

Modelo Predictivo de Regresión Lineal Múltiple: ZM Xalapa

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

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.

TABLA 2: 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	.661 ^a	.437	.393	3.10136	.437	9.791	10	126	.000	1.711
a. Variables predictoras: (Constante), Edad, Turno_matutino_dummy, INDICE DE REPRESENTACIONES SOCIALES NEGATIVAS, Sexo_femenino_dummy, INDICE DE INCIVILIDADES, ÍNDICE DE GESTIÓN ESCOLAR NEGATIVA, ÍNDICE DE CLIMA ESCOLAR, ÍNDICE DE POLÍTICAS PÚBLICAS, ÍNDICE DE BARRIO ESCOLAR, 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 **0,661**, 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 **0,437**, esto indica que podemos predecir en un **43,7% el índice de Violencia Escolar** en la Zona Metropolitana de Xalapa.

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 **0,393** (39,3%).}

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 (0,0001)**, 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 **(1,711)**.

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	941.753	10	94.175	9.791	.000 ^b
	Residual	1211.926	126	9.618		
	Total	2153.679	136			
a. Variable dependiente: ÍNDICE DE VIOLENCIA ESCOLAR						
b. Variables predictoras: (Constante), Edad, Turno_matutino_dummy, ÍNDICE DE REPRESENTACIONES SOCIALES NEGATIVAS, Sexo_femenino_dummy, ÍNDICE DE INCIVILIDADES, ÍNDICE DE GESTIÓN ESCOLAR NEGATIVA, ÍNDICE DE CLIMA ESCOLAR, ÍNDICE DE POLÍTICAS PÚBLICAS, ÍNDICE DE BARRIO ESCOLAR, 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 (H_0) 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.

TABLA 4: Coeficientes ^a								
Modelo		Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.	Intervalo de confianza de 95.0% para B	
		B	Error típ.	Beta			Límite inferior	Límite superior
1	(Constante)	7.568	4.280		1.768	.079	-.901	16.037
	ÍNDICE DE CLIMA ESCOLAR	.299	.195	.127	1.537	.127	-.086	.685
	INDICE DE INCIVILIDADES	.948	.135	.571	7.030	.000	.681	1.215
	INDICE DE REPRESENTACIONES SOCIALES NEGATIVAS	-.220	.222	-.075	-.995	.322	-.659	.218
	ÍNDICE DE GESTIÓN ESCOLAR NEGATIVA	.029	.085	.027	.336	.737	-.140	.197
	ÍNDICE DE BARRIO ESCOLAR	.025	.179	.013	.142	.887	-.329	.380
	ÍNDICE DE POLÍTICAS PÚBLICAS	.149	.145	.085	1.026	.307	-.139	.437
	Nivel_primaria_dummy	-.330	.814	-.042	-.405	.686	-1.941	1.281
	Turno_matutino_dummy	.577	.624	.064	.924	.357	-.658	1.812
	Sexo_femenino_dummy	-.260	.552	-.033	-.471	.638	-1.353	.832
	Edad	-.290	.307	-.083	-.944	.347	-.897	.317

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 la variable que posee mayor peso explicativo para el modelo es el **“Índice de incivildades” (0,571)**. Esta variable coincide con un

nivel de significación inferior a **0,05**, lo cual permite señalar que es lo suficientemente explicativa para el modelo. Esta situación se confirma al observar el intervalo de confianza, donde no se incluye el valor 0, lo cual indica con un 95% de confianza que la variable dependiente **“Índice de incivildades”** aporta significativamente al modelo.

No obstante, el resto de las variables 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 (con excepción de él **“Índice de incivildades”**) 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 0,299 unidades.**
- Por cada unidad que aumente un estudiante en el “índice de incivildades”, el “índice de violencia escolar” **aumentará en 0,948 unidades.**
- Por cada unidad que aumente un estudiante en el “índice de representaciones sociales negativas”, el “índice de violencia escolar” **disminuirá en 0,220 unidades.**
- Por cada unidad que aumente un estudiante en el “índice de gestión escolar negativa”, el “índice de violencia escolar” **aumentará en 0,029 unidades.**
- Por cada unidad que aumente un estudiante en el “índice de barrio escolar negativo”, el “índice de violencia escolar” **aumentará en 0,025 unidades.**
- Por cada unidad que aumente un estudiante en el “índice de políticas públicas deficientes”, el “índice de violencia escolar” **aumentará en 0,149 unidades.**
- Si un estudiante se encuentra en el nivel primario su “índice de violencia escolar” **disminuirá 0,330 unidades** con respecto a los estudiantes de secundaria.
- Si un estudiante se encuentra en el turno matutino su “índice de violencia escolar” **aumentará 0,577 unidades** con respecto a los estudiantes del turno vespertino.
- Si una estudiante es mujer su “índice de violencia escolar” **disminuirá en 0,260 unidades** con respecto a los estudiantes hombres.
- Por cada año de edad que aumente un estudiante su “índice de violencia escolar” **disminuirá en 0,290 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 ecuaciones⁴ que desarrollamos a continuación identifican, en primer lugar, a la constante⁵ (7.568) y los valores que continúan son las variables independientes (10) que hemos incluido en el modelo.

Con respecto a la variables dummy⁶ , se les incluyó con el valor 1 (primaria, matutino y mujer) y 0 (secundaria, vespertino, hombre).

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

⁵ Hacemos referencia a la variable dependiente “Índice de violencia escolar”.

⁶ Nos referimos a las variables nivel, turno y sexo.

ECUACIÓN 1⁷

- Ecuación 1A (primaria, matutino, hombre, 11 años, factores de riesgo altos⁸)

$$\text{Índice de violencia escolar (Y)} = 7.568 + 0.299*20 + 0.948*20 + -0.22*9 + 0.029*36 + 0.025*14 + 0.149*22 + -0.33*1 + 0.577*1 + 0*-0.260 + -0.29*11$$

$$\text{Índice de violencia escolar (Y)} = 32.257 \text{ unidades (alto)}$$

- Ecuación 1B (primaria, matutino, hombre, 11 años, factores de riesgo bajos)⁹

$$\text{Índice de violencia escolar (Y)} = 7.568 + 0.299*5 + 0.948*4 + -0.22*2 + 0.029*8 + 0.025*3 + 0.149*5 + -0.33*1 + 0.577*1 + 0*-0.260 + -0.29*11$$

$$\text{Índice de violencia escolar (Y)} = 10.524 \text{ unidades (bajo)}$$

ECUACIÓN 2

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

$$\text{Índice de violencia escolar (Y)} = 7.568 + 0.299*20 + 0.948*20 + -0.22*9 + 0.029*36 + 0.025*14 + 0.149*22 + -0.33*0 + 0.577*1 + 0*-0.260 + -0.29*13$$

$$\text{Índice de violencia escolar (Y)} = 32.007 \text{ unidades (alto)}$$

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

$$\text{Índice de violencia escolar (Y)} = 7.568 + 0.299*5 + 0.948*4 + -0.22*2 + 0.029*8 + 0.025*3 + 0.149*5 + -0.33*0 + 0.577*1 + 0*-0.260 + -0.29*13$$

$$\text{Índice de violencia escolar (Y)} = 10.274 \text{ unidades (bajo)}$$

⁷ 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.

⁸ Se les otorgó a los índices la mayor puntuación, lo cual indica factores de riesgo altos.

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

ECUACIÓN 3¹⁰

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

$$\text{Índice de violencia escolar (Y)} = 7.568 + 0.299*20 + 0.948*20 + -0.22*9 + 0.029*36 + 0.025*14 + 0.149*22 + -0.33*1 + 0.577*1 + 1 * -0.260 + -0.29*11$$

$$\text{Índice de violencia escolar (Y)} = 31.997 \text{ unidades (alto)}$$

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

$$\text{Índice de violencia escolar (Y)} = 77.568 + 0.299*5 + 0.948*4 + -0.22*2 + 0.029*8 + 0.025*3 + 0.149*5 + -0.33*1 + 0.577*1 + 1 * -0.260 + -0.29*11$$

$$\text{Índice de violencia escolar (Y)} = 10.264 \text{ unidades (bajo)}$$

ECUACIÓN 4

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

$$\text{Índice de violencia escolar (Y)} = 7.568 + 0.299*20 + 0.948*20 + -0.22*9 + 0.029*36 + 0.025*14 + 0.149*22 + -0.33*0 + 0.577*1 + 1 * -0.260 + -0.29*13$$

$$\text{Índice de violencia escolar (Y)} = 31.747 \text{ unidades (alto)}$$

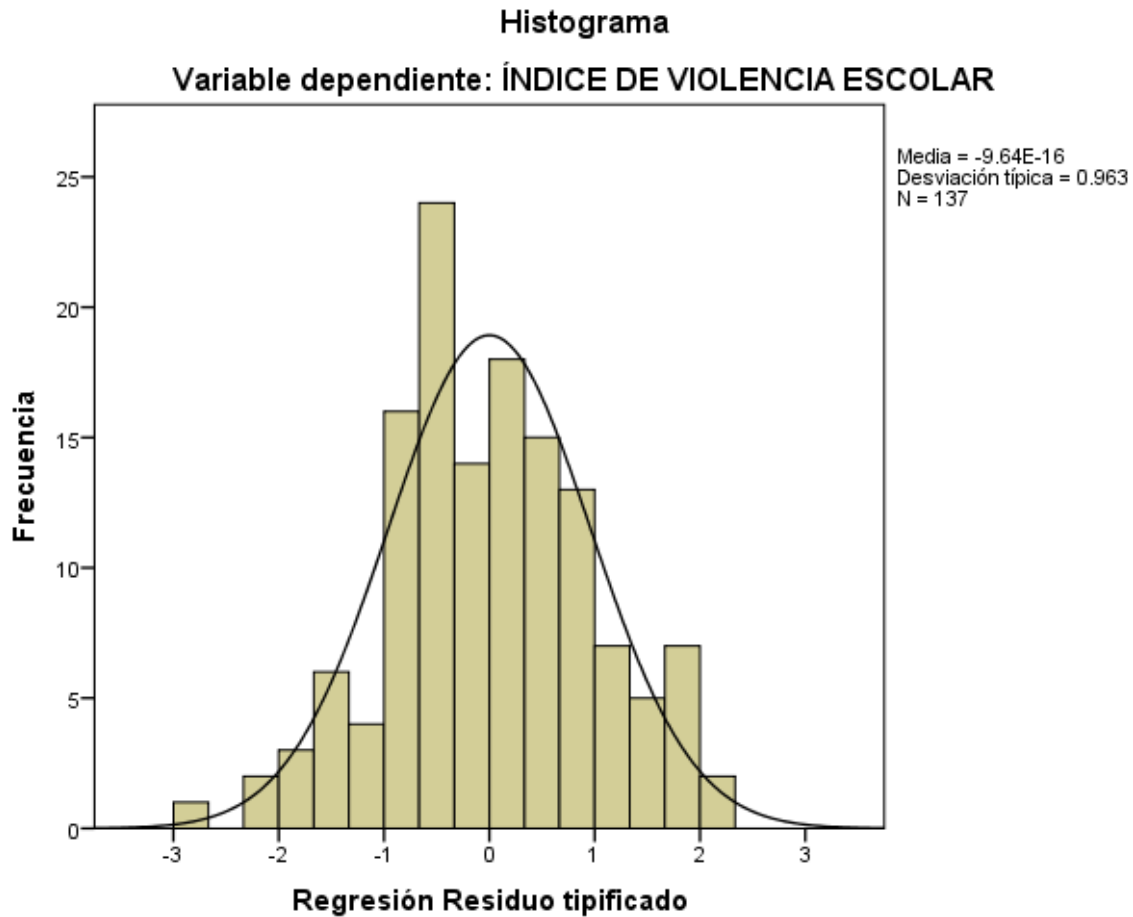
- Ecuación 4B (secundaria, matutino, mujer, 13 años, factores de riesgo bajos)

$$\text{Índice de violencia escolar (Y)} = 7.568 + 0.299*5 + 0.948*4 + -0.22*2 + 0.029*8 + 0.025*3 + 0.149*5 + -0.33*0 + 0.577*1 + 1 * -0.260 + -0.29*13$$

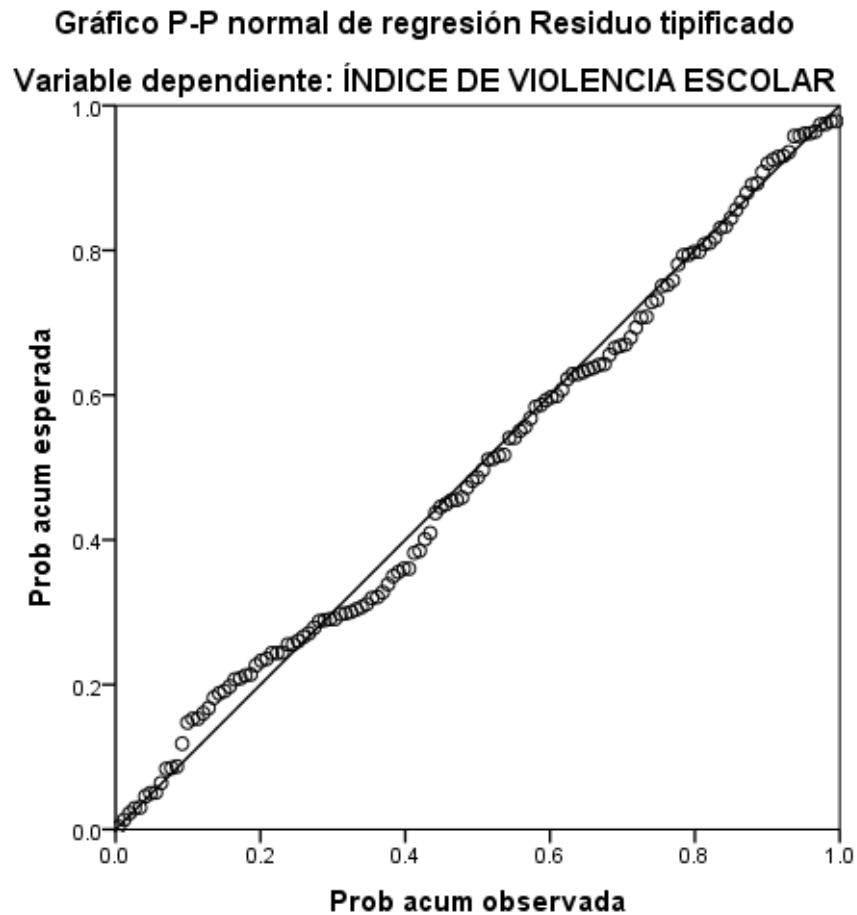
$$\text{Índice de violencia escolar (Y)} = 10.014 \text{ unidades (bajo)}$$

¹⁰ 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.

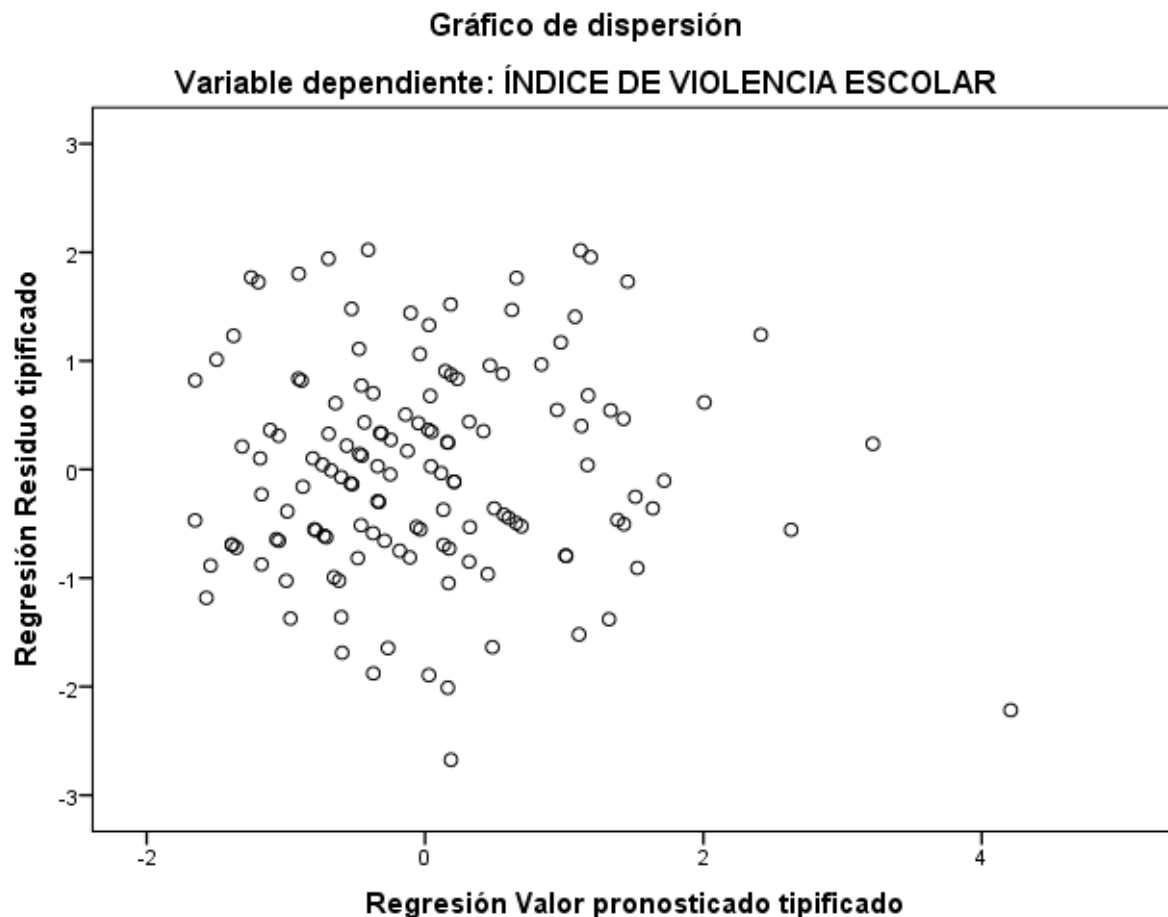
TABLA 5: Correlaciones										
ÍNDICE DE VIOLENCIA ESCOLAR	ÍNDICE DE CLIMA ESCOLAR	INDICE DE INCIVILIDADES	INDICE DE REPRESENTACIONES SOCIALES NEGATIVAS	ÍNDICE DE GESTIÓN ESCOLAR NEGATIVA	ÍNDICE DE BARRIO ESCOLAR	ÍNDICE DE POLÍTICAS PÚBLICAS	Nivel_primaria_dummy	Turno_matutino_dummy	Sexo_femenino_dummy	Edad
1.000	.273	.636	.023	.191	.315	.277	-.354	.087	-.146	.195
.273	1.000	.244	.298	.315	.318	.368	-.385	-.166	-.147	.302
.636	.244	1.000	.055	.205	.434	.282	-.493	.058	-.156	.318
.023	.298	.055	1.000	.347	-.058	.321	-.017	-.072	.007	.045
.191	.315	.205	.347	1.000	.122	.481	-.063	-.002	-.050	.167
.315	.318	.434	-.058	.122	1.000	.116	-.594	.076	-.204	.474
.277	.368	.282	.321	.481	.116	1.000	-.177	.061	-.077	.240
-.354	-.385	-.493	-.017	-.063	-.594	-.177	1.000	-.022	.245	-.612
.087	-.166	.058	-.072	-.002	.076	.061	-.022	1.000	.048	-.007
-.146	-.147	-.156	.007	-.050	-.204	-.077	.245	.048	1.000	-.154
.195	.302	.318	.045	.167	.474	.240	-.612	-.007	-.154	1.000



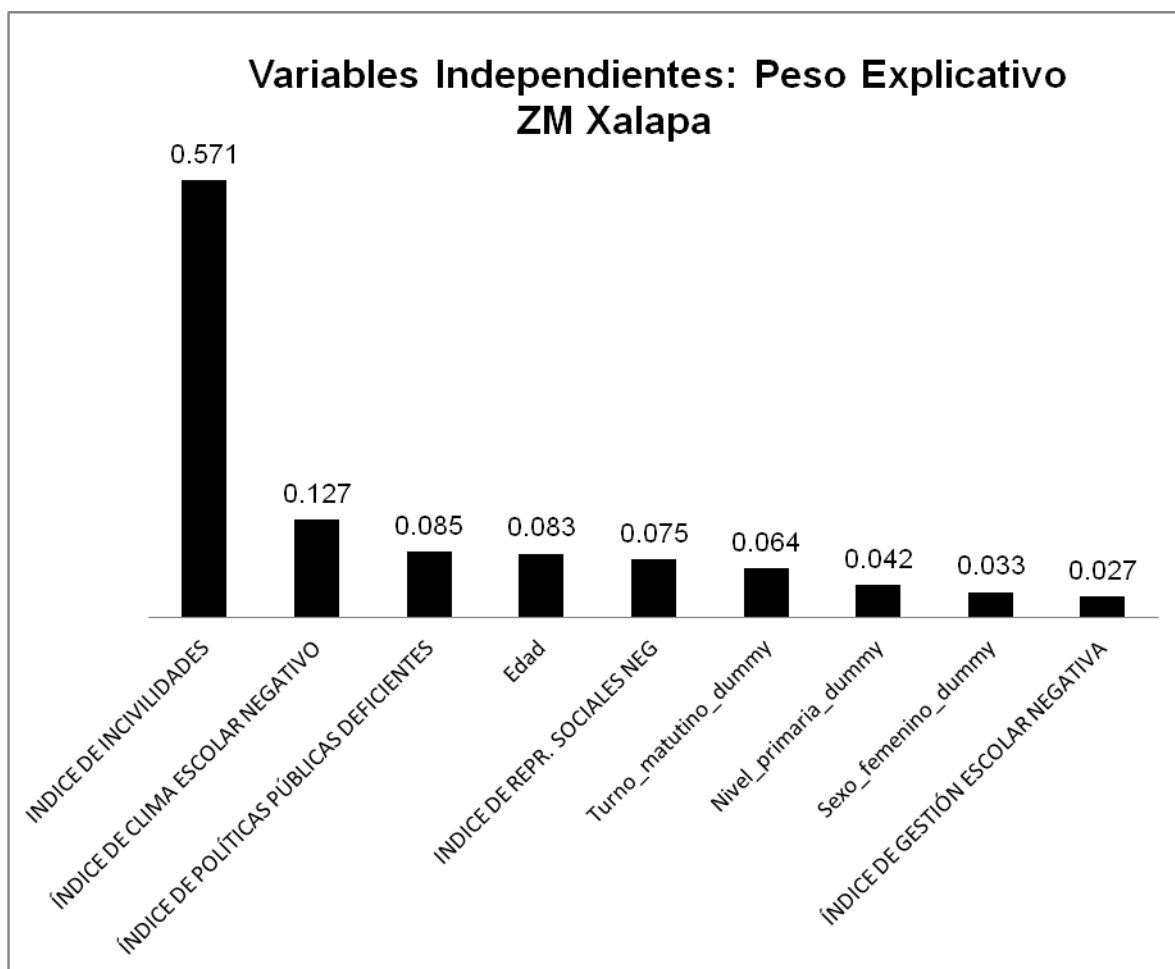
El **Histograma** de la distribución de los residuos tipificados nos permite evaluar visualmente el grado de ajuste de esta distribución a la curva normal. Es posible observar que la distribución se asemeja a una curva normal. En este caso hay sólo un par de valores que se encuentran fuera de la curva 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.



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.



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)