

Introducción a la estadística

PROFESORA MAIRA CÁCERES



Estadística

Definición: Es la ciencia de recolectar, describir e interpretar datos, como un apoyo eficiente para la toma de decisiones. (Definición ingenieril)

En este sentido la estadística es aplicable a cualquier campo de estudio.

Los métodos estadísticos pueden clasificarse en dos grandes grupos:

- **Descriptivos:** Se preocupan de describir el conjunto de datos. Generalmente están relacionados con el análisis preliminar o exploratorio de los datos.
- **Inferenciales:** Se preocupan de la obtención de conclusiones sobre la población de donde se ha extraído una muestra.

Ejemplos de problemas estadísticos



Línea del razonamiento estadístico

- **Recolección de los datos.** (encuesta, cuestionario, entrevista, observación, diagrama de flujo)
- **Descripción estadística de los datos.** (variables estadísticas)
- **Análisis estadístico de los datos.** (parámetros estadísticos)
- **Decisión o predicción.**

Estadística descriptiva

PROFESORA MAIRA CÁCERES

Estadística descriptiva

La estadística descriptiva es la técnica matemática que **obtiene, organiza, presenta y describe un conjunto de datos** con el propósito de facilitar el uso, generalmente con el apoyo de tablas, medidas numéricas o gráficas.

CONCEPTOS BÁSICOS

UNIDAD EXPERIMENTAL: es el ente que proporciona una información. También recibe el nombre de unidad de análisis.

POBLACIÓN: es el universo o conjunto total de unidades experimentales. Sobre este conjunto se obtendrán las conclusiones finales.

MUESTRA: es un subconjunto de casos o individuos de una población.

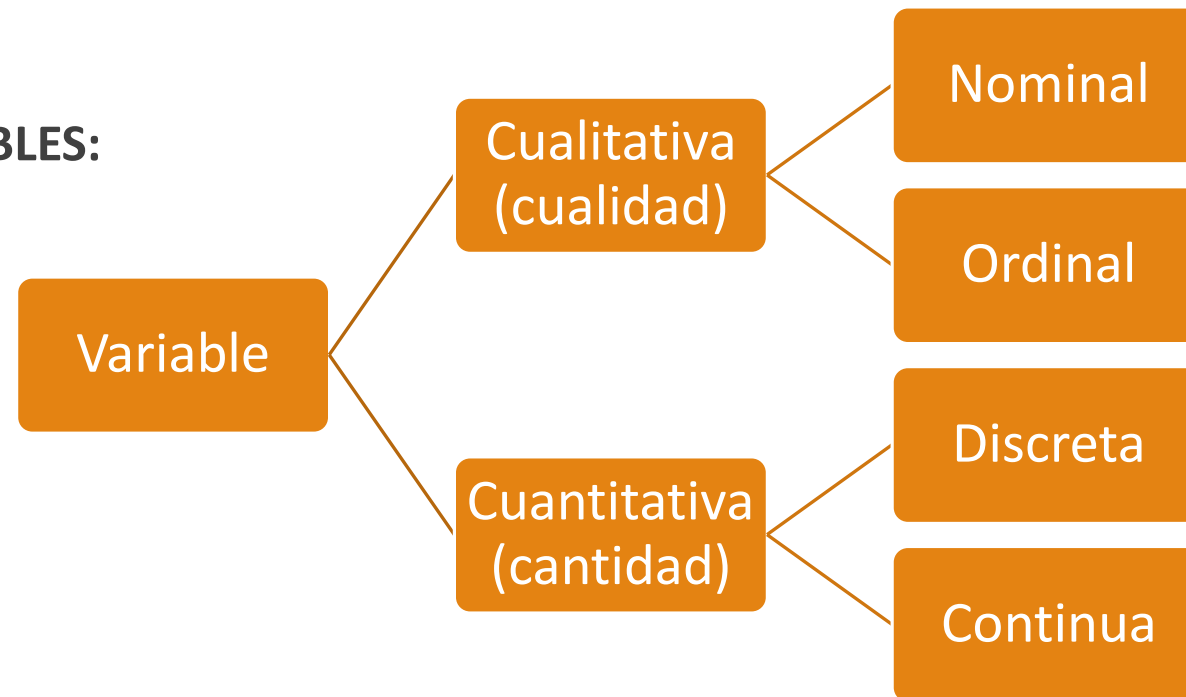
VARIABLES: son características relacionada con la población y que es de interés para el investigador, se extrae de la muestra.

PARAMETRO: es un número que resume la gran cantidad de datos que pueden derivarse del estudio de una variable estadística.

Variable estadística

DEFINICIÓN Es cualquier característica o atributo que es deseable conocer acerca de las unidades experimentales y que se espera varíe de una unidad a otra. El conjunto de todos los posibles valores de ésta se denomina recorrido de la variable.

TIPOS DE VARIABLES:



Variables cuantitativas

La variable cualitativa es aquel tipo de variable estadística la cual describe cualidades, circunstancias o características de algún tipo de objeto o persona sin hacer uso de números.

Tipos de variables:

- **Variable Nominal:** En este caso la variable no es representada por números ni tampoco tiene algún tipo de orden, así que **es menos precisa** en lo que matemáticamente se refiere.
 - Por ejemplo: los colores: negro, azul, rojo, amarillo, naranja, etc.
- **Variable Ordinal:** también conocida como variable cuasicuantitativa también es representada por una modalidad que **no requiere números pero si existe un orden o un puesto**.
 - Por ejemplo: nivel socioeconómico: alto, medio bajo.

Variable cuantitativa

Las variables cuantitativas son aquellas variables estadísticas que otorgan un resultado representado por un valor numérico.

- Por ejemplo, las variables cuantitativas pueden ser aquellas tales como peso (62 Kg, 80 Kg, etc.), altura (1.72 cm, 1.85 cm, etc.) o cantidad de miembros en una familia (2, 3 ó 4).

Tipos de variables:

- **Variable discreta:** Estas hacen referencia a aquellas variables que solo puede adquirir un valor de un conjunto de números exactos.
 - Por ejemplo: Una persona puede tener 1, 2, 3 ó más perros, nunca un perro y medio.
- **Variable continua:** la variable puede cobrar un valor con cualquier intervalo o medición, es decir, que **pueden haber otros valores en medio de dos valores exactos**, generalmente representado por valores decimales, haciendo la cifra mucho más específica que en el caso de la discreta.
 - Por ejemplo: Estatura de una persona: 1 metro y 75 centímetros.

Actividad:

- Determina en cada uno de los ejemplos si es una variable cuantitativa o cualitativa y a que tipo pertenece.

Estado civil		Sexo		Temperatura de una ciudad	
Altura de una persona		Cantidad de mascotas que posee una persona		Tes de piel	
Color de ojos		Calificación exacta de un examen		Marca del celular que usa	
Cantidad de miembros de una familia		Profesión		Grados de alcohol de una cerveza	
Comida preferida		Valor económico de un producto		Nivel de satisfacción	

Organización de datos

La **tabla de frecuencias** (o distribución de frecuencias) es una tabla que muestra la distribución de los datos mediante sus frecuencias. Se utiliza para variables cuantitativas o cualitativas ordinales.

La tabla de frecuencias es una herramienta que permite ordenar los datos de manera que se presenten numéricamente las características de la distribución de un conjunto de datos o muestra.

X_i	Frecuencia absoluta (n_i)	Frecuencia absoluta acumulada (N_i)	Frecuencia relativa ($f_i = n_i/N$)	Frecuencia relativa acumulada ($F_i = N_i/N$)
1	7	7	0,06	0,06
2	19	26	0,15	0,21
3	25	51	0,20	0,41
4	12	63	0,10	0,50
5	23	86	0,18	0,69
6	15	101	0,12	0,81
7	8	109	0,06	0,87
8	16	125	0,13	1,00
Total	125	125	1	1

Tabla de distribución de frecuencias

Construcción de la tabla de frecuencias

1. En la primera columna se ordenan de menor a mayor los diferentes valores que tiene la variable en el conjunto de datos.
2. En las siguientes columnas (segunda y tercera) se ponen las frecuencias absolutas y las frecuencias absolutas acumuladas.
3. Las columnas cuarta y quinta contienen las frecuencias relativas y las frecuencias relativas acumuladas.
4. Adicionalmente (opcional) se pueden incluir dos columnas (sexta y séptima), representando la frecuencia relativa y la frecuencia relativa acumulada como tanto por cien. Estos porcentajes se obtienen multiplicando las dos frecuencias por cien.

Tipos de frecuencias

FRECUANCIA ABSOLUTA (n_i) : La frecuencia absoluta de un valor, es el número de veces que el valor x_i está en el conjunto $(x_1, x_2, x_3 \dots x_n)$.

- La suma de las frecuencias absolutas de todos los elementos diferentes del conjunto debe ser el número total de sujetos N . Si el conjunto tiene k números (o categorías) diferentes, entonces:

$$\sum_{i=1}^k n_i = n_1 + n_2 + \dots + n_k = N$$

FRECUANCIA ABSOLUTA ACUMULADA (N_i) : La frecuencia absoluta acumulada de un valor x_i del conjunto $(x_1, x_2, x_3 \dots x_n)$ es la *suma de las frecuencias absolutas de los valores menores o iguales a x_i* , es decir:

$$N_i = n_1 + n_2 + \dots + n_i$$

Tipos de frecuencias

FRECUENCIA RELATIVA (f_i): La frecuencia relativa es la proporción que existe entre la frecuencia de un dato y el total, es decir, es la frecuencia absoluta dividida por el número total de elementos N.

$$f_i = \frac{n_i}{N}$$

siendo (X_1, X_2, \dots, X_N) el conjunto de datos y n_i el total de valores igual a X_i

FRECUENCIA RELATIVA ACUMULADA (F_i): La frecuencia relativa acumulada es la frecuencia absoluta acumulada dividida por el número total de sujetos N.

$$F_i = \frac{N_i}{N}$$

siendo (X_1, X_2, \dots, X_N) el conjunto de datos y N_i el total de valores igual o menor a X_i

Tipos de frecuencias

FRECUANCIA PONCENTUAL (h_i): la frecuencia porcentual se obtiene multiplicando la frecuencia relativa por cien, se obtiene el porcentaje (tanto por cien %).

$$h_i = f_i \times 100\%$$

FRECUANCIA PORCENTUAL ACUMULADA (H_i): se obtiene multiplicando la frecuencia relativa acumulada por cien.

$$H_i = F_i \times 100\%$$

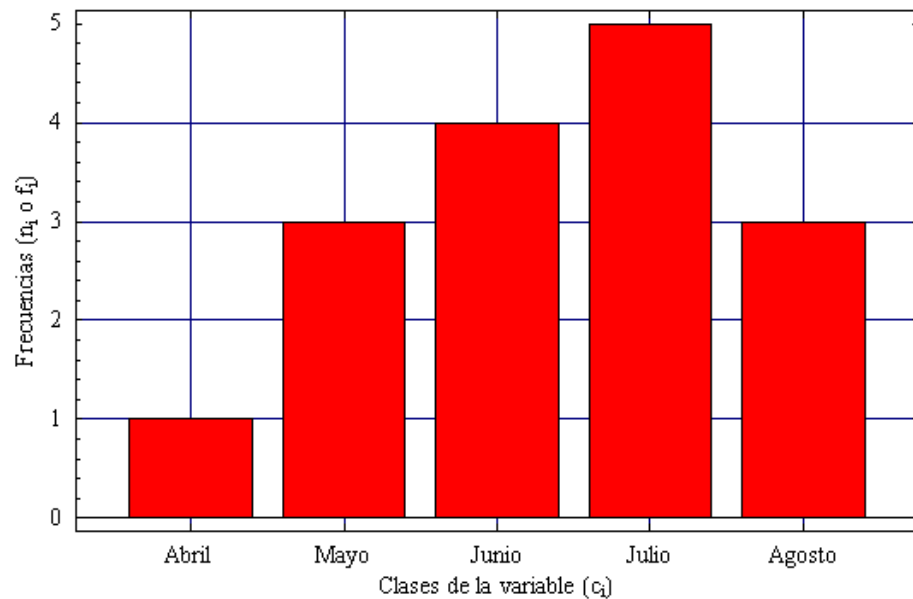
Ejemplo: cantidad de hermanos

Clase	Frecuencia absoluta n_i	Frecuencia absoluta acumulada N_i	Frecuencia relativa f_i	Frecuencia relativa acumulada F_i	Frecuencia porcentual h_i	Frecuencia porcentual acumulada H_i

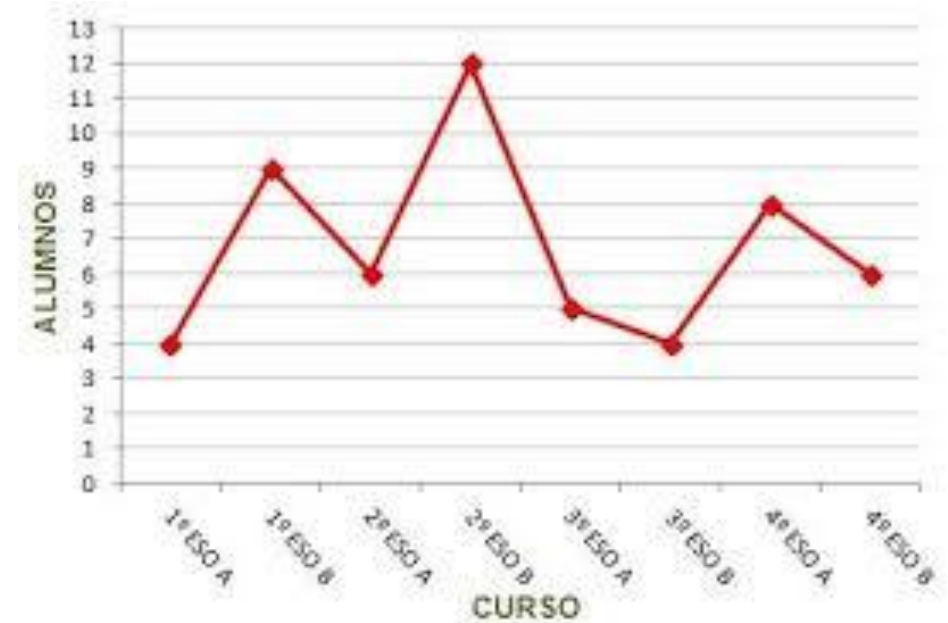
Representación grafica de datos no agrupados:

Gráficos: La principal ventaja de la construcción de gráficos con los datos de una investigación, es que nos permite visualizar más claramente la distribución de éstos, hacer una mejor comparación de resultados y un análisis objetivo de estos últimos.

Gráfico de barra:



Polígono de frecuencias



ESTO ES IMPORTANTE



SLOTHLIBRA.COM

Actividad:

En parejas encuestar a 20 compañeros utilizando una variable cuantitativa discreta

1. Realiza su tabla de frecuencia correspondiente.
 - a) Frecuencia absoluta y absoluta acumulada
 - b) Frecuencia relativa y relativa acumulada
 - c) Frecuencia porcentual y porcentual acumulada
2. Representa pictográficamente los datos utilizando gráfico de barra y polígono de frecuencia.
3. Genera un resumen estadístico.

A continuación veras el contenido de lo que debes aprender estos días

Organización de los datos continuos

Si la variable bajo estudio es cuantitativa continua (o discreta con un alto rango de variabilidad), entonces el esquema de tabla anterior sufre un leve modificación que está relacionada con la creación de los intervalos de clases. En este caso el esquema de la tabla es el siguiente:

Intervalos (Clases)	Marca de clase	n_i	f_i	N_i	F_i
$[\min; \min + A[$	m_1	n_1	f_1	N_1	F_1
$[\min + A; \min + 2A[$	m_2	n_2	f_2	N_2	F_2
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$[\min + (k - 1)A; \max]$	m_k	n_k	f_k	N_k	F_k

Donde la **marca de clase** corresponde al promedio de cada intervalo

Construcción de tabla de frecuencias

Construcción Tablas para Variables Continuas:

Se supone que la amplitud de los intervalos es la misma, se puede seguir los siguientes pasos para la construcción de tablas de frecuencias de variables continuas.

Paso 1: Contar el número n de datos.

Paso 2: Calcular el rango (R),

$$R = \max - \min,$$

donde \min y \max corresponden a los valores mínimos y máximos de los datos, respectivamente.

Paso 3: Escoger el número de clases (intervalos). Se sugiere el entero más próximo de la denominada **fórmula de Sturges**, dada por

$$k = 1 + 3,3 \log(n),$$

donde $\log(\bullet)$ es el logaritmo en base 10.

Paso 4: Calcular la amplitud (A)

$$A = \frac{R}{k}.$$

Paso 5: Para determinar los extremos de la primera clase (intervalo) se debe tomar como límite inferior el valor \min y como límite superior el valor $\min + A$.

Paso 6: Para obtener las restantes clases, se suma sucesivamente A al límite inferior, donde el límite inferior de las sucesivas clases corresponderá a límite superior de la clase anterior.

Ejemplo

Considere los siguientes datos

0,36	0,68	0,8	0,87	0,92	1,00
0,48	0,71	0,81	0,87	0,94	1,00
0,60	0,72	0,81	0,88	0,97	1,13
0,61	0,73	0,82	0,92	0,97	1,16
0,68	0,79	0,85	0,92	0,97	1,19

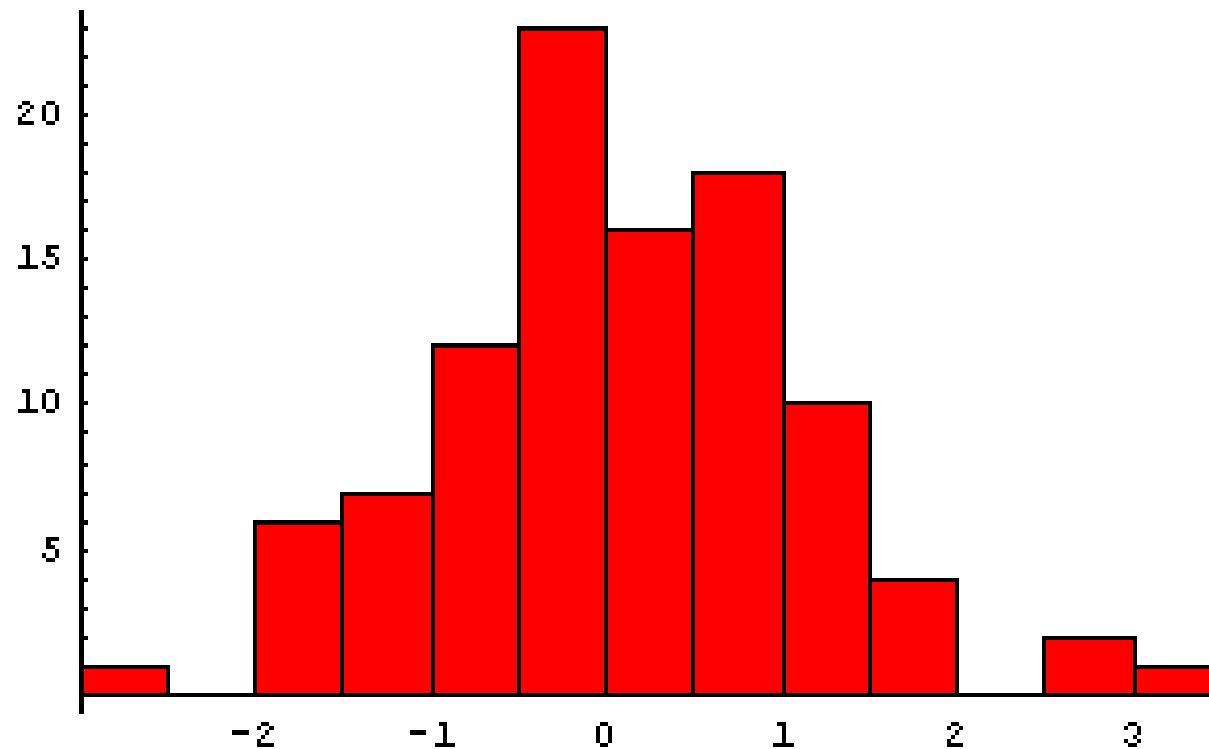
Cree una tabla de distribución de frecuencias con la metodología vista anteriormente

- 1 El valor del tamaño de muestra es $n = 30$.
- 2 El rango sería $R = 1,19 - 0,36 = 0,83$.
- 3 El número de clases a considerar es $k = 1 + 3,3 \log(30) = 5,87 \equiv 6$.
- 4 La amplitud sería $A = \frac{0,83}{6} = 0,138\bar{3}$.
- 5 La tabla quedaría de la siguiente forma

Intervalos	Marca de clase	n_i	f_i	N_i	F_i
$[0,360; 0,498[$	0,429	2	0,067	2	0,06
$[0,498; 0,637[$	0,568	2	0,067	4	0,13
$[0,637; 0,775[$	0,706	5	0,167	9	0,3
$[0,775; 0,913[$	0,844	9	0,3	18	0,6
$[0,913; 1,052[$	0,983	9	0,3	27	0,9
$[1,052; 1,190]$	1,121	3	0,1	30	1

Representación gráfica de datos agrupados

Un **histograma** es una representación gráfica de datos agrupados mediante intervalos. Los datos provienen de una variables cuantitativas continuas. Gracias a él puedes hacerte rápidamente una idea de la distribución de los datos o muestra.



Ahora tu....

En un centro comercial, se consultó la edad a todas las personas que entraban entre las 12:00 h y 12:30 h. Los resultados obtenidos fueron los siguientes:

15	73	1	65	16	3	42
36	42	3	61	19	36	47
30	45	29	73	69	34	23
22	21	33	27	55	58	17
4	17	48	25	36	11	4
54	70	51	3	34	26	10

- Crea los intervalos de una tabla de frecuencias (usa método Sturges)
- Determina la frecuencia absoluta para cada intervalo
- Realiza un histograma con la información obtenida en los pasos anteriores

Presentación de Datos: Gráficos Estadísticos

Las grandes cantidades de datos estadísticos resultan incómodos de interpretar y si éstos no están ordenados de alguna manera. La principal ventaja de la construcción de gráficos con los datos de una investigación, es que nos permite visualizar más claramente la distribución de éstos, hacer una mejor comparación de resultados y un análisis objetivo de estos últimos.

Una buena definición de lo que es un gráfico es la siguiente.

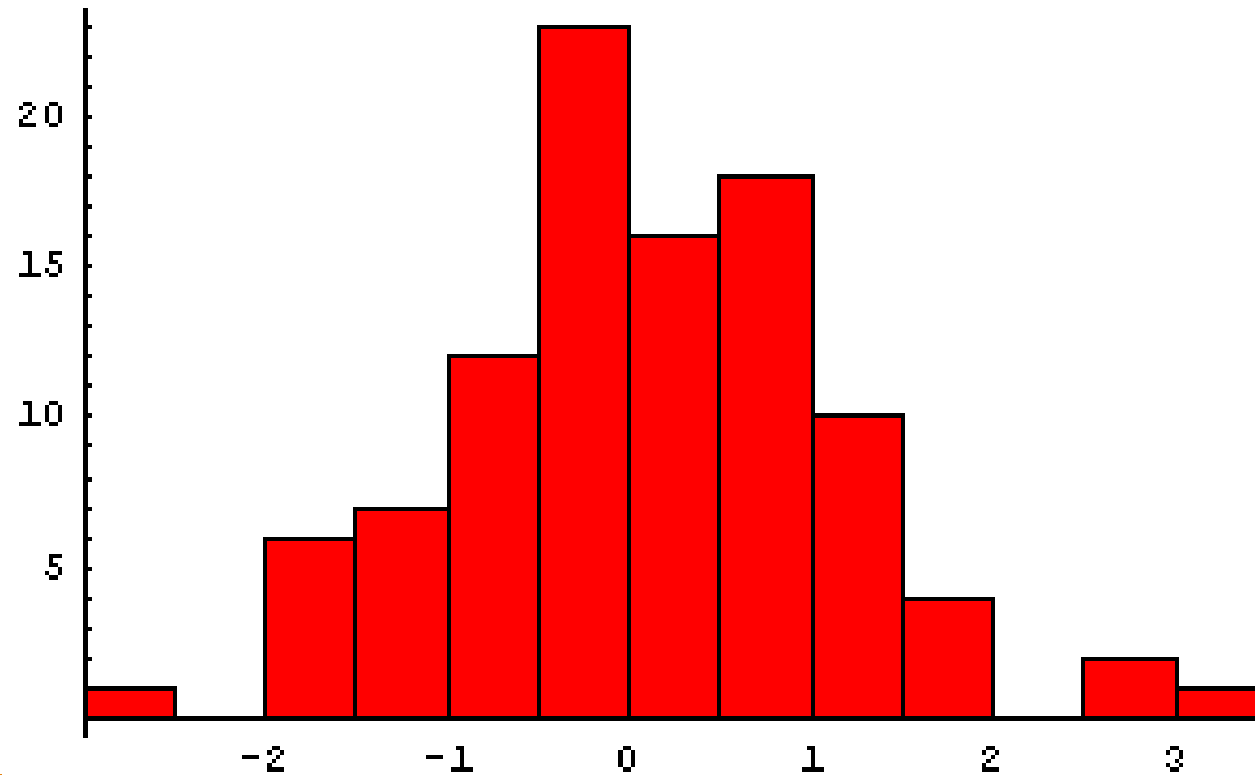
Definición de gráfico:

Es una representación pictórica, mediante figuras geométricas u otros elementos, que proporciona un resumen de la información que interesa destacar y, lo más importante, recordar.

Histograma

Se usa para variables continuas. Es un conjunto de rectángulos adyacentes.

En el eje horizontal deben ir los intervalos (clases) y en el eje vertical las frecuencias.



Polígono de frecuencias

Un polígono de frecuencia es aquel que se forma a partir de la unión de los distintos puntos medios de las cimas de las columnas que configuran lo que es un histograma de frecuencia o un gráfico de barra.

