PROSERPINA

14

- Los problemas de orden público durante el Sexenio Revolucionario: Los cuerpos de los voluntarios de la libertad y de los voluntarios de la república. La disolución de la Guardia Rural.
- Métodos cuantitativos de las Ciencias Sociales (II).
- La Educación durante el siglo XVIII en Badajoz.
- La cooperación para el desarrollo como factor de cambio de la sociedad internacional actual.
- Leer sin reglas: Conexiones entre la palabra escrita y oral en las primeras fases del aprendizaje de la lectura.
- La dirección escolar en Extremadura.
- Las Leyes de Bibliotecas de las Comunidades Autónomas (I).
- Modos de eliminación de los venenos. Tratamientos sobre intoxicaciones.
- Nulidad de Matrimonio Canónico y Civil por defecto de consentimiento y por defecto de objeto esencial de consentimiento.

DICIEMBRE, 1998

REVISTA DE LA
UNIVERSIDAD NACIONAL DE EDUCACION
A DISTANCIA

CENTRO REGIONAL DE EXTREMADURA
MERIDA

MÉTODOS CUANTITATIVOS EN LAS CIENCIAS SOCIALES (II)

José Antonio Ballesteros Díez
Profesor-tutor de Historia Moderna de España
Centro Regional de Extremadura de la UNED. Mérida

Una historieta que relatan algunos para justificar su aversión a los métodos cuantitativos es aquella de que había un lugar habitado por dos vecinos, de los cuales uno poseía dos vacas y el otro ninguna, pero que, según la Estadística, en ese lugar por cada vecino se contaba una vaca, lo que, evidentemente, no corresponde en absoluto a la realidad en lo que respecta a la propiedad del ganado, y es que, ciertamente, la historieta se utiliza de una manera tendenciosa, porque se relaciona el conjunto de ganados con el de vecinos, cuando lo correcto sería referirlo al de propietarios. Y es que aquel planteamiento, que en su simplicidad aparente se reduce a calcular un valor medio, como ya se ha indicado, no puede conducir a un análisis científicamente riguroso aplicando un método para calcular valores promedios de una manera mecanicista. Por ello, en este capítulo vamos a estudiar el cálculo de valores promedio, e indicaremos, aunque de modo somero, algunos criterios para elegir el método adecuada para cada problema.

VALOR PROMEDIO

La MEDIA ARITMÉTICA es el método más sencillo para determinar un valor que represente a un conjunto de números, y se calcula sumando todos los elementos de un conjunto y dividiendo por el número de elementos. Se puede determinar la media tanto de un conjunto de números desordenados, datos primarios, como de una serie de números ordenados según algún criterio, como el de unos números organizados en una distribución de frecuencias.

La media aritmética no sólo considera el número de elementos de un conjunto sino también el valor de cada uno, y por esta razón tiene el inconveniente de que se ve afectada por la existencia de un valor extremo; así, por ejemplo, si tenemos como datos los habitantes de todos los núcleos de población de una provincia en la que una ciudad es mucho mayor que cualquier otro núcleo de la provincia, o incluso superior al conjunto de todos los demás, como pueden ser los casos de Madrid o Barcelona, la media aritmética de habitantes por núcleo de población no representará de ningún modo la distribución de la población en la provincia.

Cuando en algún conjnunto de datos hay algunos que exceden considerablemente su valor de los restantes del conjunto, tanto en más como en menos, como puede ser el valor de la estatura de un conjunto de personas en el que hay gigantes y enanos, para obtener el valor representativo de la estatura media de la población se deben eliminar del conjunto los valores extremos, inferiores y superiores, del conjunto de datos, y al valor medio que obtengamos por el procedimiento de cálculo de la media aritmética se le denomina MEDIA ACOTADA.

El uso de los ordenadores y la formidable herramienta que son las Hojas de Cálculo, han facilitado tanto la operatoria que, muchas veces, se puede caer en errores si, al aplicar la función Promedio (en una Hoja MS Excel, por ejemplo), para calcular la media aritmética de unos datos contenidos en un conjunto de celdas no hemos tenido en cuenta la diferencia entre celdas vacías y celdas que contienen el valor cero, pues si, ciertamente, la suma de todo el conjunto de celdas es el mismo no lo es el número de celdas por el que hemos de dividir dicha suma, puesto que las celdas vacías no se deben considerar mientras que las de valor cero sí.

Aunque la sencillez de cálculo de la media aritmética no requiere mayores explicaciones, sin embargo si he visto en diferentes trabajos una aplicación errónea, y así, por ejemplo, se presentan con frecuencia relaciones de precios de un artículo determinado, tomados de actas municipales, libros de cuentas de alguna institución o cualquier otra fuente documental, en las que se relacionan los precios cronologicamente y, algunas veces, se obtiene el valor medios de un período determinado calculándolo como una media aritmética y su determinación es correcta siempre que nos refiramos exclusiva y precisamente a los precios sin extender ese concepto al de "coste", que muchos autores emplean como equivalente, y realmente no es así, pues si en las operaciones de compra o venta efectuadas en base a los precios relacionados normalmente para cada precio se verán afectadas cantidades distintas del artículo de que se trate, y entonces tendremos cuando se calcule el coste medio unitario del conjunto de las operaciones que éste no coincide con el precio medio determinado como media aritmética de los precios, pues el precio de coste medio unitario está afectado por las distintas cantidades de productos, en este caso nos encontramos con el concepto de MEDIA ARITMÉTICA PONDERADA. Veámos con un ejemplo, tomado de mi estudio sobre el Pósito de Mérida en los siglos XVI y XVII, la diferencia entre media aritmética simple y la ponderada; como datos base tenemos las compras de trigo efectuadas entre 1595 y 1603:

Año	Cantidad (fanega)	Precio (maravedí)	Importe
1.595	500	476	238.000
1.596	2.347	476	1.117.172
1.597	10.108	623	6.303.552
1.598	6.355	966	6.140.194
1.599	2.781	1.011	2.812.573
1.600	301	476	143.276
1.601	449,5	301	165.248
1.602	139,5	427	59.602
1.603	4.300	579	2.488.463
Sumas	27.281	5.355	19.468.080

la media aritmética de los precios sería: 5.355 /9 = 595

pero esta cifra no sería representativa del conjunto, y lo podemos comprobar al multiplicar la cantidad total de trigo comprado por este valor medio:

27.281 x 595 = 16.232.195

viéndose claramente que este valor no corresponde al importe total de las compras efectuadas, y así comprobamos cómo es la media ponderada la que nos da el precio medio representativo, y ésta se calcula dividiendo el importe total por la cantidad total, es decir:

19.468.080 / 27.281 = 713,6 mrs/fga.

Esta mayor precisión en la representación de un conjunto de valores que proporciona la ponderación, ha sido incorporada muy recientemente en las cotizaciones diarias de las acciones en la Bolsa de Valores de Madrid, que facilitan los medios informativos y que hasta hace poco tiempo sólo reseñaban para cada valor las cotizaciones de apertura y cierre (algunos también incluían el máximo y el mínimo diario) pero que, al no relacionar estos cambios con los voiúmenes de titulos correspondientes a las cotizaciones de las distintas operaciones realizadas durante la sesión, realmente no reflejaban con precisión la evolución del cambio medio diario de cada título, y así, ahora, el valor medio ponderado es más real que los de cierre, apertura, máximo y mínimo, e incluso más preciso que la media de estos cuatro.

Otro promedio, utilizado frecuentemente en Economía, es el de la MEDIA GEOMÉTRICA, que, en contraste con la media aritmética, tiende a asignar un menor peso

$$G = \sqrt[k]{(x_1)\,(x_2)\,.....(x_k)}$$

a los valores extremos, y su fórmula general es operación que ha de realizarse utilizando logaritmos; también, como en la media aritmética, la media geométrica requiere la ponderación para ser más precisa, y en tal caso la formula será

$$G = \sqrt[n]{(x_1)^{n_1}\,(x_2)^{n_2}.....(x_k)^{n_k}}$$

por ejemplo, en una empresa con 22 empleados, cuyos salarios por hora se distribuyen según se detalla en la tabla

Nº empleados	**Salario / hora**
10	10
5	12
4	12,5
3	14

n = 22

$$G \text{ ponderada} = \sqrt[22]{10^{10} x 12^{5} x 12'5^{4} x 14^{3}}$$

La MEDIA ARMÓNICA es otro tipo de valor promedio para datos relativos a velocidades y magnitudes análogas, y su formula de cálculo es la siguiente

$$\mathbf{H} = \frac{\mathbf{K}}{1/x_1 + 1/x_2 + ... + 1/x_k}$$

veámos más claramente su aplicación con un ejemplo, el de un vehículo que ha cubierto la distancia entre dos puntos a una velocidad media de 60 km/h en el trayecto de ida, y de 80 km /h en el de vuelta, la velocidad media no es de 70 km/h, calculada aritméticamente: (80 + 60) / 2, sino que sería

$$V = 2/(1/60+1/80) = 68'6 \text{ km/h}$$

y, también en este caso, la formula de la MEDIA ARMÓNICA PONDERADA es

$$H = \frac{n}{n_1/x_1 + n_2/x_2 + ... + n_k/x_k}$$

Otra medida de la tendencia central de un conjunto de datos es la MEDIANA, valor que tiene la propiedad de que el número de datos menores que él es igual al número de datos mayores que él, es decir, que, situados los datos por orden de magnitud, la Mediana divide en dos partes iguales el conjunto total de datos. Si el número de elementos del conjunto de datos es impar, la Mediana será el valor central una vez ordenados, así, por ejemplo, para la serie 1, 2, 3, 4, 5, la Mediana es igual a 3, pero cuando el número de datos es par la Mediana se calcula como promedio de los dos datos centrales, por ejemplo, en la serie 1, 2, 3, 4, 5, 6, la Mediana es 3'5, promedio de 3 y 4. Gráficamente la mediana será un valor de X tal que, trazando por él una vertical, quede a la derecha un área igual a la que existe a la izquierda.

La Mediana es un valor muy utilizado en estudios antropométricos, por ejemplo, en las estaturas de un conjunto de individuos de la misma edad y sexo; también en las evaluaciones de los alumnos de una misma clase. etc.

Según la posición o el orden de los datos, la Mediana es solamente una medida entre otras varias que se pueden efectuar para analizar el conjunto, y así también pueden determinarse CUARTILES, DECILES y PERCENTILES, y así como la mediana divide al conjunto en dos mitades, los cuartiles lo hacen en cuatro, los deciles en diez y los percentiles en cien; volviendo al ejemplo de las medidas antropométricas, el Cuartil nos daría la cuarta parte, con más o con menos, estatura del conjunto. En definitiva, estas medidas nos marcan la posición relativa de un valor en un conjunto de datos.

Otro valor promedio de posición es la MODA, que, como su propio nombre refiere, corresponde al valor que más veces se repite dentro de un conjunto de datos, tal como puede ser el formado por: 5, 6, 4, 4, 3, 2, 6, 4, 4, la MODA es 4.

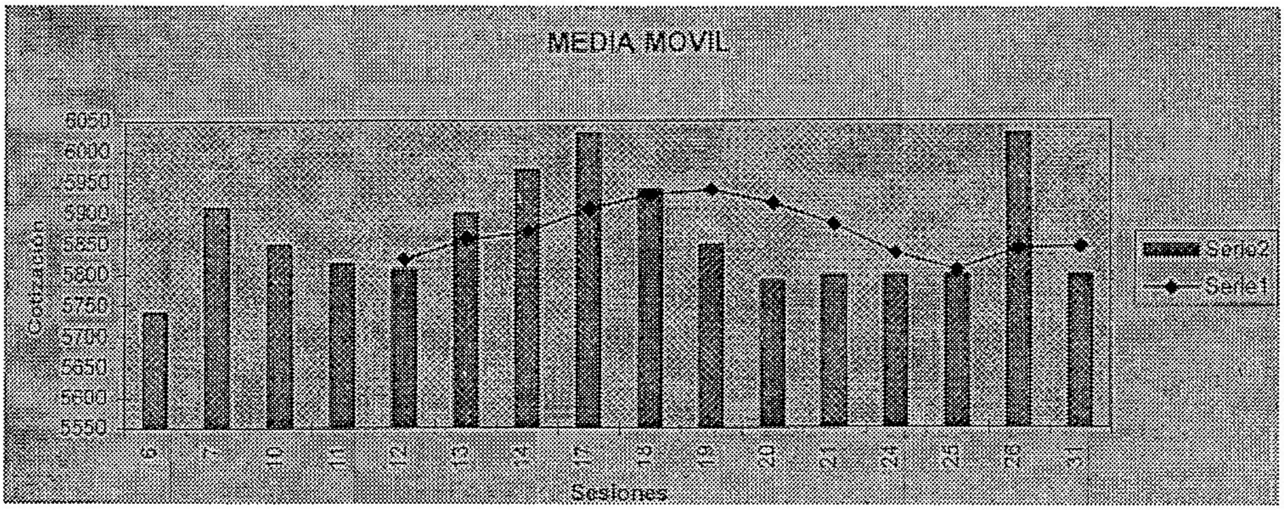
La elección de un procedimiento de cálculo determinado en el análisis cuantitaivo de unos datos ha de hacerse reflexivamente, pues si la Mediana es el valor que se repite con más frecuencia en un conjunto de datos, la Mediana representa la posición central y la Media es el valor promedio; hemos de recordar que ninguna de estas medidas, unitariamente consideradas, representan de modo absoluto al conjunto, y veámos esto con el ejemplo de un conjunto de datos agrupados en tres bloques, la mitad de los cuales son valores bajos que se repiten y la otra mitad consiste en dos valores altos, tanto Promedio como Mediana proporcionarán un valor situado en una zona central relativamente vacía, mientras que la Moda dará el valor más bajo dominante por su mayor frecuencia, lo que nos hace ver que la Moda tiene una aplicación muy limitada, aunque sea utilizada en los estudios demográficos, por ejemplo, en las edades de los contrayentes de matrimonio, o en la edad más frecuente en las mujeres para concebir su primer hijo.

En el capítulo anterior de esta serie de artículos de carácter introductorio a los métodos cuantitativos en los estudios de las ciencias sociales, al tratar de las series cronológicas, citaba a Fernand Braudel por sus análisis y explicación de los ciclos del tiempo histórico, que él establece en tres grupos: cicios cortos, medios y largos. Puede suceder que diferentes tipos de fluctuación, susceptibles de distinción analitica, aparezcan confundidos en la curva bruta, en cuyo caso habremos de recurrir a métodos que nos permitan disociar del movimiento global los distintos tipos de fluctuaciones para representar unicamente el movimiento de larga duración; ya citamos en el capítulo anterior la Línea de Tendencia representativa de un conjunto de datos ajustada por el procedimiento de los mínimos cuadrados y ncuya representación gráfica es sumamente fácil con las hojas de calculo informáticas. Pues bien, otro procedimiento de muy generalizado uso es el de la MEDIA MÓVIL, método que elimina el movimiento de corta duración y minimiza la influencia de los valores extremos; la determinación de medias móviles se puede efectuar tanto con series de números absolutos, o datos brutos, como de números índices; en este cáculo el primer paso consiste en fijar el número de ciclos a considerar estableciendo su duración media, y posteriormente se calcula el valor medio del conjunto de datos comprendidos en el primer período establecido, eliminando sucesivamente en cada serie el valor correspondiente a la cronología más antigua y añadiéndole el siguiente al del período considerado. Veámos con un ejemplo cómo se hacen éstos cálculos, y para ello vamos a tomar la evolución de la cotización de cierre de las acciones de Repsol, S.A. en la Bolsa de Madrid durante el mes de Marzo de 1.997 y calcularemos la media móvil de 5 días

MEDIA MÓVIL DE LAS ACCIONES REPSOL EN LA BOLSA DE MADRID

Marzo de 1997 *Intervalo: 5 días*

Día	Cot. cierre	'Media móvil
6	5740	
7	5910	
10	5850	
11	5820	
12	5810	5826
13	5900	5858
14	5970	5870
17	6030	5906
18	5940	5930
19	5850	5938
20	5790	5916
21	5800	5882
24	5800	5836
25	5800	5808
26	6030	5844
31	5800	5846



Como vemos, las medias móviles son representativas de un plazo dado, y aunque en el ejemplo expuesto el intervalo elegido ha sido de 5 días, en los análisis bursátiles se elaboran tres medias móviles: una, de 10 días, que representa la evolución a corto plazo; otra media móvil de 75 días, indicativa de la tendencia a medio plazo, y, por último, una tercera media móvil de 200 días como representativa de la tendencia a largo plazo.

En los estudios demográficos, sociológicos, económicos, son frecuentes las medias móviles que toman a la década como período de computo a medio plazo, e incluso la media móvil de 50 años como largo plazo. De uso generalizado en estudios de todas clases con series temporales es la media móvil de los doce últimos meses como representativa de los movimientos interanuales.

El cálculo de valores promedio puede ser también de utilidad para comprobar si, cuando estudiamos series temporales, los datos varían de un modo cíclico, que según sea la extensión en el tiempo de la serie podrá comprender ciclos cortos, medios o largos, los cuales, a su vez, habremos estratificado en intervalos más pequeños, de los que podemos calcular el valor promedio de cada uno y luego comparar la media del intervalo inicial con la del final, y midiendo esta diferencia en porcentaje tendremos una evaluación aproximada de la amplitud del movimiento. Este procedimiento rápido para determinar movimientos cíclicos es de una mayor precisión utilizando las medias móviles.

ESTRATIFICACIÓN DE DATOS

En esta breve y elemental introducción a los métodos cuantitativos en las ciencias sociales, hemos tratado los datos considerándolos de un modo individualizado dentro del conjunto, pero también se pueden analizar haciendo una estratificación de los mismos, es decir, clasificando en grupos homogéneos los datos cuantitativos de un hecho, o proceso, determinado. Supongamos, por ejemplo, los resultados obtenidos al arrojar cien veces un dado; anotaremos una serie de números comprendidos entre 1 y 6; para que éste conjunto nos proporcione algún conocimiento sobre el modo de comportarse el dado tendremos que contar cuantas veces ha salido el 1, cuantas el 2, etc., o sea, hacemos un agrupamiento o clasificación de los datos, que podemos tabular así

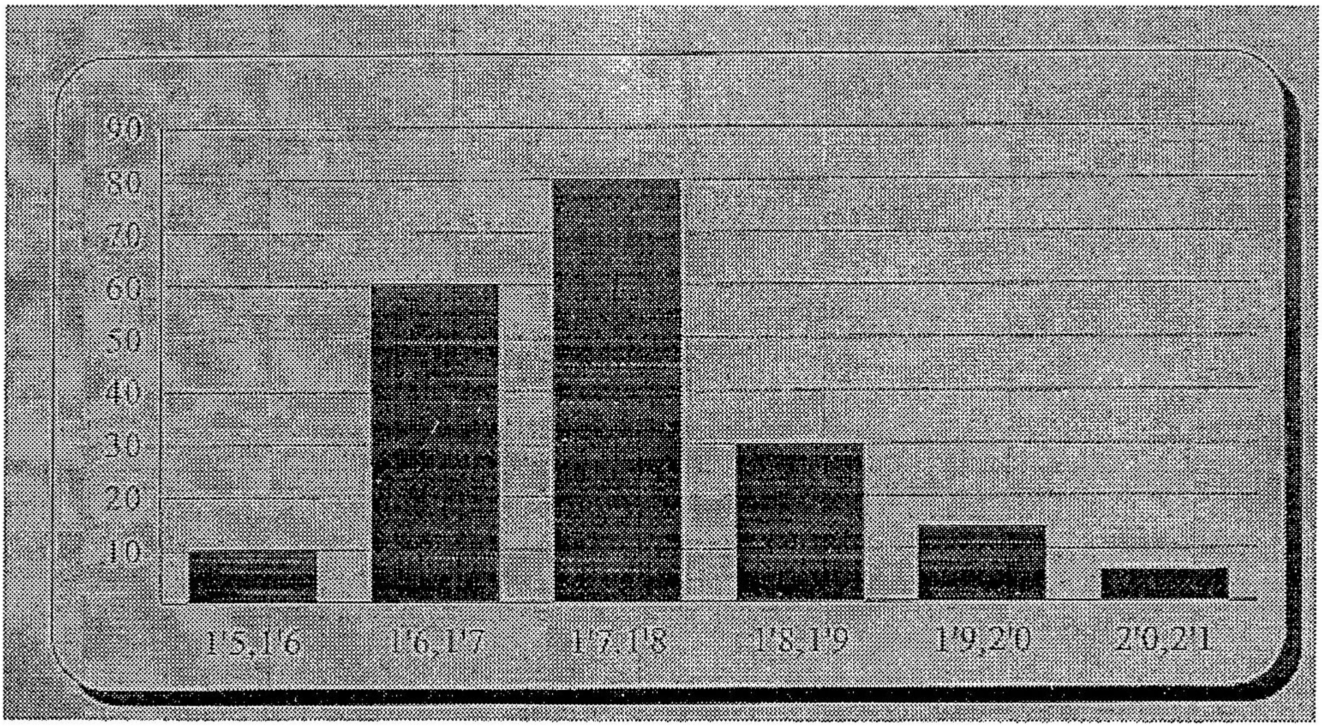
Número	1	14	veces	
"	2	17	"	
"	3	13	"	
"	4	19	"	
"	5	17	"	
"	6	20	"	Total veces = 100

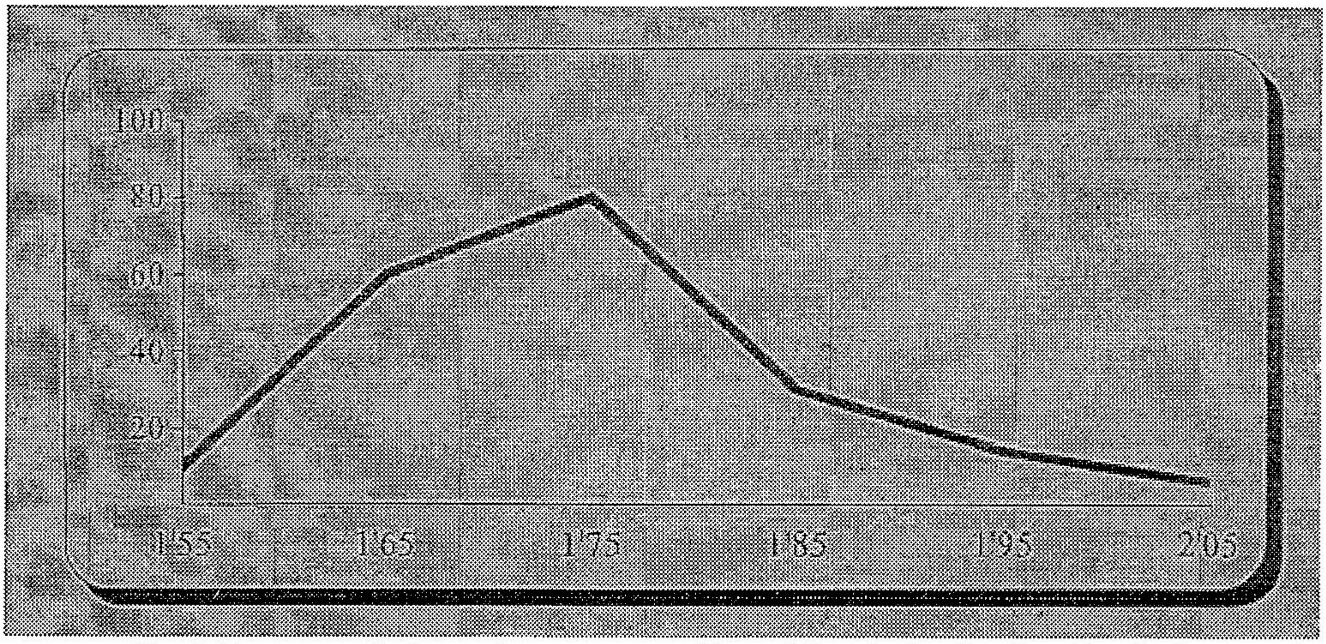
Es evidente la utilidad de la estratificacion de datos realizada, pues las cien observaciones se han clasificado solamente en seis grupos, sin que tan enorme reducción suponga ninguna merma en las posibilidades de conocimiento con precisión del hecho estudiado. Apoyándonos en este ejemplo, diremos que la cifra que hay en cada una de las caras del dado (1, 2, 3, ...6) recibe el nombre de ATRIBUTO o VARIABLE, y el número de veces que ha aparecido cada atributo o variable (los datos reseñados en la tabla anterior) se llama FRECUENCIA ABSOLUTA, pero si ésta la referimos al total de las observaciones efectuadas entonces tendremos la FRECUENCIA RELATIVA.

Veámos con una serie de datos antropomórficos un ejemplo más significativo, se trata de la medición de la estatura de todos los estudiantes que ingresan en la universidad, y se obtienen 200.000 datos con valores que oscilan entre un minimo de 1'50 m y un máximo de 2'10 m. Procedemos a clasificar las estaturas en intervalos de 10 cm y tabulamos las frecuencias que corresponden a cada intervalo, anotando también en la tabla el valor medio de cada intervalo, que se denomina MARCA DE CLASE

Intervalo	**Frecuencia** (n_i)	**Marca de Clase** (x_i)
1'50-1'60 m	10.000	1'55
1'60-1'70 m	60.000	1'65
1'70-1'80 m	80.000	1'75
1'80-1'90 m	30.000	1'85
1'90-2'00 m	14.000	1'95
2'00-2'10 m	6.000	2'05

esta tabla comprende, pues, al conjunto de datos agrupados y ordenados con sus respectivas frecuencias y las marcas de clase correspondientes a los intervalos establecidos, y recibe el nombre de DISTRIBUCION DE FRECUENCIA cuya representación gráfica se puede hacer mediante un histograma





También se puede hacer la representación gráfica de la distribución de frecuencia relacionando éstas con las marcas de clase esta otra forma del grafico se conoce como polígono de frecuencias.

Pero la utilización del histograma no debe reducirse a ser la mera representación gráfica de una tabla de datos, sino que puede, y debe ser, una herramienta para profundizar en el análisis de los procesos y hacer una correcta interpretación del hecho que estudiamos, y a tal fin debemos buscar las respuestas a unos interrogantes metódicos:

+ ¿ Hay un valor de tendencia central ?

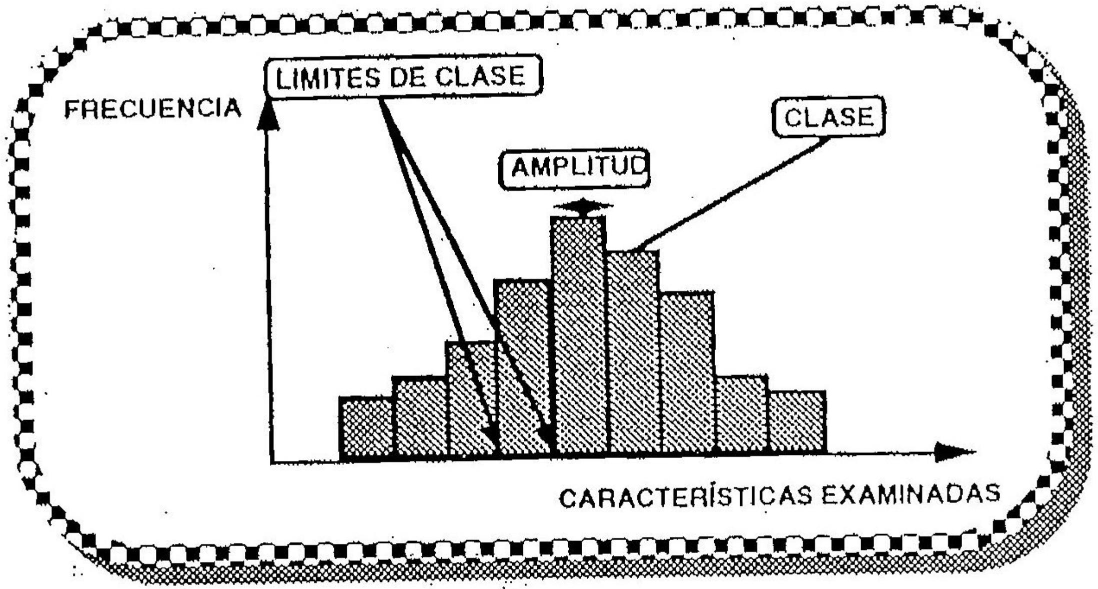
+ ¿ Se aumenta o reduce la frecuencia de las clases con regularidad antes y después de las clases de mayor frecuencia ?

+ ¿ Se produce simetría ?

+ ¿ Qué tamaño tiene la dispersión ?

+ ¿ Tienen lugar configuraciones inesperadas, es decir, aquellas en las que el polígono de frecuencias no tiene forma acampanada ?

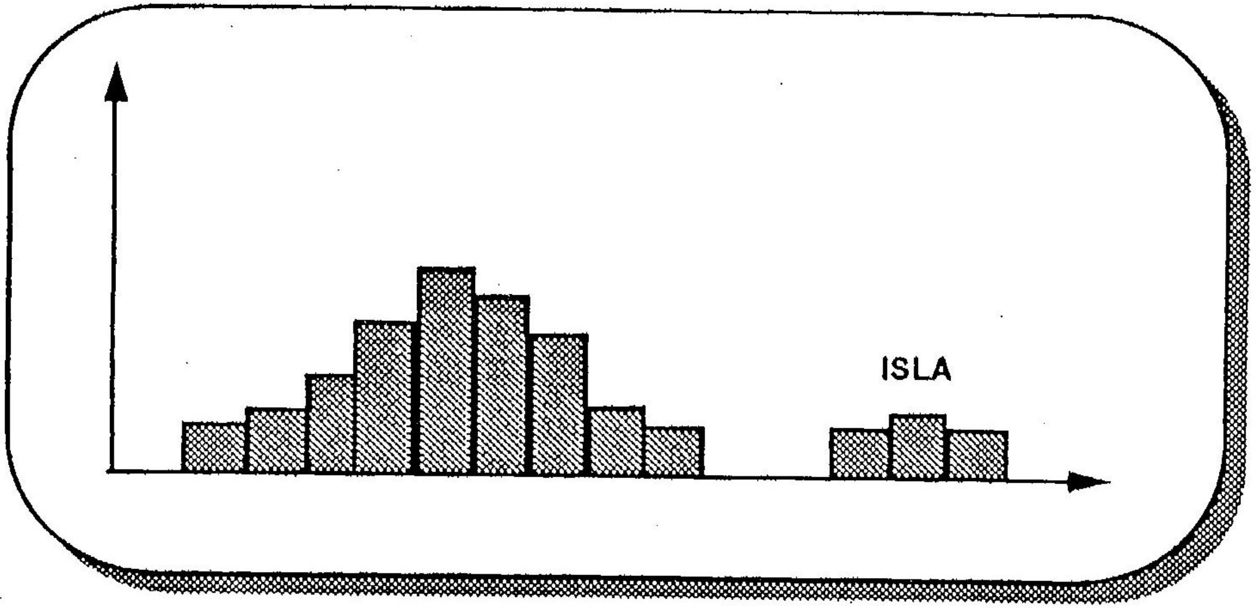
Ejemplo



CLASE: *cada barra*
AMPLITUD DE CLASE: *es el intervalo comprendido entre el valor máximo y el mínimo de cada clase.*
FRECUENCIA: *El número de observaciones de cada clase*
LÍMITES DE LA CLASE: *Los valores máximo y mínimo de cada clase*

Ejemplos de Distribuciones Específicas (I)

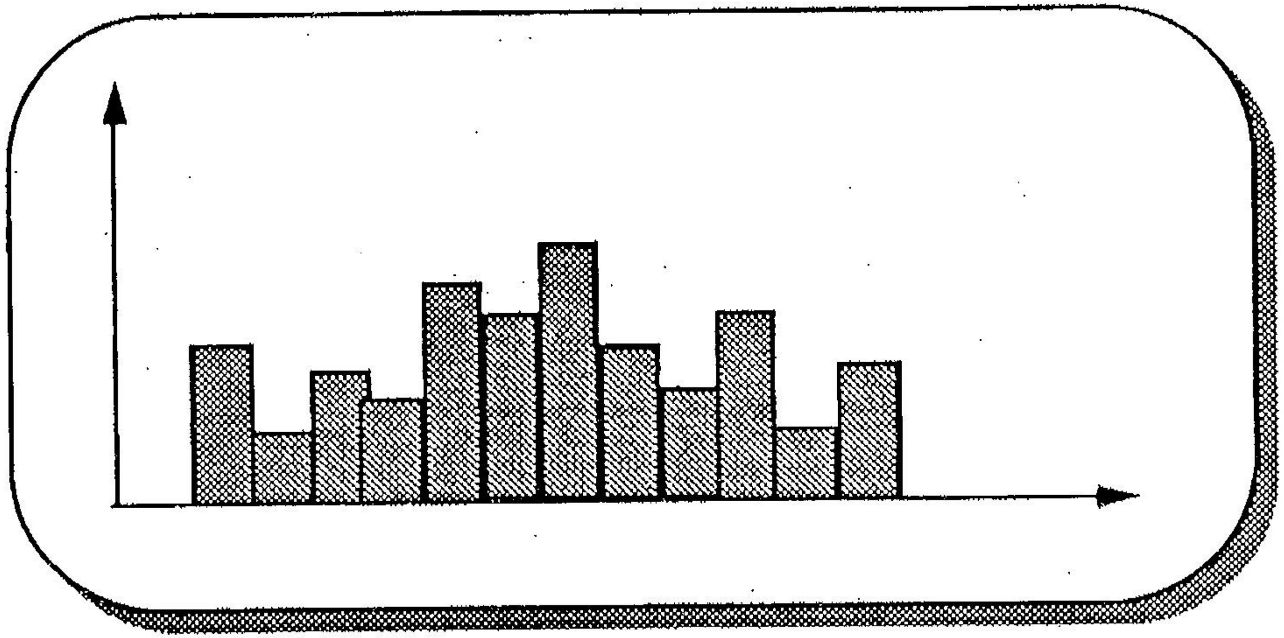
EN ISLA



•Buscar la causa por la que aparecen ese número de datos aislados utilizando la estratificación

Ejemplos de Distribuciones Específicas (II)

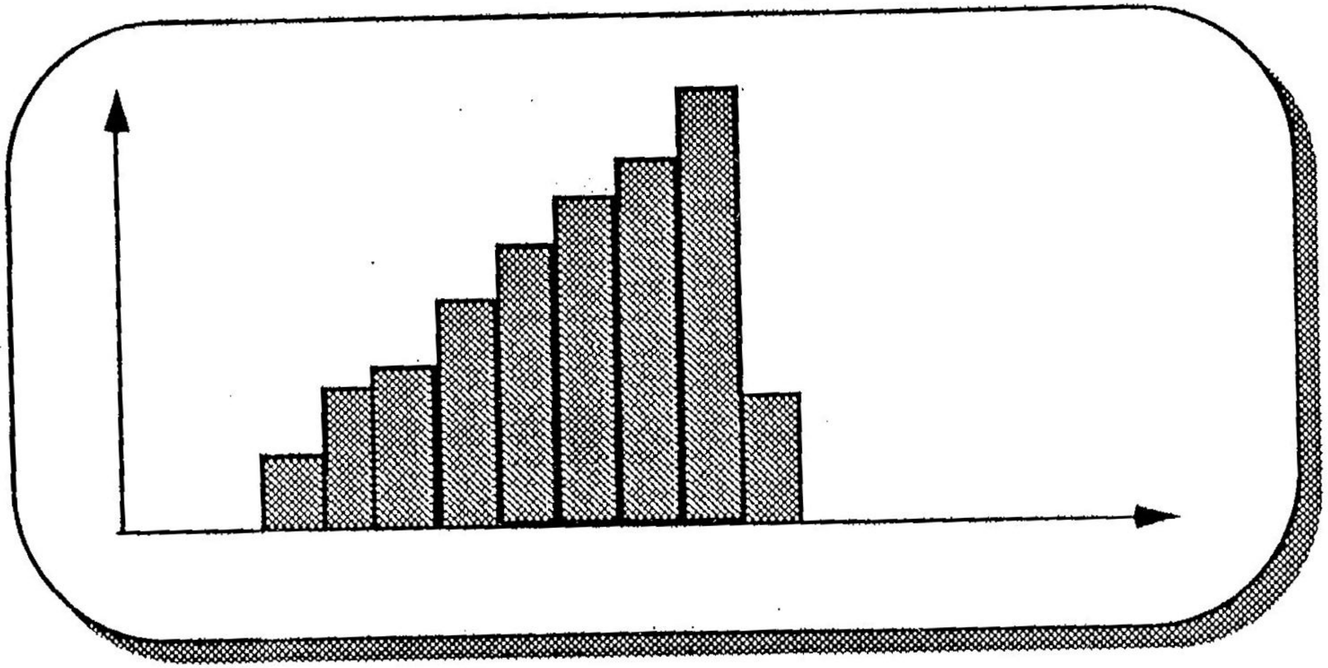
EN DIENTES DE PEINE



- Demasiadas clases con respecto al número de datos
- Redondeo inadecuado de las medidas
- Errores de medida

Ejemplos de Distribuciones Específicas (III)

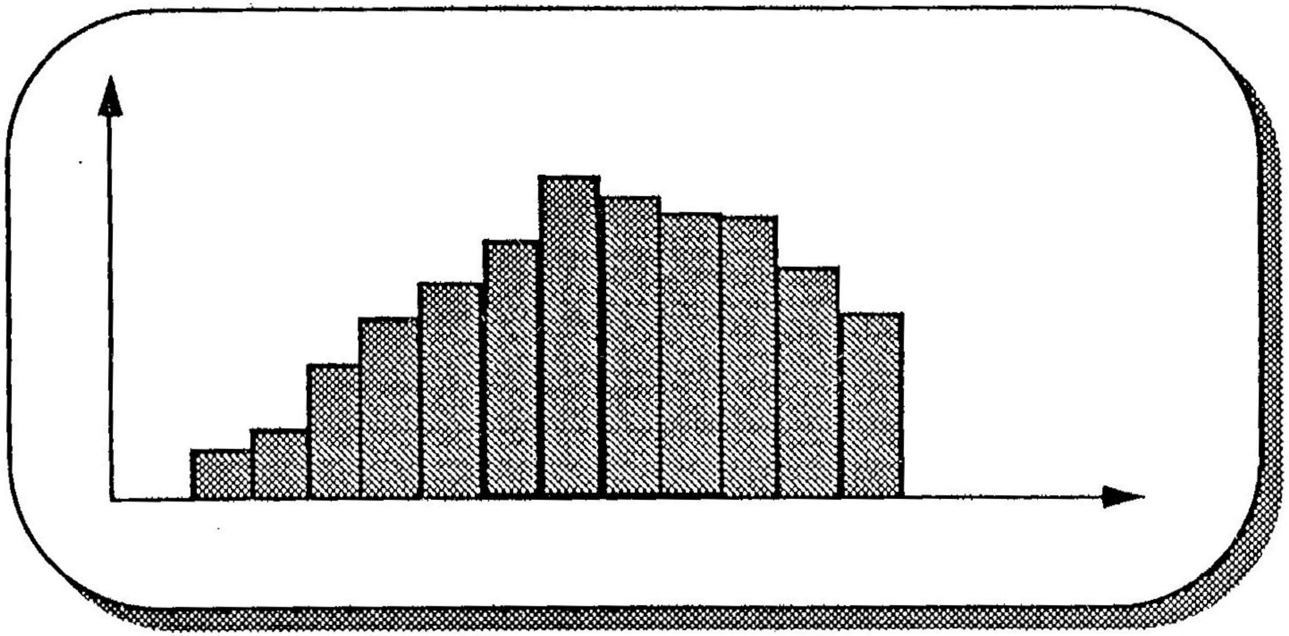
EN PRECIPICIO



•Aparición de errores de medición o valores falseados (por ejemplo, muestras tomadas en la parte final de la escala de utensilios, otros instrumnetos o desconsideración sin causa de algunos valores.

Ejemplos de Distribuciones Específicas (IV)

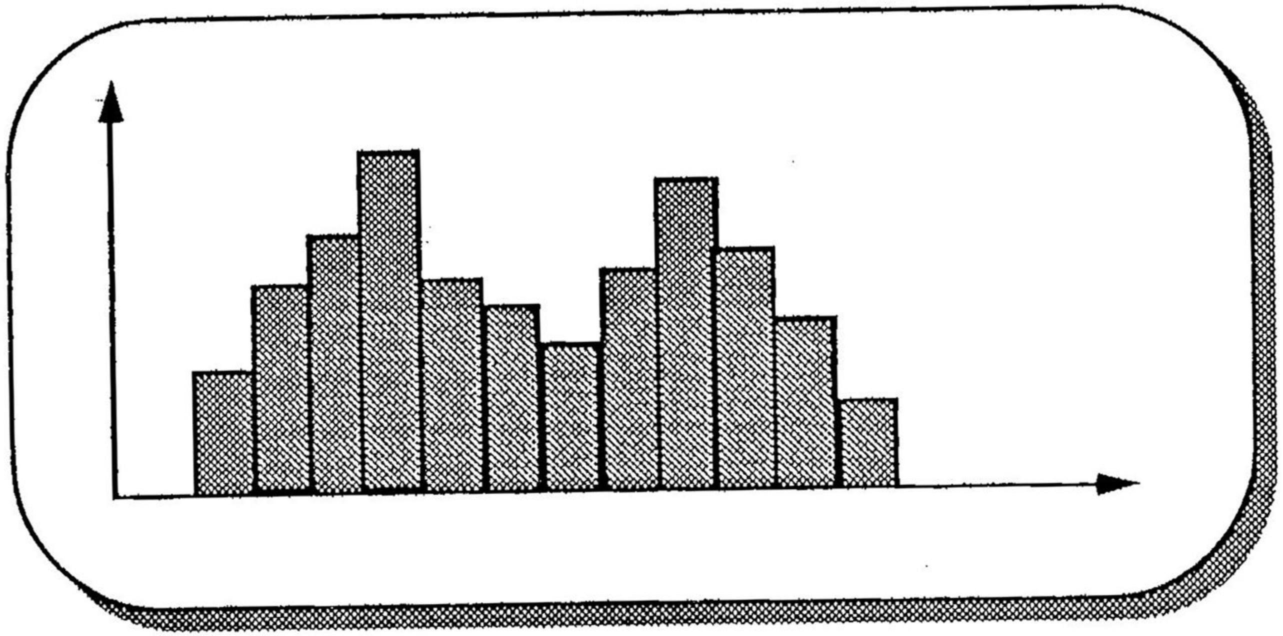
ASIMÉTRICO



- Errores de muestreo o de medición
- Superposición de datos no homogéneos (se debe estratificar)

Ejemplos de Distribuciones Específicas (V)

BIMODAL

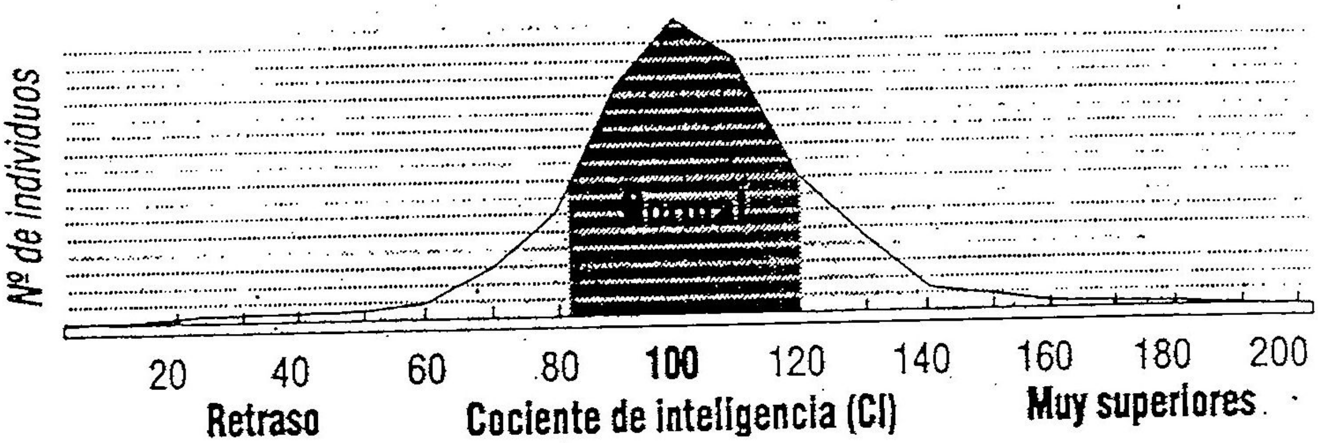


- Superposición de datos no homogéneos

Un ejemplo gráfico de una distribución de frecuencia campaniforme es el que muestra la distribución del cociente de inteligencia en la población:

DISTRIBUCIÓN DEL COCIENTE DE INTELIGENCIA DE LA POBLACIÓN

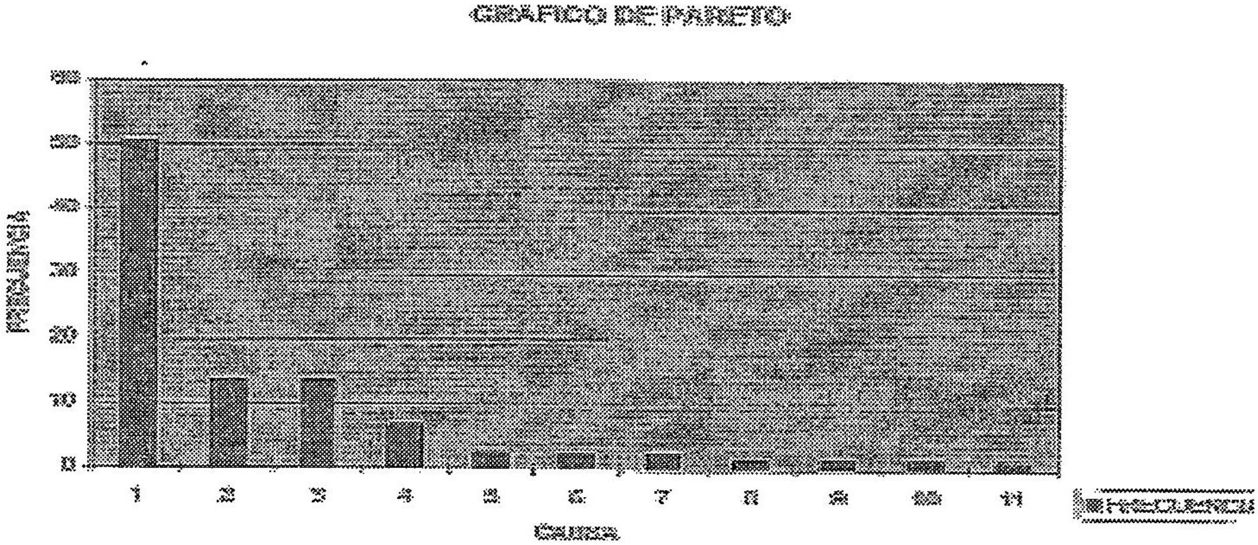
La poca altura de los extremos indica que hay pocas personas con cocientes muy altos o muy bajos, mientras que en el centro de la curva (valores normales de inteligencia) se sitúa la mayoría de los individuos.



Todos los interrogantes antes expuestos pueden ayudarnos en el proceso de análisis, y muchos de ellos se pueden cuantificar; pero también se presentan otros tipos de fenómenos en cuyo estudio encontramos distintas variables, por ejemplo, en la investigación de la mortalidad de una población, determinada tanto en el espacio como en el tiempo, podemos encontrar que el origen de la muerte se debe a un grupo de causas diversas: oncológicas, cardiovasculares, traumáticas, etc, y su representación gráfica se podrá hacer con un histograma en el cual cada barra corresponda a un grupo de causas letales; en este caso no debemos buscar una distribución de frecuencias acampanada, sino que lo más acertado es recurrir al análisis de Pareto, que consiste en representar los datos en orden de frecuencia descendente, lo que permite jerarquizar visualmente los factores más relevantes del hecho analizado al ordenar los grupos según su importancia de más a menos; a su vez, en cada uno de estos grupos se pueden establecer diversos subgrupos y representarlos de igual modo, como ejemplo veremos las defunciones ocurridas en el Hospital de San Juan de Dios, en Mérida, de 1809 a 1831, según datos del Dr. López Gómez tomados de su libro "Salud pública y medicina en Mérida, de 1700 a

CAUSA		Nº DE FALLECIDOS	FRECUENCIA
1 Enfermedades infecciosas		37	50'68 %
1.1 Calentura	35		
1.2 Terciana	1		
1.3 Disentería	1		
2 Neumología		10	13'69 %
2.1 Afecto del pecho	5		
2.2 Tisis	4		
2.3 Dolor de costado	1		
3 Cardio-circulatoria		10	13'69 %
3.1 Hidropesía	9		
3.2 Apoplejía	1		
4 Digestivo		5	6'84 %
4.1 Diarrea	5		
5 Traumatología y reumatologia		2	2'73 %
5.1 Dolores reumáticos	1		
5.2 Fractura complicada	1		
6 Cirugía	2		2'73 %
6.1 Fistula	1		
6.2 Varios	1		
7 Urología		2	2'73 %
7.1 Llaga en escroto	1		
7.2 Mal de orina	1		
8 Dermatología y venerología		1	1'36 %
8.1 Ulceras	1		
9 Neurología		1	1'36 %
9.1 Convulsiones	1		
10 Psiquiatría		1	1'36 %
10.1 Fatuo	1		
11 Varios		1	1'36 %
11.1 Escrofulas	1		

1833", págs 318-319. El gráfico de Pareto lo haremos tomando solamente los grupos principales, para mayor sencillez de la exposición



Esta simplificación que se consigue al estratificar los datos, también puede sernos de utilidad para calcular valores promedio, así, dada una distribución de frecuencias tal como

X_i	Ni
X_1	N_1
X_2	N_2
X_3	N_3
X_4	N_4
X_5	N_5

la media aritmética será $\mathbf{M} = \dfrac{\Sigma\, \mathbf{x_i\, n_i}}{\Sigma\, \mathbf{n_i}}$

Pero cuando se presente el caso de tener que operar con números muy altos, la distribución de frecuencias nos permite un método abreviado para determinar la media y evitar el engorroso cálculo de los productos Xi * Ni, esto es, de las marcas de clase por sus respectivas frecuencias, y para resolverlo con más facilidad haremos un cambio de variable del modo que seguidamente describimos: siendo h, la diferencia entre dos marcas de clase, y k la marca de clase que ocupa el lugar central del conjunto de marcas de clase, entonces se establece una nueva variable, u, que viene dada por la expresión

$$U_i = (X_i - K) / h$$

de donde

$$U_i * h = X_i - k \qquad \text{y} \qquad X_i = U_i + k$$

$$M = \frac{\Sigma x_i n_i}{\Sigma n_i}$$

y volviendo a realizar el cambio de variable

$$M = \frac{(u_1 h + k)\, n_1 + (u_2 h + k)\, n_2 + \ldots + (u_5 h + k)\, n_5}{\Sigma n_i}$$

resultará finalmente

$$M = h * \bar{u} + k$$

siendo $\bar{u}$ la media aritmética de U_i.

Veámoslo más facilmente con un ejemplo

Xi	Ni	Ui	Ni * Ui
587	5	-2	-10
687	71	-1	-71
k=787	167	0	0
887	37	1	37
987	8	2	16

$$\Sigma N_i = 288 \qquad \Sigma Ni^*Ui = -28$$

vemos que **h = 100**, es decir, la diferencia entre valores consecutivos de X, y que el valor central de la serie, o sea, k, es el correspondiente a $X_3 = 787$. Mediante la formula

$$Ui = (Xi - k) / h$$

hemos obtenido el valor que señalamos en la tabla en la columna Ui, y de ahí el producto **Ni * Ui** que corresponde a la cuarta columna de la tabla; la media de la Ui será

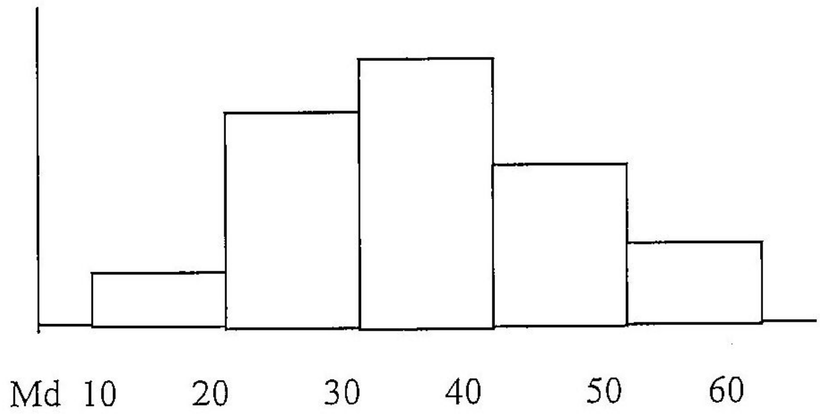
$$U = \Sigma Ui^*Ni / \Sigma Ni = -28 / 288 = -0{,}097$$

en consecuencia, la media aritmética del conjunto será

$$M = 100 * (-0{,}097) + 787 = 777{,}3$$

También es de interés la determinación de la Mediana en una distribución de frecuencia, como la a continuación desarrollamos

I	Ni
10 - 20	2
20 - 30	18
30 - 40	20
40 - 50	15
50 - 60	5



en la representación gráfica vemos cómo la mediana es el valor de X tal que divide al área representada en dos mitades, y para determinar el valor de X correspondiente a la Mediana calculamos la suma de las áreas parciales

$$20 + 180 + 200 + 150 + 50 = 600$$

por tanto, la Mediana será el valor de X que corresponde al área de 300, y estará en el tercer intervalo, pues **20 + 180 +200 = 400**, la Mediana es

$$Md = 30 + X$$

donde 30 es el extremo inferior del intervalo donde se encuentra la Mediana, y X es el valor a determinar

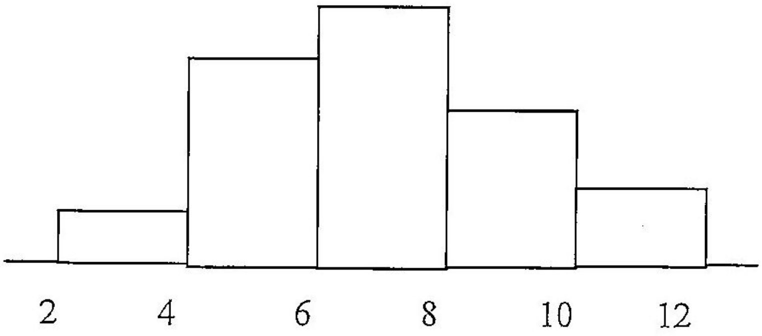
$$300 = 10^*2 + 18^*10 + 20X \qquad X = 5$$

en consecuencia, el valor de la Mediana será

$$\mathbf{Md = 30 + X = 35}$$

La Moda, como sabemos, también es un valor de posición, y en un conjunto estratificado corresponde al valor que tiene una máxima frecuencia; se puede calcular por el procedimiento que desarrollamos en este ejemplo

I	Ni
2-4	4
4-6	10
6-8	40
8-10	30
10-12	1



Es evidente que la Moda estará dentro del intervalo 6 - 8, y su posición se determina sumando al extremo inferior de este intervalo (6), la amplitud del intervalo multiplicado por el cociente de dividir la frecuencia del intervalo superior entre la del inferior más el superior, o sea

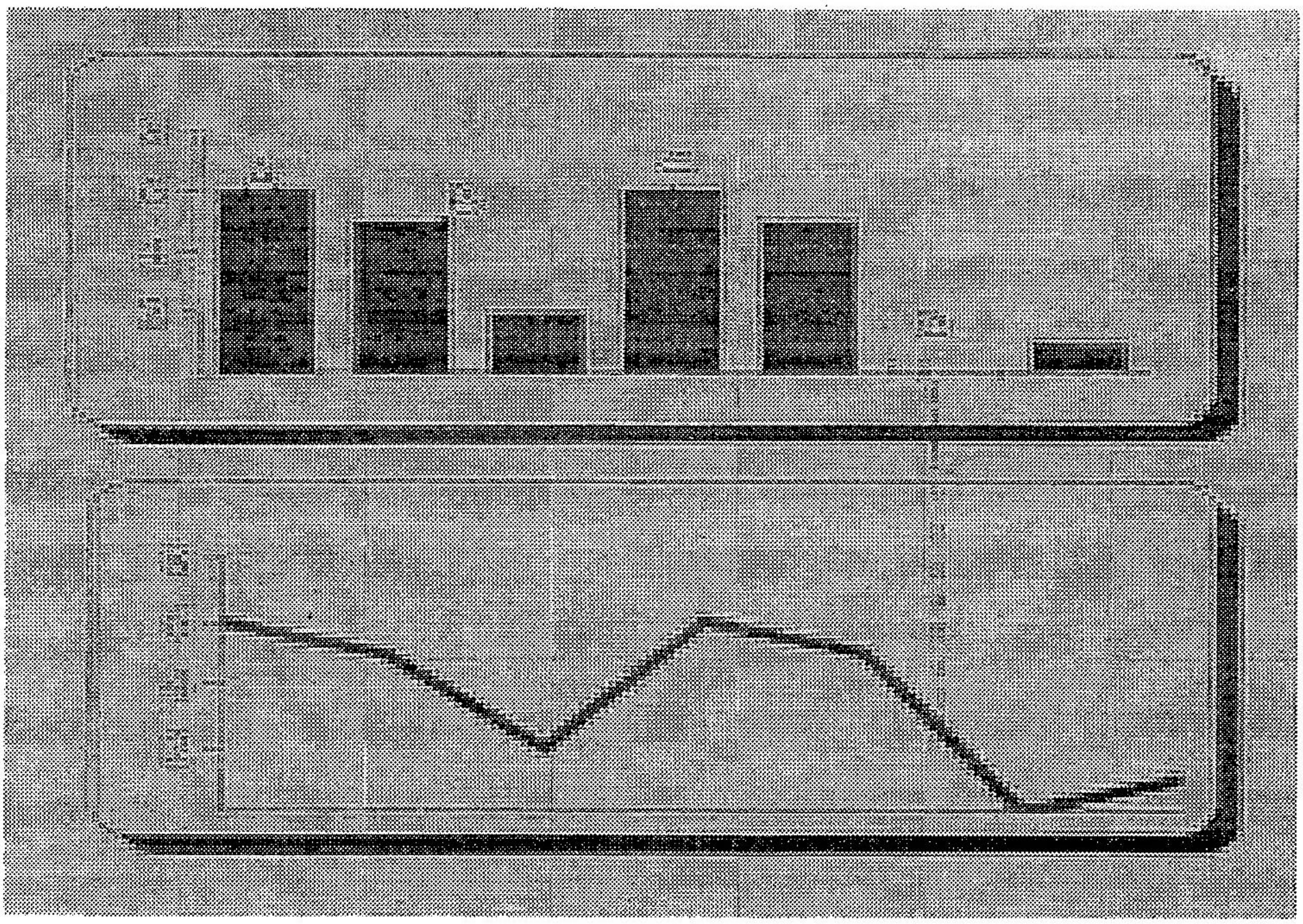
$$\mathbf{Mo = 6 + 2 * (20 \ / (20 + 10)) = 7{,}3}$$

mientras que la Media Aritmética sería

$$\mathbf{M = \Sigma Xi * Ni / \Sigma Ni = 7{,}1}$$

¿Por qué empleamos el histograma o gráfico de barras en una distribución de frecuencias en vez de utilizar un gráfico lineal ? Veámos una distribución de frecuencias como la siguiente

I	N
0-4	6
5-9	5
10-14	2
15-19	6
20-24	5
25-29	0
30-34	1



Si comparamos unos puntos del gráfico de barras con los homólogos del de línea, veremos que A corresponde al intervalo 0-4, donde tenemos 6 unidades; el B pertenece al 5-9 ,donde hay 5 unidades, el C al 15-19, con 6 unidades, y el D al intervalo de 21 a 29 que está vacío, y así se ve en el histograma, pero, sin embargo, en el gráfico lineal a la abscisa del punto D corresponde un valor en el eje de ordenadas de 1,5, lo que sabemos que es imposible; por tanto, aunque los gráficos lineales nos representan muy bien la continuidad de las variable, cuando los datos no se pueden interpretar como si fueran contínuos, lo acertado es utilizar los histogramas en las representaciones gráficas.

En la página 4 del suplemento de economía del diario "El Mundo" del domingo 26 de Enero de 1997, en un artículo de M. Irazueta y C. Llorente, escriben lo siguiente: "Los fondos de renta variable son una buena oportunidad . . . pero la fórmula de cálculo para conocer la rentabilidad que utilizan algunos de estos productos tiene truco, una de las fórmulas que aplican es la de las medias mensuales, sistema de medición que toma como referencia el Ibex de un día determinado de cada mes, y al recoger las subidas y bajadas del ciclo halla su media, lo que no coincide con el recorrido real del índice porque se suman valores positivos (subidas) y negativos (bajadas). Un claro ejemplo es la evolución del Ibex de Enero de 1993 a Diciembre de 1996; según el grupo Safei la revalorización total del índice fue algo más del 222 %, mientras que si se toma la media mensual esta fue del 48,7 %. . . Otro sistema de cálculo es el que toma como referencia el valor que el Ibex tiene en la fecha inicial del fondo y el que consigue cuando finaliza el período de garantía, el resultado final se acerca más a la realidad"

He tomado esta cita periodistica de actualidad para recordar lo expuesto en el primer artículo de esta serie (ver PROSERPINA 13), y llamar la atención sobre el contenido del presente trabajo, pues se ve con frecuencia cómo se presentan valores promedio en distintos estudios sin haber elegido con criterio correcto la media más adecuada, y también que no se utiliza la distribución de frecuencia tanto como se debiera, pues, especialmente el método de Pareto, es de una utilidad indudable. Quiero reiterar que el objetivo de estos trabajos es acercar a los estudiosos de las ciencias sociales a la utilización de los métodos cuantitativos, pues aunque su uso se puede extender por la generalización de la Informática, ésta no puede dar todas sus posibilidades si no tenemos claros los criterios de análisis y los métodos a utilizar.