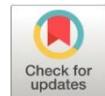


Análisis nivel de logro alcanzado con el examen ser bachiller en función de variables sociodemográficas

Analysis of the level of achievement of the bachelor's degree exam as a function of sociodemographic variables

- ¹ Tania Paulina Morocho-Barrionuevo  <https://orcid.org/0000-0002-1019-6049>
Escuela Superior Politécnica de Chimborazo (ESPOCH), Facultad de Ciencias, Carrera de Estadística, Riobamba, Ecuador
tpaulina.morochob@epoch.edu.ec
- ² Carmen Estrella Morocho-Barrionuevo  <https://orcid.org/0000-0003-0050-5888>
Universidad Nacional de Chimborazo (UNACH), Riobamba, Ecuador
carmen.morocho@unach.edu.ec
- ³ Diana Katherine Campoverde-Santos  <https://orcid.org/0000-0001-8538-6747>
Escuela Superior Politécnica de Chimborazo (ESPOCH), Facultad de Ciencias Pecuarias, Carrera de Medicina Veterinaria, Riobamba, Ecuador
diana.campoverde@epoch.edu.ec



Artículo de Investigación Científica y Tecnológica

Enviado: 12/04/2022

Revisado: 27/05/2022

Aceptado: 29/06/2022

Publicado: 11/07/2022

DOI: <https://doi.org/10.33262/ap.v4i3.229>

Cítese:

Morocho Barrionuevo, T. P., Morocho Barrionuevo, C. E., & Campoverde Santos, D. K. (2022). Análisis nivel de logro alcanzado con el examen ser bachiller en función de variables sociodemográficas. AlfaPublicaciones, 4(3), 113–129. <https://doi.org/10.33262/ap.v4i3.229>



ALFA PUBLICACIONES, es una revista multidisciplinar, **trimestral**, que se publicará en soporte electrónico tiene como **misión** contribuir a la formación de profesionales competentes con visión humanística y crítica que sean capaces de exponer sus resultados investigativos y científicos en la misma medida que se promueva mediante su intervención cambios positivos en la sociedad. <https://alfapublicaciones.com>

La revista es editada por la Editorial Ciencia Digital (Editorial de prestigio registrada en la Cámara Ecuatoriana de Libro con No de Afiliación 663) www.celibro.org.ec



Esta revista está protegida bajo una licencia Creative Commons Attribution Non Commercial No Derivatives 4.0 International. Copia de la licencia: <http://creativecommons.org/licenses/by-nc-nd/4.0/>

**Palabras
claves:**

Ecuador,
Modelos de
respuesta
Ordinal,
Modelos de
respuesta
Nominal,
modelos de
respuesta
múltiple,
Método
Stepwise

Keywords:

Ecuador,
Ordinal
Response
Models,
Nominal
Response
Models,
Multiple
Response
Models,
Stepwise
Method.

Resumen

Introducción. El presente trabajo describe el modelo de regresión logística multinomial con aplicación a datos reales, utilizando la base de datos del examen Ser bachiller – Ecuador del año lectivo 2018-2019 modelización el nivel de logro alcanzado en la nota del examen de grado, en función de las variables sociodemográficas edad, sexo, provincia, tipo de financiamiento de la institución educativa y autoidentificación étnica. **Objetivo.** Describir uso y utilidad de los modelos de respuesta nominal y de los de respuesta ordinal, obteniendo sus mejores predictores y estableciendo sus diferencias en la modelación de los datos. **Metodología.** Para el cálculo se utilizó el software libre R con datos tomados desde el Instituto Nacional de Evaluación Educativa ajustando el modelo de respuesta nominal y ordinal, además aplicando el método stepwise se halla los mejores predictores para el estudio. **Resultados.** Los resultados indican una correlación positiva y significativa entre el entorno de control y los indicadores ROE y ROA. **Conclusiones.** Se pudo concluir mediante la aplicación del método stepwise que las variables independientes utilizadas en el estudio son las adecuadas ya que mediante la selección automática en los modelos de respuesta tanto nominal y ordinal dan como predictoras a las mismas variables.

Abstract

Introduction. The present work describes the multinomial logistic regression model with application to real data, using the database of the Ser bachiller exam - Ecuador of the 2018-2019 school year, modeling the level of achievement reached in the grade exam grade, based on the sociodemographic variables age, sex, province, type of financing of the educational institution and ethnic self-identification. **Goal.** Describe the use and usefulness of nominal response models and ordinal response models, obtaining their best predictors and establishing their differences in data modeling. **Methodology.** For the calculation, the free software R was used with data taken from the National Institute of Educational Evaluation, adjusting the nominal and ordinal response model, in addition, applying the stepwise method, the best predictors for the study were found. **Results.** The results indicate a positive and significant correlation between the control environment and the ROE and ROA indicators. **Conclusions.** It was possible to conclude

through the application of the stepwise method that the independent variables used in the study are the appropriate ones, since through the automatic selection in both the nominal and ordinal response models they give the same variables as predictors.

Introducción

En el Ecuador la educación de calidad es un derecho que todas las personas poseen y que pueden acceder sin excepción alguna, según la Constitución de la Republica.

Los niveles educativos manejados están establecidos por el Sistema Nacional de Educación, mismos que se encuentran divididos en tres niveles: Inicial, Básica y Bachillerato.

- El nivel de Educación Inicial, es el proceso de acompañamiento de niños para el desarrollo integral de sus capacidades cognitivas, afectivas, psicomotrices, sociales e identitarias, que va desde los 3 hasta los 5 años y se divide en dos subniveles (Organización de las Naciones Unidas para la Educación, Ciencia y Cultura [UNESCO], 2019).

Inicial 1, que no es escolarizado y comprende a infantes de hasta tres años.

Inicial 2, que comprende a infantes de tres a cinco años.

- El nivel de Educación General Básica está compuesta por diez años de educación obligatoria en los que se refuerzan, amplían y profundizan las capacidades y competencias adquiridas, se introducen disciplinas básicas, para garantizar la diversidad cultural y lingüística, está dividida en cuatro subniveles:

Preparatoria, que corresponde al primer grado de Educación General Básica y preferentemente se ofrece a los estudiantes de cinco años.

Básica Elemental, que corresponde al segundo, tercero y cuarto grado de Educación General Básica y preferentemente se ofrece a los estudiantes de 6 a 8 años.

Básica Media, que corresponde a quinto, sexto y séptimo grado de Educación General Básica y preferentemente se ofrece a los estudiantes de 9 a 11 años.

Básica Superior, que corresponde a octavo, noveno y décimo grados de Educación General Básica y preferentemente se ofrece a los estudiantes de 12 a 14 años.

- El nivel de Bachillerato General Unificado comprende tres años de educación obligatoria brindando una formación general y una preparación interdisciplinaria desarrollando en los estudiantes capacidades permanentes de aprendizaje y competencias ciudadanas, consta de tres cursos primero, segundo, tercero de bachillerato y preferentemente se ofrece a los estudiantes de 15 a 17 años.

Las edades estipuladas son sugeridas para la educación en cada nivel, sin embargo, no se niega el acceso de los estudiantes con: necesidades educativas especiales, jóvenes y adultos con escolaridad inconclusa, repetición de un año escolar, entre otros, permitiéndoles acceder a un grado o curso (Ministerio de Educación, 2015).

Desde el 2010, el Ecuador adoptó la Ley Orgánica de Educación Superior (LOES), la cual permitió la creación de la Secretaría de Educación Superior de Ciencia y Tecnología (SENESCYT) y el Sistema Nacional de Admisión y Nivelación (SNNA) que implementó el Examen Nacional de Educación Superior (ENES) como un instrumento obligatorio para regular el acceso a la Educación Superior, pero en el período 2016-2017, se inició la aplicación de un nuevo examen, denominado Examen Unificado Ser Bachiller, a través del cual los estudiantes obtendrían su título de bachiller y el cupo a una universidad (Guadagni, 2006).

La evaluación educativa es un proceso que integra la elaboración, aplicación y análisis de instrumentos de medición, los cuales deducen las capacidades y destrezas de las personas. El instrumento de medición educativo tiene como función ofrecer información para la toma correcta de decisiones, y numerosas investigaciones consideran que el examen de ingreso a la universidad es una variable imprescindible en el éxito académico del estudiante (Carrión, 2002).

Cuando se emplean instrumentos de gran escala y de alto impacto social, tal como los exámenes de admisión y/o aprobación de la Educación a nivel de Bachillerato, su elaboración debe ajustarse a rigurosos estándares de calidad (Aiken, 1996).

El examen Ser Bachiller es tomado a los estudiantes pertenecientes al tercero de Bachillerato General Unificado (BGU). El procedimiento para la ejecución del examen parte de la autoasignación del estudiante a una sede o una sede preasignada, posteriormente deberán acercarse según el cronograma previsto, puesto que hay una fecha asignada para personas con discapacidad y personas privadas de libertad, otra para la población no escolar y otra para la población escolar y personas ecuatorianas residentes en el extranjero (Secretaría de Educación Superior [SENECYT], 2019).

La implementación del examen Ser bachiller desde el 2017 fue concebida para evaluar el desarrollo de las aptitudes y destrezas que deben alcanzar los estudiantes al terminal la

educación intermedia, y a la vez esta evaluación contribuye al proceso de admisión a la educación superior pública. La información recopilada con este examen sirve para la formulación de mecanismos de mejora para la educación a nivel inicial, básico y bachillerato, enfatizando en acciones de mejora para el acceso a la educación superior (Instituto Nacional para la Evaluación de la Educación [INEE], 2017).

Siendo así de gran importancia el Examen Unificado Ser bachiller para la población con el fin de continuar con su vida educativa y profesional, se toma como base sustentativa para el estudio la base de datos del Instituto Nacional de Evaluación educativa con su respectivo diccionario de variables denominada “Base de datos Ser Bachiller Año lectivo 2018-2019”, tomada a 514852 estudiantes pertenecientes al tercero de Bachillerato, y es oportuno identificar qué factores sociodemográficos (autoidentificación étnica, región natural, tipo de financiamiento, tipo de sexo y edad) influyen en que el estudiante de educación nivel bachillerato obtengan un nivel de logro en el rendimiento académico de tipo (insuficiente, elemental, satisfactorio o excelente).

La autoidentificación étnica es una de las variables para el estudio que vislumbra a personas que se autodefinen con alguna nacional o pueblo de manera libre y voluntaria, sea indígena, mestizo/blanco, afroecuatoriano, montubio u otra étnia.

En relación con las regiones naturales en Ecuador se encuentra definida por criterios de geografía física, principalmente con el clima, la vegetación, la hidrografía de los suelos y otros. La división territorial corresponde a Costa, Sierra, Oriente, Insular, Otro (Instituto Nacional de Estadísticas y Censos [INEC], 2018).

Región Costa: Comprende más de la cuarta parte del país, es un área geográfica que se encuentra entre el Océano Pacífico y la Cordillera de los Andes, conformada por las provincias de: Esmeraldas, Manabí, Los Ríos, Santa Elena, Guayas, Santo Domingo de los Tsáchilas y El Oro.

Región Sierra: Se extiende en una doble hilera de montañas y una estrecha meseta conocida como el valle interandino, constituida por: Azuay, Bolívar, Cañar, Carchi, Cotopaxi, Chimborazo, Imbabura, Loja, Pichincha y Tungurahua.

Región Oriente: Al este de los andes, se encuentra cubierta casi en su totalidad con selva, presenta gran cantidad de la flora y la fauna nativa, formada por: Morona Santiago, Napo, Orellana, Pastaza, Sucumbíos y Zamora Chinchipe.

Región Insular: Integrada por el Archipiélago de Colón, comprendida por la Isla Isabela, Santa Cruz y San Cristóbal (Romero, 2021).

Región Zona no delimitada: Constituida por cantones que se encuentran en zonas limítrofes o fronterizas como: Loja (Macará, Sozoranga, Espíndola, Puyango, Pindal,

Zapotillo, Calvas), Zamora Chinchipe (El Pangui, Yacuambi, Yantzaza, Nangaritzza, Palanda, Chinchipe, Paquisha), Pastaza (Pastaza, Arajuno), Orellana (Aguarico) y Morona Santiago (Taisha, Tiwintza, Limón Indanza, San Juan Bosco, Gualaquiza).

El tipo de financiamiento de las instituciones educativas está establecido según su sostenimiento como público, privado y mixto o fiscomisionales.

Públicas: Son instituciones educativas financiadas por el estado.

Mixto o Fiscomisionales: Estas Instituciones Educativas son de carácter religioso o laico, de derecho privado y sin fines de lucro, que garantizan una educación gratuita y de calidad. Los centros cuentan con financiamiento total o parcial del Estado (Gobierno Central a través del MinEduc).

Privado: Constituidas y administradas por personas naturales o jurídicas de derecho privado. La educación en estas Instituciones Educativas puede ser confesional o laica.

El tipo de sexo toma hombre y mujer, y la edad va desde los 16 a los 25 años debido a que la población mayoritaria se encuentra entre este rango de edades (Instituto Nacional de Evaluación Educativa, 2018).

El nivel de logro alcanzado existe cuatro categorías consideradas dentro de la calificación de la prueba Ser Bachiller: “Excelente” con una calificación de 950 a 1.000 puntos. En esta categoría se ubica el Grupo de Alto Rendimiento, seguido de la calificación “Satisfactoria” con un puntaje de hasta 800. Luego sigue la categoría “Elemental” con un puntaje de hasta 700, el cual es un requisito mínimo para que el estudiante se pueda graduar. Por último, una calificación “Insuficiente” de hasta 601, aunque sea el mínimo para poder postular a la educación superior, no puede aprobar como requerimiento de examen de grado (Instituto Nacional de Estadísticas y Censos [INEC], 2018).

El objetivo del presente trabajo es mostrar el uso y utilidad de los modelos de respuesta nominal y de los modelos de respuesta ordinal, para la modelización del nivel de logro alcanzado en la nota del examen de grado Ser Bachiller-Ecuador periodo 2018-2019, en función de las variables sociodemográficas edad, sexo, provincia, tipo de financiamiento de la institución educativa y autoidentificación étnica.

Metodología

A. Modelos de Respuesta Múltiple

Con una variable aleatoria de respuesta politómica Y , con categorías $Y_1, Y_2, \dots, Y_S, S > 2$ y un conjunto de variables explicativas $X = (X_1, X_2, \dots, X_R)'$, los modelos de respuesta múltiple tienen como objetivo explicar la respuesta a partir de las variables explicativas (Agresti, 2002).

Si se dispone de:

- Una muestra aleatoria de tamaño N con Q combinaciones diferentes de valores de las variables explicativas $X_1, X_2, \dots, X_R, x_q = (x_{q1}, \dots, x_{qR})' \forall q = 1, \dots, Q$.
- Una muestra aleatoria de n_q observaciones independientes de las variables de respuesta politémica Y para cada $x_q, (y_{q1}, \dots, y_{qS})' \forall q = 1, \dots, Q$ con y_{qS} el número de observaciones que caen en la categoría de respuesta $Y_s \forall s = 1, \dots, S$.

De tal manera que los vectores tienen distribuciones de probabilidad multinomiales independientes,

$$(y_{q1}, \dots, y_{qS})' \rightarrow M(n_q; p_1, \dots, p_S) \quad \text{Ecu. 1}$$

Verificando que:

- $\sum_{s=1}^S p_s = 1$, donde $p_s = (Y = Y_s | X = x_q)$
- $\sum_{s=1}^S y_s = n_q$ y $\sum_{q=1}^Q n_q = N$

1. Modelos logit de Respuesta Nominal

Al tener una variable de respuesta nominal el modelo logit se formula en base a las transformaciones logit generalizadas definidas con respecto a una categoría de referencia:

$$L_s(x) = \ln \left[\frac{p_s(x)}{p_S(x)} \right], \quad \forall t, s = 1, \dots, S - 1, \quad \text{Ecu. 2}$$

Se ha tomado como categoría de referencia la última, aunque en la aplicación los programas R suelen tomar la primera; para cualquier transformación logit para un par de categorías se puede obtener a partir de sus transformaciones logit generalizadas asociadas en la forma (Agresti, 2002):

$$\ln \left[\frac{p_t(x)}{p_s(x)} \right] = L_t(x) - L_s(x) \quad \forall t, s \quad \text{Ecu. 3}$$

2. Modelo logit de Respuesta Ordinal

En el caso de una variable de respuesta cualitativa ordinal se definen transformaciones logit que tienen en cuenta el orden entre las distintas categorías de respuesta. Se los conoce como modelos logit acumulativos y las transformaciones logit acumulativas se definen como (Agresti, 2002):

$$L_s(x_q) = \ln \left(\frac{P[Y \leq Y_s | X = x_q]}{1 - P[Y \leq Y_s | X = x_q]} \right) = \ln \left(\frac{F(x_q)}{1 - F(x_q)} \right), \quad \forall s = 1, \dots, S - 1 \quad \text{Ecu. 4}$$

Dadas R variables explicativas X_1, X_2, \dots, X_R y Q observaciones diferentes x_q el modelo logit acumulativo de efectos homogéneos es de la forma:

$$L_s(x) = \alpha_s + \sum_{r=0}^R \beta_r x_{qr} = x'_q \beta_s \quad \text{Ecu. 5}$$

Siendo $x_q = (x_{q1}, \dots, x_{qR})'$ el vector de valores observados de las variables explicativas con $\beta = (\beta_0, \beta_1, \dots, \beta_R)'$ el vector de parámetros (Agresti, 2002).

a) Estimación por máxima verosimilitud

Dada una muestra aleatoria de tamaño N con Q combinaciones diferentes de valores de las variables explicativas X_1, \dots, X_R , $x_q = (x_{q0}, x_{q1}, \dots, x_{qR})'$ con $x_{q0} = 1 \forall q = 1, \dots, Q$; y una muestra aleatoria de n_q observaciones independientes de la variable de respuesta politómica $Y, (y_{q1}, \dots, y_{qs})' \forall q = 1, \dots, Q$, la función de verosimilitud de los datos es entonces (Agresti, 2002):

$$\prod_{q=1}^Q \left(\frac{n_q!}{\prod_{s=1}^S (y_{qs})!} \prod_{s=1}^S p_{qs}^{y_{qs}} \right) \quad \text{Ecu. 6}$$

El núcleo de la log-verosimilitud este dado:

$$K = \sum_{q=1}^Q \sum_{s=1}^S y_{qs} \ln(p_s) \quad \text{Ecu. 7}$$

b) Significación de parámetros

La matriz de covarianzas de los estimadores de los parámetros de los distintos modelos aquí abordados $\hat{\beta}$ es la inversa de la matriz de información de Fisher dada por las derivadas parciales de segundo orden del núcleo de la verosimilitud. (Agresti, 2002)

$$Cov(\hat{\beta}) = \left[-E \left(\frac{\partial^2 K}{\partial \beta_j \partial \beta_k} \right) \right]^{-1} \quad \text{Ecu. 8}$$

$$Cov(\hat{\beta}) = \begin{pmatrix} Var(\hat{\beta}_1) & Cov(\hat{\beta}_1, \hat{\beta}_2) & \dots & Cov(\hat{\beta}_1, \hat{\beta}_h) \\ Cov(\hat{\beta}_1, \hat{\beta}_2) & Var(\hat{\beta}_2) & \dots & Cov(\hat{\beta}_2, \hat{\beta}_h) \\ \dots & \dots & \dots & \dots \\ Cov(\hat{\beta}_1, \hat{\beta}_h) & Cov(\hat{\beta}_2, \hat{\beta}_h) & \dots & Var(\hat{\beta}_h) \end{pmatrix} \quad \text{Ecu. 9}$$

Las raíces cuadradas de los elementos de la diagonal de esta matriz son los errores estándar (ASE) de los estimadores de los parámetros del modelo.

Como estimadores de máxima verosimilitud tienen distribución normal asintótica (Jobson, 1991).

$$\hat{\beta} \xrightarrow[N \rightarrow \infty]{d} N(\beta, Cov(\hat{\beta})) \quad \text{Ecu. 10}$$

c) *Bondad de ajuste: Test chi-cuadrado de razón de verosimilitudes*

El estadístico de Wilks de razón de verosimilitudes para el contraste de bondad de ajuste de un modelo de regresión logística multinomial M se obtiene como menos dos veces el logaritmo del cociente entre el supremo de la verosimilitud bajo la hipótesis nula y el supremo de la verosimilitud en la población, es decir (Jobson, 1991):

$$G^2(M) = 2 \left[\sum_{q=1}^Q \sum_{s=1}^S y_{qs} \ln \left(\frac{y_{qs}}{p_{qs}} \right) \right] \quad \text{Ecu. 11}$$

El contraste se puede expresar como:

$$H_0: p(x_q) = \frac{e^{(\sum_{r=0}^R \beta_{rs} x_r)}}{1 + \sum_{s=2}^S e^{(\sum_{r=0}^R \beta_{rs} x_r)}} \quad \text{Ecu. 12}$$

$$H_1: p(x_q) \neq \frac{e^{(\sum_{r=0}^R \beta_{rs} x_r)}}{1 + \sum_{s=2}^S e^{(\sum_{r=0}^R \beta_{rs} x_r)}}$$

Tasa de clasificaciones correctas

Para cuantificar la bondad del ajuste global del modelo se dispone también de otra medida como es la tasa de clasificaciones correctas. Es decir, a partir del modelo ajustado, se clasifica cada observación en la categoría más probable, construyendo así una matriz de clasificación observados-predichos y se utiliza el porcentaje de clasificaciones correctas como una medida de la calidad de predicción, del mismo modo que se hace en el análisis discriminante (Pando & San Martín, 2004). Se define como la proporción de individuos clasificados correctamente por el modelo y se calcula como el cociente entre el número de observaciones clasificadas correctamente y el tamaño muestral N . Un individuo es clasificado correctamente por el modelo cuando su valor observado de la variable respuesta Y_s coincide con su valor estimado por el modelo (Ruston, 2012).

d) *Significación de variables. Contrastes condicionales de razón de verosimilitudes*

Al suponer que tenemos un modelo de regresión logística multinomial M_G que se ajusta bien y se desea contrastar si un subconjunto de parámetros, $\beta = (\beta_1, \dots, \beta_r)$ son nulos. Sea M_P el modelo con ese subconjunto de parámetros ceros.

Así que M_P esta anidado en el modelo general M_G . Así que planteamos el contraste:

$$H_0: \beta = 0 \quad (M_P \text{ se verifica}) \quad \text{Ecu. 12}$$

$$H_1: \beta \neq 0 \quad (\text{asumiendo cierto } M_G)$$

e) *Selección de modelos logit - Stepwise*

La selección Stepwise es el método que combina los modelos adelante y atrás, el cual se puede empezar por el modelo vacío o el completo, pero en cada paso se examinan las variables incluidas, si deben salir y las no seleccionadas, si deben ingresar. El método stepwise, está basado en contrastes condicionales de razón de verosimilitudes (Hoster, 2000).

Partiendo de un modelo vacío, sólo con la constante, el método consiste en partir de este modelo inicial, de modo que en cada paso se ajustarán todos aquellos modelos que resulten de incluir cada una de las variables explicativas que no están en el modelo seleccionado en el paso anterior. Llevándose así a cabo contrastes condicionales de razón de verosimilitudes que tienen en la hipótesis alternativa el modelo resultante de la inclusión de cada variable. De modo que se seleccionarán las variables para las que el constante sea significativo, y se incluirá en modelo aquella variable asociada al mínimo p – valor de entre todos los menores o iguales que $\alpha 1$. (Jobson, 1991).

f) *Modelos con variables explicativas categóricas: variables del diseño*

Asociadas a una variable cualitativa A con categorías denotadas por $A_i (i = 1, \dots, I)$, se define un total de $(I - 1)$ variables del diseño o variables ficticias, se puede utilizar el método parcial o el método marginal. (Agresti, 2002).

Resultados

Para conocer el nivel de logro alcanzado, se analizará distintas variables sociodemográficas (edad, sexo, autoidentificación étnica, región natural y tipo de financiamiento de la institución educativa). Con las variables planteadas se desea explicar aquellos factores que mayormente influyen en el estudio.

Los datos para utilizar corresponden a la base de datos Ser Bachiller del Instituto Nacional de Evaluación Educativa del Ecuador correspondiente al año lectivo 2018-2019 proporcionado por su página web (<http://evaluaciones.evaluacion.gob.ec/BI/bases-de-datos-ser-bachiller/>).

Los archivos se encuentran divididos en diccionario de variables especificando (instituciones educativas, distrito, circuito, provincia, cantón, parroquia) y la base de datos (sin etiqueta inicial de fila y el resto de las filas contiene información registrada de cada sujeto perteneciente a la muestra). En columnas se encuentra la información de cada una de las variables recogidas mismas que están codificadas. Inicialmente la muestra se encuentra conformada de 514852 personas y 34 variables.

Se procede a filtrar la base considerando las variables para el estudio y eliminando las

respuestas vacías, con la base establecida se procede hacer el cambio de nominación tomando su valor numérico por su nombre respectivo teniendo así un total de 275498 individuos y 6 variables. El estudio se realizó con base en datos agrupados mismos que constan con 6000 filas con 7 columnas, tomando la frecuencia de cada una de las variables (ver figura 1).

La base de datos contiene:

- Variable dependiente nivel de logro alcanzado siendo: 0=insuficiente, 1=elemental, 2=satisfactorio y 3=excelente.
- Variables Independientes que ofrecen información sociodemográfica del individuo: edad, sexo, autoidentificación étnica, provincia y tipo de financiamiento de las instituciones educativas detallamos en la tabla 1.

Dentro de la base tenemos variables categóricas las regiones naturales, el tipo de financiamiento, la autoidentificación étnica, el sexo y el nivel de logro alcanzado. La herramienta que se va a utilizar para los análisis es R Studio.

Figura 1

Datos agrupados Ser Bachiller 2018-2019

edad	tpsexo	financiamiento	etnibbe	nm_regi	nl_inev	Freq	
1	16	hombre	publico	mestizo/blanco	costa	insuficiente	3
2	17	hombre	publico	mestizo/blanco	costa	insuficiente	655
3	18	hombre	publico	mestizo/blanco	costa	insuficiente	6152
4	19	hombre	publico	mestizo/blanco	costa	insuficiente	4150
5	20	hombre	publico	mestizo/blanco	costa	insuficiente	2253
6	21	hombre	publico	mestizo/blanco	costa	insuficiente	910
7	22	hombre	publico	mestizo/blanco	costa	insuficiente	577
8	23	hombre	publico	mestizo/blanco	costa	insuficiente	404
9	24	hombre	publico	mestizo/blanco	costa	insuficiente	270
10	25	hombre	publico	mestizo/blanco	costa	insuficiente	238
11	16	mujer	publico	mestizo/blanco	costa	insuficiente	15
12	17	mujer	publico	mestizo/blanco	costa	insuficiente	737
13	18	mujer	publico	mestizo/blanco	costa	insuficiente	6219
14	19	mujer	publico	mestizo/blanco	costa	insuficiente	3855
15	20	mujer	publico	mestizo/blanco	costa	insuficiente	1665
16	21	mujer	publico	mestizo/blanco	costa	insuficiente	834

Showing 1 to 17 of 6,000 entries. 7 total columns

Tabla 1

Datos agrupados Ser Bachiller 2018-2019

Edad	Sexo	Autoidentificación étnica	Regiones naturales	Tipo de financiamiento de la institución educativa
De 16 - 25 años	1: Mujer 2: Hombre	1: Afroecuatoriano	1: Costa	1: Público (Fiscal y Municipal) 2: Privado (Particular) 3: Mixto (Fiscomisional)
		2: Montubio	2: Sierra	
		3: Indígena	3: Oriente	
		4: Mestizo/ Blanco	4: Insular	
		5: Otro	90: Otras (Zona No Delimitada)	

Ajuste de regresión de respuesta nominal

Para realizar el ajuste se tomó en cuenta las categorías de referencia con las cuales se va a trabajar, siendo así: Para región natural, igual a Costa; para autoidentificación étnica mestizo/blanco, tipo de financiamiento público, tipo de sexo hombre y nivel de logro alcanzado insuficiente.

Se modeliza el nivel de logro alcanzado en función de la edad (cuantitativa), tipo de financiamiento, región natural, tipo de sexo y autoidentificación étnica (las cuatro cualitativas). En este caso la variable respuesta Y (nivel de logro alcanzado) toma cuatro valores que denotaremos por Y_0 =elemental (categoría de referencia), Y_1 =insuficiente, Y_2 =satisfactorio, Y_3 =excelente ($S=4$). En la tabla 2 se presente el ajuste del modelo

Tabla 2

Datos agrupados Ser Bachiller 2018-2019

Coefficientes:	Elemental	Satisfactorio	Excelente
nm_regisierra	$\hat{\tau}_{s1}$ -1,344	$\hat{\tau}_{12}$ 0,665	$\hat{\tau}_{13}$ 1,034
nm_regiotras	$\hat{\tau}_{s2}$ -0,453	$\hat{\tau}_{22}$ 0,148	$\hat{\tau}_{23}$ -0,028
nm_regioriente	$\hat{\tau}_{s3}$ -0,611	$\hat{\tau}_{32}$ -0,075	$\hat{\tau}_{33}$ -0,812
nm_regiinsular	$\hat{\tau}_{s4}$ -0,992	$\hat{\tau}_{42}$ 0,111	$\hat{\tau}_{34}$ -0,410
financiamientoprivado	$\hat{\tau}_{s1}$ -0,660	$\hat{\tau}_{12}$ 0,955	$\hat{\tau}_{13}$ 1,400
financiamientomixto	$\hat{\tau}_{s2}$ -0,159	$\hat{\tau}_{22}$ 0,667	$\hat{\tau}_{23}$ 1,168
tp_sexomujer	$\hat{\tau}_{s1}$ -0,042	$\hat{\tau}_{12}$ 0,147	$\hat{\tau}_{13}$ 0,245
edad	$\hat{\beta}_{11}$ 0,224	$\hat{\beta}_{21}$ -0,407	$\hat{\beta}_{31}$ -0,833
etnibbeindigena	$\hat{\tau}_{s1}$ 0,511	$\hat{\tau}_{12}$ -0,291	$\hat{\tau}_{13}$ -1,054
etnibbeafroecuatoriano	$\hat{\tau}_{s2}$ 0,281	$\hat{\tau}_{22}$ -0,451	$\hat{\tau}_{23}$ -1,049
etnibbeotro	$\hat{\tau}_{s3}$ 0,248	$\hat{\tau}_{32}$ -0,245	$\hat{\tau}_{33}$ -1,052
etnibbemontubio	$\hat{\tau}_{s4}$ 0,147	$\hat{\tau}_{42}$ 0,040	$\hat{\tau}_{34}$ -0,092
Intercepts:	insuficiente	satisfactorio	excelente
	$\hat{\beta}_{10}$ -4,515	$\hat{\beta}_{20}$ 6,693	$\hat{\beta}_{30}$ 11,311
Residual Deviance:	563519.1		
AIC:	563597.1		

La modelización de las transformaciones *logit* son de la forma:

$$L_s(x) = -4,515 + 6,693 + 11,311 + 0,224\hat{\beta}_{11} - 0,407\hat{\beta}_{21} - 0,833\hat{\beta}_{31} - 1,344\hat{\tau}_{s1} + 0,665\hat{\tau}_{12} + 1,034\hat{\tau}_{13} - 0,453\hat{\tau}_{s2} + 0,148\hat{\tau}_{22} - 0,028\hat{\tau}_{23} - 0,611\hat{\tau}_{s3} - 0,075\hat{\tau}_{32} - 0,812\hat{\tau}_{33} - 0,992\hat{\tau}_{s4} + 0,111\hat{\tau}_{42} - 0,410\hat{\tau}_{34} - 0,660\hat{\tau}_{s1} + 0,955\hat{\tau}_{12} + 1,400\hat{\tau}_{13} - 0,159\hat{\tau}_{s2} + 0,667\hat{\tau}_{22} + 1,168\hat{\tau}_{23} - 0,042\hat{\tau}_{s1} + 0,147\hat{\tau}_{12} + 0,245\hat{\tau}_{13} + 0,511\hat{\tau}_{s1} - 0,291\hat{\tau}_{12} - 1,054\hat{\tau}_{13} + 0,281\hat{\tau}_{s2} - 0,451\hat{\tau}_{22} - 1,049\hat{\tau}_{23} + 0,248\hat{\tau}_{s3} - 0,245\hat{\tau}_{32} - 1,052\hat{\tau}_{33} + 0,147\hat{\tau}_{s4} + 0,040\hat{\tau}_{42} - 0,092\hat{\tau}_{34}$$

Los parámetros significativos para el estudio son: edad, en regiones naturales (región sierra, región oriente), autoidentificación étnica, sexo, tipo de financiamiento.

Los parámetros no significativos son: región otras y región insular, ya que al ser su p-valor mayor a 0.05 se acepta la hipótesis nula asumiendo que los dos parámetros deberían ser cero, por tal motivo no se extraen del modelo puesto que tienen parámetros no nulos.

El modelo logit se puede usar como test de diagnóstico para clasificar las categorías de la variable respuesta en función de los valores de la variable explicativa (tabla 3).

Tabla 3

Tabla de clasificación del modelo nominal

	insuficiente	elemental	satisfactorio	excelente
insuficiente	9163	41681	5854	0
elemental	6237	73911	36332	0
satisfactorio	1467	35860	58276	0
excelente	31	1282	5403	0

La tabla de clasificación establece cuantas veces se repite la predicción siendo así:

- Entre los individuos de logro alcanzado insuficiente, el modelo acierta en un 16,16%.
- Entre los individuos de logro alcanzado elemental, el modelo acierta en un 63,45%.
- Entre los individuos de logro alcanzado satisfactorio, el modelo acierta en un 20.39%.
- Entre los individuos de logro alcanzado excelente, el modelo acierta en un 0%.
- Entre todos los casos, el modelo acierta en un 51,31%.

Ajuste de regresión de respuesta ordinal

Para realizar el ajuste de regresión ordinal logit de respuesta ordinal con variables de diseño cualitativas y cuantitativas para datos agrupados con la finalidad de explicar el nivel de logro a partir de todas las variables, sin embargo hay que tener en cuenta cuáles son las categorías de referencia con las cuales se va a trabajar, siendo así: Para región natural, igual a Costa; para autoidentificación étnica mestizo/blanco, tipo de financiamiento público, tipo de sexo mujer y nivel de logro alcanzado elemental. El ajuste del modelo se presenta en la tabla 4.

Tabla 4
Datos agrupados Ser Bachiller 2018-2019

Coefficients:	Value	Std. Error	t value	P-value
nm_regisierra	1,182	0,0081	145,61	0,0000
nm_regiotras	0,374	0,0939	3,982	6,841e-05
nm_regioriente	0,302	0,0168	17,944	5,366e-72
nm_regiinsular	0,580	0,0933	6,223	4,892e-10
financiamientoprivado	1,105	0,0098	112,538	0,0000
financiamientomixto	0,644	0,0139	46,122	0,0000
tp_sexomujer	0,122	0,0073	16,614	7,383e-89
edad	-0,382	0,0026	-144,110	0,0000
etnibbeindigena	-0,469	0,0165	-28,338	1,18e-176
etnibbeafroecuatoriano	-0,506	0,0183	-27,622	6,09e-168
etnibbeotro	-0,363	0,0576	-6,296	3,054e-10
etnibbemontubio	-0,095	0,0191	-4,998	5,801e-07
Intercepts:	Value	Std, Error	t value	P-value
insuficiente elemental	-7,941	0,0508	-156,222	0,0000
elemental satisfactorio	-5,702	0,0496	-115,049	0,0000
satisfactorio excelente	-2,228	0,0503	-44,288	0,0000
Residual Deviance:	5648176,48			
AIC:	568206,48			

La modelización de las transformaciones *logit* son de la forma:

$$L_s(x) = -0,382 + 1,182X_{21} + 0,374X_{22} + 0,302X_{23} + 0,580X_{24} + 1,105X_{31} + 0,644X_{32} + 0,122X_{41} - 0,469X_{51} - 0,506X_{52} - 0,363X_{53} - 0,095X_{54}$$

Los parámetros significativos para el estudio son: En Regiones Naturales (región sierra, región otras, región oriente, región insular), Tipo de financiamiento (financiamiento privado, financiamiento mixto), Tipo de Sexo (sexo hombre), Edad, Autoidentificación étnica (etnia indígena, etnia ecuatoriana, etnia otro, etnia montubia).

Tabla 5
Tabla de clasificación del modelo nominal

	insuficiente	elemental	satisfactorio	excelente
insuficiente	9626	41752	5320	0
elemental	7469	74608	34403	0
satisfactorio	1638	38798	55167	0
excelente	31	1519	5166	0

- Entre los individuos de nivel de logro alcanzado insuficiente, el modelo acierta en un 16,98%.
- Entre los individuos de nivel de logro alcanzado elemental, el modelo acierta en un 64,05%.
- Entre los individuos de nivel de logro alcanzado satisfactorio, el modelo acierta en un 18,97%.
- Entre los individuos de nivel de logro alcanzado excelente, el modelo acierta en un 0%.
- Entre todos los casos, el modelo acierta en un 50,59%.

Conclusiones

- De tres a cinco conclusiones las más relevantes Mediante la aplicación del ajuste de respuesta nominal para la proyección del nivel de logro alcanzado se analizó la incidencia de las variables sociodemográficas en la probabilidad de que un estudiante de educación media obtenga un rendimiento académico insuficiente, elemental, satisfactorio o excelente en el examen Ser Bachiller en el periodo 2018-2019. Con la construcción del modelo de regresión logística y la bondad del ajuste se validó a través del método de máxima verosimilitud, donde al mostrar significancia estadística no se desestimó ninguna variable, aceptándose la hipótesis alternativa.
- Se comprobó que para el modelo de ajuste nominal y ordinal las variables utilizadas son las adecuadas, llegando a esta conclusión mediante la aplicación del método stepwise ya que la selección automática de las variables independientes como predictoras son las mismas.
- Se observa que en el modelo ordinal hay 15 parámetros como variables estimadas, mientras que para el modelo nominal hay 39 parámetros como variables por categoría estimadas, siendo así la diferencia entre el modelo nominal y el modelo ordinal.

Referencias bibliográficas

- Agresti, A. (2002). *Categorical Data Analysis*. New York: Willey.
- Aiken, L. (1996). *Tests psicológicos y evaluación*. México: Prentice Hall Hispanoamérica.
- Carrión, E. (2002). Validación de características al ingreso como predictores del rendimiento académico en la carrera de medicina. *Educación Media Superior*, 6.

- Guadagni, A. (2006). *Ingreso a la Universidad en Ecuador, Cuba y Argentina*.
http://www.rlcu.org.ar/recursos/E_0000046_004_cea_numero_44.pdf
- Hoster D., & L. (2000). *Applied Logistic Regression*. USA: Willey & Sons.
- Instituto Nacional de Estadísticas y Censos [INEC]. (2018). *INEC*.
<https://anda.inec.gob.ec/anda/index.php/catalog/266/datafile/F1/V4118>
- Instituto Nacional de Evaluación Educativa. (2018). *CIE-Resultados Educativos*. e
https://www.evaluacion.gob.ec/wp-content/uploads/downloads/2019/02/CIE_ResultadosEducativos18_20190109.pdf
- Instituto Nacional para la Evaluación de la Educación [INEE]. (2017). *GACETA*.
<https://www.inee.edu.mx/wp-content/uploads/2018/12/G07-esp.pdf>
- Jobson, J. (1991). *Categorical & multivariate methods*. USA: Springer verlag.
- Ministerio de Educación. (2015). *Ministerio de Educación*. <https://educacion.gob.ec/wp-content/uploads/downloads/2017/02/Reglamento-General-a-la-Ley-OrgAnica-de-Educacion-Intercultural.pdf>
- Organización de las Naciones Unidas para la Educación, Ciencia y Cultura [UNESCO]. (2019). *SITEAL. ECUADOR|PERFIL DE PAÍS (N.o 05/2019)*. SITEAL.
https://siteal.iiiep.unesco.org/sites/default/files/sit_informe_pdfs/dpe_ecuador_25_09_19.pdf
- Pando, F. V., & San Martín, F. R. (2004). *Regresión Logística Multinomial*. Esp. Cien.
- Romero, S. (2021). *Ecología Verde*. <https://www.ecologiaverde.com/cuales-son-las-regiones-naturales-del-ecuador-3269.html>
- Ruston, A. (2012). *Facultad de Ciencias Agrónomas*.
https://repositorio.uchile.cl/bitstream/handle/2250/120284/Ruston_Antonio_Estadistica_descriptiva.pdf
- Secretaría de Educación Superior [SENECYT]. (2019). *SENECYT*.
<https://admission.senescyt.gob.ec/etapa/autoasignacion-de-sede/>

El artículo que se publica es de exclusiva responsabilidad de los autores y no necesariamente reflejan el pensamiento de la **Revista Alfa Publicaciones**.



El artículo queda en propiedad de la revista y, por tanto, su publicación parcial y/o total en otro medio tiene que ser autorizado por el director de la **Revista Alfa Publicaciones**.



Indexaciones

