



Description of quantitative variables

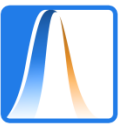
For quantitative data as with qualitative data, the first step is to construct a frequency table. If the data are discrete, this is easy.

SAMPLE: 60 adult madrileños

VARIABLE: Number of times the subject has voted in the Community of Madrid elections

OBJECTIVE: **Classification and representation of the information**

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



The table of frequencies

| Number of times voted | Absolute frequency |
|-----------------------|--------------------|
| 0 | 4 |
| 1 | 10 |
| 2 | 12 |
| 3 | 15 |
| 4 | 11 |
| 5 | 5 |
| 6 | 1 |
| 7 | 1 |
| 8 | 1 |
| >8 | 0 |
| Total | 60 |

Include an empty bar





We can also include cumulative frequencies in the table

| Times voted | Absolute frequency | Cumulative frequency | Relative frequency | Cumulative relative frequency |
|-------------|--------------------|----------------------|--------------------|-------------------------------|
| 0 | 4 | 4 | $4/60 = 0,0667$ | 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 | — |



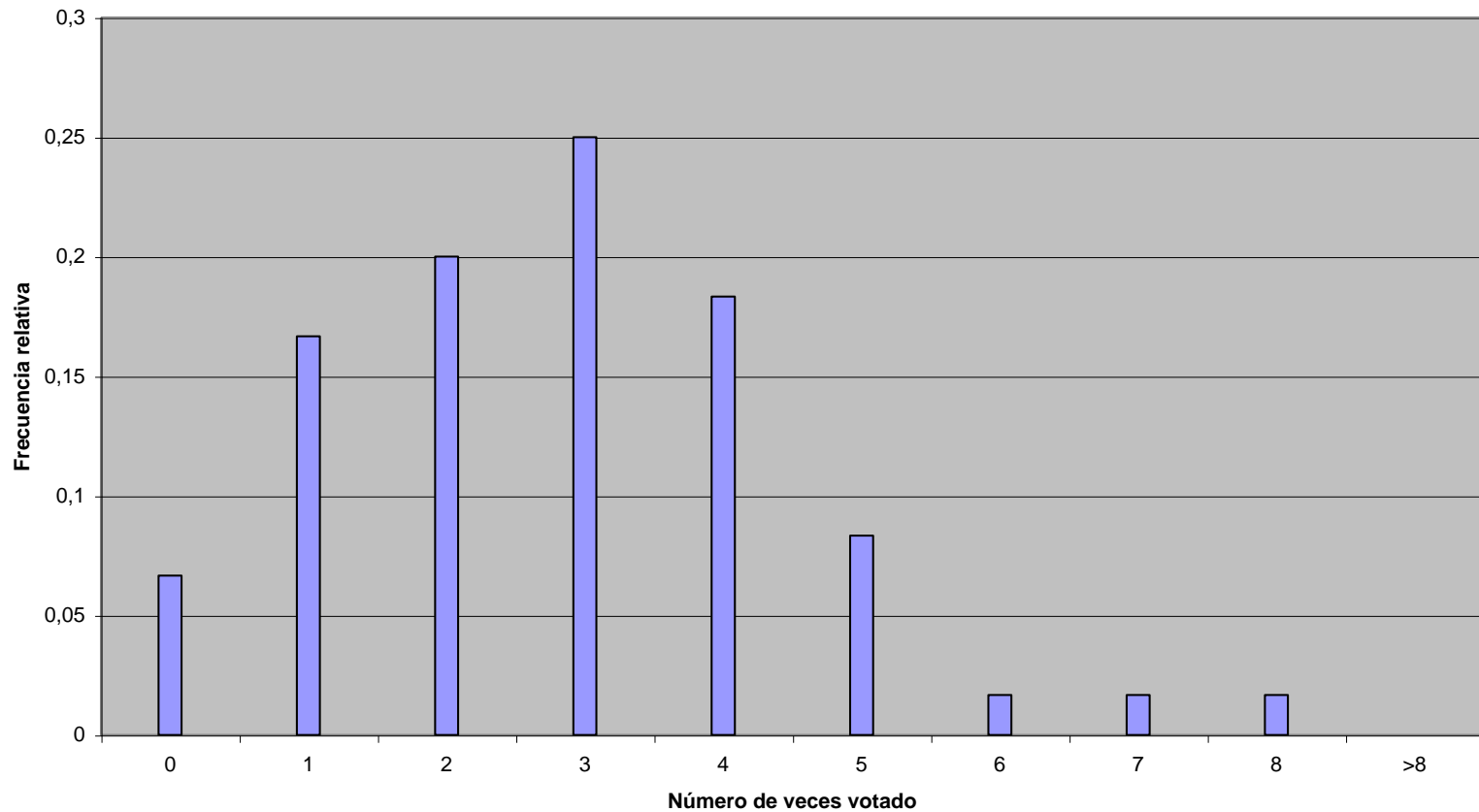
The general format of the table

| Value (x_i) | n_i | N_i | f_i | F_i |
|-----------------|----------|-------------------|---------------|-------------------|
| x_1 | n_1 | $N_1 = n_1$ | $f_1 = n_1/N$ | $F_1 = f_1$ |
| x_2 | n_2 | $N_2 = N_1 + n_2$ | $f_2 = n_2/N$ | $F_2 = F_1 + f_2$ |
| 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 | $N_k = N$ | $f_k = n_k/N$ | $F_k = 1$ |
| Total | N | — | 1 | — |



The bar chart

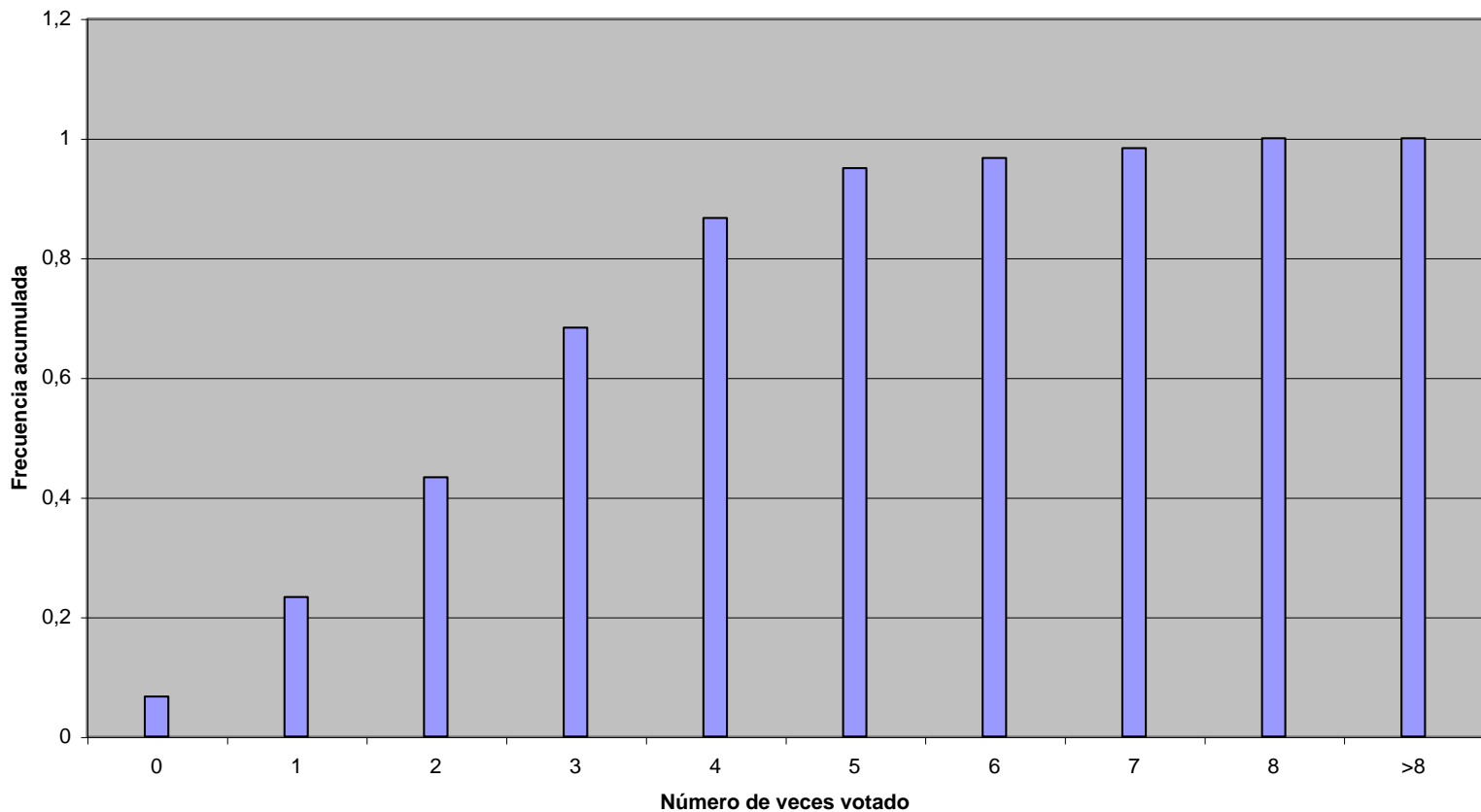
Las veces que votan los madrileños





A bar chart with cumulative frequencies

Frecuencias acumuladas del número de veces que votan los madrileños





How to construct a frequency table with continuous data

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 |

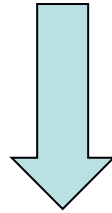
N=36

Minimum = 38360

Maximum = 114579

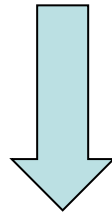


A bar chart with bars for every possible value doesn't make sense



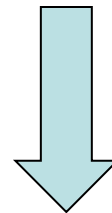
We have to group the data

How many intervals should we use?



Approximately \sqrt{N}

Where should we start and what bar widths should we use?



Use more or less 'round' numbers



The frequency table

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

Include an empty bar



Data less than or equal to 45

Data bigger than 45



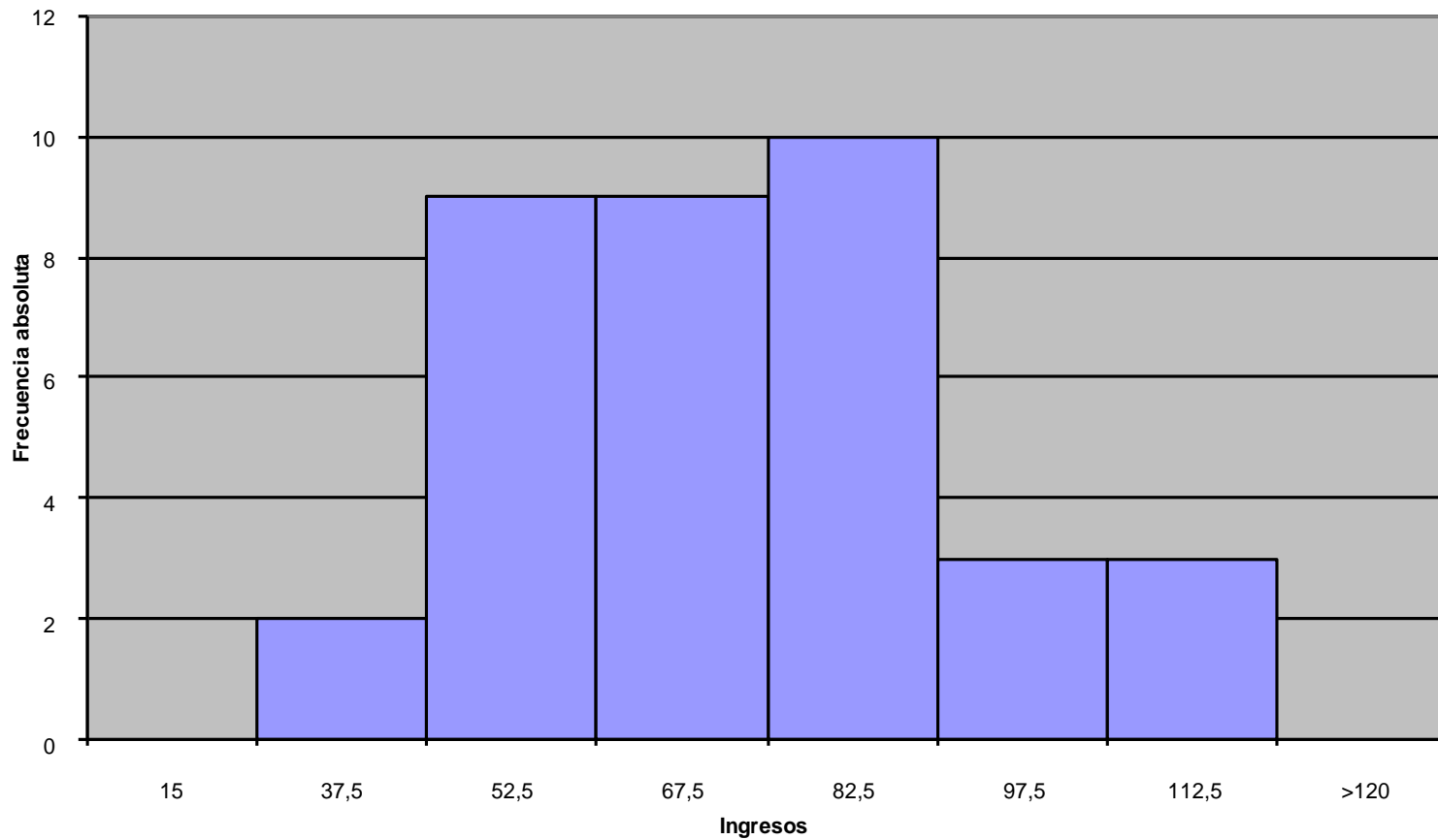
General format of the table

| Interval | Centre (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 | — |



The histogram

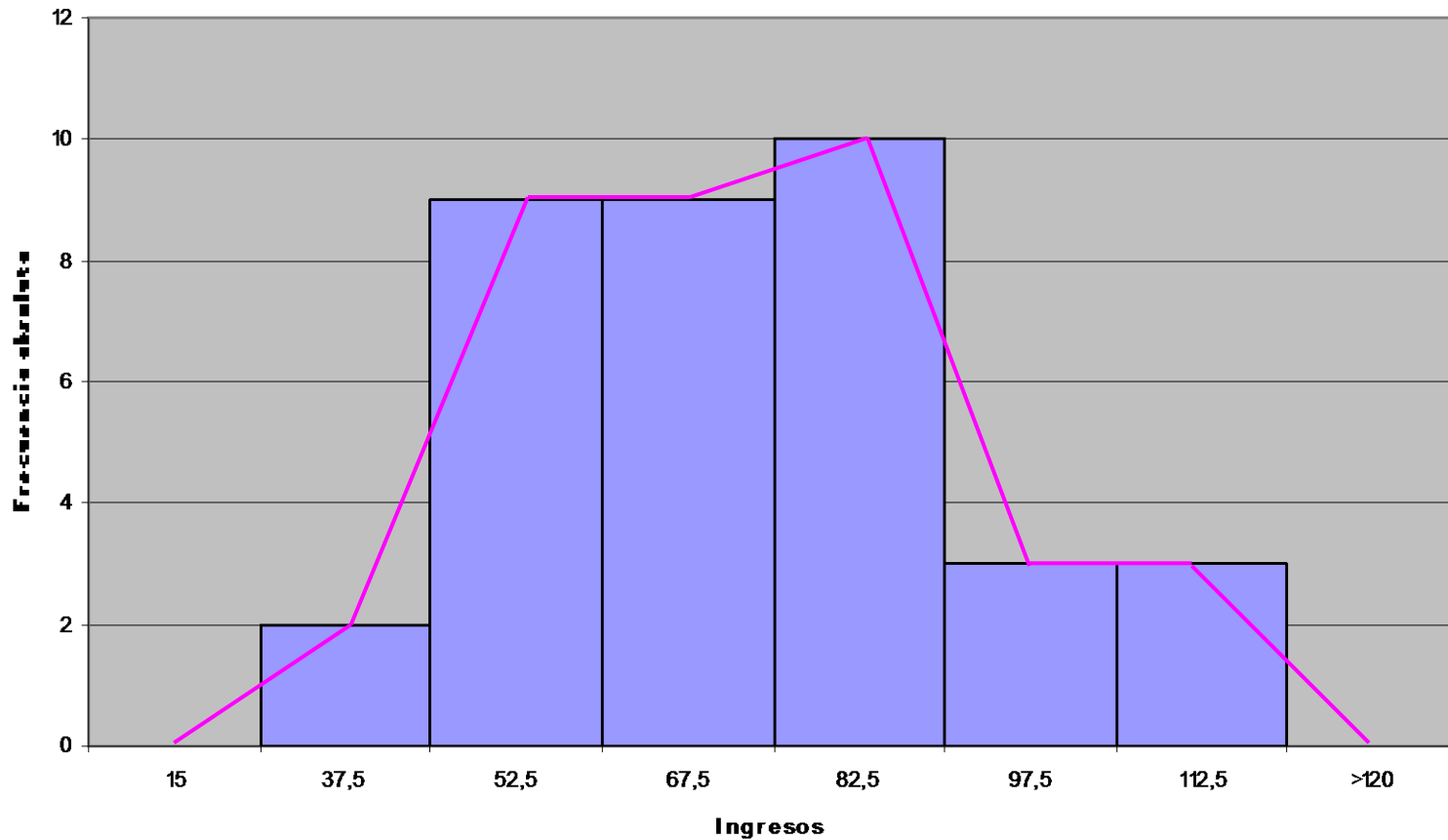
Ingresos de ayuntamientos de Madrid (millones de PTAS)





The frequency polygon I

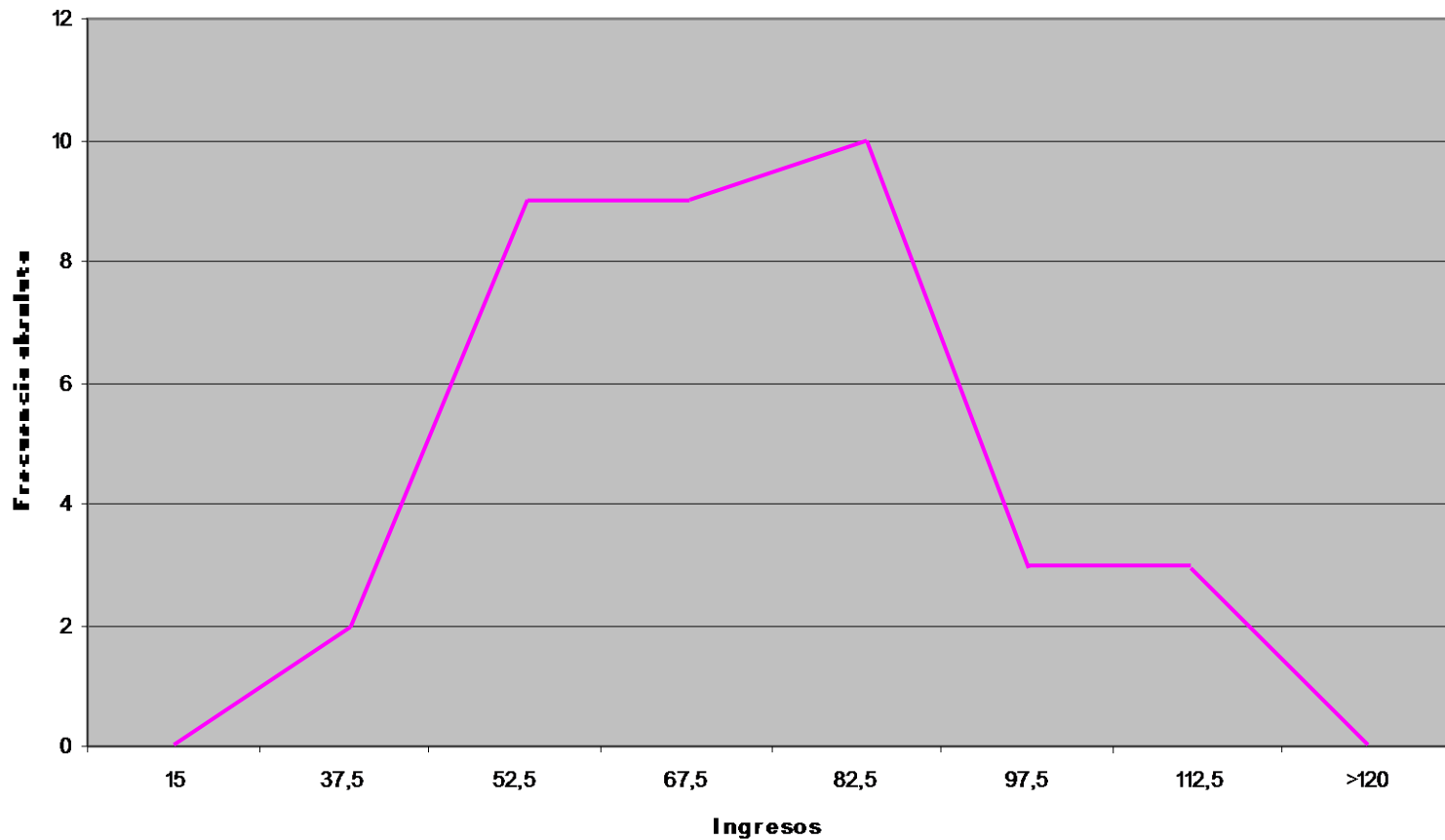
Ingresos de ayuntamientos de Madrid (millones de PTAS)

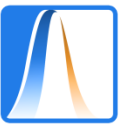




The frequency polygon II

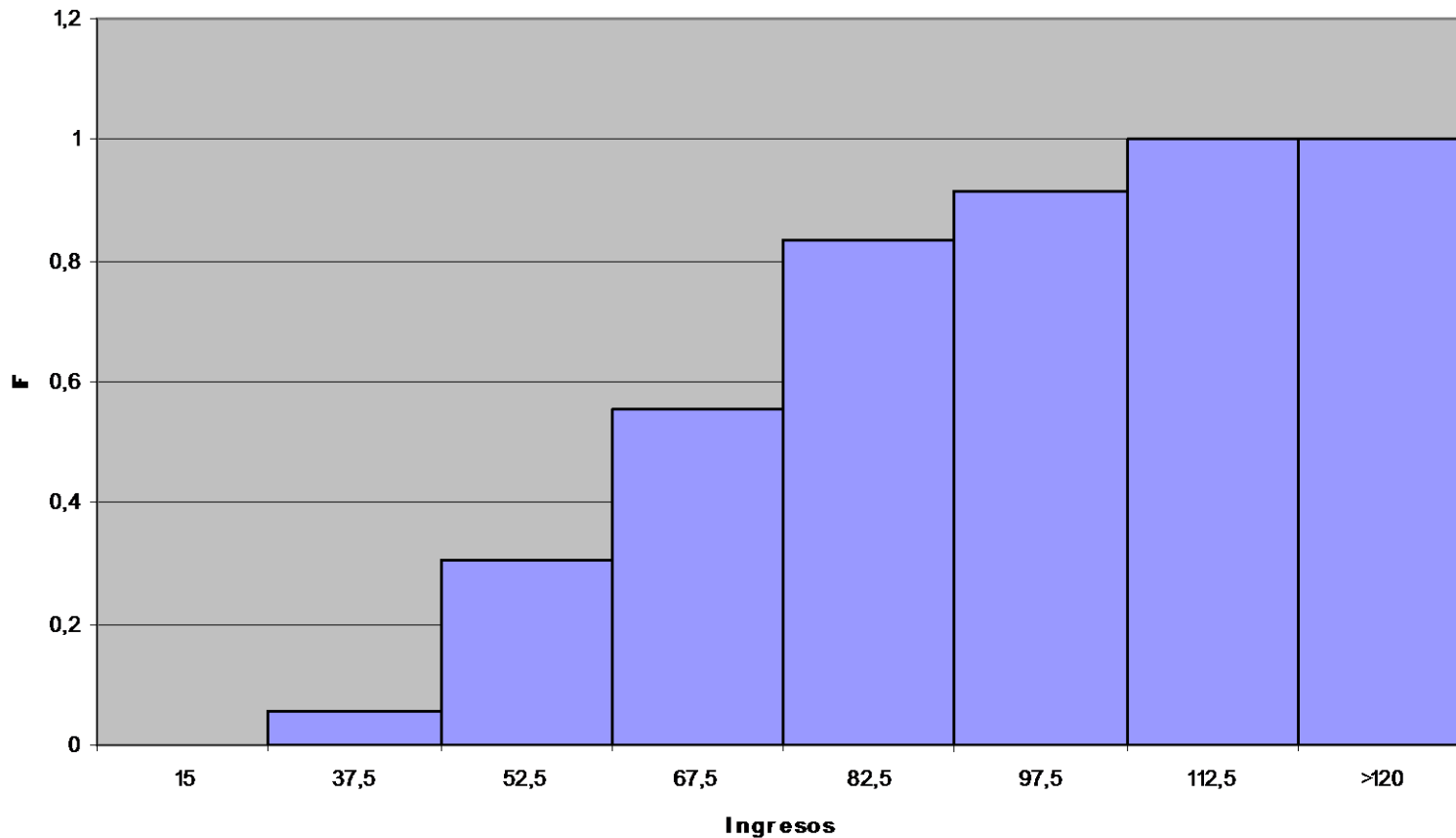
Ingresos de ayuntamientos de Madrid (millones de PTAS)





The histogram and frequency polygon for cumulative data I

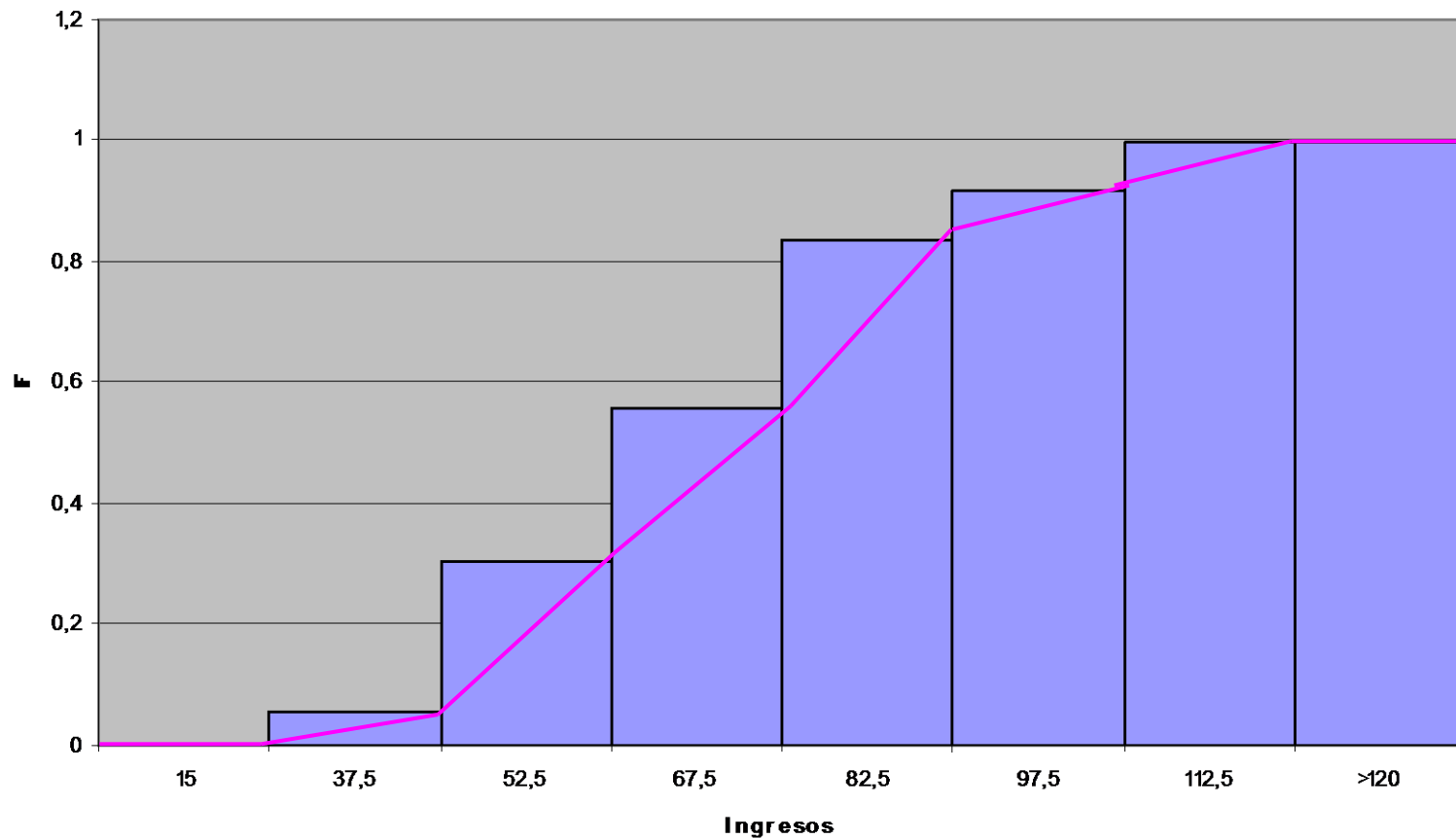
Frecuencias relativas acumuladas de los ingresos de ayuntamientos





The histogram and frequency polygon for cumulative data II

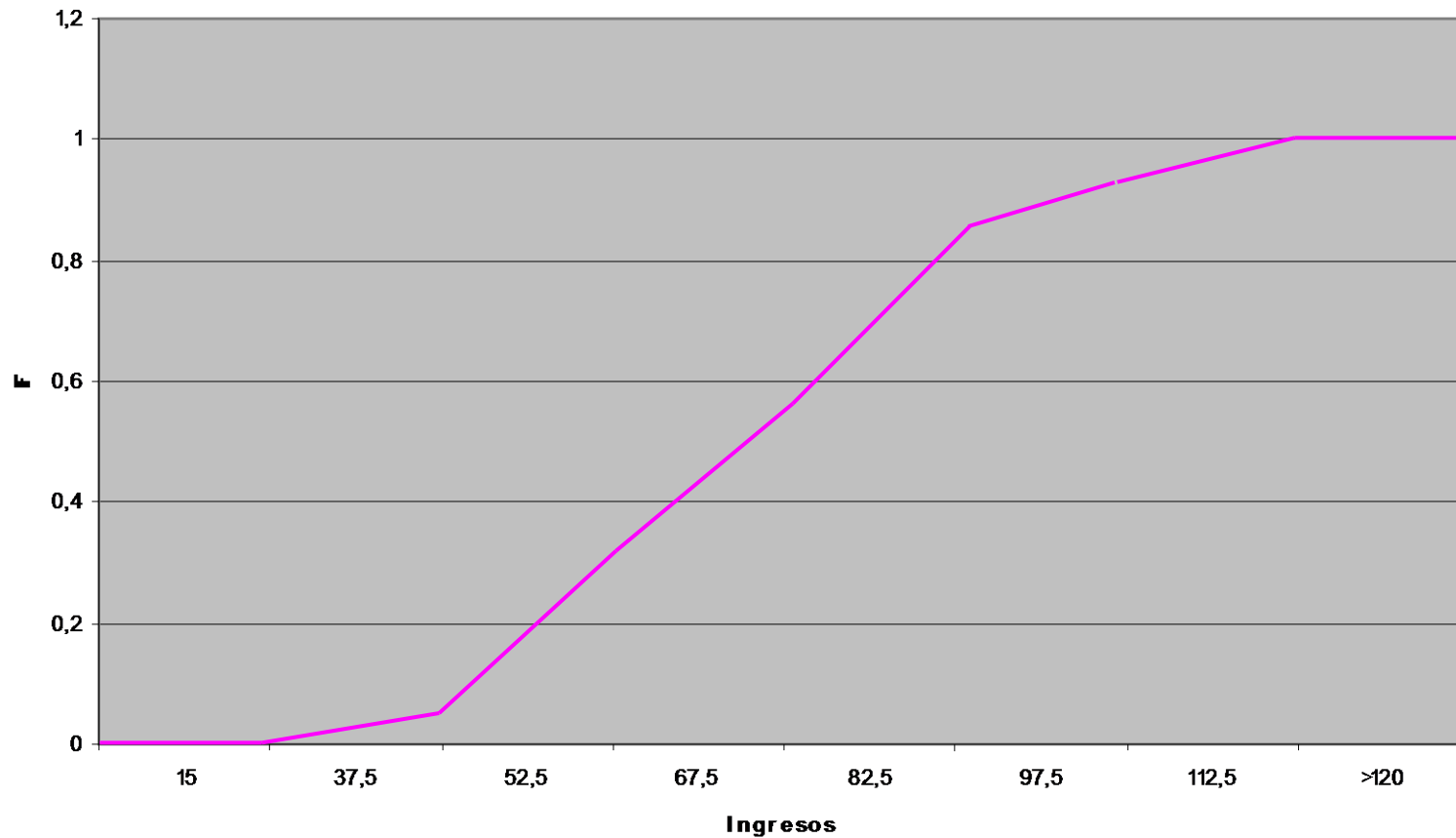
Frecuencias relativas acumuladas de los ingresos de ayuntamientos





The histogram and frequency polygon for cumulative data III

Frecuencias relativas acumuladas de los ingresos de ayuntamientos





Lying with histograms

What happens when we change the number of bars?

What happens if we use intervals of different widths?

Ejemplo 25 *Los siguientes datos son los resultados de una encuesta de usuarios sobre el número de gramas de marijuana 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.

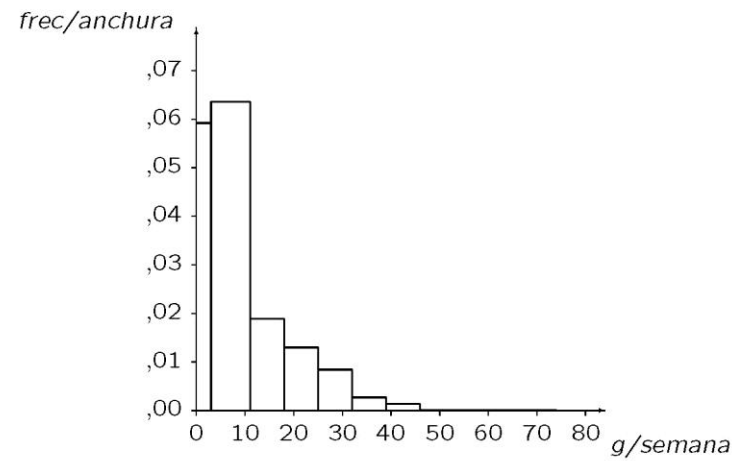
| <i>g / semana</i> | <i>anchura</i> | <i>n_i</i> | <i>f_i</i> | <i>altura</i> |
|-------------------|----------------|----------------------|----------------------|---------------|
| [0, 3) | 3 | 94 | ,178 | ,0592 |
| [3, 11) | 8 | 269 | ,509 | ,0636 |
| [11, 18) | 7 | 70 | ,132 | ,0189 |
| [18, 25) | 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

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



Histograma del consumo de marijuana semanal



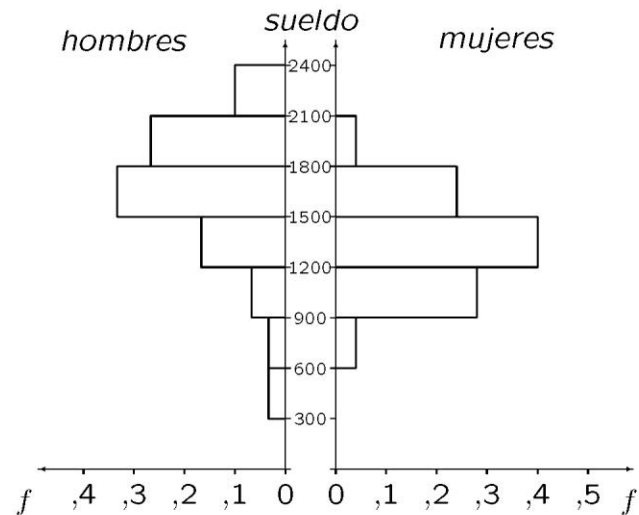
Se ve claramente que la distribución es muy asimétrica a la derecha.

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



Comparing two groups with histograms

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**.

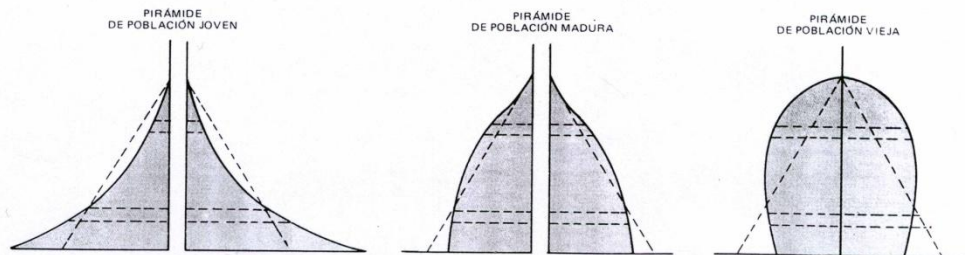
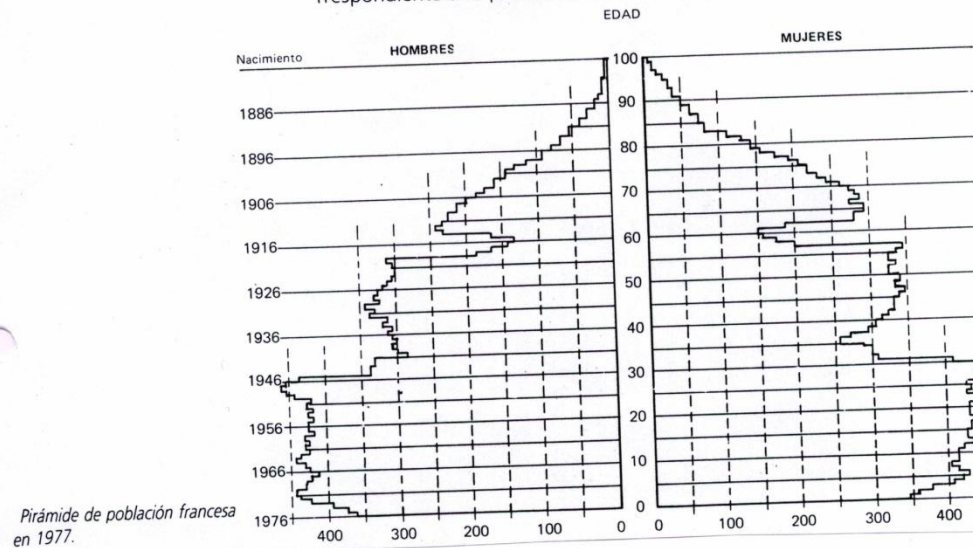
<http://www.eco.uc3m.es/temp/dollorems2.pdf>



Other graphics I

1) Population pyramid

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



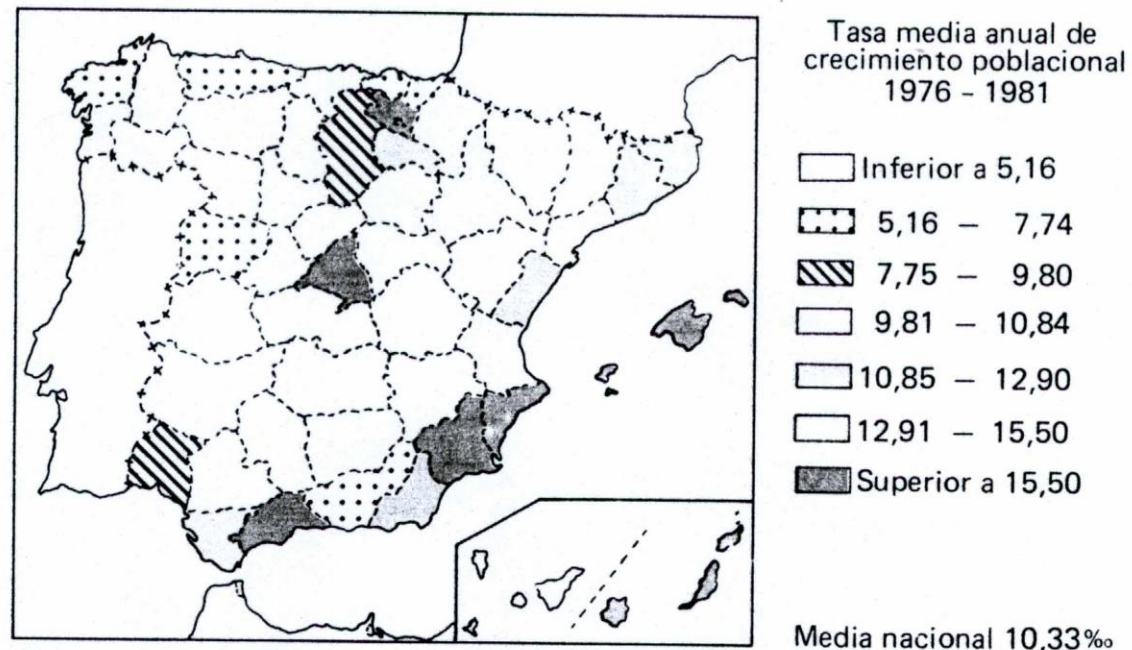
Pirámides de población joven, madura y vieja.



Other graphics II

2) Cartogram

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





Exercise

La estatura media de los integrantes de un club juvenil se ha tomado una muestra y se han obtenido los siguientes resultados:

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



Exercise

| Impact factor | Frequency |
|---------------|-----------|
| Interval | f_i |
| [0,0.3) | 22 |
| [0.3,0.6) | 25 |
| [0.6,0.8) | 17 |
| [0.8,1.0) | 10 |
| [1.0,1.5) | 11 |
| [1.5,2.0) | 9 |
| [2.0,3.0) | 4 |
| [3.0,5.0) | 1 |
| >5 | 0 |

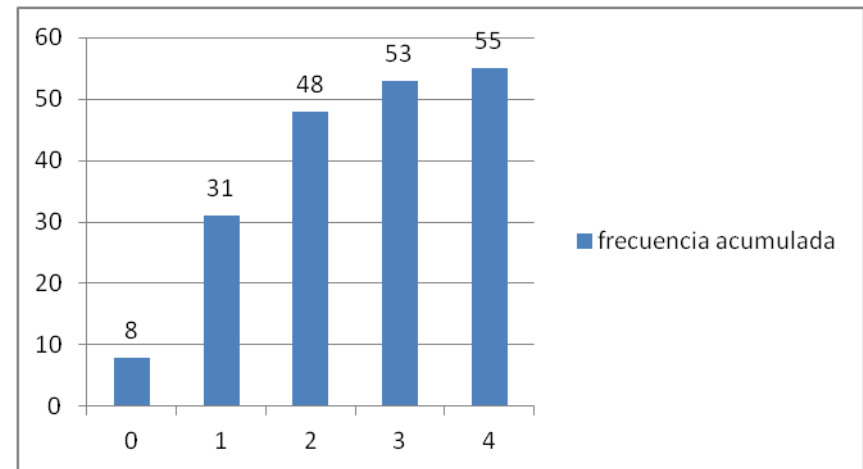
The table shows the impact factors of POLITICAL SCIENCE journals in 2008.

Construct a histogram of these data and comment.



Exercise (Test question)

The following bar chart of cumulative frequencies records the number of different news programs seen each day by a total of 55 journalism students



Which of the following affirmations is correct?

- 15% of the individuals sampled watch 3 different news programs every day.
- 31 students watch exactly 1 news program every day.
- 22 students watch between 2 and 3 news programs every day.
- 48 students watch at least 2 news programs every day.



Exercise (Test question)

In a survey on Reading habits, 60 randomly chosen university students were asked about the average amount of money that they spent on books over a 6 month period. The following results were obtained:.

| Spending on books (€) | n_i |
|-----------------------|-------|
| [0-20) | 20 |
| [20-40) | 20 |
| [40-60) | 14 |
| [60-80) | |
| [80-100] | 2 |
| Total | 60 |

What proportion of the sample spent 60 or more euros on books?

- a) 10%
- b) 6%
- c) 16%
- d) None of the previous answers.



Exercise (Test question)

The following table records the number of newspapers read each day by a sample of 50 Madrileños.

| Number of papers | Absolute Frequency |
|------------------|--------------------|
| 0 | 15 |
| 1 | 20 |
| 2 | 8 |
| 3 | 5 |
| 4 | 2 |
| Total | 50 |

Signal which of the following bar charts of cumulative, relative frequencies is correct for the above data.

