eman ta zabal zazu

Universidad del País Vasco Euskal Herriko Unibertsitatea

ALDAGAI BAKARREKO ESTATISTIKA DESKRIBATZAILEA ETA INFERENTZIALA GIZARTE ZIENTZIETAN: TEORIA ETA ADIBIDEAK

PATXI JUARISTI LARRINAGA

ISBN: 978-84-9860-880-9

EUSKARA ETA ELEANIZTASUNEKO ERREKTOREORDETZAREN SARE ARGITALPENA

Liburu honek UPV/EHUko Euskara eta Eleaniztasuneko Errektoreordetzaren dirulaguntza jaso du

AURKIBIDEA

HASIERAKO BERBAK

Estatistika zenbakizko datu multzoak sortu, zenbatu, sailkatu eta aztertu ostean ondorioak ateratzea ahalbidetzen digun jakintza da. Haren bidez, errealitate bat deskriba dezakegu (estatistika deskribatzailea), baina baita, laginen azterketetatik abiatuz, populazioen berezitasunei buruzko ondorioak atera eta zorizko aldagaiei buruzko aurresateak egin ere (estatistika inferentziala).
Gure mundu konplexu hau ezagutzeko edota gertatuko dena aurresateko, ezinbestekoa da estatistika. Merkatuen, sektore ekonomikoen edo enpresen egoerak ulertzeko, ekonomialariek ekoizpen-prozesuei, produktu eta zerbitzuen eskaerei eta eskaintzei edo Barne Produktu Gordinari buruzko denbora-serieak sortzen eta aztertzen dituzte; osasungintzako profesionalek gaixotasunei, gaixoei eta tratamenduei buruzko estatistikak biltzen dituzte, etor daitezkeen osasun-arazoak direla eta, erabaki egokiak hartzeko edota prebentzio-kanpaina arrakastatsuak bultzatzeko; eta soziologo eta politologoek gizarteei buruzko hainbat datu estatistiko biltzen dituzte, giza taldeen berezitasun eta bilakaera politiko, sozial eta demografikoak erakusteko edota aldez aurretik ezagutzeko.
Egia da estatistika askotan erabiltzen dela norbere erabaki, posizio edota portaerak zuritzeko edota besteenak manipulatzeko. Hala ere, ukaezina da estatistika gizarte modernoen motor, oinarri eta egituratzailea dela; hain zuzen ere, kalitatezko estatistikek gizarte aurreratu, demokratiko eta kohesionatuak sortzen dituztelako. Beraz, gizarte-ikerlarion erantzukizuna da estatistikaren prozedurak, teknikak eta lan egiteko modu fidagarriak eta baliagarriak ezagutzea, kalitatezko datu multzoak lortzeko; eta, azken finean, guri dagokigunaren aldetik, gizarte hobeak sortzeko. Ezin dugu ahantzi soziologook zein politologook sortzen ditugun estatistiketan oinarrituta, gobernuek, alderdi politikoek, sindikatuek, gobernuz kanpoko erakundeek zein beste hainbat eragilek giza taldeen bilakaera moldatzen duten erabakiak hartzen dituztela.

Liburuaren helburuak

Liburu hau estatistikarako lehen sarrera da, eta gizarte-ikerlari izango direnei aldagai bakarreko estatistikaren oinarriak aurkeztea du helburu nagusitzat; zehazki, soziologogaiei eta politologogaiei. Izan ere, teoriarekin eta ikerketarako metodo eta teknikekin batera, estatistika oinarrizko tresna da gizarte-errealitatea ulertzeko. Ildo horretatik, ezin dugu ahantzi E. Durkheim-ek *Metodo soziologikoaren arauak* liburuan zioena: “Estatistikak arima kolektiboaren egoera adierazten du”.
Urteetako esperientziak erakusten digu Soziologia eta Politika Zientziak eta Kudeaketa Publikoa graduak ikastera datozenetariko askok matematikak ez dituztela gustukoak, edota, askotan, zenbaki bat ikusi orduko beldurtu egiten direla. Horren ondorioz, asko kostatzen da era horretako ikasleak estatistikaren mundura erakartzea, eta, are gehiago, estatistikaren xarma erakutsaraztea.
Hori dela eta, liburu ulerterraza egiten saiatu gara. Demostrazio matematikorik ez dugu sartu, eta, gure ustez, formula eta prozedura ulerterrazenak aukeratu ditugu, betiere ezer jakintzat eman gabe, eta pauso guztiak zehatz-mehatz azalduz.
Formulak ezagutzeari eta emaitza estatistikoak ondo kalkulatzeari garrantzi handia ematen diogu, baina hori bezain garrantzitsua da kontzeptu estatistikoak argi izatea, erabili beharreko prozedurak ondo aukeratzea eta emaitzak ondo interpretatzea. Beraz, ahalegin berezia egin dugu ikerketa-egoera bakoitzean eman beharreko pausoak ondo azaltzen, eta datu multzoak, datu-diagramak eta lortzen diren emaitzak behar bezala aurkezteko eta interpretatzeko behar diren ezagutzak modu errazean agertzen. Horren haritik, teoria asko duen liburua da, eta, beraz, irakurketa sakona eskatzen du.

2013a Estatistikaren Nazioarteko Urtea dela kontuan hartuz, eta urteetan bildutako ezagutzan, esperientzian eta apunteetan oinarrituz, ikasleentzat bereziki erabilgarria izango den eskuliburua egiten saiatu gara. Irakurleak esango du lortu dugun ala ez.
Goi-Mailako Hezkuntzaren Europako Esparrura egokitutako graduetan eskaintzen diren ikasgai gehienek bezala, Soziologia eta Politika Zientziak eta Kudeaketa Publikoa graduetako bigarren mailan eskaintzen den Datuen Kudeaketa eta Analisi Estatistikoa ikasgaiak ere 15 asteko iraupena du. Astean 2 ordu teorikoren eta 2 praktikoren bidez, bigarren mailako ikasleek 32 ordu ematen dituzte irakaslearen azalpen teorikoak entzuten (eskola magistralak); 14 orduz aritzen dira praktikak egiten ikasgelan, eta beste 14 ordu ordenagailu-gelan ematen dituzte, SPSS programa estatistikoa ikasten. Horrez gain, ikasleek bakarkako edo taldekako lanak egin behar izaten dituzte ikasgelatik kanpo. Hain zuzen ere, irakaslerik gabe, 48 ordu eman behar dituzte teoria lantzen; 21 ordu, praktikak egiten, eta beste 21 ordu, beren kontura ordenagailu-gelan lanak egiten.
Liburu hau iraupen eta lan egiteko modu hori kontuan hartuta eginda dago; hau da, Datuen Kudeaketa eta Analisi Estatistikoa ikasgaia gainditzeko behar diren 150 orduei egokitutako eskuliburua da (60 ordu irakaslearekin eta 90 ordu ikasgelaz kanpo).
Ikastaro trinko horren ostean, ikasleek datu multzo kuantitatiboak prestatzeko, kudeatzeko eta behar bezala aurkezteko gai izan behar dute, baita ikerketetan sortzen diren datu estatistikoak interpretatzeko, laginak kalkulatzeko eta laginen ikerketatik abiatuz populazioei buruz ondorioak ateratzeko ere. Ikasketa-bidaia horren bukaeran, beraz, ikasleek, aldagai bakarrean oinarrituta, azterketa estatistiko sinpleak egiteko gaitasuna lortu behar dute.

Liburuaren egitura

Estatistikaren sarrera azaltzen duten lanen antzerako egitura du eskuliburuak. Lehenengo atalean, ikerketa-prozesua eta estatistika nola uztartzen diren agertzen dugu, estatistikak dituen bi adarrak aurkeztuz (estatistika deskriptiboa eta inferentziala), aldagai mota desberdinen berezitasunak azalduz, eta IBM SPSS programa informatikoari buruz hiruzpalau azalpen emanez. Estatistikaren oinarrizko kontzeptuak lantzea eta ikaslea estatistikaren munduan kokatzea da lehen atal horren helburua.
Bigarrenean, aldagai bakarreko estatistika deskribatzailearen gorabeherak ageri dira. Hain zuzen ere, azaltzen dugu maiztasun-taulak, datu-diagramak eta neurri estatistikoak nola kalkulatzen, aurkezten eta kudeatzen diren. Bukaeran, banaketa normalaren berezitasunak ere aurkezten dira. Horrekin, azterketa estatistiko deskriptiboak egiteko baliabideak eman nahi zaizkie ikasleei.
Hirugarren partean, laginaren eta populazioen arteko harremanak aztertzen dira. Laginen azterketatik abiatuz, populazioari buruz hitz egiteko tresnak lantzen dira, eta, horrela, estatistika inferentzialari ekiten. Hala, puntu-estimazioarekin, tarte-estimazioarekin eta hipotesi-frogekin lan egiteko moduak aurkezten dira. Gaia bukatzeko, laginen tamaina kalkulatzeko formula ere aztertzen da.
Eranskinetan, banaketa normalaren azpiko azalera-taula eta landu ditugun formula guztien zerrenda aurkezten ditugu.

Hartutako erabakiak

Arestian esan bezala, estatistikako eskuliburu ulerterraza egitea izan da helburua, eta, horretarako, teoria edota oinarrizko kontzeptuak ahal izan dugun modu argienean aurkezten saiatu gara: demostrazio matematikorik ez dugu sartu, eta, gure ustez, formula eta prozedura ulerterrazenak aukeratu ditugu.
Formulak liburuetan edo interneten erraz aurkitzen dira, eta, beraz, ez dira buruz ikasi behar, baina garrantzitsua da haien osagaiak ezagutzea, eta haiekin lan egiten jakitea. Horregatik,

formula bat lehenengoz agertzen denean, ikurra agertzen dugu, formulaz ohartarazteko, eta arretaz begiratzera bultzatzeko.

Hainbat adibide, buruketa edota kasu praktiko ere aurkezten ditugu, teoria behar bezala ulertzeko. Adibide bat ematen dugun bakoitzean, ikurra sartu dugu, teoria eta adibideak ondo bereizteko, eta adibideetan egiten ditugun interpretazioez ondo jabetzeko. Esan bezala, datu multzoak, datu-diagramak, neurri estatistikoak eta, oro har, estatistikaren laguntzarekin lortzen ditugun emaitzak ondo interpretatzen jakitea da liburu honen helburuetariko bat.

Bi eratako adibideak daude: datu multzoak programa informatikoen laguntzarik gabe aztertzen dituzten prozeduren ingurukoak, eta IBM SPSS programa informatikoaren bidez egindakoak. Gaur egun, estatistika ezin da ulertu programa informatikoen laguntzarik gabe, eta badira estatistika ordenagailuan oinarrituta soilik irakasten duten irakasleak, estatistika irakasteko modurik onena dela pentsatuta. Hala ere, gure ustez, estatistikaren oinarriak behar bezala barneratzeko, ezinbestekoa da formulekin eta eskuz egindako prozedura estatistikoekin lan egiten jakitea. Ahalegin hori eginda, ordenagailuaren bitartez egiten den lan estatistikoa errazago ulertzen da.

Estatistika lantzeko beste programa informatiko batzuk ere badaude, baina UPV/EHUk IBM SPSS programaren aldeko hautua egin du, eta, beraz, egokiena eta praktikoena iruditu zaigu ikasleek eskura duten programa horren bidez egindako adibideak sartzea. Hala ere, lan honetan ez dugu programa hori latzen. Nahiz eta programa hori erabiltzen ikastea ikasgaiaren xedeetariko bat den, haren gorabeherak azaltzea liburu honen helburuetatik haratago doa, eta, beraz, ez gara lan horretan sartu.

Liburua euskaraz idatzia dago, baina IBM SPSS programaren bidez egindako adibideak eta lortutako emaitzak programak eskaintzen dituen modu berean eta jatorrizko hizkuntzan (gaztelaniaz) sartu ditugu, inolako aldaketarik egin gabe, ikasleak programaren emaitzak ondo ezagut ditzan.

Izenburuak adierazten duen bezala, lan honen helburua aldagai bakarreko estatistikaren oinarriak aurkeztea da. Nahiz eta Gizarte Zientzietan bi aldagai edo aldagai anitz batera aztertzen dituzten azterketa estatistiko asko egiten diren, liburu honetatik kanpo utzi ditugu, 2. mailako ikasleek lortu behar duten ezagutza mailatik kanpo daudelako.

Irakurleak ikusiko du hainbat azalpen, ideia edota kontzeptu liburuaren pasarte batean baino gehiagotan errepikatzen direla. Batzuetan aspergarria izan daitekeela onartuta ere, pentsatu dugu errepikapenak prozedurak ondo finkatzeko eta ulertarazteko lagungarri direla. Gainera, gaiak eta kontzeptuak lotuta daudenez, gai bat jorratzerakoan ezinbestekoa da aurreko gai eta kontzeptuak azaleratzea.

Liburuan zehar, behin eta berriz errepikatzen dugu datu-diagramek izenburua eta identifikazio-zenbakia eraman behar dituztela. Adibide gisa erabiltzen ditugun datu-diagrama eta taula batzuk ez dira guk eginak, beste nonbaitetik hartuak baizik. Horrelakoetan, jatorrizko iturriko zenbakiaz gain, guk jarritako zenbakiak jarri dizkiegu, testuari osotasuna emateko.

Hartutako erabakiekin bukatzeko, esan behar dugu liburuan zehar azalpen estatistikoekin batera ikerketarako teknikekin eta txostenen idazketarekin zerikusia duten hainbat azalpen eskaintzen ditugula. Izan ere, gizarte-ikerlariok egiten ditugun azterketa estatistikoak oso lotuta daude gizarte-ikerketarako teknikekin.

Eskerrak ematen

Aitzinsolas hau bukatzeko, lan hau ahalbidetu duten guztiei eskerrak eman nahi dizkiet. Eta hasteko, 1997. urtetik aurrera, UPV/EHUko Gizarte eta Komunikazio Zientzien Fakultatean Estatistika ikasgaian ikasle izan ditudanei eskerrak eman gura nizkieke. Haien galdera, ohar eta zalantzak jakintza horretan sakontzera behartu naute, baita liburu hau egitera ere.

Eskerrik asko, orobat, UPV/EHUko Euskara eta Eleaniztasuneko Errektoreordetzari, euskarazko ikasmaterialgintza sustatzeko 2102ko deialdian lan honi ekiteko emandako bultzadagatik.
Eskerrak, halaber, testua zuzentzeko iradokizunak eta zuzenketak egin dizkidaten lagun eta lankideei, bereziki Amaia Lasherasi, euskararen ikuspuntutik egindako zuzenketengatik.
Eskerrik asko Cristina Laviari, estatistikako zalantzaren bat dudan bakoitzean beti laguntzeko prest agertzen delako. Liburu hau egitean izan ditudan zalantza asko berak argitu dizkit.
Azkenik, ez ditut ahantzi nahi nire emazte Maite eta seme-alaba Malen, Ibon eta Imanol; lan hau haiei eskaintzen diet. Hau bezalako behar luze batek behar duen oinarri, indar eta orekarik ez bainuen izango haiek inguruan izan ez banitu.

1. IKERKETA PROZESUA, ESTATISTIKAREN ADARRAK, ALDAGAIAK ETA IBM SPSS PROGRAMA

Gizarteen eta gizakien egoera eta bilakaera ezagutzeak informazio era sistematikoan biltzea eta zehaztasunez aztertzea eskatzen du. Horretarako, nahitaez ezagutu behar dira ikerketa-prozesuaren pausoak, ikerketarako tekniken berezitasunak eta horiek erabiltzeko moduak, eta, azkenik, horiek erabili ostean lortzen diren datu multzoak kudeatzeko, deskribatzeko eta inferentziak egiteko prozedura estatistikoak.

1.1. IKERKETA PROZESUA, IKERKETA TEKNIKAK ETA ESTATISTIKA GIZARTE ZIENTZIETAN

Ikerketa bakoitzaren berezitasunak eta beharrizanak ahantzi gabe, Gizarte Zientzietan egiten ditugun ikerketetan, hamar pauso eman behar dira:

- Ikergaia definitu
- Gaiari buruzko literatura aztertu
- Hipotesiak formulatu
- Aldagaiak operazionalizatu
- Ikerketa diseinatu
- Datuak bildu
- Datuak kudeatu eta aztertu
- Ondorioak interpretatu
- Txostena idatzi
- Erabilitako teknikaren fidagarritasuna eta balioa neurtu

Hamar pauso horiek ondo ezagutzeaz gain, gizarte-ikerlariek ikerketarako teknikak eta horiek erabiltzeko modu egokienak ere ezagutu behar dituzte.

Azterketaren helburua fenomenoen esanahiak sakonean aztertzea bada, alegia, pertsonek gordetzen duten informazio konplexu eta aberatsa ezagutzea, sakoneko elkarrizketa, delphi teknika, bizitza-historiak edota eztabaida-taldeak dira ikerketarako teknikarik aproposenak. Tresna horiek ez digute adierazgarritasunik ematen; hau da, ikergai diren pertsonengandik lortzen dugun informazioa ezin dugu populazio osora orokortu, baina aztergai diren pertsonentzat garrantzitsua edota adierazgarria dena azaleratzen laguntzen digute; eta beren mundua, bizitza zein ondo ezagutzen duten arlo edo gai bat nola ikusten, sailkatzen eta bizitzen duten ezagutzeko aukera ematen:

- Sakoneko elkarrizketetan eta bizitza-historietan, pertsonen bizi-esperientziak zein gai zehatz baten inguruan dituzten ikuspuntu, iritzi eta jarrerak aztertzen dira, esperientzia edo iritzi horiek berreraikitzeko asmoz. Berreraikitze hori elkarrizketatuek azpimarratu dituzten gai garrantzitsuenak, fase erabakigarrienak, inflexio-puntuak eta une edo gertaera kritikoenak aztertuz egiten dugu.
- Eztabaida-taldeetan, elkarrekin eztabaidan edo solasean aritu direnen ikuspuntu, iritzi eta jarrerak aztertzen dira, esperientzia edo iritzi desberdinak ezagutzeko asmoz; baita egon daitezkeen adostasunak, desadostasunak eta iritzi partikularrak ere.
- Delphietan, galdetegia bete duten adituen artean dauden adostasunak eta desadostasunak aztertzen dira, baita egon daitezkeen iritzi partikularrak ere.

Hiru bide horietatik lortzen dugun informazioa modu kualitatiboan kudeatu eta aztertu behar da, eta, beraz, estatistikak laguntasun gutxi ematen digu. Teknika kualitatiboak erabiltzen ditugunean, prozedura estatistikoek erabilgarritasun gutxi dute.

Estatistika baliagarria da adierazgarritasuna ziurtatzen duten eta gizarte-fenomenoen kontaketa ahalbidetzen duten teknika kuantitatiboak erabiltzen ditugunean. Horien bidez, gaiak teknika kualitatiboen bidez baino modu azalekoagoan aztertzen ditugu, baina adierazgarritasuna lortzen dugu, eta emaitzak populazio osora orokortzeko gaitasuna lortzen. Ikerketa-teknika kuantitatiboetatik lortzen ditugun datu multzoak aztertzeko, estatistikaren prozedurak ezinbestekoak dira:

- Saioan, estimulua jasan dutenen eta jasan ez dutenen arteko desberdintasunak aztertzen dira. Talde esperimentalean (estimulua hartzen duena) eta kontrolerako taldean (estimulua hartzen ez duena) mendeko aldagaiaren aldaketak estatistikoki neurtuz, estimulu esperimentalak zenbaterainoko eragina duen aztertzen da.
- Behaketa zuzenean, behatu nahi diren portaeren berezitasunak deskribatzen eta interpretatzen dira, eta beste aldagai batzuekin harremanetan jartzen. Horretarako, behaketan bildutako informazio kualitatiboa eta kuantitatiboa ordenatu behar izaten da.
- Eduki-azterketan, aztergai den testuaren berezitasunak aztertu behar dira kualitatiboki zein kuantitatiboki; kontuan izanik agerpen-maiztasunak testu batean garrantzia duten gaien adierazlea direla. Hitz edo kontzeptu batzuk zenbat eta gehiagotan agertu testuan, orduan eta garrantzitsuagoak dira testuaren sortzaileentzat.
- Iturri ez-zuzenen azterketetan, beste pertsona edo erakunde batzuk bildu, sortu edota gorde dituzten estatistika, dokumentu, grabaketa eta material pertsonalen azterketak egiten ditugu.
- Inkestan, datu multzo baten berezitasunak deskribatu (maiztasun-taulak, diagramak eta neurri estatistikoak erabiliz), egon daitezkeen erlazioak ezagutu (taula gurutzatuak, korrelazioak...) eta aldagai anitzetan dagoen informazioa laburbil dezakegu (aldagai anitzen azterketak). Horrekin batera, lagina erabiltzearen ondorioz egiten ditugun erroreak (laginketa-errorea) ezagutu behar ditugu, populazio osoari buruzko ondorioak ateratzeko asmoz. Besteak beste, honelako galdera hauei erantzun nahi izaten diegu inkesten bitartez lortutako datu multzoak aztertzen ditugunean: nolako berezitasunak ditu lortu dugun datu multzoak? Harremanik badago azaldu nahi dugun fenomenoaren (mendeko aldagaia) eta aztergai diren beste aldagai edo faktoreen artean (aldagai askeak)? Zenbatekoa da harreman hori? Nolakoa da? Zer-nolako probabilitatea dago laginean aurkitu ditugun harreman horiek populazioan aurkitzeko?

Era horretako teknika kuantitatiboak erabiltzen ditugunean eta datu multzo kuantitatiboak lortzen, oso erabilgarriak dira estatistikak eskaintzen dizkigun prozedurak eta tresnak. Izan ere, elementu, unitate edo pertsona multzo bati dagozkion zenbakizko datuak biltzen, sailkatzen eta aztertzen dituen ezagutza da estatistika.

1.2. ESTATISTIKA DESKRIBATZAILEA ETA ESTATISTIKA INFERENTZIALA

Estatistika XVII. mendean sortu zen diziplina zientifiko gisa. Dena dela, ukaezina da aurretik ere datu estatistikoak sortzen edota biltzen zirela. Hain zuzen ere, lehenengo estatistikak zentsuak eta erroldak izan ziren. Antzinaroko inperio edo zibilizazioetako agintariek (egiptoarrek, babiloniarrek, greziarrek, txinatarrek, pertsiarrek, inkek…) menpean zituzten lurraldeetako biztanleei buruzko datuak biltzen zituzten, beren boterea babesteko edota sendotzeko. Ezaguna da, adibidez, gure aroaren aurreko 2238. urtean, Txinako Yao enperadoreak zentsua egiteko agindua eman zuela.

Erromatarrek ere hainbat errolda egin zituzten beren inperioan barrena; ezagunenetarikoa Oktavio Augusto enperadoreak gure aroko 0. urtean agindutakoa da; hau da, Jesus Belenen jaiotzera bultzatu zuena. Lukasen Ebanjelioan irakur daitekeen bezala: “Garai hartan, mundu guztiko errolda egiteko agindua eman zuen Augusto enperadoreak. Lehen errolda hura Siriako gobernari Kirino zela egin zen. Beraz, erroldatzera joan ziren denak, nor bere herrira” (Lukas 2, 1-3).

Inkek ere datu estatistikoak sortzen zituzten kipuen bitartez. Kipuak kolore desberdinetako sokatxoak ziren, eta gorde nahi zen zenbakiaren araberako korapilo-kopurua zuten[1]. Erdi Aroan, hain zuzen ere, 762. urtean, Karlomagnok bere inperioaren ondasunen zerrenda abiarazi zuen, eta hainbat datu estatistiko sortu.

Ingalaterran, Gilermo I.aren aginduei jarraituz, Domesday Book izeneko datu-bilketa zabala egin zen 1086. urtean, zerga-bilketa ondo antolatzeko.

Adibide asko daude erakusten dutenak gobernuek eta agintariek antzinatik bildu dituztela datu estatistikoak, beren boterea sendotzeko, estatuak eta inperioak hobeto kontrolatzeko edota gobernatzeko .

Dena dela, XVII. mendera arte, datu kuantitatiboen bilketak egiten ziren, errealitatea deskribatzeko edo ezagutzeko, baina ez zegoen inolako inferentziarik. Ez zen laginekin lan egiten, populazioen berezitasunei buruzko ondorioak ateratzeko, eta ez ziren aldagaiei buruzko aurresateak egiten.

Errolda edo zentzuen bidez egiten zen estatistika deskribatzailetik estatistika inferentzial modernorako pausoa John Graunt-ek (1620-1674) eman zuen 1662. urtean, Londresen argitaratu zuen Natural and Political Observations Made upon the Bills of Mortality liburuarekin. Lan horretan, alde batetik, 1592tik 1662ra Londreseko parrokia-orrietan oinarrituz, jaiotzen, ezkontzen eta heriotzen berri eman zuen (estatistika deskribatzailea), eta, bestetik, biztanleriari, heriotza kopuruari eta, oro har, aldagai demografikoei buruzko aurreikuspenak egin zituen (estatistika inferentziala).

John Grauntekin estatistikaren bi adar nagusiak bereizten hasi ziren. Alegia, zenbakizko datu multzoak aztertzeko bi modu agertu ziren: estatistika deskribatzailea eta estatistika inferentziala.

Estatistika deskribatzailea zenbakizko datu multzoak bildu, sailkatu, diagrama edo grafikoen bidez irudikatu eta haien ezaugarri garrantzitsuenak azaleratzeaz arduratzen den estatistikaren adarra da. Datu-multzoak laburbiltzea eta haien ezaugarriak era ulergarri eta sinplean ematea da haren helburua. Horretarako, maiztasun-taulak, datu-diagramak eta estatistikoak erabiltzen ditu, nagusiki.

Inferentzia estatistikoak, estatistika inferentzialak edo estatistika induktiboak lagin batetik abiatuz, populazioaren berezitasun ezezagunei buruzko ondorioak ematen ditu; baita zorizko aldagai bati buruzko aurresateak egin ere. Adibidez, inferentzia estatistikoaren laguntzarekin, hiri bateko biztanle guztien batez besteko adina, hurrengo hauteskundeetan bozka nori emango dioten edo gai baten inguruko iritziak ezagut ditzakegu, hiri horretako 1.200 pertsonaz osatutako lagina ikertuz. Edo etorkizunean jaiotze-tasarekin gertatuko dena jakin dezakegu, orain arteko tasetan oinarrituta. Bi kasuetan, ezinbestekoa dugu Estimazioaren Teorian, Probabilitatearen Teorian, Limitearen Teorema Zentralean eta Kopuru Handien Legean oinarritzea.

Datu kuantitatiboen araberako ikerketa egin aurretik, beraz, argi izan behar dugu zein den gure helburua. Errealitatea deskribatu nahi dugu? Edo deskribatzeaz gain, populazio bati buruzko inferentziak ere egin nahi ditugu? Estatistika inferentzialak eta deskriptiboak datu multzoen izaera adierazteko neurri estatistiko (batez bestekoa, moda, desbideratze tipikoa…), prozedura eta teknika berdinak erabiltzen dituzte (maiztasun-taulak eta datu-diagramak), baina helburu desberdinak dituzte, eta, horren ondorioz, neurri, prozedura eta teknika horiek modu desberdinean erabiltzen dituzte. Horrekin, zera esan nahi dugu: bide estatistiko bat edo bestea hartu aurretik, argi izan behar dugu zer egin nahi dugun gure ikerketan.

1.3. ALDAGAIAK, KONSTANTEAK ETA ALDAGAI MOTAK

Aldagaiekin lan egiten dugu Gizarte Zientzietan. Neurgarriak diren eta elementu, unitate, gauza edo pertsona batzuetatik besteetara aldatzen diren ezaugarriak neurtzen ditugu (adina, bozka

[1] Kipuak sortu, gorde eta ulertzeko, Quipucamayoc (kipu-egile) zeritzen inperioko funtzionarioak zeuden; gaur egungo estatistikarien antzeko eginkizuna zuten.

politikoa, produktu baten kontsumoa, jarrerak, iritziak…). Garrantzitsua da, beraz, aldagaiak zer diren eta nola sailkatzen diren jakitea.

1.3.1. Aldagaiak eta konstanteak

Zer da neurtzea? Arau batzuei jarraituz, naturan edo gizartean dauden gauzen, elementuen edo izaki bizidunen ezaugarriei zenbakizko balioak ematea; hau da, ikergai ditugun ezaugarriei balioak ematea.
Neurketak beti balio bera ematen badigu, konstante bat izango da, hots, gauza, elementu edo izaki bakarraren ezaugarriak neurtzen ditugunean, eta neurketa horrek balio bakarra ematen duenean, konstante bat dugula esaten dugu. Gizarte Zientzietan ez dago konstanterik, baina fisikan edo matematikan, bai: grabitatea edo argiaren abiadura (300.000 km./segundoko), adibidez, konstanteak dira. Izan ere, neurketaren emaitza ez da aldatzen.
Ezaugarrien neurketak balio desberdinak ematen badizkigu, esango dugu aldagaiak ditugula. Hain zuzen ere, aldagai izena hartzen dute, neurtzen dituzten ezaugarriak aldatu egiten direlako. Adibidez, ezagutza-maila, esperientzia edo irakaskuntzarekiko jarrera aldatu egiten dira irakaslearen arabera, baita adina, sexua edo iritzi politikoa ere.
Esan bezala, Gizarte Zientzietan ez dugu konstanterik, eta, beraz, aldagaiak dira gizarte-ikerlarion neurgaiak.

1.3.2. Aldagai motak

Gizarte Zientzietan, ikerketan erabiliko ditugun aldagaien izaera oso ondo ezagutu behar dugu; horrek neurketarako erabiliko ditugun tresnak (galdeketa, behaketetan informazioa biltzeko erabiltzen ditugun fitxak, eduki-azterketak egiteko kodifikazio-hiztegiak…), ikerketa-diseinua (ikerketa-plana eta ikerketan zehar eman beharreko pausoak) eta azterketa estatistikoa baldintzatuko dituelako. Alegia, kontuan hartu behar dugu aldagai mota desberdinak daudela, eta bakoitzaren berezitasunek teknika estatistiko batzuk ahalbidetzen dituztela eta beste batzuk galarazten[2]. Aldagaiak lau irizpideren arabera sailkatu ditzakegu:

- Neurtzeko erabiltzen den eskalaren arabera
- Aldagaiaren izaeraren arabera
- Aldagaien artean sortzen diren harremanen arabera
- Aldagaien edukiaren arabera.

1.3.2.1. Aldagaiak neurketa-eskalaren arabera

Arestian esan bezala, arau batzuei jarraituz, naturan edo gizartean dauden gauzen, elementuen edo izaki bizidunen ezaugarriei zenbakizko balioak ematea da neurtzea. Horretarako, neurtzeko tresnak behar ditugu, baita neurketa-eskalak ere. Leioako Campuseko ikasleen altuera, adibidez, eskala metrikoan edo eskala ingelesean neur dezakegu, eta, horretarako, neurketarako hainbat tresna erabil ditzakegu. Tenperatura neurtzeko ere, zenbait eskala erabil ditzakegu (Celsius, Kelvin, Farenheit...), baita termometro mota desberdinak ere.

[2] Gizarte-zientzietan, neurtzen ditugun ezaugarri guztiei aldagai deitzen diegu, baina, modu zehatzean hitz eginda, aldagai estatistikoa gizabanakoaren kantitatezko ezaugarriak neurtzen dituena da, eta nolakotasunezko ezaugarriak atributuak dira.

Neurketa-eskalak

Stanley Smith Stevens-ek (1906–1973) neurketarako lau eratako eskalak bereizi zituen: nominala, ordinala, tarte-eskala eta ratio-eskala. Azter ditzagun eskala mota bakoitzaren berezitasunak:

1. **Eskala nominalean neurtzen diren aldagaiak aldagai nominalak edo atributuak dira.** Era horretako aldagaiek kategoria desberdinak dituzte[3], eta ikergai diren elementu, unitate edo pertsonak kategoria horien arabera sailkatzen ditugu; hau da, kategoria horien arabera berdinak edo desberdinak diren adierazten dugu. Kontuan izan kategoria guztiak maila berean daudela; alegia, ezin dugu esan kategoria batek beste batek baino maila edo balio handiagoa duenik. Adibidez, sexua (bi kategoria ditu: gizona eta emakumea), unibertsitatean aukeratutako ikasketak (Soziologia, Politika Zientziak eta Kudeaketa Publikoa, Kazetaritza, Medikuntza, Arkitektura, Ekonomia...), jaioterria, lan egiten duen sektorea, lanpostua, azken hauteskundeetan emandako bozka, sinesmen erlijiosoa, azken greban izandako portaera… aldagai nominalak dira.

   Adibidea. Aldagai nominala operazionalizatuta: Zein da zure lanbide edo ogibide nagusia?

   - Lan ordaindua……………………………………1
   - Langabezian………………………………………2
   - Etxeko lanak………………………………………3
   - Ikasten nagusiki…………………………………..4
   - Jubilatua, pentsioduna, ezindua…………………..5
   - Beste egoera batzuk……………………………..6
   - Ed-Ee (ez daki, ez du erantzuten)………………...7

2. **Eskala ordinalean neurtzen diren aldagaiak aldagai ordinalak dira** (kuasikuantitatiboak ere deitzen zaie, geroago ikusiko dugun bezala, kuantitatiboen berezitasun batzuk dituztelako). Aldagai ordinalek nominalen berezitasuna dute, eta, beraz, kategoria desberdinetan banatuta daude, eta ikertzen ari garen elementu, unitate edo pertsonak kategoria horien arabera sailkatu ditzakegu. Baina, nominaletan ez bezala, kasu honetan, kategoriak mailakatuta daude; hau da, kategorien artean, ordena bat dago. Adibidez, ikasketa maila neurtzen duen aldagaia ordinala da. Izan ere, gizarte bateko biztanleak egin dituzten ikasketen arabera sailkatu ditzakegu (ikasketarik gabe, oinarrizko ikasketak, Derrigorrezko Bigarren Hezkuntza (DBH), Batxilergoa, Lanbide Heziketa, Unibertsitateko Gradua eta Unibertsitateko Graduondoa, adibidez). Baina sailkatzeaz gain, ordena sortzen dugu pertsonen artean; hain zuzen ere, kategorien artean mailaketa bat dagoelako. Horrela, unibertsitate-ikasketak dituztenek Batxilergoa dutenek baino ikasketa maila altuagoa dutela esaten dugu, eta horiek, DBH egina dutenek baino maila altuagoa. Kategorien artean, ordena edo mailaketa bat badago, baina ezin dugu zehaztu kategorien artean dagoen distantzia zenbatekoa den. Ezin dugu esan batxilergoa duenak DBH duenak baino ikasketa maila bat edo bi gehiago dituenik; bakarrik esan dezakegu batxilergoa egin duenak DBH egin duenak baino ikasketa maila altuagoa duela.

   Adibidea. Aldagai ordinala operazionalizatuta: zuk zer moduz hitz egiten duzu euskaraz?

   - Oso ondo…………………………...1
   - Nahikoa ondo………………………2

[3] Kategoria aldagai nominal edo ordinal batek hartzen duen balioetako bakoitza da.

- Badakizu zerbait……………………...3
- Badakizkizu hitz batzuk……………...4
- Batere ez…………………………….5
- Ed-Ee………………………………..6

Adibidea. Aldagai ordinala operazionalizatuta: zein ikasketa amaitu dituzu?

- Batere ez, lehen mailakoak baino baxuagoak………………………….1
- Lehen mailakoak, Oinarrizko batxilergoa, OHO, DBH…………………2
- Lanbide Heziketa (LH)………………………………………………….3
- Bigarren mailakoak: Goi-mailako Batxilergoa, IEE, BBB, UBI……….4
- Erdi goi-mailakoak (Diplomaturak)……………………………………5
- Goi-mailakoak (Lizentziaturak, Ingeniaritzak, Doktoretzak,...)……….6
- Ed-Ee……………………………………………………………………7

3. **Tarte-eskalan neurtzen diren aldagaiak tarte berdinak dituen eskala baten arabera neurtzen dira.** Era horretako aldagaiekin, elementuen, unitateen edo pertsonen artean berdintasunak/desberdintasunak azter ditzakegu (nominalen berezitasuna) eta ordena zehaztu (ordinalen berezitasuna), baina, horrekin batera, eskalaren balioen artean dauden distantziak zehaztu ditzakegu. Hau da, balioen artean tarte edo distantzia berdinak daudenez, balio batean kokatzen diren elementu, unitate edo pertsonen eta beste balio batean kokatzen direnen artean dagoen distantzia zehaztu dezakegu. Tarte-eskaletan neurketarako erabiltzen den unitatea arbitrarioa da. Bestalde, era horretako aldagaiek ez dute zero absoluturik. Zeroa ere arbitrarioa da. Horren ondorioz, besteak beste, biderketa edo zatiketa ezin dira egin. Alegia, eskala horren araberako balioak batu (10º + 2º = 12º) edo ken ditzakegu (10º- 8º = 2º), baina, balioen artean, ezin dugu zatiketarik egin $\left(\frac{8^{o}}{2^{o}}\right)$, ezta biderkatu ere; beraz, ratioak ezin dira atera. Era horretako aldagaiak dira tenperatura Celsius eskalan neurtuta, adimen-test batean ateratako puntuazioa, datak edo gainpisua, alderaketarako aldez aurretik zehaztutako pisu-patroi baten arabera adierazita (eskala horrek gainpisua edo gehiegizko argaltasuna adierazten du, baina zeroa arbitrarioa da). Egia esan, Gizarte Zientzietan gutxitan erabiltzen ditugu tarte-eskalan neurtutako aldagaiak.

Adibidea. Tarte-eskalan neurtutako aldagaia: adimen-test batean atera diren puntuazioak. Testean, 100 puntu atera dituenak ez du 80 atera (nominalen berezitasuna); 80 atera duenak 100 atera duenak baino adimen gutxiago du (ordinalen berezitasuna), eta 80 atera duenak 100 atera duenak baino 20 puntu gutxiago ditu (tarte-eskalan neurtutako aldagaien berezitasuna). Aldagai hori ez da ratio-eskalaren arabera neurtuta, ezin daitekeelako esan adimen-test batean 100 puntu lortzen duenak 50 puntu lortu duenaren adimen bikoitza duela. Hau da, ezin da balio desberdinen arteko ratioa atera. Izan ere, adimen-test batean zeroa ez da zero absolutua. Zero ateratzeak ez du esan nahi adimenik ez dagoenik. Zeroa ikerlarien erabaki arbitrarioaren ondorioa da.

Adibidea. Tarte-eskalan neurtutako aldagaia: hotz-beroaren neurketa. 30º Celsiuseko tenperatura eta 10ºkoa ez dira gauza bera (nominalen berezitasuna), 30º 10º baino tenperatura handiagoa da (ordinalen berezitasuna) eta 10º daudenean, 30º daudenean baino 20º gutxiago daude. Aldagai hori ez dago ratio-eskalan neurtua, ezin daitekeelako esan 30º daudenean 10º daudenean baino 3 aldiz bero handiagoa egiten duela. Izan ere, hotz-beroa neurtzen dugunean, zeroa ez da absolutua: 0º egoteak ez du esan nahi tenperaturarik ez dagoenik, tenperatura negatiboak ere badirelako.

4. **Ratio-eskala bat dugu (ratio hitza latineko kalkulu hitzetik dator), baldin eta aldagaiak tarte berdinak dituen eskala baten arabera neurtzen badira, baina zero absolutua badugu.** Alegia, ratio-eskalek tarte-eskalen berezitasunak dituzte, baina desberdintasuna da 0 absolutua badagoela; hau da, zero balioak esan nahi du neurtzen ari garen ezaugarria ez dagoela. Horren ondorioz, balioak batu edo ken ditzakegu, baita biderkatu edo zatitu ere. Kasu horretan, balioen arteko ratioak atera ditzakegu.
   Etxe bateko diru-sarrerak neurtzen ditugunean, zeroak esan nahi du diru-sarrerarik ez dagoela, edo, adina neurtzen dugunean, jaioberri bat dela. Alegia, ratio-eskala zeroan hasten da, eta eskalan gora egiten dugu, neurtzen ari garen berezitasuna haziz doan neurri berean.

   Ratio-eskalan neurtutako aldagaien balioekin edozein operazio logiko (alderaketa eta ordena) eta aritmetiko (batuketa, kenketa, biderketa eta zatiketa) egin dezakegu. Besteak beste, bi balioren arteko alde absolutua zein erlatiboa ("ratioa" alegia) eman dezakegu. Era horretako aldagaiek neurketa-maila gorenak onartzen dituzte.

   Adibidea. Tarte-eskalan neurtutako aldagaia: esango didazu zenbat seme-alaba dituzun?

   Seme-alaba kopurua ratio-eskalan neurtzen da. Izan ere, bi seme-alaba dituenak ez du hiru seme-alaba dituenaren egoera bera (nominalaren ezaugarria). Hiru seme-alaba dituenak bi seme-alaba dituenak baino seme alaba gehiago ditu (eskala ordinalaren ezaugarria). Hiru seme-alaba dituenak bi seme-alaba dituenak baino seme-alaba bat gehiago du (tarte-eskalaren ezaugarria). Lau seme-alaba dituenak 2 seme-alaba dituenak baino seme-alaba kopuru bikoitza du (ratio-eskalen ezaugarria). Zero seme-alaba dituenak seme-alabarik ez duela esan nahi du. Kasu horretan, 0a ez da sortu ikerlarien nahien edo interesen arabera. 0a absolutua da, ez arbitrarioa.

   Adibidea. Tarte-eskalan neurtutako aldagaia: gorputzeko tenperatura termometro batekin neur dezakegu. Halere, baditugu eskala eta unitate desberdinetan neurtzen duten termometroak (Celsius eskala, Farenheit eskala, Kelvin eskala..).

   Celsius eskalan neurtzen badugu, tarte-eskalan neurtzen ari gara, eta Kelvin eskala erabiltzen badugu, ratio-eskalan. Zergatik gertatzen da hori, bi kasuetan tenperatura neurtzen ari bagara? Kelvin eskalak zero absolutua duelako eta Celsius-ek ez. Tenperaturaren zero absolutua Kelvin eskalakoa da. Izan ere, 0º Kelvin daudenean, molekulen jarduera gelditu egiten da, eta horien osagaien arteko ukiturik ez dago. Horregatik, hain zuzen ere, esan dezakegu 200º Kelvin 100º Kelvinen bikoitza dela. Kelvin eskalako zeroa -273º Celsiusetan gertatzen da. Ezin dugu esan, ordea, 200º zentigradu (473º Kelvin) 100º zentigradoren (373º Kelvin) bikoitza denik; Celsius eskalan zero absoluturik ez dagoelako.

Neurketa-eskalei buruzko ondorio batzuk

Neurketa-eskalen berezitasunak eta haien arteko harremanak aztertuta, zera ondoriozta dezakegu:

1. Neurketa-eskalak mailakatuta daude. Horren ondorioz, goi-mailako graduazio-eskalek behe-mailakoen abantailak dituzte, baina ez alderantziz:
   - Aldagaia eskala nominalean neurtuta dagoenean, ikergai diren elementu, unitate edo pertsonak aztertzen ari garen ezaugarriaren arabera (aldagaiaren kategorien arabera) sailkatu ditzakegu.
   - Aldagaia eskala ordinalean neurtuta dagoenean, elementu, unitate edo pertsonak kategoriatan sailkatzeaz gain, ordenatu egin ditzakegu (txikienetik handienera, gazteenetik zaharrenera, ahulenetik indartsuenera...).

- Aldagaia tarte-eskalan neurtuta dagoenean, elementu, unitate edo pertsonak sailkatzeaz eta ordenatzeaz gain, balio batean kokatzen direnek beste balio batean kokatzen direnek baino balio handiagoa edo txikiagoa dute.
- Aldagaia ratio-eskalan neurtua dagoenean, elementu, unitate edo pertsonak sailkatzeaz, ordenatzeaz eta beren artean dauden distantziak adierazteaz gain, balio batean kokatzen direnek beste balio batean kokatzen direnek baino hainbat aldiz balio handiagoa edo txikiagoa dute.

2. Berezitasun horiei lotuta, ezin dugu ahantzi aldagai mota bakoitzak eragiketa matematiko batzuk onartzen dituela. Eta horrek zera esan nahi du: aldagaien berezitasunek zehazten dituztela erabiliko ditugun neurri estatistiko deskribatzaileak. Sexuari, aldagai kualitatibo nominala denez, ezin zaio batez bestekoa kalkulatu (ez du zentzurik esatea 1,2ko sexua dagoela), ezta inolako sakabanatze- (desbideratze estandarra, bariantza…), kokapen- (pertzentilak, koartilak…), forma- (asimetria eta kurtosia) edo kontzentrazio- (Lorenz eta Gini) neurririk ere. Gehienez ere, sexua aldagaiari maiztasun absolutuak eta erlatiboak kalkula diezazkiokegu. Adinari, aldagai kuantitatiboa denez, batez bestekoa kalkula diezaiokegu, baita sakabanatze-, kokapen-, forma- eta kontzentrazio-neurri guztiak ere. Datu multzoak grafikoki agertzerakoan, zein maiztasunak kalkulatzerakoan ere, kontuan izan neurketa-eskala bakoitzak bere berezitasun matematiko, grafiko eta estatistikoak dituela, eta inolako kalkulurik egin aurretik aldagaiaren berezitasunei begiratu behar zaiela. Aldagai nominalak ditugunean, adibidez, ez du zentzurik maiztasun metatuak kalkulatzea edo histograma irudikatzea, baina bai tarte-eskalan neurtutako aldagaiak ditugunean.
3. Tarte-eskalen eta ratio-eskalen hurbiltasuna dela medio, Gizarte Zientzietan ez da tarte-eskalen eta ratio-eskalen arabera neurtutako aldagaien artean bereizten. Era horretako aldagai guztiak kuantitatiboak direla esaten da. IBM SPSS estatistikako programak berak ere “eskala” izenarekin biltzen ditu bi aldagai mota horiek. Hau da, IBM SPSS-rako, “eskala” motako aldagaiak dira hartzen dituzten balioak magnitudeak adierazten dituztenean, 0 absolutua duten ala ez kontuan hartu gabe. Egia esan, tarte-eskalan neurtutako aldagai gutxi erabiltzen dira Gizarte Zientzietan.
4. Azkenik, kontuan izan behar da aldagaiak neurtzeko modua aldatu egiten dela aldagaiaren operazionalizazioaren arabera. Aldagai bera eskala desberdinen arabera neur dezakegu, egiten dugun operazionalizazioaren arabera.

Adibidea. Bilboko alkateari buruzko iritzia modu desberdinetan operazionaliza dezakegu, eta, horren ondorioz, eskala desberdinen arabera neurtu:

- Bilboko alkateari buruzko iritzia ratio-eskalaren arabera neur dezakegu, era honetako galdera baten bidez: 0tik 10era, zer nota emango zenioke Bilboko alkatearen lanari, 0 guztiz txarra eta 10 guztiz ona direla kontuan hartuta? Norbaitek 6ko nota ematen badu, ez du 1 ematen (nominalaren ezaugarria). 6ko nota ematen duenak 1ekoa eman duenak baino nota gehiago ematen du (eskala ordinalaren ezaugarria). 6ko nota ematen duenak 2koa ematen duenak baino 4 puntu gehiago ematen du (tarte-eskalaren ezaugarria). 6ko nota eman duenak, 2ko nota eman duenaren hirukoitza ematen du (ratio-eskalaren ezaugarria).
- Bigarren aukera da alkatearekiko iritzia eskala ordinalaren arabera neurtzea: zer iritzi duzu Bilboko alkatearen lanari buruz?
  - Lan oso ona egiten du…….1
  - Lan ona egiten du………….2
  - Ez ona, ez txarra……………3
  - Lan txarra egiten du………..4
  - Lan oso txarra egiten du…..5
  - Ed-Ee…………………..…....6

Bigarren aukera horretan, inkestatuak 2 balioa (lan ona egiten du) aukeratzen badu, ez du 1 balioa aukeratzen (eskala nominalaren ezaugarria). Eta 2 balioa aukeratzen badu, 1 balioa aukeratu duenak baino iritzi ezkorxeagoa du alkatearen lanarekiko (eskala ordinalaren ezaugarria).

1.3.2.2. Aldagaiak neurtzeko moduaren arabera

Neurtzeko moduaren arabera, aldagai kualitatiboak (edo kategorikoak) eta kuantitatiboak ditugu.

Aldagai kualitatiboak edo kategorikoak

Era horretako aldagaiek nolakotasun bat (eta ez kopuru bat) neurtzen dute. Adibidez, ikasle batek azterketa batean atera duen kalifikazioa bost kategoria hauen arabera neur dezakegu: gutxiegi, gutxi, nahiko, oso ongi, bikain; edo sexua bi kategoria hauen arabera: neska edo mutila. Era horretako aldagaietan, kodifikaziorako erabiltzen ditugun zenbakiek (mutila 1, neska 2 edo gutxiegi 1, gutxi 2, nahiko 3, ongi 4, oso ongi 5, bikain 6) nolakotasunak edo kategoriak adierazten edo ordezkatzen dituzte, eta ikergai diren elementu, unitate edo pertsonak nolakotasun edo kategoria horien arabera sailkatzen. Badaude bi eratako aldagai kualitatiboak: nominalak eta ordinalak.

Adibidea. Aldagai kualitatibo nominala operazionalizatuta: ondoren, Espainiako estatuaren antolaketa-eredu desberdinak aurkeztuko dizkizugu. Esan iezadazu, mesedez, zein esaldik asetzen zaituen gehien:

- Gobernu zentral bakarra duen estatua, autonomiarik gabea…………………………..1
- Autonomia Erkidegoak dituen gaur egungoa bezalako estatua………………………..2
- Estatu federala, autonomia erkidegoek orain baino autonomia gehiago dutelarik…...3
- Nazionalitateak estatu independente izatea onartuko lukeen estatua…………………4
- Ed-Ee……………………………………………………………………………………….5

Adibidea. Aldagai kualitatibo ordinala operazionalizatuta: gobernu berriak hartzen dituen erabakiek zein eragina izango dute Euskadiko ekonomiaren bilakaeran?

- Eragin handia……………………………………………………………………………..1
- Nahikoa eragin…………………………………………………………………………….2
- Eragin txikia……………………………………………………………………………….3
- Batere eraginik ez………………………………………………………………………...4
- Ed-Ee……………………………………………………………………………………….5

Aldagai kualitatiboak dituzten kategorien kopuruaren arabera ere sailkatu daitezke. Hala, aldagai dikotomikoak eta politomikoak ditugu.
Aldagai dikotomikoetan, erantzun-aukera posibleak bakarrik bi dira. Sexua aldagaia bi kategoria ditu: gizona edo emakumea. Eta ez erantzun-aukerak sortzen dituzten galderak ere dikotomikoak dira.
Aldagai politomikoetan, berriz, erantzun-aukera posibleak bi baino gehiago dira.

Adibidea. Aldagai nominal dikotomikoa operazionalizatuta: zuk baduzu diru-arazorik hilabete bukaerara heltzeko?

- Bai………………………………1
- Ez……………………………….2
- Ed-EE…………………………..3

Adibidea. Aldagai nominal politomikoa operazionalizatuta: esango didazu honelako neurrien artean zein diren garrantzitsuenak krisi ekonomikoaren kontra borrokatzeko?

- Enplegua eta prestakuntza bultzatzea……………..……………………1
- Gazte ekintzaileei laguntzea……………………..……………………..2
- Iruzur fiskala kontrolatzea………………………..…………………...3
- Etxe-kaleratzeen kontra familiak babestea……….…..………………..4
- Enpresa txiki eta ertainei eta autonomoei laguntz.ea….………………..5
- Administrazio publikoan gastua murriztea………..……………………6
- Gizarte-laguntzak kontrolatzea…………………….…………………...7
- Enpresa zein partikularrei kreditua erraztea…….……………………..8
- Ikerketa, Garapena eta berrikuntza (I+G+B) sustatzea….……………...9
- Atzerriko inbertsioa erakartzea…………………………………………10
- Autogobernua babestea…………………………………………………11
- Beste batzuk……………………………………………………………12
- Ed-Ee……………………………………………………………………13

Aldagai kuantitatiboak edo eskalak

Aldagai kuantitatiboek edo “eskalek” ezaugarri bat neurtzen dute, kopuru edo zenbateko baten bidez (adina urteetan, adibidez). Hau da, ezaugarriak zenbakizko balioekin ordezkatzen dira, eta kodifikatzen duguna ezaugarria neurtzen duten zenbakiak dira. Era horretako aldagaien bidez, zenbait galderari erantzun diezaiokegu: zenbat urte dituzu? 23, 22, 21… Zenbat pertsona bizi dira zure etxean? 1,2,3,4.... Aldagai kualitatiboen kasuan ez du zentzurik zenbatekoa galdetzea. Zenbatekoa da ikastolak eskaintzen duen prestaketa akademikoa? Galdera horrek ez du zentzu handirik, aldagai kualitatiboa delako. Aldiz, honako galdera honek badu zentzua: nolakoa da ikastolak eskaintzen duen prestaketa akademikoa? Eta erantzuna oso ona; ona; txarra; ez ona, ez txarra; txarra, edo oso txarra izango da.

Badaude bi eratako aldagai kuantitatiboak: tarte-eskalaren arabera neurtzen dutenak eta ratio-eskalaren arabera egiten dutenak.

Adibidea. Aldagai kuantitatiboa operazionalizatuta: zenbat pertsonak irakurtzen dute egunkaria zure etxean? 1, 2, 3….

Aldagai kuantitatiboak edo “eskalak” beste era batera ere sailkatu daitezke. Hain zuzen ere, aldagai kuantitatibo diskretuak eta aldagai kuantitatibo jarraituak bereiz ditzakegu.
Aldagai diskretuek zenbakizko balio isolatuak bakarrik hartzen dituzte, hau da, balio osoak, eta tarteko balioek ez dute zentzurik: seme-alaben kopurua, etxeko ordenagailu kopurua, herri bateko biztanle kopurua, adibidez. Zenbat seme-alaba, ordenagailu, biztanle daude? Badu zentzua galdera hauei 1, 2, 3…1000 erantzutea, baina ez du zentzurik 1,2; 2,5 edo 10000,5 erantzutea.

Adibidea. Aldagai diskretua operazionalizatuta: zenbateko konfiantza duzu Eusko Jaurlaritza berrian? (0=Batere konfiantzarik ez; 10=Konfiantza osoa).

Aldagai jarraituak tarte batean edozein balio erreal har dezaketenak dira; hau da, balio ez-osoak ere har ditzaketenak: gizakien altuera eta pisua, tenperatura, abiadura, dentsitatea edo presioa aldagai jarraituak dira. Erosoago lan egiteko, aldagai estatistiko jarraituak tarteka taldekatzen eta aurkezten dira.
Kontuan izan, ia gehienetan, zenbaketak aldagai diskretuak izaten direla eta neurketak (luzerak eta pisuak, adibidez) jarraituak. Aipatzekoa ere bada ohikoa dela aldagai diskretuak jarraitu gisa

aztertzea; hau da, aldagai jarraitu gisa tarteka aurkezten dira, eta horietarako dauden prozedura estatistiko berak erabiltzen dira.

1.3.2.3. Aldagaiak, beren arteko harremanen arabera

Aldagai arteko erlazioak edota mendekotasunak aztertu behar direnean, aldagai askeen eta mendekoen artean bereizten da[4]:
Aldagai askea da baldin eta beraren balioak ikertzaileak berak kontrolatzen baditu edo beste aldagai baten mendean ez badago. Funtzio matematikoetan, aldagai askea x hizkiarekin adierazten da eta abzisen ardatzean irudikatzen.
Mendeko aldagaia da baldin eta beraren balioak aldagai askeak hartzen dituen balioen arabera aldatzen badira. Funtzio matematikoetan, mendeko aldagaia y hizkiarekin adierazten da, eta ordenatuen ardatzean irudikatzen.

Adibidea. Estatistika ikasten emandako ordu kopuruaren eta azterketan lortutako kalifikazioaren arteko erlazioa aztertzean, ikasten emandako ordu kopurua (x) aldagai askea da, eta lortutako nota (y) mendekoa. Matematikoki, honela adieraziko genuke harreman hori: $y = f(x)$.

1.3.2.4. Aldagaiak, edukiaren arabera

Edukiaren arabera, aldagai numerikoak eta alfanumerikoak bereiz ditzakegu.
Aldagai numerikoak dira zenbakien bidez kodifikatzen direnak. Hau da, aldagaien kategoriak (nominala denean) edo balioak zenbakizkoak direnean. Beraz, aldagai numerikoak aldagai kualitatiboak zein kuantitatiboak izan daitezke.

Adibidea. Aldagai numerikoa: sexua, aldagai kualitatiboa, nominala eta numerikoa da, modu honetan kodifikatzen badugu: 1 emakumea eta 2 gizonezkoa. Besteak beste, diru-sarrerak, adina eta ordenagailu kopurua aldagaiak ere numerikoak dira, informazioa zenbakien bidez biltzen edota kodifikatzen dugulako.

Aldagai alfanumerikoak edo kadena erakoak dira zenbakien bidez kodifikatu beharrean, karaktere alfanumerikoen bidez kodifikatzen direnak; hau da, bakarrik letrak erabiliz edo letrak eta zenbakiak nahasiz kodifikatzen direnak. Sexua aldagaia zenbakien bidez kodifikatu beharrean, gizona eta emakumea idatziz kodifika dezakegu. Ordenagailuan, beraz, zenbakien ordez, hitzak sartzen dira.

[4] Aldagai mota horiek azaltzea ikastaro honetatik haratago doa, ikastaro honetan aldagai bakarreko estatistika azaltzen dugulako; hala ere, kontzeptuak finkatze aldera, interesgarria iruditu zaigu azalpen hau ematea.

Adibidea. Aldagai alfanumeriko edo kadena erako aldagaia: Euskal AEko jaioberrien artean ohikoenak diren izenak ezagutzeko, jaioberri guztien izenak kodifikatzen ditugu, eta ostean maiztasunak atera.

1. taula: Azken hiru urteetan (2008-2010) gehien erabilitako 25 ponte-izenen zerrenda

	Emakumeak			Gizonak		
	Izena	Maiztasuna	%	Izena	Maiztasuna	%
1	Ane	928	3,02	Iker	1.001	3,06
2	June	828	2,70	Markel	925	2,83
3	Uxue	730	2,38	Oier	867	2,65
4	Irati	713	2,32	Jon	865	2,65
5	Nahia	710	2,31	Unai	843	2,58
6	Izaro	587	1,91	Ander	760	2,32
7	Naroa	538	1,75	Mikel	748	2,29
8	Nora	532	1,73	Eneko	736	2,25
9	Naia	520	1,69	Aimar	733	2,24
10	Leire	515	1,68	Unax	635	1,94
11	Lucía	490	1,60	Asier	589	1,80
12	Maialen	487	1,59	Ibai	568	1,74
13	Paula	467	1,52	Julen	541	1,65
14	Haizea	435	1,42	Aitor	492	1,50
15	Sara	429	1,40	Xabier	447	1,37
16	Eider	391	1,27	Beñat	381	1,17
17	Maddi	355	1,16	Alain	344	1,05
18	Iraia	317	1,03	Aritz	317	0,97
19	Nerea	301	0,98	Oihan	317	0,97
20	Maria/María	286	0,93	Martin/Martín	302	0,92
21	Malen	276	0,90	Danel	298	0,91
22	Maider	274	0,89	Gorka	288	0,88
23	Noa	266	0,87	Urko	284	0,87
24	Enara	265	0,86	Adrian/Adrián	277	0,85
25	Irene	259	0,84	Aner	277	0,85

Iturria: EUSTAT. Jaiotzen estatistika

Taulan ikus daitekeen bezala, 2008. urtetik 2010.era, Euskal AEn neskentzat gehien erabili den ponte-izena Ane da. Bigarren eta hirugarren postuetan June eta Uxue daude. Mutilen artean, gehien agertzen den ponte-izena Iker da, eta, horren atzetik, Markel eta Oier.

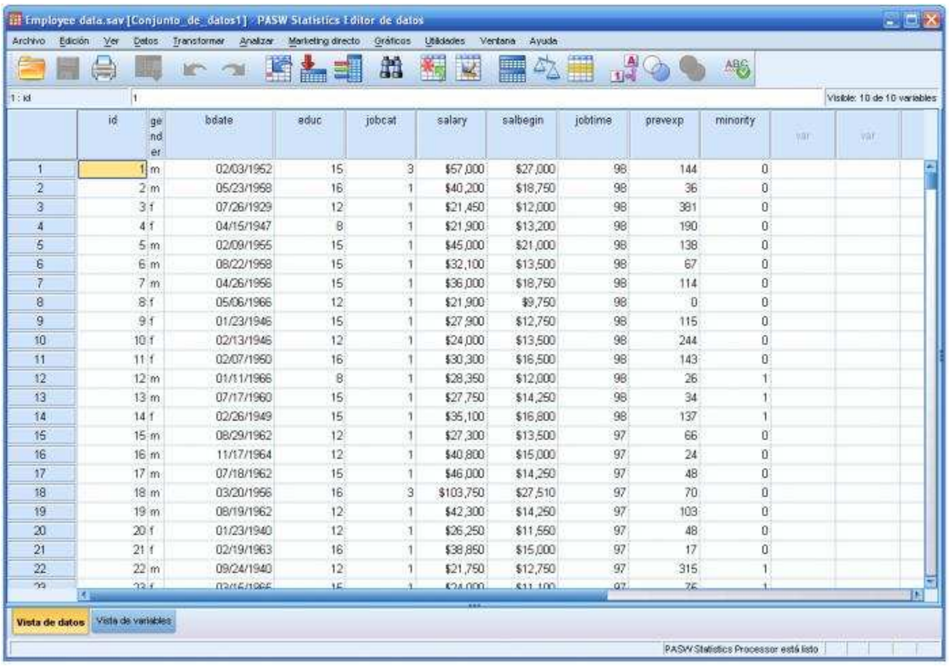
1.4. ESTATISTIKA, PROGRAMA INFORMATIKOAK ETA IBM SPSS PROGRAMA INFORMATIKOA

Arestian esan bezala, gizarteen egoera eta bilakaera ezagutzeko, ikerketa-prozesua, ikerketarako teknikak eta datu multzoak kudeatzeko eta aztertzeko prozedurak ezagutu behar ditugu, baina, horrekin batera, ordenagailurako estatistika-programak erabiltzen ere jakin behar dugu. Izan ere, ordenagailuek eta programa informatikoek elementu, unitate edo pertsonen multzo bati dagozkion zenbakizko datuak biltzeko, sailkatzeko eta aztertzeko laguntza paregabea ematen digute:

- Ordenagailuek ikerketarako erabiltzen diren prozeduren fidagarritasuna ziurtatzen dute. Alegia, egiten dituzten kontaketak guztiz fidagarriak dira.
- Lan errepikakorrak (zenbaketak, gurutzaketak, kalkuluak…) arintzen dituzte.
- Datu-diagramak erraz, bizkor eta kalitate handiarekin egiten dituzte.
- Datu kopuru erraldoiak denbora gutxian aztertzen dituztenez, ordenagailuek bide berriak urratu dituzte datuen azterketan. Besteak beste, aldagai anitzen azterketak asko erraztu eta ugaritu dituzte.
- Horren guztiaren ondorioz, dirua eta denbora aurrezten laguntzen dute.

Asko dira datu kuantitatiboak aztertzeko programa informatikoak. Besteak beste, IBM SPSS, SPAD, Sphinx Survey, R (askea), SAS, MATLAB, Statistica eta Stata aipa ditzakegu. Bakoitzak bere berezitasunak, ahultasunak, desabantailak eta abantailak dituela ahantzi gabe, aldagai kualitatibo zein kuantitatiboak aztertzeko neurri estatistikoak eta prozedurak eskaintzen dituzte; baita bildutako informazioa grafikoki irudikatzeko aukera ugari ere.
Sarreran esan bezala, liburu honetan ez ditugu estatistika programen berezitasunak agertuko. Hala ere, aurkezten ditugun adibide eta azterketa praktiko asko IBM SPSS programaren bidez eginda daude, besteak beste, UPV/EHUk irakaskuntzarako programa horren aldeko hautua egin duelako. 35 urte baino gehiago dituen IBM SPSS Statistics programaren abantaila handiena da erraz erabiltzen dela eta datu multzo handiak aztertzeko gaitasuna duela.

Adibidea. 1. diagrama: IBM SPSS programaren datuen editorea



2. ESTATISTIKA DESKRIBATZAILEA

Edozein zientziatan bezala, Gizarte Zientzietan ere, oso garrantzitsua da ikerketa ondo diseinatzea eta kanpo-lana ahal den errakuntza gutxienarekin egitea. Balioko ondorioak ateratzeko, ezinbestekoa da galdeketako galderak ondo pentsatu eta idaztea, lagina ondo diseinatzea, lagineko elementu, unitate edo pertsonak aukeratzerakoan zoria ziurtatzea eta, oro har, datu-bilketa prozesuan erroreak gutxitzea (laginketa-errorea eta errore sistematikoak gutxitzea[5]).

Ikergai izango den datu multzoa osatu dugunean hasten da ikerketa estatistikoa.

Azterketa estatistikoari ekiteko, argi izan behar dugu lortutako datu multzoa eta haren berezitasunak ezin ditugula pilatuta agertu, inolako ordenarik gabe, ondorioak ateratzea edota interpretazioak egitea zailtzen delako. Hain zuzen ere, datu multzoak modu ulerterrazean aurkeztea da estatistika deskribatzailearen helburuetariko bat, eta, horretarako, hiru baliabide nagusi eskaintzen dizkigu:

- Maiztasun-taulak edo maiztasun-banaketak
- Datu-diagramak
- Neurri estatistikoak

Bigarren atal honetan, beraz, hiru baliabide horiek aurkeztuko ditugu, estatistika deskribatzaileak eskaintzen diguna azaltzeko asmoz. Bukaeran, banaketa normalaren berezitasunak ere aztertuko ditugu.

2.1. MAIZTASUN TAULAK EDO MAIZTASUN BANAKETAK

Arestian adierazi bezala, neurtzea naturan edo gizartean dauden gauzen, elementuen edo pertsonen ezaugarriei zenbakizko balioak ematea da.

Aldagai kualitatiboek edo kategorikoek nolakotasun bat (eta ez kopuru bat) neurtzen dute; kategorietan banatuta daude, eta ikergai diren gauza, elementu edo pertsonak kategoria horien arabera sailkatzen edota ordenatzen dituzte.

Aldagai kuantitatiboek ezaugarri bat neurtzen dute, kopuru edo zenbateko baten bidez; eta ikergai diren gauza, elementu edo pertsonak eskala batekiko ematen duten balioaren arabera sailkatzen dituzte. Kategorien arabera sailkatu beharrean, gauza, elementu edo pertsonak zenbakizko balioen arabera sailkatzen dira.

Argi dago kategoriak eta balioak aldagaiak har ditzakeen aukera desberdinak direla, eta gauza, elementu edo pertsonak horien arabera neurtzen ditugula. Sexua aldagaiak, adibidez, bi kategoria ditu: emakumezkoa eta gizonezkoa. Eta 0tik 10rerako eskalak 11 balio ditu: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, eta 10.

Datuak, berriz, kategoria edo balio bakoitzerako lortu diren emaitzak edo maiztasunak dira. Adibidez, elkarte batean sexu aldagaia aztertu ostean, guztira 204 emakumezko eta 307 gizonezko daudela ikusten dugu; edo 0tik 10rerako eskalan, 205 pertsonak (%40k) 4 balioa aukeratu dutela. Maiztasun horiek dira gure datuak; edo nahi bada, ikerketarako erabiliko dugun datu multzoa.

Nahiz eta askotan nahasten diren, garrantzitsua da datuen, kategorien eta balioen artean bereiztea.

Gauza, elementu edo pertsonak aldagai kuantitatibo edo kualitatiboen arabera neurtu ditugunean, eta datu multzo bat sortu dugunean (inkesta bat egin dugu, adibidez), datu horiek ordenatu egin behar ditugu. Hau da, aldagaiaren kategoria (aldagai kualitatiboetan) edo balio

[5] Liburuaren bukaeran gai hori hobeto aztertuko dugu; hala ere, garrantzitsua da kontuan izatea laginaren tamaina zenbat eta handiago izan laginketa-errorea orduan eta txikiagoa dela. Errore sistematikoak gutxitzeko, erabili beharreko teknikak, prozedurak, galdeketak zein inkestatzaileak ondo pentsatu eta prestatu behar dira.

(aldagai kuantitatiboetan) desberdinak zein diren (zerrenda) eta horietariko bakoitza zenbat aldiz agertzen den (maiztasunak) erakutsi behar dugu[6]. Horretarako, ezinbestekoak dira maiztasun-banaketak edo maiztasun-taulak. Egiatan, maiztasun-taulak eraikitzea da azterketa estatistikoaren lehen pausoa.

2.1.1. Maiztasun-banaketen edo maiztasun-taulen berezitasunak

Maiztasun-taula edo maiztasun-banaketa datu multzo batean balio edo kategoria bakoitza (x_i hizkiarekin adierazten dena) zenbat aldiz agertzen den adierazten duen taula da. Kategoria edo balio bakoitzean zenbat datu dauden zehaztea da haren helburua, modu absolutuan, erlatiboan edo metatuan datu multzoaren izaera modu ulerterrazean erakusteko.
10 ikaslez osatutako 4. DBHko gela batean zenbat mutil eta zenbat neska dauden erakusteko, datu multzoa honela aurkez dezakegu: gizona, gizona, gizona, gizona, gizona, emakumea, emakumea, emakumea, emakumea eta emakumea. Modu berean, 10 ikasle horien adina era honetan ager dezakegu: 16, 15, 16, 15, 16, 15, 16, 15, 16, 16.
Datu multzoa horrela azalduta, ondorioak ateratzea zaila da. Balio bakoitza zenbat aldiz agertzen den jakiteko, zenbaketak egin behar ditugu, eta hori lan neketsua da, bereziki datu multzo handiak ditugunean.
Gure txostenen irakurleen lana errazteko, beraz, hobe da balio bakoitza zenbat aldiz agertzen den adierazten duten taulak egitea; hau da, maiztasun-taulak egitea.
Maiztasunak datu multzoko balio edo kategoria bakoitza zenbat aldiz agertzen den adierazten duten zenbakiak dira. Maiztasun-tauletan, lau eratako maiztasunak ager daitezke:

- Maiztasun absolutuak (f_i) kategoria edo balio bakoitza zenbat aldiz azaltzen diren adierazten duten zenbakiak dira. Aldagaiko i-garren balioari edo kategoriari (xi) dagokion maiztasun absolutua f_i hizkiaz izendatu ohi da. Maiztasun absolutu guztien batuketa datu guztien kopuruaren berdina da, eta n hizkiarekin adierazten da: $f_1 + f_2 + f_3 + \ldots + f_n = n$.
- Maiztasun erlatiboak (h_i) lortzeko, kategoria edo balio bakoitzaren maiztasuna (f_i) laginaren (n) edo populazioaren (N) kopuruarekin zatitzen dugu:

$$h_i = \frac{f_i}{n}$$

  Maiztasun erlatiboak ezin dira izan 1 baino handiagoak, eta maiztasun-tauletako maiztasun erlatibo guztien batuketa beti da 1. Aldagaiko i-garren balioari edo kategoriari (x_i) dagokion maiztasun erlatiboa h_i hizkiaz izendatu ohi da. Kontuan izan maiztasun erlatiboak bi eratakoak izan daitezkeela: proportzioak eta ehunekoak.
  Proportzioak lortzeko, kategoria edo balio bakoitzaren maiztasuna (f_i) laginaren (n) edo populazioaren (N) kopuruarekin zatitzen dugu. Esan bezala, proportzioak ezin dira izan bat baino handiagoak.
  Ehunekoa proportzioa adierazteko modua da. Ehunekoak lortzeko proportzioak bider 100 egin behar dira. Datu multzo bateko ehunekoak ezin dira izan 100 baino handiagoak.
- Maiztasun absolutu metatua (F_i): balio edo kategoria bati dagokion maiztasun absolutu metatua, balio edo kategoria horren maiztasun absolutua eta aurreko maiztasun absolutu guztiak batuta lortzen dugun zenbakia da. Aldagaiko i-garren balioari edo kategoriari (x_i) dagokion maiztasun absolutu metatua F_i hizkiaz izendatu ohi da. Kontuan izan, era horretako maiztasunak bakarrik aldagai kuantitatiboetan kalkula daitezkeela.
- Maiztasun erlatibo metatua (H_i): balio edo kategoria bati dagokion maiztasun erlatibo metatua, balio edo kategoria horren maiztasun erlatiboa eta aurreko maiztasun erlatibo guztiak batuta lortzen dugun zenbakia da. Aldagaiko i-garren balioari edo kategoriari (x_i)

[6] Lortutako emaitzak sakabanatuegiak badira, taldekatu egiten dira eta, kontrako arrazoirik ez badago, maiztasun txikiena duten balioak aukera berezi batean biltzen dira.

dagokion maiztasun erlatiboa metatua H_i hizkiaz izendatu ohi da. Era horretako maiztasunak ere aldagai kuantitatiboetan bakarrik kalkula daitezke.

Hala, maiztasun-taula batean, sei eratako maiztasunak aurkitu ditzakegu:

- maiztasun absolutuak (f_i)
- maiztasun absolutuak metatuak (F_i)
- maiztasun erlatiboak, proportzioak (h_i)
- maiztasun erlatibo metatuak, proportzioak (H_i)
- maiztasun erlatiboak, ehunekoak (%)
- maiztasun erlatibo metatuak, ehunekoak (%)

2. taula: maiztasun-tauletako maiztasun motak

x_i (balioak edo kategoriak)	f_i (maiztasun absolutuak)	F_i (maiztasun absolutuak metatuak)	h_i (maiztasun erlatiboak, proportzioak)	H_i (Maiztasun erlatibo metatuak, proportzioak)	% (maiztasun erlatiboak, ehunekoak)	% (maiztasun erlatibo metatuak, ehunekoak
X_1	f_1	F_1	h_1	H_1	h_1	H_1
X_2	f_2	F_2	h_2	H_2	h_2	H_2
X_3	F_3	F_3	h_3	H_3	h_3	H_3
.	.	.	.	.	.	.
.	.	.	.	.	.	.
X_n	F_n	N	H_n	1	H_n	100
	N		1		100	

Iturria: lanketa propioa

Sei maiztasun horien bidez, datu multzoaren ezaugarri nagusiak erakusten ditugu: besteak beste, zein den maiztasun handiena duen kategoria edo balioa, datuak zein kategoria edo balioren inguruan biltzen diren eta datuak zenbateraino sakabanatzen diren.

IBM SPSS edo antzerako programa estatistiko informatikoek oso erraz ematen dizkigute maiztasun horiek guztiak. Dena dela, kontuan izan behar dugu horiek modu desberdinean erabiltzen direla aldagai motaren arabera. Ostean ikusiko dugun bezala, maiztasun absolutuak eta erlatiboak aldagai kualitatiboetan zein kuantitatiboetan agertuko dira, eta aldagai kuantitatiboetan baita maiztasun metatuak ere.

2.1.2. Aldagai kualitatiboen maiztasun-banaketak edo maiztasun-taulak

Aldagai kategorikoen, atributuen edo kualitatiboen kasuan, maiztasun-taulak maiztasun absolutuak eta erlatiboak agertzen ditu; eta, nahi izanez gero, kategoriak maiztasun handienetik txikienera ordenatzen dira.

Adibidea. Aldagai kualitatibo baten maiztasun-banaketa: 5 urte edo gehiagokoen hizkuntza-gaitasuna Gasteizen (2006) (maiztasun absolutuak eta erlatiboak)

3. taula: 5 urte edo gehiagokoen hizkuntza-gaitasuna Gasteizen (2006)

x_i (kategoriak)	Maiztasun absolutua (f_i)	Maiztasun erlatiboa (proportzioak) (h_i)	Maiztasun erlatiboa (ehunekoak) (%)
Elebidunak	53.349	0,24	24
Elebidun hartzaileak	38.990	0,18	18
Erdaldun elebakarrak	125.257	0,58	58
5 urte edo gehiagoko biztanleria guztira (N)	217.596	1	100

Iturria: lanketa propioa, Biztanleria eta etxebizitza estatistiketatik hartutako datuetatik (www.eustat.es)

Esan bezala, maiztasun erlatiboa lortzeko, kategoria bakoitzari dagokion maiztasun absolutua (f_i) elementu, unitate edo pertsona kopuru osoarekin (N) edo laginaren tamainarekin (n) zatitzen dugu.
Taulan ikusten den bezala, bi eratako maiztasun erlatiboak daude (proportzioak eta ehunekoak), eta biak erabil ditzakegu. Hala ere, txostenak idatzi edo azterketak egiten direnean, ehunekoak hobesten dira, hobeto irakurtzen direlako.
Bestalde, maiztasun absolutuak eta erlatiboak, biak, aurkezten dira maiztasun-tauletan, baina, interpretazioak egitean, maiztasun erlatiboak gehiago erabiltzen dira. Izan ere, maiztasun absolutuetan oinarrituta zailagoa da kategorien artean edo datu multzo desberdinen artean alderaketak egitea.
Autore batzuen arabera, gutxienez, 20 datuko datu multzoa behar dugu ehunekoak ateratzeko.

Adibidea. Ez da gauza bera 53.000 elebidun izatea aztertzen ari gizartearen biztanleria osoa 217.596 denean edo 340.000 denean. 53.000 elebiduneko maiztasun absolutua, batean eta bestean, berdina da, baina kopuru batak eta besteak ez dute esanahi bera kasu bakoitzean.

Tamaina desberdina duten populazioetako maiztasunak alderatu nahi ditugunean, beraz, maiztasun erlatiboak erabiltzen dira. Hala, maiztasun absolutuen zutabeari maiztasun erlatiboen zutabea gehitzen zaio.

2.1.3. Aldagai kuantitatiboen maiztasun-banaketak edo maiztasun-taulak

Aldagai kuantitatiboen kasuan, maiztasun-taulek balio bakoitza zenbat aldiz agertzen den adierazten digute, balioak txikienetik handienera ordenatuta, gehienetan. Dena dela, taulak aurkezteko modua desberdina izan daiteke aldagai jarraituen eta diskretuen kasuan. Hori bai, aldagai kualitatiboekin ez bezala, aldagai jarraituekin eta diskretuekin, maiztasun metatuak sar ditzakegu.

Aldagai kuantitatibo jarraituak

Tarte batean edozein balio erreal har dezaketenak dira aldagai jarraituak; hau da, balio ez-osoak ere hartzen dituztenak. Era horretako aldagaiak ditugunean, maiztasun-taulak balio erreal bakoitza zenbat aldiz agertzen den adierazten digu, modu absolutuan zein erlatiboan.

Adibidea. Euskal Herriko herrien euskararen kale-erabilera: Euskal Herriko herrien lagin bat osatu ostean, aukeratutako herri bakoitzean euskaraz hitz egiten dutenen ehunekoa neurtu dugu behaketa zuzenaren bidez (%45,6; %56,3; %67,6…). Beraz, aldagai hori jarraitua da, ehunekoak balio ez-osoak ere har ditzakeelako. Lortutako balio errealak eta bakoitzari dagokion maiztasuna era honetako taula batean jar dezakegu.

4. taula: Euskara kalean erabiltzen dutenen ehunekoak	
Herri horretako elebidunen ehunekoa (x_i)	Herri kopurua (f_i)
.	f_1
.	f_2
.	f_3
.	.
56,7	f_{27}
67,4	f_{28}
78,5	f_{29}
89,7	f_{30}
.	.
.	.
X_n	f_n
Iturria: EUSTAT. Udaletako biztanle-estatistika (www.eustat.es)	

Aldagai jarraituak balio desberdin asko hartzen dituenean, maiztasun-banaketa luzea suerta daiteke, eta interpretatzen zaila. Ulerterraztasunaren eta lan egiteko erosotasunaren izenean, balio guztiak aurkeztu beharrean, tartetan biltzen dira. Hau da, aldagaiaren balioak tartetan sailkatzen eta aurkezten dira.

Adibidea. Euskal Herriko herrietan euskararen kale-erabilera aldagaiak hartzen dituen balioak (0,34, 0,35, 0,56…) 4 tartetan (25 balio tarte bakoitzean) multzokatu ditzakegu, eta lau zonalde soziolinguistiko sortu.

5. taula: kale-erabilera, zonalde linguistikoaren arabera. 2011		
Zonalde Soziolinguistikoa euskaraz hitz egiteko gaitasuna dutenen arabera	f_i	%
0-25	f_1	3,0%
25-50	f_2	14,4%
50-75	f_3	39,7%
75-100	f_4	66,5%
Iturria: lanketa propioa, 2011ko kale erabilerako datuetan oinarrituta		

Aldagai jarraituen balioak tartetan banatzeko, honako pauso hauek eman behar ditugu:

1. Tarte kopurua zehaztu (h): lehenengo eta behin, tarte kopurua erabaki behar da. Ez dago tarte kopuru hoberenik. Erabakia aldagaiaren izaeraren, datu multzoan agertzen diren balio kopuruaren, balio horiek tartetan bilduz lortu nahi diren helburuen eta ikerlariaren nahien arabera aldatzen da. Tarte kopuru desberdinek datu multzoaren ezaugarri desberdinak erakuts ditzakete, eta, beraz, komeni da tarte kopuru desberdinak sortzea, datu multzoaren ezaugarri aipagarrienak azaltzen dituena aukeratu arte. Tarte kopurua aukeratzerakoan, kontuan izan behar dugu hortik aterako diren taulek eta egingo diren grafikoek argiak eta zehatzak izan behar dutela. Ildo horretatik, tarte gutxi egiten badira, informazio-galera handia gertatzen da; eta, tarte gehiegi egiten badira, berriz, sortzen den maiztasun-taula nahasia izan daiteke, datuak behar adina bildu ez direlako. Hala, autore gehienek diote maiztasun-taulek gutxienez 5 tarte eta gehienez 15 izan behar dituztela. Izan ere, 5 tarte baino gutxiagorekin, datuak tartetan biltzeak dakarren informazio-galera handiegia da. Aldiz, 15 tarte baino gehiago eratzean, maiztasun-banaketaren bitartez lortu nahi den informazio-laburpena hutsean gelditzen da. Zenbat eta tarte gehiago sortu, orduan eta zehaztasun handiagoa izango dugu, baina argitasun gutxiago.
2. Tarte kopurua erabaki ostean, tarteek zer-nolako zabalera izango duten erabaki behar dugu (k). Horretarako, lortutako balio handienari txikiena kentzen zaio; hau da, ibiltartea kalkulatzen da. Ostean, ibiltartea aldez aurretik erabaki dugun tarte kopuruaz (h) zatitzen dugu. Kalkulu neketsuak alde batera uzteko, lortu den balioa biribildu egiten da:

$$k = \frac{x_{\max} - x_{\min}}{h}$$

Adibidea. Aldagai jarraituen balioak tartetan banatzeko prozedura: adina ikertzen ari bagara, eta gure datu multzoan balio handiena eta txikiena 90 eta 5 badira, eta 10 tarte egitea erabakitzen badugu, tartearen zabalera modu honetan kalkulatuko dugu:

$$k = \frac{90 - 5}{10} = 8{,}5$$

Emaitza biribiltzen badugu, 9 urteko zabalera izango du tarte bakoitzak: 5-13, 13-21, 21-29…

Kontuan izan adina aldagai kuantitatibo diskretua dela, baina, balio asko dituenez, aldagai jarraitu gisa aztertzen dugula.
Aurkeztu duguna tarteak egiteko prozedura orokorra da, baina datu multzoan balioren bat oso sakabanatua badago, zabalera desberdineko tarteak sor daiteke muturretan. Modu berean, ibiltarteko zatiren batean zehaztasun handiagoa lortu nahi bada, tarte batzuen zabalera txikitu egin daiteke.

Adibidea. Aldagai jarraituen balioak tartetan banatzeko prozedura: lau hilabete dituzten hiri bateko 40 umez osatutako lagina osatu dugu, eta, guztiak pisatu ostean, honako datu multzo hau lortu dugu: 6,5 – 6,3 - 5,4 – 6,5 – 6,0 – 7,5 – 6,7- 7,5 – 6,0 – 6,8 – 5,9 – 6,3 – 6,6 – 6,2 – 6,1- 6,0 – 4,9 – 5,8- 6,8 – 6,7 – 6,4 – 6,7 – 5,0 – 6,5 – 7,0 – 7,1 – 6,3 – 5,8 – 6,9 – 7,8 – 7,0 – 6,9 – 7,2 – 7,3 – 6,5 – 7,4 – 5,6 – 6,3 – 6,5 – 6,2.

Pisu horiek guztiak taldekatu gabe, era honetako taula luze eta aztergaitza lortuko genuke:

6. taula: hiri horretako umeen pisua	
Pisua (x_i)	Kopurua (f_i)
4,9	1
5,0	1
5,4	1
5,6	1
5,8	2
5,9	1
6,0	3
6,1	1
6,2	2
6,3	4
6,4	1
6,5	5
6,6	1
6,7	3
6,8	2
6,9	2
7,0	2
7,1	1
7,2	1
7,3	1
7,4	1
7,5	1
7,6	1
7,8	1
	40
Iturria: lanketa propioa	

Datu multzoa modu horretan aurkeztuta, ondorioak ateratzea edota interpretazioak egitea zaila da. Horregatik, garrantzitsua da datu multzoa tartetan banatuta aurkeztea.
Lortutako pisuak 5 tartetan banatzea erabakitzen badugu, tarte bakoitzaren zabalera kalkulatzeko zera egingo dugu:

$$k = \frac{7{,}8 - 4{,}9}{5} = 0{,}58 \approx 0{,}6$$

Biribilduz, tarte bakoitzak 0,6 kiloko zabalera izango du, eta, beraz, honelako 6 tarte aterako dira: 4,8-5,4; 5,4-6; 6-6,6; 6,6-7,2; 7,2-7,8 eta 7,8-8,4.

7. taula: hiri horretako umeen pisua	
Pisua (x_i)	Kopurua (f_i)
(4,8-5,4]	3
(5,5-6]	7
(6-6,6	14
(6,6-7,2]	11
(7,2-7,8]	5
(7,8-8,4]	-
	40
Iturria: lanketa propioa	

Horrela eginez gero, tarte bakoitzeko muturretako balioekin, arazoak izango genituzke. Hau da, zein tartetan sartzen da 5,4 balioa? Horrelako arazoak saihesteko, tarteko muturretariko bat ireki egiten da, eta bestea, itxi: (4,8-5,4]; (5,4-6]; (6,6-6].... Kasu horretan, 5,4 kilogramo dituen umea lehenengo tartean sartzen da, eta ez bigarrenean. Izan ere, kortxeteak adierazten du tarte hori 5,4 kilorekin ixten dela.

Tarteak alderantziz itxi eta irekiko bagenitu; alegia, modu honetan antolatuko bagenitu: [4,8-5,4); [5,4-6); [6,6-6)… 5,4 balioa bigarren tartean, [5,4-6) tartean, sartuko litzateke. Izan ere, kortxeteak adierazten du tarte hori 5,4 balioarekin hasten dela.

Liburuan zehar, ez dugu kortxeterik eta parentesirik erabiliko datu multzoa tartetan banatuta aurkezten badugu. Dena dela, onartuko dugu tartearen beheko mugan kortxetea dagoela eta goiko mugan, parentesia. Hurrengo adibidean, beraz, 20 balio bigarren tartean sartzen da, 30 balioa hirugarrenean, 40 balioa bosgarrenean…

8. taula: xake-elkarte bateko kideen adinak	
	f_i
[10, 20)	1
[20, 30)	8
[30,40)	10
[40, 50)	9
[50, 60)	8
[60,70)	4
[70, 80)	2
N	42
Iturria: lanketa propioa	

Egia da balioak tartetan aurkezteak informazio-galera esan nahi duela. Hain zuzen ere, 20-30 adin-tartean 8 pertsona daudela badakigu, baina ez dakigu horietariko bakoitza zein adinetakoa den. Hala ere, balioak tartetan banatuta, taula ulerterrazagoak eta lanerako erosoagoak lortzen ditugu.

Aldagai kuantitatibo diskretuak

Aldagaiak bakarrik zenbakizko balio isolatuak hartzen dituenean, aldagai diskretuak dira. Kasu horretan, aldagaiak balio osoak hartzen ditu, eta tarteko balioek ez dute zentzurik.

Balio osoak izanik, ez du zentzurik horiek elkartzea; hau da, ez du zentzurik balioak tartetan banatuta aurkeztea. Dena dela, aldagaiak balio asko dituenean, hau da, balio sakabanatuegiak ditugunean, eta, horren ondorioz, kudeaketarako eta interpretaziorako arazoak ditugunean, balio horiek taldekatu egiten dira, arestian adinarekin egin bezala. Kasu horretan, maiztasun absolutuen taulak balio multzo bakoitza zenbat aldiz azaltzen den adierazten digu. Taldekatuta, ulerterraztasuna eta argitasuna lortu nahi ditugu.

Adibidea. Euskal AEko biztanleen adina: erroldaren bidez, Euskal AEko biztanle bakoitzaren adina ezagutzen dugu, eta, beraz, posible da honelako maiztasun absolutuen taula lortzea:

9. taula: Euskal AEko biztanleria adinaren arabera (1-I-2011)	
x_i	f_i
0	20.920
1	21.017
2	21.615
3	20.925
.	.
.	.
.	.
.	.
83	12.567
98	442
99	281
100 eta gehiago	468
Guztira (N)	2.174.033
Iturria: EUSTAT. Udal biztanle-estatistika (www.eustat.es)	

Lortutako balioak sakabanatuegiak direnez (0tik 100era ibiltartea luzea da), lanerako taula luzeegia lortuko genuke. Beraz, nahitaezkoa da balioak taldekatzea, eta horrelako maiztasun absolutuen taula sortu.

10. taula: Euskal AEko biztanleria adin-talde handien arabera. 1-I-2011			
x_i	F_i	H_i	%
0-19	378.257	0,1739	17,4
20-64	1.371.478	0,6308	63
>=65	424.298	0,1951	19,6
Guztira	2.174.033	1	100
Iturria: EUSTAT. Udaletako biztanle-estatistika (www.eustat.es)			

Ikusten den bezala, aldagai diskretua izanik, aurreko tarteko azken balioa eta hurrengo tarteko balioa ez dira errepikatzen. Aldagaiak balio osoak hartzen dituenez, eta tarteko balioek zentzurik ez dutenez, tarteetan ez dira balioak gurutzatzen. Hori bakarrik gertatzen da aldagai jarraituak ditugunean, edo, esan bezala, aldagai diskretua jarraituak bezala aztertzea erabakitzen dugunean.
Bestalde, kontuan izan aldagai kuantitatibo diskretuen maiztasun-taulek ere maiztasun absolutuak eta erlatiboak erakusten dituztela.

Maiztasun bakunak eta metatuak

Orain arte ikusitakoaren arabera, aldagai kuantitatibo zein kualitatiboetan, lehenengo zutabean beti aldagaiaren kategoriak edo balioak aurkezten dira; ostean, maiztasun absolutuak eta erlatiboak (proportzioak edota ehunekoak) agertzen dira.
Aldagai kuantitatiboen kasuan, maiz agertzen dira beste maiztasun batzuk: maiztasun metatuak. Hala, aldagai kuantitatiboekin lanean ari garenean, maiztasun bakunen eta metatuen artean bereizi beharra dago.
Balio bati dagokion maiztasun bakuna balio horri, eta bakarrik berari, dagokion maiztasuna da; hau da, balio hori hartzen duten datuen kopurua da.
Balio bati dagokion maiztasun metatua balio horri eta haren azpitik egon daitezkeen balioei dagozkien maiztasunen batura da.

Maiztasun metatuak maiztasun absolutuekin zein erlatiboekin atera daitezke. Horrela, maiztasun absolutu metatuak (F_i) eta maiztasun erlatibo metatuak (H_i) kalkula ditzakegu.

Maiztasun metatuak kalkulatzeko, lehenengo eta behin, aldagaiak hartzen dituen balio desberdinak txikienetik handienera ordenatu behar dira. Eta hori aldagaia kuantitatiboak direnean, behintzat, egin daiteke, baita aldagai kualitatibo ordinalekin ere.

Balioak ordenatu eta gero, balio bakoitzaren maiztasunak eta haren azpitik gelditzen diren balioenak zenbatu behar dira. Hala, maiztasun metatuek honako propietate hau betetzen dute: balioa zenbat eta handiagoa izan, orduan eta handiagoa da maiztasun metatua; hain zuzen ere, balioa gehitu ahala datuak biltzen direlako.

Adibidea. Euskal AEko biztanleria adin-talde handien arabera: maiztasun absolutu, erlatibo eta metatuak

11. taula: Euskal AEko biztanleria adin-talde handien arabera (1-I-2011)						
x_i (balioak)	f_i (maiztasun absolutuak)	F_i (maiztasun absolutuak metatuak)	h_i (maiztasun erlatiboak, proportzioak)	H_i (Maiztasun erlatiboak metatuak, proportzioak)	% (maiztasun erlatiboak, ehunekoak)	% (maiztasun erlatiboak metatuak, ehunekoak
0-19	378.257	378.257	0,1739	0,1739	17,4	17,4
20-64	1.371.478	1.749.735	0,6308	0,8047	63	80,4
>=65	424.298	2.174.033	0,1951	1	19,6	100
Guztira	2.174.033		1		100	.
Iturria: lanketa propioa EUSTAT. Udaletako biztanle-estatistiketatik hartutako datuekin (www.eustat.es)						

Maiztasun metatuek ez dute zentzurik aldagai nominalak ditugunean. Dena dela, aldagai kualitatibo ordinaletan kategoriak ordenatu daitezkeenez, maiztasun metatuak atera daitezke. Baldin eta datuen arabera zentzua badute.

Adibidea. Aldagai ordinalen maiztasun metatuak: ikasketa-maila.

12. taula: ikasketa-maila			
Ikasketak (x_i)	f_i (maiztasun absolutuak)	% (maiztasun erlatiboak, ehunekoak)	% (maiztasun erlatiboak metatuak, ehunekoak
Batere ez, lehen mailakoak baino baxuagoak	378	22,5	22,5
Lehen mailakoak, Oinarrizko batxilergoa, OHO, DBH	371	22	44,5
Lanbide Heziketa (LH)	424	25,2	69,7
Bigarren mailakoak: Goi-mailako Batxilergoa, IEE, BBB, UBI	298	17,7	87,4
Erdi goi-mailakoak (Diplomaturak)	123	7,3	94,7
Goi-mailakoak (Lizentziaturak, Ingeniaritzak, Doktoretzak,...)	89	5,3	100
Guztira	1683	100	
Iturria: lanketa propioa			

2.1.4. Maiztasun-taulak eraikitzeko gomendio orokorrak

- Maiztasun-taulek argi eta garbi adierazi behar dute izenburua eta iturria.
- Taula guztiek identifikatzen dituen zenbakia eraman behar dute. Ikerketa aurkezteko txostenaren hasieran, aurkibide orokorrarekin batera, taulen aurkibidea sartuko dugu, taula bakoitzaren zenbakia, izenburua eta dagoen orrialdea adieraziz.
- Maiztasun-taulek maiztasun absolutuak, erlatiboak eta, aldagai kuantitatiboen kasuan, baita metatuak ere izan ditzakete. Dena dela, interpretazioak egiterakoan, maiztasun erlatiboak hobesten dira, azalpen ulergarriagoak ematen dituztelako. Edonola ere, bakarrik maiztasun erlatiboak dituzten taulak aurkezten ditugunean, ehunekoak kalkulatzeko erabili ditugun guztizko maiztasunak adierazi behar ditugu. Hau da, ikerketarako erabili den laginaren tamaina adierazi behar dugu. Hala, estatistikekin gezurrak esatea edo informazioa manipulatzea ekiditen dugu. Adibidez, iragarki batek esan dezake dentisten %75ek hortz-ore bat gomendatzen duela, aipatu gabe galdetutako dentistak 4 izan direla. Kontuz ibili laginaren tamainari buruzko informaziorik ematen ez duten ikerketekin.
- Taulek ulergarriak, argiak, ahal den sinpleenak eta zehatzenak izan behar dute. Kontuz, bada, kolore gehiegi, diseinu ulergaitz edo nahastea sortzen duten efektuekin. Baliabide horiek guztiak erabiliko ditugu, ulerterraztasuna ziurtatzen badute.
- Aldagai kuantitatiboak ditugunean, komeni da balioak txikitik handira ordenatzea, maiztasun-taula ulergarriagoa delako.
- Datu multzoaren berezitasunek eta ikerlariaren nahiek zehazten dituzte maiztasun-taulan sartuko diren maiztasunak. Hala ere, kontuan izan aldagai mota bakoitzak maiztasun mota batzuk onartzen dituela:

13. taula: aldagai motak eta maiztasunak		
Aldagai mota		Maiztasunak
Aldagai kualitatiboa	Nominala	Absolutuak eta erlatiboak
	Ordinala	Absolutuak eta erlatiboak (aldagaiaren berezitasunen arabera, maiztasun metatuak ere kalkulatu daitezke)
Aldagai kuantitatiboa	Diskretua	Absolutuak, erlatiboak eta metatuak
	Jarraitua	Absolutuak, erlatiboak eta metatuak
Iturria: lanketa propioa		

- Azkenik, kontuan izan aldagai mota desberdinak nahasten dituzten taulak ere badaudela. Alegia, aldagai kuantitatibo eta kualitatiboak nahasten dituztenak.

Adibidea. Aldagai kuantitatiboak eta kualitatiboak nahasten dituen maiztasun-taula: Euskal AEn bizi den ama batengandik bizirik jaiotakoak.

1. taula: Euskal AEn bizi den ama batengandik bizirik jaioak. Alderaketa taula

	2011ko 4. hiruhil.		Aurreko hiruhilekoa (2011ko 3. hiruhil.)		Aurreko urteko hiruh. bera (2010eko 4. hiruhil.)	
	Kop.	%	Kop.	%	Kop.	%
Euskal AE	**5.256**	**100,0**	**5.345**	**100,0**	**5.434**	**100,0**
Araba / Álava	864	16,4	886	16,6	828	15,2
Bizkaia	2.681	51,0	2.663	49,8	2.784	51,2
Gipuzkoa	1.711	32,6	1.796	33,6	1.822	33,5
Amaren adinaren arabera	**5.256**	**100,0**	**5.345**	**100,0**	**5.434**	**100,0**
≤ 24 urte	298	5,7	300	5,6	332	6,1
25 - 29 urte	748	14,2	815	15,2	717	13,2
30 - 34 urte	2.188	41,6	2.187	40,9	2.264	41,7
35 - 39 urte	1.699	32,3	1.722	32,2	1.776	32,7
≥ 40 urte	323	6,1	321	6,0	345	6,3
Amaren egoera zibilaren arabera	**5.256**	**100,0**	**5.345**	**100,0**	**5.434**	**100,0**
Ezkondua	3.352	63,8	3.497	65,4	3.559	65,5
Ezkongabea	1.904	36,2	1.848	34,6	1.875	34,5
Amaren naziotasuna	5.256	100,0	5.345	100,0	5.434	100,0
Espainiarra	4.356	82,9	4.482	83,9	4.564	84,0
Atzerritarra	900	17,1	863	16,1	870	16,0
Jaiotza-hurrenkera	**5.256**	**100,0**	**5.345**	**100,0**	**5.434**	**100,0**
Lehenengoa	2.954	56,2	2.967	55,5	2.963	54,5
Bigarrena	1.858	35,4	1.943	36,4	1.994	36,7
Hirugarrena edo gehiago	444	8,4	435	8,1	477	8,8

Iturria: EUSTAT. Euskal AEko jaiotzen estatistika

2.1.5. Maiztasun-taulak irakurtzeko gomendio orokorrak

Guk sortutako taulak interpretatzeko, modu asko daude. Txostenaren helburuen eta irakurleen berezitasunen arabera, taulak era askotara irakur daitezke. Edonola ere, honako pauso hauek lagungarri izan daitezke:

- Lehenengo eta behin, balio edo kategoria bakoitzaren maiztasunak azaldu behar ditugu. Hau da, kategoria edo balio guztien maiztasunen deskripzioa egin beharko dugu. Horretarako, egokiena da maiztasun erlatiboak erabiltzea.
- Bigarrenik, adierazi behar dugu zein diren maiztasun erlatibo handienak dituzten kategoria edo balioak. Horrekin, datuak zein kategoria edo balioren inguruan biltzen diren azalduko dugu.
- Azkenik, gutxien agertzen diren balioak edo kategoriak zein diren zehaztu behar dugu.

Beste ikerlari batzuek egin dituzten taulak irakurtzerakoan, baliagarriak izan daitezke honako aholku hauek:

- Kontuan izan maiztasun-taulak aurkezteko modu asko daudela. Alegia, maiztasun-taulak ez dira beti agertzen guk hemen aurkeztu ditugun moduan. Toki arazoak edo, besterik gabe, txosten irakurterrazagoak egiteko helburuarekin, sarritan, bakarrik maiztasun erlatiboak dituzten taulak aurkezten dira txostenetan.
- Maiztasun-taula ulertzeko, ikerlarien ikuspuntuaz fidatu behar dugu. Hasteko, beraz, ikerlariek egin dituzten interpretazioak onartu behar ditugu.
- Ezinbestekoa da taularen izenburua, kategoria, balio edo tarteak eta taulari buruzko oharrak (iturria, esaterako) irakurtzea. Horrekin, zer motatako informazioa adierazten zaigun jakingo dugu.

- Ikerlariek egin duten interpretazioa buruan dugula, maiztasunak aztertuko ditugu (maiztasun erlatibo handienak eta txikienak). Normalean, taulan dagoen informazioa eta ikerlariak egin duen interpretazioa bat etorriko dira, baina, hala ere, ezin dugu ikuspuntu kritikoa galdu.

2.2. DATU MULTZOEN ADIERAZPEN GRAFIKOAK EDO DATU DIAGRAMAK

Maiztasun-taulekin batera, datu-diagramak, irudiak edo grafikoak ere oso erabilgarriak dira datu multzoak eta horien azterketatik lortutako emaitzak modu ulerterrazean plazaratzeko; bereziki, esku artean dugun datu multzoa handia bada edota adierazi nahi duguna maiztasun-taulekin baino modu ulerterrazagoan edota argiagoan aurkeztuko dugula pentsatzen badugu.
Badaude datu-diagrama desberdinak. Datu multzoari, ikergai diren aldagaien berezitasunei eta ikerketaren helburuei begiratzea, eta diagrama egokiena aukeratzea da ikerlariaren lana.
Gaur egun, edozein kalkulu-orrik edo estatistika-programak modu errazean egiten ditu datu-diagramak, eta, beraz, ez du zentzurik lan horiek eskuz egitea. Hala ere, garrantzitsua da datu-diagramak nola egiten diren jakitea, haien logika barneratzeko.

2.2.1. Sektore-diagrama

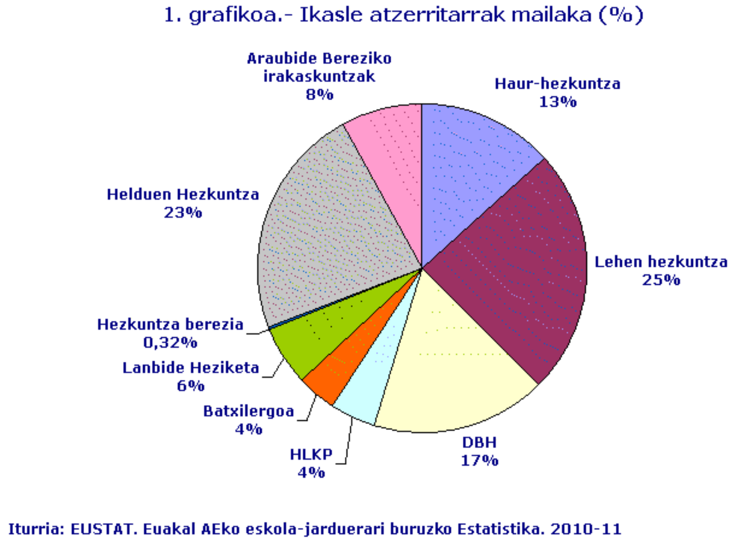
Sektore-diagrama aldagai kuantitatibo diskretuen eta, batez ere, aldagai kualitatiboen (nominal zein ordinalen) azterketatik lortzen ditugun datu multzoen maiztasun-banaketak irudikatzeko erabiltzen da; hau da, aldagaiaren balio edo kategorien maiztasunak grafikoki aurkezteko.
Diagrama hori sektoretan banatuta dagoen zirkulu bat da, eta sektore bakoitzaren azalerak aldagaiaren balioen edo kategorien maiztasuna absolutuak zein erlatiboak adierazten ditu. Alegia, sektoreen azalerak eta maiztasunak zuzenki proportzionalak dira. Kontuan izan maiztasun absolutuen zein erlatiboen sektore-diagramak irudikatu ditzakegula.

Adibidea. Aldagai nominal bati dagokion sektore-diagrama, maiztasun erlatiboak erabiliz egindakoa. Euskal AEko ikasle atzerritarren eskola-jarduerari buruzko maiztasun erlatiboen maiztasun-banaketa hau dugu:

14. taula: Euskal AEko atzerritarrak mailaka (2010-2011) %	
Haur-hezkuntza	13
Lehen hezkuntza	25
DBH	17
HLKP	4
Batxilergoa	4
Lanbide Heziketa	6
Hezkuntza berezia	0,32
Helduen Hezkuntza	23
Araubide bereziko irakaskuntzak	8
	100
Iturria: EUSTAT. Euskal AEko eskola-jarduerari buruzko estatistika, 2010-2011 (www.eustat.es)	

Maiztasun-banaketa horretatik abiatuz, sektore-diagrama hau lortzen dugu:

2. diagrama



Sektore-diagramak eratzeko prozedura

Arestian esan bezala, sektore-diagrama osatzen duten sektoreen azalerek aldagaiaren balioen edo kategorien maiztasunak adierazten dituzte. Kontuan izanik zirkulu osoak 360º dituela, hiruko erregela erabiliz, maiztasun bakoitzari dagokion angeluaren araberako azalera emango diogu. Angelua (α) kalkulatzeko, honako formula hau ere erabil dezakegu:

$$\alpha = \frac{360^{o}}{N} . f_i$$

Adibidez, aldagai kualitatibo baten kategoria bati dagokion maiztasun erlatiboa %36 bada, berari dagokion α angelua, gradutan, honela kalkulatuko dugu:

$$\alpha = \frac{36x360}{100} = 129{,}6^{o}$$

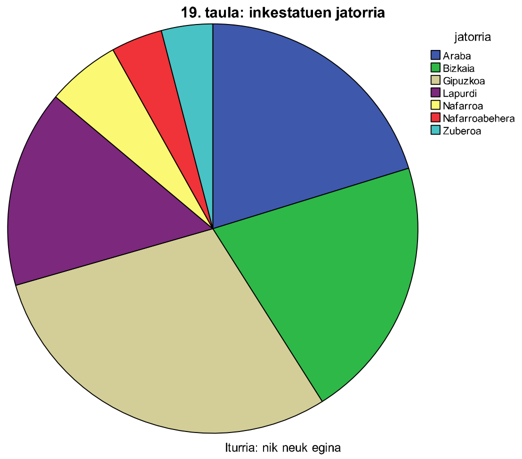
Sektore bakoitzari dagokion gradu kopurua zehaztu ostean, zirkulua zatikatzen dugu, garraiagailuaren laguntzarekin.

Adibidea. Aldagai nominal baten sektore-diagrama, sektoreek maiztasun absolutuak adierazten dituztelarik.

15. taula: inkestatuen jatorria	
Araba	35
Bizkaia	36
Gipuzkoa	51
Lapurdi	27
Nafarroa	10
Nafarroa Behera	7
Zuberoa	7
Total	173
Iturria: lanketa propioa	

Maiztasun absolutuak agertzen dituen taula horretatik abiatuta, honako sektore-diagrama hau lortzen dugu.

3. diagrama



Sektore-diagramari begiratuta, argi dago datu multzoko gehiengoa gipuzkoarra dela.

Sektore-diagrama motak

Ohikoenak aurkeztu ditugun erako sektore-diagramak dira; hau da, sektoretan banatutako zirkuluak. Dena dela, beste era bateko sektore-diagramak ere badaude. Egia esan, ikerlariaren sormenaren, helburuen eta erabiltzen duen programaren arabera, itxura eta forma askotako sektore-diagramak egin daitezke. Besteak beste, banatutako sektore-diagramak ditugu, non sektoreak elkartuta agertu beharrean, banatuta aurkezten diren. Bestetik, hiru dimentsiotako sektore-diagramak daude.

4. diagrama: Europako Legebiltzarreko diputatuen banaketa, alderdi politikoen arabera

Iturria: Europako Legebiltzarreko webgunea

Sektore-diagramak erabiltzerakoan kontuan izan beharrekoak

- Edozein aldagai motarekin erabil daiteke, baina, nagusiki, aldagai kualitatiboekin (nominal eta ordinalekin) erabiltzen da.
- Maiztasun erlatiboen zein absolutuen sektore-diagramak egin ditzakegu, baina maiztasun erlatiboen sektore-diagramak maiztasun absolutuenak baino maizago erabiltzen dira, sektoreen arteko alderaketak errazago egin daitezkeelako, eta, beraz,

ulergarriagoak direlako. Edonola ere, datu multzoaren berezitasunek eta ikerlariaren helburuek zehaztuko dute sektore-diagramaren izaera.

- Sektore-diagrama ez da egokia kategoria edo balio desberdin asko dituzten aldagaiak irudikatzeko, sektore asko dituzten diagramak zailak direlako interpretatzen. Autore batzuen arabera, bost kategoria edo balio baino gehiago ditugunean, hobe da barra-diagrama erabiltzea.
- Sektore-diagrama ez da egokia sektoreek antzerako azalerak dituztenean, sektoreen artean alderaketak egitea zaila delako.
- Sektore-diagrama oso egokia da aztergai den aldagaiak balio edo kategoria gutxi dituenean, kategoria edo balio horien artean desberdintasunak daudenean, eta informazio sinpleak modu errazean agertu nahi ditugunean.
- Autore batzuen arabera, komeni da sektoreak handitik txikira ordenatzea, ordulari baten ibilbideari jarraituz. Modu horretan, sektore-diagrama ulerterrazagoak lortzen ditugu.
- Sektoreak kolore edo tonu desberdinekin aurkeztu behar dira, sektoreen artean alderaketak errazteko. Dena dela, kontuz kolore gehiegi, diseinu ulergaitzak edo nahastea sortuko duten efektuak erabiltzearekin. Baliabide horiek guztiak erabiliko ditugu ulerterraztasuna ahalbidetzen badute.
- Ulergarritasunaren izenean, komeni da sektore bakoitzaren alboan dagokion maiztasun erlatiboa (nahi bada baita maiztasun absolutua ere) agertzea, baita sektore edo kategoria bakoitzaren izena ere.
- Beste diagrametan ez bezala, sektore-diagramarekin ezin dira maiztasun metatuak irudikatu.
- Datu-diagrama guztiek bezala, sektore-diagramek ere argi eta garbi adierazi behar dute izenburua eta iturria.
- Sektore-diagramek identifikaziorako zenbakia eraman behar dute. Datu-diagramen aurkibidean, diagramak duen zenbakia, izenburua eta dagoen orrialdera agertuko ditugu.

2.2.2. Piktograma

Piktogramak aldagai kualitatiboak irudikatzeko beste modu bat dira. Kasu honetan, agertu nahi ditugun maiztasunak irudien edo ikurren bidez adierazten dira.

Adibidea. Sexua aldagaiaren piktograma. Demagun aztertu dugun elkartean hamarretik lau gizonezkoak direla. Piktograma erraz baten bidez, era honetan adieraziko genuke gizon eta emakumeen arteko proportzio hori.

5. diagrama: elkarteko bazkideen sexua
Emakumeak: ♀♀♀♀♀♀ (%60)
Gizonak: ♂♂♂♂ (%40)
Iturria: lanketa propioa

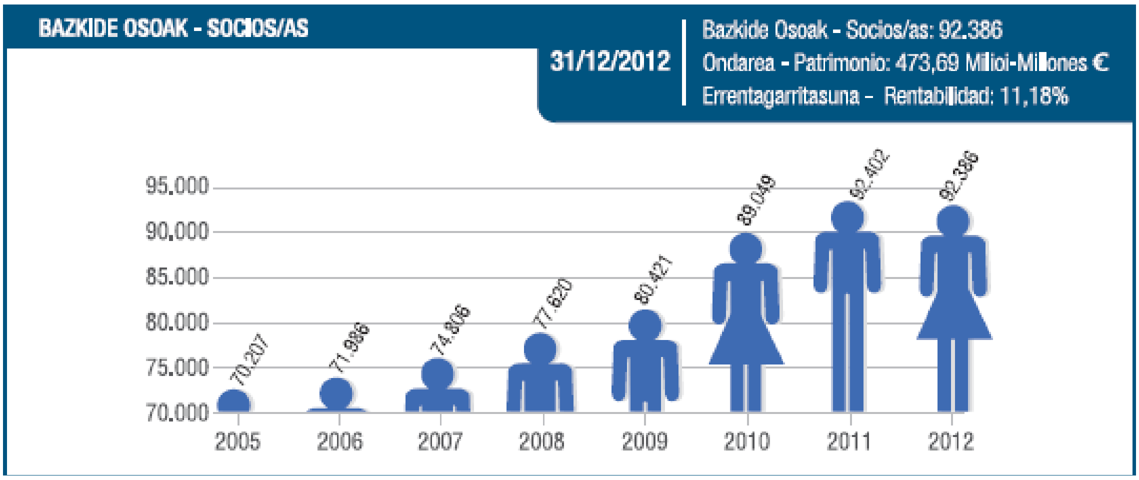
Adibidea. Piktograma: hiru herrialdek osasungintzan egiten duten gastua dolarretan.

6. diagrama: A, B eta C herrialdeek osasungintzan egiten duten gastua (dolarretan)
A herrialdea: $$$$$ (%50)
B herrialdea: $$$$ (%40)
C herrialdea: $$$ (%30)
Iturria: lanketa propioa

Piktograma eratzeko prozedura

Piktograma egiteko, gaiarekin lotuta dagoen irudi edo ikur bat aukeratzen da, eta baliabide hori errepikatzen, edo baliabide horren tamainarekin jolasten, maiztasun absolutu zein erlatiboen arabera. Ulerterraztasunaren izenean, komeni da irudi bakoitzaren ostean irudi horri dagokion maiztasun erlatiboa edo absolutua adieraztea.

Adibidea. 7. diagrama: Itzarri EPSVaren bazkide kopuruari buruzko piktograma.



Iturria: Itzarri-EPSV http://www.itzarri.com

Piktograma motak

Piktogrametan, irudiak edo ikurrak errepikatuz edota haien tamainarekin jolastuz aurkezten dira maiztasunak. Beraz, ikerlariaren sormenaren, ikerketaren helburuen eta eskuartean ditugun baliabide grafikoen arabera, nahi diren bezalako piktogramak egin ahal izango ditugu. Hori bai, ulerterrazak direla ziurtatu behar dugu.

Piktograma erabiltzerakoan kontuan izan beharrekoak

- Aldagai kuantitatibo zein kualitatiboekin erabil daiteke.
- Ondo egiten bada, informazioa emateko modu ulerterraza eta erakargarria izan daiteke. Oso erabilgarriak dira ikerketak ilustratzeko edota arintzeko.
- Lagungarri dira umeei edo gaia ulertzeko zailtasunak izan ditzaketenei datu multzoen berezitasunak azaltzeko.
- Oso egokiak dira maiztasun gutxi irudikatu behar ditugunean, eta informazio sinpleak modu errazean agertu nahi ditugunean.
- Maiztasun asko batera irudikatu behar ditugunean, edo maiztasunen artean desberdintasun argiak ez daudenean, piktograma ez da ulerterraza, eta, beraz, ez da erabilgarria.
- Oso egokia da datu multzo desberdin arteko desberdintasunak begirada batean agertzeko.
- Datu-diagrama guztiek bezala, piktogramek ere izenburua eta iturria izan behar dute.
- Datu-diagrama guztiek bezala, identifikazio-zenbakia eraman behar dute. Datu-diagramen aurkibidean, piktogramen zenbakiak, izenburuak eta dauden orrialdeak agertuko ditugu.

2.2.3. Barra-diagrama edo zutabe-diagrama

Barra-diagrama edo zutabe-diagrama laukizuzen formako barrak edo zutabeak dituen datu-diagrama da. Laukizuzenen luzeren bidez, balioen edo kategorien maiztasunak irudikatzen ditugu. Kategorien edo balioen arteko desberdintasunak modu ulerterrazean agertzea da datu-diagramen helburua. Laukizuzenek maiztasun absolutuak zein erlatiboak adieraz ditzakete.
Aldagai kualitatiboak edo kuantitatibo diskretuak irudikatzeko erabiltzen da. Aldagai jarraituekin, aurrerago aurkeztuko dugun barra-diagrama berezia erabiltzen da: histograma.

Adibidea. Aldagai nominal batekin egindako barra-diagrama; barrek maiztasun absolutuak adierazten dituzte: 173 pertsonari beren jaioterriari buruz galdetu diegu, eta honako maiztasun-taula hau lortu dugu, IBM SPSSaren laguntzarekin.

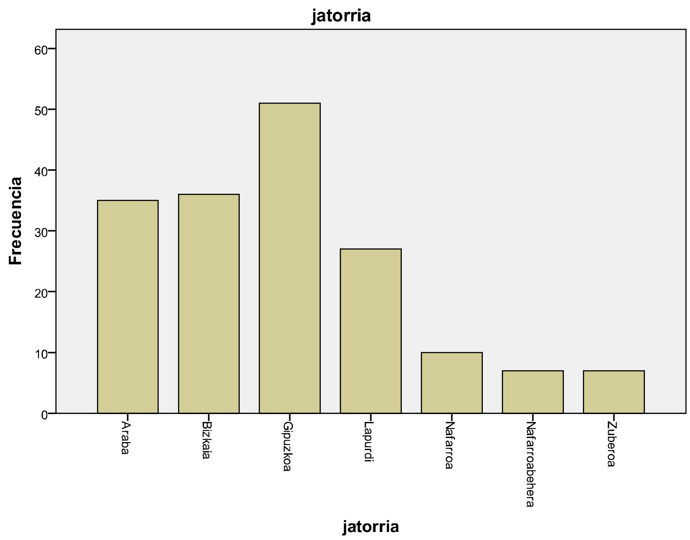
16. taula: inkestatuen jatorria

x_i	f_i	%
Araba	35	20,2
Bizkaia	36	20,8
Gipuzkoa	51	29,5
Lapurdi	27	15,6
Nafarroa	10	5,8
Nafarroa Behera	7	4,0
Zuberoa	7	4,0
Guztira (n)	173	100,0

Iturria: lanketa propioa

Maiztasun-taula horretatik abiatuta, maiztasun absolutuetan oinarritutako honako barra-diagrama hau lortzen dugu:

8. diagrama: inkestatuen jatorria



Iturria: lanketa propioa

Barra-diagrama ikusita, argi dago gehiengoa gipuzkoarra dela.

Barra-diagramak eratzeko prozedura

Barra-diagrama irudikatzeko, Ardatz Kartesiarretako ardatz horizontalean, aldagai estatistikoaren balioak edo kategoriak kokatzen ditugu, balio edo kategoria batetik bestera distantzia berdinak utziz. Ostean, balio edo kategoria bakoitzaren gainean, barra bertikal bat irudikatzen dugu. Barra horien luzerek balioaren edo kategoriaren maiztasunarekiko (absolutu zein erlatibo) zuzenki proportzionalak izan behar dute.

Adibidea. Aldagai diskretu batekin egindako barra-diagrama; barrek maiztasun erlatiboak adierazten dituzte: gailu elektronikoen kopurua etxean.

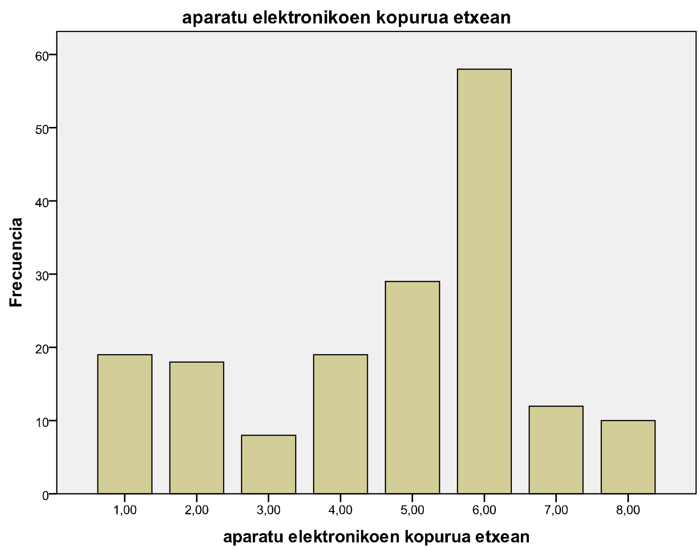
17. taula: gailu elektronikoen kopurua etxean

x_i	f_i	%	Metatutako ehunekoak
1,00	19	11,0	11,0
2,00	18	10,4	21,4
3,00	8	4,6	26,0
4,00	19	11,0	37,0
5,00	29	16,8	53,8
6,00	58	33,5	87,3
7,00	12	6,9	94,2
8,00	10	5,8	100,0
Total	173	100,0	

Iturria: lanketa propioa

IBM SPSS programaren laguntzarekin, honako barra-diagrama hau lortzen dugu:

9. diagrama



Iturria: lanketa propioa

Barra-diagramak erakusten duenez, gehienek 6 gailu elektroniko dituzte etxean.

Barra-diagrama motak

Barra-diagrama sinpleenek datu multzo bateko kategorien edo balioen maiztasunak agertzen dituzte. Dena dela, beste era bateko barra-diagramak ere egin ditzakegu.

Datu multzoak alderatzeko barra-diagramak

Datu multzoak aldagai baten arabera alderatu nahi ditugunean, bi barra-diagrama egitea da lehen aukera. Bigarren aukera, berriz, bi barra-diagrametako zutabeak diagrama berean agertzea da.

Adibidea. Datu multzoak alderatzeko barra-diagrama: araubide orokorreko irakaskuntzan ikasle kopurua irakasleko 1982/1983 eta 2010/2011. ikasturteetan.

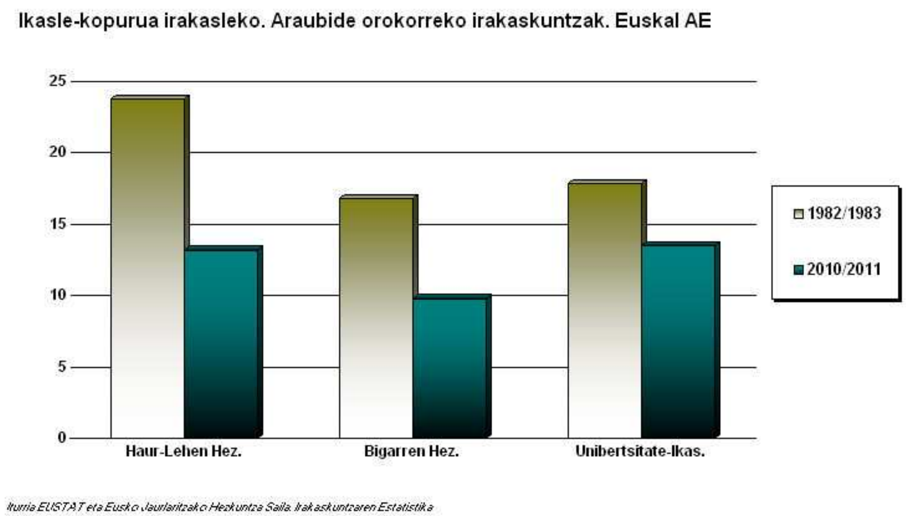
18. taula: araubide orokorreko irakaskuntzan irakasleko dagoen ikasle kopurua 1982/1983 eta 2010/2011. ikasturteetan

	1982/1983	2010/2011
Haur-lehen hezkuntza	24	12,2
Bigarren hezkuntza	17	9,8
Unibertsitate-ikasketak	23	13,5

Iturria: EUSTAT. Euskal AEko eskola-jarduerari buruzko estatistika, 2010-2011 (www.eustat.es)

Ikasturte bakoitzari dagozkion bi datu multzo horietatik abiatuta, diagrama berean, bi barra-diagrama dituen grafikoa lortzen dugu.

10. diagrama



Barra-diagrama horrek argi erakusten du araubide orokorreko irakaskuntzan irakasle bakoitzeko ikasle kopurua murriztu egin dela 1982/1983. ikasturtetik 2010/2011. ikasturtera. Halaber, erakusten du ikasle kopuru handiena irakasleko unibertsitatean dagoela.

Barra-diagrama horizontalak

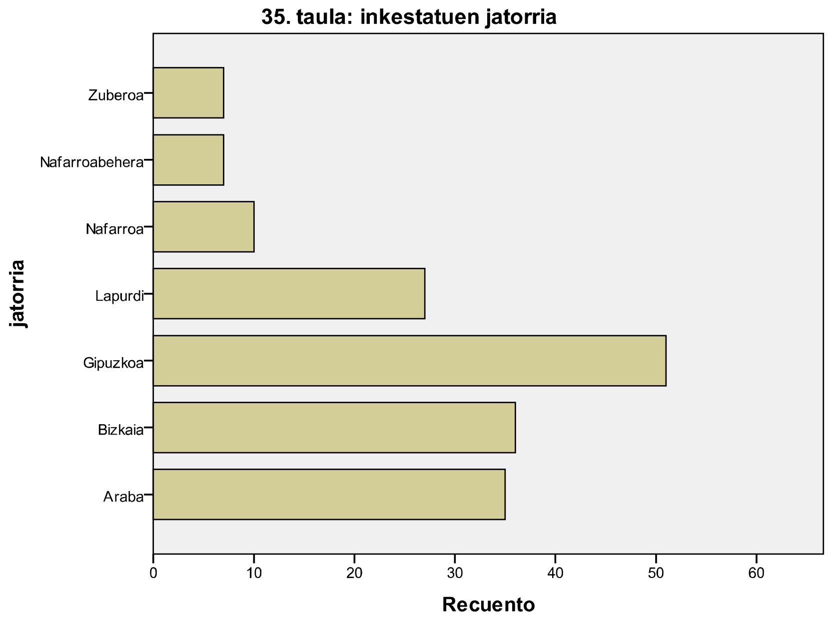
Kasu batzuetan, zutabeak modu bertikalean agertu ordez, modu horizontalean agertu nahi izaten dira. Horrelakoetan, aldagaiaren balioak edo kategoriak ardatz bertikalean jartzen dira eta, maiztasunak, ardatz horizontalean. Autore batzuen arabera, 10 zutabe baino gehiago ditugunean, hobe da barra-diagrama horizontalak erabiltzea, ulerterrazagoak direlako.

Adibidea. Barra-diagrama horizontala: inkestatuen jatorria.

19. taula: inkestatuen jatorria					
		Maiztasuna	Ehunekoa	Ehuneko balioduna	Ehuneko metatua
Baliodunak	Araba	35	6,1	20,2	20,2
	Bizkaia	36	6,2	20,8	41,0
	Gipuzkoa	51	8,8	29,5	70,5
	Lapurdi	27	4,7	15,6	86,1
	Nafarroa	10	1,7	5,8	91,9
	Nafarroa Behera	7	1,2	4,0	96,0
	Zuberoa	7	1,2	4,0	100,0
	Guztira	173	30,0	100,0	
Galduak	Sistema	404	70,0		
Guztira		577	100,0		
Iturria: lanketa propioa					

Datu multzo horretatik abiatuta, barra-diagrama horizontal hau lortzen dugu, IBM SPSS programaren laguntzarekin:

11. diagrama



Iturria: nik neuk egina

Barra-diagramak erabiltzerakoan kontuan izan beharrekoak

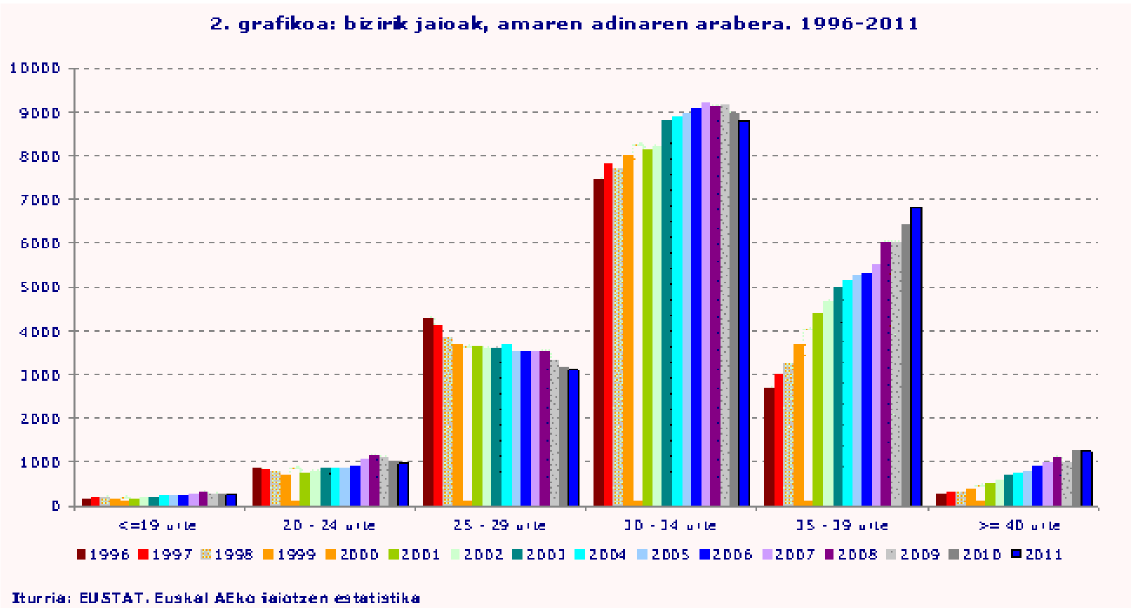
- Barra-diagramak oso egokiak dira aldagai kualitatiboen eta kuantitatibo diskretuen balio edo kategorien maiztasunak modu ulerterrazean agertzeko, baita kategorien edo balioen maiztasunen artean dauden desberdintasunak erakusteko ere.
- Arestian esan bezala, kategoria edo balio desberdin asko dituzten aldagaiak ditugunean, sektore-diagramak zailak dira interpretatzen. Autore batzuen ustez, bost kategoria edo balio baino gehiago ditugunean, hobe da barra-diagrama erabiltzea.
- Barrak handienetik txikienera edo txikienetik handienera ordenatu ditzakegu, gure nahien, hipotesien edo helburuen arabera. Hala ere, kategoria edo balioek aldez aurretik zehaztutako ordena badute, ordena hori errespetatu egin behar dugu. Adibidez, kategoriak alfabetikoki ordenatuta badaude, urtez urteko neurketak badira edo ikergaiak

zehazten duen ordena logikoren bat badago, ordena horiek kontuan hartu beharko ditugu.

- Autore batzuen arabera, aldagai kualitatiboen kasuan, kategoriak eta, horren ondorioz, baita barrak ere, asko badira, ulerterraztasuna ziurtatzeko, komenigarria da zutabeak altuera edo maiztasun handienetik txikienera ordenatzea[7]. Aldagai kuantitatiboen kasuan, ordea, komeni da aldagaiaren balioak txikienetik handienera jartzea, interpretazioa errazteko.
- Barra-diagrama oso egokia da urte desberdinetako datu multzoak eta, oro har, datu multzo desberdinak alderatzeko. Kasu horretan, datu multzo bakoitzaren zutabeak Ardatz Kartesiar berean jartzen dira; hori bai, datu multzo desberdin bakoitzaren barrei kolore desberdinak ematen zaizkie.
- Barrak modu bertikalean nahiz horizontalean jar daitezke. Ikerlari batzuen arabera, barra asko ditugunean (10 baino gehiago), hobe da barra horizontalak erabiltzea.
- Zutabe bakoitzak adierazten duena argi idatzi behar da ardatz horizontalean.
- Datu-diagrama guztiek bezala, barra-diagramek ere izenburua eta iturria adierazi behar dute.
- Datu-diagrama guztiek bezala, identifikazio-zenbakia eraman behar dute. Datu-diagramen aurkibidean, barra-diagramak duen zenbakia, izenburua eta dagoen orrialdera agertuko ditugu.

Adibidea. Barra gehiegi dituen barra-diagrama.

12. diagrama



Geroago ikusiko dugun bezala, denbora askoko datuak ditugunean, hau da, serie kronologiko luzeak ditugunean, barra-diagramak nahasiak izan daitezke. Horrelakoetan, hobe da geroago aztertuko ditugun lerro-diagramak erabiltzea.

---

[7] Barrak modu horretan ordenatuta, Paretoren diagrama izango dugu. Paretoren diagraman (80-20 kurba edo C-A-B banaketa izenarekin ere ezagutzen da), barrak ezkerretik eskuinera, maiztasun handienetik txikienera, aurkezten dira. Diagrama horren helburua erakunde batean lehentasunen artean ordena bat sortzea da, ostean erabakiak hartzeko. Izan ere, garrantzi gutxi duten arazo asko daude, eta benetako garrantzia duten arazo gutxi. Barra-diagramaren bidez, garrantzi gutxiko arazoak eskuinean jartzen dira, eta garrantzi asko duten arazoak, ezkerrean.

2.2.4. Histograma

Barra-diagramak bezala, histogramak ere laukizuzen formako barrak edo zutabeak dituzten datu-diagramak dira. Aldai jarraituetarako eta balio asko dituzten eta, horren ondorioz, tartetan banatu diren aldagai kuantitatibo diskretuetarako erabiltzen dira. Abzisen ardatzean, tartearen zabalera duten zutabeak jartzen dira, eta, altuera zehazteko, tarte bakoitzaren maiztasuna hartzen da kontuan. Kasu horretan, aldagaien izaera numerikoa adierazteko (aldagai jarraituak eta tartetan banatu diren aldagai kuantitatibo diskretuak zenbakizko balioen bidez neurtzen dira), barrak bata bestearen alboan jartzen dira.

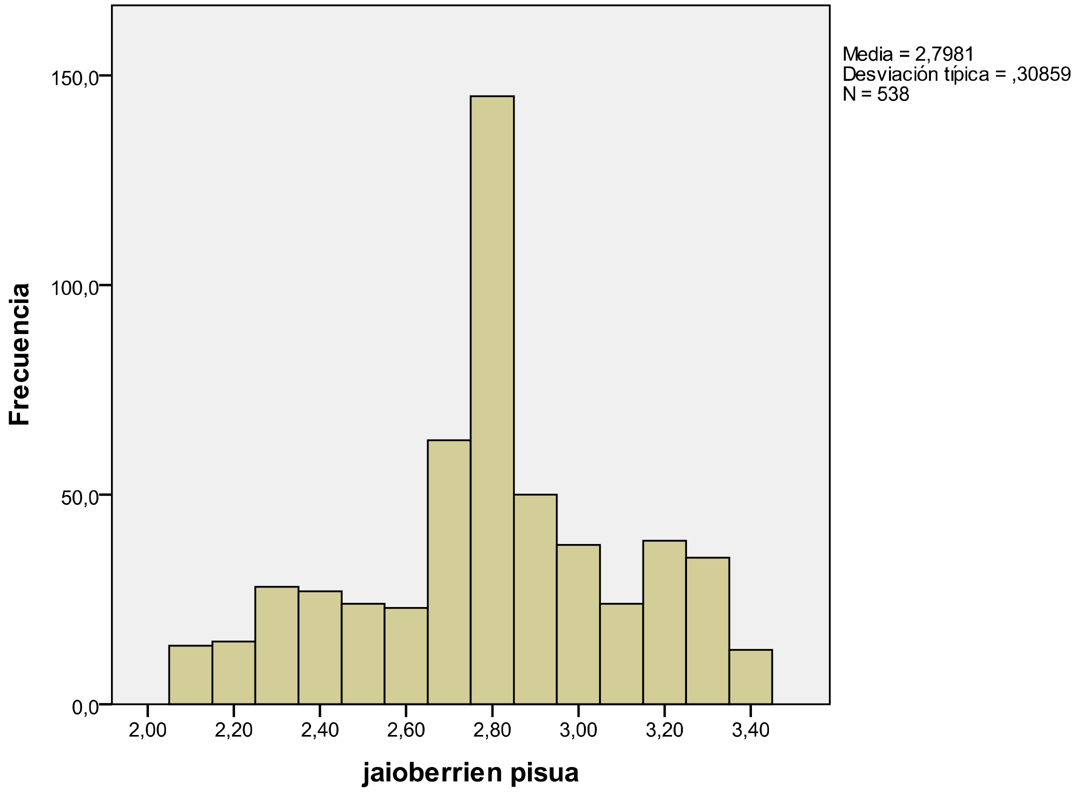
Adibidea. Aldagai jarraitu baten histograma: jaioberrien pisua. Hiru egunetan ospitale batean jaio diren 538 jaioberrien pisuak hartu ditugu, eta IBM SPSS programa informatikoaren laguntzarekin, honako maiztasun-banaketa hau lortu dugu:

20. taula: jaioberrien pisua		
2,10	14	2,6
2,20	15	2,8
2,30	28	5,2
2,40	27	5,0
2,50	24	4,5
2,60	23	4,3
2,70	63	11,7
2,80	145	27,0
2,90	50	9,3
3,00	38	7,1
3,10	24	4,5
3,20	39	7,2
3,30	35	6,5
3,40	13	2,4
Total	538	100,0
Iturria: lanketa propioa		

Datu multzo horretatik abiatuz, IBM SPSSak automatikoki sortzen ditu tarteak (2,05-2,15; 2,15-2,25; 2,25-2,35; 2,35-2,45; 2,45-2,55…), honako histograma hau sortuz:

13. diagrama

34. diagrama: jaio berrien pisua



Iturria: nik neuk egina

Histograma eratzeko prozedura

Histograma eratzeko, jatorrizko datuak tartetan biltzen dira. Horretarako, aldagai jarraituak edo diskretuak tartetan banatzeko arestian aurkeztu dugun prozedura erabiliko dugu.

Ondoren, tarte bakoitzeko maiztasun absolutu eta erlatiboak kalkulatuko ditugu. Alegia, tarte bakoitzean zenbat datu biltzen den zenbaki absolutuetan eta ehunekoetan.

Ardatz Kartesiarretako abzisetan (ardatz horizontalean), tarteak ezarriko ditugu eta ordenatuetan (ardatz bertikalean), maiztasunak (absolutuak edo erlatiboak). Hau da, tarte bakoitzaren gainean, tarte horri dagokion maiztasunarekiko zuzenki proportzionala den laukizuzena irudikatuko dugu.

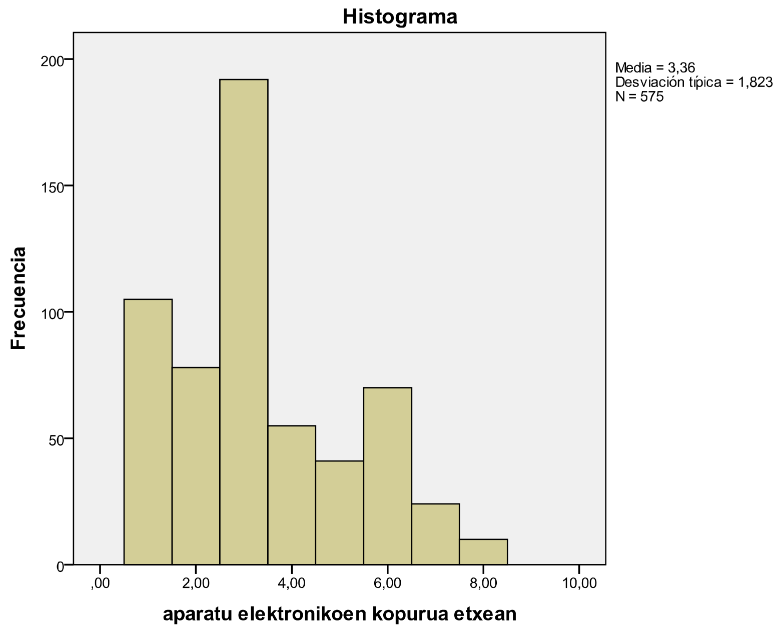
Tarte guztiek zabalera bera baldin badute, maiztasunarekiko proportzionala den altuera edo azalera hartzen da. IBM SPSS programa estatistikoak, adibidez, tarte-zabalera bereko histogramak egiten ditu.

Adibidea. Aldagai diskretu batetik egindako histograma: gailu elektronikoen kopurua etxean.

21. taula: gailu elektronikoen kopurua etxean		
Gailu elektronikoak etxean (x_i)	f_i	%
1,00	105	18,3
2,00	78	13,6
3,00	192	33,4
4,00	55	9,6
5,00	41	7,1
6,00	70	12,2
7,00	24	4,2
8,00	10	1,7
Total	575	100,0
Iturria: lanketa propioa		

Datu multzo horretatik abiatuta, honako histograma hau lortzen dugu, IBM SPSS programa informatikoaren laguntzarekin:

14. diagrama: gailu elektronikoen kopurua etxean



Iturria: lanketa propioa

Argitasunaren izenean, komeni da tarte guztiek zabalera bera izatea (aurreko kasuan IBM SPSSek zabalera bereko tarteak egin dizkigu: 0,5-1,5; 1,5-2,5; 2,5-3,5…), baina, batzuetan, histograman zehar maiztasunik gabeko hutsunerik sor ez dadin, tarte batzuk elkartu egiten dira. Beste batzuetan, hasierako eta bukaerako tarteak mugatu gabe uzten dira (>100, <25).

Dugun maiztasun-taulako tarteak zabalera desberdinekoak direnean, histograma irudikatzeko aldaketa batzuk egin behar dira, hain zuzen ere, zutabeek datuen trinkotasuna egoki irudika ditzaten. Zehatzago esanda, tarte-zabalera desberdineko maiztasun-taulak ditugunean, tarte bakoitzari dagokion zutabearen altuera (a) honela kalkulatzen da:

$$a = \frac{n}{h}$$

non n tarteko maiztasuna baita, eta h tarte-zabalera.

Kontuan izan era horretako histogramen zutabeek ez dutela maiztasuna adierazten, datuen trinkotasuna baizik. Zutabe bakoitzaren altuera horrela kalkulatuko ez bagenu, zabalera handiena duten tarteek behar baino azalera handiagoarekin agertuko lirateke.

Histograma motak

Autore batzuen arabera, hiru eratako histogramak daude:

- Histograma arruntak: abzisen ardatzean, tartearen zabalera duten zutabeak jartzen dira, eta, altuera zehazteko, tarte bakoitzaren maiztasuna hartzen da kontuan.

- ➢ Maiztasun bakunen poligonoak: barra-diagramak barra asko dituenean edo histogramak tarte asko, edota aldagaiaren balioen barruan gertatzen diren maiztasun-aldaketak erakutsi nahi ditugunean, maiztasun bakunen poligonoak erabiltzea gomendatzen da.
- ➢ Elkartutako histogramak: datu multzo desberdinak alderatu nahi ditugunean erabiltzen dira.

Histograma arruntak orain arte ikusitakoak dira, eta maiztasun bakunen poligonoak ostean ikusiko ditugu, ia gehienetan histogrametatik bereizita agertzen direlako. Dena dela, merezi du elkartutako histogramei buruz zerbait esatea.

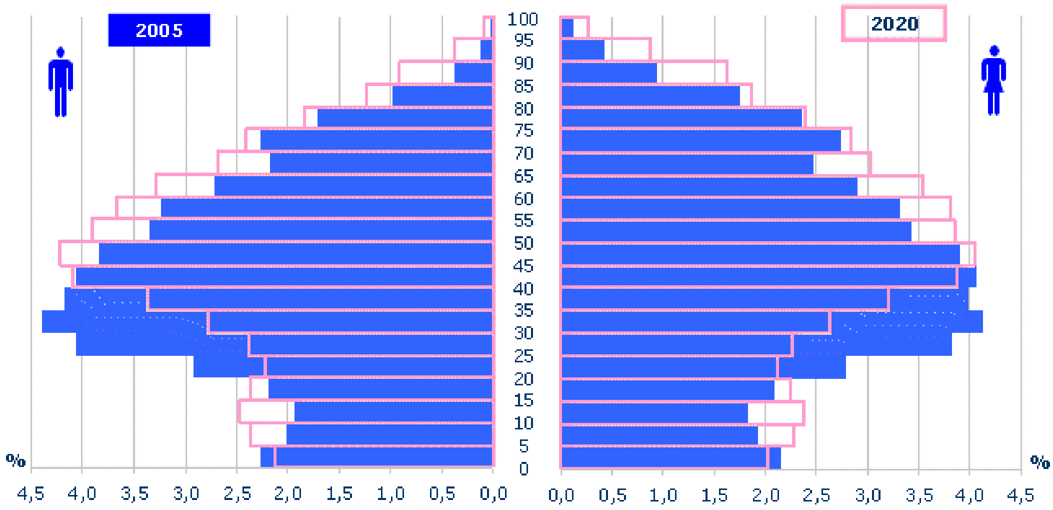
Elkartutako histogramak

Datu multzo desberdinak alderatu nahi ditugunean, histogramak elkartu egiten dira. Horrelakoetan, tarteak ardatz bertikalean jartzen dira, eta maiztasunak, horizontalean.

Elkartutako histogramak asko erabiltzen dira azterketa demografikoak egiten direnean. Hain zuzen ere, biztanleriaren piramideak elkartutako bi histograma dira, bata gizonezkoentzat eta bestea emakumezkoentzat. Esan bezala, horrelako kasuetan, tarteak ardatz horizontalean joan beharrean (abzisetan), ardatz bertikalean (ordenatuetan) jartzen dira, eta maiztasunak, ardatz horizontalean. Ostean, bi histogramak elkartzen dira.

Adibidea. Biztanleria-piramidea: Euskal AEko biztanleria piramideak 2005-2020.

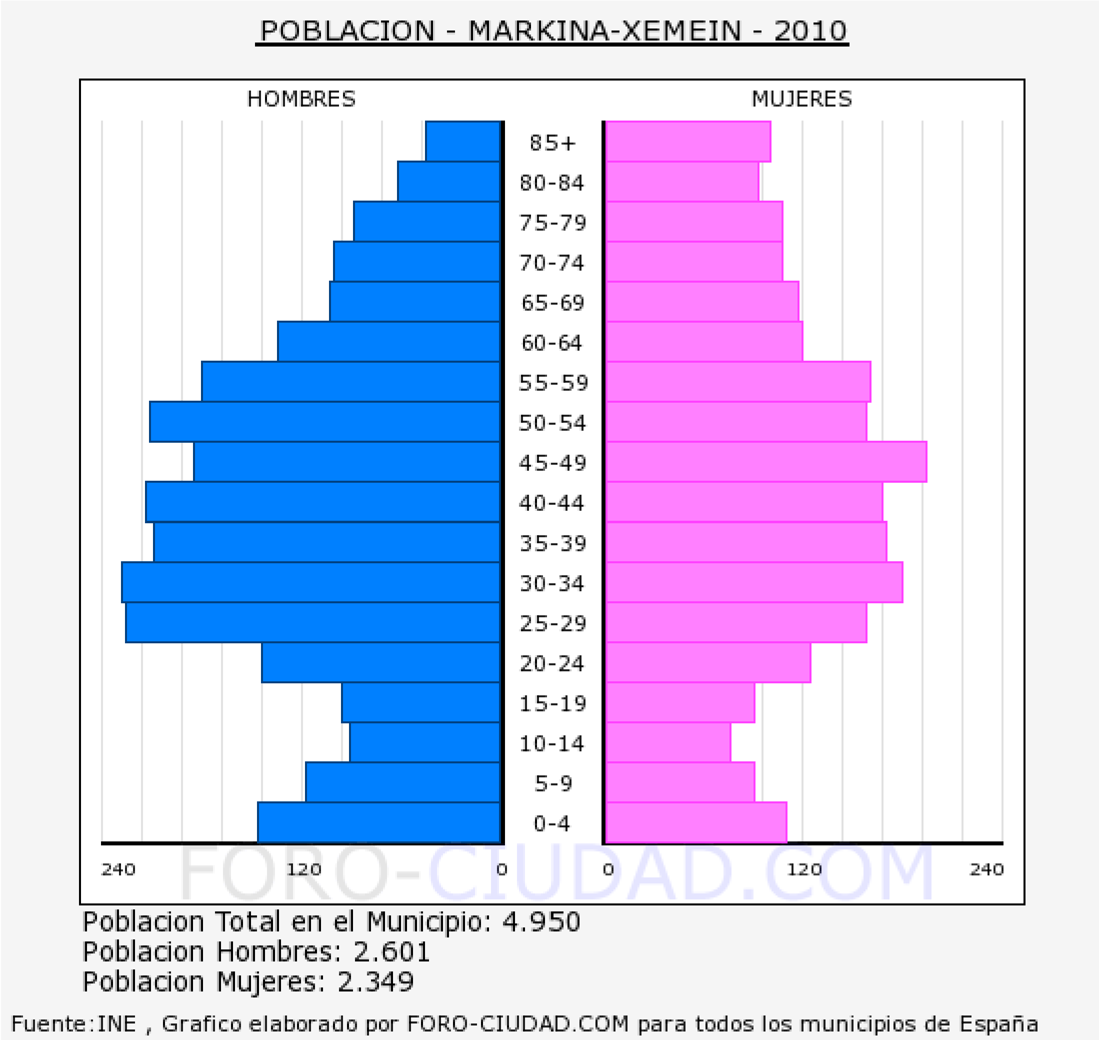
15. diagrama. Euskal AEko biztanleria piramideak 2005-2020



Iturria: EUSTAT(www.eustat.es)

Adibidea. Biztanleria-piramidea: Markina-Xemein, 2005-2020

16. diagrama: Markina-Xemeingo biztanleria-piramidea 2005-2020



Estatistikarako programa informatikoek oso erraz egiten dituzte biztanleria-piramideak. Edonola ere, eskuz egitekotan, honako pauso hauek eman behar ditugu:

- Gizonentzat eta emakumeentzat, adin tarte bakoitzean dagoen populazioaren ehunekoa kalkulatzen dugu.
- Gizonentzat eta emakumeentzat, Ardatz Kartesiarrak irudikatzen ditugu. Ardatz bertikalean, adin-tarteak zehazten ditugu, eta ardatz horizontalean, ehunekoak.
- Gizonen histograma eskuinean irudikatuz eta emakumeena ezkerrean, adin-tarte bakoitzari dagokion ehunekoaren araberako laukizuzen edo barra irudikatuko dugu.

Histogramak erabiltzerakoan kontuan izan beharrekoak

- Histograma oso egokia da aldagai jarraituak irudikatzeko, baita tartetan banatuta dauden eta aldagai jarraitu gisa aztertu nahi diren aldagai diskretuak irudikatzeko ere.
- Histograma egiteko, gutxieneko datu kopurua behar dugu. Autore batzuen arabera, histograma ez da egokia 20 datu baino gutxiago ditugunean.
- Argitasunagatik, komeni da tartearen zabalera konstantea izatea, baina histograman zehar maiztasunik gabeko tarterik sor ez dadin, tarteak elkartu egin daitezke. Esan

bezala, horrelako kasuetan, zutabeek ez dute maiztasuna adierazten, datuen trinkotasuna baizik.

- Estatistikan gehien erabiltzen den datu-diagrametako bat da, aldagai jarraituen zein tartetan banatuta dauden aldagai diskretuen ezaugarri estatistiko nagusienak (zentroa, sakabanatzea, ...) ezagutzeko aukera ematen duelako. Halaber, datu multzoak alderatzeko aukera ematen du, dagozkien histogramak batera aztertzen baititu.
- Ardatz bakoitzak adierazten duena argi idatzi behar da.
- Datu-diagrama guztiek bezala, histogramek ere izenburua eta iturria adierazi behar dute.
- Datu-diagrama guztiek bezala, identifikazio-zenbakia eraman behar dute. Datu-diagramen aurkibidean, histogramen zenbakiak, izenburuak eta dauden orrialdeak agertuko ditugu.

2.2.5. Maiztasun bakunen poligonoa

Arestian esan bezala, barra-diagramak barra asko baditu edo histogramak tarte asko, edota aldagaiaren balioen barruan gertatzen diren maiztasun-aldaketak modu grafikoak erakutsi nahi baditugu, histogramak eta barra-diagramak nahasiak izan daitezke. Horrelakoetan, hobe da maiztasun bakunen poligonoak erabiltzea.
Maiztasun bakunen poligonoak aldagai kuantitatiboen (diskretuen zein jarraituen) maiztasun-banaketak irudikatzeko erabiltzen dira. Goian adierazi bezala, autore batzuek maiztasun bakunen poligonoa histograma mota bat bezala aurkezten dute.

Adibidea. Aldagai jarraitu baten maiztasun bakunen poligonoa: bigarren ospitale bateko jaioberrien pisua.

22. taula: bigarren ospitale bateko jaioberrien pisua

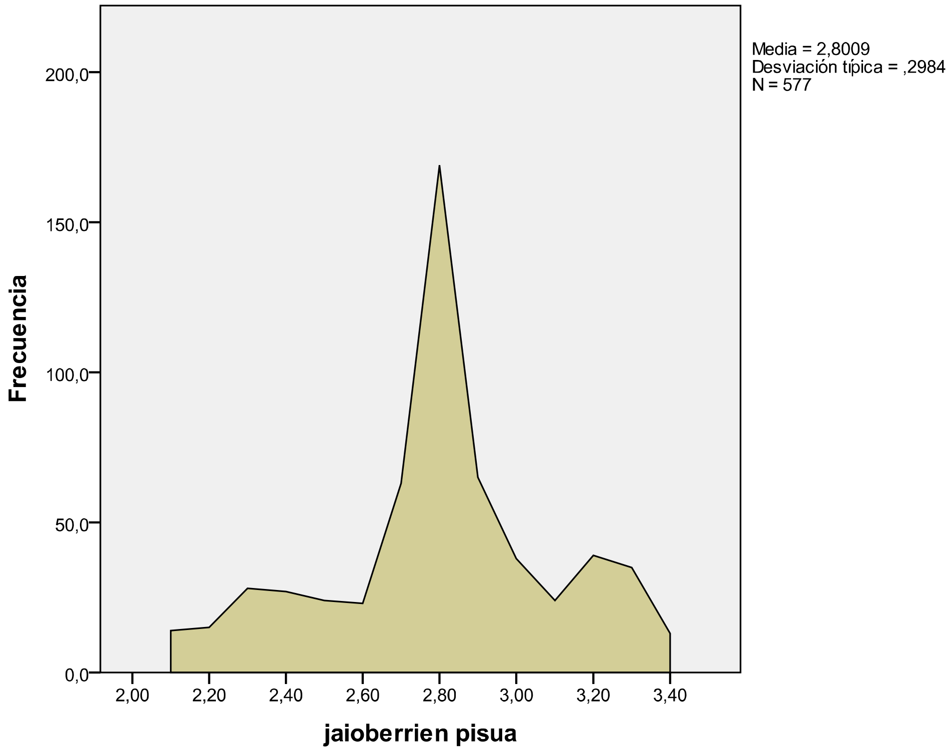
Pisua (x_i)	f_i	%
2,10	14	2,4
2,20	15	2,6
2,30	28	4,9
2,40	27	4,7
2,50	24	4,2
2,60	23	4,0
2,70	63	10,9
2,80	169	29,3
2,90	65	11,3
3,00	38	6,6
3,10	24	4,2
3,20	39	6,8
3,30	35	6,1
3,40	13	2,3
Total	577	100,0

Iturria: lanketa propioa

Datu multzo horretatik abiatuz, maiztasun bakunen honako poligono hau lortzen dugu, IBM SPSS programa informatikoaren laguntzarekin:

17. diagrama

36. taula: bigarren hospital bateko jaioberrien pisuak



Iturria: lanketa propioa

Maiztasun bakunen poligonoaren eraketa

- Maiztasun poligonoa irudikatzeko, Ardatz Kartesiarretako ardatz bertikalean, balio edo tarte bakoitzarentzako puntu bat irudikatzen dugu, balio edo tarte horri dagokion maiztasunaren arabera. Puntu horiek guztiak zuzenen bidez elkarrekin lotzean sortzen den lerro poligonala da maiztasun poligonoa. Ahantzi gabe poligonoa osatzeko zuzena zerotik hasten dela, eta zeroan bukatzen.
- Maiztasun bakunen poligonoa barra-diagrama edo histogramaren gainean ere irudika dezakegu. Horretarako, barren goiko aldeko erdiko puntuak zehazten dira, eta, ostean, puntu horiek guztiak zuzenen bidez lortzen.

Maiztasun poligonoaren eraketa desberdina da aldagai jarraituetarako eta aldagai diskretuetarako.

Eraketa aldagai jarraituentzako

- Abzisa ardatzean, tarteak kokatu beharrean, klase-markak (edo tarte-ordezkariak) zehazten dira.
- Klase-markak eta bakoitzari dagozkion maiztasunak gurutzatuz, puntu batzuk lortzen dira, Ardatz Kartesiarretan.
- Puntu horiek zuzenen bidez lotzen dira.
- Poligonoa zerotik hasten da, eta zeroan bukatzen. Horregatik da, hain zuzen ere, poligonoa. Horretarako, lehenengo emaitzaren aurreko balioa eta azken emaitzaren hurrengo balioa zero egiten dira.

Eraketa aldagai diskretuetarako

- Ardatz katesiarrean, maiztasunaren arabera, balio bakoitzerako puntu bat irudikatzen dugu.
- Puntu horiek guztiak zuzenen bidez lortzen dira.
- Poligonoa zerotik hasi eta zeroan bukatzen da. Aurreko kasuan bezala, lehenengo emaitzaren aurreko balioa eta azken emaitzaren hurrengo balioa zero egiten dira.

Maiztasun bakunen poligonoak erabiltzerakoan kontuan izan beharrekoak

- Maiztasun bakunen poligonoak oso erabilgarriak dira aldagaiak (kuantitatiboa diskretuak zein jarraituak) balio edo tarte desberdin asko dituenean, edota aldagaiaren balioen barruan gertatzen diren maiztasun-aldaketak erakutsi nahi ditugunean.
- Erraz erakuts ditzakegu datu multzoaren berezitasunak. Adibidez, modu errazean ikus dezakegu zein diren maiztasun handiena duten balioak edo tarteak; hau da, zein balio edo tarteren inguruan pilatzen diren datuak. Horregatik, hain zuzen ere, asko erabiltzen dira.
- Lagungarri onak dira bi maiztasun-banaketa edo gehiago alderatzeko. Alegia, Ardatz Kartesiarretan, datu multzo desberdinei dagozkien poligonoak irudikatzen ditugu, alderaketa errazteko. Ulerterraztasunaren izenean, horrelako kasuetan gomendatzen da maiztasun bakunen poligonoak maiztasun erlatiboekin egitea.
- Datu-diagrama guztiek bezala, maiztasun bakunen poligonoek ere izenburua eta iturria adierazi behar dute.
- Datu-diagrama guztiek bezala, identifikazio-zenbakia eraman behar dute. Datu-diagramen aurkibidean, maiztasun bakunen poligonoen zenbakiak, izenburuak eta dauden orrialdeak agertuko ditugu.

2.2.6. Maiztasun metatuen poligonoa edo ojiba

Maiztasun bakunekin eratu ordez, poligonoa maiztasun metatuekin eratzen bada, maiztasun metatuen poligonoa edo ojiba lortzen dugu. Kontuan izan ojibak maiztasun erlatiboekin zein absolutuekin egin ditzakegula.

Adibidea. Aldagai jarraitu baten maiztasun absolutu metatuen eta maiztasun erlatibo metatuen poligonoa: bigarren ospitale bateko jaioberrien pisua.

23. taula: bigarren ospitale bateko jaioberrien pisua

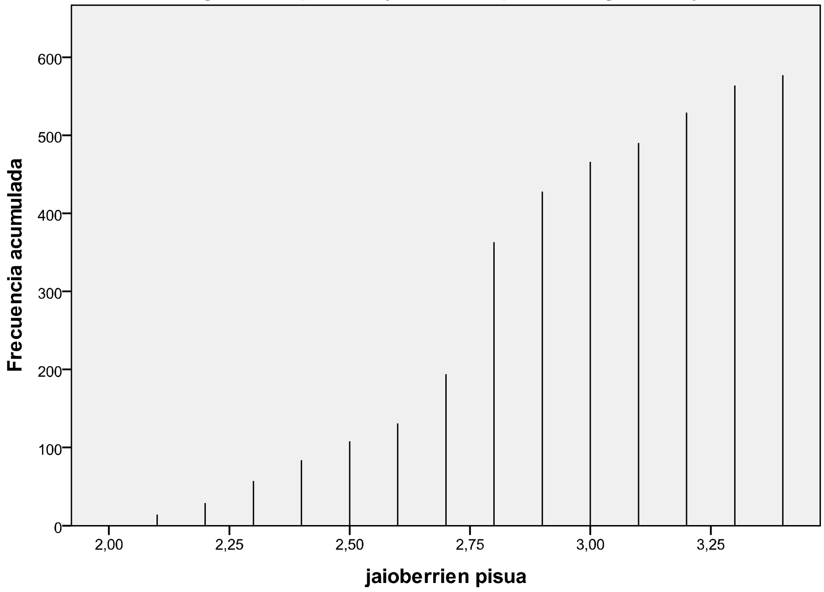
x_i	f_i	F_i	%	% metatuak
2,10	14	14	2,4	2,4
2,20	15	29	2,6	5,0
2,30	28	57	4,9	9,9
2,40	27	84	4,7	14,6
2,50	24	108	4,2	18,7
2,60	23	131	4,0	22,7
2,70	63	194	10,9	33,6
2,80	169	363	29,3	62,9
2,90	65	428	11,3	74,2
3,00	38	466	6,6	80,8
3,10	24	490	4,2	84,9
3,20	39	529	6,8	91,7
3,30	35	564	6,1	97,7
3,40	13	577	2,3	100,0
Total	577		100,0	

Iturria: lanketa propioa

Datu multzo horretako maiztasun absolutu metatuetatik abiatuz, maiztasun absolutu metatuen honako poligono hau lortzen dugu, IBM SPSS programa informatikoaren laguntzarekin:

18. diagrama

32. taula: bigarren ospitaleko jaioberrien pisuari dagokion ojiba

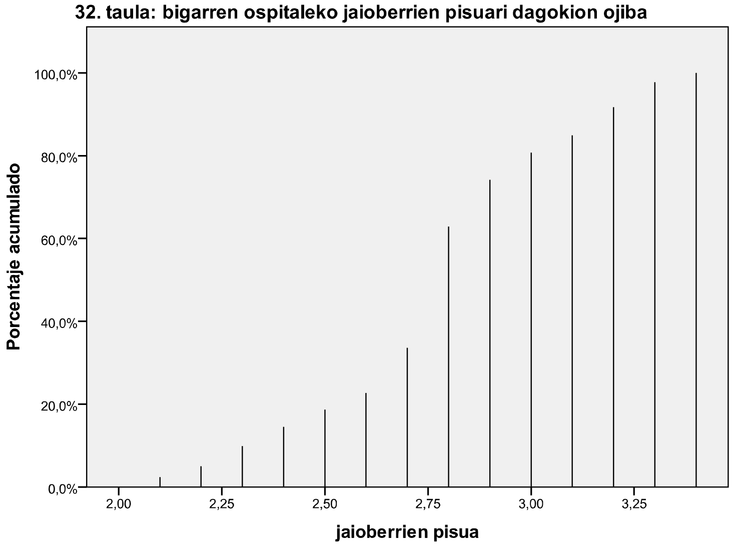


Iturria: nik neuk egina

Maiztasun absolutu metatuak erabili ordez maiztasun erlatibo metatuak erabiltzen baditugu, maiztasun erlatibo metatuen honako poligono hau lortzen dugu:

19. diagrama

32. taula: bigarren ospitaleko jaioberrien pisuari dagokion ojiba



Iturria: nik neuk egina

Maiztasun metatuen poligonoaren eraketa

- Maiztasun metatuak (absolutuak zein erlatiboak) kalkulatu ostean, histograma metatua irudikatzen dugu.
- Lehenengo zutabearen jatorritik abiatuta, zutabearen eskuineko ertzeraino zuzen bat luzatzen da, eta, hortik, hurrengo zutabearen eskuineko ertzeraino, zutabe guztiak elkartu arte.
- Mota horretako poligonoetan, zerotik hasten da, baina azken baliora iritsi ondoren ez da zerora jaisten.

Maiztasun metatuen poligonoak erabiltzerakoan kontuan izan beharrekoak

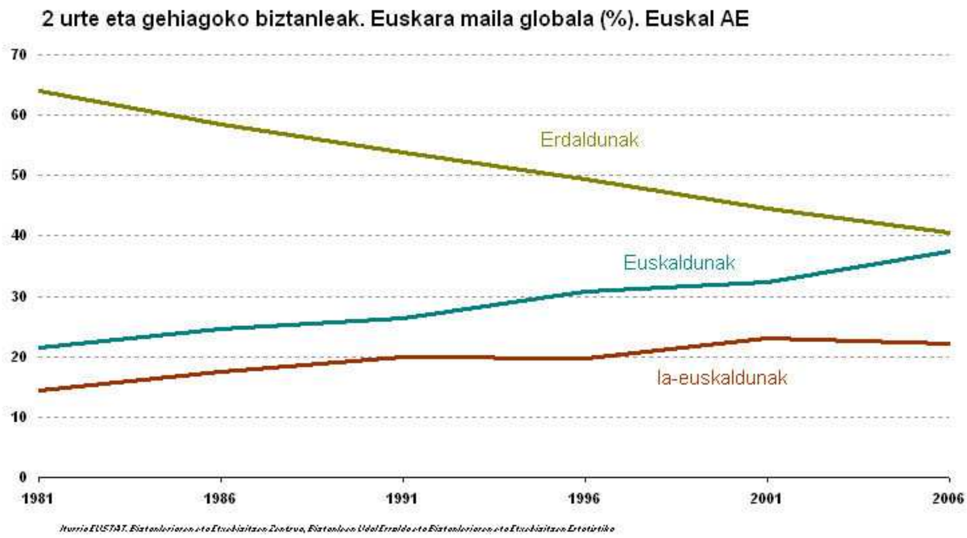
- Ojibaren bidez zenbat datu dauden tarte edo balio baten goitik eta zenbat azpitik ikus dezakegu; beraz, oso erabilgarriak dira geroago aurkeztuko ditugun koantilak irudikatzeko edota aztertzeko.
- Bi ojiba batera irudikatzen baditugu, bi datu multzoren arteko alderaketak egin ditzakegu.
- Ardatz bakoitzak adierazten duena argi idatzi behar da.
- Datu-diagrama guztiek bezala, maiztasun metatuen poligonoek ere izenburua eta iturria adierazi behar dute.
- Datu-diagrama guztiek bezala, identifikazio-zenbakia eraman behar dute. Datu-diagramen aurkibidean, maiztasun metatuen poligonoek zer zenbaki, izenburu dute eta zein orrialdetan dauden agertuko dugu.

2.2.7. Lerro-diagrama

Lerro-diagramak maiztasun bakunen poligonoak bezalakoak dira, baina, kasu honetan, sortzen den lerroa ez da zeroan hasten eta bukatzen. Alegia, ez dugu poligono bat sortzen, lerro bat baizik.
Lerro-diagrama aldagai kuantitatiboen (diskretuen zein jarraituen) maiztasun-banaketak irudikatzeko erabiltzen da. Bereziki egokia da denbora-serieak edo serie kronologikoak irudikatzeko; hau da, denboran zehar aldagaiak izan duen bilakaera erakusteko. Kontuan izan behar da denbora askoko datuak ditugunean, hau da, serie kronologiko luzeak ditugunean, barra-diagramak nahasiak izan daitezkeela.

Adibidea. Lerro-diagrama denbora-serie baterako: 2 urte eta gehiagoko biztanleak, euskara-maila globalaren arabera, Euskal AEn (1981-2006).

20. diagrama

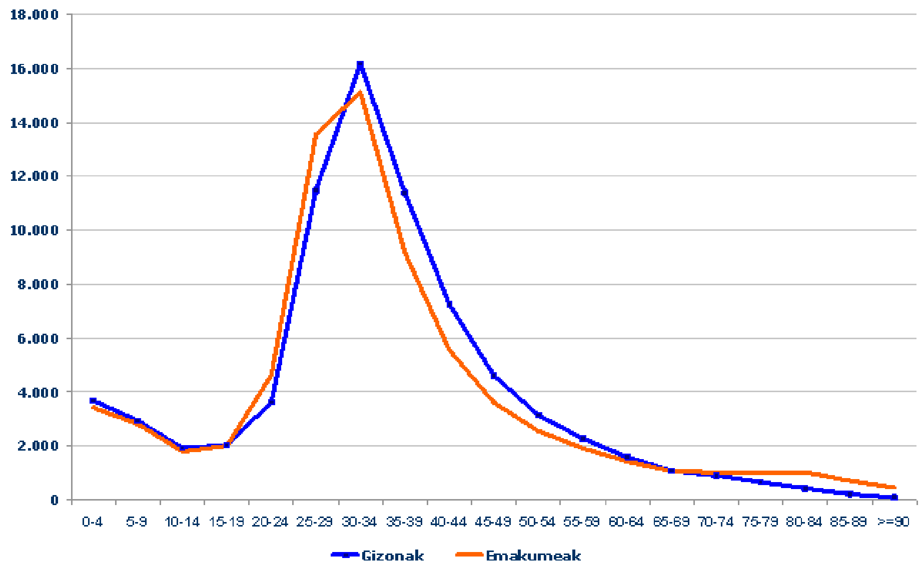


Lerro-diagramak nola eratu

- Denbora-serie bat dugunean, ardatz horizontalean, denbora zehazten da (urteak, hiruhilekoak, egunak, orduak, minutuak…) eta ardatz bertikalean, maiztasunak. Bi horien gurutzaketatik lortzen diren puntuak zuzenen bidez lotzen ditugu, denboraren arabera aldagaia nola aldatzen den erakusteko.
- Denbora ez den beste aldagai diskretu bat dugunean ere, ardatz kartesiarrean,balio bakoitzarentzako puntu bat irudikatzen dugu maiztasunaren arabera, eta, ostean, puntu horiek guztiak zuzenen bidez elkartzen.
- Aldagai jarraitu bat dugunean, abzisa ardatzean, muga errealak kokatu beharrean, klase-markak (edo tarte-ordezkariak) zehazten dira. Ostean, klase-markak eta maiztasunak gurutzatuz lortzen ditugun puntuak zuzenen bidez lortzen ditugu.

Adibidea. Lerro-diagrama aldagai jarraituetarako: 2001etik 2006ra bitartean Euskal AEn udalerriz aldatutako biztanleria, adin eta sexuaren arabera.

21. diagrama: 2001etik 2006ra bitartean, Euskal AEn, udalerriz aldatu den biztanleria, adin eta sexuaren arabera.



Iturria: EUSTAT (www.eustat.es)

Lerro-diagrama horretan ikus daitekeen bezala, Euskal AEn, 25-35 urteen artean aldatzen da jenderik gehien udalerriz. Halaber, ikusten da emakumezkoak gizonezkoak baino lehentxeago hasten direla udalerriz aldatzen.

Lerro-diagramak erabiltzerakoan kontuan izan beharrekoak

- Oso erabilgarriak dira aldagaiak (kuantitatibo diskretuak zein jarraituak) balio edo tarte desberdin asko dituenean, edota aldagaiak denboran zehar izan duten bilakaera erakutsi nahi dugunean.
- Asko erabiltzen dira datu multzo desberdinak alderatzeko; hau da, datu multzo desberdinen bilakaerak alderatzeko.
- Datu-diagrama guztiek bezala, izenburua eta iturria adierazi behar dute.
- Datu-diagrama guztiek bezala, identifikazio-zenbakia eraman behar dute. Datu-diagramen aurkibidean, lerro-diagrama bakoitzak zer zenbaki, izenburu dute eta zein orrialdetan dauden agertuko dugu.

2.2.8. Kaxa-diagrama

Kaxa-diagrama kuartiletan oinarritutako diagrama da; kaxa batez eta beso, adar, bibote edo “whiskers” batzuez osaturik dago. Hala, kaxa eta biboteen diagrama ere deitzen zaio (ingelesezko box plot, boxplot edo box and whisker diagram).
Aldagai kuantitatiboetarako diagrama da; eta haren berezitasuna da datu multzo bateko ezaugarriak era integratuan agertzen dituela. Zehatzago esanda, zentro-neurri bat (mediana), sakabanatze-neurri bat (kuartil arteko ibiltartea), alborapena eta kurtosia era integratuan eta grafikoan agertzen ditu. Horrez gain, datu arraroak modu grafikoan aurkezten dizkigu.

Kaxa-diagramak eratzeko prozedura

Diagramak laukizuzen formako kaxa bat du, eta kaxa horren luzerak (IBM SPSSan bertikalki begiratuta) kuartil arteko ibiltartea adierazten du; hau da, hirugarren eta lehengo kuartilaren arteko distantzia. Laukizuzena mediana adierazten duen marra batez banatuta dago, eta, beraz, lehen eta hirugarren kuartilekiko mediana non kokatzen den adierazten digu (kontuan izan bigarren kuartila eta mediana gauza bera direla).
Kaxatik ateratzen diren zuzenak biboteak, adarrak edo besoak dira. Adar horietatik kanpo kokatzen dira datu edo kasu bitxiak.
Hala, kaxa-diagramak datu multzo batean Q_1, Q_2 (edo mediana) eta Q_3 kuartilek hartzen dituzten balioak agertzen ditu, eta balio arraroei eta maiztasun-banaketaren simetriari eta kurtosiari buruzko informazioa adierazten.

Adibidea. Kaxa-diagrama: ospitale batean jaiotako 577 umeren pisua.

24. taula: jaioberrien pisua

		Maiztasuna	Ehunekoa	Ehunekoa	Ehunekoa metatua
Baliodun ak	2,10	14	2,4	2,4	2,4
	2,20	15	2,6	2,6	5,0
	2,30	28	4,9	4,9	9,9
	2,40	27	4,7	4,7	14,6
	2,50	24	4,2	4,2	18,7
	2,60	23	4,0	4,0	22,7
	2,70	63	10,9	10,9	33,6
	2,80	169	29,3	29,3	62,9
	2,90	65	11,3	11,3	74,2
	3,00	38	6,6	6,6	80,8
	3,10	24	4,2	4,2	84,9
	3,20	39	6,8	6,8	91,7
	3,30	35	6,1	6,1	97,7
	3,40	13	2,3	2,3	100,0
	Total	577	100,0	100,0	

Iturria: lanketa propioa

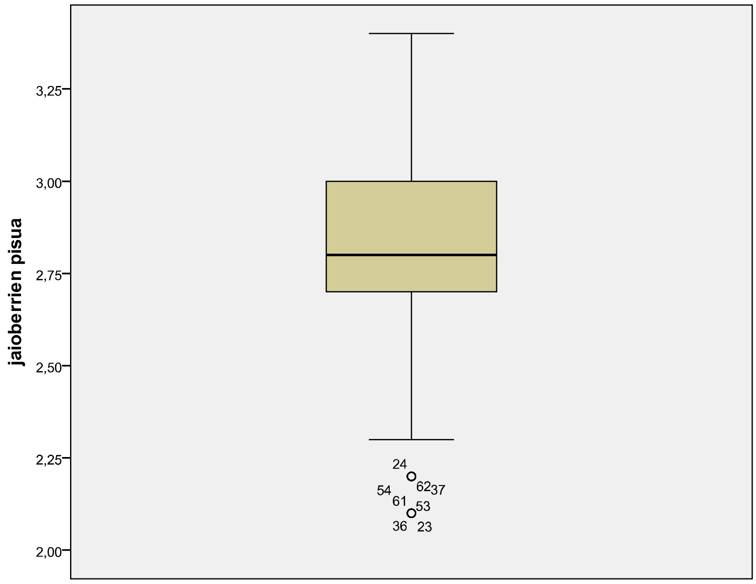
Maiztasun-taula horretatik abiatuta, honako mediana, kuartil, kurtosi eta asimetria-neurri hauek lortzen ditugu, IBM SPSS programaren laguntzarekin:

Estadísticos
Jaioberrien pisua

N	Baliodunak	577
	Galduak	0
Mediana		2,8000
Asimetria		-,236
Asim. Error tip.		,102
Kurtosis		-,111
Kurtosiaren Error tip.		,203
Pertzentilak	25	2,7000
	50	2,8000
	75	3,0000

Estatistiko horietatik abiatuta, honako kaxa-diagrama hau lortzen dugu:

22. diagrama: jaioberrien pisua



Iturria: nik neuk egina

Hala, kaxa-diagrama irudikatzeko, honako estatistiko hauek behar ditugu:

Q_1 = lehen kuartila
Q_2 = bigarren kuartila edo mediana
Q_3 = hirugarren kuartila

Estatistiko horien laguntzarekin, kaxa eta besoak irudikatuko ditugu, baita balio arraroak zehaztu ere:

- Hasteko, lehenengo kuartila (Q_1), hirugarren kuartila (Q_3) eta mediana kalkulatu behar dira.
- Ostean, kuartil arteko ibiltartea (KAI) kalkulatu behar da, hirugarren koartilari lehenengo koartila kenduta.
- Laukizuzen bat, kaxa, alegia, marraztu behar da; lehenengo koartila eta hirugarren koartila mugatzat ditu.
- Mediana marra batez irudikatzen da kaxaren barnean.
- Kaxako alde banatik, lehenengo eta hirugarren kuartilak kokatzen diren tokietatik hain zuzen ere, beso, bibote edo adarrak luzatzen dira. Beso, bibote edo adarren, hau da kaxatik ateratzen diren marren mugak zehazteko, balio arraroak edo atipikoak ezagutu behar ditugu. Hain zuzen ere, balio arraroak dira Q_1 - 1.5 x KAI (kuartil arteko ibiltartea) baino txikiagoak edo Q_3 + 1.5 x KAI baino handiak direnak. Balio arraroak zein diren jakin ostean, atipikoak edo arraroak ez diren azken balioak bilatu behar dira. Horiek biboteen, adarren edo besoen muturrak zehazteko erabiliko ditugu. Hau da, marrak egingo ditugu Q_1etik eta Q_3tik (kaxaren muturretatik) arraroak ez diren azken balioetaraino.
- Biboteen muturretatik harago dauden datuak muturreko datutzat edo datu atipikotzat hartzen dira, eta puntuen bidez adierazten dira.

Kaxa-diagrama interpretatzeko irizpideak

- Sakabanatze-neurritzat, kuartil arteko ibiltartea hartzen da. Kaxaren zabalerak adierazten du kuartil arteko ibiltartea. Esan bezala, IBM SPSS programak, kaxa eraikitzeko balioen eskala, zein kaxa bera ere, era bertikalean agertzen ditu; alegia, kuartil arteko ibiltartea ezagutzeko kaxari goitik behera begiratu behar zaio.

- Kaxaren barruan dagoen marrak mediana adierazten du. IBM SPSS programan, ardatz bertikalean agertzen da kaxa-diagrama ulertzeko balioen eskala. Beraz, kaxaren irakurketa ere modu bertikalean egin behar da. Hala, mediana adierazten duen marra era horizontalean agertzen da; beste programa batzuetan, berriz, era bertikalean agertzen da. Kontuan izan medianak kaxaren erdian egon behar duela, maiztasun-banaketa simetrikoa dela esateko. Mediana erabiltzen da, batez besteko aritmetiko sinplea ez bezala, mediana neurri jasankorra delako; hau da, muturreko datuen eraginpean ez dagoelako.
- Alborapena aztertzeko, medianatik alde banatara kuartiletaraino edo bikoteen muturretaraino dauden distantziak ere alderatzen dira. Eskuineko distantzia (medianatik gora) ezkerreko distantzia (medianatik behera) baino handiagoa bada, maiztasun-banaketak alborapen positiboa duela esaten da; ezkerreko distantzia eskuinekoa baino handiagoa bada, berriz, maiztasun-banaketak alborapen negatiboa duela esaten da.
- Kurtosiaren azterketa kaxaren eta biboteen luzera alderatuz egiten da: kaxa besoak baino zabalagoa bada, maiztasun-banaketa platikurtikoa edo zapala da; biboteak luzeagoak badira, maiztasun-banaketa leptokurtikoa edo zorrotza da.

Kaxa-diagramak erabiltzerakoan kontuan izan beharrekoak

- Diagrama horren bitartez, zentroa, sakabanatzea, alborapena eta kurtosia modu integratuan azter daitezke. Hori da, hain zuzen ere, datu-diagrama horren abantaila nagusia.
- Datu multzo desberdinak alderatu daitezke modu errazean.
- Oso erabilgarria da balio arraroak ikusteko. Izan ere, modu grafikoan erakusten ditu era horretako balioak.
- Datu-diagrama guztiek bezala, kaxa-diagramek ere izenburua eta iturria adierazi behar dute.
- Datu-diagrama guztiek bezala, identifikazio-zenbakia eraman behar dute. Datu-diagramen aurkibidean, kaxa-diagramek zer zenbaki eta izenburu duten eta zein orrialdetan dauden agertuko dugu.

Adibidea. Kaxa-diagrama: hiri bateko martxoko tenperatura. Hiri batean apirileko egunetan izan den tenperatura jaso dugu, eta honelako taula bat lortu dugu:

25. taula: apirileko hotz-beroa hiri batean

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	14,00	3	9,7	9,7	9,7
	15,00	7	22,6	22,6	32,3
	16,00	7	22,6	22,6	54,8
	17,00	4	12,9	12,9	67,7
	18,00	6	19,4	19,4	87,1
	19,00	3	9,7	9,7	96,8
	25,00	1	3,2	3,2	100,0
	Total	31	100,0	100,0	

Iturria: lanketa propioa

Taula horretatik abiatuta, mediana, asimetria, kurtosia eta kuartilak lortzen ditugu, IBM SPSS programaren laguntzarekin.

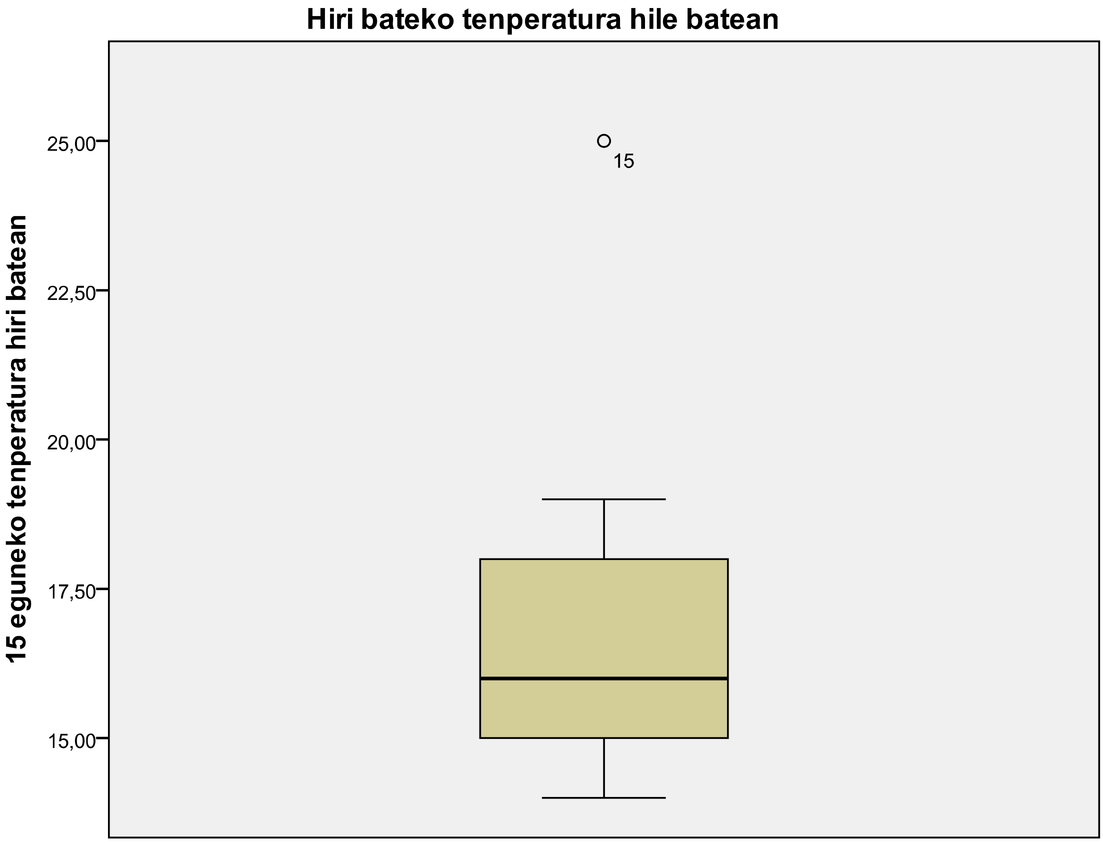
Estadísticos
15 eguneko tenperatura hiri batean

N	Válidos	31
	Perdidos	0
Mediana		16,0000
Asimetría		1,876
Error típ. de asimetría		,421
Curtosis		6,134
Error típ. de curtosis		,821
Percentiles	25	15,0000
	50	16,0000
	75	18,0000

Kaxa-diagrama egiteko behar ditugun estatistikoak ditugularik, kaxa irudikatzeko behar ditugun balioak kalkula ditzakegu:

Mediana	Me = 16	Datu ordenatuen erdian dagoen balioa.
Lehenengo kuartila	$Q_1 = 15$	Balio horren azpitik, datuen %25 dago.
Hirugarren kuartila	$Q_3 = 18$	Balio horren gainetik, datuen %25 dago (eta azpitik %75).
Kuartil arteko ibiltartea	$Q_3 – Q_1 = 3$	Lehenengo kuartiletik hirugarren kuartilera, datuen %50 dago.
Goiko bibotea	$Q_3 + 1,5 (Q_3 – Q_1) = 22,5$	Goiko biboteaz haraindiko muturreko datu bakarra dago. Hau da, 20 gradutik gorako tenperatura bakarra dugu: 25º
Goiko bibotearen muturra edo azken balioa	19	Bibotea 19. balioan moztuko dugu.
Beheko bibotea	$Q_1 + 1,5 (Q_3 – Q_1) = 19,5$	Behe-biboteaz haraindiko muturreko daturik ez dago.
Beheko bibotearen muturra edo azken balioa	15	Bibotea 15 balioan moztuko dugu.

23. diagrama



Iturria: nik neuk egina

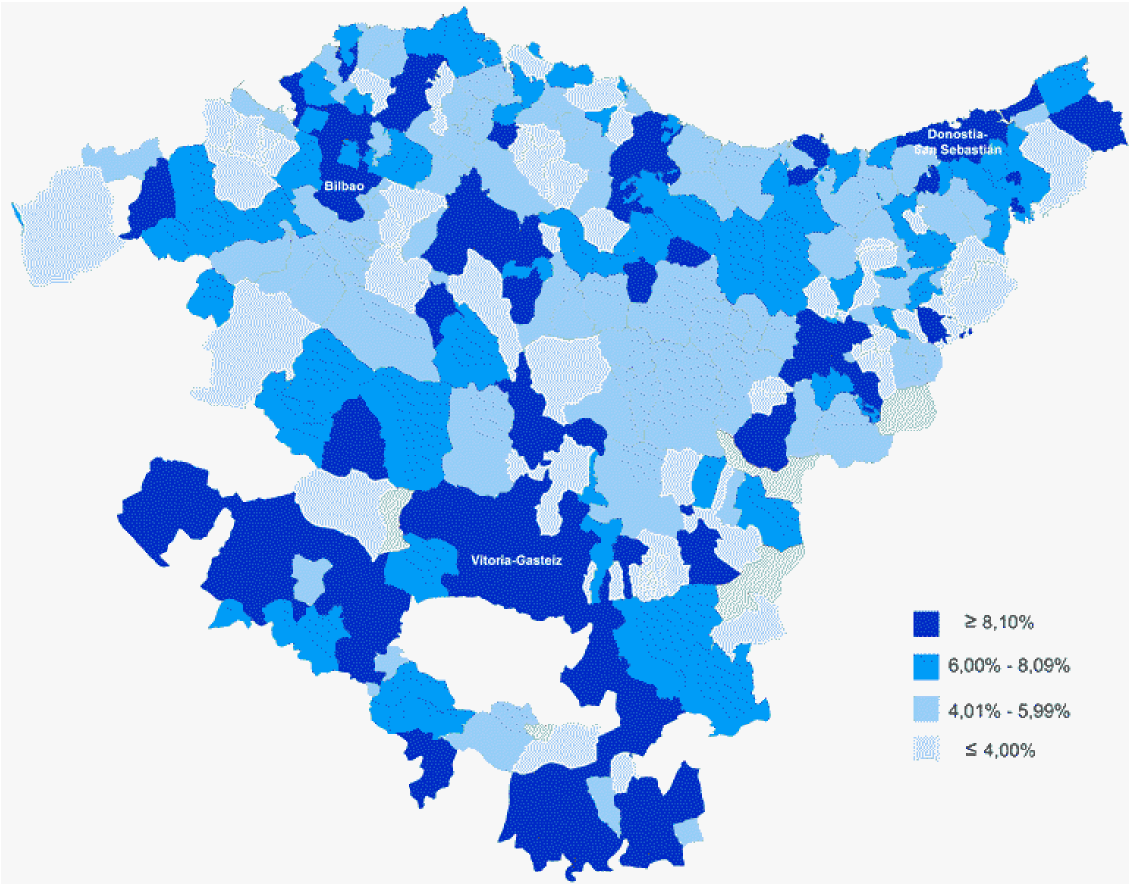
Diagraman ikus daitekeen bezala, lehenengo (15) eta hirugarren (18) kuartilek mugatzen dute kaxa, eta, horien artean, mediana kokatzen da (16). Q_3-tik bibote bat ateratzen da 19. balioraino. Hortik aurrera, 15. datua (25º) datu arraro gisa agertzen da. Q_1-etik beste bibote bat ateratzen da 15. balioraino. Balio horretatik kanpo ez dago balio arrarorik.

2.2.9. Kartograma

Eremu geografiko batean aldagai baten neurketatik lortu ditugun maiztasunak agertzeko, barra-diagrama edo sektore-diagramen ordez, kartogramak erabil ditzakegu. Alegia, maiztasunak geografikoki nola banatzen diren erakusteko, mapa bat erabili dezakegu, eta, horrela, geografikoki aldatzen diren maiztasunak modu erakargarrian agertu.

Adibidea. Kartograma morfikoa: atzerrian jaiotakoen ehunekoa, Euskal AEn udalerrietan.

24. diagrama: Atzerrian jaiotakoen ehunekoa udalerrika Euskal AEn



Iturria: EUSTAT. Biztanleen udal-estatistika 2011ko urtarrilaren 1a (www.eustat.es)

Kartograma horretako koloreek adierazten duten bezala, Euskal AEko hainbat herritan, eta baita hiru hiriburuetan ere, %8tik gorakoa da atzerrian jaiotakoen ehunekoa.

Kartograma eratzeko prozedura

Egin nahi dugun kartogramaren arabera, eratze-prozesua desberdina da:

- Maparen eremu geografikoak dagozkien maiztasun absolutu zein erlatiboen arabera koloreztatzea da kartograma egiteko modu errazena. Lehenengo, baina, maiztasun bakoitzari dagokion kolorea, hau da, idazkuna, zehaztu behar dugu.
- Maparen eremu geografiko bakoitzaren gainean sektore-diagramak edo barra-diagramak irudikatzea beste aukera bat da.
- Hirugarren aukera kartograma anamorfikoak egitea da. Horretarako, mapa erreala distortsionatu egin behar da, maiztasunen arabera; hau da, mapako eremu geografikoak handitu edo txikitu egingo ditugu, maiztasunen arabera.

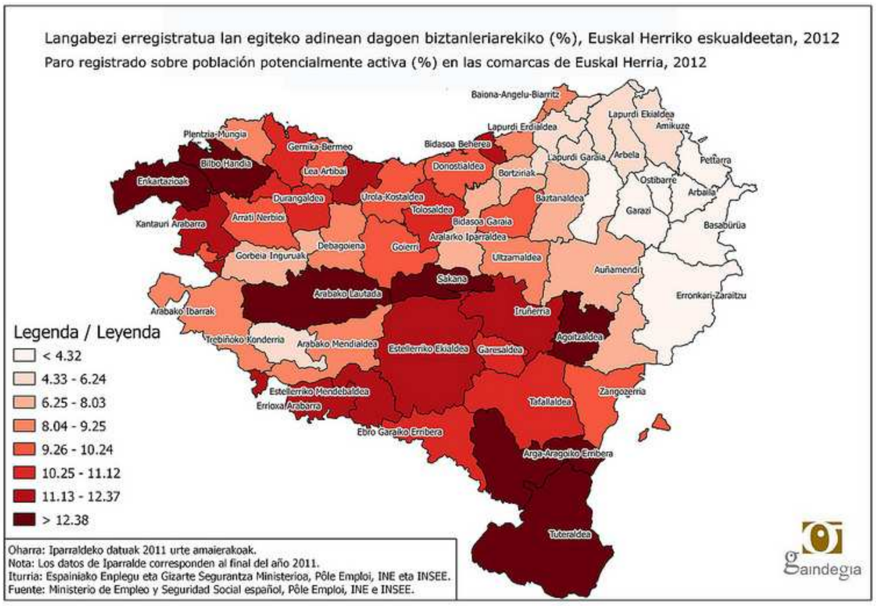
Kartograma motak

Aipatu bezala, bi eratako kartogramak daude: morfikoak eta anamorfikoak.
Kartograma morfikoetan, mapa errealak erabiltzen dira. Hau da, mapan agertzen diren herrialde, lurralde edo eskualdeetako muga errealak errespetatzen dira. Bi eratako kartograma morfikoak daude:

- Kolore edo tonu bakoitzak adierazten duen maiztasuna adierazi ostean (hau da, irudi-testua adierazi ostean), maparen eremu geografikoak dagokien maiztasunen arabera koloreztatzen ditugu.
- Maparen eremu geografiko bakoitzean sektore-diagramak edo barra-diagramak irudikatzea beste aukera bat da.

Adibidea. Kartograma morfikoa: langabezia Euskal Herriko eskualdeetan, 2012.

25. diagrama: langabezia Euskal Herriko eskualdeetan, 2012



Iturria: Gaindegia.

Kartograma morfikoez gain, anamorfikoak ditugu. Kasu horretan, eremu geografikoen tamaina distortsionatu egiten da, maiztasunen arabera; hau da, mapako eremu geografikoak handitzen edo txikitzen dira maiztasunen tamaina kontuan hartuta. Adibidez, biztanleriaren kopuruaren arabera mapako zonalde desberdinen tamainak handitu edo txikitu egiten dira. Kartograma anamorfikoek badute abantaila bat, informazioa modu ekargarri edota ikusgarrian agertzen duela. Baina, lurraldeak distortsionatzen direnez, askotan zaila izaten dela eremu geografiko zehatzak identifikatzea.

Kartogramak erabiltzerakoan kontuan izan beharrekoak

- Aldagai kuantitatiboak zein kualitatiboak irudikatzeko erabil ditzakegu.
- Geografikoki aldatzen diren maiztasunak modu erakargarrian ager ditzakegu. Beraz, oso egokiak dira ikerketak ilustratzeko edota arintzeko.
- Egokiak izan daitezke gaia ulertzeko zailtasunak izan ditzaketen pertsonei eremu geografiko desberdinetako maiztasunak edota errealitateak azaltzeko.
- Erabilgarriak dira irakurleek mapa ondo ezagutzen dutenean; hau da, mapan ondo kokatzen direnean. Kartograma irakurri behar dutenentzat mapa ezezaguna denean, ez da komeni kartogramak egitea.
- Mapa handiegia bada edo maiztasun desberdin asko irudikatu behar baditugu, kartograma ulergaitza izan daiteke, zonaldeen artean bereiztea asko kostatzen delako.
- Datu-diagrama guztiek bezala, kartogramek ere izenburua eta iturria adierazi behar dute.

- Datu-diagrama guztiek bezala, identifikazio-zenbakia eraman behar dute. Datu-diagramen aurkibidean kartogramek zer zenbaki eta izenburu duten eta zein orrialdetan dauden agertuko dugu.

2.2.10. Datu-diagramei buruzko gomendio orokorrak

- Datu-diagramak azterketa estatistikoaren osagarri onak dira. Datu multzoen izaera hobeto ulertzen laguntzen dute, eta txostenak errazago irakurtzen dira. Hala ere, ikerketa-txostenetan ez dira sartu behar datu multzo bati buruzko maiztasun-taula eta datu-diagrama, biak batera; hain zuzen ere, errepikakorra delako.
- Esan bezala, datu-diagramak azterketa estatistikoaren osagarri dira, baina diagrama edo grafikoek ezin dute ikerketa estatistikoa ordezkatu.
- Datu-diagrama aukeratzerakoan, kontuan izan aldagaiaren berezitasunak eta ikerketaren helburuak.
- Diagramak egin aurretik, begiratu zenbat datu ditugun. Izan ere, diagrama batzuk egiteko gutxieneko datu kopurua behar dugu.
- Eskala desberdinetan neurtutako aldagaiek grafiko mota desberdinak onartzen dituzte: aldagaiak kualitatiboak edo kuantitatibo diskretuak direnean, sektore-diagramak eta barra-diagramak erabiltzen dira, nagusiki; aldagai kuantitatibo jarraituak direnean, berriz, histogramak. Aldaketak nabarmendu, denboraren araberako bilakaerak adierazi edo maiztasun-banaketa desberdinen artean alderaketak egin nahi ditugunean, maiztasun bakunen poligonoak eta lerro-diagramak aukeratuko ditugu. Zenbat datu dauden tarte edo balio baten gainetik eta zenbat azpitik erakusteko, hau da, koantilak agertzeko, maiztasun metatuen poligonoak egingo ditugu. Grafiko ulerterrazagoak eta atseginagoak egin nahi baditugu, piktogramak irudikatuko ditugu, eta aldagai baten maiztasun-banaketa geografikoa mapa batean kokatu nahi badugu, kartogramak. Adin desberdinetako gizon eta emakumeen maiztasunak azaltzeko, biztanleriaren piramideak erabiltzen dira. Azkenik, kaxa- eta beso-diagramak egingo ditugu, balio arraroak ezagutzeko eta maiztasun-banaketaren zentroa, sakabanatzea, alborapena eta kurtosia modu integratuan aztertzeko.

26. taula: diagramen erabilera, aldagai motaren arabera

Diagrama mota	Aldagai nominala	Aldagai ordinala	Aldagai diskretua	Aldagai jarraitua
Sektore-diagrama	bai	Bai	bai	bai
Piktograma	bai	Bai	bai	bai
Barra-diagrama	bai	Bai	bai	bai
Kartograma	bai	Bai	bai	bai
Histograma	ez	Ez	bai	bai
Maiztasun bakunen poligonoa	ez	Ez	bai	bai
Ojiba edo maiztasun metatuen poligonoa	ez	Ez	bai	bai
Lerro-diagrama	ez	Ez	bai	bai
Kaxa-diagrama	ez	Ez	bai	bai

Iturria: lanketa propioa

- Kontuz hiru dimentsiotako grafikoekin, kono-, zilindro- eta piramide-itxura duten diagramekin, gehiegizko koloreekin edota itzalekin. Askotan, horiek denak diagramaren ulerterraztasunaren aurka joan daitezke. Diagrama bat txosten batean jarri aurretik ondo begiratu ia ondo ulertzen den. Ez ahantzi datu-diagramak ulergarriak, argiak, sinpleak eta zehatzak izan behar direla.

- Edozein delarik aukeratutako grafikoa, kontuan izan behar dugu grafiko guztiek identifikazio zenbakia, izenburua eta iturria adierazi behar dutela.
- Datu-diagrametan azaltzen diren aldagaien, balioen eta kategorien izenak argi agertu behar dira.
- Datu-diagramak egiteko, programa informatiko asko daude (Microsoft Excel, IBM SPSS, R (askea), SigmaPlot, QtiPlot (askea), Origin, Data Desk OpenOfficee.org Calc), eta nahi dugun sofistikazio-mailako diagrama egin dezakegu, baina, betiere, ulerterraztasunari eman behar zaio lehentasuna.

2.3. NEURRI ESTATISTIKO DESKRIBATZAILEAK

Aurreko bi ataletan, datu multzoak era ulerterrazean plazaratzeko bi bide landu ditugu: maiztasun-taulak eta datu-diagramak. Baliabide estatistiko bi horiek oso erabilgarriak dira datu multzoak deskribatzeko.
Dena dela, datu multzoari buruz gehiago jakin gura badugu, pauso bat aurrerago eman behar dugu; hau da, datu multzoen ezaugarriak deskribatzen dituzten neurri estatistikoak kalkulatzen eta interpretatzen jakin behar dugu.
Neurtzen dituzten ezaugarrien arabera, bost eratako neurri estatistikoak daude:

- Zentro-neurriak, joera zentraleko posizio neurriak edo joera zentraleko neurri estatistikoak: datu multzoa zein balioren inguruan biltzen den adierazten digute. Ezagunenak batez besteko aritmetiko sinplea, mediana eta moda dira.
- Sakabanatze-neurriak: datu multzoaren sakabanatzea adierazten dute. Ezagunenak desbideratze tipikoa eta bariantza dira.
- Joera zentralekoak ez diren posizio neurriak edo koantilak: alde batean datu multzoaren ehuneko zehatza uzten duten neurriak dira. Ezagunenak pertzentilak, koartilak eta dezilak dira.
- Forma-neurriak: maiztasun-banaketaren itxurari buruzko informazioa ematen digute. Ezagunenak asimetria-koefizientea (alborapena) eta kurtosia dira.
- Kontzentrazio-neurriak: aldagaia aztergai den populazioaren barruan nola banatzen den neurtzen dute. Ezagunenak Lorenz-en kurba eta Gini-ren koefizientea dira.

2.3.1. Zentro-neurriak edo joera zentraleko estatistikoak

Zentro-neurriek, joera zentraleko posizio-neurriek edo joera zentraleko neurri estatistikoek datu multzo bateko balioak zein balioren inguruan biltzen diren adierazten digute; erdigunean dauden posizioak zein diren. Asko dira zentro-neurriak. Guk hemen bost aurkeztuko ditugu[8]:

- Batez besteko aritmetiko sinplea
- Batez besteko aritmetiko haztatua
- Batez besteko aritmetiko moztua
- Mediana
- Moda

2.3.1.1. Batez besteko aritmetiko sinplea

Maiztasun-banaketako balio guztien batuketaren emaitza datu guztien kopuruaz zatitzean lortzen den zenbakiari batez besteko aritmetiko sinplea edo batez bestekoa soilik esaten zaio.
Batez bestekoa aldagai kuantitatiboetarako (diskretuak zein jarraituak) kalkula daiteke bakarrik, eta oso erabilgarria da datu multzoaren erdigunea non dagoen jakiteko; alegia, datuak zein balioren inguruan biltzen diren ezagutzeko. Horregatik, hain zuzen ere, batez bestekoa maiztasun-banaketaren “grabitate-zentroa” dela esaten da.
Populazioaren batez besteko aritmetiko sinplea μ (mu) hizki grekoaz adierazten da, eta laginaren batez bestekoa $\bar{x}$ ikurraren bidez[9].

---

[8] Batez besteko geometrikoa, batez besteko harmonikoa eta batez besteko kuadratikoa ere badaude, baina gutxi erabiltzen dira Gizarte zientzietan, eta, beraz, ez ditugu landuko liburu honetan.

[9] Neurri estatistikoek, populazioen ezaugarriei buruz aritzen direnean, parametro izena hartzen dute, eta laginekin lan egiten ari garenean, estatistiko deitzen zaie. Parametroa ezezaguna izaten da ia gehienetan, eta, horren ondorioz, estatistikoen bidez ezagutu behar izaten da, estatistika inferentzialaren prozedurak erabiliz. Parametroak hizki grekoez izendatzen dira. Adibidez, batez bestekoa parametroa denean, μ (mu) hizki grekoa erabiltzen da, eta

Adibidea. Batez besteko aritmetiko sinplea: amatasuneko batez besteko adina, urtearen arabera, Euskal AEn.

27. taula: amatasuneko batez besteko adina, urtearen arabera, Euskal AEn				
	1990/1991	1995/1996	2000/2001	2005/2006
Araba	29,7	31,1	32,0	31,9
Bizkaia	29,9	31,3	32,2	32,5
Gipuzkoa	30,2	31,4	32,2	32,4
Iturria: EUSTAT. Demografi Adierazleak (www.eustat.es)				

Taula horretan ikus daitekeen bezala, amatasuneko batez besteko adina igoz joan da Euskal AEn, 1990/1991 biurtekotik 2005/2006 biurtekora. Aipagarria ere bada Arabako emakumeek Gipuzkoakoek eta Bizkaikoek baino arinxeago izaten dituztela seme-alabak.
Batez besteko aritmetiko sinplea oso erraz ateratzen da. Dena dela, kalkulatzeko modua desberdina da aldagaiaren eta datu multzoaren arabera. Hiru egoera desberdin daude:

- Balio isolatuak ditugunean.
- Maiztasun-tauletan banatuta dauden aldagai diskretuak ditugunean.
- Tartetan banatuta dauden aldagai jarraitu zein diskretuak ditugunean.

Batez besteko aritmetiko sinplea nola kalkulatu balio isolatuetarako

Lagin bateko balioak x_1, x_2, x_3…, x_n badira, batez bestekoa formula honen bidez kalkulatzen da:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{N}$$

Formula hau beste modu honetan laburbil dezakegu:

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

Formulan ikus daitekeen bezala, batez besteko aritmetiko sinplea kalkulatzeko, ditugun balioak batu eta datu kopuruaz (N) zatitzen ditugu.

Adibidea. Batez besteko aritmetiko sinplea kalkulatzea: ikasgela bateko 6 ikasleren etxeetan bizi diren pertsona kopurua 4 – 4 – 5 – 3 - 7 eta 7 badira, batez besteko aritmetiko sinplea honela kalkulatzen dugu:

$$\bar{x} = \frac{4+4+5+3+7+7}{6} = 5$$

Beraz, ikasleak, batez beste, 5 laguneko etxeetan bizi direla ondorioztatuko dugu.

desbideratze tipikoa parametroa denean, σ (sigma) hizkia. Estatistikoak izendatzeko latindar letrak erabiltzen dira. Adibidez, batez bestekoa estatistikoa denean, $\bar{x}$ hizkia erabiltzen da, eta, desbideratze tipikoa estatistikoa denean, s hizkia.

Batez besteko aritmetiko sinplea nola kalkulatu, maiztasun-tauletan banatuta dauden aldagai diskretuetarako

Askotan, aldagai diskretuak maiztasun-tauletan banatuta egoten dira:

x_i	f_i
x_1	f_1
x_2	f_2
x_3	f_3
.	.
.	.
x_n	f_n
	N

Horrelako kasuetan, batez bestekoa honako formula honen bidez kalkulatzen da:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \ldots + x_n f_n}{N}$$

Non x_n balio bakoitza baita eta f_n, balio bakoitzaren maiztasuna.

Formula hori honela ere laburbil dezakegu:

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

Adibidea. Batez besteko aritmetiko sinplea kalkulatzea maiztasun-tauletan banatuta dauden aldagai diskretuetarako: etxean bizi den kide kopurua.

28. taula: etxean bizi den kide kopurua		
x_i	f_i	$x_i \cdot f_i$
3	5	15
4	7	28
5	6	30
6	4	24
7	3	21
8	2	16
N	27	134
Iturria: lanketa propioa		

Batez bestekoa kalkulatzeko, honako formula hau aplikatzen dugu:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{N} \qquad \bar{x} = \frac{134}{27} = 4{,}9$$

Batez besteko aritmetiko sinplea 4,9 da. Dena dela, horrelako emaitza batek ez du zentzurik aldagai diskretuekin lanean ari garenean, are gutxiago pertsonei buruz hitz egiten ari garenean. Beraz, biribilduta, etxe bakoitzean batez beste 5 pertsona bizi direla esango dugu.

Batez besteko aritmetiko sinple kalkulatzea tartetan banatuta dauden aldagai jarraitu zein diskretuetarako

Aldagai jarraituak ditugunean edota aldagai diskretua izanik balioak tartetan bilduta daudenean, tarte bakoitzeko erdiko puntua edo marka (klase-ordezkaria izenarekin ere ezagutzen da) erabiltzen da batez bestekoa kalkulatzeko.

Adibidea. Batez besteko aritmetiko sinplea kalkulatzea, tartetan banatuta dauden aldagaietarako: elkarte gastronomiko-kultural bateko bazkideen batez besteko adina.

29. taula: elkarte bateko bazkideen adinak

x_i (adin tarteak)	f_i (maiztasunak)	Markak	$x_i.f_i$
0-10	2	5	10
10-20	3	15	45
20-30	4	25	100
30-40	3	35	105
N	12		260

Iturria: lanketa propioa

Markak kalkulatu eta maiztasunekin biderkatu eta gero, erraz lortzen da batez bestekoa:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{N} \qquad \bar{x} = \frac{260}{12} = 21{,}6$$

Beraz, elkarte gastronomiko-kultural horretako kideen batez besteko adina 21 urte da.

Batez besteko aritmetiko sinplearen abantailak eta desabantailak

- Batetik, batez besteko aritmetiko sinplearen abantaila nagusia erraz kalkulatzen dela da, eta, bestetik, esanahi argia duela; alegia, interpretazio erraza du. Horregatik, hain zuzen ere, askotan erabiltzen da, batez ere talde desberdinak alderatzeko.
- Batez bestekoaren desabantaila zera da: muturretan kokatzen diren balio arraroek edo ohikoak ez direnek eragin handia dutela emaitzan, eta horrek zentzurik gabeko interpretazioetara eraman gaitzakeela. Adibidez, 5 pertsonen adinak 2 – 2 – 2 – 2 – 22 badira, talde horren batez besteko adina 6 da, baina argi dago 6ko batez bestekoa ez dela datu multzo horren erdigunean gertatzen denaren adierazle bikaina; izan ere, 22 datua handia izanik, gorantz eramaten du batez bestekoa. Horrelako kasuetan, datu multzoaren erdigunean gertatzen dena ezagutzeko, aproposagoa da ondoren ikusiko dugun zentralizazio-neurria: mediana. Hala, esaten da batez bestekoa ez dela estatistiko jasankorra edo sendoa; hain zuzen ere, muturreko datuek eragin handia dutelako beregan.

2.3.1.2. Batez besteko aritmetiko haztatua

Batez besteko aritmetiko haztatuan edo batez besteko haztatuan, balioei haztapen edo pisu desberdina ematen zaie, balioaren garrantziaren edo adierazgarritasunaren arabera.

Lagin bateko balioak x_1, x_2, ..., x_n badira, eta haiei dagozkien pisuak, w_1, w_2, w_3..., w_n, batez besteko aritmetiko haztatua honela kalkulatzen da:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i} = \frac{x_1 w_1 + x_2 w_2 + x_3 w_3 + \dots + x_n w_n}{w_1 + w_2 + w_3 + \dots + w_n}$$

Ikusten den bezala, lehenengo datu multzoko balio bakoitza haztatzeko erabili ditugun kopuruekin biderkatzen dugu, eta, horren ostean, haztatzeko erabili ditugun kopuru guztien baturarekin zatitzen.

Adibidea. Batez besteko aritmetiko haztatua kalkulatzea: estatistikako azterketa.

Estatistika ikasgaiak zati praktikoa eta zati teorikoa ditu. Ikasle batek, zati praktikoan, 6 puntu lortu ditu, eta teorikoan, 7. Ikasturte hasieran, irakasleak esan zuen zati praktikoak notaren %40 balio zuela eta teorikoak, %60. Zenbatekoa da, batez beste, ikasle horrek atera duen nota?

Batez besteko aritmetiko sinplea erabiltzen badugu, batez beste atera duen nota honela kalkulatzen dugu:

$$\bar{x} = \frac{6+7}{2} = 6{,}5$$

Hau da, batez besteko nota 6,5ekoa dela esaten dugu. Dena dela, horrela eginez, garrantzi bera ematen zaio teoriari eta praktikari (%50ekoa, alegia), eta hori ez da irakasleak ikasturtearen hasieran esandakoa. Hala, egokiagoa da batez besteko aritmetiko haztatua kalkulatzea, zati bakoitzean atera den nota irakasleak zati bakoitzari eman dion garrantziaren arabera haztatuz:

$$\bar{x}_h = \frac{(6x40) + (7x60)}{40+60} = 6{,}7$$

Ikusten den bezala, batez besteko aritmetiko haztatua erabiliz, ikasleak nota altuxeagoa lortzen du, parte teorikoan nota hobea lortu duelako, eta irakasleak zati horri garrantzi gehiago ematen diolako.

Adibidea. Batez besteko aritmetiko haztatua kalkulatzea: eskualde bateko bi udalerrien langabezia-tasak %8 eta %12 badira, ezin da, besterik gabe, batez besteko aritmetiko sinplea kalkulatu, eta langabezia-tasa %10 dela esan, bi udalerriek biztanleria aktibo desberdinak dituztelako. Batez bestekoa kalkulatzeko, biztanleria aktiboaren arabera haztatu behar da. Biztanleria aktiboak bi herriotan 2.000 eta 5.000 badira, batez besteko langabezia-tasa ezagutzeko, batez besteko aritmetiko haztatua kalkulatuko dugu:

$$\bar{x}_h = \frac{(8x2000) + (12x5000)}{2000+7000} = 10{,}85$$

Beraz, batez besteko langabezia-tasa %10,85 dela esango dugu. Batez besteko aritmetiko sinplearekin aterako duguna, handixeagoa.

2.3.1.3. Moztutako batez bestekoa

Moztutako batez bestekoa kalkulatzeko, datu multzoaren goitik eta behetik (bi muturretatik) datu kopuru zehatza kentzen da (ehunekoetan adierazita), eta gelditzen diren datuekin batez besteko aritmetiko sinplea kalkulatzen. Adibidez, %20an moztutako batez bestekoa kalkulatzeko, balio txikienen %10 eta handienen %10 baztertzen ditugu, eta, ostean, batez bestekoa kalkulatzen.

Moztutako batez bestekoaren helburua muturreko datuen eragina murriztea da, datu multzoaren zentroan gertatzen dena hobeto islatzeko asmoz. Beraz, neurri jasankorra da.
Neurri horren desabantaila da bere kalkuluan datu batzuk ezabatzen direla, eta, horren ondorioz, garrantzitsua izan daitekeen informazioa galtzen dela. Horren ondorioz, gerta daiteke batez besteko horrek ondo ez azaltzea datu multzoaren erdigunean gertatzen dena.
%5etik %25era moztutako batez bestekoak ohikoak dira. Datu multzoa gehiago mozten bada, batez bestekoak adierazgarritasuna galtzen du, datu gehiegi baztertzen direlako. Gutxiago mozten bada, baliteke moztutako batez bestekoarekin lortu nahi den helburua ez lortzea, muturreko datuak barnean gelditu direlako.

Adibidea. Moztutako batez bestekoa: 20 gazteren pisuak honako hauek dira: 50, 64, 65, 66, 64, 66, 66, 66, 65, 66, 65, 66, 66, 66, 65, 66, 65, 66, 66 eta 67. Kontuan izan gazte ia denek 64 kilotik eta 66 kilora arteko pisuak dituztela, muturretan dauden bi pertsonek izan ezik.
%10ean moztutako batez bestekoa kalkulatzeko, goitik eta behetik kendu behar ditugun datu kopurua jakin behar dugu. 20ren %10$\rightarrow \frac{20.10}{100} = 2$ da, horrek esan nahi du datu multzoaren goiko partetik bat kendu behar dugula, eta behetik, beste bat. Beraz, batez besteko pisua, hobeto esanda, moztutako batez bestekoa, erdian dauden 18 gazteen pisuekin kalkulatuko dugu.

$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{N}$	$\bar{x}$ = 65,5

20 gazteen pisuak kontuan hartu izan bagenitu, batez besteko pisua honako hau litzateke:

$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{N}$	$\bar{x}$ = 64,8

Ikusten den bezala, datu guztiak kontuan hartuta, batez besteko txikiagoa lortzen dugu, beheko muturrean dagoen pisuak asko eragiten duelako emaitzan. Kasu horretan, beraz, moztutako batez bestekoak hobeto islatzen du datu multzo horren erdigunean gertatzen dena.

2.3.1.4. Mediana

Aldagai baten balioak txikienetik handienera ordenaturik ditugunean, mediana da erdian dagoen balioa; alegia, medianak datu multzoa erdibanatzen du, alde bakoitzean utziaz datuen %50.
Mediana Me hizkien bidez adierazten da, eta kalkulatzeko balioak txikienetik handienera ordenatu behar dira. Beraz, aldagai kuantitatiboetarako eta ordinaletarako kalkula daiteke (azken kasu horretan, maiztasun metatuek badute zentzua).
Kontuan izan mediana bat datorrela geroago aztertuko ditugun bigarren kuartilarekin (Q_2), bosgarren debilarekin (D_5) eta 50. pertzentilarekin (P_{50}).

Kalkulua balio isolatuetarako

Datu multzoa bakoitia bada, balioak ordenaturik daudela, mediana erdiko datuak hartzen duen balioa da.

Adibidea. Mediana kalkulatzea balio isolatuetarako, datu multzoa bakoitia denean: demagun balio hauek ditugula: 1, 2, 2, 3, 5, 6 eta 8. Kasu horretan, Me = 3 da, balio horrek alde banatara uzten duelako datuen %50.

Datu multzo bakoitiak ditugunean, mediana formula honen bidez ere kalkulatu daiteke:

$$k = \frac{N+1}{2}$$

K zenbakiak datu multzoan medianak duen posizioa adierazten digu, eta N datu kopurua da. Aurreko adibidean, mediana dagoen posizioa modu honetan kalkulatzen dugu:

$$k = \frac{7+1}{2} = 4$$

Hala, mediana 4. posizioan dagoen 3 balioari dagokio.
Datu kopurua bikoitia bada, datuak ordenaturik daudela, erdiko datu bi ditugunez, mediana erdiko bi datu horien batez bestekoa da.
Adibidea. Medianaren kalkulua balio isolatuetarako, datu multzoa bikoitia denean: demagun balio hauek ditugula: 2, 3, 4, 5, 7, 7, 8 eta 10. Mediana modu honetan kalkulatzen dugu:

$$me = \frac{5+7}{2} = 6$$

Beraz, kasu honetan, mediana 6 da.

Nola kalkulatu aldagai diskretuetarako

Datu kuantitatiboak txikienetik handienera ordenatuta daudela, maiztasun metatu erlatiboak kalkulatzen ditugu. Maiztasun-banaketa bi zatitan (datuen %50 alde batera eta %50 beste aldera) banatzen duen balioa da mediana; hau da, %50rari dagokion balioa.

Adibidea. Mediana kalkulatzea aldagai diskretuetarako: demagun lehen hezkuntzako eskola bateko 200 umeren adina ezagutzen dugula.

30. taula: Lehen hezkuntzako eskola txiki bateko umeen adina

x_i	f_i	% metatuak
5	40	20
6	20	30
7	30	45
8	30	60
9	30	75
10	50	100
N	200	

Iturria: lanketa propioa

%50rari gehien hurbiltzen zaion ehuneko metatuari (kasu honetan, %60ari) dagokion balioa da mediana; hau da, 8 balioari. Beraz, eskola horretako umeen %50ek 8 urte edo gutxiago ditu.

Nola kalkulatu datuak tartetan bildurik daudenean

Datuak tartetan bildurik daudela, lehenengo eta behin, mediana kokatzen den tartea zehaztu behar dugu, eta, ostean, formula hau aplikatu:

$$Me = L_i + \frac{\frac{N}{2} - F_{i-1}}{f_i} \cdot a_i$$

L_i: mediana dagoen tartearen beheko muga da.
N: maiztasun absolutu guztien batura da.
F_{i-1}: mediana dagoen tartearen aurreko tartearen maiztasun metatua.
a_i: tartearen zabalera da.
F_i: mediana dagoen tartearen maiztasuna

Kontuan izan koantilen kalkuluan erabiltzen den formula bera dela. Arestian esan bezala, kontuan izan mediana bat datorrela 2. kuartilarekin (Q_2), 5. debilarekin (D_5) eta 50. pertzentilarekin (P_{50}).

Adibidea. Mediana kalkulatzea aldagai jarraituetarako: familia bateko kideen adina.

31. taula: familia bateko kideen adina			
Adin tarteak x_i	f_i	F_i	% metatuak
0-20	1	1	1,3
20-40	5	6	7,9
40-60	26	32	42,1
60-80	36	68	89,7 (mediana tarte honetan dago)
80-100	8	76	100
	76		
Iturria: lanketa propioa			

Maiztasun-taula horren kasuan, honako hauek dira formula erabiltzeko behar ditugun balioak:

$L_i = 60$
$N = 76$
$F_{i-1} = 32$
$a_i = 20$
$f_i = 36$

$$Me = L_i + \frac{\frac{N}{2} - F_{i-1}}{f_i} \cdot a_i \qquad Me = 60 + \frac{\frac{76}{2} - 32}{36}.20 = 63{,}3$$

Hala, familia horretako kideen %50ek 63 urte edo gutxiago ditu, eta %50ek, 63 urte baino gehiago.

Adibidea. Mediana kalkulatzea aldagai jarraituetarako: herri txiki bateko biztanleen adina.

32. taula: herri txiki bateko biztanleen adina			
Biztanleen adinak (x_i)	Biztanle kopurua adin tarte bakoitzean(f_i)	Maiztasun metatuak (F_i)	% metatuak
0-20	9	9	14
20-40	18	27	42,2
40-60	26	53	82,8 (mediana tarte honetan dago)
60-80	7	60	93,7
80-100	4	64	100
N		64	
Iturria: lanketa propioa			

Mediana ezagutzeko, honako hauek dira formula aplikatzeko behar ditugun balioak:

$L_i = 40$
$N = 64$
$F_{i-1} = 27$
$a_i = 20$
$f_i = 26$

$$Me = L_i + \frac{\frac{N}{2} - F_{i-1}}{f_i} \cdot a_i \qquad Me = 40 + \frac{\frac{64}{2} - 27}{26}.20 = 43{,}8 \approx 44$$

Herri horretako biztanleen %50ek 44 urte edo gutxiago ditu, eta %50ek, 44 urte baino gehiago.

Medianaren abantailak eta desabantailak

- Medianaren abantaila neurri aproposa da datu multzoan muturreko balioak ditugunean. Beraz, estatistiko jasankorra edo sendoa da. Ildo horretatik, kontuan izan behar da maiztasun-banaketa asimetrikoetan, hau da, alborapen handiko banaketetan, batez bestekoa baino neurri egokiagoa dela, maiztasun-banaketaren zentroan gertatzen dena adierazteko. Adibidez, demagun sei ikasleri poltsikoan duten diruari buruz galdetu diegula, ikasgelara batez beste zenbat diru ekartzen duten jakiteko: 200, 10, 10, 11, 15 eta 12 euroko erantzunak lortzen ditugu. Batez beste, 43 euro dituzte sei ikasle horiek. Dena dela, mediana 10,75 euro da ($me = \frac{10+11}{2} = 10{,}75$). Mediana, kasu honetan, datu multzoaren erdigunean gertatzen dena adierazteko hobea da. Alegia, medianak hobeto islatzen du datu multzo horren zentroan gertatzen dena. Izan ere, kopuruak ikusita, errealitatea desitxuratzen dugu esaten badugu ikasleak batez beste 43 eurorekin etortzen direla ikasgelara.
- Medianaren desabantailak hauek dira: eskuz kalkulatzea ez dela erraza, eta bere kalkulurako ez direla datu guztiak kontuan hartzen.

2.3.1.5. Moda

Moda Mo hizkien bidez adierazten da, eta datu multzo batean gehien agertzen den balioa edo kategoria da; alegia, maiztasun absolutu handiena duena. Kategoria edo balioa esaten dugu, zeren moda aldagai kuantitatibo zein kualitatiboetarako kalkula daiteke.
Aldagaia kualitatiboa bada edo kuantitatiboa diskretua, moda maiztasun handiena duen kategoria edo balioa da, eta maiztasun-taulari begiratzea nahikoa da hura ezagutzeko.
Aldagaia jarraitua denean, gutxi gorabeherako balio bat lor dezakegu, maiztasun absolutuen maiztasun-banaketatik abiatuta.
Histogrametan eta barra-diagrametan, moda laukizuzen altuenari dagokion tartea edo balioa da, eta maiztasun bakunen poligonoetan punturik gorenari dagokion balioari.
Gerta daiteke datu multzo batean modarik ez egotea edo moda bat baino gehiago egotea. Ildo horretatik, moda bakarreko maiztasun-banaketak, moda bikoak eta moda askoko maiztasun-banaketak aurkitu ditzakegu.

Nola kalkulatu balio isolatuetarako

Balio isolatuak ditugunean, moda gehien errepikatzen den balioa da.

Adibidea. Moda kalkulatzea balio isolatuetarako: 2-2-2-3-3-3-3-3-4-4-5-5 eta 6 datu multzoa dugularik, Mo = 3 da, gehien errepikatzen den balioa delako.

Datu multzoan baliorik errepikatzen ez denean (1-3-4-6-7 datu multzoan, adibidez), esaten da moda ez dagoela definituta edo modarik ez dagoela. Arestian esan bezala, moda bat baino gehiago izatea ere gerta daiteke: 1-3-3-4-4-5-6 datu multzoan, adibidez, bi moda daude: 3 eta 4.

Nola kalkulatu maiztasun-tauletan banatuta dauden aldagai diskretuetarako

Aldagai diskretu bati dagozkion datuak maiztasun-taula batean bilduta daudenean, maiztasunei begiratzea nahikoa da moda zein den jakiteko.

Adibidea. Modaren kalkulatzea maiztasun-tauletan banatuta dauden aldagai diskretuetarako. Ondoren aurkezten dugun maiztasun-taulan, argi ikusten da moda 10 urte dela, gehien errepikatzen den adina delako.

33. taula: ludoteka bateko umeen adina	
umeen adinak	f_i
4	5
5	4
10	16
7	7
3	3
N	35
Iturria: lanketa propioa	

Nola kalkulatu datuak tartetan bildurik daudenean

Datu multzoko balioak tartetan banatuta daudenean, moda dagoen tarteari klase modala deitzen zaio.
Moda ezagutzean, bi egoera izan ditzakegu: maiztasun-taulako balioen tarte guztiak berdinak edo desberdinak izatea.

Tarte guztiak berdinak direnean, moda kalkulatzeko, honako formula hau erabil dezakegu:

$$Mo = L_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \cdot a_i$$

L_i: moda dagoen tartearen beheko muga da.
f_i: moda dagoen tartearen maiztasun absolutua da.
f_{i-1}: moda dagoen tartearen aurreko tartearen maiztasun absolutua da.
f_{i+1}: moda dagoen tartearen hurrengo tartearen maiztasun absolutua da.
a_i: tarteen zabalera da.

Adibidea. Moda kalkulatzea aldagaia tartetan bilduta dagoenean, eta tarte guztiak berdinak direnean: nagusien egoitza bateko kideen adinaren modaren kalkulua.

34. taula: nagusien egoitza bateko kideen adina	
Adin-tarteak	f_i
60 – 63	5
63 – 66	18
66 – 69	42 (moda tarte honetan dago)
69 – 72	27
72 – 75	8
N	100
Iturria: lanketa propioa	

Formula aplikatuta, erraz lortzen dugu moda:

$$Mo = L_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \cdot a_i \qquad Mo = 66 + \frac{(42-18)}{(42-18)+(42-27)} x3 = 67{,}84$$

Adibidea. Modaren kalkulua tarte-zabalera berdinak ditugunean: herri bateko 59 biztanleren adinaren modaren kalkulua.

35. taula: herritarren adina	
Biztanleen adinak	f_i
0-20	7
20-40	18
40-60	26
60-80	7
80-100	1
N	59
Iturria: lanketa propioa	

$$Mo = L_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \cdot a_i \qquad Mo = 45{,}92$$

Aztertzen ari garen maiztasun-taulan tarteen zabalerak desberdinak direnean, lehenengo eta behin, tarte bakoitzerako altuerak (h_i) kalkulatu behar ditugu:

$$h_i = \frac{f_i}{a_i}$$

f_i: tartearen maiztasuna
a_i: tartearen zabalera
Moda altuera handiena duen tartean egongo da, eta, tarte hori oinarritzat hartuta, formula hau aplikatuko dugu:

$$Mo = L_i + \frac{h_i - h_{i-1}}{(h_i - h_{i-1}) + (h_i - h_{i+1})} \cdot a_i$$

L_i: moda dagoen tartearen beheko muga da.
h_i: moda dagoen tarteari dagokion altuera da.
h_{i-1}: moda dagoen tartearen aurreko tarteari dagokion altuera da.
h_{i+1}: moda dagoen tartearen hurrengo tarteari dagokion altuera da.
ai: moda dagoen tartearen zabalera da.

Adibidea. Moda kalkulatzea tarte-zabalera desberdinak ditugunean: herri bateko ludotekako umeen adina.

36. taula: herri bateko ludotekan dauden umeen adina		
	f_i	h_i
0 – 5	15	3
5 – 7	20	10
7 – 9	12	6
9 – 10	3	3
	50	
Iturria: lanketa propioa		

Formula aplikatuta, erraz kalkulatzen dugu moda:

$$Mo = L_i + \frac{h_i - h_{i-1}}{(h_i - h_{i-1}) + (h_i - h_{i+1})} \cdot a_i \qquad MO = 5 + \frac{10-3}{(10-3)+(10-6)}.2 = 6{,}27$$

Modaren abantailak eta desabantailak

- Aldagai jarraituen kasuan kenduta, erraz ezagutzen edota kalkulatzen da. Horregatik, hain zuzen ere, askotan erabiltzen da.
- Desabantailak: modarik ez egotea gerta daitekeela edo moda bat baino gehiago egotea.
- Beste desabantaila bat: kalkulurako datu bakar bat erabiltzen denez, informazio asko galtzen da.

2.3.1.6. Zentro-neurriei buruzko ondorio orokorrak

- Batez bestekoak, medianak eta modak datu multzoaren erdigunean gertatzen dena adierazten digute, balio bakar baten bidez.
- Zentro-neurriak kalkulatzerakoan, kontuan hartu behar da aldagaia zein eskalaren arabera neurtua dagoen. Izan ere, estatistiko guztiak ezin dira erabili aldagai mota guztiekin: eskala nominalean neurtuta dauden aldagaietan moda bakarrik kalkula daiteke. Aldagai ordinaletan, moda eta mediana kalkula daitezke. Aldagai kuantitatiboetan (diskretu zein jarraituetan), mediana, moda eta batez bestekoa ezagut ditzakegu.
- Maiztasun-banaketa bat simetrikoa dela esaten da, batez bestekotik pasatzen den zuzenak histograma bi zati simetrikoetan banatzen duenean. Ildo horretatik, zentralizazio-neurrien bidez, maiztasun-banaketa baten simetriari buruzko lehen ezagutza lor dezakegu. Alborapenik edo muturreko daturik ez badago, hau da, maiztasun-banaketa simetrikoa bada, moda, mediana eta batez besteko aritmetiko sinplea berdinak edo oso antzerakoak izango dira. Kasu horretan, batez bestekoa da zentralizazio neurri egokiena. Aldiz, alborapen handiko banaketetan (edo muturreko datuak edo ez-ohikoak ditugunean), hiru estatistiko horiek desberdinak izango dira. Horrelako kasuetan, mediana da estatistikorik egokiena, neurri jasankorra delako. Beste berba batzuekin esanda, datu multzoaren erdigunean gertatzen dena adierazteko, mediana erabiliko dugu, muturreko balioek edo ohikoak ez direnek batez bestekoa distortsionatzen dutela ikusten badugu. Adibidez, irakaslea ebaluatzerakoan, gelako ikasle talde txiki batek puntuazio oso altuak edo oso baxuak eman ditzake, eta gela osoko ikasleek irakaslearekiko duten irudia distortsionatu. Kasu horretan, hobe da mediana erabiltzea. Edonola ere, beti eman behar da batez bestekoa.
- Aurrekoarekin lotuta, aldakortasun, heterogeneotasun edo dispertsio handia duten datu multzoetan ere, hiru zentro-neurriak desberdinak izango dira. Demagun zazpi balioz osatutako datu multzoa dugula: 1, 2, 2, 3, 4, 7 eta 16. Aldakortasuna nabaria denez, hiru zentro-neurriak oso desberdinak izango dira.

Batez besteko aritmetiko sinplea	$\overline{x} = \frac{1+2+2+3+4+7+16}{7}$	5
Mediana	1, 2, 2, 3, 4, 7, 16	3
Moda	1, 2, 2, 3, 4, 7, 16	2

Horrelako kasuetan, datu multzoaren erdigunean gertatzen dena hobekien islatzen duen zentralizazio-neurria zein den ikusi beharko dugu. Aurreko adibidean, medianak islatzen du hobekien erdigunean gertatzen dena. Batez bestekoa eta moda ez dira zentro-neurri egokiak, hain zuzen ere, datu multzoan aldakortasun handia dagoelako. Muturreko balioa kenduko bagenu (16 balioa), batez bestekoa adierazle ona litzateke ($\bar{x} = 3{,}1$).

- Asimetria gutxiko eta moda bakarreko maiztasun-banaketetan erlazio hau betetzen da: $\bar{x}$ – mo = 3 . ($\bar{x}$ – me).

Adibidea. Batez bestekoa, mediana eta moda kalkulatzea IBM SPSS programaren laguntzarekin.

37. taula: gailu elektronikoen kopurua etxean

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1,00	105	18,2	18,3	18,3
	2,00	78	13,5	13,6	31,8
	3,00	192	33,3	33,4	65,2
	4,00	55	9,5	9,6	74,8
	5,00	41	7,1	7,1	81,9
	6,00	70	12,1	12,2	94,1
	7,00	24	4,2	4,2	98,3
	8,00	10	1,7	1,7	100,0
	Total	575	99,7	100,0	
Perdidos	Sistema	2	,3		
Total		577	100,0		

Iturria: lanketa propioa

Maiztasun-taula horretatik abiatuta, IBM SPSS programak erraz kalkulatzen ditu batez bestekoa, mediana eta moda.

Estatistikoak
Gailu elektronikoen kopurua etxean

N	Válidos	575
	Perdidos	2
Media		3,3565
Mediana		3,0000
Moda		3,00

Datu horien arabera, laginean aukeratu diren 575 pertsonek batez beste 3 bat gailu elektroniko dituzte etxean; gehien errepikatzen den aparatu kopurua 3 da (moda), eta mediana ere 3 da; hau da, 575 pertsona horien erdiak (%50ak) 3 aparatu edo gutxiago dituzte eta beste erdiak 3 aparatu baino gehiago.

Kasu honetan, hiru zentralizazio-neurri horiek oso ondo islatzen dute datu multzoaren erdigunean gertatzen dena. Alborapenik edo muturreko daturik ez dugunez (inor ez dago 89 aparaturekin) maiztasun-banaketa nahiko simetrikoa da, eta, horren ondorioz, moda, mediana eta batez besteko aritmetiko sinplea antzerakoak dira. Kasu honetan, batez bestekoa zentralizazio-neurri egokiena da datu multzoaren erdigunean gertatzen dena adierazteko. Asimetria gutxiko eta moda bakarreko maiztasun banaketa denez, arestian aipatutako honako erlazio hau betetzen da: $\bar{x}$ – mo = 3 . ($\bar{x}$ – me) → 3,3565 – 3 ≈ 3.(3,3565 – 3).

2.3.2. Sakabanatze-neurriak

Batez bestekoa, mediana eta moda, hau da, zentro-neurriak ez dira nahikoak datu multzoaren berezitasunak agertzeko. Izan ere, zentro-neurriek ez digute ezer esaten datu multzoaren barreiadura, sakabanatze, heterogeneotasun edo aldakortasunari buruz. Alegia, bi datu multzok antzerako zentro-neurriak (batez bestekoa edo mediana) izan ditzakete, baina errealitatean oso desberdinak izan, sakabanatze-maila desberdina dutelako.
Hala, datu multzo baten sakabanatzea neurtzeko, hau da, datuak elkarren antzerakoak edo oso desberdinak diren jakiteko, sakabanatze estatistikoak edo sakabanatze-neurriak kalkulatu behar dira. Neurri horiek zenbat eta handiagoak izan, datu multzoan orduan eta aldakortasun, barreiadura edo sakabanatze handiago dagoela ondorioztatuko dugu; hau da, datuak are eta gehiago aldentzen direla zentrotik (batez bestekotik edota medianatik). Datu multzoetan sakabanatzerik, aldakortasunik edo heterogeneotasunik ez badago, berriz, sakabanatze-neurriak zero izango dira edo zerotik hurbil egongo dira.
Horrez gain, sakabanatze-neurriak datu multzo desberdinak alderatzeko ere erabil ditzakegu. Horretarako, zentro-neurrietan oinarritu behar dugu. Adibidez, lau pertsonako bi talde ditugu: A eta B. A taldea osatzen duten umeen adina = 1 – 2 – 8 – 9 da; eta B taldea osatzen dutenena = 5 – 5 – 4 - 6. Bi multzoen batez besteko adina 5 da. Dena dela, A datu multzoa sakabanatuagoa edo heterogeneoagoa dela esaten dugu, talde horretako datuak B multzokoak baino gehiago aldentzen direlako batez bestekotik.
Sakabanatze-neurriek hirugarren erabilera ere badute: zentro-neurriak zenbateraino diren datu multzo baten adierazgarri esaten digute. Aurreko adibidean, ateratzen den batez bestekoa ($\bar{x}$ = 5) adierazgarriagoa da B taldean A taldean baino.
Horrek guztiak argi erakusten du maiztasun-banaketa baten ezaugarriak behar bezala adierazteko, zentro-neurriekin batera sakabanatze-neurriak kalkulatu behar ditugula.
Bi eratako sakabanatze-neurriak daude: sakabanatze-neurri absolutuak eta sakabanatze-neurri erlatiboak.

2.3.2.1. Sakabanatze-neurri absolutuak

Sakabanatze-neurri absolutuek datu multzo baten aldakortasun, sakabanatze edota heterogeneotasuna neurtzen dute, aldagaia neurtzeko erabili dugun unitate beretan. Horiek kalkulatzean, zentro-neurriek ez dute parte hartzen, eta, beraz, ezin dira erabili datu multzo desberdinen sakabanatzeak alderatzeko. Adibidez, familien seme-alaba kopurua aztertzen ari garenean, ez dira berdinak desbideratze tipikoa 2 seme-alabakoa izatea, batez bestekoa 3 seme-alabakoa denean edo 6koa denean. Batez bestekoa kontuan hartu gabe, ezin ditugu datu multzo desberdin bi horiek alderatu.
Sakabanatze-neurri absolutuak kasu bakar batean erabil daitezke bi datu multzo desberdinen sakabanatzeak alderatzeko: datu multzoek batez besteko bera eta unitate beretan neurtuta daudenean. Beste kasu guztietan, sakabanatze-neurri erlatiboak erabili behar dira.
Sakabanatze-neurri absolutuak hainbat dira; guk hemen gehien erabiltzen direnak aurkeztuko ditugu: Ibiltartea, kuartil arteko ibiltartea (KAI), kuartilen desbideratzea, batez besteko desbideratze absolutua, bariantza eta desbideratze estandarra.

2.3.2.1.1. Ibiltartea

Ibiltartea aldagai baten balio handienaren eta txikienaren arteko diferentzia da. Era honetan kalkulatzen da:

$$I = x_{max} - x_{min}$$

Ibiltartea zenbat eta handiagoa izan, sakabanatzea handiagoa izango da, eta zero da sakabanatzerik, aldakortasunik edo heterogeneotasunik ez dagoenean (hain zuzen ere, balio guztiak berdinak direnean). Dena dela, ezin da finkatu balio zehatz bat sakabanatze handia erakusten duenik. Bestalde, kontuan izan ibiltartea desbideratze estandarra baino bi aldiz handiagoa dela beti. Aipatzekoa ere bada, ibiltartea beste prozedura estatistiko batzuetan erabiltzen dela. Besteak beste, aldagai jarraituen balioak tartetan banatzeko prozeduran ibiltartea kalkulatu behar da.

Adibidea. Ibiltartearen kalkulua IBM SPSS programaren laguntzarekin: gailu elektronikoen kopurua etxean

38. taula: gailu elektronikoen kopurua etxean

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1,00	105	18,2	18,3	18,3
	2,00	78	13,5	13,6	31,8
	3,00	192	33,3	33,4	65,2
	4,00	55	9,5	9,6	74,8
	5,00	41	7,1	7,1	81,9
	6,00	70	12,1	12,2	94,1
	7,00	24	4,2	4,2	98,3
	8,00	10	1,7	1,7	100,0
	Total	575	99,7	100,0	
Perdidos	Sistema	2	,3		
Total		577	100,0		

Iturria: lanketa propioa

Maiztasun-taula honetan, argi ikusten dugu balio txikiena 1 dela eta handiena, 8; ibiltartea (rango hitza erabiltzen du IBM SPSSak), beraz, 7 da, eta adibide honetan esan nahi du gure laginean gailu elektroniko gutxien eta gehien dutenen artean 7 gailuren aldea dagoela.

Estadísticos
Gailu elektronikoen kopurua etxean

N	Válidos	575
	Perdidos	2
Media		3,3565
Mediana		3,0000
Moda		3,00
Rango		7,00
Mínimo		1,00
Máximo		8,00

Ibiltartearen abantailak eta desabantailak

- Ibiltartearen abantailak: oso erraz kalkulatzen da, eta aldagaiaren neurri beran adierazten da. Azken abantaila horrek esan nahi du erraz interpretatzen dela.
- Desabantaila: hura kalkulatzeko datu gutxi erabiltzen direla; alegia, bakarrik muturreko datuak erabiltzen dira. Beraz, ez da gomendagarria lehenengo balioa eta azkenengoa arraroak edo bitxiak direnean, ezta datu gutxi ditugunean ere. Datu kopurua handia dugunean eta balio arrarorik ez dugunean, ordea, erabilgarria da sakabanatzea neurtzeko.

- Esan bezala, ezin da finkatu balio bat sakabanatze handia erakusten duenik.
- Tarteko datuak kontuan hartzen ez direnez (bakarrik muturretakoak), informazio gutxi ematen digu sakabanatzeari buruz; izan ere, ezin dugu jakin muturreko balio bi horien artean datuak uniformeki banatzen diren ala ez.
- Muturreko datuek ibiltartean duten eragina saihesteko, kuartil arteko ibiltartea erabil daiteke.

2.3.2.1.2. Kuartil arteko ibiltartea edo ibiltarte kuartilikoa

Balio txikiena eta handiena beste balioengandik oso urrun daudenean (balio bitxiak edo arraroak direnean), ibiltarte laburragoa kalkula dezakegu: kuartil arteko ibiltartea (KAI). Sakabanatze-neurri horrek datu multzoaren erdian kokatzen diren datuen %50ek hartzen duen tartearen zabalera adierazten digu; hau da, kuartil arteko ibiltartean, datuen %50 dago. Sakabanatze-neurri hori era honetan kalkulatzen da :

$$KAI = Q_3 - Q_1$$

Q_1 eta Q_3 lehenengo eta hirugarren kuartilak dira. Hau da, datu multzoa lau zatitan banatu ostean, datuen %75 azpitik uzten duen balioari (hirugarren zatiari dagokion balioari) datuen %25 azpitik uzten duen balioa (lehen zatiari dagokion balioa) kentzen diogu.
Kuartil arteko ibiltartea zenbat eta handiagoa izan, sakabanatzea handiagoa izango da, eta zero da sakabanatzerik, aldakortasunik edo heterogeneotasunik ez dagoenean (Ibiltartean bezala, balio guztiak berdinak direnean). Dena dela, ezin da finkatu balio bat sakabanatze handia erakusten duenik.
Adibidea. Kuartil arteko ibilitartearen interpretazioa: Unibertsitateko fakultate batean, Q_1=20 urte da (ikasleen %25ak 20 urte edo gutxiago ditu) eta Q_3=24 (ikasleen %75ak 24 urte edo gutxiago ditu). Beraz,

$$KAI = 24 - 20 = 4$$

Lau urteko kuartil arteko ibiltarte horretan, 20 urtetik 24ra, fakultate horretako ikasleen %50 dago. Beraz, aldakortasun, sakabanatze edo heterogeneotasun txikia dago. Fakultate horretako ikasle guztiek adin bera izango balute, KAI = 0 litzateke.

Kuartil arteko ibiltartearen abantailak eta desabantailak

- Ibiltarteak baino modu zehatzagoan neurtzen du datu multzoaren sakanabatzea.
- Sakabanatze-neurri egokiagoa da maiztasun-banaketan muturreko datuak ditugunean. Hau da, egokia da maiztasun-banaketa asimetrikoak ditugunean edo zentro-neurri gisa mediana erabili dugunean.
- Kaxa-diagramak egiteko erabiltzen da.
- Desabantaila: hura kalkulatzeko datu gutxi erabiltzen direla; alegia, bakarrik datuen %50 erabiltzen dira; datu multzoaren erdian daudenak.
- Beste desabantaila bat : ezin dela finkatu balio bat sakabanatze handia duenik.

Adibidea. Kuartil arteko ibiltartea kalkulatzea IBM SPSS programaren laguntzarekin: kultura-elkarte bateko bazkideen adina. 61 pertsonak osatzen duten kultura-elkarte batean, bazkideen adina galdetu dugu, eta honelako maiztasun-taula lortu:

39. taula: kultura-elkarte batetako bazkideen adina

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	2,00	1	1,6	1,6	1,6
	24,00	12	19,7	19,7	21,3
	25,00	12	19,7	19,7	41,0
	26,00	15	24,6	24,6	65,6
	27,00	4	6,6	6,6	72,1
	28,00	12	19,7	19,7	91,8
	29,00	4	6,6	6,6	98,4
	99,00	1	1,6	1,6	100,0
	Total	61	100,0	100,0	

Iturria: lanketa propioa

Datu multzo horretatik abiatuta, eta IBM SPSS programaren laguntzarekin, honako neurri estatistiko hauek lortzen ditugu:

Estadísticos
Kultura-elkarte batetako kideen adina

N	Válidos	61
	Perdidos	0
Media		26,8689
Mediana		26,0000
Moda		26,00
Rango		97,00
Mínimo		2,00
Máximo		99,00
Percentiles	25	25,0000
	50	26,0000
	75	28,0000

Ibiltartea, kasu honetan, 97 da. Batez bestekoa 27 izanda, badirudi sakabanatze handia dagoela, eta errealitatean ez da horrela. Arazoa da goitik eta behetik ibiltartea desitxuratzen duten bi balio arraro ditugula (2 eta 99). Kasu honetan, beraz, sakabanatzea ezagutzeko, hobe da kuartil arteko ibiltartea kalkulatzea. Arestian adierazi bezala, sakabanatze-neurri hori egokiagoa da balio txikiena eta handiena beste balioetatik oso urrun daudenean (balio bitxiak edo arraroak direnean, alegia).

Kasu honetan, KAI = $Q_3 - Q_1 = 28 - 25 = 3$ da. Alegia, elkarte horretako kideen %50ak 25 urtetik 28 urtera dituzte, hau da, 3 urteko tartean daude. Emaitza horrek argi erakusten du sakabanatzea ez dela hain handia. Muturreko balioak kenduta, sakabanatzea ez da hain handia.

2.3.2.1.3. Kuartilen desbideratzea edo ibiltarte semiinterkuartilikoa

Kuartil arteko ibiltartearekin harremanetan, kuartilen desbideratzea edo ibiltarte semiinterkuartilikoa dugu. Neurri estatistiko hori balio bitxiak edo arraroak ditugunean ere erabiltzen da. Era honetan kalkulatzen da:

$$\frac{Q_3 - Q_1}{2}$$

Zenbat eta handiagoa izan, sakabanatzea handiagoa izango da, eta zero da sakabanatzerik, aldakortasunik edo heterogeneotasunik ez dagoenean (balio guztiak berdinak direnean). Dena dela, ezin da finkatu balio bat sakabanatze handia erakusten duenik.

Kuartilen desbideratzearen abantailak eta desabantailak

- Ibiltarteak baino modu zehatzagoan neurtzen du datu multzoaren sakabanatzea.
- Neurri egokia da muturreko datuak ditugunean; hain zuzen ere, neurri jasankorra delako. Ildo horretatik, neurri hori egokia da zentro-neurri gisa mediana erabili dugunean edota maiztasun-banaketa asimetrikoak ditugunean.
- Desabantaila: ez du kontuan hartzen datuetan biltzen den informazio guztia; hau da, erdian dauden datuak hartzen ditu bakarrik kontuan.
- Esan bezala, beste desabantaila bat da ezin dela finkatu balio bat sakabanatze handia erakusten duenik.

Adibidea. Ibiltarte semiinterkuartiliko kalkulatzea IBM SPSS programaren laguntzarekin: 50 estaturen errenta per capita.

40. taula: estatuetako errenta per capita

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	250,00	1	2,0	2,0	2,0
	6000,00	19	38,0	38,0	40,0
	6443,00	1	2,0	2,0	42,0
	6453,00	1	2,0	2,0	44,0
	6500,00	1	2,0	2,0	46,0
	6543,00	3	6,0	6,0	52,0
	6544,00	1	2,0	2,0	54,0
	6554,00	2	4,0	4,0	58,0
	6778,00	2	4,0	4,0	62,0
	6788,00	1	2,0	2,0	64,0
	6789,00	2	4,0	4,0	68,0
	6889,00	1	2,0	2,0	70,0
	6999,00	1	2,0	2,0	72,0
	7000,00	8	16,0	16,0	88,0
	7089,00	1	2,0	2,0	90,0
	7321,00	1	2,0	2,0	92,0
	7564,00	1	2,0	2,0	94,0
	7652,00	1	2,0	2,0	96,0
	7654,00	1	2,0	2,0	98,0
	65000,00	1	2,0	2,0	100,0
	Total	50	100,0	100,0	

Iturria: lanketa propioa

Maiztasun-banaketa horretan oinarrituta, honako estatistiko hauek lortzen ditugu:

Estadísticos
estatuetako errenta per capita

N	Válidos	50
	Perdidos	0
Media		7580,3400
Mediana		6543,0000
Moda		6000,00
Rango		64750,00
Percentiles	25	6000,0000
	50	6543,0000
	75	7000,0000

Maiztasun-taula ikusi ostean, argi dago mediana batez bestekoa baino neurri estatistiko hobea dela datu multzoaren erdigunean gertatzen dena deskribatzeko. Izan ere, muturreko datuak ditugu (250 eta 65.000), eta horiek modu artifizialean handitzen dute batez bestekoa, eta datu multzoaren deskribapen okerra egitera bultzatu. Kasu horretan, beraz, hobe da kuartil arteko ibiltartea edo ibiltarte semiinterkuartilikoa erabiltzea. Muturreko datuak ditugunez, ibiltartea neurri desegokia da desbideratzea aztertzeko, ondoren aurkezten ditugun emaitzek erakusten duten bezala:

$$I = x_{max} - x_{min} \qquad 65.000 - 250 = 64750$$

$$I_Q = Q_3 - Q_1 \qquad 7000 - 6000 = 1000$$

$$\frac{Q_3 - Q_1}{2} \qquad \frac{7000 - 6000}{2} = 500$$

2.3.2.1.4. Batez bestekoarekiko batez besteko desbideratzea edo batez besteko desbideratze absolutua

Batez besteko desbideratze absolutua $D_{\bar{x}}$ hizkiaren bidez adierazten da, eta era honetan kalkulatzen:

$$D_{\bar{x}} = \frac{|x_1 - \bar{x}| + x_2 - \bar{x}| + \ldots + |x_n - \bar{x}|}{N}$$

Formula hori modu honetan laburbil dezakegu:

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

Batez besteko desbideratze absolutuaren bidez, datu multzoko balio bakoitzak batez bestekoarekiko duen distantziaren batez bestekoa kalkulatzen da. Horretarako, datu multzoko

balio bakoitzari batez bestekoa kentzen zaio balio absolutuan[10]; ostean, emaitza guztiak batzen dira, eta, azkenik, balio guztien kopuruarekin (N) zatitzen.
Batez besteko desbideratze absolutua zenbat eta handiagoa izan, sakabanatzea handiagoa izango da, eta zero da sakabanatzerik, aldakortasunik edo heterogeneotasunik ez dagoenean (balio guztiak berdinak direnean). Dena dela, ezin da finkatu balio bat sakabanatze handia erakusten duenik.

Adibidea. Batez besteko desbideratze absolutu kalkulatzea: 4, 8 eta 15 datu multzoaren batez besteko aritmetiko sinplea ($\bar{x}$) 9 da, eta hiru balioen batez bestekoarekiko desbideratzeak, balio absolutuan, d = 5 (/4-9/), d = 1 (/8-9/) eta d = 6 (/15-9/) dira. Desbideratze horien batez besteko aritmetiko sinplea sakabanatze-neurri modura erabil dezakegu:

$$D_{\bar{x}} = \frac{5+1+6}{3} = 4$$

Datuak maiztasun-tauletan banatuta dauzkagunean, honako formula hau erabiltzen da:

$$D_{\bar{x}} = \frac{|x_1 - \bar{x}|f_1 + |x_2 - \bar{x}|f_2 + . . + |x_n - \bar{x}|f_n}{N}$$

Formula hori modu honetan laburbil dezakegu:

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

Adibidea. Batez besteko desbideratze absolutu kalkulatzea, datu multzoa maiztasun-taula batean banatuta dugunean: arraun-elkarte bateko kideen adinak.

41. taula: arraun-elkarte bateko kideen adinak					
	markak	f_i	$x_i \cdot f_i$	$\lvert x_i - \bar{x}_i \rvert$	$\lvert x_i - \bar{x}_i \rvert \cdot f_i$
10-15	12.5	3	37,5	9,28	27,85
15-20	17.5	5	87,5	4,28	21,43
20-25	22.5	7	157,5	0,71	4,99
25-30	27.5	4	110	5,71	22,85
30-35	32.5	2	65	10,7	21,42
		21	457,5		98.57
Iturria: lanketa propioa					

Batez besteko desbideratze absolutua erraz kalkulatzen da formula aplikatuta:

$$D_x = \frac{\sum_{i=1}^{n} |x_i - \bar{x}| f_i}{N} \qquad D_{\bar{x}} = \frac{98{,}57}{21} = 4{,}69$$

[10] Zenbaki baten balio absolutua zenbaki bera da, baina inolako zeinurik gabe. Zenbakiaren alde bietan zutabe bertikalak ezarriz adierazten da.

Batez besteko desbideratze absolutuaren abantailak eta desabantailak

- Sakabanatze-neurri estatistiko horren abantailak: erraz kalkulatzen da, eta desbideratze-neurrien artean ulerterrazenetarikoa da. Gainera, hura kalkulatzeko datu guztiak erabiltzen dira. Dena dela, gutxitan erabiltzen da. Gehienbat, bariantza eta desbideratze tipikoa ulertarazteko erabiltzen da.
- Desabantaila: lagin baten batez besteko desbideratze absolutua ez da populazioaren batez besteko desbideratze absolutuaren estimatzaile ona.
- Esan bezala, beste desabantaila bat da ezin dela finkatu balio bat sakabanatze handia erakusten duenik.

2.3.2.1.5. Bariantza

Bariantza, batez bestekoarekiko desbideratze absolutuen karratuen batez bestekoa da, eta haren erro karratu positiboa desbideratze estandarra (desbideratze tipikoa) da. Azken neurri horrek datu bakoitza batez besteko aritmetiko sinpletik zenbat desbideratzen den adierazten digu. Beraz, bariantza zein desbideratze tipikoa batez bestekoarekin harremanetan aztertu behar dira.
Laginekin lanean ari garenean, bariantza s^2 ikurraren bidez adierazten da, eta populazioekin ari garenean, σ^2 (sigma karratura) ikurrarekin. Bariantza beti da balio positiboa, zenbat eta handiagoa izan sakabanatzea handiagoa izango da, eta zero da sakabanatzerik, aldakortasunik edo heterogeneotasunik ez dagoenean (balio guztiak berdinak direnean). Dena dela, ezin da finkatu balio bat sakabanatze handia erakusten duenik. Gainera, desbideratzea unitate karratuetan neurtzen duenez, zaila da hura interpretatzea.

Balio isolatuetarako kalkulatzea

Ikerketarako ditugun balioak x_1, x_2, x_3,…, x_n izanik, bariantza kalkulatzeko, honako formula hau erabil dezakegu:

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{N}$$

Formula hori modu honetan laburbiltzen da:

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

Lehenengo eta behin, batez besteko aritmetiko sinplea kalkulatzen dugu. Batez besteko desbideratze absolutua kalkulatzerakoan egiten dugun bezala, balio bakoitzari batez besteko aritmetiko sinplea kentzen diogu; hau da, balio bakoitzak batez bestekora duen distantzia kalkulatzen dugu. Azkenik, distantzia horien batez besteko koadratikoa kalkulatzen dugu; alegia, distantzia karratuak ateratzen ditugu, guztiak batzen ditugu, eta lortutako emaitza datu kopuruaz zatitzen dugu.
Bariantzaren kalkulua errazteko, aurreko formulen baliokideak diren honako formula hauek ere erabiltzen dira:

$$\sigma^2 = \frac{x_1^2 + x_2^2 + \ldots + x_n^2}{N} - \bar{x}^2$$

Formula hori modu honetan laburbil dezakegu:

$$\sigma^2 = \sum_{i=1}^{n} \frac{x_i^2}{N} - \bar{x}^2$$

Egia esan, azken bi formula horiek gehiagotan erabiltzen dira, kalkulua errazagoa delako.

Adibidea. Bariantza kalkulatzea balio isolatuak ditugunean: bost ikasleren notak. Demagun estatistikako azterketa batean bost ikaslek honako nota hauek atera dituztela: 7,3 – 7 - 7,6 - 7,5 eta 7,9. Batez bestekoa, beraz, $\bar{x}$ =7,47 da. Bariantza kalkulatzeko, balio guztiak batu behar ditugu eta, bestetik, baita X_i^2 guztiak ere:

42. taula: bost ikasleren notak	
x_i	x_i^2
7,3	53,29
7	49
7,6	57,76
7,5	56,25
7,9	62,41
$\sum x_i = 37{,}3$	$\sum x_i^2 = 278{,}71$
Iturria: lanketa propioa	

Formula aplikatzen badugu, erraz kalkulatzen dugu bariantza:

$$\sigma^2 = \sum_{i=1}^{n} \frac{x_i^2}{N} - \bar{x}^2 \qquad \sigma^2 = \frac{278{,}71}{5} - 7{,}46 = 0{,}09$$

Ikusten den bezala, bariantza oso hurbil dago zerotik (0,09), eta horrek esan nahi du talde horretan dispertsio, sakabanatze edo barreiadura txikia dagoela. Egiatan, notak ikusita, talde nahiko homogeneoa da. Ez dago desberdintasun handirik ikaskideen artean.

Adibidea. Bariantza kalkulatzea balio isolatuak ditugunean: bost ikasleren notak.

Bost ikasleko beste talde batek honako nota hauek atera ditu: 9,3 – 5 - 1,6 - 7,5 eta 3,9. Batez bestekoa, beraz, $\bar{x}$ = 5,46 da.

43. taula: bost ikasleren notak	
x_i	x_i^2
9,3	86,49
5	25
1,6	2,56
7,5	56,25
3,9	15,21
27,3	185,51
Nik neuk egina	

Formula aplikatzen badugu, erraz kalkulatzen dugu bariantza:

$$\sigma^2 = \sum_{i=1}^{n} \frac{x_i^2}{N} - \bar{x}^2 \qquad\qquad \sigma^2 = \frac{185{,}51}{5} - 5{,}462 = 7{,}29$$

Ikusten den bezala, bariantza asko urruntzen da zerotik (7,29) eta horrek esan nahi du talde horretan dispertsio, sakabanatze edo barreiadura handia dagoela. Egiatan, notak ikusita, talde nahiko heterogeneo baten aurrean gaude. Desberdintasun handiak daude ikasleen artean.

Bariantza kalkulatzea maiztasun-tauletatarako (aldagai diskretuak eta jarraituak)

Datuak maiztasun-taula batean banatuta daudenean, honako formula hau erabiliko dugu:

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 f_1 + (x_2 - \bar{x})^2 f_2 + \ldots + (x_n - \bar{x})^2 f_n}{N}$$

Formula hori modu honetan laburbil dezakegu:

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

Bariantzaren kalkulua errazteko, aurreko formulen baliokideak diren honako formula hauek ere erabiltzen dira:

$$\sigma^2 = \frac{x_1^2 f_1 + x_2^2 f_2 + \ldots + x_n^2 f_n}{N} - \bar{x}^2$$

Formula hori modu honetan laburbil dezakegu:

$$\sigma^2 = \sum_{i=1}^{n} \frac{x_i^2 f_i}{N} - \bar{x}^2$$

Adibidea. Bariantza kalkulatzea, datuak maiztasun-taulan banatuta ditugunean: 12 ikasleren notak.

44. taula: 12 ikasleren notak				
x_i	f_i	$x_i.f_i$	x_i^2	$x_i^2.f_i$
5	4	20	25	110
6	5	30	36	180
8	3	24	64	192
N	12	74		482
Iturria: lanketa propioa				

Taula horretako datuak erabiliz, formula aplika dezakegu:

$$\sigma^2 = \sum_{i=1}^{n} \frac{x_i^2 f_i}{N} - \bar{x}^2$$

$$\sigma^2 = \frac{482}{12} - 6{,}12 = 2{,}13$$

Bariantza, adibide honetan, 2,13 da. Datu multzoa ikusita, eta, bariantza ez denez, 0tik asko urruntzen, eta, gainera, batez bestekoa baino txikiagoa denez ($\bar{x} = 6{,}16$), dispertsio txikia dagoela onartu behar dugu.

Adibidea. Bariantza kalkulatzea, datuak maiztasun-taulan banatuta dauzkagunean: 10 ikasleren notak.

45. taula: 12 ikasleren notak	
x_i	f_i
2	3
4	2
6	5
N	10
Iturria: lanketa propioa	

Taulako datuak erabiliz, formula aplikatzen dugu:

$$\sigma^2 = \sum_{i=1}^{n} \frac{x_i^2 f_i}{N} - \bar{x}^2$$

$$\sigma^2 = 3{,}04$$

Bigarren adibide honetan, bariantza 3,04 da. Datu multzoa ikusita, eta bariantza 0tik asko urruntzen ez denez, eta, gainera, batez bestekoa baino txikiagoa denez ($\bar{x} = 4{,}4$), dispertsioa ez da oso adierazgarria.

Adibidea. Bariantza kalkulatzea datuak tartetan banatuta daudenean: test batean 42 lagunek ateratako puntuazioa.

Kasu horretan, formula aplikatzeko, tarteen markak hartzen dira kontuan. Aldaketa hori eginda, aurretik aurkeztu ditugun formulak aplikatuko ditugu:

46. taula: test batean ateratako puntuazioa				
	markak	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
10-20	15	1	15	225
20-30	25	8	200	5000
30-40	35	10	350	12250
40-50	45	9	405	18225
50-60	55	8	440	24200
60-70	65	4	260	16900
70-80	75	2	150	11250
		42	1820	88.050
Iturria: lanketa propioa				

$$\sigma^2 = \sum_{i=1}^{n} \frac{x_i^2 f_i}{N} - \bar{x}^2$$

$$\sigma^2 = \left(\frac{88050}{42}\right) - 43{,}33^2 = 218{,}94$$

Datu multzoa ikusita, eta bariantza 0tik asko urruntzen denez, eta, gainera, batez bestekoa baino handiagoa denez ($\bar{x} = 43{,}33$), dispertsioa adierazgarria dela ondorioztatu dezakegu.

Adibidea. Bariantza kalkulatzea IBM SPSS programaren bidez: 50 estaturen errenta per capita.

47. taula: 50 estaturen errenta per capita

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	250,00	1	2,0	2,0	2,0
	6000,00	19	38,0	38,0	40,0
	6443,00	1	2,0	2,0	42,0
	6453,00	1	2,0	2,0	44,0
	6500,00	1	2,0	2,0	46,0
	6543,00	3	6,0	6,0	52,0
	6544,00	1	2,0	2,0	54,0
	6554,00	2	4,0	4,0	58,0
	6778,00	2	4,0	4,0	62,0
	6788,00	1	2,0	2,0	64,0
	6789,00	2	4,0	4,0	68,0
	6889,00	1	2,0	2,0	70,0
	6999,00	1	2,0	2,0	72,0
	7000,00	8	16,0	16,0	88,0
	7089,00	1	2,0	2,0	90,0
	7321,00	1	2,0	2,0	92,0
	7564,00	1	2,0	2,0	94,0
	7652,00	1	2,0	2,0	96,0
	7654,00	1	2,0	2,0	98,0
	65000,00	1	2,0	2,0	100,0
	Total	50	100,0	100,0	

Iturria: lanketa propioa

Maiztasun-taula horretatik abiatuta, IBM SPSS programak erraz ateratzen du bariantza:

Estadísticos
estatuetako errenta per capita

N	Válidos	50
	Perdidos	0
Media		7580,3400
Mediana		6543,0000
Moda		6000,00
Varianza		69700842,882
Rango		64750,00
Mínimo		250,00
Máximo		65000,00

Ikusten denez, bariantza oso handia da. Horren arrazoia da estatu batek oso errenta txikia duela (250) eta beste batek oro errenta altua (65.000). Muturreko datuek eragin handia dute bariantzan, eta, horren ondorioz, bariantza handia ateratzen da. Batez bestekoa baino askoz handiagoa da bariantza; beraz, estatu multzo horretan dispertsio, heterogeneotasun edo sakabanatze handia dago.

Bariantzaren abantailak eta desabantailak

- Bariantzak unitate karratuetan neurtzen du sakabanatzea. Hau da, ez du aldagaiaren unitate berean neurtzen, eta, beraz, interpretazioa zailagoa da; sakabanatzea zenbatekoa den jakiteko, zaila da bariantza batez bestekoarekin harremanetan jartzea, bat unitate karratuetan neurtzen delako eta bestea unitate errealetan. Adibidez, aldagaiak seme-alaba kopurua neurtzen badu, aldagai horren bariantzaren emaitza ber bi eginda ematen zaigu, eta interpretazioa zaildu egiten da.
- Sakabanatze-neurri absolutua den heinean, ez da egokia datu multzo desberdinetako sakabanatze-mailak alderatzeko.
- Datu arraroek edo muturrekoek eragin handia dute bariantzan. Beraz, muturreko datuak ditugunean, hobe da beste neurri jasangarriago batzuk erabiltzea.
- Esan bezala, beste desabantaila bat da ezin dela finkatu balio bat sakabanatze handia erakusten duenik.
- Batez besteko berdina duten maiztasun-banaketa desberdinak baditugu eta beren bariantzak ezagutzen baditugu, guztizko bariantza edo bariantza totala ezagut dezakegu. Lagin guztiek tamaina bera badute, guztizko bariantza edo bariantza totala modu honetan kalkulatzen da:

$$\sigma^2 = \frac{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2}{n}$$

Batez besteko berdina duten maiztasun-banaketak baditugu, baina laginen tamainak desberdinak badira, guztizko bariantza era honetan kalkulatzen da:

$$\sigma^2 = \frac{k_1 \cdot \sigma_1^2 + k_2 \cdot \sigma_2^2 + \ldots + k_n \cdot \sigma_n^2}{k_1 + k_2 + \ldots + k_n}$$

Formula horretan, k maiztasun maiztasun-banaketa bakoitzaren tamaina da

2.3.2.1.6. Desbideratze estandarra edo desbideratze tipikoa

Bariantza batez bestekoarekin harremanetan aztertu behar da. Dena dela, zaila da bariantza batez bestekoarekin harremanetan jartzea, bariantza unitate kuadratikoetan adierazten delako, eta batez bestekoa aldagaia neurtu den unitateetan. Beraz, zentro-neurriak eta desbideratze-neurriak modu errazean harremanetan jarri ahal izateko, desbideratze estandarra edo desbideratze tipikoa erabiltzen da gehiago.

Desbideratze estandarra, desbideratze tipikoa edo sakabanatze absolutua (hiru izenak erabil daitezke) batez besteko aritmetiko sinplerako datuek dituzten distantzien batez besteko koadratikoa da; alegia, desbideratze tipikoak datu multzoko balioek batez bestekoarekiko duten dispertsio maila neurtzen du. Bariantzaren erro karratu positiboa ere bada.

Laginekin lanean ari garenean, desbideratze tipikoa s hizkiaren bidez adierazten da, eta, populazioekin lanean ari garenean, σ (sigma) hizki grekoarekin. Askotan, sigma hitza soilik erabiltzen da desbideratze tipikoa izendatzeko.

Desbideratze tipikoa interpretatzea ez da erraza. Dena dela, kontuan izan behar dugu zenbat eta desbideratze estandarra txikiagoa izan, hau da, zenbat eta zerora gehiago hurbildu, datu

multzoan orduan eta sakabanatze txikiagoa dagoela, eta, beraz, datuak batez bestekoaren inguruan are eta gehiago biltzen direla, edo, hau da, homogeneotasun handiagoa dagoela.
Bestalde, ez ahantzi desbideratze tipikoa beti balio positiboa dela, eta zero dela balio guztiak berdinak direnean; hots, zero da sakabanatzerik, aldakortasunik edo heterogeneotasunik ez dagoenean.
Beste sakabanatze-neurriekin gertatzen den bezala, ezin da finkatu balio bat sakabanatze handia erakusten duenik, eta horrek zaildu egiten du hura interpretatzea.

Adibidea. Desbideratze tipikoaren berezitasunak: sei pertsonako hiru taldetan egunean zehar erretzen dituzten zigarroei buruz galdetu dugu:

- Lehenengo taldea: 0, 0, 1, 1, 14, 14.
- Bigarren taldea: 6, 5, 4, 6, 5, 4.
- Hirugarren taldea: 8, 3, 2, 5, 7, 8.

Hiru kasuetan, batez beste, 5 zigarro erretzen dituzte, eta desbideratze tipikoak dira: 6,98; 0,89 eta 2,58.
Ulergarria da bigarren kasuan desbideratze tipikoa txikiagoa izatea, eta zerora asko hurbiltzea. Izan ere, talde horretako balioak hurbilago daude batez bestekotik. Lehenengo taldean, berriz, desbideratze tipikoa handiagoa da, datuak gehiago urruntzen direlako batez bestekotik.

Balio isolatuetarako kalkulatzea

Ikerketarako ditugun balioak x_1, x_2, x_3,…,x_n izanik, desbideratze tipikoa kalkulatzeko, formula hau erabil dezakegu:

$$\sigma = \sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\ldots+(x_n-\bar{x})^2}{N}}$$

Formula hori modu honetan laburbiltzen da:

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

Adibidea. Desbideratze tipikoa kalkulatzea: 8 ikaslek ikasgelara ekarri duten diruaren desbideratze tipikoa. Demagun 8 ikaslek 9, 3, 8, 8, 9, 8, 9 eta 18 euro ekarri dituztela ikasgelara. Formula aplikatuta, erraz lortzen da desbideratze tipikoa:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{N}} \qquad \sigma = \sqrt{\frac{93}{8}} = 3{,}4$$

8 ikasle horiek ikasgelara ekarri duten dirua 9, 8, 8, 9, 9, 9, 8, 10 izan balitz, desbideratze tipikoa txikiagoa litzateke; bigarren kasu horretan, balioak gehiago hurbiltzen direlako batez bestekora (lehenengo kasuan $\bar{x} = 9$ da, eta bigarrenean, $\bar{x} = 8{,}75$); hau da, gutxiago sakabanatzen dira.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{N}} \qquad \sigma = \sqrt{\frac{3{,}496}{8}} = 0{,}66$$

Desbideratze tipikoaren kalkulua errazteko, aurreko formulen baliokideak diren honako formula hauek ere erabiltzen dira:

$$\sigma = \sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_n^2}{N} - \bar{x}^2}$$

Formula honetan modu honetan laburbil dezakegu:

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

Egia esan, azken bi formula horiek maizago erabiltzen dira, kalkulua errazagoa delako.

Desbideratze tipikoa kalkulatzea datu multzoa maiztasun-tauletan banatuta dagoenean

Datuak maiztasun-taula batean banatuta daudenean, honako formula hau erabiliko dugu desbideratze tipikoa kalkulatzeko:

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 f_1 + (x_2 - \bar{x})^2 f_2 + \ldots + (x_n - \bar{x})^2 f_n}{N}}$$

Formula hori modu honetan laburbil dezakegu:

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

Desbideratze tipikoaren kalkulua errazteko, aurreko formulen baliokideak diren honako formula hauek ere erabiltzen dira:

$$\sigma = \sqrt{\frac{x_1^2 f + x_2^2 f_2 + \ldots + x_n^2 f_n}{N} - \bar{x}^2}$$

Formula hori modu honetan laburbil dezakegu:

$$\sigma = \sqrt{\sum_{i=1}^{n} \frac{x_i^2 f_i}{N} - \bar{x}^2}$$

Adibidea. Desbideratze tipikoa kalkulatzea: 25 pertsonaren seme-alaben kopurua.

48. taula: seme-alaben kopurua				
x_i (seme-alaben kopurua)	f_i	x_i^2	$x_i.f_i$	$x_i^2.f_i$
1	3	1	3	3
2	4	4	8	16
3	5	9	15	45
4	3	16	12	48
5	6	25	30	150
6	4	36	24	144
N	25		92	406
Iturria: lanketa propioa				

Taulako datuak erabiliz, formula aplikatzen dugu:

$$\sigma = \sqrt{\sum_{i=1}^{n} \frac{x_i^2 f_i}{N} - \bar{x}^2} \qquad \sigma = \sqrt{\frac{406}{25} - 13{,}54} = 2{,}7$$

$\bar{x} = 3{,}68$ izanik, eta datu multzoko balioak ikusita, desbideratze tipikoa nahiko adierazgarria dela ondorioztа dezakegu. Hau da: pertsona talde hori nahiko heterogeneoa da, seme-alaba kopuruari dagokionez.

Adibidea. Desbideratze tipikoa kalkulatzea: 50 egunetako tenperatura herri batean.

49. taula: 50 eguneko tenperatura herri batean			
Graduak (x_i)	Egun kopurua (f_i)	$x_i.f_i$	$x_i^2.f_i$
9	1	9	81
10	4	40	400
11	9	99	1089
12	16	192	2304
13	11	143	1859
14	8	112	1568
15	1	15	225
	50	610	7526
Iturria: lanketa propioa			

Formula aplikatuta erraz lortzen da desbideratze tipikoa:

$$\sigma = \sqrt{\sum_{i=1}^{n} \frac{x_i^2 f_i}{N} - \bar{x}^2} \qquad \sigma = \sqrt{\frac{7526}{50} - 12{,}2} = 1{,}68$$

$\bar{x} = 12{,}2$ izanik, eta datu multzoa ikusita, desbideratze tipikoa txikia dela ondorioztа dezakegu. Hau da: herri horretan, tenperatura nahiko homogeneoa izan da 50 egun horietan.

Adibidea. Desbideratze tipikoa kalkulatzea: gastronomia-elkarte bateko 40 kideren adina.

50. taula: gastronomia-elkarteko kideen adina			
adina	f_i	$x_i.f_i$	$x^2_i.f_i$
51	2	102	5202
56	4	224	12544
61	11	671	40931
66	12	792	52272
71	7	497	35287
76	4	304	23104
N	40	2590	169340
Iturria: lanketa propioa			

$$\sigma = \sqrt{\sum_{i=1}^{n} \frac{x_i^2 f_i}{N} - \bar{x}^2} \qquad \sigma = \sqrt{\frac{169340}{40} - 64{,}75} = 6{,}39$$

Batez bestekoa 64,75 urte delarik, desbideratze tipikoa nahiko handia da. Izan ere, elkarte horretan, 51 urte dituztenak eta 70 urtetik gorakoak elkartzen dira. Hau da, aldakortasun, heterogeneotasun edo dispertsio adierazgarria dago adinaren ikuspuntutik.
Datu multzoko balioak tartetan banatuta badaude, desbideratze estandarra kalkulatzeko tarteen markak ezagutu behar dira, lehenengo, eta, ostean, formula berak aplikatu.

Adibidea. Desbideratze tipikoa kalkulatzea datu multzoko balioak tartetan banatuta daudenean.

51. taula: auzo bateko kideen adina				
	Markak	fi	$x_i.f_i$	$x^2_i.f_i$
10-20	15	1	15	225
20-30	25	8	200	5000
30-40	35	10	350	12 250
40-50	45	9	405	18 225
50-60	55	8	440	24 200
60-70	65	4	260	16 900
70-80	75	2	150	11 250
		42	1820	88 050
Iturria: lanketa propioa				

Formula aplikatuta desbideratze tipikoa erraz lortzen da:

$$\sigma = \sqrt{\sum_{i=1}^{n} \frac{x_i^2 f_i}{N} - \bar{x}^2} \qquad \sigma = \sqrt{\frac{88050}{42} - 43{,}33^2} = 14{,}79$$

$\bar{x} = 43{,}33$ izanik, eta datu multzoa ikusita, desbideratze tipikoa handia dela ondoriozta dezakegu. Hau da: auzo horretan heterogeneotasun handia dago adinari dagokionez.

Adibidea. Desbideratze tipikoa kalkulatzea IBM SPSS programaren laguntzarekin: 577 jaioberriren pisuaren desbideratze tipikoa.

52. taula: jaioberrien pisua

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	2,10	14	2,4	2,4	2,4
	2,20	15	2,6	2,6	5,0
	2,30	28	4,9	4,9	9,9
	2,40	27	4,7	4,7	14,6
	2,50	24	4,2	4,2	18,7
	2,60	23	4,0	4,0	22,7
	2,70	63	10,9	10,9	33,6
	2,80	169	29,3	29,3	62,9
	2,90	65	11,3	11,3	74,2
	3,00	38	6,6	6,6	80,8
	3,10	24	4,2	4,2	84,9
	3,20	39	6,8	6,8	91,7
	3,30	35	6,1	6,1	97,7
	3,40	13	2,3	2,3	100,0
	Total	577	100,0	100,0	

Iturria: lanketa propioa

Datu multzo horretatik abiatuta, IBM SPSS programak erraz kalkulatzen du desbideratze tipikoa:

Estadísticos
jaioberrien pisua

N	Válidos	577
	Perdidos	0
Media		2,8009
Mediana		2,8000
Moda		2,80
Desv. Típ.		,29840
Varianza		,089

Kasu honetan, $\bar{x} = 2{,}8$ delarik, desbideratze tipikoa (s) 0,29 da. Beraz, desbideratze tipikoa oso txikia dela ondoriozta dezakegu. Desbideratze tipiko txiki hori ulergarria da, kontuan hartzen badugu mediana, batez bestekoa eta moda oso antzerakoak direla, eta balio guztiak batez bestekoaren inguruan dabiltzala.

Desbideratze tipikoaren abantailak eta desabantailak

- Bariantzak ez bezala, sakabanatzea ez du unitate karratuetan neurtzen, aztertzen ari garen aldagaiaren unitate berdinetan baizik. Beraz, interpretazioa asko errazten da. Alegia, batez bestekoarekin harremanetan jartzea errazagoa da. Edonola ere, desbideratze tipikoak interpretazio zaila du.
- Asko erabiltzen da prozedura estatistikoetan. Banaketa normala eredutzat hartzen badugu, desbideratze tipikoaren laguntzarekin, maiztasun-banaketako balio batetik gora, behera edo bi balioren artean zenbat datu dauden oso erraz jakin dezakegu. Inferentzia

estatistikoan ere, laginketa-banaketen errore tipikoa kalkulatzen edota parametroa egon daitekeen tarteak zehazten laguntzen digu.

- Beste sakabanatze-neurriekin gertatzen den bezala, ezin da finkatu balio bat sakabanatze handia erakusten duenik. Dena dela, kontuan izan behar dugu zenbat eta desbideratze estandarra txikiagoa izan, edo, hau da, zerora zenbat eta gehiago hurbildu, orduan eta sakabanatze txikiagoa dagoela datu multzoan, eta, beraz, datuak orduan eta gehiago biltzen direla batez bestekoaren inguruan. Horrelako kasuetan, datu multzoan homogeneotasun handiagoa dagoela esaten dugu.
- Sakabanatze-neurri absolutua den heinean, desbideratze tipikoa ezin da erabili datu multzo desberdinetako sakabanatze-mailak alderatzeko.
- Datu arraroek edo muturrekoek eragin handia dute desbideratze tipikoan. Hala, muturreko datu asko ditugunean, hobe da beste neurri jasangarriago batzuk erabiltzea, adibidez, kuartilen desbideratzea.
- Batez besteko berdina duten maiztasun-banaketa desberdinak baditugu eta beren desbideratze tipikoak ezagutzen baditugu, guztizko desbideratze tipikoa ezagut dezakegu.

Lagin guztiak tamaina berekoak badira, guztizko desbideratze tipikoa edo totala modu honetan kalkulatzen da:

$$\sigma = \sqrt{\frac{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2}{n}}$$

Batez besteko berdina duten maiztasun-banaketak baditugu, baina laginen tamainak desberdinak badira, guztizko desbideratze tipikoa edo totala modu honetan kalkulatzen da:

$$\sigma = \sqrt{\frac{k_1 \cdot \sigma_1^2 + k_2 \cdot \sigma_2^2 + \ldots + k_n \cdot \sigma_n^2}{k_1 + k_2 + \ldots + k_n}}$$

Azken formula horretan, k datu multzo bakoitzaren tamaina da.

2.3.2.2. Sakabanatze-neurri erlatiboak

Sakabanatze-neurri absolutuek bezala, sakabanatze-neurri erlatiboek ere datu multzo baten sakabanatze, heterogeneotasun edo aldakortasuna neurtzen dute, aldagaia neurtzeko erabilitako unitatea kontuan hartu gabe. Alegia, horien bidez lortzen ditugun emaitzak dimentsiorik gabekoak dira. Neurri horiek kalkulatzea batez bestekoa edo beste zentro-neurri batzuk erabiltzen dira, eta, beraz, datu multzo desberdinen sakabanatzeak alderatzeko erabil daitezke. Hori da, hain zuzen ere, neurri horien abantaila handiena. Izan ere, kontuan izan behar dugu sakabanatze-neurri absolutuen arazo nagusiena dela ezin direla erabili datu multzo desberdinak alderatzeko. Sakabanatze-neurri erlatibo asko daude, baina guk ezagunenak aurkeztuko ditugu: ibiltarte erlatiboa, kuartilen aldakortasun-koefizientea eta aldakortasun-koefizientea.

2.3.2.2.1. Ibiltarte erlatiboa

Ibiltarte erlatiboa kalkulatzeko, formula honako hau da:

$$I_{\bar{x}} = \frac{I}{\bar{x}}$$

Ikusten den bezala, ibiltarte erlatiboa lortzeko, ibiltartea batez besteko aritmetiko sinplearekin zatitzen dugu. Zenbat eta ibiltarte erlatiboa handiagoa izan, sakabanatzea handiagoa izango da. Ibiltarte erlatiboa zero da sakabanatzerik, aldakortasunik edo heterogeneotasunik ez dagoenean (balio guztiak berdinak direnean).

Abantailak eta desabantailak

- Erraz kalkulatzen da, eta datu multzoen sakabanatzeak alderatzeko erabil daiteke.
- Desabantaila: ez da neurri jasankorra; izan ere, muturreko datuek eragin handia dute bere kalkuluan, ibiltartean eragiten dutelako. Beraz, muturreko datuak edo datu arraroak ditugunean hobe da ez erabiltzea.
- Beste desabantaila bat da ezin dela finkatu balio jakin bat sakabanatze handia erakusten duenik: datu multzo batean sakabanatze handia edo txikia dagoen jakiteko, beste datu multzo batekin alderatu behar dugu.

Adibidea. Ibiltarte erlatiboa: atletismoko bi talderen alderaketa. Lehen taldeko lasterkariek, batez beste, 30 minutu behar izan dituzte lasterketa bukatzeko. Denbora gutxien egin duenak 24 minutu behar izan ditu lasterketa bukatzeko, eta gehien egin duenak 32 minutu. Bigarren taldeko lasterkariek, berriz, batez beste, 32 minutu behar izan dituzte lasterketa bukatzeko, eta denbora gutxien egin duenak 23 minutu egin ditu, eta gehien egin duenak, 36 minutu. Zein da talde homogeneoena? Non dago aldakortasun txikiena?

1. taldea	$I_{\bar{x}} = \frac{I}{\bar{x}}$	$I_{\bar{x}} = \frac{32-24}{30} = 0{,}26$
2. taldea		$I_{\bar{x}} = \frac{36-23}{32} = 0{,}40$

Ibiltarte erlatiboaren emaitzak ikusita, bigarren taldean aldakortasun edo sakabanatze handiagoa dago. Horrenbestez, lehen taldea homogeneoagoa da lasterketa bukatzeko lasterkariek behar izan duten denborari dagokionez.

2.3.2.2.2. Kuartil arteko ibiltarte erlatiboa

Honako hau da kuartil arteko ibiltarte erlatiboa (KAIE) kalkulatzeko formula :

$$KAIE = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Zenbat eta sakabanatze handiagoa izan, kuartil arteko ibiltarte erlatiboa handiagoa izango da, eta zero da sakabanatzerik, aldakortasunik edo heterogeneotasunik ez dagoenean (balio guztiak berdinak direnean). Dena dela, ezin da finkatu balio bat sakabanatze handia erakusten duenik.

Abantailak eta desabantailak

- Abantaila: neurri jasankorra da, muturreko datuak baztertu eta kuartiletan bakarrik oinarritzen delako.
- Beste abantaila bat: datu multzo desberdinen sakabanatzeak alderatzeko erabil daitekeela.

- Desabantaila: ezin da finkatu balio jakin bat sakabanatze handia erakusten duenik: datu multzo batean sakabanatze handia edo txikia dagoen jakiteko, beste datu multzo batekin alderatu beharra dago.

Adibidea. Kuartil arteko ibiltarte erlatiboa kalkulatzea: aurreko adibidearekin jarraituz, lehen taldeko lasterkarien %25ek 27 minutu behar izan ditu (Q_1), eta %75ek, 31 minutu (Q_3). Bigarren taldeko lasterkarien artean, berriz, %25ek 24 minutu behar izan ditu (Q_1), eta %75ek, 33 minutu (Q_3). Bi taldeetatik, zein da talde homogeneoena? Non dago aldakortasun txikiagoa?

1. taldea	$KAIE = \frac{Q_3 - Q_1}{Q_3 + Q_1}$	$KAIE = \frac{31-27}{31+27} = 0{,}068$
2. taldea		$KAIE = \frac{33-24}{33+24} = 0{,}15$

Aurreko adibidean ikusi bezala, bigarren taldean, aldakortasun edo sakabanatze handiagoa dago.

2.3.2.2.3. Aldakortasun-koefizientea edo Pearson-en aldakuntza-koefizientea

Aldakortasun-koefizientea desbideratze estandarra batez bestekoarekin zatituz kalkulatzen da:

$$AK = \frac{\sigma}{\bar{x}}$$

Zenbat eta sakabanatze handiagoa izan, aldakortasun-koefizientea handiagoa izango da, eta zero da sakabanatzerik, aldakortasunik edo heterogeneotasunik ez dagoenean (balio guztiak berdinak direnean).
Autore batzuek Pearson-en aldakuntza-koefizientea deitzen diote koefiziente horri, eta, emaitza ehunekotan ematen bada, aldakortasun-koefizientea. Ehunekotan jarrita, datu bakoitza batez besteko aritmetiko sinpletik batez beste ehunekotan zenbat desbideratzen den adierazten digu:

$$AK = \frac{\sigma}{\bar{x}}.100$$

Adibidea. Aldakortasun-koefizientea kalkulatzea IBM SPSS programaren laguntzarekin: idazlan batean bi taldetako kideek (biak 90 kidez osatuak) egin duten akats kopurua.

53. taula: lehen taldeak egindako akats kopurua					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1,00	22	24,4	24,4	24,4
	2,00	20	22,2	22,2	46,7
	3,00	17	18,9	18,9	65,6
	4,00	19	21,1	21,1	86,7
	5,00	12	13,3	13,3	100,0
	Total	90	100,0	100,0	
Iturria: lanketa propioa					

Ikusten den bezala, lehen taldeko %65,6k 3 akats edo gutxiago egin ditu. Akats gehien egin dituenak 5 egin ditu, eta gutxien egin dituenak, 1. Lehen begiradan, talde nahiko homogeneoa da.

54. taula: bigarren taldeak egindako akats kopurua					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	3,00	12	13,3	13,3	13,3
	4,00	18	20,0	20,0	33,3
	5,00	18	20,0	20,0	53,3
	13,00	6	6,7	6,7	60,0
	44,00	6	6,7	6,7	66,7
	55,00	6	6,7	6,7	73,3
	66,00	6	6,7	6,7	80,0
	77,00	6	6,7	6,7	86,7
	88,00	6	6,7	6,7	93,3
	99,00	6	6,7	6,7	100,0
	Total	90	100,0	100,0	
Iturria: lanketa propioa					

Taulan ikus daitekeen bezala, bigarren taldeko %66,7ak 44 akats edo gutxiago egin ditu. Akats gehien egin dituenak 99 egin ditu, eta gutxien egin dituenak, 3. Lehen begiradan, talde nahiko heterogeneoa da.
Datu multzo horietatik abiatuta, IBM SPSS programak erraz ateratzen ditu aldakortasun-koefizientea kalkulatzeko behar ditugu estatistikoak:

Estadísticos			
		Lehen taldeak egindako akats kopurua	Bigarren taldeak egindako akats kopurua
N	Válidos	90	90
	Perdidos	0	0
Media		2,7667	31,6667
Mediana		3,0000	5,0000
Moda		1,00	4,00a
Desv. Típ.		1,38247	34,89245
Varianza		1,911	1217,483
Rango		4,00	96,00
a. Existen varias modas. Se mostrará el menor de los valores.			

Lehen taldearen aldakortasun-koefizientea modu honetan kalkulatuko dugu:

$$AK = \frac{\sigma}{\overline{x}} \qquad AK = \frac{1{,}38}{2{,}76} = 0{,}5$$

Bigarren taldearen aldakortasun-koefizientea modu honetan kalkulatuko dugu:

$$AK = \frac{\sigma}{\overline{x}} \qquad AK = \frac{34{,}89}{31{,}66} = 1{,}1$$

Argi dago bigarren taldean lehenengoan baino aldakortasun, dispertsio edo heterogeneotasun handiagoa dagoela.

Adibidea. Datu multzo biren dispertsioaren alderaketa aldakortasun-koefizientearen bidez. Demagun bi estatu multzo ditugula: estatu aberatsak eta estatu pobreak. Estatu aberatsek, batez beste, 40.000 euroko errenta per capita dute, 3.000 euroko desbideratze tipikoarekin; eta pobreek 4.000 euroko errenta, 2.000 euroko desbideratze tipikoarekin. Aldakortasun, heterogeneotasun edo dispertsio handiagoa non dagoen jakiteko, aldakortasun-koefizientea kalkulatuko dugu.

Estatu aberatsen aldakortasun-koefizientea honela kalkulatzen da:

$$AK = \frac{\sigma}{\bar{x}} \qquad AK = \frac{3000}{40000} = 0{,}075$$

Estatu pobreen aldakortasun-koefizientea honela kalkulatzen da:

$$AK = \frac{\sigma}{\bar{x}} \qquad AK = \frac{2000}{4000} = 0{,}5$$

Aldakortasun-koefizientea handiagoa izanik, argi dago aldakortasun, heterogeneotasun edo dispertsio handiagoa dagoela estatu pobreen artean.

Abantailak eta desabantailak

- Aldakortasun-koefizientea oso egokia da datu multzoen sakabanatze, dispertsio edo heterogeneotasun maila alderatzeko; besteak beste, hura kalkulatzeko batez bestekoak erabiltzen direlako, eta lortzen den emaitza dimentsiorik gabekoa delako.
- Batez besteko aritmetiko sinplea 0 denean edo 0tik oso gertu dagoenean, desbideratze estandarrean gertatzen den edozein aldaketak eragin handia du aldakortasun-koefizientean, eta haren emaitza baliogabetu. Beraz, batez bestekoa zero denean edo zerotik hurbil dagoenean, aldakortasun-koefizientea ezin da erabili.
- Modu berean, koefiziente hori kalkulatzeko batez bestekoak balio positiboa izan behar du beti.
- Koefiziente horren beste desabantaila bat: ezin da finkatu balio jakin bat sakabanatze handia erakusten duenik. Datu multzo batean sakabanatze handia edo txikia dagoen jakiteko, beste datu multzo batekin alderatu behar dugu.

2.3.2.3. Sakabanatze-neurriei buruzko ondorio orokorrak

- Sakabanatze-neurriek datu multzoko datuak batez bestekotik edo beste zentralizazio-neurri batetik zenbat aldentzen diren adierazten digute. Beraz, zentralizazio-neurriekin (bereziki batez bestekoarekin), harremanetan hartzen dute zentzua. Zenbat eta sakabanatze handiagoa egon, datuak orduan eta gehiago aldentzen dira batez bestekotik edota zentralizazio-neurrietatik, eta sakabanatze-neurrien balioa handiagoa da. Sakabanatzerik, aldakortasunik edo heterogeneotasunik ez duten datu multzoetan, hau da, balio guztiak berdinak direnean, sakabanatze-neurrien balioa zero da.
- Neurri horietariko batzuk aldagaiaren unitate berdinetan adierazten dira (desbideratze tipikoa, adibidez); beste batzuk, aldagaiaren unitate karratutan (Bariantza) eta, azkenik, beste batzuek ez dute dimentsiorik, eta ehunekotan adierazten dira (Pearson-en aldakuntza-koefizientea edo aldakortasun-koefizientea).
- Sakabanatze-neurri absolutuek datu multzo bakarraren aldakortasuna neurtzeko balio dute eta erlatiboek datu multzo desberdinak alderatzeko.

2.3.3. Joera zentralekoak ez diren posizio-neurriak

Joera zentraleko posizio neurriek (batez bestekoak, medianak eta modak) datu multzoaren zentroko posizioei buruzko informazio ematen digute. Batez bestekoak, adibidez, datuak zein balioren edo posizioren inguruan biltzen diren esaten digu; medianak datu multzoaren erdiko balioa edo posizioa zein den, hau da, maiztasun-banaketa zein posiziotan erdi banatzen den, eta modak gehien agertzen den balioa edo posizioa zein den erakusten digu.

Joera zentralekoak ez diren posizio-neurriek edo koantilek datu kopuru bera duten taldeetan banatzen dute datu multzoa. Alegia, koantilak datu multzoan kopuru bereko datuen azpimultzoak zehazten dituzten balioak dira; datuen ehuneko zehatza gora edo behera uzten duten posizioak.

Adibidea. Koantilen azalpena: demagun 15 pertsonak test batean lortu duten puntuazioak ditugula: 3, 2, 4, 5, 2, 7, 1, 10, 7, 9, 14, 12, 13, 15 eta 10.

Koantilak kalkulatzeko, lehenengo eta behin balioak ordenatu behar ditugu txikienetik handienera: 1, 2, 2, 3,4, 5, 7, 7, 9, 10, 10, 13, 13, 14 eta 15.
Datu multzo hori 4, 5, 8, 10 edo 100 azpimultzotan banatu daiteke, gure hipotesi, helburu edo interesen arabera. Alegia, kuartilak, kintilak, oktilak, dezilak edo pertzentilak kalkula ditzakegu.
Demagun datu multzo hori lau zatitan banatu nahi dugula; hau da, kuartilak kalkulatu nahi ditugula. 15 datu ditugunez, azpimultzo bakoitzean, 3,75 datu geldituko dira ($\frac{15}{4} = 3{,}75$), biribilduta 4 datu multzo bakoitzean; beraz, 1. koartila laugarren balioa izango da, hain zuzen ere, 3 balioa (3 balioak datuen %25 uzten du bere azpitik). Erdiko tokia 8. datuak hartzen du, 7 balioari dagokiona (7 balioak datuen %50 uzten du bere azpitik). Datu kopuruaren hiru laurden 11,25 da; hortaz, 12. balioa da hirugarren koartila, hau da,13. balioa (13. balioak datuen %75 uzten du bere azpitik).
Adibide honetan, 4 koantilek lau multzo berdinetan banatzen dituzte 15 pertsona horiek, eta, bakoitzean, testa erantzun duten pertsonen %25 kokatzen da. Q_1-ek (3), Q_2-k (7) eta Q_3-k (13) datu multzoa lau zatitan zein posiziotan banatzen den esaten digute.

2.3.3.1. Koantil erabilienen (kuartilen, pertzentilen eta dezilen) berezitasunak

Datu multzoak nahi ditugun azpimultzotan banatu ditzakegu, baina, egiatan, hiru dira gehien erabiltzen diren koantilak:

- 4-koantilak datu multzoa lau azpimultzo berdinetan banatzen duten 3 kuartil dira: Q_1, Q_2 eta Q_3 (lehen ordenako kuartila, bigarren ordenako kuartila eta hirugarren ordenako kuartila). Maiztasun-banaketaren datu guztiak txikienetik handienera ordenatu ostean, lehenengo koartilak (Q_1) ezker aldean uzten du datuen laurdena. Hau da, datuen %25 lehenengo koartilaren berdina edo txikiagoa da. Q_2 (bigarren koartila) medianarekin bat dator, eta haren azpitik uzten du datuen %50, eta Q_3-k datuen hiru laurdenak ezker aldean uzten ditu, hots, datuen %75 hirugarren koartilaren berdina edo txikiagoa da.
- 100-koantilak datu multzoa 100 azpimultzo berdinetan banatzen duten 99 pertzentil edo zentil dira: P_1, P_2, ..., P_{99}. Maiztasun-banaketaren datu guztiak txikienetik handienera ordenatuta, lehenengo pertzentilak datuen %1 ezker aldean uzten du. Alegia, datuen %1 lehenengo dezilaren berdina edo txikiagoa da. 50. pertzentila medianarekin bat dator; hau da, datuen erdia P_{50} baino txikiagoa da.
- 10-koantilak datu multzoa 10 azpimultzo berdinetan banatzen duten 9 dezil dira: D_1, D_2, ..., D_9 (Lehen ordenako dezila, bigarren ordenako dezila…). Maiztasun-banaketaren

datu guztiak txikienetik handienera ordenatu ostean, lehenengo dezilak datuen hamarrena (%10) ezkerraldean uzten du. Hau da, datuen hamarrena lehenengo dezilaren berdina edo txikiagoa da. Bosgarren dezila medianarekin bat dator; alegia, datuen %50 D_5 baino txikiagoa da.

4-koantilak, 100-koantilak eta 10-koantilak dira gehien erabiltzen diren koantilak. Dena dela, gure ikerketaren hipotesi, helburu edo nahien arabera, 5-koantilak (kintilak) eta 8-koantilak (oktilak) ere kalkula ditzakegu. Kintilek maiztasun bereko 5 azpimultzotan banatzen dute maiztasun-banaketa, eta oktilek, 8 azpimultzotan.

2.3.3.2. Kuartilen kalkulua

Esan bezala, kuartilak hiru dira (Q_1, Q_2 eta Q_3), eta lau azpimultzo berdinetan banatzen dute datu multzoa. Hiru balio horiek kalkulatzeko modua aldatu egiten da datu isolatuak ditugunean eta maiztasun-taula batean banatuta ditugunean. Halaber, kalkulua desberdina da aldagai diskretuekin eta jarraituekin.

Kuartilak kalkulatzea balio isolatuekin

Kuartilen kalkulua desberdina da datu multzoa bakoitia edo bikoitia denean.

Datu multzo bakoitia

Lehenengo eta behin balioak txikienetik handienera ordenatu behar dira. Ostean, kuartil bakoitzak duen tokia edo posizioa kalkulatzen dugu, honako formula honen bidez:

$$\frac{k.n}{4}$$

Q_1-entzat, k = 1 da; Q_2-rentzat, k = 2 da; eta Q_3-rentzat, k = 3 da.

Adibidea. Koartilen kalkulua datu multzo bakoitia dugunean: demagun, 2, 3, 4, 7, 8, 9 eta 10 balioez osatutako datu multzo bakoitia dugula. Q_1-en posizioa kalkulatzeko, formula aplikatzen dugu:

$$Q_1 = \frac{1.7}{4} = 1{,}75$$

Emaitza biribiltzen badugu, Q_1 bigarren posizioan dagoen balioari dagokio; gure kasuan, 3 balioari. Horrek esan nahi du datu multzoko datuen %25ek 3ko balioa edo gutxiago duela.
Q_2-ren posizioa era honetan kalkulatzen dugu:

$$Q_2 = \frac{2.7}{4} = 3{,}5$$

Biribiltzen badugu, Q_2 laugarren posizioan dagoen balioari dagokio; kasu honetan, 7 balioari.

Q_3-ren posizioa kalkulatzeko:

$$Q_3 = \frac{3.7}{4} = 5{,}25$$

Emaitza biribiltzen badugu, Q_3 seigarren datuari dagokio; hau da, 9 balioari.

Hala, Q_1-ek (3), Q_2-k (7) eta Q_3-k (9) lau zatitan banatzen dute datu multzoa.

Datu multzo bikoitia

Datu multzoa bikoitia denean, kuartil bakoitzak datuen artean duen tokia edo posizioa kalkulatzeko, honako formula hau erabili behar da:

$$\frac{k(n+1)}{4}$$

Aurreko kasuan bezala, Q_1-entzat, k = 1 da; Q_2-rentzat, k = 2 da; eta Q_3-rentzat, k = 3 da. Hala, Q_1-en posizioa kalkulatzeko, honako formula hau erabili behar dugu:

$$Q_1 = \frac{1(n+1)}{4}$$

Adibidea. Koartilen kalkulua datu multzo bikoitia dugunean: demagun 1, 2, 3, 4, 5, 6, 7, 8, 9 eta 10 balioekin osatutako datu multzoa dugula. Kasu honetan, Q_1-en posizioa ezagutzeko, zera egin behar dugu:

$$\frac{1(10+1)}{4} = 2{,}75$$

Emaitza biribiltzen badugu, Q_1 hirugarren posizioan dago; gure kasuan, beraz, Q_1 = 3 da.
Q_2 seigarren posizioan dagoen datuari dagokio; adibide honetan, 6 balioari.
Q_3 zortzigarren posizioan dagoen datuari dagokio; hau da, 8 balioari.

Kuartilen kalkulua aldagai diskretuekin

Aldagai diskretuak ditugunean, kuartilak kalkulatzeko, maiztasun erlatibo metatuen zutabean, bilatzen ari garen kuartilari dagokion ehunekoari dagokion balioari edo kuartil horri goitik gehien hurbiltzen zaion ehunekoari dagokion balioari hurrengo balio gehitzen zaio, eta, ostean, zati bi egiten da.

Adibidea. Koartilak kalkulatzea aldagai diskretuekin: kultura-elkarte bateko kideen adina.

55. taula: kultura-elkarte bateko kideen adina					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1,00	2	2,0	2,0	2,0
	2,00	1	1,0	1,0	3,0
	3,00	9	9,0	9,0	12,0
	4,00	7	7,0	7,0	19,0
	5,00	7	7,0	7,0	26,0
	6,00	5	5,0	5,0	31,0
	7,00	6	6,0	6,0	37,0
	8,00	5	5,0	5,0	42,0

	9,00	5	5,0	5,0	47,0
	11,00	24	24,0	24,0	71,0
	12,00	22	22,0	22,0	93,0
	13,00	7	7,0	7,0	100,0
	Total	100	100,0	100,0	
Iturria: lanketa propioa IBM SPSS programa erabiliz					

Maiztasun-taula horretan, Q_1 eta Q_3 kuartilak kalkulatzeko, zera egin behar dugu:
Q_1 kalkulatzea: maiztasun erlatibo metatuen zutabean, datuen %26 5 balioari dagokiola ikusten dugu, eta hurrengo balioa 6 da. Beraz,

$$Q_1 = \frac{x_5 + x_6}{2} = \frac{5+6}{2} = 5{,}5$$

Kultura-elkarte horretan, bazkideen %25ek 5 urte edo gutxiago ditu.
Q_2-en kalkulua: maiztasun erlatibo metatuen zutabean datuen %50 11 balioari dagokiola ikusten dugu, eta hurrengo balioa 12 da. Beraz,

$$Q_2 = \frac{11+12}{2} = 11{,}5$$

Kultura-elkarte horretan, bazkideen %50ek 11 urte edo gutxiago ditu.
Q_3 kalkulatzea: maiztasun erlatibo metatuen zutabean, datuen %75 12 balioari dagokio eta hurrengo balioa 13 da. Beraz,

$$Q_3 = \frac{12+13}{2} = 12{,}5$$

Kultura-elkarte horretan, bazkideen %75ek 12 urte edo gutxiago ditu.
Datu multzo horri dagozkion kuartilak IBM SPSS programaren bitartez kalkulatuta, honako emaitza hauek lortzen ditugu.

Estadísticos		
Kultur elkarte bateko kideen adina		
Percentiles	25	5,0000
	50	11,0000
	75	12,0000

Ikusten denez, IBM SPSS programak kuartilen balioak beherantz biribiltzen ditu. Kontuan izan programa horrek pertzentil hitza erabiltzen duela kuartilak, pertzentilak edo dezilak adieraztean.

Kuartilak kalkulatzea, datuak tartetan bildurik daudenean

Arestian egin bezala, lehenengo eta behin, kuartil bakoitzak duen tokia edo posizioa kalkulatzen dugu, honako formula honen bidez:

$$\frac{k.n}{4}$$

Gogoratu Q_1-entzat, k = 1 dela; Q_2-rentzat, k = 2 dela; eta Q_3-rentzat k = 3 dela.

Maiztasun metatuen zutabea kontuan hartuta, kuartila dagoen tartea aurkitzen dugu, eta, ostean, honako formula hau aplikatzen:

$$Q_k = L_i + \frac{\frac{k \cdot N}{4} - F_{i-1}}{f_i} \cdot a_i \qquad k = 1,2,3$$

Formularen osagaiak honako hauek dira:

- L_i: kuartila dagoen tartearen beheko muga da.
- N: maiztasun absolutu guztien batura edo datu multzoaren tamaina da.
- F_{i-1}: kuartila dagoen tartearen aurreko tartearen maiztasun metatua.
- a_i: tartearen zabalera da.
- f_i: kuartila dagoen tartearen maiztasuna.

Gogoratu, formula hori mediana kalkulatzeko erabili dugun bera dela. Izan ere, bigarren kuartilak eta medianak posizio bera adierazten dute datu multzoan.

Adibidea. Kuartilen kalkulua aldagai jarraituetarako: familia bateko kideen adina.

56. taula: familia bateko kideen adina			
Adin-tarteak	fi	Fi	% (maiztasun erlatiboak metatuak, ehunekoak)
0-20	1	1	1,3
20-40	5	6	7,9
40-60	26	32	42,1 (lehen kuartila tarte honetan dago)
60-80	36	68	89,7
80-100	8	76	100
N	76		
Iturria: lanketa propioa			

Lehen kuartila (Q_1) ezagutu nahi badugu, honako hauek dira formula aplikatzeko behar ditugun datuak:

$L_i = 40$
N = 76
$F_{i-1} = 6$
$a_i = 20$
$f_i = 26$

Formula modu honetan aplikatzen da:

$$Q_1 = 40 + \frac{\frac{1.76}{4} - 6}{26}.20 = 50$$

Q_1, kasu honetan, 50 urte da. Beraz, familia horretako %25ek 50 urte edo gutxiago ditu.

Q_2 edo mediana kalkulatzeko (gauza bera dira), modu honetan egiten dugu:

$$Q_2 = 60 + \frac{\frac{2.76}{4} - 32}{36}.20 = 63,3$$

Hala, familia horretako kideen %50ek 63 urte edo gutxiago ditu, eta %50ek 63 urte baino gehiago.

Adibidea. Kuartilen kalkulua aldagai jarraituetarako: herri txiki bateko biztanleen adina.

57. taula: herri txiki bateko biztanleen adina			
Biztanleen adinak	Biztanleak	Maiztasun metatuak	% (maiztasun erlatiboak metatuak, ehunekoak)
0-20	9	9	14
20-40	18	27	42,2
40-60	26	53	82,8 (hirugarren kuartila hemen dago)
60-80	7	60	93,7
80-100	4	64	100
N		64	
Iturria: lanketa propioa			

Mediana aztertzerakoan ikusi dugun bezala, kasu honetan, Q_2 edo mediana = 44 urte da.

Q_2 edo mediana ezagutzeko, honako balio hauek behar ditugu:

$L_i = 40$
$N = 64$
$F_{i-1} = 27$
$a_i = 20$
$f_i = 26$

Balio horiek ezagututa, Q_2 edo mediana erraz kalkulatzen da:

$$Q_2 = me = 40 + \frac{\frac{2.64}{4} - 27}{26}.20 = 43,8 \approx 44$$

Herri horretako biztanleen %50ek 44 urte edo gutxiago ditu, eta %50ek, 44 urte baino gehiago.

Q_3 edo hirugarren kuartila ezagutzeko, honako hauek dira formula aplikatzeko behar ditugun balioak:

$L_i = 40$
$N = 64$
$F_{i-1} = 27$
$a_i = 20$
$f_i = 26$

Q_2 eta Q_3 tarte berean daudenez, formula aplikatzeko behar ditugun datuak berdinak dira. Aldatzen den gauza bakarra k da.

$$Q_3 = 40 + \frac{\frac{3.64}{4} - 27}{26}.20 = 56,1$$

Emaitza ikusita, zera ondorioztatu dezakegu: herri horretako biztanleen %75ek 56 urte edo gutxiago ditu, eta %25ek, 56 urte baino gehiago.

2.3.3.3. Pertzentilen kalkulua

Pertzentilek edo zentilek datu multzoa 100 zati berdinetan banatzen dute. Beraz, 99 pertzentil edo zentil daude: P_1, P_2, ..., P_{99} (1. pertzentzila, 2. pertzentila,..., 99. pertzentzila).

Kuartilak kalkulatzean egin dugun bezala, pertzentilak kalkulatzeko modua aldatu egiten da balio isolatuak ditugunean edo maiztasun-taula batean banatuta ditugunean. Halaber, kalkulua desberdina da aldagai diskretuekin eta jarraituekin.

Pertzentilak kalkulatzea balio isolatuetarako

Pertzentilen kasuan ere, kalkulua desberdina da datu multzoa bakoitia edo bikoitia denean.

Datu multzo bakoitia

Lehenengo eta behin, balioak txikienetik handienera ordenatu behar ditugu. Ostean, pertzentil bakoitzak duen tokia edo posizioa kalkulatzen dugu, honako formula honen bidez:

$$\frac{k.n}{100}$$

k = 10 da 10. pertzentilerako; k = 35 da 35. pertzentilerako, edo k = 85, 85. pertzentilerako.

Adibidea. Pertzentilen kalkulua datu multzo bakoitia dugunean: demagun 3, 5, 9, 11, 15, 17 eta 22 balioekin osatutako datu multzo bakoitia dugula. 70. pertzentilari, maiztasun-banaketan, 5. posizioa duen 15 puntuazioa dagokio. Izan ere,

$$\frac{70.7}{100} = 4{,}9 \approx 5$$

Datu multzo bikoitia

Datu multzoa bikoitia denean, honako formula hau erabili behar da pertzentilen kalkuluan:

$$\frac{k.(n+1)}{100}$$

Aurreko kasuan bezala, P_{10} pertzentilerako, k = 10 da; P_{25} pertzentilerako, k = 25 da…

Adibidea. Pertzentilak kalkulatzea datu multzo bikoitia dugunean: demagun 1, 2, 3, 4, 5, 6, 7, 8, 9 eta 10 datu multzoa dugula. P_{75} modu honetan kalkulatzen dugu:

$$\frac{75.(10+1)}{100} = 8{,}25$$

P_{75}, beraz, datu multzoan 8. posizioa duen balioari dagokio: gure kasuan, 8 balioari. Horrek zera esan nahi du: datuen %75 8 edo balio gutxiagokoa dela.

Pertzentilen kalkulua aldagai diskretuekin

Pertzentila kalkulatzeko, maiztasun erlatibo metatuen zutabean, bilatzen ari garen pertzentzilari dagokion ehunekoaren balioari edo pertzentzil horri goitik gehien hurbiltzen zaion ehunekoaren balioari hurrengo balio gehitzen zaio, eta, zati bi egiten da.

Adibidea. Pertzentilak kalkulatzea aldagai diskretuekin: aisialdi talde bateko umeen adina.

58. taula: aisialdi-talde bateko umeen adina

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1,00	2	2,0	2,0	2,0
	2,00	1	1,0	1,0	3,0
	3,00	9	9,0	9,0	12,0
	4,00	7	7,0	7,0	19,0
	5,00	7	7,0	7,0	26,0
	6,00	5	5,0	5,0	31,0
	7,00	6	6,0	6,0	37,0
	8,00	5	5,0	5,0	42,0
	9,00	5	5,0	5,0	47,0
	11,00	24	24,0	24,0	71,0
	12,00	22	22,0	22,0	93,0
	13,00	7	7,0	7,0	100,0
	Total	100	100,0	100,0	

Iturria: lanketa propioa

Datu multzo honetako 10. eta 20. pertzentilak kalkulatzeko, zera egin behar dugu:
P_{10} kalkulatzeko, maiztasun erlatibo metatuen zutabean, datuen %10 3 balioari dagokio. Beraz,

$$P_{10} = \frac{X_3 + X_4}{2} = \frac{3+4}{2} = 3{,}5$$

Ateratzen dugun ondorioa zera da: aisialdi talde horretako umeen %10ek 3 urte edo gutxiago ditu.
Bestetik, P_{20} kalkulatzeko, maiztasun erlatibo metatuen zutabean, %20ri dagokion balioa 5 da.

$$P_{20} = \frac{5+6}{2} = 5{,}5$$

Beraz, aisialdi-talde horretako umeen %20k 5 urte edo gutxiago ditu.
Aurreko datu multzoan oinarrituta, IBM SPSS programak oso erraz kalkulatzen ditu nahi ditugun pertzentilak:

Estadísticos
Aisialdi talde bateko umeen adina

N	Válidos	100
	Perdidos	0
Percentiles	10	3,0000
	20	5,0000
	25	5,0000
	30	6,0000
	50	11,0000
	75	12,0000

Ikusten den bezala, aisialdi talde horretako umeen %75ek 12 urte edo gutxiago ditu.

Pertzentzilak kalkulatzea, datuak tartetan bildurik daudenean

Maiztasun metatuen zutabean, pertzentila dagoen tartea aurkitzen dugu, eta, ostean, honako formula hau aplikatzen:

$$P_k = L_i + \frac{\frac{k \cdot N}{100} - F_{i-1}}{f_i} \cdot a_i \qquad k = 1, 2, ..99$$

L_i: pertzentila dagoen tartearen beheko muga da.
N: maiztasun absolutu guztien batura da.
F_{i-1}: pertzentila dagoen tartearen aurreko tartearen maiztasun metatua.
a_i: tartearen zabalera da.

Adibidea. Pertzentilak kalkulatzea aldagaiaren balioak tartetan banatuta daudenean: eskola bateko irakasleek adimen-test batean lortutako puntuazioetan, P_{35} eta P_{60} kalkulatu:

59. taula: adimen-test batean lortutako puntuazioak

	fi	Fi	%	% metatuak
50-60	8	8	12	12
60-70	10	18	15	27
70-80	16	34	25	52
80-90	14	48	22	74
90-100	10	58	15	89
100-110	5	63	8	97
110-120)	2	65	3	100
	65		100	

Iturria: lanketa propioa

35. pertzentila modu honetan kalkulatzen dugu:

$$P_k = L_i + \frac{\frac{k \cdot N}{100} - F_{i-1}}{f_i} \cdot a_i \qquad k = 1, 2, ..99 \qquad P_{35} = 70 + \frac{\frac{65.35}{100} - 18}{16}.10 = 72{,}97 \approx 73$$

Emaitzak erakusten duen bezala, eskola horretako irakasleen %35ek 73 puntu edo gutxiago lortu ditu.

60. pertzentila modu honetan kalkulatzen dugu:

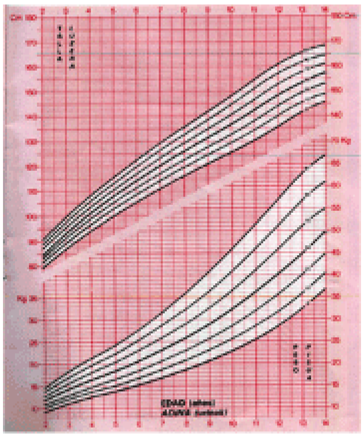
$$P_k = L_i + \frac{\frac{k \cdot N}{100} - F_{i-1}}{f_i} \cdot a_i \qquad k = 1, 2, ..99 \qquad P_{60} = 80 + \frac{\frac{65.60}{100} - 34}{14}.10 = 83{,}57 \approx 84$$

Beraz, eskola horretako irakasleen %60k 84 puntu edo gutxiago lortu ditu.

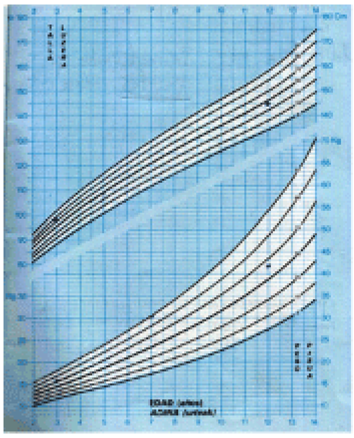
Pertzentilen erabilera

Pertzentilak askotan erabiltzen dira ume baten altuera eta pisua bere adineko beste umeenarekin alderatzeko; alegia, ume bakoitzaren altuera eta pisua altuera eta pisu guztien maiztasun-banaketaren barruan kokatzeko. Horrekin, gurasoek beren seme edo alaba adin bereko beste umeak baino lodiagoa, argalagoa, txikiagoa edo luzeagoa den jakin dezakete.
Horretarako, osasun-erakundeek sexu bakoitzari dagozkion pisua eta garaiera islatzen duten pertzentilen kurbak egiten dituzte (egiatan, ojibak dira), eta gurasoek, beren umearen adina kontuan hartuz, pisua eta altuera zehazten dute dagokion kurban. Horrekin, beren umeari dagokion pertzentil kurba (3, 10, 25, 50, 75, 90 edo 97) ezagutzen dute. Adibidez, ume baten pisuari 3. pertzentila badagokio, horrek esan nahi du adin horretako 100 umetik 3k gutxiago pisatzen dutela eta 97k, berriz, gehiago.

26. diagrama: emakumeen eta gizonen pisu eta altuera pertzentilak



Emakumeen pertzentilak



Gizonen pertzentilak

Iturria: Osakidetza

2.3.3.4. Dezilen kalkulua

Arestian esan bezala, 10-koantilak 9 dezil dira: D_1, D_2, ..., D_9 (lehen ordenako dezila, bigarren ordenako dezila…). Kuartilekin eta pertzentilekin bezala, dezilak kalkulatzeko modua aldatu egiten da datu isolatuak ditugunean edo maiztasun-taula batean banatuta ditugunean. Modu berean, kalkulua desberdina da aldagai diskretuekin eta jarraituekin.

Dezilen kalkulua balio isolatuekin

Dezilen kalkulua desberdina da datu multzoa bakoitia edo bikoitia denean.

Datu multzo bakoitia

Lehenengo eta behin, balioak txikienetik handienera ordenatu behar dira.

Ostean, dezil bakoitzak duen tokia edo posizioa kalkulatu behar da, honako formula honen bidez:

$$\frac{k.n}{10}$$

Kontuan izan k = 1 dela, 1. dezilerako; k = 3 dela, 3. dezilerako, edo k = 8 dela, 8. dezilerako.

Adibidea. Dezilak kalkulatzea datu multzo bakoitia dugunean: demagun datu multzo bakoiti hau dugula: 3, 5, 9, 11, 15, 17 eta 22. D_2-ri (bigarren dezilari) maiztasun-banaketan bigarren posizioa duen 5 puntuazioa dagokio. Izan ere,

$$\frac{2.7}{10} = 1{,}4 \approx 2$$

Datu multzo bikoitia

Datu multzoa bikoitia dugunean, honako formula hau erabili behar da:

$$\frac{k.(n+1)}{10}$$

Aurreko kasuan bezala, D_1 dezilarentzat, k = 1 da; D_2 dezilarentzat, k = 2 da…

Adibidea. Dezilen kalkulua datu multzo bikoitia dugunean: demagun 1, 2, 3, 4, 5, 6, 7, 8, 9 eta 10 datu multzoa dugula. 7. dezila kalkulatzeko, zera egin behar dugu:

$$\frac{7.(10+1)}{10} = 7{,}7 \approx 8$$

P_7, beraz, 8. posizioan dagoen balioari dagokio; gure kasuan, 8 balioari. Horrek esan nahi du datu multzoko datuen %70era 8-ko balioa edo gutxiagokoa dela.

Dezilak nola kalkulatu aldagai diskretuekin

Aldagai diskretuak ditugunean, maiztasun erlatibo metatuen zutabean, bilatzen ari garen dezilari dagokion ehunekoaren balioari edo dezil horri goitik gehien hurbiltzen zaion ehunekoaren balioari hurrengo balio gehitzen zaio, eta zati bi egiten da.

Adibidea. Dezilak kalkulatzea aldagai diskretuekin: aisialdi talde bateko umeen adina: D_5 eta D_2 kalkulatzea.

60. taula: aisialdi talde bateko umeen adina

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1,00	2	2,0	2,0	2,0
	2,00	1	1,0	1,0	3,0
	3,00	9	9,0	9,0	12,0
	4,00	7	7,0	7,0	19,0
	5,00	7	7,0	7,0	26,0
	6,00	5	5,0	5,0	31,0
	7,00	6	6,0	6,0	37,0
	8,00	5	5,0	5,0	42,0
	9,00	5	5,0	5,0	47,0
	11,00	24	24,0	24,0	71,0
	12,00	22	22,0	22,0	93,0
	13,00	7	7,0	7,0	100,0
	Total	100	100,0	100,0	

Iturria: lanketa propioa

D_5 kalkulatzeko, maiztasun erlatibo metatuen zutabean datuen %50 11 balioari dagokio. Beraz,

$$D_5 = \frac{X_{10} + X_{11}}{2} = \frac{11+12}{2} = 11{,}5 \approx 11$$

Hala, aisialdi talde horretako kideen %50ek 11 urte edo gutxiago ditu.
Bestetik, D_2 kalkulatzeko, maiztasun erlatibo metatuen zutabean, % 20ari dagokion balioa 5 da.

$$D_2 = \frac{5+6}{2} = 5{,}5 \approx 5$$

Beraz, aisialdi talde horretako kideen %20k 5 urte edo gutxiago ditu.
Aurreko datu multzoan oinarrituta, IBM SPSS programak oso erraz ateratzen ditu nahi ditugun dezilak:

Estadísticos
Aisialdi talde bateko umeen adina

N	Válidos	100
	Perdidos	0
Percentiles	10	3,0000
	20	5,0000
	25	5,0000
	30	6,0000
	50	11,0000
	75	12,0000

Dezilak kalkulatzea, datuak tartetan bildurik daudenean

Maiztasun metatuen zutabean, dezila dagoen tartea dugu, eta ostean, honako formula hau aplikatzen:

$$D_k = L_i + \frac{\frac{k \cdot N}{10} - F_{i-1}}{f_i} \cdot a_i \qquad k = 1,2,..9$$

L_i: dezila dagoen tartearen beheko muga da.
N: maiztasun absolutu guztien batura da.
F_{i-1}: dezila dagoen tartearen aurreko tartearen maiztasun metatua.
a_i: tartearen zabalera da.

Adibidea. Dezilak kalkulatzea aldagaiaren balioak tartetan banatuta daudenean: adimen-test batean lortutako puntuazioetan D_2 eta D_5-ak kalkulatzea:

61. taula: adimen-test batean lortutako puntuazioak				
	Fi	Fi	%	% metatuak
50-60	8	8	12	12
60-70	10	18	15	27
70-80	16	34	25	52
80-90	14	48	22	74
90-100	10	58	15	89
100-110	5	63	8	97
110-120)	2	65	3	100
	65		100	
Iturria: lanketa propioa				

Bigarren dezila (D_2) modu honetan kalkulatzen dugu:

$$D_k = l_i + \frac{\frac{k \cdot N}{10} - F_{i-1}}{f_i} \cdot a \qquad k = 1,2,..9 \qquad D_2 = 60 + \frac{\frac{65.2}{10} - 8}{10}.10 = 65$$

Zera ondorioztatzen dugu: lagin horretan ikertu diren kideen %20k 65 puntu edo gutxiago lortu dituela adimen-testean.

Bosgarren dezila (D_5) modu honetan kalkulatzen dugu:

$$D_k = L_i + \frac{\frac{k \cdot N}{10} - F_{i-1}}{f_i} \cdot a \qquad k = 1,2,..9 \qquad D_5 = 70 + \frac{\frac{65.5}{10} - 18}{16}.10 = 79{,}06$$

Kideen %50ek 79 puntu edo gutxiago lortu ditu adimen-testean.

2.3.3.5. Joera zentralekoak ez diren posizio-neurriei buruzko ondorio orokorrak

- Kuantilak aldagai kuantitatiboetarako eta kualitatibo ordinaletarako kalkula daitezke.
- Maiztasun-banaketetako posizioak zehazteaz gain (adimen-test bati erantzun diotenen %75 edo %80 zein puntuaziotik behera kokatzen diren, adibidez), kuantilak beste estatistiko deskriptibo batzuk kalkulatzeko edo datu-diagramak egiteko erabiltzen dira. Kuartilak, adibidez, kuartil arteko ibiltartean eta kuartilen desbideratzea edo ibiltarte semiinterkuartilikoa kalkulatzeko erabiltzen dira, baita kaxa-diagrama irudikatzeko ere. Pertzentilak, berriz, moztutako batez bestekoa kalkulatzeko erabiltzen dira.
- Kuantilak oso egokiak dira maiztasun-banaketa desberdinak alderatzeko. Alegia, datu multzo desberdinetan, kuartil, pertzentil edo dezil berdinei dagozkien balioak alderatu ditzakegu.
- Ezin da ahantzi kuantilen artean baliokidetasunak daudela: lehenengo koartila (Q_1) bat dator 25. pertzentilarekin (P_{25}); 2. koartila (Q_2), medianarekin (Me), 5. debilarekin (D_5) eta 50. pertzentilarekin (P_{50}).
- Arestian esan bezala, datu multzo bateko koantilak modu grafikoan irudikatzeko edota aztertzeko, ojiba edo maiztasun metatuen poligonoa erabil dezakegu. Datu-diagrama horren bidez, zenbat datu dauden tarte edo balio batetik gora eta zenbat azpitik ikus dezakegu.

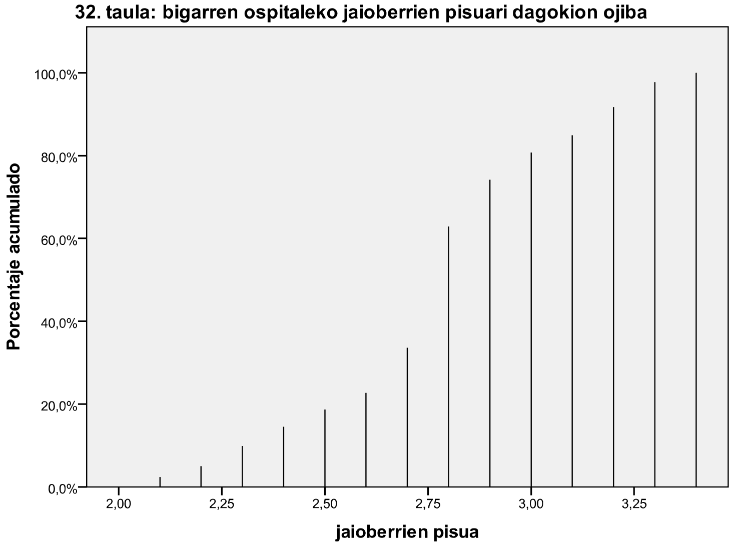
Adibidea. Aldagai jarraitu baten maiztasun erlatibo metatuen poligonoa: bigarren ospitale bateko jaioberrien pisua.

62. taula: bigarren ospitale bateko jaioberrien pisua				
x_i	f_i	F_i	%	% metatuak
2,10	14	14	2,4	2,4
2,20	15	29	2,6	5,0
2,30	28	57	4,9	9,9
2,40	27	84	4,7	14,6
2,50	24	108	4,2	18,7
2,60	23	131	4,0	22,7
2,70	63	194	10,9	33,6
2,80	169	363	29,3	62,9
2,90	65	428	11,3	74,2
3,00	38	466	6,6	80,8
3,10	24	490	4,2	84,9
3,20	39	529	6,8	91,7
3,30	35	564	6,1	97,7
3,40	13	577	2,3	100,0
Total	577		100,0	
Iturria: lanketa propioa				

Datu multzo horretako maiztasun erlatibo metatuetatik abiatuz, maiztasun erlatibo metatuen poligono hau lortzen dugu, IBM SPSS programa informatikoaren laguntzarekin:

27. diagrama

32. taula: bigarren ospitaleko jaioberrien pisuari dagokion ojiba



Iturria: nik neuk egina

Ojiban ikus dezakegunez, jaioberrien %25ek (Q_1), gutxi gorabehera, 2,65 kg. edo gutxiago ditu, %50ek (Q_2, D_5 eta P_{50}) 2,75 kg. edo gutxiago, edota %10ak ($P_{10} = D_1$) gutxi gorabehera 2,30 kg. edo gutxiago.

Adibidea. Koantilak kalkulatzea IBM SPSS programaren laguntzarekin: 100 bikoteri duten seme-alaba kopurua galdetu diegu, eta maiztasun-taula hau lortu:

63. taula: bikoteko seme alaba kopurua				
Seme alaba kopurua	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
1,00	22	22,0	22,0	22,0
2,00	38	38,0	38,0	60,0
3,00	15	15,0	15,0	75,0
4,00	15	15,0	15,0	90,0
5,00	5	5,0	5,0	95,0
6,00	5	5,0	5,0	100,0
Total	100	100,0	100,0	
Iturria: lanketa propioa				

Maiztasun-taula horretatik abiatuta, IBM SPSS programaren laguntzarekin, erraz lortzen ditugu dezilak, kuartilak eta pertzentzilak. Kontuan izan IBM SPSS programak pertzentil izena erabiltzen duela kuantil guztiak izendatzeko.

N	Válidos	100
	Perdidos	0
Media		2,5800
Mediana		2,0000
Moda		2,00
Desv. Típ.		1,37936
Varianza		1,903
Asimetría		,867
Error típ. de asimetría		,241
Curtosis		,045
Error típ. de curtosis		,478
Percentiles	10	1,0000
	20	1,0000
	25	2,0000
	30	2,0000
	40	2,0000
	50	2,0000
	60	2,6000
	70	3,0000
	75	3,7500
	80	4,0000
	90	4,9000

Ikusten den bezala, bikoteen %50ek (D_5 = Me = Q_2 = P_{50}) 2 seme alaba edo gutxiago ditu, eta bikoteen %80k, 4 seme-alaba edo gutxiago.

2.3.4. Forma- edo itxura-neurriak: alborapena eta kurtosia

Aldagai kuantitatiboen maiztasun-banaketaren formari buruzko informazioa barra-diagramek (aldagai diskretuetan), maiztasun bakunen poligonoek edo histogramek (aldagai jarraituetan) ematen digute. Adibidez, diagraman altuera gorena balio edo tarte bakarrari badagokio, moda bakarreko maiztasun-banaketa dela esaten dugu. Modari dagokion zutabea diagramaren mutur batean badago, eta, hortik abiatuta ezkerrera edo eskumara maiztasunak txikituz badoaz, maiztasun-banaketa L edo J formako dela esaten da, moda ezkerrean edo eskuinean gelditzen den.

Diagramak altuera goren bat baino gehiago baditu, moda askoko maiztasun-banaketa dela esaten dugu. Era horretako banaketen artean, bi dira aipagarriak: laukizuzen formako maiztasun-banaketak eta U formakoak.

Laukizuzen erako banaketetan, aldagaiaren balio edo tarte guztiek altuera bera dute barra-diagraman, maiztasun bakunen poligonoan edo histograman.

U formako banaketetan, bi moda daude, eta aldagaiaren lehen balioari edo tarteari eta azkenari dagozkien zutabeek dute altuera handiena; tarteko balioak edo tarteen maiztasunak txikiak dira.

Tartetan banatutako aldagai kuantitatiboak ditugunean, gehien agertzen diren maiztasun-banaketak muturretan, maiztasun gutxi eta erdian asko dituzten moda bakarreko maiztasun-banaketak dira; alegia, ohikoenak kanpai-formako histogramak ematen dituzte maiztasun-banaketak dira.

Barra-diagrama, maiztasun bakunen poligonoa edo histograma begiratzea garrantzitsua da, maiztasun-banaketaren formari buruzko informazio lortzeko. Dena dela, diagramen azterketarekin ez dugu maiztasun-banaketaren formari buruzko informazio zehatza lortzen. Banaketen forma neurtuko duten neurriak behar ditugu, modu intuitiboan ikusi duguna zehazteko.

Forma neurriek maiztasun-banaketa baten itxurari buruzko informazioa ematen dute; zehatzago esanda, zorroztasunari eta simetriari buruzko informazioa. Eta, egia esan, garrantzitsua da maiztasun-banaketaren itxura ezagutzea, ostean, banaketa normalaren diagramarekin (Gauss kanpaiarekin) alderatzeko. Izan ere, datu multzo batek banaketa normalaren antzerako zorroztasuna eta simetria baditu, maiztasun-banaketa hori jatorrizko datu multzoaren eredu teoriko gisa hartu ahal izango dugu. Eta horrek izugarri errazten du lan matematiko-estatistikoa.

Bi dira maiztasun-banaketen itxura aztertzen duten neurri estatistikoak: kurtosia eta alborapena (asimetria).

2.3.4.1. Alborapena

Esan bezala, alborapen-neurriek datu multzo baten itxurari buruzko informazioa ematen digute; hau da, maiztasun-banaketa simetrikoa den edo eskuinera edo ezkerrera alboratuta dagoen.

Histogramari edo maiztasun bakunen poligonoari begiratuta, maiztasun-banaketa simetrikoa edo asimetrikoa den (eskuinera edo ezkerrera alboratua dagoen) nahiko erraz ikus dezakegu.

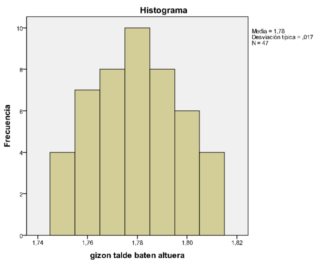
Kanpai-itxurako histograma simetrikoetan, batez bestekoaren eta medianaren ezkerretara eta eskuinetara, maiztasun eta tarte kopuru bera (edo antzerakoa) dago. Kontuan izan, horrelakoetan, mediana eta batez bestekoa berdinak edo oso antzerakoak izaten direla: Me = $\bar{x}$.

Ezkerrera alboratua dauden kanpai-itxurako banaketetan (alborapen negatiboa), batez bestekotik ezkerretara dagoen muturra luzeagoa da. Eta horrek esan nahi du batez bestekotik ezkerretara tarte kopuru handiagoa eta maiztasun handiagoa dagoela; hau da, ezkerreko buztana luzeagoa dela. Horrelakoetan, batez bestekoa mediana baino txikiagoa izaten da: $\bar{x}$ < Me.

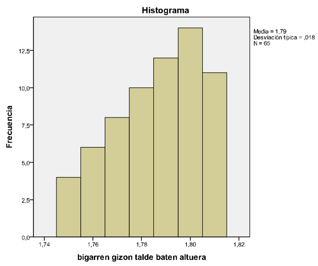
Eskuinera alboratutako kanpai-itxurako maiztasun-banaketetan (alborapen positiboa), eskuineko muturra luzeagoa da; alegia, batez bestekotik eskuinera tarte kopuru eta maiztasun handiagoa dago. Kasu horretan, batez bestekoa mediana baino handiagoa izaten da: Me < $\bar{x}$

Moda bakarreko maiztasun-banaketa bada, kanpaiaren itxurakoa eta alborapenik ez badu, alegia, maiztasun-banaketa simetrikoa bada, moda, mediana eta batez bestekoa berdinak dira: Mo = Me = $\overline{x}$. Kasu horretan, datu kopuru bera dago batez besteko aritmetiko sinpletik behera zein gora. Aldagaiaren beheko eta goiko balioak eta maiztasunak orekatuta daude.
Moda bakarreko maiztasun-banaketa bada, kanpai-itxurakoa eta alborapen negatiboa (ezkerrera alboratua) badu, batez bestekoa, mediana baino txikiagoa da eta azken hori, moda baino txikiagoa: $\overline{x}$ < Me < Mo. Kasu horretan, datu gehiago dago batez besteko aritmetiko sinpletik gora behera baino, edota, bestela esanda, moda eta mediana batez besteko aritmetiko sinplea baino handiagoak dira. Aldagaiaren goiko balioek dute maiztasun handiena, eta, beraz, histogramaren ezkerreko muturra luzeagoa da.
Moda bakarreko maiztasun-banaketa, kanpai-itxurakoa eta alborapen positiboa badu (eskuinera alboratua), moda mediana baino txikiagoa da, eta mediana batez bestekoa baino txikiagoa: Mo < Me < $\overline{x}$. Kasu horretan, datu gehiago dago batez besteko aritmetiko sinpletik behera gora baino. Aldagaiaren beheko balioek dute dira maiztasun handiena, eta, horren ondorioz, histogramaren eskuineko muturra luzeagoa da.

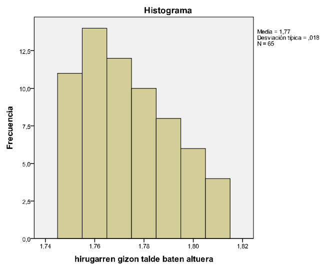
28. diagrama: maiztasun-banaketa alboratugabeen eta ezkerretara edo eskuinera alboratuen histogramak



Maiztasun-banaketa simetrikoa



Ezkerretara alboratutako maiztasun-banaketa



Eskuinera alboratutako maiztasun-banaketa

Iturria: lanketa propioa

2.3.4.1.1. Karl Pearsonen asimetria-koefizientea eta Fisher-en eta Bowley-ren alborapen-koefizientea

Histogramei begiratuta, nahiko erraz jakin dezakegu maiztasun-banaketa simetrikoa den ala alborapenik duen. Dena dela, badaude histogramak egin gabe ere alborapenari buruzko informazio ematen diguten neurri estatistikoak. Guk hiru aurkeztuko ditugu:

- Karl Pearsonen asimetria-koefizientea
- Fisher-en alborapen-koefizientea
- Bowley-en alborapen-koefizientea

Neurri horien laguntzarekin, maiztasun-banaketa irudikatu gabe ere, maiztasun-banaketa nolakoa den jakin dezakegu.

Karl Pearsonen asimetria-koefizientea

Karl Pearsonek alborapena neurtzeko, honako formula hauek eman zituen:

$$A_{K1} = \frac{\overline{x} - Mo}{s_x}$$

$$A_{K2} = \frac{3(\overline{x} - Me)}{s_{\overline{x}}}$$

Honela interpretatzen ditugu emaitzak:

- A_{k1} eta A_{k2} = 0 edo 0ra hurbiltzen badira, maiztasun-banaketa alboratugabea edo simetrikoa dela esango dugu.
- A_{k1} eta $A_{k2} > 0$ badira, alborapena eskuin alderakoa edo positiboa dela esango dugu.
- A_{k1} eta $A_{k2} < 0$ badira, alborapena ezker alderakoa edo negatiboa dela esango dugu.

Karl Pearsonen asimetria-koefizientearen abantailak eta desabantailak

Pearsonen koefiziente erraz kalkulatzen da, baina ezin da erabili maiztasun-banaketa moda bakarrekoa ez denean, edota histogramak kanpai-formarik ez duenean.

Fisher-en alborapen-koefizientea

Fisherren alborapen-koefizientea era honetan kalkulatzen da:

$$g_1 = \frac{1}{N} \cdot \frac{\sum_{i=1}^{n} (x_i - \overline{x})^3 \cdot f_i}{\sigma^3}$$

Honela interpretatuko ditugu emaitzak:

- $g_1 = 0$ bada edo 0ra asko hurbiltzen bada, maiztasun-banaketa simetrikoa edo ia simetrikoa dela esango dugu.
- $g_1 > 0$ bada, maiztasun-banaketa eskuin aldera alboratuta dagoela edo alborapen positiboa duela esango dugu.
- $g_1 < 0$ bada, maiztasun-banaketa ezkerrera alboratua dagoela edo alborapen negatiboa duela esango dugu.

Fisher-en alborapen-koefizientearen abantailak eta desabantailak

Pearsonen alborapen-koefizientea baino maizago erabiltzen da. Hala ere, kontuan izan behar dugu ez dela neurri jasankorra, batez bestekoaren bidez muturreko datuek eragina dutelako emaitzan. Datu guztiak erabiltzen direla hura kalkulatzean da abantaila.

Bowley-ren alborapen-koefizientea

Bowleyren alborapen-koefizientea modu honetan kalkulatzen da:

$$A_B = \frac{(Q_3 - Me) - (Me - Q_1)}{Q_3 - Q_1}$$

Honela interpretatuko ditugu emaitzak:

- $A_B = 0$ bada edo 0ra hurbiltzen bada, maiztasun-banaketa simetrikoa edo ia simetrikoa dela esango dugu.
- $A_B > 0$ bada, maiztasun-banaketa eskuin aldera alboratuta dagoela edo alborapen positiboa duela esango dugu;

- $A_B < 0$ bada, maiztasun-banaketa ezkerrera alboratua dagoela edo alborapen negatiboa duela esango dugu.

Abantailak eta desabantailak

Koefiziente hori jasankorra da. Izan ere, muturreko datuek edo datu arraroek ez dute emaitzan eragiten. Desabantaila da lagineko datu guztiak ez dituela kontuan hartzen.

Adibidea. Maiztasun-banaketa simetrikoa: alborapena kalkulatzea IBM SPSS programaren laguntzarekin: A gizon taldearen altuera.

64. taula: A gizon taldearen altuerak

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1,75	4	8,5	8,5	8,5
	1,76	7	14,9	14,9	23,4
	1,77	8	17,0	17,0	40,4
	1,78	10	21,3	21,3	61,7
	1,79	8	17,0	17,0	78,7
	1,80	6	12,8	12,8	91,5
	1,81	4	8,5	8,5	100,0
	Total	47	100,0	100,0	

Iturria: lanketa propioa

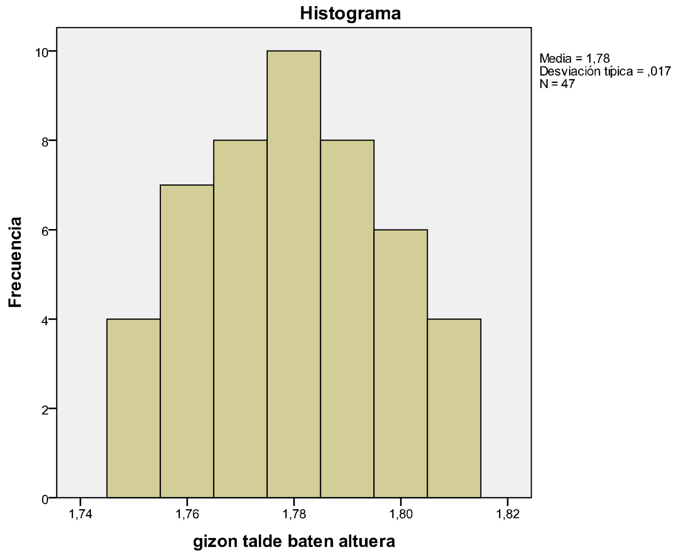
Maiztasun-banaketa horretatik abiatuta, eta IBM SPSS programaren laguntzarekin, neurri estatistiko hauek lortzen ditugu:

Estadísticos
A gizon taldearen altuera

N	Válidos	47
	Perdidos	0
Media		1,7796
Mediana		1,7800
Moda		1,78
Desv. típ.		,01744
Varianza		,000
Asimetría		,042
Error típ. de asimetría		,347
Curtosis		-,871
Error típ. de curtosis		,681
Percentiles	25	1,7700
	50	1,7800
	75	1,7900

Maiztasun-banaketa hori simetrikoa, moda bakarrekoa eta kanpai-itxurakoa izanik, moda, mediana eta batez bestekoa oso antzerakoak dira. Kasu honetan, datu kopuru bera dago batez besteko aritmetiko sinpletik behera zein gora. Alegia, aldagaiaren beheko eta goiko balioak eta maiztasunak orekatuta daude. Bestalde, asimetria gutxiko eta moda bakarreko maiztasun-banaketa izanik, honako erlazio hau betetzen da: $\bar{x}$ – mo = 3 . ($\bar{x}$ – me) → 1,7796 – 1,78 = 3 . (1,7796-1,78 → -0,0004 = -0,0012. Gainera, esan behar da asimetria-koefizientea 0,042 dela; hau da, 0tik oso hurbil dagoela. Azkenik, histogramari begiratzen badiogu, argi ikusten dugu maiztasun-banaketa simetrikoa dela.

29. diagrama: A gizon taldearen altuerak



Iturria: lanketa propioa

Adibidea. Ezkerretara alboratua dagoen maiztasun-banaketa (alborapen negatiboa): IBM SPSS programarekin egindako kalkuluak: B gizon taldearen altuera.

65. taula: B gizon taldearen altuerak					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1,75	4	6,2	6,2	6,2
	1,76	6	9,2	9,2	15,4
	1,77	8	12,3	12,3	27,7
	1,78	10	15,4	15,4	43,1
	1,79	12	18,5	18,5	61,5
	1,80	14	21,5	21,5	83,1
	1,81	11	16,9	16,9	100,0
	Total	65	100,0	100,0	
Iturria: lanketa propioa					

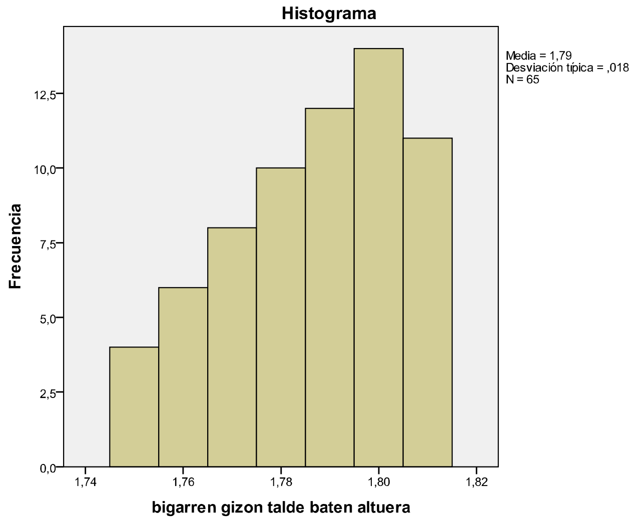
Maiztasun-banaketa horretatik abiatuta, eta IBM SPSS programaren laguntzarekin, honako neurri estatistiko hauek lortzen ditugu:

Estadísticos
B gizon taldearen altuera

N	Válidos	65
	Perdidos	0
Media		1,7863
Mediana		1,7900
Moda		1,80
Desv. típ.		,01808
Varianza		,000
Asimetría		-,423
Error típ. de asimetría		,297
Curtosis		-,837
Error típ. de curtosis		,586
Percentiles	25	1,7700
	50	1,7900
	75	1,8000

Histograma aztertzen badugu, argi ikusten dugu maiztasun-banaketa hori ezkerretara alboratuta dagoela. Izan ere, batez bestekoa, mediana baino txikiagoa da eta azken hori, moda baino txikiagoa: $\bar{x}$ < Me < Mo. Beraz, datu gehiago dago batez besteko aritmetiko sinpletik gora behera baino. Aldagaiaren goiko balioak dira maiztasun handiena dutenak, eta, horren ondorioz, histogramaren ezkerreko muturra luzeagoa da. Gainera, asimetria-koefizientea negatiboa dela ikusten dugu (-0,423).

30. diagrama: B gizon taldearen altuerak



Iturria: lanketa propioa

Adibidea. Eskuinera alboratutako maiztasun-banaketa: alborapena kalkulatzea IBM SPSS programarekin egin da: C gizon taldearen altuera.

66. taula: C gizon taldearen altuerak

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1,75	11	16,9	16,9	16,9
	1,76	14	21,5	21,5	38,5
	1,77	12	18,5	18,5	56,9
	1,78	10	15,4	15,4	72,3
	1,79	8	12,3	12,3	84,6
	1,80	6	9,2	9,2	93,8
	1,81	4	6,2	6,2	100,0
	Total	65	100,0	100,0	

Iturria: lanketa propioa

Maiztasun-banaketa horretatik abiatuta, eta IBM SPSS programaren laguntzarekin, honako neurri estatistiko hauek lortzen ditugu:

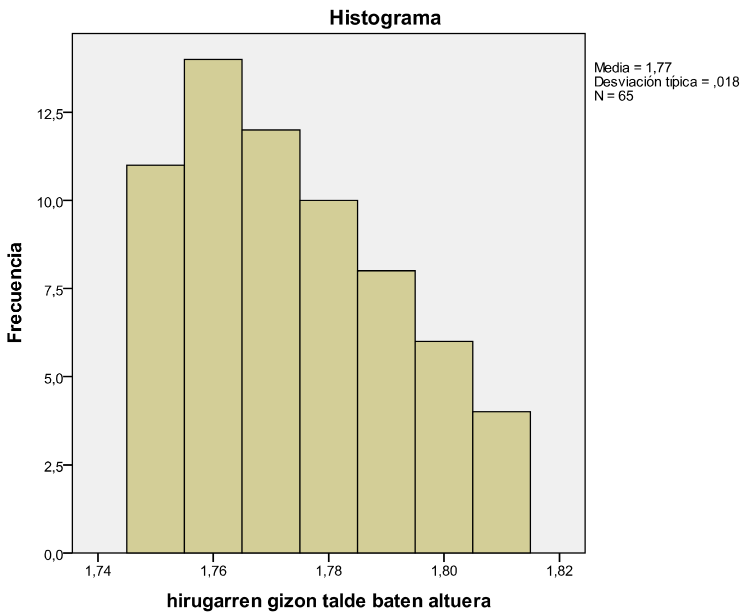
Estadísticos
C gizon taldearen altuera

N	Válidos	65
	Perdidos	0
Media		1,7737
Mediana		1,7700
Moda		1,76
Desv. típ.		,01808
Varianza		,000

Asimetría		,423
Error típ. de asimetría		,297
Curtosis		-,837
Error típ. de curtosis		,586
Percentiles	25	1,7600
	50	1,7700
	75	1,7900

Kasu honetan, moda mediana baino txikiagoa da, eta azken hori batez bestekoa baino txikiagoa: Mo < Me < $\bar{x}$. Beraz, eskuinera alboratutako maiztasun-banaketa dela esan behar da. Kasu honetan, datu gehiago dago batez besteko aritmetiko sinpletik behera gora baino. Aldagaiaren beheko balioek dute maiztasun handiena, eta, horren ondorioz, histogramaren eskuineko muturra luzeagoa da. Gainera, asimetria-koefizientea positiboa dela ikusten dugu (0,423).

31. diagrama: C gizon taldearen altuerak



Iturria: lanketa propioa

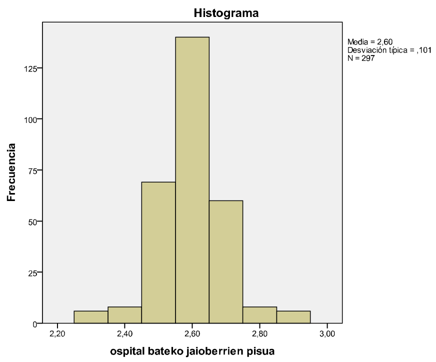
2.3.4.1.2. Alborapenari buruzko ondorio orokorrak

Hasieran esan bezala, alborapen-neurriak, ostean ikusiko ditugun kurtosi-neurriekin batera, datu multzo baterako banaketa normala eredutzat har daitekeen erabakitzeko erabiltzen dira. Datu multzoak asimetria edo alborapenik eza erakusten badu (eta ondoren ikusiko dugun bezala, kurtosi maila ertainekoa edo mesokurtikoa bada), banaketa normala onartu ahal izango dugu datu multzoa aztertzeko. Eta horrek lan matematiko-estatistikoa izugarri erraztuko digu.
Kontuan izan behar da, dena dela, guztiz simetrikoak diren datu multzo gutxi dagoela. Denek dute alborapenen bat. Hala, alborapen gutxiko maiztasun-banaketak normaltzat hartuko ditugu.

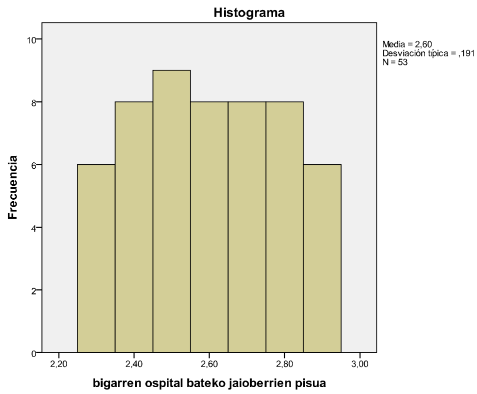
2.3.4.2. Kurtosia

Batez besteko eta bariantza berdinak dituzten bi maiztasun-banaketa simetriko forma edo itxuraz desberdinak izan daitezkeenez, garrantzitsua da kurtosia aztertzea; hau da, batez bestekoaren inguruan datuak nola biltzen diren ezagutzea.
Kurtosia maiztasun-banaketa baten zorroztasun maila adierazten duen forma-neurria da.
Histograma batean burua, sorbaldak eta besoak bereizten baditugu, kurtosi handiko banaketetan, maiztasuna sorbaldetatik beso eta buruetara mugitzen da. Alegia, kurtosi handiko banaketetan, datu kopuru handia biltzen da batez bestekoaren inguruan (maiztasun-banaketa zorrotzagoa da batez bestekoaren inguruan), eta, gainera, batez bestekotik urruntzen diren datu batzuk ere badaude besoetan. Kurtosi txikia duten banaketetan, aldiz, datuak ez dira batez bestekoaren inguruan biltzen; hau da, maiztasun-banaketa ez da zorrotza batez bestekoaren inguruan; eta, gainera, besoetan datu gutxi dago. Laburbilduz: kurtosi handia duten maiztasun-banaketek erdigune puntazorrotza dute eta beso luzeagoak. Kurtosi txikiagokoek, berriz, zentro zapala dute eta beso laburrak.

32. diagrama: kurtosi-maila desberdineko maiztasun-banaketak



Kurtosi handiko maiztasun-banaketa: batez bestekoaren inguruan zorrotza delako, eta beso luzeagoak dituelako.



Kurtosi txikiagoko maiztasun-banaketa: zentroan zapalagoa delako, eta beso motzagoak eta finagoak dituelako.

Iturria: lanketa propioa

Kurtosiak datu multzoaren sakabanatzeari buruzko informazioa ematen digu. Zenbat eta zorroztasun handiagoa izan, orduan eta sakabanatze txikiagoa dago.
Horrez gain, kurtosiaren bidez, gure datu multzoak banaketa normalaren zorroztasun maila duen ala ez jakin dezakegu. Dena dela, ezin gara irudiarekin bakarrik fidatu, hain zuzen ere, begiek baliorik gabeko ondorioetara eraman gaitzatelako. Hala, histogramari begiratzeaz gain, kurtosia kalkulatzeko neurriak ere ezagutu behar ditugu.

2.3.4.2.1. Fisherren kurtosi-neurria

Kurtosi-neurriak asko dira. Ezagunena Fisherren kurtosi-neurria da, eta formula honen bidez kalkulatzen da:

$$\alpha = \frac{\sum f(x_i - \bar{x})^4}{n\sigma^4}$$

Non,

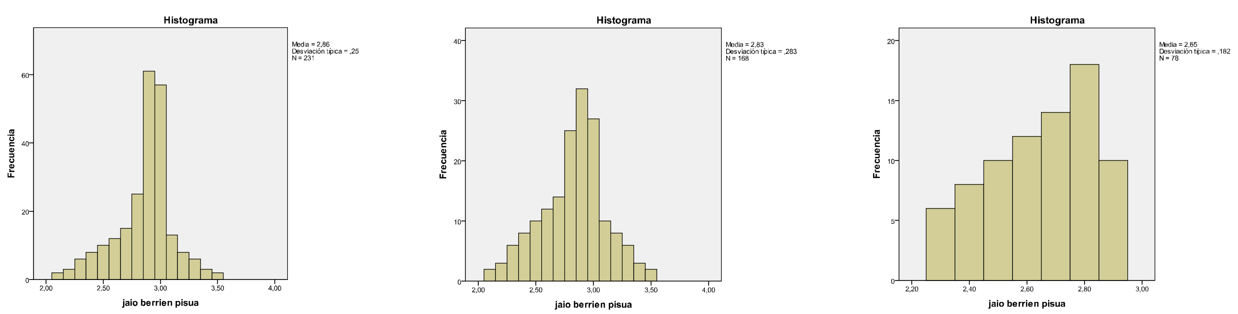
x_i= aldagaiaren balio bakoitza baita;
n = datu kopurua;
$\overline{x}$ = batez bestekoa;
σ = desbideratze tipikoa;
f_i = balio bakoitzaren maiztasun absolutua.
α zenbat eta handiagoa izan, kurtosia handiagoa izango da.
Banaketa normala datu multzo idealizatua da, eta, estatistikan, askotan erabiltzen da. Izan ere, errealitateko datu multzo asko banaketa horretara hurbiltzen dira. Kurtosiari dagokionez, banaketa normalak zorroztasun maila ezaguna eta finkoa du (α = 3), eta, horregatik, eredutzat erabil daiteke zorroztasunak sailkatzeko. Hau da, banaketa normala oinarritzat hartuta, aztertzen ari garen datu multzoak banaketa normalak baino zorroztasun handiagoa, txikiagoa edo berdina duen jakin dezakegu. Hain zuzen ere, Fisherren formulatik ateratzen den emaitza (α) 3 baino handiagoa, txikiagoa edo berdina den ikusita, aztertzen ari garen maiztasun-banaketa banaketa normalera hurbiltzen den jakin dezakegu.
Modu berean, Fisherren formulatik ateratzen den emaitzari 3 kentzen badiogu, alegia, normal erako maiztasun-banaketa baten kurtosia kentzen badiogu, banaketa normalen kurtosia = 0 da :

$$g_2 = \frac{\sum (x_i - \overline{X})^4 f_i}{ns^4} - 3$$

Hala, α–k eta g_2-k ematen dituzten emaitzen arabera, hiru eratako maiztasun-banaketak ditugu kurtosiaren ikuspuntutik:

- α > 3 eta g_2 > 0 betetzen bada, maiztasun-banaketa leptokurtikoa da.Hau da, g_2 koefizienteak balio positiboa hartzen badu, edo α 3 baino handiagoa bada, kurtosi-maila altua duela, edo banaketa normala baino zorrotzagoa dela esaten da.
- α < 3 eta g_2 < 0 betetzen bada, maiztasun-banaketa platikurtikoa da. Beraz, g_2 kurtosi-koefizienteak balio negatiboa hartzen badu, edo α hiru baino txikiagoa bada, banaketa normala baino zapalagoa da.
- α = 3 eta g_2 = 0 betetzen bada, maiztasun-banaketa mesokurtikoa, kanpai-itxurakoa edo normala dela esaten da. Kasu honetan, maiztasun-banaketak kurtosi-maila ertaina du; hau da: banaketa normalaren kurtosi-maila berdina. Kurtosia = 0 izatea oso zaila denez, 0ra hurbiltzen diren kurtosia duten maiztasun-banaketak mesokurtikotzat hartzen dira. Zehatzago esanda, gutxi gorabehera ± 0.5 tartean dauden kurtosiak mesokurtikotzat hartuko ditugu.

33. diagrama: kurtosi-maila desberdineko maiztasun-banaketak



Leptokurtikoa: banaketa normalak baino zorroztasun edo kurtosi handiagoa.

Mesokurtikoa: banaketa normalaren zorroztasun edo kurtosi berdina.

Platikurtikoa: banaketa normalak baino zorroztasun edo kurtosi txikiagoa.

Iturria: lanketa propioa

Inolako azalpenik eman aurretik, kontuan izan behar da IBM SPSSak g_2 kalkulatzen duela. Horrenbestez, IBM SPSSaren bidez lortzen dugun kurtosi-koefizientea 0 bada, edo 0tik oso hurbil badago, maiztasun-banaketa mesokurtikoa dela esango dugu; 0 baino txikiagoa bada, platikurtikoa dela, eta 0 baino handiagoa bada, leptokurtikoa.

Gogoratu, hala ere, maiztasun-banaketek beste forma batzuk ere izan ditzaketela kurtosiaren ikuspuntutik. Gehienetan, hartzen duten formaren araberako izenak hartzen dituzte. Adibidez, J formako maiztasun-banaketa, L formako maiztasun-banaketa, laukizuzen formako maiztasun-banaketa, U formako maiztasun-banaketa.

Adibidea. Kurtosia kalkulatzea Fisherren formularen bidez: demagun datu multzo hau dugula: 6, 6, 9, 9, 9, 12, 12, 12, 15 eta 17. Kurtosia kalkulatzeko, honako pauso hauek ematen ditugu:

- Batez bestekoa kalkulatu ($\overline{x}$) = 10,70
- Desbideratze tipikoa kalkulatu (s) = 3,592
- Formula aplikatu

$$g_2 = \frac{\sum (x_i - \overline{X})^4 f_i}{ns^4} - 3$$

g_2 = - 0,501

Kurtosia –0,501 izanik, maiztasun-banaketa platikurtikoa edo banaketa normala baino zapalagoa dela esan behar dugu.

Adibidea. Kurtosia kalkulatzea Fisherren formularen bidez: demagun ibai batek egunero izan duen altuera neurtu dugula 60 egunez, eta honelako maiztasun-banaketa lortu dugula.

67. taula: Ibaiaren altuera	
Altuera	Egun kopurua
1,20	20
1,21	4
1,22	4
1,23	4
1,24	4
1,25	4
1,26	4
1,27	4
1,28	4
1,29	4
1,30	4
Total	60
Iturria: lanketa propioa	

Batez bestekoa ($\overline{x}$) = 1,2367
Desbideratze tipikoa (s) = 0,3526
Fisherren formula aplikatuta, erraz lortzen da kurtosi-maila:

$$g_2 = \frac{\sum (x_i - \overline{X})^4 f_i}{ns^4} - 3$$

g_2 = - 1,269

Kurtosia –1,269 izanik, maiztasun-banaketa platikurtikoa edo banaketa normala baino zapalagoa dela esan behar dugu. Datu gehienak ez dira batez bestekoaren inguruan biltzen.

Adibidea. Kurtosia kalkulatzea IBM SPSS programaren bidez: maiztasun-banaketa leptokurtikoa: A ospitalean jaiotako umeen pisua.

68. taula: A ospitalean jaiotakoen pisua

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	2,10	2	,9	,9	,9
	2,20	3	1,3	1,3	2,2
	2,30	6	2,6	2,6	4,8
	2,40	8	3,5	3,5	8,2
	2,50	10	4,3	4,3	12,6
	2,60	12	5,2	5,2	17,7
	2,70	15	6,5	6,5	24,2
	2,80	25	10,8	10,8	35,1
	2,90	61	26,4	26,4	61,5
	3,00	57	24,7	24,7	86,1
	3,10	13	5,6	5,6	91,8
	3,20	8	3,5	3,5	95,2
	3,30	6	2,6	2,6	97,8
	3,40	3	1,3	1,3	99,1
	3,50	2	,9	,9	100,0
	Total	231	100,0	100,0	

Iturria: lanketa propioa

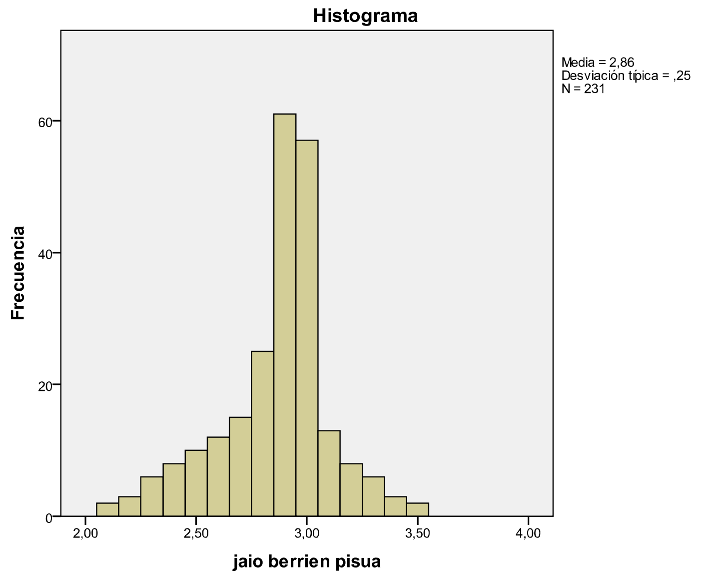
Maiztasun-banaketa horretatik abiatuta, eta IBM SPSS programaren laguntzarekin, honako neurri estatistiko hauek lortzen ditugu:

Estadísticos
jaioberrien pisua

N	Válidos	231
	Perdidos	0
Media		2,8628
Mediana		2,9000
Moda		2,90
Desv. típ.		,24987
Varianza		,062
Asimetría		-,639
Error típ. de asimetría		,160
Curtosis		,887
Error típ. de curtosis		,319
Percentiles	25	2,8000
	50	2,9000
	75	3,0000

Estatistikoen zerrenda horretan, kurtosia agertzen zaigu. Hain zuzen ere, kasu honetan, kurtosia 0,887 dela ikusten dugu. 0 baino handiagoa denez, maiztasun-banaketa leptokurtikoa dela ondorioztatu behar dugu; hau da, maiztasun-banaketa hori banaketa normala baino zorrotzagoa da. Egia esanda, histogramari begiratuta erraz ikusten da maiztasun-banaketa leptokurtiko baten aurrean daudela.

34. diagrama: maiztasun-banaketa leptokurtikoa



Iturria: lanketa propioa

Adibidea. Kurtosia kalkulatzea IBM SPSS programaren bidez: maiztasun-banaketa mesokurtikoa: B ospitalean jaiotakoen pisua.

69. taula: B ospitalean jaiotakoen pisua

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	2,10	2	1,2	1,2	1,2
	2,20	3	1,7	1,7	2,9
	2,30	6	3,5	3,5	6,4
	2,40	8	4,7	4,7	11,0
	2,50	10	5,8	5,8	16,9
	2,60	12	7,0	7,0	23,8
	2,70	15	8,7	8,7	32,6
	2,80	25	14,5	14,5	47,1
	2,90	32	18,6	18,6	65,7
	3,00	27	15,7	15,7	81,4
	3,10	13	7,6	7,6	89,0
	3,20	8	4,7	4,7	93,6
	3,30	6	3,5	3,5	97,1
	3,40	3	1,7	1,7	98,8
	3,50	2	1,2	1,2	100,0
	Total	172	100,0	100,0	

Iturria: lanketa propioa

Maiztasun-banaketa horretatik abiatuta, eta IBM SPSS programaren laguntzarekin, honako neurri estatistiko hauek lortzen ditugu:

Estadísticos

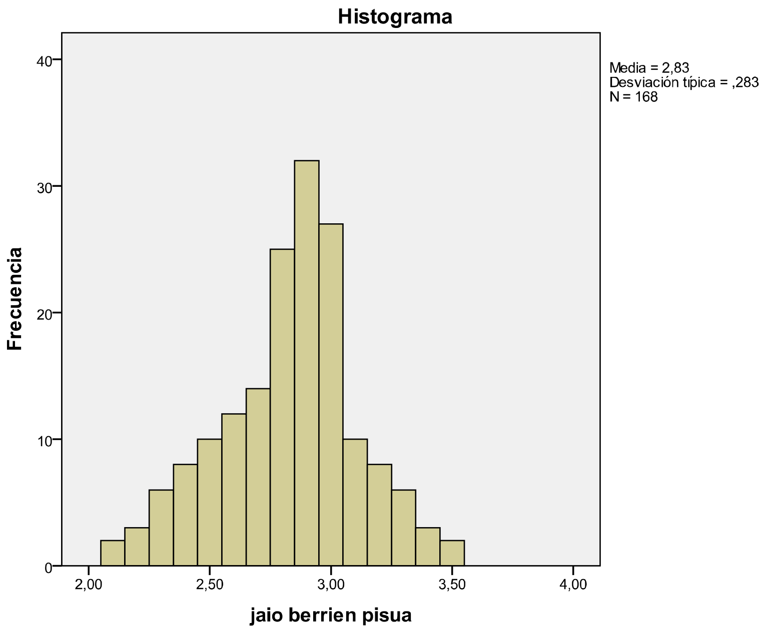
jaioberrien pisua

N	Válidos	168
	Perdidos	0
Media		2,8286

Mediana		2,9000
Moda		2,90
Desv. típ.		,28287
Varianza		,080
Asimetría		-,271
Error típ. de asimetría		,187
Curtosis		,025
Error típ. de curtosis		,373
Percentiles	25	2,7000
	50	2,9000
	75	3,0000

Bigarren adibide honetan, kurtosia 0,025 da. Ikusten den bezala, kurtosia asko hurbiltzen da zerora. Hala, maiztasun-banaketa mesokurtikoa dela esan behar dugu; hau da, banaketa normalaren kurtosi-maila berdina duela.

35. diagrama: maiztasun-banaketa mesokurtikoa



Iturria: lanketa propioa

Adibidea. Kurtosia kalkulatzea IBM SPSS programaren bidez: maiztasun-banaketa platikurtikoa: C ospitalean jaiotakoen pisua.

70. taula: C ospitalean jaiotakoen pisua

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	2,30	6	7,7	7,7	7,7
	2,40	8	10,3	10,3	17,9
	2,50	10	12,8	12,8	30,8
	2,60	12	15,4	15,4	46,2
	2,70	14	17,9	17,9	64,1
	2,80	18	23,1	23,1	87,2
	2,90	10	12,8	12,8	100,0
	Total	78	100,0	100,0	

Iturria: lanketa propioa

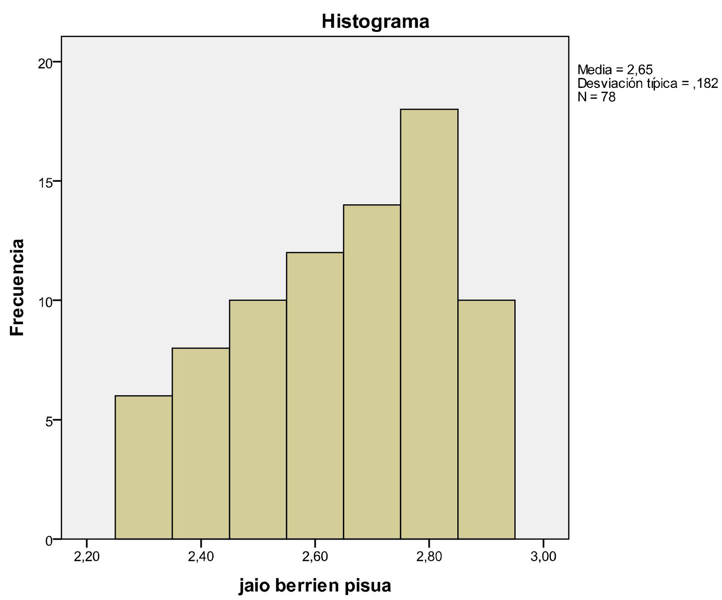
Maiztasun-banaketa horretatik abiatuta, eta IBM SPSS programaren laguntzarekin, honako neurri estatistiko hauek lortzen ditugu:

Estadísticos
jaioberrien pisua

N	Válidos	78
	Perdidos	0
Media		2,6462
Mediana		2,7000
Moda		2,80
Desv. típ.		,18212
Varianza		,033
Asimetría		-,374
Error típ. de asimetría		,272
Curtosis		-,927
Error típ. de curtosis		,538
Percentiles	25	2,5000
	50	2,7000
	75	2,8000

Kurtosi-koefizienteak, hirugarren kasu horretan, balio negatiboa hartzen du (-0,927). Beraz, maiztasun-banaketa platikurtikoa edo banaketa normala baino zapalagoa dela esan behar dugu.

36. diagrama: maiztasun-banaketa platikurtikoa



Iturria: lanketa propioa

2.3.4.2.2. Kurtosiaren abantailak eta desabantailak

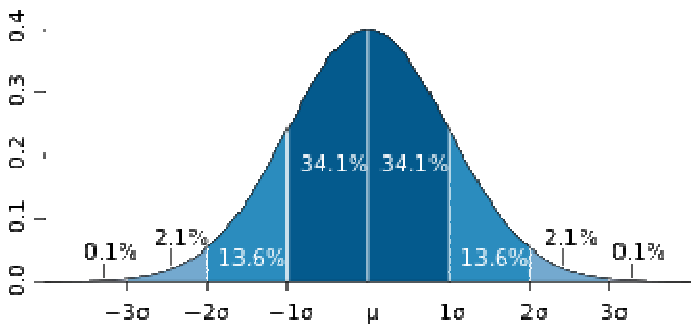
- Kurtosiaren abantailei dagokionez, aipatzekoa da kurtosiak datu multzoaren sakabanatzeari buruzko informazioa ematen digula. Zenbat eta zorroztasun handiagoa izan, orduan eta sakabanatze txikiagoa egongo da. Bestalde, kurtosiak datu multzo baterako eredu normala egokia den adierazten digu. Hain zuzen ere, banaketa mesokurtikoa dela ikusten badugu, banaketa normal bat izan genezake.

- Beste abantaila da: hura kalkulatzean datu guztiak erabiltzen direla.
- Estatistiko horren desabantaila: ez da jasankorra; alegia, muturreko datuek eragin handia dute kurtosian.

2.3.4.3. Alborapenari eta kurtosiari buruzko ondorio orokorrak

Maiztasun-banaketen alborapena eta kurtosia ezagutzea ezinbestekoa da gure datu multzoa banaketa normalera hurbiltzen den jakiteko. Izan ere, gure maiztasun-banaketak banaketa normalaren antzeko kurtosi maila eta gutxi gorabeherako simetria baditu, jatorrizko datu multzoaren (datu multzo enpirikoaren) eredu edo ordezkari gisa hartu ahal izango dugu. Eta hori oso komenigarria da ikerketa estatistikoan, banaketa normala eredutzat harturik, lan matematiko-estatistikoa asko errazten delako. Sakonago aztertuko dugu ondorengo orrietan, baina kontuan izan behar dugu banaketa normalaren abantailetariko bat dela datuen ehuneko zehatzak (%68, %95, %99…) batez bestekotik bat, bi edo hiru desbideratze estandarreko distantziaren barnean daudela; hau da, desbideratze tipikoa bider bat, bider bi edo bider hiru eginda batez bestekoari gehitzen eta kentzen badiogu, datuen ehuneko zehatzak (%68, %95, %99…) lortu ditugun tarte horietan egongo dira.

37. diagrama: maiztasun-banaketa normalaren probabilitateak



Iturria: https://commons.wikimedia.org/wiki/File:Standard_deviation_diagram.svg

2.3.5. Kontzentrazio-neurriak: Lorenz-en kurba eta Gini-ren indizea

Kontzentrazio-neurriek aldagaia (errenta, soldatak, salmentak…) aztergai den populazioaren barruan nola banatzen den neurtzen dute. Hala, populazioaren zati txiki bat aldagaiaren zati handienaren jabe bada, aldagaia oso kontzentratuta dagoela esaten dugu. Aldagaia populazioan modu orekatuan banatuta badago, berriz, maiztasun-banaketa homogeneoa dela esango dugu, edo kontzentrazio txikia dagoela. Egiatan, aldagai baten balioen maiztasun-banaketak izan ditzakeen egoera infinituetatik, bi daude muturretan:

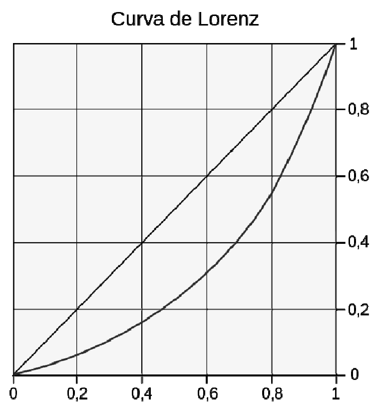
- Gehienezko kontzentrazioa: elementu, erakunde edo subjektu bakar batek guztia eskuratzen duenean eta besteek ezer ere ez dutenean. Kasu horretan, maiztasun-banaketa guztiz desorekatua dugu.
- Gutxieneko kontzentrazioa: aldagaia modu berean banatuta dagoenean populazioa osatzen duten elementu, erakunde edo subjektuen artean. Kasu horretan, maiztasun-banaketa orekatua dela esango dugu.

Kontzentrazio-neurri asko daude, baina bi dira ezagunenak: Lorenzen Kurba eta Giniren koefizientea

2.3.5.1. Lorenzen kurba

Lorenz kurbak errenta nola banatzen den adierazten digu; hau da, populazioa osatzen duten gizakien ehuneko desberdinek eskuratzen dituzten errenta osoaren ehunekoak irudikatzeko. Adibidez, herrialde bateko biztanlearen %80k errenta osoaren %20 hartzen duela esaten denean, Lorenzen kurbako puntu bat adierazten da.
Lorenz kurba errenta nola banatzen den adierazteko erabiltzen da, baina beste aldagai sozioekonomiko batzuen kontzentrazio maila ere azter daiteke. Adibidez, ekoizpenen, salmenten edo fakturazioen kontzentrazioa sektoreen arabera, lurraren kontzentrazioa herrialde batean edo langileen kontzentrazioa enpresen artean.

38. diagrama: Lorenzen kurba

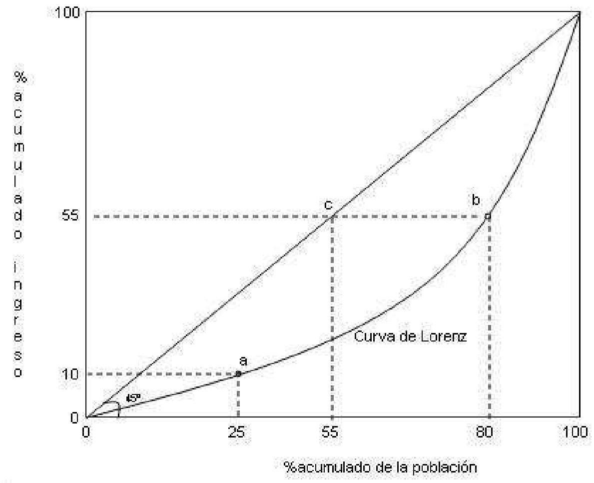


Iturria: http://commons.wikimedia.org/wiki/File:Lorenz_Curve_-_es.svg

Errenta aztertzen ari bagara, Lorenz kurbaren gaineko diagonalak egoera ideal bat irudikatzen du, non gizaki guztiek errenta edo diru-sarrera berdinak baitzituzten. Diagonalak egoera ideala adierazten du: biztanleriaren %40 ondasunen %40ren jabe litzateke, eta biztanleriaren %80, ondasunen %80rena. Kurbak adierazten du errealitatean errenta populazioan nola banatzen den. Hala, kurba zenbat eta urrunago egon diagonaletik, errentaren kontzentrazio maila altuagoa da.

Adibidea. Errentaren banaketa Lorenzen kurbaren arabera

39. diagrama: errentaren banaketa Lorenzen kurbaren arabera



Lorenzen kurba guztietan bezala, adibide honetan ere, diagonalak maiztasun-banaketa guztiz orekatua adierazten du; kurbak, berriz, errenta egiatan nola banatzen den. Adibidez, a puntuan ikusten dugu errenta gutxien duen biztanleriaren %25ek errentaren gutxi gorabehera %10 lortzen duela, eta b puntuan ikusten dugu biztanleriaren %80k errentaren %55 eskuratzen duela. Kasu honetan, errenta txarto banatuta dagoela ondoriozta dezakegu.

2.3.5.2. Giniren koefizientea

Giniren koefizienteak errentaren maiztasun-banaketa eta bestelako aldagaien kontzentrazio-maila neurtzen du. Kontzentrazioa aztertzeko erabiltzen den Lorenzen kurban oinarriturik, kontzentrazioaren neurketa absolutua ematen du, 0tik 1era. Hau da, kontzentrazio-maila txikia duen egoera batetik (Giniren koefizientea = 0) kontzentrazioa erabatekoa duen egoera batera (Giniren koefizientea = 1) neurtzen du. Errenta nola banatzen den aztertzeaz gainera, beste arlo askotan erabiltzen da, betiere kontzentrazioa aztertzeko. Koefiziente hori bider 100 egiten badugu, Giniren indizea lortzen dugu. Giniren koefizientea honako formula honen bidez kalkulatzen da:

$$C_G = \frac{\sum_{i=1}^{n-1}(p_i - q_i)}{\sum_{i=1}^{n-1} p_i}$$

Non,

$$p_i = \frac{F_i}{N}.100$$

F_i = populazioaren maiztasun metatua.

$$\mu_i = x_i . f_i$$
$$q_i = \frac{u_i(metatuak)}{\sum u_i}.100$$

Adibidea. Giniren koefizientea: enpresa bateko 40 langileren soldaten kontzentrazio-maila neurtzea.

71. taula: 40 langileren soldaten kontzentrazio-maila neurtzea.

Soldatak eurotan (x_i)	Langile kopurua (f_i)	F_i	$p_i = \frac{F_i}{N}.100$	$\mu_i = x_i . f_i$	u_i metatuak	$q_i = \frac{u_i(metatuak)}{\sum u_i}.100$
600	20	20	50	12000	12000	29'63
1000	15	35	87'5	15000	27000	66'67
2500	3	38	95	7500	34500	85'18
3000	2	40	100	6000	40500	100
N	40					
Iturria: lanketa propioa						

Formula aplikatzeko datu guztiak ditugularik, erraz lortzen da koefizientea:

$$C_G = \frac{\sum_{i=1}^{n-1}(p_i - q_i)}{\sum_{i=1}^{n-1} p_i} \qquad C_G = \frac{(50-29'63)+(87'5-66'67)+(95-85'18)}{50+87'5+95} = 0'22$$

Giniren koefizientea 0,22 izanik (Gini-ren indizea %22), enpresa horretan soldatei dagokionez kontzentrazio txikia dagoela ondorioztatu dezakegu. Hau da: soldaten banaketa egokia dago, Giniren koefizientea zerora asko hurbiltzen delako.

Adibidea. Giniren koefizientea: automobilgintzaren klusterra osatzen duten 40 enpresen fakturazioa azken urtean.

72. taula: automobilgintzaren klusterra osatzen duten 40 enpresen fakturazioa azken urtean.

Fakturazioa milioitan (x_i)	Enpresa kopurua (fi)	Fi	$p_i = \frac{F_i}{N}.100$	$u_i = x_i.f_i$	u_i metatuak	$q_i = \frac{u_i(metatuak)}{\sum u_i}.100$
3,5	10	10	25,0	35,0	35,0	13,6
4,5	12	22	55,0	54,0	89,0	34,6
6,0	8	30	75,0	48,0	147,0	57,2
8,0	5	35	87,5	40,0	187,0	72,8
10,0	3	38	95,0	30,0	217,0	84,4
15,0	1	39	97,5	15,0	232,0	90,3
25	1	40	100,0	25,0	257,0	100,0
Iturria: lanketa propioa						

Taula horretatik abiatuta, erraz lortzen dugu Gini koefizientea.

$$C_G = \frac{\sum_{i=1}^{n-1}(p_i - q_i)}{\sum_{i=1}^{n-1} p_i} \qquad Cg = \frac{83{,}99}{435} = 0{,}19$$

Giniren koefizientea 0,19 izanik (Giniren indizea %19), automobilgintzaren klusterra osatzen duten 40 enpresa horien fakturazioan kontzentrazio txikia dago. Hau da, fakturazioaren banaketa egokia dago, Giniren koefizientea zerora asko hurbiltzen delako.

Adibidea. Giniren indizea: 10 estaturen Giniren indizea Munduko Bankuaren arabera (2008-2011).

73. taula: Giniren indizea 10 estatutan 2008-2011				
Estatuaren izena	2008	2009	2010	2011
Afganistan	27,8			
Albania	34,5			
Angola		42,7		
Argentina	46,3	46,1	44,5	
Armenia	30,9		31,3	
Azerbaijan	33,7			
Bangladesh			32,1	
Bielorrusia	27,2	27,7	27,7	26,5
Bolivia	56,3			
Brasil	55,1	54,7		
Iturria: Munduko Bankua: http://databank.bancomundial.org/data/views/reports/tableview.aspx				

Ikusten den bezala, 10 estatu horien artean, Giniren indize handiena Argentinan eta Brasilen dago. Horrek esan nahi du errentaren banaketa desorekatuena estatu bi horietan dagoela.

2.3.6. Neurri estatistiko deskriptiboei buruzko ondorio orokorrak

Arestian esan bezala, estatistika deskribatzailearen helburuetariko bat datu multzoak modu ulerterrazean aurkeztea da. Horretarako, diagramarekin eta maiztasun-taulekin batera, neurri estatistikoak erabiltzen ditu.
Neurri estatistikoen helburua da datu multzoen ezaugarriak balio bakar baten bidez adieraztea. Horrela, neurri mota bakoitzak ezaugarri batzuei buruzko informazioa ematen digu:

- Zentro-neurriek, joera zentraleko posizio-neurriek edo joera zentraleko neurri estatistikoek datu multzoko datuak zein balioren inguruan biltzen diren adierazten digute.
- Sakabanatze-neurriek datu multzoko datuak batez bestekotik edo beste zentralizazio-neurri batetik zenbat aldentzen diren adierazten digute.
- Joera zentralekoak ez diren posizio-neurriek edo koantilek alde batean datu multzoaren ehuneko zehatza uzten duten balioei buruzko informazioa ematen digute.
- Forma-neurriek maiztasun-banaketaren itxurari buruzko informazioa ematen digute.
- Kontzentrazio-neurriek aztergai den aldagaia ikergai den populazioaren barruan nola banatzen den adierazten digute.

Gure helburuen arabera erabakiko dugu kalkulatu beharreko neurriak. Hala ere, gure datu multzoaren berezitasunak behar bezala ezagutzeko, egokia da neurri horiek guztiak kalkulatzea. Azken finean, guztiak osagarriak dira, eta datu multzoa ondo ezagutzeko beharrezkoak. Kontuan izan, baina, aldagaia neurtzeko erabilitako neurketa-eskalak zehazten duela kalkulatu nahi dugun neurri-estatistikoa. Hau da, aldagaiaren arabera, neurri estatistiko batzuk kalkulatu ahal izango ditugu, baina beste batzuk, ez.

74. taula: estatistikoak eta datu-diagramak aldagai motaren arabera					
Neurketa-eskala	Zentralizazio-neurriak	Sakabanatze-neurriak	Kokapen-neurriak	Forma-neurriak	Kontzentrazio-neurriak
Nominala	Moda	ez	ez	Ez	Ez
Ordinala	Moda mediana (zentzua badauka)	ez	ez	Ez	Ez
Tarte-eskala	Moda batez bestekoa mediana	bai	bai	Bai	Bai
Arrazoi-eskala	Moda batez bestekoa mediana	bai	bai	Bai	Bai
Iturria: lanketa propioa					

2.4. BANAKETA NORMALA

Banaketa normala aldagai kuantitatiboen (tartetan banatuta dauden aldagai jarraitu zein diskretuen) banaketen artean ezagunena eta ohikoena da. Are gehiago, estatistikan eredu gisa gehien erabiltzen den banaketa da, zorizko fenomeno asko hara hurbiltzen direlako. Hortik dator, hain zuzen ere, normal izena. Pertsonen ezaugarri psikologikoak (adimen-test batean lortutako puntuazioak edo portaera zehatzak), morfologikoak (altuera, pisua, hezurren luzera...), soziologikoak (azterketa bateko notak, diru-sarrerak, produktu baten kontsumoa, kontsumitutako telebista orduak, ...) edo politologikoak (eskala baten arabera neurtutako alderdi politiko edo politikariekiko iritzia, ...) aztertzen baditugu, maiztasun-banaketatik lortzen dugun histograma kanpai-itxurakoa, simetrikoa eta kurtosi-maila ertainekoa dela ikusten dugu; hau da, banaketa normalak dira.

Adibidea. Banaketa normala: zaldi-lasterketetako denborak. Demagun lasterketa-denboraldi osoan zehar hipodromo batean lehiatu duten 279 zaldik egin dituzten denborei (minutuetan) buruzko datuak ditugula.

75. taula: zaldi-lasterketetako denborak

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	135,00	2	,7	,7	,7
	136,00	2	,7	,7	1,4
	137,00	4	1,4	1,4	2,9
	138,00	6	2,2	2,2	5,0
	139,00	9	3,2	3,2	8,2
	140,00	12	4,3	4,3	12,5
	141,00	21	7,5	7,5	20,1
	142,00	28	10,0	10,0	30,1
	143,00	30	10,8	10,8	40,9
	144,00	36	12,9	12,9	53,8
	145,00	30	10,8	10,8	64,5
	146,00	27	9,7	9,7	74,2
	147,00	22	7,9	7,9	82,1
	148,00	14	5,0	5,0	87,1
	149,00	13	4,7	4,7	91,8
	150,00	10	3,6	3,6	95,3
	151,00	5	1,8	1,8	97,1
	152,00	3	1,1	1,1	98,2
	153,00	3	1,1	1,1	99,3
	154,00	2	,7	,7	100,0
	Total	279	100,0	100,0	

Iturria: lanketa propioa

Datu multzo horretatik abiatuta, honako neurri estatistiko hauek lortzen ditugu:

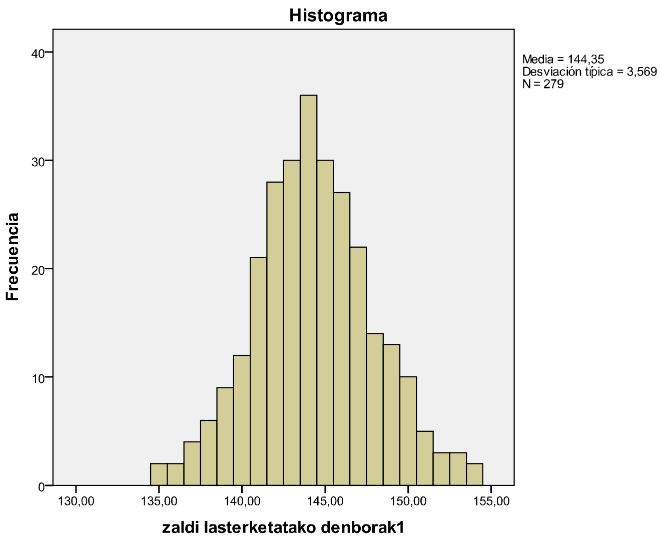
Estadísticos
Zaldi lasterketako denborak

N	Válidos	279
	Perdidos	0
Media		144,3477
Mediana		144,0000
Moda		144,00

Desv. típ.		3,56909
Varianza		12,738
Asimetría		,111
Error típ. de asimetría		,146
Curtosis		,047
Error típ. de curtosis		,291
Percentiles	25	142,0000
	50	144,0000
	75	147,0000

Ikusten den bezala, moda bakarra dugu, batez bestekoa, mediana eta moda ia berdinak dira, eta asimetria-koefizientea eta kurtosia asko hurbiltzen dira zerora. Estatistiko horien arabera, banaketa normal bat da.
Aurkeztutako maiztasun-taulak honako histograma hau ematen digu:

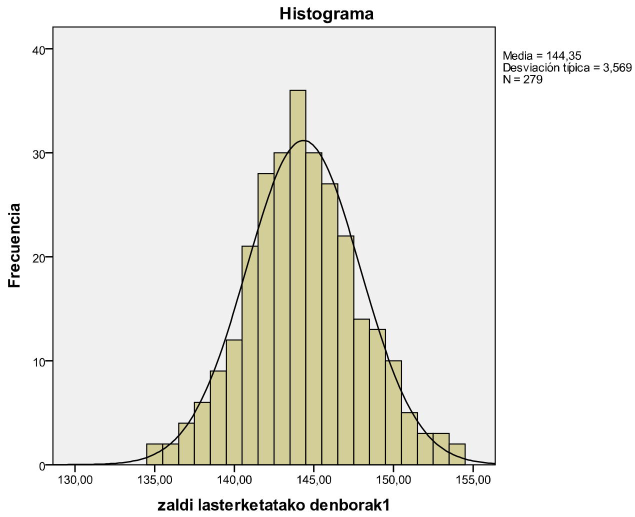
40. diagrama: zaldi-lasterketetako denboren histograma



Iturria: lanketa propioa

Histograma horrek sortzen duen kurbari begiratzen badiogu, kanpai-itxurakoa dela ikusten dugu; horrez gain, ikusten dugu simetrikoa (asimetria = 0,11) eta mesokurtikoa (0,047) dela. Beste era batera esanda, balio gehienak batez bestekoaren inguruan pilatzen dira simetrikoki. Hala, argi dago banaketa normal bat dela, histogramak sortzen duen kurbak erakusten duen bezala (kurba normal enpirikoa)
Hala, histograma horren gainean, datu multzo hori modelatzea eta, horren ondorioz, probabilitateak kalkulatzea ahalbidetzen duen kurba normala irudika dezakegu; hau da, kurba normal enpirikoaren gainean (ikerketaren ondorioz lortu dugun histogramak sortzen duen kurbaren gainean), kurba normal teorikoa irudika dezakegu (Gauss kanpaia edo kurba normala).

41. diagrama: zaldi-lasterketetako denboren histograma eta kurba normala



Iturria: lanketa propioa

Histogramaren gainean irudikatu dugun kurba teorikoari Gaussen kanpaia deritzo, eta banaketa normalaren grafikoa da. Kurba normala irudikatzeko erabili dugun trinkotasun-funtzioa honako hau da:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} ; -\infty < x < \infty$$

Non,

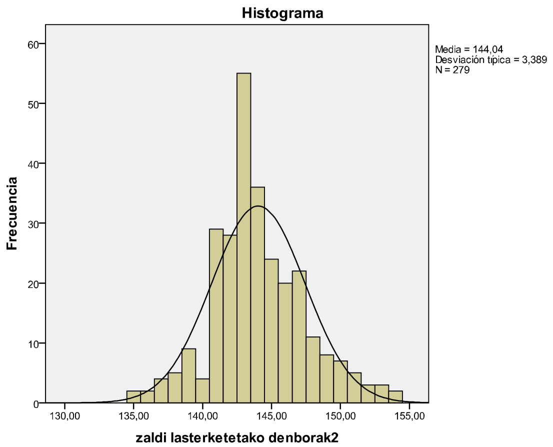
f = kurbaren altuera den, x-ren edozein baliotarako
π = 3.14159
e = 2.71828 den
μ = batez bestekoa
σ = desbideratze tipikoa

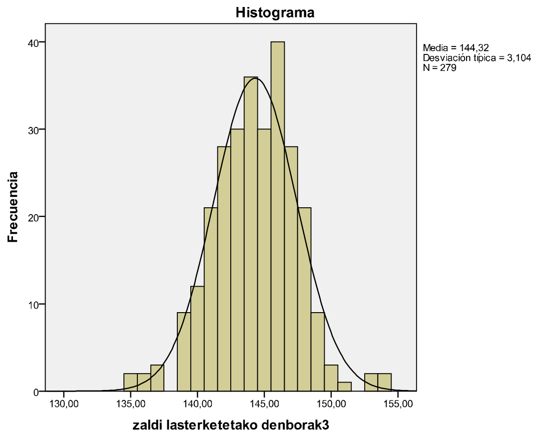
Funtzio horren laguntzarekin, Ardatz Kartesiarretako ardatz horizontalean jarriko dugun x_i balio bakoitzerako, ardatz bertikalean dagokion f balioa lortuko dugu, datu multzo horri tokatzen zaion Gauss kanpaia lortu arte. Kanpai hori datu multzo horren probabilitate-banaketa izango da, eta datu multzo horren barruan probabilitateak kalkulatzeko erabili ahal izango dugu, aurrerago ikusiko dugun bezala.
Kontuan izan behar da, batez bestekoa eta desbideratze tipikoa aldatzen diren heinean, trinkotasun-funtzio desberdinak dituzten gauss kanpai edo kurba normal desberdinak lortzen ditugula.
Formula ezagutzea beharrezkoa da, baina praktikan ez da erabiltzen. Izan ere, kurba normalarekin lan egiteko (datu multzo bati buruzko probabilitateak kalkulatzeko, alegia), kurba normalaren azpiko azalerei buruzko informazioa ematen digun taula bat dago; hau da, lor

ditzakegun taula normalaren mota guztiekin lan egiteko taula bakarra dugu. Taula hori kurba normalaren berezitasun batean oinarrituz eraikita dago: dena delako batez bestekoa eta desbideratze tipikoa duen banaketa normala dugularik, desbideratze tipikoaren arabera batez bestekoaren inguruan sortzen dugun tartean, beti dago azalera konstantea (edo datuen proportzio konstantea). Horren ondorioz, banaketa normal teorikoaren berezitasunetan oinarrituta, banaketa enpirikoaren (guk dugun banaketaren) berezitasunei buruzko baieztapenak egin ditzakegu.

Adibidea. Banaketa normala: batez besteko eta desbideratze desberdinen araberako gauss kanpaiak.

42. diagrama: Batez besteko eta desbideratze tipiko desberdinak dituzten bi banaketa normalen trinkotasun-funtzioei dagozkien Gauss kanpaiak edo kurba normalak





Iturria: lanketa propioa

Estadísticos

Zaldi-lasterketetako denborak2

N	Válidos	279
	Perdidos	0
Media		144,0394
Mediana		144,0000
Moda		143,00
Desv. Típ.		3,38881
Varianza		11,484
Asimetría		,342
Error típ. de asimetría		,146
Curtosis		,570
Error típ. de curtosis		,291
Percentiles	25	142,0000
	50	144,0000
	75	146,0000

Estadísticos

Zaldi-lasterketetako denborak3

N	Válidos	279
	Perdidos	0
Media		144,3190
Mediana		144,0000
Moda		146,00
Desv. típ.		3,10408
Varianza		9,635
Asimetría		-,094
Error típ. de asimetría		,146
Curtosis		,613
Error típ. de curtosis		,291
Percentiles	25	142,0000
	50	144,0000

Bi datu multzoetan N 279 izanik, eta batez bestekoa eta mediana antzerakoak direlarik, Gauss kanpai edo kurba normal desberdinak lortzen ditugu. Esan bezala, batez bestekoa eta desbideratze tipikoa aldatzen diren heinean, trinkotasun-funtzio desberdinak dituzten gauss kanpai desberdinak lortzen ditugu.
Gauss kanpai horiek datu multzoan probabilitateak kalkulatzeko erabiliko ditugu.

2.4.1. Banaketa normalaren berezitasunak

Banaketa normala askotan azaltzen da Gizarte Zientzietan; arestian esan bezala, zorizko fenomeno asko kanpai-forma horretara hurbiltzen direlako. Eta, egia esan, hori oso egokia da ikerlariontzat, lan estatistiko-matematikoa izugarri errazten duelako.
Dena dela, gure datu multzoetarako ezin dugu banaketa normala onartu, inolako frogarik egin gabe. Alegia, gure datu multzoa banaketa normalera hurbiltzen den ala ez esan aurretik, honako baldintza hauek betetzen diren ziurtatu behar dugu:

- Aldagaia kuantitatiboa (diskretua zein jarraitua) da; bestela, ezin dugu histograma irudikatu (kurba normal enpirikoa), ezta histogramari dagokion kurba normala ere (kurba normal teorikoa).
- Histograma kanpai-itxurakoa, simetrikoa eta mesokurtikoa da. Dena dela, kontuan izan gutxitan aurkituko dugula kanpai perfektuaren forma duen maiztasun-banaketa enpirikorik. Muga batzuen barruan, banaketa normalek asimetria eta zorroztasun maila desberdinak izan ditzakete.
- Batez bestekoa, mediana eta moda leku berdintsuan daude histograman; hau da, hiru neurri estatistiko horiek antzerakoak dira.
- Desbideratze tipikoa batez bestekoaren erdia baino txikiagoa da.
- Nahikoa handia da (30 baino gehiago)

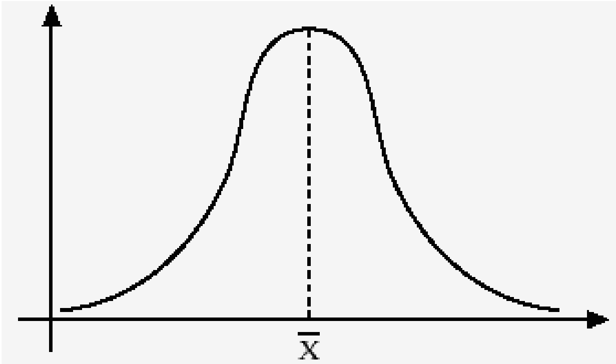
Gure datu multzoak baldintza horiek betetzen dituela ziurtatu dugunean, eta, beraz, banaketa normalera hurbiltzen dela dakigunean, zera esango dugu: maiztasun-banaketa horren batez bestekoa μ dela, eta desbideratze tipikoa, σ. Berezitasun horiek honela adierazten dira:

$$x_i \rightarrow N(\mu, \sigma)$$

Gure maiztasun-banaketa (banaketa enpirikoa) banaketa normala dela ziurtatu dugunean, maiztasun-banaketa horren gainean irudikatzen dugun Gauss kanpaiak edo kurba normalak (banaketa teorikoa) honako berezitasun hauek ditu:

- Gauss kanpaiaren edo kurba normal teorikoaren puntu gorenean, batez bestekoa dago.

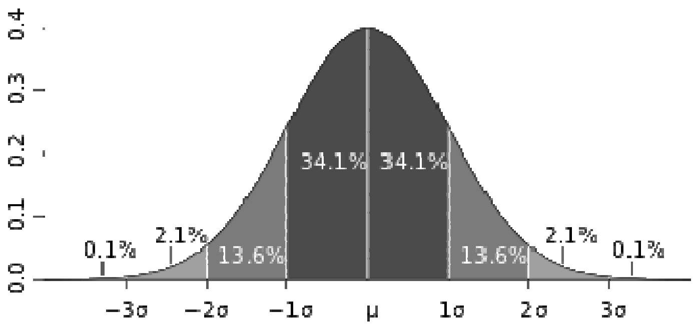
43. diagrama: Gauss kanpaia eta batez bestekoaren posizioa



- Gauss kanpaiaren edo kurba normalaren (kurba normal teorikoaren) trinkotasun-funtzioa simetrikoa da batez bestekoaren inguruan. Horrek zera esan nahi du: Gauss kanpaiaren azpian dagoen azaleraren erdia (edo datu multzoaren datu eta balio erdiak) batez bestekoa baino aurretik dagoela, eta beste erdia, batez bestekoaren ostean.
- Kurba normala, probabilitate-banaketa izanik, aldagai batek har ditzakeen balio guztiek gertatzeko duten probabilitate teorikoa adierazten du. Probabilitateak adierazten dituenez, kurbaren azpiko azalera osoak 1 balio du.
- Balio infinitu ditugunean, kanpaiaren lerromakurra asintotikoa da absize ardatzarekiko. Horrek zera esan nahi du: histogramaren besoek ez dutela inoiz bat egiten x ardatzarekin, eta -∞tik +∞ doazen balioak hartzen dituztela.
- Mediana, moda eta batez bestekoa berdinak edo oso antzerakoak dira.
- Kurba normalaren batez bestekoaren inguruan, honako probabilitate hauek aurkitzen ditugu:
    - [μ - 1σ, μ + 1σ] tartean, gutxi gorabehera, maiztasun-banaketako datuen %68,26 dago .
    - [μ - 2σ, μ + 2σ] tartean, gutxi gorabehera, maiztasun-banaketako datuen %95,44 dago.
    - [μ - 3σ, μ + 3σ] tartean, gutxi gorabehera, maiztasun-banaketako datuen %99,74 dago.

Berezitasun horri “68-95-99,7 araua” edo “arau enpirikoa” deritzo, eta banaketa normalari dagozkion probabilitateak ezagutzeko balio du. Arau horren arabera, batez bestekotik desbideratze tipiko bat, bi edo hiru aldenduta, kurbaren azpiko azaleraren %68,26, %95,44 eta %99,74 dago; edo gure datu multzoko datuen %68,26, %95,44 eta %99,74 tarte horietan daude.

44. diagrama: Gauss kanpaia eta “68-95-99,7 araua”



Iturria: https://commons.wikimedia.org/wiki/File:Standard_deviation_diagram.svg

“68-95-99,7 araua” edo “arau enpirikoa” ulertzeko, beste honako azalpen hau ere baliagarria izan daiteke: dena delako maiztasun-banaketa normala eta kurba normala dugularik, ardatz horizontaleko edozein puntutatik batez bestekora arteko distantzian (distantzia desbideratze tipikoen arabera neurtua) beti dagoela guztizko azaleraren proportzio bera (edo datuen proportzio bera).

“68-95-99,7 araua” oso egokia da banaketa normalaren azpian aurkitu ditzakegun probabilitateetara lehen hurbilketa egiteko, baina, zehaztasun gehiagorekin lan egin gura badugu, kontuan izan behar dugu Gauss kanpaiaren azpiko probabilitateak honako tarte zehatz hauetan daudela:

- [μ - 1,28155 σ, μ + 1,28155 σ] tartean, gutxi gorabehera maiztasun-banaketako datuen %80 dago.

- [μ - 1,64485 σ, μ + 1,64485 σ] tartean, gutxi gorabehera maiztasun-banaketako datuen %90 dago.
- [μ - 1,95996 σ, μ + 1,95996 σ] tartean, gutxi gorabehera maiztasun-banaketako datuen %95 dago.
- [μ - 2,32635 σ, μ + 2,32635 σ] tartean, gutxi gorabehera maiztasun-banaketako datuen %98 dago.
- [μ - 2,57583 σ, μ + 2,57583 σ] tartean, gutxi gorabehera maiztasun-banaketako datuen %99 dago.
- [μ - 2,80703 σ, μ + 2,80703 σ] tartean, gutxi gorabehera maiztasun-banaketako datuen %99,5 dago.
- [μ - 3,09023 σ, μ + 3,09023 σ] tartean, gutxi gorabehera maiztasun-banaketako datuen %99,8 dago.
- [μ - 3,29052 σ, μ + 3,29052 σ] tartean, gutxi gorabehera maiztasun-banaketako datuen %99,9 dago.
- [μ - 3,8906 σ, μ + 3,8906 σ] tartean, gutxi gorabehera maiztasun-banaketako datuen %99,99 dago.
- [μ - 4,4172 σ, μ + 4,4172 σ] tartean, gutxi gorabehera maiztasun-banaketako datuen %99,999 dago.

"68-95-99,7" arauari esker, lan estatistiko-matematikoa izugarri errazten da. Izan ere, histogramak zehazten duen kurbaren azpiko bi punturen (edo balioren) arteko azalera ezagutzeko, bi puntu horien artean integratu behar genuke, baina, banaketa normalaren berezitasun horri esker, batez bestekoa eta desbideratze tipikoa ezagututa, azalera hori erraz ezagut dezakegu. Alegia, batez bestekoa eta desbideratze tipikoa ezagututa, histogramako edo maiztasun-banaketako bi balioren artean balio bat aurkitzeko dugun probabilitatea, zein, probabilitate zehatz batekin, balio bi horien artean zenbat datu dauden jakin dezakegu.

Adibidea. Banaketa normala: histogramako bi punturen artean dagoen datuen ehunekoen ezagutza, "68-95-99,7 araua" erabiliz: Soziologiako eta Politika Zientziak eta Kudeaketa Publikoko ikasleen artean lagin bat hartu dugu, eta bakoitzari ikasgelara zenbat diru ekartzen duen galdetu diogu. Lortu dugun banaketa normal erakoa da, eta jakin dugun bakoitzak batez beste 11,3 euro ekartzen dituela ikasgelara, eta desbideratze tipikoa 3,8 eurokoa dela. Ikasleek ikasgelara ekartzen duten batez besteko diruari (11,3) bi aldeetara desbideratze tipiko bat gehitzen (11,3 + 3,8) eta kentzen badiogu (11,3 - 3,8), ondorioztatzen dugu ikasleen %68 7,5 eurotik 15,1 eurora dituztela etortzen dela ikasgelara:

%68 $\rightarrow$ $\bar{x}$ + 1.s = 11,3 + 1 . 3,8 = 7,5-15,1

Soziologiako eta Politika Zientziak eta Kudeaketa Publikoko ikasleak, guztira, 70 badira, 48 ikasle (%68) 7,5 eurotik 15,1 eurora dituztela etortzen dira ikasgelara.

Ikasleen %99 zenbat dirurekin etorriko litzatekeen jakin nahiko bagenu, hau da, gehiengoari buruzko informazioa lortu nahiko bagenu, "68-95-99,7 araua" erabiliz, zera egin behar genuke:

%99,7 $\rightarrow$ $\bar{x}$ + 3.s = 11,3 + 3 . 3,8 = 0,1-22,7

Ikusten den bezala, probabilitatea handitu dugun heinean, tartea ere asko handitu da. Hain zuzen ere, ikasleen %99,7 (70 ikasleak) 0 eurotik 22,7 eurora dituztela etortzen da ikasgelara.

Adibidea. Banaketa normala: histogramako bi punturen artean dagoen datuen ehunekoen ezagutza: Markina-Xemeingo biztanleriatik lagin adierazgarria hartu dugu (n=300), eta aukeratutakoei Gabonetako loterian zenbat diru jokatzen duten galdetu diegu. Horrela jakin dugu pertsona bakoitzak batez beste 65 euro gastatzen dutela loterian, 5 euroko desbideratze

tipikoarekin. Zer balioren artean dago markinarren %95 Gabonetan loterian jokatzen duten diruari dagokionez? Inolako eragiketarik egin gabe, maiztasun-banaketa banaketa normalera hurbiltzen dela jakinda, erraz erantzun diezaiokegu galderari:
%95 = $\bar{x}$ + 1,95996 . s = 65 + 1,95996 . 5 = 65 + 9,7998 = 55,2-74,79.
Markinarren %95ek 55,2 euro eta 74,79 euro artean jokatzen dute Gabonetako loterian.
"68-95-99,7" arauari, hau da, banaketa normalaren berezitasun horri esker, batez bestekoa eta desbideratze tipikoa ezagututa jakin dezakegu zein probabilitate dugun histogramako edo maiztasun-banaketako bi balioren artean balio bat aurkitzeko, zein probabilitate zehatz batekin balio bi horien artean zenbat datu dauden. Hori da, hain zuzen ere, banaketa normala eredutzat hartzearen abantailetariko bat.

2.4.2. Zorizko aldagaien probabilitate-banaketak eta probabilitatezko banaketa normala

Fenomeno deterministetan, emaitza erraz aurresan dezakegu, haiengan eragina duten faktore guztiak ezagutzen ditugulako. Adibidez, hutsezko baldintzetan gorputz bat zer alturatatik botatzen den baldin badakigu, lurrazalera heltzeko behar duen denbora aurresan dezakegu. Edo baldintza jakinetan bi osagai kimiko nahasita, aldez aurretik jakin dezakegu emaitza zein izango den. Oro har, baldintza ezagunetan jazotzen diren lege fisiko eta kimiko asko deterministak dira.
Fenomeno deterministez gain, zorizko edo ausazko fenomenoak ditugu (fenomeno aleatorio ere deitzen zaie). Adibidez, zorizko jokoetan ez dugu jakiten zein zenbaki, dado edo karta aterako den. Edo herrialde baten urte bateko errenta per capita ezin dugu aldez aurretik ezagutu, zorizko edo ausazko gertaerak ziurgabeak eta iragarrezinak direlako. Hala ere, epe luzera, zorizko gertaerek erregulartasunak agertzen dituzte; eta probabilitatearen teoriaren bidez azter daitezke. Zehatzago esanda, datu multzo bati buruzko probabilitateak probabilitate-banaketen bidez iker ditzakegu.
Ausaz edo zoriz gertatzen diren fenomenoetan, emaitza posible bakoitzari gertaera deitzen zaio. Zorizko aldagai baten probabilitate-banaketa aldagai horren gertaera bakoitzak (aldagaiaren balio bakoitza izan daiteke) agertzeko edo ez gertatzeko duen probabilitatea adierazten duen funtzio matematikoa da. Beste era batera esanda, probabilitate-banaketek zorizko aldagaiek har ditzaketen balio guztiak, eta balio horiek maila teorikoan gertatzeko duten probabilitateak agertzen dituzte. Ildo horretatik, probabilitate-banaketak maiztasun erlatiboko banaketen idealizazioak dira. Alegia, maiztasun erlatiboen maiztasun-banaketak enpirikoak dira (benetan gertatuak) eta probabilitateen maiztasun-banaketak teorikoak (idealizazioak edo teorikoak).
Probabilitate-banaketak taula, diagrama (barra-diagrama edo histograma) edo formula matematikoen bidez adieraz ditzakegu.
Ikuspuntu grafikotik, aldagai jarraituen (edo aldagai jarraitu gisa aztertu nahi diren aldagai diskretuen) probabilitate-banaketak histogramen bitartez eta beren gainean irudikatzen dugun kurba normalaren bitartez adierazten ditugu.
Ikuspuntu matematikotik, probabilitate-banaketak y = f(x) funtzioaren bitartez definitzen dira. Hau da, datu multzo bakoitzarentzat, probabilitate funtzio bat (aldagai diskretua denean) edo trinkotasun funtzio bat (aldagai jarraitua denean) lortzen dugu.
Probabilitate-funtzioak kurba azpian dagoen azalera adierazten du, eta, haren bidez, kurbaren azpiko balioen arteko azalerak ezagut ditzakegu. Arestian esan bezala, probabilitate-funtzioa ezagututa, histogramako edo maiztasun-banaketako bi balioren artean balio bat aurkitzeko dugun probabilitatea, zein, probabilitate zehatz batekin, balio bi horien artean zenbat datu dauden jakin dezakegu.
Banaketa normala maiztasun-banaketa jarraituen artean ezagunena eta ohikoena da. Dena dela, ezin dugu ahantzi beste probabilitate-banaketa batzuk ere badaudela, nahiz eta liburu honetan ez ditugun landuko.
Zorizko aldagai diskretuen probabilitate-banaketa garrantzitsuenak honako hauek dira:

- Banaketa binomiala
- Banaketa binomial negatiboa
- Poisson banaketa
- Banaketa geometrikoa
- Banaketa hipergeometrikoa
- Bernoulliren banaketa
- Rademacher banaketa
- Banaketa diskretua uniformea

Probabilitate-banaketa diskretuak probabilitate funtzioaren edo maiztasun-banaketa funtzioaren bitartez definitzen dira.

Aldagai diskretuen probabilitate-banaketekin batera, zorizko aldagai jarraituen probabilitate-banaketak ditugu. Honako hauek dira garrantzitsuenak:

- Banaketa esponentziala
- T Student banaketa
- Banaketa normala
- Gamma banaketa
- Beta banaketa
- F banaketa
- Banaketa jarraitua uniformea
- Weibull-en banaketa
- Paretoren banaketa

Probabilitate-banaketa jarraituak trinkotasun funtzioaren bitartez edo maiztasun-banaketa funtzioaren bitartez definitzen dira.

Banaketa normal bat dugun kasuetan, haren probabilitate-banaketa arestian aurkeztu dugun trinkotasun funtzioaren bidez lortzen dugu:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} ; -\infty < x < \infty$$

Dena dela, probabilitate-banaketa horretan probabilitateak kalkulatzeko, trinkotasun-funtzioan integratu behar genuke:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx, \quad x \in \mathbb{R}$$

Kalkulu horiek egitea konplexua denez, probabilitateak estandarizazioaren bidez kalkulatzen dira. Alegia, banaketa normal bakoitzarekin integratzen ibili beharrean, banaketa normal guztiak banaketa normal estandarrera bihurtzen dira, eta, ostean, banaketa normal estandarreko probabilitateak erraz kalkulatzen dira, N (0,1) kurba normalaren azpiko azalerak adierazten dizkigun taularen bidez. Beste era batetara esanda: banaketa normal bakoitzerako probabilitate-taula bat osatzea ezinezkoa denez, banaketa normal estandarrerako probabilitate-taula sortu da, eta, ostean, banaketa normal guztiak banaketa normal estandarrera eraldatzen dira; hala probabilitateen kalkuluak erraztu egiten dira.

2.4.3. Banaketa normal estandarra

Banaketa normala bi parametroren araberakoa da: μ batez bestekoaren eta σ desbideratze estandarraren araberakoa. x_i aldagai batek banaketa normalari jarraitzen diola modu honetan adierazten da:

$$x_i \rightarrow N(\mu, \sigma)$$

Banaketa normal estandarra μ = 0 eta σ = 1 parametroak dituen banaketa normala da, eta, banaketa normalen probabilitateak kalkulatzeko, eredu teoriko gisa erabiltzen da. Banaketa normal estandarra honela adierazten da:

$$z \rightarrow N(0,1)$$

x_i zorizko aldagaitik z banaketa tipifikatuko aldagaira aldatzeko egiten den eragiketari deitzen zaio estandarizazioa edo tipifikazioa. Hau da, estandarizazioa estandarra ez den N (μ, σ) banaketa normal batetik, N (0, 1) banaketa normal estandarrera eraldatzeko prozesua da. Prozedura horrekin, eskala-aldaketa bat egiten dugu, eta lortzen dugun maiztasun-banaketa tipifikatuaren batez bestekoa 0 da, eta desbideratze tipikoa, 1. Estandarizazioa edo tipifikazioa honako formula honen bidez egiten da:

$$z = \frac{x_i - \mu}{\sigma}$$

Ikusten den bezala, balio bat estandarizatzea edo tipifikatzea datu bat batez bestekotik zenbat aldentzen den adierazten duen balioa da, z balioa izenekoa.

Azaldutako prozesu guztia modu honetan laburbil dezakegu:

$$X \sim N(\mu, \sigma) \rightarrow Z = \frac{X - \mu}{\sigma} \sim N(0,1)$$

Adibidea. Pertsona batek 120 puntu lortu ditu test batean. Banaketa normal bat dela kontuan hartzen badugu N (120, 20), 120ko puntuazio horri 1eko z puntuazioa dagokio. Izan ere:

$$z = \frac{x_i - \mu}{\sigma} = \frac{120 - 100}{20} = 1$$

Estandarizazioari esker, datu multzoetako (maiztasun-banaketetako) datuak alderatu ditzakegula, z balioak dimentsio gabeak direlako. Adibidez, bi herrialdetako bi familiaren errentak alderatzeko, bi errentak estandarizatu behar dira. Izan ere, herrialdeen berezitasunak kontuan hartu gabe bi familiak alderatzen baditugu, ondorio okerrak atera ditzakegu. Estandarizatu ostean, z balio handiena duen errenta izango da bien artean handiena.

Adibidea. Tipifikazioa eta datu multzo desberdinetako datuak alderatzea: Soziologiako eta Politika Zientziak eta Kudeaketa Publikoko ikasleek test batean ateratako puntuazioak agertzen dira ondoren aurkezten ditugun bi maiztasun-banaketetan.

76. taula: Soziologiako taldea	
0-10	10
10-20	15
20-30	5
30-40	15
40-50	5
Iturria: lanketa propioa	

77. taula: Politika Zientziak eta Kudeaketa Publikoko taldea	
0-6	9
6-12	11
12-18	5
18-24	15
24-30	5
Iturria: lanketa propioa	

Soziologiako ikasle batek 25 puntu atera baditu, eta Politika Zientziak eta Kudeaketa Publikoko beste batek 15, bietariko zeinek du posizio hobea bere taldean?
Bi ikasle horien artean bere taldean nork duen posizio hobea ezagutzeko, lortu dituzten puntuazioak tipifikatu egin behar dira:

Soziologiako taldearen batez bestekoa: 23

Politika Zientziak eta Kudeaketa Publikoko taldearen batez bestekoa: 14,47

Soziologiako taldearen desbideratze tipikoa: 13,27

Politika Zientziak eta Kudeaketa Publikoko taldearen desbideratze tipikoa 8,08

$$z = \frac{25 - 23}{13,27}$$ 0,1507

$$z = \frac{15 - 14,47}{8,08}$$ 0,066

z biak ikusita, zera ondoriozta dezakegu: Soziologiako taldean, 25 atera duenak posizio hobea duela bere taldean Politika Zientziak eta Kudeaketa Publikoko taldean 15 atera duenak baino. Izan ere, 25 puntuazioaren estandarizatutako z balioa handiagoa da (0,1507).
Estandarizazioari esker, banaketa normaletan balio batetik behera, gora edo balio biren arteko probabilitateak oso erraz kalkulatzen dira, kurba normalaren azpiko azaleren taularen bidez. Arestian esan bezala, banaketa normal bakoitzerako probabilitate-taula bat osatzea ezinezkoa denez, banaketa normal estandarrerako probabilitate-taula bat sortu da. Hala, banaketa normal guztiak banaketa normal estandarrera eraldatzen dira (z balioetara), eta probabilitateak erraz kalkulatzen.

78. taula: kurba normalaren azpiko azalera-taula

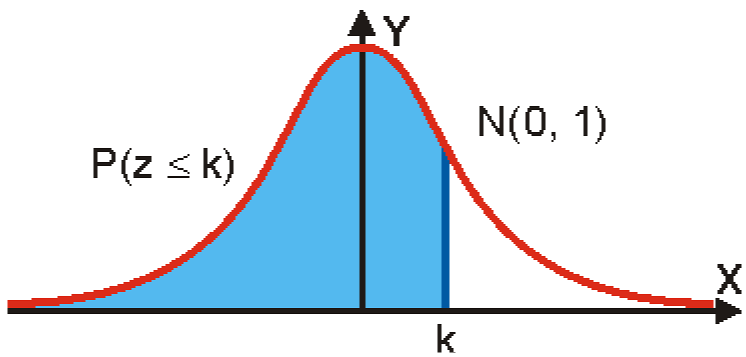
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224

0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.990097	0.990358	0.990613	0.990863	0.991106	0.991344	0.991576
2.4	0.991802	0.992024	0.992240	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613
2.5	0.993790	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.995060	0.995201
2.6	0.995339	0.995473	0.995604	0.995731	0.995855	0.995975	0.996093	0.996207	0.996319	0.996427
2.7	0.996533	0.996636	0.996736	0.996833	0.996928	0.997020	0.997110	0.997197	0.997282	0.997365
2.8	0.997445	0.997523	0.997599	0.997673	0.997744	0.997814	0.997882	0.997948	0.998012	0.998074
2.9	0.998134	0.998193	0.998250	0.998305	0.998359	0.998411	0.998462	0.998511	0.998559	0.998605
3.0	0.998650	0.998694	0.998736	0.998777	0.998817	0.998856	0.998893	0.998930	0.998965	0.998999
3.1	0.9990324	0.9990646	0.9990957	0.9991260	0.9991553	0.9991836	0.9992112	0.9992378	0.9992636	0.9992886
3.2	0.9993129	0.9993363	0.9993590	0.9993810	0.9994024	0.9994230	0.9994429	0.9994623	0.9994810	0.9994991
3.3	0.9995166	0.9995335	0.9995499	0.9995658	0.9995811	0.9995959	0.9996103	0.9996242	0.9996376	0.9996505
3.4	0.9996631	0.9996752	0.9996869	0.9996982	0.9997091	0.9997197	0.9997299	0.9997398	0.9997493	0.9997585
3.5	0.9997674	0.9997759	0.9997842	0.9997922	0.9997999	0.9998074	0.9998146	0.9998215	0.9998282	0.9998347
3.6	0.9998409	0.9998469	0.9998527	0.9998583	0.9998637	0.9998689	0.9998739	0.9998787	0.9998834	0.9998879
3.7	0.9998922	0.9998964	0.99990039	0.99990426	0.99990799	0.99991158	0.99991504	0.99991838	0.99992159	0.99992468
3.8	0.99992765	0.99993052	0.99993327	0.99993593	0.99993848	0.99994094	0.99994331	0.99994558	0.99994777	0.99994988
3.9	0.99995190	0.99995385	0.99995573	0.99995753	0.99995926	0.99996092	0.99996253	0.99996406	0.99996554	0.99996696
4.0	0.99996833	0.99996964	0.99997090	0.99997211	0.99997327	0.99997439	0.99997546	0.99997649	0.99997748	0.99997843

Kurba normalaren azpiko azalera-taulak honako berezitasun hauek ditu:

- Taulak z = 0 eta z = 4,09 z balioei dagozkien probabilitateak kalkulatzeko balio du. Izan ere, taulak z = 0 eta z = 4,09 balioen azpitik dauden azaleren proportzioak adierazten ditu. Proportzioak adierazten dituen heinean, kontuan izan kurba osoaren azpiko azalerak 1 balio duela.
- Taulak balio batetik beherako probabilitateak kalkulatzeko balio du. Alegia, taulako balioek normal erako kurbaren azpiko azalerak adierazten dituzte, z-ren balio positiboetatik behera.
- z = 4,09tik gorako probabilitateak eta z balio negatiboei dagozkien probabilitateak banaketa normalaren simetria eta probabilitateen osagarritasuna erabiliz kalkulatzen dira.
- z > 4 balioetarako probabilitateak ia 1 dira.

45. diagrama: kurba normalaren azpiko azalera



2.4.4. Probabilitateak nola kalkulatu, banaketa normal estandarraren laguntzarekin

Probabilitateak modu desberdinean kalkulatzen dira, z balioaren berezitasunen arabera. Egiatan, sei motatako z balioen probabilitateak aurkitu ditzakegu kurba normalaren azpiko azalera-taularen bidez:

- z balio batetik beherako balioen probabilitateak kalkulatzea
- z balio batetik gorako balioen probabilitateak kalkulatzea
- z balio negatiboen probabilitateak kalkulatzea
- Bi balio positiboren arteko probabilitatea kalkulatzea
- Bi balio negatiboren arteko probabilitatea kalkulatzea
- Balio positibo baten eta negatibo baten arteko probabilitateak kalkulatzea

2.4.4.1. z balio batetik beherako balioen probabilitateak nola kalkulatu

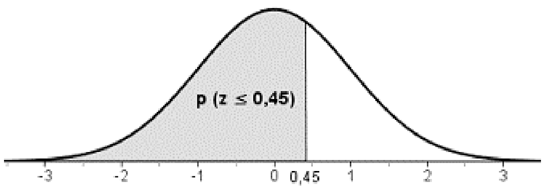
Esan bezala, taulak z balio batetik beherako probabilitateak ematen ditu. Probabilitate horiek kalkulatzeko, beraz, nahikoa da taulan begiratzea. Adibide batzuk ikusiko ditugu:

Adibidea. Banaketa normala estandarra: z balio batetik beherako balioen probabilitateak kalkulatzea.

P (z ≤ 0,45) ezagutu nahi badugu, hau da, banaketak normal estandarrean z = 0,45 edo balio txikiagoa aurkitzeko probabilitatea, 0,4 aurkitu behar dugu taulako errenkadetan eta 0,05 zutabeetan (kontuan izan errenkadetan unitateak eta hamarrenak aurkitzen ditugula eta zutabeetan ehungarrenak). Kasu honetan, 0,4 eta 0,05 zenbakiei dagozkien zutabearen eta errenkadaren gurutzaketaren ondorioz lortzen dugun laukitxoan dagoen zenbakia 0,6736 da.

P (z ≤ 0,45) = 0.6736.

46. diagrama: z balio batetik beherako balioen probabilitateak kalkulatzea



Kurba normalaren azpiko azalera-taularen zatia

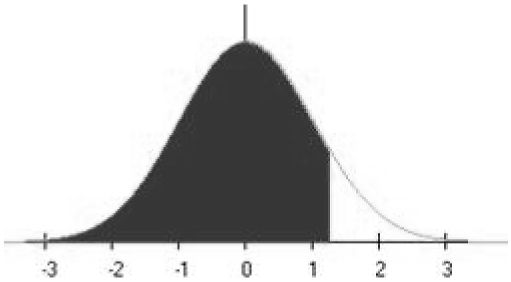
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	**0.6736**	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224

Hala, z = 0,45 edo balio txikiagoa aurkitzeko probabilitatea %67,36koa da.

Adibidea. Banaketa normal estandarra: z balio batetik beherako balioen probabilitateak kalkulatzea. Ikerketa bat egin ostean ikusi da biologiako ikasleek batez beste 8 euro ekartzen dituztela ikasgelara, 2 euroko desbideratze tipikoarekin. Banaketa normala dela kontuan hartuta [N (8,2)], zenbatekoa da norbaitek 10,2 euro edo gutxiago ekartzeko probabilitatea?

Tipifikatu ostean, 10,2 eurori dagokion z balioa 1,1 da, eta z balio hori 0 eta 4,09ren artean dagoenez, bilatzen ari garen probabilitatea zuzenean aurkitzen dugu taulan. Hain zuzen ere, 1,1eko z-ri 0,8643 probabilitatea dagokio. Beraz, 0,8643koa da ikasgelara 10,2 euro baino gutxiago ekartzeko probabilitatea. Alegia, ikasleen %86,43k 10,2 euro baino gutxiago ekarriko ditu ikasgelara. Histograma begiratzen badugu, beltzez dagoen azalerari dagokion probabilitatea bilatzen ari gara:

47. diagrama: z balio batetik beherako balioen probabilitateak kalkulatzea

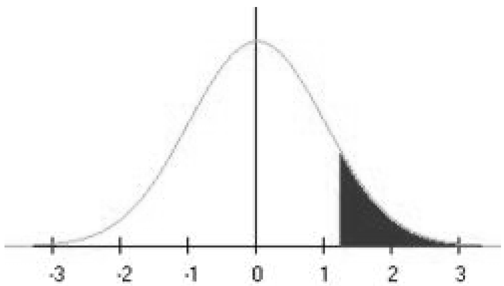


Kontuan izan probabilitate hori maiztasun absolutuetan ere adierazi dezakegula. Hala, biologiako ikasleak 150 badira, guztira, 130 (%86,43) ikasle etorriko dira ikasgelara 10,2 euro baino gutxiagorekin.

2.4.4.2. z balio batetik gorako balioen probabilitateak nola kalkulatu

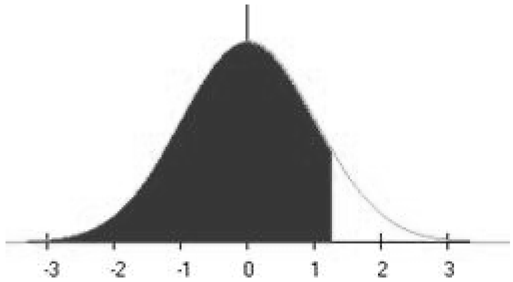
Esan bezala, taulak balio batetik beherako probabilitateak kalkulatzeko balio du. Horrek esan nahi du, balio batetik gorako probabilitateak ezin direla modu zuzenean aurkitu taulan. Dena dela, eragiketa txiki batekin, gai izango gara taulan aurkitzeko.
Balio bat agertzeko gehienezko probabilitatea 1 dela jakinda (gogoratu kurbaren azpian dagoen azalerak 1 balio duela), eta, horren ondorioz, $P(z \geq a) = 1 - P(z \leq a)$ dela kontuan hartuta, bilatzen ari garen probabilitatea erraz aurkitu dezakegu taularen laguntzarekin. Modu grafikoan honela irudika dezakegu probabilitate hori kalkulatzeko modua:
Beltzean dagoen Gauss kanpaiaren azalerari dagokion probabilitatea bilatzen ari gara, baina, taulen bidez, probabilitate hori ezin dugu ezagutu, tauletan z balio batetik beherako probabilitateak agertzen direlako:

48. diagrama: z balio batetik gorako balioen probabilitateak kalkulatzea



Azalera horren balioa ezagutzeko, z balio horretatik behera dagoen azalerari dagokion probabilitatea kalkulatzen dugu.

49. diagrama: z balio batetik gorako balioen probabilitateak kalkulatzea



Gauss kanpaiaren azpiko azaleraren probabilitatea 1 dela jakinda, 1i ezagutu nahi dugun z balio horretatik behera dagoen azalerari dagokion probabilitatea kentzen diogu, eta, horrela, z balio batetik gorako balioren probabilitateen ezagutzen dugu.

Adibidea. Banaketa normala estandarra: z balio batetik gorako balioen probabilitateak kalkulatzea: 7 urteko 500 ikasleren altuerak normal banatzen dira, N (70 zentimetro, 3 zentimetro) berezitasunekin. Kalkulatu zenbat ikasle dagoen 90 zentimetro edo gehiagorekin:

$P (x \geq 90) = P (z \geq 6{,}67) = 1 - P (z \leq 6{,}67) = 1 - 1 = 0 \times 500 = 0$

Adibidea. Banaketa normala estandarra: z balio batetik gorako balioen probabilitateak kalkulatzea: zein da politiko bati buruzko iritzia z = 0,42 baino handiago izateko probabilitatea? $P (z \geq 0{,}42)$ kalkulatu nahi badut, hau da, z = 0,42 baino handiago izateko probabilitatea, zera egin behar dut: $P (z \geq 0{,}42) = 1 – P (z \leq 0{,}42) = 1 - 0{,}6628 = 0{,}3372$.

Adibidea. Banaketa normala estandarra: z balio batetik gorako balioen probabilitateak kalkulatzea: P (z $\geq$ 1,24] kalkulatu nahi badut, hau da, z = 1,24 edo balio handiagoa aurkitzeko probabilitatea, zera egin behar dut: $P (z \geq 1{,}24) = 1 – P (z \leq 1{,}24) = 1 – 0{,}8925 = 0{,}1075$.

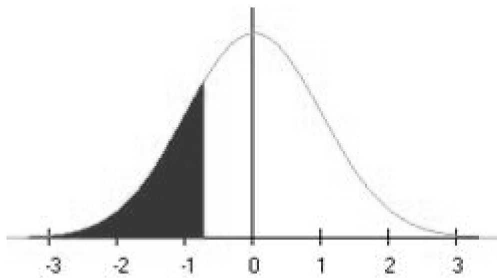
Adibidea. Banaketa normala estandarra: z balio batetik gorako balioen probabilitateak kalkulatzea: $P (z \geq 1.83) = 1 - P (z \leq 1.83) = 1 - 0.9664 = 0.0336$

Adibidea. Banaketa normala estandarra: z balio batetik gorako balioen probabilitateak kalkulatzea: $P (Z \geq 0.49) = 1 - P (z \leq 0.49) =1 - 0.6879 = 0.3121$

2.4.4.3. z balio negatiboen probabilitateak kalkulatzea

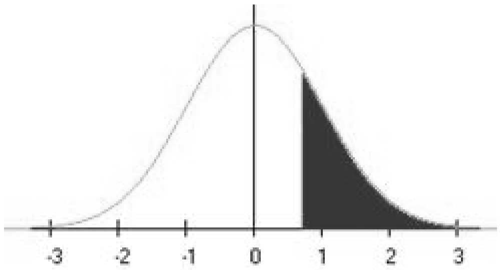
z balio negatiboei dagozkien probabilitateak aurkitu behar ditugunean, banaketa normalaren simetrian oinarrituta lan egin behar dugu. Izan ere, taulan ez dira azaltzen z balio negatiboak.

50. diagrama: z balio negatiboen probabilitateak kalkulatzea



Beraz, banaketa normalaren simetrian oinarritua, z balio negatibo horri dagokion azaleraren azalera simetrikoaren probabilitatea kalkula dezakegu, eta, hortik abiatuta, bilatzen ari garen probabilitatea aurkitu.

51. diagrama: z balio negatiboen probabilitateak kalkulatzea



Adibidea. Banaketa normala estandarra: z balio negatiboen probabilitateak kalkulatzea: zenbatekoa da norbaitek z = - 0,72 edo gutxiago ateratzeko probabilitatea? Badakigu taulan ez direla agertzen z-ren balio negatiboak. Dena dela, grafikoa simetrikoa denez, zera egin dezakegu: P (z ≤ - 0,72) = P (z ≥ + 0,72). Horrela, P (z ≥ + 0,72) kalkulatzen dugu: P (z ≥ + 0,72) = 1 - P (z < + 0,72) = 1 - 0,7642 = 0,2358. Beraz, norbaitek z = - 0,72 edo gutxiago ateratzeko probabilitatea 0,2358 da.

Adibidea. Banaketa normala estandarra: z balio negatiboen probabilitateak kalkulatzea: Fakultate batean irakasleek duten pisuak banaketa normalari jarraitzen dio N (78, 36). Zein da 72 kg. edo gehiago izateko probabilitatea?

$$P\ (x \geq 72) = P\ (\ z \geq \frac{72-78}{36}) = P\ (z \geq -\ 0{,}16\ ) = 1 - P\ (z \leq 0{,}16) = 1 - 0{,}5636 = 0{,}4364$$

Irakasleen %43,64ak 72 kg. edo gehiago izango dituzte. Fakultatean, guztira, 210 irakasle badira, 92 irakaslek 72 kg. edo gehiago izango dituzte.

Adibidea. Banaketa normala estandarra: z balio negatiboen probabilitateak kalkulatzea: 800 jaioberriren altuerak normal banatzen dira, N (66, 5) berezitasunarekin. Kalkula ezazu zenbat jaioberri dagoen 64 kg. edo gutxiagorekin:

$$P\ (x \leq 64) = P\ (\ z \geq \frac{64-70}{3}) = P\ (z \leq -\ 2\ ) = P\ (z \leq +\ 2) = 1 - P\ (z \leq 2) = 1\text{-}\ 0{,}7772$$

$= 0{,}02128 \times 500 = 11.$

Beraz, 11 jaioberri egongo dira 64 edo pisu gutxiagorekin.

Adibidea. Banaketa normala estandarra: z balio negatiboen probabilitateak kalkulatzea: Soziologiako lehen mailako ikasleek batez beste 8 euro ekartzen dituzte ikasgelara, eta desbideratze tipikoa 1,5 eurokoa da, [N (8,1.5)]. Kalkulatu ikasgelara 6 euro edo gutxiago ekartzeko probabilitatea:

$P\ (x \leq 6) = P\ (z \leq -\ 1{,}33) = P\ (z \geq +\ 1{,}33) = 1 - P\ (z \leq +\ 1{,}33) = 1 - 0{,}90824 = 0{,}091776$

Kasuen %9,1 soziologiako ikasleek 6 euro edo gutxiago ekarriko dituzte ikasgelara.

Adibidea. Banaketa normala estandarra: z balio negatiboen probabilitateak kalkulatzea: zenbatekoa da z = – 0,3 edo txikiagoa izateko probabilitatea?

$P\ (z \leq -\ 0{,}3) = P\ (z \geq +\ 0{,}3) = 1 - P\ (z \leq 0{,}3) = 1 - 0{,}617911 = 0{,}382089$

Beraz, norbaitek z = -0,3 edo gutxiago ateratzeko probabilitatea 0,38koa da.

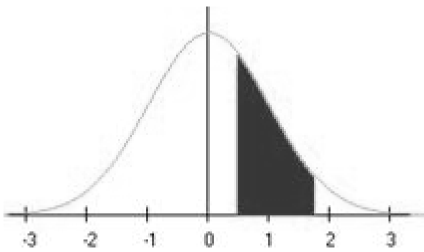
Adibidea. Banaketa normala estandarra: z balio negatiboen probabilitateak kalkulatzea: zenbatekoa da z = – 0,2 baino handiagoa izateko probabilitatea?

$P(z \le -0{,}2) = P(z \ge +0{,}2) = 1 - P(z \le +0{,}2) = 1 - 0{,}5793 = 0{,}4207$

2.4.4.4. Bi balio positiboren arteko probabilitateak kalkulatzea

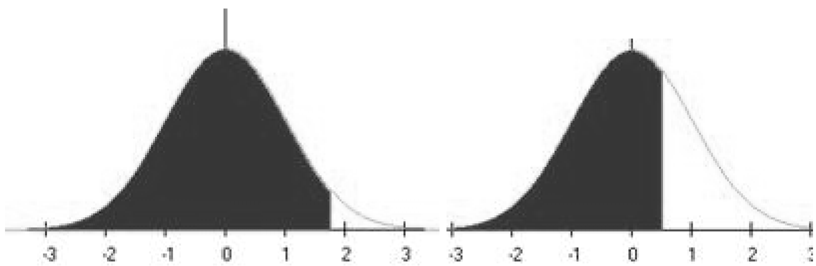
Bi balio positiboren arteko probabilitateak kalkulatzeko, balio horien probabilitateak taulan bilatzen ditugu, eta probabilitate bi horien arteko diferentzia da bilatzen ari garen probabilitatea. Alegia, guk bilatzen dugun probabilitatea beltzean dagoen azalera horri dagokiona:

52. diagrama: bi balio positiboren probabilitateak kalkulatzea



Azalera horri dagokion probabilitatea kalkulatzeko, z balio handienari dagokion azalerari z balio txikienari dagokion azalera kentzen diogu:

53. diagrama: bi balio positiboren probabilitateak kalkulatzea



Adibidea. Banaketa normala estandarra: bi balio positiboren arteko probabilitateak kalkulatzea: Zenbatekoa da z = 0,5 eta z = 1,76 artean dagoen balioa aurkitzeko probabilitatea?

$P(0{,}5 \le z \le 1{,}76) = P(z \le 1{,}76) - P(z \le 0{,}5) = 0{,}9608 - 0{,}6915 = 0{,}2693$

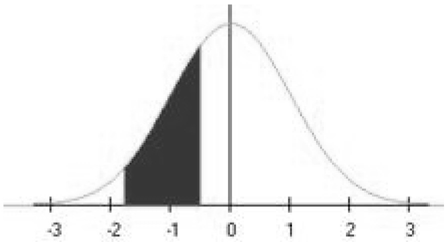
Adibidea. Banaketa normala estandarra: bi balio positiboren arteko probabilitateak kalkulatzea: zenbatekoa da z = 1 eta z = 2 artean egongo den balioa aurkitzeko probabilitatea?

$P(1 \le z \le 2) = P(z \le 2) - P(z \le 1) = 0{,}977250 - 0{,}841345 = 0{,}135905$

2.4.4.5. Bi balio negatiboren arteko probabilitatea kalkulatzea

Bi balio negatiboen arteko probabilitateak kalkulatzeko, simetriaz, balio negatiboak positiboan eraldatzen ditugu, eta probabilitateak kalkulatzen.

54. diagrama: Bi balio negatiboren arteko probabilitatea kalkulatzea

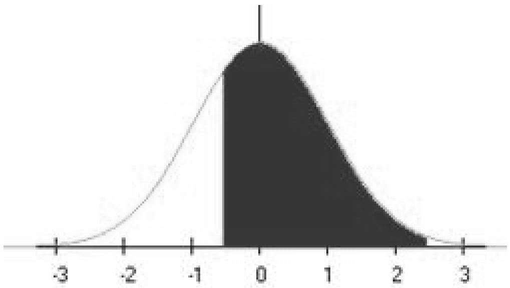


Adibidea. Banaketa normala estandarra: bi balio negatiboren arteko probabilitateak kalkulatzea: zenbatekoa da z = - 1,76 eta z = - 0,5 artean egongo den balioa aurkitzeko probabilitatea? $P(-1{,}76 \leq z \leq -0{,}5) = P(0{,}5 \leq z \leq 1{,}76) = 0{,}9608 - 0{,}6915 = 0{,}2693$

2.4.4.6. Balio positibo baten eta negatibo baten arteko probabilitateak kalkulatzea

Kasu honetan, z balio positiboari dagokion probabilitateari z balio negatiboari dagokion probabilitatea kentzen diogu. z balio negatiboei dagozkien probabilitateak aurkitzeko, banaketa normalaren simetrian oinarrituko gara.

55. diagrama: balio positibo baten eta negatibo baten arteko probabilitateak kalkulatzea



Adibidea. Banaketa normala estandarra: Balio positibo baten eta negatibo baten arteko probabilitateak kalkulatzea: zenbatekoa da z = - 0,53 eta z = 2,46 artean egongo den balioa aurkitzeko probabilitatea?

$$P(-0{,}53 \leq z \leq 2{,}46) = P(z \leq 2{,}46) - P(z \leq -0{,}53)$$
$$P(z \leq -0{,}53) = P(z \geq 0{,}53) = 1 - P(z < 0{,}53) = 1 - 0{,}7019 = 0{,}2981$$
$$P(-0{,}53 \leq z \leq 2{,}46) = P(z \leq 2{,}46) - P(z \leq -0{,}53) = 0{,}9931 - 0{,}2981 = 0{,}695$$

Kasuen %69,5ean, z = -0,53 eta 2,46 artean egongo den z balio bat aurkituko dugu.

Adibidea. Banaketa normala estandarra: z balio positibo baten eta negatibo baten arteko probabilitateak kalkulatzea:

Hiri batean irailean zehar izaten den tenperaturak banaketa normalari jarraitzen dio, eta N (23º, 5º) berezitasunak ditu. Kalkulatu 21º eta 27º arteko tenperaturak izango dituzten egunen kopurua.

$P(21 \leq x \leq 27) = P(-0{,}4 < z < 0{,}8) = P(z \leq 0{,}8) - [1 - P(z \leq -0{,}4)] = 0{,}7881 - (1 - 0{,}6554) = 0{,}7881 - 0{,}3446 = 0{,}4435.$

Kasuen %44,35ean, 21º eta 27º arteko tenperaturak egongo dira. Hilak 30 egun dituela onartuta, 30 x 0,4435 egiten badugu, guztira, 13 egunetan izango dira tenperatura horiek.

Adibidea. Banaketa normala estandarra: balio positibo baten eta negatibo baten arteko probabilitateak kalkulatzea: adimen-test batean lortutako puntuazioek banaketa normalari jarraitzen diote, N (100, 15) berezitasunekin. Populazioaren zer ehunekok izango du 110 eta 95 arteko adimena?

P $(95 \leq x \leq 110)$ = P $(- 0{,}33 < z < 0{,}67)$ = P $(z \leq 0{,}67)$ – [1 - P $(z \leq 0{,}33)$] = 0,7486 – (1- 0,6293) = 0,3779.

Populazioaren %37,79k izango du 95 eta 110 arteko puntuazioa.

Adibidea. Banaketa normala estandarra: Balio positibo baten eta negatibo baten arteko probabilitateak kalkulatzea: 800 jaioberriren altuerak normal banatzen dira, N (66, 5) berezitasunekin. Kalkulatu zenbat jaioberrik duten 65 zentimetrotik 70 zentimetrora arteko luzera.

Eragiketak egin ostean, lortzen dugun probabilitatea 0’3674 da. 800 jaioberri direnez, 0’3674 x 800 = 294 jaioberri espero ditzakegu 65 eta 70 zentimetro arteko altuerarekin.

2.4.5. Taula alderantziz erabiltzea

Balio bati dagokion probabilitatea ematen badigute, erraza da probabilitate horri dagokion z balioa ezagutzea.

Adibidea. Banaketa normala estandarra: probabilitate batetik z balioa ezagutzea: demagun probabilitate hau dugula: P $(z \leq k)$ = 0,7019
Probabilitate hori taulan bilatzen dugu, eta dagokion zenbakia lehen zutabean 0,5 dela ikusten dugu eta lehen errenkadan, 0,003. Beraz, probabilitate horri dagokion z balioa 0,53 da.
P (k) = 0,7019 ⟶ k = 0,53

2.5. Estatistika deskribatzailea eta txostenaren idazketa

Estatistika deskribatzaileak datu multzoak modu ulerterrazean aurkezteko hiru baliabide eskaintzen dizkigu: maiztasun-taulak, datu-diagramak eta neurri estatistikoak. Horiek modu egokian erabilita, balioko ikerketa deskribatzaileak egingo ditugu; kontuan izanda, betiere, baliabide horien erabilera desberdina dela aldagaia neurtzeko erabili dugun neurketa-eskalaren arabera.

79. taula: maiztasunak, neurri estatistikoak eta datu-diagramak, aldagai motaren arabera

Neurketa-eskala	**Maiztasunak**	**Zentralizazio-neurriak**	**Sakabanatze-neurriak**	**Kokapen-neurriak**	**Forma-neurriak**	**Kontzentrazio-neurriak**	**Datu-diagramak**
Nominala	absolutuak erlatiboak	moda	ez	ez	ez	Ez	sektore-diagrama, barra-diagrama, piktograma eta kartograma
Ordinala	absolutuak erlatiboak	moda mediana (zentzua badauka)	ez	ez	ez	Ez	sektore-diagrama, barra-diagrama, piktograma eta kartograma
Tarte-eskala	absolutuak erlatiboak metatuak	moda batez bestekoa mediana	bai	bai	bai	Bai	sektore-diagrama, piktograma, barra-diagrama, kartograma, histograma, maiztasun bakunen poligonoa, Ojiba edo maiztasun metatuen poligonoa, lerro-diagrama eta kaxa-diagrama

Arrazoi eskala	absolutuak erlatiboak metatuak	moda batez bestekoa mediana	bai	bai	bai	Bai	sektore-diagrama, piktograma, barra-diagrama, kartograma, histograma, maiztasun bakunen poligonoa, Ojiba edo maiztasun metatuen poligonoa, lerro-diagrama eta kaxa-diagrama
Iturria: lanketa propioa							

Maiztasun-taula, datu-diagrama eta neurri estatistikoekin batera, ikerketa deskribatzaileak egiteko lagungarri oso ona da banaketa normalaren berezitasunak ezagutzea. Izan ere, gure datu multzoa banaketa normalaren eredu teorikora hurbiltzen dela frogatzen badugu, lan estatistikoa izugarria errazten da. Hain zuzen ere, “68-95-99,7 arauaren” arabera, badakigu banaketa normal bateko balio eta datu ia denak batez bestekotik hiru desbideratze estandarreko distantziaren barnean daudela. Berezitasun horri esker, batez bestekoa eta desbideratze tipikoa ezagututa, oso erraz ezagut dezakegu histogramako edo maiztasun-banaketako balio batetik gora, behera edo bi balioren artean datuen zer ehuneko dagoen.

Bestetik, banaketa normalaren estandarizazio edo tipifikazioaren bidez, datu multzo desberdinetako datuak alderatu ditzakegu, baita balio zehatzak aurkitzeko probabilitateak kalkulatu ere.

Maiztasun-taulak kudeatzeko, datu-diagramak egiteko eta neurri estatistikoak kalkulatzeko gaitasuna badugu, eta banaketa normalaren berezitasunak ezagutzen baditugu, txosten estatistiko deskriptiboak egiteko moduan gaude. Txosten horiek egiterakoan kontuan izan formula magikorik ez dagoela. Hori bai, egiten ditugun txostenak, ulergarriak izateaz gain, argi adierazi behar dute zein diren eman diren pausoak, hartu diren erabakiak (laginaren tamaina, erabilitako teknika...) eta lortu diren emaitzak. Txarto idatzitako eta egituratutako txostenek ikerketarik onena alferrik galtzen dute.

Ondo egituratutako txostenek aurkibideak, sarrera, erabilitako teknika eta erabaki metodologikoei buruzko azalpenen atala, aurkikuntzen aurkezpenen atala, ondorioak eta eranskinak izan behar dituzte:

1. **Aurkibidea**: txostenean aurkituko ditugun gai, baliabide edo atal guztien orrialdeak adierazten dituen zerrenda da. Txostenaren aurrean edo atzean jar daiteke. Hori bai! aurkibideek edukiak zein orrialdetan aurkituko ditugun adierazi behar dute. Horrek esan nahi du orrialderik gabeko aurkibideek ez dutela ezertarako balio. Ez ahantzi, bestalde, aurkibide mota desberdinak daudela: aurkibide orokorra da garrantzitsuena, baina maiztasun-taulen eta datu-diagramen aurkibideak ere sar ditzakegu.
2. **Sarrera**: txostenaren hasieran, ikergaiaren aurkezpen gisa egin ohi den azalpena da. Sarreran, ikerketaren helburu orokor zein zehatzak, hipotesi garrantzitsuenak, txostenak dituen atalak eta txostena behar bezala ulertu eta irakurtzeko interesgarriak izan daitezkeen aholkuak edota hartu diren erabakiak aurkeztu behar dira. Sarrera ikerketa egiten lagundu digutenei eskerrak emateko ere erabiliko dugu.
3. **Erabilitako teknika eta erabaki metodologikoei buruzko atala:** lan akademikoetan, informazioa biltzeko erabilitako teknikak eta teknika horiek aukeratzearen arrazoiei buruzko atal berezia sartu behar da. Horrekin batera, lagina nola aukeratu den, izan duen tamaina eta kanpo-lanak zer berezitasun izan dituen ere aurkezten da. Hau da, atal honetan, era honetako galderei erantzun behar zaie: nondik lortu dira datuak? Zer-nolako teknikak erabilita? Zergatik? Nolakoa izan da lagina? Nola antolatu da kanpo-lana?
4. **Aurkikuntzen aurkezpena**: teknika bakoitzak era bateko datuak ematen dizkigu. Teknika kuantitatiboak erabili ditugunean (inkesta edo behaketa zuzena, adibidez), lortutako emaitzak maiztasun-banaketen, datu-diagramen eta neurri estatistikoen bidez aurkezten ditugu. Emaitzak agertzerakoan, garrantzitsua da honako aholku hauek kontuan hartzea:

- Maiztasun-taulak aurkezterakoan, maiztasun erlatiboak hobesten dira, azalpen ulergarriagoak ematen dituztelako. Hala ere, maiztasun absolutu totalak edo, beste era batera esanda, erabilitako laginaren tamaina adierazi behar da. Horrela, estatistikekin gezurrak esatea edo informazioa manipulatzea ekiditen dugu. Adibidez, egunkari bateko izenburuak dio herri bateko biztanleriaren %80 alkatearen aurka dagoela; baina ez du esaten galdeketa 10 pertsonaren artean egin dela. Arestian esan bezala, kontuz ibili laginaren tamainari buruzko informaziorik ematen ez duten ikerketekin.
- Datu-diagramek txostenak arintzen dituzte. Hala ere, ikerketa-txostenetan ez dira sartu behar datu multzo bati buruzko maiztasun-taula eta datu-diagrama, biak batera; hain zuzen ere, errepikakorra delako.
- Kontuz hiru dimentsiotako maiztasun-taula eta grafikoekin, baliabide estilistiko sofistikatuegiekin edota gehiegizko koloreekin. Askotan, era horretako baliabideak ulerterraztasunaren aurka joan daitezke. Diagrama edo maiztasun-taula bat txosten batean sartu aurretik, ondo begiratu ia ondo ulertzen den.
- Maiztasun-taula eta datu-diagrama guztiek izenburua, identifikazio-zenbakia eta iturria adierazi behar dute.

5. **Ondorioak**: aurkikuntzak aurkezteaz gain, lortutako emaitzak gaiari buruzko teoria orokorragoekin jarri behar ditugu harremanetan. Era berean, emaitzak hasieran formulatutako hipotesiekin alderatu behar ditugu, hipotesi horiek onartzeko edo alboratzeko.
6. **Bibliografia**: lana egiterakoan erabilitako liburu, artikulu eta txostenen berri emango dugu. Era berean, interesgarria izan daiteke erabili ez diren baina gaiarekin estuki loturik dauden liburu eta txostenak aurkeztea.

Adibidea. Bibliografia nola aurkeztu: liburuak eta egunkari, aldizkari eta liburuetan argitaratutako artikuluak aipatzeko aholku batzuk.

- Liburuak aipatzean, autorearen izena, liburuaren argitaratze-urtea, izenburua (letra etzanarekin edo azpimarratuta) eta edizioaren berezitasunak (argitaletxea eta argitaratu den herria) aipatu behar dira: Juaristi, P. (2001): Euskal Herria Globalizazioaren aurrean. BBK-Euskaltzaindia, Bilbo.
- Liburuetan argitaratutako artikuluak aipatzean, artikulua idatzi duenaren izena, aldizkaria argitaratu den urtea, artikuluaren izenburua (kakotx artean), liburua editatzen dutenen edo biltzaileen izenak, liburuaren izenburua (letra etzanarekin edo azpimarratuta), edizioaren berezitasunak (argitaletxea eta argitaratu den herria) eta artikulua agertzen den orrialdeak: Juaristi, P. (1995): “El análisis de contenido” in Elzo, J. (Zuzendaria): Usos del vizcaíno en la administración. Bizkaiko Foru Aldundia, Bilbo.
- Artikuluak aipatzean autorearen izena, aldizkaria zer urtetan argitaratu den, artikuluaren izenburua (kakotx artean), aldizkariaren izenburua (letra etzanarekin edo azpimarratuta) eta aldizkariaren zenbakia eta artikulua zein orrialdetan agertzen den: Juaristi, P. (2000): “Ondasunekiko harremanak Justo Mokoroaren Repertorio de Locuciones del Habla Popular Vasca esaera bilduman” in Vasconia, 28 zk., 97. orrialdetik 120.era.
- Egunkarietan argitaratutako artikuluak aipatzean, autorearen izena, artikulua zein egunetan argitaratu den, hila eta urtea, artikuluaren izenburua (kakotx artean), egunkariaren izena (letra etzanarekin edo azpimarratuta) eta artikulua zer orrialdetan agertzen den: Juaristi, P. (2000-10-22): “G. Bush eta A. Goren lehia”. Egunkaria, Andoain.

7. **Eranskinak:** eranskinetan, ikerketan erabili ditugun eta hura hobeto ulertzeko lagungarri izan daitezkeen materialak sartuko ditugu; besteak beste, ikerketaren fitxa teknikoa (informazioa bildu deneko datak, datuak biltzeko teknika, laginaren tamaina, laginketa-errorea, lagina aukeratzerakoan onartu den konfiantza-maila, ikerlarien izenak, q eta p-ren balioak...), erabilitako inkesta edo informazio biltzeko protokoloak, eta erabili ditugun material osagarri guztiak (mapak, inkestak egiteko erabilitako bide-sistema...).

Adibidea. Ikerketaren fitxa teknikoa: V. Inkesta Soziolinguistikoaren fitxa teknikoa.

- Egilea: Siadeco.
- Datu-bilketa: 2012ko ekainetik abendura bitartean. Telefono bidez egindako inkesta, galdetegi egituratua eta itxia erabiliz.
- Lagina: 16 urtetik gorako biztanleak.
- Inkesta kopurua: 7.800 inkesta egin ditu Siadeco enpresak. Horietatik, 4.100 Araban, Bizkaian eta Gipuzkoan egin ditu; Lapurdin, Zuberoan eta Nafarroa Beherean, 2.000; Nafarroan, berriz, 1.700.
- Ikerketa-eremuak: Lau eremu ikertu dituzte: herritarren hizkuntza-gaitasuna, hizkuntzaren transmisioa, euskararen erabilera hainbat esparrutan eta jarrerak.
- Akats-tartea: Euskararen lurraldeetako lagin osoaren akats-tartea 1,5ekoa da, eta 1,9koa erdaldun eta elebidunentzat. Konfiantza-maila, berriz, %95,5ekoa da.

3. ESTATISTIKA INFERENTZIALA

Bigarren atalean, maiztasun-taula, datu-diagrama eta neurri estatistikoen bitartez (zentro-neurriak, sakabanatze-neurriak, posizio-neurriak, itxura-neurriak eta kontzentrazio-neurriak) datu multzoen berezitasunak deskribatzeko gaitasuna lortu dugu.
Dena dela, estatistikak, datu multzoak deskribatzeaz gain, beste helburu batzuk ere baditu. Ikertu ez dugunari buruzko ondorioak ateratzea ere badu helburu. Alegia, estatistikaren egitekoen artean ere badago aldagai bat ikertuz aldagai horrekin etorkizunean gertatuko dena aurresatea; edo lagin baten berezitasunak aztertuz populazio osoari buruzko ondorioak ateratzea. Horretarako, baina, ezinbestekoa da inferentzia estatistikoaren prozedurak ezagutzea.
Inferentzia estatistikoa estatistikaren adar bat da, non, estatistika inferentziala edo estatistika induktiboa populazioaren lagin batetik abiatuz, populazio osoari buruzko ondorioak ateratzen baitiren, baita zorizko fenomenoetako ezaugarri ezezagunei buruzko ondorioak ateratzen dituena ere. Adibidez, inferentzia estatistikoaren bidez, hiri bateko biztanleen batez besteko errenta ezagut dezakegu, populazio horretatik lagin bat hartuz, eta laginean aukeratu direnei irabazten dutena galdetuz; beste alde batetik, estatu bateko jaiotzei buruz azken urteotako datuak bilduta, aldagai horrek etorkizunean izango duen portaera aurresan dezakegu.
Hala, inferentzia estatistikoak populazioak ikertzeko laginak diseinatzen ditu, estatistikoen bidez parametroak ezagutzeko prozedurak eskaintzen, parametro horiei buruzko konfiantza-tarteak osatzen eta estatistika-frogak egiten (non parametroei buruzko hipotesiak onartzen edo baztertzen diren). Horretarako, laginketari buruzko teorian, estimazioaren teorian, kopuru handien legean eta probabilitate teorian oinarritzen da.

3.1. LAGINA ETA LAGINKETA

Ikerketa ia guztien helburua zera da: populazioari edo unibertsoari buruzko informazio lortzea; hots, ikergai diren elementu, unitate edo pertsona guztiei (gizakiak, diskurtso politikoak, enpresak, estatuak, hiriak...) buruzko informazioa eskuratzea.
Ia gehienetan, populazioari buruzko informazioa lortzea ezinezkoa izaten da, populazioa osatzen duten elementu, unitate edo pertsona guztiak ikertzea oso garestia delako edo, besterik gabe, teknikoki oso zaila delako (estatu bateko biztanle guztiei galdetzea oso garestia da, hau da, horrelako zerbait egiteko azpiegitura eta baliabide izugarriak behar dira). Hala, Gizarte Zientzietan, ezinbestekoa da laginak erabiltzea; hau da, unibertso osoari buruzko ondorioak ateratzeko, unibertsoaren zati bat ikertu behar izaten da.

3.1.1. Zer da lagina?

Lagina populazio edo unibertsoaren zati adierazgarri bat da; haren ezaugarriak aztertuz, populazioaren berezitasunei buruz hitz egin dezakegu. Laginketa, beraz, laginak lortzeko prozedura da.
Laginen berezitasunak aurkezten hasi aurretik, argi utzi behar dugu laginak eta laginketa prozesua erabilgarriak direla, batez ere, inkestak bezala, zenbaketa onartzen duten teknika kuantitatiboetan. Kontuan izan laginak populazioaren zati adierazgarria izan behar duela, eta adierazgarritasuna lortzeko modu bakarra dela populazioaren berezitasun guztiak proportzionalki islatuko dituen populazioko elementu, unitate edo pertsona batzuk aukeratzea. Ildo horretatik, besteak beste, sakoneko elkarrizketa, bizitza-historia edo eztabaida-taldea teknikekin osotasunari buruzko ondorioak lor ditzakegu, baina ezin dugu adierazgarritasunik ziurtatu; hain zuzen ere, teknika horiekin ezin delako populazioaren berezitasunak proportzionalki islatu. Hala, sakoneko

elkarrizketetan, bizitza-historietan, eztabaida-taldeetan edo delphietan laginari buruz hitz egitea baino hobe da aztergai diren kasuei buruz hitz egitea.

3.1.2. Laginen abantailak eta desabantailak

Datu kuantitatiboetan oinarritzen den ikerketak tradizio handia du Gizarte Zientzietan. Izan ere, ikerlari askoren ustez, balioko ikerketa bakarrak dira laginketa-errorea eta fidagarritasuna neurtzea ahalbidetzen dutenak; hau da, laginak aztertuz populazioaren berezitasunak ezagutzea ahalbidetzen dutenak. Dena dela, ikerlariak argi izan behar du laginen erabilerak, abantailekin batera, desabantailak ere badituela.

Laginen abantailak

- Laginak ikertzea populazio osoak ikertzea baino merkeagoa da.
- Laginak aztertuz, bizkorrago lortzen dugu informazioa; laginak aztertzeko populazioak aztertzeko baino denbora gutxiago behar dugulako.
- Batzuetan, populazio osoen ikerketetatik baino informazio ulergarriagoa lortzen da laginen azterketatik. Laginketa-plan on batek zentsu eta errolden azterketek baino estimazio hobeak eman ditzakete; era horretako inkesta erraldoiek errore, akats edota soslai ugari izaten baitituzte, sarri askotan.

Laginen desabantailak

- Desabantailarik handiena: askotan laginen bitartez lortutako ondorioak eta emaitzak populazio osoaren neurketatik lortzen ditugunak baino zehaztasun gutxiago izaten dutela. Ezin dugu ahantzi, laginetan oinarritutako ikerketetan laginketa-errorea dagoela (estatistikoaren eta parametroaren artean dagoen desberdintasunari dagokiona), eta errore mota hori handitu egiten dela zenbat eta lagina txikiagoa izan.
- Askotan, arazoak izaten dira laginean aukeratu diren elementu, unitate edo pertsonak aurkitzeko edota aztertzeko; besteak beste, haien helbideak txarto ditugulako, telefono-zenbakia aldatu dutelako, edo, besterik gabe, azterketan parte hartu nahi ez dutelako. Horrelako kasuetan, diru eta denbora asko gastatu behar izaten da laginerako aukeratu diren elementu, unitate edo pertsonak bilatzen.

3.1.3. Laginketa motak

Laginak sailkatzerakoan, irizpide desberdinak daude. Dena dela, laginketa probabilistikoen eta laginketa ez-probabilistikoen artean bereizten duena erabiltzen da gehien.

3.1.3.1. Laginketa probabilistikoa

Laginketa probabilistikoa (zorizko edo ausazko laginketa) egiteko, ezinbestekoa da populazioa osatzen duten elementu, unitate edo pertsona guztien zerrenda lortzea. Zerrenda horretatik abiatuz, lagina osatuko duten elementu, unitate edo pertsonak aukeratuko ditugu, ausaz. Hala, populazioa osatzen duten elementu, unitate edo pertsona guztiek laginean aukeratuak izateko probabilitatea bera dute. Era horretako laginketetan kalkula dezakegu laginak zer-nolako zehaztasunarekin islatzen dituen populazioaren berezitasunak. Alegia, laginketa-errorea ezagut dezakegu. Geroago landuko dugu laginketa-errorearen kontzeptua, baina gogoratu etxez etxeko inkestetan, gutxi gorabehera ±%5eko edo gutxiagoko laginketa-erroreak dituztenak direla lagin onak. Telefono bitartez egindako inkestek laginketa-errore handiagoak onartzen dituzte.

Oro har, bost eratako laginketa probabilistikoak daude: ausazko laginketa sinplea, laginketa sistematikoa, laginketa estratifikatua, konglomeratu bitartez egindako laginketa eta laginketa telefonikoa.

Ausazko edo zorizko laginketa sinplea: ausazko laginketa sinplea erabiltzen da gehien laginketa probabilistikoen artean. Era horretako laginketa egiteko, populazioa osatzen duten elementu, unitate edo pertsona guztien zerrenda lortu behar dugu. Horrekin batera, lagina osatuko dutenak aukeratzeko teknikak ziurtatu behar du populazioa osatzen duten elementu, unitate edo pertsona guztiek dutela aukeratuak izateko probabilitate bera. Ausazko laginketa sinplea egiteko, badaude zenbait teknika:

- Zerrenda osatzen duten elementu, unitate edo pertsona guztiak loteria-ontzi batean sartzen ditugu, eta lagina osatzen dutenak aukeratzen ditugu, ausaz.
- Zerrenda osatzen duten elementu, unitate edo pertsona guztiei zenbakiak ematen dizkiegu, eta, ausazko zenbakien taula erabiliz, lagina osatuko dutenak aukeratzen. Teknika hori erabiltzeko, beraz, aldez aurretik ausazko zenbakien taula prestatu behar dugu, eta lagina osatzeko taulak ematen dizkigun zenbakiekin lotuta dauden elementu, unitate edo pertsonak aukeratzen ditugu.

Adibidea. 80. taula: ausazko zenbakien taula

02	22	34	76	334	84	110	887	347	410	625	30	730	66	634	646	725	830
85	3	23	88	456	73	132	546	234	532	756	29	269	78	556	678	925	729
00	4	94	29	56	14	566	51	447	666	53	94	545	19	756	310	567	694
64	5	42	03	78	12	159	71	878	759	80	142	142	93	87	332	242	842
94	67	09	28	134	109	234	138	557	834	137	09	09	38	134	366	278	709
42	180	54	27	51	154	345	50	899	945	159	154	154	37	951	459	746	454
09	62	25	68	169	105	456	16	416	556	123	25	25	78	769	987	987	325
54	56	28	04	67	8	635	6	65	835	65	28	28	64	24	677	765	328
25	1	64	58	135	145	678	35	356	978	130	164	164	158	735	231	988	364
98	94	61	35	567	77	889	447	477	289	546	161	161	139	267	876	459	461

Kalkulagailu batzuek RAND, RAN#, edo RANDOM izeneko tekla izaten dute. Tekla hori sakatutakoan, 0 eta 999 edo 0,000 eta 0,999 zenbakien arteko zenbaki aleatorio bat ateratzen da. Hala, kalkulagailua erabiliz, ez dugu zenbaki aleatorioen taula erabili beharko.

Laginketa sistematikoa: laginketa sistematikoetan, lagina osatuko duten elementu, unitate edo pertsonak ez dira ausaz aukeratzen, sistematikoki baizik. Hau da, populazioaren zerrendatik k elementu, unitate edo pertsona aukeratzen da, non k populazioa laginaren kopuruarekin zatitzean lortzen dugun zenbakia baita ($k = \frac{N}{n}$). Zerrendaren zein tokitan hasi behar dugun jakiteko, zerrendako lehenengo 50 elementu, unitate edo pertsonaren artetik, bat aukeratzen dugu, ausaz. Laginketa sistematikoa oso erabilgarria da populazio oso handiak ditugunean.

Laginketa estratifikatua: laginketa mota horren abiapuntua zera da: zenbat eta populazioa homogeneoagoa izan, orduan eta errazagoa da populazioaren isla izango den lagina lortzea. Ildo horretatik, laginketa estratifikatuetan, populazioa estratu edo multzoetan banatzen da, eta, ondoren, ausazko laginketa sinplea edo laginketa sistematikoa aplikatzen da multzo bakoitzaren barnean, lagina osatuko duten elementu, unitate edo pertsonak aukeratzeko. Laginketa estratifikatua bi eratakoa izan daiteke:

- Laginketa estratifikatu proportzionala: estratu edo multzo bakoitzean hartzen den elementu, unitate edo pertsona kopurua estratuaren edo multzoaren tamainaren araberakoa da.
- Laginketa estratifikatu ez proportzionala: multzo bakoitzak duen tamaina kontuan hartu gabe, multzo bakoitzetik elementu, unitate edo pertsona kopuru bera hartzen da.

Adibidea. 600 ikasle dituen fakultate batean, 20 ikasleko lagina hartu nahi da. Fakultate horretan, lau gradu eskaintzen dira: Soziologia (200 ikasle), Politika Zientziak eta Kudeaketa Publikoa (150 ikasle), Ikus-entzunezko Komunikazioa (150 ikasle) eta Publizitatea eta Harreman Publikoak (100 ikasle). Laginketa estratifikatu proportzionala egin gura badugu, zenbat ikasle hartu beharko ditugu gradu bakoitzetik 20 ikasleko lagina osatzeko? Hiruko erregela erabiliz, Soziologian, 7 ikasle hartuko ditugu:

$$\frac{20.200}{600} = 6, \approx 7$$

Modu berean, Politika Zientziak eta Kudeaketa Publikoko 5 ikasle hartuko ditugu, Ikus-entzunezko komunikazioko beste 5 eta Publizitatea eta Harreman Publikoko 3.

Konglomeratu bitartez egindako laginketa: populazioa osatzen duten elementu, unitate edo pertsona guztien zerrendarik ez dugunean, edo zerrenda egitea garestiegia denean, eta, bestalde, populazioa konglomeratu edo talde naturaletan banatuta dagoenean (eskoletako gelak, kirol-taldeak, lantegiak...), konglomeratu bitartez egindako laginketa erabil dezakegu. Teknika horren bitartez, konglomeratu guztien zerrenda prestatzen da, eta, ondoren, konglomeratu batzuk aukeratzen dira, ausaz. Ikerketa egitean, konglomeratua osatzen duten elementu guztiak har ditzakegu, edo konglomeratuaren barnean ausazko laginketa sinplea edo laginketa sistematikoa aplikatu. Azken laginketa horri laginketa polietapikoa deitzen zaio, zenbait etapatan egiten delako.

Laginketa telefonikoa: laginketa telefonikoak gero eta arruntagoak dira gure artean. Etxe gehienetan telefonoa dagoenez, gero eta jende gehiagok sakelako telefonoa duenez eta etxeetara sartzea gero eta zailagoa denez (atezain automatikoek edota inkestatzaileekiko mesfidantzak etxez-etxeko inkestak egitea zailtzen dute), asko ugaritu dira telefono bitartezko inkestak.

Hala eta guztiz ere, kontuan izan behar da lagina aukeratzeko zerrenda telefonikoak erabiltzeak ez duela, kasu askotan, lagin adierazgarriak ziurtatzen. Bi arrazoi daude hori esateko:

- Pertsona batzuek ez dute telefonorik (jende oso pobreak, adibidez) edo ez dira zerrenda telefonikoetan agertzen (jende oso aberatsa, politikariak...); eta ez dute laginean aukeratuak izateko inolako probabilitaterik.
- Zenbait pertsona behin baino gehiagotan agertzen dira telefono-zerrendetan (dendariak edo lanbide liberalak dituztenak, adibidez), eta, horren ondorioz, behin agertzen direnek baino probabilitate gehiago dute aukeratuak izateko.

Gaur egun, teknika automatikoak sortu dira telefonoa duten pertsonekin harremanetan jartzeko, telefono-zerrendetan agertu edo ez. Era horretako teknikek telefono-zenbakien lagina aukeratzen dute automatikoki, kontuan hartu gabe telefono-aurkibidean agertzen diren ala ez.

Era horretako prozedurek telefono bitartezko inkestak egiteko aurrerapausoa badira ere, ez dute inola ere konpontzen telefonorik ez duten edo telefono bat baino gehiago dutenen arazoa.

Laginketa-errorea ezagut dezakegun heinean, laginketa probabilistikoak laginketa ez-probabilistikoak baino nahiago izaten dituzte ikerlariek. Dena dela, kasu batzuetan, ezin izaten da laginketa probabilistikorik egin. Beraz, populazioaren zerrenda lortzea oso garestia denean, ikergai den populazioa zein den zehaztea zaila denean (alderdi politiko bati bozka ematen dioten guztien zerrendarik ez dago) edo ikerlariek lagin adierazgarria lortu beharrean kasu zehatz batzuk aztertuz informazio gehiago aterako dutela pentsatzen dutenean (lider politiko zehatz batzuekin sakoneko elkarrizketak eginez, adibidez), laginketa ez-probabilistikoak egiten dira.

3.1.3.2. Laginketa ez-probabilistikoa

Gizarte-ikerlari batzuen ustez (zehatzago esanda, positibisten ustez), laginketa probabilistikoa da balioko laginketa bakarra. Horien ustez, zerrendatuta ditugun fenomenoak azter daitezke bakarrik; alegia, laginketa-errorea ezagutzea ahalbidetzen duten fenomenoak.
Ikuspuntu horri jarraituta, gai asko ikertu gabe utzi behar genituzke. Izan ere, askotan ezin izaten da laginketa probabilistikorik lortu, populazioa osatzen duten elementu, unitate edo pertsona guztien zerrendarik ez dagoelako (drogamendekotasuna, homosexualitatea, gaixotasun berezi bat dutenena...) edo zerrenda lortzea asko kostatzen delako.
Hala, gure ikuspuntua da zerrendarik ez dugun kasuetan ikerketa oso zehatzak egin daitezkeela, baina probabilitatezkoa ez den laginketetan oinarritu behar dugula. Laginketa-errorea ezagutzea ezinezkoa bada ere, lagin adierazgarriak lor ditzakegu.
Era horretako laginketetan ez dakigu populazio osatzen duten elementuek laginean aukeratuak izateko duten probabilitatea. Populazioaren zerrendarik ez dagoenez, lagin mota hori ikerlariaren interesen, nahien edo aukeren arabera osatzen da. Era askotako probabilitate-laginketak daude:
Iritzi-laginketa: ikerlariak gaiari buruz duen ezagutzaren arabera, subjektiboki aukeratzen ditu lagina osatuko duten elementu, unitate edo pertsonak. Beraz, talde txiki eta askotarikoa aztertzen da, gaiaren inguruan egon daitezkeen iritzi eta ikuspuntuak sakonean ezagutzeko. Egia esan, kasu horretan, laginari buruz hitz egitea baino hobe da aztergai diren kasuen multzoari buruz hitz egitea. Arestian esan bezala, laginak populazioaren berezitasun guztiak proportzionalki islatzeaz gain, populazioaren adierazgarri izan behar du, eta iritzi-laginketetan, lagin adierazgarria lortu baino gehiago, gai bat sakonean ezagutzea nahi izaten da. Laginketa mota hori sakoneko elkarrizketetan, bizitza-historietan, delphietan eta eztabaida-taldeetan erabiltzen da nagusiki.
Komenientziazko lagina: ikerlariak oso hurbil edo eskuragarri dituen elementu, unitate edo pertsonak sartzen ditu laginean; berarengandik oso hurbil daudenak edo ikerketan parte hartzeko prest daudenak. Laginketa mota hori esplorazio-ikerketetan erabiltzen da, edo ikergai den populazioa definitzea edo bilatzea oso zaila denean. Ez ahaztu era horretako laginekin gure ikerketaren emaitzak bideratuko dituen lagin soslaitua hartzeko arrisku handia dagoela.
Kuota bitartez egindako laginketa: kasu horretan, elementu, unitate edo pertsonak populazioan duten adierazgarritasunaren arabera aukeratzen dira; beraz, laginketa estratifikatu proportzionalaren antzerakoa da. Kuota bitartez egindako laginketen eta laginketa estratifikatu proportzionalen arteko desberdintasuna da kuota bitartez aukeratutako lagina ez dela probabilitatez aukeratzen. Kasu horretan, elementu, unitate edo pertsonak aldez aurretik zehaztu diren berezitasunen arabera (sexua, adina, bizitokia...) aukeratzen dira, ausaz. Hau da, zoria ziurtatzen duen ibilbideari jarraituz, berezitasun zehatz batzuk dituzten pertsonak aurkitu eta inkestatu behar izaten dituzte inkestatzaileek. Era horretako ikerketak egiten direnean, populazioaren tamaina ezagututa, hartu nahi dugun laginaren tamaina jakinda, eta lan egiteko nahi dugun konfiantza-maila zehaztuta, aldez aurretik zehaztu dezakegu laginketa-errorea.
Elur-bolaren bitartez egindako laginketa: laginketa mota horretan laginean parte hartzen duten elementu, unitate edo pertsonek beren berezitasunak dituzten beste batzuk ezagutzeko informazioa ematen digute. Elkarrizketatzen den pertsona bakoitzari ikerketan parte hartzeko baldintzak betetzen dituzten beste pertsona batzuk ezagutzen dituen galdetzen zaio, eta, apurka-apurka, lagina handituz eta osatuz joaten da. Hau da, elur-bola handituz doa, behar adina pertsona elkarrizketatzen den arte. Era horretako laginketa oso erabilgarria da populazio oso arraroak, berezitasun oso zehatzak dituztenak edo populazio hori osatzen dutenengana hurbiltzeko arazoak ditugunean.
Laginketa mota guztiak aurkeztu ondoren, argi utzi nahi dugu oso arrunta dela laginketa probabilistikoak eta ez-probabilistikoak nahastea; hau da, askotan, laginketa mota desberdinetan oinarritutako ikerketak egiten dira. Adibidez, hiri batean dauden barruti guztien zerrendatik barruti

batzuk aukeratzen dira, ausaz (konglomeratuetan oinarritutako laginketa), eta ondoren, aukeratutako barrutietan kuota-sistemaren bitartez aukeratzen dira inkestatuak (kuota bitartez egindako laginketa).

3.2. LAGINEN ERABILERAREN ONDORIOZ SORTZEN DIREN ERROREAK

Ausazko erroreak laginak erabiltzearen ondorioz, zoriz agertzen diren erroreak dira; hain zuzen ere, laginak populazioak baino txikiagoak direlako. Ausazko erroreak neurtu egin daitezke, eta garrantzitsuena laginketa-errorea (estimazioaren errorea) da; hau da, estatistikoaren eta parametroaren artean dauden desberdintasunei dagokiona.
Kontuan hartu behar dugu laginketa-errorea gutxitu egiten dela laginaren tamaina handituta. Alegia, zenbat eta lagin handiagoa izan, orduan eta laginketa-errore txikiagoa izango dugu.
Ia gehienetan, laginketa-errorea ezezaguna izaten da. Beraz, laginketa-errorea aldez aurretik zehaztu behar izaten dugu, hartuko dugun laginaren tamainaren arabera. Zehatzago esanda, laginaren tamaina kalkulatzerakoan, aldez aurretik zehaztu behar dugu zer-nolako laginketa-errorea onartzeko prest gauden.
Gaur egun, laginketa-errorea automatikoki kalkulatzen duten programa ugari aurkitu ditzakegu interneten. Konfiantza-maila, populazioaren eta laginaren tamaina, eta, gure ustez, populazioan ikergai den berezitasuna dutenen ehunekoa adierazita (ia beti ezezaguna izanik, %50 dela adierazten da) laginketa-errorea ezagut dezakegu.

Adibidea. Laginketa-errorea kalkulatzeko programa informatikoen eredua.

Laginketa-errorea kalkulatzea	
Konfiantza-maila	95% 99%
Laginaren tamaina	1000
Populazioa edo unibertsoa	600000
Ikergai den berezitasuna dutenen ehunekoa	50
Laginketa-errorea	0

Laginen erabilerarekin zerikusia duten erroreekin jarraituz, errore tipikoa aipatu beharra dago. Sakonago azalduko dugu aurrerago, baina kontuan izan errore tipikoak estatistikoa zenbateko zehaztasunarekin den parametroaren estimatzailea esaten digula. Hala errore tipikoa ezagututa, parametroa zein tartetan egon daitekeen ezagut dezakegu.
Ausazko erroreekin lotuta, errore sistematikoak ditugu; hau da, iturri metodologiko, tekniko edota estatistikoetatik datozenak. Errore horiek edozein inferentzia distortsionatzen dute; ezin daitezke neurtu, eta beren efektua pilatu egiten da. Adibidez, datuak biltzeko teknika bat edo beste aukeratzeak errore sistematikoak sor ditzake: biztanleen portaera sexualei buruzko informazioa atez ateko inkesta baten bidez egiteak errore sistematikoak sor ditzake. Izan ere, gai oso pertsonalei buruz hitz egiterakoan, ez da erraza etxeko atean ezezagun bati egia esatea. Esanahi bikoitzeko galderek, ondo ulertzen ez direnek, edo hitz zailegiak edo teknikoegiak dituztenek ere errore sistematikoak sortzen dituzte. Beren lana gaizki egiten duten inkestatzaileek ere errore sistematikoak sortzen dituzte: pertsona bat aukeratzeko prozedurari ondo jarraitzen ez badiote, galderak gaizki egiten badituzte edo eman behar dituzten azalpenak ematen ez badituzte (erantzunak bideratzen dituzten azalpenak, adibidez) errore sistematikoak sortzen dira. Errore sistematikoak gutxitzeko, erabili beharreko teknikak, prozedurak, galdeketak zein inkestatzaileak ondo prestatu edota trebatu behar dira.

3.3. PARAMETROA, ESTATISTIKOA ETA INFERENTZIA ESTATISTIKOA

Laginekin lanean ari garenean, ondo bereizi behar ditugu parametroa eta estatistikoa.
Parametroa populazio batean neurtutako aldagai estatistiko bati buruzko informazioa laburbiltzen duen balioa edo zenbakia da (batez bestekoa, mediana, moda, desbideratze tipikoa…). Parametroa ezezaguna izaten da, ia gehienetan, eta, horren ondorioz, estatistikoen bidez ezagutu behar izaten da, estatistika inferentzialaren prozedurak erabiliz. Aurretik esan bezala, parametroak hizki grekoz izendatzen dira. Adibidez, batez bestekoa parametroa denean, μ (mu) hizkiarekin adierazten da, eta, desbideratze tipikoa parametroa denean, σ (sigma) hizkiarekin.
Estatistikoa lagin batean neurtutako aldagai bati buruzko informazioa laburbiltzen duen zenbakia da (batez bestekoa, mediana, moda, desbideratze tipikoa…). Esan bezala, estatistikoen bidez, parametroa ezagutu nahi izaten da, eta, horretarako, estatistika inferentzialaren prozedurak erabiltzen dira. Populazioren batez bestekoa ezagutzeko, adibidez, lagineko batez bestekoa ezagutu behar izaten da.
Dakigun bezala, estatistikoak izendatzeko latindar letrak erabiltzen dira. Adibidez, batez bestekoa estatistikoa denean, $\overline{x}$ hizkia erabiltzen da, eta desbideratze tipikoa estatistikoa denean, s hizkia.
Non dago inferentzia estatistikoaren benetako arazoa? Laginen azterketatik lortzen ditugun estatistikoak (batez bestekoa, mediana, ehunekoak…) aldatu egiten direla lagin batetik bestera, eta, aldiz, horien bitartez ezagutu nahi ditugun parametroak ez direla aldatzen edota konstante mantentzen direla. Hau da, aukeratzen dugun laginaren arabera, estatistikoek balio desberdinak har ditzakete, nahiz eta parametroak beti balio bera duen. Hala, nola atera ditzakegu populazioari buruzko ondorioak lagin bat ikertuz? Soziologiako eta Politika Zientziak eta Kudeaketa Publikoko ikasketak egiten dituzten 2. mailako ikasleen artean lagin bat hartuz, eta aukeratu direnei adina galdetuz, gela osoaren batez besteko adina ezagut dezakegu? Zenbateraino da laginetik lortutako batez bestekoa klase osoaren batez bestekoaren adierazgarria? Galdera horiei erantzuteko estimazioaren teorian (parametroen estimazioa), kopuru handien legean eta erabakiaren teorian (estatistikoen adierazkortasun-frogak) oinarritu behar dugu.

3.4. INFERENTZIA KLASIKOA ETA TARTE ESTIMAZIOA

Esan bezala, laginen erabileraren ondorioz sortzen diren erroreak behar bezala kontrolatuz populazioaren parametro ezezagunak ezagutzea da estatistika inferentzialaren helburuetariko ba. Horretarako, ausaz lortutako n tamainako lagina aztertzen da, eta hortik populazioaren parametroak ezagutzen. Prozesu horri estimazioa deritzo. Populazioaren parametroen estimazioa hiru bidetatik egin daiteke:

- Puntu-estimazioa (inferentzia klasikoa): populazioaren parametroa estatistikoaren bitartez estimatzen denean.
- Tarte-estimazioa (estatistika bayestarra): populazioaren parametroa tarteen bidez estimatzen denean.
- Hipotesi-frogak edo hipotesi-kontrasteak: parametroari buruzko baieztapena egiten da, eta, ostean, baieztapen hori onartzeko edo alboratzeko erabaki estatistikoa hartzen.

Lehenengo, puntu- eta tarte-estimazioak landuko ditugu, eta, ostean, hipotesi-frogak.

3.4.1. Inferentzia klasikoa edo puntu-estimazioa

Inferentzia klasikoan, parametroa populazioaren lagin bat aztertu ostean lortzen dugun estatistikoaren bidez estimatzen da. Kasu horretan, estatistikoa parametroarekin berdintzen dugu: populazioaren batez bestekoa, mediana, desbideratze tipikoa edo beste edozein neurri

estatistiko eta laginarenak berdinak direla onartzen dugu. Adibidez, hiri bateko batez besteko adina (μ = parametroa) estimatzeko, hiri horretako biztanleen lagin adierazgarria hartzen dugu, eta batez besteko adina ($\bar{x}$ = estatistikoa) kalkulatzen. Ostean, laginean atera den batez bestekoa populazioaren batez bestekoa dela onartzen dugu; alegia, $\mu = \bar{x}$ dela esaten.
Puntu-estimazioa egiteko, neurri estatistikoak parametroaren estimatzaile soslairik gabea izan behar du (estimatzailearen eta parametroaren artean desberdintasunik ez dago), efizientea edo zehatza (estimatzaile bat beste bat baino efizienteagoa da, bere bariantza bestearena baino txikiagoa denean) eta konsistentea (lagina handitzean estimatzailea are eta gehiago hurbiltzen da parametrora).

3.4.2. Estatistika Bayestarra edo tarte-estimazioa

Estatistika Bayestarra edo tarte-estimazioa egiten dugula esango dugu, laginaren azterketatik lortzen dugun estatistikoan oinarrituz, parametroa zein tartetan egon daitekeen zehazten dugunean, konfiantza-maila jakinarekin.
Lagina ikertu ostean, estatistiko bat lortzen dugu. Dena dela, lagina zorizkoa denez (aldez aurretik ez dakigu zein diren laginean aukeratuko direnak, hau da, laginean sartzen diren elementu, unitate edo pertsonak aldakorrak dira), estatistikoak balio desberdinak izan ditzake; eta, horren ondorioz, parametroaren balioak tarteen bidez ematen dira.
Kasu horretan ere, laginetik ateratzen dugun estatistikoak soslairik gabea izan behar du (estimatzailearen eta parametroaren artean desberdintasunik ez dago), efizientea edo zehatza (estimatzaile bat beste bat baino efizienteagoa da, bere bariantza bestearena baino txikiagoa denean) eta konsistentea (lagina handitzean estimatzailea are eta gehiago hurbiltzen da parametrora).
Tarte-estimazioa ulertzeko, ezinbestekoa da laginketa-banaketei buruzko teoria, limitearen teorema zentrala, errore tipikoa eta estimazioaren teoria ulertzea.

3.5. LAGINKETA BANAKETAK, LIMITEAREN TEOREMA ZENTRALA ETA ERRORE TIPIKOA

Estatistikan, laginetan oinarritutako estimazioen oinarria estatistikoen laginketa-banaketen ezagutzan dago.

3.5.1. Laginketa-banaketak

Edozein estatistikok laginketa-banaketa bat du. Laginketa-banaketa hori tamaina bereko lagin infinituetatik lortutako estatistiko infinituekin osatuta dago.

Adibidea. Laginketa-banaketak: demagun gazteen kultura-elkarte bateko bazkideen batez besteko adina ezagutu nahi dugula. Populazioa elkarte horretako bazkide guztiak direla jotzen badugu (N = 1.700 bazkide), bazkideek elkarteko kide egitean betetzen duten fitxa aztertuz, batez besteko adina (μ = 24,5) eta batez besteko horren desbideratze tipikoa (σ = 6) jakin ditzakegu. Alegia, bazkideek emandako informazioaren bidez, parametroak ezagut ditzakegu. Badakigu, halaber, datu multzo horren maiztasun-banaketa gutxi gorabehera normal erakoa dela.
Demagun elkarte horretako kideen ausazko lagina aukeratzen dugula. Hain zuzen ere, n = 100-eko tamaina duen lagina, eta batez besteko adina kalkulatzen dugula, $\bar{x} = 22$.
Kasu horretan, argi dago laginketa-errorea 2,5ekoa dela; izan ere, /22 - 22,5/ = 2,5.
Demagun elkarteko bazkideen ausazko eta tamaina bereko lagin gehiago hartzen ditugula; n = 100-eko tamainako laginak, hain zuzen ere. Hala, $\bar{x}_1 = 21$, $\bar{x}_2 = 23{,}5$, $\bar{x}_3 = 24$, ..., $\bar{x}_n = 22$ bezalako batez bestekoak lortuko genituzke.

Hau da, ziur aski, aztertutako laginetan, adinaren batez besteko desberdinak lortuko genituzke, eta oso gutxitan agertuko litzateke populazioaren benetako batez bestekoa. Hala ere, argi dago, zoriz, populazioaren batez besteko bera duten laginak bilatzeko aukera gehiago izango dugula, zenbat eta tamaina bereko lagin gehiago aztertu.

Beraz, populazio batetik atera daitezkeen tamaina bereko lagin infinitu desberdin guztiak hartu, lagin horietan guztietan batez bestekoak atera, eta batez besteko horiek maiztasun-banaketa bezala datu-diagrama batean irudikatzen baditugu (batez besteko guztiekin histograma egiten badugu, alegia), batez bestekoen laginketa-banaketa lortzen dugu (gaztelaniaz, Distribución Muestral de Medias).

Proportzioekin, gauza bera egin dezakegu. Populazio batetik atera daitezkeen tamaina bereko lagin infinitu desberdin guztiak atera, aldagai baten inguruko proportzioak aztertu eta proportzio horiek maiztasun-banaketa gisa grafiko batean irudikatzen baditugu (histograma), proportzioen laginketa-banaketa lortzen dugu.

Beste estatistiko batzuekin (mediana, moda, kurtosia…) ere, era horretako laginketa-banaketak lor ditzakegu.

Estatistiko horiekin lortu ditugun laginketa-banaketak aztertzen baditugu, argi ikusten dugu normal erakoak direla; hau da, maiztasun-banaketa simetrikoak, mesokurtikoak eta kanpai-itxurakoak direla.

Hala, batez bestekoen laginketa-banaketak zera adierazten digu: lagin infinituetatik lortutako batez besteko infinituekin laginketa-banaketa bat osa dezakegu, non batez bestekoak beren batez bestekoaren inguruan biltzen diren (batez bestekoen batez bestekoa), eta beren desbideratze tipikoaren arabera sakabanatzen.

Modu berean, proportzioen laginketa-banaketak zera adierazten digu: lagin infinituetatik lortutako proportzio infinituekin laginketa-banaketa bat osa dezakegu, non proportzioak beren batez bestekoaren inguruan biltzen diren (proportzioen batez bestekoa), eta beren desbideratze tipikoaren arabera sakabanatzen.

Arestian esan bezala, batez bestekoak edo proportzioek bezala, edozein estatistikok (medianak, kurtosiak, modak, desbideratze tipikoak...) forma zehatza (gehienetan mesokurtikoa), batez bestekoa eta desbideratze tipikoa (= errore tipiko) dituen laginketa-banaketa du. Laginketa-banaketa hori tamaina bereko lagin infinituetatik lortzen ditugun estatistikoetatik lortzen dugu.

Estatistika inferentzialaren oinarrizko estrategia zera da, laginaren estatistikoak aztertuz populazioaren parametroak ezagutzea, laginketa banaketaren bidez. Hala, estatistika inferentzialean, hiru eratako banaketekin lan egiten dugu:

- Laginaren banaketa: enpirikoa da (errealitatean existitzen da), eta ezaguna da. Ikerlariak populazioaren berezitasunak ezagutzeko asmoz osatu duen laginari dagokion banaketa da. Hainbat aldiz esan bezala, laginaren azterketatik lortzen ditugun estatistikoak izendatzeko, latindar letrak erabiltzen dira. Lagineko batez bestekoa adierazteko, $\bar{x}$ hizkia erabiltzen da, eta desbideratze tipikoa adierazteko, s hizkia.
- Populazioaren banaketa: enpirikoa da (errealitatean existitzen da) eta ezezaguna. Estatistika inferentzialaren helburua da estatistikoen bidez, populazioaren banaketaren berezitasunak ezagutzea. Parametroa populazio bateko aldagai estatistiko bati buruzko informazioa laburbiltzen duen balioa edo zenbakia da (batez bestekoa, mediana, moda, desbideratze tipikoa…). Parametroak hizki grekoz izendatzen dira. Adibidez, batez bestekoa parametroa denean, μ (mu) hizkiarekin adierazten da, eta desbideratze tipikoa parametroa denean, σ (sigma) hizkiarekin.
- Laginketa-banaketa: teorikoa da (ikerlariak ez du ateratzen, eta ez da erreala). Laginetik abiatuz, laginketa-banaketen berezitasunak ezagutzen ditugu, eta, horren bidez, populazioaren berezitasunak. Laginketa-banaketen berezitasunak ezagunak direnez (beti dira normal erakoak; hau da, simetrikoak, mesokurtikoak eta kanpai-itxurakoak dira), parametroak ezagutzea ahalbidetzen digu. Laginketa-banaketen balioak izendatzeko, hizki

grekoak erabiltzen dira, baina, azpindize gisa, letra latindarrak jarriz. Adibidez, batez bestekoa $\mu_{\bar{x}}$ hizkiarekin adierazten da, eta desbideratze tipikoa, $\sigma_{\bar{x}}$ hizkiarekin. Proportzioen kasuan, batez bestekoa μ_p hizkiarekin adierazten da, eta desbideratze tipikoa, σ_p hizkiarekin

3.5.2. Limitearen teorema zentrala

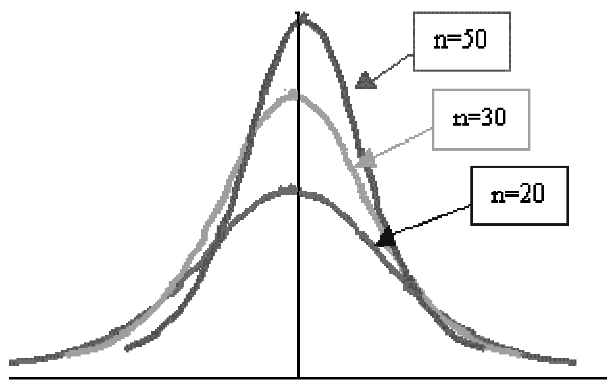
Zertan datza laginketa-banaketen erabilgarritasuna? Frogatuta dago populazioen maiztasun-banaketen eta laginketa-banaketen artean zenbait antzerakotasun daudela; horren ondorioz, laginketa-banaketen bidez populazioaren parametroak ezagut ditzakegu. Izan ere, frogatuta dago populazio batetik zenbat eta lagin gehiago hartu, lortzen ditugun batez besteko guztien batez bestekoa orduan eta gehiago hurbiltzen dela parametrora.

Gazteen kultura-elkarteko kideen (N = 1700) populaziotik zenbat eta lagin gehiago lortu, lagin horien batez bestekoen batez bestekoa orduan eta gehiago hurbilduko da parametrora (μ = 24,5). Horrenbestez, ateratzen dugun ondorioa zera da: laginketa-banaketaren batez bestekoa populazioaren benetako batez bestekora hurbiltzen dela. Alegia, lagin guztien batez bestekoen batez bestekoa = μ = parametroa.

Laginen batez bestekoekin lortzen dugun laginketa-banaketaren desbideratze tipikoa, errore tipikoa delakoa, populazioaren benetako desbideratze tipikoa baino txikiagoa da beti (gure kasuan, 6 baino txikiagoa). Horrenbestez, ateratzen dugun bigarren ondorioa zera da: laginketa-banaketaren errore tipikoa populazioaren desbideratze tipikoa (σ) baino txikiagoa da beti.

Dena dela, laginen tamainak handituko bagenitu (n = 100 izan beharrean, n = 160, adibidez), batez bestekoen laginketa-banaketaren desbideratze tipikoa; hau da, errore tipikoa txikitu egingo litzateke. Izan ere, populazioaren zati handiagoa hartzean, lagin desberdinen batez bestekoak gutxiago desberdintzen dira. Alegia, zenbat eta lagin handiagoa hartu, laginketa-banaketaren desbideratze tipikoa (=errore tipikoa) txikiagoa da, eta lortzen dugun batez bestekoa orduan eta gehiago hurbiltzen da populazioaren batez bestekora. Horrenbestez, ateratzen dugun hirugarren ondorioa zera da: errore tipikoa laginaren tamainaren (n-ren) arabera aldatzen dela (kopuru handien legea). Ondoren aurkezten dugun irudian ikus daitekeen bezala, laginaren tamaina handitzean (n =20, n=30, n=50…), laginketa-banaketaren kurba puntazorrotzagoa bihurtzen da, eta horrek esan nahi du errore tipikoa txikiagoa dela. Gogoratu kurtosia aztertzerakoan esan genuena: histogramak zenbat eta zorroztasun handiagoa izan, datu multzoan orduan eta sakabanatze txikiagoa dagoela.

56. diagrama: Lagin-tamaina desberdinekin lortutako laginketa-banaketa desberdinen kurtosi maila edo forma desberdinak



Ondorio horiek guztiak kontuan hartzen baditugu, estatistika inferentzialaren oinarri den teorema estatistikoa lortzen dugu: Limitearen Teorema Zentrala. Teorema horrek zera esaten du:
μ batez bestekoa eta σ desbideratze tipikoa duen populazio infinitu batetik, n tamainako lagin infinitu ateratzen baditugu (n>30, edo edozein tamainakoak populazioa normal erakoa bada), lagin horietatik lortutako batez bestekoen laginketa-banaketa gutxi gorabehera normal erakoa da, batez bestekoen batez bestekoa populazioaren batez bestekoaren berdina da, eta batez bestekoen desbideratze tipikoa berdin errore tipikoa =

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Non,

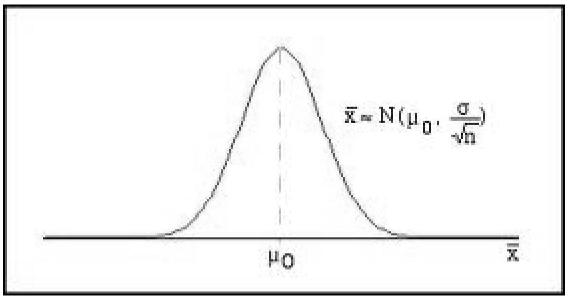
σ : populazioaren desbideratze tipikoa baita,
n: laginaren tamaina.

Hala, lagin infinituen batez bestekoen maiztasun-banaketak, hau da, batez bestekoen laginketa-banaketak, honako berezitasun hauek ditu:

$$N(\mu, \frac{\sigma}{\sqrt{n}})$$

Eta era honetako kurba normala:

57. diagrama: laginketa-banaketaren kurba normala



Esanak esan ostean, Limitearen Teorema Zentralaren honako ondorio hauek laburbil ditzakegu:

- Teorema horren erabilgarritasuna da estatistiko baten laginketa-banaketa bere parametroarekin erlazionatzen duela. Arestian esan bezala, frogatuta dago populazioaren maiztasun-banaketaren eta laginketa-banaketen artean zenbait antzerakotasun daudela; horren ondorioz, laginketa-banaketen bidez, populazioaren parametroak ezagut ditzakegu. Hala, laginaren berezitasunak aztertuz, laginketa-banaketaren berezitasunak ezagut ditzakegu, eta horren bidez populazioaren berezitasunak.
- Teorema hori populazioak normalak ez direnean ere betetzen da (errentaren kasuan, adibidez); are gehiago, batez bestekoaren kasuan ez ezik, beste estatistikoen (mediana, proportzioak, kurtosia, moda…) laginketa-banaketen kasuan ere betetzen da.
- Populazioa normala ez izanik ere, teorema bete egiten dela esan dugu. Hala eta guztiz ere, ezin da ahantzi teorema betetzen den ala ez jakiteko, laginaren tamainari begiratu behar zaiola: lagin txikiak ditugunean (n<30) eta populazioa normala dela guztiz ziur ez

bagaude, laginketa-banaketak ez dio kurba normalari jarraitzen, Student-en t maiztasun-banaketa baizik[11].

- Edozein estatistiko bere parametrora hurbiltzen den heinean, errore tipikoak estatistikoa (laginetik lortutakoa) zenbateko zehaztasunarekin den parametroaren estimatzailea neurtzeko balio du. Hau da, errore tipikoaren bidez, estatistiko bakar batekin (normalean egiten dena) parametroa ezagut dezakegu. Beraz, laginetik abiatuta, errore tipikoa kalkulatzen dugu, eta, hortik, parametroa ezagutzen. Esan bezala, laginketa-banaketaren eta errore tipikoaren bidez harremanetan jartzen ditugu lagina eta populazioa.
- Edozein estatistikoren laginketa-banaketa normal erakoa izanik, era horretako banaketak aztertzeko ditugun abantailak erabil ditzakegu. Zehatzago esanda, “68-95-99,7 araua” edo “arau enpirikoa” erabil dezakegu:
  - Laginketa-banaketako batez bestekoen %68 [μ – 1 σ, μ + 1σ] tartean dago.
  - Laginketa-banaketako batez bestekoen %95 [μ – 2 σ, μ + 2σ] tartean dago.
  - Laginketa-banaketako batez bestekoen %99 [μ – 3σ, μ + 3σ] tartean dago.

Beraz, lagin batean estatistiko bat aztertu (batez bestekoa, proportzioa, mediana…), eta bere errore tipikoa ezagutu ostean, probabilitate batekin parametroa zein tartetan egon daitekeen jakin dezakegu. Horretarako, ezinbestekoa da tarte-estimazioa nola egiten den jakitea.

3.5.3. Batez bestekoaren errore tipikoa eta populazio finituetarako faktore biderkatzailea

Esan bezala, errore tipikoak estatistikoa zenbateko zehaztasunarekin den parametroaren estimatzailea neurtzeko balio du. Horrela, errore tipikoaren bidez, estatistiko bakar batekin parametroa zein tartetan egon daitekeen ezagut daiteke. Beste era batera esanda, errore tipikoak harremanetan jartzen ditu lagina eta populazioa; horrenbestez, oso garrantzitsua da errore tipikoaren berezitasunak eta kalkulatzeko moduak ondo ezagutzea.

Adibidea. Errore tipikoa kalkulatzea parametroak ezagunak direnean: N = 1700, n = 100 eta σ = 6 izanik, populazio horren batez bestekoaren errore tipikoa kalkulatzeko, honako formula hau erabili behar dugu:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Formula aplikatuta, honako emaitza hau ateratzen dugu:

$$\sigma_{\bar{x}} = \frac{6}{\sqrt{100}} = 0{,}6$$

Beraz, datu multzo horri dagokion batez bestekoen laginketa-banaketak honako berezitasun hauek izango lituzke: N (24,5 ; 0,6). Horrek zera esan nahi du: 1.700 kide dituen elkarte horretatik har ditzakegun n = 100 tamainako lagin infinituetatik lagin bat hartuko bagenu, ateratzen dugun batez bestekoa parametroaren goitik eta behetik, gehienez jota, 0,6 urtetan desberdinduko litzatekeela.

Normalean, baina, parametroak ez dira ezagutzen, eta, horren ondorioz, populazioaren desbideratze tipikoa ez da ezaguna izaten. Hain zuzen ere, gizarte ikerlarion lana da lagin batetik abiatuz populazioaren berezitasunak ezagutzea, eta, esan bezala, horretarako, errore tipikoa

[11] Student-en t banaketa normalera hurbiltzen den datu multzo txikiak (30 baino gutxiago) aztertzeko erabiltzen den probabilitate-banaketa da.

ezagutu beharra dago (errore tipikoak, laginetik eta populaziorako bidea egitea ahalbidetzen digu). Hala, laginekin lanean ari garenean, errore tipikoa ezagutzeko, laginaren desbideratze tipikoa erabili behar izaten dugu:

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Adibidea. Errore tipikoa kalkulatzea parametroak ezezagunak direnean: 3.000 ikasle dituen unibertsitate fakultate batean, n = 345eko lagina hartu dugu, eta estatistika ikasgaian ateratako batez besteko nota 6,5 dela ikusi dugu, 2-ko desbideratze tipikoarekin. Errore tipikoa honela kalkulatzen dugu:

$$\sigma_{\bar{x}} = \frac{2}{\sqrt{345}} = 0{,}1$$

Horrek zera esan nahi du: har ditzakegun n = 345eko tamainako lagin infinituetatik bat hartuko bagenu, ateratzen dugun batez bestekoa parametroaren goitik eta behetik 0,1an desberdinduko litzatekeela.

Errore tipikoaren formulari begiratuta, argi dago errore tipikoa txikitu egiten dela lagina zenbat eta gehiago handitu. 345eko tamainako lagina hartu beharrean, 600eko tamaina duen lagina hartuko bagenu, errore tipikoa txikitu egiten dela ikusiko genuke:

$$\sigma_{\bar{x}} = \frac{2}{\sqrt{600}} = 0{,}08$$

Zalantza gabe, errore tipikoa txikiagoa da zenbat eta lagin handiagoa hartu. Dena dela, kontuan izan behar da ez dela hobekuntza konstante bat. n-ren erro karratuarekin erlazionatzean, laginaren tamaina handitzearen errendimendua gutxitu egiten da. Hau da, n-ren tamaina batetik aurrera errorea ez da gutxitzen.

Populazioa txikia bada (autore batzuek populazio finituei buruz hitz egiten dute) eta lagina ordezkapenik gabe osatzen bada[12], errore tipikoa populazio finituetarako faktore biderkatzailearekin biderkatu behar da; alegia, honako formula honetatik lortzen dugun emaitzarekin:

$$\sqrt{\frac{N-n}{N-1}}$$

Hala, populazio txikiak ditugunean, batez bestekoen laginketa-banaketaren errore tipikoa ($\sigma_{\bar{x}}$) honela kalkulatzen da:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

[12] Ordezkapena onartzen den laginetan, elementu, unitate edo pertsona bat nahi beste aldiz aukeratu daiteke; izan ere, aukeratuta eta gero, berriro botatzen da populaziora.

Non n laginaren tamaina baita, eta N, populazioarena.
Autore batzuek diote faktore zuzentzaile hori erabili behar dela lagina ordezkapenik gabe osatzen denean, eta populazioa tamaina lagina baino 20 aldiz txikiagoa denean.

Adibidea. Errore tipikoa eta populazio finituetarako faktore biderkatzailea: N = 1.700, n = 100 eta σ = 6 izanik, errore tipikoa 0,6 dela esan dugu. Populazio finituetarako biderkatzailea aplikatzen badugu, errore tipikoa 0,58koa dela ikusten dugu.

$$\sigma_{\bar{x}} = 0{,}6\sqrt{\frac{1700-100}{1700-1}} = 0{,}58$$

Adibidea. Errore tipikoa eta populazio finituetarako faktore biderkatzailea.

Ondoren aurkezten dugun maiztasun-taulan, etxe bateko hiru anaiak astero hartzen dituzten ordainsariak agertzen dira:

81. taula: etxeko umeen asteroko ordainsaria	
Umearen izena	Asteroko ordainsaria eurotan
Unai	6
Antxon	4
Aitor	2
Iturria: lanketa propioa	

Kalkulatu:

- Populazioaren batez bestekoa.
- Bi umeko lagin posible guztiak hartuta, laginketa-banaketaren batez bestekoa.
- Populazioaren desbideratze tipikoa.
- Laginketa-banaketaren desbideratze tipikoa.
- Errore tipikoa, populazio finituetarako faktore biderkatzailea kontuan hartu gabe.
- Errore tipikoa, populazio finituetarako faktore biderkatzailea kontuan hartuta.

Laginak	Ordainsariak	batez bestekoak
Unai-Antxon	6 eta 4	5
Unai-Aitor	6 eta 2	4
Antxon-Aitor	4 eta 2	3

- Populazioaren batez bestekoa: 4
- Laginketa-banaketaren batez bestekoa: 4
- Populazioaren desbideratze tipikoa: 1,63
- Laginketa-banaketaren desbideratze tipikoa (=errore tipikoa): 0,816
- Errore tipikoa, populazio finituetarako, faktore biderkatzailea kontuan hartu gabe: 1,152
- Errore tipikoa, populazio finituetarako, faktore biderkatzailea kontuan hartuta: 0,816

Ikusten den bezala, laginketa-banaketaren desbideratze tipikoa ez dator bat laginketa-banaketarako kalkulatutako errore tipikoarekin. Populazio finituetarako faktore zuzentzailea aplikatu behar dugu, datu multzoaren batez bestekoaren benetako errore tipikoa ezagutzeko:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$\sigma_{\bar{x}}$ = 0,816

Adibide horrek, halaber, erakusten digu Limitearen Teorema Zentrala bete egiten dela. Izan ere, 4 euroko batez bestekoa eta 1,63 euroko desbideratze tipikoa duen populazio horretatik [N (4, 1,63)], 2 tamainako lagin posible guztiak ateratzen baditugu, lagin horien batez bestekoen batez bestekoa populazioaren batez bestekoaren berdina da, eta batez bestekoen laginketa-banaketaren desbideratze tipikoa edo errore tipikoa (0,816) populazioaren benetako desbideratze tipikoa (1,63) baino txikiagoa da.
Errore tipikoaren faktore biderkatzailearekin bukatzeko, kontuan izan lagina zenbat eta handiagoa izan, faktore biderkatzailea orduan eta gehiago hurbiltzen dela 1era. Alegia, zenbat eta lagin handiagoa hartzen dugun, laginketa-frakzioa $\left(\frac{n}{N}\right)$ txikiagoa izango da. Beraz, lagin handiekin lan egiten dugunean ez dugu zertan faktore biderkatzailea erabili behar. Aurretik ere esan dugu, baina kontuan izan behar dugu n-ren edo laginaren tamainak zehazten duela errore tipikoa. Zenbat eta lagin handiagoa izan, orduan eta errore tipiko txikiagoa izango dugu.

Adibidea. Errore tipikoa IBM SPSS programaren bidez. Ospitaleko jaioberrien pisua.
Demagun honako datu multzo hau lortu dugula ospitale bateko jaioberrien pisuekin:

82. taula: jaioberrien pisua

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	2,10	14	2,4	2,4	2,4
	2,20	15	2,6	2,6	5,0
	2,30	28	4,9	4,9	9,9
	2,40	27	4,7	4,7	14,6
	2,50	24	4,2	4,2	18,7
	2,60	23	4,0	4,0	22,7
	2,70	63	10,9	10,9	33,6
	2,80	169	29,3	29,3	62,9
	2,90	65	11,3	11,3	74,2
	3,00	38	6,6	6,6	80,8
	3,10	24	4,2	4,2	84,9
	3,20	39	6,8	6,8	91,7
	3,30	35	6,1	6,1	97,7
	3,40	13	2,3	2,3	100,0
	Total	577	100,0	100,0	

Iturria: lanketa propioa

Hortik abiatuta, IBM SPSS programak oso erraz kalkulatzen ditu neurri estatistikoak zein batez bestekoaren, asimetriaren eta kurtosiaren errore tipikoak:

Estadísticos
Jaioberrien pisua

N	Válidos	577
	Perdidos	0
Media		2,8009
Error típ. de la media		,01242
Mediana		2,8000
Moda		2,80
Desv. típ.		,29840
Varianza		,089
Asimetría		-,236
Error típ. de asimetría		,102

Curtosis		-,111
Error típ. de curtosis		,203
Rango		1,30
Mínimo		2,10
Máximo		3,40
Suma		1616,10
Percentiles	10	2,3800
	20	2,6000
	25	2,7000
	30	2,7000
	40	2,8000
	50	2,8000
	60	2,8000
	70	2,9000
	75	3,0000
	80	3,0000
	90	3,2000

3.5.4. Proportzioen errore tipikoa

Hainbat aldiz esan bezala, batez bestekoak bezala, desbideratze tipikoak, proportzioek, modak, medianak eta beste estatistikoek ere beren laginketa-banaketa eta errore tipikoa dute. Beraz, μ batez bestekoa eta σ desbideratze tipikoa dituen populazio infinitu batetik, n tamainako lagin infinitu ateratzen baditugu (n>30, edo edozein tamainakoak populazioa normal erakoa bada), lagin horien proportzioen laginketa-banaketa gutxi gorabehera normal erakoa da; proportzioen batez bestekoa, populazioaren proportzioaren berdina da, eta laginen proportzioen desbideratze tipikoa = proportzioen errore tipikoa =

$$\sigma_p = \sqrt{\frac{p.q}{n}}$$

Non n laginaren tamaina baita, eta p eta q zerbait gertatzeko eta ez gertatzeko dauden probabilitateak direlarik. Alegia, datuak dikotomikoak direnean (1 edo 0), p bai esaten dutenen edo 1 erantzun dutenen proportzioa da, eta q, horren osagarria (q = p -1).

Adibidea. 200 pertsonako lagina badugu, eta 120k galdera bati baietz erantzuten badiote eta 80k, ezetz, $p = \frac{120}{200} = 0{,}6$ da; q = 0,4.

Adibidea. Populazioaren proportzioa ezaguna izanik, errore tipikoa kalkulatzea: demagun herrialde bateko emakume kirolariak ditugula (N = 80.000), eta, aurretik egindako ikerketen arabera badakigula, horien artean lasterka egiten dutenen proportzioa (p) 0,55 dela. Ikusten den bezala, parametroa ezagutzen dugu.
500 emakumeko lagin infinitu hartu ostean, horietariko bakoitzean lasterkarien proportzioak ateratzen ditugu: $p_1 = 0{,}5$; $p_2 = 0{,}47$; $p_3 = 0{,}62$;…;$p_n = 0{,}49$.
Laginetatik ateratako proportzioek proportzioen laginketa-banaketa osatuko lukete, non lagin guztien proportzioen batez bestekoa = μ_p = populazioaren benetako proportziora hurbilduko

litzatekeen (= 0,55=parametroa), eta proportzio guztien batez bestekoaren desbideratze tipikoa = Errore Tipikoa =

$$\sigma_p = \sqrt{\frac{pq}{n}}$$

Beraz, kasu horretan, errore tipikoa kalkulatzen dugu, eta σ_p = 0,022 (%2,2) dela ohartzen gara. Errore tipiko hori populazio finituetarako faktorearekin biderkatzen badugu $\sqrt{\frac{N-n}{N-1}}$, errore tipikoa = 0,0219 dela ikusten dugu:

$$\sigma_p = \sqrt{\frac{0{,}55.0{,}45}{500}}.\sqrt{\frac{80000-500}{80000-1}} = 0{,}0219$$

N oso handia eta laginketa-frakzioa $\left(\frac{n}{N}\right)$ oso txikia direnez, errore tipikoa ia ez da aldatzen (σ_p = 0,022).

Proportzioen errore tipikoarekin bukatzeko, zera izan behar da kontuan: proportzioen errore tipikoa p-ren eta q-ren arteko erlazioaren eta laginaren tamainaren (n-ren) arabera aldatzen dela. Lagina handitzen den heinean, errore tipikoa gutxitu egiten da. Modu berean, p eta q zenbat eta gehiago aldendu elkarrengandik (lagina homogeneoagoa da), errore tipikoa orduan eta txikiagoa da. Aldiz, p eta q zenbat eta berdintsuagoak izan (lagina heterogeneoagoa da), errore tipikoa orduan eta handiagoa da. Zentzu horretan, kontuan izan p-k eta q-k populazioa homogeneoa edo heterogeneoa den adierazten dutela. p = 0,2 bada eta q = 0,80, lagin homogeneoa da, populazioa osatzen duten gehienek q-ren alde egiten duelako. p = 0,5 eta q = 0,5 bada, berriz, hau da, p eta q berdinak badira, lagina heterogeneoa dela onartzen dugu. Eta lagin heterogeneoak baditugu, errore tipikoa handiagoa izango da.

Adibidea. Errore tipikoa eta p-ren eta q-ren arteko erlazioa. Demagun p = 0,3 eta q = 0,70 direla. Lagin homogeneoa da, gehiengoak q-ren alde egiten dutelako. 700 pertsonako lagina hartu badugu, errore tipikoa honako hau da:

$\sigma_p = \sqrt{\frac{pq}{n}}$	$\sigma_p = \sqrt{\frac{0{,}3.0{,}7}{700}} = 0{,}0173$

P eta q berdintsuagoak izango balira, adibidez 0,45 eta 0,55 errore tipikoa handiagoa izango litzateke, lagina heterogeneoagoa delako.

$\sigma_p = \sqrt{\frac{pq}{n}}$	$\sigma_p = \sqrt{\frac{0{,}45.0{,}55}{700}} = 0{,}0188$

3.6. ESTIMAZIOAREN TEORIA

Estatistiko bakarra aztertuz parametroa estimatzea da estatistika inferentzialaren helburua. Horretarako, lagina aukeratu ostean, estatistikoa ezagutu behar dugu, estatistiko horri dagokion errore tipikoa kalkulatu eta lan egin nahi dugun konfiantza-maila zehaztu.

Ezer azaltzen hasi aurretik, estatistiko bat noiz den parametroaren estimatzaile ona zehaztu behar dugu.

3.6.1. Estimatzaile on baten berezitasunak

Hasteko, esan behar dugu edozein estatistiko bere parametroaren estimatzaile ona dela. Izan ere, lagin infinitu ateratzen baditugu, eta bakoitzean estatistikoa kalkulatzen, estatistikoak parametrora hurbiltzeko joera duela ikusten dugu. Dena dela, estimatzaile on batek berezitasun batzuk bete behar ditu:

- Soslairik ez: estimatzailearen eta parametroaren artean dagoen desberdintasuna da estimatzaile baten soslaia. Komenigarria da erabiltzen ditugun estimatzaileak soslairik gabekoak izatea; hau da, estimatzailearen eta parametroaren artean desberdintasunik ez egotea. Adibidez, populazio baten batez bestekoa ezagutu nahi bada, laginaren batez bestekoa populazioaren parametroaren soslairik gabeko estimatzailea da; hain zuzen ere, laginaren batez bestekoa populazioaren batez bestekoaren berdina delako. Populazio normaletan, batez bestekoa eta mediana, populazio baten benetako batez bestekoaren soslairik gabeko estimatzaileak dira (populazioa normala ez bada, medianak badu soslaia). Laginaren desbideratze tipikoa zein bariantza populazioaren desbideratze tipikoaren estimatzaile sesgatu samarrak dira.
- Efizientzia erlatiboa edo zehaztasuna: estimatzaile bat beste bat baino efizienteagoa da, bere bariantza bestearena baino txikiagoa bada; hau da, bere errore tipikoa txikiagoa denean. Horregatik, batez bestekoa mediana baino estimatzaile efizienteagoa da.
- Konsistentzia: bariantza gutxiko estimatzaileak erabiltzea ezinezkoa bada, lagina handitzean, estimatzailea parametrora hurbiltzea da estimatzaile bati eskatzen diogun gutxieneko baldintza. Hori da, hain zuzen ere, konsistentzia. Nahiz eta sesgatua izan, estimatzaile bat konsistentea da n handitzean bere parametrora jotzeko joera badu. Argi dago soslairik gabeko estimatzaile guztiak konsistenteak direla, baina, adibidez, desbideratze tipikoa (s), sesgatua izanik, konsistentea da.

3.6.2. Tarte-estimazioa

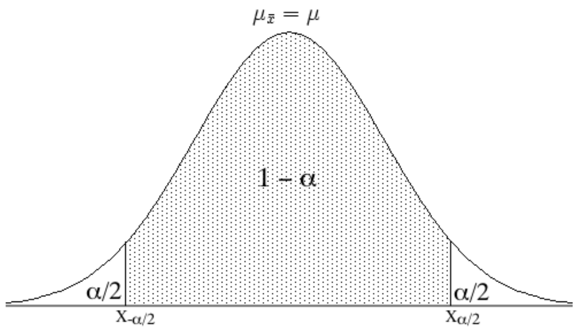
Arestian esan bezala, parametroa estimatzeko bi bide ditugu: estimazio puntuala eta tarte bitartez egindako estimazioa.
Estimazio puntualek estatistikoa bere parametroaren estimatzaile ona dela onartzen dute. Horren arabera, populazio batetik atera dezakegun lagin bateko batez besteko adina ($\bar{x}$ = 22) populazioaren benetako batez bestekoaren (μ = 22) estimatzaile puntuala da. Badakigu batez bestekoa estimatzaile ez-sesgatua, konsistentea eta efizientea dela.
Dena dela, estimazio puntualak ez dira nahikoak, ez dutelako kontuan hartzen estimatzailearen zehaztasuna edo efizientzia adierazten digun errore tipikoa. Horren ondorioz, estimazio puntuala ahantzi gabe, parametroen estimazioa tarte bitartez egiten da ia gehienetan. Hau da, errore tipikoa kontuan hartuta eta probabilitate zehatz batekin, parametroa zein tartetan egon daitekeen, haiei dagozkien balioak adierazten dira.
Tarte-estimazioa egitean, garrantzitsua da honako kontzeptu hauek argi izatea:

- Adierazgarritasun-maila (α) parametro bat zehaztutako tarteetatik kanpo egoteko dagoen probabilitatea da; hau da, egindako estimazioa okerra izateko probabilitatea. Ia gehienetan, 0,05 eta 0,01 adierazgarritasun-mailak erabiltzen dira. Adierazgarritasun-maila banaketa normalaren albo bien artean banatu behar dugu; beraz, banaketa normalaren albo bakoitzean, $_{\alpha/2}$ tamainako azalerak utzi behar ditugu.
- Konfiantza-maila (1 – α) parametro bat zehaztutako tarteen barnean egoteko dagoen probabilitatea da. α adierazitako tartetik kanpo egoteko dagoen probabilitatea bada, 1 –

α konfiantza-maila da. Ia gehienetan, %95eko eta %99ko konfiantza-mailak erabiltzen dira, hain zuzen ere, 0,05 eta 0,01 adierazgarritasun-mailei dagozkienak.

- Balio kritikoa: $z_{\alpha/2}$ ikurraren bidez adierazten da, eta, haren bidez, eremu kritikoa zehazten da. Kurba normalaren azpiko azalerak adierazten dizkigun taularen bidez kalkulatzen da. Hala, konfiantza-maila bakoitzari dagokion $z_{\alpha/2}$ balioa aurkitzen dugu, eta banaketa normalaren simetria erabiliz, balio kritikoa $-z_{\alpha/2}$etik behera eta $+ z_{\alpha/2}$ gora dagoela esango dugu. Beste era batera esanda, bere eskuinera $_{\alpha/2}$ tamainako azalera uzten duen z balioa da balio kritikoa.

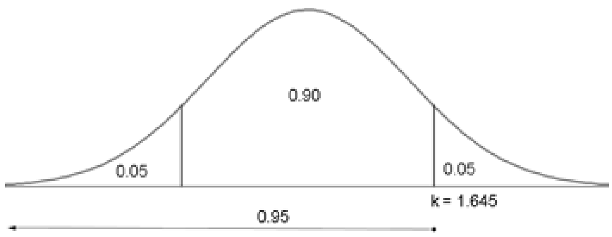
58. diagrama: adierazgarritasun-maila, konfiantza-maila eta balio kritikoa kurba normal estandarraren barruan



Adibidea. Eremu kritikoa kalkulatzea: %90eko konfiantza-mailari dagokion eremu kritikoa kalkulatzeko, banaketa normalaren banaketa funtziorako taulan begiratu behar da, baina alderantzizko bidea eginez. Alegia, bilatzen ari garen probabilitateari gehien hurbiltzen zaion z balioa aurkitu behar dugu. z balio hori izango da balio kritikoa, eta, haren bidez, eremu kritikoa zehaztuko dugu:

$P[-z_{\alpha/2} < z < z_{\alpha/2}]$

59. diagrama: adierazgarritasun-maila, konfiantza-maila eta balio kritikoa kurba normal estandarraren barruan



%90eko konfiantza-mailari 0,1-eko adierazgarritasun-maila (α) dagokio. Hau da, 0,1 da parametro bat zehaztutako tarteetatik kanpo egoteko dagoen probabilitatea; alegia: egindako estimazioa okerra izateko probabilitatea.
Batez bestekoaren inguruan dagoen %90 zer tartetan dagoen jakiteko, banaketaren albo bien artean banatu behar dugu adierazgarritasun-maila. Hala, banaketa normalaren albo bakoitzean, $_{\alpha/2}$ = 0,05 tamainako azalerak utzi behar ditugu.
Kurba normalaren azpiko azalera-taula erabiliz, 0,05eko adierazgarritasun-mailari dagokion $z_{\alpha/2}$ balio bi daude: 1,64 (h azpitik datuen %94,95 uzten du) eta 1,65 (bere azpitik datuen %95,05 uzte du). Kasu horretan, %90eko konfiantza-mailari dagokion balio kritikoa 1,645 dela onartuko

dugu, eta, banaketa normalaren simetria erabiliz, balio kritikoa -1,645etik behera eta +1,645etik gora dagoela esango dugu.

Adibidea. Eremu kritikoa kalkulatzea: %95eko konfiantza-mailari 0,05eko adierazgarritasun-maila (α) dagokio. Eremu kritiko horri dagokion azalera bitan banatzen badugu, banaketaren alde bakoitzean, 0,025eko azalera utziko dugu. Kurba normalaren azpiko azalerak adierazten dizkigun taula erabiliz, 0,025eko adierazgarritasun-mailari dagokion $z_{\alpha/2}$ balioa 1,96 da. Beraz, %95eko konfiantza-mailari dagokion balio kritikoa 1,96 dela onartuko dugu, eta, banaketa normalaren simetria erabiliz, balio kritikoa -1,96etik behera +1,96etik gora dagoela esango dugu.

Dena dela, tarteak kalkulatu behar ditugun bakoitzean ez dugu zertan aritu balio kritikoa bilatzen. Egiatan, honako hauek dira gehien erabiltzen diren konfiantza-maila, adierazgarritasun-maila eta balio kritikoak:

83. taula: konfiantza-maila, adierazgarritasun-maila eta balio kritiko erabiliena		
$1 - \alpha$	α	$z_{\alpha/2}$
0.90	0.1	1.645
0.95	0.05	1.96
0.99	0.01	2.575
Iturria: lanketa propioa		

Ikusten den bezala, adierazgarritasun-maila bakoitzari konfiantza-maila bat eta balio kritiko bat dagokio.

Demagun n tamainako lagin bakar bat dugula, eta lagin horren batez bestekotik ($\bar{x}$) abiatuz, populazioaren benetako batez bestekoa (μ) estimatu nahi dugula.

Badakigu $\bar{x}$ (laginaren batez bestekoa) μ-ren (populazioaren batez beztekoa) estimatzaile ez sesgatua, konsistentea eta efizientea dela, batez bestekoen laginketa-banaketa normala dela, eta laginketa-banaketa horrek μ batez bestekoa eta σ errore tipikoa dituela:

$$\sigma_p = \sqrt{\frac{pq}{n}}$$

Batez bestekoen laginketa-banaketak banaketa normalera hurbiltzen dela jakinda, badakigu, halaber,

- batez bestekoen %68 [μ – 1σ, μ + 1σ] tartean dagoela.
- batez bestekoen %90 [μ – 1,64σ, μ + 1,64σ] tartean dagoela.
- batez bestekoen %95 [μ – 1,96σ, μ + 1,96σ] tartean dagoela.
- batez bestekoen %99 [μ – 2,57σ, μ + 2,57σ] tartean dagoela[13].

Alegia, banaketa normalak ditugunean, lagineko batez bestekoa kalkulatuta, errore tipikoa ezagututa eta konfiantza-maila zehaztuta, populazioaren batez bestekoa (μ = parametroa) dagoen tartea modu honetan kalkula dezakegu:

$\mu = \bar{x} \pm z \cdot \sigma$

Non, $\bar{x}$ = laginetik lortu dugun batez bestekoa baita (estatistikoa), σ = errore tipikoa eta z, lan egin nahi dugun konfiantza-mailari dagokion z balioa (balio kritikoa).

84. taula: konfiantza-maila, adierazgarritasun-maila balio kritiko eta tarte erabilienak			
$1 - \alpha$	α	$z_{\alpha/2}$	Tarteak
0.90	0.1	1.645	(μ – 1.645 · σ , μ + 1.645 · σ)
0.95	0.05	1.96	(μ – 1.96 · σ , μ + 1.96 · σ)

[13] Gogoratu horiek direla kurba normalaren batez bestekoaren inguruan dauden azalerak.

0.99	0.01	2.575	(μ – 2.575 · σ , μ + 2.575 · σ)
Iturria: lanketa propioa			

Batez bestekoarekin egin dugun hori edozein estatistikorekin egin dezakegu. Proportzioen kasuan, tartea bilatzeko, honako formula hau aplikatuko dugu:

$$P = p \pm Z\sigma_p$$

Non, p = laginetik lortu dugun proportzioa baita (estatistikoa); σ = errore tipikoa, eta z, lan egin nahi dugun konfiantza-mailari dagokion z balioa.

Hala, edozein estatistikotik abiatuta, parametroa zein tartetan egon daitekeen jakin dezakegu formula hau erabiliz:

$$Parametroa = estatistikoa \pm z.errore \cdots tipikoa$$

Ez ahantzi era horretako estimazioak egiteko beharrezkoa dela estatistikoaren laginketa-banaketaren normaltasuna ziurtatzea. Normaltasunik ez badago, z balioak dagokion banaketa teorikoaren balioekin ordezkatu behar ditugu.

Adibidea. Parametroaren estimazioa tarteen bidez, estatistikotik abiatuta: demagun 1.700 kide dituen gazteen elkartea dugula, eta batez besteko adina (μ) 24,5 dela, eta desbideratze tipikoa (σ) 6. Alegia, parametroak ezagutzen ditugu.

Demagun n = 100 tamainako lagina atera dugula, eta azterketa egin ostean, batez besteko adina ($\bar{x}$) 23,8 urtekoa dela. Arestian ikusi bezala, badakigu errore tipikoa 0,58 dela.

$$\sigma_{\bar{x}} = 0{,}6\sqrt{\frac{1700-100}{1700-1}} = 0{,}58$$

Zein tartetan egongo da elkarte horretako bazkideen adina, %99ko konfiantza-mailarekin? Tartea kalkulatzeko, honako formula hau erabili behar dugu:

$$\mu = \bar{x} \pm z.\sigma_{\bar{x}}$$

μ = 23,8 $\pm$ 2,58 (0,58) = 25,29 eta 22,3 urteen artean egongo da populazioaren batez besteko adina, %99ko konfiantza-mailarekin.
Demagun beste lagin bat lortu dugula, non, $\bar{x}$ = 22 baita. Zein tartetan egongo da elkarte horretako kideen adina, %99ko konfiantza-mailarekin?
μ = 22 $\pm$ 2,58 (0,58) = 23,49 eta 20,5 bitartean egongo da populazioaren batez besteko adina, %99ko konfiantza-mailarekin.
Errealitatean, μ (parametroa) = 24,5 urte dela kontuan hartuz, lortu dugun lehen tarteak kontuan hartzen du populazioaren benetako batez bestekoa (parametroa), baina bigarrenak, ez. Eta hori horrela da bigarren kasuan lortu dugun batez bestekoa oso urrun dagoelako parametrotik. Horrek zera esan nahi du: $\bar{x}$ = 22 laginketa-banaketaren batez besteko marjinal, arraro edo ezohikoa dela.
Adibidea. Parametroaren estimazioa tarteen bidez, estatistikotik abiatuta: emakume kirolari lasterkariak.

Badakigu hiri handi bateko emakume kirolarien artean (N = 80.000) %55ek korrika egiten duela (ikusten den bezala, parametroa ezaguna da). Populazio horretatik, n = 500-eko lagina hartu eta aztertu ostean, ikusi dugu lasterkarien proportzioa = 0,59 dela. 0,59ko proportzioa estimatzaile puntuala litzateke. %99ko, konfiantza-mailarekin, zer tartetan egongo da hiri horretako emakume kirolarien artean lasterka egiten dutenen ehunekoa?
Badakigu errore tipikoa 0,0219 dela (populazio finitua denez, zuzendua dago).

$$\sigma_p = \frac{0{,}55.0{,}45}{\sqrt{500}} . \sqrt{\frac{80000-500}{80000-1}} = 0{,}0219$$

Hala, parametroa zer tartetan egongo den modu honetan estimatzen dugu:

P_p = 0,59 $\pm$ 2,58 (0,0219) = 0,6465 (%64,65) eta 0,5334 (%58,34) tarteen artean egongo da hiriko emakume kirolarien artean korrika egiten dutenen ehunekoa, %99ko konfiantza-mailarekin.
Gure lagina n = 600ekoa izan balitz, eta p = 0,64-koa, tartea desberdina litzateke: Pp = 0,64 $\pm$ 2,58 (0,019). Kasu horretan, beraz, %68,9 eta %59,09 tarteen artean egongo litzateke emakume kirolarien artean lasterka egiten dutenen ehunekoa, %99ko konfiantza-mailarekin.
Parametroa (P) = 0,55-ekoa dela jakinik, ikusten dugu da lehen tarteak parametroa kontuan hartzen duela eta bigarrenak, ez. Beraz, bigarren laginetik atera dugun proportzioa p = 0,64 laginketa-banaketaren proportzio marjinal edo arraroa da, eta, beraz, estimazioak egiteko estatistiko txarra.

Adibidea. Parametroaren tarte-estimazioa estatistikotik abiatuta: Ezker Batuari bozka.
Estatu espainiarrak 46.815.916 biztanle ditu (hain handia izanik, populazio infinitutzat hartuko dugu). 2.494 pertsonako lagina hartu dugu, eta inkesta baten bidez jakin dugu %16,5ak hurrengo hauteskundeetan Ezker Batuari emango diola bozka.
Egin beharrekoak:

a) Estimazio puntuala erabiliz, kalkulatu Ezker Batuari botoa emango diotenen ehunekoa.
b) %95eko konfiantza-mailarekin, estimatu Ezker Batuari botoa emango diotenen proportzioa zein tartetan egongo den.
c) Kalkulu bera egin 0,02 adierazgarritasun-mailarako.
d) Lortutako bi tarteak alderatu.

a) Estimazio puntualaren arabera, hurrengo hauteskundeetan Estatu espainiarreko biztanleriaren %16,5ek Ezker Batuari emango dio bozka.
b) Ezker Batuari bozka emango diotela diotenen proportzioa p = 0,165 dela ezagututa, populazioaren barruan proportzio hori zein tartetan egongo den adierazi behar dugu. Horretarako, honako formula honetan oinarritu behar dugu:

$$P = p \pm z.\sigma_p$$

Formula erabiltzeko, z eta σ_p ezagutu behar ditugu.
z balioa kalkulatzeko, banaketa normalaren azpiko azalerei buruzko taulan, 0,95eko probabilitateari dagokion z balioa aurkitzen dugu = 1,96.
Errore tipikoa (σ_p) kalkulatzeko, honako formula hau erabiltzen dugu:

$$\sigma_p = \sqrt{\frac{pq}{n}}$$

q = 1- p dela jakinda, errore tipikoa honela kalkulatzen dugu:

$$\sigma_p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0{,}165)(0{,}835)}{2.494}} = 0{,}00743$$

Kasu honetan, populazioa handia izanik (populazio infinitutzat hartzen da), lortutako errore tipikoa ez dugu zuzendu behar. Beraz, estatistikoa (laginetik lortutako proportzioa), z eta σ_p ezagututa, erraz kalkulatzen ditugu parametroaren tarteak:

$$P = p \pm Z\sigma_p = 0{,}165 \pm (1{,}96)(0{,}00743) = 0{,}165 \pm 0{,}0146$$

Hala, %95eko konfiantza-mailarekin, Estatuan Ezker Batuari bozka emango diotenen proportzioa 0,150 eta 0,180 tartean egongo da. Edo beste era batera esanda, Estatuko biztanleen %15etik %18ra emango diote bozka Ezker Batuari.

c) 0,02ko adierazgarritasun-maila %98ko konfiantza-mailari dagokio. Beraz, tarteak ezagutzeko konfiantza-maila horri dagokion z balioa ezagutu behar dugu. Hain zuzen ere, z balioa 2,33 da. Beraz,

$$P = p \pm Z\sigma_p = 0{,}165 \pm (2{,}33)(0{,}00743) = 0{,}165 \pm 0{,}0173$$

%98ko konfiantza-mailarekin, Ezker Batuari botoa emango diotela diotenen proportzioa 0,148 eta 0,182 tartean egongo da.

d) Eman ditugun bi tarteak alderatzen baditugu, ondorioztatzen dugu konfiantza-maila handitzean tartearen zabalera handitu egiten dela. Ziurtasunean irabazten dena (zoriz lortutako lagin batean lortzen dugun proportzioa tartearen barnean kokatzeko probabilitatea) zehaztasunean galtzen da (tartearen zabalera handiagoa da).

Adibidea. Parametroaren estimazioa IBM SPSS programa erabiliz: ospitale batean jaio diren 577 ume pisatu ditugu, eta, jaiotzean, batez beste, 2,8 kilo izaten dituztela ikusi dugu. Errore tipikoa 0,01242 izanik, zer tartetan egongo da gizarte horretako jaioberrien pisua, %95eko konfiantza-mailarekin?

Estadísticos para una muestra

	N	Media	Desviación típ.	Error típ. de la media
jaioberrien pisua	577	2,8009	,29840	,01242

Prueba para una muestra

	Valor de prueba = 0					
	T	Gl	Sig. (bilateral)	Diferencia de medias	95% Intervalo de confianza para la diferencia	
					Inferior	Superior
jaioberrien pisua	225,464	576	,000	2,80087	2,7765	2,8253

Ikusten den bezala, jaioberrien %95ek 2,7765 kilotik 2,8253 kilora arteko pisua izango du.

Eta %99ko konfiantza-mailarekin, zein tartetan egongo dira jaioberrien pisuak?

Prueba para una muestra

	Valor de prueba = 0					
	T	Gl	Sig. (bilateral)	Diferencia de medias	99% Intervalo de confianza para la diferencia	
					Inferior	Superior
jaioberrien pisua	225,464	576	,000	2,80087	2,7688	2,8330

Jaioberrien %99k 2,7688 kilotik 2,8330 kilora arteko pisua izango du. Ikusten den bezala, konfiantza-maila handitzean, tartea zabaldu egiten da.

Parametroaren estimazioa IBM SPSS programaren bidez egiteko, Lagin Baterako T Froga erabiliko dugu. Froga honen bidez, ikusi dezakegu populazio baten batez bestekoa balio zehatz bat den edo ez. Hau da, lagin bat hartzen dugu, batez bestekoa eta desbideratze tipikoa kalkulatzen ditugu, eta testaren bidez jakin genezake populazioaren batez bestekoa balio zehatz bat dela adierazten duen hipotesi hutsa (μ_0) onartu daitekeen edo ez. Horretarako, formula hau erabiltzen dugu:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Non, $\bar{x}$ laginaren batez bestekoa den, μ_0 frogatu nahi den hipotesi hutsa, *s* laginaren desbideratze tipikoa eta *n* laginaren tamaina.
Hipotesi frogak egiteko prozedura ostean aztertuko dugu sakonago, baina kontuan izan T froga hipotesi froga bat dela. Beraz, formula aplikatuta lortzen dugun T zenbakia, eremu kritikoan kokatzen bada, hipotesi nulua errefusatuko da, α adierazgarritasun-mailarekin edo 1 – α konfiantza-mailarekin. Kasu horretan ondorioztatuko genuke populazioaren batez bestekoa ez dela hipotesi hutsak (μ_0) adierazitako balioa. Kontrako kasuan, hau da, T zenbakia (probarako estatistikoa) onarpen-eremuan badago, hipotesi nulua onartuko da, α adierazgarritasun-mailarekin edo 1 – α konfiantza-mailarekin. Kasu horretan ondorioztatuko genuke populazioaren batez bestekoa badela hipotesi hutsak (μ_0) adierazitako balioa. Onarpen eremua zein den adierazteko, IBM SPSS programak batez bestekoa egon daitekeen goiko eta beheko mugak zehazten ditu. Eta bi muga horiek dira, hain zuzen ere, 1 – α konfiantza-mailarekin, populazioaren batez bestekoa (parametroa) zer tartetan egongo den adierazten digutenak.

3.7. HIPOTESI FROGA EDO HIPOTESI KONTRASTEA

Aurreko atalean, estatistiko baten bidez parametroa estimatzen ikasi dugu. Estimatzaile puntuala erabiliz, zein tarte-estimazioaren bidez, parametroari buruz hitz egiteko gaitasuna lortu dugu.
Dena dela, parametroa ezagutzeko beste modu bat ere badago: hipotesi-frogak edo hipotesi-kontrasteak egitea. Hain zuzen ere, populazioaren parametro bati buruzko baieztapena egiten da, eta, ostean, baieztapen hori onartzeko edo alboratzeko erabaki estatistikoa hartzen. Zehatzago esanda, parametroei buruzko hipotesiak egiten dira, eta hipotesi horiek onartzeko edo alboratzeko, zoriz aukeratutako laginean neurri estatistiko bat ateratzen da, eta laginketaren eta estimazioaren teorietan oinarrituz, baieztapena onartu edo errefusatzeko erabakia hartzen. Adibidez, fakultate bateko ikasleen (populazioa) batez besteko adina 25 urte baino txikiagoa dela baztertuko da, laginaren batez bestekoa 26 baino handiagoa dela ikusten bada.
Tarte-estimazioaren arrazoibidearen alderantzizkoa egiten da hipotesi-frogetan. Tarte-estimazioan, estatistikotik abiatuz, parametroa zer tartetan egon daitekeen zehazten da. Hipotesi-frogetan, parametroak izan ditzakeen berezitasunak zehazten dira aldez aurretik, eta, ostean,

laginaren azterketatik lortu dugun estatistikoa berezitasun horietara hurbiltzen den aztertzen da, parametroari buruz egin dugun hipotesia onartzeko edo alboratzeko.
Hala, hipotesi-kontrastea laginetik abiatuz, populazioaren ezaugarri bati buruz egindako hipotesia onargarria edo alboragarria den erabakitzeko prozedura estatistikoa da. Ildo horretatik, kontuan izan behar dugu hipotesiak populazioari buruzko proposamen edo aurresuposizioak direla, ez laginari buruzkoak. Alegia, hipotesi-frogen helburua ez da laginketa bidez lortu dugun estatistikoa aztertzea, baizik eta parametrorako adierazitako balio baten (hipotesia) eta laginetik atera den neurri estatistikoaren artean dagoen desberdintasunari buruz epaia ematea.
Demagun jaioberrien neurria aztertzea interesatzen zaigula. Zehatzago esanda, jaioberrien batez besteko neurria 50 zentimetrokoa den ala ez da gure zalantza. Modu formalean, zalantza hori era honetan adierazi dezakegu:

$H_o \rightarrow \mu = 50$ zentimetro (jaioberrien batez besteko neurria 50 zentimetrokoa da)

$H_1 \rightarrow \mu \neq 50$ zentimetro (jaioberrien batez besteko neurria ez da 50 zentimetrokoa)

Lehen proposamenari ($H_o \rightarrow \mu = 50$ zentimetro) hipotesi nulua edo hutsa deitzen zaio. Bigarren proposamena, berriz, hipotesi alternatiboa da ($H1 \rightarrow \mu \neq 50$ zentimetro).
Hipotesi alternatiboak μ 50 zentimetro baino handiagoa edo txikiagoa izan daitekeela dioenez, hipotesi alternatibo aldebikoa da. Dena dela, beste batzuetan, hipotesi alternatibo aldebakarrekoak frogatu nahi izaten dira. Hain zuzen ere, $H_o \rightarrow \mu = 50$ zentimetro hipotesi nulurako bi eratako hipotesi alternatibo aldebakarrekoak egon daitezke:

$H_1 \rightarrow \mu < 50$ zentimetro

edo

$H_1 \rightarrow \mu > 50$ zentimetro

Hipotesi bat onartzera edo errefusatzera eramaten gaituen prozedurari hipotesi-froga deitzen zaio. Hipotesi-frogen prozedurak laginen azterketatik lortzen den informazioan oinarritzen dira. Informazio hori parametroari buruz baieztapena egiten duen hipotesi hutsarekin bat badator, hipotesia egiazkoa dela onartuko dugu; aldiz, laginetik ateratzen den informazioa eta hipotesi hutsa kontrakoak badira, hipotesi hutsa faltsua dela esango dugu, eta hipotesi alternatiboa onartuko.

3.7.1. Hipotesi-frogetako erroreak

Hipotesi hutsa H_o hizkiekin adierazten da, eta populazioaren berezitasun bat (edo gehiago) egia dela adierazten duen baieztapena da. Parametroari buruz “a priori” dugun ustea edo sinesmena da.
Hipotesi alternatiboa, H_1 hizkiekin adierazten da, eta H_o-ren aurkako baieztapena da.
Hipotesi hutsa baztertzen da eta hipotesi alternatiboa onartzen, laginaren azterketak erakusten badu H_o faltsua dela. Laginaren azterketak Ho baztertzen ez badu, hipotesi hutsa baliokoa dela onartuko dugu.
Hala, hipotesi-frogen ondorio posibleak dira: H_o alboratzea edo ez alboratzea.
Adibidea. Hipotesi-frogaren azalpena: jaioberrien batez besteko neurriari dagokionez, hipotesi hutsak esaten du umeek batez beste 50 zentimetro dituztela jaiotzean. Hipotesi alternatiboak, berriz, umeek ez dutela 50 zentimetro neurtzen. Hipotesi-froga beste modu honetan ere adieraz dezakegu:

$H_o \rightarrow \mu = 50$ zentimetro

$H_1 \rightarrow \mu \neq 50$ zentimetro

10 jaioberriz osatutako lagina hartu dugu, guztiak neurtu ditugu, eta batez besteko neurria lortu dugu. Batez besteko hori parametroaren estimatzaile puntuala da. Laginaren batez bestekoa 50 zentimetrora, hau da, balio hipotetikora hurbiltzen bada, pentsa dezakegu da batez bestekoaren

benetako balioa μ = 50 zentimetro dela; alegia, laginaren azterketatik ateratzen dugun estatistikoa H_0-k adierazten duen baliora hurbiltzen bada, hipotesi nulua onartu beharko dugu. Aldiz, laginetik 50 zentimetrotik asko urruntzen den batez bestekoa lortzen badugu, hipotesi alternatiboa onartu beharko dugu.
Laginetik ateratzen dugun batez bestekoak balio desberdinak har ditzake: 48.5 eta 51.5 balioen artean dagoen batez bestekoa lortuko bagenu, H_0 onartuko genuke ($H_0 \rightarrow \mu = 50$ zentimetro); eta, aldiz, batez bestekoa 48,5 baino txikiagoa edo 51,5 baino handiagoa izango balitz, hipotesi alternatiboa onartu behar genuke ($H_1 \rightarrow \mu \neq 50$ zentimetro).
48,5 zentimetro baino txikiagoak diren batez bestekoek edo 51,5 zentimetro baino handiagoak direnek frogaren eremu kritikoa osatzen dute; 48.5 eta 51.5 tartean dauden balioek, berriz, onarpen-eremua. Eremu kritikoaren eta onarpen eremuaren arteko mugak balio kritikoak zehazten ditu. Hipotesi-froga egiten denean, ohitura da H_0 hipotesiari buruz ondorioak ateratzea. Beraz, H_0 alboratzen da, H_1-en mesedetan, frogarako estatistikoa eremu kritikoan jauzten bada. Onarpen-eremuan jauzten bada, berriz, H_0 onartzen da.
Erabakiak hartzeko prozedura horrek emaitza okerretara eraman gaitzake. Adibidez, gerta daiteke umeen batez besteko neurria 50 zentimetro izatea, baina laginean hartu diren umeen azterketatik ateratzen den batez bestekoa eremu kritikoan jauztea. Kasu horretan, hipotesi hutsa (H_0) alboratu egingo da eta alternatiboa (H_1) onartu, errealitatean H_0 egia izanda. Ondorio oker horri LEHEN MOTAKO ERROREA deitzen zaio; hau da, hipotesi hutsa egiazkoa izanik, hau alde batera uzteko erabaki okerra hartuta egiten den errorea. Ondorio oker horri adierazgarritasun-maila edo α deitzen zaio. %95-eko konfiantza-maila izango bagenu, adierazgarritasun-maila 0,05 (%5) litzateke. %90eko konfiantza-mailaren adierazgarritasun-maila 0,10 (%10) da.
Demagun, orain, umeen batez besteko neurria ez dela 50 zentimetrokoa, nahiz eta laginean lortu dugun batez bestekoa onarpen-eremuaren barruan jauzten den. Kasu horretan, H_0 onartzen da, nahiz eta faltsua izan. Ondorio oker horri BIGARREN MOTAKO ERROREA izena deitzen zaio; hau da, hipotesi nulua okerra izanik, onartu egiten da. Ondorio oker horri β ere esaten zaio.
Beraz, edozein hipotesi estatistiko frogatzerakoan, azken erabakia okerra ala zuzena den zehazten duten lau egoera daude:

85. taula: errore motak hipotesi-frogetan		
Erabakia	H_0 egia da	H_0 gezurra da
H_0 onartu	Erabaki zuzena = Ez dago errorerik Probabilitatea = 1 – α	II. motako errorea edo β
H_0 onartu	I. motako errorea edo α	Erabaki zuzena = ez dago errorerik
Iturria: lanketa propioa		

Hipotesi-froga ona egiteko, garrantzitsua da erabaki okerrak edo erroreak gutxitzeko prozedura bat diseinatzea. Horretarako, honako hauek izan behar dira kontuan:

- I. motako errorea egiteko probabilitatea (α) bat dator adierazgarritasun-mailarekin. Hala, eremu kritikoaren tamaina eta, beraz, I. motako errorea egiteko probabilitatea adierazgarritasun-mailaren arabera aldatzen dira. Hipotesi-frogetan, aldez aurretik zehaztu behar dugu I. motako errorean jauzteko probabilitatea edo adierazgarritasun-maila (α).
- II. erako errorea egiteko probabilitatea (β) parametroaren eta aurresuposatutako balioaren arteko desberdintasunaren araberakoa da. Parametroaren eta estatistikoaren arteko desberdintasuna zenbat eta handiagoa izan, II. motako errorea egiteko probabilitatea txikiagoa izango da. β ezezaguna izaten da.
- I. motako zein II. motako erroreak gertatzeko probabilitatea gutxitu egiten da, zenbat eta lagin handiagoa izan.
- Lehen motako eta bigarren motako erroreak erlazionatuta daude. Baten probabilitatea gutxitzeak bestearen probabilitatea handitzea ekartzen du. Hala, β-ren eta α-ren artean alderantzizko harremana dago. Beraz, egokia da β-ren eta α-ren balioak ezagutzea. Dena

dela, praktikan α zehazten da eta β errorea gutxitzeko lagina handitzen; izan ere, horrela, adierazitako hipotesiaren konfiantza-mailaren mugak murrizten dira.

3.7.2. Hipotesi-froga egiteko pausoak

Hipotesi-froga egiteko, bost pauso eman behar dira:

1. Lehenengo eta behin, hipotesi nulua (hipotesi hutsa) eta hipotesi alternatiboa zehaztu behar dira. Hau da, aztergai den populazioaren parametroari buruzko baieztapenak egin behar dira. Horrekin batera, lagungarri da hipotesi-froga modu grafikoan adieraztea. Kontuan izan hipotesiak, betiere, parametroaren araberakoak direla. Hau da, lagina hartu den populazioa ebaluatu nahi da. Momentu horretan zehaztu behar da egin nahi dugun hipotesi-froga aldebakarrekoa ala aldebikoa den.
2. Ostean, adierazgarritasun-maila finkatu behar da; alegia, I. motako errorea egiteko probabilitatea edo adierazgarritasun-maila (α) aldez aurretik zehaztu behar da. Adierazgarritasun-maila erabilienak 0,01 eta 0,05 dira (%99 eta %95ko konfiantza-mailei dagozkienak, hurrenez hurren). Alegia, 0,05eko adierazgarritasun-maila aukeratzen bada, horrek esan nahi du 100 kasutik bostean hipotesia alboratuko genukeela onartu behar litzatekeenean; hau da, erabaki zuzena hartzen ari gara kasuen %95ean.
3. Onarpen-eremua zehaztu: erabakitako adierazgarritasun-mailaren arabera, onarpen-eremua eta eremu kritikoa desberdinduko dituzten $z_{\alpha/2}$ (froga aldebikoa denean), edo z_{α} (froga aldebakarrekoa denean) balioak zehaztuko ditugu.
4. Egiaztatu: lagina aztertuta, hipotesi nulua egia edo gezurra den erabakitzen lagunduko digun probarako estatistikoa aterako dugu.
5. Erabaki estatistikoa hartu: lortutako probarako estatistikoa eremu kritikoan kokatzen bada, hipotesi nulua errefusatuko da, α adierazgarritasun-mailarekin. Kontrako kasuan, hau da, probarako estatistikoa onarpen-eremuan badago, hipotesi nulua onartuko da, α adierazgarritasun-mailarekin.

3.7.3. Hipotesi-froga aldebikoak eta aldebakarrekoak

Egin nahi dugun ebaluazioaren arabera, hiru eratako hipotesi-frogak egin ditzakegu:

Hipotesi-froga aldebikoa :

$H_0 = k$ $\qquad\qquad$ $H_1 \neq k$

Aldebiko hipotesi-frogetan, hipotesi nuluak esaten du parametroa ez dela aldatu, eta alternatiboak aldatu egin dela. Hipotesi nulua parametroaren berezitasunei kasu eginda osatzen da, eta alternatiboa laginaren azterketatik ateratako datuen arabera. Dena dela, kontuan izan behar dugu hipotesiak, betiere, parametroaren araberakoak direla.

Hipotesi-froga aldebakarrekoa, α ezkerrean dagoenean:

$H_0 \geq k$ $\qquad\qquad$ $H_1 < k$

Kasu honetan, hipotesi alternatiboak esaten du parametroa txikitu egin dela.

Hipotesi-froga aldebakarrekoa, α eskuinean dagoenean:

$H_0 \leq k$ $\qquad\qquad$ $H_1 > k$

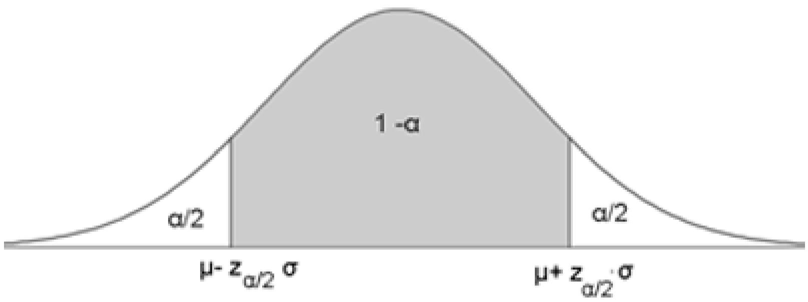
Hipotesi alternatiboak esaten du parametroa handitu egin dela.

Ikusten den bezala, hipotesi alternatiboak parametroa zein norabidetan aldatu den zehazten duenean (parametroa txikitu edo handitu egin da), hipotesi-froga aldebakarrekoak erabiliko ditugu; hipotesi alternatiboak parametroaren aldaketari buruzko norabidea zehazten ez duenean, berriz, hipotesi-froga aldebikoak.
Kontuan izan hipotesi-froga aldebikoetako eta aldebakarrekoetako eremu kritikoak tamaina berekoak direla, baina toki desberdinetan kokatuta. Aldebikoetan, adierazgarritasun-maila (α) batez bestekoarekiko bi parte simetrikotan banatzen da (bi buztanetan). α ezkerretara duten hipotesi-froga aldebakarrekoetan, adierazgarritasun-maila (α) batez bestekoaren ezkerretara kokatzen da (buztan bakarrean). α eskuinera duten hipotesi-froga aldebakarrekoetan, adierazgarritasun-maila (α) batez bestekoaren eskuinera kokatzen da (buztan bakarrean).

3.7.3.1. Hipotesi-froga aldebikoa (bi buztandun hipotesi-froga edo bi aldedun hipotesi-froga)

Ikerlariak, laginaren azterketatik ateratako ondorioak ikusita, parametroan aldaketarik egon den ikertu nahi du. Hipotesia era honetan osatzen da:
$H_0 \rightarrow \mu = k$ (edo $H_0 \rightarrow p = k$) (parametroa ez da aldatu; beraz, laginean atera dena anekdotikoa da)
$H_1 \rightarrow \mu \neq k$ (edo $H_1 \rightarrow p \neq k$) (aldaketa gertatu da parametroan, laginean atera dena ez da anekdotikoa)

60. diagrama: hipotesi-froga aldebikoa (bi buztandun hipotesi-froga edo bi aldedun hipotesi-froga)



Esan bezala, hipotesi-froga aldebikoetan, adierazgarritasun-maila (α) batez bestekoarekiko bi parte simetrikotan banatzen da (bi buztanetan). Alegia, α bi zatitan banatuta dago. Beraz, konfiantza-maila (1 - α) eta adierazgarritasun-maila (α) desberdinekiko, honako hauek dira balio kritiko ($z_{\alpha/2}$) erabilienak.

86. taula: konfiantza-maila, adierazgarritasun-maila eta balio kritiko erabilienak, bi aldedun hipotesi-frogetan

1 – α	α	$z_{\alpha/2}$
0.90	0.1	1.645
0.95	0.05	1.96
0.99	0.01	2.575

Iturria: lanketa propioa

Onarpen-eremua, kasu honetan, μ (edo parametroa) egon daitekeen tartea da.

Batez bestekoaren inguruko hipotesi-froga dugunean, onarpen-tartea honela erabakitzen da:
$(\mu - z_{\alpha/2} \cdot \sigma\ ,\ \mu + z_{\alpha/2} \cdot \sigma)$

Gogoratu batez bestekoaren errore tipikoa modu honetan kalkulatzen dela:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Proportzioen kasuan, onarpen-tartea honela kalkulatzen da:

$(P - z_{\alpha/2} \,.\, \sigma \,,\, P + z_{\alpha/2} \,.\, \sigma)$

Kasu honetan, errore tipikoa honela kalkulatzen da:

$$\sigma_p = \sqrt{\frac{pq}{n}}$$

Adibidea. Hipotesi-froga aldebikoa: 1etik 10rerako eskala bat erabilita, alderdi politiko bateko alkateari buruzko jarrera neurtu da, alkateren alderdiko militanteen artean. Politikariari emandako batez besteko puntuazioa 6koa da, 2,4-ko desbideratze tipikoarekin. Handik hiru hilabetera, alderdi politiko horretako 36 militante hartu ditugu, eta batez besteko jarrera 5,6koa dela ikusi dugu. Lagin horretako estatistikotik abiatuta, onar dezakegu alderdikideen batez besteko jarrera 6koa dela, %95eko konfiantza-mailarekin? Edo politikariarekiko jarrera aldatu egin da, laginaren azterketak erakusten duen bezala?

1. Hipotesi nulua eta alternatiboa zehaztu.

   $H_0 \rightarrow \mu = 6$; batez besteko jarrera 6koa da (jarrera ez da aldatu).

   $H_1 \rightarrow \mu \neq 6$; batez besteko jarrera 6 baino handiago edo txikiagoa da (jarrera aldatu egin da laginak adierazten duen bezala).

2. Adierazgarritasun-maila finkatu.

$\alpha = 0.05$ denean, $z_{\alpha/2} = 1.96$ da.

3. Onarpen-eremua zehaztu.

   Frogatu nahi dugun parametroa ezagututa ($\mu = 6$), %95eko konfiantza-mailari dagokion balio kritikoa ezagututa ($z_{\alpha/2} = 1{,}96$) eta errore tipikoa kalkulatuta (0,4), parametroaren onarpen-eremua zehaztu dezakegu:

   $(6 - 1{,}96 \cdot 0{,}4 \,;\, 6+1{,}96 \cdot 0{,}4) = (5{,}22 \,;\, 6{,}78)$

4. Egiaztatu.

   Laginaren azterketatik lortutako batez bestekoa = 5,6 da.

5. Erabakia hartu.

   Laginaren azterketatik lortutako batez bestekoa (5,6) onarpen-eremuan sartzen denez, hipotesi hutsa onartzen dugu, %95eko konfiantza-mailarekin. Alegia, alkatearekiko jarrera ez da aldatu.

Adibidea. Hipotesi-froga aldebikoa: badakigu eskualde bateko enpresen erdi-mailako kudeatzaileen batez besteko soldata 2.400 eurokoa dela, 300 euroko desbideratze tipikoarekin. 100 kudeatzaile hartzen dira, eta batez besteko soldata 2.320 eurokoa dela ondorioztatzen da. %95eko konfiantza-mailarekin, aldatu egin da soldata? Alegia, laginean atera den emaitza onartu, eta soldata aldatu egin dela esan behar dugu?

1. Hipotesi nulua eta alternatiboa zehaztu.

   $H_0 \rightarrow \mu = 2400$
   $H_1 \rightarrow \mu \neq 2400$

2. Adierazgarritasun-maila finkatu.

   $\alpha = 0.05$ denean, $z_{\alpha/2} = 1.96$ da.

3. Onarpen-eremua zehaztu.

$$2400 - 1{,}96.\frac{300}{\sqrt{100}};2400 + 1{,}96\frac{300}{\sqrt{100}} = (2341;2458{,}8)$$

4. Egiaztatu.

   Laginaren azterketatik lortutako batez bestekoa = 2.320 euro da.

5. Erabakia hartu.

   Laginaren azterketatik lortutako batez bestekoa (2.320) eremu kritikoan dagoenez, hipotesi hutsa (H_0) alboratzen dugu, %95eko konfiantza-mailarekin, eta ondorioztatzen dugu erdi mailako kudeatzaileen soldata aldatu egin dela.

Adibidea. Hipotesi-froga aldebikoa: herri bateko posta-zerbitzuan lan egiten duten postariek, urtean, batez beste, 800 ordu egiten dituzte etxez etxe posta banatzen, 40 orduko desbideratze tipikoarekin. 130 postariz osatutako lagina hartu dugu, eta, batez beste, posta banatzen 788 ordu ematen dituztela ikusi dugu. Esan dezakegu etxez etxe ematen dituzten orduen batez bestekoa aldatu egin dela laginak erakusten duen bezala? 0.05eko adierazgarritasun-maila erabili.

1. Hipotesi nulua eta alternatiboa zehaztu.

   $H_0 \rightarrow \mu = 800$ ordu
   $H_1 \rightarrow \mu \neq 800$ ordu

2. Adierazgarritasun-maila finkatu.

   $\alpha = 0.05$ denean, $z_{\alpha/2} = 1.96$ balio kritikoa dagokio

3. Onarpen-eremua zehaztu.

   Onarpen-eremua ezagutzeko, errore tipikoa ezagutu behar dugu:

$$\sigma = \frac{40}{\sqrt{130}} = 3{,}5$$

Frogatu nahi dugun parametroa ezagututa, %95eko konfiantza-mailari dagokion balio kritikoa ezagututa ($z_{\alpha/2}$ = 1,96) eta errore tipikoa kalkulatuta, parametroaren onarpen-tartea zehaztu dezakegu:

(800 - 1,96 · 3,5 ; 800 + 1,96 · 3,5) = (806,86 ; 793,17)

4. Egiaztatu.

Laginaren azterketatik lortutako batez bestekoa = 788 ordu.

5. Erabakia hartu.

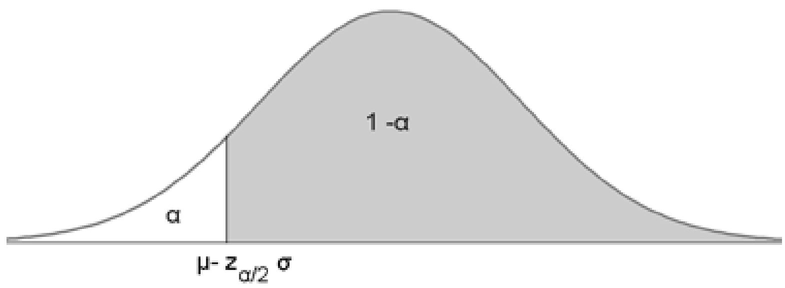
H_0 baztertzen dugu, eta onartzen dugu etxez etxe ematen dituzten orduen batez bestekoa aldatu egin dela.

3.7.3.2. Hipotesi-froga aldebakarrekoa, α ezkerretara dagoelarik

α ezkerretara dagoenean, ikerlariak parametroa txikitu egin den ikusi nahi du; alegia, laginean atera dena kontuan hartuta, parametro hasierako balioetan mantentzen den ala txikitu egin den aztertu. Hipotesia era honetan egiten da:

$H_0 \rightarrow \mu \geq k$ (edo $H_0 \rightarrow p \geq k$) erakoa da
$H_1 \rightarrow \mu < k$ (edo $H_1 \rightarrow p < k$) erakoa

61. diagrama: hipotesi-froga aldebakarrekoa, α ezkerretara dagoelarik



Kasu honetan, adierazgarritasun-maila (α) batez bestekoaren ezkerretara kokatzen da (buztan bakarrean). Beraz, konfiantza-maila (1-α) eta adierazgarritasun-maila (α) desberdinekiko, honelako hauek dira balio kritiko (z_α) erabilienak.

87. taula: balio kritiko erabilienak hipotesi-froga aldebakarrekoetan		
1 – α	A	z_α
0.90	0.10	1.28
0.95	0.05	1.645
0.99	0.01	2.33
Iturria: lanketa propioa		

Onarpen-eremua, kasu honetan, honelakoa da:

$(\mu - z_\alpha \cdot \sigma , \infty)$

Gogoratu batez bestekoaren errore tipikoa modu honetan kalkulatzen dela:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Proportzioen kasuan, onarpen-eremua honela kalkulatzen da:

$(P - z_\alpha . \sigma , \infty)$

Kasu honetan, errore tipikoa honela kalkulatzen da:

$$\sigma_p = \sqrt{\frac{pq}{n}}$$

Adibidea. Hipotesi-froga aldebakarrekoa, α ezkerretara dagoelarik: ikerketa batek adierazi du AHU alderdi politikoak botoen %40 lortuko duela hurrengo hauteskundeetan. 200 pertsonez osatutako lagina hartu dugu, eta horietariko 25ek (%12,5ek) botoa AHU alderdiari emango diotela esan dute. %1eko konfiantza-mailarekin, adierazi laginean oinarrituta egin den aurresatea onar daitekeen. Alegia, onar daiteke AHU alderdi politikoak hasieran pentsatzen zena baino boto gutxiago lortuko dituela?

1. Hipotesi nulua eta alternatiboa adierazten ditugu:

   $H_0 \rightarrow p \geq 0.40$: AHU alderdiak gutxienez botoen %40 lortuko du.

   $H_1 \rightarrow p < 0.40$: AHU alderdiak gehienez botoen %40 lortuko du;

2. Adierazgarritasun-maila finkatu.

   $\alpha = 0.01$ izanik, balio kritikoa (z_α) 2.33 da.

3. Onarpen-eremua zehaztu.

   Konfiantza-maila eta errore tipikoa ezagututa, onarpen eremua zehaztu dezakegu:

$$(0{,}4 - 2{,}33.\sqrt{\frac{0{,}4.0{,}6}{200}}; \infty) = (0{,}3192; \infty)$$

4. Egiaztatzea.

   Laginaren azterketan lortutako proportzioa = 0,125

5. Erabakia.

   Laginaren azterketatik lortu dugun proportzioa (0,125) onarpen-eremutik kanpo gelditzen da. Beraz, hipotesi nulua (H_0) alboratzen dugu. 0,01eko adierazgarritasun-mailarekin esan dezakegu AHU alderdiak botoen %40 baino gutxiago lortuko duela.

Adibidea. Hipotesi-froga aldebakarrekoa, α ezkerretara dagoelarik: eskola bateko umeek urtean, batez beste, 800 ordu ematen dituzte telebistaren aurrean. 50 umeko lagina hartu dugu, eta ikusi dugu telebistaren aurrean 750 ordu ematen dituztela. 0,01eko adierazgarritasun-mailarekin, onar dezakegu telebistaren aurrean ematen duten denbora gutxitu egin dela?

1. Hipotesi nulua eta alternatiboa adierazten ditugu:

   $H_0 \rightarrow \mu \geq 800$

   $H_1 \rightarrow \mu < 800$

2. Adierazgarritasun-maila finkatu.

   $\alpha = 0.01$ izanik, balio kritikoa ($z\alpha$)= 2.33 da.

3. Onarpen-eremua zehaztu.

Konfiantza-maila eta errore tipikoa ezagututa, onarpen eremua zehaztu dezakegu:

$$(800 - 2{,}33.\frac{120}{\sqrt{50}};\infty) = (760{,}46;\infty)$$

4. Egiaztatzea.

   Laginaren azterketan lortutako batez bestekoa = 750 ordu.

5. Erabakia.

   Hipotesi hutsa (H_0) alboratzen dugu, %99 konfiantza-mailarekin. Beraz, onar dezakegu umeek telebistaren aurrean ematen duten denbora gutxitu egin dela.

Adibidea. Hipotesi-froga aldebakarrekoa, α ezkerretara dagoelarik: enpresa batean, 300 minutu behar dituzte makina bat egiteko, 30 minutuko desbideratze tipikoarekin. 60 makinako lagina hartu da, eta ikusi da makina bat egiteko batez beste 290 minutu behar direla. %99ko konfiantza-mailarekin, onar dezakegu makina bat egiteko denbora gutxitu egin dela?

1. Hipotesi nulua eta alternatiboa adierazten ditugu:

   $H_0 \rightarrow \mu \geq 300$

   $H_1 \rightarrow \mu < 300$

2. Adierazgarritasun-maila finkatu.

   $\alpha = 0.01$; $1- \alpha = 0.99$ eta $z_\alpha = 2{,}33$.

3. Onarpen-eremua zehaztu.

$$300 - 2{,}33\frac{30}{\sqrt{60}};\infty = 291;\infty$$

4. Egiaztatzea.

   Laginaren azterketan lortutako batez bestekoa = 290 minutu

5. Erabakia.

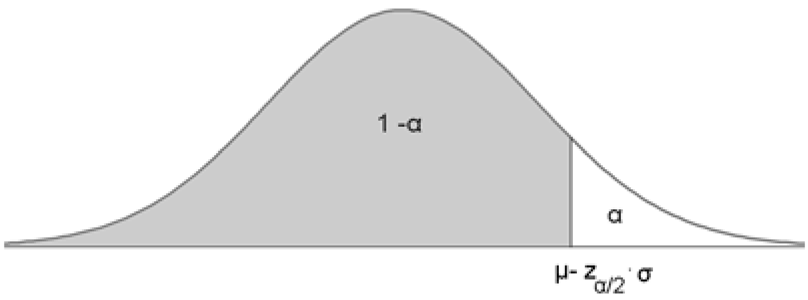
   Hipotesi nulua edo hutsa (H_0) alboratzen dugu, %99ko konfiantza-mailarekin. Beraz, onar dezakegu makina bat egiteko denbora murriztu egin dela.

3.7.3.3. Hipotesi-froga aldebakarrekoa, α eskuinera dagoelarik

α eskuinera dagoenean, ikerlariak, laginean atera dena aintzat hartuta, parametroa handitu den ikusi nahi du. Hipotesia honela egiten da:

$H_0 \rightarrow \mu \leq k$ (edo H_0: $p \leq k$)
$H_1 \rightarrow \mu > k$ (edo H_1: $p > k$)

62. diagrama: hipotesi-froga aldebakarrekoa, α eskuinera dagoelarik



Eremu kritikoa eta onarpen-eremua definitzeko, adierazgarritasun-maila batez bestekoaren eskuinera kokatzen da (buztan bakarrean). Beraz, lehengo adibidean agertu ditugun balio kritikoak erabiltzen dira (z_{α}):

88. taula: balio kritiko erabilienak		
1 – α	α	z_{α}
0.90	0.10	1.28
0.95	0.05	1.645
0.99	0.01	2.33
Iturria: lanketa propioa		

Onarpen-eremua, kasu honetan, honelakoa izango da:

$(-\infty, \mu + z_{\alpha} \cdot \sigma)$

Gogoratu batez bestekoaren errore tipikoa modu honetan kalkulatzen dela:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Proportzioen kasuan, onarpen-eremua honela kalkulatzen da:

$(-\infty, P + z_{\alpha} . \sigma)$

Kasu honetan, errore tipikoa honela kalkulatzen da:

$$\sigma_p = \sqrt{\frac{pq}{n}}$$

Adibidea. Hipotesi-froga aldebakarrekoa α eskuinera dagoelarik: txosten baten arabera, merkatuan dauden poltsikoko telefonoen batez besteko salneurria 120 eurokoa da, 40 euroko desbiderapen tipikoarekin. Azken astean, poltsikoko telefonoa erosi duten 100 pertsona hartu ditugu, eta ondorioztatu dugu pertsona horiek batez beste 129 € ordaindu dituztela beren telefonoen truke. 0,1eko adierazgarritasun-mailarekin, onar daitezke populazioari buruz hasieran eman ditugun datuak? Onar daiteke merkatuan dauden poltsikoko telefonoen salneurria gehienez jota 120 eurokoa dela?

1. Hipotesi nulua eta alternatiboa adierazten ditugu:

$H_0 \rightarrow \mu \leq 120$
$H_1 \rightarrow \mu > 120$

2. Adierazgarritasun-maila finkatu.

   $\alpha = 0.1$ adierazgarritasun-mailari $z_\alpha = 1.28$ balio kritikoa dagokio.

3. Onarpen-eremua zehaztu.

$$(-\infty;120+1{,}28.\frac{40}{\sqrt{100}}) = (-\infty;125{,}12)$$

4. Egiaztatzea.

   Laginaren azterketan lortutako batez bestekoa: 129 €.

5. Erabakia

   Hipotesi nulua baztertzen dugu, %10reko konfiantza-mailarekin.

Adibidea. Hipotesi-froga aldebakarrekoa α eskuinera dagoelarik: herriko ezkertiarrek diote herri horretako biztanleriaren %6k ezkerrari ematen diola bozka. 300ko lagina hartu da, eta 21ek esan dute ezkerrari ematen diola bozka. 0,01eko adierazgarritasun-mailarekin, onar dezakegu herriko ezkertiarrek diotena?

1. Hipotesi nulua eta alternatiboa adierazten ditugu:

   $H_0 \rightarrow p \leq 0.06$
   $H_1 \rightarrow p > 0.06$

2. Adierazgarritasun-maila.

   $\alpha = 0.01$ $z_\alpha = 2.33$

3. Onarpen-eremua zehaztu.

$$(-\infty;0{,}06+2{,}33.\sqrt{\frac{0{,}06.0{,}94}{300}}) = (-\infty;0{,}092)$$

4. Egiaztapena.

$$p = \frac{21}{300} = 0{,}07$$

5. Erabakia.

   Hipotesi hutsa (H_0) onartzen dugu %1eko konfiantza-mailarekin.

Adibidea. Hipotesi-froga aldebakarrekoa α eskuinera dagoelarik: Bilboko nagusientzako egokitzen artean, 100 pertsonako lagina hartu da, eta 71.8 urteko batez besteko adina dutela ikusi dugu. Udalak dituen datuen arabera, nagusien egoitzetako batez besteko adina 70 urtekoa da, 8,9ko desbideratze tipikoarekin. 0,05eko adierazgarritasun-mailarekin, esan daiteke populazioaren batez besteko adina 70 urtekoa dela?

1. Hipotesi nulua eta alternatiboa adierazten ditugu:

$H_0 \rightarrow \mu = 70$ urte
$H_1 \rightarrow \mu > 70$ urte

2. Adierazgarritasun-maila.

   $\alpha = 0.05$ adierazgarritasun-mailari $z_\alpha = 1,645$ balio kritikoa dagokio.

3. Hipotesiaren onarpen-eremua.

$(-\infty; 70 + 1,645.\ 0,89) = (-\infty; 71,464)$.

4. Egiaztatzea.

   Laginaren azterketan lortutako batez besteko adina 71,8 da.

5. Erabakia.

   Hipotesi hutsa baztertzen dugu, %95eko konfiantza-mailarekin.

3.8. LAGINAREN TAMAINA NOLA KALKULATU

Populazioa ezaugarri bat komunean duten bizidun, gauza, unitate edo elementuen multzoa da.
Lagina populazioaren azpimultzoa da; hain zuzen ere, populazioaren ezaugarriak ezagutzeko aukeratzen dena.
Gizarte-ikerlarion helburua lagin adierazgarri edo "lerratu gabea" lortzea da, populazioaren parametroak estimatzeko. Izan ere, lagina ikertu nahi den populazioaren adierazgarria ez bada, lagin horren azterketatik lortutako emaitzak populazio osoa aztertuz lortuko genituzkeen emaitzen oso desberdinak izango dira.
Limitearen teorema zentralak adierazi bezala, errore tipikoa laginaren tamainaren arabera aldatzen da (kopuru handien legea). Lagina handitzean, errore tipikoa txikitu egiten da, eta, beraz, egiten ditugun populazioaren estimazioak zehatzagoak dira. Hau da, laginaren azterketatik ateratzen dugun estatistikoa are eta gehiago hurbilduko da parametrora.

$\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$	$\sigma_p = \sqrt{\dfrac{pq}{n}}$

Errore tipikoa kalkulatzeko, laginaren tamainaren (n) erro karratuarekin zatitzen denez, lagina handitzean, errore tipikoa txikitu egiten da. Dena dela, ezin dugu pentsatu lagina limiterik gabe handitzean gure estimazioak ere etengabe hobeagotzen direnik. Izan ere, muga batetik aurrera, lagina handitzeak ez du errore tipikoa txikitzen. Hau da, lagina handitzearen errendimendua txikituz doa, eta kostuak, berriz, handitzen (inkesta gehiago egiteak diru gehiago gastatzea esan nahi du). Hala, gizarte ikerlarion helburua lagin-tamaina egokiena lortzea da, errore tipikoa ahal bezain txikiena izateko, baina, era berean, ikerketaren kostuak ahal den gehien murrizteko.

3.8.1. Laginaren tamainaren kalkulatzeko formula eta haren osagaiak

Estatistika inferentzialaren oinarriak aztertu ostean, laginaren kalkulurako behar ditugun osagai guztiak ulertzen ditugu, eta, horrekin, baita laginaren tamaina (n) kalkulatzeko formula ere :

$$n = \frac{z^2_{\frac{\alpha}{2}}.N.p.q}{e^2.(N-1) + z^2_{\frac{\alpha}{2}}.p.q}$$

Formula horren osagaiak azalduko ditugu ondoren:

e = laginketa-errorea

Arestian adierazi bezala, laginketa-erroreak estatistikoaren eta parametroaren arteko desberdintasuna adierazten du ehunekoetan.
Adibidez, herrialde bateko biztanleek hauteskundeetan emango dituzten botoei buruzko ikerketa egin dugu, eta ondorioztatu dugu alderdi politiko batek boto guztien %55 lortuko duela. Laginketa-errorea ±%2koa izanda, alderdi horrek bozka guztien %53tik %57ra bitartean lortuko lituzke hauteskundeetan (%55 ± %2).
Inkesta bat egin dugu hiri bateko nerabeen artean ±%3ko laginketa-errorearekin, eta %60k esaten du aste bukaeretan alkohola edaten duela. Horrek zera esan nahi du: alkohola edaten duten nerabeen ehunekoa %57 eta % 63 bitartekoa izan daitekeela (%60% ± %3ko laginketa-errorea).
Inkesta bat egin ostean, 120 pertsonak esan dute ZVP alderdi politikoaren alde bozkatuko dutela. ±%5eko laginketa-errorearekin lan egin badugu, alderdi horri bozka emango diotenak 114tik 126ra bitartean izan daitezke.
Ikusten den bezala, laginketa-errorea zenbat eta handiago izan, parametroari dagokion tartea handiago egiten da. Alegia, laginketa-errorea handituz, ziurtasunean irabazten dugu (adierazten dugun tartearen barnean, parametroa aurkitzeko probabilitatea handiagoa da), baina zehaztasunean galtzen (adierazten dugun tartearen zabalera handiagoa da).
Ia gehienetan, populazioaren berezitasunei buruz ezer jakin gabe zehaztu behar izaten dugu laginketa-errorea, eta, beraz, erabaki hori hartzea ez da erraza izaten. Hala ere, kontuan izan populazio bat zenbat eta heterogeneoagoa izan, orduan eta laginketa-errore txikiagoarekin lan egin beharko dugula.
Autore batzuen arabera, etxez etxe egindako inkestetan, ±%5ekoa da gehienez onar daitekeen laginketa-errorea; telefonoz egindako inkestetan, laginketa-errore handiagoak onartzen dira. ±%4 baino laginketa-errore txikiagoak dituztenak dira Kalitatezko ikerketak.
Bukatzeko, ez ahantzi ikerketa-txostenetan sartzen ditugun fitxa teknikoetan onartu dugun laginketa-errorea ehunekoetan adierazi behar dugula, baina formulan aplikatzerakoan, ehunekoen ordez, proportzioak erabiltzen direla.

N = populazioaren edo unibertsoaren tamaina

Arestian esan bezala, N ezaugarri bat komunean duten eta ikergai diren bizidun, gauza, unitate edo elementu guztien multzoa da. Populazioa homogeneoa edo heterogeneoa den jakitea garrantzitsua da, horrek laginketa-errore handiagoa edo txikiagoa onartzea eta lagin handiagoa edo txikiagoa hartzea esan nahi duelako.
Kontuan izan maila batetik gora populazioaren tamainak ez duela eraginik laginaren tamainan. Populazio handiak ditugunean, beraz, populazioaren tamaina handitzeak ia ez du eraginik laginaren tamainan.

$z_{\alpha/2}$ = balio kritikoa

Balio kritikoa $z_{\alpha/2}$ ikurraren bidez adierazten da, eta, banaketa normalaren banaketa funtziorako, taularen bidez kalkulatzen. Balio kritikoak ikertu nahi dugun konfiantza-maila eta adierazgarritasun-maila zehazten ditu.

Adierazgarritasun-maila (α) laginaren bidez lortutako estatistikoa parametro dagoen tarteetatik kanpo egoteko dagoen probabilitatea da; hau da, egindako estimazioa okerra izateko probabilitatea, edo estatistikoa eremu kritikoan jaustekoa. Ia gehienetan, 0,05 eta 0,01 adierazgarritasun-mailak erabiltzen dira.
Konfiantza-maila ($1 - \alpha$) laginetik lortutako estatistikoa parametroa dagoen tarteen barnean egoteko dagoen probabilitatea da; hau da, egindako estimazioa zuzena izateko probabilitatea. α adierazitako tartetik kanpo egoteko dagoen probabilitatea bada, $1 - \alpha$ konfiantza-maila da. Ia gehienetan, %95eko eta %99ko konfiantza-mailak erabiltzen dira, hain zuzen ere, 0,05 eta 0,01 adierazgarritasun-mailei dagozkienak.
%95eko konfiantza-mailaren araberako laginarekin lan egin nahi badugu, horrek esan nahi du lagin horretatik atera dezakegun edozein estatistikok kasuen %95ean bat egingo duela parametroarekin. Aldiz, kasuen %5ean, estatistikoa ez da parametroaren estimatzaile ona izango.
%99eko konfiantza-mailak esan nahi du laginetik atera dezakegun estatistikoak kasuen %99ean bat egingo duela parametroarekin. Aldiz, kasuen %1ean estatistikoa ez da parametroaren estimatzaile ona izango.
Beraz, konfiantza-maila handitzeak zehaztasun handiagoarekin lan egitea esan nahi du, eta, horren ondorioz, baita laginaren tamaina handitzea ere.
Eremu kritikoa zehazten duten balio kritikoak ezagutzeko, konfiantza-maila bakoitzari dagokion $z_{\alpha/2}$ balioa aurkitzen dugu, eta banaketa normalaren simetria erabiliz, eremu kritikoa $-z_{\alpha/2}$tik behera eta + $z_{\alpha/2}$tik gora dagoela esaten. Beste era batera esanda, balio kritikoa bere eskuinera $\alpha/2$ tamainako azalera uzten duen z balioa da, eta, banaketa normalaren simetria erabiliz, eremu kritikoa banaketa normal estandarraren $-z_{\alpha/2}$tik behera eta + $z_{\alpha/2}$tik gora dagoela esango dugu.
Dena dela, lagin bat kalkulatu behar dugun bakoitzean ez dugu zertan aritu $z_{\alpha/2}$ balio kritikoa bilatzen. Izan ere, kalkulatuta daude konfiantza-maila eta adierazgarritasun-maila bakoitzari dagozkion balio kritikoak:

89. taula: konfiantza-maila, adierazgarritasun-maila eta balio kritikoak		
$1 - \alpha$	α	$z_{\alpha/2}$
0,75	0,25	1,15
0,80	0,20	1,28
0,85	0,15	1,44
0,90	0,10	1,65
0.95	0.05	1.96
0,975	0,025	2,24
0.99	0.01	2.575
iturria: lanketa propioa		

Egiatan, gehien erabiltzen diren balio kritikoak %95eko eta %99ko konfiantza-mailari dagozkienak dira; hau da, $z_{\alpha/2} = 1{,}96$ eta $z_{\alpha/2} = 2{,}58$, hurrenez hurren.

Adibidea. 90. taula: Euskal Herrian (2.906.000 biztanle) ikerketak egiteko lagin-tamainak, laginketa-errore eta konfiantza-maila desberdinen arabera		
Lagin-tamaina desberdinak	%95eko konfiantza-mailarako laginketa-erroreak	%99ko konfiantza-mailarako laginketa-erroreak
600	± % 4	± %5,2
700	± %3,7	± % 4,8
800	± % 3,4	± % 4,5
900	± % 3,26	± % 4,2
1.000	± % 3	± % 4
1.100	± %2,95	± % 3,8
1.200	± % 2,8	± % 3,7
Iturria: lanketa propioa		

p eta q: populazioan ikergai den berezitasuna dutenen eta ez dutenen proportzioa

p eta q (1-p) zerbait gertatzeko eta ez gertatzeko dauden probabilitateak dira.

Adibidea. 100 pertsonako lagina badugu, eta 20k galdera bati baietz erantzuten badiote eta 80k, ezetz, $p = \frac{20}{100} = 0{,}2$ da; q, beraz, 0,8.

p-k eta q-k populazioa homogeneoa edo heterogeneoa den adierazten digute. P = 0,2 bada eta q = 0,80, lagin homogeneoa da, populazioa osatzen duten gehienek q-ren alde egiten dutelako. Aldiz, p = 0,5 eta q = 0,5 bada, hau da, p eta q berdinak badira, lagina heterogeneoa dela onartzen dugu.

Populazio zenbat eta heterogeneoagoa izan, laginketa-errore txikiagoak edota konfiantza-maila handiagoak onartu beharko ditugu. Hau da, populazioa zenbat eta heterogeneoagoa izan elementu, unitate edo pertsona gehiago hartu beharko ditugu populazioaren bariantza guztia behar bezala islatzeko.

Lagin bat aukeratu behar dugunean, ia gehienetan p eta q ez ditugu ezagutzen. Ez dakigu, aldez aurretik, populazio homogeneo edo heterogeneo bat den. Aukera bat da aurretesta egitea, edo aurretik egin diren ikerketa eta teorietan oinarritzea p eta q-ren balioetara hurbiltzeko. Dena dela, gehienetan, onartzen dena da p eta q berdinak direla (p = q = 0.5). Hau da, onartzen da egon daitekeen heterogeneotasun maila altuena dagoela.

Populazio baten homogeneotasun edo heterogeneotasun maila ezagutzeko, desbideratze tipikoa ere erabil dezakegu. Kasu horretan, lagina kalkulatzeko formula honako hau da:

$$n = \frac{z^2_{\frac{\alpha}{2}}.N.\sigma^2}{e^2.(N-1) + z^2_{\frac{\alpha}{2}}.\sigma^2}.$$

Gehienetan, populazioaren desbideratze tipikoa ez da ezaguna izaten, eta, beraz, beste formula batzuk erabiltzen da.

Adibidea. Lagina kalkulatzea: demagun 17.500 biztanleko herria dugula. Nolako lagina hartu beharko dugu %95eko konfiantza-mailarekin, ±%3,5eko laginketa-errorearekin eta heterogeneotasun-maila altuenarekin?

$n = \frac{z^2_{\frac{\alpha}{2}}.N.p.q}{e^2.(N-1) + z^2_{\frac{\alpha}{2}}.p.q}$	$n = \frac{1{,}96^2.17.500.0{,}5.0{,}5}{0{,}035^2(17500-1) + 1{,}96^2.0{,}5.0{,}5} = 750$

Ikusten den bezala, %95eko konfiantza-mailarekin, ±%3,5eko laginketa-errorearekin eta p = q dela onartuta, 750 pertsonako lagina hartu behar genuke.

Demagun, %95eko konfiantza-maila erabili beharrean %99ko konfiantza-maila erabili nahi dugula.

%95eko konfiantza-mailak esan nahi du lagin horren azterketatik atera dezakegun estatistikoak kasuen %95ean bat egingo duela parametroarekin, eta %99ko konfiantza-mailak estatistikoak kasuen %99an bat egingo duela parametroarekin. Beraz, %95eko konfiantza-mailatik %99ko konfiantza-mailara pasatzeak zehaztasun handiagoarekin lan egiteko nahia adierazten du.

$n = \frac{z^2_{\frac{\alpha}{2}}.N.p.q}{e^2.(N-1) + z^2_{\frac{\alpha}{2}}.p.q}$	$n = \frac{2{,}58^2.17.500.0{,}5.0{,}5}{0{,}035^2(17500-1) + 2{,}58^2.0{,}5.0{,}5} = 1260$

Ikusten den bezala, konfiantza-maila handitzean, laginaren tamaina ere handitu egin da. Hain zuzen ere, 1.260 pertsona hartu behar ditugu %99ko konfiantza-mailarekin. Lagina 510 pertsonatan handitu behar dugu zehaztasun handiagoarekin lan egin nahi badugu; eta horrek esan nahi du diru gehiago gastatu behar dugula.
Demagun %95eko konfiantza-mailarekin lan egin nahi dugula, baina laginketa-errore handiagoa onartzeko prest gaudela. Hain zuzen ere, ±%5eko laginketa-errorea onartzeko prest gaude.

$n=\dfrac{z^2_{\frac{\alpha}{2}}.N.p.q}{e^2.(N-1)+z^2_{\frac{\alpha}{2}}.p.q}$	$n=\dfrac{1{,}96^2.17.500.0{,}5.0{,}5}{0{,}05^2(17500-1)+1{,}96^2.0{,}5.0{,}5}=376$

Onartuko genukeen laginketa-errorea handitzean, laginaren tamaina txikitu egiten da. Hain zuzen ere, 376 pertsonako lagina hartu behar dugu %95eko konfiantza-mailarekin eta %5eko laginketa-errorearekin.

Adibidea. Lagina kalkulatzea populazio handiak ditugunean: 1.700.500 biztanleko hiriburu batean ikerketa egiteko laginaren tamaina kalkulatu genuen %95eko konfiantza-maila eta ±%5eko laginketa-errorearekin.

$n=\dfrac{z^2_{\frac{\alpha}{2}}.N.p.q}{e^2.(N-1)+z^2_{\frac{\alpha}{2}}.p.q}$	$n=\dfrac{1{,}96^2.1700.500.0{,}5.0{,}5}{0{,}05^2(1700500-1)+1{,}96^2.0{,}5.0{,}5}=384$

Ikusten den bezala, 1.700.500 hiriburu batean, 384 pertsonako lagina hartu behar genuke eta 17.500 biztanleko herri batean 376 pertsonakoa. Populazioa askoz handiagoa bada ere, laginaren tamaina ez da asko aldatzen.
Lehenengo ikerketa egin eta hiru urtera, hiriburu horretan, ikerketa bera egin nahi da, baina populazioa handitu egin dela ohartu gara; hain zuzen ere, populazioa 1.800.000koa dela ohartzen gara.

$n=\dfrac{z^2_{\frac{\alpha}{2}}.N.p.q}{e^2.(N-1)+z^2_{\frac{\alpha}{2}}.p.q}$	$n=\dfrac{1{,}96^2.1800.000.0{,}5.0{,}5}{0{,}05^2(1800000-1)+1{,}96^2.0{,}5.0{,}5}=384$

Ikusten den bezala, 1.700.500 biztanlerekin eta 1.800.000rekin, lagin-tamaina bera hartu behar dugu.

Adibide horiek guztiak ikusi ostean, honako ondorio hauek ateratzen ditugu:

- Laginaren tamaina oso sentikorra da laginketa-errorearekiko. Laginketa-errorea erdira murriztuta, laginaren tamaina laukoiztu egiten da.
- Laginaren tamaina aldatu egiten da konfiantza-mailaren arabera.
- Maila batetik gora, populazioaren tamainak ez du eraginik laginaren tamainan. Beraz, baliteke hiri txiki batean edo estatu oso batean lagin-tamaina bera hartu behar izatea.

3.8.2. Lagina eta fitxa teknikoa

Arestian esan bezala, gure ikerketa-txostenen eranskinetan ikerketan erabili ditugun eta hura hobeto ulertzeko lagungarriak izan daitezkeen material guztiak sartuko ditugu; besteak beste, ikerketaren fitxa teknikoa sartu behar dugu, honako informazio honekin:

- datuak bildu direneko datak
- datuak biltzeko teknika
- laginaren tamaina
- laginketa-errorea
- lagina aukeratzerakoan onartu den konfiantza-maila
- ikerlarien izenak
- q-ren eta p-ren balioak

Adibidea. Ikerketaren lagina eta fitxa teknikoa: Euskal AEko Prospekzio Soziologikoen Kabinetearen 51. soziometroa:

- Informazio-bilketa 2013ko urtarrilaren 31 eta otsailaren 7a bitartean egin zen -biak barne-, galdesorta egituratu eta itxia erabiliz, Euskal Autonomia Erkidegoko (EAEko) lurralde bakoitzerako lagin adierazgarri bati etxean egindako banakako elkarrizketen bidez.
- Lagina, 18 urte edo gehiagoko biztanleria; honela banatu zen: 490 pertsona Araban, 1.114 Bizkaian eta 597 Gipuzkoan, beraz, 2.201 pertsona elkarrizketatu ziren guztira.
- Gizabanakoen hautaketa-prozedura polietapiko eta estratifikatuaren bidez egin zen, ausazko ibilbideak erabiliz -170 laginketa abiapunturekin-, eta, ondoren, pertsonak sexuaren, adinaren eta lan egoeraren araberako kuoten bidez aukeratu ziren. Emaitzak 2012ko Hauteskunde Autonomikoetako boto-oroimenaren arabera eta hiru lurraldeetako biztanleria zentsuaren arabera haztatu dira.
- Ikerketaren diseinua, emaitzen azterketa eta txostena egitearen ardura Prospekzio Soziologikoen Kabineteari dagozkio soilik. Bestalde, informazio-bilketa Ikertalde enpresak egin zuen: Lurra eraikina, Txingurri pasealekua 28-30, 2, 20017 Donostia.
- 2.201 pertsonako lagin horri dagokion lagin-errorearen estimazioa, erabat ausazkoak diren laginketei egozgarria, ±%2,13 da, EAE osorako, %95,5eko konfiantza-mailarako, p = q = 0,5 izanik.
- Inkesten %62,4 telefono bidez kontrolatu dira, eta %18,5, berriz, etxean bertan.

ERANSKINAK

1. FORMULA GARRANTZITSUENEN ZERRENDA

Batez besteko aritmetiko sinplea nola kalkulatu, balio isolatuetarako

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \ldots + X_n}{N}$$

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{N}$$

Batez besteko aritmetiko sinplea nola kalkulatu, maiztasun-tauletan banatuta dauden aldagaietarako

$$\bar{X} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \ldots + x_n f_n}{N}$$

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i f_i}{N}$$

Batez besteko aritmetiko haztatua

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i} = \frac{x_1 w_1 + x_2 w_2 + x_3 w_3 + \ldots + x_n w_n}{w_1 + w_2 + w_3 + \ldots + w_n}$$

Mediana nola kalkulatu, datuak tartetan bildurik daudenean

$$Me = L_i + \frac{\frac{N}{2} - F_{i-1}}{f_i} \cdot a_i$$

Moda nola kalkulatu, datuak tartetan bildurik daudenean

$$Mo = L_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \cdot a_i$$

Ibiltartea

$$I = x_{max} - x_{min}$$

Kuartil arteko ibiltartea edo ibiltarte kuartilikoa

$$KAI = Q_3 - Q_1$$

Kuartilen desbideratzea edo ibiltarte semiinterkuartilikoa

$$\frac{Q_3 - Q_1}{2}$$

Batez bestekoarekiko batez besteko desbideratzea edo batez besteko desbideratze absolutua

$$D_{\bar{x}} = \frac{|x_1 - \bar{x}| + x_2 - \bar{x}| + \ldots + |x_n - \bar{x}|}{N}$$

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

Batez bestekoarekiko batez besteko desbideratzea edo batez besteko desbideratze absolutua, maiztasun-tauletan banatuta dauden aldagaietarako

$$D_{\bar{x}} = \frac{|x_1 - \bar{x}| f_1 + |x_2 - \bar{x}| f_2 + \ldots + |x_n - \bar{x}| f_n}{N}$$

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

Bariantza nola kalkulatu, datu isolatuak ditugunean

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{N}$$

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

$$\sigma^2 = \frac{x_1^2 + x_2^2 + \ldots + x_n^2}{N} - \bar{x}^2$$

$$\sigma^2 = \sum_{i=1}^{n} \frac{x_i^2}{N} - \bar{x}^2$$

Bariantza nola kalkulatu, maiztasun-tauletan banatuta dauden aldagaietarako

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 f_1 + (x_2 - \bar{x})^2 f_2 + \ldots + (x_n - \bar{x})^2 f_n}{N}$$

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

$$\sigma^2 = \frac{x_1^2 f_1 + x_2^2 f_2 + \ldots + x_n^2 f_n}{N} - \bar{x}^2$$

$$\sigma^2 = \sum_{i=1}^{n} \frac{x_i^2 f_i}{N} - \bar{x}^2$$

Desbideratze estandarra edo desbideratze tipikoa nola kalkulatu, datu isolatuak ditugunean

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{N}}$$

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

$$\sigma = \sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_n^2}{N} - \bar{x}^2}$$

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

Desbideratze tipikoa nola kalkulatu, maiztasun-tauletan banatuta dauden aldagaietarako

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 f_1 + (x_2 - \bar{x})^2 f_2 + \ldots + (x_n - \bar{x})^2 f_n}{N}}$$

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

$$\sigma = \sqrt{\frac{x_1^2 f + x_2^2 f_2 + \ldots + x_n^2 f_n}{N} - \bar{x}^2}$$

$$\sigma = \sqrt{\sum_{i=1}^{n} \frac{x_i^2 f_i}{N} - \bar{x}^2}$$

Ibiltarte erlatiboa

$$I_{\bar{x}} = \frac{I}{\bar{x}}$$

Kuartil arteko ibiltarte erlatiboa

$$KAIE = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Aldakortasun-koefizientea edo Pearson-en aldakuntza-koefizientea

$$AK = \frac{\sigma}{\bar{x}}$$

$$AK = \frac{\sigma}{\bar{x}}.100$$

Kuartilak nola kalkulatu, maiztasun-tauletan banatuta dauden aldagaietarako

$$Q_k = L_i + \frac{\frac{k \cdot N}{4} - F_{i-1}}{f_i} \cdot a_i \qquad k = 1,2,3$$

Pertzentzilak nola kalkulatu, datuak tartetan bildurik daudenean

$$P_k = L_i + \frac{\frac{k \cdot N}{100} - F_{i-1}}{f_i} \cdot a_i \qquad k = 1, 2, ..99$$

Dezilak nola kalkulatu, datuak tartetan bildurik daudenean

$$D_k = L_i + \frac{\frac{k \cdot N}{10} - F_{i-1}}{f_i} \cdot a_i \qquad k = 1, 2, ..9$$

Karl Pearsonen asimetria-koefizientea

$$A_{K1} = \frac{\overline{x} - Mo}{s_{\overline{x}}}$$

$$A_{K2} = \frac{3(\overline{x} - Me)}{s_{\overline{x}}}$$

Fisher-en alborapen- koefizientea

$$g_1 = \frac{1}{N} \cdot \frac{\sum_{i=1}^{n} (x_i - \overline{x})^3 \cdot f_i}{\sigma^3}$$

Bowley-en alborapen- koefizientea

$$A_B = \frac{(Q_3 - Me) - (Me - Q_1)}{Q_3 - Q_1}$$

Fisherren kurtosi-neurria

$$\alpha = \frac{\Sigma f(x_i - \bar{x})^4}{n\sigma^4}$$

$$g_2 = \frac{\sum (x_i - \overline{X})^4 f_i}{ns^4} - 3$$

Giniren indizea

$$C_G = \frac{\sum_{i=1}^{n-1} (p_i - q_i)}{\sum_{i=1}^{n-1} p_i}$$

Banaketa normal estandarrera eraldatzeko formula

$$z = \frac{x_i - \mu}{\sigma}$$

Batez bestekoen errore tipikoa

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Batez bestekoen errore tipikoa, populazio txikiak ditugunean

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Proportzioen errore tipikoa

$$\sigma_p = \sqrt{\frac{p.q}{n}}$$

Proportzioen errore tipikoa, populazio txikiak ditugunean

$$\sigma_p = \sqrt{\frac{p.q}{n}}\sqrt{\frac{N-n}{N-1}}$$

Tarte-estimazioa egiteko formula

$$Parametroa = estatistikoa \pm z.errore \cdots tipikoa$$

$$\mu = \bar{x} \pm z.\sigma_{\bar{x}}$$

$$P = p \pm z.\sigma_p$$

Laginaren tamaina nola kalkulatu

$$n = \frac{z^2_{\frac{\alpha}{2}}.N.p.q}{e^2.(N-1) + z^2_{\frac{\alpha}{2}}.p.q}$$

2. KURBA NORMALAREN AZPIKO AZALERA-TAULA

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.990097	0.990358	0.990613	0.990863	0.991106	0.991344	0.991576
2.4	0.991802	0.992024	0.992240	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613
2.5	0.993790	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.995060	0.995201
2.6	0.995339	0.995473	0.995604	0.995731	0.995855	0.995975	0.996093	0.996207	0.996319	0.996427
2.7	0.996533	0.996636	0.996736	0.996833	0.996928	0.997020	0.997110	0.997197	0.997282	0.997365
2.8	0.997445	0.997523	0.997599	0.997673	0.997744	0.997814	0.997882	0.997948	0.998012	0.998074
2.9	0.998134	0.998193	0.998250	0.998305	0.998359	0.998411	0.998462	0.998511	0.998559	0.998605
3.0	0.998650	0.998694	0.998736	0.998777	0.998817	0.998856	0.998893	0.998930	0.998965	0.998999
3.1	0.9990324	0.9990646	0.9990957	0.9991260	0.9991553	0.9991836	0.9992112	0.9992378	0.9992636	0.9992886
3.2	0.9993129	0.9993363	0.9993590	0.9993810	0.9994024	0.9994230	0.9994429	0.9994623	0.9994810	0.9994991
3.3	0.9995166	0.9995335	0.9995499	0.9995658	0.9995811	0.9995959	0.9996103	0.9996242	0.9996376	0.9996505
3.4	0.9996631	0.9996752	0.9996869	0.9996982	0.9997091	0.9997197	0.9997299	0.9997398	0.9997493	0.9997585
3.5	0.9997674	0.9997759	0.9997842	0.9997922	0.9997999	0.9998074	0.9998146	0.9998215	0.9998282	0.9998347
3.6	0.9998409	0.9998469	0.9998527	0.9998583	0.9998637	0.9998689	0.9998739	0.9998787	0.9998834	0.9998879
3.7	0.9998922	0.9998964	0.99990039	0.99990426	0.99990799	0.99991158	0.99991504	0.99991838	0.99992159	0.99992468
3.8	0.99992765	0.99993052	0.99993327	0.99993593	0.99993848	0.99994094	0.99994331	0.99994558	0.99994777	0.99994988
3.9	0.99995190	0.99995385	0.99995573	0.99995753	0.99995926	0.99996092	0.99996253	0.99996406	0.99996554	0.99996696
4.0	0.99996833	0.99996964	0.99997090	0.99997211	0.99997327	0.99997439	0.99997546	0.99997649	0.99997748	0.99997843