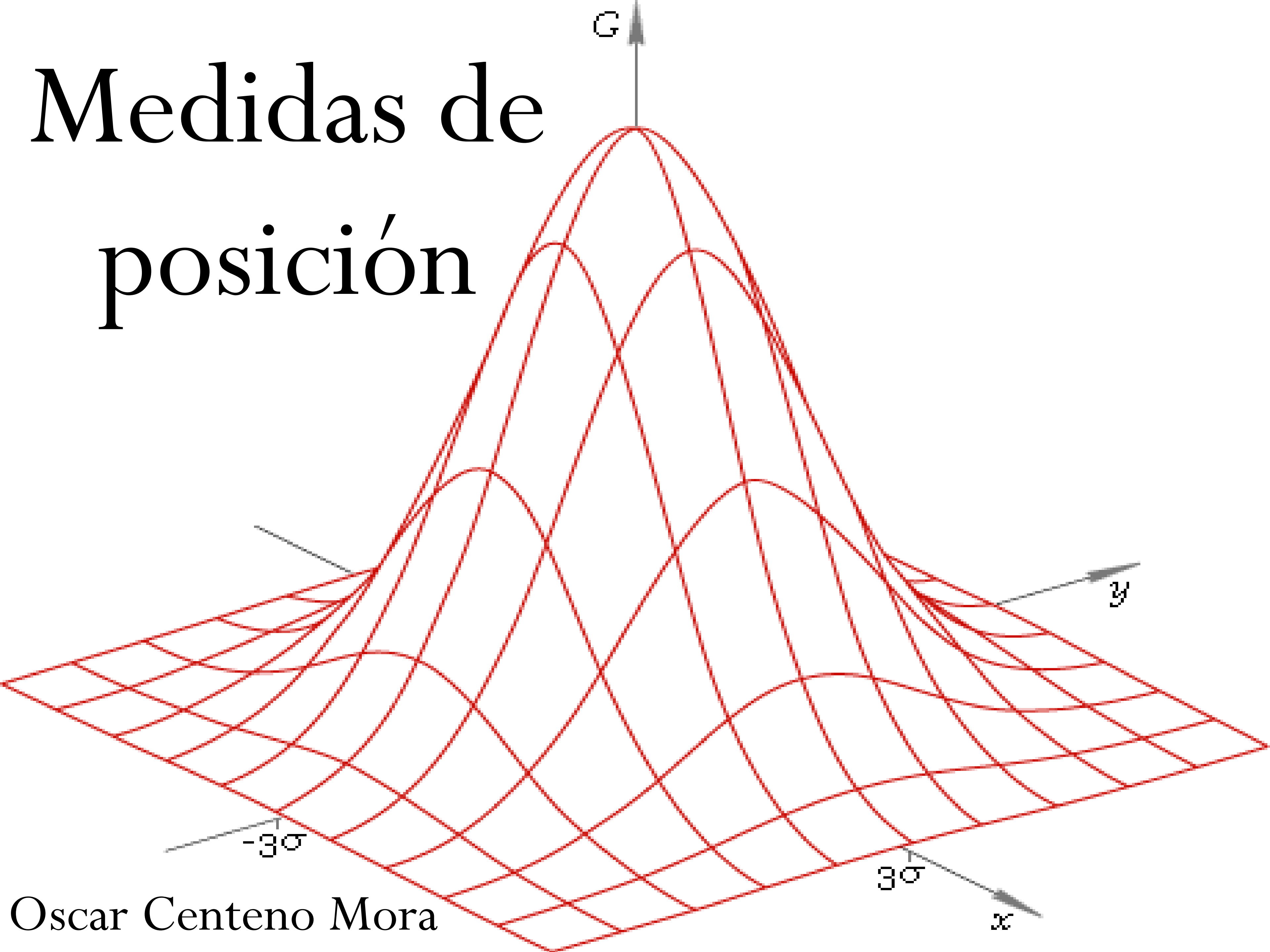
Medidas de posición



Oscar Centeno Mora

Preámbulo…

- Se le pide analizar la siguiente información referente a la información de clientes para una determinada catera del banco donde ud labora.

- Le piden que determine: días de mayor gasto, ingreso medio, más bajo y alto, el monto promedio de gastos para una determinado servicio, el 85% de los clientes que gastan más del 90% de su salario, etc.

- Y ud se pregunta : ¿cómo podría averiguar TODO esto?

Preámbulo…

¿Qué piensa hacer? ¿Cómo va a proceder?

Preámbulo…

- El presente tema se encuentra en las técnicas de análisis de la Estadística descriptiva.

- Anteriormente se estudiaron los números absolutos, los números relativos, y las distribución de frecuencias como métodos para analizar la información.
- Seguimos bajo la temática del análisis de la información, pero ahora nos enfocamos en lo primordial de los análisis: la posición y localización.

- Los temas de medidas de posición y variabilidad se deben de analizar de forma conjunta.

Índice

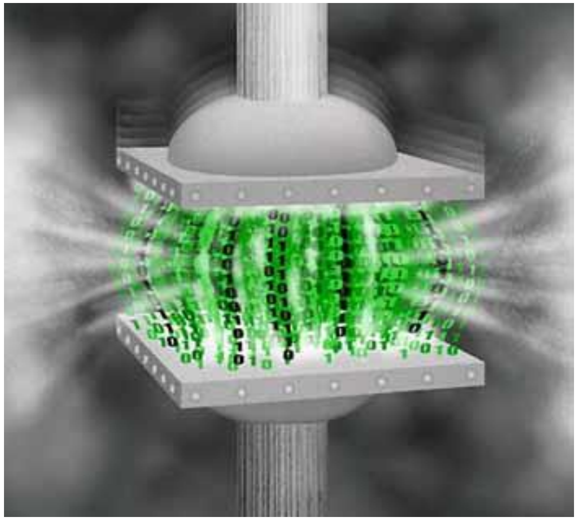
Índice

Introducción

- En los análisis estadísticos es muy importante conocer acerca de la forma o patrón de la distribución de los datos: posición, tendencia central y la dispersión o variabilidad alrededor de los valores centrales.

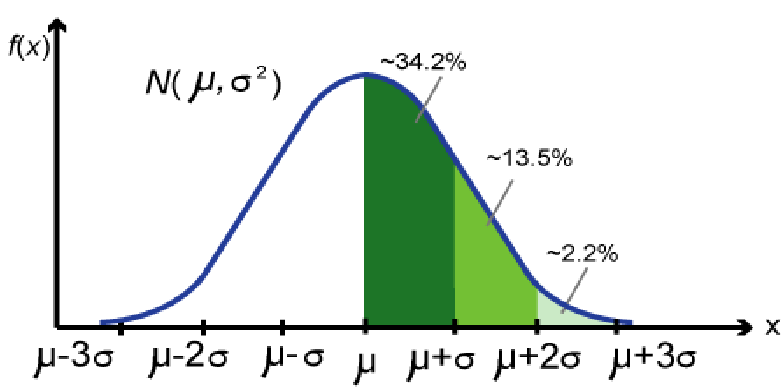
- Es necesario tanto para el análisis e interpretación del conjunto de datos, como para realizar comparaciones entre varios conjuntos de valores, el obtener medidas que resuman o condensen las características del conjunto de datos en cuanto a la posición y la variabilidad.

- El presente capítulo aborda las medidas de posición, y en próximo sobre las medidas de variabilidad

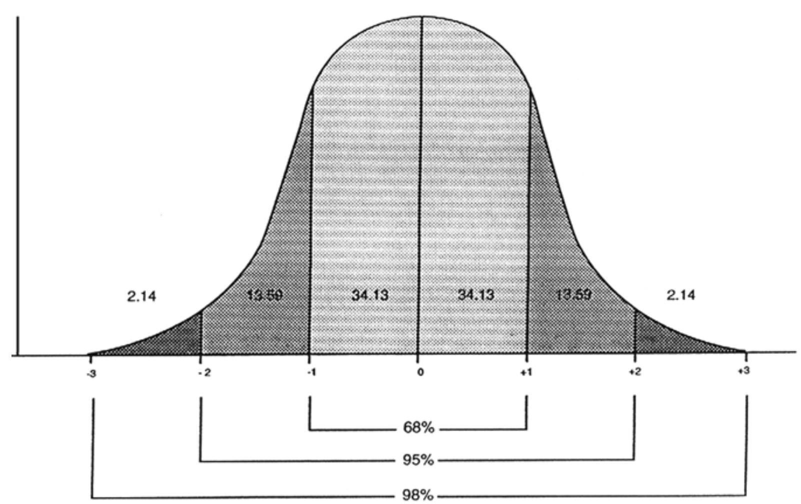


Introducción

- El propósitos de las medidas de posición es tratar de resumir, en un solo número la posición o localización de la distribución. Caso que no contemplaba las medidas relativas
- A las medidas de posición también se les conoce como medidas de posición central, ya que la mayoría de veces importa más la situación central, en donde se distribuyen los datos (este hecho es muy común).

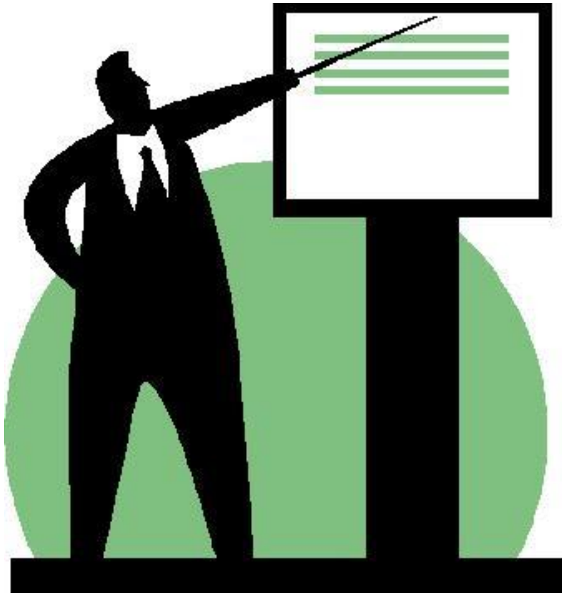


Por diversas razones, las posiciones centrales son las más utilizadas



Introducción

- En resumen, la tendencia central o posición de un conjunto de datos, puede expresarse en varias formas o tomando en cuenta diferentes dimensiones, por ello la Estadística utiliza varias medidas de posición:



a. Moda

b. Mediana

c. Media aritmética o promedio

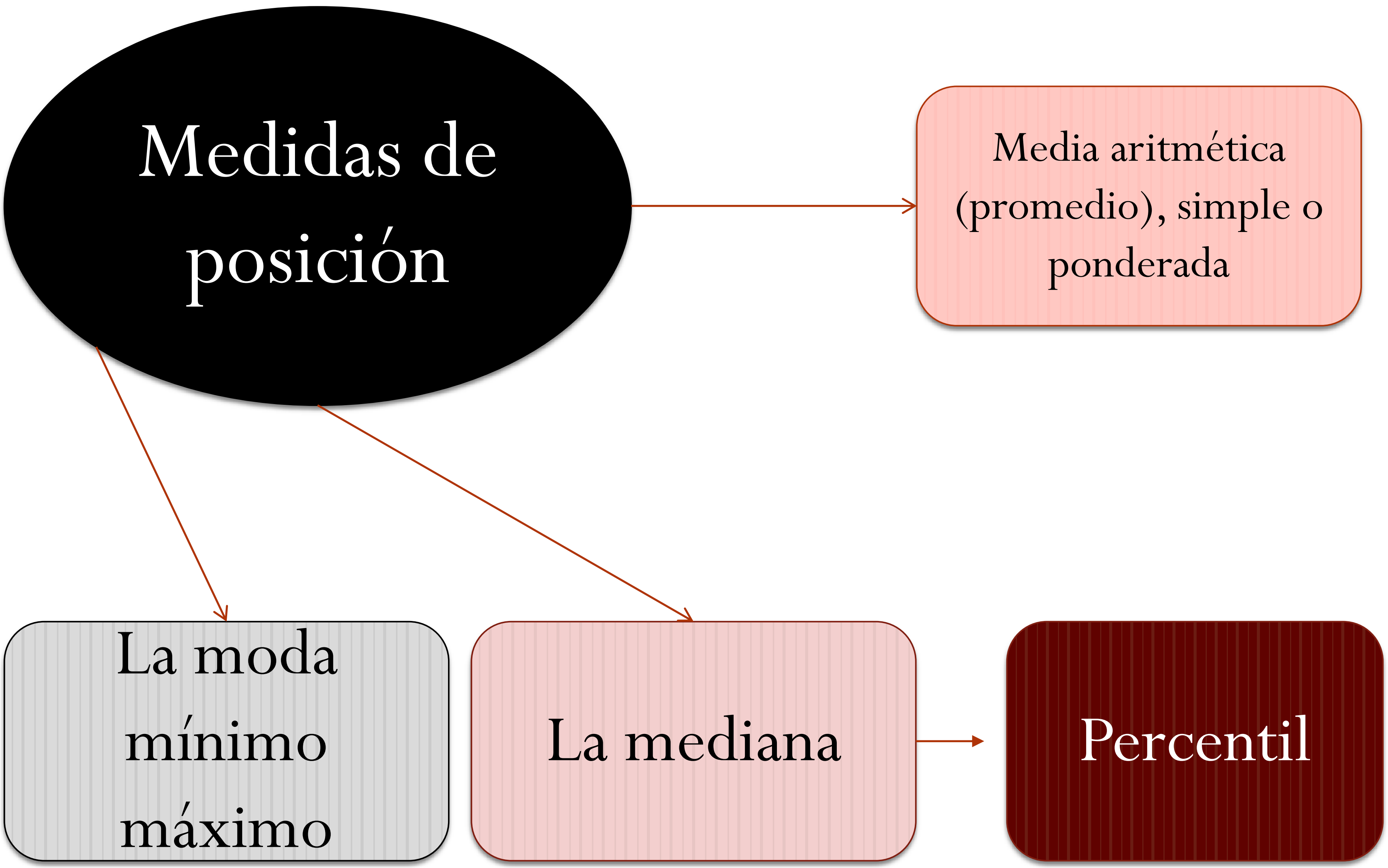
d. Media geométrica

e. Media armónica

f. Cuantilos - percentiles

No vamos a verlas

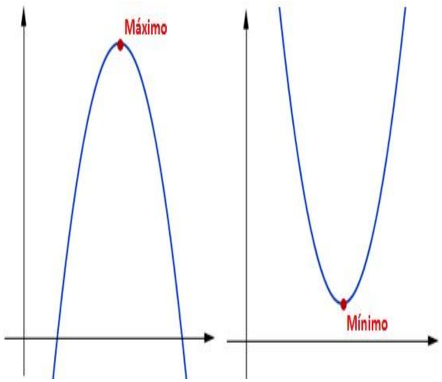
Las medidas de posición



Índice

Valor mínimo y valor máximo

- El valor mínimo se refiere al o los valores con la menor denominación en la variable de interés.



- El valor máximo se refiere al o los valores con la mayor denominación en la variable de interés.

- ¿Esto en qué ayuda a determinar la posición de una distribución?

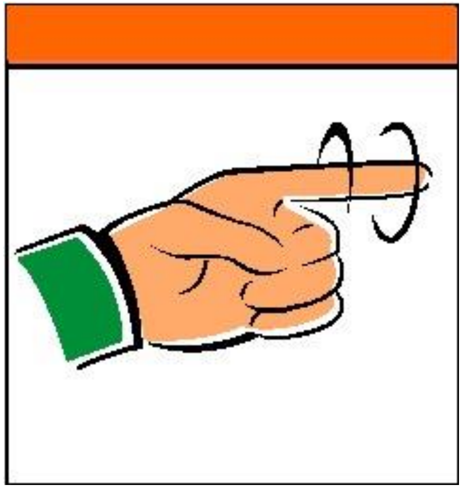
La moda (Mo)

- Esta medida de posición se asocia con el valor más común, más típico o que ocurre más frecuentemente en un conjunto de datos.

- Se define como el valor al cual corresponde la mayor frecuencia.
- Sea el siguiente conjunto de datos:
- 14, 15, 17, 17, 21, 21, 21, 21, 33, 36, 40
- ¿Cuál sería la moda en este caso?

La moda (Mo)

- La moda es una medida muy natural para describir un conjunto de datos, su concepto se adquiere fácilmente: es el dato más común.



- La principal limitación está en el hecho de que requiere un número suficiente de observaciones para que se manifieste o se defina claramente.

- La moda en ciertos casos puede no existir, no estar definida, e incluso si existe puede no ser única.

- En éste ultimo caso, la serie de datos tiene más de un valor modal.

La moda (Mo)

- Por ejemplo, sea la siguiente serie de datos:

  152, 178, 160, 148, 165, 155, 164

En este caso no hay una moda definida.

- En este otro caso:

  50, 55, 55, 55, 62, 73, 73, 73, 80

Se presentan 2 modas, ya que los valores 55 y 73 se repiten tres veces.

- En estos 2 últimos ejemplos, la utilidad de la moda es muy reducida y difícil de interpretar por tener pocos datos.

La moda (Mo)

- Finalmente, la moda se puede aplicar para datos cuantitativos como cualitativos.

- Ejemplo: marcas de computadoras

Dell	Toshiba	Toshiba	Toshiba	Toshiba	Toshiba
Acer	Toshiba	Toshiba	Toshiba	HP	Toshiba
HP	Acer	HP	HP	Dell	HP
Dell	Toshiba	HP	Dell	HP	Toshiba
HP	HP	Dell	Toshiba	Toshiba	Toshiba
Toshiba	Acer	Toshiba	HP	HP	Dell

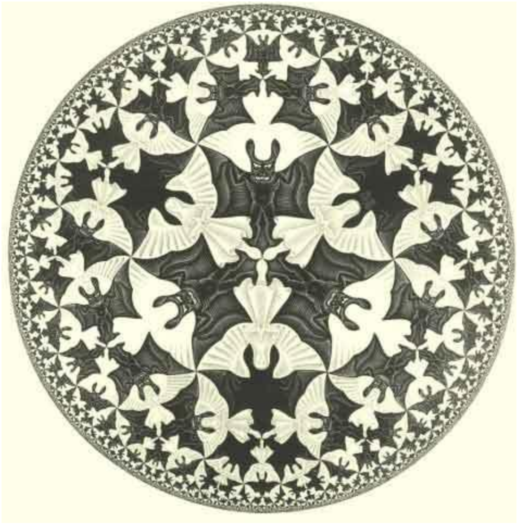
¿Cuál es la moda?

Índice

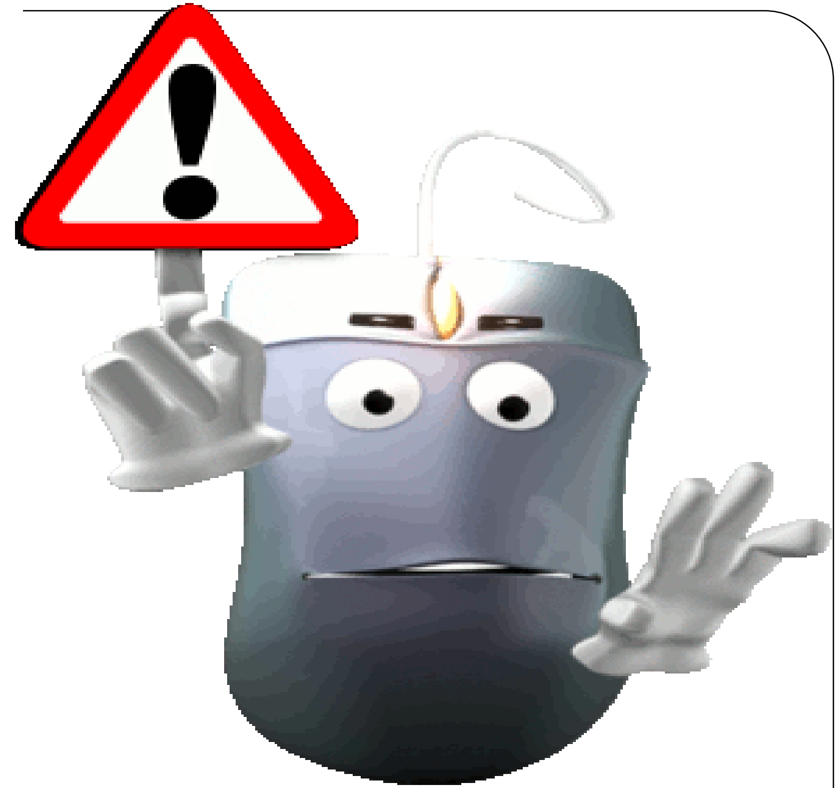
La mediana

- La mediana se define como el valor central de una serie de datos ordenados de tal forma que no más de la mitad de las observaciones son menores que ese valor, y no más de la mitad son mayores a este valor.

- Para calcular la mediana es necesario, en primero lugar, ordenar los datos de acuerdo a su magnitud. Luego se determina el valor central de la serie, y esa es la mediana.



- La mediana tiene 2 formas de calcularse: si se tiene un conjunto total de datos sea par o impar.



¡NO CONFUNDIR EL TÉRMINO EN DONDE SE ENCUENTRA LA MEDIANA CON RESPECTO AL VALOR DE ESTA!

La mediana: conjunto impar

<u>Datos impar</u>

$$6, 8, 8, 10, 12, 19, 29 \qquad n = 7$$

La mediana se encuentra en el siguiente término:

$$n = 7: \qquad \frac{n+1}{2} = \frac{7+1}{2} = 4\ término$$

Ahora, la media sería el valor del 4[to] término, lo que ahora si daría valor de la mediana:

$$Mediana = 10$$

La mediana: conjunto par

<u>Datos par</u>

$$3, 4, 4, 5, 16, 19, 25, 30 \qquad : n = 8$$

La mediana se encuentra en el siguiente término:

$$n = 8: \qquad \frac{n+1}{2} = \frac{8+1}{2} = 4,5\ \textit{término}$$

En este caso la mediana se encuentra entre el 4° y 5° término, es decir entre los valores 5 y 16. En este caso el valor de la mediana sería un promedio de estos valores.

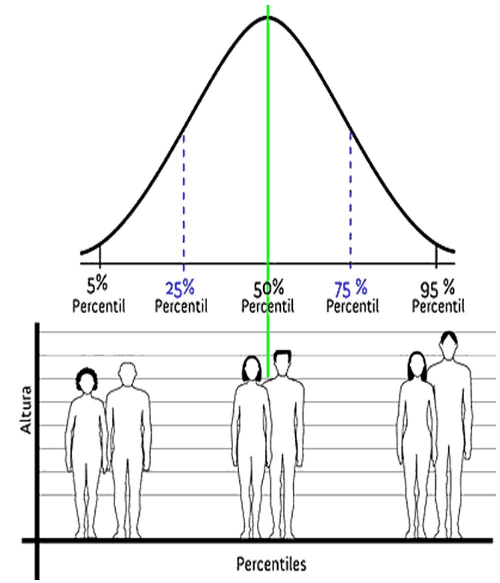
$$Mediana = \frac{5+16}{2} = 10.5$$

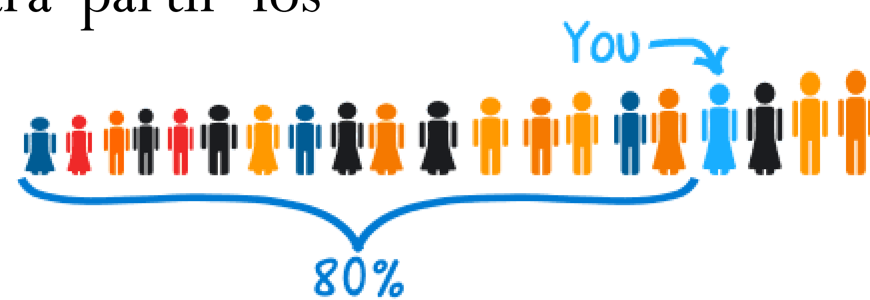
El percentil

- La mediana también es conocida como el percentil 50.

- Puede que nos interesa no solo dividir los datos en dos partes, sino en 4, o tener los grupos de 10, etc.

- Puede que nos interese también conocer el punto 17°, o el 67°, etc. Por lo que interesa conocer una formula general para partir los datos como lo hace la mediana

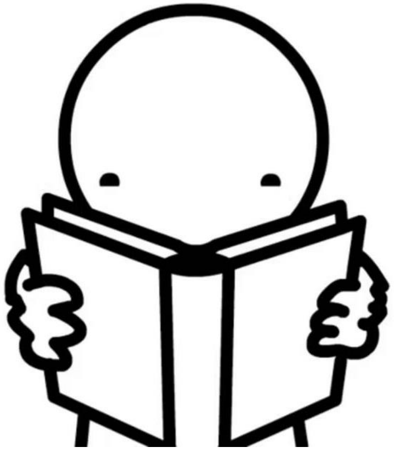




El percentil

- **El percentil** es una medida de tendencia central que indica, ordenados los datos de menor a mayor, el valor de la variable por debajo del cual se encuentra un porcentaje dado de observaciones en un grupo de observaciones. Una definición más formal:

"Dado un conjunto de n observaciones $x1, x2, \ldots, xn$, el *p-ésimo* percentil **P** es el valor de ***X***, tal que p por ciento o menos de las observaciones son menores que *P*, y (*100-P*) por ciento o menos de las observaciones son mayores que *P*."



El percentil

- Supongamos que queremos obtener el percentil 15 en un conjunto de datos, la fórmula general sería:

$$Posición\ Percentil\ ° = \frac{Percentil}{100}(n+1)\Big)$$

- Si suponemos que tenemos un muestra de tamaño 50, la posición del percentil estaría dada por lo siguiente:

$$Posición\ Percentil\ ° = \frac{15}{100}(50+1) = 7.65$$

- El valor exacto se obtiene mediante la fórmula:

$$Valor\ Percentil = L_i + \frac{\frac{k*n}{100} - f_a - 1}{f} * a_i$$

KEEP
CALM
AND
DO MORE
CALCULUS

Los cuartiles

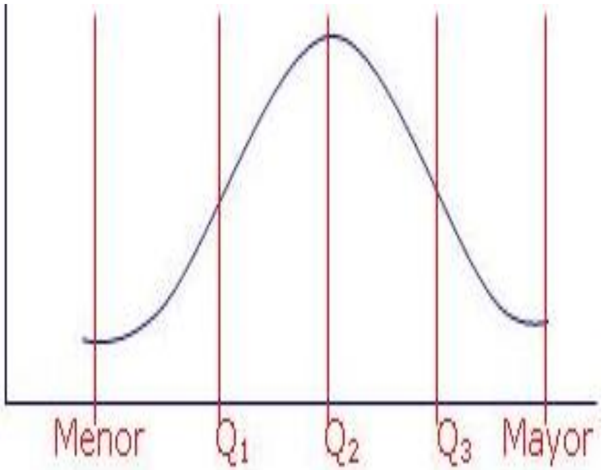
- En la práctica, los percentiles no son tan utilizados como los cuartiles.

- Los cuartiles parten la distribución de datos en 4 partes. Su representación matemática es la siguiente:

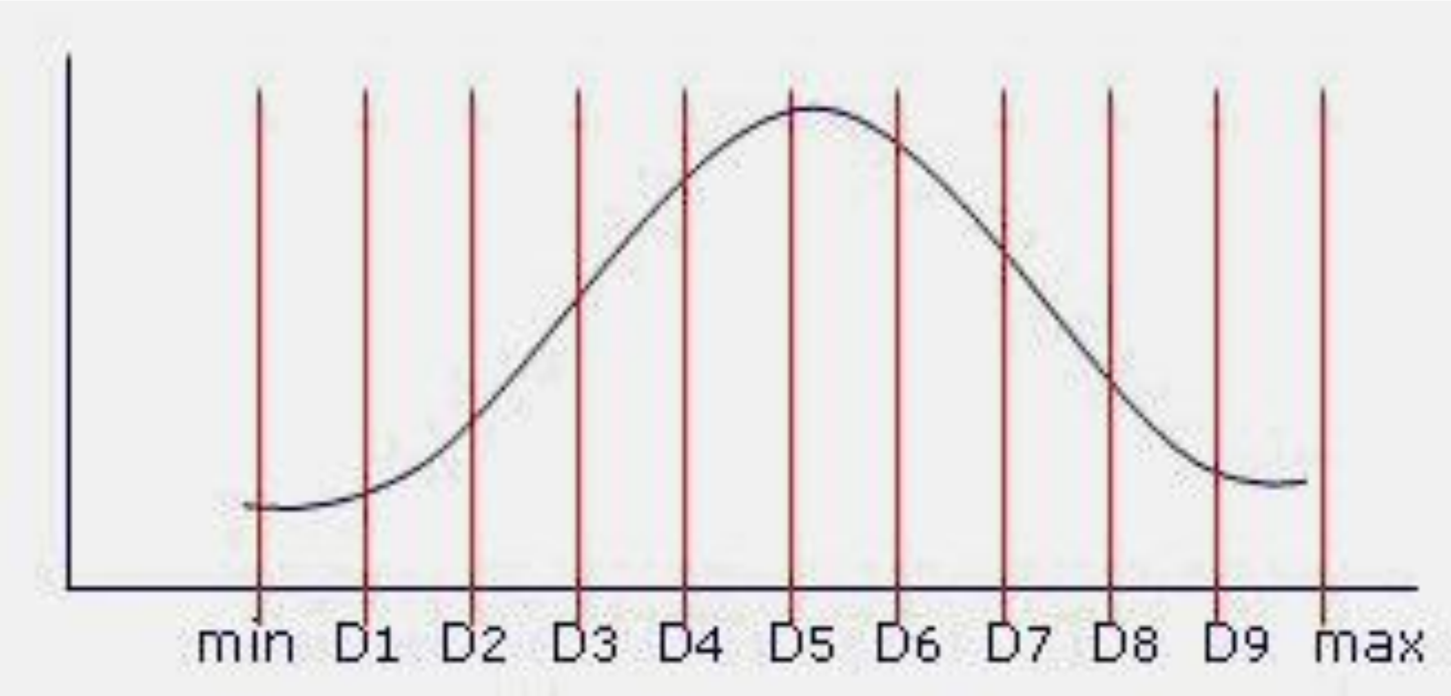
$$Q_1 = \frac{(n+1)}{4}, ésima\ observación\ ordenada$$

$$Q_2 = \frac{2(n+1)}{4} = \frac{n+1}{2}, ésima\ observación\ ordenada$$

$$Q_3 = \frac{3(n+1)}{4}, ésima\ observación\ ordenada$$



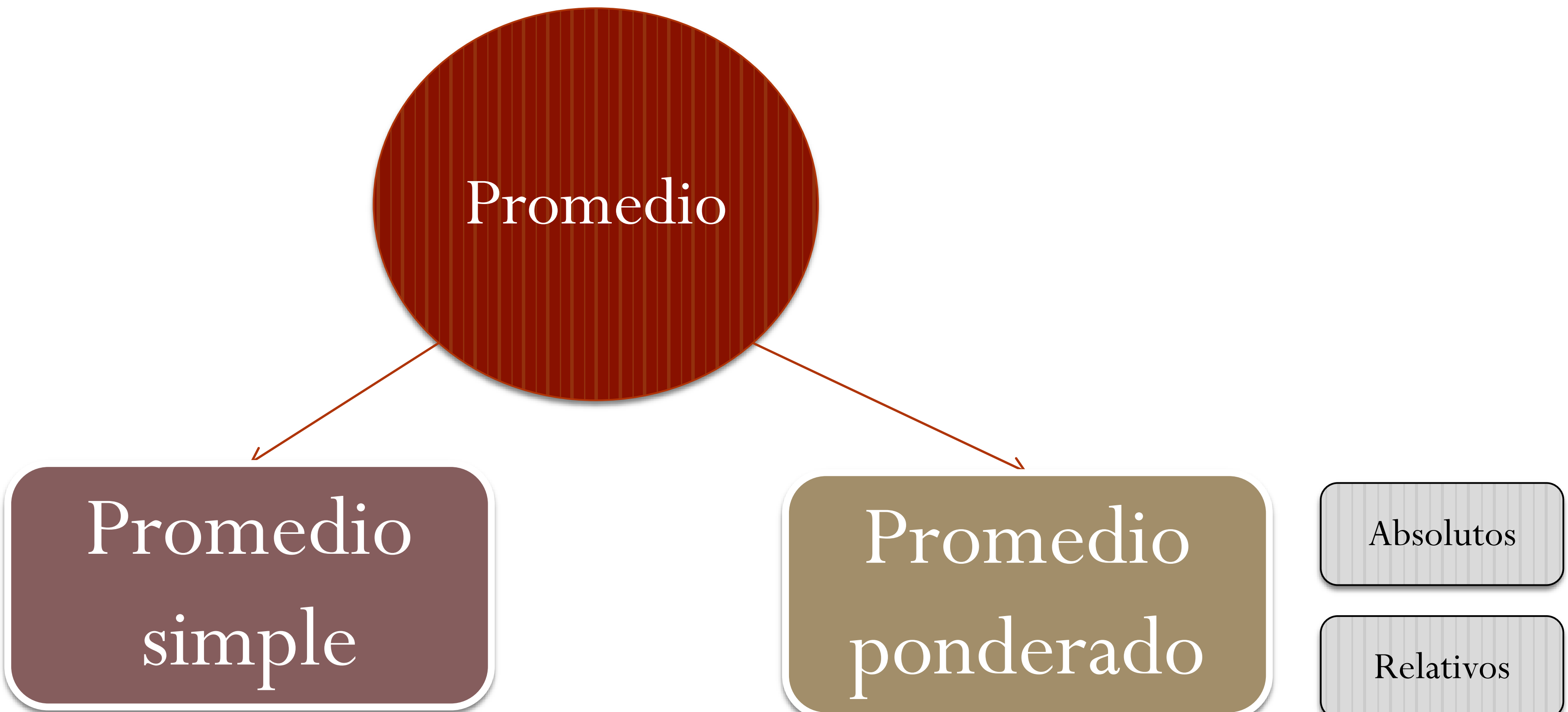
Para los deciles…. cómo sería?



Índice

La media aritmética ($\bar{x}$) o promedio

- Es la medida de posición más usada y conocida.
- Existen 2 formas alternativas para su cálculo



Promedio simple

- Es un conjunto de valores, es el resultado que se obtiene al dividir la suma de esos valores entre el número de ellos.

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

$$Promedio = \frac{Suma\ de\ los\ valores}{Número\ de\ valores}$$

- Por ejemplo, si en una clase hay 12 personas cuyas edades son:

$$20, 20, 22, 20, 30, 25, 25, 18, 20, 18, 22, 36$$

- La edad promedio de esas personas sería:

$$\bar{x} = \frac{20 + 20 + \cdots + 22 + 36}{12} = \frac{276}{12} = 23\ años$$

$$\bar{X} = \frac{\sum x}{n}$$

- La importancia del promedio es que para un conjunto de datos, esta medida es por mucho la que la gente prefiere saber (propiedades matemáticas).

Ejemplos de promedios

Promedio ponderado: absoluto

- Muchas veces los valores de interés no tienen la misma ponderación, o no tienen el mismo "valor" con respecto a otros casos.
- En este caso para determinar el promedio es necesario ponderar los resultados, por lo cual realizamos un promedio ponderado.

Ejemplo: notas de curso

Nota curso	Créditos	N*Curso
85	1	85
90	1	90
95	3	285
95	4	380
100	7	700
Total	16	1540

Promedio ponderado: absoluto

- El promedio ponderado sería las multiplicaciones entre las observaciones de la característica y el ponderador (curso y créditos respectivamente):

$$\bar{x}_{ponderado} = \frac{\sum x_i * w_i}{\sum w_i}$$

- El promedio ponderado sería:

$$PP = \frac{X_1P_1 + X_2P_2 + \cdots + X_nP_n}{(p_1 + p_2 + \cdots + p_n)}$$

$$PP = \frac{85 * 1 + 90 * 1 + \cdots + 100 * 7}{16}$$

$$PP = \frac{1540}{16} \quad \Rightarrow \quad PP = 96.25$$

Promedio ponderado: relativo

- El promedio ponderado también se puede utilizar con números relativos.

- Por ejemplo, si se quisiera un promedio, en donde en términos porcentuales se le de más énfasis a cierta característica u otro, es un caso de promedio ponderado para números relativos.

- Este tipo de promedio se utilizan mucho en estimaciones con poblaciones desbalanceadas, calificaciones son variables con diferentes pesos, etc.

- El es promedio con mayor aplicación….

Promedio ponderado : relativo

- Supóngase que una encuesta de satisfacción quiere conocer el grado de satisfacción de las carretas según la provincia de residencia.

- Participaron personas de 4 provincia, y los valores promediados de satisfacción y el porcentaje de participación por zona se presenta a continuación:

Provincia	Ponderación	Promedio
San José	0,6	78
Alajuela	0,2	79
Cartago	0,1	92
Heredia	0,1	94

…..cuál sería el nivel medio de satisfacción?

Promedio ponderado : relativo

- La fórmula del promedio ponderado sería:

$$\bar{x}_{ponderado} = \frac{\sum x_i * w_i}{\sum w_i}$$

- Para el presente caso, la satisfacción media sería:

$$\bar{x}_{ponderado} = \frac{0{,}6 * 78 + \cdots + 0{,}1 * 94}{0{,}6 + \cdots + 0{,}1} = 81{,}2$$

- Nótese que en ambos casos, la sumatoria del denominador sería la suma de la variable de ponderación

Índice

Usos de las medidas de posición

¿Cuál Medida debo utilizar? Depende……

Usos de las medidas de posición

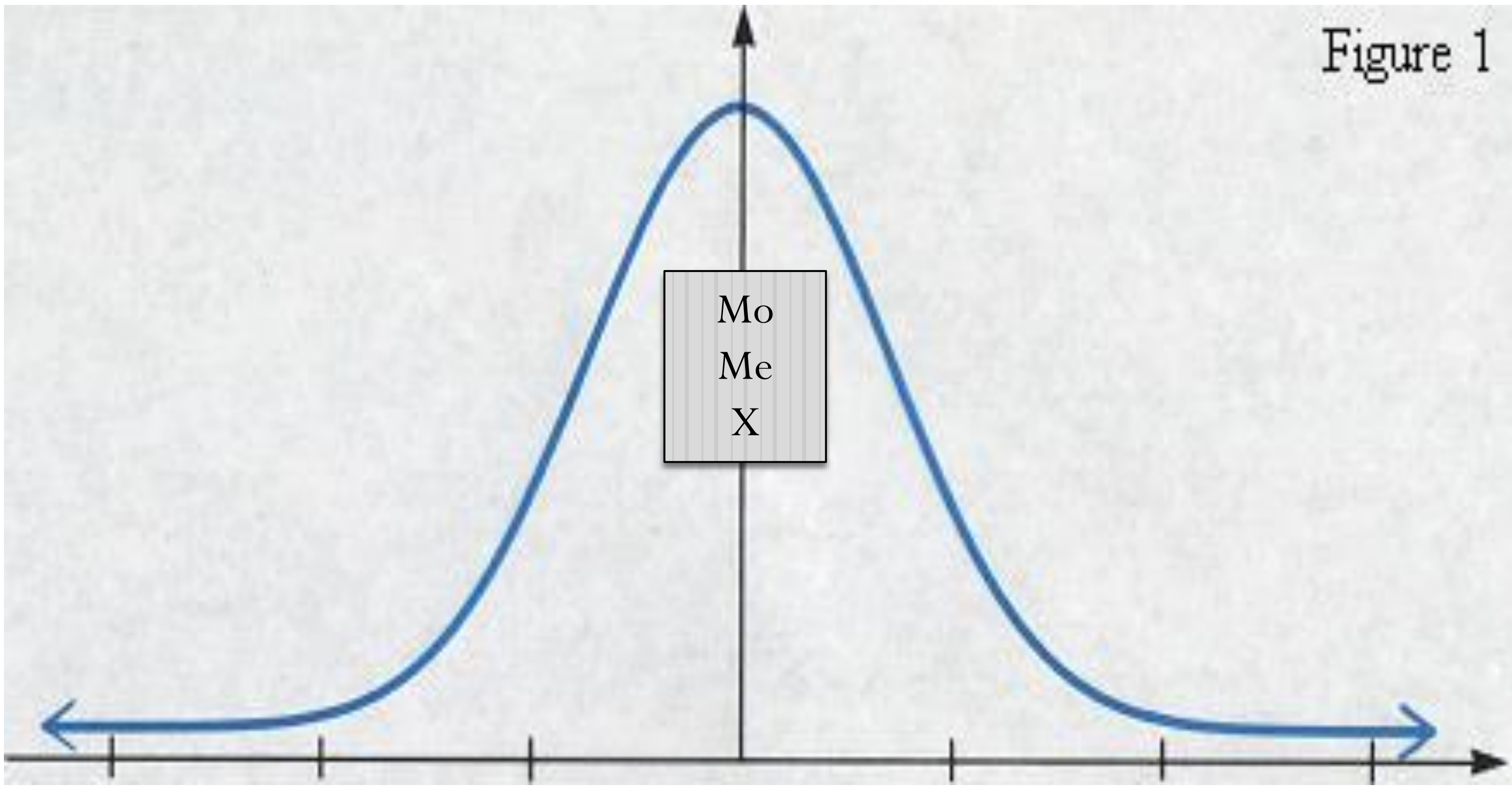
- El propósito de las M.d.P. es caracterizar y presentar un conjunto de datos; y cada una de las diferentes medias propuestas lo hacen de alguna manera.

- Dependiendo del propósito que se tenga, una medida es mejor que otra, sin embargo estas no compiten, sino que se complementan.

- Una parte del problema acerca de qué medida usar, desaparece si se tiene una idea clara de cuál es el aspecto del conjunto de datos que se quiere conocer, y si el conjunto de datos posee o no valores extremos bajos o altos

Usos de las medidas de posición

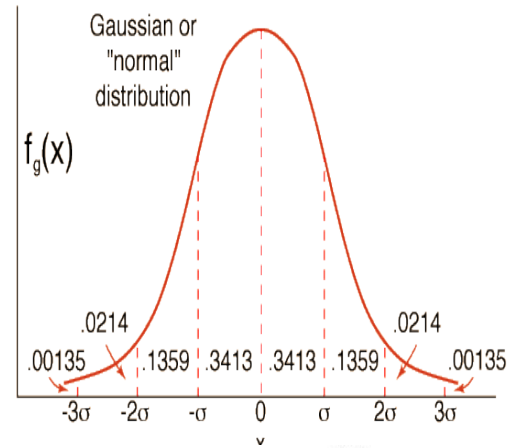
- Sea la siguiente distribución de datos:



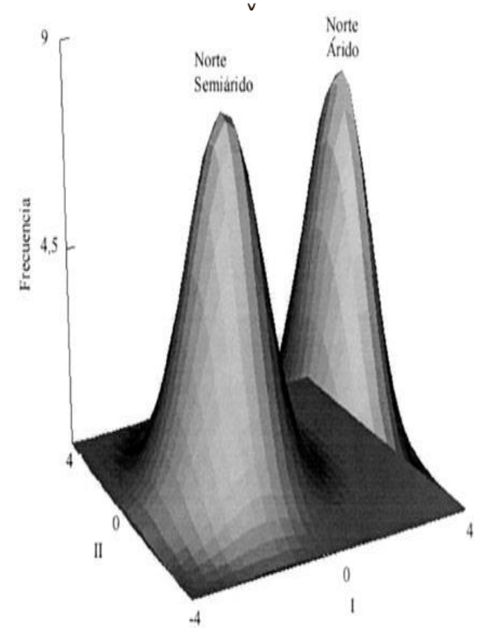
Distribución simétrica

Usos de las medidas de posición

- Cuando los datos se distribuyen normalmente, la Mo, Me, y $\bar{x}$ coinciden (valen lo mismo). En este caso, cualquier es bueno para caracterizar el grupo de datos, pero igual se prefiere al $\bar{x}$.



- De igual forma, si las 3 medidas no coinciden mucho, pero no hay un alejamiento extremo, se escoge el $\bar{x}$ por tener mejores propiedades estadísticas.

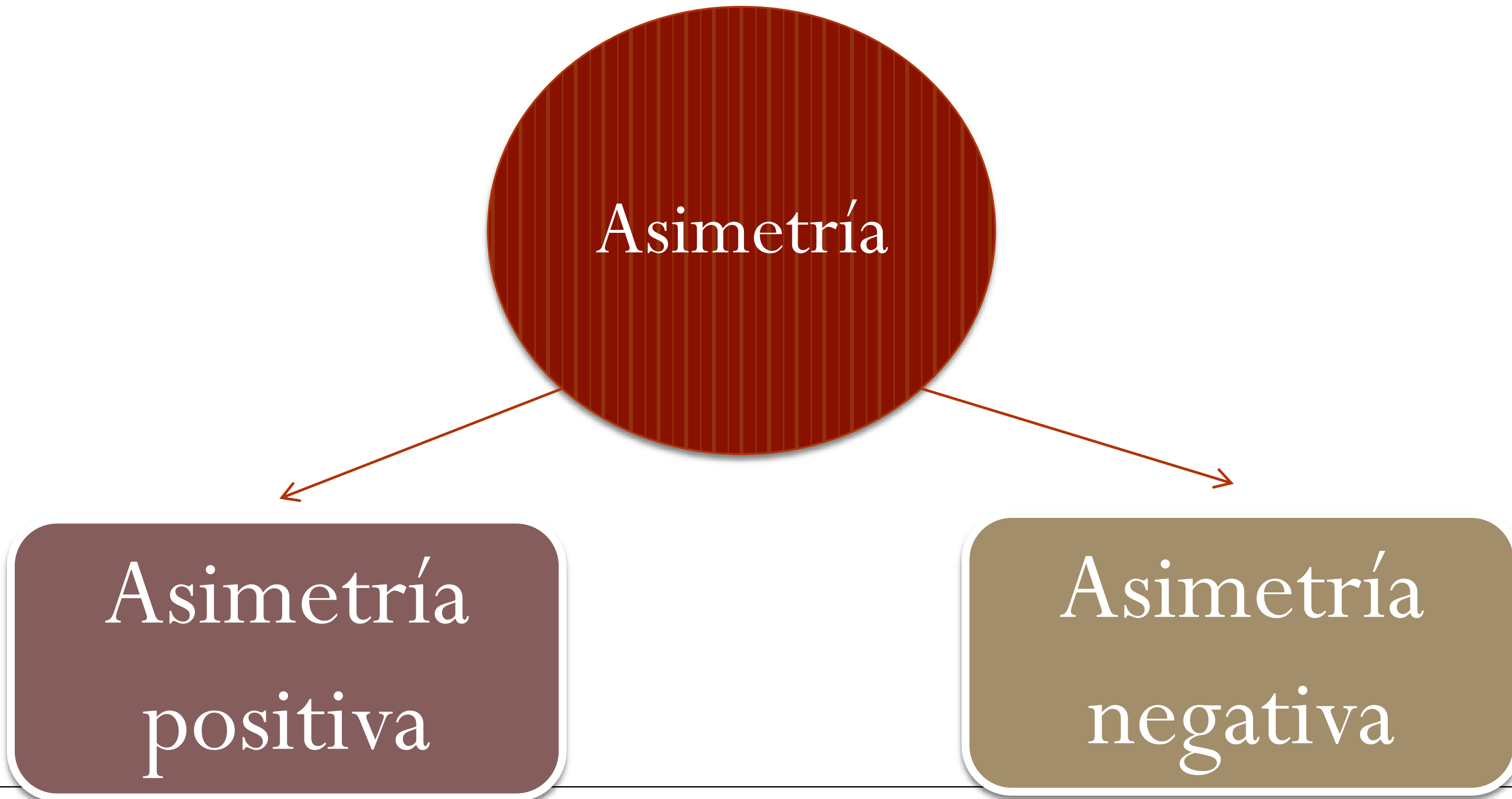


- Casi siempre, el $\bar{x}$ es la medida de preferencia para caracterizar un conjunto de datos.

Usos de las medidas de posición

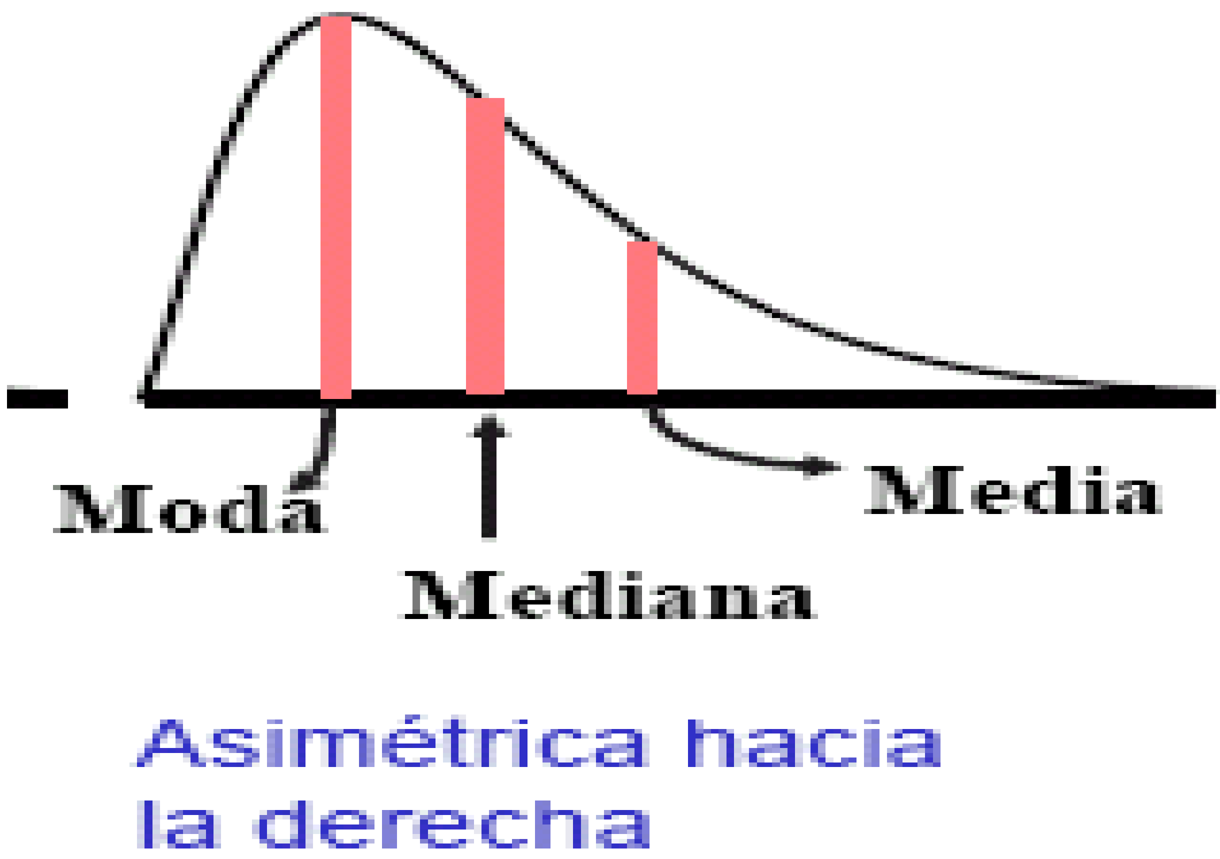
- Si la distribución de datos es asimétrica (una cola más extendida que la otra) entonces se prefiere una medida que otra, dependiendo además de lo que se quiere.

- Dos casos de distribuciones asimétricas:



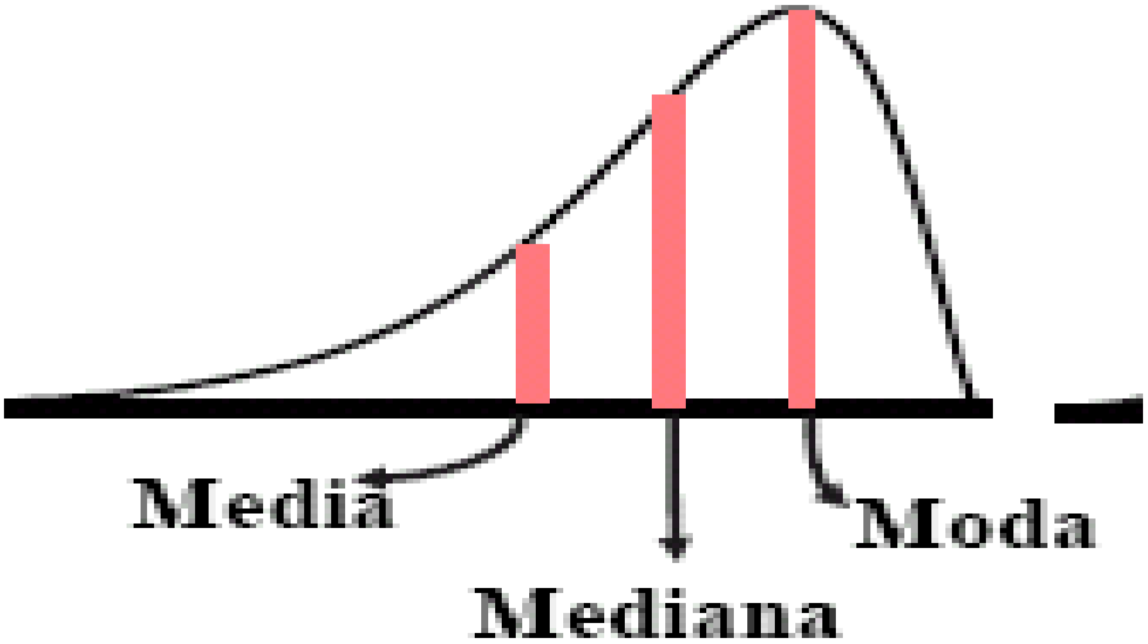
Usos de las medidas de posición

<u>Asimetría positiva</u>



Usos de las medidas de posición

<u>Asimetría negativa</u>



Asimétrica hacia
la izquierda

Usos de las medidas de posición

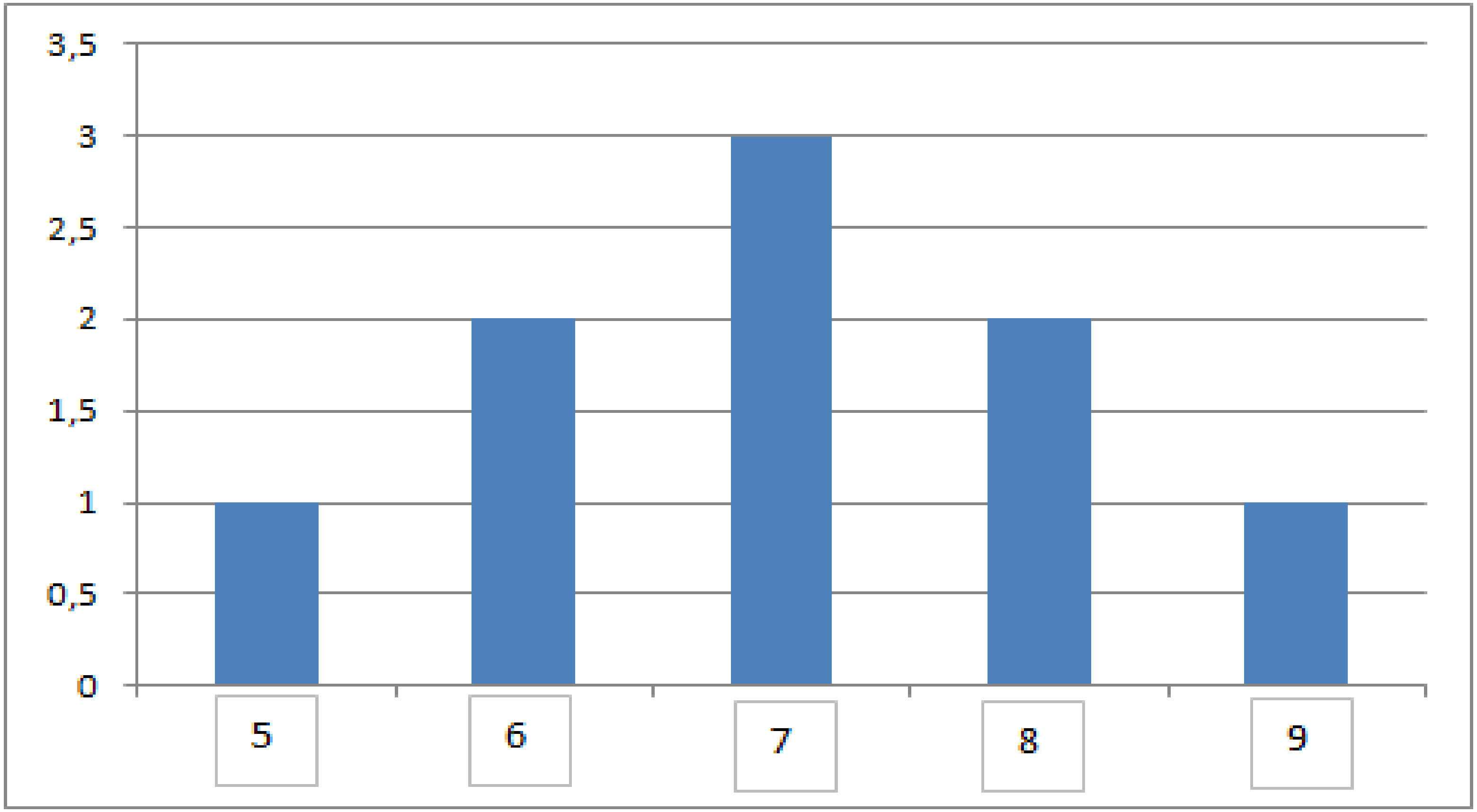
- Este comportamiento de las medidas de posición, en el caso de distribuciones asimétricas caracterizado por un alejamiento de las 3 medidas, se deben a los valores extremos.

- En tales caso se prefiere a la Me y no a las otras dos.

- Veamos unos ejemplos con datos verdaderos.

Usos de las medidas de posición

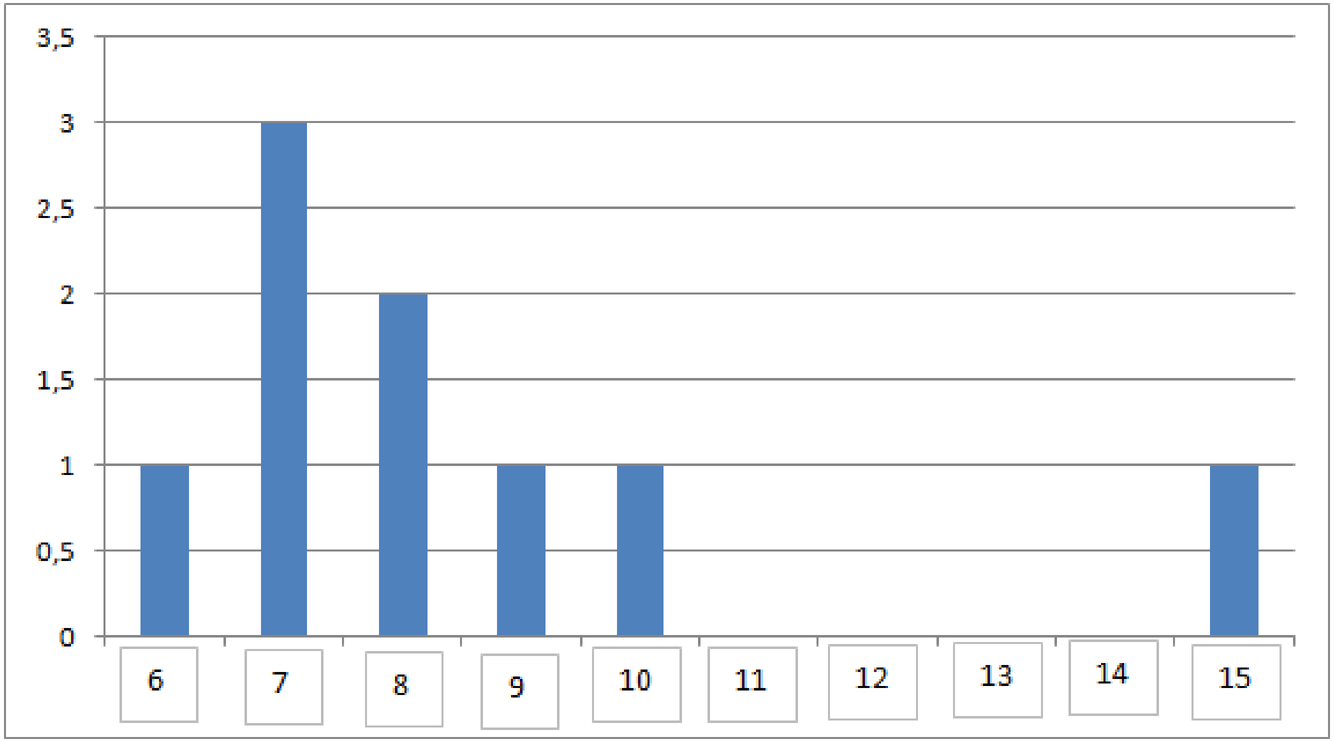
<u>Ejemplo 1</u> $Datos:$ 5,6,6,7,7,7,8,8,9



En este caso, cualquier medida es buena para representar el conjunto de datos.

Usos de las medidas de posición

Ejemplo 2 *Datos:* 6,7,7,7,8,8,9,10,15

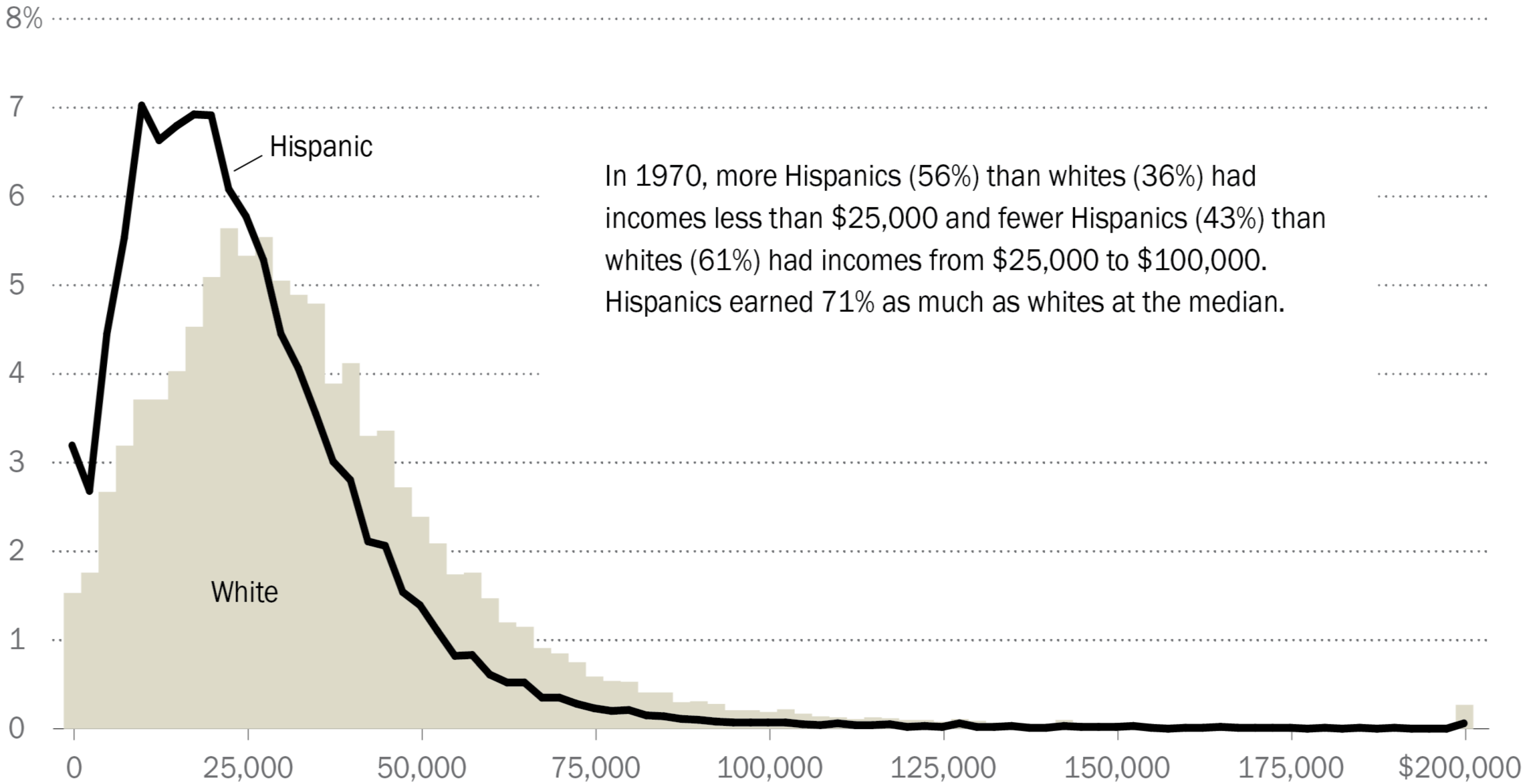


En este caso se prefiere la Mo y la Me para describir el conjunto que el

Usos de las medidas de posición

¿Qué consideran Uds. sobre el uso del salario promedio para determinar la riqueza en un país ?

Shares of whites and Hispanics with a given level of income, 1970



Note: Whites include only non-Hispanics. Hispanics are of any race. Income is adjusted for household size and expressed in 2016 dollars. See Methodology for details.

Source: Pew Research Center analysis of 1970 decennial census (IPUMS).

"Income Inequality in the U.S. Is Rising Most Rapidly Among Asians"

PEW RESEARCH CENTER

Estudio de caso

- Se estudió el salario de las personas adultas mayores. Se quiere analizar a la presente población, por lo que le solicita hacer un análisis de las siguientes estadísticas descriptivas:

Estadística	Valor
Mínimo	4%
Máximo	23%
Moda	12%
Mediana	14%
Percentil 10	6%
Percentil 90	20%
Promedio	17%

Ultimas reseñas

- Las medidas de posición son muy importantes para cualquier análisis estadístico.

- La diferencia de este con respecto a las medidas relativas es que este resume en conjunto pero enfocándose en la posición.

- Los tres más utilizados son la Moda (Mo), Mediana (Me), y Promedio.

- Estas medidas se deben realizar conjuntamente con las medidas de variabilidad (próxima clase).

Usos de las medidas de posición


FIN