CAPÍTULO 4
ANÁLISIS ESTADÍSTICO Y UNIFORMIDAD ALTIMÉTRICA

1. LA DISTRIBUCIÓN DE FRECUENCIAS

La agrupación tabular de la distribución de frecuencias[1] de una toma o levantamiento de las cotas taquimétricas (absolutas o relativas) de un terreno cuya nivelación se pretende constituye la etapa previa para cualquier otro proceso de análisis estadístico a que se hayan de someter unos valores numéricos que toma cualquier variable topográfica. De esta forma, se puede fácilmente comprender dónde están los promedios o valores centrales de la distribución antedicha y juzgar las características de cualquier nota en relación con las restantes.

Para hacer más comprensible el procedimiento, desarrollaremos el ejemplo de la parcela de terreno estudiado en el epígrafe 2 del capítulo 6 de este mismo libro, para cuya nivelación óptima se tomaron 15 puntos con sus correspondientes cotas taquimétricas relativas.

Para interpretar correctamente dicha tabla, deben formularse previamente las siguientes observaciones:

a) La columna de las frecuencias simples, encabezada por n_i, $\forall i \in (1, 2, \ldots, 15)$, recoge la *frecuencia absoluta* de cada lectura. Normalmente, esa cifra será la unidad en la mayoría de los trabajos topográficos de este tipo, salvo que hallemos varios vértices con la misma cota taquimétrica y los agrupemos, en cuyo caso obtendremos una distribución conjunta de frecuencias. Se tratará casi siempre, pues, de una distribución de frecuencias unitarias, y la frecuencia total tomará, en este caso, el valor $n = 15$.

b) Las *frecuencias ordinarias acumuladas ascendentes* las hemos designado por la notación $N_i\uparrow$ ($\forall i = 1, 2, \ldots, 15$) y a cada una de ellas corresponde el número de vértices del terreno de cota taquimétrica igual

[1] La *frecuencia absoluta* de una variable estadística es el número de veces que aparece en la muestra dicho valor de la variable; la representaremos por $\mathbf{n_i}$. La frecuencia absoluta, es una medida que está influida por el tamaño de la muestra, y al aumentar el tamaño de la muestra aumentará también el tamaño de la frecuencia absoluta. Esto hace que no sea una medida útil para poder comparar distribuciones. Para esto es necesario introducir el concepto de *frecuencia relativa*, que es justamente el cociente entre la frecuencia absoluta y el tamaño de la muestra. La denotaremos por $\mathbf{f_i}$.

o inferior; así, $N_4\uparrow = 4$, significa que se han hallado 4 vértices con una cota taquimétrica relativa igual o inferior a 100’78 m. El complementario $N_i\downarrow = n - N_i\uparrow$ representa las frecuencias ordinarias acumuladas descendentes.

c) Las *frecuencias relativas* ordinarias y acumuladas ascendentes y descendentes encabezadas, respectivamente, por las notaciones f_i, $F_i\uparrow$ y $F_i\downarrow$ vienen determinadas por el cociente de dividir la correspondiente frecuencia absoluta o acumulada (ascendente o descendente) por la frecuencia total.

La suma de las frecuencias relativas ordinarias y la última frecuencia relativa acumulada ascendente han de ser siempre iguales a la unidad, puesto que representan la probabilidad total del espacio muestral. Del mismo modo, la última frecuencia relativa acumulada descendente debe ser igual a 0. Con ello se puede elaborar la siguiente tabla, en que la columna de las cotas taquimétricas iniciales del terreno se ha dispuesto en orden creciente de magnitud, tomando como base o cota relativa inferior (100,00 m.) la del vértice 15 a los efectos de la mayor operatividad que se pretende. Así:

Tabla 1. Tabla de frecuencias.

VÉRTICES	COTA INICIAL (x_i)	n_i	$N_i\uparrow$	$N_i\downarrow$	f_i	$F_i\uparrow$	$F_i\downarrow$
15	100,00	1	1	14	0,067	0,067	0,933
14	100,42	1	2	13	0,067	0,134	0,866
10	100,45	1	3	12	0,067	0,201	0,799
13	100,78	1	4	11	0,067	0,267	0,733
9	100,87	1	5	10	0,067	0,335	0,665
5	100,98	1	6	9	0,067	0,402	0,598
12	101,03	1	7	8	0,067	0,469	0,531
8	101,22	1	8	7	0,067	0,535	0,465
4	101,29	1	9	6	0,067	0,603	0,397
11	101,30	1	10	5	0,067	0,669	0,331
3	101,37	1	11	4	0,067	0,736	0,264
7	101,48	1	12	3	0,067	0,803	0,197
6	101,80	1	13	2	0,067	0,869	0,131
2	101,93	1	14	1	0,067	0,935	0,065
1	102,30	1	15	0	0,067	1,000	0,000
TOTAL		**15**			**1,000**		

De la tabla anterior se deduce también el siguiente gráfico, donde se representan simultáneamente los diagramas acumulativos ascendente y descendente, cuyo punto de intersección debe ser teóricamente la mediana o segundo cuartil de esta distribución de frecuencias, como se verá más adelante.

2. CARACTERÍSTICAS DE LA DISTRIBUCIÓN

2.1. Medidas centrales o promedios

Comenzaremos, como es habitual en estos casos, por el cálculo de los diferentes promedios o valores de la distribución de la variable topográfica "cota taquimétrica". Para ello, y a partir de la tabla inicial, elaboraremos la siguiente:

Tabla 2. Tabla auxiliar de cálculo (I).

VÉRTICES	COTA INICIAL (x_i)	n_i	$x_i \cdot n_i$	$x_i^2 \cdot n_i$	n_i/x_i	log x_i	$n_i \cdot \log x_i$
15	100,00	1	100,00	10.000,0000	0,01000000	2,00000000	2,00000000
14	100,42	1	100,42	10.084,1764	0,00995818	2,00182022	2,00182022
10	100,45	1	100,45	10.090,2025	0,0099552	2,00194994	2,00194994
13	100,78	1	100,78	10.156,6084	0,0099226	2,00337435	2,00337435
9	100,87	1	100,87	10.174,7569	0,00991375	2,00376202	2,00376202
5	100,98	1	100,98	10.196,9604	0,00990295	2,00423537	2,00423537
12	101,03	1	101,03	10.207,0609	0,00989805	2,00445035	2,00445035
8	101,22	1	101,22	10.245,4884	0,00987947	2,00526633	2,00526633
4	101,29	1	101,29	10.259,6641	0,00987264	2,00556657	2,00556657
11	101,30	1	101,30	10.261,6900	0,00987167	2,00560945	2,00560945
3	101,37	1	101,37	10275,8769	0,00986485	2,00590945	2,00590945
7	101,48	1	101,48	10.298,1904	0,00985416	2,00638046	2,00638046
6	101,80	1	101,80	10.363,2400	0,00982318	2,00774778	2,00774778
2	101,93	1	101,93	10.389,7249	0,00981065	2,00830202	2,00830202
1	102,30	1	102,30	10.465,2900	0,00977517	2,00987563	2,00987563
TOTAL		**15**	**1.517,22**	**153.468,930**	**0,14830252**	**30,0742499**	**30,0742499**

Veamos, a continuación, los diferentes promedios o valores centrales analizados:

*** Media aritmética:** a partir de los correspondientes resultados se tiene que, tratándose de una distribución de frecuencias unitarias:

$$\overline{X} = \sum_{i=1}^{h} x_i \cdot n_i / n = \alpha = \frac{1.517'22}{15} = 101'148 \text{ m.}$$

que constituye la media aritmética o esperanza matemática ("cota relativa media") de la distribución de frecuencias de la variable topográfica analizada.

Por otra parte, y por lo que se refiere a la determinación de otros promedios o medidas centrales de esta distribución de frecuencias, veamos que, a partir de la primera columna y de la encabezada por $N_i\uparrow$ (tabla 1 anterior) se obtienen los siguientes resultados promedios:

*** Mediana:** al tratarse de una distribución unitaria con un número impar de vértices, se tendrán, realizando las interpolaciones correspondientes para los cuartiles, los siguientes valores:

$$\begin{cases} Q_0 = 100'00 \text{ m.} \\ Q_1 = 100'80 \text{ m.} \\ Q_2 = Me = 101'22 \text{ m.} \\ Q_3 = 101'40 \text{ m.} \\ Q_4 = 102'30 \text{ m.} \end{cases}$$

Otros promedios interesantes para nuestro estudio serían los siguientes:

*** Media cuadrática:**

$$C = \sqrt{\sum_{i=1}^{15} x_i^2 \cdot n_i / n} = \sqrt{153.468'93 / 15} = 101'150 \text{ m.}$$

*** Media geométrica:**

$$G = \sqrt[n]{\prod_{i=1}^{15} \cdot x_i^{n_i}} = \text{antilog.} \sum_{i=1}^{15} n_i \cdot \log x_i / n =$$

$$= \text{antilog } 30'0742499 / 15 = 101'146 \text{ m.}$$

*** Media armónica:**

$$H = n / \sum_{i=1}^{15} n_i / x_i = 15 / 0'14830252 = 101'145 \text{ m.}$$

Desde luego, como no podría ser de otra manera desde el punto de vista teórico, las cuatro medias aquí estudiadas quedan ordenadas, con arreglo a su magnitud, del modo siguiente:

armónica < geométrica < aritmética < cuadrática

(H = 101'145 m.) < (G = 101'146 m.) < ($\overline{X}$ = 101'148 m.) < (C =101'150 m.)

2.2. Medidas de dispersión o concentración

Las **medidas de dispersión**, también llamadas medidas de variabilidad, muestran la variabilidad de una distribución indicando, por medio de un número, si las diferentes puntuaciones de una variable están muy alejadas de un promedio o no. Cuanto mayor sea ese valor, mayor será la variabilidad; por el contrario, cuanto menor sea, más homogénea será a la media. Así, se sabe si todos los casos son parecidos o bien varían mucho entre ellos. Para calcular la variabilidad que una distribución tiene respecto de su media, se calcula la media de las desviaciones de las puntuaciones respecto a la media aritmética. Pero la suma de esas desviaciones es siempre cero (por una conocida propiedad de la media aritmética), así que se adoptan dos clases de estrategias para salvar este problema. Una es tomando las desviaciones en valor absoluto (obteniéndose la denominada “desviación media”) y otra es tomando las desviaciones al cuadrado (obteniéndose la “varianza” y su raíz cuadrada la “desviación típica”, “desviación standard” o “desviación cuadrática media”).

Por lo que se refiere a la desviación típica o "standard", como medida de la dispersión absoluta de la distribución por el colectivo analizado de nuestra variable topográfica “cota taquimétrica”, veamos que su valor vendrá dado por la expresión:

$$\sigma = \sqrt{C^2 - \overline{X}^2} = \sqrt{101'15^2 - 101'148^2} = 0'636 \text{ m.}$$

y un coeficiente de variación de Pearson (medida de dispersión relativa) de :

CV = $\sigma / \overline{X}$ = 0’636/101'148 = 0’0063 ≈ 0’63%, muy bajo.

Otras medidas interesantes serían:

Recorrido o rango: R = 102’30 – 100’00 = 2’30 m.

Recorrido relativo: R’ = $R / \overline{X}$ = 2’30/101’148 = 0’023

Coeficiente de apertura: $C_{ap} = x_{15}/x_1$ = 102’30/100’00 = 1’023

Recorrido intercuartílico: $Q_3 - Q_1$ = 101’40 – 100’80 = 0’60 m.

Recorrido semi-intercuartílico:

$(Q_3 - Q_1) / (Q_3 + Q_1)$ = 0’60 / (101’40 + 100’80) = 0'003 ≈ 0’3%.

es decir, que en ambos casos, la medida de dispersión empleada viene a representar aproximadamente sólo alrededor del 0’5% del correspondiente promedio. **Desde luego, ello indicaría que el grado de uniformidad y homogeneidad de las cotas del terreno en estudio resulta muy elevado. En efecto, los coeficientes de uniformidad anteriormente definidos, tomarán, en este caso, los siguientes valores:**

$$\begin{cases} CU_1 = 100 \cdot (1 - CV) = 100 \cdot (1 - 0'0063) = 99'37\% \\ CU_3 = 100 \cdot (1 - 1'27 \cdot CV) = 100 \cdot (1 - 1'27 \cdot 0'0063) = 99'20\% \\ CU_4 = 100 \cdot (1 - 0'80 \cdot CV) = 100 \cdot (1 - 0'80 \cdot 0'0063) = 99'50\% \end{cases}$$

debiendo considerarse también que:

$$CU_2 = (Q_1 / \overline{X}) \cdot 100 = (100'80 / 101'148) \cdot 100 = 99'66\%$$

, si bien otra determinación del mismo coeficiente de uniformidad altimétrica conduciría al valor:

$$CU_2 = 100\ (1- 0'68 \cdot CV) = 100\ (1 - 0'68 \times 0'0063) = 99'57\%$$,

cuya pequeña discrepancia (-0'09%) con el resultado anterior débese al propio ajuste de normalidad, o bien al proceso efectuado de cálculo decimal.

Veamos, por último, que el "coeficiente de uniformidad altimétrica medio" de la parcela en estudio, que juzgamos de mejor aplicación, ofrecerá un valor de:

$$\overline{\mathbf{CU}} = 100\ (1 - 0'92 \cdot CV) = 100\ (1 - 0'92 \times 0'0063) = \mathbf{99'42\%}$$,

mientras que también:

$$CU_5 = 100 \cdot \sqrt{\frac{Q_1}{Q_3}} = 100 \cdot \sqrt{\frac{100'80}{101'40}} = 99'70\%$$

La representación gráfica de los valores de los diferentes coeficientes de uniformidad altimétrica hallados, en definitiva, es la siguiente:

Para poder comparar la uniformidad de distintos terrenos a los efectos de su explanación, y poder tomar decisiones comparativas o de priorización respecto de su ejecución, cabe responder a la siguiente pregunta: ¿cuál de ambos, o de varios de ellos, poseen mayor uniformidad taquimétrica y, en su consecuencia, debemos priorizar su explanación, a salvo de otros condicionantes? Para ello compararemos el coeficiente de uniformidad medio obtenido en el presente problema con el que se deduce del ejemplo extraído de una publicación del antiguo Instituto Nacional de Colonización (INC)[2], que podemos encontrar tanto en el posterior capítulo 6 como en el 8 de nuestro libro.

[2] El **Instituto Nacional de Colonización y Desarrollo Rural**, fue un organismo autónomo creado en España en octubre de 1939, dependiente del Ministerio de Agricultura. Su creación estuvo motivada por

Calcularemos ahora el coeficiente de uniformidad medio del ejemplo citado. En primer lugar, habrá que establecer las nuevas cotas relativas de este último terreno sobre la base de otorgar al punto de cota taquimétrica inferior (vértice 11) el valor 100'00 m. para homogeneizar los datos y los resultados subsiguientes. Así:

Tabla 3. Tabla de cotas relativas.

Estacas	Cotas absolutas del terreno inicial	Cotas relativas del terreno inicial (x_i)	n_i	$(x_i)^2 \cdot n_i$
1	23,04	101,43	1	10.288,045
2	25,02	103,41	1	10.693,628
3	26,22	104,61	1	10.943,252
4	22,80	101,19	1	10.239,416
5	25,27	103,66	1	10.745,396
6	25,51	103,90	1	10.795,210
7	22,91	101,30	1	10.261,690
8	24,61	103,00	1	10.609,000
9	26,90	105,29	1	11.085,984
10	26,55	104,94	1	11.012,404
11	21,61	100,00	1	10.000,000
12	23,94	102,33	1	10.471,429
13	23,20	101,59	1	10.320,528
14	22,04	100,43	1	10.086,185
15	30,22	108,61	1	11.796,132
16	29,12	107,51	1	11.558,400
17	26,80	105,19	1	11.064,936
18	22,22	100,61	1	10.122,372
19	24,81	103,20	1	10.650,240
20	24,02	102,41	1	10.487,808
21	22,80	101,19	1	10.239,416
22	31,63	110,02	1	12.104,400
23	28,60	106,99	1	11.446,860
24	24,93	103,32	1	10.675,022
25	23,50	101,89	1	10.381,572

la necesidad de efectuar una reforma tanto social como económica de la tierra, después de la devastación ocasionada por la guerra civil (1936-39). El objetivo principal del mismo era efectuar la necesaria transformación del espacio productivo mediante la reorganización y reactivación del sector agrícola y el incremento de la producción agropecuaria con vistas a los planes autárquicos de la época mediante el aumento de tierras de labor y la superficie de riego. Posteriormente cambió su nombre por el de **Instituto de Reforma y Desarrollo Agrario (IRYDA)** al fusionarse con el Servicio Nacional de Concentración Parcelaria y Ordenación Rural. Para cumplir su función, el Instituto Nacional de Colonización era poseedor de tierras, las cuales eran transferidas en arrendamiento u otras formas de tenencia a los colonos, pequeños productores agropecuarios, quienes debían pagar un canon o arrendamiento hasta que finalmente adquirían la propiedad. El Instituto realizó ambiciosos proyectos de parcelación por toda España, construyendo poblados de colonización al efecto, algunos de los cuales aún subsisten en la actualidad.

26	25,70	104,09	1	10.834,728
27	24,66	103,05	1	10.619,303
28	22,59	100,98	1	10.196,960
29	29,10	107,49	1	11.554,100
30	25,11	103,50	1	10.712,250
31	23,50	101,89	1	10.381,572
32	25,13	103,52	1	10.716,390
33	24,66	103,05	1	10.619,303
34	23,81	102,20	1	10.444,840
35	22,22	100,61	1	10.122,372
TOTAL	**874,75**	**3.618,40**	**35**	**374.281,144**

Para obtenerlo debemos calcular primero las medias aritmética y cuadrática del mismo con el fin de determinar la desviación típica y el coeficiente de variación de Pearson. A saber:

*** Media aritmética:** a partir de los correspondientes resultados se tiene que, tratándose de una distribución de frecuencias unitarias, calcularemos este promedio mediante la fórmula:

$$\overline{X} = \sum_{i=1}^{35} x_i \cdot n_i / n = \alpha = \frac{3.618'40}{35} = 103'383 \text{ m.}$$

que constituye la media aritmética o esperanza matemática de la distribución de frecuencias de la variable topográfica denominada “cota taquimétrica”.

*** Media cuadrática:**

$$C = \sqrt{\sum_{i=1}^{35} x_i^2 \cdot n_i / n} = \sqrt{374.281'144 / 35} = 103'411 \text{ m.}$$

Por lo que se refiere ahora a la desviación típica o "standard", como medida de la dispersión absoluta de la distribución por el colectivo analizado de nuestra variable topográfica “cota taquimétrica”, veamos que su valor vendrá dado por:

$$\sigma = \sqrt{C^2 - \overline{X}^2} = \sqrt{103'411^2 - 103'383^2} = 2'406 \text{ m.}$$

y un coeficiente de variación de Pearson (medida de dispersión relativa) de :

CV = $\sigma / \overline{X}$ = 2’406/103'383 = 0’0233 ≈ 2’33 %, muy bajo.

Veamos, por último, que el "coeficiente de uniformidad altimétrica medio", propuesto en nuestro trabajo, de la parcela en estudio ofrecerá un valor de:

$$\overline{\mathbf{CU}} = 100\ (1 - 0'92 \cdot CV) = 100\ (1 - 0'92 \times 0'0233) = \mathbf{97'86\ \%}\ ,$$

que resulta algo inferior al del primer terreno analizado desde el punto de vista de su uniformidad taquimétrica (97’86% < 99’42%), lo que nos permite un análisis comparativo de la dificultad de explanación del expresado terreno *ex anto*, que puede resultar francamente útil para la toma de cierto tipo de decisiones de gran utilidad en este tipo de trabajos topográficos.

2.3. Otras características de la distribución de frecuencias

2.3.1. Asimetría o sesgo

Las medidas de **asimetría** o **sesgo** son indicadores que permiten establecer el *grado de simetría* (o asimetría) que presenta la distribución de probabilidad de una variable aleatoria sin tener que hacer su representación gráfica.

Como eje de simetría consideramos una recta paralela al eje de ordenadas que pasa por la media aritmética de la distribución. Si una distribución es perfectamente simétrica, existe el mismo número de valores a la derecha que a la izquierda de la media o esperanza matemática, por tanto, el mismo número de desviaciones con signo positivo que con signo negativo.

Decimos que hay asimetría positiva (o a la derecha) si la "cola" a la derecha de la media es más larga que la de la izquierda, es decir, si hay valores más separados de la media a la derecha. Por el contrario, diremos que hay asimetría negativa (o a la izquierda) si la "cola" a la izquierda de la media es más larga que la de la derecha, es decir, si hay valores más separados de la media a la izquierda.

Se podría pensar que definir la simetría usando la mediana para variables continuas y usando la media aritmética para variables discretas es una elección arbitraria. En realidad esto no es así, pues si una variable es continua, coinciden ambos criterios de simetría (con respecto a la media y a la mediana). Es más, se tiene que media y mediana coinciden para las distribuciones continuas perfectamente simétricas. Por otro lado, en el caso de variables discretas, la distribución es simétrica si el lado derecho del diagrama se obtiene por imagen especular desde la media. En este caso coincide la media con la mediana si el número de observaciones es impar. Si la variable es continua simétrica y unimodal, coinciden la media, la mediana y la moda.

Dentro de los diversos tipos de asimetría posible, vamos a destacar los dos fundamentales:

-Asimetría positiva:
Si las frecuencias más altas se encuentran en el lado izquierdo de la media, mientras que en el derecho hay frecuencias más pequeñas (*cola*).

-Asimetría negativa:
Cuando la cola está situada en el lado izquierdo, al contrario de lo que sucede en el caso anterior.

Cuando realizamos un estudio descriptivo es altamente improbable que la distribución de frecuencias sea *totalmente* simétrica. En la práctica diremos que la distribución de frecuencias es simétrica si lo es sólo de un modo aproximado. Por otro lado, aún observando cuidadosamente la gráfica de la distribución, podemos no ver claro de qué lado están las frecuencias más altas. Conviene definir, entonces, unos estadísticos que ayuden a interpretar la asimetría, a los que llamaremos **índices de asimetría**.

A continuación, emplearemos en el ejemplo que estamos desarrollando en el presente capítulo algunos de los índices de asimetría o sesgo más usuales, como son el índice basado en los tres cuartiles, el momento de tercer orden[3] y la distancia existente entre la moda y la media o la media y la mediana.

No se ha podido calcular el denominado “primer coeficiente de asimetría de Pearson” dado que está en función de la moda **Mo** de la distribución, que en este caso no existe, habida cuenta de que todas las variables tienen la misma frecuencia. Por lo que se refiere a las restantes características de la distribución de la variable topográfica “cota taquimétrica”, veamos que una de la asimetría o sesgo la constituye el denominado "2º coeficiente de asimetría de Pearson", que viene expresado por:

$$P_2 = \frac{3(\overline{X} - Me)}{\sigma} = \frac{3(101'148 - 101'22)}{0'636} = -0'34 \approx 0 \quad ,$$

[3] El conocimiento de la simetría permite precisar, de alguna manera, la forma de una distribución de frecuencias, pero tal conocimiento puede mejorarse notoriamente al disponer de otras características de dicha distribución y, de modo general, mediante los denominados “momentos de la distribución”, que pueden ser con respecto al origen (de los que son conocidos la media aritmética y el cuadrado de la media cuadrática C^2) o bien con respecto a la media (como la varianza σ^2).

luego se trata, lógicamente, de una distribución prácticamente simétrica.

También podríamos calcular el "coeficiente de sesgo cuartílico" o “coeficiente de asimetría de Yule-Bowley”, cuyo valor viene dado por la expresión:

$$\frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} = \frac{101'40 - 2 \cdot 101'22 + 100'80}{101'40 - 100'80} = -0'40 \quad ,$$

que conducen, en todos los casos, a conclusiones similares. Cabe observar, no obstante, una ligera asimetría o sesgo hacia la izquierda, puesto que: $P_2<0$ y también:

$$\overline{X} = 101'148 \text{ m.} < Me = 101'22 \text{ m.}$$

Así mismo, a partir de los resultados obtenidos, se elaborará la siguiente tabla, que deberá reunir un número de decimales suficiente para alcanzar la precisión adecuada:

Tabla 4. Tabla auxiliar de cálculo (II).

	n_i	$(x_i - 101'148)$	$(x_i-101'148)^3 \cdot n_i$	$(x_i-101'148)^4 \cdot n_i$	$\lvert x_i - 101'148\rvert \cdot n_i$
	1	-1,148	-1,51295379	1,73687095	1,148
	1	-0,728	-0,38582835	0,28088304	0,728
	1	-0,698	-0,34006839	0,23736774	0,698
	1	-0,368	-0,04983603	0,01833966	0,368
	1	-0,278	-0,02148495	0,00597282	0,278
	1	-0,168	-0,00474163	0,00079659	0,168
	1	-0,118	-0,00164303	0,00019388	0,118
	1	0,072	0,00037325	2,6874E-05	0,072
	1	0,142	0,00286329	0,00040659	0,142
	1	0,152	0,00351181	0,00053379	0,152
	1	0,222	0,01094105	0,00242891	0,222
	1	0,332	0,03659437	0,01214933	0,332
	1	0,652	0,27716781	0,18071341	0,652
	1	0,782	0,47821177	0,3739616	0,782
	1	1,152	1,52882381	1,76120503	1,152
TOTAL	**15**	**0**	**0,02193096**	**4,61185022**	**7,012**

De los resultados que se deducen de la tabla anterior, se infiere que el momento central (respecto a la media aritmética) de tercer orden es el siguiente:

$$m_3 = \frac{\sum_{i=1}^{n}(x_i - \overline{X})^3 \cdot n_i}{n} = \frac{\sum_{i=1}^{15}(x_i - 101'148)^3 \cdot 1}{15} = \frac{0'021931}{15} = 0'001462 \ ,$$

por lo que se tendrá un "coeficiente directo de asimetría" o "coeficiente de sesgo" de Fisher de:

$$g_1 = m_3/\sigma^3 = 0'001462 \ / \ 0'636^3 = 0'00568,$$

que confirma la existencia de una distribución considerablemente simétrica (en curvas simétricas como la normal, se cumple que: $g_1 = g_1^2 = 0$).

2.3.2. Apuntamiento o “curtosis”

En teoría de la probabilidad y estadística, la **curtosis** (o “kurtosis”) es una medida de la forma o apuntamiento de las distribuciones. Así, las medidas de curtosis (también llamadas de apuntamiento o de concentración central) tratan de estudiar la mayor o menor concentración de frecuencias alrededor de la media aritmética y en la zona central de la distribución. De este modo, los índices de curtosis miden la mayor o menor concentración de los datos justamente alrededor de la media aritmética.

Por otra parte, el momento central o respecto a la media aritmética de 4º orden, será:

$$m_4 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^4 \cdot n_i}{n} = \frac{\sum_{i=1}^{15}(x_i - 101'148)^4 \cdot 1}{15} = \frac{4'61185022}{15} = 0'307457 \, .$$

Con ello, se tendrá un “coeficiente de curtosis” de Fisher de:

$$g_2 = \frac{m_4}{\sigma^4} - 3 = \frac{0'307457}{0'636^4} - 3 = -1'805 < 0 \, ,$$

por lo que la distribución que nos ocupa es PLATICÚRTICA O ACHATADA, como se deduce de su propia configuración, en que todas las frecuencias $n_i = 1$.

2.3.3. Otras

Veamos, así mismo, que la anterior tabla auxiliar de cálculo, en su última columna, nos permitirá la determinación de otra medida de dispersión absoluta a la que ya nos hemos referido con anterioridad: la desviación media respecto a la media aritmética (que sería mínima con respecto a la mediana), a saber:

$$DM = \frac{\sum_{i=1}^{15} \left| x_i - \overline{X} \right| \cdot n_i}{n} = 7'012 / 15 = 0'468 \text{ m.}$$

2.4. Índice de Gini y curva de Lorenz

Para la determinación de dichos ítems, frecuentemente empleados en otro tipo de estudios como los sociológicos y/o económicos, y cuya fundamentación teórica y aplicación a los problemas altimétricos ya ha sido explicada en otro apartado de nuestro libro, resulta necesaria la elaboración de la siguiente tabla de cálculos auxiliares:

Tabla 5. Tabla auxiliar de cálculo (III).

VÉRTICES	**COTA INICIAL (x_i)**	**n_i**	**$x_i \cdot n_i$**	**$f_i \cdot 100$ (%)**	$\frac{x_i \cdot n_i}{\sum x_i n_i} \times 100$	**p_i (%)**	**q_i (%)**	**$p_i - q_i$ (%)**
15	100,00	1	100,00	6,667	6,591	6,667	6,591	0,076
14	100,42	1	100,42	6,667	6,619	13,334	13,210	0,124
10	100,45	1	100,45	6,667	6,621	20,000	19,831	0,169
13	100,78	1	100,78	6,667	6,642	26,667	26,473	0,194
9	100,87	1	100,87	6,667	6,648	33,335	33,121	0,214
5	100,98	1	100,98	6,667	6,656	40,002	39,777	0,225
12	101,03	1	101,03	6,667	6,659	46,669	46,436	0,233
8	101,22	1	101,22	6,667	6,671	53,336	53,107	0,229
4	101,29	1	101,29	6,667	6,676	60,002	59,783	0,219
11	101,30	1	101,30	6,667	6,677	66,670	66,460	0,210
3	101,37	1	101,37	6,667	6,681	73,337	73,141	0,196
7	101,48	1	101,48	6,667	6,689	80,004	79,830	0,174
6	101,80	1	101,80	6,667	6,710	86,671	86,540	0,131
2	101,93	1	101,93	6,667	6,718	93,337	93,258	0,079
1	102,30	1	102,30	6,667	6,743	100,000	100,000	0,000
TOTAL		**15**	**1.517,22**	**100**	**100**	**800,031**	**797,558**	**2,473**

Así pues, según la fórmula dada por Pulido, el valor del índice de Gini[4], en este caso, será de:

[4] **Corrado Gini** (23 de mayo de 1884 - 13 de marzo de 1965) fue un estadístico, demógrafo y sociólogo italiano que desarrolló el coeficiente de Gini, una medida de la desigualdad en los ingresos en una sociedad. Gini fue también un influyente teórico fascista e ideólogo que escribió *Las bases científicas del fascismo* en 1927.

$$G = \frac{\sum_{i=1}^{K-1}(p_i - q_i)}{\sum_{i=1}^{K-1} p_i} = \frac{2'473}{700'031} = 0'0035 = 0'35\,\%$$

Habida cuenta del bajísimo valor del índice obtenido, podemos afirmar que la distribución de las cotas taquimétricas del terreno analizado es casi perfecta desde el punto de vista estadístico.

La correspondiente curva poligonal de Lorenz[5] apenas se separa de la diagonal o bisectriz del primer cuadrante en esta ocasión, dado el bajísimo valor del índice de Gini ya obtenido, por lo que obviaremos su representación gráfica, a efectos prácticos. En cualquier caso, como podrán comprobar nuestros amables lectores, ampliamos el concepto acerca de esta importante curva e índice en otros epígrafes de este mismo libro cuya consulta encarecemos.

2.5. Índice de Williamson

En nuestro caso, la variable topográfica en estudio es la cota taquimétrica de los puntos del terreno en cuestión. Este índice, que se acostumbra a emplear con cierto éxito en trabajos económicos, sociológicos y en modelos de desigualdad regional, vendrá dado por la expresión:

$$W = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2 \times f_i}{\overline{X}}}\ ,\ \forall\, i \in (1,2,\ldots,15)$$

O sea:

$$W = \sqrt{\frac{\sum_{i=1}^{15}\left(x_i - 101'148\right)^2 \times f_i}{101'148}} = \sqrt{\frac{0'34583}{101'148}} = 0'0584$$

[5] La **curva de Lorenz** es una representación gráfica utilizada frecuentemente para plasmar la distribución relativa de una variable en un dominio determinado. El dominio constituye una primera variable. La variable cuya distribución se estudia puede ser la variable topográfica que nos ocupa. Utilizando como ejemplo estas variables, la curva se trazaría considerando en el eje horizontal el porcentaje acumulado del dominio en cuestión y en el eje vertical el porcentaje acumulado de la otra variable del problema planteado. Es, en definitiva, una polilínea o línea quebrada que tiende a convertirse en curva cuando el tamaño de la muestra analizada tiende a infinito.

como se deduce de los cálculos de la tabla siguiente y que, como era de esperar, resulta ser muy bajo, al igual que el índice de Gini anteriormente calculado:

Tabla 6. Tabla auxiliar de cálculo (IV).

Vértices	COTA INICIAL (x_i)	n_i	(x_i -101'148)	f_i	(x_i - 101'148)^2	(x_i -101'148)$^2 \cdot f_i$
15	100,00	1	-1,148	0,067	1,317904	0,08829957
14	100,42	1	-0,728	0,067	0,529984	0,03550893
10	100,45	1	-0,698	0,067	0,487204	0,03264267
13	100,78	1	-0,368	0,067	0,135424	0,00907341
9	100,87	1	-0,278	0,067	0,077284	0,00517803
5	100,98	1	-0,168	0,067	0,028224	0,00189101
12	101,03	1	-0,118	0,067	0,013924	0,00093291
8	101,22	1	0,072	0,067	0,005184	0,00034733
4	101,29	1	0,142	0,067	0,020164	0,00135099
11	101,30	1	0,152	0,067	0,023104	0,00154797
3	101,37	1	0,222	0,067	0,049284	0,00330203
7	101,48	1	0,332	0,067	0,110224	0,00738501
6	101,80	1	0,652	0,067	0,425104	0,02848197
2	101,93	1	0,782	0,067	0,611524	0,04097211
1	102,30	1	1,152	0,067	1,327104	0,08891597
TOTAL		**15**	**0**	**1,000**	**5,16164**	**0,34582988**

2.6. Curva e índice de concentración de Lorenz

2.6.1. Otras especificaciones acerca de la curva de Lorenz

En el ejemplo que desarrollamos se considera una población de N = 15 vértices o estacas y la variable estadística $X = \{x_i, n_i\}$, $\forall i = 1, 2, 3..., r$, donde x_i es la cota taquimétrica correspondiente a cada uno de los puntos analizados del terreno, siendo $N = \sum_{i=1}^{r} n_i$. Supongamos además que: $x_1 < x_2 < ... < x_r$. Llamemos $u_i = \sum_{j=1}^{i} x_j \cdot n_j$ es decir, la cota taquimétrica total de los puntos del terreno cuya cota es menor o igual que x_i. Sea $p_i = \frac{N_i^{\uparrow}}{N} \cdot 100$, es decir, el porcentaje de puntos poseedores de la cota total u_i ($N_i^{\uparrow}$ es la frecuencia absoluta acumulada ascendente, es decir, el número de puntos del terreno comprendidos en la cota total u_i); sea $q_i = \frac{u_i}{u_r} \cdot 100$, es decir, el porcentaje de la cota taquimétrica total poseído por los $N_i^{\uparrow}$ puntos o vértices anteriores.

Pues bien, la curva de Lorenz es la línea poligonal quebrada que une los puntos (0, 0), (p_1, q_1), (p_2, q_2), (p_3, q_3), ..., (p_r, q_r). En esta ocasión, se tiene que r = 15.

En nuestro caso, la correspondiente curva poligonal de Lorenz apenas se separa prácticamente de la diagonal o bisectriz del primer cuadrante, dado el bajísimo valor del índice de Gini obtenido (G = 0'0035), como hemos tenido ocasión de comprobar con anterioridad.

También puede resultar de cierto interés, en nuestro caso, la aplicación de la función del económetra italiano Vilfredo Pareto (1848-1923) a la distribución de las cotas taquimétricas del terreno en cuestión[6], aunque obviaremos dicha determinación por razones de simplicidad.

2.6.2. Índice de concentración de Lorenz

En el problema que contemplamos, veamos que un valor aproximado de este índice es el que se obtiene mediante la aplicación de la fórmula basada en los porcentajes acumulados, que se emplea comúnmente en los trabajos prácticos. Aquí, la fórmula correspondiente, explicada en este mismo libro, tomará la configuración simplificada (con n = 15 y q_n = 100):

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

El resultado que ofrece la aplicación de la fórmula anterior es el que queda reflejado en la siguiente tabla, teniendo en cuenta, como ya se ha dicho, que resulta necesario ordenar los valores de la cota taquimétrica del terreno de menor a mayor, para la aplicación correcta de la fórmula, así:

X_i	q_i
6,591	6,591
6,619	13,210
6,621	19,831

[6] Educado como ingeniero y matemático, Pareto se dedicó, en una primera etapa, a la aplicación de las matemáticas a la economía. Sucedió a León Walras en la cátedra de economía de la Universidad de Lausanne, en 1892. Utilizando las curvas de indiferencia introducidas por Edgeworth, logró establecer las condiciones matemáticas para que se produjera el Equilibrio General; creó el concepto de "óptimo", llamado luego óptimo paretiano, para referirse a una situación donde se obtiene una máxima utilidad dado un conjunto previo de bienes o servicios. Sus trabajos sirvieron como punto de partida para la llamada "Economía del Bienestar". Criticó agudamente al socialismo.

	6,642	26,473
	6,648	33,121
	6,656	39,777
	6,659	46,436
	6,671	53,107
	6,676	59,783
	6,677	66,460
	6,681	73,141
	6,689	79,830
	6,710	86,540
	6,717	93,257
	6,743	100
TOTAL	**100**	**797,558**

A esta distribución de probabilidad le corresponde, pues, el índice de Lorenz siguiente:

$$\mathbf{L} = 1 - \frac{697'558}{700} = \mathbf{0'00345 = 0'345\%}$$

que, como puede comprobarse, resulta ser bastante bajo y aproximado al índice de Gini anteriormente calculado, valiendo al respecto las mismas consideraciones efectuadas entonces.