ESTATISTIKA METODOAK INGENIARITZAN

1. Estatistika deskribatzailea

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

1. Estatistika deskribatzailea

eman ta zabal zazu

Universidad del País Vasco Euskal Herriko Unibertsitatea

1.1 Sarrera

Estatistika multzo bati dagozkion zenbakizko datuak biltzen, sailkatzen eta aztertzen dituen zientzia da. Bi arlo daude:

1. **Estatistika deskribatzailea**: Aztergai den multzoari dagozkion datuak bildu, antolatu eta egoera deskribatzeko ezaugarriak lortuko dira.

2. **Estatistika induktiboa**: Emaitzak orokortu, ondorioak atera edo aurresanak egin daitezke.

Sarrera

Populazioa eta Lagina

Aldagai Estatistikoak

Maiztasun taulak

Adierazpen grafikoa

Parametroak eta estatistikoak

Estatistiko deskribatzaileak

Eskala eta jatorri aldaketak

Aldagai tipifikatuak

Lege enpirikoa

Kutxa diagrama eta balio arraroak

1.2 Populazioa eta lagina

Populazioa: Azterketa estatistikoa multzo batean gauzatuko da, multzo oso horri populazio deritzo. (Aztertu nahi den multzo osoa)

Lagina: Populazioaren azpimultzo bat da.

Oharrak:

(1) Populazio bat ondo definituta egoteko argi izan behar da zein elementu populaziokoa den eta zein ez.

(2) Lagina populazioari buruzko informazioa lortzeko erabiltzen dugunez, populazioaren azpimultzo adierazgarri bat izan behar da.

Adibidea 1

Zergatik definitzen dugu lagina??

1.3 Aldagai estatistikoak

Sarrera
Populazioa eta Lagina
Aldagai Estatistikoak
Maiztasun taulak
Adierazpen grafikoa
Parametroak eta estatistikoak
Estatistiko deskribatzaileak
Eskala eta jatorri aldaketak
Aldagai tipifikatuak
Lege enpirikoa
Kutxa diagrama eta balio arraroak

Aztergaia: Populazioan aztertu nahi dugun ezaugarria.

Aldagai estatistikoa: Azterketa estatistikoan aztergaiak har ditzakeen balioek determinatzen dute.

Bi mota: Aldagai kuantitatiboak eta aldagai kualitatiboak.

Aldagai kuantitatiboak: Aldagaiak hartzen dituen balioak zenbakiak dira.

- Aldagai kuantitatibo diskretuak: Balio isolatuak hartzen ditu.
- Aldagai kuantitatibo jarraituak: Tarte bateko edozein balio har dezake.

Aldagai kualitatiboa: Aldagaiak hartzen dituen balioak zenbakizkoak ez badira. (Adb: Bai/Ez, Altua/Baxua,...)

1.3 Aldagai estatistikoak

Adibidea 2

1) Lantegi bateko kategori profesionala.
2) CD-ak egiten dituen lantegi batean, 50 CDko lagina erabiliz, akastunak diren CD kopurua.
3) CD-ak egiten dituen lantegi batean, 50 CDko lagina erabiliz, CD-en diametroa.

Oharra: Azterketa estatistikoa guztiz ezberdina izango da aztertu nahi den ezaugarriaren arabera.

1.4. Maiztasun taulak

Taldekatu gabeko datuak:

Modalitateak (x_i): Aldagaiak hartzen dituen balioak. Zenbakiak direnean txikienetik handienera ordenatzen dira.

Maiztasun absolutuak (f_i): x_i balioaren maiztasun absolutua, x_i balioa zenbat aldiz jaso den.

Maiztasun metatuak (F_i): $F_i = f_1+f_2+ \ldots + f_i$, maiztasun absolutuen batura.

Maiztasun erlatiboa (h_i): $h_i = \dfrac{f_i}{n}$, n laginaren tamaina izanik.

Maiztasun erlatibo metatua (H_i): $H_i = \dfrac{F_i}{n}$, n laginaren tamaina izanik.

1.4. Maiztasun taulak

Modalitateak x_i	Maiztasun absolutuak f_i	Maiztasun metatuak $F_i = f_1 + f_2 + \ldots + f_i$	Maiztasun erlatiboa $h_i = \frac{f_i}{n}$	Maiztasun erlatibo metatua $H_i = \frac{F_i}{n}$
x_1	f_1	F_1	h_1	H_1
x_2	f_2	F_2	h_2	H_2
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
x_i	f_i	F_i	h_i	H_i
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
x_m	f_m	$F_m = n$	h_m	$H_m = 1$
GUZTIRA	n		1	

1.4. Maiztasun taulak

Taldekatutako datuak:

Besterik esan ezean, datuak taldekatzeko maiztasun elkartuen metodoa erabiltzen da. Metodo honen pausuak:

1. Aldagai estatistikoaren **heina kalkulatu**: Balio handienari txikiena kendu.
2. Tartea **k azpitarteetan** edo k klaseetan **banatu**: $k = \sqrt{n}$, n laginaren tamaina izanik.
3. Klaseak eta klase-markak zehaztu:
   Klaseak: $[l_i, l_{i+1})$
   Klase-marka: $x_i = \dfrac{l_i + l_{i+1}}{2}$

Universidad del País Vasco Euskal Herriko Unibertsitatea

1.4. Maiztasun taulak

Klaseak $[l_i, l_{i+1})$	Klase-marka $x_i = \frac{l_i + l_{i+1}}{2}$	Maiztasun absolutuak f_i	Maiztasun metatuak $F_i = f_1 + f_2 + \ldots + f_i$	Maiztasun erlatiboa $h_i = \frac{f_i}{n}$	Maiztasun erlatibo metatua $H_i = \frac{F_i}{n}$
$[l_1, l_2)$	x_1	f_1	F_1	h_1	H_1
$[l_2, l_3)$	x_2	f_2	F_2	h_2	H_2
.	.	.	.	.	.
.	.	.	.	.	.
	.	.	.	.	.
$[l_i, l_{i+1})$	x_i	f_i	F_i	h_i	H_i
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
$[l_k, l_{k+1})$	x_k	f_k	$F_k = n$	h_k	$H_k = 1$
GUZTIRA		n		1	

1.4. Maiztasun taulak

Oharrak:

(1) Tarteen zabalera berdina edo desberdina izan daiteke.

Tarteen zabalera berdina izatea nahi badugu baliteke lehen eta azken tarteak zertxobait zabaldu behar izatea.

(2) Tarteen muturrak zein tarteetan dauden argi gelditzeko tarte erdi irekiak egingo ditugu.

1.5. Adierazpen grafikoak

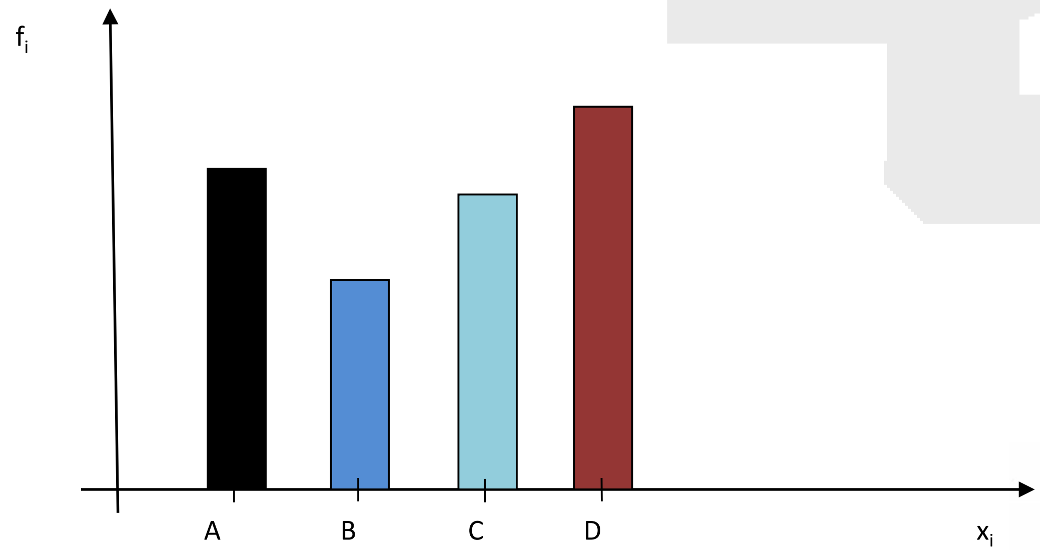
1.5.1 Aldagai kualitatiboa:

Barra diagrama eta **sektore diagrama**.

a) Barra diagrama:

Abzisa-ardatza: Aldagaiaren balioak

Ordenatu-ardatza: Maiztasun absolutuak

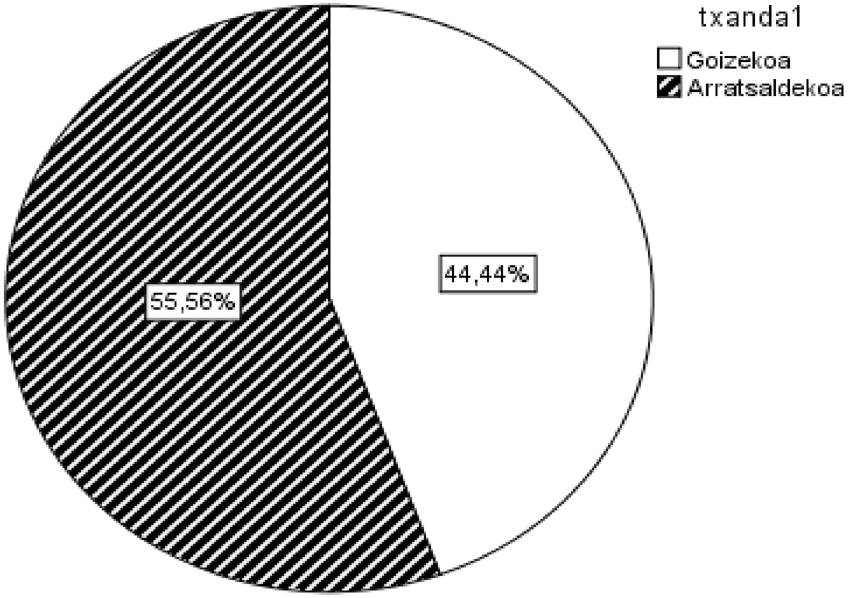


1.5. Adierazpen grafikoak

1.5.1 Aldagai kualitatiboa:

b) Sektore diagrama:

Sektore bakoitza adierazteko $\alpha_i = 360^o \dfrac{f_i}{n}$



Universidad del País Vasco
Euskal Herriko Unibertsitatea

1.5. Adierazpen grafikoak

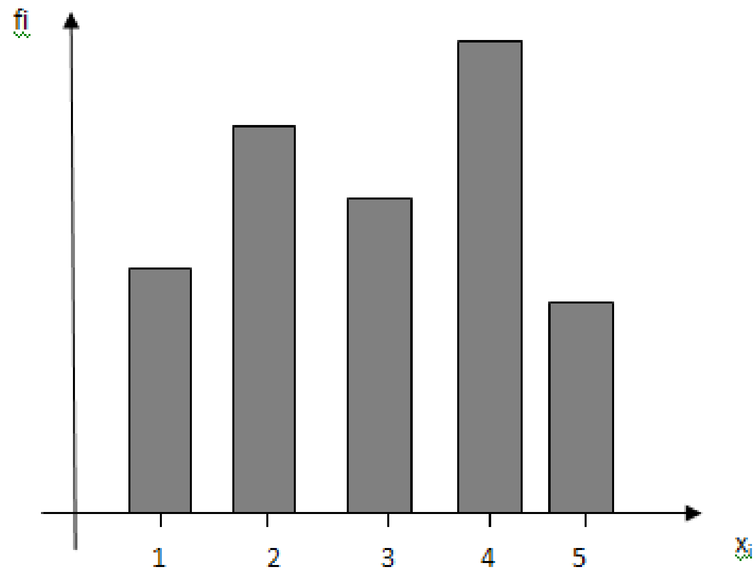
1.5.2 Aldagai kuantitatibo diskretuak:

Barra grafikoa, maiztasun metatuen grafikoa, zurtoin eta hosto grafikoa.

a) Barra grafikoa:

Abzisa-ardatza: Aldagaiaren balioak

Ordenatu-ardatza: Maiztasun absolutuak



Universidad del País Vasco Euskal Herriko Unibertsitatea

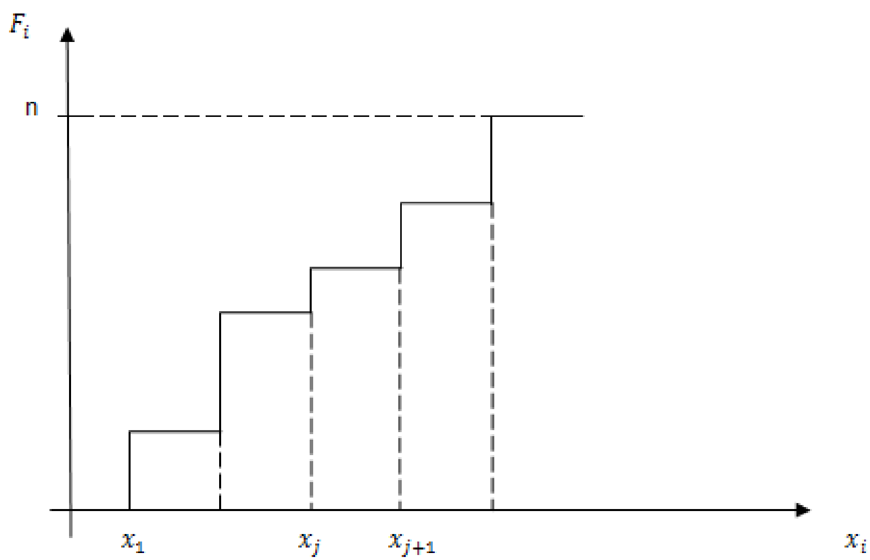
1.5. Adierazpen grafikoak

1.5.2 Aldagai kuantitatibo diskretuak:

b) Maiztasun metatuen grafikoa:

Abzisa-ardatza: Aldagaiaren balioak

Ordenatu-ardatza: Maiztasun metatuak



Oharra:

$$\left.\begin{array}{l} x_j \rightarrow F_j \\ x_{j+1} \rightarrow F_{j+1} \end{array}\right\} \Rightarrow \left(x_j,\, x_{j+1}\right) \rightarrow F_j \quad \text{altuera}$$

1.5. Adierazpen grafikoak

Adibidea 3

Ondoko balioak 30 torlojuen lodierak (milimetroetan) 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) Maiztasun taula eraiki

b) Barra-grafikoa irudikatu

c) Adierazi ezazu “oso mehea” = 1, “mehea”=2, “ertaina”=3, “lodia”=4 eta “oso lodia”=5 lodierei dagokien sektore-diagrama

d) Maiztasun metatuen grafikoa irudikatu

1.5. Adierazpen grafikoak

1.5.2 Aldagai kuantitatibo diskretuak:

c) Zurtoin eta hosto grafikoa:

Grafiko eta taula baten arteko nahasketa da. Pausoak:

1. Datuak txikienetik handienera ordenatu.
2. Datu bakoitza bi osagaitan banatu: Hosto eta zurtoin osagaiak. Orokorrean:
   a) Hosto osagaia: Azken zifra
   b) Zurtoin osagaia: Gainontzeko zatia
3. Taulan hosto osagaia eskuinaldean eta zurtoin osagaia ezkerraldean idazten dira.

1.5. Adierazpen grafikoak

Adibidea 4

Izan bitez hurrengo datuak

58	62	75	67	79	71	65	72	70
78	81	89	91	61	100	95	85	89

Egin ezazu zurtoin eta hosto grafikoa.

Universidad del País Vasco Euskal Herriko Unibertsitatea

1.5. Adierazpen grafikoak

1.5.2 Aldagai kuantitatibo jarraituak:

Histograma, maiztasun absolutuen poligonoa, maiztasun metatuen histograma eta **maiztasun metatuen poligonoa.**

a) Histograma

Abzisa-ardatza: Klaseak

Ordenatu-ardatza:

i. Klaseak zabalera berekoak badira: $f_{i,}$ (maiztasun absolutua)

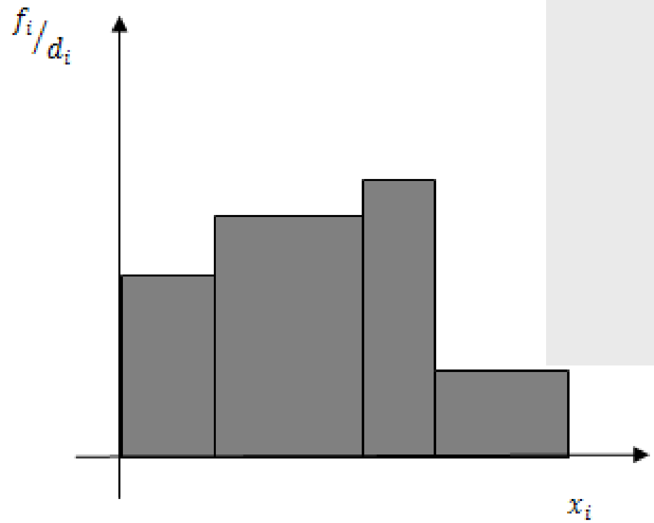
ii. Klaseak zabalera berekoak EZ badira: f_i / d_i

$$\frac{f_i}{d_i} = \frac{\text{maiztasun absolutua}}{\text{klaseen zabalera}}$$

1.5. Adierazpen grafikoak

1.5.2 Aldagai kuantitatibo jarraituak:

a) Histograma:



b) Maiztasun absolutuen poligonoa:

Histograman errektangeluetako goiko aldeetako erdiko puntuak lotuz eraikitzen da.

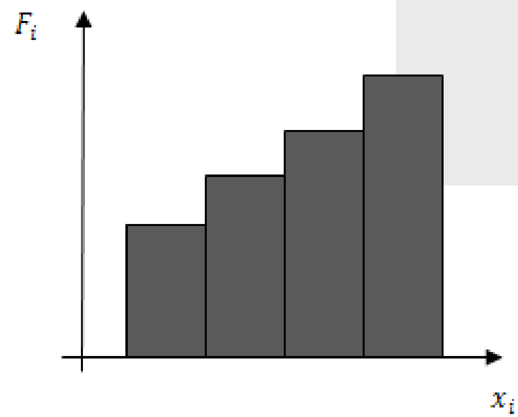
1.5. Adierazpen grafikoak

1.5.2 Aldagai kuantitatibo jarraituak:.

c) Maiztasun metatuen histograma

Abzisa-ardatza: Klaseak

Ordenatu-ardatza: Maiztasun metatuak



d) Maiztasun metatuen poligonoa:

Maiztasun metatuen histograman (l_{i+1}, F_i) puntuak lotuz eraikitzen da.

1.5. Adierazpen grafikoak

Adibidea 5

Hona hemen marka ezagun bateko 27 autoren gasolina-kontsumoa (litrotan) 100km-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

a) Maiztasun taula eraiki

b) Histograma eta maiztasun absolutuen poligonoa irudikatu

c) Maiztasun metatuen histograma eta maiztasun metatuen poligonoa irudikatu

1.6 Parametroak eta estatistikoak

- **Parametroa:** Azterketa estatistikoan **populazio osoa** erabiltzean aztergaiak hartzen duen balioa.

- **Estatistikoa:** Azterketa estatistikoan **lagina** erabiltzean aztergaiak hartzen duen balioa.

Oharra:

Ikerketa gehienetan populazioaren ezaugarriak interesatzen zaizkigu, hau da, parametro bat ezagutu nahi dugu. Baina populazio osoa aztertzea oso zaila denez, lagina erabiliz estatistikoak lortuko ditugu.

Universidad del País Vasco Euskal Herriko Unibertsitatea

1.7 Estatistiko deskribatzaileak

Estatistiko deskribatzaileak aztergai den fenomenoa deskribatzeko erabiltzen dira. Lau mota desberdinetako neurriak ikusiko ditugu:

1.7.1 Joera zentraleko neurriak:

Datuak zein balioren inguruan banatzen diren adierazten digute, hau da, laginaren balio zentrala aurkitzeko erabiltzen dira. Hauen artean: **Moda, mediana, batezbesteko aritmetikoa, batezbesteko geometrikoa, batezbesteko armonikoa eta batezbesteko koadratikoa.**

Universidad del País Vasco Euskal Herriko Unibertsitatea

1.7 Estatistiko deskribatzaileak

1.7.2 Sakabanaketa neurriak:

Lagineko elementuen balioek joera zentraleko neurriekiko duten sakabanaketa neurtzen dute: **Heina edo anplitudea, kuartilarteko heina, bariantza, desbiderazio tipikoa, aldakuntza-koefizientea.**

1.7.3 Posiziozko neurriak:

Orokorrean, aldagaiaren balioak ordenatu ondoren, lagina datu kopuru berdineko zati desberdinetan banatzen duten balioak dira: **koantilak, pertzentilak, dezilak, kuartilak.**

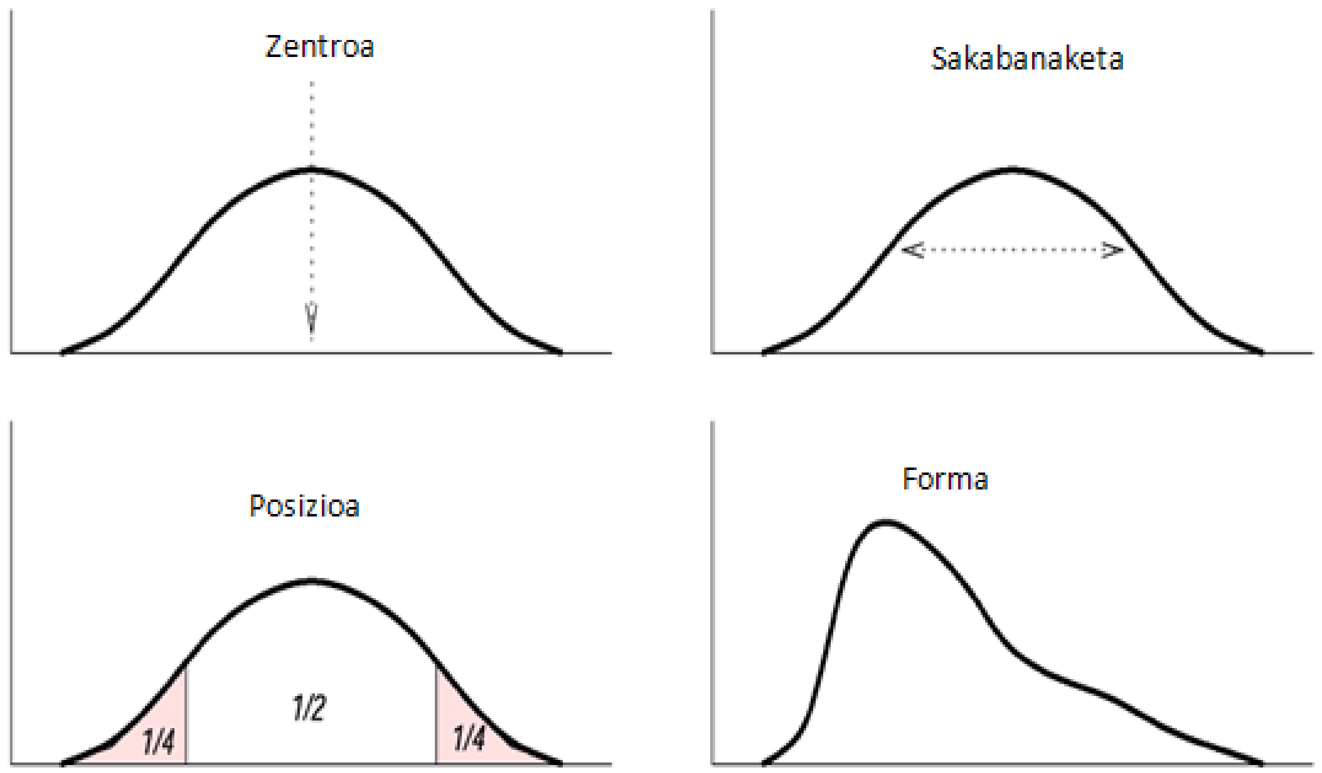
1.7.4 Formako neurriak:

Asimetria edo alborapen neurriak, kurtosia

eman ta zabal zazu

Universidad del País Vasco Euskal Herriko Unibertsitatea

1.7 Estatistiko deskribatzaileak



Helburua: Datu multzo bati dagokion informazioa ezaugarri gutxi erabiliz deskribatzea.

Universidad del País Vasco   Euskal Herriko Unibertsitatea

1.7 Estatistiko deskribatzaileak

Atal honetan erabiliko dugun notazioa:

n : Laginaren tamaina

l_i : Estatistikoa barnean daukan klasearen behe-muturra

d_i : Estatistikoa barnean daukan klasearen zabalera

f_i : Estatistikoa barnean daukan klasearen maiztasun absolutua

f_{i-1} : Estatistikoa barnean daukan klasearen aurreko klasearen maiztasun absolutua

f_{i+1} : Estatistikoa barnean daukan klasearen hurrengo klasearen maiztasun absolutua

F_{i-1} : Estatistikoa barnean daukan klasearen aurreko klasearen maiztasun metatua

Universidad del País Vasco Euskal Herriko Unibertsitatea

1.7.1 Joera zentraleko neurriak

Moda (Mo): Maiztasun absolutu handiena duen balioa da.

Kasu kualitatiboa: f_i handiena duen x_i

Kasu diskretuan: f_i handiena duen x_i

Kasu jarraituetan:

1) Moda barnean duen $[l_i, l_{i+1})$ tartea $\frac{f_i}{d_i}$ handiena duen klasea da.

2) $$Mo = l_i + \frac{\Delta_1}{\Delta_1 + \Delta_2}.d_i \; : \quad \Delta_1 = \frac{f_i}{d_i} - \frac{f_{i-1}}{d_{i-1}}; \quad \Delta_2 = \frac{f_i}{d_i} - \frac{f_{i+1}}{d_{i+1}}$$

Oharrak:

1) Banaketa batek moda bat baino gehiago izan ditzake, kasu horretan, banaketa bimodala, ..., multimodala, dela diogu.

2) Balio guztiak maiztasun berdina badute modarik ez dagoela esaten da.

1.7.1 Joera zentraleko neurriak

Mediana (Me): Aldagaiaren balioak ordenatuta daudenean, bere ezkerrean eta eskuinean gai kopuru berdina uzten duen balioa da.

Kasu diskretuan: Definizioa aplikatzea da:

$$F_{i-1} < \frac{n}{2} \text{ eta } F_i > \frac{n}{2} \text{ betetzen duen } x_i \text{ balioa.}$$

$$F_i = \frac{n}{2} \text{ betetzen bada, orduan } Me = \frac{x_i + x_{i+1}}{2}$$

Kasu jarraituetan:

1) Mediana barnean duen $\left[l_i, l_{1+1}\right)$ tartea:

$$F_{i-1} < \frac{n}{2} \text{ eta } F_i > \frac{n}{2} \text{ betetzen duen tartea.}$$

2) Formula aplikatu: $Me = l_i + \dfrac{n/2 - F_{i-1}}{f_i}.d_i$

Universidad del País Vasco   Euskal Herriko Unibertsitatea

1.7.1 Joera zentraleko neurriak

Batezbesteko aritmetikoa ($\overline{x}$):

Gehien erabiltzen den estatistikoa da. Datuen grabitate-zentroa dela esaten da.

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

k modalitate kopurua izanik.

Oharra:

Lagineko datuak taldekatuta daudenean klase-marka erabiltzen da.

1.7.1 Joera zentraleko neurriak

Batezbesteko geometrikoa $\left(\overline{x}_G\right)$

$$\overline{x}_G = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[n]{\prod_{i=1}^{k} x_i^{\,f_i}}$$

Batezbesteko armonikoa $\left(\overline{x}_H\right)$

$$\overline{x}_H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{n}{\sum_{i=1}^{k} \frac{f_i}{x_i}} = \left[\sum_{i=1}^{k} \frac{f_i}{x_i}\right]^{-1}$$

Batezbesteko koadratikoa $\left(\overline{x}_Q\right)$

$$\overline{x}_Q = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{1}{n}\sum_{i=1}^{k} f_i x_i^2$$

Universidad del País Vasco Euskal Herriko Unibertsitatea

1.7.1 Joera zentraleko neurriak

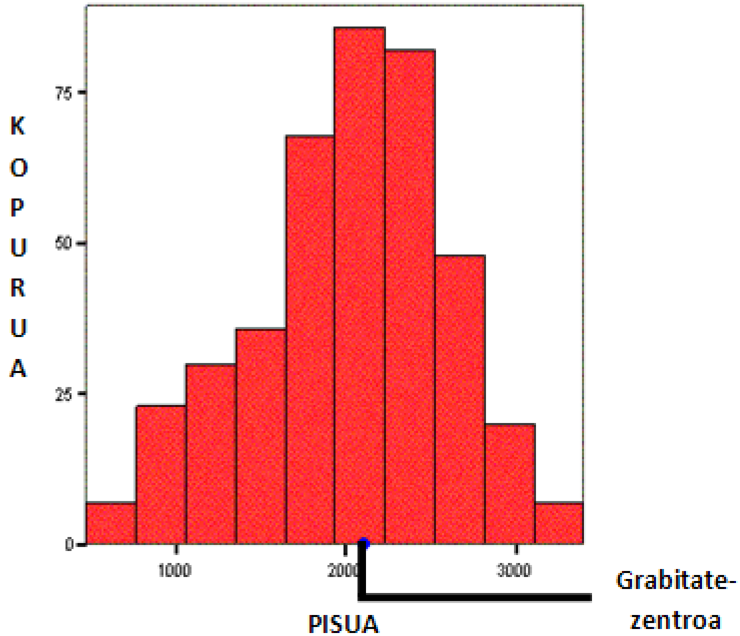
Batezbestekoa:

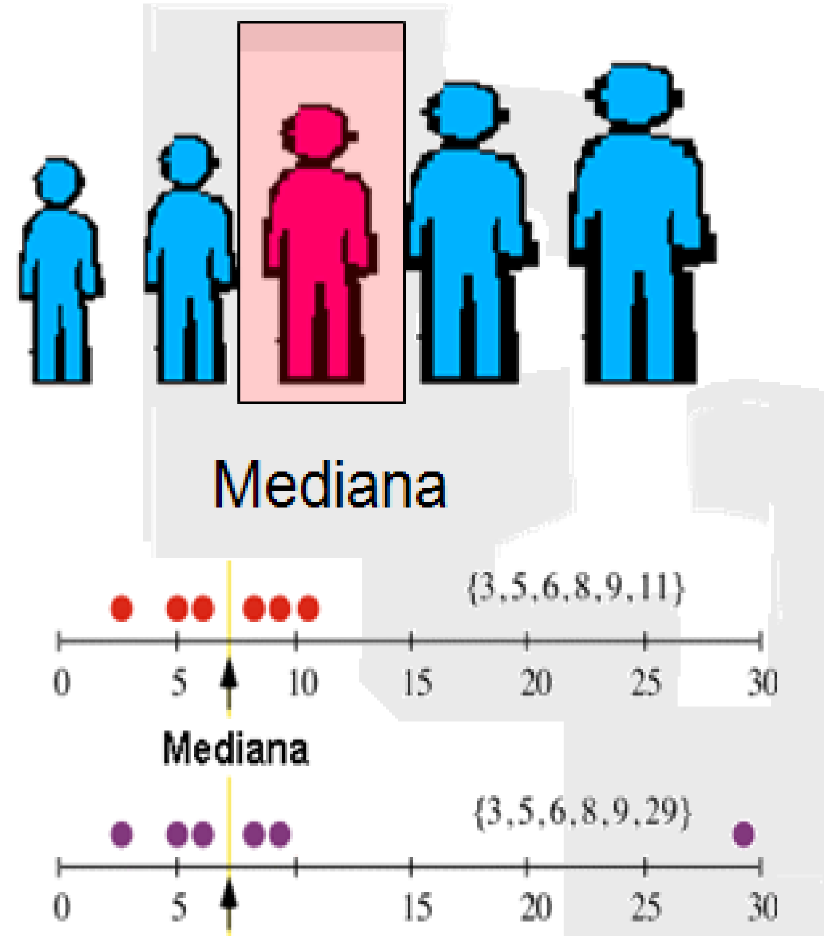
- 2,2,3,7 zenbakien batezbestekoa (2+2+3+7)/4=3,5
- Datuak berarekiko simetrikoki kontzentratzen direnean oso erabilgarria da.
- Balio arraroekiko/muturreko balioekiko sentikorra da (balio arraroek eragina dute batezbestekoan)
- Datuen grabitate-zentroa da.

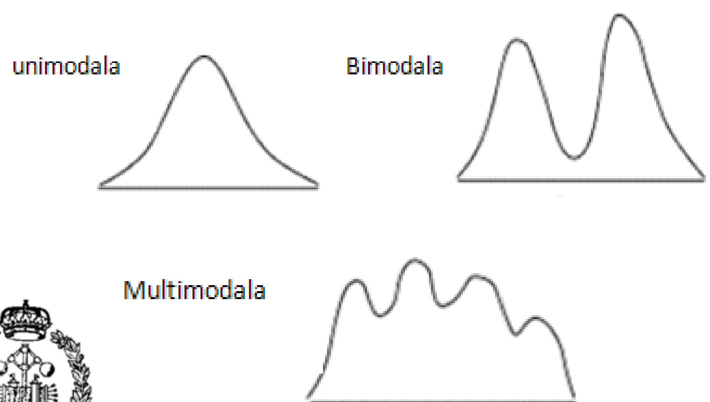
Mediana:

- 1,2,4,**5**,6,6,8 datuen mediana 5 da
- 1,2,4,**5**,**6**,6,8,9 datuen mediana (5+6)/2=5,5 da
- Datuak alboratuak direnean erabilgarria da. Ez da balio arraroekiko/muturreko balioekiko sentikorra.
- 1,2,4,**5**,6,6,800 datuen mediana 5 da, batezbestekoa berriz (1+2+4+5+6+6+800)/7=117,7

1.7.1 Joera zentraleko neurriak







Universidad del País Vasco
Euskal Herriko Unibertsitatea

1.7.1 Joera zentraleko neurriak

Adibidea 6

Pisua	Klase marka	f_i	F_i
[40, 50)	45	5	5
[50, 60)	55	10	15
[60, 70)	65	21	36
[70, 80)	75	11	47
[80, 90)	85	5	52
[90, 100)	95	3	55
[100,130)	115	3	58
		58	

- **Estatistikoak definitzeko tarteen adierazgarria den puntu bat aukeratu behar da, klase marka, hain zuzen ere.**
- **Muturreko balioek batezbestekoa eskuinerantz mugitzen dute. Ondorioz, batezbestekoa eta mediana ez datoz bat.**

Universidad del País Vasco Euskal Herriko Unibertsitatea

1.7.1 Joera zentraleko neurriak

Pisua	Klase marka	f_i	F_i
[40, 50)	45	5	5
[50, 60)	55	10	15
[60, 70)	65	21	36
[70, 80)	75	11	47
[80, 90)	85	5	52
[90, 100)	95	3	55
[100,130)	115	3	58
		58	

$$\overline{x} = \frac{\sum_{i=1}^{k} f_i x_i}{n} = \frac{45\cdot 5+55\cdot 10+\ldots+115\cdot 3}{58} = 69,3$$

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

$$= 60 + \frac{58/2 - 15}{21} 10 = 60 + \frac{29-15}{21} 10 = 66,\widehat{6}$$

$$Mo = l_i + \frac{\Delta_1}{\Delta_1+\Delta_2} .d_i ; \quad \Delta_1 = \frac{f_i}{d_i} - \frac{f_{i-1}}{d_{i-1}} ; \quad \Delta_2 = \frac{f_i}{d_i} - \frac{f_{i+1}}{d_{i+1}}$$

$$Mo = 60 + \frac{21/10 - 10/10}{\left(21/10 - 10/10\right) + \left(21/10 - 11/10\right)} 10 = 65$$

1.7.2 Sakabanaketa neurriak

Heina edo anplitudea: Laginaren balio handienaren eta txikienaren arteko kenketa da.

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

Kuartiletako heina: $RIC = Q_3 - Q_1$

(Q_3, Q_1 beranduago ikusiko dira)

Bariantza: Batezbestekoarekiko errore karratuen batezbestekoa da.

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

1.7.2 Sakabanaketa neurriak

Desbiderazio tipikoa: Bariantzaren erro karratua

$$s(x)=\sqrt{s^2(x)}$$

Aldakuntza koefizientea: $CV_x=\dfrac{s_x}{\overline{x}}$

Aldakuntza erlatiboa bezala ere ezagutzen da.

- Aldagai bat bi multzotan edo bi aldagai ezberdin konparatzeko balio du.
- Koefiziente hau erabiliz, batezbestekoarekiko duten sakabanaketa konpara dezakegu.
- Aldakuntza-koefiziente handiena duen aldagaiak sakabanaketa handiena du, ondorioz, aldagai honen batezbestekoa ez da bestearena bezain adierazgarria.

Universidad del País Vasco Euskal Herriko Unibertsitatea

1.7.2 Sakabanaketa neurriak

Heina edo anplitudea:

Muturreko balioen arteko diferentzia: $R = \max(x_i) - \min(x_i)$

- 2, 1, 4, 3, 8, 4. Heina 8 − 1 = 7 da

Balio arraroekiko/muturreko balioekiko oso sentikorra.

Bariantza:

Datuak batezbestekotik “zenbat” sakabanatzen diren neurtzen du

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

Balio arraroekiko/muturreko balioekiko oso sentikorra.

Aldakuntza koefizientea:

- Askotan ehunetan adierazten da: Batezbestekoa 80 bada eta desbiderazio tipikoa 20,orduan CV=20/80=0,25=%25
- Pisuak CV=%30 badu eta altuerak CV=%10, orduan pisua aldagaiak sakabanaketa handiagoa du.
- Aldagaiak balio negatiboak hartzen baditu ez da erabili behar.

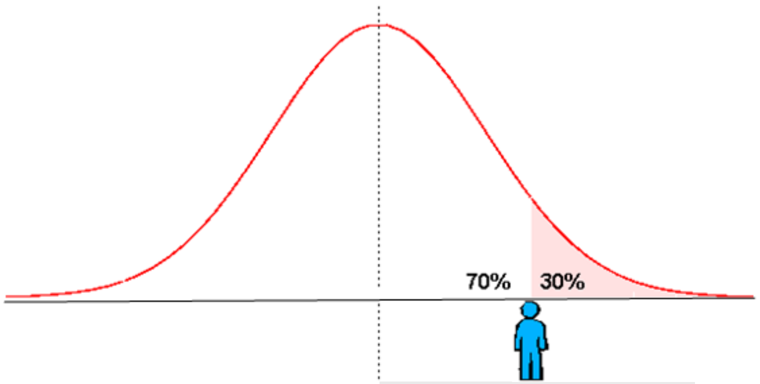
$$CV_x = \frac{S_x}{\overline{X}}$$

Universidad del País Vasco Euskal Herriko Unibertsitatea

1.7.3 Posiziozko neurriak

Koantilak:

Orokorrean, α ordenako koantila, banaketaren α balioa ezkerrean uzten du



$$C_\alpha = l_i + \frac{\alpha n - F_{i-1}}{f_i} d_i$$

Kasu partikularrak:

Pertzentilak, dezilak, kuartilak,...

Universidad del País Vasco Euskal Herriko Unibertsitatea

1.7.3 Posiziozko neurriak

Pertzentilak:

Lagina ehun zatitan banatzen dute, beraz, 99 pertzentil daude. k. ordenako pertzentilak, P_k, banaketaren %k balio ezkerrean uzten du (k=1, 2, ..., 99)

Kasu diskretuan: $F_{i-1} < \frac{k \cdot n}{100}$ eta $F_i > \frac{k \cdot n}{100}$ betetzen duen x_i balioa.

$F_i = \frac{k \cdot n}{100}$ betetzen bada, orduan $P_k = \frac{x_i + x_{i+1}}{2}$

Kasu jarraituan:

1) Pertzentila barnean duen $[l_i, l_{1+1})$ tartea:

$F_{i-1} \leq \frac{k \cdot n}{100}$ eta $F_i > \frac{k \cdot n}{100}$ betetzen duen tartea.

2) Formula aplikatu: $P_k = l_i + \dfrac{\frac{k.n}{100} - F_{i-1}}{f_i} d_i$

Universidad del País Vasco Euskal Herriko Unibertsitatea

1.7.3 Posiziozko neurriak

Dezilak:

Lagina hamar zatitan banatzen dute, beraz, 9 dezil daude. k. ordenako dezilak, D_k, banaketaren %10·k balio ezkerrean uzten du. Argi dago: $D_1 = P_{10},\ D_2 = P_{20}, \ldots, D_9 = P_{90}$

Kasu diskretuan: $F_{i-1} < \frac{10 \cdot k \cdot n}{100}$ eta $F_i > \frac{10 \cdot k \cdot n}{100}$ betetzen duen x_i balioa.

$F_i = \frac{10 \cdot k \cdot n}{100}$ betetzen bada, orduan $D_k = \frac{x_i + x_{i+1}}{2}$

Kasu jarraituan:

1) Dezila barnean duen $[l_i, l_{1+1})$ tartea:

$F_{i-1} \leq \frac{10 \cdot k \cdot n}{100}$ eta $F_i > \frac{10 \cdot k \cdot n}{100}$ betetzen duen tartea.

2) Formula aplikatu:

$$D_k = l_i + \frac{\frac{10k.n}{100} - F_{i-1}}{f_i} d_i$$

1.7.3 Posiziozko neurriak

Kuartilak:

Lagina lau zatitan banatzen dute, beraz, 3 kuartil daude. k. ordenako kuartilak, Q_k, banaketaren %25·k balio ezkerrean uzten du. Argi dago: $Q_1 = P_{25},\ Q_2 = P_{50},\ Q_3 = P_{75}$

Kasu diskretuan: $F_{i-1} < \dfrac{25 \cdot k \cdot n}{100}$ eta $F_i > \dfrac{25 \cdot k \cdot n}{100}$ betetzen duen x_i balioa.

$$F_i = \frac{25 \cdot k \cdot n}{100} \text{ betetzen bada, orduan } Q_k = \frac{x_i + x_{i+1}}{2}$$

Kasu jarraituan:

1) Dezila barnean duen $[l_i, l_{1+1})$ tartea:

$F_{i-1} \leq \dfrac{25 \cdot k \cdot n}{100}$ eta $F_i > \dfrac{25 \cdot k \cdot n}{100}$ betetzen duen tartea.

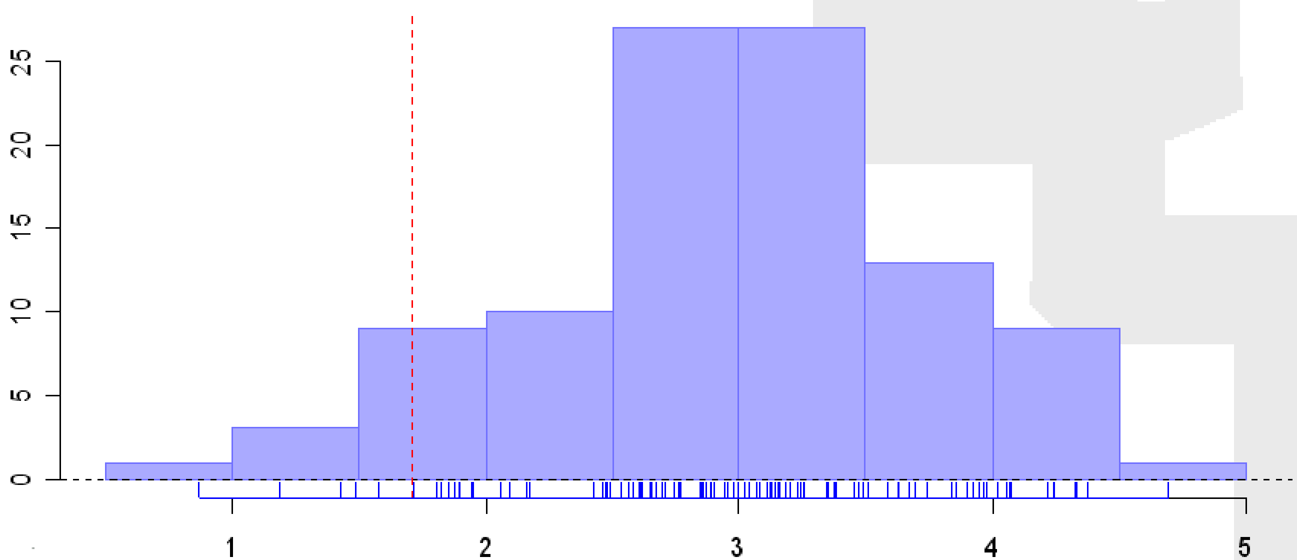
2) Formula aplikatu:

$$Q_k = l_i + \frac{\dfrac{25k.n}{100} - F_{i-1}}{f_i} d_i$$

1.7.3 Posiziozko neurriak

Adibidea 7

Jaioberrien %5ak pisu baxuegia dutela kontutan izanik, zein da baxuegia kontsideratzen den pisua? **5 ordenako pertzentila edo 0,05 ordenako koantilak**

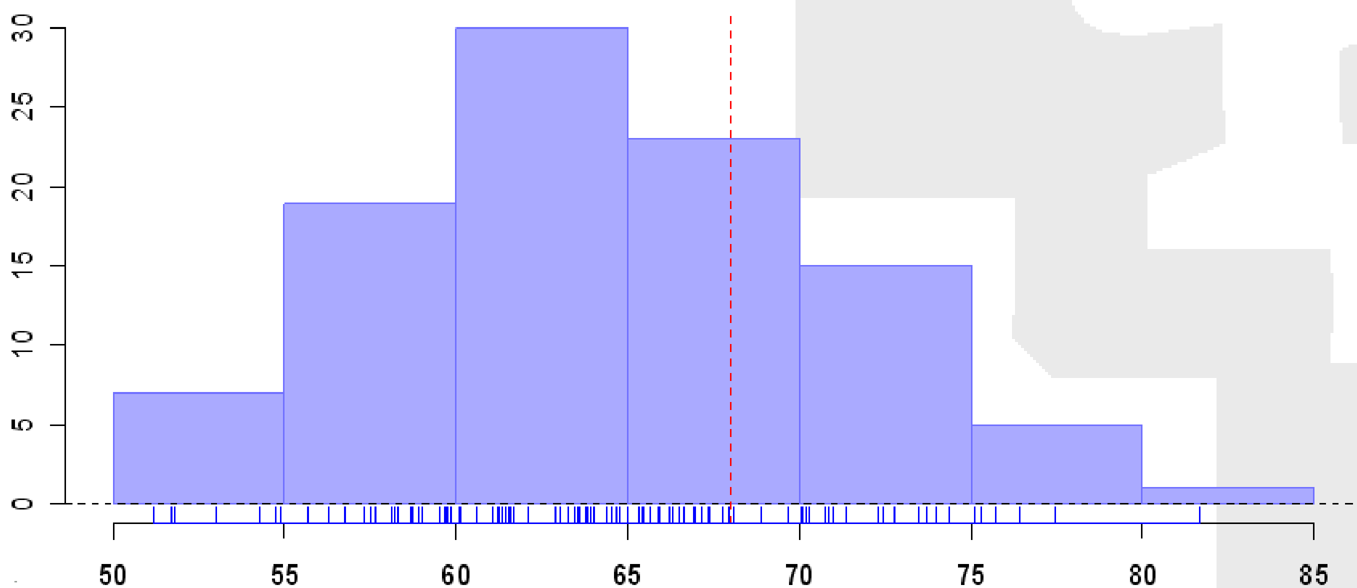


1.7.3 Posiziozko neurriak

Adibidea 8

Zein da inkestatuen %25ak bakarrik gainditzen duten pisua?

75. ordenako pertzentila edo hirugarren kuartila

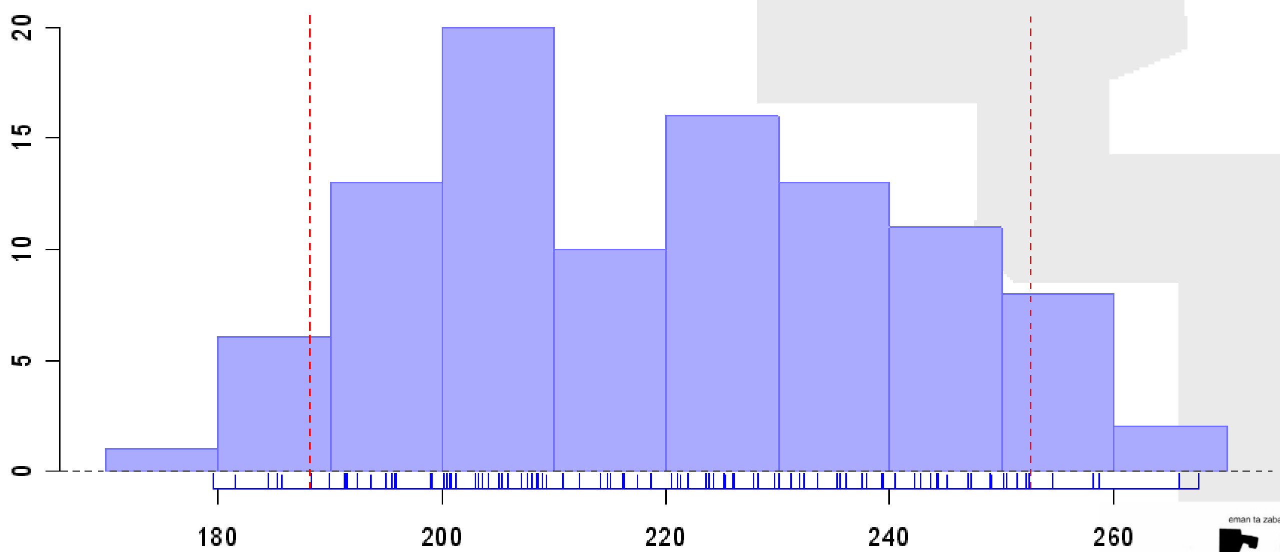


Universidad del País Vasco
Euskal Herriko Unibertsitatea

1.7.3 Posiziozko neurriak

Adibidea 9

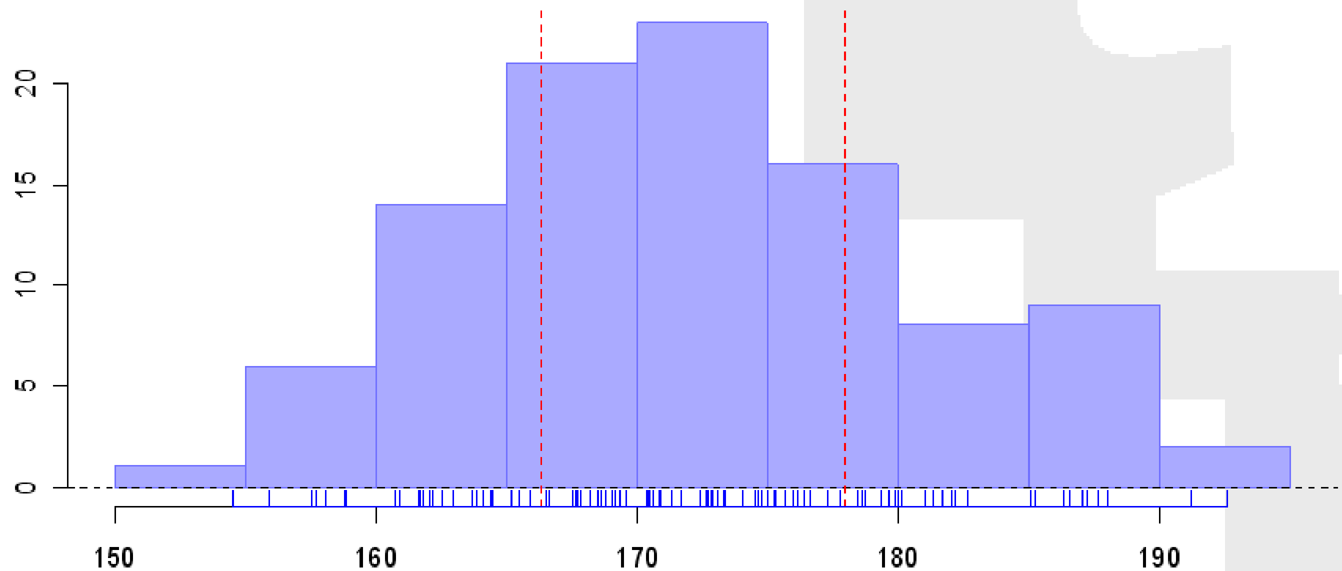
Kolesterola populazioan simetrikoki banatzen da. Muturreko balioak patologiak kontsideratzen direla eta indibiduoen %90a normalak direla jakinik, zehatz ezazu zein balioen artean dauden gaixotasunik ez duten indibiduoak: **5. eta 95. ordenako pertzentilak**



1.7.3 Posiziozko neurriak

Adibidea 10

Zein balioen artean aurkitzen dira populazioko indibiduo "normal"-en %50? **1. eta 3. ordenako kuartilak**



1.7.3 Posiziozko neurriak

Eskolatze urteak

	Frecuencia	Porcentaje	Porcentaje acumulado
3	5	,3	,3
4	5	,3	,7
5	6	,4	1,1
6	12	,8	1,9
7	25	1,7	3,5
8	68	4,5	8,0
9	56	3,7	11,7
10	73	4,8	16,6
11	85	5,6	22,2
12	461	30,6	52,8
13	130	8,6	61,4
14	175	11,6	73,0
15	73	4,8	77,9
16	194	12,9	90,7
17	43	2,9	93,6
18	45	3,0	96,6
19	22	1,5	98,0
20	30	2,0	100,0
Total	1508	100,0	

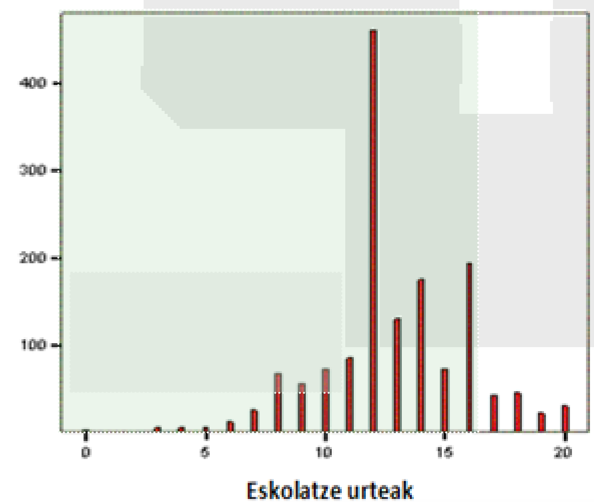
≥20%?

≥ 90%?

Estatistikoak

Eskolatze urteak

N	Válidos	1508
	Perdidos	0
Media		12,90
Mediana		12,00
Moda		12
Percentiles	10	9,00
	20	11,00
	25	12,00
	30	12,00
	40	12,00
	50	12,00
	60	13,00
	70	14,00
	75	15,00
	80	16,00
	90	16,00



1.7.3 Posiziozko neurriak

Pisua	Klase marka	f_i	F_i
[40, 50)	45	5	5
[50, 60)	55	10	15
[60, 70)	65	21	36
[70, 80)	75	11	47
[80, 90)	85	5	52
[90,100)	95	3	55
[100,130)	115	3	58
		58	

$$\bar{x} = \frac{\sum_i x_i n_i}{n} = 69,3$$

$$Me = l_i + \frac{n/2 - F_{i-1}}{f_i} d_i = 66,\widehat{6}$$

$$Mo = l_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} . d_i = 65$$

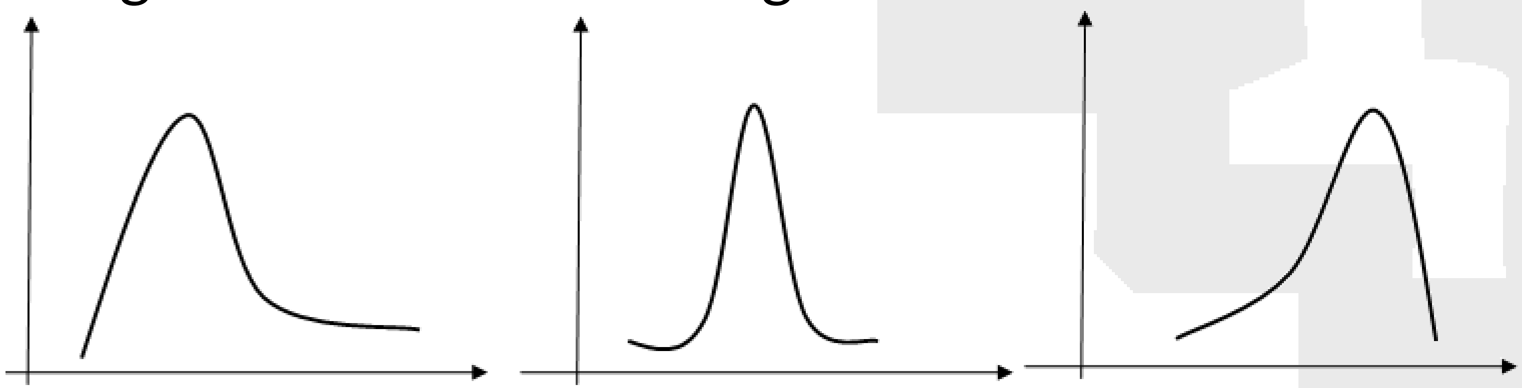
$$P_{75} = Q_3 = l_i + \frac{\frac{75.58}{100} - F_{i-1}}{f_i} d_i = 70 + \frac{43,5 - 36}{11} 10 = 76,8$$

Universidad del País Vasco
Euskal Herriko Unibertsitatea

1.7.4 Formako neurriak

Asimetria edo Alborapen neurriak:

- Banaketaren asimetria (alborapena) aztertzeko balio dute.
- Banaketa simetrikoa (alboragabea) da bere grafikoa simetrikoa denean.
- Alborapen koefizienteak grafikoa egin gabe asimetria dagoen edo ez adierazten digute.



Eskuinerantz alboratua — $Mo \leq Me \leq \overline{x}$

Alboragabea — $Mo = Me = \overline{x}$

Ezkerrerantz alboratua — $\overline{x} \leq Me \leq Mo$

1.7.4 Formako neurriak

Asimetria edo Alborapen neurriak:

Fisher-en asimetria koefizientea

$$g_1 = \frac{m_3}{s(x)^3} = \frac{1}{ns(x)^3}\sum_{i=1}^{n}(x_i - \overline{x})^3 \Rightarrow \begin{cases} g_1 > 0 & \text{eskuinerantz alboratua} \\ g_1 = 0 & \text{alboragabea} \\ g_1 < 0 & \text{ezkerrerantz alboratua} \end{cases}$$

Pearson-en koefizientea:

$$P = \frac{\overline{x} - Mo}{s(x)} \approx \frac{3(\overline{x} - Me)}{s(x)} \Rightarrow \begin{cases} P > 0 & \text{eskuinerantz alboratua} \\ P = 0 & \text{alboragabea} \\ P < 0 & \text{ezkerrerantz alboratua} \end{cases}$$

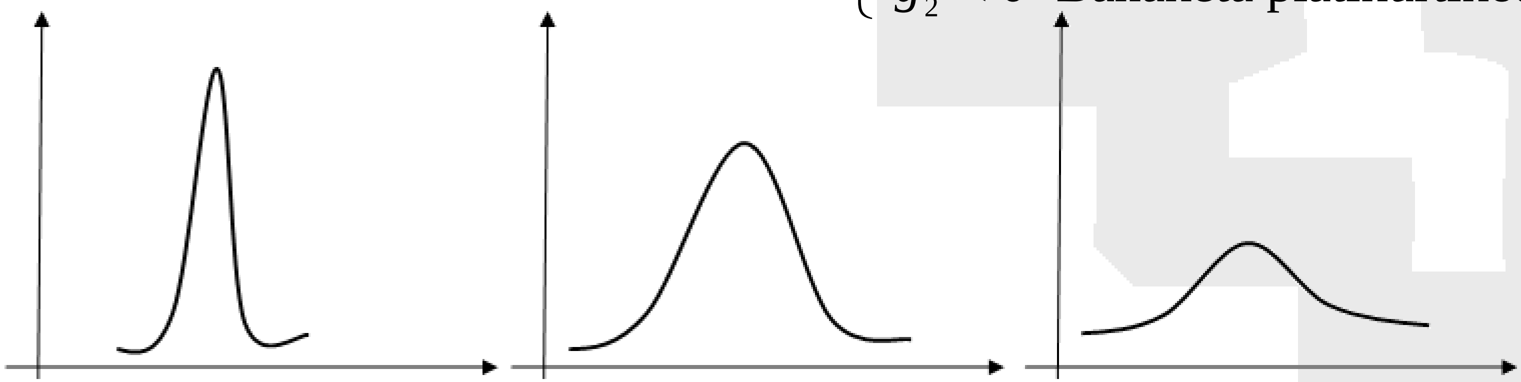
Sarrera
Populazioa eta Lagina
Aldagai Estatistikoak
Maiztasun taulak
Adierazpen grafikoa
Parametroak eta estatistikoak
Estatistiko deskribatzaileak
Eskala eta jatorri aldaketak
Aldagai tipifikatuak
Lege enpirikoa
Kutxa diagrama eta balio arraroak

1.7.4 Formako neurriak

Kurtosia:

Banaketan zorroztasuna neurtzen du (Banaketaren itxura banaketa normalarekin konparatzen du)

$$g_2 = \frac{m_4}{s(x)^4} - 3 = \frac{\sum_{i=1}^{k} (x_i - \overline{x})^4 f_i}{n s(x)^4} - 3 \Rightarrow \begin{cases} g_2 > 0 & \text{Banaketa leptokurtikoa} \\ g_2 = 0 & \text{Banaketa mesokurtikoa} \\ g_2 < 0 & \text{Banaketa platikurtikoa} \end{cases}$$



$g_2 > 0$ Leptokurtikoa

$g_2 = 0$ Mesokurtikoa

$g_2 < 0$ Platikurtikoa

1.7.5 Beste neurriak

Momentuak

c parametroarekiko r. ordenako momentua:

$$M_r(c) = \frac{\sum_{i=1}^{k} (x_i - c)^r . f_i}{n}$$

Momentu zentrala: $c = \overline{x}$ eginez lortzen den momentua

$$m_r = \frac{\sum_{i=1}^{k} (x_i - \overline{x})^r . f_i}{n}$$,ondorioz, $$s^2(x) = m_2(c)$$

Jatorriarekiko momentua: $c = 0$ eginez lortzen den momentua

$$a_r = \frac{\sum_{i=1}^{k} x_i^r . f_i}{n}$$,ondorioz, $$\overline{x} = a_1$$

1.7 Estatistiko deskribatzaileak

Adibidea 11

Hona hemen marka ezagun bateko 27 autoren gasolina-kontsumoa (litrotan) 100km-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

Kalkulatu hurrengo estatistiko deskribatzailea

a) Bataz besteko aritmetikoa, mediana eta moda

b) Desbiderazio tipikoa eta aldakuntza-koefizientea

c) Q_1 , Q_3 kuartilak eta kuartilarteko heina

d) Alborapena eta kurtosia

1.8 Eskala eta jatorri aldaketa

Ikus dezagun aldagai bati eskala eta jatorria aldatzean berari elkartutako estatistikoak nola aldatzen diren:

Eskala aldaketa

Demagun X eta Y aldagaiak ditugula

$$X = k.Y$$

Joera zentraleko neurriak:

$$\overline{x} = k.\overline{y}; \quad Mo(x) = k\,.\,Mo(y);$$

$$Me(x) = k\,.Me(y)$$

Bariantza eta desbiderazio tipikoa:

$$s^2(x) = k^2.\,s^2(y); \quad s(x) = |k|.\,s(y)$$

1.8 Eskala eta jatorri aldaketa

Jatorri aldaketa:

Demagun X eta Y aldagaiak ditugula

$$X = Y + a$$

Joera zentraleko neurriak:

$$\overline{x} = \overline{y} + a; \quad Mo(x) = Mo(y) + a;$$

$$Me(x) = Me(y) + a$$

Bariantza eta desbiderazio tipikoa:

$$s^2(x) = s^2(y); \quad s(x) = s(y)$$

eman ta zabal zazu

Universidad del País Vasco Euskal Herriko Unibertsitatea

1.8 Eskala eta jatorri aldaketa

Eskala eta jatorri aldaketa:

Demagun X eta Y aldagaiak ditugula

$$X = kY + a$$

Joera zentraleko neurriak:

$$\overline{x} = k.\overline{y} + a; \quad Mo(x) = k.Mo(y) + a;$$

$$Me(x) = k.Me(y) + a$$

Bariantza eta desbiderazio tipikoa:

$$s^2(x) = k^2.s^2(y); \quad s(x) = |k|s(y)$$

eman ta zabal zazu

Universidad del País Vasco Euskal Herriko Unibertsitatea

1.9 Aldagai tipifikatuak

- Aldagai bat zentratua dagoela esaten da bere batezbestekoa 0 bada.

  X aldagaia zentratzeko: $X - \overline{x}$

- Aldagai bat zentratua egoteaz gain bere bariantza bat bada, orduan, aldagaia tipifikatuta dagoela esaten da:

  $X - \overline{x}$ aldagaiaren bariantza bat izateko:

$$\frac{X - \overline{x}}{s(x)}$$

- $Z = \dfrac{X - \overline{x}}{s(x)}$ aldagai tipifikatua (estandarizatua) da.

Z_i balioak X_i eta $\overline{x}$ -ren arteko distantzia neurtzen du eta konparaketak egiteko balio du.

1.9 Aldagai tipifikatuak

Adibidea 12

$z_i = 1,5$ bada, x_i datua $\overline{x}$ batezbestekoa baino $1,5.s(x)$ aldiz handiagoa da.

$z_i = -1,5$ bada, berriz, x_i datua $\overline{x}$ batezbestekoa baino $1,5.s(x)$ aldiz txikiagoa da.

Bukatzeko,

$$z_i = 0 \Rightarrow x_i = \overline{x}$$

1.9 Aldagai tipifikatuak

Adibidea 13

Izan bedi ospitale batean jaiotako umeen luzera adierazten duen X aldagai estatistikoa, bere batezbesteko aritmetikoa 50 cm eta desbiderazio tipikoa 1.6 cm izanik. Bestalde, kontsidera dezagun Y 18-28 urteen bitarteko gizonen altuera adierazten duen aldagaia estatistikoa kasu honetan, batezbesteko aritmetikoa 175 cm eta desbiderazio tipikoa 6.7 cm izanik.

54 cm dituen jaioberri bat eta 190 cm neurtzen dituen mutil bat aukeratu ditugu. Bietako zein da erlatiboki (euren taldeekiko) altuagoa?

1.10 Lege enpirikoa

Lege enpirikoa:

Datuen banaketak kanpai itxura duenean, hurrengo emaitzak ditugu:

- Gutxi gorabehera datuen %68-a, (%68.27) tarte honetan dago:

$$\left(\overline{x}-s(x),\ \overline{x}+s(x)\ \right)$$

- Gutxi gorabehera datuen %95-a, (%95.45) tarte honetan dago:

$$\left(\overline{x}-2s(x),\ \overline{x}+2s(x)\ \right)$$

- Datu gehienak, (%99.73) tarte honetan daude:

$$\left(\overline{x}-3s(x),\ \overline{x}+3s(x)\ \right)$$

Universidad del País Vasco Euskal Herriko Unibertsitatea

Sarrera

Populazioa eta Lagina

Aldagai Estatistikoak

Maiztasun taulak

Adierazpen grafikoa

Parametroak eta estatistikoak

Estatistiko deskribatzaileak

Eskala eta jatorri aldaketak

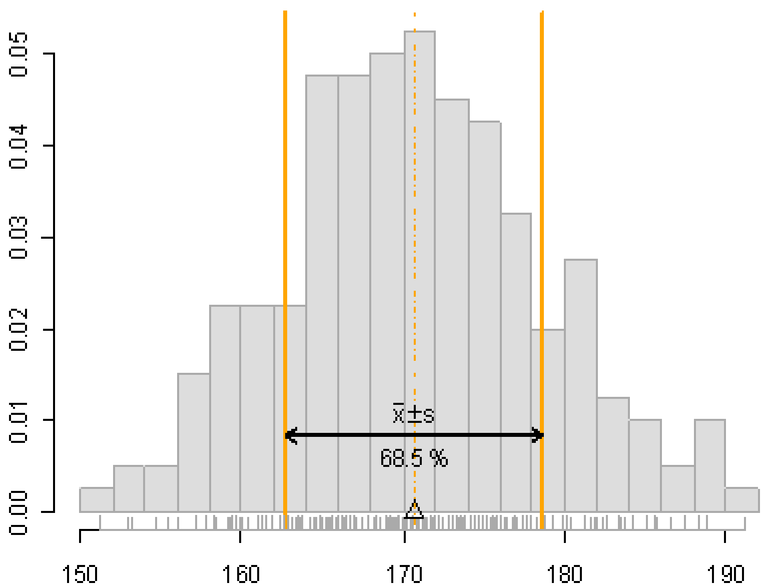
Aldagai tipifikatuak

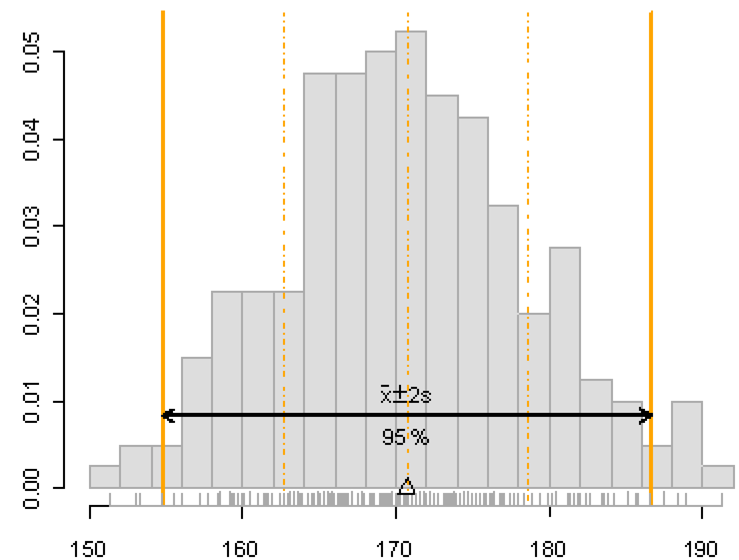
Lege enpirikoa

Kutxa diagrama eta balio arraroak

1.10 Lege enpirikoa

Grafikoki:





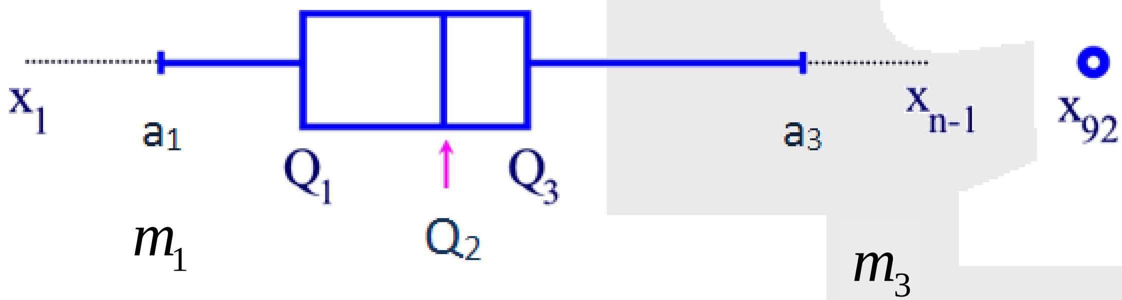
eman ta zabal zazu

Universidad del País Vasco Euskal Herriko Unibertsitatea

1.11 Kutxa diagrama eta balio arraroak

Kutxa diagrama:

$Q_1,\ Q_2$ eta Q_3 kuartilak erabiliz kutxa batean behaketen %50 a adierazten da.



Pausuak:

1. $Q_1,\ Q_2 = Me,\ Q_3$ kuartilak eta $RIC = Q_3 - Q_1$ kuartilarteko heina kalkulatu.

2. Barne hesiak kalkulatu:

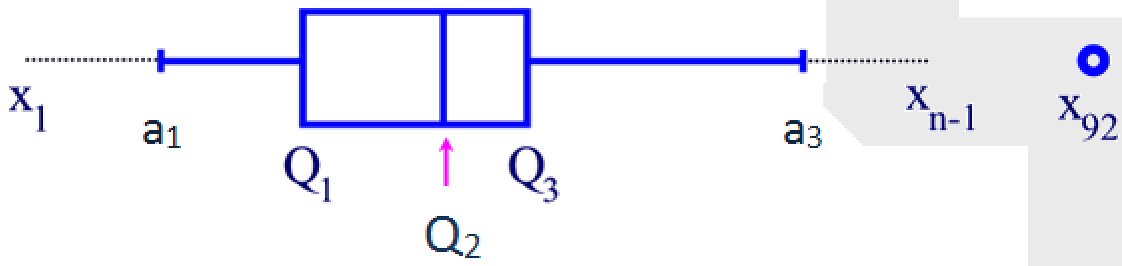
$$m_1 = Q_1 - 1.5RIC \quad \text{eta} \quad m_3 = Q_3 + 1.5RIC$$

1.11 Kutxa diagrama eta balio arraroak

3. a_1 eta a_3 balioak lortu:

$a_1 : m_1$ balioa baino handiagoa edo berdina izanik m_1-etik hurbilen dagoen datua da.

$a_3 : m_3$ balioa baino txikiagoa edo berdina izanik m_3-tik hurbilen dagoen datua da.



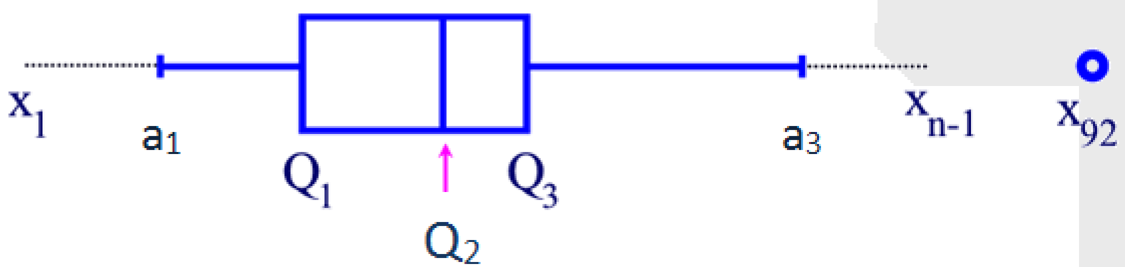
1.11 Kutxa diagrama eta balio arraroak

4. Kanpo hesiak kalkulatu:

$$M_1 = Q_1 - 3.RIC \quad \text{eta} \quad M_3 = Q_3 + 3.RIC$$

(M_1, m_1) edo (m_3, M_3) tarteetan dauden balioei *outlier* edo balio arraroak deitzen zaie, eta ∘ bidez adierazten dira.

$(-\infty, M_1)$ edo $(M_3, +\infty)$ tarteetan dauden balioei berriz izugarrizko *outlier*-ak deritze eta * bidez adierazten dira.

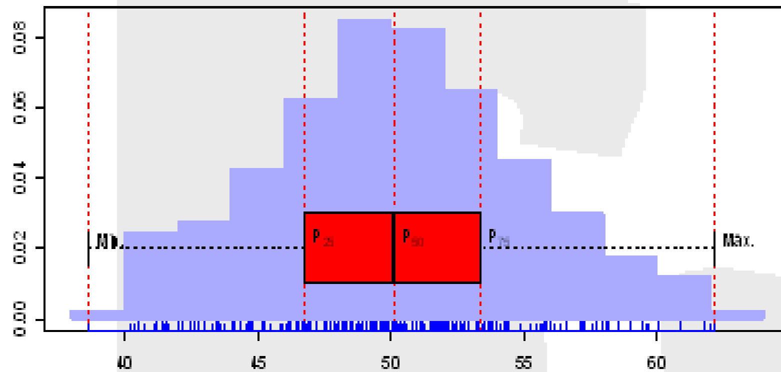


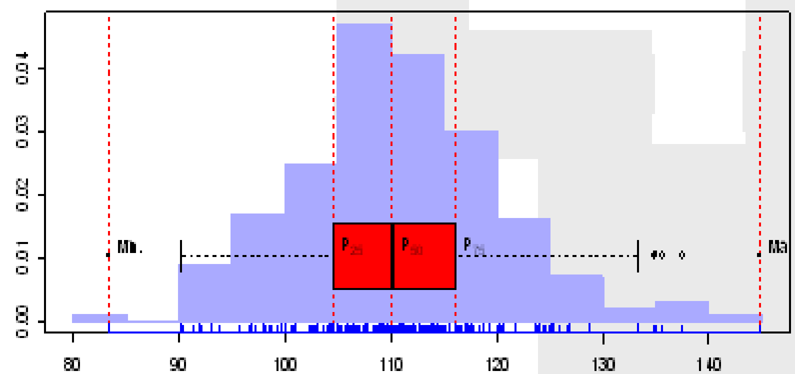
1.11 Kutxa diagrama eta balio arraroak

- 5 zenbaki erabiltzen dituen laburpena:
  - Minimoa, maximoa eta kuartilak.
  - Banaketari buruzko informazioa ematen dute.

- '**Kutxa**'-k behaketa zentralen %50a ditu.
  - Kutxaren tamaina kuartilarteko heina da (RIC)

- Orokorrean biboteak ez dira muturretako balioetaraino heltzen.
  - Biboteetatik ezkerrerago edo eskuinerago dauden balioak, balio arraroak edo outlierrak kontsideratzen dira.





1.11 Kutxa diagrama eta balio arraroak

Adibidea 14

Hurrengo taulan kalean inkestatu diren 40 pertsonen adinak agertzen dira.

xi	15	16	17	18	20	23	24	25	26	27
fi	2	2	1	2	2	2	3	4	2	3

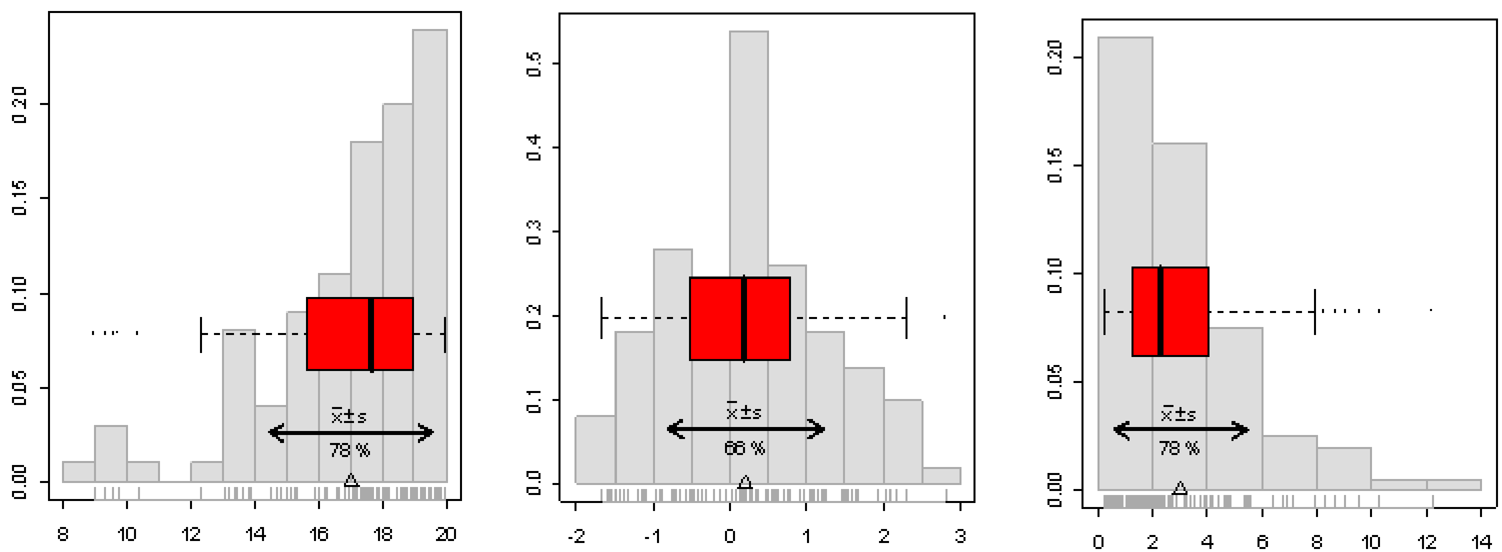
xi	28	29	38	45	47	50	54	59	63	66	85
fi	1	3	5	1	1	1	1	1	1	1	1

Irudika ezazu kutxa diagrama. Ba al dago balio arrarorik?

1.11 Kutxa diagrama eta balio arraroak

Kutxa diagrama eta alborapena:

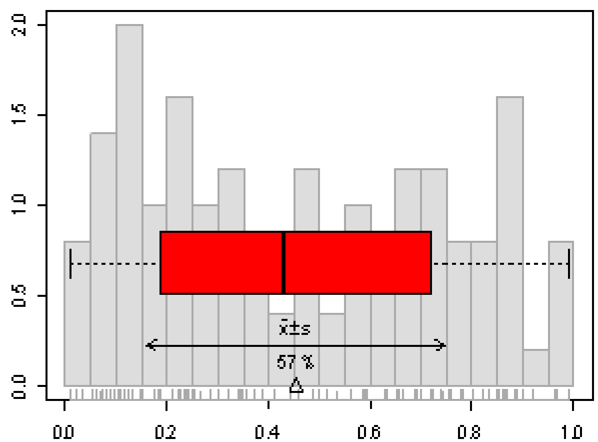
- ✓ Mediana kutxaren erdian badago simetria dago.
- ✓ Mediana Q_3 –tik Q_1 –etik baino hurbilago badago, orduan banaketa ezkerrera alboratua dago.
- ✓ Mediana Q_1 –etik Q_3 –tik baino hurbilago badago, orduan banaketa eskumara alboratua dago.
- ✓ Kutxa oso luzea ez bada, aldakortasun gutxi dago.



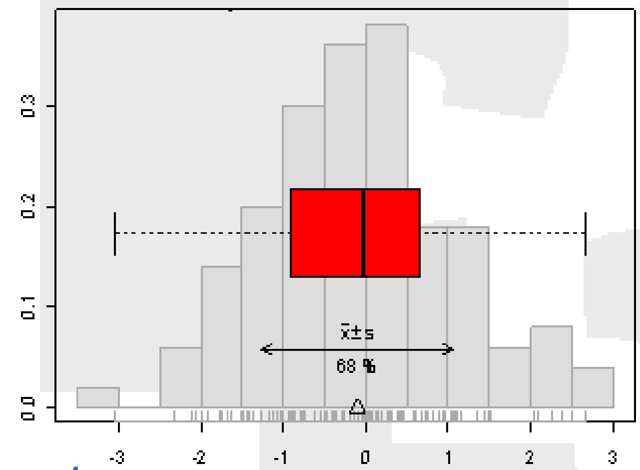
1.11 Kutxa diagrama eta balio arraroak

Kutxa diagrama eta kurtosia:

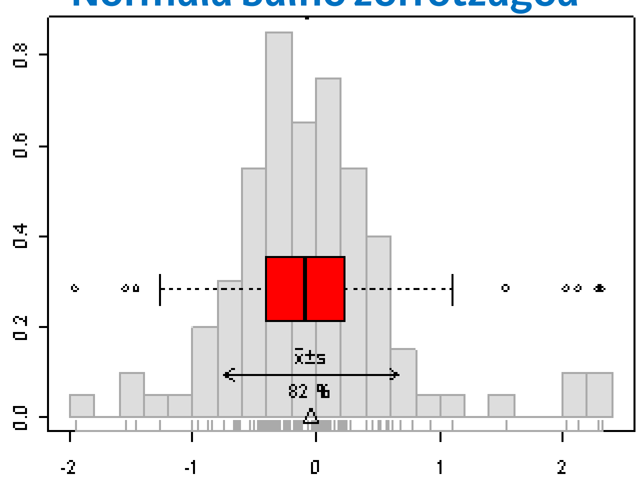
Normala baino leunagoa



Normala bezain zorrotza



Normala baino zorrotzagoa



Universidad del País Vasco
Euskal Herriko Unibertsitatea