

El uso de la regresión logística para conocer el gusto por las Matemáticas en algunas preparatorias públicas de provincia de la ciudad de México

Reyes Cervantes Hortensia Josefina¹, Cruz Armenta Diana Jazmín², Arenas Martínez Guadalupe Yoanna³, Cruz Suárez Hugo⁴, Godínez Jaimes Flaviano⁵, Ariza Hernández Francisco⁶, Vázquez Guevara Víctor Hugo⁷.

Resumen

Se usa regresión logística para investigar las variables que influyen en el gusto por las matemáticas en estudiantes de las preparatorias de la Benemérita Universidad Autónoma de Puebla. Interesó conocer si había relación entre el gusto por las matemáticas y las variables: ambiente propicio, horas de estudio, la dificultad por algún contenido en el material visto en sus cursos, gusto por el método de enseñanza de sus profesores y actitud cuando tenían dudas y otras más.

Se encontró que los alumnos que les gustan las matemáticas tienen mejores promedios y con más frecuencia estudian dos horas. Los docentes son piezas importantes para generar gusto por las matemáticas en los alumnos por tanto es importante que se preparen, que tengan formación en licenciaturas en áreas afines a las matemáticas, y tomen cursos de actualización pedagógica relacionados con las matemáticas, para que tengan argumentos suficientes para enseñar y explicar los conceptos en esos niveles y divulgar sus aplicaciones.

Palabras claves: docencia, alumnos nivel medio superior, inferencia estadística, gusto por las matemáticas, actitud de los docentes y alumnos.

Modalidad: Ponencia.

1. Introducción

Cuando uno entra por primera vez a un salón de clase de los primeros semestres de las carreras de ciencias, tiene una incertidumbre sobre cómo será el curso para esos alumnos, si tendrán la fuerza de voluntad suficiente para aguantar la lluvia de conocimiento que se avecina y de lograr desarrollar en ellos el nuevo razonamiento lógico matemático. Además, si son capaces de darse tiempo para asimilar los conocimientos y dar buenos resultados en sus tareas y exámenes. Tiempo que muchas veces ellos no están dispuestos a tolerar por sus prisas de vivir todo lo nuevo en sus vidas. Los alumnos juegan un papel importante en esta etapa porque pondrán a prueba todo su conocimiento y su aguante para tolerar sus respuestas que al principio son muy desalentadores.

Benemérita Universidad Autónoma de Puebla, México. ¹hreyes@cfm.buap.mx,
²djaz_09@hotmail.com, ³guadalupe_yam@hotmail.com, ⁴hcs@cfm.buap.mx,
⁷vvazquez@cfm.buap.mx.
Universidad Autónoma de Guerrero, México. ⁵fgodinezj@gmail.com,
⁶arizahfj@colpos.mx.

En todo el sistema de enseñanza las matemáticas han ocupado siempre un papel privilegiado y despiertan sentimientos encontrados. La gran mayoría mantiene hacia ellas una mezcla de respeto, formada durante los años escolares y producto de no haber sido capaces de dominarlas sino de sentirse dominados por ellas, para otros, pocos, son lo más bello del mundo y las aman con pasión. Como ya es conocido, la carrera de Matemáticas, en distintos países no tiene la misma demanda que las carreras del área de sociales, administrativas u otras. Por lo que se pretende indagar cuales son los factores que influyen en que a los alumnos de nivel medio superior de la Benemérita Universidad Autónoma de Puebla (BUAP) les gusten las matemáticas. En la prueba Enlace, a nivel nacional los alumnos de nivel bachillerato tienen un nivel insuficiente o elemental en habilidad matemática. Para Puebla, en el 2010 un 42.4% de los estudiantes tuvieron un dominio insuficiente en esta habilidad [7]. Existen muchos factores que contribuyen a que la Educación en México este disminuyendo el nivel educativo con respecto a años anteriores. Entre estos factores tenemos al desempleo, la falta de mejores oportunidades de trabajo, carga excesiva de trabajo, corrupción en las instituciones, modas educativas adoptadas por exigencias de organismos internacionales, estrés en profesores y alumnos, etc. Un factor importante es el docente, quien interacciona frecuentemente con los alumnos motivándolos o coartando sus gustos con sus técnicas de enseñanza. Sabemos que ha ido creciendo el número de docentes con licenciatura en matemáticas que imparten cursos en nivel medio superior, sin embargo son insuficientes para dar apoyo a todas las preparatorias [1]. Pero si el docente no cuenta con experiencia o no tiene cursos de actualización dirigidos a la enseñanza de las matemáticas, puede ser un "reto" acercar a los alumnos al gusto de las matemáticas (pero no imposible!).

Esta investigación es un ejercicio práctico en cuanto al análisis de temas relacionados con la matemática y el uso de la estadística como una herramienta para plantear hipótesis de trabajo y de investigación, lo cual nos lleva a reflexionar sobre lo que ocurre en los alumnos cuando llegan a nuestros espacios para realizar un cambio que dé mejores resultados con los alumnos que desean ingresar a una carrera de ciencias. Debemos participar en la divulgación, enseñanza y vinculación de las matemáticas en los centros de menor nivel educativo y mostrar nuestras áreas con más herramientas de apoyo didáctico (uso de software, platicas de divulgación por expertos, hacer concursos interesantes para cada nivel del conocimiento, cursos de actualización para profesores, etc.).

Objetivos:

1. Encontrar cuales son algunos de los factores que influyen en el gusto por las matemáticas de los alumnos de algunas de las preparatorias de la Benemérita Universidad Autónoma de Puebla.
2. Encontrar las actitudes de los alumnos cuando llevan materias como las matemáticas.

2. Materiales y Métodos

La población objetivo son los alumnos de las escuelas preparatorias de la Ciudad de Puebla cercanas a la Ciudad Universitaria pertenecientes a la BUAP; la Preparatoria Benito Juárez, la Enrique Cabrera, la Calderón y la Emiliano Zapata, tomando los turnos matutino y vespertino. Se elaboró un cuestionario que se validó en un piloteo previo con alumnos con las características que se deseaban. El cuestionario final tiene 13 preguntas de opción múltiple y los datos fueron analizados usando el paquete estadístico SPSS [9].

2.1 Modelo de Regresión Logística.

Sea Y una variable aleatoria binaria, esto es, Y solo tiene dos categorías o resultados: éxito o fracaso ($Y = 1, Y = 0$). Es natural suponer que Y tiene una distribución Bernoulli con probabilidad de éxito $\pi_i = P(Y_i = 1|x_i)$ y de fracaso $1 - \pi_i = P(Y_i = 0|x_i)$, donde $x_i^T = (x_{i1}, x_{i2}, \dots, x_{ip})$ es un vector de p variables independientes también conocidas como factores pronóstico, que creemos influyen en la probabilidad de éxito, π_i , donde $i = 1, \dots, N$ son los individuos [8].

El modelo de regresión logística se puede escribir de la siguiente forma:

$$\pi_i = P(Y_i = 1|x_i^T) = \frac{\exp(x_i^T \beta)}{1 + \exp(x_i^T \beta)} \quad (2)$$

donde $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ es el vector de parámetros desconocidos. El momio se define como el cociente de la probabilidad de que ocurra el evento de interés, $Y=1$, y la probabilidad de que no ocurra para los mismos valores de las variables independientes:

$$\frac{\pi_i}{1 - \pi_i} = \exp(x_i^T \beta) \quad (3)$$

Una transformación de π_i que es importante en el estudio de la regresión logística es la transformación logit, que se define como:

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = x_i^T \beta \quad (4)$$

La importancia de la transformación logit es que tiene muchas de las propiedades de un modelo de regresión lineal, es decir, es una función lineal de las variables independientes, y permite que la probabilidad, π_i , este en el rango de valores de 0 y 1 [4]. Además de que la interpretación de los parámetros es más fácil.

Los parámetros β 's se estiman mediante el método de máxima verosimilitud. Este método produce valores de los parámetros desconocidos que maximizan la probabilidad de obtener los datos observados.

Significancia individual de las variables independientes

Al estimar los coeficientes del modelo y para conocer la importancia de cada variable incluida en el modelo, es necesario probar si el parámetro β_i es diferente o igual a cero, lo cual equivale a las hipótesis:

$$H_0: \beta_i = 0 \quad \text{vs} \quad H_a: \beta_i \neq 0 = 0 \quad i = 1, \dots, p.$$

La prueba estadística utilizada para probar la hipótesis nula es la estadística de Wald, la cual se obtiene dividiendo el estimador de máxima verosimilitud del parámetro β_i entre el estimador de su error estándar (EE):

$$W = \frac{\hat{\beta}_i}{EE(\hat{\beta}_i)} \quad (5)$$

El resultado de este cociente bajo la hipótesis nula H_0 , tiene distribución normal estándar, por lo que permite probar la significación estadística del coeficiente β_i , mediante la comparación con una distribución normal estandarizada Z. Es decir, se rechaza la hipótesis H_0 si $P(|Z| > |W|) < p$, donde p es el nivel de significancia.

3. Resultados y Discusión

El análisis se realizó con el paquete estadístico SPSS versión 18. El cuestionario constaba de preguntas de opción múltiple que definen las variables independientes que se dan en la Tabla 1:

VARIABLE INDEPENDIENTES	ETIQUETA ASIGNADA
1.- Preparatoria	Prepa
2.- Semestre	Semestre
3.- Área que se les dificulta	Área
4.- Horas de estudio matemáticas	hrs_estudio
5.- Realizas algún trabajo	Trabajo
6.- Cómo se te hace más fácil aprender	fácil_aprender
7.- Gusta método de enseñanza de su profesor	gusta_método
8.- Confianza profesor	confianza
9.- Cómo apruebas el curso	Aprobar
10.- Promedio general	PromGeneral
11.- Promedio matemáticas	PromMate

Tabla 1. Variables independientes involucradas en el análisis y la etiqueta asignada.

Desafortunadamente no se pudo realizar un muestreo aleatorio, a pesar de ello, se les aplicó el cuestionario a los estudiantes de diferentes grupos y escuelas. La muestra constó

de 365 alumnos de las diferentes preparatorias, de los cuales el 28.3% eran de la Enrique Cabrera, 23.5% de la Benito Juárez, 16.9% de la Emiliano Zapata y 31.4% de la Lázaro Cárdenas.

Las frecuencias de ocurrencia para las variables independientes con su codificación (codif.) de las categorías se presentan en la Tabla 2. Se puede observar que de los alumnos encuestados el 42.7% tiene promedio en matemáticas en 6-7.9, el 27.7% en 8-8.9 y el 29.6% en 9-10; el 29.3%, no estudia ninguna hora matemáticas, el 40.8% estudia 2 horas y el 29.9% estudia más de 2 horas; al 69.3% les gusta como enseña su profesor; el 90.9% aprueban durante el curso; el 79.4% no trabaja; el 80.8% tiene confianza en su profesor y al 56.4% se les hace más fácil aprender individualmente y al 43.5% en equipo.

Valores asignados a los códigos		Frecuencia	Codif. de parámetros	
			(1)	(2)
Semestre	2do semestre	99	0	0
	4to semestre	158	1	0
	6to semestre	108	0	1
Área	Algebra	93	0	0
	Trigonometría	157	1	0
	Cálculo	115	0	1
Hrs_estudio	Ninguna	107	0	0
	2 horas	149	1	0
	Más de 2 horas	109	0	1
PromMate	6-7.9	156	0	0
	8-8.9	101	1	0
	9-10	108	0	1
PromGeneral	6-7.8	40	0	0
	8-8.9	167	1	0
	9-10	158	0	1
Gusta_método	Si	253	1	
	no	112	0	
Aprobar	Durante el curso	332	1	
	Fuera del curso	33	0	
Confianza	Si	295	1	
	No	70	0	
Trabajo	Si	75	1	
	No	290	0	
Fácil-aprender	Individualmente	206	0	
	Por equipo	159	1	

Tabla 3. Resumen de los códigos usados para las variables, en donde los números en rojo, se encuentran la categoría con mayor frecuencia.

Modelación de la variable Gusto por las matemáticas

La variable de interés **Gusto por las matemáticas** es dicotómica donde 1 expresa que SI le gusta y 0 es NO. El cuestionario incluye preguntas que se encontraron relacionadas según el criterio de los investigadores y se fundamenta en la literatura revisada. Las variables independientes con que se pretende explicar la variable respuesta son las que se muestran en la Tabla 2. El modelo de regresión logística es el apropiado para modelar la probabilidad de que a un estudiante le guste las matemáticas.

Debido a que son muchas variables independientes, se usó el procedimiento de selección de variables “hacia adelante” usando el criterio de Wald, las variables elegidas por este método, después de cuatro pasos, son *promMate*, *gusta-método*, *fácil-aprender* y *hrs-estudio* (Tabla 4).

Paso 1	Mejora			Modelo			% de clases correctas	variables
	Chi-cua	Gl.	Sig.	Chi-cua	Gl.	Sig.		
1	57.70	2	.000	57.78	2	.00	69.6	promMate
2	10.87	1	.001	68.66	3	.00	69.6	gusta-método
3	7.80	1	.005	76.46	4	.00	69.6	fácil-aprender
4	8.06	2	.018	84.53	6	.00	69.6	hrs-estudio

Tabla 4. Resumen de las iteraciones para encontrar el modelo.

Los parámetros estimados del modelo seleccionado se presentan en la Tabla 5.

	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
hrs_estudio			7.925	2	.019			
hrs_estudio(1)	-.775	.292	7.053	1	.008	.460	.260	.816
hrs_estudio(2)	-.688	.312	4.864	1	.027	.503	.273	.926
facil_aprender(1)	.637	.241	6.967	1	.008	1.891	1.178	3.035
gusta_metodo(1)	-.833	.261	10.136	1	.001	.435	.261	.726
PromMate			44.487	2	.000			
PromMate(1)	-1.330	.284	21.946	1	.000	.264	.152	.461
PromMate(2)	-1.887	.305	38.213	1	.000	.152	.083	.276
Constante	1.436	.330	18.895	1	.000	4.203		

Tabla 5. Variables finales en el modelo de regresión logística.

En la columna **Sig** (nivel de significancia) de la Tabla 5 se muestra el p-valor (en negritas) de $H_0: \beta_i = 0$, esto es, la hipótesis de que la variable respectiva no contribuye individualmente a explicar la probabilidad de que al alumno le gusten las matemáticas. Como se puede observar, para las cuatro variables se rechaza la hipótesis $H_0: \beta_i = 0, i=1, 2$,

3, 4 y por tanto las cuatro variables contribuyen individualmente a explicar la probabilidad de que al alumno le gusten las matemáticas

En la Tabla 6 se muestra una prueba Chi-Cuadrado que evalúa la hipótesis nula de que los coeficientes (β 's) de todos los términos (excepto la constante) incluidos en el modelo son cero. El estadístico Chi Cuadrado para este contraste es la diferencia entre el valor de $-2 \log$ de la verosimilitud (LL) para el modelo solo con la constante y el valor de $-2 \log$ de la verosimilitud para el modelo actual:

$$\text{Chi-cuadrado} = (-2 \text{ LL}_{\text{modelo constante}}) - (-2 \text{ LL}_{\text{modelo actual}}) = 415.396$$

Este valor indica que se rechaza la hipótesis de los cuatro parámetros son cero simultáneamente, o lo que es lo mismo, que las variables independientes *promMate*, *gustamétodo*, *fácil-aprender* y *hrs-estudio* contribuyen a explicar la probabilidad de que a un estudiante en la población estudiada le guste las matemáticas.

Paso	-2log verosimilitud	R ² de Cox y de Snell	R ² de Nagelkerke
4	415.396	.207	.277

Tabla 6. Resumen del modelo seleccionado.

En las Tabla 6 también se muestran dos medidas de bondad de ajuste del modelo resultante, R² de Cox y de Snell y R² de Nagelkerke. El R cuadrado de Cox y Snell es un coeficiente de determinación generalizado estima la proporción de varianza de la variable dependiente explicada por las variables independientes y su valor oscila entre 0 y 1 pero siempre tiene un valor máximo inferior a 1, incluso para un modelo "perfecto".

El R² de Nagelkerke es una versión corregida de la R cuadrado de Cox y Snell y corrige la escala del estadístico para cubrir el rango completo de 0 a 1. En nuestro caso, R² de Cox y de Snell= 0.207 lo que indica que el 20.7% de la variación del Gusto por las matemáticas es explicada por las variables incluidas en el modelo. Por otro lado, R² de Nagelkerke =0.277, lo que indica que el 27.7% de la variación del Gusto por las matemáticas es explicada por las variables incluidas en el modelo. Estos valores no indican que el modelo sea malo, son más útiles cuando se comparan con modelos competentes.

Un uso importante del modelo es clasificar a los estudiantes en uno de los dos grupos: Si les gustan las matemáticas o No les gustan las matemáticas. Esto se obtiene sustituyendo los valores de las variables independientes en el modelo ajustado, lo que produce una probabilidad estimada que permite predecir si el estudiante esta en el grupo uno o dos. El software automáticamente emplea un punto de corte de 0.5. Esto significa que aquellos estudiantes con probabilidad estimada < 0.5 se clasifican como ESTADO = 0 (No les gusta las matemáticas), mientras que si la probabilidad estimada es ≥ 0.5 se clasifican como ESTADO = 1 (Si les gusta las matemáticas). Con el modelo final han sido clasificados correctamente 80.1% de los estudiantes que les gustan las matemáticas y un 56.0% de los estudiantes que no les gustan las matemáticas. El porcentaje global de los estudiantes clasificados correctamente es de 69.6%.

	Pronosticado
--	--------------

			Gusto		
			No	Si	
Paso 4	Gusto	No	165	41	80.1
		Si	70	89	56.0
Porcentaje global					69.6

Tabla 7. Tabla de clasificación.

Si se usa la siguiente notación:

- $X_{11}=1$ si el alumno estudia 2 horas y 0 si no es así.
- $X_{12}=1$ si el alumno estudia más de 2 horas y 0 si no es así.
- X_2 : si al alumno se le hace más fácil aprender en grupo y 0 si individualmente.
- X_3 : si al alumno le Gusta método de enseñanza de su profesor y 0 si no es así.
- $X_{41}=1$ si el Promedio Matemáticas del alumno esta en 8-8.9 y 0 si no es así.
- $X_{42}=1$ si el Promedio Matemáticas del alumno esta en 9-10 y 0 si no es así.

Entonces el modelo ajustado es:

$$\log\left(\frac{\hat{\pi}_i}{1-\hat{\pi}_i}\right) = 1.436 - 0.775X_{11} - 0.688X_{12} + 0.637X_2 - 0.833X_3 - 1.33X_{41} - 1.887X_{42}$$

Los parámetros estimados permiten hacer las siguientes afirmaciones:

Que a un alumno le gusten las matemáticas es 1.891 veces más probable de ocurrir entre los alumnos que se les hace más fácil aprender en grupo que entre los alumnos que se les hace más fácil aprender individualmente.

Que a un alumno le guste las matemáticas es 0.435 veces más probable de ocurrir entre los alumnos que si les gusta como enseña su profesor que entre los alumnos que no les gusta.

Que a un alumno le gusten las matemáticas es 0.460 veces más probable de ocurrir entre los alumnos que estudian 2 horas que entre los alumnos que no estudian. Que a un alumno le gusten las matemáticas es 0.503 veces más probable de ocurrir entre los alumnos que estudian más de 2 horas que entre los alumnos que no estudian.

Que a un alumno le gusten las matemáticas es 0.264 veces más probable de ocurrir entre los alumnos con promedio entre 8 y 8.9 que entre los alumnos con promedio entre 6 y 7.8.

Que a un alumno le gusten las matemáticas es 0.152 veces más probable de ocurrir entre los alumnos con promedio entre 9 y 10 que entre los alumnos con promedio entre 6 y 7.8.

4. Conclusiones

Los resultados revelan que el promedio que tienen los alumnos tienen en matemáticas afecta el gusto por ellas. También muestran que las horas que le dedican a estudiar son solo de 2 horas a la semana y esto es consecuencia, de lo primero. Los resultados indican que el profesor es un factor muy relevante en que a los alumnos le gusten las matemáticas o no, su forma de enseñanza y esto se debe a que la mayoría de los profesores que imparten clases

de matemáticas no tienen la licenciatura en esta y algunos no tienen cursos de actualización dirigidos a la enseñanza de las matemáticas, por lo que esto hace que a los alumnos no les gusten las matemáticas. Los resultados de este trabajo podrían ayudar a que los profesores de las diferentes preparatorias encuestadas observen que es lo que pueden hacer para que a los alumnos les gusten las matemáticas como realizar sus clases de tal manera que para ellos no sean aburridas y les interese para así obtener mejores promedios en esta materia y dedicarle más horas a estudiarla.

Sabemos que existen muchos factores que contribuyen a la Educación en México este cada vez mal comparado con años anteriores y que hay factores multivariados que contribuyen a esta situación. Algunos de ellos, son que en los niveles de educación básicos, a los alumnos no se les enseña a aprender y razonar, sólo a aprenderse de memoria los temas, por lo cual no tienen hábitos de estudio. En la sociedad existen distracciones muy fuertes, como son los juegos de computadora, televisión, relaciones sociales demandantes que los alejan del estudio y la dedicación. También en las universidades no han creado espacios de difusión y vinculación en niveles de educación menores para concientizar a la sociedad de la existencia y creación de lugares en donde se promueva la ciencia y la tecnología para mejorar nuestro nivel de vida a futuro.

5. Referencias

- [1] Ávila L. Reyes H. & Ibarra M. (2002) Difusión de las Matemáticas en las Prepas, XVIII Congreso de Matemáticas.
- [2] Agresti, A. (1990). Categorical Data Analysis. New York: John Wiley & Sons.
- [3] Cochran, W. (1980). Técnicas de muestreo. Mexico: CECSA.
- [4] Hosmer David W. & Stanley Lemeshow (2000). Applied Logistic Regression. New York: John Wiley & Sons.
- [5] Ibarra Contreras M., Reyes Cervantes H., Hernández Guerra & Linares G. (2009). Análisis Estadístico de Algunos Factores que Afectan el Proceso de Enseñanza Aprendizaje en la FCFM-BUAP. Participación de la mujer en la ciencia, Volumen (6).
- [6] Ibarra Contreras M., Reyes Cervantes H., Hernández Guerra & Linares G. (2010). Proceso de Enseñanza Aprendizaje en la BUAP-2. Participación de la mujer en la ciencia, Volumen (7).
- [7] Prueba Enlace (2010). www.enlace.sep.gob.mx/ms/.
- [8] Rao Toutenburg & Shalabh Heumann. (2007). Linear Models and Generalizations Least Squares and Alternatives. Springer.
- [9] SPSS (2010). IBM SPSS Statistics 19 para Windows.