

MODELO LOGÍSTICO MULTINOMIAL CONDICIONES SOCIOECONÓMICAS DE LAS PERSONAS QUE HABITAN EN LA CIUDAD DE EL ALTO

MULTINOMIAL LOGISTIC MODEL SOCIOECONOMIC CONDITIONS OF PEOPLE THAT INHABIT IN EL ALTO CITY

M. Sc. Fernando Rivero Suguiura¹

Instituto de Estadística Teórica y Aplicada - UMSA, La Paz - Bolivia

✉ friverosuguiura2004@gmail.com

Artículo recibido: 2021-08-05

Artículo aceptado: 2021-09-13

RESUMEN

El objetivo es investigar las condiciones de vida de las personas que habitan en la ciudad de El Alto de La Paz, a partir de la referencia de algunas variables demográficas y socioeconómicas como son: edad, nivel de educación, seguridad alimentaria, ingreso, gasto mensual y ocupación, con la aplicación del modelo logístico multinomial que clasifica a la población en nivel socioeconómico alto, medio y bajo. Además, el modelo permite la medición probabilística de pertenecer a dichas categorías con mayor o menor influencia.

Palabras clave: *Condiciones socioeconómicas, modelo logístico multinomial, ciudad de El Alto.*

ABSTRACT

The objective is to investigate the living conditions of people living in El Alto city, located in the department of La Paz, based on the reference of some demographic and socioeconomic variables such as: age, education level, food security, income, monthly expenditure and occupation, implicating the multinomial logistic model that classifies the population into high, medium and low socioeconomic level. In addition, the model allows the probabilistic measurement of belonging to these categories with greater or lesser influence.

Keywords: *Socioeconomics conditions, multinomial logistic model, El Alto city*

INTRODUCCIÓN

La ciudad de El Alto se encuentra ubicada en la meseta del altiplano norte del departamento de La Paz en la cuarta sección de la provincia Murillo, al pie de la cordillera oriental a 12 km del centro de la ciudad de La Paz a una altura de 4.500 m.s.n.m. y tiene una extensión

de 387,56 km², según el Plan de Desarrollo de la Ciudad de El Alto. Ciudad creada el 6 de marzo de 1985 como urbe segunda más poblada de Bolivia con una población aproximada de 943.600 habitantes, donde el 51,4% son mujeres y el 48,6% son hombres, de estos, el 54,2% es menor de 30 años, según los datos del último Censo de Población y

¹ Docente de la carrera de Estadística, Facultad de Ciencias Puras y Naturales de la UMSA. Consultor en muestreo, censos y análisis estadístico en entidades nacionales e internacionales. Magister en Ciencias de la Estadística. <https://orcid.org/0000-0001-9095-7778>

Vivienda (INE, 2012).

Respecto al sector educativo, en 2018, la población matriculada en los niveles inicial, primaria y secundaria en la educación pública y privada llegó a 300.927 personas. El porcentaje de matriculados que cumplieron con el requisito mínimo para un curso inmediatamente superior (tasa de promoción) alcanza al 97%, los que abandonaron la escuela o colegio (tasa de abandono) fueron 1,6% y los que no cumplieron con la nota mínima de aprobación (tasa de reprobados) 1,3%. (El Alto en Cifras 2020 INE).

El Alto se caracteriza por su dinámico movimiento comercial y productivo basado en el crecimiento del sector de la micro, pequeña y mediana empresa y artesanía productiva, calificada por ello, como una de la segunda ciudad industrial de Bolivia (Chuquimia, 2008), tal es el caso de la feria 16 de Julio que se instaura domingo tras domingo ofreciendo el comercio de productos desde un alfiler hasta una maquinaria industrial sofisticada.

En la ciudad de El Alto se encuentran las principales vías de salida de las mercancías de exportación, siendo estas el Aeropuerto Internacional y la Zona Franca Industrial y Comercial. El movimiento de mercancías por estas aduanas llegó a un total de 1.047,7 millones de dólares en 2019, siendo la principal vía de salida el Aeropuerto de El Alto (INE, 2020).

El Producto Interno Bruto (PIB) per cápita de esta ciudad alcanza aproximadamente al 25% del PIB del departamento de La Paz que es aproximadamente de 16.558 millones de bolivianos y al 6% del PIB de Bolivia (Estrategia de Desarrollo Económico Local e Informe Estadístico del municipio de El Alto, 2020). Este porcentaje de aporte del

PIB, especialmente proviene del sector servicios, comercio y transporte, como el de manufactura en: minería, refinerías de petróleo, fábricas de azúcar y aceite, entre otros.

En el tema de pobreza, la ciudad de El Alto es una de las ciudades metrópoli más pobres de Bolivia, según el índice de Necesidades Básicas Insatisfechas (NBI) del Censo de Población y Vivienda del año 1992 al 2001, según el INE, reduce de un 73,8% a un 66% y posteriormente en el año 2005 alcanza a un 47,5%. Respecto a la desigualdad, se relaciona principalmente con la falta de servicios básicos como: alcantarillado, agua, energía eléctrica y vivienda propia.

En el tema de desigualdad habitacional, el Índice de Calidad de Vivienda reporta que el 52% de la población alteña cuenta con una vivienda catalogada como de nivel alto, el 47% de nivel medio y 1% de nivel bajo, sin embargo, este incremento en calidad no deja en expectativa la desigualdad de los habitantes de esta ciudad en este tema (INE, 2005).

A todo lo anteriormente señalado, la investigación se concluye con un estudio de las condiciones de vida de los pobladores de la ciudad de El Alto en base a información recopilada de la Encuesta de Hogares del INE Bolivia año 2018, mediante la aplicación del modelo logístico multinomial. Este último permite un análisis exhaustivo del tema de acuerdo al uso de algunas características sociales y económicas.

METODOLOGÍA

Modelo logístico multinomial

La regresión logística multinomial, es

Modelo logístico multinomial. Condiciones socioeconómicas de las personas que habitan en la ciudad de El Alto

utilizada en modelos con variable dependiente de tipo nominal con más de dos categorías y es una extensión multivariante del modelo logístico binario. Las variables independientes pueden ser tanto continuas como categóricas o no métricas. En esta aplicación sobre las condiciones socioeconómicas de las personas habitantes de la ciudad de El Alto, se construye la variable dependiente condición socioeconómica, mediante el método multivariante *cluster* análisis no jerárquico con tres categorías (alta, media y baja). Esta variable o factor, considera las subvariables tales como: a) materiales de construcción de la vivienda (techo, piso, pared); b) hacinamiento (número de personas por dormitorio); c) agua y saneamiento básico (disponibilidad de agua, alcantarillado, baño); d) insumos energéticos (combustible, energía eléctrica entre otros); e) salud y educación (acceso, atención de la salud por personal calificado, años de escolaridad y alfabetismo), fuente Encuesta de Hogares (INE, 2018).

Formulación del modelo

Sea la variable dependiente Y categórica nominal politómica con probabilidades p_1, p_2, \dots, p_k para las k categorías que compone la variable Y . Si se requiere el análisis del efecto de variables independientes X_1, X_2, \dots, X_p se define el modelo siguiente:

$$p_c = \frac{1}{1 + e^{-X\beta}} \quad (01)$$

haciendo algunos cambios en (01) se tiene

$$\begin{aligned} p_c &= \frac{1}{1 + \frac{1}{e^{X\beta}}} = \frac{1}{\frac{1 + e^{X\beta}}{e^{X\beta}}} \\ &= \frac{e^{X\beta}}{1 + e^{X\beta}} \end{aligned} \quad (02)$$

donde p_c es la probabilidad de que la categoría $Y = c$; $c = 1, 2, \dots, k$ se dé, con $\sum_{c=1}^k p_c = 1$. Además, en (02) se tiene que:

$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{p1} \\ 1 & x_{12} & \dots & x_{p2} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 1 & x_{1n} & \dots & x_{pn} \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_p \end{pmatrix}$$

Donde X es la matriz de dimensión $n \times (p+1)$ de observaciones y variables, como β que es el vector de dimensión $p+1$ de parámetros a estimar en el modelo. Si $q_c = 1 - p_c$ es la probabilidad complementaria de p_c , tal que $p_c + q_c = 1$, entonces:

$$\begin{aligned} q_c &= 1 - p_c \\ &= 1 - \frac{e^{X\beta}}{1 + e^{X\beta}} \end{aligned} \quad (03)$$

resulta de reemplazar (02) en (03). Luego

$$q_c = \frac{1 + e^{X\beta} - e^{X\beta}}{1 + e^{X\beta}} = \frac{1}{1 + e^{X\beta}}$$

La razón entre p_c/q_c se define como la razón

$$r = \frac{p_c}{q_c} = \frac{\frac{e^{X\beta}}{1 + e^{X\beta}}}{\frac{1}{1 + e^{X\beta}}} = e^{X\beta}$$

considerándose un modelo no lineal.

Método de estimación de máxima verosimilitud

La función de verosimilitud, está definida como:

$$L = \prod_{i=1}^n (p_{1i}^{Y_{1i}} p_{2i}^{Y_{2i}} \dots p_{ki}^{Y_{ki}}) = \prod_{i=1}^n \left[\left(\frac{p_{1i}}{p_{ki}} \right)^{Y_{1i}} \left(\frac{p_{2i}}{p_{ki}} \right)^{Y_{2i}} \dots \left(\frac{p_{k-1,i}}{p_{ki}} \right)^{Y_{k-1,i}} p_{ki} \right] \quad (04)$$

Aplicando a la función de razón de verosimilitud generalizada se tiene

$$\Lambda = -2\ln(L) = -2 \sum_{i=1}^n \left[Y_{1i} \ln \left(\frac{p_{1i}}{p_{ki}} \right) + Y_{2i} \ln \left(\frac{p_{2i}}{p_{ki}} \right) + \dots + Y_{k-1,i} \ln \left(\frac{p_{k-1,i}}{p_{ki}} \right) + \ln(p_{ki}) \right] \quad (05)$$

La maximización en (04) es equivalente a minimizar (05). Por la complejidad de la función Λ , esta se resuelve por métodos numéricos de forma iterativa para los estimadores $\hat{\beta}$ del vector β .

Significación del modelo

Para probar si el modelo es significativo o no en su estructura global, se aplica el test estadístico de contraste de hipótesis, teniendo en cuenta que la diferencia entre el valor inicial y el valor final de la función de razón de verosimilitud generalizada Λ tiene distribución Chi-cuadrado con grados de libertad igual al número de regresores multiplicado por el número de categorías menos uno.

La hipótesis nula (H_0) de que no existe efecto de las variables regresoras, contra la hipótesis alterna (H_1), se define como:

$$\begin{aligned} H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0 \\ H_1: \beta_1 \neq \beta_2 \neq \dots \neq \beta_p \neq 0 \end{aligned} \quad (06)$$

El p' valor del test para H_0 vendrá dado por la siguiente probabilidad

$$p' = P(\chi_{(p-1)(k-1)}^2 > \Lambda_0 - \Lambda_1)$$

se rechaza H_0 si $p' < 0,05$.

RESULTADOS

Aplicación del modelo logístico multinomial

La aplicación del modelo logístico multinomial definido anteriormente, considera las siguientes variables.

La variable 1 (Clasific_Soc) corresponde ser la dependiente del modelo con las categorías: 1 (Alto), 2 (Medio) y 3 (Bajo). Las siguientes variables de la 2 a la 5 son de condición categórica: género, edad, nivel de educación y seguridad alimentaria. Las variables 6 a la

Modelo logístico multinomial. Condiciones socioeconómicas de las personas que habitan en la ciudad de El Alto

Tabla No. 1
Variables para el modelo

Nro.	Variable	Condición	Descripción	Categoría
1	Clasific_Soc	Dependiente	Clasificación Socioeconómica	1. Alto 2. Medio 3. Bajo
2	Gen	Categórica	Genero o Sexo	0. Mujer 1. Hombre
3	Edad	Categórica	Edad en años	1. de 0 a 19 2. de 20 a 39 3. de 40 y más
4	Niv_educ	Categórica	Nivel Educativo	1. Ninguno 2. Básico 3. Superior
5	Seg_alim	Categórica	Seguridad Alimentaria	1. Inseguridad Severa y Moderada 2. Inseguridad Leve 3. Seguridad
6	I_h	Covariable	Ingreso del Hogar Mensual en Bs	
7	G_h	Covariable	Gasto del Hogar Mensual en Bs	
8	P_ocup	Covariable	Porcentaje de Ocupados por Hogar	

Fuente: Encuesta de Hogares 2018 (INE), elaboración propia.

8, son cuantitativas o llamadas también covariables tales como: ingreso y gasto del hogar mensual en Bs y el porcentaje de ocupados, estas últimas no categorizadas.

Especificación del modelo

Inicialmente se realiza la categorización de las variables independientes o explicativas para ser consideradas en el modelo logístico. Para ello, en algunos casos, se crean variables ficticias o *dummy* (dicotómicas), 1 si pertenece a la categoría y 0 cuando no pertenece, es decir:

$$X_j = \begin{cases} 1 & \text{si pertenece a la categoría} \\ 0 & \text{si no pertenece a la categoría} \end{cases}$$

X_3 : Porcentaje de ocupados del hogar

X_4 : Género mujer

X_5 : Género hombre

X_6 : Edad entre 0 y 19 años

X_7 : Edad entre 20 y 39 años

X_8 : Edad de 40 y más años

X_9 : Nivel educación ninguno

X_{10} : Nivel educación básico

X_{11} : Nivel educación superior

X_{12} : Seguridad alimentaria (inseguridad severa y moderada)

X_{13} : Seguridad alimentaria (inseguridad leve)

X_{14} : Seguridad alimentaria (seguridad)

Las variables cuantitativas ingreso y gasto del hogar siguen un proceso de estandarización para que sean comparables ambas, mediante el siguiente método de transformación de variable Z_j

$$Z_j = \frac{X_j - \bar{X}_j}{S_j} \quad j = 1, 2$$

Donde X_j tiene la media y desviación estándar muestral (\bar{X}_j , S_j). Por lo que se cuenta con las siguientes dos variables:

Z_1 : Ingreso del hogar mensual en Bs, estándar

Z_2 : Gasto del hogar mensual en Bs, estándar

Luego el modelo que genera las probabilidades de categorización, está dado por:

$$p = \frac{e^{\beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} + \beta_{11} X_{11} + \beta_{12} X_{12} + \beta_{13} X_{13} + \beta_{14} X_{14}}}{1 + e^{\beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} + \beta_{11} X_{11} + \beta_{12} X_{12} + \beta_{13} X_{13} + \beta_{14} X_{14}}} \quad (07)$$

con la variable dependiente formulada como

$$Y = X\beta$$

$$= \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} + \beta_{11} X_{11} + \beta_{12} X_{12} + \beta_{13} X_{13} + \beta_{14} X_{14}$$

luego (07) se puede escribir como

$$p_c = \frac{e^Y}{1 + e^Y}$$

X la matriz de dimensión $n=3.965$ (muestra de hogares ciudad de El Alto) por $p+1=15$ de variables, y β el vector de dimensión $p+1=15$ de parámetros a estimar en el modelo. Sin embargo, la regresión logística multinomial presenta $k-1$ modelos desagregados de acuerdo a la cantidad de categorías menos uno que tiene la variable dependiente Y , en este caso dos modelos diferentes.

Estimación y significación del modelo logístico

Aplicando la función de razón de verosimilitud generalizada (Λ), expresión (05), y realizando la maximización por métodos numéricos de forma iterativa, se consigue estimar el modelo logístico con coeficientes $\hat{\beta}$.

Se han probado diferentes combinaciones de variables categóricas y covariables, además de la inclusión y exclusión de éstas, para determinar el modelo más significativo a los datos analizados. Se han analizado 12 modelos posibles con las 14 variables y categorías descritas anteriormente, donde la variable **Genero** con categorías **mujer** y **hombre** parecían influyentes en el resultado,

sin embargo, se pudo observar que dicha variable no es relevante en el modelo, ignorándola y quedando con el modelo final estimado en su versión 12, siguiente.

Los estimadores de β para las variables y categorías consideradas en los modelos, se presentan en la segunda columna de la Tabla No. 2. Se puede observar, que el coeficiente independiente $\hat{\beta}_0$ es negativo cercano a -3 para el primer caso, sin embargo para el segundo caso es -0.24 y poco significativo individualmente. Los coeficientes estimados de las variables cuantitativas Ingreso y Gasto estándar (Z_1 , Z_2), para ambas opciones de la variable dependiente Y de los modelos Bajo y Medio, son negativos y significativos según el estadístico de contraste de Wald (valor alto) y probabilidad Sig, presente en la 6ta columna que se mide por debajo de 0,05. La variable porcentaje de ocupados por hogar (X_3) en ambas categorías de Y son relativamente cercanas a cero, pero significativas en ambos modelos, de acuerdo a los estadísticos Wald y Sig.

La columna siete de la tabla No. 2 presenta los siguientes resultados más relevantes:

Modelo logístico multinomial. Condiciones socioeconómicas de las personas que habitan en la ciudad de El Alto

Tabla No. 2
Estimadores y características de los coeficientes β

VARIABLE DEPENDIENTE CLASIFICACIÓN SOCIOECONÓMICA ^a		Estimaciones de parámetro					95% de intervalo de confianza para Exp(B)		
		B	Error estándar	Wald	gl	Sig.	Exp(B)	Límite inferior	Límite superior
Bajo	Intersección	-2,854	0,227	157,367	1	0,000			
	Ingreso estandarizado	-1,100	0,096	131,260	1	0,000	0,333	0,276	0,402
	Gasto estandarizado	-1,326	0,108	151,441	1	0,000	0,266	0,215	0,328
	PORCENTAJE DE OCUPADOS POR HOGAR	0,008	0,002	10,455	1	0,001	1,008	1,003	1,012
	[Edad recodificada dos=1]	0,171	0,137	1,568	1	0,210	1,187	0,908	1,551
	[Edad recodificada dos=2]	0,652	0,138	22,392	1	0,000	1,920	1,465	2,516
	[Edad recodificada dos=3]	0 ^b			0				
	[NIVEL DE EDUCACIÓN GENERAL=1,00]	2,038	0,232	77,250	1	0,000	7,673	4,871	12,088
	[NIVEL DE EDUCACIÓN GENERAL=2,00]	1,035	0,170	37,167	1	0,000	2,816	2,019	3,927
	[NIVEL DE EDUCACIÓN GENERAL=3,00]	0 ^b			0				
	[CODIGO SEGURIDAD ALIMENTARIA=1]	1,829	0,199	84,286	1	0,000	6,229	4,215	9,205
	[CODIGO SEGURIDAD ALIMENTARIA=2]	0,780	0,113	47,254	1	0,000	2,181	1,746	2,724
	[CODIGO SEGURIDAD ALIMENTARIA=3]	0 ^b			0				
Medio	Intersección	-0,236	0,148	2,539	1	0,111			
	Ingreso estandarizado	-0,573	0,058	96,729	1	0,000	0,564	0,503	0,632
	Gasto estandarizado	-0,516	0,060	73,214	1	0,000	0,597	0,530	0,672
	PORCENTAJE DE OCUPADOS POR HOGAR	0,006	0,002	11,304	1	0,001	1,006	1,003	1,010
	[Edad recodificada dos=1]	0,123	0,105	1,373	1	0,241	1,131	0,921	1,388
	[Edad recodificada dos=2]	0,188	0,105	3,200	1	0,074	1,207	0,982	1,482
	[Edad recodificada dos=3]	0 ^b			0				
	[NIVEL DE EDUCACIÓN GENERAL=1,00]	0,780	0,175	19,960	1	0,000	2,182	1,550	3,073
	[NIVEL DE EDUCACIÓN GENERAL=2,00]	0,232	0,111	4,342	1	0,037	1,261	1,014	1,000
	[NIVEL DE EDUCACIÓN GENERAL=3,00]	0 ^b			0				
	[CODIGO SEGURIDAD ALIMENTARIA=1]	0,809	0,180	20,124	1	0,000	2,246	1,577	
	[CODIGO SEGURIDAD ALIMENTARIA=2]	0,302	0,091	10,941	1	0,001	1,352	1,131	
	[CODIGO SEGURIDAD ALIMENTARIA=3]	0 ^b			0				

^a La categoría de referencia es: Alto.
^b El parámetro está establecido en cero por que es redundante.

Fuente: Modelo logístico multinomial SPSS.

- $\exp(\beta) = \exp(2,038) = 7,673$

Para el primer modelo, significa que pertenecer a una condición socioeconómica

baja frente a una condición socioeconómica alta, de personas que no tienen ninguna educación, están en 7,7 veces peor de los

que tienen educación básica. El valor real se encuentra en el intervalo de confianza al 95% entre (4,9; 12,1). Asimismo, se tiene que

- $\exp(\beta) = \exp(1,829) = 6,229$

para el primer modelo, significa que pertenecer a una condición socioeconómica baja frente a una condición socioeconómica alta, de personas que tienen inseguridad alimentaria severa y moderada, están en 6,2 veces peor de los que tienen inseguridad alimentaria leve. Otro caso, se tiene

- $\exp(\beta) = \exp(0,78) = 2,182$

para el segundo modelo, que pertenecer a una condición socioeconómica media frente a una condición socioeconómica alta, de personas que no tienen ninguna educación, están en 2,2 veces peor de los que tienen educación básica. El valor real se encuentra en el intervalo de confianza al 95% entre (1,6; 3,1). Finalmente

- $\exp(\beta) = \exp(0,809) = 2,246$

para el segundo modelo, significa que pertenecer a una condición socioeconómica media frente a una condición socioeconómica alta, de personas que tienen inseguridad alimentaria severa y moderada, están en 2,2 veces peor de los que tienen inseguridad

alimentaria leve.

Así se pueden interpretar los demás coeficientes $\exp(\beta)$ con la compañía de sus intervalos de confianza.

Test de hipótesis de β individual

La prueba individual de los coeficientes β , se realiza en base al estadístico de Wald como se presenta en la Tabla No. 2, definido como

$$H_0: \beta_{jc} = 0 \text{ vs } H_1: \beta_{jc} \neq 0 \quad j = 1, 2, \dots, 9 \quad c = 1, 2$$

Se supone distribución Normal en β

$$Z = \frac{\hat{\beta}_{jc} - \beta_{jc}}{\hat{\sigma}_{\hat{\beta}_{jc}}} \approx N(0, 1)$$

El estadístico de contraste de hipótesis, denominado de Wald, es

$$W = Z^2 \approx \chi^2_{1-\frac{\alpha}{2}}$$

Se rechaza H_0 si $W > \chi^2_{1-\frac{\alpha}{2}}$. Es decir, si el estadístico de Wald es superior al valor chi-cuadrado con 1 grado de libertad.

Los coeficientes estimados de β a nivel de variables con categoría **Baja** de la variable dependiente Socioeconómica (Y), son significativos excepto el coeficiente $\hat{\beta}_{6,1}$ **Edad recodificada = 1** no es significativo y

Tabla No. 3
Test conjunto de los coeficientes estimados de β

Información de ajuste de los modelos						
Modelo	Criterios de ajuste de modelo			Pruebas de la razón de verosimilitud		
	AIC	BIC	Logaritmo de la verosimilitud - 2	Chi-cuadrado	gl	Sig.
Sólo intersección	8202,238	8214,809	8198,238			
Final	7207,584	7333,289	7167,584	1030,654	18	,000

Fuente: Modelo logístico multinomial SPSS.

Modelo logístico multinomial. Condiciones socioeconómicas de las personas que habitan en la ciudad de El Alto

las categorías de las variables en el caso 0^b , para el modelo desagregado denominado categoría **Baja**.

$$H_1: \beta_1 \neq \beta_2 \neq \dots \neq \beta_p \neq 0$$

El nivel de significación p' valor del test para H_0 vendrá dado por la siguiente probabilidad

$$p' = P(\chi_{(p-1)(k-1)}^2 > \Lambda_0 - \Lambda_1)$$

Para la cualidad de **Y Media**, se observa que las categorías de los coeficientes $\hat{\beta}_{0,2}$ de Intersección y $\hat{\beta}_{6,2}$ **Edad recodificada = 1** no son significativos con Sig por encima de 0,05 y los identificados con 0^b de la Tabla No. 2.

se rechaza H_0 si $p' < 0,05$.

La Tabla No. 3, proporciona información sobre el ajuste de los modelos en forma global mediante la hipótesis de que los coeficientes de β sean significativos de manera conjunta, es decir:

La Tabla No. 3 en la columna **Pruebas de la razón de verosimilitud** demuestra que el valor del estadístico chi cuadrado es alto con p^{\wedge} (Sig) tendiente a cero, por lo cual los modelos de las cualidades Baja y Media son altamente significativas de manera conjunta. El criterio de Akaike (AIC) justifica la medición relativa de calidad de ajuste de los modelos empleados.

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$$

Los modelos estimados, son:

- **Cualidad Baja modelo 1**

Según lo señalado, la razón entre la probabilidad \hat{p}_1 y \hat{p}_3 se define como

$$r_1 = \frac{\hat{p}_1}{\hat{p}_3} = e^{X\hat{\beta}_1}$$

$$\begin{aligned} r_1 &= e^{\hat{\beta}_{0,1} + \hat{\beta}_{1,1}Z_1 + \hat{\beta}_{2,1}Z_2 + \hat{\beta}_{3,1}X_3 + \hat{\beta}_{6,1}X_6 + \hat{\beta}_{7,1}X_7 + \hat{\beta}_{9,1}X_9 + \hat{\beta}_{10,1}X_{10} + \hat{\beta}_{12,1}X_{12} + \hat{\beta}_{13,1}X_{13}} \\ &= e^{-2,854 - 1,1Z_1 - 1,326Z_2 + 0,008X_3 + 0,171X_6 + 0,652X_7 + 2,038X_9 + 1,035X_{10} + 1,829X_{12} + 0,78X_{13}} \end{aligned}$$

- **Cualidad Media modelo 2**

De igual manera la razón entre la probabilidad \hat{p}_2 y \hat{p}_3 es

$$r_2 = \frac{\hat{p}_2}{\hat{p}_3} = e^{X\hat{\beta}_2}$$

$$\begin{aligned} r_2 &= e^{\hat{\beta}_{0,2} + \hat{\beta}_{1,2}Z_1 + \hat{\beta}_{2,2}Z_2 + \hat{\beta}_{3,2}X_3 + \hat{\beta}_{6,2}X_6 + \hat{\beta}_{7,2}X_7 + \hat{\beta}_{9,2}X_9 + \hat{\beta}_{10,2}X_{10} + \hat{\beta}_{12,2}X_{12} + \hat{\beta}_{13,2}X_{13}} \\ &= e^{-0,236 - 0,573Z_1 - 0,516Z_2 + 0,006X_3 + 0,123X_6 + 1,88X_7 + 0,78X_9 + 0,232X_{10} + 0,809X_{12} + 0,302X_{13}} \end{aligned}$$

Donde las probabilidades de ocurrencia de clasificación cualidad baja, media y alta se determinan mediante las siguientes relaciones:

$$\hat{p}_1 = \frac{r_1}{1 + r_1 + r_2}$$

$$\hat{p}_2 = \frac{r_2}{1 + r_1 + r_2}$$

y la probabilidad de \hat{p}_3 se determina por el complemento de las anteriores, es decir

$$\hat{p}_3 = 1 - \hat{p}_1 - \hat{p}_2$$

tal que $\hat{p}_1 + \hat{p}_2 + \hat{p}_3 = 1$.

Clasificación observada y esperada

La Tabla No. 4 muestra la clasificación de individuos que fueron observados en las categorías condición socioeconómica alta, media y baja y los valores esperados por el modelo logístico multinomial, es decir:

Tabla No. 4
Clasificación observada y esperada

Observado	Clasificación Pronosticado			Porcentaje correcto
	Bajo	Medio	Alto	
Bajo	191	654	24	22,0%
Medio	147	1574	233	80,6%
Alto	14	774	354	31,0%
Porcentaje global	8,9%	75,7%	15,4%	53,4%

Fuente: Modelo logístico multinomial SPSS.

De acuerdo a la Tabla N° 4, el 53,4% de las personas clasifican tanto en observación y pronóstico por el modelo en las categorías baja, media y alta de condición socioeconómica. El 80,6% de las personas están en la categoría media observada y el 75,7% en la misma categoría de acuerdo al modelo. Hay un 46,6% de la población que no son coincidentes en la observación y proyección del modelo, sin embargo, no dejan de aproximarse a los resultados en la mayoría de los casos en un 99%.

RESULTADOS Y DISCUSIÓN

Al no contar con una variable categórica observada dependiente para el modelo logístico, denominada condición socioeconómica en los niveles alto, medio y bajo, se ha procedido a su determinación a partir del análisis *cluster* no jerárquico en base a las variables socioeconómicas: materiales de construcción de la vivienda, hacinamiento, agua y saneamiento básico, y otras, que son parte de las que componen el índice de necesidades básicas insatisfechas para la medición de pobreza estructural. Se puede notar luego, que al modelo logístico se le incorpora variables independientes que tienen relación con las condiciones de vida y pobreza, como: la ocupación, nivel de educación, seguridad alimentaria y otras que son de característica coyuntural como el ingreso y gasto. Al respecto, existe una predicción del modelo en un 53% en las categorías socioeconómicas de alta, media y baja para la población de la ciudad de El Alto con respecto a lo que presenta la variable observada. Sin embargo, esto no basta pues también el modelo proporciona la medición de la probabilidad de pertenecer a dichas categorías.

RECOMENDACIÓN

A lo referido anteriormente, es fundamental ampliar el análisis no solo a la categorización que pertenece la población en condición alta, media y baja; sino agrupar, además, cada condición, por intervalos de medición de probabilidad dadas por el modelo logístico e investigar a profundidad, cuál o cuáles de las variables influyen más a la determinación de la condición socioeconómica de la persona.

REFERENCIAS BIBLIOGRÁFICAS

- Agresti, A. (1990). *Categorical Data Analysis*. John Wiley and Sons. New York.
- Arguello, O. CELADE. *Desarrollo Económico, Políticas Sociales y Población*. CELADE.
- Arguello, O. (1980). *Pobreza y Desarrollo. Características Socio-Demográficas de las Familias Pobres en Venezuela*. Santiago de Chile. Centro Latinoamericano de Demografía.
- Arias, O. S., Bendini, M. (2006). *Evaluación de la pobreza en Bolivia: Sentando las bases para un crecimiento a favor de los pobres*. Programa Operacional y Analítico de la Región de América Latina y el Caribe, Banco Mundial.
- Ayaviri, N, D., Alarcón, L, S. (2014). *Clasificación socioeconómica de los municipios de Bolivia*. Universidad Católica Boliviana “San Pablo”, Unidad Académica Regional Cochabamba.
- Banco de Desarrollo de America Latina (CAF). (2017). *Agua y saneamiento en el Estado Plurinacional de Bolivia*. Buenos Aires.
- CELADE. (2002). *Vulnerabilidad sociodemográfica: viejos y nuevos riesgos para comunidades, hogares y personas*. Brasilia. CEPAL.
- CELADE. (2005). *Dinámica demográfica y desarrollo en América Latina y el Caribe*. Santiago de Chile. CEPAL.
- Cox, D. R. & Snell, E. J. (1989). *The Analysis of Binary Data*. Chapman and Hall. London.
- De la Fuente, F, S. (2011), *Análisis de conglomerados*. Universidad Autónoma de Madrid (UAM).
- UDAPE (2019). *Dossier de Estadísticas Económicas y Sociales*. Vol. 29. La Paz.
- INE Bolivia (2020). *El Alto en cifras*.
- Estado Plurinacional de Bolivia, Ministerio de Educación. (2010). *Ley de la Educación “Avelino Siñani-Elizardo Pérez”*.
- Gobierno Autónomo Municipal de El Alto. Secretaría Municipal de Desarrollo Económico del municipio de El Alto (2018). *Fortalecimiento a las iniciativas económicas: Localización Distrital*. El Alto.
- Hair, J, F., Anderson, R, E., Tatham, R, L., Black, W,C. (1999). *Análisis Multivariante*. Madrid, España. Editorial Prentice Hall.
- Hosmer, D. W. & Lemeshow, S. (1989). *Applied Logistic Regression*. Wiley Interscience. New York.
- Informe Estadístico del Municipio de El Alto (2020). Ministerio de Desarrollo Productivo y Económico Plural.
- Johnson, R. A. *Applied Multivariate Statistical Analysis*. Prentice Hall. New Jersey.
- López, J.L. (2011). *La Ecuación Diferencial Logística*.
- Medina, M. E. (2003). *Modelos de elección discreta*.
- Menard, S. (2000). *Coefficients of*

- Determination for Multiple Logistic Regression Analysis. *The American Statistician*.
- Osorio, D. (2009). Planteamiento del Modelo Logístico Multinomial a través de la Función Canónica de Enlace de la Familia Exponencial.
- Peña, Daniel (2002). *Análisis de Datos Multivariantes*. Mc Graw Hill.
- Quispe, G.M. (2009). La formación de la ciudad de El Alto y sus consecuencias. Tesis Doctoral, Universidad Autónoma de Madrid.
- UDAPE: Unidad de Análisis de Políticas Sociales y Económicas, INE: Instituto Nacional de Estadística. (2018). Migración interna en Bolivia. Organización Internacional para las Migraciones (OIM).
- Uriel, E., Aldás, J. (2005). *Análisis multivariante aplicado*. Madrid, España. Editorial Thomson.
- Villaroel, P., Hernani-Limarino, W. (2013). La evolución de la pobreza en Bolivia: un enfoque multidimensional. *Revista Latinoamericana de Desarrollo Económico (LAJED)*.