

Análisis Multivariado: Modelo Predictivo de Regresión Lineal Múltiple

El modelo predictivo utilizado para realizar el presente análisis es el de regresión lineal múltiple (RLM)¹. 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).

A continuación se presenta un completo análisis del modelo propuesto.

¹ Cuando 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.

TABLA 2: RESUMEN DEL MODELO										
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	.612 ^a	.374	.345	3.05009	.374	12.792	10	214	.000	1.819
a. Variables predictoras: (Constante), Edad, Sexo_femenino_dummy, Turno_matutino_dummy, INDICE DE REPRESENTACIONES SOCIALES NEGATIVAS, ÍNDICE DE BARRIO ESCOLAR NEGATIVO, ÍNDICE DE CLIMA ESCOLAR NEGATIVO, INDICE DE INCIVILIDADES, ÍNDICE DE GESTIÓN ESCOLAR NEGATIVA, ÍNDICE DE POLÍTICAS PÚBLICAS DEFICIENTES, Nivel_primaria_dummy										
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 **0,612**, 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 **0,374**, esto indica que podemos predecir en un **37,4% el índice de Violencia Escolar**.

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 **0,345** (34,5%) muy cercano al R cuadrado (**37,4%**).

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

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

observamos **la significación del cambio en F (0,0001)**, puesto que su valor es menor a 0,05 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 efectivamente se asume independencia entre los residuos **(1,819)**.

A partir de estos coeficientes es importante señalar que nuestro modelo se encuentra ajustado y se cumplen los criterios estadísticos de aplicabilidad.

TABLA 3: ANOVA ^a						
Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	1190.037	10	119.004	12.792	.000 ^b
	Residual	1990.852	214	9.303		
	Total	3180.889	224			
a. Variable dependiente: ÍNDICE DE VIOLENCIA ESCOLAR						
b. Variables predictoras ⁴ : (Constante), Edad, Sexo_femenino_dummy, Turno_matutino_dummy, 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, Nivel_primaria_dummy						

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 (Ho) de que R es 0 en la población. La significación que nos indica la tabla es **0,00001**, de esto se interpreta que podemos rechazar la hipótesis nula de que R es 0 en la población.

A partir de esta información podemos asumir con un 95% de confianza que existe una relación lineal significativa entre la Variable Dependiente (Índice de Violencia Escolar) y el conjunto de variables independientes del modelo.

⁴ Las variables predictoras hacen referencia a las variables independientes.

TABLA 4: COEFICIENTES ^a								
Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.	Intervalo de confianza de 95.0% para B	
		B	Error típ.	Beta			Límite inferior	Límite superior
1	(Constante)	10.043	2.951		3.404	.001	4.227	15.859
	ÍNDICE DE CLIMA ESCOLAR NEGATIVO	.330	.131	.157	2.517	.013	.072	.588
	INDICE DE INCIVILIDADES	.764	.102	.479	7.504	.000	.563	.965
	INDICE DE REPRESENTACIONES SOCIALES NEGATIVAS	-.101	.169	-.039	-.596	.552	-.435	.233
	ÍNDICE DE GESTIÓN ESCOLAR NEGATIVA	-.038	.067	-.041	-.571	.569	-.169	.093
	ÍNDICE DE BARRIO ESCOLAR NEGATIVO	.199	.126	.101	1.588	.114	-.048	.447
	ÍNDICE DE POLÍTICAS PÚBLICAS DEFICIENTES	.269	.095	.200	2.830	.005	.082	.456
	Nivel_primaria_dummy	.199	.553	.026	.359	.720	-.892	1.289
	Turno_matutino_dummy	.420	.513	.045	.817	.415	-.592	1.432
	Sexo_femenino_dummy	-.274	.421	-.036	-.651	.516	-1.103	.555
	Edad	-.544	.223	-.166	-	.016	-.983	-.104

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).

En este caso se observa que las variables que poseen mayor peso explicativo para el modelo son (en orden descendente); **“Índice de incivildades” (0,479)**, **“Índice de políticas públicas deficientes” (0,200)**, **“Edad” (-0,166)** y el **“Índice**

de clima escolar negativo (0,157). Estas variables coinciden con un nivel de significación inferior a **0,05**, lo cual permite señalar que son lo suficientemente explicativas para el modelo. Esta situación se confirma al observar los intervalos de confianza de las cuatro variables en cuestión, en ninguno de estos casos se incluye el valor 0, lo cual indica con un 95% de confianza que las variables independientes (*“Índice de Incivildades”*, *“Índice de políticas públicas deficientes”*, *“Edad”* y el *“Índice de clima escolar negativo”*) aportan significativamente al modelo.

No obstante, las variables *“Índice de representaciones sociales negativas”* (0,552), *“Índice de gestión escolar negativa”* (0,569), *“Índice de barrio escolar negativo”* (0,114), *“Nivel”* (0,720), *“Turno”* (0,415) y *“Sexo”* (0,516) 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 las variables independientes descritas no aportan significativamente al modelo.

Las variables descritas no son significativas en términos estadísticos, sin embargo no serán excluidas del modelo puesto que son relevantes teóricamente para esta investigación.

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 0,330 unidades.**
- Por cada unidad que aumente un estudiante en el “índice de incivildades”, el “índice de violencia escolar” **aumentará en 0,764 unidades.**
- Por cada unidad que aumente un estudiante en el “índice de representaciones sociales negativas”, el “índice de violencia escolar” **disminuirá en 0,101 unidades.**
- Por cada unidad que aumente un estudiante en el “índice de gestión escolar negativa”, el “índice de violencia escolar” **disminuirá en 0,038 unidades.**
- Por cada unidad que aumente un estudiante en el “índice de barrio escolar negativo”, el “índice de violencia escolar” **aumentará en 0,199 unidades.**
- Por cada unidad que aumente un estudiante en el “índice de políticas públicas deficientes”, el “índice de violencia escolar” **aumentará en 0,269 unidades.**
- Si un estudiante se encuentra en el nivel primario su “índice de violencia escolar” **aumentará 0,199 unidades** con respecto a los estudiantes de secundaria.
- Si un estudiante se encuentra en el turno matutino su “índice de violencia escolar” **aumentará 0,420 unidades** con respecto a los estudiantes del turno vespertino.
- Si una estudiante es mujer su “índice de violencia escolar” **disminuirá en 0,274 unidades** con respecto a los estudiantes hombres.
- Por cada año de edad que aumente un estudiante su “índice de violencia escolar” **disminuirá en 0.544 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

Ahora bien, las ecuaciones⁵ que desarrollamos a continuación identifican, en primer lugar, a la constante⁶ (10.043) y los valores que continúan son las variables independientes (10) que hemos incluido en el modelo.

Con respecto a la variables dummy⁷ , se les incluyó con el valor 1 (primaria, matutino y mujer) y 0 (secundaria, vespertino, hombre).

⁵ Los valores que se observan en las ecuaciones fueron obtenido de la tabla 4: coeficientes.

⁶ Hacemos referencia a la variable dependiente “índice de violencia escolar”.

⁷ Nos referimos a las variables nivel, turno y sexo.

ECUACIÓN 1⁸

- Ecuación 1A (primaria, matutino, hombre, 11 años, factores de riesgo altos⁹)

$$\text{Índice de violencia escolar (Y)} = 10.043 + 0.33 \times 20 + 0.764 \times 20 + -0.101 \times 9 + -0.038 \times 36 + 0.199 \times 14 + 0.269 \times 22 + 0.199 \times 1 + 0.420 \times 1 + 0 \times -0.274 + -0.544 \times 11$$

$$\text{Índice de violencia escolar (Y)} = 32.985 \text{ unidades (alto)}$$

- Ecuación 1B (primaria, matutino, hombre, 11 años, factores de riesgo bajos)¹⁰

$$\text{Índice de violencia escolar (Y)} = 10.043 + 0.33 \times 5 + 0.764 \times 4 + -0.101 \times 2 + -0.038 \times 8 + 0.199 \times 3 + 0.269 \times 5 + 0.199 \times 1 + 0.420 \times 1 + 0 \times -0.274 + -0.544 \times 11$$

$$\text{Índice de violencia escolar (Y)} = 10.82 \text{ unidades (bajo)}$$

ECUACIÓN 2

- Ecuación 2A (secundaria, matutino, hombre, 13 años, factores de riesgo altos)

$$\text{Índice de violencia escolar (Y)} = 10.043 + 0.33 \times 20 + 0.764 \times 20 + -0.101 \times 9 + -0.038 \times 36 + 0.199 \times 14 + 0.269 \times 22 + 0.199 \times 0 + 0.420 \times 1 + 0 \times -0.274 + -0.544 \times 13$$

$$\text{Índice de violencia escolar (Y)} = 31.698 \text{ unidades (alto)}$$

- Ecuación 2B (secundaria, matutino, hombre, 13 años, factores de riesgo bajos)

$$\text{Índice de violencia escolar (Y)} = 10.043 + 0.33 \times 5 + 0.764 \times 4 + -0.101 \times 2 + -0.038 \times 8 + 0.199 \times 3 + 0.269 \times 5 + 0.199 \times 0 + 0.420 \times 1 + 0 \times -0.274 + -0.544 \times 13$$

$$\text{Índice de violencia escolar (Y)} = 9.533 \text{ unidades (bajo)}$$

⁸ La ecuación 1 y 2 buscan comparar las variaciones del índice de violencia escolar con respecto al nivel (primaria-secundaria) de los estudiantes hombres de 11 y 13 años.

⁹ Se les otorgó a los índices la mayor puntuación, lo cual indica factores de riesgo altos.

¹⁰ Se les otorgó a los índices la menor puntuación, lo cual indica factores de riesgo bajos.

ECUACIÓN 3¹¹

- Ecuación 3A (primaria, matutino, mujer, 11 años, factores de riesgo altos)

$$\text{Índice de violencia escolar (Y)} = 10.043 + 0.33 \times 20 + 0.764 \times 20 + -0.101 \times 9 + -0.038 \times 36 + 0.199 \times 14 + 0.269 \times 22 + 0.199 \times 1 + 0.420 \times 1 + 1 \times -0.274 + -0.544 \times 11$$

$$\text{Índice de violencia escolar (Y)} = \mathbf{32.711 \text{ unidades (alto)}}$$

- Ecuación 3B (primaria, matutino, mujer, 11 años, factores de riesgo bajos)

$$\text{Índice de violencia escolar (Y)} = 10.043 + 0.33 \times 5 + 0.764 \times 4 + -0.101 \times 2 + -0.038 \times 8 + 0.199 \times 3 + 0.269 \times 5 + 0.199 \times 1 + 0.420 \times 1 + 1 \times -0.274 + -0.544 \times 11$$

$$\text{Índice de violencia escolar (Y)} = \mathbf{10.546 \text{ unidades (bajo)}}$$

ECUACIÓN 4

- Ecuación 4A (secundaria, matutino, mujer, 13 años, factores de riesgo altos)

$$\text{Índice de violencia escolar (Y)} = 10.043 + 0.33 \times 20 + 0.764 \times 20 + -0.101 \times 9 + -0.038 \times 36 + 0.199 \times 14 + 0.269 \times 22 + 0.199 \times 0 + 0.420 \times 1 + 1 \times -0.274 + -0.544 \times 13$$

$$\text{Índice de violencia escolar (Y)} = \mathbf{31.424 \text{ unidades (alto)}}$$

- Ecuación 4B (secundaria, matutino, mujer, 13 años, factores de riesgo bajos)

$$\text{Índice de violencia escolar (Y)} = 10.043 + 0.33 \times 5 + 0.764 \times 4 + -0.101 \times 2 + -0.038 \times 8 + 0.199 \times 3 + 0.269 \times 5 + 0.199 \times 0 + 0.420 \times 1 + 1 \times -0.274 + -0.544 \times 13$$

$$\text{Índice de violencia escolar (Y)} = \mathbf{9.259 \text{ unidades (bajo)}}$$

¹¹ La ecuación 3 y 4 buscan comparar las variaciones del índice de violencia escolar con respecto al nivel (primaria-secundaria) de las estudiantes mujeres de 11 y 13 años.

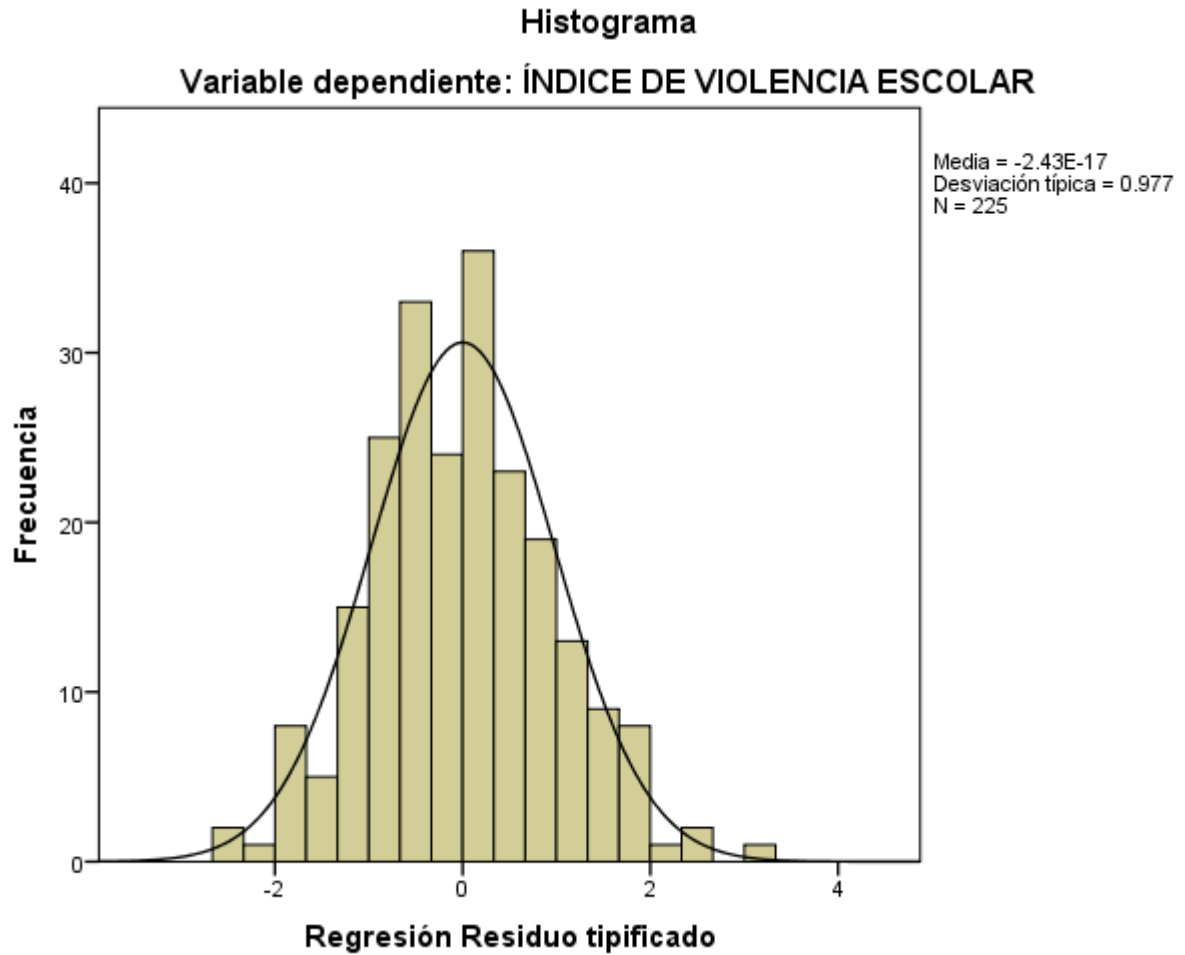
TABLA 5: Correlaciones

		Índice de violencia escolar	Índice de clima escolar	Índice de incivildades	Índice de representaciones sociales negativas	Índice de gestión escolar negativa	Índice de barrio escolar	Índice de políticas públicas	Edad	Nivel_primaria_dummy	Turno_materno_dummy	Sexo_femenino_dummy
Índice de violencia escolar	Correlación de Pearson	1	.351**	.494**	.137**	.320**	.206**	.293**	.021	-.021	.137**	-.096*
Índice de clima escolar	Correlación de Pearson	.351**	1	.321**	.283**	.404**	.154**	.407**	.214**	.192**	.105**	-.076*
Índice de incivildades	Correlación de Pearson	.494**	.321**	1	.173**	.232**	.430**	.218**	.281**	.301**	.093*	-.055
Índice de representaciones sociales negativas	Correlación de Pearson	.137**	.283**	.173**	1	.506**	-.050	.487**	.108**	.065*	-.037	-.069*
Índice de gestión escolar negativa	Correlación de Pearson	.320**	.404**	.232**	.506**	1	.100*	.526**	.173**	.073*	.112**	-.094**
Índice de barrio escolar	Correlación de Pearson	.206**	.154**	.430**	-.050	.100*	1	-.040	.306**	.422**	.121**	-.060
Índice de políticas públicas	Correlación de Pearson	.293**	.407**	.218**	.487**	.526**	-.040	1	.074	.068	.086*	-.096*
Edad	Correlación de Pearson	.021	.214**	.281**	.108**	.173**	.306**	.074	1	.525**	.161**	-.023
Nivel_primaria_dummy	Correlación de Pearson	-.021	-.192**	-.301**	.065*	.073*	-.422**	.068	.525**	1	-.221**	.018
Turno_materno_dummy	Correlación de Pearson	.137**	.105**	.093*	-.037	.112**	.121**	.086*	.161**	.221**	1	.006
Sexo_femenino_dummy	Correlación de Pearson	-.096*	-.076*	-.055	-.069*	-.094**	-.060	-.096*	-.023	.018	.006	1

** . La correlación es significativa al nivel 0,01 (bilateral).

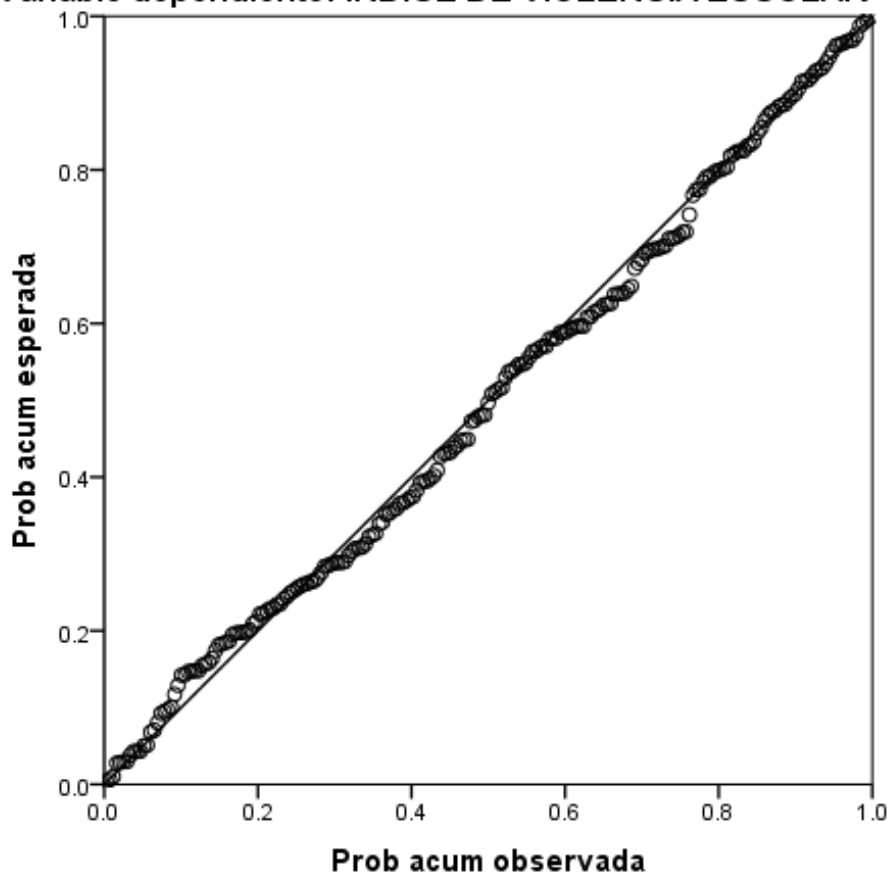
*. La correlación es significativa al nivel 0,05 (bilateral).

A partir de la salida de resultados de correlaciones de los coeficientes (**ver tabla 5**), se puede afirmar que no se genera multicolinealidad. Es decir, las correlaciones que se dan a nivel multivariado entre las variables independientes incluidas en el modelo no se acercan de manera considerable al valor 1 como para que se cumpla la condición de multicolinealidad perfecta. Esto implica que cada variable aporta con información nueva al modelo y posee poder explicativo. Como se aprecia en la tabla 5, la correlación más alta se produce entre las variables “índice de políticas públicas” y “índice de gestión escolar negativo”, con un valor de 0.526. Otras correlaciones altas se dan entre el “índice de violencia escolar” y el “índice de incivildades” (0.494), entre el “índice de violencia escolar” y “el índice de clima escolar” (0.351).

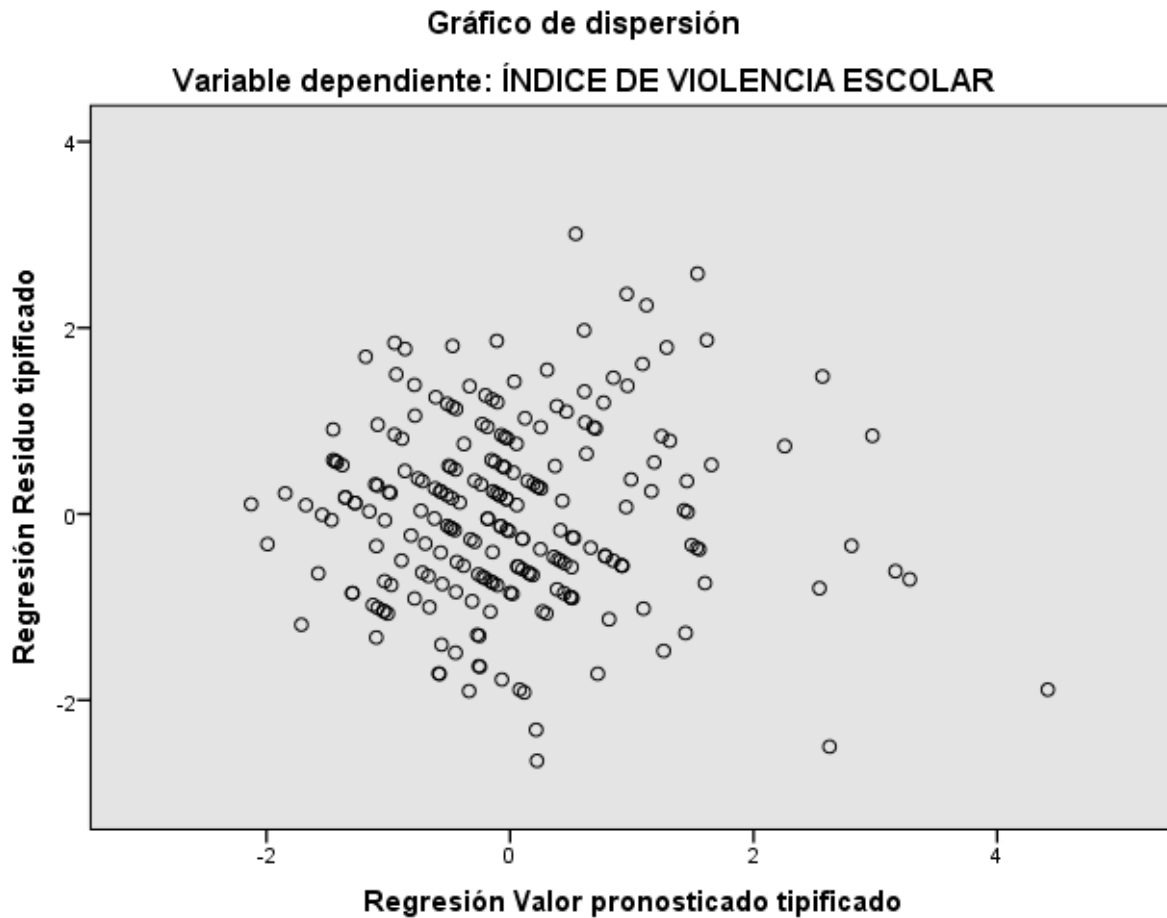


El **Histograma** de la distribución de los residuos tipificados nos permite evaluar visualmente el grado de ajuste de esta distribución a la curva normal. Es posible observar que la distribución se asemeja a una curva normal. En este caso hay sólo un par de valores que se encuentran fuera de la curva normal, por lo tanto la mayor parte de los casos se comportan con una distribución normal. En efecto, los residuos tipificados se ajustan a la curva normal superpuesta.

Gráfico P-P normal de regresión Residuo tipificado
Variable dependiente: ÍNDICE DE VIOLENCIA ESCOLAR

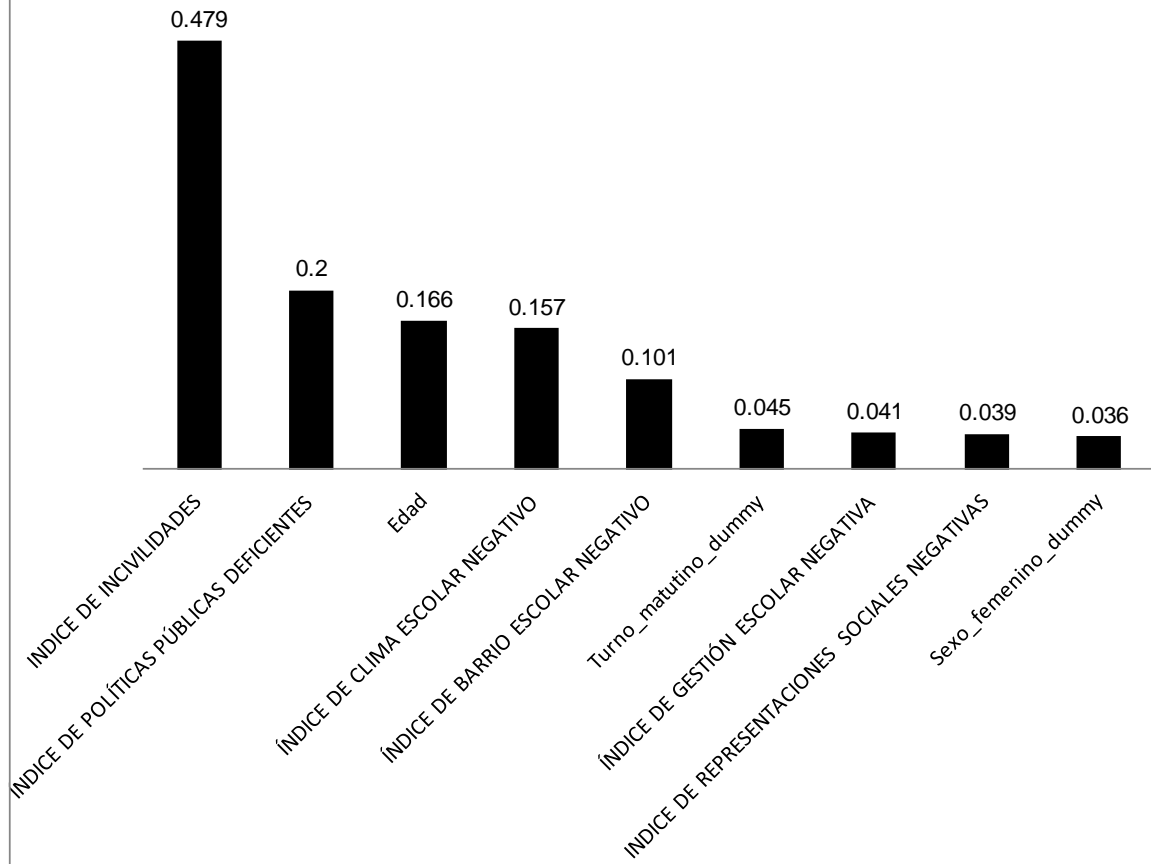


El **gráfico de probabilidades de la distribución normal**, nos muestra como la diagonal de los residuos la línea gruesa negra, se asemeja a la diagonal teórica, no hay muchos puntos de variación lo que indica que es posible afirmar que se cumple con la condición de distribución normal entre las variables.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



Como se aprecia en el **Gráfico de dispersión**, los puntos se distribuyen en un grado aceptable, ya que no se aprecian concentraciones muy grandes de puntos, es decir los datos se reparten homogéneamente, y estos se ubican en el gráfico entre los valores -2 y 2 . En ese sentido, según la noción de homocedasticidad se observa que no existe asociación entre pronósticos y residuos.

Variables Independientes: Peso Explicativo



Bibliografía

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