# Apuntes de estadística descriptiva
# Unidad 1. Estadística descriptiva.

## Conceptos básicos

### Definición de estadística.

**<u>Definición de estadística.</u>**
El término estadística tiene su raíz en la palabra Estado. Surge cuando se hace necesario para sus intereses cuantificar conceptos. En la mayoría de los casos esta cuantificación se hará en función de unos fines económicos o militares. El estado quiere conocer censo de personas, de infraestructura, de recursos en general, para poder obtener conclusiones de esta información.

Definición de estadística descriptiva:

La **estadística descriptiva** es una ciencia que analiza series de datos (por ejemplo, edad de una población, altura de los estudiantes de una escuela, temperatura en los meses de verano, etc.) y trata de extraer conclusiones sobre el comportamiento de estas variables.

La estadística descriptiva implica la abstracción de varias propiedades de conjuntos de observaciones, mediante el empleo de métodos gráficos, tabulares ó numéricos. Entre estas propiedades, están la frecuencia con que se dan varios valores en la observación, la noción de un valor típico o usual, la cantidad de variabilidad en un conjunto de datos observados y la medida de relaciones entre 2 ó más variables.

La estadística descriptiva sirve como método para organizar datos y poner de manifiesto sus características esenciales con el propósito de llegar a conclusiones

### 1.1 Medidas de tendencia central

Al describir grupos de observaciones, con frecuencia se desea describir el grupo con un solo número. Para tal fin, desde luego, no se usará el valor más elevado ni el valor más pequeño como único representante, ya que solo representan los extremos. Más bien que valores típicos. Entonces sería más adecuado buscar un valor central.

Las medidas que describen un valor típico en un grupo de observaciones suelen llamarse medidas de tendencia central. Es importante tener en cuenta que estas medidas se aplican a grupos más bien que a individuos. Un promedio es una característica de grupo, no individual.

| | |
|---|---|
| **Media aritmética** | Suma de los valores de una serie de medidas respecto del número de valores existentes. Su cálculo equivale a □ $x_i$/n, siendo **n** el tamaño de la muestra y $\mathbf{x_i}$ cada uno de los valores. |
| **Mediana** | Valor que queda en el centro tras la división de una serie de valores ordenados en dos partes iguales, una superior y una inferior. Para determinarla debe seguirse los siguientes pasos: |

| | |
|---|---|
| | -ordenar los datos de menor a mayor<br>-si el número de datos es impar corresponde al que queda en el centro<br>-si el número de datos es par corresponde al valor medio de los dos datos centrales |
| **Moda** | Valor que se presenta con más frecuencia en una serie de mediciones. |

### 1.1.1 Media aritmética, geométrica y ponderada.

**Media aritmética**

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

La medida de tendencia central más obvia que se puede elegir, es el simple promedio de las observaciones del grupo, es decir el valor obtenido sumando las observaciones y dividiendo esta suma por el número de observaciones que hay en el grupo.

En realidad hay muchas clases de promedios y ésta se la llama media aritmética para denotar la suma de un grupo de observaciones dividida por su número.

**Media aritmética**

Llamando $x_l$, ..., $x_k$ a los datos distintos de un carácter en estudio, o las marcas de clase de los intervalos en los que se han agrupado dichos datos, y $n_i$,..., $n_k$ a las correspondientes frecuencias absolutas de dichos valores o marcas de clase, llamaremos *media aritmética* de la distribución de frecuencias a

$$a = \frac{\sum_{i=1}^{k} x_i \cdot n_i}{n}$$

en donde n es la frecuencia total.

**Ejemplo 1:**

La media aritmética de las veinticinco familias encuestadas será:

$$a = \frac{\sum_{i=1}^{5} x_i \cdot n_i}{n} = \frac{0 \cdot 5 + 1 \cdot 6 + 2 \cdot 8 + 3 \cdot 4 + 4 \cdot 2}{25} = \frac{42}{25} = 1'68$$

es decir, las familias encuestadas tienen un número medio de hijos de 1'68.

**Ejemplo 2:**

Se midieron los niveles de colinesterasa en un recuento de eritrocitos en de 34 agricultores expuestos a insecticidas agrícolas, obteniéndose los siguientes datos:

| Individuo | Nivel | Individuo | Nivel | Individuo | Nivel |
|---|---|---|---|---|---|
| 1 | 10,6 | 13 | 12,2 | 25 | 11,8 |
| 2 | 12,5 | 14 | 10,8 | 26 | 12,7 |
| 3 | 11,1 | 15 | 16,5 | 27 | 11,4 |
| 4 | 9,2 | 16 | 15,0 | 28 | 9,3 |
| 5 | 11,5 | 17 | 10,3 | 29 | 8,6 |
| 6 | 9,9 | 18 | 12,4 | 30 | 8,5 |
| 7 | 11,9 | 19 | 9,1 | 31 | 10,1 |
| 8 | 11,6 | 20 | 7,8 | 32 | 12,4 |
| 9 | 14,9 | 21 | 11,3 | 33 | 11,1 |
| 10 | 12,5 | 22 | 12,3 | 34 | 10,2 |
| 11 | 12,5 | 23 | 9,7 | | |
| 12 | 12,3 | 24 | 12,0 | | |

La distribución de frecuencias las marcas de clase será:

| Intervalo | $I_i$ | 7'5-9 | 9-10'5 | 10'5-12 | 12-13'5 | 13'5-15 | 15-16'5 | |
|---|---|---|---|---|---|---|---|---|
| Marca de Clase | $x_i$ | 8'25 | 9'75 | 11'25 | 12'75 | 14'25 | 15'75 | |
| Frecuencia | $n_i$ | 3 | 8 | 10 | 10 | 1 | 2 | ?$n_i$=25 |

la cual proporciona una media aritmética de

$$a = \frac{\sum_{i=1}^{6} x_i \cdot n_i}{n} = \frac{388'5}{34} = 11'426$$

## MEDIA ARITMETICA.

Es la medida de tendencia central más utilizada en estadística y es la que se conoce como el promedio de las observaciones, sin embargo, debido a la confusión que hay con el término promedio.

La media es el valor correspondiente a una línea imaginaria que compensa los valores que se exceden de la media y los que quedan por debajo de ésta; de esta manera, la media es mayor que el valor más pequeño, y menor que el valor más grande.

Cuando se dispone de datos no agrupados, la media se puede calcular con precisión al sumar todos los valores observados y dividir el total entre el número de observaciones. Si las utilidades anuales de cinco empresas (en millones de dólares) fueron 2, 2, 4, 7 y 15, la media aritmética sería igual a:

$$\frac{2 + 2 + 4 + 7 + 15}{5} = \frac{30}{5} = 6$$

Este número (6) sería la media poblacional si el sistema de interés contuviera sólo cinco empresas, por ejemplo, un sistema de interés son todos los fabricantes de aviones en los Estados Unidos o todos los fabricantes de cerveza en Detroit. Sería una media muestral si se refiere sólo a cinco empresas de entre un grupo de interés mucho mayor, como cinco entre docenas de fabricantes de aviones en el mundo o cinco entre cientos de cervecerías en los Estados Unidos. El procedimiento anterior se resume como:

**Para una población:**

$$\mu = \Sigma x / N$$

**Para una muestra:**

$$\bar{X} = \Sigma x / n$$

en donde $\Sigma x$ es la suma de todos los valores de la población (o muestra) observados, N es el número de observaciones en la población y n es el de observaciones en la muestra.

**Propiedades de la media aritmética**

1. La suma de las desviaciones o diferencias de cada valor respecto a la media es igual a cero.
2. La suma de los cuadrados de las desviaciones de cada valor respecto a la media es un valor mínimo.
3. La media puede utilizarse para determinar el valor total de la población. (Número de elementos) * (Media) = Total de la población
4. La media se afecta sustancialmente hacia arriba o hacia abajo con la presencia de valores extremos (muy grandes o muy pequeños) respecto a la media.

**EJEMPLO** Mediante el uso de la tabla 3.1, calcule la media aritmética de las utilidades ganadas por las 100 multinacionales más grandes con oficinas en los Estados Unidos.

Tabla 3.1

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1071 | 724 | 457 | 2060 | 722 | 499 | 81 | - 353 | 115 | 108 |
| 784 | 447 | 283 | 5423 | 2276 | 3308 | 579 | 708 | 133 | 165 |
| 197 | 600 | 405 | 258 | 312 | 430 | 1169 | 524 | 592 | 489 |
| 835 | 448 | 254 | 119 | 772 | 60 | 2433 | 535 | 145 | 1177 |
| 960 | 1159 | 473 | 119 | 803 | 1576 | 754 | 1369 | 749 | 791 |
| 441 | 412 | 863 | 370 | 512 | 97 | 348 | 2762 | 2380 | 458 |
| 1671 | 215 | 522 | - 1060 | 1231 | 838 | 339 | 668 | 1992 | 580 |
| 403 | - 476 | 536 | 1652 | 1181 | 220 | 627 | 409 | 1281 | 729 |
| 445 | 531 | 916 | 532 | 338 | 519 | 437 | 98 | 1851 | 257 |
| 188 | 918 | 1467 | 333 | 580 | 252 | 2227 | 1567 | 2310 | 598 |

**SOLUCION** $\mu = \Sigma x / N$ = (78 662 / 100) = 782.62 millones de dólares

La solución se puede encontrar por cálculo manual o, mucho más rápidamente, por computadora después de que los datos de la tabla 3.2 se le hayan introducido.

**Media geométrica**

La media geométrica de un conjunto de observaciones es la raíz n ésima de su producto. El cálculo de la media geométrica exige que todas las observaciones sean positivas.

$$G = \sqrt[n]{x_1 . x_2, \ldots, x_n}$$

**MEDIA GEOMETRICA.**

Esta es una medida que puede aplicarse al crecimiento exponencial o interés compuesto, pues obtiene la raíz enésima de un grupo de n datos multiplicados entre sí, por ejemplo, la raíz cúbica del producto de 3 datos, o la raíz octava del producto de 8 datos. El resultado obtenido, al elevarse a la potencia enésima, produce el producto de todos los datos multiplicados entre sí.

**Para una población:**

$$G = \sqrt[n]{\Sigma X_i}$$

**Para una muestra:**

$$G = \sqrt[N]{\Sigma X_i}$$

**Características de la media geométrica:**

1. El cálculo de la media geométrica está basado en todos los elementos de un conjunto de datos. El valor de cada elemento de dicho conjunto afecta así el valor de la media geométrica.

2. Si uno de los valores es cero, el valor de G es cero.
3. Si uno de los valores es negativo y el número de datos es par, el valor de G es imaginario y no tiene interpretación.

   Si uno de los valores es negativo y el número de datos es impar, aunque G existe, su valor no es representativo.

4. La media geométrica es afectada por valores extremos en una menor cantidad que lo es la media aritmética. Por ejemplo, la media geométrica de los valores 1, 4 y 16 es 4, mientras que la media aritmética de los mismos valores es 7. El valor 7 es más cercano al valor alto 16 que el valor 4 lo es de 16. El valor de G es siempre menor que el valor de la media de los mismos datos, excepto cuando todos los valores en una serie son iguales, tales como la media geométrica y la media aritmética para los valores 4, 4 y 4 que son ambas 4.
5. La media geométrica da igual ponderación a las tasas de cambio iguales. En otras palabras, al promediar tasas de cambio geométricamente, la tasa que muestra el doble de su base es compensada por la otra que muestra la mitad de su base; la tasa que muestra un quinto de su base; y así sucesivamente. Las tasas de cambio son ordinariamente expresadas en porcentajes. Puesto que la base de cada proporción expresada en porciento es siempre igual a 100%, el promedio de dos proporciones las cuales se compensan deberá ser 100% también.
6. La media geométrica de las proporciones de los valores individuales con respecto a cada valor precedente en una secuencia de valores es la única medida de tendencia central apropiada para las proporciones. La media aritmética de las proporciones no dará un resultado consistente.

**EJEMPLO** Las ventas mensuales de una tienda por departamentos y las proporciones de las ventas mensuales a las ventas en cada mes previo de Enero a Mayo, están dadas en la tabla siguiente:

Tabla 3.3

| *Mes* | *Ventas mens (en millones* | *Tasa con respecto al mes previo* |
|---|---|---|
| *Enero* | 5,000 | |
| *Febrero* | 3,600 | 0.72 = (3600/5000) |
| *Marzo* | 5,760 | 1.60 = (5760/5000) |
| *Abril* | 5,184 | 0.90 = (5184/5000) |
| *Mayo* | 10,368 | 2.00 = (10368/5000) |
| *Total* | $29,912 | 5.22 |

Calcule la media geométrica así como la media aritmética de las tasas y compárelas.

**SOLUCION** La media geométrica de las tasas es 1.20 ó 120% y la media aritmética es 1.305 ó 130.5%.

Comparación de las ventas calculadas mediante la media aritmética y la media geométrica:

Tabla 3.4

| *Mes* | *Ventas reales* | *Ventas basadas en G* | *Ventas basadas en X* |
|---|---|---|---|
| *Enero* | 5,000 | | |
| *Febrero* | 3,600 | 6,000 ( = 5000X120% ) | 6,525 ( = 5,000X130.5% ) |
| *Marzo* | 5,760 | 7,200 ( = 6,000X120% ) | 8,515 ( = 6,525X130.5% ) |
| *Abril* | 5,184 | 8,640 ( = 7200X120% ) | 11,112 ( = 8,515X130.5% ) |
| *Mayo* | 10,368 | 10,368 ( = 8,640X120% ) | 14,501 (=11,112X130.5% ) |
| *Total* | $29,912 | | |

**Media armónica**
Es el inverso de la media aritmética de los inversos de las observaciones.

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

**MEDIA ARMONICA.**

La media armónica (H) de n observaciones X1, X2, ... , Xn es el inverso (multiplicativo) de la media aritmética de los inversos de las observaciones.

**Para la población:**

$$H = n \,/ \sum_{i=1}^{n} (1/x_i)$$

**Para la muestra:**

$$H = N \,/ \sum_{i=1}^{N} (1/x_i)$$

Características de la media armónica:

1. La media armónica como la media aritmética y la geométrica, se calcula usando todos los elementos en un conjunto de valores. El valor de cada elemento en todos los datos afecta, por lo tanto, el valor de la media armónica. Sin embargo, la media armónica es aún menos afectada por valores extremos que la media geométrica. La magnitud relativa de las tres diferentes medias para los mismos datos puede ser expresada como sigue:

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

2. La media armónica no es tan frecuentemente usada como una medida de tendencia central de un conjunto de datos como es la media aritmética. Sin embargo, es útil en caos especial para promediar velocidades. La razón de cambio usualmente indica la relación entre dos tipos diferentes de unidades de medida que pueden ser expresadas recíprocamente. Por ejemplo si una persona caminó 10 millas en 2 horas, la razón de su velocidad de caminar puede ser expresada:
3. 10 millas
4. ------------ = 5 millas/hora
5. 2 horas
6. ó reciprocamente,
7. 2 horas
8. ----------- = 1/5 horas/milla
9. 10 millas
10. La media armónica deberá usarse cuando un valor constante, el cual tiene la misma unidad que el numerador (millas) de cada razón dada, es igualmente aplicable a cada elemento en los datos.

**EJEMPLO** Si un automóvil recorre las primeras 10 millas a 30 mph y las segundas a 60 mph, a primera vista pareciera que la velocidad promedio de 30 y 60 es de 45 mph. Pero este tipo de media se suele definir en Física como la distancia total recorrida divida entre el tiempo total empleado en recorrerla, y como la distancia total es de 20 millas y el tiempo total es 1/3 + 1/6 de hora, se tiene que la velocidad media es:

vel = 20 / ( 1/3 + 1/6 ) = 40 mph

Es interesante observar que esta media se puede calcular como una media armónica de 30 y 60, es to es:

H = 2 / ( 1 / 30 + 1 / 60 ) = 40 mph.

**Media ponderada**

En ciertas circunstancias no todas las observaciones tienen igual peso. En general si se tienen observaciones con sus respectivos pesos es:

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

**MEDIA PONDERADA.**

La media o promedio simple es la medida de tendencia central más utilizada; sin embargo, cuando algunos de los valores por promediar son más importantes que otros, por ejemplo, al evaluar a un empleado, su calificación en conocimientos, puntualidad, presentación y otros conceptos tiene una importancia relativa diferente en función a quién, hace la evolución.

Tal vez no sea lo mismo un empleado con 10 en conocimientos, 10 en puntualidad y 7 en presentación (promedio = 9), que otro con 10 en conocimientos, 7 en puntualidad y 10 en presentación (promedio = 9).

Cuando los valores por promediar tienen diferentes grados de importancia entre sí, debe utilizarse el promedio ponderado, el cual aplica un factor de ponderación (o importancia relativa) a cada uno de los valores que se van a promediar.

**Para una población:**

$$\mu_w = \Sigma w X / \Sigma w$$

**Para una muestra:**

$$\bar{X}_w = \Sigma w X / \Sigma w$$

donde $\Sigma w X$ es la suma de todos los pesos (w) multiplicada por los valores observados (X), en tanto que $\Sigma w$ es igual a N (el número de observaciones de la población) o n (e número de observaciones de la muestra).

**EJEMPLO** En una empresa dada, el sueldo por hora es de 5 dólares para 100 trabajadores, de 10 dólares para 50 trabajadores y de 15 dólares para diez trabajadores. ¿Cuál es el sueldo promedio?

**SOLUCION**

$\mu_w$ = ((100*5) + (50*10) + (10*15)) / (100 + 50 + 10) = 7.19

El resultado dista mucho del sueldo por hora promedio no ponderado de 10 dólares.

### 1.1.2 Mediana.

Otra medida de tendencia central que se utiliza con mucha frecuencia es la mediana, que es el valor situado en medio en un conjunto de observaciones ordenadas por magnitud.

**La mediana**

La mediana (Md) es una medida de posición que divide a la serie de valores en dos partes iguales, un cincuenta por ciento que es mayor o igual a esta y otro cincuenta por ciento que es menor o igual que ella. Es por lo tanto, un parámetro que está en el medio del ordenamiento o arreglo de los datos organizados, entonces, la mediana divide la distribución en una forma tal que a cada lado de la misma queda un número igual de datos.

Para encontrar la mediana en una serie de datos no agrupados, lo primero que se hace es ordenar los datos en una forma creciente o decreciente y luego se ubica la posición que esta ocupa en esa serie de datos; para ello hay que determinar si la serie de datos es par o impar, luego el número que se obtiene indica el lugar o posición que ocupa la mediana en la serie de valores, luego la mediana será el número que ocupe el lugar de lo posición encontrada.

**Mediana**

La mediana es otra medida de posición, la cual se define como aquel valor de la variable tal que, supuestos ordenados los valores de ésta en orden creciente, la mitad son menores o iguales y la otra mitad mayores o iguales

Así, si en la siguiente distribución de frecuencias,

| $x_i$ | $n_i$ | $N_i$ |
|---|---|---|
| 0 | 3 | 3 |
| 1 | 2 | 5 |
| 2 | 2 | 7 |
| | 7 | |

Ordenamos los valores en orden creciente,

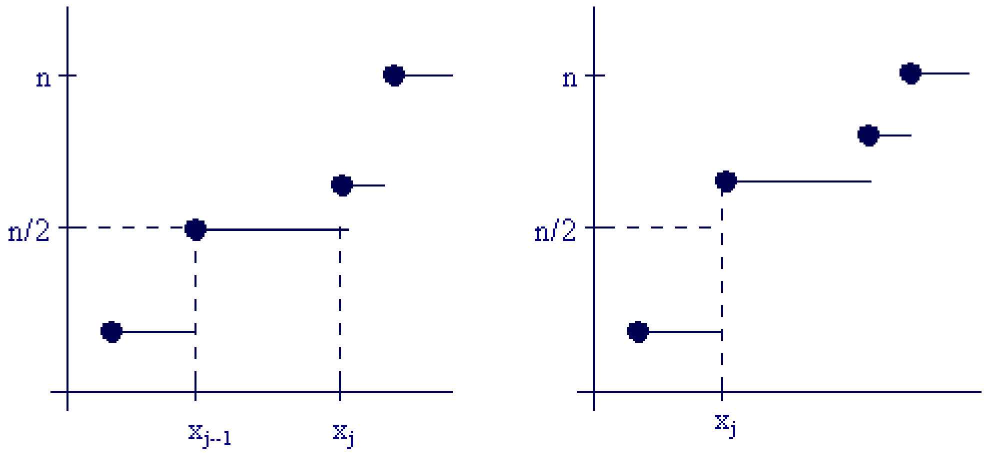
0 0 0 1 1 2 2

El 1 será el valor que cumple la definición de mediana.

Lógicamente, en cuanto el valor de la frecuencia total sea ligeramente mayor, este procedimiento resulta inviable. Por esta razón, daremos a continuación una fórmula que permita calcularla. No obstante, será necesario distinguir los casos en los que los datos vengan agrupados de aquellos en los que vengan sin agrupar.

- **Datos sin agrupar:**

Las gráficas siguientes, correspondientes a un diagrama de frecuencias absolutas acumuladas, recogen las dos situaciones que se pueden presentar:




Si la situación es como la de la figura de la derecha, es decir, si

Si la situación que se presenta es como la de la figura de la izquierda, entonces la mediana queda indeterminada, aunque en este caso se toma como mediana la media aritmética de los dos valores entre los que se produce la indeterminación; así pues, si

$N_{j-1} = n/2 < N_j$

Entonces la mediana es

$$M_e = \frac{x_{j-1} + x_j}{2}$$

**Ejemplo 1:**

La distribución de frecuencias acumuladas del ejemplo del *número de hijos* era

| **Nº de hijos($x_i$)** | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| **Frecuencias Acumuladas($N_i$)** | 5 | 11 | 19 | 23 | 25 |

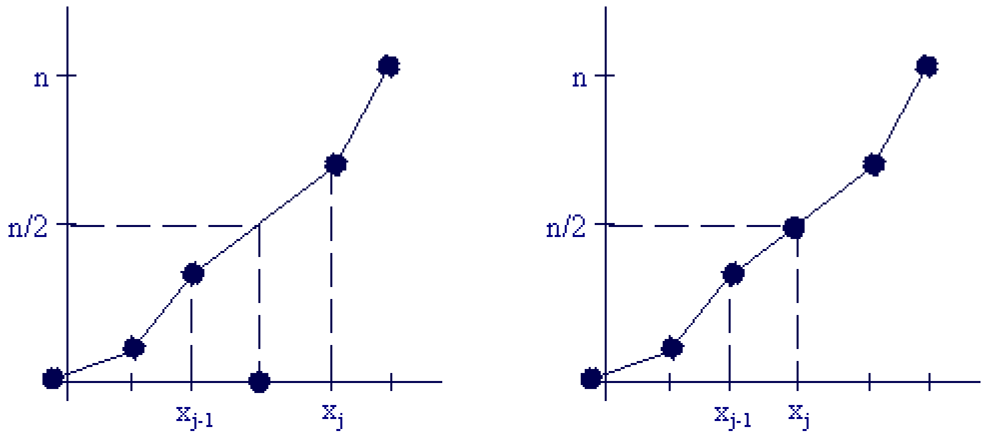
y como es n/2=12'5 y en consecuencia

11 < 12'5 < 19

la mediana será $M_e$= 2.

- **Datos Agrupados**

Las gráficas siguientes, correspondientes a *polígonos de frecuencias absolutas acumuladas*, nos plantea de nuevo dos situaciones diferentes a considerar:




El más sencillo, el de la derecha, en el que existe una frecuencia absoluta acumulada $N_j$ tal que $n/2 = N_j$, la mediana es $M_e = x_j$.

Si la situación es como la que se representa en la figura de la izquierda, en la que

$$N_{j-l} < n/2 < N_j$$

Entonces, la mediana, está en el intervalo $[x_{j-1}, x_j)$, es decir entre $x_{j-1}$ y $x_j$, tomándose en ese caso, por razonamientos de proporcionalidad, como mediana el valor

$$M_e = x_{j-1} + \frac{\frac{n}{2} - N_{j-1}}{n_j} \cdot c_j$$

Siendo $c_j$ la amplitud del intervalo $[x_{j-1}, x_j)$.

**Ejemplo:**

La distribución de frecuencias del ejemplo de los *niveles de colinesterasa* es:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Intervalo | $I_i$ | 7'5-9 | 9-10'5 | 10'5-12 | 12-13'5 | 13'5-15 | 15-16'5 |
| Frecuencia | $n_i$ | 3 | 8 | 10 | 10 | 1 | 2 |
| Frecuencia Acumulada | $N_i$ | 3 | 11 | 21 | 31 | 32 | 34 |

Al ser $n/2 = 17$ y estar

$$11 < 17 < 21$$

la mediana estará en el intervalo [10'5 , 12), y aplicando la fórmula anterior, será

$$M_e = 10'5 + \frac{\frac{34}{2} - 11}{10} \cdot 1'5 = 11'4$$

**MEDIANA.**

Es valor del elemento de la posición central de los datos individuales, ordenados de mayor a menor (o viceversa), y es el punto que marca la mitad de valores mayores que él, es decir, está a la mitad, con el 50% de valores a su derecha y el 50% de valores a su izquierda.

Es la medida de tendencia central más utilizada en estadística y es la que se conoce como el promedio de las observaciones, sin embargo, debido a la confusión que hay con el término promedio.

Para calcular la mediana:

- Ordene los datos, de mayor a menor o viceversa.
- Calcule la posición de la mediana:
- n + 1 N + 1
- Pisición de la mediana = ------- = -------
- 2 2
- Determine el elemento de la posición central, que es finalmente la mediana. (Si el número de datos es par, deberá obtener el promedio del valor de los dos elementos centrales.)

**Características básicas de la mediana**

1. El valor de la mediana se afecta por el número de datos, no por la magnitud de ningún valor extremo.
2. Es igualmente probable que cualquier observación escogida al azar sea mayor o menor que la mediana.
3. Se puede determinar, incluso en distribuciones con intervalos abiertos.
4. La suma de los cuadrados de las desviaciones respecto a la mediana es un valor mínimo.

En los casos en que los datos contengan valores extremos, y considerando la cuarta propiedad de la media, es mejor utilizar la mediana en lugar de la media como medida de tendencia central.

**EJEMPLO** Use los datos de la tabla 3.2 para calcular la utilidad mediana obtenida por las cien multinacionales más grandes de los Estados Unidos.

Tabla 3.2

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| -1060 | 119 | 257 | 405 | 458 | 535 | 708 | 835 | 1181 | 1992 |
| -476 | 133 | 258 | 409 | 473 | 536 | 722 | 838 | 1231 | 2060 |
| -353 | 145 | 283 | 412 | 489 | 579 | 724 | 863 | 1281 | 2227 |
| 60 | 165 | 312 | 430 | 499 | 580 | 729 | 916 | 1369 | 2276 |
| 81 | 188 | 333 | 437 | 512 | 580 | 749 | 918 | 1467 | 2310 |
| 97 | 197 | 338 | 441 | 519 | 592 | 754 | 960 | 1567 | 2380 |
| 98 | 215 | 339 | 445 | 522 | 598 | 772 | 1071 | 1576 | 2433 |
| 108 | 220 | 348 | 447 | 524 | 600 | 784 | 1159 | 1652 | 2762 |
| 115 | 252 | 370 | 448 | 531 | 627 | 791 | 1169 | 1671 | 3308 |
| 119 | 254 | 403 | 457 | 532 | 668 | 803 | 177 | 1851 | 5423 |

**SOLUCION** El arreglo ordenado tiene un par de observaciones. En consecuencia, hay dos valores centrales de 532 y 535 millones de dólares. La mediana es la media aritmética, es decir,

(532 millones + 535 millones) / 2 = 533.3 millones de dólares que es un número muy distinto de la media aritmética de todas las cifras

### 1.1.3 Moda.

Otra medida de tendencia central es la moda. La moda es el valor que ocurre con mas frecuencia en un conjunto de observaciones.

**La moda**

La moda es la medida de posición que indica la magnitud del valor que se presenta con más frecuencia en una serie de datos; es pues, el valor de la variable que más se repite en un conjunto de datos. De las medias de posición la moda es la que se determina con mayor facilidad, ya que se puede obtener por una simple observación de los datos en estudio, puesto que la moda es el dato que se observa con mayor frecuencia. La moda se designa con las letras Mo.

**Moda**
La *moda* se define como aquel valor de la variable al que corresponde máxima frecuencia (absoluta o relativa). Para calcularla, también será necesario distinguir si los datos están o no agrupados.

- Datos sin agrupar:

Para datos sin agrupar, la determinación del valor o valores (ya que puede haber más de uno) modales es muy sencilla. Basta observar a que valor le corresponde una mayor $n_i$. Ese será la moda.

Así en el ejemplo del número de hijos, la simple inspección de la tabla siguiente proporciona como valor para la moda el $M_d$ = 2.

| | | | | | | |
|---|---|---|---|---|---|---|
| **Nº de hijos($x_i$)** | 0 | 1 | 2 | 3 | 4 | |
| **Nº de familias($n_i$)** | 5 | 6 | **8** | 4 | 2 | ?$n_i$=25 |

- Datos agrupados:

Si los datos se presentan agrupados en intervalos es necesario, a su vez, distinguir si éstos tienen o no igual amplitud.

Si tienen amplitud constante *c*, una vez identificado el intervalo modal $[x_{j-1}, x_j)$, es decir el intervalo al que corresponde mayor frecuencia absoluta $n_j = \max\{n_l, ..., n_k\}$, la moda se define, también por razones geométricas, como

$$M_d = x_{j-1} + \frac{n_{j+1}}{n_{j-1} + n_{j+1}} \cdot c$$

**Ejemplo:**

Este ejemplo presenta un caso de distribución bimodal, ya que tanto el intervalo [10'5 - 12) como el [12 - 13'5) tienen frecuencia absoluta máxima. Deberíamos aplicar, por tanto, para cada uno de los dos intervalos la fórmula anterior, determinando así las dos modas de la distribución. No obstante, este ejemplo presenta además la peculiaridad adicional de ser ambos intervalos modales contiguos. En esta situación se considera la distribución unimodal, eligiendo como moda el extremo común, $M_d = 12$.

Si los intervalos tuvieran distinta amplitud $c_j$, primeros debemos normalizar las frecuencias absolutas $n_j$, determinando los cocientes

$$l_j = \frac{n_j}{c_j}, \quad j = 1, \ldots, k$$

y luego aplicar la regla definida para el caso de intervalos de amplitud constante a los $l_j$. Es decir, primero calcular el $l_j = \max\{l_1, ...., l_k\}$ para determinar el intervalo modal $[x_{j-1}, x_j)$ y luego aplicar la fórmula

$$M_d = x_{j-1} + \frac{l_{j+1}}{l_{j-1} + l_{j+1}} \cdot c_j$$

Siendo $c_j$ la amplitud del intervalo modal $[x_{j-1}, x_j)$.

**Ejemplo:**

Las frecuencias *normalizadas* correspondientes al ejemplo de intervalos con distinta amplitud serán,

| $I_i$ | $n_i$ | $l_i$ |
|---|---|---|
| 0-20 | 8 | 0'4 |
| 20-30 | 9 | 0'9 |
| 30-40 | 12 | 1'2 |
| 40-45 | 10 | 2 |
| 45-50 | 9 | 1'8 |
| 50-60 | 10 | 1 |
| 60-80 | 8 | 0'4 |
| 80-100 | 4 | 0'2 |

con lo que el intervalo modal es el [40 - 45) y la moda

$$M_d = 40 + \frac{1'8}{1'2 + 1'8} \cdot 5 = 43$$

A diferencia de lo que ocurre con la media o con la mediana, sí es posible determinar la moda en el caso de datos cualitativos. Así, en el ejemplo del *tratamiento de radiación seguido de cirugía* puede afirmarse que la causa modal por la que no fue completado el tratamiento es $M_d$ = rehusaron cirugía.

**MODA.**

La moda es el valor más frecuente de un conjunto de datos en ocasiones se presentan dos o más valores que se repiten con mayor frecuencia. En este caso, a los datos se les conoce como bimodales o multimodales, respectivamente.

La moda es la única medida de tendencia central que se puede aplicar a datos del tipo cualitativo, por ejemplo: analizar el color de ojos (café, negro, azul) de una población. Es muy fácil de determinar, basta con observar detenidamente al conjunto de datos y ver cuál es el que más se repite; sin embargo, no es muy útil porque puede ocurrir que una distribución tenga dos o más valores que se repitan con la misma frecuencia, en tal caso se tienen dos o más modas. También puede ocurrir que no exista ningún valor que se repita y entonces no habrá moda. Por otra parte puede ser un valor extremo el de mayor frecuencia y difícilmente podría ser considerado una medida de tendencia central.

En la práctica, la moda raras veces se usa para describir datos no agrupados, es mucho más frecuente su designación para datos agrupados.