

Análisis de datos bivariados en un ambiente basado en applets y software dinámico

Bivariate Data Analysis in an Environment Based on the use of Applets and Dynamic Software

Santiago Inzunza Cazares¹

Resumen: Se reportan resultados de una investigación sobre análisis de datos bivariados realizada con 34 estudiantes universitarios mientras tomaban un curso introductorio de estadística en el cual se utilizó un ambiente computacional integrado por applets y software dinámico para análisis de datos. En la evaluación diagnóstica los estudiantes exhibieron bajo nivel de razonamiento intuitivo sobre la covariación en diagramas de dispersión, pero en la actividad final, después de haber utilizado el ambiente computacional, lograron identificar con éxito la dirección de la relación entre dos variables y desarrollaron un buen sentido de la intensidad de la relación y el coeficiente de correlación; sin embargo, la interpretación de los coeficientes de la ecuación de regresión y el coeficiente de determinación resultó ser una tarea complicada para ellos. Las estrategias utilizadas en el análisis de datos privilegiaron el uso de representaciones gráficas y visuales sobre representaciones simbólicas de cálculo, como resultado del potencial de visualización dinámica de las herramientas tecnológicas utilizadas.

Palabras clave: Datos bivariados, razonamiento covariacional, tecnologías digitales.

Fecha de recepción: 8 de septiembre de 2015. **Fecha de aprobación:** 15 de mayo de 2016.

¹ Facultad de Informática. Universidad Autónoma de Sinaloa. México. sinzunza@uas.edu.mx

Abstract: Results are reported of a study on bivariate data analysis, with 34 university students, while taking an introductory statistics course, in which a computer environment integrated by applets and dynamic software data analysis were used, are reported. In the diagnostic evaluation, the students exhibited low level of intuitive reasoning about the covariation in scatter plots, but in the final activity, after using the computer environment, they successfully identified the direction of the relationship between the variables and developed a good sense of the strength of the relationship and correlation coefficient. However, the interpretation of the coefficients of the regression equation and the coefficient of determination were a complicated task for the students. The strategies used in the data analysis privileged the use of graphical and visual representations, instead of symbolic representations of calculation, as result of the dynamic visualization power of the technological tools used.

Keywords: *Bivariate data, covariational reasoning, digital technologies.*

INTRODUCCIÓN

El estudio de la correlación y la regresión lineal se ha vuelto parte importante en el currículo de estadística de los niveles de bachillerato y universitario, debido a que están implicadas en los análisis de relación, construcción de modelos y predicción de variables de muchos fenómenos. En el caso de dos variables cuantitativas –caso al que hacemos referencia en este estudio–, en un nivel básico, la relación entre dos variables es representada y evaluada comúnmente mediante un diagrama de dispersión, el cálculo del coeficiente de correlación y la recta de regresión con medidas de bondad de ajuste, como es el caso del coeficiente de determinación.

La covariación entre variables es condición necesaria, más no es suficiente para inferir relaciones causales, sin embargo representa un buen punto de partida para hacer inferencias inductivas causales en la ciencia (Batanero, Estepa y Godino, 1997); de ahí que el razonamiento covariacional es de suma importancia en el razonamiento científico. En la vida cotidiana, la covariación también aparece con frecuencia en los medios de comunicación, por lo general a través de descripciones verbales y datos representados en forma tabular, con dos o más variables en forma simultánea, por lo cual el tema no es sólo de interés de los científicos y profesionistas, sino que forma parte del bagaje de cultura estadística que deben tener todos los ciudadanos.

Una comprensión profunda de la correlación y la regresión lineal requiere del dominio de diversas ideas y conceptos, tales como la estructura de una relación bivariada (dirección, intensidad y forma de la relación entre las variables); tiene también interés en el aprendizaje de la construcción de modelos y en la predicción de variables. En este sentido, las nociones de correlación y regresión lineal resultan de gran ayuda en cuanto a discernir si dos variables se relacionan linealmente, con qué intensidad y en qué sentido lo hacen; y, cuando es posible, construir un modelo para predecir una variable con base en la información de otras variables. Watkins *et al.* (2004) describen las siguientes características de una relación en datos bivariados que los estudiantes requieren aprender para comprender la correlación y la regresión:

- Los valores individuales y su variabilidad.
- La forma de la tendencia de los datos en una gráfica, en términos de linealidad y datos extremos.
- Identificar si la tendencia es positiva, negativa o no existe.
- La intensidad de la tendencia: fuerte o débil, variable o constante.
- Si hay explicaciones plausibles para la tendencia. Si una variable puede tener relaciones causales, o si variables espurias causan la relación.

En algunos currículos, la noción de covariación es introducida desde los niveles elemental y medio a un nivel intuitivo, y con la idea de desarrollar desde etapas tempranas el razonamiento covariacional (por ejemplo: NCTM, 2000). Usualmente en los libros de texto y algunos currículos de bachillerato –e incluso en cursos básicos universitarios– se introduce como una extensión del análisis univariado de datos, a través de la construcción e interpretación de diagramas de dispersión, cálculo del coeficiente de correlación y la ecuación de recta de regresión con el método de mínimos cuadrados de error. A un mayor nivel de formalización, en los cursos de estadística del nivel universitario y posgrado que van más allá de lo básico, la correlación y regresión lineal se tratan después del tema de inferencia estadística, para estar en condiciones de realizar pruebas de significancia sobre la correlación y los coeficientes de la ecuación de regresión.

En México estos temas aparecen en los programas de estudio del último grado del bachillerato (15-18 años), en lo que se denomina fase propedéutica u optativa. En la más reciente reforma del bachillerato, el estudio de la covariación se reduce al tema de correlación y se deja de lado el de la regresión y la predicción (DGB, 2009). Se considera el caso de la asociación entre dos variables

cualitativas mediante tablas de continencia, y la relación entre dos variables cuantitativas a través de diagramas de dispersión, así como el cálculo del coeficiente de correlación.

La tecnología computacional posee gran potencial de cálculo para el procesamiento de las grandes cantidades de datos que tradicionalmente ha caracterizado a la estadística (Biehler *et al.*, 2013); de tal forma, el cálculo del coeficiente de correlación y la determinación de la línea de regresión con su bondad de ajuste –dos de los procedimientos más laboriosos en el análisis de datos bivariados– pueden ser elaborados con gran precisión de forma inmediata, evitando el gran esfuerzo de cálculo que se requiere en un ambiente de lápiz y papel.

Muchas tecnologías estadísticas o incluso de uso general, en particular las educativas, además del poder de cálculo descrito anteriormente, disponen de un amplio repertorio representacional (por ejemplo, gráficas, tablas, fórmulas) que permiten ver propiedades y comportamientos diversos que se integran y complementan en un análisis de datos. Además de ello, ofrecen una interactividad que genera retroalimentación inmediata ante cualquier cambio generado por el usuario en alguno de los datos, parámetros o representaciones.

La introducción de este tipo de tecnología en la clase de estadística requiere de investigación, para conocer su impacto en el aprendizaje y desarrollo del razonamiento estadístico. En el presente trabajo nos hemos propuesto explorar el potencial de un ambiente computacional que considera software ejecutable en línea (applets) y el software para análisis dinámico de datos Fathom (Finzer, Erickson y Binker, 2002). En específico, nos hemos planteado las siguientes preguntas: ¿qué estrategias y recursos de un ambiente computacional ponen en juego los estudiantes para dar sentido a la información de datos bivariados?, ¿qué elementos influyen en el desarrollo del razonamiento covariacional en un ambiente computacional con representaciones dinámicas de los datos?

ANTECEDENTES

La literatura reporta estudios de investigación sobre el razonamiento y conceptos que se involucran en el análisis de datos bivariados, con estudiantes de nivel elemental hasta universitario, con profesores y futuros profesores. En una síntesis de las investigaciones más relevantes sobre las nociones de correlación y regresión realizada por Gea (2014) se identifican diferentes líneas de trabajo desde perspectivas de la psicología y la didáctica de la estadística, tales como: estrategias

de estimación de la correlación a partir de diagramas de dispersión, sesgos y concepciones sobre la correlación, la correlación y sus representaciones, y desarrollo de razonamiento correlacional a través de la enseñanza.

En este contexto destacan los trabajos de investigadores españoles, quienes han estudiado la correlación y regresión de forma consistente desde hace varios años, analizando el significado semiótico que le atribuyen los autores de libros de texto de secundaria, bachillerato y universitarios (por ejemplo: Sánchez Cobo, 1999; Gea, 2014), así como el desarrollo del razonamiento covariacional y su evolución en una enseñanza basada en ambientes computacionales (Estepa, 1993; Cañadas, 2010). En los siguientes párrafos realizaremos un análisis de las investigaciones previas que más relación tienen con nuestro objeto de investigación.

ESTRATEGIAS DE ESTIMACIÓN DE LA CORRELACIÓN A PARTIR DE DIAGRAMAS DE DISPERSIÓN

Estepa y Batanero (1996), en un estudio con estudiantes preuniversitarios, identificaron diversas estrategias que estos utilizan para estimar la correlación a partir de diagramas de dispersión. Entre las estrategias correctas destacan la *comparación global*, la cual ocurre cuando los estudiantes se basan en una comparación global de la relación entre las dos variables; y la estrategia de *crecimiento*, que es utilizada como argumento cuando se observa el crecimiento o decrecimiento de la forma de la nube de puntos.

Entre las estrategias incorrectas identificadas se encuentra el uso de *teorías previas*, que consiste en usar el conocimiento del contexto y no la evidencia de los datos para justificar la correlación; la *interpretación incorrecta de puntos aislados*, la cual se presenta cuando los estudiantes emplean pares aislados de valores para justificar de forma incorrecta la relación entre las variables; y la estrategia de *causalidad*, que consiste en considerar que por el sólo hecho de estar correlacionadas, una variable debe ser causa de la otra.

Por otra parte Sánchez Cobo (1999), en un estudio con estudiantes universitarios sobre estimación de la correlación a partir de diferentes representaciones (descripción verbal, tabla de valores, coeficiente de correlación, diagrama de dispersión), encontró que los estudiantes estiman de forma más precisa el coeficiente de correlación a partir de un diagrama de dispersión, siendo mayor la precisión conforme la correlación entre las variables es más intensa. Otros estudios

citados en Garfield y Ben-Zvi (2008) muestran que las personas tienen dificultad para razonar cuando la relación entre los datos bivariados es negativa (Beyth-Marmor, 1982), y tienden a realizar juicios más precisos sobre la covariación cuando las variables son continuas que cuando son dicotómicas (Jennings *et al*, 1982).

CONCEPCIONES SOBRE LA CORRELACIÓN

En los trabajos de Estepa (1993) y Batanero, Estepa y Godino (1997) se documentan diversas concepciones erróneas sobre la correlación. *Concepción determinista de la correlación* se presenta cuando los estudiantes tienden a asignar un único valor de la variable independiente a cada uno de los valores de la variable dependiente; esto es, la relación de las variables sólo es considerada desde un punto de vista funcional. *Concepción local de la correlación* ocurre cuando los estudiantes se limitan a confirmar la asociación a partir de un subconjunto de datos, sin tener en cuenta la tendencia global de los mismos. *Concepción unidireccional* consiste en no admitir la correlación inversa, considerándose la intensidad de la asociación, pero no su signo. *Concepción causal* sucede si solamente consideran la dependencia entre variables, siempre que pueda adjudicarse a la presencia de una relación causal entre las mismas.

DESARROLLO DE RAZONAMIENTO CORRELACIONAL A TRAVÉS DE LA ENSEÑANZA

En esta línea de investigación sobresale el uso de herramientas computacionales para analizar la comprensión y el desarrollo del razonamiento covariacional de los sujetos de estudio. Uno de los primeros análisis de este tipo fue desarrollado por Stockburger (1982), quien en un curso introductorio con estudiantes universitarios utilizó un programa de computadora para repetir muchas veces la estimación del coeficiente de correlación de cuatro diagramas de dispersión. Al principio los estudiantes mostraron un desempeño muy pobre en la estimación de la correlación, pero su habilidad de estimación mejoró notablemente después de usar el programa de computadora.

Batanero, Estepa y Godino (1997), en una investigación con estudiantes universitarios, analizaron los cambios conceptuales y estrategias para elaborar juicios sobre la correlación, a partir de diagramas de dispersión en un experimento de enseñanza con computadoras. Observaron un notable incremento de

estrategias correctas en muchos estudiantes: la mayoría logró superar la concepción determinista de la correlación y la concepción local fue eliminada, al comprender la importancia de tener en cuenta los datos completos para el análisis. Sin embargo, en general no observaron mejoría respecto a la concepción causal de la correlación; asimismo, encontraron dificultades en los estudiantes al distinguir los roles de la variable dependiente e independiente, y para razonar con relaciones negativas.

En la misma línea sobre uso de tecnología computacional para ayudar a los estudiantes a razonar sobre la covariación, los resultados encontrados por Gravemeijer (2000) señalan que estos necesitan desarrollar la idea de tendencia global de los datos y, como alternativa didáctica en la introducción del tema de datos bivariados, sugiere que los estudiantes examinen y comparen varios conjuntos de datos univariados. Por otra parte Konold (2002), utilizando el ambiente computacional que proporciona TinkerPlots (Konold y Miller, 2005), elaboró un análisis con estudiantes de nivel medio y encontró que podían hacer mejores juicios covariacionales utilizando características del software como es la superposición de un gradiente de colores en los diagramas de dispersión, posiblemente debido a que las relaciones entre variables fueron exploradas por el uso de una dimensión (por ejemplo, el eje horizontal) para una variable y el gradiente de color para la otra.

Con estudiantes universitarios, Zieffler (2006) llevó a cabo un estudio longitudinal durante un semestre para analizar el efecto de dos secuencias de enseñanza con apoyo de tecnología, y con diferente orden en el tema de datos bivariados (antes y después de la inferencia estadística) en el desarrollo del razonamiento sobre covariación. En general se observó un crecimiento en el razonamiento de los estudiantes sobre datos bivariados, pero en forma diferenciada, el mayor cambio en el desarrollo del razonamiento ocurrió entre la primera y segunda evaluación de las cuatro efectuadas a lo largo del semestre, justo antes de que los datos bivariados fueran enseñados por métodos formales.

Fitzallen (2012), en una investigación con 12 estudiantes de primaria y secundaria, utilizó un experimento de enseñanza que constaba de una secuencia de actividades para analizar la intersección entre el pensamiento y el razonamiento de los estudiantes acerca de la covariación, y la influencia del software TinkerPlots en el proceso que los estudiantes usan para explorar conjuntos de datos, determinar relaciones entre variables e identificar tendencias mediante recursos gráficos como el diagrama de dispersión. Por medio de entrevistas individuales con cada uno de los sujetos de estudio observó que el proceso de

enseñanza experimentado, basado en TinkerPlots, proporciona un poderoso ambiente de aprendizaje para apoyar a los estudiantes en la comprensión de la covariación desde niveles tempranos.

En el caso de México no hemos encontrado antecedentes de estudios sobre los conceptos de correlación y regresión lineal, por lo que se vuelve pertinente analizar las dificultades de aprendizaje que entrañan su proceso de enseñanza y aprendizaje, y el potencial de ambientes computacionales para su enseñanza, sobre todo ante la importancia que están teniendo estos temas en las recientes reformas curriculares.

MARCO CONCEPTUAL

EL PAPEL DE LA TECNOLOGÍA EN LA ENSEÑANZA DE LA ESTADÍSTICA Y LA PROBABILIDAD

Biehler *et al.* (2013) identifican diversas categorías de herramientas tecnológicas que se emplean en la enseñanza de la estadística y la probabilidad, muchas de las cuales pueden ser utilizadas en forma individual o complementaria. Entre estas herramientas cabe destacar las siguientes: paquetes de software estadístico, software educativo, hojas de cálculo, applets, calculadoras graficadoras programables, repositorios de datos y materiales multimedia.

Con el amplio repertorio de representaciones y las características dinámicas, la interactividad y comunicación del cual algunas de estas herramientas computacionales disponen, la tecnología puede ayudar al aprendizaje de la estadística y la probabilidad bajo diversas perspectivas. Desde la perspectiva procedimental –aspecto intrínseco a la estadística y la probabilidad–, la tecnología representa un recurso con *potencial amplificador* de la capacidad de cálculo y la construcción de gráficas.

Desde una perspectiva didáctica, la tecnología constituye un recurso con *potencial reorganizador* de la mente de los usuarios, con capacidad para provocar cambios estructurales en el sistema cognitivo a través de la organización y transformación de las diversas representaciones de los datos, y la visualización de patrones que emergen al efectuar cambios en algunas de las componentes de dichas representaciones o en los datos mismos, lo cual le otorga a la tecnología computacional un estatus de herramienta cognitiva en el sentido establecido por Pea (1987: 91), quien la define como “cualquier medio que ayuda a trascender

las limitaciones de la mente, en el pensamiento, el aprendizaje y las actividades de resolución de problemas". De esta manera, algunas herramientas computacionales constituyen una potente herramienta cognitiva para aprender a razonar y pensar estadísticamente.

Con el propósito de ejemplificar lo anterior en el contexto de la enseñanza de la correlación y la regresión lineal, cuando los estudiantes utilizan una herramienta computacional como el software Fathom, tienen a su alcance una perspectiva multivariada de los datos (ver figura 1) y pueden probar la relación entre diversos pares de variables utilizando representaciones como el diagrama de dispersión y el cálculo del coeficiente de correlación. El análisis puede ser extendido superponiendo la recta de regresión sobre la nube de puntos, con su correspondiente ecuación y el coeficiente de determinación.

Es decir, se puede disponer en una misma pantalla de diversas representaciones ligadas dinámicamente de los conceptos que se involucran en los análisis de relación entre dos variables. El potencial amplificador se expresa mediante las representaciones simbólicas que permiten el cálculo de medidas descriptivas (por ejemplo, el coeficiente de correlación y la ecuación de regresión), y las

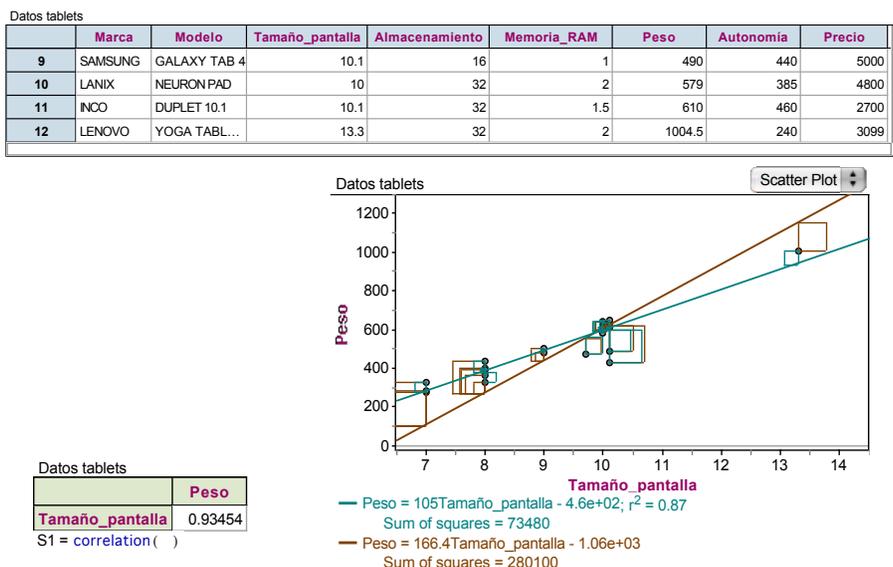


Figura 1. Coeficiente de correlación, diagrama de dispersión con línea de regresión, línea movible, coeficiente de determinación y suma de cuadrados de error.

representaciones numéricas y gráficas que ayudan a visualizar los comportamientos y tendencias en los datos.

El potencial reorganizador lo proporciona la liga entre todas las representaciones construidas sobre la misma pantalla, que las convierte en representaciones dinámicas ejecutables. Por ejemplo, las variables se pueden intercambiar en los ejes del diagrama realizando una operación de arrastre en forma directa, con lo cual se visualizan los cambios en la nube de puntos y en los coeficientes de la ecuación de regresión; con ello, los estudiantes pueden visualizar la existencia de dos ecuaciones de regresión y la importancia de distinguir entre la variable de respuesta y la variable explicativa, para calcular la ecuación correcta.

En la búsqueda para justificar la recta de regresión como recta de mejor ajuste de los datos, se puede insertar una recta móvil sobre la nube de puntos, la cual puede ser manipulada por el usuario y observar que el valor de la suma de cuadrados de error siempre será mayor a los que proporciona la recta de regresión; asimismo, realizando la acción de "arrastré" sobre cualquiera de los puntos se ve el efecto en el valor, en todas las representaciones y en los indicadores de bondad de ajuste. Esto es posible incluso en algunos applets disponibles en Internet (por ejemplo: <http://docentes.educacion.navarra.es/msadaall/geogebra/figuras/e3regresion.htm>).

Un ambiente computacional permite al usuario localizar más fácilmente patrones en la relación entre variables, seleccionar el modelo apropiado y verificar sus residuales para la desviación del modelo. Esto constituye un proceso mucho más desafiante de razonamiento estadístico que sólo aplicar el algoritmo de mínimos cuadrados (Biehler *et al.*, 2013: 647). Además, estos conceptos en un curso tradicional requieren de un tratamiento matemático que podría estar fuera del alcance de muchos estudiantes, lo cual representa un beneficio adicional de la tecnología en la educación estadística.

AMBIENTES DE APRENDIZAJE PARA EL RAZONAMIENTO Y EL PENSAMIENTO ESTADÍSTICO

En cuanto al modelo de operación de la tecnología, en el presente trabajo retomamos diversos elementos didácticos con el propósito de crear ambientes de aprendizaje para el razonamiento estadístico, basados en seis principios del diseño instruccional descritos por Cobb y McClain (2004):

1. Se enfoca en el desarrollo de las ideas estadísticas centrales (datos, distribución, variabilidad, centralidad, modelos, covariación, aleatoriedad, muestreo e inferencia), en lugar de un conjunto de herramientas, técnicas y procedimientos de presentación.
2. Utiliza datos reales y motivadores para interesar a los estudiantes en hacer y probar conjeturas.
3. Usa actividades en clase para apoyar el desarrollo del razonamiento de los estudiantes.
4. Integra el uso de las herramientas tecnológicas adecuadas que permitan a los estudiantes probar sus conjeturas, explorar y analizar datos, así como desarrollar su razonamiento estadístico.
5. Promueve un discurso en clase que incluye argumentos estadísticos e intercambios sustentados que se enfoquen en ideas estadísticas significativas.
6. Utiliza el diagnóstico para aprender lo que los estudiantes saben y monitorear el desarrollo de su aprendizaje estadístico, a fin de evaluar los planes de instrucción y su avance.

METODOLOGÍA

Los sujetos de estudio fueron 34 estudiantes (18-19 años) que tomaban un curso introductorio de probabilidad y estadística en el primer grado de la carrera de Informática. Sus antecedentes sobre el tema eran prácticamente nulos, lo cual quedó de manifiesto en un cuestionario diagnóstico aplicado antes de las sesiones de enseñanza. La investigación tuvo lugar en un aula de cómputo con 35 computadoras durante 6 sesiones de una hora. Los estudiantes recién habían visto el análisis univariado de datos, donde utilizaron el software Fathom para el cálculo de medidas descriptivas y construcción de gráficas; es decir, el análisis bivariado fue visto como una extensión del análisis univariado.

En el diseño de las actividades de enseñanza se utilizaron un par de videos y 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), de acuerdo con lo que marca el programa de estudios y siguiendo los principios instruccionales sugeridos por Cobb y McClain (2004). Las actividades consideraban conjuntos de datos reales multivariados

que se tomaron de algunos sitios de Internet (por ejemplo: <http://eleconomista.com.mx/especiales/2013/06/17/ranking-universidades-clase-2013-excelencia>) y se pedía a los estudiantes que exploraran pares de variables que pudieran estar relacionadas, para analizar las características de la relación (forma, dirección, intensidad, datos extremos).

Como punto de partida del análisis se construyeron diagramas de dispersión, y se enfocó la atención de los estudiantes en visualizar sus características; posteriormente se calcularon coeficientes de correlación buscando, conectar su valor con las características del diagrama. Finalmente se introdujo la recta de regresión y su ecuación para ajustar la tendencia de la nube de puntos, junto con el valor del coeficiente de determinación que por omisión proporciona el software. Una de las sesiones fue dedicada a la estimación del coeficiente de correlación a partir de un diagrama de dispersión en la cual se utilizó un applet del repositorio de Allan Rossman y Beth Chance (<http://www.rossmanchance.com/applets/>); en el resto de las sesiones fue utilizado el software Fathom.

Los instrumentos de recopilación de la información fueron un cuestionario para evaluar los antecedentes de los estudiantes sobre el tema, el cual constaba de tres ítems abiertos que abordaban aspectos de la correlación y regresión a partir de la interpretación de diagramas de dispersión; una hoja de trabajo para la actividad de estimación del coeficiente de correlación con el applet y otra para responder una actividad desarrollada con el software Fathom. En el análisis de la información generada, al resolver cada actividad se tuvieron en cuenta el uso de representaciones (gráficas, numéricas, simbólicas), las características de la relación entre variables identificadas visualmente de un diagrama de dispersión, y medidas numéricas como los coeficientes de correlación y de determinación.

Para la interpretación de los diagramas se definieron niveles de razonamiento basados en el modelo SOLO (Structured of Observed Learning Outcomes) desarrollado por Biggs y Collis (1982), el cual ha sido utilizado en otros estudios para definir categorías de desarrollo cognitivo de diversos conceptos estadísticos. En el modelo SOLO (ver tabla 1) los conceptos y procesos empleados por los sujetos para dar respuesta a las preguntas o tareas planteadas se pueden clasificar en un determinado nivel de los cinco que se consideran.

Todos los materiales (diapositivas, videos y documentos con contenido del tema) e instrumentos de evaluación se colocaron en una plataforma educativa que servía de soporte al curso y, en particular, al tema de datos bivariados. Cada actividad y una evaluación que se hizo a través de un mapa conceptual fueron

Tabla 1. Descripción de los niveles del modelo SOLO.

Nivel	Descripción genérica
Nivel Preestructural	Los estudiantes no se enfocan en aspectos relevantes de la tarea que les ha sido planteada. Los conceptos o procesos son utilizados en forma simplista que los conduce a errores. Pueden dejar la tarea sin resolver por falta de entendimiento.
Nivel Uniestructural	Los estudiantes se enfocan en algún aspecto relevante de la tarea planteada o en alguna etapa de la misma, realizan alguna conexión de un concepto o proceso con otro. Son capaces de desarrollar procesos simples.
Nivel Multiestructural	Los estudiantes se enfocan en más de un aspecto relevante de la tarea, pero no logran integrarlos para obtener una solución correcta.
Nivel Relacional	Los estudiantes integran todos los aspectos relevantes de la tarea como un todo coherente con estructura y significado.
Nivel Abstracto Extendido	Los estudiantes cumplen con las exigencias cognitivas del nivel relacional, pero además son capaces de transferir los conceptos y procesos fuera del contexto en que fueron aprendidos.

subidas a la plataforma en las sesiones de clase, para su correspondiente análisis por parte del investigador.

RESULTADOS Y DISCUSIÓN

EL RAZONAMIENTO DE LOS ESTUDIANTES PREVIO AL ESTUDIO DEL ANÁLISIS DE DATOS BIVARIADOS

Antes de iniciar el análisis de datos bivariados, los estudiantes respondieron un cuestionario basado principalmente en la interpretación de diagramas de dispersión, el cual evaluaba sus ideas intuitivas y nivel de razonamiento sobre la relación entre dos variables. El primer ítem requería la interpretación de dos diagramas de dispersión en forma abierta. Uno de ellos mostraba la relación entre la matrícula y la cantidad de profesores en 50 universidades mexicanas, y solicitaba además una predicción sobre la cantidad de profesores para una universidad de 100,000 estudiantes (ver figura 2).

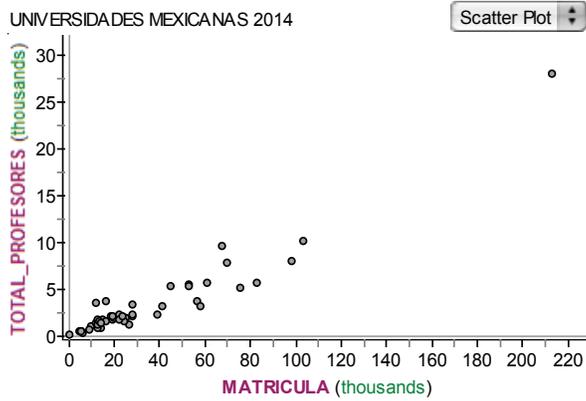


Figura 2. Diagrama de dispersión con datos de matrícula y total de profesores de 50 universidades mexicanas.

El otro diagrama mostraba la relación entre la tasa de mortalidad infantil y la esperanza de vida en 22 países (ver figura 3); además de la interpretación, solicitaba el trazado de una línea de ajuste a la nube de puntos. Una tarea similar a este último ítem ha sido utilizada en trabajos previos de Batanero *et al.* (2014).

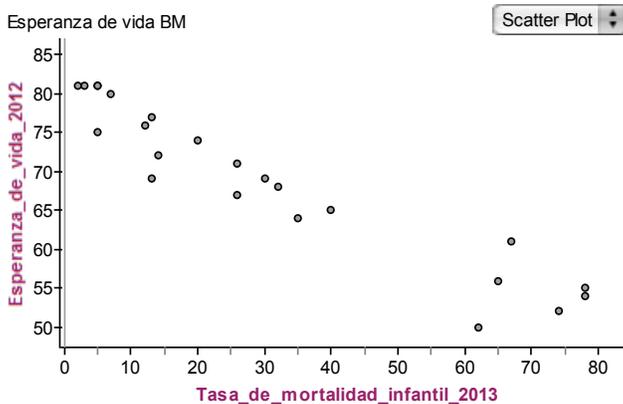


Figura 3. Diagrama de dispersión con datos de tasa de mortalidad infantil y esperanza de vida en 22 países.

En el segundo ítem se mostraban cuatro diagramas de dispersión y se pedía que identificaran el diagrama que presentaba una relación más intensa entre las variables (ver figura 4).

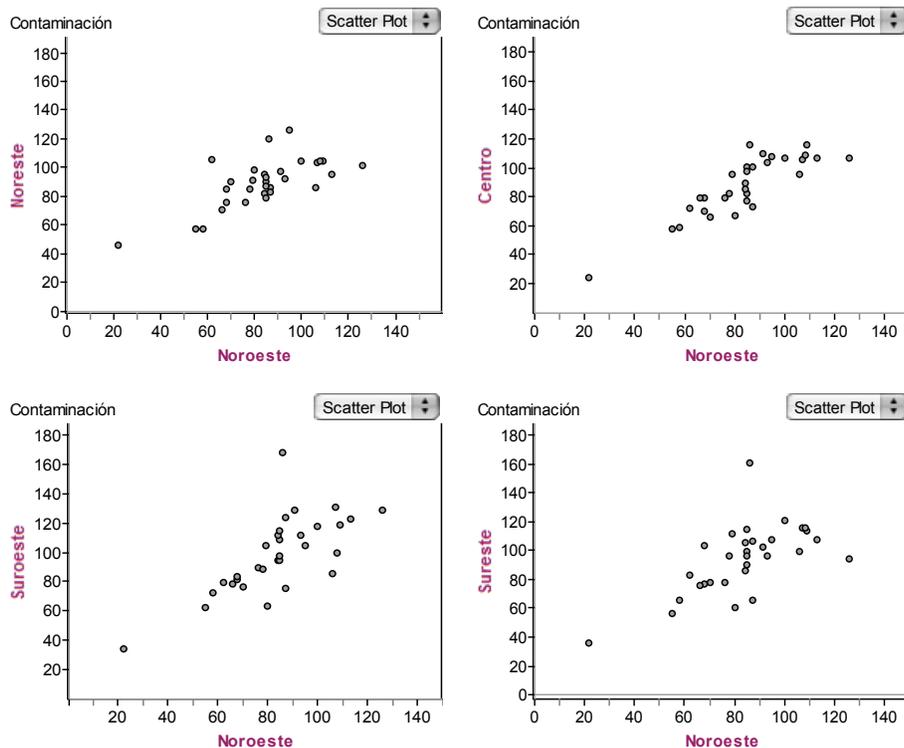


Figura 4. Diagramas de dispersión con diferentes grados de intensidad en la relación de dos variables.

El último ítem constaba de tres diagramas de dispersión que mostraban la relación entre la edad de un grupo de mujeres (eje horizontal) y la densidad de sus huesos (eje vertical). Se pedía a los estudiantes que eligieran el diagrama que mostraba la relación: a mayor edad disminuye la densidad de los huesos (ver figura 5).

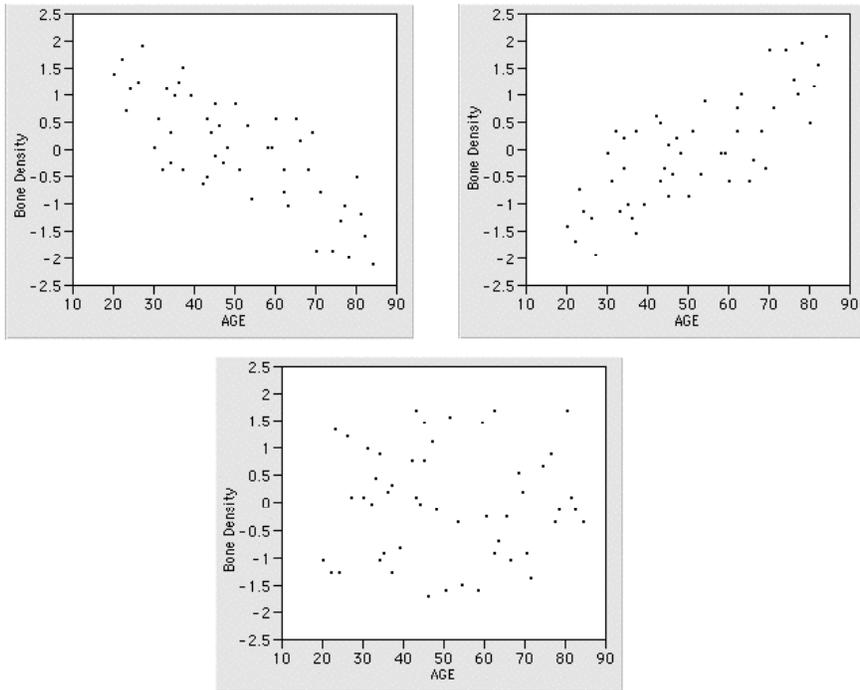


Figura 5. Diagramas de dispersión que mostraban la relación entre la edad de un grupo de mujeres y la densidad de sus huesos.

En el primer ítem sólo 5 de 34 estudiantes identificaron correctamente la tendencia entre las variables. Entre las respuestas correctas destacamos la siguiente: “los datos muestran que a mayor número de alumnos hay mayor número de profesores”. Esta estrategia ha sido denominada estrategia de *crecimiento* en el estudio de Estepa y Batanero (1996). Otro estudiante respondió: “se puede ver que son más las universidades con menor número de profesores y de igual forma con menor número de alumnos”. Esta respuesta revela información correcta de la gráfica, pero no expresa la relación entre la matrícula y la cantidad de profesores.

Respuestas de este tipo fueron comunes, ya que los estudiantes intentaron describir los diagramas de dispersión con base en el conocimiento que tenían del diagrama de puntos de del análisis univariado, sin tener en cuenta que el análisis bivariado requiere poner en coordinación la variación de dos variables en forma simultánea. En cuanto al aspecto predictivo que exploramos en este

ítem, la estimación de la mayor parte de los estudiantes estuvo alejada del valor que proporciona la recta de regresión.

La idea de que los estudiantes dibujaran una línea de regresión sobre la nube de puntos de los diagramas de dispersión (figura 3) fue para ver si identificaban la tendencia de los puntos, y en qué medida se acercaba a la recta de regresión. Un total de 24 estudiantes colocaron la recta conforme la tendencia de los puntos con un buen grado de aproximación a la recta de regresión, los otros 10 la pusieron en otra posición o no respondieron.

En la elección del diagrama de dispersión que mostraba una relación más intensa entre dos variables (ítem 2), se requería que identificaran que a medida de que los puntos tuvieran menor variación y se acomodaran más a una línea recta, habría una mayor intensidad en la relación entre las variables. Sin embargo, sólo 12 estudiantes eligieron la opción correcta, por lo que tampoco estuvieron muy acertados en identificar la intensidad de la relación.

Por último en el ítem 3, donde se solicitaba la identificación del diagrama que mostraba la relación “a mayor edad de las mujeres se tiene menor densidad en los huesos” –es decir, una respuesta basada en la estrategia de crecimiento–, 31 estudiantes identificaron el diagrama correcto, lo cual muestra que el contexto y la forma como se planteó la pregunta (opción múltiple) les ayudó a responder correctamente.

En resumen, los estudiantes tuvieron dificultades para identificar la relación entre dos variables expresadas mediante diagramas de dispersión y para estimar el valor de una variable a partir de otra. La identificación de la intensidad de la relación les resultó mucho más complicada en un ítem abierto que en uno de opción múltiple donde se proporcionaba la interpretación y sólo había que elegirla entre otras interpretaciones plausibles; resultados similares fueron encontrados por Sánchez Cobo (1999). La mayoría mostró buena intuición sobre el lugar que ocupa una recta de regresión en un diagrama de regresión. Sin embargo, consideramos de bajo a muy bajo el nivel de razonamiento intuitivo sobre los conceptos evaluados que se involucran en la correlación y la regresión.

INTERACCIONES CON EL AMBIENTE COMPUTACIONAL

a) Applets

Después de haber visto en clase cómo interpretar un diagrama de dispersión y su relación con los valores y el signo del coeficiente de correlación, en otra sesión

se procedió a desarrollar una actividad en línea con el propósito de fijar ideas sobre estos conceptos, para lo cual se utilizó un applet (ver figura 6) cuya dirección es <http://www.rossmanchance.com/applets/GuessCorrelation.html>. Se pidió a los estudiantes que elaboraran 10 estimaciones para diagramas de dispersión de 25, 50 y 100 puntos, respectivamente. Cada vez que efectuaban una estimación, el applet les proporcionaba un nuevo diagrama de dispersión.

Correlation Guessing Game

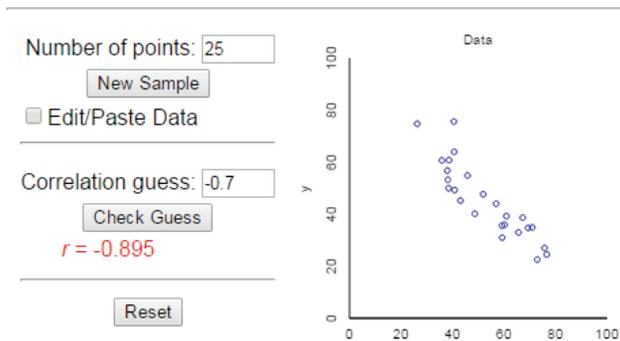


Figura 6. Applet para estimar la correlación a partir de un diagrama de dispersión.

El análisis de los resultados muestra un porcentaje de respuestas correctas en el signo de la correlación positiva de 96%, y 55% para la correlación negativa en los diagramas de dispersión de 25 puntos. Cuando el diagrama se incrementó a 50 puntos, la tasa de respuestas correctas para correlación positiva se mantuvo en 96%, y se elevó a 77% en la correlación negativa. En los diagramas de 100 puntos la tasa de respuestas correctas para correlación positiva fue de 97%, y en la correlación negativa fue de 84%. De lo anterior se desprende que a los estudiantes les resultó más sencillo identificar la correlación positiva (relación directa) que la negativa (relación inversa), lo cual coincide con resultados de investigación reportados por Beyth-Marom (1982) y Estepa (1994).

Por lo general, las dudas sobre el signo de la correlación se presentaron en los casos donde la correlación es más cercana a cero, ya que no se distingue claramente la dirección de los puntos en el diagrama, sobre todo cuando se tienen pocos datos; esta dificultad fue más crítica cuando la correlación era

negativa. Por otra parte, se observa que los estudiantes mejoraron su estimación conforme se incrementó el número de puntos en el diagrama de dispersión, particularmente en el caso de la correlación negativa. Además se observó en las primeras estimaciones que los estudiantes daban mayor importancia a la magnitud de la correlación que al signo, cuando ambos son igualmente importantes para la estimación porque el signo representa la dirección, y la magnitud a la intensidad de la relación. Sin embargo, dicha dificultad fue desapareciendo conforme elaboraron más estimaciones, al ver la divergencia entre el signo que ellos introducían y el que les proporcionaba el applet.

b) Software de análisis dinámico de datos Fathom

Al diseñar esta actividad final, nos propusimos explorar los siguientes aspectos del razonamiento covariacional: a) Interpretar información bivariada en diagramas de dispersión construidos por ellos mismos, b) Identificar en un conjunto de variables explicativas la variable de mayor influencia sobre la variable de respuesta, c) Manejo correcto de la variable de respuesta y explicativa en la recta de regresión, d) Predecir valores para la variable de respuesta a partir de un valor de la variable explicativa, e) Interpretar el significado de los coeficientes de la recta de regresión y evaluar la bondad del ajuste. Los datos analizados provenían de diversas variables que fueron medidas en 80 automóviles compactos que se vendieron en el mercado mexicano durante 2014 (ver tabla 2).

Tabla 2. Variables de 80 modelos de vehículos compactos vendidos en México en 2014.

	MARCA	MODELO	RENDIMIENTO_CIUDAD	EMISION_CO2	GASTO_ESTIMADO_COMBUSTIBLE	POTENCIA	PRECIO
1	HONDA	CIVIC HIB...	23.8	119	8700	90	266900
2	MAZDA	MAZDA3...	15.4	159	13400	153	236900
3	MAZDA	MAZDA3...	15.7	157	13200	153	236900
4	MAZDA	MAZDA3...	15	162	13800	153	258900
5	MAZDA	MAZDA3...	15.7	158	13200	153	258900
6	HONDA	CITY LX ...	13.2	182	15700	118	219000

Las unidades del rendimiento son kilómetros por litro de gasolina, la emisión se mide en gramos por kilómetro recorrido, el gasto estimado está dado en pesos por año, la potencia en caballos de fuerza y el precio se expresa en pesos.

La construcción de los diagramas de dispersión solicitados (uno con relación negativa y otro con relación positiva) fue realizada con éxito por todos los estudiantes. Sin embargo, no todos siguieron el mismo esquema, y utilizaron diferentes representaciones y estrategias. Algunos se basaron sólo en la tendencia de la nube de puntos del diagrama de dispersión, otros agregaron la recta de mínimos cuadrados que proporciona el software para tener mayor certeza del signo y la intensidad de la relación, y otros más recurrieron al cálculo del coeficiente de correlación (ver tabla 3).

Tabla 3. Representaciones utilizadas para resolver la tarea.

Tipo de representación	Frecuencia
Diagrama de dispersión solamente.	20
Diagrama de dispersión con recta de mínimos cuadrados.	5
Diagrama de dispersión y coeficiente de correlación.	9
Total	34

Con base en las representaciones utilizadas y las características de la relación entre variables identificadas por los estudiantes, clasificamos la interpretación que hicieron de los diagramas de dispersión en tres categorías del modelo SOLO (ver tabla 4).

Tabla 4. Niveles del modelo SOLO en la interpretación de diagrama de dispersión.

Niveles de interpretación	Frecuencia
Uniestructural	18
Multiestructural	15
Relacional	1
Total	34

- *Nivel uniestructural*: 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), sólo dependen de una representación para el análisis.
- *Nivel multiestructural*: 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.
- *Nivel relacional*: 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.

La única respuesta ubicada en el nivel relacional fue proporcionada por Jesús Enrique (ver figura 7): “en este diagrama observamos una asociación negativa con una relación lineal fuerte que se ve a simple vista, se puede observar un dato atípico en la parte más baja de la emisión de CO₂, esta gráfica nos muestra que entre mayor sea el rendimiento del automóvil la emisión de CO₂ será menor”.

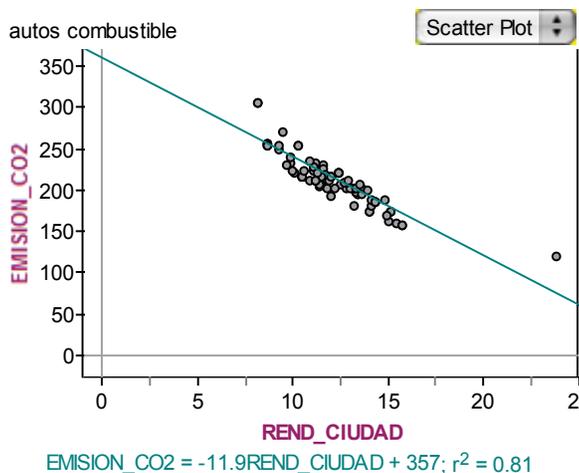


Figura 7. Diagrama de dispersión construido por Jesús Enrique.

En el análisis anterior se observa que los estudiantes no tuvieron dificultades para identificar la dirección de la relación, pero sus interpretaciones en la mayoría de los casos fueron incompletas, ya que no consideraron todas las características de la relación entre variables (por ejemplo, la forma y los datos extremos), razón por la cual su nivel de razonamiento en la interpretación del diagrama de dispersión se ubicó principalmente en los niveles uniestructural y multiestructural.

Otro apartado de la actividad solicitaba identificar la variable con mayor influencia o relación con el precio de los automóviles. 32 estudiantes establecieron correctamente que la potencia era la variable que estaba más relacionada con el precio. Sin embargo, al igual que el inciso anterior, no todos utilizaron el mismo esquema para resolver la tarea. La principal estrategia consistió en construir los cuatro diagramas de dispersión, observándose varias variantes, como se muestra en la siguiente tabla:

Tabla 5. Estrategias utilizadas por los estudiantes para resolver la tarea.

Tipo de representación	Frecuencia
Diagrama de dispersión solamente.	22
Diagrama de dispersión con recta de mínimos cuadrados.	6
Diagrama de dispersión y coeficiente de correlación.	6
Total	34

Los estudiantes dependieron más de las estrategias visuales (gráficas) que de estrategias de cálculo (simbólicas), tal fue el caso de quienes sólo usaron el diagrama de dispersión, o combinado con la recta de regresión. Estos estudiantes observaron las cuatro nubes de puntos y seleccionaron la que les parecía que tenía mayor relación con el precio, y no dependieron de ningún cálculo, aun cuando la recta de regresión estaba ligada con el valor del coeficiente de determinación r^2 (ver figura 8). Por ejemplo, José Efraín construye el siguiente diagrama y responde: “los datos de la potencia son los que están más agrupados en relación con el precio, ya que la línea de regresión pasa por mayor cantidad de datos que las demás gráficas, que se encuentran más dispersos en relación al precio”.

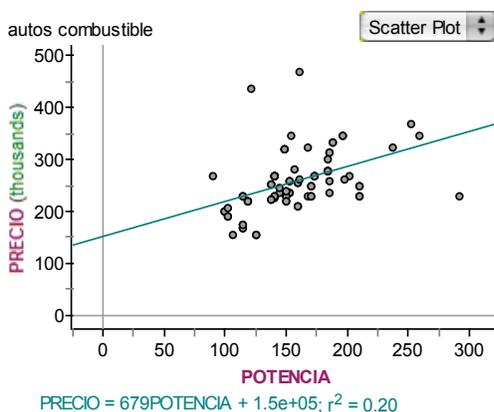


Figura 8. Diagrama de dispersión construido por José Efraín.

Por otra parte José Francisco señala: “para poder definir qué variable es la que más influye en el precio realicé cuatro diagramas de dispersión, colocando la variable del precio como variable de respuesta y comparándola con cada una de las demás variables. Después dibujé la línea de regresión para poder saber cuál es la variable que más influye en el precio. El resultado fue que las cuatro tienen poca incidencia, pero la mayor de ellas la representa el diagrama que compara el precio del automóvil con la potencia”.

Ambos estudiantes hacen referencia a aspectos visuales de los datos con la línea de regresión, como el hecho de que están más agrupados y que la línea de regresión pasa por mayor cantidad de ellos, lo cual muestra una comprensión intuitiva de la variable que más influye en la relación. Por su parte, Jonathan utiliza el diagrama de dispersión y el coeficiente de correlación para justificar la mayor relación entre precio y potencia, y señala: “la variable que tiene mayor incidencia en el precio es la potencia, que tiene una correlación de 0.44 y es la más alta de todas las demás variables”.

En las justificaciones de este inciso, 21 estudiantes se basaron en la dirección de la nube de puntos, señalando solamente que a mayor potencia el precio era mayor, lo que los ubica en un nivel de razonamiento uniestructural. Otros 13 estudiantes, además de la dirección, complementaron sus justificaciones con otras propiedades como la intensidad (ver tabla 6), por lo que fueron ubicados en un nivel multiestructural, de acuerdo con las categorías del modelo SOLO descritas anteriormente.

Tabla 6. Niveles del modelo SOLO en la determinación de la variable de mayor influencia.

Niveles de interpretación	Frecuencia
Uniestructural	21
Multiestructural	13
Relacional	0
Total	34

La última parte de la actividad estuvo enfocada en evaluar aspectos de la regresión, considerando el rendimiento como variable explicativa y la emisión de CO₂ como variable de respuesta. En lo relativo a la precisión de la ecuación de la recta para la predicción de valores de CO₂, algunos estudiantes tuvieron dificultades desde la construcción misma del diagrama de dispersión, al ubicar incorrectamente la variable explicativa y la variable de respuesta en los ejes correspondientes. La mitad de los estudiantes dieron explicaciones sobre la relación entre precisión y predicción en las que mostraron una comprensión intuitiva de la relación entre estos conceptos, atendiendo principalmente a los criterios visuales que les daban los diagramas de dispersión, a pesar de que al pie de la gráfica se proporcionaba el coeficiente de determinación. La otra mitad emitió argumentos relacionados con la interpretación del diagrama y no sobre lo que se les preguntaba en relación a la precisión. A continuación se muestran algunas de sus respuestas:

“La precisión de la recta es bastante acertada, la mayoría de los puntos están cerca de la recta, sólo un punto está bastante alejado de los demás, la predicción sobre los valores sería algo acertada porque para un próximo valor de la gráfica se pudiera saber qué es lo que seguiría” (Hugo).

“Los puntos están muy cerca de la recta de regresión o sobre ella, esto significa que la emisión de CO₂ y el rendimiento están sumamente relacionados. La predicción de datos de emisión de CO₂ sería muy buena porque están muy cerca o sobre la línea de regresión y eso nos indica que hay buena posibilidad de acertar en la predicción” (Juan Andrés).

“La precisión de la recta en este diagrama de puntos es muy alta porque todos los datos están muy cerca de la línea, por tanto, las predicciones para los valores de CO₂ con respecto al rendimiento de la ciudad son muy acertados” (Jesús Enrique).

Estos estudiantes relacionan correctamente la precisión de una predicción con la cercanía de los puntos con la línea recta, pero no utilizan indicadores como los coeficientes de correlación y de determinación para complementar su argumentación.

En cuanto a los coeficientes de la ecuación de regresión, se pedía que explicaran en el contexto del problema el significado del -11.9 que aparece en la ecuación de la parte inferior de la gráfica (ver figura 5). Este inciso fue el más complicado, pues sólo 16 estudiantes señalan correctamente que representa la pendiente de la recta, pero ninguno dio una explicación de su significado, es decir, que por cada kilómetro por litro que se incrementa el rendimiento, la emisión de CO₂ disminuirá -11.9 gramos. Dejaron en blanco el inciso 11 estudiantes, y los restantes respondieron que se trataba del coeficiente de correlación o el error. Ningún estudiante dio una respuesta satisfactoria a lo que se solicitaba en el inciso.

Por último, ante la pregunta de que si un automóvil tiene un rendimiento de 10 km por litro en la ciudad, ¿qué emisión de CO₂ tendría?, 26 estudiantes estimaron el valor de CO₂ a partir de visualizar el diagrama de dispersión, y en forma muy aproximada al valor real; el resto utilizó la ecuación de regresión, por lo cual –de nueva cuenta– los estudiantes hicieron uso de estrategias de visualización, más que de cálculo, para responder una tarea.

CONCLUSIONES

En este estudio se confirman muchos resultados encontrados en investigaciones descritas en los antecedentes, y se identifican nuevas dificultades y estrategias de los sujetos en el análisis de datos en ambientes computacionales. En la evaluación diagnóstica, los estudiantes exhibieron un bajo nivel razonamiento intuitivo sobre la covariación en el rubro de diagramas de dispersión; incluso después de la enseñanza en el ambiente computacional, la aparente sencillez de un diagrama de dispersión queda en entredicho con las respuestas que han dado, pues si bien la dirección y la intensidad fueron características que identificaron y estimaron bastante bien en la actividad final, dejaron de lado otros elementos

importantes, como son los datos extremos y la forma del diagrama, razón por la cual fueron ubicados en los niveles de razonamiento inferiores de la jerarquía construida. De esta manera se observa en las estrategias de análisis de datos bivariados de los estudiantes, un uso parcial de potencial representacional del ambiente computacional.

Respecto a la correlación entre variables, los resultados muestran que a los estudiantes les resulta más difícil identificar una correlación negativa que una positiva, lo cual ha sido reportado también por Beyth-Marom (1982) y Batanero, Estepa y Godino (1997). Sin embargo, las herramientas tecnológicas utilizadas han mostrado que tienen potencial –tanto amplificador como reorganizador– para ayudar a mejorar las estrategias de identificación y estimación de la correlación, mediante el incremento del número de datos en el diagrama de dispersión y a través de la interactividad, que proporciona una visualización inmediata del cambio del coeficiente de correlación y la forma del diagrama cuando algún dato o parámetro es modificado. Estas ventajas de la tecnología también han sido observadas con estudiantes de otros niveles educativos que participaron en los análisis de Konold (2002) y Fitzallen (2012), donde se utilizaron ambientes computacionales con características similares a las del software empleado en esta investigación.

En el análisis de la regresión, un número apreciable de estudiantes tuvo dificultades para distinguir la variable de respuesta de la variable explicativa, lo cual ha sido reportado como una de las principales dificultades para la comprensión de la regresión en el estudio de Batanero, Estepa y Godino (1997). Sin embargo, fueron muy acertados en identificar la variable explicativa con mayor relación para construir un modelo de regresión. Las mayores dificultades se presentaron al interpretar los coeficientes de la ecuación de regresión y al evaluar la bondad del ajuste con base en el coeficiente de determinación.

Los estudiantes usaron principalmente estrategias que privilegiaban el uso de representaciones gráficas sobre las representaciones de cálculo que les proporcionaba el ambiente computacional, a pesar de que era muy sencillo calcular el coeficiente de correlación y el coeficiente de determinación, esto gracias al poder de visualización del software utilizado. Esta estrategia puede llevar a respuestas correctas cuando la intensidad de la relación es de moderada a alta, pero si la intensidad es baja podría conducir a errores de estimación, así que una mejor estrategia es emplear tanto las representaciones gráficas como los cálculos correspondientes, lo cual fue puesto en práctica por una minoría de los sujetos.

En resumen, los resultados del estudio confirman que el análisis de datos bivariados es un tema complejo que requiere de un tratamiento didáctico profundo, pues las dificultades encontradas coinciden con las de otros sujetos de estudio que participaron en contextos distintos. No obstante el reducido tiempo del análisis y su carácter exploratorio, los resultados muestran que el amplio recurso representacional dinámico e interactivo de los ambientes computacionales como Fathom son un factor importante a considerar en el diseño de secuencias didácticas para el aprendizaje de la covariación. Se requiere elaborar estudios más extensos sobre el tema, con principios instruccionales que tomen en cuenta el uso de datos reales, ambientes constructivistas en el salón de clase y tecnología.

REFERENCIAS

- Batanero, C., Estepa, A., & Godino, J. D. (1997). Evolution of Students' Understanding of Statistical Association in a Computer Based Teaching Environment. In: J. B. Garfield y G. Burrill (eds.). *Research on the Role of Technology in Teaching and Learning Statistics: Proceedings of the 1996 IASE Roundtable Conference*. Voorburg, The Netherlands. International Statistical Institute.
- Batanero, C., Gea, M., Díaz, C. & Cañadas, G. (2014). Building High School Pre-Service Teachers' Knowledge to Teach Correlation and Regression. In: K. Makar, B. de Sousa y R. Gould (eds.). *Proceedings of the Ninth International Conference on Teaching Statistics*. Flagstaff, Arizona, USA. Voorburg, The Netherlands. International Statistical Institute.
- Beyth-Marom, R. (1982). Perception of Correlation Reexamined. *Memory & Cognition*, 10(6), pp. 511-519.
- Biehler, R., Ben-Zvi, D., Bakker, A. & Maker, K. (2013). Technology for Enhancing Statistical Reasoning at the School Level. In: M. A. Clements, A. Bishop, C. Keitel, J. Kilpatrick y F. Leung (eds.). *Third International Handbook of Mathematics Education*. New York: Springer, pp. 643-690.
- Biggs, J. & Collis, K. (1982). *Evaluating the Quality of Learning: the SOLO taxonomy*. Academic Press. New York.
- Cañadas, G. (2010). *Las tablas de contingencia en la formación de profesionales de psicología*. Trabajo fin de Máster. Departamento de Didáctica de la Matemática. España. Universidad de Granada.

- Cobb, P. & McClain, K. (2004). Principles of Instructional Design for Supporting the Development of Students' Statistical Reasoning. In: D. Ben-Zvi y J. B. Garfield (eds.). *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking*. Kluwer Academic Publishers, pp. 375-396.
- DGB (2009). *Programas de estudio de matemáticas*. México. Secretaría de Educación Pública. Dirección General de Bachillerato.
- Estepa, A. (1993). *Concepciones iniciales sobre la asociación estadística y su evolución como consecuencia de una enseñanza basada en el uso de ordenadores*. Tesis doctoral. España. Universidad de Granada.
- Estepa, A. & Batanero, C. (1996). Judgments of Correlation in Scatter Plots: Students' Intuitive Strategies and Preconceptions. *Hiroshima Journal of Mathematics Education*, 4, pp. 25-41.
- Finzer, W., Erickson, T. & Binker, J. (2002). *Fathom Dynamic Statistics Software*. Emeryville, CA. Key Curriculum Press Technologies.
- Fitzallen, N. (2012). *Reasoning about Covariation with TinkerPlots*. Tesis doctoral. Australia, University of Tasmania.
- Garfield, J. & Ben-Zvi, D. (2008). *Developing Students' Statistical Reasoning. Connecting Research and Teaching Practice*. Springer. The Netherlands.
- Gea, M. (2014). *La correlación y regresión en bachillerato: análisis de libros de texto y del conocimiento de los futuros profesores*. Tesis doctoral. Departamento de Didáctica de la Matemática. España. Universidad de Granada.
- Gravemeijer, K. (2000). A Rationale for an Instructional Sequence for Analyzing One-and Two-Dimensional Data Sets. *Paper presented at the annual meeting of the American Educational Research Association*. Montreal, Canadá.
- Jennings, D., Amabile, T. & Ross, L. (1982). Informal Covariation Assessment: Data-Based Versus Theory-Based Judgments. In: D. Kahneman, P. Slovic & A. Tversky (eds.). *Judgment under Uncertainty: Heuristics and Biases*. Cambridge, England: Cambridge University Press, pp. 211-230.
- Konold, C. (2002). Teaching Concepts Rather than Conventions. *New England Journal of Mathematics*, 34(2), pp. 69-81.
- Konold, C. & Miller, C. D. (2005). *TinkerPlots: Dynamic Data Exploration* [Computer software]. Key Curriculum Press. Emeryville, CA.
- NCTM (2000). *Principles and Standards for School Mathematics*. National Council of Teachers of Mathematics. Reston, VA.
- Pea, R. (1987). Cognitive Technologies for Mathematics Education. In: A. Schoenfeld (Ed.) *Cognitive Science and Mathematics Education*. Lawrence Erlbaum Associates Publishers, pp. 89-122.

- Sánchez Cobo, F. T. (1999). *Significado de la correlación y regresión para los estudiantes universitarios*. Tesis doctoral. Departamento de Didáctica de la Matemática. España. Universidad de Granada.
- Stockburger, D. W. (1982). Evaluation of Three Simulation Exercises in an Introductory Statistics Course. *Contemporary Educational Psychology*, 7(4), pp. 365-370.
- Watkins, A. E., Scheaffer, R. L. & Cobb, G. W. (2004). *Statistics in Action: Understanding a World of Data*. Key Curriculum Press. Emeryville, CA.
- Zieffler, S. A. (2006). *A Longitudinal Investigation of the Development of College Students' Reasoning about Bivariate Data During an Introductory Statistics Course*. Tesis doctoral. University of Minnesota.

