

Análisis bivariado: población y marginación en México

Emmanuel Olivera Pérez

Centro Interdisciplinario de Posgrados, Investigación y Consultoría.
Universidad Popular Autónoma del Estado de Puebla (UPAEP). Correo electrónico:
emmanuel.olivera@upaep.mx

Juan Carlos Bocarando Lara

Centro Interdisciplinario de Posgrados, Investigación y Consultoría.
Universidad Popular Autónoma del Estado de Puebla (UPAEP). Correo electrónico:
juancarlos.bocarando@upaep.edu.mx

María Catalina Ovando Chico

Centro Interdisciplinario de Posgrados, Investigación y Consultoría.
Universidad Popular Autónoma del Estado de Puebla (UPAEP). Correo electrónico:
mariacatalina.ovando@upaep.mx

Resumen

El artículo tiene como objetivo mostrar la aplicación de técnicas descriptivas e inferenciales a datos económicos indispensables para cualquier análisis de problemas sociales, económicos o políticos. Las variables de interés son *población total*, *índice de marginación* y *grado de marginación*, extraídas de proyectos estadísticos del Instituto Nacional de Estadística y Geografía (INEGI) así como del Consejo Nacional de Población (CONAPO). El universo de datos sujetos de análisis se integró de 1,382 municipios de un total de 2,457 que existen en México; las entidades federativas que conforman el análisis son aquellas que colidan con el estado de Puebla, mismas que en su conjunto representan aproximadamente el 56% del total de municipios del país. La elección de la muestra fue discrecional, ya que el interés es mostrar la utilidad de las técnicas, sin dejar de ofrecer una explicación coherente y confiable de los resultados. En este sentido, entre los hallazgos principales se encuentra que el Teorema de Chebyshev contribuye a explicar la variable *población total* mientras que la Regla Empírica la variable *índice de marginación*. Con base en el cociente de las frecuencias relativas de la variable *grado de marginación*, se espera que la probabilidad de ocurrencia de eventos que involucren a los estados que se encuentran al sur de Puebla sea de aproximadamente 47.10%, la mayor respecto a las probabilidades de cada una de las otras dos regiones consideradas.

Palabras clave: Grado de marginación; Índice de marginación; Teorema de Chebyshev; Regla Empírica.

Abstract

The article aims to show the application of descriptive and inferential techniques to economic data indispensable for any analysis of social, economic or political problems. The variables of interest are total population, index of marginalization and degree of marginalization, taken from statistical projects of Instituto Nacional de Estadística y Geografía (INEGI for its acronym in Spanish) and Consejo Nacional de Población (CONAPO for its acronym in Spanish). The universe of data subject to analysis was composed of 1,382 municipalities out of a total of 2,457 that exist in Mexico. The states included in the analysis are those that collide with the state of Puebla, which together represents approximately 56% of the total municipalities in the country. The choice of the sample was discretionary since the interest is to show the usefulness of the techniques, while still offering a coherent and reliable explanation of the results. In this sense, among the main findings is that the Chebyshev Theorem explains the total population variable and the Empirical Rule the marginal index variable. Based on the quotient of the relative frequencies of the variable degree of marginalization, it is expected that the probability of occurrence of events involving the states located south of Puebla is approximately 47.10%, the highest concerning the probabilities of each of the other two regions considered.

Keywords: Level of Marginalization; Marginalization Index; Chebyshev Theorem; Rule of Thumb.

1. Introducción

La estadística descriptiva e inferencial, representa un conjunto de métodos que nos ayudan a resumir medidas e inferir ciertos resultados a partir de muestras. En este sentido la utilidad de aprender su aplicación es relevante para explicar cualquier fenómeno económico o social, incluso de entidades como empresas.

Para efectos demostrativos, el artículo tiene como objetivo mostrar la aplicación de técnicas descriptivas e inferenciales a datos económicos indispensables para cualquier análisis de problemas sociales, económicos o políticos. El documento, expone de manera detallada algunos de hallazgos que se generaron como resultado de la aplicación de técnicas descriptivas e inferenciales a una base de datos que contiene información de tres variables, *población total*, *índice de marginación* y *grado de marginación*.

Los datos, corresponden a los municipios que conforman los estados que colindan con el estado de Puebla, mismos que de acuerdo al INEGI (2016), son los siguientes: al poniente Hidalgo, México, Tlaxcala y Morelos, al este Veracruz y al sur Oaxaca y Guerrero.

El documento se encuentra estructurado en cuatro apartados. En el primero, se describe de manera general las variables empleadas en el estudio. En el segundo, se muestran los resultados de la aplicación de técnicas de estadística descriptiva al conjunto de datos. En el tercero, se expone el resultado de emplear técnicas inferenciales a las unidades sujetas de análisis, a partir de la construcción de una tabla de probabilidades. Finalmente, en el cuarto se exponen de manera general las principales conclusiones del estudio.

2. Contexto

El estado de Puebla representa 1.75% de la superficie del país [2]. Al año 2015 estaba dividido en 217 municipios [3]. De acuerdo a cifras del Censo de Población y Vivienda, 2010, tenía una *población total* de 5'779,829 personas [4].

Con relación a los estados que se encuentran al poniente del estado de Puebla, se puede señalar que Hidalgo representa 1.06% de la superficie del país [5]. Al año 2015 estaba dividido en 84 municipios [6]. Según el Censo de Población y Vivienda, 2010, tenía una *población total* de 2'665,018 personas [7]. Mientras tanto, el Estado de México -México- representa el 0.08% de la superficie del país [8], y la *población total* está desagregada en 125 municipios [9], misma que asciende a 15'175,862 personas [10]. Por otra parte, el estado de Tlaxcala representa 0.2% de la superficie del país [11]. Al 2015 se encontraba dividido en 60 municipios [12], y tenía una *población total* de 1'169,936 personas según el Censo de Población y Vivienda, 2010 [13]. El estado de Morelos, representa solamente el 0.25% de la superficie del país [14]. Se encontraba dividido en 33 municipios al 2015 [15], y tenía una *población total* de 1'777,227 personas [16], según el Censo de Población y Vivienda, 2010.

En lo que atañe al estado de Veracruz, ubicado al este del estado de Puebla, se puede mencionar que representa el 3.66% de la superficie del país [17]. Al 2015 se encontraba dividido en 212 municipios [18], y tenía una *población total* según el Censo de Población y Vivienda, 2010 de 7'643,194 personas [19].

Finalmente con relación a los estados que se encuentran ubicados al sur del estado de Puebla se puede indicar que Oaxaca, representa 4.78% de la superficie del país [20]. Se encontraba dividido al 2015 en 570 municipios [21], y tenía una *población total* según el Censo de Población y Vivienda, 2010 de 3'801,962 personas [22]. Mientras tanto, el estado de Guerrero representa el 3.24% de la superficie del país. Al año 2015 estaba dividido en 81 municipios [23], y tenía una *población total* de 3'388,768 personas según el Censo de Población y Vivienda, 2010 [24].

En lo que respecta al *índice de marginación*, se puede indicar que es una medida-resumen que permite diferenciar el impacto de las carencias que padece la población, debido a la falta de acceso a la educación, la residencia en viviendas inadecuadas y la carencia de bienes [25].

Los valores inferiores del índice son reflejo de una menor marginación y los superiores de una mayor presencia de marginación.

Existen cinco *grados de marginación*: muy alto, alto, medio, bajo y muy bajo [26]. Debido a su naturaleza, los *grados de marginación* coadyuvan a mostrar las disparidades territoriales y dan cuenta, entre otros aspectos, de las relaciones existentes con el nivel de marginación de las entidades federativas, por lo que se constituye como un insumo directo para los tomadores de decisiones en materia de política social. Finalmente, con relación al indicador referido por estado se puede agregar que de acuerdo a cifras del Consejo Nacional de Población [27], en Guerrero y Oaxaca es muy alto, mientras que en Veracruz, Puebla e Hidalgo es alto, por otra parte en Morelos y Tlaxcala es medio, mientras que en México es bajo.

3. Análisis descriptivo

Para efectos de análisis, en este apartado se tomó en consideración solamente dos variables. La primera se refiere a *población total*, mientras que la segunda al *índice de marginación* del cúmulo de entidades federativas que colindan con el estado de Puebla, es decir, Hidalgo, México, Tlaxcala, Morelos, Veracruz, Oaxaca y Guerrero, ya que es de interés conocer su naturaleza.

La Tabla 1, muestra dos de las principales medidas de tendencia central, la media y la mediana. Es preciso aclarar que en ambos casos se partió de datos del año 2010 y de los resultados que posee cada uno de los municipios que conforman el estado de la republica que fue sujeto de análisis.

Tabla 1. Medidas de tendencia central

Ubicación /Estados (s)	Población total		Índice de marginación	
	Media	Mediana	Media	Mediana
▪Puebla	28,428.03	10,685.00	-0.02	-0.11
<i>Poniente</i>				
▪Hidalgo	34,028.08	19,187.50	0.26	0.18
▪México	129,500.86	41,333.00	0.03	-0.03
▪Tlaxcala	21,214.12	13,521.50	-0.01	0.06
▪Morelos	57,691.24	27,187.00	0.20	0.18
<i>Este</i>				
▪Veracruz	38,266.53	19,948.00	0.07	0.02
<i>Sur</i>				
▪Oaxaca	6,961.21	2,704.50	-0.08	-0.25
▪Guerrero	43,620.38	21,995.00	0.16	0.16

Fuente: elaboración propia.

Como se puede apreciar en la Tabla 1, los cocientes de la media y la mediana en la variable *población total* no se aproximan, lo que podría indicar que los datos no tienen una distribución normal. El estado de Oaxaca, con un total de 570 municipios, tiene una media de 6,961.21 y una mediana de 2,704.50 aproximadamente, es decir, esta última es menor en 2.5 veces al promedio.

Por otro lado, también se puede observar que el estado de Tlaxcala, el cual posee 60 municipios, es donde se encuentra la menor variación entre ambas medidas de tendencia central. Sin embargo, en general ninguna medida de las variables *población total* o *índice de marginación* muestran evidencia de que sean distribuciones normales, ya que el sesgo es muy amplio en todos los casos -Sesgo=Media-Mediana- [28]–[32]. Quizás, el único caso donde el sesgo es más pequeño es Hidalgo. Por lo señalado, se puede concluir que la medida de tendencia central que podría explicar de mejor forma la variable *población total* es la mediana. En el caso

del *índice de marginación* los sesgos no son amplios, de tal forma que se puede concluir que ambas medidas son buenas referencias. A continuación se muestran los histogramas de los estados de Oaxaca y Tlaxcala con la finalidad de sustentar las afirmaciones realizadas (Figuras 1 y 2).

Figura 1. Distribución de los datos de la variable *población total* para el estado de Oaxaca.

Figura 2. Distribución de los datos de la variable *población total* para el estado de Tlaxcala.

En todos los casos tratándose de la variable *población total*, como se puede observar en la Tabla 1, la media es mayor a la mediana por lo que se puede asumir que las entidades federativas sujetas de análisis tienen sesgos positivos. Asimismo, se puede apreciar en las Figuras 1 y 2, correspondientes a los estados de Oaxaca y Tlaxcala, la presencia de un sesgo a la derecha. Para confirmar lo señalado, se toman los datos del estado de Puebla que se integra por 217 municipios. En promedio, la variable *población total* es de 28,428.03 mientras que la mediana es de 10,685.00 lo cual indica que la distribución también tiene un sesgo positivo, ya que la media es mayor a la mediana como se puede apreciar en la Figura 3.

Figura 3. Distribución de los datos de la variable *población total* para el estado de Puebla.

Con base en los resultados obtenidos en las entidades de Puebla, Tlaxcala y Oaxaca, se procedió a determinar la consistencia del sesgo. La Tabla 2 muestra un breve resumen.

Tabla 2. Consistencia del sesgo

Ubicación /Estados (s)	Población total					
	Media (A)	Valor máximo (B)	Valor absoluto I A – B I	Valor mínimo (C)	Mediana (D)	Valor absoluto I C – D I
▪Puebla	28,428.03	1,576,259	1,547,830.97	469	10,685.00	10,216
<i>Poniente</i>						
▪Tlaxcala	21,214.12	95,051	73,836.88	3,103	13,521.50	10,418.5
<i>Sur</i>						
▪Oaxaca	6,961.21	264,251	257,289.79	87	2,704.50	2,617.50

Fuente: elaboración propia.

Si se comparan las diferencias absolutas entre las medias y los valores máximos (A - B) con la diferencia absoluta entre los valores mínimos y las medianas (C - D) de cada uno de los casos, el resultado siempre es el mismo. La diferencia de A - B es mayor que la diferencia entre C -D, por lo tanto hay evidencia de que los sesgos son consistentes.

En el caso del estado de Puebla, el resultado de la media menos el valor máximo es de 1'547,830.97 el cual es mayor al cociente del valor mínimo menos la mediana que es de 10,216 por lo que se confirma la consistencia del sesgo. Es importante señalar que debido a la naturaleza de la variable que se está empleando, no existe moda, es decir, no existen dos o más municipios con la misma cantidad de *población total*.

Por otra parte, tratándose de la variable *índice de marginación*, se puede observar en la Tabla 1 que los cocientes de la media y la mediana son cercanos, lo cual podría indicar que se aproximan a una distribución normal. Para confirmar lo anterior, se construyeron los histogramas correspondientes a los estados de Veracruz e Hidalgo (Figuras 4 y 5).

Figura 4. Distribución de los datos de la variable *índice de marginación* para el estado de Veracruz.

Figura 5. Distribución de los datos de la variable *índice de marginación* para el estado de Hidalgo.

Con base en los resultados obtenidos en las Figuras 4 y 5, se puede observar que en ambos casos se aproximan a distribuciones normales. Como sucede con la variable *población total* la naturaleza de la variable *índice de marginación* no arroja una moda, es decir, no existen dos o más municipios con el mismo índice. Una vez que se conoce la naturaleza de la distribución de las variables referidas, a continuación se analiza lo relativo a las medidas de dispersión para medir la distancia más corta entre una observación y cualquier otra de la variable *índice de marginación*. Para tal efecto, a manera de ejemplo, se tomarán los datos correspondientes a dos estados, Hidalgo y Puebla (Tabla 3 y 4).

Tabla 3. Estadísticos del estado de Hidalgo de la variable *índice de marginación*

N	Válido	84
	Perdidos	0
Desviación estándar		.927463548120783
Varianza		.860
Rango		4.3067114493769300
Mínimo		-1.7195502845246537
Máximo		2.5871611648522768
Percentiles	25	-.440191196851135
	50	.183971118750328
	75	.907500814152321

Fuente: elaboración propia.

Tabla 4. Estadísticos del estado de Puebla de la variable *índice de marginación*

N	Válido	217
	Perdidos	3
Desviación estándar		.994135159329003
Varianza		.988
Rango		5.7726267739599670
Mínimo		-1.9777045197436710
Máximo		3.7949222542162966
Percentiles	25	-.804588062926805
	50	-.112022605511180
	75	.599315691179271

Fuente: elaboración propia.

Como se puede apreciar en las Tablas 3 y 4, la desviación estándar en ambos casos es pequeña, lo que indica que los datos de la variable *índice de marginación* no se encuentran tan dispersos como sucedía con la variable *población total*.

La marginación de cada uno de los estados referidos tiene un comportamiento más o menos homogéneo. A continuación se muestran los diagramas de caja correspondientes para poder observar de mejor forma lo señalado y confirmar el supuesto referido (Figura 6 y 7).

Figura 6. Diagrama de caja de la variable *índice de marginación* para el estado de Hidalgo.

Figura 7. Diagrama de caja de la variable *índice de marginación* para el estado de Puebla.

En el caso del estado de Hidalgo, se puede apreciar que no existen datos atípicos, sin embargo, en el caso del estado de Puebla, en el Figura 7 se puede observar un dato que probablemente es atípico, el cual corresponde al municipio de Hueyapan, mismo que tiene un muy alto *índice de marginación* (3.79). Por lo tanto, debido a que esta observación es verdadera pero sesga la distribución, una solución en este caso lo mejor sería tomar el valor de la mediana para describir la variable *índice de marginación*.

Para confirmar lo señalado, se utiliza el puntaje Z para determinar si verdaderamente se trata de un dato atípico. Sustituyendo en la expresión correspondiente se obtiene lo siguiente:

$$\text{Puntaje } z = \frac{3.79 - (-0.02)}{0.9941}$$

$$\text{Puntaje } z = 3.83$$

Con base en el resultado, se confirma que el valor de 3.79 correspondiente al municipio de Hueyapan es atípico para el caso de una distribución normal (ya que es mayor a 2), mientras que para una distribución sesgada está en el límite superior (mayor a 4 se considera dato atípico), ya que es mayor a tres desviaciones estándar [33].

A partir de lo que se ha presentado, se espera que sea la Regla Empírica aquella que explique de mejor forma la distribución de la variable *índice de marginación* ya que como se observó en las Figuras 4 y 5 los datos “se apilan” en la conocida forma de montículo.

Mientras que el Teorema de Chebyshev sería aplicable a la variable *población total*, debido a que se puede aplicar a cualquier conjunto de mediciones, muestra o población, grande o pequeño, en forma de montículo o sesgada [34].

Tabla 5. Demostración de la aplicación de la regla empírica para describir la distribución de la variable *índice de marginación* para el estado de Puebla

K	$\bar{x} \pm ks$	Frecuencia relativa	Teorema de Chebyshev	Regla Empírica
1	$[-0.02 + 1 (0.99)] = 0.97$	147/217= 0.67	Al menos 0	Aprox. 0.68
	$[-0.02 - 1 (0.99)] = -1.01$			
2	$[-0.02 + 2 (0.99)] = 1.96$	214/217= 0.98	Al menos 0.75	Aprox. 0.95
	$[-0.02 - 2 (0.99)] = -2.00$			
3	$[-0.02 + 3 (0.99)] = 2.95$	216/217= 0.99	Al menos 0.89	Aprox. 0.997
	$[-0.02 - 3 (0.99)] = -2.99$			

Fuente: elaboración propia.

Como se puede observar en la Tabla 5, se confirma que la Regla Empírica describe de mejor forma la distribución de la variable *índice de marginación*. A una desviación estándar ambas teorías son aplicables, sin embargo, a dos y tres desviaciones estándar, ya se puede descartar el uso del Teorema de Chebyshev, debido a que las frecuencias relativas determinadas están más cercanas a lo que establece la Regla Empírica.

Antes de continuar, se considera adecuado demostrar que el Teorema de Chebyshev es aquel que mejor describe la distribución de la variable *población total*. Para tal efecto, se toman los datos correspondientes al estado de Morelos.

Tabla 6. Demostración de la aplicación de la regla empírica para describir la distribución de la variable *población total* para el estado de Morelos

K	$\bar{x} \pm ks$	Frecuencia relativa	Teorema de Chebyshev	Regla Empírica
1	$[57,691.24 + 1 (74,645.05)] = 132,336.29$	30/33= 0.91	Al menos 0	Aprox. 0.68
	$[57,691.24 - 1 (74,645.05)] = -16,953.81$			
2	$[57,691.24 + 2 (74,645.05)] = 206,981.34$	31/33= 0.94	Al menos 0.75	Aprox. 0.95
	$[57,691.24 - (74,645.05)] = -91,598.86$			
3	$[57,691.24 + (74,645.05)] = 281,626.39$	32/33= 0.97	Al menos 0.89	Aprox. 0.997
	$[57,691.24 - (74,645.05)] = -166,243.91$			

Fuente: elaboración propia.

Como se puede apreciar en la Tabla 6, a una desviación estándar ya se encuentran el 91% de los casos, con lo cual queda descartada la Regla Empírica. A pesar de que a 2 desviaciones estándar el cociente se encuentra muy cercano a lo que ésta establece, ya que sólo varía en 1% (Figura 8).

Figura 8. Gráfica de puntos de la variable *población total* para el estado de Morelos.

Se puede observar en la Figura 8, tal y como se esperaba, la variable *población total* se encuentra sesgada y, en el caso del estado de Morelos, dicho sesgo es a la derecha, por lo cual el Teorema de Chebyshev es aquel que explica la distribución.

A continuación se muestran en la Tabla 7 las covarianzas así como los coeficientes de correlación de las variables denominadas *población total* (X) e *índice de marginación* (Y), para los estados de Oaxaca y Puebla.

Tabla 7. Covarianzas para los estados de Oaxaca y Puebla

Oaxaca=570	Población total (X)	Índice de marginación (Y)	$\sum (X - \bar{x}) * (Y - \bar{y})$
Promedio	6,961.21	-0.08	-93,984.94
Desviación estándar	16,832.03	1.14	
Covarianza	-165.175647		
Coefficiente de correlación	-0.00863752		
Puebla=217	Población total (X)	Índice de marginación (Y)	$\sum (X - \bar{x}) * (Y - \bar{y})$
Promedio	28,428.03	-0.02	373,685.086
Desviación estándar	110,707.70	0.99	1
Covarianza	1,730.023547		
Coefficiente de correlación	0.015719139		

Fuente: elaboración propia.

En el caso del estado de Oaxaca, la covarianza es menor a cero lo cual indica que existe una relación negativa entre las variables X e Y [35], por lo cual se esperaría que los datos se encuentren en los cuadrantes II y IV -plano euclidiano-.

Por otro lado, ya que el coeficiente de correlación es negativo y no se aproxima a 1, se puede concluir que existe una relación lineal de 0.008 negativa entre las variables, es decir, al existir un cambio de 1 en la variable X la variable Y cambia en 0.008 de forma inversa, un cambio cercano a cero. Lo mismo ocurre cuando existen cambios en variable Y, es decir, la variable X cambia en 0.008 de forma inversa (Figura 9).

Figura 9. Diagrama de dispersión para el estado de Oaxaca.

En la Figura 9 no se puede apreciar claramente cuál es el cuadrante en donde se encuentra la mayor concentración de datos. Sin embargo, sí se toman en consideración los resultados de las medias del estado de Oaxaca determinados en la Tabla 7, es decir, *población total* (X) de 6,961.21 e *índice de marginación* (Y) de -0.08 y se ubican dentro de la Figura 9, se podrá observar que los datos se encuentran agrupados en el cuadrante IV, con lo cual se confirma lo señalado.

Por otro lado, tratándose del estado de Puebla, como se puede apreciar, la covarianza es mayor a cero, lo cual indica que existe una relación positiva entre las variables X e Y [35], además se esperaría que los datos se encuentren en los cuadrantes I y III. Por otra parte, teniendo en cuenta que el coeficiente de correlación es positivo y se aproxima a 1, se puede decir que existe una relación lineal de 0.01 positiva entre las variables, es decir, al existir un cambio de 1 en la variable X, la variable Y cambia en 0.01 en el mismo sentido. Lo mismo ocurre cuando existen cambios en variable Y, es decir, la variable X se incrementa en 0.01 aproximadamente (Figura 10).

Figura 10. Diagrama de dispersión para el estado de Puebla.

Por otra parte, tal y como ocurrió en el estado de Oaxaca, tratándose del estado de Puebla, no se puede observar claramente en la Figura 10 hacia qué cuadrante se encuentra la mayor concentración de datos. Sin embargo, sí se toman en consideración los resultados de las medias del estado de Puebla determinados en la Tabla 7, es decir, *población total* (X) de 28,428.03 e *índice de marginación* (Y) de -0.02 y se ubican dentro de la Figura 10, se podrá observar que los datos se encuentran agrupados en el cuadrante I, con lo cual se confirma lo señalado.

4. Análisis inferencial

Con base en los datos recuperados sobre los *grados de marginación* perteneciente a los estados de la república que colindan con Puebla, tales como Veracruz, Hidalgo, México, Tlaxcala, Morelos, Oaxaca y Guerrero, se construyó la Tabla 8.

Tabla 8. Distribución de los datos atendiendo la variable *grados de marginación*

Ubicación/ Estados (s)	Muy bajo	Bajo	Medio	Alto	Muy alto	TOTALES
▪Puebla	25	36	78	38	40	217
<i>Poniente</i>						
▪Hidalgo	3	11	35	16	19	84
▪México	11	21	52	21	20	125
▪Tlaxcala	5	11	28	6	10	60
▪Morelos	0	3	17	8	5	33
<i>Este</i>						
▪Veracruz	21	28	83	38	42	212
<i>Sur</i>						
▪Oaxaca	81	119	186	77	107	570
▪Guerrero	7	6	35	17	16	81
TOTALES	153	235	514	221	259	1,382

Fuente: elaboración propia.

A partir de los datos contenidos en la Tabla 8, se elaboró la Tabla 9 para asignar probabilidades a los eventos simples de los 1,382 municipios sujetos de estudio agrupados por estado de la república.

Tabla 9. Frecuencias de la variable *grado de marginación* por estado de la república

Ubicación/ Estados (s)	Muy bajo J	Bajo K	Medio L	Alto M	Muy alto N	TOTALES
▪Puebla A	1.81%	2.60%	5.64%	2.75%	2.89%	15.70%
<i>Poniente</i>						
▪Hidalgo B	0.22%	0.80%	2.53%	1.16%	1.37%	6.08%
▪México C	0.80%	1.52%	3.76%	1.52%	1.45%	9.04%
▪Tlaxcala D	0.36%	0.80%	2.03%	0.43%	0.72%	4.34%
▪Morelos E	0.00%	0.22%	1.23%	0.58%	0.36%	2.39%
<i>Este</i>						
▪Veracruz F	1.52%	2.03%	6.01%	2.75%	3.04%	15.34%
<i>Sur</i>						
▪Oaxaca G	5.86%	8.61%	13.46%	5.57%	7.74%	41.24%
▪Guerrero H	0.51%	0.43%	2.53%	1.23%	1.16%	5.86%
▪TOTALES	11.07%	17.00%	37.19%	15.99%	18.74%	100.00%

Fuente: elaboración propia.

De acuerdo a la información contenida en la Tabla 9, a continuación se describe una lista de probabilidades que son de interés con la finalidad de mostrar la aplicabilidad de la estadística inferencial.

▪ Las probabilidades de que ocurran los eventos B, C, D, E, F, G y H de forma individual son las siguientes:

- P (B)= 6.08%
- P (C)= 9.04%
- P (D)= 4.34%
- P (E)= 2.39%
- P (F)= 15.34%
- P (G)= 41.24%
- P (H)= 5.86%

Con base en lo anterior, se puede señalar que las probabilidades de que sea seleccionado Hidalgo es de 6.08%, México 9.04%, Tlaxcala 4.34%, Morelos 2.39%, Veracruz 15.34%, Oaxaca 41.24% y Guerrero 5.86%.

▪ La probabilidad de que sean seleccionados los estados que se encuentran ubicados al este o sur o poniente o todos del estado de Puebla representando por P (K).

- $P (K) = 15.34\% + 41.24\% + 5.86\% + 6.08\% + 9.04\% + 4.34\% + 2.39\% = 84.29\%$

▪ La probabilidad de que sean seleccionados los estados que se encuentran ubicados al este o sur o ambos del estado de Puebla representando por P (L).

- $P (L) = 15.34\% + 41.24\% + 5.86\% = 62.44\%$

Por lo tanto, la probabilidad de que sea seleccionados los estados de Veracruz, Oaxaca, Guerrero o los tres es de 62.44%.

- La probabilidad de que ocurra la selección de los estados de Morelos o Veracruz o ambos representando por P (M).
 - $P(M) = 2.39\% + 15.34\% = 17.73\%$
- La probabilidad de que no sean seleccionados los estados que se encuentran ubicados al este, sur y poniente del estado de Puebla.
 - $P(K^C) = 100\% - 84.29\% = 15.71\%$

Con base en lo anterior, se puede mencionar que la probabilidad de que sea seleccionado el estado de Puebla es de 15.71%.

- La probabilidad de que ocurra A y K.
 - $P(A \cap K) = 15.70\% + 17.0\% - 2.60\% = 30.10\%$
- La probabilidad de que ocurra E y N.
 - $P(E \cap N) = 2.39\% + 18.79\% - 0.36\% = 21.33\%$
- La probabilidad de que ocurra L y G.
 - $P(L \cap G) = 37.19\% + 41.24\% - 13.46\% = 64.97\%$
- La probabilidad de que ocurra M y B.
 - $P(M \cap B) = 15.99\% + 6.08\% - 1.16\% = 20.91\%$
- Se sabe que ya fue seleccionado el *grado de marginación* muy bajo (J), cuál es la probabilidad de que sea seleccionado el estado de Puebla (A), así como los estados de Hidalgo (B), México (C), Tlaxcala (D), Veracruz (F), Oaxaca (G) y Guerrero (H).
 - $P(A | J) = 1.81\% / 11.07\% = 16.33\%$
 - $P(B | J) = 0.22\% / 11.07\% = 1.98\%$
 - $P(C | J) = 0.80\% / 11.07\% = 7.21\%$
 - $P(D | J) = 0.36\% / 11.07\% = 3.24\%$
 - $P(F | J) = 1.52\% / 11.07\% = 13.72\%$
 - $P(G | J) = 5.86\% / 11.07\% = 52.92\%$
 - $P(H | J) = 0.51\% / 11.07\% = 4.60\%$

Como era de esperarse, sí se suman todas las probabilidades determinadas el cociente es igual a 100%. Nótese que se excluyó el caso del estado de Morelos (E) ya que no tiene municipios con muy bajo *grado de marginación*.

- Se sabe que ya fue seleccionado el *grado de marginación* medio (L), cuál es la probabilidad de que sea seleccionado el estado de Puebla (A) y los estados de Hidalgo (B) y México (C).
 - $P(A | L) = 1.81\% / 37.19\% = 3.17\%$
 - $P(B | L) = 0.22\% / 37.19\% = 0.59\%$
 - $P(C | L) = 0.80\% / 37.19\% = 2.15\%$
- Se sabe que ya fue seleccionado el estado de Morelos (E) ya fue seleccionado, cuál es la probabilidad de que ocurran los eventos K, L, M y N.
 - $P(K | E) = 0.22\% / 2.39\% = 9.20\%$

- $P(LIE) = 1.23\%/2.39\% = 51.46\%$
- $P(MIE) = 0.58\%/2.39\% = 24.28\%$
- $P(NIE) = 0.36\%/2.39\% = 15.06\%$

De la misma forma, como era de esperarse, si se suman todas las probabilidades, el cociente es igual a 100%. Nótese que se excluyó el caso muy bajo (J) *grado de marginación* ya que el estado de Morelos (E) no posee municipios con esa característica.

Sí se realiza un experimento con base en la información contenida en la Tabla 8 con los estados que se encuentran al sur del estado de Puebla, se obtiene el resultado que se muestra la Tabla 10.

Tabla 10. Experimento con los estados de Oaxaca y Guerrero

Ubicación/ Estados (s)	Muy bajo J	Bajo K	Medio L	Alto M	Muy alto N
<i>Sur</i>					
▪Oaxaca G	G,J	G,K	G,L	G,M	G,N
▪Guerrero H	H,J	H,K	H,L	H,M	H,N

Fuente: elaboración propia.

Sea “x” el *grado de marginación* “bajo (K)” observado (sin reemplazo).

Tabla 11. Determinación del valor esperado

Resultado	Ocurrencia	i	X	P (x)
G,J	0	1	0	8/10= 80%
G,K	1	2	1	2/10= 20%
G,L	0			
G,M	0			
G,N	0			
H,J	0			
H,K	1			
H,L	0			
H,M	0			
H,N	0			

Fuente: elaboración propia.

Por lo tanto sustituyendo en la expresión de valor esperado se obtiene lo siguiente:

$$E(x) = (0.80)(0) + (0.20)(1) = 20\%$$

Con base en el resultado obtenido, se espera que el *grado de marginación* bajo (K) ocurra 20% de las veces.

Finalmente, a continuación se abordarán algunos aspectos que se consideran de importancia sobre la distribución normal de probabilidad. Se parte de la premisa que la variable *índice de marginación* es de naturaleza continua, debido a que puede tomar valores infinitos.

Además se puede agregar, que las distribuciones de probabilidad continua pueden tomar varias formas, pero un gran número de variables aleatorias observadas en la naturaleza poseen una distribución de frecuencia que tiene más o menos la forma de montículo (Mendenhall, Beaver y Beaver, 2010). La Figura 11, muestra a través de un histograma la distribución del total de las unidades de análisis, es decir, 1,382 municipios de la variable referida.

Figura 11. Histograma de la variable *índice de marginación*.

Con base en lo anterior, se procede a determinar tres probabilidades que son de interés conocer, mismas que se enlistan a continuación:

- $P(-3 \leq X \leq 3)$
- $P(-2 \leq X \leq 1)$
- $P(1 \leq X \leq 2)$

Se debe recordar que para encontrar la probabilidad de que una variable aleatoria normal “X” se encuentre en el intervalo de “a” hasta “b”, se necesita encontrar el área bajo la curva normal entre los puntos a y b. En el primer caso, es decir, $P(-3 \leq X \leq 3)$, la probabilidad es la siguiente:

$$z \leq \frac{-3 - 0.006}{1.04} = -2.89$$

$$z \geq \frac{3 - 0.006}{1.04} = 2.88$$

Con base en el valor de Z, se emplearon las tablas “áreas bajo la curva normal” (Mendenhall, Beaver y Beaver, 2010) para conocer las probabilidades de cada uno de los casos, encontrando lo siguiente:

- $P(-2.89 \leq Z) = 0.0019$
- $P(Z \geq 2.88) = 0.9980$

Por lo tanto, la diferencia entre 0.9980 y 0.0019 es el área que se está buscando y equivale a 0.9961 aproximadamente. La Figura 12 muestra a grandes rasgos el resultado obtenido.

Figura 12. Distribución de $P(-3 \leq X \leq 3)$.

En el segundo caso, es decir, $P(-2 \leq X \leq 1)$, la probabilidad es la siguiente:

$$z \leq \frac{-2 - 0.006}{1.04} = -1.93$$

$$z \geq \frac{1 - 0.006}{1.04} = 0.96$$

Con base en el valor de Z se emplearon las tablas “áreas bajo la curva normal” (Mendenhall, Beaver y Beaver, 2010) para conocer las probabilidades de cada uno de los casos, encontrando lo siguiente:

- $P(-1.93 \leq Z) = 0.0268$
- $P(Z \geq 0.96) = 0.8315$

Por lo tanto, la diferencia entre 0.8315 y 0.0268 es el área que se está buscando y equivale a 0.8047 aproximadamente. La Figura 13 muestra a grandes rasgos el resultado obtenido.

Figura 13. Distribución de $P(-2 \leq X \leq 1)$.

En el tercer caso, es decir, $P(1 \leq X \leq 2)$, la probabilidad es la siguiente:

$$z \leq \frac{1 - 0.006}{1.04} = 0.96$$

$$z \geq \frac{2 - 0.006}{1.04} = 1.92$$

Con base en el valor de Z se emplearon las tablas “áreas bajo la curva normal” (Mendenhall, Beaver y Beaver, 2010) para conocer las probabilidades de cada uno de los casos, encontrando lo siguiente:

- $P(0.96 \leq Z) = 0.8315$
- $P(Z \geq 1.92) = 0.9726$

Por lo tanto, la diferencia entre 0.9726 y 0.8315 es el área que se está buscando y equivale a 0.1411 aproximadamente. La Figura 14 muestra a grandes rasgos el resultado obtenido.

Figura 14. Distribución de $P(1 \leq X \leq 2)$.

5. Conclusiones

El artículo tiene como objetivo mostrar la aplicación de técnicas descriptivas e inferenciales a datos económicos indispensables para cualquier análisis de problemas sociales, económicos o políticos.

Los resultados encontrados a partir de los métodos estadísticos aplicados (descriptivos y de probabilidad), sugieren que los estados con mayor marginación tienen una estrecha relación con los estados mayormente poblados. En cuanto a las distribuciones, se encontró que el *índice de marginación* exhibe una distribución normal, de tal forma que es más fácil hacer inferencias con dicha variable que de aquella que se refiere a *población total*, ya que tiene un alto sesgo (Teorema de Chebyshev).

El resultado de las frecuencias relativas (tablas de probabilidad), de la variable *grado de marginación*, sugieren que la probabilidad de ocurrencia de eventos que involucren a los estados que se encuentran al sur de Puebla es de aproximadamente 47.10%, la mayor respecto a las probabilidades de cada una de las otras dos regiones consideradas. Además, la tabla de probabilidades diseñada, sugiere que existe la probabilidad de 41.20% de que sea seleccionado el estado de Oaxaca y de 37.2% de que sean elegidos estados de la república que posean un *grado de marginación* medio.

En cuanto a las probabilidades condicionales, se encontró que la probabilidad de encontrar municipios del sur, dado que tienen una alta marginación, es mayor a encontrar un municipio del norte, dado que tiene un alto *grado de marginación*. Con lo cual se espera que este fenómeno se replique en los casos en los que la distancia norte-sur de los municipios sea mayor.

Finalmente, se destaca la necesidad de poseer ciertas habilidades para aplicar técnicas descriptivas e inferenciales. Sin embargo, también se subraya la necesidad de poseer habilidades de síntesis, análisis, interpretación, para determinar cómo se podrían aplicar diferentes técnicas estadísticas a un conjunto de datos, no de forma fragmentada, sino holística.

6. Referencias

- [1] INEGI, “Resumen de datos generales de Puebla. Obtenido de información por entidad,” 2016. [Online]. Available: <http://cuentame.inegi.org.mx/monografias/informacion/pue/default.aspx?tema=me&e=21>.
- [2] INEGI, “Superficie de Puebla. Obtenido de información por entidad,” 2016. [Online]. Available: <http://www.cuentame.inegi.org.mx/monografias/informacion/pue/territorio/default.aspx?tema=me&e=21>.
- [3] INEGI, “División municipal de Puebla. Obtenido de información por entidad,” 2016. [Online]. Available: http://www.cuentame.inegi.org.mx/monografias/informacion/pue/territorio/div_munici

- pal.aspx?tema=me&e=21.
- [4] INEGI, “Población total de Puebla. Obtenido de indicadores de población por área geográfica,” 2010. [Online]. Available: <http://www.beta.inegi.org.mx/app/areasgeograficas/?ag=21>.
- [5] INEGI, “Superficie de Hidalgo. Obtenido de información por entidad,” 2016. [Online]. Available: <http://www.cuentame.inegi.org.mx/monografias/informacion/hgo/territorio/default.aspx?tema=me&e=13>.
- [6] INEGI, “División municipal de Hidalgo. Obtenido de información por entidad,” 2016. [Online]. Available: http://www.cuentame.inegi.org.mx/monografias/informacion/hgo/territorio/div_municipal.aspx?tema=me&e=13.
- [7] INEGI, “Población total de Hidalgo. Obtenido de indicadores de población por área geográfica,” 2010. [Online]. Available: <http://www.beta.inegi.org.mx/app/areasgeograficas/?ag=21>.
- [8] INEGI, “Superficie de México. Obtenido de información por entidad,” 2016. [Online]. Available: <http://www.cuentame.inegi.org.mx/monografias/informacion/df/territorio/default.aspx?tema=me&e=09>.
- [9] INEGI, “División municipal del Estado de México. Obtenido de información por entidad,” 2016. [Online]. Available: <http://www.cuentame.inegi.org.mx/monografias/informacion/mex/default.aspx?tema=me&e=15>.
- [10] INEGI, “Población total de México. Obtenido de indicadores de población por área geográfica,” 2010. [Online]. Available: <http://www.beta.inegi.org.mx/app/areasgeograficas/?ag=21>.
- [11] INEGI, “Superficie de Tlaxcala. Obtenido de información por entidad,” 2016. [Online]. Available: <http://www.cuentame.inegi.org.mx/monografias/informacion/tlax/territorio/default.aspx?tema=me&e=29>.
- [12] INEGI, “División municipal de Tlaxcala. Obtenido de información por entidad,” 2016. [Online]. Available: http://www.cuentame.inegi.org.mx/monografias/informacion/tlax/territorio/div_municipal.aspx?tema=me&e=29.
- [13] INEGI, “Población total de Tlaxcala. Obtenido de indicadores de población por área geográfica,” 2010. [Online]. Available: <http://www.beta.inegi.org.mx/app/areasgeograficas/?ag=21>.
- [14] INEGI, “Superficie de Morelos. Obtenido de información por entidad,” 2016. [Online]. Available: <http://www.cuentame.inegi.org.mx/monografias/informacion/mor/territorio/default.aspx?tema=me&e=17>.
- [15] INEGI, “División municipal de Morelos. Obtenido de información por entidad,” 2016. [Online]. Available: http://www.cuentame.inegi.org.mx/monografias/informacion/mor/territorio/div_municipal.aspx?tema=me&e=17.
- [16] INEGI, “Población total de Morelos. Obtenido de indicadores de población por área geográfica,” 2010. [Online]. Available: <http://www.beta.inegi.org.mx/app/areasgeograficas/?ag=21>.
- [17] INEGI, “Superficie de Veracruz. Obtenido de información por entidad,” 2016. [Online]. Available:

- <http://www.cuentame.inegi.org.mx/monografias/informacion/ver/territorio/default.aspx?tema=me&e=30>.
- [18] INEGI, “División municipal de Veracruz. Obtenido de información por entidad,” 2016. [Online]. Available: http://www.cuentame.inegi.org.mx/monografias/informacion/ver/territorio/div_municipal.aspx?tema=me&e=30.
- [19] INEGI, “Población total de Veracruz. Obtenido de indicadores de población por área geográfica,” 2010. [Online]. Available: <http://www.beta.inegi.org.mx/app/areasgeograficas/?ag=21>.
- [20] INEGI, “Superficie de Oaxaca. Obtenido de información por entidad,” 2016. [Online]. Available: <http://www.cuentame.inegi.org.mx/monografias/informacion/oax/territorio/default.aspx?tema=me&e=20>.
- [21] INEGI, “División municipal de Oaxaca. Obtenido de información por entidad,” 2016. [Online]. Available: http://www.cuentame.inegi.org.mx/monografias/informacion/oax/territorio/div_municipal.aspx?tema=me&e=20.
- [22] INEGI, “Población total de Oaxaca. Obtenido de indicadores de población por área geográfica,” 2010. [Online]. Available: <http://www.beta.inegi.org.mx/app/areasgeograficas/?ag=21>.
- [23] INEGI, “División municipal de Guerrero. Obtenido de información por entidad,” 2016. [Online]. Available: http://www.cuentame.inegi.org.mx/monografias/informacion/gro/territorio/div_municipal.aspx?tema=me&e=12.
- [24] INEGI, “Población total de Guerrero. Obtenido de indicadores de población por área geográfica,” 2010. [Online]. Available: <http://www.beta.inegi.org.mx/app/areasgeograficas/?ag=21>.
- [25] CONAPO, “Índice de marginación por localidad 2010. Obtenido de Capítulo I,” 2010. [Online]. Available: http://www.conapo.gob.mx/work/models/CONAPO/indices_margina/2010/documento_principal/Capitulo01.pdf.
- [26] CONAPO, “Metodología de estimación del índice de marginación por localidad. Obtenido de anexo C.,” 2010. [Online]. Available: http://www.conapo.gob.mx/work/models/CONAPO/indices_margina/2010/anexoc/AnexoC.pdf.
- [27] CONAPO, “Índices de marginación por entidad federativa, 2010. Obtenido de anexo A,” 2010. [Online]. Available: http://www.conapo.gob.mx/en/CONAPO/Indices_de_Marginacion_2010_por_entidad_federativa_y_municipio.
- [28] S. Medina-peralta *et al.*, “Validation of models with constant bias : an applied approach Validación de modelos con sesgo constante : un enfoque aplicado,” vol. 19, no. 2, pp. 4099–4108, 2014.
- [29] F. Javier and M. Rueda, “Sesgo de deseabilidad social en medidas de valores organizacionales * Social desirability bias in measures of organizational values,” no. 2, 2017.
- [30] G. Bias, O. Surveys, and H. Economics, “Los sesgos de género en las encuestas oficiales sobre economía doméstica,” pp. 21–38, 2014.
- [31] L. P. Garc, “La explicación del sesgo en los ítems de rendimiento,” vol. 10, pp. 481–490, 1998.
- [32] D. F. Muñoz and A. Ramírez-nafarrate, “Estimación de la Esperanza , Varianza y

- Cuantiles en Simulaciones de Estado Estable Mean , Variance and Quantile Estimation for Steady-State Simulation,” vol. 26, pp. 3–13, 2015.
- [33] M. Suero, J. Privado, and J. Botella, “Methods to estimate the variance of some indices of the signal detection theory : A simulation study,” pp. 149–175, 2017.
- [34] W. Mendenhall, R. J. Beaver, and B. M. Beaver, *Introducción a la probabilidad y la estadística*, Décima ter. 2010.
- [35] E. Carlos, P. Quispe, and R. Quispe, “Componentes de varianza y repetibilidad de características productivas y textiles de la fibra en alpacas (Vicugna pacos) Huacaya criadas a nivel comunal Variance components and repeatability of productive and textil traits of the fiber of Huacaya alpacas (Vicugna pacos) raised in a communal system,” vol. 24, no. 4, pp. 217–225, 2016.

Emmanuel Olivera Pérez es Doctor en Economía por la Universidad de las Américas, Puebla. Investigador de tiempo completo en la Universidad Popular Autónoma del Estado de Puebla (UPAEP).

Juan Carlos Bocarando Lara es Maestro en Ciencias Administrativas por la Universidad Veracruzana. Estudiante de doctorado en la Universidad Popular Autónoma del Estado de Puebla (UPAEP).

María Catalina Ovando Chico es Doctora en Telecomunicaciones por la Universidad Politécnica de Madrid. Investigadora de tiempo completo en la Universidad Popular Autónoma del Estado de Puebla (UPAEP).