



Descripción de variables cuantitativas

Para datos cuantitativos tanto para datos cualitativos, el primer paso es construir una tabla de frecuencias Si los datos son discretos, es fácil

MUESTRA: 60 madrileños adultos

VARIABLE: Número de veces que han votado en las elecciones de la comunidad

OBJETIVO: **Clasificación y representación de la información**

3 3 3 4 1 2 4 5 2 3 1 1 3 8 4 1 3 4 2 5 0 0 5 4 2 1 2 3 3 2

1 4 3 2 3 5 0 6 3 1 3 5 4 1 4 1 2 4 4 3 3 0 7 2 2 1 3 4 2 2



La tabla de frecuencias

| Número de veces votado | Frecuencia absoluta |
|------------------------|---------------------|
| 0 | 4 |
| 1 | 10 |
| 2 | 12 |
| 3 | 15 |
| 4 | 11 |
| 5 | 5 |
| 6 | 1 |
| 7 | 1 |
| 8 | 1 |
| >8 | 0 |
| Total | 60 |

Se puede
incluir una
clase vacía





Ahora podemos incluir frecuencias acumuladas en la tabla de frecuencias

| Veces votado | Frecuencia absoluta | Frecuencia acumulada | Frecuencia relativa | Frecuencia acumulada relativa |
|--------------|---------------------|----------------------|---------------------|-------------------------------|
| 0 | 4 | 4 | 0,0667 | 0,0667 |
| 1 | 10 | 4+10 = 14 | 0,1667 | 14/60 = 0,2333 |
| 2 | 12 | 4+10+12 = 26 | 0,2000 | 0,4333 |
| 3 | 15 | 41 | 0,2500 | 0,6833 |
| 4 | 11 | 52 | 0,1833 | 0,8667 |
| 5 | 5 | 57 | 0,0833 | 0,9500 |
| 6 | 1 | 58 | 0,0167 | 0,9667 |
| 7 | 1 | 59 | 0,0167 | 0,9833 |
| 8 | 1 | 60 | 0,0167 | 1,0000 |
| >8 | 0 | 60 | 0,0000 | 1,0000 |
| Total | 60 | — — | 1,0000 | — — |



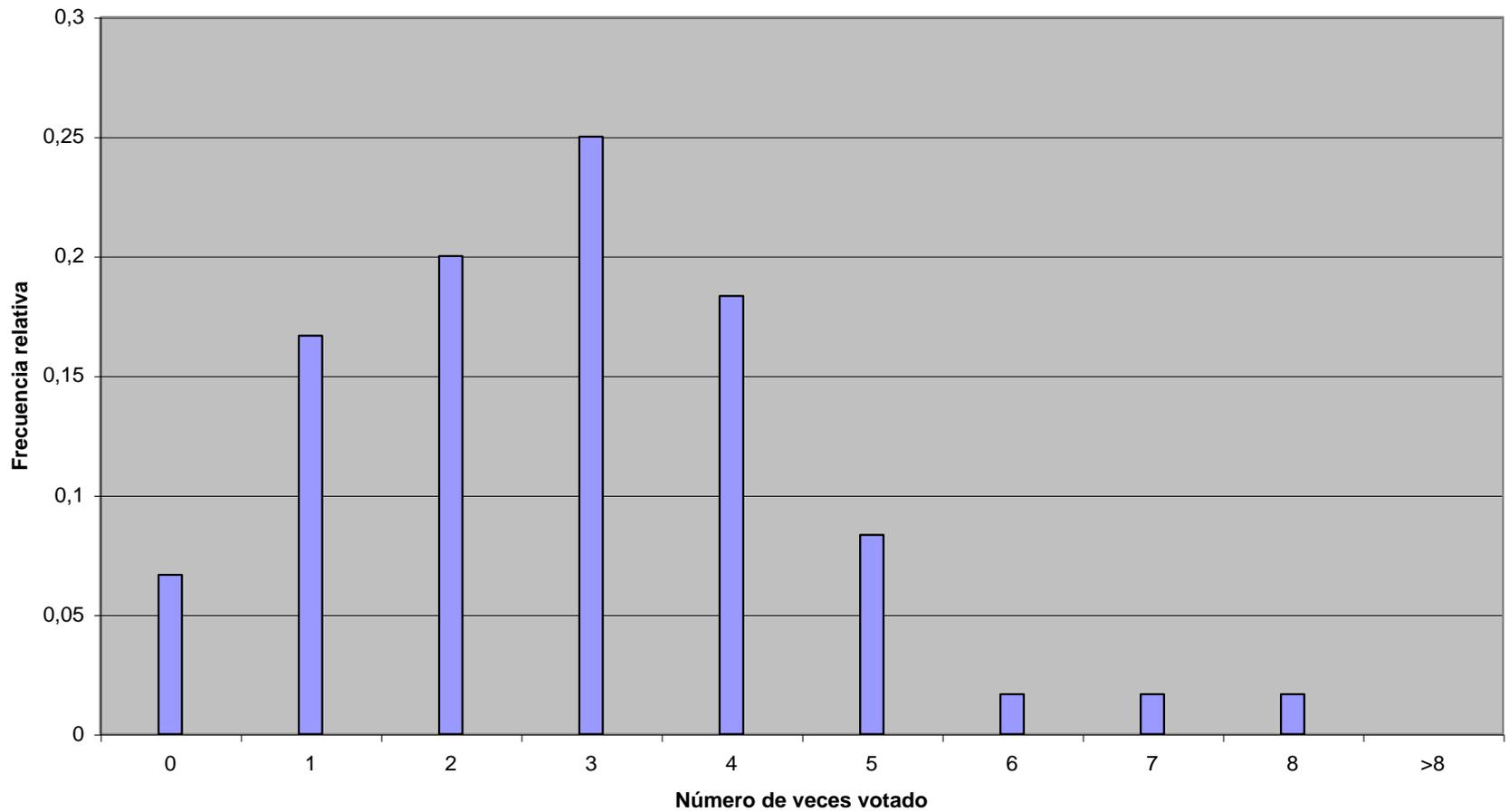
Formato general de la tabla

| Valor (x_i) | n_i | N_i | f_i | F_i |
|-----------------|----------|-------------------|----------|-------------------|
| x_1 | n_1 | $N_1 = n_1$ | f_1 | $F_1 = f_1$ |
| x_2 | n_2 | $N_2 = N_1 + n_2$ | f_2 | $F_2 = F_1 + f_2$ |
| x_3 | n_3 | $N_3 = N_2 + n_3$ | f_3 | $F_3 = F_2 + f_3$ |
| \vdots | \vdots | \vdots | \vdots | \vdots |
| x_k | n_k | $N_k = N$ | f_k | $F_k = 1$ |
| Total | N | — | 1 | — |



El diagrama de barras

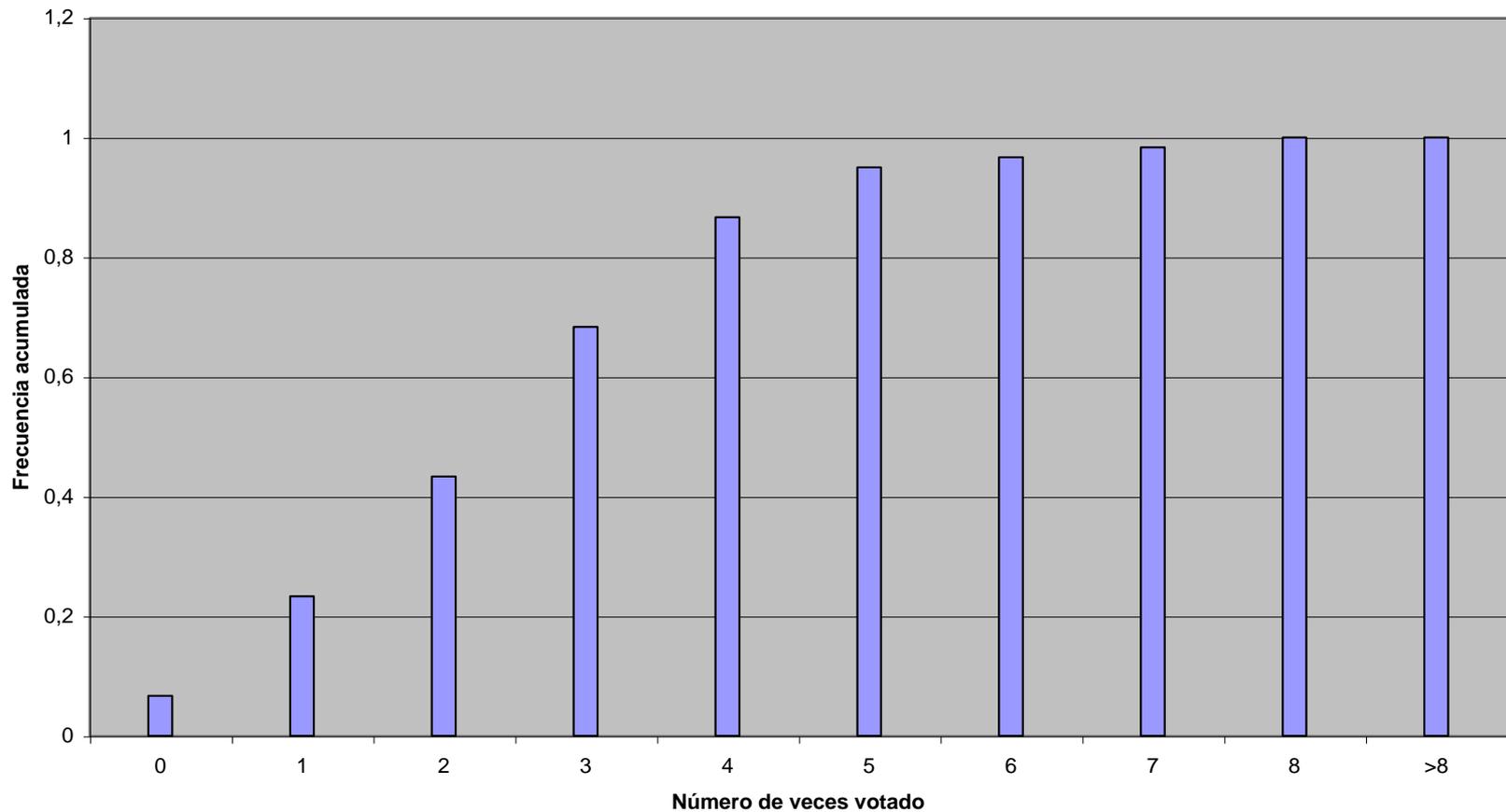
Las veces que votan los madrileños





El diagrama de barras con frecuencias acumuladas

Frecuencias acumuladas del número de veces que votan los madrileños





Como construir una tabla de frecuencias con datos continuos

Ingresos y Derechos liquidados para Operaciones corrientes en 36 municipios de Madrid en 1995 (miles de PTAS)

| | | | | | | | | |
|--------|-------|-------|-------|--------|--------|-------|-------|-------|
| 114579 | 73896 | 59003 | 86165 | 53428 | 93844 | 61536 | 90628 | 49501 |
| 56767 | 78063 | 87750 | 82409 | 107664 | 60479 | 88872 | 66325 | 78268 |
| 38360 | 82436 | 83531 | 81364 | 63210 | 112842 | 56206 | 59052 | 52660 |
| 45000 | 91562 | 66308 | 50397 | 79964 | 65369 | 71803 | 60108 | 49264 |

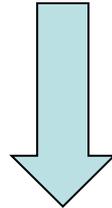
N=36

Mínimo = 38360

Máximo = 114579

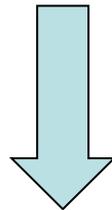


Un diagrama de barras para cada posible valor no tiene sentido



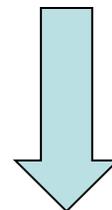
Tenemos que agrupar los datos

¿Cuántos intervalos usamos?



Más o menos \sqrt{N}

¿Dónde empezamos y qué anchura de intervalo utilizamos?



Números más o menos redondos



La tabla de frecuencias

Podemos
incluir una
clase vacía



Datos menores o iguales a 45

Datos mayores de 45

| Ingresos y Derechos liquidados (millones de PTAS) | Frecuencia absoluta |
|--|---------------------|
| ≤ 30 | 0 |
| $(30,45]$ | 2 |
| $(45,60]$ | 9 |
| $(60,75]$ | 9 |
| $(75,90]$ | 10 |
| $(90,105]$ | 3 |
| $(105,120]$ | 3 |
| > 120 | 0 |
| Total | 36 |



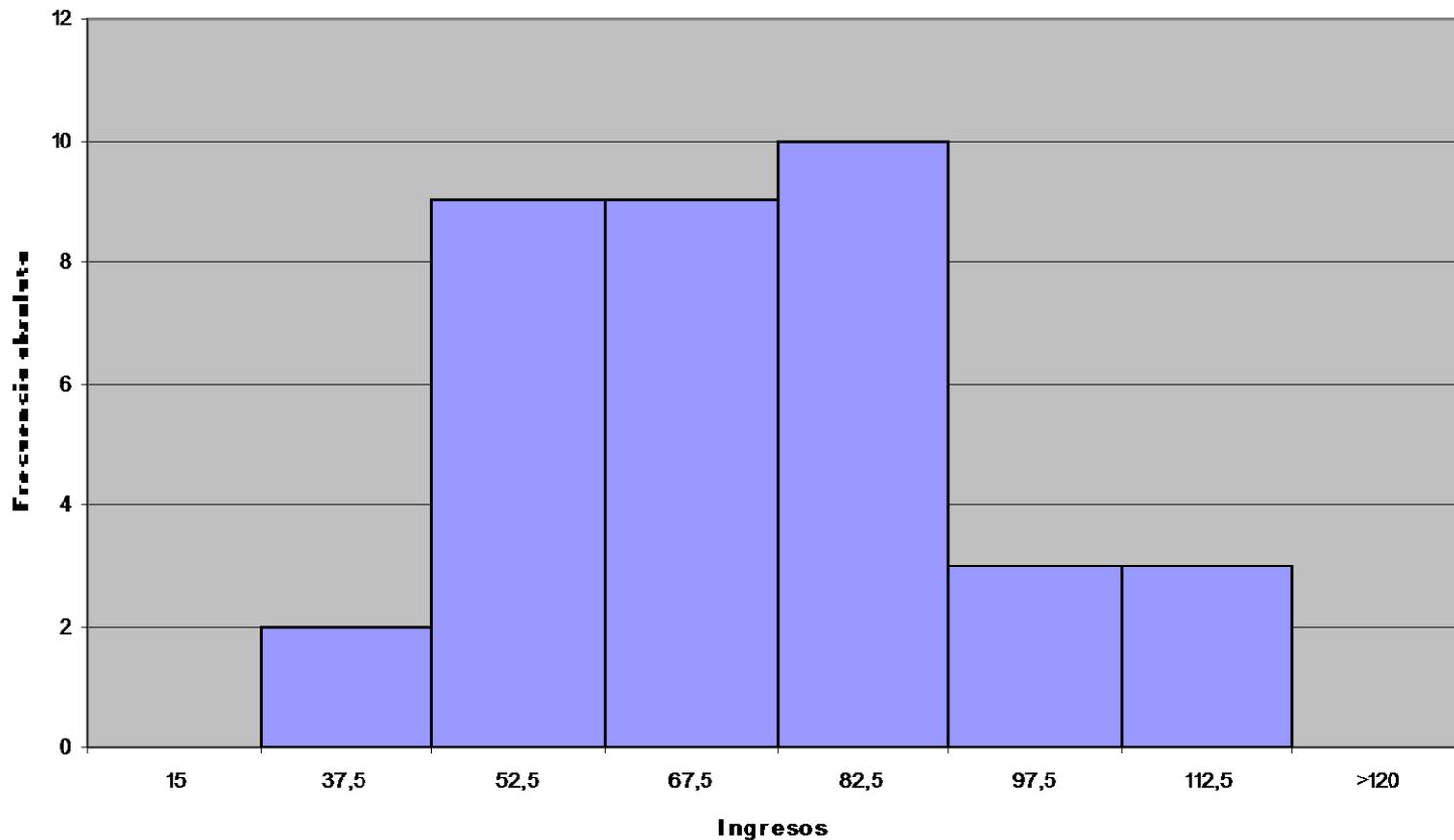
Formato general de la tabla

| Intervalo | Centro (x_i) | n_i | N_i | f_i | F_i |
|------------------|---------------------------------|----------|-----------|----------|-----------|
| $(a_0, a_1]$ | $x_1 = \frac{1}{2} (a_0 + a_1)$ | n_1 | N_1 | f_1 | F_1 |
| $(a_1, a_2]$ | $x_2 = \frac{1}{2} (a_1 + a_2)$ | n_2 | N_2 | f_2 | F_2 |
| $(a_2, a_3]$ | x_3 | n_3 | N_3 | f_3 | F_3 |
| \vdots | \vdots | \vdots | \vdots | \vdots | \vdots |
| $(a_k, a_{k+1}]$ | x_k | n_k | $N_k = N$ | f_k | $F_k = 1$ |
| Total | — | N | — | 1 | — |



El histograma

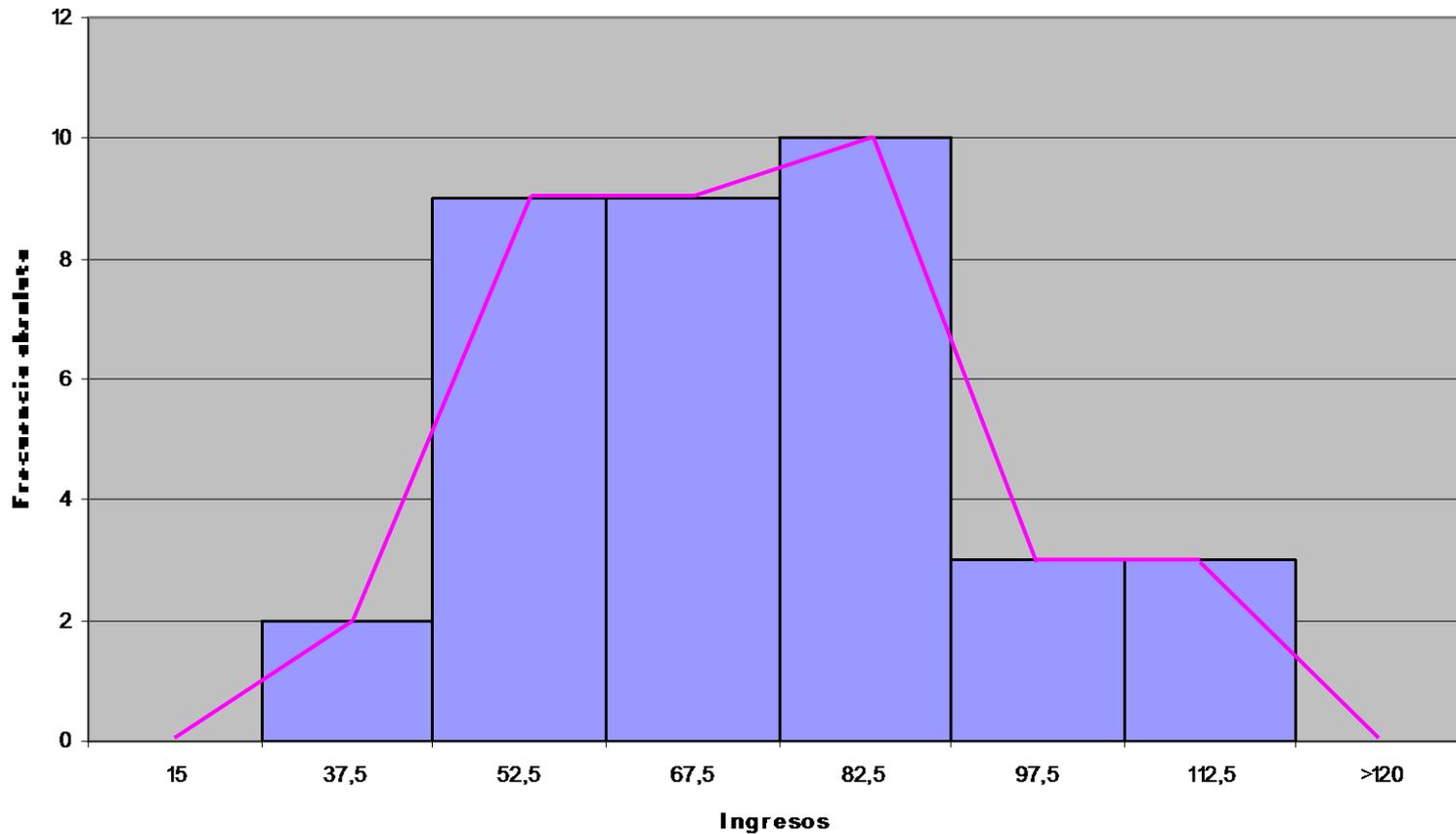
Ingresos de ayuntamientos de Madrid (millones de PTAS)





El polígono de frecuencias

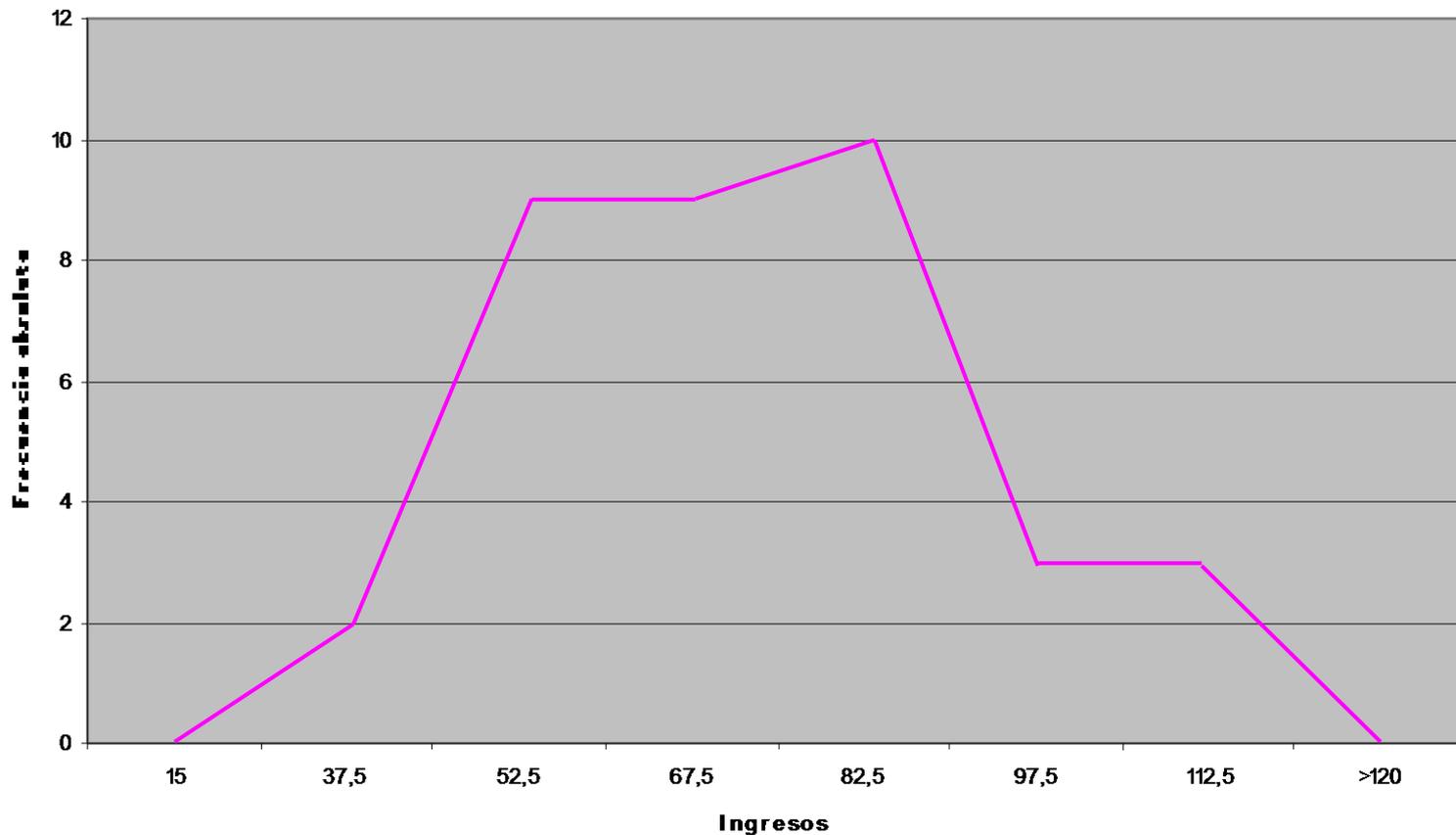
Ingresos de ayuntamientos de Madrid (millones de PTAS)





El polígono de frecuencias

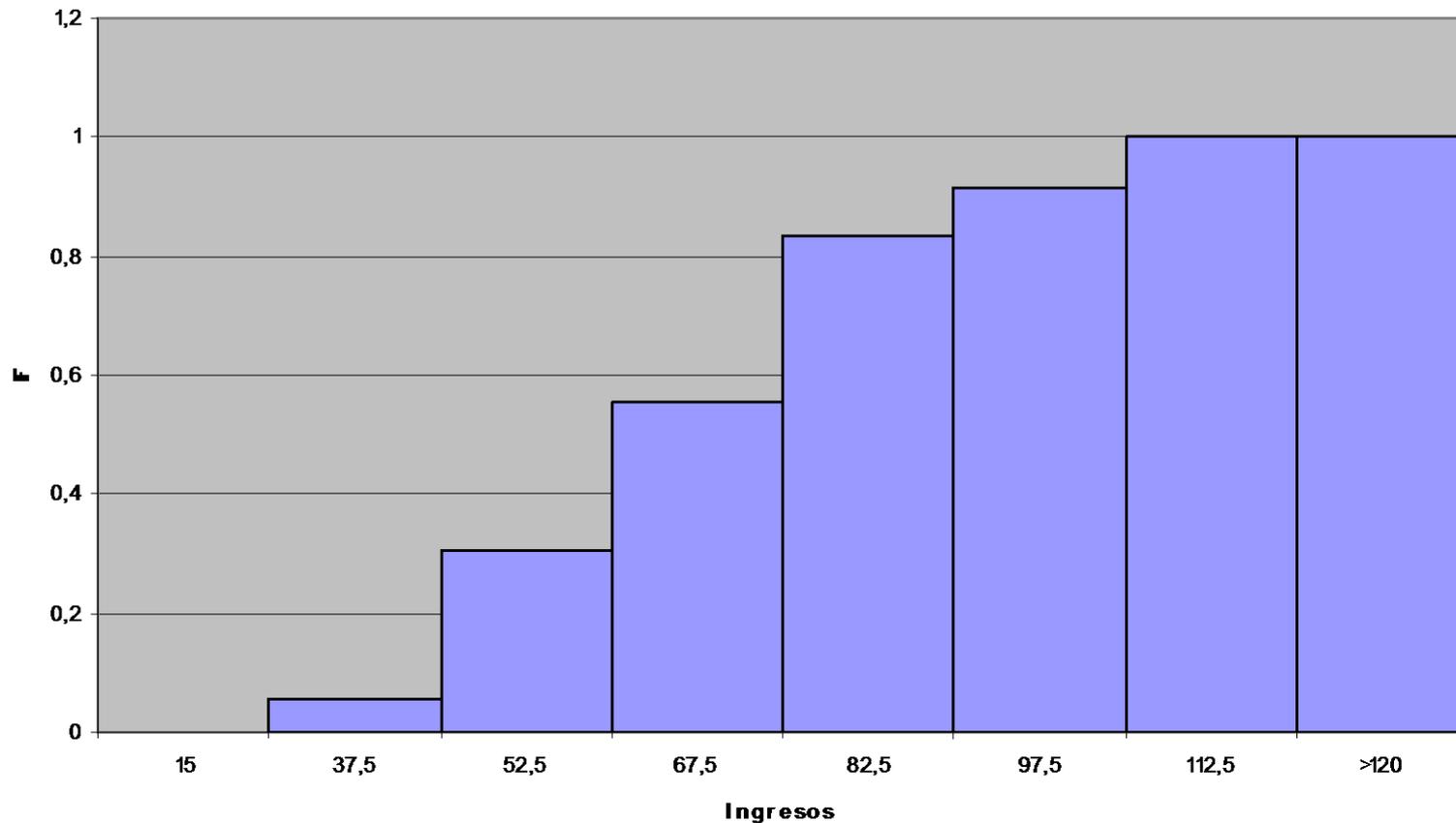
Ingresos de ayuntamientos de Madrid (millones de PTAS)





El histograma y polígono de frecuencias acumuladas

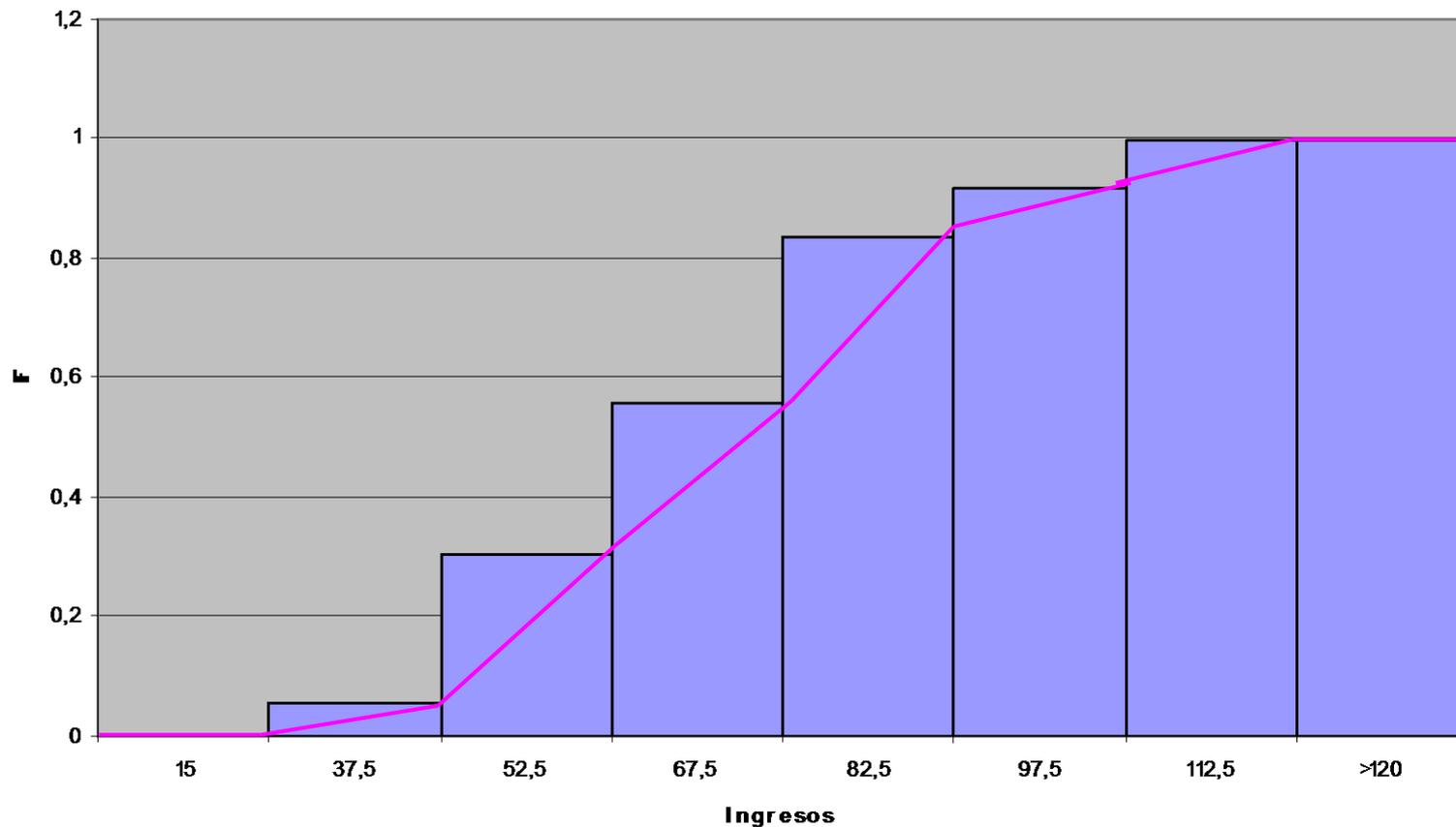
Frecuencias relativas acumuladas de los ingresos de ayuntamientos





El histograma y polígono de frecuencias acumuladas

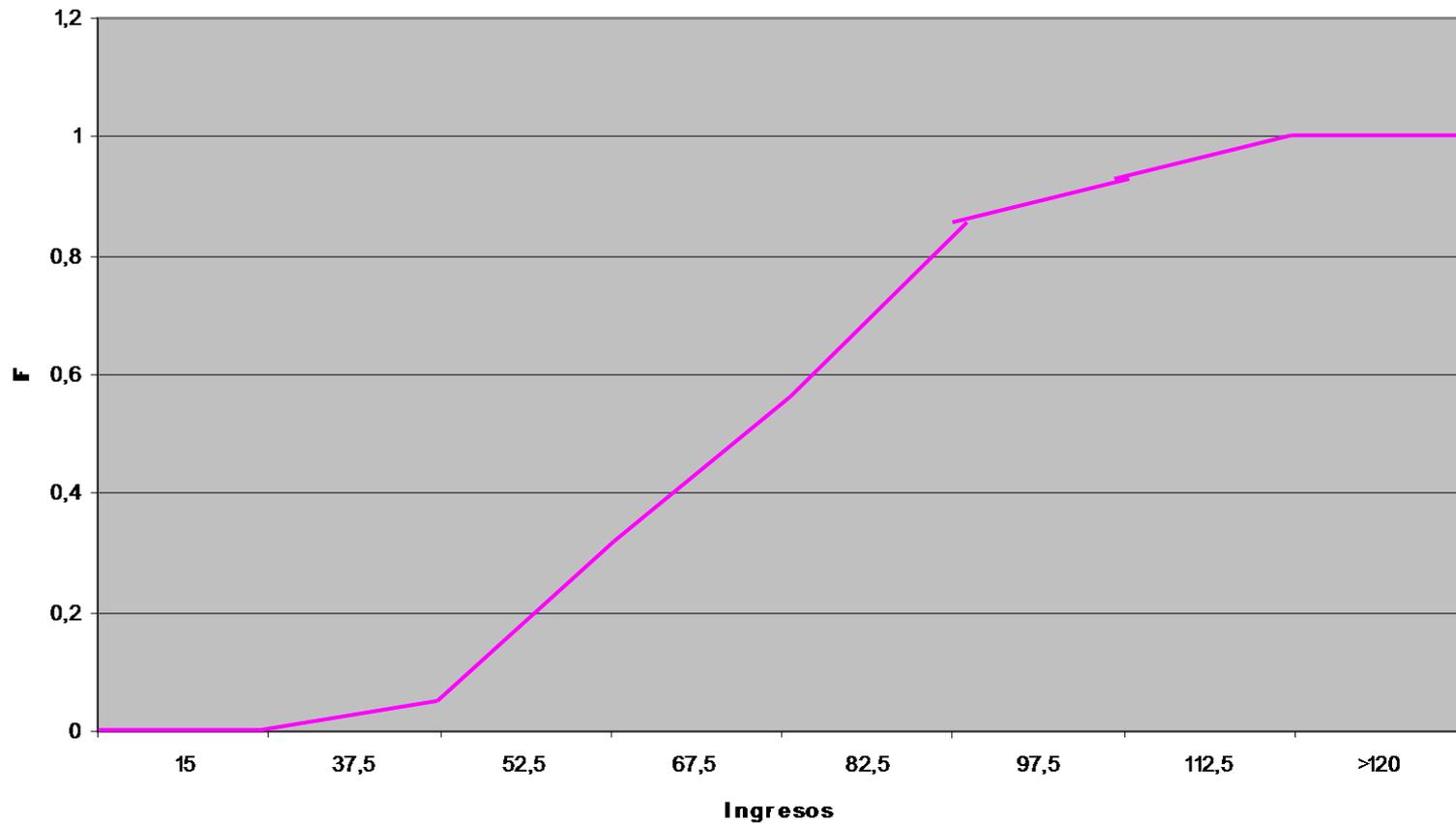
Frecuencias relativas acumuladas de los ingresos de ayuntamientos





El histograma y polígono de frecuencias acumuladas

Frecuencias relativas acumuladas de los ingresos de ayuntamientos





Mintiendo con histogramas

¿Qué pasa si cambiamos el número de barras?

¿Qué pasa si utilizamos intervalos de distintos tamaños?

Ejemplo 25 *Los siguientes datos son los resultados de una encuesta de usuarios sobre el número de gramos de marihuana que fuman cada semana.*

| <i>g / semana</i> | <i>Frecuencia</i> |
|-------------------|-------------------|
| [0, 3) | 94 |
| [3, 11) | 269 |
| [11, 18) | 70 |
| [18, 25) | 48 |
| [25, 32) | 31 |
| [32, 39) | 10 |
| [39, 46) | 5 |
| [46, 74) | 2 |
| > 74 | 0 |



Aumentamos la tabla con las frecuencias relativas y las alturas de las barras.

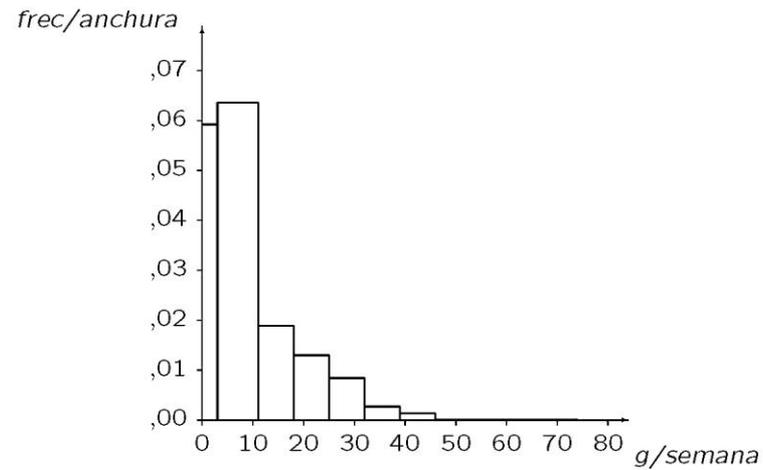
| <i>g / semana</i> | <i>anchura</i> | <i>n_i</i> | <i>f_i</i> | <i>altura</i> |
|-------------------|----------------|----------------------|----------------------|---------------|
| [0, 3) | 3 | 94 | ,178 | ,0592 |
| [3, 11) | 8 | 269 | ,509 | ,0636 |
| [11, 18) | 7 | 70 | ,132 | ,0189 |
| [18, 25) | 7 | 48 | ,091 | ,0130 |
| [25, 32) | 7 | 31 | ,059 | ,0084 |
| [32, 39) | 7 | 10 | ,019 | ,0027 |
| [39, 46) | 7 | 5 | ,009 | ,0014 |
| [46, 74) | 28 | 2 | ,004 | ,0001 |
| > 74 | 0 | 0 | 0 | 0 |
| <i>Total</i> | | 529 | 1 | |

Usamos la fórmula

$$\mathbf{altura = frecuencia / anchura \ del \ intervalo}$$



Histograma del consumo de marijuana semanal



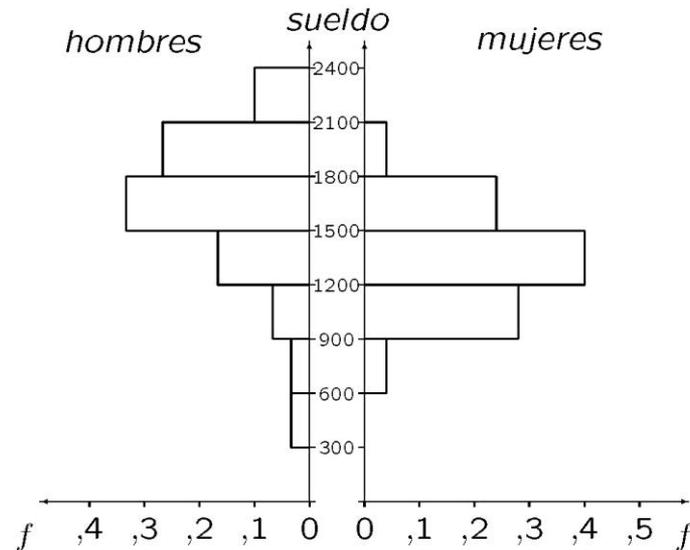
Se ve claramente que la distribución es muy asimétrica a la derecha.

Landrigan et al (1983). Paraquat and marijuana: epidemiologic risk assessment. *Amer. J. Public Health*, **73**, 784-788



Comparando dos grupos con histogramas

Histograma de los sueldos horarios de hombres y mujeres



El sueldo medio de los hombres parece un poco más alto y la distribución de sueldo de hombres es más dispersa y asimétrica.

Dolado, J. y V. LLorens (2004). Gender Wage Gaps by Education in Spain: Glass Floors vs. Glass Ceilings, *CEPR DP.*, **4203**.

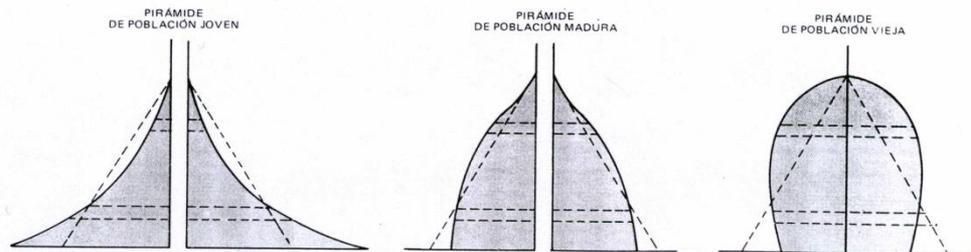
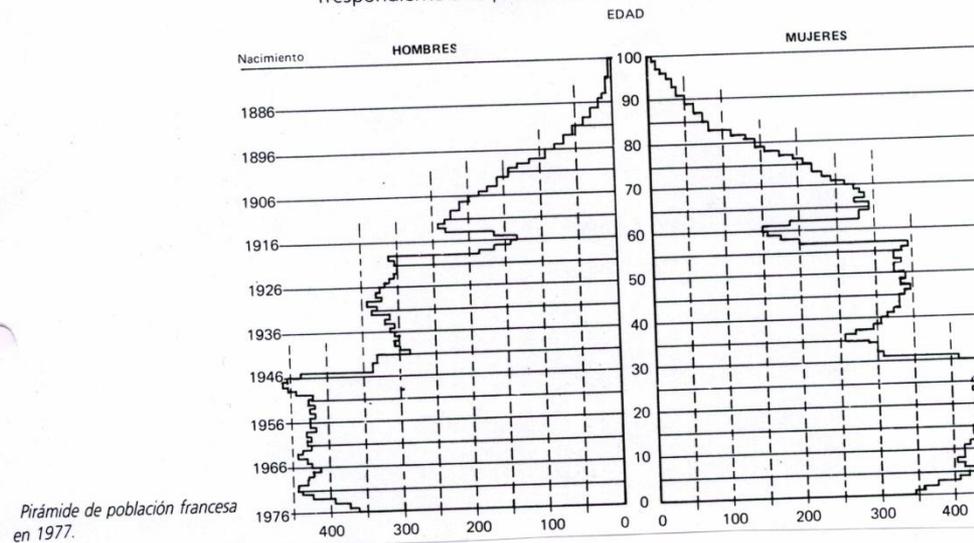
<http://www.eco.uc3m.es/temp/dolllorens2.pdf>



Otros gráficos

1) Pirámide de población

En el gráfico siguiente se representa la pirámide de población correspondiente a la población francesa en 1977:



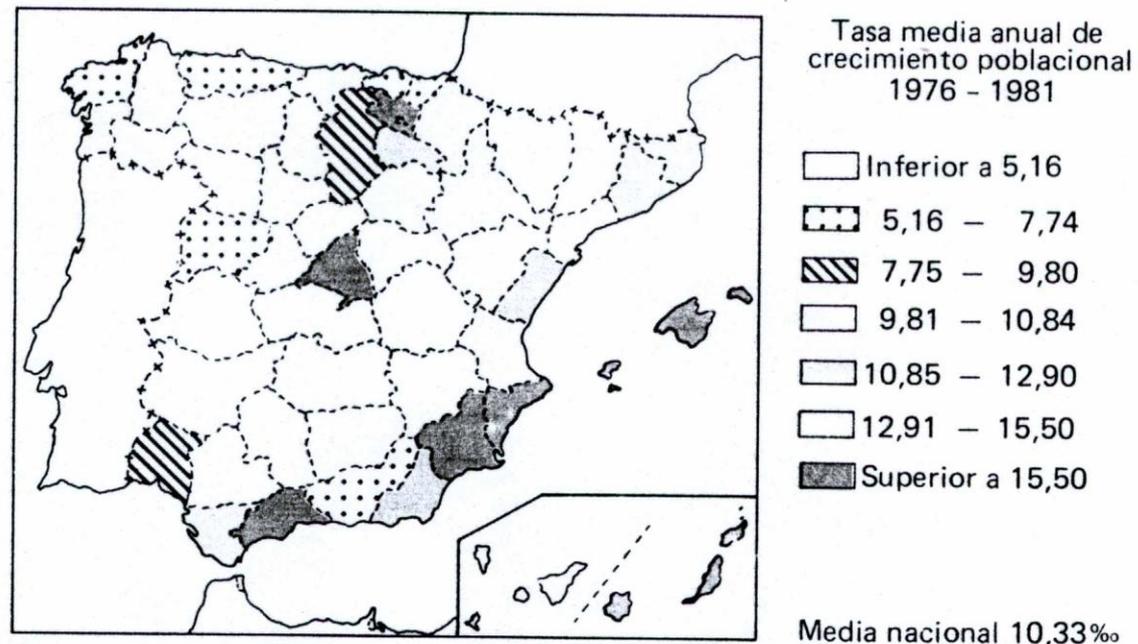
Pirámides de población joven, madura y vieja.



Otros gráficos

2) Cartograma

La distribución de la tasa media anual de crecimiento poblacional durante el período 1976-1981 fue la que se indica en el siguiente cartograma:





Ejercicio

Repasar todo lo anterior con los siguientes datos sobre las estaturas de una muestra de políticos.

| Estatura | Frecuencia |
|-----------|------------|
| Intervalo | f_i |
| 152-160 | 5 |
| 160-168 | 18 |
| 168-176 | 41 |
| 176-184 | 28 |
| 184-192 | 8 |



Ejercicio (Pregunta de examen)

El tiempo que una muestra de estudiantes de la Universidad Carlos III de Madrid tardan en llegar a clase desde sus domicilios se recoge en la siguiente tabla:

| Intervalos | Frecuencias |
|------------|-------------|
| [5-15) | 10 |
| [15-20) | 7 |
| [20-30) | 8 |
| [30-45) | 6 |
| [45-50) | 3 |

Considerando que el número de alumnos de la clase fue 34, ¿marque cuál de las siguientes afirmaciones es la correcta?:

- a) El número de estudiantes que demora menos de 20 minutos es 17.
- b) El número de estudiantes que demora exactamente 20 minutos es 7.
- c) El número de estudiantes que demora 20 ó más es 8.
- d) Ninguna de las anteriores.



Ejercicio (Pregunta de examen)

Se ha construido la siguiente tabla de frecuencias donde x_i representa el centro del intervalo, n_i la frecuencia absoluta, N_i la frecuencia absoluta acumulada, f_i la frecuencia relativa o proporción y F_i la frecuencia relativa acumulada.

Los valores representados

por **A**, **B**, **C** y **D** son:

- a) 25, 16, 16 y 0,25 respectivamente
- b) 30, 10, 12 y 0,2 respectivamente
- c) 30, 10, 16 y 0,25 respectivamente.
- d) Ninguna de las anteriores.

| Intervalos | x_i | n_i | N_i | f_i | F_i |
|------------|-------|-------|-------|-------|-------|
| [0-10) | 5 | 2 | 2 | 0,05 | 0,05 |
| [10-20) | 15 | | | | 0,15 |
| [20-40) | A | B | C | D | 0,4 |
| [40-50) | | 15 | | | 0,775 |
| [50-100) | | | 40 | | 1 |



Ejercicio (Pregunta de examen)

El diagrama de barras de frecuencias acumuladas muestra el número de programas de noticias vistas diariamente por 55 estudiantes del grado.

¿Cuál de las siguientes afirmaciones es la correcta?

- a. 15% de los individuos en la muestra ven 3 programas de noticia cada día.
- b. 48 estudiantes ven al menos 2 programas cada día.
- c. 31 estudiantes ven 1 programa de noticias cada día.
- d. 22 estudiantes ven entre 2 y 3 programas cada día.

