

**CENTRO DE CAPACITACIÓN
CÁMARA NACIONAL DE COMERCIO**

“Estadística APLICADAS para la definición de Indicadores de Gestión”

Policía de Investigaciones de Chile

Relator: Carlos Echeverría Muñoz
Ingeniero Comercial y Periodista
Master in Business Administration
Project Management Professional del PMI®
carlosecheverriam@gmail.com

MUESTREO

Métodos de selección de las muestras

Muestreo probabilístico: Es un proceso de la muestra en el cual los elementos son elegidos por métodos aleatorios, o sea, la selección de los elementos para la muestra se efectúa por procedimientos al azar y con probabilidades conocidas de selección.

MUESTREO

Métodos de selección de las muestras

1-Muestreo aleatorio simple: Es una selección de muestras en el cual los elementos o unidades se eligen individual y directamente por medio de un proceso aleatorio, en el que cada elemento no seleccionado tiene la misma oportunidad de ser elegido al igual que todos los otros elementos en cada extracción de la muestra. De modo que cada elemento en la población debe tener igual probabilidad de ser seleccionado.

MUESTREO

Métodos de selección de las muestras

2-Muestreo sistemático: Los elementos se selecciona en forma ordenada y depende del número de elementos o unidades incluidos en la población y el tamaño de la muestra. Requiere del uso de un listado de todos los elementos de la población.

MUESTREO

Métodos de selección de las muestras

3-Muestreo estratificado: La población se divide en cierto número de subgrupos o estratos, cada uno de los cuales se muestrea independientemente. El proceso a través del cual se divide la población en subgrupos o estratos, recibe el nombre de estratificación. El objeto de la estratificación es llevar a cabo selecciones separadas en cada uno de los subgrupos o estratos.

MUESTREO

Métodos de selección de las muestras

Muestreo no probabilístico: Incluye todos los métodos en que los elementos de la muestra **no se seleccionan mediante procedimientos al azar o aleatorios, o con probabilidades de selección conocidas.** Algunos procedimientos de selección del muestreo no probabilístico son:

- Muestreo de juicio:
- Muestreo por cuotas:.
- Muestreo decisional:
- Muestreo de agrupación causal:

LOS DATOS ESTADÍSTICOS

Los datos de características cuantitativas: son aquellos que se pueden expresar numéricamente y se obtienen a través de mediciones y conteos. Un dato cuantitativo se puede encontrar en cualquier disciplina; psicología, contabilidad, economía, publicidad, etc

LOS DATOS ESTADÍSTICOS

Los datos de características cuantitativas y cualitativas se clasifican a su vez en:

1-VARIABLES CONTINUAS:

2-VARIABLES DISCRETAS:

LOS DATOS ESTADÍSTICOS

Datos de características cualitativas:
Los datos de características cualitativas son aquellos que no se pueden expresar numéricamente. Estos datos se deben convertir a valores numéricos antes de que se trabaje con ellos.

LOS DATOS ESTADÍSTICOS

Los datos de características cualitativas se clasifican en:

1-Datos nominales: Comprenden categorías, como el sexo, carrera de estudio, material de los pisos, calificaciones, etc.

2-Datos jerarquizados: evaluaciones subjetivas cuando los conceptos se jerarquizan según la preferencia o logro

LOS DATOS ESTADÍSTICOS

Distribución de frecuencias: Cuando se dispone de una gran masa o cantidad de datos a veces resulta muy difícil agruparlos.

A esta forma de organizar las informaciones se le llama distribución de frecuencias y consiste en el ordenamiento de los datos a través de clases y frecuencias.

LOS DATOS ESTADÍSTICOS

Cuando los datos se presentan en una distribución de frecuencias se les denomina datos agrupados. Cuando todos los datos observados de una variable se enumeran en forma desorganizada le vamos a denominar datos no agrupados.

DISTRIBUCIÓN DE FRECUENCIAS

DISTRIBUCIÓN DE FRECUENCIAS

Introducción

La fase previa de cualquier estudio estadístico se basa en la recogida y ordenación de datos; esto se realiza con la ayuda de los resúmenes numéricos y gráficos.

DISTRIBUCIÓN DE FRECUENCIAS

Medidas de posición

Son aquellas medidas que nos ayudan a saber donde están los datos pero sin indicar como se distribuyen.

DISTRIBUCIÓN DE FRECUENCIAS

Pasos para construir una distribución de frecuencias: conocidos ya todos los elementos teóricos necesarios para la construcción y comprensión de una distribución de frecuencias vamos a proceder a mostrar los pasos requeridos para su ejecución.

DISTRIBUCIÓN DE FRECUENCIAS

1° Determinar el recorrido o rango.

$$R = X \text{ máx.} - X \text{ mín.}$$

2° Calcular el intervalo de clase, siempre que se conozca el número de clases.

$$R = R$$
$$NC$$

DISTRIBUCIÓN DE FRECUENCIAS

3° Calcular el número de clases, siempre que se conozca el intervalo de clase.

$$NC = \frac{R}{Ci}$$

Como se puede observar en el segundo y tercer paso resultaría muy difícil resolver estas ecuaciones por simples métodos matemáticos ya que cada una de ellas presenta dos incógnitas.

DISTRIBUCIÓN DE FRECUENCIAS

Como solución a este problema surge la fórmula de Sturges que se expresa así:

$$C_i = R .$$

$$1 + 3.22 \log N$$

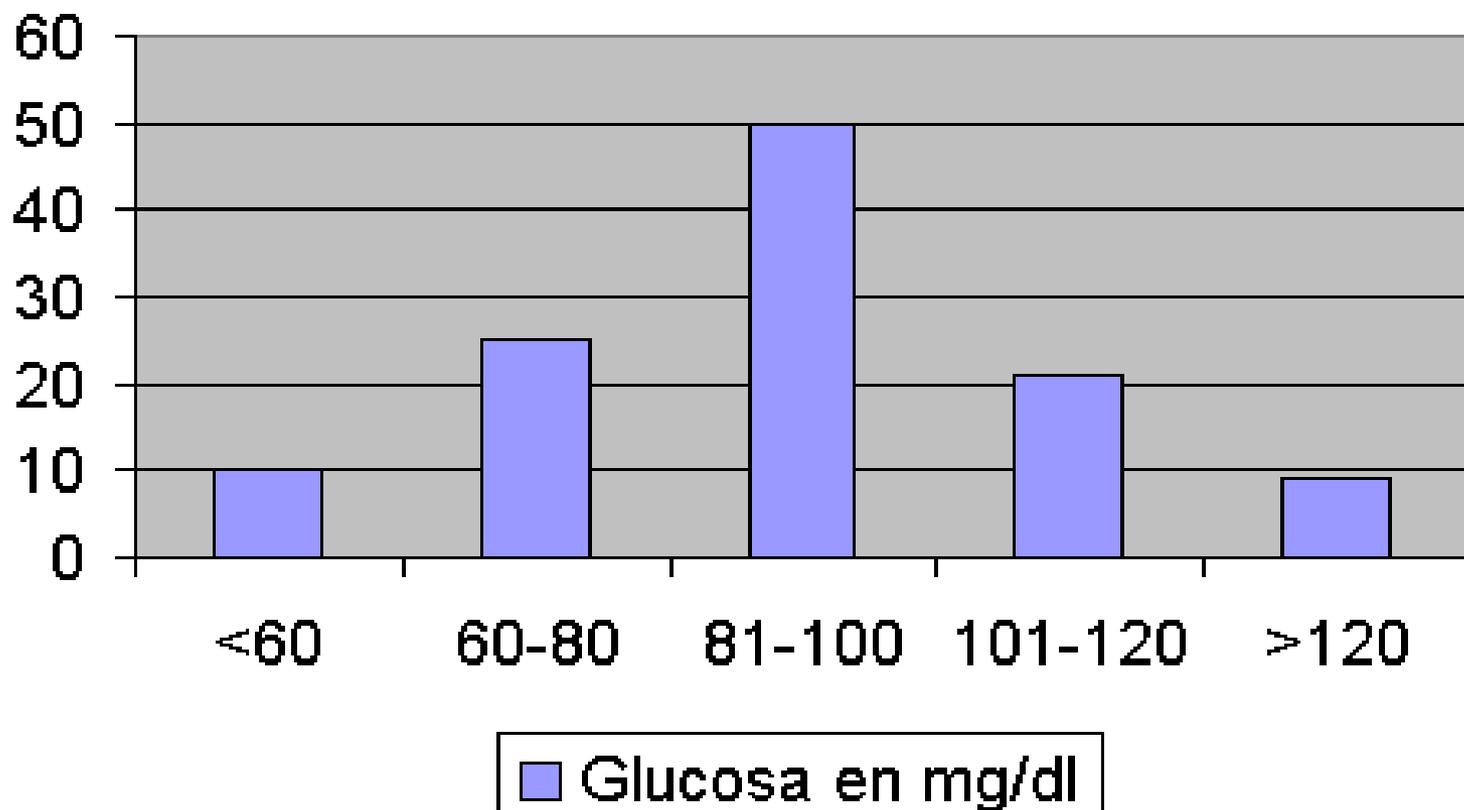
Donde R = recorrido y N = número total de valores.

HISTOGRAMA

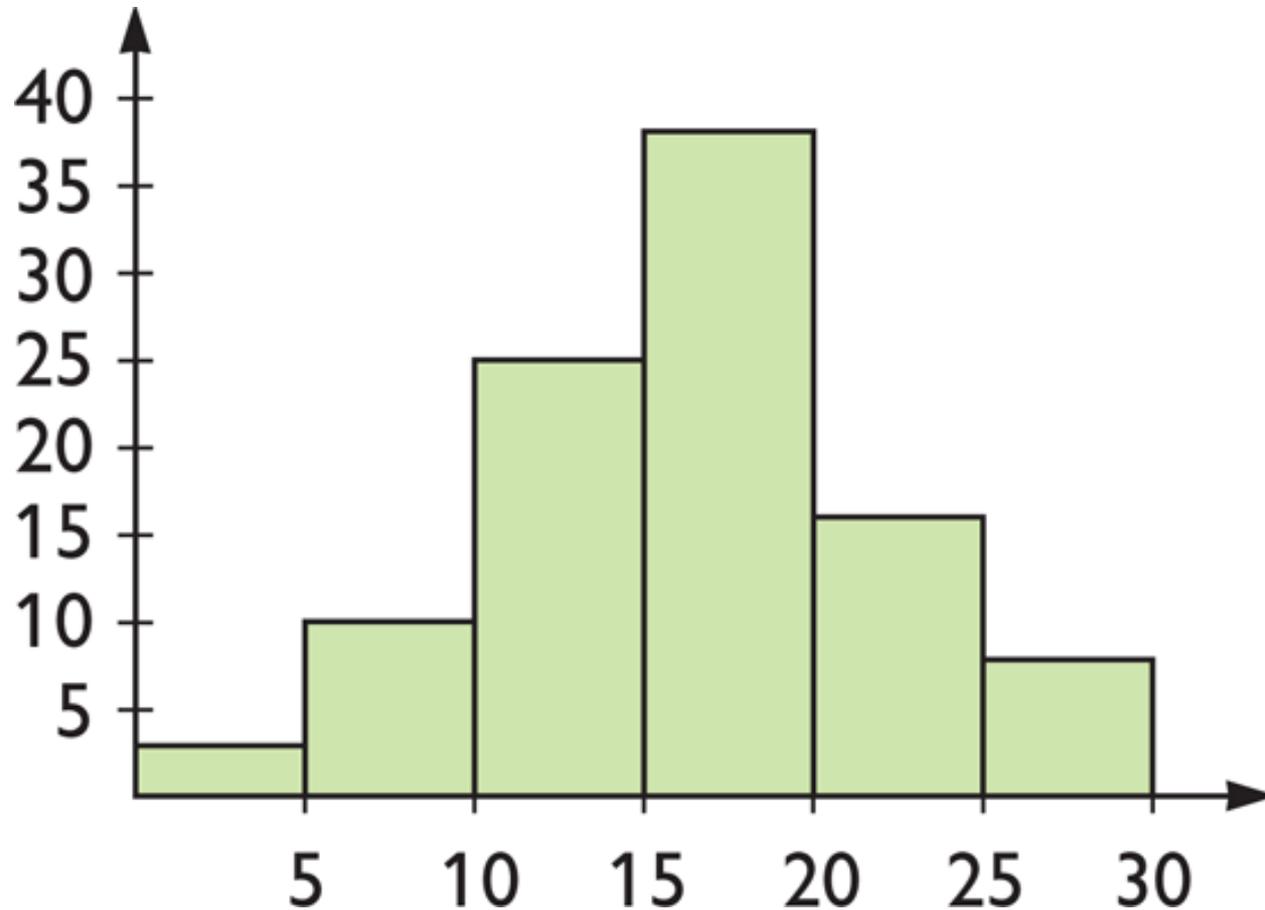
En estadística, un histograma es una representación gráfica de una variable en forma de barras, donde la superficie de cada barra es proporcional a la frecuencia de los valores representados.

HISTOGRAMA

Histograma simple



HISTOGRAMA

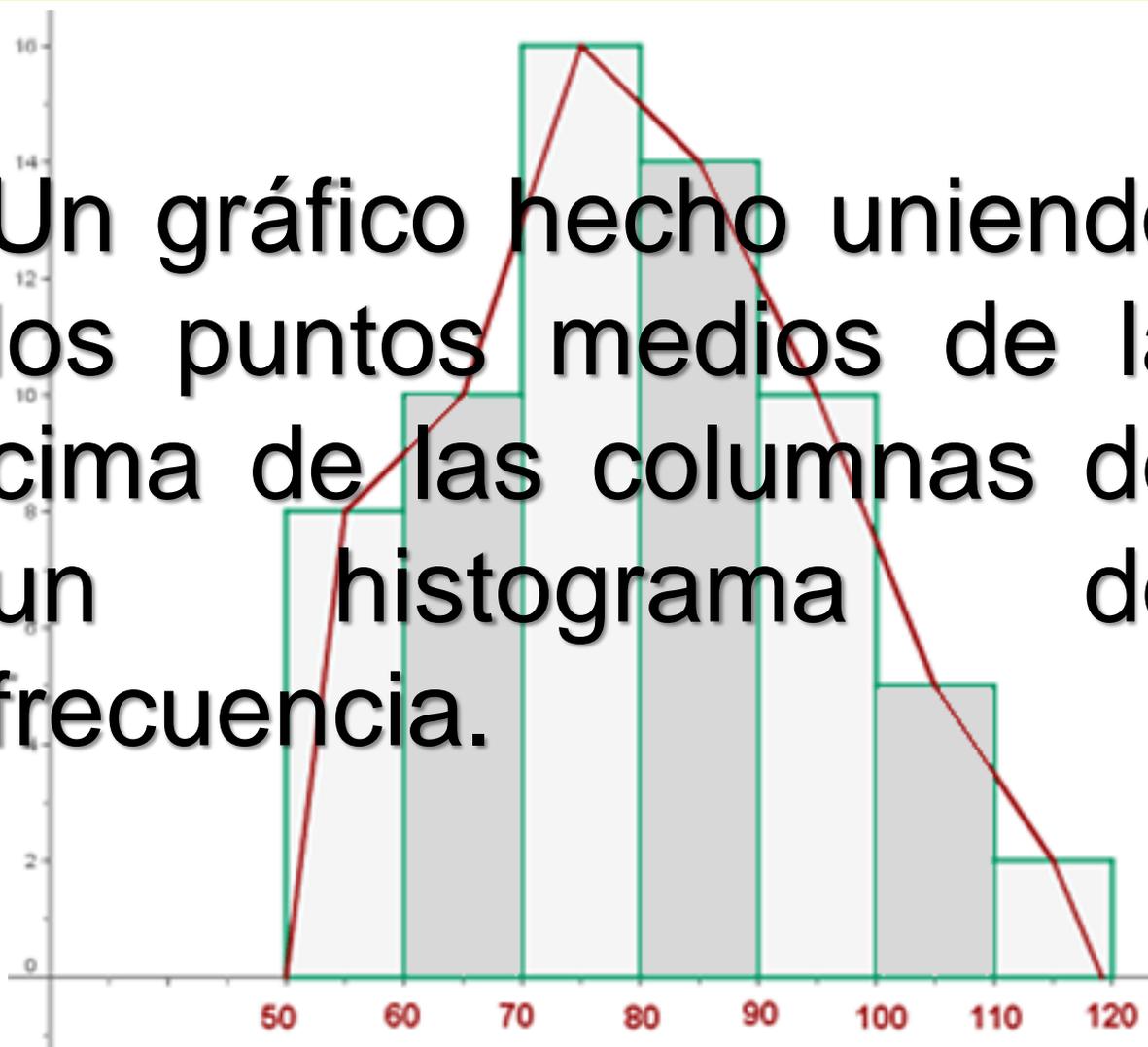


HISTOGRAMA

Se utiliza cuando se estudia una variable continua, como franjas de edades o altura de la muestra, y, por comodidad, sus valores se agrupan en clases, es decir, valores continuos. En los casos en los que los datos son cualitativos (no-numéricos), como sexto grado de acuerdo o nivel de estudios, es preferible un diagrama de sectores.

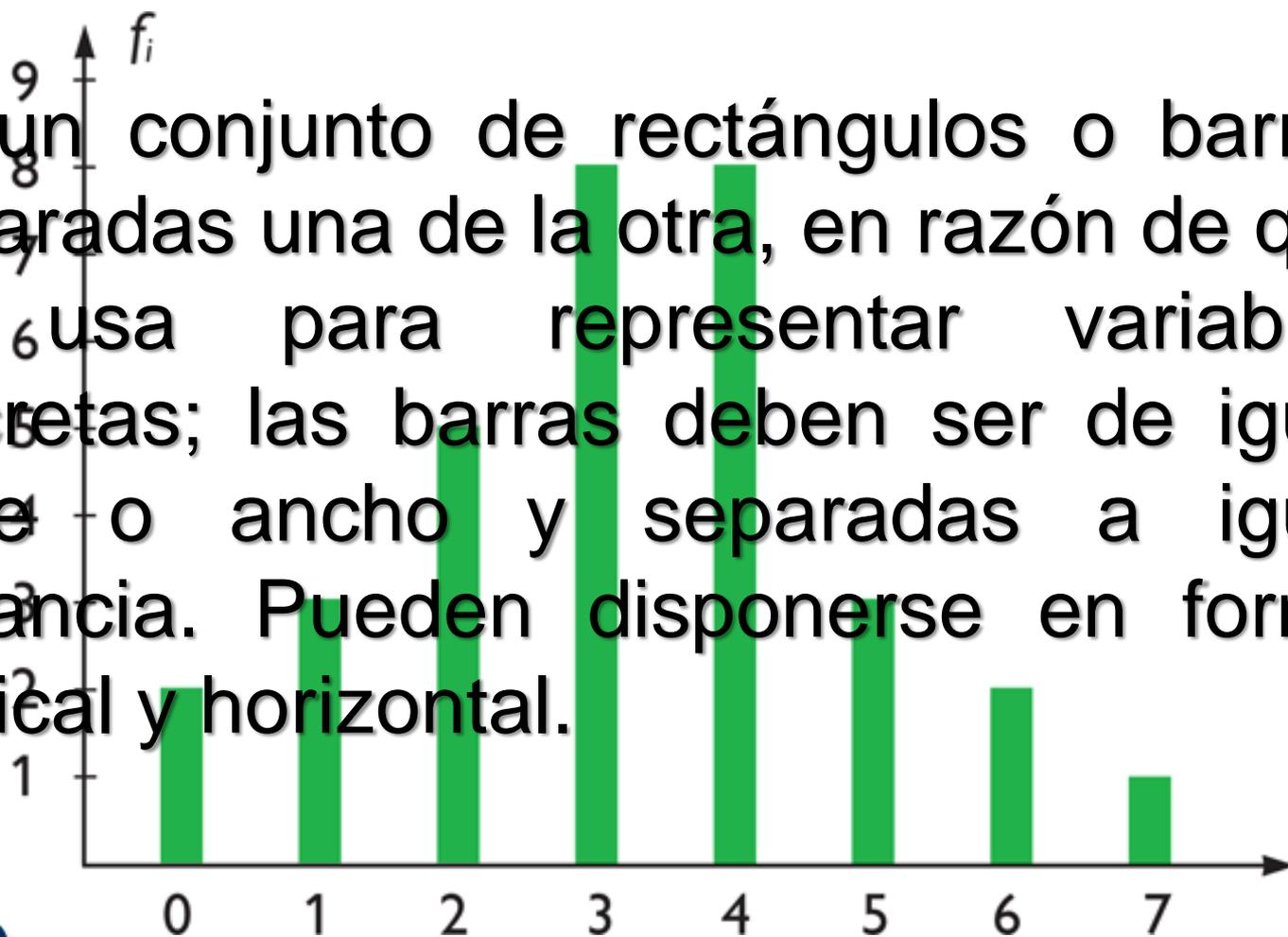
Polígono de frecuencias

Un gráfico hecho uniendo los puntos medios de la cima de las columnas de un histograma de frecuencia.



Gráfica de barras

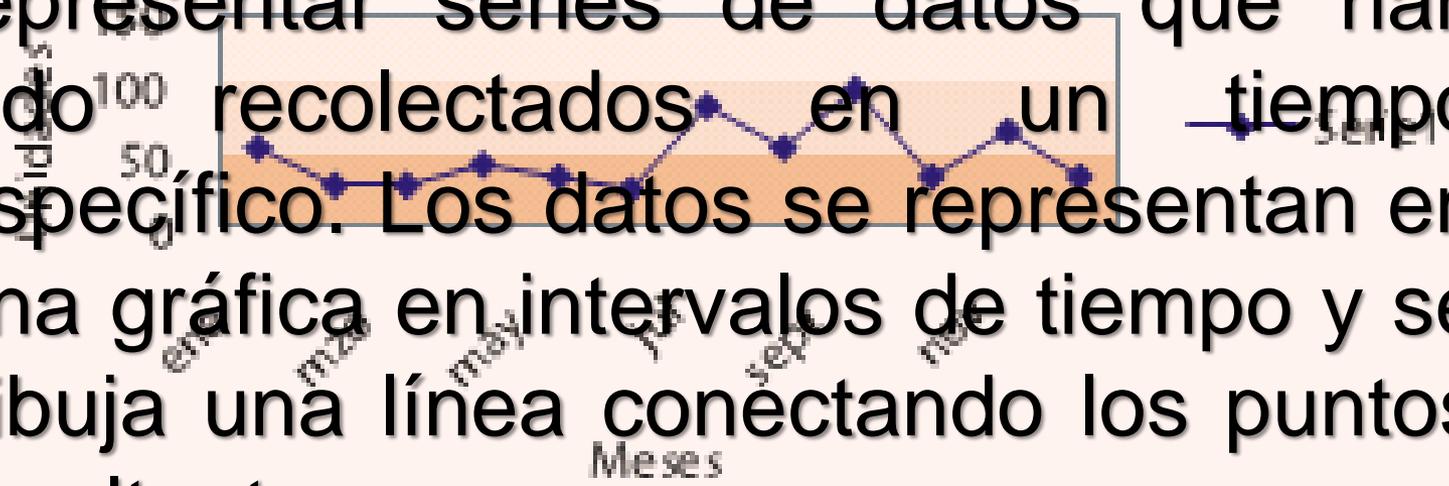
Es un conjunto de rectángulos o barras separadas una de la otra, en razón de que se usa para representar variables discretas; las barras deben ser de igual base o ancho y separadas a igual distancia. Pueden disponerse en forma vertical y horizontal.



Gráfica Lineal

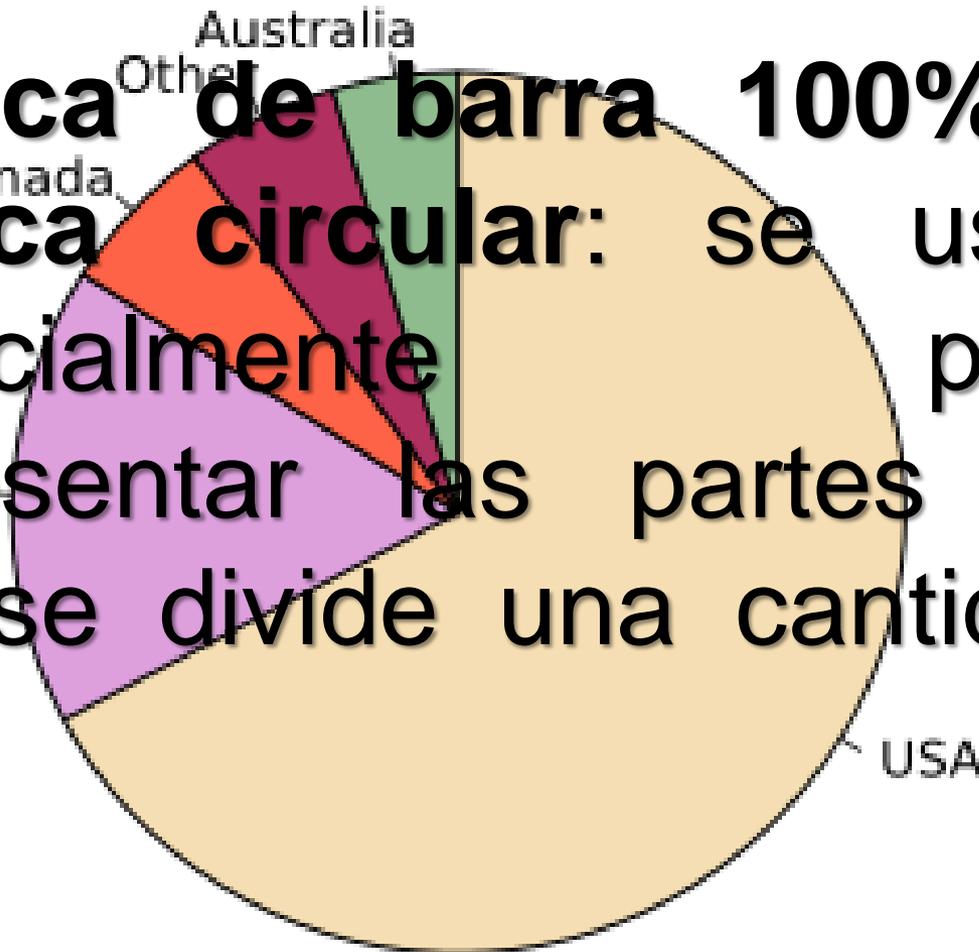
Gráfica Lineal

Una gráfica lineal se utiliza para representar series de datos que han sido recolectados en un tiempo específico. Los datos se representan en una gráfica en intervalos de tiempo y se dibuja una línea conectando los puntos resultantes.



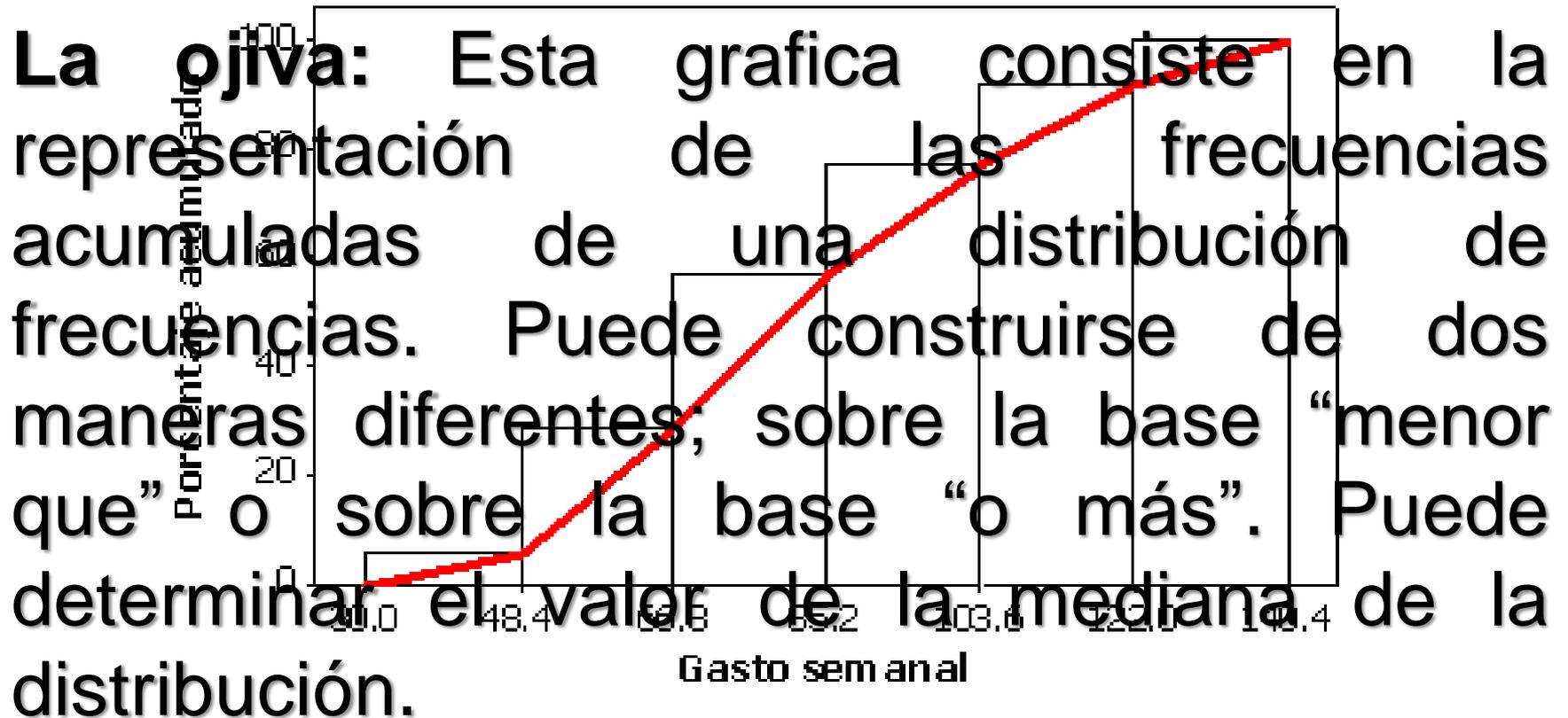
GRÁFICA CIRCULAR:

Gráfica de barra 100% y gráfica circular: se usan especialmente para representar las partes en que se divide una cantidad total.

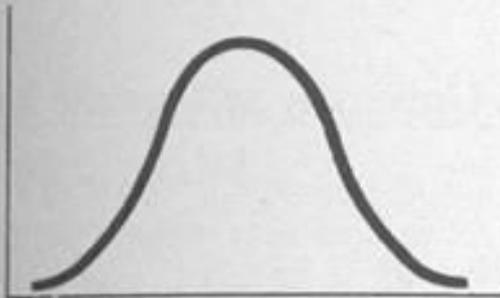


La ojiva:

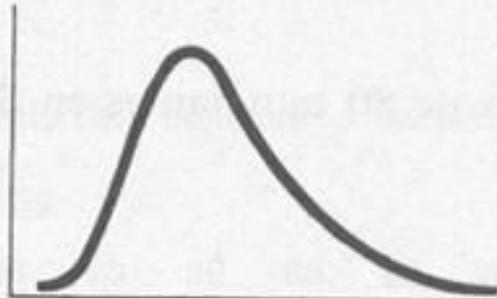
Ojiva de frecuencias



EJEMPLOS A DISCUTIR



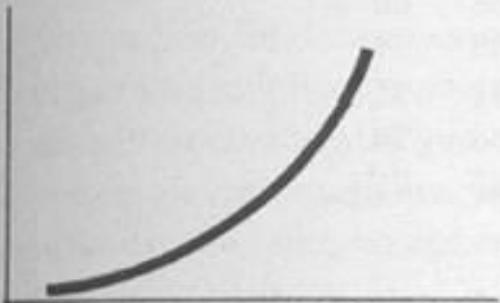
Simétrica o bien formada



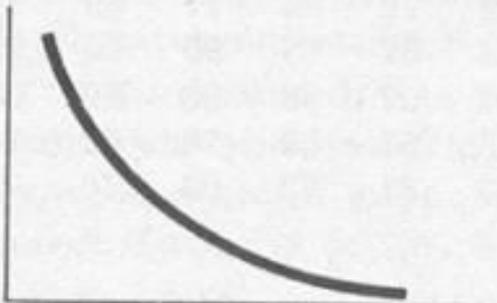
Sesgada a la derecha
(sesgo positivo)



Sesgada a la izquierda
(sesgo negativo)



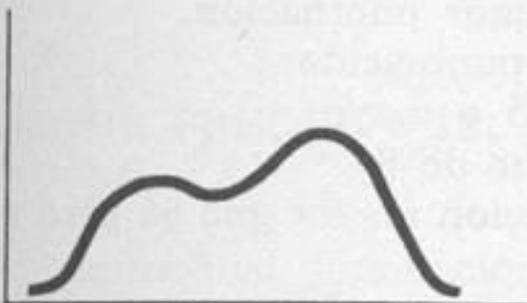
En forma de J



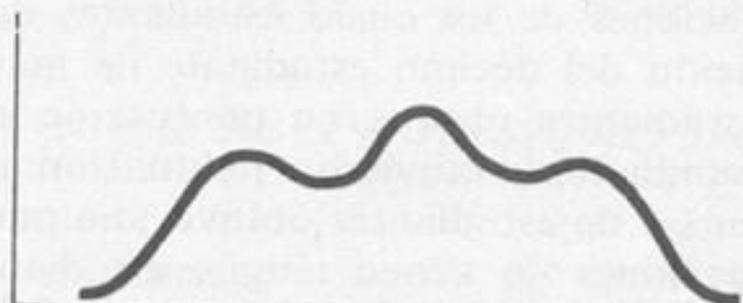
En forma de J invertida



En forma de U

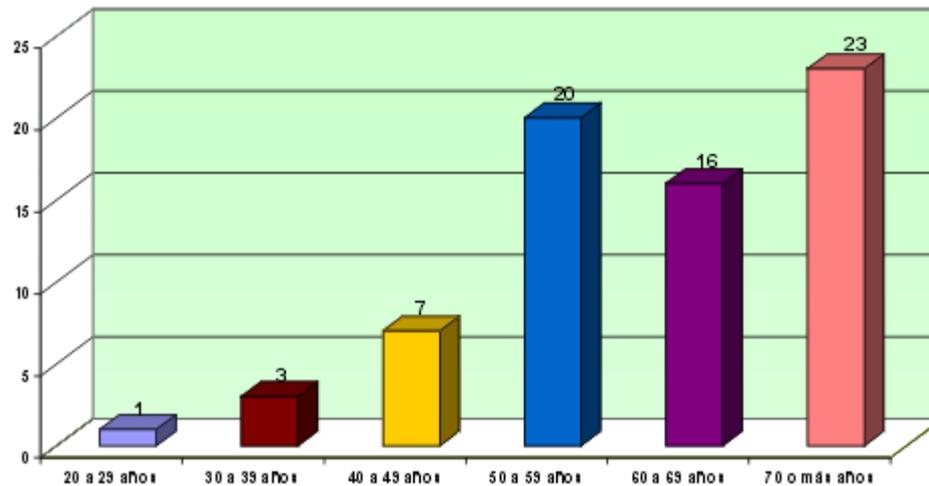
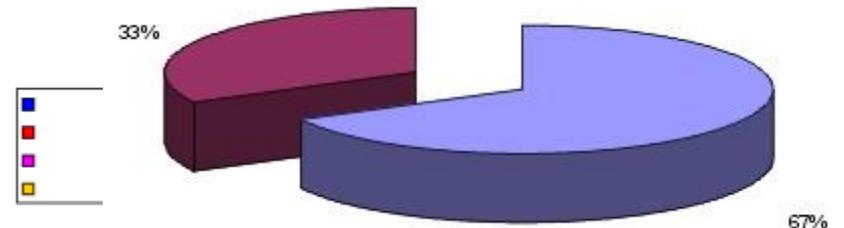
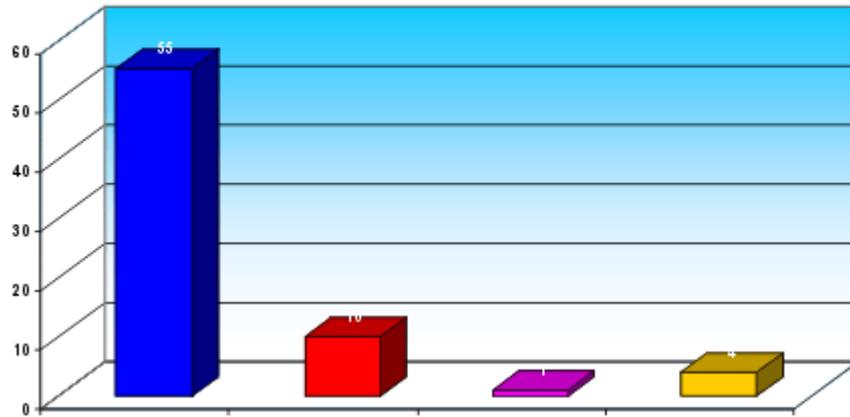


Bimodal

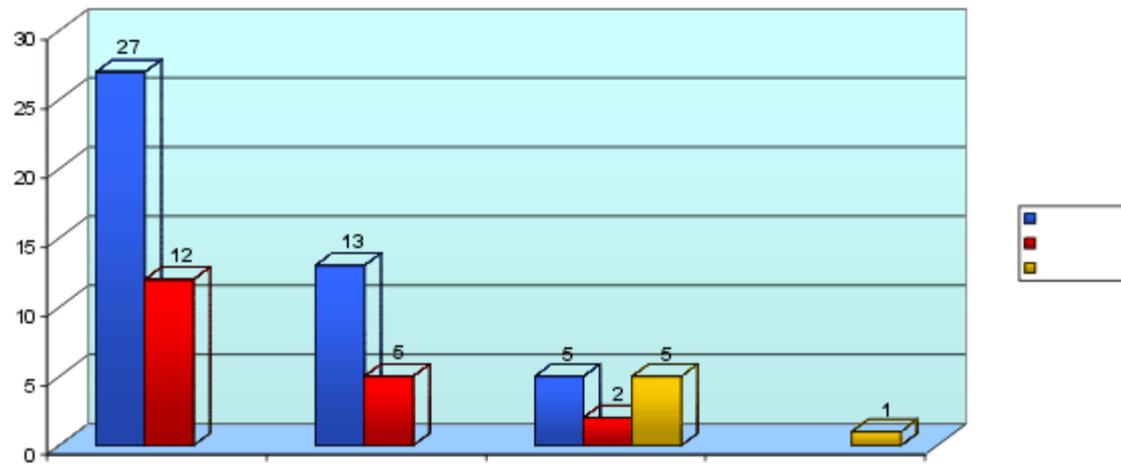
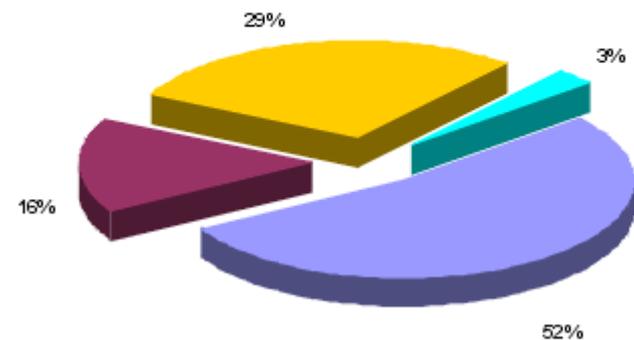
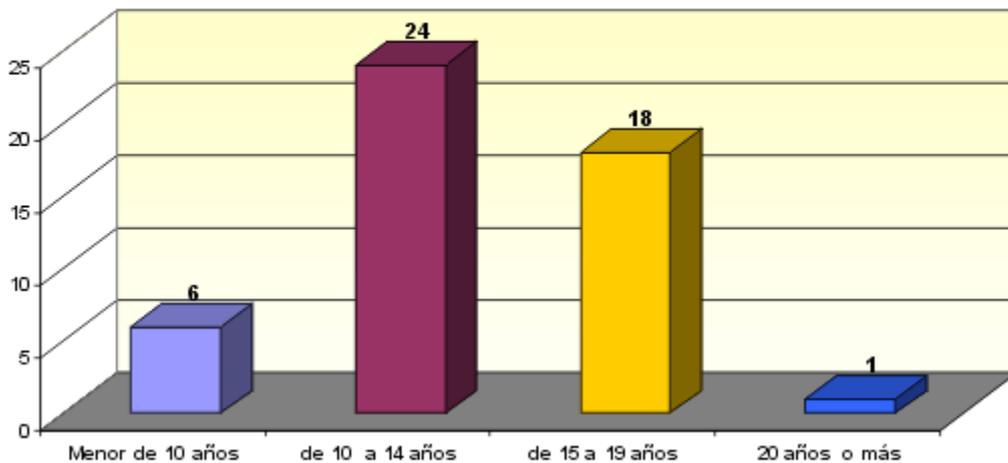


Multimodal

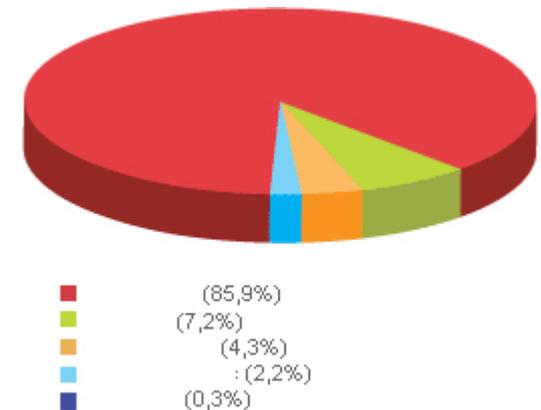
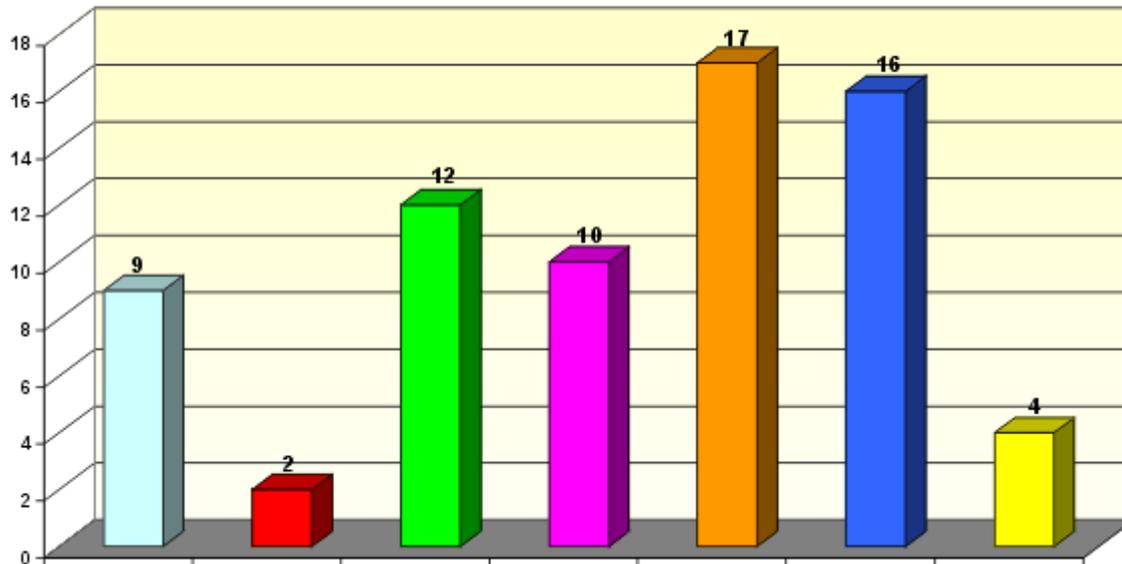
EJEMPLOS A DISCUTIR



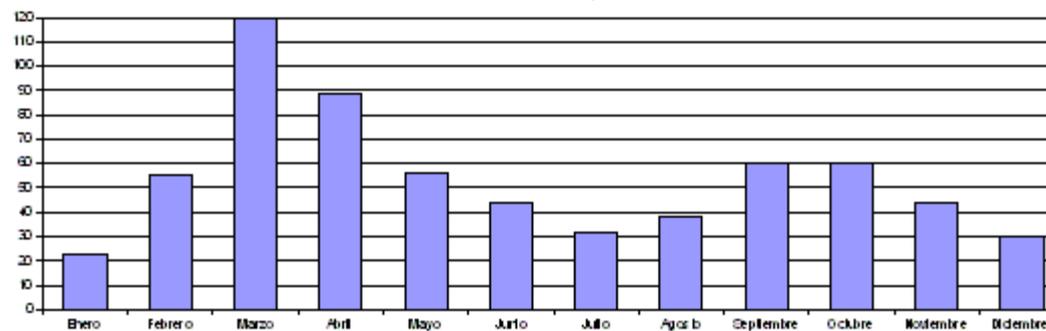
EJEMPLOS A DISCUTIR



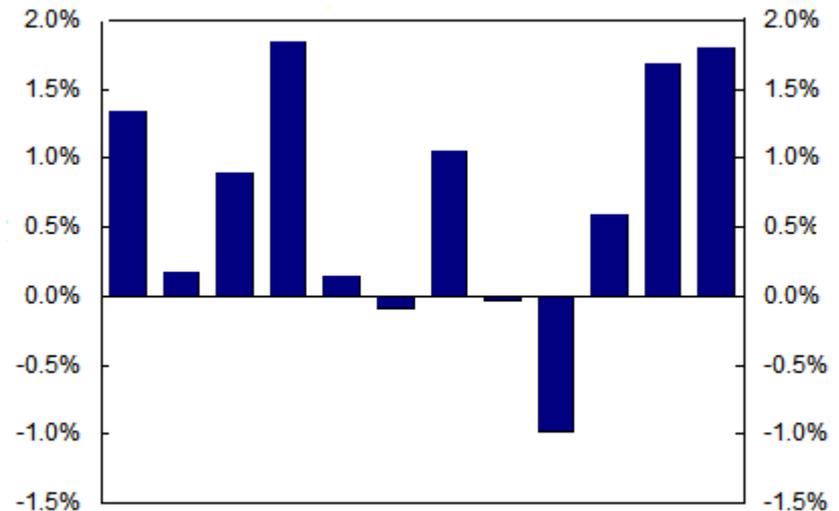
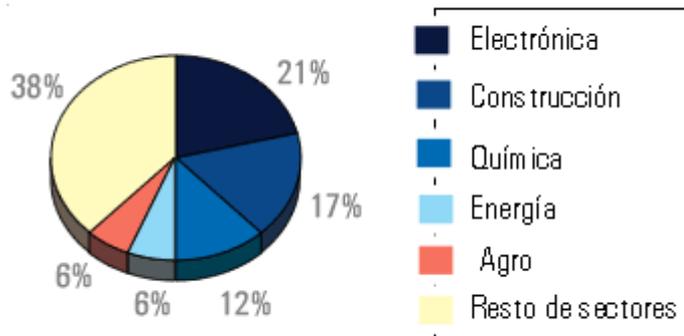
EJEMPLOS A DISCUTIR



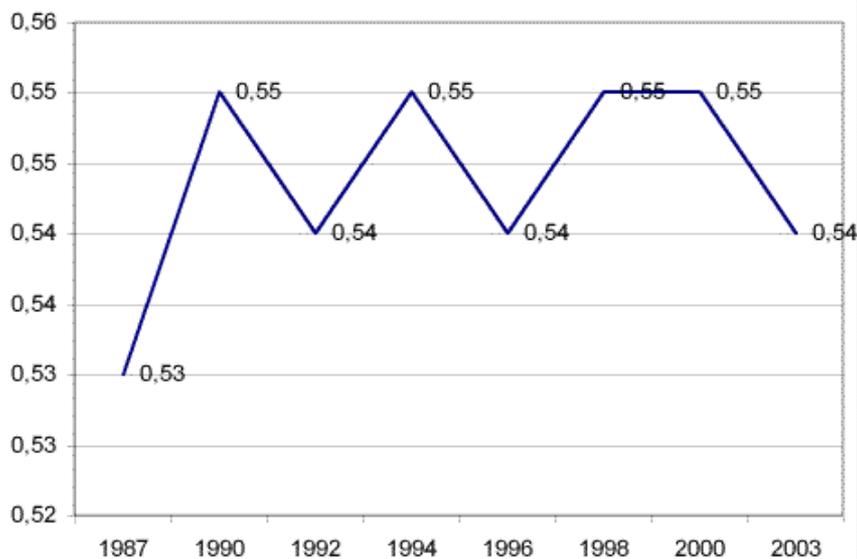
Número de árboles plantados



EJEMPLOS A DISCUTIR

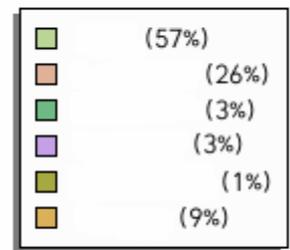
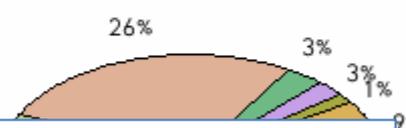
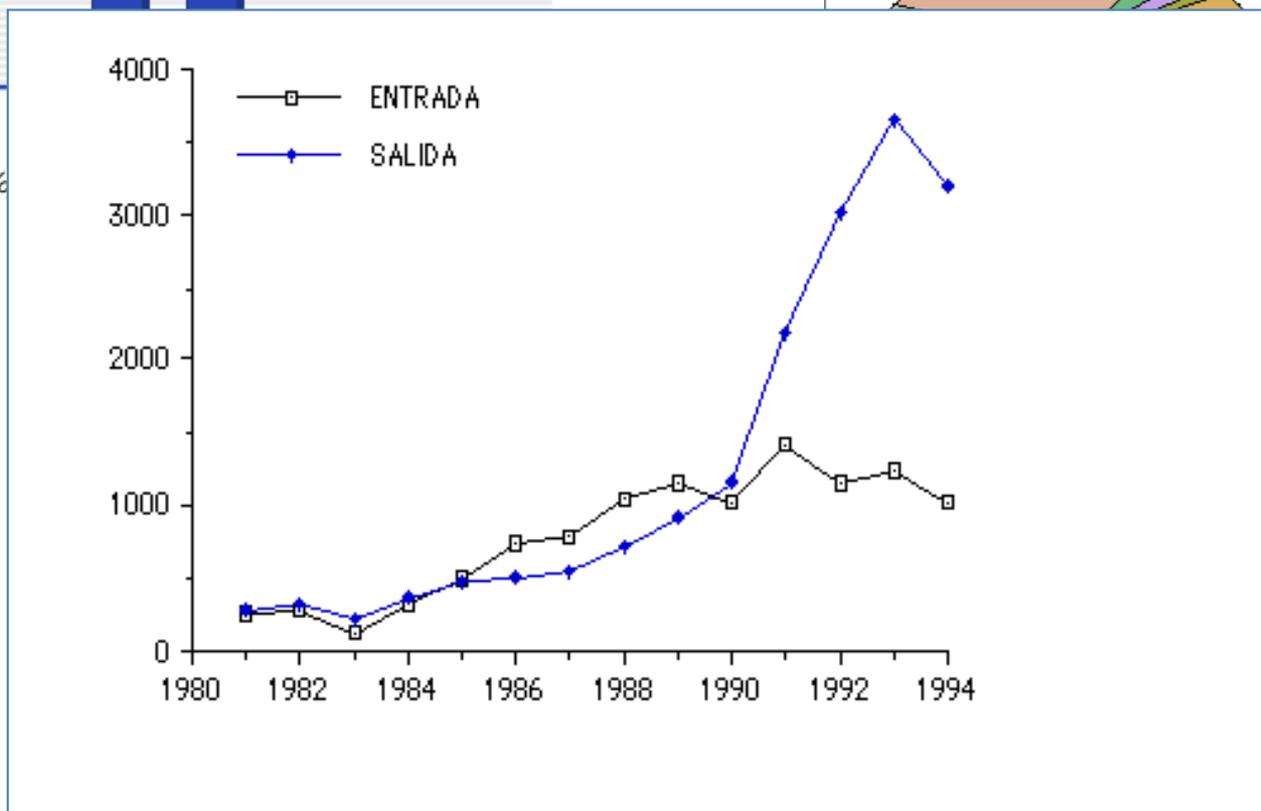
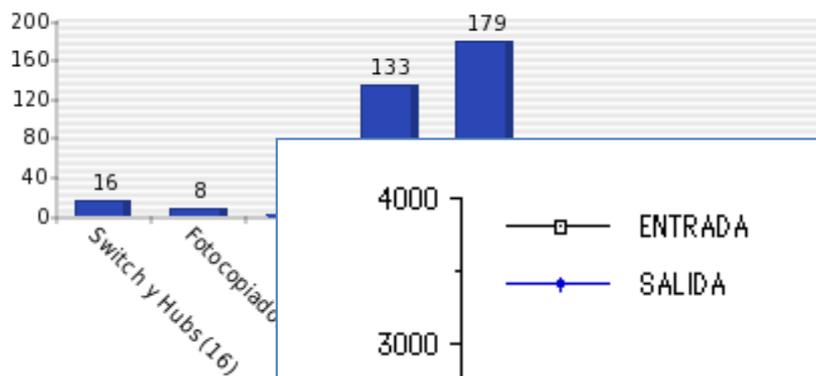


Desigualdad en Chile (Índice de GINI)

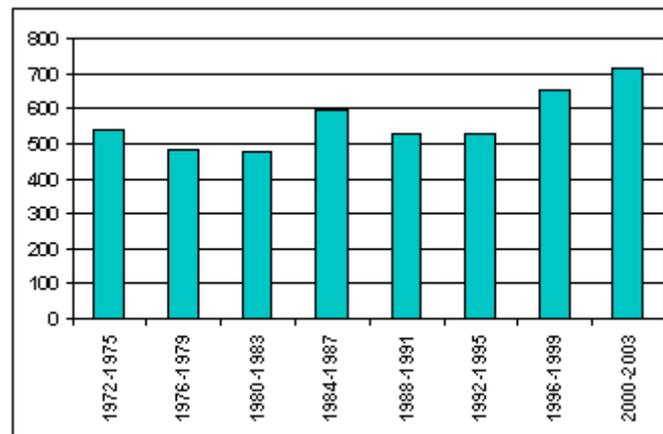
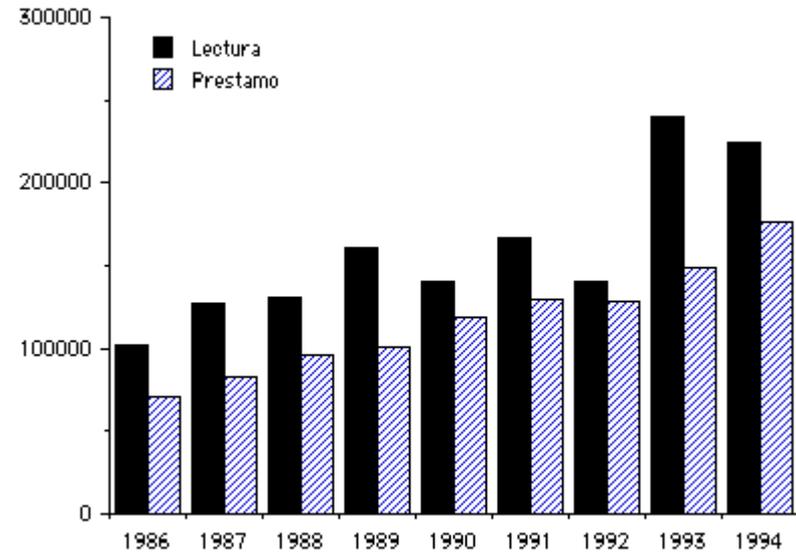
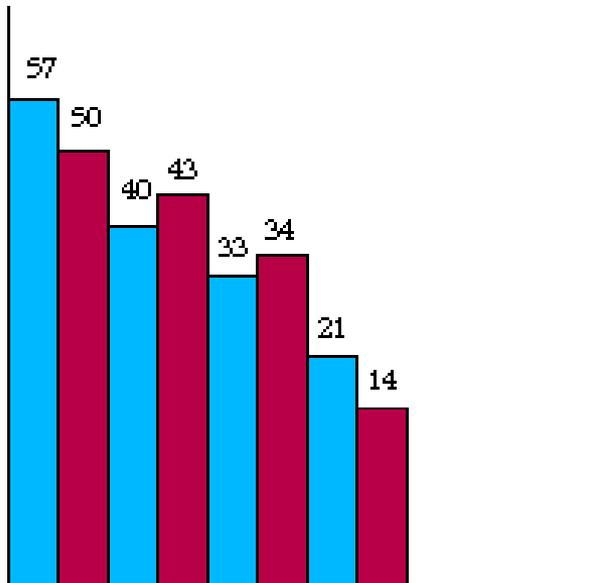


EJEMPLOS A DISCUTIR

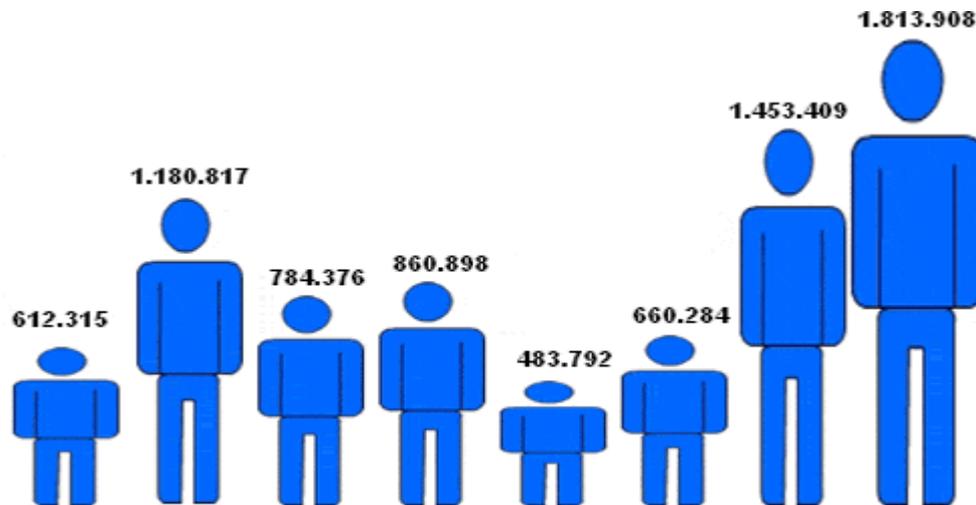
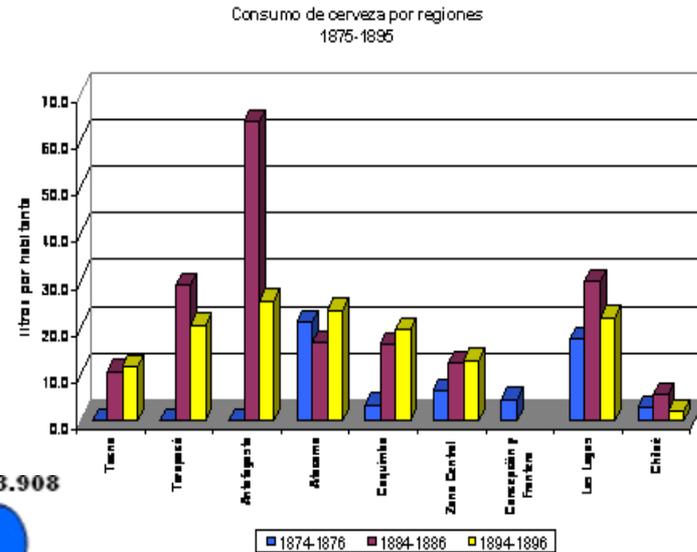
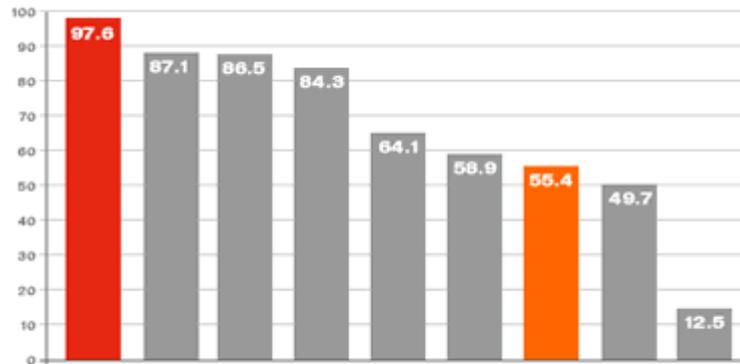
EQUIPOS DE COMPUTO Y OFICINA



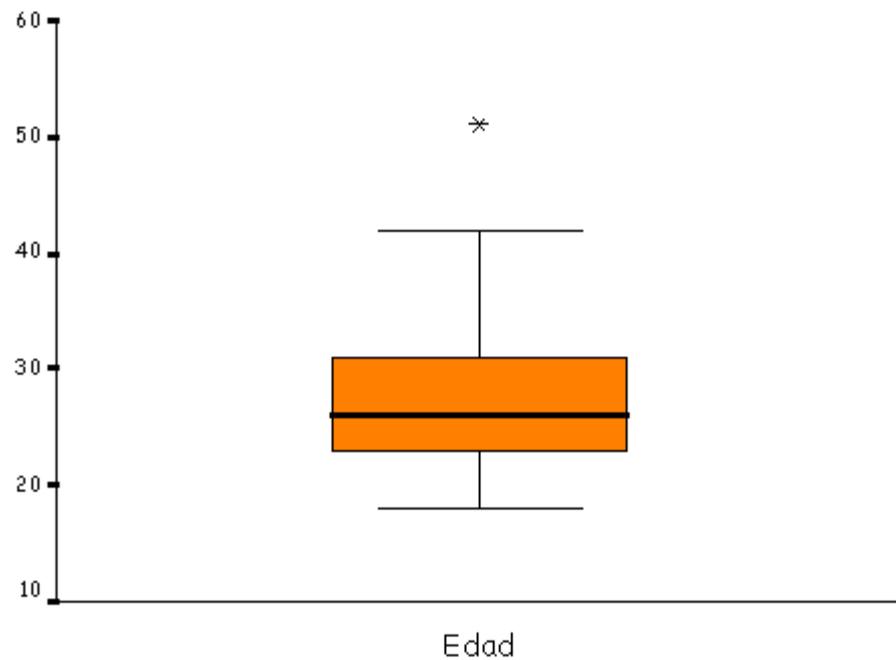
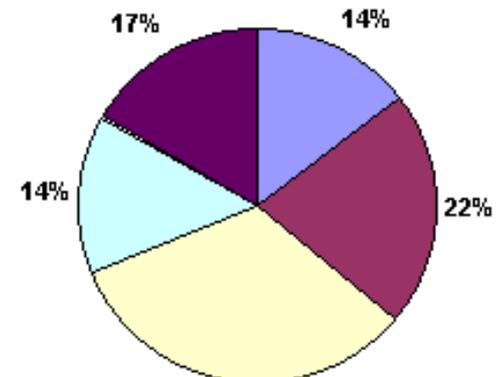
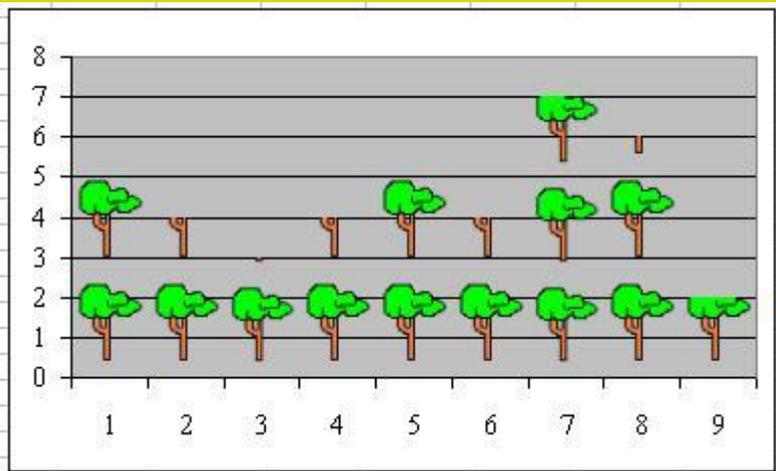
EJEMPLOS A DISCUTIR



EJEMPLOS A DISCUTIR

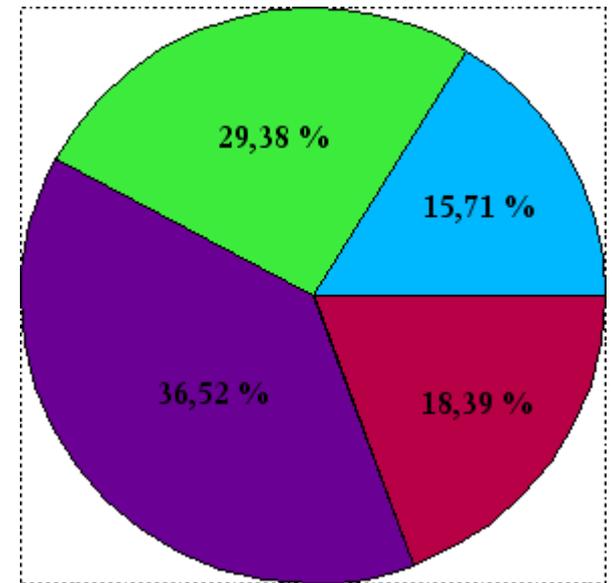
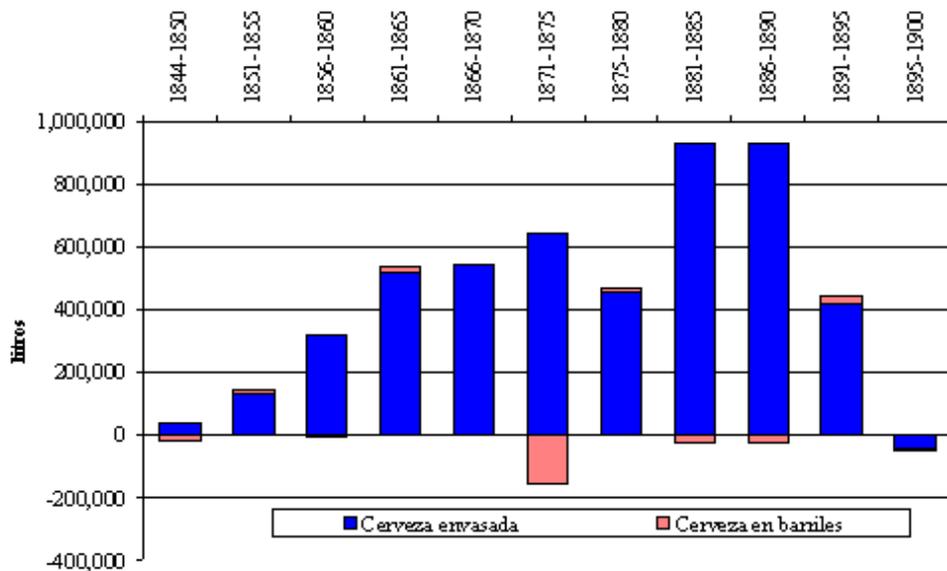


EJEMPLOS A DISCUTIR



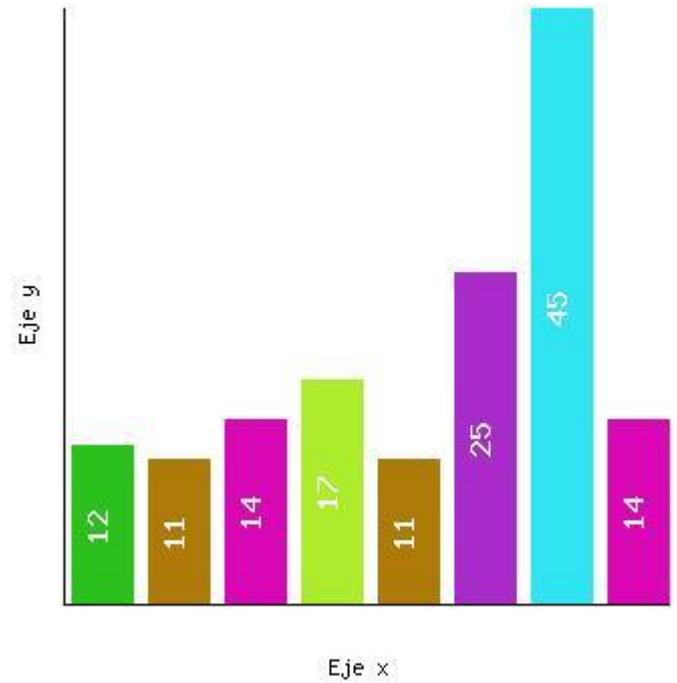
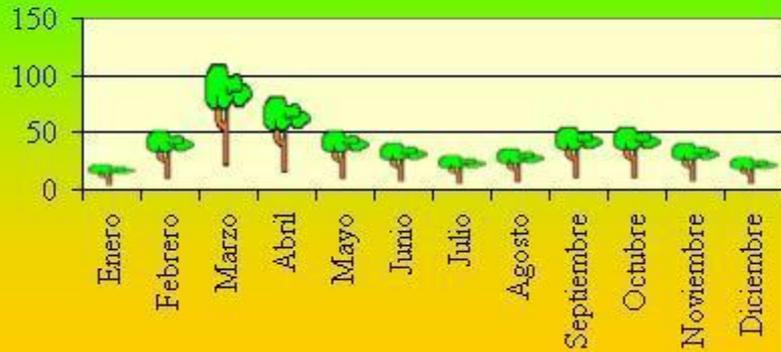
EJEMPLOS A DISCUTIR

Importaciones netas de cerveza
1844-1900

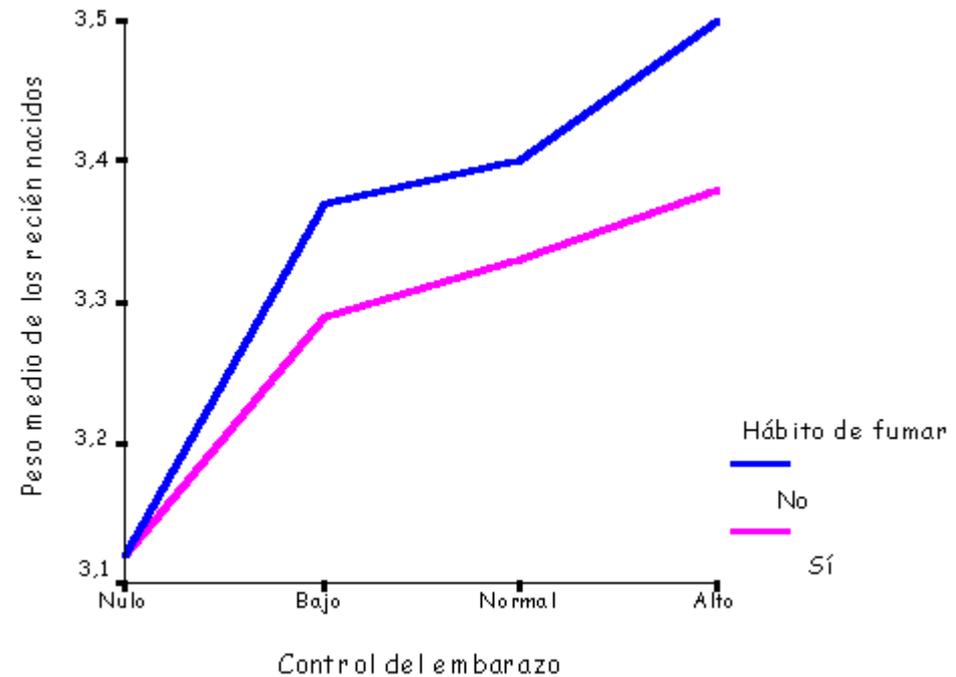
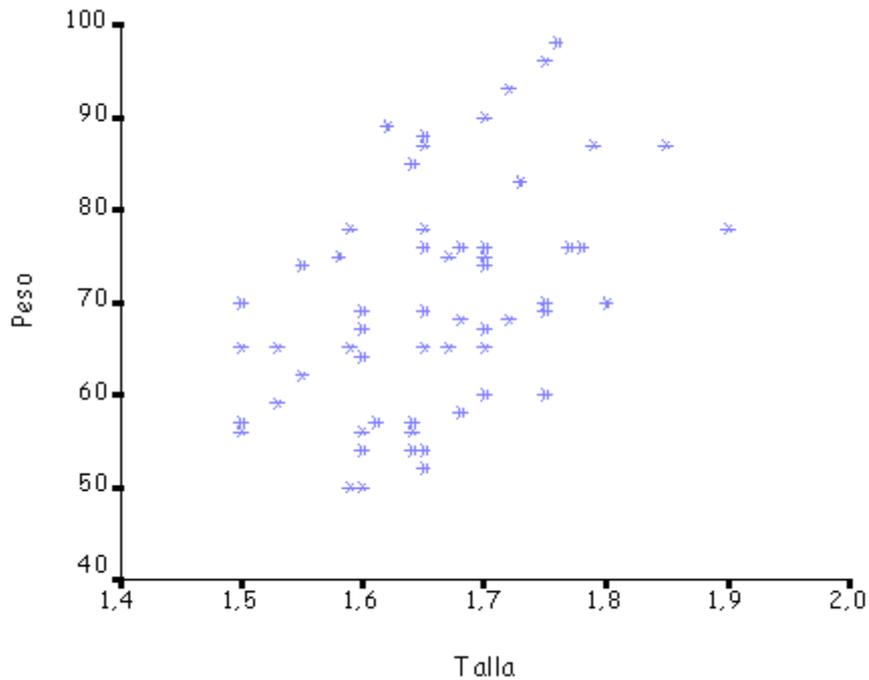


EJEMPLOS A DISCUTIR

Arboles plantados



EJEMPLOS A DISCUTIR



MEDIDAS DE POSICIÓN

MEDIA ARITMÉTICA

Una de las medidas que caracterizan a una distribución es la media aritmética, que se obtiene sumando los valores x_i introducidos, y dividiendo entre el número total N

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

LA MEDIANA ESTADISTICA

Es el valor de la variable que deja el mismo número de datos antes y después que él, una vez ordenados estos. De acuerdo con esta definición el conjunto de datos menores o iguales que la mediana representarán el 50% de los datos, y los que sean mayores que la mediana representarán el otro 50% del total de datos de la muestra

LA MEDIANA ESTADISTICA

La mediana es el valor central de la variable, es decir, supuesta la muestra ordenada en orden creciente o decreciente, el valor que divide en dos partes la muestra.



LA MEDIANA ESTADISTICA

Para comprender este concepto vamos a suponer que tenemos la serie ordenada de valores (2, 5, 8, 10 y 13), la posición de la mediana sería:

$$\textit{Posición de la mediana} = \frac{n + 1}{2} = \frac{5 + 1}{2} = 3$$

LA MEDIANA ESTADISTICA

Y si la serie ordenada fuese (2, 5, 8, 10, 13 y 15), la posición de la mediana sería:

$$\text{Posición de la mediana} = \frac{n + 1}{2} = \frac{6 + 1}{2} = 3,5$$

Es decir, la posición tres y medio. Dado que es imposible destacar la posición tres y medio, es necesario promediar los dos valores de las posiciones tercera y cuarta para producir una mediana equivalente, que para el caso corresponden a $(8 + 10)/2 = 9$. Lo que nos indicaría que la mitad de los valores se encuentra por debajo del valor 9 y la otra mitad se encuentra por encima de este valor.

LA MODA



LA MODA

La medida modal nos indica el valor que más veces se repite dentro de los datos; es decir, si tenemos la serie ordenada (2, 2, 5 y 7), el valor que más veces se repite es el número 2 quien sería la moda de los datos. Es posible que en algunas ocasiones se presente dos valores con la mayor frecuencia, lo cual se denomina *Bimodal* o en otros casos más de dos valores, lo que se conoce como *multimodal*.

Medidas de tendencia central

Medidas de tendencia central

Nos permiten identificar los valores más representativos de los datos, de acuerdo a la manera como se tienden a concentrar.

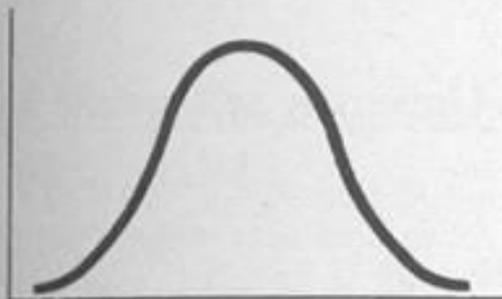
La *Media* nos indica el promedio de los datos; es decir, nos informa el valor que obtendría cada uno de los individuos si se distribuyeran los valores en partes iguales.

Medidas de tendencia central

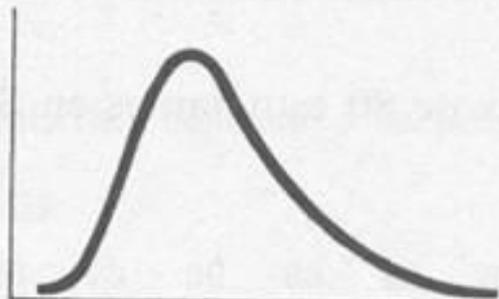
La **Mediana** por el contrario nos informa el valor que separa los datos en dos partes iguales, cada una de las cuales cuenta con el cincuenta por ciento de los datos.

Por último la **Moda** nos indica el valor que más se repite dentro de los datos

MEDIDAS DE TENDENCIA CENTRAL



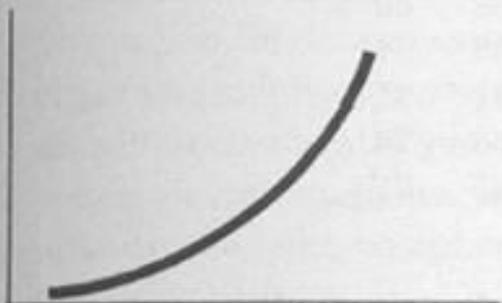
Simétrica o bien formada



Sesgada a la derecha
(sesgo positivo)



Sesgada a la izquierda
(sesgo negativo)



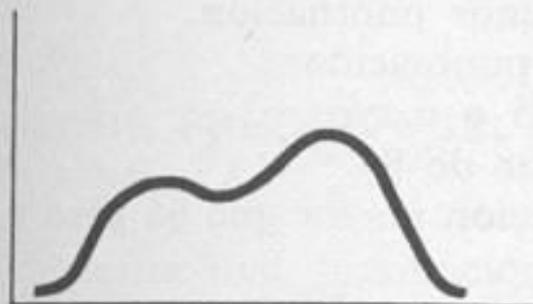
En forma de J



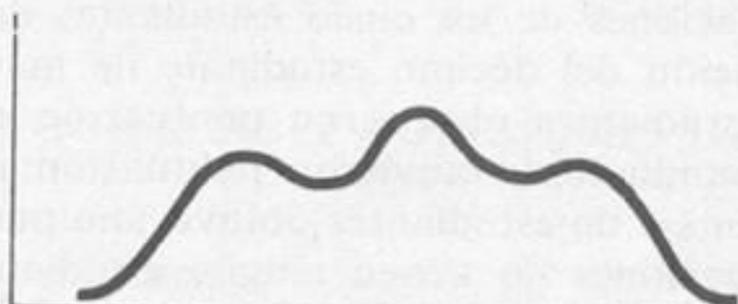
En forma de J invertida



En forma de U



Bimodal



Multimodal

MEDIDAS DE POSICIÓN

MEDIDAS DE POSICIÓN

Las medidas de posición dividen un conjunto de datos en grupos con el mismo número de individuos.

Para calcular las medidas de posición es necesario que los datos estén ordenados de menor a mayor.

MEDIDAS DE POSICIÓN

Cuartiles

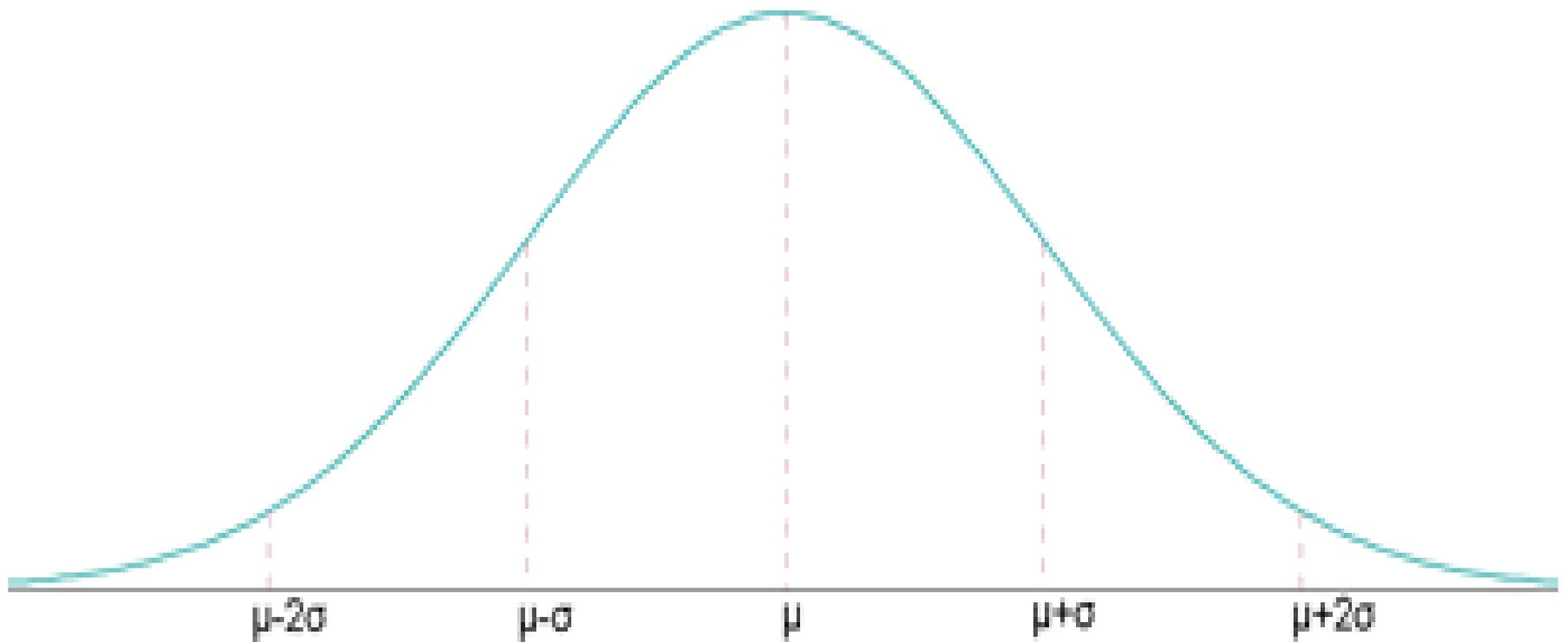
Los cuartiles son los tres valores de la variable que dividen a un conjunto de datos ordenados en cuatro partes iguales.

Q1, Q2 y Q3 determinan los valores correspondientes al 25%, al 50% y al 75% de los datos.

Q2 coincide con la mediana.

MEDIDAS DE POSICIÓN

Cuartiles



MEDIDAS DE POSICIÓN

Deciles

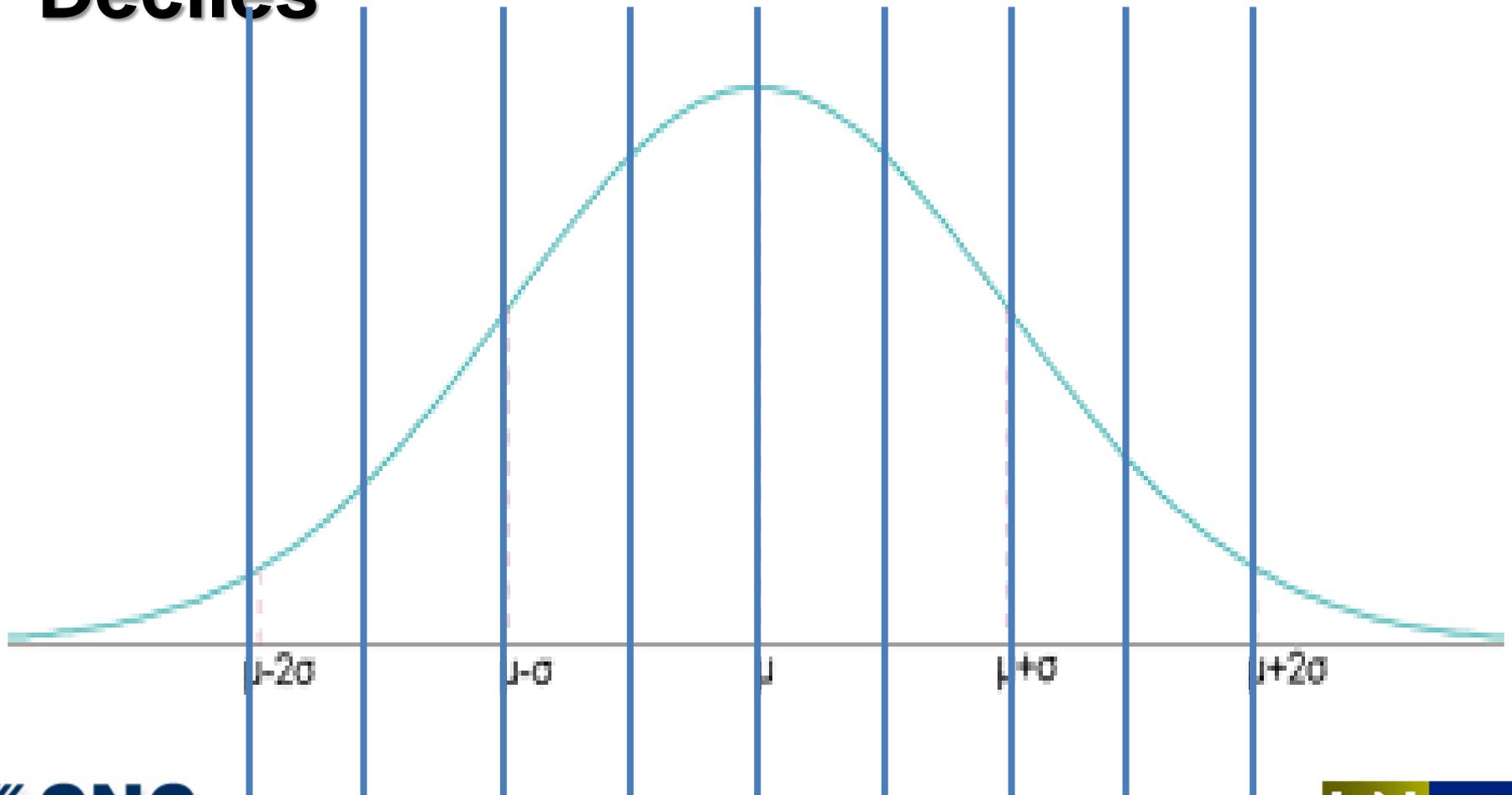
Los deciles son los nueve valores que dividen la serie de datos en diez partes iguales.

Los deciles dan los valores correspondientes al 10%, al 20%... y al 90% de los datos.

D5 coincide con la mediana.

MEDIDAS DE POSICIÓN

Deciles



MEDIDAS DE POSICIÓN

Percentiles

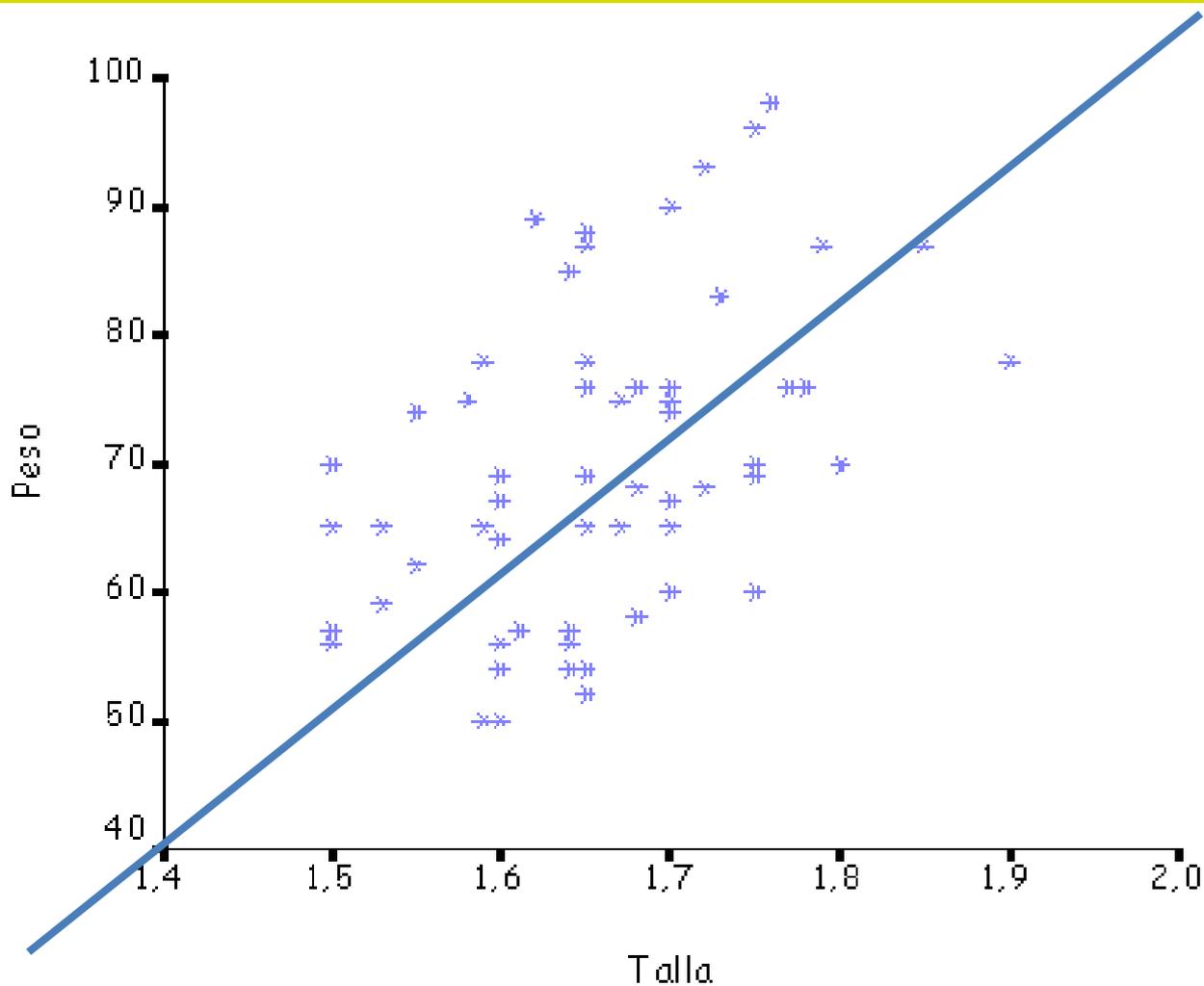
Los percentiles son los 99 valores que dividen la serie de datos en 100 partes iguales.

Los percentiles dan los valores correspondientes al 1%, al 2%... y al 99% de los datos.

P50 coincide con la mediana.

MEDIDAS DE DISPERSIÓN

MEDIDAS DE DISPERSIÓN



MEDIDAS DE DISPERSIÓN

Rango o recorrido

El rango es la diferencia entre el mayor y el menor de los datos de una distribución estadística

MEDIDAS DE DISPERSIÓN

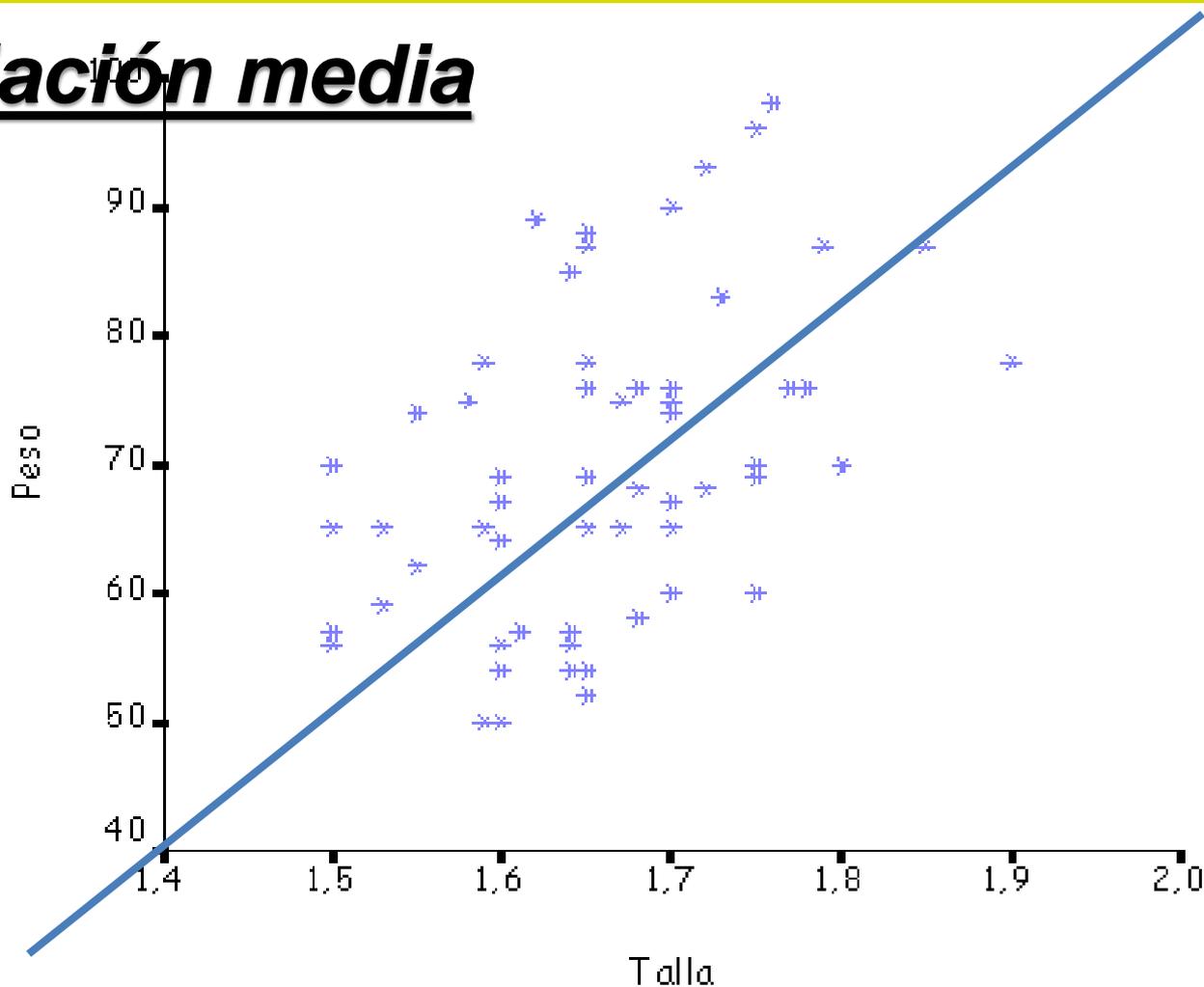
Desviación media

La desviación respecto a la media es la diferencia entre cada valor de la variable estadística y la media aritmética.

$$D_i = x - \bar{x}$$

MEDIDAS DE DISPERSIÓN

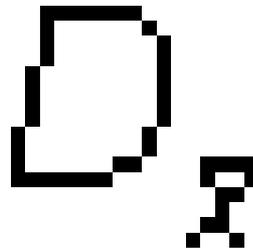
Desviación media



MEDIDAS DE DISPERSIÓN

La desviación media es la media aritmética de los valores absolutos de las desviaciones respecto a la media

La desviación media se representa por



MEDIDAS DE DISPERSIÓN

Desviación media

$$D_{\bar{x}} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \dots + |x_n - \bar{x}|}{N}$$

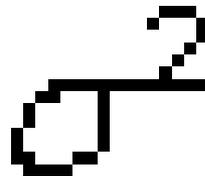
$$D_{\bar{x}} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{N}$$

MEDIDAS DE DISPERSIÓN

VARIANZA

La varianza es la media aritmética del cuadrado de las desviaciones respecto a la media de una distribución estadística.

La varianza se representa por



MEDIDAS DE DISPERSIÓN

VARIANZA

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{N}$$

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

MEDIDAS DE DISPERSIÓN

Propiedades de la varianza

1 La **varianza** será siempre un **valor positivo o cero**, en el caso de que las puntuaciones sean iguales.

2 Si a todos los **valores** de la variable se les **suma un número** la **varianza no varía**.

MEDIDAS DE DISPERSIÓN

Propiedades de la varianza

3 Si todos los **valores** de la variable se **multiplican** por un **número** la **varianza** queda **multiplicada** por el **cuadrado** de dicho **número**.

MEDIDAS DE DISPERSIÓN

Propiedades de la varianza

4 Si tenemos varias distribuciones con la misma **media** y conocemos sus respectivas **varianzas** se puede calcular la **varianza total**.

MEDIDAS DE DISPERSIÓN

Observaciones sobre la varianza

1 La **varianza**, al igual que la media, es un índice muy sensible a las puntuaciones extremas.

2 En los casos que **no se pueda hallar la media** tampoco será posible hallar la **varianza**.

MEDIDAS DE DISPERSIÓN

Observaciones sobre la varianza

3 La **varianza** no viene expresada en las mismas unidades que los datos, ya que las desviaciones están elevadas al cuadrado.

MEDIDAS DE DISPERSIÓN

Desviación estándar

La desviación estándar (σ) mide cuánto se separan los datos.

La fórmula es fácil: es la raíz cuadrada de la varianza. Así que, "¿qué es la varianza?"

MEDIDAS DE DISPERSIÓN

Desviación estándar

Varianza

La varianza (que es el cuadrado de la desviación estándar: σ^2) se define así:

Es la media de las diferencias con la media elevadas al cuadrado.

MEDIDAS DE DISPERSIÓN

Desviación estándar

En otras palabras, sigue estos pasos:

Calcula la media (el promedio de los números)

MEDIDAS DE DISPERSIÓN

Desviación estándar

Ahora, por cada número resta la media y eleva el resultado al cuadrado (la diferencia elevada al cuadrado).

MEDIDAS DE DISPERSIÓN

Desviación estándar

Ahora calcula la media de esas diferencias al cuadrado. (¿Por qué al cuadrado?)

MEDIDAS DE DISPERSIÓN

Desviación estándar

(¿Por qué al cuadrado?)

Elevar cada diferencia al cuadrado hace que todos los números sean positivos (para evitar que los números negativos reduzcan la varianza)

MEDIDAS DE DISPERSIÓN

Desviación estándar

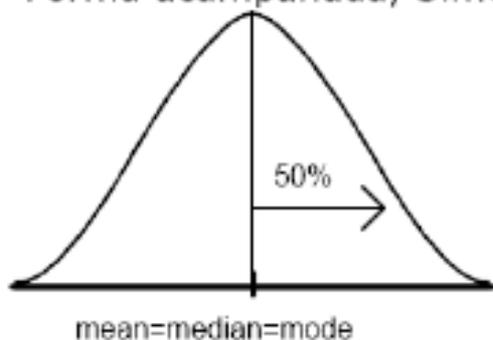
Y también hacen que las diferencias grandes se destaquen. Por ejemplo $100^2=10,000$ es mucho más grande que $50^2=2,500$.

Pero elevarlas al cuadrado hace que la respuesta sea muy grande, así que lo deshacemos (con la raíz cuadrada) y así la desviación estándar es mucho más útil

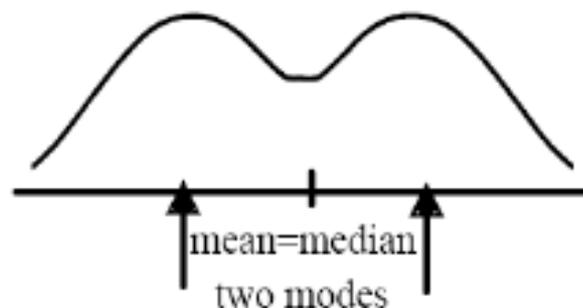
MEDIDAS DE DISPERSIÓN

¿Qué medida de tendencia utilizar?

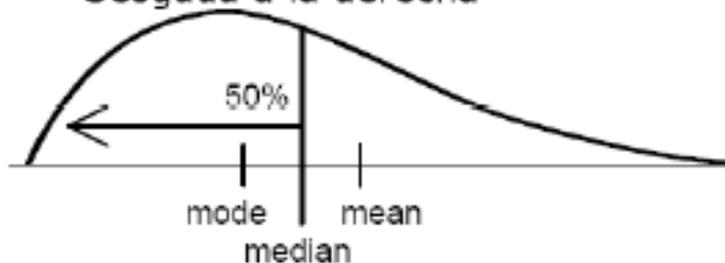
Forma acampanada, Simétrica



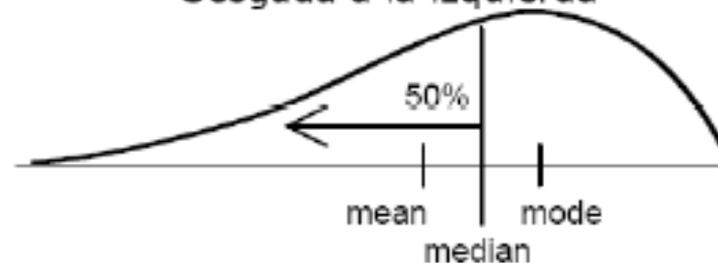
Bimodal



Sesgada a la derecha

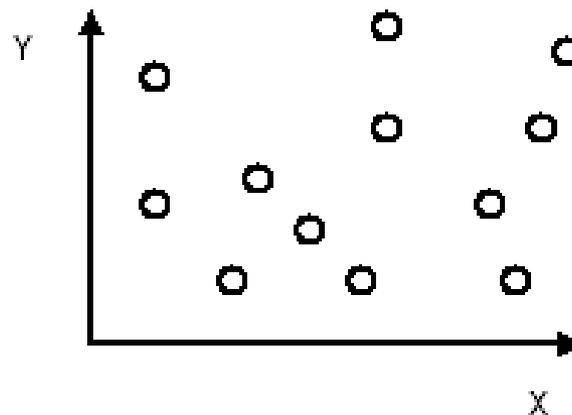
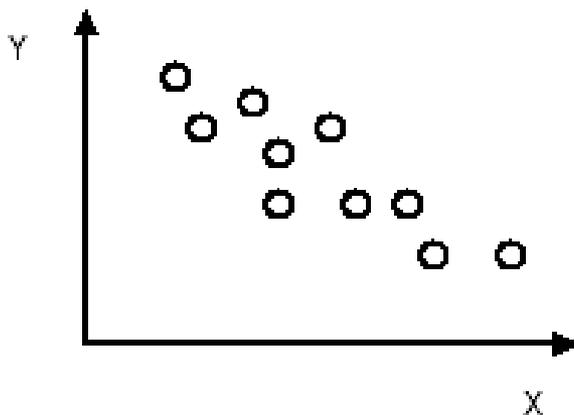


Sesgada a la izquierda



MEDIDAS DE DISPERSIÓN

RESUMEN



GRACIAS