

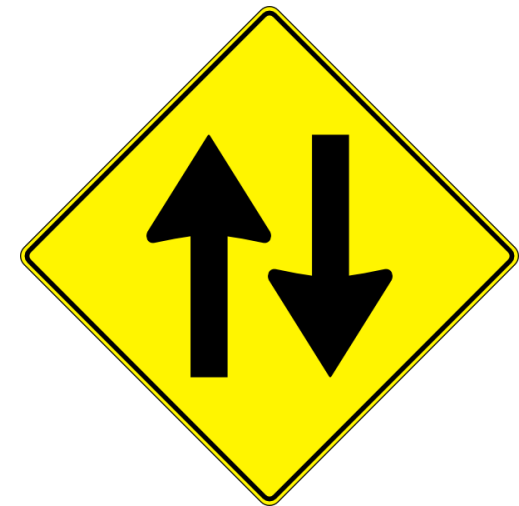
Conceptos básicos, el método científico
números relativos, representación de la
información, medidas de posición y
variabilidad.

Oscar Centeno Mora

Conceptos Básicos

¿Qué es la Estadística?

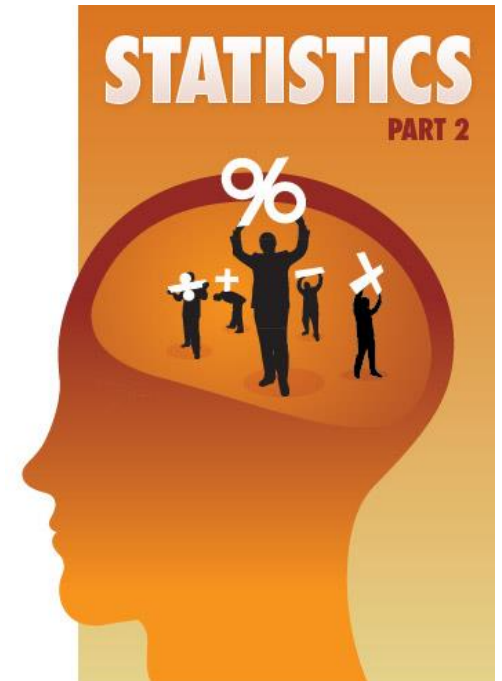
- La definición de *Estadística* se suele confundir muchas veces con lo que se entiende por “datos estadísticos”.
- La palabra *Estadística* es utilizada en varios sentidos en el habla cotidiano.
- Aún los propios estadísticos al definir su propia disciplina tienen ciertos problemas, ya que ponen de relieve unos aspectos más que otros, lo cual parece que muchas veces la definición no sea tan clara.



¿Qué es la Estadística?

- La estadística no es simplemente un conjunto de datos estadísticos, ni está interesada solamente en recolectar información, resumirlos y presentarlos.
- Por ahora brindaremos la siguiente definición de *Estadística*:

“Es la disciplina científica dedicada al desarrollo y aplicación de la teoría y técnicas apropiadas para la recolección, clasificación, presentación, análisis e interpretación de información cuantitativa, obtenida por muestreo, observación o experimentación”
(Gómez, M.)



Conceptos Básicos

```
graph TD; A([Conceptos Básicos]) --> B[UE, característica y observación]; A --> C[Población finita e infinita]; A --> D[Datos estadísticos y análisis estadísticos]; A --> E[Clasificación de la naturaleza de los datos]; A --> F[Escala de medición];
```

UE, característica y observación

Población finita e infinita

Datos estadísticos y análisis estadísticos

Clasificación de la naturaleza de los datos

Escala de medición

Unidad Estadística (UE), característica y observación

- En el análisis básico estadístico, se debe siempre conocer tres cosas:
 - La unidad de estudio*** que se quiere estudiar (**UE**).
 - Qué es lo que se quiere saber de la unidad de estudio (***característica***).
 - Qué es lo que vamos a obtener de la característica de la unidad de estudio (***la observación***).
- Ejemplo 1:

“Una oficina gubernamental desea conocer para cierta zona del país, el salario semanal promedio de los trabajadores industriales. Se llevó a cabo una encuesta telefónica para dicho fin.” Determine:

La UE:

Característica:

Observación:



Unidad Estadística (UE), característica y observación

- Ejemplo 2:

“Un Ingeniero desea verificar la calidad de una parida de bombillos producidos el día anterior. Para hacerlo, toma una muestra de 50 bombillos y determina, para cada uno, si enciente o no”

UE:

Característica:

Observación:



Población finita e infinita

- Todo estudio o investigación tiene como referencia un conjunto de unidades de estudio o elementos (personas, animales, empresas, bombillos, etc.). El estudio pretende conocer las características y generalizar los elementos de una ***la Población***.
- Una forma de definir ***la Población*** es: “el total o agregado de todas las unidades de estudio”.
- Ejemplo: todas las personas de un país, todos los peces de un estanque, todos los pacientes de un hospital.



Población finita e infinita

- A su vez, se pueden definir el término de **Población** como:

1. **Población finita:** posee un número limitado de elementos, especificada en el tiempo, y los elementos son contables.
2. **Población infinita:** hay un número ilimitado de elementos, no está especificada en el tiempo, y los no pueden ser contabilizados.

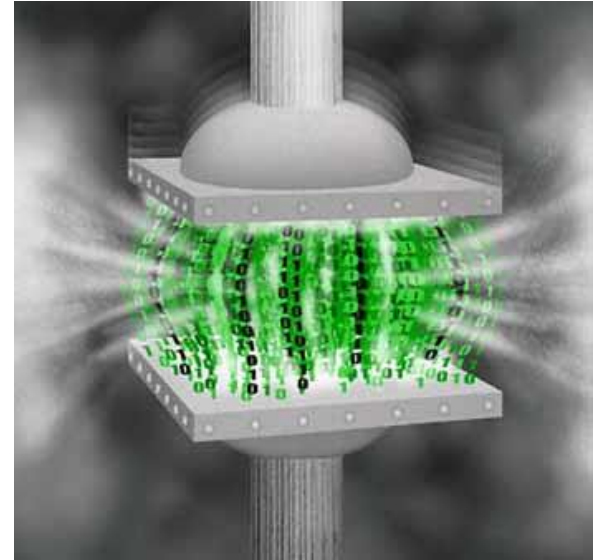


Ejemplo de PF: población a una cierta fecha, alumnos de un curso, animales de cierto parque. Todo lo anterior definido en tiempo, espacio.

Ejemplo de PI: población actual y futura que no es limitada en el tiempo, poblaciones que realmente no se pudieron contabilizar.

Datos Estadística y Análisis Estadísticos

- **Datos estadísticos:** son representativo crudas de los datos. Los cuadros, los gráficos, y datos sin un fin de aplicación de una técnica inferencial de por medio, son considerados como datos estadísticos.
- **Análisis estadísticos:** son todas las técnicas estadísticas que utilizan la matemática (probabilidades) y teoría estadística, para ser aplicada en la solución de un problema (en su mayoría inferencial).



Clasificación de la naturaleza de los datos

- En *Estadística*, normalmente la información o datos son de dos tipos:
- ***Datos cualitativos***: es toda la información que no se expresa necesariamente por números, y por lo tanto se suele denominar como características. Son muy comunes en estudios cualitativos (psicología, mercadeo, estudios de opinión, etc.).
- ***Datos Cuantitativos***: involucra necesariamente datos numéricos. Muchos ejemplos se hayan en campos como la economía, demografía, y por supuesto en el área de la salud.

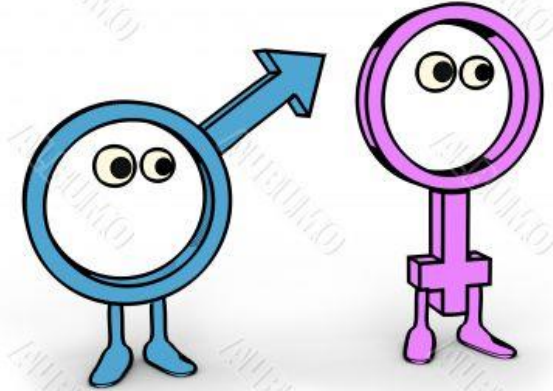


Escala de medición: nominal

- La información o los datos (cualitativos como cuantitativos) se encuentran siempre en una de las siguientes 4 escalas de medición:

Nominal: cuando se tiene una característica para la cual se pueden definir categorías diferentes, pero no es posible ordenar esas categorías, ni decir de qué manera una es superior a la otra. En este caso se tendría una escala **nominal**.

Ejemplos: sexo, estado civil, país de nacimiento, color del pelo, etc.



Escala de medición: ordinal

Ordinal: en un sistema de medición nominal, no es posible decir si dos características son iguales o diferentes. En los datos cualitativos hay situaciones en donde no se tiene una unidad de medida numérica, pero estamos en capacidad de determinar si los objetos tienen más o menos de la característica de interés, y por lo tanto es posible ordenarlos. Cuando tenemos datos cualitativos pero se pueden de cierta forma ordenar, estamos en presencia de una escala de medición **ordinal**.

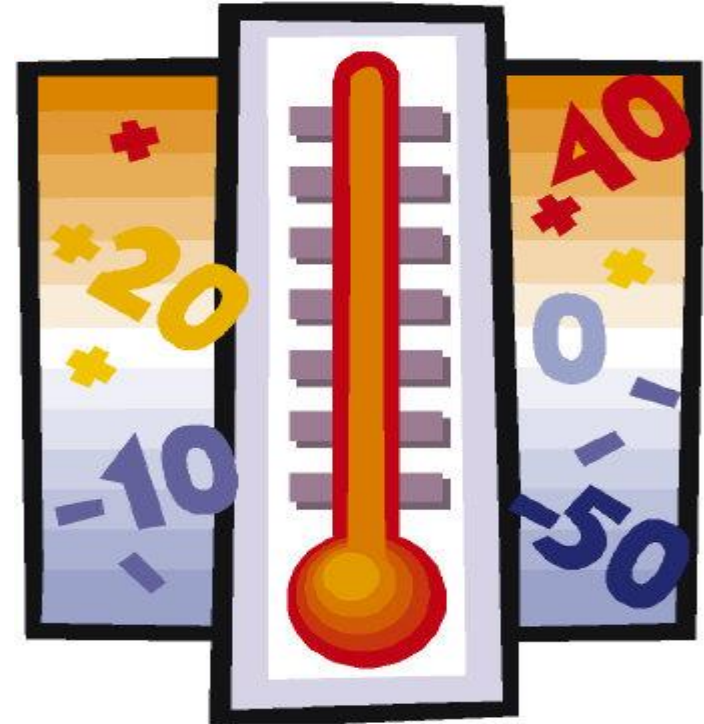
Ejemplos: grado académico, intensidad del dolor, posición militar, etc.



Escala de medición: intervalo

Intervalo: a partir de esta escala ya se tienen informaciones numéricas. La diferencia entre la escala de medición por **intervalos** y la de **razón**, es que en la de intervalos el “cero” (0) es fijado arbitrariamente, por lo que no se indica ausencia de lo que se mide. De igual forma, información proveniente de hechos subjetivos o valoraciones se pueden clasificar como escalas por **intervalo**.

Ejemplos: la temperatura, el tiempo, la medición del IQ, las escalas Likert, etc.



Escala de medición

Razón: al igual que en la escala de intervalo, las observaciones derivadas de esta contienen observaciones numéricas. Sin embargo en este caso el “cero” (0) si es verdadero y absoluto, e indica ausencia de la característica de interés. Así, por ejemplo, si tenemos una observación de cero, entonces es que no hay nada.

Ejemplo: peso, volumen, cantidad de glóbulos rojos, dinero, altura, etc.

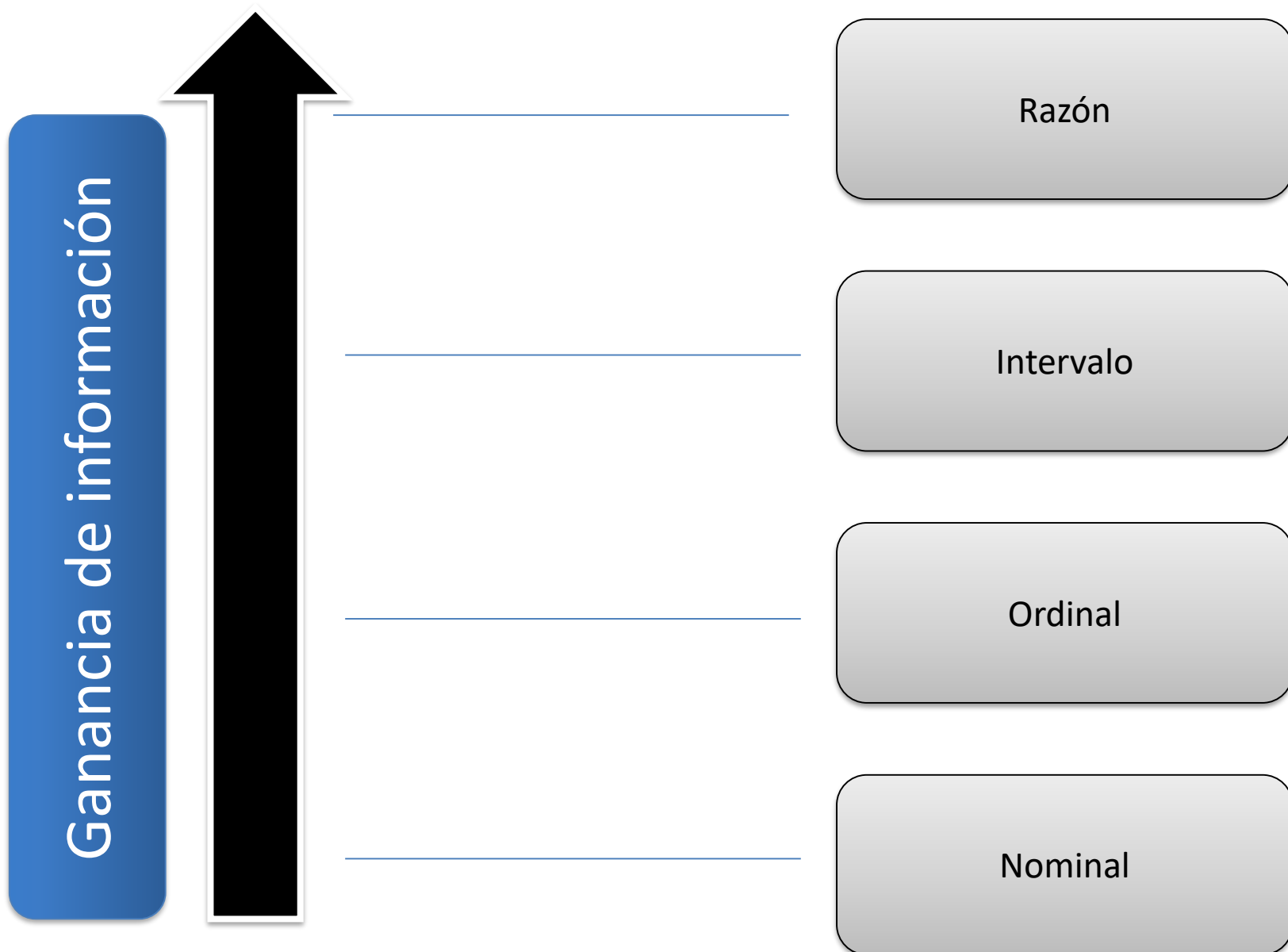


Escala de medición

- En estadística solemos llamar a las dos primeras escalas de medición como datos categóricos o cualitativos, mientras que las últimas escalas se denominan como mediciones métrica o datos cuantitativos.
- ¿Para qué sirven los niveles de medición en estadística?
- Esto ayuda a saber que tipo de pruebas o análisis estadístico se debe llevar a cabo.



Escala de medición



El método científico

Fases en una investigación en Estadística

- Formalmente consta de 9 fases o etapas:

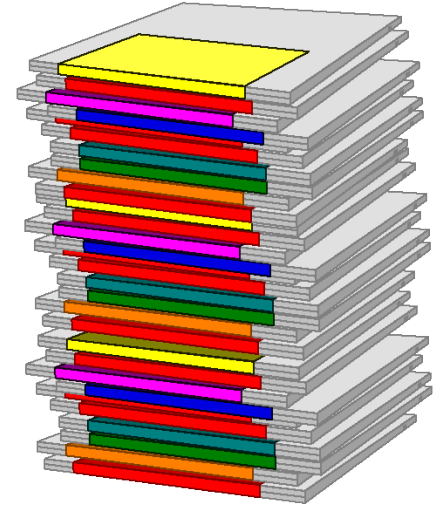
1. Definición y delimitación del problema de interés.
2. Establecimiento de los propósitos específicos del estudio.
3. Preparación de un plan de trabajo.
4. Construcción de una herramienta de recolección.
5. Diseño y selección de la muestra.
6. Preparación y ejecución del trabajo de campo.
7. Procesamiento de la información.
8. Análisis e interpretación de los datos.
9. Preparación del informe.

A todo un proceso de investigación también se le conoce como **“cadena de eslabón”**

Números relativos

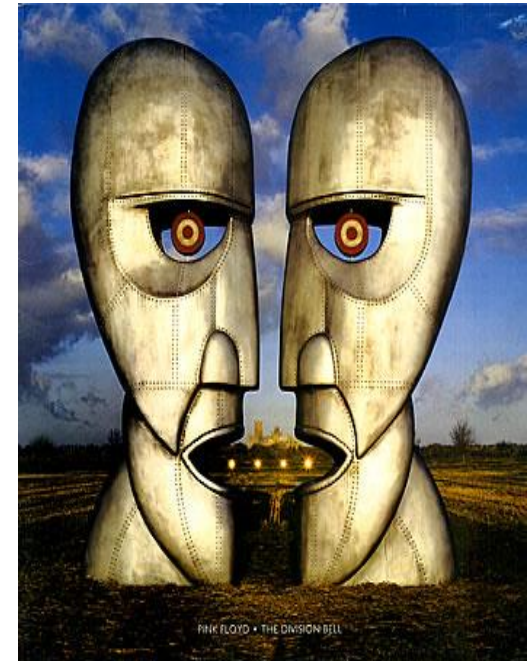
La necesidad de resumir la información

- Los número relativos más populares son:
 - Las tasas.
 - Las proporciones (porcentaje).
 - Las razones.
 - Distribución de frecuencias
 - Indicadores e Índice.
- La importancia de los número relativos consiste en poder resumir de mejor forma la información, además de mejorar la calidad de la representación de la información, lo cual es la gran debilidad de los números absolutos.



Razón

- Es la relación entre dos número positivos.
- Por ejemplo, si “A” y “B” son dos número positivos, y dividimos A entre B, obtenemos la razón $\frac{A}{B}$.
- Esto nos dice cuántas veces A es más grande que B (si el resultado es mayor a 1), o análogamente, que cantidad de A por cada cantidad de B.
- Los datos que se relacionan pueden pertenecer al mismo universo de datos o provenir de un universo distinto.



Tasas

- El término de **tasa** se confunde con muchos otros números relativos .
- Por ejemplo, se suele hablar de tasa de alfabetismo, cuando en realidad se le debe atribuir a porcentaje de alfabetismo.
- La tasa se puede definir como:

“Es la consideración de un período de referencia, o de observación, para un hecho demográfico. Las tasas indican la frecuencia relativa de un fenómeno en un período dado, generalmente un año”.

Gómez, M.



Tasas

- Un ejemplo: la tasa de natalidad.

$$\text{Tasa bruta de natalidad} = \frac{\text{Nacimientos vivos ocurridos durante el año } Z}{\text{Población total a mitad del año natural } Z} * 1000$$

- Para este ejemplo, la tasa indica cuántos nacimientos, por cada 1000 habitantes, se produjeron durante el natural considerado (Z), en **un cierta área**.
- Así, por ejemplo, en el 2000 ocurrieron un total de 89965 nacimientos vivos en CR, y la población total para ese año fue de 4 017 986.
- Con esto se puede calcular la tasa bruta de natalidad.

Tasas

- Entonces, de acuerdo a la información anterior:

$$TBN = \frac{89965}{4017986} * 1000 = 22.39\text{‰}$$

- Interpretación: “En el 2000 se dieron en Costa Rica alrededor de 22 nacimientos por cada mil habitantes.
- Nótese que el factor de amplificación que se emplea es de 1000 y que por ello la tasa se expresa por mil habitantes.



Proporción

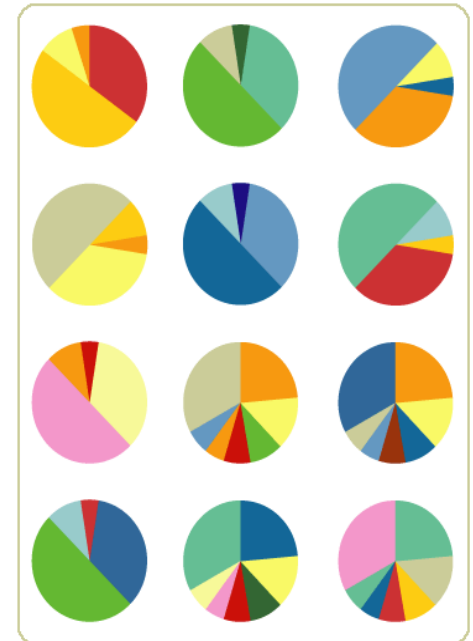
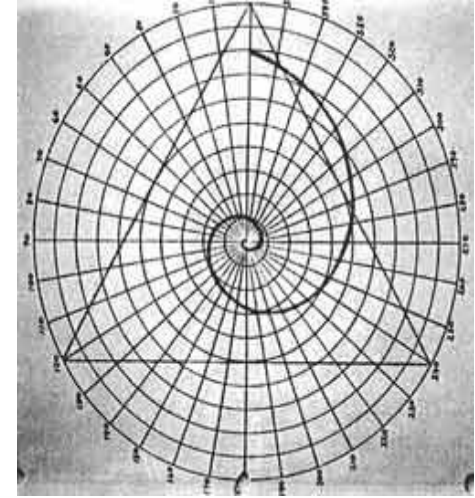
- Se puede considerar como una razón, pero con 2 características especiales:

-Relaciona 2 números del mismo universo.

-Relaciona a una parte con el todo. Por ejemplo, si tenemos los número A, B y C pertenecientes al mismo universo, una proporción sería:

$$\frac{B}{A + B + C}$$

- Como puede apreciarse, “B” numerador forma parte del total de los elementos (A,B y C). La proporción nos indica qué parte o fracción del total “A+B+C” representa B.
- Una proporción siempre va a estar entre los valores de 0 y 1.



Ejemplo de razón y proporción

- Supongamos que se desea investigar cierta población expuesta a un tratamiento para mejorar su salud. Se cuenta con un total de 2000 personas, de los cuales 1200 son mujeres y 800 son hombres. Además 700 provienen de San José, 600 de Alajuela, 400 de Heredia y 300 de Guanacaste.
- Obtenga la razón de mujeres entre hombres y la proporción de personas que provienen de Alajuela.

Razón



Porcentaje



Ejemplo de razón y proporción: interpretación

- De acuerdo a los resultados anteriores, la interpretación es la siguiente.

Razón: se tiene un total de 3 mujeres por cada 2 hombres.

Proporción: 0.3 o el 30% de las personas provienen de Alajuela.

Nota: para obtener un ***porcentaje***, lo que hacemos es multiplicar el resultado de la proporción por 100, y así obtenemos el ***porcentaje***.



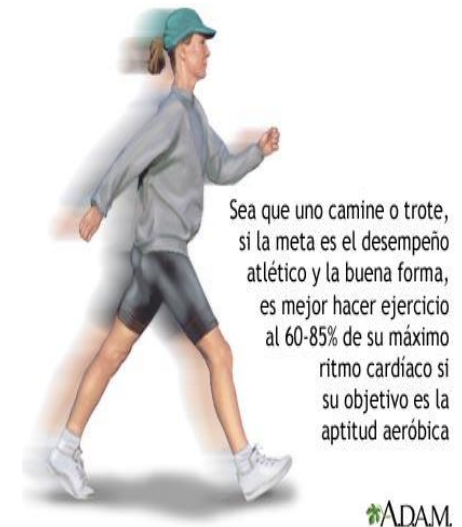
Distribución de Frecuencias

- La distribución de frecuencias es otra forma de resumir la información.
- Cuando los datos provienen de la operación de contar o de medir, y pueden haberse obtenido anotando el número de elementos que corresponde a una de las categorías definidas, entonces la utilización de las distribuciones de frecuencia es más recomendable.
- La DF son clasificaciones que se refieren a variables cualitativas o cuantitativas y constituyen un instrumento muy útil en el trabajo estadístico.



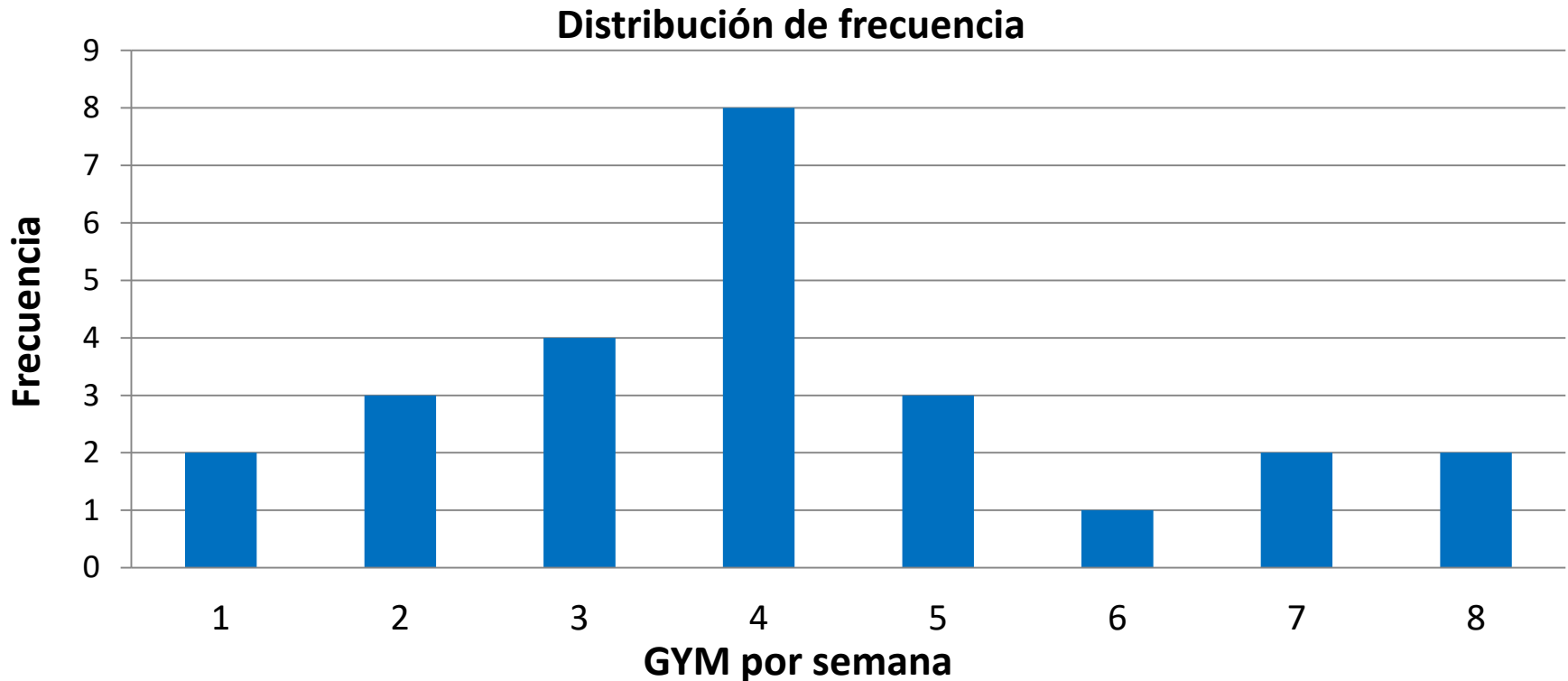
Distribución de frecuencias

- En los datos estadísticos resulta muy valioso disponer de elementos descriptivos que den información acerca de 3 aspectos:
 - La forma
 - La posición (alrededor de qué)
 - La dispersión
- Cuando los datos son numerosos, para conocer lo anterior se debe recurrir a agrupar estos en una distribución de frecuencias.
- La DF se define como una ordenación o arreglo de datos en clases o categorías que muestran, para cada una de las categorías, el número de elementos que contiene o frecuencia.



Distribución de frecuencias

- La distribución de frecuencia puede presentarse por medio de un gráfico de barras verticales, como sigue:



- En conclusión, la DF es un excelente método para hacer representaciones de la información cuando esta abunda y se quiere conocer ciertas características.

Indicadores

- Los indicadores son también número relativos, y se suelen utilizar en área académica, pero sobre todo en la toma de decisiones.
- Un indicador es una característica específica, ***observable*** y ***medible*** que puede ser usada para mostrar los cambios y progresos que está haciendo un programa hacia el logro de un resultado específico (en el tiempo).
- Un indicador debe ser definido en términos precisos, no ambiguos, que describan clara y exactamente lo que se está midiendo.



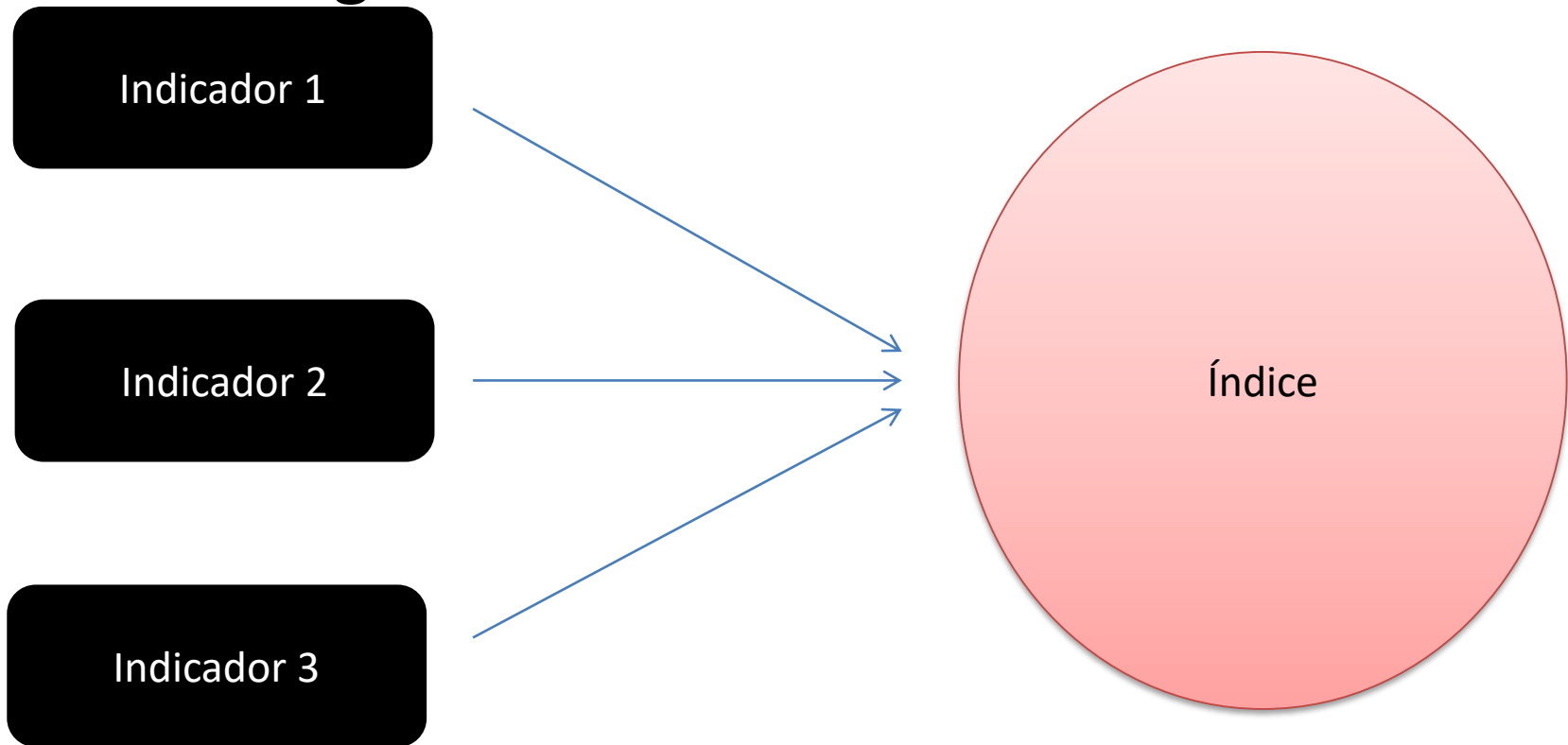
Índices

- Los índices, al igual que la razón, proporción, tasa e indicadores, busca resumir la información en un número relativo para facilitar la información de un acontecimiento o algo por el estilo.
- Los índices, por lo general, son los números relativos más complejos y difíciles de construir. Puede ser tan simple como el IMC, o tan complejo como el IDH.
- Por lo general, un índice se construye a partir de indicadores. De igual forma, hay índice que se obtienen a partir de otros índices.



Índices

- Esquema ***general*** para construir un índice es el siguiente:



Representación de la información

Presentación de la información

En un principio clásico se tienen 4 formas de presentar la información:

- a. Textual
- b. Semitabular
- c. Tabular
- d. Gráfica

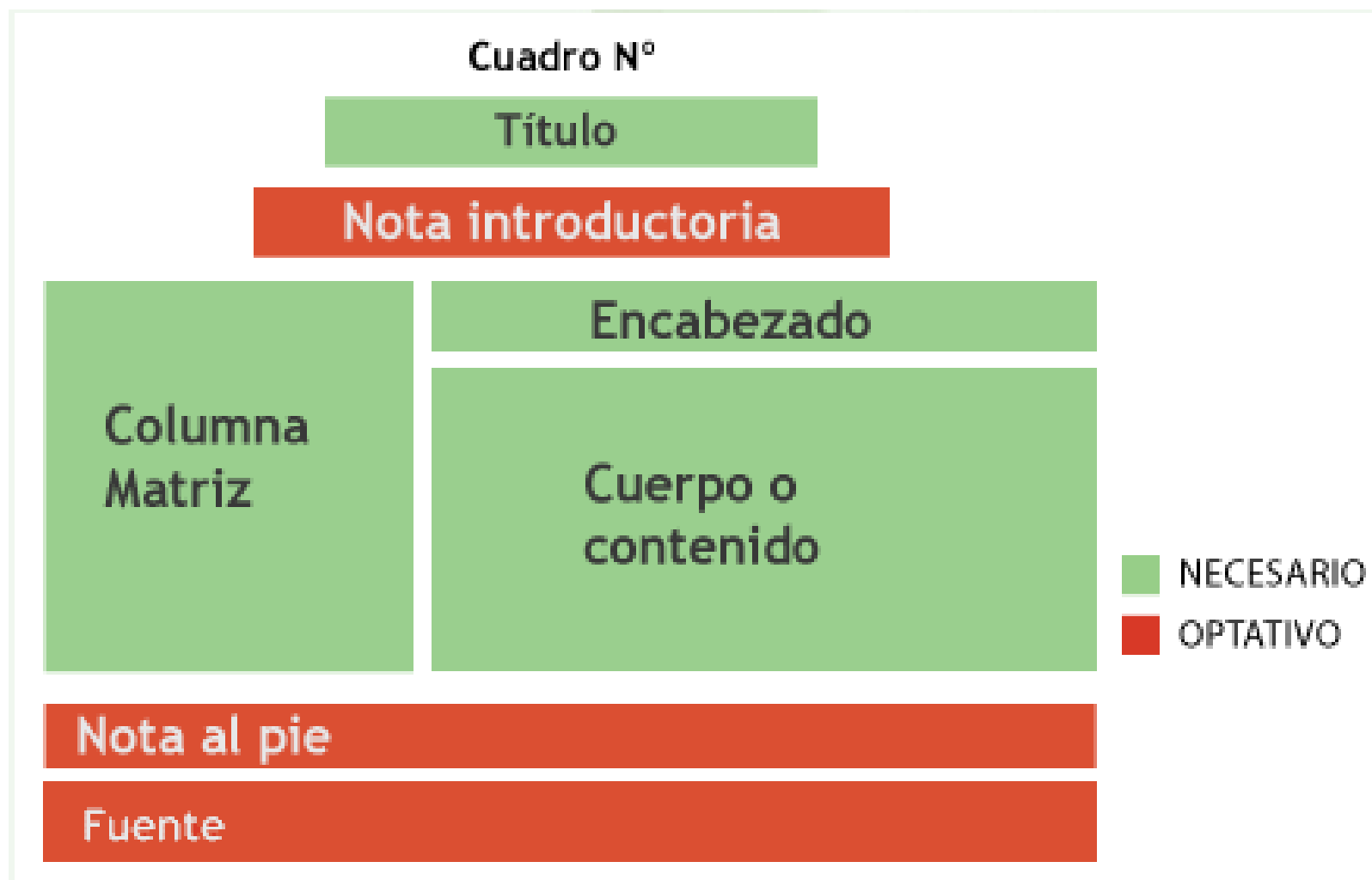


Componentes del CUADRO

- El “cuadro” en estadística está compuesto de los siguiente:
 - Título
 - Nota preliminar (nota introductoria)
 - Columna matriz
 - Encabezados o encabezamientos
 - Cuerpo o contenido
 - Nota al pie
 - Fuente
- No todos los componentes anteriores deben siempre estar en un cuadro.



Componentes de un cuadro

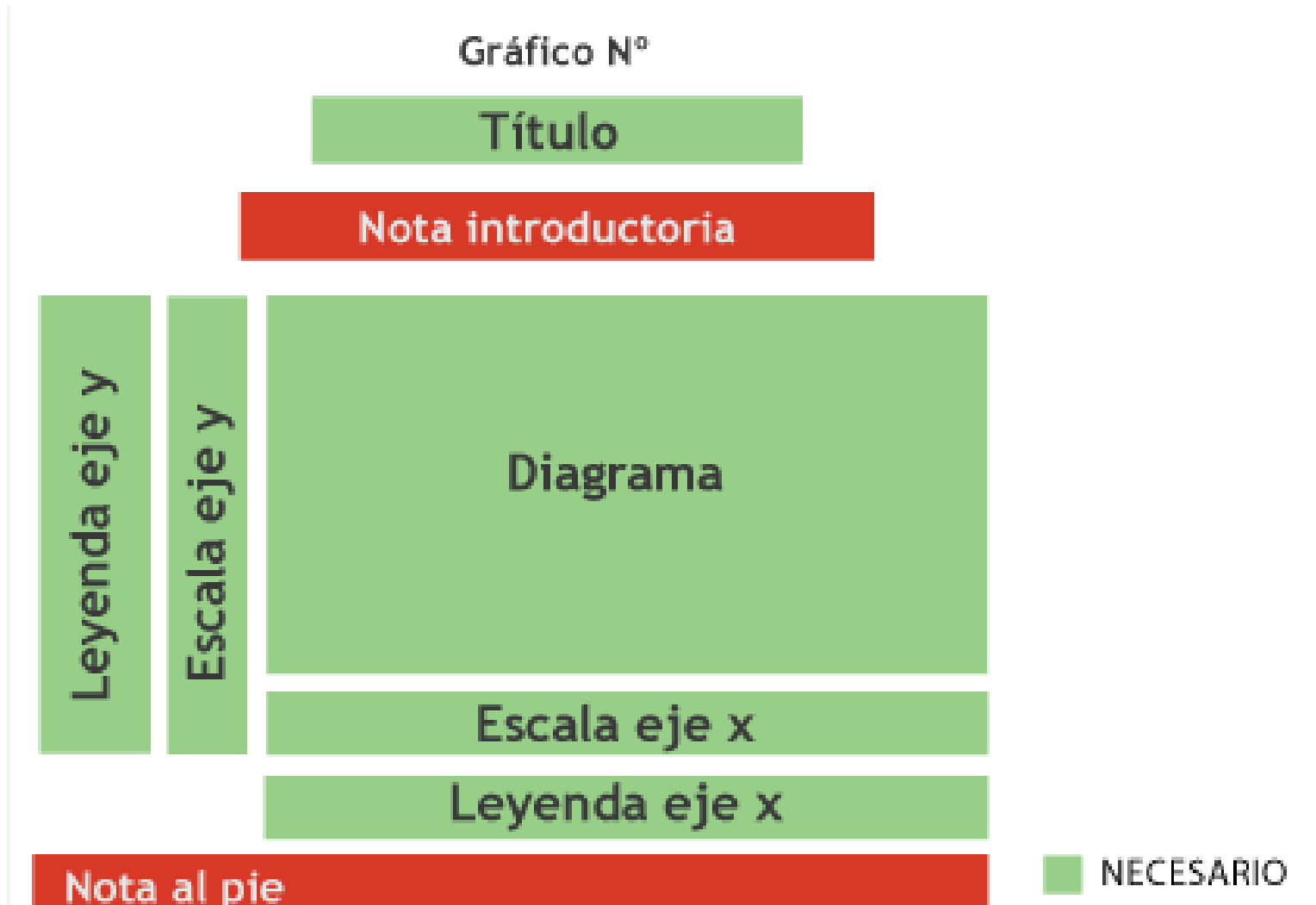


Componentes del GRÁFICO

- El “gráfico” está compuesto
 - Título
 - Nota introductoria
 - Leyenda eje “y”.
 - Escala eje “y”
 - Leyenda eje “x”
 - Escala eje “x”
 - Diagrama
 - Nota al pie
 - Fuente*



Componentes del GRÁFICO

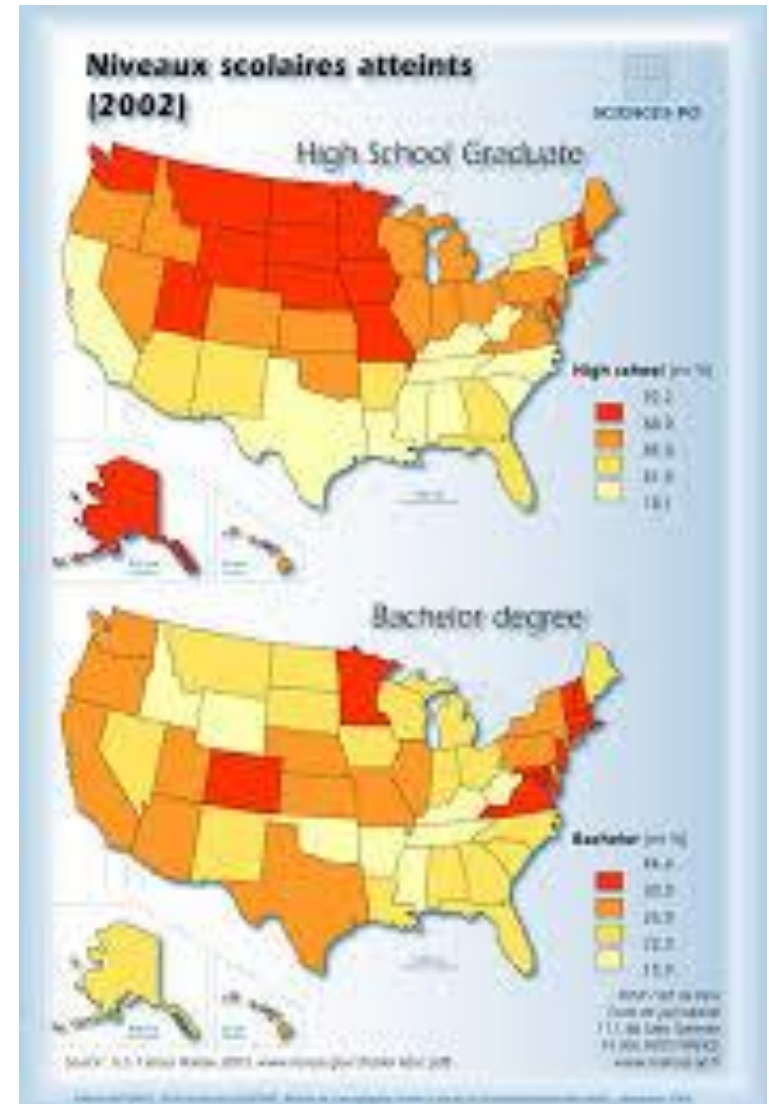
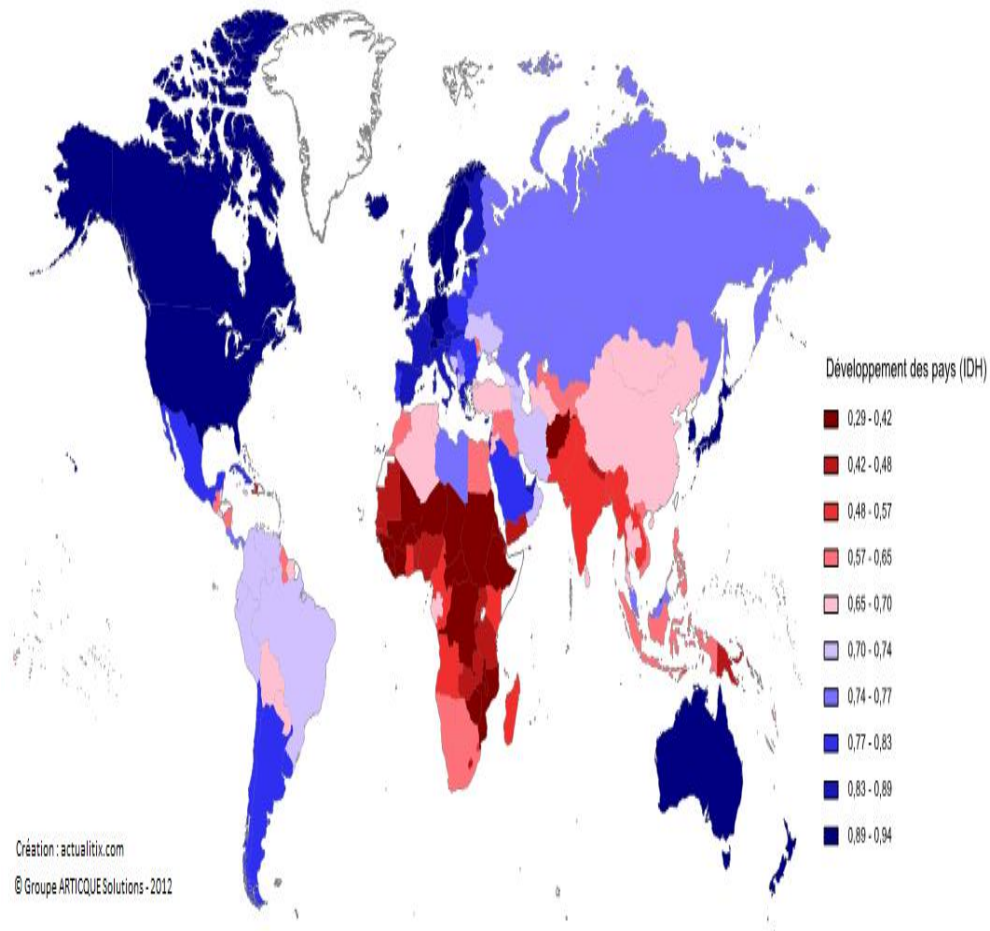


Otras representaciones gráficas

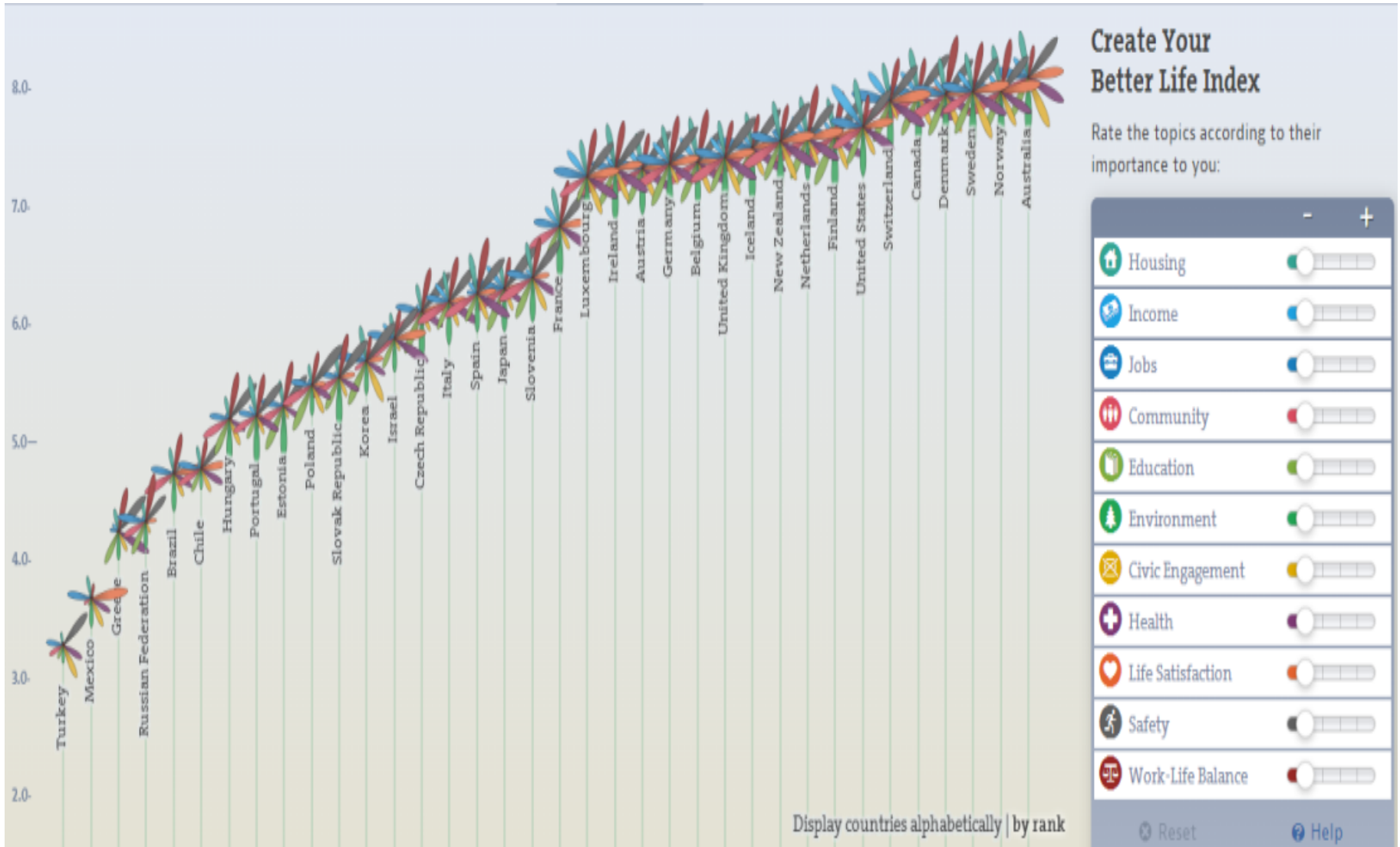
- Además de los cuadros y de los gráficos, existen otros métodos de presentación de la información.
- Los mapas es otra forma de presentar información cuando esta provienen de zonas geográficas (continentes, países, estados, provincias, etc.).
- El sistema de indicadores es otra forma de presentar la información cuando esta abunda en cantidad.
- Nube de palabras.



Otras representaciones gráficas



Otras representaciones gráficas



El Dashboard



Nube de palabras



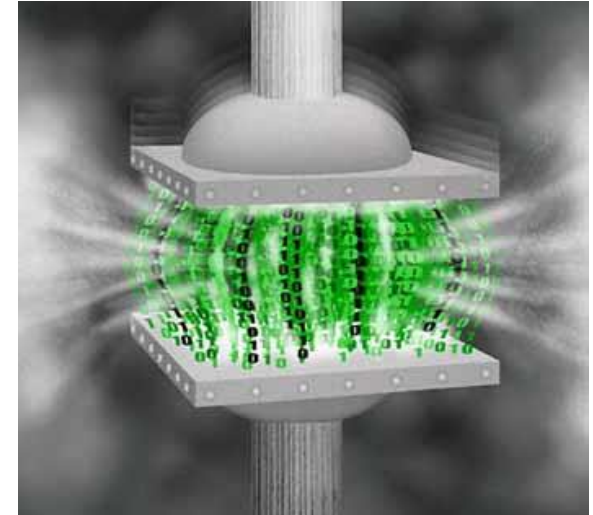
Nube de palabras



Medidas de posición

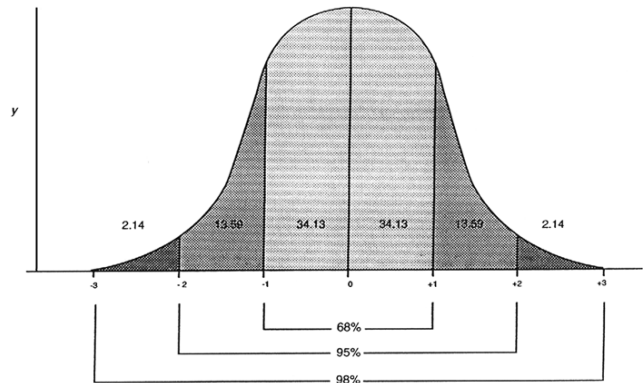
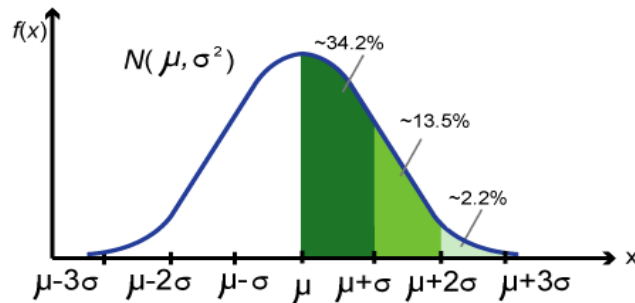
Introducción

- 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 propósito 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).



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:

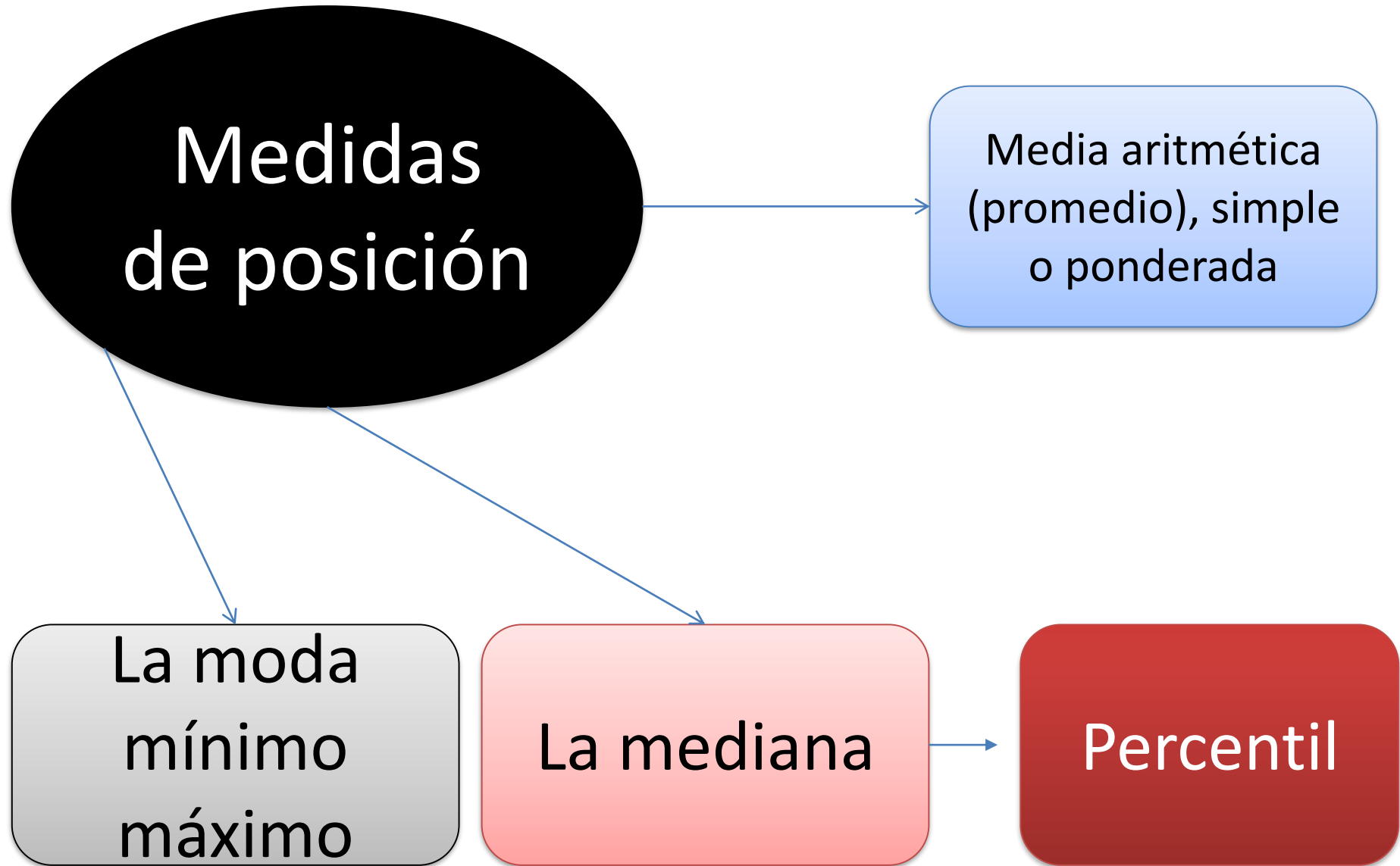


- a. Moda
- b. Mediana
- c. Media aritmética o promedio
- d. Media geométrica
- e. Media armónica
- f. Cuantiles - percentiles



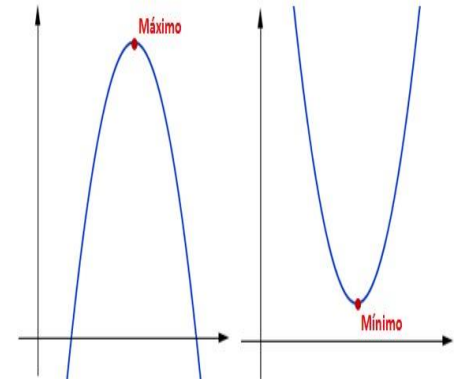
No vamos a
verlas

Las medidas de posición



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:

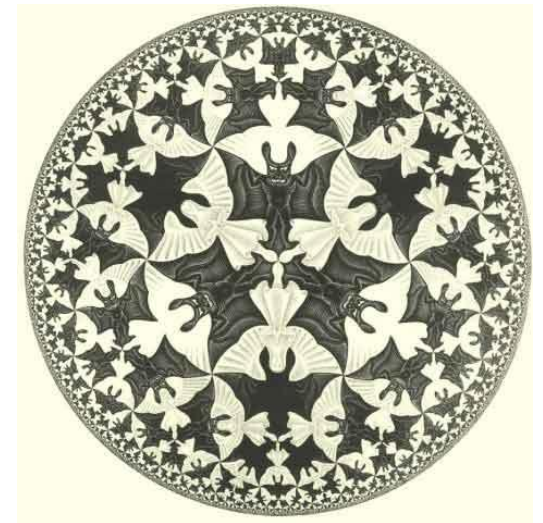
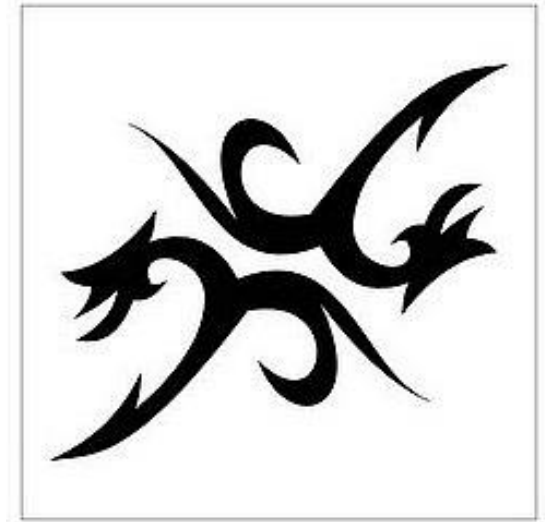
14, 15, 17, 17, 21, 21, 21, 21, 33, 36, 40

- ¿Cuál sería la moda en este caso?



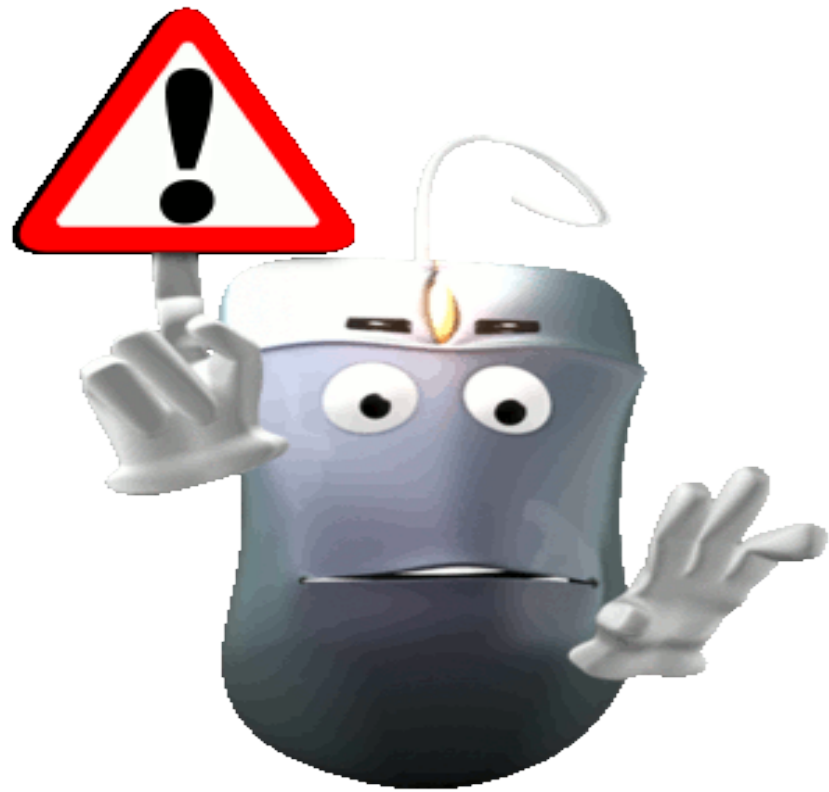
La mediana

- 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.



¡Atención! ¡atención!

por JEAN-MICHEL FRODON



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

La mediana: conjunto impar

Datos impar

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

La mediana se encuentra en el siguiente término:

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

Ahora, la media sería el valor del 4^{to} término, lo que ahora si daría valor de la mediana:

$$\text{Mediana} = 10$$

La mediana: conjunto par

Datos par

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

La mediana se encuentra en el siguiente término:

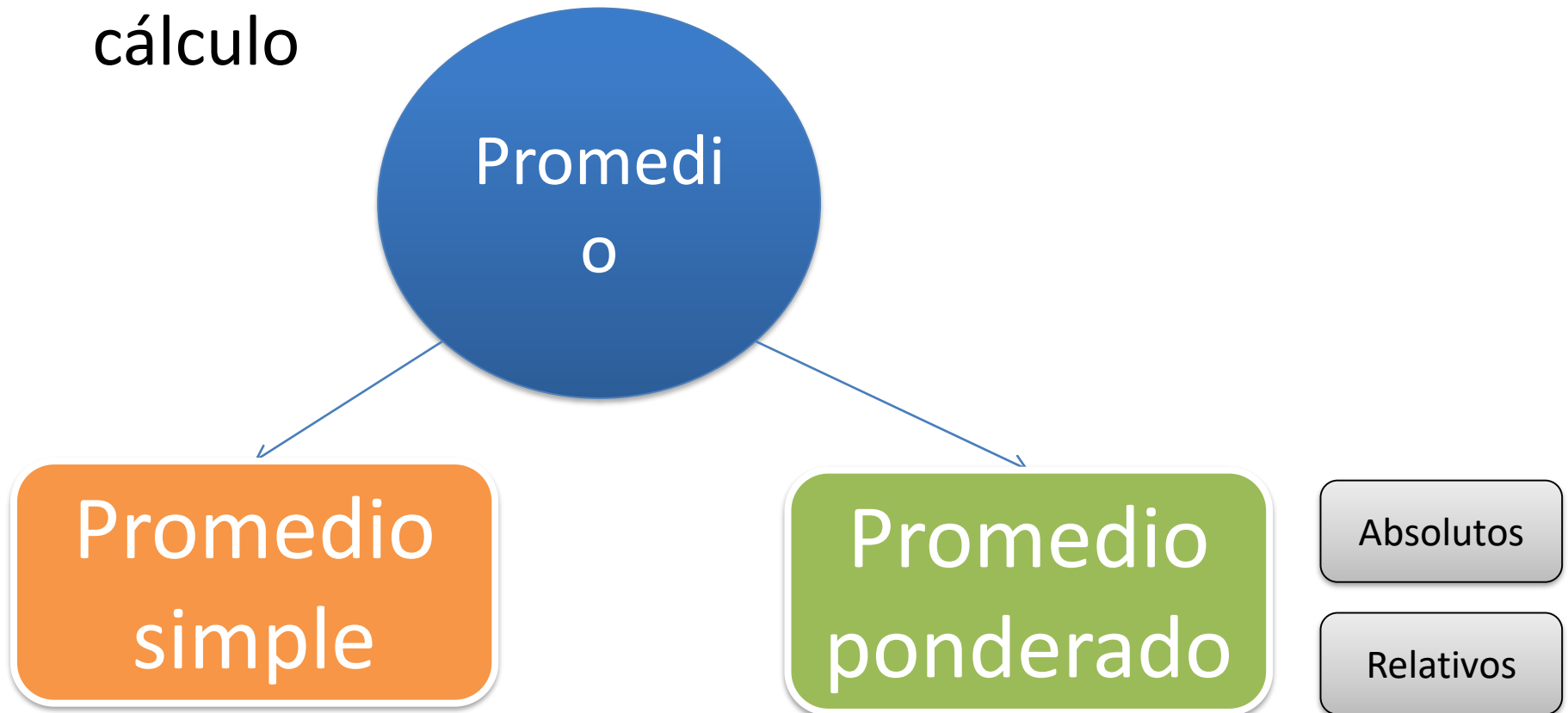
$$n = 8: \quad \frac{n + 1}{2} = \frac{8 + 1}{2} = 4,5 \text{ término}$$

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.

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

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.

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

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

- 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:

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

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



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

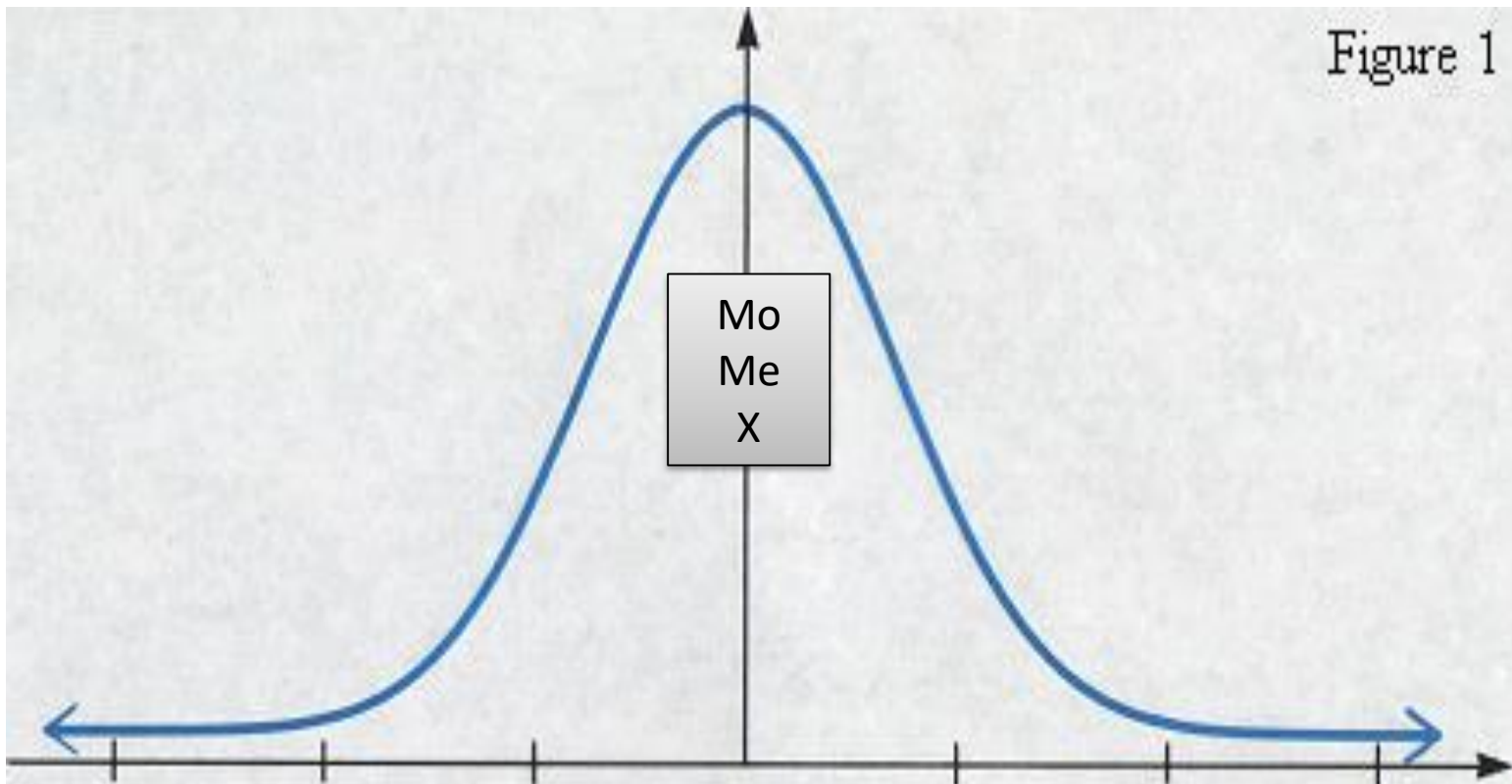
Usos de las medidas de posición

- 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

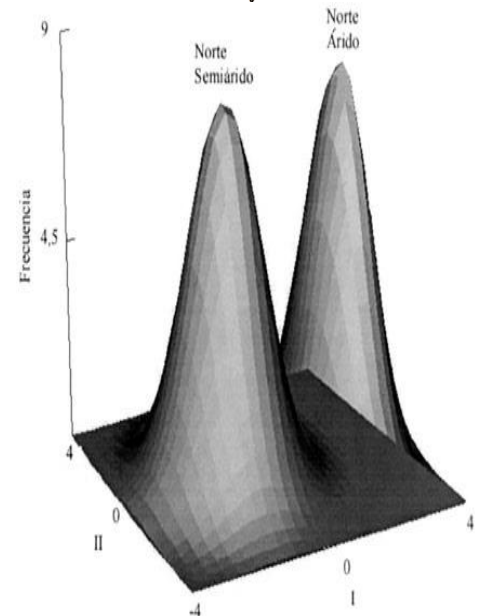
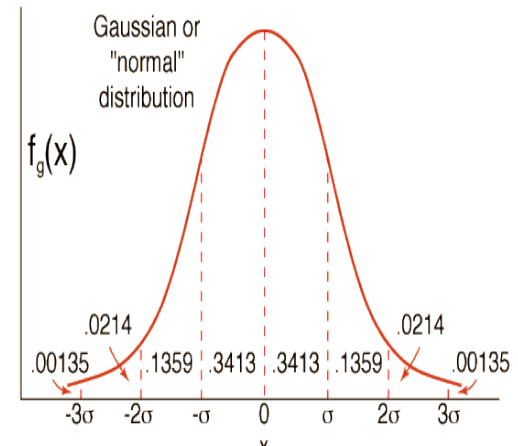
- Sea la siguiente distribución de datos:



Distribución simétrica

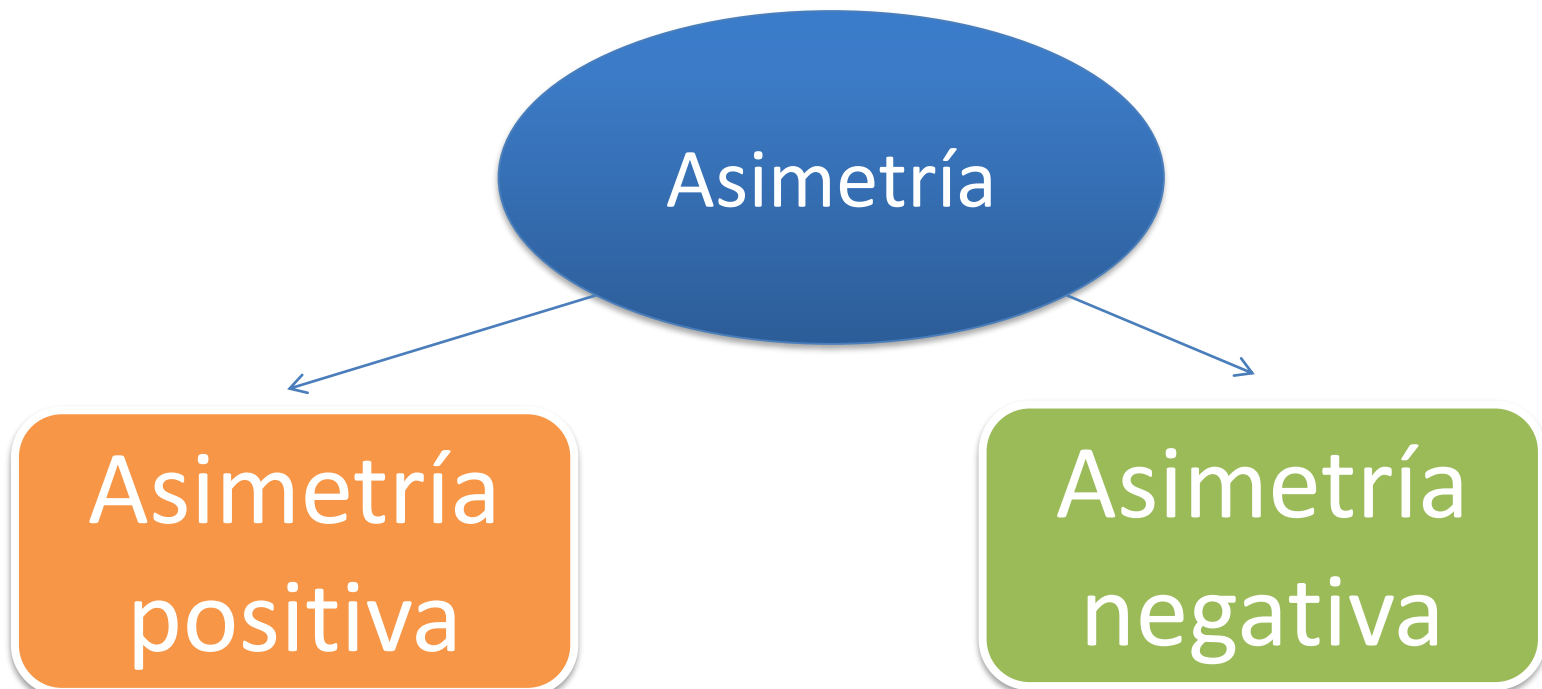
Usos de las medidas de posición

- Cuando los datos se distribuyen normalmente, la Mo , Me , y Po coinciden (valen lo mismo). En este caso, cualquier es bueno para caracterizar el grupo de datos, pero igual se prefiere al Po .
- De igual forma, si las 3 medidas no coinciden mucho, pero no hay un alejamiento extremo, se escoge el Po por tener mejores propiedades estadísticas.
- Casi siempre, el Po es la medida de preferencia para caracterizar un conjunto de datos.



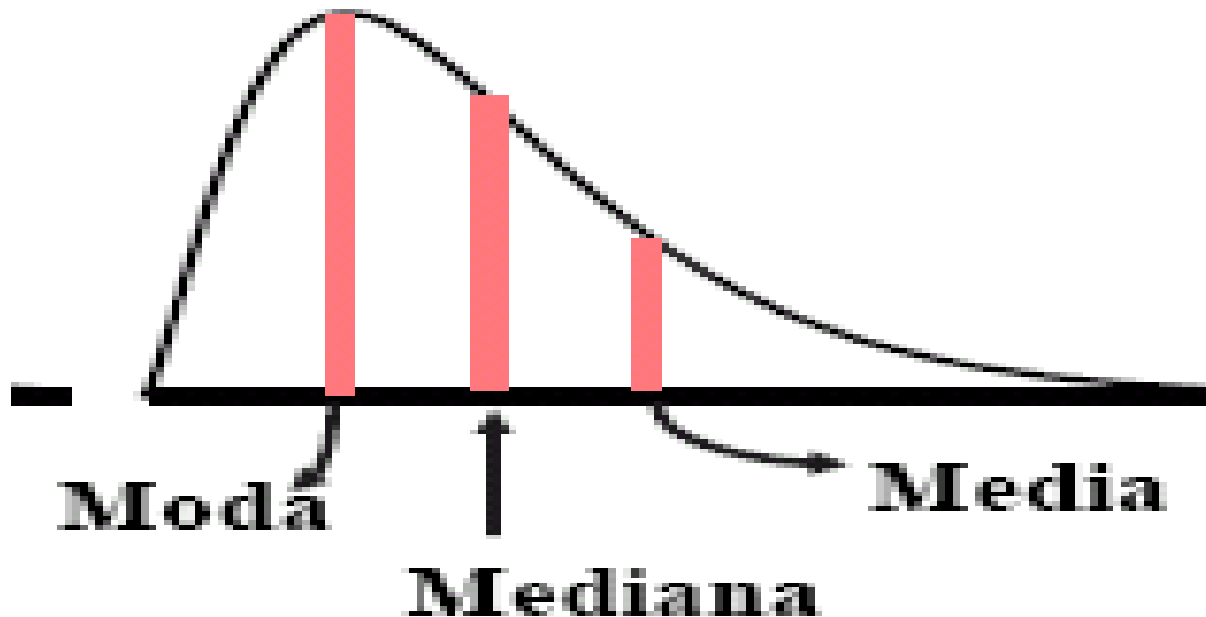
Usos de las medidas de posición

- 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

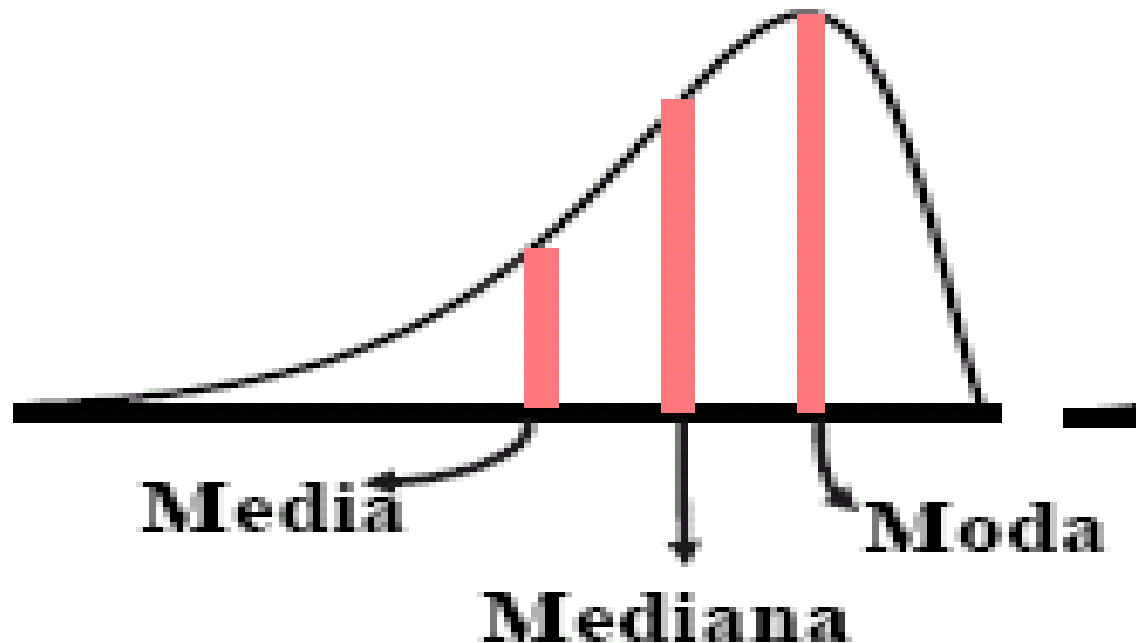
Asimetría positiva



Asimétrica hacia
la derecha

Usos de las medidas de posición

Asimetría negativa



Asimétrica hacia
la izquierda

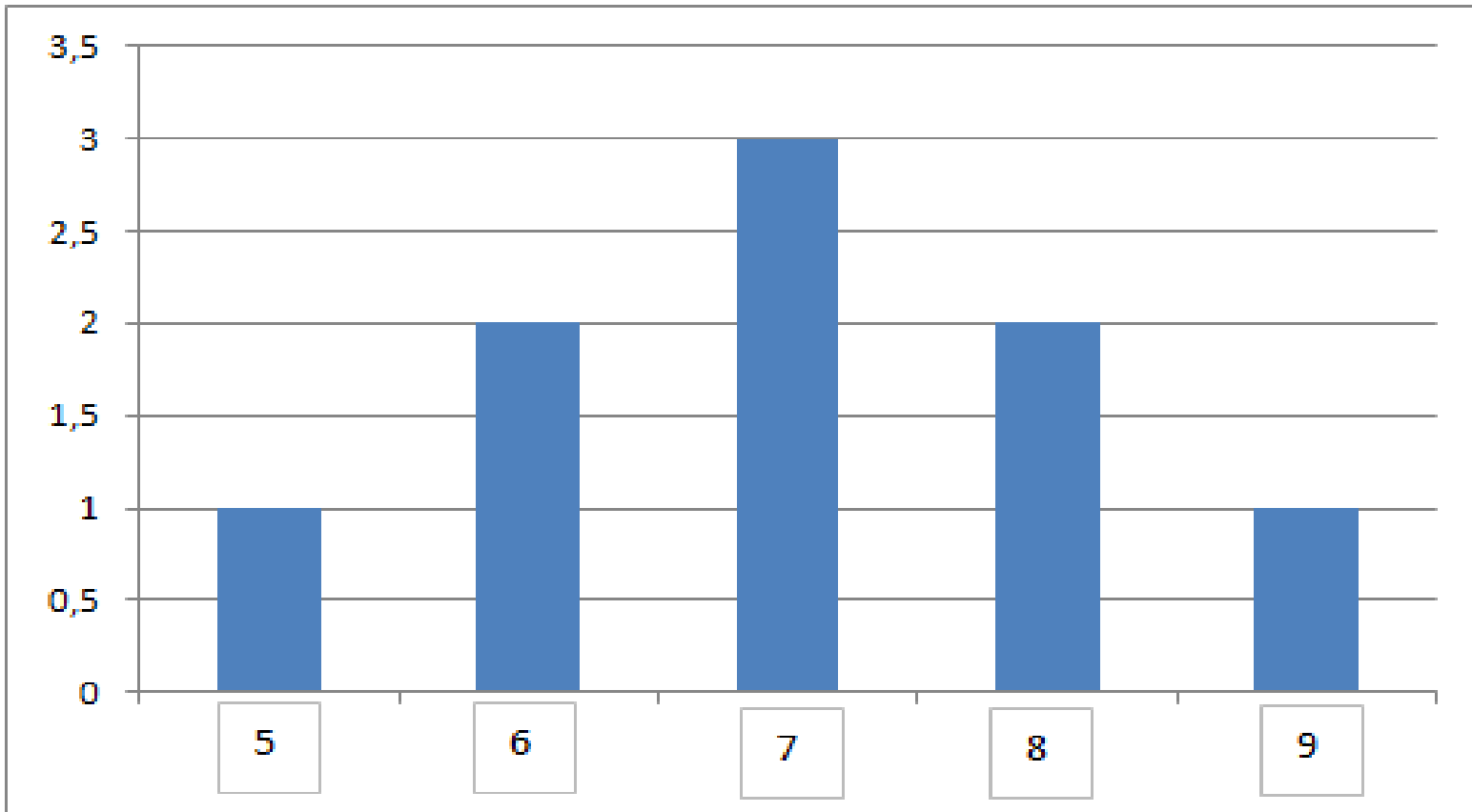
Usos de las medidas de posición

- 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

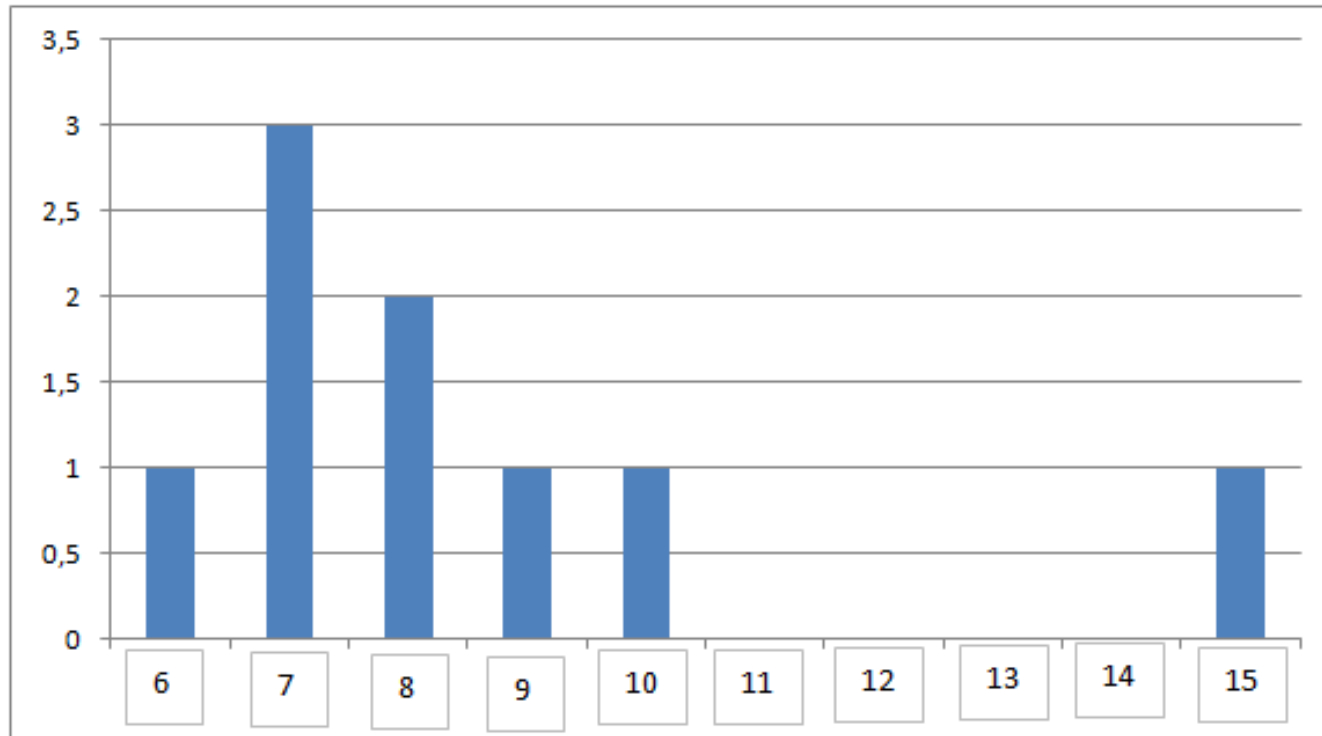
Ejemplo 1 *Datos:* 5,6,6,7,7,7,8,8,9



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

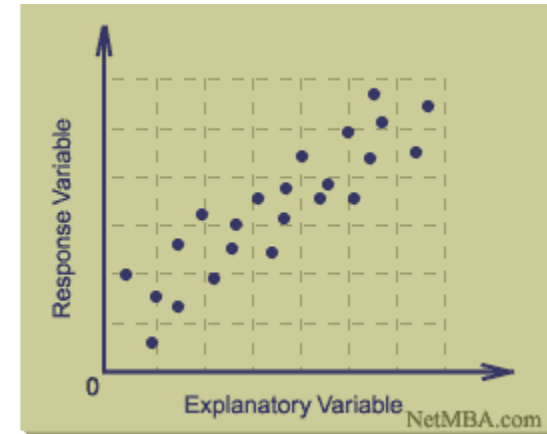


En este caso se prefiere la Mo y la Me para describir el conjunto que el

Medidas de variabilidad

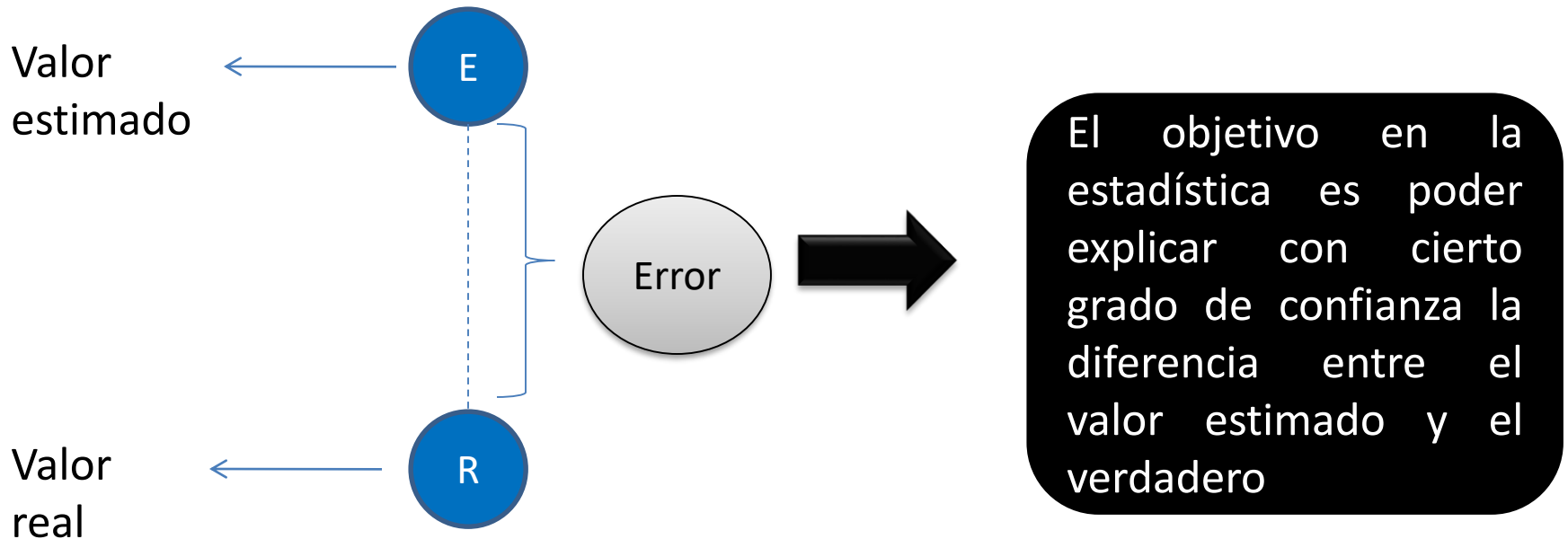
Introducción

- El concepto de variabilidad juega un papel clave dentro de la Estadística como en cualquier ciencia.
- La razón de ser de la estadística reside en que los datos son variables de mayor a menor intensidad.
- La importancia de la Estadística es suministrar procedimientos válidos y confiables para analizar los hechos variables.



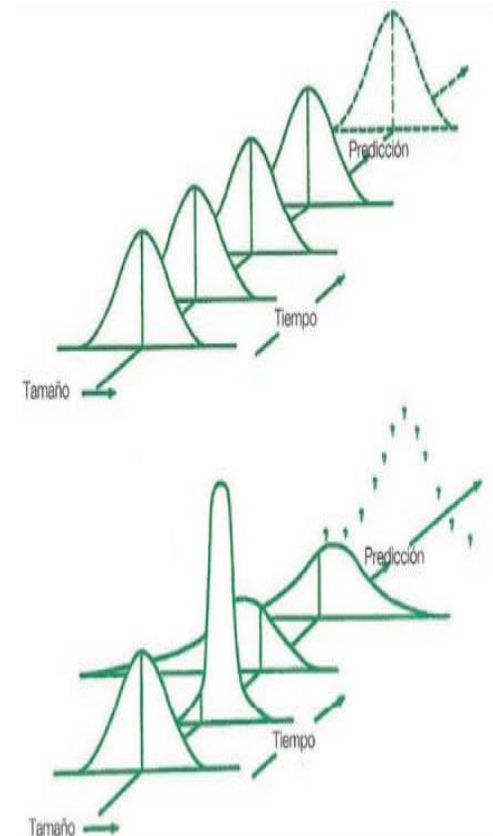
Introducción

- El objetivo está en poder explicar y mediar la variabilidad con tal de realizar inferencias fidedignas.



El concepto de variabilidad

- Cuando se analizan los datos, se debe tener en mente 2 objetivos:
 1. Se quiere conocer el fenómeno en cuestión, y la mayoría de veces se utilizan medidas de posición (MDP).
 2. Se utilizan medidas para poder medir la variabilidad en los datos. Por lo tanto se quiere saber tanto la concentración como la dispersión de estos.
- De acuerdo con el segundo objetivo, si existe la presencia de mucha variabilidad, entonces la confianza de los datos estará en juego.



Medidas de variabilidad

- Existen diferentes formas de medir la dispersión o variabilidad de los datos.
- La elección de cierta medida depende de la situación y de las posibles ventajas en relación a las otras medidas.
- Las medidas de variabilidad más corrientes son:
 - El recorrido o amplitud.
 - La desviación media (DM).
 - La variancia y desviación estándar (DE).



Recorrido o amplitud

- La forma más simple de apreciar la variabilidad es considerando los valores extremos del conjunto de datos (el valor menor y el valor mayor).
- La diferencia entre el valor mayor y el menor del conjunto de datos da origen al recorrido o amplitud.

$$\text{Recorrido} = X_{\text{máximo}} - X_{\text{mínimo}}$$

$$\text{Recorrido} = \text{Valor mayor} - \text{valor menor}$$

Recorrido o amplitud

- El cálculo es muy simple. Considérese las siguientes edades:

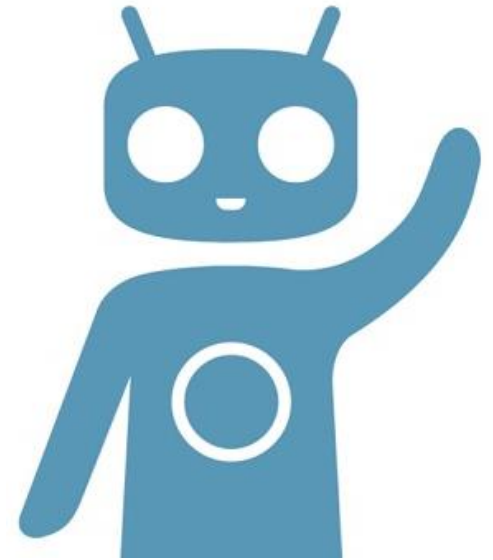
19, 20, 21, 20, 21, 19, 18, 22, 20, 19, 23.

$$\text{Recorrido} = 23 - 18 = 5$$

- En este caso se tiene que las edades tienen una dispersión o amplitud de 5 años. Dependiendo del contexto, puede ser mucho o poco.

La desviación media

- Dadas las limitaciones del recorrido, es necesario definir una medida de dispersión que tome en cuenta en el cálculo todos los datos.
- La medida tiene que estar basada en las desviaciones o diferencias de los datos individuales respecto a un valor central o típico (como el promedio).
- La mejor opción es considerar la suma de las desviaciones de los datos con respecto al promedio.



La desviación media



- Sin embargo, si esto se hiciera de buenas a primeras, esta opción siempre arrojaría valores iguales a “0”.
- Para obviar este problema, se suele usar los valores absolutos de las diferencias, y dividirlos por el número de datos para obtener una medida de dispersión promedio o por observación.
- Así se origina la desviación media:

$$DM = \frac{\sum |X_i - \bar{X}|}{n} = \frac{\text{desviación absoluta}}{\text{número de datos}}$$

La desviación media

- Sea el siguiente conjunto de datos

3, 10, 2, 8 y 7.

$$\bar{x} = 6$$

El promedio:



- La DM se expresa:

x_i	$x_i - \bar{x}$	$ x_i - \bar{x} $
3	-3	3
10	4	4
2	-4	4
8	2	2
7	1	1
30	0	14

La desviación media

- Debido principalmente a la falta de requerimientos teóricos y muchas veces prácticos, la desviación media no es la medida de dispersión más utilizada.
- Por eso se debe recurrir la gran mayoría de veces a la variancia y desviación estándar.

σ^2 and σ

Variancia y desviación estándar

- La variancia y la desviación estándar son las medidas de dispersión más cómodas y útiles que reúnen numerosas ventajas prácticas y teóricas.
- Como para la DM, se toma en cuenta las desviaciones o diferencias de todos los datos con respecto al promedio.
- Sin embargo, en esta ocasión en vez de considerar las desviaciones absolutas de cada dato con respecto al promedio, se van a considerar las desviaciones cuadráticas de los datos con respecto al promedio.

$$\sqrt{\frac{\sum(X - \bar{X})^2}{(n - 1)}}$$

where:

X = each score

\bar{X} = the mean or average

n = the number of values

Σ means we sum across the values

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$$

Variancia y desviación estándar

- Las desviaciones cuadráticas permiten obtener la medida de dispersión denominada “variancia”. Si a la variancia se le aplica una raíz cuadrada, entonces se puede obtener la desviación estándar o desviación típica.
- La variancia se formula de la siguiente forma:

$$\text{Variancia} = s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} = \frac{\text{desviación cuadrática}}{\text{número de datos}}$$

Variancia y desviación estándar

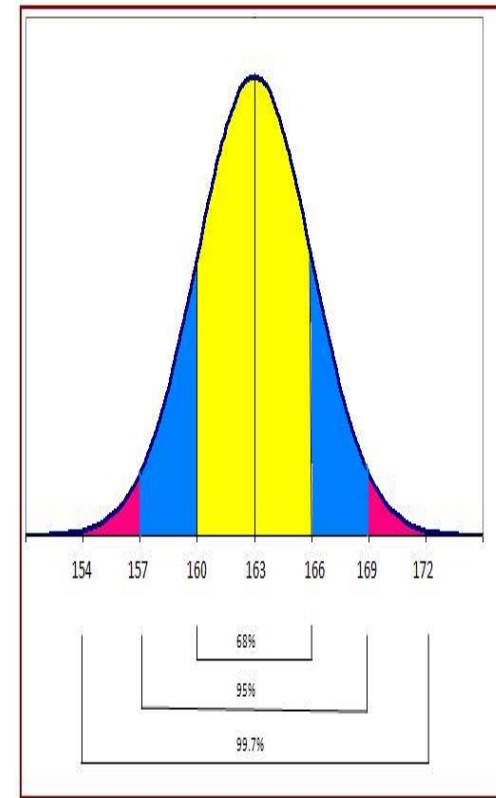
- La desviación estándar (DE) se describe de la siguiente forma:

$$DE = \sqrt{s^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}} = \sqrt{\frac{\text{desviación cuadrática}}{\text{número de datos}}}$$

- La desviación estándar nos indica cuánto se alejan, en promedio, las observaciones de la media aritmética del conjunto.
- Es la medida de dispersión más usada en estadística, tanto en aspectos descriptivos como analíticos.

Variancia y desviación estándar

- Para fines prácticos, la medida de dispersión que se utiliza en informes, reportes, descripciones y se interpreta es la desviación estándar.
- La variancia se utiliza para facilitar los cálculos, pero no se interpreta ni nada por el estilo, ya que las medidas están modificadas por el cuadro, e interesa la medida original de la variable.
- De ahí que si interpretamos algo será la DE.



Variancia y desviación estándar

- Para el cálculo de la variancia se puede proceder de dos formas. La primera es la forma “normal” y la otra es la “simplificada”.
- El cálculo de la variancia “normal”, cuando se tiene una muestra de “n” datos, se expresa como “ S^2 ”, se denota:

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

Variancia y desviación estándar

- Sea el conjunto de datos anterior,

3, 10, 2, 8, 7.

- Para obtener la variancia:

xi	xi - x	(xi-x)^2
3	-3	9
10	4	16
2	-4	16
8	2	4
7	1	1
30	0	46

$$\bar{x} = 6$$

$$s^2 = \frac{\sum (X_i - \bar{X})^2}{(n - 1)} = \frac{46}{4} = 11.5$$

Variancia y desviación estándar

- Por lo tanto la variancia es igual a 11.5.
- Para obtener la desviación estándar lo que tenemos que hacer es sacar raíz cuadrada al resultado anterior.

$$\sqrt{s^2} = s = \sqrt{11.5} = 3.39$$

- Nótese que este es el resultado que se interpreta, pero esto lo haremos más adelante.



The End

