

## **2.GAIA: DATUEN ANALISIRAKO ALDAGAI ANITZEKO TEKNIKEI SARRERA**

**2.1 Sarrera**

**2.2 Aldagai bakarreko eta bi aldagaitako datuen analisia**

**2.3 Aldagai anitzeko analisi teknikak**

**2.4 Menpekotasuneko aldagai anitzeko teknikak**

**2.5 Elkarmenpekotasuneko aldagai anitzeko teknikak**



OCW 2016

**Charterina Abando, Jon  
Rincón Diez, Virginia  
Villalba Merlo, Javier**

## 2.1 SARRERA

**Analisi teknikek ezin dute gaizki planteatutako MI lan bat konpondu.**

### **ANALISI MAILAK**

- 1. Aldagai bakarreko analisia**
- 2. Bi aldagaitako analisia**
- 3. Aldagai anitzeko analisia**

## 2.2 BI ALDAGAITAKO DATUEN ANALISIA

### BI ALDAGAITAKO KONTRASTE PROBAK ETA KOEFIZIENTEAK

1 aldagaia	2 aldagaia	Teknikak
Nominala	Nominala	Chi-karratuaren kontrastea bi aldagairako
Nominala	Ordinala	Chi-karratuaren kontrastea bi aldagairako
Nominala	Metrikoa <sup>(*)</sup>	Batez bestekoen berdintasun kontrastea Proporzioen berdintasun kontrastea Faktore bakarreko bariantzaren analisia
Ordinala	Ordinala	Spearmenen korrelazio koefizientea Chi-karratuaren kontrastea bi aldagairako
Ordinala	Metrikoa	Spearmenen korrelazio koefizientea Batez bestekoen berdintasun kontrastea Proporzioen berdintasun kontrastea Faktore bakarreko bariantzaren analisia
Metrikoa	Metrikoa	Pearsonen korrelazio koefizientea Erregresio lineal sinplea

<sup>(\*)</sup> tartezkoa edo arrazoizkoa

**a) CHI-KARRATUAREN KONTRASTEAK BI ALDAGAIRAKO:**

Bi aldagai ordinal edo nominalen artean erlaziorik dagoen zehazteko balio du.

**Adibidez: Ondoko taulak SEXUA eta LAN-KATEGORIA aldagaiei buruzko informazioa ematen du, elkarrizketatutako langile talde batentzat.**

**Tabla de contingencia: Sexua \* Lan-kategoria**

Recuento

		Lan-kategoria		Totala
		Administraria	Zuzendaria	
Sexua	Gizonezkoa	157	101	258
	Emakumezkoa	206	10	216
	Totala	363	111	474

Taulako behatutako maiztasunetan oinarrituz itxarondako maiztasunak kalkulatu dira. Hauek erlaziorik egongo ez balitz itxaron zitezkeen maiztasunak dira.

$$I_{ij} = \frac{\text{Lerroko totala}_i \times \text{Zutabeko totala}_j}{\text{Lerro eta zutabeetako totala}}$$

**Tabla de contingencia: Sexua \* Lan-kategoria**

		Lan-kategoria		Totala
		Administraria	Zuzendaria	
Sexua Gizonez.	Recuento	157	101	258
	Frecuencia esperada	197,6	60,4	258,0
Emak.	Recuento	206	10	216
	Frecuencia esperada	165,4	50,6	216,0
Totala	Recuento	363	111	474
	Frecuencia esperada	363,0	111,0	474,0

Demagun  $x$  aldagaia totalean  $i = \{1,2,3,\dots, r\}$  kategoria dituen eta demagun beste  $y$  aldagai bat  $j = \{1,2,3,\dots, s\}$  kategoria dituen

Hipotesiaren kontrasteak ondokoak dira:

$H_0$ : Ez dago erlaziorik  $x$  eta  $y$  artean

$H_1$ :  $x$  eta  $y$  erlazionaturik daude.

$$\chi^2 = \sum_{j=1}^s \sum_{i=1}^r \frac{(B_{ij} - I_{ij})^2}{I_{ij}}$$

$$\chi^2 = \sum_{j=1}^s \sum_{i=1}^r \frac{(B_{ij} - I_{ij})^2}{I_{ij}}$$

$B_{ij}$  x-en i kategoriari eta y-ren j kategoriari dagozkion behatutako kasuen edo banakoen multzoa

$I_{ij}$  x-en i kategoriari eta y-ren j kategoriari dagozkion itxarondako kasuen edo banakoen multzoa



## Kontrastearen emaitza

### Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	78,110 <sup>a</sup>	1	,000		
Corrección por continuidad <sup>b</sup>	76,197	1	,000		
Razón de verosimilitudes	89,577	1	,000		
Estadístico exacto de Fisher				,000	,000
Asociación lineal por lineal	77,945	1	,000		
N de casos válidos	474				

a. 0 casillas (,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 50,58.

b. Calculado sólo para una tabla de 2x2.

**b) SPEARMANen KORRELAZIO KOEFIZIENTEA:**

$$r_s = 1 - 6 \sum_{i=1}^n \frac{d_i^2}{n(n^2 - 1)}$$

**non:**  $d_i^2 = (1 \text{ aldagaian ordena zbk.} - 2 \text{ aldagaian ordena zbk.})^2$

## Adibidea:

<b>Herraldea</b>	<b>Lan esparrua (1)</b>	<b>Ordena zbk. (1)</b>	<b>Esparru pribatua (2)</b>	<b>Ordena zbk. (2)</b>
Alem ania	58,50	2	41,50	10
Belgika	43,40	7	56,60	5
Brasil	39,50	9	60,50	3
Kanada	58,10	3	41,90	9
España	52,60	5	47,40	7
EE .BB.	64,90	1	35,10	11
Frantzia	36,10	10	63,90	2
Italia	34,70	11	65,30	1
Herbehereak	56,20	4	43,80	8
Erresuma Batua	47,70	6	52,30	6
Suedia	41,60	8	58,40	4

Herrialdea	Ordena zbk. (1)	Ordena zbk. (2)	$d_i$	$d_i^2$
Alemania	2	10	- 8	64
Belgika	7	5	2	4
Brasil	9	3	6	36
Kanada	3	9	- 6	36
España	5	7	- 2	4
EE.BB.	1	11	- 10	100
Frantzia	10	2	8	64
Italia	11	1	10	100
Herbehereak	4	8	- 4	16
Erresuma Batua	6	6	0	0
Suedia	8	4	4	16

$$r_S = 1 - 6 \frac{440}{11(11^2 - 1)} = -1$$

$$\sum_{i=1}^{11} d_i^2 = 440$$

## **ABANTAILA NAGUSIAK**

- **Korrelazioa aztertzeko aukera ematen du bai aldagai ordinaletarako bai aldagai metrikoetarako.**
- **Balioen kopuruak ez dio eragiten, seriean balioek duten ordena besterik ez baitu kontuan hartzen.**

### c) PEARSONen KORRELAZIO KOEFIZIENTEA:

Bi aldagai metrikoen arteko erlazio lineal maila zehazten du. Demagun n banako dituen multzo bat x eta y aldagaien arabera aztertzen dugula:

$(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_{n-1}, y_{n-1}), (x_n, y_n)$

r korrelazio linealeko koefizientea honela kalkulatzen da:

$$r = \frac{\sum_{i=1}^n \frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1}}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}}} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$r_s$  eta  $r$ -ren arteko ezberdintasunak:

$r$  bi datu serieen arteko linealtasuna neurtzen du.

Grafikoki, demangun  $\{A, B, C, D, E, F\}$  banakoen talde bat.

Banako hauek  $x$  eta  $y$  aldagaietarako hartzen dituzten balioak

perfektuki eta positiboki koerlazonaturik daude eta hurrengoa

betetzen dute:

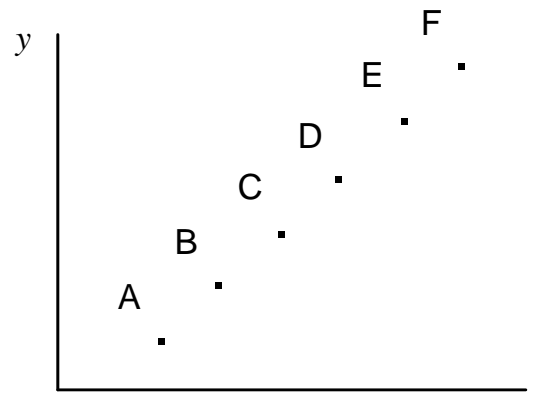
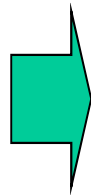
$$x_A < x_B < x_C < x_D < x_E < x_F$$

$$y_A < y_B < y_C < y_D < y_E < y_F$$

Ondoko bi egoeretatik edozein gerta daiteke eta orduan  $r$

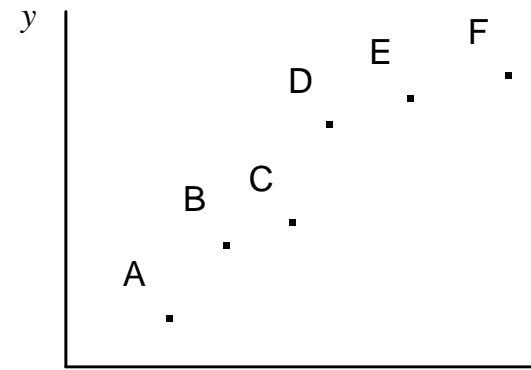
aldatu daiteke baina  $r_s$  ez:

# Korrelazio positiboa



$r_S = 1$  eta  $r = 1$

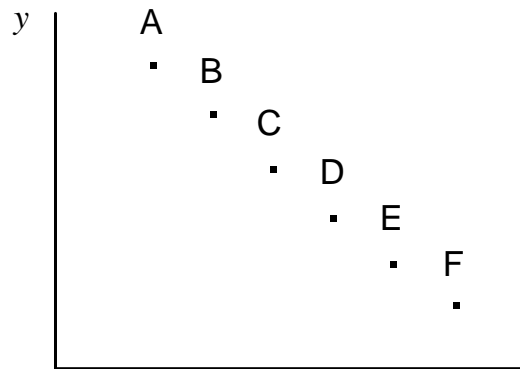
x



$r_S = 1$  baina  $r \neq 1$

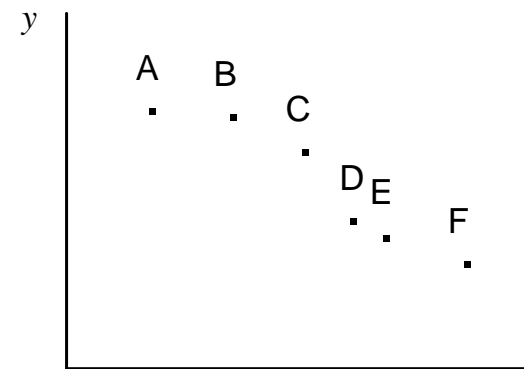
x

# Korrelazio negatiboa



$r_S = -1$  eta  $r = -1$

x



$r_S = -1$  baina  $r \neq -1$

x



## **D) BATEZ BESTEKOEN BERDINTASUN KONTRASTEAK:**

**2 egoera**

**➤ ALDAGAI BATEN BATEZ BESTEKOA BI KOLEKTIBOTAN  
KONPARATZEN DA**

**➤ BI ALDAGAI KONPARATZEN DIRA KOLEKTIBO BEREAN**

## 2.3 ALDAGAI ANITZEKO ANALISI TEKNIKAK

### Analisi teknikei esker:

- Informazioa laburtu daiteke eta informazioa interpretatu daiteke ikertzailea datu gehiegirekin nahastu dadin ekidituz.
- Arazo bati buruzko hipotesiak konfirmatu edo baztertu daitezke
- Kausa motako aldagai batzuk, enpresak aztertu behar dituen beste aldagai batzuegan duten efektua behatu daiteke.

### TRESNA ERABILGARRIAK

- Ezin dute gaizki planteatutako merkatal ikerketa lan bat konpondu.
- Informazioaren analisiak ondo diseinatutako ikerketa bat hondatu dezake.

## Bi maila

1 Datuak begiratzuz agerikoen diren erlazioak aurkitzea. (zentru joeraren neurriak, sakabanatze neurriak, forma neurriak, maiztasunak, portzentaiak e.a.)

2 Analisi sakon bat egin gabe erraztasunez aurkitu ezin daitekeen beste erlazio mota batzuk aurkitzea. **Aldagai anitzeko teknikak.**

## Analisis burutu ondoren

**Ondorioek,** era laburtuan, ikerketaren hasieran planteatutako hipotesien egiaztapena edo baztertzea adierazten dute.

**TXOSTENA:** ahozkoa eta idatzia

- Lanaren emaitzak ukiezinak izaten dira. Txostenak zerbitzua ukigarri bihurtzen du.
- Txostena eta bere aurkezpena, sarritan, marketin arduradun askok ezagutuko dituzten ikerketaren etapa bakarrak dira.

**“Txostenaren kalitatearen eta bere aurkezpenaren menpe egongo da ikertzailea etorkizunean beste lan bat egiteko deitzea”.**



Laginaren elementu edo banako bakoitzarentzat aldi berean hainbat aldagai aztertzen dituzten metodo estatistikoak.

Aldagai ezberdinen **baterako efektuak** jasotzen dituzte (aldagaiak bakoitza bere aldetik aztertuz ezingo genituzke efektu hauek antzeman)

**Oso garrantzitsua da teknikarik egokiena identifikatzen jakitea**

## **IKERTZAILEAREN HELBURUA, ZEIN DA DESKRIBATZEA EDO ESPLIKATZEA DA?**

### **TEKNIKA DESKRIBATZAILEAK**

**Kasu honetan ikertzailea datuetara hurbiltzen da hasierako hipotesirik gabe. Errealitatea behatzen da eta informazioa aztertzen da ezagupen berrien bila.**

### **TEKNIKA ESPLIKATZAILEAK**

**Metodo esplikatzaileetan ordea ikertzaileak modelo edo funtzio baten bidez adierazitako hipotesi bat konprobatu nahi du, laginaren datuekin kontrastatuz.**

# **MENPEKOTASUNEKO EDO ELKARMENPEKOTASUNEKO ERLAZIOAK AZTERTUKO DITUGU?**

## **MENPEKOTASUNEKO TEKNIKAK**

Teknika hauetan aztertutako aldagaiak bi taldetan sailkatzen dira, menpekoak eta independenteak.

Baliteke aldagai dependente bakarra egotea eta hainbat aldagai independente. Edo hainbat aldagai dependente, hainbat aldagai independenteen menpe.

## **ELKARMENPEKOTASUNEKO TEKNIKAK**

Aزتتutako aldagaiak elkarrekiko menpekotasuna duten aldagai multzo bat osatzen dute.

## **ZEIN NEURRI ESKALAN ADIERAZTEN DIRA ALDAGAIAK?**

### **TEKNIKA METRIKOAK**

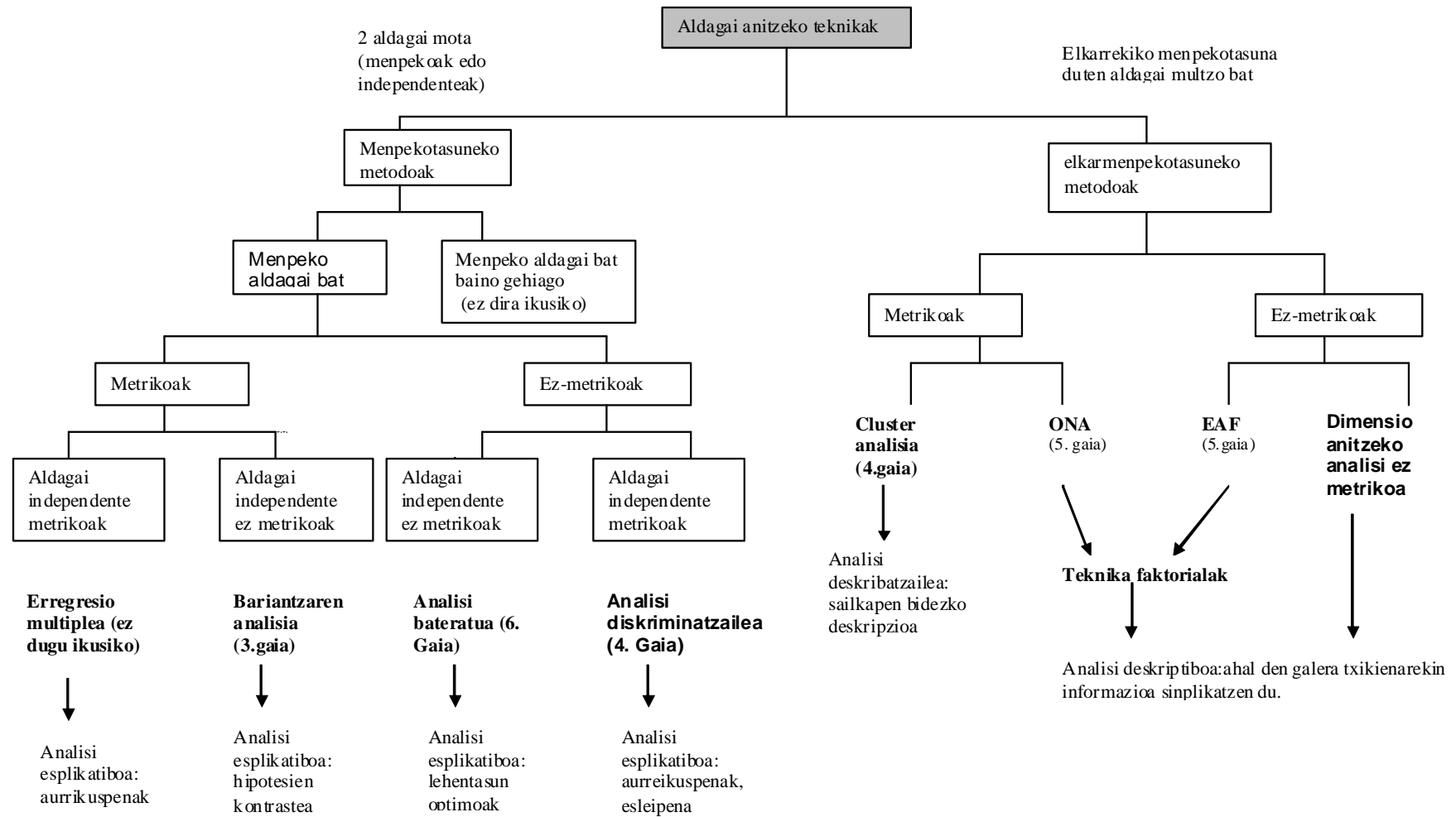
**Eskala metrikoan adierazitako datuak aztertzen dituzte**

### **TEKNIKA EZ METRIKOAK**

**Eskala ordinal edo nominalean adierazitako datuak aztertzen dituzte**



## Aldagai anitzeko tekniken sailkapena.



## 2.4 MENPEKOTASUNEKO ALDAGAI ANITZEKO TEKNIKAK

### ERREGRESIO MULTIPLEA

Aldagai independenteen aldaketen aurrean sortzen diren menpeko aldagaiaren aldaketak aurreikustea, edo orokorrean aldagai independenteek menpeko aldagaiarekin duten erlazioa aztertzeko

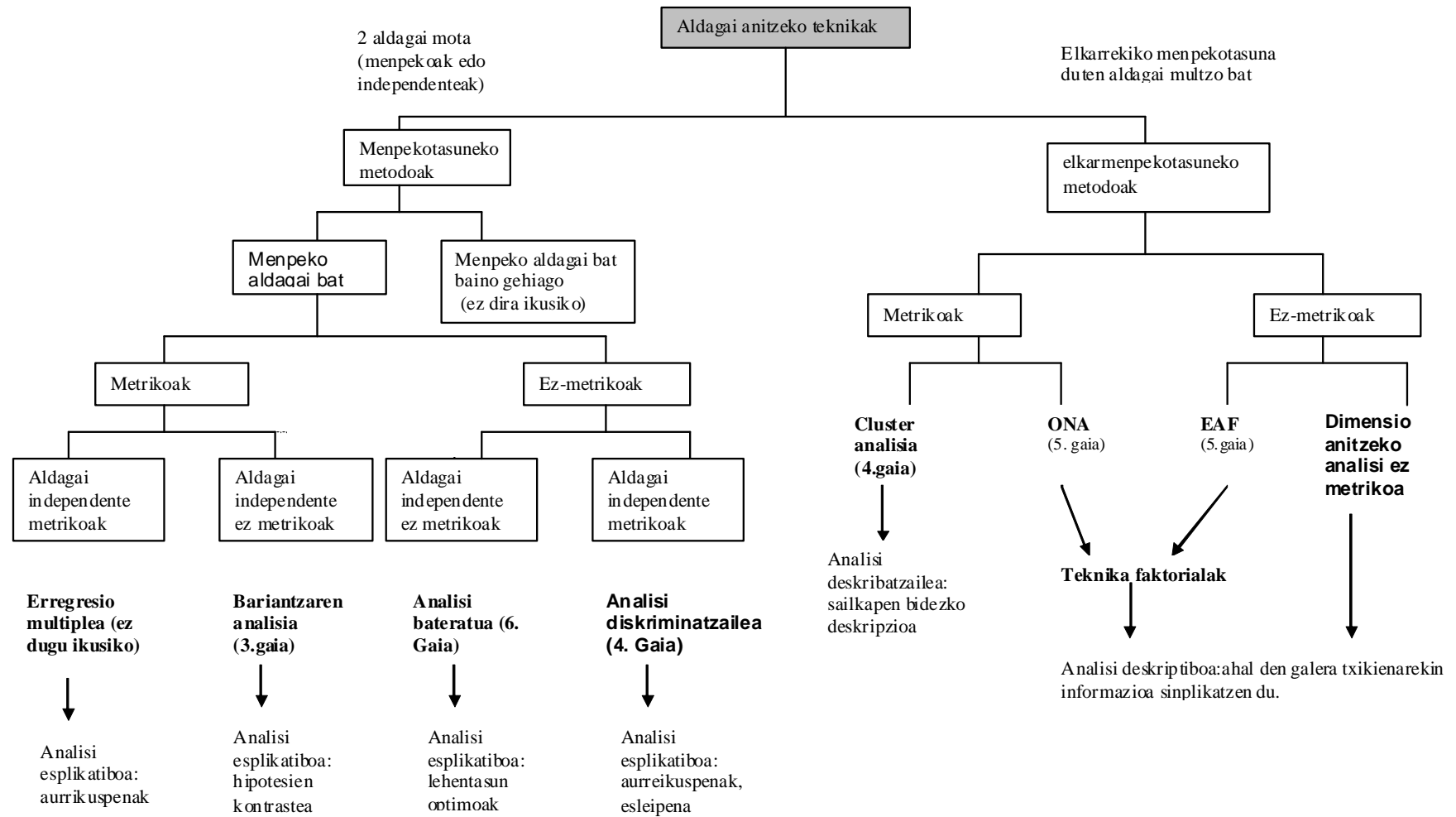
#### **Teknika esplikatiboa**

- Ekuazioa aurreikusi nahi den fenomeno adierazten duen modeloa da
- Aldagaien aukeraketa eta hauen antolaketa analistaren irizpidearen menpe daude

#### **Aurreikuspen teknika**

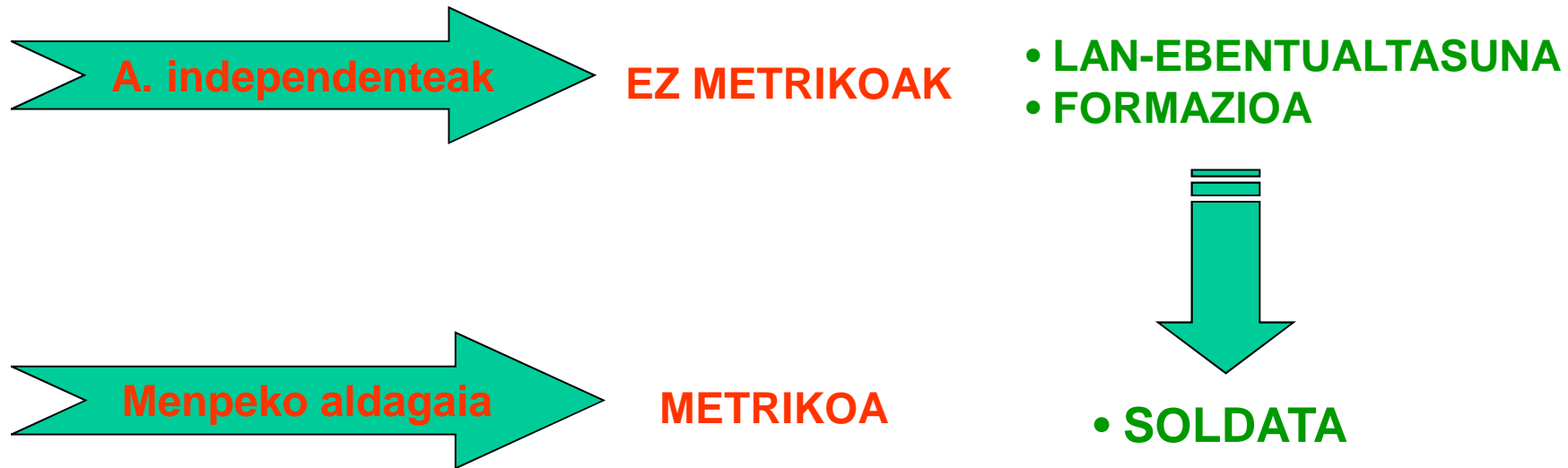
- Menpeko aldagaiaren aldaketak aurreikusi aldagai independenteen balio ezagunetatik abiatuz

## Aldagai anitzeko tekniken sailkapena.



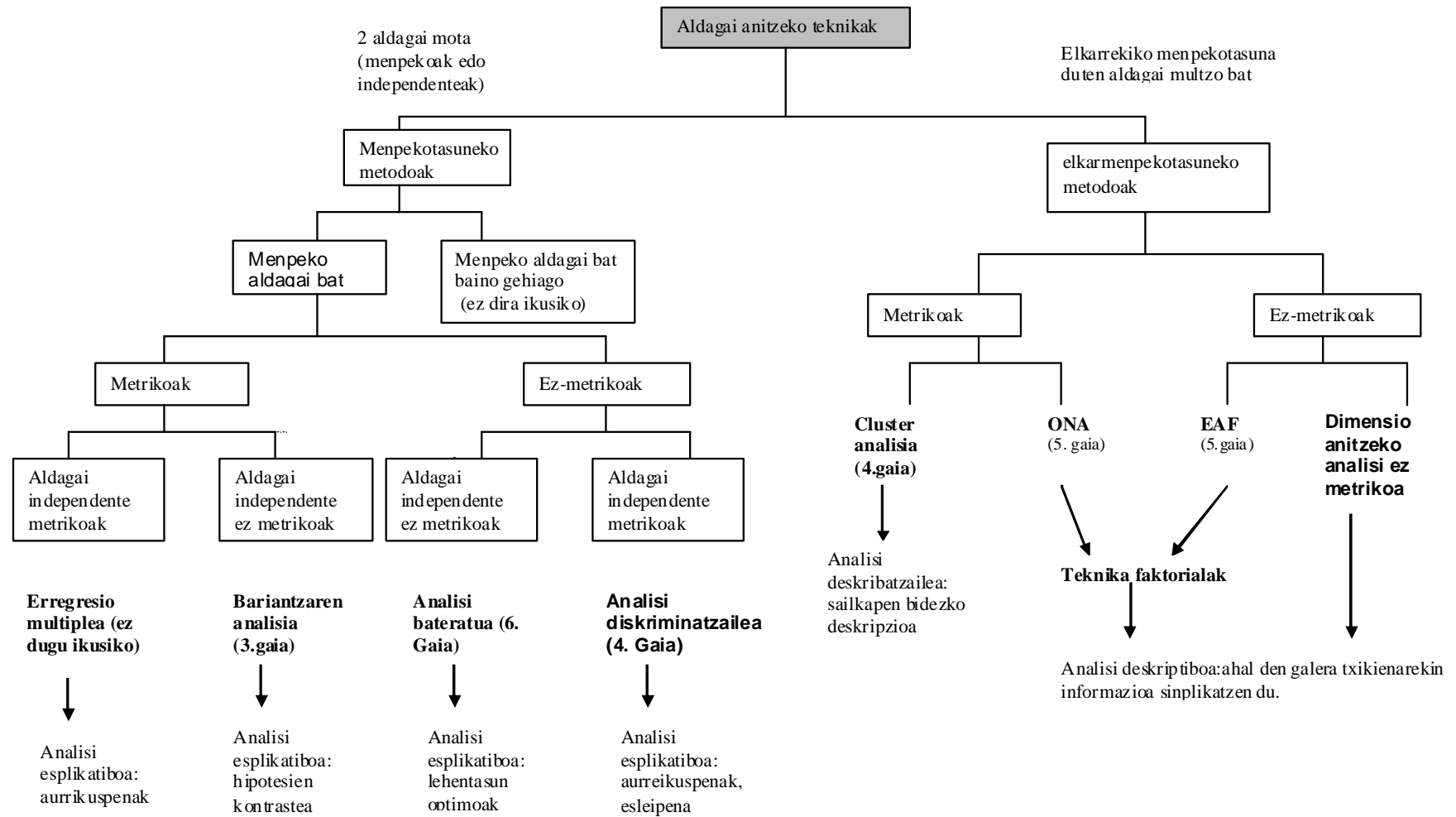
# BARIANTZAREN ANALISIA

Oso egokia analisi kausalerako edo esperimentaziorako



**Teknika esplikatzailea**

## Aldagai anitzeko tekniken sailkapena.



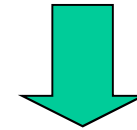
# ANALISI DISKRIMINANTEA

Oso erabilgarria, batez ere merkatuen segmentaziorako  
Taldeetara esleipena

Zeintzuk izan daitezke potentzialki berankorrak?



Hileroko soldata



Kreditu txartel kopurua

Analisi diskriminanteak erabaki arau bat ematen du. Honen arabera, oraindik sailkatu gabeko elementu berri bat emanik, honen aldagai metrikoen balioak ezagutzen direlarik, aldagai nominal batek definitutako hasierako talderen batera esleitu dezakegu.

Berankortasuna = Soldata faktorea + Kreditu txartel kopuru faktorea

$$B = a_1 S + a_2 T$$

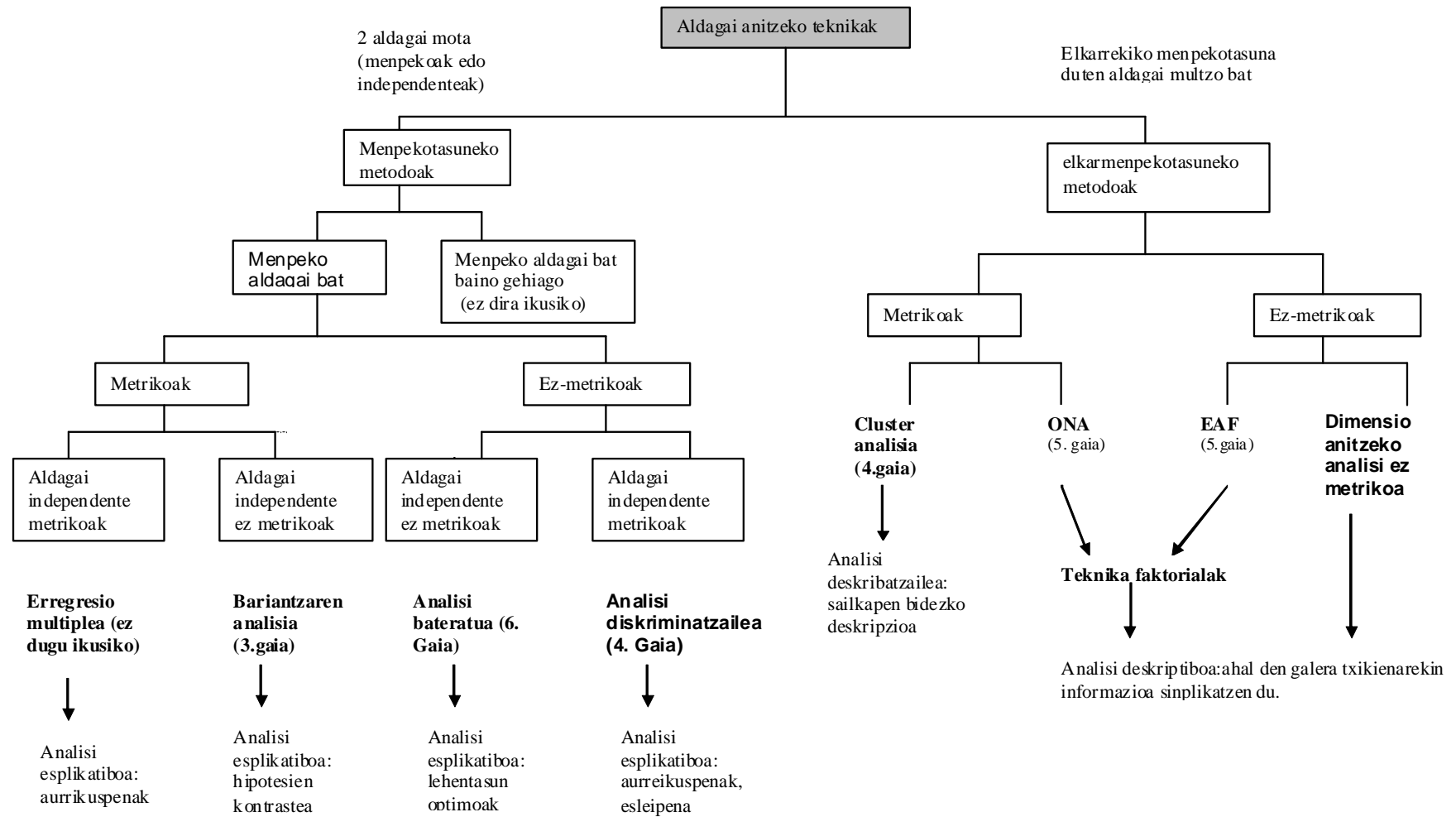
**TEKNIKA**

**ESPLIKATZAILEA**

**AURRIKUSPENEKOA**

**ESLEIPENEKOA**

## Aldagai anitzeko tekniken sailkapena.



## 2.5 ELKARMENPEKOTASUNENKO ALDAGAI ANITZEKO TEKNIKAK

Aldagai multzo batek elkarrekiko duten erlazioa aurkitzea, batzuk besteen menpe planteatu gabe.

### METODO FAKTORIALAK

Aldagai kopuru handi baten arteko erlazioak aztertzen dituzte azpian diren dimentsio edo faktore batzuen arabera azaltzeko

Aztertu beharreko jatorrizko aldagaiak faktore kopuru txikiago batean laburtzen dituzte.

### Teknika deskriptiboak dira

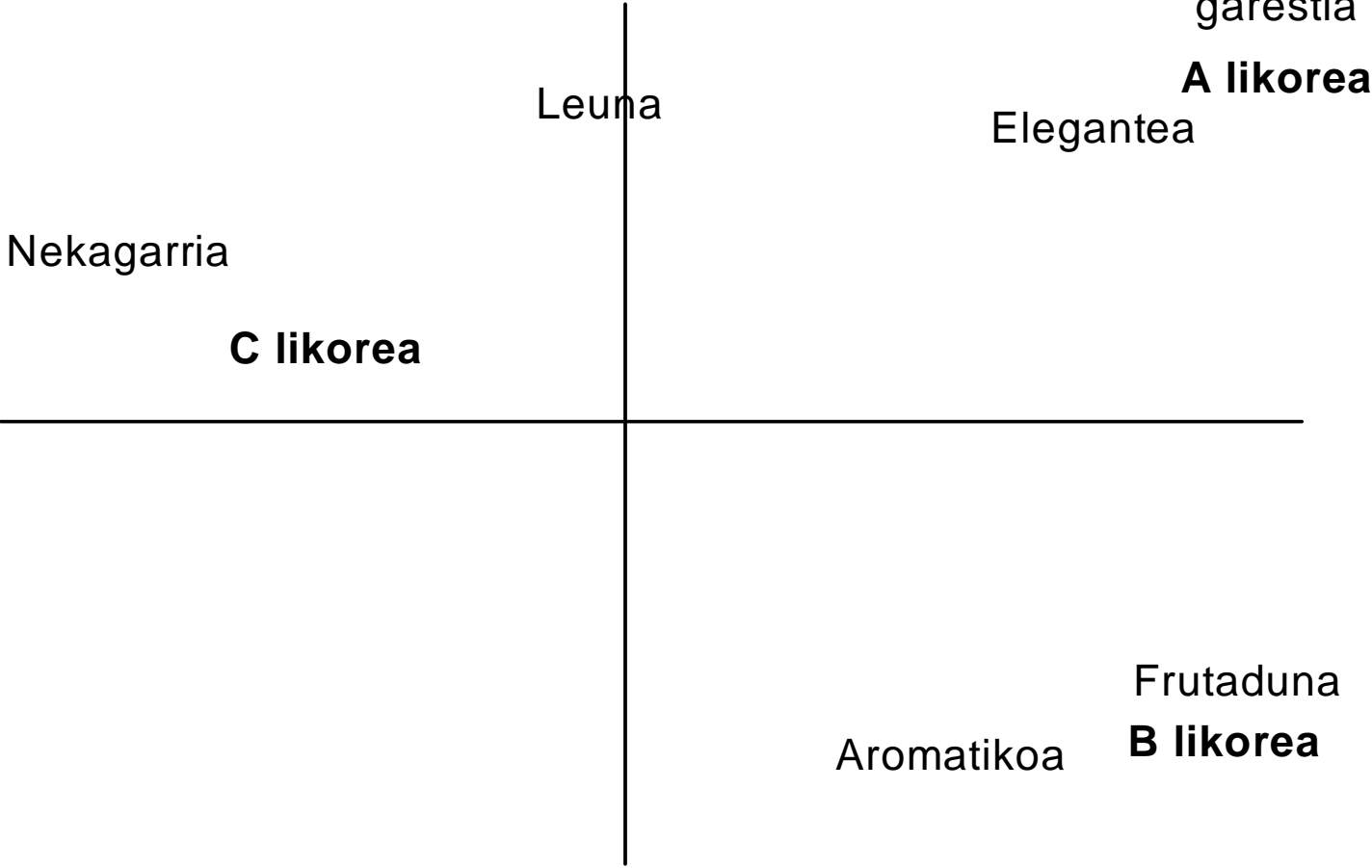
Aldagaien arteko erlazio edo patroiak aurkitu hipotesi edo modelo batetatik abiatu gabe



**Fabrikatzaile bakoitzaren posizionamendua,  
Lehiakide zuzenenak,  
Merkatuan dauden negozio aukerak,  
e.a.**







## **SAILKAPEN EDO “KLUSTER” ANALISIA**

Banako edo elementuen lagin batetatik abiatuz elkarrekiko baztertzaileak diren azpitaldeak garatzeko erabiltzen den teknika da. (segmentazioa)

Taldeak ez daude aldagai baten kategoria bezala aldez aurretik definiturik

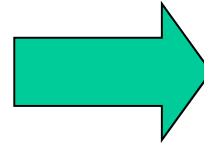
Teknika hau barnean homogeneoak eta beraien artean heterogeneoak diren taldeak eratzeko erabiltzen da

Kontuan hartu daitekeen talde kopurua analistaren irizpidearen menpe egongo da

**TEKNIKA DESKRIBATZAILEA**

## 2.4 ELKARMENPEKOTASUNENKO ALDAGAI ANITZEKO TEKNIKAK

	Prestazioa	Ekipamendua
A	1	4
B	5	4
C	1	3
D	2	1

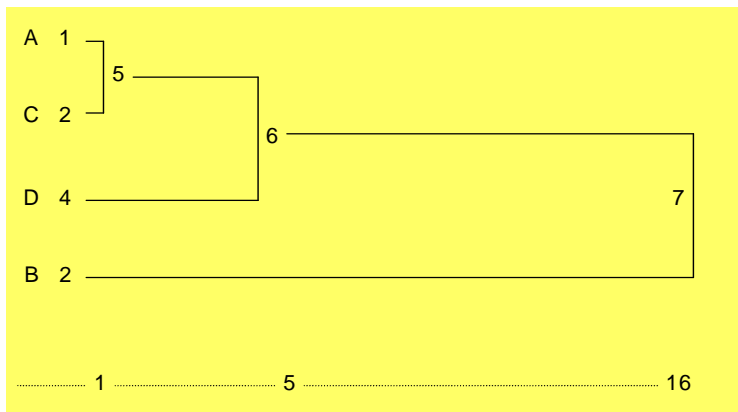
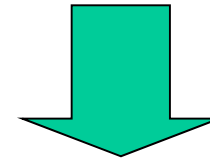


$$d^2_{A,B} = (1 - 5)^2 + (4 - 4)^2 = 16$$

	Taldea	A	B	C	D
A	1	0	16	1	10
B	2	16	0	17	18
C	3	1	17	0	5
D	4	10	18	5	0



Hurbilen dauden bi elementuak, A eta C kotxeak dira. Hauek elkartu egiten dira 5 taldea osatzeko.



	Taldea	B	D	AC
B	2	0	18	16
D	4	18	0	5
AC	5	16	5	0

6 talde berria A-k, C-k eta D-k osotuko dute.