

TEMA 2: PARÁMETROS ESTADÍSTICOS. CÁLCULO, SIGNIFICADO Y PROPIEDADES.

1. INTRODUCCIÓN

Hasta ahora hemos visto cómo se pueden resumir los datos obtenidos del estudio de una muestra (o una población) en una tabla estadística o un gráfico. No obstante, tras la elaboración de la tabla y su representación gráfica, en la mayoría de las ocasiones resulta más eficaz “condensar” dicha información en algunos números que la expresen de forma clara y concisa. Estos valores, representativos de todos los de una distribución, se llaman parámetros. Este tipo de medidas descriptivas utilizadas son, principalmente, medidas de centralización o posición y medidas de dispersión.

- **Medidas de centralización (posición).** Son coeficientes de tipo promedio que tratan de representar una determinada distribución, pueden ser de dos tipos:

1.- Centrales:

- Medias:
 - Aritmética
 - Geométrica
 - Armónica
- Mediana
- Moda

2.- No centrales:

- Cuantiles:
 - Cuartiles
 - Deciles
 - Percentiles

- **Medidas de dispersión.** Son complementarias de las de posición en el sentido que señalan la dispersión en conjunto de todos los datos de la distribución respecto de la medida o medidas de localización adoptadas.
 - Medidas de dispersión absolutas: recorrido, recorrido intercuartílico, desviación media, varianza, desviación típica.
 - Medidas de dispersión relativas: coeficiente de variación.
- **Medidas de forma.** Proporcionan una idea de la de simetría y apuntamiento de la distribución.
 - Medidas de asimetría: coeficientes de asimetría de Pearson y de Fisher.
 - Medidas de apuntamiento: coeficiente de apuntamiento o curtosis.

2. MEDIDAS DE POSICIÓN

En este proceso de síntesis buscamos unos valores que nos fijen el comportamiento global del fenómeno estudiado a partir de los datos individuales recogidos en la información disponible. Estos valores sintéticos globales son los llamados Parámetros de centralización o Medidas de Posición.

2.1. La Media Aritmética.

Definimos la Media Aritmética como la suma de todos los valores de la distribución dividida por el número total de datos. Se representa por \bar{x} .

Si la variable toma los valores x_1, x_2, \dots, x_n , siendo f_1, f_2, \dots, f_n , las frecuencias absolutas correspondientes de la distribución, la media aritmética se calcula con la expresión:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_n f_n}{N} = \frac{\sum_{i=1}^n x_i f_i}{N}$$

Ejemplo: Si tenemos la siguiente distribución, se pide hallar la media aritmética, de los siguientes datos expresados en kg.

x_i	f_i	$x_i f_i$
54	2	108
59	3	177
63	4	252
64	1	64
	$N = 10$	601

$$\bar{x} = \frac{\sum_{i=1}^n x_i f_i}{N} = \frac{601}{10} = 60,1 \text{ kg}$$

Ejercicio 1: La tabla adjunta muestra el número de faltas de asistencia en una clase a lo largo del mes:

Nº de faltas	0	1	2	3	4	5
Nº de alumnos	10	7	6	2	1	4

Calcula la media aritmética.

Solución: 1,63

En el caso de que tuviéramos una distribución con datos agrupados, se toma como valores de la variable el punto medio del intervalo de clase, esto es, las marcas de clase, y se aplica la misma fórmula para dichos valores.

Ejercicio 2: La tabla adjunta muestra los resultados de unos alumnos en la prueba de salto con pértiga:

Medida del salto (m)	[2; 2,5)	[2,5;3)	[3; 3,5)	[3,5; 4)
Nº de alumnos	6	12	15	4

Calcula la media aritmética.

Solución: 2,98

Propiedades de la media aritmética

- La media aritmética es el parámetro de centralización más utilizado.
- Si a todos los valores de la variable les sumamos una constante k , la media aritmética queda aumentada en esa constante.
- Si todos los valores de una variable los multiplicamos por una constante k , su media aritmética queda multiplicada por la misma constante.
- La suma de las desviaciones de los valores de la variable respecto a su media es cero.

$$\sum_{i=1}^n (x_i - \bar{x}) \cdot f_i = 0$$

Como ventajas podemos nombrar las tres que se le exigen a una medida de síntesis:

- Considera todos los valores de la distribución.
- Es calculable.
- Es única.

Como inconvenientes podemos indicar:

- A veces da lugar a conclusiones no muy atinadas. Esto ocurre en el caso de que la variable presente valores anormalmente extremos que pueden distorsionar la media aritmética, haciéndola poco representativa.
- Por otra parte, si consideramos una variable discreta, por ejemplo, el número de hijos en las familias españolas, el valor de la media puede no pertenecer al conjunto de valores de la variable; por ejemplo $\bar{x} = 1,2$ hijos.

Otro tema al que tenemos que hacer referencia es el de la llamada Media Aritmética Ponderada, que se produce cuando se otorga a cada valor de la variable una ponderación o peso, distinto de la frecuencia o repetición. En este caso, en el cálculo de la media aritmética ponderada tendríamos en cuenta dichas ponderaciones. En este caso, si p_i son los pesos de cada valor de la variable x_i , entonces:

$$\bar{x}_P = \frac{x_1 p_1 + x_2 p_2 + \dots + x_n p_n}{p_1 + p_2 + \dots + p_n} = \frac{\sum_{i=1}^n x_i p_i}{\sum_{i=1}^n p_i}$$

Ejemplo: Un examen consta de tres partes, un test, una parte teórica y otra de ejercicios. El profesor le asigna una importancia del 20 % al test, de un 30 % a la teoría y de un 50 % a los ejercicios. Si un alumno ha obtenido un 4 en el test, un 3 en la teoría y un 7 en los ejercicios, la nota final del examen sería la media ponderada:

$$\bar{x}_P = \frac{4 \cdot 20 + 3 \cdot 30 + 7 \cdot 50}{20 + 30 + 50} = 5,2 \quad \text{¡Aprobado!}$$

De no haber ponderado, la nota media sería 4,67. ¡Suspenso!

Ejercicio 3: Un estudiante realiza 3 exámenes de complejidad creciente, obteniendo los siguientes resultados: 5, 8 y 7. El primer examen lo hizo en media hora, el segundo en una hora y el tercero en hora y media, por lo que se les atribuye una ponderación de 1, 2 y 3 respectivamente. Se pide calcular la nota media.

Solución: $\bar{x}_P = 7$

2.2. La Media Geométrica.

Definimos la Media Geométrica, y la representaremos por G , como la raíz n -ésima del producto de los n valores de la distribución. Así:

$$G = \sqrt[n]{x_1^{f_1} \cdot x_2^{f_2} \cdot \dots \cdot x_n^{f_n}}$$

Propiedades de la media geométrica

- El logaritmo de la media geométrica es igual a la media aritmética de los logaritmos de los valores de la variable.

Como ventajas podemos señalar:

- En su determinación intervienen todos los valores de la distribución.
- Es menos sensible que la media aritmética a los valores extremos, por su carácter de producto.

Como inconvenientes tenemos:

- Es de significado estadístico menos intuitivo que la media aritmética.
- Su cómputo es más difícil.
- En ocasiones no queda determinada. Esto ocurre cuando la variable toma en algún momento el

valor 0. Si la variable toma valores negativos, se pueden presentar una amplia gama de casos particulares en los que tampoco queda determinada G . No es que G no exista, sino que no la podemos determinar.

El empleo más frecuente de la media geométrica es el de promediar porcentajes, tasas, números índices, etc. Es decir, en los casos en los que se supone que la variable presenta variaciones acumulativas.

2.3. La Media Armónica.

Definimos la Media Armónica de una distribución de frecuencias, y se representa por H , como:

$$H = \frac{N}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \dots + \frac{f_n}{x_n}} = \frac{N}{\sum_{i=1}^n \frac{f_i}{x_i}}$$

Se suele utilizar para promediar velocidades, precios, rendimientos, cambio de divisas, etc... Todas ellas son situaciones donde existen en el contexto tres variables, siendo el producto de los valores de dos de ellas igual a los valores de la tercera (velocidades y tiempos a recorrer un determinado espacio, precios y cantidades relacionados con el valor de una mercancía, etc...)

Ejemplo: Imaginemos que hacemos un recorrido de 75 km y que los primeros 25 km los recorremos a una velocidad de 50 km/h, los 25 km siguientes a 70 km/h y los últimos 25 km a 90 km/h. Nos preguntamos, ¿qué velocidad media hemos conseguido?

La media aritmética da aparentemente una respuesta clara:

$$\bar{x} = \frac{50 + 70 + 90}{3} = 70 \text{ km/h}$$

Si miramos nuestro reloj veremos que han transcurrido desde el comienzo al final de nuestro recorrido 68 minutos y 5 segundos:

$$\frac{25}{50} + \frac{25}{70} + \frac{25}{90} = 1,1349 \text{ h} = 1 \text{ h } 8 \text{ min } 5 \text{ s}$$

Pero si la velocidad media que hubiésemos tenido hubiera sido de 70 km/h, en ese tiempo de 1 h 8 min 5 s habríamos recorrido 79,44 km.

¿Qué ocurre? O el reloj está estropeado, o la estadística falla. Pues bien, ni una cosa ni la otra. Lo que ocurre es que el procedimiento utilizado no es el adecuado. Se puede trabajar directamente con velocidades si utilizamos la media armónica.

$$H = \frac{3}{\frac{1}{50} + \frac{1}{70} + \frac{1}{90}} = 66,08 \text{ km/h}$$

A la velocidad de 66,08 km/h y estando en movimiento 1 h 8 min 5 s, recorreremos una distancia de 74,99 km como queríamos probar.

Nota: La media armónica solo se utiliza para recorridos proporcionalmente subdivididos. Por ejemplo si yo hago el estudio en un tramo de 10 km y luego quiero estudiar lo que pasa en los 60 km siguientes, tomaré siete medidas, una para cada tramo de 10 km. O sea, esos 60 km los tomo como 6 veces 10 km. Y en el cálculo de la expresión matemática quedaría así, siendo v_1 y v_2 las velocidades de cada tramo:

$$H = \frac{7}{\frac{1}{v_1} + \frac{6}{v_2}}$$

Como ventajas diremos que intervienen en su cálculo todos los valores de la distribución y que, en ciertos casos, es más representativa que la media aritmética. Por otra parte, siempre se puede pasar de una media armónica a una media aritmética transformando adecuadamente los datos.

Como inconvenientes hemos de citar la influencia de los valores pequeños, y su no determinación en las distribuciones con algunos valores iguales a cero. Por ello no es aconsejable su empleo en distribuciones en las que existan valores muy pequeños.

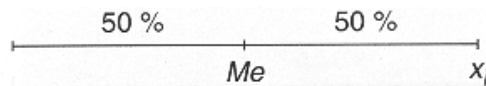
2.4. Relación entre los tres Promedios.

Para una misma distribución de frecuencias, y siempre que existan, se verifica que:

$$H \leq G \leq \bar{x}$$

2.5. La Mediana.

La Mediana es el valor de la distribución, supuesta ésta ordenada de menor a mayor, que deja a su izquierda y a su derecha el mismo número de frecuencias. Se representa por *Me*.



Si el número de datos es pequeño, es decir, trabajamos con distribuciones sin agrupar, los ordenamos y procedemos del siguiente modo:

- Cuando el número de datos es impar: tomamos el valor central. Si los valores son 4, 6, 4, 5, 7, 3, 9. Los ordenamos 3, 4, 4, 5, 6, 7, 9. Como son 7 datos cogemos el dato que ocupa el lugar central (4º) que es 5.
- Cuando el número de datos es par: tomamos la media de los dos valores que ocupan el centro de la distribución. Si los valores son 4, 6, 5, 7, 3, 9. Los ordenamos 3, 4, 5, 6, 7, 9. Como son 6 datos cogemos los datos que ocupan los lugares 3º, que es 5, y 4º, que es 6. La mediana es la media de los dos números es este caso $Me = (5 + 6) / 2 = 5,5$.

Ejercicio 4: Calcula la mediana de las siguientes distribuciones de frecuencias:

- a) 1, 5, 8, 6, 4, 8, 4 b) 0, 2, 5, 4, 8, 4, 4, 5, 9, 6

Solución: a) $Me = 5$; b) $Me = 4,5$

Si tenemos muchos datos, por ejemplo 60, ordenarlos es una tarea pesada, entonces lo que se hace es escribir los datos en forma de tabla, con las columnas de los valores x_i , frecuencia absoluta y frecuencia acumulada.

- Se empieza calculando $N/2$. La mediana es entonces el primer valor muestral x_i cuya frecuencia absoluta acumulada F_i supera a $N/2$.
- Si casualmente ocurre que hay un valor x_i cuya frecuencia absoluta acumulada F_i coincide justamente con $N/2$, la mediana es entonces la media aritmética entre ese valor y el siguiente, o sea:

$$Me = \frac{x_i + x_{i+1}}{2}$$

Ejemplo: Después de preguntar a 91 salmantinos cuántos vehículos a motor hay en sus casas, se han resumido las respuestas en la tabla de frecuencias siguiente, en la que X es la variable “número de vehículos a motor”:

x_i	f_i	F_i
0	20	20
1	35	55
2	18	73
3	12	85
4	6	91
$N = 91$		

En este caso $N/2 = 91/2 = 45,5$. Como el valor $x_2 = 1$ de la variable tiene una frecuencia absoluta acumulada $F_2 = 55$, y es el primero cuya frecuencia absoluta acumulada supera el valor 45,5, se deduce que $Me = x_2 = 1$.

Ejemplo: Tras recoger 88 datos relativos a una variable X , se han resumido en la tabla siguiente. A su lado aparece el cálculo de la mediana:

x_i	f_i	F_i
1	10	10
3	20	30
4	14	44
7	16	60
8	10	70
9	18	88
$N = 88$		

En este caso $N/2 = 88/2 = 44$. Si observamos la tabla vemos que el valor $x_2 = 1$ de la variable tiene precisamente una frecuencia absoluta acumulada $F_2 = 44$. Por eso la mediana es en este caso:

$$Me = \frac{x_3 + x_4}{2} = \frac{4 + 7}{2} = 5,5$$

Ejercicio 5: Las edades de los miembros de una asociación juvenil para la defensa del medio ambiente vienen dadas por:

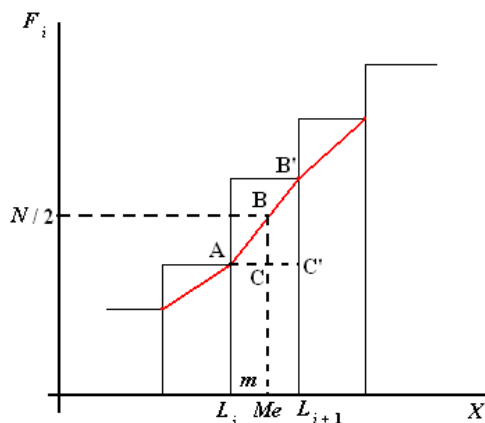
Edad (años)	13	14	15	16	17	18	19
Nº de jóvenes	10	21	14	25	37	41	22

Calcula la mediana.

Solución: $Me = 17$

En el caso de distribuciones agrupadas en intervalos, no es necesario distinguir si los intervalos se han construido de la misma o distinta amplitud. Siguiendo el método general de búsqueda del valor que ocupa el lugar $N/2$, nos podemos encontrar con los siguientes casos:

- Que $N/2$ se encuentre en la tabla. Entonces $N/2$ es la frecuencia absoluta acumulada de un cierto intervalo $[L_{i-1}, L_i)$, y la mediana será el extremo superior de dicho intervalo, L_i .
- Que $N/2$ no se encuentre en la tabla, esto es que $F_i < N/2 < F_{i+1}$. La mediana se encontrará en el intervalo $[L_{i-1}, L_i)$, al que llamaremos intervalo mediano.



Con el objeto de fijar la mediana en un valor, seleccionaremos un representante del intervalo mediano. El criterio usualmente seguido es el siguiente. Suponemos, en primer lugar, que todos los valores comprendidos dentro del intervalo mediano se encuentran distribuidos uniformemente a lo largo de él. A continuación, vamos a considerar la poligonal de frecuencias acumuladas correspondiente al intervalo mediano y a sus dos contiguos, y determinamos gráficamente la mediana.

En la figura vemos que $Me = L_i + m$. Determinaremos m en base a la hipótesis fijada que nos permite escribir:

$$\frac{\overline{AC}}{\overline{AC'}} = \frac{\overline{BC}}{\overline{B'C'}}$$

ya que los triángulos ABC y $AB'C'$ son semejantes. Pero:

$$\overline{AC} = m \quad ; \quad \overline{AC'} = L_{i+1} - L_i = c \quad ; \quad \overline{BC} = \frac{N}{2} - F_{Me-1}$$

Por tanto:

$$\overline{B'C'} = F_{Me} - F_{Me-1} = f_{Me} \quad \frac{m}{c} = \frac{\frac{N}{2} - F_{Me-1}}{f_{Me}}$$

Con lo que tenemos:

$$Me = L_i + \frac{\frac{N}{2} - F_{Me-1}}{f_{Me}} c$$

donde:

- L_i es el límite inferior de la clase mediana.
- c es la amplitud del intervalo mediano.
- N es el número total de datos.
- f_{Me} es la frecuencia absoluta de la clase mediana.
- F_{Me-1} es la frecuencia absoluta acumulada hasta llegar a la clase mediana sin incluirla.

Nota:

Ejemplo: Se ha medido la altura, en centímetros, de 15 estudiantes y los resultados se han agrupado en intervalos de clase, como aparece en la tabla siguiente, junto a la que se explica el modo de obtener la mediana:

Alturas	f_i	F_i
[160,170)	4	4
[170,180)	6	10
[180,190)	2	12
[190,200)	3	15
$N = 15$		

En este caso $N/2 = 15/2 = 7,5$. El intervalo mediano es [170,180) porque su frecuencia absoluta acumulada, 10, es la primera que está por encima de 7,5. Por lo tanto:

$$Me = 170 + \frac{7,5 - 4}{6} \cdot 10 = 175,83 \text{ cm}$$

Eso quiere decir que la mitad de los estudiantes miden menos de 175,83 cm y la otra mitad mide más.

Ejercicio 6: Los datos obtenidos en un control de velocidad realizado en una autovía han sido:

Velocidad (km/h)	[90, 100)	[100, 110)	[110, 120)	[120, 130)	[130, 140)
Nº de vehículos	16	15	35	25	10

Calcula la mediana.

Solución: $Me = 115,57$

A pesar de la fórmula que hemos estado viendo para el caso de distribuciones agrupadas en intervalos, la mediana tiene un mayor sentido en casos de distribuciones en escala ordinal (datos susceptibles de ser ordenados), de la cual es la medida más representativa por describir la tendencia central de la misma.

Entre las ventajas de la mediana, vamos a destacar las siguientes:

- Como medida descriptiva, tiene la ventaja de no estar afectada por las observaciones extremas, ya que no depende de los valores que toma la variable, sino del orden de las mismas. Por ello es adecuado su uso en distribuciones asimétricas.
- Es de cálculo rápido y de interpretación sencilla.
- A diferencia de la media, la mediana de una variable discreta es siempre un valor de la variable que estudiamos (ej. La mediana de una variable *número de hijos* toma siempre valores enteros).

El mayor defecto de la mediana es que tiene unas propiedades matemáticas complicadas, lo que hace que sea muy difícil de utilizar en *inferencia estadística*.

2.6. La Moda.

Llamaremos moda de una variable estadística al valor de la variable que tiene mayor frecuencia absoluta. Se representa por M_o .

Para calcular la moda, distinguiremos entre distribuciones no agrupadas en intervalos y distribuciones agrupadas en intervalos.

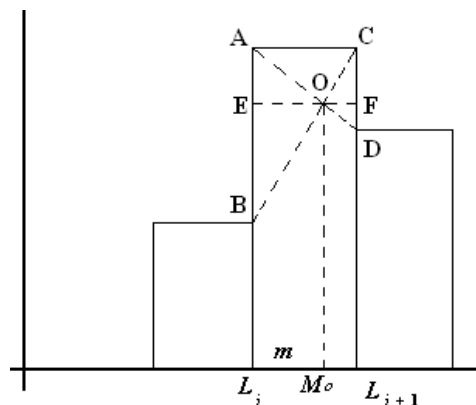
En el caso de distribuciones no agrupadas en intervalos, la determinación de la moda, M_o , es inmediata. Se observa la columna de las frecuencias absolutas y el valor de la distribución al que corresponde la mayor frecuencia será la moda. Puede ocurrir que la moda no sea única, es decir, la distribución puede tener 2, 3 o más modas, recibiendo el nombre de bimodal, trimodal, etc.

En el caso de que los datos se encuentren agrupados en intervalos de igual amplitud, la clase con mayor densidad de frecuencia se denomina clase modal, y se corresponde con el intervalo que tenga mayor altura en el histograma.

Si se desea mayor precisión en el cálculo de la moda, esta puede obtenerse suponiendo que todos los valores del intervalo están distribuidos uniformemente dentro de él. Así, la moda estará más cerca de aquel intervalo contiguo cuya frecuencia sea mayor. Para calcularla se suele utilizar la siguiente construcción geométrica.

La moda será $M_o = L_i + m$. Pero los triángulos AOB y COD son semejantes por tener sus ángulos iguales. Al ser semejantes, las alturas de éstos triángulos son proporcionales a sus bases, es decir:

$$\frac{\overline{EO}}{\overline{OF}} = \frac{\overline{AB}}{\overline{CD}}$$



Que teniendo en cuenta las propiedades de las proporciones queda:

$$\frac{\overline{EO}}{\overline{EO} + \overline{OF}} = \frac{\overline{AB}}{\overline{AB} + \overline{CD}}$$

Como:

$$\overline{EO} + \overline{OF} = c \qquad \overline{AB} = f_{M_o} - f_{M_o-1} \qquad \overline{AB} + \overline{CD} = (f_{M_o} - f_{M_o-1}) + (f_{M_o} - f_{M_o+1})$$

Se tiene que:

$$M_o = L_i + \frac{(f_{M_o} - f_{M_o-1})}{(f_{M_o} - f_{M_o-1}) + (f_{M_o} - f_{M_o+1})} c_{M_o}$$

donde:

- L_i es el límite inferior de la clase modal.
- c_{M_o} es la amplitud del intervalo modal.
- $f_{M_o}, f_{M_o - 1}, f_{M_o + 1}$ son, respectivamente, las frecuencias absolutas de la clase modal, la clase anterior y la posterior.

Ejemplo: Para el caso de la muestra de alturas de estudiantes visto anteriormente, el intervalo modal es [170,180), y por lo tanto la moda será:

$$M_o = 170 + \frac{(6-4)}{(6-4)+(6-2)} \cdot 10 = 173,33 \text{ cm}$$

Pero en el caso de que los datos se encuentren agrupados en intervalos de distinta amplitud, la clase con mayor densidad de frecuencia se denomina clase modal, y ahora puede que no se corresponda con el intervalo que tenga mayor altura en el histograma. Un razonamiento similar al que hemos hecho anteriormente para intervalos de igual amplitud no lleva a una expresión para la moda que es:

$$M_o = L_i + \frac{(d_{M_o} - d_{M_o-1})}{(d_{M_o} - d_{M_o-1}) + (d_{M_o} - d_{M_o+1})} c_{M_o}$$

donde:

- L_i es el límite inferior de la clase modal.
- c_{M_o} es la amplitud del intervalo modal.
- $d_{M_o}, d_{M_o - 1}, d_{M_o + 1}$ son, respectivamente, las densidades de frecuencia ($d_{M_o} = f_{M_o} / c_{M_o}$) de la clase modal, la clase anterior y la posterior.

Ejemplo: Consideremos la siguiente distribución de frecuencias agrupadas en intervalos (de distinta amplitud):

Variable	[0, 10)	[10, 30)	[30, 40)	[40, 70)	[70, 90)	[90, 100)
Frecuencia	30	50	35	90	40	20

En este caso los intervalos no son de igual amplitud. Por tanto la clase modal será aquella que tenga mayor densidad de frecuencia.

Variable	Frecuencia	Amplitud	Densidad de frecuencia
[0, 10)	30	10	3
[10, 30)	50	20	2,5
[30, 40)	35	10	3,5
[40, 70)	90	30	3
[70, 90)	40	20	2
[90, 100)	20	10	2

Como se observa en la tabla, la clase con mayor densidad de frecuencia es [30, 40). Esta es la clase modal. Así la moda vendrá dada por:

$$M_o = 30 + \frac{(3,5 - 2,5)}{(3,5 - 2,5) + (3,5 - 3)} \cdot 10 = 36, \hat{6}$$

De la moda destacamos las siguientes propiedades:

- Es muy fácil de calcular.
- Puede no ser única.
- Es función de los intervalos elegidos a través de su amplitud, número y límites de los mismos.
- Aunque el primero o el último de los intervalos no posean extremos inferior o superior respectivamente, la moda puede ser calculada.

Por último diremos que la moda es la medida más representativa en caso de distribuciones en escala nominal. Esto es debido a que las distribuciones de este tipo presentan los datos no susceptibles de ordenación, de tal forma que para estas distribuciones no es posible realizar operaciones elementales con sus observaciones.

Ejercicio 7: Los datos obtenidos en un control de velocidad realizado en una autovía han sido:

Velocidad (km/h)	[90, 100)	[100, 110)	[110, 120)	[120, 130)	[130, 140)
Nº de vehículos	16	15	35	25	10

Calcula la moda.

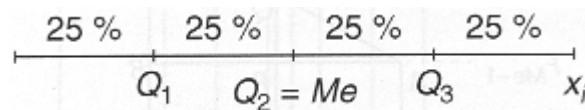
Solución: $Mo = 116,67$

2.7. Medidas de Posición no Centrales.

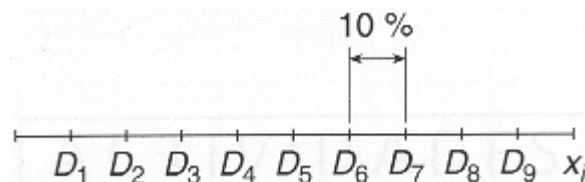
Podemos nombrar otros valores notables pero que no van a reflejar ninguna tendencia central: los Cuantiles. Son valores de la distribución que la dividen en partes iguales, es decir, en intervalos, que comprenden el mismo número de valores.

Entre los Cuantiles podemos citar, por ser de uso más frecuente:

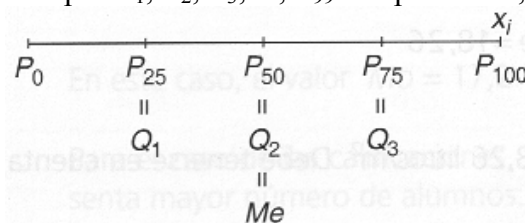
- Llamaremos cuartiles a los tres valores de la distribución que la dividen en cuatro partes iguales. Es decir, en cuatro intervalos dentro de cada cual están incluidos el 25 % de los valores de la distribución. Se denotan por Q_1 , Q_2 y Q_3 . En particular, se tiene que $Me = Q_2$.



- Llamaremos deciles a los nueve valores de la distribución que la dividen en diez partes iguales. Cada parte contendrá el 10 % de la distribución. Se representan por D_1 , D_2 , ... y D_9 . En particular, $Me = Q_2 = D_5$.



- Llamaremos percentiles a los noventa y nueve valores que dividen a la distribución en cien partes iguales. Se denotan por P_1 , P_2 , P_3 , ..., P_{99} . En particular, $Me = Q_2 = D_5 = P_{50}$.



Para obtener cada uno de estas medidas de posición no centrales, se procede como en el caso de la mediana. Para ello es necesario tener ordenada la muestra de menor a mayor.

a) Distribuciones no agrupadas en intervalos.

- Cuartiles: C_i es el valor que ocupa el lugar $\frac{iN}{4}$, con $i = 1, 2 \text{ ó } 3$.
- Deciles: D_i es el valor que ocupa el lugar $\frac{iN}{10}$ con $i = 1, \dots, 9$.
- Percentiles: P_i es el valor que ocupa el lugar $\frac{iN}{100}$ con $i = 1, \dots, 99$.

b) Distribuciones agrupadas en intervalos.

El problema que se presenta es el mismo que el que teníamos al calcular la mediana. Así:

$$\text{Cuartiles: } Q_k = L_i + \frac{\frac{kN}{4} - F_{Q_{k-1}}}{f_{Q_k}} \cdot c \quad k = 1, 2, 3.$$

$$\text{Deciles: } D_k = L_i + \frac{\frac{kN}{10} - F_{D_{k-1}}}{f_{D_k}} \cdot c \quad k = 1, 2, \dots, 9.$$

$$\text{Percentiles: } P_k = L_i + \frac{\frac{kN}{100} - F_{P_{k-1}}}{f_{P_k}} \cdot c \quad k = 1, 2, \dots, 99.$$

Las fórmulas anteriores se obtienen de forma análoga a la desarrollada para la mediana.

Ejemplo: La distribución de frecuencias de las calificaciones de una prueba de habilidad numérica se recoge en la tabla adjunta:

Calificación	[30, 40)	[40, 50)	[50, 60)	[60, 70)	[70, 80)	[80, 90)	[90, 100)
Nº de estudiantes	1	3	11	21	43	32	9

Halla los cuartiles de la distribución e interpreta su significado.

La tabla de frecuencias absolutas acumuladas es:

Calificación	Nº. de estudiantes (frecuencia absoluta)	F. absoluta acumulada
[30, 40)	1	1
[40, 50)	3	4
[50, 60)	11	15
[60, 70)	21	36
[70, 80)	43	79
[80, 90)	32	111
[90, 100)	9	120

En primer lugar, tengamos en cuenta que $N = 120$. Así:

$$\frac{N}{4} = \frac{120}{4} = 30 \quad \Rightarrow \quad Q_1 = 60 + \frac{30 - 15}{21} \cdot 10 = 60 + 7,14 = 67,14$$

$$\frac{2N}{4} = \frac{240}{4} = 60 \Rightarrow Q_2 = 70 + \frac{60-36}{43} \cdot 10 = 70 + 5,58 = 75,58$$

$$\frac{3N}{4} = \frac{360}{4} = 90 \Rightarrow Q_3 = 80 + \frac{90-79}{32} \cdot 10 = 80 + 3,34 = 83,34$$

Se tiene que entre 30 y 67 se encuentra el 25 % de estudiantes, de igual forma entre 67 y 75, entre 75 y 83 y entre 83 y 100.

Ejercicio 8: Una distribución viene dada por la siguiente tabla:

x_i	3	4	5	6	7	8	9
f_i	3	9	12	16	10	6	4

Calcula el primer cuartil y el tercer cuartil, el decil 2 y el percentil 65.

Solución: $Q_1 = 5$; $Q_3 = 7$; $D_2 = 4,5$; $P_{65} = 6$

Ejercicio 9: La siguiente tabla presenta el número de horas semanales que dedican al estudio 30 alumnos de 1º de bachillerato:

Nº de horas	[0, 4)	[4, 8)	[8, 12)	[12, 16)
Nº de alumnos	8	10	8	4

Calcula el tercer cuartil, el decil 4 y el percentil 82.

Solución: $Q_3 = 10,25$; $D_4 = 5,6$; $P_{82} = 11,3$

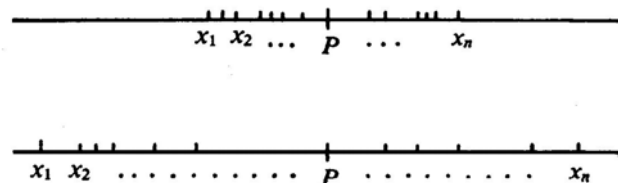
3. MEDIDAS DE DISPERSIÓN

Las medidas de centralización vistas anteriormente necesitan de otras que las complementen en el estudio de las distribuciones de frecuencias de las variables estadísticas.

Estas nuevas medidas, que denominaremos parámetros de dispersión, informan de las desviaciones que sufren los datos respecto de los valores centrales, en especial con relación a la media aritmética.

En este apartado, veremos hasta qué punto, para una distribución de frecuencias, estas medidas de tendencia central son representativas como síntesis de toda la información. Medir la representatividad de estas medidas equivale a cuantificar la separación de los valores de la distribución respecto de dicha medida.

Ejemplo: Supongamos que tenemos un promedio P del que vamos a estudiar su representatividad. Consideremos que tenemos dos distribuciones que originan este mismo promedio P (supongámoslas de frecuencias unitarias por sencillez) y que son tales como las que se representan en el siguiente gráfico:



Si queremos saber cual de los dos promedios es más representativo, a simple vista parece que el primero, porque el error que se comete utilizando P (en lugar de los valores de la distribución) es menor en la primera que en la segunda. Luego, cuanto más agrupados estén los valores alrededor del promedio, más útil será.

Para una mejor clasificación, vamos a distinguir entre medidas de dispersión absolutas y relativas.

3.1. Medidas de dispersión absolutas.

3.1.1. Rango o Recorrido.

Llamaremos recorrido a la diferencia entre el mayor valor y el menor valor de una distribución. Se representa por R .

$$R = x_{\max} - x_{\min}$$

3.1.2. Recorrido Intercuartílico.

El Recorrido Intercuartílico, es la diferencia existente entre el tercer cuartil y el primer cuartil. Se representa por R_1 :

$$R_1 = Q_3 - Q_1$$

El recorrido intercuartílico nos indica que en el intervalo de longitud R_1 están comprendidos el 50% central de los valores. Si R_1 es pequeño, siempre en términos relativos de acuerdo con las unidades en que venga dada la distribución, podemos intuir una pequeña dispersión.

Ejercicio 10: Una distribución viene dada por la siguiente tabla (ver ejercicio anterior):

x_i	3	4	5	6	7	8	9
f_i	3	9	12	16	10	6	4

Calcula el recorrido y el recorrido intercuartílico.

Solución: $R = 9 - 3 = 6$; $R_1 = 7 - 5 = 2$

Valores atípicos o alejados.

Un valor de la variable estadística se dice que es atípico, o que está alejado, cuando se encuentra muy separado del resto de los valores que toma esa variable según el siguiente criterio:

$$\begin{aligned} x \text{ es atípico por la derecha si } & x > Q_3 + 1,5(Q_3 - Q_1) \\ x \text{ es atípico por la izquierda si } & x < Q_1 - 1,5(Q_3 - Q_1) \end{aligned}$$

Ejemplo: Comprueba si hay algún dato atípico en el siguiente conjunto de datos:

600, 396, 490, 1604, 8, 606, 604, 594, 1246, 42

Si ordenamos los datos tenemos que:

8, 42, 396, 490, 594, 600, 604, 606, 1246, 1604

Así, tendremos que:

$$Q_1 = 396 \quad \text{y} \quad Q_3 = 606$$

Por tanto un valor será atípico por la derecha si es mayor que:

$$Q_3 + 1,5 \cdot (Q_3 - Q_1) = 606 + 1,5 \cdot (606 - 396) = 921$$

Y un valor será atípico por la izquierda si es menor que:

$$Q_1 - 1,5 \cdot (Q_3 - Q_1) = 396 - 1,5 \cdot (606 - 396) = 81$$

Entonces, los valores atípicos por la derecha son 1246 y 1604 y los valores atípicos por la izquierda son 8 y 42.

Ejercicio 11: ¿Hay algún dato atípico en el siguiente conjunto de datos?

63, 62, 60, 20, 65, 80, 82, 110, 70, 75, 73, 72, 108, 84, 78, 67, 19, 60, 61, 63

Solución: 19, 20, 108 y 110 son datos atípicos.

3.1.3. Desviación media.

Se denomina desviación media de una variable estadística, a la media de los valores de las desviaciones de los datos (o marcas de clase) respecto a la media aritmética. Se representa por DM .

La expresión que permite calcular la desviación media es la siguiente:

$$DM = \frac{\sum_{i=1}^n |x_i - \bar{x}| \cdot f_i}{N}$$

3.1.4. La Varianza.

De todas las medidas de dispersión absolutas respecto a la media aritmética, la varianza y su raíz cuadrada, la desviación típica, son las más importantes.

Llamamos varianza de una variable estadística a la media aritmética de los cuadrados de las desviaciones de los valores de la variable respecto de la media aritmética. Se representa por σ^2 .

Las expresiones equivalentes que permiten calcular la varianza son:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot f_i}{N} \qquad \sigma^2 = \frac{\sum_{i=1}^n x_i^2 \cdot f_i}{N} - \bar{x}^2$$

Evidentemente, σ^2 nos medirá la mayor o menor dispersión de los valores respecto a la media aritmética. Si la dispersión es muy grande, la media no será representativa.

En el caso extremo de que todos los valores de la variable fuesen iguales, la media coincidiría con el valor común de las mismas y las desviaciones serían todas nulas, dando $\sigma^2 = 0$. En general, cuanto más dispersas sean las observaciones, mayores serán las desviaciones respecto de la media, y mayor el valor numérico de la varianza.

A continuación enunciamos unas propiedades que verifica la varianza.

- La varianza nunca es negativa.
- La varianza es la medida de dispersión cuadrática óptima por ser la menor de todas.
- Si en la distribución de frecuencias sumamos a todos los valores de la variable una constante k , la varianza no varía.
- Al multiplicar los valores de una distribución de frecuencias por una constante k , la varianza queda multiplicada por el cuadrado de la constante.

3.1.5. Desviación Típica.

Se llama desviación típica de una variable estadística a la raíz cuadrada, con signo positivo, de la varianza. Se representa por σ , y es:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot f_i}{N}} = \sqrt{\frac{\sum_{i=1}^n x_i^2 \cdot f_i}{N} - \bar{x}^2}$$

Al ser la raíz cuadrada de la varianza, vendrá expresada en las mismas unidades de medida que la distribución, lo cual la hace más apta como medida de dispersión.

Sus propiedades las podemos deducir fácilmente de las propiedades de la varianza.

- La desviación típica nunca es negativa.
- Es la medida de dispersión óptima por ser la más pequeña.
- Si en la distribución de frecuencias sumamos a todos los valores de la variable una constante k , la desviación típica no varía.
- Al multiplicar los valores de una distribución de frecuencias por una constante k , la desviación típica queda multiplicada por dicha constante.

Ejemplo: Consideremos las muestras de notas obtenidas por cinco alumnos en Historia y Matemáticas:

Historia: 4, 5, 5, 5, 6

Matemáticas: 1, 1, 5, 9, 9

Es inmediato comprobar que la media aritmética en ambos casos es 5. Pero es evidente que el comportamiento que se observa en las dos asignaturas es muy distinto. Vamos a obtener la varianza y la desviación típica de las notas en Historia y Matemáticas.

$$\text{Historia: } \sigma^2 = \frac{(4-5)^2 + (5-5)^2 + (5-5)^2 + (5-5)^2 + (6-5)^2}{5} = 0,4 \quad \text{y} \quad \sigma = \sqrt{0,4} = 0,632$$

$$\text{Matemáticas: } \sigma^2 = \frac{(1-5)^2 + (1-5)^2 + (5-5)^2 + (9-5)^2 + (9-5)^2}{5} = 12,8 \quad \text{y} \quad \sigma = \sqrt{12,8} = 3,577$$

Los resultados confirman, como era de esperar, que las notas en Matemáticas están más dispersas.

Ejercicio 12: Una distribución viene dada por la siguiente tabla:

x_i	3	4	5	6	7	8	9
f_i	4	8	10	18	8	7	5

Calcula la varianza y la desviación típica.

Solución: $\bar{x} = 5,98$; $\sigma^2 = 2,69$; $\sigma = 1,64$

Ejercicio 13: La siguiente tabla preséntale número de horas semanales que dedican al estudio 30 alumnos de 1º de bachillerato:

Nº de horas	[0, 4)	[4, 8)	[8, 12)	[12, 16)
Nº de alumnos	8	10	8	4

Calcula la varianza y la desviación típica.

Solución: $\bar{x} = 7,07$; $\sigma^2 = 15,93$; $\sigma = 3,99$

3.2. Medidas de dispersión relativas.

Supongamos que tenemos dos distribuciones de frecuencias cuyas medias son \bar{x}_1 y \bar{x}_2 y queremos saber cuál de las dos es más representativa. Esta comparación no la podemos efectuar por sus respectivas medidas de dispersión, ya que las distribuciones, en general, no vendrán dadas en las mismas unidades de medida. Tampoco se podrá efectuar en el caso de que las unidades de medida sean las mismas, si los promedios son numéricamente diferentes.

Por tanto, resulta necesario construir medidas adimensionales. Estas medidas de dispersión, llamadas relativas, siempre vendrán dadas en forma de cociente.

3.2.1. Coeficiente de variación.

El coeficiente de variación de una variable estadística es el cociente entre la desviación típica y el valor absoluto de la media aritmética. Se representa por CV . Así, se tiene:

$$CV = \frac{\sigma}{|\bar{x}|}$$

El valor del coeficiente de variación suele expresarse en tanto por ciento. Cuanto más pequeño sea este coeficiente de variación, los datos están más concentrados alrededor de la media, y esta será más representativa. A mayor coeficiente de variación, más alejamiento y dispersión respecto de los valores centrales.

Este coeficiente permite comparar dos poblaciones heterogéneas. Si x e y son dos variables estadísticas cuyas medias son \bar{x} e \bar{y} , y sus desviaciones típicas σ_x y σ_y , se tiene:

- Si $\bar{x} = \bar{y}$, $\sigma_x < \sigma_y \Rightarrow \bar{x}$ es más representativa.
- Si $\bar{x} \neq \bar{y}$, $\frac{\sigma_x}{|\bar{x}|} < \frac{\sigma_y}{|\bar{y}|} \Rightarrow \bar{x}$ es más representativa.

Ejemplo: Supongamos que hemos tallado y pesado a los alumnos del primer curso de bachillerato del IES “Ramón Olleros” y después de calcular la media y la desviación típica de esas medidas se obtuvo:

Variable X = talla	Variable Y = peso
$\bar{x} = 1,70$ m	$\bar{y} = 69$ kg
$\sigma_x = 0,5$ m	$\sigma_y = 5$ kg

Nos preguntamos en cual de las dos variables hay mayor dispersión. Para ello, utilizamos el coeficiente de variación:

$$CV(\text{tallas}) = \frac{\sigma_x}{|\bar{x}|} = \frac{0,5}{1,70} \cdot 100 = 29,41 \%$$

$$CV(\text{pesos}) = \frac{\sigma_y}{|\bar{y}|} = \frac{5}{69} \cdot 100 = 7,2 \%$$

Por tanto, la variable “pesos” está menos dispersada que la variable “tallas”.

Ejemplo: Se han pesado 430 magdalenas de entre las que fabrica una determinada empresa. Los valores obtenidos se han tabulado usando seis intervalos de clase dando lugar a la tabla de frecuencias siguiente:

Peso (X)	Marcas (x_i)	f_i	F_i
[25,30)	27,5	73	73
[30,35)	32,5	80	153
[35,40)	37,5	67	220
[40,45)	42,5	98	318
[45,50)	47,5	44	362
[50,55)	52,5	68	430
		$N = 430$	

Vamos a determinar la media, la mediana, la moda, la varianza y la desviación típica de la variable $X =$ “peso en gramos de las magdalenas”:

Media:

$$\bar{x} = \frac{27,5 \cdot 73 + 32,5 \cdot 80 + 37,5 \cdot 67 + 42,5 \cdot 98 + 47,5 \cdot 44 + 52,5 \cdot 68}{430} = \frac{16945}{430} = 39,4 \text{ gramos}$$

Mediana:

Como $\frac{N}{2} = \frac{430}{2} = 215$, el intervalo mediano es el $[35,40)$, luego:

$$Me = L_i + \frac{\frac{N}{2} - F_{Me-1}}{f_{Me}} \cdot c = 35 + \frac{215 - 153}{67} \cdot 5 = 39,62 \text{ gramos}$$

Moda:

El intervalo modal es el $[40,45)$, luego:

$$Mo = L_i + \frac{f_{Mo} - f_{Mo-1}}{(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})} \cdot c = 40 + \frac{98 - 67}{(98 - 67) + (98 - 44)} \cdot 5 = 41,82 \text{ gramos}$$

Por lo tanto, la magdalenas pesan 39,4 gramos de promedio, la mitad pesa menos de 39,62 gramos y la otra mitad pesa más, y el peso más frecuente entre las magdalenas es 41,82 gramos.

Varianza:

$$\sigma^2 = \frac{(27,5 - 39,4)^2 \cdot 73 + (32,5 - 39,4)^2 \cdot 80 + \dots + (52,5 - 39,4)^2 \cdot 68}{430} = 69,5 \text{ gramos}^2$$

Desviación típica:

$$\sigma = \sqrt{\sigma^2} = \sqrt{69,5} = 8,33 \text{ gramos}$$

Coficiente de variación:

$$CV = \frac{8,33}{39,4} \cdot 100 = 21,14 \%$$

Ejercicio 14: Calcula el coeficiente de variación de las distribuciones de frecuencias de los dos últimos ejercicios anteriores.

Solución: $CV_{12} = 0,27$; $CV_{13} = 0,56$

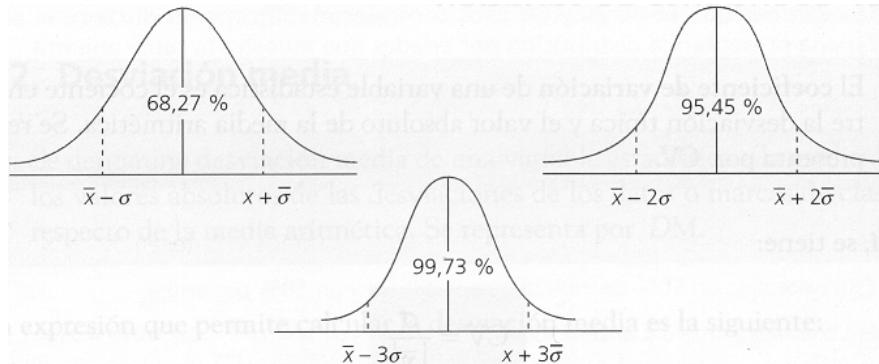
4. ESTUDIO CONJUNTO DE LA MEDIA ARITMÉTICA Y LA DESVIACIÓN TÍPICA

La media aritmética, \bar{x} , y la desviación típica, σ , son los parámetros estadísticos por antonomasia. La media es la medida central más utilizada y la desviación típica es la medida de dispersión o variabilidad por excelencia. En toda distribución estadística, el estudio del comportamiento conjunto de la media aritmética y la desviación típica nos aporta numerosa información sobre la distribución de frecuencias estudiada.

En casi todas las distribuciones estadísticas de comportamiento normal se verifican de forma aproximada los porcentajes descritos a continuación que, referidos a la media y la desviación típica,

expresan la distribución de datos. En una distribución estadística de comportamiento **normal**, se cumple:

- En $(\bar{x} - \sigma, \bar{x} + \sigma)$ está el 68,27 % del total de individuos.
- En $(\bar{x} - 2\sigma, \bar{x} + 2\sigma)$ está el 95,45% del total de individuos.
- En $(\bar{x} - 3\sigma, \bar{x} + 3\sigma)$ está el 99,73% del total de individuos.



Ejemplo: En la siguiente distribución de frecuencias:

X	[60, 76)	[76, 92)	[92, 108)	[108, 124)	[124, 140)	[140, 156)
Frecuencia	12	13	18	19	11	7

¿Cuántos valores hay en el intervalo $(\bar{x} - \sigma, \bar{x} + \sigma)$? ¿Qué porcentaje del total representan?

Calculemos la media y la desviación típica de la distribución:

$$\bar{x} = \frac{68 \cdot 12 + 84 \cdot 13 + 100 \cdot 18 + 116 \cdot 19 + 132 \cdot 11 + 148 \cdot 7}{80} = \frac{8400}{80} = 105$$

$$\sigma^2 = \frac{68^2 \cdot 12 + 84^2 \cdot 13 + 100^2 \cdot 18 + 116^2 \cdot 19 + 132^2 \cdot 11 + 148^2 \cdot 7}{80} - 105^2 = 573,4 \Rightarrow \sigma = 23,95$$

En el intervalo $(105 - 23,95, 105 + 23,95) = (81,05; 128,95)$ se encuentran, aproximadamente:

$$9 + 18 + 19 + 3 = 49 \text{ valores}$$

Estos valores representan el $\frac{49}{80} \cdot 100 = 61,25 \%$ del total.

Ejercicio 15: Una distribución viene dada por la siguiente tabla:

x_i	3	4	5	6	7	8	9
f_i	2	5	11	15	10	6	2

Calcula la media y la desviación típica y halla el porcentaje de datos incluidos en los intervalos:

$$(\bar{x} - \sigma, \bar{x} + \sigma) \quad ; \quad (\bar{x} - 3\sigma, \bar{x} + 3\sigma) \quad ; \quad (\bar{x} - 3\sigma, \bar{x} + 3\sigma)$$

Solución: $\bar{x} = 6,019$; $\sigma = 1,4$; 70,6 %, 92,2 % y 100 % respectivamente.

5. PUNTUACIONES TÍPICAS O NORMALIZADAS

Para poder comparar dos datos correspondientes a dos distribuciones distintas, hay que tipificar (o normalizar) dichos valores, es decir, calcular los valores $z = \frac{x - \bar{x}}{\sigma}$, y después, comparar los resultados.

Las puntuaciones típicas o normalizadas, también llamadas puntuaciones Z , tienen las siguientes propiedades:

- Si se transforma una distribución en puntuaciones típicas, no varía la forma de la distribución original.
- La media aritmética de las puntuaciones normalizadas es cero, es decir, $\bar{z} = 0$.
- La desviación típica de las puntuaciones típicas es la unidad, es decir, $\sigma_z = 1$.

Ejemplo: En un examen, un determinado alumno obtiene una nota de 6,5 y el conjunto de toda la clase $\bar{x} = 5,5$ y $\sigma = 2$. En otro examen, el alumno saca 7 puntos, siendo las calificaciones de la clase $\bar{x} = 6$ y $\sigma = 1$. ¿Cuál de las calificaciones es mejor respecto de la clase?

Para poder comparar ambas calificaciones, calculamos sus correspondientes puntuaciones típicas.

Para $x = 6,5$ tenemos:

$$z_{6,5} = \frac{6,5 - 5,5}{2} = 0,5$$

Para $x = 7$ tenemos:

$$z_7 = \frac{7 - 6}{1} = 1$$

Se observa que es mejor puntuación la segunda que la primera.

Ejercicio 16: En una clase hay 15 alumnos y 20 alumnas. El peso medio de los 15 alumnos es de 58,2 kg, y el de las 20 alumnas de 52,4 kg. Supongamos que las desviaciones típicas de los dos grupos son, respectivamente, 3,1 kg y 5,1 kg. El peso de Juan es de 70 kg y el de Pilar es de 65 kg. ¿Cuál de ellos se puede, dentro del grupo de alumnos de su sexo, considerarse más grueso?

Solución: $z_{Juan} = 3,81$; $z_{Pilar} = 2,47 \Rightarrow$ Juan se puede considerarse más grueso, dentro del grupo de alumnos de su sexo.

6. MEDIDAS DE FORMA

Existen otras medidas que nos permiten caracterizar la forma de la distribución. Estudiaremos los coeficientes de asimetría y de apuntamiento.

6.1 Coeficientes de asimetría.

Hacen referencia a la forma de la distribución, simétrica, asimetría a la derecha o a la izquierda. Diremos que una distribución es simétrica cuando los valores de la variable estadística equidistantes de un valor central tiene la misma frecuencia. En general la mejor manera de verlo es por la representación gráfica, pero si no la tenemos existen coeficientes que nos indican la forma de la distribución. Los más utilizados son:

6.1.1 Coeficiente de asimetría de Pearson.

Sólo se puede utilizar en distribuciones campaniformes (forma de campana) y unimodales. Cuando la moda no es única no tiene sentido calcularlo. Se define como:

$$A_p = \frac{\bar{x} - Mo}{\sigma}$$

Este coeficiente puede ser:

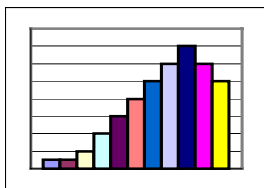
- $A_p = 0$ entonces la media igual que la moda, distribución simétrica.
- $A_p > 0$ entonces la media mayor que la moda, asimetría a la derecha.
- $A_p < 0$ entonces la media menor que la moda asimetría a la izquierda.

6.1.2 Coeficiente de asimetría de Fisher.

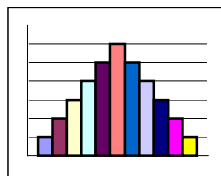
Se define el coeficiente de asimetría de Fisher como:

$$a_3 = \frac{1}{\sigma^3} \cdot \frac{\sum_{i=1}^n (x_i - \bar{x})^3 \cdot f_i}{N}$$

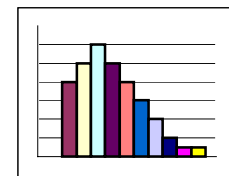
Como $(x_i - \bar{x})^3$ puede ser positivo o negativo, este coeficiente puede ser positivo o negativo; luego, atendiendo al signo, siempre que sea una distribución unimodal tenemos los casos:



$a_3 < 0$ Asimetría a la izquierda.



$a_3 = 0$ Distribución simétrica



$a_3 > 0$ Asimetría a la derecha

Este coeficiente tiene la ventaja de que se puede hallar para todas las distribuciones, aunque su cálculo es complicado y laborioso. Además no depende de las unidades de medida de las variables y es invariante por cambio de escala.

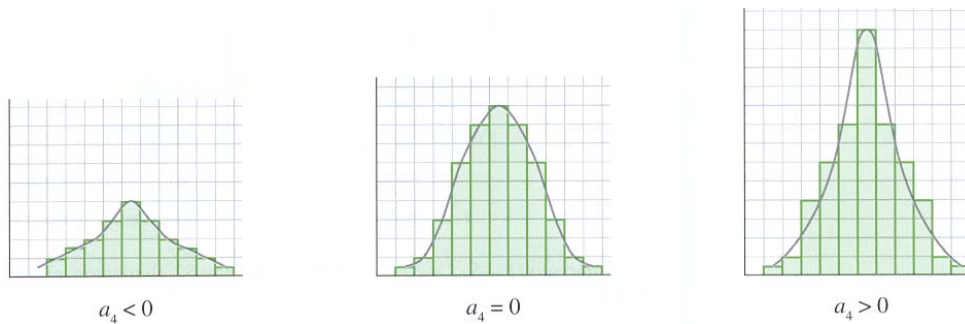
6.2 Coeficiente de apuntamiento o curtosis.

Hace referencia al mayor o menor apuntamiento que tiene una distribución de frecuencias respecto a una distribución Normal, por lo tanto sólo se estudia en distribuciones campaniformes, para compararlas con la campana de Gauss. Su cálculo también es muy laborioso.

$$a_4 = \left(\frac{1}{\sigma^4} \cdot \frac{\sum_{i=1}^n (x_i - \bar{x})^4 \cdot f_i}{N} \right) - 3$$

Este coeficiente puede ser:

- $a_4 = 0$ la curva es igual que la normal, se llama Mesocúrtica.
- $a_4 > 0$ la curva es más puntiaguda que la normal se llama Leptocúrtica.
- $a_4 < 0$ la curva es más aplastada que la normal, se llama Platicúrtica.



Ejemplo: Se eligen, al azar, 20 alumnos de una determinada facultad y se les pregunta cuántas asignaturas superaron el curso académico anterior. Los resultados obtenidos fueron estos:

Nº de asignaturas	1	2	3	4	5	6
Nº de alumnos	2	7	5	3	2	1

Estudiar la simetría (coeficientes de Pearson y de Fisher) y el apuntamiento de la distribución.

Para estudiar la simetría a través del coeficiente de Pearson necesitamos calcular la media, la moda y la desviación típica. Calculemoslas:

$$\bar{x} = \frac{1 \cdot 2 + 2 \cdot 7 + 3 \cdot 5 + 4 \cdot 3 + 5 \cdot 2 + 6 \cdot 1}{20} = 2,95 \qquad Mo = 2$$

$$\sigma^2 = \frac{1^2 \cdot 2 + 2^2 \cdot 7 + 3^2 \cdot 5 + 4^2 \cdot 3 + 5^2 \cdot 2 + 6^2 \cdot 1}{20} - 2,95^2 = 1,75 \quad \Rightarrow \quad \sigma = 1,32$$

Por tanto:

$$A_p = \frac{\bar{x} - Mo}{\sigma} = \frac{2,95 - 2}{1,32} = 0,71 > 0 \quad \Rightarrow \quad \text{Asimétrica a la derecha}$$

Calculemos ahora el coeficiente de asimetría de Fisher:

$$a_3 = \frac{1}{\sigma^3} \cdot \frac{\sum_{i=1}^n (x_i - \bar{x})^3 \cdot f_i}{N} =$$

$$= \frac{1}{1,32^3} \cdot \frac{(1 - 2,95)^3 \cdot 2 + (2 - 2,95)^3 \cdot 7 + (3 - 2,95)^3 \cdot 5 + (4 - 2,95)^3 \cdot 3 + (5 - 2,95)^3 \cdot 2 + (6 - 2,95)^3 \cdot 1}{20} =$$

$$= \frac{1}{2,31} \cdot 1,41 = 0,61 > 0 \quad \Rightarrow \quad \text{Asimétrica a la derecha}$$

Calculemos ahora el coeficiente de apuntamiento o curtosis:

$$a_4 = \left(\frac{1}{\sigma^4} \cdot \frac{\sum_{i=1}^n (x_i - \bar{x})^4 \cdot f_i}{N} \right) - 3 =$$

$$= \frac{1}{1,32^4} \cdot \frac{(1 - 2,95)^4 \cdot 2 + (2 - 2,95)^4 \cdot 7 + (3 - 2,95)^4 \cdot 5 + (4 - 2,95)^4 \cdot 3 + (5 - 2,95)^4 \cdot 2 + (6 - 2,95)^4 \cdot 1}{20} - 3 =$$

$$= \frac{1}{3,04} \cdot 8,01 - 3 = 2,63 - 3 = -0,37 < 0 \quad \Rightarrow \quad \text{Distribución platicúrtica}$$

Ejercicio 17: Se intenta hacer un estudio sobre el comportamiento del niño cuando se separa de la madre. Se estudian 20 niños de 13 meses de edad, y se obtiene la distribución de frecuencias de los metros que el niño se aleja de la madre:

Nº de metros	1	2	3	4	5
Nº de niños	3	10	4	2	1

Estudiar la simetría (coeficientes de Pearson y de Fisher) y el apuntamiento de la distribución.

Solución: $A_p = 0,39 \Rightarrow$ asimétrica a la derecha; $a_3 = 0,86 \Rightarrow$ asimétrica a la derecha; $a_4 = 0,41 \Rightarrow$ Leptocúrtica.

EJERCICIOS

1. Averigua el dato que falta en la siguiente distribución para que la media sea 18.

7 12 15 22 23 28 32

Solución: 5

2. A un conjunto de cinco números cuya media aritmética es 7,31 se le añaden 2,47 y 10,15. ¿Cuál es la media del nuevo conjunto de números?

Solución: $\bar{x} = 7,02$

3. Un instituto tiene tres grupos de bachillerato. La nota media de los alumnos del grupo A es de 5,7 puntos. La de los alumnos del grupo B es de 5,6, siendo 5,5 para los del grupo C. En el grupo A hay 30 alumnos y se sabe que en el grupo C hay 5 alumnos más que en el grupo B. Si la nota media de todos los alumnos de bachillerato es de 5,6 puntos, ¿cuántos alumnos de bachillerato hay en el instituto?

Solución: $N_A = 30$; $N_B = 25$; $N_C = 30$; Total = $30 + 25 + 30 = 85$

4. Pon un ejemplo de una distribución donde la media, la moda y la mediana coincidan.

Solución: Por ejemplo, la siguiente distribución: 1, 2, 2, 3, 3, 3, 4, 4, 5. En este caso, media = moda = mediana = 3. Puede considerarse también cualquier otro ejemplo que conserve una simetría.

5. ¿Qué habría de ocurrir para que la media aritmética de una variable estadística fuese cero? Pon un ejemplo de una distribución que verifique la propiedad anterior.

Solución: Como $\bar{x} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_n \cdot f_n}{f_1 + f_2 + \dots + f_n}$, para que $\bar{x} = 0$ se ha de verificar que $x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_n \cdot f_n = 0$. Las

frecuencias son siempre mayores o iguales que cero, por lo que debe ocurrir que la variable tome valores positivos y negativos. Ejemplo: Las temperaturas mínimas recogidas en una ciudad a lo largo de una semana del mes de enero han sido: 0, -3, 2, 0, 1, -1, 1. Por tanto, $\bar{x} = 0$, es decir, la temperatura media a lo largo de esa semana fue de 0.

6. ¿Puede ser que la media no coincida con ningún valor de la variable? ¿Y la moda? Razona tus respuestas.

Solución: La media puede no coincidir con ningún valor de la variable; sin embargo, la moda siempre estará asociada a un valor concreto de la distribución.

7. La estación meteorológica de Puebloseco registró 88 días de lluvia el año pasado, según se muestra en la siguiente tabla:

Litros/m ²	[0, 5)	[5, 10)	[10, 15)	[15, 20)	[20, 25)	[25, 30)	[30, 35)
Nº de días	3	7	19	23	18	12	6

Calcula la precipitación media, la moda y la mediana.

Solución: $\bar{x} = 18,523$ l/m²; $Mo = 17,2$ l/m²; $Me = 18,26$ l/m².

8. La tabla adjunta muestra el número de faltas de asistencia en una clase a lo largo de un mes.

Nº de faltas	0	1	2	3	4	5
Nº de alumnos	10	7	6	2	1	4

Calcula la media aritmética, la moda y los cuartiles de la distribución.

Solución: $\bar{x} = 1,63$ faltas; $Mo = 0$ faltas; $Q_1 = 0$ faltas; $Q_2 = 1$ falta; $Q_3 = 2$ faltas.

9. La tabla adjunta muestra los resultados de unos alumnos en la prueba de salto con pértiga.

Media del salto (m)	[2; 2,5)	[2,5; 3)	[3; 3,5)	[3,5; 4)
Nº de alumnos	5	12	20	11

Calcula la media aritmética, la moda y los cuartiles de la distribución.

Solución: $\bar{x} = 2,98$ m; $Mo = 3,25$ m; $Q_1 = 2,79$ m; $Q_2 = 3,18$ m; $Q_3 = 3,48$ m.

10. Para los siguientes conjuntos de datos, averigua si existe un valor atípico.

a) 30, 31, 31, 33, 34, 36, 37

b) 11, 78, 79, 81, 82, 83, 160

Solución:

a) $Q_1 = 31$ y $Q_3 = 36$; Ningún valor está alejado, ya que para los valores extremos se tiene:

$$31 - 1,5(36 - 31) = 23,5 \text{ y } 30 > 23,5 \quad ; \quad 36 + 1,5(36 - 31) = 43,5 \text{ y } 36 < 43,5$$

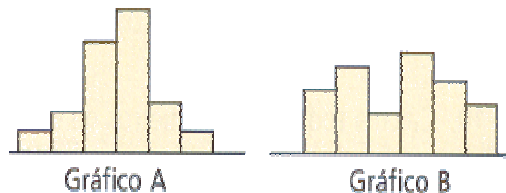
b) $Q_1 = 78$ y $Q_3 = 83$; Los valores extremos son valores alejados.

$$78 - 1,5(83 - 78) = 70,5 \text{ y } 11 < 70,5 \quad ; \quad 83 + 1,5(83 - 78) = 90,5 \text{ y } 160 > 90,5$$

11. Se nos informa que los datos correspondientes a los gráficos A y B son, aproximadamente:

$$\bar{x}_1 = 5,4 \quad ; \quad \sigma_1 = 3,3 \quad ; \quad \bar{x}_2 = 5,6 \quad ; \quad \sigma_2 = 2,5$$

Averiguar el gráfico correspondiente a cada par, explicando el razonamiento seguido.



Solución: Gráfico A \leftrightarrow Variable 2; Gráfico B \leftrightarrow Variable 1

12. Calcula la media aritmética, la varianza y la desviación típica de la distribución dada por la siguiente tabla.

x_i	3	4	5	6	7	8	9
f_i	2	5	11	15	10	6	2

Solución: $\bar{x} = 6,019$; $\sigma^2 = 1,96$; $\sigma = 1,4$

13. La siguiente tabla presenta el número de horas semanales que dedican al estudio los 30 alumnos de una clase de 1º de Bachillerato.

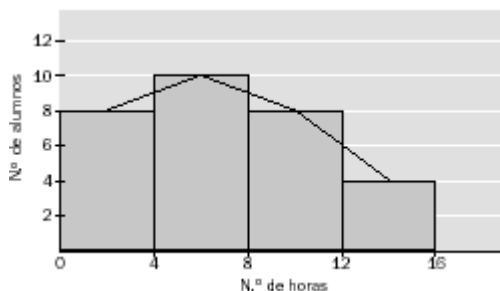
Nº de horas	[0, 4)	[4, 8)	[8, 12)	[12, 16)
Nº de alumnos	8	10	8	4

- Halla la media, la moda, la mediana y los otros dos cuartiles.
- Calcula el rango, la varianza y la desviación típica.
- Representa el histograma y el polígono de frecuencias.

Solución: a) $\bar{x} = 7,07$ horas; $Mo = 6$ horas; $Me = 6,8$ horas; $Q_1 = 3,75$ horas; $Q_3 = 10,25$ horas.

b) $R = 14 - 2 = 12$; $\sigma^2 = 15,88$; $\sigma = 3,99$

c)



14. Dados los datos de la siguiente tabla:

Clases	[10, 20)	[20, 30)	[30, 40)	[40, 50)	[50, 60)
f_i	5	12	20	11	6

Calcula la media aritmética, la varianza y la desviación típica de la distribución asociada.

Solución: $\bar{x} = 35,185$; $\sigma^2 = 124,05$; $\sigma = 11,137$

15. Si se suma a todos los valores de la variable una constante, ¿cómo quedan afectadas la media y la varianza? ¿Y si se multiplican por una constante?

Solución: Si se suma a todos los valores de la variable una constante, la media queda aumentada en esa constante, mientras que la varianza no varía. Si se multiplican todos los valores de la variable por una constante, la media queda multiplicada por esa constante, mientras que la varianza queda multiplicada por el cuadrado de dicha constante.

16. Tenemos dos variables X e Y con el mismo recorrido y media, siendo sus varianzas 4 y 9 respectivamente. ¿Para cual de las dos variables el valor de la media es más representativo?

Solución: Para X pues tiene menor dispersión.

17. Sea una variable con media 8 y desviación típica 0. ¿Qué se puede afirmar sobre el comportamiento de esta variable?

Solución: La variable sólo toma un único valor constante e igual a 8.

18. En ocasiones, la media se ajusta más que la moda a la distribución, y a veces lo contrario. En cada una de las siguientes tablas, ¿qué parámetro (\bar{x} o Mo) es más significativo y por qué?

Variable A	
x_i	f_i
138	2
254	1
351	1
2	30

Variable B	
y_i	f_i
3	8
7	6
12	7
21	5

Solución: $\bar{x} = 27,68$; $\sigma_x = 76,67$; $Mo = 2$; En esta variable, la varianza es muy grande, lo cual demuestra que la media no es muy representativa. Si nos fijamos en la tabla, se observa que de 34 datos hay 30 que son 2. Con lo cual la moda es mucho más representativa de la distribución. $\bar{y} = 9,81$; $\sigma_y = 6,44$; $Mo = 3$; Observando la varianza y la tabla, en esta variable, la media es bastante representativa, más que la moda.

19. En la tabla se recogen los datos correspondientes a una distribución estadística:

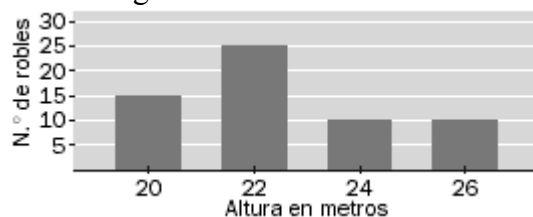
Clases	[1, 2)	[2, 3)	[3, 4)	[4, 5)	[5, 6)	[6, 7)	[7, 8)
f_i	6	12	15	20	16	11	6

Calcula la media y la desviación típica y halla el porcentaje de datos incluidos en los intervalos

$$(\bar{x} - \sigma, \bar{x} + \sigma) \quad ; \quad (\bar{x} - 3\sigma, \bar{x} + 3\sigma) \quad ; \quad (\bar{x} - 3\sigma, \bar{x} + 3\sigma)$$

Solución: $\bar{x} = 4,44$; $\sigma = 1,77$; 59 %, 93 % y 100 % respectivamente.

20. En el Parque Nacional de los Picos de Europa se ha realizado un estudio sobre la altura de sus robles, y para ello se ha tomado una muestra en una superficie de 15 kilómetros cuadrados, obteniéndose los siguientes resultados gráficos.



- Halla la media, la moda y la mediana.
- Calcula el intervalo que contenga el 95% de los robles.

Solución: a) $\bar{x} = 22,5$; $Mo = 22$; $Me = 22$; b) $(\bar{x} - 2\sigma, \bar{x} + 2\sigma) = (18,46; 26,54)$

21. En una población nórdica con 2500 habitantes adultos se ha realizado un estudio sobre su altura. La distribución de alturas es normal (unimodal y simétrica). Sabiendo que en el intervalo (172, 196) se encuentran 2375 habitantes y que la altura media es de 184 centímetros, calcula la desviación típica de la distribución.

Solución: $\bar{x} = 148$; $\sigma = 6$

22. Se ha realizado una encuesta entre los alumnos de un colegio de Enseñanza Primaria con el objeto de conocer el número de horas semanales que ven la televisión. El estudio arroja la siguiente información: el número de horas del 95% de los alumnos se encuentra en el intervalo (3,18; 17,1). Calcula la media aritmética y la desviación típica.

Solución: $\bar{x} = 10,14$; $\sigma = 3,48$

23. Tenemos una variable X de la que sabemos que: $CV = 0,5$ y que $\sigma_x = 3$. ¿Cuál es el valor de la media de X ?

Solución: $\bar{x} = 6$

24. Calcula el coeficiente de variación de la siguiente distribución.

x_i	1	2	3	4	5	6	7
f_i	5	12	18	11	7	4	1

Solución: $CV = \frac{1,41}{3,33} = 0,42 \Rightarrow 42\%$

25. El coeficiente de variación de la variable X sabemos que es 1 ¿Qué podemos decir sobre su media y su varianza?

Solución: Que son iguales.

26. Compara las dispersiones de las siguientes distribuciones.

x_i	6	3	4	3	7	5	6	8
y_i	63	39	64	55	66	70	65	62

Solución: $CV_x = \frac{1,71}{5,25} = 0,33 \Rightarrow 33\%$; $CV_y = \frac{9,04}{60,5} = 0,15 \Rightarrow 15\%$: $CV_x < CV_y \Rightarrow Y$ es menos dispersa.

27. La tabla recoge las temperaturas máximas alcanzadas en dos ciudades durante 10 días consecutivos del mes de agosto.

A	32	33	22	35	30	29	31	20	19	24
B	27	28	25	31	24	25	24	26	22	28

- a) ¿Qué ciudad ha tenido una temperatura media más alta a lo largo de esos 10 días?
- b) ¿Qué ciudad ha sufrido una variabilidad mayor de temperatura?
- c) ¿Qué parámetro has empleado para contestar el anterior apartado? ¿Por qué?

Solución: Como $\bar{x}_A = 27,5^\circ$ y $\bar{x}_B = 26^\circ$, la ciudad A ha tenido una temperatura media más alta.

- b) $CV_A = 0,198$ y $CV_B = 0,09$. Por tanto, la ciudad A ha tenido una variabilidad de temperatura mayor.
- c) El parámetro empleado es el CV , y se utiliza porque las medias y las desviaciones típicas de las dos distribuciones son distintas, y es la única manera de comparar sus dispersiones.

28. Las notas obtenidas en la asignatura de Matemáticas por los alumnos de dos clases de 1º de Bachillerato son las siguientes.

Notas	0	1	2	3	4	5	6	7	8	9	10
A	5	4	1	0	0	0	0	0	1	4	5
B	0	0	2	2	3	6	3	2	2	0	0

- a) ¿Cuál es la calificación media de cada una de las dos clases?
- b) ¿Cuál de ellas tiene las notas menos dispersas?
- c) ¿Es necesario calcular el coeficiente de variación para poder determinarlo? ¿Por qué?

Solución: Como $\bar{x}_A = 5$ y $\bar{x}_B = 5$.

- b) $\sigma_A = 4,45$ y $\sigma_B = 1,7$. Por tanto, el grupo B tiene las notas menos dispersas.
- c) No ha sido necesario calcular CV , ya que cuando las medias son iguales tiene menor dispersión la distribución que tenga menor desviación típica.

29. Dos trabajadores del mismo sector ganan 620 € y 672 €, respectivamente. El primero pertenece a la empresa A , cuya retribución media y desviación típica vienen dados por: $\bar{x}_A = 580$ € y $\sigma_A = 25$ €, mientras que para la empresa B del segundo trabajador se tiene: $\bar{x}_B = 640$ € y $\sigma_B = 33$ €. Tanto uno como el otro ganan salarios por encima de la media, y se quiere conocer cuál de los dos ocupa mejor posición relativa dentro de su empresa.

Solución: $z_A = 1,6$; $z_B = 0,97 \Rightarrow$ Por lo que, aunque en términos absolutos el trabajador de la empresa B gana más que el de A , en relación al conjunto de los empleados de cada empresa el empleado de A ocupa mejor posición.