

Modelo Predictivo de Regresión Lineal Múltiple: ZM Coatzacoalcos

El modelo predictivo utilizado para realizar el presente análisis es el de regresión lineal múltiple (RLM)¹ para la Zona Metropolitana de Coatzacoalcos. Resulta plausible la aplicación de este tipo de análisis ya que existe más de una variable predictora.

En efecto, la pertinencia del modelo se explica en la medida que nuestra propuesta contempla 10 variables predictoras o independientes² y una variable dependiente denominada “Índice de Violencia Escolar” (ver tabla 1).

TABLA 1: VARIABLES INTRODUCIDAS/ELIMINADAS			
Modelo	Variables introducidas	Variables eliminadas	Método
1	Edad, Sexo_femenino_dummy, Turno_matutino_dummy Nivel_primaria_dummy^b ÍNDICE DE REPRESENTACIONES SOCIALES NEGATIVAS ÍNDICE DE BARRIO ESCOLAR NEGATIVO ÍNDICE DE CLIMA ESCOLAR NEGATIVO ÍNDICE DE INCIVILIDADES ÍNDICE DE GESTIÓN ESCOLAR NEGATIVA ÍNDICE DE POLÍTICAS PÚBLICAS DEFICIENTES		Introducir
a. Variable dependiente: ÍNDICE DE VIOLENCIA ESCOLAR			
b. Todas las variables solicitadas introducidas.			

Fuente: Elaborado por el autor con información de base de datos SPSS

El objetivo de este análisis “es estimar el valor medio de y (variable dependiente) y/o predecir valores particulares de y, basado en la información de k variables independientes o predictoras x’s” (Santana, s/a: 171).

¹Cuando la predicción se realiza en base a solo una variable independiente se habla de regresión lineal simple, no obstante, cuando se trabaja con más de una variable independiente se habla de regresión lineal múltiple.

² El conjunto de variables independientes será denominado como factores de riesgo de violencia en las escuelas.

A continuación se presenta un completo análisis del modelo propuesto.

Resumen del modelo ^b										
Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación	Estadísticos de cambio					Durbin-Watson
					Cambio en R cuadrado	Cambio en F	gl1	gl2	Sig. Cambio en F	
1	.630 ^a	.397	-.205	3.61821	.397	.659	10	10	.739	2.789
a. Variables predictoras: (Constante), Nivel_primaria_dummy, Sexo_femenino_dummy, INDICE DE REPRESENTACIONES SOCIALES NEGATIVAS, INDICE DE INCIVILIDADES, ÍNDICE DE BARRIO ESCOLAR, ÍNDICE DE CLIMA ESCOLAR, ÍNDICE DE GESTIÓN ESCOLAR NEGATIVA, Edad, Turno_matutino_dummy, ÍNDICE DE POLÍTICAS PÚBLICAS										
b. Variable dependiente: ÍNDICE DE VIOLENCIA ESCOLAR										

Fuente: Elaborado por el autor con información de base de datos SPSS

La tabla 2 nos presenta el valor de R, este “coeficiente mide el grado de asociación entre una variable y (dep) y un grupo de variables $x_1, x_2, \dots, x_k$, (ind) (...) el coeficiente R no toma valores menores que cero” (Santana, s/a: 176-177). En este caso el valor que toma R es **0,630**, lo cual indica que existe una asociación (lineal) **alta**³ entre las variables. Por lo tanto, existe una asociación significativa (alta) entre el índice de Violencia Escolar (VD) y el conjunto de variables independientes que han sido propuestas.

El coeficiente R cuadrado se interpreta como la proporción de la varianza de la variable dependiente explicada por el conjunto de variables independientes. El valor del coeficiente R cuadrado es **0,397**, esto indica que podemos predecir en un **76,7% el índice de Violencia Escolar** en la Zona Metropolitana de Coatzacoalcos.

El R cuadrado ajustado es una corrección del coeficiente de determinación que sirve para controlar la sensibilidad de este estadístico cuando las muestras son de gran tamaño, en este caso el R cuadrado ajustado es **-0,205(-20.5%)**.

³ Baja: 0.1 a 0.3 Media: 0.3 a 0.5 Alta: 0.5 a 1

Ahora bien, vamos a someter a prueba la hipótesis nula (H_0) de que el R es 0 en la población. Para esto realizamos una prueba de análisis de varianza, de la cual observamos **la significación del cambio en F (0,739)**, puesto que su valor es mayor a 0,05 **no rechazamos la hipótesis nula** de que R es 0 en la población.

El estadístico de **Durbin- Watson** (tabla 2) informa sobre el grado de independencia entre los residuos, puede asumir valores entre 0 y 4. Sin embargo, si tiene un valor entre 1,5 y 2,5, se asume independencia. De tal manera podemos afirmar que **no se asume** independencia entre los residuos **(2,789)**.

A partir de estos coeficientes es importante señalar **que nuestro modelo no se encuentra ajustado**, puesto que no se cumplen los criterios mínimos de aplicación.

ANOVA ^a						
Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	86.323	10	8.632	.659	.739
	Residual	130.915	10	13.091		
	Total	217.238	20			
a. Variable dependiente: ÍNDICE DE VIOLENCIA ESCOLAR						
b. Variables predictoras: (Constante), Nivel_primaria_dummy, Sexo_femenino_dummy, INDICE DE REPRESENTACIONES SOCIALES NEGATIVAS, INDICE DE INCIVILIDADES, ÍNDICE DE BARRIO ESCOLAR, ÍNDICE DE CLIMA ESCOLAR, ÍNDICE DE GESTIÓN ESCOLAR NEGATIVA, Edad, Turno_matutino_dummy, ÍNDICE DE POLÍTICAS PÚBLICAS						

Fuente: Elaborado por el autor con información de base de datos SPSS.

Por su parte, la **tabla 3** nos indica Anova, la cual contrasta Hipótesis nula (H_0) de que R es 0 en la población. La significación que nos indica la tabla es **0,739**, de esto se interpreta que no rechazamos la hipótesis nula de que R es 0 en la población. **Por lo tanto, nuestro modelo para la ZM de Coatzacoalcos no se encuentra ajustado.**

A partir de esta información podemos asumir con un 95% de confianza **que no existe una relación lineal significativa** entre la Variable Dependiente (Índice de Violencia Escolar) y el conjunto de variables independientes del modelo.

Coeficientes ^a											
Modelo		Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.	Intervalo de confianza de 95.0% para B		Correlaciones		
		B	Error típ.	Beta			Límite inferior	Límite superior	Orden cero	Parcial	Semiparcial
1	(Constante)	5.398	28.044		.192	.851	-57.089	67.885			
	ÍNDICE DE CLIMA ESCOLAR	.486	.597	.335	.813	.435	-.845	1.816	.426	.249	.200
	INDICE DE INCIVILIDADES	.227	.894	.101	.254	.804	-1.765	2.219	.100	.080	.062
	INDICE DE REPRESENTACIONES SOCIALES NEGATIVAS	1.168	1.309	-.403	-.892	.393	-4.083	1.748	.169	-.272	-.219
	ÍNDICE DE GESTIÓN ESCOLAR NEGATIVA	-.015	.389	-.019	-.039	.970	-.882	.852	.247	-.012	-.009
	ÍNDICE DE BARRIO ESCOLAR	.311	.651	.150	.478	.643	-1.139	1.761	.025	.150	.117
	ÍNDICE DE POLÍTICAS PÚBLICAS	.613	.499	.663	1.228	.248	-.499	1.726	.514	.362	.301
	Edad	-.102	2.080	-.023	-.049	.962	-4.736	4.533	-.010	-.015	-.012
	Sexo_femenino_dummy	.483	2.538	.074	.190	.853	-5.172	6.138	-.291	.060	.047
	Turno_matutino_dummy	-.365	4.732	-.040	-.077	.940	-10.909	10.179	.145	-.024	-.019
Nivel_primaria_dummy	.011	5.328	.001	.002	.998	-11.860	11.881	.033	.001	.000	
a. Variable dependiente: ÍNDICE DE VIOLENCIA ESCOLAR											

Fuente: Elaborado por el autor con información de base de datos SPSS

La **tabla 4** nos presenta el coeficiente Beta (estandarizados), el cual nos indica la jerarquía explicativa que posee cada variable independiente (a partir de su peso explicativo) en relación a la variable dependiente (Índice de Violencia Escolar).

El conjunto de variables independientes incluidas **no aportan significativamente al modelo, puesto que el nivel de significación de estas variables es mayor a 0,05.**

Esta situación se confirma al observar los intervalos de confianza, los cuales incluyen el valor 0. Por lo tanto se asume con un 95% de confianza que **el conjunto de variables independientes no aportan significativamente al modelo.**

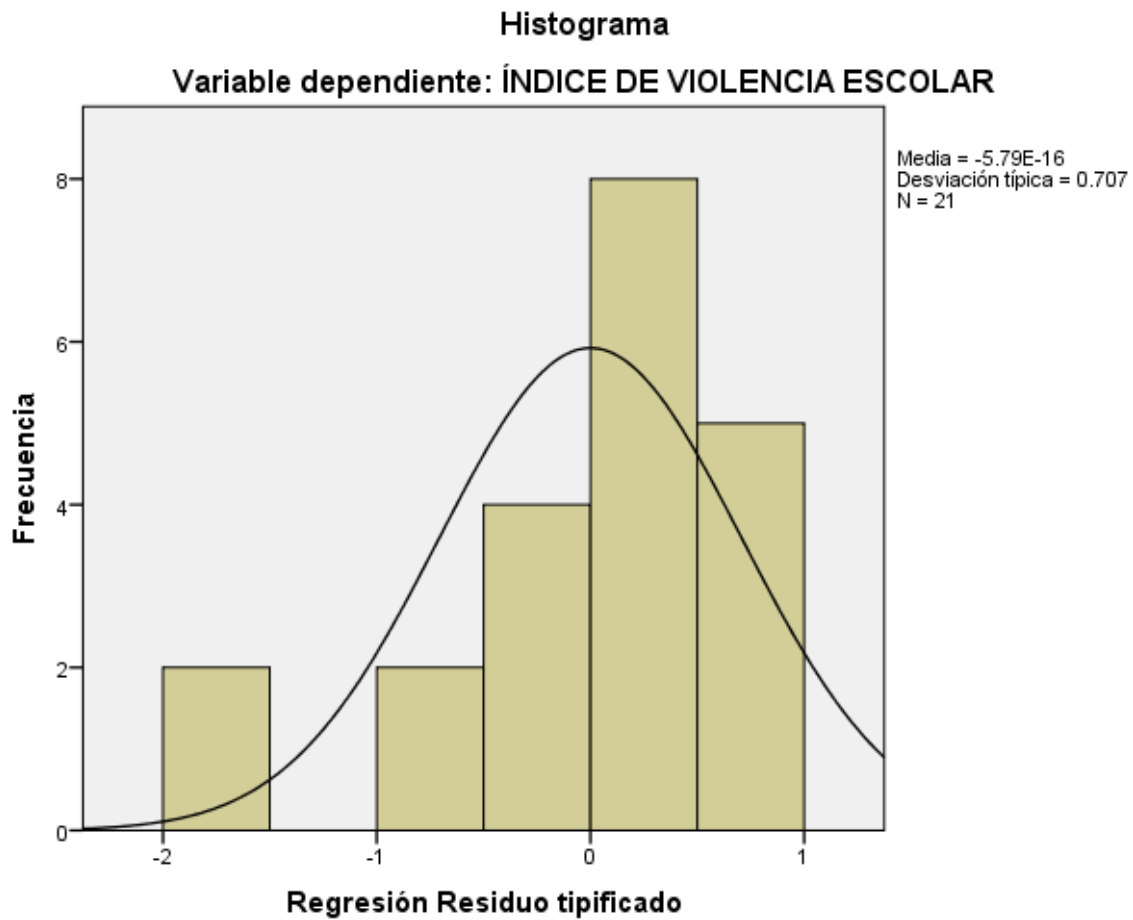
A partir del coeficiente Beta no estandarizado (tabla 4) es posible realizar varias afirmaciones al respecto:

- Por cada unidad que aumente un estudiante en el “índice de clima escolar negativo”, el “índice de violencia escolar” **aumentará en 0,486 unidades.**
- Por cada unidad que aumente un estudiante en el “índice de incivildades”, el “índice de violencia escolar” **aumentará en 0.227 unidades.**
- Por cada unidad que aumente un estudiante en el “índice de representaciones sociales negativas”, el “índice de violencia escolar” **disminuirá en 1.168 unidades.**
- Por cada unidad que aumente un estudiante en el “índice de gestión escolar negativa”, el “índice de violencia escolar” **disminuirá en 0,15 unidades.**
- Por cada unidad que aumente un estudiante en el “índice de barrio escolar negativo”, el “índice de violencia escolar” **aumentará en 0,311 unidades.**
- Por cada unidad que aumente un estudiante en el “índice de políticas públicas deficientes”, el “índice de violencia escolar” **aumentará en 0.613 unidades.**
- Si un estudiante se encuentra en el nivel primario su “índice de violencia escolar” **aumentará 0.011 unidades** con respecto a los estudiantes de secundaria.
- Si un estudiante se encuentra en el turno matutino su “índice de violencia escolar” **disminuirá 0,365 unidades** con respecto a los estudiantes del turno vespertino.
- Si una estudiante es mujer su “índice de violencia escolar” **aumentará en 0.483 unidades** con respecto a los estudiantes hombres.
- Por cada año de edad que aumente un estudiante su “índice de violencia escolar” **disminuirá en 0.102 unidades.**

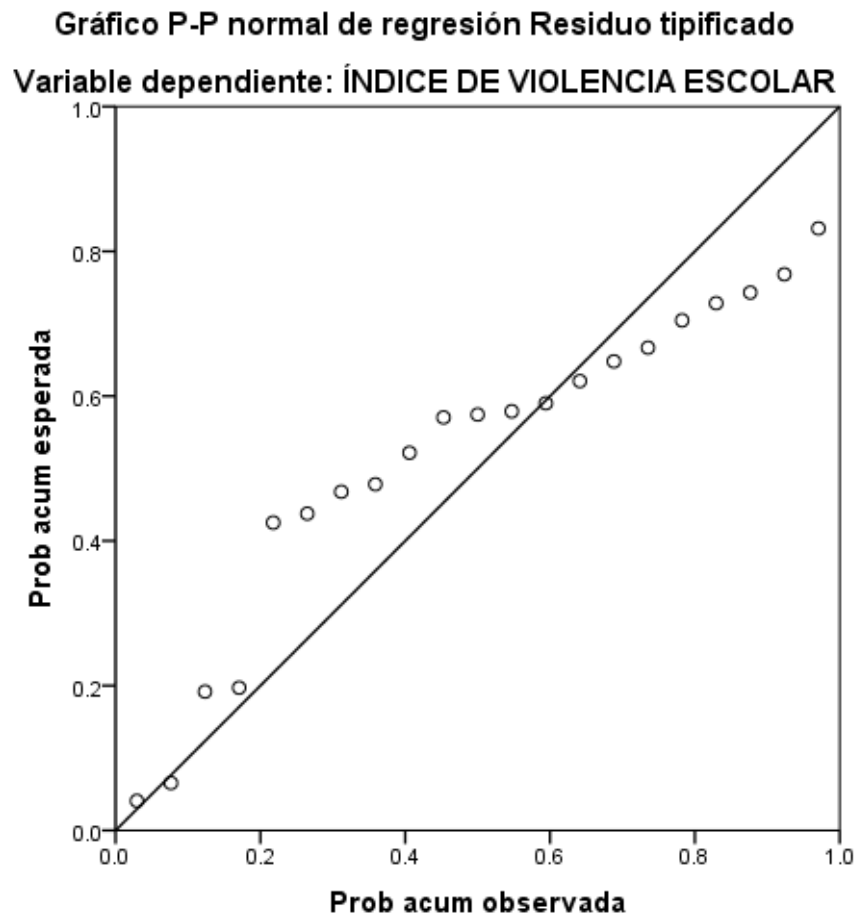
Construcción de perfiles mediante valores ajustados del modelo siguiendo la lógica de tipos ideales

A continuación presentamos un cuadro que describe los valores de la variable dependiente y las independientes.

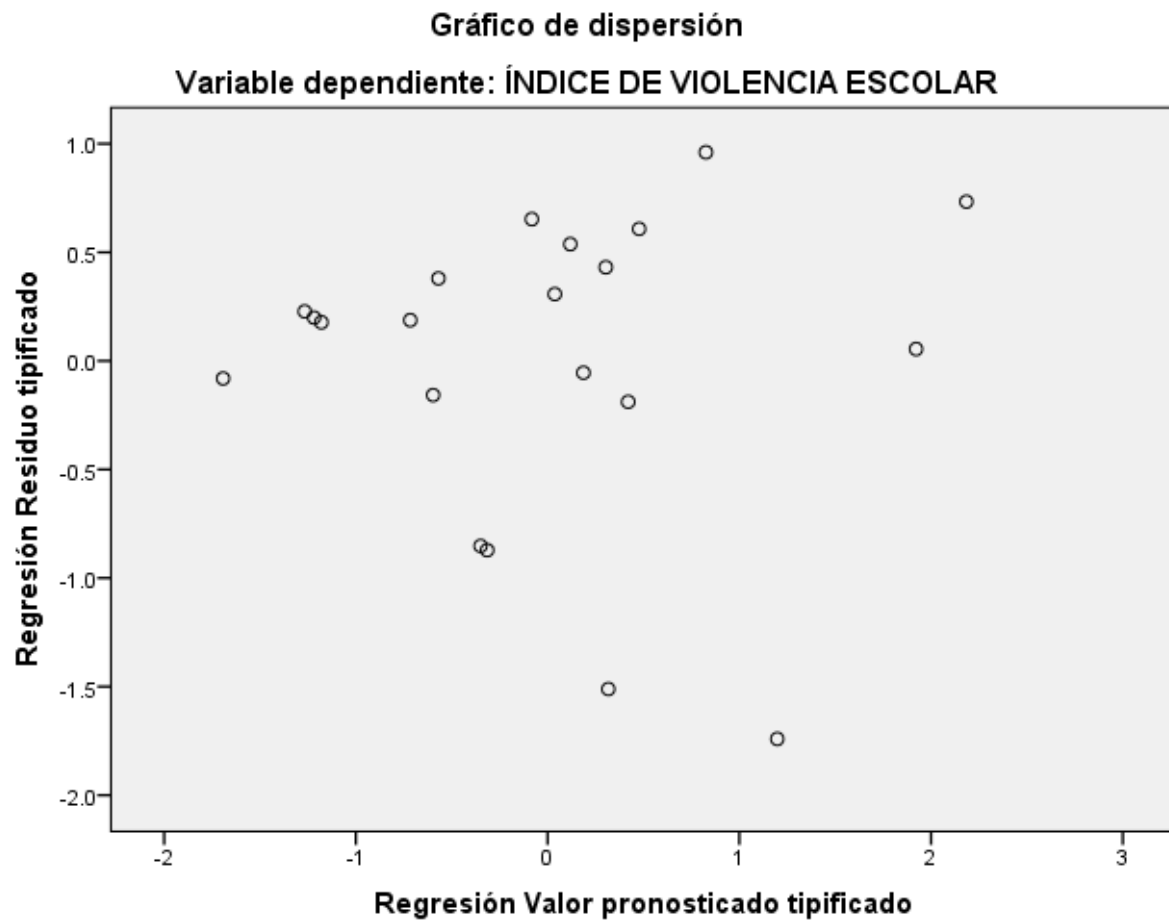
Variables incluidas en el modelo	
Índice de violencia escolar (dependiente)	Índice sumatorio: entre 7 y 35 unidades (bajo :7-16, medio: 17-26 , alto: 27-35)
Índice de clima escolar (ind)	Índice sumatorio: entre 5 y 20 unidades
Índice de incivildades (ind)	Índice sumatorio: entre 4 y 20 unidades
Índice de representaciones sociales negativa (ind)	Índice sumatorio: entre 2 y 9 unidades
Índice de gestión escolar negativa (ind)	Índice sumatorio: entre 8 y 36 unidades
Índice de barrio escolar negativo (ind)	Índice sumatorio: entre 3 y 14 unidades
Índice de políticas públicas deficientes (ind)	Índice sumatorio: entre 5 y 22 unidades
Nivel_primaria_dummy (ind)	1=primaria 0=secundaria
Turno_primaria_dummy (ind)	1=matutino 0 =vespertino
Sexo_femenino_dummy (ind)	1=femenino 0=masculino
Edad (ind)	Entre 10 y 17 años



El **Histograma** de la distribución de los residuos tipificados nos permite evaluar visualmente el grado de ajuste de esta distribución a la curva normal. No obstante, es posible observar que la distribución **no se asemeja a una curva normal**. En efecto, los residuos tipificados no se ajustan a la curva normal superpuesta.



El **gráfico de probabilidades de la distribución normal**, nos muestra como la diagonal de los residuos la línea gruesa negra, se asemeja a la diagonal teórica. Sin embargo no se cumple con la condición de distribución normal entre las variables. En este caso los residuos no se distribuyen normalmente, ya que la nube de puntos no se encuentra alineada con la diagonal del gráfico.



Como se aprecia en el **Gráfico de dispersión**, los puntos se distribuyen en un grado que no es aceptable, puesto que los datos no se reparten homogéneamente.

Bibliografía

1. Santana, M. (S/f) en:
[http://www.uesiglo21.edu.ar/site/departamentos/departamentos/economia/pdfs%20dpto%20economia/microeconomia Mirta Santana/UNIDAD07.pdf](http://www.uesiglo21.edu.ar/site/departamentos/departamentos/economia/pdfs%20dpto%20economia/microeconomia%20Mirta%20Santana/UNIDAD07.pdf)