observar y evaluar el comportamiento de dos variables cuantitativas.
- Conceptos del análisis estadístico univariado (media, mediana) influyen en las estrategias utilizadas por los estudiantes al trazar la línea de mejor ajuste.
- Conceptos previos de la función lineal en matemáticas puede interferir en la capacidad de dar sentido a la tarea de ajustar una línea.
- Al interpretar diagramas de dispersión tiene una concepción determinista de la correlación (establecen una relación funcional), una concepción local de la correlación (solo analizan una parte de los datos representados) y una concepción causal de la correlación (correlación implica causalidad entre las variables).

- La línea de mejor ajuste puede conceptualizarse como la representación de una relación general entre las variables, muestra la apariencia de los datos, algo que se usa para hacer predicciones o el promedio de los puntos.
- La concepción “cercanía” entre la línea y la nube de puntos puede surgir en los estudiantes, pero falta mayor comprensión de otros elementos como el error que corresponde a la suma de los cuadrados de las distancias verticales entre los valores observados y los valores predichos.

La mayoría de las investigaciones que han estudiado las concepciones de los estudiantes, concuerdan en la necesidad de un diseño adecuado de situaciones de enseñanza que ofrezca un camino para utilizar las concepciones previas con el objetivo de promover el razonamiento covariacional y que a su vez ayuden a superar las concepciones erróneas que se pueden presentar.

Una tendencia en educación estadística es realizar investigaciones en las que se indague sobre las relaciones entre el diseño de la enseñanza y el avance en el aprendizaje de los estudiantes. El siguiente apartado hace una distinción y discusión de los estudios que han realizado intervención de enseñanza cuyas temáticas están relacionadas con el razonamiento covariacional.

1.1. Diseños de enseñanza y el avance en el aprendizaje de los estudiantes

En investigaciones educativas que realizan intervención, el diseño de la intervención es una componente importante. Por lo tanto, el marco para la revisión de los estudios de enseñanza consistirá en distinguirlos por niveles escolares, establecer el tipo de diseño de intervención, el tipo de tareas diseñadas e implementadas, junto con la metodología que utilizaron para el análisis de datos.

La tabla 1.3 muestra los estudios que han realizado intervención de enseñanza, de los cuales todos ellos utilizan soporte tecnológico; hasta el momento no se ha encontrado ningún estudio que realice intervención de enseñanza sin implementar tecnología. Las investigaciones ordenadas por el nivel escolar de los estudiantes (primaria, secundaria, bachillerato y universitario) seguido de la edad, la temática abordada y el software utilizado.

Tabla 1.3. Estudios que realizaron intervención de enseñanza utilizando tecnología entre el 1988 y 2019 según el nivel escolar, edad de los estudiantes, temática y software utilizado

Autores	Nivel	Edad	Temática	Software
Fitzallen (2012)	Primaria	12 años	Covariación gráficamente	TinkerPlots
García y Yáñez, (2018)	Secundaria	12 y 13 años	Recta de mejor ajuste	Geogebra y Fathom
Cobb, McClain y Gravemeijer (2003)	Secundaria	14 y 15 años	Análisis de datos bivariados	Minitools
Medina, Sánchez y Silvestre (2019)	Bachillerato	17 y 18 años	Correlación y regresión	Fathom
Gil y Gibbs (2017)	Bachillerato	17 y 18 años	Correlación y regresión	Gapminder y inZight
Dierdorp, Bakker, Eijkelhof y van Maanen (2011)	Bachillerato	16 y 17 años	Correlación y regresión	TI-Nspire
Inzunza (2015, 2016)	Universitario	18 y 19 años	Correlación y regresión	Applet del repositorio de Allan Rossman y Beth Chance Fathom

Zieffler y Garfield (2009)	Universitario	-	Análisis de datos bivariados	Minitools
Batanero, Godino y Estepa (1998)	Universitario	20 años	Correlación y regresión	PRODEST Y Statgraphics
Batanero, Estepa y Godino (1996)	Universitario	20 años	Correlación y regresión	PRODEST

1.1.1. Diseños de enseñanza a nivel universitario

Los primeros diseños de enseñanza se realizaron a nivel universitario. La tabla 1.4 muestra los 4 estudios encontrados en este nivel, junto con la metodología utilizada para el diseño de la intervención y la teoría de análisis de datos implementada.

Tabla 1.4. *Estudios de intervención de enseñanza entre el 1988 y 2019 a nivel universitario, según la metodología utilizada para el diseño, la teoría empleada para el análisis.*

Autores	Metodología de diseño	Teoría de análisis
Inzunza, (2015, 2016)	Principios Instruccionales de Cobb y McClain (2004)	Modelo SOLO de Biggs y Collis (1982)
Zieffler y Garfield (2009)	Guidelines for the Assessment and Instruction in Statistics Education (GAISE, 2005)	Modelos lineales mixtos
Batanero, Godino y Estepa (1998)	Variables de tareas en la resolución de problemas estadísticos utilizando computadores (Godino, Batanero y Estepa, 1991)	Marco teórico del significado y comprensión de objetos matemáticos de Godino (1996)
Batanero, Estepa y Godino (1996)	Consideraciones de: Willet y Singer (1992) Hawkins (1990) Biehler (1990)	Análisis de correspondencia

Los primeros **experimentos de enseñanza** a nivel universitario fueron realizados por Estepa y Batanero (1996) y Batanero et al. (1998) quienes se interesaron por la comprensión de la asociación estadística y su evolución después de la intervención de enseñanza utilizando computadores. Dado que el interés fue la asociación, los estudios abordaron tareas con tablas de contingencia al igual que la asociación en variables numéricas, aclaramos de nuevo que nos enfocaremos en los ítems de asociación en variables numéricas, donde tiene sentido hablar de correlación y regresión lineal. Además, los investigadores antes de realizar los experimentos aplicaron un cuestionario con el fin de establecer las preconcepciones de los estudiantes, estas preconcepciones fueron descritas anteriormente en primer apartado de este capítulo.

El diseño de Batanero et al. (1996) tomó las siguientes consideraciones para organizar el experimento de enseñanza:

- i) Un curso orientado a la resolución de problemas de la vida real hace que las estadísticas sean más interesantes para los estudiantes (Willet y Singer, 1992).
- ii) Los estudiantes que hicieron parte del estudio estaban estudiando para ser maestros, luego existe una dificultad para capacitar a los maestros para enseñar estadísticas no solo con las matemáticas sino también con la informática y sus campos de aplicación (Hawkins, 1989).
- iii) Las computadoras modernas han revolucionado la visión de lo que es el conocimiento estadístico, por lo que los estudiantes también necesitan aprender a usar software para resolver problemas estadísticos (Biehler, 1990).
- iv) Las estadísticas requieren más que el conocimiento matemático, estos requisitos incluyen organizar e implementar proyectos, así como alentar el trabajo y la cooperación entre estudiantes y entender la computación no como herramientas, sino como medios estadísticos esenciales de conocimiento.

Los investigadores mediante el análisis de las respuestas de los estudiantes a los problemas que involucraban diagramas de dispersión evidenciaron lo siguiente: la mayoría de los estudiantes superaron la concepción determinista de la asociación (la relación entre las

variables es funcional), aceptando la dependencia aleatoria. La concepción local de la asociación (basan su juicio en solo una parte de los datos representados) también se erradicó a medida que los estudiantes notaron la importancia de tener en cuenta el conjunto de datos completo para evaluar la asociación. No hubo ninguna mejora en absoluto con respecto a la concepción causal de la asociación (correlación implica causalidad entre las variables), pues la mayoría de los estudiantes no se dieron cuenta de que una fuerte asociación entre dos variables no es suficiente para sacar conclusiones sobre la causa y el efecto. Batanero et al. (1996) sugieren que los anteriores elementos deben contextualizarse en situaciones problemáticas adecuadas que ayuden a los estudiantes a desarrollar herramientas conceptuales para resolver problemas estadísticos y comprender un significado más completo de la asociación.

El estudio de Batanero et al. (1998) con 213 estudiantes de secundaria fue una extensión del estudio de Batanero et al. (1996), por lo tanto, los elementos propuestos en esa primera investigación fueron refinados. Batanero et al. (1998) realizaron 21 sesiones para el primer experimento, y 40 sesiones para el segundo. En 7 sesiones del primer experimento y 20 del segundo los estudiantes trabajaron en el laboratorio de estadística, resolviendo problemas cuyas soluciones requerían que analizaran diferentes conjuntos de datos proporcionados por el profesor o recopilados por ellos mismos. El primer experimento fue realizado por 19 estudiantes de 20 años en su primer curso de estadística. Los estudiantes trabajaron con el software PRODEST. La intención del primer experimento fue evaluar los cambios en las concepciones de los estudiantes por lo tanto diseñaron dos versiones equivalentes de un cuestionario que fueron aplicados como pretest y post-test. El segundo experimento se realizó con 36 estudiantes quienes hicieron parte del segundo experimento y trabajaron con el paquete estadístico Statgraphis el cual permite a los estudiantes operar o representar datos bivariados de diferentes maneras. Se evaluaron las habilidades generativas e interpretativas de las soluciones de los estudiantes a las actividades de análisis de datos, pruebas de papel y lápiz, y de sus proyectos individuales. Además, al final del curso, se les aplicó una prueba a los estudiantes sobre el análisis de un nuevo conjunto de datos para evaluar el significado final que los estudiantes asignaron a la asociación. Un miembro del equipo de investigación observó el trabajo de los estudiantes, reuniendo sus respuestas por escrito a los diferentes

problemas. Esta observación también incluyó la grabación de sus discusiones, de su interacción con el profesor y la computadora.

En el estudio de Batanero et al. (1998) se establecen tres elementos adicionales a los encontrados en el estudio de Batanero et al. (1996) con el objetivo que los estudiantes dominen el concepto de asociación y lo utilicen en la resolución de problemas:

- Elemento extensional del significado: las diferentes situaciones y problemas cuya solución requiere el estudio de la asociación.
- Elementos de significado instrumentales/representativos: el uso de diferentes herramientas para tratar o representar el concepto, tales como diagramas de dispersión, tablas de dos vías o una serie de paralelos, gráficos de caja, gráficos acumulados, histogramas o gráficos de barras.
- Elementos de significado intensivos: como la diferencia entre la dependencia estadística y la dependencia funcional, las diferentes frecuencias relativas que pueden deducirse de una tabla de contingencia, el papel de las variables independientes y dependientes, y los parámetros en la ecuación de regresión, la interpretación del coeficiente de correlación y la diferencia entre correlación y causalidad.

El **estudio longitudinal** de Zieffler y Garfield (2009) buscó modelar el desarrollo del razonamiento covariacional durante un curso introductorio de estadística, como marco teórico utilizaron el modelo lineal efecto-mixto para describir el razonamiento de los estudiantes. Dos instructores distintos enseñaron a los cuatro cursos en un laboratorio de cómputo dos veces por semana por una hora y 15 minutos. Los instructores utilizaron dos secuencias aleatorias en las temáticas (ver figura 1.1) para cada curso. Primero se aplicó una encuesta sobre los conceptos previos de los estudiantes, luego se evaluó dos veces el razonamiento bivariado antes de la enseñanza del tema y la última evaluación fue al finalizar el curso.

Figura 1.1. *Secuencias de enseñanza propuesta por Zieffler y Garfield (2009)*

Secuencia 1:

Muestreo (6 Días) ⇒ EDA (7 Días) ⇒ Datos Bivariados (4 Días) ⇒ Distribución Muestral (3 Días) ⇒ Probabilidad (2 Días) ⇒ Inferencia (6 Días)

Secuencia 2:

Muestreo (6 Días) ⇒ EDA (7 Días) ⇒ Distribución Muestral (3 Días) ⇒ Probabilidad (2 Días) ⇒ Inferencia (6 Días) ⇒ Datos Bivariados (4 Días)

Para una muestra de 113 estudiantes obtuvieron un modelo cuadrático que evidenció que la mayor parte del cambio en el desarrollo del razonamiento de los estudiantes acerca de datos bivariados parece ocurrir entre las primeras dos ocasiones de medidas. Esto fue ante de la enseñanza formalmente de datos bivariados en cualquiera de las secuencias de instrucción que utilizaron.

Zieffler y Garfield (2009) afirman que el modelo que obtuvieron podría indicar que el desarrollo del razonamiento de los estudiantes acerca de datos bivariados es más un artificio de su desarrollo del razonamiento estadístico en general que un resultado de cualquier instrucción formal sobre un tema de datos bivariados, por lo que concluyen que a pesar de la buena unidad de instrucción diseñada basada en la investigación en datos bivariados, los estudiantes todavía presentaron dificultades con mucho elementos del razonamiento bivariado al final del curso de 15 semanas, lo cual confirma los hallazgos en la literatura de investigación de que las ideas de covariación suelen ser difíciles de aprender para los estudiantes. Los investigadores sugieren prestar más atención al diseño de actividades y materiales de instrucción para desarrollar el razonamiento covariacional.

Las **investigaciones de diseño** de Inzunza (2016), Inzunza y Ward (2015) exploran el razonamiento covariacional de 34 estudiantes universitarios mediante el uso de tecnología computacional y realizan un diseño de enseñanza basado en Cobb y McClain (2004) con los siguientes elementos:

- Aplicación un cuestionario diagnóstico para evidenciar los antecedentes en el tema de correlación y regresión.

- Se utilizaron un par de videos, un paquete de diapositivas que el profesor e investigador preparó con los conceptos involucrados en el análisis de datos bivariados (interpretación de diagramas de dispersión, coeficiente de correlación, la recta de regresión y el coeficiente de determinación).
- El diseño de las actividades de enseñanza: las actividades contemplaban conjuntos de datos reales multivariados que se tomaron de algunos sitios de internet, y siguieron la siguiente secuencia: Punto de partida: construir diagramas de dispersión para enfocar la atención de los estudiantes a visualizar sus características. Segundo: calcular los coeficientes de correlación buscando conectar su valor con las características del diagrama y Tercero: introducir la recta de regresión para ajustar la tendencia de los puntos y su ecuación, con el valor del coeficiente de determinación que por omisión proporciona el software. Los datos fueron las hojas de trabajo de los estudiantes donde escribían sus respuestas.

Inzunza (2016), Inzunza y Ward (2015) utilizan el marco teórico SOLO para el análisis de los datos que obtuvieron al diseñar y aplicar tres tareas. Los investigadores concluyen mediante las descripciones y análisis de las respuestas de los estudiantes que la aparente sencillez de un diagrama de dispersión queda entredicha con las respuestas de los estudiantes, dado que solo uno de ellos logró ubicarse en el nivel relacional del modelo SOLO en dos preguntas que involucraban la construcción e interpretación del diagrama. Inzunza (2016) establece los siguientes niveles del modelo SOLO (tabla 1.5) en la interpretación del diagrama de dispersión.

Tabla 1.5. *Niveles de modelado SOLO en la interpretación de diagrama de dispersión establecido por Inzunza (2016)*

<p>Nivel Uniestructural</p>	<p>Los estudiantes emplean una sola característica del diagrama de dispersión para describir la relación entre las dos variables. Aun cuando agregan o calculan algún elemento adicional al diagrama (por ejemplo, la recta de regresión y el coeficiente de correlación), solo dependen de una representación para el análisis.</p>
---------------------------------	--

<p>Nivel Multiestructural</p>	<p>Los estudiantes usan más de una propiedad del diagrama de dispersión para describir la relación entre las dos variables. Se apoyan en el diagrama de dispersión y en algún elemento o cálculo adicional (por ejemplo, la recta de regresión y el coeficiente de correlación) para la interpretación.</p>
<p>Nivel Relacional</p>	<p>Los estudiantes utilizan múltiples propiedades del diagrama de dispersión para describir la relación entre las dos variables. Se valen del diagrama y elementos o cálculos adicionales para la interpretación. Además, hacen referencia a otros elementos o al contexto para describir o justificar la relación.</p>

Además, Inzunza y Ward (2015) establecen que las respuestas de algunos estudiantes muestran que lograron apreciar en forma intuitiva la relación entre la variación y la precisión de una estimación en la variable de respuesta, pero faltó relacionarla con la suma de cuadrados y el coeficiente de determinación. Resalta la importancia de los recursos dinámicos e interactivos de los ambientes computacionales en el diseño de secuencias de aprendizaje de la covariación dado que motiva el interés de explorar relaciones por sus dinamismos en las representaciones y la interactividad que permite manipular tales representaciones.

Resumen. Las metodologías implementadas por las investigaciones que han realizado intervención de enseñanza a nivel universitario han sido experimentos de enseñanza, un estudio longitudinal e investigación de diseño, todas ellas utilizando tecnología digital.

Los experimentos de enseñanza muestran que la concepción causal prevalece en los estudiantes, para quienes la existencia de una relación implica que una variable es la causa del cambio en la otra variable. Además, los primeros experimentos de enseñanza determinaron elementos importantes para el diseño de actividades relacionadas con la correlación y regresión, entre ellos: situaciones problemas interesantes, la comprensión de la representación gráfica del diagrama de dispersión, el estudiar las diferencias entre dependencia funcional y aleatoria, la diferenciación entre variable dependiente e independiente y la interpretación del coeficiente de correlación.

El estudio longitudinal da cuenta que la ubicación de las temáticas de correlación y regresión en el plan de estudio no influye en el desarrollo del razonamiento covariacional, por lo tanto, sugieren enfocarse en el diseño de actividades y materiales de instrucción. Y las investigaciones de diseño resaltan la importancia de las tareas, así como, el prestar atención a la variación y al concepto de error, dado que son elementos explícitos en la determinación de la recta de regresión y del coeficiente de correlación.

1.1.2. Diseños de enseñanza a nivel bachillerato

A nivel de bachillerato se encontraron 3 investigaciones. La tabla 1.6 muestra estos estudios junto con la metodología utilizada para el diseño de la intervención y la teoría de análisis de datos implementada.

Tabla 1.6. *Estudios de intervención de enseñanza entre el 1988 y 2019 a nivel bachillerato, según la metodología utilizada para el diseño, la teoría empleada para el análisis y software utilizado.*

Autores	Metodología de diseño	Teoría de análisis
Medina, Sánchez y Silvestre (2019)	Trayectoria hipotética de aprendizaje	Teoría fundamentada
Gil y Gibbs (2017)	Enfoque basado en el diseño y trayectoria de aprendizaje	Análisis cuantitativo y codificación
Dierdorp, Bakker, Eijkelhof y van Maanen (2011)	Experimento de diseño y trayectoria hipotética de aprendizaje	Propias matrices de análisis

Para Simon y Tzur (2004) las **Trayectorias hipotéticas de aprendizaje (THA)** son herramientas que brindan un marco sobre los procesos hipotéticos de aprendizaje y la planificación de actividades de instrucción en las clases. Investigaciones como la realizada por Medina et al. (2019) muestran el diseño y la aplicación de una THA para la correlación y regresión con 96 estudiantes entre 17 y 18 años. Los objetivos de aprendizaje definidos por los investigadores fueron:

- El aprendizaje sobre las distribuciones de datos bivariados: que pretende que los estudiantes identifiquen qué elementos, como la forma y la tendencia de una nube de puntos de datos, determinan establecer el tipo de relación que existe entre dos variables cuantitativas.
- El aprendizaje sobre la variación de los datos con respecto al modelo de ajuste lineal: que pretende que los estudiantes identifiquen que el modelo de ajuste lineal tiene un carácter aleatorio intrínseco enraizado en los datos mismos.

Para el desarrollo de la THA aplicaron dos problemas con uso de Fathom: El primer problema consistía en analizar dos diagramas de dispersión y proponer un modelo lineal que mejor se ajustará a los datos utilizando. En el segundo problema, los estudiantes debían comparar la línea de mínimos cuadrados que proporcionaba Fathom con el modelo que estableció para la línea de ajuste del primer problema.

El método de análisis de las respuestas de los estudiantes y de grabaciones de video se analizaron empleando los principios y técnicas de codificación que brinda la Teoría Fundamentada. Entre los resultados se destaca:

- El razonamiento de los estudiantes es dividir los datos en dos partes, los casos que se comportan linealmente y los que no. Los investigadores conjeturan que los estudiantes todavía están luchando para asumir los datos como un agregado, pero a lo sumo, como un agregado particionado en dos.
- Se presentó alta frecuencia en las respuestas de los estudiantes respecto a la descripción de la tendencia y la referencia exclusiva a la línea en la etapa final de la THA. Los códigos favorables estuvieron ausentes en el estudio diagnóstico de Medina et al. (2016) por lo tanto los investigadores creen que la aparición de estos códigos favorables se dio por la incorporación del software a la secuencia de instrucciones.
- Afirman que una nueva versión del THA y futuras investigaciones podrían abordar el razonamiento de los estudiantes en aspectos relacionados con la causalidad y la variabilidad.

De los estudios más recientes a nivel de bachillerato está el realizado por Gil y Gibbs (2017), quienes examinaron el diseño de una unidad de enseñanza apoyada con tecnología con el objetivo de estudiar el cambio en el razonamiento covariacional de estudiantes de bachillerato en un contexto de grandes datos. Utilizaron en su estudio un **enfoque basado en el diseño** de Barab y Squire (2004) junto con cinco características de la investigación de diseño de Cobb et al. (2003), los elementos del diseño fueron:

- Diseñar una unidad de enseñanza de cinco actividades, las cuales se aplicaron durante seis sesiones de clase de 72 minutos cada clase.
- Los participantes fueron estudiantes de dos cursos de grado 12 entre 17 y 18 años, quienes no tenían mucha experiencia con la estadística. En grado 9 habían discutido cuestiones de tendencia y línea de mejor ajuste y relaciones entre variables dependientes e independientes.
- El profesor decidió asumir un rol secundario de apoyo y renunciar a su rol como maestro principal.
- Todas las actividades y discusiones en las dos clases fueron grabadas en video respetando las restricciones en el permiso de video para algunos estudiantes. Durante la primera y quinta actividad, se ubicaron dos cámaras de video en el aula para capturar las presentaciones de los estudiantes, dado que grupos de estudiantes presentaban sus hallazgos a toda su clase.
- Realizaron un análisis cualitativo de los datos de los videos, las presentaciones de los estudiantes y las discusiones seleccionadas se transcribieron.

La trayectoria de aprendizaje (tabla 1.7) implementada por Gil y Gibbs (2017) combinó exploraciones de datos de tamaño medio en Gapminder e inZight con una actividad colaborativa interdisciplinaria para el aprendizaje acerca de datos grandes y tomó lugar en un Espacio de Aprendizaje Orquestado Interactivo (Interactive Orchestrated Learning Space), un aula transformada que utiliza infraestructura tecnológica para conectar las fuentes de datos grandes con una plataforma colaborativa genérica para la comunicación y la construcción del conocimiento.

Tabla 1.7. *THA diseñada e implementada por Gil y Gibbs (2017)*

Contenido	Conceptos e ideas estadísticas
Introducción a las tendencias en los datos del mundo de tamaño medio en Gapminder	Examinación de las ideas iniciales del razonamiento covariacional y modelado
Experiencia interactiva con grandes datos	Aprendizaje acerca de diferentes aspectos y aplicación de datos grandes desde ejemplos en contextos interdisciplinarios
Discusión acerca de los aspectos con datos grandes	Uso de modelos para predecir tendencias en las poblaciones (Google Flu)
Recreación de los diagramas de dispersión desde el ejercicio de tendencia. Introducción a iNZight	Emerger ideas acerca del modelado de la covariación investigando datos bi- y multivariados; Aprender herramientas para la modelación de la covariación con datos de entrenamiento usando iNZight
Introducción a la página web Wellbeing Toronto y exploración de Toronto con iNZight	Investigando datos de tamaño medio bi- y multivariados a través de la visualización espacial y exploración de las relaciones entre variables usando iNZight
Gran discurso de datos de los estudiantes	Presentación de los hallazgos desde la investigación de datos de tamaño medio; Examinación del progreso del razonamiento covariacional y modelado.

Entre sus resultados evidenciaron que la noción de tendencia surgió en el razonamiento de los estudiantes y les permitió describir y explicar el comportamiento de otros sistemas en el mundo. Por ejemplo, la dirección de una tendencia o la forma de un modelo para una asociación puede describir o explicar la relación entre alfabetización y la tasa de fecundidad en los países del mundo, o las tasas de riqueza y delincuencia en los barrios de una ciudad. Además, en la recreación del diagrama de dispersión, se encontró que los gestos

representativos fue otra manera de expresar el razonamiento estadístico y demostrar el razonamiento covariacional y el modelado de este. Una asociación fuerte, descrita como “cuando puedo ponerle la mano encima”, agrega otra superposición de comprensión a nuestra interpretación del razonamiento de los estudiantes y le da vida, mediante gestos y expresiones verbales en un sistema conceptual.

Por otro lado, algunas investigaciones han diseñado e implementado THA dentro de la metodología de los **experimentos de diseño**. Por ejemplo, Dierdorp et al. (2011) con estudiantes de grado 11 proponen una THA donde buscaban apoyar el razonamiento covariacional mediante una estrategia de enseñanza y aprendizaje basada en prácticas auténticas que se enfocaban en cuatro ideas conceptuales:

- i. Apoyar a los estudiantes para coordinar visualizaciones individuales y agregadas sobre conjunto de datos.
- ii. Reconocer una tendencia que se extiende más allá de los datos.
- iii. Dar sentido a la variabilidad.
- iv. Utilizar datos como evidencia al hacer una inferencia.

La estrategia de enseñanza y aprendizaje incluía: los materiales educativos concretos (la unidad de enseñanza que consta de tres capítulos, hoja de cálculo de Excel, el software TI-Nspire) y una trayectoria hipotética de aprendizaje (tabla 1.8). Los estudiantes participaron en la discusión en el aula, pero también trabajaron en parejas y en grupos pequeños.

Tabla 1.8. *Resumen de la THA diseñada e implementada por Dierdorp et al., (2011)*

Capítulo	Conjetura THA	Contenido
Capítulo 1. Recopilación de datos de su frecuencia de frecuencia cardíaca (HRF) al aumentar la intensidad de su entrenamiento	Los estudiantes reconocerían la tendencia lineal en los datos hasta un cierto valor y encontrarían el umbral frecuencia cardíaca a partir de un diagrama de dispersión	Familiarizar a los estudiantes con la recolección de datos por ellos mismos y con diagramas de dispersión, variabilidad y correlación en un sentido informal

Capítulo 2. Monitoreando la altura de los diques.	Hacer una inferencia sobre cuándo se debe aumentar la altura de un dique, al igual que lo hacen las organizaciones de investigación profesional para el monitoreo de diques.	Generar una vista agregada de un conjunto de datos presentados en diagramas de dispersión, para ver las tendencias
Capítulo 3. Calibración de los termómetros	Los estudiantes deben tomar una decisión sobre qué termómetro es el mejor.	Evaluar la aptitud de los estudiantes para hacer inferencias con la ayuda de la regresión y la correlación

Sus resultados muestran un gran potencial para basar la instrucción estadística en prácticas auténticas, justificando que:

- En contextos auténticos, los conceptos y técnicas estadísticas se utilizan para fines con los que los estudiantes pueden relacionarse y, por lo tanto, pueden apoyar el aprendizaje de los estudiantes, siempre que la enseñanza y los materiales educativos sean de la calidad suficiente.
- Las tareas, inspiradas por problemas auténticos, parecían lo suficientemente realistas para que los estudiantes experimentaran autenticidad y se sintieran comprometidos.
- Las tareas fueron cada vez más difíciles y proporcionaron a los estudiantes muchas oportunidades para razonar sobre diagramas de dispersión, variabilidad, líneas de regresión y correlación en relación con los problemas contextuales.
- Las tareas en las que los estudiantes tenían que recopilar y modelar sus propios datos alimentaron la necesidad percibida de encontrar una solución para el problema contextual y estudiar los materiales educativos.

Resumen. Las investigaciones a nivel de bachillerato han establecido trayectorias hipotéticas de aprendizaje y algunas de ellas han utilizado el experimento de diseño, todas ellas han utilizado tecnología digital. La tabla 1.9 resume las nociones estadísticas que cada estudio abordó referente a la correlación y regresión.

Tabla 1.9. *Resume de las nociones estadísticas que abordó los estudios de enseñanza a nivel de bachillerato*

Medina et al. (2019)	Cobb, McClain, et al. (2003)	Dierdorp et al. (2011)
Forma y tendencia de la nube de puntos	Dirección y fuerza de la relación	Visualizaciones individuales y agregadas
Tipo de relación entre las variables	Los diagramas de dispersión	Tendencia de los datos
Modelo de ajuste lineal	Línea de mejor ajuste	Dar sentido a la variabilidad y hacer inferencias

De la tabla se deduce una gran dispersión en las temáticas abordadas por cada estudio: en contextos de grandes datos, en el análisis gráfico de la relación, en la recta de mejor ajuste e inferencias.

Los estudios concuerdan en la dificultad que tiene para los estudiantes la noción de agregado, es decir, ver el conjunto de datos con una totalidad, derivándose una concepción local al analizar la relación entre dos variables y al trazar el modelo lineal de ajuste. También es difícil para los estudiantes abandonar la concepción causal y desarrollar la noción de variabilidad de los datos.

Se presentan dos sugerencias opuestas, por un lado, se recomienda iniciar la instrucción con describir y explorar datos antes de la creación de datos, una perspectiva opuesta plantea que las actividades para iniciar como recopilar y modelar datos propios aumentan la necesidad en los estudiantes de encontrar una solución.

1.1.3. Diseños de enseñanza a nivel primaria y secundaria

En la tabla 1.10 se muestra que se encontró un solo estudio a nivel de primaria, realizado por Fitzallen (2012), y 2 estudios a nivel de secundaria (Cobb, McClain, et al., 2003; García y Yáñez, 2018b).

Tabla 1.10. *Estudios de intervención de enseñanza entre el 1988 y 2019 a nivel primaria, según la metodología utilizada para el diseño, la teoría empleada para el análisis y software utilizado*

Autores	Metodología de diseño	Teoría de análisis
Fitzallen (2012)	Investigación de diseño	Taxonomía SOLO
García y Yáñez, (2018b)	Trayectoria hipotética de aprendizaje	Análisis retrospectivo
Cobb, McClain y Gravemeijer (2003)	Investigación de diseño educacional y una trayectoria hipotética de aprendizaje	Marco interpretativo de significados matemáticos normativos de Cobb, Stephan, McClain y Gravemeijer (2001)

Fitzallen (2012) quien utilizó una **investigación de diseño educativo** basado en Akker et al. (2006) con un paradigma pragmático de Mackenzie y Knipe (2006) con el objetivo de captar la forma en que el software TinkerPlots contribuyó, apoyó e influyó en el desarrollo de la comprensión de la covariación de los estudiantes a medida que creaban e interpretaban gráficas. Fitzallen aclara que el término de diseño educativo se utiliza cuando se hace referencia a metodologías que se alinean con los estudios de diseño y los experimentos de diseño.

La secuencia de aprendizaje implementada por Fitzallen (2012) comprendió 6 lecciones de clase, en las cuales se abordaban distintas características de Tinkerplots, por ejemplo: creación de gráficos apilados, pilas de datos separadas y diagramas de dispersión. Para analizar las respuestas de los estudiantes utilizó el marco de Estructura de Resultados de Aprendizaje Observados (SOLO) desarrollado por Biggs y Collis (1982) junto con tres niveles de interpretación gráfica de Kosslyn (1989) los cuales se alinea con los niveles propuestos en SOLO. La teoría resultante de este estudio con 12 estudiantes de primaria es: los estudiantes de los últimos grados de primaria (12 años) son capaces de evaluar la relación entre dos atributos y desarrollar una comprensión de la covariación. En el nivel inicial de comprensión, ellos basan su conocimiento en el contexto de los datos o en elementos

individuales de los gráficos para hacer sus conclusiones. En el siguiente nivel, comienzan a considerar múltiples características o entre los gráficos y su conocimiento del contexto. Los estudiantes que operan en el nivel más alto de pensamiento son capaces de hacer las conexiones entre las partes que constituyen la gráfica, el contexto de los datos y el mensaje en los datos caracterizados por la tendencia.

La investigación de Fitzallen (2012) se centra en demostrar como el software TinkerPlots provee un ambiente de aprendizaje que favorece la oportunidad de establecer la noción de covariación en los estudiantes a través de su caracterización en los gráficos. Sugiriendo que los estudiantes de la escuela primaria superior pueden evaluar la relación entre dos atributos y desarrollar una comprensión de la covariación.

El estudio realizado por García y Yáñez (2018b) con el objetivo de analizar las formas de razonamiento covariacional informal de estudiantes de octavo grado alrededor de la recta de mejor ajuste proponen un THA que consta de dos fases: Fase I: Análisis univariado (la media como un predictor), Fase II: Análisis bivariado (la recta de mejor ajuste). Para lo cual diseñan la siguiente secuencia de actividades:

- Primera Actividad. Tuvo como objetivo que los estudiantes vieran la media, más allá de un algoritmo matemático, como un punto de “equilibrio” o “balanceador” y que dieran cuenta de una de sus propiedades: “la suma de las distancias a la derecha sobre la media es igual a la suma de las distancias a la izquierda sobre la media”.
- Segunda Actividad. Exploración de la relación entre dos variables cuantitativas, con ayuda de Fathom los estudiantes debían hacer un gráfico para cada una de las variables y observar la relación entre las dos variables cuando seleccionaban parte de los datos en el gráfico.
- Tercera Actividad. Los estudiantes debían ubicar la recta que ellos consideraban representa la relación entre las variables y verificar la ubicación de la recta. El propósito era extender las ideas que se habían visto para el caso univariado y llegar a que la suma de las diferencias de los puntos a la mejor recta era cero.

Los investigadores describen que la mayoría de los estudiantes caracterizan la media como un punto de equilibrio en el caso univariado y en el caso bivariado, los estudiantes reflejaron

una coordinación entre dos variables al asociar expresiones. En cuanto a la predicción, algunos sugieren trazar la recta, pero reflejan una concepción local al predecir los valores de los puntos más cercanos y no reflejan una visión global de los datos para que sus predicciones sean coherentes con el tipo de relación que tienen las variables. En lo concerniente a la ubicación de la recta de mejor ajuste, la mayoría de los estudiantes le atribuyen a esta recta característica como: se ubica entre los puntos, refleja la relación entre las dos variables y es la más cercana a los puntos.

Cobb, McClain, et al. (2003) diseñan e implementan una THA dentro de la metodología de los **experimentos de diseño**, cuyo objetivo es documentar una trayectoria de aprendizaje real de la comunidad en el aula. Los investigadores dividieron el proceso de realización del experimento de diseño en tres fases:

Fase I. Desarrollaron conjeturas sobre las (a) posibles trayectorias para el aprendizaje matemático y (b) los medios que podrían ser utilizados para apoyar y organizar ese aprendizaje. La idea central fue la distribución de datos bivariados que implicaba las nociones de dirección y fuerza de la relación entre dos conjuntos de medidas. El rol central que representa los diagramas de dispersión y la noción de línea de mejor ajuste mediante mínimos cuadrados. Para el análisis de datos bivariados los investigadores desarrollaron una mini herramienta (minitool) de computador.

Fase II. Experimentación en la clase: las actividades de instrucción específicas que se utilizaron en el aula se desarrollaron solo uno o dos días antes.

Fase III. Realización de un análisis retrospectivo: la intención fue mejorar el diseño de instrucción inicial mediante la prueba y modificación de conjeturas sobre el curso de la comunidad del aula y el aprendizaje matemático de los estudiantes participantes. Esto les permitió de la revisión a priori de la secuencia instruccional, establecer los siguientes componentes de una nueva trayectoria de aprendizaje para datos bivariados:

- Puntos de inicio: conjuntos de datos univariados como distribuciones: Comparar conjuntos de datos univariados al estructurarlos en términos de patrones basados en la percepción.

- Desarrollando formas de inscribir datos bivariados: Llevar a los estudiantes a que elijan el diagrama de dispersión como la mejor forma de representar datos bivariados.
- Datos apilados como distribuciones bivariadas. Involucrar datos apilados, y enmarcar la actividad de instrucción en la constitución de todo el conjunto de datos como una distribución bivariada. Los investigadores conjeturan que analizar la distribución dentro de los segmentos podría proporcionar una base para un enfoque posterior en las tendencias y patrones en un conjunto de datos completo.

Además, de la THA sugieren una teoría instruccional de dominio específico para el análisis de datos estadísticos con las siguientes características: (1) la probabilidad debería incluirse en lo que se denomina análisis de datos exploratorios, dado que evidenciaron que las preguntas que involucran remuestreo dan lugar a problemas de naturaleza probabilística. (2) Recomiendan comenzar con métodos para explorar y describir datos antes de enfocarse en la creación de datos y luego pasar a la inferencia estadística. También recomiendan que el razonamiento de los estudiantes no sea analizado en términos psicológicos individuales, sino interpretar los casos de estudiantes que no lograron aprender como indicadores de las insuficiencias del diseño instruccional.

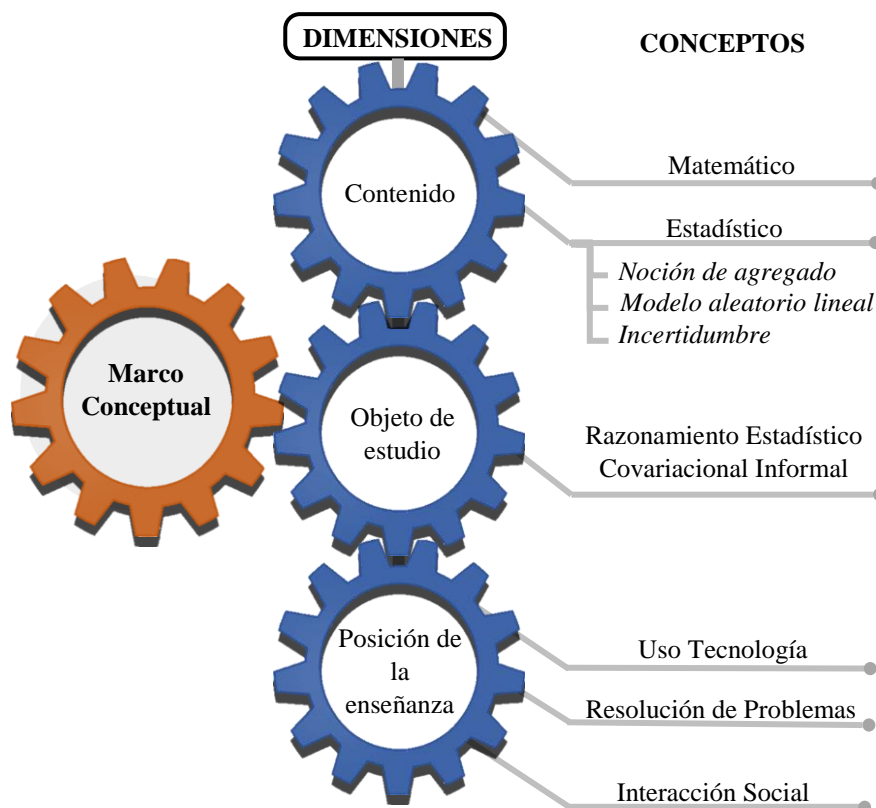
CAPÍTULO II. MARCO CONCEPTUAL

La noción de marco conceptual que aquí se maneja es entendido de una manera similar a la caracterización que proponen 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).

Jabareen (2009) propone referirse simplemente a “conceptos” en lugar de “los factores clave, las variables o los constructos” y esta simplificación es conveniente para nuestro trabajo. En consecuencia, proponemos el marco conceptual que muestra la Figura 2.1, el cual organiza los conceptos (aspectos matemáticos, aspectos estadísticos, razonamiento estadístico covariacional informal, tecnología, resolución de problemas e interacción social) en tres dimensiones (contenido, el objeto de estudio y la posición de la enseñanza). De esta manera se revela una estructura general de la presente investigación.

Figura 2.1. Marco conceptual de la investigación (elaboración propia)



Nuestro interés fue conocer los razonamientos informales que surgen en los estudiantes al enfrentar por primera vez problemas de correlación y regresión (razonamiento estadístico covariacional informal) utilizando soporte tecnológico. Una primera dimensión del marco es el contenido, el cual comprende los aspectos matemáticos y estadísticos establecidos por los estándares para el nivel de bachillerato. Los aspectos matemáticos involucrados en las actividades corresponden a los desarrollados en los cursos de álgebra y geometría analítica del bachillerato; en especial, función lineal, plano cartesiano y la ecuación de la recta. Con relación a los aspectos estadísticos destacamos tres que consideramos son clave de la estadística (agregado, modelo aleatorio lineal e incertidumbre). La segunda dimensión concierne a nuestro objeto de estudio, a saber, el razonamiento de los estudiantes en función de los elementos de diseño. La tercera se refiere a los principios de enseñanza bajo los cuales se hizo el diseño y el tipo de investigación que se realiza, a saber, el experimento de diseño.

2.1. Contenido: Aspectos matemáticos y estadísticos

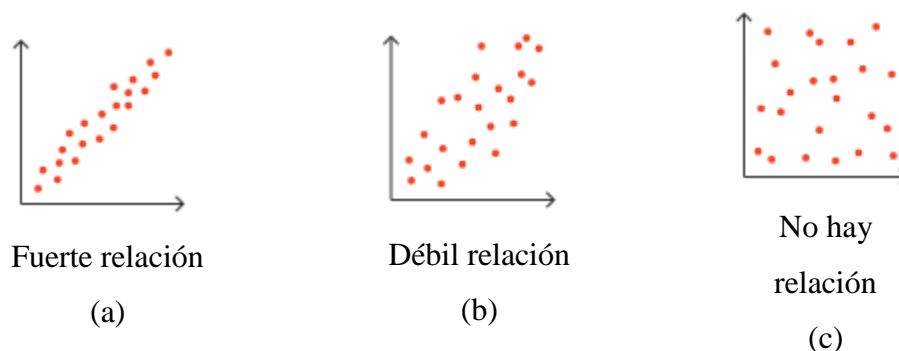
El contenido temático que un estudiante de último grado de bachillerato debe desarrollar sobre una relación bivariada de datos cuantitativos contiene conceptos y procedimientos numéricos, algebraicos y funcionales que están presentes en el análisis de la correlación y regresión lineal junto con el contenido estadístico que implica los aspectos relacionados con el contexto de los datos como la incertidumbre y la causalidad. Con base en los estándares curriculares de la National Council Of Teachers Of Mathematics (NCTM, 2000, p.327-328) para el último nivel de bachillerato y la unidad de datos bivariados en los programas del Colegio de Ciencias y Humanidades de la UNAM, México (Universidad Autónoma Nacional de México, 2016) se establecen los siguientes aspectos matemáticos y estadísticos:

2.1.1. Aspectos Matemáticos

- Expresar como pares ordenados (x, y) los datos de dos variables cuantitativas.
- Representar cada par (x, y) con un punto en un sistema de coordenadas bidimensional (plano cartesiano) conocido como diagrama de dispersión.
- Determinar la fuerza de la relación lineal entre las variables en un diagrama de dispersión, observada a través de la mayor o menor dispersión de la nube de puntos.

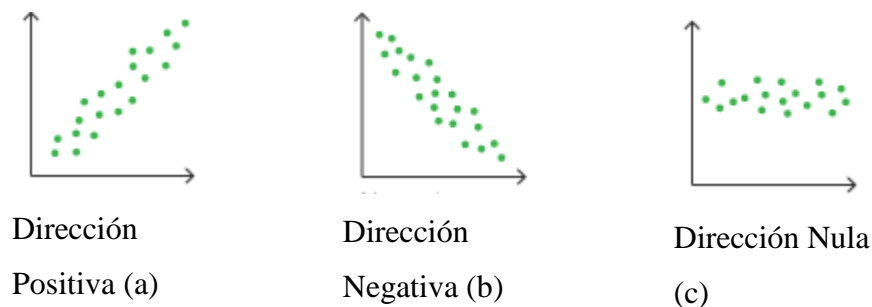
Ocurre una fuerte relación cuando la nube de puntos en un diagrama de dispersión tiende a parecerse en forma global a una línea recta imaginaria (figura 2.2-a), de lo contrario la fuerza es débil (figura 2.2-b). No existe relación, cuando la nube de puntos no tiene ningún patrón y los puntos están dispersos por todo el plano (figura 2.2-c).

Figura 2.2. *Tipos de fuerza de la relación lineal*



- Determinar la dirección de los valores de una variable en relación con los cambios en la otra variable, puede ser positiva, negativa o nula. Hay una relación positiva, si los valores de una variable tienden a aumentar conforme los valores de la otra variable también aumentan (figura 2.3-a); hay una relación negativa, si los valores de una variable tienden a aumentar conforme los valores de la otra variable tienden a disminuir o viceversa (figura 2.3-b). Hay dirección nula, cuando las variables no tienen ninguno de los anteriores comportamientos descritos (figura 2.3-c).

Figura 2.3. *Tipos de dirección de la relación lineal.*

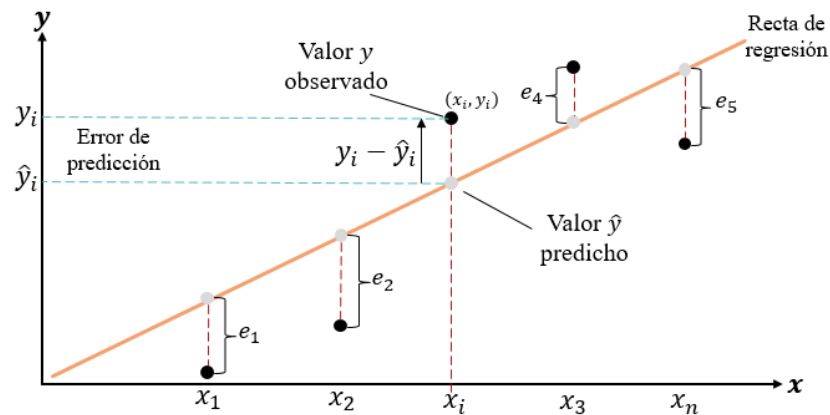


- El coeficiente de correlación r se ve fuertemente afectado por los valores atípicos o puntos extremos que se pueden observar en el diagrama de dispersión porque quedan por fuera del patrón general de la relación.

- Ajustar una línea recta a los puntos en el diagrama de dispersión significa dibujar una línea que se acerque lo más posible a los puntos, cuya métrica que se va a utilizar se deriva del método por mínimos cuadrados, ya que este método elige la línea que hace que la suma de los cuadrados de las distancias verticales entre los valores observados (y) y los valores predichos (\hat{y}) sea lo más pequeña posible.
- La métrica de los mínimos cuadrados escoge el valor de los parámetros de la recta que minimiza la suma de los errores ($e = y_i - \hat{y}_i$) de predicción al cuadrado (ver figura 2.4), es decir,

$$\min \left\{ \sum_{i=1}^n (y_i - \hat{y}_i)^2 \right\}$$

Figura 2.4. Métrica de los mínimos cuadrados



- La regresión lineal describe una relación entre una variable explicativa x y una variable respuesta y , cuyo modelo matemático es representado por $\hat{y} = \hat{b}_0 + \hat{b}_1 x$. Se utiliza \hat{y} en la expresión de la ecuación para enfatizar que la línea da una respuesta predictiva \hat{y} para cualquier x , la cual no será exactamente la misma que la respuesta realmente observada y . Si se denota por n a las parejas de datos $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, el intercepto \hat{b}_0 y la pendiente \hat{b}_1 están dadas por las siguientes fórmulas:

$$\text{Pendiente: } \hat{b}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \quad \text{Intercepto: } \hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x}$$

Para Gal y Garfield (1997) los conceptos y procedimientos matemáticos se utilizan para resolver problemas estadísticos, y se puede esperar cierta facilidad técnica con ellos en ciertos niveles educativos. Pero resaltan, que los cálculos o la ejecución de procedimientos matemáticos se está reemplazando rápidamente por la necesidad de un uso selectivo, reflexivo y preciso de la tecnología cada vez más sofisticada.

- En nuestro estudio se adopta el criterio de cercanía o lejanía entre la nube de puntos y la posible recta de ajuste, donde el procedimiento para encontrar esa medida de cercanía o lejanía se basa en minimizar el valor absoluto de las diferencias entre las ordenadas $|y_i - \hat{y}_i|$; por lo tanto, la recta L que mejor se ajusta a los datos implica determinar:

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

Behar y Grima (2010) abordan la cuestión: ¿Por qué cuando se ajusta una nube de puntos a una ecuación de regresión, se utiliza siempre el criterio de minimizar la suma de los cuadrados de los residuos, y no otros como minimizar la suma de su valor absoluto? Ellos mencionan que minimizar la suma del valor absoluto puede conducir a ajustes no adecuados, sin embargo, aclaran que no es prudente despreciar otros criterios y todo está estrechamente relacionado con el análisis de los cuadrados de los residuos. Shaughnessy y Chance (2005) responden a la pregunta ¿Por qué se elevan al cuadrado las desviaciones?, y también presentan ejemplo en los que las distancias lineales no son convenientes. No obstante, no se presentan estos casos en el problema que analizaron nuestros estudiantes. Utilizamos la distancia lineal para no desviar la atención de los estudiantes precisamente con la pregunta “¿por qué elevar al cuadrado y no tomar la distancia lineal? Quisimos centrar su atención en razonar alrededor de la medida de lejanía o cercanía de la recta a la nube de puntos, lo cual conseguimos. Esto no significa que no apreciemos la importancia de utilizar la suma del cuadrado de los residuales en lugar de la suma de los valores absolutos de las desviaciones, pero creemos que se puede realizar una extensión de la actividad para aclarar este punto, sin menoscabo de la noción de distancia de una recta a una nube de puntos.

2.1.2. Aspectos Estadísticos

- Se hace referencia a datos bivariados cuando se tiene la medición de dos variables de una sola unidad de observación.
- No hay distinción entre variables explicativas y de respuesta. Si se cree que una variable x puede explicar o causar cambio en otra variable y , x es llamada la variable explicativa e y es la variable de respuesta.
- Si dos variables están fuertemente correlacionadas, no implica que exista una relación causa-efecto entre las variables, es decir, la correlación no implica causalidad; aunque puede ser un indicador de su existencia.
- La recta de mejor ajuste es un modelo de la posible relación entre dos variables dadas.
- Los residuos son desviaciones de línea (modelo) que permite evaluar qué tan bien la línea proporciona un modelo para un conjunto de datos bivariados.
- Un valor predicho no es el mismo que el valor real (en la mayoría de los casos).

La noción de agregado, el modelo aleatorio lineal y la incertidumbre son nociones que establecemos dentro del contenido estadístico como parte del razonamiento estadístico covariacional y las cuales los instrumentos diseñados buscan promover.

2.1.2.1. *Definición de agregado en estadística*

Un agregado emerge a la hora de tratar un conjunto de datos como una entidad que tiene características no visibles en ninguno de los elementos individuales (Konold y Higgins, 2002; Mokros y Russell, 1995). Por ejemplo, cuando se ve la media y la mediana como indicadores de propiedades estables de un sistema variable, propiedades que se hacen evidentes solo en el agregado.

Bakker (2004) y Bakker y Gravemeijer (2004) hacen referencia a la vista agregada de datos cuando se perciben como un todo que tiene características agregadas 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

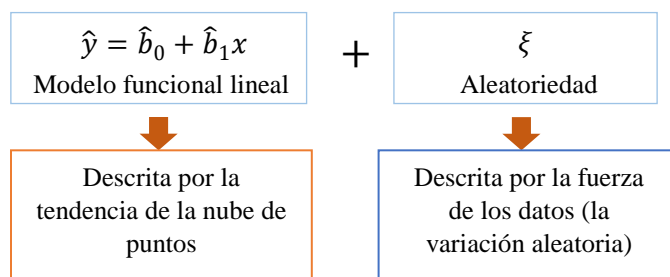
debe estar lo más cercana posible a todos los puntos. En contraste, los estudiantes que no conciben la nube de puntos como un agregado establecen la línea a partir de los puntos que, según su criterio, corresponden a un modelo lineal dejando de lado aquellos que consideran no pertenecen a dicho modelo (Medina et al., 2019).

En relación con el razonamiento acerca del agregado, se ha definido como aquel que se basa en la capacidad de construir valores representativos para grupos (Hancock et al., 1992) y se ocupa de los patrones y relaciones en el conjunto de datos como un todo (Pfannkuch y Wild, 2004). Para estos autores, el 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.

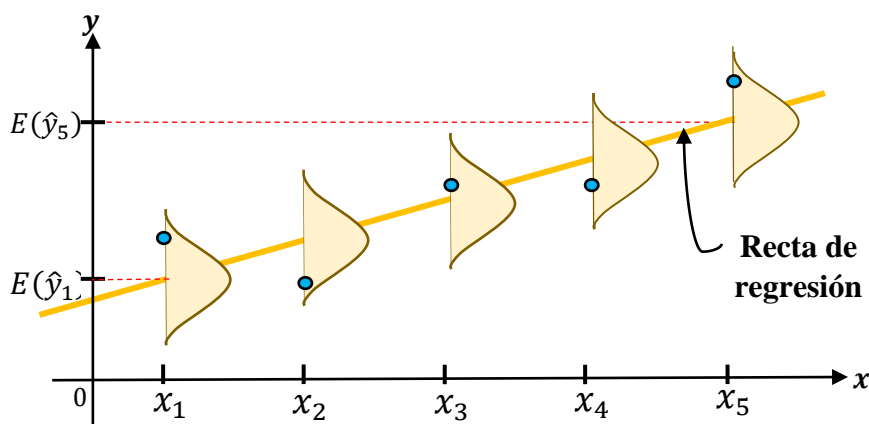
Pensar en un agregado es mucho más potente que pensar en los casos particulares (Stigler, 2016). Un dato individual, aislado, no contiene las propiedades que emergerán cuando se asocia con otros datos, es decir, cuando se trata un conjunto de datos como una entidad. Una condición para que la media de 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. 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.

2.1.2.2. *Modelo aleatorio lineal*

La covariación estadística definida por Batanero (2001) es aquella que implica comprender un modelo aleatorio lineal, el cual se compone de un modelo funcional ($\hat{y} = \hat{b}_0 + \hat{b}_1x$) más un error aleatorio (ξ). Luego, la covariación estadística lineal se puede describir como aquella que se compone de dos elementos (ver figura 2.5): i) un modelo lineal (visto como la descripción de la tendencia global de la nube de puntos), y ii) la aleatoriedad (expresada variación en los datos).

Figura 2.5. Componentes de la covariación estadística

Para profundizar más en el error aleatorio, supóngase que este componente aleatorio que es el error ξ tiene media 0 y varianza σ^2 . Dado que la variabilidad de y en un valor particular de la variable explicativa x está determinada por la varianza del error σ^2 implica que existe una distribución de valores de y para cada x , y que la varianza de esta distribución es la misma para cada x (Montgomery y Runger, 1996), esto se puede representar en la figura 2.6, donde la media de la distribución del error corresponde a la esperanza del valor \hat{y} predicho, como se muestra en la figura los valores de x están por encima y por debajo de la media de cada distribución, además como supuesto del modelo se requiere que los residuos sean independientes. Por lo tanto, cuando se establece la fuerza de los datos, es decir, la variación entre ellos, de algún modo se está pensando en los residuales presentes en los datos que se conoce como el componente aleatorio de la covariación estadística.

Figura 2.6. Modelo lineal simple (distribución de y para un valor dado de x)

2.1.2.3. Incertidumbre

La estadística y la probabilidad son las ciencias que tratan con la incertidumbre y con la variación en procesos naturales y provocados por el hombre (Moore, 1990). Para definir la incertidumbre hacemos referencia a la siguiente afirmación de Kahneman y Tversky (1982):

“La incertidumbre es un hecho con el que todas las formas de vida deben estar preparadas para enfrentarse. En todos los niveles de complejidad biológica existe incertidumbre sobre el significado de los signos o estímulos y sobre las posibles consecuencias de las acciones. En todos los niveles, se deben tomar medidas antes de que se resuelva la incertidumbre, y se debe lograr un equilibrio adecuado entre un alto nivel de preparación específica para los eventos que tienen más probabilidades de ocurrir y una capacidad general para responder adecuadamente cuando sucede lo inesperado” (p. 508).

De lo anterior se puede decir que la incertidumbre es un aspecto omnipresente de toda experiencia humana, por lo que razonar a partir de datos empíricos inciertos es un método intelectual igualmente poderoso y generalizado (Moore, 1990). Se asume que la variación estadística presente en los datos que no puede ser explicada es incertidumbre (Tal, 2001). Cuando hay presencia de variabilidad, usualmente los estudiantes suelen utilizar razonamientos determinísticos (afirmaciones de solamente certeza) y difícilmente logran juicios sobre patrones y tendencias que involucran incertidumbre (Ben-Zvi et al., 2012). También se han identificado dos tipos de incertidumbre: la incertidumbre contextual, presente cuando ocurre un conflicto entre el conocimiento del contexto de los estudiantes y los datos, y la incertidumbre estadística, cuando proviene de la variabilidad en los datos (Manor et al., 2014). No trataremos aquí con la *incertidumbre subjetiva* (Kahneman y Tversky, 1982), que proviene de deficiencias de conocimientos del individuo; este tipo de incertidumbre es muy importante, pero en nuestro trabajo es suficiente considerar la *incertidumbre objetiva*.

2.2. Objeto de estudio: Razonamiento estadístico covariacional

Inicialmente, se aborda el pensamiento y la alfabetización estadística para diferenciarlos del razonamiento estadístico, seguido se aborda el razonamiento estadístico informal, para finalmente, definir nuestro objeto de estudio, el razonamiento estadístico covariacional con enfoque informal.

Para definir el *razonamiento estadístico*, es preciso diferenciarlo de la alfabetización estadística y del pensamiento estadístico. La *alfabetización estadística* consiste en promover

la comprensión de las estadísticas lo suficientemente bien como para poder consumir la información en la que están inundados diariamente, pensar críticamente al respecto y tomar buenas decisiones basadas en esa información (Rumsey, 2002). Por otro lado, el *pensamiento estadístico* incluye lo que hace un estadístico al ir más allá de lo que enseña en el curso, para cuestionar e investigar los problemas y los datos involucrados en un contexto específico. Un pensador estadístico ve el proceso estadístico como un todo (con iteración), comprende la relación y el significado de la variación en este proceso para tener la capacidad de explorar datos más allá de lo prescrito en los textos, y generar nuevas preguntas más allá de las formuladas por un investigador principal (Chance, 2002). Respecto al *razonamiento estadístico*, Garfield (2002) menciona que puede definirse como la forma de razonar de las personas acerca de las ideas estadísticas y puede implicar conectar un concepto con otro, así como, explicar procesos e interpretar resultados estadísticos. Pfannkuch y Wild (2004) agregan que el razonamiento implica hacer interpretaciones basadas en conjuntos de datos, representaciones de datos o resúmenes estadísticos de los datos. Es decir, razonar significa comprender y ser capaz de explicar procesos estadísticos y poder interpretar completamente los resultados estadísticos. delMas (2002) propone cuestiones como: ¿Por qué o cómo se produjeron los resultados?, cuando se está interesado en el razonamiento estadístico, es decir, con preguntas que llevan a explicar el proceso; por ejemplo, cuando se busca explicar el proceso que produce la determinación de la línea de regresión. Existe una discusión amplia para llegar a un acuerdo formal y lograr distinguir el razonamiento estadístico de los demás dominios (pensamiento y alfabetización estadística). Por lo tanto, decidimos establecer nuestra propia definición de razonamiento estadístico partiendo de los siguientes conceptos:

- *El razonamiento matemático.* Es el proceso de establecer conclusiones con base en evidencia o suposiciones declaradas, específicamente, el razonamiento matemático es aquel que va desde la explicación informal y la justificación hasta la deducción formal, así como las observaciones inductivas (NCTM, 2009, p.4).
- *Las proposiciones estadísticas.* Se define como los enunciados o afirmaciones basados en datos o inferencias y que incluyen aspectos de incertidumbre.

Con lo anterior, definimos el *razonamiento estadístico* como el proceso de generar conclusiones deductivas, inductivas o tácitas presentes en los conceptos matemáticos y procesos de análisis estadísticos que se desarrollan para verificar o hacer plausibles proposiciones estadísticas. El razonamiento estadístico se evidencia a través de argumentos válidos, coherentes y con sentido que se eligen para tomar una decisión y respaldar o no la plausibilidad de las proposiciones. Como nuestro interés es el razonamiento covariacional, los procesos estadísticos corresponden a los involucrados en la correlación y regresión lineal mediante proposiciones basadas en datos bivariados.

Ahora bien, para la NCTM (2009), el razonamiento requiere niveles crecientes de comprensión: primero, los esquemas de acción resultados de la experiencia y la práctica, segundo, las explicaciones intuitivas y argumentaciones parciales de las ideas, y el tercer nivel correspondiente al formal, donde las argumentaciones son a partir de una deducción de axiomas, definiciones y postulados. Nuestro interés son los dos primeros niveles de comprensión, los cuales comprenden un enfoque informal del razonamiento matemático, donde las intuiciones y las concepciones influyen en el proceso de razonar.

Para Fischbein (1997), la dinámica del razonamiento matemático, y en general de todo tipo de razonamiento científico, incluye diversas componentes psicológicas como creencias y expectativas, imágenes inmediatas, analogías y paradigmas. Son ingredientes activos, genuinamente productivos de todo tipo de razonamiento. Así, como las formas intuitivas de razonamiento influyen en la manera de resolver e interpretar, por ello, las intuiciones son siempre un producto de la experiencia personal, de la implicación personal del individuo en determinada actividad teórica o práctica. Respecto a las concepciones, Confrey (1990) afirma que las ideas sobre el mundo, las creencias, teorías, los significados, explicaciones y estrategias forman la base del uso del término concepciones del estudiante. Para diferenciar las ideas concepto y concepción McDonald et al. (2000) realizaron una distinción entre estas: la concepción es intrapersonal, es decir, es la idea o comprensión individual; y el concepto es comunal, es decir, un concepto acordado por los matemáticos (p.78).

2.3. Posición de la enseñanza

Asumimos que la enseñanza de las matemáticas y de la estadística ofrece crear buenas oportunidades para el aprendizaje cuando está basada en la resolución de problemas, en el trabajo en equipo e integra recursos tecnológicos. No obstante, hay muchas formas de concretar esta posición general y su éxito o fracaso dependen de las formas concretas en las que se instrumenta. En el curso del presente trabajo se mostrará cómo se concretó esta postura en esta investigación. Además, abajo describimos algunos rasgos generales de las componentes que configuran dicha postura.

2.3.1. Uso de la tecnología

En este estudio, se asume 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.

La National Council of Teacher of Mathematics establece que la existencia, versatilidad, y poder de la tecnología hace posible y necesario el reexaminar qué matemáticas los estudiantes deben aprender, así como la manera de aprender (NCTM, 2000). Para la correlación y regresión en último grado de bachillerato, en los estándares se especifica:

Con relación a los datos bivariados, los estudiantes deberían ser capaces de representarlos mediante una nube de puntos, de describir su forma, y de determinar los coeficientes de regresión, las ecuaciones de regresión y los coeficientes de correlación, utilizando herramientas tecnológicas (NCTM, 2000, p.409).

Específicamente, para las temáticas de regresión y correlación la tecnología posee los siguientes rasgos relevantes:

3. La posibilidad para formar 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.
4. Obtener el valor numérico del coeficiente de correlación y determinar la expresión algebraica de la recta de regresión.
5. 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).

2.3.2. Resolución de problemas

Se acepta la hipótesis que conviene y es posible favorecer el desarrollo del razonamiento matemático de los estudiantes, cuando se comprometen en la solución de problemas matemáticos (Francisco y Maher, 2005) con el fin de alcanzar un objetivo (resultado matemático) cuyo camino no se conoce de antemano. Dado que el interés es estudiar el razonamiento de los estudiantes a enfrentar problemas sobre correlación y regresión, los problemas deben contener características que ayuden al desarrollo de los conceptos y las formas en que estos conceptos se utilizan en situaciones de la “vida real” más allá de la escuela; por lo tanto, se asume la siguiente definición de *resolución de un problema*:

Una tarea, o actividad dirigida a un objetivo, se convierte en un problema (o problemática) cuando el "solucionador de problemas" necesita desarrollar una forma más productiva de pensar sobre la situación dada (Lesh y Zawojewski, 2007, p. 782).

Para Lesh y Zawojewski (2007) la resolución de un problema es definido como el proceso de interpretación de una situación matemática, la cual implica varios ciclos iterativos de expresiones, pruebas y revisiones de interpretaciones matemática, así como el clasificar, integrar, modificar o refinar conceptos matemáticos de varios temas, donde la intención es ir más allá de la búsqueda de un procedimiento para llegar a un objetivo. Por lo tanto, se entenderá por resolución de problema a situaciones matemáticas que implican interpretar, describir y explicar; no simplemente ejecutar reglas, procedimientos o habilidades de manera experta, lo cual, se traduce como un ciclo interactivo de comprensión que buscan dar sentido

a situaciones de la vida real, donde se espera que los estudiantes aporten su propio significado personal a un problema, que prueben y revisen su interpretación durante el desarrollo de modelos matemáticos buscando que simultáneamente lleguen a una comprensión cada vez mayor de la situación del problema.

2.3.3. El contexto

Moore (1990) afirma “los datos no son simplemente números, sino números con un contexto”. El contexto juega un rol importante en las situaciones problemas que involucran datos reales, dado que puede generar una motivación y enganchar a los estudiantes donde están comprometiendo su conocimiento del contexto para interpretar y analizar más que simplemente realizar operaciones matemáticas (Moore, 1990).

Pfankuch, (2011) menciona que es necesario distinguir cómo se usa el contexto en la resolución de problemas estadísticos y cómo se ve el contexto en la resolución de problemas estadísticos. Pfankuch (2011) propone dos tipos de contextos que se deben diferenciar: el contexto de los datos y el contexto de la experiencia de aprendizaje.

El *contexto de datos* hace referencia al mundo real, a la situación de la que surgió el problema, el conocimiento de cómo se generaron los datos, el tipo de diseño del estudio y las variables que están definidas.

El *contexto de la experiencia de aprendizaje* es analizar los antecedentes que poseen los estudiantes para aportar a la solución del problema, donde sus conocimientos previos influyen en la interacción con el problema que pueden generar nuevas ideas y concepciones.

2.3.4. Trabajo colaborativo

Se asume la posición que el trabajo colaborativo entre pares puede ser motivador e interesante para los estudiantes, el trabajar juntos contribuye a maximizar su propio aprendizaje y el aprendizaje del otro, ya que el compromiso y el interés por la manera como los objetivos de los estudiantes se vinculan entre sí, ya no dependen de la curiosidad individual (Roseth et al., 2008); en esta investigación se traduce ese aprendizaje en aprender a razonar, por lo que se considera que trabajar en parejas implica un esfuerzo tanto individual

como grupal para razonar. Se es consciente que es posible que algunos estudiantes no participen y dejen que su compañero haga todo el trabajo, una forma de comprometer a los estudiantes sugerida por Roseth et al. (2008) es que los individuos tienden a buscar resultados que son beneficiosos para todos aquellos con quienes están vinculados de manera cooperativa; por lo tanto, proponen que, finalizado el primer trabajo en parejas, formar nuevos grupos donde deben comparar, explicar y apoyar sus respuestas.

Dentro de las características o pasos para llevar a cabo una lección mediante el trabajo colaborativo descritas por Roseth et al. (2008) están: (a) tomar decisiones preinstructivas sobre la lección, (b) explicar la tarea y la estructura del trabajo en parejas a los estudiantes, (c) monitorear y, si es necesario intervenir con cada pareja y (d) procesar y evaluar el rendimiento de los estudiantes. Respecto al papel del profesor, es importante mencionar que dentro del contexto social su labor es monitorear, coordinar y dar sentido a las interacciones propias de los estudiantes, así como entre los estudiantes y las herramientas tecnológicas para facilitar la resolución de problemas de los alumnos, donde se pretende que los estudiantes se centren en el proceso estadístico que precede a los cálculos y la interpretación de los resultados de estos cálculos, dejando a un lado su papel de catedrático y la enseñanza formal del tema.

CAPÍTULO III. MARCO METODOLÓGICO

El presente estudio es una investigación de diseño con estudiantes de bachillerato sobre el tema de regresión y correlación con apoyo de recursos tecnológicos. A continuación, se detallan los aspectos metodológicos que guiaron la investigación. Como se mencionó en el planteamiento del problema el objetivo es construir una teoría sustantiva sobre las posibilidades de los estudiantes de bachillerato para razonar acerca de la correlación y regresión al enfrentarse a tareas instruccionales haciendo uso de la tecnología digital; esto implica un diseño para la observación del razonamiento de los estudiantes. Este diseño se llevó a cabo con base en los siguientes elementos:

- *El tipo de investigación:* corresponde a un experimento de diseño basado en los cinco aspectos del entorno del aula de los de los "principios del diseño instruccional" de Cobb y McClain (2004).
- *El método para la recolección de datos:* Son los aspectos referentes a los participantes, instrumentos y procedimientos para recolectar la información.
- *Metodología para el análisis de datos:* Se basa en la filosofía, principios y procedimiento de la teoría fundamentada, cuyo objetivo es elaborar teorías sustantivas que emerjan de los datos y no de deducciones lógicas o de otras teorías.

El capítulo se divide en tres apartados, cada uno describe uno de los elementos metodológicos.

3.1. Tipo de investigación: El experimento de diseño

Un experimento de diseño es un método de investigación que tiene dos aspectos: a) El diseño e implementación de una experiencia de aprendizaje para los estudiantes y b) El estudio sistemático de la relación entre los elementos del diseño y el aprendizaje logrado por los estudiantes.

Los investigadores Cobb et al. (2003) se refieren a los experimentos de diseño de la siguiente manera:

“De manera prototípica, los experimentos de diseño implican tanto la "ingeniería" de formas particulares de aprendizaje como el estudio sistemático de esas formas de aprendizaje dentro del contexto definido por los medios para apoyarlos” (p. 9).

De la anterior afirmación se deduce que un experimento de diseño es un método de investigación, cuyo propósito es producir conocimientos acerca de cómo aprenden los estudiantes gracias a los elementos de diseño. Pero en el proceso de observación de lo que ocurre en la implementación y del análisis de los datos también se producen conocimientos acerca de las dificultades o falsas concepciones que frenan u obstaculizan el aprendizaje de los estudiantes, así como sobre efectos no deseados del diseño. Como consecuencia todos los conocimientos adquiridos ayudan a perfeccionar el diseño y repetir el experimento.

Un experimento de diseño busca desarrollar una teoría humilde sobre procesos de aprendizaje de un dominio específico (ej. el aprender a razonar acerca de la recta de mejor ajuste) (Cobb y McClain, 2004). El experimento de diseño pretende ir más allá de explicar por qué el diseño tiene o no éxito, y se pretende explicar el desarrollo del razonamiento de los estudiantes en relación con los medios que se diseñan para apoyar ese razonamiento.

Este proyecto propone un experimento de diseño en el dominio específico del razonamiento de estudiantes de bachillerato con algunas ideas claves de la correlación y la regresión en estadística y el apoyo de recursos tecnológicos. Para el diseño se considera el marco de los cinco aspectos del entorno del aula de los “principios del diseño instruccional” de Cobb y McClain (2004):

- i. El enfoque sobre ideas estadísticas centrales
- ii. Las tareas instruccionales
- iii. Estructura de las tareas en el salón de clase
- iv. El uso de herramientas computacionales
- v. El discurso en la clase

A continuación, se plantean cada uno de los anteriores aspectos para el experimento de diseño que se desarrolló en nuestro estudio.

3.1.1. 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. En los temas de regresión y correlación no se avanza mucho

en el desarrollo de las ideas centrales si los estudiantes solo aprenden a calcular un coeficiente de correlación, o a determinar una recta de mejor ajuste. Es necesario que se entiendan las problemáticas científicas relacionadas con ideas generales como la incertidumbre, los errores y la predicción, así como la relación de estas ideas con la regresión y la correlación.

Una idea central es una idea en torno a la cual se articula un sistema de conceptos y que se puede formular a diferentes niveles de elaboración simbólica. Frecuentemente una idea central es un objetivo de aprendizaje al que se quiere llegar desarrollando varias ideas y recursos.

En el tema de correlación y regresión hemos identificado tres ideas centrales para el razonamiento covariacional: 1) la idea acerca de los datos bivariados y su representación gráfica, 2) la idea de correlación y, 3) la idea de recta de mejor ajuste.

Los datos bivariados y sus diagramas de dispersión presentan los siguientes componentes:

- *Representación gráfica.* Es un conjunto de pares de valores numéricos (x, y) que se pueden representar en el plano cartesiano en un diagrama de dispersión, donde se puede observar la tendencia y los patrones de una nube de puntos.
- *Fuerza y dirección.* Es la representación de los datos en puntos de un plano bidimensional. Analizar el diagrama de dispersión implica describir la dirección (positiva, negativa o nula) y la fuerza (fuerte o débil) de la nube de puntos.
- *Tendencia y ajuste de modelo.* Gráficamente los puntos no se ajustan de un modo preciso a una función matemática (modelo), sino que se obtiene un conjunto de puntos más o menos dispersos. No obstante, frecuentemente se agrupan en la proximidad de una línea o curva. Las diferencias respecto a la dimensión vertical que separan a los datos de los puntos de un modelo (línea o curva) son los residuos o errores.

Los componentes de la idea central de correlación definidos son:

- *Coefficiente de correlación.* El valor del coeficiente de correlación está entre 0 y 1 o entre -1 y 0. Si se acerca a 0, concluimos que no existe una correlación lineal entre las variables, pero si el valor se acerca a 1 o -1, concluimos que hay una correlación lineal fuerte entre las variables.
- *Regresión.* Si las dos variables están fuertemente correlacionadas, una línea de regresión resume la relación, solo cuando una de las variables ayuda a explicar o predecir la otra.
- *Correlación no es causación.* Si dos variables están fuertemente correlacionadas no implica necesariamente que exista una relación causa-efecto entre las variables, aunque puede ser un indicio de su existencia.

Se han establecido los siguientes componentes relacionados con la idea de recta de mejor ajuste:

1. 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.
2. 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, $e = |y_i - \hat{y}_i|$.
3. La medida de cercanía o lejanía entre la nube de puntos y la recta de ajuste se calcula sumando todos los residuales, (le llamaremos $R(l)$ indicando que son los residuales de la línea recta), entonces:

$$R(l) = \sum_{i=1}^n |y_i - \hat{y}_i|$$

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

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

Es decir, L es tal que:

$$R(L) = \min_{l \in R^2} R(l)$$

Este concepto es difícil porque los “valores” de la variable son las rectas en el plano.

Se utilizó una definición de distancia lineal y no cuadrática como suele utilizarse en estadística porque pensamos que es más intuitiva para los estudiantes. Las dificultades de cálculo que implica el manejo de las barras del valor absoluto no son significativas en este caso pues los cálculos se realizan automáticamente por el software. Los resultados que se obtienen con la distancia lineal son muy aproximados a los que se obtienen con la distancia cuadrática.

3.1.2. Uso de la tecnología

Respecto al uso de la tecnología se utiliza la plataforma gratuita *CODAP* (<https://codap.concord.org/>) diseñada específicamente para la enseñanza y aprendizaje de la estadística. Esta reciente plataforma de código abierto creada para explorar y aprender de los datos de cualquier área de contenido, desde matemáticas y ciencias hasta estudios sociales o educación física. *CODAP* es apropiado para analizar la correlación y la regresión por rasgos relevantes como:

- (i) Arrastrar y soltar variables rápidamente dentro de un gráfico para visualizar distribuciones y relaciones entre las variables.
- (ii) Los diagramas de dispersión y la curva de mejor ajuste de grandes conjuntos de datos se pueden construir y manipular de forma rápida y flexible.
- (iii) La vinculación de múltiples representaciones que permite descubrir y observar patrones y tendencias en los datos.
- (iv) Al agregar o quitar datos se puede visualizar como las representaciones y las medidas de los parámetros cambian dinámicamente en tiempo real.
- (v) Calcular el coeficiente de correlación y determinar la expresión algebraica de la línea de regresión, así, como calcular la desviación estándar.

Además, se utiliza el software GeoGebra en el diseño de una tarea, pues sus características permiten realizar el dibujo dinámico de los elementos que consideramos relevantes para que el estudiante observe e interactúe. Específicamente, para la recta 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. Más adelante se explica las características del diseño de la tarea con uso de GeoGebra.

Como se menciona en el marco conceptual se utiliza la tecnología con el objetivo de enfocarse 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 (Biehler et al.2013). Se propone el uso de la tecnología en la clase, no con el objetivo de evaluar la habilidad que tiene el estudiante para utilizar determinado software, sino sobre la habilidad que tiene para utilizar el software en el análisis de datos y su capacidad de interpretación sobre los resultados que arroja. El profesor debe ser consciente de la extensión que representa la tecnología digital para el desarrollo del razonamiento del estudiante y también para su propio razonamiento como experto en el tema. Santos-Trigo et al. (2016) argumentan que el uso de la tecnología digital ofrece oportunidad tanto a profesores y estudiantes de ampliar y profundizar las formas de razonar sobre las estrategias matemáticas para resolver problemas. Así como el profesor adquiere o tiene autonomía para usar la tecnología digital al resolver un problema, podrá darse cuenta el momento en el que el estudiante se vuelve autónomo en el manejo del software para resolver un problema estadístico y la manera como logra articular lo arrojado por el software con el lenguaje estadístico para comunicar y expresar resultados.

3.1.3. El Discurso en el aula

Cobb y McClain (2004) también sugieren tener en cuenta el discurso en el aula, esto implica prever cómo dirigirse a los estudiantes y diseñar una estrategia para introducir la nueva terminología y lenguaje que se quiere que vayan adquiriendo. En la presente investigación hemos previsto que los estudiantes se enfrenten por primera vez con términos como diagrama de dispersión, dirección y fuerza de la relación entre dos variables, correlación, modelo y recta de mejor ajuste, además de términos técnicos propios del contexto de los problemas.

Las tareas se han diseñado con el objetivo que el estudiante incluya los términos al hacer inferencias, por lo tanto, dentro de la tarea se muestran pequeños recuadros con algunas definiciones de los términos que se quiere vayan adquiriendo, estos recuadros se muestran seguido de una pregunta que busca articular los términos al contexto del problema. Sin embargo, el profesor puede intervenir para generar una discusión entre los estudiantes acerca

de algún término, sin llegar a definir formalmente el concepto, siempre basado en la situación del problema y de generar la discusión. También conviene que el profesor ayude a los estudiantes a expresar mejor sus ideas sugiriéndoles utilizar el término técnico cuando vea que los estudiantes quieren referirse a la idea correspondiente. En el subapartado actividades se muestra cómo se incluye el lenguaje y los términos dentro del desarrollo de las tareas.

3.1.4. Las actividades y su estructura

Atender a la estructura de las actividades en el salón de clase significa definir y justificar las acciones del profesor y la manera en que los estudiantes llevarán a cabo las tareas previstas. Debe organizarse la administración de las tareas 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.

3.2. Método de recolección de datos

En este apartado se describen los criterios establecidos para los sujetos de estudio, instrumentos y procedimientos con el fin de realizar la recolección de los datos.

3.2.1. Sujetos

En el estudio participaron estudiantes de 4° 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 (4° 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 5° 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 exploración con estudiantes de 4° 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 4° 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 el que se llevó a cabo la exploración se caracteriza por dar una gran libertad a los profesores para diseñar sus cursos sin dar directrices que fueren 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.

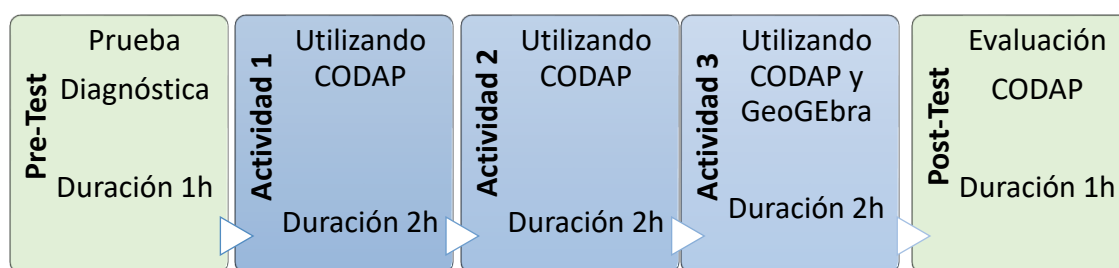
El curso fue mixto (hombres y mujeres) y de jornada matutina, la aplicación de los instrumentos se realizó en 5 clases de dos horas en enero del 2020.

3.2.2. Instrumentos y recolección de datos

Los instrumentos constan de un cuestionario inicial o diagnóstico, un cuestionario final o de evaluación y una secuencia de tres actividades. El cuestionario diagnóstico (ver apéndice F) tiene como objetivo conocer las concepciones iniciales de los estudiantes, y consta de dos partes: i) ítems que involucran cuestiones relacionadas con álgebra básica, y ii) ítems de evaluación, para explorar la forma de razonar de los estudiantes frente a problemas de correlación y regresión. El cuestionario final o evaluativo (ver apéndice G) tiene como objetivo obtener información sobre el avance de los estudiantes respecto a las ideas de correlación y regresión.

Para el diseño de la secuencia de actividades utilizando tecnología digital se establecieron tres sesiones de clase (figura 3.1). Las situaciones problema presentadas en las actividades poseen un contexto general, el cual es la información nutricional sobre productos de comida rápida, en especial, la medición de su aporte calórico. La figura 3.1 resume la secuencia de aplicación de los instrumentos.

Figura 3.1. *Secuencia de los instrumentos y duración de la aplicación.*



La obtención de datos fue por medio de hojas de trabajo para desarrollar en parejas. El cuestionario diagnóstico y final está diseñado para desarrollarse de manera individual. La aplicación de la secuencia de actividades se realizó en una sala de ordenadores, donde los estudiantes se ubicaron en parejas, y se mantuvieron conformadas las mismas parejas durante el desarrollo de las tres actividades.

Las hojas de trabajo muestran pequeños recuadros con algunas definiciones de los términos que se quiere los estudiantes vayan adquiriendo, estos recuadros se muestran seguido de una pregunta que busca articular los términos al contexto del problema, por ejemplo:

Figura 3.2. *Ejemplo de recuadro con la definición de variable respuesta y variable explicativa*

Variable de Respuesta y Variable Explicativa
Una variable explicativa puede ayudar a predecir o explicar los cambios en la variable de respuesta. Los datos se pueden representar mediante los pares ordenados (x, y) donde x es la variable independiente (o explicativa) e y es la variable dependiente (o respuesta).

Discute con tu compañero y responde:

¿Cuál es la variable de respuesta? _____ ¿Por qué?

¿Cuál es la variable explicativa? _____ ¿Por qué?

A continuación, se realiza una síntesis de cada actividad junto con su procedimiento de aplicación, la estructura de las tareas y el discurso por parte del profesor en clase. Cabe aclarar, que el formato completo de los instrumentos entregados a los estudiantes se encuentra en los apéndices B, C y D.

3.2.2.1. *Actividad 1.*

Objetivos de aprendizaje:

- Distinguir que entre dos variables puede existir una relación.
- Construir un diagrama de dispersión para describir el comportamiento (dirección y tendencia) de una relación bivariada.
- Descubrir que la recta es la función que mejor modela el conjunto de datos.
- Construir la línea de mejor ajuste definida como línea que se acerca lo más posible a todos los puntos.
- Desarrollar y evaluar estimaciones y predicciones basadas en datos.

Objetivo de investigación: Queremos encontrar evidencia acerca de los razonamientos que surgen en los estudiantes al crear y visualizar en un diagrama de dispersión un conjunto de datos bivariados utilizando tecnología, así como las estrategias previas que proponen para establecer la recta de mejor ajuste y estimar y predecir dentro del contexto de la situación.

Supuestos de aprendizaje:

Supuestos favorables:

- Describen la tendencia de la nube de puntos y su dirección.
- Establecen la idea informal que una relación se dice que es lineal si los datos en la nube de puntos se conglomeran formando una línea.

Supuestos que obstaculizan:

- La dependencia del contexto del problema, donde es difícil para los estudiantes separar sus creencias a priori temporalmente para observar los datos y evaluar la relación entre las variables (Moritz, 2004; Confrey, 2000).

- La concepción local de la asociación (basan su juicio en solo una parte de los datos representados) y la concepción determinista de la asociación (la relación entre las variables es funcional) (Batanero et al. (1996)).
- Surge el criterio de establecer una recta que contenga la mayor cantidad de puntos.
- Proponen un procedimiento aritmético o algebraico para estimar o predecir.

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>).

Situación: Quizás, oyes o creas que las grasas y las calorías son dañosas 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 3.1 muestran la cantidad de grasa en gramos y el número de calorías en algunos productos de McDonald's y Burger King.

Tabla 3.1. 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

A lápiz y papel: Observa los datos de la tabla 3.1 ¿Cuál crees que es la relación entre los gramos de grasa y las calorías? Explica tu respuesta.

- Realiza una gráfica de los datos de la tabla 3.1 en el plato Cartesiano. Etiqueta los ejes con la variable correspondiente.
- ¿Cómo harías para estimar el número de calorías de un producto que contiene 31 gramos de grasa? Explica tu respuesta.
- ¿Cómo harías para predecir el número de calorías de un producto que contiene 63 gramos de grasa? Explica tu respuesta.

Figura 3.3. Recuadro variable respuesta y variable explicativa.

Variable de Respuesta y Variable Explicativa
Una variable explicativa puede ayudar a predecir o explicar los cambios en la variable de respuesta. Los datos se pueden representar mediante los pares ordenados (x, y) donde x es la variable independiente (o explicativa) e y es la variable dependiente (o respuesta).

Discute con tu compañero y responde:

¿Cuál es la variable de respuesta? _____ ¿Por qué?

¿Cuál es la variable explicativa? _____ ¿Por qué?

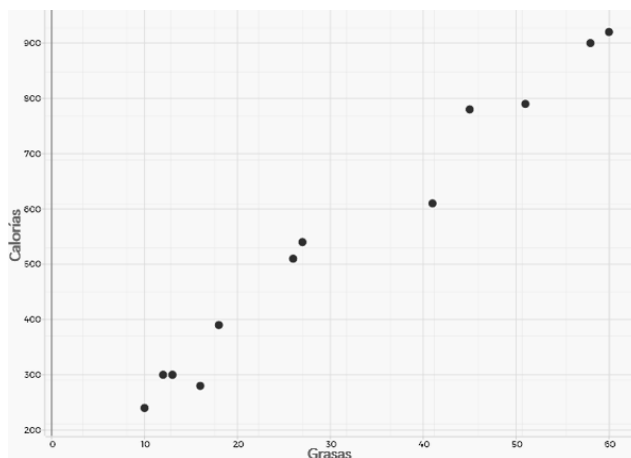
Utilizando tecnología digital: Abrir la plataforma digital *CODAP* y crear el diagrama de dispersión de los datos bivariados (grasa, calorías) presentados en la tabla 3.1. Sigue las indicaciones del profesor.

- ¿Qué características se podrían tener en cuenta al interpretar un diagrama de dispersión?

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

A lápiz y papel: La figura 3.4 muestra el diagrama de dispersión de los datos bivariados (Grasa, Calorías). Dibuja la línea que mejor se ajusta a los datos. Explica el criterio que utilizaste para trazarla.

Figura 3.4. *Diagrama de Dispersión de la Grasa y Calorías*



Utilizando tecnología digital: Sigue las indicaciones del profesor y en *CODAP* determina la línea de regresión. Compara la línea que dibujaste con la línea que muestra el software o la plataforma.




- ¿Cuál crees que es el criterio que utiliza el programa para determinar la línea de regresión?

Escribe cómo utilizarías la línea de regresión que arrojó el programa para contestar las preguntas:

- ¿Estimar el número de calorías de un producto que contiene 31 gramos de grasa?
- ¿Predecir el número de calorías de un producto que contiene 63 gramos de grasa?

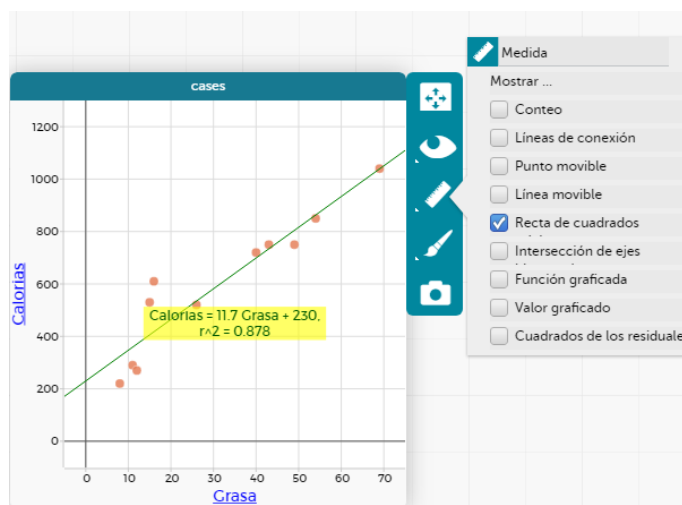
Estructura de aplicación de la actividad 1. El tiempo de aplicación del instrumento fue de 2 horas. La primera parte a lápiz y papel tuvo una duración de 20 minutos, la segunda parte utilizando CODAP tuvo una duración de unos 40 minutos, la última parte a lápiz y papel junto con tecnología tuvo una duración de una hora.

El profesor entregó la situación problema y orientó a los estudiantes a discutir acerca de la relación entre las variables, e indicándoles que escriban todo lo que ellos piensan de manera clara y extensa a lápiz y papel (sin uso de tecnología digital). Seguido, el profesor indicó cómo abrir la plataforma CODAP facilitándoles a los estudiantes el enlace de la plataforma <https://codap.concord.org/releases/latest/static/dg/es/cert/>. Y guiándolos con los siguientes pasos para tabular los datos y graficarlos:

- Seleccionar “crear nuevo documento”. El profesor mostró como ingresar los datos de la tabla 1 en la plataforma. De la siguiente manera:
- Ir a la opción tabla:  Tablas Crear nuevo conjunto de datos.
- Da clic en la palabra Atributo y en las opciones seleccionar editar propiedades del Atributo, seguido en nombre escribir Grasa y aplicar.
- En la celda debajo del nombre Grasa empezar a digitar los datos de la situación. Dando un enter cada vez que se digita un valor.
- Al lado de la palabra Casos aparece un símbolo “+” dar clic y agregar el Atributo Calorías. Seguido en la columna calorías digitar los datos correspondientes.
- Seleccionar la opción “gráfico”  Gráfico . El profesor debe aclarar que aún no están representados los datos de la tabla en el gráfico por que aún no aparecen los ejes ni la escala. Debe explicar la forma como se pueden arrastrar de la tabla cualquier de los atributos al eje X o al eje Y. Dejando que lo estudiantes arrastren a su criterio las variables. Antes de arrastrar las variables, pide a los estudiantes dar clic en el centro del gráfico, aparecerán unos iconos al lado, seleccionar el icono formato  y reducir un poco el tamaño de los puntos.

Para indicar cómo en CODAP hacer aparecer la línea de regresión se muestra al lado del diagrama de dispersión un menú, se debe seleccionar el icono de la regla que indica medida y dar clic en recta de cuadrados (ver figura 3.5, esta figura no se muestra en la hoja de trabajo del estudiante).

Figura 3.5. Captura de pantalla del diagrama de dispersión y la recta de regresión realizado en CODAP



3.2.2.2. Actividad 2.

Objetivo de aprendizaje:

- Valora, con apoyo de la computadora, el comportamiento del coeficiente de correlación.
- Entender que la correlación mide la fuerza y dirección de la relación lineal dentro del contexto del problema.

Objetivo de investigación: Con apoyo de la tecnología que permite el cálculo del coeficiente de correlación junto con la creación del diagrama de dispersión, queremos conocer los razonamientos que surgen al relacionar el diagrama de puntos con su correspondiente medida de correlación dentro del contexto del problema.

Supuestos de aprendizaje:

Supuestos favorables

- Utilizan la medida del coeficiente de correlación y el diagrama en forma conjunta (Sánchez-Cobo et al., 2000).
- De manera informal describen la medida de la correlación mediante el examen de cómo se acerca los puntos a una línea o que tan dispersos están los puntos.

- Realizan comparaciones globales entre las dos variables, explicando el crecimiento en la forma del diagrama de dispersión (Estepa y Batanero,1996).

Supuestos que obstaculizan

- No relacionan el valor de la correlación con la dirección y la fuerza de los datos representados en el diagrama de dispersión dentro del contexto del problema.

Para la segunda actividad se utilizó la información nutricional de grasa, carbohidratos, azúcar y calorías de 6 productos de McDonalds y 6 productos Burger King. Estos datos fueron descargados de las páginas oficiales de las marcas.

Situación: El aporte calórico que contiene un producto depende de muchos nutrientes entre ellos, la grasa, los carbohidratos, el azúcar, la proteína, el sodio y otros. La tabla 3.2 muestra las mediciones de algunos de los nutrientes que influyen en el aporte calórico: la grasa, carbohidratos, azúcar y proteína en gramos que contiene algunos productos de McDonald's y Burger King junto con la cantidad de calorías para cada uno.

Tabla 3.2. *Mediciones de grasa, carbohidratos y azúcar en los productos de McDonalds y Burger King*

PRODUCTO	Grasa (g)	Carbohidratos (g)	Azúcar (g)	Calorías Kcal
Big Mac	27	46	9	540
Cheeseburger	12	33	7	300
Quarter Pounder with Cheese	26	42	10	510
McDouble	18	33	7	390
Double Quarter Pounder with Cheese	45	43	10	780
Bacon Ranch Grilled Chicken Salad	13	8	3	300
Hamburger	10	26	6	240
Cheeseburger	16	27	7	320
Bacon y Cheese WHOPPER	51	50	11	790
DOUBLE WHOPPER	58	49	11	900
Double Pretzel Bacon King	60	55	2	920
Chicken Club Salad-Grilled	41	18	6	610

A lápiz y papel: Con la información en la tabla 4.2 contesta:

- ¿Qué componentes nutricionales (grasa, carbohidratos o azúcar) influyen en el aporte calórico? Los que influyen ¿Lo hacen con la misma fuerza?
- ¿Qué componente influye con mayor fuerza? ¿Por qué?

Utilizando tecnología digital: Crea un nuevo documento en *CODAP* y crea los diagramas de dispersión para los siguientes datos bivariados:

- a) (Grasa, Calorías)
- b) (Carbohidratos, Calorías)
- c) (Azúcar, Calorías)

Sigue las indicaciones del profesor para determinar las siguientes correlaciones en *CODAP* y anótalas.

Correlación entre (Grasa, Calorías) = _____

Correlación entre (Carbohidratos, Calorías) = _____

Correlación entre (Azúcar, Calorías) = _____

Figura 3.6. Recuadro definición de coeficiente de correlación

DEFINICIÓN. Coeficiente de Correlación
<p>La correlación r es una medida de la fuerza y la dirección de una relación lineal entre dos variables cuantitativas.</p> <ul style="list-style-type: none"> • La correlación r es un valor entre -1 y 1.

Observa el diagrama de dispersión y la correlación correspondiente para cada conjunto de datos bivariados. Describe la relación entre las variables teniendo en cuenta su gráfico y su correlación.

- ¿Qué observas en los datos representados en el diagrama de dispersión cuando la correlación es un valor que se acerca a 1?
- ¿Qué observas en los datos representados en el diagrama de dispersión cuando la correlación es un valor lejano de 1?

Teniendo en cuenta lo realizado anteriormente, vuelve a contestar las siguientes preguntas:

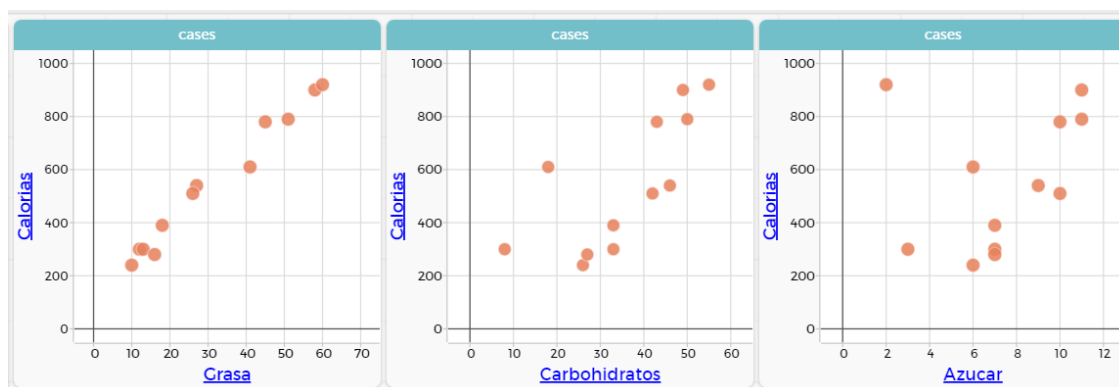
- ¿Existe algún componente nutricional (grasa, carbohidratos o azúcar) presentado en la tabla 2 que influya fuertemente en el aporte calórico (calorías)? ¿por qué?
- ¿Existe algún componente nutricional (grasa, carbohidratos o azúcar) presentado en la tabla 2 que influya débilmente en el aporte calórico? ¿por qué?

Estructura de aplicación de la actividad 2: El tiempo de aplicación del instrumento fue de 2 horas. La primera parte a lápiz y papel tuvo una duración de 20 minutos, la segunda parte utilizando CODAP tuvo una duración de una hora y 40 minutos aproximadamente.

El profesor entregó la situación problema y guió la discusión con los estudiantes acerca de la influencia de los componentes nutricionales en el aporte calórico (sin uso de tecnología digital).

Seguido, utilizando tecnología con el objetivo de optimizar tiempo les compartió el enlace donde obtenían el archivo con los datos de la situación problema ya tabulados en la plataforma CODAP. Además, les indicó como crear varios gráficos de dispersión y observarlos simultáneamente, sin necesidad de borrar alguno (ver figura 3.7). Se puede reducir el tamaño de cada gráfico de dispersión de modo que todos los que se tengan se puedan observar y no estén uno sobrepuesto del otro.

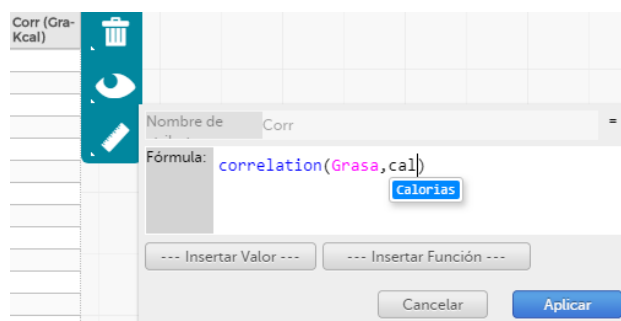
Figura 3.7. Captura de pantalla de los diagramas de dispersión realizados en CODAP.



Finalmente, el profesor explicó como determinar el coeficiente numérico de la correlación en CODAP para cada par de variables de la siguiente manera: Crear un nuevo atributo “Corr(Gra-Kcal)” y seleccionar editar fórmula. Seguido Escribir “correlation” aparecerá en color azul y un paréntesis; ubicarse dentro del paréntesis y dar clic en Insertar Valor,

seleccionar la variable Grasa luego escribir una coma (“,”) y de nuevo insertar valor, y seleccionar la variable Calorías (ver figura 3.8). Dar clic en aplicar. En la tabla aparecerá un valor sombreado de amarillo, el profesor aclaró que la correlación es un solo valor así aparezca en todas las filas.

Figura 3.8. Captura de pantalla para determinar la correlación numérica en CODAP



3.2.2.3. Actividad 3.

Objetivos de aprendizaje:

- Descubrir que la recta de mejor ajuste depende de todos los puntos del diagrama de dispersión.
- Explorar varias formas de utilizar los residuos para definir el mejor ajuste.
- Utilizar el modelo matemático de la recta para realizar predicciones y estimaciones.

Objetivo de investigación: Evidenciar que razonamientos se generan entorno a la nube de puntos y la posible recta de mejor ajuste, cuáles son las estrategias que proponen utilizando el *applet* de GeoGebra donde se muestra una línea conectada a todos los puntos de la nube junto la suma del valor absoluto de todos los residuales. Conocer que posibilidades y limitaciones de las conceptualizaciones que pueden llevar a cabo los estudiantes.

Supuestos de aprendizaje:

Supuestos favorables

- Describen la tendencia de la nube de puntos y su dirección.

- Ver 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).
- Establecen que 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.
- Utilizan la expresión algebraica de la recta de regresión para resolver problemas de estimación y predicción.

Supuestos que obstaculizan

- Prevalece la idea de tomar un par de datos para establecer algún procedimiento algebraico y estimar y predecir.
- Escoger la línea de mejor ajuste como aquella que pase por el mayor número de puntos (Casey, 2015).
- Conceptos previos de la función lineal en matemáticas puede interferir en la capacidad de dar sentido a la tarea de ajustar una línea (Casey, 2014).
- La concepción de “cercanía” de la recta a la nube de puntos (García y Yáñez, 2018)

Para el instrumento de la actividad 3 se tomó un problema del libro de texto “The practice of statistics” de Starnes et al. (2010).

Situación: ¿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.

Tabla 3.3. *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 lápiz y papel: Discute con tu compañero y responde: ¿Cuál consideras es la variable respuesta y la variable explicativa? Describe la relación entre el cambio ANE y el aumento de grasa.

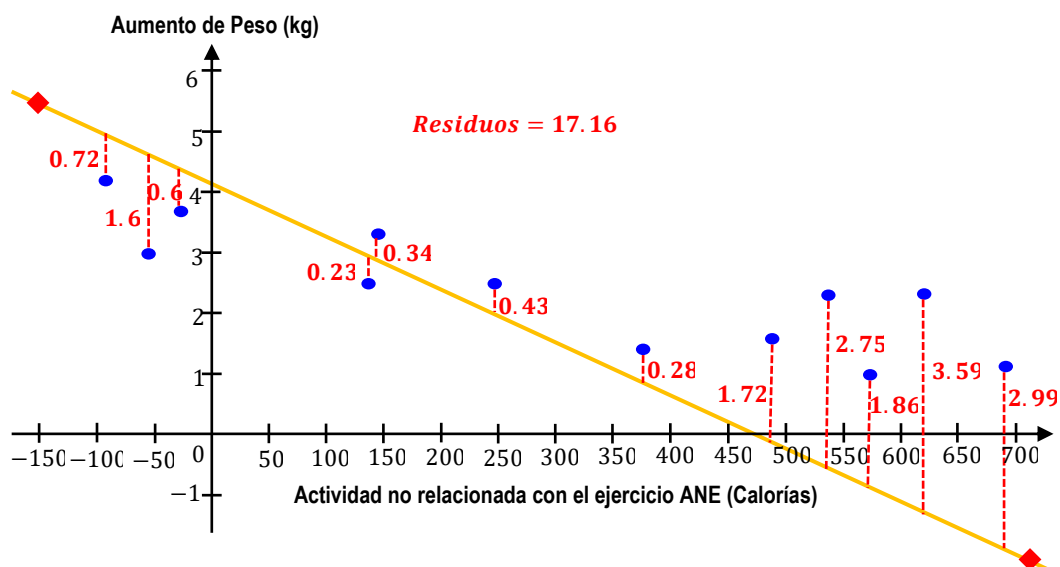
Utilizando tecnología digital: Realiza el diagrama de dispersión en CODAP (<https://codap.concord.org/>) para los datos de la tabla 4.3 y describe el comportamiento de la nube de puntos (intensidad y dirección).

Abre el archivo de GeoGebra llamado ANE_Grasas.ggb (ver figura 3.9, 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 \blacklozenge .
- 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,

$$e = \text{valor } y \text{ (punto de la nube)} - \text{valor } \hat{y} \text{ (punto posible de la línea amarilla)}.$$
- La palabra residuos muestra la suma de todos los residuales.

Figura 3.9. Captura de pantalla del archivo de GeoGebra ANE_Grasas.



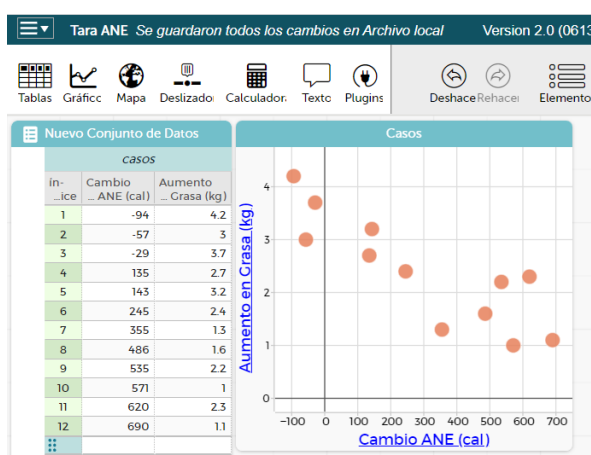
Mueve la línea amarilla, explora lo que ocurre con el valor de cada residual y responde las siguientes preguntas:

1. ¿Dónde consideras que debes ubicar la recta amarilla para que corresponda a la recta de regresión? Explica tu criterio.
2. ¿Cuál es el valor mínimo de los residuos?
3. ¿Qué relación tiene el valor numérico de los residuos con recta?
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?

Estructura de aplicación de la actividad 3: 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.

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). 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 3.10 se muestra el diagrama de dispersión realizado en CODAP (esta figura no se muestra en la hoja de trabajo del estudiante).

Figura 3.10. Captura de pantalla del diagrama de dispersión realizado en CODAP



Finalmente, les indica abrir el archivo de GeoGebra y seguir las instrucciones sin ninguna intervención.

3.3. Metodología para el análisis de resultados: Teoría Fundamentada

El análisis de los datos se basa en la filosofía, principios y procedimientos de la teoría fundamentada (Birks y Mills, 2015; Glaser y Strauss, 2008), cuyo objetivo es elaborar teorías locales que emerjan de los datos y no de deducciones lógicas o de otras teorías.

Entendemos por teoría un esquema explicativo que comprende un conjunto de conceptos relacionados entre sí a través de patrones lógicos de conectividad (Birks y Mills, 2015). Se dice que la teoría es sustantiva cuando se produce con el propósito de comprender un fenómeno tangible en una situación claramente definida (Glaser y Strauss, 2008). Se genera una teoría sustantiva mediante el proceso de elaboración de categorías, percepción y

descripción de sus propiedades, la relación entre categorías y su integración teórica, es decir, de la identificación de una categoría central que explica u organiza a las otras.

En este apartado se describe y se muestra los procesos y métodos que componen la teoría fundamentada (figura 3.11). El objetivo es construir una teoría sustantiva del razonamiento covariacional de los estudiantes de bachillerato.

Figura 3.11. Marco de Procesos y Métodos de la Teoría Fundamentada (Adaptación de Chun Tie et al., 2019)



Sensibilidad teórica. La sensibilidad teórica es la capacidad para conceptualizar y formular esquemas explicativos (teorías) emergentes de los datos (Glaser y Strauss, 2008, p. 46). La investigación que sigue una teoría fundamentada no pretende probar una hipótesis ni hay una teoría completamente desarrollada en el campo de investigación, por lo tanto, asume un alto nivel de sensibilidad teórica del investigador. El estudio y conocimiento de otras investigaciones y teorías es la fuente de la sensibilidad teórica (Birks y Mills, 2015), pero sirven como motivos de reflexión y contraste, no como guías o marcos preestablecidos para interpretar los datos. Para el presente estudio el razonamiento covariacional es definido como la unidad de análisis y está relacionado e influenciado por conceptos de la educación estadística que se estudiaron y se resumieron en la revisión de la literatura, esto permite

aumentar la sensibilidad teórica de la autora, además de la interacción intensiva con los datos durante el proceso analítico.

Muestra deliberada. Es la estrategia utilizada en la teoría fundamentada para llevar a cabo un muestreo con un propósito específico de obtener datos para aclarar, precisar o ampliar aspectos sobre categorías con potencial teórico, es decir, aptas para formar parte de un esquema explicativo.

En teoría fundamentada el muestreo suele ser abierto para recoger conjuntos de datos pertinentes al fenómeno en cuestión pues de inicio no se pretende probar una hipótesis, se desea iniciar la recopilación de datos para el desarrollo de la teoría, por lo tanto, se plantea aplicar una secuencia de actividades y recolectar datos a partir de hojas de trabajo de los estudiantes. Después de haber recolectado los datos, se realizará la transcripción de las respuestas de cada pareja de estudiantes a las tareas. Además, se deben registrar algunas notas durante la aplicación y escribir memorandos que sirvan para el análisis de las respuestas y posibles reflexiones respecto a los instrumentos aplicados.

Muestreo teórico. El investigador sigue las pistas en los datos mediante el muestreo de nuevos datos o material que proporciona información relevante. Birks y Mills (2015) definen el muestreo teórico como el proceso de identificación y seguimiento de pistas que surgen durante el análisis, este proceso se enfoca en generar datos que alimenten el proceso interactivo del análisis comparativo constante de los datos.

Codificación. El método analítico de *comparación constante* consiste en comparar entre sí datos (incidentes, frases, respuestas, etc.) para encontrar rasgos comunes y diferentes entre ellos que permita formar agrupaciones e identificar patrones; la codificación consiste en ponerle nombres de modo que pueda uno referirse a ellos. Una manera de ajustar y mejorar el proceso de codificación es que dos o más investigadores analicen de manera independiente el mismo corpus de datos. Una vez que cada uno tenga una propuesta de codificación, las comparan y discuten, de modo que lleguen a un acuerdo sobre los patrones más importantes presentes en los datos. Finalmente, negocian los nombres apropiados para dichos patrones. Se dice que una codificación así obtenida ha sido validada. Se distinguen tres tipos de

codificación: codificación inicial, codificación intermedia y codificación avanzada que se describen en el siguiente apartado después de la siguiente nota.

Codificación inicial. En la codificación inicial se generan muchos códigos, donde se utilizan palabras o grupos de palabras que se identifican 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 generan se continua el análisis comparativo constante entre las respuestas de los estudiantes, se anexan nuevos incidentes, algunos de los cuales se etiquetan con los códigos que ya se tienen establecidos, y otros incidentes generaran nuevos códigos.

Codificación intermedia. Se basa en la codificación inicial, donde 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). Durante el proceso de codificación intermedia van generando códigos principales y quizás algunos subcódigos.

Codificación avanzada. Para Birks y Mills (2015) en la codificación avanzada, los conceptos que llegan a formar las categorías serán abstractos, representando historias de muchos. Es importante resaltar que las categorías no se desarrollan para un determinado estudiante, se desarrollan con respecto a varias respuestas de los estudiantes. En la etapa final de la codificación los hallazgos se presentan como un conjunto de conceptos interrelacionados mediante enunciados explicativos que detallan las relaciones entre las categorías y la categoría central.

CAPÍTULO IV. RESULTADOS

En este capítulo se presentan los resultados de la investigación que comprenden:

- ✓ El análisis de las respuestas de los estudiantes a las tres actividades. La primera actividad (ver apéndice B) y segunda actividad (ver apéndice C) se enfoca en las ideas centrales de datos bivariados, diagramas de dispersión y coeficiente de correlación utilizando la plataforma en línea *CODAP*. La tercera actividad se centra en la idea de recta de regresión utilizando un applet diseñado en GeoGebra (ver apéndice D).
- ✓ Resultados de la aplicación de un cuestionario diagnóstico, y un cuestionario de evaluación. En el apéndice E se describe el contenido matemático y el objetivo de evaluación de cada ítem que hace parte del cuestionario diagnóstico. El cuestionario de final (ver apéndice F) fue evaluativo y el estudiante tenía la opción de utilizar el software o no para resolver los problemas.

4.1. Análisis de los datos obtenidos de las actividades

En este apartado se emplean los procesos y métodos de teoría fundamentada para el análisis de las hojas de trabajo donde los estudiantes escribieron sus respuestas a las actividades.

4.1.1. Codificación inicial.

La codificación inicial es el proceso mediante el cual por primera vez se identifican, y se les da un nombre (código), a extractos de las respuestas de los estudiantes que describen ideas o acciones que el investigador considera relevantes desde algún punto de vista, principalmente desde el punto de vista de la pregunta de investigación, aunque no necesariamente. Es decir, se busca extraer producciones que ayuden a responder la pregunta de investigación, pero sin limitarse a esto solamente, sino que también se quiere estar abierto a manifestaciones que pueden ser interesantes desde otra perspectiva. Esta actividad permite reducir significativamente los datos, pues ideas y acciones de diferentes estudiantes pueden ser descritas en lo esencial por un mismo código. A su vez, esta reducción permite comunicar lo que dicen los datos de manera eficiente. Por supuesto, “lo que dicen los datos” es una abstracción filtrada por las decisiones del analista.

A continuación, se describe el proceso realizado para la codificación de las respuestas dadas a tres actividades.

1. Primero se generaron etiquetas, donde se utilizaron palabras importantes o grupos de palabras que se identificaron. Por ejemplo, la figura 4.1 ilustra tres respuestas de los estudiantes a la pregunta indicada, que se les asignó la etiqueta “distancia (nube, recta)”, esta etiqueta también recibe el nombre de código, a su vez se establece el subcódigo “menor distancia”. Los códigos iniciales para las respuestas recopiladas de las actividades se encuentran en el apéndice 1.

Figura 4.1. Ejemplo de generación de la etiqueta distancia.

Pareja	Pregunta ¿Qué relación tiene el valor numérico de los residuos con Recta de ajuste?
1	Entre menos distancia entre todos los puntos a la línea menor va a ser el residuo.
2	Cuando están cerca todos los puntos. Entre menos distancia entre la línea amarilla y los puntos azules, el residuo disminuye.
3	Entre menos distancia hay entre la línea y los puntos el residuo es menor

Distancia (nube, recta)
Menor Distancia

2. Seguido de asignar estas primeras etiquetas, se establecieron palabras o frases claves para cada código. Por ejemplo, la tabla 4.1 muestra estas frases junto con ejemplos de respuestas de los estudiantes para los códigos de la situación utilizando tecnología digital (CODAP), que corresponde a describir la relación entre las variables presente en la actividad 3. Esto se realizó para las codificaciones iniciales de las tres actividades y se encuentran en el apéndice H.

Tabla 4.1. *Ejemplo de palabras clave o frases para los códigos.*

Código	Palabras claves o frases para codificar	Ejemplo de respuesta
RELACIÓN INVERSA	los valores de una variable aumentan y los valores de la otra variable disminuyen	<p>“al aumentar las calorías hay más quema de grasa”</p> <p>“Entre más cambio en ANE, menos aumentos de grasa presenta”</p> <p>“Entre menos actividad realices más grasa...”</p> <p>“entre menos actividades realices más grasa acumulas”</p> <p>“los movimientos que no sean ejercicios hacen que pierdas grasa y es lógico, al hacer ejercicio haces movimientos”</p>
	Se debe a que la grasa es difícil de eliminar Si no haces ejercicios no quemas grasa	<p>“el aumento de grasa depende de la cantidad de ANE que se realiza diariamente”</p> <p>“la grasa no aumenta porque, aunque no se hace ejercicio si se mantiene en movimiento.”</p>

Nota. La frase “aumentar las calorías hay más quema de grasa” podría entenderse como relación directa, sin embargo, la interpretación es inversa ya que si hay más quema de grasa significa disminución de la grasa.

- Después de codificar cada respuesta, se realizaron tablas de frecuencia para cada código. Por ejemplo, la tabla 4.2 resume la frecuencia obtenida para cada código respecto a la pregunta que involucraba la descripción de la relación entre las variables presente en la actividad 3. Las tablas de frecuencia para todos los códigos de las actividades se encuentran en el apéndice I.

Tabla 4.2. *Ejemplo de tabla de frecuencia para cada código.*

Tarea	Código	Frecuencia (Parejas)
Utilizando CODAP realizar el diagrama de dispersión, determinar la correlación y describir la relación	Relación inversa	8
	Contexto	3

A continuación, en la codificación intermedia se explica la descripción y agrupación de los códigos que emergieron en la codificación inicial.

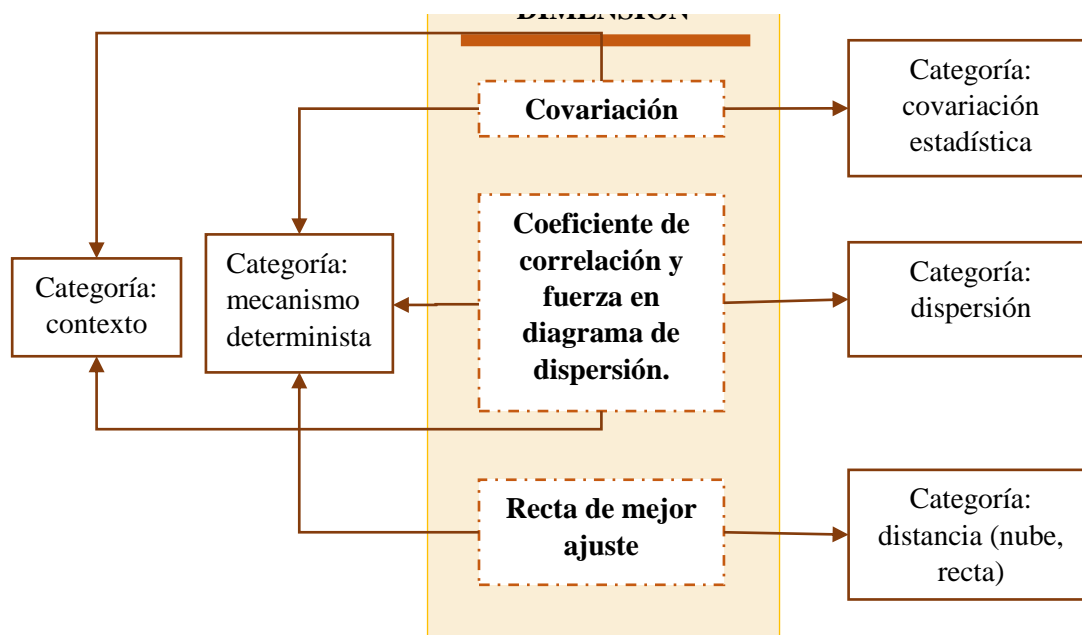
4.1.2. Codificación intermedia.

La codificación intermedia es una segunda etapa del análisis de datos en la que se identifican o construyen categorías. Una categoría es un nombre que agrupa varios códigos y, por tanto, abarca aspectos comunes de más ideas y acciones. Una categoría es más abstracta que un código. Una dimensión se forma con dos o más categorías cuando se relacionan con otras dimensiones dentro de una esquema o marco que las organice. Las tres dimensiones que proponemos se asocian a los conceptos estadísticos que tiene el mismo nombre, pero como dimensión no aludíamos al contenido estadístico, sino a las categorías que abarcan códigos relacionados con el contenido; es decir, aluden al razonamiento de los estudiantes sobre el contenido, y son:

- i) **Covariación:** comprende los razonamientos de los estudiantes cuando se les pide analizar la relación entre las variables a partir de la situación problema y utilizando tecnología.
- ii) **Coefficiente de correlación junto con la fuerza en diagramas de dispersión:** está presente en los razonamientos cuando se juzga la fuerza de datos representados en un diagrama de dispersión y se establece el coeficiente numérico de la correlación utilizando tecnología.
- iii) **Recta de mejor ajuste:** Los códigos que pertenecen a esta dimensión se relacionan con los razonamientos de los estudiantes al establecer un criterio de lo que es la recta que mejor se ajusta a los datos representados en un diagrama de dispersión.

La figura 4.2 resume las dimensiones (covariación, coeficiente de correlación y fuerza en diagramas de dispersión, recta de mejor ajuste) y sus categorías (mecanismo determinista, contexto, covariación estadística, dispersión, distancia entre la nube de puntos y la recta)

Figura 4.2. Dimensiones y categorías que emergieron de la codificación inicial e intermedia



A continuación, para cada dimensión se describen las categorías y los códigos agrupados en cada una de ellas.

4.1.2.1. Dimensión covariación.

A esta dimensión pertenecen los códigos que se relacionan con los razonamientos de los estudiantes cuando se les pide analizar la relación entre las variables. Las preguntas de covariación están presentes en la actividad 1 (12 parejas) y en la actividad 3 (11 parejas). A continuación, la tabla 4.3 muestra los códigos con su respectiva frecuencia.

Tabla 4.3. Códigos y frecuencia para los razonamientos al juzgar la relación entre las variables

Actividad	Tarea	Código	Frecuencia (Parejas)
Actividad 1	Juzgar la relación entre las variables a partir del contexto a lápiz y papel.	Tendencia global positiva	8
		Operación aritmética: vista individual	4

Actividad 3	Realizar el diagrama de dispersión a lápiz y papel	Sesgo de atención: idea funcional	5
		Nube de puntos	7
	Describir la nube de puntos que muestra el diagrama de dispersión en CODAP.	Linealidad positiva con error	5
		Orden ascendente con error	5
		Respuesta fuera del contexto	2
	Juzgar la relación entre las variables a partir del diagrama de dispersión de CODAP.	Dependencia positiva	7
		Dependencia funcional	2
		Dependencia con error	3
	A lápiz y papel describir la relación entre las variables	Covariación funcional	3
		Componente aleatorio	5
		Contexto	3
	Utilizando CODAP realizar el diagrama de dispersión, determinar la correlación y describir la relación	Relación inversa	8
		Contexto	3

Los códigos de la dimensión covariación estadística se agruparon en tres categorías llamadas *mecanismo determinista*. La segunda categoría que emergió es denominada *covariación estadística* y la tercera *covariación desde el contexto*. En seguida se describen estas categorías.

Mecanismo determinista, a esta categoría pertenecen los códigos que evidencian razonamientos donde el tratamiento de la relación entre las dos variables es de manera proporcional o funcional, es decir, donde el concepto de error está ausente y se trata a la covariación desde un punto de vista del cálculo. Los códigos que se agrupan en esta categoría son:

- *Operación aritmética* (4 de 12 parejas): Realizar operaciones aritméticas con algunos datos.
- *Sesgo de atención-idea funcional* (5 de 12 parejas): Se ubican los pares ordenados del conjunto de datos bivariados en el plano cartesiano y se traza la unión por semi-rectas entre ellos.

- *Dependencia funcional* (2 de 12 parejas): La relación entre las variables es proporcional, una variable aumenta la otra también aumenta.
- *Covariación funcional* (3 de 11 parejas): cuando el cambio en las variables es proporcional, es decir, donde los valores de la variable y quedan determinados de un modo preciso por los valores de la otra variable que se considera como independiente.

Covariación estadística: Cuando los datos no se ajustan de modo preciso a una función matemática, pero su comportamiento se representa por un modelo matemático debido a la presencia de la variabilidad del conjunto de datos bivariados.

- *Tendencia global positiva* (8 de 12 parejas): Hay una relación positiva, si los valores de una variable tienden a aumentar conforme los valores de la otra variable también aumentan.
- *Graficar nubes de puntos* (7 de 12 parejas): Se ubican los pares ordenados del conjunto de datos bivariados en el plano cartesiano.
- *Linealidad positiva con error* (5 de 12 parejas): La tendencia de la nube de puntos es positiva y su forma es casi lineal.
- *Orden ascendente con error* (5 de 12 parejas): Describen el orden ascendente de los puntos en el diagrama sin aceptar su comportamiento lineal.
- *Dependencia positiva* (7 de 12 parejas): La relación entre las variables es dependiente porque ambas aumentan.
- *Dependencia con error* (3 de 12 parejas): Es una relación donde una variable depende de la otra y ambas aumentan de manera no proporcional.
- *Componente aleatorio* (5 de 11 parejas): Se enfoca la 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 conforme leen de izquierda a derecha, notan que no hay patrones previsibles.
- *Relación inversa* (8 de 11 parejas): Describe de manera global el comportamiento de la nube, mientras una variable aumenta la otra disminuye.

Covariación explicada desde el contexto, consiste en juzgar el comportamiento u organización de los datos bivariados a partir del conocimiento previo o creencias acerca de

la relación entre las variables gracias al contenido contextual del problema. Presente en la actividad 3, de 11 parejas se evidenciaron 3 respuestas en esta categoría.

4.1.2.2. Dimensión: Coeficiente de correlación junto con diagramas de dispersión

Los códigos de esta categoría son aquellos que involucran razonamientos de los estudiantes cuando examinan el coeficiente de correlación y la información vertida en un diagrama de dispersión. Implica considerar la fuerza de la relación entre dos variables y cómo están los datos de apretados en torno a una recta lineal, dentro del contexto de un problema. Las preguntas correspondientes a esta dimensión están presentes solo en la actividad 2 (12 parejas).

Tabla 4.4. Códigos y frecuencia para los razonamientos al analizar correlación junto con diagramas de dispersión

Actividad	Pregunta	Código	Frecuencia (Parejas)	
Actividad 2	¿Qué componentes nutricionales (grasa, carbohidratos o azúcar) influyen en el aporte calórico? Los que influyen ¿Lo hacen con la misma fuerza?	Creencias	7	
		Desde el contexto	5	
	¿Qué componente influye con mayor fuerza? ¿Por qué?	Creencias	4	
		Comparar cantidades	7	
	¿Existe alguna relación entre el valor numérico de correlación y el comportamiento de los datos representados en el diagrama de dispersión?	<i>Diagrama de dispersión: Grasas-Calorías</i>		
		Mayor Correlación implica comportamiento lineal	4	
		Conglomeración	8	
		<i>Diagrama de dispersión: Carbohidratos-Calorías</i>		
		Menor Correlación un poco de dispersión	9	

	Grupos de puntos y otros alejados	3
	Diagrama de dispersión: Azúcar-Calorías	
	Menor correlación más dispersión	5
	Puntos más dispersos	7
¿Qué observas en los datos representados en el diagrama de dispersión cuando la correlación es un valor que se acerca a 1?	Puntos cerca y linealidad	10
	Conglomeración	2
¿Qué observas en los datos representados en el diagrama de dispersión cuando la correlación es un valor lejano de 1?	Dispersos y alejados	8
	Mayor dispersión pierden forma lineal	4
¿Existe algún componente nutricional presentado que influya fuertemente en el aporte calórico (calorías)? ¿por qué?	El azúcar: La nube de puntos es más dispersa	2
	Las grasas porque su correlación es cercana a uno	9
	Los carbohidratos: creencias	1
¿Existe algún componente nutricional presentado en la tabla 2 que influya débilmente en el aporte calórico? ¿por qué?	Grasas: Los puntos de la nube están juntos	2
	Azúcar	10
¿Cómo influye los valores extremos en la correlación (azúcar, calorías)?	Cerca o lejos cambia la correlación	9
	Cerca al conglomerado mayor correlación	3

Los códigos de la dimensión coeficiente de correlación junto con diagramas de dispersión se agruparon en tres categorías llamadas *mecanismo determinista*. La segunda categoría que emergió es denominada *dispersión* y la tercera categoría *creencias desde el contexto*.

Mecanismo determinista, a esta categoría pertenecen los códigos que evidencian razonamientos donde el error es ausente y se trata el coeficiente de correlación y la información vertida en un diagrama de dispersión desde una vista del cálculo.

- *Comparar cantidades* (8 de 12 parejas): se establece la relación más fuerte comparando valores numéricos de la tabla de datos que se presenta la situación problema, se enfoca la atención en un subgrupo de los datos aquellos que tienen valores grandes y un comportamiento creciente.

Dispersión. A esta categoría pertenecen los códigos que evidencian razonamientos donde la variabilidad está presente, en algunos casos se enfocan solo en la conglomeración de los puntos en el diagrama de dispersión, y otros razonamientos relacionan el valor numérico del coeficiente de correlación con la dispersión de los datos en la nube de puntos. Muchos códigos se repiten, se convierten en subcódigos, aquí se describen los códigos principales.

- *Mayor correlación implica comportamiento lineal* (8 de 12 parejas): Se establece que a mayor valor numérico del coeficiente de correlación el comportamiento de la nube de puntos se conglojera de manera lineal.
- *Menor correlación un poco de dispersión* (9 de 12 parejas): El valor de la correlación es menor y hace que los puntos de la nube estén un poco dispersos.
- *Menor correlación mayor dispersión* (5 de 12 parejas): Para las variables azúcar y calorías el coeficiente de correlación es menor, esto implica que los puntos en el diagrama están más dispersos.
- *Conglomeración* (8 de 12 parejas): Se omite el valor numérico de la correlación haciendo referencia solo a la conglomeración de los puntos en el diagrama de dispersión.
- *Mayor dispersión pierde forma lineal* (4 de 12 parejas): Los puntos en la nube poseen más dispersión y por lo tanto pierden su forma lineal.

Creencias desde el contexto, consiste en explicar la fuerza entre las variables con base en las ideas o creencias de los elementos implicados en el contexto de la situación problema.

4.1.2.3. Dimensión: recta de mejor ajuste

Los códigos que pertenecen a esta dimensión se relacionan con los razonamientos de los estudiantes acerca de la recta de mejor ajuste. Las preguntas de recta de mejor ajuste están

presentes en la actividad 1 (12 parejas) y en la actividad 3 (11 parejas). A continuación, la tabla 4.5 muestra los códigos con su respectiva frecuencia.

Tabla 4.5. Códigos y frecuencia para los razonamientos al establecer la recta de mejor ajuste

Actividad	Tarea/Pregunta	Código	Frecuencia (Parejas)
Actividad 1	Estimar a partir de la situación problema a lápiz y papel.	Aritmético siguiendo un patrón	6
		Aritmético sin patrón definido	6
	Predecir a partir de la situación problema a lápiz y papel.	Aritmético siguiendo un patrón	4
		Aritmético sin patrón definido	8
	Como utilizar la línea arrojada por CODAP para estimar y predecir.	Recta representa toda la información	10
		Proceso aritmético	2
	Cuál es el criterio que se cree utiliza CODAP para trazar la línea de mejor ajuste	Concepción funcional	1
		Cercanía	1
		Concepción local	3
		Operacional o Algebraico	5
		No describe ninguno	2
	Actividad 3	¿Dónde consideras que debes ubicar la Recta amarilla para que corresponda a la Recta de regresión? Explica tu criterio.	Cercanía
Distancia			4
Pertenencia			1
Ninguno			2
¿Qué ocurre con el valor de cada residual si mueves la recta cerca o lejos de la nube de puntos?		Distancia	4
		Valor numérico residuos= distancia (nube y la recta)	5
¿La manera como ubicaste la Recta amarilla es igual a la Recta de regresión arrojada por		Recta centrada implica menor residuos	2
		Comparación de posición	9
		Comparación de residuos	2

GeoGebra? ¿Sí? ¿No? ¿En qué se
diferencian?

Los códigos de la dimensión recta de mejor ajuste se agrupan en tres categorías llamadas *mecanismo determinista*. La segunda categoría que emergió es denominada *distancia (nube, recta)* y la tercera categoría *criterio contextual*.

Mecanismo determinista. Consiste en buscar y aislar los datos bivariados que se apeguen a un modelo (ecuación). Por ejemplo, algunos estudiantes dividen el conjunto de datos en dos partes, los que corresponden a un modelo y los que no lo hacen; ellos privilegian los puntos que expresan una dependencia funcional.

- *Aritmético siguiendo un patrón:* Buscar un patrón en la tabla de valores entre un par de datos para establecer un procedimiento aritmético.
- *Aritmético sin patrón definido:* Sin patrón definido se efectúan algunas operaciones básicas (suma, resta, multiplicación y división).
- *Proceso aritmético:* Utilizar un par de valores para realizar algún procedimiento matemático.
- *Concepción local:* La línea es aquella que contiene un par de puntos
- *Operacional o algebraico:* La línea se determina mediante una expresión algebraica.

Distancia nube recta. Se define la recta que mejor se ajusta a los datos como aquella donde la distancia de la recta a la nube de puntos sea pequeña y hace que el valor numérico de los residuos sea el menor. Esta categoría contiene las siguientes subcategorías:

- *Recta representa toda la información:* Emplear la línea arrojada por el software para estimar y predecir el valor lo cual implica asumir la línea como representante de todos los datos.
- *Cercanía:* Se evidencia cuando se describe la línea de mejor ajuste como aquella que se acerca lo más posible a la nube de datos.
- *Distancia:* muestra que el valor numérico de los residuos cuando es menor hace que la distancia entre la posible línea de mejor ajuste y la nube de puntos sea pequeña.

- *Pertenencia*: es el criterio de considerar la línea de mejor ajuste como aquella que debe pasar por algunos puntos de la nube. No es claro cuales puntos de los datos se tomaron como referencia para considerar la línea de mejor ajuste.
- *Valor numérico de los residuos*: es igual a la distancia entre la nube y la recta, es decir, se asume el valor total de los residuales como aquel valor que representa la cercanía o lejanía de la línea de ajuste.
- *Recta centrada*: implica menor residuos evidencia que se enfatiza en la línea de mejor ajuste como aquella que debe representar la tendencia de acumulación del conjunto de datos más o menos dispersos centrándose en medio de la nube de manera que el residuo es menor en esa posición.
- *Comparación de posición*: se evidencia con argumentos con relación a la cercanía o lejanía entre ambas rectas o entre las rectas y la nube de puntos.
- *Comparación de residuos*: para diferenciar las rectas se toma como elemento principal el valor de los residuos.

La codificación avanzada presenta los conceptos que forman las categorías de manera abstracta, representando historias de muchos. Es importante resaltar que las categorías no se desarrollan para un determinado estudiante, se desarrollan con respecto a varias respuestas de los estudiantes. En la etapa final de la codificación, los hallazgos se presentan como un conjunto de conceptos interrelacionados que describimos en el capítulo discusión de los resultados donde las categorías principales y códigos mostrados en este capítulo conforman la teoría que emergió.

4.2. Análisis de los datos obtenidos del cuestionario diagnóstico y el cuestionario evaluativo.

Para el análisis del cuestionario diagnóstico y evaluativo se utilizaron los procesos y métodos de teoría fundamentada, implementando las tres codificaciones: inicial, intermedia y avanzada.

4.2.1. Cuestionario diagnóstico

El cuestionario diagnóstico comprende dos partes: la primera parte consta de preguntas relacionadas con álgebra básica, y la segunda parte, problemas de correlación y regresión. El objetivo del cuestionario diagnóstico es explorar las concepciones iniciales de los estudiantes frente a las temáticas de correlación y regresión. En el apéndice 4 se encuentra la codificación inicial realizada a las respuestas obtenidas de la aplicación de este cuestionario, el cual fue contestado por 21 estudiantes. A continuación, la tabla 4.6 muestra la frecuencia para cada código del componente de álgebra básica.

Tabla 4.6. Frecuencia para cada código del componente álgebra básica

Contenido	Código	Frecuencia (Parejas)
Establecer algebraicamente la función lineal que representa el enunciado. Es decir, pasar del enunciado verbal a establecer una expresión algebraica. Dado un valor sustituir en la expresión algebraica y hallar su valor.	Lenguaje algebraico	14
	Reemplazar valores	4
	Interpretación errónea	3
Establecer la función lineal que representa la relación a partir de datos tabulares.	Lenguaje algebraico.	10
	Observación individual de variables	11
Determinar la función lineal que representa la relación de proporcionalidad y encontrar los valores	Lenguaje algebraico.	9
	Patrón Aritmético	10
	No responde	2
Identificar en una pareja ordenada la abscisa y la ordenada ubicando el punto en el plano cartesiano.	Intercambio entre abscisa y ordenada	4
	Ubicación correcta	17

La anterior tabla muestra altas frecuencias para los códigos que dan cuenta de razonamientos correctos respecto al contenido de expresiones algebraicas, estos pueden ser:

- *Lenguaje algebraico*: Se establece la forma algebraica de una expresión verbal identificando correctamente las variables. De igual manera, mediante el lenguaje algebraico se representa la relación funcional de los datos bivariados.
- *Reemplazar valores*: Sustitución correcta de un valor en una expresión algebraica.
- *Interpretación errónea*: Se establece una expresión algebraica que no corresponde al enunciado y realiza operaciones aritméticas incorrectas.
- *Observación individual de variables*: Se establece la relación estableciendo algún patrón numérico para cada variable.
- *Patrón aritmético*: Establecen un patrón aritmético utilizando sumas, restas o multiplicaciones sin uso de una expresión algebraica de esa manera establecen la proporcionalidad.
- *Intercambio entre abscisa y ordenada*: Ubicación incorrecta asume la ordenada como x y la abscisa como y .
- *Ubicación correcta*: Sustitución correcta de un valor en una expresión algebraica.

La tabla 4.7 muestra la frecuencia para cada código del componente evaluativo de correlación y regresión.

Tabla 4.7. Frecuencia para cada código del componente evaluativo de correlación y regresión

Contenido	Código	Frecuencia (Parejas)
Establecer el diagrama de dispersión que represente una correlación de 0,6.	Comportamiento lineal perfecto positivo o negativo	9
	Poca dispersión conglomeración débil	3
	Comportamiento lineal, pero conglomeración fuerte	6
	Ninguna	3
	Linealidad perfecta positiva.	4
	Linealidad perfecta negativa.	1

Reconocer gráficamente una relación fuerte entre las variables	Linealidad perfecta positiva y negativa.	8
	Linealidad positiva.	4
	Comportamiento lineal.	4
Reconocer la relación establecida en forma textual e identifique la figura correspondiente.	Comportamiento bivariado negativo	21
Establecer un criterio para representar la recta de mejor ajuste en un diagrama de dispersión.	Dependencia funcional	9
	Concepción local	9
	Medio de todos los puntos (cercanía)	2
Establecer una estrategia para predecir a partir de la situación problema.	Concepción funcional: regla de tres	1
	Patrón aritmético	4
	Promedio	2
	Comportamiento ascendente aritmético	14

Los anteriores códigos muestran los razonamientos respecto al contenido de fuerza y dirección en diagramas de dispersión, estos pueden ser:

- *Comportamiento lineal perfecto positivo o negativo:* correlación de 0.6 implica un comportamiento lineal perfecto.
- *Poca dispersión conglomeración débil:* correlación 0.6 no muestra una relación lineal perfecta pero el valor por encima de 0.5 muestra una nube de puntos con fuerza débil.
- *Comportamiento lineal, pero conglomeración fuerte:* correlación de 0.6 muestra un comportamiento lineal, pero con una nube de puntos con fuerza fuerte.
- *Comportamiento bivariado negativo:* Identificación correcta del diagrama de dispersión que representa la relación textual.

Los razonamientos respecto al contenido de diagramas de dispersión y correlación evidenciaron los siguientes códigos:

- *Linealidad perfecta positiva*: Fuerza implica una nube de puntos cuyo comportamiento es perfectamente lineal en sentido positivo.
- *Linealidad perfecta negativa*: Fuerza implica una nube de puntos cuyo comportamiento es perfectamente lineal en sentido negativo.
- *Linealidad perfecta positiva y negativa*: Fuerza implica ambos comportamientos perfectamente lineales de sentido positivo y negativo.
- *Linealidad positiva*: Fuerza implica comportamiento lineal en sentido solo positivo de la nube de puntos.
- *Comportamiento lineal*: Fuerza implica un comportamiento lineal en cualquier sentido.

La mayoría de los códigos evidencian razonamientos desde el punto de vista funcional al establecer el comportamiento lineal y fuerza de los datos, esto era de esperarse dado que los estudiantes han estudiado el tema de funciones lineales y polinomiales previamente.

Los códigos que muestran los razonamientos respecto al contenido recta de mejor ajuste son:

- *Dependencia funcional*: unir los puntos por semi-rectas.
- *Concepción local*: trazar una recta por un par de puntos o un subconjunto de puntos.
- *Cercanía*: Trazar una recta cerca y en medio de todos los puntos.

Las frecuencias para los códigos concepción local (9) y dependencia funcional (9) fueron las más altas, están concepciones ya han sido reportadas en investigaciones como las de Casey (2015, 2014).

Los códigos que muestran los razonamientos respecto al contenido predecir, estos pueden ser:

- *Concepción funcional*: Asumir proporcionalidad y realizar una regla de tres para predecir.
- *Patrón aritmético*: Establecer un patrón aritmético utilizando todos los datos para realizar sumas o diferencias.

- *Comportamiento ascendente aritmético*: Al azar establecer un valor con base en el comportamiento de los datos.

Los códigos evidencian razonamientos de dependencia funcional matemática al establecer una predicción, donde predominan las operaciones aritméticas o procedimientos al azar.

4.2.2. Cuestionario Evaluativo

El cuestionario evaluativo comprende solo comprende problemas de correlación y regresión. El objetivo del cuestionario evaluativo es obtener información sobre el avance de los estudiantes respecto a las ideas de correlación y regresión. En el apéndice 6 se encuentra la codificación inicial realizada a las respuestas obtenidas de la aplicación de este cuestionario evaluativo, el cual fue contestado por 7 estudiantes, a la última sesión de clase asistieron pocos estudiantes y posteriormente fue imposible obtener más datos debido al cierre de las instituciones por pandemia ocasionada por el COVID. A continuación, la tabla 4.8 muestra la frecuencia para cada código del componente de álgebra básica.

Tabla 4.8. Frecuencia para cada código de la componente estadística del cuestionario evaluativo.

Contenido	Código	Frecuencia (Parejas)
Establecer el diagrama de dispersión que represente una correlación de 0,6.	Poca dispersión conglomeración débil	3
	Comportamiento lineal, pero conglomeración fuerte	4
Reconocer gráficamente una relación fuerte entre las variables	Linealidad perfecta positiva	1
	Linealidad perfecta positiva y negativa	5
	Comportamiento lineal	1
Establecer la relación a partir de diagramas de dispersión y lenguaje textual e identificar	Identificación correcta del lenguaje verbal de datos bivariados al lenguaje gráfico.	5

la variable dependiente e independiente.	Confusión comportamiento ascendente con el descendente.	2
Establecer un criterio para representar la recta de mejor ajuste en un diagrama de dispersión.	Cercanía	3
	Noción distancia (nube, puntos)	2
	El comportamiento de los datos	1
	Pequeña correlación corte con el eje Y	1
Establecer una estrategia para predecir a partir de la situación problema.	Concepción funcional: regla de tres	2
	Utilizar la línea para predecir	3
	Comportamiento ascendente	2

Los códigos que muestran razonamientos respecto al contenido de fuerza y dirección en diagramas de dispersión son:

- *Poca dispersión conglomeración débil:* correlación 0.6 no muestra una relación lineal perfecta pero el valor por encima de 0.5 muestra una nube de puntos con fuerza débil.
- *Comportamiento lineal, pero conglomeración fuerte:* correlación de 0.6 muestra un comportamiento lineal, pero con una nube de puntos con fuerza fuerte.

Al finalizar los estudiantes muestran nuevos términos en su vocabulario al describir la fuerza y dirección de los datos en un diagrama, utilizan “conglomeración” para referirse a la fuerza de los datos representados. Sin embargo, sobreestiman el valor de 0,6 asociándolo a una correlación fuerte.

Los códigos que muestran razonamientos respecto al contenido de diagramas de dispersión y correlación son:

- *Linealidad perfecta positiva:* Fuerza implica una nube de puntos cuyo comportamiento es perfectamente lineal en sentido positivo.
- *Linealidad perfecta positiva y negativa:* Fuerza implica ambos comportamientos perfectamente lineales de sentido positivo y negativo.

- *Comportamiento lineal*: Fuerza implica un comportamiento lineal en cualquier sentido.

Sus razonamientos dan cuenta de una asociación entre la fuerza y el comportamiento lineal positivo y negativo de la nube de puntos en los diagramas de dispersión, los códigos evidencian que la linealidad siempre estuvo asociada a la fuerza de los datos.

Los códigos que muestran razonamientos al establecer un criterio para la recta de mejor ajuste son:

- *Identificación correcta del lenguaje verbal de datos bivariado al lenguaje gráfico*: se establece de manera adecuada la relación para cada afirmación.
- *Confusión comportamiento ascendente con el descendente*: Se confunde en el lenguaje textual el comportamiento en sentido positivo con el comportamiento en sentido negativo.

Los códigos que muestran razonamientos respecto al contenido de predecir son:

- *Cercanía*: Trazar una recta cerca y en medio de todos los puntos.
- *Noción distancia (nube, puntos)*: Se tiene en cuenta las distancias de los puntos a la recta para trazarla.
- *El comportamiento de los datos*: trazar una recta que represente el comportamiento de los puntos.
- *Pequeña correlación y corte con el eje Y*: Establece un punto de corte con el eje Y, eligiendo también tres puntos colineales para trazar la recta.

En este cuestionario surgen más códigos asociados con el comportamiento global de los datos, evidenciando “una vista agregada de los datos” al establecer el criterio de la recta de mejor ajuste teniendo en cuenta todo el conjunto de datos, es decir, la estrategia de trazar la recta por algunos puntos o unir por semi-rectas los puntos en el diagrama es superada por la evidencia del código distancia (nube, recta), la noción de cercanía, al igual que trazar una recta que representa el comportamiento de los datos.

Los códigos que muestran razonamientos respecto al contenido predecir son:

- *Concepción funcional:* Asumir proporcionalidad y realizar una regla de tres para predecir.
- *Utilizar la línea para predecir:* Ubicar la intersección de la línea vertical que pasa por el valor de x y la recta que traza de mejor ajuste.
- *Comportamiento ascendente aritmético:* Al azar establecer un valor con base en el comportamiento de los datos.

Para predecir prevale los razonamientos que involucran operaciones aritméticas (regla de tres), sin embargo, también se evidencia el utilizar la intersección entre la recta de ajuste y la línea vertical que pasa por el valor de x para establecer el valor de y correspondiente al valor predicho.

CAPÍTULO V. DISCUSIÓN DE LOS RESULTADOS

Este capítulo comprende los siguientes apartados:

1. En el primer apartado, se presenta la discusión de los resultados descritos en el anterior capítulo en relación con la literatura revisada.
2. En el segundo apartado, se aborda una introducción a los modelos de desarrollo del razonamiento estadístico, para seguido;
3. en el tercer apartado, describir la codificación avanzada que presenta los modelos de desarrollo del razonamiento covariacional.

5.1. Discusión de los resultados en relación con la literatura

La discusión de los resultados se realiza de la siguiente manera: primero, se discute las categorías (mecanismo determinista y contexto) presentes en las tres dimensiones (covariación, el coeficiente de correlación junto con los diagramas de dispersión y la recta de mejor ajuste), y segundo, se aborda las categorías que emergieron individualmente en cada dimensión (covariación estadística, dispersión y distancia entre la nube y la recta).

Al *mecanismo determinista* pertenecen los códigos que involucran el tratamiento de datos bivariados desde un punto de vista funcional del cálculo, es decir, hay ausencia de códigos que involucran razonamientos que evidencien el error. Al *mecanismo contextual*, pertenecen los códigos que muestran razonamientos donde los estudiantes no ven los datos como simplemente números, sino números con un contexto.

En la dimensión *covariación*, el mecanismo determinista consiste en juzgar la relación entre las dos variables de manera proporcional o funcional. Estepa y Batanero (1996) denominaron concepción determinista cuando los estudiantes esperan una correspondencia donde a cada valor de la variable dependiente se le asigne otro valor de la variable independiente. Este mecanismo determinista se presentó al juzgar una relación en sentido negativo; sus respuestas (3 de 11 parejas) se refieren solo 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, afirman, por ejemplo: “entre más grasa consumas más calorías aumentas”, “entre menos cambio en las calorías mediante los movimientos, menos la grasa disminuirá”. También se presenta una dependencia del *contexto* del problema, *donde es difícil para los estudiantes* separar sus creencias a priori temporalmente para observar los

datos y evaluar la relación (Watson y Moritz, 1997, 2007), por ejemplo, afirman, “la grasa aumenta porque se queda encapsulada”, “entre menos actividad realices más grasa”, “las grasas se quedan en el cuerpo”. Utilizando tecnología (3 de 11 parejas) y sin el uso tecnología (3 de 11 parejas) la descripción de la relación entre las variables se ve influenciado por el contenido contextual del problema, los estudiantes no ven los datos como simplemente números, sino números con un contexto (Moore, 1990).

Para la interpretación del *coeficiente de correlación* y los *diagramas de dispersión*, el mecanismo determinista es evidente cuando la fuerza de la relación se evalúa a través de un subgrupo de los datos, aquellos que tienen valores grandes y un comportamiento creciente (8 de 12 parejas), por ejemplo: “Los carbohidratos porque contienen mayor número de gramos”, “los carbohidratos influyen con mayor fuerza, porque a partir del cheese burger el número de carbohidratos es mayor que la grasa”. Este tipo de razonamiento también es reportado por Sánchez-Cobo et al. (2000) con estudiantes universitarios después de la intervención de enseñanza quienes hacen referencia solo a alguna característica de los datos en el diagrama de dispersión para establecer la correlación. Además, las creencias sobre el *contexto* del problema también influyen al juzgar la fuerza de la relación (7 de 12 parejas), los estudiantes afirman, por ejemplo, “los carbohidratos influyen en el aporte calórico, ya que tienen más energía y es la que más aporta de los otros datos”, “la grasa influye con más fuerza, ya que no es tan necesaria y se encapsula en el cuerpo humano”.

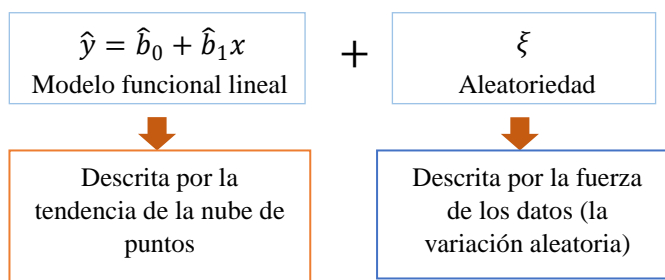
En la *recta de mejor ajuste*, el mecanismo determinista consiste en buscar y aislar los datos bivariados que se apeguen a un modelo (ecuación). Por ejemplo, algunos estudiantes dividen el conjunto de datos en dos partes, los que corresponden a un modelo y los que no lo hacen; ellos privilegian los puntos que expresan una dependencia funcional o trazan la unión por semi-rectas entre todos los puntos. En estudios previos (Casey, 2014; Medina et al., 2019) han evidenciado esta estrategia cuando los estudiantes exploran los puntos de una nube para trazar lo que para ellos es la recta de mejor ajuste es “que pase por aquellos puntos que pertenecen a un modelo lineal”. Casey (2015) resalta que el aprendizaje previo de función lineal en matemáticas puede interferir en la capacidad de dar sentido a la tarea de ajustar una línea, como una línea que no va necesariamente a través de todos los puntos. Esta afirmación de Casey (2015) de la influencia previa de la función lineal también se evidenció en algunos

códigos cuando utilizan algunos valores y establecen una expresión aritmética al enfrentar un problema de estimación y predicción. Esto ocurre con mayor frecuencia al predecir (6 de 12 parejas) y estimar (8 de 12 parejas) a partir de la situación problema a lápiz y papel, y solo 2 de 12 parejas utilizaron algunos valores para establecer expresiones aritméticas para predecir y estimar utilizando tecnología.

La *covariación estadística*, es una categoría propia de la covariación, y ocurre cuando los datos no se ajustan de modo preciso a una función matemática, pero su comportamiento se representa por un modelo matemático debido a la presencia de la variabilidad del conjunto de datos bivariados. En un inicio, sin utilizar tecnología, la descripción de los estudiantes sobre la relación entre las variables se basa en la característica de variabilidad (5 de 11 parejas) que observan en los valores de la tabla, por ejemplo: “la grasa aumenta o disminuye dependiendo de cuántas calorías queman en ANE” “las calorías que se queman aumentaron los kg o las calorías que se quemaron disminuirán los kg”; estas respuestas evidencian que los estudiantes determinan el comportamiento no constante que posee el conjunto de datos bivariados. Cuando utilizan la plataforma CODAP para realizar el diagrama de dispersión y de nuevo describir la relación entre las variables, los estudiantes dan una descripción global del comportamiento de la nube (8 de 11 parejas), por ejemplo: “cuando el cambio ANE es mayor el aumento de grasa es menor”, “Entre más cambio en ANE, menos aumento de grasa presenta”, lo cual muestra que dan cuenta de la dirección negativa de las variables, es decir, mientras una variable aumenta la otra disminuye, estableciendo esta relación a partir de una vista agregada del conjunto de puntos y no se produce una vista dirigida a casos individuales para establecer la relación entre las variables. Los estudiantes establecen la aleatoriedad en los datos bivariados a través de la tabla de valores y el contexto del problema, y luego con ayuda de la tecnología determinan la tendencia de los datos observando la nube de puntos de manera global, esto lleva a pensar que es posible promover razonamientos acerca de la covariación estadística lineal como aquella que se compone de dos elementos: i) un modelo lineal (visto como la descripción de la tendencia global de la nube de puntos) , y ii) la aleatoriedad (expresada por la variabilidad de los datos). Estos elementos hacen referencia a un modelo aleatorio lineal definido por Batanero (2001), el cual se compone de un modelo funcional más un error aleatorio ($\hat{y} = \hat{b}_0 + \hat{b}_1x$), donde, el modelo lineal hace referencia a

la relación lineal que modela el comportamiento de la nube de puntos, más el error aleatorio (ξ) que describe lo distante que está una respuesta individual de la línea de regresión poblacional. De los resultados se tiene la hipótesis que posible promover razonamientos acerca de la covariación estadística lineal como aquella que se compone de dos elementos (ver figura 6.1): i) un modelo lineal (visto como la descripción de la tendencia global de la nube de puntos), y ii) la aleatoriedad (expresada variación en los datos).

Figura 6.1. *Componentes de la covariación estadística*



La categoría *dispersión*, es propia de la dimensión correlación y diagramas de dispersión; la cual se presenta cuando se establece una relación entre la dispersión de la nube de puntos del diagrama y el coeficiente numérico de la correlación, por ejemplo, afirmaciones que dan cuenta que “a mayor valor numérico del coeficiente de correlación el comportamiento de la nube de puntos se conglo mer a de manera lineal”, “el valor de la correlación es menor y hace que los puntos de la nube estén un poco dispersos”. Estas afirmaciones se presentaron solo cuando se utiliza la plataforma para realizar los diagramas de dispersión y calcular el coeficiente de correlación. A diferencia del estudio de Inzunza (2016) quien reporta que los estudiantes en sus respuestas no relacionan la interpretación del diagrama de dispersión con el coeficiente de correlación, en esta investigación la mayoría de las respuestas evidencian que relacionan la dispersión de la nube de puntos con el valor numérico de la correlación, quizás se debe a la visualización simultánea de los diagramas de dispersión con el valor numérico de la correlación, sin embargo, a diferencia de la investigación de Inzunza en los

estudiantes en nuestro estudio no hacen referencia al contexto para describir la relación entre las variables que influye fuertemente y débilmente.

La categoría *distancia (nube y la recta)*, se define la recta que mejor se ajusta a los datos como aquella donde la distancia de la recta a la nube de puntos sea pequeña y hace que el valor numérico de la suma del valor absoluto de los residuos sea el menor, además, esta recta representa toda la información de los datos. En estudios previos (Casey, 2014; Medina et al., 2019) en los que se ha explorado la manera en que los estudiantes utilizan los puntos de una nube para trazar lo que para ellos es la recta de mejor ajuste se han encontrado estrategias como “que la recta pase por la mayoría de los puntos” o “que pase el primero y el último” o “que pase por aquellos que pertenecen a un modelo lineal”. Estos ejemplos son evidencia de una visión orientada a casos para establecer el criterio de la recta de mejor ajuste, dado que ajustan bajo características del conjunto de datos a nivel de casos individuales. Casey (2014, 2015) reporta criterios como “que los puntos estén lo más cerca de la recta” o “que haya un número igual de puntos a cada lado de la recta” lo cual identifican como una visión agregada de los datos, dado que requiere asumir el conjunto de puntos como un todo con características no necesariamente visibles en ninguno de los casos individuales. En esta investigación se observó que 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 absolutas 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”.

Una noción que emergió del proceso analítico en las categorías anteriormente descritas es “*el agregado*”. 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 ven los datos como un *agregado*. El término agregación es utilizado por Stigler (2016) para designar el primer pilar de la sabiduría estadística. Para él la agregación es el mecanismo mediante el cual se destaca la información de un conjunto de datos como un todo en detrimento de la información de los datos individuales; en la agregación hay una pérdida de información de los datos individuales a cambio de conservar propiedades globales del conjunto de datos.

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, de igual manera ocurre cuando juzgan la relación a partir de un par de datos de la tabla o para establecer la fuerza de la relación teniendo en cuenta solo un subconjunto de datos. 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. El describir la tendencia global y la fuerza de los datos al juzgar la relación bivariada. Por otro lado, también el contexto es un factor importante para ver a un conjunto de datos como agregado, 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 de una u otra forma por la variabilidad de la

población que se observa. 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).

Evidenciamos como 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, es la existencia de un mecanismo causal que ayuda a imaginar una unidad en el conjunto de datos.

5.2. Modelos de desarrollo del razonamiento estadístico

Antes de describir la teoría humilde y sustantiva que emergió en la codificación avanzada, en este apartado se muestra una introducción a los modelos cognitivos de desarrollo del razonamiento estadístico, dado que en este estudio proponemos un modelo de desarrollo para el razonamiento estadístico covariacional.

En la psicología, siempre ha habido tensión entre la perspectiva de Piaget y el marco constructivista por sus diferencias acerca del crecimiento intelectual de los niños, por un lado, un modelo del conocimiento en términos de estructuras universales biológicas y, por otro lado, la caracterización del conocimiento como producto de la actividad creativa de autoorganización del niño en contextos específicos; estas perspectivas han dejado por fuera el papel del entorno y la intervención educativa en las etapas de desarrollo cognitivo (Jones et al., 2004).

Investigadores del desarrollo cognitivo posteriores a Piaget y al constructivismo, han estudiado los modelos de desarrollo donde conocimiento se organiza dentro de dominios específicos definidos por contenido o tareas particulares como las involucradas en la exploración de datos y el razonamiento estadístico. Por ejemplo, el modelo de Biggs y Collis (1982) ha sido evolutivo desde la estructura de la taxonomía de resultados de aprendizaje observados (SOLO). Jones et al. (2004) reconocen que los modelos de desarrollo se ocupan

del conocimiento de un dominio específico, y asumen el término de *modelo cognitivo de desarrollo en el razonamiento estadístico* para hacer referencia a una teoría que sugiere diferentes niveles o patrones de crecimiento en el razonamiento estadístico que resultan de los efectos madurativos o interacciones con entornos de aprendizaje estructurado y no estructurados. Estos modelos cognitivos examinan los procesos de razonamiento estadístico, como la toma de decisiones, la predicción, la inferencia, y explican como estos procesos son aplicados a la exploración de datos univariados y multivariados.

Jones et al. (2004) establecen que los modelos de desarrollo que se relacionan con el contenido específico de relaciones bivariadas comprende dos grupos: (i) estudios previos a los modelos definitivos que han examinado el desarrollo de las concepciones de los estudiantes acerca de la asociación y regresión (Batanero et al., 1996; Meraverch y Kramarsky, 2017; Ross y Cousins, 1993; Wavering, 1989), y (ii) los modelos cognitivos más definitivos (Moritz and Watson, 2000; Moritz, 2001; Mooney, 2002). Jones et al. (2004) solo menciona los estudios para diferentes dominios específicos, pero no los describe, por lo tanto, realizamos una revisión para conocer y estudiar acerca de los modelos de desarrollo que involucran las temáticas de correlación y regresión lineal.

El estudio de Wavering (1989) se enfocó en el razonamiento lógico de graficas lineales de estudiantes de sexto grado a duodécimo grado. Solo emplea tres tipos de gráficos, una línea recta con pendiente positiva, una línea recta con pendiente negativa y una curva exponencial. El investigador clasifica las respuestas de los estudiantes en nueve categorías, las cuales las caracteriza en términos Piagetianos. Ross y Cousins (1993) aplicaron tres experimentos que involucraban problemas de correlación con más de dos variables continuas a estudiantes de 5, 7, 9 y 11 grado y un curso de 20 estudiantes de maestría en educación. Los investigadores plantean un esquema de puntuación para cuatro habilidades de correlación. Batanero et al. (1996) examinan las estrategias intuitivas y preconcepciones acerca de la asociación en tablas de contingencia de 213 estudiantes preuniversitarios. Finalmente, Mevarech y Kramarsky (1997) estudiaron las concepciones de los estudiantes de grado octavo al representar gráficamente descripciones verbales, estos investigadores establecieron tres categorías principales de respuestas.

Los estudios de Watson y Moritz (1997, 2000, 2007) y Moritz (2001, 2004) con estudiantes de 5, 6, 7, 8, grado 9 y 10 de secundaria y bachillerato se enfocaron en analizar las respuestas de los estudiantes al realizar representaciones gráficas de dos variables cuantitativas, de tal manera que fueron refinando sus modelos de desarrollo respecto a la representación gráfica de datos bivariados, el modelo más reciente de Watson y Moritz (2007) utiliza el marco de evaluación de Biggs y Collis (1982) Structure of Observed Learning Outcomes (SOLO) que comprende los niveles preestructural, uniestructural, multiestructural y relacional.

La investigación de Mooney (2002) consistió en desarrollar y validar un marco para caracterizar el pensamiento estadístico de estudiantes de bachillerato, este marco comprende 4 procesos (describir datos, organizar y reducir datos, representar datos, analizar e interpretar datos) y 4 niveles (idiosincrático, transicional, cuantitativo y analítico). Su validación la llevó a cabo con estudiantes de 6 a 8 grado. Las preguntas fueron diseñadas para que los estudiantes respondieran de manera oral o escrita, incluyen tablas y gráficos de datos bivariados y multivariados.

La tabla 5.1 resume los estudios que proponen modelos de desarrollo para el razonamiento covariacional, se observa que las temáticas son dispersas respecto a los principales conceptos que están involucrados en la correlación y regresión lineal.

Tabla 5.1. *Estudios previos de Modelo de desarrollo para el razonamiento covariacional*

	Autor	Grado Escolar	Temáticas
Estudios previos de categorización de concepciones y preconcepciones	Wavering (1989)	6 a 12	Gráficas lineales
	Ross y Cousins (1993)	5, 7, 9 y 11 grado y un curso de 20 estudiantes de maestría	Problemas de correlación con más de dos variables continuas
	Batanero et al. (1996)	Preuniversitarios	Asociación en tablas de contingencia
	Mevarech y Kramarsky (1997)	8 grado	Representación gráfica de enunciados verbales

Modelos más definitivos	Watson y Moritz (1997, 2000, 2007)	5, 6, 7, 8, grado 9 y 10 de secundaria y bachillerato	Representación gráfica de datos bivariados
	Moritz (2001, 2004)		
	Mooney (2002)	6 a 8 grado	Tablas y gráficos de datos bivariados y multivariados.

El modelo que propone Mooney (2002) está muy relacionado con el modelo de Jones et al. (2000) y ambos caracterizan el razonamiento estadístico de estudiantes de primaria y secundaria bajo dos premisas teóricas: primero, para exhibir el razonamiento estadístico de los estudiantes, ellos necesitan comprender conceptos de manejo de datos que son multifacéticos y se desarrollan con el tiempo que se pueden caracterizar bajo 4 procesos (describir datos, organizar y reducir datos, representar datos, analizar e interpretar datos); segundo, de acuerdo con el modelo de desarrollo de Biggs y Collis (1991) se asume que el razonamiento de los estudiantes puede caracterizarse como un desarrollo a través de niveles que reflejan cambios en la complejidad de su razonamiento. Es decir, para cada proceso, el razonamiento puede caracterizarse como un desarrollo a través de cuatro niveles de razonamiento (idiosincrásico, transicional, cuantitativo y analítico). A continuación, la tabla 5.2 muestra los descriptores del marco para analizar e interpretar datos a nivel de secundaria propuestos por Jones et al. (2000) y Mooney (2002).

Tabla 5.2. Marco para analizar e interpretar datos a nivel de secundaria propuesto por Jones et al. (2000) y Mooney (2002).

Proceso	Nivel 1 Idiosincrásico	Nivel 2 Transicional	Nivel 3 Cuantitativo	Nivel 4 Analítico
	Leer entre los datos			
Analizar e interpretar datos	Hace comparaciones incorrectas dentro y entre los conjuntos de datos.	Realiza una sola comparación correcta o un conjunto de comparaciones parcialmente correctas dentro	Hace comparaciones locales o globales dentro y entre los conjuntos de datos.	Hace comparaciones globales y locales dentro y entre los conjuntos de datos.

o entre los
conjuntos de
datos

Leer más allá de los datos

Hace inferencias que no se basan en los datos o inferencias basadas en cuestiones irrelevantes.	Realiza inferencias basadas parcialmente en los datos. Algunas inferencias pueden ser solo parcialmente razonables.	Hace inferencias principalmente basadas en los datos. Algunas inferencias pueden ser solo parcialmente razonables.	Hace inferencias razonables basadas en los datos y el contexto.
---	---	--	---

La anterior tabla caracteriza el desarrollo del razonamiento estadístico de los estudiantes de secundaria y pertenece al análisis e interpretación de datos con sus descriptores correspondientes a los cuatro niveles y dos subprocesos que leen entre y más allá de los datos. Es importante resaltar que el marco de Jones et al. (2000) y Mooney (2002) fue elaborado para analizar e interpretar los procesos y niveles de desarrollo cognitivo en el tiempo en los grados escolares de primaria o secundaria, por lo que sus estudios se han aplicado a más de un curso escolar. Ahora bien, nuestra investigación se realizó con un solo curso de bachillerato, y en la codificación intermedia se agruparon afirmaciones que proponen cambios en la complejidad del razonamiento covariacional de los estudiantes, por lo tanto, hemos tomado como referencia solamente los niveles (idiosincrático, transicional, cuantitativo y analítico) que propone los marcos de Jones et al. (2000) y Mooney (2002) para el razonamiento estadístico. A continuación, se describen los aspectos más relevantes para cada nivel propuesto por Jones et al. (2000) y Mooney (2002) y descritos a cabalidad en Jones et al. (2004).

En el *nivel idiosincrático*, es el razonamiento subjetivo que no está relacionada con los datos dados, y a menudo, se enfoca en experiencias personales o creencias subjetivas. Los

estudiantes que razonan a este nivel pueden distraerse o confundirse con aspectos irrelevantes de la situación problema.

En el *nivel transicional*, los estudiantes comienzan a reconocer la importancia del razonamiento cuantitativo, de tal manera que se involucran en una tarea de manera relevante, pero generalmente se enfocan solo en un aspecto de la situación problema.

En el nivel *cuantitativo*, el razonamiento de los estudiantes es consistentemente cuantitativo, en el sentido que pueden identificar las ideas matemáticas de la situación problema y no se distraen o engañan por los aspectos irrelevantes. Sin embargo, no integran necesariamente estas ideas matemáticas relevantes cuando desarrollan la tarea.

En el nivel *analítico*, el razonamiento de los estudiantes se basa en hacer conexiones entre los múltiples aspectos de una situación problema. Su razonamiento en este nivel puede integrar los aspectos relevantes de una tarea en una estructura significativa, por ejemplo, crear múltiples visualizaciones de los datos o hacer una predicción razonable.

En el siguiente apartado, se ilustran estos niveles con los resultados de esta investigación proponiendo un marco para el desarrollo estadístico covariacional de estudiantes de último grado de bachillerato.

5.3. Codificación avanzada: Modelos de desarrollo del razonamiento estadístico covariacional

De la codificación intermedia surgieron cuatro dimensiones (covariación, recta de mejor ajuste, coeficiente de correlación y los diagramas de dispersión). Para nuestra codificación avanzada observamos que las dimensiones comprendían categorías muy definidas por códigos que mostraban diferentes niveles de razonamientos de los estudiantes, por lo tanto, los modelos que proponemos es el resultado teórico de las codificaciones anteriores, y cuentan con los siguientes elementos generales:

1. Se describen todos los niveles de razonamiento (idiosincrático, transicional, cuantitativo y analítico) para los modelos de razonamiento cuando: se juzga la relación entre dos variables (corresponde a la dimensión covariación), se analiza el

coeficiente de correlación junto con el diagrama de dispersión (corresponde a la dimensión correlación), al establecer la recta de mejor ajuste (dimensión recta de mejor ajuste), y un modelo para el razonamiento al estimar y predecir.

2. Algunos de los niveles (cuantitativo y analítico) no se evidenciaron en nuestros resultados. Por lo tanto, se completaron con base en los estándares para bachillerato y en la descripción de los niveles según Jones et al. (2000). Las descripciones de los niveles que se completaron están sombreadas de color celeste.
3. Para los modelos de razonamiento que involucran juzgar la relación bivariada, establecer la recta de mejor ajuste, y el modelo para estimar y predecir surgieron en las codificaciones dos componentes muy marcados, uno referente los códigos que involucran la representación gráfica y el otro aquellos que evidencian el contexto y la tabla de datos bivariada.

5.3.1. Modelo para el razonamiento cuando se juzga la relación entre dos variables.

Situación: Conjunto de datos bivariados; cada entrada corresponde a valores de una variable. **Pregunta:** ¿Cómo están relacionadas las variables?

Tabla 5.3. Modelo para el razonamiento cuando se juzga la relación entre dos variables.

	Nivel	Gráficamente	Tabla de valores bivariada
Juzgar la relación bivariada		Ubica los pares ordenados en el plano cartesiano y se traza la unión por segmentos de recta entre ellos.	<ul style="list-style-type: none"> ○ Realiza operaciones aritméticas con un par de datos para explicar la relación. ○ Describe que la relación entre las variables es proporcional. ○ Describe que la relación entre las variables es dependiente porque ambas aumentan.
	Nivel Idiosincrático		
	Nivel Transicional	Representa los pares ordenados en el plano cartesiano.	<ul style="list-style-type: none"> ○ Describe una relación donde una variable depende de la otra y ambas aumentan de manera no proporcional.

		<ul style="list-style-type: none"> ○ Menciona la tendencia global de los datos numéricos de la tabla de valores (positiva o negativa)
Nivel Cuantitativo	Describe la tendencia global de la nube de puntos si es positiva o negativa sin aceptar su comportamiento lineal.	<ul style="list-style-type: none"> ○ Enfoca la atención en lo que ocurre en el paso de un punto al que le sigue dentro de todos los valores de la tabla.
Nivel Analítico	Acepta un comportamiento lineal de la nube de puntos.	<ul style="list-style-type: none"> ○ Reconoce la distinción entre variables explicativas y de respuesta. Si se cree que una variable x puede explicar o causar cambio en otra variable y, x es llamada la variable explicativa e y es la variable de respuesta o de lo contrario menciona que no hay distinción entre las variables, es decir, no hay variable respuesta, ni explicativa.

5.3.2. Modelo para el razonamiento cuando se analiza el coeficiente de correlación y el diagrama de dispersión.

Situación: Análisis de la fuerza entre dos variables utilizando tecnología para calcular el coeficiente de correlación y obtener el diagrama de dispersión.

Pregunta: ¿Cuál es la variable que influye con mayor fuerza, y cuál es la variable que influye con menor fuerza?

Tabla 5.4. *Modelo para el razonamiento cuando se analiza la fuerza entre las variables.*

	Nivel	Gráficamente
Analizar la fuerza entre las variables	Nivel Idiosincrático	<ul style="list-style-type: none"> ○ Compara valores numéricos de la tabla de datos para establecer la fuerza de la relación. ○ La fuerza entre las variables se establece con base en las ideas o creencias de los aspectos del contexto del problema.
	Nivel Transicional	<ul style="list-style-type: none"> ○ Omite el valor numérico de la correlación haciendo referencia solo a la conglomeración o dispersión de los puntos en el diagrama de dispersión. ○ Observa de manera sectorizada la nube de puntos e identifica los puntos más alejados. ○ Menciona que los puntos están conglomerados cuando la correlación es cercana a 1 o menos conglomerados cuando la correlación es lejana a 1.
	Nivel Cuantitativo	<ul style="list-style-type: none"> ○ Establece que a mayor valor numérico del coeficiente de correlación el comportamiento de la nube de puntos se conglo-mera de manera lineal, y cuando el valor de la correlación es menor los puntos de la nube están dispersos. ○ La correlación cercana a 1 hace que los puntos de la nube estén más cerca y formen una línea recta.
	Nivel Analítico	<ul style="list-style-type: none"> ○ Dado que la correlación es cercana a 1 y los puntos de la nube están más juntos en el diagrama de dispersión está es la variable que influye fuertemente. ○ Dado que en el diagrama de dispersión los puntos están alejados y tienen una correlación menor alejado de 1 influye débilmente.

5.3.3. Modelo para el razonamiento al establecer la recta de mejor ajuste.

Situación: La nube de puntos que representa un conjunto de datos bivariado y una recta movible para manipular en algún applet interactivo.

Pregunta: ¿Cuál criterio debe cumplir la recta movable para ser la recta de mejor ajuste a la nube de puntos?

Tabla 5.5. Modelo para el razonamiento al establecer la recta de mejor ajuste.

	Nivel	Gráficamente	Tabla de valores bivariada
Establecer la recta de mejor ajuste	Nivel Idiosincrático	<ul style="list-style-type: none"> La línea es aquella que contiene un par de puntos o un subgrupo de puntos. 	<ul style="list-style-type: none"> Utilizar un par de valores para realizar algún procedimiento matemático (hallar pendiente y punto de intersección con el eje Y)
	Nivel Transicional	<ul style="list-style-type: none"> Describe la línea de mejor ajuste como aquella que se acerca lo más posible a la nube de datos. 	<ul style="list-style-type: none"> Se emplea la media de las variables como criterio para establecer la recta de mejor ajuste
	Nivel Cuantitativo	<ul style="list-style-type: none"> Describe que el valor numérico de los residuos cuando es menor hace que la distancia entre la posible línea de mejor ajuste y la nube de puntos sea pequeña. Asume el valor total de los residuales como aquel valor que representa la cercanía o lejanía de la línea de ajuste. 	<ul style="list-style-type: none"> Encontrar la recta que pasa por el centro de gravedad (x,y) y otro punto.
	Nivel Analítico	<ul style="list-style-type: none"> Menor residuos evidencia que se enfatiza en la línea de mejor ajuste como aquella que debe representar la tendencia de acumulación del conjunto de datos más o menos dispersos centrándose en medio de la nube de manera que el residuo es menor en esa posición 	<ul style="list-style-type: none"> Según la métrica utilizada de mínimos cuadrados es aquella línea que hace que la suma de los cuadrados de las distancias verticales entre los valores observados (y) y los valores predichos

(\hat{y}) sea lo más pequeña posible.

5.3.4. Modelo para el razonamiento al estimar y predecir

Situación: Responder a una pregunta de predicción y estimación.

Preguntas: ¿Cómo estimar un valor y , dado un valor x que se encuentra entre dos datos del conjunto bivariado dado? ¿Cómo predecir un valor y , dado un valor x alejado del conjunto de datos bivariado dado?

Tabla 5.6. Modelo para el razonamiento al estimar y predecir.

	Nivel	Gráficamente	Tabla de valores bivariada
Estimar y predecir	Nivel Idiosincrático	Utiliza aspectos irrelevantes del diagrama de dispersión.	<ul style="list-style-type: none"> ○ Se enfoca en experiencias personales o creencias subjetivas de la situación problema.
	Nivel Transicional	Al azar establece un valor con base en el comportamiento de la nube de puntos.	<ul style="list-style-type: none"> ○ Buscar un patrón en la tabla de valores para establecer un procedimiento aritmético. ○ Sin patrón definido se efectúan algunas operaciones básicas (suma, resta, multiplicación y división).
	Nivel Cuantitativo	Ubican a ojo para el valor, pero no sobre la recta no sobre la recta	<ul style="list-style-type: none"> ○ Utiliza un par de valores de la tabla y realizar una regla de tres para estimar y predecir.
	Nivel Analítico	Emplear la línea arrojada por el software para estimar y predecir el valor lo cual implica asumir la línea como representante de todos los datos.	<ul style="list-style-type: none"> ○ Utilizar la recta de regresión como modelo para estimar y predecir. Un valor predicho no es el mismo que el valor real (en la mayoría de los casos).

CAPÍTULO VI. CONCLUSIONES

El capítulo de conclusiones comprende: Primero, la exposición de tres temas que consideramos son conclusiones relevantes producto de los resultados obtenidos, y son otros de los aportes de nuestra investigación respecto al razonamiento estadístico covariacional de los estudiantes, los temas son:

- El papel de la tecnología en el desarrollo del razonamiento sobre la regresión y la correlación
- La transición de ver un conjunto de datos singulares a verlo como un agregado
- De acercamientos determinísticos al reconocimiento de la incertidumbre

Segundo, la respuesta a cada pregunta de investigación. Tercero, la descripción de las limitaciones de este trabajo y finalmente, las implicaciones para la didáctica de la estadística, y la prospectiva de la investigación.

6.1. El papel de la tecnología en el desarrollo del razonamiento sobre la regresión y la correlación

En educación estadística, podemos distinguir dos perspectivas de uso de la tecnología en la enseñanza. La primera, cuando es utilizada para organizar datos, construir gráficos y realizar cálculos matemáticos, y, por otro lado, aquella cuyo enfoque es el uso de la tecnología dinámica interactiva para desarrollar la comprensión conceptual (Biehler et. al, 2013). Este estudio muestra evidencia que ambas perspectivas contribuyen a generar razonamientos acerca de la tendencia y la fuerza de la relación asociada al modelo matemático lineal cuando se realicen diseños de enseñanza apropiados.

Las investigaciones que han utilizado la tecnología para realizar cálculos tediosos y gráficos afirman que ayuda a disminuir el desfase entre la comprensión del concepto y los medios técnicos de cálculo para poder aplicarlos (Batanero y Díaz, 2011), específicamente, para los conceptos involucrados en la correlación y regresión lineal se evidencia que:

La construcción del diagrama de dispersión para un conjunto de datos bivariados genera razonamientos acerca de la tendencia de los datos, por ejemplo, se percibe el comportamiento lineal positivo de los datos al verificarse que los valores de una variable tienden a aumentar conforme los valores de la otra variable también aumentan. Además, se

percibe cuando los valores de una variable tienden a aumentar conforme los valores de la otra tienden a disminuir, lo cual corresponde a una relación con dirección negativa. Sin embargo, estas percepciones no son suficientes para incluir la variabilidad en los datos y algunos razonamientos pueden ser de tipo funcional donde la relación que establecen entre las variables es de proporcionalidad, sin poder dar cuenta de la variabilidad en los datos.

El cálculo del coeficiente de correlación y el establecimiento de la relación entre su valor numérico y los datos representados en el diagrama de dispersión generan razonamientos que dan cuenta de la noción de fuerza en la relación, donde la variabilidad está presente, por ejemplo, cuando se establece que a mayor correlación los datos representados en el gráfico se conglomeran de manera lineal y a menor correlación se presenta mayor dispersión en los datos. Sin embargo, no llegan a relacionar el signo numérico del coeficiente de correlación con la dirección de la nube de puntos.

La determinación de la ecuación lineal de la recta con tecnología no conlleva de manera inmediata su utilización para dar solución a un problema de estimación y predicción. Es difícil para los estudiantes asumir la línea como representante de todos los datos, por lo tanto, creemos que además de calcularla es necesario guiar a los estudiantes para que aprendan a sacar consecuencias de ella, por ejemplo, deben inferir que es un modelo de comportamiento general de los datos y que pueden hacer predicciones.

Por otro lado, la tecnología dinámica interactiva permite a los estudiantes relacionar múltiples representaciones (visual, simbólica, numérica y verbal) y conectar tales representaciones para apoyar el razonamiento y la solución de problemas. Por ejemplo, los elementos establecidos en el *applet* de GeoGebra permiten que al interactuar y manipular con el objetivo de establecer la posible recta de mejor ajuste se generen razonamientos relacionados con la noción de distancia entre los puntos y la posible línea, que el residuo cambia para cada valor observado, que entre más alejada la recta de los puntos es mayor el valor de la suma de todos los residuos y cuando esta cerca el valor de los residuos es menor, además, ésta pasa centrándose en medio de los puntos, evidenciándose la noción de cercanía o lejanía.

En conclusión, las dos perspectivas de uso de la tecnología promueven el desarrollo del razonamiento sobre la regresión y la correlación, pues el estudiante se vuelve autónomo en el manejo del software para resolver un problema estadístico y la manera como logra articular lo arrojado por el software con el lenguaje estadístico para comunicar y expresar resultados. Sin embargo, es un desafío para la investigación realizar buenos diseños de enseñanza utilizando tecnología, que apoye y genere pensamientos acerca de ideas estadísticas que vayan más allá de utilizarla para obtener cálculos. Un enfoque que solo atienda el uso de la herramienta digital para realizar cálculos no tiene demasiado sentido, es necesario que se utilice como un instrumento de mediación para el desarrollo cognitivo de los estudiantes. Cuando el profesor abandone la idea que es más importante la habilidad para realizar cálculos y en utilizar recetas algorítmicas, podrá aceptar otras formas de desarrollo del razonamiento en los estudiantes.

6.2. La transición de ver los datos individuales a verlos como un agregado con la correlación y regresión lineal

Establecer las propiedades de agregado en un conjunto de datos es una característica esencial del análisis de datos y la inferencia estadística (Hancock, Kaput y Goldsmith, 1992; Konold y Higgins, 2002; Ben-Zvi y Arcavi, 2001). Se considera que el desarrollo de la vista agregada de un conjunto de datos es un desafío en la educación estadística (Bakker et al. 2004). Encontramos que la transición de ver los datos individuales a verlos como un agregado no es un proceso fácil e inmediato cuando se estudia la correlación y la recta de mejor ajuste. Nuestra investigación muestra evidencia de que es posible promover una vista agregada de los datos cuando se estudia la correlación y regresión.

En el análisis de la correlación lineal, los razonamientos de los estudiantes deben involucrar deducciones que den evidencia de tratar el conjunto de datos desde una vista agregada, por ejemplo, los conceptos de fuerza y dirección de la nube de puntos representada en el diagrama de dispersión implica realizar juicios globales del conjunto de datos. No tienen sentido los conceptos de fuerza y dirección para puntos aislados ni la noción de modelo se refiere solo a una parte de los datos sino a su totalidad.

Algunos estudiantes establecen el criterio de recta de ajuste para un subconjunto de puntos de la nube en lugar de establecer el criterio asumiendo todo el conjunto de puntos, para Casey (2015) estos estudiantes están en transición de una vista orientada a casos a una vista agregada de datos. Por otro lado, un criterio que evidencia una vista agregada es aquel que involucra la “distancia de todos los puntos de la nube a la recta”; la propiedad de considerar la suma de residuos como una distancia es generada por el hecho de que, para obtenerla cada punto de la nube participa con su residuo. La noción de distancia es una propiedad global del conjunto de puntos, de tal manera que si un punto se altera o es eliminado cambia la distancia de la nube a la recta. Poseer una vista agregada del conjunto de puntos es primordial para establecer el modelo lineal aleatorio conocido como la recta de mejor ajuste donde se tienen en cuenta todos los puntos de la nube de manera articulada. Además, el modelo lineal es visto como un todo porque representa tanto los datos explícitos (nube de puntos) como los implícitos (aquellos que se pueden estimar y predecir).

Stigler propuso a la agregación como el primer pilar de la sabiduría estadística y muestra cómo es que pensar en un agregado es mucho más potente que pensar en los casos particulares (Stigler, 2016). Concebir a un conjunto de datos bivariados desde una vista agregada es una condición que favorece la comprensión de los temas de regresión y correlación. Por lo tanto, 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 solo para realizar actividades con sus estudiantes sobre el tema de correlación y regresión, sino también para diseñar actividades que involucren la emergencia de propiedades globales de conjuntos de datos.

6.3. De acercamientos determinísticos al reconocimiento de la incertidumbre

La estadística es un método general para resolver problemas en situaciones donde el sujeto trata con datos, variación e incertidumbre (Moore 1990). Particularmente, la variación estadística no puede ser explicada, es incertidumbre (Tal, 2001); un tipo de incertidumbre es conocida como la medida del error aleatorio. De nuestro análisis, mostramos que los estudiantes inicialmente muestran acercamientos determinísticos, y debido al diseño de las situaciones de los problemas emerge la conciencia de la incertidumbre; esto se manifiesta

cuando emiten juicios de correlación, así como cuando tratan de determinar la recta de mejor ajuste.

Inicialmente, los estudiantes juzgan la relación entre dos variables de manera proporcional o funcional. Estepa y Batanero (1996) denominaron concepción determinista cuando los estudiantes esperan una correspondencia donde a cada valor de la variable dependiente se le asigne otro valor de la variable independiente, y ocurre cuando sus respuestas hacen referencia solo a un par o subconjunto de datos que corresponden al modelo e ignoran aquellos que quedan fuera. También se presenta en la interpretación del coeficiente de correlación y los diagramas de dispersión cuando evalúan la fuerza de la relación solo para un subgrupo de los datos, aquellos que tienen valores grandes y un comportamiento creciente. Para la recta de mejor ajuste, se presentan acercamientos determinísticos al trazar segmentos de rectas uniendo los puntos del diagrama de dispersión, o aislar solo los datos que se apegan al modelo lineal (ecuación).

Con la exploración realizada en esta investigación, establecemos que los estudiantes reconocen la incertidumbre en tres situaciones específicas:

1. Cuando dan cuenta de la variabilidad en el conjunto de datos bivariados. Esto significa que reconocen que las medidas contienen cierta incertidumbre, pues los datos no se ajustan de modo preciso a una función matemática, pero reconocen que su comportamiento se puede representar por un modelo matemático teniendo en cuenta los errores o residuos.
2. Al comparar la medida del coeficiente de correlación junto con el diagrama de dispersión describen la tendencia de la nube de puntos junto con la fuerza presente en los datos. Es decir, al describir la tendencia establecen el modelo lineal, y con la descripción de la fuerza reconocen la aleatoriedad expresada por la variabilidad de los datos.
3. Al establecer el criterio de la recta de mejor ajuste cuando relacionan que la suma de los residuos debe tomar el mínimo valor, están reconociendo que deben minimizar el error aleatorio para establecer el criterio de mejor ajuste.

Los resultados obtenidos en esta investigación muestran que es posible promover razonamientos que involucren la incertidumbre, sin embargo, es necesario diseñar más tareas en las que los estudiantes hagan inferencias teniendo en cuenta el error aleatorio presente en la correlación y la recta de regresión. Además, falta indagar acerca del rol del contexto de los problemas como una variable de tarea que influye sobre el razonamiento de la regresión y la correlación.

6.4. Respuesta a cada pregunta de investigación.

Primera pregunta de investigación: ¿Cómo cambian los razonamientos que muestran los estudiantes frente a tareas de correlación y recta de mejor ajuste antes y después de la implementación de las tareas instruccionales?

Las jerarquías que se han construido (sección 5.3, Capítulo V), en general, dan cuenta de los cambios que pueden tener los razonamientos de los estudiantes al enfrentar tareas que involucren la correlación y recta de mejor ajuste.

Los primeros razonamientos al juzgar la relación bivariada fueron de tipo determinístico, como realizar operaciones aritméticas utilizando solo un par de valores para establecer una relación de proporcionalidad entre las variables. Estos razonamientos cambian durante la implementación de las tareas y da lugar a razonamientos a nivel cuantitativo como la descripción global del comportamiento de los datos (positiva, negativa), pero sin mencionar el comportamiento lineal.

Los razonamientos al analizar el coeficiente numérico de la correlación junto con los diagramas de dispersión para establecer la fuerza entre las variables cambiaron notoriamente al utilizar tecnología, dado que se evidenció una asociación entre el valor numérico interpretado como la fuerza entre las variables y el comportamiento de la nube de puntos en el diagrama de dispersión. Inicialmente, se obtuvieron razonamientos a un nivel idiosincrático, mostrando influencia de las creencias respecto a los rasgos del contexto del problema, luego cambiaron al evidenciarse razonamientos a nivel analítico donde se asocia una correlación cercana a 1 con la conglomeración de los datos representados en el diagrama de dispersión.

Frente a las tareas de análisis de la recta de mejor ajuste, algunos razonamientos iniciales de utilizar un subgrupo de puntos en el diagrama de dispersión para establecer el criterio de la recta de mejor ajuste cambiaron a razonamientos que evidencian el ver a una nube de puntos como un agregado a partir de un mecanismo matemático que los une con la noción de distancia de una nube a una recta, esto ocurrió con uso de tecnología. Por lo tanto, los estudiantes abandonan la 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, se evidenció que algunos pasan de establecer operaciones aritméticas utilizando un par de valores para estimar y predecir a emplear la expresión algebraica de recta arrojada por el software como representante de todos los datos para estimar y predecir.

Segunda pregunta de investigación: ¿Cómo se relacionan los cambios en los razonamientos de los estudiantes con los elementos del diseño, en especial, con las características de los problemas y la herramienta computacional?

Hay dos planos en los que conviene analizar los cambios en relación con los elementos del diseño, y son: el plano matemático y el plano estadístico.

En el *plano matemático* se destaca el impacto del diseño, al evidenciarse razonamientos que dan cuenta de la tendencia y patrones de la nube de puntos, de igual manera al establecer la dirección (positiva, negativa) y la fuerza entre las variables asociada a la medida numérica del coeficiente de correlación. Inicialmente, no se evidenciaron razonamientos que dieran cuenta que gráficamente los puntos no se ajustan de un modo preciso a una función matemática, sino que se acumula en la proximidad de una línea. Sin embargo, con el apoyo de los recursos tecnológicos y el diseño de las actividades, estos permitieron 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, evidenciándose la noción de distancia de una recta a un conjunto de puntos, la cual es producto de interactuar con el applet, pues permite calcular los residuos, donde 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 comienzan a dar cuenta que la recta que minimiza l es la recta más cercana a los puntos, es decir, la de mejor ajuste.

En el *plano estadístico*, la elección de los problemas en contextos accesibles para los estudiantes permitió promover razonamientos asociados a la noción de variabilidad de los datos bivariados. Inicialmente, los estudiantes establecen la relación entre las variables con base en la variabilidad que muestran los datos de manera tabular, es decir, notan la aleatoriedad en los datos que se dan de manera tabular, y con el implemento de la tecnología describen la relación de los datos bivariados de manera global, pues su descripción es con base en la tendencia de la nube de puntos que muestra el diagrama de dispersión, lo cual se puede interpretar en el plano estadístico como el establecimiento de un modelo aleatorio lineal que se compone de dos elementos: el análisis de tendencia lineal de las dos variables cuantitativas como el modelo matemático (función lineal) más una aleatoriedad presente en la descripción de la variación de los datos.

En conclusión, destacamos que la tecnología apoya ambos aspectos matemático y estadístico, pero con base lecciones y protocolos diseñadas con propósitos de aprendizaje bien definidos.

Tercera pregunta de investigación: ¿Qué concepciones y dificultades se revelan en las respuestas de los estudiantes cuando abordan las tareas instruccionales?

Los razonamientos evidenciados dan cuenta que la noción de “error” vista como el residual en el *applet* de GeoGebra, y el cual corresponde a la diferencia en valor absoluto entre el valor observado y el verdadero valor de la variable de interés perteneciente a la población, es difícil de promover en los estudiantes, pues prevalece la estrategia de utilizar un par de puntos para resolver un problema de predicción o estimación, en lugar, de ver la recta como representante tanto de los datos explícitos (nube de puntos) como los implícitos (aquellos que se pueden interpolar, estimar y predecir). Esto es evidencia que no es natural para los estudiantes comprender la diferencia entre cada uno los valores dados por el problema y sus valores ajustados correspondientes en la recta de regresión.

Por un lado, el contexto fue útil permitiendo modelos de pensamiento adecuados, 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, es decir, todos los datos están relacionados, por ejemplo, para la situación donde las actividades no relacionados con el ejercicio y la acumulación de grasa son parte de un proceso causal y que se manifiesta por el fenómeno de variabilidad de la población observada. Sin embargo, 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. Por otro lado, el contexto dificultó la comprensión de las ideas fundamentales para establecer la fuerza de la relación entre las variables, en algunos casos prevalece las creencias de los elementos implicados en el contexto para explicar la correlación.

La evidencia de nuestro estudio da cuenta que los aspectos matemáticos complejos detrás de los procedimientos son difíciles, pero pueden ser aligerados con la ayuda de la tecnología.

Limitaciones de esta investigación

Finalmente, identificamos ciertas limitaciones en nuestro estudio que podrían ser mejoradas en un estudio posterior. Uno de ellos, fue realizar iteraciones del experimento para mejorar el diseño. Por esto, se pretendía realizar una segunda exploración, pero debido a la situación de la pandemia ocasionada por el COVID no se logró realizar. Los instrumentos fueron diseñados con base en la experiencia de los investigadores, pero no fue posible una validación más amplia debido a que su implementación debía hacerse en un periodo escolar propicio que solo se presenta cada año. Además, el hecho de utilizar computadoras genera dificultad para la aplicación de los instrumentos, pues la disponibilidad de las salas de computadoras es limitada, así como encontrar instituciones educativas que cuenten con salas de computadores.

Con relación al diseño de los instrumentos, creemos que se pueden utilizar distancias cuadráticas en el diseño de la tarea en lugar de las distancias lineales que se utilizaron, así como enriquecer los problemas que den lugar a nube de puntos que tengan diferentes tendencias, incluyendo la independencia. Enfocarse más en la relación entre el valor de cada residual y posible recta de mejor ajuste, específicamente, en la actividad 3 en la cual se

interactúa con el *applet* de GeoGebra incluir interrogantes como: ¿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 la nube de puntos azules? ¿Qué ocurre con el valor de los residuos?

Implicaciones para didáctica de la estadística y prospectiva de la investigación

A pesar de las limitaciones descritas anteriormente en relación con esta investigación, los resultados sugieren algunas implicaciones prácticas para los profesores de los cursos de estadística. Por ejemplo, creemos que el diseño y el análisis de la experiencia que exponemos en este trabajo puede ofrecer una modesta orientación para los profesores de estadística no solo 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.

El acercamiento informal a la recta de regresión utilizando tecnología bajo el contexto de los problemas planteados ofrecen un camino para abordar conjuntos de datos más grandes y en contextos más complejos.

Al igual que Pfannkuch (2011) consideramos que los estudiantes pueden tener dificultades para establecer conexiones importantes entre lo que saben sobre el contexto de los datos y lo que están aprendiendo estadísticamente, falta profundizar en establecer estas diferencias entre lo que aprende estadísticamente y la influencia del contexto sobre la abstracción formal de los conceptos estadísticos.

REFERENCIAS BIBLIOGRÁFICAS

- Akker, J., Gravemeijer, K., McKenney, S., & Nieveen, N. (2006). *Educational design research*. Routledge.
- Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, 3(2), 64–83.
- Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. In D. Ben-Zvi & 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
- Barab, S., & Squire, K. (2004). Design-Based Research: Putting a Stake in the Ground. *Journal of the Learning Sciences*, 13(1), 1–14. https://doi.org/10.1207/s15327809jls1301_1
- Batanero, C. (2001). *Didáctica de la Estadística*. Grupo de Investigación en Educación Estadística, Departamento de Didáctica de la Matemática, Universidad de Granada.
- Batanero, C., Godino, J. D., & Estepa, A. (1998). Building the meaning of statistical association through data analysis activities. In A. Olivier & K. Newstead (Eds.), *Proceedings of the 22 Conference of the International Group for the Psychology of Mathematics Education* (pp. 221–236). University of Stellenbosch.
- Batanero, C., Estepa, A., & Godino, J. D. (1997). Evolution of students' understanding of statistical association in a computer based teaching environment. *Research on the Role of Technology in Teaching and Learning Statistics*, 191–205.
- Batanero, Carmen, Estepa, A., Godino, J. D., & Green, D. R. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education*, 27(2), 151–169. <https://doi.org/10.2307/749598>
- Behar Gutiérrez, R., & Grima Cintas, P. (2010). *55 respuestas a dudas típicas de estadística* (P. Grima Cintas (ed.); 2ª ed) [Book]. Diaz de Santos.
- Ben-Zvi, Dani, Aridor, K., Makar, K., & Bakker, A. (2012). Students' emergent articulations of uncertainty while making informal statistical inferences. *ZDM - Mathematics Education*, 44(7), 913–925. <https://doi.org/10.1007/s11858-012-0420-3>
- Biehler, R. (1990). Changing conceptions of statistics: A problem area for teacher education. In A. Hawkins (Ed.), *Training teachers to teach statistics* (pp. 20–29). International Statistical Institute.
- Biehler, Rolf, Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. In M. A. Clement, J. Bishop, C. Keitel, J. Kilpatrick, & A. Y. L. Leung (Eds.), *Third International Handbook of Mathematics Education* (pp. 643–689). Springer. https://doi.org/10.1007/978-1-4614-4684-2_21
- Biggs, J., & Collis, K. (1982). *Evaluating the Quality of Learning: The SOLO Taxonomy (Structure of the Observed Learning Outcome)*. Academic Press.

<https://doi.org/10.1177/089202068700100412>

- Birks, M., & Mills, J. (2015). *Grounded theory: A practical guide (2nd ed)*. CA: Sage.
<https://doi.org/10.13140/2.1.2982.0484>
- 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. (2014). Teacher's knowledge of students' conceptions and their development when learning linear regression. In K. Makar, B. Sousa, & 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., & 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>
- Castro Sotos, A. E., Vanhoof, S., Van Den Noortgate, W., & Onghena, P. (2009). THE TRANSITIVITY MISCONCEPTION OF PEARSON'S CORRELATION COEFFICIENT. *Statistics Education Research Journal*, 8(2), 33–55. <https://doi.org/10.52041/serj.v8i2.394>
- Chance, B. L. (2002). Components of Statistical Thinking and Implications for Instruction and Assessment. *Journal of Statistics Education*, 10(3), 1–14.
<https://doi.org/10.1080/10691898.2002.11910677>
- Chun Tie, Y., Birks, M., & 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., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design Experiments in Educational Research. *Educational Researcher*, 32(1), 9–13.
<http://www.jstor.org/stable/3699928>
- Cobb, P., & McClain, K. (2004). Principles of Instructional Design for Supporting the Development of Students' Statistical Reasoning. In Dani Ben-Zvi & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 375–395). Springer Netherlands.
https://doi.org/10.1007/1-4020-2278-6_16
- Cobb, P., McClain, K., & Gravemeijer, K. (2003). Learning About Statistical Covariation. *Cognition and Instruction*, 21, 1–78.
- Cobo, S., Castro, E., & 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.
- Confrey, J. (1990). A Review of the Research on Student Conceptions in Mathematics, Science, and Programming. In *Review of Research in Education* (Vol. 16, pp. 3–56). [Sage Publications, Inc., American Educational Research Association]. <https://doi.org/10.2307/1167350>
- Delgado, I., & Ojeda, A. M. (2007). Campo de investigación: Pensamiento relacionado con

- probabilidad, estadística. Nivel educativo: básico. In C. Crespo (Ed.), *Acta Latinoamericana de Matemática Educativa* (Vol. 20, pp. 300–305). Comité Latinoamericano de Matemática Educativa A. C.
- delMas, R. C. (2002). Statistical Literacy, Reasoning, and Learning: A Commentary. *Journal of Statistics Education*, 10(3), 1–11. <https://doi.org/10.1080/10691898.2002.11910679>
- 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>
- Duval, R. (1998). Registro de representación semiótica y funcionamiento cognitivo del pensamiento. In F. Hitt (Ed.), *Investigaciones en Matemática Educativa II* (pp. 173–201). Grupo Editorial Iberoamérica.
- Estepa, A., & Batanero, C. (1996). Judgments of correlation in scatterplots: Students' intuitive strategies and preconceptions. *Hiroshima Journal of Mathematics Education*, 4, 24–41.
- Estepa, A., & 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. <http://fehps.une.edu.au/serj>
- Fischbein, E. (1997). *Intuition in Science and Mathematics: An Educational Approach*. Reidel .
- Fitzallen, N. (2012). Interpreting graphs: Students developing an understanding of covariation. In J. Dindyal, L. P. Cheng, & S. F. Ng (Eds.), *Mathematics Education: Expanding Horizons (Proceedings of the 35th annual conference of the Mathematics Education Research Group of Australasia)* (pp. 290–297). MERGA.
- Francisco, J. M., & Maher, C. A. (2005). Conditions for promoting reasoning in problem solving: Insights from a longitudinal study. *The Journal of Mathematical Behavior*, 24(3), 361–372. <https://doi.org/https://doi.org/10.1016/j.jmathb.2005.09.001>
- Gal, I., & Garfield, J. B. (Eds.). (1997). *The assessment challenge in statistics education* (I. Gal & J. B. Garfield (eds.)). IOS press / ISI.
- García, A., & Yáñez, G. (2018a). Formas de razonamiento covariacional informal alrededor de la recta de mejor ajuste en un ambiente computacional en estudiantes de octavo grado. In I. Álvarez (Ed.), *Tercer Encuentro Colombiano de Educación Estocástica* (pp. 220–229). Asociación Colombiana de Educación Estocástica.
- García, A., & Yáñez, G. (2018b). Predicción informal en un diagrama de puntos. In L. A. Serna & D. Páges (Ed.), *Acta Latinoamericana de Matemática Educativa* (pp. 32–39). Comité Latinoamericano de Matemática Educativa.
- Garfield, J. (2002). The Challenge of Developing Statistical Reasoning. *Journal of Statistics Education*, 10(3). <https://doi.org/10.1080/10691898.2002.11910676>

- Gil, E., & 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.
- Glaser, B. G., & Strauss, A. L. (2008). *The discovery of Grounded Theory: Strategies for qualitative research*. Aldine Transaction.
- Hancock, C., Kaput, J. J., & 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
- Hawkins, A. (1989). Training Teachers to Teach Statistics. *Teaching Statistics*, 11(3), 83. <https://doi.org/10.1111/j.1467-9639.1989.tb00069.x>
- 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., & Ward, S. (2015). Explorando el razonamiento covariacional mediante un ambiente computacional en un curso introductorio de estadística. In C. Fernández, M. Molina, & N. Planas (Eds.), *Investigación en Educación Matemática XIX* (pp. 317–325). SEIEM.
- Jabareen, Y. (2009). Building a Conceptual Framework: Philosophy, Definitions, and Procedure. *International Journal of Qualitative Methods*, 8(4), 49–62. <https://doi.org/10.1177/1609406909000800406>
- Jones, G. A., Langrall, C. W., Mooney, E. S., & Thornton, C. A. (2004). Models of Development in Statistical Reasoning. *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*, 97–117. https://doi.org/10.1007/1-4020-2278-6_5
- Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. *Cognition*, 11(2), 143–157. [https://doi.org/10.1016/0010-0277\(82\)90023-3](https://doi.org/10.1016/0010-0277(82)90023-3)
- Konold, C., & Higgins, T. (2002). Highlights of related research. In & 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/>
- Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, 3(3), 185–225. <https://doi.org/10.1002/acp.2350030302>
- Kuhn, D. (1993). Connecting Scientific and Informal Reasoning. *Merrill-Palmer Quarterly*, 39(1), 74–103. <http://www.jstor.org/stable/23087301>
- Lesh, R., & Zawojewski, J. (2007). Problem solving and modeling. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 763–804). Information Age.
- Mackenzie, N., & Knipe, S. (2006). Research dilemmas: Paradigms, methods and methodology. *Issues in Educational Research*, 16(2), 193–205. <http://www.iier.org.au/iier16/mackenzie.html>
- Manor, H., Ben-Zvi, D., & Aridor, K. (2014). Students' reasoning about uncertainty while making

- informal statistical inferences in an Integrated Pedagogic Approach. In and R. G. K. Makar, B. de Sousa (Ed.), *Sustainability in statistics education (Proceedings of the Ninth International Conference on Teaching Statistics, ICOTS9, July 2014)* (Vol. 9). International Association for Statistical Education and International Statistical Institute. http://icots.info/9/proceedings/pdfs/ICOTS9_8C2_MANOR.pdf
- McDonald, M. A., Mathews, D., & Strobel, K. H. (2000). Understanding sequences: A tale of two objects. *Research in Collegiate Mathematics Education IV*, 8, 77–102.
- Medina, M., Olay, G., & Sánchez, E. (2016). Razonamiento de los estudiantes de bachillerato frente a un problema de regresión. *EUTOPIA*, 9(24), 71–78.
- Medina, M., Sanchez, E., & 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)*. <https://hal.archives-ouvertes.fr/hal-02412811>
- Meraverch, Z., & Kramarsky, B. (2017). *From Verbal Descriptions to Graphic Representations : Stability and Change in Students ' Alternative Conceptions Author (s) : Zemira R . Mevarech and Bracha Kramarsky Source : Educational Studies in Mathematics , Vol . 32 , No . 3 (Mar . , 1997) , pp . 2 . 32(3), 229–263.*
- Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis: An expanded sourcebook, 2nd ed. In *Qualitative data analysis: An expanded sourcebook, 2nd ed.* Sage Publications, Inc.
- Mokros, J., & 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>
- Montgomery, D. C., & Runger, G. C. (1996). *Probabilidad y estadística aplicadas a la ingeniería*. McGraw-Hill.
- Mooney, E. S. (2002). A Framework for Characterizing Middle School Students' Statistical Thinking. *Mathematical Thinking and Learning*, 4(1), 23–63. https://doi.org/10.1207/s15327833mtl0401_2
- Moore, D. S. (1990). Uncertainty. In *On the Shoulders of Giants: New Approaches to Numeracy* (pp. 95–137). National Academy Press.
- Moritz, J. (2004). Reasoning about Covariation. In D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 227–255). Springer Netherlands. https://doi.org/10.1007/1-4020-2278-6_10
- Nagle, C., Casey, S., & 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>
- National Council Of Teachers Of Mathematics. (2000). Principles and Standards for School Mathematics. *School Science and Mathematics*, 47(8), 868–279. <http://standards.nctm.org/>

- Pfannkuch, M. (2011). The role of context in developing informal statistical inferential reasoning: A classroom study. *Mathematical Thinking and Learning*, 13(1–2), 27–46. <https://doi.org/10.1080/10986065.2011.538302>
- Pfannkuch, M., & Wild, C. (2004). Towards an understanding of statistical thinking. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 17–46). Springer Netherlands. https://doi.org/10.1007/1-4020-2278-6_2
- Pollatsek, A., Lima, S., & Well, D. (1981). Concept or Computation: Students' Understanding of the Mean. *Educational Studies in Mathematics*, 12, 191–204.
- Roseth, C. J., Garfield, J. B., & Ben-Zvi, D. (2008). Collaboration in learning and teaching statistics. *Journal of Statistics Education*, 16(1). <https://doi.org/10.1080/10691898.2008.11889557>
- Ross, J. A., & Cousins, J. B. (1993). Patterns of student growth in reasoning about correlational problems. *Journal of Educational Psychology*, 85(1), 49–65. <https://doi.org/10.1037//0022-0663.85.1.49>
- Rumsey, D. J. (2002). Statistical Literacy as a Goal for Introductory Statistics Courses. *Journal of Statistics Education*, 10(3), 1–12. <https://doi.org/10.1080/10691898.2002.11910678>
- Sánchez-Cobo, F., Estepa, A., Batanero, C., Castro, E., & 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.
- Santos-Trigo, M., Moreno-Armella, L., & Camacho-Machín, M. (2016). Problem solving and the use of digital technologies within the Mathematical Working Space framework. *ZDM*, 48(6), 827–842. <https://doi.org/10.1007/s11858-016-0757-0>
- Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (pp. 957–1009). Information Age Publishing, Inc., and NCTM.
- Shaughnessy, M., & Chance, B. (2005). *Statistical Questions from the Classroom*. National Council of Teachers of Mathematics.
- Simon, M. A., & Tzur, R. (2004). Explicating the role of mathematical tasks in conceptual learning: An elaboration of the hypothetical learning trajectory. *Mathematical Thinking and Learning*, 6(2), 91–104. https://doi.org/10.1207/s15327833mtl0602_2
- Sorto, M. A., White, A., & 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., & 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.
- Tal, J. (2001). *Reading between the numbers: Statistical thinking in everyday life*. McGraw-Hill.

- Truran, J. (1995). Some Undergraduates' Understanding of the Meaning of a Correlation Coefficient. In B. Atweh & S. Flavel (Ed.), *Annual conference; 18th, Mathematics Education Research Group of Australasia (MERGA)* (pp. 524–529). Northern Territory University.
- Universidad Autónoma Nacional de México. (2016). *Programas del Colegio de Ciencias y Humanidades de la UNAM*. <https://www.cch.unam.mx/programasestudio>
- Watson, J. M., & Moritz, J. (1997). Student analysis of variables in a media context. In B. Phillips (Ed.), *Proceedings of the 8th International Congress of Mathematics Education (ICME8)* (pp. 129–147). Swinburne PresS.
- Watson, J., & Moritz, J. (2007). Developing aspects of distribution in response to a media-based statistical literacy task. *International Statistical Institute 56th Session*.
- Wavering, M. J. (1989). Logical reasoning necessary to make line graphs. *Journal of Research in Science Teaching*, 26(5), 373–379. <https://doi.org/10.1002/tea.3660260502>
- Willet, J. B., & Singer, J. D. (1992). Providing a statistical model: Teaching applied statistics using real-world. In F. Gordon & S. Gordon (Eds.), *Statistics for the twenty-first century* (pp. 83–98). The Mathematical Association of America.
- Yáñez, G., Rátiva, A. M., & Morgado, C. N. (2016). Un estudio sobre las intuiciones primarias asociadas al razonamiento covariacional. In I. Álvarez & C. Sua (Eds.), *Memorias del II Encuentro Colombiano de Educación Estocástica* (pp. 131–138). Asociación Colombiana de Educación Estocástica.
- Zieffler, A. S., & 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>

Apéndice A: Método de búsqueda bibliográfica

A continuación, se describe el método que se empleó para la determinación de artículos relevantes para la investigación desarrollada. Se describe mediante fases el proceso de búsqueda y selección de la información.

Fase 1. Búsqueda y selección de manuscritos.

Los artículos fueron seleccionados de tres fuentes principales: revistas, libros especializados y memorias de congresos. Con la intención de desarrollar una revisión bibliográfica amplia y actualizada, se llevó cabo la búsqueda en publicaciones en inglés y español, en período temporal entre los años 1988 y 2019. Para la búsqueda se utilizaron los siguientes términos: “covariational reasoning”, “razonamiento covariacional”, “correlation and regression”, “best fit line”, “correlación y regresión”, “línea de mejor ajuste”. Cabe aclarar que en la búsqueda se agregaron las palabras “educación estadística” y “statistics education” para diferenciar el razonamiento covariacional estudiando en el área del cálculo, también para descartar los archivos relacionados con el campo de la estadística pura.

- i. Se inicio con la búsqueda en las siguientes páginas web:

<https://www.zentralblatt-math.org/matheduc/>

<https://www.conricyt.mx/>

<https://scholar.google.com/>

- ii. Seguido se estableció el primer grupo de revistas sobre educación estadística:

- *Statistics Education Research Journal (SERJ)*, *el*
- *Journal of Statistics Education (JSE)* y *la*
- *Teaching Statistics: An International Journal for Teachers.*

- iii. En el segundo grupo se incluyeron algunas revistas del listado publicado por Williams y Leatham (2017) como las más citadas.

- *Educational Studies in Mathematics (ESM)*
- *For the Learning of Mathematics (FLM)*
- *Mathematical Thinking and Learning (MTL)*
- *Journal of Mathematics Teacher Education (JMTE)*

- *ZDM: The International Journal on Mathematics Education*
- *School Science and Mathematics (SSM)*
- iv. En el tercer grupo se incluyen las revistas de investigación en educación, algunas de las cuales se especializan en matemática educativa, y que consideramos son las más influyentes en Latinoamérica.
 - *Educación Matemática*
 - *Revista Latinoamericana de Investigación en Matemática Educativa (RELIME)*, la
 - *Revista Electrónica de Investigación Educativa, el Boletim de Educação Matemática (BOLEMA)*

Los libros especializados en estadística educativa que se revisaron fueron:

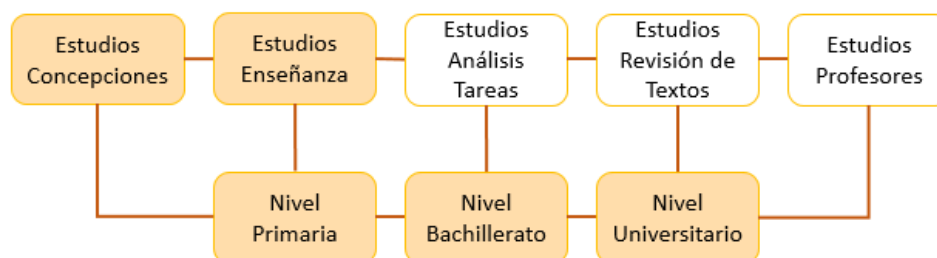
- *Ben-Zvi, Makar y Garfield (2018). International Handbook of Research in Statistics Education.*
- *Ben-Zvi y Makar (2016). The Teaching and Learning of Statistics.*
- *Garfield y Ben-Zvi (2008). Developing Students' Statistical Reasoning.*
- *Lester (2007). Second Handbook of Research on Mathematics Teaching and Learning.*
- *Ben-Zvi y Garfield (2004). The Challenge of Developing Statistical Literacy, Reasoning and Thinking.*
- *Ware y Brewer (1998). Handbook for Teaching Statistics and Research Methods.*
- v. Las memorias de congreso de matemática educativa en general, así como de estadística educativa en particular, con frecuencia son fuente de resultados de investigación que aportan elementos que quizá no se encuentren en artículos o en libros especializados, por lo que también ha parecido relevante su inclusión, por lo tanto, se revisaron las memorias de los siguientes eventos académicos.
 - *International Conference on Teaching Statistics (ICOTS)*,
 - *International Congress on Mathematical Education (ICME)*
 - *Congress of European Research in Mathematics Education (CERME)*
 - *Acta Latinoamericana de Matemática Educativa (ALME)*

Fase 2. Búsqueda de referencias bibliográficas adicionales.

Cabe aclarar que se realizó la búsqueda en las revistas, libros y memorias descritos en la fase 1. Seguido de recopilar los documentos se revisaron las referencias dentro de estas fuentes localizada con el fin de saturar la búsqueda. Además, se fueron descartando los estudios enfocados en el razonamiento covariacional centrado en datos categóricos, así como los estudios psicológicos que han estudiado este razonamiento en las personas porque en este tipo de razonamiento no tiene sentido hablar de correlación y regresión las cuales son las temáticas núcleo de la investigación.

Fase 3. Categorización de los manuscritos.

Una vez seleccionados los artículos enfocados en el razonamiento covariacional que hace referencia al juzgar la relación entre dos variables cuantitativas mediante el análisis de correlación y regresión. Seguido de manera inductiva surgieron las siguientes categorías:



- i. *Los estudios sobre las concepciones o dificultades de los estudiantes.*
- ii. *Los estudios sobre el tipo de tareas que involucran la correlación y regresión.*
- iii. *Los estudios acerca de la temática en los libros de texto.*
- iv. *Los estudios enfocados en las concepciones de los profesores.*
- v. *Los estudios de enseñanza, es decir, aquellos que implementaron diseños de clase.*

Durante la categorización surgió las subcategorías según el nivel académico en el cual se realizó el estudio: nivel de primaria, nivel bachillerato y nivel universitario.

El interés de esta investigación está en los estudios sobre la enseñanza y las concepciones respecto a los temas de correlación y regresión desde el nivel de primaria al universitario, pues para el diseño los instrumentos se desea poner atención a las concepciones, dificultades

y formas de razonar de los estudiantes frente a tareas de covariación estadística. Además, es transcendental conocer los principios y métodos que implementaron para desarrollar el razonamiento covariacional de los estudiantes, específicamente aquellos que han utilizado herramientas tecnológicas. La siguiente tabla muestra la fuente y autores de los estudios hallados sobre las concepciones (C) y enseñanza (E).

Fuente		Autores	
Statistics Education Research Journal	1	Castro-Sotos, Van Hoof, Van den Noortgate y Onghena (2009)	C
	2	Estepa y Sánchez Cobo (2003)	C
	3	Gil y Gibbs (2017)	E
	4	Zieffler y Garfield (2009)	E
Journal of Statistics Education	5	Casey (2015)	C
Teaching Statistics	6	Sorto, White y Lesser (2011).	C
Educational Studies in Mathematics	7	Casey y Nagle (2016)	C
School Science and Mathematics	8	NaglePenn, Casey y Moore-Russo (2017)	C
Cognition and Instruction	9	Gravemeijer, Cobb, y McClain (2003)	E
Research on the role of technology in teaching and learning statistics	10	Batanero, Estepa y Godino (1996)	E
Mathematical Thinking and Learning	11	Dierdorff, Bakker, Eijkelhof y van Maanen (2011)	E
Hiroshima Journal of Mathematics Education	12	Estepa y Batanero (1996)	E
Handbook: The Challenge of Developing Statistical Literacy, Reasoning and Thinking	13	Moritz (2004)	C
International Statistical Review	14	Watson y Moritz (2007)	C

Revista Educación Matemática	15	Inzunza (2016)	E
	16	Inzunza y Ward (2015)	E
Enseñanza de la Ciencias	17	Sánchez Cobo, Estepa y Batanero (2000)	C
Revista del Colegio de Ciencias y Humanidades para el Bachillerato	18	Medina, Olay y Sánchez (2016)	E
ICOTS	19	Casey (2014)	C
ALME	20	García y Yáñez (2018)	C
	21	Delgado y Ojeda Salazar (2007)	C
Encuentro Colombiano de Educación Estocástica	22	García y Yáñez (2018)	E
	23	Yáñez, Rátiva, y Morgado (2016)	C
Conference of the International Group for the Psychology of Mathematics Education	24	Batanero, Godino y Estepa (1998)	E
Mathematics Education Research Group of Australasia (MERGA)	25	Truran (1995)	C
	26	Fitzallen (2012)	E
ICME 8	27	Watson y Moritz (1996)	C
CERME 11	28	Medina, Sánchez y Silvestre (2019)	E

Fase 4. Análisis de los manuscritos seleccionados.

El análisis se realizó bajo tres ejes conductores que guiaron su desarrollo. Primero: las concepciones o dificultades que han sido detectadas. Segundo: el tipo de diseño de intervención de enseñanza y el avance en el aprendizaje de los estudiantes. Tercera: la forma de razonar de los estudiantes al enfrentar tareas que involucran variables cuantitativas. Como resultado se pretende dar respuesta al siguiente interrogante: ¿Cuáles son las descripciones y teorías sobre la enseñanza y aprendizaje que se deriva de las investigaciones?

Apéndice B: Actividad 1

Por regulación del gobierno, las marcas reconocidas de comida rápida como McDonalds y Burger King deben proveer información nutricional sobre sus productos. Por lo tanto, deben medir su aporte calórico, contenido de sodio, azúcar, carbohidratos, proteínas, grasas y otros componentes.

Quizás, oyes o creas que las grasas y las calorías son dañosas 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 1 muestran la cantidad de grasa en gramos y el número de calorías en algunos productos de McDonald's y Burger King.



¿Qué es una caloría?

Una caloría es una unidad de medida, pero esta no mide peso, ni longitud. Una caloría es una unidad de energía que mide cuánta energía proporciona un alimento al cuerpo.

<https://kidshealth.org/es/kids/calorie-esp.html>

Tabla 6. 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
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

<https://www.mcdonalds.com>
<https://www.bk.com/menu/search-by-nutrition>

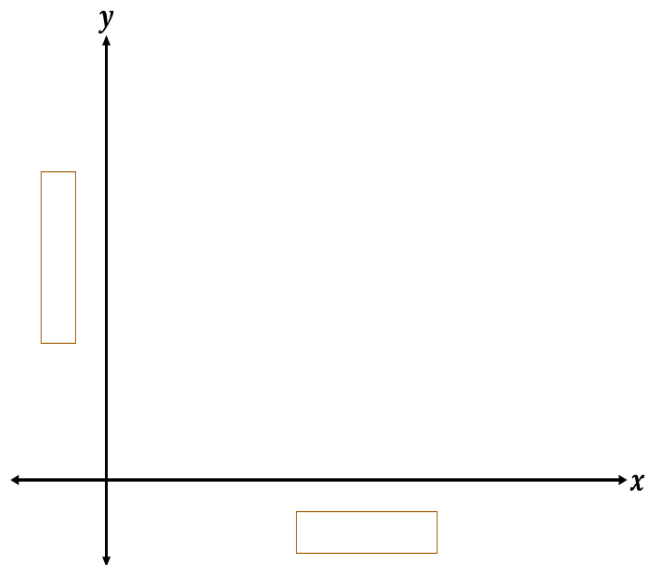
A lápiz y papel

1.1. Observa los datos de la Tabla 1. ¿Cuál crees que es la relación entre los gramos de grasa y las calorías? Explica tu respuesta.

1.2. Realiza una gráfica de los datos de la tabla 1 en el plato Cartesiano dado en la figura 1. Etiqueta los ejes con la variable correspondiente.

Producto	Grasa(g)	Calorías (Kcal)
1	27	540
2	12	300
3	26	510
4	18	390
5	45	780
6	13	300
7	10	240
8	16	320
9	51	790
10	58	900
11	60	920
12	41	610

Figura 7. Plano Cartesiano para graficar los datos de la Tabla 1



1.3.¿Cómo harías para estimar el número de calorías de un producto que contiene 31 gramos de grasa? Explica tu respuesta.

1.4.¿Cómo harías para predecir el número de calorías de un producto que contiene 63 gramos de grasa? Explica tu respuesta.

Variable de Respuesta y Variable Explicativa

Una variable explicativa puede ayudar a predecir o explicar los cambios en la variable de respuesta. Los datos se pueden representar mediante los pares ordenados (x, y) donde x es la variable independiente (o explicativa) e y es la variable dependiente (o respuesta).

1.5.Para la actividad. Discute con tu compañero y responde:

¿Cuál es la variable de respuesta? _____ ¿Por qué?

¿Cuál es la variable explicativa? _____ ¿Por qué?

Utilizando Tecnología Digital

1.6. Abrir la plataforma digital *CODAP* y crear el diagrama de dispersión de los datos bivariados (grasa, calorías) presentados en la tabla 1. Sigue las indicaciones del profesor.

DEFINICIÓN. Diagrama de Dispersión
Un diagrama de dispersión muestra la relación entre dos variables cuantitativas medidas a los mismos individuos. Los valores de una variable aparecen en el eje horizontal, y los valores de la otra variable aparecen en el eje vertical. Cada individuo en el conjunto de datos aparece como un punto en el gráfico.

1.7. ¿Qué características se podrían tener en cuenta al interpretar un diagrama de dispersión?

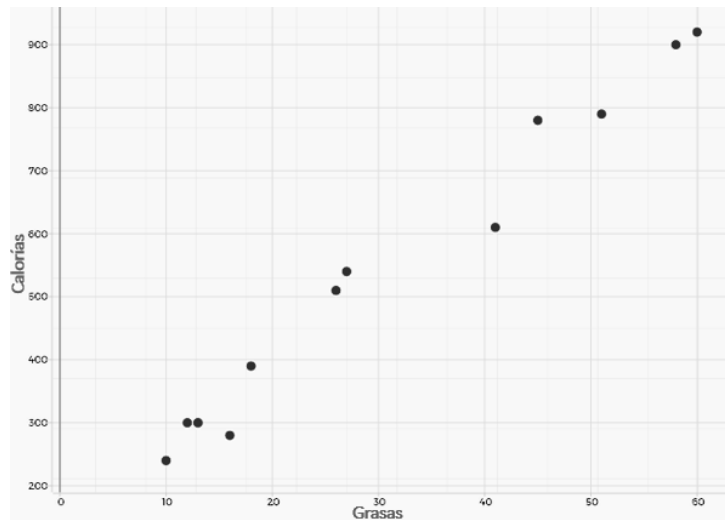
1.8. Teniendo en cuenta las características que mencionaste en 1.7. ¿Qué puedes decir de la relación entre los gramos de grasa y el número de calorías?

1.9. Socializa con tus compañeros y el profesor tus respuestas a las anteriores preguntas.

A lápiz y papel

1.10. La figura 2 muestra el diagrama de dispersión de los datos bivariados (Grasa, Calorías). Dibuja la línea que mejor se ajusta a los datos. Explica el criterio que utilizaste para trazarla.

Figura 8. Diagrama de Dispersión de la Grasa y Calorías



Explica el criterio que utilizaste para trazarla.

1.11. ¿Cómo utilizarías la línea que dibujaste para estimar el número de calorías de un producto que contiene 31 gramos de grasa?

1.12. ¿Cómo utilizarías la línea que dibujaste para predecir el número de calorías de un producto que contiene 63 gramos de grasa?

Utilizando Tecnología Digital

- 1.13.** Sigue las indicaciones del profesor y en *CODAP* determina la línea de regresión. Compara la línea que dibujaste con la línea que muestra el software o la plataforma. ¿Cuál crees que es el criterio que utiliza el programa para determinar la línea de regresión?

	DEFINICIÓN. Línea de regresión
	Una línea de regresión es una línea que describe cómo cambia una variable de respuesta y a medida que cambia una variable explicativa x . Las líneas de regresión se expresan de la forma $\hat{y} = mx + b$, donde \hat{y} (pronunciado "y gorro") es el valor predicho de y para un valor dado de x .

Apéndice C: Actividad 2

El aporte calórico que contiene un producto depende de muchos nutrientes entre ellos, la grasa, los carbohidratos, el azúcar, la proteína, el sodio y otros. La tabla 2 muestra las mediciones de algunos de los nutrientes que influyen en el aporte calórico: la grasa, carbohidratos, azúcar y proteína en gramos que contiene algunos productos de McDonald's y Burger King junto con la cantidad de calorías para cada uno.

Tabla 2. Mediciones de grasa, carbohidratos y azúcar en los productos de McDonalds y Burger King

PRODUCTO	Grasa (g)	Carbohidratos (g)	Azúcar (g)	Calorías Kcal
Big Mac	27	46	9	540
Cheeseburger	12	33	7	300
Quarter Pounder with Cheese	26	42	10	510
McDouble	18	33	7	390
Double Quarter Pounder with Cheese	45	43	10	780
Bacon Ranch Grilled Chicken Salad	13	8	3	300
Hamburger	10	26	6	240
Cheeseburger	16	27	7	320
Bacon y Cheese WHOPPER	51	50	11	790
DOUBLE WHOPPER	58	49	11	900
Double Pretzel Bacon King	60	55	2	920
Chicken Club Salad-Grilled	41	18	6	610

A lápiz y papel

Con la información en la tabla 2 contesta:

2.1.¿Qué componentes nutricionales (grasa, carbohidratos o azúcar) influyen en el aporte calórico? Los que influyen ¿Lo hacen con la misma fuerza?

2.2. ¿Qué componente influye con mayor fuerza? ¿Por qué?

--

Utilizando Tecnología Digital

Crea un nuevo documento en *CODAP* y crea los diagramas de dispersión para los siguientes datos bivariados:

- a) (Grasa, Calorías)
- b) (Carbohidratos, Calorías)
- c) (Azúcar, Calorías)

Sigue las indicaciones del profesor para determinar las siguientes correlaciones en *CODAP* y anótalas.

2.3. Correlación entre (Grasa, Calorías) = _____

2.4. Correlación entre (Carbohidratos, Calorías) = _____

2.5. Correlación entre (Azúcar, Calorías) = _____

			DEFINICIÓN. Coeficiente de Correlación
<p>La correlación r es una medida de la fuerza y la dirección de una relación lineal entre dos variables cuantitativas.</p> <ul style="list-style-type: none"> • La correlación r es un valor entre -1 y 1. 			

2.6. ¿Existe alguna relación entre el valor numérico de correlación y el comportamiento de los datos representados en el diagrama de dispersión? Observa el diagrama de dispersión y la correlación correspondiente para cada conjunto de datos bivariados. Describe la relación entre las variables teniendo en cuenta su gráfico y su correlación.

(Grasa, Calorías)	

(Carbohidratos, Calorías)	
(Azúcar, Calorías)	

2.7.¿Qué observas en los datos representados en el diagrama de dispersión cuando la correlación es un valor que se acerca a 1?

2.8.¿Qué observas en los datos representados en el diagrama de dispersión cuando la correlación es un valor lejano de 1?

Teniendo en cuenta lo realizado anteriormente, vuelve a contestar las siguientes preguntas:

2.9.¿Existe algún componente nutricional (grasa, carbohidratos o azúcar) presentado en la tabla 2 que influya fuertemente en el aporte calórico (calorías)? ¿por qué?

2.10. ¿Existe algún componente nutricional (grasa, carbohidratos o azúcar) presentado en la tabla 2 que influya débilmente en el aporte calórico? ¿por qué?

2.11. Cierra los diagramas de dispersión los datos bivariados (grasa, calorías) y (carbohidratos, calorías), deja solo el diagrama de dispersión de (azúcar, calorías) y en la tabla de datos deja la correlación de (azúcar, calorías) borra las demás.

			DEFINICIÓN. Valor atípico o valor extremo
Puntos individuales que quedan fuera del patrón general de la relación.			

a. ¿Cuáles son los pares de datos (x, y) que se son valores extremos? ¿Por qué los consideras valores extremos?

b. Mueve los valores extremos y observa lo que ocurre con el valor numérico de la correlación ¿Cómo influye los valores extremos en la correlación (azúcar, calorías)?

Apéndice D: Actividad 3

¿Estar inquieto te mantiene delgado? Un estudio que 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 de las manos y los pies, por nerviosismo o impaciencia y otras "actividades no relacionadas con el ejercicio" 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? Por lo tanto, 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 que no sea el ejercicio: inquietud, vida diaria y cosas por el estilo.

Nota: ANE es abreviatura: Actividades no relacionadas con el ejercicio.

Tabla 3. Medidas de cambio en ANE y aumento de grasa en los 16 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

Datos tomados de The Practice of Statistics by Satrnes, Yates and Moore (2010)

A lápiz y papel

Discute con tu compañero y responde:

¿Cuál es la variable de respuesta? _____ ¿Por qué?



CALCULADORA DE CALORÍAS

¿No sabes cuántas calorías necesitas diariamente? Entonces utiliza nuestra calculadora de calorías que aparece en la siguiente página web:
<https://www.foodspring.es/calculador-de-macros>

¿Cuál es la variable explicativa? _____ ¿Por qué?

¿Cuáles de las siguientes afirmaciones es correcta? Explica ¿por qué?


- Las personas con mayores incrementos en NEA tienden ganar más peso.
- Las personas con mayores incrementos en NEA tienden a ganar menos grasa.

Utilizando Tecnología Digital

3.1. Teniendo en cuenta lo que has realizado respecto al análisis de los diagramas de dispersión y el coeficiente de correlación. Describe la relación entre el cambio ANE y el aumento de grasa.



3.2. Abrir el archivo de GeoGebra ANE_Grasas_Recta.ggb. En el archivo de GeoGebra aparece el diagrama de dispersión con los siguientes elementos:

- Los puntos azules son los datos de la tabla 3.
- La Recta amarilla es movable por los símbolos en rojo .
- Los segmentos punteados de color rojo y el valor numérico en rojo corresponden a la distancia de cada dato de la tabla 3 a la recta amarilla movable.
- La palabra residuos muestra la SUMA DE TODAS las distancias.

Mueve la Recta amarilla para que corresponda a la recta de regresión

RECUERDA: DEFINICIÓN. Recta de regresión
Una Recta de regresión es una Recta que describe cómo cambia una variable de respuesta y a medida que cambia una variable explicativa x . Las Rectas de regresión se expresan de la forma $\hat{y} = mx + b$, donde \hat{y} (pronunciado "y gorro") es el valor predicho de y para un valor dado de x .

3.3. ¿Qué ocurre con el valor de cada residual si alejas la línea amarilla de la nube de puntos azules?



3.4. ¿Qué ocurre con el valor de los residuos?, ¿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?

3.5. ¿Dónde consideras que debes ubicar la recta amarilla para que corresponda a la recta de regresión? Explica tu criterio.

DEFINICIÓN. Residual

Un residual es la diferencia entre un valor real de y y el valor de y predicho por la Recta de regresión. Es decir,

$$\text{Residual} = \text{El valor real } y - \hat{y} \text{ el valor predicho} = y - \hat{y}$$

3.6. ¿Cuál es el valor mínimo de los residuos? _____

3.7. ¿Qué relación tiene el valor numérico de los residuos con recta?

3.8. Sigue las indicaciones del profesor y determina la recta de regresión (deberá aparecer de color rojo). ¿La manera como ubicaste la Recta amarilla es igual a la Recta de regresión arrojada por GeoGebra? ¿Sí? ¿No? ¿En qué se diferencian?

3.9. ¿Cuál es la expresión algebraica de la recta de regresión? Describe el criterio que consideras utiliza el programa arrojar la recta de regresión.

3.10. Queremos estimar el aumento de grasa para un individuo cuya ANE aumenta en 400 calorías cuando come en exceso. ¿Qué harías? Explica tu procedimiento

3.11. ¿Podemos predecir el aumento de grasa para alguien cuya ANE aumenta en 1500 calorías cuando come en exceso? ¿Sí? ¿No? ¿Por qué?

Apéndice E: Cuestionario diagnóstico

Parte I. Diagnóstico

Nota: Los Ítems del 1 al 8 son tomados de Zieffler (2006)

Ítem 1. El costo C , de imprimir tarjetas de felicitación consiste en un cargo fijo de 100 pesos y un cargo de 6 pesos por cada tarjeta impresa. ¿Cuál de estas ecuaciones se puede utilizar para determinar el costo de impresión de n tarjetas?

- a. $C = (100 + 6n)$ pesos
- b. $C = (106 + n)$ pesos
- c. $C = (6 + 100n)$ pesos
- d. $C = (106n)$ pesos
- e. $C = (600n)$ pesos

Contenido. Establecer algebraicamente la función lineal que representa el enunciado. Es decir, pasar del enunciado verbal a establecer una expresión algebraica.

Objetivos. Determinar si el estudiante logra establecer la forma algebraica del enunciado para la cual debe identificar el costo fijo y el costo que varía.

Ítem 2. La tabla muestra una relación entre x y y .

x	2	3	4	5
y	7	10	13	16

Escribe una ecuación que exprese está relación. Respuesta $y = 3x + 1$

Contenido. Establecer la función lineal que representa la relación a partir de datos tabulares.

Objetivo. Determinar si el estudiante logra deducir el patrón para representar algebraicamente la variación de y.

Ítem 3. Si $x = 3$, ¿Cuál es el valor de $\frac{5x+3}{4x-3}$?

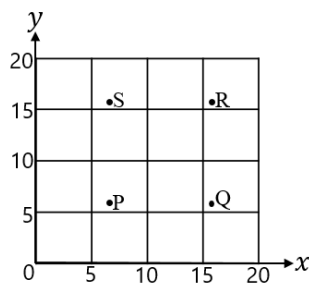
Respuesta:

$$\begin{aligned} &= \frac{5(3) + 3}{4(3) - 3} \\ &= \frac{18}{9} \\ &= 2 \end{aligned}$$

Contenido. Dado un valor sustituir en la expresión algebraica y hallar su valor.

Objetivo. Determinar si el estudiante sustituye correctamente un valor en una expresión algebraica y determina su valor numérico correctamente.

Ítem 4. ¿Cuál punto de la gráfica podría tener las coordenadas (7,16)?



- a. Punto P
- b. Punto Q
- c. Punto R
- d. Punto S**

Contenido. Identificar en una pareja ordenada la abscisa y la ordenada ubicando el punto en el plano cartesiano.

Objetivo. Determinar si el estudiante sabe ubicar un punto en el plano cartesiano.

Ítem 5. La tabla muestra algunos valores de x y y . ¿Dónde x es proporcional a y ?

x	4	8	Q
y	9	P	45

¿Cuál es el valor para P y cuál para Q?

- a. P=40 y Q=13
- b. P=18 y Q=17
- c. P=20 y Q=18
- d. P=40 y Q=18
- e. **P=18 y Q=20**

Respuesta: x es proporcional a y en todos los valores de la tabla.

$$x = ky$$

$$4 = k(9),$$

$$k = 4/9$$

$$\text{Luego, } y = \frac{9}{4}(x)$$

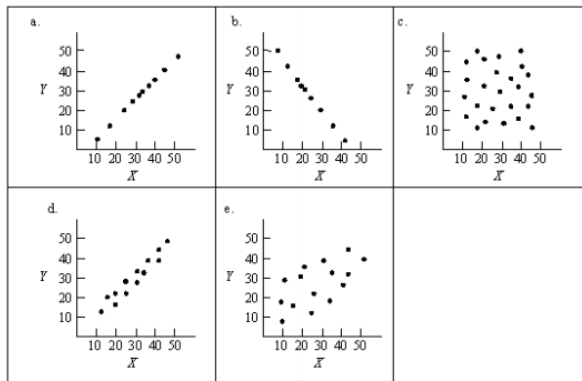
x	4	8	20
y	9	18	45

Contenido. Variables proporcionales. Determinar la función lineal que representa la relación de proporcionalidad y encontrar los valores.

Objetivo. Conocer si el estudiante sabe determinar cuándo dos variables son proporcionales y utilizar este concepto para establecer la expresión algebraica que representa la relación con el fin de hallar los valores.

Parte II. Evaluación

Ítem 6. Para los ítems i. y ii.



Contenido. Correlación gráficamente

- *A medida que los puntos están más alineados, la correlación entre las variables es más fuerte.*

Objetivos.

-*Con el inciso i. Se desea determinar si el estudiante sabe elegir el diagrama de dispersión que podría representar una correlación de 0.60, dado que 0.60 no es un valor tan cercano a 1, la nube de puntos no muestra una relación perfecta, pero el valor esta es por encima de 0.5. Por lo tanto, descartar las opciones donde las gráficas muestran una relación positiva y negativa casi perfecta y las que muestran los datos muy dispersos.*

- *Con el inciso ii. Conocer si el estudiante sabe reconocer gráficamente una relación fuerte entre dos variables.*

Ítem 7. La densidad ósea se mide típicamente como una puntuación estandarizada con una media de 0 y una desviación estándar de 1. Las puntuaciones más bajas corresponden a una densidad ósea más baja. ¿Cuál de los siguientes gráficos muestra que a medida que las mujeres envejecen tienden a tener una densidad ósea más baja?

eleccione el ii. La relación más fuerte entre las variables X y Y .

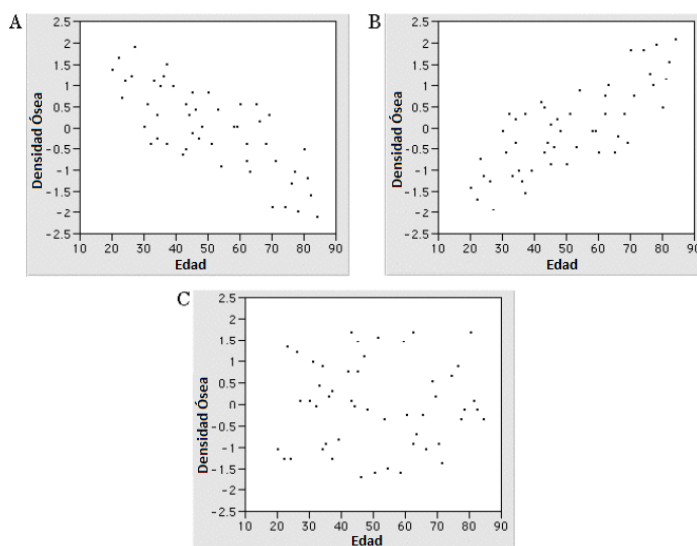
¿cuál de los siguientes muestra:

Una correlación que muestra cerca

de 0.60

- a. a
- b. b
- c. c
- d. d
- e. e

- a. a
- b. b
- c. a y b
- d. a y d
- e. a, b y d.



Respuesta (A)

Contenido.

-Identificar las variables en estudio

-Mediante la nube de puntos o diagrama de dispersión se puede identificar si hay relación es positiva o negativa.

Objetivos.

- Se pretende que el alumno reconozca la relación establecida en forma textual e identifique la figura correspondiente.

Ítem 8 (tomado de Sorto (2011)). Un artículo reciente midió la satisfacción laboral de los sujetos con una encuesta de 14 preguntas. Los datos de la tabla representan las puntuaciones de satisfacción laboral (y), y los salarios en miles de dólares(x), para una muestra de ocho individuos similares. La Figura 1 es una gráfica de dispersión de los datos.

x Miles de dólares	31	33	22	24	35	29	23	37
y Nivel de Satisfacción	17	20	13	15	18	17	12	21

Tabla 1. Nivel de Satisfacción Laboral con salarios de trabajo

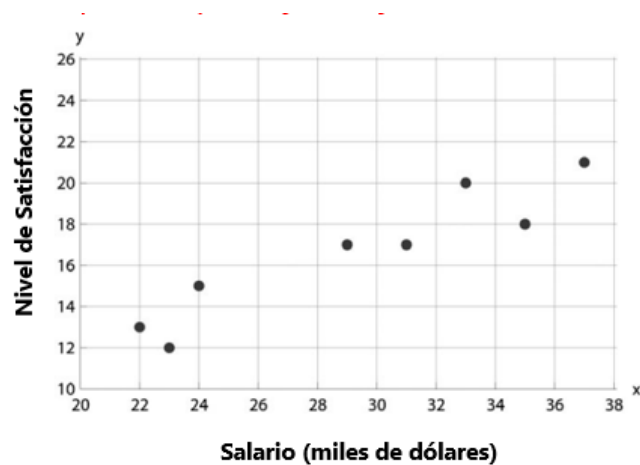


Figure 1. gráfica de dispersión de los datos

- i. Dibuje una línea en el gráfico de dispersión que mejor se ajuste a los datos. ¿Qué criterios utilizó para determinar la línea? Describa lo que estaba pensando cuando escogió una línea que se ajusta a los datos.
- ii. Prediga el nivel de satisfacción de un individuo que gana 39 mil dólares.

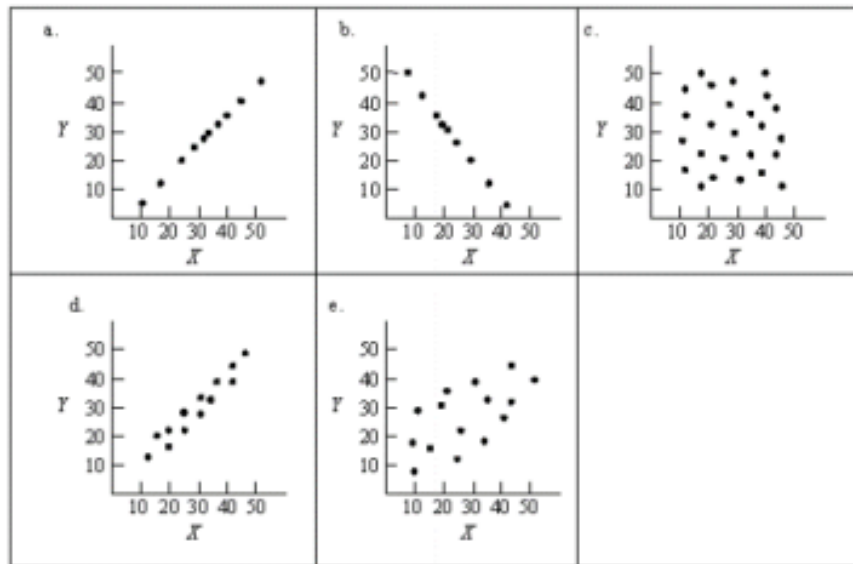
Contenido. Recta de mejor ajuste. Regresión por mínimos cuadrados.

Objetivos. Conocer las concepciones acerca de que métodos utilizan los estudiantes cuando se les pregunta por dibujar una línea que mejor se ajuste a los datos

Apéndice F: Cuestionario evaluativo

Lee atentamente cada enunciado, elabora y selecciona la respuesta que consideres correcta.

Ítem 1. Para los ítems i. y ii.



Seleccione el diagrama de dispersión que muestre:

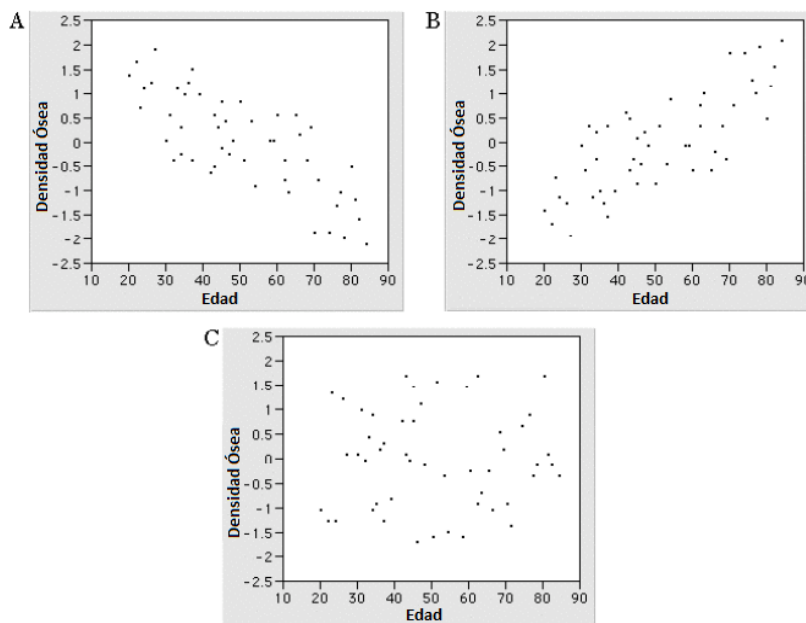
i. Una correlación que muestre cerca de 0.60

- a. a
- b. b
- c. c
- d. d
- e. e

iii. La relación más fuerte entre las variables X y Y .

- f. a
- g. b
- h. a y b
- i. a y d
- j. a, b y d.

Ítem 2. La densidad ósea se mide típicamente como una puntuación estandarizada con una media de 0 y una desviación estándar de 1. Las puntuaciones más bajas corresponden a una densidad ósea más baja. ¿Cuál de los siguientes gráficos muestra que a medida que las mujeres envejecen tienden a tener una densidad ósea más baja? _____



Ítem 3. A continuación, se presentan tres afirmaciones referidas a las conclusiones de un estudio acerca de las tasas de nacimiento, suicidio, crecimiento económico y productividad, junto con tres gráficos de dispersión.

Afirmación 1: En ciertos países con un desarrollo tecnológico alto, se tienen bajas tasas de nacimiento (TN) asociadas con altas tasas de suicidio (TS).

Afirmación 2: Algunos economistas afirman que, independientemente de los países que se estudian, a altas tasas de crecimiento (TC) se asocian altas tasas de productividad (TP).

Afirmación 3: Tanto como demógrafos afirman que las tasas de suicidio (TS) no parecen estar correlacionadas con las tasas de productividad (TP).

Desafortunadamente, en los gráficos no se colocaron los rótulos de referencia de los ejes (TS, TP, TC, TN). Asocia a cada gráfica una afirmación, completa con el nombre de los ejes.

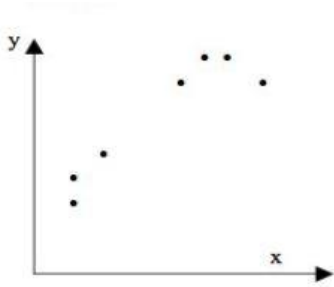


Gráfico 1

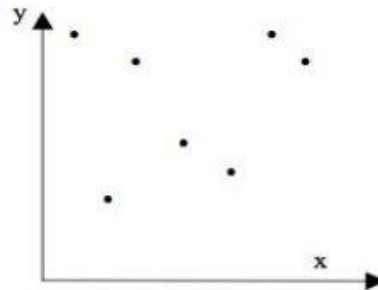


Gráfico 2

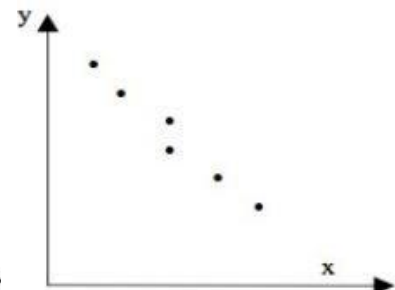


Gráfico 3

Ítem 4. Los corredores dan más pasos por segundo a medida que aumentan la velocidad. He aquí el promedio de pasos por segundo de un grupo de corredoras de élite a distintas velocidades. La velocidad se expresa en metro por segundo.

Velocidad (m/s)	4,83	5,14	5,33	5,67	6,08	6,42	6,74
Pasos por Segundo	3,05	3,12	3,17	3,25	3,36	3,46	3,55

- i) Escoja la opción que mejor describe los datos anteriores.
- La corredora con mayor velocidad dará más pasos por segundo.
 - Si las corredoras tienen velocidades altas darán menos pasos por segundo.
 - Si las corredoras tienen velocidades bajas entonces darán menos pasos por segundo.
 - No se puede establecer relación entre las velocidades de las corredoras y el número de pasos por segundo.

Escribe el procedimiento o análisis que te permitió llegar a la opción seleccionada.

- ii) Prediga el número de pasos por segundo si la velocidad de una corredora alcanza 6 m/s.

Ítem 5. Un artículo reciente midió la satisfacción laboral de los sujetos con una encuesta de 14 preguntas. Los datos de la tabla representan las puntuaciones de satisfacción laboral (y), y los salarios en miles de dólares(x), para una muestra de ocho individuos similares. La Figura 1 es una gráfica de dispersión de los datos.

x Miles de dólares	31	33	22	24	35	29	23	37
y Nivel de Satisfacción	17	20	13	15	18	17	12	21

Tabla 1. Nivel de Satisfacción Laboral con salarios de trabajo

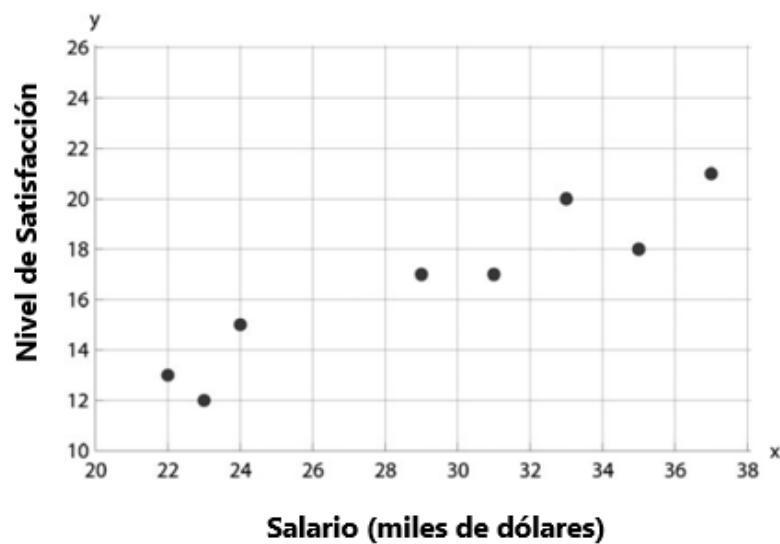


Figure 2. gráfica de dispersión de los datos

- i. Dibuje la línea que mejor se ajuste a los datos en el diagrama de dispersión. ¿Qué criterios utilizó para determinar la línea? Describa lo que estaba pensando cuando escogió una línea que se ajusta a los datos.
- ii. Prediga el nivel de satisfacción de un individuo que gana 39 mil dólares.

Apéndice G: Codificación inicial de las actividades

Codificación inicial de la actividad 1.

A continuación, se muestran para cada pregunta, los códigos que se generaron junto con las palabras clave o frases para codificar y ejemplos de respuestas

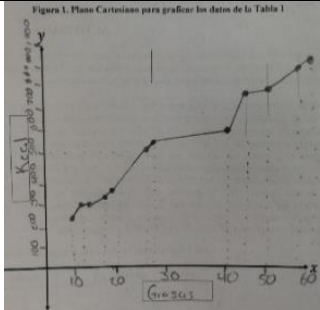
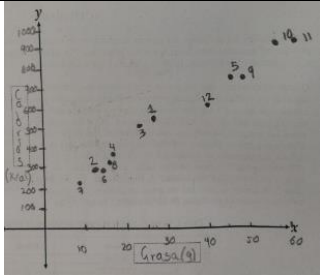
- (i) Códigos de la situación a lápiz y papel: juzgar la relación entre las variables a partir del contexto.

Código	Palabras claves o frases para codificar	Ejemplo de respuesta
Tendencia global positiva	Una cantidad aumenta la	<i>“Su relación es que entre más grasa hay también más calorías. Como se ve en la tabla mientras la cantidad de grasa aumenta la cantidad de calorías también aumenta”.</i>
	otra también aumenta La grasa hace que contenga más calorías	<i>“Se puede observar una clara relación entre grasas y calorías dando un aumento entre ambas (entre más grasa más calorías)”</i> <i>“Entre más grasa contenga un producto más calorías aporta al cuerpo del consumidor sin importar la marca de comida que sea”</i>
Operación aritmética: vista individual	Mientras la grasa aumenta menos es el aporte calórico Realizar una operación aritmética con un par de valores	<i>“Dividimos las calorías entre la grasa para sacar el equivalente de un gramo de grasa y las calorías, lo cual nos da que a menos grasa en MacDonalds dan más calorías al igual en Burger King”.</i> <i>“La relación es que entre menor cantidad de grasa mayor cantidad de calorías. Dividimos los gramos de grasa con su respectivo calorías y nos dimos cuenta de que entre menos grasa más calorías”.</i>

“Con cinco productos calculamos la pendiente de las calorías respecto a las grasas y llegamos a la conclusión que por cada gramo de grasa hay 20.7 calorías. $m=7x=540/27$; $m=y/x = 300/12 = 20$ ”.

Los anteriores códigos determinaron que la relación entre las variables puede ser:

- *Tendencia global positiva*: Hay una relación positiva, si los valores de una variable tienden a aumentar conforme los valores de la otra variable también aumentan.
 - *Operación aritmética (vista individual)*: Realizar operaciones aritméticas con algunos datos.
- (ii) Códigos de la situación a lápiz y papel: realizar el diagrama de dispersión

Código	Palabras claves o frases para codificar	Ejemplo de respuesta
Sesgo de atención: idea funcional	Trazado de semi-rectas	
Nube de puntos	Puntos en el plano	

Los anteriores códigos determinaron que al establecer el diagrama de dispersión puede presentarse dos estrategias:

- *Sesgo de atención-idea funcional*: Se ubican los pares ordenados del conjunto de datos bivariados en el plano cartesiano y se traza la unión por semi-rectas entre ellos.

- *Nube de puntos*: Se ubican los pares ordenados del conjunto de datos bivariados en el plano cartesiano.
- (iii) Códigos de la situación utilizando tecnología digital (CODAP): Describir la nube de puntos que muestra el diagrama de dispersión.

Código	Palabras claves o frases para codificar	Ejemplo de respuesta
Linealidad positiva con error	“casi alineados”, “línea imperfecta”, “desalineados”	“Siempre va a ser dos datos y el segundo va a depender del primero, están casi alineados y van de forma ascendente dirección positiva, tiene forma a línea”.
Orden ascendente con error	“todos aumentan”, “orden numérico ascendente”	“Los puntos al estar desalineados no forman una línea recta. Algunos puntos están encima de otros. Todos aumentan de manera consecutiva”.
Respuesta fuera del contexto	“determinar las variables” “forma de X y Y”	“Tomar en cuenta de escribir correctamente los datos, identificando adecuadamente cuál es la variable dependiente e independiente”.

Los anteriores códigos determinaron una tendencia para los datos bivariados representados en el plano cartesiano, esta tendencia se describe como:

Linealidad positiva con error: La tendencia de la nube de puntos es positiva y su forma es casi lineal.

Orden ascendente con error: Describen el orden ascendente de los puntos en el diagrama sin aceptar su comportamiento lineal.

Respuesta fuera del contexto: No establecen ninguna descripción

- (iv) Códigos de la situación utilizando tecnología digital (CODAP): Juzgar la relación entre las variables a partir del diagrama de dispersión.

Código	Palabras claves o frases para codificar	Ejemplo de respuesta
--------	---	----------------------

Dependencia positiva	Las calorías dependen de la grasa	<i>Las calorías son dependientes de la cantidad de grasa, pues si aumentan el otro también lo hace.</i>
Dependencia funcional	La grasa aumenta entonces las calorías también	<i>Entre más cantidad de grasa, el número de calorías va aumentado, por lo que podemos decir que son directamente proporcionales.</i>
Dependencia con error	Las calorías dependen de la grasa, pero varía	<i>Que conforme el alimento tenga más grasa las calorías van a aumentar, pero no aumentan secuencialmente, sino que las cantidades varían.</i>

Los anteriores códigos determinaron que la relación entre las variables puede ser:

Dependencia positiva: La relación entre las variables es dependiente porque ambas aumentan.

Dependencia funcional: La relación entre las variables es proporcional, una variable aumenta la otra también aumenta.

Dependencia con error: Es una relación donde una variable depende de la otra y ambas aumentan de manera no proporcional.

(v) Códigos de la situación a lápiz y papel: Estimar y predecir a lápiz y papel

Código	Palabras claves o frases para codificar	Ejemplo de respuesta
Aritmético siguiendo un patrón	Tomar dos valores cercanos o con un patrón	<i>De 58 g son 900 calorías y de 60 g son 920 calorías así que por cada gramo hay que subirle 10 calorías</i>
Aritmético sin patrón definido	Realizar una operación aritmética con cualquier par de valores	<i>Multiplicamos 63 por 20 que es aproximado a su relación grasa caloría. Esta multiplicación nos da 1260 calorías por 63 gramos.</i>

Los anteriores códigos determinaron que las estrategias para estimar y predecir a lápiz y papel son:

Aritmético siguiendo un patrón: Buscar un patrón en la tabla de valores entre un par de datos para establecer un procedimiento aritmético.

Aritmético sin patrón definido: Sin patrón definido se efectúan algunas operaciones básicas (suma, resta, multiplicación y división).

- (vi) Códigos de la situación utilizando tecnología digital (CODAP): Estimar y predecir.

Código	Palabras claves o frases para codificar	Ejemplo de respuesta
Recta representa toda la información	Utilizar la línea que arroja CODAP	<i>Ubicarla en la gráfica y así de alguna forma el punto estará cerca de la línea recta, ya que la línea representa toda la información. Colocándonos en el número 63 del eje x ascendimos por el eje "y" hasta colocarnos en la línea trazada.</i>
Proceso aritmético	Realizar un planteamiento aritmético.	<i>La utilizaría para aproximarme al número de calorías que tiene, si en 27 grasas tiene más de 500 calorías y en 40 grasa tiene menos de 600 calorías entonces para 31 tendrá más de 500 y menos de 600.</i>

Los anteriores códigos determinaron que las estrategias para estimar y predecir utilizando CODAP:

Recta representa toda la información: Emplear la línea arrojada por el software para estimar y predecir el valor lo cual implica asumir la línea como representante de todos los datos.

Proceso aritmético: Utilizar un par de valores para realizar algún procedimiento matemático.

- (vii) Códigos de la situación utilizando tecnología digital (CODAP): Establecer cuál es el criterio que se cree utiliza el software para trazar la línea de mejor ajuste.

Código	Palabras claves o frases para codificar	Ejemplo de respuesta
Concepción funcional	Pasa por todos los puntos o algunos de ellos	<i>Una línea que pase por todos los puntos para que abarque toda la información.</i>
Cercanía	Cerca a todos los puntos	<i>La línea que pase más cerca de los puntos</i>
Concepción local	Pasa por el primer y último valor	<i>Que toman el más bajo valor de grasas y lo unen con el valor más alto de grasa.</i>
Operacional o algebraico	Se halla la pendiente, se suma, se resta o multiplica.	<i>El criterio que uso la máquina para sacar los valores de la función $y=mx+b$ en donde se multiplica las calorías por la grasa y se suma el producto.</i>

Los anteriores códigos determinaron que los criterios para trazar la línea de mejor ajuste son:

Concepción funcional: La línea es aquella que pasa por todos los puntos.

Cercanía: La línea es aquella que está cerca de todos los puntos.

Concepción local: La línea es aquella que contiene un par de puntos

Operacional o algebraico: La línea se determina mediante una expresión algebraica.

La tabla resume la frecuencia obtenida para cada código respecto a las preguntas que involucraban juzgar la relación entre las variables a partir de la situación problema y del diagrama de dispersión a lápiz y papel, posteriormente utilizando tecnología. Es importante destacar que para esta actividad 24 estudiantes participaron, es decir, 12 parejas.

Tabla 1. Frecuencia para cada código

Pregunta	Código	Frecuencia (Parejas)
	Tendencia global positiva	8

Juzgar la relación entre las variables a partir del contexto a lápiz y papel.	Operación aritmética: vista individual	4
Realizar el diagrama de dispersión a lápiz y papel	Sesgo de atención: idea funcional	5
	Nube de puntos	7
Describir la nube de puntos que muestra el diagrama de dispersión en CODAP.	Linealidad positiva con error	5
	Orden ascendente con error	5
	Respuesta fuera del contexto	2
Juzgar la relación entre las variables a partir del diagrama de dispersión de CODAP.	Dependencia positiva	7
	Dependencia funcional	2
	Dependencia con error	3

La anterior tabla muestra que los códigos en las preguntas a lápiz y papel que presentaron mayor frecuencia fueron tendencia global positiva (8) y graficar la nube de puntos (7), y utilizando CODAP para graficar la nube de puntos y juzgar la relación los códigos con mayor frecuencia fueron linealidad positiva con error (5), orden ascendente con error (5) y dependencia positiva (7).

La siguiente tabla resume la frecuencia obtenida para cada código respecto a las preguntas que involucraban estimar y predecir a lápiz y papel, posteriormente utilizando tecnología.

Tabla 2. Frecuencia para cada código

Pregunta	Código	Frecuencia (Parejas)
Estimar a partir de la situación problema a lápiz y papel.	Aritmético siguiendo un patrón	6
	Aritmético sin patrón definido	6
Predecir a partir de la situación problema a lápiz y papel.	Aritmético siguiendo un patrón	4
	Aritmético sin patrón definido	8
Como utilizar la línea arrojada por CODAP para estimar y predecir.	Recta representa toda la información	10
	Proceso aritmético	2

El código de mayor frecuencia en las respuestas al estimar y predecir a lápiz y papel es aritmético sin patrón definido (8), y con uso de tecnología el código “recta” representa toda la información (10).

La siguiente tabla resume la frecuencia obtenida para cada código respecto a la pregunta que involucraba establecer la recta de mejor ajuste utilizando tecnología.

Tabla 3. Frecuencia para cada código

Pregunta	Código	Frecuencia (Parejas)
Cuál es el criterio que se cree utiliza CODAP para trazar la línea de mejor ajuste	Concepción funcional	1
	Cercanía	1
	Concepción local	3
	Operacional o Algebraico	5
	No describe ninguno	2

El código de mayor frecuencia es el operacional o algebraico (5) al establecer la línea de mejor ajuste.

Codificación inicial de la actividad 2

- (i) Códigos de la situación utilizando tecnología digital (CODAP): Establecer la fuerza de la relación entre cada par de variables.

Código	Palabras claves o frases para codificar	Ejemplo de respuesta
Creencias	Los carbohidratos o la grasa son los más dañinos y poseen más calorías	<i>Los carbohidratos influyen en el aporte calórico ya que tienen más energía y es la que más aporta de los otros datos.</i>
		<i>Las grasas y carbohidratos son los que influyen más en el aporte calórico.</i>

Relación de fuerza desde los datos	Con base en los todos influyen, pero no con la misma fuerza.	<p><i>Todos los componentes nutricionales influyen en el aporte calórico, pero no todos lo hacen con la misma fuerza.</i></p> <p><i>Los componentes que influyen en la cantidad de calorías del alimento son la grasa, carbohidratos y azúcar. No todos influyen con la misma fuerza.</i></p>
------------------------------------	--	---

Los anteriores códigos determinaron que los razonamientos que influyen al establecer la fuerza entre dos variables son:

Relación de fuerza (Creencias): la fuerza entre las variables se establece con base en las ideas o creencias.

Relación de fuerza desde el contexto: se establece que la fuerza entre cada par de variables no es la misma con base en los elementos de la situación problema.

- (ii) Códigos de la situación utilizando tecnología digital (CODAP): Establecer las variables cuya relación es la más fuerte.

Código	Palabras claves o frases para codificar	Ejemplo de respuesta
Creencias	Los carbohidratos y la grasa por que se encapsulan en el cuerpo	<p><i>Los carbohidratos, porque los carbohidratos influyen más en el importe calórico, hay casi todos los casos más carbohidratos que grasa y azúcar.</i></p> <p><i>La grasa ya que no es tan necesaria y se encapsula en el cuerpo humano.</i></p>
Comparar cantidades	Con base en los datos sus números son más elevados.	<p><i>Los carbohidratos porque comparados con los otros componentes sus números son más elevados.</i></p> <p><i>Los carbohidratos porque contienen mayor número de gramos.</i></p>

Los carbohidratos influyen con mayor fuerza, porque a partir del cheese burger el número de carbohidratos es mayor que la grasa.

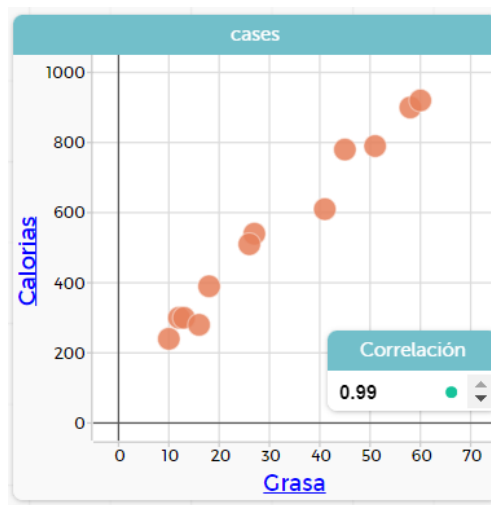
Los anteriores códigos se generaron al establecer el par de variables cuya relación es más fuerte:

Relación de fuerza (Creencias): Las variables cuya relación es la más fuerte se establece con base en las ideas o creencias.

Comparar cantidades: se establece la relación más fuerte comparando valores numéricos de la tabla de datos que se presenta la situación problema.

- (iii) Códigos de la situación utilizando tecnología digital (CODAP): Establecer la relación entre el valor numérico de la correlación y el comportamiento de los datos representados en el diagrama de dispersión para cada par de variables.

El gráfico de dispersión junto con su coeficiente de correlación para las variables grasa y calorías es:



Los códigos que surgieron son:

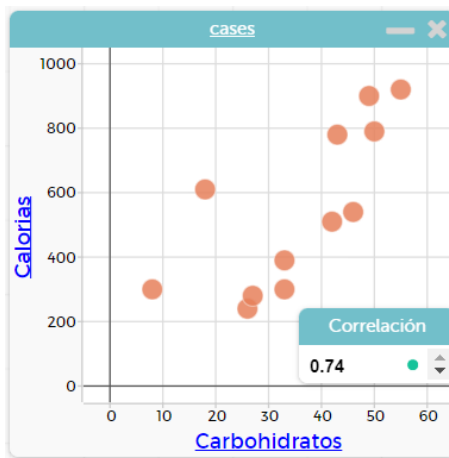
Código	Palabras claves o frases para codificar	Ejemplo de respuesta
Mayor correlación implica comportamiento lineal	Mayor correlación los puntos forman línea recta	<i>De las tres es la que más tienen correlación (x, y), ya que podrían formar una línea recta.</i> <i>Los puntos de la gráfica están más alineados que las otras, pues su valor numérico es mayor, y los puntos no están tan alejados.</i>
Conglomeración	Varía la conglomeración de los puntos Cercanos o alejados	<i>En el diagrama de grasa, se encuentran los puntos conglomerados, casi formando una línea recta, algunos puntos se unen a otros, muchos puntos forman parejas.</i> <i>Los datos se conglomeran más en el gráfico de dispersión.</i> <i>Los puntos en el diagrama se encuentran muy juntos y algunos hasta encimados.</i> <i>Los puntos se compactaron más en esta gráfica, están más cercanos.</i>

Para la relación entre el coeficiente de correlación y el gráfico de dispersión respecto a las variables grasa y calorías se generaron los siguientes códigos:

Mayor correlación implica comportamiento lineal: Se establece que a mayor valor numérico del coeficiente de correlación el comportamiento de la nube de puntos se conglopera de manera lineal.

Conglomeración: Se omite el valor numérico de la correlación haciendo referencia solo a la conglomeración de los puntos en el diagrama de dispersión.

El gráfico de dispersión junto con el valor numérico del coeficiente de correlación para las variables carbohidratos y calorías es:



Los códigos que surgieron son:

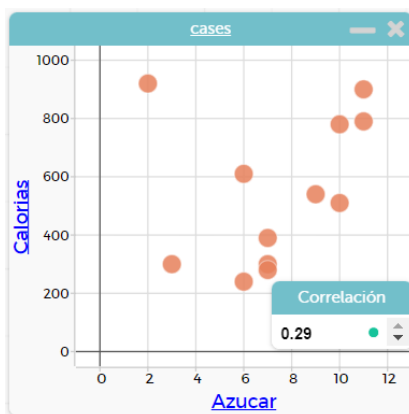
Código	Palabras claves o frases para codificar	Ejemplo de respuesta
		<i>Los puntos sufrieron un poco de dispersión.</i>
Menor correlación un poco de dispersión	A menor valor numérico de la correlación los puntos se dispersan un poco	<i>Al ser menor el valor numérico de correlación, los puntos están un poco más dispersos que en la gráfica de la grasa.</i>
		<i>Tiene sus puntos un poco más alejados/dispersos porque la correlación es como media (0.74).</i>
Grupos de puntos y otros alejados	Tiene puntos muy alejados	<i>Hay puntos que están muy juntos y forman tipo grupos, pero de ahí se alejan entre sí.</i>
		<i>Tiene sus puntos más alejados, la semi-línea que existía ya no está y ahora se ven más en grupos.</i>

Para la relación entre el coeficiente de correlación y el gráfico de dispersión respecto a las variables carbohidratos y calorías se generaron los siguientes códigos:

Menor correlación un poco de dispersión: El valor de la correlación es menor y hace que los puntos de la nube estén un poco dispersos.

Grupos de puntos y otros alejados: Se observa de manera sectorizada la nube de puntos y se identificas los puntos más alejados.

El gráfico de dispersión junto con el valor numérico del coeficiente de correlación para las variables azúcar y calorías es:



Los códigos que surgieron son:

Código	Palabras claves o frases para codificar	Ejemplo de respuesta
Menor correlación mayor dispersión	Cada vez que la correlación es menor la dispersión es mayor	<i>En esta gráfica su relación es menor que todas y entre menos cerca del número 1 más dispersos están los puntos.</i>
Puntos más dispersos	Los puntos están más separados	<i>En el caso de azúcar y calorías el diagrama de dispersión sus puntos están más separados a que su correlación es de 0.29.</i> <i>Están dispersos algunos de ellos.</i> <i>Los puntos se encuentran aún más separados.</i>

Para la relación entre el coeficiente de correlación y el gráfico de dispersión respecto a las variables azúcar y calorías se generaron los siguientes códigos:

Menor correlación mayor dispersión: Para las variables azúcar y calorías el coeficiente de correlación es menor, esto implica que los puntos en el diagrama están más dispersos.

Puntos más dispersos: Los puntos en el diagrama están más dispersos, no se establece relación alguna con el coeficiente de correlación.

- (iv) Los códigos que surgieron al establecer un argumento para el valor del coeficiente de correlación cercano a 1 y su diagrama de dispersión son:

Código	Palabras claves o frases para codificar	Ejemplo de respuesta
Puntos cerca y linealidad	Los puntos están más cerca y forman una línea	<i>Están más cerca los puntos y casi forman una línea recta.</i> <i>Cuando se acerca a 1, casi no hay dispersión pues los puntos están casi juntos y de forma lineal.</i> <i>Los puntos son más juntos y tiene una alineación ascendente formando una línea.</i>
Conglomeración de puntos	Los puntos están conglomerados	<i>Que están más conglomerados.</i> <i>Que conforman una conglomeración ascendente.</i>

Los códigos que determinan la relación entre la correlación cercana a 1 y su respectivo diagrama de dispersión son:

Puntos cerca y linealidad: La correlación cercana a 1 hace que los puntos de la nube estén más cerca y formen una línea recta.

Conglomeración de puntos: Los puntos están más conglomerados cuando la correlación es cercana a 1.

- (v) Los códigos que surgieron al establecer un argumento para el valor del coeficiente de correlación alejado de 1 y su diagrama de dispersión son:

Código	Palabras claves o frases para codificar	Ejemplo de respuesta
		<i>Están más alejados, están dispersos y no tienen una forma como tal.</i>
Dispersión (puntos alejados)	Están más alejados Hay más dispersión	<i>Sus puntos ser dispersan más, están más alejados.</i>
		<i>La dispersión es mayor pues los puntos están alejados.</i>
Mayor dispersión pierden forma lineal	Mayor dispersión y se pierde la forma lineal	<i>Que los puntos se dispersan, se nota que entre más alejado esté pierden la forma lineal y cercana, expandiéndolos todos.</i> <i>Que los puntos se dispersan y están menos alineados que cuando es un valor cercano a uno.</i>

Los códigos que determinan la relación entre la correlación alejada de 1 y su respectivo diagrama de dispersión son:

Dispersión (puntos alejados): Los puntos de la nube se encuentran más dispersos, es decir, más separados entre sí.

Mayor dispersión pierde forma lineal: Los puntos en la nube poseen más dispersión y por lo tanto pierden su forma lineal.

- (vi) Determinar cuál es la variable que influye fuertemente en el aporte calórico.

Código	Palabras claves o frases para codificar	Ejemplo de respuesta
---------------	--	-----------------------------

El azúcar: La nube de puntos está más dispersa	El azúcar	<i>Si, el azúcar ya que es lo que altera más a las calorías ya que el diagrama de puntos está muy disperso.</i>
Las grasas porque su correlación es cercana a uno	Las grasas	<i>La azúcar, porque son menos gramos, tiene más calorías que la grasa y carbohidratos y su correlación es más baja que todas las demás.</i>
Los carbohidratos: creencias	Los carbohidratos	<i>Las grasas, aunque los valores sean los más pequeños, son lo que alteran más a las calorías porque la correlación es un valor muy cercano al uno y en el diagrama los puntos están juntos.</i>
		<i>Las grasas, aunque los valores sean los más pequeños, son los que alteran más a las calorías porque la correlación s un valor muy cercano al uno y en el diagrama los puntos están juntos.</i>
		<i>Sí, los carbohidratos como ya fue antes mencionado debido a que hay carbohidratos en las calorías a diferencia del azúcar y grasa.</i>

Los códigos anteriores determinaron que la variable que influye fuertemente en el aporte calórico puede ser:

El azúcar (la nube de puntos está más dispersa): Dado que en el diagrama de dispersión la nube de puntos es la más dispersa y su correlación es la más baja luego esta es la variable que influye fuertemente.

Las grasas (su correlación es cercana a 1): Dado que la correlación es cercana a 1 y los puntos de la nube están más juntos en el diagrama de dispersión está es la variable que influye fuertemente.

Los carbohidratos (creencias): Justificando desde sus creencias establece que esta es la variable que influye fuertemente.

(vii) Determinar cuál es la variable que influye de manera débil en el aporte calórico.

Código	Palabras claves o frases para codificar	Ejemplo de respuesta
Las grasas: Los puntos de la nube están juntos	Las grasas	<p><i>Si, la grasa ya que no afecto mucho a las calorías ya que el diagrama de puntos está muy junto y casi logran formar una línea recta.</i></p> <p><i>La grasa, porque mientras más grasa, son las mismas calorías.</i></p>
Azúcar	Azúcar	<p><i>Sí, el azúcar ya que es el componente es el de más baja cantidad si lo comparamos con la grasa y carbohidratos.</i></p> <p><i>El azúcar porque no influye fuertemente y además no está dependiendo para aportar calorías, la correlación es menor.</i></p> <p><i>Influye débilmente el azúcar, porque de ella no depende tanto las calorías y si es menor o mayor no mueve/influye tanto el total de calorías y la correlación es menor.</i></p>

Los códigos anteriores determinaron que la variable que influye de manera débil en el aporte calórico puede ser:

Las grasas (los puntos de la nube están juntos): El diagrama de dispersión que muestra los puntos más juntos es la variable que influye de manera débil en el aporte calórico.

El azúcar: Dado que en el diagrama de dispersión los puntos están alejados y tienen una correlación menor alejado de 1 influye débilmente en el aporte calórico.

(viii) ¿Cómo influyen los valores extremos en el valor numérico de la correlación?

Código	Palabras claves o frases para codificar	Ejemplo de respuesta
		<i>Cuando las calorías y azúcar son modificadas, la correlación aumenta así muevas los valores más altos o bajos.</i>
Cerca o lejos cambia la correlación	Cerca o lejos cambia	<i>Está dependiendo de la forma en que se mueve hacia los puntos se acerca al 1 y a la izquierda se aleja del 1, pero al moverla abajo y se mueva a los puntos el valor se aleja del 1 y a la izquierda se acerca al 1.</i>
Cerca al conglomerado mayor correlación	Muy cerca al conglomerado mayor el coeficiente de correlación	<i>Al realizar el movimiento indicado se nota que la correlación aumenta cuando el punto se acerca a la conglomeración y disminuye, volver a su lugar de origen o se aleje más.</i> <i>Al mover el anterior punto cerca de los demás el valor numérico de correlación aumenta fuertemente, pues se encontraba en mayor relación con los demás.</i>

Los códigos anteriores surgieron al establecer la influencia de los valores extremos sobre el coeficiente de correlación.

Las grasas (los puntos de la nube están juntos): El diagrama de dispersión que muestra los puntos más juntos es la variable que influye de manera débil en el aporte calórico.

El azúcar: Dado que en el diagrama de dispersión los puntos están alejados y tienen una correlación menor alejado de 1, ésta influye débilmente en el aporte calórico.

La tabla 4 resume la frecuencia obtenida para cada código respecto a las preguntas que involucraban juzgar la fuerza entre dos variables sin utilizar tecnología digital.

Tabla 4. Frecuencia para cada código

Pregunta	Código	Frecuencia (Parejas)
¿Qué componentes nutricionales (grasa, carbohidratos o azúcar) influyen en el aporte calórico?	Creencias	7
	Desde los datos	5
¿Qué componente influye con mayor fuerza? ¿Por qué?	Creencias	4
	Comparar cantidades	7

La tabla 5 resume la frecuencia obtenida para cada código respecto a las preguntas que involucraban juzgar la nube de puntos en el diagrama de dispersión para datos bivariados junto con la influencia del coeficiente de correlación en su comportamiento utilizando CODAP.

Pregunta	Código	Frecuencia (Parejas)
¿Existe alguna relación entre el valor numérico de correlación y el comportamiento de los datos representados en el diagrama de dispersión?	<i>Diagrama: Grasas-Calorías</i>	
	Mayor Correlación implica comportamiento lineal	4
	Conglomeración	8
	<i>Diagrama: Carbohidratos-Calorías</i>	
	Menor Correlación un poco de dispersión	9

	Grupos de puntos y otros alejados	3
	Diagrama: Azúcar-Calorías	
	Menor correlación más dispersión	5
	Puntos más dispersos	7
¿Qué observas en los datos representados en el diagrama de dispersión cuando la correlación es un valor que se acerca a 1?	Puntos cerca y linealidad	10
	Conglomeración	2
¿Qué observas en los datos representados en el diagrama de dispersión cuando la correlación es un valor lejano de 1?	Dispersos y alejados	8
	Mayor dispersión pierden forma lineal	4
¿Existe algún componente nutricional (grasa, carbohidratos o azúcar) presentado que influya fuertemente en el aporte calórico (calorías)? ¿por qué?	El azúcar: La nube de puntos es más dispersa	2
	Las grasas porque su correlación es cercana a uno	9
	Los carbohidratos: creencias	1
¿Existe algún componente nutricional (grasa, carbohidratos o azúcar) presentado en la tabla 2 que influya débilmente en el aporte calórico? ¿por qué?	Grasas: Los puntos de la nube están juntos	2
	Azúcar	10
Mueve los valores extremos y observa lo que ocurre con el valor numérico de la correlación	Cerca o lejos cambia la correlación	9
	Cerca al conglomerado mayor correlación	3
¿Cómo influye los valores extremos en la correlación (azúcar, calorías)?		

Codificación inicial de actividad 3

A continuación, se muestran para cada pregunta, los códigos que se generaron junto con las palabras clave o frases para codificar y ejemplos de respuestas.

- (i) Códigos de la situación a lápiz y papel: describe la relación entre el cambio ANE y el aumento de grasa.

Tabla 7. Palabras clave o frases y ejemplos de respuestas por código

Código	Palabras claves o frases para codificar	Ejemplo de respuesta
COVARIACIÓN ESTADÍSTICA	El comportamiento de ambas variables es que aumentan o disminuyen, no es constante.	<p>“cambio ANE debido a su cambio sea negativo o positivo cambia la grasa</p> <p>“la grasa aumenta o disminuye dependiendo de cuántas calorías queman en ANE”</p> <p>“si las calorías disminuyen o aumentan dependiendo del ANE, la grasa aumentará o disminuirá también”</p> <p>“las calorías que se queman aumentaron los kg o las calorías que se queman disminuirán los kg”</p> <p>“debido a su cambio sea negativo o positivo cambia la grasa”</p>
COVARIACIÓN FUNCIONAL	El cambio en las variables es constante. Ambas variables aumentan o ambas disminuyen.	<p>“más grasa consumes más calorías aumentas”</p> <p>“entre menos cambio en las calorías mediante los movimientos, menos grasa disminuía”</p> <p>“ya que va a depender de la caloría porque si no, no puede cambiar”</p>
CONTEXTO	Se debe a que la grasa es difícil de eliminar Si no haces ejercicios no quemas grasa	<p>“la grasa aumenta porque se queda encapsulada”</p> <p>“entre menos actividad realices más grasa”</p> <p>“las grasas se quedan en el cuerpo”</p>

Los anteriores códigos determinaron una forma de *covariación* para la relación entre las variables, es decir, la relación de cambio de una variable a medida que la otra variable cambia puede ser de tipo:

- *Covariación estadística*, cuando las variables presentan un comportamiento que no es proporcional, es decir, los datos no se ajustan de modo preciso a una función matemática, se presenta variabilidad en el cambio del conjunto de datos bivariados.
 - *Covariación funcional*, cuando el cambio en las variables es proporcional, es decir, donde los valores de la variable y quedan determinados de un modo preciso por los valores de la otra variable que se considera como independiente.
 - *Contexto*, cuando el cambio en las variables se explica en términos de los elementos implicados en la situación del problema y no se tiene en cuenta el conjunto de datos.
- (ii) Códigos de la situación utilizando tecnología digital (CODAP): estudiar la relación entre las variables, describiendo la intensidad y la dirección de la nube de puntos.

Tabla 8. Palabras clave o frases y ejemplos de respuestas por código

Código	Palabras clave o frases para los codificar	Ejemplo de respuesta
RELACIÓN INVERSA	los valores de una variable aumentan y los valores de la otra variable disminuyen	“al aumentar las calorías hay más quema de grasa”
		“Entre más cambio en ANE, menos aumentos de grasa presenta”
		“Entre menos actividad realices más grasa...”
		“entre menos actividades realices más grasa acumulas”
		“los movimientos que no sean ejercicios hacen que pierdas grasa y es lógico, al hacer ejercicio haces movimientos”
		“cuando el cambio en ANE es mayor, el aumento de grasa es menor”

		<p>“más incremento en ANE más pérdida de grasa hay. Entre menos incremento en ANE menos pérdida de grasa”</p> <p>“a mayor reducción del ANE hay un incremento en la cantidad de grasa.”</p>
CONTEXTO	<p>Se debe a que la grasa es difícil de eliminar</p> <p>Si no haces ejercicios no quemas grasa</p>	<p>“el aumento de grasa depende de la cantidad de ANE que se realiza diariamente”</p> <p>“la grasa no aumenta porque, aunque no se hace ejercicio si se mantiene en movimiento.”</p>

Utilizando la tecnología digital para juzgar la relación entre las variables, surgió un nuevo código *relación inversa*, y está presente cuando se evidencia que los valores de una variable tienden a aumentar conforme los valores de la otra variable tienden a disminuir o viceversa. Se esperaba la descripción de la dirección de la nube de puntos como negativa, lo cual no se evidenció. Sin embargo, los estudiantes notaron que el comportamiento global de los datos es inverso, mientras una variable aumenta la otra disminuye.

- (iii) Códigos de la situación con el archivo de GeoGebra: ¿Dónde consideras que debes ubicar la Recta amarilla para que corresponda a la Recta de regresión? Explica tu criterio.

Tabla 9. Palabras clave o frases y ejemplos de respuestas por código

Código	Palabras clave o frases para los codificar	Ejemplo de respuesta
CERCANÍA	Es la línea que está cerca de la nube de puntos.	<p>“en ese lugar todos los puntos están cerca de la recta, es decir dónde están los puntos (750,1) y el punto (-153.5842, 3.84704)”</p> <p>“pasando centrada en los puntos”</p> <p>“Que los puntos azules estuvieron más cercanos entre sí a la recta”</p>

		“creo que es la recta que pasa cerca de todos los puntos y no los une, así cada uno tiene cierta distancia”
DISTANCIA	La línea es aquella donde el valor de los residuos es pequeño porque hace que la recta este cerca a la nube de puntos	<p>“donde el número de residuos este en su menor expresión para que la distancia sea menor”</p> <p>“Donde los residuos nos den el menor números posible, puesto que la distancia entre los puntos a la recta es la menor posible”</p> <p>“Entre menos distancia entre los puntos azules y la línea menos residuos hay.”</p> <p>“es decir, mientras menos sea la distancia con todos los puntos menor son los residuos.”</p>
PERTENENCIA	Es la línea que pasa por el primer y último punto Es la línea que contiene más puntos	“cómo cambia la variable Y cuando se mueve la variable X, es mínimo cuando pasa por casi todos los puntos.”

Para el criterio de la línea de mejor ajuste con el uso del archivo de GeoGebra surgieron los siguientes códigos:

El código *cercanía* se evidencia cuando se describe la línea de mejor ajuste como aquella que se acerca lo más posible a la nube de datos.

El código *distancia* muestra que el valor numérico de los residuos cuando es menor hace que la distancia entre la posible línea de mejor ajuste y la nube de puntos sea pequeña.

El código *pertenencia* es el criterio de considerar la línea de mejor ajuste como aquella que debe pasar por algunos puntos de la nube. No es claro cuales puntos de los datos se tomaron como referencia para considerar la línea de mejor ajuste.

- (iv) Códigos de: ¿Qué ocurre con el valor de cada residual si mueves la recta cerca o lejos de la nube de puntos?

Tabla 10. Palabras clave o frases y ejemplos de respuestas por código

Código	Palabras clave o frases para los codificar	Ejemplo de respuesta
DISTANCIA	La línea es aquella donde el valor de los residuos es pequeño porque hace que la recta este cerca a la nube de puntos	<p><i>“Entre menos distancia entre todos los puntos a la línea menor va a ser el residuo”</i></p> <p><i>“Entre menos distancia entre la línea amarilla y los puntos azules, el residuo disminuye”</i></p> <p><i>“Entre menos distancia hay entre la línea y los puntos el residuo es menor”</i></p> <p><i>“Entre menos distancia hay de los puntos azules y la recta, menor es el residuo”</i></p>
VALOR NÚMÉRICO RESIDUOS	El valor numérico de los residuos representa la distancia de la nube a la recta	<p><i>“distancia hay entre cada uno de los puntos y la recta”</i></p> <p><i>“Que la distancia que hay entre cada uno de los puntos y la recta sumadas es el resultado de los residuos”</i></p> <p><i>“Que entre menos distancia haya en los puntos y la línea, el residuo va a cambiar”</i></p> <p><i>“Entre más alejada esté la recta de ajuste de los puntos de la tabla en la gráfica es mayor el valor de residuos ya que hay una mayor distancia”</i></p> <p><i>“Que haga menos residuos, porque la línea amarilla debe estar entre menor”</i></p>
RECTA CENTRADA IMPLICA MENOR RESIDUOS	Una recta que pasa por en medio de la nube de puntos hace que el valor de los residuos sea el menor	<p><i>“Cuando la recta está mejor centrada, el nivel de residuos es menor, es decir, pasa centrándose en medio de los puntos”</i></p> <p><i>“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”</i></p>

En esta pregunta surgieron dos códigos nuevos:

El código *valor numérico de los residuos es igual a la distancia entre la nube y la recta*, es decir, se asume el valor total de los residuales como aquel valor que representa la cercanía o lejanía de la línea de ajuste.

El código *la recta centrada implica menor residuos* evidencia que se enfatiza en la línea de mejor ajuste como aquella que debe representar la tendencia de acumulación del conjunto de datos más o menos dispersos centrándose en medio de la nube de manera que el residuo es menor en esa posición.

- (v) Códigos de: ¿La manera como ubicaste la recta amarilla es igual a la recta de regresión arrojada por GeoGebra? ¿Sí? ¿No? ¿En qué se diferencian?

Tabla 11. Palabras clave o frases y ejemplos de respuestas por código

Código	Palabras clave o frases para los codificar	Ejemplo de respuesta
COMPARACIÓN DE POSICIÓN	Las rectas están ubicadas de la misma manera muy cercanas Las rectas son paralelas Las rectas están muy separadas Tienen diferente sentido	<p><i>“son casi iguales a la amarilla está un poco arriba de la recta roja per están pegaditas y casi en el mismo lugar”</i></p> <p><i>“no se encuentran en el mismo sitio”</i></p> <p><i>“ya que no están a la misma distancia”</i></p> <p><i>“se diferencian en que la distancia en el eje X es menor a partir de 800”</i></p> <p><i>“está quedó paralela y muy cercana”</i></p> <p><i>“la recta de regresión aumentas la separación en comparación de la amarilla”</i></p> <p><i>“que la recta de regresión se acerca al valor mínimo de distancias entre los puntos y la recta”</i></p> <p><i>“y solo la diferencian unos milímetros en la centración”</i></p> <p><i>“diferencia en que la recta roja pasa por el centro de todos los puntos (sin tocarlos) y la amarilla no”</i></p>

COMPARACIÓN DE RESIDUOS	Los valores de los residuos son iguales	“a diferencia de la mostrada en GeoGebra nuestra recta tenía de residuo 10 y la de regresión tenía 5”
	Los valores de los residuos son diferentes	“la línea de regresión dio como resultado un residuo de 5.25 mientras que nuestra línea daba uno de 5.1”
	Los valores de los residuos son muy cercaños	

Respecto a la comparación de la línea de mejor ajuste que se estableció y la recta arrojada por GeoGebra se generaron dos códigos: *comparación de posición*, se evidencia con argumentos con relación a la cercanía o lejanía entre ambas rectas o entre las rectas y la nube de puntos. *Comparación de residuos*, para diferenciar las rectas se toma como elemento principal el valor de los residuos.

La tabla 19 resume la frecuencia obtenida para cada código respecto a las preguntas que involucraban la descripción de la relación entre las variables.

Tabla 12 . Frecuencia para cada código

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

La anterior tabla muestra que los códigos con mayor frecuencia Del análisis más detallado de los anteriores códigos se generaron las categorías principales (*contexto, relación inversa*) para el juicio acerca de la relación entre dos variables.

La tabla 20 resume la frecuencia de los códigos obtenidos para cada pregunta que implicaba el uso del archivo de GeoGebra al determinar la línea de mejor ajuste.

Tabla 13. Frecuencia para cada código

Pregunta	Código	Frecuencia (Parejas)
¿Dónde consideras que debes ubicar la Recta amarilla para que corresponda a la Recta de regresión? Explica tu criterio.	Cercanía	4
	Distancia	4
	Pertenencia	1
	Ninguno	2
¿Qué ocurre con el valor de cada residual si mueves la recta cerca o lejos de la nube de puntos?	Distancia	4
	Valor numérico residuos= distancia (nube y la recta)	5
	Recta centrada implica menor residuos	2
¿La manera como ubicaste la Recta amarilla es igual a la Recta de regresión arrojada por GeoGebra? ¿Sí? ¿No? ¿En qué se diferencian?	Comparación de posición	9
	Comparación de residuos	2

La anterior tabla muestra que los códigos que presentaron mayor frecuencia en las respuestas relacionadas con la línea de mejor ajuste fueron: distancia (8), el valor numérico de los residuos es igual a la distancia entre la nube y la recta (5) y comparación de la posición entre la línea de mejor ajuste y la recta arrojada por GeoGebra (9), los códigos de la anterior tabla comparten la misma fuente y generan la categoría principal *distancia*.

Apéndice H: Codificación intermedia de las actividades

Durante el proceso de codificación intermedia de los tres instrumentos se fueron generando códigos principales y algunos subcódigos. Estos códigos principales se convirtieron en dimensiones que agrupan algunas categorías o subcódigos.

Dimensión: Covariación

A esta dimensión pertenecen los códigos que se relacionan con los razonamientos de los estudiantes cuando se les pide analizar la relación entre las variables. A continuación, se muestran los códigos con su respectiva frecuencia.

	Pregunta	Código	Frecuencia (Parejas)
Actividad 1	Juzgar la relación entre las variables a partir del contexto a lápiz y papel.	Tendencia global positiva	8
		Operación aritmética: vista individual	4
	Realizar el diagrama de dispersión a lápiz y papel	Sesgo de atención: idea funcional	5
		Nube de puntos	7
	Describir la nube de puntos que muestra el diagrama de dispersión en CODAP.	Linealidad positiva con error	5
		Orden ascendente con error	5
		Respuesta fuera del contexto	2
	Juzgar la relación entre las variables a partir del diagrama de dispersión de CODAP.	Dependencia positiva	7
		Dependencia funcional	2
		Dependencia con error	3
Actividad 3	A lápiz y papel describir la relación entre las variables	Covariación funcional	3
		Componente aleatorio	5
		Contexto	3
		Relación inversa	8

Utilizando CODAP realizar el diagrama de dispersión, determinar la correlación y describir la relación	Contexto	3
--	----------	---

Los códigos de la dimensión covariación estadística se agruparon en dos categorías llamadas *mecanismo determinista*, a esta categoría pertenecen los códigos que evidencian razonamientos donde el tratamiento de la relación entre las dos variables es de manera proporcional o funcional, es decir, donde el error es ausente y se trata como covariación desde una vista del cálculo. La segunda categoría que emergió es denominada *covariación estadística* dado que se evidencia razonamientos donde el error hace presencia y la aleatoriedad.

Categoría: Mecanismo determinista

- *Covariación funcional*, cuando el cambio en las variables es proporcional, es decir, donde los valores de la variable y quedan determinados de un modo preciso por los valores de la otra variable que se considera como independiente. La línea es aquella que pasa por todos los puntos.
- *Operación aritmética (vista individual)*: Realizar operaciones aritméticas con algunos datos.
- *Sesgo de atención-idea funcional*: Se ubican los pares ordenados del conjunto de datos bivariados en el plano cartesiano y se traza la unión por semi-rectas entre ellos.
- *Dependencia funcional*: La relación entre las variables es proporcional, una variable aumenta la otra también aumenta.

Categoría: Covariación estadística

- *Graficar nubes de puntos*: Se ubican los pares ordenados del conjunto de datos bivariados en el plano cartesiano.

- *Tendencia global positiva*: Hay una relación positiva, si los valores de una variable tienden a aumentar conforme los valores de la otra variable también aumentan.
- *Linealidad positiva con error*: La tendencia de la nube de puntos es positiva y su forma es casi lineal.
- *Orden ascendente con error*: Describen el orden ascendente de los puntos en el diagrama sin aceptar su comportamiento lineal.
- *Dependencia positiva*: La relación entre las variables es dependiente porque ambas aumentan.
- *Dependencia con error*: Es una relación donde una variable depende de la otra y ambas aumentan de manera no proporcional.
- *Componente aleatorio*: Se enfoca la 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 conforme leen de izquierda a derecha, notan que no hay patrones previsibles.
- *Relación inversa*: Describe de manera global el comportamiento de la nube, mientras una variable aumenta la otra disminuye.

Dimensión: Coeficiente de Correlación y diagramas de dispersión

Pregunta	Código	Frecuencia (Parejas)
¿Qué componentes nutricionales (grasa, carbohidratos o azúcar) influyen en el aporte calórico? Los que influyen ¿Lo hacen con la misma fuerza?	Creencias	7
	Desde los datos	5
¿Qué componente influye con mayor fuerza? ¿Por qué?	Creencias	4
	Comparar cantidades	7
	<i>Diagrama: Grasas-Calorías</i>	
¿Existe alguna relación entre el valor numérico de correlación y el	Mayor Correlación implica comportamiento lineal	4

comportamiento de los datos representados en el diagrama de dispersión?	Conglomeración	8
	<i>Diagrama: Carbohidratos-Calorías</i>	
	Menor Correlación un poco de dispersión	9
	Grupos de puntos y otros alejados	3
	<i>Diagrama: Azúcar-Calorías</i>	
	Menor correlación más dispersión	5
	Puntos más dispersos	7
¿Qué observas en los datos representados en el diagrama de dispersión cuando la correlación es un valor que se acerca a 1?	Puntos cerca y linealidad	10
	Conglomeración	2
¿Qué observas en los datos representados en el diagrama de dispersión cuando la correlación es un valor lejano de 1?	Dispersos y alejados	8
	Mayor dispersión pierden forma lineal	4
¿Existe algún componente nutricional (grasa, carbohidratos o azúcar) presentado que influya fuertemente en el aporte calórico (calorías)? ¿por qué?	El azúcar: La nube de puntos es más dispersa	2
	Las grasas porque su correlación es cercana a uno	9
	Los carbohidratos: creencias	1
¿Existe algún componente nutricional (grasa, carbohidratos o azúcar) presentado en la tabla 2 que influya débilmente en el aporte calórico? ¿por qué?	Grasas: Los puntos de la nube están juntos	2
	Azúcar	10
Mueve los valores extremos y observa lo que ocurre con el valor numérico de la correlación ¿Cómo influye los valores extremos en la correlación (azúcar, calorías)?	Cerca o lejos cambia la correlación	9
	Cerca al conglomerado mayor correlación	3

Dimensión: Recta de Mejor ajuste

Actividad	Pregunta	Código	Frecuencia (Parejas)	
Actividad 1	Estimar a partir de la situación problema a lápiz y papel.	Aritmético siguiendo un patrón	6	
		Aritmético sin patrón definido	6	
	Predecir a partir de la situación problema a lápiz y papel.	Aritmético siguiendo un patrón	4	
		Aritmético sin patrón definido	8	
	Como utilizar la línea arrojada por CODAP para estimar y predecir.	Recta representa toda la información	10	
		Proceso aritmético	2	
	Cuál es el criterio que se cree utiliza CODAP para trazar la línea de mejor ajuste	Concepción funcional	1	
		Cercanía	1	
		Concepción local	3	
		Operacional o Algebraico	5	
		No describe ninguno	2	
	Actividad 3	¿Dónde consideras que debes ubicar la Recta amarilla para que corresponda a la Recta de regresión? Explica tu criterio.	Cercanía	4
			Distancia	4
			Pertenencia	1
			Ninguno	2
¿Qué ocurre con el valor de cada residual si mueves la recta cerca o lejos de la nube de puntos?		Distancia	4	
		Valor numérico residuos= distancia (nube y la recta)	5	
		Recta centrada implica menor residuos	2	
¿La manera como ubicaste la Recta amarilla es igual a la Recta de regresión arrojada por		Comparación de posición	9	
		Comparación de residuos	2	

GeoGebra? ¿Sí? ¿No? ¿En qué se diferencian?		
---	--	--

Categoría: Mecanismo determinista

- *Aritmético siguiendo un patrón:* Buscar un patrón en la tabla de valores entre un par de datos para establecer un procedimiento aritmético.
- *Aritmético sin patrón definido:* Sin patrón definido se efectúan algunas operaciones básicas (suma, resta, multiplicación y división).
- *Proceso aritmético:* Utilizar un par de valores para realizar algún procedimiento matemático.
- *Concepción local:* La línea es aquella que contiene un par de puntos
- *Operacional o algebraico:* La línea se determina mediante una expresión algebraica.

Categoría: Distancia nube recta

- *Recta representa toda la información:* Emplear la línea arrojada por el software para estimar y predecir el valor lo cual implica asumir la línea como representante de todos los datos.
- *Cercanía:* Se evidencia cuando se describe la línea de mejor ajuste como aquella que se acerca lo más posible a la nube de datos.
- *Distancia:* muestra que el valor numérico de los residuos cuando es menor hace que la distancia entre la posible línea de mejor ajuste y la nube de puntos sea pequeña.
- *Pertenencia:* es el criterio de considerar la línea de mejor ajuste como aquella que debe pasar por algunos puntos de la nube. No es claro cuales puntos de los datos se tomaron como referencia para considerar la línea de mejor ajuste.
- *Valor numérico de los residuos:* es igual a la distancia entre la nube y la recta, es decir, se asume el valor total de los residuales como aquel valor que representa la cercanía o lejanía de la línea de ajuste.

- *Recta centrada*: implica menor residuos evidencia que se enfatiza en la línea de mejor ajuste como aquella que debe representar la tendencia de acumulación del conjunto de datos más o menos dispersos centrándose en medio de la nube de manera que el residuo es menor en esa posición.
- *Comparación de posición*: se evidencia con argumentos con relación a la cercanía o lejanía entre ambas rectas o entre las rectas y la nube de puntos.
- *Comparación de residuos*: para diferenciar las rectas se toma como elemento principal el valor de los residuos.

Apéndice I: Codificación del cuestionario diagnóstico y evaluativo

Codificación Inicial del cuestionario diagnóstico

Componente de álgebra básica

A continuación, se muestran para los ítems de acuerdo con el contenido, los códigos que se generaron y ejemplos de respuestas.

(viii) Códigos de los ítems cuyo contenido de álgebra es:

- *Establecer algebraicamente la función lineal que representa el enunciado. Es decir, pasar del enunciado verbal a establecer una expresión algebraica.*
- *Dado un valor sustituir en la expresión algebraica y hallar su valor.*

Código	Ejemplo de respuesta
Lenguaje algebraico	<i>Porque el cargo fijo por la tarjeta es 100 pesos y sobre ese precio se cobrarán 6 pesos más por una cantidad de n tarjetas</i>
	<i>Porque se multiplica el cargo adicional 6 por el número de tarjetas n y se le suma el costo fijo 100.</i>
Reemplazar valores	<i>Se reemplaza a $n=3$ y $n=9$ Sustituir el valor numérico en la x</i>
Interpretación errónea	<i>Ya que el costo total de la tarjeta sería 100 pesos multiplicada por n número de tarjetas daría el costo final.</i>

Los anteriores códigos muestran los razonamientos respecto al contenido de expresiones algebraicas, estos pueden ser:

- *Lenguaje algebraico:* Se establece la forma algebraica de una expresión verbal identificando correctamente las variables.
- *Reemplazar valores:* Sustitución correcta de un valor en una expresión algebraica.

- *Interpretación errónea:* Se establece una expresión algebraica que no corresponde al enunciado y realiza operaciones aritméticas incorrectas.

(ix) Códigos del ítem cuyo contenido de álgebra es:

- *Establecer la función lineal que representa la relación a partir de datos tabulares.*

Código	Ejemplo de respuesta
Lenguaje algebraico.	$3x+1$. Porque multiplicando tres veces lo que sale en x más 1 da y .
	$3x+1$. Busqué un número que explicará la relación de x y y .
	$x+y3$. La relación entre los números x es 1, y entre los de “ y ” es 3.
Observación individual de variables	$x+1=y+3$. Porque si se multiplica 3 por cualquiera de los números que se encuentren en x y le agregas 1 da como resultado los números encontrados en y .

Los anteriores códigos muestran los razonamientos respecto al contenido de representar la relación algebraica a partir de una tabla de datos, estos pueden ser:

- *Lenguaje algebraico:* Mediante el lenguaje algebraico se representa la relación funcional de los datos bivariados.
- *Observación individual de variables:* Se establece la relación estableciendo algún patrón numérico para cada variable.

(x) Códigos del ítem cuyo contenido de álgebra es:

- *Determinar la función lineal que representa la relación de proporcionalidad y encontrar los valores.*

Código	Ejemplo de respuesta
Lenguaje algebraico.	Debido a que es proporcional no se altera los resultados. 4 es a 9, entonces 8 sería a 18.
	4 es a 9, como 8 es a k . Realizando $4/9 = 8/k$; $4/k = 9/4$

Patrón Aritmético

Ya que el multiplico que hace que P sea 8 es 2. Ya que si divides 45 entre 9 es 5.

- *Lenguaje algebraico:* Mediante el lenguaje algebraico se representa la relación funcional de los datos bivariados.
- *Patrón aritmético:* Establecen un patrón aritmético utilizando sumas, restas o multiplicaciones sin uso de una expresión algebraica de esa manera establecen la proporcionalidad.

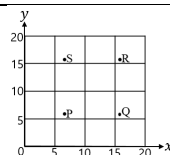
(xi) Códigos de los ítems cuyo contenido de álgebra es:

- *Identificar en una pareja ordenada la abscisa y la ordenada ubicando el punto en el plano cartesiano.*

Código

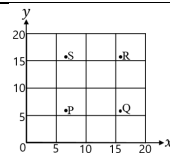
Ejemplo de respuesta

Intercambio entre abscisa y ordenada



El punto Q. Es el punto más cercano a los valores a 7 y 16

Ubicación correcta



El punto S.

Los anteriores códigos muestran los razonamientos respecto al contenido de expresiones algebraicas, estos pueden ser:

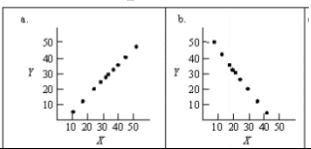
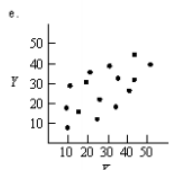
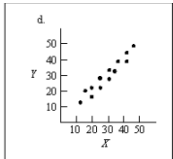
- *Intercambio entre abscisa y ordenada:* Ubicación incorrecta asume la ordenada como x y la abscisa como y.
- *Ubicación correcta:* Sustitución correcta de un valor en una expresión algebraica.

Componente estadística

A continuación, se muestran para los ítems de acuerdo con el contenido, los códigos que se generaron y ejemplos de respuestas.

(xii) Códigos de los ítems cuyo contenido de estadística es:

- *Establecer el diagrama de dispersión que represente una correlación de 0,6.*

Código	Ejemplo de respuesta
Comportamiento lineal perfecto positivo o negativo	<p><i>Una correlación que muestre cerca de 0.60</i></p> 
Poca dispersión conglomeración débil	<p><i>Una correlación que muestre cerca de 0.60</i></p> 
Comportamiento lineal, pero conglomeración fuerte	<p><i>Una correlación que muestre cerca de 0.60</i></p> 

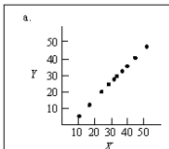
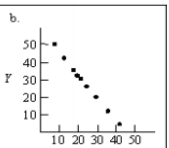
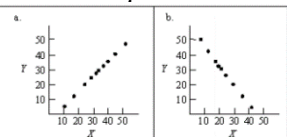
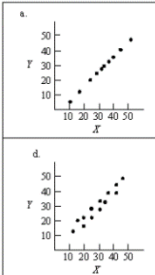
Los anteriores códigos muestran los razonamientos respecto al contenido de fuerza y dirección en diagramas de dispersión, estos pueden ser:

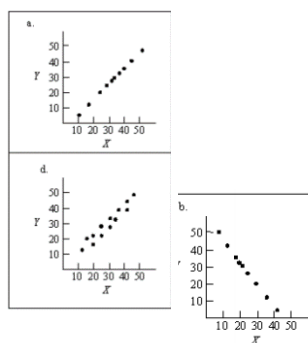
- *Comportamiento lineal perfecto positivo o negativo:* correlación de 0.6 implica un comportamiento lineal perfecto.
- *Poca dispersión conglomeración débil:* correlación 0.6 no muestra una relación lineal perfecta pero el valor por encima de 0.5 muestra una nube de puntos con fuerza débil.

- *Comportamiento lineal, pero conglomeración fuerte:* correlación de 0.6 muestra un comportamiento lineal, pero con una nube de puntos con fuerza fuerte.

(xiii) Códigos de los ítems cuyo contenido de estadística es:

- *Reconocer gráficamente una relación fuerte entre las variables*

Código	Ejemplo de respuesta
Linealidad perfecta positiva	<p><i>Fuerza implica:</i></p> 
Linealidad perfecta negativa	<p><i>Fuerza implica:</i></p> 
Linealidad perfecta positiva y negativa	<p><i>Fuerza implica:</i></p> 
Linealidad positiva	<p><i>Fuerza implica:</i></p> 
Comportamiento lineal	<p><i>Fuerza implica:</i></p>



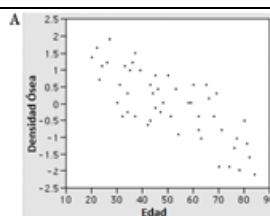
Los anteriores códigos muestran los razonamientos respecto al contenido de diagramas de dispersión y correlación, estos pueden ser:

- *Linealidad perfecta positiva:* Fuerza implica una nube de puntos cuyo comportamiento es perfectamente lineal en sentido positivo.
- *Linealidad perfecta negativa:* Fuerza implica una nube de puntos cuyo comportamiento es perfectamente lineal en sentido negativo.
- *Linealidad perfecta positiva y negativa:* Fuerza implica ambos comportamientos perfectamente lineales de sentido positivo y negativo.
- *Linealidad positiva:* Fuerza implica comportamiento lineal en sentido solo positivo de la nube de puntos.
- *Comportamiento lineal:* Fuerza implica un comportamiento lineal en cualquier sentido.

(xiv) Códigos de los ítems cuyo contenido de estadística es:

- *Reconocer la relación establecida en forma textual e identifique la figura correspondiente.*

Código	Ejemplo de respuesta
Comportamiento bivariado negativo	<i>¿Cuál de los siguientes gráficos muestra que a medida que las mujeres envejecen tienden a tener una densidad ósea más baja?</i>



Los anteriores códigos muestran los razonamientos respecto al contenido diagramas de dispersión, estos pueden ser:

- *Comportamiento bivariado negativo*: Identificación correcta del diagrama de dispersión que representa la relación textual.

(xv) Códigos de los ítems cuyo contenido de estadística es recta de mejor ajuste:

- *Establecer un criterio para representar la recta de mejor ajuste en un diagrama de dispersión.*

Código	Ejemplo de respuesta
Dependencia funcional	<i>Unir puntos</i>
Concepción local	<i>Une tres puntos (inicio de mitad en el gráfico y el último). Trazo una recta que pasaba por el punto de centro y el último punto y dice: la satisfacción más repetitiva es el 17 y es la que se acerca más. Traza una línea que pasa por un punto inicial un punto de la mitad y el último y dice: En general la satisfacción la variante más repetida es 17 y es la que se acerca más y gana29 o más es mayor la satisfacción</i>

	<i>Hice un gráfico que pasa por en medio de los puntos (traza una recta inclinada intentando que pasé por en medio de los puntos)</i>
Medio de todos los puntos (cercanía)	<i>Traza una línea en medio de todos los puntos con dirección positiva y dice: Escogí esa línea ya que, aunque no coincide con algunos puntos se acerca y se puede tener resultados con una mejor relación.</i>

Los anteriores códigos muestran los razonamientos respecto al contenido recta de mejor ajuste, estos pueden ser:

- *Dependencia funcional*: unir los puntos por semi-rectas.
- *Concepción local*: trazar una recta por un par de puntos o un subconjunto de puntos.
- *Cercanía*: Trazar una recta cerca y en medio de todos los puntos.

(xvi) Códigos de los ítems cuyo contenido de estadística es predecir:

- *Establecer una estrategia para predecir a partir de la situación problema.*

Código	Ejemplo de respuesta
Concepción funcional: regla de tres	<i>37 mil dólares ----- \rightarrow 21 Satisfacción 39 mil dólares----- \rightarrow22.13 $39*21=819/37=22.13$</i>
Patrón aritmético	<i>23 porque aumenta 2, ya que encontré la relación 3, 2, 3 y en secuencia toca 2 y a 21 se suman 2.</i>
Promedio	<i>Aproximadamente tendría una satisfacción de 25 porque en 33 es 20 y en 37 es 21 y con esos datos puede calcularse un promedio para 39.</i>
Comportamiento ascendente aritmético	<i>Según las estadísticas de la tabla al ganar 39 mil dólares el nivel de satisfacción ronda el 23.</i>

Como tiene que crecer tiene que ser más de 27, tiende a crecer y es proporcional (realiza una tabla ordenando los datos y escribe (39,24)

Los anteriores códigos muestran los razonamientos respecto al contenido predecir, estos pueden ser:

- *Concepción funcional:* Asumir proporcionalidad y realizar una regla de tres para predecir.
- *Patrón aritmético:* Establecer un patrón aritmético utilizando todos los datos para realizar sumas o diferencias.
- *Comportamiento ascendente aritmético:* Al azar establecer un valor con base en el comportamiento de los datos.

Tabla 1. Frecuencia para cada código del componente álgebra básica

Pregunta	Código	Frecuencia (Parejas)
• <i>Establecer algebraicamente la función lineal que representa el enunciado. Es decir, pasar del enunciado verbal a establecer una expresión algebraica.</i>	Lenguaje algebraico	14
	Reemplazar valores	4
• <i>Dado un valor sustituir en la expresión algebraica y hallar su valor.</i>	Interpretación errónea	3
• <i>Establecer la función lineal que representa la relación a partir de datos tabulares.</i>	Lenguaje algebraico.	10
	Observación individual de variables	11
	Lenguaje algebraico.	9
	Patrón Aritmético	10

• <i>Determinar la función lineal que representa la relación de proporcionalidad y encontrar los valores</i>	No responde	2
• <i>Identificar en una pareja ordenada la abscisa y la ordenada ubicando el punto en el plano cartesiano.</i>	Intercambio entre abscisa y ordenada	4
	Ubicación correcta	17

Tabla 2. Frecuencia para cada código del componente estadística

Pregunta	Código	Frecuencia (Parejas)
• <i>Establecer el diagrama de dispersión que represente una correlación de 0,6.</i>	Comportamiento lineal perfecto positivo o negativo	9
	Poca dispersión conglomeración débil	3
	Comportamiento lineal, pero conglomeración fuerte	6
	Ninguna	3
• <i>Reconocer gráficamente una relación fuerte entre las variables</i>	Linealidad perfecta positiva	4
	Linealidad perfecta negativa	1
	Linealidad perfecta positiva y negativa	8
	Linealidad positiva	4
• <i>Reconocer la relación establecida en forma textual e identifique la figura correspondiente.</i>	Comportamiento lineal	4
	Comportamiento bivariado negativo	21
	Dependencia funcional	9

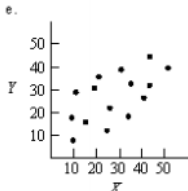
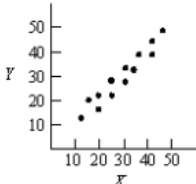
<ul style="list-style-type: none"> • <i>Establecer un criterio para representar la recta de mejor ajuste en un diagrama de dispersión.</i> 	Concepción local	9
	Medio de todos los puntos (cercanía)	2
<ul style="list-style-type: none"> • <i>Establecer una estrategia para predecir a partir de la situación problema.</i> 	Concepción funcional: regla de tres	1
	Patrón aritmético	4
	Promedio	2
	Comportamiento ascendente aritmético	14

Codificación Inicial del cuestionario evaluativo

A continuación, se muestran para los ítems que comprenden el post-test los códigos que se generaron y ejemplos de respuestas.

(xvii) Códigos de los ítems cuyo contenido es:

- *Establecer el diagrama de dispersión que represente una correlación de 0,6.*

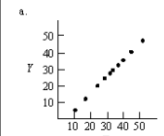
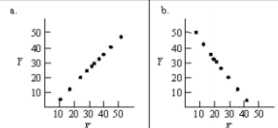
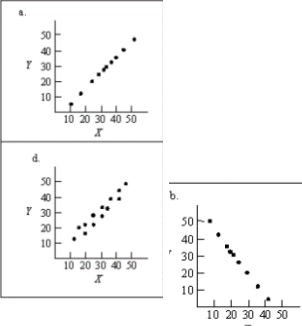
Código	Ejemplo de respuesta
	<i>Una correlación que muestre cerca de 0.60</i>
Poca dispersión conglomeración débil	 <p>A scatter plot with X and Y axes ranging from 0 to 50. The data points are widely scattered but show a slight upward trend, indicating a weak positive correlation.</p>
	<i>Una correlación que muestre cerca de 0.60</i>
Comportamiento lineal, pero conglomeración fuerte	 <p>A scatter plot with X and Y axes ranging from 0 to 50. The data points are tightly clustered along a clear upward linear path, indicating a strong positive correlation.</p>

Los anteriores códigos muestran los razonamientos respecto al contenido de fuerza y dirección en diagramas de dispersión, estos pueden ser:

- *Poca dispersión conglomeración débil:* correlación 0.6 no muestra una relación lineal perfecta pero el valor por encima de 0.5 muestra una nube de puntos con fuerza débil.
- *Comportamiento lineal, pero conglomeración fuerte:* correlación de 0.6 muestra un comportamiento lineal, pero con una nube de puntos con fuerza fuerte.

(xviii) Códigos de los ítems cuyo contenido de estadística es:

- *Reconocer gráficamente una relación fuerte entre las variables*

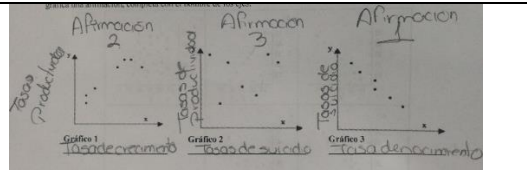
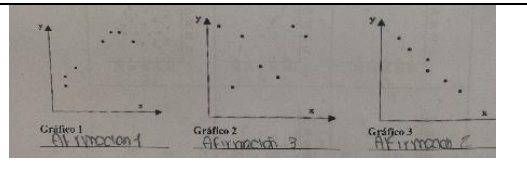
Código	Ejemplo de respuesta
Linealidad perfecta positiva	<p data-bbox="641 359 828 390"><i>Fuerza implica:</i></p> 
Linealidad perfecta positiva y negativa	<p data-bbox="641 548 828 579"><i>Fuerza implica:</i></p> 
Comportamiento lineal	<p data-bbox="641 722 828 753"><i>Fuerza implica:</i></p> 

Los anteriores códigos muestran los razonamientos respecto al contenido de diagramas de dispersión y correlación, estos pueden ser:

- *Linealidad perfecta positiva:* Fuerza implica una nube de puntos cuyo comportamiento es perfectamente lineal en sentido positivo.
- *Linealidad perfecta positiva y negativa:* Fuerza implica ambos comportamientos perfectamente lineales de sentido positivo y negativo.
- *Comportamiento lineal:* Fuerza implica un comportamiento lineal en cualquier sentido.

(xix) Códigos de los ítems cuyo contenido de estadística es:

- *Establecer la relación a partir de diagramas de dispersión y lenguaje textual e identificar la variable dependiente e independiente.*

Código	Ejemplo de respuesta
Identificación correcta del lenguaje verbal de datos bivariados al lenguaje gráfico.	
Confusión comportamiento ascendente con el descendente.	

- *Identificación correcta del lenguaje verbal de datos bivariado al lenguaje gráfico:* se establece de manera adecuada la relación para cada afirmación.
- *Confusión comportamiento ascendente con el descendente:* Se confunde en el lenguaje textual el comportamiento en sentido positivo con el comportamiento en sentido negativo.

(xx) Códigos de los ítems cuyo contenido de estadística es recta de mejor ajuste:

- *Establecer un criterio para representar la recta de mejor ajuste en un diagrama de dispersión.*

Código	Ejemplo de respuesta
Cercanía	<i>Pensé que la línea debe pasar por los puntos, pero como no se puede pues hice una línea donde todos los puntos estén cerca de ella.</i>
Noción distancia (nube, puntos)	<i>Que todos los puntos estuvieran lo más próximo de la línea posible sin crear una gran distancia.</i>
El comportamiento de los datos	<i>Solo dibujé una línea de extremo a extremo en la gráfica de esta forma se vería más claro la forma en que están los puntos.</i>

Pequeña correlación corte con el eje Y *Trace la línea de esa manera ya que el valor mínimo del eje Y es 12 y al trazar la línea se puede ver que toca tres puntos en total lo cual podría significar una pequeña correlación entre dichos puntos.*

Los anteriores códigos muestran los razonamientos respecto al contenido recta de mejor ajuste, estos pueden ser:

- *Cercanía:* Trazar una recta cerca y en medio de todos los puntos.
- *Noción distancia (nube, puntos):* Se tiene en cuenta las distancias de los puntos a la recta para trazarla.
- *El comportamiento de los datos:* trazar una recta que represente el comportamiento de los puntos.
- *Pequeña correlación y corte con el eje Y:* Establece un punto de corte con el eje Y , eligiendo también tres puntos colineales para trazar la recta.

(xxi) Códigos de los ítems cuyo contenido de estadística es predecir:

- *Establecer una estrategia para predecir a partir de la situación problema.*

Código	Ejemplo de respuesta
	<i>El nivel de satisfacción será aproximadamente de 22.13</i>
Concepción funcional: regla de tres	$37 \rightarrow 21$ $39 \rightarrow x$
	$39 * 21 = 819$
Utilizar la línea para predecir	<i>Nivel de satisfacción=24 es la proximidad que se acerca a la línea que trace con anterioridad.</i> <i>Nivel de satisfacción igual a 25, realizando la gráfica ubique en la línea de cuadrados donde se interceptaban a $x=39$</i>
Comportamiento ascendente	<i>Por obvias razones su nivel de satisfacción sería más alto del promedio al ganar más.</i>

Aproximadamente tendría una satisfacción de 24 si hacemos un cálculo comparando los datos de la tabla

Los anteriores códigos muestran los razonamientos respecto al contenido predecir, estos pueden ser:

- *Concepción funcional:* Asumir proporcionalidad y realizar una regla de tres para predecir.
- *Utilizar la línea para predecir:* Ubicar la intersección de la línea vertical que pasa por el valor de x y la recta que traza de mejor ajuste.
- *Comportamiento ascendente aritmético:* Al azar establecer un valor con base en el comportamiento de los datos.

Tabla 1. Frecuencia para cada código del componente álgebra básica

Pregunta	Código	Frecuencia (Parejas)
<ul style="list-style-type: none"> • <i>Establecer algebraicamente la función lineal que representa el enunciado. Es decir, pasar del enunciado verbal a establecer una expresión algebraica.</i> 	Lenguaje algebraico	14
	Reemplazar valores	4
	Interpretación errónea	3
<ul style="list-style-type: none"> • <i>Dado un valor sustituir en la expresión algebraica y hallar su valor.</i> 	Lenguaje algebraico.	10
	Observación individual de variables	11
<ul style="list-style-type: none"> • <i>Determinar la función lineal que representa la relación de proporcionalidad y encontrar los valores</i> 	Lenguaje algebraico.	9
	Patrón Aritmético	10
	No responde	2
<ul style="list-style-type: none"> • <i>Identificar en una pareja ordenada la abscisa y la ordenada ubicando el punto en el plano cartesiano.</i> 	Intercambio entre abscisa y ordenada	4
	Ubicación correcta	17

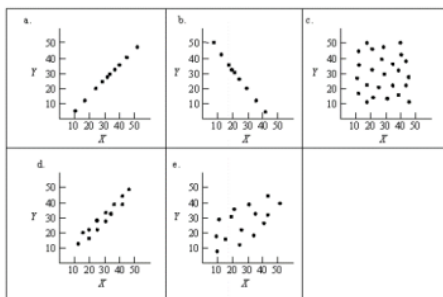
Tabla 2. Frecuencia para cada código del componente estadística del cuestionario diagnóstico

Pregunta	Código	Frecuencia (Parejas)
<ul style="list-style-type: none"> • Establecer el diagrama de dispersión que represente una correlación de 0,6. 	Comportamiento lineal perfecto positivo o negativo	9
	Poca dispersión conglomeración débil	3
	Comportamiento lineal, pero conglomeración fuerte	6
	Ninguna	3
<ul style="list-style-type: none"> • Reconocer gráficamente una relación fuerte entre las variables 	Linealidad perfecta positiva	4
	Linealidad perfecta negativa	1
	Linealidad perfecta positiva y negativa	8
	Linealidad positiva	4
	Comportamiento lineal	4
<ul style="list-style-type: none"> • Reconocer la relación establecida en forma textual e identifique la figura correspondiente. 	Comportamiento bivariado negativo	21
<ul style="list-style-type: none"> • Establecer un criterio para representar la recta de mejor ajuste en un diagrama de dispersión. 	Dependencia funcional	9
	Concepción local	9
	Medio de todos los puntos (cercanía)	2
<ul style="list-style-type: none"> • Establecer una estrategia para predecir a partir de la situación problema. 	Concepción funcional: regla de tres	1
	Patrón aritmético	4
	Promedio	2
	Comportamiento ascendente aritmético	14

Tabla 2. Frecuencia para cada código del componente estadística del post-test

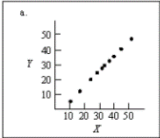
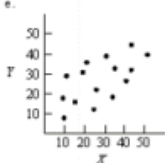
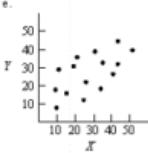
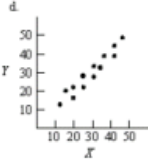
Pregunta	Código	Frecuencia (Parejas)
• <i>Establecer el diagrama de dispersión que represente una correlación de 0,6.</i>	Poca dispersión conglomeración débil	3
	Comportamiento lineal, pero conglomeración fuerte	4
• <i>Reconocer gráficamente una relación fuerte entre las variables</i>	Linealidad perfecta positiva	1
	Linealidad perfecta positiva y negativa	5
	Comportamiento lineal	1
• <i>Establecer la relación a partir de diagramas de dispersión y lenguaje textual e identificar la variable dependiente e independiente.</i>	Identificación correcta del lenguaje verbal de datos bivariados al lenguaje gráfico.	5
	Confusión comportamiento ascendente con el descendente.	2
• <i>Establecer un criterio para representar la recta de mejor ajuste en un diagrama de dispersión.</i>	Cercanía	3
	Noción distancia (nube, puntos)	2
	El comportamiento de los datos	1
	Pequeña correlación corte con el eje Y	1
• <i>Establecer una estrategia para predecir a partir de la situación problema.</i>	Concepción funcional: regla de tres	2
	Utilizar la línea para predecir	3
	Comportamiento ascendente	2

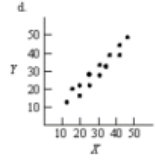
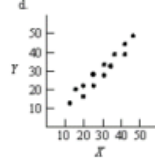
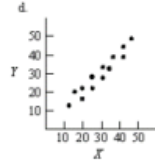
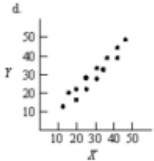
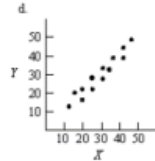
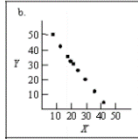
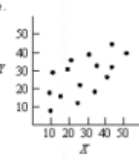
Ítem 1. Para los ítems i. y ii.



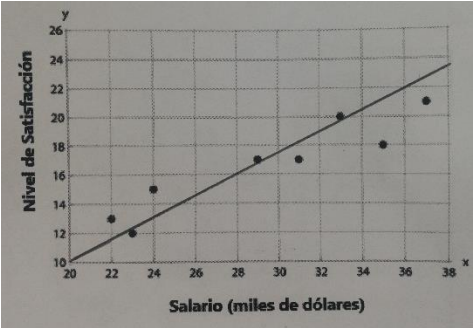
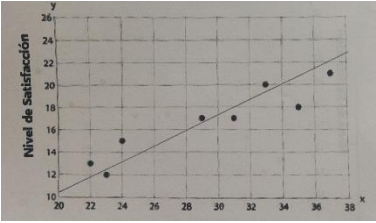
Seleccione el diagrama de dispersión que muestre:

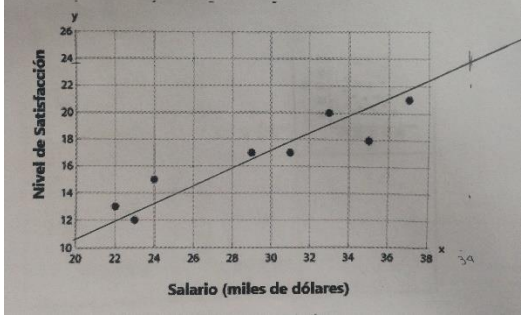
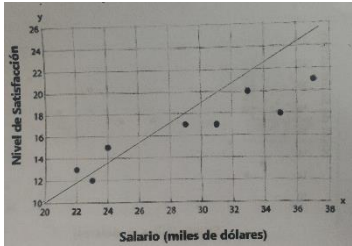
i. Una correlación que muestre cerca de 0.60

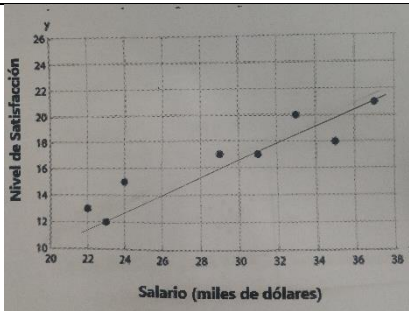
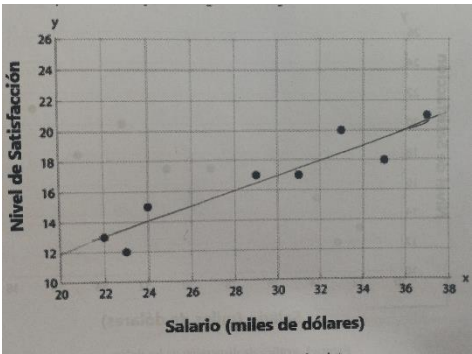
Estudiante	Cuestionario diagnóstico	Cuestionario Evaluativo
A1	 <p>Comportamiento lineal perfecto positivo</p>	 <p>Poca dispersión conglomeración débil</p>
C1		 <p>Comportamiento lineal, pero conglomeración fuerte</p>

C2	Ninguna	
D1	Ninguna	
D2	Ninguna	
H2		
K2	 <p data-bbox="467 1402 792 1495">Comportamiento lineal perfecto negativo</p>	 <p data-bbox="880 1264 1399 1306">Poca dispersión conglomeración débil</p>

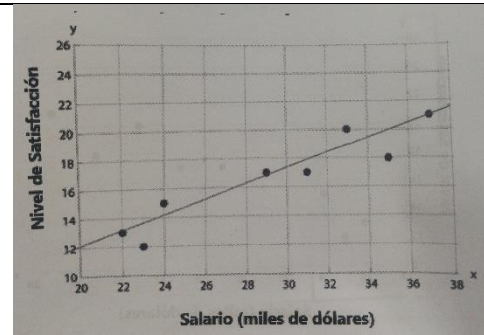
Dibuje una línea en el gráfico de dispersión que mejor se ajuste a los datos. ¿Qué criterios utilizó para determinar la línea? Describa lo que estaba pensando cuando escogió una línea que se ajusta a los datos.

Estudiante	Cuestionario diagnóstico	Cuestionario Evaluativo
A1	<p>Dependencia Funcional</p> <p>Traza semi-rectas</p>	<p>Cercanía</p>  <p>Pensé que la línea debe pasar por los puntos, pero como no se puede pues hice una línea donde todos los puntos estén cerca de ella.</p>
C1	<p>Concepción local</p> <p>Traza semi-rectas con algunos puntos</p>	<p>Cercanía</p>  <p>Debe ser una línea en la cual se pudieran observar la mayoría de los datos de la gráfica o que pasara cerca de cada uno de los puntos</p>
C2	<p>Dependencia Funcional</p> <p>Traza semi-rectas</p>	<p>Distancia (nube, recta)</p>

		 <p>Que todos los puntos estuvieran lo más próximo de la línea posible sin crear una gran distancia.</p>
<p>D1</p>	<p>Dependencia Funcional</p> <p>Traza semi-rectas</p>	<p>Comportamiento de los datos</p>  <p>Solo dibujé una línea de extremo a extremo en la gráfica de esta forma se vería más calor la forma en que están los puntos.</p>
<p>D2</p>	<p>Dependencia Funcional</p>	<p>Distancia (nube, recta)</p>

	Traza semi-rectas	 <p>Pensé en que el valor de los residuos diera la menor cifra posible, a pesar de que no sabía las distancias de los puntos a la línea, creí que, al pasar por 3 puntos, los que daría cero disminuiría la cifra total al realizar la suma.</p>
H2	<p>Cercanía</p> <p>Traza una línea en medio de todos los puntos con dirección positiva y dice: Escogí esa línea ya que, aunque no coincide con algunos puntos se acerca y se puede tener resultados con una mejor relación.</p>	<p>Cercanía</p>  <p>Escogí la línea que pensé quedaba mejor centrada entre los puntos y la verifiqué en codap</p>
K2	<p>Concepción local</p> <p>Une algunos puntos. Observar los datos y notar que no tiene un</p>	<p>Corte con el eje Y y puntos colineales</p>

orden ascendente, en algunos casos disminuye el nivel de satisfacción y en otros aumenta



Trace la línea de esa manera ya que el valor mínimo del eje Y es 12 y al trazar la línea se puede ver que toca tres puntos en total lo cual podría significar una pequeña correlación entre dichos puntos.

Prediga el nivel de satisfacción de un individuo que gana 39 mil dólares.		
Estudiante	Cuestionario diagnóstico	Cuestionario evaluativo
A1	<p>Patrón aritmético (35,18); (37,21); (39,24)</p>	<p>Regla de tres</p> <p>El nivel de satisfacción será aproximadamente de 22.13</p> <p>$37 \rightarrow 21$</p> <p>$39 \rightarrow x$</p> <p>$39 * 21 = 819$</p>
C1	<p>Promedio para predecir</p> <p>Aproximadamente tendría una satisfacción de 25 porque en 33 es 20 y en 37 es 21 y con esos datos puede calcularse un promedio para 39.</p>	<p>Comportamiento ascendente</p> <p>Aproximadamente tendría una satisfacción de 24 si hacemos un cálculo comparando los datos de la tabla.</p>
C2	<p>Comportamiento ascendente</p> <p>Según las estadísticas de la tabla al ganar 39 mil dólares el nivel de satisfacción ronda el 23</p>	<p>Utiliza la línea trazada para predecir</p> <p>Nivel de satisfacción=24 es la proximidad que se acerca a la línea que trace con anterioridad.</p>
D1	<p>Patrón aritmético</p> <p>22 23 24 29 31 33 35 37 39</p> <p>12 12 15 17 17 20 18 21 ¿?</p>	<p>Utiliza la línea trazada para predecir</p> <p>24 El punto se acerca más a este dato, respecto a la línea que trace.</p>

	<p>-1 +3 +2 0 +3 -2 +3 (diferencias entre los valores de y)</p> <p>No sé cómo resolverlo</p>	
D2	<p style="text-align: right;">Regla de tres</p> <p>37 mil dólares -----→ 21 Satisfacción</p> <p>39 mil dólares-----→22.13</p> <p style="text-align: right;">$39*21=819/37=22.13$</p>	<p style="text-align: center;">Comportamiento ascendente</p> <p>No podría por que la al ser individuos distintos no sabría la respuesta de la persona, sin embargo, por obvias razones su nivel de satisfacción sería más alto del promedio al ganar más.</p>
H2	<p style="text-align: center;">Comportamiento ascendente</p> <p>dividí la cantidad que ganan sobre la satisfacción, después promedié esos números. Luego dividí los 39 millones sobre el promedio el cual fue 1.75 me dio como resultados</p> <p style="text-align: right;">22.28</p>	<p style="text-align: center;">Utiliza la línea trazada para predecir</p> <p>Nivel de satisfacción igual a 25, realizando la gráfica ubique en la línea de cuadrados donde se interceptaban a $x=39$</p>
K2	<p style="text-align: center;">Comportamiento ascendente</p> <p>24 ya que no sería demasiada la diferencias en la relación con el caso $(35,18) - (37,21) = (2,-3) = (37,21) + (2,+3) = (39,24)$</p>	<p style="text-align: center;">Regla de tres</p> <p>22 hice una regla de tres a partir del valor más cercano a 39 en este caso sería el último dato (37) y hacer la operación para encontrar el resultado.</p>