

Para la estimación del índice de equipamiento por localidad, en este anexo se presentan los conceptos, las variables seleccionadas y la estimación de los indicadores simples que proveen información para el cálculo del índice. Una vez calculados los 13 indicadores mediante el método de Componentes con el objetivo de simplificar la información y darle un peso adecuado a cada una de las variables de equipamiento dentro de los 13 subsistemas de equipamiento. Se calcula el índice sintético mediante la técnica de distancia DP2 y después se describe la aplicación de la técnica de estratificación de Dalenius and Hodges para asignar el grado de equipamiento en las localidades que disponen información en alguno de los indicadores.

Indicadores de equipamiento

En la estimación de los indicadores parciales para el índice, se agruparon de acuerdo a los 13 subsistemas de equipamiento, de los cuales 12 fueron establecidos por el Sistema Normativo de Equipamiento elaborado en 2012, por la, en ese entonces, Secretaría de Desarrollo Social (SEDESOL), agregando además el subsistema de Banca.

Por otro lado, se consideraron fuentes de información del Directorio Estadístico Nacional de Unidades Económicas (DENUE) del Instituto Nacional de Estadística y Geografía (INEGI) y del Sistema Nacional de Información Pública (SNIE) a cargo de la Secretaría de Educación Pública (SEP); los cuales se tomaron en consideración tanto entidades públicas como privadas formando un total 46 variables que a partir de un análisis exploratorio¹ se excluyen variables como: intermediarios, almacenes y abarrotes (abasto), transporte aéreo y por agua (transporte), golf y marinas (deportes), relaciones exteriores y administración pública general (gobierno), los parques naturales y jardines (recreación), sitios históricos (cultura), cementerios (servicios urbanos), escuelas básicas (educación) y actividad bursátil (banca). (Quedándose con 32 variables equipamiento)

El universo de análisis fueron 11 786 localidades, cuyas claves están identificadas dentro del Marco Geoestadístico Nacional (MGN) elaborado por el INEGI, de las cuales 7 365 tuvieron representación cartográfica de polígono, 4093 de puntos y 328 no eran propiamente localidades.

Cálculo de los indicadores

Para cada uno de los 32 indicadores simples, se calcularon a partir de la división de establecimientos disponible en la localidad respecto al total equipamiento disponible en el municipio. Con la finalidad de poder trabajar con variables continuas y estas no sesguen los resultados en la medición del modelo.

$$w_{hjk} = \sum_{j=1}^{n_h} \frac{S_{hjk}}{S_{hk}}; \text{ para } \begin{cases} k:1,2,\dots, m \\ h:1,2,\dots, 2458 \text{ municipios} \end{cases}$$

Donde:

w_{hjk} = Razón de la variable de interés k en la j –ésima localidad del h –ésimo municipio,

n_h = Número total de la localidad en el h –ésimo municipio,

S_{hk} = Total de equipamiento en la variable de interés k del h –ésimo municipio,

S_{hjk} = Total de equipamiento en la variable de interés k en la localidad j –ésima del h –ésimo municipio.

¹ Análisis de correspondencias y Análisis de correlaciones para detectar y representar estructuras subyacentes en los datos.

También se recurrió al método de distancias DP2 para cada una de las variables de equipamiento, donde se indicaba que las variables: escuelas básicas, abarrotes y oficinas generales de gobierno, no eran relevantes debido a que eran explicada por otras variables.

Tabla 1. Distribución de establecimiento en las localidades

Variables	Localidades			
	Sin equipamiento		Con equipamiento	
	Absoluto	Relativo	Absoluto	Relativo
Mayoreo	5 913	50.17	5 873	49.83
Intermediarios	11 693	99.21	93	0.79
Almacenes	10 652	90.38	1 134	9.62
Abarrotes	5 873	49.83	5 913	50.17
Farmacias	7 681	65.17	4 105	34.83
Menudeo	5 512	46.77	6 274	53.23
Residencias	11 610	98.51	176	1.49
Asilos -Orfanatos	10 827	91.86	959	8.14
Orientación	8 897	75.49	2 889	24.51
Alimentos	10 274	87.17	1 512	12.83
Capacitación	11 476	97.37	310	2.63
Guarderías	9 675	82.09	2 111	17.91
No Bursátil	8 031	68.14	3 755	31.86
Act_Bursatil	11 198	95.01	588	4.99
Seguros	11 273	95.65	513	4.35
Correo	10 463	88.77	1 323	11.23
Mensajería	11 213	95.14	573	4.86
TELECOM	9 832	83.42	1 954	16.58
Museos	11 223	95.22	563	4.78
Sitio Histórico	11 781	99.96	5	0.04
Biblioteca	8 868	75.24	2 918	24.76
Golf	11 702	99.29	84	0.71
Marinas	11 751	99.70	35	0.30
Club Deportivo	9 532	80.88	2 254	19.12
Otros	9 747	82.70	2 039	17.30
Básica	7 338	62.26	4 448	37.74
Media	8 462	71.80	3 324	28.20
Superior	10 979	93.15	807	6.85
General	7 793	66.12	3 993	33.88
Justicia	8 988	76.26	2 798	23.74
Bienestar	8 851	75.10	2 935	24.90
Relaciones Exteriores	11 641	98.77	145	1.23
Jardín -Parque Natural	11 701	99.28	85	0.72
Parque de Diversiones	8 823	74.86	2 963	25.14
Casino	11 227	95.26	559	4.74
Cines	11 497	97.55	289	2.45
Consultorios	7 425	63.00	4 361	37.00
Centro de Salud	11 559	98.07	227	1.93
Laboratorio_ Bancos	9 890	83.91	1 896	16.09
Hospitales	8 997	76.34	2 789	23.66
Gasolineras	9 051	76.79	2 735	23.21
Cementerios	9 105	77.25	2 681	22.75
Aéreo	11 660	98.93	126	1.07
Tierra	9 911	84.09	1 875	15.91
Agua	11 630	98.68	156	1.32
Carga	10 277	87.20	1 509	12.80

Una vez calculados los indicadores simples, se cuestiona la importancia de cada variable con respecto a las demás, por lo que se recurren a un sistema de ponderación que pueda otorgar un peso a cada indicador simple que pertenece a los diferentes subsistemas de equipamiento. Por lo que se recurre al análisis de componentes principales, debido a, que permitirá reducir la información de cada subsistema en factores o componentes principales y a su vez asignarle a cada indicador simple un peso que explique la relación con los demás indicadores de acuerdo a su grado de correlación. Permitiendo analizar la variabilidad del conjunto y ordenarlos de acuerdo a su importancia, lo cual la agrupación en los componentes no condiciona el cálculo ni los resultados del índice sintético. (Resultando una medida ordinal)

La técnica de componentes principales busca transformar el espacio vectorial generado por un vector $X = (X_{n1}, X_{n2}, \dots, X_{np})$ en uno nuevo conjunto, es decir, se desea encontrar $Z = (Z_1, Z_2, \dots, Z_k), k \leq p$ que sean combinaciones lineales de los indicadores y que explique la mayor parte de la variabilidad.

$$Z = AX$$

- $A = (a_{n1}, a_{n2}, \dots, a_{np})$. – Es el vector de pesos factoriales.

Como objetivo del método es concentrar en un solo valor las magnitudes y variaciones de los indicadores y la relación entre ellos, por ello se va a agrupar cada indicador simple en los diferentes sectores de servicios en un único índice sintético. Bien, si sólo se selecciona un componente principal para cada subsistema, esté actuaría como un índice sintético del conjunto de variables, pero si se desea retener a más de un componente, ello implicaría una transformación para poder utilizarlo como indicador.

Utilizando el promedio de las puntuaciones de cada componente por la raíz cuadrada de la varianza de cada componente. En este sentido, el índice se debe calcular como:

$$\text{Indicador}_j = \frac{\sum_{i=1}^k Z_{ij} \sqrt{\lambda_i}}{\sum_{i=1}^k \sqrt{\lambda_i}} ; i = 1, 2, \dots, n(\text{componentes})$$

Siendo:

- Z_{ij} la puntuación del componente i -ésimo para la unidad de observación j -ésima.
- $\sqrt{\lambda_i}$ La raíz cuadrada del autovalor para dicho componente, garantizando así que los componentes contengan una mayor varianza explicada tenga una mayor ponderación en la calificación del índice.

Frente a esto se utilizará la técnica de componentes principales que nos ayudará a reducir la información por columnas donde se agruparan de acuerdo a los 13 subsistemas y a su vez se utilizará el método DP2 que hará las pertinentes comparaciones entre las filas.

Construcción del índice

Se define la **Distancia – P₂** de la forma:

$$DP_2 = \sum_{i=1}^n \frac{d_i}{\sigma_i} (1 - R_{i,i-1,\dots,1}^2) ; \text{ con } R_1^2 = 0$$

Donde:

- $d_i(r,*) = |x_{ri} - x_{*i}|$, es la distancia de la situación en la localidad r con respecto a la base de referencia $_*$, $x_* = (x_{*1}, x_{*2}, \dots, x_{*n})$

Se define al valor de referencia, siendo este **el valor mínimo** de cada indicador simple como referencia, representado como la “situación no deseada”.

- σ_i , es la desviación estándar del indicador parcial del componente i (I_i)
- $(1 - R_{i,i-1,\dots,1}^2)$, El factor corrector evita las redundancias, al eliminar de los indicadores parciales la información ya contenida en los indicadores precedentes.
- $R_1^2 = 0$; Porque la primera componente aporta toda la información al no existir un componente previo. Por ello su ponderación es la unidad.
- $R_{i,i-1,\dots,1}^2$, es el coeficiente de determinación de la regresión de cada indicador parcial i con respecto a los otros indicadores $(i - 1, i - 2, \dots, 1)$

Por lo tanto el DP2 implica decidir sobre el orden de entrada de los indicadores parciales en el proceso del cálculo. Es decir, que se señala que indicador parcial i es el primero en aportar su varianza al índice global, cuál será el segundo, etc.

Además cumple con las siguientes propiedades: no negatividad, homogeneidad, conmutatividad, desigualdad triangular, existencia y determinación, monotonía, unicidad, transitividad, no duplicidad de información, invariancia al cambio de origen y/o de escala en las unidades y exhaustividad.

Criterios de selección de variables

La utilización de los criterios de validación de las variables, permite identificar la importancia de cada variable en el fenómeno, la aportación de cada variable.

Coeficiente de Discriminación

$CD_i(P)$, es el **Coeficiente de Discriminación de Ivanovic**, que mide el poder discriminante de la variable i en el conjunto de observaciones (P), y se define como:

$$CD_i(P) = \frac{2}{m(m-1)} \sum_{j,l>j}^{k_j} m_{ji} m_{li} \left| \frac{x_{ij} - x_{li}}{\bar{x}_i} \right|$$

Donde

m_{ij} el número de localidades en el conjunto P

x_{ij} el valor de la variable x_i

k_i es el número de diferentes valores que x_i toma en el conjunto P .

[Zarzosa; 1994] el Coeficiente de discriminación, está comprendida entre **0 y 2**, correspondientes estos valores a los dos casos teóricos extremos en cuanto a poder de discriminación.

- $CD = 0$, La variable toma el mismo valor para todas las localidades, indicando que esa variable posee nulo poder discriminante.
- $CD = 2$, el poder discriminante de esa variable es total.

Cantidad de Información Global de Ivanovic-Pena Relativa Individual

Una vez obtenidos los coeficientes de discriminación, se ha estimado la "**Cantidad de Información Global de Ivanovic Pena Relativa Individual**" [Somarriba, (2013)] definida como:

$$\alpha_i = \frac{CD_i(1 - R_{i,i-1,\dots,1}^2)}{\sum_{i=1}^n (1 - R_{i,i-1,\dots,1}^2)}$$

Esta medida, comprendida entre 0 y 1, combina la información útil y el poder discriminante de cada indicador simple y mide la cantidad de información (combinada) relativa que aporta individualmente cada indicador simple, cuando entra de forma ordenada a formar parte del indicador sintético DP2. La suma de todos los valores de α_i es la unidad.

El cálculo del índice DP2, se realiza utilizando como variables los 13 indicadores del subsistema de equipamiento. Donde el valor mínimo que obtiene el índice DP2 es 0, su valor máximo es 18, el valor de su mediana es 0.06 y su valor medio es 1.827.

Método de Dalenius & Hodges

El método de Dalenius-Hodges (1959), forma estratos de manera que la varianza se mínima al interior de cada estrato y máxima entre cada uno de ellos. Forma estratos lo más homogéneos posibles.

Este tipo de procedimiento utiliza la raíz de las frecuencias acumuladas para la construcción de los estratos se lleva a cabo para la división de la población en el estrato L :

$$Q = \frac{1}{L} \sum_{i=1}^J f_i$$

Donde:

J es el número de clases dentro del grupo de la variable ordenada X

f_i es la frecuencia en cada clase J

L es el número de estratos

Se clasifican los valores del índice DP2 en cinco grupos, de los cuales se obtienen los grados: muy alto, alto, medio, bajo y muy bajo.

Los límites inferior y superior de los grados de equipamiento se muestran en el cuadro 2 donde los valores del índice se encuentran en el intervalo [0,18]. Señalando la variación en el comportamiento de la marginación en las localidades del país para las cuales se estimó el índice.

Tabla 2. Localidades y población según los estratos del índice de equipamiento

Grado de equipamiento	Localidades		Población 2018		Límites del estrato
	Absolutos	Relativos	Absolutos	Relativos	
Muy alto	558	4.7%	69 730 194	73.0%	(11.0382 18.0000]
Alto	810	6.9%	8 647 798	9.1%	(6.4789 11.0382]
Medio	967	8.2%	3 834 410	4.0%	(2.8795 6.4789]
Bajo	1 598	13.6%	6 815 522	7.1%	(0.4799 2.8795]
Muy bajo	7 853	66.6%	6 435 056	6.7%	[0.0000 0.4799]
Total	11 786	100.0%	95 462 981	100.0%	

Para los cálculos del índice se utilizaron las paqueterías; "p2distance", "stratification", "FactoMineR" y "corrplot" del programa R.