ESTATISTIKA ETA DATUEN ANALISIA

III: Aldagai bakunaren deskribapena: zentroa eta beste kokapenak

Egilea: Josemari Sarasola

Gizapedia

gizapedia.hirusta.io

3.1 Eranskina: batukaria

3.2 Zentro-neurriak

3.2.1 Batezbesteko aritmetiko sinplea

3.2.1.1 Kalkulua datuak maiztasun-taula batean bildurik daudenean

3.2.1.2 Kalkulua datuak tarteka bildurik daudenean

3.2.2 Mediana

3.2.2.1 Kalkulua datuak tarteka bildurik daudenean

3.2.3 Batezbestekoa eta mediana alderatuz: datu atipikoak eta sendotasuna

3.2.4 Moda

3.2.5 Batezbesteko bereziak

3.2.5.1 Batezbesteko aritmetiko haztatua

3.2.5.2 Batezbesteko kuadratikoa

3.2.5.3 Batezbesteko geometrikoa

3.2.5.4 Batezbesteko harmonikoa

3.3 Kuantilak

3.3.1 Kuantilen kalkulua

3.3.1.1 Kuantilen kalkulua tartekako datuetan

3.4 Ariketak

3. gaia: Aldagai bakunaren deskribapena: zentroa eta beste kokapenak

Aurreko ikasgaian ikasitako grafikoen bitartez aldagai kuantitatiboen ezaugarriak azter daitezkeela ikasi dugu, bereziki *zentroa* eta *sakabanatzea*. Ikasgai honetan, zentro horretarako neurri estatistikoak ikasiko ditugu, ezaugarri hori zehaztasunez neurtzearren. Zentroaz gainera, beste kokapen batzuetarako neurriak ere ikasiko ditugu. Neurri horietako askoren formuletan agertzen den ikur batukariaren erabilera ikasten hasiko gara horretarako, eranskin gisa.

3.1 Eranskina: batukaria

Batukaria zenbaki batzuen batuketa adierazteko erabiltzen den ikurra da. Greziar alfabetoko sigma maiuskulaz adierazten da. Adibidez, $x_1 = 3,\ x_2 = 1,\ x_3 = 5,\ x_4 = 3$ izanik,

$$\sum_{i=1}^{4} x_i = 3 + 1 + 5 + 4 = 11$$

Oro har, n datuen batura honela adierazten da:

$$\sum_{i=1}^{n} x_i$$

Laburrago ere adieraz daiteke: $\sum_i x_i$ edo $\sum x_i$.

Batukariak baditu zenbait propietate:

1. k konstante bat izanik,

$$\sum_{i=1}^{n} k = k + k + \ldots + k = nk$$

2. k konstante bat izanik,

$\sum_{i=1}^{n} kx_i = kx_1 + kx_2 + \ldots + kx_n = k(x_1 + x_2 + \ldots + x_n) = k \sum_{i=1}^{n} x_i$

3. $\sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$

Oro har, propietate hauek gogoratzeko, integralaren propietateak gogoratzea komeni da, integrala azalera-zati txiki-txikien batuketa bat besterik ez baita. Adibidez, bidertzen duten konstanteak batukaritik eta integraletik kanpora atera daitezke (bigarren propietatea) eta batukariaren eta integralaren barnean dauden batuketak bereizi egin daitezke batukari eta integral zenbaitetan.

3.2 Zentro-neurriak

Zentroa datuak zein balioren inguruan biltzen diren adierazten duen aldagai koantitatiboen ezaugarria da. Balio hori datu-multzo osoaren balio adierazgarri moduan erabiltzen da, orokorrean izaten den balioa zehazteko. Adibidez, ikasle batzuen puntuazioen zentro-neurri bat 6 izanik, ikasleek orokorrean edo batezbestez 6 puntuko emaitza izaten dutela ondorioztatuko da. Jarraian, zentroa neurtzeko dagoen neurri sorta zabala ikasiko dugu.

3.2.1 Batezbesteko aritmetiko sinplea

$x_1,\ x_2, \ldots,\ x_n$ datuak edukita (n lagin-tamaina izanik), honela adierazi eta kalkulatzen da haien batezbesteko aritmetiko sinplea:

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

Gehien erabiltzen den zentro-neurria da, datu guztiak barnehartzen dituelako eta aise kalkulatzen eta interpretatzen delako.

Adibidea: Bi langilek hainbat egunetan egindako ekoizpenak jaso dira (unitatetan):

A langilea: 29-40-36-35-34
B langilea: 34-38-43-39

Zein da batezbestez gehien ekoizten duen langilea? Batezbesteko aritmetiko sinplea erabili behar da horretarako.

$$\overline{x}_A = \frac{29+40+36+35+34}{5} = 34.8 \; unitate$$
$$\overline{x}_B = \frac{34+38+43+39}{4} = 38.5 \; unitate$$

B langilea da batezbestez gehien ekoizten duena, 3.7 unitate gehiago zehatzago. Hala ere, emaitza horiek *lagin-errorearen erreserbapean* daudela hartu behar da kontuan, datuek eguneko ekoizpen guztien lagin bat osatzen dutelako.

Ohartu behar da B langileak oro har ekoizpen handiagoa izateak ez duela esan nahi B langileak *beti* ekoizpen handiagoa izango duenik. Batezbestez B langileak gehiago ekoiztuta ere, baliteke egun batzuetan A langileak ekoizpen handiagoa izatea.

3.2.1.1 Kalkulua datuak maiztasun-taula batean bildurik daudenean

Datu kuantitatiboak maiztasun-taula batean bildurik daudenean, batezbesteko aritmetiko sinplea kontzeptualki berdina da, baina kalkulua beste era batera egiten da, maiztasun-taula oinarrituz.

Adibidea: Lantegi batean azken hilabetean langile bakoitzak izandako absentismo-egunak:

Egunak (x)	Langileak (n)	Portzentajeak (f)
0	14	%56
1	7	%28
2	3	%12
3	0	%0
4	1	%4
	25	%100

Kalkulatu hileko absentismo-egun kopuruaren batezbestekoa.

25 datuen zerrenda osatu eta horien batura kalkulatu ordez, maiztasun-taulan bertan oinarrituko gara kalkuluak egiteko. Horretarako nx zutabea eratuko da, lehen eta bigarren zutabeak bidertuz. nx zutabeak x balioa hartzen duten datuen batura partziala adierazten du. Azkenik, batura partzial horien baturak

datu guztien batura adieraziko du, eta datu kopuruarekin zatituz, batezbesteko aritmetiko sinplearen balioa emango digu.

Portzentaje edo maiztasun erlatiboetatik ere kalkula daiteke batezbestekoa. Kasu horretan, aski da fx kalkulatzea eta 100 balioarekin zatitzea.

Egunak (x)	Langileak (n)	Portzentajeak (f)	nx	fx
0	14	%56	0	0
1	7	%28	7	28
2	3	%12	6	24
3	0	%0	0	0
4	1	%4	4	16
	25	%100	17	68

Beraz, n maiztasun absolutuetatik abiatuz:

$$\overline{x} = \frac{17}{25} = 0.68\ egun$$

Eta maiztasun erlatiboetatik abiatuz:

$$\overline{x} = \frac{68}{100} = 0.68\ egun$$

3.2.1.2 Kalkulua datuak tarteka bildurik daudenean

Datuak tarteka bilduta daudenean, banakako datuen balio zehatza ezin da jakin, eta beraz, datuen batezbesteko zehatza ezin da kalkulatu. Irtenbidea, tarteko datu guztiek tartearen erdiko balioa hartzen dutela pentsatzea, eta batezbestekoa aurreko puntuan bezala kalkulatzea da. Horrela, ordea, eskuratzen den emaitza hurbilketa bat izango da, baina oro har nahiko fina, tarteko balio txikien eta handien ordea erdiko balioa hartzean sortzen diren erroreak konpentsatu egiten baitira.

Adibidea: Lantegi batean langileen adinak jaso dira:

Adina	Langileak (n)
15-20	2
20-25	7
25-30	8
30-35	4
35-40	1
	22

Kalkulatu langileen adinen batezbesteko aritmetiko sinplea.

Adina	Langileak (n)	x	nx
15-20	2	17.5	35
20-25	7	22.5	157.5
25-30	8	27.5	220
30-35	4	32.5	130
35-40	1	37.5	37.5
	22		580

$$\overline{x} = \frac{580}{22} = 26.36 \; urte$$

Beraz, langileen batez besteko adina 26.36 urtekoa da (26 urte, 0.36 × 12=4.32 hilabetekoa). Datuak tartetan biltzearen errorea ere kontuan hartu behar da (izan ere, baliteke bi langile gazteenak 15 eta 16 urtekoak izatea, eta ez 17.5 suposatu dugun bezala).

3.2.2 Mediana

Mediana (Me), datuak txikienetik handienera ordenaturik, datu horien erdian dagoen balioa da; hau da, alde banatara datuen %50ak uzten dituena. Gehien erabiltzen den bigarren zentro-neurria da, batezbesteko aritmetiko sinplearen ondoren.

Datu-kopurua bakoitia bada, argi dago zein den erdiko balioa (adibidez, 5 datu izanik, erdiko datua, ordenatu ondoren, 3garrena da). Datu-kopurua bikoitia denean, medianatzat hartuko dugu erdiko datuen balioen batezbestekoa) (adibidez 6 datu badaude, ez da erdiko datu bakarra, erdikoak 3garren eta 4garren datuak dira, eta horien batezbestekoa hartuko dugu medianatzat). Hala ere, bibliografia estatistikoan mediana kalkulatzeko beste metodo batzuk ere aurki daitezke.

Adibidea: Bi langilek hainbat egunetan egindako ekoizpenak jaso dira (unitatetan):

A langilea: 29-40-36-35-34
B langilea: 34-38-43-39

Zein da batezbestez gehien ekoizten duen langilea? Mediana erabili behar da horretarako.

Datuak txikienetik handienera ordenatu behar dira lehenbizi:

A langilea: 29-34-35-36-40
B langilea: 34-38-39-43

Eta zerrenda ordenatu horietatik:

$$Me_A = 35\ unitate$$

$$Me_B = \frac{38+39}{2} = 38.5$$

Medianaren arabera, B langilea da gehien ekoizten duena, 3.5 unitate gehiago zehatzago, lagin-errorearen erreserbapean (beste egun batzuetako egunak jaso izan balira, ondorioak ezberdinak izan ziratekeela kontuan harturik).

3.2.2.1 Kalkulua datuak tarteka bildurik daudenean

Datuak tarteka bildurik daudelarik, medianaren balioa hurbildu egin behar da.

Adibidea: Lantegi batean langileen adinak jaso dira:

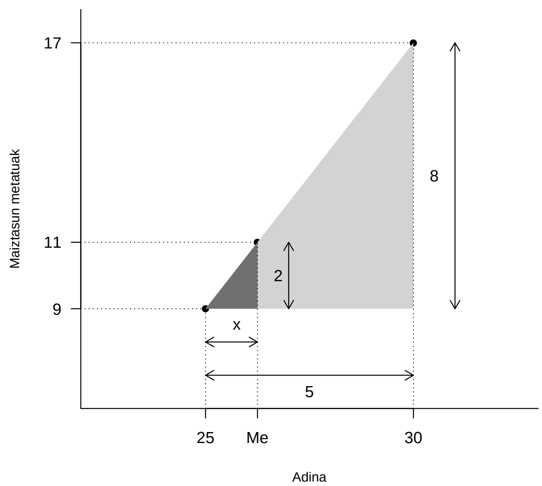
Adina	Langileak (n)
15-20	2
20-25	7
25-30	8
30-35	4
35-40	1
	22

Kalkulatu langileen adinen mediana.

Lehen pausoa datu-kopurua zati 2 egitea da: $\frac{22}{2} = 11$. Hortik mediana 11. datuaren balioa dela dakigu. Hurrengo pausoan, maiztasun metatu bakunak kalkulatzen joaten gara, 11 kopurua gainditu arte:

Adina	Langileak (n)	N
15-20	2	2
20-25	7	9
25-30	8	17
30-35	4	...
35-40	1	...
	22	

Maiztasun metatuetatik, mediana, 11garren datuaren balioa alegia, 25-30 tartean dagoela ondorioztatzen dugu. Tarte horretan, zein balio hartzen duen ez dakigu. Horretarako, jatorrizko datuak eskuratu beharko lirateke. Irtenbide bakarra balio hori hurbiltzea da. Hurbilketarako erabiltzen den metodoa interpolazio lineala da, tartean zehar datuak uniformeki banatzen direla suposatzen duena. Horrela, 11 9tik gertuago dagoenez (9 datu daude 25 baino txikiagoak) 17tik baino (17 datu daude 30 baino txikiagoak), 11. datuak 25etik gertuago beharko luke. Ikus dezagun prozesu osoa zehaztasunez modu grafikoan:



Irudia 3.1: Mediana hurbiltzeko interpolazio lineala.

Itzaldutako bi triangeluak antzekoak izanik, beraien aldeen arteko erlazioa berdina dela kontuan hartuz:

$$\frac{x}{2} = \frac{5}{8}$$

Hortik: $x = 1.25 \rightarrow Me = 25 + x = 26.25$

Beraz, medianaren arabera langileen adina 26.25 urtekoa da batezbestez.

3.2.3 Batezbestekoa eta mediana alderatuz: datu atipikoak eta sendotasuna

Estatistika-neurri bat *sendoa* dela esaten da, *datu atipikoek* neurriaren balioan eragin handirik ez dutenean. Neurri sendoak hobesten dira, sendoak ez diren neurrien balioa distortsionatua izaten baita, datu atipiko horien eraginez.

Adibidea: Herri batean bertako herritarren urteko errentak jaso dira (milaka eurotan):

20-20-20-20-160

Kalkulatu batezbesteko aritmetiko sinplea eta mediana eta emaitzak interpretatu sendotasunari buruz.

$$\overline{x} = \frac{20+20+20+20+160}{5} = 48$$
$$Me = 20$$

Harrigarria da zentroa neurtzea xede duten bi neurriok hain emaitza desberdinak izatea. Horren kausa datu-multzo nagusitik oso aparte dagoen *outlier* edo *datu atipikoa* da (160). Balio horrek gora bultzatzen du batezbestekoaren emaitza, eta batezbestekoa distortsionatu; hain zuzen, inork ez luke esango batez besteko errenta 40 mila eurokoa denik. Medianarekin ez da berdin gertatzen, eta zuzen ematen du zentroaren balioa, datu atipikoa barruan sartuta ere. Hartara, mediana sendoa dela dakusagu, eta batezbesteko aritmetiko sinplea ez. Beraz, datu atipikoak daudenean, hobe da mediana erabiltzea zentroa modu egokian neurtzearren.

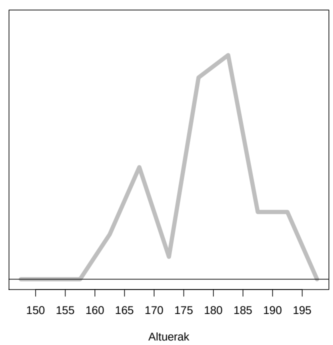
Sendoa izatea du abantaila medianak, baina oztopo bat ere badu: erdiko datuaren balio soila denez, ez du jasotzen datu-multzoan dagoen informazio guztia, ez ditu datu guztiak kontuan hartzen alegia. Batezbesteko aritmetiko sinplearen kasuan, berriz, alderantziz gertatzen da: ez da sendoa, baina datu guztiak hartzen ditu kontua.

Orokorrean, datu guztiak bildu eta kalkulatzen erraza eta interpretatzen sinplea delako, batezbesteko aritmetiko sinplea da gehien erabiltzen den zentro-neurria, eta soilik datu atipikoek bere emaitza distortsionatzen dutenean baztertuko dugu.

3.2.4 Moda

Moda (Mo) datuetan gehienetan errepikatzen den aldagaiaren balioa da, maiztasun handiena duena. Aldagai jarraituen kasuan, datuak tartetan biltzen direnean alegia, moda histograman edo maiztasun-poligonoan gailurra kokatzen den aldagaiaren balioa da. Ezaugarri berezi gisa, **aldagai kualitatiboetarako ere eman daiteke**, beste zentro neurriak ez bezala.

Datu-banaketa batzuetan badira, maiztasun handieneko *moda absolutuaz* gainera, *moda erlatiboak*, inguruko balioetan baino maiztasun handiagoa duten balioak alegia. Aurrerago ikasiko dugunez, moda anitz izateak, dela absolutuak dela erlatiboak, datuetan heterogeneotasuna egon daitekeela adierazten du.



Irudia 3.2: Hainbat gizonezkoren altueren maiztasun-poligonoa. Moda absolutua 182.5 cm balioan kokatzen da. Beste moda bat dago, erlatiboa, 167.5 cm balioan. Moda aniztasunak laginean bi jatorri etnikoetako gizonezkoak izatea adieraz dezake.

Adibidea: Tamaina bereko bi ospitaletan jasotako larrialdi-kopurua jaso da hainbat egunetan:

Larrialdiak (x)	Egunak (1. ospitalea)	Egunak (2. ospitalea)
0	1	2
1	4	5
2	6	3
3	3	4
4	-	1

Kalkulatu modak eta emaitzak interpretatu.

Lehen ospitalean:

$$Mo = 2 \; larrialdi$$

Bigarren ospitalean, berriz:

$$Mo = 1 \; larrialdi \; (absolutua); \; Mo = 3 \; larrialdi \; (erlatiboa)$$

Bigarren ospitalean moda erlatibo handi bat izateak ohiko larrialdi kopuruaz gainera ez ohiko larrialdi kopuru handia ekarri duen gertakaria izan dela, epidemia bat adibidez, adieraziko luke.

Bukatzeko, ohartarazi behar da, beste neurriak ez bezala, moda aldagai kualitatiboetarako ere kalkulatu daitekeela: gehien errepikatzen den kategoria da.

3.2.5 Magnitude desberdinetako balioen batezbestekoa: normalizazioa

Aurreko ataletan batezbestekoen aldagai jakin bati buruz elementu desberdinek hartzen duten balioen batezbestekoa nola kalkulatzen den ikusi dugu. Baina batezbestekoa elementu bakar batek aldagai desberdinetan hartzen dituen balioen batezbestekoa kalkulatzeko ere erabil daiteke, gehienetan adierazle sintetiko bat eratze aldera. Arazoa da balio horiek magnitude edo eskala desberdinetan jasota daudela eta beraz ez duela zentzurik batezbestekoa emateko horien batura kalkulatzea. Adibidez, ikasle baten aplikazio edo kalitate akademikoa neurtzeko, haren ikasketa-orduak astean (adibidez, 34 ordu) eta adimen koefizientea (110 puntuko kozientea) jasotzen baditugu, haiek nolabait bateraturik aplikazio horren balio bat lortzeko, ezin ditugu zuzenean batu gauza desberdinak jasotzen dituztelako (esan ohi den bezala, sagarren kopuru bat ezin da intxaurren kopuru batekin batu). Irtenbide gisa, **normalizazio** delako prozesua garatzen da, non elementuari dagozkion balio horiek multzoko balio maximoarekin eta minimoarekin alderatzen diren, neurri normalizatu, estandar edo homogenoak eskuratzeko, 0-100 bitarteko eskala batean:

$$n_i = \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

n_i balio normalizatu horien batezbestekoa kalkulatu baino ez dugu egin behar, aplikazio akademikoaren adierazle sintetiko bat eskuratzeko, horretarako batezbesteko aritmetiko sinplea, batezbesteko haztatua (balioren bati garrantzi handiagoa eman nahi bazaio, hurrengo atalean ikusiko den bezala) edo, hain zuzen kasu hauetan ohikoa den bezala, batezbesteko geometrikoa (hurrengo atalean ikusiko dena) kalkulatuz.

Gure adibideari helduz, gelan gehien ikasten duen ikasleak 40 ordu sartzen baditu astean eta gutxien ikasten duenak 2; eta beste alde batetik, gelako adimenkoziente handiena 120 eta txikiena 70 izanik, hauek lirateke balio normalizatuak:

$$n_{orduak} = \frac{34-2}{40-2} = 0.84 \ ; \ n_{adimena} = \frac{110-70}{120-70} = 0.8$$

Aplikazio akademikoaren A_i adierazlea kalkulatzeko, bi balio normalizatu horien batezbestekoa kalkulatzen dugu:

$$A_i = \frac{0.84 + 0.8}{2} = 0.82$$

3.2.6 Batezbesteko bereziak

3.2.6.1 Batezbesteko aritmetiko haztatua

Batezbesteko aritmetiko sinpleak haztapen edo pisu berdina ematen die datu guztiei batura kalkulatzean. Batzuetan, ordea, komeni da datuei neurri ezberdinetan haztatu edo ponderatzea; adibidez, irakasgai bateko zatikako azterketen kasuan esaterako, azterketa horien garrantzia edo edukiaren tamaina desberdina izan daiteke. Kasu horietan **batezbesteko aritmetiko haztatua** erabiltzen da.

Honela kalkulatzen da, w_i izanik, datuei eman beharreko haztapen-koefizienteak (ingelesez, $weights$):

$$\overline{x}_w = \frac{\sum_i w_i x_i}{\sum_i w_i}$$

Adibidea: Lanpostu bat betetzeko lehiaketako probetan hautagai batek puntuazio hauek lortu ditu: A proban, 6; B proban, 8; eta C proban, 9. Probei eman beharreko pisuak edo haztapenak 2, 4 eta 3 badira, hurrenez hurren, kalkulatu hautagaiaren batez besteko puntuazioa.

$$\overline{x}_w = \frac{2 \times 6 + 4 \times 8 + 3 \times 9}{2 + 4 + 3} = 7.88$$

Batezbesteko aritmetiko haztatua portzentajeen batezbestekoetarako ere erabiltzen da, portzentaje horiek total ezberdinei buruzkoak direnean.

Adibidea: Hego Euskal Herrian, langabezia tasak jaso dira lau herrialdeetarako: Araban, %8; Bizkaian, %12; Gipuzkoan %6 eta Nafarroan, %10. Lurralde horietako biztanleria aktiboak 100.000, 500.000, 300.000 eta 200.000 badira, hurrenez hurren. Hego Euskal Herriko langabezia-tasa orokorra kalkulatu behar da.

Portzentaje orokorra kalkulatzeko ezin da lurraldeetako portzentajeen batezbesteko aritmetiko sinplea erabili, portzentajeak biztanleria aktibo ezberdinei buruzkoak direlako; adibidez, Bizkaiko portzentajeari bertako biztanleriaren araberako pisu edo garrantzia eman behar zaio. Beraz, batezbesteko aritmetiko haztatua erabiliz egingo da kalkulua:

$$l = \frac{100 \times 8 + 500 \times 12 + 300 \times 6 + 200 \times 10}{100 + 500 + 300 + 200} = 9.63$$

Hego Euskal Herriko langabezia tasa %9.63 da, beraz, datu horien arabera.

Kalkulua beste era batera ere egin daiteke, batezbesteko haztatua erabiliz. Langabezia-tasa langabetu kopurua zati biztanleria aktiboa besterik ez da. Kalkula dezagun, bada, langabetu kopurua:

- Araban, $100.000 \times 0.08 = 8000$;
- Bizkaian, $500.000 \times 0.12 = 60.000$;
- Gipuzkoan, $300.000 \times 0.06 = 18.000$;
- Nafarroan, $200.000 \times 0.10 = 20.000$.

Beraz, guztira 106.000 langabetu daude Hego Euskal Herrian. Biztanleria aktiboa 1.100.000 lagunek osatzen dutenez, langabezia tasa hau izango da:

$$l = \frac{106.000}{1.100.000} = 0.0963 = \%9,63$$

Batezbesteko haztatuaren **beste aplikazio bat batezbestekoen batezbestekoa** kalkulatzea da. Izan ere, batezbesteko partzialak agregatu eta batezbesteko orokorra kalkulatu behar denean, batezbesteko partzialak haietako lagin-tamaina edo elementu kopuruaren arabera haztatu behar dira. Zehatzago $\overline{x}_1$, $\overline{x}_2$, ..., $\overline{x}_k$ batezbestekoak, n_1, n_2, ..., n_k lagin-tamainak dituztela, honela agregatzen dira, $\overline{\mathbf{x}}$ batezbesteko orokorra kalkulatzeko:

$$\overline{\mathbf{x}} = \frac{n_1\overline{x}_1 + n_2\overline{x}_2 + \cdots + n_k\overline{x}_k}{n_1 + n_2 + \cdots + n_k}$$

Adibidea: Eskola bateko maila batean hiru ikasgela daude, A, B eta C. A gelan 20 ikasle daude eta bertan batez besteko nota 6 izan da; B gelan 25 ikasle daude eta batez besteko nota 5.5 izan da, azkenik, C gelan 15 ikasle daude eta batez besteko nota 8 izan da. Kalkulatu maila osoko batez besteko nota.

$$\overline{\mathbf{n}} = \frac{20 \times 6 + 25 \times 5.5 + 15 \times 8}{20 + 25 + 15} = 6.29$$

3.2.6.2 Batezbesteko kuadratikoa

Batezbesteko kuadratikoa erroreen batezbestekoa kalkulatzeko erabiltzen da. Errore-datuak positiboak zein negatiboak izaten dira, eta horietarako batezbesteko aritmetiko sinplea erabiltzen bada, konpentsazio-efektu batengatik batezbestekoa benetakoa baino txikiagoa suertatuko da. Batezbesteko kuadratikoan datuak ber bi egiten dira, aurreko oztopoa gerta ez dadin.

$$K = \sqrt{\frac{\sum_i x_i^2}{n}}$$

Adibidea: Flasko baten eduki nominal edo estandarra 100 ml izan behar da. Lau flaskoko edukia neurtu ondoren, bolumen hauek jaso dira: 98-96-101-103. Flaskoa betetzean sortzen den batez besteko errorea kalkulatu.

Nominalaren aldean izandako erroreak kalkulatu behar dira lehenbizi, ($eduki\ erreala - eduki\ estandarra$) kalkulatuz: -2,-4,1,3.

Errore horietarako batezbesteko aritmetiko sinplea erabiliko balitz, batez besteko errorea (-2-4+1+3)/4=-0.5 ml izango litzateke. Emaitza horrek, ordea, gutxietsi egiten du errore erreala batuketan erroreak konpentsatu egiten baitira, eta flaskoak gutxiegiz betetzen direla ondorioztatzeko bakarrik balio du.

Batezbesteko kuadratikoa baliatuz:

$$K = \sqrt{\frac{-2^2 + (-4)^2 + 1^2 + 3^2}{4}} = 2.73\ ml$$

3.2.6.3 Batezbesteko geometrikoa

Batezbesteko geometrikoa batez besteko hazkunde-tasak eta interesak kalkulatzeko erabiltzen da. Adibide batez ikasiko dugu zergatik ez den egokia kasu horietan batezbesteko aritmetiko sinplea eta nola eratortzen den batezbesteko geometrikoaren formula.

Adibidea: 1000€ko kapitala inbertitu da bi urtetan, lehenengoan %10ko interes-tasan, eta %20koan bigarrenean. Zenbat da urteko batez besteko interes tasa?

Kalkula dezagun, zenbatekoa den bukaerako kapitala bi urteren buruan:

$$C_2 = 1000(1 + 0.1)(1 + 0.2) = 1320€$$

Batezbesteko aritmetikoa erabiltzen bada, batez besteko interesa %15 da. Balio hori zuzena balitz, bi urteetan interes-tasa konstante hori edukita, bukaerako kapitala berdina izan beharko litzateke. Ikus dezagun horrela den:

$$C_2 = 1000(1 + 0.15)^2 = 1322.5€$$

Alegiazko bukaerako kapitala benetakoa baino handiagoa da, eta beraz %15eko batez besteko balioa ez zuzena; zehatzago, behar baio handiagoa da, bukaerako kapital handiagoa ematen duelako.

Bukaerako kapital berdina ematen duen i batez besteko interes-tasa konstantea honela kalkulatu behar da:

$$1000(1+i)^2 = 1320 \rightarrow i = \left(\frac{1320}{1000}\right)^{\frac{1}{2}} - 1 = 0.1489 = \%14.89$$

Honela ere kalkula daiteke:

$$i = (1.10 \times 1.20)^{\frac{1}{2}} - 1 = 0.1489$$

Oro har, batezbestekoak era horretan kalkulatzeko moduari *batezbesteko geometrikoa* deritzo:

$$G = (x_1 \times x_2 \times \ldots \times x_n)^{\frac{1}{n}}$$

Formula hori aplikatzerakoan, interes- eta hazkunde-tasak +1 egin behar direla hartu behar da kontuan, eta bukaeran 1 kendu.

3.2.6.4 Batezbesteko harmonikoa

Batezbesteko harmonikoa batez besteko abiadurak kalkulatzeko erabiltzen da. Enpresa arloa interesatzen zaigu eta horretan batez besteko errendimenduak eta produktibitateak kalkulatzeko erabiltzen da. Adibide batez azalduko dugu kasu horietan batezbesteko aritmetiko sinplearen desegokitasuna:

Adibidea: Pertsona batek kilometro bana egiten du 10 km/h eta 6 km/h-ko abiaduran. Zenbat da batez besteko abiadura?

Batezbesteko aritmetiko sinplea erabiliz, erantzuna (6+10)/2=8 km/h litzateke. Baina emaitza hori zuzena izango bada, ibilbidea egiteko denbora berdina beharko lizateke 10 km/h eta 6 km/h-ko abiaduran nahiz 8 km/h-ko abiadura konstantean egitea. Ikus dezagun horrela den:

- bi kilometroak 10 km/h eta 6 km/h-ko abiaduran hurrenez hurren eginez, 6+10=16 min behar dira;
- 8 km/h-ko abiadura konstantean, bi kilometroak egiteko 15 min behar dira.

Beraz, argi dago batez besteko abiadura 8 km/h baino pixka bat motelagoa izan behar dela, denbora 15 minututik 16 minutura pasatzeko.

Batez besteko abiadura eta errendimenduak kalkulatzeko metodo laburrena abiadura eta errendimenduen oinarrizko definiziora jotzea da:

- $v = \frac{s}{t}$
- $e = \frac{ekoizpena}{t}$

Aurreko adibidean honela kalkulatuko genuke batezbestekoa, beraz:

$$v = \frac{2km}{16min} = \frac{2km}{16/60h} = 7.5km/h$$

Emaitza berdina ematen du batezbesteko harmonikoaren formulak:

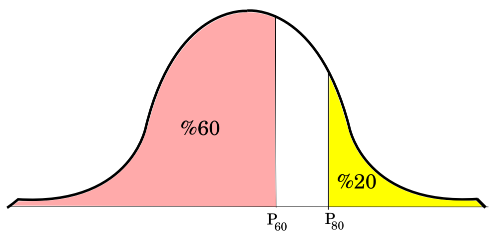
$$H = \frac{n}{\sum_i \frac{1}{x_i}}$$

Adibidera aplikatuz, ikerketa-unitateak egindako kilometroak direla kontuan hartuz:

$$H = \frac{2}{\frac{1}{10} + \frac{1}{6}} = 7.5$$

3.3 Kuantilak

Zentroaz gainera, datu-multzo batean egon daitezke beste kokapen edo posizio-neurri batzuk (egin klik). Horien artean, kuantilek (egin klik) datu baten azpitik portzentajean zenbat dauden daude adierazten duten balio dira. Adibidez, ikasleen kalifikazio-datuak ditugu. Ikasleen %60ak gainditu gabe uzteko, zein kalifikazio jarri behar da? Azpitik ikasleen %60ak uzten dituen kalifikazioa behar da. Estatistikan, 60garren pertzentila deritzo balio horri, eta P_{60} adierazten da. Eta ikasleen %20k gainditzeko, zein kalifikazio jarri behar da, berriz? Azpitik ikasleen %80ak uzten dituen kalifikazioa behar da, hau da, P_{80} edo 80. pertzentila.



3.3.1 Kuantilen sailkapena

Badira kuantilak izen berezia dutenak:

- **Pertzentilak**: $P_1,\ P_2, \ldots, P_{99}$
- **Dezilak**: $D_1 = P_{10},\ D_2 = P_{20}, \ldots, D_9 = P_{90}$
  Adibidez, lehen dezila azpitik hamar ikasleetatik ("dezil") ikasle bat ("lehen"dezila delako) uzten duen kalifikazioa da, hau da, azpitik ikasleen %10ak uzten dituena.
- **Oktilak**: $O_1 = P_{12.5},\ O_2 = P_{25}, \ldots, O_7 = P_{87.5}$
  Adibidez, lehen oktila azpitik zortzi ikasleetatik ("oktil") ikasle bat ("lehen"oktila delako) uzten duen kalifikazioa da, hau da, azpitik ikasleen 1/8=%12.5ak uzten dituena.
- **Kintilak**: $K_1 = P_{20},\ K_2 = P_{40},\ K_3 = P_{60},\ K_4 = P_{80}$
- **Kuartilak**: $Q_1 = P_{25},\ Q_2 = P_{50},\ Q_3 = P_{75}$
- **Mediana**: $Me = Q_2 = O_4 = D_5 = P_{50}$
  Hain zuzen, medianak datuen erdiak, %50ak alegia, uzten ditu azpitik.

3.3.2 Kuantilen kalkulua

Datuak banaka emanda daudenean, kuantilak kalkulatzeko pausoak hauek dira:

1. datuak txikienetik handienera ordenatzea;

2. datu-kopurua kuantilak adierazten duen ehunekoarekin bidertu, kuantila ematen duen datua zenbatgarrena den jakiteko;

3. kuantila ematen duen datuaren ordena zenbakia osoa bada, kuantila datu hori bera da; bestela, interpolazioz kalkulatzen da ondoko bi datuak harturik, d izanik eskuratutako ordenaren zati dezimala:

$$k = aurreko\ datua\ \times (1-d) + hurrengo\ datua\ \times d$$

Ohartarazi behar dugu aurreko formula kuantilak kalkulatzeko asmatu diren hainbat formuletako bat baino ez dela, guztiek antzeko emaitzak ematen badituzte ere.

Adibidea: Ikasle batzuek izandako kalifikazioak jasotzen dira ondoren:

$$5.6 - 7.8 - 8.3 - 9.2 - 6.9 - 4.2 - 7.4 - 7.2$$

Kalkulatu 1. kuartila eta 9. dezila.

Datuak ordenatuko dira lehenbizi:

$$4.2 - 5.6 - 6.9 - 7.2 - 7.4 - 7.8 - 8.3 - 9.2$$

1. kuartila (25. pertzentila) $0.25 \times 8 = 2$garren datua da: $Q_1 = P_{25} = 5.6$.

9. dezila (90. pertzentila) $0.9 \times 8 = 7.2$garren datua da. Alegiazko datu hori interpolazioz hurbiltzen da, 7garren eta 8garren datuen artean:

$$P_{90} = 8.3 \times (1-0.2) + 9.2 \times 0.2 = 8.48$$

3.3.2.1 Kuantilen kalkulua tartekako datuetan

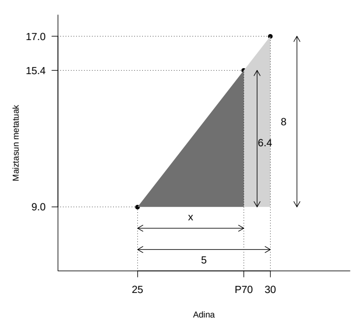
Kuantilarekin bat datorren datua zenbatgarren den jakin ondoren, datua hori kokatzen den tartea bilatu eta interpolatu behar da.

Adibidea: Lantegi batean langileen adinak jaso dira:

Adina	Langileak (n)
15-20	2
20-25	7
25-30	8
30-35	4
35-40	1
	22

Langileen %30 zaharrenari oporrak hartzeko aukera eman behar zaio. Zenbat urtetik aurrera izango da hori?

P_{70} kalkulatu behar da erantzuna emateko. Horretarako $0.7 \times 22 = 15.4$-garren datuaren balioa kalkulatu behar da. Maiztasun absolutuak metatuz (2-9-17- ...) datu hori hirugarren tartean dagoela, 25-30 tartean alegia, ikusten dugu. Datuaren balio zehatza interpolazioz bilatzen dugu: Irudian oinarritua, antzekoak



Irudia 3.3: P_{70} hurbiltzeko interpolazio lineala.

diren bi triangeluen arteko kateto-erlazioak berdinduz:

$$\frac{x}{6.4} = \frac{5}{8}$$

Hortik: $x = 4 \rightarrow P_{70} = 25 + 4 = 29$. Beraz, 29 urtetik aurrera langileen %30ak daudela estimatzen da, eta orduan adin horretatik aurrera eskainiko dira oporrak.

3.3.3 Kuantilen aplikazio bat: winsorizazioa

Jakina denez, datu atipikoak daudenean distortsionaturiko emaitza estatistikoak eskuratzeko arriskua dago. Datu atipikoen eragina saihesteko, estatistiko edo neurri sendoak erabil ditzakegu, mediana kasu, zentralizazio-neurri gisa.

Bestelako aukerak badaude, eta horietako bat da winsorizazio izenekoa (Charles P. Winsor estatistikaria -1895â1951- izan zen, hain zuzen, aukera hori planteatu zuena). Winsorizazioan goiko eta beheko muturretan dauden datuak, orokorrean kuantil baten bitartez definiturik, mutur horiek mugatzen dituzten balioekin ordezten dira. Horrela, datu atipikoak datu normalekin ordezturik, sendoak ez diren estatistikoak ere, batezbestekoa kasu, erabil daitezke, zentroa edo beste ezaugarri batzuk aztertzeko.

3.4 Ariketak

1. Industria-prozesu berbera egiten duten bi makinek egunean zehar duten geldialdi-kopurua jaso da egun batzuetan. Hona hemen emaitzak:

Geldialdiak	Egunak (A makina)	Egunak (B makina)
0	2	1
1	5	2
2	8	6
3	5	4
4	1	3
5	0	0
6	0	2
	21	18

Kalkulatu eguneko geldialdi kopuruaren batezbesteko aritmetiko sinplea eta mediana, datu atipikoen eragina azaldu eta emaitzak interpretatu.

2. 2005 eta 2015 urteetan herri bateko urteko familia-errentak jaso dira, 1000 euroko errenta minimo batekin. Honako hauek dira datuak tartetan bilduta:

Errenta	Familiak (2005)	Familiak (2015)
$<$ 5.000	8	12
5.000-10.000	14	10
10.000-20.000	22	26
20.000-30.000	6	4
30.000-50.000	2	2
	52	54

Kalkulatu familia bakoitzeko errentaren batezbesteko aritmetiko sinplea eta mediana, eta emaitzak interpretatu.

3. Ondoren hiri bateko bizilagunen adinari buruzko datuak zehazten dira, auzoaren arabera. Bost auzo handietan banatu da hiria. (a) Auzo bakoitzeko batez besteko adina eta adinen mediana kalkulatu (b) Hiri osoko batez besteko adina eta mediana kalkulatu, A, B, C, D eta E auzoetan biztanleriaren %20, %10, %20, %20 eta %30 hurrenez hurren bizi direla jakinik (c) Hiriko auzo demografikoki zaharkituak zeintzuk diren zehaztu.

Adina	A	B	C	D	E
0-20	%18	%16	%18	%16	%24
20-40	%28	%20	%24	%16	%30
40-60	%25	%22	%27	%24	%20
60-80	%19	%20	%21	%20	%14
80-100	%10	%22	%10	%24	%12
	%100	%100	%100	%100	%100

4. Enpresaren batean irizten da langilearen motibazioa islatzen duten hiru faktore nagusiak hilean burututako ekoizpena eta sarturiko ordu kopurua eta hobekuntza-ikastaroetan azken urtean emandako egun kopurua direla. Ana Basurko langileak azken hilean 400 pieza burutu ditu, 120 ordu sartu ditu eta 12 egun eman ditu trebatzen azken urtean. Lantegian ekoizpen maximoa 440 pieza izan da langileen artean, eta 180 lan ordu eta 22 trebakuntza egun maximoak. Horien kopuru minimoak 300, 80 eta 4 izan dira. Eman ezazu Ana Basurkoren motibazio-adierazle normalizatu bat.

5. 10 egunetan zehar, eguneko batez besteko ekoizpena 80 unitatekoa izan zen. Beste 20 egunetan, 100 unitatekoa izan zen.
(a) Zenbat da 30 egun horietako eguneko batez besteko ekoizpena? (Soluzioa: 93.3)
(b) 60 egunetan batez besteko ekoizpena 110era heltzea nahi bada, zenbat izan behar da azken 30 egun horietako eguneko batez besteko ekoizpena?
(c) Zenbat izan behar da 30 egun horietako ekoizpen totala?

6. Irakasgai bat gainditu dutenen portzentajeak ditugu ikasgelaren arabera, ikasle kopuruarekin batera:

Ikasgela	Ikasle kopurua	Gainditu portzentajea
A	15	%60
B	25	%80
C	10	%30

Kalkulatu gaindituen baterako portzentajea hiru ikasgeletan,
(a) portzentajeak haztatuz,
(b) gaindituen kopuru totaletik abiatuz.

7. Okindegi batek 500gr inguruko opilak egiten ditu, besteak beste. Pisatzeprozesua hobetzeko saioa egin du, eta arrakastatsua izan den jakiteko, opil zenbaiten pisuak jaso ditu pisaketa zaharrarekin nahiz berriarekin:

Aurretik	Ondoren
554	536
488	541
532	480
474	523
491	-

(a) Pisaketa aldatuta, 500 gr-ko pisu estandarrarekiko aldea batez beste hobetu den (eta baiezkoan, zenbat) kalkulatu errorearen batezbesteko koadratikoa nahiz errore absolutuen batezbestekoaren bitartez. Zein da neurri egokiena: batezbesteko kuadratiko edo balio absoutuen batezbestekoa? (b) Aurretik nahiz ondoren, opilak batez beste behar baino gutxiago edo gehiago pisatzen duen aztertu.

8. Enpresa bateko salmentak zehazten dira ondoren, urtez urte (milioka euro):

Urtea	Salmentak
2010	0.8
2011	1.4
2012	2.3
2013	3.2
2014	4.8

Salmenten urteko batez besteko hazkunde-tasa kalkulatu behar da. *(Erantzuna: Bitarteko hazkunde-tasak %75, %64, %39 eta %50 dira, eta batez bestekoa %56.40.*

9. Kapital bat 10 urtetan zehar bikoizteko, zenbat izan behar da urteko batez besteko errentagarritasun-tasa? *(Erantzuna: %7.17)* Eta 5 urtetan, kapitala %50 gehitzeko? *(Erantzuna: %8.44)*

10. Enpresa batek hiru langile ditu: A, B eta C. A langileak 12 pieza egiten ditu orduko, baina egunean soilik 4 orduz egiten du lana. B langileak 10 pieza egiten ditu orduko, eta egunean 6 orduz egiten du lana. C langileak, azkenik, 6 pieza egiten ditu orduko eta 8 orduz lanean. Hilabete batean, A, B eta C langileek 22, 20 eta 18 egunez egin dute lan, hurrenez hurren. Zenbat da hilabete horretan orduko errendimendua? *(Erantzuna: 3120/352=8.86 pieza/h)*

11. Eguneko salmentak jaso dira denda batean 26 larunbatetan (eurotan):

$$3.020 - 2.695 - 2.324 - 2.835 - 1.719 - 786 - 2.669$$

$$2.801 - 3.148 - 3.661 - 2.932 - 5.012 - 4.567$$

$$4.008 - 3.246 - 3.361 - 3.742 - 4.002 - 2.633 - 2.607$$

$$3.495 - 3.839 - 2.402 - 1.452 - 1.974 - 2.836$$

Salmenta txikieneko larunbaten %10etan, promozio-kanpaina bati ekiten zaio hurrengo asterako. Zenbat izan behar dira horretarako salmentak? *(Erantzuna: 1612,2 euro)*

12. Adimen test batean jasotako datuak taula honetan bildu dira:

Puntuazioa	Portzentajea
60-70	%4
70-80	%16
80-90	%30
90-100	%32
100-110	%10
110-120	%8

Puntuazio altuena duten pertsonen %15ei proba bereziak egingo zaizkie. Zein puntuazio behar da horretarako?

13. 20, 40, 60 eta 80 urteko pertsonen artean, azken aldiz ospitalera joan zirenetik pasa den denbora (hilabetetan) jaso da:

Denbora (hilabeteak)	20 urte	40 urte	60 urte	80 urte
0-20	8	14	25	43
20-40	16	18	34	31
40-60	34	45	42	22
60-80	45	37	36	16
80-120	42	26	22	8
>120	89	32	18	2
	234	172	177	122

Adin bakoitzean ospitalera joan gabe denbora gehien pasatzen duten pertsonen %75ari, azterketa berezia egingo zaie, horien bizimodu osasuntsuari buruzko informazioa jasotzeko.

(a) Ospitalera joan gabe zenbat denbora pasa behar da horretarako adin bakoitzean? Emaitzak azaldu eta interpretatu. *(Soluzioa: 20 urte, 60.22 hil- 40 urte, 44.8 hil - 60 urte, 31.3 hil - 80 urte, 14.1 hil)*

(b) 70 urteko pertsonen artean ere egin behar da bizimoduari buruzko ikerketa. Horien artean, zeintzuei egingo zaie azterketa? *(Soluzioa: 70 urte, 22.7 hil)* Eta 72 urtekoen artean egin behar bada? *(Soluzioa: 72 urte, 20.68 hil)*

14. 1995ean Jaizkibelgo hego maldetan landatutako pinudi bateko pinu bizien enborreko perimetroa (cm-tan) jaso da, 150cm-ko altueran. Hona hemen jasotako datuak:

Perimetroa	Pinuak
100-110	32
110-120	49
120-130	65
130-140	36
140-150	16
150-160	0
160-170	1
170-180	1
	200

Egin beharreko atazak:

(a) Kalkulatu lehen eta laugarren kintilak.

(b) Kintil horietan moztutako batezbestekoa kalkulatu.

(c) Kintil horietatik alde banatara, winsorizazio izeneko prozesua garatu, maiztasun-taula berria ezarri eta horri dagokion batezbesteko aritmetikoa kalkulatu.

(d) (i) Iritzira, zeintzuk lirateke datu atipikoak hasierako maiztasun-taulan? (ii) Datu atipiko horien eragina nabarmena al da, batezbestekoaren kalkuluan? Beraz, datu atipikotzat har al daitezke?

1. ariketa

Geldialdi kopuruaren batezbestekoa → aldagaia (x):

x	n_A	n_B	$n_A x$	$n_B x$	N_A	N_B

$\overline{x}_A = \frac{\quad}{\quad} =$

$\overline{x}_B = \frac{\quad}{\quad} =$

Interpretazioa: Batezbesteko aritmetiko sinplearen arabera,

$Me_A =$

$Me_B =$

Interpretazioa: Medianaren arabera,

2. ariketa

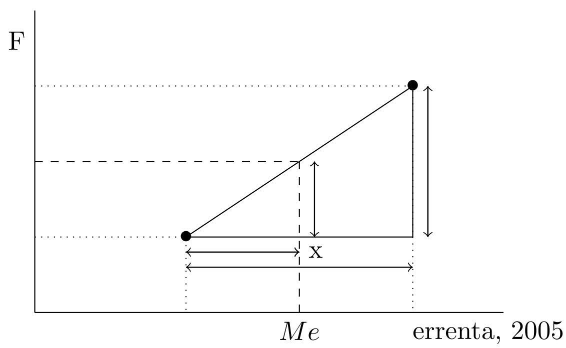
x	n_{2005}	n_{2015}	$n_{2005}x$	$n_{2015}x$	N_{2005}	N_{2015}

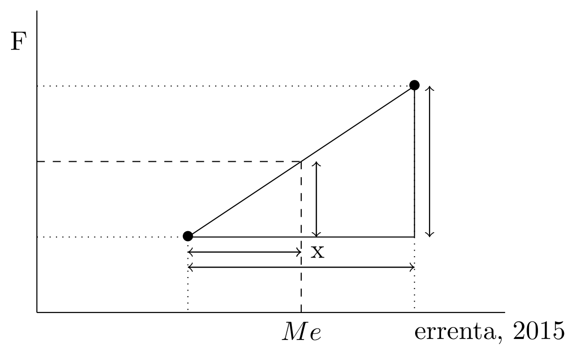
$\overline{x}_{2005} = \frac{}{} =$

Interpretazioa: Batezbestekoaren arabera,

$\overline{x}_{2015} = \frac{}{} =$

Medianak kalkula ditzagun orain:





$$\frac{x}{} = \frac{}{} \rightarrow x = \qquad\qquad \rightarrow Me =$$

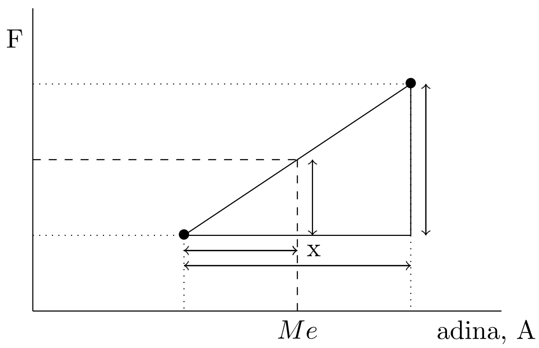
$$\frac{x}{} = \frac{}{} \rightarrow x = \qquad\qquad \rightarrow Me =$$

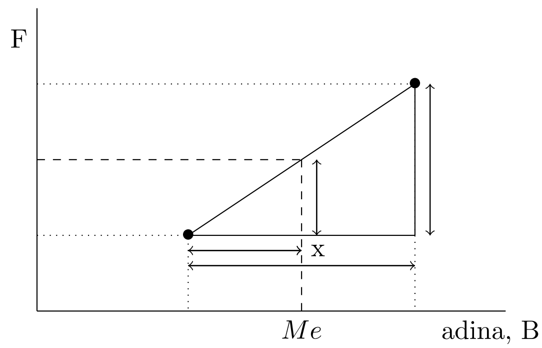
Interpretazioa: Medianaren arabera,

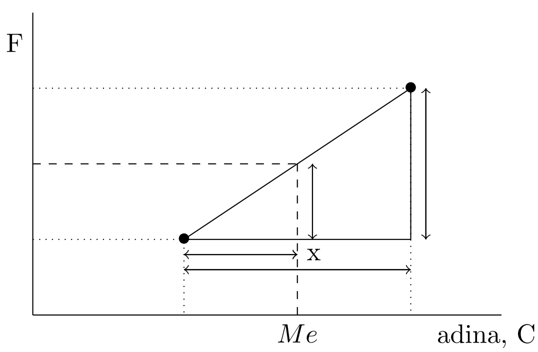
3. ariketa

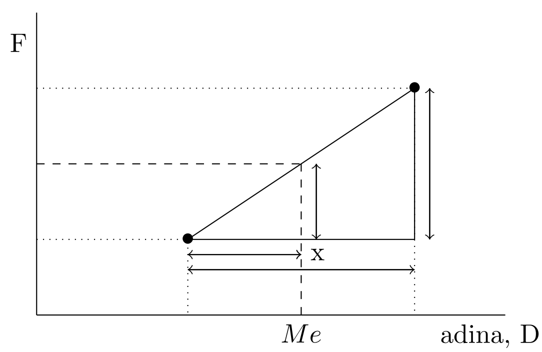
$\boxed{\text{(a)}}$

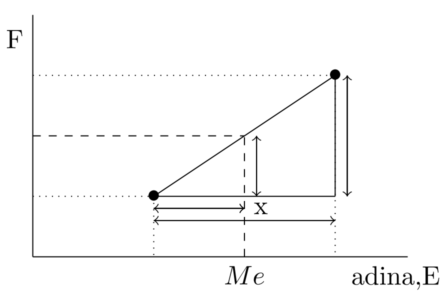
x	f_A	f_B	f_C	f_D	f_E	xf_A	xf_B	xf_C	xf_D	xf_E
	0.18	0.16	0.18	0.16		1.8				
	0.28	0.20	0.24	0.16		8.4				
	0.25	0.22	0.27	0.24		12.5				
	0.19	0.20	0.21	0.20		13.3				
	0.10	0.22	0.10	0.24		9				
	1	1	1	1	1	45				











$\boxed{A}$ $\frac{x}{\quad} = \frac{\quad}{\quad} \rightarrow x = \qquad \rightarrow$

$\boxed{B}$ $\frac{x}{\quad} = \frac{\quad}{\quad} \rightarrow x = \qquad \rightarrow$

$\boxed{C}$ $\frac{x}{\quad} = \frac{\quad}{\quad} \rightarrow x = \qquad \rightarrow$

$\boxed{D}$ $\frac{x}{\quad} = \frac{\quad}{\quad} \rightarrow x = \qquad \rightarrow$

$\boxed{E}$ $\frac{x}{\quad} = \frac{\quad}{\quad} \rightarrow x = \qquad \rightarrow$

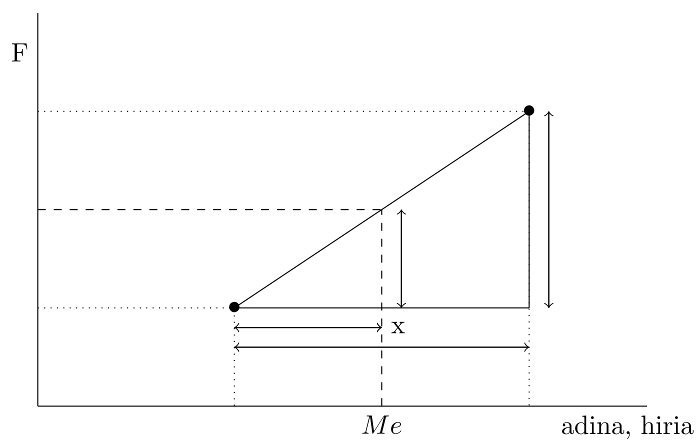
(b)

Hiri osoko adin-banaketa ez daukagu, baina enuntziatuan ematen den informaziotik atera daiteke. Nola? Adibide bat: ikasgela batean, nesken artean %70 dira aprobatuak, eta mutilen artean %40. Neskak %50 dira, eta mutilak %50. Zenbat aprobatu dira ikasgela osoan? $0.5 \times 70 + 0.5 \times 0.4 = \%55$. Hemen berdin egiten dugu: ikasgelan aprobatuen ordez, 0-20 bitartekoen portzentajea kalkulatu nahi da hiri osoan; ikasgela sexuaren arabera egin ordez, auzoaren araberako banaketa dugu hirian, eta sexu bakoitzeko aprobatuen ordez, auzo bakoitzeko 0-20 tartekoen portzentajeak.

$f(0-20) = 0.2 \times 18 + 0.1 \times 16 + 0.2 \times 18 + 0.2 \times 16 + 0.3 \times 24 = 19.2$

$f(20-40) =$
$f(40-60) =$
$f(60-80) =$
$f(80-100) =$

Tartea	f(Hiria)	x	xf
0-20	19.2		
20-40	24.6		
40-60	23.4		
60-80	18.2		
80-100	14.6		



Neurria (↓)	A	B	C	D	E	Hiria
Batezbestekoa						
Mediana						

Auzo zaharkituak, batezbestekoaren arabera:

Auzo zaharkituak, medianaren arabera:

4. ariketa

Aldagaiak	Ana Bas.	Min	Max	Ana Bas. normalizatua
Piezak	400	440	300	(400-300)/(440-300)= 0.71
Lan orduak	120			
Trebakuntza egunak	12			

$$Motibazio\text{–}adierazlea(AnaBas.) = \underline{\hspace{6cm}} =$$

Interpretazioa:

---

5. ariketa

(a)

Batezbesteko partzialak agregatu nahi ditugu. Beraz, batezbesteko haztatu erabili behar da batezbesteko orokorra kalkulatzeko:

$$\overline{\mathbf{x}}_{\mathbf{orokorra}}(\mathbf{30}) = \underline{\hspace{8cm}} = 93.3$$

(b) Hemen batezbesteko orokorra ematen digute eta batezbesteko partzial bat da ezezaguna:

$$\overline{\mathbf{x}}_{\mathbf{orokorra}}(\mathbf{60}) = 110 = \underline{\hspace{8cm}}$$

$$\longrightarrow \overline{x}_3(30) =$$

(c) 30 egun horietako ekoizpen totala egun horietako batezbestekoa bider egun kopurua da ($\sum x_i = n\overline{x}$):

6. ariketa

(a)

Ikasgela	w_i (haztapenak)	x_i (portzentajeak)	$w_i x_i$
A	15	%60	
B	25	%80	
C	10	%30	

$$\overline{p} = \frac{\quad}{\quad} =$$

(b)

Ikasgela	ikasle kop. totalak	gaindituen portzentajeak	gaindituak
A	15	%60	
B	25	%80	
C	10	%30	

$$\overline{p} = \frac{\quad}{\quad} = 0.64 = \%64$$

7. ariketa

errorea = neurri erreala - neurri estandarra

e_z (zaharra)	e_b (berria)	e_z^2	e_b^2	$\|e_z\|$	$\|e_b\|$
54	36				
-12	41				
32	-20				
-26	23				
-9	-				

Batezbesteko koadratikoaz ($K = \sqrt{\frac{\sum_i x_i^2}{n}}$):

$$\overline{E}_{zaharra} = \sqrt{\frac{\qquad}{\qquad}} = \qquad ; \overline{E}_{berria} = \sqrt{\frac{\qquad}{\qquad}} =$$

Balio absolutuez:

$$\overline{E}_{aurretik} = \frac{\qquad}{\qquad} = \qquad ; \overline{E}_{ondoren} = \frac{\qquad}{\qquad} =$$

Ondorioa da egindako saioak hobetu duela/ez duela hobetu pisatze-prozesua.

Metodo egokiena estatistikan batezbesteko kuadratikoa, balio absolutua matematikoki eta estatistikoki lantzen zaila izaten delako.

(b) Aurreko atalean errorearen batez besteko neurria eman dugu, baino jakin gabe orokorrean norako gertatzen den, gehiegiz edo gutxiegiz den. Hori jakiteko, aski da pisuen (edo erroreen) batezbesteko aritmetikoa kalkulatzea:

$$\overline{P}_{zaharra} = \frac{\qquad}{\qquad} =$$

$$\overline{P}_{berria} = \frac{\qquad}{\qquad} =$$

Pisaketa zaharrarekin nahiz berriarekin, orokorrean opil handiegiak (>500gr) egiten direla ikus daiteke.

8. ariketa

Urtea	Salmentak (y_i)	$x_i = \Delta\ y_i$
2010	0.8	-
2011	1.4	(1.4/0.8)-1=0.75
2012	2.3	
2013	3.2	
2014	4.8	

Batezbesteko geometrikoaz:

$\overline{h} =$

Hazkundearen formulaz, berriz:

---

9. ariketa

(a) Problema hau batezbesteko geometrikoaren bitartez ebaz daiteke, baina askoz hobeto ulertzen da, finantza matematikako ikuspuntutik, kapitalizazio-lege konposatua erabiliz:

$$C_n = C_0(1+i)^n \rightarrow (1+i) = (C_n/C_0)^{1/n}$$

Aise ohartzen gara (1+i) balioak baduela antza batezbesteko geometrikoaren formularekin. Finean, kapitalizazio-legea eta batezbesteko geometrikoaren formula guztiz loturik daude, batetik bestea eratortzen da alegia.

Kapitala bikoiztu nahi dugunez $Cn/Co =$ izango dugu. Beraz:

$$1+i = \qquad \rightarrow i = \qquad = \%$$

(b) %50 gehitzeko $Cn/Co =$ izango dugu.Beraz,

$$1+i = \qquad \rightarrow i = \qquad = \%$$

10. ariketa

Bi eratara kalkula daiteke: (a) batez besteko errendimenduaren kontzeptuari helduz nahiz (b) batezbesteko harmonikoaren formula baliatuz:

$\boxed{\text{(a)}}$ Batez besteko errendimendua ekoizpen totala zati lanean emandako ordu kopurua da:

$$\overline{E} = \frac{ekoizpen\ totala}{ordu\ kopurua} =$$

$$\frac{12 \times 4 \times 22 + \qquad\qquad}{4 \times 22 + \qquad\qquad} = \underline{\qquad\qquad} =$$

$\boxed{\text{(b)}}$ Formula erabilita zailagoa da. Kasu honetan, pieza guztiei datu bana dagokiela pentsatu behar da, eta beraz, arestiko kalkuluak harturik, n datu kopurua 3120 izango da. Hortik aurrera, errendimendu bakoitzeko zenbat datu edo pieza dauden kalkulatu behar da eta horren arabera formula aplikatu:

$$\overline{E} = H = \frac{n}{1/x_1 + 1/x_2 + \cdots + 1/x_n} =$$

$$\frac{3120}{\qquad\qquad\qquad\qquad} =$$

$$\frac{3120}{\qquad\qquad\qquad\qquad} =$$

Argi dago emaitza bi era horietara berdina izan behar dela.

11. ariketa

Zein pertzentil kalkulatu behar da?

Beraz, 26× → *gn. eta* *gn.* datuen artean egongo da.

Datuak ordenatzen ditugu, datu horiek eskuratu arte:

Eta orain bien arteko interpolazioa egiten dugu:

$$P_{.....} = \quad gn.\ datua\times \quad + \quad gn.\ datua\times \quad = ...$$

---

12. ariketa

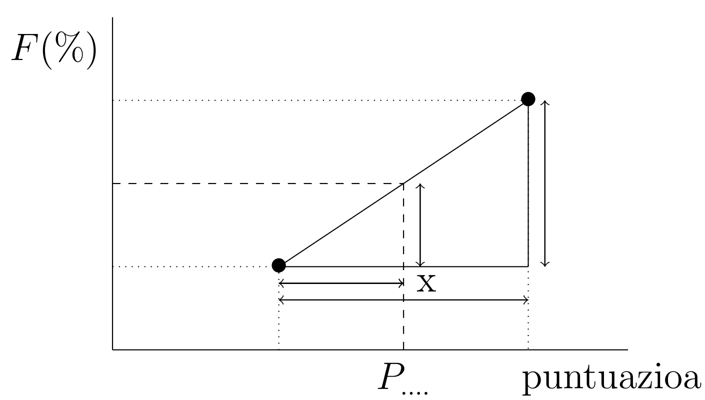
Zein pertzentil kalkulatu behar da? $P_{.....}$

Orain, datuak ez daude mailakatuta era honetan: 1.goa, 2gn.a, ...

Baina aise gainditzen dugu oztopo hori: portzentajeak zuzenean hartzen ditugu:

$$P_{.....} = \quad gn.\ datua \in$$

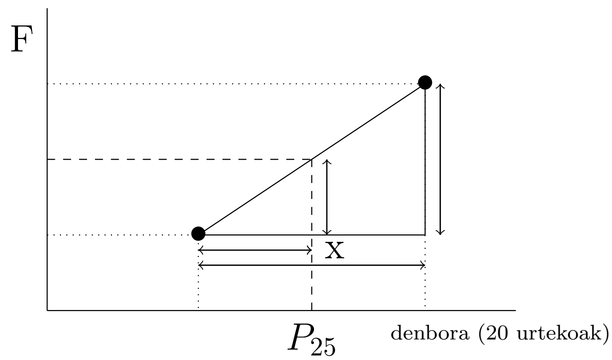
Eta interpolazioa garatzen dugu, baina maiztasun absolutuen ordez, maiztasun erlatiboak edo portzentajeak hartuz:

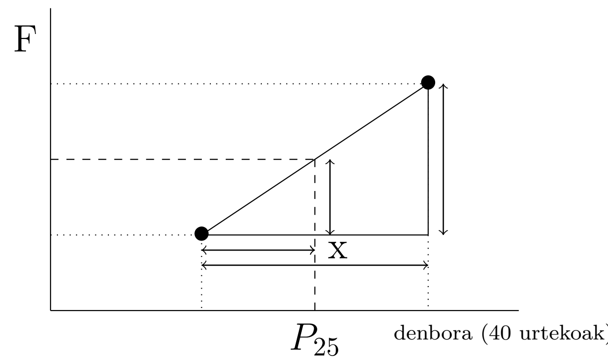


13. ariketa

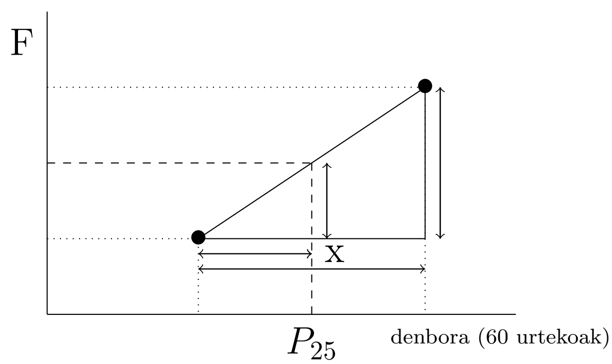
Zein pertzentil kalkulatu behar da? $P_{.....}$

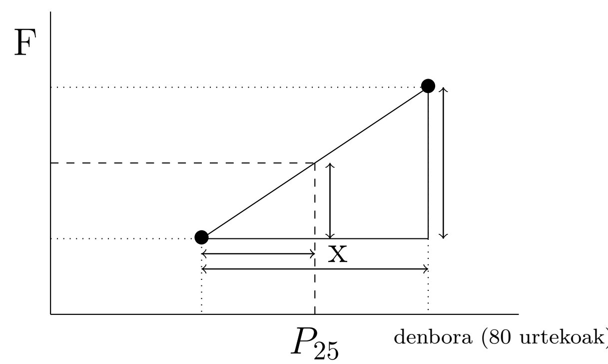
Datuak tarteka daudenez, pertzentilaren kalkulua interpolazioz egin behar da adin bakoitzeko:





$$\frac{x}{} = \frac{}{} \rightarrow x = \qquad \rightarrow P_{25} =$$





$$\frac{x}{} = \frac{}{} \rightarrow x = \qquad \rightarrow P_{25} =$$

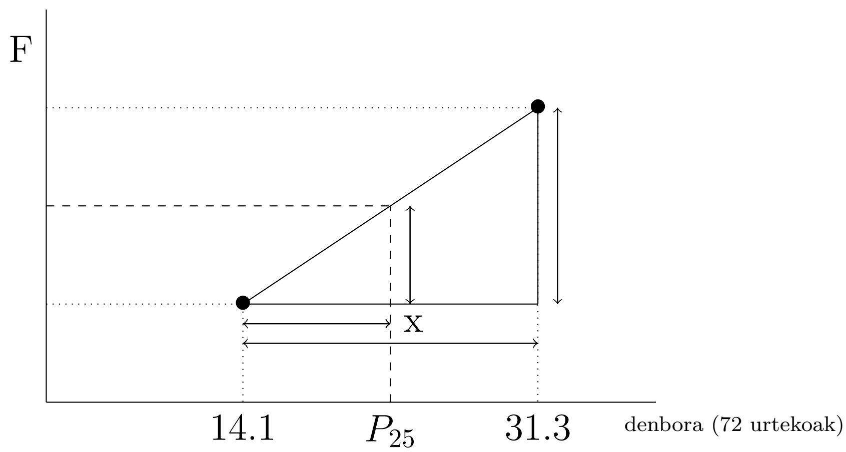
Azterketa berezia egingo zaie azken aldian ospitalera joan zirenetik pertzentil horietako balioetik gora kokatzen direnei.

Argi dago zenbat eta zaharragoa izan, osasuntsua izatearen definizioa aldatu egiten dela: gaztea denean, erlatiboki hilabete asko igaro behar dira gaztea osasuntsua kontsideratzeko; zaharra denean, ez horrenbeste.

(b) 70 urtekoetarako daturik ez dagoenez, 60 urteko eta 80 urtekoentzako emaitzen arteko interpolazioa egin beharko da. Linealtasuna suposatuz, 70 urtekoen 25. pertzentila ondoko bi pertzentilen batezbestekoa izango da noski.

$$P_{25}(70\ urte) = \frac{\qquad}{\qquad} =$$

72 urtekoetarako interpolazioa garatzen dugu baita ere, baino kasu honetan grafiko batez egin beharko dugu:



$$\frac{x}{\qquad} = \frac{\qquad}{\qquad} \to x = \qquad\qquad \to P_{25}(72urte) =$$

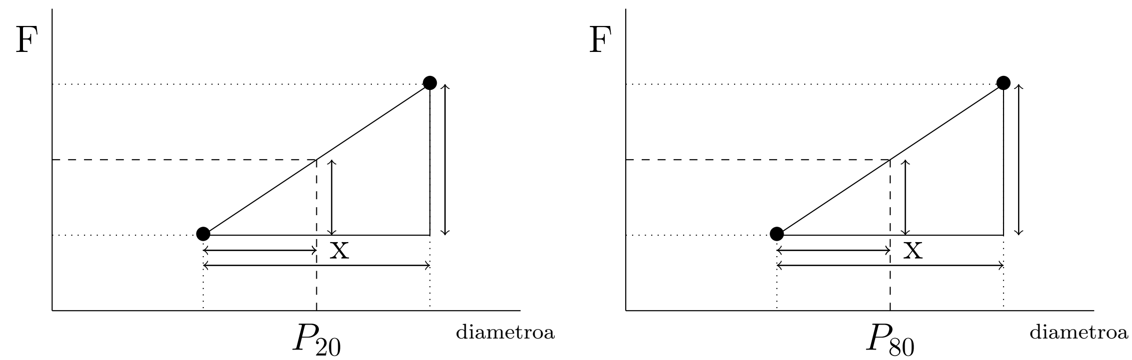
Azterketa berezia egingo zaie azken aldian ospitalera joan zirenetik:

- 70 urtekoen kasuan, hilabetetik gora pasa dutenei joan gabe;
- 72 urtekoen kasuan, hilabetetik gora pasa dutenei joan gabe.

Egindako kalkulu guztiak hurbilketak dira, noski, pertzentilen kalkuluan datuak tartearen barruan guztiz uniformeki banatzen direla , eta arestiko hurbilketen kasuan, adin batetik bestera aztertzen dugun denbora linealki aldatzen dela suposatuz.

14. ariketa

(a) Lehen eta laugarren kintilak $P_{...}$ eta $P_{...}$ pertzentilak dira hurrenez hurren.



$$\frac{x}{\quad\quad} = \frac{\quad\quad}{\quad\quad} \to x = \qquad \to P_{20} =$$

(b)

x	n	nx	x_{mozt}	n_{mozt}	$n_{mozt}x_{mozt}$
105	32	3360	-	-	-
115	49	....	115	49 - ....=	....
125	65	....	125	65	....
135	36	....	135	36 - ....=	....
145	16	....	-	-	-
155	0	....	-	-	-
165	$\boxed{1}$	165	-	-	-
175	$\boxed{1}$	175	-	-	-
	200	....		....	....

$$\overline{x} = \frac{\quad\quad}{200} =$$

$$\overline{x}_{mozt} = \frac{\quad\quad}{\quad\quad} =$$

(c)

x_{win}	n_{win}	$n_{win}x_{win}$
....	32	....
....	....	....
115	49 - ....=	....
125	65	....
135	36 - ....=	....
....	....	....
....	16	....
....	0	....
....	$\boxed{1}$	....
....	$\boxed{1}$	....
	200	....

$$\overline{x}_{win} = \frac{\qquad\qquad}{200} =$$

(d)

Moztutako eta winsortarturiko batezbestekoen helburua datu atipikoak ezabatu (moztutako batezbestekoaren kasuan) eta horien eragina mugatzea (winsortarturiko batezbestekoen kasuan) da. Batezbesteko horiek batezbesteko aritmetiko sinplearekin ....