

# Clasificación de Unidades Geográficas

## Nota metodológica

### Puntos distantes

La información nacional por entidad federativa y a nivel municipal, puede encontrarse con unidades geográficas que presentan valores alejados del resto, definidos como puntos distantes (PD), los cuales distorsionan la conformación homogénea de grupos o clases de elementos. Ante esta situación se aplica la siguiente metodología para detectar y separar dichas observaciones: Dada una variable  $x$ , entonces  $x_i$  es un punto distante, si se cumple que

$$| x_i - \text{med}(x) | > 3 \text{ mad}(x) / 0.6745$$

Donde:  $i = 1, 2, 3, \dots, n$  ( $n = 32$  para entidades federativas, y a nivel municipal  $n =$  número de municipios en la entidad);  $\text{med} =$  mediana de  $x$ ;  $\text{mad} = \text{med} ( | x_i - \text{med}(x) | )$ .

Puede darse el caso de que se obtengan dos grupos de PD (el que corresponde a los valores más altos y el de valores más bajos). En tal situación, con el resto de los valores se formarán las clases faltantes.

Adicionalmente, suele haber unidades geográficas cuyo indicador de interés presenta valores cero, éstas se agrupan en una clase, ya que podrían ser objeto de un análisis especial.

### Método de estratificación

Existen varios métodos de estratificación que, según el método seleccionado, dependen principalmente del número de observaciones, su ordenamiento, la distribución de la población, el número de estratos y de una partición inicial de entrada. Dados los grandes volúmenes de información que se manejan en el Instituto y apelando a las mejores prácticas tanto en su interior como en el ámbito internacional, uno de los métodos que proporcionan resultados muy satisfactorios es el de las K-medias, bajo ciertas variantes y condiciones que previamente define el usuario o estrategia. Este método requiere como parámetro un número de estratos determinado por el propio usuario, o bien, calculado automáticamente mediante algún procedimiento estadístico definido; por lo general, no se recomienda construir más allá de seis o siete grupos, ya que se vuelve difícil su interpretación en gráficas o mapas, además de que casi ya no hay ganancia en la homogeneización de las unidades de observación.

## Número de estratos

La determinación del número de estratos no es una práctica sencilla, se requiere de un análisis previo sobre el comportamiento de las unidades ante la variable o variables consideradas, pues se corre el riesgo de forzar el número natural de grupos. Para ello también existen diversas técnicas reportadas en la literatura, ninguna de las cuales puede generalizarse. En este producto se utiliza una opción sencilla que funciona razonablemente bien, el coeficiente de determinación  $R^2$ .

El procedimiento consiste en calcular la variabilidad relativa explicada por los estratos:

$$R_i^2 = \frac{SCE_i}{SCT} \times 100$$

Donde:  $SCE_i$  = Suma de cuadrados entre estratos ( $i = 1, 2, 3, \dots, n$ ; siendo  $n$  el número total de unidades geográficas consideradas, que coincide con el número máximo de estratos);  $SCT$  = Suma de cuadrados total.

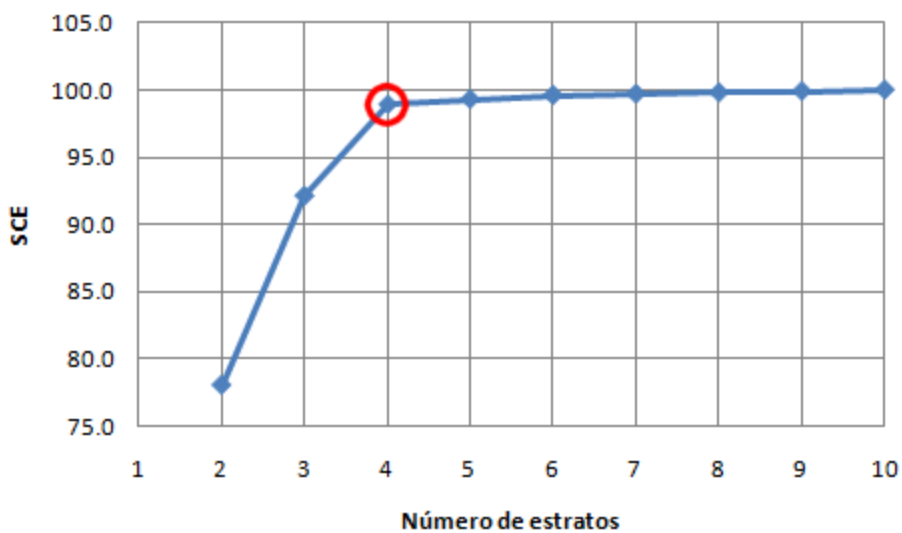
El número adecuado de estratos se obtiene cuando el incremento de dicho coeficiente pierde importancia práctica, esto es:

$$(R_i^2 - R_{i-1}^2) < \epsilon$$

En este caso se toma como  $\epsilon = 5$ , con lo que generalmente se garantiza más del 90% de variabilidad explicada entre los grupos o estratos. El coeficiente de determinación varía entre cero y cien en términos porcentuales; y en la medida en que se acerca a 100, más variabilidad se explica, es decir, los estratos son más diferentes entre ellos y, por tanto, más similares son las unidades en cada estrato, respecto a la variable de interés.

A continuación se presenta un ejemplo a fin de ilustrar el procedimiento. En este caso el número adecuado de estratos es cuatro, lo cual puede confirmarse al observar la distribución de la variable y la gráfica de sedimentación, en donde más allá de este número, prácticamente ya no hay ganancia en la homogeneidad de los estratos.

Variable(X)	Estratos (i)	SCE <sub>i</sub>	R <sub>i</sub> <sup>2</sup>	R <sub>i</sub> <sup>2</sup> -R <sub>i-1</sub> <sup>2</sup>	Óptimo
1	1	0.0	0.0		
2	2	405.6	78.0	78.0	0
3	3	479.1	92.1	14.1	0
8	4	515.1	99.0	6.9	1
9	5	516.6	99.3	0.3	0
10	6	518.1	99.6	0.3	0
15	7	518.6	99.7	0.1	0
16	8	519.1	99.8	0.1	0
21	9	519.6	99.9	0.1	0
22	10	520.1	100.0	0.1	0
<b>SCT total</b>		<b>520.1</b>			



Según sus necesidades, el usuario puede utilizar metodologías alternas, al tener la opción de exportar los datos.