

ESTADÍSTICA DESCRIPTIVA

Grado de ADE

Tema 2

Tema 2

distribuciones de frecuencia y
representaciones gráficas

Conjunto de datos: ejemplo

- Hemos preguntado a un conjunto de N personas cual consideran que ha sido su estado de salud en los últimos 12 meses. Las posibles respuestas son:
 - 1 Muy bueno
 - 2 Bueno
 - 3 Regular
 - 4 Malo
 - 5 Muy malo

ejemplo

Respuestas:

4 1 1 3 1 1 2 3 3 5 5 4 5 5 4 5 3 5 4 2 5 5 4 5 4 3 4 3 4 3 3 1 4 4 5 5 3 5 4 4 4 3 5 3 4 2 4 2 5 3 4 2 4 1 3 3 3 1 2
4 4 5 2 4 4 4 2 4 5 4 4 5 4 5 3 5 4 3 3 3 3 4 5 4 5 4 1 5 5 5 4 4 4 4 5 2 5 4 5 2 4 4 3 5 3 2 4 3 5 3 4 3 2 4 4 4 4 2
1 3 2 4 4 5 5 4 5 5 1 5 3 2 5 4 5 5 3 2 5 1 5 3 5 3 2 2 2

Conjunto original de datos

- Nos hacemos unas preguntas sobre los datos:

¿Cuántas personas fueron encuestadas ?

¿Cuál fue la respuesta más frecuente ?

¿Cuántas personas tienen mala o muy mala salud?

ejemplo

Conjunto original de datos:

4 1 1 3 1 1 2 3 3 5 5 4 5 5 4 5 3 5 4 2 5 5 4 5 4 3 4 3 4 3 3 1 4 4 5 5 3 5 4 4 4 3 5 3 4 2 4 2 5 3 4 2 4 1 3 3 3 1 2 4 4 5
2 4 4 4 2 4 5 4 4 5 4 5 3 5 4 3 3 3 3 3 4 5 4 5 4 1 5 5 5 4 4 4 4 5 2 5 4 5 2 4 4 3 5 3 2 4 3 5 3 4 3 2 4 4 4 4 2 1 3 2 4 4
5 5 4 5 5 1 5 3 2 5 4 5 5 3 2 5 1 5 3 5 3 2 2 2

- Preguntas

¿Cuántas personas fueron encuestadas ?

¿Cuál fue la respuesta más frecuente ?

¿Cuántas personas tienen mala o muy mala salud?

*Las **preguntas** son difíciles de responder así ... ¿por qué?*

¿y si ordenamos la información?

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 4
4 4 4 4 4 4 5

Datos reorganizados: no perdemos información

ejemplo

Datos reorganizados:

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 3 3 3 3 3 3 3
3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 4
4 4 4 4 4 4 4 5

Pero aun así, se puede mejorar (hay que contar el numero de datos, de unos, de doses...)

1 (16), 2 (19), 3 (22),
4 (38),
5 (29)

*Conjunto ordenado
de "valores" y "frecuencias"*

ejemplo

1 (16), 2 (19), 3 (22),
4 (38),
5 (29)

Conjunto ordenado
de "valores" y "*frecuencias*"

Lo ponemos en forma de tabla:

respuesta	n
1	16
2	19
3	22
4	38
5	29
	124

ejemplo

Preguntas:

¿Cuántas personas fueron encuestadas ? 124

¿Cuál fue la respuesta más frecuente ? 4 mala salud

¿Cuántas personas tienen mala (4) o muy mala(5) salud? 38+29

respuesta	n
1	16
2	19
3	22
4	38
5	29
	124

ejemplo

¿Cómo respondemos a la pregunta ..”¿Cuántas personas contestan que su salud es mejor que “regular”?

- Hay que sumar los que contestan muy buena, buena:
frecuencias acumuladas

salud	n	N
1	16	16
2	19	35
3	22	57
4	38	95
5	29	124
	124	

Recogemos datos en otra población y los queremos comparar:

salud	n
1	200
2	170
3	120
4	60
5	40
	590

salud	n	N
1	16	16
2	19	35
3	22	57
4	38	95
5	29	124
	124	

Ahora son 590, frente a los 124 del colectivo anterior. Una misma frecuencia no tiene la misma interpretación en ambas tablas.

¿Qué ha cambiado?: **la importancia relativa de la frecuencia**, puesto que $n=40$ frente a $N=124$ es diferente a $n=40$ frente a $N=590$. De hecho, el valor 4 pasa incluso de ser el más frecuente a ser el segundo menos presente.

La solución se encuentra en expresar las frecuencias en términos relativos en vez de absolutos. Esto es precisamente lo que consiguen las **proporciones**: expresar una cantidad con respecto al total. Así, añadimos una nueva columna, conteniendo las **frecuencias relativas** (f) que surgen de hacer la operación $f = n / N$.

El resultado, al añadir las frecuencias relativas $f_i = n_i/N$:

salud	n	f
1	16	0,13
2	19	0,15
3	22	0,18
4	38	0,31
5	29	0,23
	124	1

salud	n	f
1	200	0,34
2	170	0,29
3	120	0,20
4	60	0,10
5	40	0,07
	590	1

Por lo general, cuando se exponen los resultados de una encuesta en un medio de comunicación, lo habitual es utilizar otro tipo de frecuencias relativas: **los porcentajes**. El principio que rige la utilización de los porcentajes es el mismo que para las proporciones o frecuencias relativas: utilizar un referente fijo de tal forma que no sea necesario contar con el número total de datos para interpretar una frecuencia. La diferencia entre los porcentajes y las frecuencias relativas es que los primeros utilizan el referente 100, mientras que las frecuencias relativas utilizan el 1.

Conseguir los porcentajes a partir de las frecuencias relativas es muy fácil, basta con multiplicar a éstas por 100:

salud	n	f	%
1	16	0,13	12,90
2	19	0,15	15,32
3	22	0,18	17,74
4	38	0,31	30,65
5	29	0,23	23,39
	124	1	100,00

salud	n	f	%
1	200	0,34	33,90
2	170	0,29	28,81
3	120	0,20	20,34
4	60	0,10	10,17
5	40	0,07	6,78
	590	1	100,00

Conceptos

- **Frecuencias absolutas:** Contabilizan el número de individuos de cada modalidad/categoría: n_i

$$N = n_1 + n_2 + \dots + n_k$$

(donde N es el total de individuos de la muestra. k es el tamaño de la muestra)

- **Frecuencias relativas** (porcentajes): el número de individuos de cada modalidad/categoría, dividido por el total : $f_i = \frac{n_i}{N}$

(para obtener los porcentajes se multiplica por 100)

- **Frecuencias acumuladas:** Sólo tienen sentido para variables numéricas y ordinales: absolutas y relativas. Son muy útiles para calcular cuantiles (lo veremos más adelante)
 - **Frecuencias absoluta acumulada:** hasta una modalidad, al número de individuos o número de datos, N_i , que presentan una modalidad menor o igual que ésta: $N_i = n_1 + n_2 + \dots + n_i$
 - **Frecuencias relativa acumuladas:** al cociente $F_i = \frac{N_i}{N}$
 $= f_1 + f_2 + \dots + f_i$

Se cumple que:

- La última frecuencia acumulada es igual a N (el total)

$$N_k = N (= n_1 + n_2 + \dots + n_k)$$

- La suma de las frecuencias relativas es 1.

$$f_1 + f_2 + \dots + f_k = 1$$

- La última frecuencia relativa acumulada es 1

$$F_k = 1 \quad \left(= \frac{N_k}{N} \right)$$

Distribución de frecuencias

Ejercicio 2.1 (a y b)

x_i	n_i	N_i	f_i	F_i
1	3	3	0,10	0,10
2	6	9	0,20	0,30
3	8	17	0,27	0,57
4	5	22	0,17	0,73
5	4	26	0,13	0,87
6	2	28	0,07	0,93
7	1	29	0,03	0,97
8	1	30	0,03	1,00
	30		1,00	

frecuencia absoluta n_i

frecuencia relativa $f_i = n_i/N$

frecuencia absoluta acumulada $N_i = n_1 + \dots + n_i$

frecuencia relativa acumulada $F_i = N_i/N$

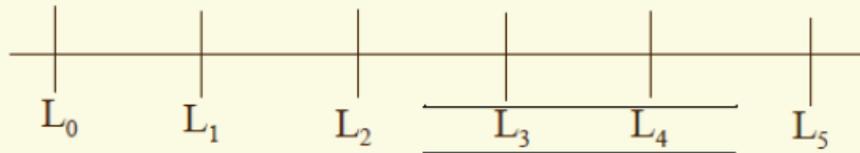
Ejercicio 2.2 (a,b y c)

	número de turistas ni	fi	Ni	Fi
América	163.000.000	0,16	163.000.000	0,15733591
Europa	534.000.000	0,52	697.000.000	0,67277992
Oriente Medio	52.000.000	0,05	749.000.000	0,72297297
África	53.000.000	0,05	802.000.000	0,77413127
Asia y Pacífico	234.000.000	0,23	1.036.000.000	1
	1.036.000.000			

Agrupación en clases

En ocasiones, y con objeto de facilitar la toma o presentación de datos cuantitativos, estos se agrupan en **intervalos o clases**.

Por ejemplo, es más sencillo anotar cuántos individuos hay en una muestra con una estatura entre 1.70 y 1.80, que anotar exactamente la estatura de todos. No obstante, siempre se producirá una pérdida de información al agrupar los datos en intervalos (no deseable).



Los datos se agrupan en clases o intervalos (datos agrupados):

$$L_1 < L_2 < \dots < L_k$$

Agrupación en clases: distribución de frecuencias

Límites de clase: valores extremos que se pueden observar dentro de una clase: L_i

Marcas de clase: punto medio entre fronteras $x_i = \frac{L_{i-1} + L_i}{2}$

Amplitud ó longitud de clase: distancia entre fronteras $c_i = L_i - L_{i-1}$

Densidad de clase $d_i = \frac{n_i}{c_i}$

Ejercicio 2.1. apartado d

$[L_{i-1}, L_i)$	L_{i-1}	L_i	x_i	n_i	f_i	N_i	F_i	c_i	d_i
[0,5 2,5)	0,5	2,5	1,5	9	0,30	9	0,30	2	4,5
[2,5 4,5)	2,5	4,5	3,5	13	0,43	22	0,73	2	6,5
[4,5 6,5)	4,5	6,5	5,5	6	0,20	28	0,93	2	3
[6,5 8,5)	6,5	8,5	7,5	2	0,07	30	1,00	2	1
				30	1,00			8	

Ejercicio 2.3 b,c,d,e

Tema 2

GRÁFICOS

CARTE FIGURATIVE des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite.

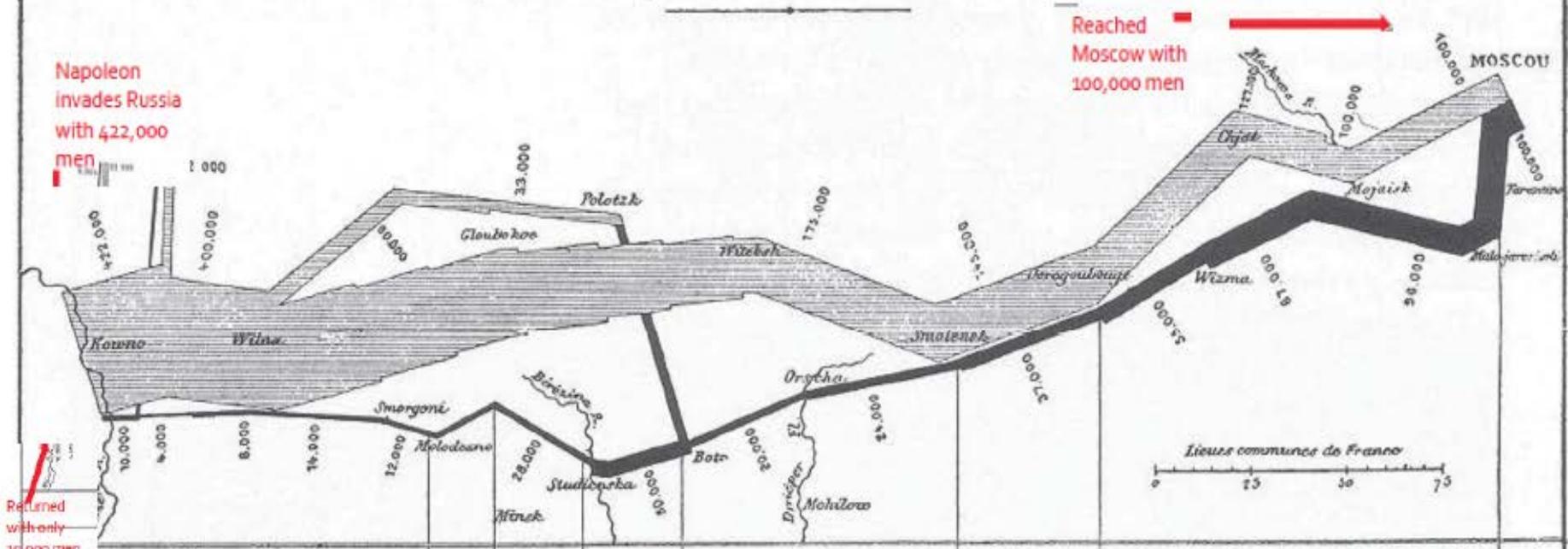
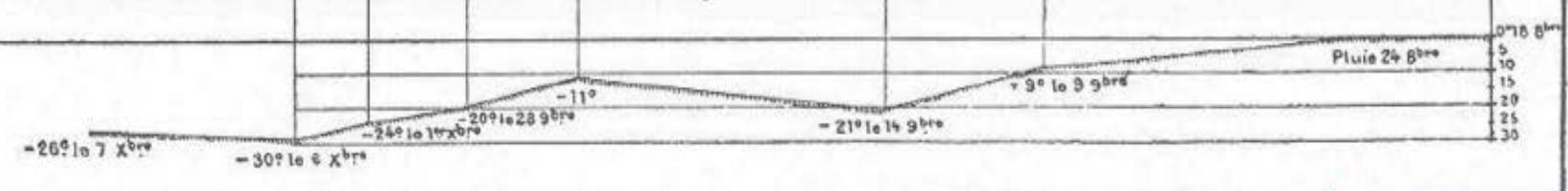


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro



- Six variables are plotted:
- The size of the army
 - It's location on a 2D surface
 - Direction of army movement
 - Temperature on various dates during the retreat



Deaths from cholera in central London, 1854: (John Snow)

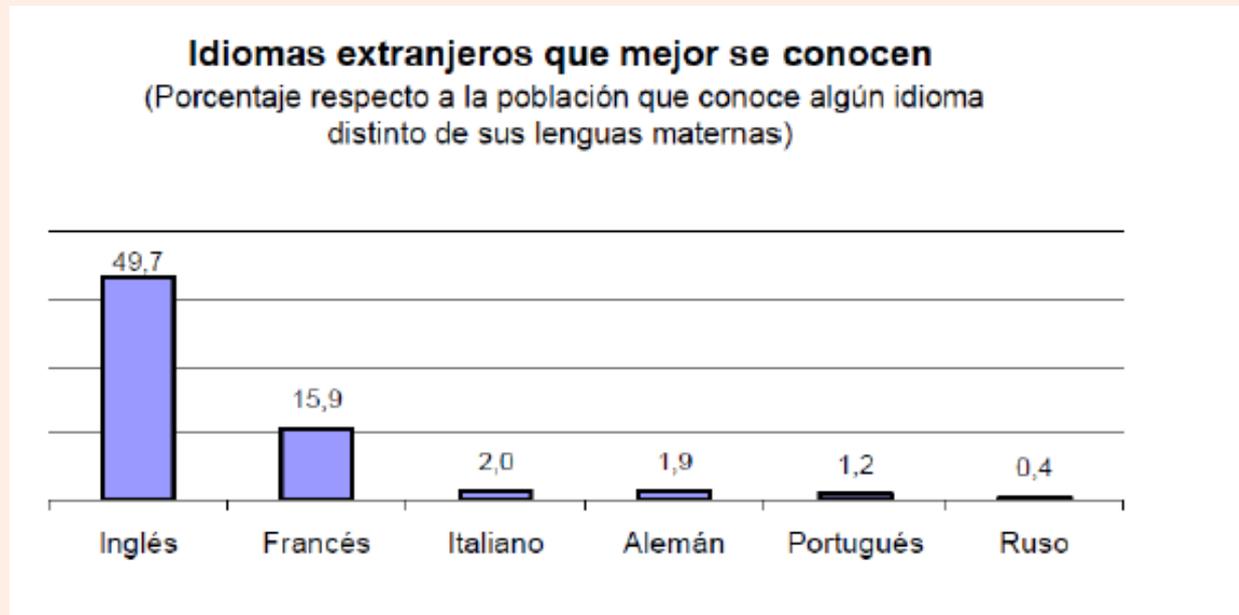


John Snow (Inglaterra 1813 – 1858) fue un médico inglés precursor de la epidemiología, hasta el punto de ser considerado padre de la epidemiología moderna. Demostró que el cólera era causado por el consumo de aguas contaminadas con materias fecales, al comprobar que los casos de esta enfermedad se agrupaban en las zonas donde el agua consumida estaba contaminada con heces, en la ciudad de Londres en el año de 1854. Ese año cartografió en un plano del distrito de Soho los pozos de agua, localizando como culpable el existente en Broad Street. Snow recomendó a clausurar la bomba de agua, con lo que fueron disminuyendo los casos de la enfermedad. Este episodio está considerado como uno de los ejemplos más tempranos en el uso del método geográfico para la descripción de casos de una epidemia.

Gráficos para variables cualitativas

- Diagramas de barras

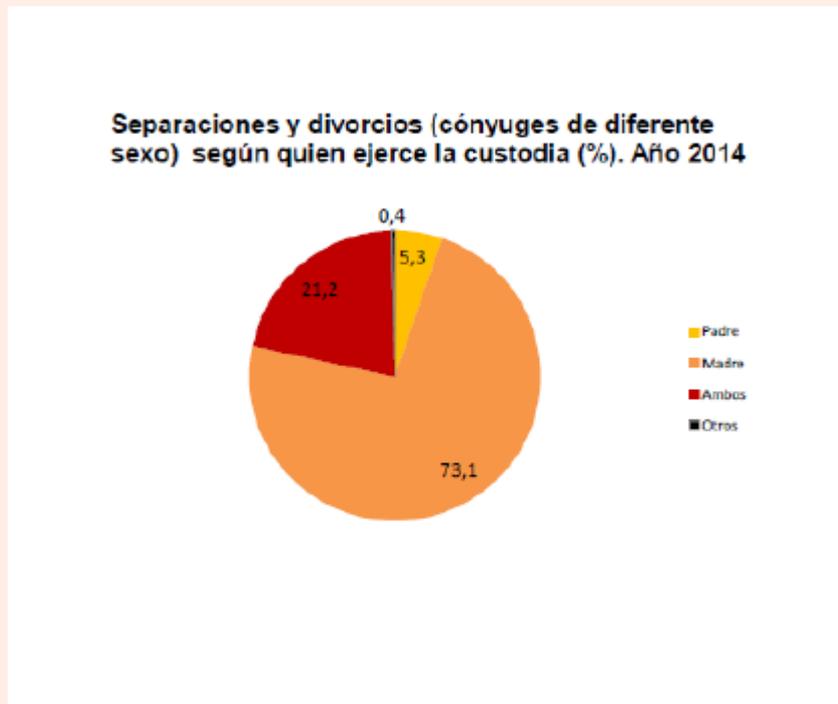
Altura proporcional a la frecuencia (absoluta o relativa)
(Se puede aplicar a vbles cuantitativas discretas también)



Gráficos para variables cualitativas

- Diagramas de sectores

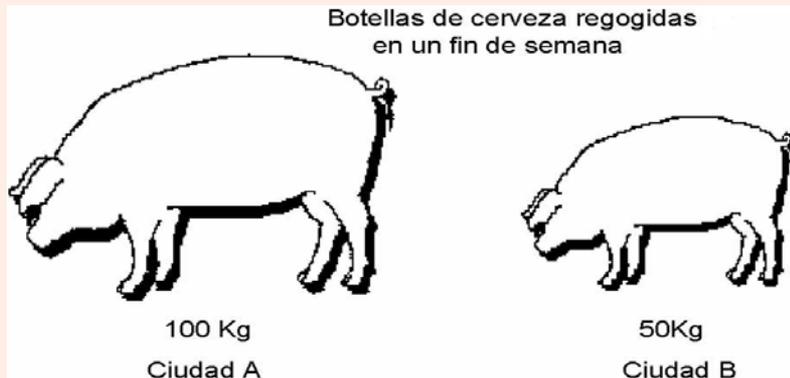
El área de su sector es proporcional a su frecuencia (absoluta o relativa)



Gráficos para variables cualitativas

- Pictogramas

- El área de cada modalidad debe de ser proporcional a su frecuencia (absoluta o relativa).
- Cada uno de los dibujos o símbolos equivale a ciertas unidades de frecuencia relativa o absoluta.

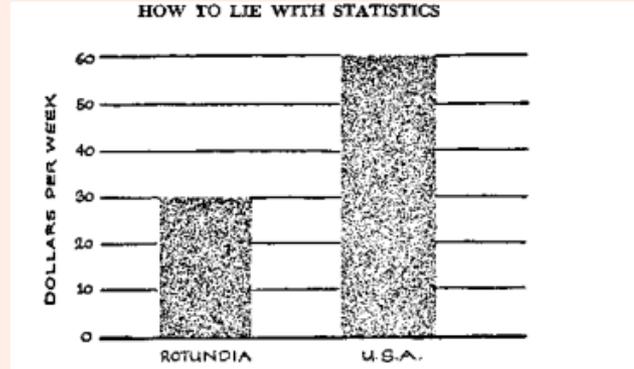


Municipio	Árboles
Chalco	
Coacalco	
Ecatepec	
Toluca	
Villa Victoria	

 = 1000 arboles

Pictogramas

Histograma correcto para el salario medio de dos países (salario es el doble en Rotundia que en EEUU):



¿Cuál será el pictograma correcto?



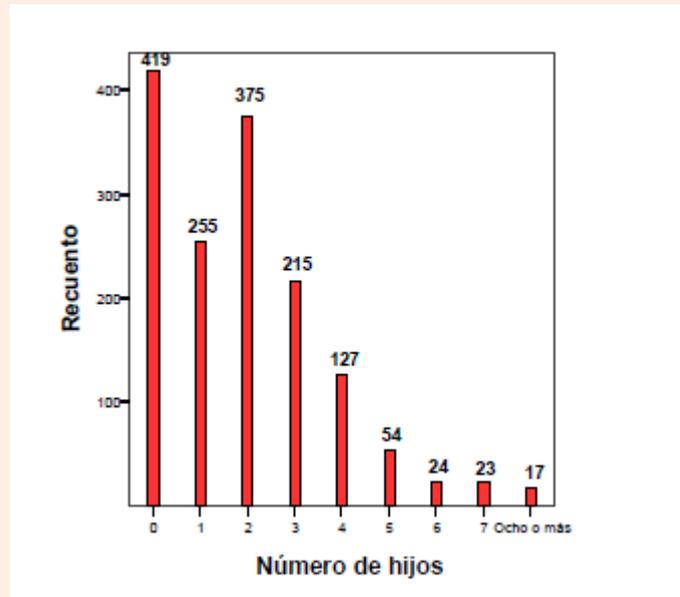
Gráficos para variables numéricas/cuantitativas:

- Diagrama de barras

Variables discretas/datos sin agrupar

Frecuencias absolutas o relativas.

Alturas proporcionales a las frecuencias (abs. o rel.)



Gráficos para variables numéricas/cuantitativas:

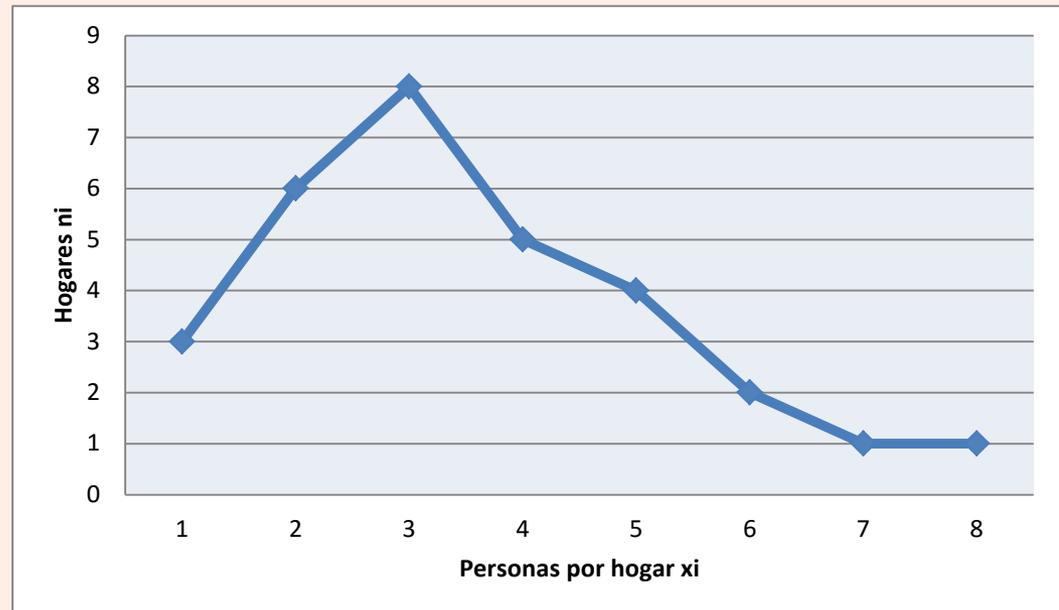
- Polígono de frecuencias

Variables discretas/datos sin agrupar

En ordenadas pueden ir frecuencias absolutas o relativas (n_i o f_i)

Ejercicio 2.1

x_i	n_i	N_i	f_i	F_i
1	3	3	0,10	0,10
2	6	9	0,20	0,30
3	8	17	0,27	0,57
4	5	22	0,17	0,73
5	4	26	0,13	0,87
6	2	28	0,07	0,93
7	1	29	0,03	0,97
8	1	30	0,03	1,00
	30		1,00	



Gráficos para variables numéricas/cuantitativas:

- Polígono de frecuencias acumuladas

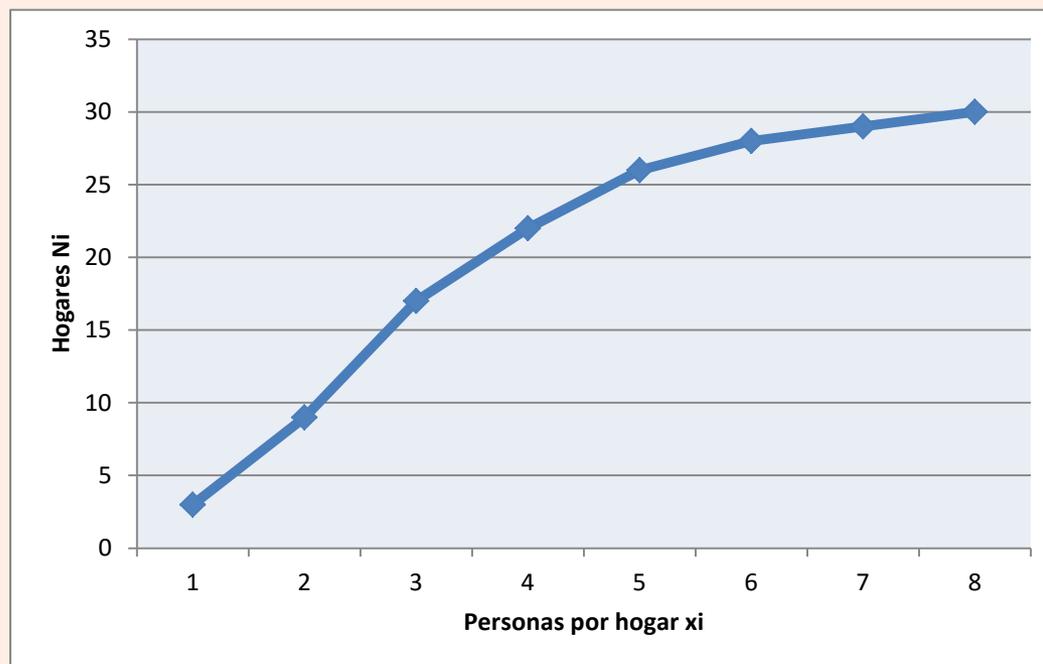
Variables discretas/datos sin agrupar

En cada valor que toma la variable se produce un incremento de la frecuencia acumulada.

En ordenadas pueden ir frecuencias absolutas o relativas (N_i o F_i)

Ejercicio 2.1

x_i	n_i	N_i	f_i	F_i
1	3	3	0,10	0,10
2	6	9	0,20	0,30
3	8	17	0,27	0,57
4	5	22	0,17	0,73
5	4	26	0,13	0,87
6	2	28	0,07	0,93
7	1	29	0,03	0,97
8	1	30	0,03	1,00
	30		1,00	

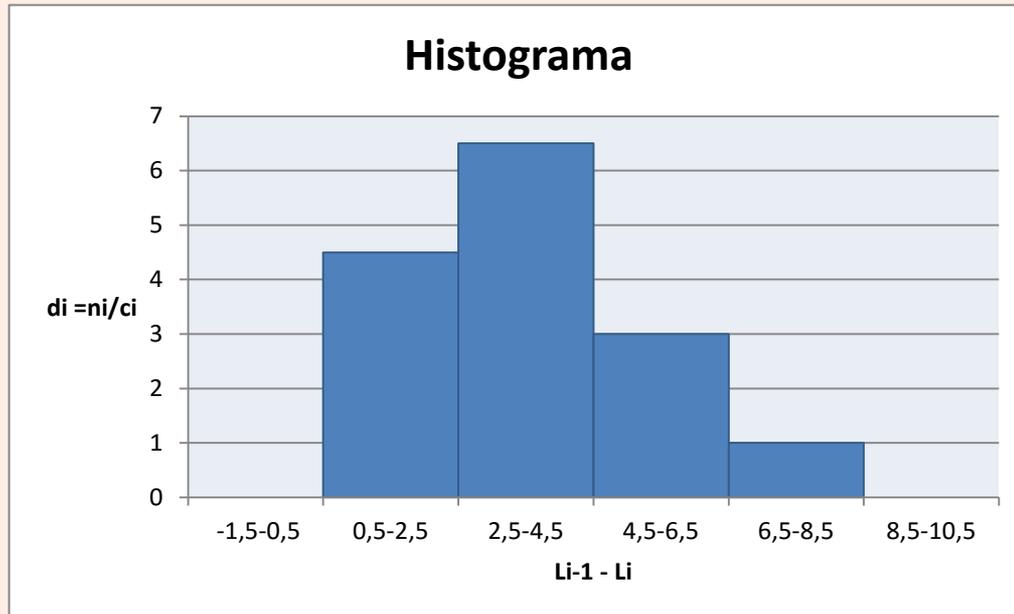


Gráficos para variables numéricas/cuantitativas:

- **Histograma** (datos agrupados)

Variables continuas: El área que hay bajo el histograma entre dos puntos cualesquiera indica la cantidad (porcentaje o frecuencia) de individuos en el intervalo.

Gráfico más frecuente para datos agrupados



Gráficos para variables numéricas/cuantitativas:

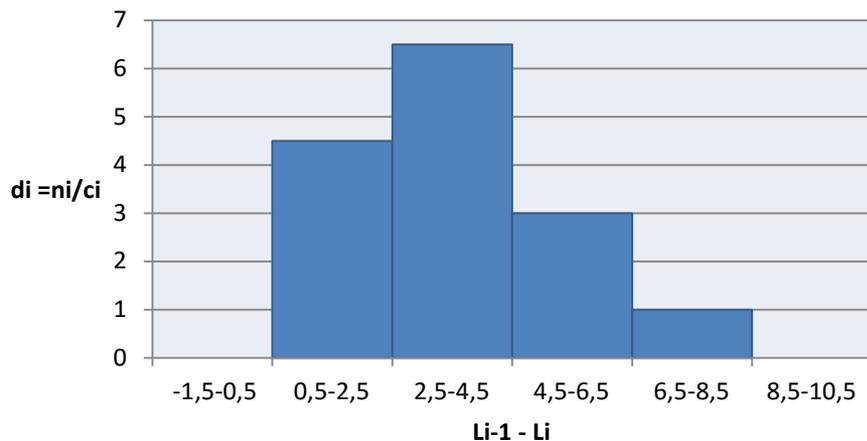
- **Histograma** (datos agrupados)

Ejercicio 2.1

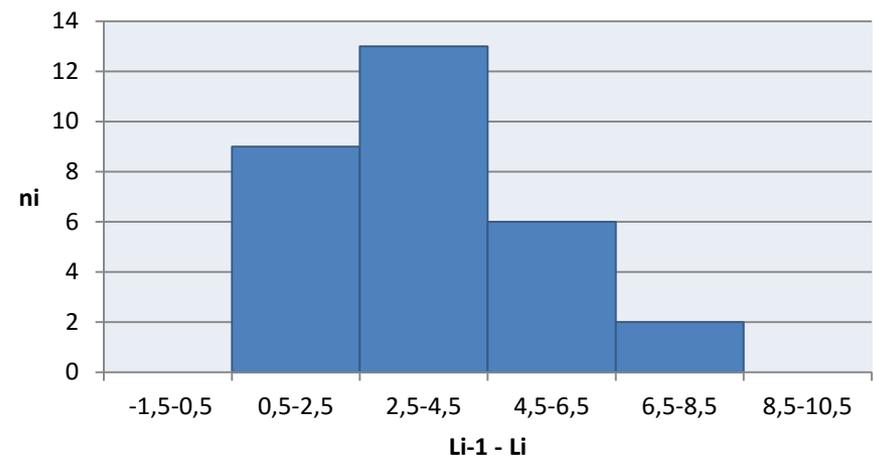
Li-1	Li	xi	ni	fi	Ni	Fi	ci	di
0,5	2,5	1,5	9	0,30	9	0,30	2	4,5
2,5	4,5	3,5	13	0,43	22	0,73	2	6,5
4,5	6,5	5,5	6	0,20	28	0,93	2	3
6,5	8,5	7,5	2	0,07	30	1,00	2	1
			30	1,00			8	0

Si ci es constante, en el eje de las Y se puede poner di o ni

Histograma ($ci = cte$)



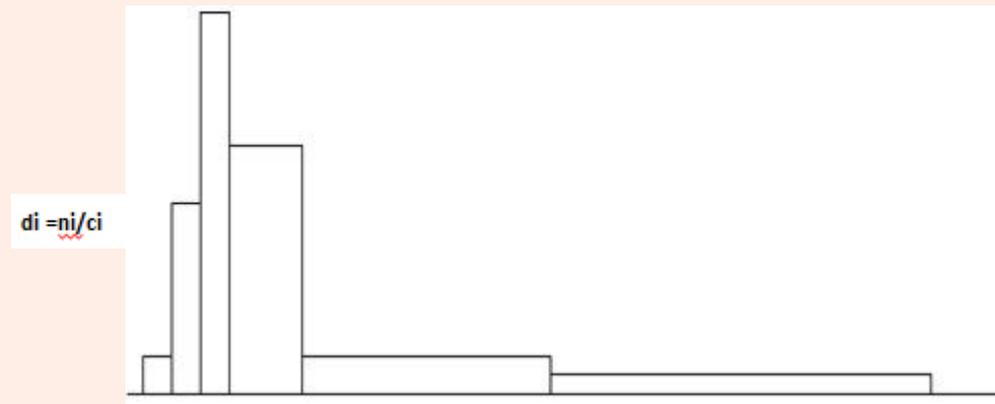
Histograma ($ci = cte$)



Gráficos para variables numéricas/cuantitativas:

- **Histograma** (datos agrupados)

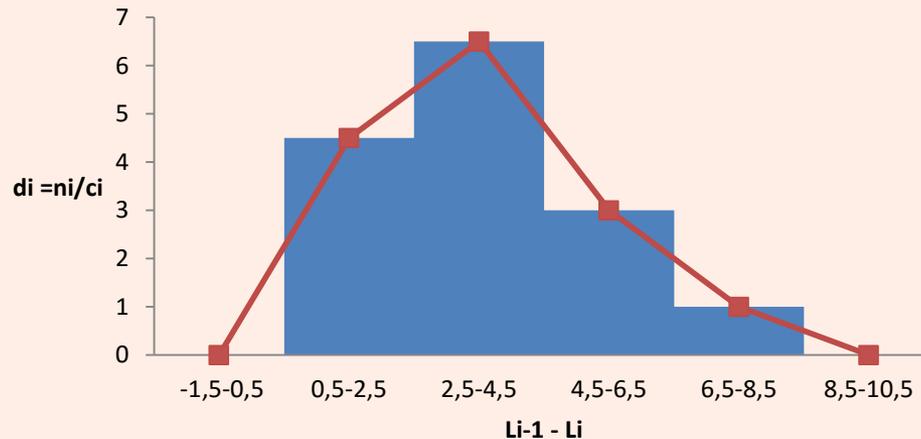
Si c_i no es constante, en el eje de las Y se debe poner d_i



Gráficos para variables numéricas/cuantitativas:

- Polígono de frecuencias

Li-1	Li	xi	ni	fi	Ni	Fi	ci	di
0,5	2,5	1,5	9	0,30	9	0,30	2	4,5
2,5	4,5	3,5	13	0,43	22	0,73	2	6,5
4,5	6,5	5,5	6	0,20	28	0,93	2	3
6,5	8,5	7,5	2	0,07	30	1,00	2	1
			30	1,00			8	



Se unen las marcas de clase y se obtiene el polígono de frecuencias

Gráficos para variables numéricas/cuantitativas:

- Polígono de frecuencias acumuladas

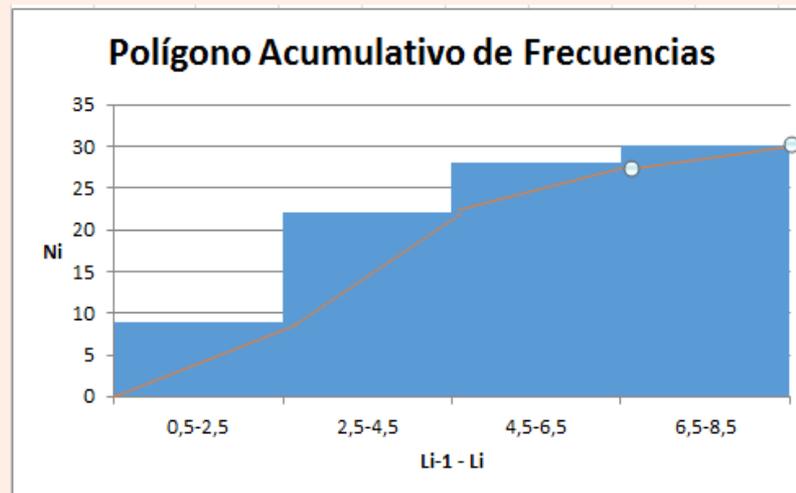
Datos agrupados

En cada valor que toma la variable se produce un incremento de la frecuencia acumulada.

En ordenadas pueden ir frecuencias absolutas o relativas (N_i o F_i)

Ejercicio 2.1

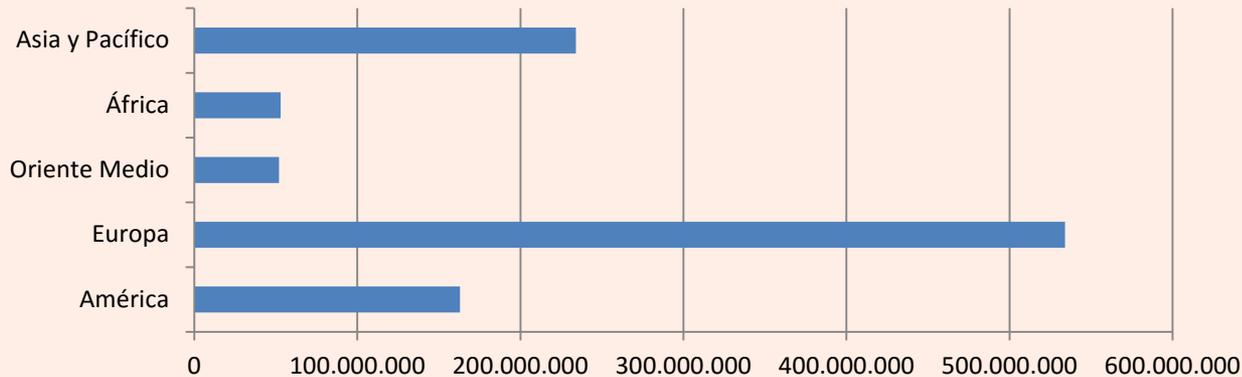
Li-1	Li	x_i	n_i	f_i	N_i	F_i	c_i	d_i
0,5	2,5	1,5	9	0,30	9	0,30	2	4,5
2,5	4,5	3,5	13	0,43	22	0,73	2	6,5
4,5	6,5	5,5	6	0,20	28	0,93	2	3
6,5	8,5	7,5	2	0,07	30	1,00	2	1
			30	1,00			8	0



Ejercicio 2.2. ejemplo de diagrama de barras horizontales

	número de turistas n_i	f_i	N_i	F_i
América	163.000.000	0,16	163.000.000	0,15733591
Europa	534.000.000	0,52	697.000.000	0,67277992
Oriente Medio	52.000.000	0,05	749.000.000	0,72297297
África	53.000.000	0,05	802.000.000	0,77413127
Asia y Pacífico	234.000.000	0,23	1.036.000.000	1
	1.036.000.000			

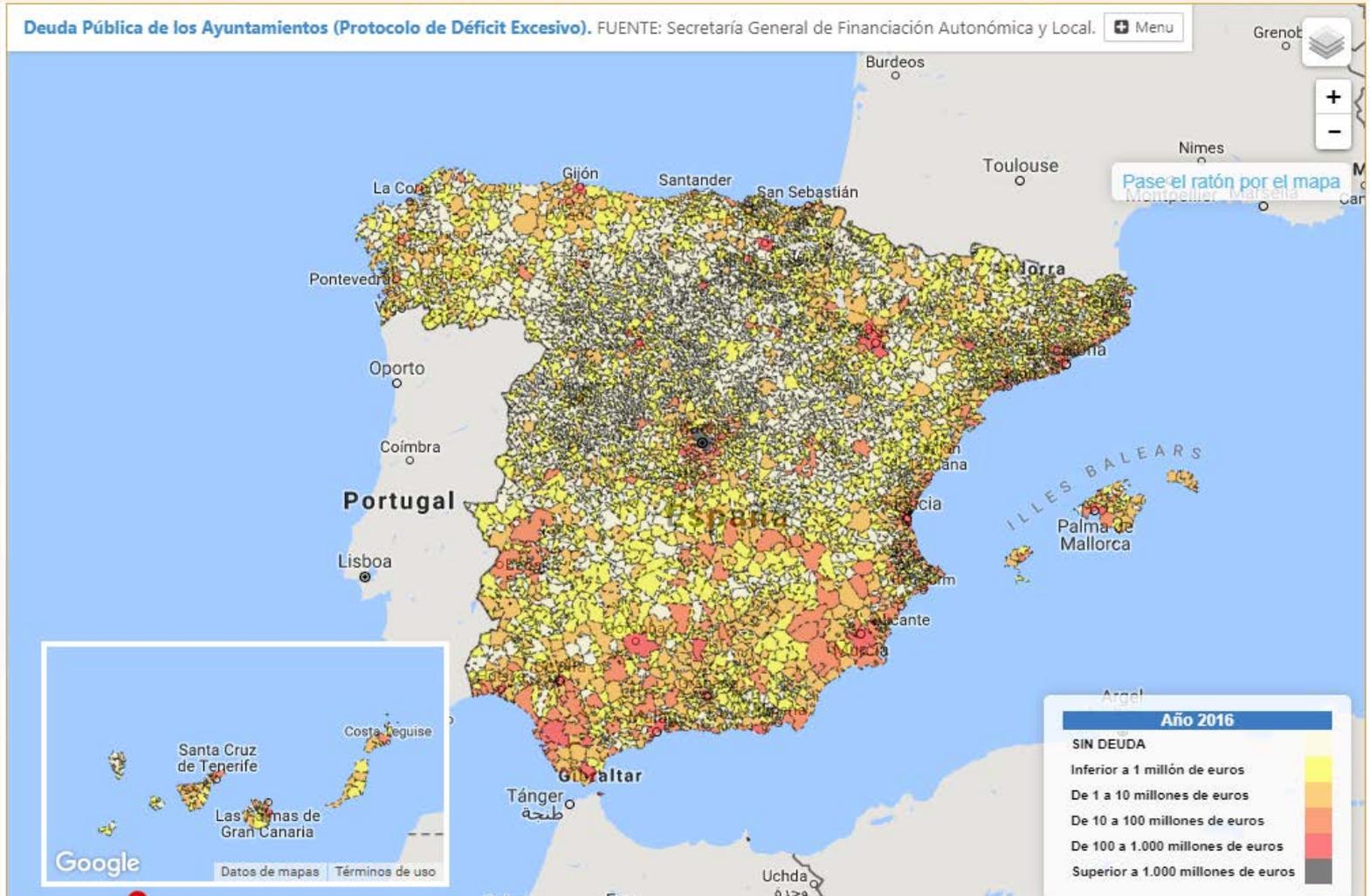
Número de turistas



Mapas



Mapas



infografías

Nombres

más frecuentes de los recién nacidos en España. 2016

Datos Provisionales

Nº total de nacimientos: **408.384**

Niños: 51,4 %

Niñas: 48,6%



Por comunidades autónomas



Gráficos tendenciosos...

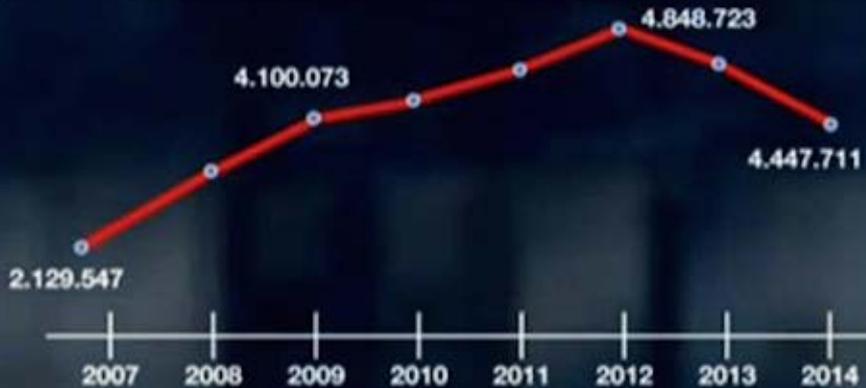


#elDBTeconomía

rtve.es

[D]

REGISTRO DESEMPLEO



rtve.es

FUENTE: Min **redifusión**

5.040.222

4.698.783

Febrero

Agosto

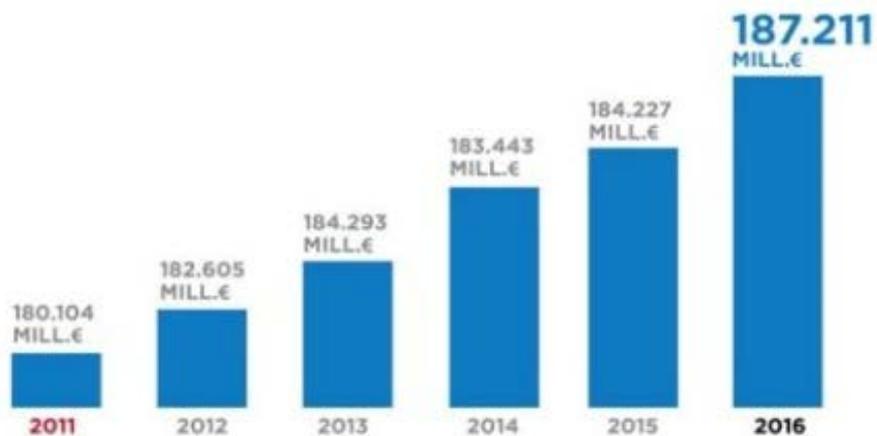
Paro registrado 2013

15:33

24h

GASTO SOCIAL

EDUCACIÓN, SANIDAD Y PROTECCIÓN SOCIAL



Venezolana de Televisión **VENEZUELA ¡de verdad!** Lunes, 15 de Abril 2013

Buscar en el Sitio

SEÑAL EN VIVO RADIOWEB VTV

INICIO NOTICIAS PROGRAMAS MULTIMEDIA OPINIÓN EL CANAL ELECCIONES HUGO CHÁVEZ

VOTO 2013 El pueblo decide

ELECCIÓN PRESIDENCIAL 2013

PORCENTAJES

50,66 % 49,07

Ganó el hijo de Chávez: Nicolás Maduro es el Presidente Electo de Venezuela (+Video)

ELECCIONES

La tendencia es irreversible, señaló Tibisay Lucena, presidenta del ente comicial / Maduro obtuvo 50.66% y Capriles 49.07%

Leer más

Minnée Mark en Míame

Manirua Conillas Barinani

Ganó el hijo de Chávez: Nicolás Maduro es el Presidente Electo de Venezuela (+Video)

Nicolás Maduro rechazó "falso" que le propusieron Capriles antes de conocerse resultados oficiales

Hechos sobre cuentas Twitter del Presidente Nicolás Maduro y del Psuv

Nicolás Maduro: Hoy es un día de historia, vamos con Cristo Redentor (oficial)

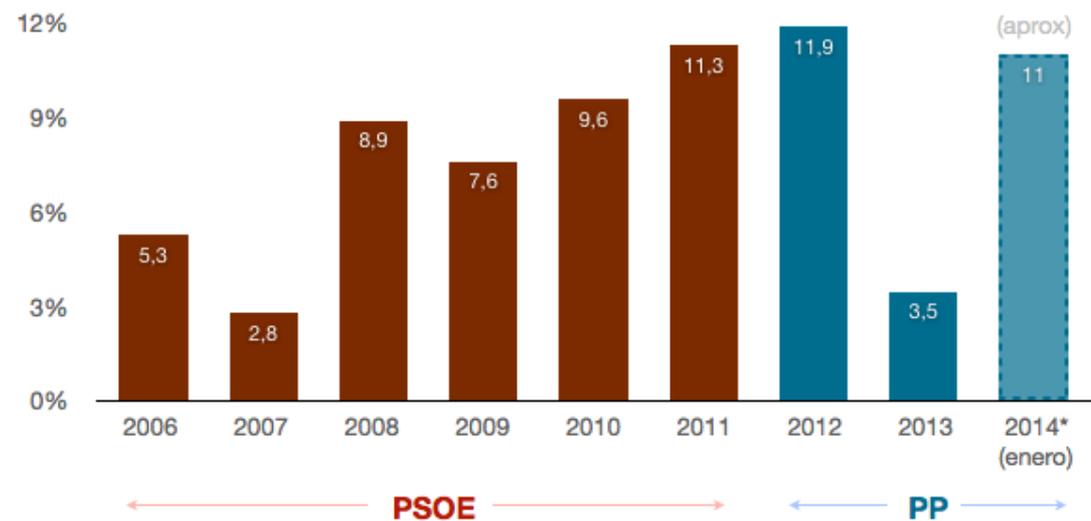
Fuente: Mala prensa

Variación del precio de la luz

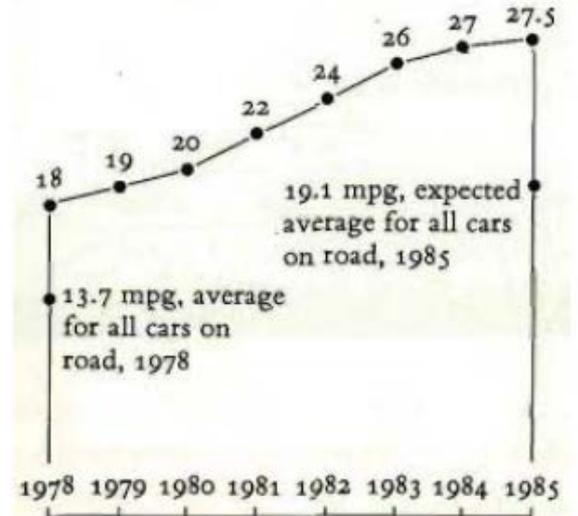


Variación del precio de la luz en España (%)

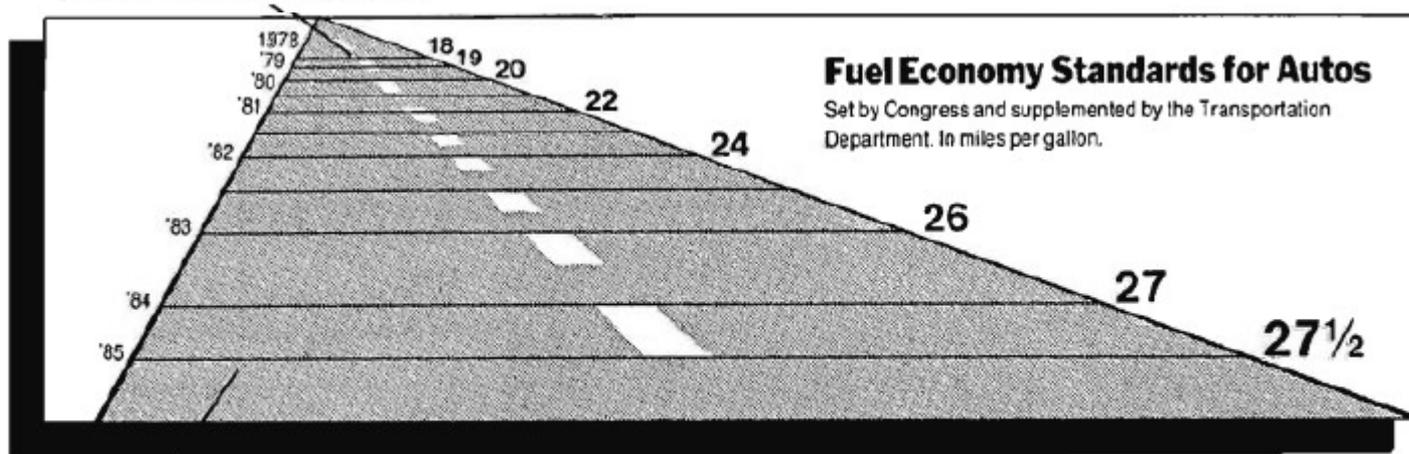
Por Kiko Llaneras — Politikon.es



REQUIRED FUEL ECONOMY STANDARDS:
NEW CARS BUILT FROM 1978 TO 1985

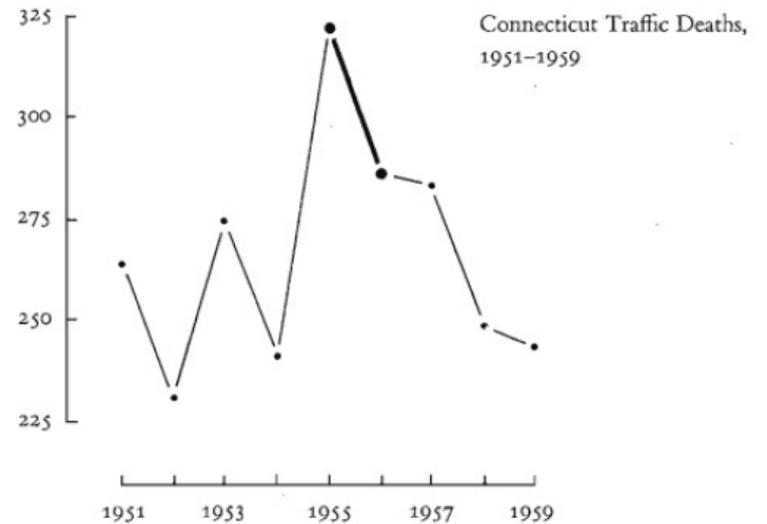
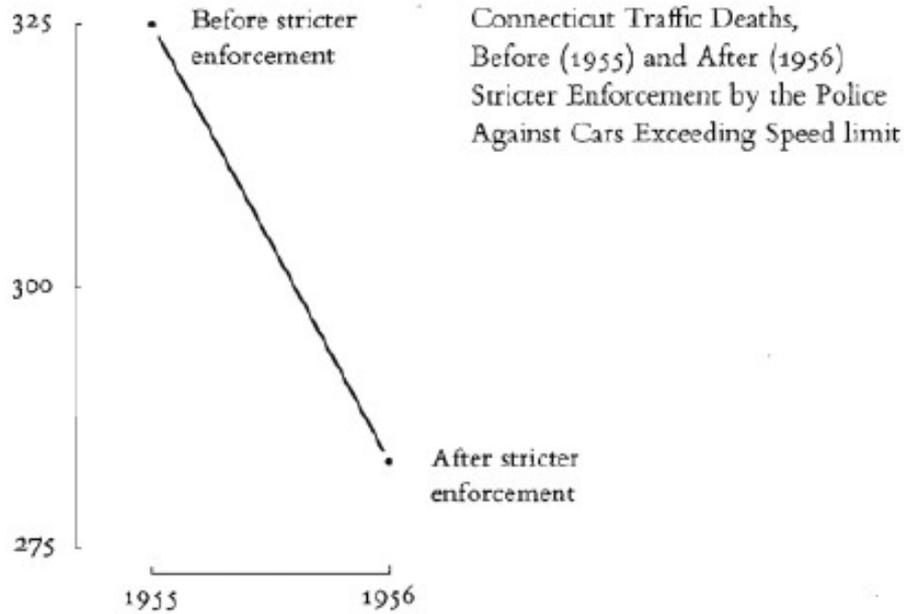


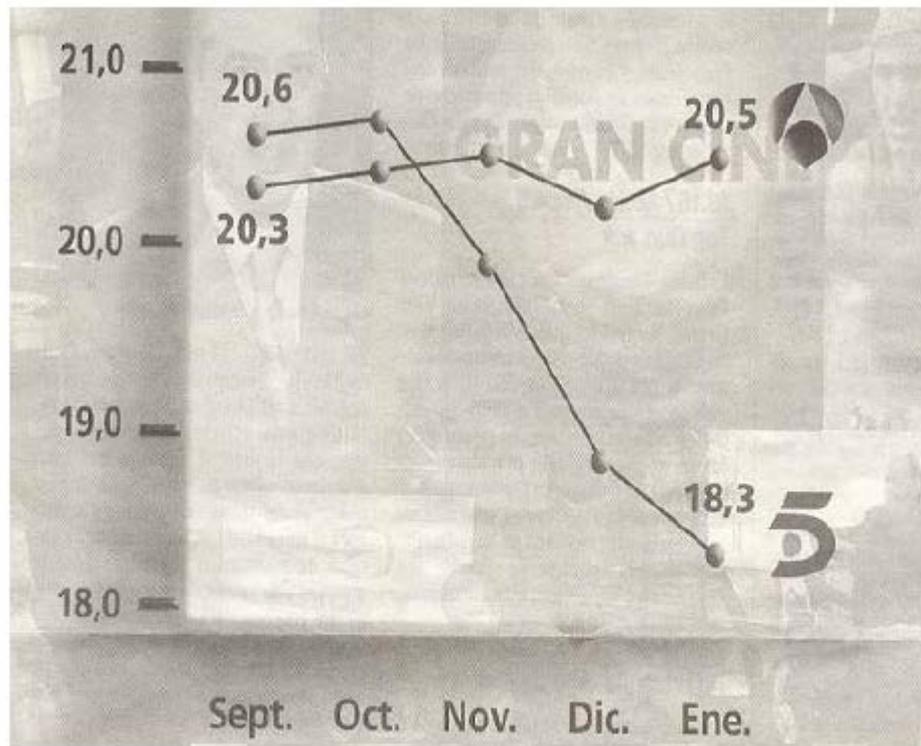
This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



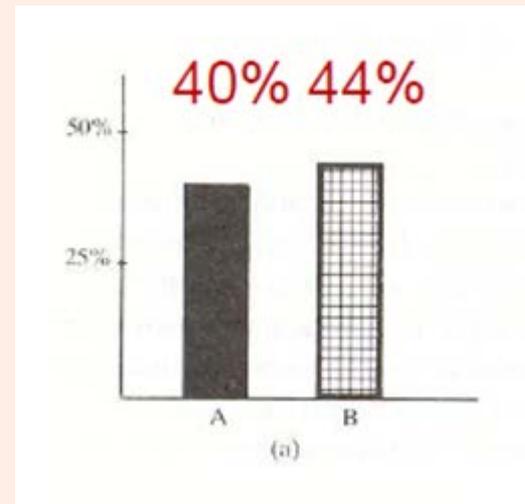
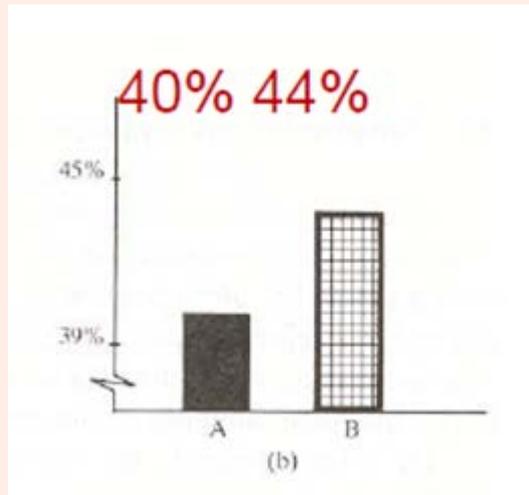
This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

Muertes por trafico





ANTENA 3
La cadena privada líder



Porcentajes de curación de una enfermedad con los fármacos A y B