

## CAPÍTULO 3

# LA UNIFORMIDAD EN LA DISTRIBUCIÓN ALTIMÉTRICA DEL TERRENO

### 1. EL CONCEPTO DE “COEFICIENTE DE UNIFORMIDAD ALTIMÉTRICA”

La uniformidad es una magnitud que caracteriza a la distribución de cualquier variable altimétrica por la superficie de un terreno determinado. **Por ello, hemos creído conveniente ampliar el concepto de “uniformidad altimétrica” al estudio del comportamiento de ciertas variables topográficas.** Los nuevos coeficientes de uniformidad aquí propuestos se podrán utilizar, indistintamente, para la *evaluación* altimétrica de los terrenos analizados.

Por lo que se refiere a los antecedentes, veamos que ya en el libro de J. M. FRANQUET titulado *Análisis territorial (División, organización y gestión del territorio)*, citado en la bibliografía, y concretamente en su capítulo 12 (“Uniformidad y equilibrio del territorio”), se propone y define el concepto de “coeficiente de uniformidad territorial” como medida de la uniformidad de la distribución de una variable socioeconómica por un cierto territorio, precisamente de sentido contrario a su grado de variabilidad. Pues bien, creemos que una extensión de dicho concepto a la evaluación de la distribución de las variables altimétricas resulta perfectamente posible y provechosa para el análisis y conceptualización de los terrenos a explicar.

El proceso de cálculo que aquí se propone comienza con la determinación del coeficiente de variación (CV) de Pearson<sup>1</sup> (que, como es sabido, trátase de una medida abstracta de dispersión relativa de los valores de la variable aleatoria estadística, profusamente utilizada) de las cotas de un terreno determinado; de hecho, en el capítulo siguiente de “Análisis estadístico y uniformidad altimétrica” lo hemos aplicado a un caso concreto. Es obvio que el terreno en cuestión se hallará tanto más equilibrado desde el punto de vista, por ejemplo, de la distribución de las

---

<sup>1</sup> **Karl Pearson** (Londres 27 de marzo de 1857- Londres, 27 de abril de 1936) fue un prominente científico, matemático, historiador y pensador británico, que estableció la disciplina de la *estadística matemática*. Desarrolló una intensa investigación sobre la aplicación de los métodos estadísticos en la biología y se le considera el fundador de la bioestadística. Fue un positivista radical, en la tradición de Berkeley y Ernst Mach. Era partidario de la eugenesia y un protegido y biógrafo de Sir Francis Galton. En 1911 fundó el primer departamento de estadística en la Universidad de Londres, donde fue profesor y donde dirigió el *Laboratory of National Eugenics* creado por Sir Francis Galton. Fundó en 1902 la revista *Biometrika*, desde entonces una de las más importantes en el campo de la estadística.

cotas taquimétricas, cuanto menores sean los valores de su CV (“coeficiente de variación” de Pearson) referido a la variable “cota”. Destaca, del coeficiente elegido como medida de la variabilidad, su adimensionalidad, es decir, su independencia de las unidades de medida, permitiendo establecer la comparación entre terrenos diferentes, lo que no resulta posible efectuar mediante el exclusivo empleo de la varianza o de su raíz cuadrada: la desviación típica o “standard” de la correspondiente distribución de frecuencias.

Al respecto, y como medida de la uniformidad en la distribución de las cotas taquimétricas o de cualquier otra variable topográfica en un terreno determinado, pueden utilizarse los diversos coeficientes que propondremos a continuación (expresados en %), de sentido contrario a la variabilidad antedicha.

El primero de ellos podría ser el siguiente:

**$CU_1 = 100(1 - CV)$** , de gran sencillez y aplicabilidad, siendo:  $CV = \sigma/\bar{X}$ , en que  $\bar{X}$  es la media aritmética de los valores de la variable analizada (altitud o cota taquimétrica) y  $\sigma$  es su desviación típica o “standard” (desviación cuadrática media).

El significado físico del CV se deduce claramente si aceptamos que todos los valores de la variable “cota”, o cualquier otra suficientemente significativa elegida para su aplicación, se distribuyen de acuerdo con la curva campaniforme de una distribución normal (véase la posterior figura 2 y el anexo nº: 3) y, por lo tanto, se tendrá lo siguiente en base a los conceptos ya señalados en el epígrafe inicial del mencionado anexo:

a) Prácticamente, todos los valores observados se hallarán comprendidos en el entorno:  $(1 \pm 3 \cdot CV)\bar{X}$ .

b) Aproximadamente, el 95% de las observaciones se encuentran comprendidas en el entorno:  $(1 \pm 2 \cdot CV)\bar{X}$ .

c) Si se toman las  $n/4$  observaciones de valores más bajos del total de los  $n$  valores medidos de la variable en cuestión (cuyo valor superior será el primer cuartil  $Q_1$  de la distribución de frecuencias), su media aritmética será igual a:  $q_{25} = (1 - 1'27 \cdot CV)\bar{X}$ .

d) El 68'27% de las observaciones realizadas estarán comprendidas en el intervalo:  $(1 \pm CV)\bar{X}$ .

Otros coeficientes de uniformidad topográfica podrían definirse a partir de las siguientes expresiones:

$$CU_2 = (Q_1/\bar{X}) \times 100 \text{ (de menor aplicabilidad) y } CU_3 = (q_{25}/\bar{X}) \times 100,$$

siendo  $q_{25}$ , como ya se ha visto, el valor medio del cuarto inferior de los valores de la variable topográfica analizada.

En relación a la uniformidad topográfica a la que nos venimos refiriendo en el presente epígrafe, veamos que la propiedad más interesante de la distribución normal de los valores de la variable topográfica analizada es que si se toma el 25% de los valores más bajos, su valor medio, es decir, lo que hemos denominado  $q_{25}$ , valdrá, según se deduce del estudio de la distribución normal:

$$q_{25} = (1 - 1'27 \cdot CV) \cdot \bar{X} \quad ,$$

con lo que el coeficiente de uniformidad  $CU_3$  anteriormente definido, tomará el valor:

$$CU_3 = 100 (1 - 1'27 \cdot CV) < CU_1$$

Si suponemos, v.gr., un cierto terreno en el que analizando la distribución de las cotas taquimétricas obtenemos un  $CV = 0'32$ , veamos que:

$$\begin{aligned} CU_1 &= 100 (1 - 0'32) = 68'00\% \\ CU_3 &= 100 (1 - 1'27 \times 0'32) = 59'36\% \quad , \end{aligned}$$

aunque dependería de las circunstancias el escoger uno u otro índice para la medida de la uniformidad topográfica (altimétrica) que se analiza, lo que constituye una responsabilidad del topógrafo experimentador o de la reglamentación que pudiera elaborarse al respecto. De hecho, el  $CU_3$  siempre ofrecerá, expresado en %, por su propia configuración analítica, valores absolutos más bajos que el correspondiente  $CU_1$ , tanto si se trata de valores positivos como negativos (véase, al respecto, el gráfico de la figura 3). Por otra parte, según se deduce del estudio ya realizado de la distribución normal, se cumplirá que:  $Q_1 = (1 - 0'68 \cdot CV) \cdot \bar{X}$ , que será el intervalo correspondiente al 50% de los casos o "rango intercuartílico" ( $Q_3 - Q_1$ ) de la distribución de probabilidad, con lo que también:

$$CU_2 = 100 (1 - 0'68 \cdot CV),$$

que, lógicamente, será el mayor de los cuatro coeficientes de uniformidad topográfica aquí definidos (ver figura 3).

Así pues, y en base a dichos coeficientes, resulta un  $\overline{CU}$  (medio) de :  $Z = -0'9375$  (media aritmética), o bien  $Z = -0'9117$  (media geométrica), por lo que podríamos considerar, como medida "standard" de la uniformidad de un colectivo cualquiera, un  $\overline{CU} = 100 (1 - 0'92 \cdot CV)$ , cuyo intervalo, bajo la hipótesis de normalidad en la distribución de los valores de la

variable topográfica analizada, abarcaría un 64'24% de los casos, según puede comprobarse mediante las tablas de las figuras más completa de áreas y ordenadas bajo la función normal, que adjuntamos en el anexo nº: 3 de este mismo libro (ver figuras A-3.7., A-3.8. y A-3.9 del mencionado anexo).

En el ejemplo anteriormente propuesto, se tendrá:

$$\left\{ \begin{array}{l} CU_2 = 100 (1 - 0'68 \times 0'32) = 78'24\% \\ \overline{CU} = 100 (1 - 0'92 \times 0'32) = 70'56\% \end{array} \right.$$

pudiendo, en la práctica, escoger cualquiera de ellos como medida de la uniformidad topográfica que deseamos realizar.

## 2. OTROS COEFICIENTES DE UNIFORMIDAD TOPOGRÁFICA

### 2.1. Basados en la desviación media absoluta

La "desviación media" es la media aritmética de las desviaciones absolutas de los  $n$  valores de la variable topográfica analizada respecto a un promedio cualquiera. Si tomamos, como dicho promedio, la media aritmética o esperanza matemática  $\bar{X} = \alpha$ , su expresión será, en el caso de una distribución de frecuencias unitarias:

$$DM = \frac{|x_1 - \bar{X}| + |x_2 - \bar{X}| + \dots + |x_n - \bar{X}|}{n} = \frac{\sum_{i=1}^n |x_i - \bar{X}|}{n}$$

(este valor resultaría mínimo si en vez de considerar la  $\bar{X}$  hubiéramos tomado la mediana  $M_e = Q_2$  o valor central de la correspondiente distribución de frecuencias).

Este valor estadístico no es de mucha utilidad en Estadística debido a que no es fácil manipular dicha función al no ser derivable. Siendo más formales, veamos que la "desviación media" debería llamarse "desviación absoluta respecto a la media", para evitar confusiones con otra medida de dispersión, la "desviación absoluta respecto a la mediana",  $DMe$ , cuya fórmula es la misma, sustituyendo la media aritmética  $\bar{x}$  por la mediana  $Me$ . Pero tal precisión no resulta relevante, porque la desviación absoluta respecto a la mediana es de uso todavía menos frecuente.

Por otra parte, en el caso de operar con frecuencias agrupadas o conjuntas, lo que sucederá cuando se opte por agrupar los valores de la variable topográfica analizada por intervalos de clase, se tendrá que:

$$n_1 + n_2 + \dots + n_h = \sum_{i=1}^h n_i = n$$

$$DM = \frac{|x_1 - \bar{X}| \cdot n_1 + |x_2 - \bar{X}| \cdot n_2 + \dots + |x_h - \bar{X}| \cdot n_h}{n} =$$

$$= \frac{\sum_{i=1}^h |x_i - \bar{X}| \cdot n_i}{n} = \sum_{i=1}^h |x_i - \bar{X}| \cdot f_i$$

Pues bien, en base a ella, podríamos definir el siguiente nuevo coeficiente de uniformidad:

$$CU_4 = 100 (1 - DM/\bar{X}),$$

que, en realidad, resulta similar al  $CU_1$ , habiendo substituido la desviación típica o "standard" por la desviación media absoluta, como medida absoluta de dispersión, por el terreno analizado, de los valores de la variable aleatoria topográfica (cota taquimétrica, volúmenes de desmote o de terraplén, ...). Normalmente, para un mismo terreno, se cumplirá que:

$$CU_3 < CU_1 < CU < \overline{CU}_4 < CU_2 < CU_5$$

estando los valores de todos estos coeficientes de uniformidad limitados o acotados superiormente en el 100%, según podrá comprobarse, de modo gráfico, en la figura 3.

## 2.2. Basados en otras medidas de dispersión y concentración

### 2.2.1. Índice de Gini y curva de Lorenz

Teóricamente, la distribución perfecta de la variable altimétrica tendrá lugar cuando, por ejemplo, en un determinado terreno, todos sus puntos tengan la misma cota taquimétrica, lo que podría constituir un *desideratum* ideal en el caso de buscar una nivelación horizontal cuasi-perfecta (por ejemplo, para un campo de fútbol o una parcela arrozal) pero, en cualquier caso, presenta una medida de la uniformidad en la distribución de las cotas taquimétricas por dicho terreno. En este caso, al representar los porcentajes acumulados de las cotas frente a los porcentajes acumulados de las estacas o vértices, se obtendrá la recta de ecuación:  $q_i = p_i$ , coincidente con la bisectriz del primer cuadrante, y el índice de GINI valdrá 0. Obviamente, este índice se encuentra más próximo a 1 cuanto peor está distribuida, por el terreno, la variable topográfica que estamos evaluando.

En los libros de A. PULIDO SAN ROMÁN<sup>2</sup> y de A. ALCAIDE INCHAUSTI<sup>3</sup>, podemos encontrar presentaciones diferentes de la medida que hemos empleado para parametrizar la concentración de las cotas o alturas en el ejemplo práctico desarrollado al respecto: el índice de GINI. Para interpretar correctamente su significado, resulta suficiente con observar que  $G$  varía entre los valores extremos 0 y 1, tomando el valor mínimo o nulo cuando cada  $p_i$  es igual a su correspondiente  $q_i$ , lo que provoca la anulación del numerador de su expresión definitoria; es decir, cuando cualquier porcentaje de puntos del terreno posee un porcentaje igual de cota taquimétrica sobre el global. O bien, dicho de otra manera,  $G = 1$  tendría lugar en el supuesto teórico o hipotético de que todas las  $q_i$  fuesen nulas, excepto la última o  $k$ -ésima (correspondiente al último intervalo de clase considerado) que concentraría todas las cotas de los puntos del terreno que nos ocupa, lo que señalaría la menor uniformidad u homogeneidad en la distribución posible.

Todos estos conceptos pueden precisarse mucho mejor representando en un diagrama la función:  $p_i = f(q_i)$ , o bien su inversa:  $q_i = \varphi(p_i)$ , que permite obtener una línea poligonal construida por encima (o por debajo) de la diagonal de un cuadrado que tiene un extremo en el centro u origen de coordenadas cartesianas rectangulares y el otro extremo en el punto de coordenadas (100,100). Esta figura, denominada CURVA DE LORENZ, frecuentemente usada en el Análisis estructural económico, pondrá de manifiesto una distribución de las cotas o alturas más equitativa en la medida que la línea poligonal resultante (que tenderá a convertirse en una curva al aumentar el número de puntos en estudio) se sitúe más próxima a la citada diagonal (o bien  $G$  más próximo a 0) y también recíprocamente<sup>4</sup>.

Otra forma de observar la curva de Lorenz es estimando el área de la superficie que se encuentra comprendida entre la curva y la diagonal del primer cuadrante (la recta  $p = q$ ). Esa superficie se denomina **área de concentración**.

El **índice Gini** constituye un índice de concentración de los valores de la variable aleatoria estadística o topográfica que estamos manejando y equivale al doble del área de concentración. Su valor estará siempre comprendido entre cero y uno.

---

<sup>2</sup> *Estadística y Técnicas de Investigación Social*. Ed. ANAYA. Madrid, 1971, pág. 111.

<sup>3</sup> *Estadística Económica*. Ed. SAETA. Madrid, 1973, pág. 294.

<sup>4</sup> Así pues, cuanto más pequeña sea el área rayada comprendida entre la curva de Lorenz y la diagonal del primer cuadrante del círculo, mejor será también la distribución de la variable topográfica (cota taquimétrica) que es objeto de nuestro estudio.

Se considera, en definitiva, que existe equidistribución de las cotas del terreno cuando  $p_i = q_i$ , y en esta circunstancia, la expresada curva de Lorenz descansa sobre la diagonal antedicha, o sea, se trata de una recta que descansa sobre la bisectriz del primer cuadrante del círculo. En un caso general adopta, por ejemplo, la configuración gráfica siguiente:

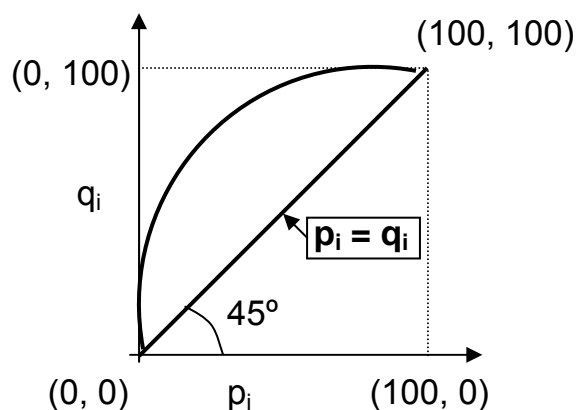


Fig. 1. Curva poligonal de Lorenz.

### 2.2.2. Índice de Williamson

Por otra parte, en el mismo orden de ideas, juzgamos recomendable la utilización, a los efectos de medir el grado de concentración/dispersión de la variable topográfica en estudio, del denominado "índice de Williamson", que nos ofrecerá una buena información en cuanto al nivel de agrupación de los valores de la variable aleatoria estadística ( $q_i$ ) en relación al valor central o media de la correspondiente distribución de frecuencias.

En el caso que se desarrolla extensamente en el capítulo siguiente de nuestro libro, con 15 vértices o estacas, la fórmula pertinente vendrá dada por la expresión:

$$W = \sqrt{\frac{\sum_{i=1}^{15} (x_i - \bar{X})^2 \times \frac{n_i}{n}}{\bar{X}}}, \forall i \in (1, 2, \dots, 15)$$

con el siguiente significado:

$x_i$  = cotas taquimétricas observadas de los puntos del terreno.  
 $n_i$  = número de puntos por cota altimétrica (en nuestro caso,  $\forall n_i = 1$ ).  
 $\bar{X}$  = media aritmética de las cotas del terreno.  
 $n$  = número total de puntos del terreno analizado.

### 2.2.3. Índice de concentración de Lorenz

A mayor abundamiento, desarrollaremos el cálculo de este nuevo índice desde el mismo diagrama o curva que hemos propuesto anteriormente. Tal como se ha venido considerando, se obtendrán siempre curvas cóncavas hacia las  $y$  positivas, y que se hallan situadas por debajo de la diagonal del cuadrado que pasa por el origen de coordenadas y por el punto (100,100).

Así pues, tendremos:

$$L = \frac{(a - q_1) + (2a - q_2) + \dots + [(n - 1)a - q_{n-1}]}{a + 2a + \dots + (n - 1)a} \quad (1),$$

donde  $a$  es la media aritmética de los porcentajes de las cotas de los diferentes puntos del terreno correspondientes a cada intervalo de clase, o sea:

$$X_i = \frac{x_i \cdot n_i}{\sum_{i=1}^n x_i \cdot n_i} \times 100 \quad ; \quad a = \frac{\sum_{i=1}^n X_i}{n} = \frac{q_n}{n} \quad ;$$

De esta manera, se cumplirá también que:

$$\begin{aligned} q_1 &= X_1 \\ q_2 &= X_1 + X_2 \\ q_3 &= X_1 + X_2 + X_3 \\ q_4 &= X_1 + X_2 + X_3 + X_4 \\ q_5 &= X_1 + X_2 + X_3 + X_4 + X_5 \\ &\dots\dots\dots \\ q_n &= X_1 + X_2 + \dots + X_n \end{aligned}$$

o sea:  $q_i = \sum_{j=1}^i X_j$

que es justamente el criterio que hemos seguido en todo momento para la elaboración de la tabla correspondiente. **Debe tenerse bien presente que, en este caso, la ordenación de los valores de las  $X_j$  es preciso realizarla de menor a mayor.**

Desarrollando la expresión anterior (1), obtendremos:



$$\begin{aligned}
 L &= \frac{a + 2a + \dots + (n-1)a - (q_1 + q_2 + \dots + q_{n-1})}{a + 2a + \dots + (n-1)a} = \\
 &= 1 - \frac{q_1 + q_2 + \dots + q_{n-1}}{a + 2a + \dots + (n-1)a} = 1 - \frac{\sum_{i=1}^{n-1} q_i}{\frac{n(n-1)}{2}a} = \\
 &= 1 - \frac{2}{n-1} \times \frac{\sum_{i=1}^{n-1} q_i}{q_n} \quad ,
 \end{aligned}$$

ya que:  $1 + 2 + \dots + (n-1) = n \cdot (n-1)/2$  , dado que se trata de la adición de los  $(n-1)$  primeros términos consecutivos de una progresión aritmética de razón igual a la unidad (demostrable por inducción completa<sup>5</sup>), y además:  $n \cdot a = q_n$ , por la propia definición que hemos considerado de la media aritmética **a**.

Veamos, entonces, los valores que adopta este nuevo índice en los casos extremos posibles. Efectivamente, **si la concentración de las cotas o alturas del terreno es máxima**, tendremos que:

$$X_1 = X_2 = \dots = X_{n-1} = 0, \quad \text{y también : } q_n = \sum_{i=1}^n X_i$$

$$L = 1 - \frac{2}{n-1} \times \frac{\sum_{i=1}^{n-1} q_i}{q_n} = 1 - \frac{2}{n-1} \times \frac{0}{q_n} = 1 \quad ,$$

dado que:  $\sum_{i=1}^{n-1} q_i = 0$  .

Sin embargo, **si la concentración de dichas cotas es mínima**, o sea, la distribución de la misma variable altimétrica es teóricamente perfecta desde el punto de vista estadístico (no necesariamente debe suceder esto en la realidad, sino más bien acontecerá muy raramente, habida cuenta de las irregularidades propias del relieve del terreno en estudio), se tendrá lo siguiente:

$$X_1 = X_2 = \dots = X_n = a,$$

$$\sum_{i=1}^{n-1} q_i = \frac{n(n-1)}{2} a$$

<sup>5</sup> Tipo de razonamiento, opuesto a la **deducción** que partiendo de enunciados particulares, concluye uno de extensión general. Si se consideran todos los casos particulares que incluye el universal, la **inducción** se llama completa; si sólo algunos se llama **inducción** incompleta. La demostración por inducción completa es, en realidad, el desarrollo de un proceso de deducción. El nombre que le damos se debe a una similitud aceptada con los procesos de inducción propios de las ciencias naturales.

en cuyo caso, el índice de concentración de Lorenz será:

$$L = 1 - \frac{2}{n-1} \times \frac{\sum_{i=1}^{n-1} q_i}{q_n} = 1 - \frac{2}{n-1} \times \frac{n \cdot (n-1) \cdot a}{2 \cdot n \cdot a} = 1 - 1 = 0$$

De hecho, estos valores extremos del índice analizado se corresponden con similares valores del índice de Gini anteriormente estudiado. Podemos ver que, cuando:  $L = 0$  ( $X_1=X_2=...=X_n=a$ ), sucede justamente que:  $q_n = n \cdot a$ , razón por la cual la curva pertinente es el segmento recto coincidente con la diagonal del cuadrado al que nos hemos referido con anterioridad. En el caso de la concentración máxima, resulta:  $L = 1$  ( $X_1=X_2=...=X_{n-1}=0$ ), y la curva poligonal de Lorenz, que constituye un triángulo rectángulo, viene dada por los dos lados normales o perpendiculares del cuadrado construido al objeto de trazar el diagrama en cuestión. Obviamente, cuanto más se aproxime la curva a la diagonal relacionada, más perfecta será -al menos desde el punto de vista estadístico- la distribución de la variable topográfica en estudio. Incluso podemos dar una interpretación geométrica del índice de Lorenz de esta manera: el numerador de la fórmula (1) se puede considerar como la adición de las áreas de  $(n-1)$  rectángulos de base igual a la unidad y altura:  $(h \cdot a - q_h)$ ,  $\forall h \in [1, 2, \dots, (n-1)]$ . El denominador, en este caso, es la suma de las áreas de  $(n-1)$  rectángulos de base unidad y altura:  $h \cdot a$ ,  $\forall h \in [1, 2, \dots, (n-1)]$ . Si observamos lo que representa la suma de estos rectángulos, deduciremos que el numerador de la expresión (1) es el área comprendida entre la curva poligonal de Lorenz y la diagonal del cuadrado, mientras que el denominador es precisamente el área de la mitad de dicho cuadrado<sup>6</sup>.

Este índice es equivalente al anteriormente estudiado de Gini y obliga a la realización del cálculo de la superficie rayada de la figura, comprendida entre la diagonal y la correspondiente curva o poligonal de Lorenz. Un valor aproximado es el que se obtiene mediante la aplicación de la fórmula basada en los porcentajes acumulados, que resulta muy empleada en los trabajos prácticos. En nuestro caso, como se puede ver en el ejemplo desarrollado al efecto en el capítulo siguiente de este mismo libro, la anterior fórmula (1) tomará la configuración simplificada (con  $n = 15$  y  $q_n = 100$ ):

$$L = 1 - \frac{2}{14} \times \frac{\sum_{i=1}^{n-1} q_i}{100} = 1 - \frac{\sum_{i=1}^{14} q_i}{700}$$

<sup>6</sup>Así pues, también el índice de concentración de Lorenz será tanto más pequeño cuanto menor sea el valor del área limitada por la diagonal del primer cuadrante y la misma curva poligonal.

### 3. OTRAS CARACTERÍSTICAS INTERESANTES DE LA DISTRIBUCIÓN DE LAS VARIABLES TOPOGRÁFICAS

#### 3.1. Ecuaciones de ligadura entre los diferentes coeficientes de uniformidad

Las medidas de tendencia central ofrecen una idea aproximada del comportamiento de una serie estadística. No obstante, no resultan suficientes para expresar plenamente sus características: una misma media puede provenir de valores cercanos a la misma o resultar de la confluencia de datos estadísticos enormemente dispares. Para conocer en qué grado las medidas de tendencia central son representativas de la serie, se han de complementar con medidas de dispersión como la varianza, la desviación media con respecto a un promedio o la desviación típica.

Normalmente, en Topografía puede revestir interés ocuparse de la dispersión de la distribución de las variables topográficas, es decir, si los datos aparecen sobre todo alrededor de la media o si están distribuidos por todo el rango. Una medida bastante usada de la dispersión es la diferencia entre dos percentiles  $P_r$ , por lo general entre el 25 y el 75, que equivale al rango intercuartílico. El percentil  $r$  es un número tal que un  $r$  por ciento de los datos son menores o iguales que  $r$ . En particular, los percentiles 25 y 75 se denominan cuartiles inferior o primer cuartil ( $Q_1$ ) y superior o tercer cuartil ( $Q_3$ ) respectivamente. La desviación típica o "standard" es otra medida absoluta de la dispersión, pero resulta más útil su empleo que los percentiles, pues está definida en términos aritméticos.

Las medidas de centralización ayudan, en definitiva, a determinar el «centro de gravedad» de una distribución estadística. Para describir el comportamiento general de la serie se necesita, sin embargo, una información complementaria para saber si los datos están dispersos o agrupados. Así, las **medidas de dispersión** pueden definirse como los valores numéricos cuyo objeto es analizar el grado de separación de los valores de una serie estadística con respecto a las **medidas de tendencia central** consideradas.

Las medidas de dispersión, básicamente, son de dos grandes tipos:

- **Medidas de dispersión absoluta:** como recorrido, desviación media, varianza y desviación típica, que se usan en los análisis estadísticos generales.

- **Medidas de dispersión relativa:** que determinan la dispersión de la distribución estadística independientemente de las unidades en que se exprese la variable. Se trata de parámetros más técnicos y utilizados en

estudios específicos, y entre ellas se encuentran los coeficientes de apertura, el recorrido relativo, el coeficiente de variación (índice de dispersión de Pearson), el índice de dispersión mediana y el coeficiente de uniformidad aquí definido.

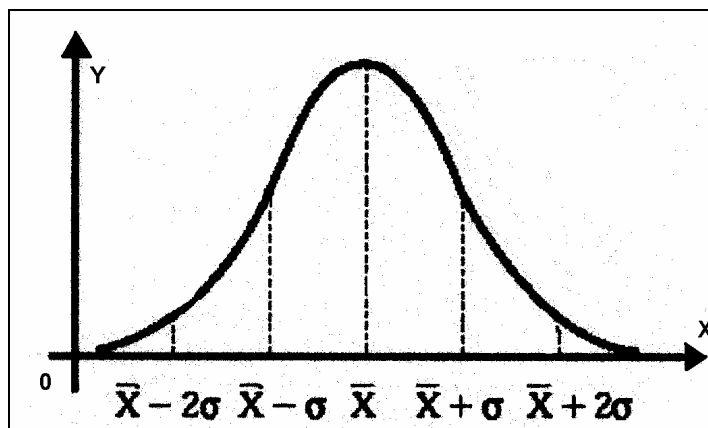


Fig. 2. Curva de distribución normal.

La distribución de probabilidad normal, o campana de Gauss, es una función simétrica con respecto a la ordenada o frecuencia máxima (con la media aritmética y la moda en el centro de la serie) con un grado de dispersión bajo, dado que la mayoría de los valores de la variable topográfica en estudio estarán comprendidos dentro del entorno cuyo centro es la media aritmética y de radio la desviación típica, como puede comprobarse de la contemplación del gráfico anterior.

Evidentemente, existen en la metodología estadística otras medidas del grado de concentración y/o dispersión de las variables topográficas que pueden emplearse eficazmente en la medida de la uniformidad altimétrica (como, por ejemplo, el recorrido "semi-intercuartílico", el "coeficiente de apertura", el "recorrido relativo", etc...), debiéndose tener en cuenta que, para distribuciones moderadamente asimétricas, se pueden aplicar, con buena aproximación, las fórmulas empíricas siguientes (donde  $Q_1$  y  $Q_3$  son, respectivamente, el primer y tercer cuartil de la correspondiente distribución de frecuencias):

$$DM \approx (4/5) \cdot \sigma \quad ; \quad (Q_3 - Q_1)/2 \approx (2/3) \cdot \sigma$$

, que no son más que consecuencias directas del hecho de que, para distribuciones normales, se tiene que la desviación media absoluta DM y el "rango semiintercuartílico" son, respectivamente, iguales a 0'7979 y 0'6745 veces la desviación típica o "standard"  $\sigma$ .

Desde esta perspectiva, y para distribuciones aproximadamente normales con suficiente número de valores de la variable topográfica en

estudio ( $n \geq 30$ ), los coeficientes de uniformidad definidos en el presente libro pueden representarse, geoméricamente, por rectas o funciones lineales cuya variable independiente o explicativa sea el coeficiente de variación de Pearson CV. En concreto, se tendrá que:

$$CU_4 = 100 (1 - 0,7979 \cdot s/\bar{X}) \approx 100 \cdot (1 - 0,80 \cdot CV)$$

La representación gráfica resultante será la siguiente:

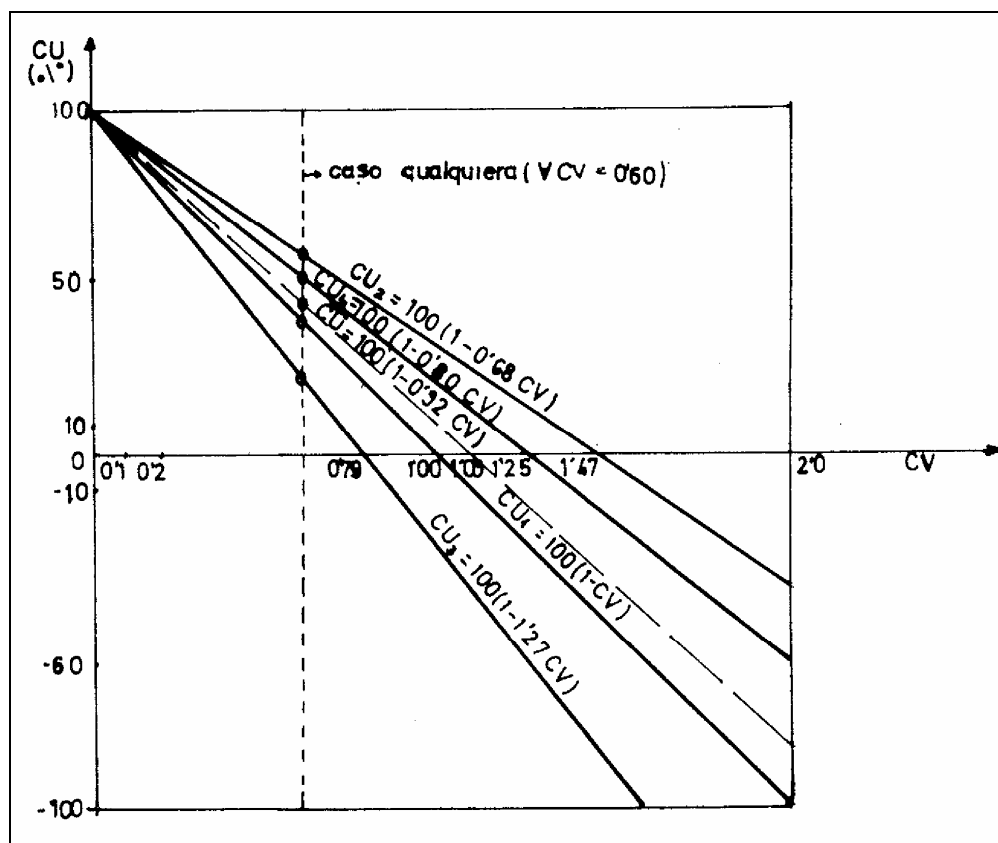


Fig. 3. Coeficientes de uniformidad en función del coeficiente de variación de Pearson.

A su vez, las relaciones que ligan entre sí los diferentes coeficientes de uniformidad topográfica aquí definidos, pueden deducirse del siguiente modo:

$$\begin{aligned} CU_1 &= 100 (1 - CV) = 100 - 100 \cdot CV \\ CU_2 &= 100 (1 - 0,68 \cdot CV) = 100 - 68 \cdot CV \\ CU_3 &= 100 (1 - 1,27 \cdot CV) = 100 - 127 \cdot CV \\ CU_4 &= 100 (1 - 0,80 \cdot CV) = 100 - 80 \cdot CV \\ \overline{CU} &= 100(1 - 0,92 \cdot CV) = 100 - 92 \cdot CV \end{aligned}$$

De donde se tiene que:

$$\begin{array}{rcl}
 CU_1 - CU_3 & = & 100 - 100 \cdot CV - 100 + 127 \cdot CV = 27 \cdot CV \\
 CU_3 - CU_4 & = & 100 - 127 \cdot CV - 100 + 80 \cdot CV = -47 \cdot CV \\
 \hline
 CU_1 - CU_4 & = & \dots\dots\dots(27 \cdot CV - 47 \cdot CV) \dots\dots\dots = -20 \cdot CV
 \end{array}$$

Se tendría que:

$$\begin{array}{l}
 CU_1 / CU_3 = (1 - CV) / (1 - 1,27 \cdot CV) ; \\
 CU_1 - 1,27 \cdot CV \cdot CU_1 = CU_3 - CV \cdot CU_3 ; \\
 CU_1 - CU_3 = 27 \cdot CV = 1,27 \cdot CV \cdot CU_1 - CV \cdot CU_3 ; \\
 27 = 1,27 \cdot CU_1 - CU_3 ; \quad CU_3 + 27 = 1,27 \cdot CU_1 ; \text{ con lo que:} \\
 \mathbf{CU_1 = (CU_3 + 27) / 1,27}
 \end{array}$$

Así mismo:

$$\begin{array}{l}
 CU_1 / CU_4 = (1 - CV) / (1 - 0,8 \cdot CV) ; \\
 CU_1 - 0,8 \cdot CV \cdot CU_1 = CU_4 - CV \cdot CU_4 ; \\
 CU_1 - CU_4 = -20 \cdot CV = 0,8 \cdot CV \cdot CU_1 - CV \cdot CU_4 ; \\
 -20 = 0,8 \cdot CU_1 - CU_4 ; \quad CU_4 - 20 = 0,8 \cdot CU_1 ; \text{ y entonces:} \\
 \mathbf{CU_1 = (CU_4 - 20) / 0,8}
 \end{array}$$

Si observamos la representación gráfica adjunta de la posterior figura 4, la convergencia de ambas rectas tendrá lugar para los valores:

$$(CU_3 + 27) / 1,27 = (CU_4 - 20) / 0,8 \quad \text{y} \quad CU_3 = CU_4$$

, lo que implica que, en dicho punto, tendrá lugar la máxima uniformidad altimétrica posible, con:

$$\mathbf{CU_1 = CU_3 = CU_4 = 100\% = CU_2 = \overline{CU}}$$

También se cumplirá que:

$$\begin{array}{l}
 CU_3 / CU_4 = (1 - 1,27 \cdot CV) / (1 - 0,8 \cdot CV) ; \\
 CU_3 - 0,8 \cdot CV \cdot CU_3 = CU_4 - 1,27 \cdot CV \cdot CU_4 ; \\
 CU_3 - CU_4 = -47 \cdot CV = 0,8 \cdot CV \cdot CU_3 - 1,27 \cdot CV \cdot CU_4 ; \\
 1,27 \cdot CU_4 - 47 = 0,8 \cdot CU_3 ; \text{ y entonces:} \\
 \mathbf{CU_3 = (1,27 \cdot CU_4 - 47) / 0,8}
 \end{array}$$

Las relaciones existentes entre los diferentes coeficientes de uniformidad topográfica anteriormente definidos, de utilidad en los procesos de cálculo, y que ya han sido expresadas analíticamente, pueden verse gráficamente a continuación:

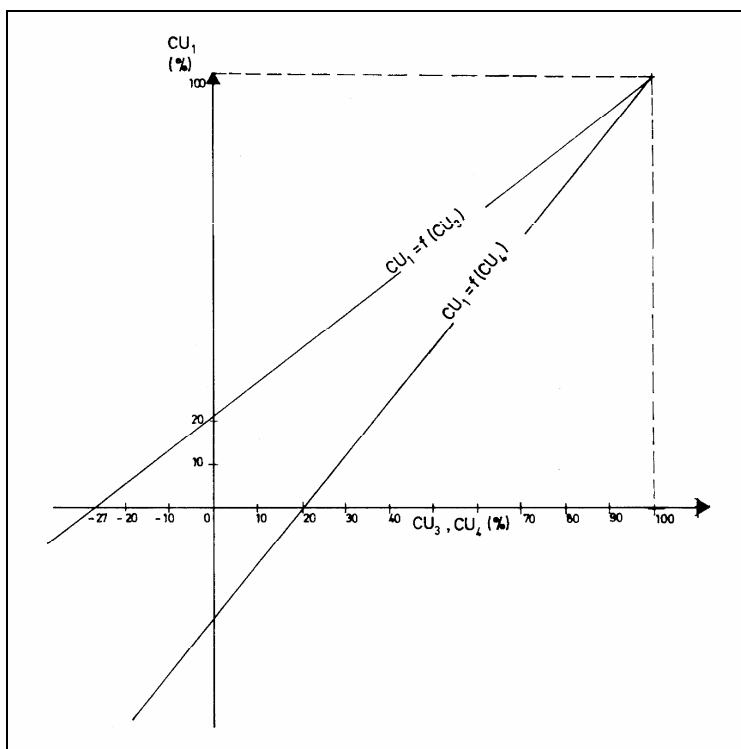


Fig. 4. Relaciones existentes entre los diferentes coeficientes de uniformidad, para distribuciones topográficas aproximadamente normales (I).

El gráfico anterior puede complementarse, por su elevado interés práctico, con el siguiente, que relaciona el coeficiente de uniformidad  $CU_3$  con el  $CU_4$  del siguiente modo:

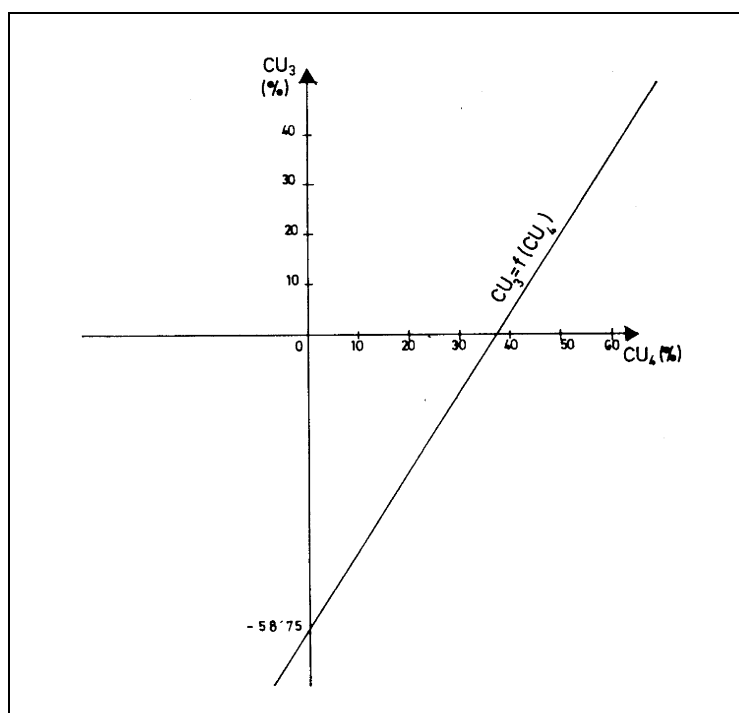


Fig. 5. Relaciones existentes entre los diferentes coeficientes de uniformidad, para distribuciones topográficas aproximadamente normales (II).

De las expresiones anteriores, se deducen inmediatamente las otras tres siguientes, hasta completar las seis relaciones posibles existentes entre los índices topográficos de tal suerte definidos, esto es:

$$\begin{aligned} \text{CU}_3 &= 1,27 \cdot \text{CU}_1 - 27 ; \text{CU}_4 = 0,8 \cdot \text{CU}_1 + 20 ; \\ \text{CU}_3 &= (0,8 \cdot \text{CU}_3 + 47) / 1,27 \end{aligned}$$

Idénticas consideraciones podríamos realizar respecto a  $\text{CU}_2$  y a  $\overline{\text{CU}}$  en relación con los tres restantes coeficientes de uniformidad topográfica, y cuya materialización brindamos, como ejercicio recapitulatorio, a nuestros amables lectores.

### **3.2. Agrupamiento en "clases" y otras características de las distribuciones topográficas**

#### *3.2.1. Los intervalos de clase*

Así mismo, cuando el número  $n$  de los valores de la variable topográfica analizada sea grande, lo que tendrá lugar en aquellos casos de levantamiento de terrenos suficientemente extensos y/o en que se haya tomado un número elevado de medidas, resultarán poco manejables las tablas estadísticas que recojan todos los valores con sus correspondientes frecuencias. En tales casos, se agruparán los valores de la variable (por ejemplo "cota") en "clases", que podrán ser de la misma o diferente amplitud; una norma práctica genérica pudiera ser el establecer una misma amplitud equivalente al 10% de la observación mayor, con lo que el número de clases oscilará alrededor de la decena. Cuando esto acontezca, el cálculo de la desviación típica necesaria para el hallazgo de los CV y de los pertinentes coeficientes de uniformidad registrará algo de error, debido, precisamente, al "error de agrupamiento" en clases. Para ajustarnos mejor a la realidad, se utilizará entonces la varianza corregida, ofrecida por la denominada "corrección Sheppard", a saber:

$$\sigma_c^2 = \sigma^2 - C^2/12$$

siendo  $C$  la amplitud del intervalo de clase escogido y  $\sigma^2$  la varianza de los datos agrupados, y ello tendrá lugar en distribuciones continuas donde las "colas" van gradualmente convergiendo a 0 en ambas direcciones ( $-\infty$  y  $+\infty$ ).

En líneas generales, veamos que un número excesivo de "clases" reduce las ventajas de la agrupación, pero también resulta cierto que un número escaso de ellas puede llegar a anular la significación de los datos. Un criterio usado frecuentemente es que el número de clases debe ser



aproximadamente igual a la raíz cuadrada del número de datos. Por ejemplo, la raíz cuadrada de 35 (número de lecturas en el ejemplo desarrollado en el tutorial del capítulo 8) es mayor que cinco, por lo que se podrían seleccionar seis clases, aunque en este caso se haya optado por establecer una distribución de frecuencias unitarias. También la regla de Sturges no es sino una recomendación acerca del número deseable de clases que deben considerarse a la hora de elaborar un histograma. Éste viene dado por la siguiente expresión:

$$k = \text{número de clases} = 1 + 3.3 \cdot \log_{10} n,$$

siendo  $n$  el tamaño del colectivo. Así, por ejemplo, en un levantamiento altimétrico de una extensa zona constituido por  $n = 1.000$  observaciones, con esta fórmula se aconsejaría el establecimiento de:

$$1 + 3.3 \times \log. 1.000 = 10.9 \approx 11 \text{ clases.}$$

Respecto a la amplitud de las "clases" establecidas, conviene observar que, en general, es conveniente que sea la misma para todas; sin embargo, esto dependerá mucho de los propios datos y del objetivo final de la distribución de la variable en estudio. En principio, si la distribución es más o menos uniforme, todas las "clases" serán de igual amplitud, y si, por el contrario, presenta grandes oscilaciones, puede ser interesante considerar intervalos de amplitud diferente.

De hecho, la construcción de una distribución numérica -como la mayoría de las que elaboraremos aquí- consta de tres etapas fundamentales: 1) determinar las "clases" con sus intervalos más procedentes, tal como ya hemos expresado antes, en las que se han de agrupar los datos de la variable topográfica en estudio, 2) clasificar (o distribuir) los datos en las clases apropiadas, y 3) contar el número de casos de cada clase. Como sea que las dos últimas etapas son puramente mecánicas, así como el establecimiento de la correspondiente "marca de clase" (obtenida normalmente, a falta de más datos, como la semisuma de los valores extremos del intervalo de clase), nos fijaremos sólo en la primera. Por esto, hace falta determinar el número de clases así como la amplitud del intervalo de los valores de la variable aleatoria estadística con la que trabajamos (cota taquimétrica, superficie de parcela, volumen de desmonte y/o terraplén, distancia de transporte, ...). Por esto, en términos generales, se pueden observar al efecto las siguientes normas:

- a) Pocas veces emplearemos menos de 6 ó más de 15 clases; el número exacto de las mismas dependerá de la naturaleza, cuantía e intervalo que cubren los datos.
- b) Siempre escogeremos las clases de manera que todos los datos queden comprendidos.

- c) Se procurará, siempre que sea posible, que todos los intervalos de clase tengan la misma amplitud, lo que obviará la determinación de las “densidades de frecuencia” -que determinan la altura de los rectángulos yuxtapuestos del histograma- para el cálculo de algunas medidas centrales de la correspondiente distribución de frecuencias (como la “moda”) o la representación gráfica de los histogramas.

Veamos, por último, que mediante el razonamiento ya expuesto que sirve para definir la “desviación típica o standard” como una medida de dispersión absoluta de los valores de la variable topográfica, se puede afirmar que si este estadístico resulta pequeño, los valores se encuentran concentrados en el entorno de la media aritmética y, además, si la desviación típica es grande, los valores están mucho más esparcidos o dispersos en relación a los centrales. Para comprender este razonamiento sobre una base algo menos intuitiva, nos referiremos brevemente al importante *Teorema de Tchebyshev*, que expresa que **para cualquier clase de datos (poblaciones o muestras), al menos el 75% de los datos se encuentran sobre el intervalo que se extiende a cada lado de la media aritmética en dos veces el valor de la desviación típica ( $\pm 2\sigma$ )**. Según este teorema, también se puede afirmar que por lo menos el 88'8% de los datos se encuentran dentro del intervalo de tres veces ( $\pm 3\sigma$ ) la desviación típica (a ambos lados de la media aritmética) y que al menos el 96% de los mismos se hallan comprendidos dentro del intervalo de amplitud de cinco veces la desviación típica ( $\pm 5\sigma$ ).

Genéricamente, este teorema indica que cualquiera que sea la forma de una distribución de frecuencia de población, la proporción de observaciones que caen dentro de  $k$  desviaciones estándar de la media es, al menos, de:  $1 - (1/k^2)$ , siempre que  $k$  sea 1 o más.

Precisamente, una característica importante del teorema de Tchebyshev<sup>7</sup> es que resulta válido para cualquier tipo de datos, incluidos los topográficos. No obstante, si se dispone de alguna información adicional en relación a la forma global de la distribución que estamos trabajando, también se pueden realizar afirmaciones mucho más estrictas. Por ejemplo, si una distribución es *campaniforme* o gaussiana, se puede esperar que aproximadamente el 95% de los datos (en lugar de al menos el 75%) se encuentren dentro del intervalo  $\pm 2\sigma$  y el 99% de los datos (en lugar de al menos el 88'8%) se encuentran dentro del intervalo  $\pm 3\sigma$ .

---

<sup>7</sup> Otra aportación importantísima de este mismo autor estriba en la denominada “Desigualdad de Tchebyshev”, que permite generar un intervalo de valores de la variable aleatoria dada la probabilidad de que esto ocurra, o también es posible realizar el hallazgo de la probabilidad de que ocurran los valores que toma la característica aleatoria incluidos en un intervalo dado. En cualquier caso es necesario disponer de la varianza poblacional.

Estos porcentajes, en definitiva, corresponden a la llamada *distribución normal*, que es objeto de estudio en diversas partes de nuestro libro.

### 3.2.2. Forma de la distribución de frecuencias

También existen otras características de menor interés práctico, que tratan de precisar la *forma de la distribución* de la variable que se estudia en relación con una distribución normal. Así, la curva de Gauss sabemos que es simétrica respecto de la ordenada  $x = \alpha$  (el parámetro  $\alpha$  es la media aritmética o esperanza matemática de una distribución normal de frecuencias) y la distribución observada puede ser *asimétrica* o sesgada respecto a la ordenada correspondiente e, incluso, dicha asimetría puede representar una mayor área bajo la curva, a la derecha o a la izquierda de dicha ordenada; por otra parte, a la distribución observada puede corresponderle un área bajo la curva más achatada o más alargada que la correspondiente área de una distribución normal; a tal característica la denominaremos *curtosis* o *medida de apuntamiento*.

Todas estas características, bien conocidas por otra parte, de una distribución de frecuencias originan medidas exactas si están utilizando todos los posibles valores de la variable topográfica, es decir, si corresponden a la población o universo de datos; pero un objetivo esencial de la Inferencia Estadística es el de estimar dichas características poblacionales a partir de una muestra o subconjunto poblacional, lo que en la Topografía aplicada sucederá en numerosos casos. Por ejemplo, cuando establecemos la malla o red para la toma de datos de las cotas de un terreno, estamos analizando sólo una muestra de la totalidad del mismo que deseamos sea lo suficientemente representativa de su relieve y configuración planimétrica. En esta tesitura, las características de la *distribución muestral* suelen ser, en general, los mejores *estimadores* de las características de la distribución de la población, pero las *estimaciones* que originan han de presentarse, como sabemos, acompañadas de los errores de muestreo y de otras medidas de naturaleza probabilística, que permiten apreciar el grado de confianza o fiabilidad de la correspondiente estimación de la característica poblacional, y que no procede analizar aquí con mayor profundidad por comprensibles razones de espacio y oportunidad.

### 3.2.3. Otros coeficientes de uniformidad altimétrica

Veamos, por último, que en base a los mismos o parecidos conceptos, sería posible la definición de otros coeficientes de uniformidad altimétrica. Y así, valga como ejemplo el que tendría en cuenta el valor del primer y tercer cuartil de la distribución de frecuencias de la variable contemplada, a saber:

$$CU_5 = 100 \times \sqrt{\frac{Q_1}{Q_3}}$$

que, en el caso de una distribución moderadamente asimétrica (aproximadamente normal), ofrecerá un valor en función de  $Q_3$  y de  $\sigma$  equivalente a:

$$Q_3 - Q_1 \approx 4\sigma/3 \quad ; \text{ esto es:}$$

$$(Q_3 - Q_1)/Q_3 \approx 4\sigma/3Q_3 \approx 1 - Q_1/Q_3 \quad ; \text{ de donde:}$$

$$\sqrt{\frac{Q_1}{Q_3}} = \sqrt{1 - \frac{4\sigma}{3Q_3}} \quad , \text{ con lo que:}$$

$$CU_5 = 100 \cdot \sqrt{\frac{Q_1}{Q_3}} = \sqrt{10.000} \cdot \sqrt{1 - \frac{4\sigma}{3Q_3}} = \sqrt{10.000 - \frac{40.000 \cdot \sigma}{3Q_3}}$$

Puede resultar de interés, en fin, el análisis de la distribución de probabilidad “gamma”, puesto que son numerosas las aplicaciones de esta función a experimentos o fenómenos aleatorios que tienen asociadas variables aleatorias que siempre resultan ser no negativas (positivas o nulas) y cuyas distribuciones son sesgadas a la derecha (asimetría positiva), es decir, cuya área bajo la función de densidad disminuye progresivamente a medida que nos alejamos del origen de coordenadas. Estas circunstancias suelen presentarse en los trabajos de Topografía; de ahí el interés de su estudio.

