

Análisis de los factores más significativos que afectan el rendimiento académico de los estudiantes de la UNASAM en el curso de estadística, mediante el uso de la regresión logística- 2013

Analysis of the most significant factors affecting the academic performance of the students of the UNASAM in the course of statistics, using regression logistic – 2013

Fernando Arce Z.¹, Edwin Asnate S.¹

RESUMEN

Se basa en la técnica estadística de la regresión logística binaria para determinar los factores más significativos que afectan el rendimiento académico de los estudiantes de la UNASAM, en el curso de estadística, por ser este uno de los cursos más preponderantes en la actualidad para la formación de los futuros profesionales en todos los campos del saber humano; los resultados de la investigación dieron como conclusión que la técnica estadística mencionada es la más adecuada para determinar la probabilidad de éxito; es decir, la de aprobar el curso de estadística en base a las factores que afectan el rendimiento académico de los estudiantes y también se determinó que las variables más significativas son: la edad, el hecho de trabajar y estudiar, el nivel educativo del padre, el curso previo de matemática, la asistencia a clases, la visita a la biblioteca, la motivación en sus carrera y la influencia del docente motivador. Variables que deben ser consideradas por los docentes y directores de escuela para que el estudiante tenga un buen rendimiento en el curso de estadística.

Palabras claves: Regresión logística; rendimiento académico.

ABSTRACT

It is based on the statistical technique of binary logistic regression to determine the most significant factors affecting the academic performance of students UNASAM in the statistics course, as this is one of the most prevalent courses today for the training of future professionals in all fields of human knowledge; the results of the investigation led to conclude that the statistical technique mentioned is the right to determine the likelihood of success; ie to approve the statistics course based on the factors affecting the academic performance of students and also determined that the most significant variables are: the age, the fact work and study, the educational level of the father, the previous course of mathematics, attendance, visiting the library, motivation in their career and influence of motivating teachers. Variables that should be considered by teachers and principals for the student will perform well in the statistics course.

Key words: Logistic regression; academic achievement.

¹Facultad de Ciencias. Universidad Nacional Santiago Antúnez de Mayolo, Huaraz, Ancash - Perú

INTRODUCCIÓN

La regresión logística binaria es una técnica estadística de análisis multivariante, de uso tanto explicativo como predictivo. Resulta útil su empleo cuando se tiene una variable dependiente categórica en especial dicotómica, un atributo cuya ausencia o presencia puede ser codificado con los valores cero y uno, respectivamente (aprobó o desaprobó) y un conjunto de k variables predictoras o independientes, que pueden ser cuantitativas (que se denominan covariables) o categóricas. En este último caso, si las variables tienen más de dos categorías se requiere que sean transformadas en variables ficticias o simuladas (“*dummy*”). La regresión logística puede ser aplicada a un rango más amplio de situaciones de investigación que el análisis discriminante, o que la regresión lineal, por cuanto no exige que las covariables tengan distribución normal, pero tampoco lo impide.

Por otro lado, el rendimiento académico es considerado como la capacidad del alumno para responder a los estímulos educativos, en este sentido, el rendimiento académico está vinculado a una serie de factores que inciden en dicho rendimiento, desde la dificultad propia de algunas asignaturas como la Estadística, la capacidad intelectual y actitudinal del alumno, hasta el ambiente social que lo rodea.

Uno de los problemas más álgidos que enfrentan los estudiantes en la actualidad, es precisamente el bajo rendimiento académico que se aprecia en las diferentes universidades de nuestro país, entonces buscar alternativas de solución frente a este problema se ha convertido hoy en día, en primera prioridad de los responsables de este proceso.

Resolver los problemas educativos de un país, como argumenta Salvador Reyes R.[2004, 4] no es una tarea fácil, este tipo de problemas vienen ligados a estructuras más complejas y diversas que tienen sus apéndices radicales en

las macro variables de la sociedad, la política, la economía y la voluntad de los gobiernos.

Por otra parte, la Estadística como ciencia, forma parte del acervo cultural de todo ciudadano educado de hoy en día, pues constituye una herramienta metodológica indispensable para todo profesional, así como para todo aquel que quiera hacer investigación científica en cualquier campo del saber humano. El alumno cuando lleva el curso de estadística suele ser un principiante en la universidad y no está familiarizado ni con la investigación ni con el ejercicio de su profesión lo que dificulta su aprendizaje. Para comprender e interpretar los datos estadísticos se tiene que asimilar muchos conceptos de la ciencia estadística, con diversos niveles de abstracción. Es conocido que la mayoría de los estudiantes universitarios tienen serias dificultades para comprender y hacer conexiones entre los diferentes conceptos, generando con ello frustración, aburrimiento, disgusto y hasta temor por el curso; motivo por el cual, el presente trabajo tiene los siguientes objetivos:

MATERIALES Y MÉTODOS

El presente trabajo de investigación tiene las siguientes características según la clasificación hecha por Roberto Hernández Sampieri (2010), el diseño es del tipo no experimental; transeccional, pues se usó los datos de una encuesta hecha en un determinado momento, además es

correlacional por cuanto se relacionó la variable dependiente (rendimiento académico) con las variables independientes o factores que influyen en el rendimiento académico. El enfoque usado es el cuantitativo y por la finalidad es aplicada.

La población investigada, fueron todos los estudiantes matriculados en el Semestre Académico 2013-I y que habían llevado el primer curso de estadística; la unidad de análisis fue un estudiante con las características dadas. La muestra fue del tipo probabilístico, para obtener el tamaño de muestra se usó la fórmula del muestreo aleatorio simple para proporciones, pero distribuyendo las

submuestras en forma proporcional a la cantidad de alumnos de cada escuela de formación profesional, el tamaño obtenido fue de 213 estudiantes, a los cuales se les aplicó una encuesta para obtener información referente a las 26 variables escogidas que supuestamente afectaban su rendimiento académico en el curso de estadística. El instrumento de recolección de datos fue la hoja de encuesta, el cual fue validado usando el Alfa de Cronbach, previa corrección del análisis de los datos de la encuesta piloto. El procesamiento de los datos se hizo con el software estadístico SPSS versión N° 22 y el STATA.

RESULTADOS

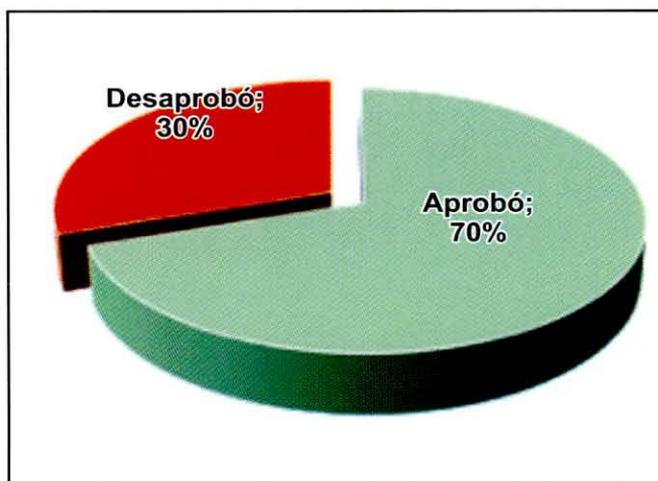


Figura 1. Porcentaje de alumnos aprobados y desaprobados en el curso de estadística.

Las variables consideradas en el estudio fueron codificadas, la mayoría de ellas fueron dicotomizadas para facilitar el trabajo, en algunos casos se usó las variables dummy, por su parte los resultados de la encuesta nos indican que: En la variable independiente: Nota final del primer curso de estadística en primera matrícula (R1); el 70% de los estudiantes aprobaron el primer curso de estadística en primera matrícula; En las variables dependientes (Factores): Edad (AP2), el 56,8% son menores de 20 años, de ellos el 73,8% aprobó el curso de estadística. Sexo (AP3), el 50,2% son varones y de ellos el 55% aprobó el curso de estadística. Trabajo (AP4), el 31,9% trabaja y estudia y de ellos sólo el 32,2%

aprobó el curso de estadística. Ingreso económico (AP5), el 56,8% tienen ingresos económicos suficientes, de ellos el 75,8% aprobó el curso de estadística. Nivel educativo del padre (AP6), el 54,5% tienen padres con un nivel educativo superior, de ellos el 72,5% aprobó el curso de estadística. Nivel educativo de la madre (Ap7), el 44,6% tienen madres con un nivel educativo superior y de ellos el 59,7% aprobó el curso de estadística. Tipo de colegio (AS8), el 70% de los estudiantes proceden de colegios estatales, de ellos el 59,7% aprobó el curso de estadística. Ubicación del colegio (AS9), el 74,2% de los estudiantes proceden de colegios de zonas urbanas y de ellos el 88,6% aprobó el curso de estadística.

Lugar de procedencia del alumno (As10), el 71,8% de los estudiantes proceden de zonas urbanas y de ellos el 85,9% aprobó el curso de estadística. Habitación (AS11), el 69,5% de los estudiantes viven con sus padres, de ellos el 84,6% aprobó el curso de estadística. Curso previo de matemática (AA12), el 70,4% de los estudiantes aprobó en primera matrícula el curso previo de matemática y de ellos el 94% aprobó el curso de estadística. Ingreso a la Universidad (AA13), el 63,4% de los estudiantes ingresó a la UNASAM en el primer intento y de ellos el 78,5% aprobó el curso de estadística. Conocimiento de informática (AA14), el 77% de los estudiantes tienen conocimiento de informática y de ellos el 94,6% aprobó el curso de estadística. Créditos aprobados (AA15), el 71,4% de los estudiantes aprobó más de 20 créditos y de ellos el 90,6% aprobó el curso de estadística. Horas de estudio (AA16), el 62% de los estudiantes estudian 5 o más horas al día y de ellos el 83,2% aprobó el curso de estadística. Asistencia a clases (AA17), el 73,2% de los estudiantes asiste a clases en forma normal y de ellos el 96% aprobó el curso de estadística. Hábitos de estudio (AA18), el 59,5% de los estudiantes estudian en grupo y de ellos el 78,5% aprobó el curso de estadística. Visita a la biblioteca (AA19), el 63,8% de los estudiantes visita a la biblioteca normalmente y de ellos el 83,2% aprobó el curso de estadística. Motivación en su carrera (AA20), el 77,5% de los estudiantes tiene motivación para seguir su carrera y de ellos el 96% aprobó el curso de estadística. Docente capacitado (DD21), el 79,8% de los estudiantes consideran que su profesor está capacitado en el área y de ellos el 96,6% aprobó el curso de estadística. Docente motivador (DD22), el 67,1% de los estudiantes manifiestan que su docente de estadística fue motivador y de ellos el 86,6% aprobó el curso de estadística. Metodología del docente (DD23), el 57,8% de los estudiantes manifiestan que la metodología del docente de estadística es regular y de ellos el 51,7% aprobó el curso de estadística. Evaluación del docente (DD24), 58,7% de los estudiantes manifiestan que la forma de evaluación del docente es regular y de ellos el

53,7% aprobó el curso de estadística. Asistencia a clases del docente (DD25), el 85,9% de los estudiantes indican que el docente de estadística asistió normalmente a clases y de ellos el 95,3% aprobó el curso de estadística. Aspecto logístico del campus universitario (EL26), el 69,5% de los estudiantes manifiestan que el campus universitario es regular para el estudio y de ellos el 69,8% aprobó el curso de estadística. Ambiente de estudio (EL27), el 58,2% de los estudiantes manifiestan que el ambiente universitario es mala en la parte Política para el estudio y de ellos el 49,7% aprobó el curso de estadística.

Usando el software STATA, se encontró que algunas variables tenían problemas de colinealidad, por lo que se eliminó las variables dummy DD23 y EL27, sin estas variables el software nos proporcionó el modelo inicial cuyos valores se presentan en la Tabla N° 1, en dicho cuadro se observa que la mayoría de las variables no son significativas, pues el P valor (nivel de significancia observado) es superior al nivel de significancia teórico planteado del 5%, lo que implica que dichas variables no se incluirán en el modelo. A continuación se realizó la prueba de la bondad de ajuste mediante la técnica de Hosmer-Lemeshow, los resultados de la prueba (ver Tabla N° 2) nos indican que el modelo de regresión logística se ajusta muy bien al conjunto de variables con una seguridad del 99%. Con este aval se usó el método del Stepwise para seleccionar las variables más significativas para ser incluidas en el modelo.

Con las ocho variables escogidas, nuevamente se usó la prueba de Hosmer-Lemeshow para determinar si el modelo reducido es adecuado o no; los resultados de la prueba nos indicaron que el modelo es adecuado con un 95% de seguridad. En

seguida se verificó si había multicolinealidad entre las variables escogidas, para ello se analizó el VIF (Factor de Inflación de la Varianza) para cada variable independiente, cuyos resultados se muestran en la siguiente tabla:

Tabla 2.

Logistic model for R1, goodness-of-fit test

number of observations = 213
 number of covariate patterns = 164
 Pearson chi2(137) = 81 .65
 Prob > chi2 = 1.0000

:

p = 0.8857 >= 0.0500 removing AS8 p =
 0.7705 >= 0.0500 removing AP7 p =
 0.7607 >= 0.0500 removing AA16 p =
 0.7784 >= 0.0500 removing AA18 p =
 0.8105 >= 0.0500 removing AP3 p =
 0.7218 >= 0.0500 removing AA14 p =
 0.4982 >= 0.0500 removing AS9 p =
 0.4214 >= 0.0500 removing DD25 p =
 0.5029 >= 0.0500 removing AA15 p =
 0.2282 >= 0.0500 removing AS11 p =
 0.3277 >= 0.0500 removing AS10 p =
 0.1432 >= 0.0500 removing AP5 p =
 0.0947 >= 0.0500 removing DD21 p =
 0.1160 >= 0.0500 removing AA13

Logistic regression

Number of obs = 213
 LR chi2(8) = 202.80
 Prob > chi2 = 0.0000
 Pseudo R2 = 0.7788

Log likelihood = -28.798876

R1	Coef.	Std. Err.	z	P> z	95% Conf. Interval]	
Ap2	2.272407	.8250124	2.75	0.006	.655412	3.889401
AA19	-2.648599	1.075572	-2.46	0.014	-4.756681	-.5405159
AP4	-2.491827	.9062509	-2.75	0.006	-4.268046	-.715608
AA12	3.95094	.8603539	4.59	0.000	2.264678	5.637203
AP6	2.345289	.865023	2.71	0.007	.6498749	4.040703
DD22	2.011226	.8662537	2.32	0.020	.3134002	3.709052
AA20	3.052439	1.072803	2.85	0.004	.9497829	5.155095
AA17	2.057933	.9794418	2.10	0.036	.1382627	3.977604
_cons	-5.158779	1.146974	-4.50	0.000	-7.406806	-2.910752

Tabla 03. Resultado del software STATA

Variable	VIF	1/VIF
!A17	3.36	0.297439
AA19	3.10	0.322120
AA12	2.38	0.420685
DD22	2.36	0.424223
AA20	2.32	0.431758
AP6	1.72	0.582603
AP2	1.49	0.670863
AP4	1.05	0.949409
Mean VIF	2.22	

Como se puede apreciar, todos los VIF tienen valores por debajo de 10, con lo cual se concluye que no existen problemas de multicolinealidad. Luego se analizó la cantidad de instancias clasificadas correctamente, los resultados nos indican que el modelo tiene una alta especificidad (92,19%) y una alta sensibilidad (95,97%), asimismo se concluye que el 94,84% de los registros han sido clasificados correctamente.

Por otro lado, se analizó la curva ROC (Receiver operating characteristic), donde se observa que el área bajo la curva es de 0,9849 (ver Fig. 2) que es muy cercano a la unidad, por lo que podemos concluir que el modelo hallado por el método del stepwise es excelente para el conjunto de datos en estudio.

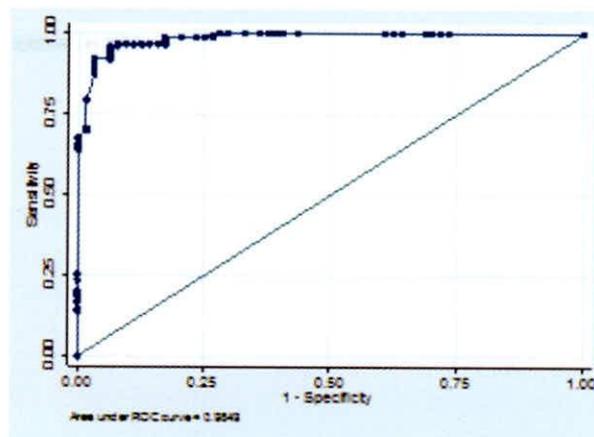


Figura 02: Curva ROC

Asimismo se analizó, la gráfica de dispersión (Ver Fig. 3), con la finalidad de determinar si existían puntos influyentes en el modelo, usando por su puntos alejados de los demás puntos, por lo que se concluye que no existen puntos influyentes que puedan distorsionar los resultados.

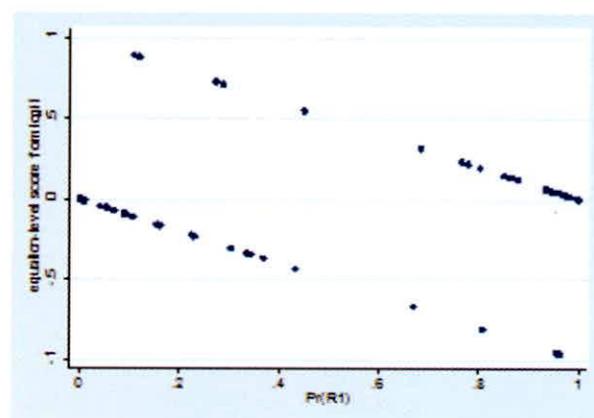


Figura 3:
Gráfica de dispersión de residuos.

Finalmente el modelo obtenido es el siguiente:

$$P(x) = \frac{1}{1 + e^{-(-5.16+2.27AP2-2.49AP4+2.35AP6+3.95AA12+2.06AA17-2.65AA19+3.05AA20+2.01DD22)}}$$

DISCUSIÓN

Después de procesar los datos recolectados mediante el cuestionario de la encuesta, utilizando los software estadístico SPSS versión 22 y el STATA, se observan variables confundentes ni de interacción, tal como se suponía que podrían existir. Asimismo al hacer la prueba de Hosmer Lemeshow para determinar si el modelo de regresión logística era el adecuado para el presente caso, se encontró que con toda seguridad era el indicado, pues el software arrojó una probabilidad de prácticamente uno, tal como se observa en la Tabla 1 en esta tabla también se observa que la gran mayoría de las variables usadas no son significativas para nuestro modelo de regresión logística, a pesar de ser variables predominantes en estos casos, tal como esperábamos apoyados en las conclusiones de los investigadores mencionados en la parte de antecedentes, motivo por el cual se eliminó 14 variables, quedando finalmente sólo 8 variables independientes los que deben intervenir en el modelo, y son los que se muestran en la Tabla 2. Por otro lado, analizando los antecedentes mencionados, se puede notar que existen muchas coincidencias entre las conclusiones dadas y la del presente trabajo, así por ejemplo: El Dr. Juan Manuel Izar Landeta (2011) de la Universidad Autónoma de San Luis Potosí- México, indicó que entre los factores que afectan el rendimiento académico del educando eran la motivación del alumno y el nivel educativo del padre, que coincide con dos de nuestras ocho variables significativas. Por su parte, Chue, Barreno y Millones [2007, 79-80] profesores de la Universidad de Lima-Perú, coinciden en señalar que la mejor técnica para analizar el rendimiento académico de los alumnos es la Regresión Logística. Asimismo, Tonconi Quispe (2010), docente de la Universidad Nacional del Altiplano-Puno-

Perú, concuerda que entre los factores que afectan el rendimiento académico de los alumnos están la asistencia a clases y el nivel educativo de los jefes del hogar. Ponsot, Sinha, Varela y Valera (2009) de la Universidad de Los Andes de Venezuela concluyeron en primer término, que los modelos de regresión logística demuestran ser una herramienta muy poderosa para indagar sobre la explicación de variables categóricas de respuesta, como es el presente caso. En tanto que, Tejedor y García-Valcárcel (2007) de la Universidad de Salamanca- España concluyeron que hay un conjunto de variables que afectan el rendimiento académico, de los cuales coinciden con el presente trabajo, las variables: edad, cursos previos, motivación, estudios de los padres, capacitación docente, estrategias de aprendizaje (asistencia a la biblioteca). Ibarra y Michalus (2005) de la Facultad de Ingeniería de la Universidad Nacional de Misiones- UNAM –México, también concluyen que entre las variables que afectan el rendimiento académico están: el nivel educativo de los padres. Bartual y Poblet, (2009) profesores de la Universidad de Barcelona, indican que entre las variables que afectan el rendimiento académico de los alumnos está la asistencia a clases. Como se puede apreciar hay una buena gamma de variables que coinciden con las variables incluidas en nuestro modelo.

CONCLUSIONES

Las conclusiones a las que se llega en el presente trabajo de investigación, como respuesta a los objetivos e hipótesis, son los siguientes:

1. Los factores más influyentes que afectan el rendimiento académico de los estudiantes de la UNASAM en el curso de estadística son:
 - La edad.
 - El trabajo con remuneración.
 - El nivel educativo del padre.

- El curso previo de matemática. La asistencia a clases.
- La visita a la biblioteca.
- La motivación en su carrera.
- El docente motivador.

2. Análisis de los factores más significativos que afectan el rendimiento académico de los estudiantes de la UNASAM en el curso de Estadística.

De los 26 factores que se escogieron y que se consideraron que podrían afectar el rendimiento académico del alumno en el curso de estadística, por ser estos preponderantes y que en otros casos sí afectaron el rendimiento académico de un estudiante, la mayoría fue descartada por no ser significativos para el presente estudio, quedando finalmente los mencionados en la primera conclusión. El análisis completo de los factores se realizó en el documento original.

REFERENCIAS BIBLIOGRÁFICAS

Barriga, y Hernández. (2000) Estrategias Docentes para un Aprendizaje Significativo: Una Interpretación Constructivista. México: Edit. McGraw Hill.

Cea D'Ancona, Ángeles. (2002) Análisis Multivariable. Teoría y Práctica en la Investigación Social. España: Edit. Síntesis. S.A.

Chue G. Jorge, Barreno V. Emma y Millones R. Rosa. (2007) Sistema para el Análisis Estadístico de Técnicas Multivariadas del Rendimiento Académico de los Estudiantes de una Institución de Enseñanza Superior. Universidad de Lima.

Izar Landeta, Juan M. (2011) Factores que afectan el desempeño académico de los estudiantes de nivel superior en Rio verde, San Luis Potosí, México.

Johnson, Dallas E. (2000) Métodos Multivariados Aplicados al Análisis de Datos. México: Edit. Internacional Thomson Editores S.A.

Evy, Jean P., y Varela, Jesús. (2003) Análisis Multivariante para las Ciencias Sociales. España: Edit. Pearson Prentice Hall.

Pérez López, César. (2004) Técnicas de Análisis Multivariante de Datos. España: Edit. Pearson Prentice Hall.

Tonconi Q, Juan. (2010) Factores que Influyen en el Rendimiento Académico y la Deserción de los Estudiantes de la Facultad de Ingeniería Económica de la UNA-Puno, Periodo 2009

Correspondencia

Fernando Arce Zuñiga
fearazu@yahoo.es