ESTATISTIKA I. PROGRAMA

I. ATALA: ALDAGAI BAKARREKO ANALISI ESTATISTIKO DESKRIBATZAILEA.

1. GAIA: Aldagai bakarreko maiztasun-banaketa I.

1.1. Sarrera orokorra. Estatistika kontzeptua. Datu iturburuak.
1.2. Aldagai estatistikoa. Sailkapena.
1.3. Zenbakizko deskribapena. Adierazpen grafikoa: barrak, histograma, sektoreak, biztanleriaren piramidea.

2. GAIA: Aldagai bakarreko maiztasun-banaketa II.

2.1. Momentuak. Propietateak.
2.2. Batezbesteko aritmetikoa eta harmonikoa. Propietateak.
2.3. Mediana, koantilak eta moda.
2.4. Sakabanatze-neurriak: ibiltartea, bariantza, desbideratze tipikoa, aldakuntza-koefizientea.
2.5. Simetria- eta kurtosia- neurriak
2.6. Aldagai zentratua eta tipifikatua.

3. GAIA: Desberdintasun neurriak.

3.1. Gini-ren indizea.
3.2. Lorenz-en kurba.

II. ATALA: BI ALDAGAIREN ANALISI ESTATISTIKO DESKRIBATZAILEA.

4. GAIA: Maiztasun-banaketa dimentsio bikoa.

4.1. Zenbakizko deskribapena eta adierazpen grafikoa.
4.2. Bazter-banaketak eta banaketa baldintzatuak. Independentzia estatistikoa.

5. GAIA: Aldagai kuantitatiboen arteko erlazioaren analisia.

5.1. Momentuak. Batezbestekoa, bariantza, kobariantza.
5.2. Transformazio eta konbinazio lineala.
5.3. Korrelazioa R^2-an.

6. GAIA: Aldagai ez kuantitatiboen arteko erlazioaren analisia.

6.1. Asoziazioa: koefizientea.
6.2. Korrelazio ordinala: koefizientea.

7. GAIA: Bi aldagairen arteko erregresioa.

7.1. Karratu txikienen erregresio lineala R^2-an. Doikuntza-egokitasunaren analisia: mugatze-koefizientea.
7.2. Erregresio parabolikoa, potentziala eta esponentziala. Doikuntza-egokitasunaren analisia: mugatze-koefizientea.

III. ATALA: ALDAGAI ANITZEKO ANALISI ESTATISKO DESKRIBATZAILEA

8. **GAIA: Bi aldagai baino gehiagoren arteko erregresioa eta korrelazioa**.

8.1. Karratu txikienen erregresio lineala R^n-an. Doikuntza-egokitasunaren analisia.
8.2. Korrelazio-koefizienteak: totala, anizkoitza eta partziala.

IV. ATALA: ZENBAKI INDIZEAK

9. **GAIA: Zenbaki-indizeak**.
9.1. Definizioa.
9.2. Zenbaki indize bakun eta konplexua.
9.3. Kontsumo prezioen indizea (KPI) eta beste indize batzuk.
9.4. Aldakuntza eta ondorioa.
9.5. Oinarri-aldiaren aldaketa eta serieen lotura.
9.6. Serie estatistikoen deflazioa.

V. ATALA. PROBABILITATE ESPAZIOAK

10. GAIA: Probabilitatea espazioak (I)
10.1. Sarrera
10.2. Probabilitate-definizioa.
10.3. Probabilitatearen lehen propietateak.
10.4. Probabilitateen esleipena.

11. GAIA: Probabilitate espazioak (II)
11.1. Probabilitate baldintzatua.
11.2. Ebaketaren teorema, partiketaren teorema eta Bayes-en teorema.
11.3. Dependentzia eta independentzia estatistikoa.

Oinarrizko bibliografia

Fernandez Agirre, K. (1996): *Estatistirako sarrera, koerlazioa, erregresioa eta datu anizkoitzen analisia.* U.E.U

Narvaiza, J.L. (1998): *Estadística aplicada a la gestión y a las ciencias sociales: estadística descriptiva y probabilidad*. Desclée de Brouwer.

Bibliografia gehigarria.

. **Escuder Vallés, R** (1987): *Métodos estadísiticos aplicados a la Economia*. Ariel Economía.

. **Martin-Guzman, P y Martin-Pliego, F.J.** (1992): *Cuso básico de estadística económica*. A.C.

. **Fernandez Aguirre, K.** (1989): *Estatistika deskribatzailea*. U.E.U.

I. GAIA: ALDAGAI BAKARREKO MAIZTASUN BANAKETA (I).

1.1. Sarrera orokorra. Estatistikaren kontzeptua.
1.2. Aldagai estatistikoa. Sailkapena.
1.3. Zenbakizko deskripzioa. Adierazpen grafikoa.
1.4. Ariketak.

1.1. Sarrera orokorra. Estatistikaren kontzeptua.

Edozein ikerketa egiteak fenomeno jakin batean sakontzea du helburu. Alabaina, fenomeno guztiak ez dira berdinak . Hau da, egunero eguzkia nondik irten den edo burtsa-indizea igoko den galdetuz gero, haien arteko desberdintasunaren bat aurkituko dugu, behintzat. Lehen kasuan beti erantzun bera izango da, hots, ekialdetik. Bigarren kasuan, berriz, bai edo ez izan daiteke erantzuna gertatu baino lehen ez baitago baieztatzerik. Estatistikak bigarren kasuko fenomenoak aztertzen ditu, hau da, lege finkoen bitartez ezar ezin daitezkeenak.

Estatistikak, zenbakizko informazioa lortzeko ez ezik, lortutako informazioa aztertzeko ere metodo zehatzak errazten dizkigu. Oro har, Estatistika erabiliko dugu datu-multzo handiak eskurakoi bihurtu nahi dugunean. Garbi dago datu guztietatik informazioa ateratzea ezinezkoa dela; beraz, Estatistikak eskueran jartzen dizkigun teknika grafikoak zein zenbakizkoak erabili beharko ditugu fenomenoa deskribatzeko eta datu-multzoan dagoen informazioa laburtzeko.

Hasteko, oso ondo menperatu behar ditugu Estatistikaren zenbait kontzeptu:

Populazioa: azterlan jakin baten helburua definitu orduko, aztertzeko daukagun informazioaren elementuen multzo bat da.
Lagina: kontuan hartutako populazioaren azpimultzo adierazgarri bat da.
Banakoa: populazioaren edo laginaren elementu bakoitza da, eta horien kopuruak populazioaren edo laginaren **tamaina** definitzen du.

Orain, informazioa non dagoen dakigunez gero, elementuen ezaugarri bakoitzak **aldagai** bat definitzen du populazio horretan.

1.2. Aldagai estatistikoa. Sailkapena.

Lehen esan dugunez, estatistikaren bidez fenomeno jakin baten zenbait ezaugarri azter ditzakegu. Beraz, behatutako ezaugarriak aldagai estatistikoak dira, eta X, Y, Z, ...letraz adierazten dira.

Ikertzen duguna → X

1. behaketa..... x_1
2. behaketa..... x_2

.
.

i. behaketa..... x_i

.
.

k. behaketa..... ..... x_k

SAILKAPENA

Ikerketa estatistiko batean populazioa edo lagina sakonki ikertzea ez dugu helburu; beraren ezaugarri batera edo bestera mugatzea baizik. Beraz, **dimentsio** horren arabera, honako aldagai hauek bereiziko ditugu:

a. *Aldagai bakuna* ---ikertu nahi dugun ezaugarria bakarra denean.
b. *Aldagai bikoitza*---ikertu nahi duguna bi ezaugarriz adierazten denean.
c. *Aldagai anizkoitza*---ikertu nahi duguna bi baino ezaugarri gehiagotan adierazten denean.

Bestalde, beste sailkapen bat aipa dezakegu, aldagaiek hartzen dituzten balioen **izaeraren araberakoa.**

a. *Kualitatiboak edo atributuak*---Aldagaiak hartzen dituen balioak zenbakizkoak ez direnez gero, beren arteko hurrenkera ez dago zehazterik. Balioen bidez, elkarren artean desberdinak direla baino ez da adierazten. Adibidez, jarduera ekonomikoa arlo batzuetan, sexua, ... Kasu horretan, aldagaiaren balioak modalitateak deitu ohi dira.

b. *Ordinalak*---Balioen artean, elkarren baztertzaileak dira, aurreko kasuan bezala, baina, orain, haien arteko hurrenkera defini daiteke. Esaterako,

edozeini buruzko zaletasuna adieraztekotan, aldagaiaren balioak honako hauek daitezke: gutxi, hala-nola eta asko.

c. *Kuantitatiboak*---Aldagaiak hartzen dituen balioak zenbakiak direnez gero, hurrenkera zehaztu ahal izateaz gain eragiketa matematikoak gara ditzakegu. Horren ondorioz, estatistika erabiltzeko aldagairik komenigarrienak izaten dira.

Aldagai **kuantitatiboak** honako era honetan sailkatzen dira:

a. *Diskretuak*--- Bi balio hurbilen artean aldagaiak har dezakeen balio kopurua finitua denean. Adibidez, gela bateko ikasle kopurua.
b. *Jarraituak*---Bi balio hurbilen artean aldagaiak har dezakeen balio kopurua infinitua denean. Adibidez, pertsona batzuen altuera edo pisua.

1.3. Zenbakizko deskripzioa. Adierazpen grafikoa.

Analisi estatistikoaren helburua lortzeko datu-multzo handiak aztertu behar ditugunez gero, ondo eta era berezian antolatu behar ditugu datu horiek.

Adibidez: zenbakizko behaketa izanez gero, begi-bistan dago datu horietatik zer edo zer eratortzekotan antolamendu egokia topatzea komenigarria dela.

0,2,1,2,2,3,6,5,3,1,1,2,3,0,2,4,1,2,3,2,0,8,2,2,3.

Datuek anai-arreben kopurua adierazteko, horrela idatzita ez dakigu banaketa nola dagoen. Beraz, irakurketa errazteko datuak modu egokian antolatu behar ditugu.

X_i	n_i	f_i	N_i	F_i
Anai-arreba kopurua	Pertsona kopurua	Maiztasun erlatiboa	Maizt.abso. metatua	Maizt. erlat. metatua
0	3	0´12	3	0´12
1	4	0´16	7	0´28
2	9	0´36	16	0´64
3	5	0´20	21	0´84
4	1	0´04	22	0´88
5	1	0´04	23	0´92
6	1	0´04	24	0´96
8	1	0´04	25	1

Aurreko taula aldagaiaren balioak, txikitik handira ordenatuz osatu da; horretaz gain, zenbatu behar da zenbat aldiz agertzen den aldagaiaren balio bakoitza; hala, aldagaiaren maiztasunak definitzen dira. Bestela esanda, aldagaiaren maiztasunek adierazten dute balio bakoitza zenbat aldiz errepikatzen den.

Beraz, aurreko adibidea kontuan hartuta:

- X : Aldagaia: neba-arreben kopurua.

- X_i : Aldagaiak hartzen dituen balioak $(i = 1,2,.....,k)$

- N : Behatutako multzo edo populazioaren tamaina.

- n_i : Maiztasun absolutuak $\Rightarrow$ aldagaian balioa edo datu bera hartzen duten objektu edo indibiduoen kopurua adierazten du.

$$\sum_{i=1}^{k} n_i = N$$

- f_i : Maiztasun erlatiboa $\Rightarrow$ maiztasun absolutuak erlatiboki kontuan hartuta.

$$f_i = \frac{n_i}{N} \text{ beraz } \sum_{i=1}^{k} f_i = 1 \text{ eta } 0 \leq f_i \leq 1$$

Maiztasun mota hori ehunekotan ere adieraz daiteke.

- N_i : Maiztasun absolutu metatuak $\Rightarrow$ populazioaren X_i balioa eta txikiagoak hartzen dituzten indibiduo kopurua adierazten du.

$$N_i = n_1 + n_2 + ... + n_i \text{ beraz } N_k = \sum_{i=1}^{k} n_i = N \text{ eta } N_i - N_{i-1} = n_i$$

- F_i : Maiztasun erlatibo metatuak $\Rightarrow$ maiztasun absolutu metatuak erlatiboki kontuan hartuta.

$$F_i = \frac{N_i}{N} = f_1 + f_2 + ... + f_i \text{ beraz } F_k = \frac{N_k}{N} = 1$$

Maiztasun mota hori ehunekotan ere adieraz daiteke.

Egoera batzuetan, komenigarria izaten da aldagaiaren balioak elkartzea, eta, horren ondorioz, **banaketa elkartuak edo jarraituak** erabiltzen ditugu. Hau da, balioak tarteka adierazten dira.

$L_{i-1} - L_i \rightarrow$ Aldagaiaren *i.* tartea edo klasea. L_{i-1} tartearen behe-muga. L_i tartearen goi-muga.

$\frac{L_{i-1} - L_i}{2} \rightarrow$ klase-marka. Tartearen zentroko puntua. $X_i^{'}$ adierazten da.

ADIERAZPEN GRAFIKOA

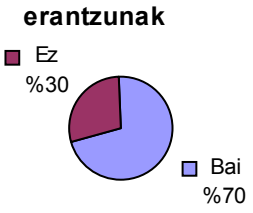
1. Sektore-grafikoa

Biribilak edo tarta itxurakoak deitzen zaie. Sektore-grafikoek aldagaiaren maiztasunak adierazten dituzte, eta maiztasunaren arabera zatitzen dute zirkuluaren azala.

Demagun talde bati honako hau galdetu zaiola: Aurten oporretara joango zara? eta horren erantzunak hauek direla:

Erantzuna	f_i	**360°**
Bai	0´7	252
Ez	0´3	108

Beraz, banaketa horri dagokion sektore-grafikoa honako hau izango da:



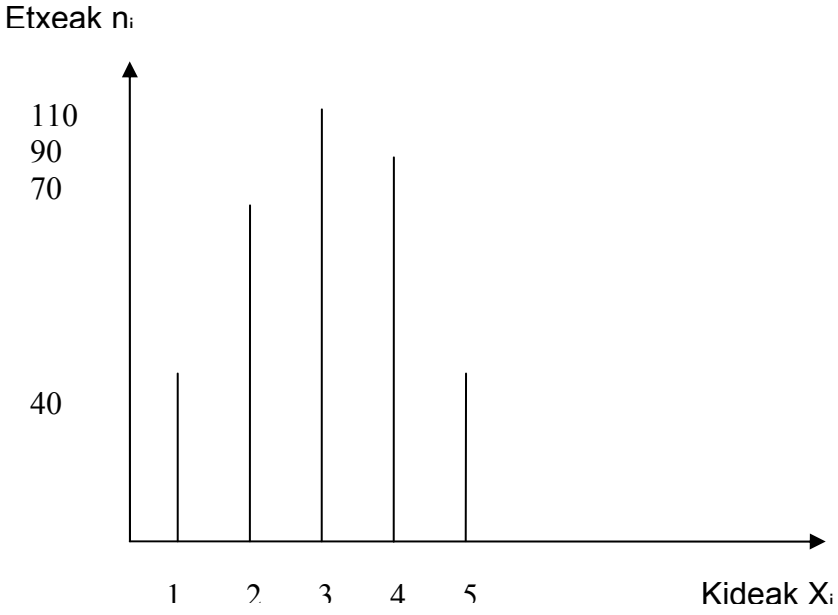
2. Barra-diagrama.

Kartesiar ardatz-sisteman maiztasun-banaketa adierazteko erabiltzen da, abzisean aldagaiaren balioak adieraziz eta ordenatuan, berriz, maiztasun absolutuak edo erlatiboak.

Demagun 350 etxeren sailkapena familia-kide kopuruaren arabera honako hau dela:

Kide kopurua X_i	**Etxe kopurua** n_i
1	40
2	70
3	110
4	90
5	40

Banaketa horri dagokion barra-diagrama honako hau izango da:



3. Histograma eta maiztasun-poligonoa.

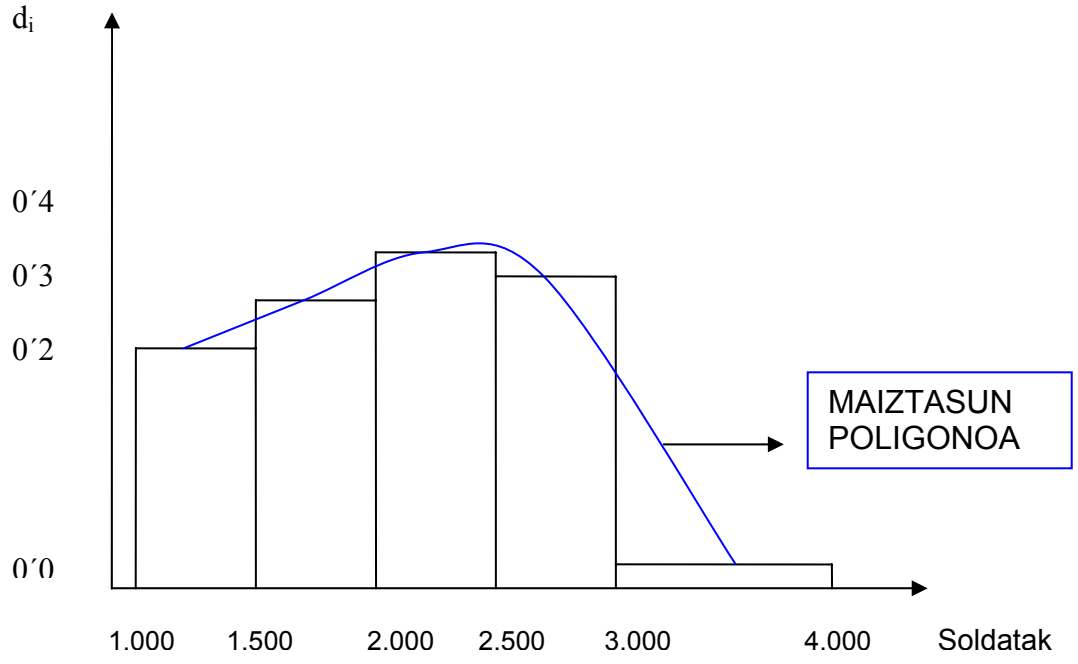
Histograma kartesiar ardatz-sisteman maiztasun-banaketa jarraituak adierazteko erabiltzen da, abzisean aldagaiaren balioak tarteka adieraziz eta ordenatuan, berriz, maiztasun-dentsitatea. (d_i)

Bestalde, **maiztasun-poligonoa** eratzeko klase-marken balioak lerro baten bidez lotzen ditu;hala, poligono itxurako lerro zabala lortuko da.

Demagun 670 langileren soldatak honako banaketa hau duela:

Soldata $L_{i-1} - L_i$	**Langile kop.** n_i	**Klase – marka** X'_i	**Maiztasun - dentsitatea** d_i
1.000-1.500	100	1.250	0´2
1.500-2.000	150	1.750	0´3
2.000-2.500	200	2.250	0´4
2.500-3.000	180	2.750	0´36
3.000-4.000	40	3.500	0´04

Banaketa horri dagokion histograma zein maiztasun-poligonoa honako hau izango dira:



1.4. Ariketak.

1. Taula honek 1995ean Fasa-Renaultek modelo bakoitzean ekoiztutako auto kopurua adierazten du:

Modeloak	Ekoizpena
19	37.788
Clio	91.991
Laguna	104.179
Megane	19.543
Twingo	91.534

a) Aztertzen ari den aldagaia zein izango da? Zein aldagai mota izango da?
b) Eratu sektore-diagrama.

2. Hiri handi bateko ospitale baten gastuak honako hauek direla adierazteko irudika ezazu sektore diagrama bat: soldatetan % 73, horniduretan % 13, mantentzean % 8 eta administrazio gastuetan % 6.

3. 500 familiako lagin baten hilabeteko diru-sarrerak honela banatzen dira:

DIRU SARRERAK €	FAMILIA KOPURUA
500 baino gutxiago	25
500-600	61
600-900	72
900-1.200	75
1.200-1.500	128
1.500-1.800	64
1.800-3.000	45
3.000-3.600	30

Irudika ezazu banaketa hori grafikoki

4. Enpresa bateko merkataritzako agenteak, 25 egunetan, lortutako eguneroko salmentak honako hauek dira: 8, 10, 12, 12, 10, 10, 11, 11, 10, 9, 11, 10, 9, 9, 11, 12, 9, 10, 9, 10, 9, 10, 9, 8, 10.

Irudika ezazu banaketa hori maiztasun absolutuen eta erlatiboen taulak erabiliz.

2. GAIA: ALDAGAI BAKARREKO MAIZTASUN BANAKETA (II).

2.1. Momentuak. Propietateak.
2.2. Batezbesteko aritmetikoa.
2.3. Mediana, koantilak eta moda.
2.4. Sakabanatze-neurriak: ibiltartea, bariantza, desbideratze tipikoa eta aldakuntza-koefizientea.
2.5. Simetria eta kurtosia- edo zapaltasun-neurriak.
2.6. Aldagai zentratua eta tipifikatua

2.1. Momentuak. Propietateak.

Banaketa baten nolakotasun guztiak kontuan hartzen ditugunean, banaketaren balio guztietatik ondorioztatzen diren balio zehatzak aipatuko ditugu. Balio zehatz horiek momentu izena dute.

Beraz, momentuen bidez bereizten dira banaketak. Hau da: bi banaketak momentu guztiak berdinak badituzte, orduan esan dezakegu banaketa bera dela. Eta zenbat eta momentu berdin gehiago izan hainbat eta antzekoagoak izango dira.

Edozein abiabururekiko x_0, *h* ordenako momentua, honela definitzen da:

$$a_{h,x_0} = \frac{1}{N}\sum_{i=1}^{k}(x_i - x_0)^h n_i \qquad \forall h \in \mathrm{N}$$

Bi momentu mota bereiziko ditugu:

a. Momentu arruntak edo jatorriarekiko momentuak, $x_0 = 0$.
b. Momentu zentralak edo batezbestekoarekiko momentuak, $x_0 = \bar{x}$.

a. Momentu arruntak edo jatorriarekiko momentuak .

Oro har, jatorriarekiko *h* ordenako momentua a_h, izendatuko dugu, beraz,

$$a_h = \frac{1}{N}\sum_{i=1}^{k} x_i^h n_i$$

- Jatorriarekiko zero ordenako momentua: $a_0 = \frac{1}{N}\sum_{i=1}^{k}(x_i - 0)^0 n_i = 1$
- Jatorriarekiko lehen ordenako momentua: $a_1 = \frac{1}{N}\sum_{i=1}^{k} x_i^1 n_i = \bar{x} \rightarrow$ banaketaren batezbesteko aritmetikoa.

---

b. Momentu zentralak edo batezbestekoarekiko momentuak

Oro har, *h* ordenako momentu zentrala m_h izendatuko dugu, beraz,

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

- Zero ordenako momentu zentrala: $m_0 = \frac{1}{N}\sum_{i=1}^{k}(x_i - \bar{x})^0 n_i = 1$
- Lehen ordenako momentu zentrala: $m_1 = \frac{1}{N}\sum_{i=1}^{k}(x_i - \bar{x})^1 n_i = 0$
- Bigarren ordenako momentu zentrala:

$$m_2 = \frac{1}{N}\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i = S_x^2 \rightarrow \text{banaketaren bariantza da.}$$

2.2. Batezbesteko aritmetikoa.

Banaketa baten balioak batuz eta populazioaren tamainaz zatituz kalkulatuko dugu banaketa horren batezbesteko aritmetikoa. Batezbestekoaren ikurra hauxe da: $\bar{X}$. Beraz,

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

Maiztasunak masatzat hartzen baditugu, batezbesteko aritmetikoak banaketaren grabitate-zentroa adieraziko du.

1. Batezbestekoarekiko aldagaiaren balioen desbideratzearen batura zero da beti. Hau da:

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

Propietate horrek batezbestekoaren esangura adierazten du; hau da, batezbestekoaren alde bateko balioak eta beste aldekoak batezbestekoarekiko orekatzen dira eta.

2. *Banaketa bitarra*→ Aztertzen dugun aldagaiak bi balio baino ez du, bi besterik ez. Oro har, honela adieraziko dugu: arrakasta (1 kodea) edo porrota (0 kodea); arrakastaren proportziorako, edo, haren maiztasun erlatiborako "*p*" idatziko dugu eta porrotarako "*q*". Beraz, banaketa bitarren batezbestekoa kalkulatzeko, honako egingo dugu:

X	f_i
0	q
1	p

$$\bar{x}=1.p+0.q=p$$

Horrenbestez, banaketa bitarraren batezbestekoa arrakastaren maiztasun erlatiboa izango da.

3. *Transformazio linealaren* batezbestekoa.

Demagun $Y=a+bX$ *X*-ren transformazio lineala dela; beraz, *Y*-ren batezbestekoa $\bar{y}=a+b\bar{x}$ izango da. Hau da, batezbestekoaren definizioaren arabera, badakigu $\bar{y}=\frac{1}{N}\sum_{i=1}^{k}y_i n_i$ dela eta $Y=a+bX$ dela; beraz, *Y*-ren batezbestekoa honela adieraz dezakegu:

$$\bar{y}=\frac{1}{N}\sum_{i=1}^{k}(a+bx_i)n_i=a+b\bar{x}=\bar{y}$$

Transformazio linealetan jatorriaren eta neurriaren aldaketek eragiten diote batezbestekoari.

4. Balioen multzo batetik azpimultzo disjuntuak lortzen baditugu, multzo osoaren batezbesteko aritmetikoa honela erlazionatzen da beste azpimultzoen batezbesteko aritmetiko guztiekin.

$$\bar{x} = \frac{\bar{x}_A N_A + \bar{x}_B N_B + \bar{x}_C N_C}{N}$$

5. ▪ Banaketaren maiztasun guztiak *c* parametroaz biderkatu edo zatitzen baditugu, batezbestekoa ez da aldatzen.

   ▪ Aldagaiaren balio guztiak *c* parametroaz biderkatu edo zatitzen baditugu batezbestekoa *c*-z ere biderkatuta edo zatituta edukiko dugu.

Abantailak:

a. Banaketaren balio guztiak hartzen ditu kontuan.
b. Aldagai kuantitatiboekin kalkula daiteke.
c. Banaketa bakoitzari batezbesteko bat baino ez dagokio.

Eragozpenak:

a. Banaketak muturreko balioak dituenean, batezbestekoak ez du banaketaren zentroa ondo adierazten, muturreko balioen eraginagatik.
b. Aldagai kualitatibo edo ordinalekin ezin da kalkulatu.

2.3. Mediana, koantilak eta moda.

MEDIANA---

Banaketaren balioak txikitik handira ordenatuta, banaketa-erditik zatitzen duen balioa banaketaren mediana da. Hau da, banaketaren balioen erdia mediana bezalakoa edo txikiagoa izango da, eta beste erdia, mediana bezalakoa edo handiagoa. M_e ikurraz adieraziko dugu.

● Banaketa tarteka emanda baldin badago, *medianadun-tartea* definituko da, eta tarte horren barruan egongo da mediana. Beraz, taldearen erdia duen maiztasun metatuari dagokion tarteari deituko zaio medianadun tartea.

KOANTILAK---

Oro har, banaketa zenbat ataletan zatitzen duten balioak dira. Hau da, atal bakoitzak duen balioen kopurua bera izango da. Banaketa zatitzen den atalen kopuruari erreparatzen badiogu, hiru koantil mota defini daitezke: kuartila, dezila eta pertzentila.

Kuartilek lau ataletan zatitzen dute banaketa, eta banaketa bakoitzak hiru kuartil ditu. Hau da:

● Lehen kuartila (q_1)---taldearen kopuru-laurdena $\left(\frac{N}{4}\right)$ duen maiztasun metatua dagokion balioa da lehen kuartila; hau da, banaketaren balioen % 25 lehen kuartila edo txikiagoak dira eta beste balioen % 75 lehen kuartila bezalakoak edo handiagoak dira.

● Bigarren kuartila (q_2)---maiztasun metatuan taldearen kopuru-erdian $\left(\frac{2N}{4}\right)$ duen aldagaiaren balioa da bigarren kuartila; hau da, banaketa erditik zatitzen duen balioa. Beraz, kontuan hartu behar dugu bigarren kuartila eta mediana balio bera direla.

● Hirugarren kuartila (q_3)---maiztasun metatuan taldearen kopuruaren hiru laurden $\left(\frac{3N}{4}\right)$ duen aldagaiaren balioa da hirugarren kuartila; hau da, banaketaren balioen % 75 hirugarren kuartila bezalakoak edo txikiagoak dira, eta beste balioen % 25 hirugarren kuartila bezalakoak edo handiagoak dira.

Pertzentilek--- banaketa ehun ataletan zatitzen dute. Atal bakoitzean banaketaren balioen % 1 dugu. Pertzentilak adierazteko *p* letra erabiliko dugu. 99 pertzentil ditugu: $p_1, p_2, \ldots\ldots, p_{99}$.

MODA---

Maiztasun-banaketa jakin batean gehien errepikatzen den balioa da moda; hau da, maiztasun-dentsitaterik handieneko balioa da. M_o ikurraz adieraziko dugu.

● Banaketa tarteka emanda baldin badago, *modadun-tartea* definituko dugu, eta tarte horren barruan egongo da moda. Beraz, maiztasun-dentsitaterik handieneko tarteari deituko diogu modadun tartea.

2.4. Sakabanatze-neurriak: ibiltartea, bariantza, desbideratze tipikoa eta aldakuntza-koefizientea.

Orain arte ikusitako neurriak zentrorako joeraren neurriak dira eta informazioa laburtzea dute helburu. Orain azalduko diren neurriek, berriz, zentrorako joeraren neurriek, informazio-laburpena ondo adierazten dutenetz aztertzea dute helburu.

Adibidez, jakin nahi badugu batezbestekoak banaketaren balio guztien zentro-joera ea zehatz neurtzen duen, kontuan hartu beharko dugu batezbestekoaren arabera balio bakoitzaren distantzia zein den. Balio guztiak batezbestekoaren alboan badaude, batezbesteko hori adierazgarria izango da. Distantzia nagusiei edo txikiei sakabanatze edo aldakortasuna deitzen zaie.

IBILTARTEA---

Aldagaiaren balioak txikitik handira ordenatuta, balio handienaren eta txikienaren arteko diferentzia da banaketaren ibiltartea. Hau da,

$$I = X_k - X_1$$

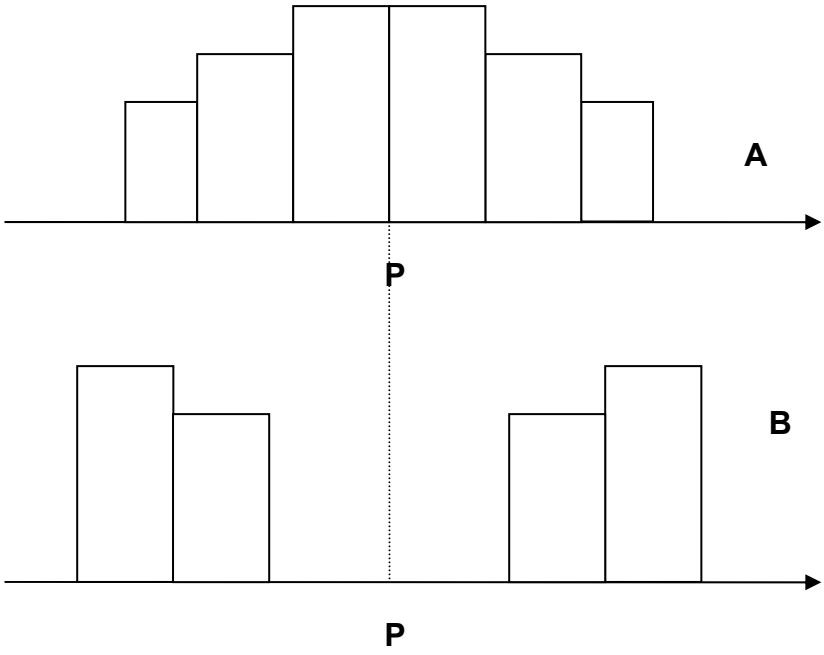
Demagun bi banaketa ditugula:
a. 3,5,6,7,10,12,15,18
b. 3,8,8,8,9,9,9,18

Bi kasuetan, ibiltartea 15 da; baina lehen banaketak bigarrenak baino sakabanatze handiago du. Beraz, edozein banaketak muturreko balioak dituenean erator dezakegu ibiltartearen bidez sakabanatzea neurtzea oso fidagarria ez dela. Adibidez, muturreko balioa kenduz gero, informazioa hobetzen da. Beraz, bi kasuetan 3 eta 18 kenduz gero, lehen kasuaren

ibiltartea 15-5=10 izango da eta bigarren kasuarena, berriz, 9-8=1. Orain bai, lehenak bigarrenak baino sakabanatze handiagoa duela ikusten da.

Dena dela, gure azken helburua zentralizazio-neurrien adierazgarritasuna aztertzea denez, azaldutako aurreneko neurriak ez du helburua betetzen, ez duelako posizio-neurri bat bera ere erabiltzen. Beraz, zentralizazio-neurriak kontuan hartzen dituzten sakabanatze-neurriak behar dira; hau da, bariantza eta desbideratze tipikoa.

Demagun bi banaketaren **P** zentralizazio-neurriak honako hauek direla eta bakoitzaren adierazgarritasuna ezagutu nahi dugula.



Agerian dago, lehen banaketaren zentralizazio-neurria bigarrena baino adierazgarriagoa dela, lehenengoan aldagaiaren balioak bigarrenean baino hurbilago baitaude. Balioak *P*-tik zenbat eta hurbilago egon, zentralizazio-neurria orduan eta adierazgarriagoa izango da.

Nola neur daiteke edozein banaketako sakabanatzea?

BARIANTZA---

Batezbestekoaren arabera, aldagaiaren balioen distantzien kuadroaren batezbestekoa da, eta S^2 ikurraz adieraziko da.

$$S^2 = \frac{1}{N}\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i \geq 0$$

Bariantzak batezbestekoaren arabera neurtzen du sakabanatze handiena edo txikiena. Beraz, sakabanatzea oso handia baldin bada batezbestekoa ez da oso adierazgarria izango.

Muturreko kasuan, aldagaiaren balio guztiak berdinak direnean, batezbestekoaren balioa bera izango da; hori dela eta, barantziaren balioa zero izango da.

Propietateak

1. Bariantza zera da: jatorriarekiko bigarren ordenako momentua ken lehen ordenakoa ber koadroa:

$$S^2 = \frac{1}{N}\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i = \frac{1}{N}\sum_{i=1}^{k}\left(x_i^2 + x^2 - 2x_i\bar{x}\right)n_i = a_2 - a_1^2$$

2. Transformazio linealak: $Y = aX + b$

$$S_y^2 = \frac{1}{N}\sum_{i=1}^{k}(y_i - \bar{y})^2 n_i = a^2 S_x^2$$

Transformazio linealetan bariantzari eskala-aldaketek eragiten diote, baina ez jatorri-aldaketek.

3. Bigarren ordenako momentu guztietatik, momenturik txikiena da bariantza. Edozein abiabururekiko x_0 bigarren ordenako momentua honela idatz daiteke.

$$a_{2,x_0} = \frac{1}{N}\sum_{i=1}^{k}(x_i - x_0)^2 n_i = S^2 + (\bar{x} - x_0)^2$$

Bigarren zatia beti zenbaki positiboa izango denez gero, bigarren ordenako momenturik txikiena izango da bariantza.

DESBIDERATZE TIPIKOA---

S ikurraz adierazten da eta bariantzaren erro karratu positiboa da. Beraz, aldagaiaren neurri-unitatetan emanda dago desbideratze tipikoa .

$$S = +\sqrt{S^2}$$

ALDAKUNTZA KOFEZIENTEA---

Demagun bi banaketaren batezbestekoak eta desbideratze estandarrak badakizkigula:

$$\overline{X}_1 = 10kg. \quad S_1 = 4kg \qquad \overline{X}_2 = 10kg \quad S_2 = 6kg.$$

Galdetuz gero zein banaketatan den batezbestekorik adierazgarriena, lehena erantzungo dugu; izan ere, batezbestekoak berdinak direnez gero desbideratze tipikoeii erreparatzen badiegu lehena bigarrena baino txikiagoa da. Beraz, adierazgarriagoa izango da.

Baina, batezbestekoak desberdinak direnean, desbideratze estandarrak konparatzea ez du balio sakabanatzea azter dadin. Beraz, batezbestekoak desberdinak izanez gero, batezbestekoaren eraginpean dagoen neurria behar da. Aldakuntza-koefizientea izango da neurri hori eta g_0 ikurraz adieraziko da.

$$g_0 = \frac{S}{\overline{X}}$$

Koefiziente horri erreparatzen badiogu, ikusten dugu dimentsio gabea dela. Hau da, kontuan hartu behar dugu desbideratze tipikoak eta batezbestekoak dimentsioa bera dutela.

Aldakuntza-koefizienteak adierazten du desbideratze tipikoak batezbestekoa zenbat aldiz edukitzen duen, eta desbideratze estandarrak g_0 zenbat eta handiago izan, orduan eta aldi gehiagotan edukiko du batezbestekoa. Beraz, aldakuntza-koefizienteak zenbat eta balio handiago izan hainbat eta txikiago izango da adierazpena.

<u>Propietatea</u>

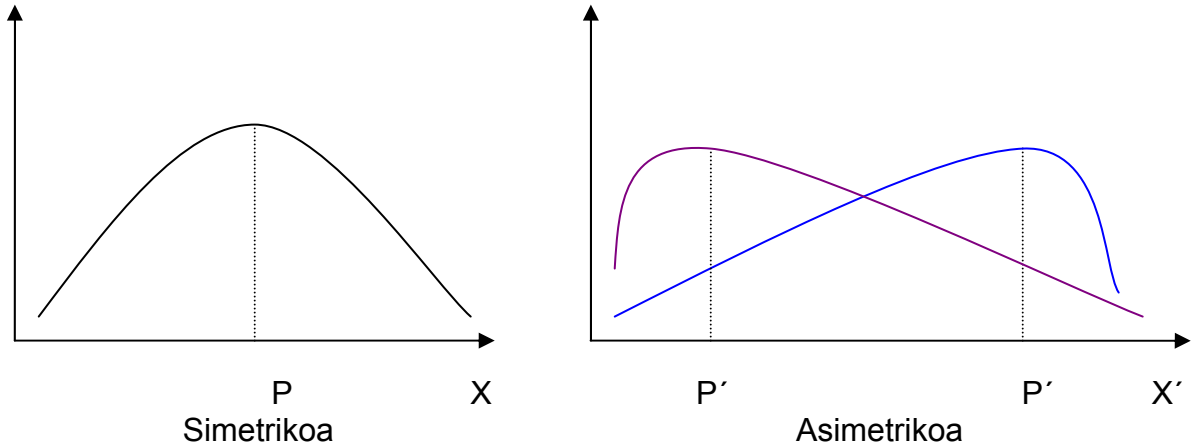
Aldakuntza-koefizienteari jatorri-aldaketek eragiten diote, baina ez eskala-aldaketek.

2.5. Simetria-eta kurtosia-neurriak

SIMETRIA-NEURRIAK---

Banaketaren adierazpen grafikoa egin gabe, neurri baten bidez simetriaren maila jakin daiteke. Demagun banaketaren irudikapen grafikoa jakina dela eta batezbestekoaren puntuan abzisarekiko lerro-zut bat marraztu dela; lerro-zutaren alde baten edo bestearen balioen kopurua berdina denean esango da banaketa simetrikoa dela.

Kontrako kasuan, banaketak asimetrikoak izango dira; izan ere, banaketa jakin batean, simetriarik ezari asimetria deitzen zaio. Hau da,



Kontuan hartu behar dugu erabiliko ditugun neurriak batezbestekoaren araberakoak izango direla. Beraz, banaketa simetrikoetan bai batezbestekoaren ezkerraldean bai eskuinaldean balioen kopurua bera izango da. Hau da, desbideratze kopurua bera izango da minus ikurraz eta plus ikurraz. Hori neurtzeko, honako koefiziente hau erabiliko dugu:

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

Dimentsio gabeko koefizientea lortzeko, lehengoa beste kantitateez zatitu behar dugu, eta biak neurri unitate bera izan behar dute. Beraz, desbideratze tipikoa ber hiru aukeratuko da. Horren bidez, Fisher-en simetria-neurria definituko dugu.

Koefizienteak m_3-k edukitako zeinua hartuko du, desbideratze tipikoa beti positiboa baita. Beraz,

$g_1 = 0 \Rightarrow$ banaketa simetrikoa izango da.
$g_1 \prec 0 \Rightarrow$ banaketa asimetriko negatiboa edo ezkerrerantz
$g_1 \succ 0 \Rightarrow$ banaketa asimetriko positiboa edo eskuinerantz

Propietatea

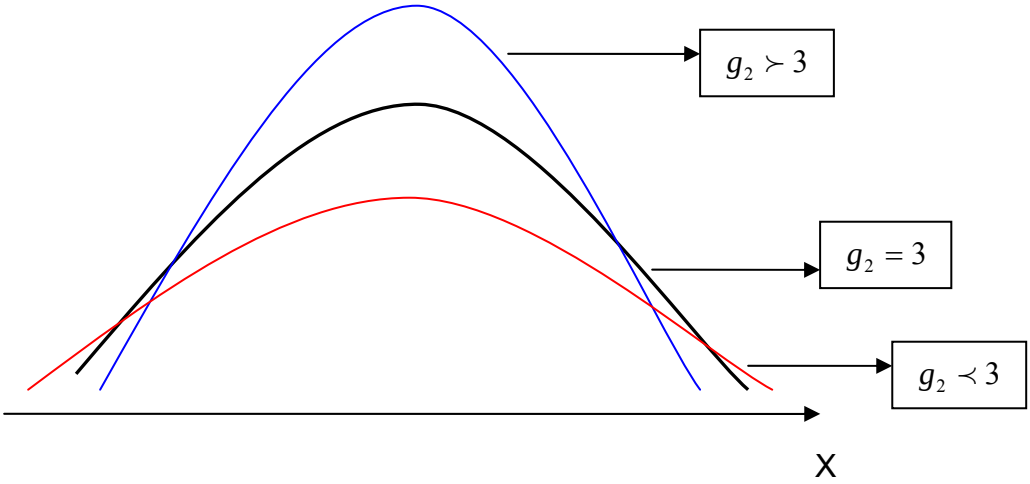
Asimetria-koefizienteari ez diote eragiten ez jatorri-aldaketek ezta eskala-aldaketek ere.

KURTOSIA EDO ZAPALKUNTZA NEURRIAK

Neurri horiek, gehienetan kanpai-forma duten banaketei aplikatzen zaie; hots, moda bat besterik ez dutenei, simetrikoak edo pittin bat asimetrikoak diren arren.

Banaketaren kurtosia aztertzeko, alde batetik, banaketa mota bat definitu beharko da eta, beste aldetik, kurtosia-koefizientea. Banaketa hori-normala delakoa-, erreferentzia izango denez gero, beraren formari erreparatuz aztertuko da beste banaketen kurtosia. Horretaz gain, honako adierazle hau kalkulatuko da:

$$g_2 = \frac{m_4}{S^4} \begin{cases} \succ 3 & \textit{zorroztasuna} & \textit{handiagoa} \\ = 3 & \textit{normalari} & \textit{dagokio} \\ \prec 3 & \textit{zapalkuntza} & \textit{txikiagoa} \end{cases}$$



Kurtosi-koefizientearen bidez azter daiteke ea banaketaren forma kanpaia edo *U* izango den. Hau da, ezagun da banaketa uniformearen (banaketa angeluzuzen) kurtosi-koefizientearen balioa gutxi gorabehera 1´8 izaten denez, banaketak kanpai-forma edukiko duela banaketajakin baten koefizientea 1'8 baino handiagoa baldin bada; banaketaren koefizientea, berriz, 1'8 baino txikiagoa baldin bada, banaketa, *U* formaduna izango da.

2.6. Aldagai zentratua eta aldagai tipifikatua.

Lehenengo eta behin, aldagai zentratua definitu behar da. Aldagai baten batezbesteko aritmetikoa zero baldin bada, aldagai zentratua deitzen zaio.

Edozein *X* aldagairi bere batezbestekoa kenduz, $(X - \bar{x})$ aldagai zentratua lortzen da.

Gainera, esaten da aldagai zentratu baten desbideratze estandarra 1 denean aldagai tipifikatua edo estandarizatua dela. Aldagai zentratuen

kasuan bezala, edozein X aldagairi bere batezbesteko aritmetikoa kenduz eta desbideratze tipikoaz zatituz $t = \frac{X - \bar{x}}{S_x}$ aldagai tipifikatua lortzen da. Horrek honako propietate hauek beteko ditu beti.

1. Aldagai tipifikatuaren batezbestekoa 0 da.

$$\bar{t} = \frac{1}{S_X}(X - \bar{x}) = 0$$

2. Aldagai tipifikatuaren bariantza 1 izango da.

$$Bar(t) = Bar\left(\frac{X - \bar{x}}{S_x}\right) = \frac{1}{S_x^2} S_x^2 = 1$$

Ariketak

1. Taula honetan, 1989ko auto-enpresen salmentak (mila milioitan) eta langile kopurua (milatan) agertzen dira:

ENPRESA	SALMENTAK	LANGILEAK
SEAT	457´3	23´8
RENAULT	449´7	19´2
G.MOTORS	372´7	9´4

a. Kalkula itzazu batez besteko salmentak eta enpresako langile kopuruaren batezbestekoa.
b. Aztertu non dagoen sakabanatzerik txikiena, salmentetan ala langile kopuruan.
c. Kalkula itzazu batez besteko salmentak langile bakoitzeko.

2. Taula honetan X aldagaiaren maiztasun-banaketa agertzen da:

x_i	n_i
-4	5
-2	4
0	2
2	4
4	5

Zera eskatzen da: datu horietan oinarrituta, azter ezazu zein izango den banaketaren forma (asimetria eta kurtosia), eta komenta itzazu emaitzak.

3. Merkataritza-bidaiari batek bi enpresatan lan egiten du, eta enpresa batetik hilero, 165€ irabazten du komisio bakoitzean eta, bestetik, berriz, 216€ irabazten du. Lehenengo enpresak bere agenteei ordaindutako komisioen diru kantitateak 114 €-ko batezbestekoa eta 37 €-ko desbideratzea dituen banaketari jarraitzen dio; bigarrenak, berriz, 191 €-ko batezbestekoa eta 56 €-ko desbideratzea dituenari jarraitzen dio.

Bi enpresa horien artean non dago hobeto merkataritza-agente hori erlatiboki?.

4. Zinema-ekoizle bat film baterako estra talde bat aukeratzen ari da. Lehenengo 10 pertsonaren adinak honako hauek dira:

50, 51, 56, 55, 55, 61, 49, 60, 52, 51

Filmaren zuzendariak 54 urte inguruko gizonak nahi ditu. Estatistika gustatzen zaionez gero, batezbestekoaren arabera adin bakoitza gehienez desbideratze batera egotea nahi du. Estra talde horrek baldintza hori betetzen du?. Zein adinek betetzen dute baldintza?.

5. Bi auto-industriatan, azken urteotan, honako auto kopuru hauek ekoitzi dira:

Lauhilekoa	Lantegi 1	Lantegi 2	1.aren lanegunak	2.aren lanegunak
1	600	650	68	68
2	750	1.200	78	78
3	850	1.250	54	54
4	400	800	74	54

Zera eskatzen da:

a. Zein lantegitan izango da eguneroko batez besteko produkziorik handiena?.
b. Datu-taulari begiratuz, ikusten da lehenengo hiru lauhilekoetan ekoizpena handitzen dela eta laugarrenean, berriz, txikitu, baina zein lantegitan izango da laugarren lauhilekoaren beherakadarik handiena erlatiboki?.

6. . Maiztasun banaketa hau ezagututa:

Adina	f_i
18	0,3
19	0,2
20	0,25
21	0,25

m_3 = 0´04 eta m_4 = 2´8 direla jakinik; aztertu eta komenta itzazu:

a. Zentrorako joera.
b. Sakabanatzea.
c. Banaketaren forma.
d. Adieraz ezazu grafikoki.

7. 2000ko bi seihilekoetan honako hauek izan dira informatika-denda bateko ordenagailu-salmentak :

Prezioak €-tan	1. seihilekoa	2. seihilekoa
600	20	20
1200	30	30
1800	25	40
2400	15	20

a. Lor ezazu bi seihilekoen batez besteko prezioa eta esan bietatik zein den adierazgarriena.
b. Urteko salmenta osorako, kalkula ezazu batez besteko prezioa, preziorik maizena eta maiztasun-banaketa bi zati berdinetan banatzen duen prezioa.

8. Atleta talde baten pultsazioak neurtu dira lasterketa bat egin ondoren:

Pultsazioak	Atleta-kopurua
$[70,75)$	3
$[75,80)$	3
$[80,85)$	7
$[85,90)$	10
$[90,95)$	12
$[95,100)$	8

a. Kalkula itzazu zentrorako joeraren neurriak.
b. Azter ezazu aldagaiaren sakabanatzea, desbideratze tipikoa eta aldakuntza-koefizientearen emaitzak komentatuz.

9. 25 egunetan enpresa bateko merkataritzako agenteak lortutako eguneroko salmentak honako hauek dira: 8, 10, 12, 12, 10, 10, 11, 11, 10, 9, 11, 10, 9, 9, 11, 12, 9, 10, 9, 10, 9, 10, 9, 8, 10.

a. Kalkula itzazu batezbestekoa, moda eta mediana.
b. Aurreko datuei begira, merkataritzako agenteak hirugarren egunean 12 unitate saldu zituen; baina beste agente batek, berriz, 20 unitate saldu zituen 25 egun horietan haien batez besteko salmentak eta desbideratze tipikoa 18´5 eta 2´5 izanik, hurrenez hurren, bietariko zeinek izan du emaitzarik onena bakoitzaren salmenta multzoarekin konparatuz?

10. Honako taula honetan 1. lauhilabetekoan Enpresa Eskolako 1. mailako ikasle bik lortutako notak azaltzen dira:

	Finantza-Kontab. I	**Ekonom. Politikoa I**	**Industrializazio eta E.G.**	**E.G. Aplik. Informatika**	**Finantza-Matemat. I**	**Enpresarako Matemat. I**
Kepa	6	5’5	4	7’5	7	6
Jon	4	6’5	8	7	5	5’5

a. Nork jasan du traumarik handiena bere gutxiegirekin? Hau da, gutxiegien artean, zein izan da, erlatiboki, emaitzarik txarrena?
b. 1. lauhilabetekoan zehar Jonek 20, 34, 45, 30, 30 eta 27 orduz ikasi zuen irakasgai bakoitza prestatzeko, hurrenez hurren. Kalkula ezazu orduen karratu txikienen erregresio-zuzena noten arabera.
c. Jonek, Estatistika I irakasgaian 5 ateratzeko, zenbat orduz ikasi beharko du?

3. GAIA: DESBERDINTASUN NEURRIAK.

3.1. Giniren indizea.
3.2. Lorenzen kurba.

3.1. Giniren indizea.

Hizkuntza arruntean sakabanatze eta kontzentrazio hitzek kontrako esanahiak izan arren, estatistikan erabilitakoei esanahi hauek ez zaizkie egokitzen.

Estatistikaren ikuspuntutik, sakabanatzea datuen aldakortasunari dagokio, hots, batezbestekoen adierazgarritasun txiki edo handiari. Kontzentrazioaren neurketek, ordea, aldagaiaren balio guztien arteko desberdintasunaren maila neurtzea dute helburu. Esaterako, herri bateko biztanleen errentak. Demagun, biztanleen arteko errentaren banaketa berdina ote den aztertu nahi dela; bestela esanda, errenten artean dagoen desberdintasun maila neurtu nahi dela.

Hau da, errentak $x_1 \leq x_2 \leq ..... \leq x_k$ izanik, jakin nahi da $\sum x_i n_i$ nola banatzen den biztanleriaren artean.

Beraz, kontzentrazioari erreparatzen bazaio, gertatu daitezkeen egoera guztiak honako bi muturreko egoera hauetan kokatuko dira beti:

---

a. **Kontzentraziorik handiena:** biztanle guztietatik batek, soilik, irabazten du errenta guztia.

$$x_1 = x_2 = ... = x_{k-1} = 0 \quad \text{eta} \quad x_k \neq 0$$

b. **Kontzentraziorik txikiena edo berdinbanatzea:** biztanle guztiek soldata berdina jasotzen dutenean. Hau da, $x_1 = x_2 = ... = x_k$ denean.

Beraz, egoera horien zenbakizko neurketa egitekotan, Giniren indizea definitu behar da. Aldagai jakin baten indizea kalkulatzeko honako hau egin behar da:

Demagun *X* aldagaiaren maiztasun-banaketaren kontzentrazioa aztertu nahi dela.

X_i	n_i	$X_i.n_i$	N_i	U_i	p_i	q_i
x_1	n_1	$x_1.n_1$	N_1	U_1	p_1	q_1
.	.	.	.	.	.	.
.	.	.	.	.	.	.
x_k	n_k	$x_k.n_k$	N_k	U_k	p_k	q_k

1. $x_i.n_i$, biderkadurek x_i soldatako biztanleek jasotzen duten errenta osoa adierazten dute.

2. N_i maiztasun absolutu metatuak.

3. U_i, kantitate metatuak.

$$U_1 = x_1 n_1$$
$$U_2 = x_1 n_1 + x_2 n_2$$
.....
$$U_k = \sum_{i=1}^{k} x_i n_i \quad \text{Errenta osoa.}$$

4. Zutabe horretan maiztasun metatuak ehunekotan adieraziko dira. $p_i = \frac{N_i}{N} 100$

5. Masa edo ehuneko metatua. Hots, $q_i = \frac{U_i}{U_k} 100$.

Aurreko bost urrats kalkulatuz gero honako formula hau erabiliz kalkulatzen da Giniren indizea:

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

Kontzentraziorik handienaren egoeran, $I_G = 1$ izango da, indizearen zenbakitzaileak eta izendatzaileak balio bera baitute. Kontzentraziorik txikienaren egoeran, berriz, $I_G = 0$ izango da, indizearen zenbakitzaileak *0* balioa hartzen baitu. Hau da, edozein *i*-tarako p_i eta q_i balioak berdinak dira eta.

Beraz, edozein banaketaren indizea *0*tik hurbil egoteak kontzentrazio txikia adierazten du eta *1*etik hurbil kokatzeak kontzentrazio handia erakusten du.

PROPIETATEAK

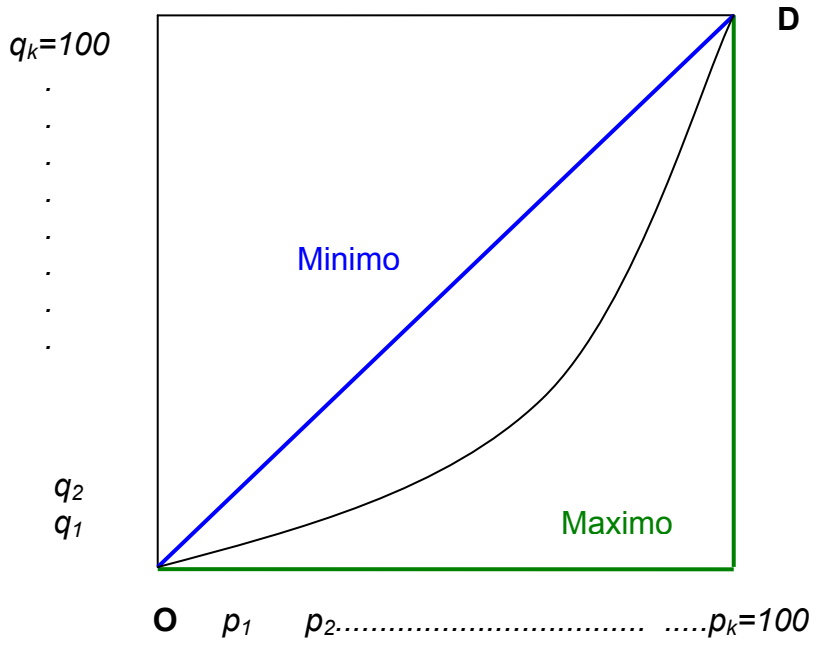
1. I_G indizeari unitate-aldaketak ez dio eragiten. Hau da, kontzentrazioari dagokionez, X aldagaiaren honako transformazio hau izateak ez du egoera aldatzen: $Y = aX$.

2. Translazioa, ordea, indizeari eragiten dio. Esaterako, honako transformazio hau egitean: $Y = X + a$, desberdintasun-egoerak murriztu daitezke. Bestalde, $Y = X - a$ transformazioa eginez gero, kontrakoa gertatuko da; desberdintasuna handitu daiteke, alegia.

3.2. Lorenzen kurba.

Banaketaren kontzentrazioa, grafikoki ikusarazten duen teknika bat da. Lorenzen kurba eraikitzeko, aurreko taula hartuko dugu kontuan. Taula egin ondoren, (p_i, q_i) bikote-balio guztiak sistema kartesiar batean grafikoki adieraziz eta bikote guztiak elkarren artean zuzen baten bidez lotuz lortzen dugu Lorenzen kurba.

Eskalak bi ardatzetan berdinak direnez gero, Lorezen kurba karratu baten barruan aurkitzen da beti. Jatorritik irteten den OD diagonala erabiliko dugu uniformetasunaren-hots, kontzentrazio minimoaren-egoera adierazteko. Hau da, multzoko puntu guztiak OD diagonalean badaude, kontzentrazio minimoa izango dugu. Beraz, kurbaren eta diagonalaren artean zenbat eta tarte handiago egon, orduan eta kontzentratuagoa izango da banaketa; bestela esanda, kurba diagonaletik zenbat eta hurbilago egon, orduan eta uniformeagoa edo uniformetasunerako joera handiagoa du.



<u>Adibidea</u>.

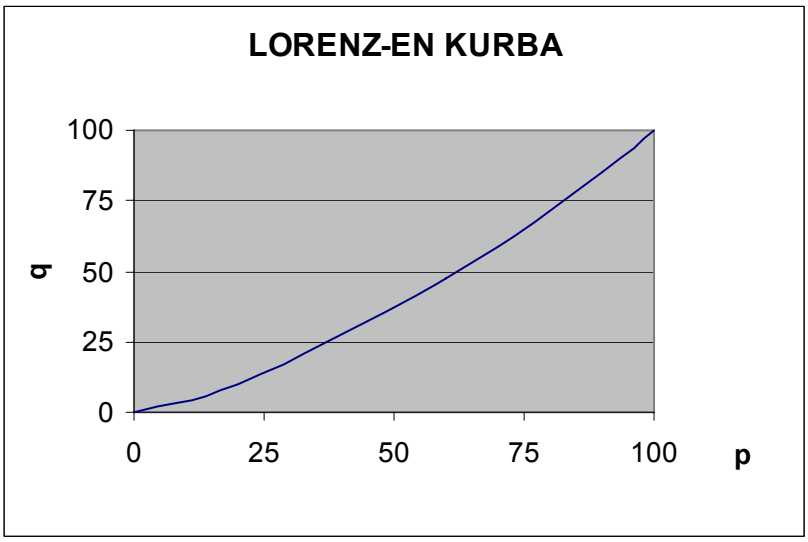
Suposa dezagun enpresa bateko langile-soldaten banaketa honako hau dela:

X Soldatak (€-tan)	n_i langile-kopurua
480	10
900	20
1.200	15
1.450	5

Banaketari begira, zer esan dezakegu soldaten kontzentrazioari buruz?

$X_i n_i$ Soldaten masa	N_i langile kopuru metatua	p_i	q_i
4.800	10	20	9´99
18.000	30	60	47´45
18.000	45	90	84´91
7.250	50	100	100

Lorenzen kurba eraikitzeko, alde batetik, p_i-ren balioak markatuko ditugu abzisa-ardatzean eta, bestetik, ordenatu-ardatzean q_i-renak. Bikote bakoitzaren balioak koadrantean markatu eta gero, lortutako puntuak lotuz margotzen da Lorenzen Kurba.



Grafikoari begira esan dezakegu gure banaketan kontzentrazio minimoan ez dagoela baina maximoan ere ez. Grafikoaren arabera kontzentrazio minimotik maximotik baino hurbilago dago, antza. Beraz, esandakoa egia izanez gero, Giniren indizea kalkulatzen badugu 0tik hurbil egongo da indizearen balioa.

$$I_G = \frac{\sum_{i=1}^{k-1}(p_i - q_i)}{\sum_{i=1}^{k-1} p_i} = 0'16$$

Izan ere, indizearen balioa (0´16) 0tik 1etik baino hurbilago dago.

Ariketak

1. Suposa ezazu Ceres enpresako etekina 6.000 € dela eta honela banatzen dela: % 22 A bazkidearentzat; % 28 B bazkidearentzat; % 24 C bazkidearentzat, eta gainerakoa, D bazkidearentzat.

Ulises enpresak bere 12.000 €-ko etekina honela banatzea erabaki du: 2.400 € lehenengo bazkidearentzat; 3.600 bigarrenarentzat, eta gainerakoa, beste bi bazkideentzat erdi bana.

Banaketa bietatik, zein da bidezkoena?

2. Enpresa baten etekin-banaketan, langileen % 10ek etekinaren % 25 hartu du; % 20k etekinaren % 15 hartu du, eta % 35ek etekinaren %20.
Kalkula ezazu honako hau:

a) Banaketa horren berdintasun edo desberdintasuna komentatzeko indize egokia.
b) Etekinaren zein ehuneko hartzen du langileen % 90ek?.

3. Enpresa txiki baten soldata-banaketa honako hau da:

Kategoria	Soldata €	Langile kopurua
Langileak	300-600	9
Administrariak	600-850	8
Zuzendariak	850-1200	3

a) Lor ezazu soldataren kontzentrazio-maila.
b) Enpresak enplegatu guztien soldatak % 5 igotzea erabakitzen badu, zein da soldata berrien kontzentrazio-maila?.
c) Soldataren zein ehuneko hartzen dute langileek?
d) Soldataren zein ehuneko hartzen dute gehien irabazten duten % 55ek?.

4. GAIA: MAIZTASUN BANAKETA DIMENTSIO BIKOA

4.1. Zenbakizko deskribapena.
4.2. Bazter-banaketak eta banaketa baldintzatuak.
4.3. Independentzia estatistikoa.

4.1. Zenbakizko deskribapena.

Orain arte aztertu diren fenomenoetan, banako bakoitzari ezaugarri estatistiko bati buruzko neurketa bat besterik ez zitzaion egiten. Hau da, dimentsio bakarreko aldagaiarekin lan egin da.

Ikasgai honetan aipatuko diren ikerketetan, populazioko banako bakoitzari bi neurketa egingo zaizkio, bakoitza ezaugarri edo aldagai estatistiko bati dagokiona; hots, dimentsio biko aldagaiarekin lan egingo da.

Demagun populazio baten errenta (X) eta gastua (Y) batera ikertu nahi direla, zein harreman mota dagoen haien artean, alegia. Horretarako, batera aztertzekotan aldagai bikoa (X, Y) definitu behar da. Beraz, haren zenbakizko deskribapena (X, Y) aldagai bikoa bada oro har bi aldagai kontuan hartuz, honako dimentsio biko taula hau osatuko da.

Y \ X	x_1	x_2	.....	x_i	.....	x_k	
y_1	n_{11}	n_{21}	.....	n_{i1}	.....	n_{k1}	$n_{\bullet 1}$
y_2	n_{12}	n_{22}	.....	n_{i2}	.....	n_{k2}	$n_{\bullet 2}$
.....	.....	.....	.....	.....	.....	.....	.....
y_j	n_{1j}	n_{2j}	.....	n_{ij}	.....	n_{kj}	$n_{\bullet j}$
.....	.....	.....	.....	.....	.....	.....	.....
$y_{k'}$	$n_{1k'}$	$n_{2k'}$	.....	$n_{ik'}$	.....	$n_{kk'}$	$n_{\bullet k'}$
	$n_{1.}$	$n_{2.}$	.....	$n_{i.}$	.....	$n_{k.}$	N

n_{ij} → balio-bikotearen maiztasun absolutuak. Aldagai bikoan (x_i, y_j) balio bera hartzen duten banako kopurua.

$$\sum_{i=1}^{K}\sum_{j=1}^{K'} n_{ij} = N$$

Taula horri korrelazio-taula deitzen zaio, aldagaiak kuantitatiboak direnean; aldagaiak kualitatiboak izanez gero, berriz, kontingentzia-taula.

4.2 Bazter-banaketak eta banaketa baldintzatuak.

Demagun errenta (*X*) aldagaiak eta gastu (*Y*) aldagaiak honako dimentsio biko maiztasun-banaketa hau osatzen dutela:

Y \ X	10	15	30	$n_{.j}$
1	4	2	0	6
3	1	5	1	7
7	0	1	6	7
$n_{i.}$	5	8	7	20

Gastuaren zein errentaren ikerketa bakartua eginez gero, bakoitzaren dimentsio bakarreko banaketa hartu beharko da kontuan. Bestela esanda, aldagai biko analisian X eta Y aldagaien bazter banaketak, hurrenez hurren.

BAZTER BANAKETAK

X eta Y aldagaiek hartzen dituzten balioak eta maiztasunak aintzakotzat hartuz lortzen dira *X*-ren eta *Y*-ren bazter-banaketak :

Errenta (X)	Perts. kopurua (n_i)
10	5
15	8
30	7

Gastua (Y)	Perts. kopurua (n_j)
1	6
3	7
7	7

Nolabait, bazter-banaketa definitzen dugunean besteen portaerak alde batera utziko dira.

$$n_{i.} = \sum_{j=1}^{K'} n_{ij} \Rightarrow \sum_{i=1}^{k} n_{i.} = \sum_{i=1}^{k} \sum_{j=1}^{k'} n_{ij} = N$$

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

Oro har, honela izango dira:

X	$n_{i.}$	$f_{i.} = n_{i.}/N$
X_1	$n_{1.}$	$f_{1.} = n_{1.}/N$
X_2	$n_{2.}$	$f_{2.} = n_{2.}/N$
.....	.....	.....
X_i	$n_{i.}$	$f_{i.} = n_{i.}/N$
.....	.....	.....
X_k	$n_{k.}$	$f_{k.} = n_{k.}/N$
	N	1

BANAKETA BALDINTZATUAK

Demagun aurreko aldagaiei begira (gastua eta errenta), 3 gastatu dutenean errentaren banaketa jakin nahi dela; hots, baldintza hori- 3 gastatzea- betetzen dutenen banaketa.

$X/Y = 3$	Perts. kopurua
10	1
15	5
30	1

Oro har:

$X/Y=y_j$	$n(X/Y=y_j)$	**f (X/Y=y$_j$)**
x_1	n_{1j}	$f_{1j} = n_{1j}/n_{.j}$
x_2	n_{2j}	$f_{2j} = n_{2j}/n_{.j}$
.	.	.
.	.	.
x_k	n_{kj}	$f_{kj} = n_{kj}/n_{.j}$
	$n_{.j}$	1

Modu berean, errentaren edozein baliok gastuaren banaketa baldintzapean jar dezake. Adibidez:

$Y/X=10$	Perts. kopurua
1	4
3	2
7	0

Bazter-banaketak bezala, banaketa baldintzatuak dimentsio bakarreko banaketak dira; beraz, orain arte azaldutako neurketak banaketa baldintzatuetan ere kalkula daitezke. Hori bai, kontuan hartu beharko dugu, baldintza hori egiaztatzen duten banakoen kopurua mugatua dela.

Oro har, banaketa baldintzatuetan batezbestekoa eta bariantza honela adieraziko dira.

$$\bar{x}/y = \frac{1}{n_{i.}}\sum_{i=1}^{k} x_i n_{ij} \qquad S^2_{x/y} = \frac{1}{n_{i.}}\sum (x_i - \bar{x}/y)^2 n_{ij}$$

Aurreko adibidean, gastu aldagaiaren balio guztiez baldintzatutako errenta aldagaiaren batezbestekoak kalkulatuz gero, hiru azpipopulazioren errentaren batezbestekoak lortuko dira populazio osorako.

$$\bar{x}/y_1 = \frac{70}{6}; \qquad \bar{x}/y_2 = \frac{115}{7}; \qquad \bar{x}/y_3 = \frac{195}{7}$$

Batezbestekoaren propietatetik bat aplikatuz, populazio osoaren batezbestekoa kalkula daiteke.

$$\bar{x} = \frac{6\frac{70}{6} + 7\frac{115}{7} + 7\frac{195}{7}}{20} = 19$$

4.3. Independentzia estatistikoa.

X eta *Y* aldagaiak kontuan hartuz, aldagai baten balioak aldatzerakoan bestearena aldatzen bada, esango dugu biak elkarren arteko mendekoak direla.

Bi aldagai estatistiko askeak izango dira honako propietate hau betetzen badute:

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

Bi aldagai askeak izateak bata bestearen baldintzapean ez dagoela adierazten du, hots, batak ez du bestea baldintzatzen. Beraz, X eta Y elkarrekiko askeak baldin badira bakoitzaren banaketa baldintzatua haren bazter-banaketa izango da.

Ariketak

1. Egunkari bateko asteroko gehigarriak lanean dagoen jendearen poztasuna eta egoera zibila ezaugarriak aztertzen ditu. Artikuluan, honako taula hau agertzen da:

	OSO POZIK	ERDI POZIK	POZTASUN GUTXI	POZTASUN BATERE EZ
EZKONDUAK	351 % 63	242 % 57	73 % 50	36 % 43
ALARGUNAK	44 % 8	31 % 7	7 % 5	5 % 6
DIBORTZIAT.	68 % 12	55 % 13	16 % 11	13 % 15
BEREIZIAK	15 % 3	12 % 3	5 % 3	6 % 7
INOIZ EZ EZKONDUAK	77 % 14	83 % 20	45 % 31	24 % 29

Artikulu horren egileek honako ondorio hau eratortzen dute: ezkonduta dagoen jendea besteak baino alaiago dago lanean, oso pozik dagoen jendearen % 63 baita. Zuzena da ondorio hori?.

2. Taula honetan 50 familiak dituzten seme-alabak eta auto kopuruak laburtzen dira:

SEME-ALABA KOP.	0	0	0	1	1	2	2	2	3	3
AUTO KOP.	0	1	2	1	2	1	2	3	2	3
FAMILIA KOP.	2	4	3	7	9	5	10	3	6	1

(a) Kalkula ezazu:

(i) Familia horien auto kopuruaren batezbestekoa.

(ii) Seme-alabarik ez duten familien auto kopuruaren batezbestekoa.

(b) Komenta ezazu baieztapen hauek egiazkoak edo faltsuak diren arrazoibide estatistikoetan oinarrituz:

(i) Auto bat duten familien ehunekorik handiena, 2 seme-alaba dituztenen artean gertatzen da.

(ii) 2 auto dituzten familien % 20k 2 seme-alaba ditu.

(iii) Familien % 20k 2 auto eta 2 seme-alaba ditu.

3. Euskal Urtekari Estatistikotik (EUSTAT), honako taula hau atera dugu:

2002ko BPG	**Euskadi**	**Araba**	**Bizkaia**	**Gipuzkoa**
Nek. eta Arran.	64.190	17.014	25.723	21.453
Industria	1.315.517	236.886	633.530	445.101
Zerbitzuak	1.395.207	184.521	785.425	425.261
Ekonomia osoa	2.774.914	438.421	1.444.678	891.815

Datu horietan oinarrituta, erantzun itzazu galdera hauek:

(a) Azter ezazu nola dagoen banatuta Barne Produktu Gordina sektoreen artean, (ehunekotan) herrialde bakoitzerako eta konpara itzazu banaketa baldintzatu horiek, Euskadi osoaren bazter-banaketarekin.

(b) Aurreko emaitzak kontuan hartuta, aldagai biak estatistikoki askeak direla uste duzu?

(c) Zein ehuneko adierazten du Arabako industria sektorearen BPGk Euskadiko Ekonomia Osoarekiko? Gipuzkoakoa baino handiago edo txikiagoa da?

5. GAIA: ALDAGAI KUANTITATIBOEN ARTEKO ERLAZIOAREN ANALISIA.

5.1. Aldagai kuantitatiboen arteko erlazioaren analisia: kobariantza.
5.2. Transformazio eta konbinazio linealak.
5.3. Korrelazioa R^{2-}an.

5.1. Aldagai kuantitatiboen arteko erlazioaren analisia: kobariantza.

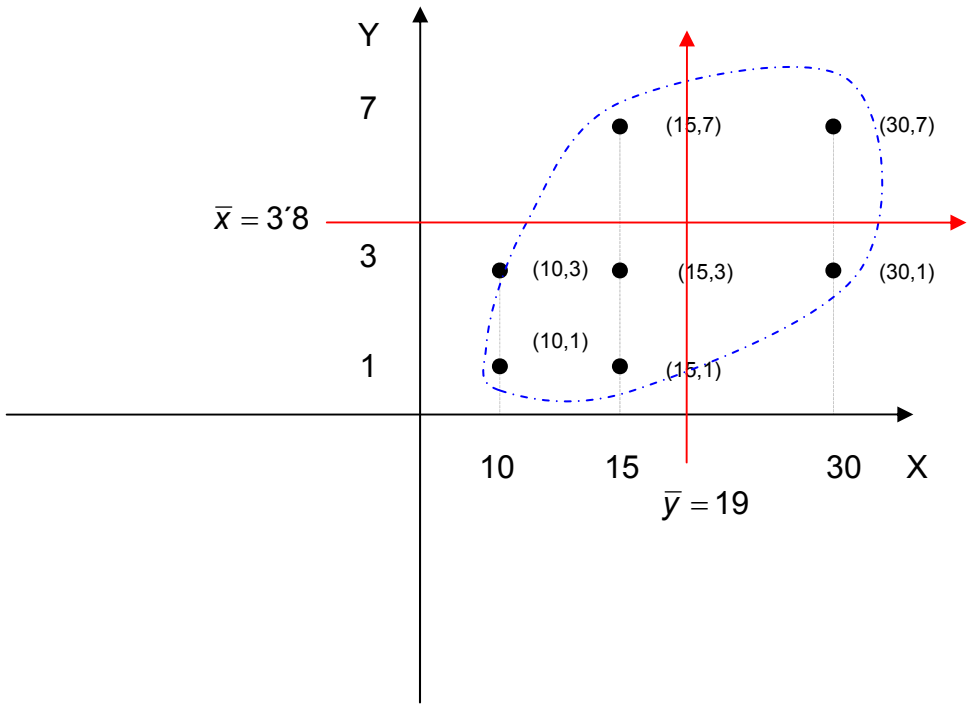
Bi muturreko egoeren artean (mendekotasun funtzionala eta independentzia estatistikoa) bitarteko egoerak ere egon daitezke; eta nahiz eta egoera horietan aldagaien arteko erlazioa funtzio baten bidez adieraztea zaila izan, horien arteko erlaziorik egon badago.

Oro har, fenomeno ekonomikoak ez dira deterministak izaten; hau da, ez dira eredu zehatzaren bidez azaltzen. Beraz, bi aldagairen arteko erlazioa adierazterakoan erabiliko da mendekotasun estatistikoaren kontzeptua.

Mendekotasun estatistikoan zenbait maila dago, aldagaien arteko erlazioa arinagoa edo indartsuagoa izan baitaiteke. Beraz, jarraian, bi aldagairen arteko mendekotasuna neurtzen duten neurriak definituko dira.

Aurreko gaiaren errentaren eta gastuen banaketaren puntu-hodeia kontuan hartzen bada, X eta Y aldagaiak positiboki erlazionatuta daudela nabari daiteke, errentaren balio handienei gastuaren balio handiak ere baitagozkie. Erlazioaren neurria lehen urratsa eraikitzeko honako hau izango da: aldagai zentratuen balioak gogoan hartuko dira, jatorrizko balioak kontuan hartu beharrean. Hala,

desbideratze positiboak zein negatiboak aztertuz, berehala jakingo da nola mugitzen diren aldagaiak.



Orain, $(x_i - \bar{x})(y_j - \bar{y})$ desbideratzeen biderkadura kontuan hartuz gero, zeinu positiboak adierazten du bi balioak batera beren batezbestekoen gainetik edo azpitik kokatzen direla; zeinu negatiboak, berriz, aldagai bakoitzaren norabidea kontrakoa dela, eta bata handitzen den bitartean bestea txikitzen dela; hau da, bata batezbestekoaren gainean kokatzen da eta bestea, berriz, azpian.

Izan ere, biderkadura horien bidez, banaketaren balio bikote bakoitzerako konparaketak egin daitezke; alabaina, datu guztietarako balio bakarra ematen

duen neurria lortu nahi da. Neurri horri kobariantza deituko zaio eta S_{XY} ikurraz adieraziko da. Honako hau izango da:

$$S_{XY} = \sum_{i,j} (X_i - \bar{x})(Y_j - \bar{y}) \frac{n_{ij}}{N} = \sum_{i,j} (X_i - \bar{x})(Y - \bar{y}) f_{ij}$$

edo honela ere adieraz daiteke:

$$S_{XY} = \frac{1}{N} \sum_{ij} x_i y_j n_{ij} - \bar{x}\bar{y}$$

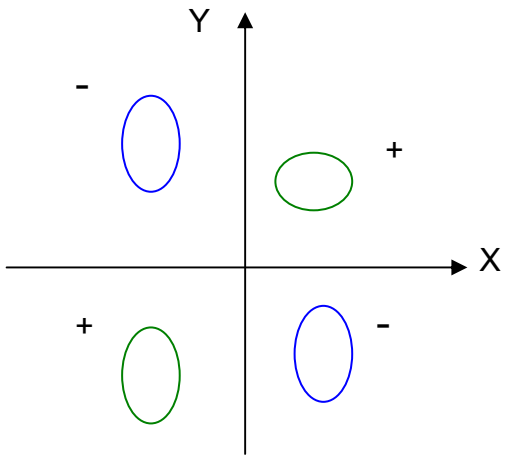
Aurreko adibidean:

$(x_i - \bar{x})$	$(y_j - \bar{y})$	f_{ij}	$(x_i - \bar{x})(y_j - \bar{y}) f_{ij}$
-9	-2´8	0´2	5´04
-9	-0´8	0´05	0´36
-4	-2´8	0´1	1´12
-4	-0´8	0´25	0´8
-4	3´2	0´05	-0´64
11	-0´8	0´005	-0´44
11	3´2	0´3	10´56

Beraz, bi aldagairen arteko baterako aldakuntzaren neurria da kobariantza, kontuan hartuta bi aldagairen arteko erlazio lineala baino ez duela adieraziko.

Bi aldagai desberdinen arteko kobariantzaren zeinua positiboa zein negatiboa izan daiteke, eta zeinuak erlazioaren norabidea adieraziko du.

Kobariantzaren zeinua positiboa izateak honako hau adierazten du: batugai positiboek negatiboek baino gehiago pisatzen dute. Kobariantzaren zeinu negatiboak, berriz, kontrakoa ageri du.



Puntu gehienak nola kokatzen diren arabera, S_{xy}-k zeinu positibo edo negatiboa hartuko du. Kontuan hartu behar da bi aldagairen artean dagoen erlazio lineal mota baino ez duela ematen; hots, norabide bera edo kontrakoa.

Norabide bera edo mendekotasun lineala zuzena izatean: $S_{xy}>0$
Norabide kontrakoa edo alderantzizko mendekotasun lineala izatean: $S_{xy}<0$.

X eta *Y* aldagaiak askeak izanez gero, haien artean ez da harremanik egongo; beraz, erlazio linealik ere ez bereziki. Hori dela eta, kobariantzaren balioa 0 izango da.

Oro har, *X* eta *Y* aldagaiak askeak izanez gero, $S_{XY} = 0$ izango dela frogatuko da. Hau da, *X* eta *Y* askeak baldin badira, $f_{ij} = f_i \bullet f_j$ betetzen da, beraz:

$$S_{XY} = \left[\sum_i (x_i - \bar{x}) f_i \right] \bullet \left[\sum_j (y_j - \bar{y}) f_j \right] = 0$$

Baina, alderantzizkoa ez da betetzen; hau da, bi aldagairen kobariantza *0* izan arren ez dute askeak izan behar.

5.2. Transformazio eta konbinazio linealak.

Bi aldagairen transformazio linealen kobariantza--- Jatorri-aldaketari dagokionez S_{xy} askea izango da; alabaina, eskala-aldaketari erreparatzen bazaio ez. Hau da:

$$U = aX + b$$
$$V = cY + d$$

$$S_{UV} = \frac{1}{N} \sum_{ij} (u_i - \bar{u}) \bullet (v_j - \bar{v}) n_{ij} = \frac{1}{N} \sum_{ij} (ax_i + b - a\bar{x} - b) \bullet (cy_j + d - c\bar{y} - d) n_{ij} =$$

$$= \frac{1}{N} ac \sum_{ij} (x_i - \bar{x}) \bullet (y_j - \bar{y}) n_{ij} = acS_{XY}$$

Bi aldagairen konbinazio linealaren bariantza--- Z aldagaia honela definituz gero $Z = aX + bY$, honako hau izango da Z aldagaiaren bariantza:

$$S_Z^2 = \frac{1}{N}\sum(z_i - \bar{z})^2 = \frac{1}{N}\sum\ [(ax_i + by_i) - (a\bar{x} + b\bar{y})]^2 = \frac{1}{N}\sum[a(x_i - \bar{x}) + b(y_i - \bar{y})]^2 =$$

$$\frac{1}{N}\sum\left[a^2(x_i - \bar{x})^2 + b^2(y_i - \bar{y})^2 + 2ab(x_i - \bar{x})(y_i - \bar{y})\right] = a^2 S_X^2 + b^2 S_Y^2 + 2abS_{XY}$$

5.3. Korrelazio lineala R^2-an.

Korrelazio-koefizienteak, kobariantza bezala, *X* eta *Y* aldagaiek duten erlazioa neurtzen du. Baina, kobariantzaren alde, dimentsio gabeko neurria denez gero aldagaien arteko erlazio mota adierazteaz gain, erlazioaren maila ere adierazten du. Formula honen bidez kalkulatzen da:

$$r_{XY} = \frac{S_{XY}}{S_X S_Y}$$

PROPIETATEAK

- Neurri-unitaterik ez du, hots, dimentsiogabea da. Aldiz, kobariantza aldagaien neurri-unitateen araberakoa denez gero, banaketen neurri-unitateak desberdinak baldin badira, dependentzia linealaren maila ez dago alderatzerik.

- Ez unitate-aldaketak ez jatorri aldaketak ez diote eragiten. Hau da, *(X,Y)* aldagaien ordez horiekin erlazionatuta dauden *(U, V)* aldagaiak erabiliz gero, azalduko da honako erlazio hauek ez diotela korrelazio-koefizienteari eragiten:

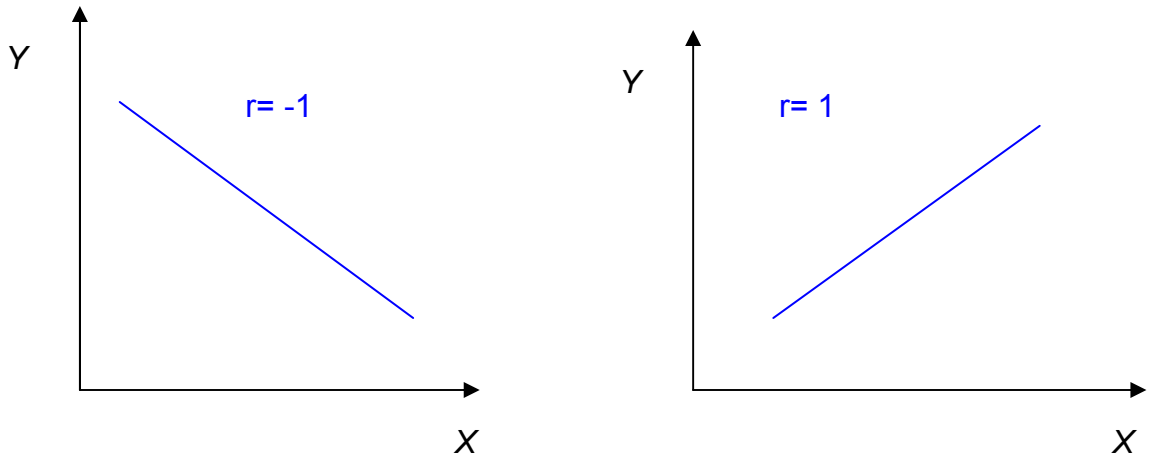
U=aX+b eta V=*cY+d* izanik, lehenengo eta behin gogora ekarri behar da desbideratze tipikoetan eta kobariantzan *(U,V)* aldagaien eraginak zein diren:

$$S_U=aS_X; \quad S_V=cS_Y; \quad S_{UV}=acS_{XY}$$

Hiru kasuetan, eragina unitate-aldaketa baino ez denez, korrelazio-koefizientea erabiliz:

$$r_{UV}=\frac{S_{UV}}{S_U S_V}=\frac{acS_{XY}}{aS_X cS_Y}=r_{XY}$$

- Korrelazio-koefizienteak har ditzakeen balioak $[-1\ ,\ 1]$ tartekoak dira. Zero balioa hartzeak adierazten du aldagaien arteko erlazio linealik ez dagoela eta 0ra hurbilduz gero erlazioa ahula dela. Bestalde, balio absolututan, 1etik zenbat eta hurbilago egon, orduan eta handiagoa izango da erlazioa. Korrelazio-koefizienteak -1 edo 1 balioa duenean, bi aldagairen arteko erlazio lineala zehatza izango da, alderantzizkoa edo zuzena. Hau da:



Orain arte, mendekotasun lineala hartu da aintzat, ikerketa honek zentzurik ba ote daukan kontuan hartu gabe. Kasu batzuetan, estatistikoki bi aldagairen artean korrelazioa izan arren, ez dago justifikatuta. Egoera batzuetan, bi aldagairen arteko korrelazioa hirugarren baten eragina baino ez da. Beraz, kontuz eta zentzuz ibili behar da.

Ariketak

1. 100 familiaren kontsumoa (k) eta hileroko errenta (y) honako hauek dira:

k / **y**	**15**	**25**	**35**	**45**
30	10	15	--	--
40	5	20	25	--
50	--	15	5	5

Kalkula ezazu korrelazio-koefiziente lineala eta azaldu. Korrelazio-koefizientea kontuan hartuz, esan daiteke aldagaiak askeak direla?

2. Demagun aldagai biko maiztasun-banaketa hau:

y / **x**	**3**	**4**	
0	42	28	70
1	18	12	30
	60	40	100

(a) Askeak al dira bi aldagai horiek?.

(b) Eragiketarik egin gabe, jakin dezakegu zein den korrelazio-koefiziente lineala?. Azal ezazu zergatik.

3. Ordezkoen diren bi produkturen prezioak jaso dira: X, A artikuluaren prezioa; eta Y, B artikuluaren prezioa. Honako taula honetan, 50 behaketa laburtzen dira:

Y / **X**	**80**	**90**	**96**
60	7	8	2
70	3	12	8
80	**-**	2	8

Zera eskatzen da:

(a) Aldagaien arteko kobariantza eta korrelazioa.

(b) Prezioetan honako aldaketa hauek gertatuz gero:

Y' = 0´8 Y

X' = 10 + 1´1 X

maiztasun-banaketa berdin mantenduz, X' eta Y' arteko kobariantza eta korrelazioa.

4. Pertsona talde bati honako ezaugarri hauetaz galdetzen zaio:

X : 1993ko errentaren aitorpenean aitortutako diru-sarrera.

Y : 1993ko benetako diru-sarrera.

(a) Nola espero duzu izatea aldagai bien arteko kobariantza? Zergatik?

(b) Azal ezazu zein izango den korrelazio-lineala koefizientearen balioa honako kasu hauetan:

i) guztiz zintzoak izan diren pertsonentzat.

ii) errealitatearekiko guztiz independenteak diren sarrerak aitortzen dituzten pertsonentzat.

(c) Ogasunari dioten beldurra dela-eta, pertsonek irabazten dutenaren % 10 baino gutxiago aitortzen badute, zein izango da korrelazio-koefiziente lineala?

6. GAIA: ALDAGAI EZ KUANTITATIBOEN ARTEKO ERLAZIOAREN ANALISIA.

6.1. Elkarketa-koefizientea.
6.2. Korrelazio ordinala: koefizientea

6.1. Elkarketa-koefizientea.

Orain arte, bi aldagai kuantitatiboren arteko erlazioa, hots korrelazioa, azaldu da. Gai honetan, berriz, bi aldagai kualitatiboren edo kualitatibo baten eta kuantitatibo baten arteko erlazioa jorratuko da.

Aldagai kuantitatiboetan bezala, kontingentzia-taula erabiliz, bi atributu askeak diren edo ez jakin daiteke; independentzia-baldintzaren bidez, baldintza horretan aldagaien balioek ez baitute parte hartzen, baterako maiztasunek eta bazter-maiztasunek baizik.

Bi atributu independenteak ez direla frogatu eta gero, haien arteko elkarketa-maila edo mendekotasuna kuantifika daiteke, zenbait elkarketa-neurri erabiliz. Horrelakoak *Pearson-en* koefizientea azalduko da:

$$C = \left(\frac{\chi^2}{N + \chi^2} \right)^{\frac{1}{2}}$$

$$\chi^2 = \sum_{ij} \frac{\left[n_{ij} - \frac{n_{i.} n_{.j}}{N} \right]^2}{\frac{n_{i.} n_{.j}}{N}}$$

Koefiziente horrek, behatutako banaketa eta independentzia-baldintzaren menpe dagoen banaketa alderatzen ditu. Independentziaren egoerarekiko zenbat diferentzia dagoen, erlatiboki, neurtuko du; beraz, χ^2 balioa zenbat eta handiagoa izan, orduan eta indartsuagoa izango da atributuen arteko elkarketa-maila. Dena dela, kontuan hartu behar da neurri hori ez dagoela bornatuta, banaketaren zutabe eta errenkaden araberakoa baita.

- *C*-ren zeinua positiboa izango da. $0 \leq C \prec 1$
- Aldagaiak askeak (independenteak) izanez gero, $\chi^2 = 0$ izango da.
- *C*-ren balioa zenbat eta handiagoa izan, orduan eta handiago izango da atributuen arteko elkarketa-maila.

6.2.Korrelazio ordinala (komunztadura): koefizientea.

Demagun aztertzen ari diren ezaugarriak *A* eta *B* direla. Bestalde, zehaztutako eskalaz txikitik handira *A* ezaugarria ordenatuz gero, kontuan hartuko da *A*-ri legokiokeen ordenako zenbakia edo heina X aldagaia izango litzatekeela; modu berean, *Y*-k *B*-ren heina adieraziko luke.

Ordenamenduaren irizpideak *A* eta *B* deitzen baditugu eta *X* eta *Y* haiei dagozkien heinei, Spearmanen korrelazio-koefizientea erraz lortuko da, r_{XY} korrelazio-koefizientetik.

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

$$r_s = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N^3 - N} \qquad d_i = x_i - y_i \text{ izanik}$$

- koefizientea *-1 eta 1*en artean aldatuko da.
- Komunztadura egokiena izatean, orduan $\sum_{i=1}^{k} d_i^2 = 0 \Rightarrow r_s = 1$
- Desadostasuna egokiena izatean, orduan $r_s = -1$

Adibidez: honako hauek dira 5 ikasleren zailtasunen araberako heinak estatistikaz eta ekonomiaz:

IKASLEAK	Estatistika x_i	Ekonomia y_i	d_i	di^2
A	1	5	-4	16
B	2	3	-1	1
C	3	4	-1	1
D	4	2	2	4
E	5	1	4	16

$$r_s = 1 - \frac{6\sum d_i^2}{N^3 - N} = 1 - \frac{6 \bullet 38}{120} = -0'9$$

Ariketak

1. Taula honetan adierazten dugu 20 pertsona etxe baten jabeak direnetz (1 = jabeak; 2 = alokatuak) eta udal zerbitzuez pozik dauden ala ez (1 = pozik ez; 2 = pozik; 3 = oso pozik):

JABEA/ALOKATUA	1	1	1	2	2
POZTASUNA	1	2	3	2	3
PERTSONAK	5	1	2	2	10

(a) Idatz itzazu datu horiek baterako maiztasunak ematen dituen taula gurutzatua.
(b) Laginaren zein ehuneko izango dira jabeak?
(c) Jabeen artean, zein ehuneko dago oso pozik udal zerbitzuekin?
(d) Kalkula daiteke korrelazio-koefizientea? Interpretaziorik ba al du? Ba al dago beste koefizienteren bat aldagaien arteko erlazioa dagoen edo ez hobeto azaltzen duena? Zein? Beste bat egonez gero, kalkula ezazu.

2. 150 gazteri inkesta bat egin ondoren, maiztasun-banaketa bateratu hau lortu da, X-k adina adierazita:

Y **X**	**1**	**2**	**3**
18	0	50	0
19	40	0	40
20	10	0	10

Y aldagaiaren esangurari buruzko bi aukera hauek kontuan hartuz, kalkula ezazu erlazio-koefiziente egoki bat,

(a) Proiektu arkitektoniko bati buruzko lehentasuna, kodifikazio honen arabera:

1 = eraikin bat zaharberritzea.

2 = bota eta berri bat egitea.

3 = egitura modernoa gainetik jartzea.

(b) Herri bateko jaietarako rock-talde bat kontratatzeko aurrekontu egokiaren lehentasuna, honako kode hauen arabera:

1 = 10.000 €.

2 = 20.000.€

3 = 30.000 €

3. Freskagarri enpresa batek urtearen 8 aste aukeratu ditu zoriz. Taula honetan, batez besteko tenperatura eta saldutako freskagarri kopuruak ageri dira:

Tenperatura	10	18	12	31	30	19	39	5
Freskag. kop.	21	65	19	72	75	39	67	11

(a) Irudika ezazu puntu-hodeia eta komenta ezazu.

(b) r_{xy} koefizientea erabiliko zenuke bi aldagaien arteko erlazioa aztertzeko? Baietz erantzunez gero, kalkula ezazu eta azaldu.

(c) Freskagarri kopurua izan beharrean, freskagarri mota izango bazenu, zein koefiziente erabiliko zenuke aldagai bien arteko erlazioa aztertzeko?Arrazoi ezazu zure erantzuna.

4. Oposizio-epaitegi batek A, B, C, D eta E opositoreak ordenatu ditu onenetik txarrenera, azterketa teorikoan eta azterketa praktikoan lortutako puntuen arabera. Esan daiteke azterketa teorikoa ondo egitean praktikoa ere ondo egingo dutela?

Azterketa teorikoa	**Azterketa praktikoa**
A	E
D	C
E	A
B	D
C	B

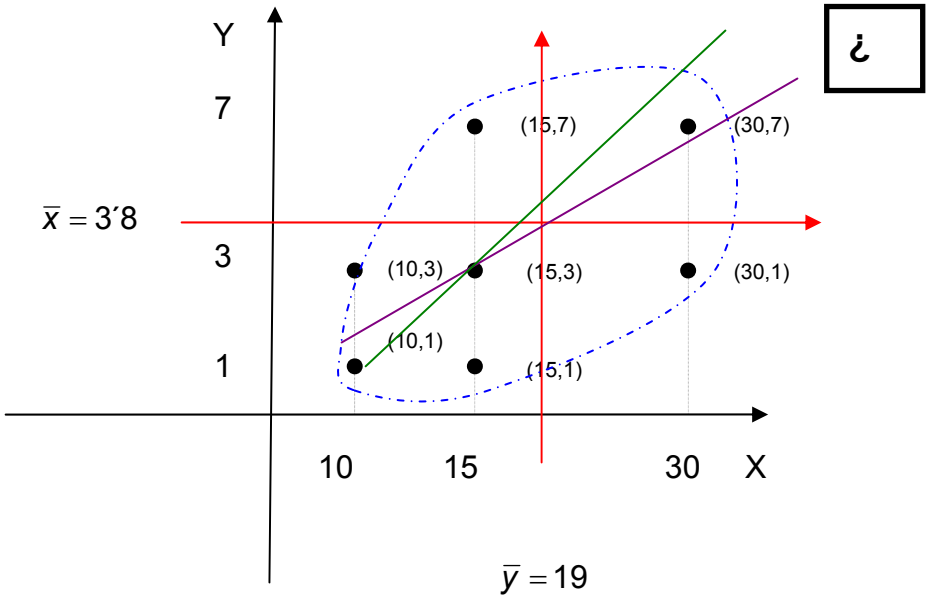
7. GAIA : BI ALDAGAIREN ARTEKO ERREGRESIOA

7.1. Karratu txikienen erregresio lineala R^2an.

7.2. Doikuntza-egokitasunaren analisia: mugatze-koefizientea.

7.1. Karratu txikienen erregresio lineala R^2an.

Korrelazio-analisiak populazio jakin baten zenbait aldagairen arteko mendekotasuna aztertzea du helburu. Haatik, erregresio-analisiak eredu matematikoa bilatzea du helburu; eredu horren bidez, alde batetik, aldagaien arteko mendekotasuna frogatzeko eta bestalde iragarpenak egiteko bereziki. Erabilitako adibidearen bidez korrelazioa azaltzeko, erregresio lineala jorratuko da.



Grafikoa aztertu eta gero, geure buruari galdetu behar diogu zein lerro izango den ezin hobea. **Zein lerro da ezin hobea?** Lerroa aukeratzerakoan, X aldagaiaren arabera Y aldagaiaren ahalik eta portaerarik onena azaltzen duen lerroa hartuko da gogoan.

Izan ere, X aldagaiaren arabera Y aldagaiaren portaera lerro zuzen baten bitartez azaldu nahi izanez gero, puntu-hodeiak mendekotasun hori adierazten du, y = a + bx zuzena bilatu beharko dugu; baina, x_i balio bakoitzerako Y aldagaiak bi balio desberdin hartuko ditu: y_i balio behatua eta $\hat{y}_i$ doitutako zuzenaren bitartez lortzen den iragarpena edo balio zenbatetsia. Beraz, aldagai berri bat sortzen da:

$$\hat{y}_i = a + bx_i$$

Grafikoan argi ikusten da balio behatuaren (y_i) eta balio aurresanaren ($\hat{y}_i$) artean diferentzia bat dagoela, x_i balio bakoitzerako. Diferentzia hori erregresioaren errorea edo hondarra da; errore hori erregresioan agertzen den beste aldagai berri bat izango da:

$$e_i = y_i - \hat{y}_i = y_i - a - bx_i$$

Karratu txikienen irizpidea zera da: errore horien karratuen (ikurren konpentsazioagatik) batura minimoa lortzea.

$$\sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} [y_i - a - bx_i]^2 .$$

a-ren eta *b*-ren arabera deribatu partzialak eginez gero:

$$\frac{\delta \sum e_i^2}{\delta a} = -2\sum [y_i - a - bx_i] = 0$$

$$\frac{\delta \sum e_i^2}{\delta b} = 2\sum [y_i - a - bx_i](-x_i) = 0$$

Bi atalak *(-2)*z zatituz gero:

$$\left.\begin{array}{r} \sum (y_i - a - bx_i) = 0 \\ \sum (y_i - a - bx_i)x_i = 0 \end{array}\right\} \Rightarrow$$

$$\left.\begin{array}{r} \sum y_i = Na + b\sum x_i \\ \sum x_i y_i = a\sum x_i + b\sum x_i^2 \end{array}\right\} \Rightarrow$$

N-z zatituz :

$$\left.\begin{array}{l} \bar{y} = a + b\bar{x} \\ \frac{1}{N}\sum x_i y_i = a\bar{x} + b\frac{1}{N}\sum x_i^2 \end{array}\right\} \Rightarrow$$

Lehenengo ekuazioak adierazten du erregresio zuzenak $(\bar{x}, \bar{y})$ batezbesteko balioen puntutik pasatzen dela, eta hortik *a* koefizientea lortzen da:

$$a = \bar{y} - b\bar{x} \qquad (1)$$

Eta bigarren ekuazioan *(1)* ordezkatuz:

$$\frac{1}{N}\sum x_i y_i = (\bar{y} - b\bar{x})\bar{x} + b\frac{1}{N}\sum x_i^2$$

$$\frac{1}{N}\sum x_i y_i = \overline{yx} - b\bar{x}^2 + b\frac{1}{N}\sum x_i^2$$

$$\frac{1}{N}\sum x_i y_i - \overline{yx} = b(\frac{1}{N}\sum x_i^2 - \overline{x}^2)$$

$$S_{XY} = bS_X^2 \quad \Rightarrow \quad b = \frac{S_{XY}}{S_X^2} \qquad (2)$$

$b = \frac{S_{xy}}{S_x^2} = r_{xy}\frac{S_y}{S_x}$ **erregresio-koefizientea**: *X*-ren unitate-aldaketaren arabera, *Y*-ren itxarondako aldakuntza.

(1) eta *(2)* emaitzekin erregresio zuzena lortzen dugu:

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

PROPIETATEAK.

1. $\overline{\hat{y}} = \overline{y}$

$$\overline{\hat{y}} = a + b\overline{x} = \overline{y} - b\overline{x} + \frac{S_{xy}}{S_x^2}\overline{x} = \overline{y} - \frac{S_{xy}}{S_x^2}\overline{x} + \frac{S_{xy}}{S_x^2}\overline{x} = \overline{y}$$

2. $S_{\hat{y}}^2 = bS_{xy}$

$$S_{\hat{y}}^2 = S_{a+bx}^2 = b^2 S_x^2 = \left(\frac{S_{xy}}{S_x^2}\right)^2 S_x^2 = \frac{(S_{xy})^2}{S_x^2} = bS_{xy}$$

3. $\overline{e} = 0$

$$e = y - \hat{y} \qquad \overline{e} = \overline{y} - \overline{\hat{y}} = 0$$

4. Hondar-bariantza:

$$S_e^2 = \sum \frac{(e_i - \overline{e})^2}{N} = \frac{\sum e_i^2}{N}$$

5. $S_{ex} = 0$

$$S_{ex} = \frac{1}{N}\sum(e_i - \overline{e})(x_i - \overline{x}) = \frac{1}{N}\sum e_i(x_i - \overline{x}) = \frac{1}{N}\sum e_i x_i - \overline{x}\frac{1}{N}\sum e_i = \frac{1}{N}\sum e_i x_i =$$

$$= \frac{1}{N}\sum(y_i - \hat{y}_i)x_i = \frac{1}{N}\sum(y_i - a - bx_i)x_i = 0$$

6. $S_{e\hat{y}} = 0$

$$S_{e\hat{y}} = \frac{1}{N}\sum e_i(\hat{y}_i - \overline{y}) = \frac{1}{N}\sum e_i(a + bx_i - \overline{y}) = \frac{1}{N}\sum e_i a + b\frac{1}{N}\sum e_i x_i - \overline{y}\frac{1}{N}\sum e_i = 0$$

7.2. Doikuntzaren egokitasun-analisia: mugatze-koefizientea.

Doikuntzaren egokitasun-analisia miatuz, $\hat{y} = f(x)$ ereduak ondo azaltzen duen aldagaien portaera jakin nahi da. Hori aztertzeko erregresioaren errorea izango da abiapuntua:

$$e_i = y_i - \hat{y}_i$$

Baina doikuntzaren egokitasuna ezin daiteke kalkula erroreen batezbestekoa erabiliz, *0* baita. Hori dela eta, erroreen bariantza edo hondar-bariantza erabiliko da; hots, **batez besteko errore koadratikoa** deritzoguna

$$\text{BEK} = \frac{\sum e_i^2}{N} = \frac{\sum (y_i - \hat{y}_i)^2}{N} \geq 0$$

BEK = 0 izanez gero, errore guztiek *0* balioa hartzen dutela ondorioztatuko da; beraz, $y_i = \hat{y}_i$.

Datu berdinak azaltzeko hainbat eredu erabil daitezke; batez besteko errore koadratikorik txikiena duen eredua izango da horietariko onena.

Alabaina, *BEK* ezin izango dugu beti erabili, hots:

- datu desberdinei doitutako ereduak konparatzeko, *BEK* neurri-unitateen araberakoa delako.
- datu multzo bati doitutako ereduaren egokitzapena ezagutzeko, *BEK*aren balioa mugagabea delako.

Kasu horietan, beste koefiziente bat bilatu beharko dugu:

$y = \hat{y} + e$ oinarrizko erlazioa izanik, $\hat{y}$ eta e korrelazio gabeak direnez gero, honako hau erator dezakegu:

$$s_y^2 = s_{\hat{y}}^2 + s_e^2$$

s_y^2: azaldu nahi dugun aldagaiaren bariantza

$s_{\hat{y}}^2$: erregresioak azaltzen duen bariantza

s_e^2: azaldu gabe geratzen den bariantza, hondar-bariantza

Erlazio horren bidez, *Y* aldagaiaren aldakuntza bi osagaitan oinarrituz interpreta daiteke; osagai batek erregresio-ereduak azaldutako bariantza adierazten du, eta besteak, azaldu gabe gelditzen den bariantza.

Erlazio hori S_y z zatituz, honako hau lortzen da:

$$1 = \frac{S_{\hat{Y}}^2}{S_Y^2} + \frac{S_e^2}{S_Y^2}$$

Hala, $\frac{S_{\hat{Y}}^2}{S_Y^2}$ kozienteak, ereduak azaldutako *Y*-ren aldakuntzaren zatia adierazten du; beraz, doikuntzaren egokitasuna egokia izango da neurritzat, eta **mugatze-koefizientea** esaten zaio.

$$R^2 = \frac{s_{\hat{y}}^2}{s_y^2}$$

Hau da, azaldu nahi dugun bariantza totalaren doitutako erregresioaz azaldutako bariantzaren proportzioa.

PROPIETATEAK

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

$$R^2 = \frac{S_{\hat{y}}^2}{S_y^2} = \frac{\frac{(S_{xy})^2}{S_x^2}}{S_y^2} = \frac{S_y^2 - S_e^2}{S_y^2} = 1 - \frac{S_e^2}{S_y^2}$$

2.- $R^2 = r_{xy}^2$

$$R^2 = \frac{S_{\hat{y}}^2}{S_y^2} = \frac{b^2 S_x^2}{S_y^2} = \frac{\left(\frac{S_{xy}}{S_x^2}\right)^2 S_x^2}{S_y^2} = \frac{S_{xy}^2}{S_x^2 S_y^2} = \left(\frac{S_{xy}}{S_x S_y}\right)^2 = r_{xy}^2$$

3.- $0 \le R^2 \le 1$

$$-1 \le r_{xy} \le 1 \quad \Rightarrow \quad 0 \le R^2 \le 1$$

Ariketak

1. Erregresio linealeko eredu batek gutxienez bariantzaren % 60 azalduz gero, ikertzaile batek zuzentzat hartuko du eredua. Haren lanean, honako emaitza hauek lortzen ditu:

$$\hat{y} = 7'36 + 0'18\,x \qquad S_{xy} = 248 \qquad S_y^2 = 49'60$$

Emaitza horiekin, zein litzateke ikertzailearen erabakia?

2. 5 familiaren iazko sarrerak (X) eta kontsumo-gastua (Y) (mila €tan) honako hauek izan dira:

Sarrera (X)	2	4	6	8	10
Kontsumoa(Y)	1	3	6	7	8

$\bar{x} = 6$; $\bar{y} = 5$; $S_x^2 = 8$; $S_y^2 = 6'8$; $S_{xy} = 7'2$

(a) Doi ezazu kontsumoa sarreraren arabera, karratu txikienen erregresio-zuzena erabiliz.
(b) Kalkula itzazu erroreak.
(c) Kalkula ezazu doikuntzaren egokitasuna. Komenta ezazu emaitza eta azaldu erregresio linealak azaltzen ez duen kontsumoaren bariantzaren ehunekoa.

3. Honako aldagai hauek aztertu dira:

X: biztanlerian dagoen analfabetoen ehunekoa.

Y: errenta biztanleko (mila eurotan)

Emaitza hauek lortuta:

$\overline{X} = 36$ $\overline{Y} = 10$ $S_X = 27$ $S_Y = 5$ $S_{XY} = -120$

Bi aldagaien arteko erlazio lineala ezartzen dela kontuan hartuta:

Baldin eta herri jakin batean, analfabeto-populazioaren % 17 bada, zein izango da biztanleen errenta (per capita errenta)?

Zein izango da eredu linealak azaldutako bariantzaren ehunekoa?

Baldin eta herri bateko biztanleriaren %15 analfabetoa bada eta biztanleko errenta 14 (mila eurotan), zein izango da erregresioaren errorea?

3. Laku batean, oxigenoaren edukia (Y) eta sakonera (X) neurtu dira honako datu hauek lortuta:

X	15	20	30	40	50	60	70
Y	6´5	5´6	5´4	6	4´6	1´4	0´1

a. Kalkula eta interpreta ezazu bi aldagai horien arteko korrelazio lineala.
b. Doi ezazu *X*-rekiko *Y*-ren zuzena karratu txikienen metodoa erabiliz.
c. Sakonera 80m izanik, oxigenoaren zer edukia aurresan daiteke?
d. Zein izango da aurreko doikuntzaren egokitasuna?

8. GAIA: BI ALDAGAI BAINO GEHIAGOREN ARTEKO ERREGRESIOA ETA KORRELAZIOA

8.1. Karratu txikienen erregresio lineala R^n-an. Doikuntzaren egokitasun-analisia.
8.2. Korrelazio-koefizienteak: totala, anizkoitza eta partziala.

8.1. Karratu txikienen erregresio lineala R^n-an. Doikuntzaren egokitasun-analisia.

KARRATU TXIKIENEN ERREGRESIO LINEALA R^n-AN---

Oro har, Ekonomian, bi aldagairen arteko harremana isolatzen saiatzea nahiko zaila izaten da, beste aldagai batzuen menpe ere egon baitaitezke, neurri batean.

Izan ere, logikoa da pentsatzea familia-gastua errentaren menpe dagoela, antza; baina, familiako pertsona kopuruaren menpe ere egon daiteke, adibidez. Hori dela eta, aurreko gaian azaldutako erregresio-zuzenak ez digu balio hiru aldagai horien arteko harremana azaltzeko.

Beraz, erregresio anizkoitzaren egoera dugu. Hau da, gastuaren portaera (X_1 aldagaia) azaltzen saiatuko gara, bai errenta (X_2 aldagaia) bai familiako pertsona-kopurua (X_3 aldagaia) kontuan hartuta. Oro har, erregresio anizkoitzak kausazko aldagai multzoaren arabera ($X_2, X_3, ..., X_n$) X_1en portaera azaltzea du helburu. Guk bi aldagaitara mugatuko dugu.

Gure planteamendua bi aldagairen arteko erregresioa orokortzea izango denez gero, X_2 eta X_3 aldagai askeen arabera X_1 aldagaiaren portaera azalduko dugu plano baten bitartez. Beraz, honako funtzio hau $\hat{X}_1 = b_{10} + b_{12}X_2 + b_{13}X_3$ definitu beharko dugu, eta R^3 espazioan puntu-hodeia doitzen duen planorik onena izango da.

X datu-matrize honetan X_1, X_2, X_3 aldagaien N banakori buruzko datuak izango ditugu:

$$X = \begin{pmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \\ \dots & \dots & \dots \\ x_{1N} & x_{2N} & x_{3N} \end{pmatrix}$$

Beraz, erroreen karratuaren batezbestekoa minimoa egiten duen $\hat{X}_1 = b_{10} + b_{12}X_2 + b_{13}X_3$ planoa aurkitu beharko dugu. Hau da:

$$E = \frac{1}{N}\sum_{i=1}^{N}(x_{1i} - b_{10} - b_{12}x_{2i} - b_{13}x_{3i})^2$$

Dakigunez, honako hau da beharrezko baldintza *E*-k minimoa izan dezan honako hau da: b_{10}, b_{12} eta b_{13} parametro bakoitzarekiko lehen deribatu partzialak 0 izatea. Hau da:

$$\frac{\partial E}{\partial b_{10}} = 0, \quad \frac{\partial E}{\partial b_{12}} = 0 \quad \text{eta} \quad \frac{\partial E}{\partial b_{13}} = 0$$

Hortik, honako hau lortzen da:

$$\left.\begin{aligned} \bar{x}_1 &= b_{10} + b_{12}\bar{x}_2 + b_{13}\bar{x}_3 \\ S_{x_2x_1} &= b_{12}S^2_{x_2} + b_{13}S_{x_2x_3} \\ S_{x_3x_1} &= b_{12}S_{x_2x_3} + b_{13}S^2_{x_3} \end{aligned}\right\}$$

Lehen ekuaziotik, erregresio-planoa batezbestekoen bektoretik $(\bar{x}_1, \bar{x}_2, \bar{x}_3)$ pasatzen dela ondorioztatzen da, eta horretaz gain $b_{10} = \bar{x}_1 - b_{12}\bar{x}_2 - b_{13}\bar{x}_3$ dela

Matrizeen bidez adierazita, hauxe izango dugu:

$$\begin{pmatrix} S_{x_2x_1} \\ S_{x_3x_1} \end{pmatrix} = \begin{pmatrix} S^2_{x_2} & S_{x_2x_3} \\ S_{x_2x_3} & S^2_{x_3} \end{pmatrix} \begin{pmatrix} b_{12} \\ b_{13} \end{pmatrix}$$

Eta Cramer-en erregelaz b_{12} eta b_{13} kalkulatuko ditugu:

$$b_{12} = \frac{\begin{vmatrix} S_{x_2x_1} & S_{x_2x_3} \\ S_{x_3x_1} & S^2_{x_3} \end{vmatrix}}{\begin{vmatrix} S^2_{x_2} & S_{x_2x_3} \\ S_{x_2x_3} & S^2_{x_3} \end{vmatrix}} \qquad b_{13} = \frac{\begin{vmatrix} S^2_{x_2} & S_{x_2x_1} \\ S_{x_2x_3} & S_{x_3x_1} \end{vmatrix}}{\begin{vmatrix} S^2_{x_2} & S_{x_2x_3} \\ S_{x_2x_3} & S^2_{x_3} \end{vmatrix}}$$

Honako idazkera hau erabiliz:

$$L = \begin{pmatrix} S^2_{X_1} & S_{X_1X_2} & S_{X_1X_3} \\ S_{X_1X_2} & S^2_{X_2} & S_{X_2X_3} \\ S_{X_1X_3} & S_{X_2X_3} & S^2_{X_3} \end{pmatrix} = \begin{pmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \text{kobariantza-matrizea}$$

Hauxe dugu:

$b_{12} = \dfrac{-L_{12}}{L_{11}}$ eta $b_{13} = \dfrac{-L_{13}}{L_{11}}$ erregresio-koefizienteak, l_{1j}-ri dagokion adjuntua L_{1j} izanik.

b_{1j} erregresio-koefizientea--- X_j, j=2,3 aldagaiaren igoera unitate batekoa denean X_1 aldagaiaren igoera da, gainerako aldagaia konstantea izanda.

PROPIETATEAK

1. Erroreen batezbestekoa *0* izango da: $\bar{e} = 0$

Batez besteko errore koadratikoa edo hondar-bariantza hauxe izango da:

$$S_e^2 = \frac{\sum (e_i - \bar{e})^2}{N} = \frac{\sum e_i^2}{N} = \frac{\sum (x_{1i} - \hat{x}_{1i})^2}{N}$$

2. Balio zenbatetsien batezbestekoa eta balio behatuen batezbestekoa bat

datoz: $\bar{\hat{x}}_1 = \bar{x}_1$

3. Errore aldagaia eta aldagai askeak korrelazio gabekoak dira.

$$S_{eX_2} = 0 \quad \text{eta} \quad S_{eX_3} = 0$$

4. Aldagai zenbatetsia eta errore-aldagaia korrelazio gabekoak dira. $S_{e\hat{x}_1} = 0$

DOIKUNTZAREN EGOKITASUN ANALISIA---

Aurreko propietateetan ikusi dugunez, $\hat{X}_1$ eta *e* korrelazio gabekoak dira; hori dela eta, honako hau erator dezakegu:

$$e = X_1 - \hat{X}_1 \quad \Rightarrow \quad X_1 = \hat{X}_1 + e \quad \Rightarrow \quad S^2_{X_1} = S^2_{\hat{X}_1} + S^2_e$$

$S^2_{X_1}$ -z zatituz, honako hau lortzen da:

$$1 = \frac{S^2_{\hat{X}_1}}{S^2_{X_1}} + \frac{S^2_e}{S^2_{X_1}}$$

$\frac{S^2_{\hat{X}_1}}{S^2_{X_1}}$ kozienteak erregresio-ereduak azaldutako $S^2_{X_1}$ en proportzioa adierazten du.

Beraz, azaldutako aldakuntzaren proportzioa zenbat eta handiago izan, orduan eta egokiagoa izango da eredua. Hori dela eta, doikuntzaren egokitasun-neurri aproposa izango da, eta **mugatze-koefizientea** deituko zaio.

$$R^2 = \frac{S^2_{\hat{X}_1}}{S^2_{X_1}} = 1 - \frac{S^2_e}{S^2_{x_1}}$$

8.2. Korrelazio-koefizienteak: anizkoitza, totala eta partziala.

KORRELAZIO ANIZKOITZAREN KOEFIZIENTEA---

Mugatze-koefizientearen erro karratuari deitzen zaio korrelazio anizkoitzaren koefizientea, eta $r_{1.23}$ ikurraz adierazten da. Honela ere adieraz daiteke:

$$r_{1.23} = \sqrt{R^2} = r_{X_1\hat{X}_1}$$

Korrelazio anizkoitzaren koefizientearen definizioa eta bi aldagairen arteko korrelazio bakunaren koefizientearen definizioa antzekoak dira; hau da, X_1 menpeko aldagaiaren eta $\hat{X}_1 = f(X_2, X_3)$ aldagai aske guztien arteko elkar harremana adierazten du.

KORRELAZIO TOTALAREN KOEFIZIENTEA---

Bi aldagairen arteko korrelazioa adierazten du, beste aldagaien eraginari uko egin gabe.

$$r_{ij} = \frac{S_{ij}}{S_i S_j}$$

KORRELAZIO PARTZIALAREN KOEFIZIENTEA---

Bi aldagairen arteko korrelazioa adierazten du, beste aldagaien eraginari uko eginik.

Demagun X_1 eta X_2 aldagaien arteko korrelazio partzialaren koefizientea kalkulatu nahi dugula; hau da, X_3 aldagaiaren eraginari uko egin behar diogu. $r_{12.3}$ ikurraz adieraziko da eta X_3 aldagaiarekiko X_1 eta X_2 aldagai bakoitzaren erregresio-hondarren arteko korrelazio totala izango da. Hau da:

$$r_{12.3} = r_{e_{1.3},e_{2.3}} = \frac{\mathrm{kob}(e_{1.3}, e_{2.3})}{\sqrt{\mathrm{bar}(e_{1.3})\mathrm{bar}(e_{2.3})}}$$

Ariketak

1. *Y, X_1* eta X_2 aldagaitarako hiru behaketa egin dira:

Y	X_1	X_2
4	6	8
5	6	3
2	3	6

Kalkula ezazu X_1-ekiko eta X_2-rekiko Y aldagaiaren erregresio-planoa, eta konproba ezazu planoa ondo doitzen den emandako hiru puntuetara.

2. Demagun X_1, X_2 eta X_3 aldagaiak ditugula:

X_1	1	6	14	12
X_2	3	8	2	10
X_3	2	5	9	10

Honako hauek dira haien batezbesteko balioen bektorea eta kobariantza-matrizea:

$$\bar{x} = \begin{bmatrix} 8'25 \\ 5'75 \\ 6'5 \end{bmatrix} \qquad L = \begin{bmatrix} 26'18 & 2'31 & 15'87 \\ 2'31 & 11'18 & 3'62 \\ 15'87 & 3'62 & 10'25 \end{bmatrix}$$

Zera eskatzen da:

a. Lor ezazu X_2-rekiko eta X_3-rekiko X_1-en erregresio-planoa.
b. Kalkula itzazu honako balio hauek:
   - Azaldu behar den bariantza totala.
   - Azaldutako bariantza.
   - Azaldu gabeko bariantza
   - Azaldutako bariantzaren ehunekoa.
   - Azaldu gabeko bariantzaren ehunekoa.

3. Honako taula honetan, 6 herritako NPG gasolinaren eta energiaren elektrikoaren kontsumoaren datuak agertzen dira:

NPG	**GASOLINA**	**ENERGIA ELEKTRIKOA**
8	4	20
10	3	22
12	6	23
13	6	26
15	7	27
18	8	30

Batezbesteko balioen bektorea eta kobariantza-matrizea honako hauek dira:

$$\bar{x} = \begin{bmatrix} 12'667 \\ 5'667 \\ 24'667 \end{bmatrix} \qquad L = \begin{bmatrix} 10'555 & 5'055 & 10'722 \\ 5'055 & 2'888 & 5'055 \\ 10'722 & 5'055 & 11'222 \end{bmatrix}$$

a. Kalkula ezazu NPGren eta gasolinaren kontsumoaren arteko korrelazio-koefiziente partziala, eta azaldu horren esangura.
b. NPG aldagai menpekotzat hartuta, lor itzazu karratu txikienen erregresio-planoa, korrelazio anizkoitza eta mugatze-koefizientea.

4. Demagun X_1, X_2 eta X_3 hiru aldagai ekonomiko direla eta hiru herrialdetan hartu dituzten balioak honako hauek direla:

X_1	X_2	X_3
6	3	0
2	10	1
1	20	2

Batezbestekoen bektorea eta kobariantza-matrizea honako hauek dira :

$$\bar{x} = \begin{bmatrix} 3 \\ 11 \\ 1 \end{bmatrix} \qquad L = \begin{bmatrix} \frac{14}{3} & -\frac{41}{3} & -\frac{5}{3} \\ -\frac{41}{3} & \frac{146}{3} & \frac{17}{3} \\ -\frac{5}{3} & \frac{17}{3} & \frac{2}{3} \end{bmatrix}$$

Zera eskatzen da:

a. X_2-rekiko eta X_3-rekiko X_1-en erregresio-planoa.
b. X_2-rekiko X1-en erregresio-zuzena.
c. X_1-en eta X_2-ren arteko korrelazio-koefiziente totala eta partziala. Komenta ezazu emaitza.

9. GAIA: ZENBAKI INDIZEAK

9.1. Definizioa.
9.2. Zenbaki-indize bakuna eta konplexua.
9.3. Kontsumo-prezioen indizea (KPI) eta beste indize batzuk.
9.4. Aldakuntza eta ondorioa.
9.5. Oinarri-aldiaren aldaketa eta serieen lotura.
9.6. Serie estatistikoen deflazioa.

9.1. Definizioa.

Fenomeno ekonomiko batzuen esanahiek eta ikerketek konplexutasun maila desberdinak izaten dituzte. Egoera gehienetan, fenomeno horiek aztertzeak behartzen gaitu aldagaiak alderatzera; konparaketa hori zatiketaz egin ohi da, hau da erlatiboki; hala, neurri-unitateen eragina balio gabe geldituko da. Zatiketaren bidez, zenbaki-indizea definitzen du konparaketa horrek.

Izan ere, zenbaki indizeak tresnarik egokienak izango dira honako galdera hauei erantzuteko: Egoera ekonomikoa positiboa edo negatiboa da? Inflazio maila egokia ote da?,...

DEFINIZIOA→ zenbaki-indizea da, lekuan edo denboran zehar magnitude jakin baten aldaketa erlatiboki adierazten digun neurri estatistikoa.

Aldagaiaren bilakaera aztertu nahi denean, zenbait aldiren balioak izan beharko ditugu: X_0, X_1, ..., X_T, eta honako kontzeptu hauek definitu:

- **Oinarri-aldia**, X_0 aldagaiaren bilakaera alderatzeko erreferentziatzat hartzen den aldia.
- **Konparazio-aldia**, aldagaiaren balioa X_t (t= 0,1,2,...,T) oinarri-aldikoarekiko alderatzen den aldia.

Ehunekotan ***0***tik ***t***-ra bilakaera indizea: $I_0^t = \frac{X_t}{X_0} 100$

- Aldagaiak kuantitatiboak badira, orduan, zenbaki indizeek badute zentzurik.
- Indizea, zatiketaren bidez definituta dagoenez gero, aldagaiaren neurri-unitatearekiko askea izango da. Beraz, indize desberdinak batu ahal izango dira. Hala, fenomeno ekonomikoen bilakaera adierazleak eraiki daitezke oro har.

9.2. Zenbaki-indize bakuna eta konplexua.

Zenbaki-indizeak honako hau kontuan hartuta sailkatzen dira:

- Aldagai kopuruaren arabera : →bakunak
  →konplexuak
- Aldagai bakoitzaren garrantzi erlatiboaren arabera: →haztapenik gabekoa
  → haztatua

9.2.1. Zenbaki-indize bakunak

Aldagai bat besterik ez duen magnitudearen bilakaera aztertzen du.
Enpresek ekoizpenak zein salmentak aztertu nahi dituztenean erabili ohi dituzte.

Ehunekotan, *0*tik *t*ra bilakaera indizea: $I_0^t \frac{X_t}{X_0}.100$

I_0^t***: t*** denboraldian ***X*** aldagaiaren zenbaki-indizea.
X_t ***t*** denboraldian ***X*** aldagaiaren balioa
X_0 ***0*** oinarri-aldian ***X*** aldagaiaren balioa

I_0^tk oinarri-aldian 100 balioa hartzen du.

$I_0^t \prec 100$ denean $X_t \prec X_0$ izango da, hau da, ***0***tik ***t***-ra eboluzioa negatiboa izan dela adierazten du.

$I_0^t \succ 100$ denean $X \succ X_0$ izango da, hau da, ***0***tik ***t***-ra eboluzioa positiboa izan dela adierazten du.

9.2.2. Indize konplexuak.

Aldagaia bat baino gehiago dituen magnitudearen bilakaera aztertzea du helburu. Aldagai bakoitzari garrantzia edo haztapena w_i esleitzen zaio, eta haztapenak berdinak $(w_1 = w_2 = ... = w_N)$ edo desberdinak $(w_1 \neq w_2 \neq ... \neq w_N)$ izan daitezke; horren arabera, haztapenik gabeko indizeak eta indize haztatuak bereiziko dira. Bestalde, indize konplexuak eraikitzeko, bi metodo erabil daitezke, batezbesteko aritmetikoarena eta batezbesteko gehituarena, hain zuzen.

- ***N-*** aldagai kopurua (1,2,..., ***N***)
- $I_{i,0}^t$ ***-t*** denboraldian, ***i*** aldagai bakoitzaren zenbaki-indize bakunak.
- X_{it}***- t*** denboraldian ***i*** aldagaiaren balioa.
- X_{i0}***-0*** oinarri-aldian ***i*** aldagaiaren balioa.

- **Batezbesteko aritmetikoaren indizea (Sauerbeck)→**

$$S_{t,0} = \frac{\sum I_{i,0}^t \bullet w_i}{N} = \frac{\sum \frac{X_{it}}{X_{i0}} \bullet w_i}{N} 100$$

- **Batezbesteko gehituaren indizea (Bradstreet-Dutot) →**

$$B_{t,0} = \frac{\sum X_{it} \bullet w_i}{\sum X_{i0} \bullet w_i} 100$$

<u>Prezio-indizea</u>--- Denboran zehar, prezioen bilakaera neurtzen dute.

Prezio-indize bakuna.

P_0^t ***-t*** aldian prezio-indize bakuna.
P_0***- 0*** oinarri-aldian prezioa.
P_t***- t*** aldian prezioa.

$$P_0^t = \frac{P_t}{P_0} 100$$

Haztapenik gabeko prezio-indize konplexuak.

a) Batezbesteko aritmetikoaren indizea. Aldagai bakoitzerako prezio-indize bakunen batezbesteko aritmetikoa da.

$$P_S = \frac{1}{N} \sum_{i=1}^{N} P_{i0}^t = \frac{1}{N} \sum_{i=1}^{N} \frac{P_{it}}{P_{i0}} 100$$

b) Batezbesteko gehituaren indizea. t aldiko zein oinarri-aldiko N prezioren batezbestekoen arteko kozientea.

$$P_B = \frac{\frac{1}{N} \sum_{i=1}^{N} P_{it}}{\frac{1}{N} \sum_{i=1}^{N} P_{i0}} 100 = \frac{\sum_{i=1}^{N} P_{it}}{\sum_{i=1}^{N} P_{i0}} 100$$

Prezio-indize konplexu haztatuak

Indizeen arteko desberdintasuna ***wi*** haztapenaren irizpidea izango da.

$$P_X = \frac{\sum_{i=1}^{N} P_{it} \bullet w_i}{\sum_{i=1}^{N} P_{i0} w_i} 100$$

a) P_L Laspeyresen prezio-indizea.

Oinarri-aldiko kantitateek haztapentzat jotzen dute. Denboraldi guztietarako haztapenak tinko mantentzea abantaila izango da; alabaina, denborarekiko zenbat eta urrunago egon, hainbat eta txikiagoa izango da adierazgarritasuna.

$$W_i = q_{i0}$$

b) P_P Paascheren prezio-indizea.

Konparaziozko aldiko kantitateak izango dira haztapenak. Alde batetik, denboraldi bakoitzerako aldagai guztien garrantzi erlatiboa eguneratzea izango da abantaila; baina, beste aldetik, kalkuluaren konplexutasunak eta eratorritako kostuek astungarritzat jotzen dute.

$$W_i = q_{it}$$

PROPIETATEAK

- **Existentzia**: indizeak zenbaki errealera eta finitura mugatu behar du. $I_0^t \neq 0$
- **Identitatea:** oinarri-aldia eta alderaketazkoa bat badatoz, $I_0^t = 100$ izango da.
- **Alderantzikatzea:** bi denboraldiko bi indize alderantzikaturen biderkadura $I_0^t \bullet I_t^0 = 1$ izango da.
- **Zirkularra:** denboraldi batzuetarako alderantzikatzea orokortzea. $I_0^{t'} \bullet I_{t'}^0 \bullet I_t^0 = 1$
- **Proportzionaltasuna:** Aldagaia ***1+K*** proportzioan aldatzen bada, indizea proportzio berean aldatuko da.

Izan ere, indize bakuna gehienek betetzen dituzte, konplexuek, berriz, zenbait propietate ez dituzte betetzen. Indizerik garrantzitsuenei buruz honako hau esango dugu:

- **Betetzen dituzte**: existentzia, identitatea zein proportzionaltasuna.
- **Ez dituzte betetzen**: inbertsioa zein zirkularra.

Bai prezioetan bai ekoizpenetan Laspeyresen indizea izan da, oro har, gehien erabili ohi den adierazlea. Haren diseinuak zein geroko kalkuluak indizearen aldagai eta haztapenen hautapena zehatza izatera behartzen dute. Dena dela,

oinarri-alditik zenbat eta gehiago urrundu, orduan eta txikiagoa izaten da haztapen-egituraren adierazgarritasuna; hori dela eta, beste egitura berri bat ezarri eta oinarri-aldi berria finkatu beharko da.

9.3. Kontsumo-prezioen indizea (KPI) eta beste indize batzuk.

9.3.1. Kontsumo-prezioen indizea (KPI)

• FITXA TEKNIKOA

- Inkesta mota: etengabea (FAI). Hiruhileroko oinarrizko informazioa.
- Oinarri-aldia: 2006.
- Haztapenen erreferentzia-aldia: 1999ko 2. hiruhilabetetik 2001eko 1.ra.
- Udalerrietako lagina: elikadurarako 177 eta gainontzekorako 97.
- Artikulu-kopurua: 491.
- Behaketa kopurua: hilero 220.000 prezio gutxi gorabehera..
- Sailkapen funtzionala:12 talde, 37 azpitalde, 80 Klase eta 117 azpiklase; 57 sinadura eta 37 talde berezi.
- Kalkulu-metodo orokorra: Laspeyresen kateatua.
- Datu-bilketako metodoa: inkestagileak saltokietan eta artikulu berezietarako bilketa zentralizatua.

- **EZAUGARRI BEREZIAK**

Kontsumo-prezioen indizea da, Espainiako biztanleriak kontsumitzen dituen ondasunen eta zerbitzuen multzoaren bilakaera.

Kontsumo-prezioen indize harmonizatua da, Europar Batasuneko herrien arteko konparaketak egiteko adierazlea. Indize horrek inflazio-neurria ematea du helburu.

Espainian, KPI (92ko oinarri-aldia) **1993ko urtarriletik** egon da indarrean. KPI kalkulatzeko erabiltzen den erosketa-saskiaren oinarria une jakin bateko familien kontsumoa da eta noizean behin eguneratu behar da. Gaurko kasuan erosketa-saskia eguneratu ez ezik, KPI kalkulatzeko era berria sortu ere egiten da. Hori dela eta, ez da oinarri-aldiaren aldaketa gertatu sistemaren aldaketa baizik; **2001**

oinarri-aldiko indize-sistema, hain zuzen ere, eta 2002ko urtarrilean sartu zen indarrean. Gaur egun kontuan hartu behar dugu 2006. urtea dela oinarri-aldia.

Gaurko indizea aurrekoa baino dinamikoagoa da, sarritan haztapenak egunera baitaiteke eta merkatu-bilakaerara hobeto moldatzen baita. Horretaz gain, produktu baten kontsumoa garrantzitsua edo adierazgarria izaten hasten den unean produktu hori sartu ahal izango da.

- **EZAUGARRI GARRANTZITSUENAK**

- **Oinarri-aldia.**

Hileroko indizeen batezbestekoa 100 den aldia izango da oinarri-aldia. Sistema berrian, oinarri-aldia 2006a da. Beraz, kalkulatzen diren indize guztiek urte hori hartuko dute kontuan.

- **Haztapenen erreferentzia-aldia.**

Haztapenak familia-aurrekontuen inkestatik (FAI) ateratzen direnez gero, KPIren erreferentzia-aldia FAI egiten den aldia bera izango da. Beraz, 2004ko 2. lauhilabetearen eta 2006eko 1. lauhilabetearen arteko aldia da sistema berriaren erreferentzia-aldia. Haztapenak kalkulatzeko, eguneratze-momentutik hurbilago dauden lauhilabeteen informazioak besteenak baino garrantzi handiagoa izan du.

- **Lagina.**

Adierazle adierazgarriak lortzeko, hiru ataletan egitaratuta dago maila guztietarako hautapen-prozesua eta prozesu bakoitzak bakoitzaren osagaiak aukeratzea du helburu. Honako hauek dira:

  - Udalerriak hautatzea.
  - Merkataritzaguneak eta saltokiak hautatzea.
  - Behaketa-kopuruaren zehaztea.

- **Kontsumo-eredua.**

Etxebizitzetan kontsumora xedatzen diren ondasunen eta zerbitzuen multzoak osatzen du.

- **Erosketa-saskia.**

Familia-aurrekontuen inkestaren arabera KPI kalkulatzeko kontuan hartzen diren ondasunen eta zerbitzuen multzoa da.

- **Kalkulatzeko metodoa.**

Laspeyresen indizean datza.

- **Kalitate-aldaketa.**

Ondasun bateko ezaugarriren baten aldaketak kontsumitzaileari emandako baliagarritasunari eragiten badio, aztertu behar da ondasunaren prezioari ere eragiten dion. Hala bada, aldaketa hori islatu beharko da KPIn.

- **Merkealdiak barruan egotea.**

Prezio-beheratuak hartzen dira kontuan.

- **KPIren egokitzapen jarraitua.**

Haztapenak urtero berrikusten dira.

- **KPIren egitura-berrikustea.**.

Oinarri-aldia bost urtean behin aldatzen da.

9.3.2. Abagune ekonomikoaren beste adierazle batzuk.

- **Industria-ekoizpenaren indizea. (IPI)**

Industria-sektoreetako ekoizpenaren bolumena aztertzen du. 600 produktu behatzen dira eta hilero kalkulatzen da. INE da erakunde arduraduna.

- **Industria-prezioen indizea. (IPRI)**

Aurreko indizearekin, industria-abagunearen informazioa osatzen da. Fabrikatik ateratzean eskatzen da prezioa.

- **Txikizkako merkataritzaren indizea.**

- **Salmenta-indizea. (azalera handietako merkataritzaren indizea)**

9.4. Aldakuntza eta ondorioa.

ALDAKUNTZA--- Indizearen aldakuntza zera da: indizeak bi dataren artean duen igoera edo beherapenaren ehunekoa. Beraz, ***t*** eta ***t´***-ren arteko aldakuntza neurtzekotan:

$$A_{(t\to t')} = \frac{I_{0t'} - I_{0t}}{I_{0t}} \bullet 100$$

$A_{(t\to t')} \to$ daten arteko aldakuntza

$I_{0t} \to$ ***t*** aldiko indizea, oinarri-aldia ***0***.

$I_{0t'} \to$ ***t´*** aldiko indizea, oinarri-aldia ***0.***

ONDORIOA--- Indizeko artikulu jakin baten aldakuntzaren ondorioa da, artikulu hori dagokion indizearen aldakuntza. Hori dela eta, indizean parte hartzen duten artikulu guztien ondorioen batura indizearen aldakuntza da. Beraz, $A_{(t\to t')}$n ***i*** artikuluaren ondorioa kalkulatzekotan:

$$O^i_{(t\to t')} = \frac{I^i_{0t'} - I^i_{0t}}{I_{0t}} w_i \bullet 100 \Rightarrow \sum_i O^i_{(t\to t')} = A_{(t\to t')}$$

9.5. Oinarri-aldiaren aldaketa.

Nahiz eta Laspeyresen indizeak inbertsio-propietatea ez bete, oinarri-aldia desberdinak dituzten zenbaki-indizeak lotu behar direnean, inbertsio-propietatean oinarrituko gara.

Demagun honako zenbaki-indize hauen segida hau dela eta horien oinarri-aldia ***0*** dela:

$$I^0_0, I^1_0, ..., I^n_0$$

t aldia ***0*** konparazio-alditik oso urrun dagoenean interesgarria izan daiteke ***0*** oinarria aldatzea. Horretarako, zirkular- eta inbertsio-propietateetan oinarrituz, segida ezaguna ***t´*** oinarri-aldia duen segida izatera pasatzen da.

a) *t'* oinarri-berrian indize bakoitzaren balioa.

Segida transformatzeko, ***t'*** aldirako lotura teknikoa izango den I_0^t indizea egongo da.

Antzinako indize bakoitza aldi berrian adierazteko, honako hau betetzen da:

$$I_0^t \bullet I_{t'}^t \bullet I_t^0 = 1 \rightarrow I_0^{t'} \bullet I_{t'}^t = \frac{1}{I_t^0} = I_0^t \quad \rightarrow \quad I_{t'}^t = \frac{I_0^t}{I_0^{t'}} 100$$

b) Transformazio-koefizientearen kalkulua.

Koefiziente hori antzinako koefizienteaz biderkatuz oinarri berriko indizea lortzen da.

$$I_{t'}^0 = \frac{I_0^0}{I_0^{t'}} 100 \Rightarrow$$

<u>Oinarri desberdineko zenbaki indizeen segida-lotura edo berriztapena.</u>

Egoera batzuetan aztertu nahi den aldagaia edo fenomenoa bi segidatan islatuta egon daiteke, eta bakoitzak bere oinarri-aldia izan. Beraz, zer edo zer ikertzeko, erreferentzia-aldia bera izan beharko dute.

Oraingo alditik hurbilen dagoen segidaren oinarri-aldia aintzat hartuko da eta segidarik zaharrenari lotura-koefizientea aplikatuko zaio. Demagun segidarik zaharrena zenbaki indizeak ***0*** oinarri-aldia izanda, honako hau dela :

$$I_0^0, I_0^1, I_0^2, ..., I_0^n$$

Demagun ***t'*** segidarik zaharrenaren aldia dela. Baldin eta antzinako segidaren indize bakoitza lotura-koefizienteaz biderkatzen badugu, indizea oinarri berrian lortuko dugu.

$$I_{t'}^0 = \frac{I_0^0}{I_0^{t'}} \Rightarrow \text{Lotura koefizientea } I_{t'}^0 = \frac{100}{I_0^{t'}}$$

9.6 .Balio-indizeak eta Ekonomia-segidaren deflazioa.

Ondasun ekonomikoek heterogeneotasun handia izaten dutenez gero, ondasunak gehitzekotan, homogeneizatu egin beharko dira. Balio indizeen bidez lortuko dugu hori.

Balio-indizearen bidez, ondasun multzo jakin baten balioa aztertu ahal izango dugu.

0 oinarri-aldiko balioa: $B_0 = \sum_{i=1}^{N} p_{i0} q_{i0}$

t aldiko balioa $B_t = \sum_{i=1}^{N} p_{it} q_{it}$

N aldagai kontuan hartuz gero, honako hau izango da indize konplexua:

$$I_{V_t} = \frac{\sum_{i=1}^{N} p_{it} q_{it}}{\sum_{i=1}^{N} p_{i0} q_{i0}}$$

Ikusten dugunez, denboran zehar, indize-bilakaeraren arrazoia izaten da kantitateen eta prezioen baterako bilakaera, eta bakoitzaren eragina ez dago isolatzerik.

Ekonomian, interesgarria izaten da ***N*** ondasun-multzoaren bilakaera aztertzea prezio-konstanteetan; hau da, aldagai desberdinen prezioetan aldaketarik egon gabe.

Aldi desberdinetan, balio-segidan konparazio-analisia egiteko, eguneko balioko moneta-unitateek oinarri-balioko moneta-unitateak izatera pasatu behar dute, prezioen indizerik egokienaz; hots, segida indize deflaktatzaileaz zatituz. Prozesu horri deflazioa deitzen zaio.

Zenbaki-indizeen artean Laspeyresen eta Paascheren indizeak erabiltzen dira gehien segidak deflaktatzeko.

a) Laspeyresen indizearen bidezko deflazioa.

Baldin eta eguneko balioa Laspeyresen prezioen indizeaz zatitzen bada, honako hau izango da:

$$\frac{B_t}{P_L} = \frac{\sum_{i=1}^{N} p_{it} q_{it}}{\frac{\sum_{i=1}^{N} p_{it} q_{i0}}{\sum_{i=1}^{N} p_{i0} q_{i0}}} = \sum_{i=1}^{N} p_{it} q_{i0} \frac{\sum_{i=1}^{N} p_{it} q_{it}}{\sum_{i=1}^{N} p_{it} q_{i0}} = B_0 Q_P$$

Ez da oinarri-prezioko balio-segida lortzen, $\boldsymbol{B_0}$ $\boldsymbol{Q_P}$ baizik; hau da, oinarri-aldiko balioa eta Paascheren kantitate-indizearen arteko biderkadura. Beraz, $\boldsymbol{P_L}$ indizea ez da benetako deflaktatzailea izango, praktikan deflaktatzailetzat jo arren.

b) Paascheren indizearen bidezko deflazioa.

Benetako indize deflaktatzailea da hori. ***N*** ondasun multzoaren eguneko balioa oinarri-aldiko eguneko prezioetan adieraziko baitugu. Nahi dugun ekonomia segidaren arabera, prezioen indizerik egokiena aukeratuko dugu.

$$\frac{B_t}{P_P} = \frac{\sum_{i=1}^{N} p_{it} q_{it}}{\frac{\sum_{i=1}^{N} p_{it} q_{it}}{\sum_{i=1}^{N} p_{i0} q_{it}}} = \sum_{i=1}^{N} p_{i0} q_{it}$$

• Urte jakin batean, familia baten errenta erabilgarria moneta-unitate konstantetan deflaktatzeko, KPI izaten da deflaktatzaile egokia.

• Industriako prezioen multzoa deflaktatzeko, berriz, IPI (Industriako prezioen indizea) izango da deflaktatzaile egokia.

Ariketak.

1. Martinez jaunak jakin nahi du zein izan den gosariaren prezioaren eboluzioa. Goizero kafea, ogi xigortuak eta laranja zukua hartzen ditu. Urteetan zehar bere ohiturak aldatu dira:

- 2000n, 250 gr. laranja, 4 ogi xigortu eta kafe bat hartzen zituen.
- 2002an, 500 gr. laranja, 2 ogi xigortu eta kafe bat hartzen zituen.

Taula honetan ageri da 3 produktu horien prezioen eboluzioa:

Gosaria	2000	2002	2004
Laranja 1 kg.	0´39	0´60	0´69
Ogi xigortu 1	0´18	0´30	0´38
Kafe 1	0´21	0´35	0´42
KPI	100	149	189

Zera eskatzen da:

a) 2000 oinarri–aldia hartuta, kalkula itzazu Martinez jaunaren 2002ko gosariaren prezioen indizeak (Paasche eta Laspeyres). Azaldu zergatik diren indize biak desberdinak.
b) Konpara ezazu KPIren eboluzioa, laranjen prezioan.

2. Hiru artikulu hauetan azken 3 urteotan kontsumitutako kantitateak honako hauek izan dira:

URTEAK	OGIA(Kg)	ESNEA(l.)	ARRAUTZAK(dozena)
2002	200	500	80
2003	240	610	87
2004	275	870	92´5

Zera eskatzen da:

a) Hiru urte horietan kontsumitutako kantitateen baterako eboluzioa ikusteko, haztatu gabeko indize konplexurik egokiena (oinarria 2002).
b) Laspeyresen indizea (oinarria 2002), 2002ko prezioak honako hauek direla jakinik: ogia: 6€./kg. esnea: 1´1€./l. arrautzak: 1´5€./dozena
c) Emaitzak.

3. Konpainia batek bi kalkulagailu mota merkaturatzen ditu: 16-C eta 43-E, haien prezioen segida (eurotan) eta saldutako kantitateak (mila unitatetan) honako taula honetan agertzen dira:

URTEAK	16-C		43-E	
	p	q	p	q
2000	8	1	10	4
2001	9	2	10	3
2002	10	3	10	2
2003	11	4	10	1

Zera eskatzen da:

a) 2000 oinarri-aldia hartuta, Laspeyresen prezio indizearen segida.
b) Orain 2002 oinarri-aldia hartuta, segida hori berriztatu eta lotura-eragiketa egitea.

4. Herri bateko KPI (oinarria 1992) hiru ondasunek osatzen dute:

urtea	jantzia (w_1 = 0´4)	elikadura (w_2 = 0´3)	aisia
1994	108	106´6	102
1995	107	111	103

a) Kalkula ezazu 94. eta 95. urteetako KPI.
b) Kalkula ezazu 94. urtearekiko 95. urteko aldakuntza erlatiboa.
c) Kalkula ezazu 94. urtearekiko 95. urteko aisia ondasunaren ondorio erlatiboa.

5. Har dezagun segida hau:

URTEA	SOLDATA NOMINALA (milak)	KPI
1987	80	100
1988	90	105
1989	95	107
1990	105	108
1991	120	110
1992	122	113
1993	125	117

a) Deflaziona ezazu soldata nominalaren segida 87. urtea oinarritzat hartuta. Zein da jatorrizko eta deflazionatutako segiden arteko diferentzia?
b) 87.urtearekiko soldata erreala ez aldatzeko, 93.urteko soldata nominala zein litzateke?

10. GAIA: PROBABILITATE ESPAZIOAK I

10.1. Sarrera.
10.2. Probabilitate-definizioa.
10.3. Probabilitatearen lehen propietateak.
10.4. Probabilitateen esleipena.

10.1. Sarrera.

Batzuetan, errealitatean topatzen ditugun egoerak ez ditugu ikertzerik, haien konplexutasunagatik. Beraz, egoera horietan, fenomeno erreal horiek aztertzeko errazten dituen ereduren bat beharko dugu. Horretarako, kontuan hartu behar dugu faktore kopuru murriztuak errealitatea azaltzen duenean, erlazio matematikoen bidez azaltzea erraza izaten dela.

Dena dela, beste egoera batzuetan errealitatea errazten duen eredu matematikoak ez du errealitatea zehatz azaltzen, emaitza ausazkoa da eta. Hori dela eta, eredua ziurgabetasunean oinarrituko da, hau da probabilitatean.

Beraz, probabilitate-propietateak ikastea beharrezkoa izango da errealitaterik probableenak ezagutzeko→**PROBABILITATE TEORIA.**

Eredu bat errealitatearen idealizazioa besterik ez denez gero, eredutik erakarriko diren ondorioak errealitatera zenbat eta gehiago hurbildu orduan eta hobeak izango dira. Beraz, aplikazio-eremua ondo ezagutzea funtsezkoa izango da ahalik eta aukera hoberen egiten saiatzeko.

Hurrengoan, probabilitate-teoriaren oinarrizko hainbat kontzeptu definituko dira:

1. AUSAZKO ESPERIMENTUA.

Honako ezaugarri hauek bete behar ditu defini dadin:

a. Baldintza berdinetan eta mugatu gabe gertatu arren emaitzak ezin daitezke aurresan. Aldiro, emaitzak desberdinak lor daitezke.

Adibidea: baldintza berdinetan dado bat jaurti dezakegu eta emaitzak desberdinak lortu.

b. Proba bakoitzak esperimentuaren gertaera posibleen multzo barnekoa den emaitza ematen du.

Adibidea: $\{5\} \in \{1,2,3,4,5,6\}$

c. Proba kopurua handitzean, emaitza bakoitzaren maiztasun erlatiboaren joera zenbaki finkoa da. $\rightarrow$ "Erregulartasun estatistikoa".

Esperimentu determinista**: Kausa berek emaitza bera ere lortzen dute.**

2. LAGIN-ESPAZIOA (Ω):

Ausazko esperimentuaren gertaera posibleen multzoa.

Adibidea: dado bat jaurtitzearena $\Omega = \{1,2,3,4,5,6\}$

txanpon bat jaurtitzearena $\Omega = \{a,+\}$

bi txanpon jaurtitzearena $\Omega = \{aa, a+,++,+a\}$

3. GERTAERAK.

Lagin-espazioaren azpimultzo bakoitzari ausazko gertaera estatistikoa esaten zaio. Honako gertaera hauek bereiz daitezke:

. Oinarrizkoa---- Esperimentu aleatorio baten emaitza bakoitza.

. Konposatua----Bi gertaera elemental edo gehiagoz osatuta.

. Ziurra--- Beti gertatzen da. Matematikoki, bat dator lagin-espazioarekin. Ω-z adierazten da.

. Ezinezkoa---- Ez da inoiz gertatzen**.** $\varnothing$-z adierazten da.

. Gertaeren bilketa--- $(A \cup B)$, A edo B gertatzea edo biak.

. Gertaeren ebaketa--- $(A \cap B)$, A eta B gertatzea.

. Gertaera berdinak---A, B gertaeren arteko parekotasun edo berdintasuna

$A \Rightarrow B \qquad A \subset B$

$A = B$ baldin eta $A \subset B$ eta $B \subset A$

. Aurkako gertaerak edo osagarriak --- A eta A^o $\rightarrow A \cup A^o = \Omega$ A gertaera betetzen bada, bere osagarria den A^o ez da betetzen, eta alderantziz.

. Gertaera bateraezinak --- $A, B \rightarrow A \cap B = \varnothing$; Aurkako gertaerak bateraezinak dira, baina ez alderantziz.

10.2. Probabilitatearen definizioa.

DEFINIZIOA----funtzio bat da, non ausazko esperimentu jakin baten gertaera bakoitzari $[0,1]$ tartetik zenbaki erreala esleitzen baitzaio.

$$Pr: A \rightarrow [0,1]$$

eta honako axioma hauek bete behar dira:

I. $\forall A \qquad Pr(A) \geq 0$

II. $Pr(\Omega) = 1$

III. Baldin eta bi gertaera bateraezinak badira $\rightarrow A \cap B = \varnothing$

$$P_r(A \cup B) = P_r(A) + P_r(B)$$

10.3. Probabilitatearen lehen propietateak.

1. $\forall A$ $\boxed{P_r(A^o) = 1 - P_r(A)}$

$\Omega = A \cup A^o$; $A \cap A^o = \varnothing$, bateraezinak dira.

$$P_r(\Omega) = P_r\left(A \cup A^o\right) = P_r(A) + P_r(A^o)$$
$$1 = P_r(A) + P_r(A^o) \Rightarrow P_r(A) = 1 - P_r(A^o)$$

2. $P_r(\varnothing) = 0 \Rightarrow \Omega = \Omega \cup \varnothing \Rightarrow \Omega \cap \varnothing = \varnothing$

$$P_r(\Omega) = P_r(\Omega \cup \varnothing) = P_r(\Omega) + P_r(\varnothing)$$
$$1 = P_r(\Omega) + P_r(\varnothing) \Rightarrow P_r(\varnothing) = 1 - P_r(\Omega) = 1 - 1 = 0$$

3. $\forall A, B \Rightarrow A \subset B \Rightarrow P_r(A) \leq P_r(B)$

B gertatzen denean *A* gertatu behar bada, *B A* bezain probablea izango da.

$A \subset B \Rightarrow B = A \cup \left(B \cap A^o\right)$ eta $A \cap \left(B \cap A^o\right) = \varnothing$, bateraezinak dira.

$$P_r(B) = P_r(A) + P_r(B \cap A^o) \Rightarrow P_r(B) \geq P_r(A)$$

4. $\forall A, B$ $\quad P_r(A \cup B) = P_r(A) + P_r(B) - P_r(A \cap B)$

$A \cup B = A \cup \left(A^o \cap B\right)$ eta $A \cap \left(A^o \cap B\right) = \varnothing$

$$P_r(A \cup B) = P_r\left(A \cup \left(A^o \cap B\right)\right) = P_r(A) + P_r(A^o \cap B)$$
$$B = (A \cap B) \cup \left(A^o \cap B\right) = B \cap \left(A \cup A^o\right)$$

$$P_r(B) = P_r(A \cap B) + P_r(A^o \cap B)$$

$$P_r(A^o \cap B) = P_r(B) - P_r(A \cap B)$$

$$P_r(A \cup B) = P_r(A) + P_r(B) - P_r(A \cap B)$$

10.4. Probabilitateen esleipena.

Matematikoki, badakigu zer den probabilitatea. Alabaina, edozein probabilitate esleipenek probabilitate axiomak bete ditzakeen arren, huts egin dezakegu.
Beraz, probabilitatea esleitzeko probabilitatearen hiru interpretazio ikusiko ditugu.

I. INTERPRETAZIO KLASIKOA.

Ausazko jokoekin uztartzen da. Esaterako, bi txanpon batera jaurtitzea: $\Omega = \{(a,a),(a,+),(+,a),(+,+)\}$. Txanponak txarto ez badaude, edozein emaitza irteteko probabilitatea bera da. Beraz, intuitiboki, pentsa dezakegu, aurpegi eta gurutzea irteteko probabilitatea $\frac{2}{4}$ izango dela. Horrek ez du esan nahi lau jaurtiketetan bi aldiz *(a,+)* irtengo dela, baina esperimentua aldi askotan egin eta gero, ikusiko dugu emaitza horretaranz jotzen duela .

Definizioa $\rightarrow$ Esperimentu jakin bat ausazkoa izanez gero, elkarren artean baztertzaileak eta probabilitate bereko n emaitza gerta daitezke eta emaitza

horietariko n_A-k A ezaugarria edukiz gero, A-ren probabilitatea $\frac{n_A}{n}$ proportzioa izango da.

Adibidea: dado bat jaurtiz gero, zenbaki bikoitia ateratzeko probabilitatea, zein izango da?: $\frac{3}{6}$.

II. **MAIZTASUN ERLATIBOA.**

Definizioa→ Baldintza berdinetan esperimentu jakin bat n aldiz errepikatuz gero eta emaitzetatik n_B B emaitzaren aldekoak baldin badira, n handitzen den neurrian, $\frac{n_B}{n}$ limitea izango da B gertaeraren probabilitatea, hau da, $\frac{ALDEKO \quad KASUAK}{KASU \quad POSIBLEAK}$.

III. **SUBJEKTIBOA.**

Aurreko interpretazioek esperimentua baldintza berdinetan errepikatzea dute oinarria. Baina, fenomeno batzuk errepikatu ezin diren arren beharrezkoa izango da probabilitatearen bidez eredua egitea.

Egoera horretan, gertaera jazotzeko sineskera mailaren arabera probabilitatez interpretatzen da, hots, fenomeno jakin batez bakoitzaren iritzia adierazten du. Fenomeno jakin baten alde zenbateko apustu egingo genuke?, hau da, 10 eurotik zenbat jokatuko genuke?

Adibidea: Bi politikariri *(A, B)* hautesleen boto-asmoaz galde dakieke eta demagun honako erantzun hauek izango direla: *A* politikariak 8 biren kontra jokatzeko prest

dago, *B*, berriz, 7 hiruren kontra. Horrela, probabilitate subjektiboak definitzeko: $\frac{8}{8+2}$ eta $\frac{7}{7+3}$.

Ariketak.

1. Karta sorta batetik bi karta ateratzen dira (berriz sartu gabe). Zera eskatzen da:

a) urrerik ez lortzeko probabilitatea.
b) urre bat ateratzeko probabilitatea.
c) bi urre ateratzeko probabilitatea.

2. 5 txanpon jaurtitzen dira. Zein izango da 3 aurpegi lortzeko probabilitatea?

3. Izan daitezen A eta B bi gertaera, non $P_r(A)=0'3$; $P_r(B)=0'4$; $P_r(A\cap B)=0'1$ baitira. Esan honako baieztapen hauek egia direnetz:

a. $P_r(A\cap B^c)=0'2$
b. $P_r(A^c\cap B)=0'3$

4. Ontzi batean, a, e, i, o eta u letraz markaturiko bost bola ditugu. Hiru bola batera atera ditugu. Zein izango da a, e eta i letraz markaturiko bolak izateko probabilitatea?

5. 2 bola zuri, 1 gorri eta 3 urdin dituen ontzi batetik 3 bolako lagin bat ateratzen da berriz sartu gabe eta ordenarik gabe.

a. Zenbat dira horrela lor daitezkeen lagin desberdinak?
b. Zein da ateratako hiru bola kolore desberdinetakoak izateko probabilitatea?

6. Dado bat bi aldiz botatzen dugu airera. Eskatzen da:

a. Lehen jaurtiketan bi zenbakia edo bi baino txikiagoko zenbaki ateratzea A gertaera izanik, kalkulatu A gertatzeko probabilitatea .
b. Bigarren jaurtiketan 5 zenbakia baino txikiagoa ateratzea B gertaera izanik. Kalkula ezazu B gertatzeko probabilitatea..
c. Kalkulatu aurreko gertaeren bilketaren probabilitatea.

7. Liburu batek 400 orrialde ditu; horien artean 100 orrialdek argazkiak dituzte eta 50ek taulak. Horrez gain, 10 orrialdek taulak eta argazkiak dituzte. Liburua zoriz aukeratutako orrialde batetik zabaldu dugu. Kalkula itzazu ondorengo gertaeren probabilitateak:

a. Orrialdeak argazkiak edo taulak izatea.
b. Orrialdeak argazkiak izatea eta ez taularik.
c. Orrialdeak argazkirik ez izatea ezta taularik ere.

11. GAIA: PROBABILITATE ESPAZIOAK II

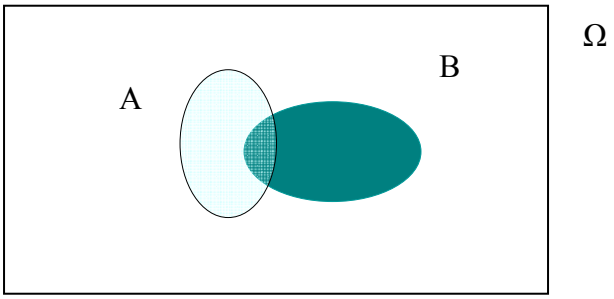
11.1. Probabilitate baldintzatua.
11.2. Ebaketa-teorema, partiketa-teorema eta Bayes-en teorema.
11.3. Mendekotasun eta independentzia estatistikoa.

11.1. Probabilitate baldintzatua.

Gertaera jakin bat gauzatzea, beste gertaera bat betetzearen mendean badago baldintza agertzen da. Horren bidez, informazio garrantzitsua gehitu ohi da gertaeraren probabilitatea kalkulatzeko.

Ez da gauza bera kreditua eskatzea, interes-tasa kontuan hartu gabe, edo eskatzea interes-tasa finkoa denean. (interes-tasa *A* izanez gero eskatuko da, bestela ez)

Grafikoki, probabilitate baldintzatua ikus daiteke, hau da:



Baldin eta B gertaera gauzatu bada, Ω B-ra mugatzen da eta B gertaera betez gero, A-ren probabilitatea (A∩B)-rena izango da.

$$P_r(A/B) = \frac{P_r(A \cap B)}{P_r(B)}; \qquad P_r(B) \geq 0$$

Propietatea→ $P_r(A/B) = \dfrac{P_r(A \cap B)}{P_r(B)} \Rightarrow P_r(B/A) = \dfrac{P_r(B \cap A)}{P_r(A)}$

$$P_r(A \cap B) = P_r(B \cap A)$$

$$P_r(B)P_r(A/B) = P_r(A)P_r(B/A) \Rightarrow P_r(B/A) = \frac{P_r(B \cap A)}{P_r(A)}$$

11.2. Ebaketa-teorema, partiketa-teorema eta Bayesen teorema

A. EBAKETA TEOREMA---

$$P_r(S_1 \cap S_2 \cap S_3 \cap S_4) = P_r(S_1)P_r(S_2 / S_1)P_r(S_3 / S_1 \cap S_2)$$

Lehenengo eta behin, teoreman sartutako probabilitate baldintzatuak ondo definituta daudela egiaztatuko dugu.

$H = S_1 \cap S_2 \cap S_3$ izanik:

$$P_r(S_1 \cap S_2 \cap S_3 \cap S_4) = P_r(H)P(S_4 / H) =$$
$$= P_r(S_1 \cap S_2 \cap S_3)P_r(S_4 / S_1 \cap S_2 \cap S_3) \ .$$

$H' = S_1 \cap S_2$ izanik:

$$P_r(S_1 \cap S_2 \cap S_3) = P_r(H')P_r(S_3 / H') =$$
$$= P_r(S_1 \cap S_2)P_r(S_3 / S_1 \cap S_2) = P_r(S_1)P_r(S_2 / S_1)P_r(S_3 / S_1 \cap S_2)$$

$$P_r(S_1 \cap S_2 \cap S_3 \cap S_4) = P_r(S_1)P_r(S_2 / S_1)P_r(S_3 / S_1 \cap S_2)\ P_r(S_4 / S_1 \cap S_2 \cap S_3)$$

<u>Adibidea</u>: Herri jakin bateko erroldaren bidez, honako probabilitate hauek estimatu dira:
Arraza zurikoa izateko probabilitatea: $P_r(Z) = 0'8$.
Arraza zurikoa izanik emakumezkoa izateko probabilitatea: $P_r(E/Z) = 0'52$
Arraza zuriko emakumezko bat ezkongabea izateko probabilitatea: $P_r(EG/Z \cap E) = 0'4$.
Kalkula ezazu herri horretan zoriz aukeratutako pertsona bat emakumezkoa, arraza zurikoa zein ezkongabea izateko probabilitatea.

$$P_r(Z \cap E \cap EZ) = P_r(Z)P_r(E/Z)P_r(EZ/Z \cap E) = 0'8 \bullet 0'52 \bullet 0'4 = 0'0764$$

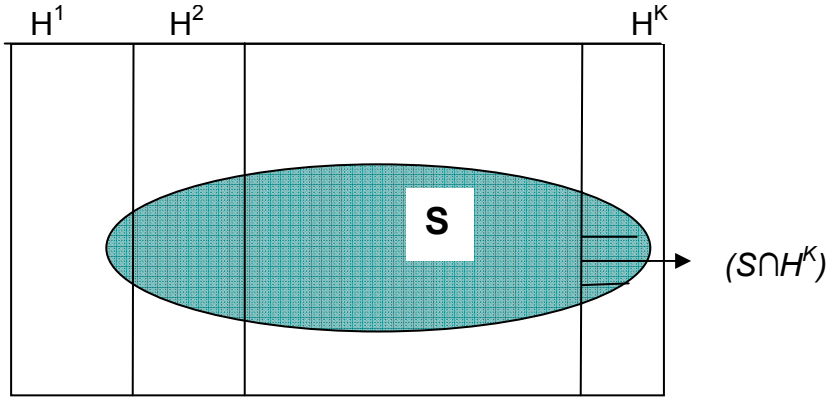
B. PARTIKETA TEOREMA.---

H^1, H^2, …, H^k gertaera multzoa, non $P_r(H^i)>0$ baita $\forall$ irako, Ω-ren partiketa da binaka bateraezinak eta haien bildura Ω direnean:

1) $H^i \cap H^j = \phi \quad \forall i \neq j$
2) $H^1 \cup H^2 \cup \ldots \cup H^k = \Omega$

<u>Teorema</u>→ Baldin eta gertaerek Ω-ren partiketa osatu eta probabilitate positiboak badituzte, honako hau egiaztatzen da:

$$P_r(S) = P_r(H^1)P_r(S/H^1) + P_r(H^2)P_r(S/H^2) + \ldots + P_r(H^k)P_r(S/H^k)$$



$$S = \left(H^1 \cup H^2 \cup \ldots \cup H^k\right) \cap S = \left(S \cap H^1\right) \cup \left(S \cap H^2\right) \cup \ldots \cup \left(S \cap H^k\right)$$

H^i gertaerak binakako bateraezinak direnez gero ebakidurak S-rekiko ere izango dira.

$$P_r(S) = P_r\left(S \cap H^1\right) + P_r\left(S \cap H^2\right) + \ldots + P_r\left(S \cap H^k\right)$$

Eta probabilitatea baldintzatua kontuan hartuz:

$$P_r(S) = P_r(H^1)P_r(S/H^1) + ... + P_r(H^k)P_r(S/H^k)$$

Adibidea: Sistema batek funtzionamendu-denboraren % 30ean energia elektrikoa jasotzen du, % 60an hidraulikoa eta gainontzeko denboran, % 10, energia mekanikoa. Elektrikoki ibiltzean, matxuratzeko probabilitatea 0´002 izaten da, hidraulikoki ibiliz gero 0´001 eta mekanikoki, berriz, 0´05. Kalkula ezazu matxuratzeko probabilitatea.

$$P_r(Mt) = P_r(M)P_r(Mt/M) + P_r(H)P_r(Mt/H) + P_r(E)P_r(Mt/E) = 0'0062$$

C. BAYESEN TEOREMA.---

Adibidea: demagun, produktu bat gaizki dagoen jakiteko analisi bat dagoela eta analisiaren fidagarritasuna % 95 dela. Hau da, produktua gaizki baldin badago, analisiak aldien % 95etan emango du positiboa, edo ondo egonez gero, aldien % 95etan ere emango du negatiboa. Halaber, demagun, produktuen % 10 gaizki dagoela. Kalitate-kontroleko atalburuak produktu batek positiboa eman duela jakinarazten badigu, artikulu akastuna izateko probabilitatea zein izango da orain?

Aurreko egoeran, 1.000 artikulu ditugunez gero, horietariko 900 zuzenak izango dira eta gainerako 100ak akastunak. Akastunei kalitate-analisia egiten badiegu honako emaitza hauek izango ditugu: ***95 artikulutan positiboa*** (analisiak asmatu egin du) eta 5 artikulutan negatiboa (analisiak huts egin du). Zuzenei kalitate-analisia eginez gero, berriz, emaitzak hauek izango dira: ***45 artikulutan positiboa*** (analisiak huts egin du) eta 855 artikulutan negatiboa (analisiak asmatu egin du). Beraz, analisiak 140 artikulu akastun dagoela dio, zein izango da benetakoa izateko probabilitatea? Probabilitatea honela kalkulatuz gero:

$$\frac{A.K}{K.P} = \frac{95}{95+45} = 0'68$$

Teoremaren ideia honako hau da: ***S*** gertatu da (artikulua akastuna dela esan du analisiak), alabaina, benetakoa den ez dakigu. Teoremaren bidez, benetako akastuna izateko probabilitatea jakingo dugu; aldez aurretiko probabilitatea (a priori probabilitatea) ondorengo probabilitatea (a posteriori probabilitatea) izatera pasatzen da.

$$P_r(H^n/S) = \frac{P_r(H^n \cap S)}{P_r(S)} = \frac{P_r(S/H^n)P_r(H^n)}{\sum_i P_r(S/H^i)P_r(H^i)}$$

11..3. Dependentzia eta independentzia estatistikoa.

A eta *B* gertaerek $P_r(A/B) = P_r(A)$ betetzen badute, interpretatu behar dugu *A* gertaera ez dagoela *B* gertaeraren menpean. Hau da, gertaerak elkarrekiko askeak dira, A B-rekiko askea baldin bada B A-rekiko ere askea baita.

Bi gertaera askeak izango dira baldintza hauetariko bat betetzen badute

1. $P_r(A/B) = P_r(A)$
2. $P_r(B/A) = P_r(B)$
3. $P_r(A \cap B) = P_r(A)P_r(B)$;

$$P_r(A/B) = \frac{P_r(A \cap B)}{P_r(B)} = \frac{P_r(A)P_r(B)}{P_r(B)} = P_r(A)$$

A. B eta *C* elkarrekiko askeak izango dira baldin eta:

1. $P_r(A \cap B) = P_r(A)P_r(B)$
2. $P_r(A \cap C) = P_r(A)P_r(C)$
3. $P_r(B \cap C) = P_r(B)P_r(C)$
4. $P_r(A \cap B \cap C) = P_r(A)P_r(B)P_r(C)$

A eta B askeak izanez gero, **A^c eta B, A eta B^c , A^c eta B^c** ere izango dira.

Ariketak

1. Karta sorta batetik karta bat ateratzen dugu:
a) Figura atera dela jakinik, zein da bikoitia izateko probabilitatea?
b) Bastoia atera dela jakinik, zein da bikoitia izateko probabilitatea?
c) Urrea atera dela jakinik, zein da figura bat izateko probabilitatea?
d) Figura atera dela jakinik, zein da ezpata bat izateko probabilitatea?
e) Bikoitia atera dela jakinik, zein da figura izateko probabilitatea?

2. Karta sorta batetik karta bat ateratzen dugu.
a) Urrea eta figura gertaera askeak al dira?
b) Ezpata eta bikoitiak gertaerak askeak al dira?
c) Bikoitia eta figura gertaerak askeak al dira?

3. Ogasunera heldutako errentaren aitorpenen % 15ek akatsen bat duela eta ikuskaritzaren fidagarritasuna % 95ekoa dela jakinik, zera eskatzen da:
a) Ikuskaritzak ez zuzentzat hartuz gero, aitorpen bat zuzena ez izateko probabilitatea
b) Aitorpen bat zuzentzat hartzeko probabilitatea.
c) Ikuskaritzak erratzeko probabilitatea.

4. Demagun HIESarako diagnostiko-proba batek honako hau betetzen duela:
$P_r(A/H)=0'95$ $P_r(A^c/H^c)=0'95$
H: "pertsona HIESak jota dago" eta A: "testak dio pertsona HIESak jota dagoela" gertaerak izanik.
Kalkula ezazu testean positibo eman duen pertsona bat benetan HIESak jota egoteko probabilitatea, ondorengo baldintza desberdin bakoitzaren pean:
a) Testa $P_r(H)=0'0005$ betetzen duen populazio baten pertsonei aplikatzen zaie.
b) Testa $P_r(H)=0'5$ betetzen duen populazio baten pertsonei aplikatzen zaie.

5. Mekanika lantegi batean hiru makina erabili ziren pieza sorta bat ekoizteko. Demagun artikuluen % 20 M1 makinaz eta % 30 M2z ekoiztu zirela. Horretaz gain, badakigu M1 makinaz fabrikatutakoen % 1, M2z fabrikatutakoen % 2 eta M3z fabrikatutakoen % 3 akastunak direla.
a) Pieza bat zoriz ateratzen da, pieza hori akastuna izateko probabilitatea zein izango da?
b) Pieza bat zoriz aukeratzen da eta ez da akastuna. M2 makinaz fabrikatutakoa izateko probabilitatea zein izango da?
c) Pieza bat zoriz aukeratzen da eta akastuna da, M2 makinaz fabrikatutako pieza izateko probabilitatea zein izango da?