

## Capítulo 2. Procesamiento de los datos

### 1. Introducción

El primer paso en el proceso cartográfico es decidir una jerarquía para las clases a cartografiar, así como el modelo de distribución que se tomará – el conjunto de datos va a concebirse como continuo y uniforme por ejemplo—. El cartógrafo debe utilizar técnicas estadísticas al utilizar escalas nominales, ordinales, de intervalo y de índice en la manipulación de datos. También las utilizará en la selección de las categorías; por ejemplo, la realización de un mapa de suelos exige la elaboración de las categorías modales.

Tras estas decisiones básicas se puede proceder a valorar una serie de cuestiones relacionadas con los datos.

Como se apuntó anteriormente los datos estadísticos a menudo se obtienen a partir de distintas fuentes y es necesario homogeneizarlos de modo que proporcionen valores comparables. Por ejemplo, diversos países utilizan distintas unidades de medidas tales como las métricas, las toneladas largas o cortas, las hectáreas o los kilómetros cuadrados etc.

También con frecuencia es necesario *depurar* las estadísticas de modo que queden eliminados aquellos aspectos que no se desean reflejar, como al preparar por ejemplo un mapa de población rural, donde es necesario eliminar los datos referentes a la población urbana del conjunto de datos total.

En algunos casos, el siguiente paso es convertirlos a datos que puedan utilizarse en la elaboración cartográfica como índices, rendimientos por hectárea, densidades, porcentajes y otras varias, que deben calcularse antes de proceder al diseño del mapa. Asimismo es habitual que posteriormente sea necesario proceder a la clasificación de los datos en intervalos de clase.

### 2. Datos absolutos y datos relativos

Los mapas cuantitativos se hacen representando en ellos valores de dos tipos: absolutos o relativos. Son ejemplos del primer grupo los mapas que muestran la producción o el consumo de bienes, las elevaciones de la superficie terrestre sobre el nivel del mar... Una representación de datos en forma absoluta significa que los valores se muestran tal y como se toman. Las cantidades se observan respecto a un solo tipo de datos, y podemos expresarlas sobre el mapa en términos absolutos, como por ejemplo la producción de energía hidroeléctrica por provincias, o la población por regiones. En ningún caso de los expuestos, los datos expresan relaciones.

Los mapas que representan valores relativos expresan algún tipo de resumen o alguna clase de relación entre dos o más juegos de datos. Entre los ejemplos de este segundo grupo se pueden citar la densidad de población, los ingresos per cápita, la tasa de paro...

En definitiva, se trata de resumir los datos brutos con el fin de facilitar su posterior representación gráfica, así como de proporcionar nuevos datos que resulten esclarecedores en cuanto al tema que tratan, o que expliquen situaciones mediante dicha combinación de

variables. Los tipos de relaciones más utilizadas en cartografía son las que se tratan más adelante bajo los epígrafes de *índices* y *densidades*.

### 3. Promedios

Los promedios constituyen probablemente el tipo más común de variables derivadas empleadas, ya que utilizan una cualidad o cantidad seleccionada para caracterizar una serie de datos que normalmente son numerosos. Existen muchos tipos de promedios, pero en términos generales, en cartografía interesan principalmente tres de ellos que son, la media aritmética, la mediana y la moda (de la que ya se habló anteriormente).

#### La media aritmética

La mayoría de los mapas de clima, ingresos, producción, y otros elementos tratados en el estudio del carácter físico y humano de las regiones se basan en medias aritméticas obtenidas mediante la reducción de grandes cantidades de datos estadísticos.

Sin querer abundar en este concepto, sí es interesante destacar que en cartografía a menudo la media debe considerarse en función de una superficie. Si en un mapa se representa el valor de la tierra de cultivo por hectárea en cada comunidad por ejemplo, éste ha de prepararse a partir de datos obtenidos por promedios regionales, entre las que existirán (grandes) diferencias de tamaño.

Si se da la misma importancia a todas las regiones se falsea el promedio de las comunidades, por lo que siempre que los valores ( $x$ ) en una distribución estén relacionados de cualquier modo con una extensión de superficie, éstos deben considerarse en función de su frecuencia superficial.

La expresión general para cualquier media considerada en función del área es, por lo tanto:  $\bar{x} = \sum x/A$ ; donde  $\sum x$  representa la suma de los productos de cada valor  $x$  por su superficie, y  $A$  es la superficie total. La media considerada en función de la superficie también se denomina media geográfica.

#### La mediana

Si ordenamos todos los valores que toma una variable desde el inferior hasta el superior, la mediana es el valor situado en el centro, de forma que la mitad de los valores serán superiores y la otra mitad inferiores a dicho valor. Es otro tipo de medida relativa de la tendencia central, utilizable por ejemplo en un mapa que represente, en este caso, la mediana de los valores de tierra de cultivo por hectárea en cada provincia, obtenida a partir de los datos municipales.

Al igual que en el caso anterior, si las regiones varían muchísimo en cuanto a extensión, debe considerarse la mediana en función de la superficie. Haciéndolo así, la mediana geográfica será el valor por encima y por debajo del cual se halla la mitad de la superficie total. El valor

de la mediana geográfica es aquel cuya superficie acumulada asociada, resulta igual a la mitad de la superficie total.

## 4. Índices

Otra clase de cantidad relativa es la consistente en medidas como razones o índices, proporciones y porcentajes, en las que algo se mide por unidades de otro elemento, o en las que algún elemento de los datos se individualiza para compararlo con el resto.

Los mapas que representan el porcentaje de días de lluvia, la proporción de ganado vacuno dentro de la ganadería en general, las tasas o índices de mortalidad, o la tasa de crecimiento o decrecimiento de algún fenómeno son algunos ejemplos. En este grupo el valor numérico representado en el mapa será normalmente el resultado de una de las siguientes operaciones:

### Razón o índice

La razón es una expresión de relación entre datos. que se expresa de la forma  $f_a/f_b$ , en donde  $f_a$  es la frecuencia de una clase  $a$ , y  $f_b$  es la frecuencia o número de elementos de otra clase  $b$ .

*Como ejemplo podemos considerar la razón de sexos de una pequeña población. Saber que existen 3000 hombres da una idea al respecto, pero si además se sabe que el número de mujeres es de 1500 la idea cambia y vemos que la razón es de dos hombres por mujer.*

### Proporción

La proporción es la razón entre la frecuencia de una clase ( $f_a$  para la clase  $a$ ) y el total ( $f_a/N$ ) siendo  $N$  la frecuencia total. En el ejemplo anterior, la proporción de hombres es de  $3000/4500$ , es decir de 0,66.

### Porcentaje

Normalmente las proporciones se dan multiplicadas por 100, o lo que es lo mismo en porcentajes. En este caso decir que el 66% de la población son hombres, es más sencillo y fácil de entender que hablar de una proporción de hombres de 0,66.

También estas estadísticas presentan a veces las características de un promedio espacial. Este tipo de razones son la base del concepto de densidad que se trata más adelante. Una razón típica de la geografía es la de densidad de población, definida como el número de habitantes por kilómetro cuadrado o por otra unidad superficial. Si el área de la población anterior es de 10 kilómetros cuadrados, la densidad de población sería de  $4.500/10$ , es decir 450 hab/km<sup>2</sup>.

La cartografía de este tipo de cantidades relativas se elabora para mostrar las variaciones de un lugar a otro de la relación definida, y normalmente se prepara a partir de resúmenes de datos estadísticos. Cuando los porcentajes, razones y tasas se simbolizan en unos límites, el usuario supone que el valor representado se extiende de modo más o menos uniforme a través de dicha unidad.

Si esto no se produce en la realidad, entonces los datos representados pueden resultar un tanto equívocos. Otro tanto sucede si hay muy pocos elementos, de modo que el 100% de niños escolarizados puede ser el resultado de que los diez jóvenes de una región despoblada

asistan a clase. Por otro lado nunca deben servir de base para una razón cantidades no comparables o que carezcan de sentido –dentistas por kilómetro cuadrado–, el sentido común normalmente indicará maneras de asegurar la posibilidad de comparación.

## 5. Densidades

Las densidades se utilizan cuando lo que se quiere reflejar es la acumulación o escasez geográfica relativa de datos discretos. Se calcula dividiendo el dato por la superficie en la que se encuentra, pero en muchos casos este valor no es tan significativo como el que expresa la razón entre otros factores que están más estrechamente relacionados. Por ejemplo, la relación entre número de personas y superficie productiva en sociedades predominantemente agrícolas. Esta relación se considera más útil que lo que lo es el simple cálculo de la población con respecto al área total, productiva o no.

Al trabajar con densidades el cartógrafo está limitado en el detalle que puedan presentar los tamaños de las unidades de enumeración (municipios, regiones o países) en las que se ha elaborado el recuento de elementos, aunque en muchos casos los datos iniciales deben complementarse con los proporcionados por otras fuentes con el fin de presentar una distribución más cercana a la realidad.

Otra categoría de cantidades relativas son los potenciales (o criterio de gravedad), que también se utilizan en mapas, y que suponen que los elementos de una distribución –personas o precios por ejemplo– influyen entre sí directamente con las magnitudes del fenómeno e inversamente con la distancia entre sus ubicaciones. De este modo el valor del potencial en un punto, es la suma de la influencia de todos los demás puntos sobre él, más su propia influencia.

## 6. Clasificación de los datos en intervalos de clase

En páginas anteriores se menciona la repercusión que tiene en un mapa el sistema de clasificación de datos que se haya empleado en su confección, ya que por absolutos o relativos que sean los datos que describen la distribución, a menudo éstos son demasiado numerosos como para ser representados gráficamente uno a uno.

Si se piensa en lo que respecta a las poblaciones de un mapa general, lo normal es que éstas se presenten categorizadas, en *clases* ordenadas y con diferencias cuantitativas expresas, de modo que éstas se agrupan en función de que tengan menos de 25.000 habitantes, o de 25.000 a 100.000 y así sucesivamente hasta alcanzar el máximo, o la ciudad más habitada. En este caso será raro que se represente con signo propio la población con 26.023 habitantes, o la que cuenta con 26.102 etc.

La forma de categorizar las poblaciones anteriores es desde luego arbitraria, en el sentido de que los límites pueden ser cualesquiera, aunque siempre se debe buscar la forma más representativa de mostrar la distribución que se trate.

Este aspecto es especialmente decisivo al emplear la técnica coroplética, donde se verá que hay que resumir en un número de clases reducido el total de los datos que hay que representar.

Son dos las decisiones fundamentales a tomar al clasificar los datos; por un lado el número de clases que se representarán, y por otro los límites de cada uno de los intervalos. Cada subdivisión de datos estadísticos en intervalos de clase puede ser comparada con un proceso de generalización, asumiendo que esta generalización afecta a la superficie estadística correspondiente, uniformizando sus irregularidades.

### *La superficie estadística*

*Una superficie estadística es la superficie formada al asignar a cada punto del territorio  $(x,y)$ , una  $z$  proporcional al valor que toma una variable cuantitativa en dicho punto.*

*Al formar esta superficie, pueden no existir saltos bruscos de modo que tal y como ocurre con la topografía del terreno, el cambio de las diferentes  $z$  es más o menos suave, pero siempre gradual. Así sucede con las temperaturas, precipitaciones, presión atmosférica...*

*Entre dos datos puntuales puede obtenerse el valor correspondiente a un punto intermedio sin más que interpolar. Además existen*

*valores para todos los puntos de la superficie. No existen saltos bruscos, estamos hablando de una superficie estadística suavizada, que será la que corresponda a fenómenos continuos, como la temperatura, las presiones atmosféricas....*

*Por el contrario a distribuciones discretas les corresponderá una superficie estadística escalonada, donde los valores pueden variar de forma brusca, sin que existan valores intermedios. La mayoría de las distribuciones socio-económicas como la densidad de población, las ocupaciones profesionales... son de este tipo y, por continuar con el ejemplo anterior, entre dos puntos de control no tiene por qué existir un valor intermedio.*

La realización de intervalos también puede compararse con la obtención de curvas de nivel que son el resultado de intersectar planos horizontales con un modelo del terreno. Igualmente los intervalos pueden obtenerse intersectando planos horizontales con la superficie estadística que describe una distribución determinada. Todas las  $z$  comprendidas entre dos planos consecutivos, pertenecerán a la misma clase.

No cabe la menor duda acerca de que un sistema de clasificación no adecuado para los datos de partida puede proporcionar un mapa alejado de la realidad, y que incluso puede distorsionarla, por lo que la clasificación final que se realice para un mapa, requiere de un estudio previo orientado a encontrar aquel sistema que se adapte bien a los datos iniciales. Se trata de proporcionar una imagen clara y legible, sin desvirtuar la distribución original.

## El número de clases

El número de clases es función del detalle necesario para mostrar adecuadamente el contenido temático, pero viene limitado por aspectos perceptivos (así como de la técnica a emplear en la representación y el proceso de impresión). Lógicamente con tres intervalos no es posible detallar tanto el comportamiento de la variable en el espacio como lo haríamos con doce, pero elegir un número demasiado alto no está exento de problemas; hay que buscar un punto de equilibrio según cada caso particular.

## Los límites de clase

Una vez decidido el número de clases que se van a emplear en una representación, el siguiente paso es establecer por dónde se realizarán los *cortes* en la distribución. Los sistemas que se pueden emplear son muy numerosos por lo que sólo algunos de ellos se muestran en los ejemplos más adelante.

Como guía general a seguir, se deben buscar límites que reduzcan al mínimo las diferencias entre los datos de una misma clase, y que a su vez hagan máximas las diferencias entre clases. Esto puede estar reñido también con la lectura del mapa, ya que intervalos iguales y con límites redondos siempre son más rápidos de interpretar que otros irregulares y con límites menos memorizables.

## 7. Límites de clase

Es este un tema que se trata de manera diferente según la bibliografía cartográfica que se consulte, donde diferentes autores realizan sus propias clasificaciones para los distintos sistemas de realización de intervalos de clase. Por otro lado, los sistemas de clasificación son tan numerosos que parece necesario categorizarlos de algún modo.

Hay quien distingue entre matemáticos/estadísticos y empíricos (*Gorkin & Gocham 1974*). *Dickinson* habla de 5 formas principales de las que 4 de ellas serían matemáticas/estadísticas. *Robinson (1985)* sólo diferencia tres; intervalos iguales, sistemáticamente desiguales e irregulares, y es el esquema que se presenta en el siguiente apartado. *Wonka (1980)* también habla de tres formas principales, pero subdivide la categoría de los empíricos en *exógenos* y los determinados *en base a su distribución espacial*.

*Evans (1977)* habla de cuatro métodos principales; exógenos, arbitrarios (buscan límites de fácil lectura sin preocuparse de la distribución original de los datos), ideográficos (se basan en detalles específicos del conjunto de datos para representar los puntos de ruptura de la distribución), y por último menciona las series (*intervalos iguales, en progresión...*).

Pasamos ahora a mostrar la clasificación empleada por Robinson en su clásico, *Elementos de Cartografía*, que como se ha indicado distingue tres categorías: Intervalos iguales, sistemáticamente desiguales e irregulares.

## 7.1. Intervalos de clase iguales

### 7.1.1. Iguales según la amplitud de los datos

Consiste en dividir la amplitud máxima existente entre los datos, entre el número de clases que se haya elegido  $(nf-ni)/n$ . Este tipo de clasificación puede ser útil para realizar mapas de isolíneas por ejemplo. Puede proporcionar clases con muchas observaciones y otras con muy pocas o ninguna cuando la distribución se asemeja a una distribución normal. Se adapta mejor a distribuciones rectangulares y es más representativa cuando la amplitud del conjunto de datos es pequeña.

### 7.1.2. Iguales según los parámetros de la distribución normal

Para elegir los intervalos de clase, pueden utilizarse los parámetros de una distribución normal. Basta obtener la media del conjunto de datos y su desviación estándar, que puede ser sumada y restada desde la media (en fracciones o múltiplos). Cuanto más normal sea la distribución de los datos, mejores resultados se obtendrán de la aplicación de este método, que además resulta útil para mostrar los desvíos respecto a dicha media.

También se puede buscar equilibrar la distribución dividiendo los datos por medias sucesivas, de forma que se halla primero la media del conjunto total, y a su vez las medias de los grupos que ésta deja por encima y por debajo. Para cada una de estas cuartas partes se realiza la operación anterior, y así sucesivamente. Este sistema puede ser útil cuando el número de intervalos requerido sea  $2n$ .

### 7.1.3. Iguales según el número de observaciones; cuantiles

Consiste en dividir el número de observaciones en partes iguales al número de clases que queramos. Son habituales los cuartiles (4 clases), los quintiles (5), los septiles (7), los deciles (10). Para obtener cuartiles se ordenan los datos y se dividen en cuatro partes con igual número de observaciones cada uno.

Los percentiles muestran los valores donde se encuentran ciertos porcentajes de casos por encima y por debajo. Así el *percentil 25* es el valor por debajo del cual se encuentra el 25% de los valores observados de una variable, y por encima del que se encuentran el 75% de los valores restantes.

Los *cuantiles de superficie* consisten en intervalos que en el mapa ocupan una cantidad de superficie similar. La superficie total se divide en el número de clases elegido y los límites de cada intervalo se desprenden en función de esta igualdad de superficies. Su ventaja consiste en que en el mapa los colores se distribuyen por igual, pero por otro lado su utilización puede desnaturalizar considerablemente la distribución original y presentar mapas alejados de la realidad que se desea describir.

## 7.2. Intervalos en progresión

Generalmente las series de datos que tienen una amplitud menor causan menos problemas cartográficos que las series que abarcan un rango mayor. En el último caso los intervalos de clase tienen que ser grandes y con ellos no se pueden mostrar detalles en toda la amplitud de datos.

Sin embargo, a menudo se necesita detallar la distribución en los valores más bajos, ya que pequeñas diferencias absolutas, pueden tener una gran importancia relativa, tal y como sucede en la elección de los intervalos de altura para la representación del relieve por medio de tintas hipsométricas. Una forma de detallar más la distribución en los valores más bajos es utilizando series en progresión, donde los intervalos crecen sistemáticamente. Se describen algunas de ellas a continuación.

### 7.2.1. Intervalos en progresión aritmética

En este sistema el tamaño de cada intervalo aumenta progresivamente con un valor constante. Conociendo el número de clases en que se ha de dividir una serie de datos, el cálculo de los intervalos de clase se hace como sigue:  $a+x+2x+3x+4x+\dots+n x = b$ ; donde  $a$  es valor más bajo y  $b$  el valor más alto,  $n$  el número de clases y  $x$  el tamaño del primer intervalo.

Puesto que  $a$ ,  $b$ , y  $n$  son conocidos, el valor del incremento  $x$  se puede despejar, y definir por tanto los intervalos como siguen.

Es posible considerar que el valor del incremento también aumente cada vez, obteniendo una serie en progresión aritmética con un índice en aumento, de forma que el intervalo crezca más rápido.

$a - (a+x)$	<i>diferencia</i> $x$
$(a+x) - (a+3x)$	$2x$
$(a+3x) - (a+6x)$	$3x$
$(a+6x) - (a+10x)$	$4x$

### 7.2.2 Intervalos empleando progresiones geométricas

Este caso es igual que el anterior, sólo que el intervalo va aumentando cada vez siguiendo una progresión geométrica, con un crecimiento más rápido del tamaño del intervalo. Una forma sencilla de emplear progresiones de este tipo es fijar que sean los límites los que respondan a esta clase de progresión. Siendo  $a$  el valor más bajo,  $x$  la razón de la progresión y  $n$  el número total de clases, estas quedan definidas tal y como se muestra abajo.

El mayor valor de la última clase iguala al valor más alto en el conjunto de datos, es decir que  $b = ax^n$ . Así es posible calcular el valor de  $x$ , y con sólo multiplicar obtener los límites de clase.

$a$	-	$ax$
$ax$	-	$ax^2$
$ax^2$	-	$ax^3$
$ax^{n-1}$	-	$ax^n$



### 7.3. Intervalos irregulares

En los sistemas anteriores los límites de clase son impuestos al utilizar la regla matemática del sistema seleccionado, de forma que del cálculo del tamaño de los intervalos se desprenden unos valores, que son los que rompen en el continuo de la distribución, y son los que se emplean después como límites de los intervalos. De algún modo, son unos límites impuestos o forzados.

Sin embargo antes de elegir el sistema de clasificación es sin duda muy útil la observación de los datos, utilizando por ejemplo gráficos, como la curva de frecuencias. Se trata de conocer cómo se comportan los datos, observar qué tendencias siguen, ver si su crecimiento es constante o si hay cambios bruscos en el comportamiento del dato. Nos interesa también saber en qué regiones hay más observaciones, dónde se acumulan y dónde se dispersan.

#### Puntos de ruptura

En la observación anterior quizá se manifiesten claramente los llamados puntos de ruptura naturales de la distribución. Son puntos de ruptura aquéllos que representan puntos significativos de ésta, irregularidades que pueden corresponderse con puntos de inflexión, cambios de pendiente, ausencia del dato...

Una vez determinados estos puntos críticos los intervalos son los que de se desprenden naturalmente. De este modo podrán ser completamente irregulares, por ejemplo: *0-10; 10-25; 25-40; 40-60; 60-75* donde los intervalos de clase son respectivamente de: *10, 15, 15, 20 y 15* (de ahí el incluirlos en este epígrafe de intervalos irregulares) sin atender a ninguna regla lógica.

Si estos límites propios de la distribución son claros y están bien definidos, resulta interesante seleccionar los intervalos empleando como guía dichos puntos de ruptura, procediendo en sentido contrario al del caso de los intervalos de igual tamaño o de los intervalos en progresión: En ellos los límites son consecuencia de la aplicación del tamaño del intervalo, y son ajenos a la distribución. En este otro caso se parte del conocimiento de los límites –que no son ajenos a la distribución– desprendiéndose de ellos los intervalos.

Este tipo de clasificación puede utilizarse para realzar elementos que con otros sistemas pasarían desapercibidos. Para observar la distribución de los eventos en el conjunto de la información, pueden construirse una serie de gráficos que proporcionen una imagen adecuada de la distribución de las características.

Tales gráficos son por ejemplo los gráficos de dispersión, la curva de frecuencias, la curva de frecuencias acumulativas, la curva clinográfica, etc. Todos ellos ayudan a señalar las irregularidades de la distribución de los datos.

Solamente si los gráficos muestran claramente las irregularidades, pueden elegirse sin dificultad los puntos que delimitarán las clases. Muchas veces sin embargo, las irregularidades no son tan importantes o no están tan inequívocamente señaladas, en cuyo caso nos decidiremos por elegir intervalos de clase más sistemáticos, y más sencillos de interpretar.

También es posible emplear técnicas estadísticas en la definición de intervalos según los puntos de ruptura naturales. El llamado *Método de Optimización de Jenks*, es el que implementan diferentes aplicaciones informáticas bajo la opción de realizar intervalos según los puntos de ruptura naturales de la distribución (*natural breaks*). Este método persigue el doble propósito de obtener clases de gran homogeneidad interna, con máximas diferencias entre las clases para el número de intervalos que se haya especificado previamente.

Para ello realiza la clasificación basándose en la prueba de la bondad del ajuste –*Goodness of Variance Fit (GVF)*– que indica cómo de bien describen las clases al conjunto. Dicho indicador toma diferentes valores según los agrupamientos que se hagan de un mismo conjunto de datos, siendo más representativos aquellos agrupamientos que proporcionen los valores más altos. Se trata de un proceso iterativo que calcula la media de cada clase con las respectivas varianzas, y traslada observaciones entre clases hasta obtener el valor máximo del GVF.

### Límites exógenos

En ocasiones resulta interesante incluir como límites de intervalo ciertos valores que siendo ajenos a los datos observados, son significativos para la variable. Se trata de valores importantes para el tema en cuestión, aunque no representen un límite natural en la distribución que corresponda representar. Tienen un significado en sí mismos y son ajenos al conjunto observado, de donde viene su denominación.

En una mapa de renta per cápita por ejemplo, resultará expresivo introducir el valor que se corresponde con el índice de pobreza, o el valor crítico de densidad de población en el caso de un mapa de densidad de población... A menudo existe la posibilidad de delimitar ciertos intervalos con valores exógenos interesantes; éstos deben considerarse antes de realizar la clasificación definitiva, y si en el tema que corresponde describir pueden definirse valores de este tipo, hay que considerar la opción de incluirlos en el mapa.

## 8. Conclusiones

La clasificación de datos en intervalos es un proceso de generalización que afecta a la superficie estadística correspondiente, donde se uniformizan sus pequeñas irregularidades. No obstante se deben intentar mantener los rasgos más destacados de cada distribución, de modo que si entre los datos existe alguno excepcional, éste debe constar como tal en el mapa.

Se debe intentar seleccionar las clases de manera que se mantengan las características más significativas de la distribución, abarcando todo el rango de datos y sin que existan clases vacías. Asimismo resulta conveniente dividir los datos en grupos de números de observaciones razonablemente similares, así como buscar una relación lógica en los tamaños de los intervalos pues facilitará la lectura.

A modo de guía se puede decir que los intervalos iguales según rango son útiles en el caso de que el histograma del conjunto de datos sea más o menos rectangular, lo que no habitual en los fenómenos geográficos, y los corogramas (en el caso de tratarse de mapas de corople-

tas) fueran de tamaños similares. En estos casos la clasificación proporciona un mapa bien ordenado y metódico, de fácil lectura.

La clasificación basada en la distribución normal sólo debería utilizarse cuando la distribución de datos se asemejara a una distribución normal, y es particularmente útil cuando el propósito del mapa es mostrar la desviación respecto a la media del conjunto. En estos casos normalmente las categorías quedan limitadas a seis clases.

Los cuantiles no deberían utilizarse en el caso en que los tamaños de los corogramas varían considerablemente. Para utilizar cuantiles de superficie, precisaríamos también de corogramas de tamaños similares.

Una progresión aritmética será adecuada cuando la gráfica de la distribución del conjunto de datos se asemeje a una progresión aritmética, y lo mismo ocurriría con las progresiones geométricas, que serán aplicables cuando la gráfica tienda a mostrar una progresión geométrica.

Los puntos de ruptura de la distribución nos proporcionarán grupos de valores homogéneos, lo que sin duda es deseable, y su utilización será interesante cuando dichas rupturas queden claramente definidas. Además este sistema puede utilizarse junto con algún otro sistema de clasificación.

Vistos los distintos tipos de clasificación para un mismo conjunto de datos, sólo queda ver cuál de ellos es el que mejor se adecúa a la distribución real de los mismos. Para ello se puede construir la superficie estadística correspondiente a la distribución real y las correspondientes a las clasificaciones. La que parecida al modelo real es la correspondiente a la mejor clasificación para el conjunto de datos.

*A continuación se muestran los efectos que tendría la aplicación de algunos sistemas de clasificación de datos, sobre un supuesto mapa de coropletas, donde los datos de partida varían de cero a cien. Con esto se pretende mostrar la realización práctica de intervalos de clase, y más aún dejar claro su manifestación directa en los mapas. Se elige una base cartográfica de delimitación provincial y la realización de cuatro intervalos de clase.*

## Los Datos

1	Lleida	1	37	Almería	29,
2	Castellón	1,	3	Vizcaya	2 ,9
3	Baleares	,2	39	Salamanca	2 ,4
4	Girona	7 ,2	4	Jaén	2 ,1
5	Soria	77,2	41	Melilla	25,
6	Teruel	7 ,1	42	Cáceres	23,4
7	Huesca	74,9	43	Asturias	21,4
8	Navarra	73,3	44	León	2 ,4
9	Tarragona	71,5	45	Ceuta	2 ,3
10	Alava	9,2	4	Málaga	19,9
11	Zaragoza	,	47	Granada	1 ,
12	Rioja	7,5	4	Badajoz	1 ,9
13	Cuenca	1,1	49	Sevilla	15,2
14	Segovia	,2	5	Córdoba	9,1
15	Guipúzcoa	57,4	51	Huelva	,3
16	Guadalajara	57,1	52	Cádiz	
17	Lugo	55			
18	Toledo	52,2			
19	Burgos	49,			
20	Avila	49,4			
21	Barcelona	47			
22	Zamora	4 ,9			
23	Coruña (A)	45,			
24	Alicante	44,7			
25	Madrid	43,4			
26	Palencia	41,			
27	Palmas (Las)	3 ,7			
28	Murcia	37,7			
29	Cantabria	3 ,5			
30	S.C.Tenerife	35,4			
31	Pontevedra	34,5	máx		1
32	Valladolid	33,7	mín		
33	Ourense	33,	media	54, 9	44444
34	Ciudad Real	32,	dev.tip	17,7	1 9
35	Valencia	32	mediana	5 ,9	
36	Albacete	31,4			

## Cuatro intervalos de 0 a 100

### INTERVALOS IGUALES

#### Rango de los datos

$$\Delta = \frac{\text{Máx} - \text{Mín}}{\text{n}^\circ \text{ Interv.}}$$

clase 1 0 a 25 <sup>(11)</sup>  
 clase 2 25 a 50 <sup>(23)</sup>  
 clase 3 50 a 75 <sup>(9)</sup>  
 clase 4 75 a 100 <sup>(9)</sup>

#### Parámetros dist.normal



$< a - \sigma$  clase 1 0 a 37,12 <sup>(24)</sup>  
 $-\sigma$  a  $\bar{X}$  clase 2 37,12 a 54,89 <sup>(11)</sup>  
 $\bar{X}$  a  $\sigma$  clase 3 54,89 a 72,66 <sup>(9)</sup>  
 $> a \sigma$  clase 4 72,66 a 100 <sup>(8)</sup>

utilizando la desviación x 1.5 (26,66)

$< a - \sigma$  clase 1 0 a 28,23  
 $-\sigma$  a  $\bar{X}$  clase 2 28,23 a 54,89  
 $\bar{X}$  a  $\sigma$  clase 3 54,89 a 81,52  
 $> a \sigma$  clase 4 81,52 a 100

#### Medias Anidadas



clase 1 0 a 31,35 <sup>(25)</sup>  
 clase 2 31,35 a 54,89 <sup>(11)</sup>  
 clase 3 54,89 a 71,12 <sup>(7)</sup>

## Cuartiles

52 observaciones  
4 clases > clases de 13

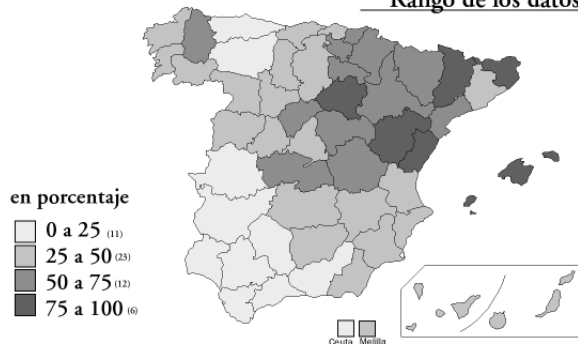
clase 1 0 a 28,1  
 clase 2 28,4 a 38,7  
 clase 3 41,8 a 60,2  
 clase 4 61,1 a 100

## Superficies Iguales

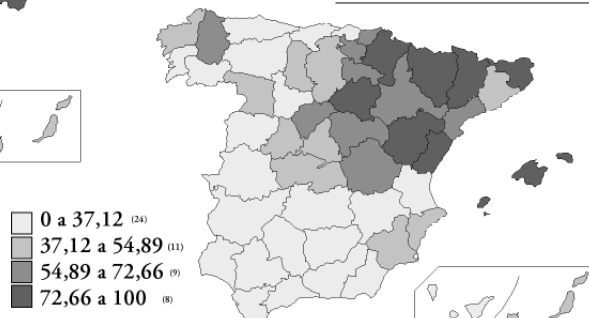
Cada clase cubre aproximadamente la cuarta parte

### INTERVALOS IGUALES

#### Rango de los datos



#### Parámetros dist.normal



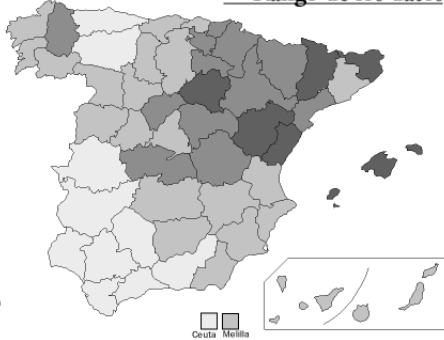
**PROVINCIAS / 4 intervalos / Rango de 0 a 100**

**INTERVALOS IGUALES**

Rango de los datos

en porcentaje

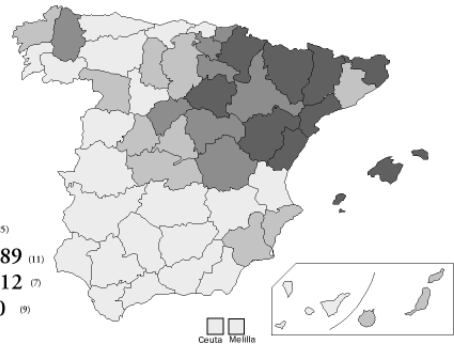
- 0 a 25 <sup>(11)</sup>
- 25 a 50 <sup>(23)</sup>
- 50 a 75 <sup>(12)</sup>
- 75 a 100 <sup>(6)</sup>



Medias anidadas

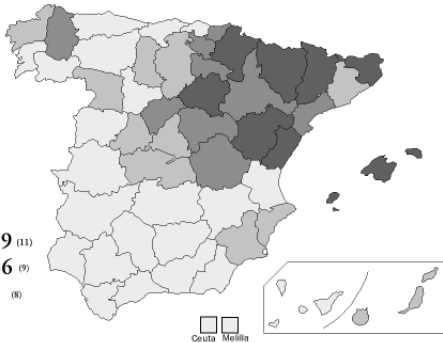
en porcentaje

- 0 a 31,35 <sup>(25)</sup>
- 31,35 a 54,89 <sup>(11)</sup>
- 54,89 a 71,12 <sup>(7)</sup>
- 71,12 a 100 <sup>(9)</sup>



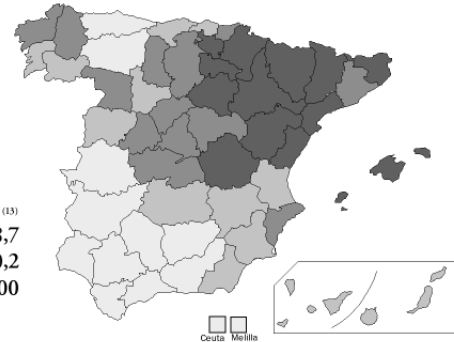
Parámetros dist.normal

- 0 a 37,12 <sup>(24)</sup>
- 37,12 a 54,89 <sup>(11)</sup>
- 54,89 a 72,66 <sup>(9)</sup>
- 72,66 a 100 <sup>(8)</sup>

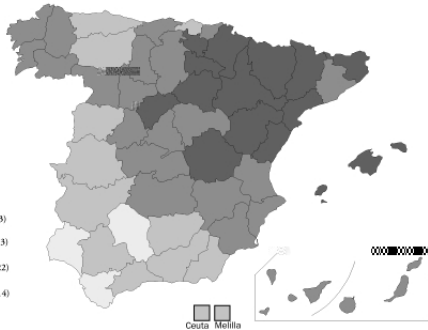


Cuartiles

- 0 a 28,1 <sup>(13)</sup>
- 28,4 a 38,7
- 41,8 a 60,2
- 61,1 a 100



- 0 a 10 <sup>(3)</sup>
- 10 a 30 <sup>(13)</sup>
- 30 a 60 <sup>(22)</sup>
- 60 a 100 <sup>(14)</sup>



**PROVINCIAS / 4 intervalos / Rango de 0 a 100 SISTEMATICAMENTE DESIGUALES**

Prog. Aritmética incremento constante

$$A + d + 2d + 3d + \dots + nd = B$$

$$d = \frac{B - A}{1+2+3+\dots+n}$$

$$d = \frac{100 - 0}{10} = 10$$

Intervalo	Límites
d	A a (A+d)
2d	(A+d) a (A+3d)
3d	(A+3d) a (A+6d)
4d	(A+6d) a (A+10d)

intervalos	clases
10	0 a 10
20	10 a 30
30	30 a 60
40	60 a 100

**PROVINCIAS / 4 intervalos / Rango de 0 a 100 SISTEMATICAMENTE DESIGUALES**

Límites en Prog. Geométrica

a, ax, ax<sup>2</sup>, ax<sup>3</sup>, ..., ax<sup>n</sup>

Límite superior x veces el inferior

siendo 'a' el mínimo distinto de cero

$$B = ax^n$$

$$a - ax$$

$$ax - ax^2$$

$$ax^2 - ax^3$$

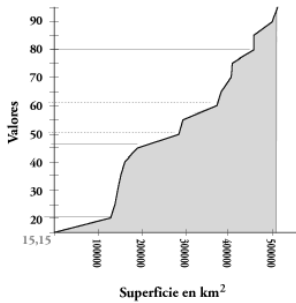
$$ax^3 - ax^4$$

$$x = \sqrt[n]{B/a}$$

$$x = \sqrt[4]{100} = 3,16228$$

Clases	Intervalos	Clases
0 a 3,16	3,16	0 a 3
3,16 a 10,00	6,84	3 a 10
10,00 a 31,50	21,5	10 a 31
31,50 a 100	68,3	31 a 100

### 3. FRECUENCIAS ACUMULADAS / SUPERFICIES



	sup. (km2)	SupAcumulada
1 ANDALUCIA	15,15	87200
2 ASTURIAS	20,15	41602
3 CANTABRIA	21,4	10545
4 CASTILLA Y LEON	56,5	6899
5 CASTILLA LA MANCHA	37,05	7542
6 CATALUÑA	57,7	11517
7 GALICIA	42,25	29484
8 E. LEÓN	42,25	94193
9 MADRID	43,4	7995
10 E. LA MANCHA	46,92	79250
11 PAIS VASCO	51,8333	7261
12 C. VALENCIANA	52,7666	28505
13 LA RIOJA	57,5	5034
14 ARAGON	73,4	47830
15 NAVARRA	73,5	10421
16 CATALUÑA	74,175	51950
17 BALEARES	80,2	5014
Media	25,8	
Moda	20,3	

Son puntos críticos los inferiores y superiores de los grandes 'escarpes' sobre la curva, que determinan superficies de valores semejantes.

### 4. METODOS ITERATIVOS

Establecer algún criterio estadístico lógico de manera que los datos se agrupen de la manera más próxima al criterio establecido: p.ej. minimizar las desviaciones ajustadas alrededor de la media de cada clase (Jenks).

- Iguals
- Desv. estándar
- Aritmética
- Geométrica
- Cuantiles
- Puntos de Ruptura\*

