1. Estatistika deskribatzailea. Aldagai bakuna

1.1. SARRERA

Estatistika multzo bati dagozkion zenbakizko datuak biltzen, sailkatzen eta aztertzen dituen zientzia da. Estatistikaren barruan bi arlo nagusi bereiz daitezke: **Estatistika Deskribatzailea** eta **Estatistika Induktiboa.** Estatistika deskribatzailearen bidez, aztergai den multzoari dagozkion datuak bildu, antolatu eta egoera deskribatzeko ezaugarriak lortuko dira. Estatistika induktiboaren bidez, aldiz, emaitzak orokortu, ondorioak atera edo aurresanak egin daitezke.

Azterketa estatistikoa multzo batean gauzatuko da. Multzo oso horri **populazioa** deritzo. Hala ere, kasu gehienetan azterketa populazioaren azpimultzoetan egiten da. Populazioaren edozein azpimultzori **lagina** deritzo. Populazioaren elementu bakoitza **unitate estatistiko** edo **ale estatistikoa** da. Populazioaren (edo laginaren) **tamaina** populazioaren (edo laginaren) elementu-kopurua da.

Azterketa estatistikoan aztergaiak har ditzakeen balioek **aldagai estatistikoa** determinatzen dute. Aldagaiak hartzen dituen balioak zenbakiak badira, **aldagai kuantitatiboa** dugu. Aldagaiak hartzen dituen balioak zenbakizkoak ez badira, aldiz, **aldagai kualitatiboa** dugu.

Aldagai aleatorio kuantitatiboak diskretuak edo jarraituak izan daitezke. **Aldagai kuantitatibo diskretua**k balio isolatuak hartzen ditu eta **aldagai kuantitatibo jarraitua**k tarte bateko edozein balio har dezake.

1.2. MAIZTASUN-TAULAK

Datuak bildu ondoren, komeni da datuak ordenatu eta tauletan adieraztea. Horretarako, honako kontzeptuak hartuko dira kontuan:

1. **Modalitateak:** x_i gaiak dira aldagaiaren balioak. Balio hauek zenbakiak direnean, oro har, txikienetik handienera ordenatzen dira.

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

3. **Maiztasun metatuak:** $F_i = f_1 + f_2 + \dots + f_i$ maiztasun absolutuen batura da.

Batzuetan interesgarria izan daiteke **maiztasun erlatiboa** $h_i = \dfrac{f_i}{n}$ eta **maiztasun metatu erlatiboa** $H_i = \dfrac{F_i}{n}$ kalkulatzea, non laginaren tamaina n eta $i = 1, 2, \ldots, k$ diren.

Askotan datu-kopurua handia denean, komenigarria da **maiztasun elkartuen metodoa** erabiltzea. Horretarako, ondoko pausoak eman daitezke:

i. Aldagai estatistikoaren *heina* kalkulatuko da, heina balio handienaren eta txikienaren arteko kendura delarik.

ii. Tartea, azpitartetan edo klasetan banatuko da. Klase-kopurua zehazteko orduan klase gutxi hartuz gero, informazioa gal daiteke, eta, aldiz, klase gehiegik azterketa zaildu dezakete. Beraz, klase-kopurua $k = \sqrt{n}$ aukera daiteke, non n gaia laginaren (populazioaren) tamaina den.

iii. Klaseak zehaztuko dira. Praktikan erosoa da $[l_i\ ,\ l_{i+1})$ erako klaseak aukeratzea. Klase bakoitzeko balioen ordezkari moduan klase-marka dugu, hots, tarte bakoitzeko erdiko puntua: $x_i = \dfrac{l_i + l_{i+1}}{2}$.

iv. Maiztasunak zenbatu eta taula eratuko da.

$[l_i\ ,\ l_{i+1})$	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	$F_k = f_1 + f_2 + \ldots + f_k = n$	$h_k = f_k/n$	$H_k = F_k/n = 1$
		$\sum_{i=1}^{k} f_i = n$		$\sum_{i=1}^{k} h_i = 1$	

1.3. ADIERAZPEN GRAFIKOAK

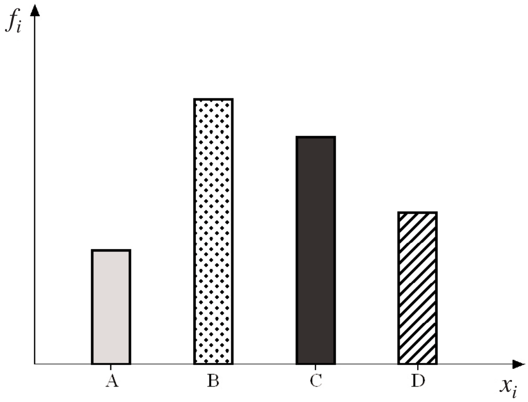
Adierazpen grafikoak oso lagungarriak izaten dira. Zenbaitetan begirada soil batez banaketaren ideia ematen dute. Aldagaiaren izaeraren arabera, hurrengo adierazpen grafikoak azter daitezke:

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

1.3.1. Aldagai kualitatiboaren adierazpen grafikoak

1.3.1.1. Barra-diagrama

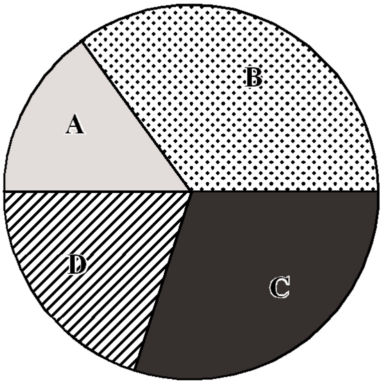
Barrez osaturiko diagrama bat da, non abzisa-ardatzean aldagaiaren balioak eta ordenatu-ardatzean maiztasun absolutuak adieraziko diren. Barrak ez dira elkarturik egoten.



1.1. irudia.

1.3.1.2. Sektore-diagrama

Zirkulu baten barruan aldagai kualitatiboaren balio bakoitzari dagokion sektorea adierazten da. Sektore bakoitzaren kalkulurako $\alpha_i = 360° \cdot \frac{f_i}{n}$ adierazpena erabiliko da, non f_i maiztasun absolutua eta n laginaren tamaina diren.

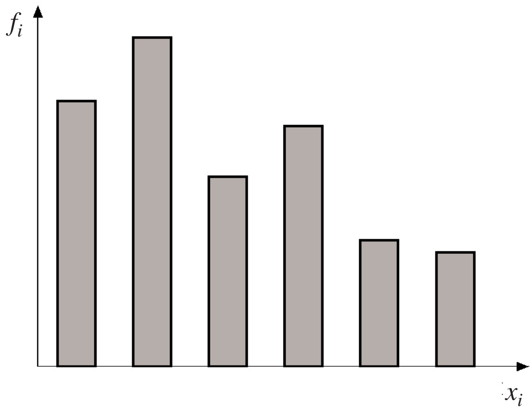


1.2. irudia.

1.3.2. Aldagai kuantitatibo diskretuaren adierazpen grafikoak

1.3.2.1. Barra-grafikoa

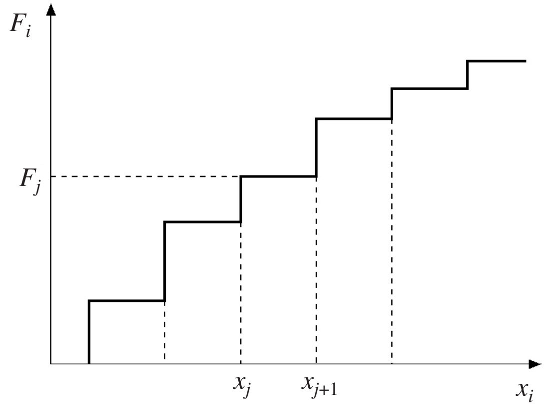
Abzisa-ardatzean aldagaiaren balioak eta ordenatu-ardatzean maiztasun absolutuak adierazten dituzten barrez osaturiko grafikoa da.



1.3. irudia.

1.3.2.2. Maiztasun metatuen grafikoa

Grafiko honetan abzisa-ardatzean aldagaiaren balioak eta ordenatu-ardatzean maiztasun metatuak adieraziko dira. Grafiko honek eskailera-itxura du. x_j balioari F_j altuera eta x_{j+1} balioari F_{j+1} altuera emango zaizkie, hurrenez hurren, (x_j, x_{j+1}) tarteko balioei F_j altuerako maila dagokielarik.

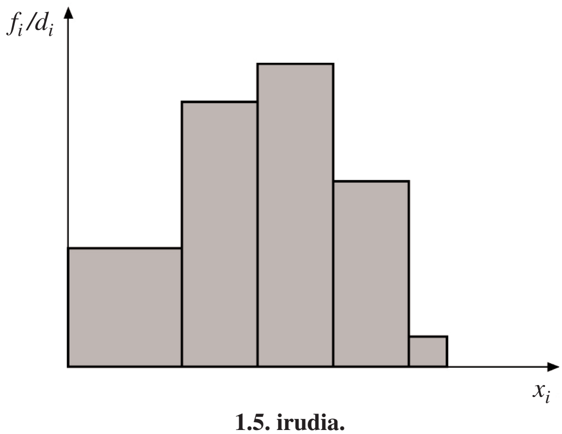


1.4. irudia.

1.3.3. Aldagai kuantitatibo jarraituaren adierazpen grafikoak

1.3.3.1. Histograma

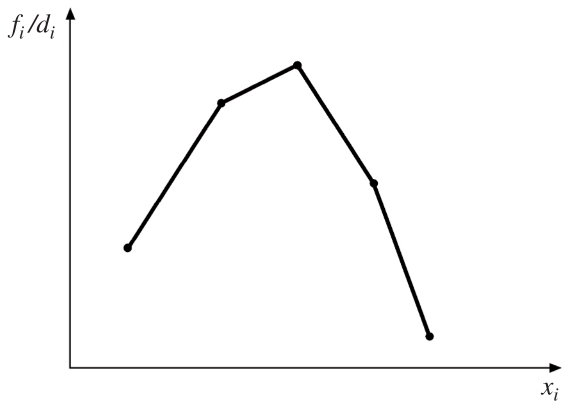
Oinarrian klaseak dituen errektangelu-multzoa da. Klase bakoitza errektangelu baten oinarria izango da. Klase guztiak anplitude berekoak badira, errektangelu bakoitzaren altuera klaseari dagokion maiztasun absolutua izango da. Klaseak anplitude berekoak ez direnean, altuerak f_i / d_i balioak izango dira, non kasu bakoitzean f_i maiztasun absolutua eta d_i klasearen anplitudea izango diren.



1.5. irudia.

1.3.3.2. Maiztasun absolutuen poligonoa

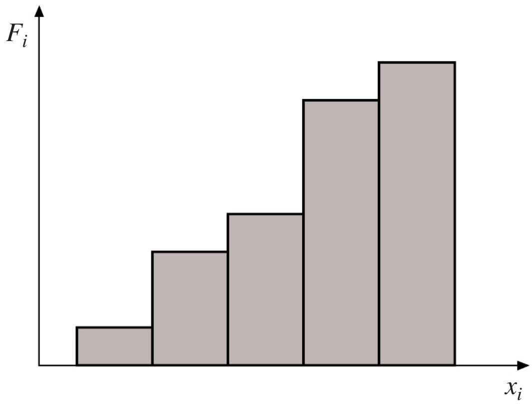
Aurreko histograman errektangeluetako goiko aldeetako erdiko puntuak lotuz, maiztasun absolutuen poligonoa eraikiko da.



1.6. irudia.

1.3.3.3 Maiztasun metatuen histograma

Errektangelu-multzo bat da, zeinaren oinarriak klaseak eta altuerak maiztasun metatuak diren.



1.7. irudia.

1.3.3.4 Maiztasun metatuen poligonoa

Aurreko histograman (l_{i+1}, F_i) puntuak lotuz lortuko den poligonoa da.

1.8. irudia.

1.4. ESTATISTIKO DESKRIBATZAILEAK

Estatistiko deskribatzaileak laginaren menpeko funtzioak dira. Aztergai den fenomenoaren ezaugarriak era laburtuan deskribatzeko balio dute. Hurrengo taulan zenbait estatistiko adieraziko dira era laburtuan. Taula azaldu baino lehen ohar batzuk kontuan hartzea komeni da:

1. X aldagai estatistiko kuantitatiboak $x_1, x_2, ..., x_k$ modalitateak $f_1, f_2, ..., f_k$ maiztasun absolutuz hartzen ditu eta laginaren tamaina n da.

2. Estatistikoek hurrengo sailkapena onartzen dute: i) joera zentralekoak (batez bestekoa, mediana, moda), ii) sakabanatzekoak (heina, kuartilarteko heina, bariantza, desbideratze tipikoa, aldakuntza-koefizientea), iii) posiziokoak (pertzentilak, dezilak eta kuartilak) eta iv) formakoak (alborapena, kurtosia).

3. Kasu jarraituan erabilitako notazioa:

   l_i = estatistikoa daukan klasearen behe-muturra

   f_i = estatistikoa daukan klasearen maiztasun absolutua

   d_i = estatistikoa daukan klasearen luzera

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

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

   n = laginaren tamaina

4. Bariantza adierazteko $V(x)$, s^2 edo $s^2\,(x)$ notazioak erabil daitezke.

   Kurtosiaren adierazpenean, bigarren ordenako momentua bariantza da, $m_2 = s^2$ alegia.

Estatistikoa	*Adierazpena*	*Esanahia*
Batez besteko aritmetikoa	$\bar{x} = \dfrac{\sum_{i=1}^{k} x_i f_i}{n}$	Aldagaiak hartzen dituen balio guztien baturaren eta laginaren arteko proportzioa da.
Mediana	Kasu diskretua: • n bakoitia denean, mediana $(n+1)/2$ posizioan dagoen balioa da. • n bikoitia denean, mediana $n/2$ eta $(n/2)+1$ posizioetako balioen erdiko puntua da. Kasu jarraitua: $M_e = l_i + \dfrac{\frac{n}{2} - F_{i-1}}{f_i} \cdot d_i$	Bere ezkerrean eta bere eskuinean ale-kopuru berdina uzten duen balioa da, eta aldez aurretik aldagaiaren balioak ordenatuta izango dira.
Moda	Kasu diskretua: f_i handiena duen x_i Kasu jarraitua: $M_o = l_i + \dfrac{\Delta_1}{\Delta_1 + \Delta_2} \cdot d_i$ $\Delta_1 = \dfrac{f_i}{d_i} - \dfrac{f_{i-1}}{d_{i-1}}, \ \Delta_2 = \dfrac{f_i}{d_i} - \dfrac{f_{i+1}}{d_{i+1}}$	Maiztasun handieneko balioa da.

Heina edo anplitudea	$R = max(x_i) - min(x_i)$	Laginaren balio handienaren eta txikienaren arteko kendura da.
Bariantza	$V(x) = \dfrac{\sum_{i=1}^{k} (x_i - \bar{x})^2 f_i}{n} = \dfrac{\sum_{i=1}^{k} x_i^2 f_i}{n} - (\bar{x})^2$	Batez bestekoarekiko errore karratuen batez bestekoa da.
Desbideratze tipikoa	$s(x) = \sqrt{V(x)}$	Bariantzaren erro karratu positiboa da.
c parametroarekiko r. ordenako momentua	$M_r(c) = \dfrac{\sum_{i=1}^{k} (x_i - c)^r f_i}{n}$	Parametroarekiko erroreen potentzien baturaren eta laginaren tamainaren arteko proportzioa da.
Momentu zentrala	$m_r = \dfrac{\sum_{i=1}^{k} (x_i - \bar{x})^r f_i}{n}$	Aurreko momentuan $c = \bar{x}$ eginez lortzen da.
Jatorriarekiko momentua	$a_r = \dfrac{\sum_{i=1}^{k} x_i^r f_i}{n}$	c-rekiko momentuan $c = 0$ eginez lortzen da.
Aldakuntza-koefizientea	$CV_x = \dfrac{s_x}{\bar{x}}$	Desbideratze tipikoaren eta batez bestekoaren arteko zatidura da. Aldagai batek bi multzotan edo bi aldagai ezberdinen aldakuntza konparatzeko balio du.
k. ordenako pertzentila	Kasu diskretua: Esanahia aplikatu. Kasu jarraitua: $P_k = l_i + \dfrac{\frac{kn}{100} - F_{i-1}}{f_i} d_i$	Banaketaren % k balio bere ezkerrean uzten du, non $k = 1, 2, \ldots, 99$ den.

k. ordenako dezila	$D_1 \equiv P_{10}$ $D_2 \equiv P_{20}$ .... $D_9 \equiv P_{90}$	Banaketaren % 10k balio bere ezkerrean uzten du, non $k = 1, 2, \ldots, 9$ den.
k. ordenako kuartila	$Q_1 \equiv P_{25}$ $Q_2 \equiv P_{50}$ $Q_3 \equiv P_{75}$	Banaketaren % 25k balio bere ezkerrean uzten du, $k = 1, 2, 3$ delarik.
Kuartilarteko heina	$R_I = Q_3 - Q_1$	3. ordenako eta 1. ordenako kuartilen arteko diferentzia da.
Alborapena	$v = \dfrac{\bar{x} - M_0}{s_X}$	Banaketaren simetria neurtzen du. $v > 0$ →eskuinerantz alboratua $v < 0$ →ezkerrerantz alboratua $v = 0$ →banaketa, alboragabea
Kurtosia	$g_2 = \dfrac{m_4}{m_2^2}$	Banaketaren zorroztasuna neurtzen du. $g_2 > 3$→banaketa leptokurtikoa $g_2 = 3$→banaketa mesokurtikoa $g_2 < 3$→banaketa platikurtikoa

1.5. ARIKETA EBATZIAK

1.5.1. ariketa

Ondoko balioak 30 torlojuren lodierak (mm) dira:

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

a. Eraiki bedi maiztasun-taula.

b. Irudika bedi barra-grafikoa.

c. Adieraz bedi “oso mehea” = 1, “mehea” = 2, “ertaina” = 3, “lodia” = 4 eta “oso lodia” = 5 lodierei dagokien sektore-diagrama.

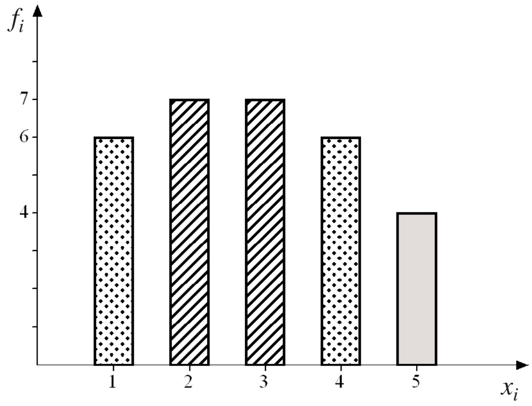
d. Esanahia azalduz, kalkula bitez batez bestekoa, moda, mediana eta desbideratze tipikoa.

Ebazpena

a. Hasteko, datuak ordenatu eta aldagai kuantitatibo diskretu honi dagokion maiztasun-taula eratuko da.

x_i	f_i	F_i	h_i	H_i
1	6	6	6/30 = 0,20	6/30 = 0,20
2	7	13	7/30 ≈ 0,23	13/30 ≈ 0,43
3	7	20	7/30 ≈ 0,23	20/30 ≈ 0,67
4	6	26	6/30 = 0,20	26/30 ≈ 0,87
5	4	30	2/15 ≈ 0,13	30/30 = 1

b. Barra-grafikoa:

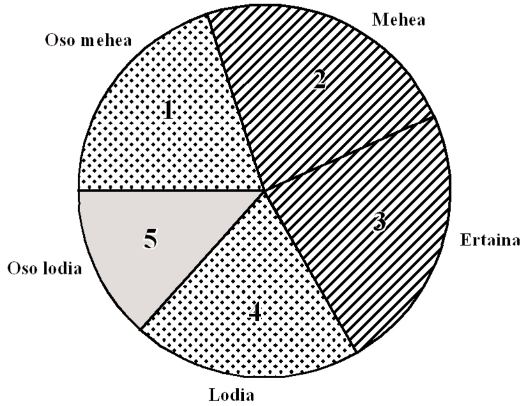


1.9. irudia.

c. Sektore-diagrama:

Hasteko, gai bakoitzari dagokion sektorea kalkulatuko da:

$$\alpha_1 = f_1 \cdot \frac{360}{n} = 6 \cdot \frac{360}{30} = 72° = \alpha_4, \alpha_2 = f_2 \cdot \frac{360}{n} = 7 \cdot \frac{360}{30} = 84° = \alpha_3, \alpha_5 = 48° .$$



1.10. irudia.

d. Batez bestekoaren kalkulua:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{n} =$$

$$= \frac{1 \cdot 6 + 2 \cdot 7 + 3 \cdot 7 + 4 \cdot 6 + 5 \cdot 4}{30} = \frac{85}{30} = 2,83 \text{ mm}.$$

Beraz, torloju-multzo horren batez besteko lodiera 2,83 mm-koa da. Batez bestekoak, beste estatistikorik kalkulatu gabe, multzoaren adierazle den lodiera azaltzen du.

Banaketak bi moda ditu, 2 eta 3 balioak hain zuzen ere, maiztasun handieneko balioak baitira. Kasu honetan, 2 eta 3 mm-ko lodierako torlojuak dira gehien azaltzen direnak laginean.

Medianaren balioa $M_e = 3$ da, neurtutako lodieren erdiak 3 mm baino txikiagoak eta beste erdiak gutxienez 3 mm-koak direlako. Beraz, torlojuen erdien lodiera [1, 3) mm-koa eta beste erdien lodiera [3, 5] mm-koa da.

Bariantzaren balioa hurrengoa da:

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

$$= \frac{1^2 \cdot 6 + 2^2 \cdot 7 + 3^2 \cdot 7 + 4^2 \cdot 6 + 5^2 \cdot 4}{30} - (2,83)^2 = 1,76 \text{ mm}^2.$$

Bariantzaren erro karratu positiboa kalkulatuz,

$$s = \sqrt{1,76} = 1,33 \text{ mm}$$

lor daiteke. Desbideratze tipikoak torlojuen batez besteko lodierarekiko erroreen batez bestekoa adierazten du eta bere balioa 1,33 mm-koa da.

1.5.2. ariketa

Hurrengo taulan 110 enpresaren fakturazioa azaltzen da:

Fakturazioa (milaka euro)	[0,10)	[10,20)	[20,30)	[30,40)	[40,50]
Enpresa-kopurua	15	25	32	23	15

a. Kalkula bedi gehien fakturatu den kantitatea zein den.

b. Zenbat enpresak fakturatzen dituzte gutxienez 32,17 mila euro? Adieraz ezazu emaitza maiztasun metatuen poligonoan.

Ebazpena

a. Adierazitako enpresen fakturazioa aldagai kuantitatibo jarraitua da. Hona hemen dagokion maiztasun absolutuen eta maiztasun metatuen taula:

$[l_i, l_{i+1})$	[0,10)	[10,20)	[20,30)	[30,40)	[40,50]
f_i	15	25	32	23	15
F_i	15	40	72	95	110

Bestalde, modaren kalkuluak gehien fakturatu den kantitatea adieraziko du.

$$M_o = l_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot d_i = 20 + \frac{\frac{32-25}{10}}{\left(\frac{32-25}{10}\right) + \left(\frac{32-23}{10}\right)} \cdot 10 = 20 + 4,375 = 24,375.$$

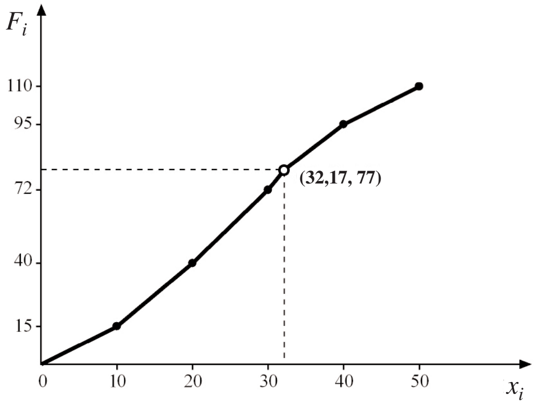
Beraz, maiztasun handieneko fakturazioa 24.375 eurokoa da.

b. Bigarren atal honetan eskatzen dena posizio-estatistiko bat zehaztea da. Horrela, $P_j = 32{,}17$ balioaren bidez % j enpresak 32,17 mila euro baino gutxiago fakturatzen dituztela adierazten da. Datuak ordezkatuz, pertzentilaren ordena zehaztuko da.

$$P_j = l_i + \frac{\frac{j \cdot n}{100} - F_{i-1}}{f_i} \cdot d_i = 30 + \frac{\frac{j \cdot 110}{100} - 72}{23} \cdot 10 = 32,17 \quad \rightarrow \quad j = 69,99.$$

Beraz, $j \approx 70$ da, hots, enpresen % 70ek (77 enpresak, alegia) 32.170 euro baino gutxiago fakturatzen ditu. Ondorioz, enpresen beste % 30ak gutxienez 32.170 euro fakturatuko ditu, kasu horretan 33 enpresa izango direlarik.

Grafikoki ikus daitekeenez, (32,17, 77) puntuak 77 enpresak 32,17 mila euro baino gutxiago fakturatzen dituztela adierazten du.



1.11. irudia.

1.5.3. ariketa

Ingurumenari kalteak eragiteagatik, 1.000 egunetan izandako isun-kopurua neurtu da eta egun bakoitzean 0 eta 5 isun bitartean jaso dira, hurrengo taulan ikus daitekeen moduan:

Isun-kopurua	0	1	2	3	4	5
Egun-kopurua	?	260	150	190	100	90

a. Kalkula bitez batez bestekoa, mediana, moda eta desbideratze tipikoa.

b. Lor itzazu 1. eta 3. ordenako kuartilak.

c. Adieraz bitez barra-grafikoa eta maiztasun metatuen grafikoa.

Ebazpena

a. Guztira 1.000 egunetako isun-kopurua neurtu denez, taulan falta den datua $f_1 = 210$ da.

$$1.000 = f_1 + 260 + 150 + 190 + 100 + 90 \rightarrow f_1 = 210.$$

Hori kontuan hartuz,

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{n} =$$

$$= \frac{0 \cdot 210 + 1 \cdot 260 + 2 \cdot 150 + 3 \cdot 190 + 4 \cdot 100 + 5 \cdot 90}{1.000} = \frac{1.980}{1.000} = 1{,}98 \text{ isun}$$

jaso dira, batez beste.

Mediana 2 da. Hau da egunen erdietan bi isun baino gutxiago eta beste egunen erdietan 2 eta 5 isun bitartean jaso dira.

Modaren balioa 1 da. Hau da, gehien jaso den isun-kopurua 1 da.

Bariantzaren balioa hurrengoa da:

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

$$= \frac{1^2 \cdot 260 + 2^2 \cdot 150 + 3^2 \cdot 190 + 4^2 \cdot 100 + 5^2 \cdot 90}{1.000} - (1,98)^2 = 2,5 \text{ (isun)}^2.$$

Bariantzaren erro karratu positiboa kalkulatuz, desbideratze tipikoa 1,58 isunekoa dela ondorioztatzen da.

$$s = \sqrt{2,5} = 1,58 \text{ isun}.$$

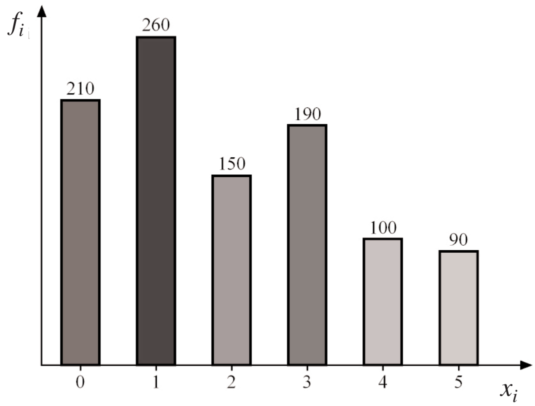
b. Kalkula ditzagun Q_1 eta Q_3 kuartilak. Horretarako osa ditzagun taulako datuak, maiztasun metatuen zutabea erantsiz.

x_i	0	1	2	3	4	5
f_i	210	260	150	190	100	90
F_i	210	470	620	810	910	1.000

Orduan, $\frac{25 \cdot n}{100} = \frac{25 \cdot 1.000}{100} = 250 < F_2$ denez, 1. ordenako kuartila $Q_1 = 1$ balioa da.

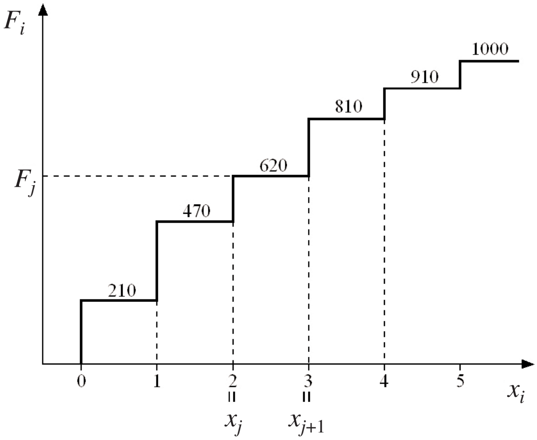
Era berean, $\frac{75 \cdot n}{100} = \frac{75 \cdot 1.000}{100} = 750 < F_4$ denez, 3. ordenako kuartila $Q_3 = 3$ balioa da.

c. Barra-grafikoa:



1.12. irudia.

Maiztasun metatuen grafikoa:



1.13. irudia.

1.5.4. ariketa

Ondoko taulan 80 langileren hileko soldatak azaltzen dira:

Soldatak (euro)	Ingeniariak (%)
1.500-1.560	0,15
1.560-1.600	0,15
1.600-1.680	0,25
1.680-1.740	0,20
1.740-1.800	0,15
1.800-1.860	0,10

a. Kalkula bedi langile horien zein proportziok irabazten duen hilero $\bar{x} + s$ euro baino gutxiago.

b. Lagin horretako 4 langileren soldata gutxienez A eurokoa da. Estatistiko egoki bat erabiliz, zehatz bedi A kantitatea.

c. Azal bedi aurreko ataleko A kantitatearen interpretazio grafikoa aldagaiaren histograman.

Ebazpena

a. Enuntziatuko datuei honako maiztasunen taula dagokie:

$[l_i, l_{i+1})$	h_i	x_i	f_i	F_i
[1.500,1.560)	0,15	1.530	12	12
[1.560, 1.600)	0,15	1.580	12	24
[1.600, 1.680)	0,25	1.640	20	44
[1.680, 1.740)	0,20	1.710	16	60
[1.740, 1.800)	0,15	1.770	12	72
[1.800, 1.860]	0,10	1.830	8	80

Batez bestekoaren kalkulua:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{n} = \frac{1.530 \cdot 12 + 1.580 \cdot 12 + \ldots + 1.830 \cdot 8}{80} = \frac{133.360}{80} = 1.667 \text{ euro.}$$

Bariantzaren kalkulua:

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

$$= \frac{(1.530)^2 \cdot 12 + (1.580)^2 \cdot 12 + \ldots + (1.830)^2 \cdot 8}{80} - (1.667)^2 = 8.751 \text{ (euro)}^2.$$

Ondorioz, desbideratze tipikoaren balioa hurrengoa da:

$$s = \sqrt{8.751} = 93,55 \text{ euro.}$$

$\bar{x}$ + s = 1.667+ 93,55 = 1.760,55 euro da. Orduan, hilero 1.760,55 euro baino gutxiago irabazten dituzten langileen proportzioa kalkulatu behar da. Horretarako P_k pertzentilaren ordena zehaztu behar da.

$$P_k = l_i + \frac{\frac{k \cdot n}{100} - F_{i-1}}{f_i} \cdot d_i.$$

Kasu honetan, hurrengo ekuazioa ebatziko da:

$$1.760,55 = 1.740 + \frac{\frac{k \cdot 80}{100} - 60}{12} \cdot 60 \quad \rightarrow$$

$$k = 80,14.$$

Beraz, 80 langileen % 80,14k hilero 1.760,55 euro baino gutxiago irabazten ditu.

b. Erabiliko den estatistikoa berriro ere pertzentila da. 4 langilek laginaren % 5 osatzen dutenez, erantzuna 95. ordenako pertzentilaren bidez kalkula daiteke.

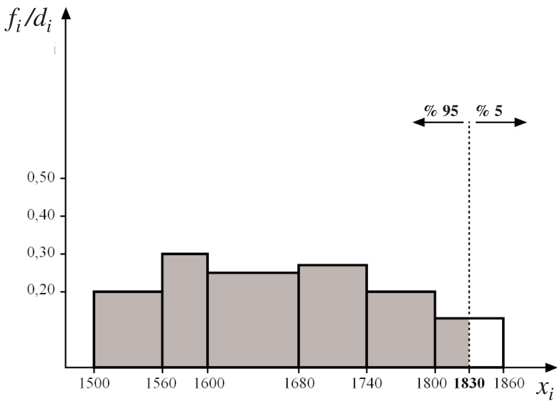
$$P_{95} = l_i + \frac{\frac{95 \cdot n}{100} - F_{i-1}}{f_i} \cdot d_i = 1.800 + \frac{\frac{95 \cdot 80}{100} - 72}{8} \cdot 60 = 1.830 \text{ euro.}$$

Hau da, 4 langileren hileko soldatak gutxienez A = 1.830 eurokoak dira.

c. Aldagai estatiskoari dagokion histograman abzisa-ardatzean klaseak eta ordenatu-ardatzean f_i/d_i balioak adieraziko dira, non f_i gaiak maiztasun absolutuak eta d_i balioak klaseen anplitudeak diren.

$[l_i, l_{i+1})$	f_i	d_i	f_i/d_i
[1.500,1.560)	12	60	0,20
[1.560, 1.600)	12	40	0,30
[1.600, 1.680)	20	80	0,25
[1.680, 1.740)	16	60	0,27
[1.740, 1.800)	12	60	0,20
[1.800, 1.860]	8	60	0,13

Histogramaren adierazpen grafikoa honako hau da:



1.14. irudia.

Bertan ikus daitekeenez, A = 1.830 balioa posizio estatistiko bat da, 95. ordenako pertzentila hain zuzen ere. Beraz, kasu honetan banaketaren % 95 geratzen da 1.830 balioaren ezkerrean eta beste % 5a bere eskuinean dago.

1.5.5. ariketa

Hurrengo datuek pertsona baten iazko elektrizitate-gastuak (euro) adierazten dituzte:

Urt.	Ots.	Mar.	Api.	Mai.	Eka.	Uzt.	Abu.	Ira.	Urr.	Aza.	Abe.
35,6	36	36,2	35,6	27,6	25,8	27,6	32,4	32,4	32,4	32,4	33,6

a. Nolakoa da aldagai hau, kuantitatiboa ala kualitatiboa?

b. Zein da urteko batez besteko elektrizitate-gastua?

c. Lor bedi desbideratze tipikoaren balioa.

d. Kalkula bedi medianaren balioa eta adierazi bere esanahia.

Ebazpena

a. Pertsona baten iazko elektrizitate-gastuak zenbakizko balioak direnez, aldagai kuantitatiboa da. Gainera, balio puntualak daudenez, aldagai kuantitatibo diskretua dela erantsi daiteke.

Halaber, aldagaiari hurrengo maiztasun-taula dagokio:

x_i	25,8	27,6	32,4	33,6	35,6	36	36,2
f_i	1	2	4	1	2	1	1
F_i	1	3	7	8	10	11	12

b. Urteko batez besteko elektrizitate-gastua:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{n} = \frac{25{,}8\cdot 1 + 27{,}6\cdot 2 + \ldots + 36{,}2\cdot 1 +}{12} = \frac{387{,}6}{12} = 32{,}3 \text{ euro.}$$

Bariantzaren kalkulua:

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

$$= \frac{(25{,}8)^2\cdot 1 + (27{,}6)^2\cdot 2 + \ldots + (36{,}2)^2\cdot 1}{12} - (32{,}3)^2 = 11{,}57 \text{ (euro)}^2.$$

Desbideratze tipikoa, bariantzaren erro karratu positiboaren emaitza da:

$$s = \sqrt{11{,}57} = 3{,}4 \text{ euro.}$$

d. Mediana kalkulatzeko, kasu honetan n = 12 bikoitia denez, mediana seigarren eta zazpigarren balioen erdiko puntua da. Ondorioz,

$$M_e = 32{,}4 \text{ euro}$$

da. Horrek esan nahi du, sei hilabetetako gastuak 32,4 euro baino gutxiagokoak eta beste sei hilabeteetako gastuak gutxienez 32,4 eurokoak izan direla.

1.5.6. ariketa

Lan-talde batek produkzio-plantako produkzio-unitate bakoitza egiteko behar izan den denbora neurtzen du. Produkzio-planta batean 30 egunetako datuak bildu dira eta datu horien bidez egun bakoitzean produkzio-unitateak egiteko guztira behar izan dituzten orduak neurtu dira:

128	119	96	97	124	128	142	98	108	120
114	109	124	132	97	138	133	136	120	112
144	128	103	135	114	109	100	111	131	113

Datuok anplitude bereko bost klasetan bilduz:

a. Kalkula bitez produkzio-unitate horien ekoizpenerako batez besteko denbora eta maiztasun handieneko ordu-kopurua.

b. Alborapen-koefizientearen balioaren arabera, zer esan daiteke?

Ebazpena

a. Emandako banaketa, datuak bost klasetan bilduz, honela adieraz daiteke:

$[l_i, l_{i+1})$	x_i	f_i
[95, 105)	100	6
[105, 115)	110	8
[115, 125)	120	5
[125, 135)	130	6
[135, 145)	140	5

Produkzio-unitateen ekoizpenerako, batez beste, 118,67 ordu behar dira:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{n} = \frac{3.560}{30} = 118{,}67 \text{ ordu.}$$

Bestalde, modaren balioa kalkulatuko da.

$$M_o = l_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot d_i = 105 + \frac{\frac{8-6}{10}}{\left(\frac{8-6}{10}\right) + \left(\frac{8-5}{10}\right)} \cdot 10 = 105 + 4 = 109 \text{ ordu.}$$

Orduan, maiztasun handienez 109 ordu behar dira pieza horiek ekoizteko.

b. Alborapen-koefizientearen balioa zehazteko, batez bestekoa eta modaren balioak aurreko atalean lortu direnez, orain desbideratze tipikoaren balioa kalkulatu behar da.

Lehenago, bariantzaren kalkulua egin behar da:

$$s^2 = \frac{\sum_{i=1}^{k} x_i^2 f_i}{n} - (\bar{x})^2 = \frac{428.200}{30} - (118,67)^2 = 190,76 \text{ (ordu)}^2.$$

Orduan, desbideratze tipikoa $s = \sqrt{190,76} = 13,81$ ordu da.

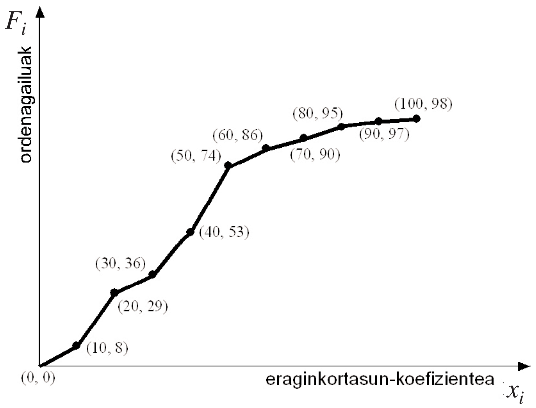
Bukatzeko, alborapen-koefizientearen balioa kalkulatuko da.

$$v = \frac{\bar{x} - M_o}{s} = \frac{118,67 - 109}{13,81} = 0,7.$$

Alborapen-koefizientea positiboa denez, banaketa eskuinerantz alboratua dela esan daiteke.

1.5.7. ariketa

Ordenagailuen eraginkortasuna neurtzeko 0 eta 100 bitarteko koefizienteak ezarri dira. Horrela, ondoko grafikoan 98 ordenagailuren eraginkortasuna neurtzen duen koefizientea maiztasun metatuen poligonoaren bidez adierazten da:



1.15. irudia.

a. Eraiki bedi aldagai kuantitatibo jarraitu honi dagokion maiztasun absolutu eta metatuen taula.

b. Kalkula itzazu bariantza, desbideratze tipikoa eta aldakuntza-koefizientearen balioak.

c. Jakinda 49 ordenagailuren eraginkortasunaren koefizientea x balioa baino txikiagoa dela, zenbatekoa da x balioa?

Ebazpena

a. Maiztasun metatuen grafikoan (l_{i+1}, F_i) puntuak adierazten dira, l_{i+1} balioa klasearen goi-muturra eta F_i gaia x_i balio bakoitzari dagokion maiztasun metatua izanik. Orduan, maiztasun absolutu eta metatuen taula osatzeko, aldagai kuantitatibo jarraitua hamar anplitudeko klasetan bana daiteke.

$[l_i, l_{i+1})$	x_i	f_i	F_i
[0, 10)	5	8	8
[10, 20)	15	21	29
[20, 30)	25	7	36
[30, 40)	35	17	53
[40, 50)	45	21	74
[50, 60)	55	12	86
[60, 70)	65	4	90
[70, 80)	75	5	95
[80, 90)	85	2	97
[90, 100)	95	1	98

b. Lehenengo batez bestekoaren balioa, eta gero, eskatutako estatistikoen balioak kalkulatuko dira.

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{n} = \frac{3.630}{98} = 37,04.$$

Bariantza:

$$s^2 = \frac{\sum_{i=1}^{k} x_i^2 f_i}{n} - (\bar{x})^2 = \frac{177.450}{98} - (37,04)^2 = 438,69.$$

Desbideratze tipikoa:

$$s = \sqrt{438,69} = 20,94.$$

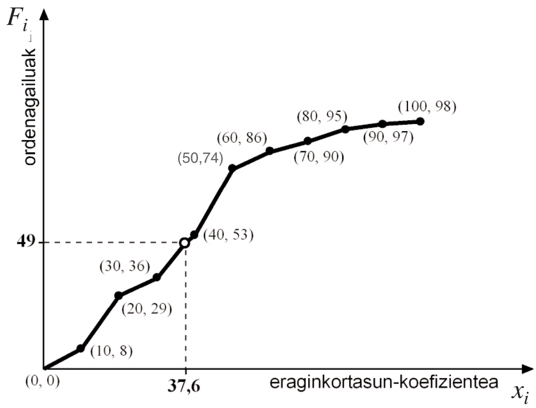
Aldakuntza-koefizientea:

$$CV_x = \frac{s_x}{\bar{x}} = \frac{20,94}{37,04} = 0,565.$$

c. P_k pertzentilaren bidez, laginaren % k ordenagailuk duten eraginkortasuna P_k balioa baino txikiagoa dela adierazten da. Kasu honetan, 49 ordenagailu laginaren erdia direla ikus daiteke. Beraz, problema mediana kalkulatuz ebatz dezakegu:

$$M_e = l_i + \frac{\frac{n}{2} - F_{i-1}}{f_i} \cdot d_i = 30 + \frac{49 - 36}{17} \cdot 10 = 37,6.$$

Ondorioz, ordenagailu erdien eraginkortasunaren koefizientea 37,6 baino txikiagoa da.



1.16. irudia.

1.5.8. ariketa

A termometroarekin neurtutako tenperaturen erroreek 1 ºC-eko batez bestekoa eta 0,01 ºC-ko desbideratze tipikoa erakutsi dituzte. Bestalde, B termometroarekin neurtutako tenperaturen erroreek 0,2 ºC-ko batez bestekoa eta 0,08 ºC-ko desbideratze tipikoa adierazi dituzte.

a. Zein termometro da erlatiboki zehatzagoa, A ala B?

b. Gaur goizeko zortzietan 8 ºC-ko tenperatura zegoenean, A termometroak 9,5 ºC eta B termometroak 9,6 ºC neurtu dituzte. Zein termometrok erakutsi du erlatiboki errore txikiagoa goizeko zortzietan?

Oharra: errorea = $|$benetako tenperatura − neurtutako tenperatura$|$

Ebazpena

a. Erlatiboki zehatzagoa den termometroa zehazteko, aldakuntza-koefizienteak kalkulatuko dira:

$$CV_A = \frac{s_A}{\bar{x}_A} = \frac{0,01}{1} = 0,01,$$

$$CV_B = \frac{s_B}{\bar{x}_B} = \frac{0,08}{0,2} = 0,4.$$

B termometroarekin egindako neurketei dagokien aldakuntza-koefizienteak % 40ko aldakuntza adierazten du. A termometroari dagokionez, aldiz, % 1eko aldakuntza du. Ondorioz, $CV_A < CV_B$ denez, A termometroa zehatzagoa da.

b. Errorea benetako tenperaturaren eta neurtutako tenperaturaren arteko diferentzia denez, A eta B termometroek goizeko zortzietan jasotako neurrien erroreak

$$x_A = |8 - 9,5| = 1,5,$$

$$x_B = |8 - 9,6| = 1,6$$

dira, hurrenez hurren. Tipifikazioa aplikatuz, balio bakoitzaren posizio erlatiboa kalkula daiteke:

$$z_A = \frac{x_A - \bar{x}_A}{s_A} = \frac{1,5 - 1}{0,01} = 50,$$

$$z_B = \frac{x_B - \bar{x}_B}{s_B} = \frac{1,6 - 0,2}{0,08} = 17,5$$

$z_B < z_A$ denez, B termometroaren goizeko zortzietako neurriaren errorea erlatiboki txikiagoa da.

1.5.9. ariketa

Lan-talde batean kualifikazioa eta produktibitatea kalifikatu dira: kualifikazioaren batez besteko puntuazioa 10 eta desbideratze tipikoa 2,2 eta produktibitatearen batez besteko puntuazioa 30 eta desbideratze tipikoa 13 izan dira. Marta eta Mikel lan-talde horretako bi partaide dira. Martak kualifikazioan 12 eta produktibitatean 35 puntuazioak lortu ditu. Mikelen puntuazioak kualifikazioan eta produktibitatean 11 eta 38 dira, hurrunez hurren. Kualifikazioa eta produktibitatea batera hartuz, globalki, nork du puntuazio hobea?

Ebazpena

Enuntziatuko datuen arabera, bi alderdi neurtzen dira:

Kualifikazioa	Produktibitatea
$\bar{x}_1 = 10$	$\bar{x}_2 = 30$
$s_1 = 2,2$	$s_2 = 13$

Martaren eta Mikelen puntuazioak kualifikazioan eta produktibitatean:

	Kualifikazioa	Produktibitatea
Mikel	$x_1^1 = 11$	$x_2^1 = 38$
Marta	$x_1^2 = 12$	$x_2^2 = 35$

Bi alderdiak batera tratatzeko puntuazioak tipifikatu behar dira.

$$\text{Mikel: } z_1^1 = \frac{11-10}{2,2} = 0,45, \qquad z_2^1 = \frac{38-30}{13} = 0,62,$$

$$\text{Marta: } z_1^2 = \frac{12-10}{2,2} = 0,91, \qquad z_2^2 = \frac{35-30}{13} = 0,38.$$

Puntuazio globala lortzeko, bi puntuazioen batez bestekoa kalkulatuko da.

$$\text{Mikelen puntuazio globala: } \frac{z_1^1 + z_2^1}{2} = \frac{0,45+0,62}{2} = 0,535,$$

$$\text{Martaren puntuazio globala: } \frac{z_1^2 + z_2^2}{2} = \frac{0,91+0,38}{2} = 0,645.$$

Ondorioz, Martaren puntuazio globala Mikelena baino hobea da.

1.5.10. ariketa

Hogei hiritako hileko euri-kantitateak (l/m^2) neurtu dira eta emaitzak anplitude ezberdinetako sei klasetan bildu dira. Anplitudeak honakoak dira: $d_1 = 8$, $d_2 = 8$, $d_3 = 4$, $d_4 = 4$, $d_5 = 8$, $d_6 = 10$, non d_i balioa *i.* klasearen anplitudea den. Bestalde, dagozkien maiztasun erlatibo metatuak hauek dira, hurrenez hurren: H_1= 0,10, H_2 = 0,10, H_3 = 0,55, H_4 = 0,75, H_5 = 0,95, H_6 = 1.

a. Jaso den euri-kantitate txikiena 65 l/m^2-koa dela jakinda, kalkula bedi zein den hogei hirietako batez besteko hileko euri-kantitatea.

b. Egia al da hamar hiritako euri-kantitatea 73 l/m^2 eta 86 l/m^2 bitartekoa izan zela?

Ebazpena

a. Hasteko, euri-kantitate txikiena x_i balio txikiena 65 denez, eta klaseen anplitudeak ezagunak direnez, klaseak eta klase-markak zehatz daitezke. Bestealde, H_i maiztasun erlatibo metatuak ezagunak direnez eta n = 20 denez, maiztasun erlatiboak, maiztasun absolutuak eta maiztasun absolutu metatuak kalkula daitezke. Horretarako kontuan hartuko diren adierazpenak hurrengoak dira: $h_i = H_i - H_{i-1}$, $F_i = 20H_i$ eta $f_i = F_i - F_{i-1}$.

$[l_i, l_{i+1})$	x_i	H_i	h_i	F_i	f_i
[65, 73)	69	0,10	0,10	2	2
[73, 81)	77	0,10	0	2	0
[81, 85)	83	0,55	0,45	11	9
[85, 89)	87	0,75	0,20	15	4
[89, 97)	93	0,95	0,20	19	4
[97, 107)	102	1	0,05	20	1

Batez besteko hileko euri-kantitatea hauxe da:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{n} = \frac{69 \cdot 2 + 77 \cdot 0 + 83 \cdot 9 + 87 \cdot 4 + 93 \cdot 4 + 102 \cdot 1}{20} = \frac{1.707}{20} = 85,35 \text{ l/m}^2.$$

b. Pertzentilak erabiliz, 10. ordenako pertzentila kalkulatuz, bi hiritako euri-kantitateak 65-73 l/m^2-koak izan zirela ikus daiteke.

$$73 = P_k = l_i + \frac{\frac{k \cdot n}{100} - F_{i-1}}{f_i} \cdot d_i = 65 + \frac{\frac{k \cdot 20}{100} - 0}{2} \cdot 8 \quad \rightarrow \quad k = 10.$$

Era berean, 60. ordenako pertzentila kalkulatuz, 12 hiritako euri-kantitateak 86 l/m^2 baino txikiagoak izan zirela nabari da.

$$86 = P_m = l_j = \frac{\frac{m \cdot n}{100} - F_{j-1}}{f_j} \cdot d_j = 85 + \frac{\frac{m \cdot 20}{100} - 11}{4} \cdot 4 \quad \rightarrow \quad m = 60.$$

Ondorioz, hamar hiritako euri-kantitateak 73-86 l/m^2 bitartekoak izan ziren.

1.5.11. ariketa

Berrogei pertsonari test psikometriko bat aplikatu zaie eta hurrengo puntuazioak lortu dira:

Puntuazioak	0-4	4-8	8-12	12-16	16-19
Pertsona-kopurua	6	15	10	5	4

a. Kalkula bitez mediana eta moda. Azaldu beraien esanahia.

b. Erabil ezazu estatistiko egoki bat 13 puntu baino gutxiago lortu duen pertsona-kopurua zehazteko.

Ebazpena

a. Biz X = "test psikometrikoaren puntuazioak" aldagai estatistikoa. Taulako puntuazioak klasetan bil daitezke. Ondoren maiztasun metatuak kalkulatuko dira.

$[l_i, l_{i+1})$	[0, 4)	[4, 8)	[8, 12)	[12,16)	[16,19)
f_i	6	15	10	5	4
F_i	6	21	31	36	40

Mediana da bere eskuinaldean banaketaren % 50 uzten duen balioa.

$$M_e = l_i + \frac{\frac{n}{2} - F_{j-1}}{f_i} \cdot d_i = 4 + \frac{\frac{40}{2} - 6}{15} \cdot 4 = 7,73.$$

Beraz, lagineko pertsonen erdiek ez dute gainditu 7,73 puntuazioa.

Moda, maiztasun absolutu handieneko balioa da. Kasu honetan, hurrengo balioa du:

$$M_o = l_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot d_i = 4 + \frac{\frac{15-6}{4}}{\frac{15-6}{4} + \frac{15-10}{4}} \cdot 4 = 6,57.$$

Beraz, gehien lortu den puntuazioa 6,57 izan da.

b. P_k pertzentilaren bidez, banaketaren % k pertsonak lortu duten puntuazioa P_k baino txikiagoa dela adierazten da.

$$P_k = l_i + \frac{\frac{k \cdot n}{100} - F_{i-1}}{f_i} \cdot d_i = 12 + \frac{\frac{k \cdot 40}{100} - 31}{5} \cdot 4 = 13,$$

$$k = 80,625.$$

Ondorioz, lagineko pertsonen % 80,625ek lortu duen puntuazioa 13 baino txikiagoa da, hots, 32 pertsonak 13 baino puntuazio txikiagoa lortu dute.

1.5.12. ariketa

Gasteizen eta Bilbon kokatutako bi enpresa ekoizpena aztertzen ari dira. Ondoko taulan egunean ekoitzitako pieza-kopurua neurtu da:

Pieza-kopurua	Enplegatu-kopurua Gasteiz	Enplegatu-kopurua Bilbo
0-30	94	85
30-50	140	160
50-70	160	160
70-90	98	89
90-100	8	6

Enpresen plangintzetan, ekoizpenaren arabera (ekoizpen txikienetik handienera), lau langile-kategoria ezarriko dira: A kategorian enplegatuen % 65, B kategorian enplegatuen % 20, C kategorian enplegatuen % 10 eta D kategorian enplegatuen % 5.

a. D kategoriako langile izateko ekoizpen minimoa, non da handiagoa, Gasteizen ala Bilbon?

b. Irudika bitez kasu bakoitzeko histograma eta maiztasun metatuen poligonoa, aurreko atalean kalkulatutako estatistikoen balioak adieraziz.

Ebazpena

a. Hasteko, hona hemen kasu bakoitzari dagokion taula:

Gasteiz				Bilbo		
f_i^G	f_i^G / d_i	F_i^G	$[l_i, l_{i+1})$	f_i^B	f_i^B / d_i	F_i^B
94	3,13	94	[0, 30)	85	2,83	85
140	7	234	[30, 50)	160	8	245
160	8	394	[50, 70)	160	8	405
98	4,9	492	[70, 90)	89	4,45	494
8	0,8	500	[90, 100)	6	0,6	500

Lehenengo, kalkula dezagun Gasteizko kasurako 95. ordenako pertzentila.

$$\frac{500 \cdot 95}{100} = 475 \text{ denez, orduan } P_{95}^{G} \in [70, 90) \text{ dago.}$$

$$P_{95}^{G} = 70 + \frac{475 - 394}{98} \cdot 20 = 86,53.$$

Ondorioz, Gasteizko langileen artean 87 eta 100 pieza bitartean ekoizten dituzten langileak D kategorian daude.

Orain, kalkula dezagun Bilboko langileen kasurako 95. ordenako pertzentilaren balioa.

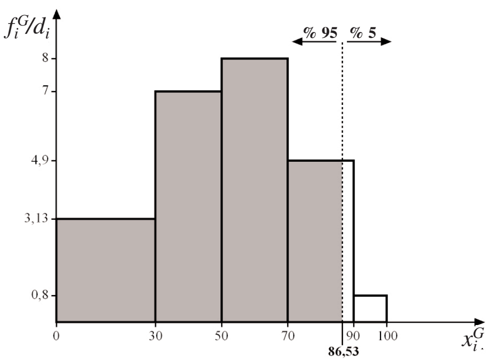
$$\frac{500 \cdot 95}{100} = 475 \text{ denez, orduan } P_{95}^{B} \in [70, 90) \text{ dago.}$$

$$P_{95}^{B} = 70 + \frac{475 - 405}{89} \cdot 20 = 85,73.$$

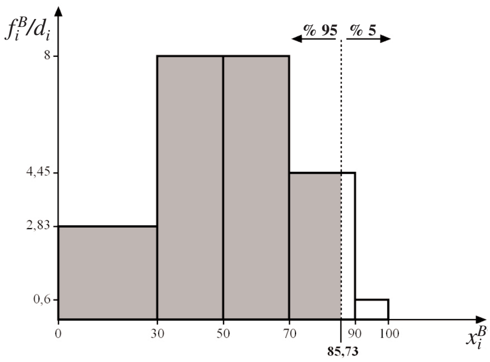
Hau da, Bilboko langileen artean 86 eta 100 pieza bitartean ekoizten dituztenak daude D kategorian.

Gasteizko kasuan 95. ordenako pertzentilaren balioa handiagoa denez, D kategoriako langile izateko ekoizpen minimoaren balioa Gasteizen handiagoa da.

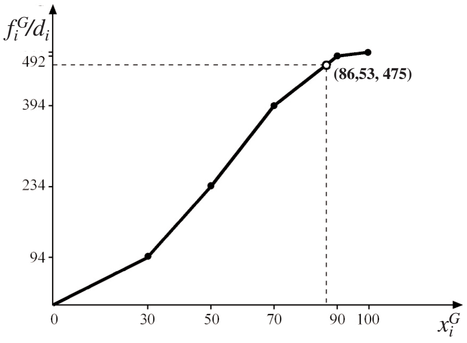
b. 1.17. eta 1.18. irudietan kasu bakoitzeko histogramak adierazi dira. Era berean, 1.19. eta 1.20. irudietan Gasteizko eta Bilboko kasuetarako maiztasun metatuen poligonoak nabari daitezke.



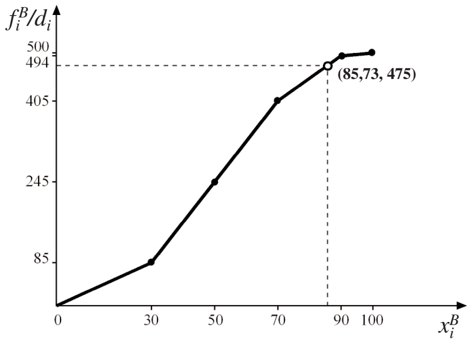
1.17. irudia



1.18. irudia



1.19. irudia



1.20. irudia

1.5.13. ariketa

Ondoko taulan 40 lanpararen iraupenak azaltzen dira:

Iraupena (urte)	2	6	10	14	18
Lanpara-kopurua	6	15	10	5	4

a. Kalkula bedi lanparen batez besteko iraupena, mediana eta desbideratze tipikoa.

b. Beste mota bateko lanparen batez besteko iraupena 11 urte eta desbideratze tipikoa 5,5 urte izan dira. Zein datu daude erlatiboki kontzentratuago?

c. Enuntziatuko 40 lanparetako batek 10 urte eta b) ataleko beste lanpara batek 12 urte iraun dute. Zein lanparak izan du erlatiboki iraupen handiagoa?

Ebazpena

a. Biz X = “lanparen iraupena” aldagai aleatorio diskretua. Zehatz dezagun modalitateaz, maiztasun absolutuaz eta maiztasun metatuaz osaturiko taula:

x_i	2	6	10	14	18
f_i	6	15	10	5	4
F_i	6	21	31	36	40

Lanparen batez besteko iraupena honela kalkulatuko da:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{n} = \frac{344}{40} = 8,6 \text{ urte.}$$

Medianaren kalkulua:

n = 40 bikoitia denez, mediana 20. eta 21. posizioetako balioen erdiko puntua da. Beraz, M_e = 6 urte da. Ondorioz, lanparen erdiek 2 edo 6 urteko iraupena eta beste lanparen erdiek 6, 10, 14 edo 18 urteko iraupena dute.

Bariantza:

$$s_x^2 = \frac{\sum_{i=1}^{6} x_i^2 f_i}{n} - (\bar{x})^2 =$$

$$= \frac{2^2 \cdot 6 + 6^2 \cdot 15 + 10^2 \cdot 10 + 14^2 \cdot 5 + 18^2 \cdot 4}{40} - (8,6)^2 = 22,04 \text{ (urte)}^2.$$

Bariantzaren erro karratu positiboa kalkulatuz, desbideratze tipikoaren balioa lortuko da:

$$s = \sqrt{22,04} = 4,69 \text{ urte}.$$

b. Kalkula ditzagun aldakuntza-koefizienteen balioak bi lanpara-multzoetarako.

Lehenengo kasuan, aldakuntza-koefizientearen balioa hauxe da:

$$CV_x = \frac{s_x}{\bar{x}} = \frac{4,69}{8,6} = 0,545.$$

Bigarren atal honetako lanparei dagokien aldakuntza-koefizientea honakoa da:

$$CV_y = \frac{s_y}{\bar{y}} = \frac{5,5}{11} = 0,5.$$

Ondorioz, bigarren multzoko lanparen iraupenari dagokion aldakuntza-koefizientea txikiagoa denez, datu hauek erlatiboki kontzentratuago daude.

c. Galdera honi erantzuna emateko tipifikazioa aplikatuko da.

I. multzoko lanparak	II. multzoko lanparak
$\bar{x}_1 = 8,6$	$\bar{x}_2 = 11$
$s_1 = 4,69$	$s_2 = 5,5$
$x_1 = 10$	$x_2 = 12$

I. multzoko lanpararen balio tipifikatua: $z_1 = \dfrac{x_1 - \bar{x}_1}{s_1} = \dfrac{10 - 8,6}{4,69} = 0,299.$

II. multzoko lanpararen balio tipifikatua: $z_2 = \dfrac{x_2 - \bar{x}_2}{s_2} = \dfrac{12 - 11}{5,5} = 0,182.$

$z_1 > z_2$ denez, lehenengo multzotik hartutako lanparak erlatiboki gehiago irauten du.

1.5.14. ariketa

Hona hemen marka ezagun bateko 27 autoren gasolina-kontsumoa (litro) 100 km-ko ibilbidean:

2,1	3,3	4,4	3,0	4,0	5,0	2,7	2,6	4,8
4,7	2,8	4,8	3,9	2,3	3,8	2,8	3,0	3,7
3,3	4,4	3,1	4,0	3,7	2,5	2,7	5,1	4,7

Aldagai kuantitatibo jarraitu hau [2,1, 5,1] tartean definituta dagoela jakinda, kalkula bitez batez bestekoa, mediana, moda, desbideratze tipikoa, aldakuntza-koefizientea, Q_1 eta Q_3 kuartilak, heina, kuartilarteko heina, alborapena eta kurtosia.

Ebazpena

Egin dezagun azterketa aldagai kuantitatiboa era jarraituan tratatuz. Heina $R = 5{,}1 - 2{,}1 = 3$ denez eta klase-kopurua $k = \sqrt{n} = \sqrt{27} \approx 5$ har daitekeenez, datuak $\frac{R}{k} = \frac{3}{5} = 0{,}6$ anplitudeko 5 klasetan banatuko dira. Argudio hauen arabera, hona hemen aldagai kuantitatibo jarraituari dagokion maiztasun-taula:

$[l_i, l_{i+1})$	x_i	f_i	F_i	h_i	H_i
[2,1, 2,7)	2,4	4	4	0,15	0,15
[2,7, 3,3)	3,0	7	11	0,26	0,41
[3,3, 3,9)	3,6	5	16	0,18	0,59
[3,9, 4,5)	4,2	5	21	0,18	0,78
[4,5, 5,1]	4,8	6	27	0,22	1

Batez bestekoa:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{n} = \frac{2{,}4 \cdot 4 + 3{,}0 \cdot 7 + 3{,}6 \cdot 5 + 4{,}2 \cdot 5 + 4{,}8 \cdot 6}{27} = \frac{98{,}4}{27} = 3{,}64 \text{ litro.}$$

Mediana: $F_3 = 16 > \frac{27}{2} = 13{,}5$ denez, orduan $M_3 \in [3{,}3, 3{,}9)$ dago.

$$M_e = l_i + \frac{\frac{n}{2} - F_{i-1}}{f_i} \cdot d_i = 3{,}3 + \frac{13{,}5 - 11}{5} \cdot 0{,}6 = 3{,}6 \text{ litro.}$$

Moda:

Maiztasun handiena $f_2 = 7$ da. Orduan, moda [2,7, 3,3) tartean dago.

$$M_o = l_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot d_i = 2,7 + \frac{\frac{7-4}{0,6}}{\left(\frac{7-4}{0,6}\right) + \left(\frac{7-5}{0,6}\right)} \cdot 0,6 = 3,06 \text{ l.}$$

Desbideratze tipikoa:

Desbideratze tipikoa bariantzaren erro karratu positiboa denez, lehenengoz bariantzaren balioa lortuko da.

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

$$= \frac{(2,4)^2 \cdot 4 + 3^2 \cdot 7 + (3,6)^2 \cdot 5 + (4,2)^2 \cdot 5 + (4,8)^2 \cdot 6}{40} - (3,64)^2 = 0,72 \text{ (litro)}^2.$$

Beraz, desbideratze tipikoaren balioa hurrengoa da:

$$s = \sqrt{0,72} = 0,85 \text{ litro.}$$

Aldakuntza-koefizientea:

$$CV = \frac{s_x}{\bar{x}} = \frac{0,85}{3,64} = 0,23.$$

Q_1 eta Q_3 kuartilen kalkulua:

$$F_2 = 11 > \frac{25 \cdot n}{100} = \frac{25 \cdot 27}{100} = 6,75 \text{ denez, orduan } Q_1 \in [2,7, 3,3) \text{ dago.}$$

$$Q_1 = l_i + \frac{\frac{25 \cdot n}{100} - F_{i-1}}{f_i} \cdot d_i = 2,7 + \frac{6,75 - 4}{7} \cdot 0,6 = 2,94 \text{ litro.}$$

$$F_4 = 21 > \frac{75 \cdot n}{100} = \frac{75 \cdot 27}{100} = 20,25 \text{ denez, orduan } Q_3 \in [3,9, 4,5) \text{ dago.}$$

$$Q_3 = l_i + \frac{\frac{75 \cdot n}{100} - F_{i-1}}{f_i} \cdot d_i = 3,9 + \frac{20,25 - 16}{5} \cdot 0,6 = 4,41 \text{ litro.}$$

Heina:

$$R = max(x_i) - min(x_i) = 5,1 - 2,1 = 3.$$

Kuartilarteko heina:

$$Q_3 - Q_1 = 4{,}41 - 2{,}94 = 1{,}47.$$

Alborapen-koefizientea:

$$v = \frac{\bar{x} - M_o}{s_X} = \frac{3{,}64 - 3{,}06}{0{,}85} = 0{,}68.$$

Koefiziente horren balioa positiboa denez, banaketa eskuinerantz alboratua da.

Kurtosia:

$$g_2 = \frac{m_4}{m_2^2}.$$

Lehenengo, laugarren ordenako momentuaren balioa lortuko da.

$$m_4 = \frac{\sum_{i=1}^{k} \left(x_i - \bar{x}\right)^4 f_i}{n} = \frac{\left(2{,}4 - 3{,}64\right)^4 \cdot 4 + \left(3 - 3{,}64\right)^4 \cdot 7 + \ldots + \left(4{,}8 - 3{,}64\right)^4 \cdot 6}{27} =$$

$$= \frac{21{,}97}{27} = 0{,}81 \text{ (litro)}^4.$$

Bestalde, bariantzaren balioa $m_2 = 0{,}72$ (litro)^2 denez, orduan $m_2^2 = 0{,}52$ (litro)^4 da.

Ondorioz, kurtosiaren balioa lortuko da:

$$g_2 = \frac{0{,}81}{0{,}52} = 1{,}56.$$

Kurtosi-koefizientearen balioa 3 baino txikiagoa denez, banaketa platikurtikoa da.

1.6. ARIKETA PROPOSATUAK

1.6.1. ariketa

Airearen kutsatzaile batzuen kontzentrazioen arabera, 1, 2, 3, 4 eta 5 indizeak adierazten dira, indize horiek airearen kalitatea hurrengo eran definitzen dutelarik:

Indizea	Airearen kalitatea
1	Oso ona
2	Ona
3	Egokia
4	Txarra
5	Oso txarra

Hurrengo zenbakiek hemezortzi egunetako airearen kalitatea adierazten dute:

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

a. Nolakoa da aldagaia?

b. Adieraz bitez barra-grafikoa eta sektore-diagrama.

c. Kalkula bitez batez besteko indizea eta mediana. Azaldu esanahia.

1.6.2. ariketa

X aldagai estatistikoari hurrengo taula dagokio:

x_i	f_i	F_i	h_i
10		2	
12		5	0,12
15	3		
20		13	
35	6		0,24
38			0,24

a. Osa bedi taula.

b. Eraiki bitez barra-grafikoa eta maiztasun metatuen diagrama.

b. Kalkula bitez mediana eta moda.

1.6.3. ariketa

Aleazio forjatuko 40 barra mozteko hurrengo bihurdurak behar izan dira:

33	21	32	44	35	22	40	36	22	37
20	37	42	31	23	44	32	30	44	44
42	35	40	36	32	31	37	43	24	40
25	30	26	35	33	41	25	44	36	27

a. Bil itzazu balio hauek anplitude bereko sei klasetan eta osa ezazu maiztasun-taula.

b. Adieraz bitez histograma eta maiztasun metatuen poligonoa.

c. Konpara bitez heina eta kuartilarteko heina.

1.6.4. ariketa

Berrogei mekanikariren gaitasuna neurtzeko, proba berezi bat burutu da. Proba horretako puntuazioak hurrengoak izan dira:

Aptitudea	37	37’20	37’50	38	38’10	38’50
Mekanikari-kopurua	1	5	13	6	10	5

a. Kalkula bitez batez bestekoa, desbideratze tipikoa, mediana eta moda.

b. Lor bitez 65. eta 90. ordenako pertzentilak. Zer adierazten dute kasu honetan?

1.6.5. ariketa

Hona hemen makina-multzo batek egun batean kontsumitzen duen erregai-kantitatea:

Erregai-kantitatea (litro)	Makina-kopurua
[10, 20)	8
[20, 30)	12
[30, 40)	10
[40,50)	22
[50,60)	13

a. Irudika bitez histograma eta maiztasun metatuen poligonoa.

b. Kalkula bitez heina, kuartilarteko heina, desbideratze tipikoa, eta aldakuntza-koefizientea.

c. Zein bi pertzentil hurbilen artean dago 42,5 litroko erregai-kantitatea?

1.6.6. ariketa

Berrogei hiritako urteko batez besteko tenperaturak (ºC) hauexek dira:

Tenperaturak	10	14	17	20	25
Hiri-kopurua	5	10	15	7	3

a. Kalkula bedi H_4 maiztasun erlatibo metatua eta adierazi beraren esanahia kasu honetan.

b. Kalkula bitez batez besteko tenperatura, mediana, moda eta desbideratze tipikoa.

c. Zein da 65. ordenako pertzentilaren balioa? Zer esan nahi du?

d. Beste hogeita bost hiri hartu dira, batez besteko tenperatura 18 ºC eta desbideratze tipikoa 4,2 ºC izanik. Zein hirik dituzte tenperatura erlatiboki kontzentratuagoak, lehenengo berrogeiek edo oraingo hogeita bostek?

1.6.7. ariketa

Ondoko taulan hogei enpresaren urteko irabaziak eta dagozkien maiztasun erlatiboak azaltzen dira:

Irabaziak (milaka euro)	Enpresak (%)
[65, 81)	10
[81, 85)	45
[85, 89)	20
[89, 97)	20
[97, 107)	5

a. Irudika bitez histograma eta maiztasun metatuen poligonoa.

b. 65 mila euroko irabazia minimoa izanda, zein da enpresen erdiek irabazitako kantitate maximoa urte horretan? Lortutako emaitza adieraz bedi histograman.

c. Enpresak errentagarria izateko urtean gutxienez 95 mila euro irabazi behar baditu, hauetariko zenbat dira errentagarriak? Adieraz bedi lortutako emaitza maiztasun metatuen poligonoan.

1.6.8. ariketa

Herrialde batean etxebizitzaren batez besteko alokairua 600 euro eta desbideratze tipikoa 20 euro dira. Beste herrialde batean, etxebizitzaren batez besteko alokairua 400 euro eta desbideratze tipikoa 40 euro dira. Lehenengo herrialdeko etxe bat 520 eurotan alokatu da eta bigarren herrialdeko beste etxe bat 335 eurotan alokatu da. Zein etxek du erlatiboki prezio baxuagoa?

1.6.9. ariketa

Hurrengo taulan industrien sufre dioxidoaren emisioak daude:

Sufre dioxidoa (tona)	Industria (%)
[6, 14)	% 15
[14, 22)	% 25
[22, 30)	% 40
[30, 42)	% 20

a. Industrien zein portzentajek emititzen ditu gutxienez 18 tona sufre dioxido?

b. Jakinda emisio minimoa 6 tonakoa dela, zein da industrien % 75ek emititzen duen sufre dioxidorako goi-muga?

1.6.10. ariketa

Neurgailu batek hezetasun erlatiboko hurrengo datuak jaso ditu:

29	12	34	17	26	28	23	28	29	34
17	23	12	26	34	29	12	34	28	17

a. Eraiki bedi aldagai estatistiko kuantitatibo diskretu honi dagokion maiztasun-taula.

b. Kalkula bitez batez bestekoa, mediana, moda, desbideratze tipikoa, aldakuntza-koefizientea, Q_1 eta Q_3 kuartilak, heina, kuartilarteko heina, alborapena eta kurtosia.