

# Ciencia de Datos

---

## Tarea 2

### Ejercicio 2



Miranda Belmonte Hairo Ulises  
hairo.miranda@cimat.mx

**Institución:** Centro de Investigación en Matemáticas Unidad Monterrey

**Dirección:** Alianza Centro 502, 66629 N.L

**Programa:** Maestría en Computo Estadístico

**Fecha:** 15 de Febrero del 2019

**Clase:** Computo Estadístico

# Contenido

1	Antecedentes . . . . .	4
2	Descripción de la base . . . . .	5
3	Indicadores Marginación . . . . .	7
	Dimensiones de la marginación en localidades . . . . .	10
4	Resultados . . . . .	12
5	Determinación de los grados de marginación . . . . .	14
6	Contaste respecto al índice de marginación oficial . . . . .	17
7	Recomendaciones . . . . .	19
8	Conclusión . . . . .	20
	Anexo . . . . .	20
	Diccionario de datos . . . . .	21
	Cálculo de los indicadores . . . . .	22
	Construcción del índice . . . . .	23
	Construcción del grado de marginación . . . . .	23

## Ciencia de Datos

---

### Tarea 2

### Ejercicio 2

# Índices de Marginación de las localidades en Nuevo León, México, para el año 2010

Miranda Belmonte Hairo Ulises

## Resumen

Se calcula el índice de marginación para 2,037 localidades en Nuevo León con base al Censo de Población y Vivienda de 2010 que realiza el Instituto Nacional de Estadística y Geografía (INEGI). La estimación se realiza calculando ocho indicadores socioeconómicos que componen al índice de marginación, con ellos se realiza la técnica de componentes principales y se obtiene el grado de marginación. Entre los resultados principales se tiene que 517 localidades tienen un grado de marginación muy bajo, 672 bajos, 437 medio, 291 alto y sólo 120 muy alto.

# 1 Antecedentes

De acuerdo con cifras del Instituto Nacional de Estadística y Geografía (INEGI), para el año 2017, Nuevo León es el tercer estado con mayor participación en el producto interno bruto (PIB), aportando el 7.3% (1.5 billones de pesos) a la producción nacional, y cuya producción per cápita registró una variación anual del 3%.

Gran parte de la producción estatal se contribuye al dinamismo de su gran capacidad productiva en la industria manufacturera (actividades terciarias). Según cifras de INEGI, para el año 2016, el 61.9% de su PIB se debía a este sector, el 37.5% actividades secundarias y un .6%, a actividades primarias.

El estado de Nuevo León cuenta con 135,482 unidades económicas, el 3.2%, del país. Empleando a un total de 1,399,230 de personas, que equivale al 6.5% del personal ocupado de México. No esta por más indicar, que dado a la dinámica en la economía de Nuevo León, oportunidades de desarrollo en términos sociales y económicas se presentan viables para años futuros.

Figure 1: Estado de Nuevo León

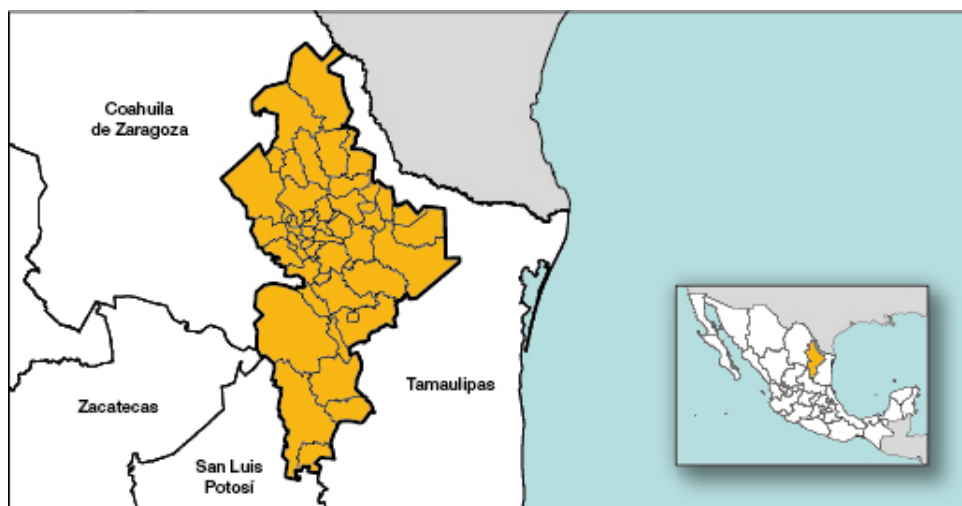


Imagen del Instituto Nacional de Estadística y Geografía (INEGI)

## 2 Descripción de la base

Los censos de población y vivienda generan fuentes de información detallada sobre indicadores económicos y sociales, cuya relevancia radica en la política pública para la toma de decisión. Se compone de datos que proporcionada sectores socio-territoriales, como las localidades, e información geo-estadística, de los estados y municipios.

El Censo de Población y Vivienda se realiza a manera decenal; no obstante, con el fin de actualizar la estadística demográfica y socioeconómica en periodos más cortos que los decenales, con el fin de satisfacer la demanda en información, la INEGI realiza conteos de población y vivienda en periodos de cinco años.

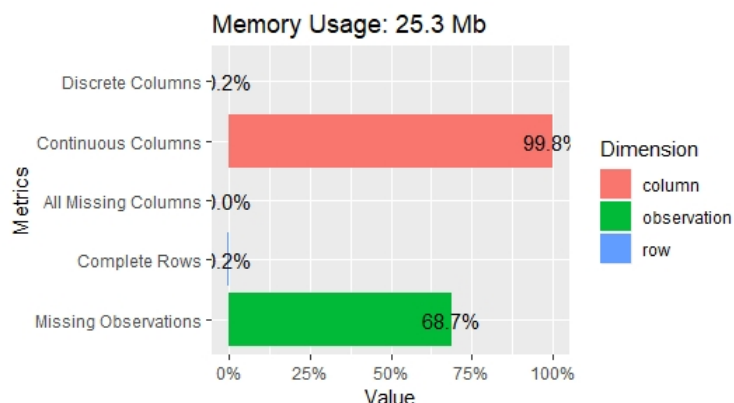
En base al Censo de Población y Vivienda, el Consejo Nacional de Población y Vivienda construye el índice de marginación para distintos niveles territoriales en México, con el fin de establecer un parámetro que analice cuando un sector de la sociedad se encuentre en una situación donde las oportunidades de desarrollo no se presenten, y cuya capacidad de encontrarlas no parece viable.

Para analizar el índice de marginación en el estado de Nuevo León se toma la base del Censo de Población y Vivienda del 2010 para el estado de Nuevo León<sup>1</sup>. La base se conforma de un total de 52 municipios, compuestos por 5,265 localidades (filas), y 625 características (columnas). El problema de trabajar con tantas localidades radica en el alto porcentaje de valores perdidos que cuenta la base, como se observa en la tabla 1, al rededor de 68.7% son valores perdidos.

---

<sup>1</sup>Nota: la base fue proporcionado por el Doctor a cargo la clase de Ciencia de Datos del CIMAT Monterrey

Table 1: Censo de Población y Vivienda 2010: Nuevo León



Para tratar el problema de los valores perdidos se utiliza el reporte oficial de los indicadores socioeconómicos, e índice de marginación de Nuevo León que proporciona el CONAPO en su sitio oficial<sup>2</sup>. El cual no consideran 3,228 localidades y cuenta con una población de 1,3470 habitantes, para las cuales no fue posible calcular el índice de marginación, debido a que estas localidades cuentan con una o dos viviendas o no registran información.

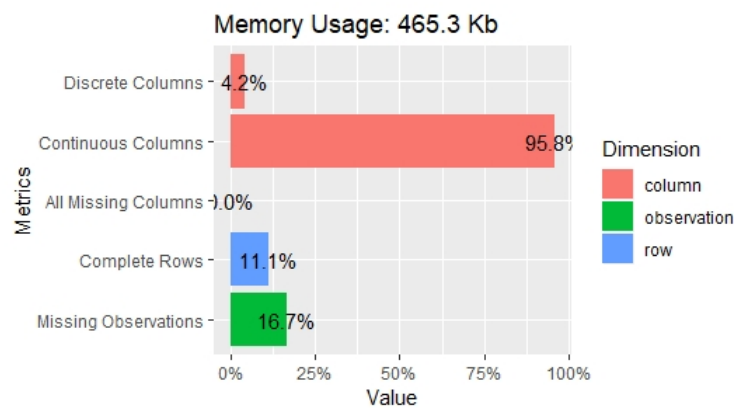
Una vez realizado el filtro, la base con la que se trabaja cuenta con 2,037 localidades - las cuales se centra el trabajo- y 626 características; sin embargo, no todas las características (columnas) se requieren para el índice de marginación, siendo la razón por la cual se implementa un filtrado de aquellas variables que no se requieren en el cálculo, obteniendo un total de 2,037 localidades y 24 características.

Al reducir el número de variables se evita descarta un mayor número de localidades que no cuentan con información para la construcción del índice de marinalización. Se puede apreciar en la figura 2 que el porcentaje de valores perdidos es solo del 16.7%. Con la figura 3 se resume las variables con las cual se trabajan<sup>3</sup>, y el porcentaje de valores no reportados en cada característica de la localidad.

<sup>2</sup>Vea: [http://www.conapo.gob.mx/es/CONAPO/Indices\\_de\\_Marginacion\\_publicaciones](http://www.conapo.gob.mx/es/CONAPO/Indices_de_Marginacion_publicaciones)

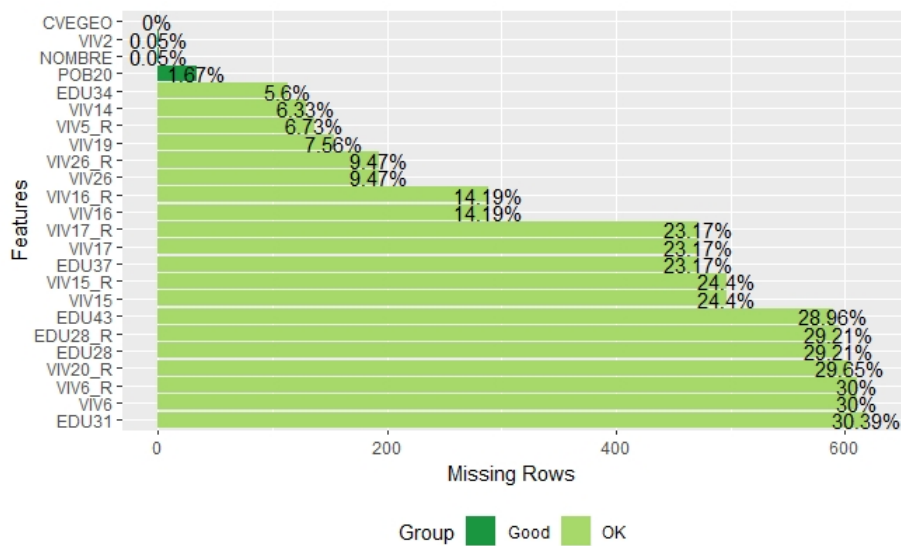
<sup>3</sup>Vea el anexo para una descripción de ellas

Table 2: Censo de Población y Vivienda 2010: Nuevo León



Filtración de información de acuerdo con CONAPO

Table 3: Base de datos



Filtración de información de acuerdo con CONAPO

### 3 Indicadores Marginación

De acuerdo con CONAPO, la marginación es un problema estructural de la sociedad, donde no se presentan oportunidades de desarrollo ni capacidades de obtenerlas, impidiendo

a familias y comunidades alcanzar determinadas condiciones de vida. La marginación, a su vez, se entiende como aquellos problemas sociales que se presentan tanto en las comunidades como en las localidades.

La problemática en la marginación se pretende cuantificar en base al índice que construye CONAPO en base al Censo de población que realiza INEGI, con el fin de entender cuándo un sector de la sociedad se encuentra en una situación donde no están presentes las oportunidades para el desarrollo, ni la capacidad para encontrarlas.

El estado por medio de políticas sociales es el que atiende las desigualdades y las oportunidades de desarrollo por medio de transferencias y aportaciones de servicios a los estados. De esta forma, las políticas sociales la mayoría de la veces son orientadas a reducir las inequidades de los grupos vulnerables. Es por esto, que es necesario contar con instrumentos analíticos que permitan identificar los lugares en los cuales existen dichas desigualdades, y es donde el índice de marginación toma relevancia.

## **Dimensiones de la marginación en localidades**

El índice de marginación se construye en base al método de reducción de dimensiones llamada componentes principales o PCA por sus siglas en inglés. A grandes rasgos es una técnica estadística que sintetiza la información en nuevas variables, creadas como resultado de una relación entre las variables originales, los factores que resultan de sintetizar la información son llamados componentes principales.

La reducción en la información (reducción de dimensión), se obtiene de ocho indicadores socioeconómicos que se construyen en base a la información del Censo de Población y Vivienda. Para el interés del trabajo, estos ocho indicadores se construyen para las 2,037 localidades. de las cuales, no todos aportan información suficiente para la construcción de algún indicador.

Para afrontar la problemática de falta de información, seguimos la metodología del



anexo técnico que proporciona CONAPO para el cálculo del índice de marginación, en el cual sugieren que aquellos municipios cuyo indicador no pueda ser calculado se impute con el valor medio del resto de las localidades; no obstante, la mayoría de los ocho indicadores presentan varianzas relativamente alta <sup>4</sup>, y es por eso que utilizamos la mediana en lugar de la media.

A continuación se presentan a manera general -vea el anexo para algo más detallado- las dimensiones en las que se clasifican los ocho indicadores que componen el índice de marginación a nivel localidad, y las variables que se utilizan del Censo de Población y Vivienda para el cálculo de los indicadores socioeconómicos.

1. *Porcentaje de población de 15 años o más analfabeta*

- \* Población de 15 años o más
- \* Población de 15 años o más analfabeta

2. *Porcentaje de población de 15 años o más sin primaria completa*

- \* Población de 15 años o más sin escolaridad
- \* Población de 15 años o más con primaria incompleta
- \* Población de 15 años o más con primaria completa
- \* Población de 15 años o más con secundaria incompleta
- \* Población de 15 años o más con secundaria completa
- \* Población de 18 años o más con educación pos-básica

**Indicadores sobre las viviendas**

3. *Porcentaje de viviendas particulares habitadas sin excusado*

- \* Viviendas particulares habitadas totales
- \* Viviendas particulares habitadas que disponen de sanitario

---

<sup>4</sup>Vea tabla 4.

4. *Porcentaje de viviendas particulares habitadas sin energía eléctrica*

- \* Viviendas particulares habitadas que disponen de luz eléctrica
- \* Viviendas particulares habitadas que no disponen de luz eléctrica

5. *Porcentaje de viviendas particulares habitadas sin agua entubada*

- \* Viviendas particulares habitadas que disponen de agua entubada fuera de la vivienda
- \* Viviendas particulares habitadas que disponen de agua entubada dentro de la vivienda

6. *Promedio de ocupantes por cuarto en viviendas particulares habitadas*

- \* Promedio de ocupantes por cuarto en viviendas particulares habitadas

7. *Porcentaje de viviendas particulares habitadas con piso de tierra*

- \* Viviendas particulares habitadas con piso de tierra
- \* Viviendas particulares habitadas con piso diferente de tierra

**Indicador sobre disponibilidad de bienes**8. *Porcentaje de viviendas particulares habitadas que no disponen de refrigerador*

- \* Viviendas particulares habitadas totales
- \* Viviendas particulares habitadas que disponen de refrigerador

A cada componente lo nombraremos con la letra  $I_i$ , donde  $i = 1, 2, \dots, 8$ , se refiere al indicador socioeconómico que componen al índice de marginación, esto con la finalidad de facilitar la lectura. De esta forma, si observamos la tabla 4, se presentan estadísticos descriptivos por indicador socioeconómico para el estado de Nuevo León para el año 2010. Se puede ver, que si se realiza la imputación -de los valores de las localidades que no registran algún indicador- con la media, la cifra es propensa a subestimar, siendo así que se opta por utilizar un estadístico más robusto como es la mediana (invariante a

valores atípicos).

Con la tabla 4 obtenemos un panorama general de Nuevo León; el porcentaje medio de población analfabeta *I1*, ronda en 7.743%; el porcentaje medio de población sin primaria completa *I2*, es de 39.33%, valor muy elevado para un estado como el de Nuevo León; el porcentaje medio de viviendas sin excusado es de *I3*, es de 7.387%; y para las que no cuentan con energía eléctrica de *I4*, es de 4.487%; el nivel medio del promedio de ocupantes por cuarto en vivienda es de *I5*, es de 40.47%; el porcentaje medio de población sin primaria completa *I6*, es de 39.33%; el porcentaje medio de viviendas con piso de tierra es *I7*, es de 5.77%; por último, el porcentaje promedio de población que no dispone de refrigerador *I8*, es 22.27%<sup>5</sup>.

Table 4: Estadísticos descriptivos por indicador socioeconómico, 2010

Indicador	1er Qtl	3er Qtl	Min	Media	Mediana	Max
I1	2.400	9.200	0.000	7.743	5.600	77.800
I2	26.50	50.00	0.00	39.33	38.20	100.00
I3	0.000	8.800	0.000	7.387	1.400	100.000
I4	0.000	0.000	0.000	4.487	0.000	100.000
I5	6.40	82.40	0.00	40.47	29.10	100.00
I6	1.000	1.400	0.400	1.235	1.200	10.000
I7	0.00	1.40	0.00	5.77	0.00	100.00
I8	2.60	25.90	0.00	22.27	12.20	100.00

(a) Información para el estado de Nuevo León

La relación entre indicadores socioeconómicos se mide por medio de la correlación. En la tabla 5, se presenta una matriz de correlación para los indicadores que componen al índice de marginación de las localidades de Nuevo León. Entre lo relevante se encuentra la correlación de 0.55 entre el porcentaje de la población que no dispone de refrigerador

<sup>5</sup>El indicador de porcentaje de población que no dispone refrigerador es análogo al indicador municipal que registra si la persona gana menos de dos salarios mínimos.

*I8*, y aquellos que no cuentan con energía eléctrica en el hogar *I4*, cuya explicación está de más.

Table 5: Matriz de correlaciones de los indicadores socioeconómicos, 2010



(a) Información para el estado de Nuevo León

## 4 Resultados

Se utiliza la matriz de correlación de los datos para el cálculo de los valores y vectores propios, con el fin de estandarizar valores que producen efectos de escala sobre las variables. Para calcular los componentes principales se utiliza el software R (versión, 3.5.2), con la librería *princom*, como resultado se obtienen ocho componentes principales, uno para cada indicador socioeconómico, cada uno con sus respectivos valores propios, los cuales se presentan en la tabla 6.

Los valores que se encuentran en la columna 4, en la tabla 6, determinan el porcentaje de variabilidad que explica cada componente; CONAPO, toma el primer componente para caracterizar al índice de marginación, el cual explica la mayor parte en el comportamiento de las ocho variables, y como se observa, su proporción de varianza es de 31%.

Table 6: Valores propios de la matriz de correlaciones y porcentaje de varianza explicada, 2010

Componentes	Valores Propios	Desviación Estándar	% Varianza	% Varianza Acum.
Comp.1	2.46	1.57	0.31	0.31
Comp.2	1.25	1.12	0.16	0.46
Comp.3	1.03	1.01	0.13	0.59
Comp.4	0.88	0.94	0.11	0.70
Comp.5	0.80	0.90	0.10	0.80
Comp.6	0.68	0.83	0.09	0.89
Comp.7	0.51	0.71	0.06	0.95
Comp.8	0.38	0.62	0.05	1.00

(a) Información para el estado de Nuevo León

En el cuadro 7, se presentan los coeficientes de los vectores propios estimados mediante los indicadores socioeconómicos; i.e., el peso relativo que tomó cada indicador. A su vez, se presenta una columna con el ponderado de lectura (no es porcentaje), siendo un vector que se calcula realizando el producto de valor propio por vector propio para el primer componente principal.

Se puede observar que el indicador que más aporta a la marginación es la carencia de refrigerador *I8*, con 1.231. Le sigue las viviendas que no cuentan con luz eléctrica con 0.89, y aquellas personas que no registran primaria completa 0.888.

Table 7: Valores propios de la matriz de correlaciones y porcentaje de varianza explicada, 2010

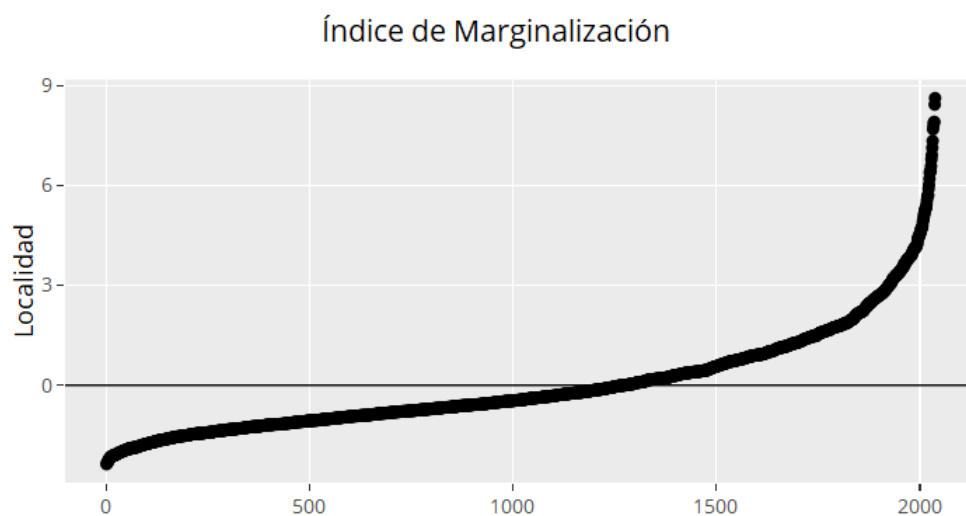
<b>Indicador socioeconómico</b>	<b>Pesos</b>	<b>Ponderador de lectura</b>
11	0.312	0.768
12	0.361	0.888
13	0.345	0.848
14	0.365	0.898
15	0.347	0.854
16	0.151	0.372
17	0.356	0.877
18	0.500	1.231

(a) Información para el estado de Nuevo León

## 5 Determinación de los grados de marginación

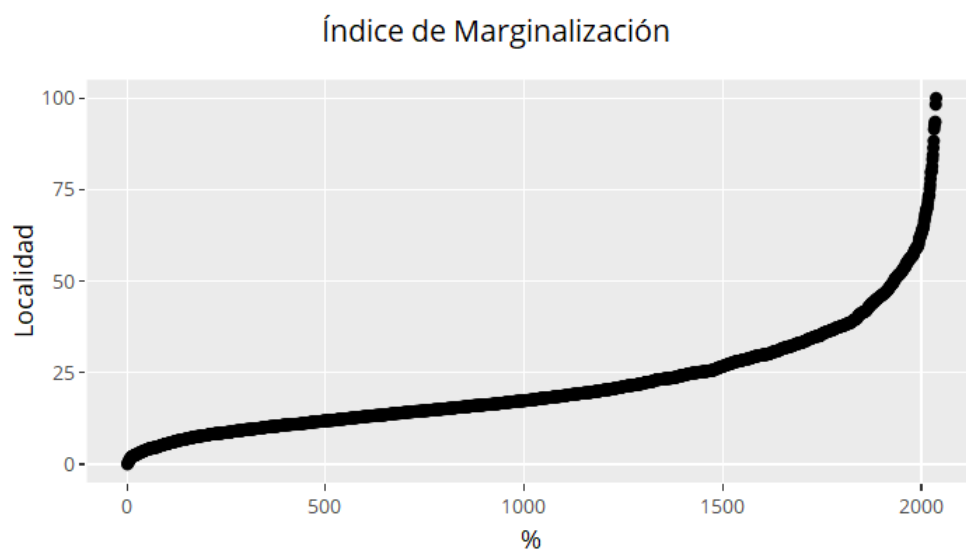
Una vez que se calcula el índice para las 2,037 localidades de Nuevo León, se realizan cinco grupos que representan los grados de marginación. La figura 2, muestran los valores de los índices de marginación. El índice considera a una localidad con problemas estructural y escasas oportunidades de desarrollo, si su valor es muy alto, caso contrario, si su valor es muy bajo.

Figure 2: Distribución del índice de marginación de Nuevo León a nivel localidad, 2010



Con la idea de mejorar la lectura e interpretación del índice, se transforma el indicador a una escala que va de 0 a 100. La figura 3, explica la distribución que tiene el índice en unidades porcentuales, se puede apreciar que aproximadamente más del de mil localidades cuentan con un nivel de marginación menor al 25%.

Figure 3: Distribución del índice de marginación de Nuevo León en porcentaje a nivel localidad, 2010



Para la estimación de los grados se emplearon los cortes de la técnica de Dalenius y Hodges en dos fases<sup>6</sup>. Se utiliza la librería *statification* que contiene a la función *strata.cumrootf*, cuya salida son los límites superiores del intervalo y la frecuencia de las observaciones que caen dentro de él. Los límites inferior y superior de los grados de marginación estimados se muestran en el cuadro 8. Se ve que de 2,037 localidades, 537, presentan niveles muy bajos, 672 bajos, 437 medios, 291 altos, y sólo 120 reportan niveles muy altos.

Los resultados indican, que en Nuevo León, aproximadamente existen un 20% de localidades con problemas estructurales, y cuyas oportunidades de desarrollo son escasas, impidiendo a las familias y comunidades aspirar a determinadas condiciones de vida.

Table 8: Localidades y población según estratos del índice de marginación, 2010

Grado de marginación	Número de localidades	Límite inferior	Límite superior
MuyBajo	517	0	12
Bajo	672	12	20
Medio	437	20	31
Alto	291	31	48
MuyAlto	120	48	100

## 6 Contaste respecto al índice de marginación oficial

En este apartado, se contrasta el índice de marginación calculado, respecto al que reporta el CONAPO; asimismo, se utilizan los componentes de los indicadores socioeconómicos que se incluyen en el reporte, con el fin de realizar la técnica PCA y obtener un tercer punto de comparación.

Primero, vamos a comparar aquellos indicadores socioeconómicos que reporta el CONAPO respecto al elaborado, de manera propia, en base al Censo de Población y Vivienda. En la tabla 9 se tienen los estadísticos básicos de los indicadores reportados

---

<sup>6</sup>Ver anexo



en el CONAPO. Si se comparan contra los de la tabla 4, se observa que se son parecidos.

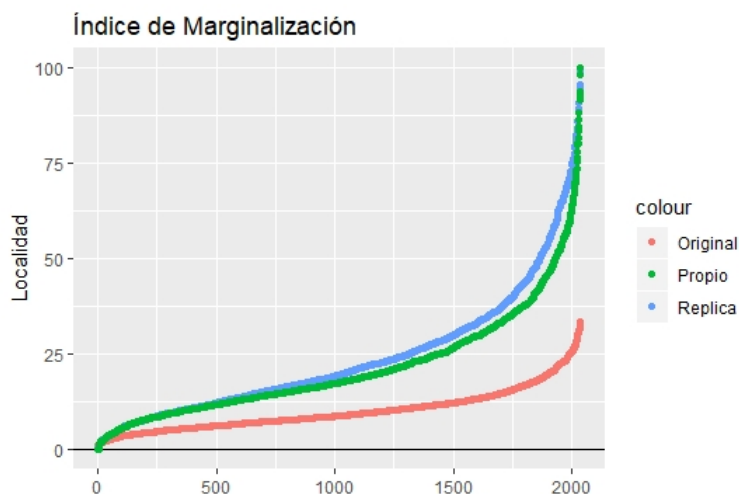
Table 9: Estadísticos descriptivos por indicador socioeconómico a nivel estatal, 2010

Indicador	1er Qtl	3er Qtl	Max	Media	Mediana	Min
I1	2.400	9.200	77.800	7.743	5.600	0.000
I2	26.50	50.00	100.00	39.33	38.20	0.00
I3	0.000	8.800	100.000	7.387	1.400	0.000
I4	0.000	0.000	100.000	4.487	0.000	0.000
I5	6.40	82.40	100.00	40.47	29.10	0.00
I6	1.000	1.400	10.000	1.235	1.200	0.400
I7	0.00	1.40	100.00	5.77	0.00	0.00
I8	2.60	25.90	100.00	22.27	12.20	0.00

A continuación, en la figura 4, se presentan tres series del índice de marginación en forma porcentual; el primero, es el original reportado por el CONAPO para Nuevo león (Original); el segundo (Replica), se utilizan los indicadores socioeconómicos del CONAPO y se aplica componentes principales -siguiendo su metodología- para obtener el índice, paso seguido se transforma a escala de 0 a 100; el tercero es el calculado de forma propia, realizado con el Censo de Población y Vivienda del 2010 (Propio).

Calculando el índice de marginación con los indicadores socioeconómicos del CONAPO, el índice no se aproxima al que ellos reportan. La línea rosa, el índice oficial, subestima la línea azul que representa la replica. No obstante, la línea verde, que representa el índice calculado en este trabajo, se asemeja al que se calcula utilizando los indicadores socioeconómicos contruidos por el CONAPO.

Figure 4: Contraste de la distribución del índice de marginación a nivel estatal, 2010



Viendo la estratificación que se calcula para el índice elabora en este trabajo, y utilizando el oficial (original), las localidades de Nuevo León presentan mejor perspectiva de vida, registrando solo 9 localidades con un índice alto, y ninguno como muy alto. Por otro lado, al utilizar el índice replicado, la frecuencia de las localidades se aproximan a los de este trabajo.

Table 10: Estratificación del índice de marginación a nivel estatal, 2010

Grado de marginación	Frecuencia Original	Frecuencia Propia	Frecuencia Replica	Límite inferior	Límite superior
MuyBajo	1465	517	482	0	12
Bajo	440	672	558	12	20
Medio	123	437	490	20	31
Alto	9	291	320	31	48
MuyAlto	0	120	187	48	100

## 7 Recomendaciones

La sección de recomendaciones brinda sugerencias a una forma alterna de calcular el índice de marginación. Cabe resaltar que las dimensiones que componen el índice

abarcan de manera correcta - a mi perspectiva- lo que se pretende medir.

La dimensión de educación, con indicadores de población analfabeta y aquellos que no finalizan la primaria, capturan la relación entre años de estudio y nivel de ingreso esperado. La parte del índice que se enfoca a la vivienda, engloba la acumulación de bienes y servicios, que a manera macroeconómica, hace analogía de la acumulación del capital. Inclusive que el último indicador hable de acumulación de bienes, medir el tener o no tener refrigerador, equivale -de acuerdo a CONAPO- a la población que recibe a lo menos dos salarios mínimos, brindando un panorama del poder adquisitivo de las localidades.

Con lo anterior se deduce que los indicadores pueden ser medidas indirectas de algún tipo de ingreso; entonces, el índice de marginación aproxima a un índice de desigualdad económica. Siendo así, una de las recomendación para la construcción alterna es; tener en cuenta otro indicador - a parte de los ocho existentes- que cuantifique la desigualdad en términos de salud; la medida de salud, puede elaborarse con el mismo Censo de Población y Vivienda, ya que cuenta con apartados en el cual se realiza la pregunta al entrevistado de si tiene o no algún servicio de salud.

El indicador de salud es la recomendación que se da por la parte de los componentes a utilizar. Para la cuestión estadísticas, se recomienda no utilizar la media muestral de los indicadores socioeconómicos como medida de imputación. En este trabajo, se utiliza la mediana muestral para evitar posibles valores atípicos; no obstante, existen métodos de imputación más sofisticados.

Entre otras sugerencias, se debe contemplar un mayor número de componentes principales, y construir una medición que pondere de acuerdo a su proporción de varianza explicada. Por último, se necesita que dispongan del algoritmo que implementan para el cálculo del índice, con el fin replicar el indicador y reducir incertidumbre sobre las cantidades que reportan.

## 8 Conclusión

Se utilizaron datos del Censo de Población y Vivienda del 2010 para el Estado de Nuevo León, y con base a la metodología de la CONAPO se realizó el cálculo del índice de

marginación para 2,037 localidades.

Algunas de las limitantes del trabajo son; el índice sólo se calcula para el 38% de las localidades de Nuevo León; y la forma tan poco confiable de realizar la imputación -poner valor medio- en los indicadores socioeconómicos llegan a sobre estimar la cifra.

Entre los resultados se encuentra que solo 20% de las localidades presentan grados de marginación altos o muy altos, de acuerdo con el criterio de estratificación del CONAPO. De esta forma, 80% de las localidades registran grados aceptables de marginación, brindando certidumbre en el desarrollo y crecimiento tanto económico como social, y permitiendo a las familias mejorar en sus condiciones de vida.

## Anexo

### Diccionario de datos

Nota: las variables para la construcción del porcentaje de población de 15 años a más sin primaria completa ( $p15ym_{se}$ ,  $p15pri_{in}$ ,  $p15pri_{co}$ ,  $p15se_{in}$ ,  $p15se_{in}$ ,  $p18ym_{pb}$ ), no se encuentra en la base que se proporciona en los insumos de la clase de Ciencia de Datos, y lo que se hace es descargar el Censo de Población y Vivienda, nuevamente, para realizar la misma limpieza que se le da a la base que se proporciona en la clase. La base puede encontrarse en la página oficial de la INEGI:

<https://www.inegi.org.mx/programas/ccpv/2010/default.html>.

## Descripción de variables

Descripción	Variable
Población de 15 años o más	EDU28
Población de 15 años o más analfabeta	POB20
Población de 15 años o más sin escolaridad	p15ym_se
Población de 15 años o más con primaria incompleta	p15pri_in
Población de 15 años o más con primaria completa	p15pri_co
Población de 15 años o más con secundaria incompleta	p15sec_in
Población de 15 años o más con secundaria completa	p15sec_co
Población de 18 años o más con educación pos-básica	p18ym_pb
Viviendas particulares habitadas totales	VIV2
Viviendas particulares habitadas que disponen de sanitario	VIV19
Viviendas particulares habitadas que disponen de luz eléctrica	VIV15
Viviendas particulares habitadas que no disponen de luz eléctrica	VIV14
Viviendas particulares habitadas que disponen de agua entubada fuera de la vivienda	VIV17
Viviendas particulares habitadas que disponen de agua entubada dentro de la vivienda	VIV16
Promedio de ocupantes por cuarto en viviendas particulares habitadas	VIV5_R
Viviendas particulares habitadas con piso de tierra	VIV6
Viviendas particulares habitadas con piso diferente de tierra	VIV2-VIV6
Viviendas particulares habitadas totales	VIV2
Viviendas particulares habitadas que disponen de refrigerador	VIV26

Nota :las variables de la fila 3 a 8 se toman de una base distinta.

## Cálculo de los indicadores socioeconómicos

$I1$ , porcentaje de población de 15 años a más analfabeta.

$$I1 = \frac{EDU28}{POB20} * 100$$

$I2$ , porcentaje de población de 15 años a más sin primaria completa.

$$I2 = \frac{p15ym_{se} + p15pri_{in}}{p15ym_{se} + p15pri_{in} + p15pri_{co} + p15se_{in} + p15se_{in} + p18ym_{pb}} * 100$$

$I3$ , porcentaje de viviendas particulares habitadas sin excusado.

$$I3 = \frac{VIV2 - VIV19}{VIV2} * 100$$

$I4$ , porcentaje de viviendas particulares habitadas sin energía eléctrica.

$$I4 = \frac{VIV15}{VIV15 + VIV14} * 100$$

$I5$ , porcentaje de viviendas particulares habitadas que no disponen de agua entubada.

$$I5 = \frac{VIV17}{VIV16 + VIV17} * 100$$

$I6$ , promedio de ocupantes por cuarto de viviendas particulares habitadas sin excusado.

$$I6 = VIV5_R$$

$I7$ , porcentaje de viviendas particulares habitadas con piso de tierra.

$$I7 = \frac{VIV6}{VIV6 + (VIV2 - VIV6)} * 100$$

$I8$ , porcentaje de viviendas particulares habitadas que no disponen de refrigerador.

$$I1 = \frac{VIV2 - VIV26}{VIV2} * 100$$

## Construcción del índice

El índice de marginación son los scores del método PCA, la forma de calcularlos es la siguientes:

Sea  $\mathbf{w}_1$  el vector de pesos o dirección (loadings) del primer componente principal, y sea  $X$  la matriz que contienen las observaciones, si se realiza la multiplicación

$$z_{1j} = \mathbf{w}_1' X$$

se obtienen los  $z_{1j}$  que son los valores del índice de marginación para la  $j$ -ésima localidad, o también conocidos como scores.

Para obtener el índice de marginación en una escala de 0 al 10, se realiza la siguiente operación:

$$IMG\% = \frac{IM - \min(IMG)}{\max(IMG) - \min(IMG)}$$

La proporción de la varianza explicada está dada por la siguiente expresión:

$$\%Var = \frac{\lambda_i}{\sum_{j=1}^n (\lambda_i)}$$

con  $\lambda_i$  como el  $i$ -ésimo valor propio de la matriz de covarianzas o correlación de los datos.

## Construcción del grado de marginación

El método de Dalenius - Hodges consiste en la conformación de estratos de manera que la varianza obtenida sea mínima para cada estrato. El procedimiento es el siguiente:

Sea  $n$  el número de observaciones y  $L$  el número de estratos.

- 1– Ordena las observaciones de manera ascendente
- 2– Agrupa las observaciones en  $J$  clases, donde  $J = \min(L * 10, n)$
- 3– Calcula los límites para cada clase de la siguiente manera:

$$\liminf C_k = \min(x_i) + (k + 1) \frac{\max(x_i) - \min(x_i)}{J}$$

$$\limsup C_k = \min(x_i) + (k) \frac{\max(x_i) - \min(x_i)}{J}$$

Los intervalos se toman abiertos por izquierda y cerrados por derecha.

- 4– A partir de los límites, obtén la frecuencia en cada caso
- 5– Obtén la raíz cuadrada de la frecuencia para cada clase
- 6– Acumula la suma de la raíz cuadrada de las frecuencias
- 7– Divide el último valor acumulado entre el número de estratos

$$Q = \frac{C_J}{L}$$

- 8– Los puntos de corte de cada estrato se toman sobre el acumulado de la raíz cuadrada de las frecuencias en cada clase de acuerdo a los siguiente  $Q, 2Q, \dots, (h - 1)Q$