

Sociología

Estadística MUY aplicada

Teresa Villagarcía



¿Para qué sirve la estadística?

PARA SUFRIR MUCHO



Estudiaremos

- Formas eficaces de obtener información de los datos: medir.
- Muchos análisis gráficos
- Usaremos ordenadores: Statgraphics
- ***CASOS REALES QUE habrá que analizar por equipos***
- Software disponible en el mercado: Statgraphics, S, R, SPSS, SAS....

Índice de la primera parte

- Tipos de datos
- Análisis gráficos
- Medidas resumen numéricas
- Tablas




Tipos de datos:

- Datos cualitativos
- Datos cuantitativos
 - Datos transversales
 - Datos temporales



Datos Cualitativos: (No son números)

- Sexo de una persona
 - Nacionalidad
 - Estado civil
 - Hoteles: céntrico o extrarradio
 - ¿Datos de todo tipo?.....
- 


Datos Cuantitativos

- Altura de una persona
- Peso
- Ingresos
- Hotel: Número de habitaciones.
- ¿Datos de todo tipo?.....

Datos transversales:

Se obtienen de muchos individuos en el mismo instante de tiempo o **en tiempos equivalentes.**

Típico de encuestas

- Número de trabajadores en 45 hoteles
 - Altura de 200 personas
 - Ingresos de 3000 familias
 - ¿Datos de todo tipo?.....
- 

Datos temporales:

Evolución de **una** variable en el tiempo

- Evolución de la inflación en España desde 1980
- Evolución de las ventas de la empresa desde hace 5 años
- Evolución de la calidad percibida por nuestros clientes.
- ¿....?



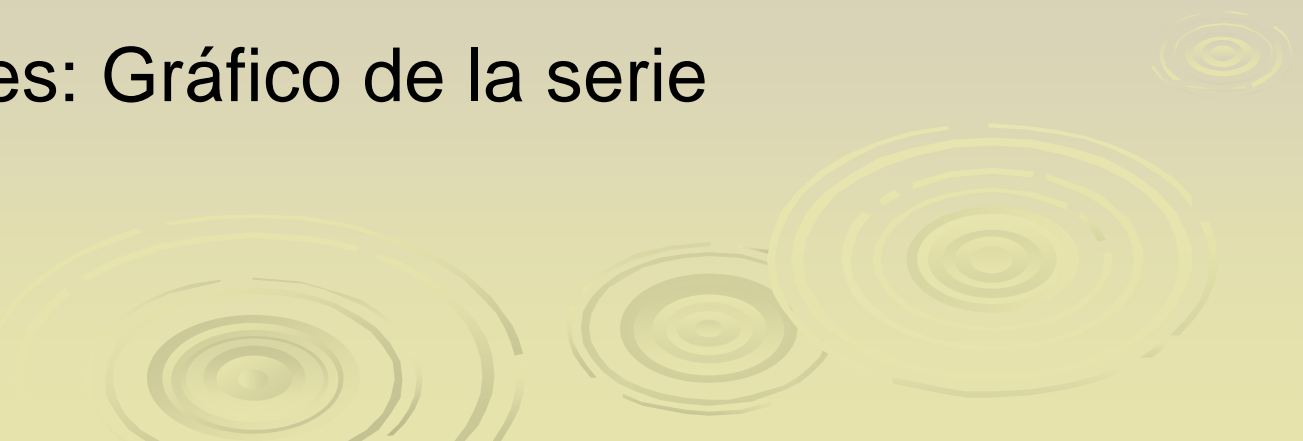
Vamos a estudiar cómo se representan gráficamente

➤ Datos cuantitativos

- Transversales: Histograma, Box plot.
- Temporales: Gráfico de la serie

➤ Datos cualitativos

- Transversales: Tarta, Barras
- Análisis de Pareto
- Temporales: Gráfico de la serie



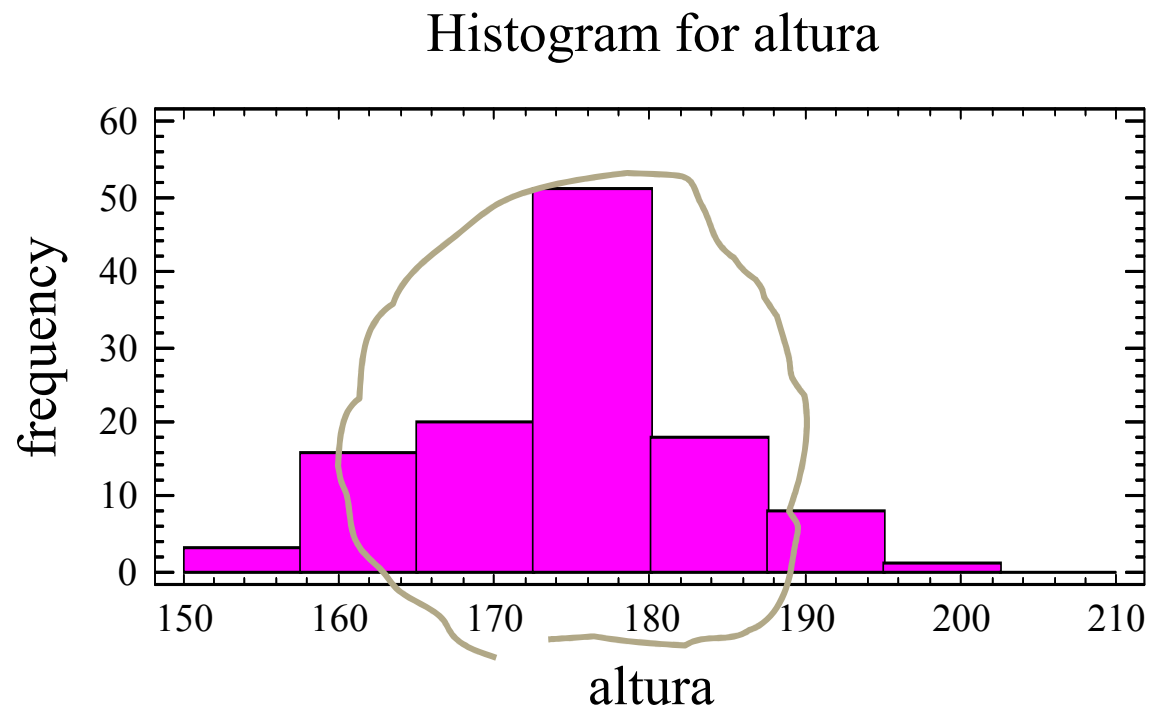
Datos cuantitativos transversales: Alturas de 117 alumnos míos:

HORRRRRRRRR

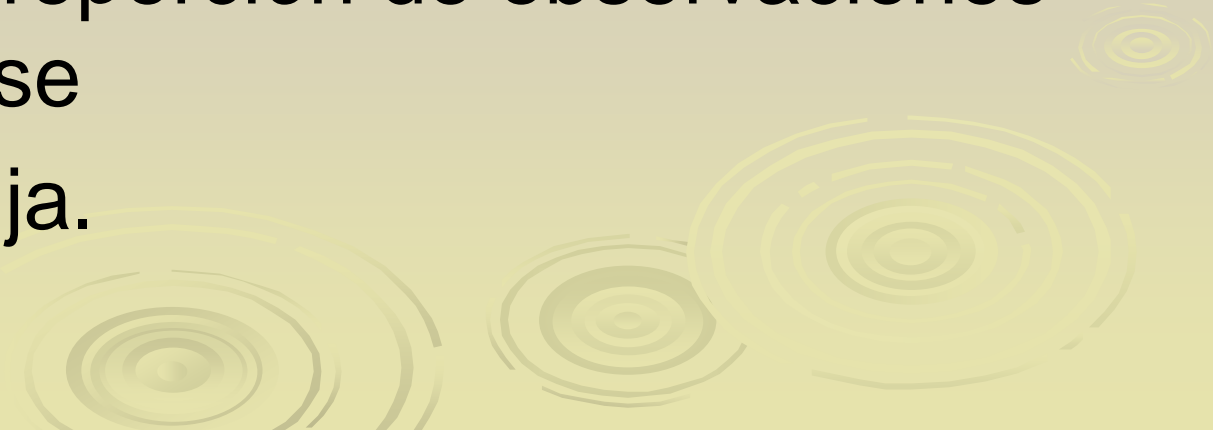
180	178	192	180	162	183	168	160	182	172	162	175
163	182	179	174	182	178	159	157	155	171	176	171
189	180	182	179	178	175	170	171	172	178	178	173
171	175	173	170	171	174	190	178	163	170	180	189
181	175	167	167	173	172	175	175	165	180	173	165
163	169	162	169	178	163	184	172	169	176	164	178
187	181	199	190	169	179	184	187	175	176	179	161
178	178	169	179	175	177	169	175	178	177	184	180
175	175	184	156	173	192	186	180	169	171	172	180
193	182	185	177	170	173	192	166	173			

Hacemos un histograma

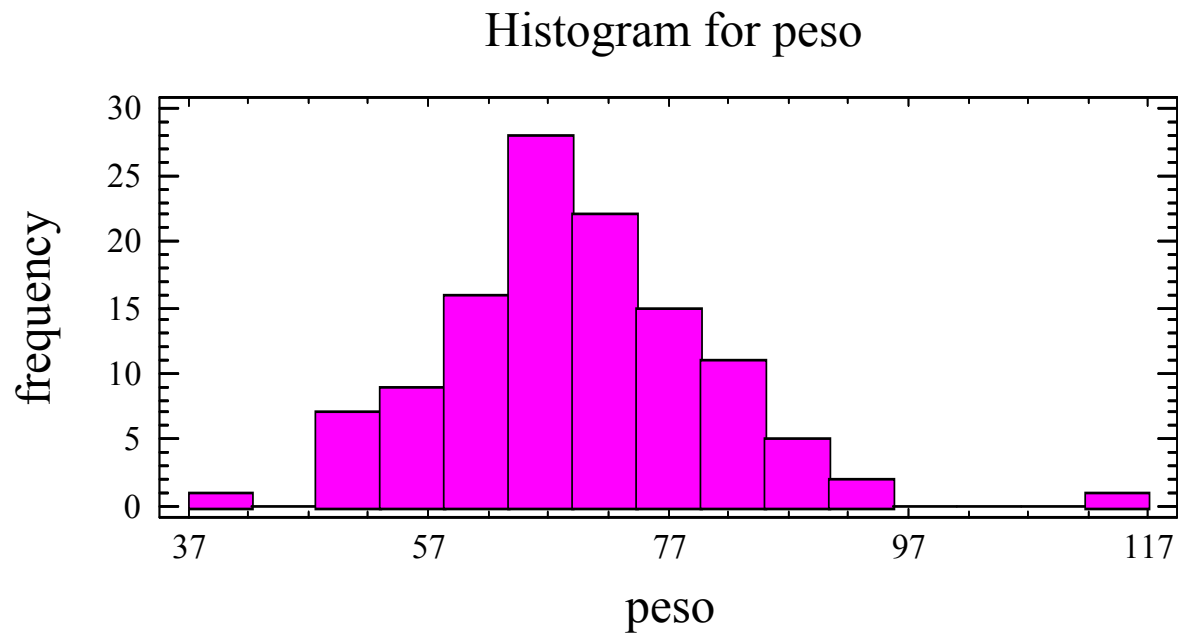
(Bueeno lo hace el ordenador que para eso está)



¿Cómo ha hecho el ordenador el histograma?

1. Calcula el valor mínimo y máximo observado
 2. Divide ese rango en una serie de clases
 3. Cuenta cuantas personas hay en cada una de esas clases
 4. Calcula la proporción de observaciones en cada clase
 5. Y.....lo dibuja.
- 

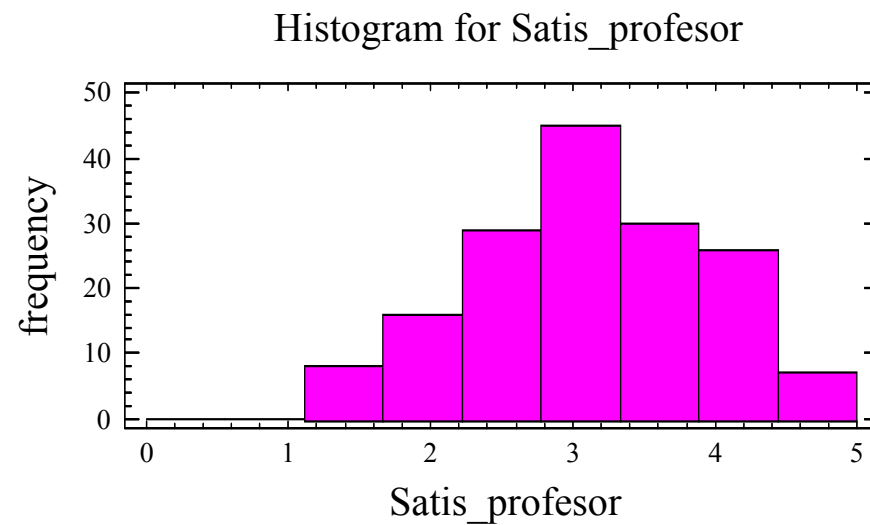
Aprender a ver histogramas:



Aprender a ver histogramas:

Satisfacción con los profesores (de 1 a 5).

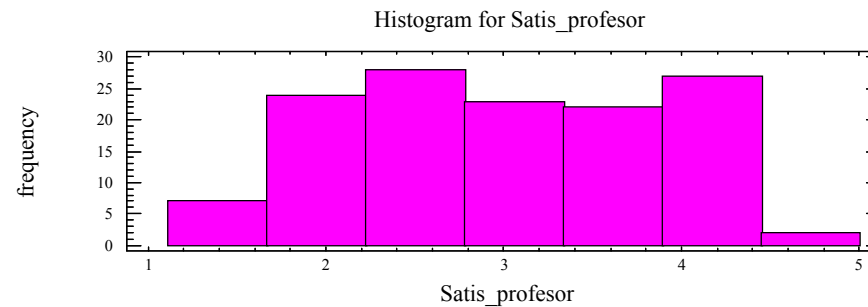
Encuesta en Ingeniería técnica mecánica



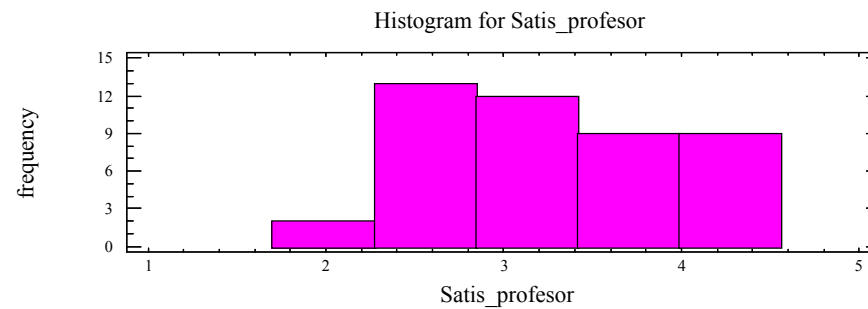
Aprender a ver histogramas:

Satisfacción con los profesores.
Encuesta en Ingeniería técnica mecánica

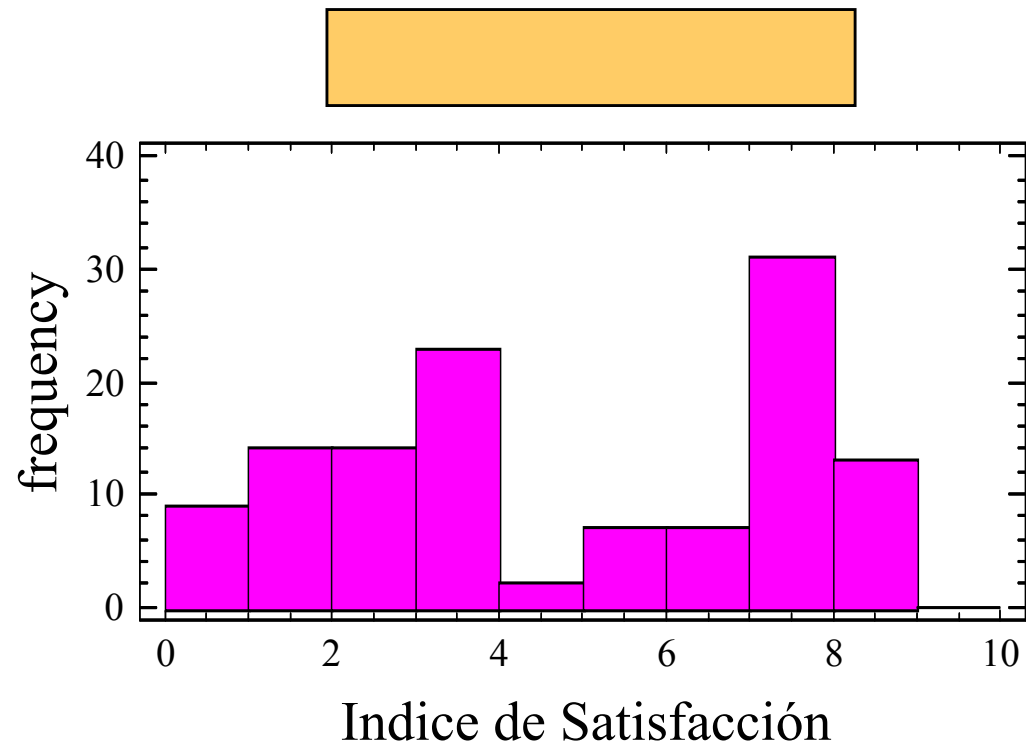
**Primer
Curso**



**Tercer
Curso**

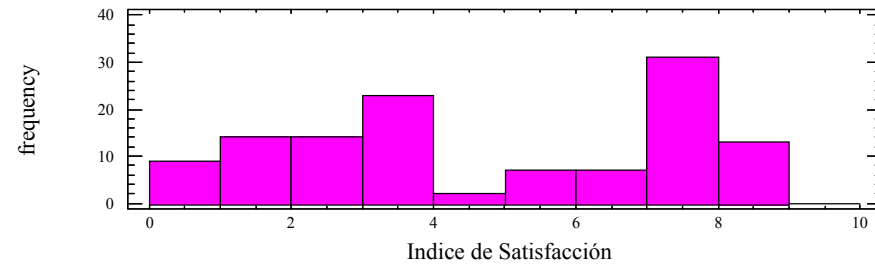


Aprender a ver histogramas: Distribuciones bimodales



Aprender a ver histogramas: Índice de satisfacción de dos empresas

¿Qué hacer?



¿Qué hacer?



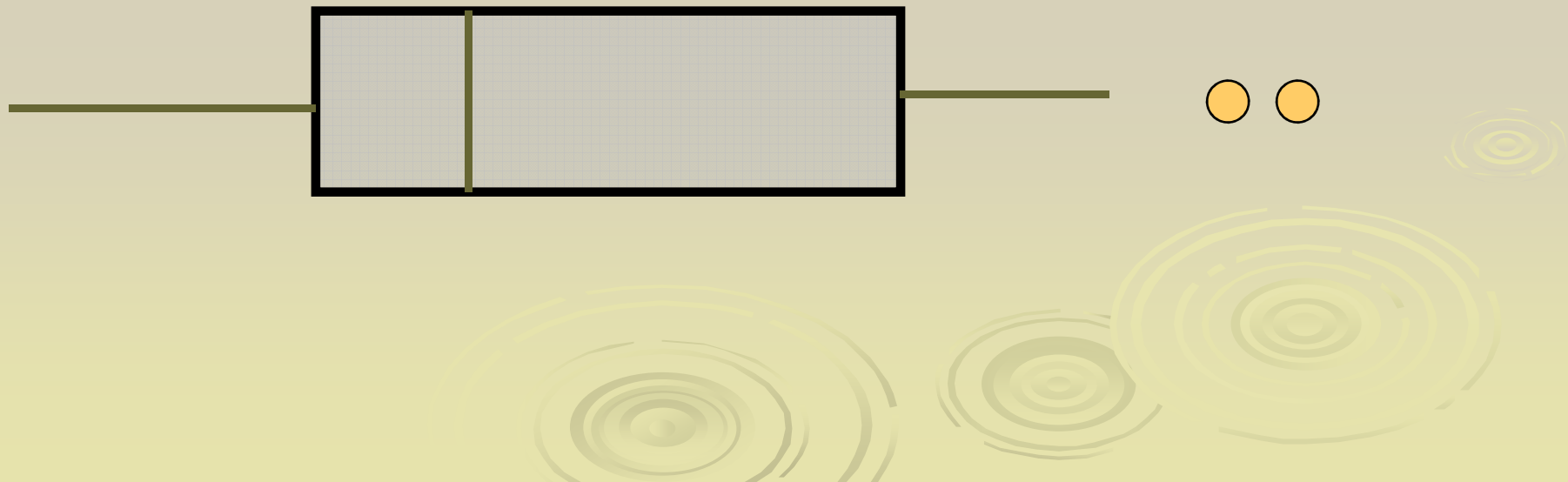
Diagrama de tallo y hojas

- Para hacer a mano en un momento de necesidad

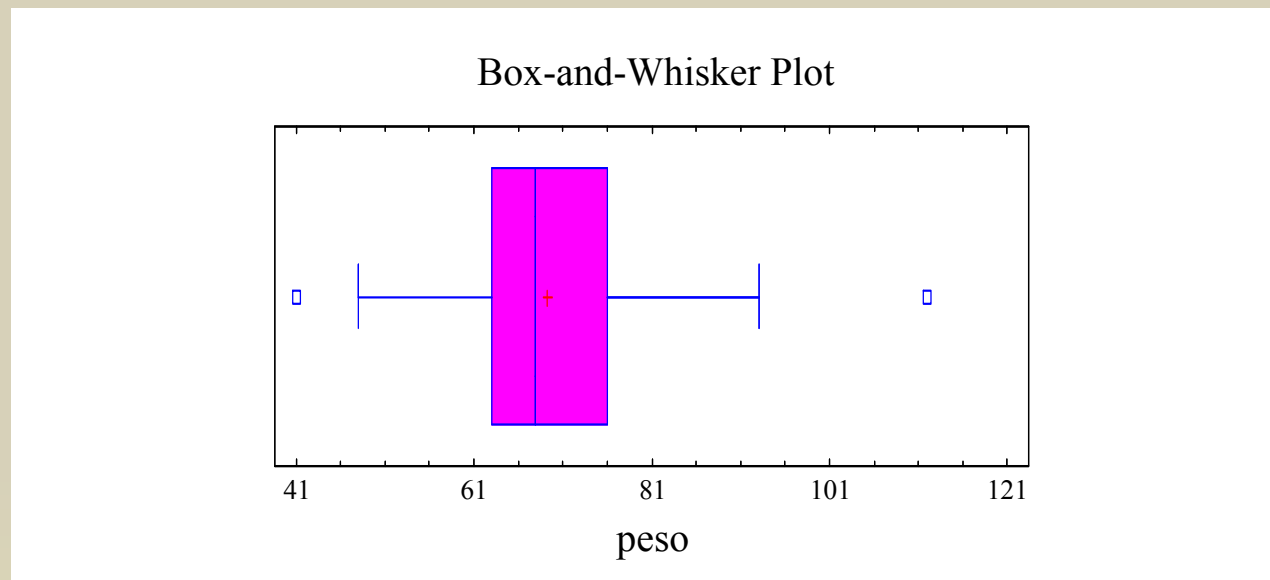


Un gráfico fantástico: El diagrama de caja: Box-plot

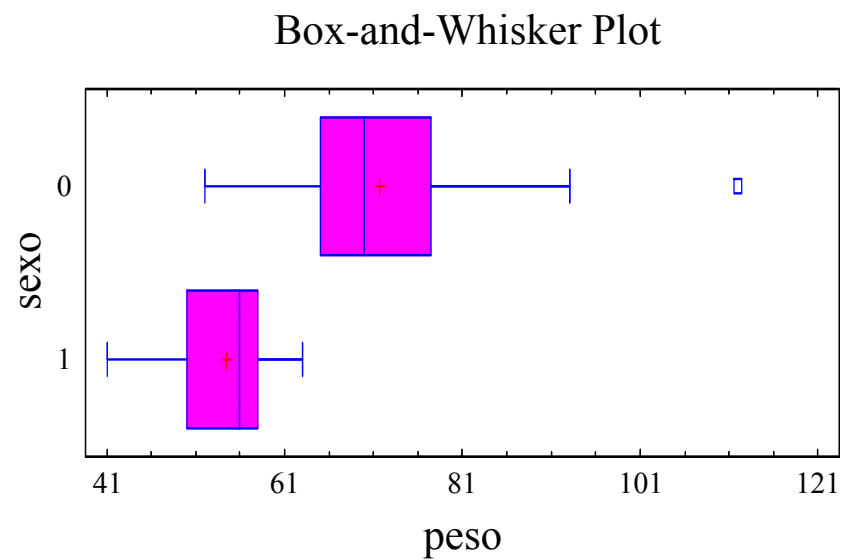
1. Se construye una caja que contiene el 50% central de los datos
2. Se dibuja la mediana
3. Se dibujan dos líneas hasta los puntos de corte (que calcula el ordenador)
4. Se dibujan los puntos que quedan fuera: Puntos atípicos



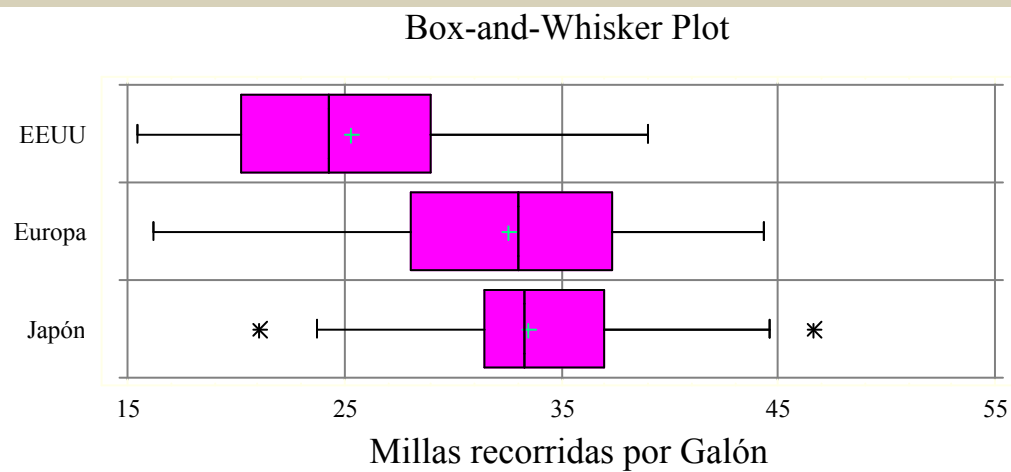
Interpretar Box-plots



Interpretar Box-plots por sexo

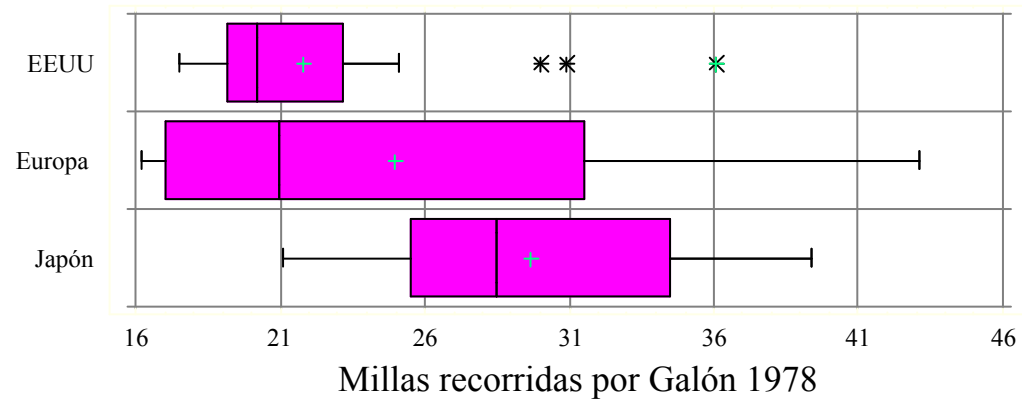


Consumo de los automóviles vendidos en Estados Unidos desde 1978 a 1981 por origen

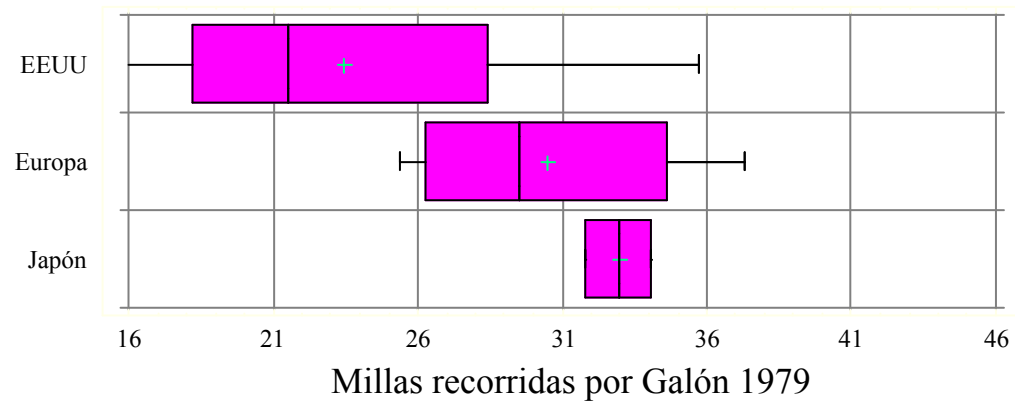


Consumo de los automóviles vendidos en Estados Unidos por origen separando por años.

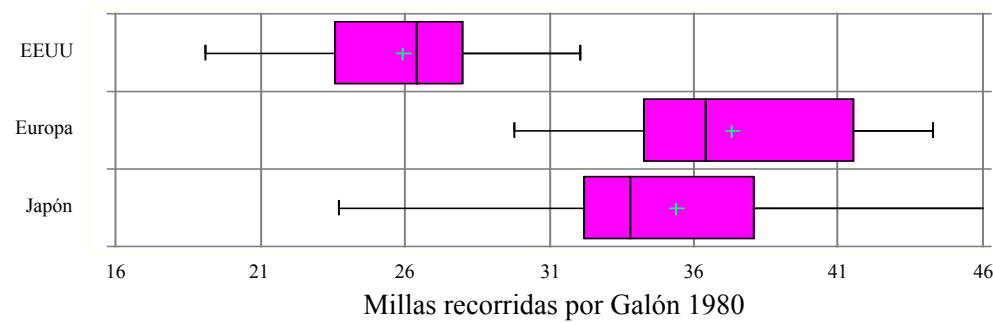
Box-and-Whisker Plot



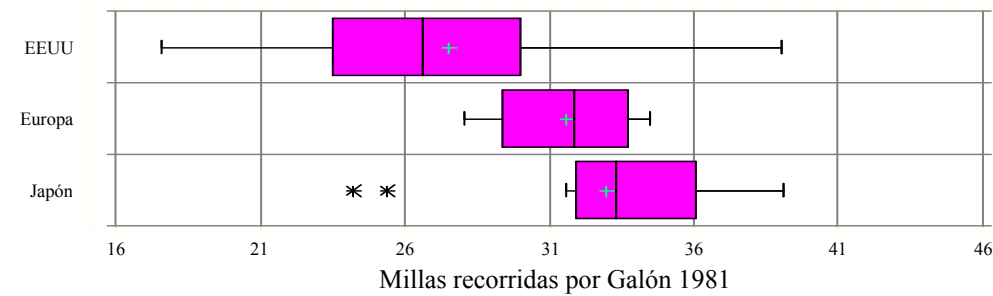
Box-and-Whisker Plot



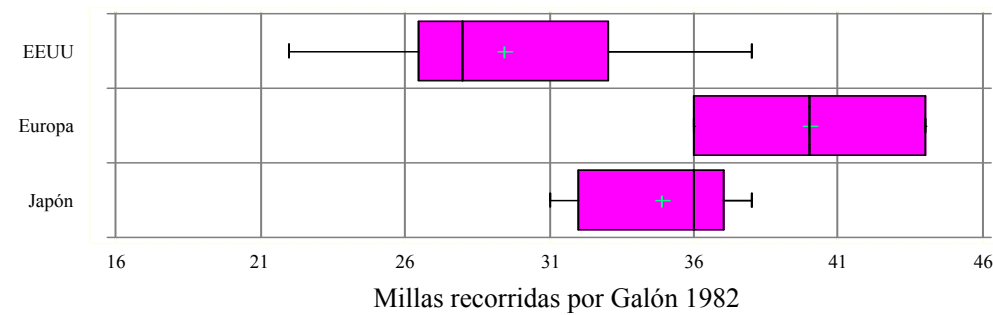
Box-and-Whisker Plot



Box-and-Whisker Plot

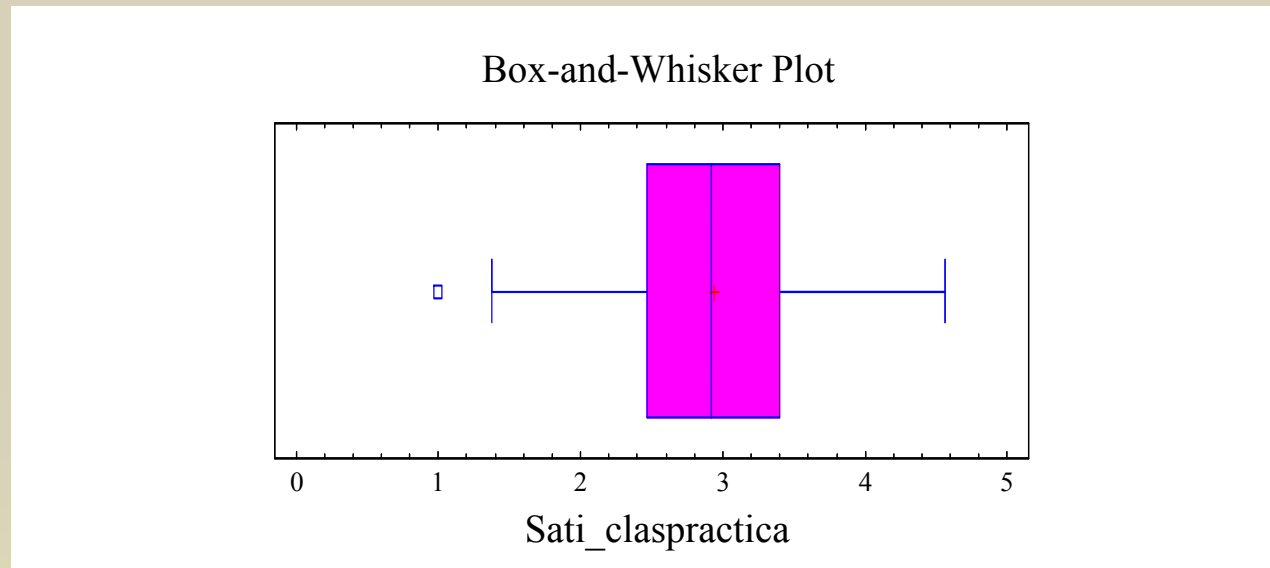


Box-and-Whisker Plot



Interpretar Box-plots:

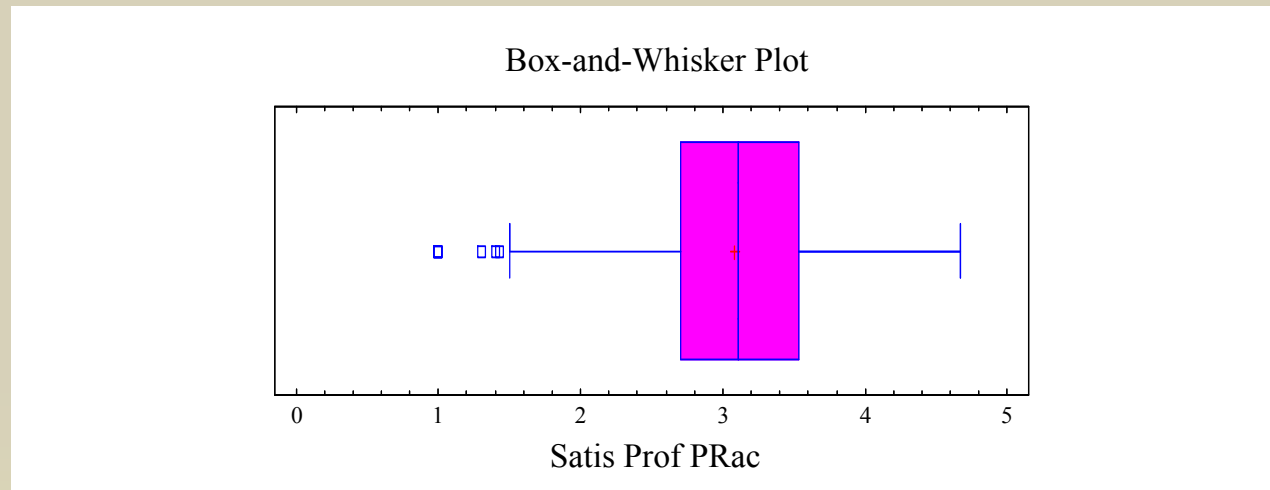
Satisfacción prácticas: Ing. Técnica mecánica



Hay que identificar esas asignaturas complicadas.

Interpretar Box-plots:

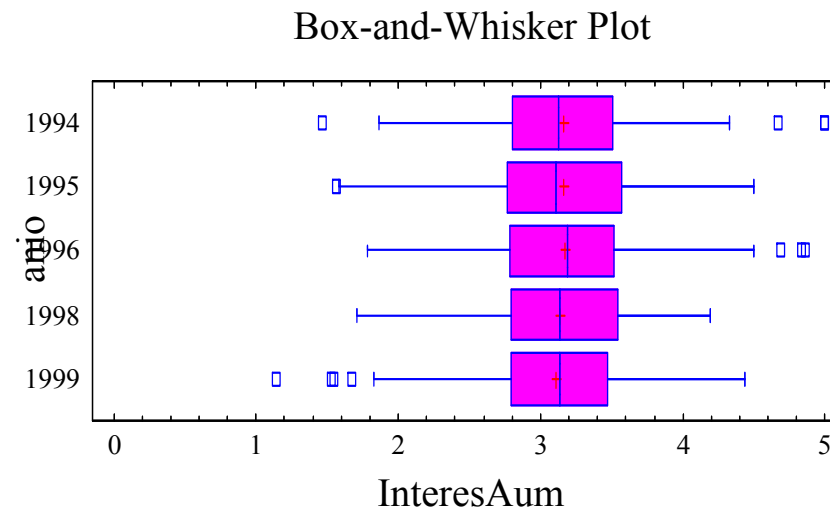
Satisfacción prácticas: Ing. Industrial



Hay que identificar esas asignaturas complicadas.

Interpretar Box-plots:

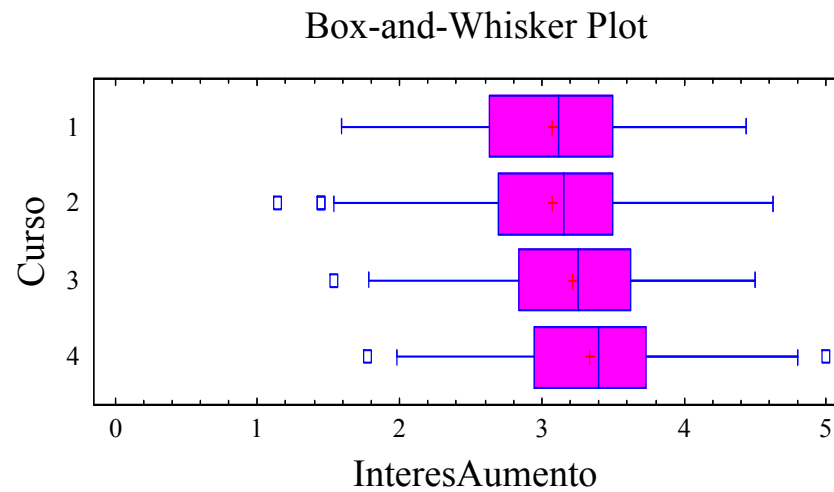
Aumenta el interés: Ing. Industrial



Hay que identificar esas asignaturas complicadas

Interpretar Box-plots:

Aumenta el interés:Admón empresas



Hay que identificar esas asignaturas complicadas

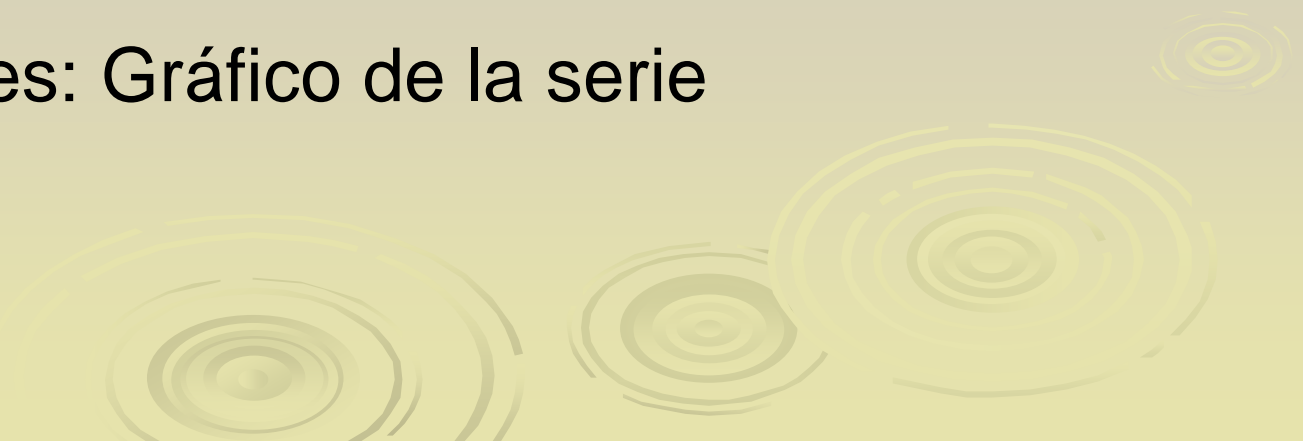
Vamos a estudiar cómo se representan gráficamente

➤ Datos cuantitativos

- Transversales: Histograma, Box plot.
- Temporales: Gráfico de la serie

➤ Datos cualitativos

- Transversales: Tarta, Barras
- Análisis de Pareto
- Temporales: Gráfico de la serie



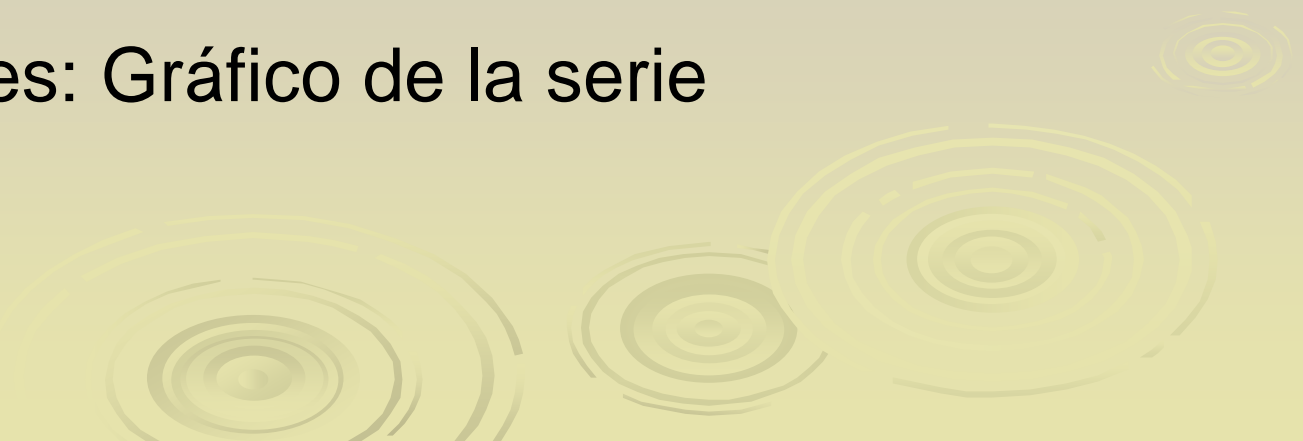
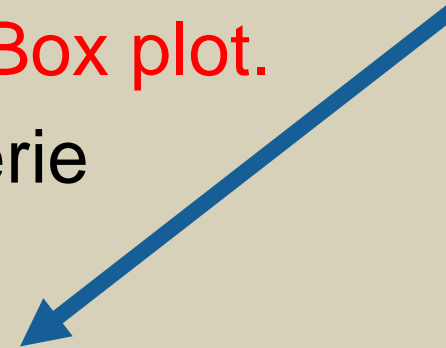
Vamos a estudiar cómo se representan gráficamente

➤ Datos cuantitativos

- Transversales: Histograma, Box plot.
- Temporales: Gráfico de la serie

➤ Datos cualitativos

- Transversales: Tarta, Barras
- Análisis de Pareto
- Temporales: Gráfico de la serie



Datos cualitativos: Bar chart

Encuesta de satisfacción de clientes:

- Muy satisfechos
- Bastante satisfechos
- Medianamente satisfechos
- Descontentos

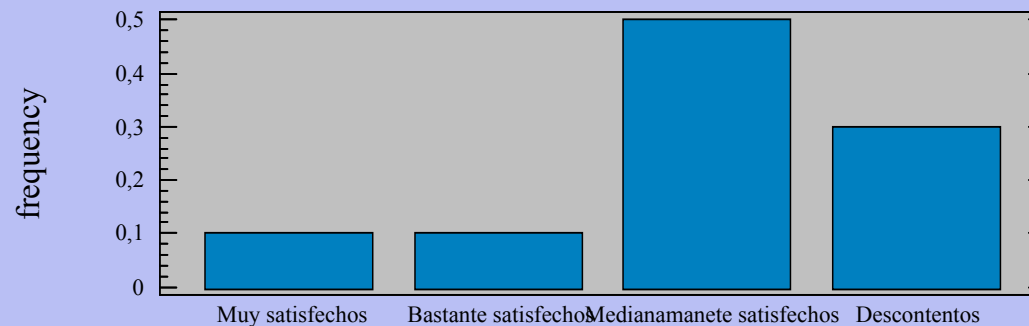


Diagrama de barras

¿Diferencia con el histograma?

Datos cualitativos: Bar chart

Nivel de estudios de 75 personas:

1 Analfabeto

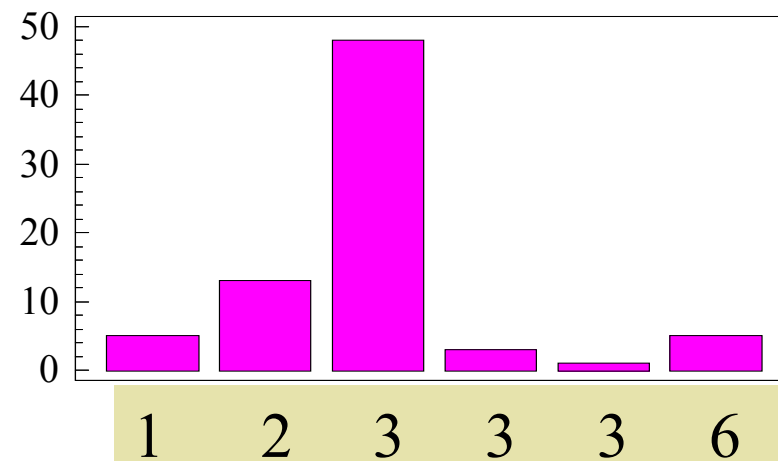
2 Sin estudios

3 Primarios

4 Secundarios

5 Medios

6 Estudios superiores

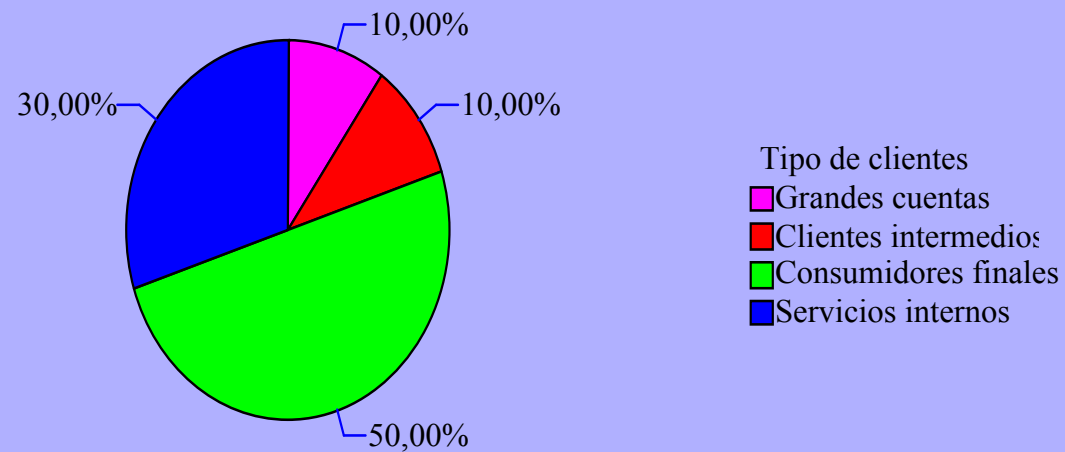


Datos cualitativos transversales : Pie chart

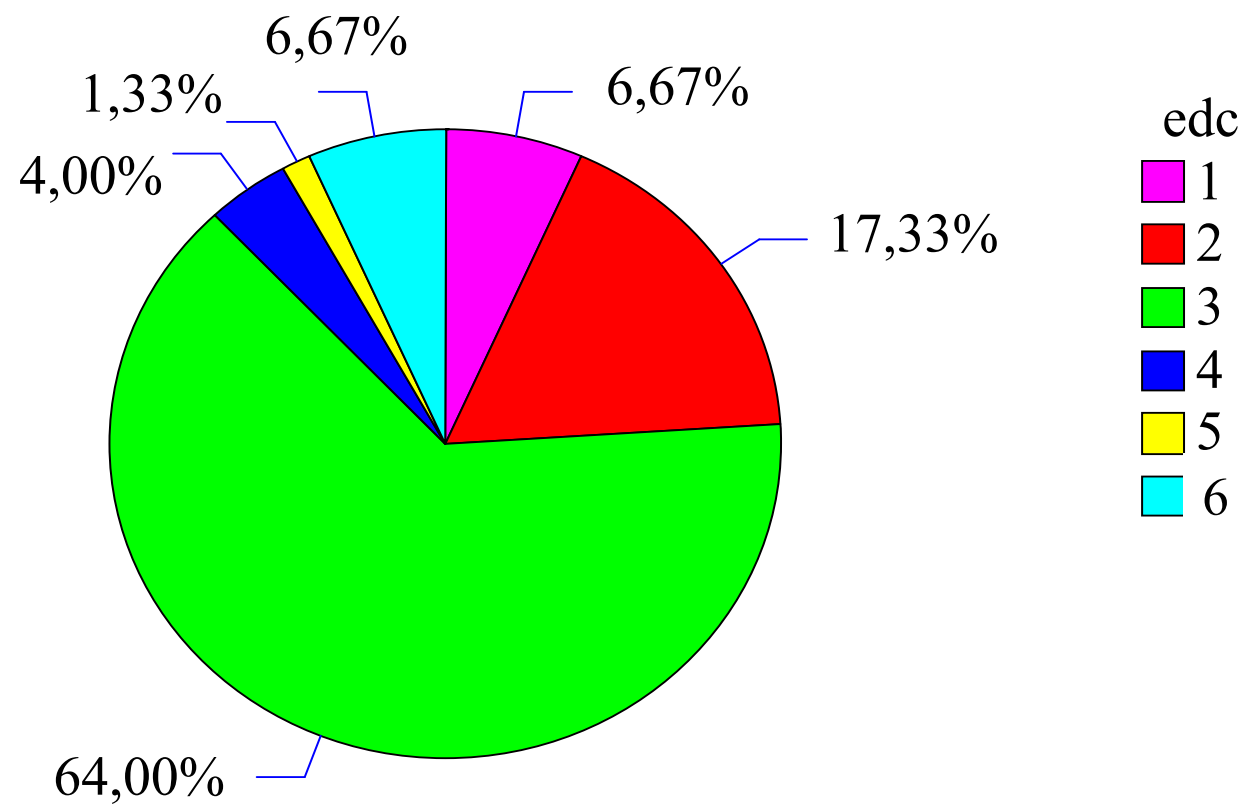


Datos cualitativos transversales : Pie chart

Gráfico de proporciones



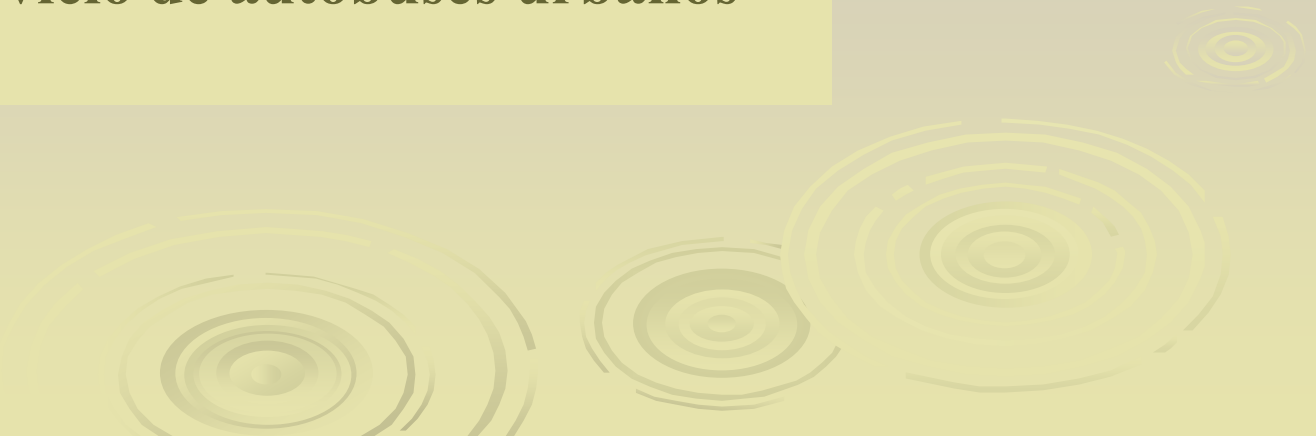
Datos cualitativos transversales : Pie chart



Datos cualitativos: Análisis de Pareto

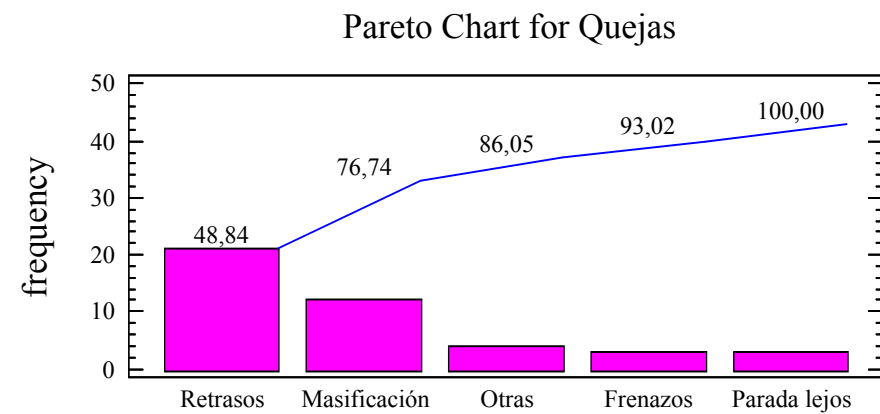
- **Muy apropiado para estudiar causas de problemas.**
- **Suele haber algunas causas que se repiten mucho.**
- **Varias causas son responsables de un importante número de problemas**

**Estamos estudiando las reclamaciones
en un servicio de autobuses urbanos**



Datos cualitativos: Análisis de Pareto

Causa de la Queja	Número de quejas
Retrasos	21
Masificación	12
Parada lejos	3
Frenazos	3
Otras	4



Vamos a estudiar cómo se representan gráficamente

➤ Datos cuantitativos

- Transversales: Histograma, Box plot.
- Temporales: Gráfico de la serie

➤ Datos cualitativos

- Transversales: Tarta, Barras
- Análisis de Pareto
- Temporales: Gráfico de la serie

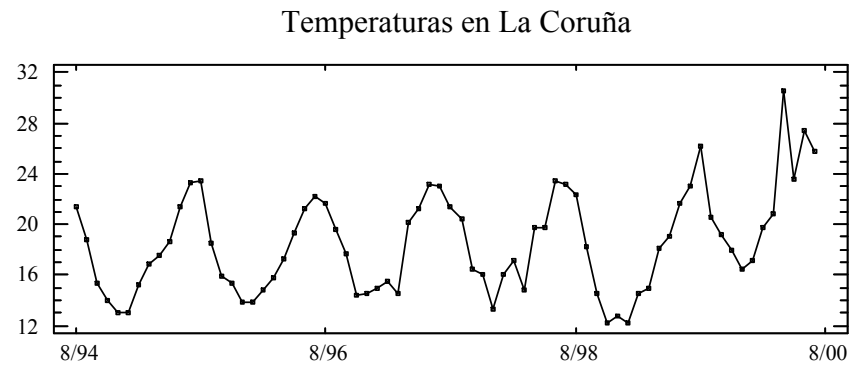
Series temporales:

➤ Datos de evolución de variables en el tiempo:

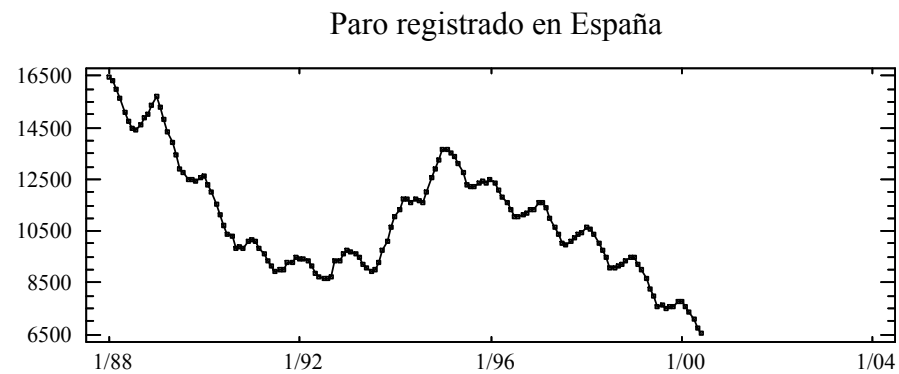
1. Periodicidad: Frecuencia de recogida de datos. Anual, mensual
2. Tendencia: Si aumenta o disminuye con el tiempo
3. Variabilidad-Volatilidad: Su variación (grosor)
4. Ciclo estacional Se observa un ciclo ligado al momento del año en que se ha recogido el dato

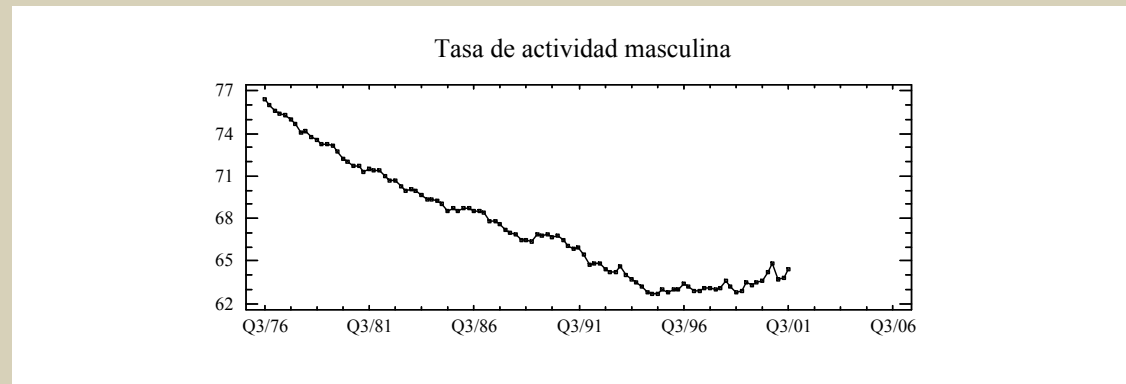


Periodicidad mensual- Ciclo estacional



Periodicidad mensual- Tendencia- Ciclo estacional





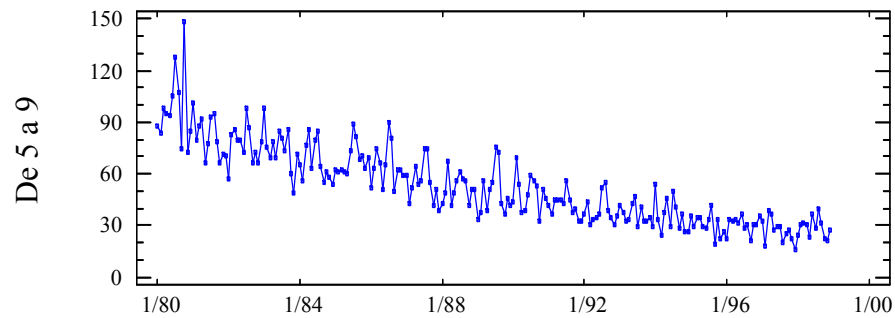
¿Tiene tendencia?

Mortalidad Española por edades e infantil en la UE desde 1975

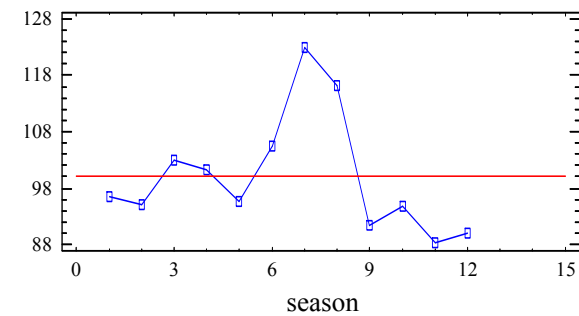
- Fuente de datos: INE. www.ine.es
INEBASE

Mortalidad española. Número de personas Por edades desde Enero de 1980. Datos mensuales Jóvenes: de 5 a 9 y de 20 a 24 años

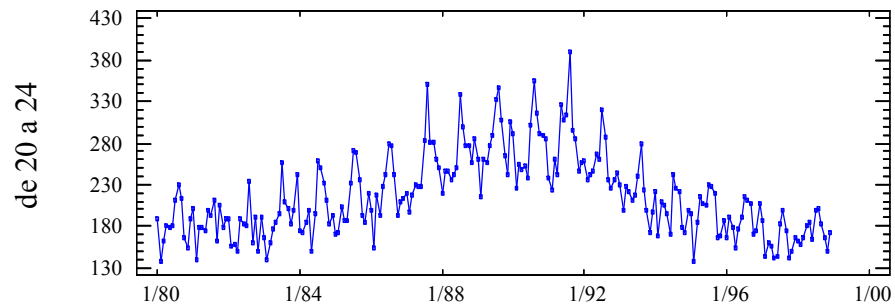
Mortalidad personas de 5 a 9 años



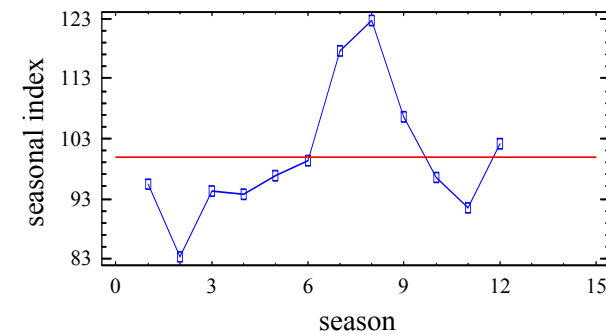
Seasonal Index Plot for De 5 a 9



Mortalidad personas de 20 a 24 años

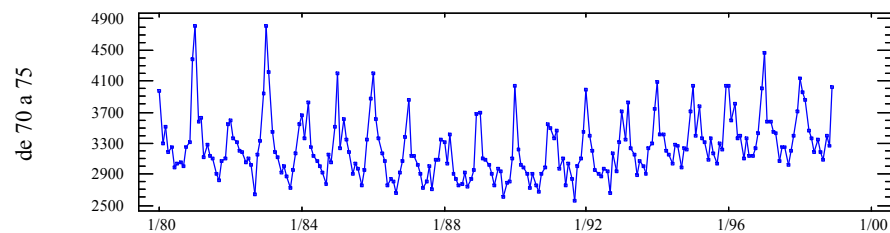


Seasonal Index Plot for de 20 a 24

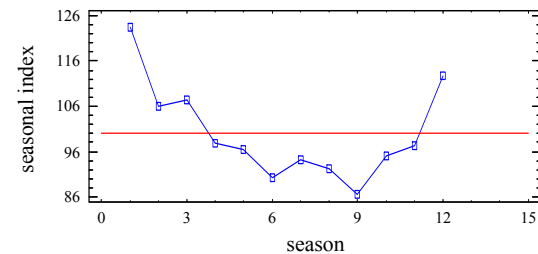


Personas mayores: 70 a 75 y 80 a 85

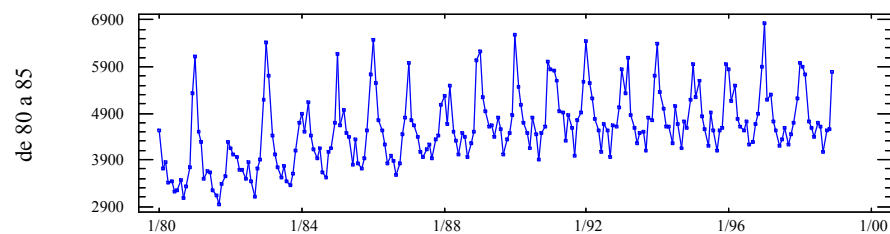
Mortalidad personas de 70 a 75 años



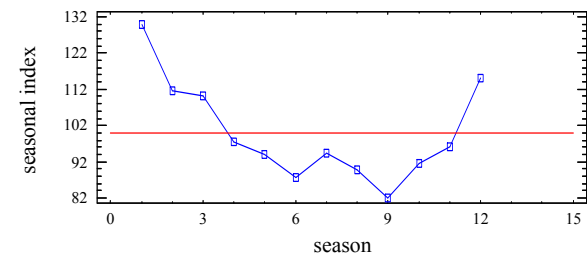
Seasonal Index Plot for de 70 a 75



Mortalidad personas de 80 a 85 años



Seasonal Index Plot for de 80 a 85

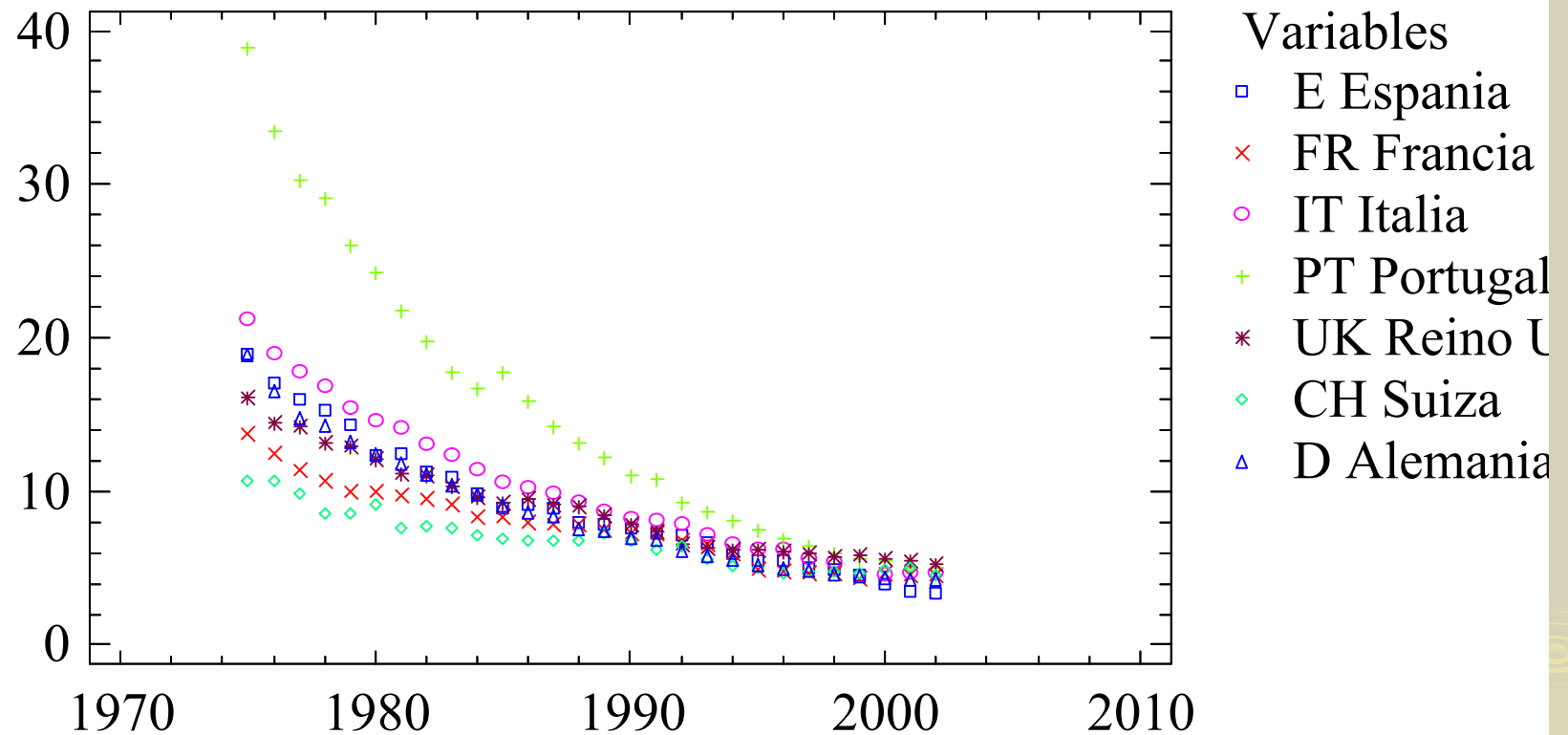


Mortalidad infantil en la UE

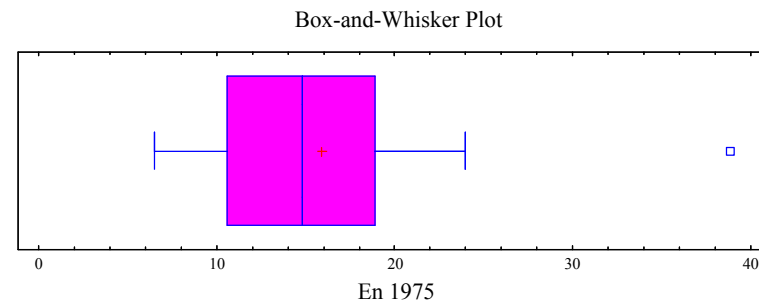
- Fuente: INE.
- Muertos por 1000 nacidos
- Desde 1975



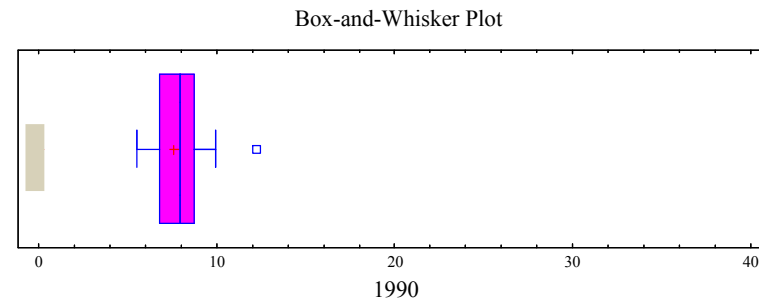
Mortalidad por 1000 nacidos UE



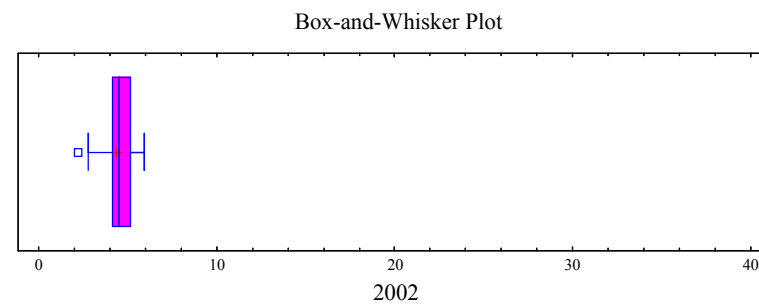
1975



1990



2002



¿A qué se debe el descenso?

Vamos a estudiar cómo se representan gráficamente

➤ Datos cuantitativos

- Transversales: Histograma, Box plot.
- Temporales: Gráfico de la serie

➤ Datos cualitativos

- Transversales: Tarta, Barras
- Análisis de Pareto
- Temporales: Gráfico de la serie

Datos especiales

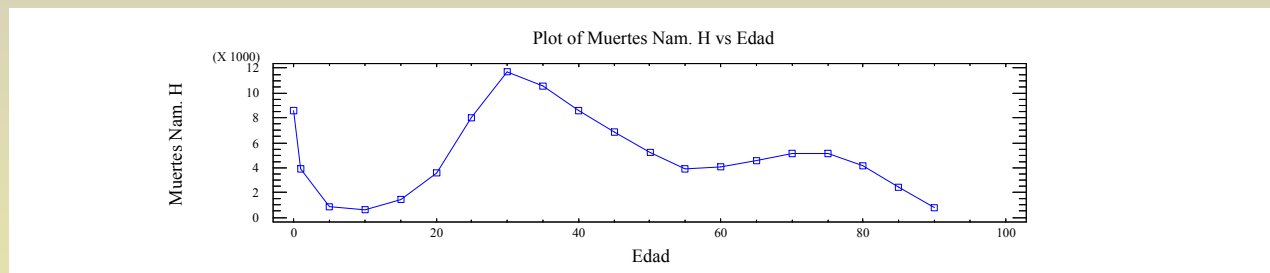
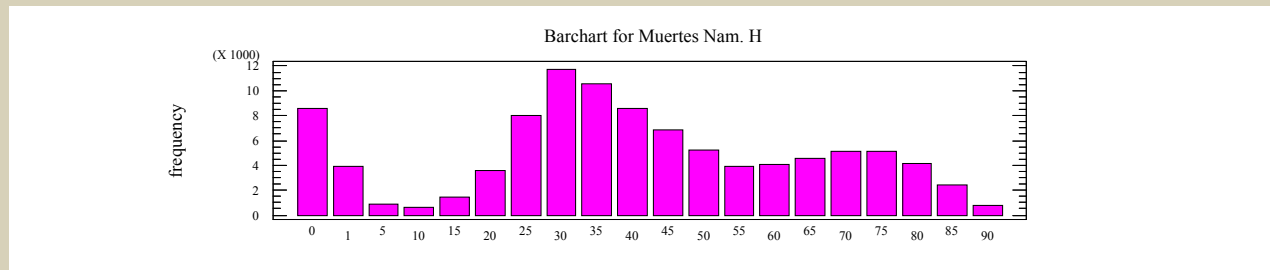
- Mortalidad por edades y sexo en España, Namibia, Afganistán y Francia
- Fuente OMS

Fichero

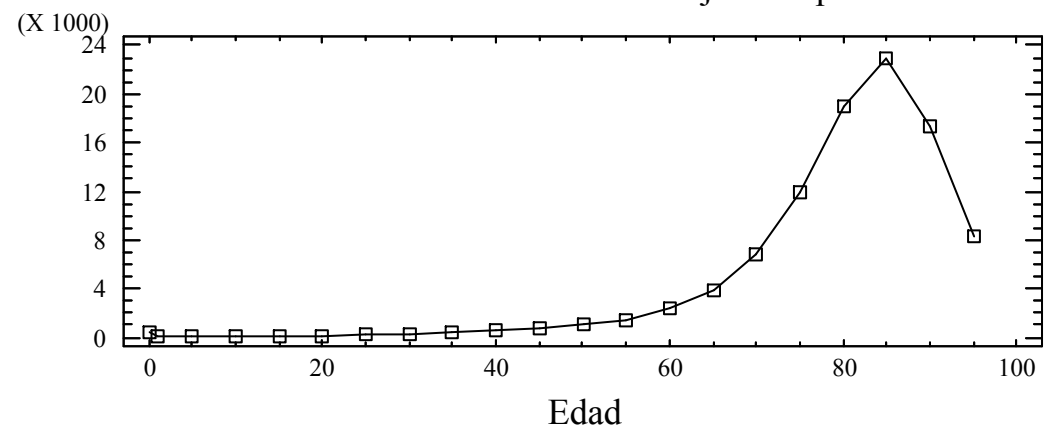
	Edad	Muertes Nam. H	Muertes Nam. Muj	Muertes Esp. Muj.	Muertes Esp. Hom	Afg. Hom. H	Afg. Muj. H	
1	0	8568,71019	7790,24988	401,170012	493,422997	-1,64864527368	-1,74131769828	1
2	1	3912,44484	4107,31234	98,7814833	106,683515	-2,61238244887	-2,4104291345	8
3	5	846,644932	844,917404	58,6835066	103,190634	-3,8621068088	-3,65925851788	7
4	10	607,438922	595,769925	69,277617	100,701737	-4,44583326371	-4,09666184505	7
5	15	1401,76613	2566,52814	134,241022	306,698648	-4,31376853974	-3,75840703996	7
6	20	3539,26358	7223,45537	132,909449	464,127578	-3,88355314538	-3,36870357039	7
7	25	7997,76138	11521,0723	178,594326	536,556225	-3,42991110309	-3,25342327475	6
8	30	11676,4829	11779,1118	241,194616	746,40871	-3,15785081263	-3,21792377821	6
9	35	10561,505	8252,09356	341,77619	905,359915	-2,90140730618	-3,0483721584	6
10	40	8574,53286	5831,77448	514,966231	1159,0383	-2,67929351659	-3,07100187398	6
11	45	6870,26084	3988,70023	734,236283	1711,47067	-2,4852210192	-2,90567733291	5
12	50	5225,70922	3253,95203	1070,53235	2573,49593	-2,26603110547	-2,61564965928	5
13	55	3942,84009	2882,22982	1471,09577	3649,15817	-2,03507417224	-2,31226327588	4
14	60	4070,05564	3245,54025	2410,65046	5809,92519	-1,74339442398	-1,94037591124	4
15	65	4535,8508	4105,83566	3914,70783	8268,1462	-1,43793777995	-1,57197663079	3
16	70	5097,25668	5298,41787	6784,40154	11656,9745	-1,08724089869	-1,19106253498	2
17	75	5108,89624	5987,14219	11875,4926	15569,2711	-0,770242999578	-0,863698068283	1
18	80	4154,55779	5474,55505	18983,9284	18043,4616	-0,494502900528	-0,567854140161	9
19	85	2392,89566	3587,80105	22977,2191	15714,5998	-0,264567615063	-0,319352267181	3
20	90	749,552453	1319,21434	17381,7253	8459,99563	-0,165893947882	-0,201224170824	8
21	95	157,550154	321,563331	8380,40652	3102,57404	-0,0302067238975	-0,0528348820234	1
22								
23								
24								

¿Cómo representamos estos datos?

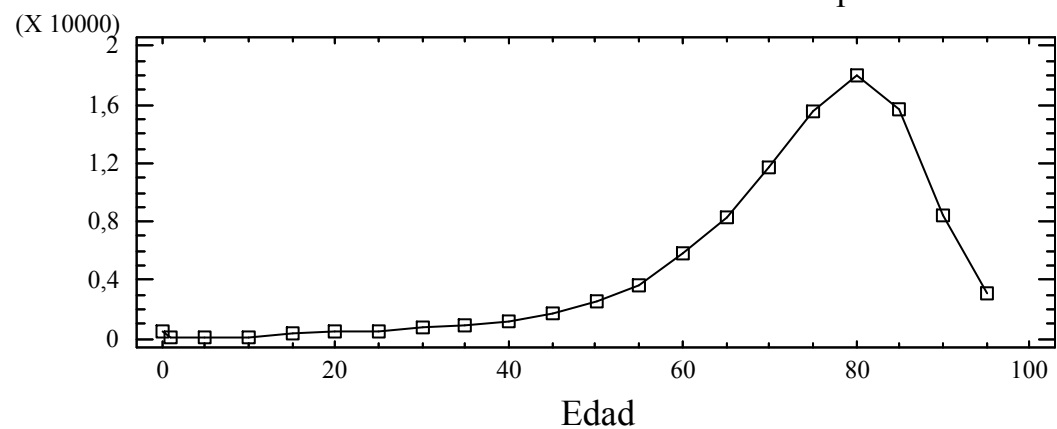
- Mediante un diagrama de barras
- O de puntos



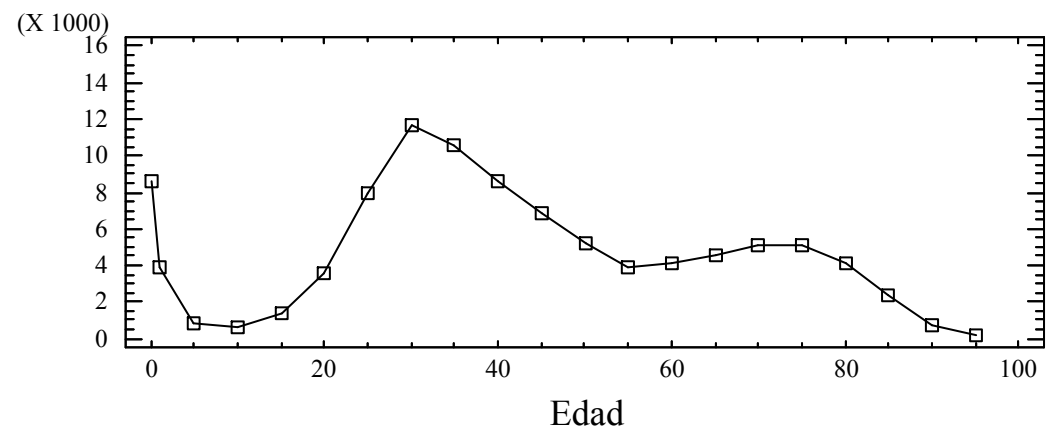
Distribución de muertes Mujeres España



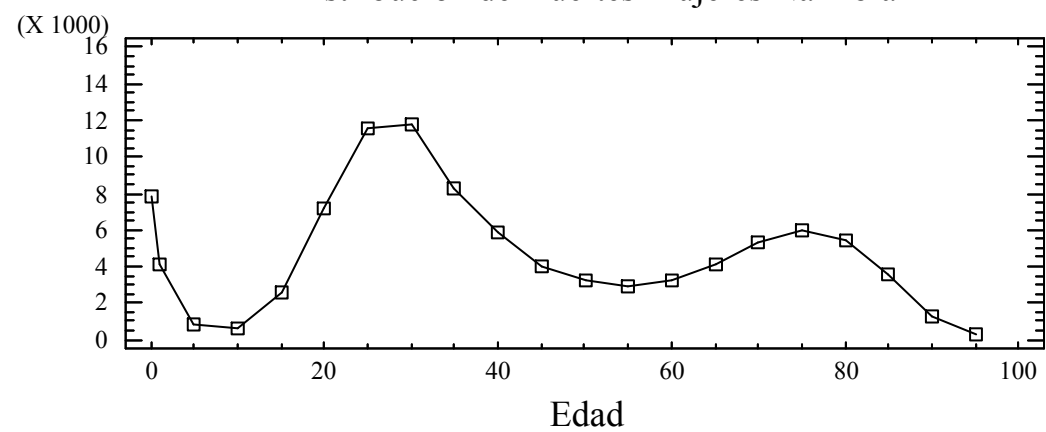
Distribución de muertes Hombres España

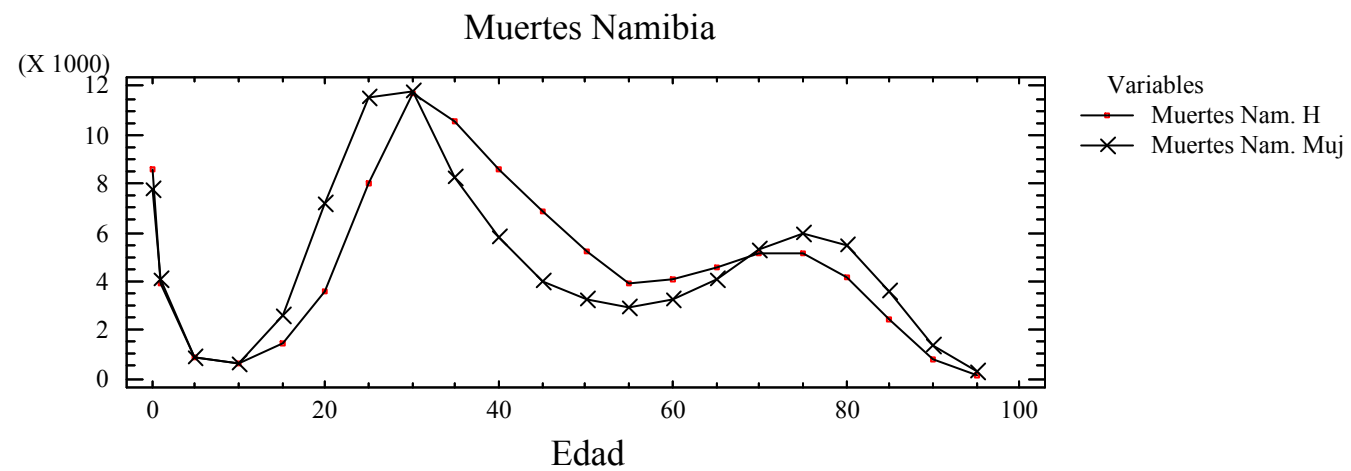
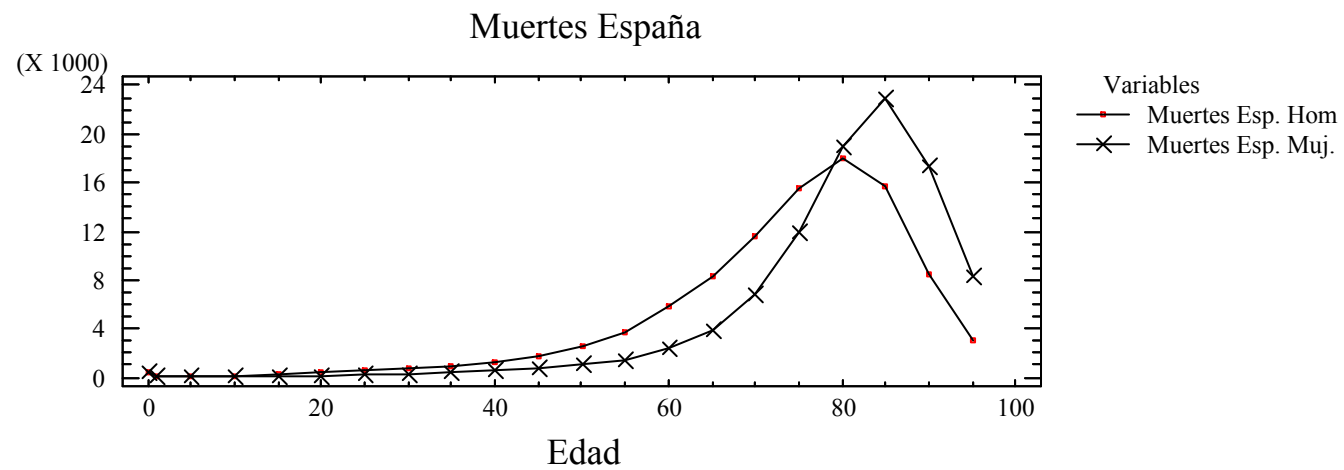


Distribución de muertes Hombres Namibia



Distribución de muertes Mujeres Namibia





Medidas analíticas (números) para describir los datos

- Posición de la variable:
 - Media (No la explico por obvia)
 - Mediana (Observación del medio)
- Dispersión o amplitud de la variable
 - Desviación típica: Mide la amplitud de los datos



Medidas analíticas (números) para describir los datos

Mediana (Observación del medio de los datos)

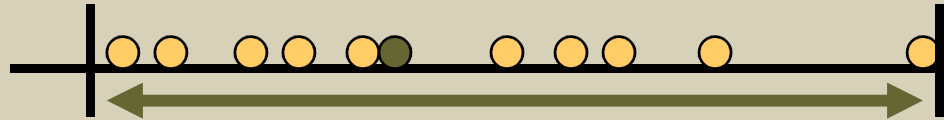


Representa estupendamente la ubicación de los datos.
Es una medida con ventajas respecto a la media

Medidas analíticas (números) para describir los datos

Dispersión: mide la amplitud de los datos

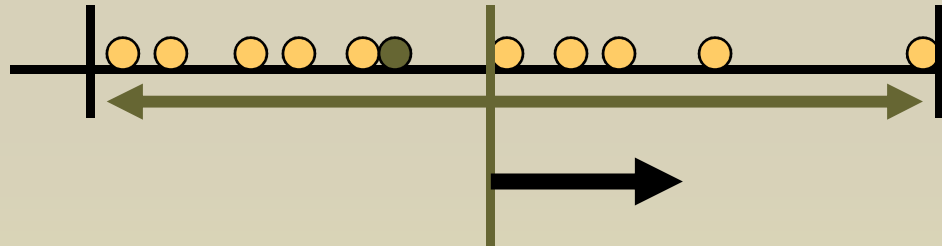
Rango mide la amplitud de los datos



Medidas analíticas (números) para describir los datos

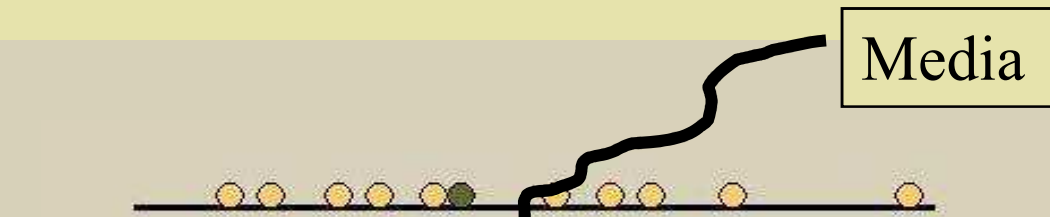
Dispersión: mide la amplitud de los datos

Rango mide la amplitud de los datos



Desviación típica mide la amplitud de los datos

¿Por qué la mediana es muy útil?



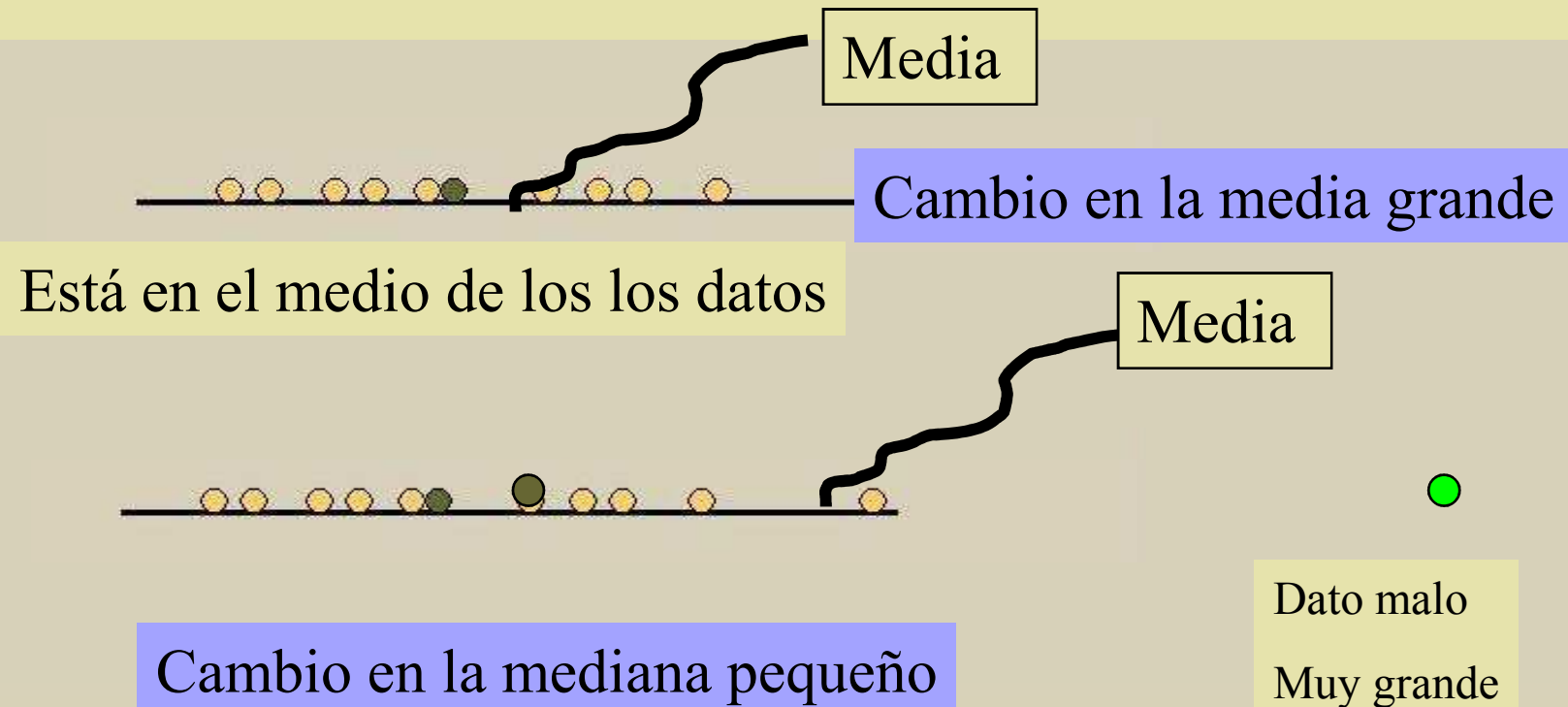
Está en el medio de los los datos



Dato malo

Muy grande

¿Por qué la mediana es muy útil?



Medidas analíticas para describir los datos

- ***La media no siempre es una medida satisfactoria.
Ejemplo***

**En un hotel se pregunta a los clientes por su
valoración de:**

- Limpieza de la habitación
- Rapidez en los trámites de entrada
- Iluminación del BAR

Las valoraciones para dos hoteles han sido:

Variable – Atributo de Calidad	Valor medio obtenido: 6 HOTEL 1	Valor medio obtenido: 6 HOTEL 2
Limpieza Habitación	3	8
Rapidez trámites entrada	5	7
Iluminación BAR	10	3

¿A qué hotel irías?

¿Son todos los atributos de calidad igualmente importantes?

Si pensamos que la importancia es:

- Limpieza habitación: 50%
- Recepción: 40%
- Iluminación del Bar: 10%

➤ HOTEL 1: $0.5 \times 3 + 0.4 \times 5 + 0.1 \times 10 = 4.5$

➤ HOTEL 2: $0.5 \times 8 + 0.4 \times 7 + 0.1 \times 3 = 7.1$

Relación entre dos variables

- Gráfico de dispersión
- Gráfico de dispersión múltiple
- Correlaciones

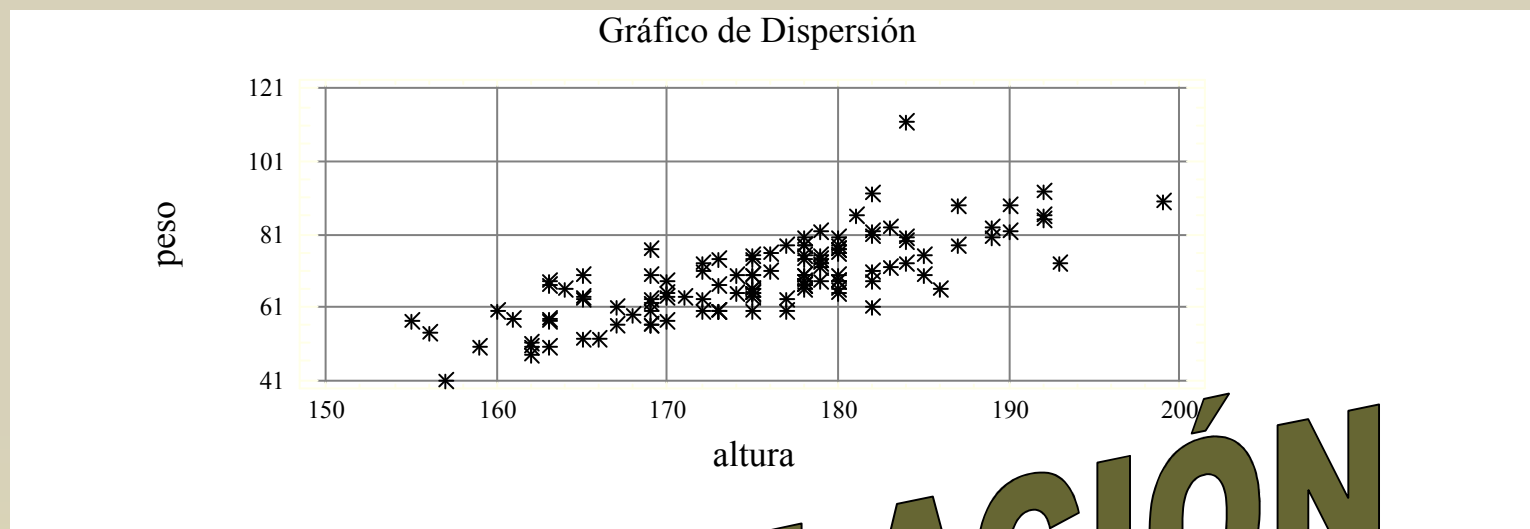


Relación entre dos variables:

- Normalmente hay que analizar más de un aspecto de interés:
 - VARIAS VARIABLES.
- Gráfico de dispersión (*Scatterplot*) resulta muy útil.

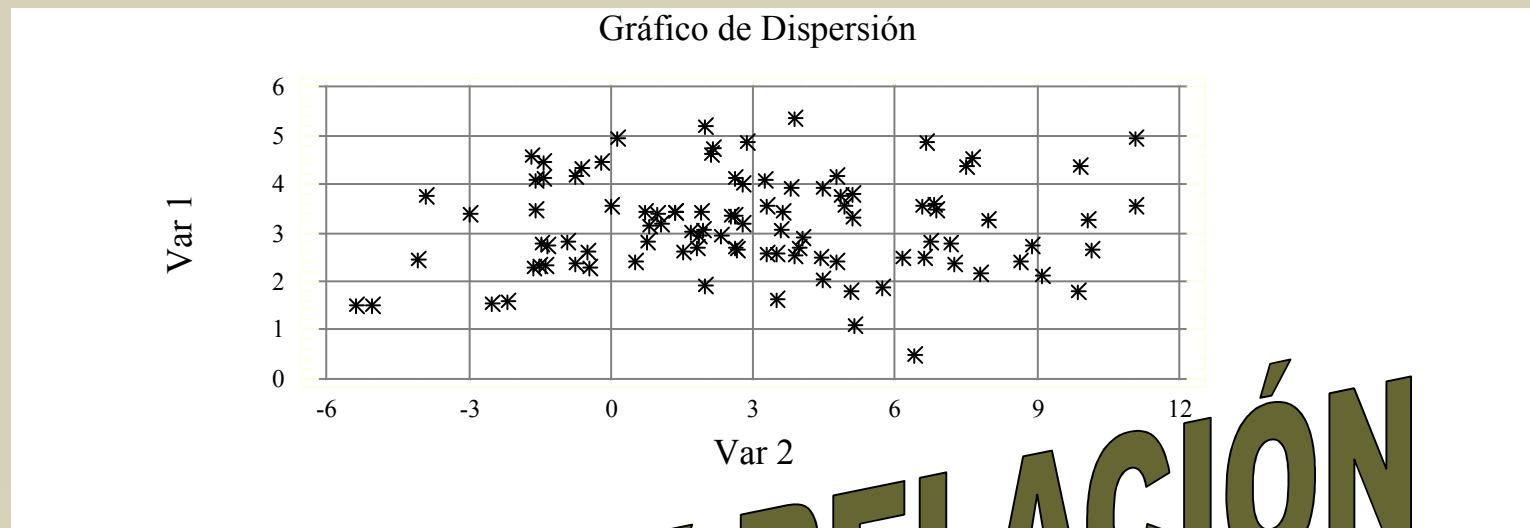


Relación entre dos variables: PESO Y ALTURA



HAY RELACIÓN

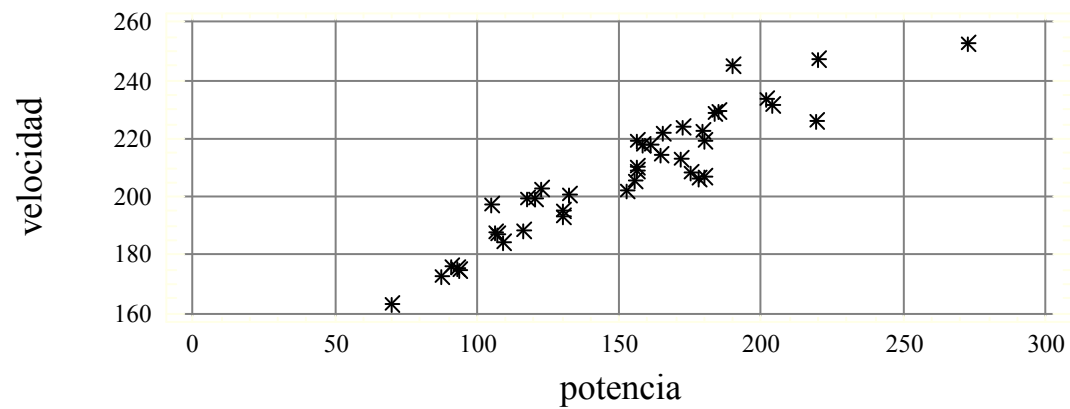
No hay relación entre dos variables:



NO HAY RELACIÓN

¿Hay relación entre estas variables?

Gráfico de Dispersión



Para medir el grado de relación entre variables

- Utilizamos la correlación.
- Varía entre -1 y $+1$

Interpretación de la correlación:

-1

0

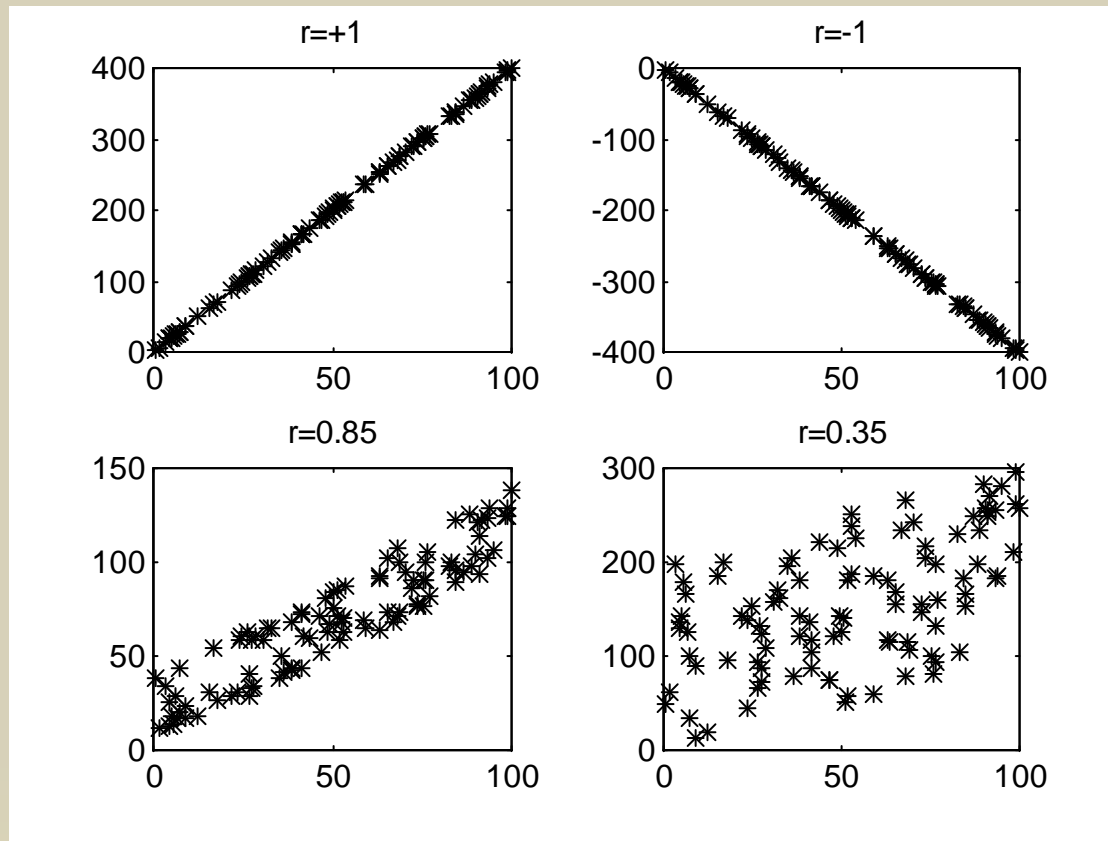
+1

**Mucha relación
Decreciente**

No hay relación

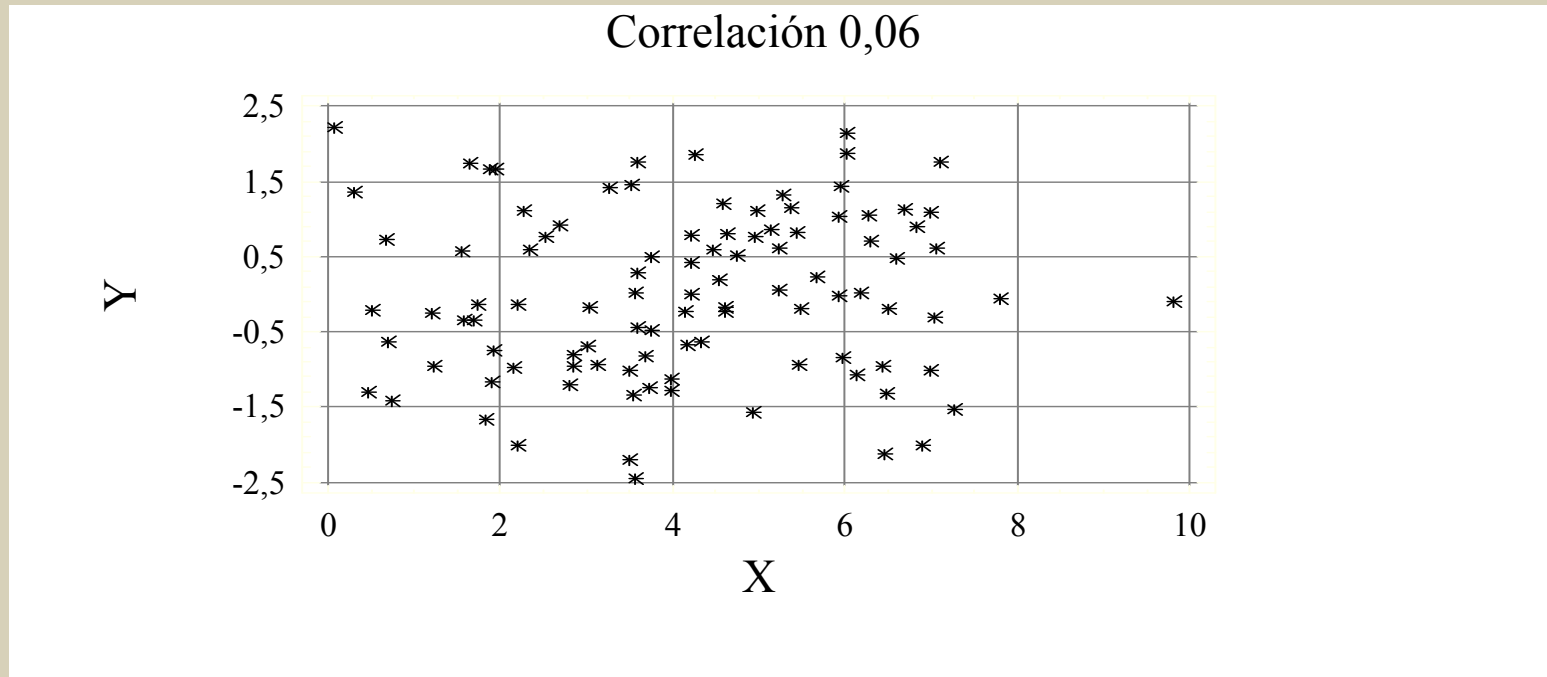
**Mucha relación
Creciente**

Interpretación de la correlación



- + Relación creciente: Si una variable aumenta, la otra también
- Relación decreciente: Si una variable aumenta, la otra disminuye

Interpretación de la correlación



Si la correlación es muy pequeña indica falta de relación entre las variables.

Dispersión múltiple

- Cuando tenemos muchas variables hacer los gráficos de dos en dos es muy latoso
- Los gráficos múltiples hacen de golpe todas los graficos:



Dispersión múltiple

- Saca todos los gráficos de dispersión entre un grupo de variables
- Para países del mundo en 1995



