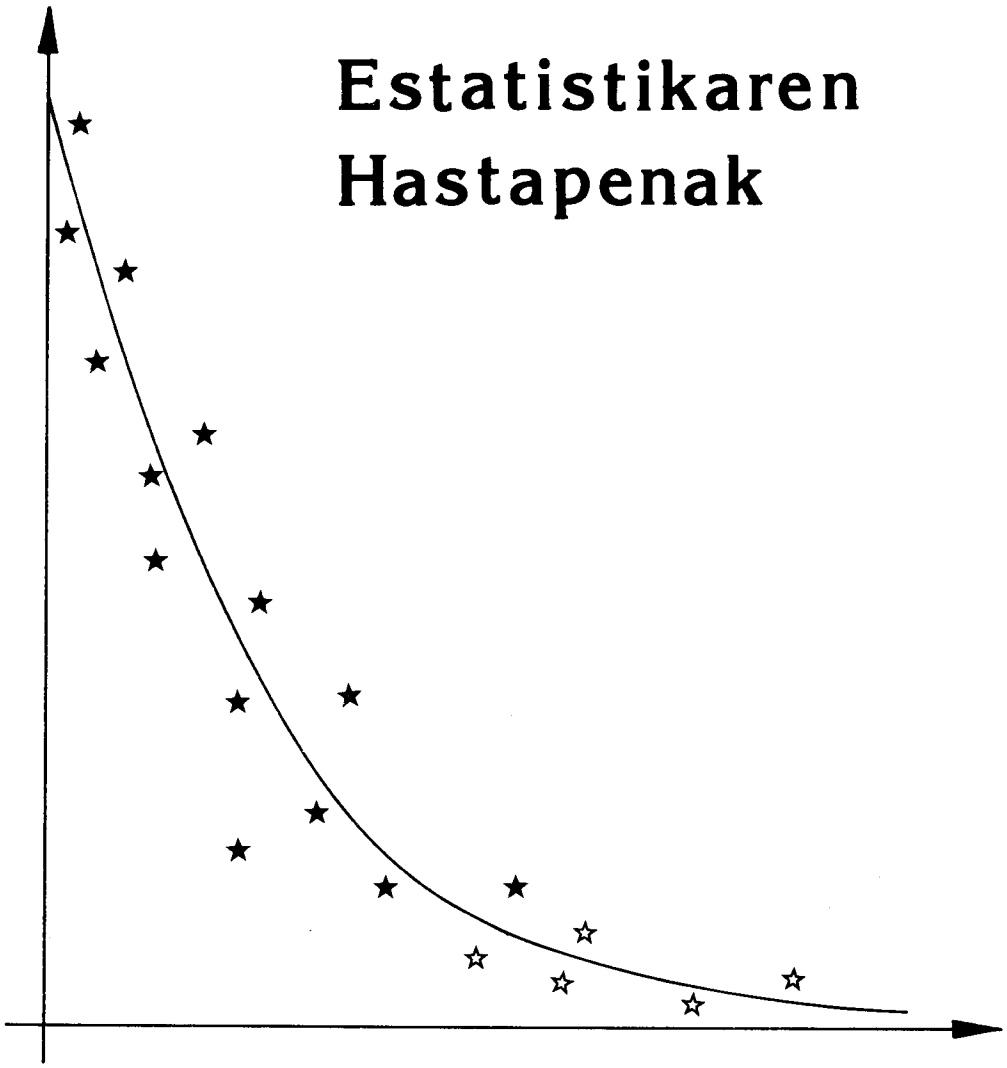
Estatistikaren Hastapenak



udako euskal unibertsitatea

IRUÑEA 1983

E S T A T I S T I K A R E N H A S T A P E N A K

K. Fernandez Agirre
1982-1983 ikasturtea
SARRIKO (BILBO)

JABEGOA: U.E.U.ko Estatistika Saila
I.S.B.N.: 84-300-9855-0
LEGE-GORDAILUA: BI-1363-83
INPRIMATEGIA: BOAN S.A.

A U R K I B I D E A

HITZAURREA IV

I - EZAUGARRI ESTATISTIKO BAKUNAK, BANAKETAK, TAULAK, ADIERAZPIDE GRAFIKOAK 1

1. ESTATISTIKAREN HISTORIA LABURRA. ESTATISTIKA ETA - PROBABILITATEA.
2. OROKORTASUNAK.
3. EZAUGARRI ESTATISTIKOAK.
4. MAIZTASUN-BANAKETA BAKUNAK, TAULAK.
5. MAIZTASUN-BANAKETA BAKUNEN ADIERAZPIDE GRAFIKOAK.

II - EZAUGARRI BAKUNEN BALIO TIPIKOAK 19

1. MOMENTUAK.
2. MAIZTASUN-BANAKETA BAKUNEN BALIO TIPIKO EDO ESTATISTIKOAK.
   2.1. ZENTRU-JOERAREN BALIO TIPIKOAK.
   2.2. SAKABANATZEAREN BALIO TIPIKOAK.
   2.3. ITXURAREN BALIO TIPIKOAK.
   2.4. KONTZENTRAZIO-NEURKETAK.

III - ZENBAKI-INDIZEAK 53

1. INDIZE SINPLEAK.
2. PONDERAZIO GABEKO INDIZE KONPLEXUAK.
3. INDIZE KONPLEXU PONDERATUAK.
4. INDIZE KONPLEXUEN ERAIKETAN SORTZEN DIREN ZENBAIT - ERAGOZPEN.
5. ERAGOZPEN BEREZI BATZU.
6. ZENBAKI-INDIZEEN APLIKAPENAK.

IV - EZAUGARRI ESTATISTIKO BIKOITZAK,BANAKETAK,TAULAK, ADIERAZPIDE GRAFIKOAK 87

1. EZAUGARRI ESTATISTIKO BIKOITZAK.
2. MAIZTASUN-BANAKETA BIKOITZAK.
3. MAIZTASUN-BANAKETA BIKOITZEN ADIERAZPIDE GRAFIKOAK.
4. MAIZTASUN-BANAKETA BALDINTZATUAK.
5. TAULA BATEN DEPENDENTZIA EDO INDEPENDENTZIA ESTATISTIKOA.

V - EZAUGARRI BIKOITZEN BALIO TIPIKOAK 103

1. MOMENTUAK.
2. MAIZTASUN-BANAKETA BIKOITZEN BALIO TIPIKO EDO ESTATISTIKOAK.
3. ALDAGAIEN ALDAKETAK.
4. SAKABANATZE ETA KORRELAZIO-MATRIZEAK.
5. OROKORPENA.

VI - KORRELAZIOA ETA ERREGRESIOA R^2-n 119

1. KORRELAZIOA ETA ERREGRESIOA.
2. ERREGRESIO TEORIA.
   2.1. BATEZBESTEKOAREN ERREGRESIOA.
   2.2. KARRATU TXIKIENEN ERREGRESIOA.
      2.2.1. KARRATU TXIKIENEN ERREGRESIO LINEALA.
      2.2.2. KARRATU TXIKIENEN ERREGRESIO PARABOLIKOA.

VII - ERREGRESIO ANIZKOITZA 141

1. SARRERA.
2. ERREGRESIO-PLANOA R^3-n.
3. ERREGRESIO-HIPEPLANOA R^n-n.
   3.1. KARRATU TXIKIENEN ERREGRESIOAREN PROPIETATEAK.
   3.2. KOEFIZIENTE MUGATZAILEA ETA KORRELAZIO-KOEFIZIEN

TE ANIZKOITZA.

4. "BETA" ERREGRESIO-KOEFIZIENTE STANDARDIZATUAK.

VIII - KORRELAZIO PARTZIALA 159

1. SARRERA.
2. KORRELAZIO PARTZIALA : KONTZEPTU MATEMATIKOA.
3. KORRELAZIO-KOEFIZIENTE PARTZIALA ETA "β" KOEFIZIENTEAK.

E R A N S K´I N A 175

S.P.S.S. GIZARTE ZIENTZIETARAKO ESTATISTIKA-PAKETEA

1. SARRERA.
2. INFORMAZIOAREN SARRERA.
3. SPSS-rako DATUEN PRESTAKETA.
4. SPSS PROGRAMAREN PRESTAKETA.
5. SPSS KONTROL-TXARTELAK.
6. SPSS TRATAMENDU-TXARTELAK.
7. ANALISI ESTATISTIKOAK LORTZEKO AZPIPROGRAMAK.

H I T Z A U R R E A

Urte mordoxka bat ba da, Sarrikoko Ekonomi eta Enpresa - Zientzien Fakultatean, Matematika eta Estatistika arloetan klaseak euskaraz emateko saioak egiten hasi ginenetik.

Garai hartan Estatistika sailera Tolosar baten etorrera , Anjeles Iztuetarena hain zuzen, bultzatzailea izan zen, biok - boluntarismo hutsez eta Jose Mari Piris-en laguntzaz hasi bait ginen.

Bide batez, U.Z.E.I.-ko Matematika Hiztegia Mikel Zalbiderren zuzendaritzapean moldatzen ari zen eta halaber gu baino dinamika aurreratuago eta indartsuagoarekin Leioako Zientzi Fakultateko taldeak han zihardutenFisika, Kimika, Matematika, - Geologia eta Biologia euskaraz egiten.

Gure lehenbiziko saioak ondo atera zirela eta Fakultateko euskararako batzordeak proposaturik, bigarren kurtsoko Estatistikarako Sarrera ofizialki euskaraz ematea onartu zen 1980-1981 ikasturtean.

Eskutartean daukazun textua, azkeneko hiru ikasturteetan idatziriko apunteen ondorioa da, baina historia apur bat ere - ba dauka, asignatura berak izan duen historiarekin loturik.

"Introducción a la Estadistica" delako asignatura lehen lau hilebetez bakarrik ematen zen; gero urte osorako asignatura bihurtuz, programa horren gaiak euskaraz idatzi nituen, - ikasleentzako euskarazko apunteak edukitzea premiazkoa bait - zen.

Hurrengo ikasturtean, 1981-1982 ikasturtean alegia, asignaturaren programa birmoldatu egin zen, lehen ematen zen Probabilitate-Teoria kenduz eta bide batez Estatistika Deskribatzaileko gai batzuren tratamenduan sakonduz.

Azkenik, asignatura beste alde batetik ere birmoldatu da, batez ere ikasturte honetan 1982-1983.an alegia, konputagailuen erabilpena aztertuz. Honela, SPSS estatistikaren programa-paketeari buruzko eranskina sartu dugu.

Ikasgai hauek Lizentziaturako ikasketetan ikasle guztiek ikasten duten Estatistika Orokorraren lehenengo zatia besterik ez dira; horregatik Estatistikaren aldetik guztiz ezosoak dira Estatistika Deskribatzailearen aldetik osoagoak izan arren.

Amaitu aurretik eskerrik beroenak neure saileko lankide guztiei, beriziki Jose Mari Piris-i, hainbat ikasgaietan beraren laguntza baliotsua izan zaidalako.

Modu berean, eskerrik beroenak Martxel Aizpuruari, Xabier Kintanari eta Jose Ramón Etxebarriari, hiru ikasturte hauetan zehar euskararen aldetik nire lanak zuzentzen iragan dituzten orduengatik.

Aipamen berezia merezi dute originalen mekanografiatzean hamaika ordu eman dituzten Mari Carmen Menikak eta Feli Verdejo-k. Mila esker beraiei.

Sarrikon, 1983.eko Ekainaren 30ean.

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

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

I.2. OROKORTASUNAK

2.1. POPULAZIOA ETA LAGINA
2.2. UNITATE ESTATISTIKO EDO INDIBIDUOAK

I.3. EZAUGARRI ESTATISTIKOAK

3.1. EZAUGARRI ESTATISTIKO BAKUNAK

3.1.1. EZAUGARRI KUANTITATIBOAK
3.1.2. EZAUGARRI KUALITATIBOAK

I.4. MAIZTASUN-BANAKETA BAKUNAK

4.1. MAIZTASUN-BANAKETAK (ABSOLUTUAK,ERLATIBOAK ETA METATUAK)
4.2. TAULAK

I.5. MAIZTASUN-BANAKETA BAKUNEN ADIERAZPIDE GRAFIKOAK

5.1. BARRAZKO DIAGRAMAK
5.2. HISTOGRAMAK
5.3. MAIZTASUN POLIGONOAK
5.4. DIAGRAMA LINEALAK
5.5. BESTE ADIERZAPIDE 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 neolitikoan herriak eta salerosketak gero eta ugariagoak egin zirenez, estatistikaren beharra sortu zen.

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

Baina "cinquecento" italiarrean (XVI. mendean) izena eta bere buruaren kontzientzia pixka bat 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 bultzaketa dela medio.

Etapa honetan, oraindik Estatistika ez da zientzia, eta gertaera aleatoriotatik eta probabilitate teoriatik urrun zegoen arte bat zen.

Baina geroztik "estatistikariei" problemak agertzen hasi zitzaizkien:

- estatistiketatik ondorio zientifikoak nola lortu?

- kolektibo aldakor edo zentsatzeko zail direnetatik informazio fidagarriak nola lortu?

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

Estattistika 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 KONKRETUAK ⟷ KOLEKTIBO ABSTRAKTUAK

Fenomenoetatik bere kausetarainoko bidea agiten duen ezaupide hau, INDUKZIOA deitzen da eta, hementxe, INFERENTZIA ESTATISTIKOA.

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

Hau dena INFERENTZIA ESTATISTIKOARI, dagokio.

Apuntu-liburu honekan ikusten diren gaiak Estatistika Deskribatzaileari dagozkio, hau da, oinarrizko Estatistika ikusiko dugu edo Estatistikaren hastapenak.

Dakusagun, bada, gaian sartu aurretik, ESTATISTIKA DES - KRIBATEAILEAREN definizio bat.

ESTATISTIKA DESKRIBATZAILEA, datu-multzo HANDIAK aurkez di tzakeen teknika bat da, multzo hau eskurakoia eginez beraren - egitura gardenduz eta beraren barne-harremanak neurtuz.

I.2. OROKORTASUNAK

2.1. POPULAZIOA ETA LAGINA

Ikerketa estatistiko bakoitzean objektu nagusi diren pertsona, ondasun, saioaren emaitza (adibidez: datoaren aurpegia) probintzia baten herri, hauteskundearen alderdi politiko, edo ta 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 edo infinituak (adz: boltsa bate - tako erauzketak) ala bukakorrak edo finituak (eskualde batetako enpresa kooperatiboak) izan daitezke.

LAGINA: Lagina azpipopulazio errepresentagarri bat besterik ez da; populazio batetaz 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 populazio 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-TEORIAri dagokio.

Lagina osatzen duten elementuen kopuruari laginaren tamainua deritzo.

Sinbolikoki:

Ω = Populazioa

Ω' = Lagina $\quad \Omega' = \{\omega_1\ \omega_2 \quad \omega_N\} \subset \Omega$

N = Laginaren tamainua.

2.2. UNITATE ESTATISTIKO EDO INDIBIDUOAK

Populazioa edo lagina osatzen duten elementu bakoitzari unitate estatistiko edo indibiduo deritzogu.

Indibiduo izena, estatistika deskribatzailearen jatorri demografikoari zor diogu.

I.3. EZAUGARRI ESTATISTIKOAK

Ikerketa estatistiko batetan helburua ez dugu populazioa - edo lagina exaustiboki ikertzea, baizik eta beraren ezaugarri - batetara edo batzutura 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:

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

$$\Omega = \{\omega_1, \omega_2, \ldots\ldots \omega_N\} \longrightarrow \{(x_1y_1), (x_2y_2) \ldots (x_ny_n)\}$$

3.1. EZAUGARRI ESTATISTIKO BAKUNAK

Ezaugarri estatistiko bakunen sailkanena bere balio mul - tzo desberdinen arabera:

- EZAUGARRI ESTATISTIKO BAKUNAK
  - KUANTITATIBOAK (edo ZENBAKIZKOAK)
    - JARRAIAK
    - DISKRETUAK
  - KUALITATIBOAK (edo NOLAKOTASUNEZKOAK)

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. (¡)

Balio biren artean edozein balio har dezakenean ezaugarri jarraia deritzo.

3.1.2. Ezaugarri kualitatiboak

Ezaugarriak har ditzakeen balio indibidualak kategoria edo

(¡) balio-multzoko elementuak puntu isolatuak direnean balio-mul tzo diskretu bat dugu.

atributuak direnean, edo eta "balio-multzoa" kategoria edo a - tributuek osatzen dutenean (diskretuak beraz), ezaugarri kuali tatiboa deritzo.

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

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

- Exklusibitatea, hau da, indibiduo bat ezin liteke bi ka tegorietan egon.

- Exaustibitatea, hau da, indibiduo guztiek nonbait klasi fikatuak egon behar dute.

ABIDIDEAK: Eskualde batetako enpresa kooperatiboen lagin bat aztertzean:

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

I.4. MAIZTASUN-BANAKETA BAKUNAK

4.1. MAIZTASUN-BANAKETAK (absolutuak, erlatiboak, eta me tatuak).

Ezaugarri diskretuetan:

Suposa dezagun N indibiduoz osatutako lagin bat, non ezau garri bakun batek "m" balio edo kategoria desberdin hartzen - baititu.

Hots:

$$\{\omega_1, \omega_2, \ldots \omega_N\} \longrightarrow \{x_1, x_2, \ldots\ldots, x_j \ldots\ldots x_m\}$$

$$\{\omega_1, \omega_2, \ldots \omega_N\} \longrightarrow \{A, B, \ldots\ldots, J, \ldots\ldots M\}$$

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 indibiduoen kopurua erlatiboki kontutan hartuz lortzen den proportzioa da.

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

Orduan:

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

Laginari dagokion balio-multzoa ordenatu ahal badugu (ezau garria kuantitatiboa bada, beti; bestela ez beti), txikitik han dira ordenatuko dugu.

Hots:

$$x_1 \leqslant x_2 \leqslant \cdots\cdots \leqslant x_j \cdots\cdots \leqslant x_m$$

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

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

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

eta: x_j balioaren MAIZTASUN ERLATIBO METATUA sinbolikoki F_j honela definituko dugu, non:

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

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

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

Edozein maiztasun erlatibo bezala, portzentaiatan ere a - dierazten dira.

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

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.erlatibo metatuak
1	x_1 A	n_1	N_1	f_1	F_1
2	x_2 B	n_2	N_2	f_2	F_2
..	.. ..	..		..	
::	x_j J	n_j	N_j	f_j	F_j
::	.. ..	..		..	
N	x_m M	n_m	N	f_m	1

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

Ezaugarri jarraietan

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

Hots:

$$\{\omega_1 , \omega_2 , \ldots , \omega_N\} \xrightarrow{X} I = (a,b) \subset R$$

bainan praktikan:

$$\{\omega_1 , \omega_2 , \ldots , \omega_N\} \xrightarrow{X} \{x_1 , x_2 , \ldots , x_n ..\} \subset (a,b) \subset R$$

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

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

non:

$$a \leqslant a_o \leqslant a_1 \leqslant a_2 \leqslant \dots \leqslant a_{n-1} \leqslant a_n \leqslant b$$

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

- Ondo definituak.
- Exklusibitatea.
- Exaustibitatea.

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

Normalki ordezkaria, maila tartearen erdiko puntua da,bainan goi ala behe muturrean ere izan liteke.

Honela bada, balio-multzoko maila bakoitza, balio batetara mugatzen dugu (ordezkariaren balioara) 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
1	(a_o, a_1)	x_1	n_1	N_1	f_1	F_1
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
..	.. ..	..	..	..	..	..
N	(a_{m-1}, a_m)	x_m	n_m	N	f_m	1
			$\sum_{j}^{m} n_j = N$		$\sum_{j}^{n} f_j = 1$	

ADIBIDEAK

a) X ezaugarri kualitatiboa da.

Eskualde batetako pupulazioaren banaketa produkzio-sektoreen arabera:

Sektoreek	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) Diskretua

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 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 baino gehiago			2	100

(Kasu honetan, mailaketan, maila guz - tiek tamainu berdina dute, aurrenekoa eta azkenekoa kenduz).

I.5. MAIZTASUN-BANAKETA BAKUNEN ADIERAZPIDE GRAFIKOAK

Nahiz eta maiztasun-banaketen taulak informazio guztia har tu (balioak mailetan laburtuak ez badira izan behintzat), grafi koki adierazteak, asko laguntzen du ikusarazten eta uletzen.

Adierazpide grafiko asko erabili izan arren, batzu baka - rrik ikusiko ditugu.

5.1. BARRAZKO DIAGRAMAK

Oso baliagarriak zaizkigu ezaugarriak diskretuak ditugu - nean (kuantitatiboak zein kualitatiboak).

Abzisa-ardatzean balio edo kategoria desberdinak (ahal ba da ordenaturik), ezarriko ditugu, eta ordenatu-ardatzean be - rriz laginari dagozkion maiztasunak (absolutuak, erlatiboak ala taldekatuak).

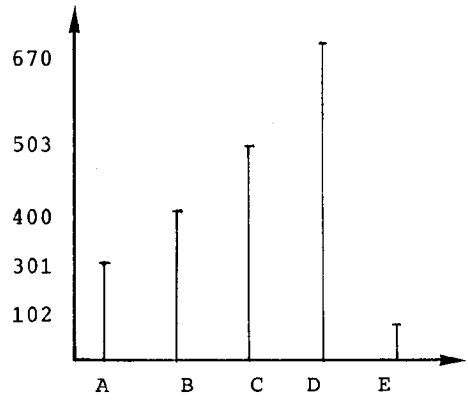
ADIBIDEAK

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

- taula -



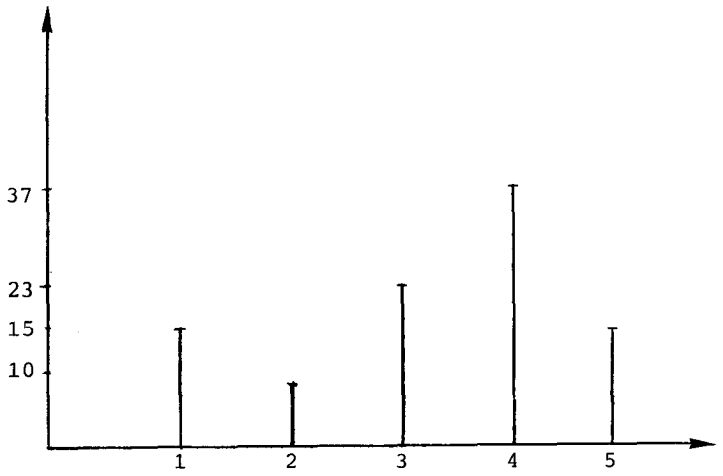
- barrazko diagrama -

b) Ezaugarri kuantitatibo diskretua.

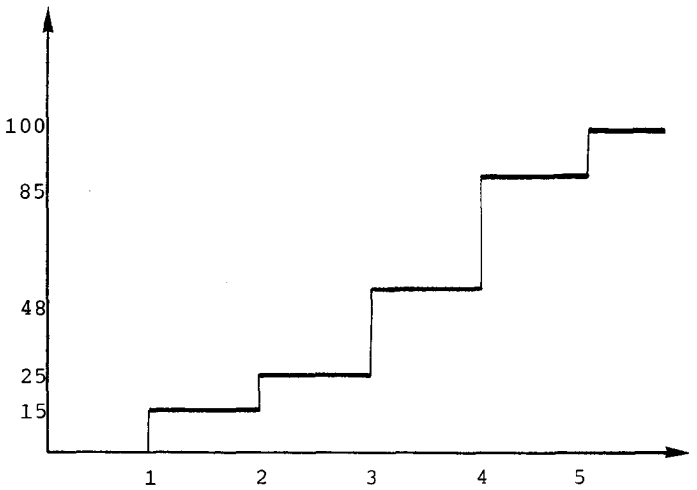
Loteen banaketa,beren baitan dituzten pieza txarren kopuruaren arabera:

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

- taula -



- Maiztasun erlatiboen barrazko diagrama -



- maiztasun metatuen diagrama -

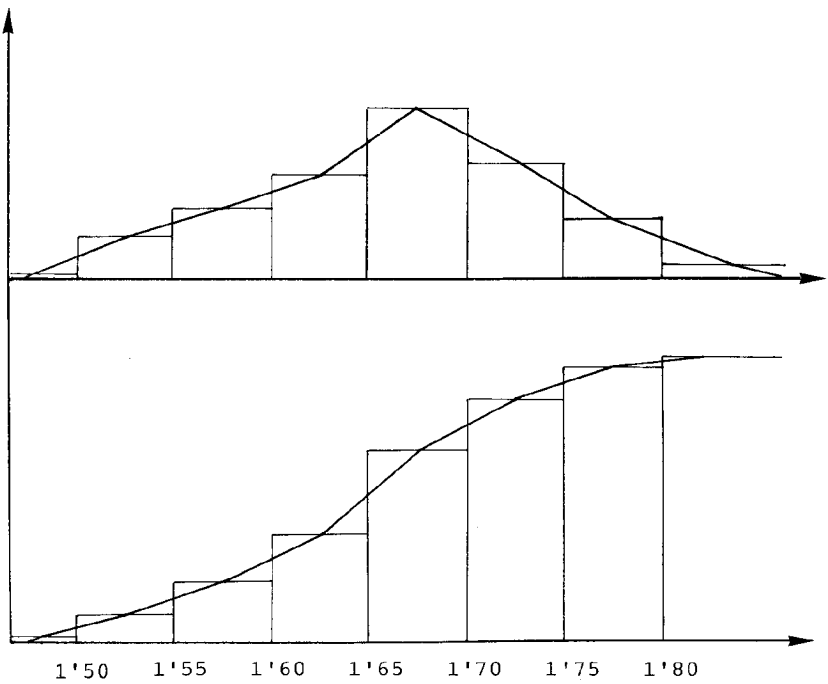
5.2 HISTOGRAMAK

Oso baliagarriak zaizkigu ezaugarriak jarraiak direnean.

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

Maila guztiak tamainu berdinekoak baldin badira, maila bakoitzari dagokion laukizuzenaren altuera, bere maiztasun absolutu edo erlatiboa da.

Ondoan b-2) adibideari dagozkien histogramak eta maiztasun-poligonoak ditugu (Ikus b-2) 12.orrialdean).



- histogramak eta maiztasun poligonoak -

5.3. MAIZTASUN POLIGONOAK

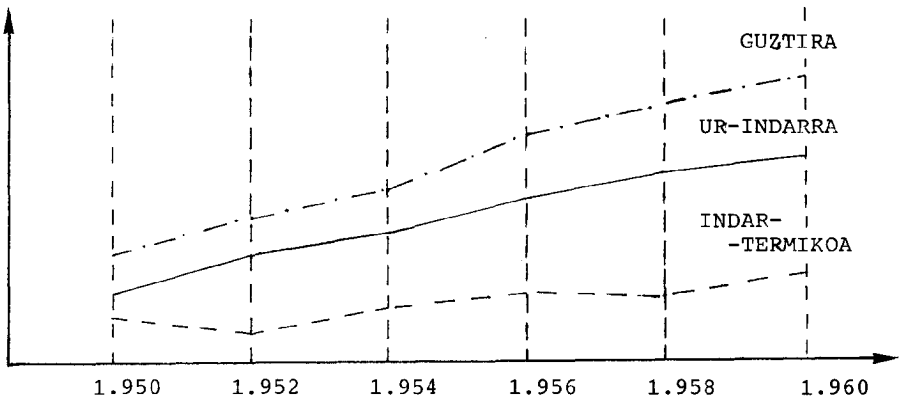
Maiztasun-poligonoa laukizuzenen goiko oinarriko erdiko - puntuak marra batez loturik eraikitzen da.

Honela egin dugu aurreko adibidean.

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:

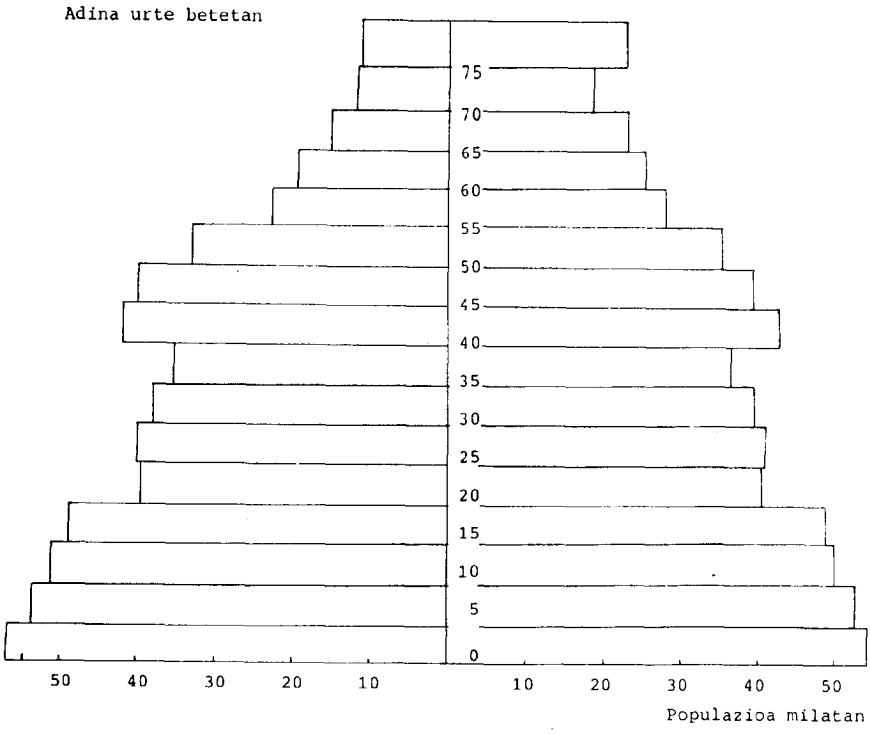


5.5. BESTE ADIERAZPIDE GRAFIKOAK

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

- POPULAZIO-PIRAMIDEAK.

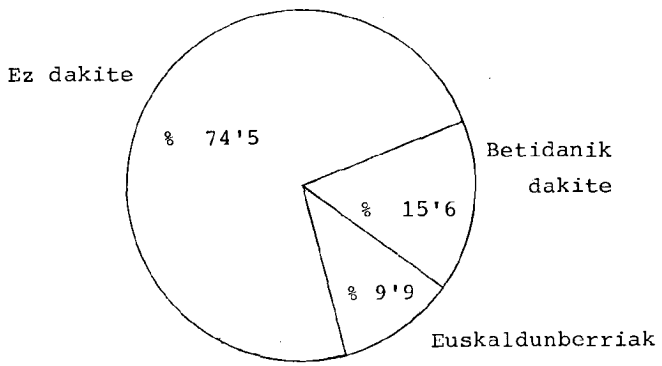
Histograma bikoitzak dira, eta honela eraikitzen dira: ordenatu-ardatzean adin-taldeak adierazten dira eta abzisa- - ardatzean ezkerrerantz eta eskuinerantz gizonezko eta emakumezkoen kopuruak (maiztasun absolutuak) edo erlatiboak neurtzen dira, dagozkien laukizuzenak eraikiz.



- Bizkaiko Populazio-Piramidea 1975.urtean -

- SEKTORE GRAFIKOAK

Grafiko hauek honela eraikitzen dira: zirkunferentziako - 360 gradutik, aldagaien balio ala kategoria bakoitzaren maizta sunari proportzionalki zati bat dagokio.



- Euskeraren ezaguera Sarrikoko Fakultatean 1979.urtean -

II. EZAUGARRI BAKUNEN BALIO TIPIKOAK

II.1. MOMENTUAK.

1.1. MOMENTU ARRUNTAK EDO JATORRIAREKIKOAK

1.2. MOMENTU ZENTRALAK EDO BATEZBESTEKOAREKIKOAK

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

2.1. ZENTRU-JOERAREN BALIO TIPIKOAK

2.1.1. BATEZBESTEKO ARITMETIKOA. PROPIETATEAK.

2.1.2. BATEZBESTEKO GEOMETRIKOA.

2.1.3. BATEZBESTEKO HARMONIKOA.

2.1.4. MODA.

2.1.5. MEDIANA. KOANTILAK.

2.2. SAKABANATZEAREN BALIO TIPIKOAK.

2.2.1. BARIANTZA.

2.2.2. BATEZBESTEKO DESBIDAZIOA.

2.2.3. IBILERA.

2.3. ITXURAREN BALIO TIPIKOAK

2.3.1. ASIMETRI KOEFIZIENTEA.

2.3.2. KURTOSI KOEFIZIENTEA.

2.4. KONTZENTRAZIO-NEURKETAK.

2.4.1. LORENZ-EN KURBA

2.4.2. GINI-ren INDIZEA

ARIKETAK

II.1. MOMENTUAK

Datuek berekin duten informazioa, balio tipiko edo esta - tistikoen bidez sintetizatzea komeni da.

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

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

Momentuak: - arruntak edo jatorriarekikoak.
- zentralak edo batezbestekoarekikoak.

izan daitezke.

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:

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

non: N= Laginaren edo Populazioaren indibiduoen kopurua

x_i= Ezaugarriaren i. balioa.

m= Ezaugarriak hartzen dituen balio desberdinak;

$m \leq N$ izango da.

Momentu arrunten bi berezitasun ikusiko ditugu:

1) 0-ordenako momentuak 1 balioa hartzen du:

Hots :

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

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.

Hala ere, Estatistikan eragile matematiko bezala satzen - dira, teoriaren garapenean oso erabilgarriak bait dira.

1.2. MOMENTU ZENTRALAK EDO BATEZBESTEKOAREKIKOAK

Maiztasun-banaketa bakun baten h-ordenako momentu zentrala edo batezbestekoarekikoa, sinbolikoki m_h ikurraz adierazi ko 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:

$$m_h = \frac{1}{N} \sum^{m} (x_1 - \bar{x})^h \, n_1 =$$

$$= \frac{1}{N} \sum^{m} \left(\binom{h}{0} x_i^h - \binom{h}{1} x_i^{h-1} \bar{x} + \binom{h}{0} x_i^{h-2} \bar{x}^2 - \ldots . + \right.$$

$$\left. + (-1)^{h-1} (h-1)\, h\, x_i \bar{x}^{h-1} + (-1)^h \binom{h}{h} \bar{x}^h \right) n_i =$$

$$= \binom{h}{o} a_h - \binom{h}{1} a_{h-1} \bar{x} + \binom{h}{2} a_{h-2} \bar{x}^2 - \ldots + (-1)^{h-1} h\bar{x}^h + (-1)^h \bar{x}^h$$

Hortik:

$$m_2 = a_2 - \bar{x}^2$$

$$m_3 = a_3 - 3a_2\bar{x} + 2\bar{x}^3$$

$$m_4 = a_4 - 4a_3\bar{x} + 6a_2\bar{x}^2 - 3\bar{x}^4$$

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

$$m_2 = \frac{\sum (x_1 - \bar{x})^2 \, n_i}{N} = a_2 - \bar{x}^2 = \frac{\sum x_i^2 \, n_i}{N} - \left(\frac{\sum x_i \, n_i}{N} \right)^2$$

Eta azkenez momentuak edozein punturekiko izan daitezke:

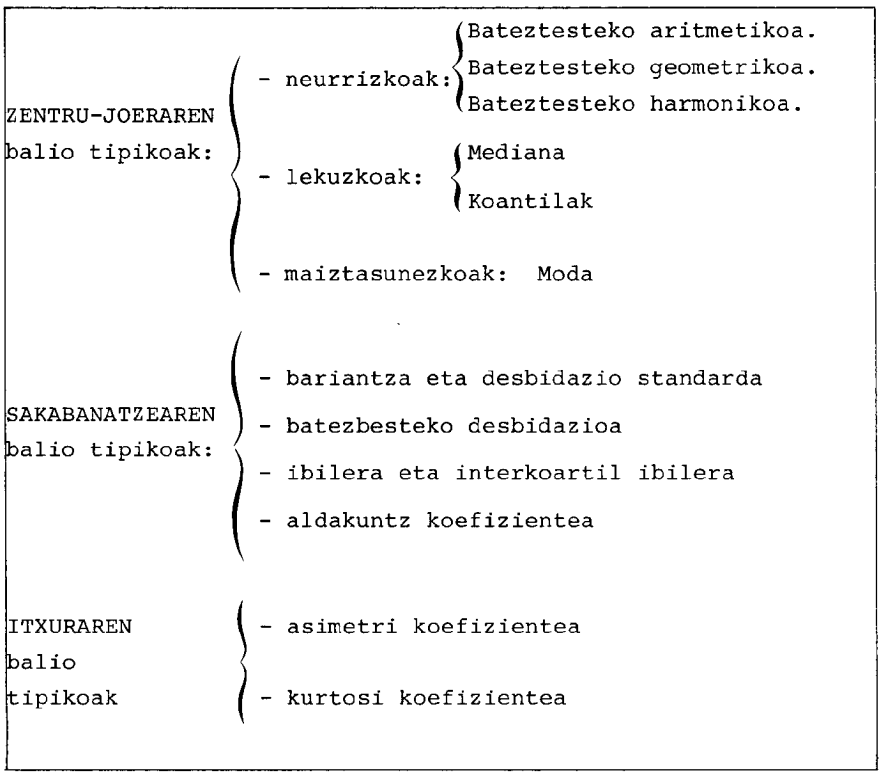
$$a_{h,x_o} = \frac{\sum^{N} (x_i - x_o)^h}{N} = \frac{\sum^{m} (x_i - x_o)^h \, n_i}{N}$$

$x_o = 0$ baldin bada, momentu arruntak izango dira,

eta $x_o = \bar{x}$ 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.



2.1. ZENTRU-JOERAREN BALIO TIPIKOAK

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

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 errealagoa eduki nahi baldin badugu.

2.1.1. BATEZBESTEKO ARITMETIKOA

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

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

Maiztasun erlatiboak, f_1 , f_2 , ... , f_m, kontutan hartuz

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

Ezaugarriak hartzen dituen balio guztiak desberdinak bali ra, maiztasun absolutuek 1 balioa hartuko lukete, hau da, - $n_1 = n_2 = ... = n_N = 1$ eta orduan:

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

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

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

x_i balio bakoitza $h.x_i$ balioaz ordezkatzen 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:

$$\bar{t} = \frac{\sum_{i=1}^{N} t_i}{N} = \frac{\sum_{i=1}^{N} (x_i + y_i + z_i)}{N} =$$

$$= \frac{\sum_{i=1}^{N} x_i}{N} + \frac{\sum_{i=1}^{N} y_i}{N} + \frac{\sum_{i=1}^{N} z_i}{N} = \bar{x} + \bar{y} + \bar{z}$$

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:

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

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

Batezbesteko ponderatua, zera izango da:

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

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.

Orduan $\bar{x}_{(pon)} = \dfrac{\sum x_i \, w_i}{\sum w_i}$

Oharra: kontutan hartu behar da, bien formulak baliokideak badira ere ponderazioak maiztasunak ez direla.

Beste adibide bat:

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

Hileroko gastuak / familia, 1.968. urtean

	Familiak	Gastuak (mila pezetatan)
Araba	24	4'4
Gipuzkoa	53	5'3
Nafarroa	57	5'1
Bizkaia	84	5'1

Batezbesteko aritmetikoa: $\bar{x} = \dfrac{19'9}{4} = 4'975$

Baina, familien kopurua kontutan hartuz:

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

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

2.1.2. BATEZBESTEKO GEOMETRIKOA:

Ondoko berdintasunaren bidez adierazten dugu

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

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.

2.1.3. BATEZBESTEKO HARMONIKOA

Ondoko berdintasunaren bidez adierazten dugu

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

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.

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

2.1.4. MODA

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

$x_1 , x_2 , \ldots x_i \ldots\ldots x_m$ edo A, B, ... , I, ... M

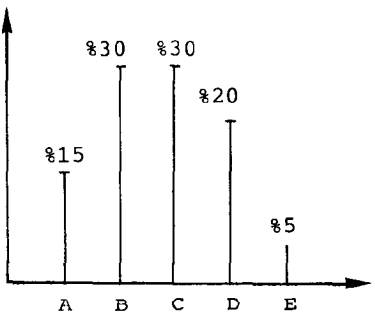
balio edo kategorien multzo bati dagozkien maiztasun erlati - boak hurrenez hurren $f_1, f_2, \ldots, f_m$ baldin badira, maiztasun erlatio gehien duen x_i balioari edo I kategoriari "moda" Mo deritzo.

Hots: Mo = x_i edo Mo = I n_i eta f_i maiztasun handienak - izanik.

Banaketa, modagabea izan liteke $f_1 = f_2 = \ldots = f_m$, edo moda bat baino gehiago ere eduki dezake, eta honen arabera, - "bimodala", "hirumodala" ... orokorki anitz-modala izango da.

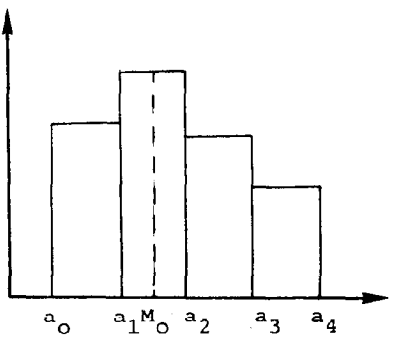
ADIBIDEAK

a)



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

b)



a_1 , a_2 klasea da kasu honetan moda, klase honen ordezkaria , Mo, har daiteke modatzat.

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

Klase edo tarte modala L_{i-1}, L_i izanez, balio modala finkatzeko bi formula desberdin proposatuko ditugu.

1) Klaseak zabaltasun berdinekoak baldin badira, ondoko formula erabil daiteke.

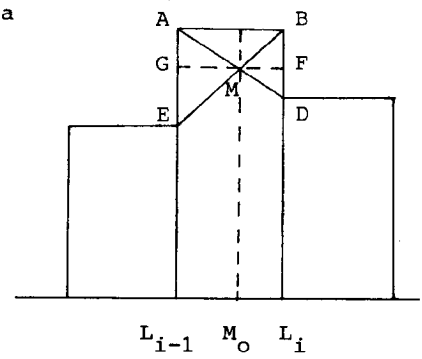
$$M_o = L_{i-1} + \frac{d_{i-1}}{d_{i-1} + d_{i+1}} \cdot C$$

non: C = klaseen zabaltasuna

$d_{i-1} = n_i - n_{i-1}$

$d_{i+1} = n_i - n_{i+1}$

Frogapen grafikoa :



AME eta BMD triangeluak antzekoak dira, hau da, beraien altuerak oinarriekiko proportzioanalak dira.

Hots:

$$\frac{MG}{MF} = \frac{AE}{BD}$$

Edo ondoko proportzio baliokidea.

$$\frac{MG}{MG + MF} = \frac{AE}{AE + BD} \text{ hau da, } \frac{MG}{C} = \frac{d_{i-1}}{d_{i-1} + d_{i+1}}$$

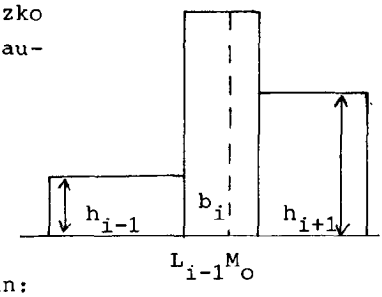
eta, $M_o = L_{i-1} + MG$ izatean, proposatu dugun formula frogatuta gelditzen da.

2) Klaseak zabaltasun desberdinekoak direnean, ondoko formula erabil daiteke:

$$M_o = L_{i-1} + \frac{h_{i+1}}{h_{i-1} + h_{i+1}} C_i$$

non: h_{i-1} = aurreko mailaren altuera
h_{i+1} = hurrengo mailaren altuera

Ondoko klaseen altuerekiko alderantzizko banaketa proportzioanal bat 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, $M_o = L_{i-1} + b_i$ izatean, proposatu dugun formula frogatuta gelditzen da.

Klaseak zabaltasun berdinekoak balira, hau da, $C_i = C$ C zabaltasunaz biderkatzen eta zatitzen badugu eta $n_i = C \times h_i$ dela kontutan hartuz, ondoko hau izango da formula:

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

2.1.5. MEDIANA

Medianak banaketaren balioak bi zati berdinetan zatitzen ditu.

Balio-multzoa ordenatuz gero, maiztasun absolutuen histo - grama area berdineko bi zatitan ebakitzen duen balioari, media na 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 indibiduoen kopurua bakoi tia baldin bada, aurreko baldintza bete egingo da, baina ez - beste edozein kasutan.

Banaketa diskretua eta N bikoiti baten aurrean bagaude.

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

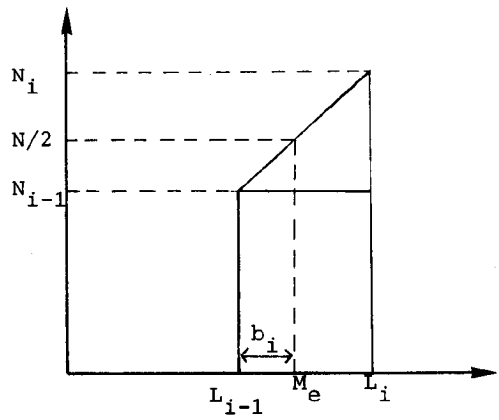
Hots: $M_e = \dfrac{\text{erdiko balioak}}{2}$

Banaketa taldekatuetan eta jarraietan: $\frac{N}{2}$ balioa duen - tartea finkatuz gero eta tartean n_i balioak linealki gehitzen direla suposatuz kalkulatuko 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:



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

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

Oharra: maiztasun eraltiboak erabiliz, era berdinean kalkula daiteke M_e ; kasu horretan, $\frac{N}{2} = \frac{1}{2}$.

KOANTILAK

Koantilak bestelako lekuzko neurriak dira. Mota desberdinetakoak dira: koartilak, dezilak eta zentilak . (¡)

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

(¡)Medianaren antzera definitzen dira, baina bi parte hartu ordez lau (koartilak) hamar (dezilak) edo ehun (zentilak) parte hartuz.

Dezilak d_1 .... d_9 izango dira eta zentilak, C_1 ... C_{99} .

Argiro ikusten da:

$q_2 = d_5 = C_{50} = M_e$ dela.

OHARRAK:

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

Ondoko erlazio hau betetzen da:

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

Eta gutxi gorabehera, hiruretako edozein besteen bidez kalkula daiteke.

2.2. SAKABANATZEAREN BALIO TIPIKOAK

Datu-multzoaren ideia orokor ematen ziguten zentru-joeraren balio tipikoak. Ideia hau osatzeko zentru-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.

2.2.1. BARIANTZA

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

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

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

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

Orduan, sakabanatze-neurri bezala bariantzaren erro karratua hartzea egokiagoa da, desbidazio tipikoa edo standarda, hain zuzen.

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

eta ezaugarria neurtuta dagoen unitateetan, sakabanatze - neur - keta daukagu.

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

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

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

Bariantzak garrantzizko propietate bat betetzen du, biga - rren ordenako momentu zentralik txikiena dela, hain zuzen.

Hots:

$$a_2 , x_o = \frac{1}{N} \sum_{i=1}^{m} (x_i - \bar{x}_o)^2 n_i =$$

$$= \frac{1}{N} \sum_{i=1}^{m} \left((x_i - \bar{x}) + (\bar{x} - x_o) \right)^2 n_i =$$

$$= \frac{1}{N} \sum_{i=1}^{m} \left((x_i - \bar{x})^2 + 2 (x_i - \bar{x}) (\bar{x} - x_o) + (\bar{x} - x_o)^2 \right) n_i =$$

$$= \frac{1}{N} \sum_{i=1}^{m} (x_i - \bar{x})^2 n_i + 2 (\bar{x} - x_o) \sum_{i=1}^{m} (x_i - \bar{x}) \frac{n_i}{N} +$$

$$+ \frac{1}{N} (\bar{x} - x_o)^2 \sum_{i=1}^{m} n_i = S_x^2 + (\bar{x} - x_o)^2$$

Eta argiro dagoenez, $x_o = \bar{x}$ baldin bada, txikiena izango da.

2.2.2. BATEZBESTEKO DESBIDAZIOA

Balio absolutuak harturik (konpensatziorik izan ez dadin), batezbesteko batekiko balioek duten desbidazioen batezbesteko aritmetikoari, "batezbesteko desbidazioa" deritzo.

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

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

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

Neurri honek indibiduoen 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:

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

Batezbestekoa mediana denean, batezbesteko desbidazioarik txikiena daukagu.

2.2.3. IBILERA

Ezaugarri zenbakizko batek hartzen dituen balio maximo eta minimoaren arteko diferentziari ibilera deritzogu.

Hots:

$$x_o \leqslant x_1 \leqslant \cdots\cdots \leqslant x_m \quad \text{izanik}$$

$$R = x_m - x_o$$

Interkoartil ibilerak ere kontsidera daitezke, adibidez, hirugarren eta lehenengoen artekoena, edota, interdezil ibilerak e.a.

2.3. ITXURAREN BALIO TIPIKOAK

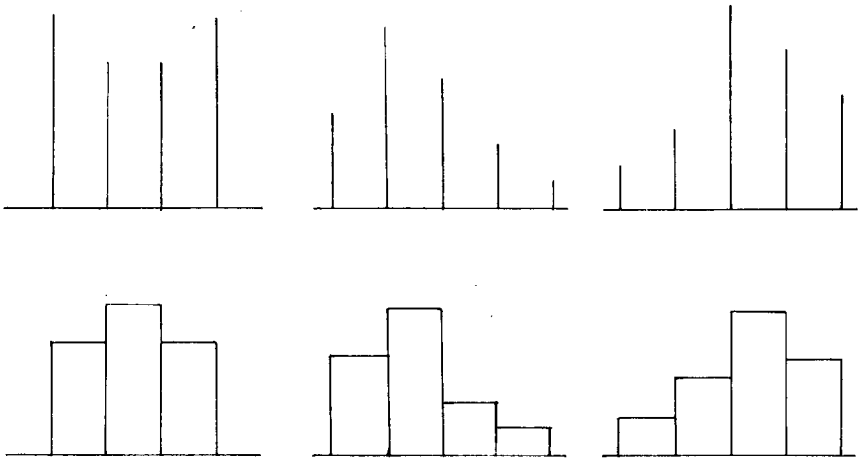
Aztertzen ari garen laginaren ezaugarria eta berari dagokion datu-multzoa , zentru-joeraren eta sakabanatze-balioen bi-

dez nahiko argi geratu bada ere, gehiago osa daiteke itxura - ren balio tipikoak aztertzean; hauen artean asimetri koefi - zientea eta kurtosi koefizientea ikusiko ditugu.

2.3.1. ASIMETRI KOEFIZIENTEA

Maiztasun-banaketaren grafikoa simetrikoa denean, banake ta simetrikoa dela esango dugu, hau da, barrazko diagrama edo histograma egitean simetrikoa bada, banaketaren elementuak ba tezbestekoarekiko bi aldeetan berdin kokatzen direla esan nahi du.

Aurrekoa betetzen ez bada, banaketa asimetrikoa izan - go da.



Simetrikoak

Eskuinerantz asimetrikoak

Ezkererantz asimetrikoak

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

Hots: $\bar{x} = M_o = M_e$

Asimetri koefizientea g_1 ikurraz adieraziko dugu sinboliko, non:

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

$m_3 = \sum_{i=1}^{m} (x_i - \bar{x})^3 \frac{n_i}{N}$ izanik, maiztasunen balioak $\bar{x}$ - arekiko simetrikoki kokatuta badaude, $m_3 = 0$ izango da; baina maiztasunen balioak eskuinaldean handiagoak badira, hau da, - $x_i > \bar{x}$ denean, orduan $m_3 > 0$ izando da, eta, alderantziz,ezkerraldean 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

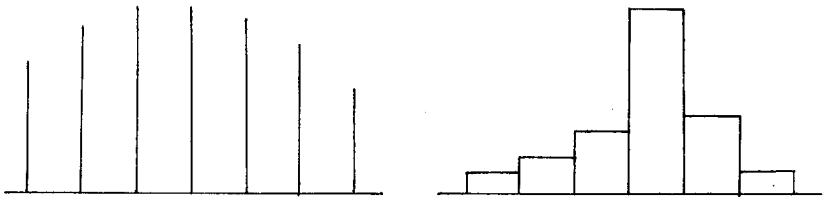
$$g_1 = \frac{m_3}{S^3}$$

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

2.3.2. KURTOSI 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 zukeen lerroa diskretuaren kasuan kasu hauek dauzkagu:



zapaldua puntaduna

Batezbesteko aritmetikoa eta bariantza berdina daukaten - bi banaketa aztertuz, itxuraren aldetik bat oso zapaldua da eta oso puntaduna bestea.

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

Orduan, m_4 zapalduraren adierazle bezala har daiteke,baina neurri-unitateak ber lau izanez, zuzenkiago.

g_2 ikurraz adierazten den kurtosi koefizientea, zapalduraren adierazle bezala erabiliko dugu.

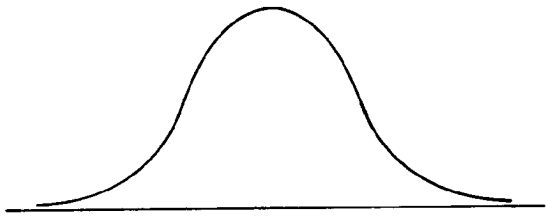
$$g_2 = \frac{m_4}{S_x^4} = \frac{\sum_{i=1}^{m} (x_i - \bar{x})^4 \; n_i}{N \; S_x^4}$$

"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 puntaduna izango da eta 3 baino txikiagoa baldin bada, zapaldua 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 probabilititate-ere dua baita.



banaketa Normala : $g_2 = 3$

2.4. KONTZENTRAZIO - NEURKETAK

Batezbesteko aritmetikoa agitean zenbakitzailean atera - tzen zaigun kopurua, abiapuntu estatistiko batetatik ikusita, batzutan ez du zentru argirik eta interesgarririk. Adibidez , pertsona batzuren altuera-banaketa aztertzean, zenbakitzaile hori altueren batuketa izango da; baina, beste kasu batzutan, bereziki ezaugarri sozio-ekonomikoetan, aztergarria izaten da adibidez, alokairu-banaketa aztertzean, alokairu guztien kopurua edo "alokairu-masa" da.

Kontzentrazio - Neurketak zahazki "masa" baten uniformetasuna neurtzea dute helburutzat.

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

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

Banaketak uniformetasun absoluturantz jotzen duenean, - maiztasun-banaketaren batezbesteko aritmetikoa guztiz adieraz garria da eta, alderantziz gertatuko da, kontzentrazio maxi - moa denean.

2.4.1. LORENZ-en KURBA

Grafikoki banaketen kontzentrazioa ikusarazten digun neu rrian, interesgarria zaigu.

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

Alokairuak: x_i	Langile-ko purua: 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

$N = 50$ $\qquad M = \sum_{i=1}^{k} x_i\, n_i = 80$

azken zutabeak honela kalkulatzen dira:

N_i : (n_i) zutabea metatuz

hots: $n_1 + n_2 + ... + n_i = N_i$

M_i : ($x_i n_i$) zutabea metatuz

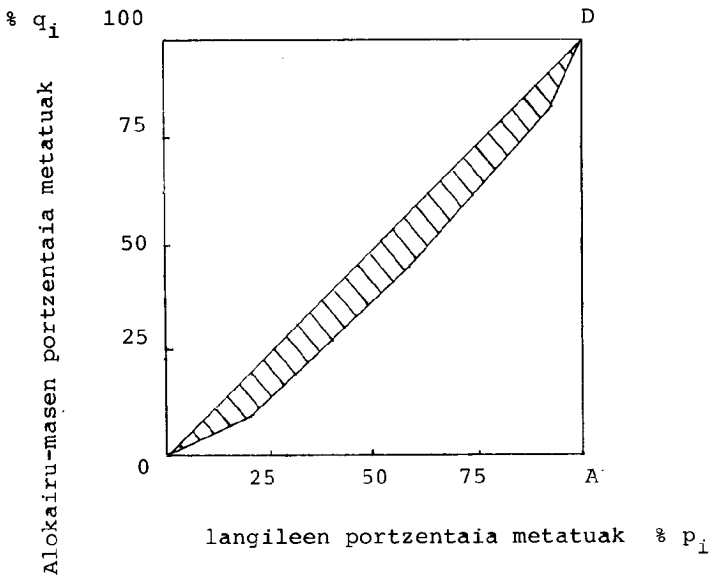
hots: $M_i = x_1 n_1 + x_2 n_2 + ... + x_i n_i$

$p_i = \frac{N_i}{N} 100$ langileen portzentaia metatua

$q_i = \frac{M_i}{M} 100$ alokairu-masaren portzentaia metatua

$\{(p_i , q_i)\}_{i \in I}$ bikote guztiak sistema cartesiar batetan grafikoki adieraziz , eta $(p_1 q_1)$, (0,0)- jatorriarekin eta beste bikote guztiak elkarren artean zuzenen bitartez lotuz LORENZ-en kurba lortzen dugu.

Lehenengo 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 diagonala erabiliko dugu uniformetasunaren erreferentzia moduan.

Zeren eta $(p_i q_i)$ multzoko punto 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 % 25 langileek alokairu-masaren % 25-a inoiz lortu).

2.4.2. GINI-ren indizea

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

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:

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

non $0 \leqslant I_G \leqslant 1$

Kontzentrazioa maximoa bada, orduan OAD azalera = marraztutako azalera, eta $I_G = 1$.

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

Kalkulua: I_G GINI-ren indizea kalkulatzean, azalera horiek zuzenki neurtuz kalkula litezke; bainan beti era honetan erraza izaten ez denez gero, gutxi gorabehera ondoko formula erabiliz kalkulatzen da I_G:

$$I_G = \frac{\sum_{i=1}^{k-1} (p_i - q_i)}{\sum_{i=1}^{k-1} p_i}$$

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

Gutxi gorabeherako formula honetaz ere egiazta daiteke, $0 \leqslant I_G \leqslant 1$ 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 sozio-ekonomikoetarako erabilgarriak dira. Adibide bezala hauek aipatuko ditugu: errenta pertsonalaren banaketa, nekazal - propietatearen banaketa, sektore bateko enpresa-produkzioaren banaketa eta abar.

A R I K E T A K

1) Ondoko taulan Bizkaiatik Karlos II.aren gudetara joaten ziren gizasemeak ikusten ditugu urtez urte, bere 35 urtetako agite-garaian.

200	300	100	100	0
200	0	0	200	0
0	400	0	0	400
0	1.000	300	0	0
0	0	0	0	100
0	300	0	200	200
0	0	100	300	0

Eskatzen da:

1) Maiztasun-banaketaren adierazpide grafikoa.
2) Zentru-joeraren balio ezagunenak.
3) Sakabanatze-balio ezagunenak.

2) Urbanizazio baten zentsotik balio hauek hartzen ditugu.

10 urte baino gutxiagokoak: 500 emakumezko eta 500 gizonezko
10 urtetik 30 urterartekoak:700 emakumezko eta 800 gizonezko
30 urtetik 50 urterartekoak:500 emakumezko eta 500 gizonezko
50 urtetik 80 urterartekoak:300 emakumezko eta 200 gizonezko

Eskatzen da:

1) Histogramak edo populazio-piramidea.
2) Batezbesteko aritmetikoa eta desbidazioa tipikoa gizonezko, emakumezko eta populazio totalarekiko, hurrenez hurren.

3) Erostetxe edo magazin batetako erosleriaren gastuak aztertu ondoren, honela izan ziren banatuak gastuak mila pezetatan:

Gastu-tarteak	Maiztasunak
0- 4'99	5
5- 9'99	10
10- 14'99	14
15- 19'99	20
20- 24'99	35
25- 29'99	15
30- 34'99	1

Eskatzen da:

Kalkula itzazu neurri hauek, beren esanahia adieraziz: batezbestekoa, desbidazio tipikoa, aldakuntz koefizientea, moda.

4) Jaiotzean normalak izan diren 20 haurren pisua neurtu da eta emaitzak hauek dira.

Pisua (kilotan)	Haurren Kopurua
2'15 - 2'55	1
2'55 - 2'95	3
2'95 - 3'35	9
3'35 - 3'75	5
3'75 - 4'15	1
4'15 - 4'55	1

Eskatzen da: 1) Banaketaren adierazpide grafikoa.

2) Batezbestekoa, moda, mediana, eta asimetri - koefizientea.

5) Populazio batetan 20 pertsonatako lagin bat hartu da. Altuerak neurtu dira eta ondoko emaitzak lortu ditugu:

1'65 - 1'70 - 1'68 - 1'82 - 1'80 - 1'45 - 1'56 - 1'35 - 1'90 - 1'56 - 1'88 - 1'66 - 1'65 - 1'55 - 1'57 - 1'60 - 1'62 - 1'90 - 1'82 - 1'65 .

Kalkula itzazu: zentru-joeraren balio tipikoak (batezbestekoa, moda,mediana) eta sakabanatzearenak (Bariantza, debidazio tipikoa eta ibilera).

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

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

Histogramen, zentru-joeraren eta sakabanatzearen balio tipikoen bidez, maiztasun-banaketa hauek gonbara itzazu.

7) Ondoko taulan herrialde baten udalen banaketa ematen dugu - beraien biztanleen arabera.

Biztanleak : X	Udal-kopurua: n
$X \leqslant 10.000$	90
$10.000 \leqslant X \leqslant 50.000$	600
$50.000 \leqslant X \leqslant 150.000$	300
$150.000 \leqslant X \leqslant 450.000$	10

Eskatzen da:

1) Banaketa grafikoki adieraztea.

2) $\overline{X}$, M_e , M_o, batezbesteko desbidazioa, desbidazio tipikoa eta g_o aldakuntz koefizientea.

8) Ondoan hileroko alokairu-banaketa ematen da, mila pezetatan neurtua.

x_i	n_i
9	9
15	6
12	10
20	5

Kalkula itzazu: Lorenz-en kurba eta GINI-ren indizea.

9) Zure eritzia emanez, ondoko alokairu-banaketari (mila pezetatan) dagokion Lorenz-en kurba eta GINI-ren indizea eskatzen dira.

10	Lanbide - ikasle	20
25	Berezilari edo espezialista.....	50
5	Maisu	100
10	Ofizial	70

10) Goi-mailako ikastetxe batetan, langileak honela banatzen dira. 15 katedratiko, 4 irakasle agregatu, 45 adjuntu, 25 administrazio laguntzaile eta 5 menpeko. Hileroko alokairuak mila pezetatan hauek dira.

Lanbidea	Alokairua
Adjuntuak	65
Menpekoak	25
Agregatuak	85
Katedratikoak	105
Adm. laguntzaileak	32

Kalkula itzazu, ezagutzen dituzun kontzentrazio-neurketak, lortutako emaitzari buruz zure eritzia emanez.

11) I.M.N.E.B. - ren ordezkaritza probintzialeko enplegatuen - hileroko soldatak honela banatzen dira.

40	enplegatu 50.000	pezeta bakoitzak.
100	enplegatu 100.000	pezeta bakoitzak.
40	enplegatu 200.000	pezeta bakoitzak.
20	enplegatu 500.000	pezeta bakoitzak.

Eskatzen da:

1) Barrazko diagrama, enplegatuen banaketa soldaten arabera grafikoki adierazteko.
2) Banaketa honen g_o, aldakuntz koefizientea hain zuzen.
3) Lorenz-en kurba, soldaten banaketa enplegatuen artean aurkeztеko.
4) GINI-ren indizea, kontzentrazio- maila neurtzeko.

III. ZENBAKI-INDIZEAK.

III.1. INDIZE SINPLEAK

III.2. PONDERAZIO GABEKO INDIZE KONPLEXUAK

2.1. BATEZBESTEKO ARITMETIKOA SINPLEAREN METODOA
2.2. BATEZBESTEKO GEHIKORTU SINPLEAREN METODOA

III.3. INDIZE KONPLEXU PONDERATUAK

3.1. BALIOEN, PREZIOEN ETA KOPURUEN INDIZEAK
3.1.1. LASPEYRES -EN INDIZEAK
3.1.2. PAASCHE-REN INDIZEAK
3.1.3. FISHER-EN INDIZEAK
3.1.4. PROPIETATE ETA ERLAZIO BATZU
3.1.5. KALKULAKETA

III.4. INDIZE KONPLEXUEN ERAIKETAN SORTZEN DIREN ZENBAIT ERAGOZPEN

4.1. ALDAGAIEN HAUTAPENA.
4.2. SOMATUTAKO LEKU ETA DENBORAREN HAUTAPENA
4.3. TALDE ETA AZPITALDEEN HAUTAPENA
4.4. OINARRI-DENBORAREN HAUTAPENA
4.5. FORMULA ETA DEMBORAREN HAUTAPENA
4.6. INDIZEAREN ESANGURA ETA ZABALDURA

III.5. ERAGOZPEN BEREZI BATZU

5.1. OINARRI-ALDAKETA INDIZE SIMPLEETAN
5.2. BERRIZTAPEN ETA LOTURAK INDIZE KONPLEXUETAN

III.6. ZENBAKI-INDIZEEN APLIKAPENAK

6.1. KONTSUMO-PREZIOEN INDIZEAK

ARIKETAK

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 hauenbidez, neurtzeko unitatea desagertzen da eta aldagai baten gorabeherakadak autonomikoki azter daitezke. Honela indizeak aldagai segiden gorabeherakadak erraztatzen di tu (originalki unitate desberdinetakoak).

Bedi x_0 , x_1 .... 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 kalkula - tzen 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 uztabilketa bat ona edo txarra dela esatean, guretzat uztabilketa "normalak" direnekin gonbaraturik egiten dugu.

III.2. PONDERAZIO GABEKO INDIZE KONPLEXUAK

n- Balioren 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, ba, ponderazio gabeko indize konplexuak kalkulatzeko bi metodo ikustera.

2.1. BATEZBESTEKO ARITMETIKOA SINPLEAREN METODOA

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

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

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

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

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

2.2. BATEZBESTEKO GEHIKORTU 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 daiteke erabil, aldagaiak unitate desberdinetan neurtuak direnean, ez baitu zentzu askorik unitate desberdinetan neurtuak izan diren balioak batzea.

Ondoko taulan ponderatu gabeko indize konplexu hauen formulak dauzkagu hedaturik.

x / t	$x_1, x_2, \ldots x_i, \ldots x_n$	I_i (Indize sinpleak)	S_t	$\sum_i x_{it}$ B_t
0	$x_{10} x_{20} \ldots x_{io} \ldots x_{no}$	$I_{10} I_{20} \ldots I_{no}$	$S_o = \frac{\sum_i I_{io}}{n}$	$\sum_i x_{io}$ $B_o = \frac{\sum_i x_{io}}{\sum_i x_{io}} 100$
1	$x_{11} x_{21} \ldots x_{il} \ldots x_{nl}$	$I_{11} I_{21} \ldots I_{nl}$	$S_1 = \frac{\sum_i I_{il}}{n}$	$\sum_i x_{il}$ $B_1 = \frac{\sum_i x_{il}}{\sum_i x_{io}} 100$
.				. .
.				. .
.				. .
t	$x_{1t} x_{2t} \ldots x_{it} \ldots x_{nt}$	$I_{1t} I_{2t} \ldots I_{nt}$	$S_t = \frac{\sum_i I_{it}}{n}$	$\sum_i x_{it}$ $B_t = \frac{\sum_i x_{it}}{\sum_i x_{io}} 100$
.				. .
.				. .
m	$x_{1m} x_{2m} \ldots x_{im} \ldots x_{nm}$	$I_{1m} I_{2m} \ldots I_{nm}$	S_m	B_m

Adibidea: Hego Esuskal Herriko populazioaren datu hauen - bidez, populazio-igoerak ikusiko ditugu herrial de 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

Taula honi 1. Grafikoa dagokio.

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

Taula honi 2. Grafikoa dagokio.

Azkenik, hamarkada bakoitzeko urtez urteko hazkunde-tasa bilatu eta 3. grafikoan ikusarazten dizuegu 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 o uneko unitate bat, $(1 + \alpha)^{10}$ bihurtu da 10 urtetan, eta exenplu honetan datuak hamarkadaz di tugunez, urteko hazkunde-tasa ondoko formulen bidez kalkulatzen 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, urteko haztunde-tasak adierazten dira:

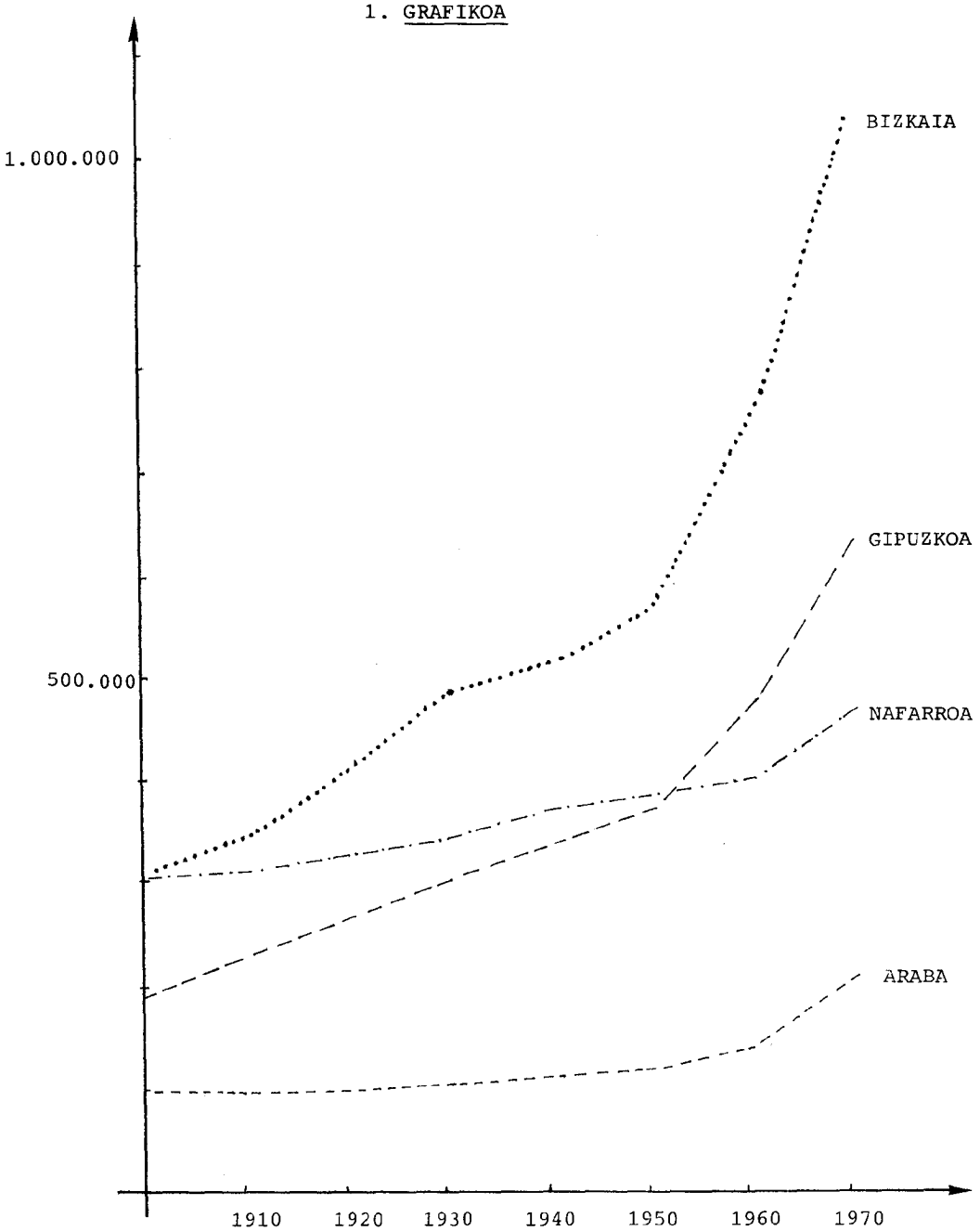
(3)

α x100	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

Taula honi 3. Grafikoa dagokio.

Lortu ditugun emaitza hauek, datozen era desberdinetako hi ru grafikoetan adierazi ditugu.

1. GRAFIKOA


1.000.000
500.000
BIZKAIA
GIPUZKOA
NAFARROA
ARABA
1910
1920
1930
1940
1950
1960
1970

1. GRAFIKOAN IKUSTEN DIREN GAUZA NABARMENENAK

(1) Taulari dagokio.

Grafiko honek, herrialde bakoitzeko populazioaren hazkunde absolutua adierazten digu.

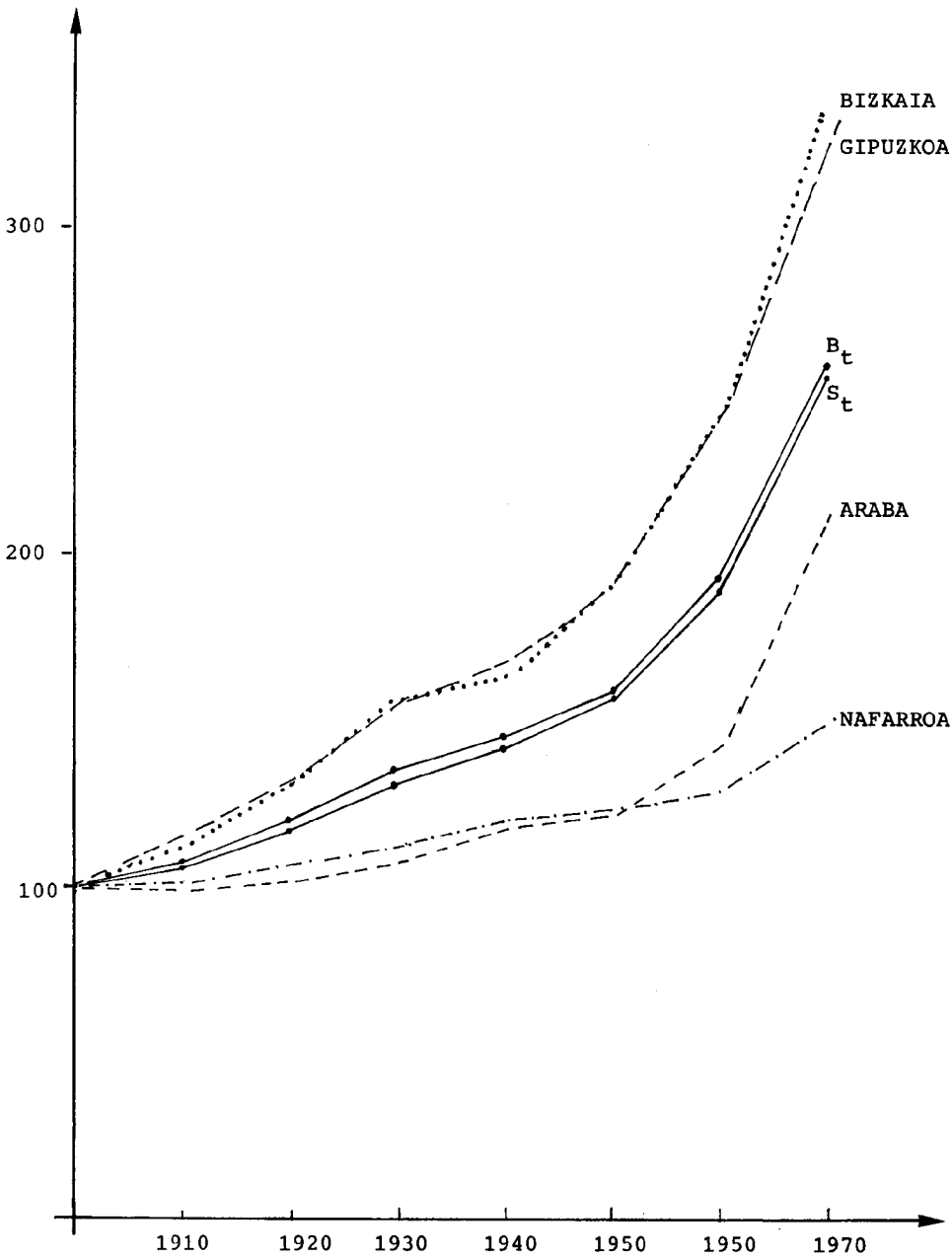
Ondorio hauek atera genitzake:

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 araberakoa da, (non gehiketa hamarkadaren hasierako populazio absolutuarekikoa da).

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

2. GRAFIKOA


BIZKAIA
GIPUZKOA
300
B_t
S_t
ARABA
200
NAFARROA
100
1910
1920
1930
1940
1950
1950
1970

2. GRAFIKOAN IKUSTEN DIREN GAUZA NABARMENENAK

(2) Taulari dagokio.

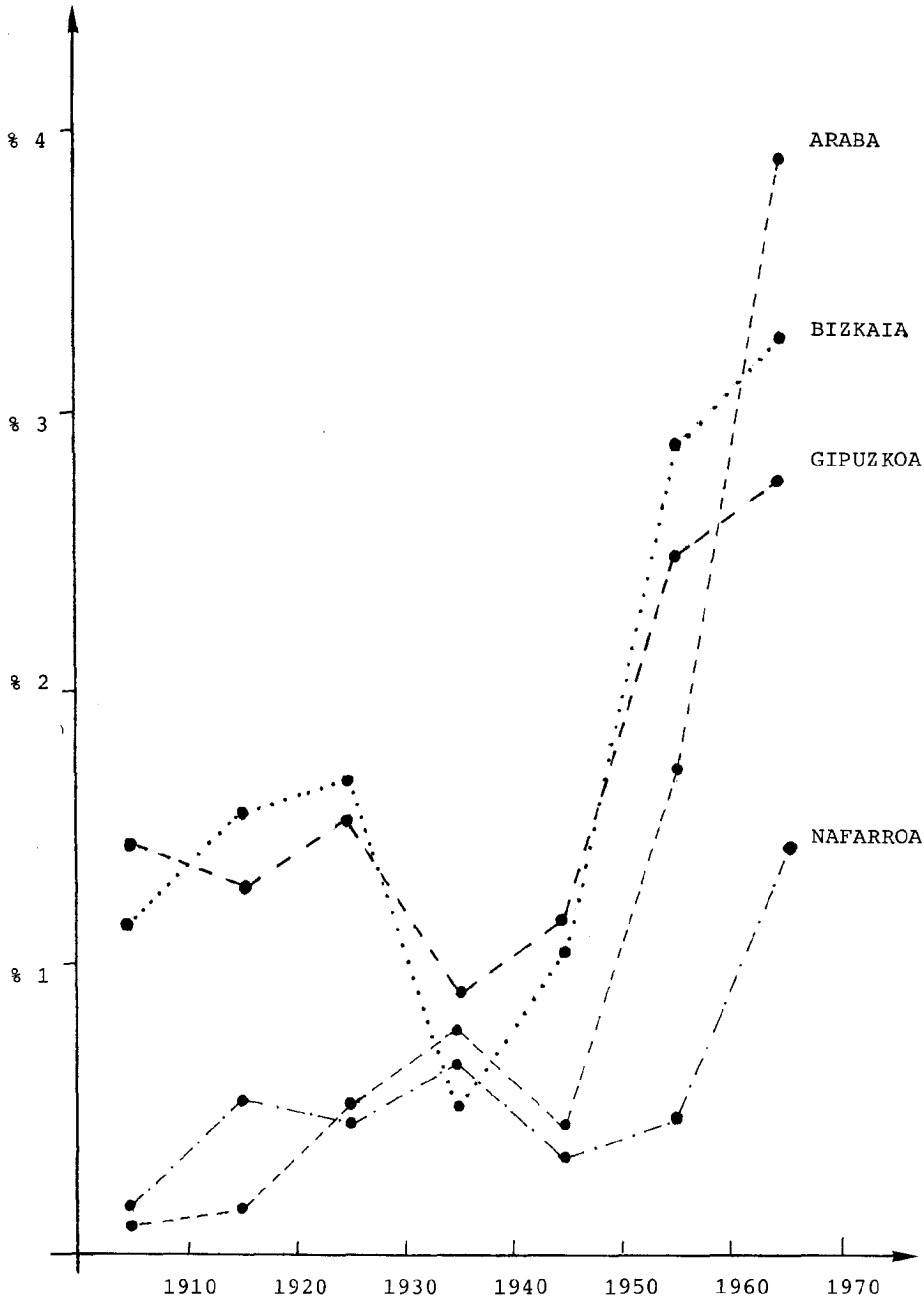
Alde batetik grafiko honetan, herrialde bakoitzari dagokion populazio-indize sinpleen ibilbidea daukagu.

Indize sinple horien ibilbideak, populazioaren gehiketa erlatiboa, 1.900 urteko populazioarekiko, adierazten digute. Herrialde guztietako populazioa 1.900 urtean, 100 dela suposaturik.

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).

3. GRAFIKOA


% 4
% 3
% 2
% 1
ARABA
BIZKAIA
GIPUZKOA
NAFARROA
1910
1920
1930
1940
1950
1960
1970

3. GRAFIKOAN IKUSTEN DIREN GAUZA NABARMENENAK

(3) Taulari dagokio.

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

Hazkunde-tasan ibilbideek populazioaren gehiketa erlatiboa adierazten digute, (kasu honetan gehiketa hamarkada bakoitzarekikoa delarik).

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

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

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

Azken Hamarkadetan izugarrizko azelerazioa nabaritzen da, lau herrialdeetan, kezkagarria 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, pauta batzu jarraituz egingo dugu, eta artikulu 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 familili bati dagokiona baldin bada, (senitarte-kopuru maugatu eta errenta-tarte baten barnekoa), orduan, ogiari dagokion ponderazioa ere erizpide berdinetan oinarritua izan behar du.

3.1. BALIOEN PREZIOEN ETA KOPURUEN INDIZEAK

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

$$b = p \cdot k$$

0,1,...t,...,m denboretan, ondasun baten prezioak $p_o, p_1, \ldots p_t, \ldots p_m$ eta dagozkien kopuruak $k_o, k_1, \ldots k_t, \ldots k_m$ baldin badira, orduan "balio-segida denborala" hau izando da:

b_t

$b_o = p_o\ k_o$

$b_1 = p_1\ k_1$

.
.
.

$b_t = p_t\ k_t$

.
.
.

$b_m = p_m\ k_m$

Baina denbora iragatean k=kte bada, b_t balio-segida denboralak, prezioaren gorabeherakadak adieraziko ditu bakarrik. Berdin, denbora iragatean p=kte 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)
$p_o\ k_o$	$p\ k_o$	$p_o\ k$
$p_1\ k_1$	$p\ k_1$	$p_1\ k$
.	.	.
.	.	.
.	.	.
$p_t\ k_t$	$p\ k_t$	$p_t\ k$
.	.	.
.	.	.
.	.	.
$p_m\ k_m$	$p\ k_m$	$p_m\ k$

"O" ondasun bati dagozkion (balio, prezio eta kopuruen) segida horiek (¡) balio errealetan adieraziak direnez gero, beste "O_i" ondasun batzuri dagozkienak ere batu ditzakegula, argi i - kusten da.

Suposa dezagun O_1 O_2 ..... O_i ....O_n n- ondasun desberdin ditugula, (non O_i ondasun bakoitzaren balio, prezio eta kopuruen balio-segida denboralak ezagunak zaizkigun), n-ondasun - horiek, batera osatzen duten "konplexuaren indizeak" edo "indize konplexuak" kalkulatzerakoan, balio-segida desberdinen batuketak, laburki, honela adieraz litezke:

$\sum_{i=1}^{n} b_{it}$	$\sum_{i=1}^{n} b_{it}(k)$	$\sum_{i=1}^{n} b_{it}(p)$
$\sum_i p_{io} k_{io}$	$\sum_i p_i k_{io}$	$\sum_i p_{io} k_i$
$\sum_i p_{il} k_{il}$	$\sum_i p_i k_{il}$	$\sum_i p_{il} k_i$
:	:	:
$\sum_i p_{it} k_{it}$	$\sum_i p_i k_{it}$	$\sum_i p_{it} k_i$
.	.	.
.	.	.
.	.	.
$\sum_i p_{im} k_{im}$	$\sum_i p_i k_{im}$	$\sum_i p_{im} k_i$

(Bi azkenetako segidetan, p_i = kte, $\forall i$ eta k_i = kte, $\forall i$ dira).

(¡) Lehenengoa prezio eta kopuruaren funtzioa da, bigarrena, kopuruaren funtzioa eta hirugarrena prezioarena.

Beraz, "indize konplexuak"lortzeko: konplexuari dagozkion (balio, prezio eta kopuru) hiru balio segida denboral horien indiza sinpleak kalkulatu behar ditugu.

Hots oinarri-denbora t =0 bada:

I_t^b	I_t^k	I_t^p
$\frac{\sum_i p_{io}k_{io}}{\sum_i p_{io}k_{io}} 100$	$\frac{\sum_i (p_i)k_{io}}{\sum_i (p_i)k_{io}} 100$	$\frac{\sum_i p_{io}(k_i)}{\sum_i p_{io}(k_i)} 100$
$\frac{\sum_i p_{il}k_{il}}{\sum_i p_{io}k_{io}} 100$	$\frac{\sum_i (p_i)k_{il}}{\sum_i (p_i)k_{io}} 100$	$\frac{\sum_i p_{il}(k_i)}{\sum_i p_{io}(k_i)} 100$
⋮	⋮	⋮
$\frac{\sum_i p_{it}k_{it}}{\sum_i p_{io}k_{io}} 100$	$\frac{\sum_i (p_i)k_{it}}{\sum_i (p_i)k_{io}} 100$	$\frac{\sum_i p_{it}(k_i)}{\sum_i p_{io}(k_i)} 100$
⋮	⋮	⋮
$\frac{\sum_i p_{im}k_{im}}{\sum_i p_{io}k_{io}} 100$	$\frac{\sum_i (p_i)k_{im}}{\sum_i (p_i)k_{io}} 100$	$\frac{\sum_i p_{im}(k_i)}{\sum_i p_{io}(k_i)} 100$

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

Bainan 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 hainbeste ebazpide edo soluzio aurki daitezke, bainan praktikan bi inposatu dira: LASPEYRES-ena eta PAASCHE-rena.

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

3.1.1. LASPEYRES - en indizeak

I_t^p eta I_t^k, prezio eta kopuruen indize konplexuak kalkulatzeko, O_i ondasun bakoitzaren kasuan konstantetzat hartzen ditugun kopuru eta prezioak, oinarri-denboran sortutakoak dira. Hots, $k_i = k_{io}$ eta $p_i = p_{io}$.

Honela lortutako indize konplexuei, Laspeyres-en prezio eta kopuruen indize konplexuak deritzegu eta sinbolikoki L_t^p , L_t^k adieraziko ditugu.

Hots:

$$L_t^p \qquad\qquad L_t^k$$

$$\frac{\sum_i p_{io}(k_{io})}{\sum_i p_{io}k_{io}}100 \qquad \frac{\sum_i (p_{io})\, k_{io}}{\sum_i p_{io}k_{io}}100$$

$$\frac{\sum_i p_{il}(k_{io})}{\sum_i p_{io}k_{io}}100 \qquad \frac{\sum_i (p_{io})\, k_{il}}{\sum_i p_{io}k_{io}}100$$

$$\vdots \qquad\qquad \vdots$$

$$\frac{\sum_i p_{it}\,(k_{io})}{\sum_i p_{io}\,k_{io}}\,100 \qquad \frac{\sum_i (p_{io})k_{it}}{\sum_i p_{io}\,k_{io}}\,100 \quad (1)$$

$$\vdots \qquad\qquad \vdots$$

$$\frac{\sum_i p_{im}\,(k_{io})}{\sum_i p_{io}\,k_{io}}\,100 \qquad \frac{\sum_i (p_{io})\,k_{im}}{\sum_i p_{io}\,k_{io}}\,100$$

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

$$\frac{L_t^p}{100} \qquad \frac{L_t^k}{100}$$

$$\frac{\sum_i \frac{p_{il}}{p_{io}}\,(p_{io}k_{io})}{\sum_i p_{io}\,k_{io}}\,100 \qquad \frac{\sum_i \frac{k_{il}}{k_{io}}(p_{io}k_{io})}{\sum_i p_{io}\,k_{io}}\,100$$

$$\vdots \qquad\qquad \vdots$$

$$\frac{\sum_i \frac{p_{it}}{p_{io}}\,(p_{io}k_{io})}{\sum_i p_{io}\,k_{io}}\,100 \qquad \frac{\sum_i \frac{k_{il}}{k_{io}}\,(p_{io}k_{io})}{\sum_i p_{io}\,k_{io}}\,100 \quad (2)$$

$$\vdots \qquad\qquad \vdots$$

$$\frac{\sum_i \frac{p_{it}}{p_{io}}\,(p_{io}k_{io})}{\sum_i p_{io}\,k_{io}}\,100 \qquad \frac{\sum_i \frac{k_{im}}{k_{io}}\,(p_{io}k_{io})}{\sum_i p_{io}\,k_{io}}\,100$$

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

Era honetara, (2) formulak oso erabilgarriak zaizkigu : O_i ondasun bakoitzaren indize sinpleak erabiliz Laspeyres-en prezio edo kopuruen indize konplexuak kalkulatzeko.

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 deborako, 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:

$$\frac{p_t^p}{100} \qquad \frac{p_t^k}{100}$$

$$\begin{matrix}
\dfrac{\sum_i p_{il}(k_{il})}{\sum_i p_{io}(k_{il})}100 & \dfrac{\sum_i (p_{il})\, k_{il}}{\sum_i (p_{il})\, k_{io}}100 \\
\vdots & \vdots \\
\dfrac{\sum_i p_{it}(k_{it})}{\sum_i p_{io}(k_{it})}100 & \dfrac{\sum_i (p_{it})\, k_{it}}{\sum_i (p_{it})\, k_{io}}100 \\
\vdots & \vdots \\
\dfrac{\sum_i p_{im}(k_{im})}{\sum_i p_{io}(k_{im})}100 & \dfrac{\sum_i (p_{im})\, k_{im}}{\sum_i (p_{im})\, k_{io}}100
\end{matrix} \qquad (1)$$

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

$$\frac{p_t^p}{100}\qquad\qquad \frac{p_t^k}{100}$$

$$\frac{\sum_i \frac{p_{il}}{p_{io}}(p_{io}k_{il})}{\sum_i p_{io}k_{il}}100 \qquad \frac{\sum_i \frac{k_{il}}{k_{io}}(p_{il}k_{io})}{\sum_i p_{il}k_{io}}100$$

$$\vdots \qquad\qquad \vdots$$

$$\frac{\sum_i \frac{p_{it}}{p_{io}}(p_{io}k_{it})}{\sum_i p_{io}k_{it}}100 \qquad \frac{\sum_i \frac{k_{it}}{k_{io}}(p_{it}k_{io})}{\sum_i p_{it}k_{io}} \qquad (2)$$

$$\vdots \qquad\qquad \vdots$$

$$\frac{\sum_i \frac{p_{im}}{p_{io}}(p_{io}k_{im})}{\sum_i p_{io}k_{im}}100 \qquad \frac{\sum_i \frac{k_{im}}{k_{io}}(p_{im}k_{io})}{\sum_i p_{im}k_{io}}$$

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

3.1.3. FISHER-en indizeak

Fisher-en indize konplexu edo "indize ideal" deritzona, Laspeyres eta Paasche-ren indizeen batezbesteko geometrikoa da.

Hots:

$F_t^p = \sqrt{L_t^p \quad p_t^p}$ Fisher-en prezioen indize konplexua.

$F_t^k = \sqrt{L_t^k \quad p_t^k}$ Fisher-en kopuruen indize konplexua.

3.1.4. PROPIETATE ETA ERLAZIO BATZU

Laspeyres eta Paasche-ren indizeak (1) formulen bidez batez besteko gehikortu ponderatu bezala adierazten dira, eta (2) formulen bidez batezbesteko aritmetikoa ponderatu bezala.

Indizeen propietate garrantzitsu bat elkargarri 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.

Hots: $L_t^p \cdot p_t^k = B_t$ $\qquad$ $p_t^p \cdot L_t^k = B_t$

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

3.1.5. KALKULAKETA

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

	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:

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

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

$$L^p_{64} = \frac{5.680}{4.120}100 = 138$$

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

Bestalde, Paasche-ren prezio indizeaz:

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

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

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

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

III.4. INDIZE KONPLEXUEN ERAIKETAN SORTZEN DIREN ZENBAIT ERAGOZPEN

4.1. ALDAGAIEN HAUTAPENA

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

Beraz, aipatu dugun multzoaren hautapena, lehendabiziko arrazoa 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 izendatzen zehatz batetara eramaten gaitu. Honela, denbora iragatean, datu segidak erizpide berdinekoak izango baitira.

4.2. SOMATUTAKO LEKU ETA DENBORAREN HAUTAPENA

Arduraz produktuen kalitateak definitzea eta somaketen 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 hilebete 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.

4.3. TALDE ETA AZPITALDEEN HAUTAPENA

Indize konplexua kalkulatzean, honetaz gainera aldagai,talde eta azpitalde batzuren "konplexua" kalkulatzen 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 sailketa batetan, barne kontsumorako eta exportaketarako produktuak izan zitezkeen.

Talde eta azpitalde hauen hautapena, egin nahi den ikerketaren menpe egongo da.

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 nakazal 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.

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, Fischer-ena dakigunez, beste bien artean batezbesteko geometrikoa da.

4.6. INDIZEAREN ESANGURA ETA ZABALDURA

Ondasunen kopuruaren menpe zabaldura 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 kalkulatzen badi tugu, emaitzen desberdintasuna gehitu egingo da oinarri denbora tik 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

5.1. OINARRI-ALDAKETA INDIZE SINPLEETAN

Adibide batez ikus dezagun

Urtea	L_t (oinarri = 1.955)	I_t^i (oinarri = 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 hortan segida denborala hautsi egiten da eta segida denboral berrian atzerantz erregla 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 salduzela 1.962. urtean 1955. ean baino eta % 51.5 gutxiago salduzela 1962. urtean 1965. ean baino.

5.2. BERRIZTAPEN ETA LOTURAK INDIZE KONPLEXUETAN

Produkzio, aktibitate eta gastuen aldaketaren ondorio beza la, 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 arruntak egiten 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), kontsumoprezioen indizea berriztatu egiten du eta gaur egun bi segidaude, bata 1936. urtean oinarritua eta bestea 1958. ean.

KONTSUMO-PREZIOEN INDIZEA:

Urte	Oinarria 1.936 = 100	Oinarria 1.958 = 100
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 bakar bat lortzen duenez gero, egin liteke 1.936 urtean edo 1.958. ean oinarrituz . Bietan indize sinpleetan ikusi genuen erregla 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, bider 8'077 zatituko da.

Ondoko taulan, loturik dauden bi segida ikusiko ditugu.

Urte	Oinarria 1.936 = 100	Oinarria 1.958 = 100
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 gonbaragarriak, aldagai eta ponderazio desberdinetaz eginak baitaude, baina, ez dugu beste biderik, indize konplexuari ja -rraitasuna eman nahi badiogu.

Ez dugu ahaztu behar, indizeak fenomeno baten eboluzioa i-kusteko 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, exportazioen indizeak eta abar, gehienak ekonomi aktibitatearen eboluzioa ikusteko baliagarriak direlarik.

Ikasgai honetan, kontsumo-prezioen indizea ikusiko dugu , garrantzitsuena baita. (¡)

6.1. KONTSUMO-PREZIOEN INDIZEAK

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

Egikera erabiliena "erosketa-otarrea" deitzen dena da; hau inkesta baten bidez, aurrekontuan sartzen diren ondasunen kopuruak, zerbitzuak eta industri produkuak osatzen dute.

(¡) Besteak U.E.U.-n 1.979 urtean argitaratu zen "ZENBAKI-INDIZEAK" liburuskan ikus daitezke.

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

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

Bitez:

k_{10} k_{20} .... k_{no} "otarre"- ko ondasunen kopuruak (bizi-maila presuposatzen dute - nak).

P_{10} P_{20} .... P_{no} beren prezioak oinarri-denboran.

eta

P_{11} P_{21} .... P_{n1} beren prezioak oraingo unean.

Orduan kontsumo-prezioen indizea, ondoko formulaz edukiko dugu.

$$I_{01} = \left(\sum_{i=1}^{n} p_{i1}k_{i0} \; / \; \sum_{i=1}^{n} p_{i0}k_{i0} \right) 100$$

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

A R I K E T A K

1) 1) 1.970 urtea oinarritzat harturik, kalkula itzazu Laspeyres-en indizea prezioetarako eta Paasche-ren indizea kopuruetarako ondoko datuen bidez:

URTEAK	A ONDASUNA		B ONDASUNA	
	kop.	Prez.	kop.	Prez.
1.970	3	2	5	2
1.971	4	2	6	3
1.972	5	3	4	3
1.973	6	4	5	3

2) 1.972 urtea oinarritzat harturik, 1.973 urtekoa berritu, eta kalkula itzazu 1.970 eta 1.971 urtekoak, lotura-eragiketa erabiliz.

3) Zer eragozpen du loturak?

2) Ondoko taulan datuak mila toneladatan adierazten dira.

URTEAK	Bakailo eta kidekoak	Sardina eta kidekoak
1.965	282	249
1.966	312	220
1.967	349	197
1.968	350	194
1.969	297	185

Ezkatzen da:

1) kalkula itzazu ahal dituzun indizeak.
2) Komenta itzazu laburki indize hauek dituzten eragozpenak.

3) 1) Azal itzazu, indize sinple eta konplexuen arteko desberdintasunak; baita ponderatu eta ezponderatuen artekoak ere.

Zein indize erabil daiteke ondasun batzu neurri desberdine tan adieraziak izan badira?

2) Kalkula ezazu kontsumo-prezioen indizea ondoko datuen - bidez.

Ondasunak	Kopuruak	Prezioak 1.969-1.979	
A	80	8	7
B	40	10	14
C	30	5	11
D	50	20	25

4) Suposa dezagun neurri-konbentziozkoetan emandakoak diren on doko informazio-estatistikaz baliatzen garela.

	A		B		C	
	kop.	prez.	kop.	prez.	kop.	prez.
1.976	5	4	2	2	2	4
1.977	3	5	4	6	3	6
1.978	2	6	4	10	5	8

1.976 urtea oinarritzat harturik, zera eskatzen da:

1) Prezio-indize sinpleak, hiru ondasunen kasuan.
2) Laspeyres,Paasche eta Fisher-en prezio-indize konplexu ponderatuak.

5) Zeintzu indize kalkula daitezke, litro bat olio, ehun gramo gurin, litro bat esne eta dozena bat arraultzeri dagozkien - prezio-segidak ondokoak badira 1.970-77 urte-tartean ?

Urteak	O	G	E	A
1.970	54	20	17	43
1.971	58	27	18	45
1.972	60	30	19	50
1.973	62	31	20	53
1.974	67	34	23	56
1.975	70	36	28	61
1.976	74	37	31	64
1.977	80	40	32	70

6) Automobilgintzan diharduen enpresa batek bi auto-mota kaleratzen ditu,"A" eta "B" motakoak alegia.

1.979-82 urteetako produkzio-indize bat ateratzeko ondoko datuak dauzkate.

	"A" MOTA		"B" MOTA	
	Prezio/unitate milako pezetatan	Salmentak milako unitatetan	Prezio/unitate milako pezetatan	Salmentak milako unitatetan
1979	700	300	450	400
1980	800	350	600	425
1981	950	270	600	600
1982	1.200	150	600	500

Eskatzen da:

1. Paasche-ren prezio eta kopuru-indizeak, 1.979 urtean oinarriturik.
2. Lortutako emaitzak interpretatzea, zeintzu konklusio aterako zenituzkeen azalduz.

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

IV.1. EZAUGARRI ESTATISTIKO BIKOITZAK.

IV.2. MAIZTASUN-BANAKETA BIKOITZAK

2.1. MAIZTASUN-BANAKETAK: BIKOITZAK ETA BAZTER-MAIZTASUNAK.

2.2. MAIZTASUN-BANAKETA BIKOITZAK: TAULAK

IV.3. MAIZTASUN-BANAKETA BIKOITZEN ADIERAZPIDE GRAFIKOAK.

3.1. SAKABANATZE-DIAGRAMA EDO PUNTU-HODEIA

IV.4. MAIZTASUN-BANAKETA BALDINTZATUAK

IV.5. TAULA BATEN DEPENDENTZIA EDO INDEPENDENTZIA ESTATISTIKOA

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

2.1. MAIZTASUN-BANAKETAK BIKOTZAK ETA BAZTER-MAIZTASUNAK

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

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

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

Hots:

$$\omega_p \longrightarrow (x_i, y_j) \qquad \begin{array}{l} p = 1, \ldots\ldots, N \\ i = 1, \ldots\ldots, K \\ j = 1, \ldots\ldots, 1 \end{array}$$

eta (X,Y) ezaugarri bikotzak, laginean hartuko lituzkeen balio-bikote desberdinen kopurua, k x l izango da; hemendik aurrera horiek bakarrik hartuko ditugu kontuan.

Hots:

$$\{\omega_1 \ , \ \omega_2 \ \ldots\ \omega_N\} \xrightarrow{(X,Y)} \{(x_1y_1) \ (x_1y_2) \ \ldots (x_iy_1) \ \ldots (x_iy_1) \ \ldots \ (x_ky_1) \ldots \ \ldots (x_ky_1) = (x_iy_j)\}$$

(x_iy_j) balio-bikotearen MAIZTASUN ABSOLUTUA

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

Non:

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

(x_iy_j) balio-bikotearen MAIZTASUN ERLATIBOA

Sinbolikoki f_{ij} ikurraz adieraziko dugu, non:

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

eta, portzentaietan:

$$p_{ij} = \%f_{ij} \ . \ 100 \qquad \sum_{j} \sum_{i} p_{ij} = 100$$

x_i balioaren BAZTER-MAIZTASUN ABSOLUTUA

sinbolikoki $n_{i\bullet}$ ikurraz adieraziko dugu, non:

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

y_j balioaren BAZTER-MAIZTASUN ABSOLUTUA

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

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

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 $$f_{..} = \sum_{i} f_{i.} = \sum_{j} f_{.j} = 1$$

2.2. MAIZTASUN-BANAKETA BIKOITZAK: TAULAK

Lagin bati dagozkion maiztasun-banaketen taulak, ezaugarria bikotza bada, ondoan azaltzen ditugunak dira:

BANAKETA BIKOTZA

X \ Y	y_1	y_2	..	..	y_j	..	..	y_l	Guztira
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_{21} f_{21}	$n_{2.}$ $f_{2.}$
. .			.. ..	.. ..	.. ..	.. ..			
x_i	n_{il} f_{il}	n_{i2} n_{i2}	..	..	n_{ij} f_{ij}	..	..	n_{il} f_{li}	n_i f_i
. .	. .	. .	.. ..	.. ..	. .	.. ..	. .		. .
x_k	n_{k1} f_{k1}	n_{k2} k_{k2}	..	..	n_{kj} f_{kj}	..	..	n_{kl} f_{kl}	n_k f_k
Guz-tira	$n_{.1}$ $f_{.1}$	$n_{.2}$ $f_{.2}$			$n_{.j}$ $f_{.j}$			$n_{.l}$ $f_{.l}$	$n_{..}=N$ $f_{..}=1$

BAZTER-BANAKETAK

X	M. Ab.	M. Er.
x_1	$n_{1.}$	$f_{1.}$
x_2	$n_{2.}$	$f_{2.}$
⋮	⋮	⋮
x_i	$n_{i.}$	$f_{i.}$
⋮	⋮	⋮
x_i	n_k	f_k

Y	M. AB.	M. ER.
y_1	$n_{.1}$	$f_{.1}$
y_2	$n_{.2}$	$f_{.2}$
⋮	⋮	⋮
y_j	$n_{.j}$	$f_{.j}$
⋮	⋮	⋮
y_l	n_l	f_l

$$\sum_i n_{i.} = N \qquad \sum_j n_{.j} = N$$

$$\sum_i f_{i.} = N \qquad \sum_j f_{.j} = N$$

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

Adibidea: 750 familia duen talde batetan 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	(20 tik , 30-ra)	(30tik,40 ra)	(40tik , 50ra)	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 dagokionak), 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 bikotz baten maiztasun-banaketa ikusterazteko bi adierazpide grafiko ditugu, ESTEREOGRAMA eta SAKABANATZE-DIAGRAMA edo PUNTU-HODEIA hain zuzen.

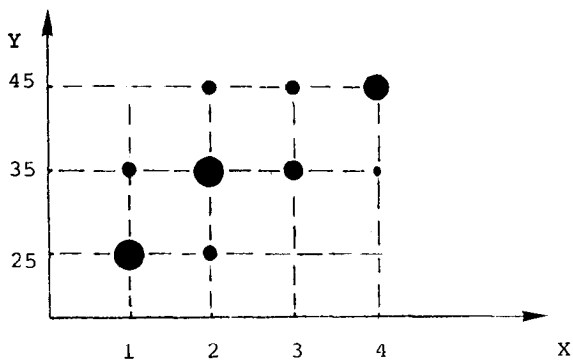
Hauetariko bigarrena ondoan azalduko dugu, bestea ez baita ia erabiltzen.

3.1. SAKABANATZE-DIAGRAMA EDO PUNTU-HODEIA

Bedi (X,Y) ezaugarri estatistiko bikotz bat, eta bitez $\left\{(x_i y_j)\right\}_{\substack{i=1,\ldots,k \\ j=1,\ldots,l}}$ (X,Y) ezaugarriak N tamainuko lagin batetan hartzen dituen balio-bikote desberdinak.

Orduan, ardatz cartesiarra sistema batetan puntu batez adierazten badugu, $(x_i y_j)$ balio-bikote bakoitza, $\left\{(x_i y_j)\right\}_{\substack{i=1,\ldots,k \\ j=1,\ldots,l}}$ multzo osoaren adierazpideari, sakabanatze-diagrama edo puntu-hodeia deritzo.

Ondoa ; 2.2. ataleko adibideari dagokion puntu-hodeia daukagu.



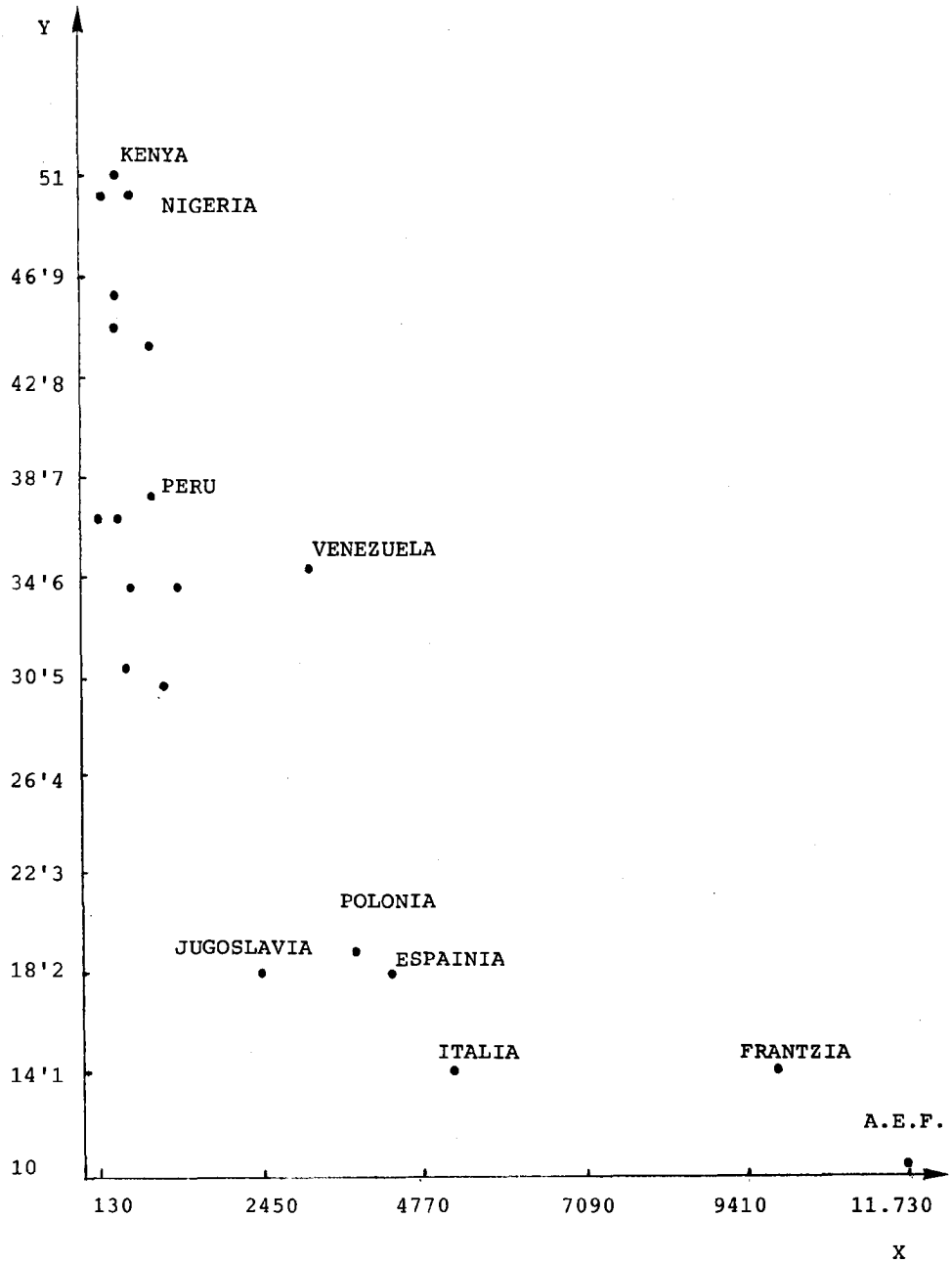
Ikusten dugunez, $(x_i y_j)$ balio-bikoteak errepikatzen direnean sakabanatze-diagramako puntuetariko batzu besteak baino lo diagoak dira, beraien azalerak eta balio-bikoteei dagozkien - maiztasunak proportzionalak izanik.

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

Batez ere datu-multzo handiak ditugunean, datu-isolatuenak edo ezer gutxi errepikatzen diren datuenak dira taulak; eta holako kasuetan, puntu-hoideak oso baliagarriak zaizkigu, multzoren ibilbidea ikusterazteko.

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

ESTATUA	IZENA	NPG	JAIOKORTASUNA
1	ETHIOPIA	130	50
2	BIRMANIA	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	AL.ERR.FEDERATUA	11730	10


Y
KENYA
51
NIGERIA
46'9
42'8
38'7
PERU
VENEZUELA
34'6
30'5
26'4
22'3
POLONIA
JUGOSLAVIA
ESPAINIA
18'2
ITALIA
FRANTZIA
14'1
A.E.F.
10
130
2450
4770
7090
9410
11.730
X

IV.4. MAIZTASUN-BANAKETA BALDINTZATUAK

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

Y / X	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 ondokiak 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	$n_{2/i} = n_{i2}$	$\frac{n_{2/i}}{n_{i.}}$
.	.	.
.	. (1)	. (2)
.	.	.
y_j	$n_{j/i} = n_{ij}$	$\frac{n_{j/i}}{n_{i.}}$
.	.	.
.	.	.
.	.	.
y_1	$n_{1/i} = n_{il}$	$\frac{n_{1/i}}{n_{i.}}$

eta (1) x_i balioaz baldintzatutako Y aldagaiaren 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 baldintzatzaileak izanez.

Adibidea:

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

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

X \ 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

Holako banaketa baldintzatuak zera adierazten digute, hain zuzen: aldagai baten balio bat finkoa 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:

X \ Y	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_i/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 hoiek, bada, ez daude berdin banatuta akastunen artean eta akasgabeen artean, hots, nolabaiteko menpekotasun edo dependentzia ikusten da taula honetan.

Demagun, orain, honako taula hau

X \ Y	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.

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

Independentziaren baldintza beharrezkoa eta nahikoa ondokoa da:

$$\frac{n_{ij}}{N} = \frac{n_i}{N} \cdot \frac{n_j}{N} \qquad \forall i,j$$

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

V. EZAUGARRI BIKOITZEN BALIO TIPIKOAK

V.1. MOMENTUAK

1.1. MOMENTU ARRUNTAK EDO JATORRIAREKIKOAK

1.2. MOMENTU ZENTRALAK EDO BATEZBESTEKOARE-KIKOAK

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

2.1. KOBARIANTZA : S_{xy}

2.2. KORRELAZIO-KOEFIZIENTEA: r_{xy}

V.3. ALDAGAIEN ALDAKETAK

3.1. ALDAGAI TIPIFIKATUA

V.4. SAKABANATZE ETA KORRELAZIO-MATRIZEAK

V.5. OROKORPENA

ARIKETAK

V.1. MOMENTUAK

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

1.1. MOMENTU ARRUNTAK EDO JATORRIAREKIKOAK

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

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

1-ordenako edo lehenengo ordenako momentu arruntak hauexek dira:

$$a_{10} = \frac{1}{N} \sum_i \sum_j x_i n_{ij} = \frac{1}{N} \sum_i x_i \sum_j n_{ij} =$$

$$\frac{1}{N} \sum_i x_i n_{i.} = \bar{x}$$

$$a_{01} = \frac{1}{N} \sum_i \sum_i y_j n_{ij} = \frac{1}{N} \sum_j y_j \sum_i n_{ij} =$$

$$\frac{1}{N} \sum_j y_j n_{.j} = \bar{y}$$

2-orderako edo bigarren ordenako momentu arruntak hauexek dira

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

$$\frac{1}{N}\sum_i x^2_i \quad n_{i.} = a_2$$

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

$$a_{02} = \frac{1}{N} \sum_i \sum_j y^2_j \; n_{ij} = \frac{1}{N} \sum_j y^2_j \sum_i n_{ij} =$$

$$= \frac{1}{N} \sum_j y^2_j \; n_{.j} = a_2$$

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:

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

1.2. MOMENTU ZENTRALAK EDO BATEZBESTEKOAREKIKOAK

Maiztasun-banaketa bikoitz baten h-ordenako momentu zentrala edo batezbestekoarekikoa, 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:

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

$$m_{20} = \frac{1}{N} \sum_i \sum_j (x_i - \bar{x})^2 \; n_{ij} = \frac{1}{N} \sum_i (x_i - \bar{x})^2 \sum_j n_{ij} =$$

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

Hori da, bestetik, X ezaugarriaren bariantza (ikus 34. orrialdea)

$$m_{20} = \frac{1}{N} \sum_i \sum_j (y_j - \bar{y})^2 n_{ij} = \frac{1}{N} \sum_j (y_j - \bar{y})^2 \sum_i n_{ij} =$$

$$\frac{1}{N} \sum_j (y_j - \bar{y})^2 n_{.j} = S_y^2$$

Y ezaugarriaren barantza 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 ESTATISTI - KOAK

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

2.1 KOBARIANTZA: S_{xy}

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

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

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

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

$$\frac{1}{N} \sum_i \sum_j x_i y_j n_{ij} - \frac{1}{N}\ \bar{y} \sum_i \sum_j x_i\ n_{ij} - \frac{1}{N}\ \bar{x}\ \sum_i \sum_j\ y_j n_{ij} +$$

$$\frac{1}{N} \bar{x}\ \bar{y}\ \sum_i \sum_j n_{ij} = a_{11} - \bar{y}\ \bar{x} - \bar{x}\ \bar{y} + \bar{x}\ \bar{y} = a_{11} - \bar{x}\ \bar{y}$$

2.2. KORRELAZIO KOEFIZIENTEA: r_{xy}

Estatistiko honek ere X eta Y aldagaiek laginean duten erlazioa neurtzen digu, bainan kasu honetan berdin da X eta Y indibiduoen gain zer unitatetan neurtzen diren , neurketan era biltzen diren unitateek ez baitute eraginik korrelazio-koefizientean.

$$r_{xy} = \frac{S_{xy}}{S_x S_y} = \frac{\sum_i \sum_j (x_i - \bar{x})\ (y_j - \bar{y})\ n_{ij} / N}{(\frac{1}{N} \sum_i (x_i - \bar{x})^2\ n_{i.})^{1/2} (\frac{1}{N} \sum_j (y_j - \bar{y})^2 n_{.j})^{1/2}} =$$

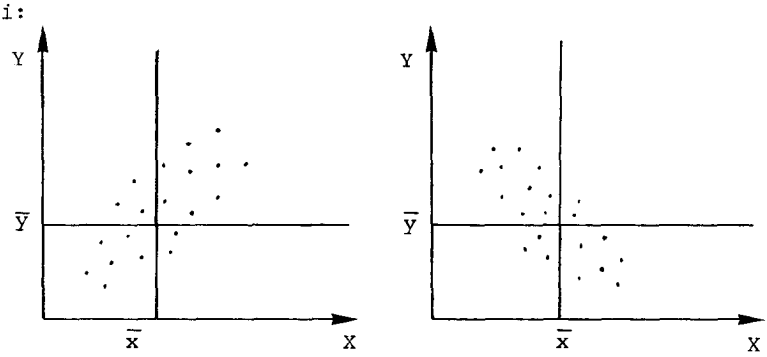
$$= \frac{a_{11} - \bar{x}\ \bar{y}}{(a_{20} - \bar{x}^2)^{1/2} (a_{02} - \bar{y}^2)^{1/2}}$$

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

Intuitiboki honetaz ohar daiteke: $(x_i - \bar{x})$ 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. koadratean egongo dira) $S_{xy} > 0$ izango da; aldiz, zeinu kontrakoak baldin badira (puntuak 2. eta 4. koadrantean agongo dira) $S_{xy} < 0$.

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

Grafikoki:



Kobariantzak nahiz korrelazio-koefizienteak bi ezaugarrien artean dagoen erlazioa adierazten digute; baina ikusi dugu zein kasutan izandо den $S_{xy} > 0$, hots, $(x_i - \bar{x})$ eta $(y_j - \bar{y})$ kendurak biak positiboak direnean (hau da bi ezaugarriak gorakoarrak dira) edo biak negatiboak direnean (hau da, bi ezaugarriak beherakorrak) $S_{xy} < 0$ 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.

Aldiz, r_{xy} koefizienteak ez du erlazio hau nolakoa den bakarrik esaten (r_{xy} - ren zeinua S_{xy} -rena baita) baizik eta zer mailatan erlazionatuta dauden ere bai.

Korrelazio-koefizienteak har ditzakeen balioak, $\left[-1, +1 \right]$ tartekoak dira.

Zerora hurbiltzen den balio bat hartzen badu, bi ezauga - rrien arteko erlazio 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 erreferitzen ari gatzaizkio; baina ez - du zerikusirik estatistika teorikoko aldagai aleatorio delakoarekin.

Ikus dezagun, bada, (X,Y) aldagaien ordez, haiekin erlazio natuta dauden (U,V) aldagaiak erabiltzen ditugunean, kobariantza aldatu egiten dela eta korrelazio-koefizientea ez.

Horrexegatik eta baita ere erregresio-teorian betetzen - duen funtzioagatik, korrelazio-koefizientea, balio tipiko ga - rrantzitsuena 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

badugu (¡), (ikus 106. orrialdean)

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

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

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

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

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

$$S_x^2 = 1/N\left(\sum_i (x_i - \bar{x})^2 n_i\right) = 1/N\left(\sum_i (au_i + b - a\bar{u} - b)^2 n_i\right) =$$

$$= a^2 . 1/N \left(\sum_i (u_i - \bar{u})^2 n_i\right) = a^2 . S_u^2$$

eta: $S_x = a.S_u$

Analogikoki $S_y = e.S_v$

$$S_{xy} = 1/n\left(\sum_i \sum_j (x_i - \bar{x}) (y_j - \bar{y}) n_{ij}\right) =$$

$$= 1/N\left(\sum_i \sum_j (au_i + b - a\bar{u} - b) (ev_j + f - e\bar{v} - f) n_{ij}\right) =$$

$$= a.e . 1/N \left(\sum_i \sum_j (u_i - \bar{u}) (v_j - \bar{v}) n_{ij}\right) = a.e.S_{uv}$$

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

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

Hots:

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

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

Adibidea: Lagin batetan X eta Y ezaugarriak aztertu dira. Maiz tasun-banaketaren taula ondokoa bada, kalkula ditzagun: batez-bestekoak, bariantzak, kobariantza eta korrelazio-koefizientea

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

Aldaketa hauek egiten ditugu: $u_i = \frac{x_i - 243}{4}$,

$v_j = \frac{y_j - 1'740}{0'005}$

hots: $x_i = 4u_i + 243$ 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:

$$\bar{u} = \frac{\sum_i u_i n_{i.}}{N} = \frac{55}{100} = 0'55$$

$$\bar{x} = 4 \cdot 0'55 + 243 = \underline{245'3}$$

$$\bar{v} = \frac{\sum_i v_j n_{.j}}{N.} = 8/100 = 0'08$$

$$\bar{y} = 0'005 . 0'08 + 1'740 = 1'7404$$

$$s_u^2 = a_{20} - \bar{u}^2 = \frac{\sum_i u_i^2 n_{i.}}{N} - \bar{u}^2 = 1'2475$$

$$s_x^2 = 4^2 . 1'2475 = \underline{19'96}$$

$$s_y^2 = a_{02} - \bar{v} = \frac{\sum_i v_i^2 n_{.j}}{N} - \bar{v}^2 = 0'6736$$

$$s_y^2 = 0'005^2 . 0'6736 = \underline{0'00001}$$

$$s_{uv} = 1/N \sum_i \sum_j u_i v_j n_{ij} - \bar{u}\,\bar{v} = 0'126$$

$$s_{xy} = 4 . 0'005 . 0'126 = \underline{0'00252}$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{s_{uv}}{s_u s_v} = \underline{0'178}$$

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

3.1. ALDAGAI TIPIFIKATUA :

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

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

Hau da:

$$\bar{z} = \frac{1}{N} \sum_i z_i n_i = \frac{1}{N} \sum_i (x_i - \bar{x}) n_i =$$

$$= \frac{1}{N} \sum_i x_i n_i - \frac{1}{N} \bar{x} \sum_i n_i = \bar{x} - \bar{x} = 0$$

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

$$\bar{z} = \frac{1}{S_x} \sum_i (x_i - \bar{x}) \frac{n_i}{N} = 0$$

$$S_z^2 = \frac{1}{N} \sum_i (z_i - \bar{z})^2 n_i = \frac{1}{N} \sum_i \frac{(x_i - \bar{x})^2}{S_x^2} n_i =$$

$$= \frac{1}{S_x^2} \frac{1}{N} \sum_i (x_i - \bar{x})^2 n_i = \frac{1}{S_x^2} S_x^2 = 1$$

KORRELAZIO-KOEFIZIENTEA, aldagai tipifikatuen arteko kobarian tza da.

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

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

$$= \frac{1}{S_x S_y} S_{xy} = r_{xy}$$

V. 4. SAKABANATZE ETA KORRELAZIO-MATRIZEAK

Momentu zentralak honela ere adierazten dira:

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

letraz laburki adierazten ditugun ondoko matrizeak, sakabanatze edo kobariantza eta korrelazio-matrizeak dira.

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

Ikusten denez, simetrikoak dira biak.

Lehengo adibidekoak hauek dira:

$$\underline{\underline{L}} = \begin{bmatrix} 19'96 & 0'0025 \\ 0'0025 & 0'0001 \end{bmatrix} \qquad \underline{\underline{R}} = \begin{bmatrix} 1 & 0'178 \\ 0'178 & 1 \end{bmatrix}$$

V. 5. ORÒKORPENA

Suposa dezagun n ezaugarri edo aldagai ditugula. Binaka hartuz, kobariantzak eta korrelazio-koefizienteak atera daitezke; orduan, $\bar{\bar{L}}$ eta $\bar{\bar{R}}$ matrizeak hauexek izango lirateke:

$$\underline{\underline{L}} = \begin{bmatrix} l_{11} & l_{12} & \cdots & l_{1n} \\ l_{21} & l_{22} & \cdots & l_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{bmatrix} \qquad \underline{\underline{R}} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1n} \\ r_{21} & 1 & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & 1 \end{bmatrix}$$

non l_{11} , l_{22}, ... 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 korrelazio-koefizientea da eta diagonaleko 1 balioak aldagai bakoitzak berekiko duen korrelazioa.

A R I K E T A K

1) Euskal Herriko lau probintzi hauen kasuan, X aldagaiak populazio-igoera ematen digu ehunekotan 1.973-1.976 tartean eta Y aldagaiak "per capita" famili-errentaren igoera urte tarte berean.

	X	Y
ARABA	6	49
GIPUZKOA	2'5	46
NAFARROA	1	43
BIZKAIA	3'5	44

Eskatzen da:

1. X eta Y aldagaien batezbestekoak eta desbidazio tipikoak.
2. (X,Y) aldagaien kobariantza eta korrelazio-koefizientea, emaitza irazkinduz edo komentatuz.

2) Ondoko datuak 1.956, 1.957, ... 1.964 urteetako diru-joanetorriaren kopuruak eta prezio-indizeak baldin badira,

Urteak	Diru kop.	Pre.- ind.
1.956	4.377	174'8
1.957	4.244	167'9
1.958	4.284	161'4
1.959	4.346	167'7
1.960	4.499	167'0
1.961	5.013	169'8
1.962	4.879	166'6
1.963	4.788	159'1
1.964	4.678	163'4

Kalkula ezazu korrelazio-koefizientea,datuak sinplifikatzeko ahal duzun aldaketa eginez.

3) Ondoko taulan maiztasun-banaketa bat ematen da.

X \ Y	1	2	3
10	0	4	10
11	5	9	1
12	9	2	0

1. Lor itzazu bazter-banaketak eta banaketa-baldintzatu - guztiak.

2. U, V aldagaiak honela atera tzen badituzu:

   $X = Ua + b$, $Y = Ve + f$

   Zein da S_{xy} eta S_{uv} kobariantzen arteko erlazioa?

3. Kalkula ezazu korrelazio-koefizientea, laburrena den kalkulua eginez.

4) 1.976 urteko datu hauek ditugu Iran, Irak, eta Israel he - rrietan.

X_1 aldagaiak ospitale-oheen kopurua/ /biztanle adierazten du.

X_2 - k zinetoki-jesarlekuak/ 1.000 biztanle.

X_3 -k energi-kontsumoa/biztanle.

	x_1	x_2	x_3
IRAN	2	5	0'7
IRAK	1'6	8'2	1'5
ISRAEL	7	44'5	2'5

Kalkula ezazu korrelazio-matrizea, korrelazio horien esana - hia azalduz.

VI. KORRELAZIOA ETA ERREGRESIOA R^2-n

VI.1. KORRELAZIOA ETA ERREGRESIOA

VI.2. ERREGRESIO-TEORIA

2.1. BATEZBESTEKOAREN ERREGRESIOA

2.2. KARRATU TXIKIENEN ERREGRESIOA

2.2.1. KARRATU TXIKIENEN ERREGRESIO LINEALA

2.2.1.1. KORRELAZIO-KOEFIZIENTE LINEALA, HONDAR-BARIANTZA ETA KOEFIZIENTE MUGATZAILEA

2.2.1.2. ERREGRESIO LINEALEAN BILAKATZEN DIREN BESTE BI KASU

2.2.2. KARRATU TXIKIENEN ERREGRESIO PARABOLIKOA

ARIKETAK

VI.1. KORRELAZIOA 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:

i) Erregresio tekniken bidez, $y = \emptyset\ (x, a, b \ldots l)$ funtzio matematikoa lortuko dugu, zeinak, x-en balioak ordezkatuz gero, y-ren balioak emango baitizkigu.

   $\emptyset$ funtzioa puntu-hodeian ondo doitzen bada, Y aldagaiaren balioak aurresateko baliagarria izango zaigu.

ii) Korrelazio tekniken bidez, ordea, koefiziente batzu lortuko ditugu, aldagaien elkar-aldaketak eta beren norantza - ikusaraz diezazkigutenak.

Erregresio-teknikak zenbakizkoak edo kuantitatiboak deitzen dira, dagoen menpekotasuna zenbatzen baitute; aldiz, nolakotasunezko korrelazio - teknikek, menpekotasun hori handiago edo txikiago den bakarrik esaten digute. Hala ere, korrelazio-koefi - zienteak 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 dispertsio-diagrama edo puntu-hodeia kontutan - harturik.

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

Erregresio-metodo asko daude, hauetariko batzu hauek dire larik.

- Batezbestekoaren erregresioa.
- Karratu txikienen erregresioa { Erregresio lineala / Erregresio parabolikoa }
- Erregresio ortogonala.

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

Edozein metodori jarraituz, bi erregresio desberdin kontsi dera daitezke: 1) X aldagaia independentetzat hartuz, Y -ren X-ekiko erregresioa deitzen dena. 2) Y aldagaia independentetzat hartuz, X-ren Y-rekiko erregresioa.

2.1. BATEZBESTEKOAREN ERREGRESIOA

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

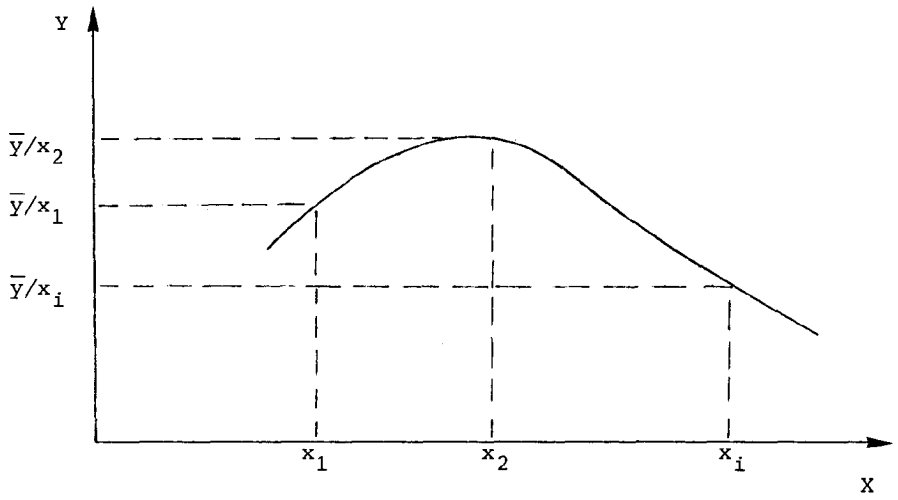
$x_1, x_2, \ldots\ldots, x_n$ balioek baldintzatutako Y aldagaiaren banaketak hurrenez hurren $f(y/x_1)$, $f(y/x_2)$, ... $f(y/x_n)$, baldin badira, batezbesteko balioak honela lortuko ditugu:

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

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

.
.
.

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



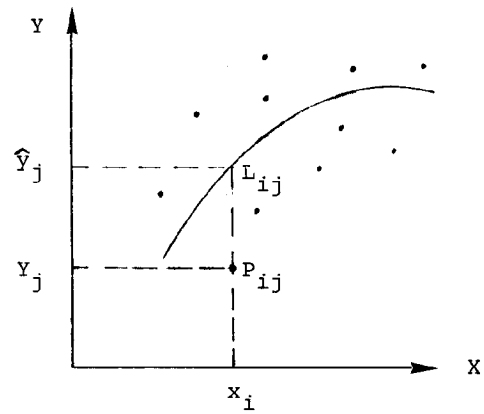
2.2. KARRATU TXIKIENEN ERREGRESIOA

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

Suposa dezagun kurba hau, $y = \emptyset\ (x, a, b, \ldots, l)$ kurba-familia honetakoa dela, arazoa orduan a, b, ...,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.

$$E = \sum_i \sum_j \overline{P_{ij}\ L_{ij}}^2\ f(x_i y_j) = \sum_i \sum_j (y_j - \emptyset\ (x_i, a, b, \ldots 1))^2 f(x_i y_i)$$



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

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

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

2.2.1. KARRATU TXIKIENEN 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.

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

txikiena izateko baldintzaz, $\hat{y} = a + bx$ zuzeneko a eta b ba - lioak mugatuko ditugu.

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

Dakigunez,

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

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

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

hau da: $\sum_{ij} f_{ij} (y_j - bx_i - a) = 0$; $\sum_{ij} f_{ij} (y_j - bx_i - a)\, x_i = 0$

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

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

Hots:

$$\boxed{\begin{array}{l} \bar{y} = a + b\bar{x} \\ \\ a_{11} = a\bar{x} + ba_{20} \end{array}}$$

EKUAZIO NORMALAK

Ekuazio normalak deitzen dira eta, ikusten dugunez, lehe - nengoa $y = a+bx$ bilatzen ari garen zuzena da $(\bar{x},\bar{y})$ puntuan har tuta; hau da, zuzena banaketaren grabitate-zentrutik pasatzen - da eta $a = \bar{y} - b\bar{x}$ bigarren ekuazioan ordezkatuz:

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

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

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

Orduan:

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

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

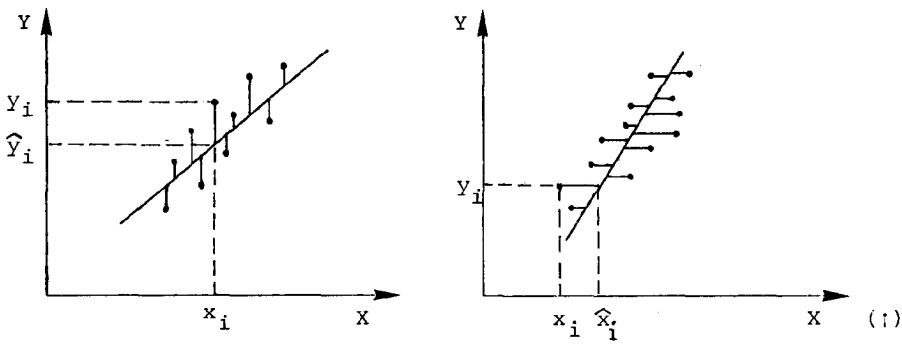
$$\hat{y} - \bar{y} = \frac{S_{xy}}{S_x^2}(x - \bar{x})$$

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} dei - tzen da).

Analogikoki X-ren Y-rekiko erregresio-zuzena ondokoa izan go 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



(¡)

KORRELAZIO-KOEFIZIENTEA ETA ERREGRESIO-KOEFIZIENTEEN ARTEKO ER-LAZIOA:

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

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

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

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, korrelazio-koefizientearen karratua da.

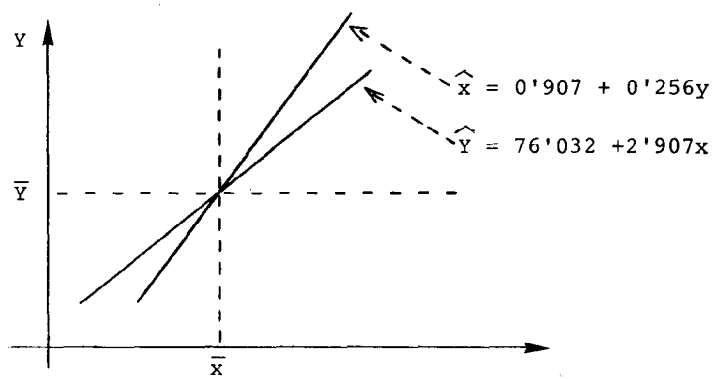
Adibidea:

X aldagaia lantegi batetan egindako lan-orduak izanik eta Y aldagaia produzitu diren unitateak, kalkula dezagun korrelazio-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

Emaitzak hauek izan dira:

$r = 0'863$ $\quad \bar{x} = 79'9$ $\quad \bar{y} = 308'36$



2.2.1.1. KORRELAZIO-KOEFIZIENTE LINEALA HONDAR-BARIANTZA - ETA KOEFIZIENTE-MUGATZAILEA

Datu estatistikoen multzoan zuzen bat doituta dagoela supo satzen badugu, bi aldagaien arteko korrelazio-koefizienteak era biltzeak zentzu asko dauka, aldagaien menpekotasuna erakus diezagun.

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

Honela, m eta n bi zenbaki erreal izanik, har dezagun $\left[m(x_i - \bar{x}) + n(y_j - \bar{y}) \right]^2$ expresio arbitrario honen batezbes teko balioa.

$$\sum_{ij} \left[m(x_i - \bar{x}) + n(y_j - \bar{y}) \right]^2 f_{ij} = m^2 \sum_{ij} (x_i - \bar{x})^2 f_{ij} + n^2 \sum_{ij} (y_j - \bar{y})^2 f_{ij} + 2mn \sum_{ij} (x_i - \bar{x})(y_j - \bar{y}) f_{ij} =$$

$$= m^2 S_x^2 + n^2 S_y^2 + 2\,mn\, S_{xy} \geqslant 0$$

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

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

$$S_y^2 S_x^2 - S_{xy}^2 \geqslant 0 \quad ; \quad S_{xy}^2 \leqslant S_y^2 S_x^2$$

$$\frac{S_{xy}^2}{S_y^2 S_x^2} \leqslant 1 \quad ; \quad r^2 \leqslant 1 \qquad \boxed{-1 \leqslant r \leqslant 1}$$

Balio teoriko eta balio errealen arteko desbidazioak (hondar-desbidazioak) kontsideratzen baditugu, beste garrantzizko neurri bat aztertuko dugu: Hondar-bariantza deitzen dena.

Desbidazio karratuen batezbesteko aritmetikoa hondar-bariantza deitzen dena da.

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

Kalkula dezagun, bada, S_e^2 hondar-bariantza.

$$S_e^2 = \sum_{ij} n_{ij} \ (y_j - bx_i)^2 \ /N =$$

$$= \sum_{ij} n_{ij} \ (y_j^2 - 2\, bx_i y_j + b^2 x_i^2) \ /N =$$

$$= S_y^2 - 2\, b\, S_{xy} + b^2\, S_x^2 = \quad (¡)$$

$$= S_y^2 - 2\, \frac{S_{xy}^2}{S_x^2} + \frac{S_{xy}^2}{S_x^4}\ S_x^2 =$$

$$= S_y^2 - \frac{S_{xy}^2}{S_x^2} = S_y^2 \ (1 - \frac{S_{xy}^2}{S_x^2\, S_y^2}) =$$

$$= S_y^2 \ (1-r^2) \ \geqslant \ 0$$

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

Hots:

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

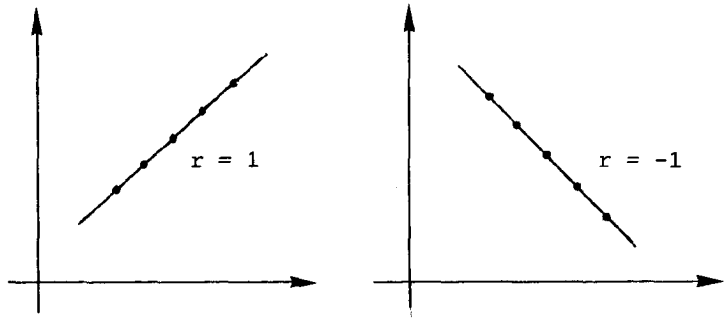
$r^2 \leqslant 1$ da eta hondar-bariantza Y aldagaiaren bariantza bai no txikiagoa da beti.

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

---

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

Grafikoki:



Puntuak zenbat eta zuzenetik gutxiago desbidatu, S_e^2 txi - kiago 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 koefiziente mugatzailea deitzen zaio.

Beste modu batetara esateko:

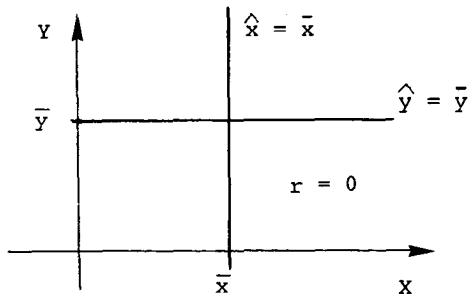
$r^2 = 1$ baldin bada, Y aldagaiaren aurresana ahal den ge - hiena hobetu dugu erregresio analisiaz.

100 r^2 kontsideratuz, aurresana ehunekotan zenbat hobetzen den jakingo dugu, erlazio lineala suposatzen ez genuen situa - zio batekin konparatuz.

Adibidez, $r^2 = 0'64$ baldin bada, Y-ren aurresana, X-aren - aldaketak kontuan izanik, erregresio-zuzenaren bidez % 64 ean hobetu dugula esango da.

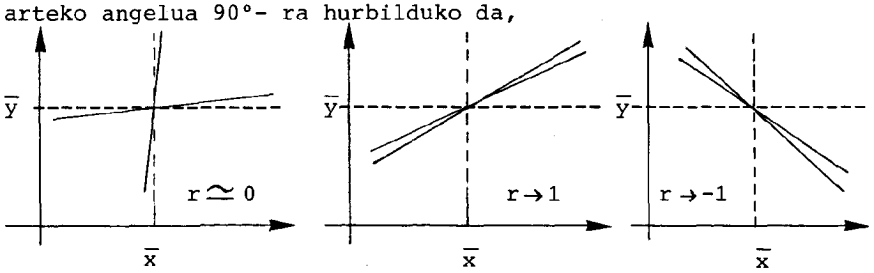
Kasu berezi batetan, $r^2 = 0$ baldin bada, situazioa ezertxo ere ez da hobetu erregresioaren bidez; aldagaiak ezkorrelatuak direla esango da eta erregresio-zuzenak ardatzekin paraleloak dira.

Batezbesteko balioak independenteak dira eta bazter-baztez bestekoen gain(zeinekin koitziditzen baitute) ez dute aurresana hobetzen.



Halako erregresio-zuzenak guztiz sakabanatuak dauden da - tuei dagozkie eta, zerbait aurresan bahar badugu, zuhurrena ba_ tezbestekoa 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 ezkorrelatuak (r=0) diren ala ez esan dezakegu eta baita joera bat noiz da - goen ere, baina batzutan gehiagorik ez.

Erlazio linealen batean pentsatzen hasteko, r 0'70-tik go_ rakoa izatea eskatzen da (hau da $r^2 = 0'49$) eta askok _are ge - hiago eskatzen du.

2.2.1.2. ERREGRESIO LINEALEAN BILAKATZEN DIREN BESTE BI KASU

Puntu-hodeia kontsideratuz, berreketazko erlazio bat - $y = ax^b$ edo exponentzial bat $y = a\, e^{bx}$ suposatzen badugu, logaritmoak hartuz, aldagai transformatu batzuren erlazio linealean bilakatzen dira.

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

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

2.2.2. KARRATU TXIKIENEN ERREGRESIO PARABOLIKOA

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

$$\hat{y} = b_o + b_1 x + b_2 x^2 + \ldots + b_k x^k$$

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

Bigarren graduko polinomio bat (parabola bat) kontsideratuz:

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

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

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

$$\frac{\partial E}{\partial b_1} = -2 \sum_{ij} f_{ij} \, (y_j - b_0 - b_1 x_1 - b_2 x_i^2) \, x_i = 0$$

$$\frac{\partial E}{\partial b_2} = -2 \sum_{ij} f_{ij} \, (y_j - b_0 - b_1 x_1 - b_2 x_i^2) \, x_i^2 = 0$$

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

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

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

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 baditugu (n=1 beti izanik)

$$\frac{\sum y_j}{N} = b_0 + b_1 \frac{\sum x_i}{N} + b_2 \frac{\sum x_i^2}{N}$$

$$\frac{\sum x_i y_j}{N} = b_0 \frac{\sum x_i}{N} + b_1 \frac{\sum x_i^2}{N} + b_2 \frac{\sum x_i^3}{N}$$

$$\frac{\sum x_i y_i}{N} = b_0 \frac{\sum x_i^2}{N} + b_1 \frac{\sum x_i^3}{N} + b_2 \frac{\sum x_i^4}{N}$$

$$\left.\begin{aligned} \sum y_j &= N b_0 + b_1 \sum x_i + b_2 \sum x_i^2 \\ \sum x_i y_j &= b_0 \sum x_i + b_1 \sum x_i^2 + b_2 \sum x_i^3 \\ \sum x_i^2 y_j &= b_0 \sum x_i^2 + b_1 \sum x_i^3 + b_2 \sum x_i^4 \end{aligned}\right\} \quad (1)$$

Ikusten denez,(1) ekuazioak erraz gogoratuko ditugu, ze - ren lehenengoa, batukaria sartuz eta b_0 koefizientea N-z bider katuz, parabolarena baita eta beste biak, x_i eta x_i^2 - z bider katuz, 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_iy_i	x_i^3	x_i^4	$x_i^2y_i$
-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

$$\left.\begin{array}{l} \bar{y} = a + b\bar{x} \\ a_{11} = a\bar{x} + ba_{20} \end{array}\right\} \quad \begin{array}{l} a = 4 \\ b = 0 \end{array}$$

$$\hat{y} = a + bx \quad ; \quad \hat{y} = 4 \qquad S_{xy} = r = r^2 = 0$$

Ikusten denez, doitzen dugun zuzena, OX ardatzarekin paralela da eta edozein x-entzat y-ren balioa 4 da beti. Aldagaiak ezkorrelatuak dira, hau da, independentzia lineala agertzen zaigun kasu baten aurrean gaude. Beste aldetik, X aldagaiak Y-ren aldaketak ezertxo ere ez dituela azaltzen esan dezakegu.

Parabola doituz:

$\bar{y} = b_0 + b_1\bar{x} + b_2 a_{20}$ $\qquad 4 = b_0 + 2b_2$

$a_{11} = b_0\bar{x} + b_1 a_{20} + b_2 a_{30}$ $\qquad 0 = 2b_1$

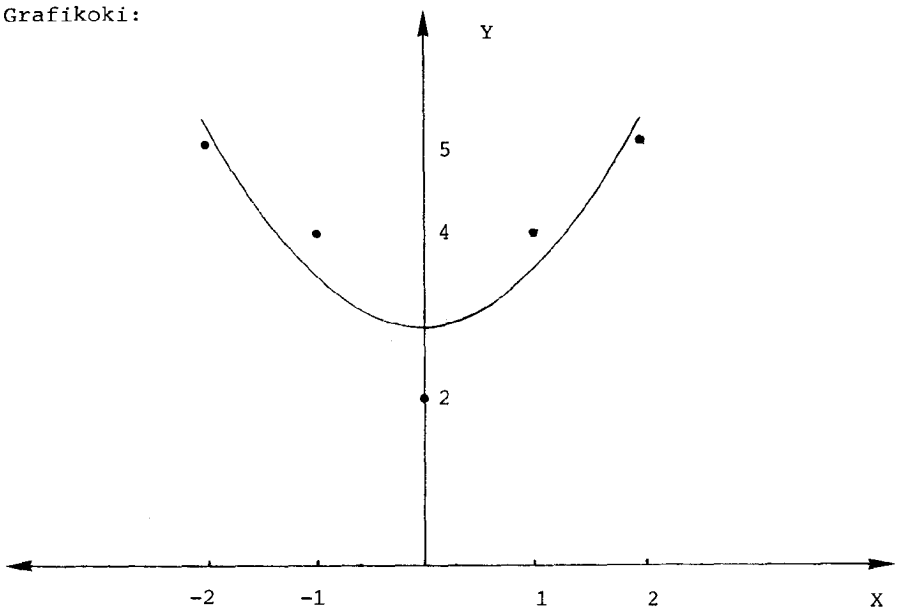
$a_{21} = b_0 a_{20} + b_1 a_{30} + b_{20} a_{40}$ $\qquad 40 = 10b_0 + 34b_2$

$b_1 = 0 \qquad b_0 = 20/7 \qquad b_2 = 4/7$

eta $\hat{y} = b_0 + b_1 x + b_2 x^2$ parabolan ordezkatuz

$$\hat{y} = 20/7 + 4/7\ x^2$$

Grafikoki:



- Puntu-hodeia eta doitutako parabola -

A R I K E T A K

1) Suposatzen badugu: $y = 0'90x + 0'40$ Y-ren X-ekiko erregresio zuzena dela eta $S_x = 0'48$, $S_y = 0'60$, $\bar{x} = 30$ direla, eskatzen da:

   1. Korrelazio-koefizientea.
   2. X-ren Y-ekiko erregresio-zuzena.

2) Ondoko datuak 1.970....1.973 urteetako jaiotzak/1.000 biztanle izanik,

Urteak	1.970	1.971	1.972	1.973
Jaiotzak/1.000 biztanle	59	64	69	74

   galdetzen da:

   1. Erregresio-zuzen baten bidez, zenbat jaiotza aurresango dugu 1.983 urterako?
   2. Zein da korrelazio-koefizientearen balioa? Zer azaltzen digu balio honek?

3) Ondoko taulak x ordu igarotzean kultibo batetan bolumen-unitateko landu diren bakterioak ematen dizkigu.

Orduak	0	1	2	3	4	5
Bakt./bolumen-unit.	12	19	23	34	56	63

   1. Zenbat ordutan edukiko dugu 100 bakterio?
   2. Zenbat bakterio edukiko dugu 6 ordutan?

4) Bost urtetan, hurrenez hurren eta mila pezetatan, ondoko sarrerak eta gastuak izan baditugu,

Sarrerak:	35	40	50	70	80
Gastuak :	33	37	40	52	60

eskatzen da:

1. Korrelazio-koefizientea, beraren esanahia azalduz.
2. Aldagai baten erregresio - zuzena.
3. Zertarako balio du aurreko kalkuluak?

5) X eta Y bi aldagai ezkorrelatu izanik, froga ezazu ondoko - berdintasuna:

$$S^2_{(x + y)} = S^2_x + S^2_y$$

6) X,Y aldagaietarako ondoko datuak emanez

X	0	1	2	3
Y	3	0	2	4

Eskatzen da:

1. Karratu txikienen metodoa erabiliz bigarren graduko para bola doitzea.
2. Hondar-bariantza eta doikuntzaren egokitasunaren ikasketa.
3. Zein izango litzateke Y aldagaiaren bariantzaren mugatasunaren portzentaia, doikuntza lineala egingo bagenu?

7) 1. Froga ezazu R^2 planoan bi erregresio-ekuazioak ekuazio - berbera direla, datuak eredu linealari ezin hobeki doi - tzen zaizkionean.
2. Froga ezazu, halaber, bi erregresio - zuzenak perpendikularrak direla X, Y aldagaiak ezkorrelatuak direnean.

8) (X,Y) aldagai bikotzaren hogei balio ohartuak banatuta digu gu ondoko taulan.

X \ Y	1	2	3
4	5	2	0
5	0	5	1
6	2	0	5

1. Lor itzazu X aldagariaren Y-rekiko eta Y aldagaiaren - X-ekiko batezbestekoen erregresio - lerroak.

2. Adieraz itzazu grafikoki.

9) Enpresa batetan, egiten diren piezen diametroak eta pisuak neurtzen dira.

Ondoko 5 piezetarako neurketak dauzkagu, eta seigarren piezaren diametroa 4 m. 10 mm. izan dela dakigu, baina pisua - ren neurria galdu zaigu.

Pieza	Diametroa	Pisua
1.	3m 995 mm	114 kg.
2.	4m 0 mm	118 kg.
3.	4m 5 mm	120 kg.
4.	4m 20 mm	122 kg.
5.	4m 30 mm	126 kg.

1. Zein izango da aurresan daiteken pisua , dauzkagun datuen bidez, karratu txikieneko erregresio lineala egiten badu gu?

2. Eman ezazu doikuntza honen neurriren bat, zenbateko ona den esateko.

10) Hamar ikaslek M-I eta T-I asignaturetan atera dituzten notak, ondokoak dira.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
M-I :	9	7'5	3	5	7	6'5	4'5	2	8	7'5
T-I :	6'5	8	4	4'5	6'5	6	4	3	7'5	8

Kalkula itzazu korrelazio - koefizientea eta koefiziente mugatzailea.

Zein proportziotan azalduko litzateke bariantza erregresio linealaren kasuan.

11) 1. Lor ezazu, arrazonatuz, hondar-bariantzaren adierazpena Y-ren X-ekiko erregresio linealaren kasuan.

2. Erlaziona ezazu, halaber, hondar - bariantza eta koefiziente mugatzailea.

12) GREENFIELD HISPANIA enpresa TK-32 eultzikailu (trilladora) merkaturatzen ari da, azkeneko hiru urteetan, 200 salmenta agenteren bidez.

Zuzendaritzak, agenteen gaitasun profesionalaren aurreratzea aztertu nahiean, ondoko maiztasun- taula bikoitza lortu du.

		Urtean zehar lortutako kontratuak			
		2	3	4	5
Enpresan daramatzaten urteak	1	10	20	10	0
	2	10	10	60	20
	3	0	10	30	10

Eskatzen da:

1. Bi aldagaien arteko korrelazio-koefizientea.
2. Bi aldagaien arteko erlazio-maila adieraztea.
3. Karratu txikieneko bi erregresio-zuzenak lortzea eta - grafikoki adieraztea.
4. Lortutako kontratuen aldagaiak enpresan daramatzaten - urteen aldagaiekiko duen batezbestekoaren erregresioa lortzen eta grafikoki adieraztea.
5. Komenta itzazu laburki erregresio-mota bakoitzaren - abantailak.

VII . ERREGRESIO ANIZKOITZA

VII.1. SARRERA

VII.2. ERREGRESIO-PLANOA R^2-n

VII.3. ERREGRESIO-HIPERPLANOA R^3-n

3.1. KARRATU TXIKIENEN ERREGRESIONAREN PROPIETATEAK.

3.2. KOEFIZIENTE MUGATZAILEA ETA KORRELAZIO-KOEFIZIENTE ANIZKOITZA.

VII.4. "BETA" ERREGRESIO KOEFIZIENTE STANDARDIZATUAK

ARIKETAK

VII.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 deitutako 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 korrelazio-koefizientea zen.

Korrelazio 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 korrelazio-koefizientea zen eta hondar-bariantza bat geratzen zen.

Nahiz eta, koefiziente mugatzaileak 1 balioari hurbiltzen zitzaion balio bat hartu, ezin genuen horregatik aldagaien artean kausa-efektu erlazio bat baieztatu, hau da, korrelazioa edukitzea,kasualitatez, izan daiteke zein argigarri diren beste kausa batzurengatik.

Baina, alderantziz,kausa- efektu erlazio bat denean, korrelazioa edukitzea probablea izango da; hau da, korrelazioak estatistikoki erlazioa frogatuko du.

Maiz, erlazio funtzionalak aztertzeko aldagai bakar bi gu txiegi dira errealitatean 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, errealitatearen sinplifikazio bat egiten ari gara.

Honela, haur baten pisua bere altueraren, adinaren, bularriguruaren etabarren menpe dagoela esan dezakegu; bai eta, ha laber, lur eremu baten emankortasuna bertako euriaren, hazi- kantitatearen, ongarriaren etabarren menpe dagoela.

Ikasgai honetan segituko dugun bidea, bereiztasun batzure kin, aurrekoarekiko 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, korrelazio-koefiziente anizkoitza emango dugu.

VII.2. ERREGRESIO-PLANOA R^3-n

Hiru aldagai kontsideratzen baditugu eta erregresio linea la egin nahian bat menpekotzat hartzen badugu, arazoa R^3-ko puntu-hodeian hobeki dotzen den planoa aukeratzea izango da.

Hau da: $\hat{x}_1 = b_{12}x_2 + b_{13}x_3 + a$ planoa finkatu behar dugu, desbidazio karratuen batura, batezbeste , txikiena izan da din.

Hots:

$$E = \sum_{ijk} (x_{1,i} - b_{12}x_{2,j} - b_{13}x_{3,k} - a)^2 f_{ijk} \text{ minimoa}$$

a(jatorriarekiko ordenatua) eta b_{12} , b_{13} (erregresio-koefizien teak) hiru alditan bilatuko ditugu E txikiena egin dezaten.

- Erregresio-planoa betezbesteko-bektoretik, hau da, hodeia ren 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:

$$\frac{\partial E}{\partial a} = -2 \sum_{ijk} (x_{1,i} - b_{12}x_{2,j} - b_{13}x_{3,k} - a)\ f_{ijk} = 0$$

$$\bar{x}_1 - b_{12}\bar{x}_2 - b_{13}\bar{x}_3 - a = 0$$

$$\boxed{\bar{x}_1 = b_{12}\bar{x}_2 + b_{13}\bar{x}_3 + a}$$

Aldagai zentratuetarako erregresio-koefizienteak:

Dakigunez, edozein aldagairi bere batezbestekoa kenduz ge ro, aldagai zentratu bat lortzen dugu.

Aldagai zentratuaren batezbestekoa zero da eta une hone - tan $\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = 0$ kontsideratuko dugu, frogaketa - errezago egin dezagun.

Erregresio-planoa, bada, (0,0,0) puntutik pasatzen da eta orduan $a = 0$ izanez:

$$E = \sum_{ijk} (x_{1,i} - b_{12}x_{2,j} - b_{13}x_{3,k})^2\ f_{ijk} \quad \text{minimoa}$$

Baldintza beharrezkoak idatziz:

$$\frac{\partial E}{\partial b_{12}} = -2 \sum_{ijk} (x_{1,i} - b_{12}x_{2,j} - b_{13}x_{3,k})\, x_{2,j}\, f_{ijk} = 0$$

$$\frac{\partial E}{\partial b_{13}} = -2 \sum_{ijk} (x_{1,i} - b_{12}x_{2,j} - b_{13}x_{3,k})\, x_{3,k}\, f_{ijk} = 0$$

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.

$$b_{12} = \frac{\begin{bmatrix} s_{12} & s_{23} \\ s_{13} & s_3^2 \end{bmatrix}}{\begin{bmatrix} s_2^2 & s_{23} \\ s_{23} & s_3^2 \end{bmatrix}} \qquad b_{13} = \frac{\begin{bmatrix} s_2^2 & s_{12} \\ s_{23} & s_{13} \end{bmatrix}}{\begin{bmatrix} s_2^2 & s_{23} \\ s_{23} & s_3^2 \end{bmatrix}}$$

Soluzioan aurkitzen diren determinanteak kobariantza-matrizekoak direnez, sinpleki aurkez daitezke erregresio - koefizienteak.

$$\begin{pmatrix} l_{11} & l_{12} & l_{13} \\ l_{12} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \qquad \begin{pmatrix} s_1^2 & s_{12} & s_{13} \\ s_{12} & s_2^2 & s_{23} \\ s_{13} & s_{23} & s_3^2 \end{pmatrix}$$

Hots:

$$b_{12} = \frac{-L_{12}}{L_{11}} \qquad b_{13} = \frac{-L_{13}}{L_{11}}$$

L_{ij}, l_{ij} elementuaren adjuntua izanik.

eta ekuazioan ordezkatuz:

$$\hat{x}_1 = -\frac{L_{12}}{L_{11}} x_2 - \frac{L_{13}}{L_{11}} x_3$$

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 ordez $X_1 - \bar{x}_1$, $X_2 - \bar{x}_2$, $X_3 - \bar{x}_3$ ipintzen baditugu, X_1, X_2, X_3 berri hauek zentratugabe izango dira eta planoaren ekuazio 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 kobariantza-matrizeaz eta batezbesteko-bektoreaz idatzi dugu.

Ondoko ekuazioa, ordez korrelazio-matrizeaz eta batezbesteko-bektoreaz.

$$(\hat{x}_1 - \bar{x}_1) = - \frac{S_1 R_{12}}{S_2 R_{11}} (x_2 - \bar{x}_2) - \frac{S_1 R_{13}}{S_3 R_{11}} (x_3 - \bar{x}_3)$$

$\bar{\bar{L}}$ eta $\bar{\bar{R}}$ matrizeen arteko elementuetan $S_{ij} = S_i S_j r_{ij}$ erlazioa betetzen dela ikusirik, ekuazio batetik erraz lortzen dugu bestea.

VII.3. ERREGRESIO-HIPERPLANOA R^n-n

$(X_1, X_2 \ldots\ldots X_i \ldots\ldots X_n)$ aldagaietatik X_i aldagaia beste n-1 aldagaien menpean kontsideratzen badugu, frogaketa baliokide - baten bidez, hiperplanoaren ekuazioa lortzen dugu.

Hots:

$$(\hat{x}_i - \bar{x}_i) = - \frac{L_{i1}}{L_{ii}} (x_1 - \bar{x}_1) - \ldots - \frac{L_{ik}}{L_{ii}} (x_k - \bar{x}_k) - \ldots\ldots - \frac{L_{in}}{L_{ii}} (x_n - \bar{x}_n)$$

$$b_{ik} = - \frac{L_{ik}}{L_{ii}}$$ erregresio-koefizienteak izanik.

Erregresio-hiperplanoa ondoko beste bi ekuazioez ere adieraz daiteke:

$$L_{i1} (x_1 - \bar{x}_1) + \ldots + L_{ii} (\hat{x}_i - \bar{x}_i) + \ldots + L_{ik} (x_k - \bar{x}_k) + \ldots +$$

$$+ L_{in} (x_n - \bar{x}_n) = 0$$

edo

$$\frac{R_{il}}{S_l} (x_1 - \bar{x}_1) + \ldots + \frac{R_{ii}}{S_i} (\hat{x}_i - \bar{x}_i) + \ldots + \frac{R_{ik}}{S_k} (x_k - \bar{x}_k) + \ldots . +$$

$$\frac{R_{in}}{S_n} (x_n - \bar{x}_n) = 0$$

Dakigunez, $R^2 - n$:

$$\bar{\bar{L}} = \begin{pmatrix} l_{xx} & l_{xy} \\ l_{xy} & l_{yy} \end{pmatrix} = \begin{pmatrix} S_x^2 & S_{xy} \\ S_{xy} & S_y^2 \end{pmatrix}$$

eta $b = - \frac{-S_{xy}}{S_x^2} = - \frac{L_{12}}{L_{22}}$

hots; erregresio-zuzena: $S_x^2 (\hat{y} - \bar{y}) - S_{xy} (x - \bar{x}) = 0$

3.1. KARRATU TXIKIENEN ERREGRESIOAREN PROPIETATEAK

Errore karratuen batezbestekoa deribatzean ateratzen ditugun ekuazioetan, inplizituki erregresioaren propietateak aurkitzen dira.

Hauekin batera, orain arte aipatu ez ditugun aldagai ezkorrelatuen bi propietate ikusiko ditugu.

1. Aldagai estatistiko ezkorrelatuen baturaren bariantza, aldagaien bariantzen batura da.

   Hots, $X_1, X_2 \ldots\ldots X_k$ binaka ezkorrelatuak badira, $S_{x_i x_j} = 0$ $\forall i \neq i$

$$S^2_{(x_1 + x_2 + \ldots + x_k)} = S^2_{x_1} + S^2_{x_2} + \ldots\ldots + S^2_{x_k}$$

2. Aldagai estatistiko bat beste batzurekiko ezkorrelatua baldin bada, beraien edozein konbinazio linealekiko ezkorrelatua da.

$X_1 , (X_2, X_3 \ldots\ldots X_k)$ $S_{x_1 x_i} = 0$ $\forall i \in \{2, \ldots, k\}$ baldin -

bada $S_{x_1, (ax_2 + bx_3 + \ldots + lx_k)} = 0$ / a, b, ..., l $\in$ R

E adierazpenaren a-rekiko deribatua egitean:

$$\bar{e} = \sum_{ijk} (x_{1,i} - b_{12}x_{2,j} - b_{13}x_{3,k} - a)\, f_{ijk} = 0$$

Hauxe da kolorario 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:

$$x_1 = \hat{x}_1 + e \qquad \bar{x}_1 = \overline{\hat{x}_1} + 0 \qquad \overline{\hat{x}_1} = \bar{x}_1$$

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:

$$\sum_{ijk} (x_{1,i} - b_{12}x_{2,j} - b_{13}x_{3,k} - a)\, x_{2,j}\, f_{ijk} = 0$$

$$\sum_{ijk} (x_{1,i} - b_{12}x_{2,j} - b_{13}x_{3,k} - a)\, x_{3,k}\, f_{ijk} = 0$$

hots,

$$\left.\begin{array}{l} \sum_{ijk} e\, x_{2,j}\, f_{ijk} = 0 \\ \sum_{ijk} e\, x_{3,k}\, f_{ijk} = 0 \end{array}\right\} \Rightarrow \begin{array}{l} s_{e,x_{2,j}} = 0 \\ s_{e,x_{3,k}} = 0 \end{array}$$

hau da, $\bar{e} = 0$ denez zera esan dezakegu:

iii) Hondarrak eta edozein aldagai independente edo erregreso re elkarren artean ezkorrelatuak dira.

Aipatu dugun aldagai ezkorrelatuen 2. propietatea kontutan hartuz, zera esan daiteke:

iv) Hondarrak eta balio estimatuak elkarren artean ezkorrela- tuak dira.

Berriro $x_1 = \hat{x}_1 + e$ dela kontsideratuz eta aldagai ezko rrelatuen 1. propietatea kontutan hartuz, korolario bezala, ba riantzaren deskonposaketa batukorra idatz dezakegu.

v) Balio estimatuen bariantzaren (erregresioan azaldurik da - goen bariantza) eta hondar-bariantzaren batura, balio errea len bariantza da.

$$s^2_{\hat{x}_1} + s^2_e = s^2_{x_1}$$

s^2_e , hondar-bariantza deitutakoa, $\bar{e} = 0$ izatean, errore - karratuen batezbestekoa da.

$$s_e^2 = (e - \bar{e})^2 f = \sum_{ijk} (x_{1,i} - b_{12}x_{2,j} - b_{13}x_{3,k})^2 f_{ijk}$$

Hau da, karratu txikienen metodoaz erregresioa egitean, -hondar-bariantza txikiena egiten ari gara.

Orain arte ikusitako bost korolarioak, eskematikoki honela labur ditzakegu.

$$\frac{\partial E}{\partial a} = 0 \longrightarrow \bar{e} = 0 \longrightarrow \hat{\bar{x}}_1 = \bar{x}_1$$

$$\frac{\partial E}{\partial b_{1i}} = 0 \longrightarrow s_{e,x_i} = 0 \longrightarrow s_{e,\hat{x}_1} = 0$$

$$\longrightarrow s_{\hat{x}_1}^2 + s_e^2 = s_{x_1}^2$$

Azkenez, kalkulu matrizialaren bidez, hondar-bariantzaren adierazpen orokorra lortzen dugu.

$$s_e^2 = \frac{L}{L_{11}} = \frac{R}{R_{11}} s_{x_1}^2$$

non L eta R, $\overset{=}{L}$ eta $\overset{=}{R}$ kobariantza eta korrelazio-matrizeen determinanteak baitira hurrenez hurren.

R^2-n aurreko ekuazioa honela konkretatzen da:

$$s_e^2 = \frac{s_x^2 s_y^2 - s_{xy}}{s_x^2} = s_y^2 - \frac{s_{xy}^2}{s_x^2} = s_y^2 (1 - r^2)$$

non: $s_x^2 s_y^2 - s_{xy} = L$ eta $s_x^2 = L_{22}$ (hau da, s_y^2 - ren adjuntua) diren.

3.2. KOEFIZIENTE MUGATZAILEA ETA KORRELAZIO-KOEFIZIENTE ANIZKOITZA

Bariantzaren deskonposaketa batukorra($s_{\hat{x}_1}^2 + s_e^2 = s_{x_1}^2$)

frogatu dugunez gero:

$$S^2_{\hat{x}_1} = S^2_{x_1} - \frac{L}{L_{11}} = \frac{1}{l_{11}} - \frac{L}{L_{11}} \text{ edo } = S^2_{x_1} - \frac{R}{R_{11}} S^2_{x_1}$$

Ondoko koefizienteak, azaldu behar den bariantzarekiko mugapenaren proportzioa ematen digu.

$$r^2_{1.2,3} = \frac{S^2_{\hat{x}_1}}{S_{x_1}} = 1 - \frac{L}{l_{11}L_{11}} = 1 - \frac{R}{R_{11}}$$

Eta beronen erro karratuari, K. Pearson-n koefizientearen analogiaz, korrelazio-koefiziente anizkoitza deituko diogu.

$$r_{1.2,3} = +\sqrt{1 - \frac{L}{l_{11}L_{11}}} = +\sqrt{1 - \frac{R}{R_{11}}}$$

Aurreko formulak, balio orokorra dauka, hau da, edozein n-ren kasurako baliagarria zaigu eta honela idatz dezakegu:

$$r_{1.23...n} = +\sqrt{1 - \frac{L}{l_{11}L_{11}}} = +\sqrt{1 - \frac{R}{R_{11}}}$$

X_1 delakoa, azaldu behar den aldagaia izanik.

R^n - n daukagun korrelazio-koefiziente anizkoitza eta - R^2 - n daukaguna, berbera da, zeinua salbu.

Hots,

$$r_{1.2} = \sqrt{1 - \frac{1 - r^2_{12}}{1}} = +\sqrt{r^2_{12}} = |r_{12}|$$

R^n-n beste koefiziente batzuk, "Beta" koefizienteek hain zuzen, erregresioan aldagai independenteek daukaten indarra - eta zeinua azalduko digute.

VII.4. "BETA" ERREGRESIO-KOEFIZIENTE STANDARDIZATUAK

Zergatik ez dauka zeinurik $r_{1.23...n}$ koefizienteak?

Galdera honi, honela erantzun behar diogu:

Alde batetik, koefiziente-mugatzailearen erro karratua bakarrik izatean, zeinua justifikatzeko ezer ez dagoelako; eta - bestaldik, aldagai independente batzuk funtzio 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: $S_1 = S_2 = 1$ 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.

Dakingunez, aldagaiak tipifikaturik daudenean, korrelazio matrizea kobariantza-matrize bera da, eta ondorioz erregresio ekuazioa:

$$\hat{T}_1 = - \frac{R_{12}}{R_{11}} T_2 - \frac{R_{13}}{R_{11}} T_2 - \ldots\ldots - \frac{R_{1n}}{R_{11}} T_n$$

Lehenengo erregresio-planoa edo hiperplano bera da, baina aldagaiak tipifikaturik hartzean, eskala berri batetan dauzkagu, hau da, R^n-n transformazio lineal bat egina daukagu,"tipifikazioa" hain zuzen.

Erregresio-koefiziente partzial tipifikatu edo standardizatu hauei, "β" letra grekoz deritzegu, korrelazio-matrizearen bidez kalkulatuz:

Hots:

$$\beta_{1i} = - \frac{R_{1i}}{R_{11}}$$

edo eta kobariantza-matrizearen bidez.

Hots:

$$\beta_{1i} = - \frac{S_i}{S_1} \frac{L_{1i}}{L_{11}} \qquad \text{edo} \quad \beta_{1i} = \frac{b_{1i}}{S_1/S_i}$$

Azkenez, koefiziente mugatzaile anizkoitzaren adierazpen matematikoan, R determinantea lehenengo lerroko adjuntuen bidez idazten badugu:

$$\frac{S^2_{\hat{x}_1}}{S^2_{x_1}} = r^2_{1.23...i...n} = 1 - \frac{R}{R_{11}} =$$

$$= 1 - \frac{R_{11} + r_{12} R_{12} + ... + r_{1i} R_{1i} .... r_{1n} R_{1n}}{R_{11}} =$$

$$= \beta_{12} r_{12} + ... + \beta_{1i} r_{1i} + ... \beta_{1n} r_{1n}$$

Azalduta dagoen bariantzaren proportzioa, batukorki descomposatuta lortzen dugu honela.

Hau da, era honetara ere doitzen dugun hiperplanoan, aldagai independenteek aldagai menpekoaren bariantza zein proportziotan azaltzen duten jakin dezakegu.

A R I K E T A K

1) 490 ikasle dituen lagin bati buruzko ikerketa bat egingo bi faktoreen bidez beraien emaitzak nola azaltzen diren ikus - dezagun.

X_1 delakoa lortu zituzten puntuak, X_2 adimen-koefizientea - eta X_3 ikasketa-orduak/asteko izango dira.

Eta: $\begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{bmatrix} = \begin{bmatrix} 18'5 \\ 100'6 \\ 24 \end{bmatrix}$ $S_1 = 11'2$ $r_{12} = 0'60$

$S_2 = 15'8$ $r_{13} = 0'32$

$S_3 = 6$ $r_{23} = 0'35$

Kalkula ezazu:

1. X_1 aldagaiaren X_2 eta X_3-ekiko erregresio-planoa.
2. Koefiziente mugatzailea eta korrelazio-koefiziente anizkoitza.
3. Emaitzen irazkina.

2) O.N.U. - ko urtekari estatistikotik ondoko datuak atera ditugu, 1.976 urtean harturik.

	x_1	x_2	x_3
Birmania	0'69	5	96
Bangla Desh	0'23	5	112
Sri Lanka	3'85	38	214
Indonesia	0'75	61	234
Thailandia	1'56	69	351
Malasia	2'93	322	825

x_1 aldagaia, ospitale-ohe/1000 biztanle

x_2 exportatutako balioa/biztanle (dolarretan) eta

x_3 Nazio-errenta erabilgarria (dolarretan) izanik.

Kalkula ezazu:

1. X_3-ren X_1 eta X_2- rekiko erregresio lineala.

2. Korrelazio-koefiziente anizkoitza.

3. Zein da aldagai independenteek X_3-ren bariantzatik azaltzen duten proportzioa?

4. Komenta itzazu laburki lortutako emaitzak.

3) Lagin aleatorio batetan bost familia hartu ditugu. Aurrezkiak sarrerak eta kapitalak neurtuz, ondoko datuak ditugu:

Fam.	S: Aurrezkia	Y: Sarrera	W: Kapitala
1	600 $	8.000 $	12.000 $
2	1.200	11.000	6.000
3	1.000	9.000	6.000
4	700	6.000	3.000
5	300	6.000	18.000

Kalkula ezazu:

1. S-ren, Y eta W-rekiko erregresio lineala.

2. Zein izango da familia baten aurrezkia, kapitala 5.000$ eta sarrera 8.000 $ izanik?.

3. Zein proportziotan goratuko da aurrezkia, kapitala konstantea izanik, sarrerak 2.000 $-tan goratzen badira?

4. Zer gerta daiteke, sarrerak 1.000 $-tan goratzen badira eta kapitala 3.000 $-tan?

5. Zein proportziotan azaltzen da edo eta azaldu gabe geratzen S-ren bariantza?

4) Aurreko ariketan daukagun lagin berbera kontsideratuz, - beste aldagai bat kontutan hatzen dugu: N, haurren kopurua, - hain zuzen.

Familia	N: Haurrak
1	5
2	2
3	1
4	3
5	4

Kalkula ezazu:

1. S-ren Y,W,N-ekiko hiperplano tipifikatua.
2. $r^2_{s.y,w,n}$ eta beronen esangura .
3. Komenta itzazu laburki lortutako emaitzak.

5) Dakigunez, $S_e^2 = \frac{L}{L_{ii}}$ adierazpena X_i aldagaiaren ($X_1, X_2 \ldots$ .... X_{i-1}, $X_{i+1}, \ldots . X_n$) aldagaiekiko karratu txikieneko erregresio linealaren hondar-bariantza da, non L kobariantza-matrizearen determinantea den.

1. Adierazpen hori kontutan hartuz, lor ezazu x_i aldagaiaren beste (n-1) aldagaiekiko korrelazio-koefiziente - anizkoitza.
2. Komenta ezazu laburki koefiziente horren esangura.
3. (X_1, X_2, X_3)aldagai hirukoitza kontsideratuz, froga ezazu $r_{1.23} = 1$ izatean $r_{2.13} = 1$ eta $r_{3.12} = 1$ direla.

VIII. KORRELAZIO PARTZIALA

VIII.1. SARRERA

VIII.2. KORRELAZIO PARTZIALA: KONTZEPTU MATEMATIKOA

VIII.3. KORRELAZIO-KOEFIENTE PARTZIALA ETA "β" KOEFIENTEAK

ARIKETAK

VIII.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 errealitateranzko hurbilpena bilatzean, aldagai bi baino gehiago kontsideratuko ditugu, korrelazio partzialak aztertuz.

Zertzu dira korrelazio-koefiziente partzialak?

Korrelazio-koefiziente partzialak hauexek dira:

Bi aldagairen arteko egiazko erlazioa neurtu nahi duten koefiziente batzu, beste aldagai guztiak konstanteak kontsideratuz edo beste modu batetara esanda, beste guztien eragina baztertuz.

Bi aldagairen arteko korrelazioa, besteak kontutan izan gabe, korrelazio totala deitzen da, baina, korrelazio 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 korrelazio positibo bat dela jakinik, kalkulatzean negatiboa ateratzen zaigu, hots, fenomenoa bi aldagaien artekoa balitz bezala hartu dugu, eta ez dugu beste aldagai batzuren eragina baztertu, korrelazioa 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	Tenperatura 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:

$$\hat{x}_1 = 60 - 1'67\ x_2 \qquad \text{eta} \qquad r_{12} = -\ 0'17$$

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:

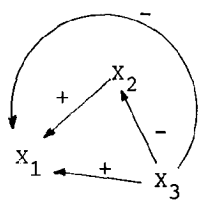
$$\hat{x}_1 = 60 + 5'71\ x_2 + 2'95\ x_3$$

Kasu honetan b = 5'71 da, erregresio-koefiziente positi - boa daukagu, hots, euri-kantitatearen eta produkzioaren erla - zioa 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 1970.urtean euri-kantitatea handia eta tenperatura baxua izan zen; ondorioz, uztabilketa txarra izan zen, baina ez x_2 balioagaitik, x_3-renagatik baizik.

X_1, X_2 aldagaien arteko erlazioa aztertu nahi badugu, b_{12} rregresio-koefiziente partziala lortu behar dugu edo x_3 aldagaiaren balioak konstanteak balira bezala, honen eragina baztertuz, korrelazio-koefiziente partziala lortu.

Datu hauetan: $r_{12.3} = 0'68$

Halaber (X_1,X_2,X_3)aldagaien elkarren arteko korrelazioak zeinu berkoak izanez X_1,X_2 aldagaien arteko korrelazioa ateratzean X_3-ren eragina kontutan hartzen ez badugu, handiegi den gezurrezko balio bat lortuko dugula ikusiko dugu.

Dakusagun, bada, beste adibide hau:

Ondoko datuak, $X_1X_2X_3$ aldagaiei dagozkie:

x_1: Hileroko famili gastu totalak.

x_2: Hileroko famili gastu totalak janaritan.

x_3: Hileroko famili sarrerak

X_1	X_2	X_3
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

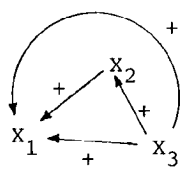
Argi dago x_1 eta x_2 aldagaien arteko korrelazioa positiboa izango dela; korrelazio totala kalkulatzean, emaitza, r_{12} =0'9356 da.

Baina bi aldagai horiek sarrerekin korrelazio positiboa daukate eta beraien balioak sarrera horien menpe bezala ikusdaitezke.

Orduan, egiazko korrelazioa ikusteko, X_3 aldagaiaren eratina baztertu behar dugu, hau konstante balitz bezala kontsideratuz.

Kalkuluak eginez: $r_{12.3}$ = 0'669 da.

Sinbolikoki



Zeharkako Erlazioa (+)

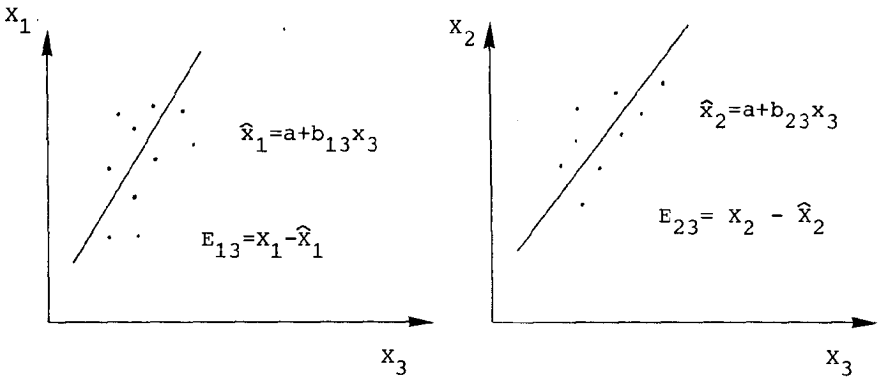
Korrelazio partziala kalkulatzean, zeharkako erlazioak duen eragin positiboa, baztertua izan da.

VIII.2. KORRELAZIO PARTZIALA : KONTZEPTU MATEMATIKOA

Demagun, $X_1X_2X_3$ hiru aldagai direla.

Nola baztertuko dugu X_1,X_2 aldagaien arteko korrelazioa 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 korrelazio-koefiziente partziala, E_{13} eta E_{23} aldagaien arteko korrelazio-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 hauen korrelazio-koefizientea.

$$r_{12.3} = \frac{S_{e_{13},\, e_{23}}}{(S^2_{e_{13}} \; S^2_{e_{23}})^{1/2}}$$

Hots, $\bar{e}_{13} = \bar{e}_{23} = 0$ eta $\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = 0$ izanik:

$$S_{e_{13}e_{23}} = \frac{1}{N} \sum \left(x_1 - \frac{S_{13}}{S^2_3} x_{3i}\right) \left(x_{2i} - \frac{S_{23}}{S^2_3} x_{3i}\right) =$$

$$= \frac{1}{N} \sum \left(x_1 \; x_{2i} - \frac{S_{13}}{S^2_3} x_{2i}x_{3i} - \frac{S_{23}}{S^2_3} x_1 \, x_{3i} + \frac{S_{13}S_{23}}{S^4_3} x^2_{3i}\right)$$

$$= S_{12} - \frac{S_{13}\, S_{23}}{S^2_3} - \frac{S_{13}\, S_{23}}{S^2_3} + \frac{S_{13}\, S_{23}}{S^4_3} S^2_3 =$$

$$= S_{12} - \frac{S_{13}\, S_{23}}{S^2_3}$$

Orduan:

$$r_{12.3} = \frac{S_{12} - \dfrac{S_{13}\, S_{23}}{S^2_3}}{\left(S^2_1 - \dfrac{S^2_{13}}{S^2_3}\right)^{1/2} \left(S^2_2 - \dfrac{S^2_{23}}{S^2_3}\right)^{1/2}}$$

Eta, zenbakitzailea eta izendatzailea, $S_1S_2 = \left(S^2_1\right)^{1/2} \left(S^2_2\right)^{1/2}$ balioaz zatituz.

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\left(1 - r^2_{13}\right)^{1/2} \left(1 - r^2_{23}\right)^{1/2}}$$

Korrelazio-matrizearen adjuntuen bidez idatziz:

$$r_{12.3} = - \frac{R_{12}}{(R_{11} \; R_{22})^{1/2}}$$

Ikusten dugu, bada, nola lortzen den X_1, X_2 aldagaien korrelazio partziala. Biek X_3-ren bidez dauzkaten estimazioak kontutan hartuz eta kondarren korrelazioa eginez, X_3 aldagaiarekin daukaten korrelazioa kontrolatu dugu.

Demagun orain, X_1 eta X_2 aldagaien arteko korrelazioa interesatzen zaigula, baina, (X_1 X_2.... X_k) aldagaiak aztertu nahi dugun fenomenoan sarturik.

Kasu honetan, X_3... X_k aldagaiak duten eragina baztertu behar dugu eta, analogikoki, X_1 eta X_2 aldagaiek hiperplanoen bidez dauzkaten estimazioak kontutan hartuz hondarren korrelazioa kalkulatuko dugu.

Hots:

$$E_{1.34...k} = x_1 - \hat{x}_1 = x_1 - a_1 - b_{13}x_3 - ... - b_{1k}x_k$$

$$E_{2.34...k} = x_2 - \hat{x}_2 = x_2 - a_2 - b_{23}x_3 - ... - b_{2k}x_k$$

Eta honela X_1, X_2 aldagaien arteko korrelazio-koefiziente partziala lehenengokoa dela frogatzen da.

Orduan:

$$r_{12.3...k} = - \frac{R_{12}}{(R_{11} \quad R_{12})^{1/2}}$$

VIII.3. KORRELAZIO-KOEFIZIENTE PARTZIALA ETA "β" KOEFIZIENTEAK

Sarreran esan dugun bezala, bi aldagaien arteko erlazioa erregresio-koefizienteen bidez ikusten da eta "β" erregresio-koefiziente partzial standardizatuek, bereziki, aldagai independente edo erregresore bakoitzak erregresioan duen indar erlatiboa eta zeinua adierazten digute.

Dakusagun, bada, β koefizienteen eta korrelazio-koefizien te partzialen arteko erlazioa:

Matematikoki:

$$\left|r_{12.3}\right| = - \frac{R_{12}}{(R_{11}R_{22})^{1/2}} = \left(\frac{-R_{12}}{R_{11}} \cdot \frac{-R_{12}}{R_{22}}\right)^{1/2} = (\beta_{12}\beta_{21})^{1/2}$$

Hots: $$\left|r_{12.3..k}\right| = (\beta_{12}\,\beta_{21})^{1/2}$$

Korrelazio-koefiziente partziala bi erregresio-koefizien-te standardizatuen batezbesteko geometrikoa da.

Lehenengo kasuan, hau da, X_3 aldagaiaren eragina baztertzen dugunean, korrelazio-matrizearen adjuntuak idatziz, honela izango da:

$$\left|r_{12.3}\right| = \frac{r_{12} - r_{13}\,r_{23}}{(1-r_{13}^2)^{1/2}(1-r_{23})^{1/2}} = \left(\frac{r_{12}-r_{13}\,r_{23}}{1 - r_{23}^2} \cdot \frac{r_{12}-r_{13}r_{23}}{1 - r_{13}^2}\right)^{1/2}$$

Berdin ikus daiteke, $X_3 \cdots X_k$ aldagaien eragina baztertzen dugunean, korrelazio-matrizearen adjuntuak idatziz.

A R I K E T A K

1) Ikasgaian datorren lehenengo adibiderako kalkula itzazu.

   1. X_1 aldagaiaren X_2,X_3-rekiko plano tipifikatua.

   2. Korrelazio total eta partzial guztiak.

   3. Komenta itzazu lortutako emaitzak.

2) Berdin bigarren adibidearen kasuan.

3) Ondoko sei urtetako segidetan, X_1,X_2,X_3 aldagaiak hartzen dituzten balioak dauzkagu:

X_1 :(Nazio-Produktu Gordina)	X_2: (Gasolina-kontsumoa)	X_3:(Elektrizitate-kontsumoa)
8	4	20
10	3	22
12	6	23
13	6	26
15	7	27
18	8	30

Eskatzen dira:

1. $r_{12.3}$,$r_{13.2}$,$r_{12.1}$

2. Gasolina-kontsumoaren bidez elektrizitate-kontsumoa konstantea izanik, N.P.G. estimatzean, kalkula ezazu batezbesteko errorea edo erregresio-zuzenarekiko desbidazio tipikoa.

3. Berdin X_1 estimatzean, X_3-ren bidez eta X_2 konstantea izanik.

4. Berdin X_2 estimatzean, X_3-ren bidez eta X_1 konstantea izanik.

5. X_1 aldagaiaren erregresio-planoa eta korrelazio-koefiziente anizkoitza.

4) Bost urtetako ondoko segidak emanez:

Urteak	Y	X_1	X_2	X_3
1.955	20	4	10	2
1.956	22	5	12	3
1.957	23	7	13	5
1.958	24	6	16	6
1.959	26	8	19	9
1.960	?	8	22	12

Non: Y = Nazio-Errenta, mila miloi pezetatan.

X_1= Nekazal sektoreko produktu gordina, mila miloi pezetatan.

X_2= Industri sektoreko produktu gordina,mila miloi pezetatan.

X_3= Hezkuntza sektoreko aurrekontua, mila miloi pezetatan.

Eskatzen dira:

1. Nazio-Errenta eta Nekazal produkzioaren arteko korrelazioa, industri produkzioa eta hezkuntza sektorea konstanteak izanik.

2. Nazio-Errenta eta Industri produkzioaren arteko korrelazioa nekazal sektorea eta hezkuntza sektorea konstanteak izanik.

3. Nazio-Errenta eta Hezkuntza sektorearen arteko korrelazioa X_1,X_2 konstanteak izanik.

4. Nekazal sektorearen bidez, X_2,X_3 konstanteak izanik, Nazio-Errenta aurresatean, kalkula ezazu batezbesteko errorea edo desbidazio tipikoa erregresio-zuzenarekiko.

5. Berdin X_2-ren bidez X_1, X_3 konstanteak izanik, Y aurresatean.

6. Berdin X_3-ren bidez X_1,X_2 konstanteak izanik,Y aurresatean.

7. Erregresio-planoa.

8. Y = N-E estimatua 1.960. urterako, beste aldagaien balioak taulakoak izanik.

9. Y eta (X_1,X_2,X_3) multzoaren arteko korrelazioa.

10. (X_1,X_2,X_3)ren bidez, N-E aurresatean, kalkula ezazu batez-besteko errorea.

5) Bizkaian egoitza sozialak dituzten ondoko bost enpresetako X_1,X_2,X_3 aldagaiek, 1979. eta 1980. urteetan balio hauek hartzen dituzte.

	1.979			1.980		
	X_1	X_2	X_3	X_1	X_2	X_3
IBERDUERO	7	6	+2	10	7	1'6
PETRONOR	120	11	+1'5	225	13	0'2
FIRESTONE	4	1'5	+0'1	4	2	0'1
ALTOS HORNOS	4	1'5	-0'8	4	1'5	-0'9
ECHEVARRIA	2	0'6	-1	3'6	1'4	-1

X_1 : "per capita"ko salmentak.

X_2 : "per capita"ko balio erantsia.

X_3 : "per capita"ko mozkina.

eta datu hurreratuak edo hurbilduak izanik, bi urte horietarako eskatzen dira:

1. X_3-ren X_1 X_2-ekiko erregresio-planoa.

2. Korrelazio-koefiziente anizkoitza.

3. $r^2_{3.12}$ eta beraren esaugura.

4. Korrelazio-koefiziente partzial guztiak.

5. Komenta itzazu lortutako emaitzak, urte batetik bestera ikusten duzun eboluzioa azpimarratuz.

6) X_1, X_2, X_3 aldagaien ondoko estatistikoak ezagutuz:

$$\bar{x} = \begin{bmatrix} 1 \\ 10 \\ 100 \end{bmatrix} \qquad \begin{matrix} S_1 = 1 \\ S_2 = 5 \\ S_3 = 20 \end{matrix} \qquad = \qquad R = \begin{bmatrix} 1 & 0'4 & -0'4 \\ 0'4 & 1 & 0'4 \\ -0'4 & 0'4 & 1 \end{bmatrix}$$

1. Lor itzazu X_3 aldagaiaren X_1, X_2 aldagaiekiko erregresio-planoa eta erregresio-plano tipifikatua.

2. Komenta itzazu lortutako b eta β koefizienteak, beraien arteko desberdintasunak adieraziz.

3. Lor ezazu $r_{12.3}$ korrelazio-koefizientea eta komenta ezazu beraren esaugura.

7) Froga itzazu karratu txikienen erregresio linealaren ondoko propietateak.

* Hondarrak edo erroreak zentratuak dira eta balio estimatu eta errealen batezbesteko aritmetikoa berdina da.

* Hondarrak eta balio estimatuak elkarren artean ezkorrelatuak dira.

* Balio estimatuen bariantzaren eta hondar-bariantzaren batura, balio errealen bariantza da.

8) Ondoko adierazpena kontutan hartuz:

$$r_{12.3} = - \frac{L_{12}}{L_{11}^{1/2} \; L_{22}^{1/2}}$$

1. Lor ezazu korrelazio horren adierazpena korrelazio-matrizearen adjuntuen bidez.

2. Erlaziona ezazu aurrean lortutako adierazpena β_{12} eta - β_{21} koefizienteekin.

9) Herri-multzo batetarako hiru aldagai sozioekonomiko aztertu ondoren, emaitza hauek lortu ditugu.

$$\bar{x} = \begin{bmatrix} 20 \\ 2 \\ 200 \end{bmatrix} \qquad L = \begin{bmatrix} 100 & 9 & -800 \\ 9 & 1 & -90 \\ -800 & -90 & 10000 \end{bmatrix}$$

Eskatzen da:

1. $\bar{\bar{R}}$ korrelazio-matrizea.

2. $\hat{x}_3$ $(x_1 x_2)$ erregresio-planoa.

3. $\hat{x}_3$ $(x_1 x_2)$ erregresio-planoa, hiru aldagaiak tipifikatuak direnean.

4. $r_{3.12}$ korrelazio-koefiziente anizkoitza.

5. $r_{12.3}$ korrelazio-koefiziente partziala.

6. Komenta itzazu laburki hiru aldagaien arteko erlazioak, lortu dituzun koefizienteen arabera.

10) Ondoko datu-matrizean, lau indibiduo eta X_1, X_2, X_3 alda - gaiei dagozkien datuak ditugu.

X_1	X_2	X_3
-1	+3	+6
-1	+3	+4
-1	+1	+2
+3	+1	0

Eskatzen da:

1. X_3 aldagaiaren X_1, X_2 -ekiko - karratu txikieneko erregresio - planoa.

2. Erregresio-plano berbera baina aldagaiak tipifikaturik dau denean.

3. Erregresio-koefiziente guztien irazkin edo komentario labu rra.

4. $r_{3.12}$ korrelazio-koefiziente anizkoitzaren balioa, r_{13} - eta r_{23} balioekin parekatuz.

5. Erregresio honen hondarrak edo erroreak. Zein izango da be raien balioen karratuen batezbestekoa?

6. $r_{12.3}$ korrelazio-koefiziente partziala eta gero r_{12} koe - fizientearekin parekatzea.

E R A N S K I N A

S.P.S.S. STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES

GIZARTE ZIENTZIETARAKO ESTATISTIKA-PAKETEA

1. SARRERA

2. INFORMAZIOAREN SARRERA

3. SPSS-RAKO DATUEN PRESTAKETA

   3.1. DATU BILKETA
   3.2. KASUA, ALDAGAIA ETA NEURRIA
   3.3. SARRERAKO DATUEN PRESTAKETA

4. SPSS PROGRAMAREN PRESTAKETA

   4.1. OROKORTASUNAK
   4.2. PROGRAMA OSATZEN DUTEN TXARTELAK
   4.3. KONTROL-TXARTELEN TANKERA

5. SPSS KONTROL-TXARTELAK

6. SPSS TRATAMENDU-TXARTELAK

   6.1. ERABIDE DEFINIZIORAKO TXARTELAK
   6.2. OPTIONS TXARTELA

7. ANALISI ESTATISTIKOAK LORTZEKO AZPIPROGRAMAK

   7.1. ARGIBIDE EDO INFORMAZIO-PROBLEMAK(Aldagai bat)
   7.2. IKERKETA PROBLEMAK (Aldagai batzu)

1. SARRERA

Guztiok ezagutzen dugu, azken urte hauetan konputagailu - ren erabilketak ekarri duen aurrerapena. Aurrerapen hau, bi ar lo desberdinetan ematen da, administralgoan eta zientziatan, - hain zuzen ere.

Egunero ohar gintezke administralgoan ematen denaz edo - zein bulegotan, banketxetan, aurrezki kutxatan e.a.

Berdin esan dezakegu zientzia arloan ematen denari buruz: gaur egun edozein zientziatan erabiltzen dira konputagailuak ikerketak egiterakoan.

Zientziatan duen erabilketan zentraturik eta, konkretuki Gizarte zientzetan duenean, hauxe azpimarratu behar dugu: Datu estatistikoak kalkulatzean (estatistiko edo balio tipiko deitu takoak) suposatzen duen abantaila erabatekoa dela.

Lehenago, ikertzaileek denbora asko pasatzen zuten datu - estatistikoak eskuz kalkulatzen, esate baterako, taula batzuk egitea izugarrizko lana izaten zen : gaur egun konputagailuren bidez segundu batzutan egiten dena aste batzutarako lana izaten zen.

Berdin esan dezakgu, aldagai estatistiko askoren erla - zioak ikertu nahi dituzten ikasketei buruz. Azken urte hauetan Estatistikaren adar batek, Datu-analisiak hain zuzen, garapen handia eduki du konputagailuen erabilketari esker, hauen bidez aldagai asko batera aztertzen baitira.

Baina, lehenbiziko urteetan programatzaile trebeak behar genituen konputagailuen bidez emaitzak ateratzeko. Arazo hau - konpontzeko,programa-paketeak lortu dira; eta gaur egun len - goaia erraz baten bidez, horiekek erabil ditzakegu, estatistikoak kalkula ahal izateko.

Gizarte zientzetan ikertzaileek behar dituzten kalkulu estatistikoak egiteko pentsatuta dagoen programa-pakete horietariko bat, SPSS deiturikoa da.

Dakusagun, bada, nota txiki hauetan SPSS-ri dagokion lengoaia berezia.

Irakurlea estatistika-paketeen erabiltzaileak dituen abantailetaz kontura dadin, hurrengo adibidea kontsideratuko dugu.

Demagun, N balio dituen segida baten batezbesteko aritmetikoa atera nahi dugula. Ondoan, arlo zientifikoan erabiltzen diren lengoaiarik errezanetariko batean, BASIC lengoaian hain zuzen, baharko genukeen programa daukagu.

```
10:  "A" INPUT  "DATUAK" ; N
20:  LET  T = 0
30:  FOR  I= 1 TO  N
40:  INPUT  "X=  "; X
50:  LET  T = T + X
60:  NEXT  I
70:  LET  M = T/N
80:  PRINT  "BATEZBESTEKOA = " ; M
90:  END
```

Aldiz, SPSS lengoaian programa hauxe izango litzateke:

```
1.
 DATA LIST
 CONDESCRIPTIVE
 STATISTICS
 READ INPUT DATA
 ======
 ======
 ======   Datuak
 ======
 ======
 ======
 END INPUT DATA
 FINISH
```

```
16.
  FIXED/ 1  ALDAGAIA 1-6
  ALDAGAIA
  ALL
```

Eta agindu hauen bidez, ez bakarrik batezbesteko aritmetikoa, baizik eta bariantza, desbidazio standarda, balio handiena, balio txikiena, ibilera, kurtosi koefizientea eta asimetri koefizientea lortuko ditugu.

Ikusten dugunez, estatistika-pakete bat erabiltzean ez ditugu estatistikoen formulak lengoaia baten bidez banan banan garatu behar. (Ikus adibidean, nola BASIC lengoaian batezbesteko aritmetikoa lortzeko daukagun formula garatu behar izan dugun, baina ez SPSS-n).

Konputagailuan SPSS sartuta dugunez gero, eragiketa guzti horiek batera egiteko azpiprograma prestatua daukagu , hau da, erabiltzaileari estatistikoak lortzeko bidea asko erraztu - zaio.

2. INFORMAZIOAREN SARRERA

Konputagailua, bere zirkuitu elektronikoen bidez eraikitako lanak agiteko prestatua izan da.

Konputagailu bakoitzak, dituen lengoaietan emandako agindu batzuren bidez, giza-erabiltzaile batek nahi dituen lanak egiten ditu;konputagailuari agindu-sekuentzia egokia emanez gero, eskuz egiteko neketsu gertatuko lirateken kalkuluak, segundu batzutan egiten ditu.

Agindu-sekuentzia edo agindu-segida hori, programa deritzoguna da.

Ordinatorea, gailu zentral batetaz (Prozesuaren unitate - zentrala) eta inguraldea eratzen duten beste gailu batzutaz (pantailak, txartel zulatuen irkurgailua, inprimagailuak) osatuta dago.

Gailu honetariko batzuren bidez informazioaren sarrera - egingo da; lehenbizi, programa eta datuak irakur araziko dizkiogu, gero konputagailuak agindu guztiak beteko ditu, eta az kenik nahi genituen emaitzen zerrenda lortuko dugu.

Informazioaren sarrera egiteko ohizko euskarrietariko bat txartel zulatua da, horregatik inguaraldeko gailuen artean - txartelen irakurgailua edukiko dugu.

Txartel bakoitzeko goiko atalean, agindua (programa-txartela denean) edo datuak (datu-txartela denean) inprimatuta - agertuko dira, baina, konputagailuak zuloak irakurriko ditu.

Txartel bakoitzean 80 karaktere sar daitezke, bakoitza zu tabe batean sarturik; zutabe bakoitzean 12 posizio dauzkagu - eta beraietan zuloak egin daiztezke.

Zulatzen ditugun karaktereak, zifrak, letrak edo ikurrak izan daitezke.

Zuloak egiteko makina berezi batzuk ditugu, zulagailuak - deritzogunak hain zuzen. Era berean, txartelen irakurgailua be harko dugu txartel zulatuak itzuli ahal izateko. Horretarako - irakurgailuak zutabe bakoitzetako zuloen esahahia itzuli eta zinta edo diskotan "idazten" du, tarte elektroniko egokietan.

Bestalde, programa baten emaitzak inprimatuta lortzea - ohizko gauza da. Horregatik inguraldean inprimagailuak edukiko ditugu.

3. S.P.S.S.-rako DATUEN PRESTAKETA

3.1. DATU BILKETA

Datu enpirikoak, ezaugarri edo gertaeren oharpenen bidez erregistratutako gertakizunak dira.

Oharpen bakoitza ekintza sinplea edo nahiko konplexua - izan daiteke; adibidez, famili baten seme-alabak zenbatzea e - kintza sinplea izango litzateke, baina ez herri baten N.P.G. - (Nazio-Produktu Gordina) zehaztea.

Gizarte zientzietan ikertzaileak erabil ditzaken datuak, iturri desberdinetatik erator daitezke: Inkesta publiko edo - pribatuetatik, hauteskundeen emaitzetatik, zentsu-estatistiketatik, finantza-estatistiketatik e.a.

"Datuak" eta "Programa estatistikoa" argi desberdintzea , garrantzi handikoa da. Programa batek, era egoki batean emanda ko edozein datu prozesatuko du.

Datuak eta Programa,bada,elkarren artean independenteak - dira; datuak erabiltzaileak lorturikoak eta prestaturikoak - izango dira.

3.2. KASUA, ALDAGAIA ETA NEURRIA

Analisiaren oinarrizko unitatea, Kasua da; estatistika - deskribatzailean :indibiduo estatistikoa" deritzoguna da. Edozein izaki edo erakunde izango da eta beraien gain neurriak - hartuko dira.

Aldagaia ohar edo beha daiteken ezaugarri bat da. Alda - gaiak balio bat edo batzu har ditzake eta datu-bilketan normalki bat baino gehiago neurtuko ditugu.

Oharpen baten erregistrapena neurria da; fenomeno batek - hartzen duen balioa izango da. Neurria zenbaki bat izan daiteke (Hiri baten populazioa,herri baten N.P.G., enpresa baten - salmentak e.a.) baina bai aldagai kualitatibo bati egokitu diogun kodifikazio baten balio zenbakizkoa edo alfanumerikoa ere (sexua, kokapena e.a.)

Datu guztiak (kasu guztien gain) ez ditugunean, faltatzen zaizkigun datuak, datu-eza izango dira, ingelesez MISSING DATA deitzen direnak.

Aldagai horren analisia egitean kasu horiek utzi egin behar dira.

3.3. SARRERAKO DATUEN PRESTAKETA

Dakigunez, txartelek 80 zutabe dituzte eta txartel bakoitzean gure programaren lerro bat zulatuko da ala kasu baterako ditugun neurriak.

Datuak, hutsunerik gabe zutabetan zula daitezke 80 zutabeak ongi bete nahi izanez gero, hala ere, sartuko ez balira , kasu bakoitzerako txartel bat baino gehiago beharko genuke.

Aldagaiak zutabe finkoen tankeraz zulatuko dira, hau da , aldagai baten balioak beti kasu bakoitzerako zutabe berdinetan zulatuko dira. Kasu bakoitzak txartel bat baino gehiago behar duenean, aldagai baten balioak beti txartel berdinean zulatuko dira, eta noski aldagai baten balioa ezin da txartel batetik bestera segitu; balio bakoitzak txartel bakar batean egon behar du.

Kasu bakoitzaren txartelak sakuentzian kokatu behar dira, baina ez nahitaez kasu desberdinak.

Azkenez esango dugu, kasuak identifikatzeko letra edo zenbaki batzu zula daitezkeela eta kasu bakoitzerako sekuentziaren zenbakia.

Orduan, kasu bakoitzerako txartel bat gutxienez eta beti aldagaien neurriak zutabe berdinetan.

4. S.P.S.S. PROGRAMAREN PRESTAKETA

4.1. OROKORTASUNAK

Izenak: Aldagai bakoitzari izen bat dagokio SPSS-n eta gehienez 8 karaktere edukiko ditu, lehenengoa nahitaez alfabetikoa izanik.

Adibidez:
ALD23 ⟶ Ongi
23ALD ⟶ Gaizki
ALDAGAIA ⟶ Ongi
ALDAGAIA23 ⟶ Gaizki

Zenbaki hamartarrak: Aldagai batek hartzen dituen balioetan zenbaki hamartarrak dituenean, bi era desberdinez joka dezakegu.

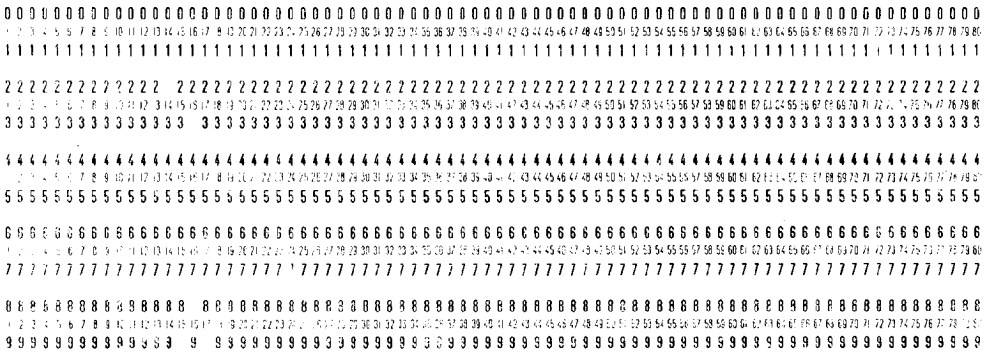
1. Datu-txarteletan puntu hamartarra zulatuz, puntu hori zutabe finko batean zulatua izanik.
2. Datu-txarteletan puntu hamartarra zulatu gabe utziz, hau da, zenbakiak osoak balira bezala kontsideratuz, kasu honetan programako DATA LIST aginduan zenbaki hamartarrak zeintzu diren esan beharko dugu.

Demagun adibidez, 29.9 balioa zulatu behar dugula eta aldagaiaren balioak konkretuki (00.0,99.9) tartean izan daitezkeela.

Puntu hamartarra zulatuz jokatzen budugu, aldagai horren balioak lau zutabe finkoren tankeraz zulatuko dira.

Hots, konkretuki:

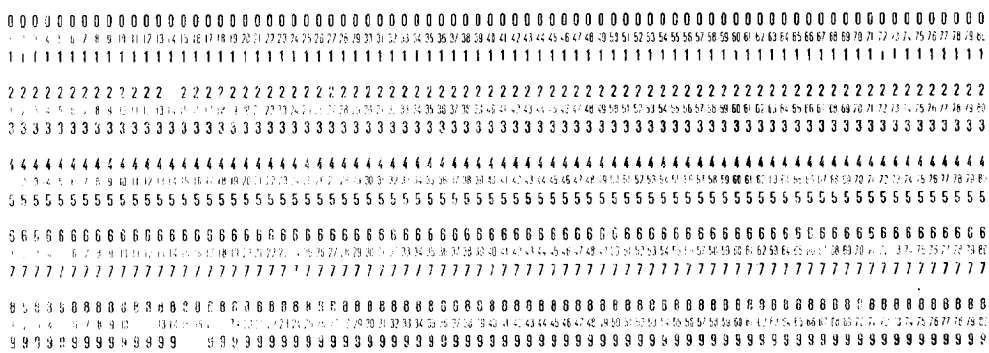
29.9



Puntu hamartarra zulatu gabe utziz jokatzen badugu,alda - gai horren balioak hiru zutabe finkoren tankeraz zulatuko dira.

Hots:

249



Geroago, DATA LIST agindua ikastean, konputagailuari , az keneko zenbakia hamartarra dela nola jakinarazten diogun ikusi ko dugu.

Demagun, orain, aldagai horren 3.5 balioa zulatu behar dugula.

Puntu hamartarra zulatuz, 03.5 edo 3.5 zulatuko genuke eta zulatu gabe utziz, 035 edo 35, zeren konputagailuak zuri (zulo gabe) geratzen diren tokiak 0 zulatua balute bezala ulertuko baitu.

Datu-eza: Batzutan, aldagaiak ez du baliorik hartzen kasu bat edo batzuren gain, datua ezezaguna delako edo inkesta baten datuak izanik, " ez daki / ez du erantzuten = ED/ EE" eran agertzen delako; hau da, datu-eza daukagu: SPSS-n MISSING VALues.

Dakusagun nola jokatuko dugun:

Atributu kasutan beste klase bat bezala kodifikatuko dugu adibidez: Emakumezkoa = 1, Gizonezkoa = 2 ED/EE =3.

Balioak zenbakizkoak balira, datu-ezak ezin ditugu besterik gabe zuri (zulatu gabe) utzi, konputagailuak zero balira bezala ulertuko lituzke eta.

Arazo hau, aurrerago ikusiko dugun MISSING VALUES agindua ren bidez konponduko da.

4.2. PROGRAMA OSATZEN DUTEN TXARTELAK

Esan genuen bezala, datuak eta programa ez ditugu nahastu behar, hots, datu-txartelak eta programa-txartelak ditugu; azkeneko hauetan berriz, kontrol-txartelak eta tratamendu-txartelak ditugu. Horrez gainera, sarrera-txartelak eta irteera-txartelak;hauek ez dira SPSS-renak sistemarenak baizik.

Sarrera-txartelen bidez, konputagailuak badaki SPSS pro - grama bat irakurri behar duela; irteera-txartelen bidez, ordea informazioa amaitu dela eta agindutakoa bete behar duela.

1. Atalean idatzi genuen adibidean SPSS txartel-mota desberdinak ikusten ditugu:

Bi lehenengoak kontrol-txartelak dira eta ondoan doazen - beste biak, tratamendu-txartelak (hauetan tratamendu estatistiko bat eskatzen da zehazki).

Ikusten dugu, ostera, eskatu dugun tratamenduen atzetik, datuak segitzen dutela, READ INPUT DATA aurrean dutelarik.

4.3. KONTROL-TXARTELEN TANKERA

Kontrol-txartelak funtzio desberdinak betetzen dituzte - SPSS-n.

Batzuk datuak deskribatzen eta definitzen dituzke, beste batzuk erabiltzen diren estatistikoak kontrolatzen dituzte,beste batzuk datuak aldatzen dituzte.

Guztiek bi zati edo arlo desberdinak dituzte.

Kontrolerako arloa eta zehaztapenen arloa, hain zuzen.

Kontrolerako arloa : 15 lehenengo zutabeek osatzen dute ; arlo honetan hitz batzuren bidez konputagailuak egin behar dituen ekintzak eskatzen dira.

Zehazkapenen arloa : beste zutabeek osatzen dute, hots, 16. zutabetik 80. zutaberaino. Arlo honetan ekintza bakoitzerako behar ditugun zehaztapenak emango ditugu.

Lehenengo 15 zutabeak zuri utziz gero, zehaztapenak txartel batetik bestera segi daitezke.

Zehaztapenetan edukiko ditugu:

Izenak: Esan genuen bezala, gehienez 8 karaktere dituzte lehenengoa alfabetikoa izanik; hauen bidez aldagaiak definitzen ditugu.

Balioak: Zehaztapenen arloan edozein kausagatik agertzen diren aldagaien balio espezifikoei dagozkien kantitate edo kode alfanumerikoak dira.

Giltza-hitzak: SPSS-rako adierazpen berezia duten hitzak dira, adibidez : TO, BY, WITH .....

Etiketak: Emaitzekin batera inprimatuta agertzen direla - rik dokumentazioarako baliagarriak zaizkigu .

Mugatzaileak: Izenak, giltza-hitzak, etiketak eta balioak bereizten dituzten ikurrak dira. SPSS lengoaiaren puntuazio ikurrak dira. Arruntenak, zuria, koma, ( ) parentesiak, = berdin ikurra eta / barra direlarik.

5. S.P.S.S. KONTROL-TXARTELAK

RUN NAME: Agindu honen bidez, programaren izena gehienez 64 karaktere erabiliz idatz dezakegu. Gure zerrendako orrialde guztietan idatzita agertuko da.

Programaren lehenengo txartela izango da, baina ez da nahitaezkoa hautabidezkoa baizik.

	1.zutabea	16.zut.
TANKERA :	RUN NAME	Etiketa
ADIBIDEAK :	RUN NAME	ALOKAIRUEN BANAKETA
	RUN NAME	SIDEROMETALURGIAREN AZTERKETA.
	RUN NAME	BIZKAIARI BURUZKO IKASKETA

DATA LIST: Agindu honen bidez, aldagaiak definitzen ditugu, - bide batez datu-txarteletan bakoitzak betetzen dituen zutabeak azalduz.

1. 16.
TANKERA: DATA LIST FIXED (n) / m aldagaia(k) $\left\{\begin{matrix} a - b \\ c \end{matrix}\right\}$(f).

n: kasu bakoitzerako ditugun txartelak izango da.

Dakigunez, 80 zutabetan datuak sartzen ez direnean n=2 - izango da, era berean 160 zutabetan sartzen ez direnean n = 3 e.a.
n = 1 denean, ez dugu ezer idatzi beharrik.
m : n horietatik deskribatzen ari garena izango da, m beti idatzi behar da.

Aldagaiaren neurketa bakoitzak a zutabetik b zutaberaino hartuko du.

Parentesi barruan daukagun f, zenbaki arrunt bat edo A letra izan daiteke; lehenengo kasuan aldagaiak zenbat zifra ha - martar duen adieraziko du eta aldagaia alfanumerikoa (¡) dela bigarrenean.

(¡) Aldagai alfanumerikoa esaten dugunean, zenbakizkoa ez dela esan nahi dugu, hau da, alfabetikoa edo propioki alfanumerikoa izango da.

Aldagai batzuren balioak errepikatzen direnean, datu-txarteletan zutabe kopuru berdina hartuz, lehenengo eta azkeneko zutabea adieraziz gero, konputagailuak aldagai bakoitzari dagokion zutabe kopurua zuzenki banatuko dio.

Dakusagun adibide baten bidez:

Demagun inkesta batean hiru erantzun ditugula, eta erantzun bakoitzarentzat hiru zutabe hartzen ditugula.

Honela egingo genuke:

```
1.              16.
DATA LIST       FIXED /1 .... ITEN 6, ITEN 7, ITEN 8,16-24
```

edo

```
DATA LIST       FIXED /1 .... ITEN 6, TO ITEN 8 16-24
```

Beste adibide batzu:

```
DATA LIST       FIXED /1 ALOK 3-6,DEN 7, KAT 8
DATA LIST       FIXED /1 SAR79  3-8 LAN79  9-13 SAR80 ,
                14-19, LAN80  20-24,SARAL 25-28 (1) LANAL
                29-32 (1)
DATA LIST       FIXED /1 IZEN 1-3 (A) KOKAP 4, BIZT 5-10
                AZAL 11-13, AZHAZ 14-17, EUSK 18-20
                X1 TO X14  21-48, ALOG 49-51, Y1 TO Y4
                52-59, UCD 60, EE,AP,PC 61-66
```

RUN NAME aginduan deitutako hiru lanei DATA LIST hauek dagozkie. Bi azkenekoetan zehaztapenak ez dira guztiak (16 - 80) zutabeetan sartzen ; hau da, DATA LIST aginduak bi txartel behar ditu bigarren adibidean eta hiru hirugarrenengoan.

SPSS programa batean DATA LIST agindua nahitaezkoa izango da.

N OF CASES: N of cases aginduak, datu-txarteletan zenbat kasu diren azalduko du.

Zenbat kasu diren ezezagun denean, UNKNOWN hitzaz adieraz daiteke.

	1.	16.
TANKERA:	N OF CASES	n (kasuen kopurua) UNKNOWN
ADIBIDEA:	N OF CASES	50

ALOKAIRUEN BANAKETA deitutako lanean 50 langile izan ziren kasuak.

Ez da nahitaezko txartel bat; ipini gabe ere utz dezakegu eta orduan konputagailuak UNKNOWN ulertuko du.

Baina kasu horietan eta baita UNKNOWN txartel baten bidez zehazten dugun kasuetan, datuen atzean END INPUT DATA txartela ipini beharko da.

Hots:

	1.
TANKERA:	END INPUT DATA

Txartel hau, ordea, ez da beharrezkoa, zenbat kasu diren adierazten dugunean.

Ikusten dugunez, bada, N OF CASES eta END INPUT DATA kasu batzutan ipiniko ditugu eta beste batzutan ez, hau da, agindu baldintzazkoak izango dira.

MISSING VALUES: Ikusi genuenez aldagai bat edo batzuetan datu-ezak dauzkagunean, ezin ditugu besterik gabe datu-txarteletan zuri utzi, konputagailuak zero balira bezala ulertuko li tuzkeelako eta ondorioz emaitzak okerrak aterako lituzkeelako.

Arazo hau konpontzeko bi bide ditugu:

1. Ezezagun diren datuei balio arraro bat eman diezaiekegu balio hau, aldagaiak hartzen dituen balioetatik kanpo egongo - da.

Balio hori aldagai baten edo batzuren balio guztiak positiboak direnean, balio negatiboa izan daiteke.

Balio arraroa, elegia, datu-txarteletan zulatuko da, baina, konputagailuak aldagai baten neurketak hartzen dituenean datu-eza duten kasuak utzi egin behar ditu kalkuluak egiteko.
Horretarako MISSING VALUES agindua daukagu, hau da, agindu honen bidez konputagailuak balio ez eragingarriak zeintzu aldagaiek dituzten eta hauek zeintzu diren ba daki.

```
          1.                    16.
TANKERA : MISSING VALUES        Aldagaia(k)   (balioa) /....

ADIBIDEA: MISSING VALUES        SEXUA (3) / ADINA (-1) / ....
```

2. Datu ezezagunak ez ditugu zulatzen, hau da, kasu hone - tan, zuriune batzuk ditugu datu-txarteletan.

Zuri horiek (dakigunez, zero balira bezala eragiketak egiterakoan) RECODE aginduaren bidez balio arraro batetan bihurtuko ditugu.

Hori eginez gero 1. kasuan bezala gaude, hots, lehen beza la jokatuko dugu MISSING VALUES agindua emanez, hain zuzen ere.

RECODE : Agindu honen bidez kodea aldatzen dugu, hau da,bir kodifikatu egiten dugu.

Adibidez: Aldagai baten balioak taldeka hartu nahi ditugu nean talde bakoitzaren balio guztiak beraien ordezkari den kla se-ordezkari balioan bihurtuko ditugu.

Modu berean egingo dugu datu-txarteletan ditugun zuriak - balio arraro batean bihurtu nahi ditugunean, gero MISSING VA - LUES agindua erabil ahal izateko.

RECODE agindua erabiltzeko ondoko hitzak behar ditugu:

LOWEST, LO ..... Txikiena

HIWEST, HI ..... Handiena

THRU ..... Raino

ELSE ..... Beste edozein

BLANK ..... Zuria.

1. 16.
TANKERA: RECODE {Aldagaia edo aldagaiak} (balioen zerrenda= balio berria): .../

ADIBIDEAK: RECODE ALOK(LO THRU 1250=1125)(1251 THRU (1500=1375) (1501 THRU 1750=1625) (1751 THRU 2500 = 2375)

RECODE SAR79,SAR80 (LO THRU 1499=1400) (1500 THRU 1999=1750)(2000 THRU 2999=2500)(3000 THRU 4999=4000) (5000 THRU 7999=6500)(8000 THRU 11999=10000)/ LAN79,LAN80(LO THRU 249=200)(250 THRU 499=350)(500 THRU 999=850)(1000 THRU 2499=1500)(2500 THRU HI = 7000)

RECODE EUSK (LO THRU 25=6)(26 THRU 50=37) (51 THRU 75 =65)(ELSE = 90)

Ondoko adibideetan zuriak balio arraro batean bihurtzeko RECODE erabiltzen dugu, gero MISSING VALUES agindua joango da.

```
1.                  16.
RECODE              SAR79  TO LANAL (BLANK =999)
MISSING VALUES      SAR79 TO LANAL  (999)

RECODE              EUSK (BLANK = -1)
MISSING VALUES      EUSK (-1)
```

Lehenbiziko adibidean, aldagai guztietarako (ikus DATA LIST) ditugun zuriak, balio berberan bihurtu ditugu. Aldagaien zerrenda TO hitza erabiliz laburtu dugu, baina, zerrendako aldagai guztiak koma ikurraz bereiziz ipin genitzakeen.

*RECODE ipintzen badugu birkodifikazioa tratamendu estatistiko baterako bakarrik zaigu baliagarria. Hurrengoko tratamendu estatistikotarako datu-txarteletan sartu diren balioak , berehala hartuko ditu konputagailuak.

Agindu hau, hautabidezkoa da.

RECODE eta *RECODE aginduak dituzten mugapenak hauxek dira:

- RECODE, *RECODE AGIDUAREN AURRETIK JOAN BEHAR DA.

- RECODE lehenbiziko tratamendu estatistikoaren aurretik joan behar da.

-*RECODE aplikatzen den tratamendu estatistikoaren aurretik joan behar da.

VAR LABELS: SPSS-n aldagai bakoitzari etiketa bat egokitzeko bidea daukagu.

Etiketa hauen bidez erabiltzaileak aldagaien esanahia argiro dauka,lorturiko estatistikoak aztertzean.

Aldagaia agertzen den toki bakoitzean, ondoan etiketa agertuko da idatzita beti.

Etiketaren luzera, gehienez, 40 karakteretakoa izan daiteke.

	1.	16.
TANKERA:	VAR LABELS	Izena, etiketa /....
ADIBIDEAK:	VAR LABELS	ALOK,ALOKAIRUA/DEN,DENBORA/ KAT,KATEGORIA/
	VAR LABELS	SAR79 ,1979 SARRERAK/LAN79,1979 LANGILEAK/SAR80 ,1980SARRERAK/ LAN80 1980 LANGILEAK/SARAL,79-80 SARREREN ALDAKETAK/ LANAL,79-80 LANGILEEN ALDAKETAK/
	VAR LABELS	IZEN,IZENA/ KOKAP,KOKAPENA/ BIZT,BIZTANLEAK/ AZAL,AZALERA/ AZHAZ,AZKENEKO HAZKUNDEA/..... ...X1, GIZONEZKOAK..X14,AUTONOMOAK/.....

Agindu hau hautabidezkoa da.

READ INPUT DATA: Agindu honek konputagailuari datuak irakur erazten dizkio; datu-txartelen aurretik doa.

	1.
TANKERA:	READ INPUT DATA

Ikusten dugunez, zehaztapenen arloan ez dauka ezer.
Agindu hau nahitaezkoa da.

FINISH: SPSS programaren azkeneko agindua da, prozesuaren a - maiera adierazten du.

```
          1.
TANKERA:  FINISH
```

Aurrekoa bezalatsukoa da eta,noski, nahitaezkoa.

PRINT FORMATS: Aldagaiak alfanumerikoak edo zenbakizkoak iza - nik zenbaki hamartarrekin ditugunean, agindu hau erabiliko - dugu.

```
              1.                    16.
TANKERA:      PRINT FORMATS         [izena edo zerrenda]  (f)/....

ADIBIDEAK:    PRINT FORMATS         SARAL(1) /LANAL (1)

              PRINT FORMATS         IZEN (A)
```

Ikusten dugunez, (f) delakoa zenbaki bat edo A letra izan daiteke. Lehenbiziko kasuan aldagaiak zenbat zifra hamartar - duen eta bigarrenean alfanumerikoa dela adieraziko du.

Agindu hau, aldagai alfanumerikotarako bakarrik nahitaez koa da.

Aldagaien zerrenda batean zenbaki hamartar kopuru berbera edo A idatzi behar badugu, agindua koma ikurraren bidez edo TO hitzaren bidez labur dezakegu.

```
              1.                    16.
      HOTS:   PRINT FORMATS         SARAL,LANAL (1)

       edo    PRINT FORMATS         SARAL TO LANAL (1)
```

LIST CASES: Agindu honen bidez, konputagailuak aldagai bat - edo batzuren zerrenda ematen digu.

Zerrendak, aldagaien balioak tartekatuta emango dizkigu , puntu hamartarrak idatzita edukiko dituelarik. Lehenbiziko zu- tabean kasu bakoitzaren ordinala edukiko du.

	1.	16.
TANKERA:	LIST CASES	CASES =n/ VARIABLES=Aldagaien zerrenda
ADIBIDEA:	LIST CASES	CASES= 97/ VARIABLES=IZEN TO PC

agindu hau hautabidezkoa da.

COMMENT: Agindu honen bidez lerro batetako irazkinak edo komentarioak idatz daitezke.

	1.	16.
TANKERA:	COMMENT	Irazkina

Programa bakoitzean COMMENT agindu bat edo batzu eska ditzakegu.

Noski, hautabidezko agindua da.

DOCUMENT: Agindu honen bidez limiterik gabeko irazkin bat idatz daiteke.

	1.	16.
TANKERA:	DOCUMENT	Irazkina
ADIBIDEA:	DOCUMENT	ZAZPI ESKUALDE KONTSIDERATUKO DITUGU: 1 = ENKARTAZIOETA, 2= BILBO HANDIA,3= URIBE BUTROE, 4= BUSTURIA, 5= LEA ARTIBAI, 6= DURANGALDEA, 7= ARRATIA-NERBIOI

Ikusi ditugun kontrol-txartelak, ez dira, noski, SPSS paketeko guztiak; oso interesgarri diren asko ez dira apuntehauetan azaldu,hasieratik gure helburua ikasleei SPSS-aren oinarrizko ideia bat ematea eta beraiek programa txiki batzu egitera bultzatzea izan baita.

Utzi ditugun aginduen artean, aldagai berriak sortzeko diren batzu aipa ditzakegu:

CONPUTE,IF,COUNT aginduak benetan interesgarriak dira. Baita , kasu batzutan beraiekin erlazioanaturiko beste batzu ere: ASSING MISSING, DO REPEAT, FILE NAME, SAVE FILE, GET FILE e.a. agin - duak adibidez.

Erabiltzaileak (ikasleak kasu honetan) SPSS-aren ezaguera ren edozein gehitze nahiko balu, bere irakaslea edo SPSS eskuliburua kontsultatuz lortuko du.

6. S.P.S.S. TRATAMENDU-TXARTELAK

6.1. ERABIDE DEFINIZIORAKO TXARTELAK

Erabili nahi ditugun azpiprogramak hautatzeko eta aktibatzeko dira.

Beraien tankera eta edukia erabili nahi ditugun estatistikoen menpe egongo da.

Analisi estatistiko hauek eta dagozkien tankerak 7. ata - lean ikusiko ditugu.

6.2. OPTIONS TXARTELA

Erabiltzaileak txartel honen bidez aukerabide batzuren artean, "OPTIONS" hain zuzen, hautatzeko bidea dauka.

Tratamendu bakoitzerako aukerabide batzu ditugu konpuntagailuak kalkulu estatistikoak nahi dugun erara egin ditzan.

	1.	16.
TANKERA	OPTIONS	(aukerabideen zerrenda edo ALL (guztiak))

OPTIONS TXARTELA lehen eta STATISTICS TXARTELA gero, tratamendu-txartelaren atzean doaz.

7. ANALISI ESTATISTIKOAK LORTZEKO AZPIPROGRAMAK

7.1. ARGIBIDE EDO INFORMAZIO-PROBLEMAK (Aldagai bat)

Aldagai bakun baten estatistikoak lortzeko ditugun azpiprogramak deskribatuko ditugu lehenbizi:

FREQUENCIES: Agindu honen bidez, maiztasun-banaketen taulak lortuko ditugu.

Azpiprograma honek dituen aukerabideetatik 8.a aipatuko dugu, OPTIONS 8 hain zuzen. Honen bidez maiztasun-banaketari dagokion "histograma" (¡) lortuko dugu.

Aldagai bakunetarako ezagutzen ditugun estatistikoetatik, hauek ditugu: batezbesteko aritmetikoa, bariantza, desbidazio standarda (¡¡), ibilera, balio handiena, balio txikiena , moda mediana, asimetri eta kurtosi koefizienteak.

(¡) Ez da egiazko histograma bat; abzisen ardatzean balioak edo klase-ordezkariak heinak balira bezala harturik barrazko diagrama bat egiten du.

(¡¡) SPSS programak bariantza eta desbidazio standarda deitzen dituenak, estatistikan Quasi-bariantza eta Quasi-desbidazio standarda deitzen ditugunak dira.

Hots: $S_x^{*2} = \frac{(x_i - \bar{x})^2}{N - 1}$; $S_x^{*} = \frac{(x_i - \bar{x})^2}{N - 1}$

Bakoitzari zenbaki bat dagokio azpiprograma honetan eta STATISTICS aginduan zenbakia edo ALL (guztiak) idatziz zehaztu behar dugu.

```
         1.                  16.
TANKERA  FREQUENCIES         GENERAL=( Aldagaien zerrenda )
                                     ( edo ALL            )
```

```
ADIBIDEAK: FREQUENCIES     GENERAL= ALOK
           OPTIONS         8
           STATISTICS      ALL

           FREQUENCIES     GENERAL =SAR79  TO LAN80
           OPTIONS         8
           STATISTICS      ALL

           FREQUENCIES     GENERAL = LEK, EUSK
           OPTIONS         8
           STATISTICS      ALL
```

FREQUENCIES azpiprograma, aldagai estatistiko diskretutarako (kualitatibo nahiz kuantitatibo) edota aldagai estatistiko jarraitarako(lehen klasetan taldekatuz) erabiliko dugu.

Noski, tratamendu estatistiko hau edo beste edozein eskatzea hautabidezkoa da.

CONDESCRIPTIVE: klasetan taldekatu gabeko aldagai jarrai bat - daukagunean, edo diskretua izanik balioak errepikatu gabe kontsideratzen ditugunean, maiztasun-banaketen taulak ez zaizkigu interesatzen.

Orduan, CONDESCRIPTIVE agindua erabiliko dugu.

Azpiprograma honek dituen estatistikoak, moda eta mediana izan ezik, FREQUENCIES azpiprogramak dituenak dira.

Azpiprograma honek dituen aukerabideetariko hirugarrena - aipatuko dugu, OPTIONS 3 hain zuzen. Honen bidez aldagaiei dagozkien balio tipifikatuak lortuko ditugu.

```
          1.                    16.
TANKERA:  CONDESCRIPTIVE        aldagaien zerrenda
                                ALL

ADIBIDEAK: CONDESCRIPTIVE       ALOK
           STATISTICS           ALL

           CONDESCRIPTIVE       ALL
           STATISTICS           ALL

           CONDESCRIPTIVE       BIZ TO PC
           OPTIONS              3
           STATISTICS           ALL
```

BREAKDOWN: Agindu honen bidez, aldagai bat edo batzuren estatistikoak beste batzuren funtzioan geruzatuz lortuko ditugu.

Azpiprograma honek, aldagai baten batura, batezbesteko aritmetikoa, desbidazio standarda, bariantza eta kasuen kopurua kalkulatzen du, kasuen azpitaldeen artean.

Adibidez, estatistiko horiek emakumezkoentzako eta gizonezkoentzako edo gizarte-maila desberdinentzako batera errepikatu nahi ditugunean erabiliko dugu agindu hau.

```
          1.                    16.

TANKERA:  BREAKDOWN             TABLES= Aldagaien zerrenda BY aldag.
                                        zerr. BY..../
```

Aldagaien lehenbiziko zerrenda aldagai menpekoek (normalki kuantitatiboak) osatuko dute, horien estatistikoak kalkulatu nahi ditugularik. Beste zerrendak aldagai independenteek(- normalki kualitatiboak) osatuko dituzte eta beraiekiko geruzak hartu nahi ditugu.

```
          1.                    16.
ADIBIDEA: BREAKDOWN             TABLES = ALOK BY SEX BY KAT/
```

Demagun bi kategoria desberdin direla. SPSS-n lortuko genukeen taula, ondokoa izango litzateke.

	SUM	MEAN	STD DEV	VARIANCE	N
FOR ENTIRE POPULATION	---	----	-------	--------	(..)
SEX - - - - - - - -1	---	----	-------	--------	(..)
KAT - - - - - - 1	---	----	-------	--------	(..)
KAT - - - - - - 2	---	----	-------	--------	(..)
SEX - - - - - - - - 2	---	----	-------	--------	(..)
KAT - - - - - - 1	---	----	-------	--------	(..)
KAT - - - - - - 2	---	----	-------	--------	(..)

ADIBIDEAK:

1.	16.
BREAKDOWN	TABLES = ALOK BY KAT
BREAKDOWN	TABLES = BIZT TO PC BY KOKAP

Azpiprograma honek bi estatistiko ditu, Bariantz Analisia eta linealtasun test bat, baina, ez dizkiogu oraingoz konputagailuari eskatuko, estatistika deskrebatzailean ez baitira sartzen.

Aurrerago, aldagai bikoitz eta aldagai anizkoetarako ditugun tratamendu estatistikoak eta hauei dagozkien azpiprogramak ikusiko ditugu.

Orain, bada, dakusagun gure programaren txartelak gutxigorabehera jarraituko duten ordena .

RUN NAME	Hautabidezkoa
DATA LIST	Nahitaezkoa
N OF CASES	Baldintzazkoa
MISSING VALUES	Baldintzazkoa
RECODE	Hautabidezkoa
VAR LABELS	Hautabidezkoa

DOCUMENT ..... Hautabidezkoa
COMMENT ..... Hautabidezkoa
PRINT FORMATS ..... Baldintzazkoa
LIST CASES..... Hautabidezkoa
Lehenbiziko tratamendu est. Nahitaezkoa (baina edozein)
OPTIONS ..... Hautabidezkoa
STATISTICS ..... Hautabidezkoa
READ INPUT DATA ..... Nahitaezkoa
---] DATU-TXARTELAK
---]
---]
END INPUT DATA ..... Baldintzazkoa
Beste tratamendu estatistikoak.. Hautabidezkoak
FINISH ..... Nahitaezkoa

7.2. IKERKETA-PROBLEMAK (aldagai batzu)

CROSSTABS: Aldagai bakoitzaren banaketa aztertu ondoren, ikertzailea, normalki, bi edo aldagai gehiagoren arteko erlazioak bilatzen hasiko da.

CROSSTABS azpiprogramak 2-tik 10 dimentsiotarainoko tabulazio gurutzatu bat egingo du.

Lortzen ditugun taulak, maiztasun-banaketenak dira.

	1.	16.
TANKERA:	CROSSTABS	TABLES = Aldagaien zerr. BY aldag. zerr./
ADIBIDEAK:	CROSSTABS	TABLES = ALOK BY DEN/ALOK BY KAT DEN BY KAT/ ALOK BY DEN BY KAT
	CROSSTABS	TABLES = SAR79 BY LAN79 / SAR80 BY LAN80 / SAR79 BY SAR80 /LAN79 BY LAN80/ EUSK,X9 BY KOKAP/
	CROSSTABS	TABLES= EUSK,X9 BY KOKAP/EUSK BY X9/ EUSK BY X9 BY KOKAP

Bi dimentsiotako taulak lortuko ditugu, adibidez:(EUSK, KOKAP) (X9, KOKAP) (SAR79 , SAR80) e.a.

Baita, ALOK BY DEN BY KAT eta EUSK BY X9 BY KOKAP eskat - zean, hiru dimentsiotako taulak lortuko ditugu. Hauek (ALOK, - DEN) taulak izango dira kategoria desberdinetako eta (EUSK, X9) taulak KOKAP aldagaiaren balio desberdinetarako (¡).

Lortutako maiztasun-banaketak, STATISTICS batzuren bidez azter ditzakegu, baina hauek estatistika deskrebatzailetik kan pokoak dira.

PEARSON CORR: Dakigunez, bi aldagaien arteko elkar erlazioaren maila, korrelazio koefienteak ematen digu.

PEARSON CORR azpiprogramak, Karl Pearson-en r korrelazio-koefizienteak ematen dizkigu edo, aldagai guztien arteko korre lazio-koefizenteak eskatzen ditugunean, korrelazio-matrizea - emango digu.

	1.	16.
TANKERA :	PEARSON CORR	Aldagaien zerrenda[WITH aldag.zerr.]
ADIBIDEAK:	PEARSON CORR	EUSK WITH AZHAZ, X9
	PEARSON CORR	ALOK TO KAT
	PEARSON CORR	BIZT TO PC

Ikusten dugunez, korrelazio konkretu batzu nahi ditugunean "aldagaien zerrenda WITH aldagaien zerrenda" ipiniko dugu.Bai- na, lehenbiziko "aldagaien zerrenda" bakarrik aipatzen badugu aldagaien arteko korrelazio-matrizea lortuko dugu, (Azkeneko - bi adibideen kasua).

(¡) Oharra: CORSSTABS aldagai taldekatu edota diskretuetarako eskatzen dira.

Aldagai bat edo batzurentzako datu-ezak ditugunean lortzen duguna ez da egiazko korrelazio-matrize bat.

OPTIONS eta STATISTICS txartelen bidez aukerabide batzu - ditugu.

Konkretuki STATISTICS 1,2 eskatuz, batezbesteko aritmetikoak, biderkadura gurutzatuak, bariantza eta desbidazio stan - dardak lortuko ditugu.

SCATTERGRAM: Agindu honen bidez bi dementsiotako disper - tsio-diagrama edo puntu-hodeia lortuko dugu.

Abszisa eta ordenatuen ardatzetan bi aldagai definitzen direnez gero, konputagailuak, puntu edo zenbaki baten bidez bi - ardatzetan dauden balioei dagozkien maiztasun absolutuak adierazirik, puntu-hodeia ematen digu.

	1.	16.
TANKERA:	SCATTERGRAM	Aldag. zerre.[WITH aldag.zerr.]
ADIBIDEAK:	SCATTERGRAM	ALOK TO KAT
	SCATTERGRAM	IN79 TO LANAL
	SCATTERGRAM	BIZT WITH AZAL,AZHAZ,EUSK,X3 TO X6,X9 TO X12, ALOG TO Y4, UCD,EE, AP,PC/ AZAL WITH AZHAZ,EUSK, X5, x9,X12/

Ditugun aukerabideetariko hauek aipatuko ditugu:

OPTIONS 5 eta OPTIONS 7. Lehenengoaren bidez puntu-hodeiaren grafikan diagonala marraztuta agertuko da, honela honen - gainetik eta azpitik kokatzen diren puntuak ikusi ahal izateko

bigarrenaren bidez, ordea abszisen ardatzean, balio osoak agertuko dira.

STATISTICS ALL eskatuz, r, r^2, estimazio-errore standarda jatorriko ordenatua, erregresio-zuzenaren malda e.a. lortuko ditugu.

Orain arte ikusi ditugun azpiprogramen bidez analisi elementalenak egingo ditugu.

REGRESSION:

Dakigunez, erregresio anizkoitzaren bidez (bikoitza barruan sarturik menpeko edo dependente bezala hartzen dugun aldagai baten erlazioa aztertzen dugu beste bat edo batzurekiko.

SPSS programa-paketearen REGRESSION azpiprogramak, erregresio lineal anizkoitza egiten du.

Azpiprograma ahaltsu honek, erregresio linealaz gainera,-erregresio polinomikoak (parabolikoa, exponentziala e.a.) ere egiten ditu.

	1.	16.
TANKERA:	REGRESSION	VARIABLES = aldagaien zerrenda/
		REGRESSSION= diseinuaren agindua/

Aldagaien zerrendaren bidez erregresioan sartzen diren aldagaiak zehazten ditugu, ondoren, aldagai horiei dagokien korrelazio-matrizea egin ahal izateko.

Diseinuaren aginduaren bidez, ordea, korrelazio-matrizetik nahi dugun kalkulurako egokiak diren korrelazio-koefizienteak aukeratzen dira.

Diseinu honen bidez erregresio-eredua finkatzen dugu, hau da, aldagaietariko zein hartzen den menpekotzat eta zein edo zein tzu independentetzat.

Ikus ditzagun, bada, ditugun posibilitateak adibide batzuren bidez.

ADIBIDEAK:

1)
```
1.                  16.
REGRESSION          VARIABLES = AAL, BAL, FAL, ZAL /
                    REGRESSION= ZAL WITH AAL TO FAL (2)/
```

Kasu honetan, erregresioa ZAL menpekotzat eta AAL, BAL,FAL independententetzat harturik egingo du; parentesi barruan dagoen zenbaki bikoiti batek (2 kasu honetan) ohizko erregresioa nahi dugula adierzaten dio konputagailuari.

2)
```
1.                  16.
REGRESSION          VARIABLES= AAL,BAL,FAL,ZAL/
                    REGRESSION= ZAL WITH AAL (4),BAL(6),FAL(2)/
```

Diseinu honen bidez aldagaiak hierarkikoki sartzen dira erregresio-ekuazioan.

Agindu honen bidez hiru erregresio-ekuazio eskatzen zaizkio konputagailuari: lehenengoan ZAL aldagaia BAL aldagaiaren menpe, bigarrenean ZAL aldagaia BAL eta AAL aldagaien menpe,hirugarrenean ZAL aldagaia BAL,AAL eta FAL aldagaien menpe izanik.

Dakusagunez, aldagaiak banan banan sartzen dira erregresio ekuazioan, segitzen duten ordena parentesi barruan dauden zenbaki bikoitiak emandakoa, eta lehenengoa zenbaki bikoitirik haundiena daukana izanik.

3)
```
1.                  16.
REGRESSION          VARIABLES= AAL,BAL,FAL,ZAL/
                    REGRESSION= ZAL WITH AAL TO FAL (3)/
```

Diseinu honen bidez ere, aldagaiak erregresio-ekuazioan hierarkikoki sartzen dira, baina ez aurreko kasuan bezala;kasu honetan, hierarkia ez zaio konputagailuari ematen, honek alda-

gai independente bakoitzak dependentearen bariantzaz azaltzen duen proportzioa ikusiz aukeratzen baitu.

Honela, aldagai independenteak banan banan sartuko dira erregresioan, lehenengoa aldagai dependentearen bariantzak proportzio gehien azaltzen duena izanik. Bigarrena aurrekoarekin, hau da, biek elkarrekin proportzio gehien azaltzen dutena izango da, e.a.

Dakusagunez, parentesi barrua dagoen zenbaki bakoiti batek (3 kasu honetan) adierazten dio erregresio-mota hau konputagailuari .

```
    1.              16.
4)  REGRESSION      VARIABLES= AAL,BAL,DAL,JAL,KAL,LAL,MAL,NAL
                               ZAL/
                    REGRESSION= ZAL WITH AAL TO DAL(3), JAL TO
                                NAL (1)/
```

Diseinu honetan aldagai independenteen zerrendak ditugu, parentesi barruan zenbaki bakoitiak dituztelarik.

Disenu hau, 2) eta 3) kasuen sintesi bat da; parentesi barruan degoen zenbaki bakoiti bakoitzaren zerrendan, aldagai independenteak erregresioan 3) kasuan bezala sartuko ditu, hau da, azaltzen duten bariantzaren proportzioaren arabera; baina aurretik konputagailuak zenbaki bakoiti handienera jo du.

```
    1.              16.
5)  REGRESSION      VARIABLES = AAL TO ZAL/

                    REGRESSION = diseinuaren agindua,AAL,LAL(0)
```

Zero zenbakiak, aldagai bat edo batzuren ondoan kokatuta, aldagai hau edo hauek erregresio-ekuazioan ez direla sartuko esan nahi du, baina azpiprogramak orain dik ekuazioan sartu ezdiren aldagaietarako ematen dituen ikasketetan, kontutan hartzen ditu konputagailuak.

BESTE POSIBILITATE BATZU

a) Giltza-zenbaki bakoiti batzuren bidez, erregresio-ekuazioan sartzeko ditugun aldagaiei, limiteak ipini diezazkiegu.

(n,F,t) zenbakiak, parentesi barruan honela kokaturik, menpeko aldagaiaren izenaren ondoren ipini behar dira.

- n zenbakiak aldagaien kopurua handiena adieraziko du (n zenbakiaren bidez adierazten ez dugunean, 80 aldagaietakoa izango da).

- F: Adierazpen-parametro baten balio txikiena da (inferentzia estatistikoari dagokion kontzeptu bat da, hots, estatistika deskribatzailetik at geratzen da).

- t: Aldagai independente bakoitzari ekuazioan sartzeko eskatzen zaion jasapen minimoa da.

Aldagai independente baten jasapen edo tolerantzia, ekuazioan sarturik dauden beste aldagai independenteen bidez linealki azal ez daiteken aldagai horren bariantzaren proportzioa litzateke.

(t zenbakiaren bidez adierazten ez dugunean, tolerantzia txikiena 0'001-ena da, hots, aldagaiaren bariantzaren milakobat).

b) Aurresandako balioen eta hondarren zerrendak lor daitezke, hala nola hondar hauen adierazpide grafikoak ere.

Diseinuaren aginduaren ondoren RESID = 0 edo RESID=(Datu-eza onargarrien proportzioa).

Adibidez, RESID = 0 denean, datu-eza kasutarako ez ditu aurresandako balioak eta hondarrak kalkulatuko; baina, RESID = 0.4 balitz, % 40 batean datu-ezak onartuko lituzke eta ondorioz dagozkien balioak kalkulatuko.

Beste azpiprograma guztietan bezala, OPTIONS eta STATIS - TICS eska ditzakegu.

Azpiprograma honek dituen estatistiko gehigarriak, besteren artean, ahuexek dira:

1.- Korrelazio-matrizea, 2-Batezbestekoak eta desbidazio-standardak, 3-korrelazio-koefizente ez kalkulagarriak izango balira ohartaraziko lizkiguke, 6-balio teoriko eta hondar tipifikatuen grafikoa e.a.

STATISTICS eskatzen ez badira, erregresioaren ekuazioekin batera ohizko estatistikoak lortuko ditugu, hau da, r korrelazio-koefizientea, r^2 koefiziente mugatzailea eta inferentzia - estatistikoari dagozkion beste batzu.

PARTIAL CORR:

Dakigunez, bi aldagaien arteko egiazko elkar erlazioa - neurtu nahi badugu, baztertu egin beharko dugu fenomenoan sartzen diren beste aldagaien eragina.

Adibidez, uztabilketa eta euri-kantitatearen egiazko elkar erlazioa lortzeko neurketak egiterakoan, uztabilketetan eragina duten beste aldagai guztiak balio konstanteak hartu beharko lituzkete; baina, aldagai guztien balioek gorabeherakadak dituztenez gero, matematikoki oztopo hau korrelazio-koefizente partziala kalkulatuz gainditzen dugu.

PARTIAL CORR aginduaren bidez korrelazio-koefiziente partzialak lortuko ditugu.

Tratamendu estatistiko hau eskatzean, bi eratara joka dezakegu: Bata, gure ikerketarako dugun datu-matrizea erabiliz,- beste tratamendu estatistiko guztiekin bezala; bestea, korrelazio-matrizea erabiliz, korrelazio-koefiziente totalen matrizea alegia.

Apunte hauetan lehenengo bidea azalduko dugu .

Honako hau izango da, bada, tratamendu honen tankera:

```
1.              16.
PARTIAL CORR    Aldagaien zerrenda BY aldagaien zerr.(ordenako
                zenbakiak)/ [beste ald.zerr.BY... ]
```

Lehenengoko zerrendan dauden aldagaien korrelazio partzialak bigarren zerrendako aldagaien eraginak kontrolatuz edo baztertuz eskatzen ari gara. Parentesi barruan dauden zenbakien - bidez, banaka, binaka, hirunaka.... aldagaiak kontrolatuz joatea eskatzen dugu.

ADIBIDEAK:

```
1.              16.
PARTIAL CORR    AAL WITH BAL,DAL BY JAL,KAL (1,2)/
```

Adibide honetan, (AAL,BAL) eta (AAL,DAL) bikoteen korrelazioak eskatzen ditugu, JAL,KAL aldagaiak fenomenoan daukaten eragina baztertuz: lehenengo bakoitzaren eragina eta gero biena,(1,2) zenbakien arabera.

```
1.              16.
PARTIAL CORR    AAL TO DAL BY JAL(1)/ DAL,JAL BY AAL,BAL,KAL -
                (2,3)/
```

Adibide honetan, ordea, ALL,BAL,DAL aldagaien arteko korrelazio guztiak eskatzen ditugu, hots, korrelazio-matrizea, JAL aldagaiaren eragina baztertuz; eta bigarren diseinuan, DAL,JAL - aldagaien korrelazioa,AAL,BAL,KAL aldagaien eragina baztertuz, lehenengo binaka hartuz eta gero hirurak,(2,3) zenbakien arabera.

Azpiprograma honek dituen estatistiko gehigarriak hauexek dira:

1.- Korrelazio totalen matrizea, 2-batezbesteko aritmetiko eta desbidazio standarden zerrendak, 3-korrelazio koefizientе totalak ez kalkulagarriak.

SPSS programa-paketean aurki ditzakegu azpiprograma egokiak beste tratamendu estatistikotarako; besteak beste, hauexek aipa ditzakegu.

ANALYSIS OF VARIANCE AND COVARIANCE (ANOVA ETA ONEWAY azpiprogramak).

DISCRIMINANT ANALYSIS (DISCRIMINANT azpiprograma).

FACTOR ANALYSIS (FACTOR azpiprograma).

CANONICAL CORRELATION ANALYSIS (CANCORR azpiprograma).

Guztiak benetan interesgarriak dira, baina gai teoriko horiek "Estatistikaren Hastapenak" textu honetan ikusten ez direnez gero, ez ditugu dagozkien azpiprogramak eranskin honetan azaltzen.

Dena dela, irakurlea interesaturik balego ez luke zailtasun askorik dagozkien instrukzioak SPSS eskuliburuan aurkitzeko.

H I Z T E G I A

Agindu :	Sentencia, instrucción.
Agindu-segida:	Sucesión de instrucciones.
Agindu-sekuentzia:	Secuencia de instrucciones.
Ahaltsu:	Potente.
Aldagai:	Variable.
Aukerabide:	Elección.
Baldintzazko:	Condicional.
Batera:	Simultaneamente.
Berkodifikatu:	Recodificar.
Datu-eza:	Datos ausentes (Missing data)
Diseinu:	Diseño.
Emaitza:	Resultado (aritm.).
Erabide:	Procedimiento.
Eragiketa:	Operación (aritm.).
Eragin:	Operar.
Eragingarri:	Operable.
Eratorri:	Derivar.
Eraiki:	Construir.
Erabiltzaile:	Usuario.
Errore:	Error.
Eskuliburu:	Manual.
Etiketa:	Etiqueta.
Euskarri:	Soporte.
Gailu:	Dispositivo.
Garatu:	Desarrollar.
Garapen:	Desarrollo.
Geruza:	Estrato.

Giza-erabiltzaile:	Usuario humano.
Giltza-hitza:	Palabra clave.
Hautabidezkoa:	Opcional.
Hein:	Rango.
Hierarkikoki:	Jerarquicamente.
Inguraldea:	Periferia.
Irazkin:	Comentario.
Irazkin(du):	Comentar.
Itzuli:	Traducir.
Karaktere:	Caracter.
Kasu:	Caso.
Kode:	Codigo.
Kodifikatu:	Codificar.
Kodifikazio:	Codificación.
Konputagailu:	Computador(a).
Kontrol:	Control.
Memoria:	Memoria.
Mugatzaile:	Delimitador.
Nahitaezko:	Obligatoria, imprescindible.
Neurri:	Medida.
Neurketa:	Medida, evaluación.
Ordinatore:	Ordenador (conputer).
Programa:	Programa.
Programa (tu):	Programar.
Programatzaile:	Programador.
Prozesuaren Unitate Zentrala:	Unidad Central del Proceso.
Prozesatore:	Procesador.

Prozesa(tu):	Procesar.
Puntu hamartar:	Punto decimal.
Tankera:	Formato.
Txartel zulatu:	Ficha o tarjeta perforada.
Tratamendu:	Tratamiento.
Zehaztapen:	Especificación.
Zenbaki hamartar:	Número decimal.
Zenbaki oso:	Número entero.
Zerrenda:	Lista, listado.
Zirkuito elektroniko:	Circuito electronico.
Zulagailu:	Perforadora (máquina).
Zula:	Perforar.
Zulatzaile:	Perforista.
Zuri:	Blanco.