

Índice de desigualdad

Metodología

Construcción del índice:

De manera general, a cada Área Geoestadística Básica (AGEB) se le asigna un valor del índice, que depende de tres factores, detallados a continuación:

- 1) Promedio ponderado de las variables involucradas en el índice.
- 2) Cantidad de grupos en que se divide a la población.
- 3) Filtro geográfico elegido por el usuario.

1. Promedio ponderado: Para construir la ponderación o peso que debe representar cada variable del AGEB, se emplea la técnica de *Análisis de Componentes Principales* (PCA por sus siglas en inglés) con la información de todas las AGEB del estado de Jalisco, consultada del Censo de población y vivienda 2020 del INEGI.

En primer lugar, se estandariza cada columna de la matriz de datos, restando medias muestrales y dividiendo por las desviaciones estándar de las observaciones de cada variable. A continuación se muestran los primeros tres valores (para tres AGEB) de cada variable:

```
head(X, 3)
```

```
##      sin_disc escolaridad  ocupados      salud      piso      energia
## 1  0.6449160  -1.053778 -0.1407936 -0.2547810 0.09564022 0.02140901
## 2 -0.8623501  -1.053778  0.2036481 -0.6821894 0.08106616 0.06721624
## 3  0.4924194  -1.107944  0.6916008  0.4198326 0.31117423 0.11228760
##           agua  tinaco  escusado  drenaje  refri  lavadora      auto
## 1 -0.06074670 0.5396155 -0.08090781 0.0178194 -0.3258712 0.6518579 -0.06980178
## 2  0.09944589 0.6514380  0.11190758 0.1611460  0.3471814 0.5621573 -0.31891643
## 3  0.18006569 0.8701728  0.21630486 0.2124088 -0.2790131 0.3179721 -0.42122286
##      celular  internet  invfec  invhacin
## 1 -0.1722984 -0.19707921 -0.8151127 -0.1969912
## 2 -1.1987500 -0.09792668 -1.0347070  0.4954140
## 3  0.2867014 -0.74970352 -0.3440454 -0.7495839
```

Las variables involucradas son:

- 1) “sin_disc” = “Porcentaje de habitantes sin discapacidades, limitaciones, problemas o condiciones mentales.”,
- 2) “escolaridad” = “Suma del total de años escolares aprobados por las personas mayores a 15 años de edad, entre las personas del mismo grupo de edad.”,
- 3) “ocupados” = “Proporción de personas mayores a 12 años que tenían trabajo en la semana de referencia.”,
- 4) “salud” = “Proporción de personas con derecho a recibir servicios médicos en alguna institución de salud pública o privada”,

- 5) “piso” = “Porcentaje de viviendas particulares habitadas con piso de cemento o firme, madera, mosaico u otro material.”,
- 6) “energía” = “Proporción de viviendas particulares habitadas que tienen luz eléctrica.”,
- 7) “agua” = “Porcentaje de viviendas particulares habitadas que tienen disponibilidad de agua entubada dentro de la vivienda, o fuera de la vivienda, pero dentro del terreno.”,
- 8) “tinaco” = “Proporción de viviendas particulares habitadas que disponen de tinaco.”,
- 9) “escusado” = “Porcentaje de viviendas particulares habitadas que tienen escusado, retrete, sanitario, letrina u hoyo negro.”,
- 10) “drenaje” = “Proporción de viviendas particulares habitadas que tienen drenaje conectado a la red pública, fosa séptica, barranca, grieta, río, lago o mar.”,
- 11) “refri” = “Proporción de viviendas particulares habitadas que tienen refrigerador.”,
- 12) “lavadora” = “Proporción de viviendas particulares habitadas que tienen lavadora.”,
- 13) “auto” = “Proporción de viviendas particulares habitadas que tienen automóvil o camioneta.”,
- 14) “celular” = “Proporción de viviendas particulares habitadas que tienen teléfono celular.”,
- 15) “internet” = “Porcentaje de viviendas particulares habitadas que tienen servicio de internet.”,
- 16) “invfec” = “Total de mujeres de 12 o más años de edad entre el total de hijos nacidos vivos de mujeres del mismo grupo.”,
- 17) “invhacin” = “Número de cuartos de viviendas particulares entre el número de personas que residen en ellas.”

Esta nueva matriz \mathbf{X} es entonces la base a partir de la que se calculan los componentes principales, construyendo la matriz de covarianzas de la misma, para obtener su descomposición en valores singulares y extraer así los componentes principales, ordenando los *eigenvectores* de forma descendente, de acuerdo con sus *eigenvalores*. A continuación se presentan las desviaciones estándar y las proporciones que los componentes abarcan de la varianza:

```
summary(cp)
```

```
## Importance of components:
##              PC1    PC2    PC3    PC4    PC5    PC6    PC7
## Standard deviation  2.6730 1.8265 1.22979 1.06213 0.9625 0.84658 0.64452
## Proportion of Variance 0.4203 0.1963 0.08896 0.06636 0.0545 0.04216 0.02444
## Cumulative Proportion 0.4203 0.6166 0.70551 0.77187 0.8264 0.86853 0.89296
##              PC8    PC9    PC10    PC11    PC12    PC13    PC14
## Standard deviation  0.64115 0.5683 0.5032 0.47743 0.40572 0.3688 0.32578
## Proportion of Variance 0.02418 0.0190 0.0149 0.01341 0.00968 0.0080 0.00624
## Cumulative Proportion 0.91714 0.9361 0.9510 0.96444 0.97413 0.9821 0.98837
##              PC15    PC16    PC17
## Standard deviation  0.31828 0.23505 0.20292
## Proportion of Variance 0.00596 0.00325 0.00242
## Cumulative Proportion 0.99433 0.99758 1.00000
```

Los primeros cuatro componentes principales explican aproximadamente el 77% de la varianza de la información, contando además con *eigenvalores* mayores a 1, con lo que se procede a descartar al resto de los componentes y a considerar sólo estos.

Tomada la decisión anterior, para calcular el vector \bar{w} de ponderaciones para las variables, se realiza la siguiente operación:

$$\bar{w} = |PC_4|\lambda$$

Donde PC_4 es la matriz 18×4 con los primeros 4 componentes principales y λ es el vector 4×1 de varianzas de cada uno de estos 4 componentes. Con lo que se obtiene un vector $\bar{w}_{18 \times 1}$, con la ponderación de cada variable:

```
##          [,1]
## sin_disc    1.663845
## escolaridad 3.010386
## ocupados    2.150306
## salud       1.674539
## piso        3.055044
## energia     3.273747
## agua        2.959951
## tinaco      2.083660
## escusado    3.284890
## drenaje     3.265727
## refri       2.702096
## lavadora    3.099630
## auto        3.243121
## celular     2.768719
## internet    2.960248
## invfec      2.781231
## invhacin    2.916623
```

Una vez obtenido el vector de ponderaciones, se calcula el promedio ponderado de las observaciones de la matriz X de la siguiente manera:

$$P = \frac{X\bar{w}}{\sum_{i=1}^{18} \bar{w}_i}$$

Donde \bar{w}_i es el elemento i del vector \bar{w} .

2. Cantidad de grupos: Para incorporar el elemento de desigualdad en el índice, se emplea la opción de *cantidad de grupos*, que permite dividir a las AGEB seleccionadas, agrupando a cada una dependiendo del cuantil al que pertenece en la distribución de la población, con respecto al promedio ponderado calculado.

De forma detallada, suponiendo que el usuario elige k grupos, el proceso de asignación es el siguiente:

- i) Las AGEB se ordenan de manera ascendente, de acuerdo con el valor del índice.
- ii) Se crea una nueva variable $F_X(x)$, que considera la proporción de la población (mayor a 3 años de edad) con un índice menor o igual que el de cada AGEB.
- iii) Se divide al intervalo $[0, 1]$ en k particiones del mismo tamaño. Generando así los grupos $1, 2, \dots, k-1, k$.
- iv) A cada AGEB se le asigna un grupo i , donde $i = 1, 2, \dots, k$, dependiendo del valor calculado en el **paso ii**; perteneciendo así al grupo cuya cota derecha en la partición es la menor, tal que la fracción que hay entre ésta y 0, es mayor o igual que $F_X(x)$.

3. Filtro geográfico: Así como la clasificación de los promedios depende de la cantidad de grupos, la selección geográfica que realiza el usuario también tiene implicaciones en el grupo al que pertenecerá cada AGEB. Una vez que se filtra cierta región, la división de la población tomará en cuenta sólo a aquella que habite dentro de la misma. Eso implica, por ejemplo, que un AGEB en Zapopan puede pertenecer a un grupo distinto si se toma en cuenta sólo el mismo municipio, el Área Metropolitana de Guadalajara o todo el estado de Jalisco, ya que la agrupación relativizará con respecto a los habitantes del filtro geográfico.

Desigualdad municipal:

Para determinar cuáles son los municipios con mayor desigualdad en Jalisco, se calculó la distancia que hay entre el índice de su AGEB con el mayor valor y aquel en el que se observa el mínimo. De esta manera, el municipio en el que difieren más, sin importar si el valor mínimo es “grande” o “pequeño”, será el más desigual.

Por ejemplo, el municipio con mayor nivel de desigualdad, Zapopan, cuenta con un AGEB con el máximo índice igual a 1.54, mientras que el área con el mínimo resulta en -7.70, por lo que la diferencia o distancia entre estos valores es de 9.24.