



## Conocimiento del contenido sobre correlación y regresión de futuros profesores

### Content knowledge of correlation and regression in prospective teachers

 **Carmen Batanero** batanero@ugr.es  
Universidad de Granada, España

 **María M. Gea** mmgea@ugr.es  
Universidad de Granada, España

 **Pedro Arteaga** parteaga@ugr.es  
Universidad de Granada, España

 **J.M. Contreras** jmcontreras@ugr.es  
Universidad de Granada, España

 **Carmen Díaz** carmen.diaz@dpsi.uhu.es  
Universidad de Huelva, España

Conocimiento del contenido sobre correlación y regresión de futuros profesores

Revista Latinoamericana de Investigación en Matemática Educativa, RELIME, vol. 21, núm. 3, 2018

Comité Latinoamericano de Matemática Educativa

#### Cita recomendada:

Batanero, C., Gea, M., Arteaga, P., Contreras García, J.M. & Díaz, C. (2018). Conocimiento del contenido sobre correlación y regresión de futuros profesores. *Revista Latinoamericana de Investigación en Matemática Educativa*, 21(3), 325-348. <https://doi.org/10.12802/relime.18.2134>

Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial 4.0 Internacional.

**Recepción:** 28 Febrero 2018

**Aprobación:** 28 Octubre 2018

**DOI:** <https://doi.org/10.12802/relime.18.2134>

#### Financiamiento

Fuente: MINECO

Nº de contrato: EDU2013-41141-P

#### Financiamiento

Fuente: AEI, FEDER

Nº de contrato: EDU2016-74848-P

#### Financiamiento

Fuente: FECYT

Nº de contrato: FCT-16-10974

#### Financiamiento

Fuente: Junta de Andalucía

Nº de contrato: FQM126

Conocimiento del contenido de correlación y regresión en futuros profesores



#### Resumen:

El objetivo de esta investigación fue identificar el conocimiento del contenido: correlación y regresión (común, avanzado y especializado) en una muestra de 65 futuros profesores de Educación Secundaria Obligatoria y Bachillerato. Para ello se analizaron sus respuestas escritas en diferentes tareas relacionadas con un proyecto estadístico basado en datos de las Naciones Unidas. Los resultados indican buena estimación de la correlación e identificación de la función ajuste (conocimiento común), corrección en la ordenación de las variables explicativas según su poder predictivo y mayor dificultad para explicar la correlación por relaciones diferentes a la causalidad (conocimiento avanzado) y un análisis razonable de la idoneidad didáctica del proyecto (conocimiento especializado). Se informa también de algunos sesgos y del efecto de las variables de tarea sobre las respuestas.

**Palabras clave:**

Correlación y regresión, futuros profesores, conocimiento del contenido, evaluación, Educación Secundaria Obligatoria y Bachillerato.

**Abstract:**

This research aimed to identify content knowledge (common, advanced and specialised) of correlation and regression in a sample of 65 prospective secondary and high school teachers. The participants' written productions in different tasks proposed within a statistical project based on United Nations Data were analysed. Results suggest a good estimation of correlation and identification of the model of fit (common knowledge), correct ordering of explanatory variables according to their predictive value, some difficulty in explaining correlation by relationships different from causality (advanced knowledge), and a reasonable analysis of the project didactic suitability (specialised knowledge). We also inform of some biases, as well as the effect of task variables in the responses.

**Keywords:**

Correlation and regression, prospective teachers, content knowledge, assessment, secondary and high school levels.

**Resumo:**

Esta investigação tinha por finalidade identificar o conhecimento do conteúdo correlação e regressão (comum, avançado e especializado) numa amostra de 65 futuros professores do 3.º ciclo do ensino básico e ensino secundário. Para tal, analisaram-se as suas respostas escritas a diferentes tarefas, dentro de um projeto estatístico baseado em dados das Nações Unidas. Os resultados indicam uma boa capacidade de estimação da correlação e de identificação da função de ajuste (conhecimento comum), correção na ordenação das variáveis explicativas segundo o seu poder preditivo e maior dificuldade em explicar a correlação por relações diferentes da causalidade (conhecimento avançado) e razoável análise da idoneidade didática do projeto (conhecimento especializado). Relatam-se também alguns enviesamentos e o efeito das variáveis de tarefa sobre as respostas.

**Palavras-chave:**

Correlação e regressão, futuros professores, conhecimento do conteúdo, avaliação, 3º ciclo do ensino básico e ensino secundário.

**Résumé:**

L'objectif de cette recherche était d'identifier la connaissance du contenu (commune, avancé et spécialisée) de la corrélation et la régression de 65 futurs enseignants de l'éducation secondaire et du baccalauréat. Pour atteindre cet objectif, leurs réponses écrites à différentes questions dans un projet statistique basé sur les données des Nations Unies ont été analysées. Leur travail écrit indique une bonne capacité pour estimer la corrélation et identifier la fonction d'ajustement (connaissance commune), la correcte identification de l'ordre des variables explicatives selon son pouvoir prédictif, une grande difficulté à expliquer la corrélation par relations autres que la causalité (connaissance avancé), et un raisonnable analyse de la pertinence didactique du projet (connaissance spécialisée). On informe également de quelques biais et l'effet des variables de tâche sur les réponses des enseignants.

**Mots clés:**

Corrélation et régression, futurs enseignants, connaissance du contenu, évaluation, enseignement secondaire obligatoire et baccalauréat.

**1. INTRODUCCIÓN**

El objetivo de esta investigación fue identificar algunos elementos del conocimiento matemático común, avanzado y especializado (Ball, Lubienski y Mewborn, 2001) requerido por los futuros profesores de Educación Secundaria y Bachillerato españoles para enseñar la correlación y la regresión. Dicho tema amplía el conocimiento sobre estadística descriptiva unidimensional y permite aplicar las funciones conocidas por el estudiante en una actividad de modelización (Engel y Sedlmeier, 2011; Gea, 2014; Puig y Monzó, 2013). Es también un método ampliamente utilizado en ciencia y tecnología y base de muchas técnicas estadísticas.

A pesar de esta relevancia, la investigación previa advierte de errores y sesgos en la estimación de la correlación (Estepa y Sánchez Cobo, 2001); el más conocido, denominado *correlación ilusoria* (Chapman, 1967), consiste en que los sujetos se guían por sus ideas previas, más que por los datos, al estimar la correlación. Otras investigaciones describen concepciones erróneas (Batanero, Godino y Estepa, 1998; Castro-Sotos, Vanhoof, Van Den Noortgate y Onghena, 2009), como la *concepción causal*, por la que el sujeto equipara correlación con causa-efecto; la concepción unidireccional, por la que no se acepta una correlación inversa, o la concepción local, consistente en estimar la correlación con sólo una parte de los datos (Estepa, 2007, 2008). Por otra parte, en el análisis de regresión, algunos estudiantes confunden las variables dependiente e independiente (Estepa y Sánchez Cobo, 2003).

La enseñanza de la correlación y regresión en España se realiza en el primer curso de Bachillerato, en las modalidades de Ciencias y de Humanidades y Ciencias Sociales (MECD, 2015). Además, se introduce la construcción e interpretación de diagramas de dispersión en cuarto curso de la Educación Secundaria Obligatoria.

Puesto que una buena enseñanza requiere una preparación adecuada del profesorado, en esta investigación nos hemos interesado por identificar los conocimientos matemáticos (común, avanzado y especializado) de los futuros profesores sobre este tema. En las siguientes secciones se desarrollan los fundamentos y metodología del estudio, se analizan sus resultados y se discuten las implicaciones para la formación de profesores.

## 2. FUNDAMENTOS

### 2.1 MARCO TEÓRICO

En su análisis del conocimiento requerido por los profesores para tratar con éxito la enseñanza, Shulman (1986) describió el conocimiento del contenido, el conocimiento pedagógico del contenido y el conocimiento del currículo.

Centrándonos únicamente en el conocimiento del contenido matemático, nos basamos en la descomposición que Ball y sus colaboradores (Ball, Lubienski y Mewborn, 2001; Hill, Ball y Schilling, 2008) proponen en su modelo de conocimiento matemático para la enseñanza (MKT):

- *El conocimiento común del contenido* es el que posee una persona (no necesariamente profesor) después de haber estudiado el tema; se trataría del conocimiento que el profesor ha de enseñar a sus alumnos. Por ejemplo, ser capaz de identificar el signo y estimar la intensidad de la correlación o identificar el tipo de función de ajuste, a partir de un diagrama de dispersión.
- *El conocimiento matemático avanzado*, que los autores denominan conocimiento en el horizonte matemático, va más allá del conocimiento común; incluye conocimiento del tema a un nivel superior y las conexiones con otras materias. En nuestro estudio, consideraremos como parte de este conocimiento ordenar una serie de variables explicativas según su poder de predicción de una variable respuesta y diferenciar correlación y causalidad, al explicar la existencia de correlación mediante diferentes tipos de relaciones entre las variables.
- *El conocimiento especializado del contenido* es el que aplica el profesor para articular tareas de enseñanza (Hill, Ball y Schilling, 2008); un ejemplo, según Verdú, Callejo y Márquez (2014), se presenta al examinar y comprender procedimientos de resolución de problemas no usuales. En nuestro estudio, se pidió a los futuros profesores que analizaran la idoneidad epistémica de un proyecto estadístico, para lo cual debieron identificar el contenido matemático utilizado en el mismo.

La idoneidad epistémica de un proceso de estudio es parte de la *idoneidad didáctica* (Godino, Contreras y Font, 2006) que los autores introducen para valorar situaciones de aprendizaje y enseñanza de las matemáticas y tiene los siguientes componentes: idoneidad epistémica, cognitiva, interaccional, mediacional, emocional y ecológica. La idoneidad epistémica se define como la adecuación de los contenidos matemáticos en un proceso de enseñanza respecto a un contenido de referencia, fijado en las directrices curriculares y por el significado del tema dentro de la matemática. Para estudiarla, Godino (2013) propone una serie de descriptores que se presentan en la Sección 6. Como último punto de nuestro trabajo, analizaremos las respuestas escritas de los futuros profesores al utilizar tales descriptores para valorar la idoneidad epistémica de un proyecto estadístico para identificar algunos elementos de su conocimiento especializado del contenido.

### 2.2. ANTECEDENTES

Como se indicó en la introducción, la investigación previa ha descrito numerosas dificultades con el tema, aunque son pocas las investigaciones que han analizado el conocimiento de los profesores sobre correlación y regresión.

El primero de estos trabajos fue realizado por Estepa y sus colaboradores (Batanero et al., 1998; Estepa y Batanero, 1995), quienes llevaron a cabo una experiencia de enseñanza (incluyendo la correlación y la regresión) con 22 futuros profesores de educación primaria. Entre otras dificultades, citan la identificación de la correlación inversa, y la confusión entre la variable dependiente e independiente en la regresión. También reportaron estrategias incorrectas en la interpretación de los diagramas de dispersión, por ejemplo, usar sólo parte de los datos. Identifican las siguientes variables relevantes en el estudio del tema: intensidad y signo de la correlación y tipo de función de ajuste. Otras variables que inciden en la dificultad del tema son la existencia o no de teorías previas sobre el tipo de relación en los datos (Estepa y Sánchez-Cobo, 2001) y tipo de relación entre las variables que puede explicar la correlación (Barbancho, 1992).

Casey (2010) usa el marco teórico del MKT para analizar la enseñanza sobre correlación y regresión desarrollada por tres profesores en ejercicio. Realiza un estudio muy completo del conocimiento del contenido, abarcando incluso los aspectos inferenciales, que no tenemos en cuenta en esta investigación. Casey y Wasserman (2015) analizan el conocimiento de 19 profesores en servicio sobre la enseñanza del ajuste informal de una recta de regresión. Sin embargo, los autores no diferencian en sus resultados los componentes común, avanzado y especializado de este contenido, ni tratan de identificar el conocimiento de los profesores, sino sólo el conocimiento matemático teórico que les sería necesario en la enseñanza. Por otro lado, nosotros usaremos tareas no tenidas en cuenta por Casey.

Quintas, Ferreira y Oliveira (2015) analizan el conocimiento pedagógico del contenido (en el modelo MKT) de dos profesoras con amplia experiencia, mientras explican la correlación y la regresión en un curso de secundaria. Las autoras describen ejemplos de cómo las profesoras observadas usaron su conocimiento de la enseñanza y del estudiante en las lecciones desarrolladas. Observaron también errores de las profesoras en el conocimiento común del contenido; más concretamente, ninguna de ellas advirtió la existencia de dos rectas diferentes de regresión y pidieron a los alumnos usar siempre la recta de regresión de Y sobre X, independientemente de cuál fuera la variable que se quisiera predecir.

Utilizamos también el estudio de Arteaga, Batanero, Cañadas y Gea (2012), quienes evalúan los conocimientos de futuros profesores sobre gráficos estadísticos siguiendo el modelo MKT. Como parte de su estudio, utilizaron el análisis de la idoneidad epistémica de un proyecto estadístico realizado por futuros profesores de educación primaria para valorar su conocimiento especializado del contenido. Nosotros emplearemos el mismo método para evaluar el conocimiento especializado sobre la correlación y regresión de los participantes en nuestra investigación. Igualmente, consideramos el trabajo de Puig y Monzó (2013), que incluye la regresión y la correlación como parte de la modelización.

Aunque los trabajos citados fundamentan el nuestro y proporcionan 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, la explicación de la relación entre las variables —diferente a la causalidad— o el análisis de los contenidos sobre correlación y regresión de las tareas por parte de los participantes. Éstas serán aportaciones originales en nuestro trabajo.

### 3. MÉTODO

#### 3.1 MUESTRA Y CONTEXTO

El estudio se desarrolló con 65 estudiantes españoles que se preparaban para ser futuros profesores dentro del Máster de Formación del Profesorado de Educación Secundaria y Bachillerato, especialidad de Matemática. De dicho conjunto de estudiantes, 56% eran licenciados en Matemáticas o Estadística, y el resto había cursado alguna ingeniería, arquitectura u otras ramas de ciencias. Todos habían estudiado alguna asignatura de estadística, y 57% tenían experiencia docente. Estuvieron divididos en dos grupos de aproximadamente el mismo tamaño, que trabajaron con el mismo profesor y método.

Las tareas propuestas se realizaron dentro de una asignatura de Didáctica de la Matemática, a lo largo de tres sesiones de dos horas cada una. En las dos primeras sesiones, dirigidas a identificar y desarrollar el conocimiento común y avanzado del contenido de la correlación y la regresión, los participantes resolvieron un proyecto estadístico (véase sección 3.2). En la tercera, dedicada a identificar y desarrollar su conocimiento especializado, los estudiantes analizaron la idoneidad didáctica de dicho proyecto. En la sección 3.2 se analizan las tareas propuestas y en las secciones 4 a 6 las respuestas a las mismas.

#### 3.2 MATERIAL Y TAREAS PROPUESTAS

El proyecto estadístico propuesto a los participantes expuso la pregunta “¿cuáles son los factores que más influyen en la esperanza de vida al nacer en un país?”, la cual permitió plantear diferentes tareas. Los datos para elaborar los gráficos mostrados a los participantes se tomaron de los utilizados en la elaboración de los Informes sobre Desarrollo Humano por las Naciones Unidas, disponibles en el servidor (<http://hdr.undp.org/es/data>).

Las variables utilizadas en el proyecto son indicadores internacionales de desarrollo humano y se presentan en la Tabla I junto a los criterios que motivaron su selección y tomaron en cuenta los factores que inciden en la dificultad de las tareas de correlación y regresión (Batanero et al., 1998; Castro-Sotos, et al., 2009; Estepa, 2007; 2008 y Estepa y Sánchez-Cobo, 2001). En primer lugar, se tuvo en cuenta el signo y la intensidad de la correlación, brindando así la posibilidad de analizar correlaciones de signo positivo y negativo y de diferentes grados de intensidad. La relación de dependencia entre las variables no se limitó a la regresión lineal, sino incluyó otros modelos de ajuste con funciones que se estudian en la Educación Secundaria. Igualmente, se consideraron variables para las que la correlación observada entre los datos contradice o confirma las posibles expectativas (teorías previas) de los participantes (Chapman, 1967). Siguiendo a Barbancho (1992), se utilizaron también relaciones que pueden ser explicadas por causa y efecto, y otras debidas a interdependencia o dependencia indirecta.

| Variable                                    | $r^1$ | Modelo de ajuste | Explicación de la dependencia | Teorías previas frente a los datos |
|---|-------|------------------|-------------------------------|------------------------------------|
| Índice de Desarrollo Humano                 | 0,91  | Lineal           | Interdependencia              | Coincide                           |
| PIB per cápita                              | 0,61  | Logarítmica      | Dependencia indirecta         | Coincide                           |
| Tasa de fertilidad en adolescentes          | -0,73 | Lineal           | Dependencia indirecta         | No hay                             |
| Tasa de mortalidad en menores de cinco años | -0,92 | Exponencial      | Causa-efecto                  | Coincide                           |
| Gasto en salud pública                      | 0,38  | Polinómica       | Causa-efecto                  | Más débil de lo esperado           |
| Índice de educación                         | 0,78  | Lineal           | Dependencia indirecta         | No hay teoría previa               |
| Población total                             | 0,004 | Independencia    | Independencia                 | No hay teoría previa               |
| Población urbana                            | 0,62  | No hay modelo    | Dependencia indirecta         | Contraria                          |

<sup>1</sup> Coeficiente correlación entre la esperanza de vida y la variable en el conjunto de datos.

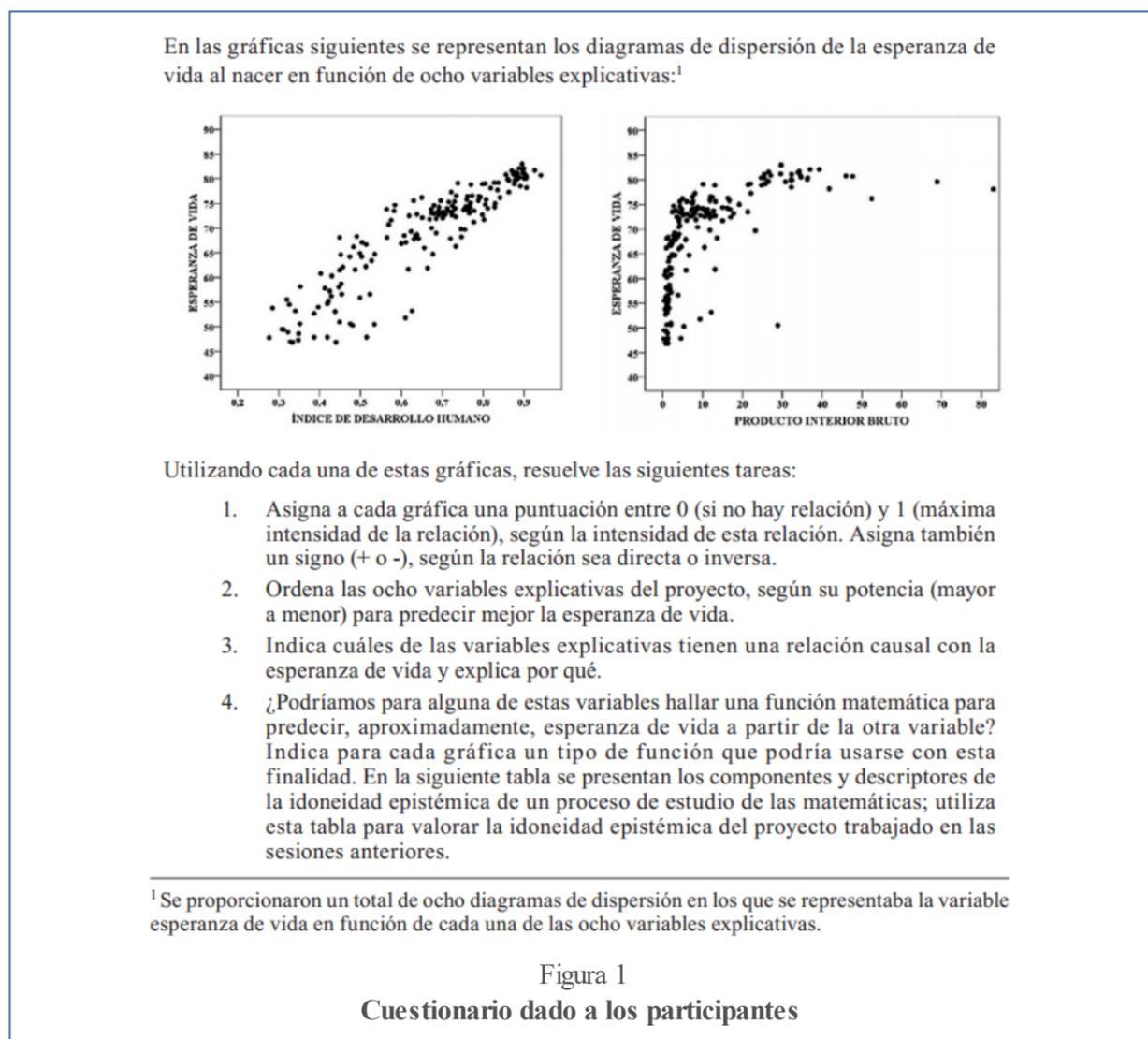
En la primera sesión, el formador de profesores explicó a los estudiantes la forma en que se recogieron los datos y se aseguró de que se comprendía la utilidad y el método de cálculo de cada variable. Se realizaron también algunas actividades iniciales de interpretación de gráficos y resúmenes estadísticos de la distribución de la esperanza de vida, que no son objeto de este artículo. Una vez que los estudiantes se familiarizaron con las variables, se les proporcionó, sucesivamente, cada una de las tareas del cuestionario mostrado en la figura 1, en un formato con espacio suficiente para incluir cada solución. La última media hora de la primera sesión se dedicó a resolver la tarea 1; las tareas 2 a 4 se resolvieron en la segunda, y la tarea 5, en la tercera. Cada participante las respondió por escrito, individualmente. Una vez recogidas las respuestas a cada una de las tareas, se debatieron las soluciones con el fin de resolver las posibles dificultades y desarrollar el conocimiento matemático de los participantes.

En la tarea 1 se pidió a los participantes estimar el valor absoluto del coeficiente de correlación y asignarle un signo. Esto resultó similar a otras tareas propuestas en las investigaciones sobre comprensión de la correlación (por ejemplo, Estepa, 2007, 2008 y Moritz, 2004). Al igual que la tarea 4 (identificar una función de ajuste), aparece con frecuencia en los textos de Bachillerato. Por tanto, nos posibilitaron identificar una parte del *conocimiento común* del tema de los futuros profesores (y posteriormente desarrollarlo en la discusión conjunta).

En la tarea 2 se solicitó a los futuros profesores que ordenaran las variables analizadas según su poder de predicción de la variable esperanza de vida, para lo cual los futuros profesores debían observar la mayor o menor dispersión de los puntos del diagrama de dispersión respecto a la tendencia, siendo la bondad de ajuste (o poder de predicción) mayor cuanto menor es la dispersión. La proporción de varianza explicada por la regresión se mide por el coeficiente de determinación. Dicho coeficiente es igual al cuadrado del coeficiente de correlación, en caso de dependencia lineal. En la tarea 3 se pidió a los participantes analizar si la correlación observada en cada caso era una relación causal, y se buscara, si las hubiera, otras posibles explicaciones de correlación. Estas tareas son algo más complejas de lo que se incluye habitualmente en la enseñanza, pero el profesor debe saber realizarlas. Las utilizamos para valorar su *conocimiento avanzado* del contenido.

La tarea 5 es específica del trabajo del profesor, quien ha de ser capaz de valorar la idoneidad de cualquier actividad planteada a sus alumnos; para resolverla ha de identificar los contenidos matemáticos implicados en el proyecto. Siguiendo a Verdú, Callejo y Márquez (2014), se utilizaron los resultados de este análisis para evaluar el *conocimiento especializado* de los participantes del contenido.

Una vez recogido el informe con las respuestas a las tareas que entregó cada participante, se realizó un estudio de dichas respuestas que se recoge en los siguientes apartados.



#### 4. CONOCIMIENTO COMÚN DEL CONTENIDO

##### 4.1 ESTIMACIÓN DE LA CORRELACIÓN A PARTIR DEL DIAGRAMA DE DISPERSIÓN

En primer lugar, se analizaron el valor absoluto y el signo del coeficiente de correlación que los futuros profesores asignaron a cada una de las gráficas (tarea 1). En la Tabla II se proporcionan los resultados globales: media del valor absoluto estimado para el coeficiente de correlación en las diferentes variables y porcentaje de participantes que asignó un signo correcto. Se acompaña de los valores reales del coeficiente de correlación en los datos (segunda columna) para una mejor interpretación de los resultados.

Observamos que el signo de la correlación en los datos fue identificado fácilmente. En el caso de independencia (población total), la mayoría indicó correctamente que la independencia supone ausencia de signo. Sin embargo, algunos futuros profesores no respondieron o no identificaron correctamente el signo para esta variable o para el gasto en salud pública, que es la única variable con intensidad baja. Globalmente, los resultados son mucho mejores que los proporcionados con el mismo tipo de tarea por otros estudiantes (e.g., Estepa, 2007, 2008 y Moritz, 2004).

Las estimaciones de la intensidad del coeficiente son también buenas, pues los valores medios en la muestra de futuros profesores se aproximan, en general, a la correlación observada en los datos. La precisión de la estimación es mayor para las correlaciones altas, y dentro de ellas, de menor calidad para las negativas que para las positivas, lo que coincide con los resultados expuestos por Estepa (2007), Sánchez Cobo, Estepa y Batanero (2000) y Moritz (2004).

**Tabla II**  
**Valor absoluto estimado y signo asignado al coeficiente de correlación para cada variable explicativa**

| Variable explicativa                        | Valor r | Valor absoluto estimado |                   | Porcentaje de futuros profesores que identifican el signo |
|---|---------|-------------------------|-------------------|---|
|   |         | Media                   | Desviación típica |   |
| Índice de Desarrollo Humano                 | 0,91    | 0,88                    | 0,08              | 98,5  |
| PIB per cápita                              | 0,61    | 0,71                    | 0,18              | 95,4  |
| Tasa de fertilidad en adolescentes          | -0,73   | 0,52                    | 0,22              | 93,8  |
| Tasa de mortalidad en menores de cinco años | -0,92   | 0,82                    | 0,23              | 96,9  |
| Gasto en salud pública                      | 0,38    | 0,39                    | 0,23              | 83,1  |
| Índice de educación                         | 0,78    | 0,68                    | 0,18              | 96,9  |
| Población total                             | 0       | 0,05                    | 0,14              | 80,8  |
| Población urbana                            | 0,62    | 0,39                    | 0,21              | 92,3  |

Los futuros profesores estimaron un valor medio del coeficiente de correlación cercano a cero en el caso de independencia (población total). En este punto, los resultados son mejores que en el estudio de Estepa (2008), en el que sólo 72% de los estudiantes identificó la independencia perfecta en un diagrama de dispersión en el que asignaron un coeficiente de correlación cercano a cero.

Se puede observar cierta influencia de las teorías previas sobre la menor precisión de la estimación de la correlación, cuando los datos se muestran en contra de esas teorías (población urbana). Igualmente se aprecia una sobreestimación de la correlación en el caso del PIB per cápita, para el que los participantes esperaban una correlación fuerte. Estos dos resultados se explican por el fenómeno de la correlación ilusoria (Chapman, 1967), es decir, algunos futuros profesores basaron su estimación de la correlación atendiendo, de algún modo, a expectativas o esquemas propios referidos a las variables que estudiaban (se confirmó este hecho con la puesta en común de las respuestas al finalizar la tarea). Por otra parte, en caso de las correlaciones negativas (tasa de fertilidad en adolescentes, tasa de mortalidad), encontramos una estimación a la baja, lo que coincide con lo expuesto en investigaciones previas (e.g. Estepa, 2008).

#### 4.2 RECONOCIMIENTO DE LA FUNCIÓN DE AJUSTE

En la tabla III presentamos la función de ajuste sugerida por los futuros profesores para cada variable explicativa trabajada en el proyecto, a la que se añadió en la segunda columna el tipo de función de ajuste más adecuada a los datos. Se identificaron correctamente los tipos de funciones de ajuste convenientes, sobre todo para las variables índice de desarrollo humano, índice de educación, PIB per cápita y tasa de fertilidad en adolescentes, todas (salvo el PIB), con función de ajuste lineal.

En general, los participantes reconocieron bien la forma de la función lineal y logarítmica, y más de 50% la exponencial, con lo que mostraron su alta preparación matemática; incluso en muchos casos se indicó el signo de los parámetros, su ecuación y el tipo de crecimiento de una variable respecto a la otra. En la variable población urbana, los participantes advirtieron la alta dispersión de los datos y lo indicaron en sus respuestas, y además sugirieron que no sería muy preciso el resultado de la predicción si utilizaban una función de ajuste a los datos. El alto índice de omisión de respuesta para esa variable se debió a que algunos participantes no asignaron una función porque no consideraban que hubiera una adecuada (así lo manifestaron en la puesta en común de respuestas a la tarea). Este efecto también se presentó con la variable población total (Tabla II), para la que un alto porcentaje de participantes identificó correctamente la independencia.

**Tabla III**  
**Porcentaje de participantes, según la función de ajuste considerada en cada variable**

| Variable explicativa                        | Tipo de función prevista | Correcta | Lineal | Otra | No responde |
|---|--------------------------|----------|--------|------|-------------|
| Índice de Desarrollo Humano                 | Lineal                   | 96,9     |        | 1,5  | 1,5         |
| PIB per cápita                              | Logarítmica              | 86,2     |        | 10,8 | 3,1         |
| Tasa de fertilidad en adolescentes          | Lineal                   | 83,1     |        |      | 16,9        |
| Tasa de mortalidad en menores de cinco años | Exponencial              | 52,3     | 21,5   | 24,6 | 1,5         |
| Gasto en salud pública                      | Polinómica               | 44,6     | 6,2    | 9,2  | 40          |
| Índice de educación                         | Lineal                   | 96,9     |        |      | 3,1         |
| Población total                             | Independencia            | 24,6     | 1,5    |      | 73,8        |
| Población urbana                            | Lineal                   | 53,8     |        | 9,2  | 36,9        |

Como resumen de este apartado, se observó un muy buen “conocimiento común” sobre correlación y regresión en esta muestra de futuros profesores que resolvieron con éxito las tareas.

## 5. CONOCIMIENTO AVANZADO DEL CONTENIDO

### 5.1 ORDEN DE FIABILIDAD EN LA PREDICCIÓN

Para analizar los resultados de la tarea 2, se consideraron parcialmente correctas las respuestas que asignaron un orden de predicción que difirió en una unidad del orden correcto. En la Tabla IV se presentan

la media y la desviación típica del orden asignado por los futuros profesores y el porcentaje de participantes, según su grado de corrección en la tarea. Se añade una columna con los valores correctos del orden de las variables, según su poder de predicción, para facilitar la lectura de los resultados.

Los futuros profesores, en general, ordenaron de modo correcto o parcialmente correcto las variables según su poder para predecir la esperanza de vida. Las respuestas se distribuyeron en torno al valor correcto, como se deduce de la proximidad de la media del orden asignado y del orden de predicción en los datos, así como la baja desviación típica. Las tres variables que mejor predicen la esperanza de vida, tasa de mortalidad, índice de desarrollo humano y PIB per cápita, en ese orden, se suelen identificar correctamente o en forma parcialmente correcta, aunque para el PIB per cápita aparezca mayor variabilidad de órdenes asignados. Igualmente encontramos una buena asignación del orden de predicción en la población total y la tasa de fertilidad en adolescentes, para las que se obtuvo la mayor proporción de respuestas correctas.

Observamos mayor dificultad en las asignaciones de orden para correlaciones menos intensas, como el gasto en salud pública, pero en general, son bastante precisas. Además, confirmamos la buena capacidad de los futuros profesores para identificar la independencia, ya que la mayor proporción de respuestas correctas se refirieron a la asignación de la población total al último orden para predecir la esperanza de vida.

**Tabla IV**  
**Media y desviación típica del orden de predicción asignado y porcentaje de participantes según corrección de la respuesta**

| Variable                                    | Orden de predicción según datos | Orden asignado |           | Grado de corrección (porcentaje de participantes) |                        |             |             |
|---|---------------------------------|----------------|-----------|---|------------------------|-------------|-------------|
|   |                                 | Media          | D. Típica | Correctas   | Parcialmente correctas | Incorrectas | No responde |
| Índice de Desarrollo Humano                 | 2                               | 1,6            | 0,71      | 30,8  | 67,7                   |             | 1,5         |
| PIB per cápita                              | 3                               | 3,3            | 1,44      | 23,1  | 46,1                   | 29,2        | 1,5         |
| Tasa de fertilidad en adolescentes          | 5                               | 4,8            | 1,09      | 50,8  | 29,2                   | 18,5        | 1,5         |
| Tasa de mortalidad en menores de cinco años | 1                               | 1,9            | 1,07      | 41,5  | 36,9                   | 20          | 1,5         |
| Gasto en salud pública                      | 7                               | 5,4            | 1,50      | 26,2  | 32,3                   | 38,5        | 3,1         |
| Índice de educación                         | 4                               | 4              | 1,21      | 38,5  | 35,4                   | 24,6        | 1,5         |
| Población total                             | 8                               | 7,5            | 0,95      | 70,8  | 16,9                   | 9,2         | 3,1         |
| Población urbana                            | 6                               | 6,5            | 1,04      | 24,6  | 55,4                   | 18,5        | 1,5         |

## 5.2 ORDEN DE FIABILIDAD EN LA PREDICCIÓN

Previo al desarrollo de la tarea 3, el formador indicó a los futuros profesores que su finalidad era ayudarles a diferenciar correlación y causalidad, lo que es importante en situaciones tales como de diagnóstico médico o de evaluación. Se pidieron ejemplos de situaciones causales y no causales en las que aparece correlación; entre otras, algún participante aportó: “la talla o peso de un padre, sabiendo la talla o peso del hijo adulto”; variables que están claramente correlacionadas, pero la posible relación causal iría de padre a hijo, y no al contrario. Se explicó a los participantes los tipos de explicaciones de la correlación sugeridos por Barbancho (1992) con ejemplos que clarificasen las diferentes categorías.

En nuestro análisis previo (Tabla I), la relación sería de tipo causal unilateral sólo para las variables tasa de mortalidad y gasto en salud pública, siendo la correspondiente al PIB per cápita y tasa de fertilidad en adolescentes de dependencia indirecta, la del índice de desarrollo humano de interdependencia, el índice de educación y población urbana dependencia indirecta y la población total independencia.

Los resultados de la tarea se presentan en la Tabla V, en donde se observa mayor porcentaje de reconocimiento de la causalidad en la variable tasa de mortalidad que en gasto en salud pública. También se asignó con gran frecuencia causalidad al índice de desarrollo humano, aunque en este caso la relación es de interdependencia, pues la esperanza de vida se incluye en el cálculo de este índice, como así se hizo saber en la presentación de las variables a los participantes. Hay una proporción variable de estudiantes que asignaron incorrectamente una relación de causalidad en las otras variables, como ocurrió con el PIB per cápita, que además, en este caso, coincidió con sus teorías previas, con lo que pudo presentarse en algunos la concepción causal de la asociación (Estepa y Batanero, 1995).

Los argumentos utilizados por los futuros profesores para justificar la causalidad se han clasificado según las categorías que se resumen a continuación, con ejemplos tomados de las respuestas de los participantes, cuyas iniciales indicamos entre paréntesis. Entre ellos hemos encontrado los siguientes argumentos correctos:

- *Influencia en la variable respuesta.* Cuando se justifica la causalidad por la influencia directa de la variable explicativa sobre la esperanza de vida, por ejemplo: “Si gastas más en salud e investigación se encuentran curas a nuevas enfermedades y por tanto la esperanza de vida aumenta” (GM).

- *Modo en que se calcula la esperanza de vida.* En algunos casos se justifica la causalidad por el método de cálculo de la esperanza de vida; generalmente se aplica correctamente para la variable de mortalidad, como se muestra en el siguiente ejemplo: “Es causal, pues a menor mortalidad infantil, mayor será el índice de esperanza de vida puesto que crecerá la media” (CL).

También se han identificado algunos argumentos incorrectos para justificar una relación causal entre las variables:

- *Crecimiento/decrecimiento.* Se identifica la relación de tipo causal con el crecimiento (decrecimiento) conjunto de la variable explicativa y la esperanza de vida. Aunque en algunas gráficas se presentan estas dos propiedades, el argumento no es correcto, en general, pues una causa puede actuar sobre un efecto en forma no lineal, como se indica en el siguiente ejemplo: “Cuanto más avances tiene una población o mayor desarrollo humano posee, mayor será la esperanza de vida” (PP).

También se aplica este argumento erróneamente, al intercambiar la variable explicativa con la variable explicada, por ejemplo: “Nos indica que cuanto menor sea la esperanza de vida mayor será la mortalidad” (ChC).

- *Existencia de terceras variables.* Cuando se justifica la causalidad por el efecto de otras variables que influyen sobre las dos que se analizan, sin diferenciar causalidad de relación indirecta, por ejemplo: “Mientras más alto sea el PIB per cápita por persona, más bienestar se tendrá y, por tanto, la esperanza de vida mejorará; es una relación causal” (LT).

En algunos casos se utiliza erróneamente este tipo de argumentación, al justificar una esperanza de vida baja por una superpoblación, como indica CM: “A mayor fecundidad de la población, más población habrá. Menor será la esperanza de vida por superpoblación”, o intercambiar la variable explicativa con la variable explicada, por ejemplo, PG: “Va directamente ligado con el índice de desarrollo humano, pues cuanto más desarrollo tiene un país, más escolaridad, más educación”.

- *Correlación.* Cuando se identifica explícitamente correlación y causalidad, lo que implicaría una concepción causal de la correlación (Batanero et al., 1998), como se muestra en el siguiente ejemplo: “El índice de correlación está cercano a 1, así que las variables son causa y efecto” (MIH).

En la Tabla V se presenta el porcentaje de futuros profesores que sugirieron que la relación de cada una de las variables explicativas con la esperanza de vida era causal (segunda columna). Observamos que las variables que se citan con mayor frecuencia fueron el índice de desarrollo humano y la tasa de mortalidad, seguida del PIB per cápita. Por tanto, se identificó correctamente la tasa de mortalidad como relación causal con la esperanza de vida, pero no el gasto en salud pública y se supusieron causales relaciones que no lo eran. Por otra parte, y como ocurrió con la estimación y asignación del coeficiente de correlación, la población total se consideró por todos los participantes como independiente de la esperanza de vida.

| Variable                                    | Asumen relación causal | Justificación de la relación causal |         |             |                    | Correlación |
|---|------------------------|-------------------------------------|---------|-------------|--------------------|-------------|
|   |                        | Correcta                            |         | Incorrecta  |                    |             |
|   |                        | Influencia                          | Cálculo | Crecimiento | Terceras variables |             |
| Índice de Desarrollo Humano                 | 89,2                   | 7,7                                 |         | 27,7        | 46,2               | 7,7         |
| PIB per cápita                              | 56,9                   | 4,6                                 |         | 7,7         | 40                 | 4,6         |
| Tasa de fertilidad en adolescentes          | 13,8                   | 1,5                                 | 1,5     | 3,1         | 7,7                |             |
| Tasa de mortalidad en menores de cinco años | 80                     | 10,8                                | 26,2    | 35,4        | 4,6                | 3,1         |
| Gasto en salud pública                      | 32,3                   | 12,3                                |         | 3,1         | 15,4               | 1,5         |
| Índice de educación                         | 33,8                   | 15,4                                |         | 6,2         | 10,8               | 1,5         |
| Población total                             |                        |                                     |         |             |                    |             |
| Población urbana                            | 3,1                    | 1,5                                 |         | 1,5         |                    |             |

En resumen, el conocimiento avanzado del contenido correlación y regresión fue bueno en la tarea relacionada con ordenar las variables explicativas en función de su poder de predicción sobre la esperanza de vida. Respecto a la diferenciación entre correlación y causalidad, el conocimiento mostrado fue menor, pues no se identificaron algunas relaciones causales, mientras se supusieron otras que no eran tales. Además, se observaron participantes que identificaron la causalidad con el crecimiento/decrecimiento conjunto de las variables, la confundieron con la dependencia indirecta o bien la identificaron con la correlación, aunque en pocos casos.

## 6. CONOCIMIENTO ESPECIALIZADO DEL CONTENIDO

Para valorar este componente del conocimiento matemático de la enseñanza se proporcionó a los profesores la pauta de análisis de la idoneidad didáctica, adaptada de Godino (2013), y utilizada por Arteaga et al. (2012), y se les pidió que la aplicasen para valorar el proyecto estadístico en el que habían trabajado las sesiones anteriores. Su aplicación requería que el futuro profesor reconociera la presencia de los distintos tipos de objetos matemáticos en ese proyecto y valorara la pertinencia de cada uno. Respecto al componente de la idoneidad epistémica, los futuros profesores respondieron por escrito a cada una de las nueve cuestiones planteadas (Tabla VI).

**Tabla VI**  
**Pauta de análisis de la idoneidad epistémica proporcionada a los participantes**

| <i>Descriptores de la idoneidad epistémica</i> |   |
|--|---|
| <i>Situaciones-problemas</i>                   | D1. Te parece que los problemas que se presentan en el proyecto son útiles para contextualizar, aplicar y ejercitar los contenidos de correlación y regresión? ¿Por qué? D2. ¿Se proponen situaciones para que el alumno de Bachillerato invente nuevos problemas?  |
| <i>Lenguaje matemático</i>                     | D3. ¿Se usan diferentes modos de expresión matemática (verbal, gráfica, simbólica...), traducciones y conversiones entre dichas expresiones? D4. ¿Hay actividades de representación e interpretación del lenguaje matemático?                                       |
| <i>Conceptos, propiedades, procedimientos</i>  | D5. ¿Qué conceptos, propiedades y procedimientos sobre la correlación y la regresión habría que explicar previamente para desarrollar este proyecto? D6. ¿Se proponen tareas en las que los alumnos tengan que reconocer definiciones propiedades o procedimientos? |
| <i>Argumentos</i>                              | D7. ¿Son las explicaciones, comprobaciones y demostraciones apropiadas para Bachillerato? D8. ¿Se incluyen situaciones en las que el alumno tenga que argumentar?   |
| <i>Relaciones</i>                              | D9. ¿Se relacionan y conectan entre sí los contenidos matemáticos (problemas, definiciones, propiedades, etcétera)?   |

Para realizar esta tarea, los participantes debían ser capaces de reconocer los conceptos, propiedades y procedimientos, lenguajes usados en el proyecto, los tipos de justificaciones, relaciones, así como las diferentes situaciones-problemas implicados. Es decir, debieron identificar las configuraciones epistémicas de objetos y procesos matemáticos requeridas en el proceso de estudio llevado a cabo, y compararlas con las pretendidas en la enseñanza y dictaminadas por las orientaciones curriculares (Godino, Bencomo, Font y Wilhelmi, 2006).

Para evaluar la respuesta a cada pregunta se asignó a cada participante un nivel y se atendió al conocimiento mostrado en la aplicación del descriptor. Estos niveles desarrollan los propuestos por Arteaga et al. (2012), quien utilizó únicamente tres niveles para evaluar el conocimiento especializado sobre gráficos estadísticos de futuros profesores de educación primaria; los dos primeros coinciden con los nuestros, el tercero lo hemos desglosado en tres niveles en nuestro trabajo. Añadimos un primer nivel “0” cuando no se alcanza el primero. A continuación se describen los niveles de análisis que son comunes a todos los descriptores y se incluyen algunos ejemplos tomados de las respuestas de los participantes para clarificarlos.

- *Nivel 0:* El futuro profesor deja en blanco la respuesta a un descriptor, no habiendo comprendido la pregunta, o no siendo capaz de aplicarla en el contexto del proyecto.
- *Nivel 1.* Responde a la pregunta, aunque el participante se limita a copiar casi literalmente el descriptor, sin aplicarlo al proyecto analizado.
- *Nivel 2.* Responde a la pregunta, aplicando el descriptor, pero no se centra específicamente en el proyecto desarrollado, sino en aspectos anecdóticos o no estrictamente matemáticos. Por ejemplo, en lugar de valorar los contenidos matemáticos se fija en el contexto utilizado (esperanza de vida). También consideramos en este nivel a los futuros profesores que aplican una parte del descriptor correctamente y otra incorrectamente, por no haberlo interpretado adecuadamente.

Creo que sí, porque al tomar los datos de la página *web* de la ONU, trabajamos en el contexto público que los alumnos están acostumbrados a ver en los medios de comunicación (ME, descriptor D1).

Sí, ya que se pueden hacer más deducciones y los conceptos tienen más sentido y se entienden mejor (EGO; descriptor D7).

- *Nivel 3.* El futuro profesor aplica el descriptor a contenidos matemáticos del proyecto y a la forma en que se trabajó con el proyecto en el aula, pero la aplicación es incompleta, porque se centra en un aspecto no recogido por el descriptor. Por ejemplo, se le pide valorar las situaciones-problema propuestas en el proyecto, y en su lugar valora una propiedad.

Sí, porque la nube de puntos representa el grado de correlación entre dos variables (ChC, descriptor D1).

Sí, son conceptos básicos que deben saber de años anteriores, no sólo en Bachillerato (media, varianza y desviación típica, por ejemplo) (CL; descriptor D7).

- *Nivel 4.* Aplica correctamente el descriptor al utilizar contenidos matemáticos del trabajo en el proyecto en forma consistente con la pregunta, pero razona mediante un único ejemplo.

Sí, por ejemplo, el cuadro de datos dado en la clase anterior de la media, moda, mediana... Para saber las relaciones entre ellas deben saber su significado (CM; descriptor D6).

Además, cuando nos dan la nube de puntos tenemos que interpretar si hay correlación o no (JFM; descriptor D4).

- *Nivel 5.* El futuro profesor hace una aplicación correcta y consistente del descriptor y utiliza contenidos matemáticos en forma consistente con la pregunta y razona a partir de dos o más ejemplos. Sería idéntico, pero más completo que el caso anterior.

Sí. Por ejemplo, en la actividad en la que se habla de qué variables sirven para predecir mejor la esperanza de vida. O en las actividades de ampliación cuando se pide estudiar “otras variables que te interesen e influyan en el desarrollo humano” (ME, descriptor D2).

Sí, al tener que asignar una relación positiva-negativa, fuerte-débil e incluso el coeficiente de correlación, se le está pidiendo que reconozca los conceptos relacionados con esto. También lo es la tarea de ajustar nubes de puntos con un tipo de función (lineal, exponencial, logarítmica). (ATL; descriptor D6).

Para resumir los resultados de esta tarea, se han calculado las puntuaciones medias obtenidas por el total de participantes en cada pregunta planteada (Tabla VII). Los resultados son buenos, ya que la media global (3.6) supera en un punto el nivel medio teórico (2.5), y es mucho mejor que el mostrado por los futuros profesores de educación primaria en el trabajo de Arteaga et al. (2012) (puntuación media 1.11). Por lo tanto, los participantes muestran su capacidad para identificar los objetos matemáticos implícitos, o que se han hecho explícitos en una cierta situación de enseñanza, lo que sería parte de su conocimiento especializado del contenido, según Hill, Ball y Chilling (2008). Estos resultados son coherentes con los anteriores debido a la relación entre el conocimiento especializado del contenido y el conocimiento común, y al hecho de que el conocimiento avanzado sobre el tema también fue razonable.

**Tabla VII**  
**Medias y desviaciones típicas en los descriptores de la idoneidad epistémica**

| <i>Descriptor</i> | <i>Contenido</i>                    | <i>Media</i> | <i>D, Típica</i> |
|-------------------|-------------------------------------|--------------|------------------|
| <i>D1</i>         | Problemas contextualizados          | 3,2          | 1,3              |
| <i>D2</i>         | Invención de nuevos problemas       | 2,7          | 1,5              |
| <i>D3</i>         | Variedad de lenguaje                | 4,3*         | 1,3              |
| <i>D4</i>         | Interpretación /traducción          | 3,3          | 1,2              |
| <i>D5</i>         | Conocimientos previos requeridos    | 2,8          | 1,2              |
| <i>D6</i>         | Tareas de reconocimiento de objetos | 4,2*         | 1,1              |
| <i>D7</i>         | Nivel adecuado argumentación        | 3,5          | 1,3              |
| <i>D8</i>         | Argumentación por el alumno         | 4,2*         | 1,1              |
| <i>D9</i>         | Relaciones entre objetos            | 4,4*         | 1,1              |
| <i>Total</i>      |                                     | 3,6          | 1,4              |

\* Media (en la aplicación del descriptor) superior a 4.

**Tabla VIII**  
**Ejemplos de conocimientos especializados sobre la correlación y regresión en los futuros profesores**

| <i>Objetos matemáticos</i>                    | <i>Ejemplos de conocimientos especializados</i>  |
|---|--|
| <i>Situaciones-problema</i>                   | 1.1. Identifica ejemplos de problemas en los que se contextualiza la correlación y la regresión en el proyecto, como la búsqueda de la variable que mejor explique la variabilidad de la esperanza de vida. 1.2. Identifica situaciones en que los estudiantes tengan que proponer o modificar problemas relacionados con la correlación y la regresión; en particular, el análisis de otras variables dependientes en el fichero de datos.  |
| <i>Lenguaje matemático</i>                    | 1.3. Reconoce que los diagramas de dispersión, las ecuaciones de las funciones de ajuste y los términos en el proyecto son distintos modos de expresión matemática ligados a la correlación y la regresión. 1.4. Reconoce las representaciones gráficas que permiten visualizar mejor la correlación y la regresión; por ejemplo, ve la necesidad de un cambio de escala en los datos. 1.5. Identifica situaciones de interpretación y traducción de representaciones de la correlación, como la búsqueda de un coeficiente de correlación o la estimación de éste, a partir del diagrama de dispersión.                 |
| <i>Conceptos, propiedades, procedimientos</i> | 1.6. Reconoce los conceptos, propiedades y procedimientos asociados a la correlación y a la regresión que el alumno ha de aplicar al resolver cada problema, relacionando, entre otros, la dependencia estadística y funcional, la intensidad y el sentido, la correlación, el ajuste y el modelo. 1.7. Reconoce cuáles de ellos han de ser explicados previamente y cuáles pueden ser construidos por el alumno con la ayuda del profesor. 1.8. Identifica y propone nuevas tareas en las que el alumno ha de reconocer o recordar conceptos, propiedades y procedimientos asociados a la correlación y a la regresión. |
| <i>Argumentos</i>                             | 1.9. Identifica argumentaciones adecuadas e inadecuadas para el nivel del estudiante para justificar una solución o una propiedad. 1.10. Identifica situaciones que requieran una argumentación por parte del estudiante, como por ejemplo las referidas a la justificación de situaciones causales y no causales en el proyecto. 1.11. Identifica relaciones entre objetos matemáticos en el contexto del proyecto, como la relación entre la dirección de la dependencia y el signo del coeficiente de   |

Otra aportación del análisis es la identificación, a partir de las respuestas de los estudiantes, de ejemplos de conocimientos especializados requeridos por parte del profesor sobre los distintos objetos matemáticos presentes en el estudio de la correlación y la regresión que mostramos en la Tabla VIII. Esta tabla se ha construido con ejemplos específicos de conocimientos manifestados por los participantes en las respuestas del análisis de la idoneidad epistémica (niveles 4 y 5), algunas de las cuales hemos presentado a lo largo de la sección.

## 7. DISCUSIÓN E IMPLICACIONES PARA LA FORMACIÓN DE PROFESORES

La principal contribución de la investigación reseñada es la identificación de conocimientos matemáticos (común, avanzado y especializado) de la correlación y la regresión en una muestra de futuros profesores de Educación Secundaria y Bachillerato que complementa los escasos antecedentes sobre el tema.

Los resultados muestran buenos conocimientos de los participantes sobre el significado y los tipos de correlación, así como buena estimación del valor absoluto y el signo de la correlación a partir de diagramas de dispersión e identificación de la tendencia de los datos en el caso de ajuste lineal, así como de tendencia logarítmica, exponencial o polinómica por un porcentaje alto de futuros profesores. Además, los participantes en este estudio mostraron mejor identificación de la independencia estadística en diagramas de dispersión y estimación de su correlación que los participantes en estudios previos. Todo ello indica un buen conocimiento común del contenido.

Los futuros profesores, en general, ordenan de modo correcto o parcialmente correcto las variables según su poder para predecir la esperanza de vida. Fue más difícil aceptar explicaciones de la correlación diferentes de las causales, pero, en general, los futuros profesores mostraron un conocimiento avanzado razonable del tema. Igualmente fue bueno su análisis de la idoneidad epistémica del proyecto, pues el nivel medio alcanzado (en una escala de 5 puntos) se sitúa en todos los descriptores de la pauta de análisis sobre el valor medio teórico (2.5). Otras aportaciones de la investigación son la propuesta de niveles para valorar la competencia de análisis de la idoneidad didáctica por parte de los futuros profesores, que amplía la de Arteaga et al. (2012) y la identificación de ejemplos de conocimiento especializado sobre la correlación y la regresión a partir de las respuestas de los participantes.

Por la limitación de espacio, en este trabajo sólo se describen resultados de las tareas que se muestran en la figura 1, cuyas soluciones se discutieron una vez realizadas con los futuros profesores y posteriormente se ampliaron con otras, con objeto de reforzar su conocimiento sobre el tema. Para completar las nuevas tareas se proporcionaron a los participantes los datos empleados en la elaboración de los diagramas de dispersión en una hoja Excel (193 países con datos recogidos de nueve variables). Usando estos datos, se trabajó colectivamente con la hoja Excel para identificar la ecuación de la función de regresión de la esperanza de vida respecto a cada una de las variables explicativas propuestas. Se realizaron actividades de interpretación de los parámetros de las funciones de regresión y se analizaron los valores del coeficiente de determinación como indicador de la bondad de ajuste de cada modelo. Todos estos resultados se compararon con las respuestas de los participantes en las tareas 1 a 4.

Las tareas planteadas resultaron interesantes 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 correlación, principalmente la causalidad y la dependencia indirecta (*conocimiento avanzado*). Las variables seleccionadas para trabajar en el proyecto estadístico les permitieron observar diversos aspectos que influyen en la comprensión del tema (*conocimiento especializado*); la determinación del signo e intensidad de la correlación entre las variables, e identificación de funciones más allá de la dependencia lineal, permitió trabajar el *conocimiento común*. Al trabajar con fuentes internacionales aumentó el interés de los participantes y su reconocimiento de la utilidad de la estadística, lo que mejoró sus actitudes. La actividad de valoración de la idoneidad epistémica también mejoró su *conocimiento especializado*.

Además, se les mostró un ejemplo del trabajo con proyectos que pueden utilizar con sus futuros alumnos para recorrer un estudio estadístico completo: problema, datos, análisis y conclusiones (Batanero y Borovcnick, 2016). Proporciona igualmente una actividad de modelización en la que el interés se centra en analizar la dependencia y, en caso de ser alta, encontrar el modelo de ajuste más adecuado (lineal, logarítmico, etc.). Precisamente Puig y Monzó (2013) incluyen la regresión y la correlación como una de las lecciones en su propuesta de enseñanza de la modelización. Con esta actividad el alumno puede comprender que, como explica Henry (1997), “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” (p. 78). Las actividades permiten también analizar la diferencia entre el modelo (ecuación de regresión) y la realidad reflejada en los datos bivariantes. Asimismo, traducir los resultados del trabajo matemático realizado con el modelo a la realidad modelizada.

Estos últimos puntos forman parte del conocimiento didáctico del tema, que era escaso en los futuros profesores de la muestra, pues casi ninguno había trabajado previamente con datos reales con la regresión utilizando Excel, ni el método de proyectos, cuya importancia en la modelización resaltan Puig y Monzó (2013). Desconocían asimismo las investigaciones didácticas sobre el tema, que fueron discutidas con ellos a la vez que se debatieron las soluciones. Nuestra reflexión final es que la mejora de la enseñanza de la correlación y la regresión depende de la formación de los profesores, que puede mejorarse con actividades como las descritas en este trabajo.

### Agradecimientos

Trabajo realizado en el marco de los proyectos EDU2013-41141-P (MINECO), EDU2016-74848-P (AEI, FEDER) y FCT-16-10974 (FECYT) y grupo de investigación FQM126 (Junta de Andalucía).

### Referencias

- Arteaga, P., Batanero, C., Cañadas, G. R. y Gea, M. (2012). Evaluación del conocimiento especializado de la estadística en futuros profesores mediante el análisis de un proyecto estadístico. *Educação Matemática Pesquisa*, 14(2) 279-297. Disponible en: <https://revistas.pucsp.br/index.php/emp/article/view/9317>
- Ball, D. L., Lubienski, S. T., & Mewborn, D. S. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. En V. Richardson (Ed.), *Handbook of research on teaching* (pp. 433-456). Washington, DC.: American Educational Research Association.
- Barbancho, A. G. (1992). *Estadística elemental moderna*. Barcelona: Ariel.

- Batanero, C. & Borovcnik, M. (2016). *Statistics and probability in high school*. Londres: Sense Publishers. <https://doi.org/10.1007/978-94-6300-624-8>.
- Batanero, C., Godino, J. D., y Estepa, A. (1998). Building the meaning of statistical association through data analysis activities. En A. Olivier & K. Newstead, (Eds.), *Proceedings of the 22nd Conference of the International Group for the Psychology of Mathematics Education* (pp. 221-236). Stellenbosch, South Africa: Universidad de Stellenbosch.
- Casey, S. A. (2010). Subject matter knowledge for teaching statistical association. *Statistics Education Research Journal*, 9(2), 50-68. Disponible en [https://iase-web.org/documents/SERJ/SERJ9\(2\)\\_Casey.pdf](https://iase-web.org/documents/SERJ/SERJ9(2)_Casey.pdf)
- Casey, S. A. y Wasserman, N. H. (2015). Teachers' knowledge about informal line of best fit. *Statistics Education Research Journal*, 14(1), 8-35. Disponible en [https://iase-web.org/documents/SERJ/SERJ14\(1\)\\_Casey.pdf](https://iase-web.org/documents/SERJ/SERJ14(1)_Casey.pdf)
- Castro-Sotos, A. E., Vanhoof, S., Van Den Noortgate, W. & Onghena, P. (2009). The transitivity misconception of Pearson's correlation coefficient. *Statistics Education Research Journal*, 8(2), 33-55. Disponible en [https://iase-web.org/documents/SERJ/SERJ8\(2\)\\_Sotos.pdf](https://iase-web.org/documents/SERJ/SERJ8(2)_Sotos.pdf)
- Chapman, L. J. (1967). Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior*, 6(1), 151-155.
- Engel, J. & Sedlmeier, P. (2011). Correlation and regression in the training of teachers. En C. Batanero, G. Burrill & C. Reading (Eds.), *Teaching statistics in school mathematics-challenges for teaching and teacher education: A Joint ICMI/IASE study* (pp. 247-258). New York: Springer. [https://doi.org/10.1007/978-94-007-1131-0\\_25](https://doi.org/10.1007/978-94-007-1131-0_25).
- Estepa, A. (2007). Caracterización del significado de la correlación y regresión de estudiantes de Educación Secundaria. *Zetetiké*, 15(28), 119-151.
- Estepa, A. (2008). Interpretación de los diagramas de dispersión por estudiantes de Bachillerato. *Enseñanza de las Ciencias*, 26(2), 257-270. Disponible en <https://ddd.uab.cat/record/39831?ln=ca>
- Estepa, A. y Batanero, C. (1995). Concepciones iniciales sobre la asociación estadística. *Enseñanza de las Ciencias*, 13(2), 155-170. Disponible en <https://www.raco.cat/index.php/ensenanza/article/viewFile/21404/93364>
- Estepa, A. y Sánchez Cobo, F. T. (2001). Empirical research on the understanding of association and implication for the training of researchers. En C. Batanero (Ed.) *Training researchers in the use of statistics* (pp. 37-51). Granada, España: International Association for Statistical Education and International Statistical Institute.
- Estepa, A. y Sánchez-Cobo, F. T. (2003). Evaluación de la comprensión de la correlación y regresión a partir de la resolución de problemas. *Statistics Education Research Journal*, 2(1), 54-68. Disponible en [https://www.stat.auckland.ac.nz/~iase/serj/SERJ2\(1\).pdf#page=56](https://www.stat.auckland.ac.nz/~iase/serj/SERJ2(1).pdf#page=56)
- Gea, M. 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. Universidad de Granada. Disponible en <http://www.ugr.es/~batanero/pages/librotesis.html>
- Godino, J. D. (2013). Indicadores de la idoneidad didáctica de procesos de enseñanza y aprendizaje de las matemáticas. *Cuadernos de Investigación y Formación en Educación Matemática*, 11, 111-132. Disponible en <https://revistas.ucr.ac.cr/index.php/cifem/article/view/14720>
- Godino, J. D., Bencomo, D., Font, V. y Wilhelmi, M. R. (2006). Análisis y valoración de la idoneidad didáctica de procesos de estudio de las matemáticas. *Paradigma*, 27(2), 221-252.
- Godino, J. D., Contreras, A. y Font, V. (2006). Análisis de procesos de instrucción basado en el enfoque ontológico-semiótico de la cognición matemática. *Recherches en Didactiques des Mathématiques*, 26(1), 39-88.
- Henry, M. (1997). Notion de modèle et modélisation dans l'enseignement. En M. Henry (Ed.), *Enseigner les probabilités au lycée* (pp. 77-84). Reims: Commission Inter-IREM.
- Hill, H. C., Ball, D. L. & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education*, 39(4), 372-400.
- MECD. (2015). *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: Ministerio de Educación, Cultura y Deporte.
- Moritz, J. (2004). Reasoning about covariation. En D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 221-255). Dordrecht: Kluwer.
- Puig, L. y Monzó, O. (2013). Fenómenos y ajustes. Un modelo de enseñanza del proceso de modelización y los conceptos de parámetro y familia de funciones. En T. Rojano (Ed.). *Las tecnologías digitales en la enseñanza de las matemáticas* (pp. 9-35) México: Trillas.

Quintas, S., Ferreira, R. y Oliveira, H. (2015). O conhecimento didático de estatística de duas professoras de matemática sobre dados bivariados. *Bolema*, 29(51), 284-306. <https://doi.org/10.1590/1980-4415v29n51a15>

Sánchez Cobo, F. T., Estepa, A. y Batanero, C. (2000). Un estudio experimental de la estimación de la correlación a partir de diferentes representaciones. *Enseñanza de las Ciencias*, 18(2), 297-310. Disponible en <https://www.raco.cat/index.php/Ensenanza/article/viewFile/21674/21508>

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4-14. Disponible en <https://doi.org/10.1590/1980-4415v29n51a15>

Verdú, C., Callejo, M. L. y Márquez, M. (2014). Conocimiento de los estudiantes para maestro cuando interpretan respuestas de estudiantes de primaria a problemas de división-medida. *Enseñanza de las Ciencias*, 32(3), 407-424. <https://doi.org/10.5565/rev/ensciencias.1235>