

¿Qué es la ESTADÍSTICA?

Es la ciencia que trata de cómo obtener información a partir de una observación incompleta de la realidad.

Herramienta principal de la ESTADÍSTICA

Los DATOS ← *Valores numéricos a partir de los cuales se extraen las conclusiones*

Los datos se pueden obtener de:

1. *Experimentos*
2. *Encuestas*
3. *Observaciones*

**La INFERENCIA de conclusiones generales a partir de DATOS
es siempre INCIERTA**



**Uno de los objetivos principales de la ESTADÍSTICA es medir
esa incertidumbre**

En **ESTADÍSTICA** utilizamos el método **INDUCTIVO** para **INFERIR conclusiones**

De lo PARTICULAR → A lo GENERAL

¿Qué es particular? *LA MUESTRA*

¿Qué es general? *LA POBLACIÓN*

Inferencia DEDUCTIVA

general → particular

Todos los loros son pájaros

Este animal es un loro

Entonces concluimos que este animal es un pájaro

Inferencia INDUCTIVA

particular → general

Observamos 1000 loros y todos ellos hablan

Entonces concluimos que todos o **CASI TODOS** los loros hablan

¿Qué es la población?

Es el conjunto de todas las medidas que nos interesan

1. *Preferencias de cada votante (en España)*
2. *Producción eléctrica anual en todas las centrales térmicas europeas*

¿Qué es la muestra?

Es el conjunto de todas las medidas que se observan.

1. *Preferencias de 1000 votantes entrevistados en una encuesta*
2. *Producción eléctrica anual en 100 centrales térmicas elegidas al azar*
3. *Producción eléctrica anual en cada central térmica de Francia*

La muestra es un subconjunto de la población

Objetivo: extraer conclusiones sobre la población basándonos en la información que proporciona la muestra

¿Por qué hablamos de INCERTIDUMBRE en las CONCLUSIONES?

Porque existen muchas muestras posibles, pero nosotros sólo OBSERVAMOS UNA !!!

¿Qué nos interesa de la ESTADÍSTICA APLICADA?

- El análisis descriptivo y exploratorio de datos
- La obtención de muestras y el diseño de experimentos.
- La inferencia estadística para obtener conclusiones generales y su grado de fiabilidad.

Métodos estadísticos

1. Descriptivos - Resumen

- (a) Gráficos
- (b) Numéricos

2. Inferencia

- (a) Estimaciones puntuales
- (b) Intervalos de confianza
- (c) Contrastes de hipótesis

Consejos sabios:

1. Intenta comprender que dicen los datos en un determinado contexto. Todos los métodos que aprenderás sólo son herramientas que ayudan a comprender los datos.
2. Deja que una calculadora o un ordenador haga la mayor parte de los cálculos y los gráficos, de manera que puedas concentrarte en qué haces y por qué lo haces.
3. Céntrate en las grandes ideas de la estadística, no te centres sólo en las reglas y fórmulas.

David S. Moore

ESTADÍSTICA DESCRIPTIVA

Tipos de datos

- Variables cualitativas o categóricas
- Variables cuantitativas discretas
- Variables cuantitativas continuas

Descripción gráfica:

- **Diagrama de barras y pictograma** ← Variables cualitativas o discretas
 - **Diagrama de tallo y hojas**
 - **Histograma**
 - **Diagrama de cajas**
 - **Gráfico temporal**
- ← variables cuantitativas

Descripción numérica:

- **Medidas de posición** ← *media, mediana, media recortada, moda.*
- **Medidas de dispersión** ← $\left\{ \begin{array}{l} \text{recorrido, varianza y desviación típica} \\ \text{MEDA y rango intercuartílico} \end{array} \right.$
- **Otras medidas de forma:** ← $\left\{ \begin{array}{l} \text{coeficiente de asimetría} \\ \text{coeficiente de kurtosis (o apuntamiento)} \end{array} \right.$

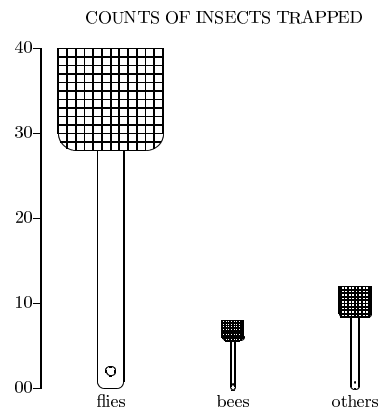
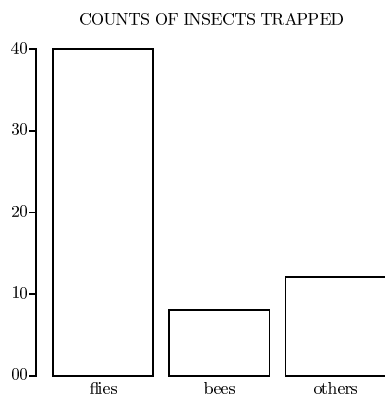
Diagrama de barras

Se representan mediante barras las frecuencias (o porcentajes) de cada suceso

Se utiliza para: **Variables cualitativas y cuantitativas discretas**

En las discretas, los sucesos se ordenan de menor a mayor

En las cualitativas, sólo cuando exista un orden natural



Pictograma

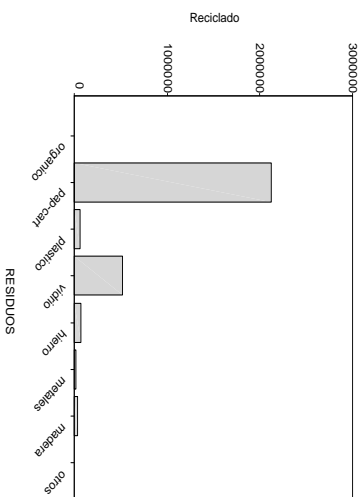
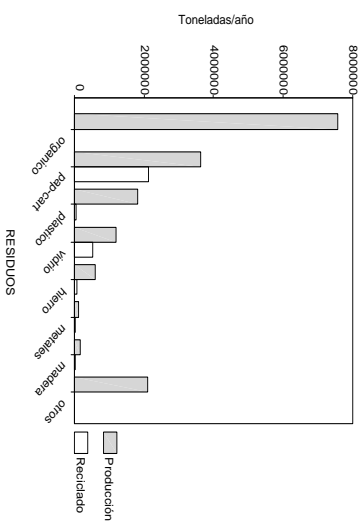
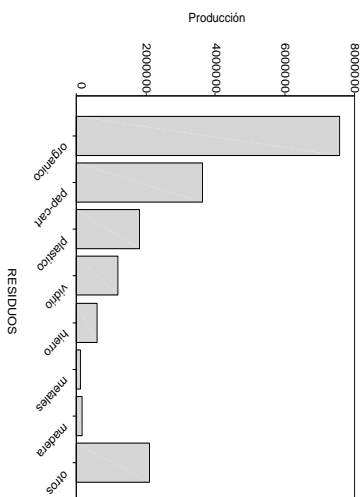
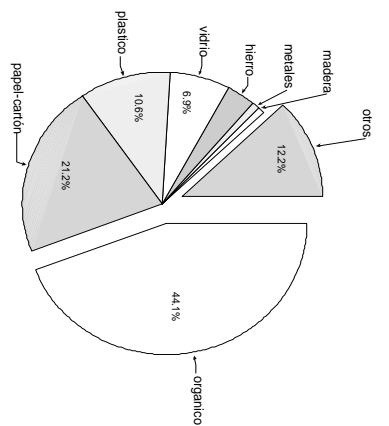
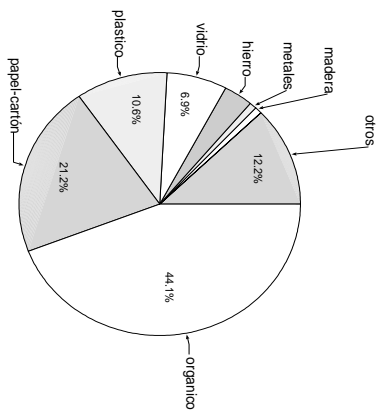
Las frecuencias se representan mediante la parte proporcional de un círculo

Se utiliza para: **Variables cualitativas y cuantitativas discretas**

Producción actual de residuos	Tm / año	%	Reciclado actual	%	Objetivos de reutilización / reciclaje	
					2003	2006
Orgánicos	7.567.397	44			doméstico	80%
Papel-cartón	3.637.704	21,1	2.125.000	41	envases	90%
					comercio	90%
Plástico	1.818.852	10,5	59.000	5	agua	25%
Vidrio	1.190.240	6,9	521.492	37	refrescos	35%
					cerveza	70%
					vinos	15%
Hierro	589.109	3,4	72.250	23	acero	90%
Metales no férricos	116.791	0,6	21.067	19	aluminio	90%
Madera	164.882	0,9	34.200	9,6		50%
Citros	2.090.220	12,1	2	2	tetra-brick	50%
					aceite	80%
TOTAL	17.175.186					

Tratamiento actual de los residuos urbanos, destino final

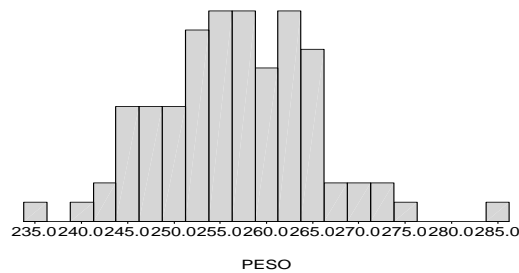
Sistema	Residuos tratados (Tm / año)	%
Vertido controlado	9.989.386	58,2
Vertido incontrolado	2.101.250	12,2
Reciclaje	1.985.040	11,6
Compostaje	2.384.162	13,9
Incineración (con recuperación de energía)	627.949	3,7
Incineración (sin recuperación de energía)	77.399	0,5
TOTALES	17.175.186	100,0



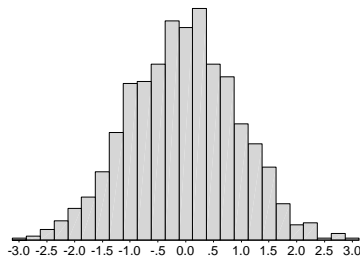
Histograma

Se representan mediante cajas las frecuencias de los resultados agrupados en intervalos.

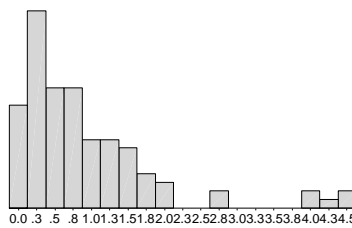
Se utiliza para: Variables cuantitativas continuas



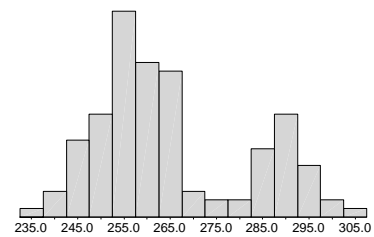
Describe la forma en que se DISTRIBUYEN los datos



normal



asimétrico



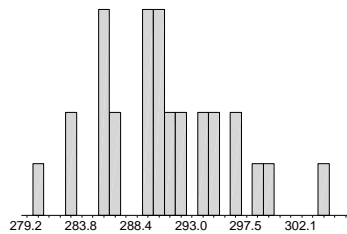
bimodal

Las frecuencias se pueden representar en:

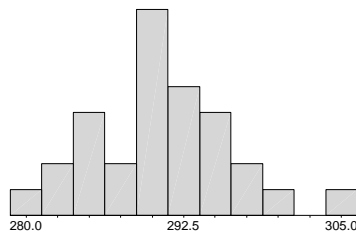
- *valores absolutos en la altura*
- *valores relativos en la altura*
- **valores relativos en el área**
- *valores acumulados ...*

Consejos:

1. Usar intervalos de la misma longitud
2. Los intervalos no pueden solaparse
3. Cada observación sólo puede pertenecer a un intervalo
4. Todos los datos deben pertenecer a algún intervalo
5. La forma del histograma depende de la amplitud del intervalo que se elija. Regla: amplitud correspondiente a \sqrt{n} intervalos cubriendo todo el rango de valores



30 clases



automático

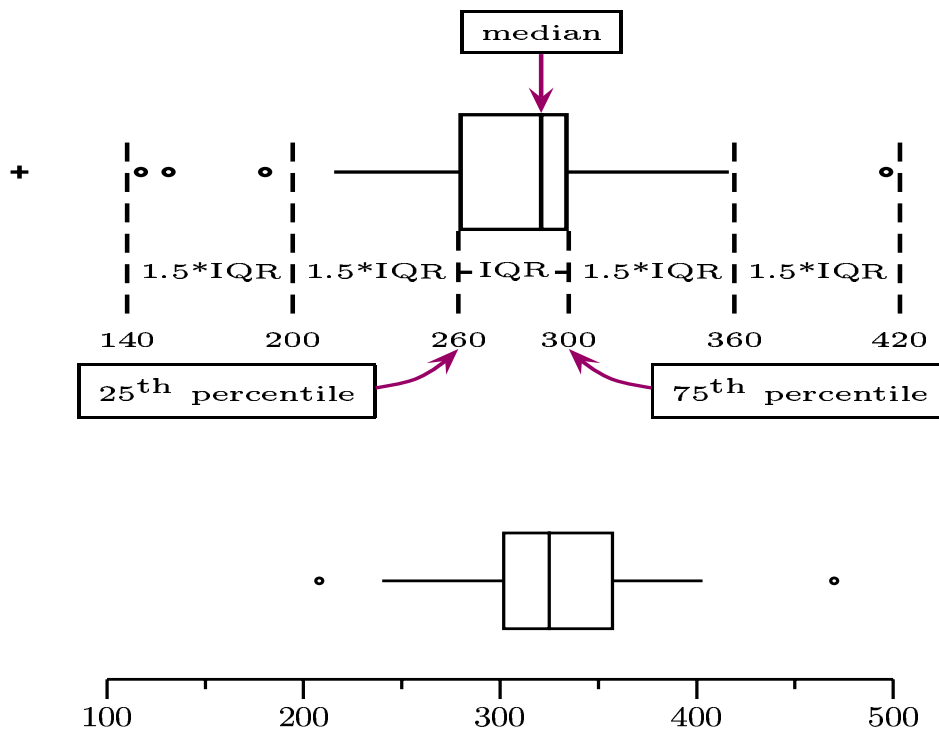


1 clase

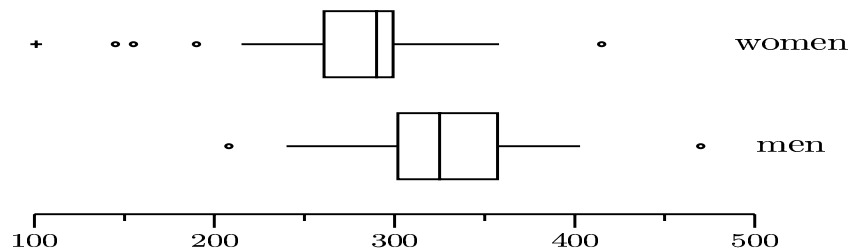
Diagrama de caja (boxplot)

Pasos para construir el boxplot:

1. Ordenar la muestra
2. Calcular la mediana, el primer y el tercer cuartil
3. Calcular el rango y el rango intercuartílico (IQR)



Boxplot múltiple en un gráfico

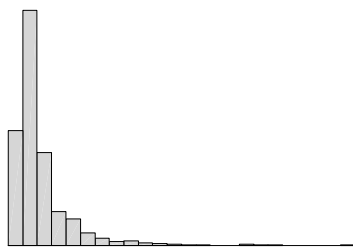


Transformación de datos

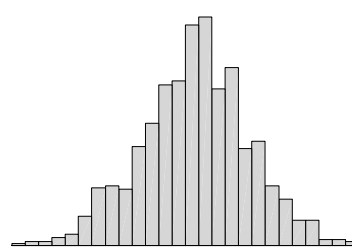
Se aplican transformaciones a los datos para **SIMPLIFICAR** y **FACILITAR** su estudio.

Consejos prácticos:

1. **Hacer cambios de unidades para facilitar la comparación de muestras**
2. **Buscar transformaciones que den una distribución SIMÉTRICA de los datos**
3. **Tomar logaritmos de los datos hace más simétricos conjuntos de datos referidos a tamaños.**



x



$y = \log x$

Análisis de dos conjuntos de datos

COMPARACIÓN: Los datos corresponden a la misma medición en dos tipos de individuos

$$\boxed{(x_1^1, \dots, x_n^1) \quad (x_1^2, \dots, x_m^2)}$$

¿Son los dos tipos distintos según esta característica?

Métodos gráficos: - *Diagrama de tallo y hojas*
- *Boxplot multiples*

RELACIÓN: Los datos corresponden a distintas mediciones en los mismos individuos

$$\boxed{(x_1, y_1), \dots, (x_n, y_n)}$$

¿Influyen las medidas de una variable en las medidas de la otra?

Métodos gráficos: - *Histograma tridimensional*
- *Diagrama de dispersión*
- *Recta de mínimos cuadrados*

Resúmenes numéricos: - *Covarianza muestral*
- *Coefficiente de correlación muestral*