

ESTATISTIKA

DESKRIBATZAILEA

Koerlazioa, erregresioa eta
datu anizkoitzen analisisia

2. argitarapena

Karmele Fernandez Agirre

© Karmele Fernandez Agirre
© Udako Euskal Unibertsitatea

I.S.B.N.: 84-86967-53-8
Lege-gordailua: BI-1832-93

Inprimategia: BOAN S.A. Padre Larramendi 2. BILBO
Azalaren diseinua: Julio Pardo

Banatzaileak: UEU. General Concha 25, 6. BILBO
Zabaltzen: Igarabide, 88. DONOSTIA

AURKIBIDEA

HITZAURREA	XI
-------------------------	-----------

I. EZAUGARRI ESTATISTIKO BAKUNAK, BANAKETAK, TAULAK, ADIERAZPIDE GRAFIKOAK

I.1. ESTATISTIKAREN HISTORIA LABURRA. ESTATISTIKA ETA PROBABILITATEA	3
I.2. OROKORTASUNAK	5
I.2.1. Populazioa eta lagina	5
I.2.2. Unitate estatistiko edo indibiduoak	6
I.3. EZAUGARRI ESTATISTIKOAK	6
I.3.1. Ezaugarri estatistiko bakunak	6
I.3.1.1. Ezaugarri kuantitatiboak	7
I.3.1.2. Ezaugarri kualitatiboak	7
I.4. MAIZTASUN-BANAKETA BAKUNAK	8
I.4.1. Maiztasun-banaketak (absolutuak, erlatiboak, eta metatuak)	8
I.4.2. Taulak	9
I.5. MAIZTASUN-BANAKETA BAKUNEN ADIERAZPIDE GRAFIKOAK	13
I.5.1. Barra-diagramak	13
I.5.2. Histogramak	15
I.5.3. Maiztasun-poligonoak	17
I.5.4. Diagrama linealak	18
I.5.5. Beste adierazpide grafikoak	18

II. EZAUGARRI BAKUNEN BALIO TIPIKOAK

II.1. MOMENTUAK	23
II.1.1. Momentu arruntak edo jatorriarekikoak	23
II.1.2. Momentu zentralak edo batezbestekoarekikoak	24

II.2. MAIZTASUN-BANAKETA BAKUNEN BALIO

<i>TIPIKO EDO ESTATISTIKOAK</i>	26
<i>II.2.1. Zentru-joeraren balio tipikoak</i>	26
<i>II.2.1.1. Batezbesteko aritmetikoa</i>	27
<i>II.2.1.2. Batezbesteko geometrikoa</i>	30
<i>II.2.1.3. Batezbesteko harmonikoa</i>	31
<i>II.2.1.4. Moda</i>	31
<i>II.2.1.5. Mediana</i>	34
<i>II.2.2. Sakabanatzearen balio tipikoak</i>	36
<i>II.2.2.1. Bariantza</i>	37
<i>II.2.2.2. Batezbesteko desbidazioa</i>	38
<i>II.2.2.3. Ibiltarte</i>	39
<i>II.2.3. Itxuraren balio tipikoak</i>	40
<i>II.2.3.1. Asimetri koefizientea</i>	40
<i>II.2.3.2. Kurtosi-koefizientea edo</i> <i>zapaltasun/zorroztasun-koefizientea</i>	41
<i>II.3. KONTZENTRAZIO-NEURKETAK</i>	43
<i>II.3.1. LORENZ-en kurba</i>	43
<i>II.3.2. GINI-ren indizea</i>	46

III. ZENBAKI INDIZEAK

<i>III.1. INDIZE SINPLEAK</i>	51
<i>III.2. PONDERAZIO GABEKO INDIZE KONPLEXUAK</i>	52
<i>III.2.1. Batezbesteko aritmetikoa sinplearen metodoa</i>	52
<i>III.2.2. Batezbesteko agregatu sinplearen metodoa</i>	52
<i>III.3. INDIZE KONPLEXU PONDERATUAK</i>	58
<i>III.3.1. Balioen, prezioen eta kopuruen indizeak</i>	58
<i>III.3.1.1. LASPEYRES-en indizeak</i>	61
<i>III.3.1.2. PAASCHE-ren indizeak</i>	63
<i>III.3.1.3. FISHER-en indizeak</i>	65
<i>III.3.1.4. Propietate eta erlazio batzu: Batera-</i> <i>garritasuna eta alderantzizkotasuna</i>	65
<i>III.3.1.5. Kalkulaketa</i>	66
<i>III.4. INDIZE KONPLEXUEN ERAIKETAN SORTZEN</i> <i>DIREN ZENBAIT ERAGOZPEN</i>	68

III.4.1. Aldagaien hautapena	68
III.4.2. Somatutako leku eta denboraren hautapena	68
III.4.3. Talde eta azpitaldeen hautapena	68
III.4.4. Oinarri-denboraren hautapena	68
III.4.5. Formula eta ponderazioen hautapena	69
III.4.6. Indizearen esangura eta zabaldua	69
III.5. ERAGOZPEN BEREZI BATZU	70
III.5.1. Oinarri-aldaketa indize sinpleetan	70
III.5.2. Berriztapen eta loturak indize konplexuetan	70
III.6. ZENBAKI-INDIZEEN APLIKAPENAK	72
III.6.1. Kontsumo-prezioen indizeak	72

IV. EZAUGARRI ESTATISTIKO BIKOITZAK, BANAKETAK, TAULAK, ADIERAZPIDE GRAFIKOAK

IV.1. EZAUGARRI ESTATISTIKO BIKOITZAK	77
IV.2. MAIZTASUN-BANAKETA BIKOITZAK	77
IV.2.1. Maiztasun-banaketa bikoitzak eta bazter-maiztasunak	77
IV.2.2. Maiztasun-banaketa bikoitzak: Taulak	79
IV.3. MAIZTASUN-BANAKETA BIKOITZEN ADIERAZPIDE GRAFIKOAK	81
IV.3.1. Sakabanatze-diagrama edo puntu-hodeia	81
IV.4. MAIZTASUN-BANAKETA BALDINTZATUAK	83
IV.5. TAULA BATEN DEPENDENTZIA EDO INDEPENDENTZIA	86
IV.6. KONTINGENTZI TAULA BATEN ERRENKADA ETA ZUTABEEN BATEZBESTEKO SOSLAIK	88

V. EZAUGARRI BIKOITZEN BALIO TIPIKOAK

V.1. MOMENTUAK	91
V.1.1. Momentu arruntak edo jatorriarekikoak	91
V.1.2. Momentu zentralak edo batezbestekoarekikoak	92

V.2. MAIZTASUN-BANAKETA BIKOITZEN BALIO	
TIPIKO EDO ESTATISTIKOAK	93
V.2.1. Kobariantza: S_{xy}	93
V.2.2. Koerlazio-koefizientea: r_{xy}	94
V.3. ALDAGAIEN ALDAKETAK	96
V.3.1. Aldagai zentratua	99
V.3.2. Aldagai tipifikatua edo standardizatua	100
V.4. ALDAGAI KOERLAZIO GABEEN BI PROPIETATE	101
V.5. SAKABANATZE ETA KOERLAZIO-MATRIZEAK	102
V.6. OROKORPENA	102

VI. KOERLAZIOA ETA ERREGRESIOA $\mathbb{R}^2$ -n

VI.1. KOERLAZIOA ETA ERREGRESIOA	107
VI.2. ERREGRESIO-TEORIA	107
VI.2.1. Batezbestekoaren erregresioa	108
VI.2.2. Karratu txikiaren erregresioa	109
VI.2.2.1. Karratu txikiaren erregresio lineala	110
VI.2.2.2. Karratu txikiaren erregresio parabolikoa.	117

VII. ALDAGAI ESTATISTIKO n -KOITZEN MATRIZE-ESTATISTIKOAK

VII.1. DATU-MATRIZEAK	123
VII.2. BATEZBESTEKO-BEKTORRA: PROPIETATEAK	124
VII.3. KOBARIANTZ MATRIZEA	125
VII.3.1. Kobariantz matrizearen propietateak	126
VII.4. KOERLAZIO-MATRIZEA	127
VII.4.1. Koerlazio-matrizearen propietateak	129

VIII. ERREGRESIO ANIZKOITZA

VIII.1. SARRERA	133
VIII.2. ERREGRESIO-PLANOA $\mathbb{R}^3$ -n	134

VIII.3. ERREGRESIO-HIPERPLANOA $\mathbb{R}^n$ -n	137
VIII.3.1. Karratu txikiaren erregresioaren propietateak ..	138
VIII.3.2. Mugatze-koefizientea eta koerlazio-koefiziente anizkoitza	142
VIII.4. “BETA” ERREGRESIO-KOEFIZIENTE STANDARDIZATUAK	143

IX. KOERLAZIO PARTZIALA

IX. 1. SARRERA	147
IX.2. KOERLAZIO PARTZIALA: KONTZEPTU MATEMATIKOA	150
IX.3. ERREPIKAPEN-FORMULA	154
IX.4. KOERLAZIO-KOEFIZIENTE PARTZIALA ETA “ β ” KOEFIZIENTEAK	155

X. DESKRIBAPEN ESTATISTIKOAREN ADIERAZPIDE GEOMETRIKOAK

X.1. OROKORTASUNAK	159
X.2. X DATU-MATRIZEAREN BI ADIERAZPIDE GEOMETRIKO	159
X.3. BI ALDAGAI ETA BI OHARPEN	160
X.4. BI ALDAGAI ETA HIRU OHARPEN	167
X.5. ERREGRESIO BAKUNA ALDAGAI ESTATISTIKOEN BALIO-MULTZOEN ESPAZIOAN	173

XI. ERREGRESIO ORTOGONALA ETA KOBARIANTZ MATRIZEAREN AUTODIREKZIOAK

XI.1. ELKARREN MENPEKO ANALISIA	179
XI.2. ERREGRESIO-ZUZEN ORTOGONALA ETA HONDAR-BARIANTZA	179

XI.3. BATERAKO BARIANTZAREN DESKONPOSAKETA	
ERREGRESIO ORTOGONALEAN	188
XI.4. ERREGRESIO ORTOGONALAREN PROPIETATEAK	190

XII. DATU ANIZKOITZEN ANALISIA

XII.1. SARRERA	195
XII.2. ANALISI OROKORRA	196
XII.2.1. $\mathbb{R}^n$ espazioaren $\mathbb{R}^q$ azpiespazio baten bidez egindako doikuntza	197
XII.2.2. $\mathbb{R}^m$ espazioaren $\mathbb{R}^q$ azpiespazio baten bidez egindako doikuntza	200
XII.2.3. $\mathbb{R}^n$ eta $\mathbb{R}^m$ espazioen arteko erlazioa	201
XII.2.4 X datu-taularen berreraketa	203
XII.3. OSAGAI NAGUSIZKO ANALISIA	205
XII.3.1. $\mathbb{R}^n$ espazioan egindako analisia	206
XII.3.2. $\mathbb{R}^m$ espazioan egindako analisia	207
XII.3.3. Baterako aurkezpen grafikoak	211
XII.3.4. Adibidea	213
XII.4. KORRESPONDENTZI ANALISI FAKTORIALA	221
XII.4.1. Hodeiak, masak eta distantziak	224
XII.4.2. $\mathbb{R}^n$ espazioan egindako analisia	227
XII.4.3. $\mathbb{R}^m$ espazioan egindako analisia	228
XII.4.4. $\mathbb{R}^n$ eta $\mathbb{R}^m$ espazioen arteko erlazioa	229
XII.4.5. Maiztasun-taularen berreraketa	231
XII.4.6. Interpretaziorako laguntzak	232
XII.4.7. Korrespondentzia Anitzeko Anlisi Faktoriala	234
XII.5. ELEMENTU GEHIGARRIAK	237
XII.6. SAILKAPEN-METODOAK	238
XII.6.1. Goranzko Sailkapen Hierarkikoa	238
XII.6.1.1. Bariantzaren metodoa	239
XII.6.2. Adibidea	241

LEHENENGO ARGITARAPENENKO HITZAURREA

Duela 6 urte, 1983. urtean alegia, *Estatistikaren Hastapenak* izenaz lehen apunte-liburua argitaratu genuen Ekonomi eta Enpresa-Zientzien Fakultateko “Estatistikarako Sarrera” asignaturaren testuliburu gisa.

Une honetan, eskuartean daukazu *Estatistika Deskribatzailea. Koerlazioa, Erregresioa eta Datu Anizkoitzen Analisia* asignatura berberaren testuliburua izango da, baina aurrekoarekin parekatuz koalitatezko saltoa suposatzen du. Alde batetik, *Estatistikaren Hastapenak* hark biltzen zituen gaiak zuzenduak eta osatuak izan dira, eta bestaldetik hiru ikasgai berriren bitartez Datu Anizkoitzen Analisia ikasgaira iristen gara, non eskola frantsesaren Datu-Analisiaren metodologiaren gunea azaltzen dugun.

Esan beharra daukagu, aurkezpen geometrikoen eta erregresio ortogonalaren bitartez metodo faktorialera iristea guztiz originala dela “Estatistikarako Sarrera” asignatura batetan. Era honetan, datuen eta estatistikoaren zentzu bektoriala eta grafikoa ikasleari ikustaraziz bide tinkoa jorratzen dugu Datu Anizkoitzen Analisisaraino.

Guzti hau, egunez egun asignaturaren programan egindako talde-lana izan da eta meritua ez da gurea bakarrik, hein handi batetan hainbat urtetan lankide daukagun J.M. Piris-ena baizik. Aipamen berezia merezi du berak, bai eta edizio honetan hainbeste lagundu diguten UEUko Alfontso Mujikak eta Mari Karmen Menikak ere, hain zuzen. Mila esker guztiei.

Egileak
Sarriko, 1989.eko Urria

BIGARREN ARGITARAPENENKO HITZAURREA

Lau urte pasatu dira, liburu honen lehenengo argitarapena egin zenetik. Tarte horretan pauso asko eman dira irakaskuntzan, Euskararen Erabilera Normaltzeko Plangintzaren arabera.

Plangintza hori, Euskal Herriko Unibertsitateko Fakultate eta Eskoletako Euskara Batzordeek eta hauen bidez sortu zen Unibertsitateko Euskara Batzordeak bultzaturik, Gobernu-Juntak onartu zuen 1990.eko Irailaren 14ean.

Nahiz eta arazo asko gainditu behar izan, ikastegi batzutan beste batzutan baino gehiago bada ere, orokorki oinarritzko asignatura asko gaur egun Euskaraz irakasten direla esan beharra dago, horrela, Estatistikaren irakaskuntza ikastegi gehienetara zabaldu da eta honen ondoren Ekonometria asignatura, Ekonomi eta Enpresa-Zientzien Fakultatean.

Aspaldidanik UZEIko "Matematika Hiztegia" eta Euskal Gobernuak argitaratutako "Termino Estatistikoen Hiztegia" lagungarri izan ditugu eta une honetan UZEIko "Euskalterm" terminologi bankua ere badaukagu. Hemendik aurrera Estatistikaren irakaskuntza eta erabilera egokitzen dugun bitartean, Ekonometria, Denborazko Serien Teoria eta beste gai aurreratuetan ihardungo dugu, gai horietan lan egiten duten beste irakasle euskaldunekin batera.

Bukatzeko, bigarren argitarapen honetan gaiek berean segitzen badute ere, zuzendu eta egokitu egin ditugu horietariko batzu.

Egileak
Sarriko, 1993.eko Iraila

***I. EZAUGARRI ESTATISTIKO BAKUNAK,
BANAKETAK, TAULAK,
ADIERAZPIDE GRAFIKOAK***

*I.1. ESTATISTIKAREN HISTORIA LABURRA.
ESTATISTIKA ETA PROBABILITATEA*

I.2. OROKORTASUNAK

I.2.1. Populazioa eta lagina

I.2.2. Unitate estatistiko edo indibiduoak

I.3. EZAUGARRI ESTATISTIKOAK

I.3.1. Ezaugarri estatistiko bakunak

I.3.1.1. Ezaugarri kuantitatiboak

I.3.1.2. Ezaugarri kualitatiboak

I.4. MAIZTASUN-BANAKETA BAKUNAK

*I.4.1. Maiztasun-banaketak (absolutuak,
erlatiboak, eta metatuak)*

I.4.2. Taulak

*I.5. MAIZTASUN-BANAKETA BAKUNEN
ADIERAZPIDE GRAFIKOAK*

I.5.1. Barra-diagramak

I.5.2. Histogramak

I.5.3. Maiztasun-poligonoak

I.5.4. Diagrama linealak

I.5.5. Beste adierazpide grafikoak

I.1. ESTATISTIKAREN HISTORIA LABURRA. ESTATISTIKA ETA PROBABILITATEA

Estatistika hitza gutxienez hiru zentzu desberdinetan erabil daiteke.

Esan ohi da:

- “estatistika batzu” izen arrunta bezala erabiliz edo
- “..... txosten estatistiko bat.....” adjetibotzat hartuz edo.
- “.... Estatistika ikastea.... “ izen propioa bezala azkenik.

Hitz honen adierazpen orokor, zahar eta ez-espezializatua hauxe da:

MULTZO HANDIEN “Ikasketa” edo “erabilera”, MULTZO ZENTZUAN HARTURIK.

Honela, **hasieran** Estatistika artea eta teknika izan zen.

Mundu neolitikoa herriak eta salerosketak gero eta ugariagoak egin zirenez, estatistikaren beharra sortu zen.

Gizartearen hasiera historikotik praktikatu zen beraz. Israeldar, Erromatar, Txinatar eta Inken aparteko inperioarenak, adibide eder batzu besterik ez dira.

Baina “cinquecento” italiarrean (XVI. mendean) izena eta bere buruaren kontzientzia hartzen hasi zela dirudi, kontabilitatearekin nahiko lotuta hasi ere.

XVIII. mendetik aurrera argiago eta garbiago agertzen da, beharbada. Estatu-zientziaren alboan hasi zen, gehienbat, Alemaniak sortutako bultzada zela medio.

Eta honetan, oraindik Estatistika ez da zientzia, eta gertaera aleatorioetatik eta probabilitate-teoriatik urrun zegoen artea zen.

Baina geroztik “estatistikariei” problemak agertzen hasi zitzaizkien:

- estatistiketarik ondorio zientifikoak nola lortu?
- kolektibo aldakor edo zentsatzeko zail diren eraketarik informazio fidagarriak nola lortu?

Britaniar giroan XIX. mendearen amaieran eta XX.aren hasieran, oztopo hauetatik **irteten zen heinean**, Estatistika zientzia bilakatu zen.

Estatistika probabilitate-teoriarekin esentziazko harremanetan dagoen zientzia da orain.

Honela, hurrengo galderetan ikusiko dugun maiztasun kontzeptutik probabilitate kontzeptura pasatu zen.

Eskematikoki ikusiz:

MAIZTASUNAK	↔	PROBABILITATEAK
(gertapenen arloa)		(joerak, gertagarritasunak)

KOLEKTIBO	}	↔	{	KOLEKTIBO
KONKRETUAK				ABSTRAKTUAK

Fenomenoetatik bere kausetarainoko bidea egiten duen ezagupide honi, **INDUKZIOA** deritzo Estatistikaren alorrean eta, hementxe, **INFERENTZIA ESTATISTIKOA**.

Estatistika zientziaren oinarri trinkoan bermatutako estimazioak egiten dira gaur egun, beraien fidagarritasuna ere neurtzen delarik.

Hau dena **INFERENTZIA ESTATISTIKOARI** dagokio.

Apunte-liburu honetan ikusten diren gaiak Estatistika Deskribatzaileari dagozkio, hau da, oinarritzko Estatistika ikusiko dugu edo Estatistikaren hastapenak, halaber, Estatistika Deskribatzaile anizkoitza edo Datu-Analisia ikusiko dugu XII. gaian.

Dakusagun, bada, gaian sartu aurretik, ESTATISTIKA DESKRIBATZAILEAREN definizio bat.

ESTATISTIKA DESKRIBATZAILEA, datu-multzo HANDIAK aurkez ditzakeen teknika bat da, multzo hau eskurakoi bihurtuz, beraren egitura gardenduz eta beraren barne-harremanak neurtuz.

1.2. OROKORTASUNAK

1.2.1. Populazioa eta lagina

Ikerketa estatistiko bakoitzean objektu nagusi diren pertsona, ondasun, saioaren emaitza (adibidez: datuaren aurpegia) probintzia baten herri, hauteskudearen alderdi politiko, edota fabrikazio-unitateek osatzen duten multzoari POPULAZIO edo UNIBERTSO deritzogu.

Populazioa ondo definitzeko edozein elementu partikular populaziokoa den ala ez jakin ahal izatea, beharrezkoa zaigu.

Populazioak **bukaezinak** (adb: boltsa batetako erauzketak) ala **bukakorrak** (eskualde batetako enpresa kooperatiboak) izan daitezke.

LAGINA: Lagina azpipopulazio errepresentagarri bat besterik ez da; populazio batez konklusio fidagarriak atera nahi baditugu, ahal dugun azpipopulazio errepresentagarriena aukeratu beharko dugu lagintzat.

LAGIN errepresentagarri hau erauzi egiten duen prozesuari **zorizko laginketa** deritzo.

Lagina erauztean populazioaren elementu guztiei posibilitate berdina ematen badiegu, **zorizko lagin sinplea** deritzogu honela aukeratutako laginari.

Lagina erauzteko beste era eta jokabide asko daude, geruzatuz, mordoketaz,... baina hau **LAGINKETA-TEORIA**ri dagokio.

Lagina osatzen duten elementuen kopuruari laginaren **tamainua** deritzo.

Sinbolikoki:

Ω = Populazioa

$\Omega' =$ Lagina $\Omega' = \{\omega_1, \omega_2, \dots, \omega_N\} \subset \Omega$

N = Laginaren tamaina.

1.2.2. Unitate estatistiko edo indibiduoak

Populazioa edo lagina osatzen duen elementu bakoitzari **unitate estatistiko** edo indibiduo deritzogu.

Indibiduo izena, estatistika deskribatzailearen jatorri demografikoari zor diogu.

1.3. EZAUGARRI ESTATISTIKOAK

Ikerketa estatistiko batetan helburua ez dugu populazioa edo lagina eshaustiboki ikertzea, baizik eta beraren ezaugarri batetara edo batzutara mugatuko gara.

Ikertu nahi dugun ezaugarria bakarra bada, **ezaugarri estatistiko bakuna** deritzo.

Populazioaren bi edo ezaugarri gehiago batera ikertu nahi baditugu, **ezaugarri estatistiko anizkoitza** deritzo.

Ezaugarri estatistikoek (bakun ala anizkoitzek) populazio edo laginen multzoko indibiduo bakoitzean **balio indibidual** bat har dezakete.

Lagineko indibiduo-multzoari, ezaugarriaren **balio-multzo** bat dagokio.

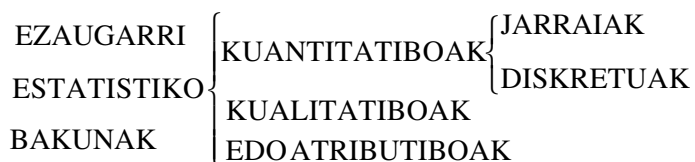
Sinbolikoki:

$\Omega = \{ \omega_1, \omega_2, \dots, \omega_N \} \longrightarrow \{ x_1, x_2, \dots, x_n \}$

$\Omega = \{ \omega_1, \omega_2, \dots, \omega_N \} \longrightarrow \{ (x_1 y_1), (x_2 y_2) \dots (x_n y_n) \}$

1.3.1. Ezaugarri estatistiko bakunak

Ezaugarri estatistiko bakunen sailkapena bere balio multzo desberdinen arabera:



1.3.1.1. Ezaugarri kuantitatiboak

Ezaugarriaren balio indibidualak populazioan edo laginean **zenbakiak** baldin badira, edota balio-multzoa zenbakizkoa bada, ezaugarri kuantitatiboa deritzo.

Ezaugarriari dagokion balio-multzoa, **multzo diskretua** denean, **ezaugarri diskretua** deritzo.¹

Balio biren artean edozein balio har dezakenean ezaugarri jarraia deritzo.

1.3.1.2. Ezaugarri kualitatiboak

Ezaugarriak har ditzakeen balio indibidualak kategoria edo atributuak direnean, edota **balio-multzoa** kategoria edo atributuek osatzen dutenean (diskretuak beraz), ezaugarri kualitatiboa deritzo.

Ondo klasifikatu ahal izateko, ezaugarri kualitatiboak hiru baldintza hauek bete behar ditu:

- **Ondo definitua**, hau da, kategoria edo atributu bakoitzak zer ulertarazten duen argiro azaldu behar du.

- **Esklusibotasuna**, hau da, indibiduo bat ezin liteke bi kategoriatan egon.

- **Eshaustibotasuna**, hau da, indibiduo guztiek nonbait klasifikatuak egon behar dute.

ADIBIDEAK: Eskualde bateko enpresa kooperatiboen lagin bat aztertzean:

- Langile-kopurua, ezaugarri kuantitatibo diskretua da;
- Produkzioa, ezaugarri kuantitatibo jarraia da;
- Lekutasuna, ezaugarri kualitatiboa da.

1. balio-multzoko elementuak, puntu isolatuak direnean, balio-multzo diskretu bat dugu.

I.4. MAIZTASUN-BANAKETA BAKUNAK

I.4.1. Maiztasun-banaketak (absolutuak, erlatiboak, eta metatuak)

Ezaugarri diskretuetan:

Suposa dezagun N indibiduo osatutako lagin bat, non ezaugarri bakun batek “ m ” balio edo kategoria desberdin hartzen baititu.

Hots:

$$\begin{aligned} \{ \omega_1, \omega_2, \dots, \omega_N \} &\longrightarrow \{ x_1, x_2, \dots, x_j, \dots, x_m \} \\ \{ \omega_1, \omega_2, \dots, \omega_N \} &\longrightarrow \{ A, B, \dots, J, \dots, M \} \end{aligned}$$

x_j **Balioaren MAIZTASUN ABSOLUTUA**, sinbolikoki n_j **ikurraz** adieraziko dugu, eta laginean x_j balioa hartzen duten indibiduen kopurua da.

Non:

$$\sum_{j=1}^m n_j = N$$

x_j **Balioaren MAIZTASUN ERLATIBOA**, sinbolikoki f_j **ikurraz** adieraziko dugu, non:

$$f_j = \frac{n_j}{N} \qquad \sum_{j=1}^m f_j = 1$$

hau da, laginean x_j balioa hartzen duten indibiduen kopurua erlatiboki kontutan hartuz lortzen den proportzioa da.

Maiztasun erlatiboak portzentaian ere adierazten dira, batzutan nahiko komenigarria baita horrela egitea.

Orduan:

$$P_j = \% f_j \cdot 100 \qquad \sum_{j=1}^m p_j = 100$$

Laginari dagokion **balio-multzoa ordenatu** ahal badugu (ezaugarria kuantitatiboa bada, beti; bestela ez beti), txikitik handira ordenatuko dugu.

Hots:

$$x_1 \leq x_2 \leq \dots \leq x_j \dots \leq x_m$$

Kasu honetan, x_j balioaren MAIZTASUN ABSOLUTU METATUA, sinbolikoki N_j , honela definituko dugu:

$$N_j = n_1 + n_2 + \dots + n_j \quad N_m = \sum_{j=1}^m n_j = N$$

hau da, laginean x_j balioa eta txikiagoak hartzen dituzten indibiduen kopurua da.

Eta: x_j **balioaren MAIZTASUN ERLATIBO METATUA**, sinbolikoki F_j , honela definituko dugu:

$$F_j = \frac{n_1 + n_2 + \dots + n_j}{N} = \frac{N_j}{N} = f_1 + \dots + f_j$$

$$F_m = \sum_{j=1}^m f_j = 1$$

hau da, x_j balioa eta txikiagoak hartzen dituzten indibiduen kopurua N indibiduo guztiekiko kontsideratutako proportzioa (batekotan).

Edozein maiztasun erlatibo bezala, portzentaiatan ere adierazten dira.

$$\text{Kasu honetan, } F_m = \sum_{j=1}^m p_j = 100$$

1.4.2. Taulak

Datu edo balioen ordenaketa eta elkarketari TABULAZIOA deritzo, eta, azkeneko datu-disposizioari, TAULA ESTATISTIKOA.

Lagin bati dagokion **maiztasun-banaketen taula** (ezaugarri bakun eta diskretua izanik) ondokoa da.

Lagina	Balio- -multzoa	MAIZTASUN-BANAKETA			
		M. absolutuak	M. absolutu metatuak	M. erlatiboak	M. erlatib metatuak
ω_1	x_1 A	p_1	$p_1 N$	$\frac{p_1}{N}$ f	$\frac{p_1}{N}$
ω_2	x_2 B	p_2	$p_2 N$	$\frac{p_2}{N}$ f	$\frac{p_2}{N}$
..		..		..	
ω_j	x_j J	p_j	$p_j N$	$\frac{p_j}{N}$ f	$\frac{p_j}{N}$
..		..		..	
ω_N	x_m M	p_m	$p_m N$	$\frac{p_m}{N}$ f	$\frac{p_m}{N}$

Ezaugarri jarraietan

Nahiz eta **teorikoki** ezaugarria etengabea izan eta fenomeno askok ezaugarri-mota honi erantzun, **praktikan** zehaztasun gehiago edo gutxiago duten tresnez neurtzen direnez gero, diskretu bukaezinak gertatzen zaizkigu.

Hots:

$$\{\omega_1, \omega_2, \dots, \omega_N\} \xrightarrow{x} I = (a, b) \subset \mathbb{R}$$

baina praktikan:

$$\{\omega_1, \omega_2, \dots, \omega_N\} \xrightarrow{x} \{x_1, x_2, \dots, x_N\} \subset (a, b) \subset \mathbb{R}$$

Lehen urratsean: Ezaugarriari dagokion **balio-multzoa mailatan zatituko dugu**, mailen muturrak ondo definituz.

$$I = (a_0, a_1) \cup (a_1, a_2) \cup \dots \cup (a_{n-1}, a_n)$$

non:

$$a \leq a_0 \leq a_1 \leq a_2 \leq \dots \leq a_{n-1} \leq a_n \leq b$$

Mailen kopurua, tamainu desberdinekoak hartu ala ez azterketa bakoitzean erabaki beharreko gauzak dira, baina beti ere mailaketak hiru baldintza hauek bete beharko ditu:

- Ondo definituak.
- Esklusibotasuna
- Eshaustibotasuna

Bigarren urratsean: Maila bakoitza balio batetara mugatuko dugu, **klase-ordezkari** deritzoguna.

Normalki ordezkaria, maila tartearen **erdiko** puntua da, baina goi ala beheko muturrean ere izan liteke.

Honela bada, balio-multzoko maila bakoitza, balio batetara mugatzen dugu (ordezkariaren baliora) eta laburpen honetaz informazio estatistikoa galdu arren, hasieran ezaugarri jarraia zena, ezaugarri diskretu bukakor bihurtzen dugu, eta dagokion maiztasun-banaketa kalkulatu ondoren, ondoan dagoen taula bezalako batetan adieraz liteke.

Lagina	Mailak edo klaseak	Klase-ordezkariak	M. absolutuak	M. absolutu metatuak	M. erlatiboak	M. erlatibo metatuak
ω_1	(a_0, a_1)	x_1	n_1	N_1	f_1	F_1
ω_2	(a_1, a_2)	x_2	n_2	N_2	f_2	F_2
..		..	..		..	
	(a_{j-1}, a_j)	x_j	n_j	N_j	f_j	F_j
..		..	..		..	
ω_N	(a_{m-1}, a_m)	x_m	n^m	N	f_m	1
$\sum_{j=1}^m n_j = N$					$\sum_{j=1}^n f_j = 1$	

ADIBIDEAK

a) X ezaugarri kualitatiboa da.

Eskualde batetako populazioaren banaketa produkzio-sektoreen arabera:

Sektoreak	Portzentaiak	Portzentaia metatuak
Nekazaritza	70	70
Industrigintza	25	95
Zerbitzuak	5	100

(kasu honetan, sektore-kategoria bakoitzari dagozkien maiztasun erlatiboak portzentaietan bakarrik jarri ditugu).

b) X ezaugarria kuantitatiboa da.

b-1) Diskretu-taldekatua

Enpresen banaketa, langile-kopuruaren arabera herri batetako enpresetatik lortutako lagin batetan.

Langileen kopurua	Enpresen kopurua	
	m. absolutuak	m. abs. metatuak
0-tik 9-ra	3.876	3.876
10-tik 49-ra	2.028	5.904
50-tik 99-ra	976	6.880
100-tik 499-ra	689	7.569
500-tik 999-ra	320	7.889
1.000-tik 4.999-ra	304	8.193
5.000-tik 9.999-ra	295	8.488
10.000-tik aurrera	12	8.500
N = 8.500		

(kasu honetan, laginean langile-kopuru ezaugarriari zegozkion balio desberdinak asko zirenez gero, taldekatu egin ditugu mailaka edo klaseka).

b-2) Ezaugarri jarraia

Soldaduzkara urte batetan doazen gazteen banaketa, beren altueraren arabera:

Klaseak (altuera m.)	Klase- -ordezkariak	Portzentaiak	Portzentaia Metatuak
1'5 b. gutxiago		1	1
1'5-tik 1'55-ra	1'525	7	8
1'55 1'60	1'575	12	20
1'60 1'65	1'625	18	38
1'65 1'70	1'675	30	68
1'70 1'75	1'725	20	88
1'75 1'80	1'775	10	98
1'80 b. gehiago		2	100

(Kasu honetan, mailaketan, maila guztiek tamainu berdina dute, aurrenekoa eta azkenekoa kenduz).

***1.5. MAIZTASUN-BANAKETA BAKUNEN ADIERAZPIDE
GRAFIKOAK***

Nahiz eta maiztasun-banaketen taulak informazio guztia hartu (balioak mailetan laburtuak ez badira izan behintzat), grafikoki adierazteak asko laguntzen du ikusarazten eta ulertzen.

Adierazpide grafiko asko erabili izan arren, batzu bakarrik ikusiko ditugu.

1.5.1. Barra-diagramak

Oso baliagarriak zaizkigu ezaugarriak **diskretuak** ditugunean (kuantitatiboak zein kualitatiboak).

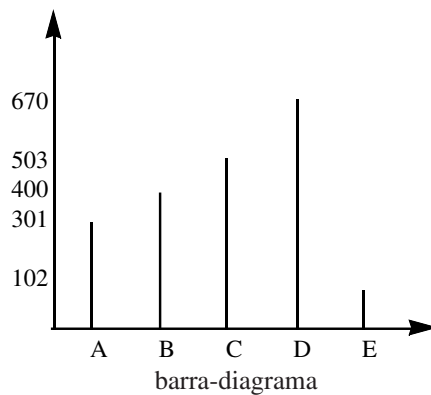
Abzisa-ardatzean balio edo kategoria desberdinak (ahal bada ordenaturik), ezarriko ditugu, eta ordenatu-ardatzean berriz laginari dagozkion maiztasunak (absolutuak, erlatiboak ala metatuak).

ADIBIDEAK

a) Ezaugarri kualitatibo diskretua.

Enpresen banaketa bere kokapenaren arabera:

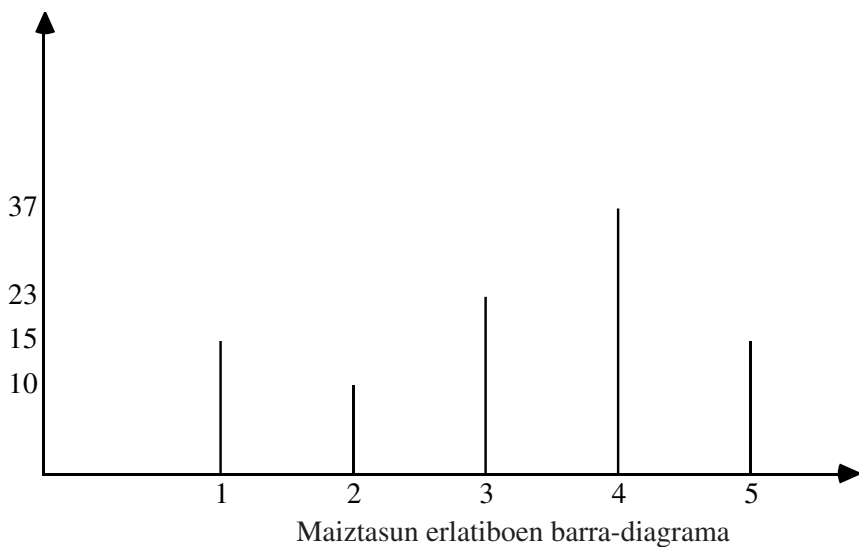
kokapena	Enpresen kopurua
A arloa	301
B arloa	400
C arloa	503
D arloa	670
E arloa	102
	1.976

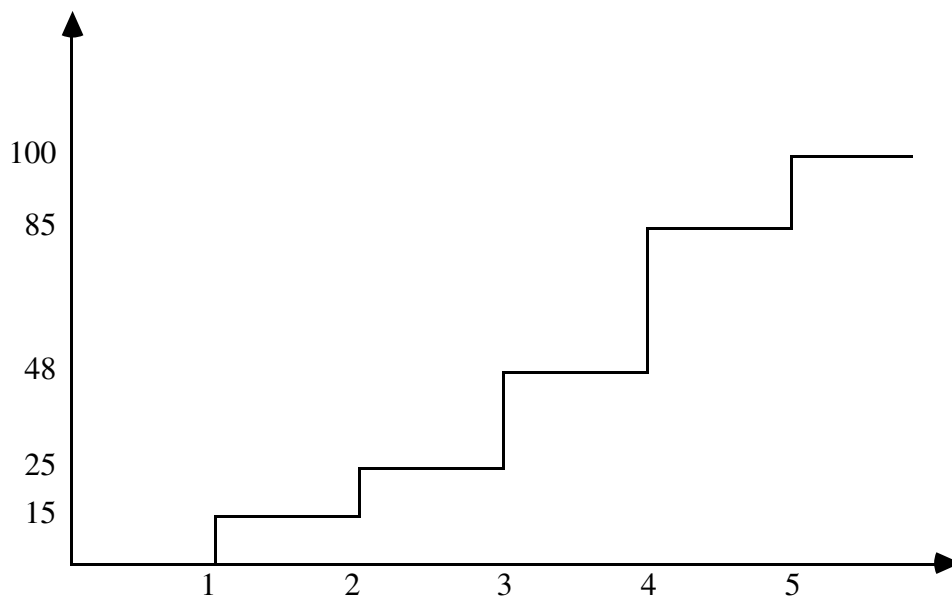


b) Ezaugarri kuantitatibo diskretua.

Loteen banaketa, beren baitan dituzten pieza akastunen kopuruaren arabera:

Txarren kopurua:	Maiztasun-erlatiboa: (portzentatetan)	Maiztasun erlatiboa: Portzentaia metatuak
1	15	15
2	10	25
3	23	48
4	37	85
5	15	100





Maiztasun metatuen diagrama

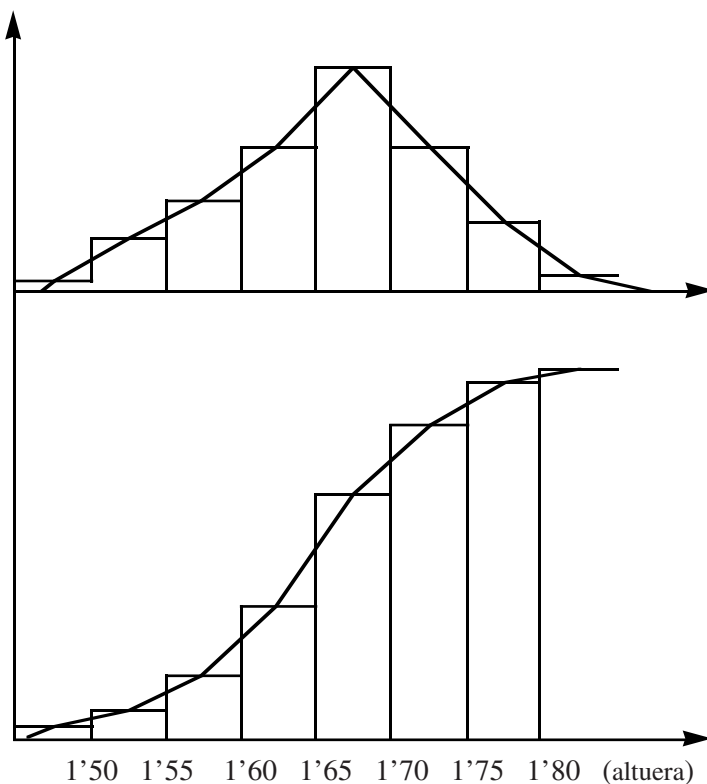
1.5.2. Histogramak

Oso baliagarriak zaizkigu ezaugarriak jarraiak direnean.

Ezaugarriaren balio-multzoa mailetan zatitu ondoren eta maila edo klase bakoitza oinarritzat harturik, dagokion maiztasunaren (nahiz absolutua, erlatiboa edo metatua) **araberako azalera duen laukizuzen bat eraikitzen da.**

Klase guztiak zabalera berdinekoak baldin badira, klase bakoitzari dagokion laukizuzenaren altuera, bere maiztasun absolutu edo erlatiboa da. Honela, soilik, laukizuzenen oinarriak berdinak izatean, azalerak altueren menpe daude.

Ondoan b-2) adibideari dagozkion histogramak eta maiztasun-poligonoak ditugu (Ikus b-2) 13. orrialdean).



Histogramak eta maiztasun-poligonoak

Klaseen zabalerak, ordea, desberdinak baldin badira, laukizuzenen altuerak, maiztasun absolutuak edo erlatiboak zabaleraz zatituz, kalkulatueto ditugu.

h_i , n_i eta c_i , i . klasearen altuera, maiztasuna eta zabalera, hurrenez hurren, izanik:

$$h_i = \frac{n_i}{c_i}, \text{ orduan } S_i \text{ azalera: } S_i = c_i \frac{n_i}{c_i} = n_i$$

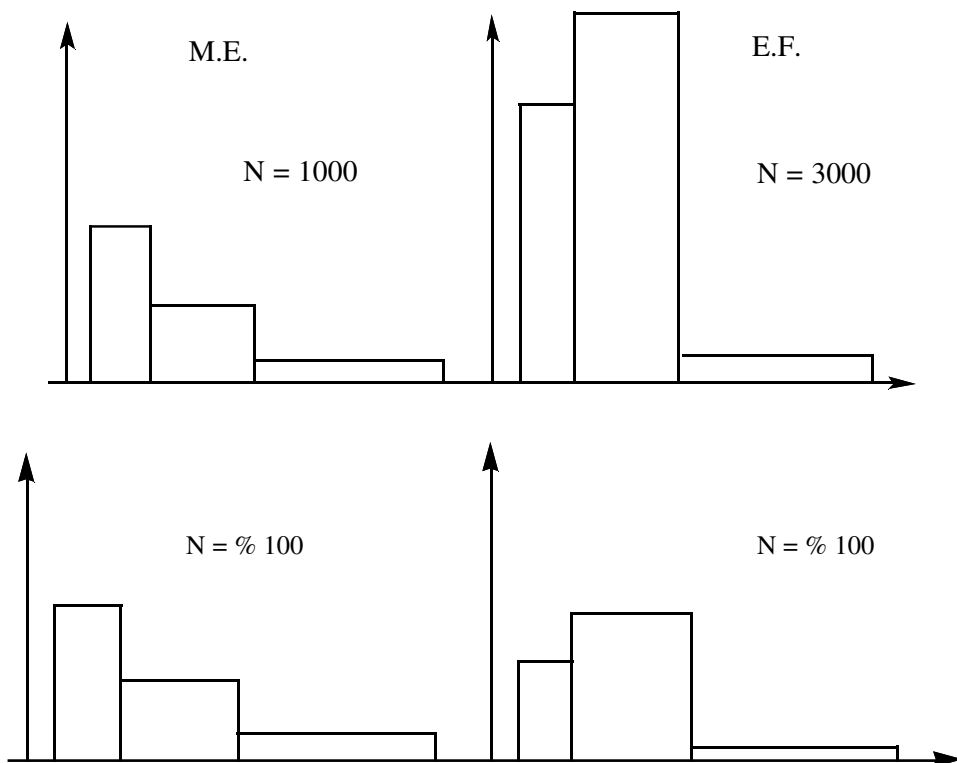
h_i altuera d_i bezala ere izendatzen da eta maiztasun-dentsitate bezala ezagutzen.

Adibidea:

Bi goi-mailako ikastetxetan, ikasleen adinak honela banatzen dira:

<u>Urteak (mailak)</u>	<u>M.E. Ikastetxea</u>	<u>E.F. Ikastetxea</u>
17 - 19	400	750
20 - 25	400	2.000
26 - 37	200	250

Adibidearen histogramak egin ditugu a) kasuan maiztasun absolutuak kontsideratuz eta b) kasuan, berriz, maiztasun erlatiboak portzentaietan.



Dakusagunez, maiztasun erlatiboak (ehunekotan edo ez) adieraztean totalaren ikusketa, N-rena hain zuzen, galtzen da, baina ordea bi banaketak konparagarriago bihurtzen dira.

1.5.3. Maiztasun-poligonoak

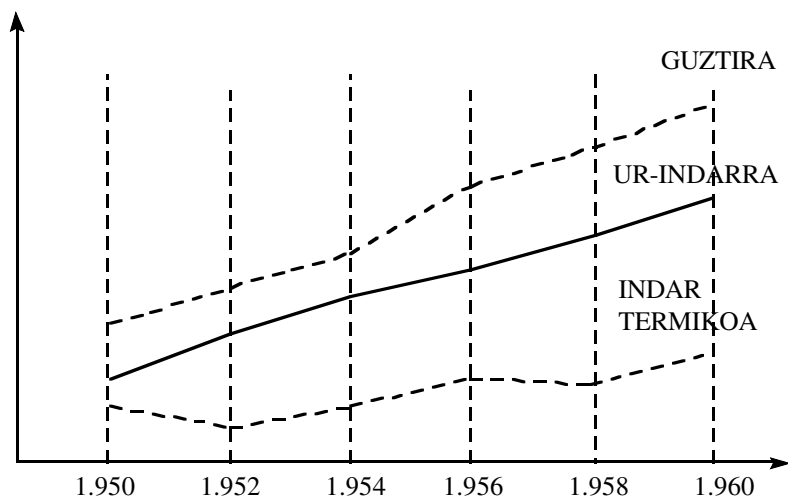
Maiztasun-poligonoa laukizuzenen goiko oinarriko erdiko puntuak marra batez lotuz eraikitzen da.

Honela egin dugu 16. orrialdeko histograman.

1.5.4. Diagrama linealak

Oso erabilgarriak izan ohi dira, balioen aldaketetan denborak izan lezakeen eragina aztertzeko.

Adibidez: Energi produkzioa 1.950-tik, 1.960 urterarte:



1.5.5. Beste adierazpide grafikoak

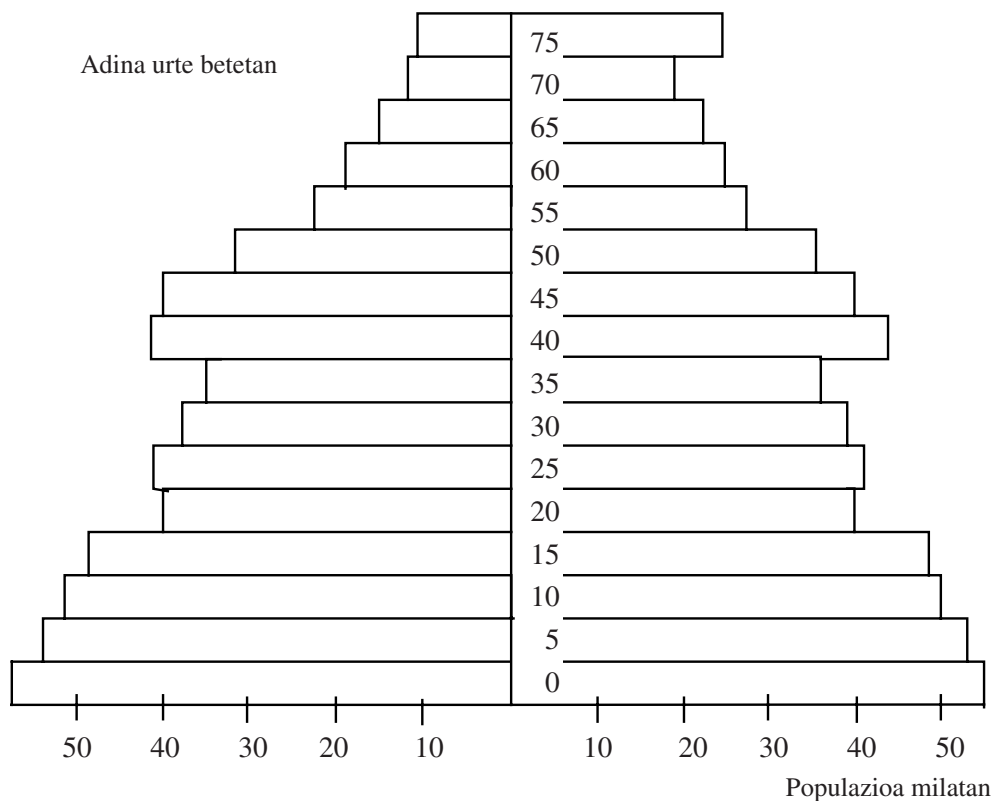
Beste batzuren artean, ikus ditzagun, azkenez, bi hauek:

- POPULAZIO-PIRAMIDEAK.

Histograma bikoitzak dira, eta honela eraikitzen dira: ordenatu-ardatzean adin-taldeak adierazten dira eta abzisa-ardatzean, ezkererantz eta eskuinerantz gizonezko eta emakumezkoen maiztasun erlatiboak populazio totalarekiko adierazten dira, dagozkien laukizuzenak eraikiz.

Maiztasun erlatiboak populazio totalarekiko kontsideratzean, gizonezkoen eta emakumezkoen laukizuzenak konpara daitezke adin-talde bakoitzerako.

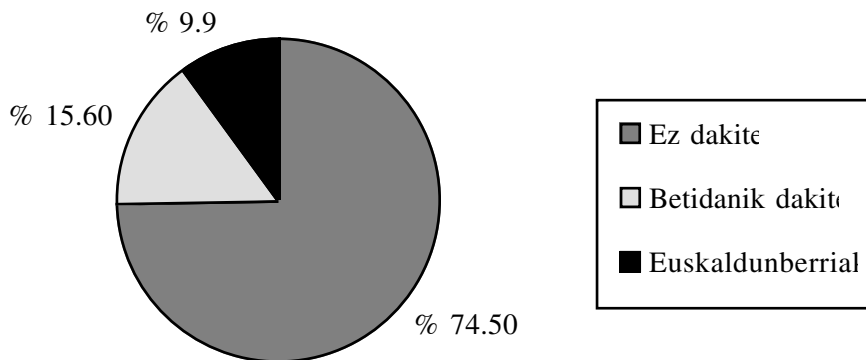
Halaber, maiztasun erlatiboak kontsideratzean, eskala berdinean eraikiz gero, tamainuaren aldetik oso desberdinak diren populazioak konpara daitezke.



Bizkaiko Populazio-Piramidea 1.975. urtean

- SEKTORE GRAFIKOAK

Grafiko hauek honela eraikitzen dira: zirkunferentziako 360 graduetatik, aldagaien balio ala kategoria bakoitzaren maiztasunari proportzionalki zati bat dagokio.



Euskararen ezaguera Sarrikoko Fakultatean 1979. urtean.

II. EZAUGARRI BAKUNEN BALIO TIPIKOAK

II.1. MOMENTUAK

II.1.1. Momentu arruntak edo jatorriarekikoak

*II.1.2. Momentu zentralak edo
batezbestekoarekikoak*

II.2. MAIZTASUN-BANAKETA BAKUNEN BALIO TIPIKO EDO ESTATISTIKOAK

II.2.1. Zentru-joeraren balio tipikoak

II.2.1.1. Batezbesteko aritmetikoa

II.2.1.2. Batezbesteko geometrikoa

II.2.1.3. Batezbesteko harmonikoa

II.2.1.4. Moda

II.2.1.5. Mediana

II.2.2. Sakabanatzearen balio tipikoak

II.2.2.1. Bariantza

II.2.2.2. Batezbesteko desbidazioa

II.2.2.3. Ibiltartea

II.2.3. Itxuraren balio tipikoak

II.2.3.1. Asimetri koefizientea

*II.2.3.1. Kurtosi-koefizientea edo
zapaltasun/zorroztasun-koefizientea*

II.3. KONTZENTRAZIO-NEURKETAK

II.3.1. LORENZ-en kurba

II.3.2. GINI-ren indizea

II.1. MOMENTUAK

Datuek berekin duten informazioa, balio tipiko edo estatistikoen bidez sintetizatzea komeni da.

Noski, laburketa guztiei informazio-galtze bat dagokie; errakuntzarako motibo hau, bada, kontuan izan beharko dugu estatistiko bakoitza aztertzerakoan.

Baina, estatistiko edo balio tipikoak aurkeztu baino lehen, momentuak ikusi behar ditugu, balio tipikoak momentuak bait dira edo momentuen bidez definitzen.

Momentuak bi motatakoak izan daitezke:

- arruntak edo jatorriarekikoak.
- zentralak edo batezbestekoarekikoak.

II.1.1. Momentu arruntak edo jatorriarekikoak

Maiztasun-banaketa bakun baten h-ordenako momentu arrunta edo jatorriarekikoa, sinbolikoki a_h ikurraz adieraziko da eta honela definituko: non:

N = Laginaren edo Populazioaren indibiduen kopurua

x_i = Ezaugarriaren i. balioa.

m = Ezaugarriak hartzen dituen balio desberdinak; $m \leq N$ izango da.

$$a_h = \frac{\sum_{i=1}^N (x_i)^h}{N} = \frac{\sum_{i=1}^m (x_i)^h n_i}{N}$$

Momentu arrunten bi berezitasun ikusiko ditugu:

- 1) 0-ordenako momentuak 1 balioa hartzen du:

Hots:

$$a_0 = \frac{\sum_{i=1}^m (x_i)^0 n_i}{N} = 1$$

2) 1- ordenako edo lehenengo ordenako momentua batezbesteko aritmetikoa da:

Hots:

$$a_1 = \frac{\sum_{i=1}^m x_i n_i}{N} = \bar{x}$$

eta normalki $\bar{x}$ ikurraz adierazten da.

Ikusten ari garen momentu arruntek nahiz ondoren ikusiko ditugunek, ez dute berekiko esangurarik, bidebatez balio tipiko direnean izan ezik, adibidez, batezbesteko aritmetikoa, bariantza, e.a.

Halere, Estatistikan eragile matematiko bezala sartzen dira, teoriaren garapenean oso erabilgarriak baitira.

II.1.2. Momentu zentralak edo batezbestekoarekikoak

Maiztasun-banaketa bakun baten h-ordenako momentu zentrala edo batezbestekoarekikoa, sinbolikoki m_h ikurraz adieraziko da eta honela definituko:

$$m_h = \frac{\sum_{i=1}^N (x_i - \bar{x})^h}{N} = \frac{\sum_{i=1}^m (x_i - \bar{x})^h n_i}{N}$$

Eragiketak eginez, ikus dezagun nola kalkulatzen diren momentu zentralak momentu arrunten bidez:

$$\begin{aligned}
m_h &= \frac{1}{N} \sum^m (x_i - \bar{x})^h n_i = \\
&= \frac{1}{N} \sum^m \left(\binom{h}{0} x_i^h - \binom{h}{1} x_i^{h-1} \bar{x} + \binom{h}{2} x_i^{h-2} \bar{x}^2 - \dots + \right. \\
&\quad \left. + (-1)^{h-1} \binom{h}{h-1} x_i \bar{x}^{h-1} + (-1)^h \binom{h}{h} \bar{x}^h \right) n_i = \\
&= \binom{h}{0} a_h - \binom{h}{1} a_{h-1} \bar{x} + \binom{h}{2} a_{h-2} \bar{x}^2 - \dots + (-1)^{h-1} h \bar{x}^{h-1} + (-1)^h \bar{x}^h
\end{aligned}$$

Hortik:

$$\begin{aligned}
m_2 &= a_2 - \bar{x}^2 \\
m_3 &= a_3 - 3a_2 \bar{x} + 2\bar{x}^3 \\
m_4 &= a_4 - 4a_3 \bar{x} + 6a_2 \bar{x}^2 - 3\bar{x}^4
\end{aligned}$$

Momentu hauen artean bigarren ordenakoa (m_2), BARIANTZA delakoa da garrantzitsua.

$$m_2 \bar{x}_{(\text{pon})} = \frac{\sum (x_i - \bar{x})^2 n_i}{N} = a_2 - \bar{x}^2$$

Eta azkenez, momentuak edozein punturekiko izan daitezke:

$$a_h, x_0 = \frac{\sum^N (x_i - x_0)^h}{N} = \frac{\sum^m (x_i - x_0)^h n_i}{N}$$

$x_0 = 0$ baldin bada, momentu arruntak izango dira.

eta $x_0 = \bar{x}$ bada, zentralak.

II.2. MAIZTASUN-BANAKETA BAKUNEN BALIO TIPIKO EDO ESTATISTIKOAK

Organigrama batetan azalduko ditugu banaketa bakunetan erabili ahal diren estatistiko edo balio tipikoak; gero erabilienak banaka ikusiz joango gara.

ZENTRU - JOERAREN edo posizioaren balio tipikoak	- neurrizkoak:	- Batezbesteko aritmetikoa - Batezbesteko geometrikoa - Batezbesteko harmonikoa
	- lekuzkoak:	- Mediana - Koantilak
	- maiztasunezkoak:	Moda
SAKABANATZEAREN balio tipikoak:	- bariantza eta desbidazio standarda	
	- batezbesteko desbidazioa	
	- ibiltartea eta koartilarteko ibiltartea	
	- aldakuntz koefizientea	
ITXURAREN balio tipikoak	- asimetri koefizientea	
	- kurtosi koefizientea	

II.2.1. Zentru-joeraren edo Posizioaren balio tipikoak

Lagin batetan ezaugarri bati dagozkion balioak ordenatzen baditugu, zentruan kokatzen diren balioak erabiltzen dira datu guztiak errepresentatzeko. Posizio-neurriak dira.

Balio batetan datu-multzo baten ideia ematea argiro ikusten dugu, sarritan gehiegizko sinplifikazioa izanik.

Horregatik, zentruko estatistikoak beste estatistiko batzurekin (sakabanatzearenak,..) osatzen dira, aztertzen ari garen populazio edo laginaren ideia erreala goa eduki nahi baldin badugu.

II.2.1.1. Batezbesteko aritmetikoa

Lagin batetan $x_1, x_2, \dots, x_i, \dots, x_m$, zenbakizko balio-multzo bati dagozkion maiztasun absolutuak hurrenez hurren $n_1, n_2, \dots, n_i, \dots, n_m$ baldin badira, beraren batezbesteko aritmetikoa, $\bar{x}$, honako hau da:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^m x_i n_i \quad \text{non} \quad \sum_{i=1}^m n_i = N$$

Maiztasun erlatiboak, $f_1, f_2, \dots, f_m$, kontutan hartuz

$$\bar{x} = \sum_{i=1}^m x_i f_i$$

Ezaugarriak hartzen dituen balio guztiak desberdinak balira, maiztasun absolutuek 1 balioa hartuko lukete, hau da, $n_1 = n_2 = \dots = n_N = 1$ eta orduan:

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

Maiztasun erlatiboak masak bezala kontsideratzen baditugu, batezbesteko aritmetikoak **grabitrate-zentrua** finkatuko du. Hau da, barra baten gainean eta $x_1, x_2, \dots, x_m$ puntuetan maiztasunak bezalako pisuak kokatzen baditugu sistemak oreka lor dezan, esekitze-puntuak $\bar{x}$ puntuan izan beharko du.

PROPIETATEAK:

1.- Edozein maiztasun-banaketatan, bere maiztasunaz biderkatuz, batezbestekoarekiko desbideratzeen batukaria zero egiten da.

Hots:

$$\sum_{i=1}^m (x_i - \bar{x}) n_i = 0$$

Frogapena:

$$\sum_{i=1}^m (x_i - \bar{x}) n_i = \sum_{i=1}^m x_i n_i - \bar{x} \sum_{i=1}^m n_i = N \bar{x} - N \bar{x} = 0$$

2.- Ezaugarri edo aldagai baten balio guztiak h parametroaz biderkatzen baditugu, batezbesteko aritmetikoa ere h-z biderkatuta edukiko dugu.

Hots:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^m x_i n_i \quad \text{izanez}$$

x_i balio bakoitza h x_i balioaz ordezkatzeko badugu:

$$\frac{1}{N} \sum_{i=1}^m h x_i n_i = h \frac{1}{N} \sum_{i=1}^m x_i n_i = h \bar{x}$$

3.- Aldagai bi edo gehiagoren batura den aldagai baten batezbesteko aritmetikoa, batezbesteko aritmetikoen batura da.

Hau da: $t = x + y + z$ baldin bada.

$\bar{t} = \bar{x} + \bar{y} + \bar{z}$ izango da.

Suposa dezagun oposizio-lehiaketa bat hiru azterketaz osatzen dela eta i. lehiakideak x_i, y_i, z_i puntuak ateratzen dituela, puntuazio totala $t_i = x_i + y_i + z_i$ izanez; N lehiakideren puntuazioen batezbesteko aritmetikoa ondokoa izango da:

$$\begin{aligned} \bar{t} &= \frac{\sum_{i=1}^N t_i}{N} = \frac{\sum_{i=1}^N (x_i + y_i + z_i)}{N} = \\ &= \frac{\sum_{i=1}^N x_i}{N} + \frac{\sum_{i=1}^N y_i}{N} + \frac{\sum_{i=1}^N z_i}{N} = \bar{x} + \bar{y} + \bar{z} \end{aligned}$$

BATEZBESTEKO ARITMETIKO PONDERATUA

Batzutan batezbesteko aritmetikoa ateratzean, aldagaien garrantzi erlatiboa desberdina dela ohartzen gara.

Suposa dezagun hiru sagar mota saltzen direla 60, 80 eta 90 pezeta/kilo prezioetan hurrenez hurren.

Batezbesteko prezioa honela atera daiteke:

$$\bar{x} = \frac{60 + 80 + 90}{3} = 76'6, \text{ hau da, batezbesteko aritmetikoa sinplea eginik.}$$

Baina mota bakoitzarentzat saltzen diren kiloak, hurrenez hurren, hauek baldin badira: 125, 72, 3

Batezbesteko ponderatua, zera izango da:

$$\bar{x}_{(\text{pon})} = \frac{60 \cdot 125 + 80 \cdot 72 + 90 \cdot 3}{125 + 72 + 3} = 67'6$$

Ikusten dugunez, batezbesteko aritmetiko ponderatua egitean, aldagaiaren balio bakoitza zenbaki batez biderkatzen dugu, balio-multzo barnean daukan garrantziaren arabera.

Zenbaki hauek pisuak edo ponderazioak dira eta sinbolikoki w_i deituko ditugu.

$$\text{Orduan } \bar{x}_{(\text{pon})} = \frac{\sum x_i w_i}{\sum w_i} \quad \text{edota} \quad \bar{x}_{(\text{pon})} = \sum x_i w_i \quad \sum w_i = 1 \quad \text{denean}$$

Oharra: kontutan hartu behar da, bien formulak baliokideak badira ere ponderazioak maiztasunak ez direla. Hala ere, maiztasunak ponderazioak dira aldagaiaren balio desberdinek duten garrantzia maiztasunen arabera izango baita.

Beste adibide bat:

Ondoko taulan Hego Euskal Herriko lau herrialdeen kasuan hileroko gastuak/familia datuak dauzkagu.

Hileroko gastuak / familia, 1.968. urtean

	<u>Familiak</u>	<u>Gastuak (mila pezetatan)</u>
Araba	24	4,4
Gipuzkoa	53	5,3
Nafarroa	57	5,1
Bizkaia	84	5,1

$$\text{Batezbesteko aritmetikoa: } \bar{x} = \frac{19,9}{4} = 4,975$$

Baina, familien kopurua kontutan hartuz:

$$\bar{x}_{(\text{pon})} = \frac{4,4 \cdot 24 + 5,3 \cdot 53 + 5,1 \cdot 57 + 5,1 \cdot 84}{24 + 53 + 57 + 84} = 5,07$$

Noski, batezbesteko ponderatuen bidez errealitatearen azterketari gehiago hurbiltzen gatzaizkio, baina normalki datu guztiak ezagutzen ez ditugunez, batezbesteko aritmetiko sinplea egiten dugu.

II.2.1.2. Batezbesteko geometrikoa:

Ondoko berdintasunaren bidez adierazten dugu:

$$G = \sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \cdot \dots \cdot x_m^{n_m}} = \sqrt[n]{\prod_{i=1}^m x_i^{n_i}}$$

Arrazoi edo ehunekoen batezbesteko bat ateratzeko, batezbesteko aritmetikoa baino aproposagoa da eta ondoko propietatea betetzen du:

$$\text{Log } G = \frac{1}{N} \log \prod_{i=1}^m x_i^{n_i} = \frac{1}{N} \sum_{i=1}^m (\log x_i) n_i$$

Hots:

Batezbesteko geometrikoaren logaritmoa, aldagaiaren balioen logaritmoen batezbesteko aritmetikoa da.

II.2.1.3. Batezbesteko harmonikoa

Ondoko berdintasunaren bidez adierazten dugu

$$H = \frac{N}{\sum_{i=1}^m \frac{n_i}{x_i}} \quad \text{non} \quad N = \sum_{i=1}^m n_i$$

Ohar daiteke batezbesteko harmonikoa zera dela, x_i balioen alderantzizko balioen batezbesteko aritmetikoaren alderantzizko balioa.

Batezbesteko abiaduren batezbesteko bat ateratzeko, batezbesteko harmonikoa da egokiena.

Ikusitako batezbestekoak eta beste batzuk: batezbesteko koadratikoa, batezbesteko kubikoa, e.a. sar daitezke k ordenako batezbestekoaren kontzeptuan.

$$\text{Hots:} \quad M_{(k)} = \sqrt[k]{\sum x_i^k f_i}$$

II.2.1.4. Moda

Maiztasun-banaketa batetan gehien errepikatzen den aldagai edo ezaugarriaren balioari, hau da, maiztasunik handiena duenari, moda eta balio modala deritzo.

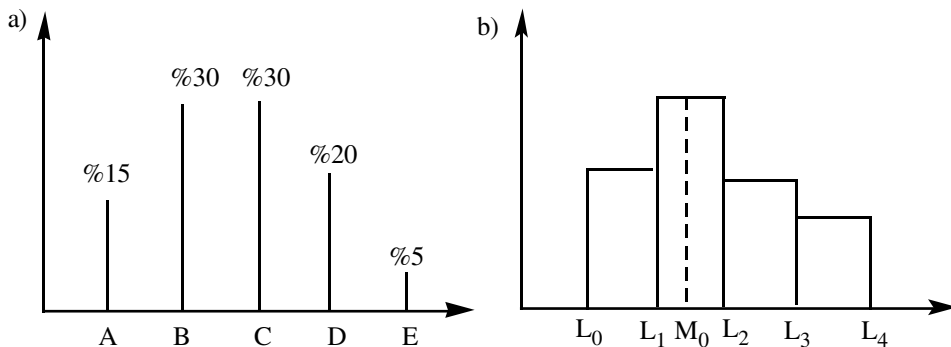
$$x_1, x_2, \dots, x_i, \dots, x_m \quad \text{edo} \quad A, B, \dots, I, \dots, M$$

balio edo kategorien multzo bati dagozkien maiztasun erlatiboak hurrenez hurren $f_1, f_2, \dots, f_m$ baldin badira, maiztasun erlatibo gehien duen x_i balioari edo I kategoriari “moda” M_0 deritzo.

Hots: $M_0 = x_i$ edo $M_0 = I$, n_i eta f_i maiztasun handienak izanik.

Banaketa, modagabea izan liteke $f_1 = f_2 = \dots = f_m$, edo moda bat baino gehiago ere eduki dezake, eta honen arabera, “bimodala”, “hirumodala” ... orokorki moda-anitza izango da.

ADIBIDEAK



Kasu honetan banaketa bimodala dugu, B eta C arloak modak direlarik.

L_1 , L_2 klasea da kasu honetan moda, klase honen ordezkaria, M_0 , har daiteke modatzat.

Banaketa jarraia edo klasetan taldekatua bada, dakigunez, maiztasun handiena, M_0 , duen klasean kokatuko da moda, eta klase-ordezkaria modatzat hartzen ez badugu, gutxi gorabeherako kalkuluak egin daitezke.

Klase edo tarte modala L_{i-1} , L_i izanez, balio modala estimatzeko bi erizpide desberdinetan oinarritutako formulak proposatuko ditugu.

1) Altueren alderaketaren erizpidea:

$$M_0 = L_{i-1} + \frac{d_{i-1}}{d_{i-1} + d_{i+1}} \cdot c_i$$

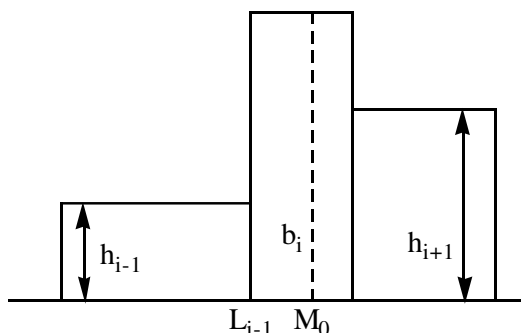
non: h_{i-1} = aurreko klasearen altuera

h_{i+1} = hurrengo klasearen altuera

Ondoko klaseen altueretikiko alderantzizko banaketa proportzionala egitean, aurreko formula lortzen dugu.

Hots:

$$\frac{b_i}{h_{i+1}} = \frac{c_i - b_i}{h_{i-1}}$$



edo aurrekariak eta atzekariak batzean:

$$\frac{b_i}{h_{i+1}} = \frac{c_i}{h_{i+1} + h_{i-1}}$$

eta, $M_0 = L_{i-1} + b_i$ izatean, proposatu dugun formula frogatuta gelditzen da.

Klaseak zabalera berdinekoak balira, hau da, $c_i = c$ c zabaleraz biderkatzen eta zatitzen badugu eta $n_i = c \cdot h_i$ dela kontutan hartuz, ondoko hau izango da formula:

$$M_0 = L_{i-1} + \frac{h_{i+1} \cdot c}{(h_{i+1} + h_{i-1})} c = L_{i-1} + \frac{n_{i+1}}{n_{i+1} + n_{i-1}} c$$

II.2.1.5. Mediana

Mediana banaketaren balioak bi zati berdinetan zatitzen dituen lekuzko estatistikorik garrantzitsuen da.

Balio-multzoa ordenatuz gero, maiztasun absolutuen histograma area berdineko bi zatitan ebakitzen duen balioari mediana deritzo, eta hori zehazki betetzen duenik ez badago, aurrean dagoena hartzen dugu batzutan medianatzat.

Ondoko baldintza betetzen duen balio bat da:

$$M_e = x_i / N_{i-1} < \frac{N}{2} \wedge N_i > \frac{N}{2}$$

Hala ere, banaketa diskretua eta indibiduen kopurua bakoitia baldin bada, aurreko baldintza bete egingo da, baina ez beste edozein kasutan.

Banaketa diskretua eta N bikoiti baten aurrean bagaude.

$$M_e = \frac{x_i + x_{i+1}}{2} / N_i = \frac{N}{2}$$

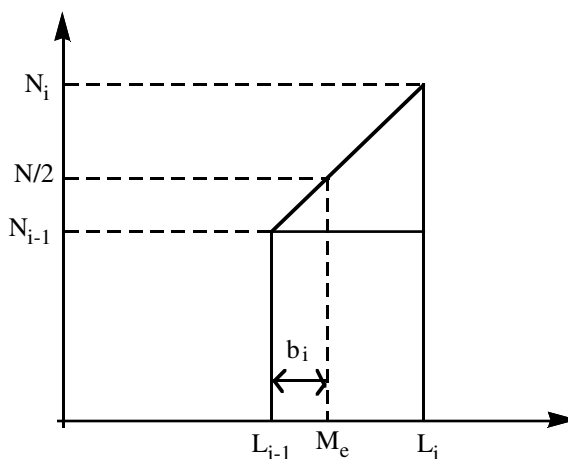
$$\text{Hots: } M_e = \frac{\text{erdiko balioak}}{2}$$

Banaketa taldekatuetan eta jarraietan: $N/2$ balioa duen tartea finkatuz gero eta tartean n_i balioak linealki gehitzen direla suposatuz kalkulatu dugu M_e .

Orduan: (L_{i-1}, L_i) M_e duen tartea izanez:

$$M_e = L_{i-1} + b_i$$

Irudian dauden bi triangeluak antzekoak izanez, zera idatz daiteke:



$$\frac{N_i - N_{i-1}}{c_i} = \frac{\frac{N}{2} - N_{i-1}}{b_i}$$

$$\text{eta: } b_i = \frac{N/2 - N_{i-1}}{N_i - N_{i-1}} c_i$$

Oharra: maiztasun erlatiboak erabiliz, era berdinean kalkula daiteke M_e ; kasu horretan,

$$\frac{N}{2} = \frac{1}{2} \cdot$$

KOANTILAK

Koantilak beste lekuzko neurriak dira. Mota desberdinetakoak dira: koartilak, dezilak eta zentilak.

Koartilak, dezilak eta zentilak medianaren antzera definitzen dira, baina bi parte hartu ordeztu lau (koartilak) hamar (dezilak) edo ehun (zentilak) parte hartuz.

Koartilak q_1, q_2, q_3 izango dira eta (q_1, q_3) tartean banaketaren erdia daukagu eta justu erdiko balioak, hau da, bazterreko balioak kenduz.

Dezilak $d_1 \dots d_9$ izango dira eta zentilak, $c_1 \dots c_{99}$.

Argiro ikusten da:

$$q_2 = d_5 = c_{50} = M_e \text{ dela.}$$

OHARRAK:

Erdiko baliorik garrantzitsuenak, batezbesteko aritmetikoa, moda eta mediana dira eta bereziki batezbesteko aritmetikoa, guztiak aldagaia edo ezaugarria neurtua dagoen unitateetan neurtuak direlarik.

Ondoko erlazio hau betetzen da:

$$M_0 \simeq 3 M_e - 2 \bar{x}$$

Eta gutxi gorabehera hiruretako edozein, besteen bidez kalkula daiteke.

II.2.2. Sakabanatzearen balio tipikoak

Datu-multzoaren ideia orokorra ematen ziguten zentzurako joeraren balio tipikoek. Ideia hau osatzeko, zentzurako joeraren balioen inguruan datu-multzoa

guztiz bilduta edo oso sakabanatuta dagoen adierazten diguten estatistiko batzu ere behar ditugu.

Batezbesteko aritmetikoaren inguruan datuek duten sakabanatze-neurri bezala, estatistiko garrantzitsuenak hauek dira: Bariantza, Desbidazio Standarda eta Batezbesteko Desbidazioa.

II.2.2.1. Bariantza

Batezbestekoarekiko, sakabanatze-neurririk garrantzitsuenak da bariantza eta bera erabiliz lortzen ditugu desbidazio standarda eta aldakuntz koefizientea.

$\{x_1, x_2, \dots, x_i, \dots, x_m\}$ zenbakizko balio-multzo bati dagozkien maiztasun absolutuak hurrenez hurren $n_1, n_2, \dots, n_i, \dots, n_m$ baldin badira, bariantza, S_x^2 , deritzona hau da:

$$S_x^2 = \frac{1}{N} \sum_{i=1}^m (x_i - \bar{x})^2 n_i \quad \text{edota} \quad S_x^2 = \sum_{i=1}^m (x_i - \bar{x})^2 f_i$$

Bariantza neurri-unitate karratuetan daukagula, erraz ikusten dugu formularen.

Orduan, sakabanatze-neurri bezala bariantzaren erro karratua hartzea egokiagoa da, **Desbidazio Tipikoa edo Standarda**, hain zuzen.

$$S_x = \sqrt{\frac{1}{N} \sum_{i=1}^m (x_i - \bar{x})^2 n_i} = \sqrt{\sum_{i=1}^m (x_i - \bar{x})^2 f_i} = \sqrt{a_2 - \bar{x}^2}$$

eta ezaugarria neurtuta dagoen unitateetan, sakabanatze-neurketa daukagu.

Askotan, estatistika inferentzialean estimatzaile bezala dituzten propietateengatik, S_x^{*2} Quasi-bariantza eta S_x^* Quasi-desbidazio standarda erabiltzen dira.

$$S_x^{*2} = \frac{\sum_{i=1}^m (x_i - \bar{x})^2 n_i}{N - 1} \quad \text{eta} \quad S_x^* = \sqrt{\frac{\sum_{i=1}^m (x_i - \bar{x})^2}{N - 1}}$$

Konputagailuko programa-paketeak S_x^{*2} , S_x^* ematen dizkigute Bariantza eta Desbidazio Standarda deitzen badituzte ere.

Desbidazio tipikoak batezbestekoarekiko suposatzen duen batekoa (edo ehunekoa) **aldakuntz koefizientea** deitzen da.

$$g_0 \text{ ikurraz adieraziko dugu sinbolikoki, non: } g_0 = \frac{S_x}{\bar{x}}$$

$\bar{x}$ batezbestekoa zerora hurbiltzen bada, g_0 horrek ez du zentzu askorik (infiniturantz jotzen baitu kasu honetan).

Bariantzak garrantzizko propietate bat betetzen du, bigarren ordenako momentu zentralik txikiena dela, hain zuzen.

Hots:

$$\begin{aligned} a_2, x_0 &= \frac{1}{N} \sum_{i=1}^m (x_i - \bar{x}_0)^2 n_i = \\ &= \frac{1}{N} \sum_{i=1}^m \left((x_i - \bar{x}) + (\bar{x} - x_0) \right)^2 n_i = \\ &= \frac{1}{N} \sum_{i=1}^m \left((x_i - \bar{x})^2 + 2 (x_i - \bar{x}) (\bar{x} - x_0) + (\bar{x} - x_0)^2 \right) n_i = \\ &= \frac{1}{N} \sum_{i=1}^m (x_i - \bar{x})^2 n_i + 2 (\bar{x} - x_0) \sum_{i=1}^m (x_i - \bar{x}) \frac{n_i}{N} + \\ &+ \frac{1}{N} (\bar{x} - x_0)^2 \sum_{i=1}^m n_i = S_x^2 + (\bar{x} - x_0)^2 \end{aligned}$$

Eta argiro dagoenez, $x_0 = \bar{x}$ baldin bada, txikiena izango da, edozein beste kasutan bigarren batugaia positiboa baita.

II.2.2.2. Batezbesteko desbidazioa

Balio absolutuak harturik (konpentsaziorik izan ez dadin), batezbesteko batekiko balioek duten desbidazioen batezbesteko aritmetikoari, “batezbesteko desbidazioa” deritzo.

$$\text{Hau da: } D = \frac{1}{N} \sum_{i=1}^m |x_i - \text{batezbestekoa}| n_i$$

a) Aukeratutako batezbestekoa, batezbesteko aritmetikoa bada, batezbesteko desbidazioa hauxe izango da:

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

Neurri honek indibiduen homogenotasuna adierazten digu, hau da, desbidazioak txikiak badira, batezbesteko desbidazioa txikia izango da eta, alderantziz, handiak badira, batezbesteko desbidazioa handia izango da.

b) Aukeratutako batezbestekoa mediana bada, batezbesteko desbidazioa hauxe izango da:

$$D_{M_e} = \frac{1}{N} \sum_{i=1}^m |x_i - M_e| n_i$$

Batezbestekoa mediana denean, batezbesteko desbidaziorik txikiena daukagu.

II.2.2.3. Ibiltartea

Ezaugarri zenbakizko batek hartzen dituen balio masimo eta minimoaren arteko diferentziari ibiltarte deritzogu.

Hots:

$$x_0 \leq x_1 \leq \dots \leq x_m \text{ izanik}$$

$$R = x_m - x_0$$

Koartil arteko ibiltarteak ere kontsidera daitezke, adibidez, $q_3 - q_1$ hirugarren eta lehenengo artekoena hain zuzen ere, edota, dezil arteko ibiltarteak e.a.

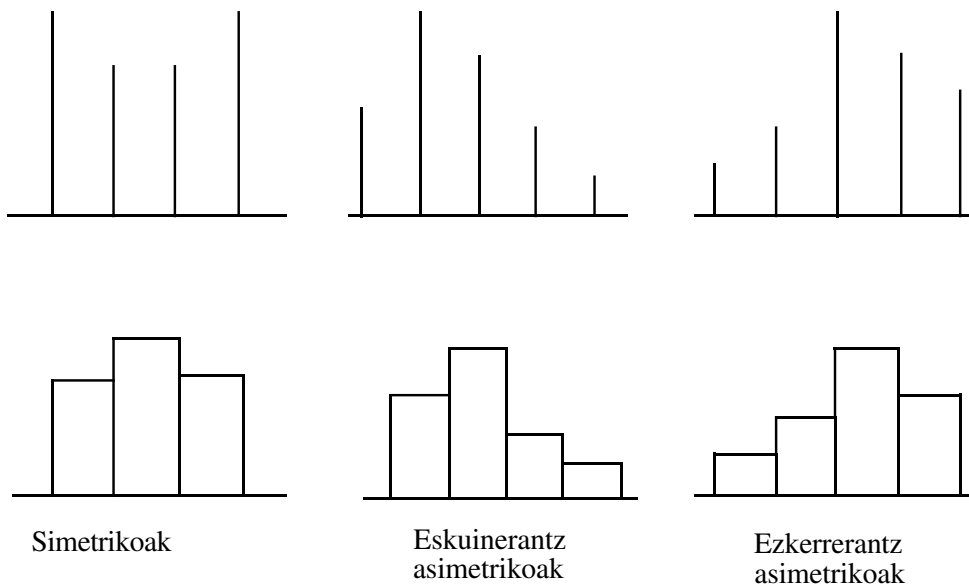
II.2.3. Itxuraren balio tipikoak

Aztertzen ari garen laginaren ezaugarria eta berari dagokion datu-multzoa, zentzurako joeraren eta sakabanatze-balioen bidez nahiko argi geratu bada ere, gehiago osa daiteke itxuraren balio tipikoak aztertzean; hauen artean asimetri koefizientea eta kurtosi-koefizientea ikusiko ditugu.

II.2.3.1. Asimetri koefizientea

Maiztasun-banaketaren grafikoa simetrikoa denean, banaketa simetrikoa dela esango dugu, hau da, barra-diagrama edo histograma egitean simetrikoa bada, banaketaren elementuak batezbestekoarekiko bi aldeetan berdin kokatzen direla esan nahi du.

Aurrekoa betetzen ez bada, banaketa asimetrikoa izango da.



Banaketa bat moda batekoa eta simetrikoa baldin bada, batezbesteko aritmetikoa, moda eta mediana balio berean daude.

$$\text{Hots: } \bar{x} = M_0 = M_e$$

Asimetri koefizientea g_1 ikurraz adieraziko dugu sinbolikoki, non:

$$g_1 = \frac{m_3}{S_x^3}$$

$$m_3 = \sum_{i=1}^m (x_i - \bar{x})^3 \frac{n_i}{N} \text{ izanik, maiztasunen balioak } \bar{x} \text{ balioarekiko}$$

simetrikoki kokatuta badaude, $m_3 = 0$ izango da; baina maiztasunen balioak eskuinaldean **handiagoak** badira, hau da, $x_i > \bar{x}$ denean, orduan $m_3 > 0$ izango da, eta, alderantziz, ezkerrealdean handiagoak badira, hau da, $x_i < \bar{x}$ denean, orduan $m_3 < 0$.

Ikusten dugu, bada, m_3 asimetriaren balio adierazle bezala har dezakegula, baina neurri-unitateak ber hiru izanez zuzenkiago.

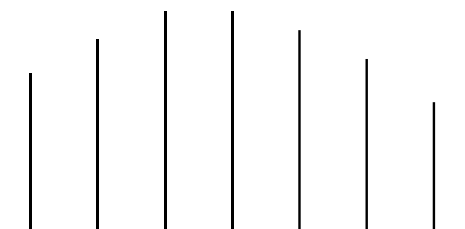
$$g_1 = \frac{m_3}{S^3}$$

erabiltzen da: $g_1 > 0$ bada, maiztasun-banaketa eskuinerantz asimetrikoa dela esaten da; $g_1 < 0$ bada, alderantziz, ezkererantz.

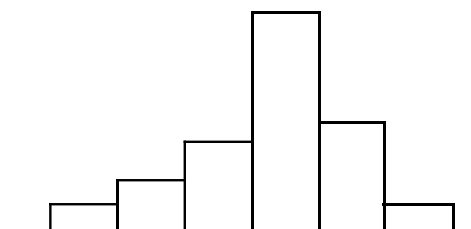
II.2.3.2. Kurtosi-koefizientea edo zapaltasun/zorroztasun-koefizientea

Simetriarekin batera, itxuraren beste balio tipiko bat ikusiko dugu, kurtosi-koefizientea hain zuzen.

Maiztasun-banaketaren grafikoa kontutan hartuz, konkretuki, maiztasun-poligonoa jarraiaren kasuan edo barrazko diagramaren goi-puntuak lotuko zituzkeen lerroa diskretuaren kasuan kasu hauek dauzkagu:



zapala



zorrotza

Batezbesteko aritmetikoa eta bariantza berdina daukaten bi banaketa aztertuz, itxuraren aldetik bat oso zapala da eta oso zorrotza bestea.

Banaketaren maiztasunak batezbesteko aritmetikoaren inguruan handiak baldin badira, momentu zentratu bikoitiek, (adibidez, laugarren ordenakoa), balio handiagoak hartuko dituzte.

Orduan, m_4 zapaltasunaren adierazle bezala har daiteke, baina neurri-unitateak ber lau izanez, zuzenkiago g_2 ikurraz adierazten den kurtosi-koefizientea, zapaltasunaren adierazle bezala erabiliko dugu.

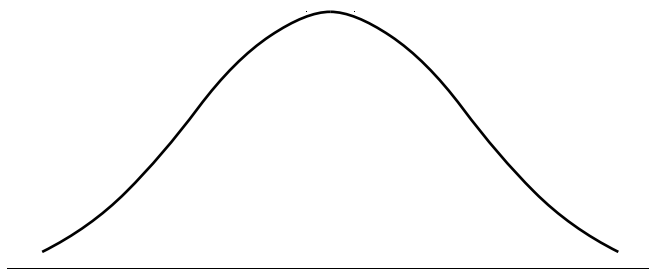
$$g_2 = \frac{m_4}{S_x^4} = \frac{\sum_{i=1}^m (x_i - \bar{x})^4 n_i}{N S_x^4}$$

“Normal” deitutako banaketaren kurtosi-koefizientea 3 da eta balio hau kurtosia aztertzean konparazio-puntu bezala hartuko dugu.

g_2 , 3 baino handiagoa baldin bada, banaketa zorrotza izango da eta 3 baino txikiagoa baldin bada, zapala izango da.

g_2 , -3 kurtosi-koefizientetzat hartzen badugu, kurtosi positiboa edo negatiboa kontsidera daiteke.

Oharra: Aipatu dugun banaketa “Normala” 3. kurtsoan ikasiko da, inferentzia estatistikoan, oinarrizko probabilitate-eredua baita.



banaketa Normala : $g_2 = 3$

II.3. KONTZENTRAZIO-NEURKETAK

Batezbesteko aritmetikoa egitean zenbakitzailean ateratzen zaigun kopuruak, abiapuntu estatistiko batetatik ikusita, batzutan ez du zentzu argirik eta interesgarririk. Adibidez, pertsona batzuren altuera-banaketa aztertzean, zenbakitzaile hori altueren batuketa izango da; baina, beste kasu batzutan, bereziki ezaugarri sozioekonomikoetan, aztergarria izaten da adibidez, alokairu-banaketa aztertzean, alokairu guztien kopurua edo “alokairu-masa” da.

Kontzentrazio-neurketek zehazki “masa” baten uniformetasuna neurtzea dute helburutzat.

Normalki errenta eta alokairu-banaketetan erabili ohi dira, baina beste edozein aldagairen banaketatan erabil daitezke.

Suposa dezagun alokairu-banaketa batetan langile guztiek alokairu berdina jasotzen dutela; orduan, banatzearen uniformetasuna osoa litzateke. Baina, ordez, alokairu-masa guztia langile batek jasoko balu, **uniformetasunik eza** ere osoa izango litzateke edo **kontzentrazioa masimoa** dela esango genuke kasu honetan.

Banaketak uniformetasun absoluturantz jotzen duenean, maiztasun-banaketaren batezbesteko aritmetikoa guztiz adierazgarria da eta, alderantziz gertatuko da, kontzentrazioa masimoa denean.

II.3.1. LORENZ-en kurba

Grafikoki banaketen kontzentrazioa ikusarazten digun neurrian, interesgarria zaigu.

Lorenz-en kurba eraikitzeko, ondoko taulan ikusten ditugun zutabeak kalkulatu beharko ditugu.

x_i	n_i	$x_i n_i$	N_i	M_i	P_i	Q_i
x_1	n_1	$x_1 n_1$	N_1	M_1	P_1	Q_1
x_2	n_2	$x_2 n_2$	N_2	M_2	P_2	Q_2
.	.	.	.	.	.	.
.	.	.	.	.	.	.
x_i	n_i	$x_i n_i$	N_i	M_i	P_i	Q_i
.	.	.	.	.	.	.
.	.	.	.	.	.	.
x_k	n_k	$x_k n_k$	N	M	100	100
$\sum_i n_i = N$		$\sum_i x_i n_i = M$				

azken zutabeak honela kalkulatzeko dira:

N_i : (n_i) zutabea metatuz

hots: $N_i = n_1 + n_2 + \dots + n_i$

M_i : ($x_i n_i$) zutabea metatuz

hots: $M_i = x_1 n_1 + x_2 n_2 + \dots + x_i n_i$

$P_i = \frac{N_i}{N} \cdot 100$ maiztasunen (langileen esate baterak) portzentaia metatua

$Q_i = \frac{M_i}{M} \cdot 100$ masaren portzentaia metatua

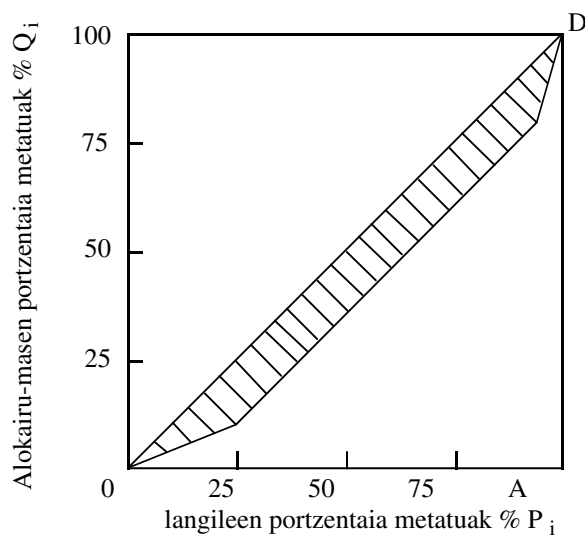
$\{(P_i, Q_i)\}_{i \in I}$ bikote guztiak sistema kartesiar batetan grafikoki adieraz, eta (P_1, Q_1) , $(0,0)$ - jatorriarekin eta beste bikote guztiak elkarren artean zuzenen bitartez lotuz LORENZ-en kurba lortzen dugu.

Adibidez: Suposa dezagun enpresa batetako langileen banaketa bere asteroko alokairuaren arabera (mila pezetatan) ondoko taularen lehenengo bi zutabeetan emana dela:

Alokairuak: x_i	Langile- -kopurua: n_i	Alokairu- -masa: $x_i n_i$	Metaketak N_i M_i		Portzentaia metatuak: P_i Q_i	
8	10	80	10	80	20	10
15	20	300	30	380	60	47,5
20	15	300	45	680	90	85
24	5	120	50	800	100	100

$$N = 50 \quad M = \sum_{i=1}^k x_i n_i = 800$$

Adibideari dagokion LORENZ-en kurba ondoko hau izango da.



Eskalak, bi ardatzetan berdinak direnez gero, LORENZ-en kurba, karratu baten barruan aurkitzen da beti. Jatorritik irteten den OD bere diagonal erabiliko dugu uniformetasunaren erreferentzia moduan.

Zeren eta (P_i Q_i) multzoko puntu guztiak, OD-n aurkitzen badira orduan uniformetasuna osoa baitugu.

Hots: langileen arteko % 10-ek, alokairu-masaren % 10-a jasotzen dute, % 30-ek, alokairu-masaren % 30-a, e.a.

Zenbat eta tarte handiagoa egon kurba eta diagonalaren artean, orduan eta **banaketa kontzentratuagoa** izango da, (edota zenbat eta kurba diagonaletik hurbilago egon, orduan eta **banaketa uniformeagoa**).

Abzisa ardatzean P_i kokatzen badugu eta ordenatu ardatzean Q_i , LORENZ-en kurba, karratuaren beheko triangeluan edukiko dugu (alokairuak txikienetatik handienetarantz ordenatu ditugunez gero, adibidez ezin dute langileen % 25-ak alokairu-masaren % 25-a inoiz lortu).

II.3.2. GINI-ren indizea

LORENZ-en kurba banaketen kontzentrazioaren argitzaile bada ere, zenbakizko indize batetaz kontzentrazioa neurtzea, komenigarria zaigu; esate baterako, erabilgarria izango zaigu banaketak konbaratzerakoan.

Helburu honi erantzuten dio **Gini-ren indizeak** edo **kontzentrazio-indizeak**.

Marraztutako azalera diagonal eta kurbaren artean dagoena bada, honela definituko dugu I_G , GINI-ren indizea:

$$I_G = \frac{\text{Marraztutako azalera}}{\triangle \text{ OAD azalera}}$$

$$\text{non } 0 \leq I_G \leq 1$$

Kontzentrazioa masimoa bada, orduan $\triangle \text{ OAD azalera} = \text{marraztutako azalera}$, eta $I_G = 1$.

Ordez, kontzentrazioa minimoa bada edo eta banaketa uniforme, orduan marraztutako azalera = 0 izango da eta $I_G = 0$.

Kalkulua: I_G Gini-ren indizea kalkulatzeko, azalera horiek zuzenki neurtuz kalkula litezke; baina beti era honetan erraza izaten ez denez gero, **gutxi gorabehera** ondoko formula erabiliz kalkulatu da I_G :

$$I_G = \frac{\sum_{i=1}^{k-1} (P_i - Q_i)}{\sum_{i=1}^{k-1} P_i}$$

kontura gaitezen, batukariak $k-1$ batugai bakarrik dituela, $p_k - q_k = 0$ baita.

Gutxi gorabeherako formula honetaz ere egiazta daiteke $0 \leq I_G \leq 1$ dela.

Adibidez: Lehenengo taularen datuekin jarraitzen badugu, aurreko formula erabiliz, I_G GINI-ren kontzentrazio-indizea ondoko hau izango da:

P_i	Q_i	$P_i - Q_i$
20	10	10
60	47'5	12'5
90	85	5
170		27'5

$$I_G = \frac{\sum_{i=1}^{k-1} (P_i - Q_i)}{\sum_{i=1}^{k-1} P_i} = \frac{27'5}{170} = 0'16$$

LORENZ-en kurba nahiz GINI-ren indizea aldagai sozioekonomikoetarako erabilgarriak dira. Adibide bezala hauek aipatuko ditugu: errenta pertsonalaren banaketa, nekazal-propietatearen banaketa, sektore batetako enpresa-produkzioaren banaketa eta abar.

III. ZENBAKI INDIZEAK

III.1. INDIZE SINPLEAK

III.2. PONDERAZIO GABEKO INDIZE KONPLEXUAK

III.2.1. Batezbesteko aritmetikoa sinplearen metodoa

III.2.2. Batezbesteko agregatu sinplearen metodoa

III.3. INDIZE KONPLEXU PONDERATUAK

III.3.1. Balioen, prezioen eta kopuruen indizeak

III.4. INDIZE KONPLEXUEN ERAIKETAN SORTZEN DIREN ZENBAIT ERAGOZPEN

III.4.1. Aldagaien hautapena

III.4.2. Somatutako leku eta denboraren hautapena

III.4.3. Talde eta azpitaldeen hautapena

III.4.4. Oinarri-denboraren hautapena

III.4.5. Formula eta ponderazioen hautapena

III.4.6. Indizearen esangura eta zabaldura

III.5. ERAGOZPEN BEREZI BATZU

III.5.1. Oinarri-aldaketa indize sinpleetan

III.5.2. Berriztapen eta loturak indize konplexuetan

III.6. ZENBAKI-INDIZEEN APLIKAPENAK

III.6.1. Kontsumo-prezioen indizeak

III.1. INDIZE SINPLEAK

Aldagai baten gorabeherakadak aztertzeko, **indize sinpleak** erabiltzen dira, eta aldagai batzuren gorabeherakadak “batera” aztertzeko **indize konplexuak**.

Hemen denboran zeharreko gorabeherakadak aztertuko dira. Erreferentzia gisa hartzen dugun denborari, **oinarria** deritzogu.

Erreferentziaren balioarekiko aldagaiaren balio bakoitzaren portzentaiak baino ez dira indize sinpleak.

Portzentaia hauen bidez, neurtzeko unitatea desagertzen da eta aldagai baten gorabeherakadak autonomoki azter daitezke. Honela indizeak aldagai-segiden gorabeherakadak errazten ditu (originalki unitate desberdinetakoak).

Izan bedi $x_0, x_1 \dots x_t$ aldagaiaren balio-segida denboral bat; hots, 0, 1, t denboretan neurtutako balioak.

$t = 0$ oinarritzat harturik eta x_0 balio erreferentziala, indize sinpleak ondoko taulan agertzen diren bezala kalkulatu dira:

Denbora: t	Balio-segida: x_t	Indize Sinpleak
0	x_0	$I_0 = 100$
1	x_1	$I_1 = \frac{x_1}{x_0} \cdot 100$
2	x_2	$I_2 = \frac{x_2}{x_0} \cdot 100$
.	.	.
.	.	.
.	.	.
t	x_t	$I_t = \frac{x_t}{x_0} \cdot 100$

x_0 balio erreferentzial gisa “urte normal” bati dagokiona hartzea, oso garrantzitsua da.

Adibidez: Instintiboki uzta-bilketa bat ona edo txarra dela esatean, guretzat uzta-bilketa “normalak” direnekin konparaturik egiten dugu.

III.2. PONDERAZIO GABEKO INDIZE KONPLEXUAK

n balioen aldakuntzak batera aztertu nahi ditugunean, indize konplexuak eraiki behar ditugu. Horiek n balio-segidak, balio-segida bakar batetara laburtzen dituzte, non balioak oinarritzat hartzen den urte batetara erreferiturik dauden eta “konplexu” guztiaren mugimendua ikusarazten diguten.

Goazen, bada, ponderazio gabeko indize konplexuak kalkulatzeko bi metodo ikustera.

III.2.1. Batezbesteko aritmetikoa sinplearen metodoa

Indize konplexu hau, denbora bakoitzari dagozkion **indize sinpleen** batezbesteko aritmetikoa da.

Hots: t denboran $x_1, x_2, \dots, x_n$ aldagaiei dagozkien $x_{1t}, x_{2t}, \dots, x_{it}, \dots, x_{nt}$ n balio-segidak baldin badira, non $t = 0, 1, 2, \dots, m$ den $I_{1t}, I_{2t}, \dots, I_{nt}$ indize sinpleak kalkulatu ondoren, indize konplexua formula honetaz kalkulatzen da:

$$S_t = \frac{\sum_{i=1}^n I_{it}}{n} = \frac{\sum_{i=1}^n \frac{x_{it}}{x_{i0}}}{n} 100$$

Momentu batetan, indize sinpleen elkarrekiko garrantzia ez hartzeagatik, ponderazio gabeko indizea da, orduan, **zenbat eta aldagai gehiago izan parte hartzen dutenak, hobe gertatzen da metodo hau.**

Estatistika-institutuan, urtez urteko aldagai-kopuru handi batez Estatu espainolerako metodo honetaz edo metodo ponderatuen bidez kalkulatzen diren indizeak, ez dira hain desberdinak.

III.2.2. Batezbesteko agregatu sinplearen metodoa

Metodo honen bidez, une bakoitzerako aldagai guztien balioen batura egiten da eta lortzen den segidan indize sinpleak bakarrik kalkulatzen dira.

Formula honetaz kalkulatzen dira:

$$B_t = \frac{\sum_{i=1}^n x_{it}}{\sum_{i=1}^n x_{i0}} 100$$

Metodo hau, **ezin erabil daiteke, aldagaiak unitate desberdinetan neurtuak direnean**, ez baitu zentzu askorik unitate desberdinetan neurtuak izan diren balioak batzea.

Ponderazio gabeko indize hau aurkez daiteke indize sinpleen indize ponderatu bezala ere, ponderazioak oinarri denborako balioak direlarik.
(Ikus batezbesteko aritmetiko ponderatua 27. orrialdean)

$$B_t = \frac{\sum_{i=1}^n I_{it} x_{i0}}{\sum_{i=1}^n x_{i0}}$$

Ondoko taulan ponderazio gabeko indize konplexu hauen formulak dauzkagu hedaturik.

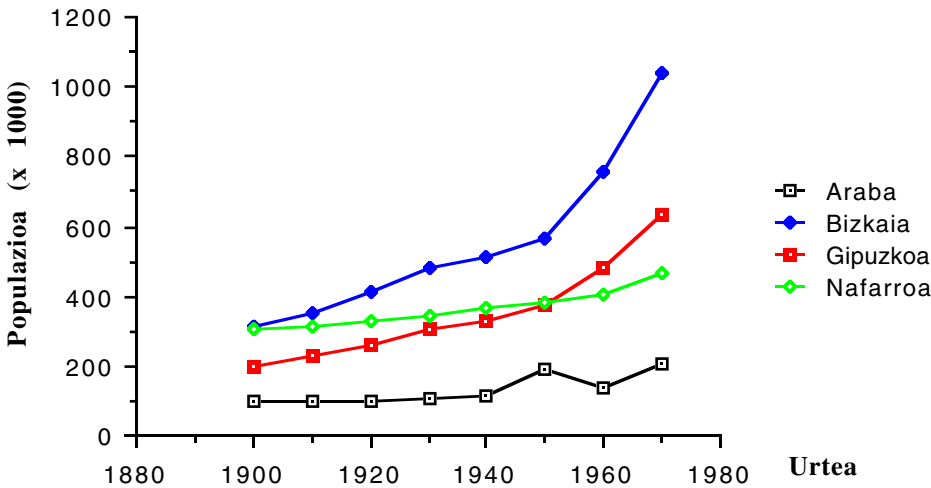
$\begin{matrix} x \\ t \end{matrix}$	$x_1, x_2, \dots x_i, \dots x_n$	I_i (Indize sinpleak)	S_t	$\sum_i x_{it} B_t$
0	$x_{10} x_{20} \dots x_{i0} \dots x_{n0}$	$I_{10} I_{20} \dots I_{n0}$	$S_0 = \frac{\sum_i I_{i0}}{n}$	$\sum_i x_{i0} B_0 = \frac{\sum_i x_{i0}}{\sum_i x_{i0}} 100$
1	$x_{11} x_{21} \dots x_{i1} \dots x_{n1}$	$I_{11} I_{21} \dots I_{n1}$	$S_1 = \frac{\sum_i I_{i1}}{n}$	$\sum_i x_{i1} B_1 = \frac{\sum_i x_{i1}}{\sum_i x_{i1}} 100$
.	.	.	.	.
t	$x_{1t} x_{2t} \dots x_{it} \dots x_{nt}$	$I_{1t} I_{2t} \dots I_{nt}$	$S_t = \frac{\sum_i I_{it}}{n}$	$\sum_i x_{it} B_t = \frac{\sum_i x_{it}}{\sum_i x_{it}} 100$
.	.	.	.	.
m	$x_{1m} x_{2m} \dots x_{im} \dots x_{nm}$	$I_{1m} I_{2m} \dots I_{nm}$	S_m	B_m

Adibidea: Hego Euskal Herriko populazioaren datu hauen bidez, populazio-igoerak ikusiko ditugu herrialde bakoitzean.

Datuak:

Urtea	Araba	Bizkaia	Gipuzkoa	Nafarroa
t	x_1	x_2	x_3	x_4
1.900	96.385	311.361	195.850	307.669
1.910	97.181	349.932	226.684	312.235
1.920	98.668	409.550	258.557	329.875
1.930	104.176	485.205	302.329	345.883
(1) 1.940	112.876	511.135	331.753	369.618
1.950	188.012	569.188	374.040	382.932
1.960	138.934	754.383	478.337	402.042
1.970	204.323	1.043.311	631.003	464.867

1. Taula



1. Grafikoa

1. GRAFIKOAN IKUSTEN DIREN GAUZA NABARMENENAK

Grafiko honek, herrialde bakoitzeko **populazioaren hazkunde absolutua** adierazten digu.

Ondorio hauek atera genitzake:

1.900. urtean Nafarroako eta Bizkaiako populazioak berdintsuak ziren, Nafarroak XIX. mendean zuen populazioa askoz handiagoa izan arren.

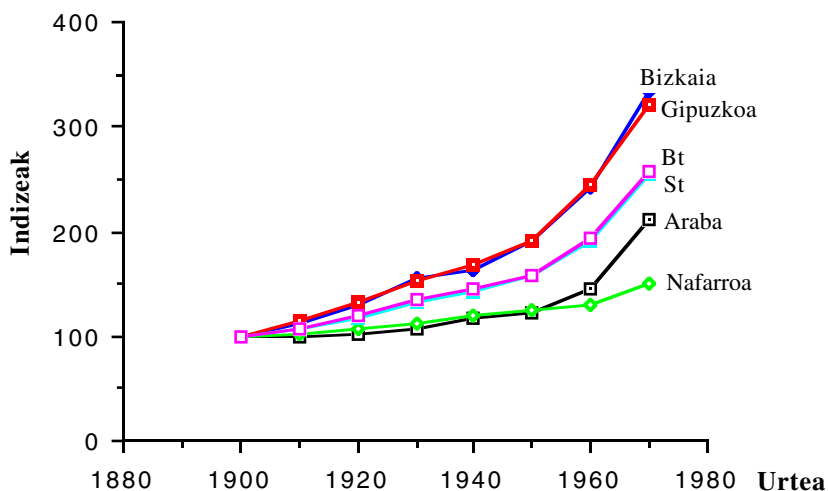
Hamarkada bakoitzean, ibilbidearen malda, gehitzen den biztanle-kopuruaren arabera da, (non gehiketa hamarkadaren hasierako populazio absolutuarekiko baita).

Ikusten denez, Nafarroak eta Arabaren ibilbideak antzekoak dira alde batetik, eta bestetik, Bizkaia eta Gipuzkoarenak ere oso antzekoak dira.

Indizeak:

Urteak	I_{1t}	I_{2t}	I_{3t}	I_{4t}	S_t	$\sum_i x_{it}$	B_t
1.900	100'00	100'00	100'00	100'00	100'00	911'265	100'00
1.910	100'83	112'38	115'74	101'48	107'61	986'032	108'20
1.920	102'37	131'53	132'02	107'22	118'28	1.086'670	119'24
(2) 1.930	108'08	155'83	154'37	112'42	132'67	1.237'593	135'81
1.940	117'11	164'16	169'39	120'13	142'70	1.325'364	145'44
1.950	122'44	191'48	190'98	124'46	157'34	1.444'172	158'48
1.960	144'15	242'28	244'23	130'67	190'33	1.773'696	194'64
1.970	211'98	335'08	322'18	151'09	255'08	2.343'504	257'17

2. Taula



2. Grafikoa

2. GRAFIKOAN IKUSTEN DIREN GAUZA NABARMENENAK

Alde batetik grafiko honetan, herrialde bakoitzari dagokion populazio-indize sinpleen ibilbidea daukagu.

Indize sinple horien ibilbideek, **populazioaren gehiketa erlatiboa, 1.900 urteko populazioarekiko**, adierazten digute, herrialde guztietako populazioa 1.900 urtean, 100 dela suposatuz.

1.900 urtetik zenbat eta gehiago urruntzen garen, orduan eta ibilbide horien maldak esangura gutxiago izango dute.

Beste batetik, Hego Euskal Herriko populazio-indize konplexu, bien (S_t , B_t) ibilbideak ditugu (erdikoak).

Azkenik, hamarkada bakoitzeko urtez urteko hazkunde-tasa bilatu eta 3. grafikoan adierazten dira grafikoki.

Hots:

$$i_{t,t-1} = \frac{x_{i,t}}{x_{i,t-1}} = (1 + \alpha)$$

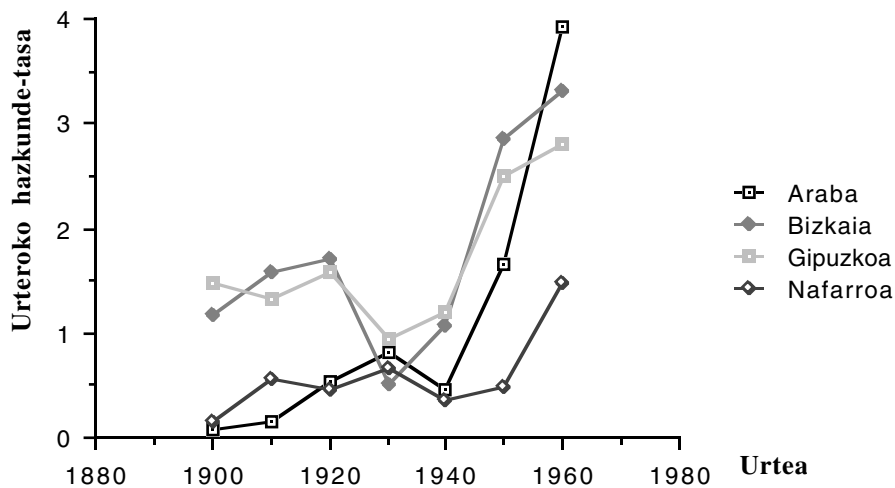
Hemen $i_{t,t-1}$ delakoa momentu baten aurrekoarekiko indize sinplea da. Hau da, $t-1$ uneko unitate bat, $1 + \alpha$ bihurtu da t unean, orduan hamarkada batetan 0 uneko unitate bat, $(1 + \alpha)^{10}$ bihurtu da 10 urtetan, eta exenplu honetan datuak hamarkadaz ditugunez, **urteroko hazkunde-tasa** ondoko formulen bidez kalkulatzeko dugu:

$$\alpha + 1 = \sqrt[10]{\frac{x_{i,t}}{x_{i,t-10}}} \quad \text{eta} \quad \alpha = \sqrt[10]{\frac{x_{i,t}}{x_{i,t-10}}} - 1$$

Ondoko taulan, urteroko hazkunde-tasak adierazten dira:

$\alpha \times 100$	1.900/ 1.910	1.910/ 1.920	1.920/ 1.930	1.930/ 1.940	1.940/ 1.950	1.950/ 1.960	1.960/ 1.970
ARABA	0'08	0'15	0'54	0'81	0'45	1'65	3'93
BIZKAIA	1'17	1'59	1'71	0'52	1'08	2'86	3'30
GIPUZKOA	1'47	1'32	1'58	0'93	1'21	2'49	2'80
NAFARROA	0'15	0'55	0'47	0'67	0'35	0'49	1'47

3.Taula.



3. Grafikoa

3. GRAFIKOAN IKUSTEN DIREN GAUZA NABARMENENAK

Grafiko honetan, herrialde bakoitzari dagokion hazkunde-tasen ibilbidea daukagu.

Hazkunde-tasen ibilbideek **populazioaren gehiketa erlatiboa adierazten** digute, kasu honetan gehiketa **hamarkada bakoitzarekikoa** delarik.

Globalki 7 hamarkada horietan, Bizkaia eta Gipuzkoari dagozkien hazkunde-tasen ibilbideek dute azelerazio gehien.

Geldipen handia nabaritzen da lau herrialdeetan, 1.936 urte inguruan (gerra izan liteke arazo honen zergatia).

Lehenengo hamarkadan nahiko geldikorrak dira hazkunde-tasak bai Araban, bai Nafarroan; gero bietan hazkunde handia nabaritzen da, ordea, 1.930-1.940 hamarkadan, justu Bizkaia eta Gipuzkoan geldipen handia denean (migrazioa eta gerra izan litezke arazo honen zergatia?).

Azken hamarkadetan izugarrizko azelerazioa nabaritzen da lau herrialdeetan, kezagarria Bizkaia eta Gipuzkoaren kasua, azkenean Arabak, Bizkaia eta Gipuzkoaren saturazioa jasotzen duelarik.

III.3. INDIZE KONPLEXU PONDERATUAK

Orain arte, ponderazio gabeko indizeak aztertu ditugu. Baina, adibidez, arroza garia baino bi aldiz gehiago kontsumitzen bada, prezioen indize baten kalkulaketan garrantzi handiagoa eman behar zaie arrozaren prezioaren aldaketei gariarenei baino.

Hau ponderaketaren bidez lortzen da.

Prezioen indize bat kalkulatzean, ondasun bakoitza, kontsumitzen den kopuruaz ponderatzen bada, pautu batzu jarraituz egingo dugu, eta artikulua edo ondasun batentzat aukeratuak izan diren pautak besterentzat ere errespetatu behar dira ahal bada.

Adibidez, hileroko janari prezioen indize batetan, esnearen prezioa bere kantitateaz ponderatzen bada eta kopuru hau famili bati dagokiona baldin bada, (senitarte-kopuru mugatu eta errenta-tarte baten barnekoa), orduan, ogiari dagokion ponderazioak ere erizpide berdinetan oinarritua izan behar du.

III.3.1. Balioen prezioen eta kopuruaren indizeak

“p” delako ondasun baten prezioa bada eta “k” beraren kopurua (saldua, produzitua,...) orduan, “b” delako ondasunaren balioa dela esango dugu eta honela kalkulatu dugu:

$$b = p \cdot k$$

0,1,...,t,...,m denboretan, ondasun baten prezioak

$p_0, p_1, \dots, p_t, \dots, p_m$ eta dagozkien kopuruak

$k_0, k_1, \dots, k_t, \dots, k_m$ baldin badira, orduan “balio-segida denborala” hau izango da:

$$\text{---}b_t\text{---}$$

$$b_0 = p_0 k_0$$

$$b_1 = p_1 k_1$$

$$\vdots$$

$$\vdots$$

$$b_t = p_t k_t$$

$$\vdots$$

$$\vdots$$

$$b_m = p_m k_m$$

Baina denbora iragatean $k = k_t$ bada, b_t balio-segida denboralak, prezioaren gorabeherakadak adieraziko ditu bakarrik. Berdin, denbora iragatean $p = p_t$ bada, b_t balio segidak kopuruaren gorabeherakadak adieraziko ditu bakarrik.

Orduan, ondasun bati dagozkion hiru “balio-segida denboral” desberdin idatz ditzakegu.

b_t	$b_t(k)$	$b_t(p)$
$p_0 k_0$	$p k_0$	$p_0 k$
$p_1 k_1$	$p k_1$	$p_1 k$
$\vdots$		
$p_t k_t$	$p k_t$	$p_t k$
$\vdots$		
$p_m k_m$	$p k_m$	$p_m k$

“O” ondasun bati dagozkion (balio, prezio eta kopuruaren) segida horiek balio errealetan adieraziak direnez gero, beste “ O_i ” ondasun batzuri dagozkienak ere batu ditzakegula, argi ikusten da.

Suposa dezagun $O_1 O_2 \dots O_i \dots O_n$ n ondasun desberdin ditugula, (non O_i ondasun bakoitzaren balio, prezio eta kopuruaren balio-segida denboralak ezagunak zaizkigun), n ondasun horiek, batera osatzen duten “konplexuaren indizeak” edo “indize konplexuak” kalkulatzekoan, balio-segida desberdinen batuketak, laburki, honela adieraz litezke:

$$\begin{array}{ccc}
\underline{\sum_{i=1}^n b_{it}} & \underline{\sum_{i=1}^n b_{it}(k)} & \underline{\sum_{i=1}^n b_{it}(p)} \\
\\
\sum_i p_{i0} k_{i0} & \sum_i p_i k_{i0} & \sum_i p_{i0} k_i \\
\\
\sum_i p_{i1} k_{i1} & \sum_i p_i k_{i1} & \sum_i p_{i1} k_i \\
\vdots & \vdots & \vdots \\
\sum_i p_{it} k_{it} & \sum_i p_i k_{it} & \sum_i p_{it} k_i \\
\vdots & \vdots & \vdots \\
\sum_i p_{im} k_{im} & \sum_i p_i k_{im} & \sum_i p_{im} k_i
\end{array}$$

(Bi azkenetako segidetan, $p_i = k_{it}$, $\forall_i$ eta $k_i = k_{it}$, $\forall_i$ dira).

Beraz, **indize konplexuak** lortzeko konplexuari dagozkion (balio, prezio eta kopurua) hiru **balio segida denboral** 1. horien **indize sinpleak kalkulatu** behar ditugu.

Hots oinarri-denbora $t = 0$ bada:

$$\begin{array}{ccc}
\underline{I_t^b} & \underline{I_t^k} & \underline{I_t^p} \\
\\
\frac{\sum_i p_{i0} k_{i0}}{\sum_i p_{i0} k_{i0}} 100 & \frac{\sum_i (p_i) k_{i0}}{\sum_i (p_i) k_{i0}} 100 & \frac{\sum_i p_{i0}(k_i)}{\sum_i p_{i0}(k_i)} 100 \\
\\
\frac{\sum_i p_{i1} k_{i1}}{\sum_i p_{i0} k_{i0}} 100 & \frac{\sum_i (p_i) k_{i1}}{\sum_i (p_i) k_{i0}} 100 & \frac{\sum_i p_{i1}(k_i)}{\sum_i p_{i0} k_{i0}} 100 \\
\vdots & \vdots & \vdots
\end{array}$$

-
1. Lehenengoa prezio eta kopuruaren funtzioa da, bigarrena, kopuruaren funtzioa eta hirugarrena prezioarena.

$$\begin{array}{ccc}
 \frac{\sum_i p_{it} k_{it}}{\sum_i p_{i0} k_{i0}} 100 & \frac{\sum_i (p_i) k_{it}}{\sum_i (p_i) k_{i0}} 100 & \frac{\sum_i p_{it}(k_i)}{\sum_i p_{i0}(k_i)} 100 \\
 \vdots & \vdots & \vdots \\
 \frac{\sum_i p_{im} k_{im}}{\sum_i p_{i0} k_{i0}} 100 & \frac{\sum_i (p_i) k_{im}}{\sum_i (p_i) k_{i0}} 100 & \frac{\sum_i p_{im}(k_i)}{\sum_i p_{i0}(k_i)} 100
 \end{array}$$

I_t^b I_t^k I_t^p aurreko indizeei, balio, prezio eta kopuruen **indize konplexu agregatuak** deritze, erabiltzen den agregazioa ponderatua bada ere.

Baina orain problema bat sortzen zaigu:

I_t^k “kopuruen indize konplexua” lortzeko, O_i ondasun bakoitzaren zein prezio hartuko dugu konstantetzat?

Edota:

I_t^p “prezioen indize konplexua” lortzeko, O_i ondasun bakoitzaren zein kopuru hartuko dugu konstantetzat?

Problema honetarako nahi beste ebazpide edo soluzio aurki daitezke, baina praktikan bi inposatu dira: LASPEYRES-ena eta PAASCHE-rena.

I_t^k eta I_t^p , lortzeko, horiek urratutako bide desberdinak ikus ditzagun ondoren.

III.3.1.1. LASPEYRES-en indizeak

I_t^k eta I_t^p , prezio eta kopuruen indize konplexuak kalkulatzeko, O_i ondasun bakoitzaren kasuan konstantetzat hartzen ditugun kopuru eta prezioak, oinarri-denboran sortutakoak dira.

Hots, $k_i = k_{i0}$ eta $p_i = p_{i0}$

Honela lortutako indize konplexuei, **LASPEYRES-en prezio eta kopuruen indize konplexuak** deritzegu eta sinbolikoki I_t^k adieraziko ditugu

Hots:

L_t^p	L_t^k	
$\frac{\sum_i p_{i0}(k_{i0})}{\sum_i p_{i0}(k_{i0})} 100$	$\frac{\sum_i (p_{i0})^{k_{i0}}}{\sum_i (p_{i0})^{k_{i0}}} 100$	
$\frac{\sum_i p_{i1}(k_{i0})}{\sum_i p_{i0}(k_{i0})} 100$	$\frac{\sum_i (p_{i0})^{k_{i1}}}{\sum_i (p_{i0})^{k_{i0}}} 100$	
$\vdots$	$\vdots$	
$\frac{\sum_i p_{it}(k_{i0})}{\sum_i p_{i0}(k_{i0})} 100$	$\frac{\sum_i (p_{i0})^{k_{it}}}{\sum_i (p_{i0})^{k_{i0}}} 100$	(1)
$\vdots$	$\vdots$	
$\frac{\sum_i p_{im}(k_{i0})}{\sum_i p_{i0}(k_{i0})} 100$	$\frac{\sum_i (p_{i0})^{k_{im}}}{\sum_i (p_{i0})^{k_{i0}}} 100$	

Oinarrizko eragiketa batez, erraz ikus daiteke (1) formulak, ondoren datozen (2) formula horien baliokideak direla:

L_t^p	L_t^k
$\frac{\sum_i \frac{p_{i1}}{p_{i0}}(p_{i0} \ k_{i0})}{\sum_i (p_{i0} \ k_{i0})} 100$	$\frac{\sum_i \frac{k_{i1}}{k_{i0}}(p_{i0} \ k_{i0})}{\sum_i (p_{i0} \ k_{i0})} 100$
$\vdots$	$\vdots$

$$\begin{array}{cc}
 \frac{\sum_i \frac{p_{it}}{p_{i0}} (p_{i0} k_{i0})}{\sum_i (p_{i0} k_{i0})} 100 & \frac{\sum_i \frac{k_{i1}}{k_{i0}} (p_{i0} k_{i0})}{\sum_i (p_{i0} k_{i0})} 100 \\
 \vdots & \vdots \\
 \frac{\sum_i \frac{p_{it}}{p_{i0}} (p_{i0} k_{i0})}{\sum_i (p_{i0} k_{i0})} 100 & \frac{\sum_i \frac{k_{im}}{k_{i0}} (p_{i0} k_{i0})}{\sum_i (p_{i0} k_{i0})} 100
 \end{array} \quad (2)$$

Ikusten denez, oinarri-denborako prezio edo kopuruaz biderkatzen eta zatitzen da zenbakitzailearen batugai bakoitzean.

Era honetara, batezbesteko aritmetiko ponderatu bezala, (2) formulak oso erabilgarriak zaizkigu: O_i ondasun bakoitzaren indize sinpleak erabiliz LASPEYRES-en prezio edo kopuruen indize konplexuak kalkulatzen ditugu.

III.3.1.2. PAASCHE-ren indizeak

I_t^p eta I_t^k prezio eta kopuruen indize konplexuak kalkulatzeko, O_i ondasun bakoitzaren kasuan indizea kalkulatzen dugun denborako, kopurua eta prezioa hartzen ditugu konstantetzat.

Hots: $k_i = k_{it}$ eta $p_i = p_{it}$

Honela lortutako indize konplexuei, **PAASCHE-ren prezio eta kopuruen indize konplexuak** deritzegu eta sinbolikoki P_t^p P_t^k adieraziko ditugu

Hots:

$$\begin{array}{cc}
 \frac{P_t^p}{100} & \frac{P_t^k}{100} \\
 \frac{\sum_i p_{i1}(k_{i1})}{\sum_i p_{i0}(k_{i1})} 100 & \frac{\sum_i (p_{i1}) k_{i1}}{\sum_i (p_{i1}) k_{i0}} 100 \\
 \vdots & \vdots
 \end{array}$$

$$\begin{array}{ccc}
\frac{\sum_i p_{it}(k_{it})}{\sum_i p_{i0}(k_{it})} 100 & \frac{\sum_i (p_{it}) k_{it}}{\sum_i (p_{it}) k_{i0}} 100 & (1) \\
\vdots & \vdots & \\
\frac{\sum_i p_{im}(k_{im})}{\sum_i p_{i0}(k_{im})} 100 & \frac{\sum_i (p_{im}) k_{im}}{\sum_i (p_{im}) k_{i0}} 100 &
\end{array}$$

Lehen egin genuen bezala, zenbakitzailearen batugai bakoitzean, oinarri-denborako prezio edo kopuruaz biderkatzen eta zatitzen bada (1) formuletatik datozen (2) formuletara iritsiko gara.

$$\begin{array}{ccc}
\frac{P_t^p}{100} & \frac{P_t^k}{100} & \\
\frac{\sum_i \frac{p_{i1}}{p_{i0}} (p_{i0} k_{i1})}{\sum_i (p_{i0} k_{i1})} 100 & \frac{\sum_i \frac{k_{i1}}{k_{i0}} (p_{i1} k_{i0})}{\sum_i (p_{i1} k_{i0})} 100 & \\
\vdots & \vdots & \\
\frac{\sum_i \frac{p_{it}}{p_{i0}} (p_{i0} k_{it})}{\sum_i (p_{i0} k_{it})} 100 & \frac{\sum_i \frac{k_{it}}{k_{i0}} (p_{it} k_{i0})}{\sum_i (p_{it} k_{i0})} 100 & (2) \\
\vdots & \vdots & \\
\frac{\sum_i \frac{p_{im}}{p_{i0}} (p_{i0} k_{im})}{\sum_i (p_{i0} k_{im})} 100 & \frac{\sum_i \frac{k_{im}}{k_{i0}} (p_{im} k_{i0})}{\sum_i (p_{im} k_{i0})} 100 &
\end{array}$$

Ikusten denez, (2) formula hauek ere, indize sinpleen bidez konplexuak kalkulatzera eramaten gaituzte.

III.3.1.3. FISHER-en indizeak

FISHER-en indize konplexu edo “indize ideal” deritzona, LASPEYRES eta PAASCHE-ren indizeen batezbesteko geometrikoa da.

Hots:

$$F_t^p = \sqrt{L_t^p P_t^p} \text{ FISHER-en prezioen indize konplexua}$$

$$F_t^k = \sqrt{L_t^k P_t^k} \text{ FISHER-en kopuruen indize konplexua}$$

III.3.1.4. Propietate eta erlazio batzu: Bateragarritasuna eta alderantzizkotasuna.

LASPEYRES eta PAASCHE-ren indizeak (1) formulen bidez batezbesteko gehikortu ponderatu bezala adierazten dira, eta (2) formulen bidez batezbesteko aritmetikoa ponderatu bezala.

Bateragarritasuna

Indizeen propietate garrantzitsu bat **bateragarri izatea da**, alegia, $B = P \cdot K$ berdintasuna betetzea; B, P eta K balio, prezio eta kopuruen indizeak izanik.

LASPEYRES eta PAASCHE-ren indizeek ez dute betetzen erlazio hau, baina bai FISHER-enak.

Hots:

$$L_t^p \cdot L_t^k \neq B_t$$

$$F_t^p \cdot F_t^k = B_t$$

$$P_t^p \cdot P_t^k \neq B_t$$

LASPEYRES eta PAASCHE-ren konbinaketaz ere betetzen da.

$$\text{Hots: } L_t^p \cdot P_t^k = B_t \quad P_t^p \cdot L_t^k = B_t$$

Sinboloen ordeztu formula osoak ipintzen baditugu, aurreko erlazioak erraz frogatuko ditugu.

Alderantzizkotasuna

t denborako o denborarekiko indizea eta o denborako t denborarekiko indizea elkarren artean balio alderantzizkoak izateari deritzogu alderantzizkotasuna.

Indize sinpleek ondoko erlazio hau betetzen dute:

$$\frac{I_{t,o}}{100} = \frac{1}{I_{o,t} / 100}$$

LASPEYRES eta PAASCHE-ren prezio edota kopuru-indizeek ez dute betetzen, baina bai FISHER-enek.

$$\frac{L_{t,o}^p}{100} \neq \frac{1}{L_{o,t}^p / 100}$$

$$\frac{P_{t,o}^p}{100} \neq \frac{1}{P_{o,t}^p / 100}$$

$$\frac{F_{t,o}^p}{100} = \frac{1}{F_{o,t}^p / 100}$$

LASPEYRES eta PAASCHE-ren prezio edota kopuru-indizeen konbinaketaz ere betetzen da.

$$\frac{L_{t,o}^k}{100} = \frac{1}{P_{o,t}^k / 100} \quad \frac{P_{t,o}^k}{100} = \frac{1}{L_{o,t}^k / 100}$$

III.3.1.5. Kalkulaketa

Ondoko taulan labore batzuren prezio eta kopuruak ditugu, 1.960 eta 1.964 urtetakoak hain zuzen.

	Arroza		Artoa		Garia	
	p	k	p	k	p	k
1.960	24	100	18	40	20	50
1.964	32	80	12	40	40	70

LASPEYRES-en prezioen indizeaz:

$$\sum_i p_{it} k_{io} = 32(100) + 12(40) + 40(50) = 5.680$$

$$\sum_i p_{io} k_{io} = 24(100) + 18(40) + 20(50) = 4.120$$

$$L_{64}^p = \frac{5.680}{4.120} 100 = 138$$

Alegia, oinarri-denborako kopuruaren balioa, prezioen bidez %38 gehitu dela, baina gehienetan, “prezioak 1.960-1.964 tartean, %38 inguruan goratu direla” esango da.

Bestalde, PAASCHE-ren prezio indizeaz:

$$\sum_i p_{it} k_{it} = 32(80) + 12(40) + 40(70) = 5.840$$

$$\sum_i p_{io} k_{it} = 24(80) + 18(40) + 20(70) = 4.040$$

$$P_{64}^p = \frac{5.840}{4.040} 100 = 144$$

Indize honek LASPEYRES-en bezalako esanahia du, ordea, eraiketa desberdina duenez, oraingo emaitza %44 da.

III.4. INDIZE KONPLEXUEN ERAIKETAN SORTZEN DIREN ZENBAIT ERAGOZPEN

III.4.1. Aldagaien hautapena

Dakigun bezala, aldagaien multzo bat batera aztertzea, indize konplexuari dagokio.

Beraz, aipatu dugun multzoaren hautapena, lehendabiziko arazoa izango da. Adibidez, aldagaiak nekazal prezioak baldin badira, beren kopurua nahiko handia izango da, areago, kalitate desberdinak kontuan hartzen baditugu; beraz, gehienetan, ondasun garrantzitsuenetatik azpipopulazio bat hautatuko da.

Honek, produktuen eta beren kalitateen izendatze zehatz batetara eramaten gaitu, honela, denbora iragatean, datu-segidak erizpide berdinekoak izango baitira.

III.4.2. Somatutako leku eta denboraren hautapena

Arduraz produktuen kalitateak definitzea eta somaketan leku eta denborak jakinaraztea komeni da, somatutako zenbakizko datuak hartzean. Prezioen kasuan, lekutzat produkzio, salketa edo kontsumo-lekua har liteke; denboratzat une bat (aste, egun edo hilabete bat) har liteke. Ohituraz, maiztasun, adiera edo garrantzi gehieneko erizpideari jarraituz aukeratzen dira leku eta denborak; beraz ez da posible, diren kasu guztien arteko hautapen aleatorio bat.

III.4.3. Talde eta azpitaldeen hautapena

Indize konplexua kalkulatzeko, honetaz gainera aldagai, talde eta azpitalde batzuren “konplexua” kalkulatzeko badugu, argibide gehiago izango dugu.

Adibidez, nekazal prezioen arloan bi talde handi kontsidera genitzake, sekain lurrena eta lur ureztatuena. Lehenengo taldean azpitaldeak har litezke, adibidez, laboreak, lekariak,..... eta bigarrenean, adibidez, barazkiak, fruituak,... Beste sailkapen batetan, barne-kontsumorako eta esportaketarako produktuak izan zitezkeen. Talde eta azpitalde hauen hautapena, egin nahi den ikerketaren menpe egongo da.

III.4.4. Oinarri-denboraren hautapena

Oinarriari dagozkion datuak, beti ehun bezala hartzen dira, orduan, esan genuen bezala, garrantzitsuena “urte normal” bat hartzea da.

Indizearen barnean dauden aldagaiak, aldaketa nabarmenik ez badute (adibidez, industri produktuetan gertatzen den bezala), urte normal batetakoak oinarritzat har daitezke, bestalde, aldaketa handiak baldin badituzte (adibidez nekazal produktuetan bezala), hiruzpalau urtetako erdineurri bat hartuko da oinarritzat.

Beste arazo garrantzitsu bat zera da, oinarri-denbora oraingo denboratik ez dela oso urrun egon behar, baina, dena dela, beste galdera batetan (oinarri-aldaketari buruzkoan) honetaz gehiago hitz egingo da.

III.4.5. Formula eta ponderazioen hautapena

Formula eta ponderazioek elkarbide bat dute. Ponderazioen informazioa lortu ezin bada, normalki, indize sinplearen batezbesteko aritmetikoa egingo da, baina, lortu ahal bada, gehienetan LASPEYRES, PAASCHE eta FISHER-en indizeak erabiliko dira.

LASPEYRES-en indizean, ponderazioa oinarri-denborakoa denez, informazio gutxiena behar du, PAASCHE-renak gehiago behar du, ponderazioa aldakorra baita eta azkenik, FISHER-ena dakigunez, beste bien artean batezbesteko geometrikoa da.

III.4.6. Indizearen esangura eta zabaldua

Ondasunen kopuruaren menpe zabaldua dago, adibidez, %75 edo %90 bat izango da. Batzutan, hautatu gabe gelditu diren ondasunetariko informazioa baldin badugu, zuzenketak sar daitezke indizeetan.

Indizearen esangura, batezbestekoen ikuspegitik begiratuko dugu; indize konplexuarekin batera sakabanatzearen neurri bat edukitzen dute eta hau handiago da oinarri-denboratik urruntzen garenean.

LASPEYRES eta PAASCHE-ren indizeak batera kalkulatzeko baditugu, emaitzen desberdintasuna gehitu egingo da oinarri-denboratik urruntzen garenean eta hau sakabanatzearen aipamen bezala har dezakegu.

Desberdintasun hau nahiko handia egiten denean, oinarrien berriztapena egitea komeniko da.

III.5. ERAGOZPEN BEREZI BATZU

III.5.1. Oinarri-aldaketa indize sinpleetan

Adibide batez ikus dezagun

Urtea	I_t (oinarria = 1.955)	I_t (oinarria = 1.965)
1.955	100	36'4
1.956	116'7	42'4
1.957	150	54'5
1.958	150	54'5
1.959	158'3	57'6
1.960	125	45'5
1.961	100	36'4
1.962	133'3	48'5
1.963	166'7	60'5
1.964	200	72'7
1.965	275	100

Salmenten indize bat baldin bada eta oinarria 1.965. urtean berritzen bada, momentu horretan segida denborala hautsi egiten da eta segida denboral berrian atzerantz erregela proportzional baten bidez (I_t -ren zenbaki bakoitza bider 100 / 275 biderkatzean) egiten ditugu kalkulaketak.

Alegia, 1.955 urteko salmentak %100 baldin baziren, orain 1.965 urtekoak %100 dira.

Bi segida hauek adierazten digute: %33'3 gehiago saldu zela 1.962. urtean 1.955.ean baino eta %51'5 gutxiago saldu zela 1.962. urtean 1.965.ean baino.

III.5.2. Berriztapen eta loturak indize konplexuetan

Produkzio, aktibitate eta gastuen aldaketaren ondorio bezala, berriztapena inposatzen da. Momentu batetan ondasunik erabilgarrienak zirenak, bigarren postu batetan gelditzen dira denbora iragatean edo desagertu egiten dira. Bestaldetik ondasun berriak agertzen dira, gero arrunt bihurtzen direlarik.

Horregatik ondasun multzoa ez dezakegu utz aldatu gabe. Denbora iragatean, aldagaiak, oinarriak eta ponderazioak aukeratu beharko ditugu berriro.

Indize konplexuen berriztapenak, segidaren hausturaren arazoa sortzen du.

Ikus dezagun nola egiten den lotura kasu erreal batetan:

Estatistika-Institutuak (Estatu espainolekoak), kontsumo-prezioen indizea berriztatu egiten zuen eta bi segida daude, bata 1.936. urtean oinarritua eta bestea 1.958.ean.

Kontsumo-prezioen indizea:

<u>Urtea</u>	<u>Oinarria 1.936 = 100</u>	<u>Oinarria 1.958 = 100</u>
1.956	643'1	—
1.957	712'4	—
1.958	807'7	100
1.959	866'7	—
1.960	876'9	—
1.961	—	111'3
1.962	—	117'3

Lotura-eragiketak segida bakat bat lortzen duenez gero, 1.936. urtean edo 1.958.ean oinarrituz egin liteke. Bietan indize sinpleetan ikusi genuen erregela proportzionalen bidez egingo da.

Bi segidak, bi indize desberdinak ditugun urtean (1.958 urtea kasu honetan) berdinak egiten dira. Oinarri bakartzat 1.936. urtea hartzen badugu, orduan 1.958.eko 100, bider 8'077 biderkatzean 807'7 bihurtzen da, eta honela egingo da segida osoaz, aldiz, oinarri bakartzat 1.958 urtea hartzen badugu, orduan, 1.936.ean oinarriturik dagoen segida, 8'077 balioaz zatituko da.

Ondoko taulan, loturik dauden bi segida ikusiko ditugu.

<u>Urtea</u>	<u>Oinarria 1.936 = 100</u>	<u>Oinarria 1.958 = 100</u>
1.956	643'1	79'5
1.957	712'4	82'2
1.958	807'7	100
1.959	866'7	107'3
1.960	876'9	108'6
1.961	899'0	111'6
1.962	947'2	111'4

Argiro, indize sinpleen alderantziz, bi segida hauek ez dira konbaragarriak, aldagai eta ponderazio desberdinez eginak baitaude, baina, ez dugu beste biderik, indize konplexuari jarraitasuna eman nahi badiogu.

Ez dugu ahaztu behar, indizeak fenomeno baten eboluzioa ikusteko erabiltzen diren indikadore batzu besterik ez direla.

III.6. ZENBAKI-INDIZEEN APLIKAPENAK

Indizeen formulak era general batez ikusi ondoren, bi multzotan sailka ditzakegu, prezio eta kopuruenak, baina, formula hauek, indize batzutan zehaztu egiten dira, adibidez, industri produkzioen indizeak, alokairuen indizeak, kontsumo-prezioen indizeak, balore mugigarrien indizeak, inportazioen indizeak, esportazioen indizeak eta abar, gehienak ekonomi aktibitatearen eboluzioa ikusteko baliagarriak direlarik.

Ikasgai honetan, kontsumo-prezioen indizea ikusiko dugu, garrantzitsuena baita.¹

III.6.1. Kontsumo-prezioen indizeak

Hipotesi bezala familia baten bizimaila konstantetzat hartzen badugu, denbora-une desberdinetan familia horrek kontsumoan ordaintzen duenaren gain konparaketak egitea da indize honen helburua. Alegia, aurrekontu (familia baten bizimaila deritzona) baten aldaketak neurtu nahi ditu.

Egikera erabiliena “erosketa-otarrea” deitzen dena da; hau inkesta baten bidez, famili aurrekontuen inkesta, alegia, aurrekontuan sartzen diren ondasunen kopuruek, zerbitzuek eta industri produktuek osatzen dute.

Hartzen den familia, “normala” da, hots, bere aurrekontua batezbestekoari hurbiltzen zaio.

Oinarri-denborako prezioaz eta oraingo prezioaz “otarre”-aren balioa kalkulatzeko, arazo bakarra izango da, eta bi hauen zatidura kontsumo-prezioen indizea da.

Izan bitez:	
$k_{10} \ k_{20} \ \dots \ k_{n0}$	“otarre”-ko ondasunen kopuruak (bizimaila presuposatzen dutenak).
$P_{10} \ P_{20} \ \dots \ P_{n0}$	haien prezioak oinarri-denboran.
eta	
$P_{11} \ P_{21} \ \dots \ P_{n1}$	haien prezioak oraingo unean.

1. Besteak UEU-k 1979. urtean argitaratu zuen ZENBAKI-INDIZEAK liburuxkan ikus daitezke.

Orduan **kontsumo-prezioen indizea**, ondoko formulaz edukiko dugu.

$$I_{1,0} = \frac{\sum_{i=1}^n p_{i1} k_{i0}}{\sum_{i=1}^n p_{i0} k_{i0}} 100$$

Ikusten denez, indize hau LASPEYRES-en indize bat da.

Lehen esan dugunez, denbora iragatean indizearen oinarria aldatuko ez balitz eta honekin batera indizearen estruktura gaurkotuko ez balitz indizea balio gabe geratuko litzateke. Zer esanik ez indizea kontsumo-prezioena bada eta azken urteetan jasaten ditugun aldaketa teknologiko eta ekonomikoen egoeraz.

Espainiako Estatistika Institutuak KPI kalkulatzeko oinarri hauek finkatu ditu: 1936, 1958, 1976, 1983 eta 1985.eko Uztaila.

Dakusagunez lehen aldaketa 22 urte iragan ondoren izan zen, bigarrena 18 urte iragan ondoren, azkeneko urteetan gero eta hurbilago dago aldaketa, kontsumoaren estruktura aldaketak beharturik.

***IV. EZAUGARRI ESTATISTIKO BIKOITZAK,
BANAKETAK, TAULAK, ADIERAZPIDE
GRAFIKOAK***

IV.1. EZAUGARRI ESTATISTIKO BIKOITZAK

IV.2. MAIZTASUN-BANAKETA BIKOITZAK

***IV.2.1. Maiztasun-banaketak bikoitzak eta
bazter-maiztasunak***

IV.2.2. Maiztasun-banaketa bikoitzak: Taulak

IV.3. MAIZTASUN-BANAKETA BIKOITZEN

ADIERAZPIDE GRAFIKOAK

IV.3.1. Sakabanatze-diagrama edo puntu-hodeia

IV.4. MAIZTASUN-BANAKETA BALDINTZATUAK

***IV.5. TAULA BATEN DEPENDENTZIA EDO
INDEPENDENTZIA***

***IV.6. KONTINGENTZI TAULA BATEN ERRENKADA
ETA ZUTABEEN BATEZBESTEKO SOSLAIK***

IV.1. EZAUGARRI ESTATISTIKO BIKOITZAK

Orain arte ikusi ditugun fenomenoetan, indibiduo bakoitzari ezaugarri estatistiko bakoitzarekiko neurketa bat besterik ez zitzaion egiten.

Orain aipatuko ditugun ikerketetan, lagineko **indibiduo bakoitzari bi neurketa** egiten dizkiogu, bakoitza ezaugarri estatistiko bati dagokiona.

Ezaugarri estatistiko bikoitzak edo ezaugarri estatistiko bidimentsionalak ikusiko ditugu, bada, ondoren.

Adibidez: Urte batetan soldaduzkara doazen gazteei bi neurketa egiten zaizkie, edo beraien bi ezaugarri aztertzen dira batera: pisua eta altuera.

IV.2. MAIZTASUN-BANAKETA BIKOITZAK

IV.2.1. Maiztasun-banaketak bikoitzak eta bazter-maiztasunak

Suposa dezagun N indibiduo osatutako lagin bat, non X ezaugarri batek $x_1 x_2 \dots x_i \dots x_k$ balio desberdinak hartzen dituen.

Suposa dezagun, halaber, beste Y ezaugarri bakun batek lagin berdineko indibiduoetan $y_1 y_2 \dots y_j \dots y_l$ balio desberdinak hartzen dituela.

Orduan, (X,Y) ezaugarri bikoitzak, lagineko indibiduo bakoitzean balio-bikote bat hartzen du.

Hots:

$$\omega_p \rightarrow (x_i, y_j) \quad \begin{array}{l} p = 1, \dots, N \\ i = 1, \dots, k \\ j = 1, \dots, l \end{array}$$

eta (X,Y) ezaugarri bikoitzak, laginean hartuko lituzkeen **balio-bikote desberdinen** kopurua, $k \times l$ izango da; hemendik aurrera horiek bakarrik hartuko ditugu kontuan.

Hots:

$$\{ \omega_1, \omega_2 \dots \omega_N \} \xrightarrow{(X,Y)} \{ (x_1 y_1) (x_1 y_2) \dots (x_1 y_l) \dots \\ \dots (x_i y_l) \dots (x_k y_1) \dots \\ \dots (x_k y_l) = (x_i y_j) \}$$

(x_i, y_j) **balio-bikotearen MAIZTASUN ABSOLUTUA**

Sinbolikoki n_{ij} ikurraz adieraziko dugu, eta laginean (x_i, y_j) balio-bikotea hartzen duten indibiduen kopurua da.

Non:

$$\sum_{j=1}^l \sum_{i=1}^k n_{ij} = n_{..} = N$$

(x_i, y_j) **balio-bikotearen MAIZTASUN ERLATIBOA**

Sinbolikoki f_{ij} ikurraz adieraziko dugu, non:

$$f_{ij} = \frac{n_{ij}}{N} \quad \sum_{j=1}^l \sum_{i=1}^k f_{ij} = 1$$

eta portzentaetan:

$$p_{ij} = \%f_{ij} \cdot 100 \quad \sum_j \sum_i p_{ij} = 100$$

x_i **balioaren BAZTER-MAIZTASUN ABSOLUTUA**

Sinbolikoki $n_{i.}$ ikurraz adieraziko dugu, non:

$$n_{i.} = \sum_{j=1}^l n_{ij} \quad \sum_i n_{i.} = n_{..} = N$$

y_j **balioaren BAZTER-MAIZTASUN ABSOLUTUA**

Sinbolikoki $n_{.j}$ ikurraz adieraziko dugu, non:

$$n_{.j} = \sum_{i=1}^k n_{ij} \quad \sum_j n_{.j} = n_{..} = N$$

x_i **balioaren BAZTER-MAIZTASUN ERLATIBOA**

Sinbolikoki $f_{i.}$ ikurraz adieraziko dugu, non:

$$f_{i.} = \frac{n_{i.}}{N} = \frac{\sum_j n_{ij}}{N}$$

y_j balioaren **BAZTER-MAIZTASUN ERLATIBOA**

Sinbolikoki $f_{.j}$ ikurraz adieraziko dugu, non:

$$f_{.j} = \frac{n_{.j}}{N} = \frac{\sum_i n_{ij}}{N}$$

eta

$$f_{..} = \sum_i f_{i.} = \sum_j f_{.j} = 1$$

IV.2.2. Maiztasun-banaketa bikoitzak: Taulak

Lagin bati dagozkion maiztasun-banaketen taulak, ezaugarria bikoitza bada, ondoan azaltzen ditugunak dira.

$\begin{matrix} Y \\ X \end{matrix}$	y_1	y_2	..	..	y_j	..	..	y_l	
x_1	n_{11} f_{11}	n_{12} f_{12}	..	..	n_{1j} f_{1j}	..	..	n_{1l} f_{1l}	$n_{1.}$ $f_{1.}$
x_2	n_{21} f_{21}	n_{22} f_{22}	..	..	n_{2j} f_{2j}	..	..	n_{2l} f_{2l}	$n_{2.}$ $f_{2.}$
.			..	..		..	..		
.			..	..		..	..		
x_i	n_{i1} f_{i1}	n_{i2} f_{i2}	..	..	n_{ij} f_{ij}	..	..	n_{il} f_{il}	$n_{i.}$ $f_{i.}$
.	.	.	..	..	.	..	..	.	.
.	.	.	..	..	.	..	..	.	.
x_k	n_{k1} f_{k1}	n_{k2} f_{k2}	..	..	n_{kj} f_{kj}	..	..	n_{kl} f_{kl}	$n_{k.}$ $f_{k.}$
Guztira	$n_{.1}$ $f_{.1}$	$n_{.2}$ $f_{.2}$			$n_{.j}$ $f_{.j}$			$n_{.l}$ $f_{.l}$	$n_{..}=N$ $f_{..}=1$

BAZTER-BANAKETAK

X	M. Ab.	M. Er.	Y	M. Ab.	M. Er.
x_1	$n_{1.}$	$f_{1.}$	y_1	$n_{.1}$	$f_{.1}$
x_2	$n_{2.}$	$f_{2.}$	y_2	$n_{.2}$	$f_{.2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	$n_{i.}$	$f_{i.}$	y_j	$n_{.j}$	$f_{.j}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	$n_{k.}$	$f_{k.}$	y_l	$n_{.l}$	$f_{.l}$

$$\sum_i n_{i.} = N$$

$$\sum_j n_{.j} = N$$

$$\sum_i f_{i.} = 1$$

$$\sum_j f_{.j} = 1$$

(X edo Y ezaugarriak edo biak jarraiak badira, ezaugarri bakunetan bezala mailakatuz diskretu bihurtuko ditugu).

Adibidea: 750 familia duen talde batean bi ezaugarri aztertu dira batera. Bata aktibo diren pertsonen kopurua, eta bestea hileroko soldata (mila pezetatan).

Ondoko taulan dauzkagu maiztasun absolutuen banaketak.

Hileroko soldata mila pezetatan

Pertsona aktiboak	(60,80)	(80,140)	(140,250)	Guztira
1	143	59	--	202
2	46	170	32	248
3	----	93	86	179
4	----	17	104	121
Guztira	189	339	222	750

Bazter-maiztasunaren esangura, bereziki azpimarratzea komeni zaigu.

Ikus dezagun nola maiztasun absolutuen kopuruetatik (ezaugarri bikoitzari dagozkionak), batuketa batzu eginez, bazter-maiztasunen banaketak lortzen

ditugun; hau da, errenkadak batuz ezaugarri batenak (pertsona aktiboak) eta zutabeak batuz beste ezaugarriarenak (hileroko soldata).

Ohargarria da, bada, bazter-maiztasunen banaketak, maiztasun bakunen banaketa batzu besterik ez direla.

IV.3. MAIZTASUN-BANAKETA BIKOITZEN ADIERAZPIDE GRAFIKOAK

Aldagai bikoitz baten maiztasun-banaketa ikusterazteko bi adierazpide grafiko ditugu, ESTEREOGRAMA eta SAKABANATZE-DIAGRAMA edo PUNTU-HODEIA hain zuzen. Hauetariko bigarrena ondoan azalduko dugu, bestea ez baita ia erabiltzen.

IV.3.1. Sakabanatze-diagrama edo puntu-hodeia

Izan bedi (X,Y) ezaugarri estatistiko bikoitz bat, eta izan bitez

$$\{(x_i y_j)\} \quad \begin{matrix} i = 1 \dots k \\ j = 1 \dots l \end{matrix}$$

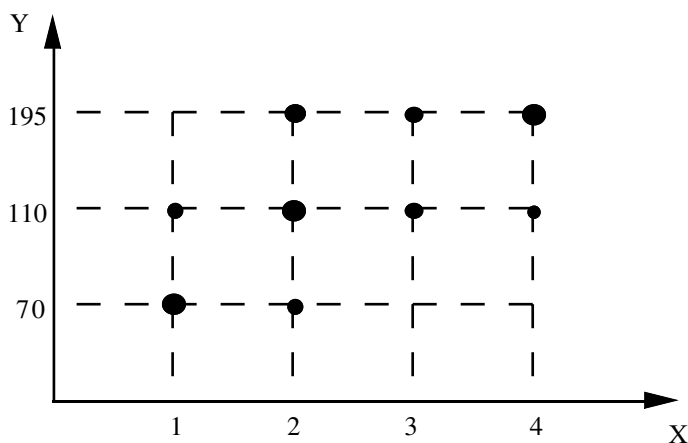
(X,Y) ezaugarriak N tamainuko lagin batetan hartzen dituen balio-bikote desberdinak.

Orduan, ardatz kartesiarra sistema batetan puntu batez adierazten badugu, $(x_i y_j)$ balio-bikote bakoitza,

$$\{(x_i y_j)\} \quad \begin{matrix} i = 1, \dots, k \\ j = 1, \dots, l \end{matrix}$$

multzo osoaren adierazpenari, sakabanatze-diagrama edo puntu-hodeia deritzo.

Ondoan, 2.2. ataleko adibideari dagokion puntu-hodeia daukagu.



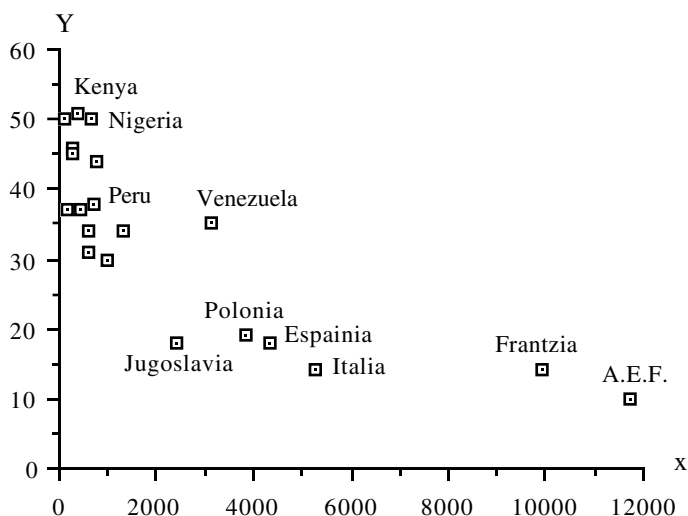
Ikusten dugunez, (x_i, y_i) balio-bikoteak errepikatzen direnean sakabanatze-diagramako puntuetariko batzu besteak baino lodiagoak dira, beraien azalerak eta balio-bikoteei dagozkien maiztasunak proportzionalak izanik.

Dena dela, taula bikoitzak edo datu taldekatuen taulak eta beraren dispertsio-diagramak informazio berdintsua ematen digute.

Batez ere datu-multzo handiak ditugunean, daturik isolatuenak edo ezer gutxi errepikatzen diren datuenak dira taulak; eta horrelako kasuetan, puntu-hodeiak oso baliagarriak zaizkigu, multzoaren ibilbidea ikusterazteko.

X ezaugarria Nazio-Produktu Gordina eta Y jaiokortasuna izanik, ondoan 20 estaturi dagozkien datuak eta beraien puntu-hodeia ditugu.

<u>ESTATUA</u>	<u>IZENA</u>	<u>NPG</u>	<u>JAIOKORTASUNA</u>
1	Ethiopia	130	50
2	Birmanian	160	37
3	Tanzania	270	46
4	Uganda	290	45
5	Kenya	380	51
6	Egipto	460	37
7	Thailandia	590	31
8	Filipinas	600	34
9	Nigeria	670	50
10	Peru	730	38
11	Marokko	740	44
12	Kolonbia	1010	30
13	Turkia	1330	34
14	Jugoslavia	2430	18
15	Venezuela	3130	35
16	Polonia	3830	19
17	Espainia	4340	18
18	Italia	5240	14
19	Frantzia	9940	14
20	A. Err. Federatua	11730	10



IV.4. MAIZTASUN-BANAKETA BALDINTZATUAK

(X,Y) ezaugarri edo aldagaien banaketa bikoitzaren kasuan eta X aldagaiaren x_i balioarentzat, Y aldagaiaren maiztasun-banaketa kontsidera daiteke.

$\begin{array}{c} Y \\ \diagdown \\ X \end{array}$	y_1	y_2		y_j		y_l	
.	.	.		.		.	
.	.	.		.		.	
.	.	.		.		.	
x_i	n_{i1}	n_{i2}		n_{ij}		n_{il}	$n_{i.}$
.	.	.		.		.	
.	.	.		.		.	
.	.	.		.		.	

x_i balioa baldintzatzailea izanez, maiztasun-banaketa baldintzatuak ondokoak ditugu:

$Y/X = x_i$	$n_{j/i}$	$f_{j/i}$
y_i	$n_{1/i} = n_{i1}$	$\frac{n_{1/i}}{n_{i.}}$
y_2	$n_{2/i} = n_{i2}$	$\frac{n_{2/i}}{n_{i.}}$
$\cdot$	$\cdot$	$\cdot$
$\cdot$	$\cdot$	$\cdot$
$\cdot$	$\cdot$	$\cdot$
y_j	$n_{j/i} = n_{ij}$	$\frac{n_{j/i}}{n_{i.}}$
$\cdot$	$\cdot$	$\cdot$
$\cdot$	$\cdot$	$\cdot$
$\cdot$	$\cdot$	$\cdot$
y_l	$n_{l/i} = n_{il}$	$\frac{n_{l/i}}{n_{i.}}$

eta (1) x_i balioaz baldintzatutako Y aldagaiaren maiztasun absolutuen banaketa deitzen da.

Maiztasun-banaketa honen balioak, $n_{i.}$ bazter-maiztasunaz zatitzen baditugu (2), maiztasun-banaketa erlatiboa lortzen dugu, eta:

$$f_{j/i} = \frac{n_{j/i}}{n_{i.}} = \frac{n_{ij}}{n_{i.}}$$

X aldagaiak hartzen dituen balio guztiak kontsideratuz, beste hainbeste Y aldagaiaren banaketa baldintzatu edukiko ditugu.

Analogikoki, X aldagaiaren banaketa baldintzatuak edukiko ditugu, Y aldagaiak hartzen dituen balio guztiak baldintzatzailerak izanez.

Adibidea:

Ondoko taulan (X,Y) aldagaien banaketa daukagu. Y aldagaiak pieza baten luzera adierazten du.

X aldagaiak, ordea, akastuna (B) edo akasgabea (A) izatea.

$X \backslash Y$	71	72	73	$n_{i.}$
B	3	1	5	9
A	2	4	6	12
$n_{.j}$	5	5	11	21

Maiztasun-banaketa baldintzatuak (erlatiboak) hauek dira:

$f(y/x = B)$	3/9	1/9	5/9
$f(y/x = A)$	2/12	4/12	6/12

$f(x/y = 71)$	$f(x/y = 72)$	$f(x/y = 73)$
3/5	1/5	5/11
2/5	4/5	6/11

Horrelako banaketa baldintzatuak zera adierazten digute: aldagai baten balio bat finko kontsideratuz, beste aldagaiaren balioek dauzkaten proportzioak zeintzu diren.

Hau da: Pieza horiek osatzen duten populazioan akastunen azpipopulazioa kontsideratuz, zer proportziotan daude luzera desberdinetakoak akastunen artean? Erantzuna $f(y/x = B)$ banaketak ematen digu.

Analogikoki, akasgabeen azpipopulazioa kontsideratzen badugu, $f(y/x = A)$ edukiko dugu.

Aldiz, Y aldagaiaren balio bat baldintzatzailea baldin bada, dagokion azpipopulazioa kontsideratuko dugu.

Adibidez, Y aldagaiak luzera milimetrotan adierazten badu eta 73 mm neurtzen dituztenean azpipopulazioa kontsideratuz, zer proportziotan dira akastunak ala akasgabeak pieza horiek? Erantzuna kasu honetan, $f(x/y = 73)$ banaketak ematen digu.

IV.5. TAULA BATEN DEPENDENTZIA EDO INDEPENDENTZIA

X eta Y aldagaiak kontsideratuz, aldagai baten aldaketarekin bestearen banaketa aldatzen bada, biak elkarren artean dependenteak dira.

Dakusagun ondoko taula:

$\begin{matrix} Y \\ X \end{matrix}$	71	72	73	$n_{i.}$	$f_{i.}$
B	3	1	5	9	9/21
A	2	4	6	12	12/21
$n_{.j}$	5	5	11	21	
$f_{.j}$	5/21	5/21	11/21		1

X aldatzean, Y aldagaiaren banaketa aldatzen da, hau da:

y_j/x_i	$n(y/x = B)$	$n(y/x = A)$	$f(y/x = B)$	$f(y/x = A)$
71	3	2	1/3	1/6
72	1	4	1/9	1/3
73	5	6	5/9	1/2

Luzera horiek, bada, ez daude berdin banatuta akastunen artean eta akasgabeen artean, hots, nolabaiteko menpekotasun edo dependentzia ikusten da taula honetan.

Demagun, orain, honako taula hau:

$\begin{array}{c} Y \\ \diagdown \\ X \end{array}$	71	72	73	$n_{i.}$	$f_{i.}$
B	3	6	9	18	18/30
A	2	4	6	12	12/30
$n_{.j}$	5	10	15	$30 = N = n_{..}$	
$f_{.j}$	5/30	10/30	15/30		

X aldatzean, Y aldagaiaren banaketa ez da aldatzen, hau da:

y_j/x_i	$n(y/x = B)$	$n(y/x = A)$	$f(y/x = B)$	$f(y/x = A)$
71	3	2	3/18	2/12
72	6	4	6/18	4/12
73	9	6	9/18	6/12

Ikusten dugunez, Y aldagaiaren banaketa baldintzatuak berdinak dira, luzerak proportzio berdinetan daude akastunen eta akasgabeen artean, eta Y aldagaiaren bazter-banaketa (5/30, 10/30, 15/30) ere berdina da.

Hau da, x_i balioaz baldintzatutako y_j balioaren maiztasun erlatibo baldintzatua eta y_j balioaren bazter-maiztasun erlatiboa berdinak dira:

$$\frac{n_{ij}}{n_{i.}} = \frac{n_{.j}}{N}$$

X aldagaiaren banaketa baldintzatuak eta bazter-banaketa kontsideratuz, gauza bera gertatzen da, hots, Y aldagaiaren balioak aldatzean, X aldagaiaren banaketa ez da aldatzen eta X eta Y independenteak dira.

Aurreko erlazioaren simetrikoa edukiko dugu, hau da:

$$\frac{n_{ij}}{n_{.j}} = \frac{n_{i.}}{N}$$

Bi erlazio horietatik erraz ateratzen den independentziaren baldintza beharrezkoa eta nahikoa ondokoa da:

$$f_{ij} = f_{i.} f_{.j} \quad \forall (i,j)$$

Hots, edozein (x_i, y_j) bikoterentzat, bazter-maiztasun erlatiboen biderkadura eta bikote horren maiztasun erlatiboa berdinak dira.

IV.6. KONTINGENTZI TAULA BATEN ERRENKADA ETA ZUTABEEN BATEZBESTEKO SOSLAIK

X	Y	$Y_1 \dots y_j \dots y_l$	$n_{i.}$
x_1		$n_{11} \dots n_{1j} \dots n_{1l}$	$n_{1.}$
$\vdots$		$\vdots$	$\vdots$
x_i		$n_{i1} \dots n_{ij} \dots n_{il}$	$n_{i.}$
$\vdots$		$\vdots$	$\vdots$
x_k		$n_{k1} \dots n_{kj} \dots n_{kl}$	$n_{k.}$
$n_{.j}$		$n_{.1} \dots n_{.j} \dots n_{.l}$	N

Y eta X aldagaien maiztasun erlatiboen bazter-banaketak, Y aldagaiaren banaketa baldintzatuen eta X aldagaiaren banaketa baldintzatuen, batezbesteko soslaiak dira. Y aldagaiaren banaketa baldintzatuen edo errenkada soslaien batezbestekoa hauxe da:

$$\sum_i \frac{n_{ij}}{n_{i.}} f_{i.} = f_{.j}$$

Hots: $f_{.j}, \frac{n_{ij}}{n_{i.}}$ balioen batezbesteko ponderatua da $f_{i.}$ maiztasuna x_i elementu baldintzaile bakoitzaren pisu erlatiboa izanik.

$$\text{Halaber, } \sum_i \frac{n_{ij}}{n_{.j}} f_{.j} = f_{i.}$$

Hots: $f_{i.}, \frac{n_{ij}}{n_{.j}}$ balioen batezbesteko ponderatua da $f_{.j}$ maiztasuna y_j elementu baldintzaile bakoitzaren pisu erlatiboa izanik.

V. EZAUGARRI BIKOITZEN BALIO TIPIKOAK

V.1. MOMENTUAK

V.1.1. Momentu arruntak edo jatorriarekikoak

V.1.2. Momentu zentralak edo batezbestekoarekikoak

V.2. MAIZTASUN-BANAKETA BIKOITZEN BALIO TIPIKO EDO ESTATISTIKOAK

V.2.1. Kobariantza: S_{xy}

V.2.2. Koerlazio-koefizientea: r_{xy}

V.3. ALDAGAIEN ALDAKETAK

V.3.1. Aldagai zentratua

V.3.2. Aldagai tipifikatua edo standardizatua

V.4. ALDAGAI KOERLAZIO GABEEN BI PROPIETATE

V.5. SAKABANATZE ETA KOERLAZIO-MATRIZEAK

V.6. OROKORPENA

V.1. MOMENTUAK

II. ikasgaian, kasu bakunerako gertatu zen bezala, balio tipikoak edo estatistikoak aurkeztu baino lehen, momentuak ikusi behar ditugu eta arrazoi berdinetatik, hots, balio tipikoak momentuak direlako edo momentuen bidez definitzen direlako.

V.1.1. Momentu arruntak edo jatorriarekikoak

Maiztasun-banaketa bikoitz baten h . ordenako momentu arrunta edo jatorriarekikoa, ondoko formularen bidez definitzen da:

$$a_{h_1 h_2} = \frac{1}{N} \sum_i \sum_j x_i^{h_1} y_j^{h_2} n_{ij} \quad h_1 + h_2 = h$$

1. ordenako edo lehenengo ordenako momentu arruntak hauexek dira:

$$\begin{aligned} a_{10} &= \frac{1}{N} \sum_i \sum_j x_i n_{ij} = \frac{1}{N} \sum_i x_i \sum_j n_{ij} = \\ &= \frac{1}{N} \sum_i x_i n_{i.} = \bar{x} \\ a_{01} &= \frac{1}{N} \sum_i \sum_j y_j n_{ij} = \frac{1}{N} \sum_j y_j \sum_i n_{ij} = \\ &= \frac{1}{N} \sum_j y_j n_{.j} = \bar{y} \end{aligned}$$

2. ordenako edo bigarren ordenako momentu arruntak hauexek dira:

$$a_{20} = \frac{1}{N} \sum_i \sum_j x_i^2 n_{ij} = \frac{1}{N} \sum_i x_i^2 \sum_j n_{ij} =$$

$$= \frac{1}{N} \sum_i x_i^2 n_{i.} = a_2(x)$$

hau da, X ezaugarriaren bazter-banaketa kontsideratuz, banaketa bakun honen bigarren ordenako momentu arrunta, aurrekoa da.

$$\begin{aligned} a_{02} &= \frac{1}{N} \sum_i \sum_j y_j^2 n_{ij} = \frac{1}{N} \sum_j y_j^2 \sum_i n_{ij} = \\ &= \frac{1}{N} \sum_j y_j^2 n_{.j} = a_2(y) \end{aligned}$$

Y ezaugarriaren bazter-banaketaren bigarren ordenako momentu arrunta, aurrekoa izango da.

$$a_{11} = \frac{1}{N} \sum_i \sum_j x_i y_j n_{ij}$$

Hirugarren ordenako momentuak hauexek izango lirateke:

$$a_{30}, a_{03}, a_{21}, a_{12}$$

V.1.2. Momentu zentralak edo batezbestekoarekikoak

Maiztasun-banaketa bikoitz baten h-ordenako momentu zentrala edo batezbestekoarekiko, ondoko formularen bidez definitzen da:

$$m_{h_1 h_2} = \frac{1}{N} \sum_i \sum_j (x_i - \bar{x})^{h_1} (y_j - \bar{y})^{h_2} n_{ij}$$

Bereziki, bigarren ordenakoak oso inportanteak dira:

$$m_{20}, m_{02}, m_{11}$$

$$\begin{aligned} m_{20} &= \frac{1}{N} \sum_i \sum_j (x_i - \bar{x})^2 n_{ij} = \frac{1}{N} \sum_i (x_i - \bar{x})^2 \sum_j n_{ij} = \\ &= \frac{1}{N} \sum_i (x_i - \bar{x})^2 n_{i.} = S_x^2 \end{aligned}$$

Hori da, bestetik, X ezaugarriaren bariantza (ikus 39. orrialdea)

$$\begin{aligned} m_{20} &= \frac{1}{N} \sum_i \sum_j (y_j - \bar{y})^2 n_{ij} = \frac{1}{N} \sum_j (y_j - \bar{y})^2 \sum_i n_{ij} = \\ &= \frac{1}{N} \sum_j (y_j - \bar{y})^2 n_{.j} = S_y^2 \end{aligned}$$

Y ezaugarriaren bariantza da.

$$m_{11} = \frac{1}{N} \sum_i \sum_j (x_i - \bar{x}) (y_j - \bar{y}) n_{ij}$$

m_{11} momentua S_{xy} KOBARIANTZA deitzen dena da, balio tipiko inportanteenetariko bat, hain zuzen.

V.2. MAIZTASUN-BANAKETA BIKOITZEN BALIO TIPIKO EDO ESTATISTIKOAK

(X,Y) ezaugarri estatistiko bikoitz bat bada, eta bere maiztasun-banaketa lagin batetan kalkulatu baldin badugu, kalkula genitzake X eta Y-ren arteko erlazioa neurtzen duten estatistiko batzu.

V.2.1. Kobariantza: S_{xy}

X eta Y ezaugarriek, laginean duten erlazio-neurri bat da (non X eta Y ezaugarriak indibiduen gain neurtzen direneko unitateek eragina baitute).

Esan dugunez, m_{11} bigarren ordenako momentu hau kobariantza da; ikus dezagun nola adierazten den momentu arrunten bidez:

$$\begin{aligned} S_{xy} &= m_{11} = \frac{1}{N} \sum_i \sum_j (x_i - \bar{x}) (y_j - \bar{y}) n_{ij} = \\ &= \frac{1}{N} \sum_i \sum_j (x_i y_j - x_i \bar{y} - y_j \bar{x} + \bar{x} \bar{y}) n_{ij} = \\ &= \frac{1}{N} \sum_i \sum_j x_i y_j n_{ij} - \frac{1}{N} \bar{y} \sum_i \sum_j x_i n_{ij} - \end{aligned}$$

$$\begin{aligned}
& - \frac{1}{N} \bar{x} \sum_i \sum_j y_j n_{ij} + \frac{1}{N} \bar{x} \bar{y} \sum_i \sum_j n_{ij} = \\
& = a_{11} - \bar{y} \bar{x} - \bar{x} \bar{y} + \bar{x} \bar{y} = a_{11} - \bar{x} \bar{y}
\end{aligned}$$

V.2.2. Koerlazio-koefizientea: r_{xy}

Estatistiko honek ere X eta Y aldagaiek laginean duten erlazioa neurtzen digu, baina kasu honetan berdin da X eta Y indibiduen gain zer unitatetan neurtzen diren, neurketan erabiltzen diren unitateek ez baitute eraginik koerlazio-koefizientean.

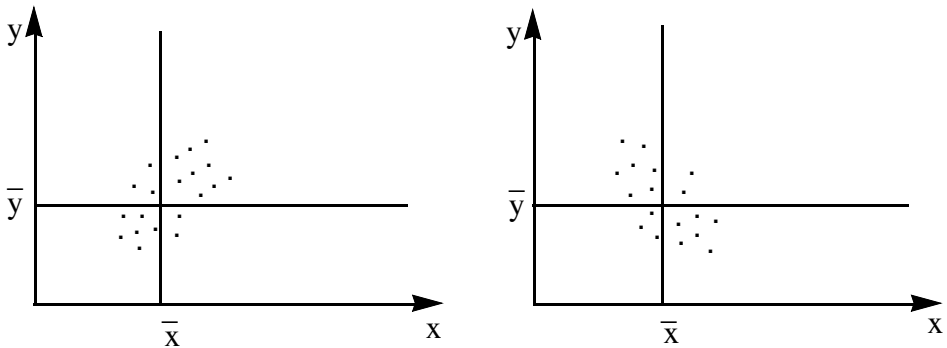
$$\begin{aligned}
r_{xy} &= \frac{S_{xy}}{S_x S_y} = \frac{\sum_i \sum_j (x_i - \bar{x}) (y_j - \bar{y}) n_{ij} / N}{\left(\frac{1}{N} \sum_i (x_i - \bar{x})^2 n_{i.} \right)^{1/2} \left(\frac{1}{N} \sum_j (y_j - \bar{y})^2 n_{.j} \right)^{1/2}} = \\
&= \frac{a_{11} - \bar{x} \bar{y}}{(a_{20} - \bar{x}^2)^{1/2} (a_{02} - \bar{y}^2)^{1/2}}
\end{aligned}$$

Dakigunez S_x eta S_y desbidazio tipikoak, beti positiboak dira; orduan, S_{xy} kobariantza positiboa (negatiboa) baldin bada, r_{xy} koerlazio-koefizientea positiboa (negatiboa) izango da.

Intuitiboki honetaz ohar daiteke: puntu gehienentzako $(x_i - \bar{x})$ eta $(y_j - \bar{y})$ kendurak biak positiboak edo negatiboak baldin badira, (hau da, $(\bar{x}, \bar{y})$ puntua jatorritzat hartzen badugu, sakabanatze-diagramako puntuak 1. eta 3. koadrantean egongo dira) $S_{xy} > 0$ izango da; aldiz, puntu gehienentzako zeinu kontrakoak baldin badira (puntuak 2. eta 4. koadrantean egongo dira), $S_{xy} < 0$.

Hots: kobariantzaren zeinuak puntu-banaketaren ideia bat ematen digu.

Grafikoki:



Kobariantzak nahiz koerlazio-koefizienteak bi ezaugarrien artean dagoen erlazioa adierazten digute; baina ikusi dugu zein kasutan izango den $S_{xy} > 0$, hots, $(x_i - \bar{x})$ eta $(y_j - \bar{y})$ kendurak biak positiboak direnean (hau da, bi ezaugarriak gorakorrek dira) edo biak negatiboak direnean (hau da, bi ezaugarriak beherakorrek). $S_{xy} < 0$ den kasuan, kenduren zeinuak kontrakoak direnez, ezaugarri bat gorakorra da eta beherakorra bestea. Azkenez $S_{xy} = 0$ baldin bada, bi ezaugarriek ez dute erlazio linealik elkarren artean, hau da, koerlazio gabeak dira.

Ohar daiteke taula baten independentziak koerlazio eza halabehartzen duela baina ez alderantziz.

Independentzia $(f_{ij} = f_{i.} \cdot f_{.j}) \Rightarrow$ koerlazio eza $\nRightarrow$

Hots: Independentziaren kasuan

$$S_{xy} = \sum_i \sum_j (x_i - \bar{x})(y_j - \bar{y}) f_{i.} f_{.j} = \sum_i (x_i - \bar{x}) f_{i.}$$

$$\sum_j (y_j - \bar{y}) f_{.j} = m_1(x) \cdot m_1(y) = 0$$

r_{xy} koefizienteak ez du aldagaien arteko erlazioa nolakoa den bakarrik esaten (r_{xy} -ren zeinua S_{xy} -rena baita) baizik eta zer mailatan erlazionatuta dauden ere.

Koerlazio-koefizienteak har ditzakeen balioak, $[-1, +1]$ tartekoak dira.

Zerora hurbiltzen den balio bat hartzen badu, bi ezaugarrien arteko erlazioa ahula da.

r_{xy} koefizientea $+1$ baliora hurbiltzen bada, erlazioa sakona eta zuzena da.

Aldiz, -1 baliora hurbiltzen bada, erlazioa sakona da, baina alderantzizkoa.

V.3. ALDAGAIEN ALDAKETAK

Atal honetan ohar batekin hasiko gara: Aldagai hitza erabiltzean, ezaugarriari gataizkio; baina ez du zerikusirik estatistika teorikoko aldagai aleatorio delakoarekin.

Ikus dezagun, bada, (X,Y) aldagaien ordeztu, haiekin erlazionatuta dauden (U,V) aldagaiak erabiltzen ditugunean, kobariantza aldatu egiten dela eta koerlazio-koefizientea ez.

Horrexegatik eta baita erregresio-teorian betetzen duen funtzioagatik ere, koerlazio-koefizientea, balio tipikorik garrantzitsuena da, bi edo ezaugarri gehiago aztertzen ditugun ikerketetan.

Batzutan, jatorria eta unitatea aldatzea komenigarria da, honela, kalkuluak errazten baitira.

Hots: $x_i = au_i + b$ eta $y_j = ev_j + f$ egiten badugu¹.

$$\bar{x} = \frac{1}{N} \sum_i x_i n_i = \frac{1}{N} \sum_i (au_i + b) n_i =$$

$$= a \frac{1}{N} \sum_i u_i n_i + b \frac{1}{N} \sum_i n_i = a \bar{u} + b$$

$$\text{berdin eginez} \quad \bar{y} = e\bar{v} + f$$

Ikusten dugunez, “X” eta “Y” ordeztu “U” eta “V” aldagaiak hartzen baditugu, batezbesteko aritmetikoak ez dira berdinak; pentsa dezagun estatistiko hau, ezaugarriak neurtzen ditugun unitateetan, neurri bat dela.

1. $b = 0$ baldin bada, unitate-aldaketa egin dugu soilik.
 $a = 1$ baldin bada, jatorri-aldaketa soilik zeina berdin b baita.

Ikus dezagun, koerlazio-koefizientea kalkulatzeko erabiltzen dugun desbidazio tipikoetan eta kobariantzan, zer eragina duen aldaketa honek.

$$\begin{aligned} S_x^2 &= \frac{1}{N} \sum_i (x_i - \bar{x})^2 n_i = \frac{1}{N} \sum_i (au_i + b - a\bar{u} - b)^2 n_i = \\ &= a^2 \cdot \frac{1}{N} \sum_i (u_i - \bar{u})^2 n_i = a^2 \cdot S_u^2 \end{aligned}$$

eta: $S_x = a \cdot S_u$

Analogikoki $S_y = e \cdot S_v$

$$\begin{aligned} S_{xy} &= \frac{1}{N} \sum_i \sum_j (x_i - \bar{x})(y_j - \bar{y}) n_{ij} = \\ &= \frac{1}{N} \sum_i \sum_j (au_i + b - a\bar{u} - b)(ev_j + f - e\bar{v} - f) n_{ij} = \\ &= a \cdot e \cdot \frac{1}{N} \sum_i \sum_j (u_i - \bar{u})(v_j - \bar{v}) n_{ij} = a \cdot e \cdot S_{uv} \end{aligned}$$

Jatorria aldatzeak eraginik ez duela ikusten da baina, bai baduela unitate-aldaketa horrek.

Aldiz, koerlazio-koefizientean ez du eraginik ez jatorri aldaketak, ezta unitate-aldaketak ere.

Hots:

$$r_{xy} = \frac{S_{xy}}{S_x S_y} = \frac{a \cdot e \cdot S_{uv}}{a \cdot S_u \cdot e \cdot S_v} = \frac{S_{uv}}{S_u S_v} = r_{uv}$$

Ikusten dugunez, bi aldagaien arteko koerlazio-koefizientea inbariantea da edozein transformazio linealen aurrean.

Adibidea: Lagin batetan X eta Y ezaugarriak aztertu dira. Maiztasun-banaketaren taula ondokoa bada, kalkula ditzagun: batezbestekoak, bariantzak, kobariantza eta koerlazio-koefizientea, kalkulurik laburrena eginez.

X \ Y	1'735	1'740	1'745	n _{i.}
239	7	9	9	25
243	8	7	5	20
247	11	9	10	30
251	4	7	14	25
n _{.j}	30	32	38	100

$$\text{Aldaketa hauek egiten ditugu } u_i = \frac{x_i - 243}{4}, \quad v_j = \frac{y_j - 1'740}{0'005}$$

$$\text{hots: } x_i = 4u_i + 243 \quad \text{eta } y_j = 0'005 v_j + 1'740$$

honela beste taula honetara iristen gara:

u \ v	-1	0	1	n _{i.}
-1	7	9	9	25
0	8	7	5	20
1	11	9	10	30
2	4	7	14	25
n _{.j}	30	32	38	100

Eta kalkuluak eginez:

$$\bar{u} = \frac{\sum_i u_i n_i}{N} = 0'55$$

$$\bar{x} = 4 \bar{u} + 243 = 245'2$$

$$\bar{v} = \frac{\sum_i v_j n_{.j}}{N} = 0'08$$

$$\bar{y} = 0'005 \quad \bar{v} + 1'740 = 1'7404$$

$$S_u^2 = 1'2475$$

$$S_x^2 = 4^2 \quad S_u^2 = 19'96$$

$$S_v^2 = 0'6736$$

$$S_y^2 = 0'005^2 \cdot S_v^2 = 1'684 \cdot 10^{-5}$$

$$S_{uv} = 0'126$$

$$S_{xy} = 4 \cdot 0'005 \cdot S_{uv} = 0'00252$$

$$r_{xy} = \frac{S_{xy}}{S_x S_y} = \frac{S_{uv}}{S_u S_v} = 0'1374$$

Ikusten dugunez, X eta Y ezaugarrien arteko erlazioa zuzena da eta oso ahula; ez dute ia zerikusirik batak eta besteak.

V.3.1. Aldagai zentratua

Aldagai baten batezbesteko aritmetikoa zero baldin bada, **aldagai zentratua** deitzen da; aldez, batezbesteko aritmetikoa zero eta desbidazio tipikoa bat balioak baldin badira, **aldagai tipifikatua** deritzogu.

Edozein X aldagairen kasuan bere batezbestekoa kenduz $Z = X - \bar{x}$, aldagai zentratua lortzen dugu.

Hau da:

$$\begin{aligned} \bar{z} &= \frac{1}{N} \sum_i z_i n_i = \frac{1}{N} \sum_i (x_i - \bar{x}) n_i = \\ &= \frac{1}{N} \sum_i x_i n_i - \frac{1}{N} \bar{x} \sum_i n_i = \bar{x} - \bar{x} = 0 \end{aligned}$$

Kontutan hartu behar dugu zentratze-eragiketak ez duela unitate-aldaketarik suposatzen, bakarrik jatorri-aldaketa edota aldagaiaren translazioa ardatzean suposatzen duelarik.

Aldagaiaren grabitate-zentrua, translazio honetan, jatorrira eraman dugu, $\bar{z} = 0$ baita.

V.3.2. Aldagai tipifikatua edo standardizatua

Modu berean, edozein X aldagairen kasuan batezbestekoa kenduz eta desbidazio tipikoaz zatituz $Z = (X - \bar{x}) / S_x$, aldagai tipifikatua lortzen dugu.

Ikus dezagun 3.1. atalean emandako definizioa betetzen duela:

$$\begin{aligned}\bar{z} &= \frac{1}{S_x} \sum_i (x_i - \bar{x}) \frac{n_i}{N} = 0 \\ S_z^2 &= \frac{1}{N} \sum_i (z_i - \bar{z})^2 n_i = \frac{1}{N} \sum_i \frac{(\bar{x}_i - \bar{x})^2}{S_x^2} n_i = \\ &= \frac{1}{S_x^2} \frac{1}{N} \sum_i (x_i - \bar{x})^2 n_i = \frac{1}{S_x^2} S_x^2 = 1\end{aligned}$$

Tipifikazio-eragiketak, zentratzeaz gainera unitatez aldatzea suposatzen du. Aldagai bakoitzaren balio zentratuak, bere desbidazio tipikoaz zatitzean, honen arabera neurturik ditugu, jatorrizko neurri-unitateak desagertu direlarik.

KOERLAZIO-KOEFIZIENTEA, aldagai tipifikatuen arteko kobariantza da.

$$\text{Hots: } Z = \frac{(X - \bar{x})}{S_x} \quad \text{eta} \quad Z' = \frac{(Y - \bar{y})}{S_y} \quad \text{baldin badira:}$$

$$\begin{aligned}S_{zz'} &= \frac{1}{N} \sum_i \sum_j \left(\frac{x_i - \bar{x}}{S_x} \right) \left(\frac{y_j - \bar{y}}{S_y} \right) n_{ij} = \\ &= \frac{1}{S_x S_y} S_{xy} = r_{xy}\end{aligned}$$

V.4. ALDAGAI KOERLAZIO GABEEN BI PROPIETATE

2.2. atalean esaten genuen bezala, $S_{xy} = 0$ denean ez dago inongo erlazio linealik X,Y aldagaien artean, horrexegatik $r_{xy} = 0$, hau da, koerlazio-koefiziente lineala zero izatean koerlazio gabeak direla esango dugu.

Dakusagun orain bada, aldagai koerlazio gabeen bi propietateak.

1. X eta Y aldagaiak koerlazio gabeak direnean, beraien baturaren bariantza bariantzen batura da.

$$S_{xy} = 0 \quad S_{(x+y)}^2 = S_x^2 + S_y^2$$

Frogapena:

$$\begin{aligned} S_{(x+y)}^2 &= \frac{1}{N} \sum (x_i + y_i - \bar{x} - \bar{y})^2 = \frac{1}{N} \sum [(x_i - \bar{x}) + (y_i - \bar{y})]^2 = \\ &= \frac{1}{N} \sum [(x_i - \bar{x})^2 + (y_i - \bar{y})^2 + 2(x_i - \bar{x})(y_i - \bar{y})] = \\ &= \frac{1}{N} \sum (x_i - \bar{x})^2 + \frac{1}{N} \sum (y_i - \bar{y})^2 + 2 \frac{1}{N} \sum (x_i - \bar{x})(y_i - \bar{y}) = \\ &= S_x^2 + S_y^2 + 2S_{xy} \end{aligned}$$

Hots:

$$S_{xy} = 0 \Rightarrow S_{(x+y)}^2 = S_x^2 + S_y^2$$

2. X aldagaia Y,Z,... aldagaiekin koerlazio gabea denean, beraien edozein konbinazio linealekin koerlazio gabea da.

$$\left. \begin{array}{l} S_{x,y} = 0 \\ S_{x,z} = 0 \end{array} \right\} \Rightarrow S_{x, ay+bz+c} = 0$$

Frogapena:

$$S_{x, ay+bz+c} = \frac{1}{N} \sum [(x_i - \bar{x})(ay_i + bz_i + c - a\bar{y} - b\bar{z} - c)] =$$

$$\begin{aligned}
 &= \frac{1}{N} \sum_i (x_i - \bar{x}) [a(y_i - \bar{y}) + b(z_i - \bar{z})] = a \frac{1}{N} \sum_i (x_i - \bar{x})(y_i - \bar{y}) + \\
 &+ b \frac{1}{N} \sum_i (x_i - \bar{x})(z_i - \bar{z}) = a S_{x,y} + b S_{x,z} = 0
 \end{aligned}$$

V.5. SAKABANATZE ETA KOERLAZIO-MATRIZEAK

Momentu zentralak honela ere adierazten dira:

$$l_{11} = m_{20} = S_x^2, \quad l_{22} = m_{02} = S_y^2, \quad l_{12} = l_{21} = m_{11} = S_{xy}$$

L eta **R** letraz laburki adierazten ditugun ondoko matrizeak, sakabanatze edo kobariantz matrizea eta koerlazio-matrizeak dira.

$$\mathbf{L} = \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix} \quad \mathbf{R} = \begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix}$$

Ikusten denez, simetrikoak dira biak.

Lehengo adibidekoak hauek dira:

$$\mathbf{L} = \begin{bmatrix} 19'96 & 0'0025 \\ 0'0025 & 1'684 \cdot 10^{-5} \end{bmatrix} \quad \mathbf{R} = \begin{bmatrix} 1 & 0'1374 \\ 0'1374 & 1 \end{bmatrix}$$

V.6. OROKORPENA

Suposa dezagun n ezaugarri edo aldagai ditugula. Binaka hartuz, kobariantzak eta koerlazio-koefizienteak atera daitezke; orduan, **L** eta **R** matrizeak hauxek izango lirarteke.

$$\mathbf{L} = \begin{bmatrix} l_{11} & l_{12} & \dots & l_{1n} \\ l_{21} & l_{22} & \dots & l_{2n} \\ \dots & \dots & \dots & \dots \\ l_{n1} & l_{n2} & \dots & l_{nn} \end{bmatrix} \quad \mathbf{R} = \begin{bmatrix} 1 & r_{12} & \dots & r_{1n} \\ r_{21} & 1 & \dots & r_{2n} \\ \dots & \dots & \dots & \dots \\ r_{n1} & r_{n2} & \dots & 1 \end{bmatrix}$$

non $l_{11}, l_{22}, \dots, l_{nn}$, n aldagaien bariantzak diren hurrenez hurren; $l_{1n} = l_{n1}$ n eta 1 aldagaien arteko kobariantza da.

$r_{1n} = r_{n1}$ n eta 1 aldagaien arteko koerlazio-koefizientea da eta diagonaleko 1 balioak aldagai bakoitzak berekiko duen koerlazioa.

VI. KOERLAZIOA ETA ERREGRESIOA $\mathbb{R}^2$ -n

VI.1. KOERLAZIOA ETA ERREGRESIOA

VI.2. ERREGRESIO-TEORIA

VI.2.1. Batezbestekoaren erregresioa

VI.2.2. Karratu txikien erregresioa

VI.2.2.1. Karratu txikien erregresio lineala

VI.2.2.2. Karratu txikien erregresio parabolikoa

VI.1. KOERLAZIOA ETA ERREGRESIOA

Bi (edo gehiago) aldagai ditugunean, horien elkar-aldaketa oso garrantzizkoa da analisi-estatistiko baten aurrean.

Elkar-aldaketa zehazteko ditugun teknikak, bi motatakoak dira:

- 1) Erregresio tekniken bidez, $y = \emptyset(x, a, b \dots l)$ funtzio matematikoa lortuko dugu, zeinak, x -en balioak ordezkatzuz gero, y -ren balioak emango baitizkigu.
 $\emptyset$ funtzioa puntu-hodeian ondo doitzen bada, Y aldagaiaren balioak auresateko baliagarria izango zaigu.
- 2) Koerlazio-tekniken bidez, ordea, koefiziente batzu lortuko ditugu, aldagaien elkar-aldaketak eta beren norantza ikusaraz diezazkigutenak.

Erregresio-teknikei kuantitatiboa deritze, dagoen menpekotasuna zenbatzen baitute; aldiz, koerlazio-teknikek kualitatiboak izanik, menpekotasun hori handiago edo txikiago den bakarrik esaten digute. Hala ere, koerlazio-koefizienteak guztiz baliagarriak dira erregresio-teorian.

VI.2. ERREGRESIO-TEORIA

Maiztasun-banaketa bikoitz batetan, aldagai baten balio bakoitzari beste aldagaiaren balio batzu dagozkio.

Behin aldagai bat independentetzat jotzen dugunean, balio bakoitzari funtzio matematiko batetan beste aldagaiaren balio bat dagokio, lehenengokoaren funtzioa hain zuzen.

Funtzio matematiko hau erizpide hoberen bati jarraituz lortuko dugu eta dispersio-diagrama edo puntu-hodeia kontutan harturik.

Funtzio honek elkartzen dituen balioak, ez dira normalki maiztasun-banaketarenak.

Erregresio-metodo asko daude, hauetariko batzu hauek direlarik:

- Batezbestekoaren erregresioa
 - Karratu txikien erregresioa
 - Erregresio ortogonala
- $$\left\{ \begin{array}{l} \text{Erregresio lineala} \\ \text{Erregresio parabolikoa} \end{array} \right.$$

Egin dugun erregresioaz baliatuko gara gero, aldagai baten bidez beste bat auresateko, ahal badugu behintzat haxe baita, azken finean, aldagaien erlazioak aztertzean bilatzen dena.

Edozein metodori jarraituz, bi erregresio desberdin kontsidera daitezke: 1) X aldagaia independentetzat hartuz, Y-ren X-ekiko erregresioa deritzona. 2) Y aldagaia independentetzat hartuz, X-ren Y-ekiko erregresioa.

VI.2.1. Batezbestekoaren erregresioa

Y-ren X-ekiko erregresioa kontsideratuz, batezbestekoaren erregresioak X aldagaiaren x_i balio bakoitzari ($Y/X = x_i$) banaketa baldintzatuaren $\bar{y}/x_i$ batezbesteko balioa elkartzen dio.

$x_1, x_2, \dots, x_n$ balioek baldintzatutako Y aldagaiaren banaketak hurrenez hurren $f(y/x_1), f(y/x_2), \dots, f(y/x_n)$, baldin badira, batezbesteko balioak honela lortuko ditugu:

$$\bar{y}/x_1 = \sum_i y_i/x_1 \quad f(y_i / x_1)$$

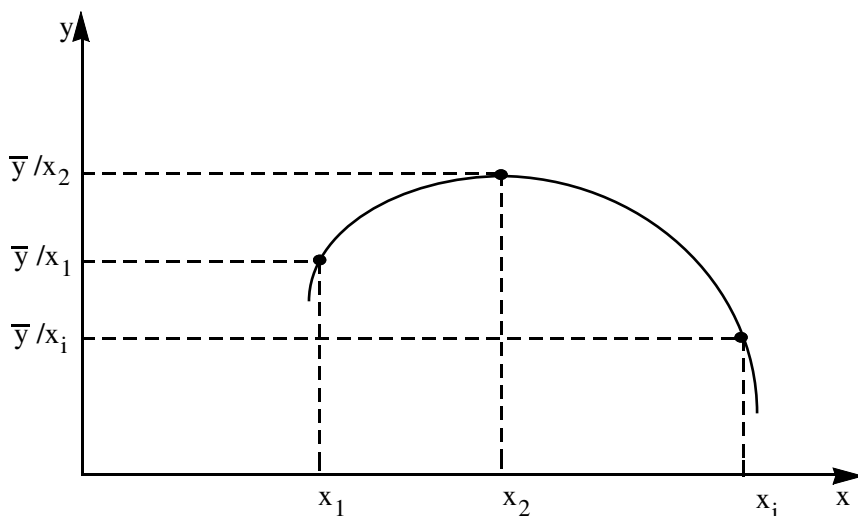
$$\bar{y}/x_2 = \sum_i y_i/x_2 \quad f(y_i / x_2)$$

.

.

.

Eta $(x_1, \bar{y}/x_1), (x_2, \bar{y}/x_2) \dots$ puntuak anitz eta hurbilak balira, kurba bat osatuko lukete. Kurba honi, batezbestekoaren erregresioarena edo erregresio enpirikoarena deritzo, hain zuzen.



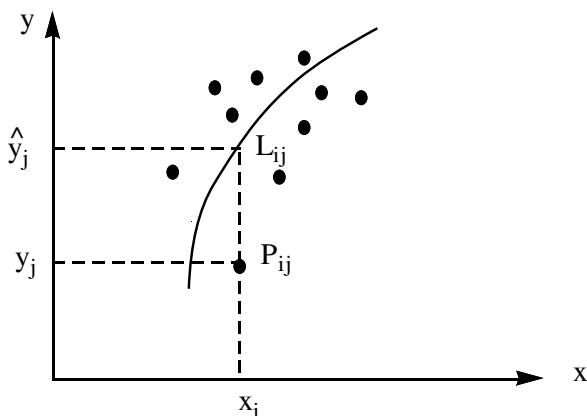
VI.2.2. Karratu txikiaren erregresioa

Karratu txikiaren zentzuan puntu-hodeian ondoen doitzen den kurba lortu behar dugu.

Suposa dezagun kurba hau, $y = \varnothing(x, a, b, \dots, l)$ kurba-familia honetako dela, arazoa orduan $a, b, \dots, l$ parametroak determinatzean datza eta puntu-hodeian ondoen doitzen den kurbari dagozkionak izango dira.

Hau da, kurba bat aukeratu nahi dugu, zeinean ondoko batukaria txikiena izan dadin.

$$\begin{aligned}
 E &= \sum_i \sum_j P_{ij} L_{ij}^2 f(x_i y_i) = \\
 &= \sum_i \sum_j (y_j - \varnothing(x_i, a, b, \dots, l))^2 f(x_i y_i)
 \end{aligned}$$



$P_{ij} = (x_i, y_j)$: balio ohartua edo balio erreala.

$L_{ij} = (x_i, \hat{y}_j)$: balio teorikoa.

$e = y_j - \hat{y}_j = y_j - \emptyset(x_i, a, b, \dots)$: erregresio-errorea.

VI.2.2.1. Karratu txikiaren erregresio lineala

Puntu-hodeia zuzen baten inguruan kondentsatzen da, orduan Y-ren X-ekiko erregresioa egitean zuzen bat lortuko dugu, zeinak abzisak jakinez gero ordenatuak ematen baitizkigu.

$$E = \sum_{ij} \overline{P_{ij} L_{ij}}^2 \quad f_{ij} = \sum_{ij} e^2 f_{ij} = \sum_{ij} (y_j - bx_i - a)^2 f_{ij}$$

txikiena izateko baldintzaz $\hat{y} = a + bx$ zuzeneko a eta b balioak mugatuko ditug

Balio teoriko eta balio errealen arteko diferentzien baturak txikiena izan behar du.

Dakigunez, a-rekiko eta b-rekiko deribatuak zero egitea, baldintza beharrezkoa da E txikiena izan dadin.

Hots:

$$\frac{\partial E}{\partial a} = -2 \sum_{ij} (y_j - bx_i - a) f_{ij} = 0$$

$$\frac{\partial E}{\partial b} = -2 \sum_{ij} (y_j - bx_i - a) x_i f_{ij} = 0$$

hau da:

$$\sum_{ij} (y_j - bx_i - a) f_{ij} = 0; \quad \sum_{ij} (y_j - bx_i - a) x_i f_{ij} = 0$$

edo:

$$\sum_{ij} y_j f_{ij} = \sum_{ij} (bx_i + a) f_{ij}$$

$$\sum_{ij} y_j x_i f_{ij} = \sum_{ij} (bx_i + a) x_i f_{ij}$$

Hots:

$$\begin{aligned}\bar{y} &= a + b\bar{x} \\ a_{11} &= a\bar{x} + ba_{20}\end{aligned}$$

Ekuazio normalak deritze eta, ikusten dugunez, lehenengoa $y = a + bx$ bilatzen ari garen zuzena da $(\bar{x}, \bar{y})$ puntuan hartuta; hau da, zuzena banaketaren grabitate-zentrutik pasatzen da eta $a = \bar{y} - b\bar{x}$ bigarren ekuazioan ordezkatzuz:

$$a_{11} - (\bar{y} - b\bar{x}) \bar{x} - ba_{20} = 0$$

$$a_{11} - \bar{x} \bar{y} - b(a_{20} - \bar{x}^2) = 0$$

$$S_{xy} - bS_x^2 = 0 \Rightarrow b = \frac{S_{xy}}{S_x^2}$$

Orduan:

$$a = \bar{y} - \frac{S_{xy}}{S_x^2} \bar{x}$$

eta $\hat{y} = a + bx$ ekuazioan ordezkatzuz

$$\hat{y} - \bar{y} = \frac{S_{xy}}{S_x^2} (x - \bar{x})$$

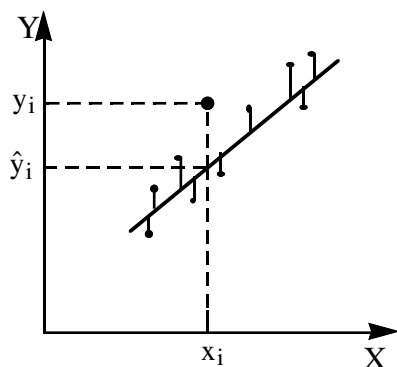
$$\text{Hau Y-ren X-ekiko erregresio-zuzena da, non } b = \frac{S_{xy}}{S_x^2}$$

Y-ren X-ekiko erregresio-koefizientea den (batzutan b_{yx} deritzo).

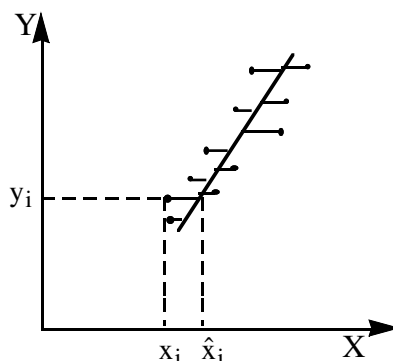
Analogikoki X-ren Y-ekiko erregresio-zuzena ondokoa izango da:

$$\hat{x} - \bar{x} = \frac{S_{xy}}{S_y^2} (y - \bar{y})$$

eta $b_{xy} = \frac{S_{xy}}{S_y^2}$ X-ren Y-rekiko erregresio-koefizientea



(1)



(2)

KOERLAZIO-KOEFIZIENTEA ETA ERREGRESIO-KOEFIZIENTEEN ARTEKO ERLAZIOA:

Dakigunez:

$$r = \frac{S_{xy}}{S_x S_y}$$

Orduan:

$$r^2 = \frac{S_{xy}^2}{S_x^2 S_y^2} = \frac{S_{xy}}{S_x^2} \cdot \frac{S_{xy}}{S_y^2} = b_{yx} \cdot b_{xy}$$

bi erregresio-koefizienteen biderkadura, koerlazio-koefizientearen karratua da.

Adibidea:

X aldagaia lantegi batean egindako lan-orduak izanik eta Y aldagaia produzitu diren unitateak, kalkula dezagun koerlazio-koefizientea eta bi erregresio-zuzenak.

X	80	79	83	84	78	60	82	85	85	79	84
Y	300	302	315	330	300	250	300	340	315	330	310

(1) Y-ren X-ekiko erregresio-zuzena eta balio erreal eta balio teorikoen arteko desbidazioak.

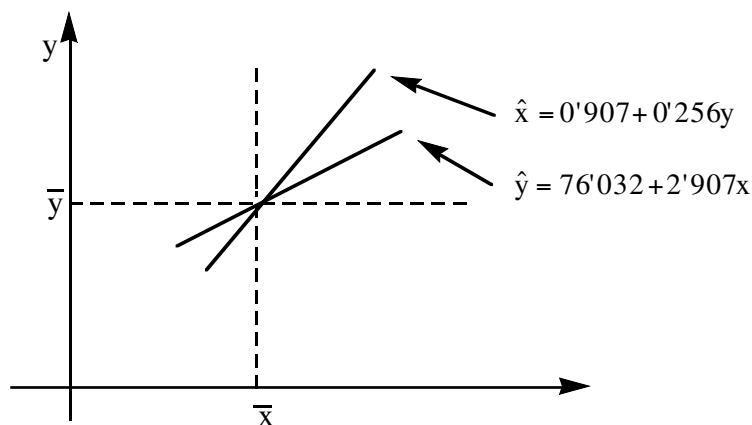
(2) X-ren Y-rekiko erregresio-zuzena eta desbidazioak.

Emaitzak hauek izan dira:

$$r = 0'863$$

$$\bar{x} = 79'9$$

$$\bar{y} = 308'36$$



Aurreko adibidean bi erregresio zuzenak egin baditugu ere, aldagaien arteko menpekotasuna alderantzizkoa da, hau da:

Bi galdera desberdin erantzuteko balio luketen zuzenak dira.

- 1) Zenbat lan-ordu beharko lirateke produzitutako unitateak kopuru jakina izateko?
- 2) Zein izango litzateke produzitutako unitateen kopurua egindako lan-orduak kopuru jakina bada?

VI.2.2.1.1. *Koerlazio-koefiziente lineala, hondar-bariantza eta mugatze-koefizientea*

Datu estatistikoen multzoan zuzen bat doituta dagoela suposatzen badugu, bi aldagaien arteko koerlazio-koefizientea erabiltzeak zentzu asko dauka, aldagaien menpekotasunaren maila erakus diezagun.

Froga dezagun jakina den erlazio hau: $-1 \leq r \leq +1$

Honela, m eta n bi zenbaki erreal izanik, har dezagun

$[m(x_i - \bar{x}) + n(y_j - \bar{y})]^2$ expresio arbitrario honen batezbesteko balioa

$$\begin{aligned}
& \sum_{ij} \left[m(x_i - \bar{x}) + (y_j - \bar{y}) \right]^2 f_{ij} = \\
& = m^2 \sum_{ij} (x_i - \bar{x})^2 f_{ij} + n^2 \sum_{ij} (y_j - \bar{y})^2 f_{ij} + 2mn \sum_{ij} (x_i - \bar{x})(y_j - \bar{y}) f_{ij} = \\
& = m^2 S_x^2 + n^2 S_y^2 + 2mn S_{xy} \geq 0
\end{aligned}$$

Egin dezagun: $m = S_y^2$ eta $n = -S_{xy}$

Hots: $S_y^4 S_x^2 + S_{xy}^2 S_y^2 - 2 S_y^2 S_{xy}^2 \geq 0$

$$S_y^2 S_x^2 - S_{xy}^2 \geq 0 ; \quad S_{xy}^2 \leq S_y^2 S_x^2$$

$$\frac{S_{xy}^2}{S_y^2 S_x^2} \leq 1 ; \quad r^2 \leq 1$$

$-1 \leq r \leq 1$

Balio teoriko eta balio errealeen arteko desbidazioak (hondar-desbidazioak) kontsideratzen baditugu, beste garrantzizko neurri bat aztertuko dugu: **Hondar-bariantza** deritzona.

Desbidazio karratuen batezbesteko aritmetikoa hondar-bariantza deritzona da.

Aldagaiak zentratuak kontsideratuz, hau da, $\bar{x} = \bar{y} = 0$ Y-ren X-ekiko erregresio-zuzena hauxe izango da: $y - \bar{y} = b(x - \bar{x}) \Rightarrow y = bx$, hots, jatorritik pasatzen den zuzen bat.

Kalkula dezagun, bada, S_e hondar-bariantza.

$$\begin{aligned}
S_e^2 &= \sum_{ij} (y_i - bx_i)^2 f_{ij} = \sum_{ij} (y_i^2 - 2bx_i y_i + b^2 x_i^2)^2 f_{ij} = \\
&= S_y^2 - 2bS_{xy} + b^2 S_x^2 = S_y^2 - 2 \frac{S_{xy}^2}{S_x^2} + \frac{S_{xy}^2}{S_x^4} S_x^2 = \\
&= S_y^2 - \frac{S_{xy}^2}{S_x^2} = S_y^2 \left(1 - \frac{S_{xy}^2}{S_x^2 S_y^2} \right) = S_y^2 (1 - r^2) \geq 0
\end{aligned} \tag{1}$$

(1) $\bar{x}$ eta $\bar{y}$ zero direla kontuan izanik.

Ikusten dugunez $S_e^2 = S_y^2(1 - r^2)$ da eta berdintasun honetatik ondoko bi proportzio osagarriak lortzen ditugu, S_e^2/S_y^2 eta r^2 hain zuzen.

Hots:

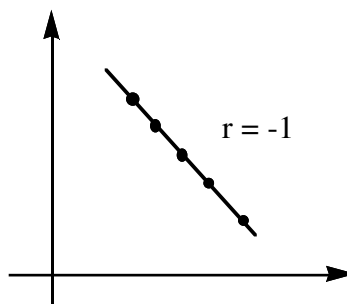
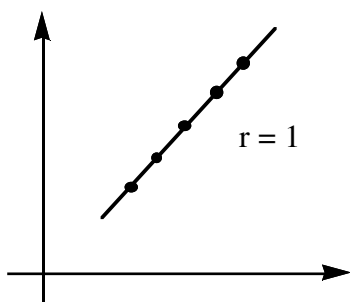
$$\frac{S_e^2}{S_y^2} = 1 - r^2$$

$$r^2 = 1 - \frac{S_e^2}{S_y^2}$$

$r^2 \leq 1$ da eta hondar-bariantza Y aldagaiaren bariantza baino txikiagoa da beti.

Bereziki $S_e^2 = 0$ baldin bada, batugai bakoitzak zero balio du eta (x_i, y_i) puntu guztiak zuzenaren gainean daude, bestalde, Y aldagaiaren bariantza guztiz azalduta geratzen da.

Grafikoki:



Puntuak zenbat eta zuzenetik gutxiago desbidatu, S_e^2 txikiago izango da eta r^2 1 baliora hurbilduko da, hau da, r^2 erlazio lineal batetara hurbiltzen direneko maila ematen digu, horregatik r^2 horri **mugatze-koefizientea** deritzo.

Beste modu batetara esateko:

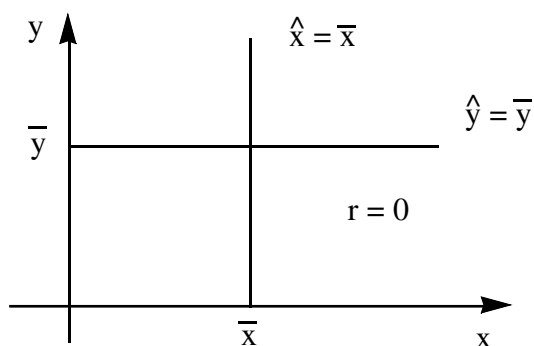
$r^2 = 1$ baldin bada, Y aldagaiaren auresana ahal den gehiena hobetu dugu erregresioaren analisiaz.

100 r^2 kontsideratuz, auresana ehunekotan zenbat hobetzen den jakingo dugu, erlazio lineala suposatzen ez genuen situazio batekin konparatuz.

Adibidez, $r^2 = 0.64$ baldin bada, Y -ren auresana, X -aren aldaketak kontuan izanik, erregresio-zuzenaren bidez %64ean hobetu dugula esango da.

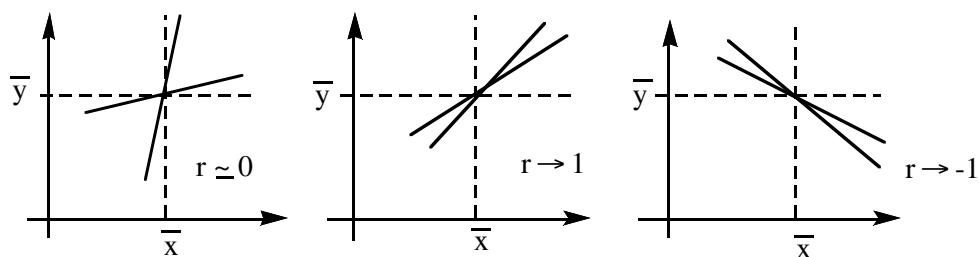
Kasu berezi batetan, $r^2 = 0$ baldin bada, situazioa ezertxo ere ez da hobetu erregresioaren bidez; aldagaiak koerlazio gabeak direla esango da eta erregresio-zuzenak ardatzekiko paraleloak dira.

Batezbesteko balioak independenteak dira eta bazter-batezbestekoen gain (zeinekin kointziditzen baitute) ez dute auresana hobetzen.



Halako erregresio-zuzenak guztiz sakabanatuak dauden datuei dagozkien eta, zerbait auresan behar badugu, zuhurrena batezbestekoa ematea da.

$r^2 > 0$ baina txikia baldin bada, bi erregresio-zuzenen arteko angelua 90° -ra hurbilduko da,



Aldiz, $r^2 > 0$ eta 1 baliora hurbiltzen bada, angelua oso txikia izango da. Azkenez, $r^2 = 1$ baldin bada, angelua 0° da eta bi zuzenak bat egiten dira. Ikus aurreneko orrialdeko irudiak.

Analisi estatistiko hau egitean, aldagaiek koerlazio gabeak ($r = 0$) diren ala ez esan dezakegu eta baita joera bat noiz dagoen ere, baina batzutan gehiagorik ez.

Erlazio linealen batetan pentsatzen hasteko, r koerlazio-koefizientearen balioa 0'70-tik gorakoa izatea eskatzen da (hau da $r^2 = 0'49$) eta askok ere gehiago eskatzen du.

VI.2.2.1.2. Erregresio lineal bilakatzen diren beste bi kasu

Puntu-hodeia kontsideratuz, erlazio potentziala $y = ax^b$ edo esponentziala $y = a e^{bx}$ suposatzen badugu, logaritmoak hartuz, aldagai transformatu batzuren erlazio lineal bilakatzen dira.

Hau da: $\log y = \log a + b \log x$
eta,

$$\log y = \log a + bx$$

VI.2.2.2. Karratu txikiaren erregresio parabolikoa

Puntu-hodeia kurba batetan kondentsatzen bada, funtzio polinomiko batetara doitzen da normalki.

$$\hat{y} = b_0 + b_1 x + b_2 x^2 + \dots + b_k x^k$$

$b_0, b_1, b_2, \dots, b_k$ parametroak lortu behar ditugu, desbidazio karratuen batura, minimoa izan dadin.

Hots:

$$E = \sum_{ij} e^2 f_{ij} = \sum_{ij} (y_j - \hat{y}_j)^2 f_{ij}$$

Bigarren graduko polinomioa (parabola) kontsideratuz:

$$E = \sum_{ij} (y_j - b_0 - b_1 x_i - b_2 x_i^2)^2 f_{ij}$$

txikiena izateko baldintzaz $\hat{y} = b_0 + b_1 x + b_2 x^2$ parabolako b_0, b_1, b_2 balioak mugatuko ditugu.

Lehen egin genuen bezala, deribatuak zero egin behar ditugu, E minimoa izateko baldintza beharrezkoa baita.

$$\frac{\partial E}{\partial b_0} = -2 \sum_{ij} (y_j - b_0 - b_1 x_i - b_2 x_i^2) f_{ij} = 0$$

$$\frac{\partial E}{\partial b_1} = -2 \sum_{ij} (y_j - b_0 - b_1 x_i - b_2 x_i^2) x_i f_{ij} = 0$$

$$\frac{\partial E}{\partial b_2} = -2 \sum_{ij} (y_j - b_0 - b_1 x_i - b_2 x_i^2) x_i^2 f_{ij} = 0$$

$$\sum_{ij} y_j f_{ij} = \sum_{ij} (b_0 + b_1 x_i + b_2 x_i^2) f_{ij}$$

$$\sum_{ij} y_j x_i f_{ij} = \sum_{ij} (b_0 + b_1 x_i + b_2 x_i^2) x_i f_{ij}$$

$$\sum_{ij} y_j x_i^2 f_{ij} = \sum_{ij} (b_0 + b_1 x_i + b_2 x_i^2) x_i^2 f_{ij}$$

Hau da:

$$\bar{y} = b_0 + b_1 \bar{x} + b_2 a_{20}$$

$$a_{11} = b_0 \bar{x} + b_1 a_{20} + b_2 a_{30}$$

$$a_{21} = b_0 a_{20} + b_1 a_{30} + b_2 a_{40}$$

EKUAZIO NORMALAK

Datuak banan banan hartzen badiugu ($n = 1$ beti izanik)

$$\frac{\sum y_j}{N} = b_0 + b_1 \frac{\sum x_i}{N} + b_2 \frac{\sum x_i^2}{N}$$

$$\frac{\sum x_i y_j}{N} = b_0 \frac{\sum x_i}{N} + b_1 \frac{\sum x_i^2}{N} + b_2 \frac{\sum x_i^3}{N}$$

$$\frac{\sum x_i^2 y_j}{N} = b_0 \frac{\sum x_i^2}{N} + b_1 \frac{\sum x_i^3}{N} + b_2 \frac{\sum x_i^4}{N}$$

$$\left. \begin{aligned} \sum y_j &= Nb_0 + b_1 \sum x_i + b_2 \sum x_i^2 \\ \sum x_i y_j &= b_0 \sum x_i + b_1 \sum x_i^2 + b_2 \sum x_i^3 \\ \sum x_i^2 y_j &= b_0 \sum x_i^2 + b_1 \sum x_i^3 + b_2 \sum x_i^4 \end{aligned} \right\} (1)$$

Ikusten denez, (1) ekuazioak erraz gogoratuko ditugu, zeren lehenengoa, batukaria sartuz eta b_0 koefizientea N -z biderkatuz, parabolarena baita eta beste biak, x_i eta x_i -z biderkatuz, lehenengo ekuaziotik ateratzen ditugu.

Adibidea:

Ondoko datuak kontsideratuz eta X aldagaia independentetzat hartuz, doikuntza lineala egin dezagun. Ona al da doikuntza hau? Eta parabola bat doitzen badugu?

x_i	y_i	x_i^2	y_i^2	$x_i y_i$	x_i^3	x_i^4	$x_i^2 y_i$
-2	5	4	25	-10	-8	16	20
-1	4	1	16	-4	-1	1	4
0	2	0	4	0	0	0	0
1	4	1	16	4	1	1	4
2	5	4	25	10	8	16	20
0	20	10	86	0	0	34	48

$$\left. \begin{aligned} \bar{y} &= a + b\bar{x} \\ a_{11} &= a\bar{x} + ba_{20} \end{aligned} \right\} \begin{aligned} a &= 4 \\ b &= 0 \end{aligned}$$

$$\hat{y} = a + bx \quad ; \quad \hat{y} = 4 \quad S_{xy} = r = r^2 = 0$$

Ikusten denez, doitzen dugun zuzena, OX ardatzarekin paraleloa da eta edozein x -entzat y -ren balioa 4 da beti. Aldagaiak koerlazio gabeak dira, hau da, independentzia lineala agertzen zaigun kasu baten aurrean gaude. Bestaldetik, X aldagaiak Y -ren aldaketak ezertxo ere ez dituela azaltzen esan dezakegu.

Parabola doituz:

$$\left. \begin{aligned} \bar{y} &= b_0 + b_1 \bar{x} + b_2 a_{20} \\ a_{11} &= b_0 \bar{x} + b_1 a_{20} + b_2 a_{30} \\ a_{21} &= b_0 a_{20} + b_1 a_{30} + b_2 a_{40} \end{aligned} \right\}$$

$$b_1 = 0$$

$$b_0 = 20/7$$

$$b_2 = 4/7$$

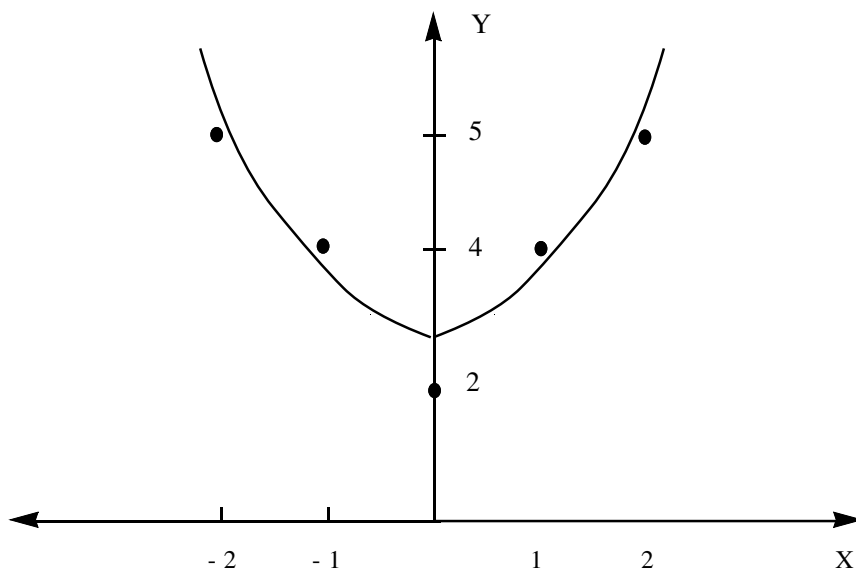
Konkretuki:

$$\left. \begin{aligned} 4 &= b_0 + 2b_2 \\ 0 &= 2b_1 \\ 40 &= 10b_0 + 34b_2 \end{aligned} \right\}$$

eta $y = b_0 + b_1 x + b_2 x^2$ parabolari ordezkatuz

$$\hat{y} = 20/7 + 4/7 x^2$$

Grafikoki:



Puntu-hodeia eta doitutako parabola

***VII. ALDAGAI ESTATISTIKO n -KOITZEN
MATRIZE-ESTATISTIKOAK***

VII.1. DATU-MATRIZEAK

*VII.2. BATEZBESTEKO-BEKTOREA:
PROPIETATEAK*

VII.3. KOBARIANTZ-MATRIZEA

VII.3.1 Kobariantz matrizearen propietateak

VII.4. KOERLAZIO-MATRIZEA

VII.4.1. Koerlazio-matrizearen propietateak

VII.1. DATU-MATRIZEAK

Azkeneko ikasgaietan bi aldagaien arteko erlazioak aztertzeke tresna batzuk eman ditugu, baina hemendik aurrera gure ikasketako objektuak, hiru, lau,..... n aldagai izango dira eta matrizialki ikasketak egitea komenigarria da.

Has gaitezen, bada, matrizeak definitzen: $X_1, X_2, \dots, X_n$ n aldagai kontsideratuz $X = [X_1, X_2 \dots X_n]$ beraiek osatzen duten zutabe-bektorea da:

Hots:

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

Aldagai horiek m indibiduen gain neurketa konkretu batzuk edo balio ohartu batzuk hartzen dute, hauexek $\mathbf{X}$ datu-matrizea osatzen dutelarik.

Hau da:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix} = [x_{ij}]$$

j-garren aldagaiaren batezbesteko aritmetikoa eta desbidazio tipikoa

$$\bar{x}_j = \frac{\sum_{i=1}^m x_{ij}}{m} \quad \text{eta} \quad S_j = \sqrt{\frac{1}{m} \sum_{i=1}^m (x_{ij} - \bar{x}_j)^2} \quad \text{izanik}$$

Datu-matrize zentratua eta datu-matrize tipifikatua edo standardizatua idatz ditzakegu, **C** eta **T** matrizeak hurrenez hurren.

$$\mathbf{C} = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \dots & x_{1n} - \bar{x}_n \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \dots & x_{2n} - \bar{x}_n \\ \dots & \dots & \dots & \dots \\ x_{m1} - \bar{x}_1 & x_{m2} - \bar{x}_2 & \dots & x_{mn} - \bar{x}_n \end{bmatrix} = [x_{ij} - \bar{x}_j]$$

$$\mathbf{T} = [x_{ij} - \bar{x}_j / S_j]$$

Datu-matrizeak, **X** alegia, eta aldagai bikoitz baten maiztasun-banaketaren matrizeak ez dute zerikusirik, biak matrizeak direla baizik; lehenengo kasuan x_{ij} elementua i . indibiduoari dagokion balioa j . aldagaiarekiko da eta bigarreanean, ordea, n_{ij} (x_i , y_j) balio bikotearen maiztasun absolutua.

VII.2. BATEZBESTEKO-BEKTOREA: PROPIETATEAK

$\bar{\mathbf{x}}$ -ikurraz adieraziko dugu eta aldagaien batezbestekoen zutabe-bektorea da.

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_n \end{bmatrix} = [\bar{x}_j]$$

Aldagai bakun baten batezbestekoaren linealtasun-propietateak, batezbesteko-bektorerako zabaltzen dira berehala.

1. Propietatea

$$\left. \begin{array}{l} \mathbf{Y} = \mathbf{AX} + \mathbf{b} \\ \mathbf{Y} = \mathbf{X} \mathbf{A}^T + \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \mathbf{b}^T \end{array} \right\} \bar{\mathbf{y}} = \mathbf{A} \bar{\mathbf{x}} + \mathbf{b}$$

non: $\mathbf{A}$ matrizea $\mathbb{R}^n \rightarrow \mathbb{R}^n$ aplikazio lineal baten matrizea eta $\mathbf{b}$ lekualdatze-edo traslazio-bektorea dira.

Honen ondorioz zera esan dezakegu:

Aldagaien jatorri-aldatzeak eta unitate- edo eskala-aldatzeak batezbestekoetan daukaten eragina, zera da: beraiek ere linealki aldatzea. Hau da, batezbestekoetan transformazio lineala daukagu.

$$\mathbf{Z} = \mathbf{X} + \mathbf{Y} \Rightarrow \bar{\mathbf{z}} = \bar{\mathbf{x}} + \bar{\mathbf{y}}$$

Aldagaien bektore bat beste bi aldagaien bektoreen batura denean baturaren batezbesteko-bektorea bi batugaien batezbesteko-bektoreen batura da.

VII.3. KOBARIANTZ-MATRIZEA

$(X_1, X_2 \dots X_n)$ aldagai n -koitz baten bigarren ordenako momentu zentratuak bi eratakoak izan daitezke.

Aldagai bakun baten bazter-momentuak, adibidez:

$$m_{002..0} = \frac{1}{m} \sum_{i=1}^m (x_{i3} - \bar{x}_3)^2 = S_{x_3}^2$$

Aldagai bikoitz baten bazter-momentuak, adibidez:

$$m_{101..0} = \frac{1}{m} \sum_{i=1}^m (x_{i1} - \bar{x}_1) (x_{i3} - \bar{x}_3) = S_{x_1 x_3}$$

Hots bariantzak edo kobariantzak dira.

Aldagai baten bariantza aldagaiak berekiko duen kobariantza bezala uler dezakegu, baita definizioz $S_{ij} = S_{ji}$ da.

Guzti hau kontutan hartuz aldagai n -koitzaren 2. ordenako momentu zentratu guztiek matrize simetriko bat osatzen dute, kobariantz matrizea hain zuzen.

$\mathbf{L}$ ikurraz sinbolikoki adierazten dugu eta matrizialki honela lortzen da:

$$\mathbf{L} = \frac{1}{m} \mathbf{C}^T \mathbf{C} = \begin{bmatrix} S_1^2 & S_{12} & \dots & S_{1n} \\ S_{12} & S_2^2 & \dots & S_{2n} \\ \dots & \dots & \dots & \dots \\ S_{1n} & S_{2n} & \dots & S_n^2 \end{bmatrix} = \begin{bmatrix} l_{11} & l_{12} & \dots & l_{1n} \\ l_{12} & l_{22} & \dots & l_{2n} \\ \dots & \dots & \dots & \dots \\ l_{1n} & l_{2n} & \dots & l_{nn} \end{bmatrix}$$

halaber: $\mathbf{L} = \mathbf{X}^T \mathbf{X}$ $\mathbf{X} = \left[\frac{x_{ij} - \bar{x}_j}{\sqrt{m}} \right]$ matrizea izanik

l_{ii} , l_{ij} ikurraz ere elementuak adierazten direlarik. $\mathbf{X}$ datu-matrizea osatzen duten aldagaiak zentratuak daudenean kobariantz matrizea ondokoa da.

$$\mathbf{L} = \frac{1}{m} \mathbf{X}^T \mathbf{X}$$

VII.3.1. Kobariantz matrizearen propietateak

1. propietatea

$$\left. \begin{array}{l} \mathbf{Y} = \mathbf{A} \mathbf{X} + \mathbf{b} \\ \mathbf{Y} = \mathbf{X} \mathbf{A}^T + \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \mathbf{b}^T \end{array} \right\} \Rightarrow \mathbf{L}(\mathbf{Y}) = \mathbf{A} \mathbf{L}(\mathbf{X}) \mathbf{A}^T$$

Ikusten dugunez eskala-aldatzeak kobariantz matrizean dauka eragina, baina ez jatorri-aldatzeak.

Oharra: Gogora dezagun $\mathbf{Y} = \mathbf{a}\mathbf{X} + \mathbf{b}$ izanik, bariantzen arteko erlazioa zera zela:

$$S_y^2 = a^2 S_x^2$$

2. propietatea:

Kobariantz matrizeak definitu positiboak edo erdidefinitu positiboak dira.

Matrize-algebran honek suposatzen du azpimatrizen diagonal guztiak definitu positiboak edo erdidefinitu positiboak direla.

VII.4. KOERLAZIO-MATRIZEA

Aldagai tipifikatuen arteko kobariantz matrizea da.
Hau da:

$$\mathbf{L}(\mathbf{T}) = \frac{1}{m} \mathbf{T}^T \mathbf{T} = \frac{1}{m} \left[\frac{x_{ij} - \bar{x}}{S_j} \right]^T \left[\frac{x_{ij} - \bar{x}_j}{S_j} \right] = \mathbf{R}$$

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \dots & r_{1n} \\ r_{12} & 1 & \dots & r_{2n} \\ \dots & \dots & \dots & \dots \\ r_{1n} & r_{2n} & \dots & 1 \end{bmatrix}$$

halaber: $\mathbf{R} = \mathbf{X}^T \mathbf{X}$ $\mathbf{X} = \left[\frac{x_{ij} - \bar{x}_j}{S_j \sqrt{m}} \right]$ izanik

$\mathbf{X}$ datu-matrizea eta $\mathbf{T}$ datu-matrize tipifikatuen arteko erlazioa hauxe da:

$$\mathbf{X} = \mathbf{T} \mathbf{D}_S^T + \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} \bar{\mathbf{x}}^T$$

Halaber, $\mathbf{X}$ aldagaien bektorea eta $\mathbf{T}$ aldagai tipifikatuen artekoa:

$$\mathbf{X} = \mathbf{D}_S \mathbf{T} + \bar{\mathbf{x}}$$

Espresso honek adierazten duena zera da: Transformazio lineal diagonal bat dugula ($\mathbb{R}^n$ espazio bektorialean definitua); transformazio hau, $\mathbf{D}_S$ matrize diagonalak definitu baitu.

$$\mathbf{D}_S = \begin{bmatrix} S_1 & 0 & \dots & 0 \\ 0 & S_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & S_n \end{bmatrix}$$

Dakusagun $\mathbf{L}(X)$ eta $\mathbf{L}(T)$ kobariantz matrizeen arteko erlazioa:

$$\mathbf{L}(X) = \mathbf{D}_S \mathbf{L}(T) \mathbf{D}_S^t$$

Hots:

$$\mathbf{L} = \mathbf{D}_S \mathbf{R} \mathbf{D}_S^t$$

Kobariantza eta koerlazio-matrizeen arteko erlazioa.

Estentsiboki:

$$\mathbf{L} = \begin{bmatrix} S_1 & 0 & \dots & 0 \\ 0 & S_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & S_n \end{bmatrix} \begin{bmatrix} 1 & r_{12} & \dots & r_{1n} \\ r_{12} & 1 & \dots & r_{2n} \\ \dots & \dots & \dots & \dots \\ r_{1n} & r_{2n} & \dots & 1 \end{bmatrix} \begin{bmatrix} S_1 & 0 & \dots & 0 \\ 0 & S_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & S_n \end{bmatrix}^T$$

Dakusagun azkenik matrize hauen determinanteen eta adjuntuen arteko erlazioak.

$\mathbf{L}$ eta $\mathbf{R}$ matrizeen arteko elementuetan $S_{ij} = S_i S_j r_{ij}$ erlazioa betetzen dela ikusirik, ondoko erlazioak, erraz ateratzen ditugu.

$$|\mathbf{L}| = S_1^2 S_2^2 \dots S_n^2 |\mathbf{R}|$$

$$\mathbf{L}_{11} = S_2^2 S_3^2 \dots S_n^2 \mathbf{R}_{11}$$

$$\mathbf{L}_{12} = S_1 S_2 S_3^2 \dots S_n^2 \mathbf{R}_{12}$$

$$|\mathbf{L}| \text{ eta } |\mathbf{R}|, \mathbf{L} \text{ eta } \mathbf{R}$$

matrizeen determinanteak izanik

$$\mathbf{L}_{11} = l_{11} \text{ elementuaren adjuntua}$$

$$\mathbf{R}_{11} = r_{11} \text{ elementuaren adjuntua}$$

$$\mathbf{L}_{12} = l_{12} \text{ elementuaren adjuntua}$$

$$\mathbf{R}_{12} = r_{12} \text{ elementuaren adjuntua}$$

VII.4.1. Koerlazio-matrizearen propietateak

Gogoratu behar dugu (V.3.2.) koerlazioak, aldagai tipifikatuen arteko kobariantzak direla.

Honela, koerlazio-matrizea kobariantz matrize-mota bat izatean kobariantz matrizeen propietateak ditu (3.1.).

Bestalde, koerlazio-matrizea, bi aldagaien arteko koerlazioa bezala (V.3) inbariantea da edozein transformazio linealen aurrean.

VIII. ERREGRESIO ANIZKOITZA

VIII.1. SARRERA

VIII.2. ERREGRESIO-PLANOA $\mathbb{R}^3$

VIII.3. ERREGRESIO-HIPERPLANOA $\mathbb{R}^n$ -n

*VIII.3.1. Karratu txikiaren erregresioaren
propietateak*

*VIII.3.2. Mugatze-koefizientea eta koerlazio-
koefiziente anizkoitza*

*VIII.4. "BETA" ERREGRESIO-KOEFIZIENTE
STANDARDIZATUAK*

VIII.1. SARRERA

Aurreko ikasgaietan magnitudeen bikoteak erlazionatzen hasiak ginen. Edozein erlazio sumatzen genuenean, magnitudeen mugimenduetan elkarrekiko aldaketa neurtzen saiatzen ginen.

Elkar-aldatze honek, batez ere, kobariantza deritzon momentuan eragina zuenez gero, hura momentu honexen bidez neurtzen genuen.

Baina, nola neur elkar-aldatze honen pisu erlatiboa, aldagai bakoitzaren aldaketaren tamainuarekiko? Horretarako, lortutako kobariantza bi aldagaien bariantzen batezbesteko geometrikoaz zatitzen genuen, Karl Pearson-ek asmatu zuen bezala: X eta Y -ren arteko koerlazio-koefizientea zen.

Koerlazio hau ikusi eta gero, gure ustez, funtzio-familiarik egokiena hartuz eta kontrolagaitz eta aldakor diren kausa askorengatik desdoitzeak zirela aurrez onartuz, doikuntza egiten saiatzen ginen.

Aukeratutako funtzio-familia zuzenen multzokoa zenean, bi aldagaiek elkarren artean zuten mugapenaren erro karratua koerlazio-koefizientea zen eta hondar-bariantza bat geratzen zen.

Nahiz eta mugatze-koefizienteak 1 balioari hurbiltzen zitzaion balio bat hartu, ezin genuen horrengatik aldagaien artean kausa-efektu erlazio bat baieztatu, hau da, koerlazioa edukitzea kasualitatez izan daiteke zein argigarri diren beste kausa batzurengatik.

Baina, alderantziz, kausa-efektu erlazio bat denean, koerlazioa edukitzea probablea izango da; hau da, koerlazioak estatistikoki erlazioa frogatuko du.

Maiz, erlazio funtzionalak aztertzeko aldagai bakar bi gutxiegi dira, errealtatean aztertzen diren arazoak ez baitira normalki bi aldagairen artekoak. Izan ere, gehienetan, bi baino gehiago daude erlazionaturik; eta beti, nahiz eta asko kontsideratu aldagai baten kausatzat, errealtatearen sinplifikazio bat egiten ari gara.

Honela, haur baten pisua bere altueraren, adinaren, bular inguruaren etabarren menpe dagoela esan dezakegu; bai eta, halaber, lur eremu baten emankortasuna bertako euriaren, hazi-kantitatearen, ongariaren etabarren menpe dagoela.

Ikasgai honetan segituko dugun bidea, bereiztasun batzurekin, VI. ikasgaiarekiko baliokidea izango da.

Lehendabizi, erregresio anizkoitzaren eredua planteatuko dugu. Aldagaiak bi baino gehiago izango dira, baina erregresio anizkoitz lineala egingo dugu bakarrik.

Gero, azaldu behar den aldagaiaren bariantzaz, aldagai independenteek azaltzen duten portzentaia neurtuko dugu eta ondoren, koerlazio-koefiziente anizkoitza emango dugu.

VIII.2. ERREGRESIO-PLANOA $\mathbb{R}^3$ -n

Hiru aldagai kontsideratzen baditugu eta erregresio lineala egin nahian bat menpekotzat hartzen badugu, $\mathbb{R}^3$ espazio hirudimentsionalean kokatutako puntu-hodeian, hobeki doitzen den plano aukeratzea izango da arazoa. Ondoko $\mathbf{X}$ datu-matrizean, m indibiduen gain ohartutako balioak ditugu X_1, X_2, X_3 aldagaietarako.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ \dots\dots\dots \\ x_{i1} & x_{i2} & x_{i3} \\ \dots\dots\dots \\ x_{m1} & x_{m2} & x_{m3} \end{bmatrix}$$

Orduan: $\hat{X}_1 = b_{12}X_2 + b_{13}X_3 + a$ plano finkatu behar dugu, desbidazio karratuen batura, batezbeste, txikiena izan dadin.

Hots:

$$E = \frac{1}{m} \sum_i (x_{i1} - b_{12} x_{i2} - b_{13} x_{i3} - a)^2$$

minimoa a (jatorriarekiko ordenatua) eta b_{12}, b_{13} (erregresio-koefizienteak) hiru alditan bilatuko ditugu E txikiena egin dezaten.

- Erregresio-planoa batezbesteko-bektoretik, hau da, hodeiaren grabitate-zentrutik pasatzen dela frogatuko dugu.
- Aldagai zentratutarako koefizienteak lortuko ditugu.
- Erregresio-planoaren adierazpen orokorra idatziko dugu.

ERREGRESIO-PLANOA ($\bar{x}_1$ $\bar{x}_2$ $\bar{x}_3$) BEKTORETIK PASATZEN DA:

Minimoaren baldintza beharrezko bat, zera da:

$$\frac{\partial E}{\partial a} = -2 \frac{1}{m} \sum_i (x_{i1} - b_{12} x_{i2} - b_{13} x_{i3} - a) = 0$$

$$\bar{x}_1 - b_{12} \bar{x}_2 - b_{13} \bar{x}_3 - a = 0$$

$$\boxed{\bar{x}_1 = b_{12} \bar{x}_2 + b_{13} \bar{x}_3 + a}$$

ALDAGAI ZENTRATUETARAKO ERREGRESIO-KOEFIZIENTEAK:

Dakigunez, edozein aldagairi bere batezbestekoa kenduz gero, aldagai zentratu bat lortzen dugu.

Aldagai zentratuaren batezbestekoa zero da eta une honetan $\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = 0$ kontsideratuko dugu, frogaketa errezago egin dezagun.

Erregresio-planoa, bada, (0, 0, 0) puntutik pasatzen da eta orduan $a = 0$ izanez:

$$E = \frac{1}{m} \sum_i (x_{i1} - b_{12} x_{i2} - b_{13} x_{i3})^2 \text{ minimoa}$$

Baldintza beharrezkoak idatziz:

$$\frac{\partial E}{\partial b_{12}} = -2 \frac{1}{m} \sum_i (x_{i1} - b_{12} x_{i2} - b_{13} x_{i3}) x_{i2} = 0$$

$$\frac{\partial E}{\partial b_{13}} = -2 \frac{1}{m} \sum_i (x_{i1} - b_{12} x_{i2} - b_{13} x_{i3}) x_{i3} = 0$$

Hots:

$$\left. \begin{aligned} S_{12} - b_{12}S_2^2 - b_{13}S_{23} &= 0 \\ S_{13} - b_{12}S_{23} - b_{13}S_3^2 &= 0 \end{aligned} \right\}$$

Matrizialki ordenatuz:

$$\begin{bmatrix} S_2^2 & S_{23} \\ S_{23} & S_3^2 \end{bmatrix} \begin{bmatrix} b_{12} \\ b_{13} \end{bmatrix} = \begin{bmatrix} S_{12} \\ S_{13} \end{bmatrix}$$

Eta azkenez, Cramer-en erregelaren bidez, b_{12} b_{13} askatuko ditugu.

$$b_{12} = \frac{\begin{vmatrix} S_{12} & S_{23} \\ S_{13} & S_3^2 \end{vmatrix}}{\begin{vmatrix} S_2^2 & S_{23} \\ S_{23} & S_3^2 \end{vmatrix}} \quad b_{13} = \frac{\begin{vmatrix} S_2^2 & S_{12} \\ S_{23} & S_{13} \end{vmatrix}}{\begin{vmatrix} S_2^2 & S_{23} \\ S_{23} & S_3^2 \end{vmatrix}}$$

Soluzioan aurkitzen diren determinanteak kobariantz matritzekoak direnez, sinpleki aurkez daitezke erregresio-koefizienteak.

$$\begin{bmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{bmatrix} = \begin{bmatrix} S_1^2 & S_{12} & S_{13} \\ S_{12} & S_2^2 & S_{23} \\ S_{13} & S_{23} & S_3^2 \end{bmatrix}$$

Hots:

$$b_{12} = \frac{-L_{12}}{L_{11}} \quad b_{13} = \frac{-L_{13}}{L_{11}}$$

L_{ij} , l_{ij} elementuaren adjuntua izanik.

Eta ekuazioan ordezkatzuz:

$$\hat{x}_1 = -\frac{L_{12}}{L_{11}} x_2 - \frac{L_{13}}{L_{11}} x_3$$

Hots: $L_{11}\hat{x}_1 + L_{12}x_2 + L_{13}x_3 = 0$

Hori da aldagai zentratuen kasuan daukagun soluzioa.

ERREGRESIO-PLANOAREN ADIERAZPEN OROKORRA

Lehenengo ekuazioan X_1 , X_2 , X_3 zentratuak baziren, ondokoan beraien ordeztuz $X_1 - \bar{x}_1$, $X_2 - \bar{x}_2$, $X_3 - \bar{x}_3$ ipintzen baditugu, X_1 , X_2 , X_3 berri hauek zentratugabe izango dira eta planoaren ekuazioa ondokoa izango da:

$$(\hat{x}_1 - \bar{x}_1) = -\frac{L_{12}}{L_{11}} (x_2 - \bar{x}_2) - \frac{L_{13}}{L_{11}} (x_3 - \bar{x}_3)$$

Adierazpen hau kobariantz matrizeaz eta batezbesteko-bektoreaz idatzi dugu.

$$(\hat{x}_1 - \bar{x}_1) = -\frac{S_1 R_{12}}{S_2 R_{11}} (x_2 - \bar{x}_2) - \frac{S_1 R_{13}}{S_3 R_{11}} (x_3 - \bar{x}_3)$$

Ondoko ekuazioa, ordeztuz, koerlazio-matrizeaz eta batezbesteko-bektoreaz.

Ekuazio batetik bestea ateratzeko kobariantz matrizearen eta koerlazio-matrizearen adjuntuen arteko erlazioak gogoratu behar ditugu (ikus VII.4).

VIII.3. ERREGRESIO-HIPERPLANOA $\mathbf{R}^{n-n}$

$(X_1, X_2, \dots, X_i, \dots, X_n)$ aldagaietatik X_i aldagaia beste $n-1$ aldagaien menpean kontsideratzen badugu, frogaketa baliokide baten bidez, hiperplanoaren ekuazioa lortzen dugu.

Hots:

$$(\hat{x}_i - \bar{x}_i) = -\frac{L_{i1}}{L_{ii}} (x_1 - \bar{x}_1) - \dots - \frac{L_{ik}}{L_{ii}} (x_k - \bar{x}_k) - \dots - \frac{L_{in}}{L_{ii}} (x_n - \bar{x}_n)$$

$$b_{ik} = - \frac{L_{ik}}{L_{ii}} \text{ erregresio-koefizienteak izanik.}$$

Erregresio-hiperplanoa ondoko beste bi ekuazioez ere adieraz daiteke:

$$\begin{aligned} L_{i1}(x_1 - \bar{x}_1) + \dots + L_{ik}(\hat{x}_i - \bar{x}_i) + \dots + L_{ik}(x_k - \bar{x}_k) + \dots + \\ + L_{in}(x_n - \bar{x}_n) = 0 \end{aligned}$$

edo

$$\begin{aligned} \frac{R_{i1}}{S_1}(x_1 - \bar{x}_1) + \dots + \frac{R_{ii}}{S_i}(\hat{x}_i - \bar{x}_i) + \dots + \frac{R_{ik}}{S_k}(x_k - \bar{x}_k) + \dots + \\ + \frac{R_{in}}{S_n}(x_n - \bar{x}_n) = 0 \end{aligned}$$

Dakigunez, $\mathbb{R}^2 - n$:

$$L = \begin{bmatrix} l_{xx} & l_{xy} \\ l_{xy} & l_{yy} \end{bmatrix} = \begin{bmatrix} S_x^2 & S_{xy} \\ S_{xy} & S_y^2 \end{bmatrix}$$

$$\text{eta } b = -\frac{S_{xy}}{S_x^2} = -\frac{L_{12}}{L_{22}}$$

$$\text{hots, erregresio-zuzenak } S_x^2(\hat{y} - \bar{y}) - S_{xy}(x - \bar{x}) = 0$$

VIII.3.1. Karratu txikiaren erregresioaren propietateak

Errore karratuen batezbestekoa deribatzean ateratzen ditugun ekuazioetan, inplizituki erregresioaren propietateak aurkitzen dira.

Hauekin batera, V.4. atalean ikusi genituen aldagai koerlazio gabeen bi propietateak orokortuko ditugu.

1. Aldagai estatistiko koerlazio gabeen baturaren bariantza, aldagaien bariantzen batura da.

Hots, $X_1, X_2, \dots, X_k$ binaka koerlazio gabeak badira, $S_{x_i x_j} = 0$
 $\forall i \neq j$

$$S_{(x_1 + x_2 + \dots + x_k)}^2 = S_{x_1}^2 + S_{x_2}^2 + \dots + S_{x_k}^2$$

2. Aldagai estatistiko bat beste batzuekiko koerlazio gabea baldin bada, beraien edozein konbinazio linealekiko koerlazio gabea da.

$X_1, (X_2, X_3, \dots, X_k)$ $S_{x_1 x_i} = 0 \quad \forall i \in \{2, \dots, k\}$ baldin bada
 $S_{x_1}, (ax_2 + bx_3 + \dots + kx_k) = 0 / a, b, \dots, k \in \mathbb{R}$

E adierazpenaren a-rekiko deribatua egitean:

$$\bar{e} = \frac{1}{m} \sum_i (x_{i1} - b_{12} x_{i2} - b_{13} x_{i3} - a) = 0$$

Hauxe da korolario gisa lortzen genuena:

I) ERREGRESIO-HONDARREN BATEZBESTEKO ARITMETIKOA ZERO DA

Dakigunez, balio erreal bakoitza, balio teoriko eta hondar edo errore batetan deskonposatzen da, eta, hortaz, batezbesteko aritmetikoaren propietateak aplikatzean:

$$x_1 = \hat{x}_1 + e \qquad \bar{x}_1 = \bar{\hat{x}}_1 + 0 \qquad \bar{\hat{x}}_1 = \bar{x}_1$$

II) BALIO ESTIMATU EDO TEORIKOEN BATEZBESTEKO ARITMETIKOA ETA BALIO ERREALENAK BERDINAK DIRA

I) eta II) propietateak lehenengo ekuazio deribatutik lortu ditugu korolario gisa.

Bi erregresio-koefizienteekiko deribatuak egitean:

$$\frac{1}{m} \sum_i (x_{i1} - b_{12} x_{i2} - b_{13} x_{i3} - a) x_{i2} = 0$$

$$\frac{1}{m} \sum_i (x_{i1} - b_{12} x_{i2} - b_{13} x_{i3} - a) x_{i3} = 0$$

hots,

$$\left. \begin{aligned} \frac{1}{m} \sum_i e x_{i2} &= 0 \\ \frac{1}{m} \sum_i e x_{i3} &= 0 \end{aligned} \right\} \Rightarrow \begin{aligned} S_{e, x_2} &= 0 \\ S_{e, x_3} &= 0 \end{aligned}$$

hau da, $\bar{e} = 0$ denez zera esan dezakegu:

III) HONDARRAK ETA EDOZEIN ALDAGAI INDEPENDENTE EDO ERREGRESORE ELKARREN ARTEAN KOERLAZIO GABEAK DIRA

$$S_{e, x_i} = 0 \quad x_i \in \{x_2, x_3\}$$

Aipatu dugun aldagai koerlazio gabeen 2. propietatea kontutan hartuz, zera idatz daiteke:

$$S_{e, \hat{x}_1} = 0 \quad x_1, x_i \text{ erregresoreen konbinazio lineala baita.}$$

IV) HONDARRAK ETA BALIO ESTIMATUAK ELKARREN ARTEAN KOERLAZIO GABEAK DIRA

Berririo $x_1 = \hat{x}_1 + e$ dela kontsideratuz eta aldagai koerlazio gabeen 1. propietatea kontutan hartuz, korolariorik bezala, bariantzaren deskonposaketa batukorra idatz dezakegu.

$$S_{\hat{x}_1}^2 + S_e^2 = S_{x_1}^2$$

V) BALIO ESTIMATUEN BARIANTZAREN (ERREGRESIOAN AZALDURIK DAGOEN BARIANTZA) ETA HONDAR-BARIANTZAREN BATURA, BALIO ERREALEN BARIANTZA DA

S_e^2 , hondar-bariantza deitutakoa, $\bar{e} = 0$ izatean, errore karratuen batezbestekoa da.

$$S_e^2 = \frac{1}{m} \sum_i (e - \bar{e})^2 = \frac{1}{m} \sum_i (x_{i1} - b_{12} x_{i2} - b_{13} x_{i3})^2$$

Hau da, karratu txikiaren metodoaz erregresioa egitean, hondar-bariantzarik txikiena bilatzen ari gara.

Orain arte ikusitako bost korolariorak, eskematikoki honela labur ditzakegu.

$$\begin{array}{lll} \frac{\partial E}{\partial a} = 0 & \rightarrow & \bar{e} = 0 \quad \rightarrow \quad \bar{\hat{x}}_1 = \bar{x}_1 \\ \frac{\partial E}{\partial b_{li}} = 0 & \rightarrow & S_{e, x_i} = 0 \quad \rightarrow \quad S_{e, \hat{x}_1} = 0 \\ & \rightarrow & S_{\hat{x}_1}^2 + S_e^2 = S_{x_1}^2 \end{array}$$

Azkenez, kalkulu matrizialaren bidez, hondar-bariantzaren adierazpen orokorra lortzen dugu.

$$S_e^2 = \frac{|\mathbf{L}|}{\mathbf{L}_{11}} = \frac{|\mathbf{R}|}{\mathbf{R}_{11}} S_{x_1}^2$$

non $|\mathbf{L}|$ eta $|\mathbf{R}|$, $\mathbf{L}$ eta $\mathbf{R}$ kobariantz eta koerlazio-matrizeen determinanteak baitira hurrenez hurren.

$\mathbb{R}^2$ -n aurreko ekuazioa honela konkretatzen da:

$$S_e^2 = \frac{S_x^2 S_y^2 - S_{xy}^2}{S_x^2} = S_y^2 - \frac{S_{xy}^2}{S_x^2} = S_y^2 (1 - r^2)$$

non: $S_x^2 S_y^2 - S_{xy}^2 = |L|$ eta $S_x^2 = L_{22}$ (hau da, S_y^2 -ren adjuntua) diren.

VIII.3.2. Mugatze-koefizientea eta koerlazio-koefiziente anizkoitza

Bariantzaren deskonposaketa batukorra ($S_{\hat{x}_1}^2 + S_e^2 = S_{x_1}^2$)

frogatu dugunez gero:

$$S_{\hat{x}_1}^2 = S_{x_1}^2 - \frac{|L|}{L_{11}} = 1 - \frac{|L|}{L_{11}} \quad \text{edo} \quad = S_{x_1}^2 - \frac{|R|}{R_{11}} S_{x_1}^2$$

Ondoko koefizienteak, azaldu behar den bariantzarekiko mugapenaren proportzioa ematen digu.

$$r_{1,2,3}^2 = \frac{S_{\hat{x}_1}^2}{S_{x_1}^2} = 1 - \frac{|L|}{l_{11}L_{11}} = 1 - \frac{|R|}{R_{11}}$$

Eta beronen erro karratuari, K. Pearson-en koefizientearen analogiaz, koerlazio-koefiziente anizkoitza deritzogu.

$$r_{1,2,3} = +\sqrt{1 - \frac{|L|}{l_{11}L_{11}}} = +\sqrt{1 - \frac{|R|}{R_{11}}}$$

Aurreko formulak balio orokorra dauka, hau da, edozein n-ren kasurako baliagarria zaigu eta honela idatz dezakegu:

$$r_{1,2,3,\dots,n} = +\sqrt{1 - \frac{|L|}{l_{11}L_{11}}} = +\sqrt{1 - \frac{|R|}{R_{11}}}$$

X_1 delakoa, azaldu behar den aldagaia izanik.

$\mathbb{R}^n$ -n daukagun koerlazio-koefiziente anizkoitza eta $\mathbb{R}^2$ -n daukaguna, berbera da, zeinua salbu.

Hots,

$$r_{1.2} = +\sqrt{1 - \frac{1-r_{12}^2}{1}} = +\sqrt{r_{12}^2} = |r_{12}|$$

$\mathbb{R}^n$ -n beste koefiziente batzuk, “Beta” koefizienteek hain zuzen, erregresioan aldagai independenteek daukaten indarra eta zeinua azalduko digute.

VIII.4. “BETA” ERREGRESIO-KOEFIZIENTE STANDARDIZATUAK

Zergatik ez dauka zeinurik $r_{1.23\dots n}$ koefizienteak?

Galdera honi, honela erantzun behar diogu:

Alde batetik, mugatze-koefizientearen erro karratua bakarrik izatean, zeinua justifikatzeko ezer ez dagoelako; eta bestaldetik, aldagai independente batzuk eragin negatiboa eta beste batzuk, aldiz, positiboa izan dezaketelako.

r_{12} ez da bakarrik mugapenaren indize bat, aldagaiak tipifikaturik edo standardizaturik daudeneko erregresio-koefizientea baizik.

Hots: $S_1 = S_2 = 1$ eta $S_{t_1 t_2} = r_{12}$ izatean:

$$\hat{T}_2 = r_{12} T_1 \quad \text{eta} \quad \hat{T}_1 = r_{12} T_2$$

Ez du, bada, bakarrik azaltzen den bariantza erlatiboki neurtzen, honekin batera erregresioaren zeinua ere ematen baitigu.

Dakigunez, aldagaiak tipifikaturik daudenean, koerlazio-matrizea kobariantz matrize bera da, eta ondorioz erregresio ekuazioa:

$$\hat{T}_1 = -\frac{\mathbf{R}_{12}}{\mathbf{R}_{11}} T_2 - \frac{\mathbf{R}_{13}}{\mathbf{R}_{11}} T_3 - \dots - \frac{\mathbf{R}_{1n}}{\mathbf{R}_{11}} T_n$$

Lehenengo erregresio-planoa edo hiperplano bera da, baina aldagaiak tipifikaturik hartzean, eskala berri batetan dauzkagu, hau da, $\mathbb{R}^n$ -n transformazio lineal bat egina daukagu, “tipifikazioa” hain zuzen.

Erregresio-koefiziente partzial tipifikatu edo standardizatu hauei, “ β ” letra grekoz deritzegu, koerlazio-matrizearen bidez kalkulatzuz:

Hots:

$$\beta_{1i} = - \frac{\mathbf{R}_{1i}}{\mathbf{R}_{11}}$$

edo eta kobariantz matrizearen bidez.

Hots:

$$\beta_{1i} = - \frac{S_i}{S_1} \frac{L_{1i}}{L_{11}} \quad \text{edo} \quad \beta_{1i} = \frac{b_{1i}}{S_1/S_i}$$

Azkenez, mugatze-koefiziente anizkoitzaren adierazpen matematikoan $|\mathbf{R}|$ determinantea lehenengo lerroko adjuntuen bidez idazten badugu:

$$\begin{aligned} \frac{S_{\hat{x}_1}^2}{S_{x_1}^2} &= r_{1.23\dots i\dots n}^2 = 1 - \frac{|\mathbf{R}|}{\mathbf{R}_{11}} = \\ &= 1 - \frac{\mathbf{R}_{11} + r_{12}\mathbf{R}_{12} + \dots + r_{1i}\mathbf{R}_{1i} \dots r_{1n}\mathbf{R}_{1n}}{\mathbf{R}_{11}} = \\ &= \beta_{12}r_{12} + \dots + \beta_{1i}r_{1i} + \dots + \beta_{1n}r_{1n} \end{aligned}$$

Azalduta dagoen bariantzaren proportzioa, batukorki deskonposatuta lortzen dugu honela.

Hau da, era honetara ere doitzen dugun hiperplanoan, aldagai independenteek aldagai menpekoaren bariantza zein proportziotan azaltzen duten jakin dezakegu, batugai guztiak positiboak direnen kasutan.

IX. KOERLAZIO PARTZIALA

IX.1. SARRERA

***IX.2. KOERLAZIO PARTZIALA: KONTZEPTU
MATEMATIKOA***

IX.3. ERREPIKAPEN-FORMULA

***IX.4. KOERLAZIO-KOEFIZIENTE PARTZIALA
ETA " β " KOEFIZIENTEAK***

IX.1. SARRERA

Erregresio anizkoitza ikastean, aldagai dependentearen estimazioan aldagai independente batzuk eragin positiboa eta beste batzuk negatiboa eduki dezaketela ikusi genuen.

Eragin hori erregresio-koefizienteen bidez aztertzen genuen, hau da, aldagai independente edo erregresore bakoitzak menpekoan zuen eragina, positiboa edo negatiboa eta beraien arteko erlazioa zuzena edo alderantzizkoa ote zen ikusiz.

Une honetan, gauza berbera aztertuko dugu, hau da, gure ikerketan errealtateranzko hurbilpena bilatzean, aldagai bi baino gehiago kontsideratuko ditugu, koerlazio partzialak aztertuz.

Zertzu dira koerlazio-koefiziente partzialak?

Koerlazio-koefiziente partzialak hauexek dira:

Bi aldagaien arteko egiazko erlazioa neurtu nahi duten koefiziente batzu, beste aldagai guztiak konstanteak kontsideratuz edo beste modu batetara esanda, beste guztien eragina baztertuz.

Bi aldagaien arteko koerlazioari, besteak kontutan izan gabe, koerlazio totala deritzo, baina, koerlazio honek ez du zentzu handirik aldagai horietatik bat menpekotzat kontsideratzen ez bada, eta bestea, menpeko edo dependente horretan eraginik nabariena duten bakarretako bat ez bada.

Honela, batzutan aurretik koerlazio positibo bat dela jakinik, kalkulatzearn negatiboa ateratzen zaigu, hots, fenomeno bi aldagaien artekoa balitz bezala hartu dugu, eta ez dugu beste aldagai batzuren eragina baztertu, koerlazioa ateratzean.

Ikus dezagun adibide bat:

Herrialde batetan urte batzutarako datuak hartuz, nekazal produkzioaren eta euri-kantitatearen arteko erlazioa ikertu nahi da.

Ondoko taulan dauzkagun datuen bidez, bada, erregresio lineala egingo dugu.

Urtea	Uztabilketa: X_1	Euri-kantitatea Udaberrian: X_2	Temperatura X_3
1.963	60	8	56
1.964	50	10	43
1.965	70	11	53
1.966	70	10	53
1.967	80	9	56
1.968	50	9	47
1.969	60	12	44
1.970	40	11	44

Erregresio-zuzena hauxe da:

$$\hat{x}_1 = 60 - 1'67x_2 \quad \text{eta} \quad r_{12} = -0'17$$

Koefizienteek, $b_{12} = -1'67$ eta $r_{12} = -0'17$ izanik erlazio negatibo edo alderantzizko bat dagoela adierazten digute.

Baina, aldez aurretik, X_1 eta X_2 -ren arteko erlazioa zuzena zela genekienez gero, X_1 eta X_2 -ren erregresioa egitean beraien erlazioa neurtzea, bide okerra da.

Guk jakin nahi duguna, hauxe da: Nola erlazionatzen dira X_1 eta X_2 beste garrantzizko aldagai guztiak konstantetzat hartuz gero?

Egin dezagun, bada, erregresio lineala X_3 aldagaia sartuz, tenperatura hain zuzen.

Orduan:

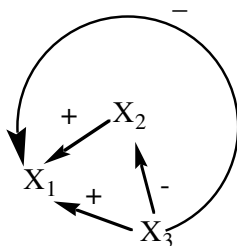
$$\hat{x}_1 = 60 + 5'71x_2 + 2'95x_3$$

Kasu honetan $b_{12} = 5'71$ da, erregresio-koefiziente positiboa daukagu, hots, euri-kantitatearen eta produkzioaren erlazioa zuzena da.

Zergatik ateratzen zitzaigun, bada, negatiboa?

Konkretuki fenomenoan aldagai garrantzizkoa izan arren X_3 kontutan hartzen ez genuelako.

Sinbolikoki

Zeharkako
Erlazioa (-)

Azal dezagun:

X_2 eta X_3 -ren arteko erlazioa negatiboa izatean, zeharkako erlazio negatibo bat sortzen da X_1 eta X_2 -ren artean, datu hauen artean dagoen erlazio zuzen positiboa gailenduz edo nagusituz.

Dakusagun, adibidez, nola 1.970. urtean euri-kantitate handia eta tenperatura baxua izan zen; ondorioz, uztabilketa txarra izan zen, baina ez X_2 balioagatik, X_3 -renagatik baizik.

X_1 , X_2 aldagaien arteko erlazioa aztertu nahi badugu, b_{12} erregresio-koefiziente partziala lortu behar dugu edo X_3 aldagaiaren balioak konstanteak balira bezala, honen eragina baztertuz, koerlazio-koefiziente partziala lortu.

Datu hauetan: $r_{12.3} = 0.68$

Halaber (X_1 , X_2 , X_3) aldagaien elkarren arteko koerlazioak zeinu berekoak izanez X_1 , X_2 aldagaien arteko koerlazioa ateratzean X_3 -ren eragina kontutan hartzen ez badugu, handiegi den gezurrezko balio bat lortuko dugula ikusiko dugu.

Dakusagun, bada, beste adibide hau:

Ondoko datu-matrizean, X_1 , X_2 , X_3 aldagaiei dagozkien oharpenak ditugu:

$$X = \begin{bmatrix} 90 & 40 & 100 \\ 95 & 60 & 120 \\ 100 & 50 & 130 \\ 100 & 65 & 110 \\ 105 & 80 & 115 \\ 95 & 70 & 135 \\ 120 & 85 & 140 \\ 140 & 110 & 180 \\ 130 & 100 & 160 \\ 150 & 120 & 175 \end{bmatrix}$$

X_1 : Hileroko famili gastu totalak

X_2 : Hileroko famili gastu totalak janaritan

X_3 : Hileroko famili sarrerak

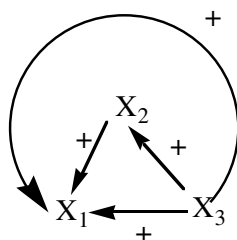
Argi dago X_1 eta X_2 aldagaien arteko koerlazioa positiboa izango dela; koerlazio totala kalkulatzean, emaitza, $r_{12} = 0.9356$ da.

Baina bi aldagai horiek sarrerekin koerlazio positiboa daukate eta beraien balioak sarrera horien menpe bezala ikus daitezke.

Orduan, egiazko koerlazioa ikusteko, X_3 aldagaiaren eragina baztertu behar dugu, hau konstante balitz bezala kontsideratuz.

Kalkuluak eginez: $r_{12.3} = 0.669$ da.

Sinbolikoki



Zeharkako
Erlazioa (+)

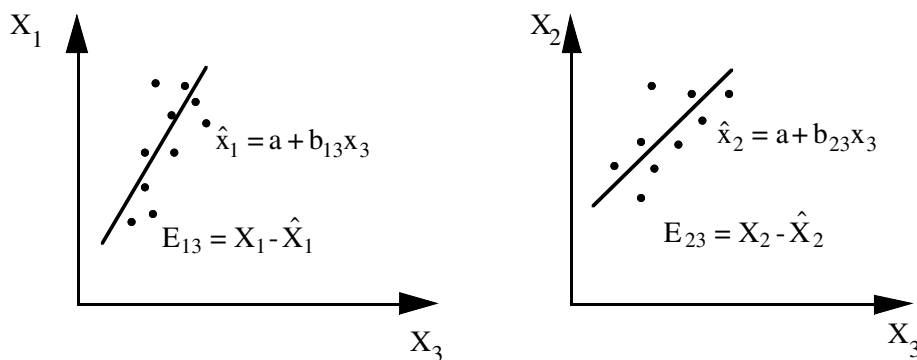
Koerlazio partziala kalkulatzean, zeharkako erlazioak duen eragin positiboa, baztertua izan da.

IX.2. KOERLAZIO PARTZIALA: KONTZEPTU MATEMATIKOA

Demagun, X_1 X_2 X_3 hiru aldagai direla.

Nola baztertuko dugu X_1 , X_2 aldagaien arteko koerlazioa ateratzean X_3 -k duen eragina?

X_1 eta X_2 aldagai bakoitzari, X_3 -rekiko bere erregresio-zuzeneko estimazioa kendu behar zaio.



Orduan, X_1 eta X_2 aldagaien arteko koerlazio-koefiziente partziala, E_{13} eta E_{23} aldagaien arteko koerlazio-koefizientea da.

Hots, aldagaiak zentraturik kontsideratuz:

$$E_{13} = x_1 - \hat{x}_1 = x_1 - \frac{S_{13}}{S_3^2} x_3$$

$$E_{23} = x_2 - \hat{x}_2 = x_2 - \frac{S_{23}}{S_3^2} x_3$$

Eta hondar-banaketa edo aldagi depuratu hauen koerlazio-koefizientea:

$$r_{12.3} = \frac{S_{e_{13}, e_{23}}}{(S_{e_{13}}^2 S_{e_{23}}^2)^{1/2}}$$

Hots, $\bar{e}_{13} = \bar{e}_{23} = 0$ eta $\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = 0$ izanik:

$$\begin{aligned} S_{e_{13}e_{23}} &= \frac{1}{m} \sum \left(x_{i1} - \frac{S_{13}}{S_3^2} x_{i3} \right) \left(x_{i2} - \frac{S_{23}}{S_3^2} x_{i3} \right) = \\ &= \frac{1}{m} \sum \left(x_{i1} x_{i2} - \frac{S_{13}}{S_3^2} x_{i2} x_{i3} - \frac{S_{23}}{S_3^2} x_{i1} x_{i3} + \frac{S_{13}S_{23}}{S_3^4} x_{i3}^2 \right) \\ &= S_{12} + \frac{S_{13}S_{23}}{S_3^2} - \frac{S_{13}S_{23}}{S_3^2} + \frac{S_{13}S_{23}}{S_3^4} S_3^2 = \\ &= S_{12} - \frac{S_{13}S_{23}}{S_3^2} \end{aligned}$$

Orduan:

$$r_{2.3} = \frac{S_{12} - \frac{S_{13}S_{23}}{S_3^2}}{\left(S_1^2 - \frac{S_{13}^2}{S_3^2} \right)^{1/2} \left(S_2^2 - \frac{S_{23}^2}{S_3^2} \right)^{1/2}}$$

Eta, zenbakitzailea eta izendatzailea, $S_1 S_2 = (S_1^2)^{1/2} (S_2^2)^{1/2}$ balioaz zatituz.

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{(1 - r_{13}^2)^{1/2} (1 - r_{23}^2)^{1/2}}$$

Koerlazio-matrizearen adjuntuen bidez idatziz:

$$r_{12.3} = - \frac{\mathbf{R}_{12}}{(\mathbf{R}_{11} \mathbf{R}_{22})^{1/2}}$$

Ikusten dugu, bada, nola lortzen den X_1 , X_2 aldagaien koerlazio partziala. Biek X_3 -ren bidez dauzkaten estimazioak kontutan hartuz eta hondarren koerlazioa eginez, X_3 aldagaiaarekin daukaten koerlazioa kontrolatu dugu.

Demagun orain, X_1 eta X_2 aldagaien arteko koerlazioa interesatzen zaigula, baina, $(X_1 X_2 \dots X_k)$ aldagaiak aztertu nahi dugun fenomenoan sarturik.

Kasu honetan, $X_3 \dots X_k$ aldagaiak duten eragina baztertu behar dugu eta, analogikoki, X_1 eta X_2 aldagaiek hiperplanoen bidez dauzkaten estimazioak kontutan hartuz hondarren koerlazioa kalkulatu dugu.

Hots:

$$E_{1.34\dots k} = x_1 - \hat{x}_1 = x_1 - a_1 - b_{13}x_3 - \dots - b_{1k}x_k$$

$$E_{2.34\dots k} = x_2 - \hat{x}_2 = x_2 - a_2 - b_{23}x_3 - \dots - b_{2k}x_k$$

Eta honela X_1 , X_2 aldagaien arteko koerlazio-koefiziente partziala lehenengokoa dela frogatzen da.

Orduan:

$$r_{12.3\dots k} = - \frac{\mathbf{R}_{12}}{(\mathbf{R}_{11} \mathbf{R}_{22})^{1/2}}$$

IX.3. ERREPIKAPEN-FORMULA

Aurkeztu dugun matrize-formulak guztiz errazak dira, batez ere formulazio teorikotarako. Praktikan, konputagailutarako programatzen den errepikapen-formula erabiltzen da, kontutan hartu behar dugu $r_{12.3\dots k}$ koerlazio-partziala ateratzeko formula matrizealean ditugun adjuntuak $(k-1)$ ordenakoak direla.

Errepikapen-formularen bidez k -ordenako koerlazioak $(k-1)$ ordenakoen bidez kalkulatzen dira.

$(X_1 X_2 \dots X_k \dots X_n)$ aldagai n -koitza badaukagu.

$r_{12}, r_{12} \dots r_{2k}, \dots r_{n-1,n}$ zero ordenako koerlazioak dira. Hauen bidez, 1 ordenako edo lehen ordenako koerlazio partzialaren koefizienteak:

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - r_{13}^2} \sqrt{1 - r_{23}^2}}$$

$$r_{13.2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{1 - r_{12}^2} \sqrt{1 - r_{23}^2}}$$

$r_{14.5} \ r_{56.k} \dots \dots r_{12.n} \dots \dots$ atera daitezke.

1 ordenako koerlazioen bidez 2 ordenakoak edo bigarren ordenakoak:

$$r_{12.34} = \frac{r_{12.3} - r_{14.3} r_{24.3}}{\sqrt{1 - r_{14.3}^2} \sqrt{1 - r_{24.3}^2}} \text{ edo}$$

$$r_{12.34} = \frac{r_{12.4} - r_{13.4} r_{23.4}}{\sqrt{1 - r_{13.4}^2} \sqrt{1 - r_{23.4}^2}}$$

eta honela bi bide desberdin jarraituz $r_{13.56} \ r_{14.56} \dots$ atera daitezke.

2-ordenako koerlazio partzialaren koefizienteen bidez 3 ordenakoak edo hirugarren ordenakoak:

$$r_{12.345} = \frac{r_{12.34} - r_{15.34} r_{25.34}}{\sqrt{1 - r_{15.34}^2} \sqrt{1 - r_{25.34}^2}} \text{ edo}$$

$$r_{12.345} = \frac{r_{12.35} - r_{14.35} r_{24.35}}{\sqrt{1 - r_{14.35}^2} \sqrt{1 - r_{24.35}^2}} \text{ edo}$$

beste lau era desberdinetan atera daitezke.

Noski SPSS bezalako pakete-programa baten bidez oso erraz lortuko ditugu guztiak, PARTIAL CORR aginduaren bidez hain zuzen.

IX.4. KOERLAZIO-KOEFIZIENTE PARTZIALA ETA “ β ” KOEFIZIENTEAK

Sarreran esan dugun bezala, bi aldagaien arteko erlazioa erregresio-koefizienteen bidez ikusten da eta “ β ” erregresio-koefiziente partzial standardizatuak, bereziki, aldagai independente edo erregresore bakoitzak erregresioan duen indar erlatiboa eta zeinua adierazten digute.

Dakusagun, bada, β koefizienteen eta koerlazio-koefiziente partzialen arteko erlazioa:

Matematikoki:

$$|r_{12.3}| = \left| -\frac{\mathbf{R}_{12}}{(\mathbf{R}_{11}\mathbf{R}_{22})^{1/2}} \right| = +\sqrt{\frac{-\mathbf{R}_{12}}{\mathbf{R}_{11}} \cdot \frac{-\mathbf{R}_{12}}{\mathbf{R}_{22}}} = +\sqrt{(\beta_{12}\beta_{21})}$$

Hots: $|r_{12.3\dots k}| = (\beta_{12}\beta_{21})^{1/2}$

Koerlazio-koefiziente partziala bi erregresio-koefiziente standardizatuaren batezbesteko geometrikoa da.

Lehenengo kasuan, hau da, X_3 aldagaiaren eragina baztertzen dugunean, koerlazio-matrizearen adjuntuak idatziz, honela izango da:

$$|r_{12.3}| = \left| -\frac{r_{12} - r_{13}r_{23}}{(1 - r_{13}^2)^{1/2} (1 - r_{23}^2)^{1/2}} \right| = + \sqrt{\frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2} \frac{r_{12} - r_{13}r_{23}}{1 - r_{13}^2}}$$

Berdin ikus daiteke, $X_3 \dots X_k$ aldagaien eragina baztertzen dugunean, koerlazio-matrizearen adjuntuak idatziz.

Matrizialki idatziz

$$r_{12.3\dots k} = \frac{-\mathbf{R}_{12}}{\mathbf{R}_{11}^{1/2} \mathbf{R}_{22}^{1/2}} \quad \text{dela}$$

kontutan hartuz “ β ” erregresio-partzial standardizatuaren koefizienteen ondoko adierazpenak idatz ditzakegu.

$$\beta_{12.3\dots k} = \frac{\mathbf{R}_{22}^{1/2}}{\mathbf{R}_{11}^{1/2}} r_{12.3\dots k}$$

$$\beta_{21.3\dots k} = \frac{\mathbf{R}_{11}^{1/2}}{\mathbf{R}_{22}^{1/2}} r_{12.3\dots k}$$

***X. DESKRIBAPEN ESTATISTIKOAREN
ADIERAZPIDE GEOMETRIKOAK***

X.1. OROKORTASUNAK

*X.2. X DATU-MATRIZEAREN BI ADIERAZPIDE
GEOMETRIKO*

X.3. BI ALDAGAI ETA BI OHARPEN

X.4. BI ALDAGAI ETA HIRU OHARPEN

*X.5. ERREGRESIO BAKUNA ALDAGAI ESTATIS-
TIKOEN BALIO-MULTZOEN ESPAZIOAN*

X.1. OROKORTASUNAK

Ikasgai honetan batezbestekoa, bariantza eta kobariantza geometrikoki ikustea hartzen dugu helburutzat.

Matrize-estatistikoak azaltzen genuen ikasgaian, matrizialki ikasketak egitea estatistika anizkoitzean beharrezko bihurtzen dela esaten genuen.

Errealitatea estatistikoki¹ ikasi nahi badugu, bada, formulazio matematiko bektoriala edota matriziala nahitaezkoak eta bide batez paregabeak direla esan daiteke, errealitatea bektoriala baita.

$\mathbf{X}_{(m \cdot n)}$ datu-matrizearen bektoreak, $\mathbb{R}^n$, $\mathbb{R}^m$ espazioetako bektoreak alegia, aurkezten hasiko gara, ondoren, $\mathbb{R}^2$, $\mathbb{R}^3$ espazioetan zentratutako estatistikoak geometrikoki ikusiko ditugu.

Zoritxarrez n edo m 3 baino handiago direnean iruditzen ezer gutxi ikusiko dugu, baina, $\mathbb{R}^2$, $\mathbb{R}^3$ espazioak ondo aztertu baditugu espazio anizkoitzean gertatzen dena errazago ulertuko dugu.

X.2. X DATU-MATRIZEAREN BI ADIERAZPIDE GEOMETRIKO

Lehen aurkeztutako $\mathbf{X}_{(m \cdot n)}$ datu-matrizean bi bektore mota kontsidera daitezke:

1) $\mathbb{R}^n$ espazioan, indibiduoek dituzten n koordenatuak n aldagaien arabera $\mathbb{R}^n$ espazioko bektoreak osatzen dituzte.

2) Bestalde, $\mathbb{R}^m$ espazioan, aldagaiek dituzten m koordenatuak m indibiduen gain $\mathbb{R}^m$ espazioko bektoreak osatzen dituztelarik.

1. Estatistika Deskribatzailean kokaturik, eta konkretuki, Datu-Analisiaren atarian, hau da, probabilitate-indukzioan oinarritzen den Inferentzia Estatistikotik at.

Hots:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & \dots & x_{mj} & \dots & x_{mn} \end{bmatrix}$$

Errenkadak, $\mathbb{R}^n$ espazioko bektoreak dira, i. indibiduorako edo oharpenerako $(x_{i1} \ x_{i2} \ \dots \ x_{ij} \ x_{in})$ koordinatuak direlarik. Errenkadetan, bada, m **indibiduo-puntu** edo oharpen-puntu ditugu, orain arte edozein hodeitan genituen puntuak alegia.

Zutabeak, aldiz, $\mathbb{R}^m$ espazioko bektoreak dira j. aldagairako $(x_{1j} \ x_{2j} \ \dots \ x_{ij} \ \dots \ x_{mj})$ koordinatuak direlarik. Zutabetan n **aldagai-puntu** ditugu hemendik aurrera kontutan hartuko dugun puntu-hodei berri bat osatzen dutelarik.

X.3. BI ALDAGAI ETA BI OHARPEN

Bi aldagai bi indibiduen gain neurtu ondoren honako emaitza hauek ditugu.

$$\begin{matrix} & X_1 & X_2 \\ \omega_1 & \begin{bmatrix} x_1(1) & x_2(1) \end{bmatrix} \\ \omega_2 & \begin{bmatrix} x_1(2) & x_2(2) \end{bmatrix} \end{matrix} = (\text{konkretuki}) = \begin{bmatrix} 1 & 5 \\ 3 & 2 \end{bmatrix} \text{ datu-matrizea}$$

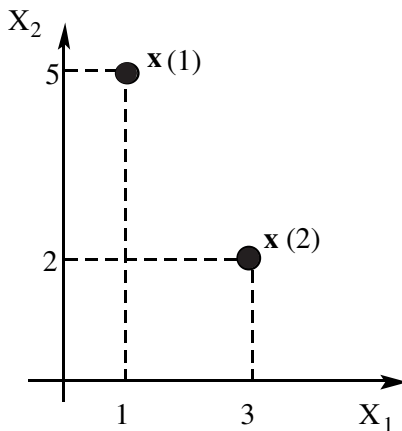
adibide bezala erabiliko dugu.

Kasu honetan $\mathbf{x}(1) = (1,5)^T$, $\mathbf{x}(2) = (3,2)^T$, $\mathbb{R}^2$ -ko bi bektoreak indibiduo-puntuak edo oharpen-puntuak dira. $\mathbf{x}_1 = (1,3)$, $\mathbf{x}_2 = (5,2)$, $\mathbb{R}^2$ -ko bi bektoreak, aldiz, aldagai-puntuak.

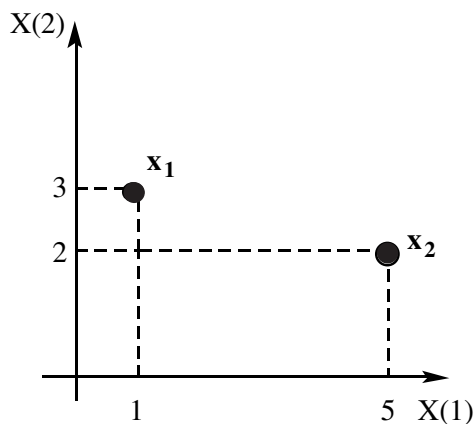
Konturatu behar dugu datu-matrizeko bektoreak ez direla desberdintzen errenkada-bektoreak edo zutabe-bektoreak izateagatik bakarrik; daukaten desberdintasuna benetan garrantzitsuagoa da, lehenengoak balio-multzo ez-homogenoak dira aldagai desberdinen neurriak indibiduo baten gain izanik, bigarrenak, ordea, balio-multzo homogenoak dira aldagai bakoitzaren neurriak izatean.

Datu-matrizearen (i,j) elementua adierazteko bi ikur ezberdin aukeratu ditugu. Hauexek dira: $x_{ij} = x_j(i)$

Biek X_j aldagaiak ω_i indibiduoaren gain hartzen duen balioa adierazten digute. Dakusagun, bada, adierazpide grafikoak:



Indibiduoaren Hodeia



Aldagaiaren Hodeia

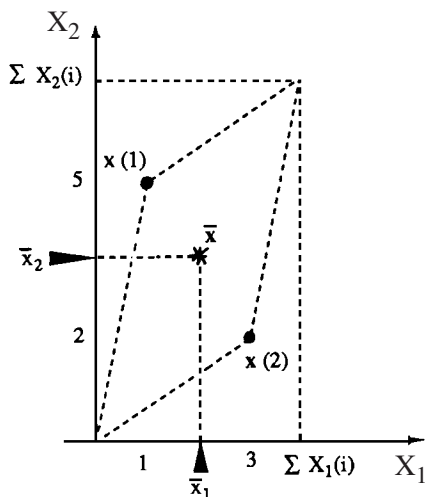
X_1 , X_2 , aldagaien balioen batezbestekoek betebeharrak desberdinak betetzen dituzte bi hodei horietan

$$\bar{x}_1 = \frac{1}{m} \sum_i x_1(i) = 2$$

$$\bar{x}_2 = \frac{1}{m} \sum_i x_2(i) = 3.5$$

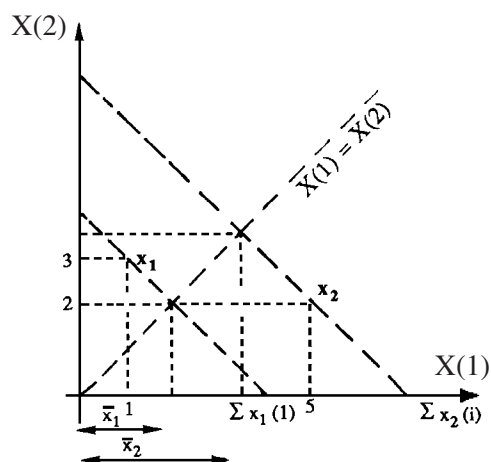
INDIBIDUOEN HODEIAN

Indibiduo-puntuaren grabitate-zentrua finkatzen dute.



ALDAGAIEN HODEIAN

Aldagai-puntuen osagaien batezbestekoa



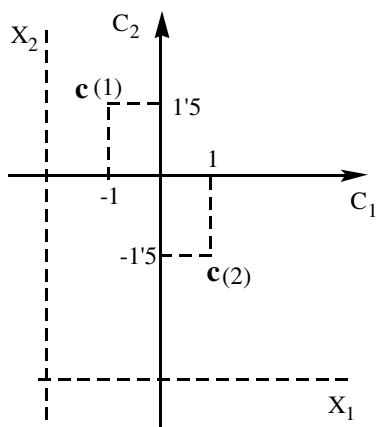
Aldagai-bektorearen batezbestekoa, $X(1) = X(2)$ erdikarian aldagai-puntuak daukan proiektatutako puntuaren osagai bakoitza bezala ikusten da.

Aldagaien balio bakoitzaren **zentratze-eragiketa**:

$C_i(i) = X_j(i) - \bar{x}_j$ zentratze-eragiketa eginez, C datu-matrize zentratua lortzen dugu.

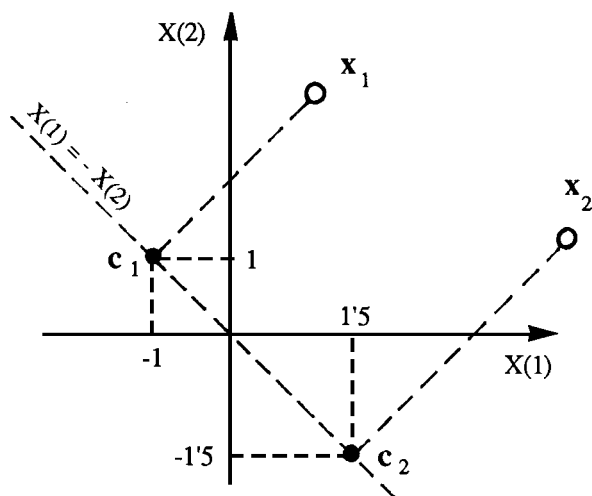
$$C = \begin{bmatrix} c_1(1) & c_2(1) \\ c_1(2) & c_2(2) \end{bmatrix} = \begin{bmatrix} -1 & 1,5 \\ 1 & -1,5 \end{bmatrix}$$

INDIBIDUOEN HODEIAN



Zentratzeak ardatzen aldaketa paralelo bat egitea suposatzen du hodei honetan.

ALDAGAIEN HODEIAN



Hodei honetan, ordea, batezbestekoz eta baturaz zero den $X(1) = -X(2)$ erdikarian aldagai-puntuak proiektatzea da zentratze-eragiketak suposatzen duena.

Aldagai batek bi oharpen dituenean, zentratzea $X(1) = -X(2)$ erdikarian proiektatzea da, bi balio zentratuak, berdinak balio absolutuz eta aurkakoak baitira. Ikus grafikoan.

Horrexegatik $X(2 \times 2)$ datu-matrizeak ezaugarri oso bereziak ditu.

Dakusagun:

1) Desbidazio standarda aldagai zentratuaren koordinatuen balio absolutua da.

$$S_j = |c_j(1)| = |c_j(2)|$$

$$\text{Kasu honetan: } S_1 = |-1| = |1| = 1$$

$$S_2 = |1'5| = |-1'5| = 1'5$$

Dakigunez:

$$S_1^2 = \frac{1}{m} \sum_1 c_1^2 \quad (i) \quad S_2^2 = \frac{1}{m} \sum_1 c_2^2 \quad (i)$$

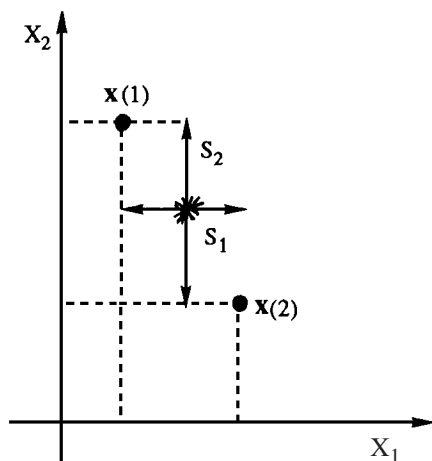
$$S_1 = \frac{\|c_1\|}{\sqrt{m}} \quad S_2 = \frac{\|c_2\|}{\sqrt{m}}$$

$m = 2$, hots bi indibiduo edo bi oharpen izanik

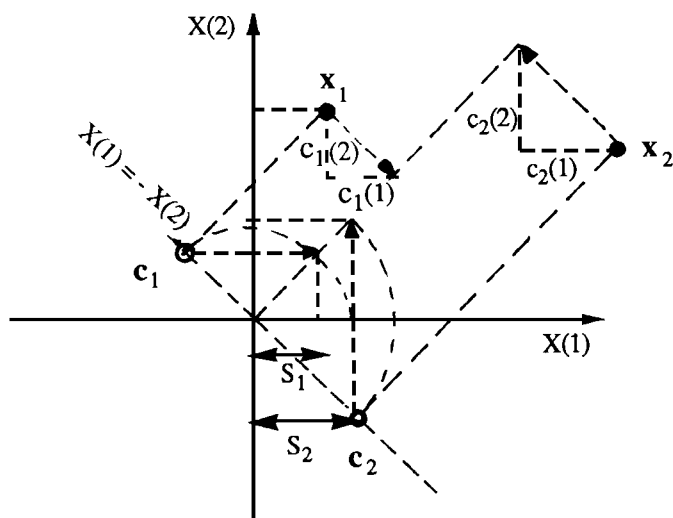
$$\mathbf{L} = \begin{bmatrix} S_1^2 & S_{12} \\ S_{12} & S_2^2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} -1 & 1 \\ 1'5 & -1'5 \end{bmatrix} \begin{bmatrix} -1 & 1'5 \\ 1 & -1'5 \end{bmatrix} = \begin{bmatrix} 1^2 & -1.1'5 \\ -1.1'5 & 1'5^2 \end{bmatrix}$$

Kobariantz matrizea lortzean, matrize hauen bigarren berezitasuna agertzen zaigu, baina horretaz geroxeago hitz egingo dugu.

Dakusagun, orain bada, desbidazio standarda nola agertzen den indibiduen hodeian.



Kasu berezi honetan ($X(2 \times 2)$) desbidazio standarda edo tipikoa “desbidazioa” balio absolutuan da.



$X(1) = X(2)$ erdikarian aldagai-puntua proiektatzean, proiektatutako puntuaren koordinatu bakoitza batezbestekoa zen bezalaxe, orain erdikarian aldagai-puntu zentratua proiektatzean, proiektatutako puntuaren koordinatu bakoitza desbidazio standarda da.

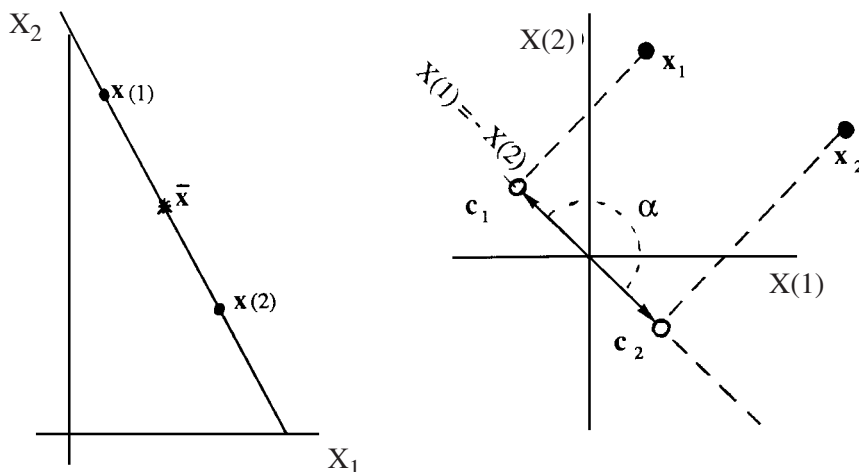
C_j bektorea eta bere proiektzioa norma bera daukate:

$$\|c_j\| = \sqrt{m} S_j$$

aldagai zentratuaren norma sakabanatzearen neurri on bat da, askotan erabilia izanik.

Azkenik azpimarratu behar dugu aldagai zentratuaren norma, jatorrizko aldagaiak erdikarira daukan distantzia euklidearra dela.

2) Bi oharpen ditugunean aldagaiak guztiz korrelatuak dira beti, hots koerlazio koefizientea 1 edo -1 (gure adibidean bezala) izango da. Ikus **L** aurreko orrialdean.



$$r_{12} = \frac{S_{12}}{S_1 S_2} = \frac{\langle c_1, c_2 \rangle / m}{\frac{\|c_1\|}{\sqrt{m}} \frac{\|c_2\|}{\sqrt{m}}} = \frac{\langle c_1, c_2 \rangle}{\|c_1\| \|c_2\|} = \cos \alpha$$

Dakusagunez, aldagai zentratuen azpiespazio bektoriala $X(1) = -X(2)$ zuzena da, horrexegatik edozein aldagai bikotek osatzen duen angelua zero edo ehun eta larogei gradukoa izango da.

Horrek, dakigunez, erregresio perfektua halabehartzen du.

Indibiduo-puntuek ezin dutela zuzen baten gainean ez izan ikusten dugu ezkeraldeko hodeian, lerrokatuak daude.

Aldagai zentratu bakoitza bestearen eskalarea dela ikusten dugu, ordea, eskuinaldekoan.

Ondoren, hau baino errepresentagarriago den kasu bat aztertuko dugu. Bi aldagai eta hiru oharpenen kasua, alegia.

Ariketa: $\mathbf{X}$ datu-matrizea $\begin{bmatrix} 1 & 4 \\ 5 & 6 \end{bmatrix}$ matrizea izanik. Ikus ezazu, grafikoki, $\mathbf{c}_1$, $\mathbf{c}_2$ bektoreak zero graduko angelua osatzen dutela edota $r_{12} = 1$ dela.

X.4. BI ALDAGAI ETA HIRU OHARPEN

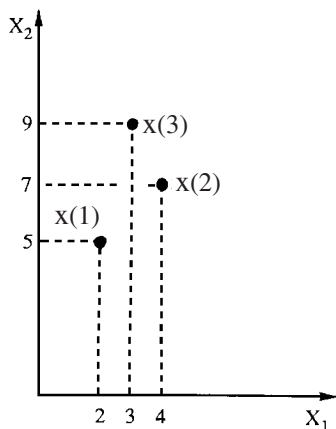
Ondoan, bi aldagai eta hiru indibiduoren kasuko matrizea daukagu.

$$\mathbf{X} = \begin{bmatrix} x_1(1) & x_2(1) \\ x_1(2) & x_2(2) \\ x_1(3) & x_2(3) \end{bmatrix}$$

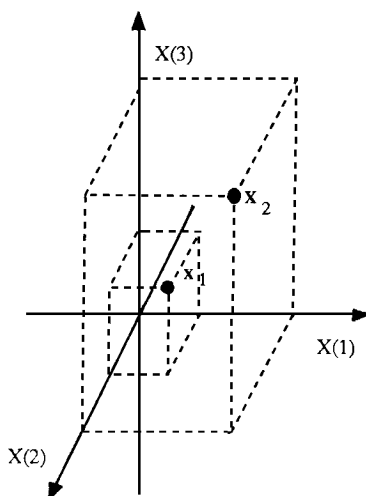
Eta adibide bezala kontsideratzen duguna.

$$\mathbf{X} = \begin{bmatrix} 2 & 5 \\ 4 & 7 \\ 3 & 9 \end{bmatrix}$$

Indibiduo-puntuen hodeia



Aldagai-puntuen hodeia

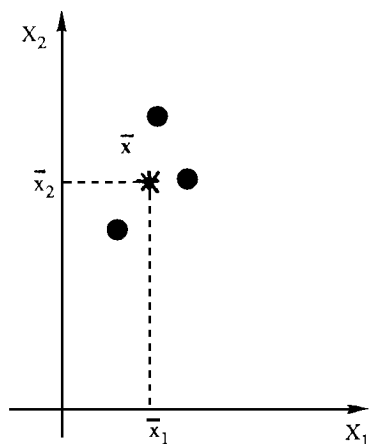


$$\bar{x}_1 = \sum_1 x_1(i) / m = 3$$

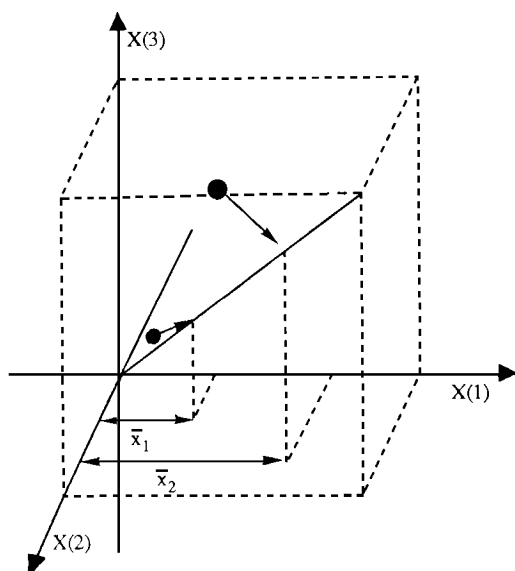
$$\bar{x}_2 = \sum_1 x_2(i) / m = 7$$

$$\text{Hots } \bar{\mathbf{x}} = \begin{bmatrix} 3 \\ 7 \end{bmatrix}$$

Dakusagun, bada, adierazpide grafikoak:



Batezbesteko bektorea indibiduen barizentrua edo grabitate-zentrua da.



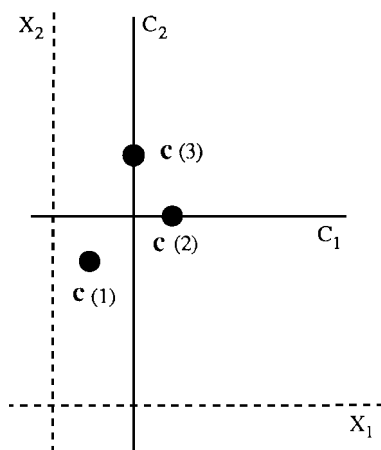
Aldagai bektorearen batezbesteko osagaia $X(1) = X(2) = X(3)$ erdikarian proiektatutako puntuaren osagai bakoitza da.

Zentratze-eragiketa:

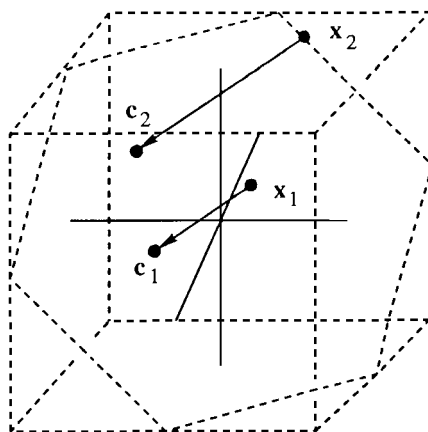
$$c_1(i) = x_j(i) - \bar{x}_j \text{ eginez} \quad C = \begin{bmatrix} -1 & -2 \\ 1 & 0 \\ 0 & 2 \end{bmatrix}$$

datu-matrize zentratua lortzen dugu.

Grafikoki:



Berriro ere ardatzen aldaketa paralelo bat daukagu.

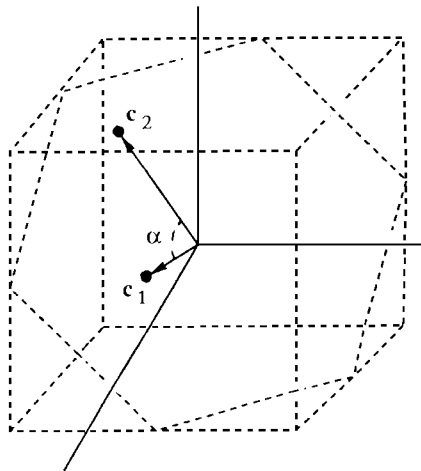


Lehen $X(1) = -X(2)$ erdikarian proiektatzen ziren aldagai-puntuak, hots, baturaz eta batezbestekoz zero diren bektoreen azpiespazioan.

Orain, bada, berdin gertatzen da baina kasu honetan baturaz eta batezbestekoz zero diren bektoreen azpiespazioa plano bat da, $X(1) + X(2) + X(3) = 0$ plano a alegia.

Irudian, kubo baten barruan dagoen hexagonoaren bidez ikusarazi nahi da azpiespazio hori, hau da, hexagonoak $X(1) + X(2) + X(3) = 0$ plano a ikusarazteko besterik balio ez duelarik.

Aldagai zentratuen grafika handituz aldagaien arteko koerlazioa geometrikoki ikusiko dugu.



Ikusi genuenez:

$$\|c_1\| = \sqrt{m} \quad S_1$$

$$\|c_2\| = \sqrt{m} \quad S_2$$

$$r_{12} = \frac{\langle c_1, c_2 \rangle}{\|c_1\| \|c_2\|} = \cos \alpha \quad \alpha = \arccos r_{12}$$

Aldagaien osagaiak ez dira berdinak balio absolutuan ez eta aldagaiek ez dute zergatik guztiz korrelatuak izan behar ere.

Darabilgun adibidean **L** eta **R** matrizeak hauek dira:

$$\mathbf{L} = \begin{bmatrix} 2/3 & 2/3 \\ 2/3 & 8/3 \end{bmatrix} \quad \mathbf{R} = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix} \quad \begin{array}{l} \text{hots } \alpha = 60^\circ \\ \alpha = \arccos 1/2 \end{array}$$

Ariketa: $\mathbf{c}_2$ bektorearen ordezkio berean dagoen $\mathbf{c}_2 = (-1, 0, 1)$ bektorea kontsideratuz. Ikus ezazu $\mathbf{c}_1, \mathbf{c}_2$ bektoreen arteko angelua 60° koa dela.

Oharra: $\mathbf{c}_1, \mathbf{c}_2, 0$ puntuek alde berdineko triangelu bat osatzen dute, alde bakoitzak 2 balio duelarik.

DATU-MATRIZE TIPIFIKATUA ETA ZENTRU-NORMALIZATUA

Dakigunez aldagaien balioak tipifikatuak edo standardizatuak dituen datu-matrizeari tipifikatua deritzogu.

Adibidean:

$$\mathbf{C} = \begin{bmatrix} -1 & -2 \\ 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad S_1 = \sqrt{\frac{2}{3}}, \quad S_2 = 2\sqrt{\frac{2}{3}} \quad \text{izanik}$$

$$\mathbf{T} = \left[\frac{x_j(i) - \bar{x}_j}{S_j} \right] = \begin{bmatrix} \frac{\sqrt{3}}{\sqrt{2}} & \frac{\sqrt{3}}{\sqrt{2}} \\ \frac{\sqrt{3}}{\sqrt{2}} & 0 \\ 0 & \frac{\sqrt{3}}{\sqrt{2}} \end{bmatrix}$$

Era berean aldagai zentratuen balioak dagozkien normaz zatituz datu-matrize zentru-normalizatua lortuko dugu.

Hau da:

$$\|\mathbf{c}_1\| = S_1 \sqrt{3} = \sqrt{2}$$

$$\|\mathbf{c}_2\| = S_2 \sqrt{3} = 2\sqrt{2}$$

$$\mathbf{N} = \begin{bmatrix} \frac{x_j(i) - \bar{x}_j}{S_j \sqrt{m}} \\ \frac{-1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

Aldagai-puntu tipifikatuen norma $\sqrt{m}$ da.

Hots:

$$\|t_j\| = \left\| \frac{x_j(i) - \bar{x}_j}{S_j} \right\| = \left\| \frac{\mathbf{c}_j}{S_j} \right\| = \frac{\|\mathbf{c}_j\|}{S_j} = \frac{|\mathbf{c}_j|}{|\mathbf{c}_j|/\sqrt{m}} = \sqrt{m}$$

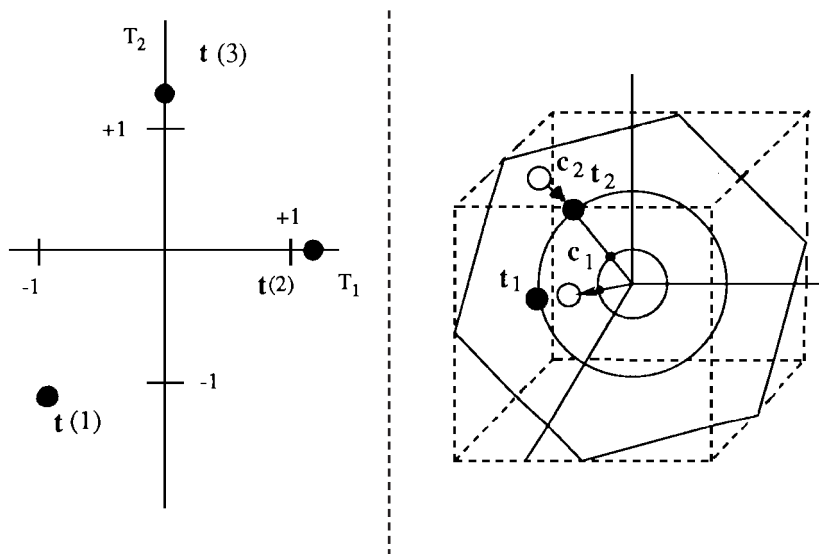
Noski, $\mathbf{N}$ matrizeko zutabe-bektoreek 1 daukate normaz, bektoreak normalizatuak baitaude.

Edozein kasutan, aldagaiak zentratzeak translazio bat bakarrik suposatzen badu ere, tipifikatzeak edota zentratu ondoren normalitzeak eskala aldatzea suposatzen du.

Horrexegatik, $\mathbb{R}^3$ espazioan aldagai-puntu tipifikatuak $\sqrt{3}$ erradioz duen zirkulu batetan edota $m = 4$, $m > 4$ direnean $\sqrt{m}$ erradioz duen esfera edo hiperesfera batetan kokatzen dira hurrenez hurren.

Aldagai-puntu zentru-normatuak, berriz, erradioz 1 neurtzen duen zirkulu, esfera, edo hiperesferan kokatzen dira.

Bi kasu horietan, balioak erlatibizatzen ditugu desbidazio standardaren arabera eta aldagai guztiak eskala berean ditugunez beraien direkzioak interesatzen zaizkigu.



Indibiduen balioak tipifikaturik ditugu ezkerraldeko hodeian.

Eskumaldeko hodeian $\sqrt{3}$ erradioz duen zirkunferentzian aldagai-bektore tipifikatuak ikusten ditugu, baita, aldagai-bektore normatuak 1 erradioz duen zirkunferentzian ere.

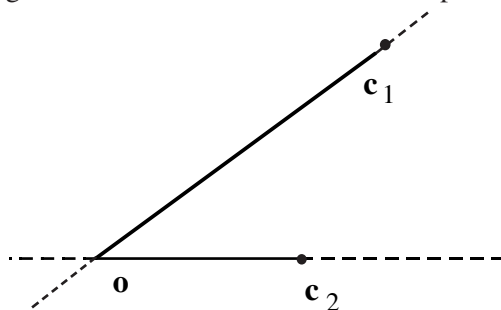
X.5. ERREGRESIO BAKUNA ALDAGAI ESTATISTIKOEN BALIO-MULTZOEN ESPAZIOAN

X.2. atalean ikusi dugun bezala $\mathbf{X}(m \times n)$ datu-matrizearen zutabeak $\mathbb{R}^m$ espazioko aldagai-bektoreak dira, m indibiduen gain aldagai bakoitzaren balioak hartu baititugu.

Aldagai-bektore zentratuak, ordea, beraien osagaien edo balioen batura zero denez, $\mathbb{R}^{m-1}$ espazioko bektoreak dira. Konkretuki, X.3. eta X.4. ataletan, $\mathbb{R}^1$ eta $\mathbb{R}^2$ espazioan kokaturik ikusi ditugun bi indibiduen eta hiru indibiduen kasuetarako, hurrenez hurren.

Bi aldagaien arteko erregresio bakuna (aldagai independente edo erregresore bat) beti, plano batean adieraz daiteke, alegia, $\mathbb{R}^{m-1}$ dimentsioko bi aldagai-bektore zentratuek sortarazten duten espazio bidimentsionalean.

Bi aldagai eta m oharpen suposatuz, ikus ditzagun, bada, $X(m \times 2)$ datu-matritzeko bi aldagai zentratuak beraiek osatzen duten planoan adieraziak.



X_1 aldagaiaren X_2 aldagaiarekiko erregresioa suposatuz eta aldagai-bektore zentratuen bidez adieraziz:

$$\hat{c}_1 = b_{12} c_2$$

Hau da, aldagaiaren balio estimatuen bektorea c_2 bektorearen eskalarea da.

$b_{12} c_2$, bada, $O c_2$ segmentua kokatuta dagoen zuzenean kokatuko da, zuzen hau c_2 bektorearen eskalare guztien leku geometrikoa edo c_2 direkzioa baita.

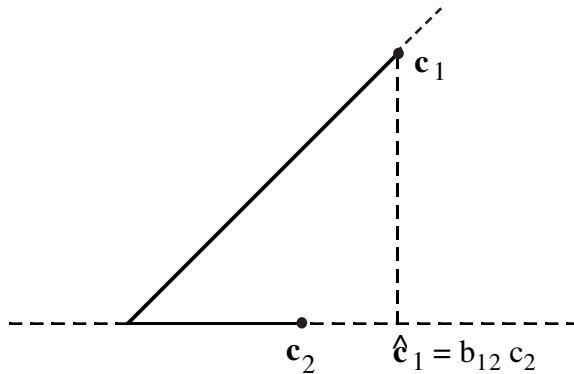
$\mathbb{R}^m$ espazioan, c_1 eta $b_{12} c_2$ bektoreen arteko distantzia, norma euklidearra kalkulatu daiteke.

$$d = + \sqrt{\sum_i [c_1(i) - b_{12} c_2(i)]^2} = \sqrt{\sum_i [c_1(i) - \hat{c}_1(i)]^2}$$

eta $c_1(i) - \hat{c}_1(i)$ i . indibiduoaren hondarra edo errorea izanik, distantzia, errore karratuen baturaren erro karratua da, hau da, $\|e\|$ norma. Balio hau txikiena eta bide batez, S_e^2 hondar-bariantza txikiena izatea da erregresioaren helburua.

Erregresioa, bada, geometrikoki honela planteatu daiteke:

$\hat{c}_1 = b_{12} c_2$ eta c_1 balio-multzoen arteko distantzia txikiena izatea da bilatzen dena. Balio hau, dakigunez, c_1 eta c_1 -ek c_2 direkzioan, daukan proiektzio ortogonalaren arteko distantziari dagokio.



Honela, geometrikoki, b_{12} erregresio-koefizientea lor dezakegu:

$$\text{proj. } \mathbf{c}_1 = b_{12} \|\mathbf{c}_2\|$$

eta dakigunez,

$$\text{proj. } \mathbf{c}_1 = \langle \mathbf{c}_1 | \frac{\mathbf{c}_2}{\|\mathbf{c}_2\|} \rangle = \frac{\langle \mathbf{c}_1 | \mathbf{c}_2 \rangle}{\|\mathbf{c}_2\|}$$

orduan,

$$b_{12} = \frac{\langle \mathbf{c}_1 | \mathbf{c}_2 \rangle}{\|\mathbf{c}_2\|^2} = \frac{\langle \mathbf{c}_1 | \mathbf{c}_2 \rangle / m}{\|\mathbf{c}_2\|^2 / m} = \frac{S_{12}}{S_2^2}$$

azkenik, balio estimatuen bektorea

$$\hat{\mathbf{c}}_1 = \frac{S_{12}}{S_2^2} \mathbf{c}_2$$

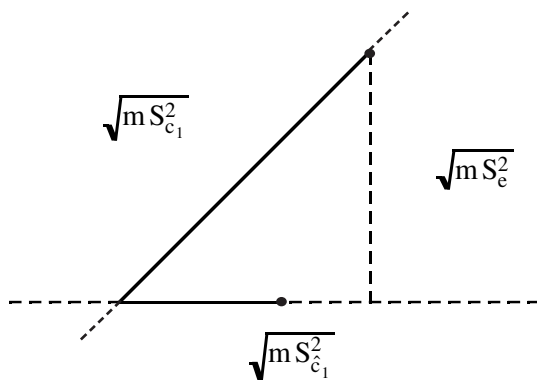
$\mathbf{c}_1$ bektorearen bektore-posizioaren osagaiak $\hat{\mathbf{c}}_1$ bektorearekiko $\mathbf{c}_1$ -ren $\mathbf{c}_2$ -rekiko erregresioaren hondarrak dira.

Dakusagunez, $\mathbf{e}$ bektorearen direkzioa $\mathbf{c}_2$ eta $\hat{\mathbf{c}}_1$ bektoreen direkzioarekiko ortogonal da, honela geometrikoki erregresioaren bi propietate hauek ikusten ditugu:

$-\mathbf{r}_{\mathbf{c}_2} = \mathbf{r}_{\mathbf{e}_2} = 0$, hau da, hondarrak eta aldagai independentea koerlazio gabeak dira.

$-r_{e\hat{c}_1} = r_{e\hat{x}_1} = 0$, hau da, hondarrak eta balio estimatuak koerlazio gabeak dira.

Azkenik, hain garrantzitsua den bariantzaren deskonposaketa batukorra, geometrikoki ikus dezakegu:



Pitagoras-en teorema aplikatuz

$$S_{c_1}^2 = S_{\hat{c}_1}^2 + S_e^2$$

edota

$$S_{x_1}^2 = S_{\hat{x}_1}^2 + S_e^2$$

***XI. ERREGRESIO ORTOGONALA ETA
KOBARIANTZ MATRIZEAREN
AUTODIREKZIOAK***

XI.1. ELKARREN MENPEKO ANALISIA

*XI.2. ERREGRESIO-ZUZEN ORTOGONALA ETA
HONDAR-BARIANTZA*

*XI.3. BATERAKO BARIANTZAREN DESKONPOSAKETA
ERREGRESIO ORTOGONALEAN*

*XI.4. ERREGRESIO ORTOGONALAREN
PROPIETATEAK*

XI.1. ELKARREN MENPEKO ANALISIA

Orain arte ikasi dugun ohizko erregresioan, ikasketa egin aurretik aldagaiak bi multzo desberdinetan honela banatzen genituen.

Batean, azaldu nahi genuen aldagaia kontsideratuz.

Bestean, aldagai independenteak edo azaltzaileak.

Erregresio Ortogonalak (bi aldagaien kasuan) eta Osagai Nagusizko Analisiak (bi baino aldagai gehiagoren kasuan), ordez, aldagaiak multzo bakar batean kontsideraturik ikasten ditu.

Aldagai guztiak dira, bada, bide batez, azaltzaileak eta azalduak, hau da, betebeharrak baldin badira.

XI.2. ERREGRESIO-ZUZEN ORTOGONALA ETA HONDAR-BARIANTZA

Ohizko erregresioan bezala, erregresio ortogonalaren helburua $\mathbb{R}^2$ espazioaren puntu-hodeian zuzen bat doitzea da.

Doikuntza, lehen bezala berriro ere, erizpide hoberen bati jarraituz egingo dugu, aukeratutako erizpidea errore koadratikoa txikiena izatea izango delarik.

Zertan datza, bada, erregresio honen ezberdintasuna? Erregresio-errorearen definizioan, hain zuzen. Errorea kasu honetan balio ohartu eta balio estimatuaren arteko desberdintasuna ez baita.

Erregresio ortogonalean, definitzen den errorea, balio ohartutik zuzenera dagoen distantzia euklidearra da.

Izan bedi $\mathbb{R}^2$ espazioan definitutako (X, Y) aldagai estatistiko bikoitza eta izan bitez (x_i, y_i) ; $\alpha x + \beta y + \gamma = 0$ oharpena eta doitzen dugun zuzenaren ekuazioa hurrenez hurren.

Errorea (x_i, y_i) puntutik zuzenera dagoen distantzia da.

Hots:

$$e_i = \frac{|\alpha x_i + \beta y_i + \gamma|}{\sqrt{\alpha^2 + \beta^2}} \quad (1)$$

Dakigunez, zuzenaren ekuazio inplizitu desberdinak idatz ditzakegu zuzen bakar bat adierazteko.

Honen ondorioz, ondoko berdintasuna betetzen duten α , β koefizienteak aukera ditzakegu.

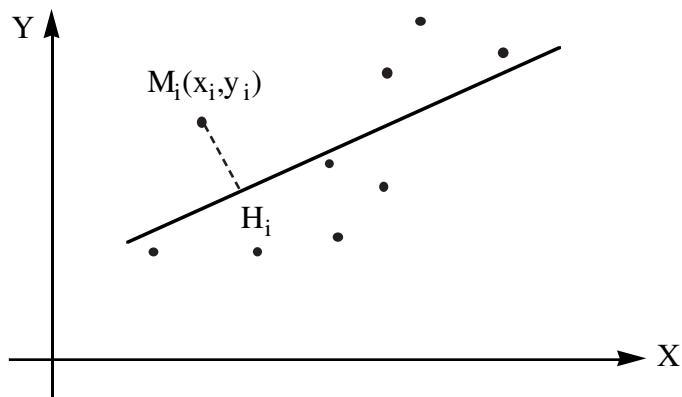
$$\alpha^2 + \beta^2 = 1$$

Betebehar baliokideak dituzten aldagaien erlazio linealik hoberena bilatzen dugu, konkretuki, puntu-hodeiak zuzen horrekiko edukiko duen sakabanatzerik txikiena.

Zuzenaren ekuazio inplizitutik abiatuz, konbinazio lineal horren koefizienteak finkatu behar ditugu, osotzen dute $[\alpha, \beta]$ bektorea normatua izatearen baldintzaz.

Kasu honetan, bada.

$$e_i = |\alpha x_i + \beta y_i + \gamma|$$



Batezbesteko txikienerako errore koadratikoa lortu nahi dugunez, idatz dezagun

(1) Fernandez de Trocaniz, A. (1979): *Matemáticas Generales*, c 1.5 orrialdean edo Geometria Analitikozko edozein testuliburutan.

bere adierazpena.

Ondoren, minimo baldintzatu bat ebatzi behar dugu, $\alpha^2 + \beta^2 = 1$ baldintzaz

$$\overline{e_i^2} = \frac{1}{m} \sum_i e_i^2 = \frac{1}{m} \sum_i \overline{M_i H_i}^2 = \frac{1}{m} \sum_i (\alpha x_i + \beta y_i + \gamma)^2$$

errorea idatzi baitugu.

Lagrange-ren biderkatzaileen metodoa erabiliz:
txikiena izateko baldintzaz $\alpha, \beta, \gamma, \lambda$ balioak mugatuko ditugu.

$$\mathcal{L} = \frac{1}{m} \sum_i (\alpha x_i + \beta y_i + \gamma)^2 - \lambda (\alpha^2 + \beta^2 - 1)$$

Dakigunez, balioekiko deribatu partzialak zero egitea baldintza beharrezkoa da $\mathcal{L}$ txikiena izan dadin.

Lehenbizi γ -rekiko deribatu partziala egiten dugu.
Hau da:

$$\frac{\partial \mathcal{L}}{\partial \gamma} = 2 \frac{1}{m} \sum_i (\alpha x_i + \beta y_i + \gamma) = 0$$

Ondoko korolararioak lortzen ditugularik:

$$\frac{1}{m} \sum_i (\alpha x_i + \beta y_i + \gamma) = 0$$

I) Erregresio-erroreen edo hondarren batezbesteko aritmetikoa zero da.

Hots: $\bar{e} = 0$

II) Erregresio-zuzena hodeiaren grabitate-zentrutik pasatzen da.
Aldagai zentratutarako erregresio-koefizienteak.

$$\alpha \bar{x} + \beta \bar{y} + \gamma = 0 \qquad \gamma = -\alpha \bar{x} - \beta \bar{y}$$

Lagrangearrean γ ordezkatzeko α, β balioekiko deribatuak egin aurretik, baina ekuazioen sinpletasuna bilatuz X, Y aldagaiak zentratuak kontsideratuko ditugu emaitzak ez baitira aldatzen.

Orduan: $\bar{x} = \bar{y} = \gamma = 0$

Hots:

Deribatuak, bada:

$$\overline{e_i^2} = \frac{1}{m} \sum_i (\alpha x_i + \beta y_i)^2 \quad \alpha^2 + \beta^2 = 1 \text{ baldintzaz}$$

$$\mathcal{E} = \frac{1}{m} \sum_i (\alpha x_i + \beta y_i)^2 - \lambda (\alpha^2 + \beta^2 - 1)$$

Hau da:

$$\frac{\partial \mathcal{E}}{\partial \alpha} = 2 \frac{1}{m} \sum_i (\alpha x_i + \beta y_i) x_i - 2\lambda \alpha = 0$$

$$\frac{\partial \mathcal{E}}{\partial \beta} = 2 \frac{1}{m} \sum_i (\alpha x_i + \beta y_i) y_i - 2\lambda \beta = 0$$

Edota:

$$\frac{1}{m} \sum_i (\alpha x_i + \beta y_i) x_i - \lambda \alpha = 0$$

$$\frac{1}{m} \sum_i (\alpha x_i + \beta y_i) y_i - \lambda \beta = 0$$

$$\alpha S_x^2 + \beta S_{xy} = \lambda \alpha$$

$$\alpha S_{xy} + \beta S_y^2 = \lambda \beta$$

Matrizialki:

$$\begin{bmatrix} S_x^2 & S_{xy} \\ S_{xy} & S_y^2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \lambda \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

Hots, laburki idatziz: $\mathbf{L} \mathbf{u} = \lambda \mathbf{u}$

Dakusagunez α, β koefizienteek kobariantz matrizearen autobektore bat osatzen dute; honi dagokion autobalioa λ Lagrange-ren koefizientea izanik.

Hau da

Azkenik $\frac{\partial \mathcal{L}}{\partial \lambda}$ deribatua eginez, $\alpha^2 + \beta^2 = 1$ daukagu.

$$\| \mathbf{u} \| = (\alpha^2 + \beta^2)^{1/2} = 1$$

Orduan:

$\mathbf{u} = [\alpha, \beta]$ kobariantz matrizearen autobektore unitarioa da.

Matrize-algebran, kobariantz matrizea bezalako matrize simetrikoak autobalio errealak eta autodirekzioak ortogonalak dituela frogatzen da.

Suposa dezagun kobariantz matrizearen autobalioak λ_1, λ_2 direla $\lambda_1 \geq \lambda_2$ izanik.

Zein auto-direkziori dagokio $\mathbf{u}$ autobektore unitarioa?

Zalantza hau askatzeko errore koadratikoen batezbestekoa, hau da hondar-bariantza ($\bar{\epsilon} = 0$ baita) lortuko dugu.

Berriro ere aldagai zentratuetarako kontsideratuz.

Hots:

$$\begin{aligned} S_e^2 &= \overline{e^2} = \frac{1}{m} \sum (\alpha x_i + \beta y_i)^2 = \\ &= \frac{1}{m} \sum (\alpha^2 x_i^2 + \beta^2 y_i^2 + 2\alpha\beta x_i y_i) = \\ &= \alpha^2 s_x^2 + \beta^2 s_y^2 + 2\alpha\beta s_{xy} \end{aligned}$$

Hondar-bariantza, bada, $\mathbf{u}$ autobektoreari dagokion λ autobalioa da eta

$$S_e^2 = [\alpha \ \beta] \begin{bmatrix} S_x^2 & S_{xy} \\ S_{xy} & S_y^2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

$$= \begin{bmatrix} \alpha & \beta \end{bmatrix} \lambda \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \lambda (\alpha^2 + \beta^2) = \lambda$$

txikiena bilatzen ari bait gara.

ondorioz $\lambda = \lambda_2$, S_e^2 Dagoeneko, emaitza hauek azpimarra ditzakegu:

- Erregresio-zuzen ortogonalaren koefizienteak α_2 , β_2 dira, kobariantz matrizearen $\mathbf{u}_2$ autobektore unitarioaren osagaiak hain zuzen.

- $\mathbf{u}_2$ autobektoreari dagokion autobalioa λ_2 hondar-bariantza neurtzen duen balioa da.

Dakusagunez, $\alpha x + \beta y = 0$ erregresio-zuzen ortogonalaren koefizienteak α_2 , β_2 dira, konkretuki, zuzen honekin ortogonal edo perpendikularra den direkzioaren koefizienteak.

Azal dezagun:

Puntu-hodeiaren doikuntza ortogonal egitean, hodeia $\alpha x + \beta y = 0$ zuzenaren bidez ordezkaturik daukagu, karratu txikiaren zentzuan puntuetatik zuzenera doazen distantzia euklidearrak minimoak egiten ditugularik.

Atera dugun emaitza $S_e = \lambda_2$ izanik, berari dagokion $\mathbf{u}_2$ autodirekzioa e_i errorearen direkzio bera da eta hauek erregresio-zuzen ortogonalarekiko perpendikularrak dira.

$\mathbf{u}_2$ autobektorearen koefizienteak, bada, erregresio-zuzen ortogonalaren koefizienteak dira, erregresio-zuzen hau $[\alpha_2, \beta_2]$ bektorearekin ortogonalak diren $[x, y]$ bektoreen leku geometrikoa baita.

Hots:

Noski, problema planteatzean α , β koefizienteak zirenak, α_2 , β_2 deitzen

$$\begin{bmatrix} \alpha_2 & \beta_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0 \Leftrightarrow \alpha_2 x + \beta_2 y = 0$$

ditugu, λ_2 autobalioari dagokion autobektorea osatzen baitute.

Bestalde zuzena barizentrutik pasatzen da, jatorritik hain zuzen X,Y aldagaiak zentratuak ditugunez.

Erregresio-zuzen ortogonalak aldagaiak zentratu gabe izanik $\alpha_2 (x - \bar{x}) + \beta_2 (y - \bar{y}) = 0$ da $(\bar{x}, \bar{y})$ puntuak barizentrua finkatuz.

Baina, $\mathbf{L}\mathbf{u} = \lambda\mathbf{u}$ sistemak bi ebazpide dauzka lehen aditzera eman dugun bezala λ_1, λ_2 autobalioei dagozkienak.

Dakigunez, kobariantz matrizeak simetrikoak eta definitu positiboak edo erdidefinitu positiboak dira, horrexegatik λ_1, λ_2 positiboak (edo oso gutxitan $\lambda_2 = 0$) eta $\mathbf{u}_1, \mathbf{u}_2$ autobektoreak ortogonalak dira.

Hau da:

$$\mathbf{u}_1^T \mathbf{u}_2 = 0$$

Eta ondorioz $\alpha_1 = -\beta_2$ $\beta_1 = \alpha_2$ dira.

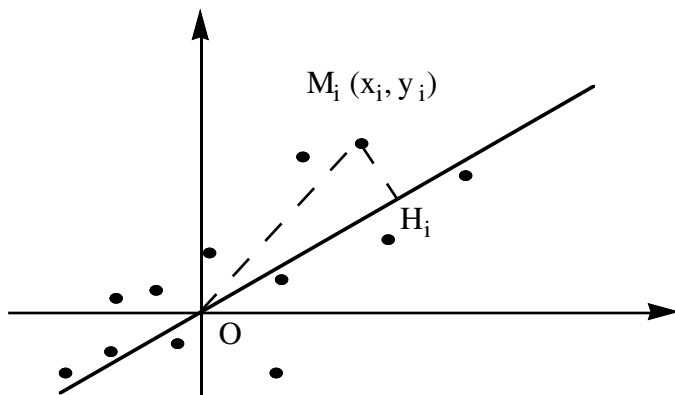
$$\begin{bmatrix} \alpha_1 & \beta_1 \end{bmatrix} \begin{bmatrix} \alpha_2 \\ \beta_2 \end{bmatrix} = 0$$

λ_1 autobalio masimoari dagokion $\mathbf{u}_1$ autobektorearen direkzioa erregresio-zuzen ortogonalaren direkzioa eta bide batez, bere balioak berizentrutik pasatzen den beronen zuzen perpendikularren koefizienteak dira. Hau da, zuzen honen ekuazioa $-\beta_2 x + \alpha_2 y = 0$ da.

$\lambda_1, \mathbf{u}_1$ bikoteak, sistemaren ebazpide edo soluzio bezala, hondar-bariantza masimoa duen zuzena finkatzen du.

Dakusagun ondoko grafikoaren laguntzaz:

(x_i, y_i) puntu bakoitzetik jatorrira daukagun distantzia karratua bi distantzia karratutan deskonposatzen da, proiektzioarenean eta zuzenera daukanean, hau da:



Eta puntu bakoitzetik jatorrira dagoen distantzia finkoa denez, zuzenera ditugun distantzia karratuak batezbestekoz minimizatzea edo proiekzioen karratuak batezbestekoz masimizatzea baliokidea da.

$$\overline{OM}_i^2 = \overline{OH}_i^2 + \overline{M}_i H_i^2$$

Dakusagunez, bada, minimo edo masimoaren baldintza beharrezkoak erabiliz $\mathbf{Lu} = \lambda \mathbf{u}$ sistemara iristen gara eta honek bi puntu ditu ebazpidetzat, bata minimo eta masimoa bestea; lehenengoak hondar-bariantza minimoa eta hodeiaren forma luzatua biltzen duen zuzena ematen dizkigu, bigarrenak, ordea, zuzen horren perpendikularra eta hondar-bariantza masimoa duena, hain zuzen.

Hau da, konkretuki eta notazioaren sinpletasuna bilatuz, hurrengo ataletan datozen formulazio matematikoetan $F = \alpha x + \beta y = 0$ erregresio-zuzen ortogonalak eta $F^\perp = -\beta x + \alpha y = 0$ barizentrutik pasatzen den beronen zuzen perpendikularra izango dira.

XI.3. BATERAKO BARIANTZAREN DESKONPOSAKETA ERREGRESIO ORTOGONALEAN

PUNTU-HODEIAREN INERTZIA

“Inertzia” kontzeptua beste hainbat bezala (grabitate-zentrua, bektorea...) Fisikatik datorkigu.

Edozein P punturekiko “inertzi momentua” Fisikan honela lortzen da; partikulen masen eta berauetako bakoitzetik P puntura dagoen distantzia koadratuen biderkaketa eta biderkaketa hauen batukariaz, hain zuzen.

Guretzat, edozein punturekiko puntu-hodeiaren “inertzi momentua” edo “inertzia” hauxe izango da: hodeiaren puntuetatik puntu horretara dauden distantzia koadratuen batukaria.

Halaber, grabitate-zentruarekiko hodeiaren “inertzia” puntuetatik grabitate-zentrura dauden distantzia koadratuen batukaria da.

Hau da:

Eta X, Y aldagaiak zentratuak direnean:

Kasu honetan, grabitate-zentrua jatorria denez, “inertzia” grabitate-zentruarekiko eta jatorriarekiko batera etortzen dira.

$$I = \sum_i (x_i - \bar{x})^2 + \sum_i (y_i - \bar{y})^2$$

$$I = \sum_i x_i^2 + \sum_i y_i^2$$

PUNTU-HODEIAREN BATERAKO BARIANTZA

Hodeiaren puntuak grabitate-zentruarekiko daukaten batezbesteko inertziari “baterako bariantza” deritzogu.

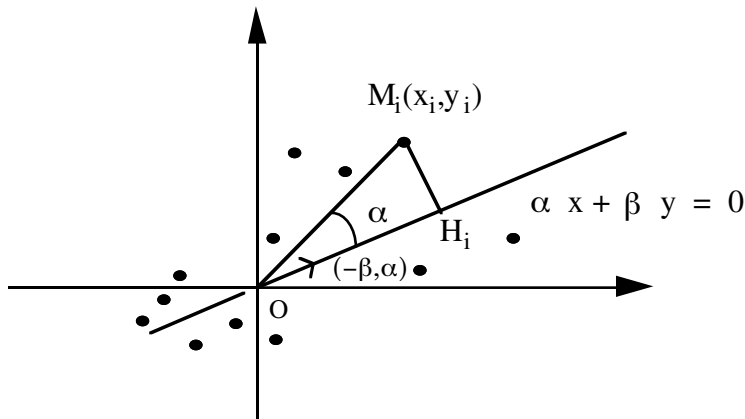
Hau da:

Aldagaiak zentratuak direnean:

$$\begin{aligned} S_{(x,y)}^2 &= \frac{I}{m} = \frac{1}{m} \sum_i (x_i - \bar{x})^2 + \frac{1}{m} \sum_i (y_i - \bar{y})^2 = \\ &= S_x^2 + S_y^2 = \text{tr } \mathbf{L} \quad (\mathbf{L}\text{-ren Aztarren}) = \frac{1}{m} \text{tr } \mathbf{C}^T \mathbf{C} \end{aligned}$$

F zuzen zentratu baten gainean $\mathbf{x} = (x_i \ y_i)$ puntuaren O H_i proiektzioa, F zuzena, $\alpha x + \beta y = 0$ eta beronen bektore unitarioa $\mathbf{u} = (-\beta, \alpha)$ izanik, ondokoa izango da.

$$S_{(x,y)}^2 = \frac{1}{m} \sum_i x_i^2 + \frac{1}{m} \sum_i y_i^2 = S_x^2 + S_y^2 = \text{tr } \mathbf{L}$$



Orduan $\langle \mathbf{u} | \mathbf{x} \rangle = \|\mathbf{x}\| \cos \alpha = O H_i$

Hots: $O H_i = \langle \mathbf{u} | \mathbf{x} \rangle \quad \mathbf{u}^T \mathbf{x} = -\beta x_i + \alpha y_i$

$$O H_i = \mathbf{u}^T \mathbf{x} = \begin{bmatrix} -\beta, \alpha \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} = -\beta x_i + \alpha y_i$$

Alde batetik $O H_i = \|\mathbf{x}\| \cos \alpha$ da
 Dakusagun: Beste batetik $\cos \alpha = \frac{\langle \mathbf{u} | \mathbf{x} \rangle}{\|\mathbf{u}\| \|\mathbf{x}\|} = \frac{\langle \mathbf{u} | \mathbf{x} \rangle}{\|\mathbf{x}\|}$

Eta F zuzenean proiektatutako hodeiaren inertzia hauxe izango da:

Proiektatutako hodeiaren bariantza, ordea, I_F/m izango da.

Ikusi genuenez, (x_i, y_i) puntutik koefiziente normatuak duen F zuzenari

$$I_F = \sum_i O H_i^2 = \sum_i (-\beta x_i + \alpha y_i)^2$$

dagoen distantzia erregresio ortogonalaren e_i hondarra da.

Hau da: $e_i = | \alpha x_i + \beta y_i |$

Halaber, erregresio ortogonalean hodeiaren hondar-inertzia edo hodeiaren F -rekiko hondar-inertzia kontsideratuko dugu.

Dakusagun, orain, F zuzenean proiektatutako hodeiaren inertziak eta hondar-inertziak hodeiaren baterako inertzia edo grabitate-zentruarekiko inertzia osatzen dutela.

$$I_E = \sum_i e_i^2 = \sum_i (\alpha x_i + \beta y_i)^2$$

Hots:

Hau da, hodeiaren baterako inertzia, F zuzenean proiektatutako inertzian eta hondar-inertzian deskonposatzen da batukorki.

Era berean puntuen kopuruaz aurreko berdintasuna zatitzen badugu, hauxe esan daiteke:

$$\begin{aligned} I_F + I_E &= \sum_i (-\beta x_i + \alpha y_i)^2 + \sum_i (\alpha x_i + \beta y_i)^2 = \\ &= \sum_i (\beta^2 x_i^2 + \alpha^2 y_i^2 - 2\alpha\beta x_i y_i) + \\ &\quad + \sum_i (\alpha^2 x_i^2 + \beta^2 y_i^2 + 2\alpha\beta x_i y_i) = \\ &= \sum_i (\alpha^2 + \beta^2)(x_i^2 + y_i^2) = \sum_i (x_i^2 + y_i^2) = \\ &= \sum_i x_i^2 + \sum_i y_i^2 = I \end{aligned}$$

Hodeiaren baterako bariantza, F zuzenean proiektatutako bariantzan eta hondar-bariantzan deskonposatzen da batukorki.

Hots:

Baina, Algebran ikasten dugunez $\mathbf{L}$ matrizea eta $\mathbf{D}$ matrize diagonalaren antzekoak dira, biek aztarren berdina daukatelarik.

Ondorioz, bada: $\text{tr } \mathbf{L} = \lambda_1 + \lambda_2$

$$S_{(x,y)}^2 = \frac{I}{m} = \frac{\sum_i x_i^2}{m} + \frac{\sum_i y_i^2}{m} = S_x^2 + S_y^2 = \text{tr } \mathbf{L}$$

Dakigunez, erregresio-zuzen ortogonalaren definizioz $\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$ hondar-bariantza txikiena duen zuzena da.

Orduan:

Baterako bariantzaz, bada, hauxe esan daiteke:

- $\mathbb{R}^2$ espazioan kontsidera daitezkeen dimentsio bakarreko azpiespazioetatik, erregresio-zuzen ortogonalaren gain proiektatutako bariantzarik handiena

$$\frac{I}{m} = S_{(x,y)}^2 = S_F^2 + S_E^2 = \lambda_1 + \lambda_2$$

duena da.

- Erregresio-zuzen ortogonalaren gain proiektatutako bariantza, kobariantz matrizearen λ_1 autobalio handiena da eta honi dagokion autodirekzioa erregresio-zuzen ortogonalaren direkzioa da.

Lortu ditugun ondorioak oso garrantzitsuak dira, $\mathbb{R}^2$ espazioan ikusi dugun Erregresio Ortogonalaren eta espazio anizkoitzean Osagai Nagusizko Analisiaren arteko zubi naturala baitira.

XI.4. ERREGRESIO ORTOGONALAREN PROPIETATEAK

1. F proiektzioen multzoa eta E hondarren multzoa aldagai estatistiko zentratuak dira.

Dakusagun:

Aldagaiak zentratuak izatean $\sum x_i, \sum y_i$ zero baitira.

Lagrangearraren γ -rekiko deribatua egitean $\bar{e} = 0$ berdintasuna lortu genuen korolariora gisa.

$$\tilde{f} = \frac{1}{m} \sum_i (-\beta x_i + \alpha y_i) = -\beta \frac{1}{m} \sum x_i + \alpha \frac{1}{m} \sum y_i = 0$$

2. F proiektzioak eta E hondarrak aldagai estatistiko koerlazio gabeak dira.

$$f_i = -\beta x_i + \alpha y_i$$

$\forall i$ denez

$$F = -\beta x + \alpha y$$

X,Y aldagaien konbinazio lineala da

Lagrangearraren α eta β -rekiko deribatuak berdin zero egitean, S_{XE} , S_{YE} kobariantzak genituen.

Eta konbinazio linealaren kobariantza aplikatuz.

Dakusagunez, bada, F, E koerlazio gabeak dira.

$$S_{XE} = \frac{1}{m} \sum (\alpha x_i + \beta y_i) x_i = \alpha \lambda$$

$$S_{YE} = \frac{1}{m} \sum (\alpha x_i + \beta y_i) y_i = \beta \lambda$$

$$S_{FE} = S_{(-\beta) x + \alpha y, E} = -\beta S_{XE} + \alpha S_{YE} = -\beta \alpha \lambda + \beta \alpha \lambda = 0$$

XII. DATU ANIZKOITZEN ANALISIA

XII.1. SARRERA

XII.2. ANALISI OROKORRA

XII.2.1. $\mathbb{R}^n$ espazioaren $\mathbb{R}^q$ azpiespazio baten bidez egindako doikuntza

XII.2.2. $\mathbb{R}^m$ espazioaren $\mathbb{R}^q$ azpiespazio baten bidez egindako doikuntza

XII.2.3. $\mathbb{R}^n$ eta $\mathbb{R}^m$ espazioen arteko erlazioa

XII.3. OSAGAI NAGUSIZKO ANALISIA

XII.3.1. $\mathbb{R}^n$ espazioan egindako analisia

XII.3.2. $\mathbb{R}^m$ espazioan egindako analisia

XII.3.3. Baterako aurkezpen grafikoak

XII.3.4. Adibidea

XII.4. KORRESPONDENTZI ANALISI

FAKTORIALA

XII.4.1. Hodeiak, masak eta distantziak

XII.4.2. $\mathbb{R}^n$ espazioan egindako analisia

XII.4.3. $\mathbb{R}^m$ espazioan egindako analisia

XII.4.4. $\mathbb{R}^n$ eta $\mathbb{R}^m$ espazioen arteko erlazioa

XII.4.5. Maiztasun-taularen berreraketa

XII.4.6. Interpretaziorako laguntzak

XII.4.7. Korrespondentzia Anitzeko Analisi

Faktoriala

XII.5. ELEMENTU GEHIGARRIAK

XII.6. SAILKAPEN-METODOAK

XII.6.1. Goranzko Sailkapen Hierarkikoa

XII.6.2. Adibidea

XII.1. SARRERA

Azken hogeï edo hogeitamar urteetan, konputagailu eta kalkulu automatikoarekin batera, estatistika deskribatzaile anizkoitza asko garatu da J.P. Benzecri-ren eskola frantsesan.

Datu-Analisiak ezaguten diren metodo hauek, bi motakoak dira:

- Metodo faktorialak
- Sailkapen-metodoak

Datu-multzo handi berdinei aplikatzen zaizkie, berorien ikasketa egiterakoan osagarri bezala erabiltzen direlarik.

Liburu honen lehen gaiaren sarreran esaten genuenez, estatistika deskribatzailearen helburua datu-multzo handien ikasketak egitea da, haien barne egitura gardenduz.

Ikasketa horietan, metodo faktorialak eta sailkapen-metodoak, datu-taula handien analisi bakoitzean bata besteari lagunduz, dira urrunago eramaten gaituztenak.

Zenbakizko datuen bidez, errealitate anizkoitzaz partzialki ohartu ahal badugu ere, metodo estatistiko egokiak aplikatzen dituzten kalkulu-programen bidez, errealitatean dauden estrukturak aztertu ahal ditugu.

Laburki esanez, metodo faktorialak diagonalizazio-metodoak dira, hauen bidez datutan dauden ardatz nagusi batzu eskurakoi bihurtzen direlarik.

Normalki, deskribatu nahi ditugun objektuak aurkezpen grafiko jarraietan kokatzen ditugu, ondoren, taldekatzen saiatzen gara puntu-hodeian dauden konstelazioak ikusi ahal izateko eta sailkapen metodoen bidez eskurakoi bihurtzen dira.

Sailkapen-metodoak, metodo algoritmikoak dira, egokia den algoritmo bat era errepikariaz aplikatuz oharpenen edo indibiduen sailak edo klaseak lortzen ditugu.

Gai honetan, metodo faktorialen oinarrian dagoen Analisi orokorra aurkezten da, ondoren, Osagai Nagusizko Metodoa eta Korrespondentzi Analisi

Faktoriala ikusiko dira, ondoren Korrespondentzia Anitzeko Analisi Faktorialaren ideia aurkezten da eta azkenik, sailkapen-metodoaz arituz Bariantzaren Metodoa ere laburki aurkeztuko da.

XII.2. ANALISI OROKORRA

Atal honetan, analisi faktorialen doikuntza-metodoek duten gune teorikoa azalduko dugu.

Edozein metrika edo Analisi Kanonikoa erabiltzeak aurkezpen dotoreagoa egiteko aukera emango liguke, baina oso utilak ez diren garapen matematikotan sartuz, Datu-Analisiaren oinarrian dagoen ideia nagusia ilunduta utziko genuke ikasleentzat.

Idea hau da: Datu-matrize batek, bi espazio bektorial desberdinetan, hodei-itxurako aurkezpen grafikoak egiteko aukera ematen du eta hodei horietan egindako doikuntzak, erlazio sinple batzuren bidez loturik daude.

Izan bedi $\mathbf{X}(m,n)$ datu-matrizea. Bere elementu orokorraz adieraziz $\mathbf{X} = [x_{ij}]$ izango da eta m,n balioez osaturik dago.

m,n balio-kopurua oso handia izan daiteke eta honen aurrean gure galdera hau da: Aurkez ditzakegu m,n balioak balio-kopuru txiki baten bidez taulan daukagun sakabanatze adieragarrienaz jabetuz?

Demagun $\mathbf{u}_1$ eta $\mathbf{v}_1$, hurrenez hurren, m eta n osagai dituzten zutabe-bektoreak direla eta berorien bidez $\mathbf{X}$ datu-matrizea $(m+ n)$ balio horien bidez berrerratu ahal dugula.

$$\mathbf{X}_{(m,n)} = \mathbf{u}_1_{(m,1)} \mathbf{v}_1^T_{(1,n)}$$

kasu honetan $\mathbf{X}$ datu-matrizearen heina 1 da. Normalki, oso zaila izango da $\mathbf{X}$ horrela berrerratea, orduan, q heinaren hurbilketa bilatuko da.

Hau da:

$$\mathbf{X} = \mathbf{u}_1 \mathbf{v}_1^T + \mathbf{u}_2 \mathbf{v}_2^T + \dots + \mathbf{u}_q \mathbf{v}_q^T + \mathbf{E}$$

$\mathbf{E}(m,n)$ hondar-matrizea izanik.

$\mathbf{X}$ datu-taularen errenkadetan $\mathbb{R}^n$ espazioko m puntuak ditugunez $\mathbf{u}$ bektore unitarioaz atzebiderkatuz F_1 zuzenaren puntu horien proiektzioak edo koordinatuak lortuko ditugu

$$\text{Hau da:} \quad \underset{(m,n)}{\mathbf{X}} \underset{(n,1)}{\mathbf{u}} = \underset{(m,1)}{\mathbf{f}}$$

$\mathbf{f}$ bektorearen m osagaiak F_1 zuzenaren gain lortzen duten luzerak dira.

$\mathbb{R}^2$ -espazioan planteatzen genuen erregresio ortogonalaren gaian gertatzen zen bezalaxe (ikus XI.2) hemen ere, $\mathbb{R}^n$ espazioan, hain zuzen, puntu bakoitzetik jatorrira dagoen distantzia karratua bi distantzia karratutan deskonposatzen da, hauek F_1 zuzenera daukan e_i distantzia karratua eta F_1 gainean daukan f_i proiektzioaren karratua direlarik.

$$\text{Irudian: } \overline{OM}_i^2 = \overline{MH}_i^2 + \overline{OH}_i^2 \text{ edo } \overline{OM}_i^2 = e_i^2 + f_i^2$$

Ondorioz, puntuek ardatzera daukaten distantzia karratuak minimizatzea (karratu txikien erizpidea) edo zuzenaren gain daukaten proiektzio karratuak maximizatzea batera ematen da.

Puntu guztiak batera hartuz:

$$\sum_i \overline{OM}_i^2 = \sum_i \overline{MH}_i^2 + \sum_i \overline{OH}_i^2$$

$$\sum_i e_i^2 \text{ minimoa edo } \sum_i f_i^2 \text{ maximoa bilatzea baliokidea da.}$$

Arazoa, bada, honela aurkez daiteke:

Zein da, $\mathbf{f}^T \mathbf{f} = \sum_i f_i^2$, proiektzio karratuen batukaria maximoa egiten duen F_1 ardatzaren $\mathbf{u}_1$ bektore unitarioa?.

$$\text{Hots: } \mathbf{f}^T \mathbf{f} = \mathbf{u}_1^T \mathbf{X}^T \mathbf{X} \mathbf{u}_1 \text{ maximoa } \mathbf{u}_1^T \mathbf{u}_1 = 1 \text{ baldintzaz.}$$

Honela, hodeia hoberen doitzen duen dimentsio bateko azpiespazioa lortzen da, hau da, $\mathbf{u}_1$ -ek sortarazten duen azpiespazioa, $\mathbf{u}_1^T \mathbf{X}^T \mathbf{X} \mathbf{u}_1$ forma koadratikoa maximoa egiten duelarik.

Hodeia hoberen doitzen duen bi dimentsioko azpiespazioak $\mathbf{u}_1$ daukala, absurdora eramanez erraz froga daiteke: $\mathbf{u}_1$ ez balu, $\mathbf{u}_1$ edukiz hobe doituko lukeen bi dimentsioko beste azpiespazioa lortuko genuke.

Honela, $\mathbf{u}_2$ bektore unitarioa eta $\mathbf{u}_1$ -ekiko ortogonal ($\mathbf{u}_2^T \mathbf{u}_2 = 1$ eta $\mathbf{u}_2^T \mathbf{u}_1 = 0$) lortuz, $\mathbf{u}_2^T \mathbf{X}^T \mathbf{X} \mathbf{u}_2$ forma koadratikoa masimoa egiten duen bi dimentsioko azpiespazioa sortaraziko da.

Analogikoki, era errepikariaz arrazonatuz zera ikusiko dugu: $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_q$ bektore unitarioek, non $\mathbf{u}_q$, aldez aurretik lortutako $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{q-1}$ bektoreekiko ortogonal baita, $q \leq n$ dimentsioko azpiespazioa sortarazten dute, $\mathbf{u}_q^T \mathbf{u}_q = 1$ baldintzaz, $\mathbf{u}_q^T \mathbf{X}^T \mathbf{X} \mathbf{u}_q$ forma koadratikoa masimoa egiten delarik.

Masimo baldintzatuaren ebazpidea: Izan bedi λ Lagrange-ren biderkatzailea, $\mathbf{u}^T \mathbf{u} = 1$ izanik, $\mathbf{u}^T \mathbf{X}^T \mathbf{X} \mathbf{u} - \lambda (\mathbf{u}^T \mathbf{u} - 1)$ adierazpenaren masimoa lortu behar dugu.

Dakigunez, optimizazio-baldintza beharrezkoak aplikatuz, $\mathbf{u}$ bektorearen osagai bakoitzarekiko deribatuak zero egin behar dira.

Osagai bakoitzarekiko deribatuak eginez, zero balioari berdinduz, eta gero ditugun ekuazioak matrizialki idatziz ondoko ekuazio matriziala daukagu: (berdin litzateke matrizialki deribatzea)

$$2 \mathbf{X}^T \mathbf{X} \mathbf{u} - 2 \lambda \mathbf{u} = 0$$

$$\text{Hau da:} \quad \mathbf{X}^T \mathbf{X} \mathbf{u} = \lambda \mathbf{u} \quad (1)$$

Dakusagunez, λ balioa eta $\mathbf{u}$ bektorea $\mathbf{X}^T \mathbf{X}$ matrize simetrikoaren autobalioa eta berari loturiko autobektorea, hurrenez hurren, dira.

(1) berdintasuna, $\mathbf{u}^T$ bektoreaz aurrebiderkatuz:

$$\mathbf{u}^T \mathbf{X}^T \mathbf{X} \mathbf{u} = \lambda$$

Hau da, bilatzen ari garen masimoa λ autobalioa da. Honela, $\mathbf{u}$ bektorea λ_1 autobalio handienari dagokion $\mathbf{X}^T \mathbf{X}$ matrizearen $\mathbf{u}_1$ autobektorea izango da.

Hoberen doitzen den bi dimentsioko azpiespazioa lortu nahi badugu, bigarren zuzen bat behar dugu $\mathbf{t}$ berorren bektore unitarioa izanik.

$$\mathbf{t}^T \mathbf{X}^T \mathbf{X} \mathbf{t} \text{ masimoa } \mathbf{t}^T \mathbf{u}_1 = 0 \text{ eta } \mathbf{t}^T \mathbf{t} = 1 \text{ izanik lortu behar dugu.}$$

Hau da: $\mathbf{t}^T \mathbf{X}^T \mathbf{X} \mathbf{t} - \lambda (\mathbf{t}^T \mathbf{t} - 1) - \mu \mathbf{t}^T \mathbf{u}_1$, bi biderkatzaile dituen lagrangearra idazten dugu.

Berriro ere, $\mathbf{t}$ bektorearen osagaiekiko deribatu partzialak zero balioari berdinduz, ondoko ekuazio matritziala daukagu.

$$2 \mathbf{X}^T \mathbf{X} \mathbf{t} - 2 \lambda \mathbf{t} - \mu \mathbf{u}_1 = 0 \quad (2)$$

(2) berdintasuna bektoreaz aurre biderkatuz, ondoren ordezkatuz eta $\mathbf{u}_1^T$ direla jakinik $\mu = 0$ dela ikusten dugu. $\mathbf{u}_1^T \mathbf{X}^T \mathbf{X} \mathbf{t} = \lambda \mathbf{u}_1^T \mathbf{t}$

Hau da: $\mathbf{u}_1^T \mathbf{t} = 0, \mathbf{u}_1^T \mathbf{X}^T \mathbf{X} \mathbf{t} = \lambda \mathbf{t}$

Dakusagunez, $\mathbf{t}$ bektore unitarioa, $\mathbf{X}^T \mathbf{X}$ matrize simetrikoaren λ_2 bigarren autobalioari loturiko $\mathbf{u}_2$ bigarren autobektorea izango da. Prozesua jarraituz, $k = 1, 2, \dots, r$ autoelementuetara edo elementu propioetara emaitza zabaltzen da.

Honela, $\mathbf{X}^T \mathbf{X} \mathbf{u}_k = \lambda_k \mathbf{u}_k \quad \forall k \leq r \quad (3)$

r , $\mathbf{X}^T \mathbf{X}$ matrizearen heina izanik.

Azkenik, karratu txikien zentzuan hodeian hoberen doitzen den q dimentsioko azpiespazioaren oinarria, $\mathbf{X}^T \mathbf{X}$ matrize simetrikoaren q handien autobalioei loturiko q autobektoreez osaturik dago.

$\mathbf{X}^T \mathbf{X}$ matrizea, simetrikoa eta definitu edo erdidefinitu positiboa izanez, hauek dituzten propietateetatik abiatuz, beste frogapen baten bidez, aurreneko emaitzara iritsiko ginateke.

XII.2.2 $\mathbb{R}^m$ espazioaren $\mathbb{R}^q$ azpiespazio baten bidez egindako doikuntza

Era berean, $\mathbb{R}^m$ espazioan ditugun n zutabe-puntuak $\mathbb{R}^q$ azpiespazioko ardatzetan dituzten koordenatuen bidez eta ardatz berri hauek dituzten osagaien bidez berrera daitezke.

Izan bedi G_1 jatorritik pasatzen den zuzena eta $\mathbf{v}$ beronen direkzio-kosinuen bektorea, G_1 zuzenaren gainean puntuen proiektzioen karratuen batura masimoa egitean, hodeian hoberen doitzen den zuzena, karratu txikien zentzuan lortzen dugu.

Hau da: $\mathbf{X}^T \mathbf{v} = \mathbf{g}$

$$(n,m)(m,1) \quad (n,1)$$

$\mathbf{g}$ zutabe-bektorearen n errenkadak proiektzioen n balioak dira, orduan, $\mathbf{v}^T \mathbf{X} \mathbf{X}^T \mathbf{v}$ forma koadratikoa, masimoa egin nahi dugun balioa, izango da

$\mathbf{v}^T \mathbf{v} = 1$ baldintzaz, $\mathbf{v}^T \mathbf{X} \mathbf{X}^T \mathbf{v}$ masimoa egiten duen $\mathbf{v}$ bektorea aurkituz, problema ebatzia edukiko dugu.

$\mathbb{R}^n$ espazioaren kasuan bezala, matrize simetriko baten diagonalizazioaren aurrean gaude, $\mathbb{R}^m$ espazioan, $\mathbf{X} \mathbf{X}^T$ matrize simetrikoaren autobektorea $\mathbf{v}$ baita.

μ_1 autobalio handienari dagokion autobektorea $\mathbf{v}_1$ izango da. Ondoren, $\mathbf{v}_2 \dots \mathbf{v}_q$ finkatuz, hodeian hoberen doitzen den q dimentsioko azpiespazioa sortarazten duten bektoreak lortu ditugu.

Dakigunez, $\mathbf{X}^T \mathbf{X}$ eta $\mathbf{X} \mathbf{X}^T$ matrizeak $\mathbf{X}$ matrizearen heina berbera daukate r , hain zuzen, eta $q < r$ izango da.

XII.2.3 $\mathbb{R}^n$ eta $\mathbb{R}^m$ espazioen arteko erlazioa

$\mathbf{v}_k$ eta $\mathbf{u}_k$ bektoreen arteko erlazioak aurkitu nahi ditugu.

$$\text{Definizioz} \quad \mathbf{X} \mathbf{X}^T \mathbf{v}_k = \mu_k \mathbf{v}_k \quad (4)$$

μ_k eta $\mathbf{v}_k$, $\mathbf{X} \mathbf{X}^T$ matrizearen k . autobalioa eta berari loturiko k . autobektorea, hurrenen hurren, izanik.

Diagonalizazioaren bidez, desberdin zero diren r autobalio lortuko dira eta $r \leq (m,n)$, hau da, $\mathbf{X}$ matrizearen dimentsiorik txikiena baino txikiagoa edo berdina izango da r , $\mathbf{X}$ matrizearen heina r delako.

(4) berdintasuna $\mathbf{X}^T$ matrizeaz aurrebiderkatuz:

$$\mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{v}_k) = \mu_k (\mathbf{X}^T \mathbf{v}_k)$$

Baina, $\mathbf{X}^T \mathbf{X}$ diagonalizazioan $\mathbf{X}^T \mathbf{X} \mathbf{u}_k = \lambda_k \mathbf{u}_k$ ((3) berdintasuna) daukagunez, $\mathbf{X} \mathbf{X}^T$ matrizearen $\mathbf{v}_k$ bakoitzari ($k \leq r$) $\mathbf{X}^T \mathbf{X}$ matrizearen $\mathbf{X}^T \mathbf{v}_k$ dagokio μ_k autobalio berdinari lotuta.

Dakusagunez, $\mathbf{u}_k = h_k \mathbf{X}^T \mathbf{v}_k$ ($h_k = \text{konstantea}$) idatz dezakegu bektore

proportzionalak baitira.

λ_k k. autobalioa, handiena izatean, nahitaez:

$$\mu_k \leq \lambda_k \quad \text{izango da, eta } \{\mu_k\} \subset \{\lambda_k\}$$

Hots, μ_k autobalioen multzoa λ_k autobalioen multzoaren azpimultzoa da.

Era berean, (3) berdintasuna $\mathbf{X}$ matrizeaz aurrebiderkatuz:

$$\mathbf{X} \mathbf{X}^T (\mathbf{X} \mathbf{u}_k) = \lambda_k (\mathbf{X} \mathbf{u}_k)$$

Honela, $\mathbf{X}^T \mathbf{X}$ matrizearen $\mathbf{u}_k$ bakoitzari $\mathbf{X} \mathbf{X}^T$ matrizearen $\mathbf{X} \mathbf{u}_k$ dagokio λ_k autobalio berdinari lotuta.

Orduan, $\mathbf{v}_k = h_k' \mathbf{X} \mathbf{u}_k$ (h_k konstantea) eta μ_k , $\mathbf{X} \mathbf{X}^T$ matrizearen k. autobalio handiena denez

$$\lambda_k \leq \mu_k \quad \text{eta } \{\lambda_k\} \subset \{\mu_k\}$$

Hots: $\lambda_k = \mu_k \quad \forall k \leq r$ idatz dezakegu.

eta $\mathbf{X}^T \mathbf{X} \mathbf{u}_k = \lambda_k \mathbf{u}_k$ ordezkatuz.

Dakigunez, $\mathbf{u}_k^T \mathbf{u}_k = 1, \quad \mathbf{v}_k^T \mathbf{v}_k = 1$

Honela, $\mathbf{v}_k^T \mathbf{v}_k = h_k^2, \quad \mathbf{u}_k^T \mathbf{X}^T \mathbf{X} \mathbf{u}_k$

Analogikoki, dela lortzen da.

$$\mathbf{v}_k^T \mathbf{v}_k = h_k^2, \quad \mathbf{u}_k^T \lambda_k \mathbf{u}_k = h_k^2 \lambda_k = 1$$

$$h_k' = \frac{1}{\sqrt{\lambda_k}}$$

Orduan, $h_k = \frac{1}{\sqrt{\lambda_k}}$

$\mathbf{u}_k$ bektore unitarioaren euskarria den F_k ardatza, $\mathbb{R}^n$ espazio bektorialaren k.

ardatz faktoriala deitzen da eta $\mathbf{v}_k$ bektore unitarioaren euskarria den G_k ardatza

$$\mathbf{u}_k = \frac{1}{\sqrt{\lambda_k}} \mathbf{X}^T \mathbf{v}_k \quad (5)$$

(n,1) (n,m)(m,1)

$$\mathbf{v}_k = \frac{1}{\sqrt{\lambda_k}} \mathbf{X} \mathbf{u}_k \quad (6)$$

(m,1) (m,n)(n,1)

$\mathbb{R}^m$ espazio bektorialaren k . ardatz faktoriala, hain zuzen ere, biek, λ_k autobalio berdinari dagokion autodirekzioa finkatzen dutelarik.

Dakusagunez, $\mathbb{R}^n$ (edo $\mathbb{R}^m$)-ren k .ardatzaren gain hodeiaren puntuek dituzten koordenatuak $\mathbf{X} \mathbf{u}_k$ (edo $\mathbf{X}^T \mathbf{v}_k$) dira. Orduan, (5), (6) erlazioetan ikus daitekeenez, koordenatu horiek espazio batean eta ardatzaren osagai unitarioak beste espazioan proportzionalak dira.

XII.2.4 X datu-taularen berreraketa

Dakusagun, nola, $\mathbb{R}^n$ eta $\mathbb{R}^m$ espazioetan ditugun hodeien puntuak beraien posizioetan, gutxi gora behera, berreratzen ditugun karratu txikiaren zentzuan hoberen doitzen den azpiespazio baten oinarrian dituzten koordenatuen bidez.

Suposa dezagun, λ_1 , lehen autobalioa, besteak baino askoz garrantzitsuagoa dela, kasu honetan, balioak proiektzioz jatorriarekiko distantzia karratuak neurtzen ditu eta lehen ardatzak puntuen posizioak ongi berreratuko ditu. Normalki, q lehen ardatz faktorialen bidez, “hurbilketa onaz”, puntuen posizioak berreratuko ditugu, horretarako $\lambda_1 + \lambda_2 + \dots + \lambda_q$ baturak, $\mathbf{X}^T \mathbf{X}$ matrizearen aztarnaren proportzio λ_1 gaurkiz λ_1 aurkeztu behar du. $\mathbf{X}^T \mathbf{X}$ matrize simetrikoa eta, normalki, definitu positiboa izango denaren $n = r \leq m$ autobalioak errealak dira eta autobektoreak ortogonalak.

$\mathbf{U}$ autobektore unitarioak osatutako matrize diagonalizatzaila eta $\mathbf{X}^T \mathbf{X}$ matrizea antzekoak dira, hau da, oinarri desberdinetan transformazio lineal berbera definitzen dute.

Hots, $\mathbf{\Lambda}$ matrizea, $\lambda_1, \lambda_2 \dots \lambda_n$ autobalioak osaturiko matrize diagonal izanez, $\mathbf{X}^T \mathbf{X} \mathbf{U} = \mathbf{U} \mathbf{\Lambda}$ eta $\mathbf{U}^T \mathbf{X}^T \mathbf{X} \mathbf{U} = \mathbf{U}^T \mathbf{U} \mathbf{\Lambda}$ edota $\mathbf{U}^{-1} \mathbf{X}^T \mathbf{X} \mathbf{U} = \mathbf{\Lambda}$ $\mathbf{U}$ ortogonal baita.

$$\begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

Dakusagunez, bada, $\mathbf{X}^T \mathbf{X}$ matrizearen antzekotasun¹ez transformatu diagonalan da $\mathbf{\Lambda}$ matrizea eta honelako matrize antzekoen propietate bat aplikatuz¹:

$$\text{tr}(\mathbf{X}^T \mathbf{X}) = \text{tr} \mathbf{\Lambda}$$

Hasierako zenbakizko balioak, hau da, $\mathbb{R}^n$ espazioan puntuak dituzten koordenatuak, q ardatzen bidez lortu nahi izanez gero, ardatz faktorialetan dituzten koordenatuak eta ardatzen osagai unitarioak (direkzio-kosinuak) batera erabiliz lortuko ditugu.

(6) erlazioa honela idatziz:

$$\text{tr}(\mathbf{X}^T \mathbf{X}) = \sum_i \sum_j x_{ij}^2 = \sum_k \lambda_k = \text{tr} \mathbf{\Lambda} \quad (7)$$

Ondoren, berdintasuna bektoreaz atzebiderkatuz.

eta k balio guztientzako batukaria hartuz (autobalio nuluak balira beraiei dagozkien

$$\mathbf{u}_k^T \mathbf{X} \mathbf{u}_k = \sqrt{\lambda_k} \mathbf{v}_k^T \mathbf{u}_k$$

autobektoreek $\mathbb{R}^n$ espazioaren oinarria osatuko lukete).

$$\mathbf{X} \mathbf{u}_k \mathbf{u}_k^T = \sqrt{\lambda_k} \mathbf{v}_k \mathbf{u}_k^T$$

q lehen ardatzak hartzen baditugu, lortzen dugun $\mathbf{X}^*$ matrizea $\mathbf{X}$ datu-matrizearen hurbilketa izango da.

balioak txikiak balira hurbilketa ona izango litzateke.

$$\mathbf{X} \sum_{k=1}^n \mathbf{u}_k \mathbf{u}_k^T = \sum_{k=1}^n \sqrt{\lambda_k} \mathbf{v}_k \mathbf{u}_k^T$$

$$\text{eta } \mathbf{X} = \sum_{(m,n)} \sqrt{\lambda_k} \mathbf{v}_k \mathbf{u}_k^T \quad \sum_k \mathbf{u}_k \mathbf{u}_k^T = \mathbf{U}^T \mathbf{U} = \mathbf{I} \quad (\text{unitate matrize})$$

Berreraketaren kalitatea, ondoko proportzioaz neur daiteke:

1. Lau propietate dituzte, diskriminatzailea, autobalioak, aztarrena (tr) eta heina (r) berbera edukitzea, hain zuzen. Ikus FZD. DE TROCONIZ A. *Matemáticas Generales* I, 1.18.2

$$\mathbf{X} \approx \mathbf{X}^* = \sum_{k=1}^q \sqrt{\lambda_k} \mathbf{v}_k \mathbf{u}_k^T$$

$$\sqrt{\lambda_{q+1}} \dots \sqrt{\lambda_n}$$

$\mathbf{X}$ datu-matrizearen $m \cdot n$ balioak $q(m+n) = qm + qn$ balioen bidez berrerratu ditugu, q bektoreen m balioen bidez eta $\mathbf{u}_k$, q bektoreen n balioen bidez.

$$\tau_q = \sum_{k=1}^q \lambda_k / \sum_{k=1}^n \lambda_k$$

λ_k autobalio bakoitzak k . ardatzaren gain puntuen proiektzio karratuen batura neurtzen du, hau da, k . ardatzaren gain proiektatutako hodeiaren inertzia (ikus XI.3 $n=2$ kasurako). Honela, inertzi tasa deitutako τ_q balioak q lehen ardatzetan projektatutako inertziaren proportzioa edo bariantzaren proportzioa ematen digu.

Hoberen doitutako $\mathbb{R}^q$ azpiespazioan biltzen den hodeiaren sakabanatzearen proportzioa, inertzi tasak neurtzen digu.

XII.3. OSAGAI NAGUSIZKO ANALISIA

Osagai nagusizko analisia, aztertu dugun analisi orokorraren barrutian aurkezten dugu.

Osagai nagusizko analisia “aldagai-indibiduo” taula-motatan aplikatuko da, kasu honetan, errenkada-puntuen hodeia indibiduen hodeia eta zutabe-puntuen hodeia aldagaien hodeia izango direlarik.

Datu zentratutatik abiatuz eta hodei batean nahiz bestean zentratze-eragiketak suposatzen duena kontutan hartuz, osagai nagusizko analisia bi eratara planteatu daiteke.

Osagai Nagusizko Analisia

$\mathbf{X}$ matrizearen gai orokorra denean, diagonalizatzen den matrize simetrikoa $\mathbf{X}^T \mathbf{X} = \mathbf{L}$ kobariantz matrizea da.

$$\frac{x_{ij} - \bar{x}_j}{\sqrt{m}}$$

Osagai Nagusizko Analisi Normatua.

Aldagaiaren neurriak, batezbestekoen aldetik eta sakabanatzearen aldetik heterogenoak direnean $\mathbf{X}$ matrizearen gai orokorra izango da, kasu honetan diagonalizatzen den matrize simetrikoa $\mathbf{X}^T \mathbf{X} = \mathbf{R}$ koerlazio-matrizea da.

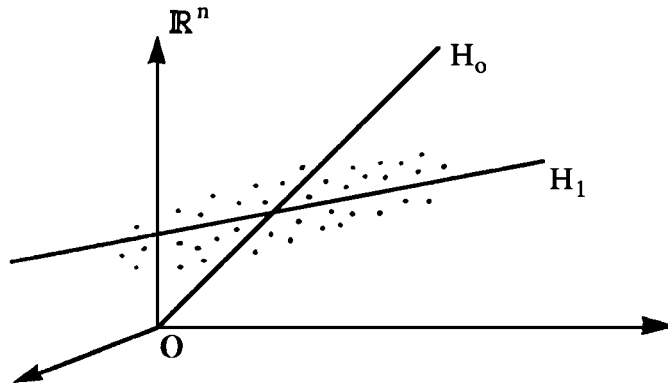
XII.3.1 $\mathbb{R}^n$ espazioan egindako analisisia

Datu zentratutatik abiatzeak dituen abantailak hauexek dira:

- Diagonalizatzen den matrizea $\mathbf{L}$ edo $\mathbf{R}$ da.
- Proiekzio karratuen batura, masimoa egiten duen lehen ardatza jatorritik pasatzen da, honela aztertu dugun analisi orokorraren barrutian azaldu ahal dugu osagai nagusizko analisisia.

Ondoko grafikoan ikus dezakegunez, hodeiaren grabitate-zentrua jatorrian ez balego, proiekzio karratuen batura masimoa egingo lukeen ardatza jatorritik ez litzateke pasatuko.

H_1 ardatzarentzako balio hori masimoa izango litzateke eta ez H_0 ardatzarentzako.



$\mathbb{R}^n$ espazioan, indibiduen hodeian egindako analisisian, grabitate-zentrua jatorrira eramaten da edota ardatzen translazio paraleloa egiten da. koefizientearen bidez, diagonalizatzen den $\mathbf{X}^T \mathbf{X}$ matrizea kobariantz matrizea izatea lortzen da.

Aldagaien sakabanatzeak oso desberdinak direnean, zentratzeaz gainera, $1/\sqrt{m}$ beste aldaketa egitea beharrezkoa da, honela, osagai normatutako analisisian indibiduen arteko distantziak aldagai desberdinekiko, ondoan ikusten den

bezala “eskalan berdindu ditugu”.

Hots:

Dakusagunez, indibiduen hurbiltasunak kalkulatzean, aldagaiek ekarpen analogoa daukate, desbidazio tipikoen arabera koordenatuak hartu bait ditugu.

$$d^2(i, i') = \sum_{j=1}^n \left[\frac{x_{ij} - \bar{x}_j}{S_j \sqrt{m}} - \frac{x_{i'j} - \bar{x}_j}{S_j \sqrt{m}} \right]^2 =$$

$$= \frac{1}{m} \sum (x_{ij} - x_{i'j})^2 / S_j^2$$

Kasu honetan, dakigunez, diagonalizatzen den $\mathbf{X}^T \mathbf{X}$ matrizea koerlazio-matrizea da.

Indibiduo-puntuak k. ardatzaren gain dituzten koordenatuak, $\mathbf{X}\mathbf{u}_k$ zutabe-bektorearen errenkadak dira, eta (6) formularen arabera.

XII.3.2 $\mathbb{R}^m$ espazioan egindako analisisa

$$\mathbf{X}\mathbf{u}_k = \sqrt{\lambda_k} \mathbf{v}_k$$

Espazio honetan ditugun n puntuak aldagai-puntuak dira. Dakigunez, $\mathbb{R}^n$ espazioan egindako analisiak $\mathbb{R}^m$ espazioaren analisisa indutitzen du. Halere, i eta j azpiindizeek betebeharrak desberdinak dituzte eta $\mathbf{X}$ datu-matrizean egindako transformazioek interpretazio geometriko desberdinak dituzte (ikus X. gaia).

$\mathbb{R}^m$ espazioan, aldagaien hodeian, zentratze-eragiketaren bidez aldagai-puntuak $\mathbb{R}^{m-1}$ azpiespazio batera proiektatzen ditugu, konkretuki, lehen erdikariarekiko proiektzio paraleloa eginez, koordenatuen baturaz zero ematen duen $\mathbb{R}^{m-1}$ azpiespaziora proiektatzen ditugu.

Aldagai-puntuak jatorrira daukan distantzia karratua aldagaiaren bariantza da: j, j' bi aldagaien arteko distantzia karratua:

Aldiz, Osagai normatutako analisisan ardatzen eskalak aldatzeak, hau da, $\mathbb{R}^m$ espazioan ditugun koordenatuak, aldagaiaren normaren balioaz zatituz,

$$d^2(j, o) = \frac{1}{m} \sum_{i=1}^m (x_{ij} - \bar{x}_j)^2 = S_j^2$$

aldagai-puntuak normatzen ditugu eta guztiak jatorriarekiko 1 distantziaz kokatzen

$$\begin{aligned} d^2(j, j') &= \frac{1}{m} \sum_{i=1}^m \left[(x_{ij} - \bar{x}_j) - (x_{ij'} - \bar{x}_{j'}) \right]^2 \\ &= S_j^2 + S_{j'}^2 - 2S_{jj'} \end{aligned}$$

dira.

$$S_j \sqrt{m}$$

Hots:

n aldagai-puntuak, koordenatuen jatorrian zentraturiko bat erradioz daukan hiperesfera batean kokatzen dira.

$$d^2(j, o) = \frac{1}{m} \sum_{i=1}^m (x_{ij} - \bar{x}_j)^2 / S_j^2 = 1$$

Ondokoak, j eta j' bi aldagai puntuen arteko distantzi karratua: karratua eginez eta batukaria aplikatuz, hauxe da ateratzen dugun emaitza:

Honela, aldagai-puntuen hurbiltasunak koerlazioen arabera interpreta

$$d^2(j, j') = \frac{1}{m} \sum_{i=1}^m \left[\frac{x_{ij} - \bar{x}_j}{S_j} - \frac{x_{ij'} - \bar{x}_{j'}}{S_{j'}} \right]^2$$

daitezke. Oso hurbilak dauden bi aldagai-puntuen arteko koerlazioa positiboa eta

$$d^2(j, j) = S_j^2 / S_j^2 + S_{j'}^2 / S_{j'}^2 - 2 r_{jj'} = 2(1 - r_{jj'})$$

sakona izango da ($r_{jj'} \simeq 1$), eta oso urrun dauden bi aldagai-puntuen koerlazioa negatiboa eta sakona izango da ($r_{jj'} \simeq -1$). Aldagai-puntuek k ardatzaren gain dituzten koordenatuak $\mathbf{X}^T \mathbf{v}_k$ zutabe-bektorean osagaiak dira, eta (5) formularen arabera

Dakusagunez, bada, aldagai-puntuen koordenatuak k ardatzaren gain, $\mathbf{R}$ koerlazio-matrizearen diagonalizazioan kalkulaturako λ_k , $\mathbf{u}_k$, kantitateen funtzioan kalkulatzeko dira.

Dakigunez, (ikus X. gaia) bi aldagai-puntuak osatzen duten angeluaren kosinua aldagaien arteko koerlazio-koefizientea da. Analisi normatuaren kasuan, aldagai-puntuak zentru-normatuak direnez, kosinua aldagaien biderkadura eskalarra da.

Aldagai-puntuek k . ardatzaren gain dituzten abzisak $\mathbf{X}^T \mathbf{v}_k$ zutabe-bektorearen osagaiak dira.

Dakusagunez, abzisa bakoitza aldagai-bektore bakoitzaren ($\mathbf{X}^T$ matrizearen errenkada) eta $\mathbf{v}_k$ bektorearen arteko biderkadura eskalarra da. Honela, aldagai-puntu bakoitzak ardatz baten gain daukan abzisa, aldagaia eta aldagai guztien konbinazio lineala den $\mathbf{v}_k$ aldagai berriaren arteko koerlazio-koefizientea da.

Analisi Orokorrean ikusten genuenez, λ_k autobalio bakoitzak k . ardatzaren gain proiektzio karratuen batura neurtzen du, hau da, k . ardatzaren gain proiektatutako hodeiaren inertzia.

$\mathbf{X}^T \mathbf{X}$ matrizeak, inertzi matrizea izanik, bere aztarrenean hodeiaren jatorriarekiko inertzia biltzen du.

Bestalde, diagonalizazioan egindako transformazio lineal ortogonalaren bidez, jatorriarekiko baterako inertzia, n ardatz ortogonaletan proiektatutako inertzitan deskonposatzen da (ikus (7) berdintasuna).

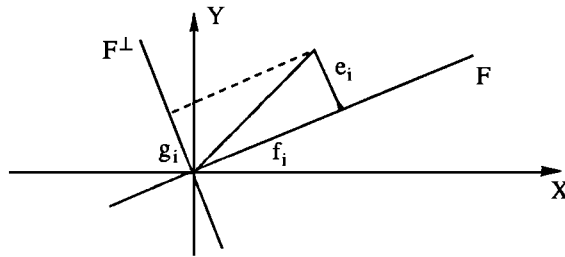
$\mathbb{R}^n$, n -dimentsioko espazio bektoriala, dimentsio bateko n azpiespazio ortogonaletan deskonposatu da, n ardatz faktorialetan, hain zuzen.

$\mathbb{R}^n$, n -dimentsioko espazio bektoriala, bi espazio ortogonaletan deskonposa dezakegu, hauek betegarriak direlarik.

Honela, q lehen ardatzetan hodeiaren sakabanatzearen zatirik adierazgarriena biltzen denean, $\mathbb{R}^n$ espazioa $\mathbb{R}^q$ eta $\mathbb{R}^{n-q}$ azpiespazio betegarritan deskonposatzen da.

$n = 2$ denean (ikus XI gaia) oso erraz ikus daiteke hodeiaren baterako inertzia F zuzen baten gainean proiektatutako inertzian eta zuzen horrekiko daukan hondar-inertzian deskonposatzen dela, konkretuki, F zuzenaren ortogonal den $F^\perp$ zuzenaren gainean proiektatutako inertzia eta F -rekiko hondar-inertzia berdinak dira.

Puntu bakoitzerako $\mathbf{g}_i = \mathbf{e}_i$ eta puntu guztientzako hodeiaren baterako inertzia batukorki deskonposatzea, Pitagoras-en teoremaren aplikatzeaz zuzenki eratortzen delarik.



$n > 2$ denean eta $\mathbb{R}^n$ espazioa $\mathbb{R}^q$ eta $\mathbb{R}^{n-q}$ azpiespazio betegarritan

$$\sum_i g_i^2 = \sum_i e_i^2$$

deskonposatuta daukagunean, hodeiaren inertzia bi espaziotan proiektatutako inertzien batura da, baina $\mathbb{R}^{n-q}$ espazioan proiektatutako inertzia eta $\mathbb{R}^q$ espazioarekiko hondar-inertzia berdinak dira.

Era berean, hodeiaren baterako inertzia batukorki deskonposatzea, “Pitagoras-en Teorema Espazioan” aplikatzeaz zuzenki eratortzen da.

Osagai Nagusien Analisisan, Erregresio Ortogonalean gertatzen zen bezala, diagonalizatzen dugun matrizea $\mathbf{L}$ kobariantz matrizea edo $\mathbf{R}$ aldagai tipifikatuen kobariantz matrizea da. Halere, aldagai transformatuen inertzi matrizeak bezala lortu direla gogoratu behar dugu.

Edozein kasutan, λ_k autobalio bakoitzak k . ardatzaren gain puntuen proiektzioen karratuen batura batezbestekoa $(\sum_j (x_{jk} - \bar{x}_j)^2) / \sqrt{n}$ edo $(\sum_j (x_{jk} - \bar{x}_j)^2) / \sqrt{n}$ aren gain proiektatutako hodeiaren bariantza.

Hodeiaren baterako bariantza:
edo

kasu honetan, koerlazio-matrizearen elementu diagonalak 1 baitira.

Era berean, q lehen ardatzetan hodeia ondo doituta kontsideratzen badugu,

$$\text{tr } \mathbf{L} = \sum_{j=1}^n S_j^2 = \sum_{k=1}^n \lambda_k = \text{tr } \mathbf{A}$$

$$\text{tr } \mathbf{R} = n = \sum_{k=1}^n \lambda_k = \text{tr } \mathbf{A}$$

proiektatutako bariantza eta

hondar-bariantza izango dira.

$$\sum_{k=1}^q \lambda_k$$

$$\sum_{k=q+1}^n \lambda_k$$

bariantz tasa izango da, proportzio horrek, askotan portzentaiatan emanik, q lehen faktoreetan proiektatutako bariantzaren portzentaia neurtzen du, edo q lehen

$$\tau_q = \frac{\sum_{k=1}^q \lambda_k}{\sum_{k=1}^n \lambda_k}$$

faktoreetan proiektatutako inertziaren portzentaia, honela, era berean, inertzi tasa deritzogu.

XII.3.3. Baterako aurkezpen grafikoak

Osagai nagusizko analisia mende honen hasieratik ezaguna da, baina, azken hogeitamar urte hauetan eduki duen bultzada oso handia izan da.

Arlo askotan aplikaturik, Soziologian, Ekonomian, Biologian, Geologian, Antropologian, Linguistikan... datu-taula anizkoitzetatik laburpen deskribatzaileak ateratzeko guztiz erabilgarria da.

Analisi-teknika honen garapena, gaiaren hasieran esaten genuen bezala, konputagailuen eta kalkulu automatikoen garapenari loturik dago, baina praktikan, erabilpenaren garapena, baterako aurkezpen grafikoetatik eratorritako interpretagarritasunean datza.

$\mathbf{X}^T \mathbf{X}$ eta $\mathbf{X} \mathbf{X}^T$ matrize simetrikoen analisisetan lortutako emaitza paraleloak, hau da, autobalio berdinak eta autodirekzio baliokideak edukitzeak, errenkada-puntuak eta zutabe-puntuak hodeiaren ebaketa-plano berean proiektatzearen ideia sortu zuen Datu-Analisiaren eskola frantzesan.

Trantsizio-erlazioak deitzen diren (5) eta (6) erlaziotan daukagu indibiduo-puntuak eta aldagai-puntuak erlazionatzeko bidea.

(6) erlazioa ikusiz zera esan dezakegu: k . ardatzaren gain koordinatu handienak dauzkaten indibiduo-puntuek koefiziente handiagoz hartzen dute parte aldagai-puntuen hodeian k . ardatza finkatzean. Eta, (5) erlazioa ikusiz, hodeia eta puntu-mota aldatuz, ondorio berdina atera daiteke.

Bi ardatz faktorialak osatutako plano $\mathbb{R}^n$ edo $\mathbb{R}^m$ espaziotan, jatorritik pasatuz egindako ebaketa da, eta lehen plano faktoriala edo plano faktorial nagusia

algoritmoaren menpe izango da. $\mathbb{R}^n$ espazioan k. ardatzaren bektore unitarioa aukeratzeko, bi posibilitate daude $\mathbf{u}_k$ edo $-\mathbf{u}_k$, hain zuzen, baina bata aukeratuz gero $\mathbb{R}^m$ espazioarena finkatuta dago, dakigunez $\mathbf{v}_k = \mathbf{h}_k$, $\mathbf{X} \mathbf{u}_k$ eta honen ondorioz (5), (6) trantsizio-erlazioak espazio baten elementuen abzisak k. ardatzaren gain, beste espazioan, k. ardatzaren bektore unitarioaren osagaien bidez ematen dizkigutenak, koherenteki aplikatutako dira.

Bi hodeien puntuak plano faktorialetan proiektatu ondoren eta adierazpide grafiko hauetatik ondorioak atera baino lehen, ondokoa kontutan hartu behar dugu:

Aldagai batzuren posizioek, inguruan dauden indibiduen hurbiltasunaren interpretazioaren ikuspegia emango digute, baina aldagai bat indibiduo baten ondoan egoteak ez du esangurarik, puntu baten kokapena beste hodeiaren puntu guztien kokapenarekin dago erlazionaturik eta hortik datorkio bere posizioa. Ez dugu inoiz ahaztu behar ikuspegi estatistikoa multzokoa delarik.

XII.3.4. Adibidea

Aurreko garapen teorikoen argibide gisa erabil daitekeen adibidea hauxe da:

Suposa dezagun Hego eta Ertamerikako ikasketa bat egin nahiean, hamazortzi estatutako sei aldagaien ohartutako balioak ditugula.

Aukeratu ditugun ondoko aldagaietan, munduko kontinente honen errealitatea partzialki ohartzen dugu, halere, errealitateari hurbiltzen zaion adibide honetan, ikasleak SPAD paquete-programaren azpiprograma batzuren bidez lortutako balioak jarrai ditzake kalkuladora baten bidez.

Aldagaiak

Datu-matrizeko zutabeetan ditugun aldagaiak eta identifikatzaileak hauek dira:

Nazio-Produktu Gordina / biztanle, dolarretan:	NPGB
Urteroko inflazio-tasa, 1980-1985 tarterako:	UINT
Biztanle-kopurua, mila biztanletan, 1985 urterako:	BIZK
Populazioaren hazkunde-tasa, 1973-1985 tarterako:	POHT
Hiri-populazioaren ehunekoa, 1985 urterako:	HIPO
Eskolako matrikula-tasa, 1984 urterako:	ESMT

Datu-matrizea:

Estatuak	Ident.							
VENEZUELA	VEN	3110	9.2	17323	3.3	85	74	
ARGENTINA	ARG	2130	342.8	30531	1.6	84	88	
MEXICO	MEX	2080	62.2	78820	2.8	69	88	
PANAMA	PAN	2020	3.7	2180	2.3	50	83	
URUGUAY	URU	1660	44.6	3004	0.6	85	86	
BRASIL	BRA	1640	147.7	135539	2.3	73	78	
CHILE	CHI	1440	19.3	11990	1.7	83	95	
COLOMBIA	COL	1320	22.5	24418	1.9	67	73	
COSTA RICA	CRI	1290	36.4	2593	2.8	45	69	
GUATEMALA	GUA	1240	7.4	7966	2.8	41	47	
ECUADOR	ECU	1160	29.7	9367	2.9	52	85	
PERU	PER	960	98.6	18653	2.4	68	85	
PARAGUAY	PAR	940	15.8	3388	2.5	41	73	
NICARAGUA	NIC	850	33.8	3263	3.1	56	71	
REPUBLICA DOMINICANA	RDO	810	14.6	6261	2.4	56	81	
HONDURAS	HON	730	5.4	4366	3.5	39	67	
EL SALVADOR	ELS	710	11.6	5564	3.0	43	56	
BOLIVIA	BOL	470	569.1	6383	2.7	44	71	

Argiro ikusten denez, aldagaiak oso heterogenoak dira eta honek Osagai Nagusizko Analisi Normatua egitea eskatzen du.

Programaren bidez jasotako lehen estatistiko bakunetan, batezbestekoen eta desbidazio tipikoen desberdintasunak ikus daitezke, halaber, berorien bidez atera ditzakegun aldakuntz koefizienteen desberdintasunak.

Estatistiko bakunak

Ondoren jasotzen dugun koerlazio-matrizea oso adierazgarria da, adibide txiki

1 DESCRIPTION SOMMAIRE DES VARIABLES SUR 18 INDIVIDUS

	VARIABLE	MOYENNE	ECART-TYPE	MINIMUM	MAXIMUM	ABSENTS
1 /	NPGB	1364.4446	639.2376	470.0000	3110.0000	0
2 /	UINT	81.9111	142.0133	3.7000	569.1000	0
3 /	BIZK	20644.9414	32960.8437	2180.0000	oooooooo	0
4 /	POHT	2.4778	.6762	.6000	3.5000	0
5 /	HIPO	60.0556	16.4096	39.0000	85.0000	0
6 /	ESMT	76.1111	11.6232	47.0000	95.0000	0
1	MARICE DES CORRELATIONS SUR 18 INDIVIDUS					

honetan, sei aldagai besterik ez bait ditugu.

Koerlazio-matrizea:

Hiru azpimatrizen azpimarratu ditugu, 1. eta 2. aldagai ekonomikoak, 3. eta 4.

	NPGB	UINT	BIZK	POHT	HIPO	ESMT
NPGB	1.00	-.14	.32	-.17	.67	.35
UINT	-.14	1.00	.15	-.14	.06	.11
BIZK	.32	.15	1.00	-.06	.37	.22
POHT	-.17	-.14	-.06	1.00	-.59	-.53
HIPO	.67	.06	.37	-.59	1.00	.67
ESMT	.35	.11	.22	-.53	.67	1.00
	NPGB	UINT	BIZK	POHT	HIPO	ESMT

populaziokoak, 5. eta 6. sozialak baitira.

Bikote hauen erlazio mailak eta NPGB aldagaiak besteekin dituen erlazio mailak eta zeinuak asko esaten du.

Koerlazio-matrizearen diagonalizazioaz egindako Osagai Nagusizko Analisisan, ondoko $\lambda_1, \dots, \lambda_6$ autobalioak ditugu.

Autobalioak

Dakusagunez:

EDITION DES VALEURS-PROPRES

SOMME DES VALEURS-PROPRES

HISTOGRAMME DES PREMIERS VALEURS-PROPRES

	VALEUR-PROPRE	POURCENTAGE	POURCENTAGE CUMULE	
1	2.68648791	44.77	44.77	*****
2	1.14504230	19.08	63.86	*****
3	1.01897109	16.98	80.84	*****
4	.55648887	9.27	90.12	*****
5	.43828008	7.30	97.42	*****
6	.15473048	2.58	100.00	**

hiru lehen ardatz faktorialetan proiektatutako bariantza % 80'84 da.

$$\sum_k \lambda_k = 6 = \text{tr } \mathbf{R}$$

Hots: $\tau_3 = 80'84$

Aldagaien koordinatuak eta interpretaziorako laguntzak.

Aldagaien kasuan, $g_k(j)$ proiektzioak edo koordenatuak ditugu sei

1 EDITION DES COORDONNEES ET DES CONTRIBUTIONS DES VARIABLES

NOMS	E.TYPE	*	COORDONNES						*	PROJECTION ANCIENS AXES UNITE						*	CORRELATION VARIABLE-FACTEUR						*																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										

ardatzetarako, hau da $k \in \{1, \dots, 6\}$

Dakigunez, k . ardatz faktorialaren gain proiektatutako inertzia

da, kasu honetan $\mathbf{X}^T \mathbf{X} = \mathbf{R}$ diagonalizatzen denez, proiektatutako bariantza tipifikatua, hain zuzen.

$$\sum_j g_k^2(j) = \lambda_k$$

Aldagai bakoitzari dagokion proiektatutako bariantzaren proportzioari aldagaiaren ekarpen absolutua deritzo. Hau da, . aldagaiaren ekarpen absolutua da.

Dakusagunez, programak balio hauen erro karratuak $g_k^2(j)/\lambda_k$ emanen dizkigu edo $\mathbf{u}_k$ bektore unitarioen osagaiak.

Hots:

Azken sei zutabeetan aldagai eta faktoreen arteko koerlazioak ditugu, hauek $\mathbf{g}_k = \mathbf{X}^T \mathbf{v}_k$ bektoreen balioak direnez, aldagai normatuen eta $\mathbf{v}_k$ bektore unitarioaren arteko biderkadura eskalarrak, hain zuzen.

$$\mathbf{u}_k = \frac{1}{\sqrt{\lambda_k}} \mathbf{X}^T \mathbf{v}_k = \frac{\mathbf{g}_k}{\sqrt{\lambda_k}}$$

Balio hauek, balioen erro karratuak dira, analisi normatuan $d^2(j,0) = 1$ baita.

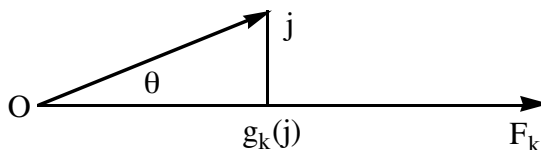
Ardatz faktorial bakoitzari dagokion aldagaiaren jatorriarekiko daukan distantzia karratuaren proportzioari, ardatzaren ekarpen erlatiboa deritzo.

Hau da, $g_k^2(j)/d^2(j,0)$ k . ardatzaren ekarpen erlatiboa da, faktoreak

Dakigunez, $g_k^2(j)/d^2(j,0)$ “Pitagoras-en Teorema Espazioan” kasu honetan $\mathbb{R}^{18}$ espazioan aplikatuz (aldagaien espazioa) lortutako sei ardatzetarako.

Ohartu behar dugu, $g_k(j) = d^2(j,0)$ aldagaiak ardatzarekin osatzen duen θ angeluaren kosinua dela, bien arteko koerlazioa, hain zuzen.

Indibiduen koordinatuak eta interpretaziorako laguntzak.



Indibiduen kasuan, $d^2(i,0)$ jatorriarekiko daukan distantzia karratua daukagu,

$$\cos_k^2(\theta) = g_k^2(j)/d^2(j,0)$$

$$\sum_k \cos_k^2(\theta) = 1$$

ondoren koordinatu zuzenduak sei ardatzetarako.

1 EDITION DES COORONNEES ET DES CONTRIBUTIONS DES INDIVIDUS

NOMS MASES DISTO			* COORDONNES							* CONTRIBUTIONS ABSOLUES*100							* CONTRIBUTIONS RELATIVES *							

			* F1	F2	F3	F4	F5	F6	*	F1	F2	F3	F4	F5	F6	*	F1	F2	F3	F4	F5	F6	*	

VEN	1.000	1.93	*	-56	.94	-.23	.81	.03	.18	*	3.8325	.57	1.7739	.05	.690	*	.16	.46	.03	.34	.00	.02	*	
ARG	1.000	1.63	*	-1.06	-.53	-.13	.43	-.11	-.04	*	13.93	8.26	.5411	.28	1.01	.37	*	.69	.17	.01	.12	.01	.00	*
MEX	1.000	.99	*	-.65	.36	-.57	-.15	.28	-.12	*	5.19	3.81	10.36	1.35	6.12	3.19	*	.42	.13	.23	.02	.08	.01	*
PAN	1.000	.41	*	-.11	.26	.30	.17	.13	.44	*	.15	1.99	2.86	1.66	1.27	4.64	*	.03	.17	.31	.07	.04	.48	*
URU	1.000	1.89	*	-1.00	-.25	.78	-.10	-.45	.00	*	12.39	1.87	20.01	.65	15.20	.01	*	.53	.03	.32	.01	.11	.00	*
BRA	1.000	2.21	*	-.76	.06	-1.10	-.61	-.22	-.01	*	7.20	.11	39.64	1.95	3.72	.05	*	.26	.00	.55	.17	.02	.00	*
CHI	1.000	1.03	*	-.82	-.07	-.52	-.16	.20	.16	*	8.30	.16	8.76	1.55	2.98	5.39	*	.65	.01	.26	.03	.04	.02	*
COL	1.000	.20	*	-.18	.03	.18	-.18	-.31	.06	*	.40	.03	1.06	1.90	7.22	.84	*	.16	.00	.17	.16	.49	.02	*
CRI	1.000	.31	*	.51	.13	.08	.08	-.01	-.14	*	3.23	.52	.19	.40	.02	3.98	*	.84	.06	.02	.02	.00	.06	*
GUA	1.000	1.38	*	.94	.30	-.10	.10	-.62	-.04	*	11.03	2.67	.30	.6329	11	.37	*	.64	.07	.01	.01	.28	.28	*
ECU	1.000	.26	*	.17	.06	.13	-.08	.45	.07	*	.38	.12	.53	.3515	25	.94	*	.12	.02	.06	.02	.77	.02	*
PER	1.000	.21	*	-.17	-.22	.13	-.15	.23	.19	*	.37	1.37	.52	1.43	4.15	8.10	*	.14	.23	.08	.11	.26	.18	*
PAR	1.000	.39	*	.52	-.02	.23	-.17	.03	-.20	*	3.36	.01	1.68	1.69	.07	8.75	*	.69	.00	.13	.07	.00	.10	*
NIC	1.000	.36	*	.51	.08	.06	-.02	.15	.25	*	3.24	.17	.11	.02	1.72	13.73	*	.73	.02	.01	.00	.06	.18	*
RDO	1.000	.24	*	.17	-.09	.31	-.25	.18	.05	*	.36	.26	3.21	3.84	2.56	.57	*	.12	.04	.42	.27	.14	.01	*
HON	1.000	1.01	*	.96	.18	-.06	.07	.23	.06	*	11.32	.92	.12	.33	4.12	.74	*	.90	.03	.00	.01	.05	.00	*
ELS	1.000	1.03	*	.96	.12	-.02	-.09	-.26	.13	*	11.46	.43	.01	.45	5.08	3.42	*	.90	.01	.00	.01	.06	.02	*
BOL	1.000	2.53	*	.56	-1.33	-.50	.44	.07	-.01	*	3.85	.73	8.0211	.46	.35	.02	*	.12	.70	.10	.08	.00	.00	*

Era $\sqrt{m/n} f_k(i,0)$ indibiduo bakoitzari dagokion proiektatutako bariantzaren proportzioari, indibiduoaren ekarpen absolutua deritzo.

Hau da, $\sum_k f_k^2(i)/\lambda_k$ -i. indibiduoaren ekarpen absolutua da, kasu honetan, balio hauek ehunekotan ditugu.

Azken sei zutabeetan faktoreek indibiduoari egindako ekarpen erlatiboak ditugu.

Ekarpen erlatiboa, aldagaien kasurako definitu dugun kontzeptu berbera denez, hau da:

eta

$$f_k^2(i) / d^2(i,0)$$

Indibiduoak ardatzarekin osatzen duen angeluaren kosinua $f_k(i)/d(i,0)$ da.

$$\sum_{k=1}^6 f_k^2(i) = d^2(i,0)$$

Hau da:

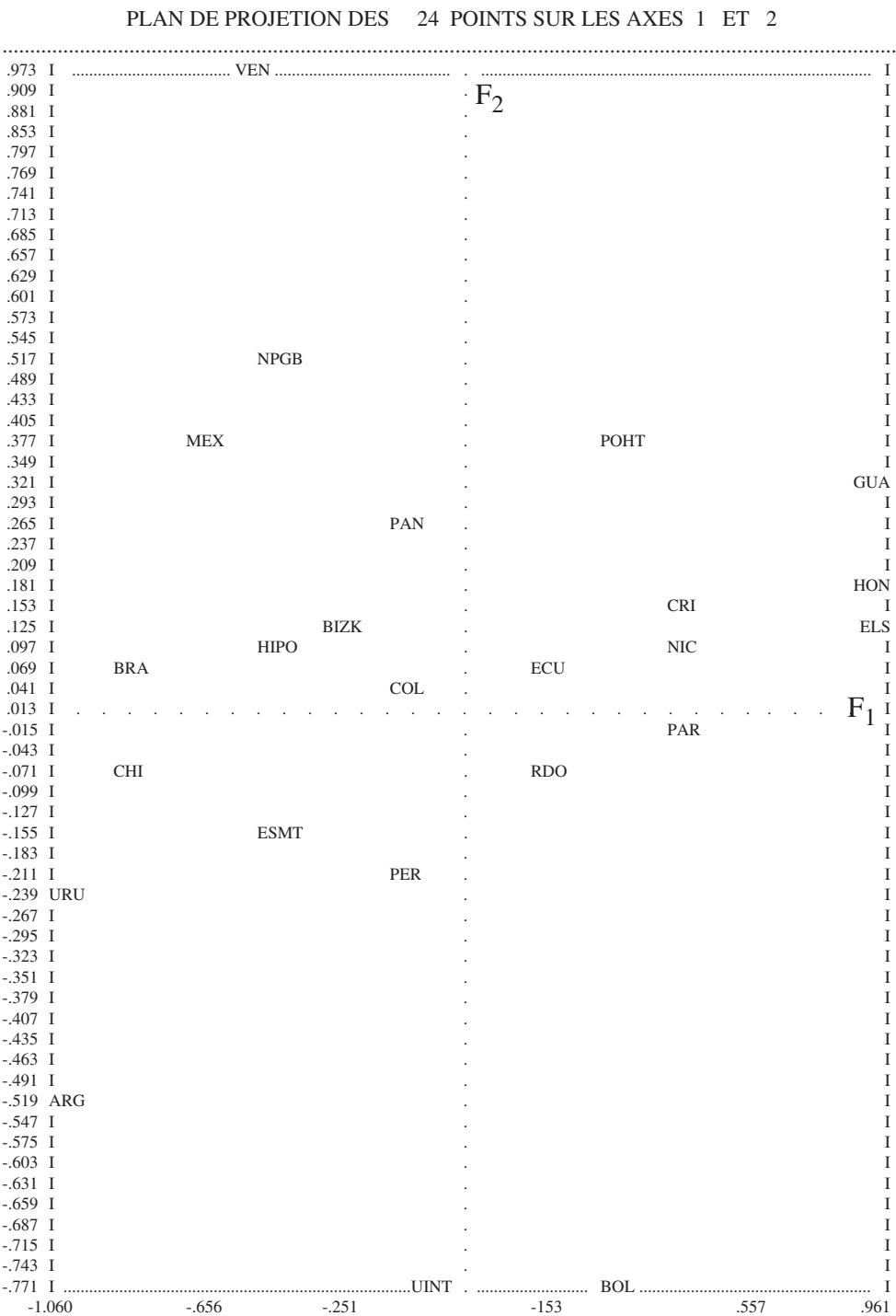
$$\cos_k^2(\theta) = f_k^2(i) / d^2(i,0)$$

$$\sum_k \cos_k^2(\theta) = 1$$

Baterako aurkezpen grafikoa: Plano Faktorial Nagusia

Grabitate-zentrutik gehien urruntzen diren aldagai eta indibiduoak dira koordenatu eta ekarpen absolutu eta erlatibo handienak dauzkatenak ardatz

bakoitzerako, hau da, puntu horiek ardatz faktoriala aldagai berri bezala definitzera



eramaten gaituzte.

Lehen ardatzaren kasuan HIPO (hiri-populazioaren ehunekoa) da aldagairik nabariena alde batean, bere koordenatua edota faktorearekin daukan koerlazioa -0.94 da eta honen arabera faktorearen bariantzari egindako ekarpen absolutua handiena da $(-0.57)^2$ (gogoratu behar dugu ardatzaren alde batean edo bestean egotea edonolakoa dela). ESMT (eskolako matrikula-tasa) eta NPGB (Nazio-Produktu Gordina) alde berean nabariak dira.

Aldagai hauen inguruan, planoaren ezker aldean, dauden indibiduorik nabariena ARG, URU, CHI, eta atzetik MEX, BRA, dira. Azken bi hauek ekarpenetan dituzten ezberdintasunak aipagarriak dira; Brasilek faktoreari egindako ekarpen absolutua handiagoa dauka bere koordenatua handiagoa delako, baina faktoreak Brasili egindako ekarpen erlatiboa txikiagoa da Brasilek grabitate-zentrutik daukan distantzia handiagoa delako. Guzti honek zera esan nahi du: Brasil beste ardatz batean hobeto aurkezten dela, kasu honetan, konkretuki, hirugarren ardatzarekiko dituen balioak adierazgarriagoak dira.

Ardatzaren beste aldean POHT (populazioaren hazkunde-tasa) da ekarpen absolutu eta erlatibo nabariak daukagun aldagaia. Indibiduen artean ELS, HON, GUA eta atzetik PAR, NIC, guzti hauek planoaren eskuin aldean ikus daitezkeelarik.

Lehen ardatza, bada, garapen-ardatza bezala defini daiteke, garapen-mota batena behintzat, eta honen arabera kontinentearen hegokonoa eta Ertamerika kontrajarriak daude.

Bigarren ardatzaren kasuan, UINT (urteroko inflazio-tasa) da aldagairik nabariena eta bere inguruan BOL eta ARG daude, ardatzean proiektatzen den bariantzaren erdia baino gehiago Bolibiari dagokiolarik. Puntu hauek planoaren beheko aldean ikus daitezke.

Ardatzaren beste aldean, hau da, planoaren goiko aldean, NPGB eta POHT daude, biak lehen ardatzean kontrajarriak eta koerlazio hobeagorekin genituenak eta hauen inguruan VEN, ardatzean proiektatutako bariantzaren laurdena berari dagokiona, eta MEX, azken honek ekarpen absolutu eta erlatiboak txikiak dituelarik.

Bigarren ardatza edo beste edozein beti hondar-ardatza da aurrenekoekiko, zentzu honetan bigarrenak lehenengoak azaltzen ez dituen datu-matrizearen sakabanatzearen alderdiak azaltzen ditu.

Adibide honen kasuan bigarren ardatzak, gehienbat, kontinente honetan hain berezia den inflazio-tasaren arabera dauden desberdintasunak azaltzen ditu. Halere, nazio-produktu gordina eta populazioaren hazkunde-tasa aldagaien konbinazio lineala den bigarren faktorean garrantziaz sartu dira.

Hirugarren ardatzean populazioko aldagaiak dira garrantzitsuenak, nabariena BIZK (biztanle-kopurua) izanik. Faktore honen arabera populazio handia edo gehikorra duten estatuak (BRA, MEX) eta populazio ez hain handia eta gehikorra duten estatuak (URU, CHI) kontrajartzen dira. Azkenik esan behar dugu, interpretatu nahi izan dugun hirugarren ardatza beste biekiko hondar-ardatza dela. Datu-matrizearen desberdintasun globalik garrantzitsuenak lehen ardatzean proiektatzen ziren, sakabanatzearen % 44'77 jasoz eta bigarren garrantzitsuenak eta ez hain globalak bigarrenean, sakabanatzearen % 19'08 jasoz, hain zuzen ere.

XII.4. KORRESPONDENTZI ANALISI FAKTORIALA

Datu-Analisiaren eskola frantsesaren aita bezala ezaguna den J.P. Banzécri-k korrespondentzi Analisi Faktoriala (KAF) garatu zuen 60.eko hamarkadaren lehen urteetan.

Estatistika inferentzialetik at lehen ezer gutxi ikasiak izan ziren kontingentzi tauletan arrakasta handiz aplikatzen da orduz gero KAF-ren metodologia¹ eta honek Datu-Analisiaren eskolaren bultzada suposatu zuen.

Teknika honi buruz egin daitezkeen aurkezpen ezberdinekin loturik daude KAF-ren helburuak.

Halere, helburu orokor bat kontsidera daiteke: Populazio baten gain ohartutako bi aldagaien arteko menpekotasun-erlazioen azterketa egitea, hain zuzen.

KAF taula-mota ezberdinetara aplikagarria da, kontingentzi tauletara, alegia, eta zentzu zabalean kontsideraturik kontingentzi taulak bezala har daitezkeen tauletara.

(1) K. FERNANDEZ AGIRRE. *Análisis multivariante de tablas de gran inercia. Aplicación al corpus con terminología económica en euskara*. Doktoradutza-tesia (1.988)

Kontingentzi taula

$$\begin{matrix} & \begin{matrix} 1 & 2 & \dots & j & \dots & n \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ \vdots \\ i \\ \vdots \\ m \end{matrix} & \begin{bmatrix} k_{11} & k_{12} & \dots & k_{1j} & \dots & k_{1n} \\ k_{21} & k_{22} & \dots & k_{2j} & \dots & k_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ k_{i1} & k_{i2} & \dots & k_{ij} & \dots & k_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ k_{m1} & k_{m2} & \dots & k_{mj} & \dots & k_{mn} \end{bmatrix} \end{matrix} = [k_{ij}]$$

i. errenkadaren eta j. zutabearen ebakiduran daukagun k_{ij} balioa, kontaketa baten emaitza izango da, normalki, (i,j) bikotearen maiztasun absolutua.

Halere, zentzu zabalean, kontingentzi taula definituz gero, errenkadaren baturak, $k_{i.}$ balioak hain zuzen, eta zutabeen baturak, $k_{.j}$ balioak, zentzu argia daukaten balioak izan daitezke nahiz eta berez taula horiek neurri-taulak izan. Adibidez, enpresa batzuren salmenta-kopuruak, esportatutako balioak... izan daitezke balio horiek.

Askotan, kontingentzi taulei menpekotasun-taulak edo taula gurutzatuak deritze eta batzutan maiztasun-taulak ere. (Ikus IV. ikasgaia).

Taula hauetan, errenkadak nahiz zutabeak populazio beraren partizioak dira, atribututan kategoriatan edo balio ordinaletan egindako partizioak alegia, horrela errenkadek nahiz zutabeek betebeharrak berdina dauzkate.

Kontingentzi taulatik errenkada-soslaiak eta zutabe-soslaiak eratortzen dira errenkada-puntuaren eta zutabe-puntuaren arteko distantziek zentzua eduki dezaten.

Honela, errenkada-soslaien taulan, $k_{i.}$ errenkadaren totalarekiko proportzioak eta zutabe-soslaien taulan, $k_{.j}$ zutabearen totalarekiko proportzioak ditugu.

Hots:

$$\begin{aligned} k &= \sum_i \sum_j k_{ij} \text{ kontingentzitaulararen totala} \\ f_{ij} &= k_{ij}/k \quad \text{maiztasun erlatiboak} \\ \left. \begin{aligned} f_{i.} &= \sum_j f_{ij} \\ f_{.j} &= \sum_i f_{ij} \end{aligned} \right\} \quad \text{bazter-maiztasunak} \end{aligned}$$

Errenkada-soslaien taula

	1	2	...	j	...	n	Totala
1	$\frac{f_{11}}{f_{1.}}$	$\frac{f_{12}}{f_{1.}}$	...	$\frac{f_{1j}}{f_{1.}}$	...	$\frac{f_{1n}}{f_{1.}}$	1
2	$\frac{f_{21}}{f_{2.}}$	$\frac{f_{22}}{f_{2.}}$	...	$\frac{f_{2j}}{f_{2.}}$	...	$\frac{f_{2n}}{f_{2.}}$	1
...							
i	$\frac{f_{i1}}{f_{i.}}$	$\frac{f_{i2}}{f_{i.}}$	...	$\frac{f_{ij}}{f_{i.}}$	...	$\frac{f_{in}}{f_{i.}}$	1
...							
m	$\frac{f_{m1}}{f_{m.}}$	$\frac{f_{m2}}{f_{m.}}$	...	$\frac{f_{mj}}{f_{m.}}$	...	$\frac{f_{mn}}{f_{m.}}$	1
Totala	$f_{.1}$	$f_{.2}$	...	$f_{.j}$	...	$f_{.n}$	

Zutabe-soslaien taula

	1	2	...	j	...	n	Totala
1	$\frac{f_{11}}{f_{.1}}$	$\frac{f_{12}}{f_{.2}}$	...	$\frac{f_{1j}}{f_{.j}}$	...	$\frac{f_{1n}}{f_{.n}}$	$f_{1.}$
2	$\frac{f_{21}}{f_{.1}}$	$\frac{f_{22}}{f_{.2}}$	...	$\frac{f_{2j}}{f_{.j}}$	...	$\frac{f_{2n}}{f_{.n}}$	$f_{2.}$
...							
i	$\frac{f_{i1}}{f_{.1}}$	$\frac{f_{i2}}{f_{.2}}$	...	$\frac{f_{ij}}{f_{.j}}$	...	$\frac{f_{in}}{f_{.n}}$	$f_{i.}$
...							
m	$\frac{f_{m1}}{f_{.1}}$	$\frac{f_{m2}}{f_{.2}}$	...	$\frac{f_{mj}}{f_{.j}}$	...	$\frac{f_{mn}}{f_{.n}}$	$f_{m.}$
Totala	1	1	...	1	...	1	

XII.4.1. Hodeiak, masak eta distantziak

Osagai Nagusizko Analisisian bezala, bi espazio desberdinetan koka gaitzke.

a) $\mathbb{R}^n$ espazio bektorialean m errenkada-soslaien puntuak ditugu. Honela, $f_{i.}$ masaz hornitutako i puntuaren koordinatuak ondoko multzokoak dira:

$$\left\{ \left(f_{ij} / f_{i.} \right) ; j = 1, 2, \dots, n \right\}$$

Konkretuki, errenkada-soslaien hodeiaren m puntuak $\mathbb{R}^{n-1}$ azpiespazio batean kokaturik daude, puntu bakoitzaren koordinatuen batura 1 baita.

Hau da:

$$\sum_j \left(f_{ij} / f_{i.} \right) = 1 \quad \forall i = 1, 2, \dots, m$$

Honela, puntuen hurbiltasunak soslaien hurbiltasunak bezala interpreta daitezke

b) $\mathbb{R}^m$ espazio bektorialean n zutabe-soslaien puntuak ditugu. Honela, $f_{.j}$ masaz hornitutako j puntuaren koordinatuak ondoko multzokoak dira:

$$\left\{ \left(f_{ij} / f_{.j} \right) ; i = 1, 2, \dots, m \right\}$$

Konkretuki, zutabe-soslaien hodeiaren n puntuak $\mathbb{R}^{m-1}$ azpiespazio batean kokaturik daude, puntu bakoitzaren koordinatuen batura 1 baita.

Hau da:

$$\sum_i \left(f_{ij} / f_{.j} \right) = 1 \quad \forall j = 1, 2, \dots, n$$

Doikuntza egin aurretik puntu bakoitzaren masa kontutan hartzea guztiz zilegia da, horrela, masa txikiak dituzten puntuak ez dira gehiegi nagusitzen eta banaketa bere baitan errespetatzen da.

$\mathbb{R}^n$ eta $\mathbb{R}^m$ espazioetan definitzen diren distantziak hauexek dira:

$$d^2(i, i') = \sum_{j=1}^n \frac{1}{f_{\cdot j}} \left(\frac{f_{ij}}{f_{i \cdot}} - \frac{f_{i'j}}{f_{i' \cdot}} \right)^2$$

$$d^2(j, j') = \sum_{i=1}^m \frac{1}{f_{i \cdot}} \left(\frac{f_{ij}}{f_{\cdot j}} - \frac{f_{ij'}}{f_{\cdot j'}} \right)^2$$

Horrela definituriko distantziari χ^2 (ji karratua) deritzo.

Distantzia ponderatu honek daukan abantaila guztiz berezkoa izateaz gain **banaketa-baliokidetasun** izeneko propietatea betetzea da.

Propietate honen arabera i_1, i_2 errenkada-puntuak soslai berbera badira (proportzio berdinak zutabe guztietarako) berorien masen batura daukan i_0 errenkada-puntuaz ordezkatu daitezke eta $\mathbb{R}^n$ nahiz $\mathbb{R}^m$ espazioetan distantziak ez dira aldatzen. Berdin esan daiteke j_1, j_2 soslai berbereko bi zutabe-puntuentzako.

Propietate honek analisiaren emaitzak oso egonkor bihurtzen ditu, honela, datu-taula handi batean berdin antzeko soslaiak daukaten lerroak emaitzak mantenduz batu daitezke eta honek nomenklatura baten arbitraltasunaren aurrean emaitzen garrantzia suposatzen du.

Soslaien artean definitutako χ^2 distantzia ohizko distantzia euklidear bihurtu daiteke, ponderazioak parentesien barruan sartuz.

Hau da:

$$d^2(i, i') = \sum_{j=1}^n \left(\frac{f_{ij}}{f_{i \cdot} \sqrt{f_{\cdot j}}} - \frac{f_{i'j}}{f_{i' \cdot} \sqrt{f_{\cdot j}}} \right)^2$$

Kasu honetan, ardatzen eskala aldatuz i puntuaren koordenatuak $f_{ij}/f_{i \cdot} \sqrt{f_{\cdot j}}$ balioak dira.

Era berean:

$$d^2(j, j') = \sum_{i=1}^m \left(\frac{f_{ij}}{f_{.j} \sqrt{f_{.i}}} - \frac{f_{ij'}}{f_{.j'} \sqrt{f_{.i}}} \right)^2$$

kasu honetan, ardatzen eskala aldatuz j puntuaren koordenatuak $f_{ij}/f_{.j} \sqrt{f_{.i}}$ balioak dira.

Eskala aldatu ondoren, errenkada-soslaien grabitate-zentrua zutabeen bazter-banaketaren erro karratua da.

i puntu bakoitzaren masa edo pisua $f_{.i}$ denez:

$$g_j = \sum_{i=1}^n \frac{f_{ij}}{f_{.i} \sqrt{f_{.j}}} f_{.i} = \sqrt{f_{.j}}$$

Eta zutabe-soslaien grabitate-zentrua errenkaden bazter-banaketaren erro karratua da. j puntu bakoitzaren masa edo pisua $f_{.j}$ denez:

$$h_i = \sum_{j=1}^n \frac{f_{ij}}{f_{.j} \sqrt{f_{.j}}} f_{.j} = \sqrt{f_{.i}}$$

Azkenik:

$$\frac{f_{ij}}{f_{.i} \sqrt{f_{.j}}} - \sqrt{f_{.j}} \quad \text{eta} \quad \frac{f_{ij}}{f_{.j} \sqrt{f_{.i}}} - \sqrt{f_{.i}}$$

bi hodeien koordenatu zentratuak, hurrenez hurren, dira.

Dakusagunez, bada, puntu-hodeiaren grabitate-zentrua jatorrira eramán dugu eta ardatzen eskala-aldaketa baten bidez χ^2 distantzia ohizko distantzia euklidearra bezala idatzi dugu, honela, korrespondentzi analisi faktoriala analisi orokorraren barrutian aurkezten da.

XII.4.2. $\mathbb{R}^n$ espazioan egindako analisia

$\mathbb{R}^n$ espazioan diagonalizatzen den kobariantz matrizearen gai orokorra i puntuaren masa $f_{i.}$ dela kontutan hartuz, hauxe da:

$$S_{jj'} = \sum_{i=1}^m \left(\frac{f_{ij}}{f_{i.} \sqrt{f_{.j}}} - \sqrt{f_{.j}} \right) \left(\frac{f_{ij'}}{f_{i.} \sqrt{f_{.j'}}} - \sqrt{f_{.j'}} \right) f_{i.}$$

j, j' puntuen balio zentratuen arteko biderkadura da i guztietarako eta i puntuen masaz ponderaturik.

Azkenik, biderkatzaile bakoitza izendatzaile komunean ipiniz eta $\sqrt{f_{i.}}$ biderkatuz.

$$S_{jj'} = \sum_{i=1}^m \left(\frac{f_{ij} - f_{i.} f_{.j}}{\sqrt{f_{i.} f_{.j}}} \right) \left(\frac{f_{ij'} - f_{i.} f_{.j'}}{\sqrt{f_{i.} f_{.j'}}} \right)$$

Horrela diagonalizatzen den $L(J)$ kobariantz matrizea $X(m,n)$ matrizearen funtzioan idatz daiteke:

$$\begin{matrix} L(J) & = & X^T & X \\ (n,n) & & (n,m) & (m,n) \end{matrix}$$

non, X matrizearen gai orokorra $\frac{f_{ij} - f_{i.} f_{.j}}{\sqrt{f_{i.} f_{.j}}}$ da.

Hots:

$$X = \left[\frac{f_{ij} - f_{i.} f_{.j}}{\sqrt{f_{i.} f_{.j}}} \right]$$

Froga daiteke datuak zentratuz edo zentratu gabe erabiltzeak emaitza berdinetara eramaten duela.

Hau da, datuak zentratu gabe hartzen baditugu

$$X^* = \left[\frac{f_{ij}}{\sqrt{f_{i.} f_{.j}}} \right]$$

eta kobariantz matrizea $\mathbf{L}^*(J) = \mathbf{X}^{*T} \mathbf{X}^*$ izango da.

$\mathbf{u}_n$ bektorea, non j . osagaia $\sqrt{f_{.j}}$ bait da, $\mathbf{L}(J)$ kobariantz matrizearen 0 autobalioari dagokion autobektorea da, halaber, $\mathbf{u}_n$ (grabitate-zentrua) $\mathbf{L}^*(J)$ kobariantz matrizearen 1 autobalioari dagokion autobektorea da.

$\mathbf{u}_n$ ez den edozein $\mathbf{u}_k$ autobektore $\mathbf{L}(J)$ edota $\mathbf{L}^*(J)$ matrizeen autobalio berberari dagokion autobektorea da.

k. ardatzaren gain m errenkada-puntuek dauzkaten proiektzioak, datu zentratuaz nahiz zentratu gabeaz kalkula daitezke

$$\begin{matrix} \hat{\psi}_k &= \mathbf{X} & \mathbf{u}_k &= \mathbf{X}^* \mathbf{u}_k \\ (m,1) & & (m,n)(n,1) & (m,n)(n,1) \end{matrix}$$

$\mathbf{X} \mathbf{u}_k$ eta $\mathbf{X}^* \mathbf{u}_k$ bektoreen abzisak berberak baitira.

Dakusagun i puntuaren koordinatua bektore batean nahiz bestean berbera dela.

$$\begin{aligned} \hat{\psi}_{ki} &= \sum_{j=1}^n \left(\frac{f_{ij}}{f_{i.} \sqrt{f_{.j}}} - \sqrt{f_{.j}} \right) u_{kj} = \\ &= \sum_{j=1}^n \frac{f_{ij}}{f_{i.} \sqrt{f_{.j}}} u_{kj} \end{aligned} \quad (8)$$

parentesia kentzean bigarren batugaia $\sum_{j=1}^n \sqrt{f_{.j}} u_{kj} = 0$ da, $\mathbf{u}_n$ eta $\mathbf{u}_k$ autobektoreak ortogonalak baitira.

XII.4.3. $\mathbb{R}^m$ espazioan egindako analisia

$\mathbb{R}^m$ espazioan diagonalizatzen den kobariantz matrizearen gai orokorra j puntuaren masa $f_{.j}$ dela kontutan hartuz, hauxe da:

$$S_{ii'} = \sum_{j=1}^n \left(\frac{f_{ij}}{f_{.j} \sqrt{f_{i.}}} - \sqrt{f_{i.}} \right) \left(\frac{f_{i'j}}{f_{.j} \sqrt{f_{i'.}}} - \sqrt{f_{i'.}} \right) f_{.j}$$

Eta biderkatzaile bakoitza izendatzaile komunean ipiniz eta $\sqrt{f_{.j}}$ biderkatuz

$$S_{ii'} = \sum_{j=1}^n \left(\frac{f_{ij} - f_{i.} f_{.j}}{\sqrt{f_{i.} f_{.j}}} \right) \left(\frac{f_{i'j} - f_{i'.} f_{.j}}{\sqrt{f_{i'.} f_{.j}}} \right)$$

Hau da:

$$\mathbf{L} \begin{pmatrix} \text{(I)} \\ \text{(n,n)} \end{pmatrix} = \mathbf{X} \begin{pmatrix} \mathbf{X}^T \\ \text{(m,n)} \end{pmatrix} \begin{pmatrix} \text{(n,m)} \end{pmatrix}$$

Dakusagunez korrespondentzian jartzen diren errenkaden eta zutabeen multzoek betebeharrak berdinak dauzkate eta espazio batetik bestera pasatzeko i, j azpiindizeen betebeharrak aldatzea aski da.

$\mathbf{v}_m$ bektorea, non i . osagaia $\sqrt{f_{i.}}$ da, $\mathbf{L}(\mathbf{I})$ kobariantz matrizearen 0 autobalioari dagokion autobektorea da, halaber $\mathbf{v}_m$ (grabitate-zentrua) $\mathbf{L}^*(\mathbf{I}) = \mathbf{X}^* \mathbf{X}^{*T}$ kobariantz matrizearen 1 autobalioari dagokion autobektorea da.

$\mathbf{v}_m$ ez den edozein $\mathbf{v}_k$ autobektore $\mathbf{L}(\mathbf{I})$ edota $\mathbf{L}^*(\mathbf{I})$ matrizeen autobalio berberari dagokion autobektorea da.

Azkenik, k . ardatzaren gain n zutabe-puntuek dauzkaten proiektzioak:

$$\hat{\varphi}_k = \mathbf{X}^T \mathbf{v}_k = \mathbf{X}^{*T} \mathbf{v}_k$$

Eta j puntuaren koordinatua bere adierazpen errazenean:

$$\hat{\varphi}_{kj} = \sum_{i=1}^m \frac{f_{ij}}{f_{.j} \sqrt{f_{i.}}} v_{ki} \quad (9)$$

XII.4.4. $\mathbf{R}^n$ eta $\mathbf{R}^m$ espazioen arteko erlazioa

Analisi orokorrean ikusi genuenez $\mathbf{L}^*(\mathbf{J}) = \mathbf{X}^{*T} \mathbf{X}^*$ eta $\mathbf{L}^*(\mathbf{I}) = \mathbf{X}^* \mathbf{X}^{*T}$ matrize simetrikoek desberdin zero diren λ_k autobalio berberak dituzte eta λ_k autobalioari dagozkion $\mathbf{L}^*(\mathbf{J})$ -ren $\mathbf{u}_k$ eta $\mathbf{L}^*(\mathbf{I})$ -ren $\mathbf{v}_k$ autobektoreen arteko erlazioak hauexek dira:

$$\mathbf{v}_k = \frac{1}{\sqrt{\lambda_k}} \mathbf{X}^* \mathbf{u}_k \quad (10)$$

$$\mathbf{u}_k = \frac{1}{\sqrt{\lambda_k}} \mathbf{X}^{*T} \mathbf{v}_k \quad (11)$$

$\mathbf{v}_k$ bektorearen i . errenkada (10) kontutan hartuz, honela idatz daiteke:

$$v_{ki} = \frac{1}{\sqrt{\lambda_k}} \sum_{j=1}^n \frac{f_{ij}}{\sqrt{f_{i.} f_{.j}}} u_{kj}$$

Eta (8) erlazioan u_{kj} ordezkatzuz:

$$\hat{\psi}_{ki} = \sqrt{\lambda_k} \frac{v_{ki}}{\sqrt{f_{i.}}} \quad (12)$$

Era berean, $\mathbf{u}_k$ bektorearen j . errenkada (11) kontutan hartuz honela idatz daiteke:

$$u_{kj} = \frac{1}{\sqrt{\lambda_k}} \sum_{i=1}^m \left(\frac{f_{ij}}{\sqrt{f_{i.} f_{.j}}} \right) v_{ki}$$

Eta (9) erlazioan v_{ki} ordezkatzuz:

$$\hat{\varphi}_{kj} = \sqrt{\lambda_k} \frac{u_{kj}}{\sqrt{f_{.j}}} \quad (13)$$

Azkenik, (9) erlazioan (12) erlaziotik ateratako v_{ki} ordezkaturik:

$$\hat{\varphi}_{kj} = \frac{1}{\sqrt{\lambda_k}} \sum_{i=1}^m \frac{f_{ij}}{f_{.j}} \hat{\psi}_{ki} \quad (14)$$

eta (8) erlazioan (13) erlaziotik ateratako u_{kj} ordezkaturik:

$$\hat{\psi}_{ki} = \frac{1}{\sqrt{\lambda_k}} \sum_{j=1}^n \frac{f_{ij}}{f_{i.}} \hat{\varphi}_{kj} \quad (15)$$

(14), (15) erlazioetan errenkada-puntuek eta zutabe-puntuek k . ardatzaren gain dituzten koordinatuen arteko erlazioak ditugu.

Dakusagunez ($f_{ij} / f_{i.}$) gai orokorrak daukan matrizearen bidez, hau da, errenkada-soslairen matrizearen bidez eta j zutabe-puntuaren koordinatuen bidez ((15) erlazioa), i errenkada-puntuaren koordinatuak lortzen dira.

Honela, i errenkada-puntuaren koordinatua k ardatzaren gain ($\hat{\psi}_{ki}$), $1/\sqrt{\lambda_k}$ balioa salbu, j zutabe-puntuaren koordinatuen barizentrua edo grabitate-zentrua da, pisuak i errenkada-soslairen balioak direlarik.

Era berean, j zutabe-puntuaren koordinatua k ardatzaren gain ($\hat{\varphi}_{kj}$), $1/\sqrt{\lambda_k}$ balioa salbu, i errenkada-puntuaren koordinatuen barizentrua da, pisuak j zutabe-soslairen balioak direlarik ((14) erlazioa).

Erlazio barizentrikoak bezala ezagutzen diren (14), (15) erlazioak ez dira analisi orokorraren (5), (6) erlazioen kasu bereziak, KAF-ean transizio-erlazioaren matrizeak bata bestearen iraulea ez baitira.

Erlazio hauetan, inplizituki, edozein k -rako $\lambda_k \leq 1$ dela daukagu ($1/\sqrt{\lambda_k} < 1$ ezin baita izan).

XII.4.5. Maiztasun-taularen berrerraketa

Gai honen 2.4. atalean, analisi orokorrerako ikusi genuenez, $\mathbf{X}^* = [f_{ij}/\sqrt{f_{i.} f_{.j}}]$ matrizearen berrerraketa hau da:

$$\mathbf{X}^* = \sum_{k=1}^n \sqrt{\lambda_k} \mathbf{v}_k \mathbf{u}_k^T$$

(12), (13) erlazioetatik ateratako $\mathbf{v}_k, \mathbf{u}_k$ balioak ordezkatzuz gero:

$$\frac{f_{ij}}{\sqrt{f_{i.} f_{.j}}} = \sum_{k=1}^n \left[\frac{1}{\sqrt{\lambda_k}} \psi_{ki} \varphi_{kj} \right] \sqrt{f_{i.} f_{.j}}$$

Eta dakigunez $\lambda_1 = 1$ autobalioari $\mathbf{u}_{1j} = \sqrt{f_{.j}}$, $\mathbf{v}_{1i} = \sqrt{f_{i.}}$ autobektoreak dagozkio, orduan, (12) (13) erlazioetan dakusagunez $\hat{\varphi}_{1j} = 1$ eta $\hat{\psi}_{1i} = 1$

$$f_{ij} = f_{i.} f_{.j} \left[1 + \sum_{k=2}^n \frac{1}{\sqrt{\lambda_k}} \psi_{ki} \varphi_{kj} \right]$$

Berreratze-formula honetan, $\hat{\psi}_{ki}$, $\hat{\varphi}_{ki}$ eta λ_k hirurek zerorantz jotzen dute.

Askotan, berreratze-formula, faktore normatuen bidez idazten da ere.

ψ_{ki} eta φ_{ki} faktore normatuak:

$$\psi_{ki} = \frac{1}{\sqrt{\lambda_k}} \hat{\psi}_{ki} \quad \text{eta} \quad \varphi_{ki} = \frac{1}{\sqrt{\lambda_k}} \hat{\varphi}_{ki}$$

Orduan berreratze-formula:

$$f_{ij} = f_i \cdot f_j \left[1 + \sum_{k>1} \sqrt{\lambda_k} \psi_{ki} \varphi_{kj} \right]$$

XII.4.6. Interpretaziorako laguntzak

Osagai Naguzizko Analisiaren adibidean ikusi genuen bezala, ardatzen interpretaziorako, bi koefiziente-segidak kalkulatzeko dira korrespondentzia ipinitako bi multzoen elementu bakoitzerako.

1) Ekarpen absolutuak

Dakigunez, elementu bakoitzari dagokion proiektatutako bariantzaren proportzioari elementuaren ekarpen absolutua deritzo.

k. ardatzaren gain m errenkada-puntuen bariantza proiektatua, i errenkada-puntu bakoitzaren masa f_i dela jakinik, hauex da:

$$\sum_{i=1}^m f_i \cdot \hat{\psi}_{ki}^2$$

Eta $\hat{\psi}_{ki}$ balioa ordezkatzuz:

$$\sum_{i=1}^m f_i \cdot \hat{\psi}_{ki}^2 = \lambda_k \sum_{i=1}^m v_{ki}^2 = \lambda_k \quad (v_k \text{ unitate bektorea baita})$$

Era berean j zutabe-puntuen bariantza proiektatua.

$$\sum_{j=1}^n f_{.j} \hat{\varphi}_{kj}^2 = \lambda_k \sum_{j=1}^n u_{kj} = \lambda_k$$

Honela, k . ardatzaren gain proiektatutako bariantzan i elementuari dagokion proportzioa:

$$Ca_k(i) = f_{i.} \hat{\psi}_{ki}^2 / \lambda_k$$

$Ca_k(i)$ balioa, i elementuak k . ardatzean daukan ekarpen absolutua izanik.

Ohar daiteke:

$$\sum_{i=1}^m Ca_k(i) = 1 \quad \forall_k \text{ dela}$$

Analogikoki, j elementuak k . ardatzean daukan ekarpen absolutua:

$$Ca_k(j) = \frac{f_{.j} \hat{\varphi}_{kj}^2}{\lambda_k}$$

2) Ekarpen erlatiboak

Dakigunez, ardatz faktorial bakoitzari dagokion elementuaren jatorriarekiko daukan distantzia karratuaren proportzioari, ardatzaren ekarpen erlatiboa deritzo.

Ardatz faktorialek, bi espaziotan, oinarri ortonormalak osatzen dituztenez, bi espazioko G , H , grabitate-zentruetikiko elementuek dituzten distantzia karratuak ardatzetan dituzten koordinatuen karratueta deskonposatzen dira batukorki.

Honela, $\mathbb{R}^n$ espazioko i puntu bakoitzerako:

$$d_n^2(i, G) = \sum_{j=1}^n \left(\frac{f_{ij}}{f_{i.} \sqrt{f_{.j}}} - \sqrt{f_{.j}} \right)^2$$

Halaber, $\mathbb{R}^m$ espazioko j puntu bakoitzerako:

$$d_m^2(j, H) = \sum_{i=1}^m \left(\frac{f_{ij}}{f_{.j} \sqrt{f_{i.}}} - \sqrt{f_{i.}} \right)^2$$

Puntu baten soslaiak batezbesteko soslaiarekin kointziditzen badu distantzia zero izango dela ohartu behar dugu.

k. ardatzaren ekarpen erlatiboa i puntuarekiko:

$$Cr_k(i) = \frac{\hat{\psi}_{ki}^2}{d_n^2(i, G)}$$

eta i puntuak k. ardatzarekin osatzen duen angeluaren cosinu karratua denez, koerlazio-koefiziente karratua bezala interpreta daiteke.

Ohar daiteke

$$\sum_k Cr_k(i) = 1 \quad \forall i \text{ dela}$$

Analogikoki, k. ardatzak j elementuaren posizioarekiko daukan ekarpen erlatiboa:

$$Cr_k(j) = \frac{\hat{\varphi}_{kj}^2}{d_m^2(j, H)}$$

Eta, era berean:

$$\sum_k Cr_k(i) = 1 \quad \forall j$$

XII.4.7. Korrespondentzia Anitzeko Analisi Faktoriala

Korrespondentzi Analisi Faktorialaren orokorpena bezala Korrespondentzia Anitzeko Analisi Faktoriala sortzen da.

Bi aldagai baino gehiago, kualitatiboak nahiz kuantitatiboak izan, azken kasu honetan bada kategoriatan mailakatuta, ditugunean, Korrespondentzia Anitzeko Analisi Faktorialaren metodologia analisi ikuskatzailea egiteko guztiz baliagarria da.

Teknika honen aplikazio-eremu handi bat galdera itxien inkestetakoa da; halaber, edozein inkesta motatan lortutako erantzunak aztertzeo, oso egokia da.

Korrespondentzia Anitzeko Analisiaren eragiketaren teknika, deskribapen logikozko taula batean Korrespondentzi Analisi Faktoriala aplikatzea besterik ez da.

Deskribapen logikozko taula

Taula hauen elementuak 0 eta 1 dira. k_{ij} taularen gai orokorra izanik, i elementuak j propietatea duenean $k_{ij} = 1$ izango da, aldiz, i elementuak j propietaterik ez duenean $k_{ij} = 0$ izango da.

Deskribapen logikozko taula batek ondoko baldintza betetzen baldin badu aukera osoko taularen erara dagoela esango dugu: J zutabeen (edo modalitateen) multzoa J_q azpimultzoen (edo galderen) Q segida da non:

$$\forall i \in I, \forall J_q \in Q, \exists j \in J_q \mid k_{ij} = 1 \wedge (j' \in J_q, j' \neq j) \Rightarrow k_{ij'} = 0$$

Honek zera esan nahi du: I indibiduen multzoko i elementu bakoitzak J_q klasean modalitate bat eta bakarrik bat edukiko duela.

Deskribapen logikozko taula batek aukera osoko taularen erara honelako itxura dauka:

	J_1		J_2		J_q		J_Q	Totala
1	010		1000		0010		010	Q
.								
.								
$Z = i$	001		0100		0100		100	Q
(m,J)								
.								
m	100		0001		0100		010	Q

Dakusagunez i indibiduoak J_q azpimultzoan bigarren modalitatea aukeratu du, inkesta baten kasua balitz q . galderarentzako bigarren erantzuna aukeratu duela esango genuke.

Korrespondentzi Anitzeko Analisia eta Burt-en taula batetan Korrespondentzi Analisi Faktoriala aplikatzea baliokideak dira.

Burt-en taula

m indibiduen gain Q aldagai (edo galdera) ohartu baditugu, Burt-en taula ($J_q, J_{q'}$) binaka gurutzatutako taula guztiez osaturik dago.

B =

	J_1	J_2	...	J_q	...	J_Q
J_1	0	J_{12}		J_{1q}		J_{1Q}
J_2	J_{21}	0		J_{2q}		J_{2Q}
$\vdots$						
J_q	J_{q1}	J_{q2}		0		J_{qQ}
$\vdots$						
J_Q	J_{Q1}	J_{Q2}		J_{Qq}		0

Diagonalean ditugun azpitaulak bloke diagonalak dira, honelako (J_q, J_q) blokeek, diagonalean, aukeratutako modalitateen kopuruak dituzte eta beste elementuak zero dira bi modalitate (edo erantzun) desberdin batera ezin baitira aukeratu.

($J_q, J_{q'}$) blokeak, benetazko kontingentzi taulak dira bi galderen erantzunak gurutzatu baitira.

B deitutako Burt-en taula, matrizialki, **Z** taularen bidez honela lortuko da:

$$\mathbf{B} = \mathbf{Z}^T \mathbf{Z}$$

Korrespondentzi Analisia aplikatu ondoren lortutako faktoreak interpretatzerako, KAF-ean lortutako erlazioak eta interpretaziorako laguntzak (ekarpen absolutuak eta erlatiboak) baliagarriak dira.

XII 5. ELEMENTU GEHIGARRIAK

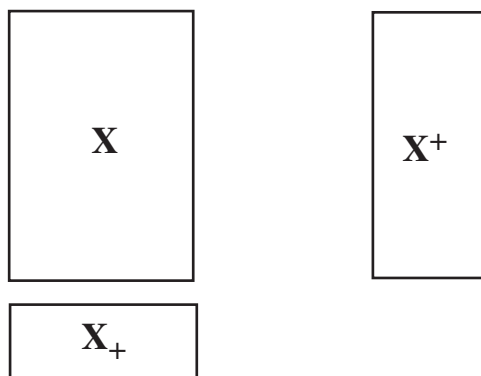
Korrespondentzi Analisi Faktorialean (bikoitza nahiz anizkoitza) eta Osagai Nagusizko Analisisan *elementu gehigarrien teknikak* asko laguntzen du analisiaren interpretazioan.

Laginetik at, n aldagaien gain ditugun indibiduo batzuren balioak ezaguturik, analisiaren lekuko bezala, hodeian, talde hori kokatzea komenigarria gerta daiteke.

Batzutan, ezaugarri interesgarri bat daukaten indibiduo-multzoak izanik, berorien grabitate-zentruak indibiduo gehigarri bezala hartuko dira. Azkenik, jatorrizko laginerako beste aldagai batzuren balioak baztertuak izanik datu-multzoan homogenoak ez direlako, gehigarri bezala aurkeztea komenigarria izatea gerta daiteke.

Arazoa, bada, hauxe da: Errenkada-puntu edota zutabe-puntu horien koordinatu faktorialak lortzea, honela ardatz eta plano faktorialetan aurkeztu ahal dira.

Eskematikoki,



X^+ nahiz X_+ tauletan ditugun elementuen koordinatu faktorialak, elementu aktiboenak bezala oso erraz kalkulatu dira eta eskola frantsesaren datu-analisiaren esparruan egiten diren estatistikaren programa-paketetan ahalbide bezala aurki daitezke.

Elementu hauek nahiz eta analisisan, aktiboki, parterik hartu ez, ardatz eta planotan proiektatzean, batzutan, kokapen pribilegiatua edukiz asko lagundu ahal dute emaitzen azalpenetan.

XII.6. SAILKAPEN-METODOAK

Gaiaren hasieran esaten genuenez estatistika deskribatzaile anizkoitzean bi analisi-mota egin daitezke eta biak datu-hodei berberan aplikagarriak dira.

Orain arte ikusi ditugun analisi faktorialen bidez hodeiaren direkzio luzatuenak bilatzen dira eta beraien gain bildutako sakabanatzerik adierazgarriena azaltzen da.

Sailkapen-metodoen bidez, ordea, hodeiaren elementu (indibiduo) hurbilenak sailetan kokatzen saiatuko gara. Lortutako sailak beraien elementutan homogenoak izango dira eta sailen arteko desberdintasunak ahal den hoberen finkatuko dira.

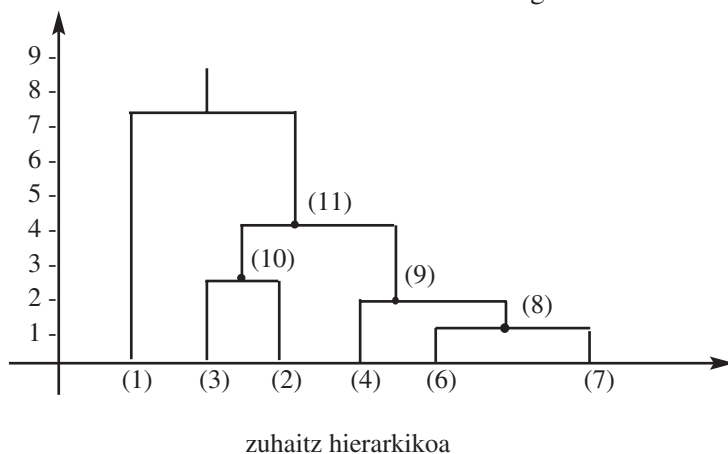
Normalki, sailkapen-metodoak analisi faktorialaren emaitzak osotzeko eta aberasteko erabiltzen dira, kasu hauetan indibiduoek dituzten lehen koordinatu faktorialetatik abiatuz eta sailkapen-algoritmo egoki bat aplikatuz sailak lortuko dira.

XII.6.1. Goranzko Sailkapen Hierarkikoa

Sailkapen automatikoko moten artean, zentruaren ingurukoa, beheranzko sailkapen hierarkikoa eta goranzko sailkapen hierarkikoa ditugu, azken hau da garrantzitsuena duen koherentzia logikoagatik, hain zuzen.

Goranzko sailkapena egitean elementuak binaka elkartzen dira, honela, lehen etapan, bi hurbilenak elkartu ondoren $m-1$ elementu geratzen dira. Elkartuz lortu den elementua besteak bezala izanik, bigarren etapan, berriro bi hurbilenak elkartu ondoren $m-2$ elementu geratzen dira; honela jarraituz, azkenean, sail bat besterik ez da geratzen, honek elementu guztiak dituelarik.

Sortzen den zuhaitz hierarkikoan ebaketa bat eginez sailak lortzen dira.



Dakusagunez, lehen etapan 6, 7 elementuak elkartu dira eta elkarketa honi, (8) adabegiaz, ordenatuen ardatzean elkartze-indize txikiena dagokio, bigarren etapan 4, (6,7) elementuak elkartu dira (9) adabegian, hirugarrenean 3,2,.....

(8), (9), (10),..... hierarkiaren adabegiak deitzen dira.

Elementuen arteko hurbilketak lehendik definitutako distantzia baten bidez finkatuko dira.

Goranzko sailkapen hierarkikoa erizpide desberdinen arabera egin daiteke, (saltu txikiena, batezbesteko distantzia, barizentrua,...), hauen artean bariantzaren metodoa delarik garrantzitsuen.

XII.6.1.1. Bariantzaren metodoa

Bariantzaren metodoaren baldintza bakarra sailkatu nahi ditugun elementuak espazio euklidear batean aurkeztuak izan ahal izatea da. Honek zera esan nahi du: I multzoko m elementuak n aldagaien arabera edo n zutabeen arabera kontingentzi taulatan bezala, karakterizatuak daude. m elementuak $\mathbb{R}^n$ espazio bektorialeko bektoreak dira, non distantzia euklidear bat definitua baita.

Bi dimentsioko espazioan hiruki zuzen batean Pitagoras-en teorema betetzen duen edozein distantziari, distantzia euklidearra deritzo.

“Aldagai-indibiduo” taulatan ohizko distantzia euklidearra definitzen da (ikus gai honen 3. atala) eta kontingentzi taulatan distantzia euklidear berezi bat χ^2 distantzia da (ikus gai honen 4. atala), hain zuzen. Ondorioz, Goranzko Sailkapen Hierarkikoa bariantzaren metodoaz edozein kasutan aplikatu daiteke.

Bariantzaren metodoa ez da bakarrik distantzien kalkuluan oinarritzen, inertzien kalkuluetan ere oinarritzen da, inertzia elementuen masa bider distantzia karratuen batura bezala definiturik. Kalkuluan sartzen diren elementuen masen batura bat baldin bada, inertzia bide batez bariantza izango da.

Köning-Huygens-en teorema

Suposa dezagun sailkatu nahi ditugun $N(I)$ hodeiaren m puntuei x_i deritzegula eta p_i berorien masak direla.

$g \in \mathbb{R}^n$ eta $y \in \mathbb{R}^n$ hodeiaren grabitate-zentrua eta hodeiaren edozein puntu izanik:

$$I_N(y) = \sum_{i=1}^m p_i(x_i, y)^2$$

Hau da, hodeiaren inertzia y puntuarekiko, puntuen masak bider puntu horrekiko distantzia karratuen batura da.

Köening-Huygens-en teoremaren bidez zera dakigu:

$$I_N(y) = d^2(g, y) + I_N(g)$$

hau da, $I_N(g)$ grabitate-zentruarekiko hodeiaren sakabanatzea minimoa dela.

Sortutako sailak beraien elementuetan homogenoak izatea eta sailen arteko desberdintasunak ahal den hoberen finkatuta izatea nahi dugunez, homogenotasun neurria bezala “sail barneko inertzia”-ren kontzeptua eta desberdintasun-neurria bezala “sail arteko inertzia”-ren kontzeptua hartuko dira.

$N_1, N_2, \dots, N_k, \dots, N_h$; $N(I)$ hodeiko h sail disjuntu izanik eta $g_k, P(N_k), N_k$ sailaren grabitate-zentrua eta masa izanik:

$$I_{N_k}(g_k) = \sum_{x_i \in N_k} p_i d^2(x_i, g_k)$$

Hau da, N_k sailaren grabitate-zentruarekiko daukan “sail barneko inertzia” $I_{N_k}(g_k)$ balioa da.

$N(I)$ hodeiaren g grabitate-zentruarekiko “sail arteko inertzia” hauxe izango da:

$$\sum_{k=1}^h p(N_k) d^2(g_k, g)$$

Köening-Huygens-en teorema aplikatuz:

$$I_N(g) = \sum_{k=1}^h I_{N_k}(g_k) + \sum_{k=1}^h p(N_k) d^2(g_k, g)$$

hau da, $N(I)$ hodeiaren inertzia g grabitate-zentruarekiko sail barneko inertzien baturan gehi sail arteko inertzian deskonposatzen da batukorki.

$I_N(g)$ konstante izanik, sail barneko inertzien batura minimizatzea eta sail arteko inertzia masimizatzea baliokidea da.

Algoritmoaren konkrezioa

Elkartzearen erizpidea, $N(I)$ hodeiaren inertiaren galtzea minimoa izatea da. Honela, lehen, klaseak elementu guztiak izanik, sail barneko inertzia minimoa daukaten bi klaseak elkartuko dira.

x_i, x_j lehen fasean ditugun m klaseen bi elementu izanik, sail barneko inertzia hau da:

$$\Delta I_{ij} = \frac{p_i p_j}{p_i + p_j} d^2(x_i, x_j)$$

honela, ΔI_{ij} minimoa daukaten bi elementuak elkartuko dira.

Sailkapenaren hurrengo klasetan z_k eta x_i, x_j elkartzean sortu den beste edozein klaseren arteko sail barneko inertzia, azken honen grabitate-zentrua x izanik ondoko errepikapen-formulaz lortzen da.

$$\Delta I_{z,x} = \frac{1}{p_k + p_i + p_j} [(p_k + p_i) \Delta I_{ki} + (p_k + p_j) \Delta I_{kj} + p_k \Delta I_{ij}]$$

Era honetan, fase bakoitzean inertiaren galtze minimoa eragiten duten bi sailak elkartzean, hodeiarekiko sail barneko inertzia minimoa eta bide batez sail arteko inertzia masimoa edukitzea lortzen da.

Dakigunez, elementu guztiak sail batean elkartzean prozesua amaituko da.

Hodeiaren partizio ezberdinak zuhaitz hierarkikoaren ebaketaz lortu ahal dira.

XII.6.2. Adibidea

Hamazortzi Hego eta Ertamerikako estatuei buruz egindako Osagai Nagusien Analisiaren azterketan (3.4 atala) hiru lehen faktoreak kontutan harturik, hau da, hauen gain estatuek dituzten koordenatuen arabera sailkapen automatikoa egiteaz ikasketa osotu nahi dugu.

Goranzko sailkapen hierarkikoa bariantzaren metodoaz aplikatu ondoren lortutako emaitzak ondokoak dira:

Elkartze- Masa -indizea		Indibiduoa	
1.000	.013	6	
1.000	.042	3	
1.000	.174	1	
1.000	.005	7	
1.000	.035	5	
1.000	.416	2	
1.000	.000	14	
1.000	.002	9	
1.000	.025	13	
1.000	.000	17	
1.000	.001	16	
1.000	.116	10	
1.000	.002	12	
1.000	.007	8	
1.000	.009	4	
1.000	.002	15	
1.000	.150	11	
1.000	----	18	

Elkartze-indizeak salto handia eman duelako zuhaitzaren adarrak nahiko luzatzen direnean ebaketa eginez bost sail ditugu.

Dakusagunez, zuhaitzaren oinarrian sailetan ditugun indibiduoek datu-taulan daukaten posizioaren arabera ditugu; halaber, indibiduoen masa 1 dela ikus daiteke.

Aplikaturako SPAD programen bidez sailen deskripzio zehatza eta lehen plano faktorialean estatu guztiak bere sailen zenbakiaz adierazita lortzen ditugu:

FIN DE L - ETAPE ** SEMIS **

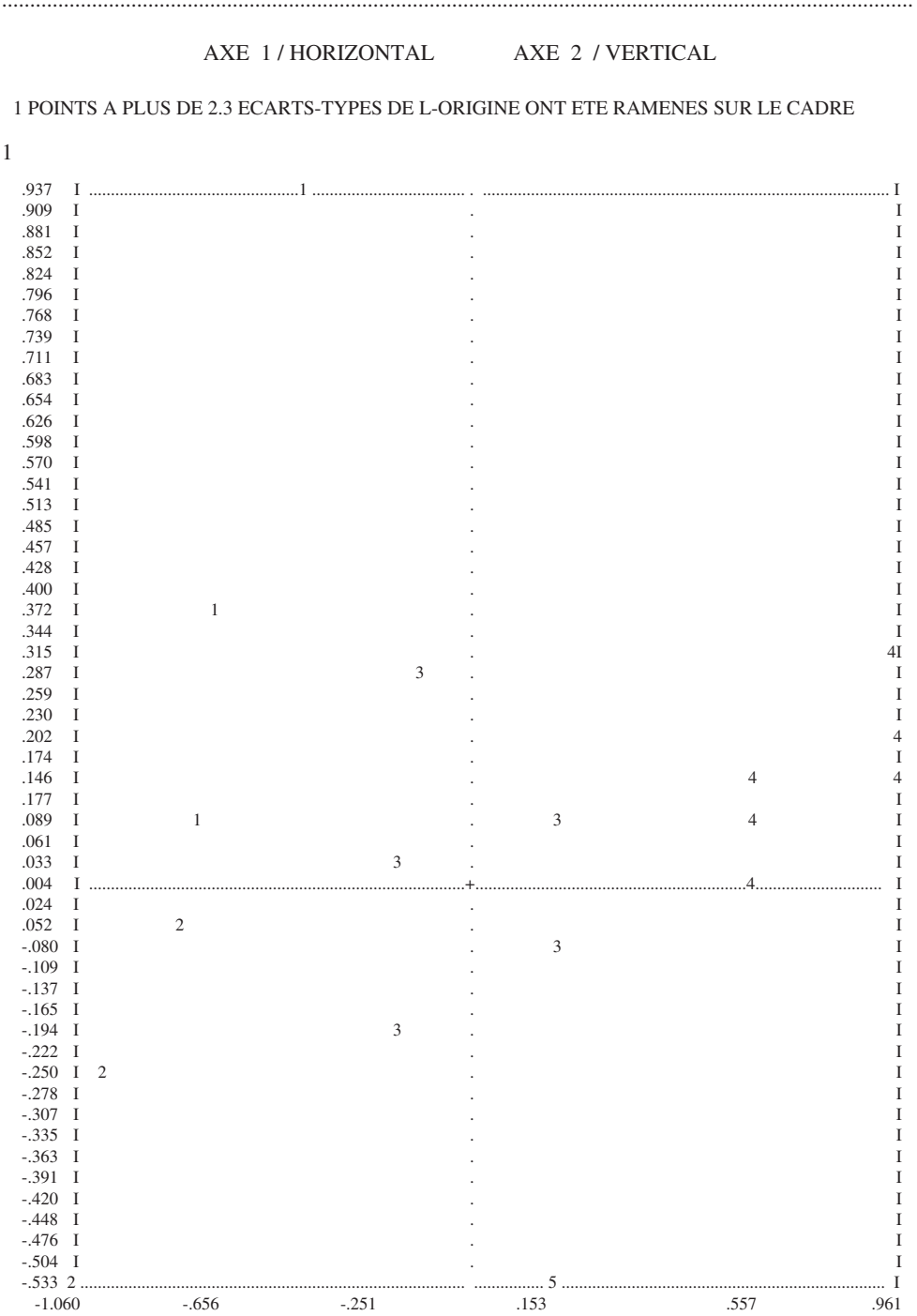
CLASSE EFFECTIF POIDS CONTENU (ANCIENNES CLASSES)

1	3	3.000	1	3	6				
2	3	3.000	2	5	7				
3	5	5.000	4	8	11	12	15		
4	6	6.000	9	10	13	14	16	17	
5	1	1.000	18						

VEN / 1 ARG / 2 MEX / 1 PAN / 3 URU / 2 BRA / 1 CHI / 2 COL / 3 CRI / 4 GUA / 4 ECU / 3
PER / 3 PAR / 4 NIC / 4 RDO / 3 HON / 4 ELS / 4 BOL / 5

GRAPHIQUES DE DENSITE DES 18 INDIVIDUS

PLAN DE PROJECTION DES 18 POINTS SUR LES AXES 1 ET 2



Dakusagunez, plano faktorial nagusian, sailak oso ondo bereiziaz agertzen dira. Lehen sailean eta goitik behera, Venezuela, Mexiko eta Brasil. Bigarrenean, Chile, Uruguay eta Argentina, hego-konoa hain zuzen. Hirugarrenean, Panama, Colombia, Peru, Ecuador eta R. Dominicana. Laugarrenean C. Rica, Nicaragua, Paraguay, Guatemala, Honduras eta El Salvador. Bostgarrenean, azkenik, duen inflazioagatik hain berezia den Bolivia.