

Conocimiento Tecnológico sobre la Correlación y Regresión: un estudio exploratorio con Futuros Profesores

Technological Knowledge of Correlation and Regression: an exploratory study with Prospective Teachers

María M. Gea*

Pedro Arteaga**

Carmen Batanero***

Juan Jesús Ortiz****

Resumen

Se evalúa el conocimiento tecnológico sobre la correlación y regresión en veinticinco estudiantes que se preparan para ser profesores de matemáticas de Educación Secundaria y Bachillerato en España. Nos basamos en el modelo de Niess (2005) de conocimiento del profesor para enseñar con ayuda de la tecnología. Utilizando datos obtenidos del servidor de las Naciones Unidas, se propone a los futuros profesores ajustar un modelo de regresión a la variable *esperanza de vida*, en función de diversos indicadores del desarrollo, en 194 países. Se propone, también, elegir otra variable dependiente o tomar nuevos datos de las Naciones Unidas para repetir el análisis. Los resultados sugieren un buen conocimiento estadístico y un buen conocimiento tecnológico estadístico sobre la correlación y regresión en la mayoría de los participantes. No obstante, algunos muestran dificultades en las ideas de independencia e incorrelación y el uso de la tecnología, o realizan un uso acrítico de la misma.

Palabras clave: Conocimiento tecnológico. Correlación y regresión. Futuros profesores de Educación Secundaria y Bachillerato. Evaluación.

Abstract

We assessed the technological knowledge on correlation and regression in 25 students who are preparing to become secondary and high school education mathematics teachers in Spain. We relied on the theoretical model by Niess (2005) of teacher's knowledge to teach with technology. Using data from the United Nations server, the prospective teachers were requested to fit a regression model to the life expectancy according to various development indicators in 194 countries. They were also proposed to choose another dependent variable or to take new data from the United Nations to repeat the analysis. The results suggested a good statistical knowledge

* Doctora en Ciencias de la Educación por la Universidad de Granada (UGR), España. Profesora del Departamento de Didáctica de la Matemática de la Universidad de Granada (UGR), Granada, España. Dirección postal: Campus de Cartuja, 18071, Granada, España. E-mail: mmgea@ugr.es.

** Doctor en Didáctica de la Matemática por la Universidad de Granada (UGR), España. Profesor del Departamento de Didáctica de la Matemática de la Universidad de Granada (UGR), Granada, España. Dirección postal: Campus de Cartuja, 18071, Granada, España. E-mail: parteaga@ugr.es.

*** Doctora en Matemática por la Universidad de Granada (UGR), España. Catedrática de la Universidad de Granada (UGR), Granada, España. Dirección postal: Campus de Cartuja, 18071, Granada, España. E-mail: batanero@ugr.es.

**** Doctor en Matemáticas por la Universidad de Granada (UGR), España. Profesor Titular de la Universidad de Granada (UGR), Melilla, España. Dirección postal: Campus Universitario de Melilla (Universidad de Granada), 52005, Melilla, España. E-mail: jortiz@ugr.es.

and a good statistical technological knowledge about correlation and regression in most participants. However, some of them showed difficulties in the ideas of independence and incorrelation, in the use of technology or made an uncritical use of it.

Keywords: Technological knowledge. Correlation and regression. Prospective secondary and high school teachers. Assessment.

1 Introducción

La correlación y regresión son ideas básicas en estadística, utilizadas en las ciencias y un gran número de campos de la actividad humana. Estas técnicas amplían la idea de dependencia funcional a situaciones aleatorias y, en España, se presentan en el primer curso de Bachillerato (alumnos de 16 años) en las dos modalidades en que se enseña matemáticas: *Ciencias y Humanidades* y *Ciencias Sociales* con contenidos muy similares. Por ejemplo, en la modalidad de Humanidades y Ciencias Sociales encontramos (MECD, 2015, p. 385):

Estadística descriptiva bidimensional: Tablas de contingencia. Distribución conjunta y distribuciones marginales. Distribuciones condicionadas. Medias y desviaciones típicas marginales y condicionadas. Independencia de variables estadísticas. Dependencia de dos variables estadísticas. Representación gráfica: Nube de puntos. Dependencia lineal de dos variables estadísticas. Covarianza y correlación: Cálculo e interpretación del coeficiente de correlación lineal. Regresión lineal. Predicciones estadísticas y fiabilidad de las mismas. Coeficiente de determinación.

La importancia y utilidad de este tema para los estudiantes es clara, así como en la mayoría de estudios universitarios; aun así, su enseñanza y aprendizaje no está exenta de problemas didácticos que la investigación ha descrito. Destacamos, entre otros, las dificultades de comprensión en torno a los conceptos de covarianza y correlación como, por ejemplo, no distinguir una distribución bidimensional de dos conjuntos de datos independientes; la concepción unidireccional de la correlación (aceptar únicamente la correlación directa); la oposición entre el razonamiento numérico y el gráfico en la estimación de la correlación; o la concepción causal (confundir correlación y causalidad) al estimar una correlación significativa (ESTEPA, 2008; ESTEPA; BATANERO, 1996; ZIEFFLER; GARFIELD, 2009).

Un recurso para ayudar a los estudiantes a superar muchas de estas dificultades es la tecnología. Hoy en día, existe una gran variedad de recursos tecnológicos como la calculadora, hoja de cálculo, applets y programas de ordenador específicos, que pueden facilitar la realización de cálculos y gráficos (PRATT; DAVIES; CONNOR, 2011).

Además, proporciona al estudiante recursos de visualización y exploración de las ideas estadísticas (BIEHLER et al., 2013). El aprendizaje de y a través de la tecnología es esencial

en esta etapa educativa y este hecho se especifica en los documentos curriculares, donde se incluye el siguiente objetivo “Utilizar con solvencia y responsabilidad las tecnologías de la información y la comunicación” (MECD, 2015, p. 188). Y, en relación con las matemáticas, se indica:

El uso de herramientas tecnológicas tendrá un papel esencial en el currículo de la materia, tanto para la mejor comprensión de conceptos o en la resolución de problemas complejos, como para contrastar con mayor rigor las hipótesis propuestas y presentar y comunicar los resultados obtenidos. Además, estas herramientas contribuyen a la preparación para el aprendizaje a lo largo de la vida y apoyan el trabajo fuera del aula (MECD, 2015, p. 381)

Para seguir estas recomendaciones es importante asegurar un conocimiento adecuado de los profesores de los recursos tecnológicos; sin embargo, la investigación relacionada con el conocimiento del profesor de dichos recursos en el tema de la correlación y regresión es escasa. Con objeto de contribuir a este punto, el objetivo de este trabajo fue describir y desarrollar el conocimiento tecnológico sobre el tema de correlación y regresión en una muestra de estudiantes españoles, que se preparan para ser profesores de matemáticas en Educación Secundaria y Bachillerato. Los resultados servirán como base para la organización de actividades de formación sobre este tema.

2 Fundamentos

Nos basamos en dos tipos de fundamentos. En primer lugar, analizamos el papel de la tecnología en la enseñanza y el aprendizaje de la estadística. En segundo lugar, la conceptualización del conocimiento del profesor de matemáticas, centrándonos en los modelos que consideran el conocimiento específico del profesor para enseñar con apoyo tecnológico.

2.1 Papel de la tecnología en la enseñanza y aprendizaje de la estadística

El uso de la tecnología en la enseñanza de la estadística ha sido reconocido, entre otros, por Pratt, Davies y Connor (2011), que destacan la reducción del tiempo de cálculo y la ampliación del tipo de gráficos que el alumno puede realizar interactivamente. Igualmente, señalan la posibilidad de trabajar con proyectos, en que el alumno parte de un problema de investigación y completa todos los pasos de una investigación. Este tipo de trabajo fomenta los componentes del razonamiento estadístico: reconocimiento de la necesidad de datos,

percepción de la variación, transnumeración, uso de modelos estadísticos e integración de la estadística y el contexto (WILD; PFANNKUCH, 1999).

Además, si se trabaja con datos reales, como en el caso del proyecto usado en este trabajo, se potencia la interdisciplinariedad en clase de estadística, permitiendo aprender contenidos que no se adquieren habitualmente con problemas tomados de los libros de texto (HALL, 2011). Al facilitar el cálculo y la representación gráfica, la tecnología disminuye el problema tradicional en la enseñanza de la estadística en cuanto al desfase entre la comprensión de los conceptos y los medios técnicos de cálculo para poder aplicarlos (BATANERO; BOROVCNIK, 2016).

Otra aplicación, característica del uso de la tecnología, es la simulación y visualización (BIEHLER et al., 2013). Las representaciones icónicas de conceptos (por ejemplo, la desviación de los puntos de un diagrama de dispersión a la recta de regresión) pueden ayudar a los estudiantes a pensar a través de modelos concretos, cuando todavía no son capaces de generalizar sus ideas matemáticas. Por otro lado, permiten explorar objetos abstractos, creando micromundos virtuales donde los estudiantes pueden experimentar con las diferentes variables que intervienen. Fernandes, Batanero, Contreras y Díaz (2009) añaden el interés de la simulación en el aprendizaje de la modelización, pues una simulación es ya un modelo matemático (al haber simplificado la realidad), pero es menos abstracto que el modelo matemático y constituye un puente entre éste y la realidad (EICHLER; VOGEL, 2014).

2.2 Conocimiento del profesor para enseñar matemáticas con apoyo tecnológico

Hoy en día, hay una amplia línea de investigación centrada en el análisis del conocimiento requerido por el profesor para la enseñanza de las matemáticas (descrita, por ejemplo, en BLÖMEKE et al., 2014; LLINARES; KRAINER, 2006; TATTO; SENK, 2011).

Dentro de esta línea de investigación, algunos autores tienen en cuenta los conocimientos específicos que necesita el profesor para enseñar con apoyo de la tecnología. Así, Lee y Hollebrands (2011) adaptan a la estadística el modelo de Niess (2005), que toma de Shulman (1986), que incluye el Conocimiento del Contenido (CC) y Conocimiento Pedagógico (CP) y añade el Conocimiento Tecnológico (CT), y tiene en cuenta todas las posibles intersecciones de estos tres componentes (Figura 1), de entre las que destacamos:

- CPC: *Conocimiento pedagógico del contenido o cómo enseñar un cierto contenido*. Por ejemplo, el relacionado con la enseñanza de la correlación y regresión.

- **CTP:** *Conocimiento tecnológico-pedagógico* o conocimiento del uso de la tecnología en la enseñanza, en general (no para un contenido concreto). Por ejemplo, cómo intercalar actividades con tecnología en el proceso de enseñanza para motivar a los estudiantes.
- **CTC:** *Conocimiento tecnológico del contenido*. Implica el conocimiento de los recursos tecnológicos y de su uso para un contenido concreto; en nuestro caso, un ejemplo sería el conocimiento de la hoja Excel como recurso para el análisis estadístico en cuanto a la producción de gráficos y resúmenes estadísticos, entre otros.
- **CTPC:** *Conocimiento tecnológico-pedagógico del contenido*. Este estaría en el corazón del modelo y sería el conocimiento pedagógico para enseñar un contenido particular con ayuda de la tecnología. Incluiría el conocimiento de las estrategias de enseñanza y las representaciones para enseñar temas particulares con apoyo de la tecnología; sobre la comprensión, razonamiento y aprendizaje de los estudiantes con la tecnología; así como del currículo y materiales curriculares que integran la tecnología en el aprendizaje.

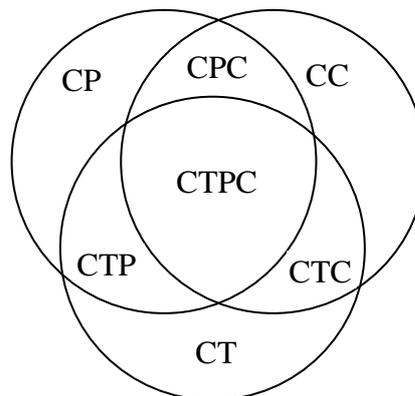


Figura 1 – Modelo de Niess

Fuente: Adaptado de Lee y Hollebrands (2011, p. 361)

Siguiendo este modelo, en este trabajo analizamos específicamente el conocimiento tecnológico del contenido sobre la correlación y regresión (que sería una parte del CTC). Específicamente, analizamos su conocimiento del ajuste de modelos de regresión utilizando la hoja Excel, que es el recurso didáctico específicamente sugerido en el currículo español (MECD, 2015). Puesto que el CTC es intersección del conocimiento tecnológico (CT) con el conocimiento del contenido (CC), en este trabajo, indirectamente, también se describe el conocimiento sobre la correlación y regresión de los futuros profesores.

3 Antecedentes

Son pocas las investigaciones previas que han analizado los componentes del conocimiento de los profesores sobre correlación y regresión.

Sobre el conocimiento del contenido correlación y regresión, el primer trabajo fue realizado por Estepa y sus colaboradores (BATANERO; GODINO; ESTEPA, 1998; ESTEPA; BATANERO, 1996). Los autores realizaron una experiencia de enseñanza de la asociación (incluyendo las tablas de contingencia, correlación y regresión) con una muestra de veintidós futuros profesores de educación primaria. Entre otras dificultades con las que se encontraron los futuros profesores, los investigadores citan la identificación de la correlación inversa, la confusión entre la variable dependiente e independiente en la regresión, así como estrategias incorrectas en la interpretación de los diagramas de dispersión, por ejemplo, usar sólo parte de los datos.

Casey (2010) analiza las clases desarrolladas por tres profesores de secundaria en ejercicio y determina veinte descriptores observados del conocimiento para la enseñanza de la correlación y regresión, según el modelo propuesto por Ball, Thames y Phelps (2008). Los descriptores más observados fueron los relacionados con las propiedades de la recta de regresión, así como la interpretación y dominio del lenguaje asociado al cálculo del coeficiente de correlación. La autora incluye, en sus descriptores, los aspectos inferenciales (estimación de los coeficientes de correlación y regresión), que no tenemos en cuenta en este trabajo al no ser contenidos de enseñanza en Bachillerato (MECD, 2015).

En Casey y Wasserman (2015) se analiza el conocimiento de diecinueve profesores (once en formación y ocho en servicio) sobre la enseñanza del ajuste informal de una recta de regresión. Sin embargo, estos trabajos no se centran en el uso de la tecnología. Tampoco tienen como finalidad describir el conocimiento de los profesores en estudio, sino solo delimitar el conocimiento que es necesario para la enseñanza del tema. Por otro lado, nosotros usamos tareas no tenidas en cuenta por la autora.

Quintas, Ferreira y Oliveira (2015) analizan el conocimiento de dos profesoras de secundaria, con amplia experiencia, mientras explican la correlación y regresión en un curso de secundaria. Describen ejemplos del uso, por estas profesoras, de su conocimiento del contenido y su conocimiento pedagógico. Se observan, también, errores por parte de las profesoras; más concretamente, ninguna de ellas es consciente de la existencia de dos rectas diferentes de regresión y piden a los alumnos usar siempre la recta de regresión de Y sobre X, independientemente de cuál sea la variable que se quiere predecir.

Tendremos en cuenta, también, algunas investigaciones sobre la comprensión de la correlación y regresión con alumnos universitarios. Por ejemplo, Truran (1997) evalúa el

aprendizaje de estudiantes de economía y negocios sobre la recta de regresión y el coeficiente de determinación. Indica que el aprendizaje de este último concepto es rutinario, pero que menos del 10% asumen la existencia de modelos no lineales de regresión.

Estepa y Sánchez-Cobo (2003) encuentran que el 84 de 197 estudiantes de las diplomaturas de empresariales y enfermería, en una universidad española, tienen dificultad para discriminar la variable dependiente e independiente al resolver problemas relacionados. Otros confunden la intensidad y la dirección de la correlación al interpretar que un valor moderado del coeficiente de determinación corresponde a una relación inversa. Igualmente, hay una tendencia a pensar que la relación conjunta debe ser principalmente de tipo lineal, desechando otros posibles modelos de ajuste, sobre todo si el coeficiente de correlación es de pequeña intensidad. Alrededor del 11% de los estudiantes confunden los coeficientes de correlación r y de determinación r^2 .

Aunque los trabajos citados fundamentan el nuestro y aportan puntos de comparación, ninguno realiza un estudio con futuros profesores de secundaria y bachillerato; tampoco analizan su competencia en la identificación de la función de ajuste utilizando la tecnología y en la capacidad de proponer nuevos problemas utilizando datos reales. Estos son puntos originales en nuestro trabajo.

4 Metodología

Nuestra investigación es de tipo cualitativa, exploratoria y descriptiva, pues la muestra de participantes es intencional, por lo que nuestra intención no es generalizar a otras muestras o contextos. Está basada en el análisis de contenido, que estudia la naturaleza del discurso y se basa en el análisis sistemático de documentos escritos (ZAPICO, 2006), en nuestro caso las producciones de los participantes a las tareas propuestas. En la sección 4.1, se describen el contexto y muestra participante, así como el método de trabajo en el aula. En la sección 4.2, se descompone el objetivo general en dos objetivos específicos y se analiza el instrumento utilizado y la forma en que las tareas propuestas operacionalizan el conocimiento tecnológico del contenido sobre correlación y regresión para el *software* específico utilizado.

4.1 Contexto, muestra y método de trabajo con los futuros profesores

El estudio se desarrolló en un curso del Máster de Formación del Profesorado de Educación Secundaria y Bachillerato, especialidad de matemáticas, que tiene un año de

duración y es obligatorio para todos los que quieren concursar para una plaza de profesor de matemáticas de secundaria, en España. El máster consta de un total de sesenta créditos (cada crédito corresponde a veinticinco horas de trabajo, incluyendo docencia en el aula y trabajo personal del alumno). Estos créditos se dividen en doce créditos de materias de Psicología y Didáctica General, seis de complementos de matemática, doce de Enseñanza y Aprendizaje de las Matemáticas, seis de Innovación Docente en Matemáticas, ocho de libre disposición (atención a la especificidad del alumnado o gestión educativa, así como de otros másteres), diez de prácticas en el aula y seis de elaboración de un trabajo fin de Máster.

La muestra estuvo constituida por un grupo de veinticinco estudiantes, quienes fueron informados de la finalidad del estudio, y cedieron sus datos voluntariamente. Sólo 14 (56%) de ellos eran licenciados en Matemáticas o Estadística y el resto habían cursado ingenierías, arquitectura u otras ramas de ciencias. Todos habían cursado una o más asignaturas de estadística y quince (50%) tenían experiencia de enseñanza en colegios privados o clases particulares. Todos ellos conocían el manejo de la hoja Excel, cuyo aprendizaje es recomendado en la enseñanza de las matemáticas en el currículo español (MECD, 2015).

La recogida de datos se llevó a cabo como parte de un taller formativo realizado dentro de la asignatura de Innovación Docente en Matemáticas, único curso en que los estudiantes trabajan con la tecnología. Se dedicaron tres sesiones de dos horas cada una, de las cuales las dos primeras estaban dirigidas a describir y desarrollar el *Conocimiento tecnológico del contenido* (CTC) sobre la correlación y regresión de los participantes, mientras la tercera estuvo dedicada a describir y desarrollar su *Conocimiento tecnológico-pedagógico del contenido* (CTPC) sobre el tema.

En estas sesiones se trabajó con un proyecto estadístico, alrededor de una pregunta de investigación (¿Cuáles son los factores que más influyen en la esperanza de vida al nacer en un país?) que permitió plantear diferentes tareas. Los datos para este proyecto se tomaron de los utilizados en la elaboración de los Informes sobre Desarrollo Humano por las Naciones Unidas (ONU, 2016), que están disponibles en su servidor (<http://hdr.undp.org/es/data>).

Las variables utilizadas en el proyecto son indicadores internacionales de desarrollo humano y, al seleccionarlas, se trató de no limitar la relación de dependencia entre las variables a la regresión lineal e incluir otros modelos de ajuste con funciones que se estudian en la Educación Secundaria. Igualmente, se brindó la posibilidad de observar el signo positivo y negativo y los diferentes grados de intensidad en la correlación.

El formador proporcionó a los futuros profesores una hoja Excel que contenía los datos correspondientes a las variables del fichero que ofrece el servidor de las Naciones

Unidas, junto con los resúmenes estadísticos de su distribución y algunos gráficos. El formador de profesores recordó a los participantes las herramientas de regresión de Excel realizando preguntas para comprobar su conocimiento previo sobre el programa y mejorarlo en aquel que hubiesen olvidado, utilizando como ejemplo la primera de las variables independientes (Cuadro 1). Dichas herramientas permiten visualizar, junto con el diagrama de dispersión, diferentes funciones de ajuste junto con su expresión algebraica y el valor del coeficiente de determinación. Se consideran los modelos exponencial, lineal, logarítmico, polinómico (grado 2 a 6) o potencial.

4.2 Objetivos específicos e instrumento de recogida de datos

Sobre el proyecto trabajado en el aula, se propusieron a los futuros profesores tres tareas sobre el uso de la tecnología en el tema de correlación y regresión que se muestran en el Cuadro 1.

Tarea 1. Utilizando Excel encuentra la función que mejor describa la esperanza de vida a partir de cada una de las variables (entre las disponibles en el programa) y escribe la expresión algebraica	
Variante independiente	Expresión algebraica de la función que usaríamos para predecir la esperanza de vida $Y=f(X)$
Índice de Desarrollo Humano	
PIB per cápita	
Tasa de fecundidad adolescentes	
Tasa de mortalidad de niños	
Gasto público en salud	
Índice de educación	
Población, total	
Población urbana	

Tarea 2. Elige otra variable del fichero como variable dependiente. Utilizando los datos dados en la hoja Excel, estudia su posible relación con el resto de variables, incluida la esperanza de vida. Utilizando Excel encuentra la función que mejor describa dicha variable a partir de cada una de las otras variables (entre las disponibles en el fichero) y escribe la expresión algebraica.

Tarea 3. *Otros datos disponibles en Internet.* Visita el servidor de las Naciones Unidas en (<http://hdr.undp.org/es/data>). Elige algunas otras variables que te interesen; descárgalas en formato Excel y estudia su relación con otras variables. Elige una de ellas como variable dependiente y encuentra la función de ajuste con cada una de las otras variables elegidas.

Cuadro 1 – Tareas propuesta a los participantes
Fuente: Elaboración propia para la investigación

El objetivo general de la investigación se descompuso en dos objetivos específicos a los que se trata de responder mediante las tareas propuestas:

O1. Describir el *Conocimiento Tecnológico del Contenido* (CTC) de los futuros profesores para ajustar un modelo de regresión para cada una de las variables independientes,

utilizando la hoja Excel, así como identificar aquella que mejor describa a la *esperanza de vida* (Tareas 1 y 2).

Indirectamente, se pretendía reforzar en los futuros profesores el conocimiento de la herramienta que proporciona Excel, muy sencilla de utilizar para el estudio de la correlación y regresión, no sólo lineal, pues incluye diversas funciones de ajuste. Asimismo, las dos tareas permiten, indirectamente, reforzar el Conocimiento del Contenido (CC) de los futuros profesores; específicamente sobre la regresión, análisis del coeficiente de determinación, como medida de la bondad del ajuste, y la proporción de la varianza explicada por el modelo. Igualmente, se encuentra latente en ambas tareas el análisis de la correlación pues se pide a los participantes reflexionar sobre la conveniencia del uso de la función de ajuste obtenida para predecir una variable.

De este modo, los participantes tuvieron ocasión de relacionar los valores del coeficiente de correlación y de determinación en casos en que la dependencia fuese o no lineal. En cualquiera de estas dos tareas, para decidir el mejor ajuste, se toma en cuenta la simplicidad de la expresión y la mejora sustancial de la precisión, es decir, el incremento del valor del coeficiente de determinación. Asimismo, se refuerza su conocimiento del contenido de diversas funciones elementales y del significado de sus parámetros, así como la interpretación del coeficiente de determinación para valorar el ajuste de diferentes modelos matemáticos.

Mientras en la Tarea 1 se trabaja con la principal variable del proyecto, en la Tarea 2 se propone a los participantes cambiar la variable dependiente y repetir el análisis con la finalidad de promover su creatividad en la formulación de nuevos problemas y su iniciativa en el trabajo autónomo con variables de su interés.

O2: Describir y desarrollar algunos elementos del *Conocimiento Tecnológico-Pedagógico del Contenido* (CTPC) sobre correlación y regresión de los futuros profesores.

Para lograr este objetivo, la Tarea 3 ofrece al futuro profesor espacios de investigación y reflexión, para poder elegir sus propios datos, analizarlos y establecer conclusiones sobre otras variables disponibles en la web de las Naciones Unidas. En particular, se trata de desarrollar su capacidad de elaboración de proyectos de estadística en que pueda trabajarse el tema; dichos proyectos son recomendados para el trabajo con estadística en las orientaciones curriculares españolas (MECD, 2015).

Como ayuda en esta tarea, el formador de profesores mostró a los participantes el servidor, y explicó la forma en que podrían elegirse variables para introducirlas en la hoja Excel, al igual que se había hecho con la utilizadas en el proyecto. El formador resaltó la

cantidad de variables disponibles para analizar su posible dependencia con la *esperanza de vida*, o plantear nuevos estudios. También explicó que este es sólo uno de los muchos servidores que proporcionan datos. Se explicó que en el trabajo con estudiantes de Bachillerato, el objetivo es que el estudiante formule nuevas preguntas que requieran disponer de datos para resolverlas, desarrollando su razonamiento estadístico. También se espera que conozcan mejor la forma en que se recogen estas variables, y aprecien la complejidad del proceso de elaboración de las estadísticas demográficas o económicas, así como la utilidad de la información que aportan para el desarrollo de un país. Todas las explicaciones del formador de profesores fueron complementadas con presentaciones de *Power Point* que describían el uso de la tecnología y que se pusieron a disposición de los participantes.

5 Resultados

Los estudiantes resolvieron individualmente las tareas y prepararon un fichero *word* o *pdf* en el que incluyeron los gráficos y análisis de datos elaborados con Excel, así como la interpretación de los mismos. El análisis de contenido de dichos documentos permitió describir los conocimientos que eran objeto de la investigación en los participantes en el estudio y que permiten abordar los objetivos específicos de la misma.

En las siguientes secciones se describen los resultados del análisis, incluyendo tablas resumen de los resultados del grupo, así como algunos ejemplos de respuestas que permiten describir su conocimiento, así como mostrar algunas de las dificultades.

5.1 Conocimiento tecnológico del contenido (CTC) sobre correlación y regresión

La descripción y desarrollo de este conocimiento fue objeto del objetivo específico O1 y se aborda mediante el análisis de las dos primeras tareas que se realiza a continuación.

5.1.1 Ajuste de funciones a la esperanza de vida

De los veinticinco participantes, veinticuatro ajustaron, con ayuda de Excel, diferentes modelos de ajuste a la *esperanza de vida* de una o más de las variables independientes proporcionadas en el fichero. Otro participante no ajusta las funciones usando Excel, aunque indica el tipo de función de ajuste correctamente; por tanto, este participante muestra un

conocimiento del contenido correlación y regresión, pero no un conocimiento tecnológico del mismo.

La Tabla 1 expresa un resumen de los modelos de ajuste presentados, junto con el modelo o modelos previstos por el formador. El número de participantes que llega a ajustar una función plausible varía en las diferentes variables independientes.

Tabla 1 – Tipos de modelos ajustados por algunos participantes

Variable independiente	Modelo óptimo de función de ajuste a la Esperanza de vida	R ²	Número de participantes según modelo ajustado			
			Lineal	Logarítmico/ Exponencial	Polinómico	No ajusta
Índice desarrollo humano	Lineal	0,82	13		11	1
Producto Interior Bruto	Logarítmica	0,63	1	20	2	2
Tasa de fecundidad	Lineal	0,54	9	10	5	1
Tasa mortalidad niños	Exponencial	0,86	4	4	14	3
Gasto Público en salud	Polinómica	0,23	1	6	15	3
Índice de educación	Lineal	0,61	10		11	4
Población total	Incorrelacionada	0,00	4	6	4	11
% Población urbana	Lineal	0,38	11	2	9	3

Fuente: Elaboración propia para la investigación

Por ejemplo, la futura profesora CM determina todas las funciones de ajuste, y para decidir el mejor modelo utiliza elementos numéricos (coeficiente de determinación) y gráficos (Figura 2), como podemos ver en su argumento, que reproducimos a continuación.

Como vemos la variable que mejor describe la esperanza de vida es la tasa de mortalidad de niños, cuya función tiene como expresión algebraica $y = 0,0008x^2 - 0,3096x + 79,411$, puesto que si nos fijamos en el coeficiente de correlación R^2 es la que tiene mayor coeficiente de correlación. Además, lo podemos ver en la gráfica siguiente ya que se aprecia perfectamente como la línea de tendencia ajusta los datos con una precisión bastante buena (Alumno CM).

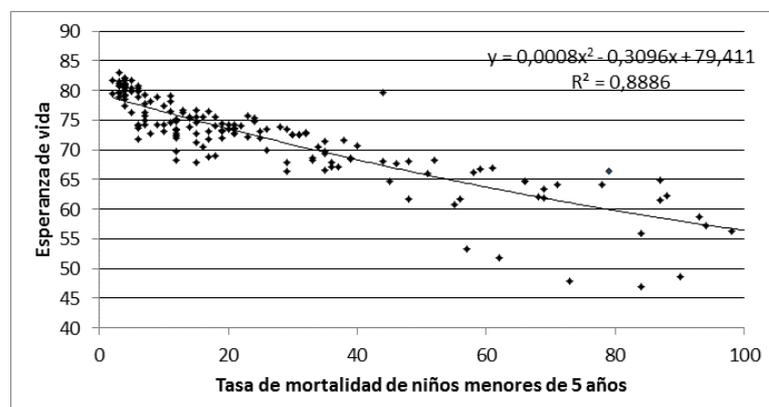


Figura 2 – Gráfica proporcionada por CM en su informe sobre funciones de ajuste
 Fuente: Alumno CM, participante de la investigación

No obstante, del buen conocimiento tecnológico del contenido mostrado, ha confundido coeficiente de correlación y de determinación (pues el programa ofrece el valor del coeficiente de determinación R^2 y no el valor del coeficiente de correlación R). Dicha confusión, que constituye un fallo en su conocimiento del contenido, también apareció en la investigación de Estepa y Sánchez-Cobo (2003). Otro fallo es que CM trata de ajustar un modelo a todas las variables, incluso a la *población total*, que no correlaciona con la *esperanza de vida* y tiene un coeficiente de correlación cercano a 0, por lo que no sería pertinente ajustar una función a esta variable. Aunque sólo incluimos acá su análisis de una variable, por razones de espacio, también en las otras variables ajustadas eligió un modelo no lineal, incluso en los casos en que el ajuste lineal es muy fuerte; posiblemente, trata de mejorar el ajuste y no percibe que el ajuste lineal ya es bueno para alguna de las variables, lo que atribuimos a su confusión mostrada entre coeficiente de determinación y coeficiente de correlación.

Los modelos ajustados por los futuros profesores no se reducen al lineal (Tabla 1) sino que se incluyen funciones logarítmicas, exponenciales y polinómicas de diferente grado, lo que muestra su conocimiento del contenido de diversas funciones elementales y su conocimiento tecnológico del uso de la hoja Excel para el trabajo con la regresión y correlación. En general, los participantes comparan el coeficiente de determinación de varios modelos para elegir el que maximice su valor, como muestra el siguiente ejemplo de MP:

Ahora probamos con las distintas líneas de tendencia para ver cuál es la que mejor describe la relación de la esperanza de vida con el índice de desarrollo humano, esto es, cuál tiene mayor coeficiente de determinación:

Exponencial: $R^2=0.795$

Lineal: $R^2=0.821$

Logarítmica: $R^2=0.8206$

Polinómica de orden 2: $R^2=0.8262$

Polinómica de orden 3: $R^2=0.8264$

Polinómica de orden 4: $R^2=0.8286$

Polinómica de orden 5: $R^2=0.8296$

Polinómica de orden 6: $R^2=0.8298$

Potencial: $R^2=0.8076$

La función que mejor se ajusta, es decir, que mayor coeficiente de determinación tiene es la polinómica de orden 6, aunque todas las polinómicas, la lineal y la logarítmica rondan en torno al 0.82 que es muy buen coeficiente de determinación: $y = -5713,9x^6 + 18609x^5 - 23406x^4 + 14038x^3 - 3919x^2 + 422,66x + 45,655$ (Alumno MP).

El futuro profesor no observa que aunque aumente la proporción de varianza explicada (dada por el coeficiente de determinación) del modelo lineal con los modelos polinómicos, dicho aumento no es significativo, por lo que sería, entonces, preferible utilizar el modelo lineal. Por tanto, realiza un uso acrítico del *software*, aplicando una regla rutinaria que puede

no ser la más adecuada para resolver un problema (BEN-ZVI; FRIEDLANDER, 1997). De hecho, MP ajustó finalmente ecuaciones polinómicas de orden seis a todas las variables, independientemente de la forma de la gráfica. En el ejemplo se muestra sólo su argumentación respecto a la primera variable, por razones de espacio.

El modelo no siempre coincide con el previsto en nuestro análisis (ver Tabla 1), pues nosotros hemos elegido entre los modelos que proporcionan un valor prácticamente idéntico (las dos primeras cifras decimales idénticas) el modelo de expresión algebraica más simple, generalmente el lineal o exponencial /logarítmico. Por el contrario, algunos futuros profesores prueban todos los modelos posibles que proporciona Excel y se centran, únicamente, en el aumento de este coeficiente utilizando modelos polinómicos, en su mayoría de grado seis (con un razonamiento similar al ejemplo mostrado anteriormente), que es el máximo grado que ofrece Excel, realizando lo que Ben-Zvi y Friedlander (1997) denominan un uso acrítico de la tecnología pues, aunque esta solución proporciona mayor varianza explicada, sería parcialmente correcta al ofrecer un modelo complejo. Observamos, en estos casos, cómo los futuros profesores se dividen en dos grupos: los que usan el criterio de máxima varianza explicada y los que optan por un modelo más simple (por ejemplo, el lineal) si la varianza explicada cambia poco.

Todos los participantes realizan un ajuste correcto, representando gráficamente los datos y cambiando la escala del diagrama de dispersión producido por Excel para visualizar los datos con facilidad. En consecuencia, muestran un Conocimiento Tecnológico del Contenido correlación y regresión adecuado (NIESS, 2005) y un nivel de lectura muy avanzado (lectura crítica de datos, descrita por Friel, Curcio y Bright (2001) del diagrama de dispersión.

Únicamente MJG muestra dificultades en la tarea, porque no cambia la escala para las diferentes variables, con lo cual, visualiza con dificultad la relación con la *esperanza de vida* con las otras variables y, además, no alcanza el nivel citado de lectura de gráficos. Este futuro profesor muestra menor Conocimiento Tecnológico sobre la correlación y regresión (NIESS, 2005) y opta por aplicar el modelo lineal en todas las variables, incluso en casos como el producto interior bruto en que la gráfica muestra claramente que el ajuste no es lineal. Por otro lado, no utiliza el coeficiente de determinación para valorar la bondad del ajuste. Recordemos que tanto en la investigación de Truran (1997) como en la de Estepa y Sánchez-Cobo (2003) hay una tendencia a pensar que la relación conjunta debe ser principalmente de tipo lineal, desechando otros posibles modelos de ajuste. Esto ocurre en nuestro estudio únicamente en este participante.

Varios participantes tratan de encontrar un modelo de ajuste para la relación con la Población total, a pesar de que existe independencia en los datos (por lo que el coeficiente de correlación es cero), que ellos mismos interpretaron correctamente, por lo que no parecen comprender que la regresión es útil únicamente cuando las variables tienen una correlación fuerte o al menos moderada. Otros participantes indican expresamente:

No hay expresión algebraica representativa (Alumnos AG y HD)

No hay función (Alumno AA).

Y el resto, simplemente no lo ajusta; todo ello confirma la buena percepción de la independencia por estos futuros profesores. Aun así, gran parte de la muestra (60%) trata de ajustar un modelo a esta variable, mostrando dificultad con las ideas de independencia e incorrelación así como el uso acrítico de la tecnología (BEN-ZVI; FRIEDLANDER, 1997).

5.1.2 Análisis de otras variables dependientes

Los veinticuatro futuros profesores que realizaron ajustes en la actividad anterior eligieron una variable dependiente diferente entre las proporcionadas por el formador en la hoja Excel, y ajustaron modelos considerando su relación con alguna de las independientes, usualmente en forma correcta. En general, una vez elegida una variable dependiente, la cruzan con el resto y determinan la mejor función de ajuste. En total se realizan 130 ajustes, lo que da una media de 5,4 ajustes por participante.

No se observa, en nuestros participantes, la confusión entre variable dependiente e independiente sugerida en la investigación de Estepa y Sánchez-Cobo (2003). Atribuimos esta diferencia a la mayor formación de nuestros participantes, todos ellos licenciados o ingenieros /arquitectos, que han trabajado la noción de variable dependiente e independiente en sus estudios tanto de estadística como de análisis matemáticos en sus carreras, mientras en el estudio citado se trataba de estudiantes de enfermería y empresariales en primer o segundo curso universitario.

La mayoría de los futuros profesores determinan correctamente un modelo de ajuste a sus variables dependientes, y generalmente, obtienen una conclusión sobre cuál de ellas proporciona mejor predicción. También, como en el siguiente caso, si ninguna sirve para predecir la variable dependiente, muestran una buena comprensión del significado del coeficiente de correlación y su relación con la bondad de ajuste:

Como podemos observar, no podemos predecir el gasto público de un país en función de ninguna de las variables, ya que, como podemos ver, el coeficiente de correlación es muy

pequeño en todos los casos (Alumno MA).

En la Tabla 2 se presenta una síntesis de los modelos ajustados por los futuros profesores. Como se ha indicado, cada uno eligió libremente la variable dependiente y, de hecho, todas las variables del fichero han sido tomadas como dependientes por algún participante excepto la Población total, que es independiente del resto de las variables. Los ajustes realizados son aceptables, lo que de nuevo muestra, en general, un buen conocimiento tecnológico y estadístico de la correlación y regresión por estos futuros profesores.

Tabla 2 – Número de análisis de otras variables dependientes

Variable Dependiente	Variable independiente*								Total
	Esperanza de vida	PIB	TF	TM	GPS	IE	PT	PU	
Índice desarrollo humano	4	5	4	4	4	4	3	4	32
Producto Interior Bruto	6		6	6	5	6	4	8	41
Tasa de fecundidad	1	1		1	1	3	1	1	9
Tasa mortalidad niños	1	1	1		3	1		1	8
Gasto Público en salud	2	2	2	2		4	1	4	17
Índice de educación	1	1	1	1	1			3	8
Población total									0
% Población urbana	3	2	2	2	2	2	2		15

*PIB: Producto Interior Bruto; TF: Tasa de fecundidad; TM: Tasa mortalidad niños; GPS: Gasto Público en salud; IE: Índice de educación; PT: Población total; PU: Población urbana.

Fuente: Elaboración propia para la investigación

También observamos que once futuros profesores ajustan la variable elegida con la población total, que presenta correlación cercana a cero. Así, por ejemplo, MA la cruza con el *gasto público en salud*, obteniendo una función exponencial $y = 6,7927 \cdot x^{0,086}$, aunque con coeficiente de determinación casi nulo (0,081). Observamos, de nuevo, dificultad en relacionar la idea de independencia con la utilidad de la regresión que sería una carencia en el Conocimiento del Contenido correlación y regresión.

5.2 Conocimiento Tecnológico-Pedagógico del Contenido (CTPC) sobre correlación y regresión

En la Actividad 3, quince de los veinticinco participantes seleccionaron nuevos datos del servidor de las Naciones Unidas y los descargaron en la hoja Excel para realizar nuevos estudios de correlación y regresión, mostrando, por tanto, su Conocimiento Tecnológico-Pedagógico del Contenido (CTPC) sobre correlación y regresión. Fueron menos que los que realizaron las actividades anteriores, pues resultó más complicado el obtener nuevos datos y

exportarlos para tratarlos con Excel. Encontraríamos en este sentido una carencia en el conocimiento tecnológico en general (NIESS, 2005), es decir, de la obtención de datos de Internet y su exportación a formato Excel de dichos participantes, que les puede dificultar la enseñanza de la estadística por medio de proyectos que tengan como base obtener datos reales desde Internet.

Variable Dependiente	Variáveis independientes
Esperanza de vida (MG)	Esperanza de vida en 1980, Esperanza de vida en 2012.
Esperanza de vida (ME)	Años de educación promedio, Ingreso nacional bruto (INB) per cápita, Tasa bruta de matriculación, Población bajo la línea de pobreza de ingresos, Superficie forestal, Tasa de alfabetización de adultos, Tasa de mortalidad maternal.
Índice desarrollo humano (IE)	Años esperados de escolarización, Años reales de escolarización, Esperanza de vida (mujeres), Hombres con al menos Educación Secundaria, Índice de desigualdad de género, Mujeres en el parlamento nacional, Mujeres con al menos Educación Secundaria, Tasa de escolaridad, Tasa de participación laboral (Hombres), Tasa de participación laboral (Mujeres), Ranking según Producto Interior Bruto.
Índice desarrollo humano (AG)	Desempleo juvenil, Hombres con al menos Educación Secundaria, Mujeres con al menos Educación Secundaria, Satisfacción con el trabajo, Tasa de escolaridad, Tasa de participación laboral (Hombres), Tasa de participación laboral (Mujeres).
Índice desarrollo humano (PG)	Años de escolarización Adultos, Tasa bruta de matriculación, Gasto en educación.
Tasa mortalidad infantil (HD)	Índice de desigualdad de género, Mujeres en el parlamento nacional, Población con al menos Educación Secundaria, Población con al menos Educación Secundaria (Mujeres), Tasa de participación laboral (Mujeres).
Coefficiente de desigualdad humana (MP)	Esperanza de vida, Gasto público en salud, Índice de desarrollo humano, Índice de educación, Tasa de fecundidad, Producto interior bruto, Tasa de mortalidad infantil, Población total.
Gasto Público en Investigación, Desarrollo e Innovación (GO)	Grado de electrificación, Número de abonados a servicios de telefonía móvil y fija, Número de investigadores, Número de Ordenadores Personales, Número de usuarios de Internet, Patentes registradas, Tasa de graduados en Ciencias e Ingeniería.
Índice de ingresos (JPA, VC)	Años educación promedio, Índice de salud, Población mujeres, Tasa participación trabajo mujeres/hombres.
Índice de pobreza multidimensional (PA)	Población en pobreza, Población en pobreza extrema, Población en riesgo de pobreza.
Porcentaje de población con estudios secundarios (CC)	% de Población vacunada, Número de usuarios de Internet, Producto interior bruto, Tasa de desempleo juvenil.
Porcentaje de mujeres en parlamento (JP)	Índice de desigualdad, Número de abonados a telefonía móvil, Tasa de mortalidad maternal.
Puntuación PISA en Comprensión lectora (BS)	% de la Población con al menos estudios secundarios.
Puntuación PISA en Ciencias (BS)	% de la Población con al menos estudios secundarios.
Puntuación PISA en Matemáticas (BS)	% de la Población con al menos estudios secundarios.
Tasa de empleo (AA)	Confianza en la gente, Confianza en el Gobierno Nacional, Desempleo juvenil, Pérdida en el IDH debido a la desigualdad, Percepción de la seguridad, Satisfacción general en la vida, Satisfacción con la libertad de elección, Satisfacción con el trabajo, Satisfacción con la comunidad, Tasa de homicidios, Trabajo infantil.

Cuadro 2 – Modelos de regresión ajustados a nuevos datos

Fuente: Elaboración propia para la investigación

Los participantes que completaron la actividad eligieron diferentes variables como dependiente; generalmente, el futuro profesor elige una sola y la cruza con varias independientes, como vemos en el Cuadro 2, donde, en la primera columna indicamos la variable dependiente y en la segunda, todas las independientes analizadas por el mismo participante. Seis de ellos usan variables proporcionadas por el formador como dependientes, pero la cruzan con nuevas variables, cuyos datos han descargado del servidor de las Naciones Unidas; entre ellas, variables relacionadas con la educación, la medida de la pobreza o el empleo. Así, IE selecciona siete variables independientes, nuevas, sobre igualdad de género, que cruza con el *índice de desarrollo humano* y obtiene todos los modelos de regresión correctamente, aunque no comenta sus conclusiones.

Once futuros profesores proponen otros temas de estudio, relacionados con la igualdad de género, como por ejemplo JP, quien toma como variable dependiente la proporción de mujeres en el parlamento y HD, quien relaciona la mortalidad infantil con varios indicadores de igualdad de género. Por su parte, GO elige como campo la investigación y el desarrollo tecnológico, tomando como variable dependiente el gasto en I+D+i, y una serie de variables independientes relacionadas con la tecnología e investigación.

Observamos un interés por el empleo y la educación en estos participantes. Así, JPA y VC relacionan el *índice de ingresos* con varias variables educativas; y AA analiza la *tasa de empleo* en función de varios indicadores de *desigualdad de género, educación y confianza en diversas instituciones*. Incluso, un participante se interesa por explicar los resultados de los estudios PISA, analizando las puntuaciones en matemáticas, comprensión lectora y ciencias respecto al abandono en la Educación Secundaria, como expone en sus comentarios:

Dentro de la gran masa de datos que podemos obtener de esta página, me resultan de especial interés aquellos referentes a educación. En concreto, vamos a estudiar cómo afecta un temprano abandono o fracaso escolar, reflejado con la variable referente al porcentaje de población que tiene, al menos, la Educación Secundaria Obligatoria, o equivalente, a los diferentes resultados que los diferentes países obtienen en el ámbito de matemáticas, lectura o ciencia en el informe PISA, que quizás sea el más conocido en el ámbito informativo en los países que lo realizan (Alumno BS).

En resumen, este grupo de futuros profesores ha mostrado, no sólo conocimiento tecnológico para seleccionar diferentes datos del servidor de las Naciones Unidas, sino también conocimiento del contexto para seleccionar variables relevantes, que les sirvan para estudiar fenómenos de su interés, planteando y resolviendo nuevos problemas. Con todo ello, muestran un alto grado de conocimiento del contenido estadístico, y en particular, de la correlación y regresión; asimismo, conocimiento tecnológico-pedagógico del contenido, pues

muestran la competencia para proponer nuevas investigaciones que pueden servir de base a la enseñanza de la estadística mediante proyectos que es, hoy, recomendada en las orientaciones curriculares (FRANKLIN et al., 2007; MECD, 2015).

6 Discusión e implicaciones para la formación de profesores

La investigación reseñada proporciona nuevos resultados sobre el conocimiento estadístico y conocimiento tecnológico de la estadística (NIESS, 2005) sobre correlación y regresión en una muestra de futuros profesores de Educación Secundaria y Bachillerato, complementando, de este modo, los escasos antecedentes sobre el tema.

Los participantes, en su mayoría, muestran buena comprensión de las propiedades de la correlación y regresión, así como de la bondad del ajuste y el coeficiente de determinación. Así mismo, utilizan las opciones de la hoja Excel para ajustar modelos de ajuste plausibles, basados en los anteriores conceptos. Aun así, algunos de ellos han hecho un uso acrítico de la tecnología o tratan de ajustar un modelo incluso en casos de clara independencia entre las variables.

Por la limitación de espacio, en este trabajo sólo se describen resultados de las tareas que se muestran en la Figura 1, que fueron parte del taller realizado en un curso de formación docente. Las soluciones a dichas tareas se discutieron una vez realizadas con los futuros profesores y se ampliaron con otras, con objeto de reforzar su conocimiento sobre el tema y proporcionar a los participantes en el taller oportunidades de aprendizaje que les permitieran superar las dificultades que hemos descrito.

Las actividades, junto con su posterior corrección y discusión en el aula, fueron útiles para desarrollar el Conocimiento del Contenido, Conocimiento Tecnológico y Conocimiento Tecnológico-Pedagógico de los participantes respecto a la correlación y la regresión, además de identificar sus conocimientos iniciales como sobre el tema.

El taller resultó interesante a los futuros profesores, que pudieron razonar con datos reales sobre variables relevantes en la sociedad actual y materializar la distinción entre distintos tipos de función de ajuste (lineales, exponenciales, polinómicas), dando sentido al coeficiente de determinación como medida de la bondad de dicho ajuste. Al trabajar con fuentes internacionales, se aumentó el interés de los participantes y su percepción de la utilidad de la estadística en el desarrollo de un país, mejorando sus actitudes y promoviendo su iniciativa hacia la innovación en la enseñanza.

Como se ha indicado, en el taller planteado se les mostró un ejemplo del trabajo con proyectos, que pueden utilizar con los estudiantes para recorrer un estudio estadístico completo: problema, datos, análisis y conclusiones (BATANERO; BOROVCNICK, 2016). Proporciona, igualmente, una actividad de modelización, donde el interés se encuentra en analizar la dependencia y, en caso de ser alta, encontrar el modelo de ajuste más adecuado (lineal, logarítmico etc.). Con ello, el futuro profesor puede comprender que, como explica Henry (1997, p.78) “un modelo es una interpretación abstracta, simplificada e idealizada de un objeto del mundo real, de un sistema de relaciones o de un proceso evolutivo que surge de una descripción de la realidad”.

Las actividades permiten, también, analizar la diferencia entre el modelo (ecuación de regresión) y la realidad reflejada en los datos bivariantes. Asimismo, permiten ejercitar la actividad de traducción de los resultados del trabajo matemático realizado con el modelo (función de ajuste) a la realidad modelizada que sería la intensidad y tipo de dependencia entre las variables y el uso de la función de regresión para la predicción de valores de la variable dependiente en función de la independiente (EICHLER; VOGEL, 2014).

Estos últimos puntos forman parte del Conocimiento Tecnológico-Pedagógico del tema, que era escaso en los futuros profesores de la muestra, pues casi ninguno había trabajado previamente con datos reales ni el método de enseñanza por proyectos. Desconocían, asimismo, las investigaciones didácticas sobre el tema, que fueron discutidas con ellos a la vez que se debatieron las soluciones a las tareas. Nuestra reflexión final es que la mejora de la enseñanza de la correlación y regresión depende de la formación de los profesores. Talleres y actividades, como las descritas en este trabajo, pueden contribuir a reforzar dicha formación.

Reconocimiento

Trabajo realizado en el marco del proyecto de investigación EDU2016-74848-P (AEI, FEDER) y grupo de investigación FQM126 de la Junta de Andalucía.

Referencias

BALL, D. L.; THAMES, M. H.; PHELPS, G. Content knowledge for teaching. What makes it special? **Journal of Teacher Education**, Londres, v. 99, n. 5, p. 389-407. 2008.

BATANERO, C.; BOROVCNIK, M. **Statistics and probability in high school**. Rotterdam: Sense Publishers, 2016.

- BATANERO, C.; GODINO, J. D.; ESTEPA, A. Building the meaning of statistical association through data analysis activities. En 22 Conference of the International Group for the Psychology of Mathematics Education. **Proceedings**. Stellenbosch, South Africa: Universidad de Stellenbosh, 1998. p. 221-236.
- BEN-ZVI, D.; FRIEDLANDER, A. Statistical thinking in a technological environment. En: GARFIELD J.; G. BURRILL (Ed.). **Research on the role of technology in teaching and learning statistics**. Voorburgo: International Statistical Institute, 1997. p. 54-64.
- BIEHLER, R et al. Technology for enhancing statistical reasoning at the school level. En: CLEMENTS, M. A. et al. (Ed.). **Third international handbook of mathematics education**. New York: Springer, 2013. p. 643-689.
- BLÖMEKE, S. et al. **International perspectives on teacher knowledge, beliefs and opportunities to learn**. 1ª. ed. Dordrecht: Springer, 2014.
- CASEY, S. A. Subject matter knowledge for teaching statistical association. **Statistics Education Research Journal**, New Zealand, v. 9, n. 2, p. 50-68. 2010.
- CASEY, S. A.; WASSERMAN, N. H. Teachers' knowledge about informal line of best fit. **Statistics Education Research Journal**, New Zealand, v. 14, n. 1, p. 8-35, 2015.
- EICHLER, A.; VOGEL, M. Three approaches for modelling situations with randomness. En: CHERNOFF, E. J.; SRIRAMAN, B. (Ed.). **Probabilistic Thinking**. Presenting plural perspectives. Berlin, Heidelberg: Springer, 2014. p. 75-99.
- ESTEPA, A. Interpretación de los diagramas de dispersión por estudiantes de Bachillerato. **Enseñanza de las Ciencias**, Barcelona, v. 26, n. 2, p. 257-270, 2008.
- ESTEPA, A.; BATANERO, C. Judgments of correlation in scatter plots: Students' intuitive strategies and preconceptions. **Hiroshima Journal of Mathematics Education**, Tokio, v. 4, p. 25-41, 1996.
- ESTEPA, A.; SÁNCHEZ-COBO, F. Evaluación de la comprensión de la correlación y regresión a partir de la resolución de problemas. **Statistics Education Research Journal**, New Zealand, v. 2, n. 1, p. 54-68, 2003.
- FERNÁNDES, J. A.; BATANERO, C.; CONTRERAS, J. M.; DÍAZ, C. A simulação em Probabilidades e Estatística: potencialidades e limitações. **Quadrante**, Lisboa, v. XVIII, n. 1-2, p. 161-183, 2009.
- FRANKLIN, C. et al. **Guidelines for assessment and instruction in statistics education (GAISE) report: A preK- 12 curriculum framework**. 1º ed. Alexandria, VA: American Statistical Association. Disponible en: www.amstat.org/education/gaise/, 2007. Acceso en: 3 de Junio de 2017.
- FRIEL, S.; CURCIO, F.; BRIGHT, G. Making sense of graphs: critical factors influencing comprehension and instructional implications. **Journal for Research in Mathematics Education**, Reston, VA, v. 32, n. 2, p. 124-158, 2001.
- HALL, J. Engaging teachers and students with real data: Benefits and challenges. En: BATANERO, C.; BURRILL, G.; READING, C. (Ed.). **Teaching statistics in school mathematics**. Challenges for teaching and teacher education. A joint ICMI/IASE Study. New York: Springer, 2011. p. 335-346.
- HENRY, M. Notion de modèle et modélisation en l'enseignement. En: **Enseigner les probabilités au lycée**. Reims: Commission Inter-IREM, 1997. p. 77-84.

- LEE, H. S.; HOLLEBRANDS, K. F. Characterising and developing teachers' knowledge for teaching statistics with technology. En: BATANERO, C.; BURRILL, G.; READING, C. (Ed.). **Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education**. Springer Netherlands, 2011. p. 359-369.
- LLINARES S.; KRAINER K. Mathematics (student) teachers and teacher educators as learners. En: GUTIERREZ, A.; BOERO, P. (Ed.). **Handbook of research on the psychology of mathematics education**. Rotherdam: Sense Publishers, 2006. p. 429- 459.
- MECD, MINISTERIO DE EDUCACIÓN, CULTURA Y DEPORTE. **Real Decreto 1105/2014**, de 26 de diciembre por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato. Madrid: Autor, 2015.
- NISS, M. L. Preparing teachers to teach science and mathematics with technology: Developing a technology pedagogical content knowledge. **Teaching and Teacher Education**, Netherlands, v. 21, p. 509 -523, 2005.
- ORGANIZACIONE DE LAS NACIONES UNIDAS. **Informe sobre el desarrollo humano. Desarrollo para todas las personas**. Nueva York: Autor, 2016.
- PRATT, D.; DAVIES, N.; CONNOR, D. The role of technology in teaching and learning statistics. En: BATANERO, C.; BURRILL, G.; READING, C. (Ed.). **Teaching statistics in school mathematics. Challenges for teaching and teacher education**. A Joint ICMI/IASE Study. New York: Springer, 2011. p. 97-107.
- QUINTAS, S.; FERREIRA, R.; OLIVEIRA, H. O conhecimento didático de estatística de duas professoras de matemática sobre dados bivariados. **Bolema**, Rio Claro, v. 29, n. 51, p. 284-306, 2015.
- SHULMAN, L. S. Those who understand: Knowledge growth in teaching. **Educational Researcher**, Routledge, v. 15, n. 2, p. 4-14, 1986.
- TATTO, M. T.; SENK, S. The mathematics education of future primary and secondary teachers: Methods and findings from the Teacher Education and Development Study in Mathematics. **Journal of Teacher Education**, New York, v. 62, n. 2, p. 121-137, 2011.
- TRURAN, J. M. Understanding of association and regression by first year economics students from two different countries as revealed in responses to the same examination questions. En: GARFIELD, J.; TRURAN, J. M. (Ed.). **Research papers on stochastics educations from 1997**. Minneapolis, MN: University of Minnesota, 1997. p. 205-212.
- WILD, C. J.; PFANNKUCH, M. Statistical thinking in empirical enquiry. **International Statistical Review**, Voorburg, v. 67, n. 3, p. 223-265, 1999.
- ZAPICO, M. Interrogantes acerca de análisis de contenido y del discurso en los textos escolares. En: MINEDUC (Ed.). **Primer Seminario Internacional de Textos Escolares**. Santiago: Ministerio de Educación, 2006. p. 149-155.
- ZIEFFLER, A.; GARFIELD, J. Modeling the growth of students' covariational reasoning during an introductory statistics course. **Statistics Education Research Journal**, New Zealand, v. 8, n. 1, p. 7-31, 2009.

Submetido em 09 de Maio de 2017.
Aprovado em 10 de Outubro de 2017.