

Publicación Cuatrimestral. Vol. 7, No 3, Septiembre/Diciembre, 2022, Ecuador (p. 1-14). Edición continua
<https://revistas.utm.edu.ec/index.php/Basedelaciencia/index>
revista.bdlaciencia@utm.edu.ec
Universidad Técnica de Manabí
DOI: <https://doi.org/10.33936/revbasdelaciencia.v7i3.5197>

MODELO DE REGRESIÓN LOGÍSTICA PARA LA ESTIMACIÓN DE LA DESERCIÓN ESCOLAR DEL POSGRADO EN LA UNIVERSIDAD TÉCNICA DE MANABÍ, ECUADOR

Julio César Solís Ventura¹, Santiago Quiroz Fernández¹, Osvaldo Fosado Téllez^{1*}

¹Estudiante de la maestría en Estadística, mención Estadística Aplicada. Instituto de Posgrado. Universidad Técnica de Manabí. Portoviejo. Ecuador. E-mail: juliosolis1302@gmail.com

²Instituto de Posgrado. Universidad Técnica de Manabí. Portoviejo. Ecuador. E-mail: santvqf@gmail.com

³Instituto de Posgrado. Universidad Técnica de Manabí. Portoviejo. Ecuador. E-mail: osvaldo.fosado@gmail.com

*Autor para la correspondencia: osvaldo.fosado@gmail.com

Recibido: 27-4-2022 / Aceptado: 01-07-2022 / Publicación: 01-09-2022

Editor Académico: Wilmer Orlando López González

RESUMEN

En el presente trabajo se ofrece un modelo de regresión logística binaria para la predicción de la deserción estudiantil de posgrado en la Universidad Técnica de Manabí, Portoviejo, Ecuador. Se obtuvieron registros de variables predictoras disponibles en el Sistema de Gestión Académica, detectándose, a través de las pruebas de Wald, como altamente significativas el estado civil, situación laboral, edad y salario. El modelo fue probado desde el punto de vista analítico a través de las pruebas Chi cuadrado, R^2 de Nagelkerke y Hosmer-Lemeshow alcanzando resultados significativos en todas ellas. Posteriormente se realizó una clasificación de los resultados de una muestra piloto con el modelo estimado, obteniéndose el 98,6% de respuestas correctas. Se concluyó evaluando el modelo con una muestra de validación (30% de los datos) alcanzándose el 99,06% de aciertos.

Palabras clave: Modelo logit, Contrastes de significancia, Deserción escolar.

LOGISTIC REGRESSION MODEL FOR THE ESTIMATION OF SCHOOL DROPOUT OF POSTGRADUATE IN THE TECHNICAL UNIVERSITY OF MANABÍ, ECUADOR

ABSTRACT

In the present work, a binary logistic regression model to predict postgraduate student dropout at the Technical University of Manabí, Portoviejo, Ecuador is offered. Records of available predicting variables in the Academic Management System were obtained, detecting marital status, employment status, age and salary as highly significant through the Wald test. The model was tested from an analytical point of view through Chi square tests, Nagelkerke's R^2 , and Hosmer-Lemeshow,

reaching significant results in all of them. Subsequently, a classification of the results of a pilot sample was carried out with the estimated model, obtaining 98.6% of correct answers. Finally, the model was evaluated with a validation sample (30% of the data) reaching 99.06% correct answers.

Keywords: Logit model, Significance tests, School desertion.

MODELO DE REGRESSÃO LOGÍSTICA PARA ESTIMATIVA DE EVASÃO ESCOLAR NA PÓS-GRADUAÇÃO DA UNIVERSIDADE TÉCNICA DE MANABÍ, EQUADOR

RESUMO

No presente trabalho, apresenta-se um modelo de regressão logística binária para a previsão da evasão estudantil na pós-graduação da Universidad Técnica de Manabí, Portoviejo, Equador. Obtiveram-se registros das variáveis preditoras disponíveis no Sistema de Gestão Acadêmica, detectando como altamente significativas por meio dos testes de Wald, o estado civil, situação empregatícia, idade e salário. O modelo foi testado do ponto de vista analítico através dos testes Qui-quadrado, R^2 de Nagelkerke, Hosmer-Lemeshow, alcançando resultados significativos em todos eles. Posteriormente, foi realizada uma classificação dos resultados de uma amostra piloto com o modelo estimado, obtendo-se 98,6% de respostas corretas. Concluiu-se avaliando o modelo com uma amostra de validação (30% dos dados) atingindo 99,06% de acertos.

Palavras chave: Modelo Logit, Testes de significância, Evasão escolar.

Citación sugerida: Solís, J., Quiroz, S., Fosado, O. (2022). Modelo de regresión logística para la estimación de la deserción escolar del posgrado en la Universidad Técnica de Manabí, Ecuador. Revista Bases de la Ciencia, 7(3), 1-14. DOI: <https://doi.org/10.33936/revbasdelaciencia.v7i3.5197>



1. INTRODUCCIÓN

La educación es considerada como uno de los sectores cruciales en la economía de los países, el sistema educativo que estos adopten debe estar diseñado para cubrir las necesidades de desarrollo y aspiraciones de la nación. En otras palabras, la competitividad se logra con el desarrollo educacional de la sociedad, principalmente la educación superior y el posgrado, como el grado de especialización de mayor impacto. La educación superior pública es vista como un sistema planeado, organizado, estructurado y dirigido por el Estado, reconocido como un bien público, a través del cual se forman las personas para una vida social y laboral. Dicha tarea debe responder a indicadores de calidad que aseguren, en la medida de lo posible, los objetivos que se plantean para el desarrollo de la nación y que sean reconocidos por la sociedad. (Hernández, Pérez, & González, 2014).

La formación de posgrado, dentro de la formación académica de las universidades, tiene una alta responsabilidad social, garantizando el perfeccionamiento y actualización de los conocimientos de los profesionales a través de especialidades, programas de maestrías y doctorados.

En Ecuador, los posgrados empezaron a demandarse de forma masiva a finales de la década de los 90's, actualmente se han convertido en requisitos para acceder a cargos en diferentes instituciones, inclusive para desempeñar cargos académicos en las instituciones de Educación Superior es un requisito indispensable poseer título académico de cuarto nivel (Molina & González, 2019, págs. 9-24).

Dentro de estos estudios una problemática presente en todo el mundo es la deserción escolar, esta se considera como un sinónimo de abandono, que se define como estudiantes que suspenden, repiten, cambian de carrera o programas, o abandonan antes de obtener el título. Según Sinchi y Gómez (2018) la deserción contiene algunas modalidades que diferencian su comprensión y sus diversas motivaciones, puede ser retiro temporal, completo o en el caso de las universidades particulares o programas costeados, no poseer capacidad para cubrir los costos de colegiatura.

A pesar de toda la investigación desarrolladas desde distintas disciplinas, el fenómeno de la deserción sigue ocurriendo y pocas son las herramientas que se han generado para mitigar sus efectos negativos. Esto genera una oportunidad para que nuevas disciplinas, principalmente aplicadas, respondan al desafío del mejoramiento de la gestión de la deserción (Vásquez, 2016).

En América Latina, existe un índice de deserción universitaria general cercano al 57%, mientras que en algunos de sus países alcanzan las alarmantes cifras que se citan a continuación: Guatemala 82%, Bolivia 73,3%, Uruguay el 72%, Brasil el 69%, Costa Rica el 54%, Chile el 53,7%, México el 53%, Panamá el 42%, Argentina el 40% y en Honduras el 38%. El país donde la tasa de deserción es menor

es en Cuba, con un 25%, no considerándose como valor despreciable para su análisis (Fernández, Solís, Hernández, & Moreira, 2019, págs. 73-97).

De acuerdo con Madrid (2017), las estadísticas de deserción y graduación presentadas en el Sistema para la Prevención de la Deserción de la Educación Superior (SPADIES) en el año 2015 en Colombia fue del 46.1%, es decir que de cada 100 estudiantes que son admitidos a una IES aproximadamente la mitad no culminaban sus estudios universitarios.

El fracaso de estudiantes en programas de estudios de posgrado es un problema social serio y se ha vuelto un tema importante para los profesionales de la educación determinar el por qué existen personas que no culminan estos estudios. Por lo tanto, se puede considerar que el problema de la deserción estudiantil afecta a todas las instituciones de educación superior y en los estudios recientes se nota que la problemática va en aumento (Lopera, 2008). Algunos intentos se han hecho para frenar esta problemática como el otorgamiento de becas y créditos educativos, pero estas alternativas son insuficientes en correspondencia con las necesidades” (Sinchi & Gómez, 2018).

Mifflin (2017), en su investigación menciona que las cifras de las tasas de matrícula y graduación indican que el 12 % de estudiantes de universidades públicas y el 11 % de estudiantes de universidades privadas logran terminar sus estudios de posgrado (egresados). Y de ellos, logran titularse el 63 % y 72 % de los estudiantes de la universidad pública y privada respectivamente. De igual manera señala que el mayor problema para alcanzar la graduación en el posgrado suele ser la elaboración y sustentación de los trabajos de investigación conducentes al grado académico de maestría o doctorado.

Adicional a la cuantificación del problema de la deserción es importante entender por qué se genera este fenómeno, que desde la perspectiva de las ciencias sociales significa identificar los factores y predictores de la deserción. La identificación de estos elementos puede ser realizada a través de la aplicación de distintas herramientas y metodologías, siendo las más comunes las provenientes de la econometría y los modelos de minería de datos utilizados en los últimos años (Vásquez, 2016, pág. 11).

Diversos son los estudios y técnicas desarrolladas para tratar esta problemática, dentro de ellas destacan los siguientes trabajos: En un estudio realizado por Manandhar & Sthapit (2012) utilizaron la regresión logística para el estudio de la deserción escolar en niños de Nepal a partir de variables que involucraban características de los estudiantes y del entorno social que les rodeaba, demostrando cómo estos factores pueden ser identificados y poder trabajar sobre ellos. A similares conclusiones arribaron Abdellatif Fadul & Abd Elghafar Mohammed (2018) en estudios de deserción escolar

primaria de Sudán, la regresión logística les permitió detectar los factores sociales, económicos y de infraestructura que influían significativamente en los altos índices de abandono existentes.

Barahona y Veres (2016) determinaron rendimiento académico y variables influyentes en la deserción a través de modelos de regresión múltiple en carreras de ingeniería de la Universidad de Atacama, específicamente a través de regresión logística. Madrid (2017) por su parte utiliza algunas técnicas estadísticas dentro de las que destacan Máquinas Vector Soporte, Análisis Discriminante, K-vecinos más próximos y Regresión Logística para clasificar a los estudiantes de pregrado de la Universidad Nacional de Colombia sede Medellín en dos poblaciones (con posibilidad o no de desertar). Alban Taipe & Sánchez (2018), proponen un enfoque de aprendizaje automático basado en las técnicas de regresión logística y árboles de decisión y Sandoval Palis, Naranjo, Vidal, & Gilar Corbi (2020), comparan modelos de regresión logística y redes neuronales con el mismo propósito de estimar la deserción escolar.

Los estudios de posgrado de la Universidad Técnica de Manabí, Ecuador, no están ajenos a la situación de deserción estudiantil por diferentes causas. Sin llegar a considerarse como una situación grave, es de interés identificar cuáles son las causas que más prevalecen y de qué manera puede identificarse al estudiante proclive a desertar.

En el presente estudio se pretende validar un modelo de regresión logística para la estimación de la deserción escolar en estudiantes de posgrado de la Universidad Técnica de Manabí, identificándose los principales factores influyentes que permitan tomar acciones sobre ellos.

El artículo presenta las fuentes, definición de las variables y los métodos utilizados en la sección 2. Posteriormente se exponen los resultados obtenidos y su análisis en la sección 3, para finalizar con las conclusiones del estudio en la sección 4.

2. MATERIALES Y MÉTODOS

La muestra fue conformada por 729 estudiantes que cursaban en el mes de julio de 2021 estudios de posgrado en niveles superiores o iguales al 50% del período académico de sus programas y que cumplieran con la disponibilidad de la información requerida.

Los datos se obtuvieron del Sistema de Gestión Académica (SGA), donde se registra la información personal de cada estudiante y de los registros académicos del Instituto de Posgrado, específicamente de este último se obtuvo la variable dependiente dicotómica “Deserción”.

Las variables independientes disponibles en el SGA fueron las siguientes:

- Sexo (Femenino, Masculino)
- Estado civil (Casado, Soltero, Divorciado, Unión libre, No definido)
- Trabajo (Si, No)
- Tipo de programa (Profesionalizante, Investigación)
- Salario
- Edad
- Distancia (Distancia entre lugar de residencia y la Universidad Técnica de Manabí)

Se excluyó de los análisis variables como Disponibilidad de internet, Nacionalidad, Etnia y Número de miembros en la familia por no disponerse de datos suficientes de las mismas.

Las variables Salario y Distancia fue necesaria su transformación a través del logaritmo neperiano con el objetivo de disminuir su rango de variación.

A los datos originales se les realizó un análisis descriptivo, posteriormente de la matriz resultante se seleccionó el 70% de los casos de manera aleatoria para el modelado (511 estudiantes), reservándose aproximadamente el 30% restante para su validación (215 estudiantes).

Las variables categóricas fueron codificadas tal como se muestra la **Tabla 1**:

Tabla 1. Codificación de las variables categóricas

		Frecuencia	Codificación de parámetros			
			(1)	(2)	(3)	(4)
Estado Civil	Casado	21	1	0	0	0
	Soltero	477	0	1	0	0
	Divorciado	6	0	0	1	0
	No Definido	6	0	0	0	1
	Unión Libre	1	0	0	0	0
Trabaja	No	258	1			
	Si	253	0			
Programa	Profesionalizante	291	1			
	Investigación	220	0			
Sexo	Femenino	190	1			
	Masculino	321	0			

Posteriormente se procede a la construcción del modelo de regresión logística binaria (I) (Santos Peña, Muñoz Alamillos, Juez Martel, & Cortiñas Vázquez, 2003; King, 2008; Allison, 2012):

$$P \left[y = 1 / x_1, x_2, x_3, \dots, x_k \right] = \frac{e^{f(X)}}{1 + e^{f(X)}} = \frac{1}{1 + e^{-f(X)}} \quad (I)$$

Donde:

y: variable binaria (Deserción)

$x_1, x_2, x_3, \dots, x_k$: Conjunto de variables exploratorias, donde se incorporan las variables dummy y las cuantitativas.

Una vez estimada la ecuación se procede al cálculo de los Odds (ratio del riesgo) (II):

$$Odds \text{ (ratio de riesgo)} = \frac{P[y=1/x_1, x_2, x_3, \dots, x_k]}{1-P[y=1/x_1, x_2, x_3, \dots, x_k]} = \frac{p(x_1, x_2, x_3, \dots, x_k; \beta)}{1-p(x_1, x_2, x_3, \dots, x_k; \beta)} = e^{\beta_1 + \beta_2 x_2 + \dots + \beta_k x_k} \quad (II)$$

Aplicando logaritmo neperiano en la expresión anterior, se obtiene la expresión lineal para el modelo (III)

$$Logit[P(y = 1)] = Ln \left[\frac{P[y=1/x_1, x_2, x_3, \dots, x_k]}{1-P[y=1/x_1, x_2, x_3, \dots, x_k]} \right] = \beta_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (III)$$

Para realizar el análisis del efecto de un incremento unitario de una de las variables explicativas del modelo en la variable dependiente, se calcula el Odds_ratio (Exp(B)) (IV):

$$Odds_{ratio} = \frac{Odds_2}{Odds_1} = e^{\beta_i} \quad (IV)$$

De aquí se deduce que valores en los coeficientes de la ecuación β_i próximos a cero (Odds_ratio ≈ 1), la variable explicativa asociada al coeficiente tendrá un efecto débil sobre la variable dependiente y (Deserción), valores altos del $|\beta_i|$ generaría un efecto inverso.

Para la significancia de los coeficientes del modelo estimado se utilizó el estadístico de Wald (Wald, 1943), este estadístico nos permite verificar si los coeficientes de las variables predictoras β_i son significativamente diferentes de cero o no.

$$H_0: \beta_i = \beta_{i0}$$

$$H_1: \beta_i \neq \beta_{i0}$$

El estadístico correspondiente para esta prueba es $W = \frac{(\hat{\beta}_i - \beta_{i0})^2}{\hat{\sigma}_{\hat{\beta}_i}^2}$, donde $\hat{\beta}_i$ representa los estimadores de máxima verosimilitud de los coeficientes β_i y $\hat{\sigma}_{\hat{\beta}_i}^2$ el error estándar del estimador $\hat{\beta}_i$.

$$Z = \frac{\hat{\beta}_i}{\hat{\sigma}_{\hat{\beta}_i}} \sim N \left(\left(\frac{\hat{\beta}_i}{\hat{\sigma}_{\hat{\beta}_i}} \right), 1 \right) \text{ por tanto, } Z^2 \sim \chi_{(\xi, 1)}^2, \text{ donde } \xi = \frac{(\hat{\beta}_i)^2}{\hat{\sigma}_{\hat{\beta}_i}^2}$$

En caso de variables categóricas, los grados de libertad para la prueba chi cuadrado corresponden al número de categorías de esa variable menos uno.

Para el análisis del ajuste del modelo se realizaron diferentes pruebas:

- La prueba ómnibus muestra estadístico chi cuadrado que evalúa la hipótesis nula de que los coeficientes β_i de los términos que aparecen en el modelo son cero. El estadístico de contraste en esta prueba es la diferencia existente entre el valor de $-2 \log$ de la verosimilitud ($-2LL$) sólo para la constante y el valor de $-2LL$ para el modelo actual, es decir, es usada para determinar si hay una influencia significativa al incluir las variables explicativas al modelo.
- El de Nagelkerke, es una versión corregida del coeficiente de determinación de Cox y Snell, tiene un recorrido entre cero y uno y estima la proporción de varianza de la variable dependiente explicada por las variables explicatorias.
- La prueba de Hosmer-Lemeshow, realiza una comparación entre las frecuencias esperadas y observadas partiendo de la idea de que si el ajuste es bueno se le asocia un alto valor de probabilidad a la variable explicada (Deserción), asignándole un valor uno (1) de la variable binomial dependiente, mientras que si el valor de probabilidad asociado es bajo el resultado de esa variable es cero (0).
- Clasificación de los resultados: Estimado el modelo y evaluado sobre la muestra asignada para su construcción, se evalúa el porcentaje de respuestas correctas (Observado Deserción Si entonces $y=1$ u Observado Deserción No entonces $y=0$). Un alto porcentaje (superior a 70%) garantiza fiabilidad del modelo.
- Por último, se valoró la eficiencia del modelo con el 30% de los datos de prueba, comprobándose el porcentaje de casos acertados en la clasificación realizada por el modelo estimado.

Para el procesamiento de los datos se usó el programa estadístico IBM SPSS Statistics versión 21 (IBM Corp. Released, 2012).

3. RESULTADOS Y DISCUSIÓN

Un análisis descriptivo de las variables analizadas (explicada y explicatorias) se refleja en las **Tablas 2 y 3** con el 100% de la muestra (729 estudiantes). Primeramente, se plasman las variables categóricas analizadas, incluyendo los resultados globales de deserción y posteriormente las variables numéricas.

Tabla 2. Análisis descriptivo de variables categóricas

Variable	Categorías	Frecuencia absoluta	Frecuencia relativa (%)
Sexo	Femenino	268	36,8
	Masculino	461	63,2
	Total	729	100,0
Estado civil	Casado	30	4,1
	Soltero	681	93,4
	Divorciado	9	1,2
	No Definido	7	1,0
	Unión Libre	2	0,3
	Total	729	100,0
Programa	Profesionalizante	417	57,2
	Investigación	310	42,5
	Total	727	99,7
	Perdidos	2	0,3
	Total	729	100,0
Trabaja	No	386	52,9
	Si	343	47,1
	Total	729	100,0
Deserción	No	689	94,5
	Si	40	5,5
	Total	729	100,0

Tabla 3. Descripción de las variables cuantitativas

Variable	Media	Error típico	Mínimo	Máximo
Salario (USD)	483,67	27,21	0	5111,00
Edad (Años)	35,88	0,30	22	61
Distancia (Km)	46,31	3,02	0	674,6

Existe un predominio de estudiantes del sexo masculino en los registros obtenidos, representando el 63,2% del total de estudiantes, también se observa un predominio de estudiantes solteros, siendo muy poco significativo el resto de los estados civiles registrados (6,6%). En cuanto al tipo de maestrías estudiadas existe un ligero predominio de estudiantes de las profesionalizantes, lo que se corresponde con el número de este tipo que se ofertan en la Universidad Técnica de Manabí.

Una de las principales variables que pueden tributar a la deserción estudiantil está relacionada con los ingresos producto del trabajo. Como se observa hay un ligero predominio de estudiantes que no trabajan de manera estable (52,9%), lo cual repercute en los bajos ingresos promedio obtenidos (483,67 USD mensuales), un poco por encima de los salarios mínimos del Ecuador. La edad promedio 35,88 años evidencia una discontinuidad entre los estudios de tercer nivel y cuarto nivel.

En cuanto a la deserción estudiantil presentada en el período no debe considerarse como alta (5,5%), lo cual puede atribuirse a un conjunto de acciones que realiza la Universidad Técnica de Manabí como son ayudas al 100% de los estudiantes en un porcentaje de sus matrículas y variedad de becas por diferentes conceptos como situación económica, discapacidad, enfermedades catastróficas, titulación, desarrollo institucional y convenios intra e inter institucionales.

Al estimarse el modelo de regresión logística binaria con el 70% de la muestra destinada al efecto, la prueba omnibus basada en el estadístico chi-cuadrado demuestra alta significancia de los coeficientes de las variables añadidas en la construcción del modelo ($p\text{-valor} < 0,001$).

El R^2 de Nagelkerke obtenido es de 0,789, demostrando una alta proporción de varianza de la variable dependiente (Deserción) explicada por las variables predictoras incluidas. La prueba de Hosmer-Lemeshow demostró que no hay diferencias entre los valores observados y los valores pronosticados ($p\text{-valor} = 0,896$).

Por último, la **Tabla 4** nos proporciona la clasificación realizada por el modelo de los casos usados para su construcción (70% de la muestra total), garantizándose con el mismo un 98,6% de respuestas correctas, demostrando así la validez del modelo estimado.

Tabla 4. Clasificación de los resultados

Observado		Pronosticado ^a		
		Deserción		Porcentaje correcto
		No	Si	
Deserción	No	479	2	99,6
	Si	5	25	83,3
Porcentaje global				98,6
a. El valor de corte es ,500				

La **Tabla 5** presenta los resultados globales del modelo, en una primera columna los coeficientes de la ecuación de regresión (B), los errores estándar de los mismos (ET), el estadístico de Wald con sus grados de libertad y significancia de los parámetros estimados, y por último el Odds_ratio (e^{β_j}) acompañado de un intervalo de confianza del 95% para los mismos, estos últimos explicarán el efecto que tendrán en la deserción cambios unitarios en las variables explicativas manteniendo constante el resto de las variables. Debe tenerse en cuenta en dicho análisis que las variables Salario y Distancia se encuentran transformadas para la estimación del modelo.

Tabla 5. Resultados globales del modelo estimado.

Variables	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
Sexo(S)	-,523	,855	,374	1	,541	,593	,111	3,167
E_Civil			44,012	4	,000			
E_Civil (EC1)	-21,689	40193,205	,000	1	1,000	,000	,000	.
E_Civil (EC2)	-28,002	40193,205	,000	1	,999	,000	,000	.
E_Civil (EC3)	-20,884	40193,205	,000	1	1,000	,000	,000	.
E_Civil (EC4)	-19,931	40193,205	,000	1	1,000	,000	,000	.
Programa (P)	-,388	,805	,232	1	,630	,678	,140	3,287
Trabaja (T)	-6,274	2,467	6,467	1	,011	,002	,000	,237
Edad (E)	,182	,054	11,299	1	,001	1,200	1,079	1,334
Ln_Distancia (LnD)	,098	,187	,275	1	,600	1,103	,764	1,593
Ln_Salario (LnS)	-1,002	,371	7,312	1	,007	,367	,178	,759
Constante	22,002	40193,205	,000	1	1,000	3593209760,073		
Variables introducidas: Sexo, E_Civil, Programa, Trabaja, Edad, Ln_Distancia, Ln_Salario.								

Un análisis del estadístico de Wald arroja baja significancia de variables como sexo (p-valor=0,541), programa (p-valor=0,630) y el Ln_Distancia, no obstante, se valoró mantenerlas dentro del modelo, permitiendo incrementar la precisión del mismo y no ser variables de difícil acceso.

Las variables que mayor aportan a la explicación de la deserción estudiantil fueron: el estado civil, situación laboral, edad y salario, detectándose de esta manera cuáles son los factores que pueden atribuirse como causas más incidentes en la problemática. Coinciden estos resultados con los obtenidos por Manandhar & Sthapit (2012), Abdellatif Fadul & Abd Elghafar Mohammed (2018) y Alban Taipe & Sánchez (2018), en cuyos estudios presentan a los factores económicos y sociales como los más incidentes en la deserción escolar.

La ecuación resultante para los pronósticos es la siguiente (V):

Los signos de los coeficientes de las variables significativas demuestran que, disponer de trabajo disminuye la probabilidad de deserción, de igual manera influyen los salarios y en sentido inverso la edad.

Una vez comprobadas las bondades de modelo desde el punto de vista analítico, así como estimada la ecuación logística binaria de pronóstico, se procede a su validación con el 30% de la muestra restante. Para ello se aplica el modelo estimado a estos datos y se obtienen los resultados que se muestran en la **Tabla 6**.

Tabla 6. Validación de los resultados con el 30% de la muestra

Observado		Pronosticado		
		Deserción		Porcentaje correcto
		No	Si	
Deserción	No	203	2	99,02
	Si	0	10	100
Porcentaje global				99,06

Como se puede observar se ha alcanzado un 99,06% de eficiencia, lo que demuestra que el modelo estimado cumple las funciones para lo cual fue establecido.

4. CONCLUSIONES

Los modelos de regresión logística binaria han estado avalados en la literatura como una herramienta útil que permite identificar las posibles causas de deserción estudiantil y realizar pronósticos de las mismas.

Se demostró, desde el punto de vista analítico y a través de los pronósticos realizados en la muestra de modelado (70%) y la muestra de validación (30%), que el modelo estimado es capaz de garantizar fiabilidad en sus estimaciones, alcanzándose el 98,6% de respuestas correctas en la muestra piloto y 99,06% en la muestra de validación.

El modelo estimado, permitió pronosticar aquellos estudiantes proclives a desertar de sus estudios de posgrado en la Universidad Técnica de Manabí, identificando como las principales causas el estado civil, situación laboral, edad y salario.

5. AGRADECIMIENTOS

Se agradece el apoyo recibido por el personal del Instituto de Posgrado de la Universidad Técnica de Manabí al brindar la información necesaria para el desarrollo de este trabajo y al proyecto “Diseño de la Calidad de los procesos de posgrado de la Universidad Técnica de Manabí”, al cual se asocia esta investigación.

6. DECLARACIÓN DE CONFLICTO DE INTERÉS DE LOS AUTORES

Los autores declaran no tener conflicto de intereses

7. REFERENCIAS

- Abdellatif Fadul, A., & Abd Elghafar Mohammed, K. (2018). Using Logistic Regression to Model Factors Associated With Basic School Drop-Out in Kassala State, Sudan. *Global Journal of Pure and Applied Mathematics*, 14(2), 301-313. Obtenido de https://www.ripublication.com/gipam18/gipamv14n2_11.pdf
- Alban Taípe, M. S., & Sánchez, D. M. (2018). Prediction of university dropout through technological factors: a case study in Ecuador. *Espacios*, 39(52), 8. Obtenido de <http://www.revistaespacios.com/a18v39n52/a18v39n52p08.pdf>
- Allison, P. D. (2012). *Logistic regression using SAS: Theory and application* (Second ed.). SAS Institute Inc.
- Barahona, P., Veres, E., & Aliaga, V. (2016). Deserción académica de la Universidad de Atacama, Chile. *Comuni@cción*, 7(2), 27-37. Obtenido de <http://www.scielo.org.pe/pdf/comunica/v7n2/a03v7n2.pdf>
- Carpio, A. (2017). Frecuencia y Factores Asociados a la Deserción Universitaria en Estudiantes de la Escuela de Posgrado Modalidad a Distancia de la Universidad Católica de Santa María. Arequipa 2013-2015. Obtenido de <http://tesis.ucsm.edu.pe/repositorio/handle/UCSM/6696>
- Fernández, T., Solís, M., Hernández, M., & Moreira, T. (2019). Un análisis multinomial y predictivo de los factores asociados a la deserción universitaria. *Educare*, 23(1), 73-97. doi: <http://dx.doi.org/10.15359/ree.23-1.5>
- Hernández, O., Pérez, R., & González, G. (2014). La deserción en los posgrados, un problema no menor. *Diálogos Sobre Educación*, 5(8), 2. doi:DOI: 10.32870/dse.v0i8.309
- IBM Corp. Released. (2012). IBM SPSS Statistics for Windows, versión 21.0. Armonk, NY.
- King, J. E. (2008). Binary Logistic Regression. En J. W. Osborne, *Best practice in quantitative methods* (pág. 358). SAGE Publications.
- Lopera, O. (2008). *Determinantes de la deserción universitaria en la Facultad de Economía de la Universidad del Rosario*. Rosario, Argentina: Universidad del Rosario. Obtenido de <https://ideas.repec.org/p/col/000091/004558.html>
- Madrid, J. (Julio de 2017). Propuesta de un modelo estadístico para caracterizar y predecir la deserción estudiantil universitaria. *Tesis de maestría*. Medellín, Colombia: Universidad Nacional de Colombia. Obtenido de <https://repositorio.unal.edu.co/handle/unal/60097>
- Manandhar, N., & Sthapit, A. (2012). Logistic regression model for primary school dropout children. *International Journal of Asian SociScience*, 2(2), 145-151. Obtenido de https://www.researchgate.net/publication/227368344_Logistic_Regression_Model_For_Primary_School_Drop_out_Children_Of_Chitwan_District_Of_Nepal
- Mifflin, M. (2017). La deserción en el posgrado: estudio comparativo entre maestrías de una universidad pública y privada. Córdoba, Argentina: Congresos CLABES VII. Obtenido de <https://revistas.utp.ac.pa/index.php/clabes/article/view/1681>
- Molina, E., & González, M. (2019). Analíticas en la calidad de la gestión del posgrado. *Didasc@lia: Didáctica y Educación*, 10(1), 9-24. Obtenido de <https://dialnet.unirioja.es/servlet/articulo?codigo=7241785>
- Sandoval Palis, I., Naranjo, D., Vidal, J., & Gilar Corbi, R. (2020). Early Dropout Prediction Model: A Case Study of University Leveling Course Students. *Sustainability*, 12(22), 4-17. doi:doi:10.3390/su12229314
- Santos Peña, J., Muñoz Alamillos, Á., Juez Martel, P., & Cortiñas Vázquez, P. (2003). *Diseño de encuestas para estudios de mercado. Técnicas de muestreo y análisis multivariante*. Centro de estudios Ramón Areces, S.A.

- Sinchi, E., & Gómez, G. (2018). Acceso y deserción en las universidades. Alternativas de financiamiento. *Alteridad. Revista de educación*, 13(2), 274-287. doi: <https://doi.org/10.17163/alt.v13n2.2018.10>
- Vásquez, J. (Mayo de 2016). Modelo predictivo para estimar la deserción de estudiantes en una institución de educación superior. *Tesis de maestría*, 06. Santiago de Chile, Chile. Obtenido de <https://repositorio.uchile.cl/bitstream/handle/2250/144169/V%C3%A1squez%20Verdugo%20Jonathan.pdf?sequence=1>
- Wald, A. (1943). Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations Is Large. *Transactions of the American Mathematical Society*. doi: <https://doi.org/10.1090/S0002-9947-1943-0012401-3>

Contribución de autores

Autor	Contribución
Julio César Solís Ventura	Procesamiento de los datos, revisión de la literatura y escritura del manuscrito
Santiago Quiroz Fernández	Propuesta del proyecto, obtención de la información y escritura y revisión del manuscrito
Osvaldo Fosado Téllez	Propuesta del proyecto, obtención de la información, procesamiento de datos y revisión del manuscrito