

EXPLICACIÓN DEL RENDIMIENTO A PARTIR DEL CONTEXTO. ALGUNAS PROPUESTAS DE ANÁLISIS GRÁFICO Y ESTADÍSTICO¹

Explanation of performance by context. Some proposals from graphic and statistical analysis

EMELINA LÓPEZ-GONZÁLEZ Y JOSÉ GONZÁLEZ-SUCH
Universidad de Valencia

LUIS LIZASOAIN
Universidad del País Vasco

La evaluación de sistemas educativos adolece de algunas debilidades, entre las que se encuentra el limitado tratamiento estadístico de las informaciones que se obtienen. Una mejora consiste en prestar mayor atención al análisis explicativo del producto a partir del contexto. El presente trabajo tiene como objeto ejemplificar de manera sencilla posibles modelos estadísticos que pongan en relación variables antecedentes (variables de contexto) con resultados de aprendizaje (rendimiento) y hacerlo empleando algunas de las herramientas gráficas que aportan los modelos de dependencia basados en la regresión. De manera específica se construye un modelo lineal clásico y un modelo lineal generalizado de regresión de Poisson. Para ello se revisan las condiciones de aplicación de los modelos que se trabajan empleando fundamentalmente los elementos de diagnóstico que proporciona el estudio de residuos: gráficos básicos de diagnóstico en regresión; datos inusuales —*outliers*, valores *hat*, valores de influencia— y gráficos de residuos en sus distintas versiones. El programa estadístico que facilita la tarea propuesta es R, dado su cómodo *interface* gráfico y la facilidad de las salidas de análisis que aporta. La información que se desprende de las representaciones gráficas de los residuos y de los indicadores de diagnóstico orienta la toma de decisiones de las etapas posteriores de los análisis.

Palabras clave: *Modelo lineal, Modelo lineal generalizado, Diagnóstico en regresión, Regresión de Poisson.*

Introducción

La evaluación de sistemas educativos tiene hoy una importancia reconocida, particularmente por el ambiente de reforma generalizada de los sistemas para adaptarse a los vertiginosos cambios sociales. Sin lugar a dudas sus consecuencias son trascendentales, especialmente por residir su finalidad última en la toma de decisiones asociada a diversas funciones: de diagnóstico de la situación, de retroalimentación de la información, de control, de acreditación y/o certificación y de prospectiva (De la Orden y Jornet, 2012). No cabe duda de que tal proceso encierra una complejidad importante, tanto por la necesaria combinación de elementos por tratar en el marco de un modelo sistémico que los relacione, como por la adaptación de los procedimientos metodológicos y de análisis estadístico que deban emplearse.

El diseño de instrumentos para abordar la medida de los constructos que acontecen en un sistema educativo constituye una importante área de trabajo de la que dan cuenta los artículos de Joaristi, Lizasoain y Gamboa (2012); Jornet, González-Such y Perales (2012) y López-González, Tourón y Tejedor (2012), incluidos también en este monográfico. En líneas generales se estudia construir cuestionarios de contexto a partir de indicadores simples y complejos representados por micro-instrumentos derivados de escalas bien diseñadas. Son propuestas que se orientan en el marco general de actuación que perfilan De la Orden y Jornet (2012) y que toman como referencia: las aportaciones de la investigación educativa, un modelo sistémico que ayude a dar coherencia a los planes de evaluación, las necesidades de información de las audiencias implicadas en la evaluación y un procedimiento más riguroso de diseño de instrumentos. Sin embargo, tal actuación resulta incompleta si no se tiene en cuenta un mayor conocimiento de las posibles estrategias estadísticas que emplear con las informaciones que se obtienen. Algo tan sencillo como saber responder a la pregunta ¿qué hacer con los datos?

De la Orden y Jornet (2012) señalan cinco debilidades que caracterizan a las evaluaciones de sistemas educativos. Dos de ellas hacen referencia al limitado tratamiento estadístico de los datos y ambas redundan en el bajo aprovechamiento de la información recogida. Como indican, a pesar de que la medida de los productos de aprendizaje contemple procedimientos de seguimiento y valoración de la evolución del sistema, no se realizan los análisis longitudinales precisos². Tampoco se presta una adecuada atención al análisis explicativo del producto a partir del contexto. Con todo, no se conectan los procesos con los productos. Se evalúa el rendimiento, pero sin explorar posibles vínculos que lo relacionan con otros elementos antecedentes del sistema. Como resultado, no se rentabiliza suficientemente la inversión en esfuerzos y recursos efectuada en la recogida de la información y, lo que es más notorio, disminuye el valor de la evaluación misma.

Una posible causa del limitado tratamiento estadístico de los datos en las evaluaciones educativas quizá se encuentre en el desconocimiento de algunos modelos explicativos (modelos estadísticos de dependencia) que no siempre resultan comprensibles, lo que se concreta en el extendido hábito de analizar someramente los datos de forma descriptiva, sencillamente con porcentajes, puntuaciones típicas, percentiles o correlaciones. Es probable que se produzca también un cierto temor a *manipular* los datos, sostenido tal vez en la creencia de que el *aparataje* estadístico trastoca la información original. Pero si se desea potenciar la utilidad de las evaluaciones de sistemas educativos, es deseable emplear herramientas de análisis adecuadas que permitan hacer emerger en lo posible la información contenida en los datos. A nuestro entender, las mayores dificultades en una evaluación residen en la definición de un modelo teórico que sostenga el sistema y en el abordaje de los indicadores de las variables de entrada, contexto y producto con instrumentos de calidad. Una vez obtenidos los datos, existe una oferta amplia de recursos estadísticos —algunos sencillos y otros algo más

sofisticados— y una propuesta de *software* estadístico lo suficientemente desarrollada para que tal cometido se lleve a cabo sin grandes dificultades.

El presente trabajo tiene como objeto ejemplificar de manera sencilla dos modelos de dependencia que ponen en relación variables antecedentes (variables de contexto) con resultados de aprendizaje (rendimiento), empleando algunas de las herramientas gráficas que aportan los modelos explicativos basados en la regresión. Particularmente se construye un modelo lineal clásico y un modelo lineal generalizado de regresión de Poisson. Para tal cometido se recurre a la ayuda del programa R, cuya capacidad para el análisis gráfico está demostrada (Ruiz-Soler y López-González, 2009). Los análisis se efectúan con datos recogidos en el proyecto de investigación anteriormente referenciado¹.

La oferta actual de modelos de dependencia es muy amplia. Entre los más complejos, un ligero repaso obliga a señalar las extensiones de los modelos de ecuaciones estructurales (Coenders; Batista y Saris, 2005; Kolenikov, Steinley y Thombs, 2010), las extensiones de los modelos lineales generalizados (Hardin y Hilbe, 2007) y la variedad de modelos jerárquicos, multinivel o anidados, lineales (Bickel, 2007; Bliese, 2009; De la Orden, 1999; Gaviria y Castro, 2005; Gelman y Hill, 2007) y no lineales (Raudenbush y Bryk, 2002; De Leeuw y Meijer, 2008)³. En la idea de animar al sector de evaluadores al uso de modelos explicativos, las estrategias que aquí se utilizan son más sencillas, aunque no por ello menos meritorias. El modelo lineal (ML) —regresión clásica—, por ejemplo, suele emplearse con frecuencia, pero tal vez sin explorar suficientemente las informaciones que se desprenden de su diagnóstico. Una mejor opción la constituyen los modelos lineales generalizados (MLGs), más adaptados que el modelo lineal a las características métricas de las habituales escalas de los cuestionarios de contexto (normalmente escalas Likert —ordinales—, o de recuentos, como el número de

respuestas correctas en pruebas de rendimiento). Otras soluciones estadísticas poco habituales son las transformaciones de variables, muy útiles a la hora de potenciar el cumplimiento de las condiciones del modelo lineal (normalidad, linealidad, independencia de los errores y ausencia de autocorrelación).

Un elemento clave es la decisión acerca de qué modelo explicativo emplear en cada situación, decisión que debe descansar en el conocimiento de sus características matemáticas, lo que afecta a las condiciones de aplicación, los objetivos del modelo, el diagnóstico, los indicadores por interpretar, las transformaciones y alternativas posibles, etc. Para ello es necesario entender algunos conceptos básicos en los que se basa, tanto la elección del modelo que mejor se ajuste a los datos como la interpretación acertada de los indicadores y gráficos que se obtienen. Por esta razón, antes de comentar los análisis exponemos una pequeña síntesis de algunos de estos conceptos y de los elementos que vinculan un modelo y otro. Después empleamos y diagnosticamos ambos modelos. Terminamos discutiendo las ventajas e inconvenientes de uno y otro.

Elementos de diagnóstico en el modelo lineal

La mayoría de las decisiones que se toman en torno al uso de modelos de regresión están basadas en el análisis de residuos. La consideración de estos elementos para el diagnóstico del modelo es trascendental, tal y como se argumenta en el trabajo de López-González (1994). Puede decirse que constituyen una *radiografía* bastante fiel del estado interno del modelo.

En un modelo lineal de regresión, dos son las variables que están sujetas a restricciones (además de la deseable relación lineal entre los predictores y el criterio): los valores pronosticados del criterio o respuesta y los errores del modelo o componente aleatorio. Los residuos son los elementos que relacionan ambas variables. Las

distribuciones de los valores pronosticados (o distribuciones condicionadas) están determinadas por los errores del modelo, y estos últimos son, no obstante, desconocidos. De modo que solo el análisis de los residuos hace posible el estudio de las características de las distribuciones condicionadas de la respuesta, permitiendo con ello valorar qué condiciones matemáticas reúne el modelo en su cometido de representar bien los datos.

Si la respuesta y depende de un conjunto de k predictores x a través de una función lineal $E(y|x) = b_0 + b_1x_1 + \dots + b_kx_k + \varepsilon$, en términos muestrales dicha relación queda establecida como $\hat{y}_i = b_0 + b_1x_{i1} + \dots + b_kx_{ik}$. Los *residuos ordinarios* se representan como las diferencias entre las respuestas observadas y los valores ajustados: $e_i = y_i - \hat{y}_i$, $i = 1, \dots, n$.

Estos residuos deben reunir dos características: a) no correlacionar con los valores pronosticados o, en su lugar, con cualquier combinación de los predictores —condición de *independencia*— y b) constituir variables aleatorias con media cero y varianza constante: $\text{Var}(e_i) = \sigma^2(1 - h_i)$ —condición de *homoscedasticidad*—. Junto con estas dos *hipótesis básicas del modelo lineal*, deben producirse igualmente otras dos: la *normalidad* de las distribuciones de la variable de respuesta (tanto en su versión de valores observados como pronosticados o ajustados) y la *linealidad* entre predictores y criterio antes mencionada. Aunque estas dos últimas condiciones no afectan directamente a los residuos, también ellos dan testimonio de su cumplimiento.

El análisis de residuos se efectúa observando las propiedades de sus distribuciones (medias, varianzas), registrando *anomalías* en sus tamaños y signos —lo que fácilmente da lugar a arbitrariedades—, pero sobre todo observando los patrones gráficos en los que quedan representados. Para tal cometido pueden utilizarse distintas versiones: residuos ordinarios, estandarizados, estudentizados, de Pearson, etc.

Habitualmente, cuando los residuos se representan *enfrentados* a la variable de respuesta o a los predictores del modelo, o transformados en algún valor de influencia o en alguna distancia (distancia de Cook, de Mahalanobis, valores *hat*, etc.), dan fiel cuenta del *estado* del modelo y, por tanto, de si resulta acertado para representar bien los datos. No se obtiene de este modo una medida *exacta* del cumplimiento de las hipótesis básicas ni de la bondad del ajuste, pero sí se posee una información suficiente para orientar la actuación del evaluador a la hora de construir un buen modelo explicativo.

Unos elementos asociados a los residuos y que son también muy útiles en el diagnóstico son los valores *hat*⁴, contenidos en la expresión $\text{Var}(e_i) = \sigma^2(1 - h_i)$. En los modelos lineales, los h_i son valores no aleatorios que oscilan entre 0 y 1. Los valores altos de h_i corresponden a los casos con puntuaciones en los predictores x_i relativamente inusuales, es decir, a las respuestas observadas que más influyen en la *iésima* respuesta predicha: las observaciones influyentes.

Las distintas versiones de los residuos se obtienen a partir del uso de estos valores *hat*. Por ejemplo, si los residuos ordinarios se dividen por un estimador de su varianza, tal que

$$e_{Si} = \frac{e_i}{\hat{\sigma}\sqrt{1 - h_i}},$$

se obtienen los *residuos estandarizados* o escalados. El uso de estos residuos tiene la ventaja de corregir posibles problemas de heteroscedasticidad: al quedar divididos por un estimador de su varianza, esta se estabiliza. Los residuos estandarizados permiten además eliminar las diferencias en la variabilidad de los predictores. Al construirlos se estandarizan también las puntuaciones de los predictores, lo que facilita las comparaciones necesarias para el diagnóstico.

Los *residuos estudentizados* se plantean ajustando el modelo una vez eliminada una de las

observaciones iniciales de los datos y pronosticada su respuesta:

$$e_{Ti} = \frac{e_i}{\hat{\sigma}_{(-i)}\sqrt{1-h_i}} .$$

La ventaja se encuentra en que al no intervenir dicha observación en el cálculo de la varianza de los residuos $\hat{\sigma}_{(-i)}^2$, el residuo puro e_i es independiente de $\hat{\sigma}_{(-i)}^2$ y el numerador y denominador del residuo estudentizado quedan también independientes. Otra ventaja añadida es que se distribuyen como estadísticos t de Student con $n-k-2$ grados de libertad, lo que resulta ser un test muy útil para detectar los valores más alejados (*outliers*).

Por último, los *residuos de Pearson* se establecen realizando el ajuste del modelo con mínimos cuadrados ponderados en lugar de con mínimos cuadrados ordinarios, cuya expresión es:

$$e_{Pi} = \sqrt{w_i}e_i .$$

Cuando se efectúa un análisis de residuos conviene detectar además los valores más alejados, ya que si su presencia es considerable pueden alterar de manera importante la estimación de los verdaderos parámetros del modelo. Lo habitual es localizar los residuos estandarizados que no se encuentren incluidos en el intervalo de ± 3 desviaciones típicas. Una segunda exploración de datos de este tipo debe permitir detectar las observaciones influyentes, es decir, aquellos casos que sin presentar un residuo estandarizado mayor a ± 3 , pueden tener una notable influencia en los resultados del análisis. Para ello se utiliza la distancia de Mahalanobis o la distancia de Cook. Esta segunda analiza los cambios que se producen en todos los residuos si el caso i fuera eliminado. Si la distancia es elevada, cabe pensar que el cambio sería grande y se habría eliminado una medida de notable influencia.

La mayor parte de estos sencillos elementos de diagnóstico son también comunes a los modelos

lineales generalizados (aunque con alguna ligera modificación). Al fin y al cabo el modelo lineal no deja de ser el caso más simple de modelo lineal generalizado, como veremos a continuación.

Del modelo lineal a los modelos lineales generalizados

En un artículo sobre historia de la estadística, Efron (2003) comenta que uno de los principales avances en el uso de modelos estadísticos durante la segunda mitad del siglo xx ha sido el desarrollo de los modelos lineales generalizados. La idea original de estos modelos se formaliza en el trabajo de Nelder y Wedderburn (1972). A juicio de Fox y Weisberg (2011: 229), suponen una extensión del *elegante y bien comprendido* modelo lineal a situaciones en las que la variable de respuesta es categórica o discreta, en lugar de continua. Unas notas sobre su estructura permitirán entender mejor sus potenciales beneficios. Para una explicación más detallada pueden consultarse los textos de Agresti (2002: 116-117); Ato y López-García (1996: 84-87); Ato, Losilla, Navarro, Palmer y Rodrigo (2005: 6-9); Fox (2008: 379-385) y López-González y Ruiz-Soler (2011) —por citar solo algunos—.

La estructura de los modelos lineales generalizados posee tres componentes: el componente aleatorio, el predictor lineal y la función de enlace:

- a) El *componente aleatorio* especifica la distribución condicionada de la variable de respuesta, y, dados los x predictores. Así como en el modelo lineal este componente es siempre del mismo tipo —una distribución gaussiana (normal)—, aquí puede adoptar un conjunto de distintas funciones que Nelder y Wedderburn (1972) denominan familia exponencial. Entre ellas, la distribución binomial y la de Poisson son las más usuales.

- b) Al igual que en los modelos lineales, en un MLG los predictores conforman un conjunto de k variables regresoras, $\mathbf{x} = (x_1, \dots, x_k)$, que pueden ser *dummy*, polinomiales, *splines*⁵, transformaciones, etc. La respuesta depende de los predictores por medio de una función lineal, que en el entorno de los MLGs se llama *predictor lineal*:

$$\eta_i(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k.$$

- c) Como hemos indicado antes, en el modelo lineal la conexión entre la media condicional de la respuesta $E(y|\mathbf{x})$ y el predictor lineal $\eta_i(\mathbf{x})$ es directa, tal que: $E(y|\mathbf{x}) = \eta_i(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$, de manera que la media es igual a una combinación lineal de los regresores. Sin embargo, tal relación directa no es correcta en todos los MLGs porque $\eta(\mathbf{x})$ puede tomar cualquier valor entre $(-\infty, +\infty)$. Por esta razón se lleva a cabo una inversión de la *función de enlace* que *traduce* la escala de la media de la respuesta a la escala del predictor lineal (tabla 1). Si la media condicional de la respuesta es $\mu(\mathbf{x}) = E(y|\mathbf{x})$, tenemos que $g = [\mu(\mathbf{x})] = \eta(\mathbf{x})$. Invertiendo esta relación se produce la función de enlace inversa, $g^{-1} = [\eta(\mathbf{x})] = \mu(\mathbf{x})$, denominada también como *función media* o *función media nuclear* (Fox y Weisberg, 2011: 231)⁶.

Otra característica diferencial de los MLGs respecto al modelo lineal es el *método de estimación* de los coeficientes de la ecuación. A diferencia de la estimación mínimo cuadrática del modelo lineal, los MLGs realizan una estimación máximo verosímil, utilizando de manera iterativa el procedimiento de mínimos cuadrados ponderados en lugar de los mínimos cuadrados ordinarios del modelo lineal (una explicación con más detalle se encuentra en Fox y Weisberg, 2011: 283-284 y López-González, Pelegrina y Ruiz-Soler, 2002). Tomando las estimaciones máximo verosímiles de los parámetros de regresión como b_0, b_1, \dots, b_k , el valor estimado del predictor lineal es $\hat{\eta}_i(\mathbf{x}) = b_0 + b_1 x_1 + \dots + b_k x_k$ y la media estimada de la respuesta $\hat{\mu}(\mathbf{x}) = g^{-1} = [\hat{\eta}(\mathbf{x})]$.

Las *varianzas* de los modelos también son un aspecto que tener en cuenta. La varianza en la familia exponencial es el producto de un parámetro de dispersión positivo — ϕ — y la función de la media, establecida a partir del predictor lineal $Var(y|\mathbf{x}) = \phi V[\mu(\mathbf{x})]$.

Para el modelo lineal (que sigue una distribución de Gauss) la varianza es $V[\mu(\mathbf{x})]=1$, lo que la hace dependiente únicamente de la dispersión del parámetro ϕ . Normalmente se sustituye ϕ por σ^2 , dando un valor constante. Este dato es importante, ya que mientras que la varianza es constante en el modelo lineal, en los MLGs depende de $\mu(\mathbf{x})$.

TABLA 1. Algunas características del modelo lineal (ML) y del caso de modelo lineal generalizado (MLG) que empleamos aquí: μ es el valor esperado de la respuesta; η es el predictor lineal; ϕ es el parámetro de dispersión; $\mu = \mu(\mathbf{x})$ es la media condicional de y dados los valores de los predictores⁷

Modelo	Enlace	$\eta = g(\mu)$	$\mu = g^{-1}(\eta)$	Inversa del enlace
ML	identidad	μ	η	identidad
MLG	log	$\log_e \mu$	e^η	exponencial
	Familia	Enlace canónico	Rango de y	$Var(\mathbf{v} \mathbf{x})$
ML	Gauss	identidad	$(-\infty, +\infty)$	ϕ
MLG	Poisson	log	0, 1, 2, ...	μ

Por último, la *desviación* o discrepancia entre los valores empíricos y los ajustados de un MLG está basada en el valor maximizado de la log-verosimilitud y tiene un papel similar a la suma de cuadrados residual del modelo lineal. Es útil, por tanto, para valorar el ajuste del modelo. Si tomamos $p[y; \mu(x), \phi]$ como la función de densidad-probabilidad de una sencilla respuesta y a partir de los predictores x , el valor log-verosímil de las estimaciones de los coeficientes de regresión para una dispersión fija es: $\log_e L_0 = \sum \log_e p[y; \hat{\mu}(x), \phi]$, aplicando el sumatorio para los n casos independientes.

De manera similar, imaginemos el ajuste de otro modelo, que llamamos *modelo saturado*, con un parámetro para cada uno de los n casos. El valor estimado de la media de respuesta para cada caso es justo su valor observado. Así, la log-verosimilitud de este modelo saturado es $\log_e L_1 = \sum \log_e p[y_i; y_i, \phi]$ y la *desviación residual* será la diferencia entre las log-verosimilitudes de ambos modelos: $D(y, \hat{\mu}) = 2(\log_e L_1 - \log_e L_0)$. Dado que el modelo saturado debe ajustar los datos al menos tan bien como cualquier otro modelo, la desviación nunca es negativa. Cuanto mayor es, menor coincidencia hay entre el modelo y los datos. Es decir, peor ajuste. La discrepancia dividida por los grados de libertad suele usarse para detectar sobre o baja dispersión, lo que también es síntoma de un ajuste incorrecto. En la regresión de Poisson la media y la varianza son iguales, lo que implica que la discrepancia dividida por los grados de libertad debe aproximarse a uno.

Método

Participantes

Se trabajó a partir de una investigación realizada con 528 alumnos de 18 centros de enseñanza de Valencia capital, de nivel de sexto de primaria. Las chicas representaron un 46% (244 casos) y los chicos un 54% (284). El muestreo fue

accidental: las pruebas se pasaron tal y como estaban conformados los grupos en los centros correspondientes.

Instrumentos

Se aplicaron los mismos instrumentos a todos los alumnos encuestados: un primer cuestionario para medir variables de contexto y una segunda prueba de rendimiento en Matemáticas. Esta última consistió en un test objetivo de 20 preguntas adaptado al nivel de sexto de primaria. En su análisis se contabilizó el número de respuestas correctas.

El cuestionario de contexto estaba referido a diversos constructos o variables, de los que se tomaron para trabajar aquí los siguientes: *autoimagen*, *autoconcepto académico*, *clima de aprendizaje*, *clima social*, *metodología didáctica*, *evaluación* y *satisfacción general con el contexto escolar*. El contenido de los ítems de cada constructo se muestra en la tabla 2. La respuesta fue recogida en formato Likert de cuatro puntos, con cuantificadores lingüísticos de frecuencia o de cantidad, según el caso. Para los análisis, la puntuación de cada variable de contexto se tomó como la suma de respuestas de los ítems correspondientes. Es preciso aclarar que el conjunto de ítems de cada constructo constituía un microinstrumento resultado de un largo proceso previo de selección, reducción de ítems y validación de la escala. Ejemplos de este proceso para distintos constructos se describen con detalle en los trabajos: Biencinto, González-Barbera, García-García, Sánchez-Delgado y Madrid (2009); Joaristi, Lizasoain y Gamboa (2012); López-González, Pérez-Carbonell y Ramos-Santana (2011) y López-González, Tourón y Tejedor (2012).

Análisis estadísticos

Los análisis estadísticos se han efectuado con el programa R 2.14.0, específicamente con el entorno RGui⁸. Para ello ha sido necesario utilizar el

TABLA 2. Constructos e ítems empleados en el cuestionario de contexto

N. Ítem	Variables de contexto
Autoimagen (auto_im)	
1	¿Cómo crees tú que eres? OPCIONES: inteligente, trabajador, responsable, buen/a estudiante, simpático, tímido/a, ayudo a los demás, gusto a la gente
Autoconcepto académico (auto_acad)	
2	Sobre mí mismo/a como estudiante pienso que... OPCIONES: me gusta estudiar, soy bueno/a en la mayoría de las asignaturas, creo que soy capaz de obtener buenas notas
3	Creo que mis padres piensan que...
4	Creo que mis profesores piensan que... OPCIONES: soy inteligente, soy buen/a estudiante, puedo sacar buenas notas
Clima de aprendizaje (clim_apren)	
5	Al finalizar el curso voy a tener un buen nivel para continuar en los cursos siguientes
6	Creo que los profesores están satisfechos con la marcha general de los alumnos de este grupo de clase
7	Creo que mi clase es un lugar agradable (me gusta estar en mi clase)
8	En esta clase los alumnos tenemos muy buena comunicación con los profesores
9	En esta clase los profesores y los alumnos nos preocupamos unos de otros
10	En esta clase los alumnos nos llevamos muy bien
11	Estoy satisfecho con lo que aprendo en clase
12	La distribución de la clase permite trabajar a gusto
13	La mayor parte de mis compañeros dejan trabajar a los demás
14	La mayoría de los compañeros estamos contentos con lo que aprendemos en clase
Clima social (clim_soc)	
15	La relación entre los padres y los alumnos es cordial
16	Las relaciones entre nosotros y los profesores son agradables
17	Los alumnos colaboramos muy bien entre nosotros
18	Los alumnos nos sentimos orgullosos de esta clase
19	Los compañeros nos ayudamos para hacer las tareas de clase
20	Los profesores se interesan por lo que aprendemos
21	Los profesores se sienten orgullosos de esta clase
22	Me parece interesante lo que aprendemos en las clases
23	Me siento orgulloso de mi trabajo en clase
24	Pongo interés en los trabajos y actividades de clase
Metodología didáctica (met_doc_mat)	
25	Las actividades que me gustan más son aquellas en las que realizamos trabajos de investigación y descubrimiento
26	El profesor nos premia cuando atendemos durante las explicaciones
27	El profesor nos premia cuando cuidamos de los materiales
28	Estamos más cómodos y aprendemos más cuando las mesas están en forma de U
29	Cuando el profesor comienza un tema nuevo suele repasar los conceptos de temas anteriores
30	Cuando el profesor comienza un tema nuevo suele buscar ejemplos de la vida diaria que faciliten la explicación del tema

TABLA 2. Constructos e ítems empleados en el cuestionario de contexto (cont.)

N. Ítem	Variables de contexto
Metodología didáctica (met_doc_mat)	
31	Cuando el profesor comienza un tema nuevo suele entregar al alumnado material relacionado con el tema que se va a desarrollar
32	En clase se utilizan libros de la biblioteca del colegio o del aula
33	En clase se utilizan los ordenadores
34	En clase se utilizan materiales elaborados en equipo (por el equipo de profesores/as)
35	En clase se utilizan materiales impresos editados (libros de texto...)
36	En clase se utiliza material manipulativo (de laboratorio, juegos, instrumentos...)
Evaluación (eva_mat)	
37	El profesor/a nos evalúa a través de controles orales después de cada tema o lección
38	El profesor/a nos evalúa a través de controles en los que podemos utilizar material
39	El profesor/a nos evalúa a través de observarnos cuando trabajamos en clase
40	Nuestro profesor/a hace un seguimiento personal de cada alumno/a
41	Nuestro profesor/a explica la evaluación realizada revisándola si considera que puede haber un error
42	Cuando el alumno/a obtiene calificaciones muy bajas nuestro profesor/a habla con él
43	Cuando el alumno/a obtiene calificaciones muy bajas nuestro profesor/a suele darle tareas adicionales
44	Cuando el alumno/a obtiene calificaciones muy bajas nuestro profesor/a suele avisar a los padres y madres
Satisfacción general con el contexto escolar (satis_gene)	
45	Mis compañeros de clase son...
46	Mis profesores/as son...
47	Mi colegio es...
OPCIONES: muy malo, bastante malo, bastante bueno, muy bueno	

lenguaje de programación propio de R, cuyas instrucciones no incluimos en este documento por resultar excesivo el espacio necesario. Para el lector interesado en reproducirlos, sugerimos el apoyo de los siguientes textos: Crawley (2007); Dobson y Barnett (2008); Faraway (2006); Fox (2002) y Fox y Weisberg (2011). Es también muy recomendable el libro de Sheather (2009) que, aunque no incorpora en el texto el lenguaje necesario para los análisis, sus instrucciones pueden consultarse fácilmente en la *web* asociada al libro: www.stat.tamu.edu/~sheather/book. Por último, resulta de gran ayuda la consulta de la *web* del entorno R: <http://cran.r-project.org/>.

Resultados

Regresión clásica

Se estableció en primer lugar la solución habitual para construir un modelo de dependencia. Al tratarse de variables cuantitativas (la suma de respuestas Likert de los ítems correspondientes a los constructos) se intentó un modelo lineal: una regresión múltiple clásica. No obstante, el nivel cuantitativo de las variables no caracteriza también su continuidad. En realidad se trata de variables ordinales. Este hecho es esencial, dado que la asociación entre variables con esta métrica condiciona especialmente el cumplimiento de las condiciones del modelo

elegido: la linealidad entre los predictores y el criterio, la igualdad de varianzas (homoscedasticidad), la independencia de los errores y la ausencia de autocorrelación. El incumplimiento de esta última condición es más frecuente con modelos longitudinales, pero las tres anteriores no suelen darse si las variables no son continuas, como es el caso. El interés se centró en establecer un modelo lineal con todos los

predictores considerados, a saber: *autoimagen*, *autoconcepto*, *clima de aprendizaje*, *clima social*, *metodología didáctica*, *evaluación y satisfacción general*, averiguando al mismo tiempo el orden de entrada más adecuado en la ecuación de regresión. Para ello se realizó una regresión *paso a paso (stepwise)*, hacia adelante, con criterio de entrada *AIC*⁹. Al parar el proceso los resultados del modelo fueron los siguientes:

```
lm(formula = rto_mate ~ auto_acad + auto_im + clim_apren + clim_soc +
eva_mat + met_doc_mat + satis_gene, data = rto_mate)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-11.1461	-2.2436	-0.0671	2.3294	8.4995

Coefficients:

	Estimate	Std. Error	t value	Pr (> t)
(Intercept)	6.831377	1.561806	4.374	1.47e-05 ***
auto_acad	0.269984	0.042943	6.287	6.86e-10 ***
auto_im	0.094028	0.058677	1.602	0.10966
clim_apren	0.033017	0.051750	0.638	0.52375
clim_soc	-0.053508	0.057040	-0.938	0.34865
eva_mat	-0.121327	0.044624	-2.719	0.00677 **
met_doc_mat	-0.002059	0.032042	-0.064	0.94879
satis_gene	-0.169556	0.120949	-1.402	0.16155

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.528 on 520 degrees of freedom

Multiple R-squared: 0.1511, Adjusted R-squared: 0.1397

F-statistic: 13.22 on 7 and 520 DF, p-value: 1e-15

Como puede observarse, solo dos fueron los predictores seleccionados: *autoconcepto académico* y *evaluación*. Realizado en ajuste con ambas variables, la ecuación de regresión múltiple resultó ser la siguiente:

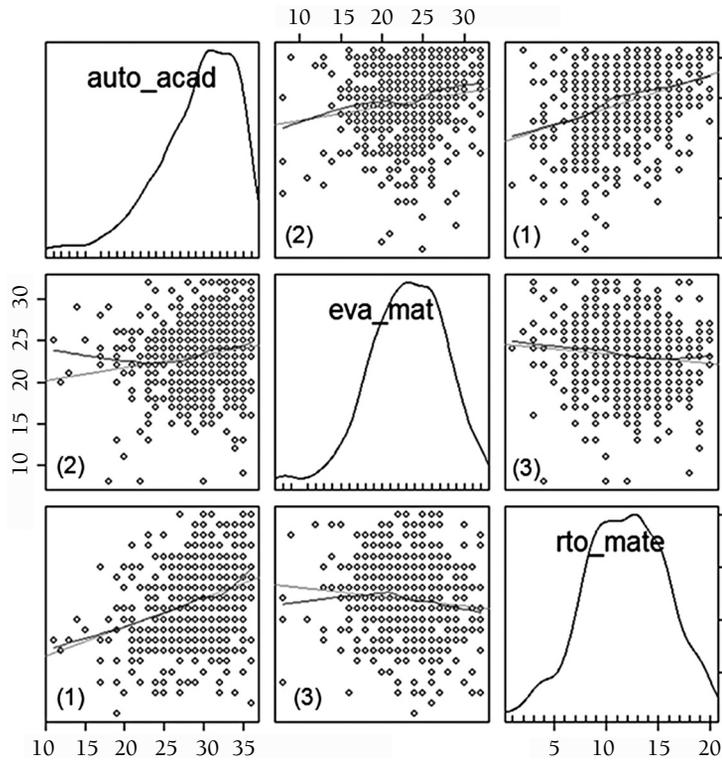
$$rto_mate = 6.3554 + 0.298 (auto_acad) - 0.1379 (eva_mat).$$

La proporción de varianza explicada por el modelo — R^2 — fue francamente baja: apenas

un 14%, a pesar de que el contraste de regresión resultó significativo ($p < .01$).

En el gráfico matricial de la figura 1 se representan los diagramas de dispersión entre las parejas de variables asociadas. Se observa una cierta tendencia lineal ascendente entre las variables *rendimiento* y *autoconcepto académico* (1), una ausencia de relación lineal entre *rendimiento* y *evaluación* (2) y una ligera asociación lineal nada aconsejable entre los dos regresores —*autoconcepto académico* y

FIGURA 1. Gráfico matricial de las variables seleccionadas en la regresión lineal múltiple

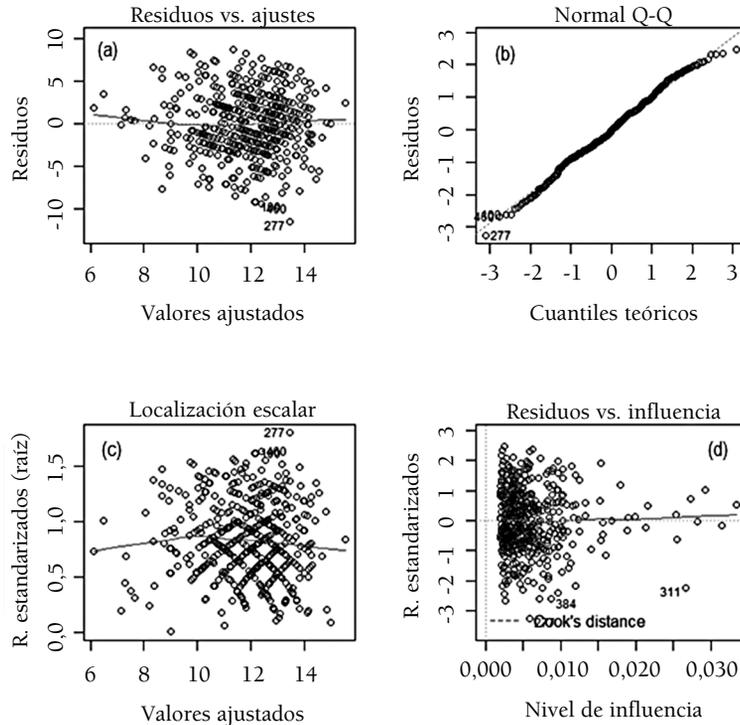


evaluación (3) —, que parece anunciar un problema de colinealidad. En los demás recuadros se dibujan los perfiles de las funciones de probabilidad de las variables: la distribución del *autoconcepto académico* es asimétrica negativa, lo que sugiere una transformación, quizá logarítmica, mientras que las distribuciones de la *evaluación* y el *rendimiento* tienden a la normalidad.

Partiendo de la idea de que cualquier patrón o tendencia en los gráficos de residuos estandarizados es un claro indicador de que el modelo no es adecuado (Sheather, 2009: 156), los gráficos básicos de diagnóstico terminaron por revelar la inconveniencia del modelo lineal construido (figura 2). En el gráfico de residuos *versus* valores ajustados (2a) tendría que haberse obtenido una línea completamente horizontal, y no fue el

caso. El gráfico Q-Q (2b) permitió juzgar la condición de *normalidad* de los residuos estandarizados y, por tanto, de los valores pronosticados del criterio. Cabría esperar una diagonal perfecta. Esta información puede completarse también con el histograma de los residuos estandarizados de la figura 3d, donde se observa una tendencia a la normalidad. El gráfico de valores ajustados *vs.* residuos estandarizados (2c) mostró un perfil claramente heteroscedástico: además de la característica nube de puntos, la línea trazada correspondiente a la media de estos residuos tendría que haber sido cero en todos los casos, pero su pequeña curvatura desvela la falta de igualdad de varianzas en algunos ellos. Por último, se produjeron numerosas *observaciones influyentes* con una mayor distancia de Cook (figura 2d), lo que se completó también con la información de la figura 5.

FIGURA 2. Gráficos básicos de diagnóstico del modelo lineal



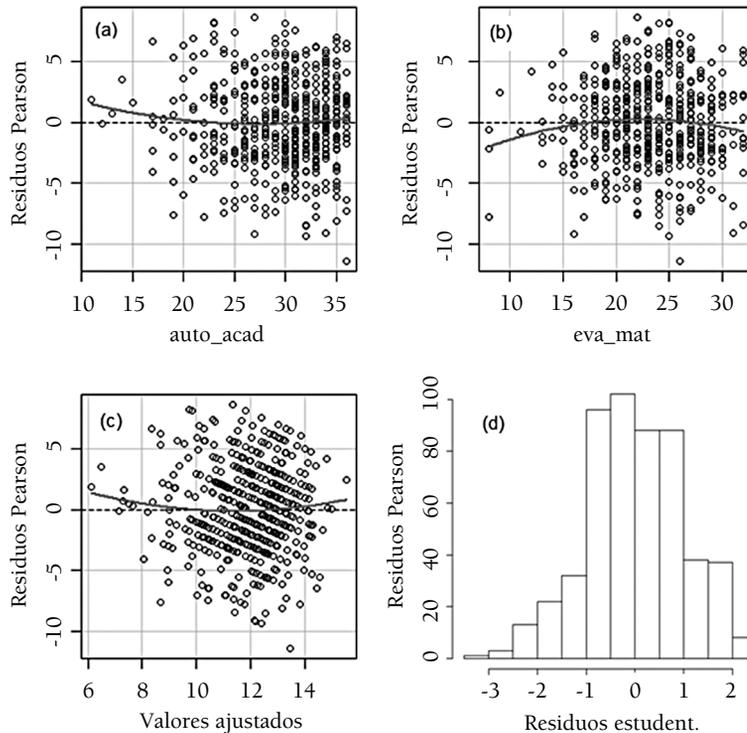
La falta de adecuación del modelo quedó confirmada igualmente en los gráficos de residuos de Pearson *versus* valores pronosticados (figura 3). La línea curva representada en el gráfico 3c corroboró de nuevo que el modelo no se adecuaba por incumplimiento de la condición de linealidad, lo que también sucedía con el predictor *evaluación* (3b). En el caso del *autoconcepto académico* (3a) se consiguió una recta casi horizontal, excepto en la zona del extremo izquierdo (muy probablemente por el efecto de alguna de las observaciones influyentes que tendría que eliminarse). Este último gráfico puede considerarse *nulo* por no mostrar ningún patrón definitivo, lo cual es un buen indicador de que el predictor mantiene una relación lineal con el criterio. No obstante, aunque desde el punto de vista de la condición de linealidad es interesante obtener gráficos nulos, el dato es insuficiente para poder juzgar la adecuación del modelo de forma definitiva (Fox y Weisberg, 2011: 288).

La información de los anteriores gráficos se combinó con la prueba de no aditividad de Tukey (Tukey, 1949, trabajada en Fox y Weisberg, 2011: 289). Se realizó el test añadiendo al modelo un término cuadrático —el cuadrado de los valores ajustados— y reajustándolo de nuevo. Sus niveles de significación se compararon con la distribución normal estándar:

Test	stat	Pr(> t)
auto_acad	1.188	0.235
eva_mat	-1.930	0.054
Tukey test	1.147	0.251

Un valor significativo hubiera desvelado falta de linealidad, pero el test confirmó la impresión visual de la relación lineal entre el *autoconcepto académico* y el *rendimiento* (figura 3a), así como la ausencia de esta condición entre el predictor *evaluación* y el *rendimiento* (figura 3b). Aun así, el modelo globalmente superó el

FIGURA 3. Gráficos de residuos de Pearson e histograma de residuos estudentizados



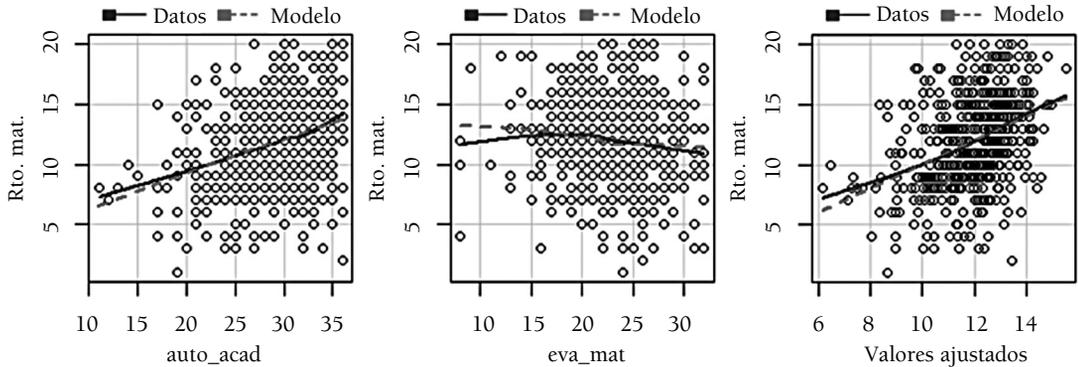
test, dado que la probabilidad asociada no resultó significativa ($p = .251$). Puede entenderse, por tanto, que la falta de linealidad de la función de regresión era debida sobre todo al predictor *evaluación*, lo que aconsejaba su transformación. Otra posible solución consistía en explorar un ajuste con otro modelo más adaptado a la métrica de la variable de respuesta. No debe olvidarse que el *rendimiento* era un criterio no continuo, ordinal; en realidad una variable de conteo: número de aciertos en las respuestas. Retomaremos enseguida este punto.

Una variante de los gráficos básicos de residuos son los gráficos marginales del modelo (propuestos por Cook y Weisberg, 1997). En ellos se representa la variable de respuesta en el eje vertical y los predictores o valores ajustados en el horizontal —figura 4—. Estos gráficos son marginales en el sentido de que ajustan los

datos con un predictor ignorando las demás variables de la función. Permiten comparar dos tipos de ajuste: uno paramétrico —representado por líneas discontinuas y que asume una relación lineal— y otro semiparamétrico con regresión suavizada —con líneas continuas y no lineal—. Las diferencias pueden valorarse por medio del test que proponen Azzalini y Bowman (1993), o bien visualmente con los gráficos de la figura 4. En este caso se confirmó de nuevo la falta de linealidad del predictor *evaluación*.

En la figura 5 se presentan dos gráficos para detectar casos de influencia. En el gráfico de la derecha aparecen relacionados los residuos estudentizados y los cuantiles de una distribución $t(n-k-2)$. La figura se construye empleando unos estimadores paramétricos *bootstrap*¹⁰ que generaran una especie de *envoltura* de un

FIGURA 4. Gráficos marginales del modelo lineal



95% de confianza alrededor de los residuos estudentizados (Atkinson, 1985). Los valores que se sitúan fuera corresponden a los casos *outliers* con residuos estudentizados mayores que ± 3 (los sujetos identificados con las etiquetas 100, 277 y 460). Esta información se completó

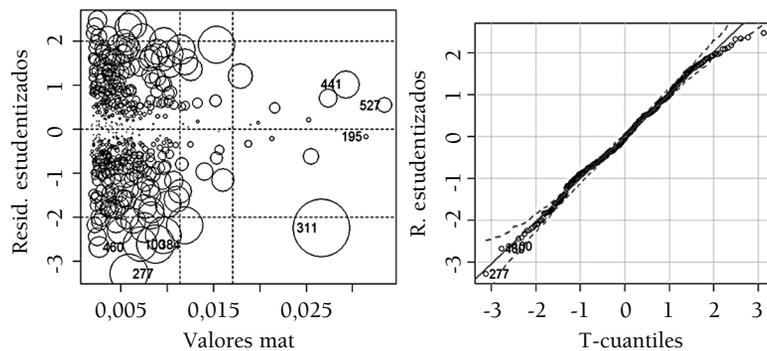
aplicando sobre el residuo estudentizado de mayor tamaño el test corregido de Bonferroni (en este caso, el del sujeto 277). Su valor fue -3.2844 y resultó no significativo, lo que indicaba que no tenía una influencia determinante en el modelo:

```
No Studentized residuals with Bonferroni p < 0.05
Largest |rstudent|:
rstudent unadjusted p-value Bonferonni p
277 -3.284426 0.0010902 0.57562
```

El gráfico de *burbujas* de la figura 5 combinó los valores *hat* con los residuos estudentizados y las distancias de Cook, detectando cuáles eran las observaciones influyentes: los casos 100, 195,

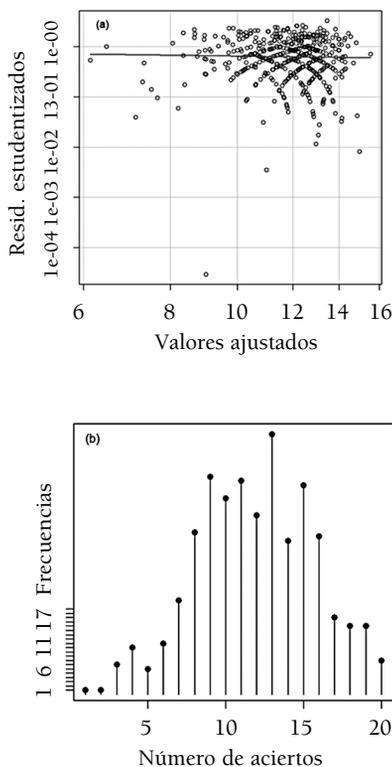
277, 311, etc. Atendiendo a la información analizada hasta el momento, se confirmaba que el modelo lineal construido no ofrecía una solución correcta. Utilizando argumentos estadísticos,

FIGURA 5. Gráficos para detectar observaciones influyentes y valores alejados



correspondía entonces explorar qué transformaciones de los datos podían facilitar un mejor ajuste sin variar el tipo de modelo empleado. Otra solución se encontraba en cambiar el modelo directamente. Respecto a la primera opción, las transformaciones podían efectuarse sobre los predictores —en este caso sobre la variable *evaluación*, que había resultado con una relación no lineal con el criterio— o sobre la misma variable de respuesta, el *rendimiento*.

FIGURA 6. Gráfico de nivel de difusión y distribución de la variable rendimiento



Un fácil diagnóstico que sirviera para sugerir alguna posible transformación del criterio se obtuvo al construir el gráfico de nivel de difusión —*spread level plot*—. Este gráfico (figura 6a) permitió comprobar si el tamaño de los residuos cambiaba sistemáticamente

con los valores pronosticados. La ligera pendiente de la línea sugería una pequeña asociación que podría resolverse con una transformación potencial de la variable de respuesta, tal que: $[rto_mate]^{1.1967}$. Sin embargo, al ser el valor de la potencia prácticamente uno, los cambios no fueron apreciables. Una información complementaria se consiguió con el test de varianza no constante, que exploró la relación entre la difusión residual y los valores ajustados a partir de cualquier combinación posible de predictores (aunque en este caso eran solo dos). La falta de significación del estadístico *Chi* cuadrado descartó tal asociación:

Suggested power transformation: 1.196765
 Non-constant Variance Score Test
 Variance formula: ~ fitted.values
 Chisquare = 0.3292709 Df = 1 p = 0.5660887

Regresión de Poisson

Comprobada la inadecuación del primer modelo construido, una segunda posibilidad consistía en ajustar otro tipo modelo que atendiera a la métrica de la variable de respuesta. Como ha quedado señalado, el *rendimiento* había sido operacionalizado como el número de respuestas correctas, esto es, una variable de recuento. En la figura 6b se muestra la distribución que siguieron sus valores observados para los 528 alumnos encuestados. El modelo adecuado para una respuesta de este tipo es un modelo lineal generalizado, concretamente una regresión de Poisson.

Se ejecutó el modelo de Poisson del mismo modo que se procedió con la regresión clásica: se buscó establecer una óptima combinación de predictores (tanto en el orden de entrada en la ecuación como en relación a los parámetros correspondientes) con la idea de explicar al máximo la variabilidad del *rendimiento*. Se inició el proceso con los siete predictores y se trabajó con un enlace *log*:

```
glm(formula = rto_mate ~ auto_acad + auto_im + clim_apren + clim_soc +
eva_mat + met_doc_mat + satis_gene, family = poisson(log), data = rto_mate)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.8312	-0.6867	-0.0287	0.6544	2.3802

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	7.219334	1.487363	4.854	1.21e-06 ***
auto_acad	0.256975	0.039477	6.509	7.54e-11 ***
auto_im	0.097944	0.055695	1.759	0.07865 .
clim_apren	0.029840	0.050053	0.596	0.55107
clim_soc	-0.046937	0.054584	-0.860	0.38984
eva_mat	-0.122506	0.043556	-2.813	0.00491 **
met_doc_mat	0.003566	0.031021	0.115	0.90848
satis_gene	-0.205422	0.118095	-1.739	0.08195 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 686.64 on 527 degrees of freedom
Residual deviance: 587.77 on 520 degrees of freedom
AIC: 2856.8
Number of Fisher Scoring iterations: 5

El programa terminó por seleccionar cuatro predictores de los siete: *autoconcepto académico*, *evaluación*, *satisfacción general* y *autoimagen*. Ajustando el modelo de Poisson con estas cuatro variables, las ecuaciones que se obtuvieron fueron las siguientes:

$$\log \hat{\mu} = 1.9968 + .0234 (\text{auto_acad}) - .0106 (\text{eva}) - .0162 (\text{satis_gene}) + .0074 (\text{auto_im})$$

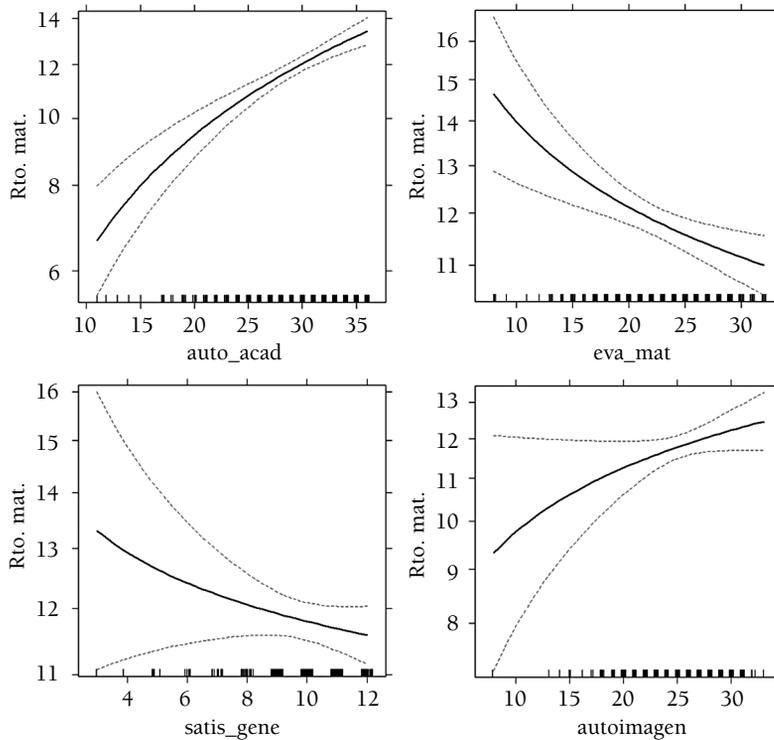
$$\hat{\mu} = \exp [1.9968 + .0234 (\text{auto_acad}) - .0106 (\text{eva}) - .0162 (\text{satis_gene}) + .0074 (\text{auto_im})]$$

Los coeficientes del modelo de la primera ecuación se interpretan como los efectos de los predictores sobre el logaritmo de la escala de recuento (la escala del predictor lineal). La segunda ecuación es la función exponencial de los coeficientes y recoge su efecto multiplicativo

sobre la escala de recuento (la escala de la respuesta). Estos componentes del MLG se distinguen en los gráficos de los elementos de la regresión de Poisson que se muestran en la figura 7. Los ejes verticales están medidos en la escala del predictor lineal, o sea, son los logaritmos de la escala de recuento. Los ajustes son curvas que se dibujan en la escala original del predictor correspondiente, en lugar de las líneas rectas que se obtendrían en la escala transformada. Las líneas intermitentes alrededor de los efectos fijos delimitan los intervalos confidenciales de estimación al 95%.

La revisión de estos gráficos permitió comparar el distinto comportamiento de los predictores seleccionados por el programa. Mientras que en los gráficos del *autoconcepto académico* y de la *evaluación* las franjas del 95% resultaron próximas a las curvas, no sucedió lo mismo con los

FIGURA 7. Gráficos de los componentes de la regresión de Poisson



predictores *satisfacción general* y *autoimagen*: las dispersiones que mostraron a lo largo del eje vertical eran muy amplias en relación a las pendientes de las curvas. Podría incluso haberse encajado entre ellas una recta en lugar de una curva, lo que resultaba un claro indicador de su flojo poder explicativo para con el *rendimiento* (a pesar de haber cumplido con los indicadores correspondientes para entrar a formar parte del modelo). De hecho, los coeficientes de regresión de estos predictores (.0819 para *satisfacción* y .0786 para *autoimagen*) habían

obtenido probabilidades asociadas entre .10 y .05, lo que apuntaba a un error de tipo I (establecer significación cuando en realidad no se produce).

Un análisis de varianza entre los dos modelos —la regresión clásica y la regresión de Poisson— permitió comprobar que la distinta proporción de variabilidad explicada por uno y otro era significativa. Es decir, se producía una ganancia importante en el ajuste si se empleaba la regresión de Poisson:

Analysis of Variance Table

Model 1: rto_mate ~ auto_acad + eva_mat

Model 2: rto_mate ~ auto_acad + eva_mat + satis_gene + auto_im

	Res.Df	RSS	Df	Sum of Sq	Pr(>Chi)
1	525	6541.2			
2	523	587.8	2	5953.3	< 2.2e-16 ***

No obstante, cuando se realizó el análisis de desviación tipo II sobre el MLG, solo dos de los cuatro predictores seleccionados obtuvieron

una clara significación, conclusión acorde con lo comentado anteriormente en la lectura de los gráficos de la figura 7.

```
Analysis of Deviance Table (Type II tests)
Model: poisson, link: log
Response: rto_mate
Terms added sequentially (first to last)
```

Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			527	686.64
auto_acad	1	77.282	526	609.36 < 2.2e-16 ***
eva_mat	1	16.494	525	592.87 4.881e-05 ***
satis_gene	1	2.602	524	590.26 0.1067
auto_im	1	2.451	523	587.81 0.1175

En este segundo modelo se revisaron las medidas de influencia, quedando señalados con mayores distancias de Cook los casos 295, 311 y 384, y los sujetos 6, 195 y 311 con los mayores valores *hat* (figura 8). Para valorar la trascendencia de su influencia, el programa R permite construir en paralelo otro modelo sin contenerlos, ajustando después las dos funciones de Poisson (con y sin medidas de influencia). Se interpretan los resultados juzgando los

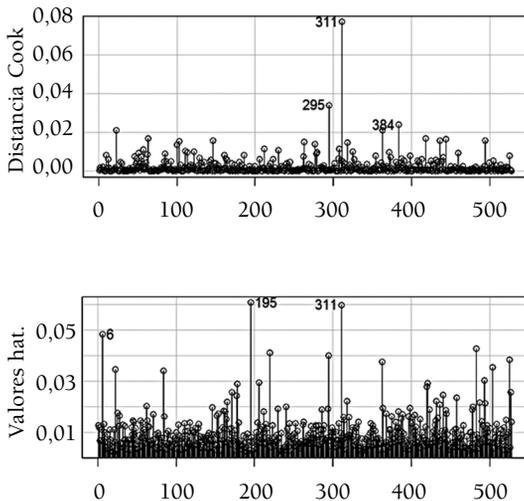
posibles cambios en los tamaños de los coeficientes de regresión de ambos modelos. En este caso no fueron importantes, por lo que no vemos necesario incluirlos aquí.

Discusión

Los dos modelos de dependencia contruidos para explicar el rendimiento amplían las posibilidades de análisis estadístico más allá de los habituales indicadores descriptivos y correlacionales propios de la evaluación de sistemas educativos. Existe una amplia variedad de modelos, opciones analíticas y gráficos que facilitan la explicación del producto a partir de variables antecedentes. El presente trabajo ha querido ser solo una pequeña muestra de tales posibilidades. El interés se ha centrado en animar al evaluador a aprovechar mejor la información de sus datos, pero cuidando los procedimientos y los recursos que utilice para ello.

En la construcción de los modelos hemos utilizado el procedimiento de regresión *paso a paso* por considerarlo más conveniente que la estrategia de incluir los predictores solo con algún criterio sustantivo. No hay que olvidar que las variables de un modelo teórico una

FIGURA 8. Gráficos de diagnóstico de las medidas de influencia en la regresión de Poisson



vez medidas adquieren unas propiedades matemáticas que es necesario tener en cuenta (Townsend y Ashby, 1984). Así, la selección de los predictores ha podido resultar decepcionante en los dos modelos (dos predictores en la regresión clásica y cuatro en la de Poisson), pero es claro que el resto de los constructos (*clima de aprendizaje, clima social y metodología didáctica*) no han añadido ninguna explicación significativa al *rendimiento*.

Sin embargo, el objetivo no se ha limitado a la construcción de un modelo sin más. El diagnóstico del modelo lineal ha desvelado la falta de normalidad del *autoconcepto académico*, la no linealidad entre la *evaluación* y el *rendimiento* y el incumplimiento de la condición de homoscedasticidad de los residuos del modelo, por tanto, de las distribuciones condicionadas de los pronósticos. Y ello dando por hecho que todas las variables (predictores y criterio) poseían una métrica continua cuando, tal y como habían sido operacionalizadas en la investigación previa, su medida era ordinal. Esta falta de cumplimiento de las hipótesis básicas del modelo lineal no se habría desvelado si únicamente se hubieran empleado los indicadores de bondad de ajuste.

Al construir el modelo lineal generalizado con la regresión de Poisson hemos dado paso a un ajuste *correcto* desde un punto de vista matemático. Las dos ecuaciones obtenidas respetan la métrica de las variables implicadas. Esto ha permitido trabajar con el logaritmo

del número de respuestas acertadas, evitando forzar la construcción de un plano de ajuste lineal *irreal*. Según el análisis de varianza efectuado, el modelo de Poisson ha resultado eficaz para explicar el *rendimiento*. La desviación dividida por los grados de libertad ha mostrado un parámetro de dispersión próximo a uno, lo que es un claro indicador de buen ajuste. No obstante, respecto a la selección definitiva de predictores, la *satisfacción general* y la *autoimagen* no han resultado adecuadas para explicar el *rendimiento* tampoco en esta segunda opción, a la vista del análisis de desviación y de los gráficos de la figura 7.

Como conclusión, hemos de insistir en las siguientes ideas: a) las actuales herramientas de construcción de modelos permiten un tratamiento explicativo del producto a partir de variables antecedentes del sistema —en este caso, de variables de contexto—, b) la identificación de modelos debe realizarse atendiendo a sus características matemáticas y a las propiedades métricas de las variables que se combinan, c) el estudio de los residuos de una regresión y el análisis de sus gráficos proporcionan una información sencilla, cómoda y orientativa, que debe irse contrastando y completando con el conjunto de los elementos de diagnóstico y d) cualquier modelo explicativo es solo una *representación* de los datos; su interpretación ha de combinarse con los elementos sustantivos del modelo teórico que sostienen el sistema.

Notas

¹ Este trabajo se desarrolló en el contexto del proyecto de investigación *Análisis de Variables de Contexto: Diseño de cuestionarios de contexto para la evaluación de sistemas educativos* (Proyecto AVACO, I+D+I, 2006-2008. Código SEJ 2005-05 923). Investigador principal: Jesús Jornet Meliá, del Departamento de Métodos de Investigación y Diagnóstico en Educación de la Universitat de València.

² La literatura monográfica sobre modelos longitudinales es actualmente muy amplia y en crecimiento constante. Por citar solo algunos ejemplos, cabe mencionar los recientes textos de Berridge y Crouchley (2011) y Heck, Thomas y Tabata (2010).

³ Otras opciones se encuentran en la regresión no paramétrica, como la regresión suavizada (Venables y Ripley, 2002: 8), la regresión por *splines* y la regresión *Kernel* (ambas expuestas en Faraway, 2006, 11).

⁴ Los valores *hat* son los elementos de la diagonal principal de la matriz *hat*, matriz que relaciona los residuos con el componente aleatorio del modelo (representado por ε). La expresión es: $\hat{\varepsilon} = (\mathbf{I}-\mathbf{H}) \boldsymbol{\varepsilon}$, donde \mathbf{I} es una matriz identidad. En regresiones con más de dos predictores los valores *hat* son determinantes para detectar las observaciones de influencia, debido a que el diagnóstico por medio de gráficos se complica bastante.

⁵ En la regresión por *splines* se construye una curva de valores pronosticados no de forma continua, sino por segmentos discretos, realizando interpolaciones no paramétricas en determinados puntos llamados nodos. Pueden consultarse ejemplos muy ilustrativos en Wood (2006: 3).

⁶ Según comentan Fox y Weisberg (2011), las matemáticas que encierra esta expresión no son fáciles de entender. Una explicación en pocas palabras es la siguiente: una transformación no lineal de la media, dada por la expresión $g = [\mu(x)]$, puede modelarse como una combinación lineal de las variables explicativas, dada por $\eta(x)$. Para entender mejor este *cambio* recomendamos la lectura del artículo de López-González y Ruiz-Soler (2011), donde se explica con un sencillo ejemplo.

⁷ Una presentación completa de los MLGs se encuentra en Dobson y Barnett (2008: 3) y en López-González y Ruiz-Soler (2011).

⁸ Hemos respetado el formato habitual del programa R en las *salidas* de los análisis que se realizan. De este modo es como habitualmente se muestran en los numerosos manuales sobre modelos de dependencia que trabajan con este programa.

⁹ El criterio de información *AIC* fue propuesto por Akaike (1974) como una alternativa distinta a la identificación de modelos con hipótesis nulas. En el contexto de la regresión se utiliza para la selección de predictores. Responde a la expresión: $AIC = 2k - 2\ln(L)$, donde k es el número de parámetros incluidos en el modelo y L es el valor maximizado de la función de verosimilitud del modelo. En cada paso de la regresión *stepwise* se selecciona el predictor que presenta un menor *AIC*.

¹⁰ El *bootstrap* es un método de remuestreo para construir intervalos de confianza de un estadístico aproximando su sesgo o varianza.

Referencias bibliográficas

- AGRESTI, A. (2002). *Categorical Data Analysis* (2ª ed.) New York: Wiley.
- AKAIKE, H. (1974). A New Look at Statistical Model Identification, *IEEE Transactions on Automatic Control*, 19 (6), 716-723.
- ATKINSON, A. C. (1985). *Plots, Transformations and Regression: An introduction to Graphical Methods of Diagnostic Regression Analysis*. Oxford: Clarendon Press.
- ATO, M. y LÓPEZ-GARCÍA, J. J. (1996). *Análisis estadístico para datos categóricos*. Madrid: Síntesis.
- ATO, M., LOSILLA, J. L., NAVARRO, J., PALMER, A. y RODRIGO, M. (2005). *Modelo lineal generalizado*. Girona: EAP.
- AZZALINI, A. y BOWMAN, A. (1993). On the use of nonparametric regression for checking linear relationships, *Journal Royal Statistical Society, Ser. B.*, 55, 549-559.
- BERRIDGE, D. M. y CROUCHLEY, R. (2011). *Multivariate Generalized Linear Models using R*. Boca Ratón: CRC/Press.
- BIENCINTO, C., GONZÁLEZ-BARBERA, C., GARCÍA-GARCÍA, M., SÁNCHEZ-DELGADO, P. y MADRID, D. (2009). Diseño y propiedades psicométricas del AVACO-EVADIE. Cuestionario para la evaluación de la atención a la diversidad como dimensión educativa en las instituciones escolares, *Relieve*, 15 (1).
- COENDERS, G., BATISTA, J. M. y SARIS, W. E. (2005). *Temas avanzados en modelos de ecuaciones estructurales*. Madrid: La Muralla.
- COOK, R. D. y WEISBERG, S. (1997). Graphics for assessing the adequacy of regression models, *Journal of the American Statistical Association*, 92 (438), 490-499.
- CRAWLEY, M. J. (2007). *The R Book*. Chichester: Wiley & Sons, Ltd.

- DE LA ORDEN, A. (1999). Los modelos jerárquicos lineales y su aplicación a la investigación educativa, *Revista de Investigación Educativa*, 17 (2) 425-428.
- DE LA ORDEN, A. y JORNET, J. (2012). La utilidad de las evaluaciones de sistemas educativos: el valor de la consideración del contexto, *Bordón*, 64 (2), 69-88.
- DE LEEUW, J. y MEIJER, E. (eds.) (2008). *Handbook of multilevel analysis*. New York: Springer.
- DOBSON, A. J. y BARNETT, A. (2008). *An Introduction to Generalized Linear Models* (3ª ed.) Boca Raton, FL.: Chapman and Hall/CRC.
- EFRON, B. (2003). The statistical century. En PANARETOS, J. (ed.), *Stochastic Musings: Perspectives from the Pioneers of the Late 20th Century*. New Jersey: Lawrence Erlbaum Associates, 29-44.
- FARAWAY, J. J. (2006). *Extending the Linear Model with R. Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Boca Raton, FL.: Chapman & Hall/CRC.
- FOX, J. (2008). *Applied regression analysis and generalized linear models* (2ª ed.). Thousand Oaks: Sage.
- FOX, J. y WEISBERG, S. (2011). *An R companion to applied regression* (2ª ed.). Thousand Oaks: Sage.
- GAVIRIA, J. L. y CASTRO, M. (2005). *Modelos jerárquicos lineales*. Madrid: La Muralla.
- HARDIN, J. W. y HILBE, J. M. (2007). *Generalized linear models and extensions*. College Station, TX: Stata Press.
- HECK, R. H., THOMAS, S. L. y TABATA, L. N. (2010). *Multilevel and longitudinal modeling with IBM SPSS*. New York: Routledge.
- JOARISTI, L., LIZASOAIN, L. y GAMBOA, E. (2012). Construcción y validación de un instrumento de medida del Nivel Socioeconómico y Cultural (NSE) de estudiantes de Educación primaria y secundaria, *Bordón*, 64 (2), 151-172.
- JORNET, J. M., GONZÁLEZ-SUCH, J. y PERALES, M. J. (2012). Diseño de cuestionarios de contexto para la evaluación de sistemas educativos: optimización de la medida de constructos complejos, *Bordón*, 64 (2), 89-110.
- KOLENIKOV, S., STEINLEY, D. y THOMBS, L. (eds.) (2010). *Statistics in the social sciences. Current methodological developments*. Hoboken, NJ: Wiley.
- LÓPEZ-GONZÁLEZ, E. (1994). Importancia del estudio de residuos para el análisis de las condiciones de aplicación de la regresión, *Bordón*, 46 (1), 53-68.
- LÓPEZ-GONZÁLEZ, E., PELEGRINA, M. y RUIZ-SOLER, M. (2002). Estimación de parámetros en el modelo lineal general y en MLGs: diferencias e interpretación. *Metodología de las Ciencias del Comportamiento* (vol. especial), 341-345.
- LÓPEZ-GONZÁLEZ, E., PÉREZ-CARBONELL, A. y RAMOS-SANTANA, G. (2011). Modelos complementarios al análisis factorial en la construcción de escalas ordinales: un ejemplo aplicado a la medida del clima social aula, *Revista de Educación*, 354, 369-397.
- LÓPEZ-GONZÁLEZ, E. y RUIZ-SOLER, M. (2011). Análisis de datos con el modelo lineal generalizado: una aplicación con R, *Revista Española de Pedagogía*, 248, 59-80.
- LÓPEZ-GONZÁLEZ, E., TOURÓN, J. y TEJEDOR, F. J. (2012). Diseño de un micro-instrumento para medir el clima de aprendizaje en cuestionarios de contexto, *Bordón*, 64 (2), 111-126.
- NELDER, J. y WEDDERBURN, R. (1972). Generalized Linear Models, *Journal of the Royal Statistical Society* (a), 135, 370-384.
- RAUDENBUSH, S. W. y BRYK, A. S. (2002). *Hierarchical linear models* (2ª ed.) London: Sage.
- RUIZ-SOLER, M. y LÓPEZ-GONZÁLEZ, E. (2009). El entorno R para el análisis estadístico de datos: ventajas de su uso en la docencia y la investigación, *Revista Española de Pedagogía*, 243, 255-274.
- SHEATHER, S. J. (2009). *A Modern Approach to Regression with R*. New York: Springer.
- TOWNSEND, J. T. y ASHBY, F. G. (1984). Measurement scales and statistics: The misconception misconceived, *Psychological Bulletin*, 96, 394-401.
- VENABLES, W. N. y RIPLEY, B. D. (2002). *Modern Applied Statistics with S* (4ª ed.) New York: Springer.
- WOOD, S. N. (2006). *Generalized Additive Models. An Introduction with R*. Boca Ratón, FL.: Chapman & Hall/CRC.

Abstract

Explanation of performance by context. Some proposals from graphic and statistical analysis

The assessment of education systems suffers from certain weaknesses, among which is the limited statistical treatment of the information obtained. An improvement could be to pay more attention to explaining the product from the context. This paper aims to illustrate, in a simple way, possible statistical models for linking antecedent variables (context variables) with learning outcomes (performance) using some of the graphical tools that provide the dependency models based on regression. More specifically, an ordinary linear model and a generalized linear model of Poisson regression are carried out. For this objective, the conditions of application for these models are reviewed examining some elements of diagnostic in regression: basic diagnostic plots; unusual data (outliers, hat values, influence measures) or residual plots, for example. R is the software program used in all statistics analyses, because we think it has a good graphic interface and very simple outputs. Then we offer some suggestions for making decisions based on the information provided by graphical representations from model residuals.

Key words: *Linear model, Generalized linear model, Regression diagnostics, Poisson regression.*

Résumé

Explication de la performance à partir du contexte. Quelques propositions d'analyse graphique et statistique

L'évaluation des systèmes éducatifs souffre de quelques faiblesses, parmi lesquelles nous trouvons le traitement statistique limité de certaines informations obtenues. Un progrès pourrait consister à porter plus d'attention aux analyses explicatives du produit à partir du contexte. Ce travail cherche à exemplifier d'une façon simple des possibles modèles statistiques capables de relier les variables antécédentes (variables de contexte) avec les résultats d'apprentissage (la performance). Et le faire en utilisant certains outils graphiques offerts par les modèles de dépendance basés sur la régression. Plus spécifiquement, nous construisons un modèle linéaire classique et un modèle linéaire généralisé de régression de Poisson. Pour cela, nous révisons les conditions d'application des modèles proposés en employant fondamentalement des éléments diagnostics fournis par l'étude des déchets : des graphiques basiques d'analyse de régression, des données pas habituelles —*outliers*, valeurs *hat*, valeurs d'influence— et des graphiques des déchets dans ses différentes versions. Le programme statistique qui fait possible accomplir cette tâche est le R, compte tenu de son convenable *interface* graphique et sa maniable sortie d'analyses qu'il nous rapporte. L'information fournie par les représentations graphiques des déchets et les indicateurs diagnostics, orientent la prise de décisions au cours des étapes ultérieures d'analyse.

Mots clés: *Modèle linéaire, Modèle linéaire généralisé, Analyse de régression, Régression de Poisson.*

Perfil profesional de los autores

Emelina López-González

Profesora titular de Psicoestadística y de Métodos de Investigación en Educación. Departamento de Métodos de Investigación y Diagnóstico en Educación. Facultad de Filosofía y Ciencias de la Educación, Universidad de Valencia.

Sus principales líneas de investigación son: desarrollo y aplicación de los modelos lineales generalizados y modelos multivariados para el análisis de datos categóricos en el ámbito de las Ciencias Sociales.

Correo electrónico de contacto: emelina.lopez@uv.es

José González-Such

Profesor titular en el departamento MIDE-UVEG. Facultad de Filosofía y CC. Educación. Universidad de Valencia.

Coordinador de la Unidad de Innovación InnovaMide del grupo GEM (MIDE-UVEG). Su trabajo se centra en la Medición y Evaluación Educativas, siendo sus líneas preferentes de investigación: diseño de pruebas e instrumentos de medición educativa, evaluación de la docencia y diseño y evaluación de materiales de innovación docente sustentados sobre nuevas tecnologías.

Correo electrónico de contacto: gonzalej@uv.es.

Luis Lizasoain

Profesor titular de Estadística. Departamento de Métodos de Investigación y Diagnóstico en Educación. Facultad de Filosofía y Ciencias de la Educación UPV-EHU.

Su labor investigadora está centrada en la evaluación de programas y sistemas educativos, con numerosas publicaciones en revistas especializadas. Además de su labor docente en la UPV-EHU, imparte cursos de postgrado en universidades latinoamericanas.

Correo electrónico de contacto: luis.lizasoain@ehu.es

