1. ESTATISTIKA DESKRIBATZAILEA. ALDAGAI BAKUNA

1.1. Kontzeptu orokorrak
1.2. Maiztasun-taulak
1.3. Adierazpen grafikoak
1.4. Estatistiko deskribatzaileak

1.1. KONTZEPTU OROKORRAK

- ***Estatistika,*** multzo bati dagozkion zenbakizko datuak biltzen, sailkatzen eta aztertzen dituen zientzia da. Bi alor ditu:

- ***Estatistika Deskribatzailea:*** aztergai den multzoari dagozkion datuak bildu, antolatu eta egoera deskribatzeko ezaugarriak lortu.

- ***Estatistika induktiboa edo inferentzia:*** emaitzak orokortu, ondorioak atera edo aurresanak egin daitezke.

- ***Populazioa:*** azterketa estatistikoa egiten deneko multzoa.

- ***Lagina***: populazioaren edozein azpimultzo.

- ***Unitate estatistikoa:*** populazioaren elementu bakoitza.

- Populazioaren (edo laginaren) ***tamaina:*** populazioaren (edo laginaren) elementu-kopurua.

- ***Aldagai estatistikoa:*** aztergaia.

- ***Aldagai kuantitatiboa***: aldagaiak har ditzakeen balioak zenbakiak direnean.

- ***Aldagai kualitatiboa:*** aldagaiak hartzen dituen balioak zenbakizkoak ez badira.

- Aldagai aleatorio kuantitatiboak diskretuak ala jarraituak izan daitezke.

  - ***Aldagai kuantitatibo diskretuak*** balio isolatuak hartzen ditu.
  - ***Aldagai kuantitatibo jarraituak*** tarte bateko edozein balio har dezake.

1.2. MAIZTASUN-TAULAK

Bira $x_1, x_2, \ldots, x_k$ balioak, $f_1, f_2, \ldots, f_k$ *maiztasun absolutuak eta n* laginaren tamaina.

$[l_i, l_{i+1})$	x_i	f_i	F_i	h_i	H_i
$[l_1, l_2)$	x_1	f_1	$F_1 = f_1$	$h_1 = f_1/n$	$H_1 = F_1/n$
$[l_2, l_3)$	x_2	f_2	$F_2 = f_1 + f_2$	$h_2 = f_2/n$	$H_2 = F_2/n$
$[l_k, l_{k+1}]$	x_k	f_k	$F_k = f_1 + f_2 + \ldots + f_k = n$	$h_k = f_k/n$	$H_k = F_k/n = 1$

- ***Modalitateak***: x_i gaiak, aldagaiak hartzen dituen balioak dira.

- ***Maiztasun absolutuak***: f_i gaiak, balio bakoitza zenbat aldiz jaso den adierazten du.

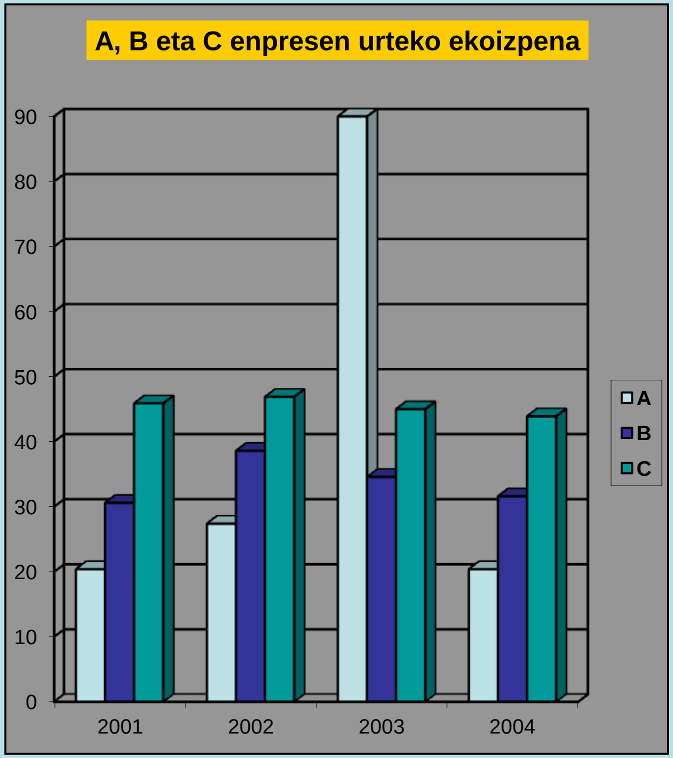
- ***Maiztasun metatuak***: $F_i = f_1 + f_2 + ... + f_i$, maiztasun absolutuen batura da.

- ***Maiztasun erlatiboa:*** $h_i = f_i/n$.

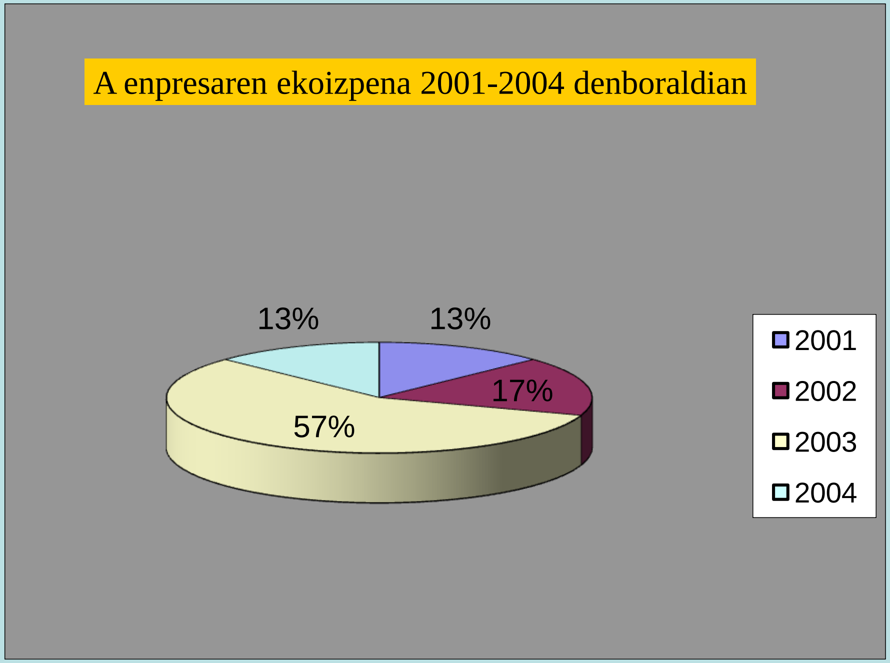
- ***Maiztasun erlatibo metatua:*** $H_i = F_i/n$.

1.3. ADIERAZPEN GRAFIKOAK

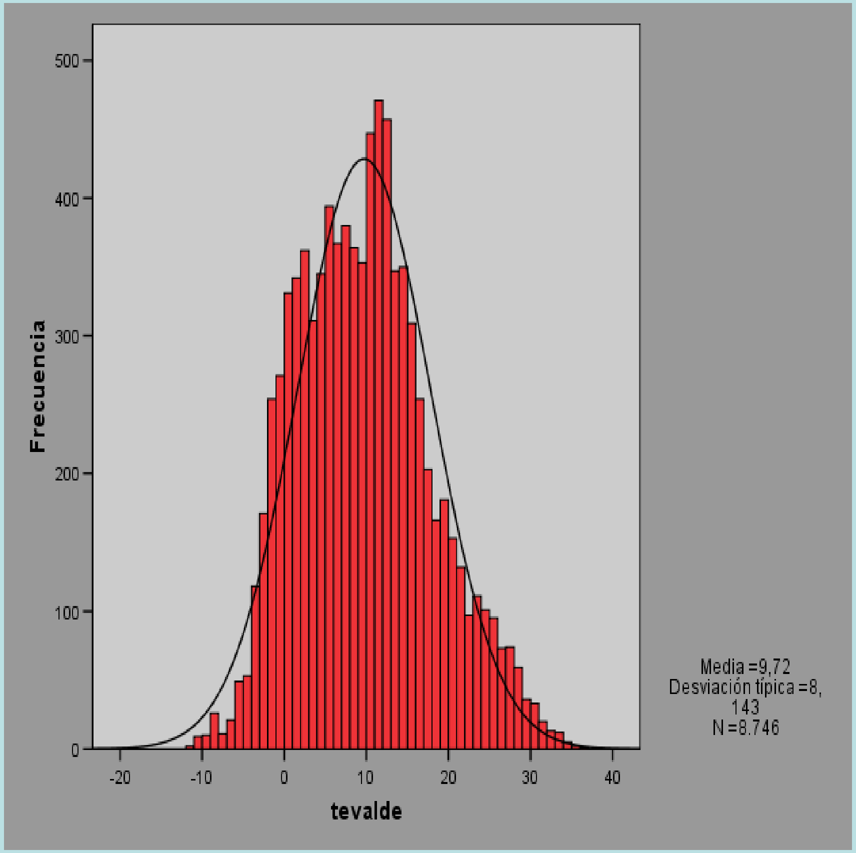
Barra-diagrama



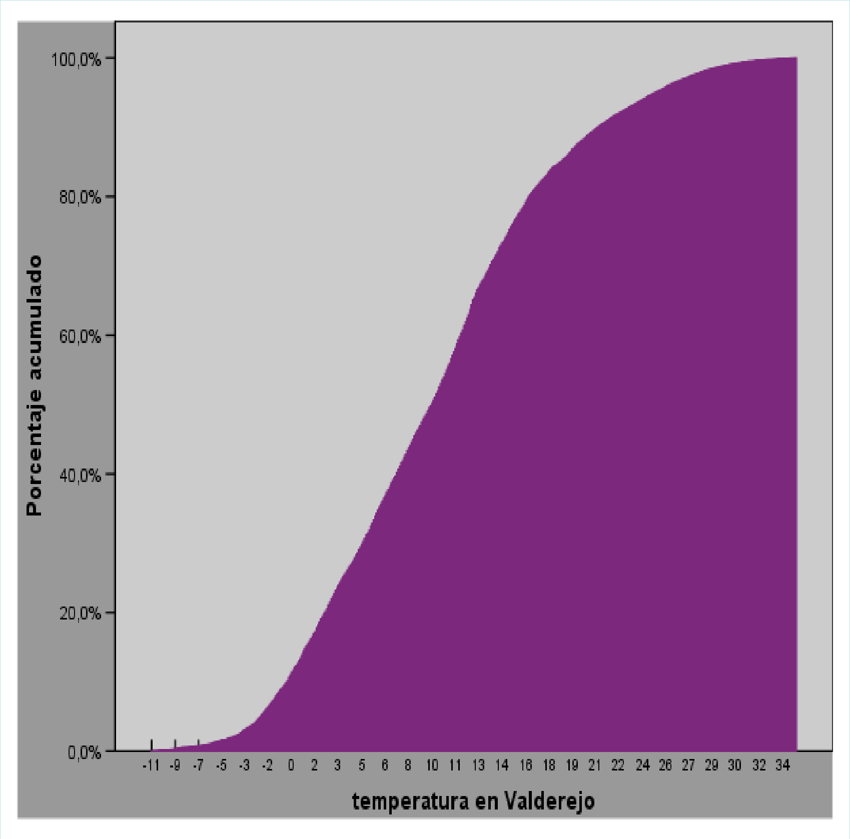
Sektore-diagrama



Histograma



Maiztasun metatuen poligonoa

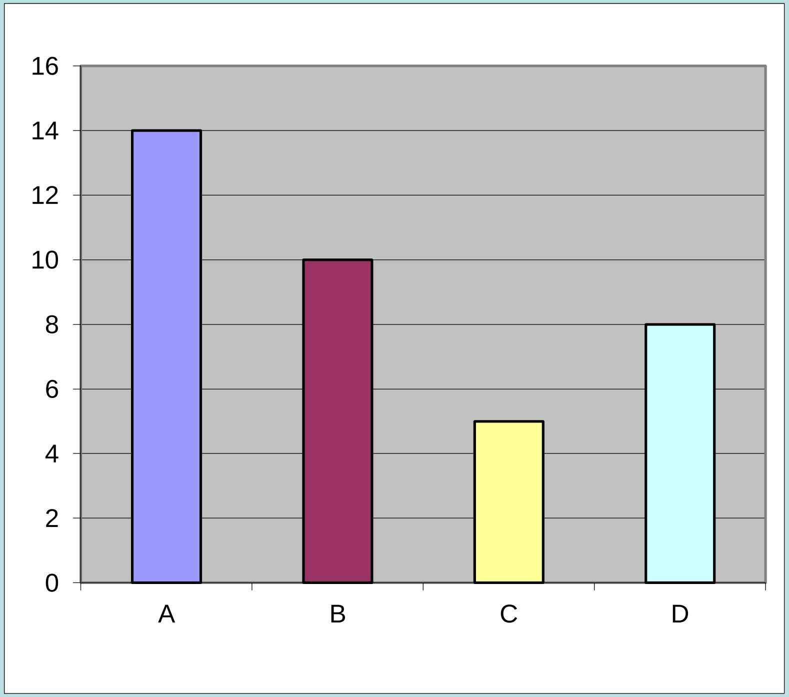


Zenbait adierazpen grafiko

- Kualitatiboa
  - Barra-diagrama
  - Sektore-diagrama
- Kuantitatiboa
  - Diskretua
    - Barra-grafikoa
    - Maiztasun metatuen grafikoa
  - Jarraitua
    - Histograma
    - Maiztasun absolutuen poligonoa
    - Maiztasun metatuen histograma
    - Maiztasun metatuen poligonoa

Barra-diagrama

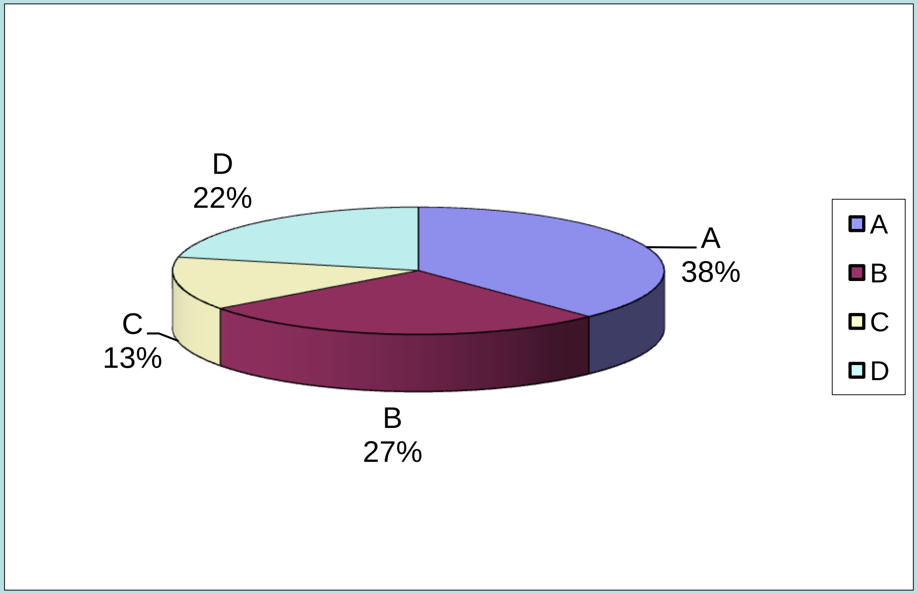
Barrez osaturik,
abzisa-ardatzean aldagaiaren balioak,
ordenatu-ardatzean maiztasun absolutuak.



Sektore-diagrama

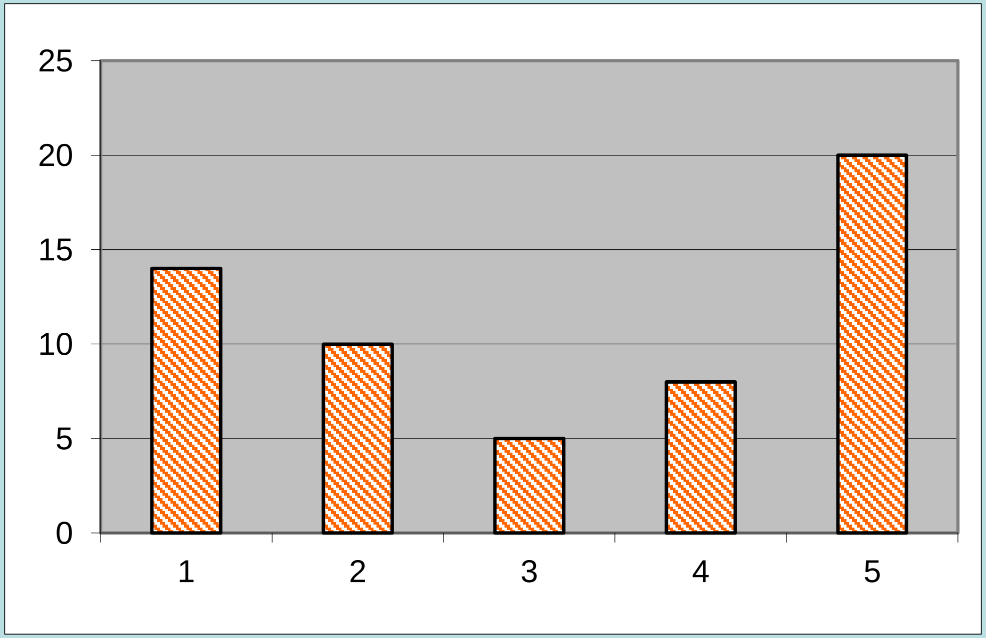
Zirkulu baten barruan, sektoreak

Sektore bakoitza, $\alpha_i = 360^{o} \cdot \frac{f_i}{n}$



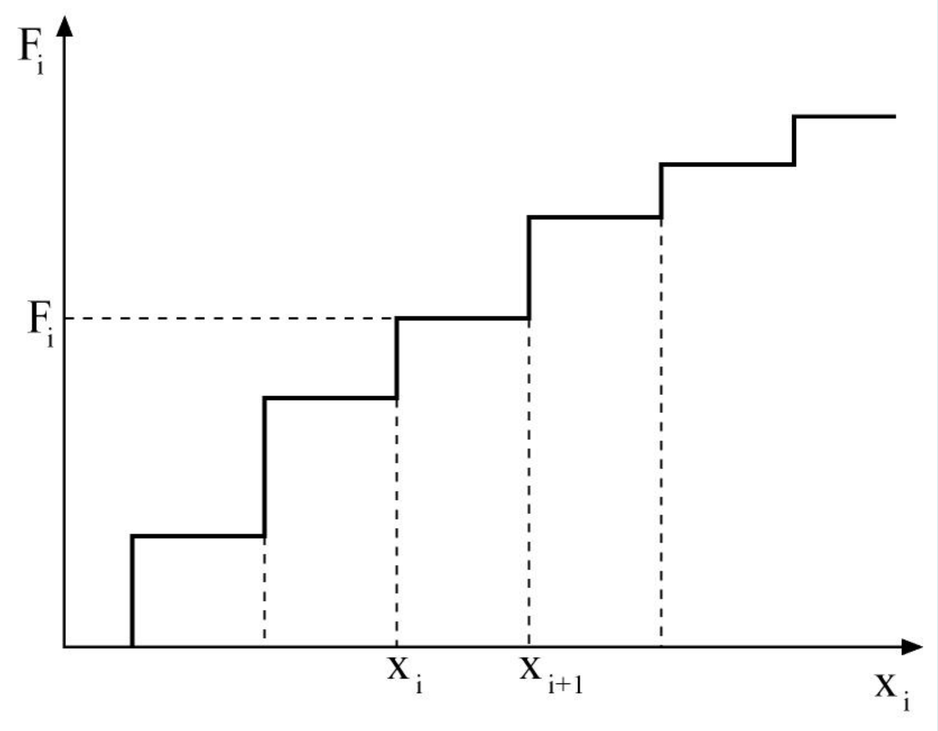
Barra-grafikoa:

Barrez osaturik,
abzisa-ardatzean aldagaiaren balioak (ordenatuta)
ordenatu-ardatzean maiztasun absolutuak.



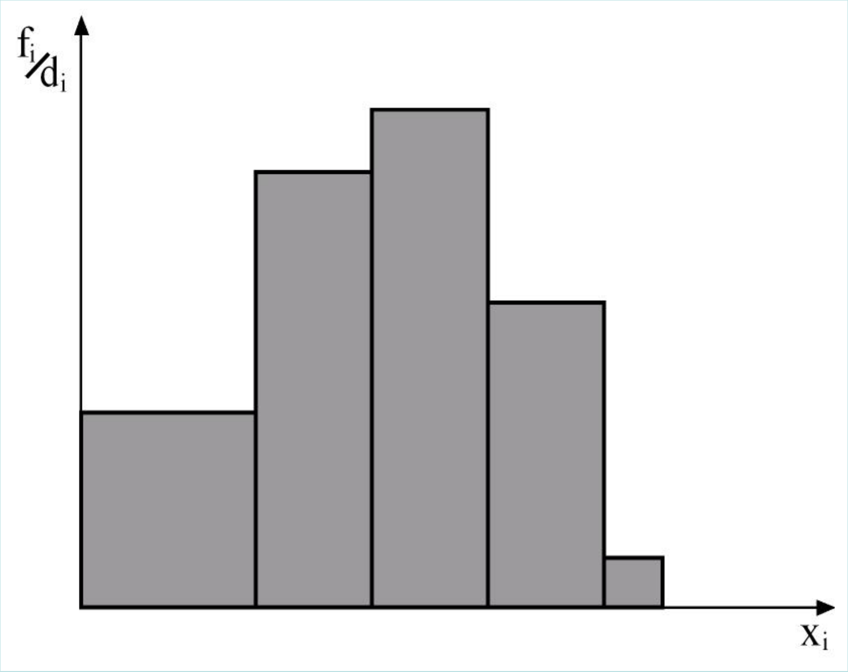
Maiztasun metatuen grafikoa:

•Abzisa-ardatzean aldagaiaren balioak eta ordenatu-ardatzean maiztasun metatuak.
•Eskailera itxura du.



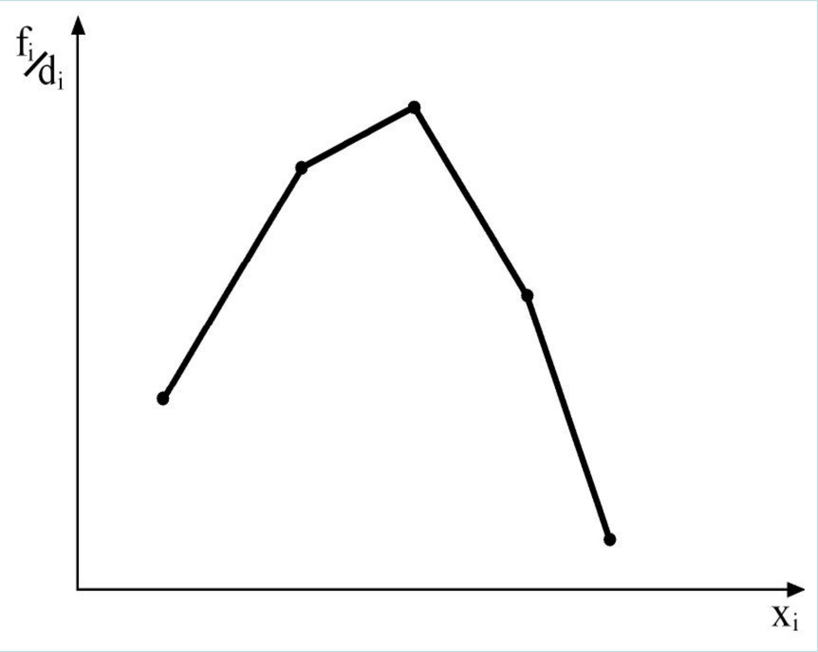
Histograma

Elkarturiko barrez osaturik, abzisa-ardatzean aldagaiaren balioak
ordenatu-ardatzean f edo f/d, d klasearen luzera izanik.



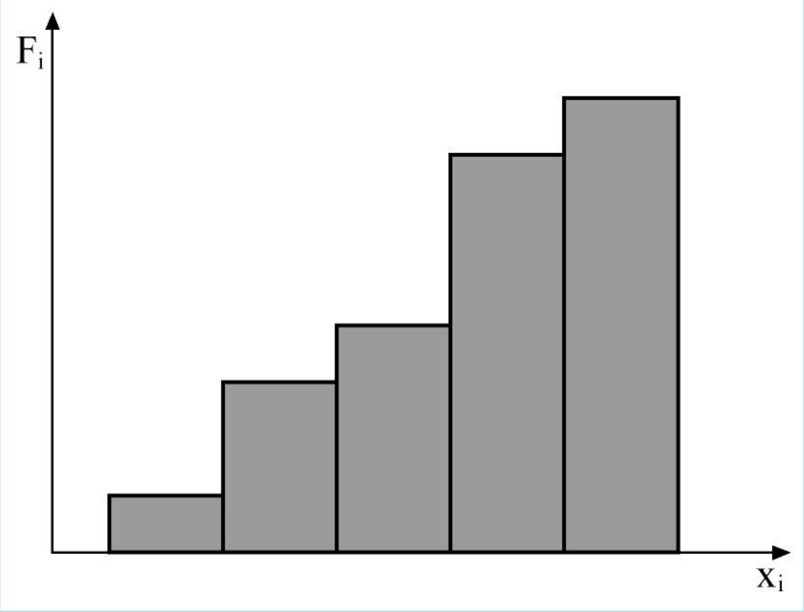
Maiztasun absolutuen poligonoa

Aurreko histograman errektangeluetako goiko aldeetako erdiko puntuak lotuz.



Histograma metatua

Elkarturiko barrez osaturik, abzisa-ardatzean aldagaiaren balioak
ordenatu-ardatzean F.

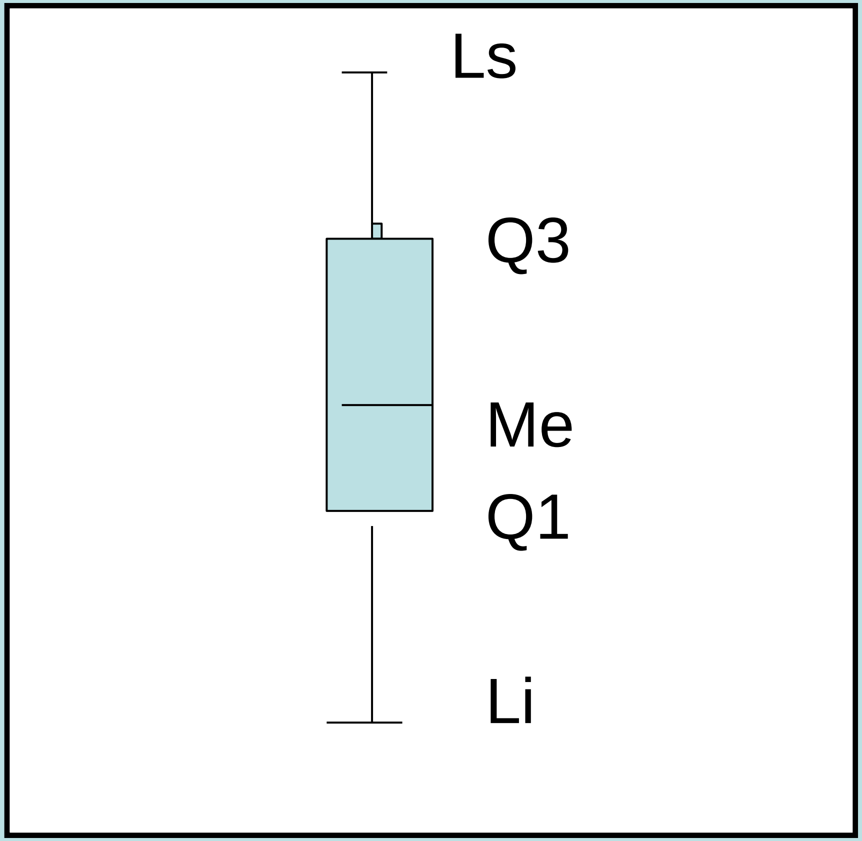


Maiztasun metatuen poligonoa

Aurreko histograman errektangeluetako goiko aldeetako eskubiko erpinak lotuz.

kutxa diagrama

Koartilen bidez egindako grafikoa da.



Zurtoin eta hosto diagrama

Datuak bi zutabeetan ezkerrekoan zurtoina eta eskuinekoan azken zifra hostoa

- **11 | 7**

  12 | 2 4 8 9

  13 | 0 2 2 4 6 7 8

  14 | 0 1 2 7 8

  15 | 1 8

  16 | 2

1.4. ESTATISTIKO DESKRIBATZAILEAK

- Bira X aldagai estatistiko kuantitatiboa, x_1, x_2, …, x_k balioak eta f_1, f_2, …, f_k maiztasunak, hurrenez hurren.

- ***Estatistiko deskribatzaileak*** laginaren menpeko funtzioak dira. Hurrengo sailkapena onartzen dute:

1.4.1. **Joera zentraleko estatistikoak**

- ***Batez besteko aritmetikoa***: aldagaiak hartzen dituen balio guztien batura eta laginaren arteko proportzioa da.

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_k f_k}{n} = \frac{\sum_{i=1}^{k} x_i f_i}{n}$$

- **Mediana, M_e** : Bere ezkerrean eta bere eskuinean ale-kopuru berdina uzten duen balioa da, aldez aurretik aldagaiaren balioak ordenatuta izango direlarik.

$$Me = \begin{cases} x_{n+1/2} & n \ bakoitia \\ \dfrac{x_{n/2} + x_{(n/2)+1}}{2} & n \ bikoitia \end{cases}$$

Kasu jarraituan:

$$\boxed{M_e = l_i + \frac{\frac{n}{2} - F_{i-1}}{f_i} d_i}$$

l_i = mediana daukan klasearen behe-muturra,

f_i = mediana daukan klasearen maiztasun absolutua,

d_i = mediana daukan klasearen luzera,

f_{i-1} = mediana daukan aurreko klasearen maiztasun absolutua,

F_{i-1} = mediana daukan aurreko klasearen maiztasun metatua,

n = laginaren tamaina

Moda, M_o: maiztasun handieneko balioa da.

Kasu jarraituan:

$$M_o = l_i + \frac{\dfrac{f_i}{d_i} - \dfrac{f_{i-1}}{d_{i-1}}}{\left(\dfrac{f_i}{d_i} - \dfrac{f_{i-1}}{d_{i-1}}\right) + \left(\dfrac{f_i}{d_i} - \dfrac{f_{i+1}}{d_{i+1}}\right)} d_i$$

l_i = moda daukan klasearen behe-muturra,

f_i = moda daukan klasearen maiztasun absolutua,

d_i = moda daukan klasearen luzera,

f_{i-1} = modaren aurreko klasearen maiztasun absolutua,

F_{i-1} = moda daukan aurreko klasearen maiztasun metatua

1.4.3. **Posizio-estatistikoak**

- ***P_k*, *k. ordenako pertzentilak*** banaketaren %*k* balio bere ezkerrean uzten du, non *k* = 1,2, …,99 den.

- Kasu jarraituan:

$$\boxed{P_k = l_i + \frac{\frac{kn}{100} - F_{i-1}}{f_i} d_i}$$

- ***D_k*, *k. ordenako dezilak***: *b*anaketaren %10*k* balio bere ezkerrean uzten du, non *k* = 1,2, …,9 den.

$$D_1 \equiv P_{10,} \quad D_2 \equiv P_{20}, \quad \ldots, D_9 = P_{90}$$

- ***Q_k*, *k. ordenako kuartilak***: *b*anaketaren %25*k* balio bere ezkerrean uzten du, non *k* = 1,2, 3 den.

$$Q_1 \equiv P_{25,} \quad Q_2 \equiv P_{50}, \quad Q_3 = P_{75}$$

- 1.4.2. **Sakabanapen-estatistikoak**

- ***Heina***: aldagaiaren balio handienaren eta balio txikienaren arteko diferentzia da.

$$\boxed{R = \max(x_i) - \min(x_i)}$$

- ***Bariantza***: $x_1, \ldots, x_k$ datuetarako $(x_1 - \bar{x}), (x_2 - \bar{x}), \ldots, (x_k - \bar{x})$ diferentzia edo erroreen bidez batez bestekoarekiko aldakuntza kalkula daiteke. Horrela, s^2 bariantzaren bidez errore karratuen batez bestekoa kalkulatzen da.

$$\boxed{s^2 = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^2 f_i}{n}}$$

•***Desbideratze tipikoa*** edo ***desbideratze estandarra, s***:

Bariantzaren erro karratu positiboa da. Halaber, batez besteko bereko bi datu multzoren batez bestekoarekiko sakabanapena konparatzeko erabil daiteke.

$$\boxed{s = \sqrt{Var(x)}}$$

•***Aldakuntza-koefizientea, CV:***

Unitate bereko bi multzoren aldakuntza edota unitate desberdineko aldakuntza aztertzea baimentzen du.

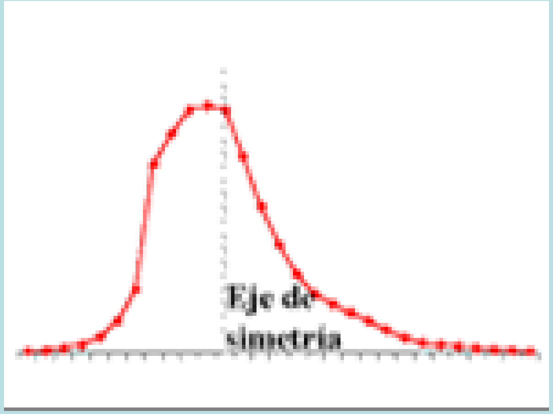
$$\boxed{CV = \frac{s}{\bar{x}}}$$

1.4.4. **Forma-estatistikoak**

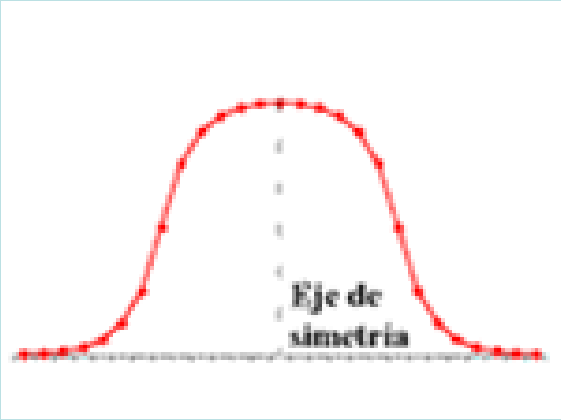
- **Alborapen-koefizientea,** batez bestekoarekiko simetria neurtzeko erabiltzen da.

$$\boxed{\nu = \frac{\overline{x} - M_o}{s}}$$

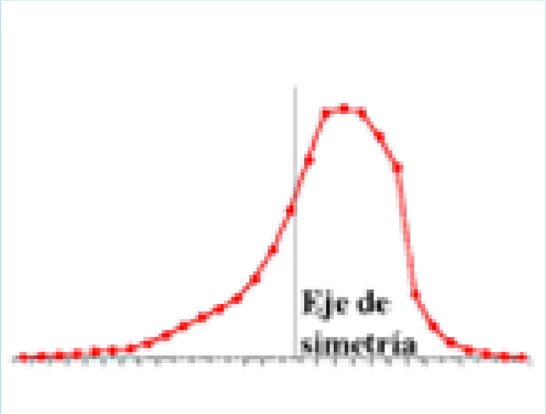
ν **> 0** *eskuinerantz alboratua*



ν **= 0** *simetrikoa*



ν **< 0** *ezkerrerantz alboratua*

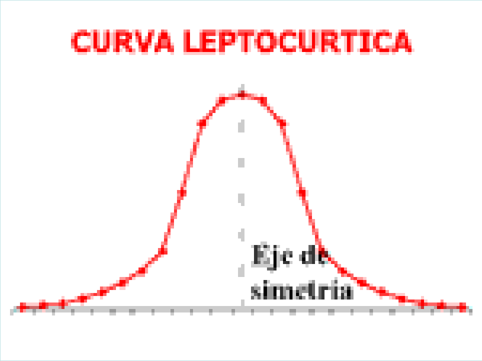


- **Kurtosia** honela definitzen da:

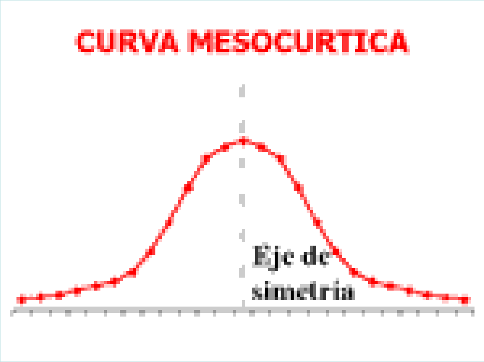
$$\boxed{g_2 = \frac{(1/n)\sum_{i=1}^{k}(x_i - \overline{x})^4 f_i}{s^4} - 3}$$

Banaketaren zorroztasuna neurtzen du.

$g_2 > 0$



$g_2 = 0$



$g_2 < 0$