

**LA REPITENCIA EN LA FACULTAD DE CIENCIAS EXACTAS Y
NATURALES DE LA UDENAR APLICANDO REGRESION LOGISTICA**

LEONEL DELGADO ERASO
Código 832232

UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ, 2009

**LA REPITENCIA EN LA FACULTAD DE CIENCIAS EXACTAS Y
NATURALES DE LA UDENAR APLICANDO REGRESION LOGISTICA**

**LEONEL DELGADO ERASO
Código 832232**

**Trabajo de grado presentado para optar al título de
Especialista en Estadística**

**Directora:
MARTHA PATRICIA BOHORQUEZ CASTAÑEDA**

**UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ, 2009**

DEDICATORIA

A mi esposa Yalila por su comprensión y apoyo, a mi hija Juliana que me regalo su tiempo para cumplir con esta meta, a mi hijo Juan Esteban que es mi ángel de la guarda que desde el cielo me llena de bendiciones, a mis familiares, amigos, compañeros y profesores.

Leonel

CONTENIDO

	Pág.
Introducción	10
1 Marco Teórico	12
1.1 La repitencia	12
1.2 Evaluación Académica	12
1.3 Régimen académico	12
1.4 El modelo de regresión logístico múltiple	13
1.5 Ajuste del modelo de regresión logístico múltiple	14
1.6 Prueba para la significancia del modelos	16
1.7 Estimación de intervalos de confianza	17
1.8 Variables independientes dicotómicas	19
1.9 Determinación del ajuste del modelo	22
1.9.1 Área bajo la curva ROC	22
2 Metodología	23
2.1 Elaboración y aplicación de la encuesta	23
2.2 Procesamiento de la información	23
3 Presentación de los resultados	25
3.1 Estimación del modelo de regresión	25
3.2 Pruebas de bondad y ajuste	27
3.3 Validación del modelo	29
3.4 Área bajo la curva ROC	29
4 Conclusiones y Recomendaciones	31
4.1 Conclusiones	31
4.2 Recomendaciones	31
Bibliografía	32
Anexos	33

LISTA DE TABLAS

		Pág.
1	Valores del modelo de regresión logística cuando la variable independiente es dicotómica	20
2	Distribución de estudiantes según la variable dependiente	23
3	Estimación de los coeficientes	25
4	Probabilidades de repitencia	26
5	Pruebas de bondad y ajuste	28
6	Prueba de Hosmer y Lemeshow	27
7	Resumen del modelo	28
8	Tabla de Clasificación	29

LISTA DE FIGURAS

1	Area bajo la curva ROC	Pág. 30
---	-------------------------------	-------------------

RESUMEN

En este trabajo se presentan los resultados de un estudio realizado en la Facultad de Ciencias Exactas y Naturales de la Universidad de Nariño aplicando el modelo de regresión logística para predecir la probabilidad de repitencia de al menos dos materias durante la permanencia de un estudiante en la universidad. Los datos se obtuvieron de una encuesta aplicada a los estudiantes de la facultad con el fin de determinar los factores que inciden en la repitencia de materias. Se presentan datos correspondientes a 367 estudiantes matriculados en el periodo A de 2008¹, los cuales hacen parte de la población de estudiantes de las carreras de Licenciatura en Matemáticas, Física, Química y Biología con una edad media de 22 años y donde se presenta un alto índice de repitencia de materias.

Descriptores: repitencia, factores, predecir, regresión logística múltiple

¹ Febrero a Junio

SUMMARY

The results of a stated study in the Faculty of Natural Sciences at Nariño University is shown in this work, applying the model of logistic regression to predict the probability of repetitiveness of at least two subjects during the stay of a student at the university. The scores were taken from a survey applied to the students of the faculty, aiming to determine the factors that had to do with the repetitiveness of the subjects. The correspondent data of 367 registered students in the A 2008 period, that belong to the population of students of the Mathematics degree, Physics, Chemistry and Biology twenty-two year old students were the subject repetitiveness was very high.

Keywords: repetitiveness, factors, predict, multiple logistic regression

INTRODUCCION

Además de la falta de acceso a la escuela y la deserción, la repitencia constituye uno de los mayores problemas de los sistemas escolares contemporáneos. A pesar de que, a nivel internacional, la repitencia tiende a verse como un fenómeno típicamente latinoamericano - región, en efecto, con índices muy elevados de repitencia y en la que ésta viene mereciendo atención en los últimos años- se trata en verdad de un fenómeno que, de manera abierta, afecta a la mayor parte de sistemas escolares en el mundo².

Este autor también afirma que la repitencia es la “solución” interna que ha encontrado el sistema escolar para lidiar con el problema del no-aprendizaje o de la mala calidad de dicho aprendizaje. Analizar las fuentes y la naturaleza de la repitencia es, de este modo, analizar la misión misma de la universidad, incluyendo la gama de variables y procesos que inciden sobre el aprendizaje en el medio escolar, su calidad, contextos y resultados.

Los directivos superiores asociados a la actividad docente de instituciones de educación superior asumen que la mayoría de los estudiantes que ingresan a una carrera universitaria lo hacen con la intención de permanecer en ella hasta su graduación y no pueden aceptar que esos estudiantes, en altos porcentajes, puedan no tener éxito en varias de las asignaturas que cursan en los diferentes semestres, debiendo prolongar sus estudios hasta casi duplicar sus tiempos normales para titulación. A su vez, la mayoría de ellos asume que las altas tasas de reprobación estudiantil se deben principalmente a métodos de enseñanza y evaluación inadecuados, contenidos complejos, por su nivel de abstracción, contenidos innecesarios (a su juicio) para el futuro ejercicio profesional y mal diseño de la estructura de prerrequisitos entre asignaturas³.

El mismo autor asegura que los encargados de realizar la actividad docente asumen que las altas tasas de reprobación y repitencia se deben a fallas de formación de los estudiantes en enseñanza media, carencia de responsabilidad y de perseverancia del estudiante, problemas de inteligencia y carencia de aptitudes verbales y/o matemáticas.

Ante la situación descrita en forma general se hace necesario definir las variables que influyen con la repitencia de asignaturas y por ende, en el mal rendimiento estudiantil

²Torres, Rosa María. Repetition: A Major Obstacle to Education for All, en Education New, No 15. Nueva York: Unicef, Abril 1995. p1a5. Disponible en internet, URL: www.mineducacion.gov.co/1621/articles-85774_archivo_pdf.pdf.

³ Ramírez Peradotto, Pedro. Reyes Rocabado, Jimmy. Escobar Flores, Carlos. Duarte Vargas, Juan. Una Aplicación del Modelo de Regresión Logística en la Predicción del Rendimiento Estudiantil. Antofagasta, Chile, 2004. p1. Disponible en internet, URL: www.scielo.cl/scielo.php?pid=S0718-07052007000200006&script=sci_arttext - 54k -.

durante los semestres de su carrera y mediante un modelo de regresión logística relacionarlas con la repitencia de tal forma que se pueda predecir con alta precisión antes de que este ocurra, y tomar medidas preventivas.

Rosa María Torres concluye que la repitencia es un indicador claro de la disfuncionalidad e ineficiencia interna del sistema escolar, la sociedad en general y la comunidad educativa en particular (profesores, padres, estudiantes, directores, decisores de políticas a los distintos niveles) tienden a aceptar la repitencia como "natural", como un componente inherente y hasta inevitable de la vida escolar.

En la Facultad de Ciencias Exactas y Naturales de la Universidad de Nariño, la repitencia de materias se convierte en un problema, cuando el número de estudiantes aumenta semestre tras semestre, hasta el punto de volverse inmanejable con aquellos estudiantes que repiten una asignatura por tercera o cuarta vez, a los cuales se les debe hacer un seguimiento académico con ayuda profesional en lo referente a sus aspectos psicosociales. (artículo 87 del estatuto estudiantil, párrafo).

De acuerdo con los anteriores argumentos, el presente trabajo pretende proponer un modelo que permita estimar la probabilidad de que un estudiante repita dos o más de dos asignaturas en su carrera.

1. MARCO TEORICO

1.1 LA REPITENCIA

La repitencia se entiende como la acción de cursar reiterativamente un ciclo lectivo o bloque de asignaturas, o bien, cada asignatura cuando se trata de un currículum flexible.

1.2 EVALUACIÓN ACADÉMICA

La evaluación académica es el conjunto de procesos y actividades mediante los cuales se valora el grado o medida en que el estudiante:

- Alcanzó los objetivos de la asignatura, núcleo temático o actividad académica.
- Desarrolló su habilidad para aprender a aprender
- Fundamentó y desarrolló los valores precolonizados por la Universidad (artículo 89 del est. estudiantil).

La calificación en todas las asignaturas serán numéricas de cero(0) a cinco (5), en unidades y décimas. (artículo 97 del est. estudiantil).

En todos los programas de pregrado de la Universidad, la nota aprobatoria mínima es de tres (3.0). (artículo 98 del est. estudiantil).

Cuando un estudiante no presente ninguna de las pruebas programadas en una asignatura en la cual se matriculó y no la canceló oportunamente, la Oficina de Registro Académico la registrará con una calificación de cero (0.0). (artículo 102 del est. estudiantil).

1.3 RÉGIMEN ACADÉMICO

El estudiante que haya perdido por dos veces la misma asignatura podrá matricular en el siguiente período académico o año, ésta más aquellas que el Comité curricular y de Investigaciones del Departamento le determine, una vez estudie su situación académica. Si una vez cursada, nuevamente la reprueba, el estudiante tiene dos opciones:

- Si la asignatura se ofrece en el siguiente período académico o año, debe cursar únicamente ésta y de no aprobarla perderá definitivamente el derecho a continuar estudios en el programa.
- Si la asignatura no se ofrece en el siguiente período académico o año, podrá matricular hasta un máximo de tres asignaturas autorizadas por el Comité Curricular y de Investigaciones. En el siguiente período o año, obligatoriamente deberá matricular

solamente la asignatura perdida por tercera vez, de no aprobarla perderá definitivamente el derecho a continuar en el programa.

1.4 EL MODELO DE REGRESIÓN LOGÍSTICO MÚLTIPLE

Para el cálculo de la probabilidad de que un estudiante repita dos o más de dos materias durante su carrera universitaria, se definirá el modelo de regresión logística múltiple, de acuerdo con Hosmer y Lemeshow⁴:

Consideremos p variables independientes denotadas por el vector $x' = (x_1, x_2, \dots, x_p)$, La probabilidad condicional que la respuesta este presente será denotado por $P(Y=1|x) = \pi(x)$. El logit del modelo de regresión logística múltiple es dado por la ecuación

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_p x_p \quad (1)$$

en este caso el modelo de regresión logístico es

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}} \quad (2)$$

Si algunas de las variables independientes son discretas, las variables nominales tales como sexo, raza, tipo de colegio, etc., es inadecuado incluirlas en el modelo como si fueran variables de intervalo. Los números usados para representar los varios niveles de estas variables nominales son simplemente identificadores, y no tienen ninguna significación numérica. En esta situación el método de opción es utilizar una colección de variables diseño (o de *variables dummy*).

Generalmente si una variable nominal tiene k posibles valores, se construyen $k-1$ variables dummy. Para ilustrar la notación usada para las variables diseño, suponga que la j -ésima de la variable independiente de x_j tiene k_j niveles. Las k_j-1 variables dummy se denotan como D_{jl} y los coeficientes para estas variables dummy serán denotados como B_{jl} , donde $l = 1, 2, \dots, k_j-1$. Así, el logit para un modelo con p variables y la j -ésima variable siendo discreta serían

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \sum_{l=1}^{k_j-1} \beta_{jl} D_{jl} + \beta_p x_p$$

y el modelo de regresión logística múltiple es como lo indica la ecuación (2)

⁴ Hosmer, David W. Lemeshow, Stanley. Applied Logistic Regression. New York, Second Edition. Interscience Publication, John Wiley & Sons, Inc, 2000. p. 31.

1.5 AJUSTE DEL MODELO DE REGRESIÓN LOGÍSTICO MÚLTIPLE

Hosmer y Lemeshow⁵, describe el procedimiento de la siguiente manera:

Asumamos que tenemos una muestra de n observaciones independientes (x_i, y_i) , $i=1,2,3,\dots, n$. Ajustar el modelo requiere que obtengamos las estimaciones del vector $\beta' = (\beta_0, \beta_1, \dots, \beta_p)$.

Una manera de expresar la función de verosimilitud para los pares (x_i, y_i) viene dada por la expresión

$$\pi(x_{ij})^{y_i} [1 - \pi(x_i)]^{1 - y_i} \quad (3)$$

Puesto que las observaciones son independientes, la función de verosimilitud es obtenida como el producto de los términos dados en la expresión (1.2) como sigue

$$l(\beta) = \prod_{i=1}^n \pi(x_{ij})^{y_i} [1 - \pi(x_i)]^{1 - y_i}. \quad (4)$$

El principio de máxima verosimilitud indica que utilizamos como nuestra estimación de β el valor que maximiza la expresión en la ecuación (1.3). Sin embargo, es más fácil matemáticamente trabajar con el logaritmo de la ecuación (1.3). Esta expresión, del logaritmo de verosimilitud, esta definida como

$$L(\beta) = \ln[l(\beta)] = \sum_{i=1}^n \{y_i \ln[\pi(x_{ij})] + (1 - y_i) \ln[1 - \pi(x_i)]\} \quad (5)$$

Para encontrar el valor de β que maximice $L(\beta)$ diferenciamos $L(\beta)$ con respecto a β_j y hacemos las expresiones resultantes iguales a cero. Estas ecuaciones, conocidas como las ecuaciones de máxima verosimilitud, son:

$$\sum_{i=1}^n [y_i - \pi(x_i)] = 0$$

y

$$\sum_{i=1}^n x_{ij} [y_i - \pi(x_i)] = 0.$$

Para $j = 1, 2, \dots, p$

⁵ Ibid., p. 33

Denotemos $\hat{\beta}$ la solución a estas ecuaciones. Así, el valor ajustado para el modelo de regresión logística múltiple es $\hat{\pi}(x_i)$, el valor de la expresión en la ecuación (2) calculada usando, $\hat{\beta}$, y x_i .

El método de estimar las varianzas y las covarianzas de los coeficientes estimados viene de la teoría de la valoración de la máxima verosimilitud. Esta teoría indica que los estimadores están obtenidos de la matriz de las segundas derivadas parciales de la función log de verosimilitud. Estas derivadas parciales tienen la forma general siguiente

$$\frac{\partial^2 L(\beta)}{\partial \beta_j^2} = \sum_{i=1}^n x_{ij}^2 \pi_i (1 - \pi_i) \quad (6)$$

y

$$\frac{\partial^2 L(\beta)}{\partial \beta_j \partial \beta_l} = - \sum_{i=1}^n x_{ij} x_{il} \pi_i (1 - \pi_i) \quad (7)$$

Para $j, l = 1, 2, \dots, p$; donde π_i denota $\pi(x_i)$. Luego la matriz $(p+1) \times (p+1)$ que contiene los términos negativos dados en las ecuaciones (6) y (7) se denoten como $\mathbf{I}(\beta)$. Esta matriz se llama la *matriz de información observada*. Las varianzas y las covarianzas de los coeficientes estimados se obtienen de la inversa de esta matriz que denotamos como $\text{var}(\beta) = \mathbf{I}^{-1}(\beta)$. Excepto en casos muy especiales no es posible anotar una expresión explícita para los elementos en esta matriz. Por lo tanto, utilizaremos la notación $\text{Var}(\beta_j)$ para denotar el j -ésimo elemento de la diagonal de esta matriz, que es la varianza de $\hat{\beta}_j$, y de la $\text{Cov}(\beta_i, \beta_l)$ para denotar un elemento arbitrario fuera de la diagonal, que es la covarianza de $\hat{\beta}_j$ y del $\hat{\beta}_l$. Los estimadores de las varianzas y de las covarianzas, que serán denotadas por $\hat{\text{Var}}(\hat{\beta})$, son obtenidos evaluando la $\text{var}(\beta)$ en el $\hat{\beta}$. Utilizaremos la $\hat{\text{Var}}(\hat{\beta}_j)$ y la $\hat{\text{Cov}}(\hat{\beta}_j, \hat{\beta}_l)$, $j, l = 1, 2, \dots, p$ para denotar los valores en esta matriz.

En general, tendremos que utilizar solamente los errores estándar estimados de los coeficientes estimados, que denotaremos como

$$\hat{SE}(\hat{\beta}_j) = [\hat{\text{Var}}(\hat{\beta})]^{1/2} \quad (8)$$

para $j = 0, 1, 2, \dots, p$. Utilizaremos esta notación en el desarrollo de los métodos para la prueba del coeficiente y la estimación de intervalo de confianza.

Una formulación de la matriz de información que será útil al discutir el ajuste del modelo y la asignación del ajuste es $\hat{I}(\hat{\beta}) = \mathbf{X}'\mathbf{V}\mathbf{X}$, donde \mathbf{X} es una matriz n por $p+1$ que contiene los datos para cada individuo, y \mathbf{V} es una matriz diagonal n por n con elemento genérico $\hat{\pi}_i(1-\hat{\pi}_i)$. Es decir, la matriz \mathbf{X} es

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

y la matriz \mathbf{V} es

$$\mathbf{V} = \begin{bmatrix} \hat{\pi}_1(1-\hat{\pi}_1) & 0 & \dots & 0 \\ 0 & \hat{\pi}_2(1-\hat{\pi}_2) & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \hat{\pi}_n(1-\hat{\pi}_n) \end{bmatrix}$$

1.6 PRUEBA PARA LA SIGNIFICANCIA DEL MODELO⁶

Una vez que hemos ajustado un modelo de regresión logístico múltiple particular, comenzamos el proceso de evaluación del modelo. El primer paso en este proceso es generalmente determinar la significación de las variables en el modelo. La prueba del cociente de verosimilitud para la significación total de los p coeficientes se basa en la estadística G dada en la ecuación

$$G = -2\ln \left[\frac{\binom{n_1}{n} \binom{n_0}{n}}{\prod_{i=1}^n \hat{\pi}_i^{y_i} (1-\hat{\pi}_i)^{(1-y_i)}} \right]$$

⁶ Ibid., p. 36

Bajo la hipótesis nula que los p coeficientes para las variables independientes en el modelo son iguales a cero, la distribución de G será ji-cuadrado con p grados de libertad.

Antes de concluir que cualquiera o todos los coeficientes son diferentes de cero, podemos mirar la estadística de *prueba univariante de Wald*,

$$W_j = \frac{\hat{\beta}_j}{\widehat{SE}(\hat{\beta}_j)}$$

Si el objetivo es obtener el mejor modelo ajustado mientras que reduzca al mínimo el número de parámetros, el paso lógico siguiente es ajustar un modelo reducido que contiene solamente esas variables probablemente significativas, y lo compara con el modelo completo que contiene todas las variables.

Debido a los grados de libertad múltiples debemos tener cuidados en nuestro uso de las estadísticas de Wald (W) para determinar la significación de los coeficientes. Por ejemplo, si las estadísticas de W para ambos coeficientes exceden de 2, podríamos concluir que las variables dummy son significativas. Alternativamente, si un coeficiente tiene una estadística de W de 3.0 y la otra un valor de 0.1, entonces nosotros no podemos estar seguros sobre la contribución de la variable al modelo

En el modelo logístico múltiple la prueba de Wald se obtiene del cálculo siguiente del vector-matriz:

$$W = \hat{\beta}' \left[\widehat{Var}(\hat{\beta}) \right]^{-1} \hat{\beta} = \hat{\beta}' (X'VX) \hat{\beta},$$

el cuál será distribuido como ji-cuadrado con $p+1$ grados de libertad bajo la hipótesis que cada uno de los coeficientes $p+1$ es igual a cero. Las pruebas para los p coeficientes son obtenidas eliminando $\hat{\beta}_0$ de $\hat{\beta}$ y la fila y columna de $(X'VX)$. Puesto que la evaluación de esta prueba requiere la capacidad de realizar operaciones de la vector-matriz y obtener $\hat{\beta}$, no hay aumento sobre la prueba del cociente de probabilidad de la significación del modelo.

1.7 ESTIMACION DE INTERVALOS DE CONFIANZA⁷

Los límites al $100(1-\alpha)\%$ de un intervalo de confianza para los coeficiente son

⁷ Ibid., p. 40

$$\hat{\beta}_j \pm z_{1-\alpha/2} \hat{SE}(\hat{\beta}_j), \quad (9)$$

y para el término constante son:

$$\hat{\beta}_0 \pm z_{1-\alpha/2} \hat{SE}(\hat{\beta}_0), \quad (10)$$

donde $z_{1-\alpha/2}$ es el punto superior $100(1-\alpha/2)\%$ de la distribución normal estándar y $\hat{SE}(\cdot)$ denota un estimador del error estándar del respectivo parámetro estimado.

De la ecuación (1) se obtiene una expresión general para la estimación del logit para un modelo que contiene p variables independentes

$$\hat{g}(x) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_p x_p. \quad (11)$$

Una manera alternativa de expresar el estimador del logit en (11) en forma vectorial es, $\hat{g}(x) = x' \hat{\beta}$, donde el vector $\hat{\beta}' = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)$ denota el estimador de los $p+1$ coeficientes y el vector $x' = (x_0, x_1, x_2, \dots, x_p)$ representa la constante y un conjunto de valores de las p -variables independientes en el modelo, donde $x_0=1$.

La varianza estimada de la varianza del estimador del logit en (11) es

$$\hat{Var}[\hat{g}(x)] = \sum_{j=0}^p x_j^2 \hat{Var}(\hat{\beta}_j) + \sum_{j=0}^p \sum_{k=j+1}^p 2x_j x_k \hat{Cov}(\hat{\beta}_j, \hat{\beta}_k). \quad (12)$$

Podemos expresar este resultado mucho más fácil usando la expresión matricial para la varianza estimada del estimador de los coeficientes. De la expresión para la matriz de información observada, tenemos que

$$\hat{Var}(\hat{\beta}) = (X' V X)^{-1}. \quad (13)$$

Se sigue de (13) que una expresión equivalente para el estimador en (12) es

$$\hat{Var}[\hat{g}(x)] = x' \hat{Var}(\hat{\beta}) x = x' (X' V X)^{-1} x. \quad (14)$$

Afortunadamente, todos los paquetes de programas informáticos de regresión logística proporcionan la opción para que el usuario cree una nueva variable que contiene los valores estimados de (14) o el error estándar para todos los individuos en el conjunto de datos. Esta característica elimina el cómputo asociada a los cálculos de la matriz en (14) y permite que el usuario calcule fácilmente valores ajustados y estimaciones del intervalo de confianza.

1.8 VARIABLE INDEPENDIENTE DICOTÓMICA⁸

Comenzamos nuestra consideración de la interpretación de los coeficientes de regresión logística con la situación donde la variable independiente es nominal y dicotómica⁹ (es decir, medido en dos niveles). Este caso proporciona la base conceptual para el resto de situaciones.

Asumimos que la variable independiente, x , está codificada como cero o uno. La diferencia en el logit para un individuo con $x=1$ y $x=0$ es

$$g(1) - g(0) = [\beta_0 + \beta_1] - [\beta_0] = \beta_1$$

La deducción algebraica en esta ecuación es algo directa. Se presenta en este nivel de detalle para acentuar que el primer paso en la interpretación del efecto de una variable independiente en un modelo es expresar la diferencia deseada del logit en términos del modelo. En este caso la diferencia del logit es igual al β_1 . Para interpretar este resultado necesitamos introducir y discutir una medida de asociación llamada el *odds ratio*

Los valores posibles de las probabilidades logísticas se pueden observar convenientemente en una tabla 2x2 según las indicaciones de la tabla A. Los odds del resultado que está presente entre individuos con $x=1$ se definen como $\pi(1)/[1 - \pi(1)]$. Similarmente, los odds del resultado que está presente entre individuos con $x=0$ se definen como $\pi(0)/[1 - \pi(0)]$. Los odds ratio, denotado OR, se define como el cociente de las probabilidades para $x=1$ y los odds para $x=0$, y es dado por la ecuación

$$OR = \frac{\frac{\pi(1)}{1 - \pi(1)}}{\frac{\pi(0)}{1 - \pi(0)}} \quad (15)$$

⁸ Ibid, p. 48.

⁹ Para las variables independientes en la regresión logística, no se establece ninguna restricción, pueden ser cuantitativas, tanto continuas como discretas (con dos o más niveles), o cualitativas, nominales u ordinales transformadas a variables dummy.

Substituyendo las expresiones para el modelo de regresión logístico mostrado en la tabla A en (15) obtenemos

$$OR = \frac{\left(\frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}\right) / \left(\frac{1}{1 + e^{\beta_0 + \beta_1}}\right)}{\left(\frac{e^{\beta_0}}{1 + e^{\beta_0}}\right) / \left(\frac{1}{1 + e^{\beta_0}}\right)} = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{(\beta_0 + \beta_1) - \beta_0} = e^{\beta_1}$$

Tabla 1. Valores del modelo de regresión logística cuando la variable independiente es dicotómica.

Variable Respuesta(Y)	Variable independiente	
	x=1	x=0
y=1	$\pi(1) = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}$	$\pi(0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$
y=0	$1 - \pi(1) = \frac{1}{1 + e^{\beta_0 + \beta_1}}$	$1 - \pi(0) = \frac{1}{1 + e^{\beta_0}}$
Total	1.0	1.0

Por lo tanto, para la regresión logística con una variable independiente dicotómica codificada 1 y 0, la relación entre los *odds ratio* y el coeficiente de regresión es

$$OR = e^{\beta_1} \quad (16)$$

Esta relación simple entre el coeficiente y los odds ratio es la razón fundamental por la que la regresión logística ha demostrado ser tal herramienta analítica de gran alcance.

Los odds ratio son una medida de la asociación que ha encontrado amplio uso, especialmente en epidemiología, pues aproxima cuánto es más probable (o inverosímil) para que la respuesta esté presente entre éstos con x=1 que entre éstos con x=0. Por ejemplo, si y denota la presencia o la ausencia de cáncer de pulmón y si x denota si la persona es

fumadora, entonces $\hat{OR} = 2$ estima que el cáncer de pulmón es dos veces tan probable ocurrir entre fumadores que entre no fumadores en la población del estudio.

La interpretación dada para los odds ratio se basa en el hecho de que en muchos casos aproxima una cantidad llamada el riesgo relativo. Este parámetro es igual al cociente $\pi(1)/\pi(0)$. De la ecuación (15) se tiene que los odds ratio se aproximan al riesgo relativo si $[1-\pi(0)]/[1-\pi(1)] \approx 1$. Esto se sostiene cuando $\pi(x)$ es pequeño para ambos $x=1$ y 0 .

Los odds ratio, OR, es generalmente el parámetro de interés en una regresión logística debido a su facilidad de la interpretación. Sin embargo, su estimación, \hat{OR} , tiende a tener una distribución que es sesgada. La asimetría de la distribución muestra que \hat{OR} es debido al hecho que los valores posibles extienden entre 0 y ∞ , con el valor nulo igual a 1 . En teoría, para muestras de tamaños bastante grandes, la distribución de \hat{OR} es normal. Desafortunadamente, este requisito del tamaño de muestra excede típicamente de la mayoría de estudios. Por lo tanto, las inferencias se basan generalmente en la distribución de muestra del $\ln(\hat{OR}) = \hat{\beta}_1$, que tiende a seguir una distribución normal para tamaños de muestra mucho más pequeños. Una estimación al $100(1-\alpha)\%$ del intervalo de confianza (CI) para el odds ratio es obtenida primero calculando los límites de un intervalo de confianza para el coeficiente, β_1 , y después exponenciando estos valores. En general los límites están dados por la expresión

$$\exp[\hat{\beta}_1 \pm z_{1-\alpha/2} SE(\hat{\beta}_1)].$$

Antes de concluir el caso de variable dicotómica, es importante considerar el efecto que la codificación de la variable tiene en el cómputo de la estimación de los odds ratio. En la discusión anterior observamos que los odds ratio estimados fueron $\hat{OR} = \exp(\hat{\beta})$. Esto está correcto cuando la variable independiente se codifica como 0 o 1 . La otra codificación puede requerir que calculemos el valor de la diferencia del logit para la codificación específica usada, y después exponenciar esta diferencia estimar los odds ratio.

Ilustramos estos cálculos detalladamente, pues muestran el método general para calcular estimaciones de los odds ratio en la regresión logística. La estimación del log del odds ratio para cualquier variable independiente en dos diversos niveles, digamos $x=a$ contra $x=b$, es la diferencia entre los logists estimados computada en estos dos valores,

$$\ln[\hat{OR}(a,b) = \hat{g}(x=a) - \hat{g}(x=b) = (\hat{\beta}_0 + \hat{\beta}_1 \times a) - (\hat{\beta}_0 + \hat{\beta}_1 \times b) = \hat{\beta}_1 \times (a-b) \quad (17)$$

La estimación de los odds ratio es obtenida por exponenciación de la diferencia del logit,

$$\hat{OR}(a,b) = \exp[\hat{\beta}_1 \times (a-b)] \quad (18)$$

Observe que esta expresión es igual al $\exp(\hat{\beta}_1)$ solamente cuando $(a-b) = 1$. En (17) y (18) la notación $\hat{OR}(a,b)$ es usada para representar los odds ratio

$$\hat{OR}(a,b) = \frac{\hat{\pi}(x=a)/[1-\hat{\pi}(x=a)]}{\hat{\pi}(x=b)/[1-\hat{\pi}(x=b)]} \quad (19)$$

y cuando $a=1$ y $b=0$ tenemos $\hat{OR} = \hat{OR}(1,0)$.

1.9 DETERMINACIÓN DEL AJUSTE DEL MODELO

1.9.1 AREA BAJO LA CURVA ROC¹⁰

La sensibilidad y la especificidad confían en solo punto de corte para clasificar un resultado de la prueba como positivo. Una descripción más completa de la exactitud de la clasificación es dada por el área bajo la curva ROC¹¹. Traza la probabilidad de detectar la señal verdadera (sensibilidad) y la señal falsa (1-especificidad) para un rango entero de posibles puntos de corte.

El área bajo la curva ROC, que se extiende a partir de la cero a uno, proporciona una medida del modelo tiene una capacidad de *discriminar* entre los sujetos que experimentan el resultado de interés contra los que no lo hagan.

Si nuestro objetivo era elegir un punto de corte óptimo para los propósitos de la clasificación, una pudo seleccionar un punto de corte que maximiza sensibilidad y especificidad.

Como regla general:

- Si $ROC = 0.5$ esto no sugiere ninguna discriminación
- Si $0.7 \leq ROC < 0.8$ esto se considera discriminación aceptable
- Si $0.8 \leq ROC < 0.9$ esto se considera discriminación excelente
- Si $ROC \geq 0.9$ esto se considera discriminación excepcional.

¹⁰ Ibid., p. 160

¹¹ Curva de características operativas = Receiver Operating Characteristic.

2. METODOLOGIA

2.1 ELABORACION Y APLICACIÓN DE LA ENCUESTA

La encuesta (Anexo A) que se aplicó en la Facultad de Ciencias Exactas y Naturales consta de cuatro partes: Identificación, Información académica, Información socio-económica e información personal, en los cuales se indaga en los siguientes aspectos: Rendimiento académico, hábitos de estudio, acompañamiento en el hogar, relaciones personales estudiante-maestro, y los principales determinantes psicosociales como: necesidad de que el estudiante trabaje, desintegración del hogar, preparación del maestro y condiciones mínimas para la investigación.

La encuesta no se aplicó a los primeros semestres, ya que si un estudiante se encuentra repitiendo alguna asignatura se lo registro en el segundo semestre, o al que se encuentre matriculado.

2.2 PROCESAMIENTO DE LA INFORMACION

Se consideró como variable dependiente a la que toma los siguientes valores: 1 si el estudiante ha repetido dos o más de dos asignaturas, en el transcurso de su carrera, 0 si el estudiante no ha repetido ninguna o máximo una asignatura.

Tabla 2. Distribución de estudiantes según la variable dependiente

REPETICION					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	142	38.7	38.7	38.7
	1	225	61.3	61.3	100.0
	Total	367	100.0	100.0	

Las variables independientes o predictoras inicialmente incluidas fueron: *Identificación*(semestre, edad, estrato), *Información académica*(colegio, horas de estudio, razones por las cuales ingreso al programa, bases del bachillerato,..), *Información socio-económica* (sostenimiento, matrícula y materiales), *Información personal* (convivencia, lug_estudio,...) (Anexo B).

Un resumen descriptivo de los resultados que se obtuvieron en las encuestas es el siguiente: el 55.3% de los estudiantes piensan que sus bases académicas del bachillerato no han sido suficientes; por otro lado, y contrario a lo esperado un 78.2% de los encuestados opina que

la relación con docentes es buena. El 64.9% de los estudiantes le dedica al estudio de las diferentes materias menos de tres horas diarias y el 35.1% le dedica más de tres horas diarias. El 36.5% de los estudiantes ha tenido dificultades en adquirir los materiales necesarios para el desempeño de sus actividades académicas y el 63.5% no las ha tenido.

3. PRESENTACIÓN DE LOS RESULTADOS

3.1 ESTIMACION DEL MODELO DE REGRESION

Se estimo los coeficientes para el modelo de regresión logística

Tabla 3. Estimación de los coeficientes

		Variables in the Equation					95.0% C.I. for EXP(B)		
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1	H_ESTUDIO(1)	.620	.233	7.064	1	.008	1.859	1.177	2.936
	Constant	.253	.131	3.761	1	.052	1.288		
Step 2	H_ESTUDIO(1)	.568	.236	5.811	1	.016	1.764	1.112	2.800
	DOCENTES(1)	.598	.258	5.358	1	.021	1.819	1.096	3.018
	Constant	-.191	.233	.676	1	.411	.826		
Step 3	H_ESTUDIO(1)	.563	.237	5.661	1	.017	1.757	1.104	2.794
	DOCENTES(1)	.601	.260	5.350	1	.021	1.824	1.096	3.035
	MATERIALES(1)	.447	.225	3.931	1	.047	1.563	1.005	2.431
	Constant	-.470	.274	2.954	1	.086	.625		
Step 4	H_ESTUDIO(1)	.607	.239	6.434	1	.011	1.835	1.148	2.935
	BASES(1)	.448	.226	3.937	1	.047	1.566	1.005	2.438
	DOCENTES(1)	.534	.262	4.135	1	.042	1.705	1.019	2.851
	MATERIALES(1)	.457	.227	4.059	1	.044	1.579	1.012	2.461
	Constant	-.633	.287	4.868	1	.027	.531		

- a. Variable(s) entered on step 1: H_ESTUDIO.
- b. Variable(s) entered on step 2: DOCENTES.
- c. Variable(s) entered on step 3: MATERIALES.
- d. Variable(s) entered on step 4: BASES.

Con estas estimaciones de los coeficientes de regresión se puede, a su vez, calcular la probabilidad¹² de que un estudiante repita dos o más de dos asignaturas en los semestres que hasta la fecha haya cursado en la carrera.

$$p(REPETICION = 1) = \frac{1}{1 + e^{-[-0.633 + 0.607H_ESTUDIO + 0.448BASES + 0.534DOCENTES + 0.457MATERIALES]}} \quad (20)$$

Por ejemplo, consideremos un estudiante cuyas características fuesen: le dedica al estudio más de tres horas diarias, opina que las bases del bachillerato son suficientes para las materias de la universidad, que los docentes si tienen buenas relaciones con los estudiantes y que pocas veces o nunca ha tenido dificultades en adquirir materiales para el desempeño de sus actividades académicas, es decir, un estudiante para el cual las variables asumen los valores siguientes:

¹² El modelo de regresión logística se puede transformar fácilmente a: $\pi(x) = \frac{1}{1 + e^{-g(x)}}$

H_ESTUDIO=0, BASES=0, DOCENTES=0, MATERIALES=0

Aplicando la fórmula (20) se tiene que la probabilidad de repetir dos o más de dos materias es igual a 34.68%

Supongamos que a un docente le asignan un grupo de 16 estudiantes y que se desea determinar a quienes hay que dedicarles más tiempo en proceso enseñanza-aprendizaje (por ejemplo, quienes necesitan más asesoría, quienes necesitan que se les nombre un monitor, quienes necesitan que se los nivele en los conocimientos básicos). En la siguiente tabla se exponen los 16 perfiles de entrada hipotéticos y las respectivas probabilidades de repetencia.

Tabla 4. Probabilidades de repetencia

Estudiante	Variables de Entrada					Prob. De repetencia
	H_estudio	Bases	Docentes	Materiales	Suma	
1	1	1	1	1	1.413	0.80423869
2	1	1	1	0	0.956	0.72232022
3	1	1	0	1	0.879	0.70661495
4	1	1	0	0	0.422	0.60396173
5	1	0	1	1	0.965	0.72412177
6	1	0	1	0	0.508	0.62433751
7	1	0	0	1	0.431	0.60611243
8	1	0	0	0	-0.026	0.49350037
9	0	1	1	1	0.806	0.69125647
10	0	1	1	0	0.349	0.58637506
11	0	1	0	1	0.272	0.56758384
12	0	1	0	0	-0.185	0.45388146
13	0	0	1	1	0.358	0.5885562
14	0	0	1	0	-0.099	0.47527019
15	0	0	0	1	-0.176	0.45611323
16	0	0	0	0	-0.633	0.34683061

Como se desprende de estos resultados, si hubiese que fijar una prioridad, ésta sería la siguiente, el estudiante primero, con una probabilidad de repetencia de 80.42%, es aquel que merece por parte del docente mayor atención y dedicación, le seguirían el segundo, el tercero, el quinto, con los riesgos más altos de repetencia de materias. Los cuales tienen como indicadores comunes, dedican menos de tres horas diarias al estudio de las asignaturas y consideran que las bases académicas del bachillerato no han sido suficientes para sus estudios universitarios.

3.2 PRUEBAS DE BONDAD Y AJUSTE

La prueba Chi-cuadrado sobre los coeficientes del modelo indica que a un 95% de confianza los coeficientes son significativos, lo que indica que de manera conjunta las variables son significativas. Por otro lado, la prueba de Hosmer y Lemeshow indica que no existe una diferencia significativa entre los valores observados y los valores predichos, por lo tanto, se puede afirmar que el modelo es significativo. De igual manera, el estadístico de razón de verosimilitud indica que este modelo ajusta de manera adecuada los datos.

Tabla 5. Pruebas de bondad y ajuste

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	7.290	1	.007
	Block	7.290	1	.007
	Model	7.290	1	.007
Step 2	Step	5.343	1	.021
	Block	12.633	2	.002
	Model	12.633	2	.002
Step 3	Step	3.929	1	.047
	Block	16.562	3	.001
	Model	16.562	3	.001
Step 4	Step	3.978	1	.046
	Block	20.541	4	.000
	Model	20.541	4	.000

Tabla 6. Prueba de Hosmer y Lemeshow

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	.000	0	.
2	2.815	2	.245
3	2.747	6	.840
4	4.546	7	.715

Tabla 7. Resumen del modelo

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	482.545 ^a	.020	.027
2	477.202 ^a	.034	.046
3	473.273 ^b	.044	.060
4	469.295 ^b	.054	.074

a. Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

b. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Por último, se estimó el R^2 de Nagelkerke¹³, indica que el modelo de regresión logística explica el comportamiento de la variable dependiente en un 7.4%. Aunque este valor es pequeño, este tipo de modelos clasifican de manera adecuada a los estudiantes.

¹³ El R^2 de Nagelkerke es un estadístico que permite identificar el grado de ajuste del modelo (Montgomery 2002).

3.3 VALIDACION DEL MODELO.

Tabla 8: Tabla de clasificación.

Classification Table^a

Observed			Predicted		Percentage Correct
			REPETICION		
			0	1	
Step 1	REPETICION	0	0	142	.0
		1	0	225	100.0
Overall Percentage					61.3
Step 2	REPETICION	0	30	112	21.1
		1	30	195	86.7
Overall Percentage					61.3
Step 3	REPETICION	0	30	112	21.1
		1	30	195	86.7
Overall Percentage					61.3
Step 4	REPETICION	0	51	91	35.9
		1	40	185	82.2
Overall Percentage					64.3

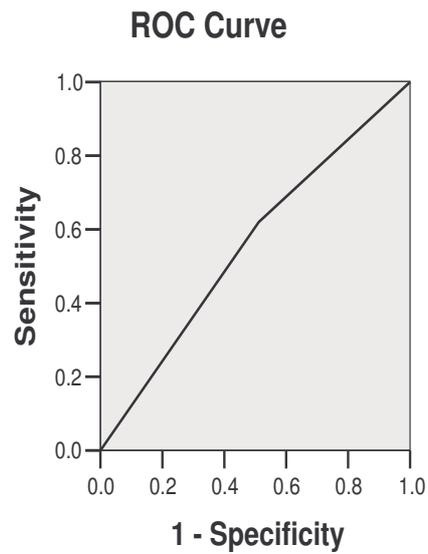
a. The cut value is .500

De la tabla 8 podemos observar que el 35.9% de los estudiantes que no repiten ninguna materia o repiten máximo una y el 82.2 % de los estudiantes que repiten dos o más de dos asignaturas en lo que llevan de su carrera están bien clasificados por el modelo. El porcentaje global de buena clasificación es de 64.3%.

3.4 AREA BAJO LA CURVA ROC

Se puede apreciar una buena eficiencia predictiva del modelo, el área bajo la curva, medida de resumen de tal eficiencia es de 0.654.

Figura 1. Área bajo la curva ROC.



Diagonal segments are produced by ties.

4. CONCLUSIONES Y RECOMENDACIONES

4.1 CONCLUSIONES

Mediante el uso de la regresión logística se identificó los factores que más inciden en la repitencia de materias en la Facultad de Ciencias Exactas y Naturales, los cuales son: Las bases académicas del bachillerato, la actitud de los docentes en el proceso enseñanza aprendizaje, el número de horas diarias que los jóvenes universitarios le dedican al estudio y las dificultades en adquirir los materiales para el desempeño de sus actividades académicas.

Todas las variables incluidas en el modelo de regresión logística tienen un efecto positivo en la variable dependiente, dado el signo de los coeficientes asociados, por lo que un cambio de una unidad en la variable provoca un aumento en el logaritmo de la odds de la probabilidad de que un estudiante repita dos o más de dos materias en su permanencia en la universidad, en el valor del coeficiente respectivo.

4.2 RECOMENDACIONES

El cálculo de la probabilidad de que un estudiante repita dos o más asignaturas en su carrera permite a la facultad diseñar estrategias para que el grupo de estudiantes con similares características reciban tratamiento especial con el fin de motivarlos, nivelarlos y asesorarlos en el ambiente académico universitario.

BIBLIOGRAFÍA

HOSMER, David W. y LEMESHOW, Stanley. Applied Logistic Regression. Second Edition. New York: Interscience Publication, John Wiley & Sonc, 2000. 163 p.

PEREZ, Cesar. Técnicas estadísticas con SPSS. Madrid: Prentice Hall, 2001. 420 p.

SILVA, Luis Carlos. Excursión a la regresión logística en Ciencias de la Salud. Ediciones Díaz de Santos.

UNIVERSIDAD DE NARIÑO. Estatuto Estudiantil. Pasto, Colombia: Centro de Publicaciones, 2007.

MONTGOMERY, Duglas . Introduction to linear regresión analysis. New York: John Wiley & Sonc. 145 p.

TORRES, Rosa María. Repetition: A major obstacle to education for all. In: Education. New York: Unicef, No. 15 (Abril, 1995). Disponible en internet, URL: www.mineducacion.gov.co/1621/articles-85774_archivo_pdf.pdf.

RAMÍREZ, Pedro; REYES, Jimmy; ESCOBAR, Carlos y DUARTE, Juan. Una aplicación del modelo de regresión logística en la predicción del rendimiento estudiantil. Antofagasta, Chile: s.n.e., 2004. Disponible en interner, URL: www.scielo.cl/scielo.php?pid=S0718-07052007000200006&script=sci_arttext - 54k -.

BRENES, María Isabel. Deserción y repitencia en la educación superior universitaria de Costa Rica. s.n.e., 2005. Disponible en Internet, URL: www.catalogos/doctextcomp/opes/2005/INFORMEFINALDESERCIONYPREITENCIACOSTARIC.

ANEXOS

Anexo A: Encuesta aplicada a los cuatro programas de la facultad de Ciencias Exactas y Naturales de la Universidad de Nariño

UNIVERSIDAD DE NARIÑO FACULTAD DE CIENCIAS EXACTAS Y NATURALES

Objetivo: Determinar los factores que inciden en la repetición de asignaturas en los diferentes programas de la Facultad de Ciencias Exactas y Naturales de la Universidad de Nariño.

Nota: Por favor contestar las siguientes preguntas con gran seriedad, sinceridad y responsabilidad, para así, poder hacer sugerencias a las directivas de tu facultad y de la Universidad para mejorar tu situación o por lo menos buscar una asesoría profesional que te oriente y te guíe.

IDENTIFICACION

1. Programa al que pertenece _____
2. Semestre que cursa _____
3. Código Estudiantil _____
4. Edad. _____
5. Estrato socio-económico _____

Por favor marca con una X, una única opción

INFORMACION ACADEMICA

6. En lo que lleva de su carrera universitaria, ha repetido materias:
 - a. ninguna vez
 - b. una vez
 - c. dos veces: cuales: _____
 - d. tres o más veces: Cuales: _____
7. El colegio en el cual cursó y aprobó la mayoría de grados es de carácter:
 - a. oficial
 - b. privado
 - c. semi-oficial
8. Cuántas horas diarias, fuera de los horarios de clase le dedica al estudio de las materias que cursa?
 - a. menos de una hora
 - b. entre una y tres horas
 - c. entre tres y cinco horas
 - d. más de 5 horas
9. Por cual de las siguientes razones esta estudiando la carrera:
 - a. Influencia directa de tus padres o alguna otra persona
 - b. Por decisión propia y gusto por la carrera
 - c. El puntaje icfes no te alcanzó para ingresar a la carrera que deseabas
 - d. Ingresaste a esta carrera porque tu idea es (o era) hacer transferencia a otra carrera que en realidad quieres (o querías) estudiar.
 - e. Las condiciones económicas no te permiten ingresar a una carrera que devengue altos gastos en: prácticas, instrumentos, materiales, etc.

10. El grupo de compañeros(as) con los cuales estudia tiene la disciplina necesaria para desarrollar las actividades académicas de la universidad (aprovecha el tiempo para lo planeado)
- siempre
 - frecuentemente
 - pocas veces
 - nunca
11. Las bases académicas del bachillerato han sido suficientes para sus estudios universitarios?
- siempre
 - frecuentemente
 - pocas veces
 - nunca
12. Con qué frecuencia asiste a una biblioteca a consultar o estudiar
- siempre
 - frecuentemente
 - pocas veces
 - nunca
13. La relación personal con los docentes es satisfactoria
- siempre
 - frecuentemente
 - pocas veces
 - nunca
14. Los docentes de las diferentes asignaturas representan para usted una autoridad académica
- siempre
 - frecuentemente
 - pocas veces
 - nunca
15. El programa ofrece condiciones óptimas para la investigación (infraestructura, bibliografía, asesoría, etc.)
- siempre
 - frecuentemente
 - pocas veces
 - nunca

INFORMACION SOCIO-ECONOMICA

16. Su sostenimiento económico durante la mayor parte de sus estudios ha dependido de:
- papá, mamá o ambos
 - apoyo familiar o de otros
 - usted mismo
17. El pago de matrícula depende de:
- papá, mamá o ambos
 - apoyo familiar o de otros
 - crédito
 - beca
 - usted mismo

18. Ha tenido dificultades en adquirir los materiales necesarios para el desempeño de sus actividades académicas
- a. siempre
 - b. frecuentemente
 - c. pocas veces
 - d. nunca

INFORMACION PERSONAL

19. Con quién vive?
- a. papá y mamá
 - b. uno de padres
 - c. con otra persona diferente a papá y mamá
 - d. vive solo
20. El lugar de estudio en el hogar consta de:
- a. lo necesario ubicado en algún lugar de la casa
 - b. lo necesario en un sitio independiente
 - c. lo necesario, más biblioteca, computador, internet.
 - d. No tienes estudio en casa.
21. Su novio(a) o amigo(a) sentimental ha afectado el rendimiento académico de sus estudios
- a. siempre
 - b. frecuentemente
 - c. pocas veces
 - d. nunca
22. El repetir una materia en más de 2 ocasiones, aplaza el tiempo de culminación de estudios esta situación es:
- a. preocupante
 - b. no preocupante, eres joven y tienes mucho tiempo para lograr lo que te propones
23. Además de las actividades académicas, a qué otra actividad se dedica?
- a. deportivas
 - b. culturales
 - c. políticas
 - d. otra.Cuál?. _____
24. Según su criterio, qué otro factor tiene que ver con que un estudiante repruebe una materia?

OBSERVACIONES:

Muchas Gracias por su colaboración.

Anexo B: Variables independientes iniciales

* Identificación.

SEMESTRE. Semestre que cursa

EDAD. Edad en años cumplidos del estudiante

ESTRATO. Estrato socio-económico donde vive el estudiante

* Información académica.

COLEGIO. Tipo de colegio donde realizo el bachillerato(oficial o privado)

H_ESTUDIO. Horas diarias dedicadas al estudio

RAZONES. Razones por las cuales esta en el programa

COMPAÑEROS. El grupo de estudio tiene la disciplina de estudio

BASES. Las bases el bachillerato son suficientes

BIBLIOTECA. Frecuencia con que asiste a la biblioteca

DOCENTES. Relaciones personales con los docentes

AUTORIDAD. El docente es una autoridad para el estudiante

INFRAESTRUCTURA. Infraestructura de la facultad

* Información Socio-económica.

SOSTENIMIENTO. Sostenimiento económico del estudiante

MATRICULA. Responsable de la matricula

MATERIALES. Materiales para sus actividades académicas

* Información Personal.

CONVIVENCIA. con quien vive

LUG_ESTUDIO. Lugar de estudio

NOVIO. El novio(a) afecta el rendimiento académico.

REPET_MATERIAS. La repitencia de materias es preocupante?.

OTRA_ACT. Otra actividad que realiza el estudiante