

BÚSQUEDA DE RELACIÓN ENTRE DIVERSOS INDICADORES

DESIGUALDAD ECONÓMICA Y EDUCATIVA EN MÉXICO

25 de Noviembre de 2025

Eric E. Moreles
Universidad de Guanajuato
Departamento de Matemáticas
`eric.moreles@cimat.mx`

1. Introducción

La desigualdad económica y educativa es un problema central para el diseño de políticas públicas; entender cómo se relacionan los ingresos, las condiciones sociales y el nivel de escolaridad a escala estatal permite identificar territorios con necesidades prioritarias. En este reporte se intenta encontrar una relación entre el nivel educativo, el ingreso económico y el bienestar social, para el año 2020, en México. Los datos fueron obtenidos de fuentes oficiales (INEGI, CONEVAL, CONAPO) y normalizados para facilitar su comparación y análisis conjunto.

2. Recolección de datos

Se usan fuentes y páginas del gobierno de México para buscar y recolectar datos. Seleccionamos datos del año 2020, pues es uno de los años con mayor información disponible. Se usan principalmente datos del Instituto Nacional de Estadística y Geografía, del Consejo Nacional de Evaluación de la Política de Desarrollo Social y del Consejo Nacional de Población. Se tratan datos de las 32 entidades federativas, entonces se tienen 32 datos para cada variable.

2.1. Educación

En la página de indicadores del INEGI, existen muchos indicadores de educación de la población mexicana. Nosotros usamos el *Grado promedio de escolaridad de la población de 15 y más años*, lo cual nos da un panorama de cuanto tiempo duran en la escuela las personas por estado. Los datos vienen para todos los estados de la nación, y no se tuvo que hacer imputación o limpieza.

2.2. Ingreso

Para tener un indicador económico, usamos el *Ingreso corriente a nivel entidad federativa, por año*. Cabe notar que lo anterior es una cifra per capita. Igual que los datos de educación, no hizo falta limpieza ni imputación de datos, y por lo tanto tenemos datos para todos los estados.

2.3. Condición social

En este ámbito usamos el *Índice de Marginación estatal 2020*, publicado por la CONAPO. Este índice se estima mediante varios indicadores:

- Porcentaje de la población analfabeta de 15 años o más.
- Porcentaje de población sin educación básica de 15 años o más.
- Porcentaje de ocupantes en viviendas particulares sin drenaje ni sanitario.
- Porcentaje de ocupantes en viviendas particulares sin energía eléctrica.

- Porcentaje de ocupantes en viviendas particulares sin agua entubada.
- Porcentaje de ocupantes en viviendas particulares con piso de tierra.
- Porcentaje de viviendas particulares con hacinamiento.
- Porcentaje de población en localidades con menos de cinco mil habitantes.
- Porcentaje de población ocupada con ingresos de hasta dos salarios mínimos.

Se explica en la nota técnico-metodológica (CONAPO 2020) como se calculan exactamente estos indicadores. Una vez se tienen todos, se usa la Distancia de Pena Trapero (Trapero 1977), también conocida como la distancia DP_2 , la cual se define de la siguiente manera:

$$DP_2 = \sum_{i=1}^n \frac{d_{ij}}{\sigma_j} (1 - R_{j,j-1,\dots,2,1}^2); \text{ con } R_1^2 = 1.$$

Donde:

- $d_{ij} = |x_{rj} - x_{*j}|$, es la distancia de la j -ésima variable de la entidad federativa r con respecto a la base de referencia.
- σ_j , es la desviación estándar de la variable j .
- $R_{j,j-1,\dots,1}$, es el coeficiente de determinación de la regresión del indicador parcial j con respecto a los otros indicadores $(j-1, j-2, \dots, 1)$.
- $(1 - R_{j,j-1,\dots,1}^2)$, es el factor corrector que evita la duplicidad, al eliminar la información parcial de los indicadores ya contenidos en los indicadores precedentes.
- $R_1^2 = 0$, porque es la primer componente, la cual aporta toda la información y no existe un componente previo.

La anterior distancia hace uso de una base de referencia, pues al ser una distancia necesitamos algo con lo que medir. En este caso se usa el peor caso de la información censal de 2010. Esto significa que entre mayor sea la distancia, mejor será la condición social, pues esta mas alejado del “peor caso”. Los datos, otra vez, no necesitan imputación no limpieza. Se tienen, por lo tanto, datos para todos los estados.

3. Metodología

Normalizamos todas las variables numéricas para mejorar la estabilidad numérica, pues el ingreso se reporta en miles de pesos, el índice de marginación esta en un rango aproximado de 10 a 20 y el promedio de años de escolaridad esta entre 7 y 10, y al normalizar todo se reporta con aproximadamente las mismas unidades. Tenemos, en la Figura 1, la matriz de correlación de las tres variables

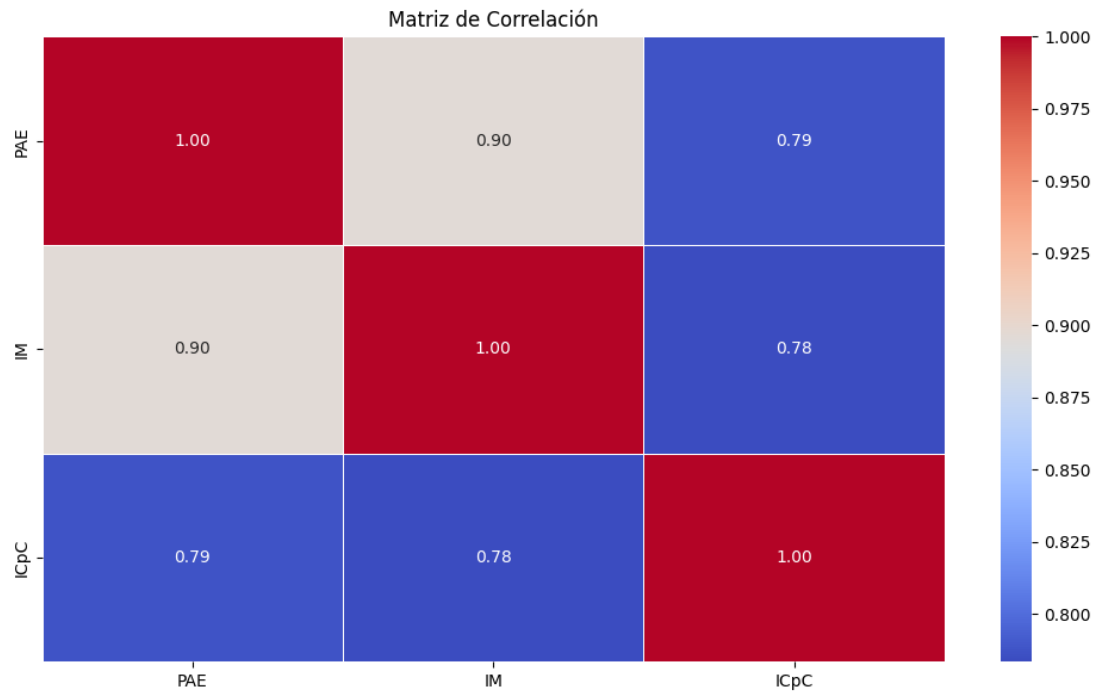


Figura 1: Matriz de correlación, IM vs PAE vs ICpC

Observamos que las variables están altamente correlacionadas, en específico PAE, el Promedio de Años de Escolaridad, el Índice de Marginación están altamente relacionados. Esto es de esperarse, pues en el IM se toma en cuenta porcentaje de población analfabeta y porcentaje de la población sin educación básica, ambos indicadores de la escolaridad de la población, pero también hay que considerar que se consideran otros 7 indicadores a parte de los dos mencionados. EL ICpC, Ingreso Corriente per Capita, esta correlacionado con ambas IM y PAE prácticamente con la misma magnitud.

Si ajustamos un modelo de regresión lineal, usando OLS, con PAE como variable respuesta e ICpC e IM como variables predictiva, usando statmodels en python, tenemos un valor de R^2 de 0.821, y la siguiente tabla de estadísticas de las variables:

	coef	std err	t	$P > t $
IM	0.7213	0.126	5.71	0.0
ICpC	0.225	0.126	1.762	0.89

Los resultados sugieren que PAE se predice de buena manera con el modelo, pero vemos de las estadística que IM es la variable con mayor efecto en el modelo. En efecto, si hacemos regresión lineal solo con PAE e IM, tenemos un valor $R^2 = 0.802$, las estadísticas:

	coef	std err	t	$P > t $
IM	0.8956	0.081	11.028	0.0

Y la gráfica de la línea ajustada en la Figura 2. Vemos de esto que el Índice de Marginación es un buen predictor para el nivel de escolaridad, salvo algún que otro outlier. Cabe mencionar que ambos modelos tienen un valor de Prob (F-statistic) $< 10^{-11}$, entonces ambos modelos son significativos.

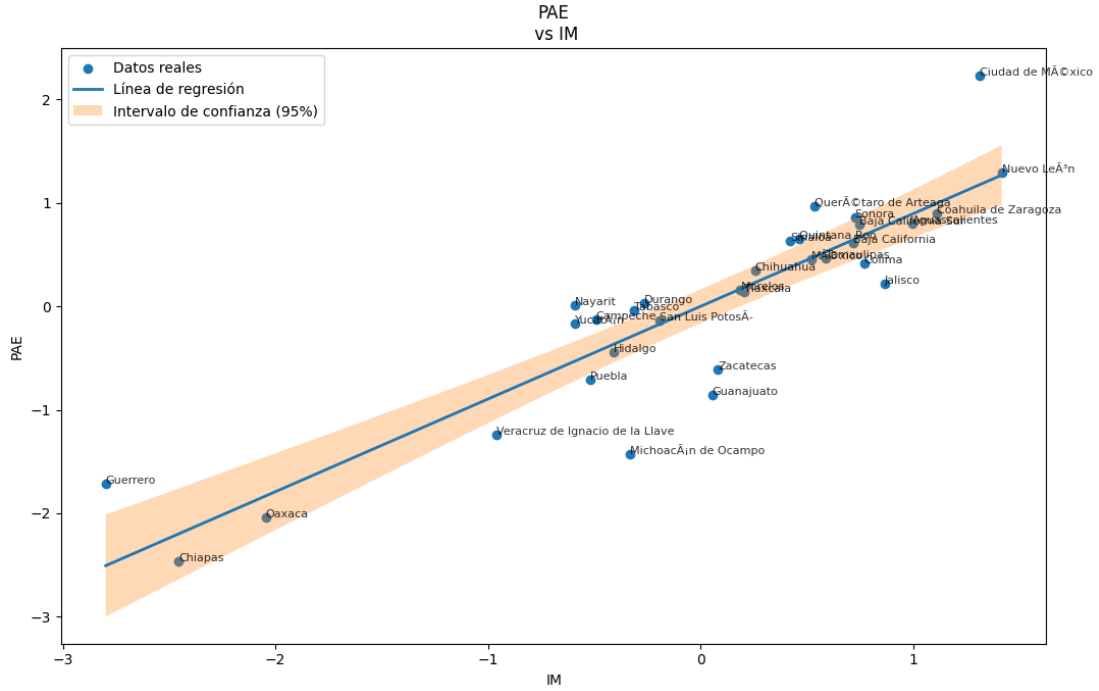


Figura 2: Línea ajustada de regresión lineal, PAE vs IM, IM como variable predictora.

Si, por otro lado, intentamos hacer regresión con ICpC como nuestra variable respuesta, tenemos los resultados:

	coef	std err	t	$P > t $
PAE	0.4345	0.247	1.762	0.089
IM	0.3943	0.247	1.599	0.121

Con una $R^2 = 0.65$. En este caso, es PAE la variable más significativa, y si hacemos también la regresión lineal ICpC vs PAE, tenemos la línea ajustada en la Figura 3.

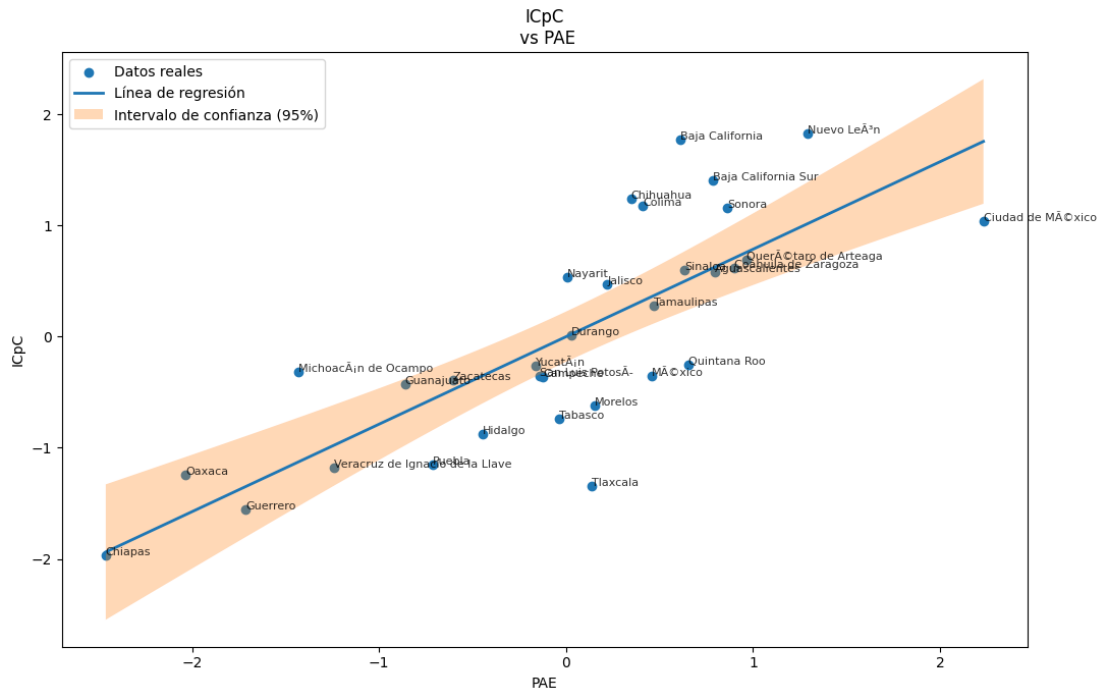


Figura 3: Línea ajustada de regresión lineal, ICpC vs PAE, PAE como variable predictora.

4. Comentarios Finales

Los resultados muestran una relación clara y consistente entre condición social, nivel de ingresos y escolaridad a nivel estatal en México para 2020. La matriz de correlación evidencia una fuerte asociación entre PAE, IM e ICpC; de hecho, el Índice de Marginación (IM) resulta ser el predictor más sólido del Promedio de Años de Escolaridad (PAE).

Los modelos de regresión aportan evidencia cuantitativa: un modelo con PAE como variable respuesta y con IM e ICpC como predictores obtiene un R^2 cercano a 0.821, mientras que un modelo simple $PAE \sim IM$ mantiene un R^2 de aproximadamente 0.802 y un coeficiente altamente significativo ($p < 0.001$). Esto sugiere que las dimensiones que captura el IM (analfabetismo, rezago educativo, carencias en servicios, hacinamiento, ruralidad) explican la mayor parte de la variación en escolaridad entre entidades. Por otro lado, al modelar ICpC como respuesta, perdemos un poco del poder predictivo, lo que indica que la relación ingreso-educación es real pero más compleja y probablemente mediada por otros factores no incluidos en este análisis.

También cabe notar que este fue un análisis bastante superficial, con un enfoque en manejo y análisis de datos, mas que en búsqueda de causas y soluciones a la desigualdad educativa, económica y social en México.

Referencias

1. **INEGI** (2022). *Grado promedio de escolaridad de la población de 15 y más años (Años de escolaridad), Estados Unidos Mexicanos, 2020*. Datos seleccionados mediante menús desplegables: “Demografía y Sociedad → Educación → Características educativas de la población → Grado promedio de escolaridad de la población de 15 y más años (Años de escolaridad)”. inegi.org.mx/indicadores.
2. **CONAPO** (2025). *Índice de marginación estatal 2020*. Base de datos. datos.gob.mx/indices_marginacion.
3. **CONEVAL** (2025). *Ingreso corriente a nivel entidad federativa, por año*. Base de datos. datos.gob.mx/pobreza_mexico.
4. **CONAPO** (2020). *Índice de marginación por entidad federativa y municipio 2020*. Consejo Nacional de Población (CONAPO), México. Gobierno de México. gob.mx/Nota_tecnica_marginacion_2020.
5. Trapero, X. B. P. (1977). *Problemas de la medición del bienestar y conceptos afines: (una aplicación al caso español)*. Presidencia del Gobierno, Instituto Nacional de Estadística. ISBN: 9788426001788. books.google.com.mx.