



TÉCNICAS ESTADÍSTICAS UTILIZADAS ACTUALMENTE EN LA GENERACIÓN DE MODELOS EXPLICATIVOS DEL LOGRO EDUCATIVO A GRAN ESCALA

Alicia Alelí Chaparro Caso López
Universidad Autónoma de Baja California

Fernando Martínez Abad
Universidad de Salamanca

Área temática: 12. Evaluación educativa

Línea temática: 10. Evaluación a gran escala

Tipo de ponencia: Aportaciones teóricas

Resumen:

La forma como son analizados los distintos grupos de factores asociados al logro académico, varía en función del estudio y los intereses de la investigación. En una revisión del estado de la cuestión en bases de datos especializadas de los últimos 15 años a partir de metabuscadores como EBSCO para publicaciones en inglés y Redalyc para publicaciones en español, se lograron identificar multitud de estudios que analizan las variables asociadas al logro. Sin embargo, centrándonos sólo en los estudios a gran escala, cuyo tamaño de muestra es superior a 1,000 sujetos, se pueden identificar investigaciones con la aplicación de diferentes técnicas estadísticas, elegidas en función tanto de los objetivos de la investigación como, principalmente, de las características de las variables analizadas. En este trabajo presentamos una revisión del estado de la cuestión de las actuales técnicas estadísticas aplicadas en el estudio del logro académico, a fin de tener una perspectiva sobre el aporte de estas técnicas en el campo de la investigación educativa. De maneja general, las técnicas que se pueden identificar principalmente son la regresión múltiple, modelamiento de ecuaciones estructurales (o SEM, por sus siglas en inglés) y análisis multinivel. De forma incipiente está surgiendo la aplicación de otras técnicas relacionadas con la econometría y la minería de datos. La revisión realizada nos lleva a identificar que existen técnicas que por su facilidad son más populares y de mayor uso, sin embargo, el investigador educativo debe considerar la amplia gama de técnicas aplicables a los diferentes tipos de datos.

Palabras clave: Evaluación educativa, Análisis estadístico, Modelos educativos, Logro Educativo, Métodos de evaluación

Introducción

Como señaló Ballesteros (2014), actualmente los sistemas educativos han acumulado gran cantidad de información en bases de datos, para el manejo de esa información se requieren procesos de análisis que permitan identificar y analizar la información importante, pues el uso de herramientas estadísticas clásicas es casi imposible cuando se manejan volúmenes grandes de información. Es así, que cada vez más la investigación educativa, sobre todo la dedicada a los estudios a gran escala, se ha apropiado de técnicas y procedimientos provenientes de otras ciencias y disciplinas, para aplicarlos al estudio de los fenómenos educativos. Pues la naturaleza de los objetos de estudio en la Educación es tan compleja y diversa, que hace necesario la implementación de herramientas que garanticen un tratamiento adecuado a los datos que se recogen, a fin de llegar a conclusiones válidas y comprensivas de los distintos fenómenos.

Uno de los temas de estudio que más atención ha generado en la investigación educativa es precisamente el de los factores asociados al logro educativo, acumulando a la fecha un corpus de conocimiento amplio. El estudio de los factores asociados al logro educativo es una línea de investigación que ha prevalecido en el ámbito de la Educación desde hace varias décadas. El interés por comprender las variables que explican los resultados que los estudiantes obtienen en evaluación de logro, ha promovido el desarrollo de distintas propuestas metodológicas y técnicas de análisis. El logro académico, al igual que muchos otros objetos de estudio en Educación, es un fenómeno multicausal en el que intervienen muchas variables (Tejedor, 2003) interrelacionadas (Lee & Shute, 2010), por lo que su identificación es una labor compleja; y es necesario la utilización de diversos y eclécticos procedimientos (Tejedor, 2003). El uso de cada procedimiento obedece, a su vez, tanto a los objetivos de cada investigación, como a la naturaleza misma de las variables analizadas.

Desarrollo

Las técnicas estadísticas aplicadas a la contrastación empírica de modelos explicativos del logro educativo, a gran escala, se pueden agrupar principalmente en Regresión múltiple, Análisis de senderos (o SEM), Análisis multinivel, Modelos econométricos y Minería de datos:

Regresión múltiple. Es quizás el procedimiento de análisis más utilizado, tanto en estudios a gran escala, como con tamaños de muestra inferiores. En esta técnica se busca identificar la relación entre una variable dependiente, en este caso el logro o rendimiento académico, y un conjunto de variables explicativas, que generalmente pueden ser cuantitativas continuas o cualitativas dicotómicas (Guimaraes & Sampaio, 2013; Nath, 2012; Stull, 2013). El método de estimación de parámetros más habitual es la regresión por mínimos cuadrados ordinarios (OLS por sus siglas en inglés), aunque, también es común la estimación por máxima verosimilitud (ML, por sus siglas en inglés). En caso de incumplimiento de los supuestos previos de normalidad o homocedasticidad, es necesario recurrir a estimaciones robustas, transformaciones en las variables e incluso la aplicación del método de estimación por mínimos cuadrados generalizados (GLS por sus siglas

en inglés). Existen interesantes variaciones a la técnica de la regresión múltiple, como la regresión cuantil (Guimaraes & Sampaio, 2013; Haile & Nguyen, 2008), que es una técnica de modelamiento que permite la descripción del cuantil condicional de una variable dependiente dadas las variables explicativas. Otro tipo de regresión reportada en la literatura es el análisis truncado (Kaighobad & Allen, 2008), que muy útil en los casos donde ciertos casos, de acuerdo con valores de las variables, están sistemáticamente truncados de la muestra. Finalmente, las regresiones lógicas binomial y multinomial, técnicas utilizadas (Porchea, Robbins & Phelps, 2010) para predecir las probabilidades de los diferentes resultados posibles de una variable categórica dependiente, dado un conjunto de variables independientes

Modelamiento de ecuaciones estructurales (SEM). A pesar de estar fundamentada en las técnicas de regresión anteriores, el SEM es la técnica estadística más empleado en los estudios de factores asociados al logro académico (Caso & Hernández, 2010; Barbero, Holgado, Villa & Chacón, 2007; Diniz, Dias & Silva, 2011; Kristjánsson & Sigfúsdóttir, 2009; Rosário, Nuñez, González-Pineda, Almeida, Soares & Rubio, 2005; King et al., 2005), pues aumenta su potencial, permitiendo establecer series de hipótesis simultaneas relacionadas con el efecto de variables latentes, observadas y otro tipo de variables. Dentro de esta técnica, el método de estimación de parámetros por ML es el más usual (Cardoso, Ferreira, Abrantes, Seabra & Costa, 2011; Liem, Martin, Porter & Colmar, 2011; Miñano, Castejón & Gilar, 2012). Sin embargo, restricciones del ML son tamaños de muestra pequeños y falta de normalidad univariante y/o multivariante de las variables incluidas, casos en los que los estimadores no son robusto. Así, es usual estimar parámetros mediante otras técnicas, como la estimación paramétrica Bayesiana (Payandeh, Omid & Reza, 2013) o los no paramétricos por mínimos cuadrados no ponderados (Bollen, 1989; Kline, 2005) y asintóticamente libre de distribución (Browne, 1984; Hawkins, 1982). Otros métodos relacionados con este planteamiento son los SEM multinivel (Mih & Mih, 2013), que permiten la identificación de los efectos directos e indirectos de múltiples variables, incluyendo el anidamiento de los datos, aunque este procedimiento en realidad es poco común.

Análisis jerárquico lineal (multinivel). A pesar de que el análisis de regresión múltiple “clásico” o “de un nivel” ha sido utilizado en los estudios de factores asociados, éste es limitado cuando los datos presentan una estructura anidada, por ejemplo, si hay variables del estudiante y variables de un grupo o conglomerado de estudiantes (p.ej. las características de la escuela). Por lo tanto, de acuerdo con Montero, Villalobos & Valverde (2007), si las variables independientes están medidas en diferentes niveles de agregación, el supuesto de muestreo simple al azar para hacer inferencia estadística, se viola al utilizar el modelo clásico de regresión. En consecuencia, se produce una subestimación de los errores estándar cuando el modelo de regresión ignora el efecto de la conglomeración. Sin embargo, a pesar de que muchos explicativos contemplan variables agregadas, en realidad son pocos los estudios que hacen uso de este tipo de análisis (Cervini, Dari & Quiroz, 2014; Cordero & Manchón, 2014; Mahimuang, 2005; Montero, Villalobos & Valverde, 2007; Strayhorn, 2010; Yil dirim, 2012).

Análisis discriminante. No es propiamente un procedimiento para el establecimiento de modelos explicativos del rendimiento, sino que se trata de una técnica estadística multivariante cuya finalidad es describir (si existen) diferencias significativas entre g grupos de objetos ($g > 1$) sobre los que se observan p variables (variables discriminantes). Sin embargo, la identificación de estas variables clasificatorias se puede asumir como variables explicativas. Por ello, varios estudios reportan esta técnica en la comprensión de los factores asociados al logro académico (Risso, Peralbo, & Barca, 2010; Tejedor, 2003). Al igual que otros procedimientos, tiene variantes que permiten ajustarse a la diversidad de variables que se analizan. Por ejemplo, Santos, Godás y Lorenzo (2013) realizaron un análisis discriminante con el Método de la Distancia de Mahalanobis, el cual proporciona una poderosa herramienta para determinar la similitud entre dos variables aleatorias multidimensionales.

Técnicas econométricas. Aunque la mayor parte de las técnicas aplicadas en el ámbito de la econometría están íntimamente relacionadas con el análisis de la estructura de varianzas-covarianzas de las variables (regresión, SEM o análisis multinivel), muchos estudios han adoptado otros procedimientos de diversa complejidad, con aportaciones interesantes. Por ejemplo, el estudio de Win & Mille (2005), quienes aplicaron una aproximación al enfoque input-output y un modelo de parámetros aleatorios. En otro estudio, De Witte & Kortelainen (2013) utilizaron técnicas bayesianas no paramétricas mediante una adaptación de la función kernel para la estimación de las funciones de densidad y técnicas de bootstrapping para la estimación de los intervalos de confianza de los parámetros.

Minería de datos. Timarán, Calderón y Jiménez (2013) afirmaron que el uso de la minería de datos permite predecir cualquier fenómeno dentro del ámbito educativo. Es un conjunto de procedimientos estadísticos que, de acuerdo con Nghe et al. (2007), permite determinar la agudeza de la discriminación en la identificación de variables predictoras del rendimiento. La minería de datos se refiere al proceso de extraer relación entre variables desde grandes cantidades de datos almacenados en distintos formatos. En términos más simples, se trata de un método de explotación de datos y extracción de información valiosa. Estas técnicas han tenido su mayor uso en el ámbito de las ingenierías, las matemáticas y la economía, empezándose a aplicar en la Educación prácticamente en este nuevo siglo (Alcover et al. 2007; Porcel, Dapozo & López, 2010; Hsieh, 2013; Ma, 2005; Nghe et al., 2007; Schumacher, Olinsky, Quinn & Smith, 2010). Así ha surgido el término Minería de Datos Educativos (MDE), para hacer referencia a la exploración de datos que provienen del contexto educativo; y cuyo objetivo es la búsqueda, análisis y extracción de información para la generación de modelos explicativos que generen información para la mejora de los procesos educativos (Ballesteros, 2014).

Al ser un conjunto de técnicas, son muchos los procedimientos que forman parte de él. Los procedimientos identificados se pueden clasificar en tres grupos:

1. *Árboles de decisión*, conocidos como técnicas de inducción (Luan, 2002) se utilizan para detectar efectos de interacción en subgrupos específicos de casos, en capas o niveles profundos del árbol o en algunas regiones del mismo (Castro y Lizasoain, 2012). Dentro de éstas encontramos

los procedimientos: ID3, C4.5, árboles de clasificación (CART o CR&T, por sus siglas en inglés), algoritmo CHAID, bosques aleatorios, entre otros.

2. *Redes neuronales*, son modelos de regresión adaptados para realizar estimaciones sobre un conjunto de unidades de input/output, en donde cada conexión tiene un peso. Las redes neuronales permiten extraer patrones o tendencias de difícil identificación (Baradwaj & Pal, 2011). En este grupo se encuentran procedimientos como: redes de Kohonen, perceptron multicapa (MLP por sus siglas en inglés), máquinas de aprendizaje extremo (ELM por sus siglas en inglés).
3. *Conglomerados*. Es un método exploratorio, cuyo objetivo es encontrar conjuntos de datos que están asociados. Es útil cuando se tiene poca información sobre la estructura de categorías de un conjunto de datos. Existen dos técnicas que pueden usarse en el algoritmo de conglomerados. El primero, es el K-medias, en el que cada caso es asignado a un solo conglomerado, éste es denominado conglomerado duro, mientras que el análisis de expectación máxima (EM) permite que un caso sea asignado a más de un conglomerado, de ahí su nombre de conglomerado suave (Kiray, Gok y Bozkir, 2015). Según DeFreitas y Bernard (2015), los algoritmos de conglomerados se clasifican de acuerdo con las técnicas utilizadas. Particularmente estos autores definen tres: métodos de partición (por ejemplo, k-medias), métodos jerárquicos y métodos basados en densidad.

Conclusiones

La revisión del estado de la cuestión respecto a las técnicas de análisis para el procesamiento de datos a gran escala, principalmente en el estudio de los factores asociados al logro educativo, se puede establecer como una primera conclusión general, que efectivamente, como señaló Tejedor (2003), los procedimientos aplicados son eclécticos, en muy pocos casos se hace uso de un único procedimiento de análisis de datos, por el contrario, existe una combinación de diferentes técnicas.

No obstante, es claro que los investigadores educativos se decantan por la aplicación de la regresión múltiple, así como el análisis de senderos o SEM, encontrándose, en esta investigación documental, como la técnica de mayor uso en los estudios de identificación de variables asociadas al logro. Probablemente, esto se debe a la naturaleza de las técnicas, pues son procedimientos más amigables, para los cuales existen diferentes softwares y que, casi cualquier persona con nociones básicas de estadística puede poner en práctica. Sin embargo, esta técnica tiene muchas limitaciones, o restricciones para su uso, que en muchos casos son ignorados en los estudios. Por ejemplo, la escala de medición, el fenómeno de la colinealidad, el anidamiento de los datos, por mencionar sólo algunos, lo que implica necesariamente el uso de técnicas más apropiadas, que garanticen la validez de las conclusiones derivadas de la investigación.

Es así que en algunas investigación se ha hecho uso de las técnicas multinivel que consideran la naturaleza anidada de las variables de estudio, situación muy común en la investigación educativa. Sin embargo, su uso es en realidad limitado, si se toma en cuenta que prácticamente cualquier estudio a gran escala debería tener en cuenta esta situación del anidamiento. Por ejemplo, en los estudios sobre factores explicativos de los resultados de PISA, se puede notar que menos de la mitad de las investigaciones hacen uso de las técnicas multinivel. Por lo tanto, es importante, que los investigadores educativos dedicados a la evaluación a gran escala, tengan en cuenta este fenómeno.

Así también se puede hablar del caso de las técnicas de minería de datos, ricas en la versatilidad de las posibilidades de análisis que ofrecen, pero pobres aún en su uso. Al igual que en el caso del análisis multinivel, las técnicas de minería de datos han sido subutilizadas si se tienen en cuenta la naturaleza de los datos en los estudios a gran escala. Como señaló Ballesteros (2014), existe una gran cantidad de volúmenes de datos, que no siempre es posible analizar de forma completa. En tal caso la minería de datos es útil para la identificación de relaciones complejas u ocultas, que no se pueden derivar de forma tan simple con técnicas como el SEM.

Se espera que esta revisión pueda ser útil para los investigadores educativos noveles, así como estudiantes interesados en la investigación de variables explicativas del logro educativo, y de fenómenos similares, de manera que puedan tener en cuenta hacia donde se orientan las técnicas actuales y la relevancia que tiene su uso y explotación apropiadas.

Referencias

- Alcover, R., Benlloch, J., Blesa, P., Calduch, M.A., Celma, M., Ferri, C., Zúñiga, L.R. (2007). Análisis del rendimiento académico en los estudios de informática de la Universidad Politécnica de Valencia aplicando técnicas de minería de datos. Trabajo presentado en las *XIII Jornadas de Enseñanza universitaria de la Informática, Teruel, España*. Recuperado de <http://bioinfo.uib.es/~joemiro/aenui/procjenui/jen2007/alanal.pdf>
- Ballesteros, A., Sánchez-Guzmán, D., & García, R. (2014). Minería de datos educativa: Una herramienta para la investigación de patrones de aprendizaje sobre un contexto educativo. *Latin-American Journal of Physics Education*, 7, 662–668. Recuperado de http://www.lajpe.org/dec13/22-LAJPE_814_bis_Alejandro_Ballesteros.pdf
- Baradwaj, B. K., & Pal, S. (2011). Mining educational data to analyze students' performance. *International Journal of Advanced Computer Science and Applications*, 2(6), 63–69. Recuperado de <https://arxiv.org/ftp/arxiv/papers/1201/1201.3417.pdf>
- Barbero, M.I., Holgado, F.P., Villa, E., y Chacón, S. (2007). Actitudes, hábitos de estudio y rendimiento en Matemáticas: diferencias por género. *Psicothema*, 19(3), 413–421.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley-Interscience.
- Browne, M. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *The British journal of mathematical and statistical psychology*, 37, 62–83.
- Cardoso, A. P., Ferreira, M., Abrantes, J.L, Seabra, C., & Costa. C. (2011). Personal and pedagogical interaction factors as determinants of academic achievement. *Social and Behavioral Sciences* 29. 1596 – 1605. doi: 10.1016/j.sbspro.2011.11.402

- Caso, J., y Hernández, L. (2010). Modelo explicativo del bajo rendimiento escolar: Un estudio con adolescentes mexicanos. *Revista Iberoamericana de Evaluación Educativa*, 3(2), 146-159
- Castro, M., & Lizasoain Hernández, L. (2012). Las técnicas de modelización estadística en la investigación educativa: minería de datos, modelos de ecuaciones estructurales y modelos jerárquicos lineales. *Revista española de pedagogía*, 70(251), 131-148.
- Cervini, R., Dari, N., & Quiroz, S. (2014). Estructura familiar y rendimiento académico en países de América Latina. *Revista Mexicana de Investigación Educativa*, 19(61), 569-597.
- Cordero, J.M., & Manchón, C. (2014). Factores explicativos del rendimiento en educación primaria: un análisis a partir de TIMSS 2011. *Estudios sobre Educación*, 27, 9-35.
- DeFreitas, K., & Bernard, M. (2015). Comparative performance analysis of clustering techniques in educational data mining. *IADIS International Journal on Computer Science and Information Systems*, 10(2), 65-78. Recuperado de <http://www.iadisportal.org/ijcsis/papers/2015180205.pdf>
- De Witte, K., & Kortelainen, M. (2013). What explains the performance of students in a heterogeneous environment? Conditional efficiency estimation with continuous and discrete environmental variables. *Applied Economics*, 45(17), 2401-2412. <http://doi.org/10.1080/00036846.2012.665602>
- Diniz, A., Dias, M., y Silva, L. (2011). Cognitive abilities, sociocultural background and academic achievement. *Psicothema*, 23(4), 695-700.
- Guimaraes, J., & Sampaio, B. (2013). Family background and student's achievement on a university entrance exam in Brazil. *Education Economics*, 21(1), 38-59.
- Haile, G.A., & Nguyen, A.N. (2008). Determinants of academic attainment in the United States: A quantile regression analysis of test scores. *Education Economics*, 16(1), 29-57.
- Hawkins, D. (1982). *Topics in applied multivariate analysis*. Cambridge [etc.]: Cambridge University Press.
- Hsieh, M. (2013). The data mining form education database examine the factors impacting the school performance in the United States. *International Journal of Intelligent Technologies and Applied Statistics*, 6(2), 135-143. doi: 10.6148/IJITAS.2013.0602.03
- Kaighobad, M., & Allen, M.T. (2008). Investigating academic success factors for undergraduate business students. *Decision Sciences Journal of Innovative Education*, 6, 427-436.
- King, G., McDougall, J., DeWitt, D., Hong, S., Miller, L., Offord, D., Meyer, K., & LaPorta, J. (2005). Pathways to children's academic performance and prosocial behaviour: Roles of physical health status, environmental, family, and child factors. *International Journal of Disability*, 52(4), 313-344.
- Kline, R. (2005). *Principles and practice of structural equation modeling*. New York: Guilford Press.
- Kristjánsson, A.L., & Sigfúsdóttir, I. D. (2009). The role of parental support, parental monitoring, and time spent with parents in adolescent academic achievement in Iceland: A structural model of gender differences. *Scandinavian Journal of Educational Research*, 53(5), 481-496.
- Kiray, S. A., Gok, B., & Bozkir, A. S. (2015). Identifying the factors affecting science and mathematics achievement using data mining methods. *Journal of Education in Science, Environment and Health (JESEH)*, 1, 28-48. Recuperado de <https://www.jeseh.net/index.php/jeseh/article/view/4>
- Lee, J., & Shute, V.J. (2010). Personal and social-contextual factors in K-12 academic performance: An integrative perspective on student learning. *Educational Psychologist*, 45(3), 185-202. doi: 0.1080/00461520.2010.493471
- Liem, G.A., Martin, A.J., Porter, A.L., & Colmar, S. (2011). *Asian Journal of Social Psychology*, 15, 1-13. doi: 10.1111/j.1467-839X.2011.01351.x
- Luan, J. (2002). Data mining and its applications in higher education. *New Directions for Institutional Research*, 113, 17-36.
- Ma, X. (2005). Growth in Mathematics achievement: Analysis with classification and regression trees. *The Journal of Educational Research*, 99(2), 78-86.

- Mahimuang, S. (2005). Factors influencing academic achievement and improvement: A value-added approach. *Educational Research for Policy and Practice*, 4, 13-26. doi: 10.1007/s10671-005-0677-1
- Mih, V. & Mih, C. (2013). Perceived autonomy-supportive teaching, academic self-perceptions and engagement in learning: Toward a process model of academic achievement. *Cognition, Brain & Behavior*, 17(4), 289-313.
- Miñano, P., Castejón, J.L., & Gilar, R. (2012). An explanatory model of academic achievement based on aptitudes, goal orientation, self concept and learning strategies. *The Spanish Journal of Psychology*, 15(1), 48-60.
- Montero, E., Villalobos, J., & Valverde, A. (2007). Factores institucionales, pedagógicos, psicosociales y sociodemográficos asociados al rendimiento académico en la Universidad de Costa Rica: Un análisis multinivel. *Revista Electrónica de Investigación y Evaluación Educativa*, 13(2), 215-234.
- Nath, S. R. (2012). Factors influencing primary students' learning achievement in Bangladesh. *Research in Education*, 88, 50-63.
- Nghe, N.T., Janecek, P., & Haddawy, P. (2007) *37th ASEE/IEEE Frontiers in Education Conference*. Milwaukee, WI, octubre de 2007. A comparative analysis of techniques for predicting academic performance. Recuperado de http://www.researchgate.net/publication/224299912_A_comparative_analysis_of_techniques_for_predicting_academic_performance
- Payandeh, A.T., Omidi, M., & Reza, M. (2013). Factors contributing to academic achievement: a Bayesian structure equation modelling study. *International Journal of Mathematical Education in Science and Technology*, 44(4), 490-500.
- Porcel, E., Dapozo, G. y López, M. (2010). Predicción del rendimiento académico de alumnos de primer año de la FACENA (UNNE) en función de su caracterización socioeducativa. *Revista Electrónica de Investigación Educativa*, 12(2). Recuperado de <http://redie.uabc.mx/voll2no2/contenido-porcel>
- Porchea, S.F., Allen, J., Robbins, S., & Phelps, R.P. (2010). Predictors of long-term enrollment and degree outcomes for community college students: Integrating academic, psychosocial, social-demographic, and situational factors. *The Journal of Higher Education*, 81(6), 680-708.
- Risso, A., Peralbo, M., y Barca, A. (2010). Cambios en las variables predictoras del rendimiento escolar
- Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers and Education*, 51(1), 368-384. <http://doi.org/10.1016/j.compedu.2007.05.016>
- Rosário, P., Nuñez, J.C., González-Pineda, J. A., Almeida, L., Soares, S., y Rubio, M. (2005). El aprendizaje escolar examinado desde la perspectiva del «Modelo 3P» de J. Biggs. *Psicothema*, 17(1), 20-30.
- Santos, A., Godás, A., y Lorenzo, M. (2013). Rendimiento académico y diversidad cultural: el eje lingüístico. *Revista Española de Pedagogía*, 71(256), 461-478.
- Schumacher, P., Olinsky, A., Quinn, J., & Smith, R. (2010). A comparison of logistic regression, neural networks, and classification trees predicting success of actuarial students. *Journal of Education for Business*, 85, 258-263.
- Strayhorn, T. L. (2010). The role of schools, families, and psychological variables on math achievement of black high school students. *The High School Journal*, 93(4), 177-194.
- Stull, J. (2013). Family socioeconomic status, parent expectations, and a child's achievement. *Research in Education*, 90, 53-67.
- Tejedor, F.J. (2003). Poder explicativo de algunos determinantes del rendimiento en los estudios universitarios. *Revista Española de Pedagogía*, 224, 5-32.
- Timarán, R., Calderón, A., y Jiménez, J. (2013). Descubrimiento de perfiles de deserción estudiantil con técnicas de minería de datos. *Revista Vínculos*, 10(1), 373-383, doi: <https://doi.org/10.14483/2322939X.4687>
- Win, R., & Miller, P.W. (2005). The effects of individual and school factors on university students' academic performance. *The Australian Economic Review*, 38 (1), 1-18