1.Gaia:
Estatistika Deskribatzailea

Cristina Alcalde - Arantxa Zatarain

Donostiako Unibertsitate Eskola Politeknikoa - UPV/EHU

Definizioa

Populazioa

Elementu multzo baten ezaugarriren bat ezagutu nahi dugunean elementu guztiek populazio edo unibertsoa osatzen dute.

Aldagaia

Aztertzen ari garen ezaugarria aldagaia da.

Definizioa

Lagina

Elementu-multzoa oso handia denean edo banan-banan elementu bakoitzaren ezaugarriak aztertzea oso zaila denean elementu batzuk aukeratzen dira eta elementu horiek lagina osatzen dute.

Definizioa

Datu motak

Estatistikaren bidez aztertzen diren ezaugarrien motak:

1. Kualitatiboak edo atributuak.
2. Kuantitatiboak edo aldagaiak.
   - Diskretuak.
   - Jarraituak. (Aldagaiak bi balioen artean edozein balio har dezake.)

Definizioa

Maiztasun absolutua

x_i balio baten maiztasuna (edo maiztasun absolutua) n_i balio hori ditugun datu multzoan zenbat aldiz azaltzen den adierazten duen zenbakia da. Balioen maiztasuna datu kopuruaz zatitzen dugunean maiztasun erlatiboa f_i lortzen dugu.

Maiztasun erlatiboa

Balioen maiztasuna datu kopuruaz zatitzen dugunean maiztasun erlatiboa f_i lortzen dugu.

Definizioa

Maiztasun absolutu metatua

Balio bat edo bere aurrekoak (txikiagoak) agertzen diren aldi kopurua balioaren maiztasun (absolutu) metatua da.

Maiztasun erlatibo metatua

Maiztasun metatua zati datu kopurua maiztasun erlatibo metatua da.

Maiztasun-taulak

Aldagaia diskretua bada:

Aldagaien balioak x_i	maiztasun absolutuak n_i	maiztasun erlatiboak $f_i = n_i/n$	maiztasun absolutuak metatuak N_i	maiztasun erlatiboak metatuak $F_i = N_i/n$
x_1	n_1	$f_1 = n_1/n$	$N_1 = n_1$	$F_1 = N_1/n$
x_2	n_2	$f_2 = n_2/n$	$N_2 = N_1 + n_2$	$F_2 = N_2/n$
$\vdots$				
x_k	n_k	$f_k = n_k/n$	$N_k = N_{k-1} + n_k = n$	$F_k = N_k/n = 1$
totalak	n	1		

Maiztasun-taulak

Datu ugari jaso direnean edo aldagai jarraitua dugunean, azterketa egiteko datuak mailatan edo klaseetan sailkatzea komeni da. Klasearen erdiko puntua klase-ordezkaria da eta azterketa egiterakoan tartearen elementu guztiak klase-ordezkariaz ordezkatzen dira. Horrela egiterakoan maiztasun taula era honetakoa gelditzen da.

klasearen mugak $a_i - a_{i+1}$	ordezkaria c_i	maiztasun absolutuak n_i	maiztasun erlatiboak $f_i = n_i/n$	maiztasun absolutuak metatuak N_i	maiztasun erlatiboak metatuak $F_i = N_i/n$
$a_1 - a_2$	c_1	n_1	f_1/n	$N_1 = N_1$	N_1/n
$a_2 - a_3$	c_2	n_2	f_2/n	$N_2 = N_1 + n_2$	N_2/n
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$a_k - a_{k+1}$	c_k	n_k	$f_i = n_1/n$	$N_k = N_{k-1} + n_k = n$	$N_k/n = 1$
totalak		n	1		

Maiztasun-taulak

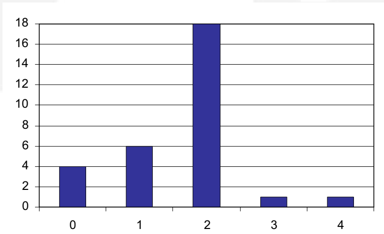
Maiztasun-banaketa lortzeko lehendabizi klaseak definitu beharko ditugu, eta horretarako honako pausu hauek egingo ditugu:

1. Datuen balio handiena eta txikiena aurkitu, bien kendura datuen heina edo ibiltartea da.
2. Ibiltartea klase edo tarte kopuruaz zatitu, klaseen zabalera lortzeko,
   Tarte kopurua 5 eta 20 bitartean aukeratuko dugu.
   Aukeratzeko bi irizpide erabil dezakegu, datu kopuruaren erro karratuaren hurbilketa osoa izan dadin edo Struges-en erregela $1 + 3.322 \ln N$ balioaren hurbilketa osoa. Klaseen zabalera lortzerakoan lortzen dugun zatidura berriro ere borobildu egingo dugu datuen arabera.

Adierazpen grafikoak

Barra-diagramak

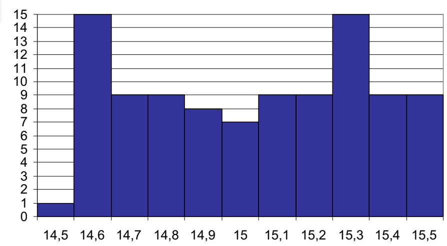
Ezaugarri diskretuak direnean erabiltzen dira, abzisa-ardatzean balio ezberdinak ipintzen dira eta ordenatu ardatzean maiztasunak.



Adierazpen grafikoak

Histogramak

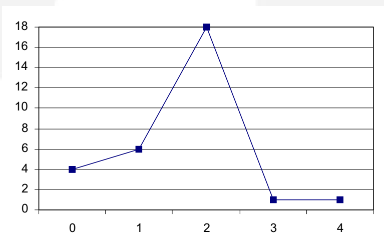
Ezaugarri jarraituak ditugunean oso baliagarriak dira. Datuak mailetan zatitu ditugunean, maila bakoitzari lauki zuzen bat dagokio. Maila edo klasea oinarritzat hartuta, lauki zuzenen azalerak maiztasunen araberakoak dira. Klase guztiak zabalera berdinekoak direnean, lauki zuzenen altuerak maiztasuna adierazten du.



Adierazpen grafikoak

Maiztasun poligonoak

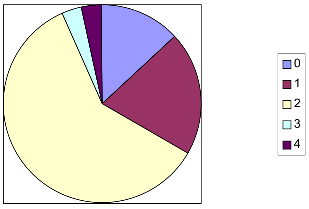
Barra diagramen bezala abzisa balioa eta ordenatua maiztasuna duten puntuak zuzenkien bidez lortzen ditugun lerro poligonalaren bidezko adierazpenak dira.



Adierazpen grafikoak

Sektore grafikoak

Grafiko honetan zirkulu baten sektoreen bidez adierazten ditugu maiztasunak, zirkuluaren 360 graduak proportzionalki banatzen dira maiztasunen arabera sektorek dituzten graduak lortzeko.



Joera zentraleko neurriak

Definizioa

Joera zentraleko neurriak aldagaia balio bakar baten bidez deskribatzeko erabiltzen diren zenbakiak dira.

Joera zentraleko neurriak: Batezbestekoa

Batezbestekoa

Joera zentraleko indizerik erabiliena da.
Behatutako balio guztiak batuz eta balio kopuruarekin zatituz lortzen den zenbakia da:

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

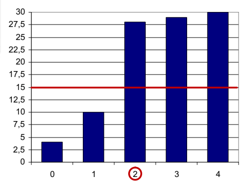
- Maiztasun taula lortu badugu eta x_i $(i = 1, \ldots k\ )$ balioen maiztasun absolutuak n_i badira: $\bar{x} = \frac{\sum_{i=1}^{k} x_i \cdot n_i}{\sum_{i=1}^{k} n_i}$.
- Maiztasun erlatiboak f_i badira: $\bar{x} = \sum_{i=1}^{k} x_i \cdot f_i$.
- Datuak klasetan sailkatu baditugu, batezbestekoa lortzeko x_i bezala klasearen ordezkaria hartzen dugu.

Joera zentraleko neurriak: Mediana

Mediana

Balio kopuruen erdia azpitik eta beste erdia gainetik dituen balioa da, hau da, balioak ordenatuz mediana erdiko posizioan dagoen balioa da.

Maiztasun metatuen taula dugunean $n/2$ maiztasuna duena da.
Maiztasun metatu erlatiboen poligonoa hartzen dugunean, %50 maiztasun puntutik OX ardatzarekin paraleloa den zuzena eta poligonoaren ebakigunearen abzisa da mediana. Datu kopurua bikoitia bada erdiko bien bitarteko balioa da.



Joera zentraleko neurriak: Mediana

Aldagaia jarraituaren medianaren kalkulua:
Balioak tartekatuta ditugunean maiztasun metatua $n/2$ duen tartearen mugen artean interpolatu egiten dugu era honetan:

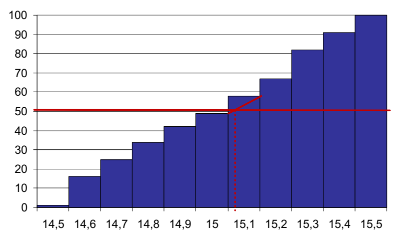
1. maiztasun metatuen taulan $n/2$ balioa zer tartetan dagoen ikusten dugu.
2. tartearen mugak a_i eta a_{i+1} hartzen ditugu
3. tarte horren maiztasuna n_i
4. aurreko tartearen maiztasun metatua N_{i-1}

balioekin formula honen bidez mediana lortzen dugu

$$\text{Mediana} = a_i + \left(\frac{n/2 - N_{i-1}}{n_i} \right) (a_{i+1} - a_i)$$

Joera zentraleko neurriak: Mediana

Grafikoki:



$$\text{Mediana} = a_i + \left(\frac{n/2 - N_{i-1}}{n_i} \right) (a_{i+1} - a_i)$$

Joera zentraleko neurriak: Moda

Moda
Moda maiztasun handieneko balioa da. Gehien errepikatzen den aldagaia edo ezaugarriaren balioa da.
Gehien errepikatzen diren balioak bi direnean banaketa bi modala da, eta gehien errepikatzen diren balioak bi baino gehiago direnean multimodala.
Izan daiteke modarik ez izatea, balioak behin azaltzen direnean adibidez.

Joera zentraleko neurriak: Moda

Datuak tartetan sailkatuak ditugunean moda era honetan lortzen dugu

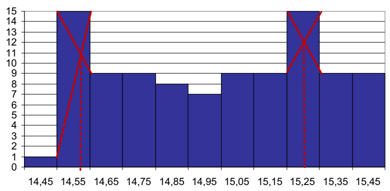
1. maiztasun taulan maiztasun handieneko tartea aurkitzen dugu.
2. tartearen mugak a_i eta a_{i+1} hartzen ditugu
3. tarte horren maiztasuna n_i
4. aurreko tartearen maiztasuna n_{i-1} eta hurrengo tartearena n_{i+1}

balioak formula honetan ordezkatzen ditugu:

$$\text{Moda} = a_i + \left(\frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} \right) \cdot (a_{i+1} - a_i)$$

Joera zentraleko neurriak: Moda

Grafikoki:



Kasu honetan banaketa bimodala dugu.

Lehengo moda: $\text{Moda} = 14,5 + \left(\frac{15-1}{(15-1)+(15-9)} \right) \cdot (14,6 - 14,5) = 14,56$

Bigarren moda: $\text{Moda} = 15,2 + \left(\frac{15-9}{(15-9)+(15-9)} \right) \cdot (15,2 - 15,1) = 15,25$

Joera zentraleko neurriak: Pertzentilak

Pertzentilak
Beste lekuzko neurriak dira, mota desberdinekoak dira eta datuak ordenatzen ditugunean, datuen portzentaia berezi bat azpitik duten balioak dira.
Adibidez, koartilak hiru dira, lehenengoa koartilaren azpitik datuen %25a gelditzen da, bigarrenaren azpitik %50a eta hirugarrenaren %75a.
Era berean, dezilak %10, %20,. . . ,%90 portzentajeekin definitzen dira, zentilak %1, %2,. . . ,%99 portzentajeekin.

Sakabanatze neurriak

Definizioa
Aldagaiaren balioak zentro baten inguruan biltzen diren edo harengandik sakabanatzen diren jakiteko erabilitako neurriak dira.

Sakabanatze neurriak: Ibiltartea

Ibiltartea
Zenbakizko datuak direnean balio maximo eta minimoaren arteko kendura da.

Sakabanatze neurriak: Bariantza

Bariantza eta desbiderapen tipikoa sakabanatze-neurririk garrantzitsuenak dira.

Bariantza

- Era honetan definitzen da,

$$\text{Bar} = S_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

- Lagina dugunean, era honetan definitzen da,

$$\text{Bar} = S_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Maiztasun taula dugunean:

$$\text{Bar} = S_x^2 = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 \cdot n_i}{n}$$

Sakabanatze neurriak: Bariantza

Propietatea

Bariantza era honetan kalkula daiteke,

$$\text{Bar} = S_x^2 = \frac{\sum_{i=1}^{k} x_i^2 \cdot n_i}{n} - \bar{x}^2$$

Sakabanatze neurriak: Desbiderapen tipikoa

Desbiderapen tipikoa

Desbiderapen tipikoa bariantzaren erro karratua da, era honetan batezbestekoa eta desbiderapen tipikoa antzerako magnitudeko neurriak dira, bariantzarekin gertatzen ez dena.

$$\sigma = S_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$$

- Maiztasun taula dugunean, $S_x = \sqrt{\frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 \cdot n_i}{n}}$
- Aurreko propietateagatik, $S_x = \sqrt{\frac{\sum_{i=1}^{k} x_i^2 \cdot n_i}{n} - \bar{x}^2}$

Sakabanatze neurriak: Batezbesteko desbiderapena

Batezbesteko desbiderapena

Datuak eta batezbestekoaren arteko diferentziaren balio absolutuen batezbesteko aritmetikoa da.

$$D_x = \frac{\sum_{i=1}^{n}|x_i - \bar{x}|}{n}$$