



Modelo Logit para predecir competencias óptimas de matemáticas preuniversitarias en estudiantes de nuevo ingreso al propedéutico de la UACH

Logit model to predict optimal competences in pre-university mathematics in new students entering the UACH propedeutical

Manuel Jesús Negrete Quijano^{1*} y Zaztal Marigly de Atocha Santos Magaña¹

¹Universidad Autónoma Chapingo, Centro Regional Universitario Península de Yucatán, Km 10 carretera Merida-Progreso, Comisaría Temozón Norte, CP 97300 Mérida, Yucatán, México.

*Corresponding author:
mnegreteq@chapingo.mx

Resumen: En este trabajo se propone un modelo de Regresión Logística (Logit) como técnica de análisis multivariado, para la predicción del nivel óptimo de competencias matemáticas preuniversitarias, a manera de que se pueda determinar si los estudiantes ingresan a la Universidad con competencias básicas o si requieren apoyo y en qué áreas, para así poder garantizar el éxito en sus estudios superiores; siendo ésta menester de la Universidad Autónoma Chapingo, se lleva a cabo esta investigación con estudiantes de nuevo ingreso al propedéutico de dos Centros Regionales, en donde se aplicó una prueba diagnóstica y se obtuvo un modelo que contiene 4 variables: calificaciones de lenguaje algebraico, álgebra, funciones y sus gráficas, y sexo, luego de realizar el método hacia atrás con un modelo inicial de 8 variables que contemplaban además de las anteriores, calificación final de la prueba, aritmética, trigonometría y tiempo de resolución de la prueba.

Palabras clave: Modelo Logit, análisis multivariado, predicción, diagnóstico.

Abstract: In this work, a Logistic Regression model (Logit) is proposed as a multivariate analysis technique, for the prediction of the optimal level of pre-university mathematical skills, so that it can be determined whether students enter the University with basic skills or if they require support and in what areas, in order to guarantee success in their higher studies; This being the responsibility of the Chapingo Autonomous University, this research is carried out with new students at two Regional Centers, where a diagnostic test was applied and a model was obtained that contains 4 variables: algebraic language qualifications, algebra, functions and their graphs, and sex, after carrying out the backward method with an initial model of 8 variables that included, in addition to the previous ones, final

grade of the test, arithmetic, trigonometry and time to solve the test.

Keywords: Logit model, multivariate analysis, prediction, diagnosis.

I. INTRODUCCIÓN

La formación preuniversitaria es fundamental para que los estudiantes de bachillerato puedan ingresar a la universidad, en este sentido Sánchez et. al. (2020) indica que también garantiza el éxito en su trayectoria en el programa educativo de su elección. En este sentido, el dominio de competencias matemáticas a nivel preuniversitario es una necesidad que da soporte y garantiza el éxito o fracaso en los estudios superiores de corte ingenieril y ciencias. Stone (2003) indica que las habilidades matemáticas son necesarias para todos aquellos que quieren avanzar en su camino profesional. Por su parte, Herzog (2005), citado en Sánchez et. al. (2020) encontró que la falta de dominio de las habilidades matemáticas es un indicador fuerte de la posible deserción del primer y segundo semestres de la universidad.

Particularizando en los programas educativos preuniversitarios que se imparten en la Universidad Autónoma Chapingo (UACH): la *preparatoria agrícola y el propedéutico*; el primero es un programa de tres años con ingreso después de la secundaria, y el segundo, es un curso obligatorio de un año para todo aspirante a ingresar en alguna ingeniería o licenciatura agronómica que haya concluido el bachillerato en otra institución. La preparatoria Agrícola se imparte en la sede central que se encuentra en Texcoco, Estado de México, mientras que el programa de Propedéutico, además, se imparte en cuatro de ocho Centros Regionales distribuidos a nivel nacional. Dos de esos Centro Regionales se encuentran en Yucatán y en Guerrero. El Centro Regional Universitario Península de Yucatán (CRUPY) y el Centro Regional de San Luis Acatlán (SLA), reciben estudiantes para el nivel propedéutico de los estados respectivos y de otros, como Chiapas, Campeche y Quintana Roo.

A nivel institucional, se comparte la preocupación del alto índice de reprobación y deserción escolar (Acuerdo 1219-1, del H. Consejo Universitario en su sesión ordinaria del día 16 de junio del 2023). Dichos índices se reflejan principalmente en el área de matemáticas y física, entre otras disciplinas. Se ha observado que, el

nivel de dominio de habilidades matemáticas en la prueba institucional de ingreso a la UACH, la cual es presentada por todo estudiante que se incorpora al propedéutico, no garantiza éxito en los cursos subsecuentes de matemáticas. En atención a esta problemática surge la necesidad de evaluar el área de matemáticas antes de iniciar con el propedéutico a modo de prueba diagnóstica, de acuerdo con Cámara (2013), quien expone que los resultados de las pruebas diagnósticas, son altamente predictivos de las calificaciones obtenidas en la universidad.

En línea con la idea de las pruebas diagnósticas en el área de Matemáticas Ngo (2019), Corbishley & Truxaw (2010) mencionan que las competencias matemáticas necesarias para tener acceso a cursos de nivel superior, susceptibles de evaluarse, son la fluidez en los procedimientos con fracciones, la resolución de ecuaciones algebraicas, y resolución de problemas de lenguaje algebraico, razonamiento y generalización; en ese tenor Negrete (2010) consideró que las competencias previas a los cursos de matemáticas del programa educativo del Propedéutico son: aritmética, lenguaje algebraico, álgebra, trigonometría, funciones y sus gráficas, cada una de ellas considerada como factor predictivo en nuestro análisis.

En términos de esta problemática y en vías de subsanarla, nos dimos a la tarea de dar respuesta a los siguientes cuestionamientos ¿en qué nivel de aprovechamiento de matemáticas estarán los estudiantes de nuevo ingreso al propedéutico?, ya que provienen de diferentes entidades del país y modalidades del bachillerato ¿cuáles son las competencias matemáticas que dominan y pueden demostrar?, ¿Qué porcentaje de los estudiantes necesitarán seguimiento y asesoría académica complementaria?, ¿cuál es la probabilidad de que un estudiante tenga un nivel óptimo de competencias matemáticas?, y de manera general, ¿se puede generar un modelo matemático predictivo que pueda dar respuesta a estas interrogantes a través de una herramienta diagnóstica?.

Considerando la articulación de esta problemática se derivó el objetivo de investigación:

A. Objetivo

Encontrar un modelo para diagnosticar y predecir la probabilidad del nivel óptimo y no óptimo de competencias matemáticas en estudiantes

preuniversitarios, que inician el curso propedéutico de la UACH, mediante la regresión logística (modelo logit) multivariada con los datos de una herramienta diagnóstica.

II. MARCO TEÓRICO

Entre las técnicas de análisis multivariado de dependencia se encuentra el análisis de la regresión logística o modelo logit. Para Obregón et al. (2015) el objetivo de estas pruebas, al igual que la regresión lineal, es encontrar un modelo de buen ajuste para los datos que describa la relación entre variables independientes, métricas o no métricas, con una variable dependiente categórica.

A. Modelo Logit

La regresión logística predice directamente la probabilidad de ocurrencia de un suceso (Hair et al. 1999), considerando el vector x_i con k variables predictoras, tal que $x_i = (1, x_{i1}, x_{i2}, \dots, x_{ik})'$, así como el vector de observaciones $Y = (y_1, y_2, \dots, y_n)'$ donde Y es una variable categórica binaria para los $i = 1, 2, \dots, n$ unidades experimentales que conforman la muestra. Se asume que $y_i = 1$ tiene probabilidad p_i , cuando $y_i = 0$ entonces su probabilidad es $1 - p_i$.

El modelo logit, está dado por $p_i = f(x_i' \alpha)$ donde $f(z) = \frac{1}{1 + e^{-z}}$ es la función de distribución logística, con $-\infty < z < \infty$, y $\alpha = (\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_k)'$ es el vector de parámetros a estimar, donde α_j representa el efecto de la j -ésima variable predictora, para $j = 1, 2, \dots, k$, y α_0 es el término independiente del modelo.

La estimación de los parámetros α se lleva a cabo con el método de máxima verosimilitud, considerando los datos precedentes y la función:

$$L(f(x_i' \alpha) | y_i) = \prod_{i=1}^n [f(x_i' \alpha)]^{y_i} [1 - f(x_i' \alpha)]^{1 - y_i}$$

Luego, se maximiza el logaritmo natural de la misma

$$\frac{\delta \ln L(f(x_i' \alpha) | y_i)}{\delta \alpha_i} = 0$$

Una vez obtenida la estimación de los parámetros $\alpha_0, \alpha_1, \dots, \alpha_k$ se prueba el nivel de significación de cada variable predictora. Para ello, Ibarra & Michalus (2010) consideran los Odds Ratio y el estadístico de

Walt; y para analizar la bondad de ajuste del modelo se utiliza el valor de verosimilitud, el coeficiente de determinación (Pseudo-R² de Mc Fadden) y la prueba de Hosmer y Lemeshow.

A continuación, se describe brevemente cada elemento: En Pérez (2004) se define un Odds (ventaja) como aquella razón que indica que tan probable es la ocurrencia de un evento usando probabilidades, si p es la probabilidad de éxito y q la probabilidad de fracaso, entonces Odds = p/q . De esta manera, el Odds Ratio (ventaja comparativa o riesgo relativo) es el cociente entre dos Odds, o bien, puede expresarse como el exponencial del coeficiente de una variable predictora, entonces OR = e^{α_j} , donde α_j es el parámetro a estimar de la j -ésima variable del modelo.

El Estadístico de Walt indica el nivel de significancia de cada variable en el modelo logit, si el p -valor asociado es menor al nivel de significancia seleccionado, entonces la variable predictora puede ser considerada en el modelo.

El análisis de la verosimilitud es equivalente a la prueba F en regresión lineal múltiple, consiste en comparar las probabilidades predichas con las observadas mediante la prueba Chi-cuadrada. Indica si la diferencia es significativa en la reducción del $-2 \ln L(f(x_i' \alpha) | y_i)$ entre el modelo inicial y el final. Un modelo ideal tendría valor de verosimilitud nulo $-2 \ln L(f(x_i' \alpha) | y_i) = 0$, por lo tanto se esperan valores pequeños cercanos a cero para que exista un buen ajuste del modelo.

El coeficiente Pseudo-R² de Mc Fadden indica un buen ajuste del modelo mientras más cercano se encuentre de la unidad, sus valores están en el intervalo $I = (0, 1)$.

Por último, la prueba de ajuste del modelo que encontramos en Hosmer & Lemeshow (2000) muestra si existe diferencia significativa entre los valores observados y los predichos para la variable respuesta, mediante la prueba Chi-cuadrada. Sin embargo, esta prueba no siempre es segura para tamaños grandes de muestras, es por eso que también se usa el análisis de la verosimilitud, el Criterios de Información Akaike (AIC) y el Criterio de Información Bayesiano (BIC).

B. Instrumento de medición

Para diagnosticar y predecir la probabilidad del nivel óptimo de las competencias matemáticas, se adaptó el

instrumento de medición utilizado en Negrete (2010) que abarca las cinco áreas de conocimiento: aritmética, el lenguaje algebraico, el álgebra, la trigonometría, las funciones y sus gráficas.

El diseño del instrumento de medición diagnóstico consideró las competencias en las cinco áreas de matemáticas. Consiste en una prueba con 28 reactivos integradores cuya resolución refleja el dominio de las competencias pertinentes para un estudiante que concluye bachillerato e inicia un programa de nivel superior. Estos reactivos están divididos de la siguiente manera: 6 para el área de aritmética, 5 de lenguaje algebraico, 7 de álgebra, 4 de trigonometría, 6 de funciones y sus gráficas.

Se aplicó el instrumento a una muestra de 68 estudiantes de nuevo ingreso al nivel propedéutico de los Centros Regionales de la UACH, en los estados de Yucatán y Guerrero. La calificación mínima aprobatoria para demostrar el dominio en las competencias en cada área fue de 66 puntos, en una escala de 0 a 100.

C. Variables

La variable dependiente, nombrada y=óptimo, es el nivel óptimo de competencias matemáticas de cada estudiante. Es de tipo categórica con dos niveles, el primero y₁=1 es el éxito o nivel óptimo, cuando el estudiante aprueba con una calificación de 66 puntos al menos tres de las cinco áreas matemáticas; el segundo, y₂=0 es el fracaso o nivel no óptimo, cuando el estudiante aprueba con una calificación de 66 puntos dos o menos de las cinco áreas matemáticas.

Las variables independientes que se consideraron se describen en la Tabla 1.

Tabla 1. Descripción de variables. Fuente: Elaboración propia.

Variable (x _i)	Descripción
x ₁ =arit	métrica de 0 a 100, calificación del área aritmética.
x ₂ =lal	métrica de 0 a 100, calificación del área lenguaje algebraico.
x ₃ =alg	métrica de 0 a 100, calificación del área álgebra.
x ₄ =trigo	métrica de 0 a 100, calificación del área trigonometría.
x ₅ =fncg	métrica de 0 a 100, calificación del área funciones y sus gráficas.
x ₆ =tiempo	métrica de 10 a 90, duración de la prueba en minutos.
x ₇ =sexo	no métrica, dicotómica, género del estudiante: h, m.
x ₈ =cfinal	métrica de 0 a 100, promedio de calificaciones de todas las áreas.

Entonces, Consideramos el modelo de Regresión Logit Multivariada para analizar la base de datos con las variables definidas en la Tabla 1, y la variable dependiente categórica binaria, con la función:

$$f(x_i' \alpha) = \frac{1}{1 + e^{-[\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 + \alpha_5 x_5 + \alpha_6 x_6 + \alpha_7 x_7 + \alpha_8 x_8]}}$$

para los coeficientes α_j donde j = 0,1,...,8.

Se utilizó el lenguaje de programación en R, software libre en su versión 4.3.1, para la estimación de los parámetros, la generación de tablas y gráficos.

III. RESULTADOS

Los resultados obtenidos del modelo logit, se presentan en la Tabla 2.

Tabla 2. Resultados del modelo logit con todas las variables. Fuente: Elaboración propia.

Variable	Estimación	Error Est	z value	Pr(> z)
Intercept	-6.231e+01	4.169e+01	-1.495	0.135
arit	3.718e+00	3.899e+00	0.954	0.340
lal	3.789e+00	3.930e+00	0.964	0.335
alg	5.485e+00	5.517e+00	0.994	0.320
trigo	2.860e+00	3.003e+00	0.952	0.341
fncg	5.263e+00	5.465e+00	0.963	0.336
tiempo	-8.801e-05	3.408e-02	-0.003	0.998
sexoM	-4.027e+00	2.951e+00	-1.364	0.172
cfinal	-2.013e+01	2.128e+01	-0.946	0.344

Las variables métricas arit, lal, alg, trigo, fncg, y cfinal se analizaron de forma univariada en la Figura 1, mediante los gráficos boxplot que reflejaron sesgos asimétricos, pertinentes para el análisis logit, y de forma bivariada con el mapa de correlaciones Figura 2.

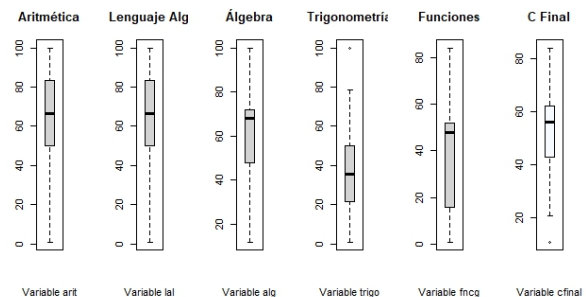


Figura 1. Variables métricas. Fuente: Elaboración propia.

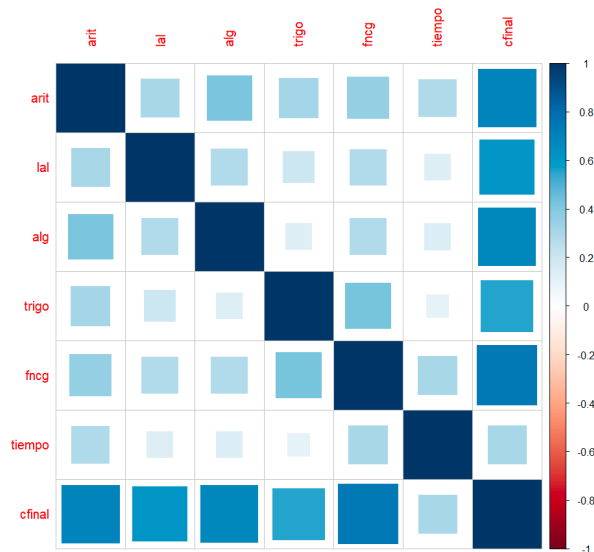


Figura 2. Mapa de correlaciones entre variables.
Fuente: Elaboración propia.

Para la muestra de estudiantes que aplicaron la prueba, se encontró que el 43% de los hombres y el 28% de las

mujeres cuentan con el nivel óptimo de competencias para iniciar el curso de Matemáticas I. Dato que conlleva a considerar que el 67% de los estudiantes requiere seguimiento académico con asesoría matemática en sus distintas áreas. Principalmente, las que necesitan mayor reforzamiento son aritmética y trigonometría, con promedios de 60 y 36 puntos, respectivamente, en una escala de 0 a 100.

Además, se encontró que las variables arit, trigo, tiempo y cfinal tienen las correlaciones más altas (> 0.31) con las cuatro métricas restantes, atribuyendo el color azul respectivamente en la escala señalada en la acotación derecha de la Figura 2. Por lo que se sospecha que todas o algunas de ellas pueden omitirse en el modelo, sin perder información de ajuste.

Es así, que realizamos la técnica hacia atrás del modelo logit, la cual consistió en eliminar variables al modelo completo y analizar la bondad de ajuste de los nuevos modelos. En la Tabla 3, se presenta la comparación entre modelos, utilizando las principales pruebas estadísticas que definen el mejor ajuste.

Tabla 3. Comparación de modelos logit. Fuente: Elaboración propia.

Mod.	x_i	X^2	X^2 p-value	Pseudo-R ² (Cragg-Uhler)	Pseudo-R ² (Mc Fadden)	AIC	BIC
I	Con todas las variables	56.83	0.00	0.90	0.82	30.71	48.78
II	Sin tiempo	56.83	0.00	0.90	0.82	28.71	44.77
III	Sin sexo	53.15	0.00	0.86	0.76	32.40	48.46
IV	Sin cfinal	55.27	0.00	0.88	0.79	30.27	46.33
V	Sin cfinal y tiempo	55.19	0.00	0.88	0.79	28.36	42.41
VI	Sin cfinal, tiempo arit, y trigo	54.26	0.00	0.87	0.78	25.29	35.32
VII	Sin cfinal, tiempo, arit, trigo, sexo	47.19	0.00	0.80	0.68	30.36	38.38

Ubicamos la atención en el modelo VI, con un resultado de la prueba de Hosmer & Lemeshow de $X^2 = 55$ y aunque tiene un p-valor menor a 0.05, lo determinante fue considerar las pruebas de verosimilitud $X^2 = 54.26$ con un p-valor menor a 0.05, en conjunto con los Criterios de Información Akaike de 25.29 y de Información Bayesiano de 35.32 menores a los demás, por lo tanto, estos datos comprueban que es el mejor modelo para la variable dependiente nivel óptimo de competencias matemáticas, con un 78% de variabilidad

explicada resultante de la prueba Pseudo-R² de Mc Fadden.

La estimación de los parámetros α (estimate), el error estándar (std. error), el valor de wald (z value), y el p-valor del z de wald ($Pr(>z)$) se muestran en la Figura 3.

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -36.26094   16.11795  -2.250   0.0245 *
lal          0.13260    0.06484   2.045   0.0408 *
alg          0.31610    0.14766   2.141   0.0323 *
fncg         0.12737    0.05818   2.189   0.0286 *
sexoM        -4.30650    2.29439  -1.877   0.0605 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 69.545  on 54  degrees of freedom
Residual deviance: 15.287  on 50  degrees of freedom
AIC: 25.287
    
```

Figura 3. Resumen del modelo logit en R. Fuente: Elaboración propia.

De los Odds Ratio (OR) y los intervalos de confianza del 95% que se observan en la Figura 4, se obtuvieron los siguientes resultados:

- El coeficiente α_1 de la variable lal tiene un valor positivo y su Odds Ratio es de 1.14, lo que indica que por cada incremento puntual en la calificación del área de lenguaje algebraico, tiene 14% de probabilidad de tener nivel óptimo.
- El OR del coeficiente $\alpha_2=0.316$ de la variable alg, indica que por cada incremento puntual en la calificación del área de álgebra, tiene 37% de probabilidad de tener el nivel óptimo de competencias.
- El OR del coeficiente $\alpha_3=0.127$ de la variable fncg, nos indica que por cada incremento en la calificación del área funciones y sus gráficas, el estudiante tiene 14% de probabilidad de tener nivel óptimo.
- El coeficiente α_4 de la variable sexo tiene un valor negativo y OR de 0.01, lo que indica que las mujeres tienen sólo el 1% de probabilidad de tener el nivel óptimo de competencias, en comparación con los hombres.

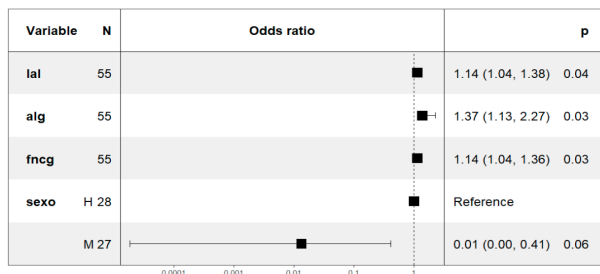


Figura 4. Odds Ratio del modelo VI. Fuente: Elaboración propia.

Para evaluar la predicción del modelo se obtuvo de la prueba ROC Figura 5, un 67% de área bajo la curva, y

con la Matriz de Confusión Figura 6, un nivel de exactitud (accuracy) del 69.23%.

```

Call:
roc.default(response = test.data$optimo, predictor = probabilidades, ci = T)

Data: probabilidades in 9 controls (test.data$optimo no) < 4 cases (test.data$optimo si).
Area under the curve: 0.6667
95% CI: 0.3216-1 (DeLong)
    
```

Figura 5. Prueba de predicción. Fuente: Elaboración propia.

Confusion Matrix and Statistics

```

Reference
Prediction no si
no      8  3
si      1  1

Accuracy : 0.6923
95% CI : (0.3857, 0.9091)
No Information Rate : 0.6923
P-Value [Acc > NIR] : 0.6310

Kappa : 0.1613

McNemar's Test P-Value : 0.6171

Sensitivity : 0.8889
Specificity : 0.2500
Pos Pred Value : 0.7273
Neg Pred Value : 0.5000
Prevalence : 0.6923
Detection Rate : 0.6154
Detection Prevalence : 0.8462
Balanced Accuracy : 0.5694

'Positive' Class : no
    
```

Figura 6. Matriz de confusión. Fuente: Elaboración propia.

Se complementó el análisis con las gráficas de predicción óptima versus lenguaje algebraico Figura 7, vs. álgebra Figura 8, vs. funciones y sus gráficas Figura 9.

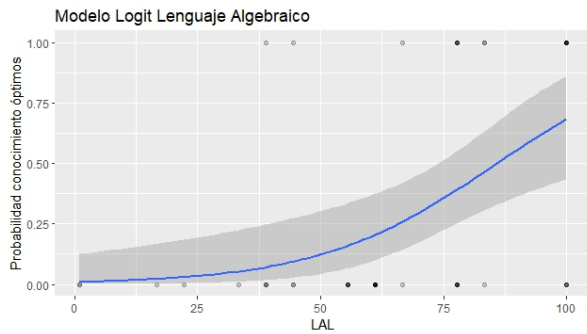


Figura 7. Predicción *lal*. Fuente: Elaboración propia.

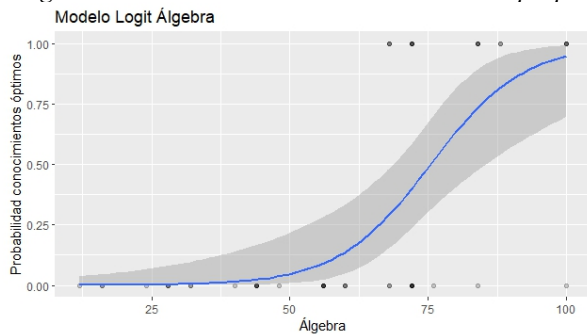


Figura 8. Predicción *alg*. Fuente: Elaboración propia.

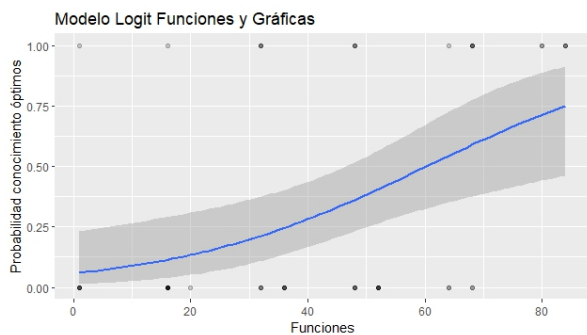


Figura 9. Predicción *fncg*. Fuente: Elaboración propia.

IV. CONCLUSIONES

Después de analizar los datos de una herramienta diagnóstica mediante la regresión logística multivariada, que consideró ocho variables independientes, descritas en la *Tabla 1*, y una variable dependiente de tipo categórica y binaria, observamos que no todas las variables consideradas eran significativas para describir la dependencia (*Tabla 2*).

Encontramos que, el modelo logit de mejor ajuste contiene cuatro de las ocho variables, cuya expresión es:

$$f(x_i, \alpha) = \frac{1}{1 + e^{-[-36.260 + 0.132|lal| + 0.316|alg| + 0.127|fncg| - 4.306|sexo]}}$$

donde $f(x_i, \alpha)$ predice la probabilidad de que un estudiante preuniversitario tenga o no el nivel óptimo de competencias matemáticas, al iniciar el curso propedéutico de la UACH, las variables independientes fueron las calificaciones obtenidas en el área de lenguaje algebraico, álgebra, funciones y sus gráficas, y el sexo del estudiante.

Este modelo de regresión logística lo consideramos un buen modelo para la predicción y diagnóstico, ya que su capacidad discriminativa es adecuada con un 67% de probabilidad para diagnosticar si un estudiante cuenta o no con el nivel óptimo de competencias matemáticas, y con una precisión mayor al 69% de los casos que sí cuentan con el nivel óptimo.

De acuerdo con el modelo encontrado, se concluye que los estudiantes que obtengan menos de 70 puntos en las áreas de álgebra, lenguaje algebraico, funciones y sus gráficas del instrumento diagnóstico, tienen una baja probabilidad del nivel óptimo de competencias matemáticas, por lo que deberán llevar un seguimiento académico complementario para incrementarla, y así contribuir al nivel óptimo que garantice el éxito en los cursos de matemáticas de la UACH.

Atendiendo a la preocupación de la universidad, con respecto al alto índice de reprobación y rezago, este modelo sería de utilidad al inicio de cada curso para determinar la pertinencia de estrategias de aprendizaje con un enfoque centrado en el alumno, asesorías para las áreas de matemáticas que no contribuyen al nivel óptimo, así como la adecuación del programa y la incorporación de tecnología educativa al fenómeno didáctico.

Sugerimos para futuros estudios, considerar un complemento del análisis mediante el incremento del tamaño muestral, además de un estudio comparativo con otra técnica multivariada que incluya más variables predictoras, como son, el resultado de la prueba de ingreso a la UACH, el tipo de sistema educativo inmediato anterior, la localidad de procedencia, la edad, si habla alguna lengua indígena, entre otras posibles. Todo ello en vías de aumentar la capacidad predictiva del modelo logit multivariado, y coadyuvar de esa manera a minimizar la problemática.

REFERENCIAS

- Camara, W. (2013). Defining and Measuring College and Career Readiness: A Validation Framework. *Educational Measurement: Issues and Practice*, 32, 16-27.
- Corbishley, J. & Truxaw, M. (2010). Mathematical Readiness of Entering College Freshmen: An Exploration of Perceptions of Mathematics Faculty. *School Science and Mathematics*, 110(2).
- Hair, J., Anderson, R., Tatham, R., & Black, W. (1999). *Análisis Multivariante*. Prentice Hall, Madrid. 5ª ed.
- Herzog, S. (2005). Measuring Determinants of Student Return vs. Dropout/Stopout vs. Transfer: A First-to-Second Year Analysis of New Freshmen. *Research in Higher Education*, 46(8), 883-928. Retrieved from <http://www.jstor.org/stable/40197396>.
- Hosmer, W., & Lemeshow, S. (2000). *Applied Logistic Regression*. John Wiley & Sons Inc.
- Ibarra, M., & Michalus, J. (2010). Análisis del rendimiento Académico mediante un modelo logit. 4º Congreso Argentino de Ingeniería Industrial COINI 2010, Misiones, Argentina.
- Ngo, F. (2019). Fractions in College: How Basic Math Remediation Impacts Community College Students. *Research in Higher Education* 60(4), 485-520. doi: [10.1007/s11162-018-9519-x](https://doi.org/10.1007/s11162-018-9519-x).
- Negrete, M. (2010). Prueba diagnóstica de matemáticas para alumnos de nuevo ingreso al propedéutico: reporte general. *Artículo no publicado del CRUPY -UACH*. Recuperado de <https://n9.cl/4cn50>
- Obregón, Biosca, S. A., Romero Navarrete, J. A., & Betanzo Quezada, E. (2015). La estadística en los estudios movilidad urbana en la Zona Metropolitana de Querétaro. *Facultad de Ingeniería de la Universidad Autónoma de Querétaro*, 1-30.
- Pérez, C. (2004). *Técnicas de Análisis Multivariante de Datos. Aplicaciones con SPSS*. Prentice Hall. 646 pp.
- Sánchez, S., Ramírez, E., Calderón, C., & Sánchez, R. (2020). Conocimientos matemáticos de los estudiantes de primer ingreso del CUValles de la Universidad de Guadalajara. *Revista Iberoamericana para la Investigación y el Desarrollo Educativo*. 10(20). doi:10.23913/ride.v10i20.675
- Stone, J. R. (2003). Improving Math Skills in CTE: How You Can Help. *Techniques: Connecting Education & Careers*, 78(3), 58. Retrieved from <https://search.ebscohost.com/login.aspx?direct=true&db=tfh&AN=9203716&site=ehost-live>.