

# The Anáhuac Journal

Business and Economics

The Academic Journal of the Universidad Anáhuac-México Sur

---

Volume 16, Number 1 • First Semester 2016

ISSN 1405-8448

Edited by Antonio García de la Parra Motta



LIDeditorial  
.com

# Estudio de indicadores de bienestar por entidad federativa en México usando análisis de componentes principales

Iñigo Aranzábal Martín, Carlos Carredano Riega, Hugo Alexer Pérez Vicente

## Resumen

*La calidad de vida en un país es discutible y su estudio ha sido controversial dadas las diferentes perspectivas de medición planteadas a lo largo del tiempo. En este trabajo se propone un estudio acerca de los indicadores de la calidad de vida de las entidades federativas en México. La importancia recae en el interés de proponer una metodología que apunte al adecuado establecimiento de estrategias para mejorar el bienestar de vida de las personas. Para ello se ha planteado usar la técnica de análisis de componentes principales para describir el patrón en las variables consideradas y que están directamente relacionadas a la calidad de vida en cierto estado. Los resultados preliminares y las conclusiones nos darán a conocer cuáles y a qué nivel estas características están relacionadas.*

**Palabras clave:** análisis de componentes principales (PCA), calidad de vida, correlación lineal, prueba de normalidad.

**Clasificación JEL:** C02, C81, H10 y J17.

## Abstract

*The quality of life in a country is debatable, and controversial study has been given the different perspectives of measurement for many years. In this paper a study on living standards by state in Mexico is presented. The importance lies in the interest of proposing a methodology that points to the proper establishment of strategies to improve the welfare of people's lives. For it has been raised using the technique of principal component analysis to describe the pattern in the variables considered and that are directly related to the quality of life in a certain State in Mexico. The preliminary results and conclusions will let us know what and at what level these characteristics are related.*

**Keywords:** *Principal Component Analysis (PCA), Quality of Life, Linear Correlation, Normality Test.*

**JEL Classification:** *C02, C81, H10 and J17.*

## 1. Introducción

Los métodos actuales para dar una medición de bienestar en México indican que se han logrado grandes avances en los niveles de calidad de vida durante las últimas décadas. Según se menciona en García Vega y Sales Heredia (2011), se han registrado mejores resultados en los niveles de pobreza, niveles de educación, tasas de mortalidad infantil y esperanza de vida, entre otros. En este mismo estudio se argumenta que los indicadores son comparables con los usados para medir el bienestar en América Latina y Asia, y muestran que, aunque con diferentes patrones, estos se mueven según los casos específicos de cada país (crisis económicas y factores internos); México presenta niveles muy similares a los promedios de estas regiones.

Para estudiar los avances en los niveles de bienestar se utilizan los distintos índices que generan, a partir de datos de indicadores macroeconómicos y encuestas de percepción hechas a la población de forma directa para su elaboración. Algunos de estos índices, como el Índice Nacional de Calidad de Vida (INCAV), se usan como indicadores del avance social que planean y modifican las políticas sociales del país o de ciertas regiones (Gabinete de Comunicación Estratégica, 2015). En particular, el INCAV considera que una región tiene rezago social o que alguna otra tiene mejores niveles de bienestar. Una de las finalidades de los indicadores como este último, es resumir grandes cantidades de información en un solo número, para que sea comparable entre regiones o a través de los años en una misma región.

En este trabajo se usa el análisis de componentes principales (PCA, por sus siglas en inglés) –técnica de análisis multivariado– con la intención de encontrar tendencias y patrones en un conjunto de datos indicadores de bienestar sin tener que resumir la información en algún índice general. Se pretenden lograr conclusiones específicas sobre las relaciones entre un indicador y otro, o también entre una región y otra, a partir de los datos originales recabados por el Instituto Nacional de Estadística y Geografía (INEGI). Esta forma de analizar las cifras nos podría proporcionar una herramienta diferente y crucial para comprender los datos disponibles y usarlos en forma útil y contundente.

## 2. Antecedentes

En la actualidad, existen muchas formas diferentes de medir y estudiar el bienestar de las personas, estas han evolucionado a lo largo del tiempo y se han vuelto cada vez más complejas. Además, persiste la polémica que

cuestiona si los métodos usados son los adecuados a lo que se encuentra en la literatura como diferentes propuestas de mejora. Durante la década de los sesenta en el siglo XX se utilizaban solamente variables macroeconómicas para determinar el bienestar relativo de una población. No fue hasta 1974 que Drewnowski, pionero de la medición de la calidad de vida, comenzó a destacar la importancia de medir más variables (Drewnowski, 1974). La Organización para la Cooperación y el Desarrollo Económico (OCDE, 1976) publicó el reporte sobre el desarrollo de indicadores de bienestar, y Morris (1979) construyó el Índice de Calidad Física de Vida de su clase, con el cual buscaba medir y relacionar diferentes variables de bienestar para calcular índices de pobreza, desigualdad y educación, entre otros, como se comenta en Berumen (2004). Hoy en día, diferentes instituciones y estudiosos proponen alternativas a la medición del bienestar que difieren entre ellas, como se menciona en García Vega y Sales Heredia (2011).

Tradicionalmente, en México el INCAV, propuesto por la Universidad de Monterrey, se utiliza para medir la calidad de vida. Este considera la percepción ciudadana en los rubros de suficiencia y precios de vivienda, suficiencia de escuelas, de espacios de esparcimiento y diversión, de espacios culturales y bellezas naturales, movilidad y aire limpio, que después de un tratamiento estadístico da como resultado un valor entre 0 y 100 comparable con el resultado de la misma región en años anteriores o con el resultado de otras regiones en ese mismo año (Gabinete de Comunicación Estratégica, 2015). Muchos autores presentan alternativas a esta medición y proponen métodos para encontrar resultados más realistas y acertados. Por ejemplo, el estudio *Bienestar y calidad de vida en México*, editado por la Universidad de Monterrey y el Centro de Estudios Sociales de Opinión Pública de la Cámara, presenta cinco textos en los que los autores critican las fallas que tienen los métodos usados y en algunos casos proponen mejoras. Cada autor tiene su propia crítica, pero todos ellos tienen una tendencia común: la búsqueda de un índice que represente mejor la calidad de vida de una población.

Lo crucial que se presenta en los estudios realizados en el campo de los Índices de Calidad de Vida está precisamente en la definición de calidad de vida y la subjetividad que conlleva (García Vega, 2011). Es así que sigue siendo un reto para los analistas saber qué medir, ya que hay una infinidad de factores que se podrían considerar y cada persona los ponderaría de forma diferente de acuerdo a su importancia. Por otro lado, se tienen cantidades inmensas de información sobre calidad de vida y bienestar, por lo que el reto también radica en saber qué hacer con ella. La motivación del presente trabajo se centra en ese intento por extraer un conocimiento que ayude a comprender la calidad de vida desde una perspectiva diferente y, en la medida de lo posible, con las herramientas que se tengan al alcance.

### 3. Descripción del estudio

Para la elección de variables, en este estudio se optó por considerar ciertos criterios particulares. Sin ser exhaustivo y para los fines de este trabajo, se eligió un grupo de variables con miras a que existiera alguna relación entre los indicadores de bienestar en México. Los datos se obtuvieron del INEGI<sup>1</sup> porque: 1) dan confiabilidad al estudio, 2) están basados en información real, 3) tuvieron la característica de no tener datos faltantes y 4) es posible su repetitividad para estudios posteriores.

Las ocho variables consideradas fueron: accesos a servicios de salud de 2014, niveles de educación de 2010, GINI del ingreso disponible de hogares per cápita de 2014, contaminación del aire de 2012, tasa de pobreza de 2014, esperanza de vida al nacer de 2015, tasa de homicidios de 2013 y participación electoral de 2012; todos ellos fueron obtenidos para cada entidad federativa; es decir, 31 estados de la República Mexicana y el Distrito Federal (ahora Ciudad de México). Es cuestionable la incompatibilidad en el tiempo de observación de cada variable; no obstante, se espera que los resultados, lejos de arrojar conclusiones similares, no pierdan objetividad en el presente estudio. A continuación, en la tabla 1 se muestra la descripción de las variables analizadas:

**Tabla 1.** Descripción de las variables de estudio

Dimensión	Escala de medición	Unidad
Servicios de salud	R	%
Niveles de educación	R	%
GINI del ingreso	R	coef.
Contaminación aire	R	µg*m <sup>3</sup>
Tasa pobreza	R	%
Esp. de vida al nacer	I	año
Tasa de homicidios	I	persona
Participación electoral	R	%

Nota: Se evidencia que todas las variables de estudio son variables continuas. En particular, refiriéndose a la tabla 1, se tiene que para Escala de Medición (I = Intervalo, R= Razón); Unidad (% = porcentaje, coef. = coeficiente entre 0 y 1, µg\*m = microgramos por metro cúbico, año = años de vida, persona = homicidios por cada 100,000 habitantes).

Fuente: elaboración propia.

<sup>1</sup> Fuente de datos disponible en: <http://www3.inegi.org.mx/app/bienestar/>.

### *Materiales y métodos*

El procedimiento de análisis inicia con la exploración de las variables para revisar posibles relaciones lineales entre pares de ellas. Esto precisó la realización e interpretación de las matrices de correlación y covarianza, respectivamente. Posteriormente, se procedió a la aplicación de la técnica PCA (*Principal Component Analysis*) que sirve para analizar la varianza total de un conjunto de datos (Hair, JF., Anderson, E., Tatham, L., y Black, C., 1999). El análisis PCA es una técnica multivariante que busca combinaciones lineales entre las variables. Su objetivo es describir la variación en un conjunto de variables correlacionadas, en términos de un nuevo conjunto de variables no correlacionadas (Everitt, B. y Hothorn, T., 2011). La idea general es que los primeros pocos componentes cuenten con una proporción sustancial de la variación observada en las variables originales, esto también como un método para reducir la dimensionalidad de un conjunto de datos (De la Garza, J., Morales, B.N. y González, B.A., 2013). Intuitivamente, la técnica sirve para hallar las causas de la variabilidad de un conjunto de datos y ordenarlas por importancia (Winston, 2014). El análisis PCA sirve para los fines de este estudio, ya que puede indicar qué relaciones de bienestar están relacionadas con cuáles. Además, se pretende encontrar relaciones que no sean triviales a simple vista y así concluir que, para mejorar algún tipo de factor, se necesita trabajar en alguno, o que dos factores no están relacionados entre sí. Por tanto, se lograría entender de manera más detallada lo que estos factores o indicadores significan, lo cual es el objetivo del presente trabajo. Para la realización del estudio se utilizó el programa estadístico Minitab 17 y el programa comercial MS Excel.

El tratamiento de los datos empieza con su exploración; su naturaleza define qué se va a usar: puede ser la matriz de covarianza (si los datos están en las mismas unidades) o la matriz de correlación (si los datos están en distintas unidades). De este mismo modo, para nuestro análisis ocuparemos diagramas de cajas y bigotes para detectar datos atípicos, diagramas de dispersión para visualizar relación lineal entre variables y las pruebas de normalidad (univariada y multivariada).

Después de la etapa exploratoria se procede a realizar el PCA, considerando la elección del número de componentes a utilizar. Como se menciona en Everitt, B. y Hothorn, T. (2011), el método más popular consiste en revisar el eigenvalor, que implica retener aquellos componentes principales cuyo eigenvalor sea mayor o igual a uno. Esta elección también se puede basar en la varianza acumulada o en un diagrama de sedimentación. Finalmente se debe analizar cada diagrama o gráfica, así como las pruebas pertinentes, para obtener conclusiones y poder interpretar de manera adecuada los resultados.

En este tipo de estudios es muy importante comprender los pasos de lo que se va a hacer, como también conocer las variables y sus connotaciones para poder hacer una predicción de lo que es más posible que ocurra. En el PCA, la interpretación se realiza por componente y se estudian principalmente los tipos de relación entre las variables dentro de cada «grupo» formado. Por lo tanto, es factible y no imposible suponer de alguna forma los resultados antes de llevar a cabo la metodología.

Es muy probable que los servicios de salud estén directamente relacionados con los de educación. Así como el GINI del ingreso con la tasa de pobreza. Por otro lado, se podría suponer que la tasa de homicidios está fuertemente relacionada con la esperanza de vida. Estas aseveraciones se hacen de acuerdo a lo que cualquier persona podría sugerir o inferir. No son más que simples opiniones o predicciones de lo que el analista cree que el estudio mostrará.

#### *Análisis previo de los datos*

La tabla 2 muestra los valores de covarianza de cada una de las variables con respecto a todas las otras y a sí misma. Cabe señalar que el primer renglón corresponde a la primera variable descrita y así sucesivamente. La covarianza es un concepto muy parecido al de varianza, es por eso que estos datos no son comparables entre sí, ya que las unidades en las que se encuentran influyen en el resultado (Everitt, B. y Hothorn, T., 2011). Por definición, los datos en la diagonal representan las varianzas de las variables.

**Tabla 2.** Matriz de covarianza

$$\begin{pmatrix} 10.756 & 7.303 & -0.014 & -4.607 & \mathbf{-22.413} & 1.285 & 0.618 & 1.734 \\ 7.303 & 43.320 & -0.033 & -2.012 & \mathbf{-72.473} & 4.124 & -0.466 & 0.488 \\ -0.014 & -0.0335 & 0.0009 & -0.010 & 0.136 & -0.005 & -0.077 & 0.060 \\ -4.607 & -2.012 & -0.010 & 22.140 & 14.081 & 0.478 & 0.840 & 2.505 \\ \mathbf{-22.413} & \mathbf{-72.473} & 0.137 & 14.081 & 180.967 & -8.300 & -6.164 & 16.366 \\ 1.286 & 4.124 & -0.006 & 0.478 & -8.300 & 0.951 & -4.761 & 0.609 \\ 0.619 & -0.466 & -0.078 & 0.840 & -6.164 & -4.761 & 214.665 & \mathbf{-34.001} \\ 1.734 & 0.488 & 0.061 & 2.505 & 16.366 & 0.609 & \mathbf{-34.001} & 26.221 \end{pmatrix}$$

Fuente: elaboración propia.

En la tabla 2 los valores más altos a consideración del analista, y sin seguir ninguna regla específica, están marcados en negritas, con excepción de las varianzas que se encuentran en la relación de cada variable con ella misma. Las relaciones marcadas son:

- Tasa de pobreza con acceso a servicios de salud
- Tasa de pobreza con niveles de educación
- Participación electoral con tasa de homicidios

La tabla 3 muestra los coeficientes de correlación de Pearson ( $\rho$ ) de cada una de las variables con respecto al resto de las variables, suponiendo que existe algún tipo de correlación lineal aparente. Estos datos son comparables entre sí ya que se realizan a partir de datos estandarizados (con media cero y varianza uno).

**Tabla 3.** Matriz de correlación

1	0.388	-0.141	-0.299	-0.508	0.402	0.013	0.103
0.338	1	-0.169	-0.065	-0.819	0.642	-0.005	0.014
-0.141	-0.169	1	-0.074	0.337	-0.191	-0.175	0.391
-0.299	-0.065	-0.074	1	0.222	0.104	0.012	0.104
<b>-0.508</b>	<b>-0.819</b>	0.337	0.222	1	-0.633	-0.031	0.238
<b>0.402</b>	<b>0.642</b>	-0.191	0.104	<b>-0.633</b>	1	-0.333	0.122
0.013	-0.005	-0.175	0.012	-0.031	-0.333	1	-0.453
0.103	0.014	0.391	0.104	0.238	0.122	<b>-0.453</b>	1

Fuente: elaboración propia.

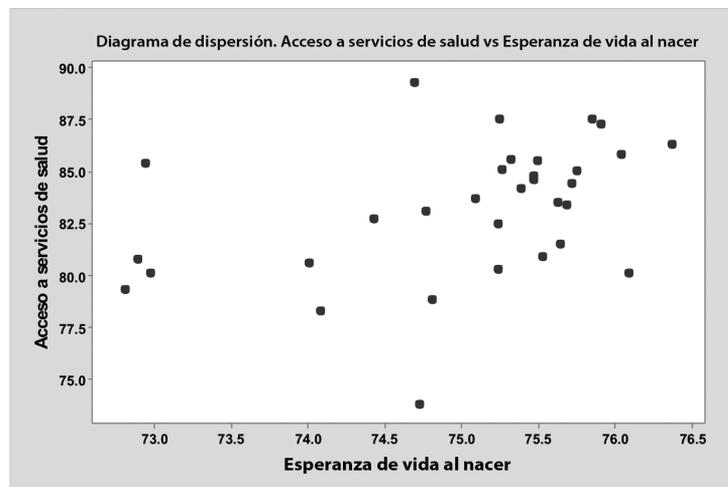
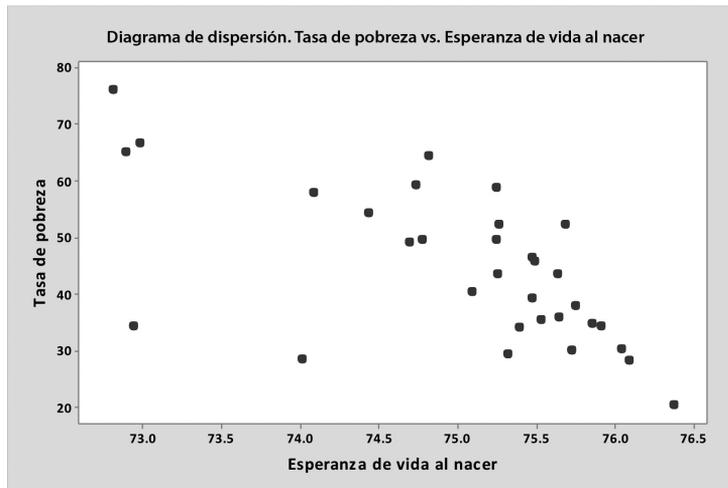
Considerando  $\rho > |0.5|$  como alta (marcados en negritas de la tabla 3) se encontraron las siguientes correlaciones:

- Tasa de pobreza con acceso a servicios de salud
- Tasa de pobreza con niveles de educación
- Esperanza de vida al nacer con acceso a servicios de salud
- Esperanza de vida al nacer con niveles de educación
- Esperanza de vida al nacer con tasa de pobreza
- Participación electoral con tasa de homicidios

Estas relaciones serán analizadas más adelante por medio de la técnica PCA; sin ser contundentes a nivel de par de variables, la idea es intentar generalizar o extraer conclusiones en conjunto.

Los diagramas de dispersión muestran de forma visual la relación entre dos variables. Estos grafican puntos que para nuestro caso representan los diferentes estados con el valor de una variable en el eje  $x$  y el de otra en el eje  $y$  en un plano cartesiano. Se realizaron todos los diagramas de dispersión



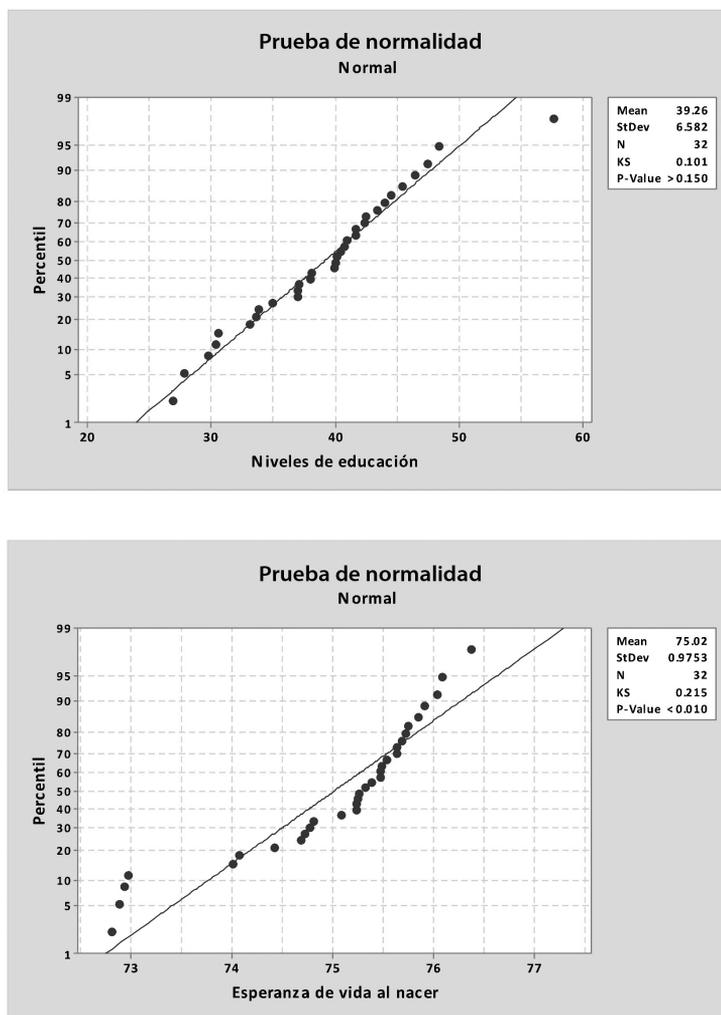


Fuente: elaboración propia.

Pese a que una de las bondades del PCA es que no supone alguna distribución específica en los datos debido a que es un análisis del tipo no-paramétrico, se realizó una prueba de normalidad a cada una de las variables, usando la prueba Kolmogorov-Smirnov (KS). El programa usado nos muestra el gráfico P-P, el estadístico KS y su correspondiente valor-p. Si los puntos se asemejan a la línea recta del gráfico y mayor sea el valor-p dado cierto

nivel de significancia –usualmente un valor de 0.05– no podría rechazarse la hipótesis de normalidad en los datos. En la figura 2 se muestran dos gráficos P-P, y a continuación los resultados de los valores-p en la tabla 4.

**Figura 2.** Gráficos P-P para la variable nivel de educación (arriba) y esperanza de vida al nacer (abajo)



Fuente: elaboración propia.

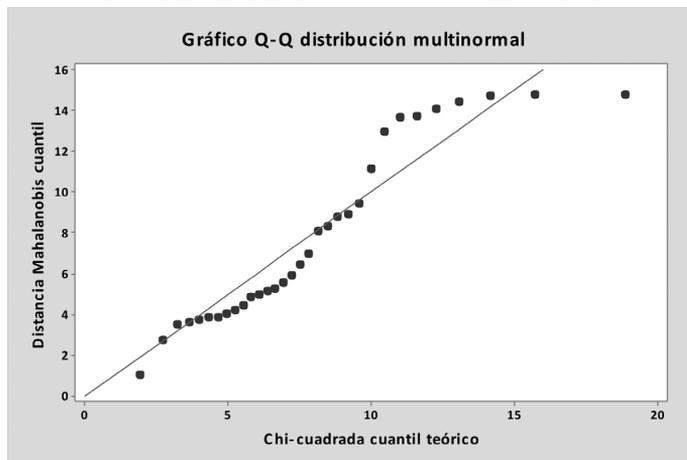
**Tabla 4.** Resultados de la prueba de normalidad KS

Variable	Valor-p	¿Sigue una distribución normal?
Servicios de salud	>0.150	SÍ
Niveles de educación	>0.150	SÍ
GINI del ingreso	>0.150	SÍ
Contaminación aire	>0.094	SÍ
Tasa pobreza	>0.150	SÍ
Esp. de vida al nacer	>0.100	NO
Tasa de homicidios	0.480	NO
Participación electoral	>0.150	SÍ

Fuente: elaboración propia.

La prueba de la normalidad multivariada se realizó con ayuda del gráfico Q-Q. Como se sugiere (Everitt, B. y Hothorn, T., 2011), se observa si la Distancia Mahalanobis (MD) sigue una distribución Chi-cuadrada con grados de libertad igual al número de variables. Entre más se acerquen las distancias Mahalanobis a la distribución Chi-cuadrada, se aproximarían a una distribución normal multivariada en una equivalencia 1:1. Es por eso que en los casos en los que la distribución es multinormal, el gráfico Q-Q se aproximará a una línea recta (ver figura 3).

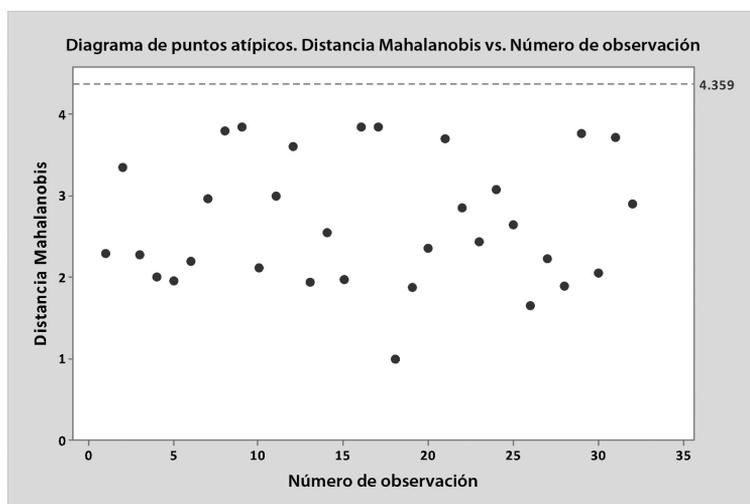
**Figura 3.** Gráfico Q-Q usado para detectar una distribución multinormal



Fuente: elaboración propia.

En la figura 4 se muestra la dispersión de las MD, y con el punto de corte que calcula Minitab basado en la distribución F, no podría observarse ningún punto atípico multivariado. Adicionalmente, se calculó un punto de corte por medio de la distribución Chi-cuadrada sin cambiar las conclusiones obtenidas.

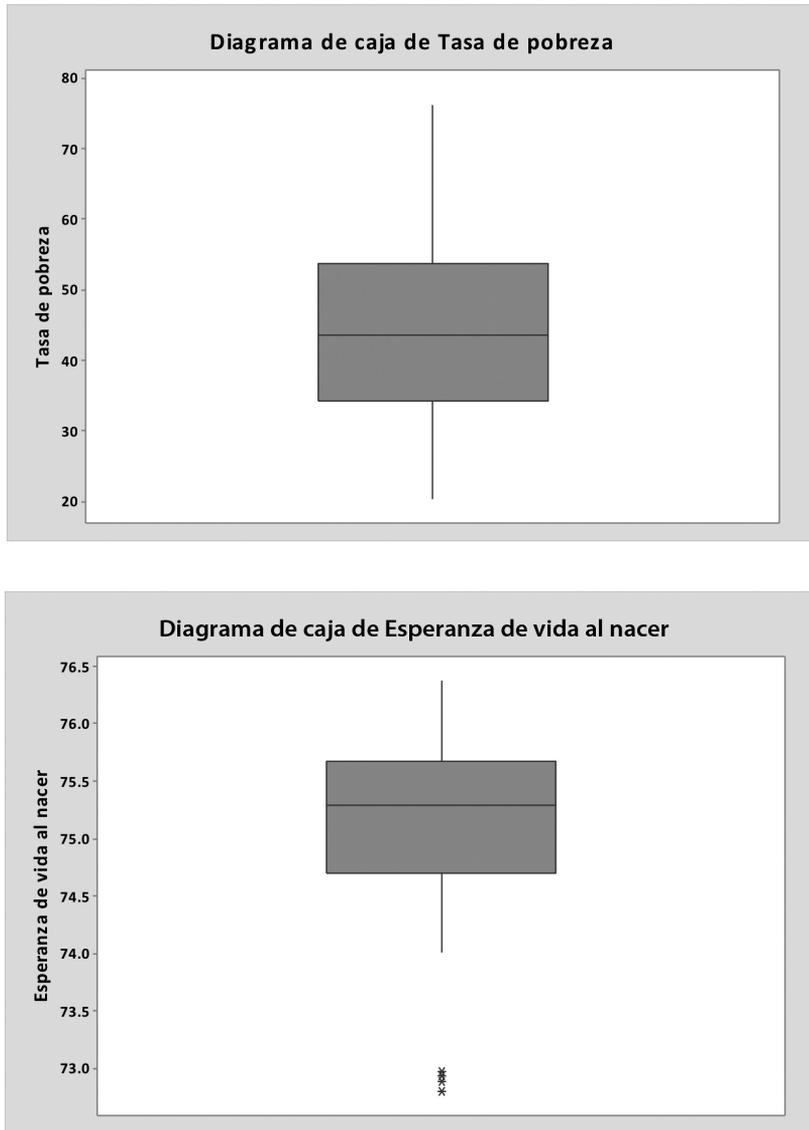
**Figura 4.** Diagrama de dispersión de las MD respecto al número de observación



Fuente: elaboración propia.

Para encontrar los valores atípicos por variable se aplicaron dos métodos: 1) el rango intercuartil (IQR, por sus siglas en inglés) mostrado en un diagrama de cajas y bigotes y el 2) el z-score. En los gráficos de la figura 5 se muestran dos diagramas de cajas y bigotes, uno sin puntos atípicos y otro con puntos atípicos, respectivamente. En la tabla 5 se muestran los datos atípicos encontrados en los diagramas.

**Figura 5.** Diagramas de caja de algunas de las variables



Fuente: elaboración propia.

**Tabla 5.** Datos atípicos de diagramas de caja

Variable	Punto atípico	Valor
Esperanza de vida al nacer	Oaxaca	72.98
Esperanza de vida al nacer	Chihuahua	72.94
Esperanza de vida al nacer	Guerrero	72.89
Esperanza de vida al nacer	Chiapas	79.30
Participación electoral	Yucatán	77.40
Tasa de homicidios	Chihuahua	58.80
Tasa de homicidios	Guerrero	64.80
Contaminación del aire	Morelos	26.20
GINI del ingreso disponible	Puebla	0.572
Educación	Distrito Federal	57.60

Fuente: elaboración propia.

Usando el segundo método que se basa términos de desviaciones estándar, se definieron como datos atípicos aquellos valores que sobrepasan  $\pm 3\sigma$ , con los siguientes resultados:

**Tabla 6.** Datos atípicos por método z-score

Variable	Punto atípico	Valor
Tasa de homicidios	Guerrero	3.1315
Contaminación del aire	Morelos	3.5020
Ingresos de hogares p/c	Puebla	3.0959

Fuente: elaboración propia.

### *Aplicación de la técnica PCA*

Una vez realizada la parte exploratoria, se procedió a aplicar la técnica multivariada. El análisis de los resultados sugiere determinar el número de

componentes principales a considerar. En este caso, aplicando el criterio del eigenvalor, se considerarán tres componentes. Debido a la presencia de valores atípicos se optó por dividir el análisis en dos casos: I) en el que se usa la información completa y II) en el que se quitaron los valores atípicos detectados con el método 2. En la tabla 7 se muestran los resultados del análisis, incluidos los puntos atípicos.

**Tabla 7.** Resultados del PCA: caso I

	PC1	PC2	PC3	PC4
Eigenvalor	2.8412	1.7715	1.2003	0.7548
Proporción	0.355	0.221	0.150	0.094
Acumulado	0.355	0.577	0.727	0.821

Fuente: elaboración propia.

Únicamente se tomaron los tres primeros componentes principales (PC1, PC2 y PC3), ya que a partir del cuarto el eigenvalor es menor que 1. También se obtuvieron los valores de dichos componentes principales, que se muestran en la tabla 8.

**Tabla 8.** Carga de cada componente para las variables de estudio: caso I

Dimensión	PC1	PC2	PC3	Com. (%)
Servicios de salud	-0.381	-0.066	-0.381	29.5
Niveles de educación	-0.505	-0.056	0.098	26.8
GINI del ingreso	0.230	-0.405	-0.324	32.2
Contaminación aire	0.114	-0.072	0.812	67.7
Tasa pobreza	0.554	-0.103	0.034	31.9
Esp. de vida al nacer	-0.470	-0.259	0.271	36.1
Tasa de homicidios	0.038	0.581	-0.071	34.4
Participación electoral	0.061	-0.639	-0.035	41.3

Fuente: elaboración propia.

En la tabla 8 se muestran los valores de cada componente principal en las primeras tres columnas y en la última la comunalidad (Com.) de cada variable, se llega a la conclusión de que la variable con mayor peso en la medición del bienestar de la población es la contaminación del aire.

Posterior a la obtención de los resultados, se decidió expandir el análisis y se realizó, como se mencionó anteriormente, el PCA sin incluir las observaciones que contienen datos atípicos. Los resultados obtenidos se muestran en la tabla 9.

**Tabla 9.** Resultados del PCA: caso II

	PC1	PC2	PC3	PC4
Eigenvalor	2.7648	1.8417	1.3847	0.7352
Proporción	0.346	0.230	0.173	0.092
Acumulado	0.346	0.576	0.749	0.841

Fuente: elaboración propia.

Como se puede observar, se volvieron a obtener los mismos resultados que con los datos con valores atípicos. Nuevamente, se obtuvieron los valores de los tres componentes a analizar, ahora sin valores atípicos univariados. Esto se muestra en la tabla 10:

**Tabla 10.** Carga de cada componente para las variables de estudio: caso II

Dimensión	PC1	PC2	PC3	Com. (%)
Servicios de salud	0.353	0.188	0.373	29.9
Niveles de educación	0.509	0.148	-0.089	28.9
GINI del ingreso	-0.186	0.420	0.403	37.3
Contaminación aire	-0.115	0.010	-0.725	53.9
Tasa pobreza	-0.567	0.027	0.015	32.2
Esp. de vida al nacer	0.453	0.274	-0.280	35.9
Tasa de homicidios	0.148	-0.537	0.294	39.7
Participación electoral	-0.143	0.634	0.023	42.3

Fuente: elaboración propia.

La tabla 10 muestra los valores de cada componente principal en las primeras tres columnas y en la última, la comunalidad. Al comparar con la tabla 8 vemos que los valores y comunalidades son muy similares para cada componente principal y para cada variable. Como se observa en la tabla 10, la contaminación del aire sigue siendo el factor que más afecta el bienestar de la población.

#### 4. Discusión y conclusiones

En el presente trabajo se estudian variables relacionadas con la calidad de vida en México, con el uso de la técnica PCA. La interpretación de los componentes principales muestra relaciones interesantes que podrían utilizarse para definir estrategias de carácter público. Los resultados indican la posibilidad de ser estudiados usando una técnica multivariada descriptiva disponible en cualquier programa informático de uso libre o comercial.

La discusión de los resultados se basará en la interpretación de los valores de los componentes principales de la tabla 8. En términos generales, se podría sustentar que: en PC1 las variables de nivel de educación y tasa de pobreza están inversamente relacionadas, lo que podría significar que a mayor educación se tendría menor tasa de pobreza. Asimismo, considerando la esperanza de vida al nacer se observa que también está relacionada con estas dos variables. Para el caso de PC2, las variables tasa de homicidios y participación electoral están inversamente relacionadas, dando a entender que a menor participación electoral de la población, los homicidios en la comunidad aumentan. Finalmente, en PC3 la variable de contaminación de aire tiene una relación fuerte con el nivel de bienestar de la población, inherente a lo que dicta el sentido común.

De acuerdo a la interpretación de los resultados del análisis PCA, el presente trabajo acusa lo siguiente:

- En cuanto al primer componente, el nivel de educación de un estado y su tasa de pobreza estarán relacionadas de forma inversa; es decir, cuando uno sea alto, el otro será bajo. La esperanza de vida al nacer está relacionada con menor fuerza a estas variables, en el mismo sentido que el nivel de educación. Además, estas variables tendrán un peso elevado en los indicadores de bienestar. Esta conclusión es bastante trivial y el análisis no hace más que reafirmar con sustento matemático que al reducir o aumentar una de estas variables la otra se afectará de forma inversa. Se podría decir entonces que para disminuir la pobreza, se tendría que aumentar el nivel de educación en el estado o que para mejorar la educación y alargar la esperanza de vida al nacer, se tendrían que tomar medidas para reducir la pobreza.

- Del componente dos puede verse una relación un poco menos trivial y que sin un análisis adecuado es más difícil detectar. La tasa de homicidios en una región está inversamente relacionada con la participación electoral de la población en esa región. Esto sugiere que los estados en los que vota más porcentaje de gente son aquellos en los que menos homicidios por habitante suceden y viceversa. Es una relación que, desde nuestro punto de vista, resulta interesante y que no se sospechaba antes de hacer este análisis. Pareciera que existe un factor cívico y ético que provoca que estas dos medidas de bienestar estén fuertemente relacionadas.
- Del tercer componente puede deducirse que la calidad del aire tiene mucha influencia en el bienestar de las personas. Aunque no se haya relacionado con ninguna otra variable analizada, esta registró un peso notablemente alto. Este resultado es más difícil de conceptualizar pues significa que aunque no se registran patrones que relacionen esta variable con el resto, la variable tuvo una fuerte influencia en el bienestar general de las entidades que se representan por medio de estos datos.

Este tipo de análisis pretende dar bases para estudiar a un nivel más detallado respecto de las medidas que los diferentes estados de la República Mexicana deben tomar para generar mejores condiciones de bienestar de las personas. En este sentido, el diseño de políticas públicas resulta un tema crucial en el estudio de la calidad de vida en México. No obstante, esta tarea seguirá siendo importante para los organismos de gobierno como la Secretaría de Desarrollo Social o de Desarrollo Económico de las entidades federativas, así como en foros mundiales de la Organización para la Cooperación y el Desarrollo Económico (OCDE). Los resultados preliminares de este estudio sugieren enfocarse en generar una mayor participación electoral y una cultura de la no violencia o la reducción del homicidio. Adicionalmente, se evidencia la importancia del fortalecimiento de la educación, por estar relacionada con los niveles de pobreza en México.

Como trabajo a futuro, se pretende repetir el análisis con un mayor número de variables con características sugeridas por un grupo de especialistas de diversas áreas de la Economía y la Sociología, apuntando a la realización de un análisis comparativo con otras investigaciones y dar sentido a las variables elegidas en este trabajo. Además, resulta de vital importancia proponer alguna estrategia de análisis para estudiar variables que provienen de diferentes periodos de tiempo con la finalidad de dar una mejor consistencia a las conclusiones obtenidas.

## Los autores

Iñigo Aranzábal Martín es alumno de noveno semestre de Ingeniería Industrial de la Facultad de Ingeniería en la Universidad Anáhuac Norte. Ha sido distinguido con beca de excelencia por alto desempeño a lo largo de sus estudios de licenciatura. En 2015 realizó un intercambio académico en la Universidad Católica de Lyon, en Francia. Sus líneas de interés son Optimización y Minería de Datos. Actualmente está colaborando en un proyecto de investigación para dicha Facultad de Ingeniería que será presentado en el Congreso Internacional PATAT 2016.

[inigoaran@hotmail.com](mailto:inigoaran@hotmail.com)

Carlos Carredano Riega es alumno de noveno semestre de Ingeniería Industrial de la Facultad de Ingeniería en la Universidad Anáhuac Norte. Ha sido distinguido con beca de excelencia por alto desempeño. Sus líneas de interés son Optimización y Minería de Datos.

[carrecar93@hotmail.com](mailto:carrecar93@hotmail.com)

Hugo Alexer Pérez Vicente es maestro en Ciencias en Ingeniería de Sistemas por la Universidad Autónoma de Nuevo León (UANL) e ingeniero industrial por el Instituto Tecnológico de Tuxtla Gutiérrez (ITTG). Ha expuesto en congresos nacionales e internacionales como el MICAI, CLAIO-SMIO y SIMMAC. Además, ha publicado artículos en revistas indexadas y *proceedings* en *IIE Annual Conference* y CIINDET. Es profesor en la Universidad Anáhuac Norte donde estudia el Doctorado en Ingeniería Industrial. Sus líneas de especialidad son Minería de Datos, Cadena de Suministro y Estadística Aplicada.

[hugo.perez@anahuac.mx](mailto:hugo.perez@anahuac.mx)

### Referencias

- Berumen, Sergio A. (2004). Construcción y análisis del índice de la calidad de vida en Guatemala (PQLI). *Estudios Económicos de Desarrollo Internacional*. 4(2), 73-92.
- De la Garza, J., Morales, B.N. y González, B.A. (2013). *Análisis estadístico multivariante*. México: Mc Graw Hill.
- Drewnowski, J. (1974). *On Measuring and Planning the Quality of Life*. La Haya: Mouton.
- Everitt, B. y Hothorn, T. (2011). *An Introduction to Applied Multivariate Analysis with R*. Nueva York: Springer.
- Gabinete de Comunicación Estratégica. (2015). *Las ciudades más habitables de México 2015*. México: GCE.
- García Vega, J.J. y Sales Heredia, F.J. (Eds.) (2011). *Bienestar y calidad de vida en México*. México: Centro de Estudios Sociales y de Opinión Pública y Universidad de Monterrey.
- García Vega, J.J. (2011). Hacia un nuevo sistema de indicadores de bienestar. *Revista internacional de Estadística y Geografía*, 2(1), 78-95.
- Hair, J.F., Anderson, E., Tatham, L. y Black, C. (1999). *Análisis multivariante*. Madrid: Prentice Hall International, Inc.
- Morris, M. D. (1979). *Measuring the Condition of the World's Poor: The Physical Quality of Life Index*. Nueva York: Pergamon Press.
- Organización para la Cooperación y el Desarrollo Económico [OCDE]. (1976). *Measuring Social Well-Being: A Progress Report on the Development of Social Indicators*. París: OCDE.
- Winston, W. (2014). *Marketing Analytics: Data-Driven Techniques with Microsoft Excel*. Indianápolis: John Wiley&Sons.