udako
euskal
unibertsitatea

ESTATISTIKA

IRUINEA 1978

Banco de Vizcayak
BABESTUTAKO ARGITARAPENA

EDICION PATROCINADA POR EL
Banco de Vizcaya

udako euskal unibertsitatea

UDAKO VI. EUSKAL UNIBERTSITATEA

IRUINEA 1.978

ESTATISTIKA

ERREALITATEAREN IKERKETA OSAGAI NAGUSIEN ANALISIAREN BIDEZ

Egile: Anjeles Iztueta

A U R K I B I D E A

= = =

0. SARRERA

Gizonaren adimenak, ezagutzera iristen den edozein gertaera edo bibentzia, beti, aztertzeko eta argitzeko joera du.

Esan genezake erru haundirik gabe, geure pentsakera dimentsio bakarrekoa dela, errealitatea aztertzerakoan, ikuspuntu baten, egitasuna ala faltsutasuna sendotzea, besterik ez dugula helburutzat hartzen.

Bein bitartean, errealitatea infinitu dimentsio dituen espazio batetakoa da, eta, bera aztertzerakoan, milaka ikuspuntuen, egitasuna edo faltsotasuna sendotu behar genituzke.

Guttienez, pentsakera dialektiko batez, errealitatea aztertzean, beste ikuspuntu bat, ere kontuan hartzen dugu; kontrako ikuspuntua.

Baina behar genukeena, "pentsakera multidialektiko" bat izango zen.

"MULTIDIMENTSIONAL ANALISI METODOETAZ, ETA, GEHIENBAT, OSAGAI NAGUSIEN ANALISIAZ, LOR GENEZAKE, ERA ARRAZIONAL BATETAZ, GEURE PENTSAMOLDEARI, AZTERGAITZA ZAION, R^m ESPAZIO MULTIDIMENTSIONAL BAT, BESTE DIMENTSIO BAKARREKO, EDO, DIMENTSIO GUTXIKO, ESPAZIO BATETARA, ERREDUZITZEA".

Giza zientzietan, lehen urratsa, ikergai dugun errealitatea definitzea da.

Egungo pentsakera zientifikoan, errealitatea: "estruktura" bat bezala definitzen da, non errealitatea osatzen duten elementuen arteko errelazio bereziak, estrukturaren ezaugarri garrantzitzuenak bezala hartuko dira.

Estrukturak bere beitan hertsiak ez izan arren, (zeren eta lotura asko baitituzte inguruko kanpo estrukturekin), komeni da, ikerketa bakoitzean, aztergai dugun estruktura ahal dugun neurrian isolatzea eta definitzea.

" GERTAKIZUNEN ITSASOA ZEHARKATZEN DUTEN;
LEGE KORRONTEAK EZAGUTZEKO
MILAKA DATU ERABIL BEHARREAN AURKITZEN GARA
ETA NAHIZ DATU HOIEK BAKARKI INFORMAZIO GUTXI ERAMAN
ESTRUKTURA BAT ERATZEN DUTE
MULTIDIMENTSIONAL ANALISI METODOETAZ ETA GEHIEN BAT
OSAGAI NAGUSIEN ANALISIAZ
ESTRUKTURA AZTERTU ETA DEFINI DEZAKEGU ".

1. AZTERKETAREN TAULA

I. Suposa dezagun: egin ditugula p-test x_1 x_2 -- x_p edo p-aldagaien azterketa N indibiduei buruz.

A/ Aldagaiek aurkezpenean hartzen dituzten balioak zenbatu garriak direla suposatuko dugu (unitate berdinetan ala desberdinetan - neurtuak).
Adibidez: garaiera, adina, pisua, alokairua . . . aldagai zenbatugarriak dira, unitate desberdinetan neurtuak, (cm, urte, kgr, pzta . . .)
Inteligentzia, trebetasuna, mina . . . aldagai zenbatugaitzak dira.

B/ Guren aurkezpeneko N-indibiduoak equiprobableak suposatuko ditugu.
Indibiduoak izan litezke, pertsonak, herrialdeak, zabalera, geografikoak . . .

- Azterketaren erresultatuak "taula" batetan azalduko ditugu:

	1	2	—	j	—	N
X_1	X_1^1	X_2^2	—	X_1^j	—	X_1^N
X_2	X_2^1	X_2^2	—	X_2^j	—	X_2^N
|	|	|		|		|
X_i	X_i^1	X_i^2	—	X_i^j	—	X_i^N
|	|	|		|		|
X_p	X_p^1	X_p^2	—	X_p^j	—	X_p^N

"TAULA"

Non: X_i^j , " X_i " aldagaiak "j" indibiduoan hartzen duen balioa da.

X_1 aldagai bakoitzen kalkula dezakegu:

- Aurkezpenean duen media aritmetikoa

$$\bar{X}_i = \frac{1}{N} \sum_{j=1}^{N} X_i^j \qquad \forall i = 1, \text{---}, p$$

- Aurkezpenean duen bariantza, sakabanatze neurri bat bezala.

$$S_{x_i}^2 = \frac{1}{N} \sum_{j=1}^{N} (x_1^j - \bar{x}_i)^2 \quad \forall i = 1, \text{---}, p$$

$x_1, x_2 — x_p$ aldagai multzoan kalkula dezakegu, aldagaien arteko errelazio edo sinkronizazio neurri batzuk bezala:

- aldagai bikoteen kobariantzak:

$$\text{Kob}\ (x_i, x_N) = \frac{1}{N} \sum_{j=1}^{N} (x_i^j - \bar{x}_i)(x_N^j - \bar{x}_N) \quad \forall i \neq k \quad i,k = 1, \text{---} p$$

$$\text{Kob}\ (x_i\ x_i) = \frac{1}{N} \sum_{j=1}^{N} (x_i^j - \bar{x}_i)^2 = S_{x_i} \qquad \forall i = 1, \text{---}, p$$

- aldagai bikoteen korrelazioak:

$$r(x_i\ x_N) = \frac{\text{Kob}\ (x_i\ x_N)}{S_{x_i} \cdot S_{x_N}} = \frac{\sum (x_i^j - \bar{x}_i)(x_N^j - \bar{x}_N)}{\left[\sum_j (x_i^j - \bar{x}_i)\right]^{1/2} \cdot \left[\sum_j (x_k^j - \bar{x}_k)\right]^{1/2}} \qquad \forall i \neq k \quad i, K = 1, \text{---} p$$

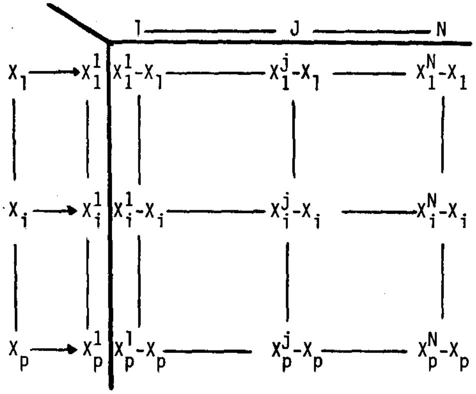
$$r(x_i\ x_i) = \frac{\text{Kob}\ (x_i\ x_i)}{S_{x_i} \cdot S_{x_i}} = \frac{S_{x_i}^2}{S_{x_i}^2} = 1 \qquad \forall i = 1, \text{---}, p$$

II. TAULAREN LEHEN TRANSFORMAZIOAK

(A) TAULA ZENTRATUA

$x_1\ x_2$ — x_p aldagaiak unitate berdinetan neurtuak badira, taula zentratu batetan transformatuko dugu, non x_i balioak beren media aritmetikotik neurtzen ditugu.

Transformazio honetan ez dugu informazio estadistikorik galtzen.



"TAULA ZENTRATUA"

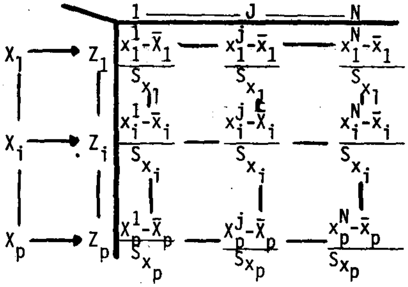
$$X_i^1 = X_i - \bar{X}_i \quad \forall i = 1, \text{—} p$$

$$\bar{X}_i^1 = \frac{1}{N} \sum_j (X_i^j - \bar{X}_i) = 0 \qquad \forall i = 1, \text{—} p$$

(B) TAULA TIPIFIKATUA

$x_1\ x_2$ — x_p aldagaiak unitate desberdinetan neurtuak badira, hasierako taula, taula tipifikatu betetan transformatuko dugu (zentratua eta standarizatua)

Transformazio honetan informazio estadistikoa galduarren (aldagai bakoitzaren sakabanakuntza neurria galtzen baitugu, indibiduoen arteko sakabanakuntza errelatiboa gordetzen badugu ere), unitate desberdinak eraginik ez dute.



"TAULA TIPIFIKATUA"

$$Z_i = \frac{x_i - \bar{x}_i}{S_{x_i}} \quad \forall_i = 1, \text{—} p$$

$$\bar{Z}_i = 0 \quad \forall_i = 1, \text{—} p; \qquad s^2_{Z_i} = 1 \quad \forall_i = 1, \text{—} p; \qquad \frac{\text{Kob } (Z_i\ Z_N)}{1 \cdot 1} = r(Z_i Z_N) \quad \forall i \neq k$$

2. TAULAREN ADIERAZPIDEAK

2.1. MATRITZEZKO ADIERAZPIDEA:

(A) - Taula zentratu baten matritzezko adierazpidea:

$$X_{p.N} = \begin{bmatrix} x_1^1 & x_1^2 & \text{---} & x_1^j & \text{---} & x_1^N \\ | & | & & | & & | \\ x_i^1 & x_i^2 & \text{---} & x_1^j & \text{---} & x_1^N \\ | & | & & | & & | \\ x_p^1 & x_p^2 & \text{---} & x_p^j & \text{---} & x_p^N \end{bmatrix}$$

Non: $x_i^j = X_i^j - \bar{X}_i$

- Media aritmetikoen zutabe matritza:

$$\bar{X}_{p.1} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ | \\ \bar{x}_i \\ | \\ \bar{x}_p \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ | \\ 0 \\ | \\ 0 \end{bmatrix}$$

non : $\bar{X}_i = \frac{1}{N} \sum_j x_i^j \quad \forall i$

- $X_1\ X_2$ — X_p -en kobariantza matritza:

$$K_{p.p}^{x_1 - x_p} = \begin{bmatrix} s_{x_1}^2 & K(x_2x_2) & \text{---} & K(x_1x_p) \\ K(x_2x_1) & s_{x_2}^2 & \text{---} & K(x_2x_p) \\ | & | & \ddots & | \\ K(x_px_2) & K(x_px_2) & \text{---} & s_{x_p}^2 \end{bmatrix}$$

$K_{p.p}$ matritza simetrikoa da:

$$\forall\, i \neq k \qquad k(x_ix_N) = k(x_Nx_i) = \frac{1}{N} \sum_j x_i^j \cdot x_N^j \quad \text{delako.}$$

- $x_1 x_2 — x_p$ -en korrelazio matritza:

$$R_{p.p}^{x_1 - x_p} = \begin{bmatrix} 1 & r(x_1x_2) & \cdots & r(x_1x_p) \\ r(x_2x_1) & 1 & \cdots & r(x_2x_p) \\ \vdots & & \ddots & \vdots \\ r(x_px1) & r(x_px_1) & \cdots & 1 \end{bmatrix}$$

$R_{p.p}$ matritza simetrikoa da:

$$\forall i \neq k \qquad r(x_ix_k) = r\,(x_kx_i) = \frac{k(x_kx_i)}{S_{x_i}\; S_{x_k}} = \frac{\sum x_i^j \cdot x_k^j}{\left[\sum (x_i^j)^2\right]^{1/2}\left[\sum (x_k^j)^2\right]^{1/2}} \text{ delako.}$$

(B) - Taula tipifikatu baten matritzezko adierazpidea:

$$Z_{p.N} = \begin{bmatrix} Z_1^1 . Z_2^2 & — & Z_1^j & — & Z_1^N \\ | \quad | & & | & & | \\ Z_i^1 , Z_i^2 & — & Z_i^j & — & Z_i^N \\ | \quad | & & | & & | \\ Z_p^1 , Z_p^2 & — & Z_p^j & — & Z_p^N \end{bmatrix} \qquad \text{non: } Z_i^j = \frac{x_i^j - \bar{X}_i}{S_{x_i}}$$

- Media aritmetikoen matritze zutabea:

$$\bar{Z}_{p.1} = \begin{bmatrix} \bar{Z}_1 \\ | \\ \bar{Z}_i \\ | \\ \bar{Z}_p \end{bmatrix} = \begin{bmatrix} 0 \\ | \\ 0 \\ | \\ 0 \end{bmatrix} \qquad \text{non: } \bar{Z}_i = \frac{1}{N} \sum_j Z_i^j \quad \forall i$$

- $Z_1\, Z_2 — Z_p$ -en kobariantza matritza:

$$K_{p.p}^{Z_1 - Z_p} = \begin{bmatrix} 1 & k(Z_1Z_2) & — & K(Z_1Z_p) \\ K(Z_2Z_1) & 1 & — & K(Z_2Z_p) \\ | & | & \ddots & | \\ K(Z_pZ_1) & K(Z_pZ_2) & & 1 \end{bmatrix}$$

$K_{p.p}^{Z_1 . Z_p}$ matritze simetrikoa da: $K(Z_iZ_k) = \frac{1}{N} \sum_j Z_i^j . Z_k^j = K(z_kZ_i) \quad \forall i \neq k$

eta $K(Z_iZ_i) = S_{Z_i}^2 \equiv \frac{1}{N} \sum (Z_i^j)^2 = 1 \quad \forall i$

2.2 GEOMETRIAZKO ADIERAZPIDEA

Taula bat emanik non:

	1	2	—	J	—	N
X_1	X_1^1	X_1^2	—	X_1^j	—	X_1^N
X_2	X_2^1	X_2^2	—	X_2^j	—	X_2^N
|	|	|		|		|
X_i	X_i^1	X_i^2	—	X_i^j	—	X_i^N
|	|	|		|		|
X_p	X_p^1	X_p^2	—	X_p^j	—	X_p^N

bere ikerketa askoz errezago bihurtzen zaigu, geometrikoki multidimensional espazio batetan adierazten badugu.

I. Gogora desagun: R^N espazio bektorial euklidearra (distantzia euklideardun espazio bektorial metrikoa):

Definitzen badugu, R^N (e.b.) -ean, eskalar biderkadura honela:

$$\langle \vec{X} . \vec{Y} \rangle = (X_1 X_2 — X_N) \begin{pmatrix} 1 & & 0 \\ & 1 \ddots & \\ 0 & & 1 \end{pmatrix} \begin{pmatrix} Y_2 \\ Y_2 \\ | \\ Y_N \end{pmatrix} = \sum_{j=1}^{N} X_j . Y_j$$

Orduan definizioz (ondorio bezala):

a) bektore baten euklidear norma:

$$\|\vec{X}\| = \langle \vec{X}.\vec{X} \rangle^{1/2} = \left[(X_1 X_1 — X_N) \begin{pmatrix} 1 & & \\ & 1 \ddots & \\ & & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \\ X_N \end{pmatrix} \right]^{1/2} = \left(\sum_{j=1}^{N} X_j^2 \right)^{1/2}$$

izango da.

b) bi bektoreen arteko euklidear distantzia:

$$d(\vec{X}_1 \vec{Y}) = \|\vec{X} - \vec{Y}\| = \left[\sum_{j=1}^{N} (X_j - Y_j)^2 \right]^{1/2}$$

c) bi bektoreek eratzen duten angeluaren kosinua:

$$\cos \varphi(\vec{X}_1 \vec{Y}) = \frac{\langle \vec{X} \cdot \vec{Y} \rangle}{\|\vec{X}\| \cdot \|\vec{Y}\|} \quad ; \quad -1 \leq \cos \varphi(\vec{X}_1 \vec{Y}) \leq 1$$

Hots: $\langle \vec{X} \cdot \vec{Y} \rangle = \|\vec{X}\| \cdot \|\vec{Y}\| \cdot \cos \varphi(\vec{X}_1 \vec{Y})$

II. R^N - (espazio bektorial euklidear)-ean, adierazten baditugu X_1, X_2 — X_p aldagaiak (suposatuko dugu guttienez zen tratuak direla):

$\vec{X}_1\ \vec{X}_2$ — $\vec{X}_p$ bektore bezala, non X_1 aldagai (zentzatu) bakoitzari,

R^N - n

$\vec{X}_i = (X_i^1\ X_i^2$ — X_i^j - $X_i^N)$ bektorea dagokio.

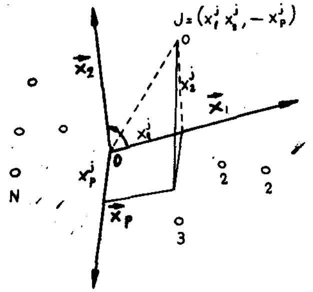
($\vec{X}_i$ bektorearen koordenadak, X_i aldagaiak indibiduo bakoitzean hartzen dituen balioak dira).

- $\langle \vec{X}_1, \vec{X}_2$ —, $\vec{X}_p$ / $\vec{X}_i = (X_i^1$ — X_i^j — $X_i^N) \rangle$ bektore multzoak azpiespazio euklidear bat eratzean dute R^N espazio euklidearrean.

- eta $S = \langle 0, \vec{X}_1\ \vec{X}_2$ — $\vec{X}_p \rangle$ erreferentzial sistema bezela harturik, "J" indibiduo bakoitza, puntu batez adieraziko dugu.

(Non "J" puntuaren p-Koordenadak: $(X_1^j\ X_2^j$ — $X_p^j)$ aldagai guztiek "J" indibiduoen hartzen dituzten balioak dira).

Hots: R^N



III. Adierazpide honetan ikus dezakegu:

(1) $\vec{X}_i$, $\vec{X}_k$, bektoreen e. <u>eskalar biderkadura</u>, N bider $X_i X_k$ aldagaien <u>kobariantza</u> dela:

Hots: $\boxed{\langle \vec{X}_i \cdot \vec{X}_k \rangle = N \cdot \text{kob}\,(X_i\, X_k)}$ $\forall$ i, k

$$\left.\begin{array}{l} \langle \vec{X}_i \cdot \vec{X}_k \rangle = \sum_{j=1}^{N} X_i^{\,j} \cdot X_k^{\,j} \\ \text{Kob}(X_i X_k) = \frac{1}{N} \sum_{j=1}^{N} (X_i^{\,j}-0)(X_k^{\,j}-0) \end{array}\right| \Rightarrow \langle \vec{X}_i \vec{X}_k \rangle = N \cdot \text{kob}(X_i X_k)$$

f.n.c.b.

(2) $\vec{X}_i$ bektorearen <u>euklidear norma,</u> $\sqrt{N}$ bider X_i aldagaiaren <u>desbidazio standarra</u> dela:

Hots: $\boxed{\|\vec{X}_i\| = \sqrt{N} \cdot S_{X_i}}$ $\forall$i

$$\|\vec{X}_i\| = \langle \vec{X}_i \vec{X}_i \rangle^{1/2} \underset{(1)}{=} \left[N \cdot \text{kob}\,(X_i X_i)\right]^{1/2} = \left[N \cdot S^2_{X_i}\right]^{1/2} = \sqrt{N}\; S_{X_i}$$

f.n.c.b.

(3) $\vec{X}_i$, $\vec{X}_k$ bektoreak eratzen duten <u>angeluen kosinua,</u>

X_i, X_k aldagaien arteko <u>korrelazio estadistikoa da</u>

Hots: $\boxed{\cos \varphi\, (\vec{X}_i\, \vec{X}_k) = r(X_i\, X_k)}$ $\forall$i, k

$$\cos \varphi\, (\vec{X}_i, \vec{X}_k) = \frac{\langle \vec{X}_i\, \vec{X}_k \rangle}{\|\vec{X}_i\| \cdot \|\vec{X}_k\|} \underset{(1)(2)}{=} \frac{N \cdot \text{kob}(X_i X_k)}{\sqrt{N} \cdot S_{X_i} \cdot \sqrt{N}\; S_{X_k}} = \frac{\text{kob}(X_i X_k)}{S_{X_i}\, S_{X_k}} = r(X_i X_k)$$

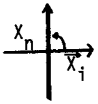
f.n.c.b.

$$\boxed{-1 \leq \cos \varphi\, (\vec{X}_i \vec{X}_k) = r\,(X_i\, X_k) \leq 1} \quad \forall \text{i, k}$$

Orduan esan dezakegu:

- <u>$\vec{X}_1 \vec{X}_n$ bi bektore adierazleek eratzen duten angelua, zenbat eta txikiago den, dagozkion aldagaien korrelazioa hainbat handiagoa dela.</u>

(a) $X_i X_n$ aldagaiak korrelaziogabeak izango dira, soilik eta baldin soilik, bere bektore adierazleak ortogonalak badira $R^N - n$

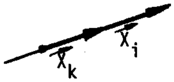


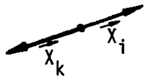
Hots: $r(X_i X_k) = 0 \longleftrightarrow \cos \varphi(\vec{X}_i \vec{X}_k) = 0$ edo $\varphi(\vec{X}_i \vec{X}_k) = 90°$;

$\vec{X}_i \perp \vec{X}_k$ $\langle \vec{X}_i \cdot \vec{X}_k \rangle = 0$

(b) $X_i X_n$ aldagaien korrelazioa ± 1 izango da, soilik eta baldin soilik beren artean menpekotasun lineal bat badago.

Hots: $r(X_i X_k) = \pm 1 \longleftrightarrow \cos \varphi(\vec{X}_i \vec{X}_k) = \pm 1$ edo $\varphi(\vec{X}_i \vec{X}_k) = 0°$, $180°$;





$\exists \lambda / \vec{X}_i = \vec{X}_k$

* Taulan existitzen bada, X_i aldagaien bat, non menpekotasun lineal bat du beste batekin X_n edo beste batzuekin X_h X_k - orduan ikerketarako X_i aldagaia baztertu egingo dugu.

IV. Orduan taulen adierazpide geometrikoa, bi kasuetan honela izango litzateke:

(A) $X_1\ X_2 — X_p$ aldagaiek zentratuak baldin badira, hots:

$\bar{X}_1 = 0,\ \bar{X}_2 = 0,\ — \bar{X}_p = 0$, lehenago esan dugunez beti lortu genezake hau, eta gainera informazio estadistikorik galdu gabe.

Eta dagokion taula zentratua; bada:

	1	2 —	j —	N
X_1	X_1^1	X_2^2 —	X_1^j —	X_1^N
X_2	X_2^1	X_2^2 —	X_2^j —	X_2^N
X_i	X_i^1	X_i^2 —	X_i^j —	X_i^N
Xp	X_p^1	X_p^2 —	X_p^j —	X_p^N

Orduan, adierazten badugu taula geometrikoki, R^N indibiduoen espazio euklidearrean:

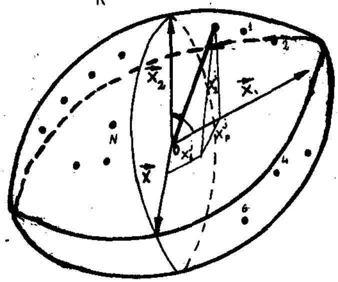
Eta baldin badira $\vec{X}_1,\ \vec{X}_2 — \vec{X}_p,\ X_1, X_2 — X_p$ aldagaieri dagozkien koordenadatako bektore adierazleak:

$$\vec{X}_1 = (X_1^1 X_2^1 — X_1^N) \in R^N$$
$$\vec{X}_i = (X_i^1\ X_i^2 — X_i^N) \in R^N$$
$$\vec{X}_p = (X_p^1\ X_p^2 — X_p^N) \in R^N$$

$\vec{X}_1, \vec{X}_2, — \vec{X}_p$ bektoreen muturrak, jatorrian zentratuta dagoen elipsoide batetan aurkitzen dira, non elipsoide honen p- ardatzak, aldagaien desbidazio standarrari proportzionalak dira, eta ez dute zergatik ortogonalak beren artean izan behar.

Hots:

R^N



$$\vec{X}_i = (X_i^1 X_i^2 — X_i^N) \in R^N$$

$$S = \langle 0, \vec{X}_1, \vec{X}_2, — \vec{X}_p \rangle$$

$$\|\vec{X}_i\| = \sqrt{N} \cdot S_{X_i} \qquad \forall_i = 1, — p$$

$$\cos \varphi (\vec{X}_i \vec{X}_k) = r(X_i X_k) \quad \forall i, k$$

$$J = (X_1^j, — X_p^j) \qquad \forall J = 1, — N$$

(B) $X_1\ X_2 — X_p$ aldagaiak tipifikatzen baldin badira, hots:

$$Z_i = \frac{X_i - \overline{X}_i}{S_{X_i}} \quad \forall i \quad ; \quad \overline{Z}_1 = 0 \ , \ \overline{Z}_2 = 0, — \overline{Z}_p = 0 \quad \text{eta} \quad S^2_{Z_1} = 1,\ S^2_{Z_2} = 1, — S^2_{Z_p} = 1$$

Dakigunez aldagaiek tipifikatzean informazioa galtzen dugu, baina honela aldagaiek neurtuta duten unitate desberdinak ez dute eraginik. Dagokion taula tipifikatua bada:

	1	2	— j —	— N
$X_1 \longrightarrow Z_1$	$Z_1^{\ 1}$	$Z_1^{\ 2}$	$Z_1^{\ j}$	$Z_1^{\ N}$
$X_2 \longrightarrow Z_2$	$Z_2^{\ 1}$	$Z_2^{\ 2}$	$Z_2^{\ j}$	$Z_2^{\ N}$
$X_i \longrightarrow Z_i$	$Z_i^{\ 1}$	$Z_i^{\ 2}$	$Z_i^{\ j}$	$Z_i^{\ N}$
$X_p \longrightarrow Z_p$	$Z_p^{\ 1}$	$Z_p^{\ 2}$	$Z_p^{\ j}$	$Z_p^{\ N}$

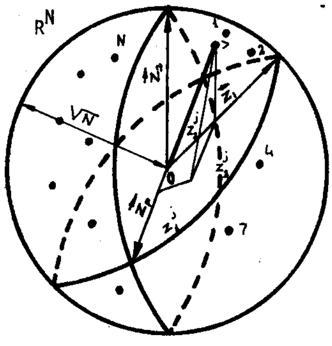
Orduan, adierazten badugu taula, geometrikoki, R^N indibiduoen espazio **euklidearrean**:

Eta **baldin** badira $\vec{Z}_1\ \vec{Z}_2 — \vec{Z}_p$, $Z_1\ Z_2 — Z_p$ aldagai tipifikatuai **dagoz**kien bektore adierazleak:

$$\vec{Z}_1 = (Z_1^{\ 1}\ Z_1^{\ 2} — Z_1^{\ N}) \quad \in R^N$$
$$\vec{Z}_i = (Z_i^{\ 1}\ Z_i^{\ 2} — Z_i^{\ N}) \quad \in R^N$$
$$\vec{Z}_p = (Z_p^{\ 1}\ Z_p^{\ 2} — Z_p^{\ N}) \quad \in R^N$$

$\vec{Z}_1\ \vec{Z}_2 — \vec{Z}_p$ bektoreen muturrak, jatorrian zentratuta dagoen esfera batetan aurkitzen dira, non esfera honen arradioa $\sqrt{N}$ izango da.

Hots:



$$\vec{Z}_i = (Z_i^{\ 1} Z_i^{\ 2} — Z_i^{\ j} — Z_i^{\ N}) \in R^N$$

$$S = \langle 0, \vec{Z}_1\ \vec{Z}_2, — \vec{Z}_p \rangle$$

$$\| \vec{Z}_i \| = \sqrt{N} \cdot S_{Z_i} = \sqrt{N}.1 \quad \forall_i = 1, — p$$

$$\cos \varphi(\vec{Z}_i \vec{Z}_k) = r(Z_i Z_k) = k(Z_i Z_k) = N(\vec{Z}_i \vec{Z}_k)$$

$$J = (Z_i^{\ j} Z_2^{\ j} — Z_p^{\ j}) \quad \forall J = 1, — N$$

3. OSAGAI NAGUSIEN ANALISIA

Suposa dezagun egin ditugula p-test $X_1\ X_2 — X_p$ edo p- aldagaien azterketa N-indibiduei buruz.

Eta erresultatuen "taula" hau dugula:

	1	2	—	J	—	N
X_1	X_1^1	X_1^2	—	X_1^j	—	X_1^N
X_2	X_2^1	X_2^2	—	X_2^j	—	X_2^N
|	|	|				
X_p	X_p^1	X_p^2	—	X_p^j	—	X_p^N

I. <u>Zer lortu genezake osagai nagusien analisitaz?</u>

Azterketako p-testak $X_1X_2 — X_p$, non gehienetan beren artean alkar korrelazionaturik daude, beste r-testara ($r < p$) <u>erreduzitu</u>, bitez $Y_1\ Y_2 — Y_r$ non hoiek <u>ez dira alkar korrelazionatuak</u>, eta <u>sakabanakuntza guztiaren % asko esplikatzen dute</u>.

Gero azterketa bakoitzean, "Osagai nagusiak", deituko ditugun "$Y_1Y_2 — Y_r$" aldagai berri hoiei, esanahi bat **emango diogu**, haserako $X_1\ X_2 — X_p$ aldagaien esanahiarekin loturik noski.

Eta osagai nagusietan gure N-indibiduoek hartzen dituzten balioak taula batetan azalduz; hots:

	1	2	—	J	—	N
Y_1	Y_1^1	Y_1^2	—	Y_1^j	—	Y_1^N
Y_2	Y_2^1	Y_2^2	—	Y_2^j	—	Y_2^N
|	|	|				
Y_r	Y_r^1	Y_r^2	—	Y_r^j	—	Y_r^N

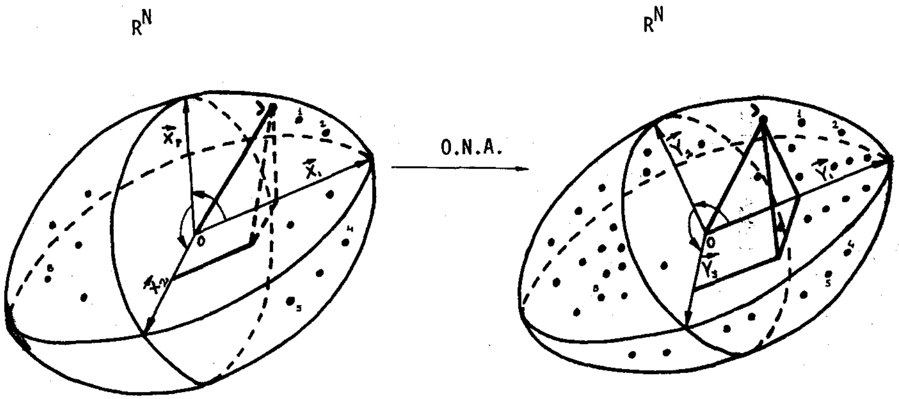
$r < p$

$$r(Y_iY_k) = 0 \quad \forall_{i \neq k} \quad i,k = 1, — r$$

$$r < p$$

N-indibiduoak aztertuko ditugu.

II. PLANTEIAMENDU GEOMETRIKOA



$S = \langle 0, \vec{X}_1 \vec{X}_2, — \vec{X}_p \rangle$

$J = (X_1^{j}\ X_2^{j}, — X_p^{j})\quad \forall J = 1, — N$

$S' = \langle 0, \vec{Y}_1\ \vec{Y}_2, — \vec{Y}_p \rangle$

non:
- a) $\vec{Y}_1 \perp \vec{Y}_2 \perp \ldots \perp \vec{Y}_p$
- b) $\|\vec{Y}_1\| \geqslant \|\vec{Y}_2\| \geqslant \ldots \geqslant \|\vec{Y}_p\|$

  edo

  $D^2_{Y_1} \leq D^2_{Y_2} \leq \ldots \leq D^2_{Y_p}$ /

  $/ D^2_{Y_i} = \Sigma_j\ Y^2(J, \vec{Y}_i)$

$J = (Y_1^{j}\ Y_2^{j}, — Y_p^{j})\quad \forall J = 1, — N$

Geometrikoki adierazten baditugu N-indibiduoak, N puntutaz, $S = \langle 0, \vec{X}_1, \vec{X}_2, — \vec{X}_p \rangle$ sisteman non ardatzak ez dute zergaitik ortogonalak izan behar beren artean, orduan "osogai nagusien analisiak" transformatzen du S sistema, beste jatorri berdina duen $S' = \langle 0, \vec{Y}_1, \vec{Y}_2, — \vec{Y}_p \rangle$ sisteman, non "$\vec{Y}_1, \vec{Y}_2 — \vec{Y}_p$": osagai nagusien bektore adierazleak izango dira, eta ondoko propietateak beteko dituzte:

- $\vec{Y}_1, \vec{Y}_2 — \vec{Y}_p$, jatorrian zentratua dagoen, elipsoide baten ardatz ortogonalak dira, garrantzi beherakoiaz ordenaturik edo - $\vec{Y}_1, \vec{Y}_2 — \vec{Y}_p$, jatorrian pasatzen diran, ingurakoi-zuzen sorta ortogonal bat da.

Hots:

<u>Lehenengo osagai nagusien bektore adierazlea:</u> $\vec{Y}_1$ jatorritikan pasatuaz, indibiduoek gehien inguratzen dituen zuzena izango da.
Hots: $\sum_{J=1}^{N} D^2(J, \vec{Y}) = D_y^2$ minimatzen duena.

Ala baliokide dena:

Jatorritikan pasatuaz, sakabanakuntza edo bariantza gehien jasotzen duen zuzena, norma euklidearra gehien duena.

Hots: $S_y^2 \simeq \|\vec{Y}\|^2$ maximatzen duena.

<u>Bigarren osagai nagusien bektore adierazlea:</u> $\vec{Y}_2$ jatorritikan pasatuaz, eta $\vec{Y}_1$ -ri ortogonala izanik, indibiduoek gehien inguratzen dituen zuzena izango da, edo ortogonaletatik norma euklidear gehien - duena edo sakabanakuntza (bariantza) gehien jasotzen duena.

<u>P-garrengo osagai nagusiaren bektore adierazlea:</u> $\vec{Y}_p$ jatorritikan pasatuaz, eta $\vec{Y}_1\ \vec{Y}_2$ — $\vec{Y}_{p-1}$-ri ortogonala izanik, indibiduoek gehien inguratzen dituen zuzena izango da, edo ortogonaletik norma euklidear ....

<u>OSAGARRIAK</u>

a) Zergatikan jatorritikan pasa arazten ditugu $\vec{Y}_1\ \vec{Y}_2$ — $\vec{Y}_p$?
Honela metrika S, ta S' -n berdina delako, euklidearra geure kasuan.

b) Zergatikan $\vec{Y}_1 \perp \vec{Y}_2 \perp$ —$\perp \vec{Y}_p$? Honela zuzen batek hartzen ez duen sakabanakuntza, besteek jasotzen duelako.

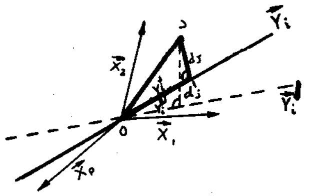
$$\|\vec{J}\|^2 = (Y_1^j)^2 + (Y_2^j)^2 + \dots + (Y_p^j)^2 = \sum_{J=1}^{N} (Y_i^j)^2$$ (Pitagorasen Teoremataz)

c) <u>Euklidear espazio</u> **batetan, ta** $\vec{Y}_1$ <u>jatorritik pasatzen</u> bada, baliokideak dira:

(a) $D_{Y_i}^2 = \sum_{J=1}^{N} d^2 (J,\vec{Y}_i)$ minimizatu

edo

(a') $\|\vec{Y}_i\|^2 = N \cdot S_{y_i}^2$ maximalizatu



Frogaketa:

(Pitagorasen teoremaz)

$$\|J\|^2 = d_j^2 + (Y_i^j)^2 = d_j'^2 + (Y_i'^j)^2 \quad \forall J$$

orduan edozein jatorritikan pasatzen diren $\vec{Y}_i$, $\vec{Y'}_i$ bi zuzenentzat; $d_j < d'_j \Rightarrow Y_i^j > Y'^j_i$

eta baliokidea da d_j minimizatu ala Y_i^j maximalizatu.

J-n gehiketa eginik:

$$\sum \|J\|^2 = \sum_j d^2(J,\vec{Y}_i) + \sum_j (Y_i^j)^2 = \sum_j d^2(J,\vec{Y}_i) + \|\vec{Y}_i\|^2 = \text{Kte}$$

orduan jatorritikan pasatzen diren $\vec{Y}_i$ zuzenentzat, baliokidea da:

$$\sum_j d^2(J, \vec{Y}_i) = D^2_{\vec{y}_i} \quad \text{minimazatu}$$

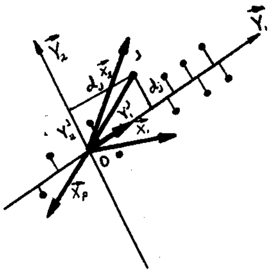
ala

$$\sum_j (Y_i^j)^2 = \|\vec{Y}_i\|^2 = N\, S^2_{y_i} \quad \text{maximalizatu}$$

j.n.g.b.

eta $\|\vec{Y}_i\|^2$ maximalizatzea, $\|\vec{Y}_i\|$ maximalizatzea da.

III. Ikus dezagun <u>Osagai nagusiek aurkitzeko metodoa</u>:



Planteamenduan esan dugunez, $\vec{Y}_1$ lehenengo osagai nagusiaren adierazleak, ondoko bi propietate bete behar ditu:

1.- Jatorritikan pasa:

$\vec{Y}_1 = L_{11}\vec{X}_1 + L_{12}\vec{X}_2 + \ldots + L_{1p}\vec{X}_p$ izango da

non: $=(L_{11}\, L_{12} - L_{1p})$, $\vec{Y}_1$-n bertsore direktzionala da

$=\sum_{i=1}^{p} i_1^{\,2}$; $1in = \cos\varphi(\vec{Y}_1\vec{X}_k) = r(Y_1X_k)$

2.- $\sum_j d^2(J_1\vec{Y}_1)$ mínimo da, edo

$S^2_{y_1} = \frac{1}{N}\sum_j (Y_1{}^j)^2 = \frac{1}{N}\sum_{j=1}^{N}\langle\vec{J}\cdot\vec{L}_1\rangle^2$ maximo da

Lehen pausoan aurkitu nahi dugu:

(1)

$\vec{L}$ $\vec{Y}_1$-n bertsore direktzionala ?
non: $s^2_{y_1} = \frac{1}{N}\sum_j \langle \vec{J}\cdot\vec{L}_1\rangle^2 \overset{*}{=} L_1 \cdot k \cdot L_1^T$ maximo da

$$S^2_{y_1} = \frac{1}{N}\sum_{j=1}^{N}\langle \vec{J}\cdot\vec{L}_1\rangle^2 = \frac{1}{N}\sum_{j=1}^{N}(L_{11}X_1^j + L_{12}X_2^j + \quad + L_{1p}X_p^j)^2 =$$

$$= \frac{1}{N}\sum_{j=1}^{N}\left(\sum_{k=1}^{p}\sum_{i=1}^{d} L_{1i}L_{1k}\cdot X_i^j\cdot X_k^j\right) = \sum_{k=1}^{p}\sum_{i=1}^{d} L_{1i}L_{1k}\frac{1}{N}\sum_{j=1}^{N} X_i^j\cdot X_k^j =$$

$$= \sum_{k=1}^{p}\sum_{i=1}^{p} L_{1i}L_{1k}\frac{1}{N}\langle\vec{X}_i\vec{X}_k\rangle = (L_{11} \text{---} L_{1p})\begin{pmatrix} K \end{pmatrix}\begin{pmatrix} L_{11}\\ |\\ L_{1p}\end{pmatrix} = L_1\cdot K\cdot L_1^T$$

* Metrika euklidearrean gabiltzelako.

Lagrange-ren funtzio definitzen badugu:

$$F = \frac{1}{N}\sum_{j=1}^{N} \langle \vec{J}.\vec{L}\rangle^{2} - \lambda(\|\vec{L}\|^{2} - 1) = \frac{1}{N}\sum_{j}(X_1^{j}L_1 + \ldots + X_p^{j}L_p)^2 - \lambda(\sum_{i=1}^{p} L_i^2 - 1)$$

eta F deribatuaz $L_1 L_2 \ldots L_p$ ezezagunarekiko, ta zerora berdinduaz; (p + 1) ekuazioko sistema batekin aurkitzen gera:

$$S_{(1)}: \begin{cases} \dfrac{\partial F}{\partial L_i} = \dfrac{1}{N}\sum_{r=1}^{N} 2\,(X_1^{J} + X_2^{J}L_2 + \ldots X_p^{J}L_p)X_i^{J} - 2\lambda L_i = 0 & \forall i = 1-p \\ \dfrac{\partial F}{\partial \lambda} = \sum_{i=1}^{p} L_i^2 - 1 = 0 \end{cases}$$

edo baliokide dena:

$$S'_{(1)}: \begin{cases} \sum_{k=1}^{p} \dfrac{1}{N} L_k \sum_{j=1}^{N} x_k^{j} \cdot x_i^{j} \overset{*}{=} \sum_{k=1}^{p} L_k \ \mathrm{kob}(X_k X_i) = \lambda L_i & \forall i=1,\ldots p \\ \sum_{i=1}^{p} L_i^{2} = 1 \end{cases}$$

eta matritzialki idatziaz:

$$S''_{(1)}: \begin{cases} (L_1 L_2 \ldots L_p)\begin{pmatrix} \ldots k(X_1X_i)\ldots \\ k(X_2X_i) \\ \vdots \\ k(X_pX_i) \end{pmatrix} = \lambda\,(L_1 L_2 \ldots L_p) \qquad \boxed{(L_1 L_2 \ldots L_p)(K - I)=0} \\ (L_1 L_2 \ldots L_p)\begin{pmatrix} L_1 \\ L_2 \\ \vdots \\ L_p \end{pmatrix} = 1 \end{cases}$$

$$S''_{(1)}: \begin{cases} (L_1 L_2 \ldots L_p)\begin{pmatrix} \ldots k(X_1X_i)\ldots \\ k(X_2X_i) \\ \vdots \\ k(X_pX_i) \end{pmatrix} = \lambda\,(L_1 \ldots L_i \ldots L_p) \\ S_y^2 = (L_1 L_2 \ldots L_p)\,(k)\begin{pmatrix} L_1 \\ L_2 \\ \vdots \\ L_p \end{pmatrix} = \lambda\,(L_1 \ldots L_i \ldots L_p)\begin{pmatrix} L_1 \\ L_2 \\ \vdots \\ L_p \end{pmatrix} = \lambda \end{cases}$$

erraz ikus daiteke orduan:

geure haserako (1) planteaimendua:

(1)

> Zein dira : $(L_1 \ldots L_p)$? , λ ?
>
> non: $\begin{cases} (L_1 \ldots L_p)(K - I\lambda) = 0 \\ \text{eta } S_y^{2} = \lambda \text{ maximo da.} \end{cases}$

Baina (1) sistema **homogeneo** honek hutsa ez den **soluzio** bat izango du $\forall \lambda \,/\, | K - \lambda I | = 0$; "λ"-ri K-ren autobalio propioa deritza.

Gure problema honetan datza: $K_{p \cdot p}^{x_1 - x_p} = [K(X_i X_k)]$ <u>kobariantza</u> matritzaren edo aldagaiek tipifikatuak **badira** $R_{p.p}^{z_1 z_p} = [r(Z_i Z_k)] = [k(Z_i Z_k)] = K_{p.p}^{z_1 z_p}$ (R = K) <u>korrelazio matritzaren autobalio propioak aurkitu</u>, (gaur egun ordinadoretaz erraz egitan da, ia (400 x 400) matritzaren autobalio propioak kalkulatzen dira ordinadoretaz). K ala R, matritze simetrikoak direnez, dagozkien autobalio propioak desberdinak dira.

Orduan erresolbitu dugu (gehienetan **ordinadoretaz**):

$$\begin{vmatrix} k(X_1 X_p) - \lambda & X(X_1 X_2) & \cdots & K(X_i X_p) \\ K(X_i X_2) & K(X_2 X_2) - \lambda & \cdots & k(X_2 X_p) \\ \vdots & \vdots & \ddots & \vdots \\ K(X_1 X_p) & K(X_1 X_p) & \cdots & K(X_p X_p) - \lambda \end{vmatrix} = 0$$

$$| K - \lambda I | = 0$$

... eta suposa dezagun ekuazio berezi honen soluzioak ordenaturik hoiek direla:

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \ldots \geq \lambda_p \geq 0$$

non: $| K - \lambda_i I | = 0 \qquad \forall \lambda_i$

Orduan hartuko dugu:

<u>Lehenengo osagai nagusiaren bertsore direktzionaletaz</u>: $\vec{L}_1$ / $\vec{L}_1 = (L_{11} L_{12} \ldots L_{1p})$, kobariantza matritzaren λ_1 autobalio propio haundienari dagokion bektore propioa, hots: $(L_{11} L_{12} \ldots L_{1p})(K - \lambda_1 I)=0$ betetzen duena.

a) $\vec{Y}_1 = L_{11}\vec{X}_1 + L_{12}\vec{X}_2 + \ldots + L_{1p}\vec{X}_p$ lehenengo osagaiaren bektore adierazlea izango da.

b) "J" indibiduoak, Y_1 osagai nagusian hartzen duen balioa:

$$Y_1 = L_{11}X_1^j + L_{12}X_2^j + \ldots + L_{1p}X_p^j \qquad \forall J = 1 \longrightarrow N$$

c) Lehenengo osagai nagusiek jasotzen duen indibiduoen sakabanakuntza edo bariantza:

$$S^2_{y_1} = \frac{1}{N} \sum_j (Y_1^j)^2 = \lambda_1$$

Bigarren pauso batetan, hartuko dugu:

<u>Bigarren osagai nagusiaren bertsore direktzionaletaz</u>: $\vec{L}_2$ / $_2 = (L_{21} L_{22} \ldots L_{2p})$, kobariantza matritzaren λ_2 bigarren autobalio propio haundienari dagokion bektore propioa, hots: $(L_{21} L_{22} \ldots L_{2p})(K-\lambda_2 I)= 0$ betetzen duena, zergatikan $\vec{L}_2 \perp \vec{L}_1$ den?

Algebraz badakigu,

Matritze bat simetrikoa bada (geure kasuan K ala R simetrikoak dira), orduan matritzaren balio propio desberdinari dagozkien bektore propioak beren artean ortogonalak dira.

a) $\vec{Y}_2 = L_{21}\vec{X}_1 + L_{22}\vec{X}_2 + \ldots + L_{2p}\vec{X}_p$ bigarren osagaiaren bektore adierazlea izango da. $\vec{Y}_2 \perp \vec{Y}_1$.

b) "J" indibiduoak Y_2 osagai nagusiaren hartzen duen balioa:

$$Y_2^j = L_{21} X_1^j + L_{22}X_2^j + \ldots + L_{2p}X_p^j \qquad \forall J=1_1 \longrightarrow N$$

C)Bigarren osagai nagusiak jasotzen duen indibiduoen sakabanakuntza edo bariantza:

$$S^2_{y_2} = \frac{1}{N} \sum_j (Y_2^j)^2 = \lambda_2$$

r-garren osagai nagusiaren bertsore direktzionaletaz: $\vec{L}_r$ / $\vec{L}_r = (L_{r1}\, L_{r2} \ldots L_{rp})$, kobariantza matritzaren, λ_r, r-garren autobalio propio haundienari dagokion bektore propioa hartuko dugu.

Hots: $(L_{r1}\, L_{r2} \ldots L_{rp})(K - \lambda_r I) = 0$ betetzen duena

Eta * -taz $\vec{L}_1 \perp \vec{L}_2 \perp \ldots\ldots \perp \vec{L}_r$ izango da.

a) $\vec{Y}_r = L_{r1}\vec{X}_1 + L_{r2}\vec{X}_2 + \ldots + L_{rp}\vec{X}_p$, r-garren osagaiaren bektore adierazlea izango da. $\vec{Y}_1 \perp \vec{Y}_2 \perp \ldots\ldots \perp \vec{Y}_r$

b) "J" indibiduoak Y_r osagai nagusiaren hartzen duen balioa:

$$Y_r^j = L_{r1}X_1^j + L_{r2}X_2^j + \ldots\ldots + L_{rp}X_p^j \qquad \forall\, J = 1, \text{———} N$$

c) r-garren osagai nagusiak jasotzen duen sakabanakuntza edo bariantza:

$$S_{y_r}^2 = \frac{1}{N} \sum_j (Y_r^j)^2 = \lambda_r$$

IV. Osagai nagusien analisian ezaugarri berezi bat zera da:

"Indibiduoen sakabanakuntza osoa, p-osagai nagusiek jasotzen dutela"

Hots:
$$\sum_{j=1}^{N} \| J \|^2 = S_{x_1}^2 + S_{x_2}^2 + \ldots\ldots + S_{x_p}^2 = \lambda_1 + \lambda_2 + \ldots\ldots + \lambda_p$$

Orduan lehenengo r-osagai nagusiek, $Y_1 Y_2$ ——— Y_r, sakabanakuntzaren

$$\frac{\lambda_1 + \lambda_2 + \ldots + \lambda_r}{\sum_{i=1}^{p} \lambda_i} \;\%$$ jasotzen dute.

V. <u>Zer esanahi emango diegu osagai nagusi bakoitzari, azterketan?</u>

"Y_i osagai nagusiaren esnahiak, ekarpen gehiena ematen dien aldagaien esanahiarekin zerikusi handia izango du".

*Gogora dezagun:

$\vec{Y_i} = L_{i1}\vec{X_1} + L_{i2}\vec{X_2} + \dots + L_{ip}\vec{X_p} = r(Y_iX_1)\,\vec{X_1} + r(Y_iX_2)\vec{X_2} + \dots = r(Y_iX_p)\vec{X_p}$ dela.

Adibidez:

Baldin bada: $\vec{Y_i} = 0{,}19\ \vec{X_1} + \underline{0{,}98}\ \vec{X_2} + 0{,}36\ \vec{X_3} + \dots\ \underline{-\ 0{,}85}\ \vec{X_p}$

Orduan:

Y_i -n ekarpen gehiena duten aldagaiak X_2, X_p dira, batek ekarpen positiboa eta besteek negatiboa.

4. "Pentsa dezagun Euskal Herriko egoerataz ikerketa bat egin nahi dugula":

- Aurkezpen bezala 140 herri aukeratu ditugu.

Herri hoiek errepresentagarri izan daitezen, metodo arrazional bat jarraitu behar dugu aukeratzekoan (aurkezpen teoría oso lagungarri zaigu urrats honetan).

- Aurkezpeneko herri bakoitzean 40 test zenbatugarri egin ditugu, eta suposa dezagun balioak ondoko taulan azaltzen direnak direla:

	1	2	3		J		140
X_1	20	500	30		X_1^j		80
X_2	5	30	5		X_2^j		4
X_{40}	90	50	70		X_{40}^j		25

X_1 = telefonoen kopurua / mila biztanlekiko

X_2 = bi liburu baino gehiago urtean zeinek irakurtzen dituen / / ehun biztanlekiko.

X_3 = renta per kapita

⋮

X_{40} = Ehunetik zenbat diren etxearen jabe.

Lehen urratsean: taula tipifikatu egingo dugu, testak unitate desberdinetan neurtuak baitira.

	1	2	3	J	140
Z_1				Z_1^j	
Z_2				Z_2^j	
$\vert$				$\vert$	
Z_{40}				Z_{40}^j	

$(Z_i^j > 0$ ala $Z_i^j < 0$ ala $Z_i^j = 0)$

Bigarren urratsean: korrelazio matritza kalkulatuko dugu: Suposa desagun honela dela:

$$R_{40x4o}^{Z_1-Z_{40}} = \begin{bmatrix} 1 & 0,8 & -0,7 & \cdots & 0,3 \\ 0,8 & 1 & 0,5 & \cdots & 0,8 \\ -0,7 & 0,5 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0,3 & o,8 & 0 & \cdots & 1 \end{bmatrix} = K_{40x40}^{Z_1-Z_{40}}.$$

(Obserba genezake korrelazio handia dagoela aldagaien artean).

Hirugarren urrtsean; ordinadoretaz, korrelazio matritzaren auto-balio propioak kalkula genezake.

Suposa dezagun hoiek direla:

- $\lambda_1 = 3,45 > \lambda_2 = 2,15 > \lambda_3 = o,40 > \ldots\ldots\ldots\ldots > \lambda_{40} = 0,001$

... eta $\sum_{i=1}^{40} \lambda_i = 3,45 + 2,15 + \ldots\ldots + o,001 = 7,1$ baldin bada:

- lehengo bi osagai nagusiek sakabanakuntzaren %80 jasotzen dute. (Hots: $\frac{3,45 + 2,15}{7,1} = 0,8$)

* Hau ez da oso erreala, praktikan sakabanakuntzaren %80 jasotzeko bi osagai nagusi bainan gehiago kontuan hartu behar dira, bainan exemplu honetan honela suposatuko dugu. (errextako).

Laugarren urratsean: ordinadoretaz, korrelazio matritzaren bi bektore propio kalkulatuko ditugu, $\lambda_1 = 3{,}45$, $\lambda_2 = 2{,}15$ balio propioeri dagozkienak.

Orduan lehen bi osagai nagusien bektore adierazleak ezagunak izango zaizkigu.

Hots:

1. $(L_1\ L_2 \ldots L_{40})(R - 3{,}45\ I) = 0$ sistema homogeneoa askatuaz:

suposa dezagun: $_1 = (\ _{11}\ _{12} \cdots\ _{140}) = (0{,}60,\ 0{,}15,\ 0{,}98, \ldots 0{,}69)$ dela

orduan: $\vec{Y}_i = 0{,}60\ \vec{Z}_1 + 0{,}15\ \vec{Z}_2 + 0{,}98\ \vec{Z}_3 + \ldots + 0{,}69\ \vec{Z}_{40}$

2. $(L_1\ L_2 \cdots L_{40})(R - 2{,}15\ I) = 0$ sistema homogeneoa askatuaz:

suposa dezagun: $_2 = (L_{21}\ L_{22} \ldots L_{240}) = (0{,}25,\ 0{,}70,\ 0{,}10, -0{,}80 \ldots 0{,}10)$

orduan: $\vec{Y}_2 = 0{,}25\ \vec{Z}_1 + 0{,}70\ \vec{Z}_2 + 0{,}10\ \vec{Z}_3 - 0{,}80\ \vec{Z}_4 + \ldots = 0{,}10\ \vec{Z}_{40}$

Bostgarren urratsean: esanahi bat ematen saiatuko gera osagai nagusieri. Urrats honetan, noski espezialisten laguntza behar beharrezkoa zaigu (soziologo, ekonomilari, sendagile...)

Exenplu honetan: $\vec{Y}_1 = \underline{0{,}60}\ \vec{Z}_1 + 0{,}15\ \vec{Z}_2 + \underline{0{,}98}\ \vec{Z}_3 + \ldots + \underline{0{,}69}\ \vec{Z}_{40}$ bada.

Y_1 -n zein aldagaiek dute ekarpen gehiena?

ordenaturik Z_3, Z_{40}, Z_1 ...

Orduan: Y_1 osagai nagusiak esan genezake herrien maila ekonomikoarekin zerikusi handia zuela.

Exenplu honetan: $\vec{Y}_2 = 0{,}25\ \vec{Z}_1 + \underline{0{,}70}\ \vec{Z}_2 + 0{,}10\ \vec{Z}_3 - 0{,}80\ \vec{Z}_4 + \ldots + 0{,}10\ \vec{Z}_{40}$ bada

Y_2 -n zein aldagaiek dute ekarpen gehiena?

ordenaturik: Z_4, Z_2, ...

orduan: Y_2 osagai nagusiak esan genezake herrien maila kulturalarekin zerikusi handia duela.

Seigarren urratsean: gure aurkezpeneko 140 herriak, Y_1, Y_2 osagaietan hartzen dituzten balioak kalkula genezake.

Hots:

$$Y_1^j = 0{,}60\ Z_1^j + 0{,}15\ Z_2^j + \ldots\ldots \quad + 0{,}69\ Z_{40}^j$$

$$Y_2^j = 0{,}25\ Z_1^j + 0{,}70\ Z_2^j + \ldots\ldots \quad + 0{,}10\ Z_{40}^j$$

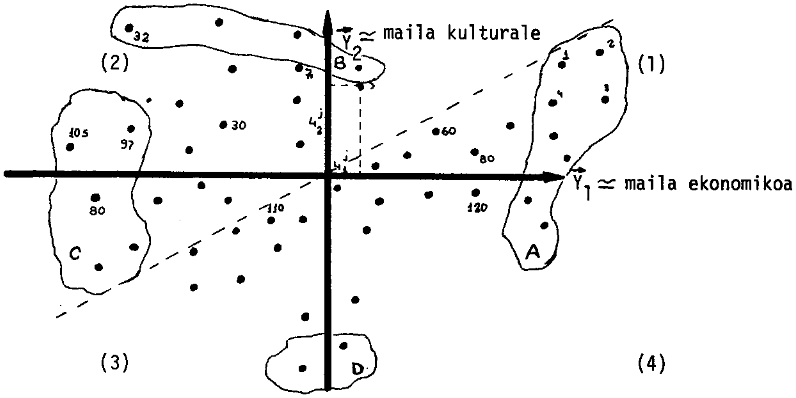
$$\forall J : (Y_1^j,\ Y_2^j)$$

Zazpigarren urratsean; taula batez eta geometrikoki azalduko ditugu;

(A)

	1,	2,	J,	100	
Y_1	Y_1^1	Y_1^2	Y_1^j	Y_1^{100}	$\bar{Y}_1=0;\ S^2_{y_1} = \lambda_1 = 3{,}45$
Y_2	Y_2^1	Y_2^2	Y_2^j	Y_2^{100}	$\bar{Y}_2=0;\ S^2_{y_2} = \lambda_2 = 2{,}15$

(B)



Azterketaren ondorio nabarienak:

a) 1. arloan kokatu diren herriak maila ekonomiko eta maila kulturala mediatikan gora dute.

2. arloan kokatu diren herriak, maila ekonomikoa mediatikan bera dute, eta maila kulturala mediatikan gora.

3. arloan kokatu diren herriak maila ekonomikoa eta maila kulturala mediatikan behera dute.

4. arloan kokatu diren herriak maila ekonomikoa mediatikan gora dute eta maila kulturala mediatikan behera.

b) Maila ekonomikoa handiena duten herriak A multzokoak dira.

c) Maila kulturala handiena duten herriak B multzokoak dira.

d) Maila ekonomiko txikiena duten herriak C multzokoak dira.

e) Maila kultural txikiena duten herriak D multzokoak dira.

f) Maila ekonomikoa kulturala baino handiagoa da Euskal Herrian zeren eta ia herri guztiak $Y_1 > Y_2$ arloan kokatzen dira.

g) Nola explika "82" herriaren egoera?

h) Hurbil kokatzen diren herriak, egoera berdintsukoak dira ...

... Eta honela jarrai genezake. ...