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 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

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

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). Bestelako irizpideak ere aurki daitezke estatistika-testuetan.

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

Eta zerrenda ordenatu horietatik:

$$Me_A = 35$$

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

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:

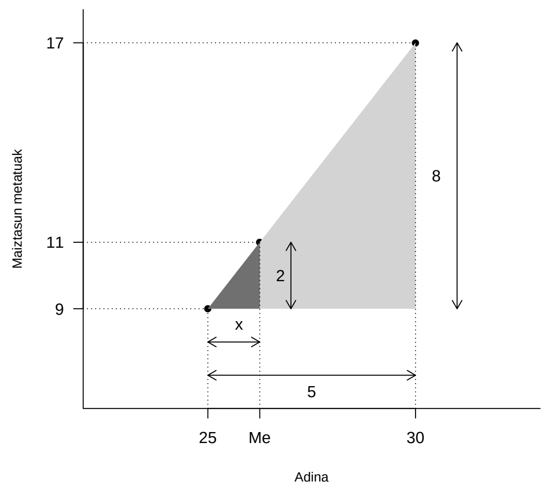
Adina	Langileak (n)
15-20	2
20-25	7
25-30	8
30-35	4
35-40	1
	22

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, 11. 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. 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.

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

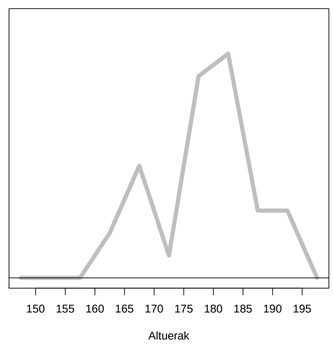
Harrigarria da zentroa neurtzea xede duten bi neurrik hain emaitza desberdinak izatea. Horren kausa datu-multzo nagusitik oso aparte dagoen *outlier* edo *muturreko datua* da (160). Balio horrek gora bultzatzen du batezbestekoaren emaitza, eta batezbestekoa distortsionatu; hain zuzen, inork ez luke esango 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 kontuan.

3.2.4 Moda

Moda (Mo) gehienetan errepikatzen den datuak hartzen duen balioa da, maiztasun handiena duena. Aldagai jarraituen kasuan, datuak tartetan biltzen direnean alegia, moda histograman edo maiztasun-poligonoan gailurra kokatzen den aldagaiaren balioa da.

Batzuetan, moda gailurtzat hartzen den kasuetan, badira, maiztasun handieneko *moda absolutuaz* gainera, *moda erlatiboak*, inguruko balioetan baino maiztasun handiagoa duten balioak alegia. Aurrerago ikasiko dugunez, moda anitz izateak 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.

3.2.5 Batezbesteko bereziak

3.2.5.1 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 betezeko lehiaketan hautagai batek puntuazio hauek eskuratu ditu (0tik 10erako eskala batean): A frogan, 6; B frogan, 8; eta C frogan, 9. Frogei eman beharreko pisuak 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 erabili. 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:

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

3.2.5.2 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:

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

3.2.5.3 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.

3.2.5.4 Batezbesteko harmonikoa

Batezbesteko harmonikoa batez besteko abiadura eta errendimenduak 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:

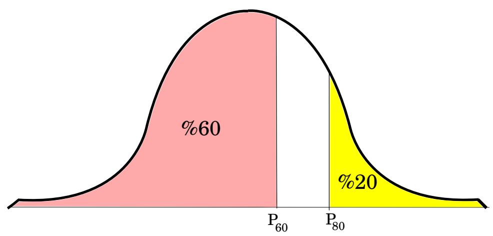
$$H = \frac{n}{\sum_i \frac{1}{x_i}}$$

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 jakingarri batzuk. 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:

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

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

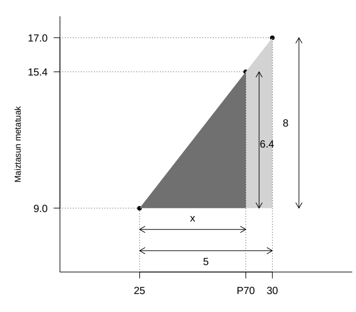
Kuantilarekin bat datorren datua zenbatgarren den jakin ondoren, datua hori kokatzen den tartea bilatu eta interpolatu behar da.

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.

3.4 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 batezbesteko aritmetiko sinplea eta mediana, muturreko datuen eragina aztertu eta emaitzak interpretatu.

2. 2005 eta 2015 urteetan herri bateko urteko familia-errentak jaso dira. Honako hauek dira datuak tartetan bilduta:

Errenta	Familiak (2005)	Familiak (2015)
< 5000	8	12
5000-10000	14	10
10000-20000	22	26
20000-30000	6	4
30000-50000	2	2

   Kalkulatu 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. Hiri osoarekin alderaturik, auzo zaharkituak zein diren ezarri behar da, horretarako bizilagunen batez besteko adina kalkulatuz. A, B, C, D eta E auzoetan biztanleriaren %20, %10, %20, %20 eta %30 bizi dira, hurrenez hurren. Hiritar guztien adinen mediana ere kalkulatu.

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

4. Datu hauek eskuratu dira aldagai bati buruz: 22-24-16-18-20. Gaineratu lau datu, mediana batezbesteko aritmetiko sinplea baino handiagoa izan dadin.

5. 10 egunetan zehar, eguneko batez besteko ekoizpena 80 unitatekoa izan zen. Beste 20 egunetan, 100 unitatekoa izan zen. Zenbat da 30 egunetako eguneko batez besteko ekoizpena? 40 egunetan batez besteko ekoizpena 110era heltzea nahi bada, zenbat izan behar da azken 10 egunetan eguneko batez besteko ekoizpena?

6. 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.

7. Kapital bat 10 urtetan zehar bikoizteko, zenbat izan behar da urteko batez besteko errentagarritasun-tasa? Eta 5 urtetan, kapitala %50 gehitzeko?

8. 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?

9. Eguneko salmentak jaso dira denda batean 26 larunbatetan:

   $3.020-2.695-2.324-2.835-1.719-786-2.669-2.801-3.148-3.661-2.932-5.012-4.567$

   $4.008-3.246-3.361-3.742-4.002-2.633-2.607-3.495-3.839-2.402-1.452-1.974-2.836$

   Salmenta txikieneko larunbaten %10etan, promozio-kanpaina bati ekiten zaio hurrengo asterako. Zenbat izan behar dira horretarako salmentak?

10. 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 froga bereziak egingo zaizkie. Zein puntuazio behar da horretarako?

11. 20, 40, 60 eta 80 urteko pertsonen artean, azken ospitalaratzera arteko denborak (hilabetetan) jaso dira:

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

Hemendik aurrera adin bakoitzean ospitalaratua izan arteko denbora gehien pasatzen duten pertsonen %75ari, azterketa berezia egingo zaie, horien bizimodu osasuntsuari buruzko informazioa jasotzeko.

(a) Ospitalaratu arte zenbat denbora pasa behar da horretarako adin bakoitzean?

(b) 70 urteko pertsonen artean ere egin behar da bizimoduari buruzko ikerketa. Horien artean, zeintzuei egingo zaie azterketa?