2. DISTRIBUCIONES DE FRECUENCIAS

Estadística Descriptiva

Dr. Francisco Rabadán Pérez

Índice del tema

- 1. Introducción y notación
- 2. Distribuciones de frecuencias
  - *Absolutas y relativas*
  - *Acumuladas y acumuladas relativas*
- 3. Representaciones gráficas

1. Introducción y notación

- Idea: ordenar, y contar (resumir)
- i: variable de orden
- N: nº total de datos de la población
- n: Nº total de datos de la muestra
- x_i: frecuencias absolutas
- n_i: frecuencias absolutas
- N_i: frecuencias acumuladas
- f_i: frecuencia relativa
- F_i: frecuencia relativa acumulada

frecuencias	No acumuladas	Acumuladas
absoluta	n_i frecuencias absolutas	N_i frecuencias acumuladas
relativa	f_i frecuencia relativa	F_i frecuencia relativa acumulada

2.1. Frecuencia absoluta y relativa

- **Frecuencia absoluta (n_i)** o repetición: número de veces que se repite cada valor o dato de la variable.
- **Frecuencia relativa (f_i):** tanto por uno (o tanto por ciento) de ocasiones en que se repite cada valor de la variable.

  Por tanto:

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

frecuencias	No acumuladas
absoluta	n_i frecuencias absolutas
relativa	f_i frecuencia relativa

- Ver tabla pág. 15 (Martín-Pliego 2011)

2.2. Frecuencia acumulada absoluta y relativa

Frec.	Acum.
absoluta	N_i frecuencias acumuladas
relativa	F_i frecuencia relativa acumulada

- **Frecuencia Absoluta Acumulada (N_i)** : número de veces en que se da un valor x_i de la variable y valores inferiores a este.

$$N_c = \sum_{i=1}^{c} n_i = n_1 + n_2 + \cdots + n_c\,;c \in [1, n]$$

- **Frecuencia Relativa Acumulada (F_i)** : tanto por uno de ocasiones en que se da un valor x_i y valores inferiores a este.

$$F_c = \sum_{i=1}^{c} f_i = \sum_{i=1}^{c} \frac{n_i}{N} = f_1 + f_2 + \cdots + f_c\,;c \in [1, n]$$

Algunas curiosidades:

$\sum_{i=1}^{n} f_i = 1$

$f_i = \mathrm{F}_i - \mathrm{F}_{i-1}$

$N_n = N$

$\sum_{i=1}^{n} \mathrm{n}_i = N$

2.3. Distribución de frecuencias de una sola variable

- **Distribución de frecuencias:** conjunto de pares de valores formado por los distintos datos ordenados de menor a mayor y la frecuencia asociada a cada uno de ellos.
- **Valor de la variable** (x_i): número que representa a los datos observados o potenciales y que puede repetirse.

Renta (euros) x_i	Frec. absoluta n_i	Frec. relativa f_i	Frec. abs. acum. N_i	Frec. rel. acum. F_i
901	1	1/20	1	1/20
931	3	3/20	4	4/20
949	2	2/20	6	6/20
961	3	3/20	9	9/20
973	5	5/20	14	14/20
991	1	1/20	15	15/20
1 081	1	1/20	16	16/20
1 117	2	2/20	18	18/20
1 202	1	1/20	19	19/20
1 232	1	1/20	$20 = N$	$20/20 = 1$
	$N = 20$	1	—	—

Fuente: (Martín Pliego,2011, pág. 15)

2.4. Recorrido, rango, intervalos y marcas de clase

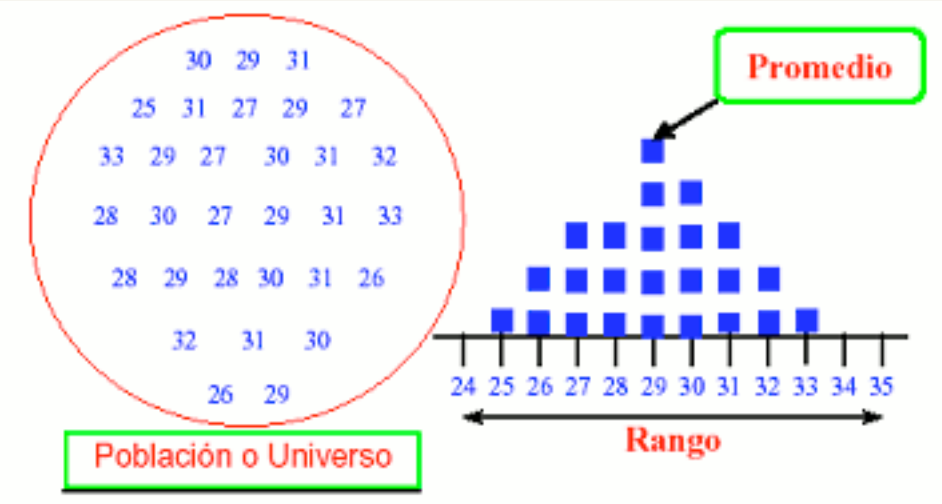
- **Recorrido** = valor máximo menos valor mínimo de la variable.

$$RE = R(x) = \max(x_i) - min(x_i)$$

- **Rango** = intervalo entre el valor mínimo y el valor máximo de la variable

$$Rango(x) = [\, min(x_i), \max(x_i)]$$

- Es la primera medida de dispersión que nos encontramos en la asignatura.



Fuente:
http://www.matematicasypoesia.com.es/Estadist/ManualCPE02p2.htm

2.4. Recorrido, rango, intervalos y marcas de clase

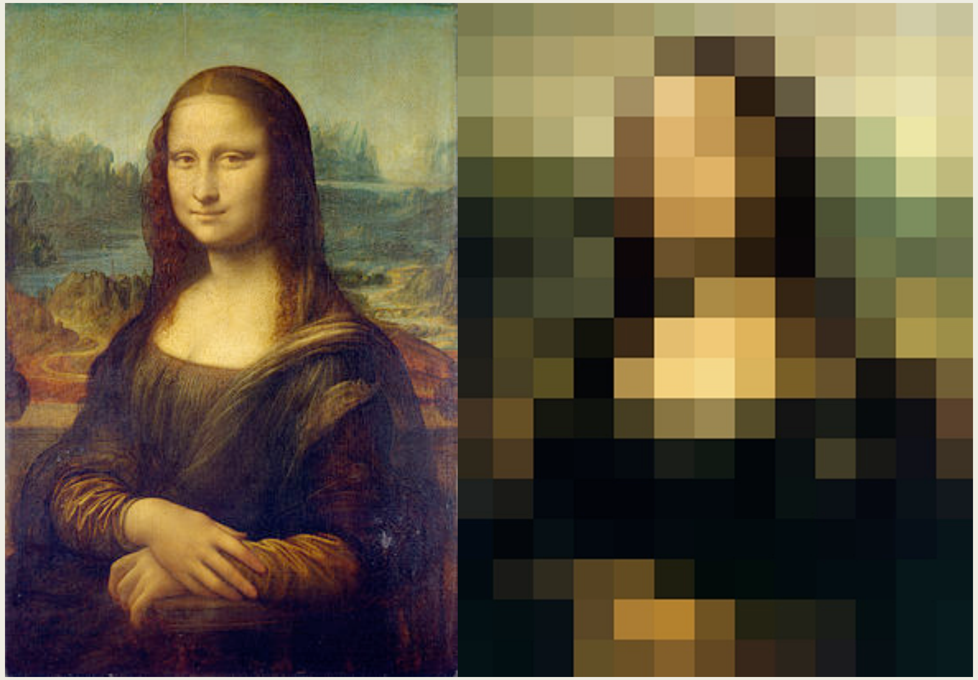
- **Al agrupar es importante tener en cuenta:**

– *perdemos información: no conocemos con precisión que ocurre en cada punto interior del intervalo (pixelamos).*

– *Se produce un **sesgo investigador*** (elegimos los intervalos que no tienen porque tener la misma amplitud)

– *Para el **análisis de datos no utilizaremos datos agrupados.***

– *Son útiles para **presentar la información** de forma fácilmente comprensible.*



Fuente: https://es.wikipedia.org/wiki/La_Gioconda

2.4. Recorrido, rango, intervalos y marcas de clase

$$Intervalo = [x_i, x_j] \,;\forall i,j \in [1,n]$$

- *Una vez que hemos decidido el par de valores $[x_i, x_j]$, pasaremos a notar el intervalo como $[L_i, L_j]$ siendo L_i el límite inferior y L_j el límite superior, o de* ***forma genérica*** $[\mathbf{L_i}, \mathbf{L_s}]$ *(limite inferior y superior)*

- **Amplitud de intervalo (c_i):** la diferencia entre los extremos del intervalo $L_{i-1} - L_i$
- Si los intervalos tienen la misma amplitud

Recorrido = nº intervalos * c_i

Edades	Enfermos	fa
50 – 54	10	105
45 – 49	5	95
40 – 44	12	90
35 – 39	9	78
30 – **34**	**22**	69*
25 – 29	15	47
20 – 24	11	32
15 – 19	7	21
10 - 14	4	14
5 – 9	10	10
	105	

Fuente:
http://www.monografias.com/trabajos66/estadistica-basica/estadistica-basica2.shtml

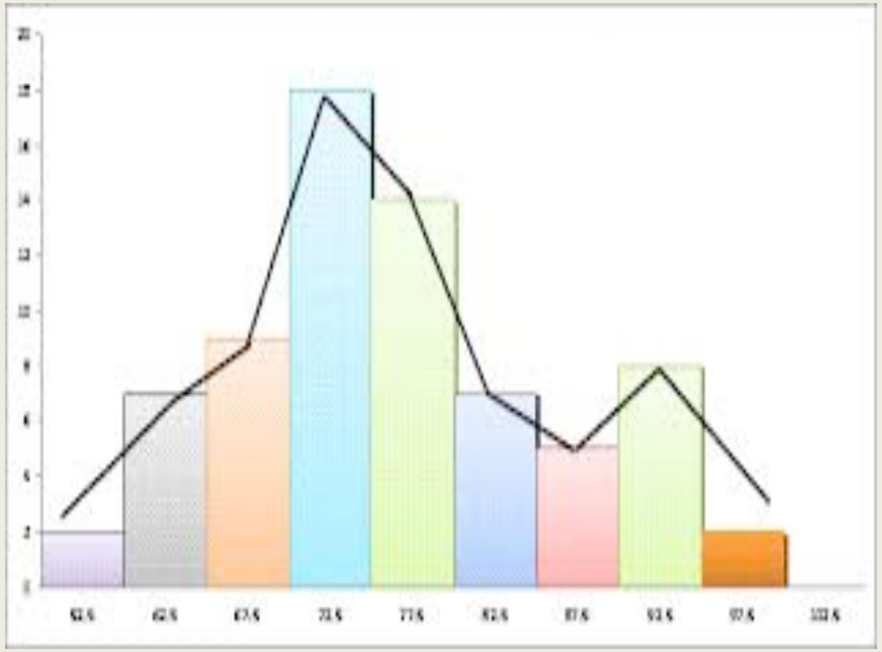
2.4. Recorrido, rango, intervalos y marcas de clase

- **Marca de clase (m_i):**
  - *es el punto central del intervalo*
  - *Pretende representar a todos los puntos del intervalo*
  - *y se calcula como*

$$x_i = \frac{L_{inf} + L_{sup}}{2}$$

- **Distribución de frecuencias:** Es el conjunto de intervalos con sus frecuencias asociadas.

Polígono de frecuencias (m_i)

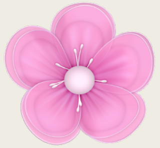


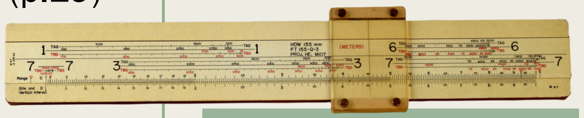
Fuente:

http://www.monografias.com/trabajos66/estadistica-basica/estadistica-basica2.shtml

3. Representaciones gráficas

- Tipos
  - Para fenómenos cualitativos
    - Diagramas de sectores
    - Diagrama de rectángulos
    - Cartogramas
    - Pictogramas
  - Para fenómenos cuantitativos:
    - para distribuciones no agrupadas
      - Diagrama de barras
      - Polígono de frecuencias
      - Diagrama de escalera (N_i)
    - Para distribuciones agrupadas
      - Histograma

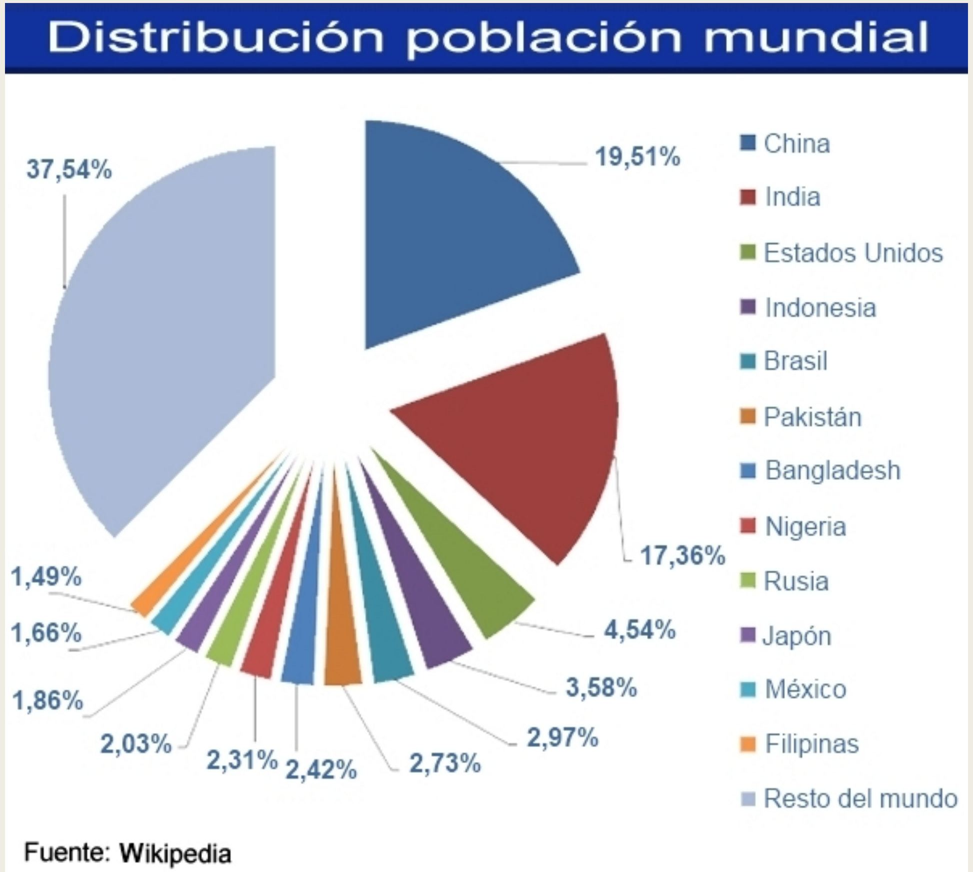


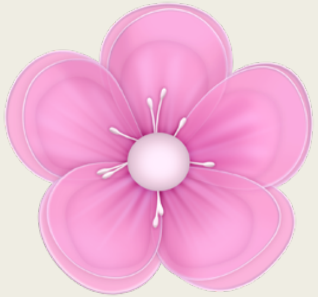


Martín Pliego, 2011 (p.19)

3. Representaciones gráficas

Diagrama de sectores





El tamaño de cada sector es proporcional a la frecuencia del valor, siendo su número de grados:

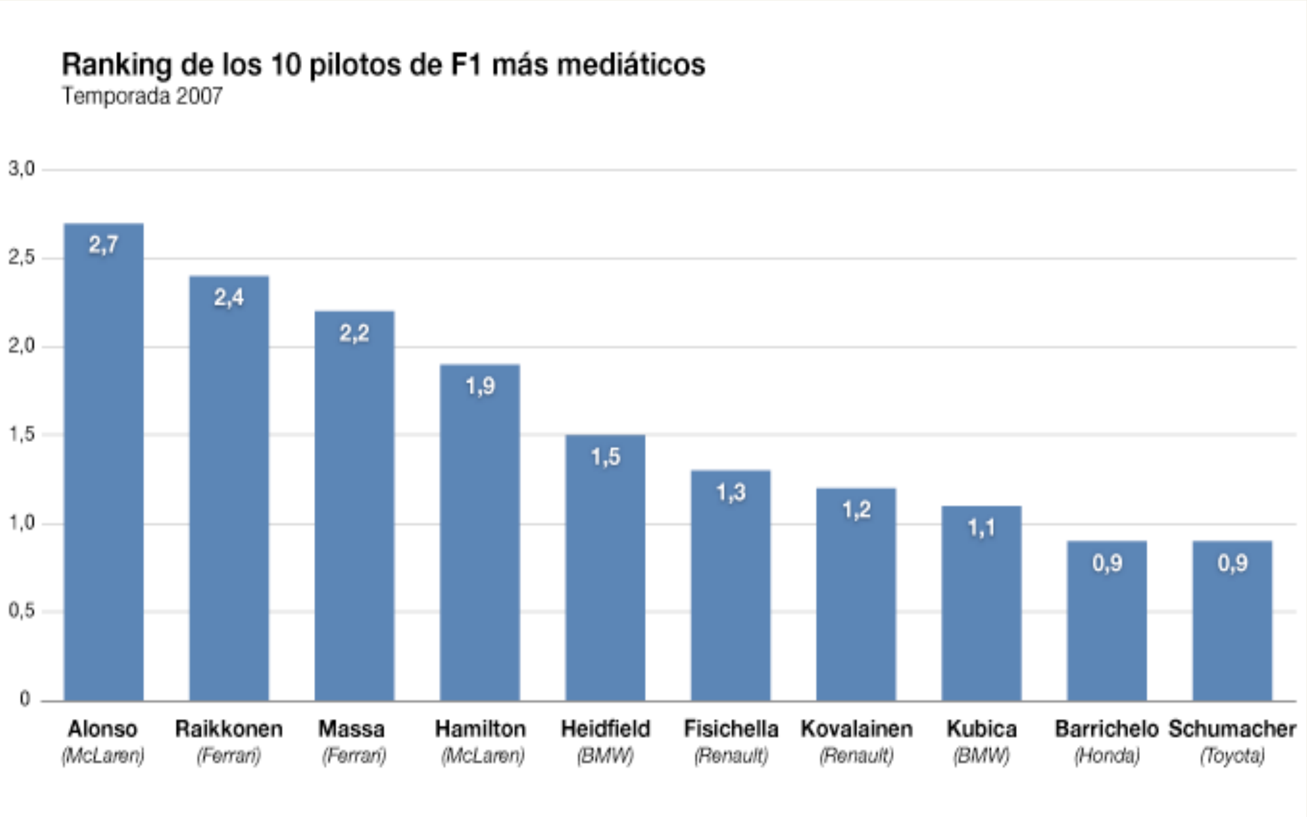
$$g = \frac{n_i}{N} \cdot 360^o = \frac{n_i}{N} \cdot 2\Pi$$

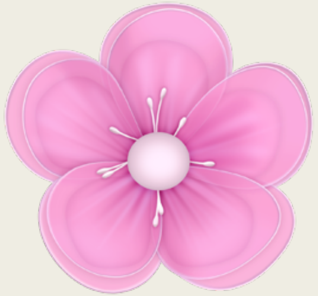
N es a 360º como n_i es a xº (regla de tres)

Fuente: http://www.iesmateoaleman.es/espa/act/unidad3/tema_2/contenido/ODE-03e4bbd3-a721-3391-bcdb-82310620501a/42_diagramas_de_barras.html

3. Representaciones gráficas

Diagrama de rectángulos



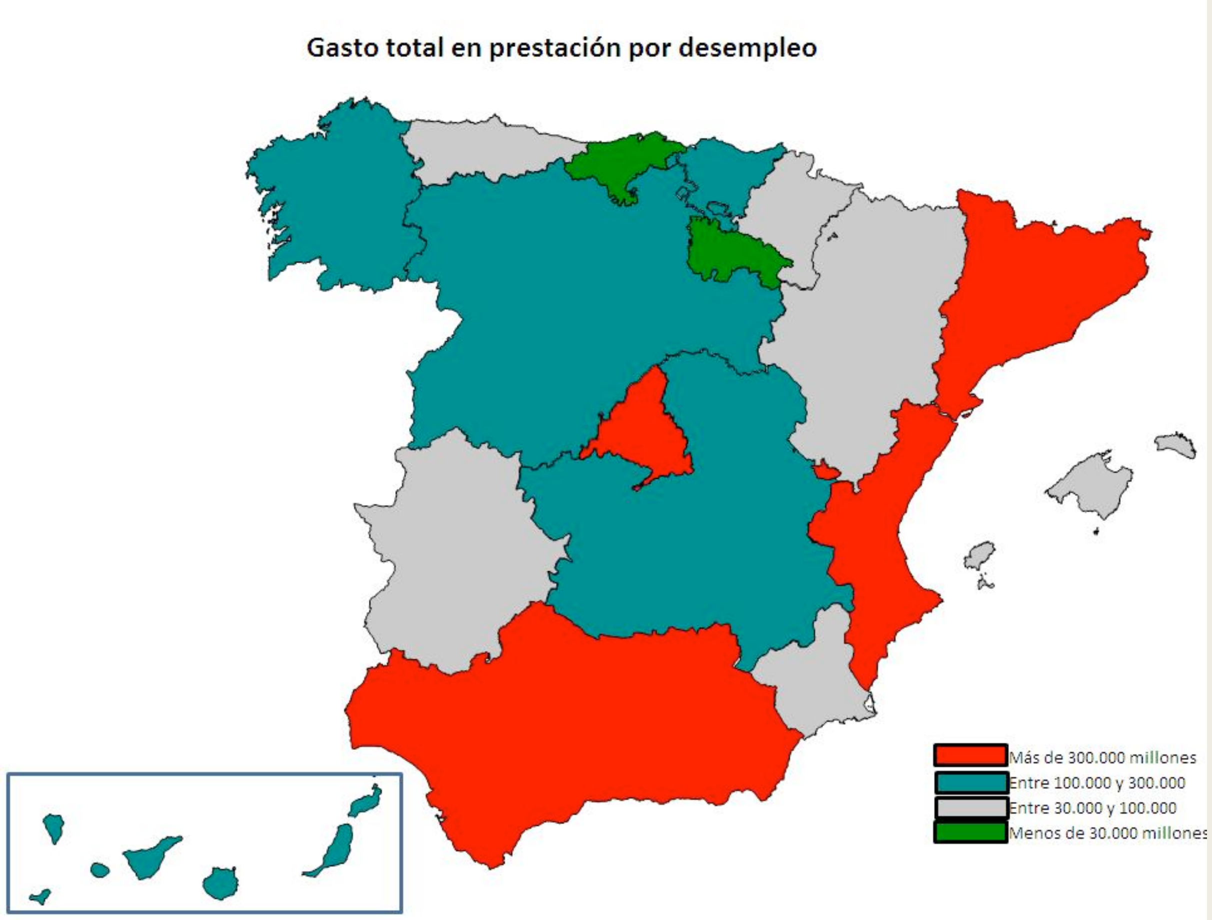


La altura de cada rectángulo coincide con la frecuencia absoluta n_i o relativa f_i

Fuente: http://www.iesmateoaleman.es/espa/act/unidad3/tema_2/contenido/ODE-03e4bbd3-a721-3391-bcdb-82310620501a/42_diagramas_de_barras.html

3. Representaciones gráficas

Cartograma

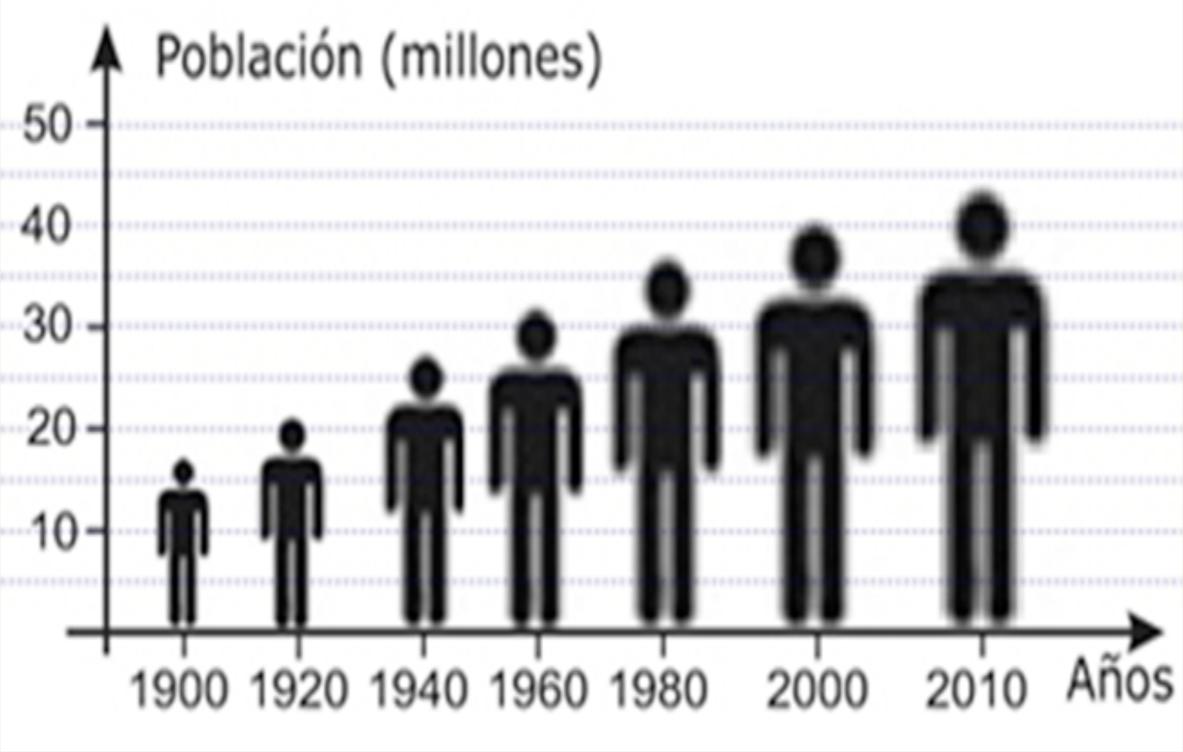


Fuente: http://analisisydecision.es/trucos-excel-mapa-de-espana-por-comunidades-autonomas/

- Cada color se identifica con un valor (x_i)
- Se puede utilizar la escala cromática para indicar intensidades

3. Representaciones gráficas

Pictograma



- El tamaño de la figura es proporcional a la frecuencia absoluta

Fuente: http://www.soystaff.com/2015/10/estadistica-descriptiva-graficas.html

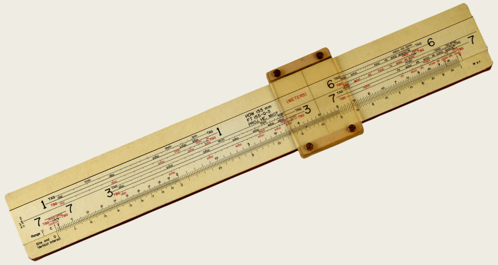
3. Representaciones gráficas

Diagrama de barras

- principio general: proporcionalidad de alturas o áreas respecto a las frecuencias absolutas.

3. Representaciones gráficas

Diagrama de barras



- Barra: línea gruesa que parte del valor xi en el eje de abscisas (x), y alcanza la altura de la frecuencia en el eje de ordenadas[f(x)].
- Hay cuatro tipos según el tipo de frecuencias: n_i, N_i, f_i, F_i

x_i	n_i	N_i	f_i	F_i
x_1	1	1	0,1	0,1
x_2	2	3	0,2	0,3
x_3	1	4	0,1	0,4
x_4	3	7	0,3	0,7
x_5	3	10	0,3	1,0

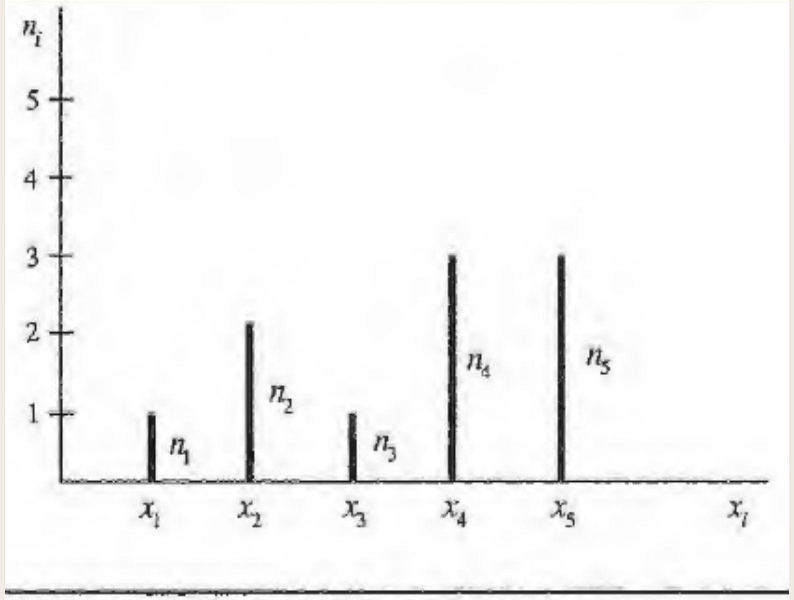
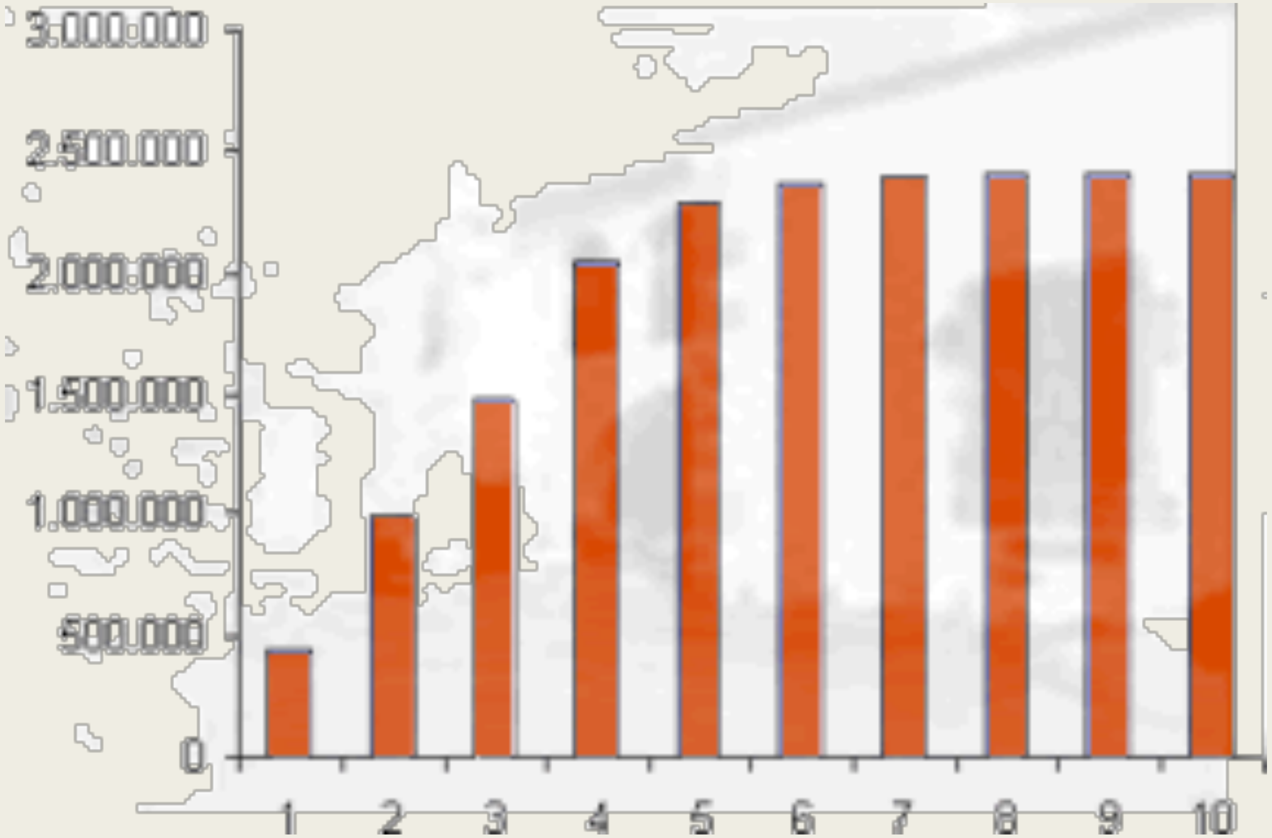


FIGURA 2.4. Diagrama de barras

Fuente: Martín Pliego, 2011 (p.22)

3. Representaciones gráficas

Diagrama de barras de frecuencias acumuladas



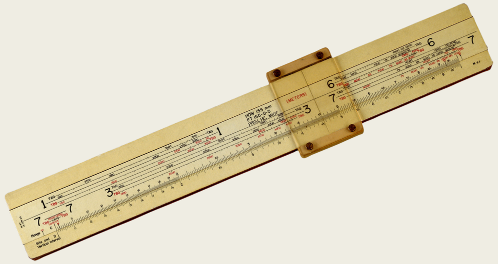
Fuente: http://www.ematematicas.net/graficas_estadistica.php?tipo=barras

En lo gráficos del software estadístico habitual se diferencia del diagrama de rectángulos en:

1. La variable es cuantitativa
2. El área del rectángulo no importa, sólo su altura (frecuencia).

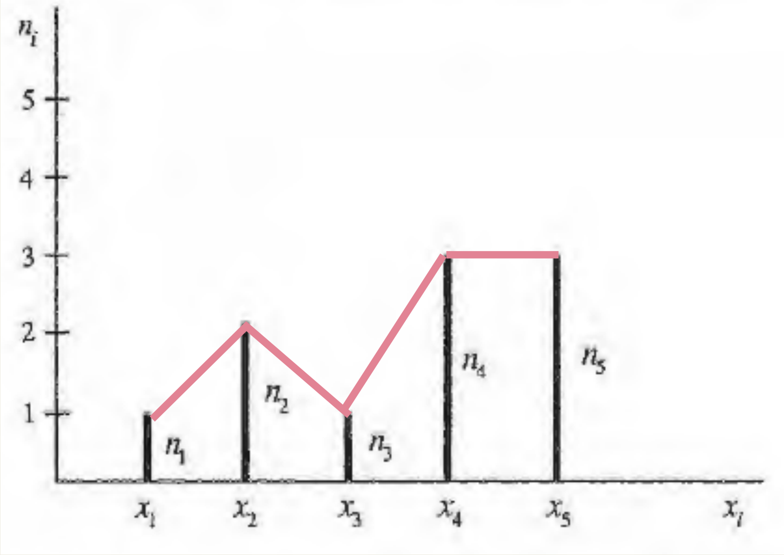
3. Representaciones gráficas

Polígono de frecuencias en datos no agrupados



- Bastaría con unir los puntos mas altos de las barras para conseguir el polígono de frecuencias.
- También habrá cuatro tipos según el tipo de frecuencias: n_i, N_i, f_i, F_i

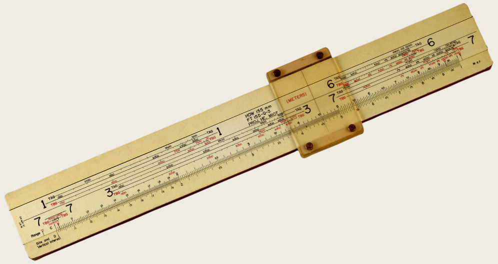
x_i	n_i	N_i	f_i	F_i
x_1	1	1	0,1	0,1
x_2	2	3	0,2	0,3
x_3	1	4	0,1	0,4
x_4	3	7	0,3	0,7
x_5	3	10	0,3	1,0



Esta línea forma el polígono de frecuencias

3. Representaciones gráficas

Polígono de frecuencias acumuladas



En los polígonos de frecuencias acumulados, en lugar de la marca de clase, se utiliza el extremo superior de cada intervalo para cargar la frecuencia acumulada.

$L_{i-1}-L_i$	n_i	N_i
L_0-L_1	n_1	N_1
L_1-L_2	n_2	N_2
L_2-L_3	n_3	N_3
$\vdots$	$\vdots$	$\vdots$
$L_{i-1}-L_i$	n_i	N_i
$\vdots$	$\vdots$	$\vdots$

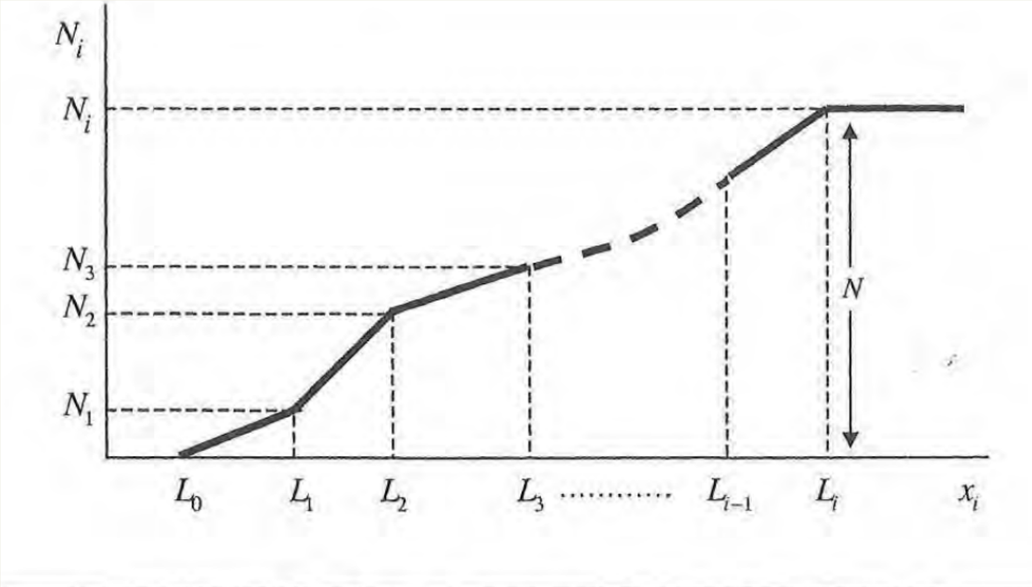


FIGURA 2.7. Polígono de frecuencias

Fuente: Martín Pliego, 2011 (p.25)

3. Representaciones gráficas

Diagrama de Escalera

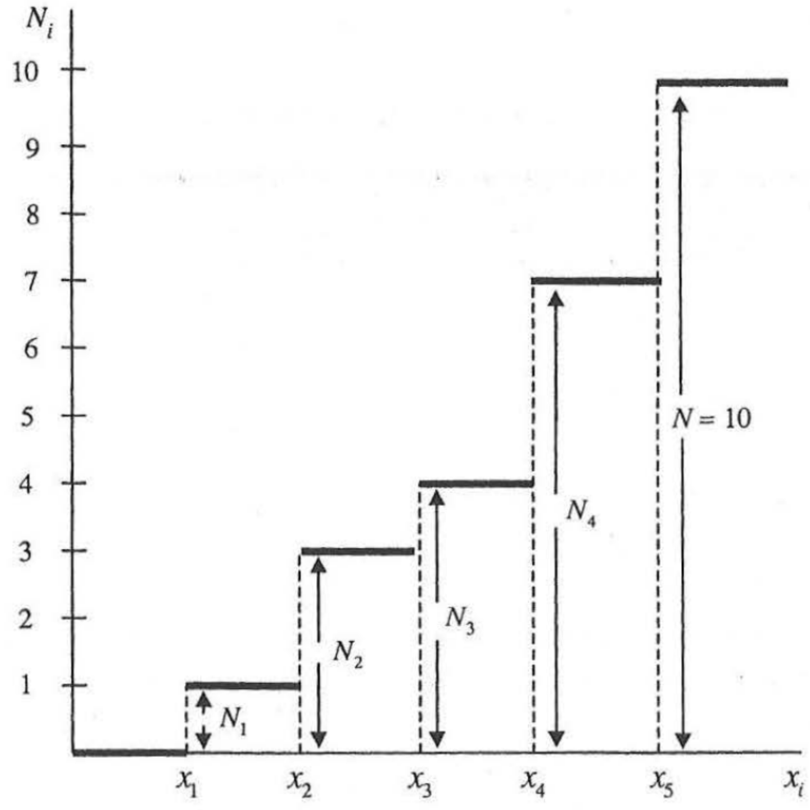
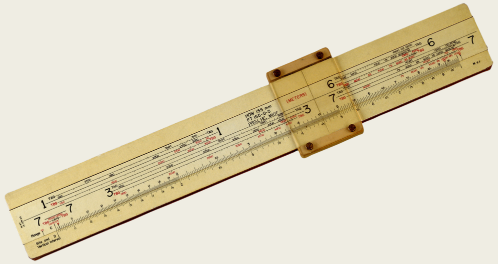


FIGURA 2.5. Diagrama de escalera



- Se utiliza para frecuencias acumuladas o para variables que varían de forma constante a lo largo de periodos.
- Es especialmente interesante en el caso de las frecuencias relativas acumuladas.

Fuente: Martín Pliego, 2011 (p.23)

2.5. Representaciones gráficas

En datos agrupados

- Si utilizamos la marca de clase podríamos construir diagramas de barras o polígonos de frecuencias, pero:
  - **No seríamos del todo "<u>sinceros</u>"** porque sustituimos un intervalo por un solo punto.
- Lo correcto es considerar <u>el intervalo como un todo</u>:
  - el intervalo será la base del tramo del polígono, o del histograma
  - El área será representativa de la frecuencia. Esto nos hace <u>transformar las frecuencias en densidades</u> *(ponderamos la altura del rectángulo respecto a la longitud del intervalo)*

Densidad de frecuencia

$$\boxed{d_i = \frac{n_i}{c_i}}$$

C_i: amplitud de intervalo

$C_i = L_{i+1} - L_i$

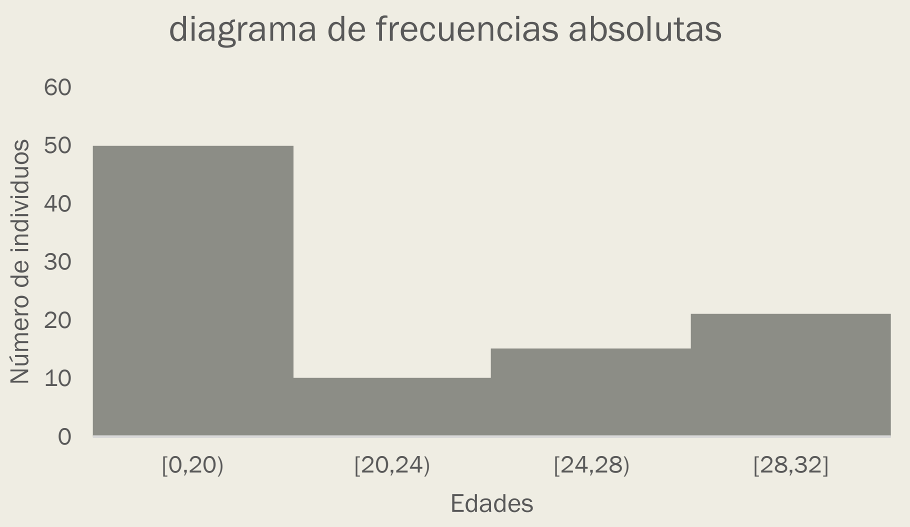
3. Representaciones gráficas

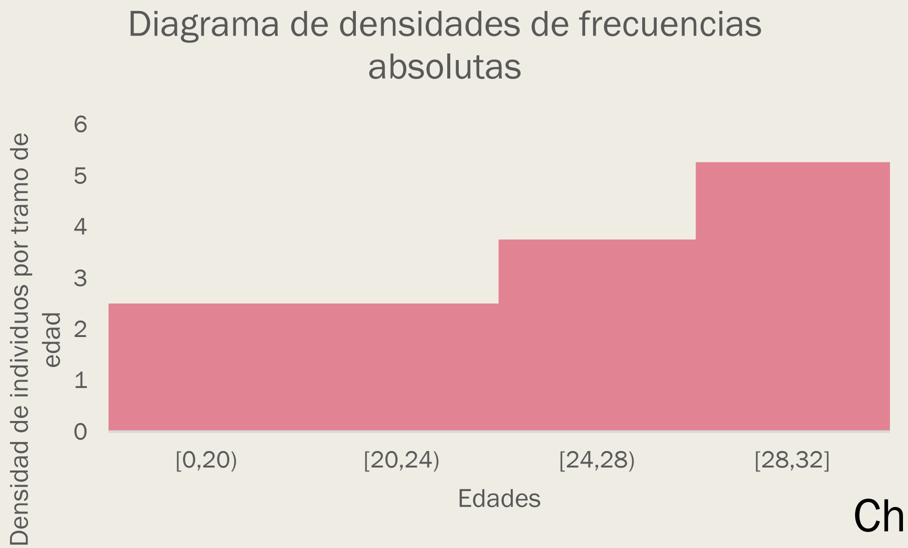
Comprendiendo la densidad de frecuencia

$$d_i = \frac{n_i}{c_i}$$

L_i	L_{i+1}	n_i	c_i	d_i
0	20	50	20	2,5
20	24	10	4	2,5
24	28	15	4	3,75
28	32	21	4	5,25

- X: Edad de un conjunto de individuos
- Si no trabajamos con densidades "*da la sensación*" de que entre 0 y 20 años hay comparativamente más individuos que entre 20 y 28 por tramo de edad (ej: 4 años)





3. Representaciones gráficas

Densidad de frecuencia

$$\boxed{d_i = \frac{n_i}{c_i}}$$

C_i: amplitud de intervalo

Tabla de datos agrupados

L_{i-1} - L_i	n_i	c_i	d_i
L_0 - L_1	n_1	c_1	d_1
L_1 - L_2	n_2	c_2	d_2
L_2 - L_3	n_3	c_3	d_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$
L_{i-1} - L_i	n_i	c_i	d_i
$\vdots$	$\vdots$	$\vdots$	$\vdots$

Histograma

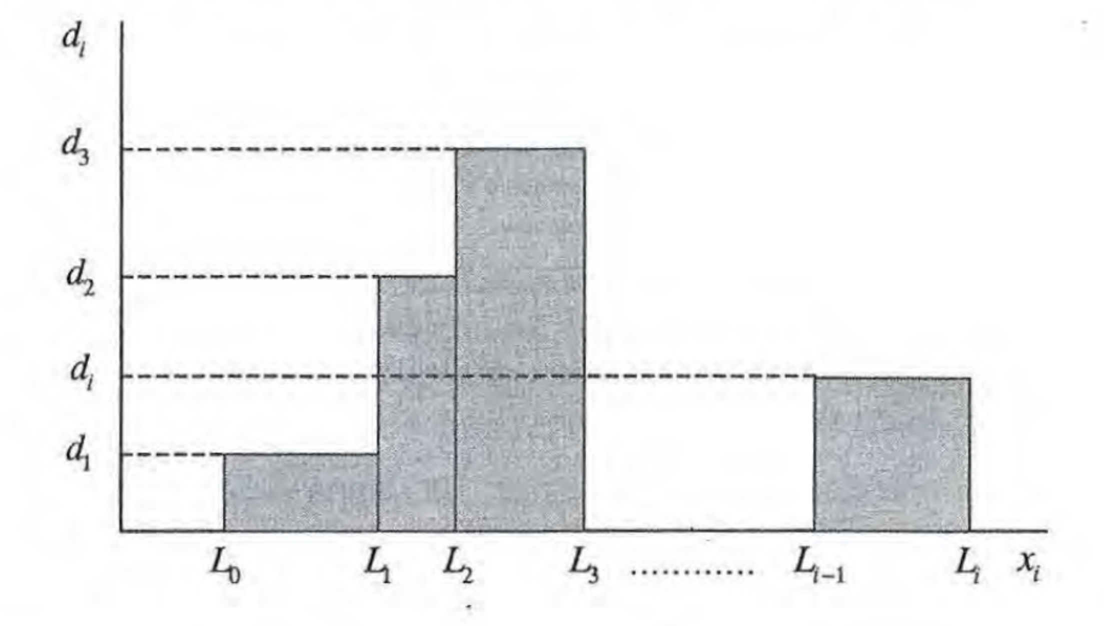


FIGURA 2.6. Histograma de frecuencias

Fuente: Martín Pliego, 2011 (p.24)

Apéndice

EL OPERADOR *SUMA* O SUMATORIO

Una variable es un símbolo matemático que representa a un conjunto de valores. Si este conjunto tiene un número infinito de valores, la variable se representa por el símbolo X. Si, por el contrario, tenemos un conjunto finito representaremos la variable X_i. De una variable obtenemos n observaciones

$$x_1, x_2, x_3, \ldots, x_n .$$

Se llama suma o sumatorio de x_i para i desde 1 hasta n a:

$$x_1 + x_2 + x_3 + \cdots + x_n = \sum_{i=1}^{n} x_i .$$

El operador Σ representa, pues, una suma y goza, por lo tanto, de todas las propiedades de ésta. No obstante, vamos a desarrollar algunas de ellas:

(*a*) La suma de una constante desde 1 hasta n es n veces la constante.

En efecto,

$$\sum_{i=1}^{n} k = k + k + \cdots + k = n \cdot k .$$

(*b*) La suma de una constante por una variable es igual a la constante por la suma de la variable

$$\sum_{i=1}^{n} kx_i = kx_1 + kx_2 + \cdots + kx_n = k(x_1 + x_2 + \cdots + x_n) = k\sum_{i=1}^{n} x_i .$$

(*c*) La suma desde 1 hasta n de una suma de variables es la suma de las sumas desde 1 hasta n de cada una de las variables

$$\sum_{i=1}^{n} (x_i + y_i) = (x_1 + y_1) + \cdots + (x_n + y_n) =$$

$$= (x_1 + x_2 + \cdots + x_n) + (y_1 + y_2 + \cdots + y_n) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i .$$

Fuente: Martín Pliego, 2011 (p.33-34)

Apéndice

EL OPERADOR *PRODUCTO*

El producto de los diferentes valores de x_i para i desde 1 hasta n se representa por el siguiente operador, llamado **producto**

$$\prod_{i=1}^{n} x_i = x_1 \cdot x_2 \cdot x_3 \cdots x_n .$$

El operador Π representa, pues, un producto y goza de las propiedades de éste:

(*a*) El producto desde 1 hasta n de una constante es la constante elevada a n

$$\prod_{i=1}^{n} k = k \cdot k \cdot k \cdots k = k^n .$$

(*b*) El producto de una constante por una variable es igual a la constante elevada a n por el producto de la variable

$$\prod_{i=1}^{n} kx_i = kx_1 \cdot kx_2 \cdots kx_n = k \cdot k \cdots k \cdot x_1 \cdot x_2 \cdots x_n = k^n \prod_{i=1}^{n} x_i .$$

(*c*) El logaritmo de un producto es la suma de los logaritmos de cada uno de los valores

$$\log \prod_{i=1}^{n} x_i = \log (x_1 \cdot x_2 \cdots x_n) =$$

$$= \log x_1 + \log x_2 + \cdots + \log x_n = \sum_{i=1}^{n} \log x_i .$$

Fuente: Martín Pliego, 2011 (p.33-34)

Textos recomendados

- Martín-Pliego, *Introducción a la Estadística Económica y Empresarial*, Editorial AC, 2011, 3ª Edición

Prácticas recomendadas

- Ejercicio 1, pág. 29 (Martín-Pliego, 2011)
- Ver recursos en aula virtual.
- Prácticas y ejercicios resueltos en clase.