ANÀLISI EN COMPONENTS PRINCIPALS. EXEMPLE D'APLICACIÓ

Quaderns d'enginyeria
2(1980)1 p.93-111

per Oriol Presas i Moises Valls *

RESUM

L'anàlisi de dades multivariants constitueix un grup de diferents mètodes entre els quals hi figura el de les components principals, motiu d'aquest article. Amb aquestes tècniques estadístiques tenim una forma de descriure i interpretar el que s'amaga darrera una matriu de dades de dimensions més o menys considerables. Aquestos mètodes abarquen un extens camp d'aplicació, com per exemple, la indústria, marketing, biologia, sociologia...

Amb les tècniques d'anàlisi factorial es disposa de mitjans de control i de reducció de dades. De manera sencilla podem reconstruir aproximadament la taula inicial a partir de les dimensions retingudes, les quals ens permeteran entendre millor el fenómen estudiat. L'anàlisi en components principals pretèn extreure la informació continguda en la taula de dades i suministrar una representació gràfica més fàcil d'interpretar.

A efectes de divulgació de la metodologia, hem escollit com a cas pràctic una taula de característiques socio-econòmiques preses dels països de la O.C.D.E.

SUMMARY

Multivariate data analysis is a set of different methodes. One of those is the so called Principal Components, object of this paper. With this technique we can recover what lays behind a data matrix of high dimension. The scope of applications of Principal Components analysis is very wide: marketing, biology, sociology, industry.

With factorial analysis techniques we have a device for control and data reduction. In a simple way, it unable us to reconstruct the original high dimension data matrix from reduce dimension one. This permits us to improve our knowledge of the object under study.

On the other hand, Principal Components analysis give us a useful graphical representation of the raw data. In our exercise, we explain how to use this technique using socio-economic data from several O.C.D.E. Countries.

*Becaris de la Catedra d'Estadística

1. INTRODUCCIO

1.1. Classificació de l'anàlisi factorial

L'anàlisi factorial abarca un extens camp de diferentes tecniques. Anem a veure com es classifiquen segons la taula de dades i l'objectiu que persegueixen.

A) En els casos en que la matriu inicial sigui del tipus individus x variables se'ns poden presentar els casos seguents:

a.1) Els individus formen un grup homogeni. Si les variables formen un grup homogeni hi han dugues tècniques: anàlisi en components principals i anàlisi factorial de correlacions. En la primera, l'objectiu que es persegueix és reduir el nombre de variables de manera que expliquin el màxim de la variabilitat total, mentres que en la segona, les noves variables han d'explicar el màxim de les correlacions entre les antigues.

Si les variables es divideixen en varis grups s'utilitza l'anàlisi de correlacions canóniques per a explicar les relacions existents entre aquests grups.

a.2) Els individus procedeixen de varies poblacions.

Si volem estudiar la significació de les diferències entre els grups d'individus s'utilitza l'anàlisi factorial discriminant. Si en canvi l'objectiu és trobar criteris per a classificar nous individus, recurrim a l'anàlisi discriminant.

B) Si la matriu inicial no és la del cas A), podem utilitzar les següents tècniques:

b.1) La matriu inicial és de diferències entre individus.

Quan es pretén definir grups homogenis d'individus, es fa servir l'anàlisi cluster. Si volem representar les relacions entre individus en un espai de poques dimensions, utilitzem els escalogrames multidimensionals.

b.2) Els elements de la matriu de partida són les frecuències que ens relacionen dos conjunts.

En aquest cas utilitzem l'anàlisi factorial de correspondències per a trobar les relacions existents dintre els dos conjunts i entre ells mateixos.

1.2. Descripció del problema.

Es disposa d'una taula amb n files i p columnes, on les files representen individus u observacions i les columnes variables o característiques. Els n individus poden ser considerats com n punts d'un espai R^p. La proximitat entre dos d'ells en aquest espai ens mesura

la semblança entre els dos individus.

Ara tenim n x p elements, pero existiran dependències entre les p variables. Es tracta de reduir de manera sistemàtica les p variables en un nombre inferior de valors tenint en compte els lligams entre elles i sense perdre excessiva informació.

2. BASES TEORIQUES

2.1 Interpretació geométrica

Designarem per R la matriu de partida d'ordre n x p on r_i^j sera la i-èsima observació per la variable j. Volem reduir la dimensió de l'espai R^P i despres fer una descripció del fenomen que s'ajusti als resultats obtinguts.

El primer pas consisteix en colocar el centre dels nous eixos en el centre de gravetat del núvol de punts de partida. Això es consegueix mitjançant el canvi:

$$x_i^j = r_i^j - \bar{r}^j$$

On r_i^j és la i-èsima observació per la variable "j" i $\bar{r}_j$ és la mitja de les mesures per la variable "j".

Degut a la diferència d'escales en les variables, haurem de realitzar encara una altre transformació que pot consistir en dividir per la desviació tipus de la variable j o per la diferència entre el valor màxim i mínim d'aquesta variable. Normalment es fa lo primer amb lo qual el terme general de la matriu de partida transformada serà:

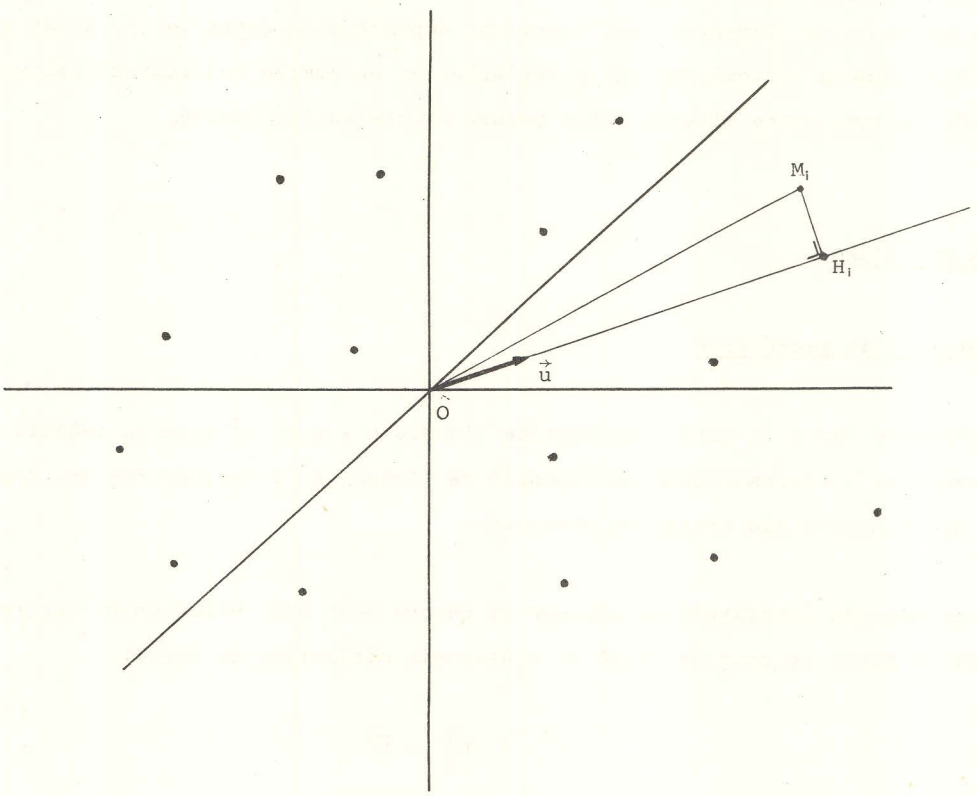
$$x_i^j = \frac{r_i^j - \bar{r}^j}{s_j \sqrt{n}}$$

Amb aquest canvi conseguim ajustar un subespai que ens expliqui més aproximadament el fenómen. La raó que apareixi $\sqrt{n}$ és la posterior formació de coeficients de correlació.

2.2. Búsqueda del subespai de R^P

Començarem per buscar un subespai d' una sola dimensió que realitzi el millor ajustament possible del núvol de punts. Això sera una recta que evidentment passarà pel centre de

gravetat d'aquest núvol,



Sigui u el vector unitari en aquesta direcció. Les n projeccions OH_i dels punts M_i sobre aquesta recta seran les n components del vector Xu on cada una de les línees de la matriu X és un vector-individu de R^P.

Per a fer l'ajustament, utilitzem el criteri clàssic dels "mínims cuadrats", lo qual es tradueix en fer mínima la suma $\sum_{i=1}^{n} (M_iH_i)^2$. Si apliquem el teorema de Pitágoras a cada un dels n triangles rectangles del tipus H_iOM_i trobem la relació:

$$\Sigma(M_iH_i)^2 = \Sigma(OM_i)^2 - \Sigma(OH_i)^2$$

com $\Sigma(OM_i)^2$ és una constant característica del núvol, haurem de fer maxima la suma $\Sigma(OH_i)$ que es pot escriure en funció de X i u de la forma:

$$\Sigma(OH_i)^2 = (Xu)'Xu = u'X'Xu \qquad \{1\}$$

Per a trobar u_1, hem de buscar el màxim de u'X'Xu amb la condició que u'u = 1.

La matriu X'X és exactament la matriu de correlacions entre variables, cuadrada, simétrica, per tant amb tots els valors propis reals no negatius i els vectors propis ortogonals.

El subespai de dugues dimensions que millor s'ajusta al núvol ha de contindre forçosament el subespai engendrat per $\vec{u}_1$. Podem buscar donç $\vec{u}_2$, segon vector de la base d'aquest subespai, que sigui ortogonal a u_1 i que faci màxim $u_2'X'Xu_2$.

De forma anàloga buscarem les següents direccions que ens donguin el millor subespai en el sentit dels mínims cuadrats.

Per a trobar la solució a {1} s'utilitzen els multiplicadors de Lagrange, que porten a l'expressió:

$$2.\ X'X\vec{u}_1 - 2\lambda\ \vec{u}_1 = 0$$

Per tant:

$$X'X\vec{u}_1 = \lambda\vec{u}_1$$

On λ és un multiplicador de Lagrange.

$\vec{u}_1$ és donç el vector propi de la matriu X'X corresponent al valor propi més gran λ_1. A continuació, hem de trobar el vector $\vec{u}_2$, ortogonal a $\vec{u}_1$ tal que $\vec{u}_2'X'X\vec{u}_2$ sigui màxim. El resultat obtingut pels multiplicadors de Lagrange, és altre cop:

$$X'X\vec{u}_2 = \lambda_2\ \vec{u}_2$$

On $\vec{u}_2$ és el vector propi corresponent al segon valor propi més gran λ_2, i per la α-èsima direcció principal es compleix la relació:

$$X'X\vec{u}_\alpha = \lambda_\alpha\ \vec{u}_\alpha$$

Així donç, la búsqueda de la base del subespai es redueix a trobar els vectors propis corresponents als valors propis més grans de la matriu de correlacions X'X.

2.3. Relació dels ajustaments entre R^p i R^n

Considerem are l'espai R^n on els p punts són punts-variables. Com en el cas R^p, les projeccions dels punts-variables sobre els eixos principals seran les components del vector $X'\vec{v}_\alpha$ on $\vec{v}_\alpha$ és el vector unitari corresponent a la direcció α. Per tant, per a trobar els vectors i valors propis hem de maximitzar:

$$(X'v)'\ X'v = v'XX'v$$

amb la condició de que $v'v = 1$

En R^p teniem $X'X\vec{u}_1 = \lambda_1\vec{u}_1$ {2}

En R^n tenim $XX'\vec{v}_1 = \mu_1\vec{v}_1$ {3}

On λ_1 i μ_1 són els valors propis més grans de les matrius X'X i XX' respectivament. Si multipliquem {2} per X, tenim $(XX')X\vec{u}_1 = \lambda_1 X\vec{u}_2$. El vector $X\vec{u}_1$ és el vector propi corresponent al valor propi λ_1 de la matriu XX'. Per tant $\lambda_1 < \mu_1$. Si multipliquem {3} per X' tenim $(X'X)X'\vec{v}_1 = \mu_1 X'\vec{v}_1$ on $X'\vec{v}_1$ és el vector propi corresponent al valor propi μ_1 de la matriu X'X. Per tant $\mu_1 < \lambda_1$, amb lo qual arribem a la conclusió que $\lambda_1 = \mu_1$. En general, $\lambda_\alpha = \mu_\alpha$.

El vector Xu_α té per norma λ_α, ja que $u'\alpha\ X'Xu_\alpha = \lambda_\alpha$. Per tant, de les relacions $(XX')Xu_\alpha = \lambda_\alpha(Xu\ )$ i $XX'v_\alpha = \lambda_\alpha v_\alpha$ deduïm que $v_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} Xu_\alpha$ per $\lambda_\alpha \neq 0$. De la mateixa forma trobem que $u_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} X'v_\alpha$. Amb aquestes dugues relacions, ajustant en R^p podem trobar directament l'ajust en R^n i viceversa.

Direm que les coordenades dels punts en els eixos factorials són les components principals.

La α-èsima component principal Xu_α serà donç les projeccions dels n punts sobre el α-èsim eix principal. Aquestes seran les n components del vector Xu_α. Com que $Xu_\alpha = \sqrt{\lambda_\alpha} \cdot v_\alpha$, les coordenades dels punts en un eix factorial en R^p són donç proporcionals a les components de l'eix principal en R^n corresponent al mateix valor propi.

2.4. Correlacions

Com hem dit abans, la matriu X'X és la matriu de correlacions entre variables, ja que el terme general és:

$$c_{jj'} = \frac{1/n \sum_i (r_{ij} - \bar{r}_j)(r_{ij'} - \bar{r}_{j'})}{S_j S_{j'}}$$

$c_{jj'}$ és el coeficient de correlació entre les variables j i j'. Es pot comprovar que la distància entre dos punts variables j i j' en R^n és:

$$d^2(j,j') = 2 - 2\,c_{jj'}$$

per tant, si $c_{jj'} \simeq 1$ (fortament correlacionades), la distància serà molt petita; si $c_{jj'} \simeq -1$ (inversament correlacionades), la distància serà molt gran i si $c_{jj'} \simeq 0$ (no estan correlacionades) estaran a una distància mitjana.

Els p coeficients de correlació existents entre les coordenades dels punts sobre un eix i les p variables són les p components del vector $X'v_\alpha = u_\alpha\sqrt{\lambda_\alpha}$ ja que l'abcisa d'un punt-variable sobre un eix és el coeficient de correlació entre aquesta variable i la variable artificial.

2.5. Percentatge de variança explicada.

Tenim la relació $Xu_\alpha = \sqrt{\lambda_\alpha}\, v_\alpha$ pels α-èsims eixos de R^P i R^n. Si multipliquem els dos membres per $u'\alpha$ i sumem, obtenim:

$$X \sum_{\alpha=1}^{P} u_\alpha \cdot u'_\alpha = \sum_{\alpha-1}^{P} \sqrt{\lambda_\alpha} \cdot v_\alpha \cdot u'_\alpha$$

Però $\sum_{\alpha=1}^{P} u_\alpha \cdot u'_\alpha = I$ ja que els vectors u_α són unitaris i ortogonals. Ens queda:

$$X = \sum_{\alpha=1}^{P} \sqrt{\lambda_\alpha} \cdot v_\alpha \cdot u'_\alpha \qquad \{4\}$$

Per tant, tenim una manera de reconstruir la taula X a partir dels λ_α, u_α i v_α associats.

Si despreciem els (p-q) valors propis més petits, tenim:

$$X \simeq X^* = \sum_{\alpha=1}^{q} \sqrt{\lambda_\alpha} \cdot v_\alpha \cdot u'_\alpha \qquad \{5\}$$

Una manera d'avaluar la qualitat de reconstrucció de la taula X és:

$$\tau q = \sum_{i,j} X_i^{j\,2} / \sum_{i,j} X_i^{j\,2} = \operatorname{tr} X^{*\prime}X^* / \operatorname{tr} X'X$$

on τq és la tasa d'inèrcia o porcentatge de variança relatiu als q primers factors.

Sustituint X i X* pels seus valors a {4} i {5}, s'obté:

$$\tau q = \sum_{\alpha \leq q} \lambda_\alpha \Big/ \sum_{\alpha=1}^{P} \lambda_\alpha$$

El coeficient τq és inferior o igual a 1, i ens diu el percentatge d'informació retinguda.

3. EXEMPLE

3.1. Descripció de les taules de l'anàlisi.

L'exemple que exposem a continuació consta d'una matriu de 18 variables socioeconómiques mesurades en 18 països de l'O.C.D.E. (Veure Bertier).

En les taules I i Ia) hi figuren la llista de variables i d'individus amb els seus códis respectius. La taula II conté les dades del problema. En la taula III a) s'ha trobat la matriu de correlacions i en la III b) les mitges i desviacions tipus de les variables.

Les direccions principals vindran donades pels vectors propis corresponents als valors propis més grans.

En la taula IV a) figuren els valors propis i el percentatge de variança explicada per cada un d'ells. Tal com s'ha explicat, això es troba dividint el valor propi corresponent per la traça de la matriu de correlacions que coincidira sempre amb el nombre de variables, en aquest cas.

	PERCENTATGE DE VARIANÇA EXPLICADA PER CADA FACTOR	PERCENTATGE DE VARIANÇA EXPLICADA ACUMULADA
1er. factor :	$\frac{7.13902}{18} = 39,66123$	39,66123
2on. factor :	$\frac{3.04133}{18} = 16,89628$	56,55751
3er. factor :	$\frac{2.31512}{18} = 12,86177$	69,41928

Es a dir, amb aquestos tres factors principals expliquem casi el 70% de la variabilitat total.

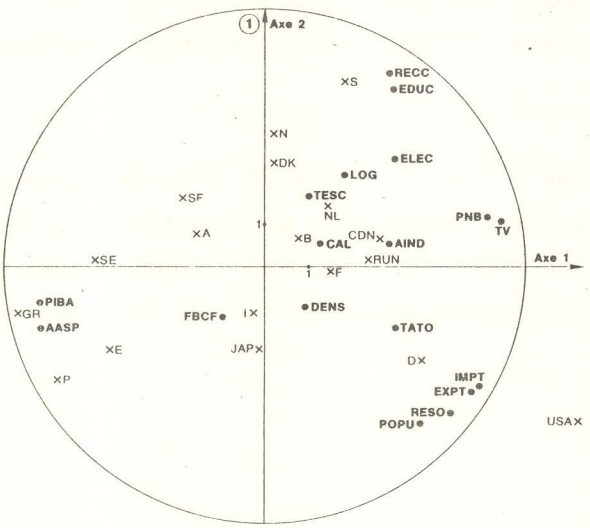
En la taula IV b) hi han les correlacions entre variables i factors que es troben de la forma següent:

Correlació de la variable "i" amb el factor "j" = $\sqrt{\lambda_j}\ u_j^i$; siguent u_j^i la component "i" del vector unitari u_j en la direcció principal "j" i λ_j el valor propi "j". Analitzant les correlacions de les variables en cada eix, veurem quines ens expliquen millor cada direcció principal.

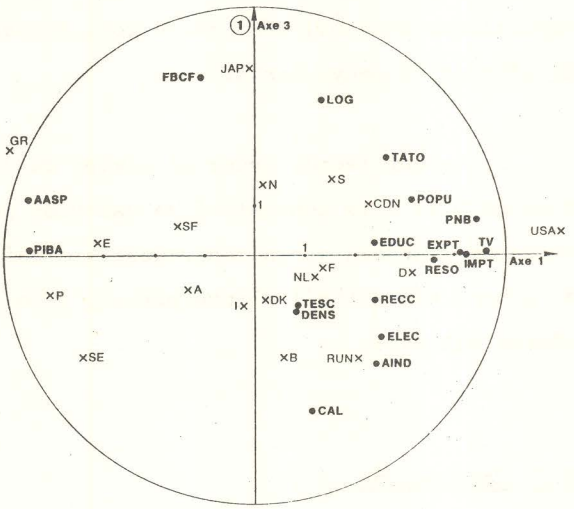
La taula V a) ens dona les components principals, que són les projeccions de cada observació en els eixos principals. Aquestes s'obtenen fent $\bar{v}_\alpha = \bar{\bar{X}}\,\vec{u}_\alpha$, siguent $\vec{u}_\alpha$ el vector unitari en la direcció α i $\bar{v}_\alpha$ la α-esima component principal.

Les columnes de la taula V b) representen els quadrats dels cosinus acumulats dels angles que formen les observacions amb els eixos principals, lo qual ens mesura la bondat de la representació de cada país.

A partir de les taules IV b) i V a) s'obtenen els gràfics següents:



Eixos 1 i 2



Eixos 1 i 3

3.2. Comentaris de l'exemple

Amb tres factors hem conseguit tenir en compte el 69,42% del total de la variança del núvol de punts. Així doncs, hem lograt passar d'un espai R^{18} a un espai R^3 perdent el 30% de la informació. Passem ara a analitzar cada un dels factors.

Per el primer factor s'observa que les correlacions més elevades són amb el PNB (0,893),

Importacions (0,848), Exportacions (0,847), TV (0,934). En contraposició tenim percentatge de PIB en l'agricultura (-0,905) i percentatge d'actius en agricultura i pesca (-0,887). El primer factor contraposa els països més industrialitzats amb els més agrícoles. Països com Estats Units i Alemania s'oposen a països com Grècia, Portugal, Irlanda i Espanya. Per recolzar aquesta explicació veiem en la taula III a) que les correlacions entre les variables explicatives del primer factor són altes, com per exemple, la correlació entre T.V. i PNB val 0.915, entre actius de pesca i agricultura amb percentatge de PIB en agricultura val 0.922 i entre aquesta i TV val -0.806. Es comprova també en la matriu de partida que, per exemple, per USA i Grècia respectivament, el PNB val 4.660 i 950 i el percentatge de PIB en agricultura val 2,90 i 20,30.

El segon factor oposa els ingressos corrents (0.759) i els gastos d'educació (0,713) a la població total (-0,640) i a les reserves oficials (-0,606). La interpretació en aquest cas és més difícil i les correlacions menys próximes a 1. Aquest factor oposa els països nòrdics de tendència socialista (Suècia, Noruega, Dinamarca i Finlàndia) amb els països de capitalisme lliberal (USA sobre tot, Alemania, Japó, Portugal* i Espanya) on la intervenció de l'Estat és menys forta.

El tercer factor oposa el nombre de vivendes (0,739) i la formació bruta de capital fixe (0,714) a la consumició de calories per habitant (-0,734). Es pot observar que països com Japó i Grècia venen contraposats a països com Anglaterra, Bélgica e Irlanda. Aquest eix és encara més difícil d'interpretar, pero d'alguna manera explica el grau de joventut econòmica del país.

3.3. Conclusions

Considerant la proximitat entre països, observem que es formen diferents grups. Per una banda tenim USA aillat. Per altre banda veiem que els països del mercat comú estan bastant agrupats, encara que Alemania cau entre ells i USA. Junt amb aquestos ens apareix Canadà. Un altre agrupament el formen els països Nordics.

Finalment, tenim els països més pobres com un altre grup d'individus.

* Aquestes dades estan preses abans del canvi de règim a Portugal.

TAULA I

DICCIONARI DE LES VARIABLES

1.	POBLACIO TOTAL	POP
2.	DENSITAT PER KM^2	DEN
3.	TASA CREIXEMENT. TOTAL POBLACIO	TAT
4.	PERCENTATGE ACTIUS AGRICULTURA	AAS
5.	TANT PER CENT ACTIUS INDUSTRIA	AIN
6.	PROD. NACIONAL BRUT EN DOLARS PER HAB.	PNB
7.	TANT PER CENT DE PIB EN AGRICULTURA	PIB
8.	FORMACIO BRUTA DE CAP FIXE EN PNB/100	FBC
9.	INGRESSOS CORRENTS EN PNB/100	REC
10.	RESERVES OFICIALS (EN MILL DOL)	RES
11.	TASA DE DESCOMPTE OFICIAL	TES
12.	IMPORTACIONS	INP
13.	EXPORTACIONS	EXP
14.	CALORIES PER HABITANT I DIA	CAL
15.	VIVENDES PER 1.000 HABITANTS	LOG
16.	CONSUM ELECTRIC EN KWH/HAB. ANY	ELE
17.	DESPENSES PUBLIQUES EDUCACIO EN PNB/100	EDU
18.	NOMBRE DE TV. PER 1.000 HABITANTS	TV

NOMBRE DE LES OBSERVACIONS 18

NOMBRE TOTAL DE QUESTIONS 18
NOMBRE TOTAL DE MODALITATS 18

TAULA I a)

1.	ALEMANIA	D
2.	AUSTRIA	A
3.	BELGICA	B
4.	CANADA	CDN
5.	DINAMARCA	DK
6.	ESPANYA	E
7.	USA	USA
8.	FINLANDIA	SF
9.	FRANÇA	F
10.	GRECIA	G
11.	IRLANDA	SE
12.	ITALIA	I
13.	JAPO	JAP
14.	NORUEGA	N
15.	PAISOS BAIXOS	NL
16.	PORTUGAL	P
17.	ANGLATERRA	RUN
18.	SUECIA	S

TAULA II

VARIABLES / PAISOS	1	2	3	4	5	6	7	8	9
D	60848.00	245.00	1.05	9.60	49.10	2520.00	3.60	24.40	37.90
A	7373.00	88.00	.50	19.10	39.90	1690.00	7.00	23.20	37.50
B	9984.00	332.0	.60	5.40	44.80	2352.00	5.40	23.10	35.10
CON	21089.00	2.00	1.85	8.20	32.30	3460.00	5.90	21.70	35.20
DK	4893.00	114.00	.75	11.90	38.50	2860.00	8.90	22.00	37.10
E	32949.00	65.00	.95	30.70	37.10	870.00	15.00	22.00	22.40
USA	203213.00	22.00	1.35	4.60	33.70	4660.00	2.90	16.70	31.50
SF	4706.00	14.00	.70	24.50	34.60	1940.00	14.70	23.80	35.90
F	50325.00	91.00	1.05	15.10	40.60	2770.00	6.00	25.40	38.10
GR	8866.00	67.00	.70	48.20	22.50	950.00	20.30	29.70	26.90
SE	2921.00	42.00	.25	28.40	29.70	1040.00	19.70	19.90	30.70
I	54120.oo	180.00	.85	21.50	43.10	1520.00	11.30	20.50	33.30
JAP	102380.00	277.00	1.05	18.80	35.00	1630.00	8.70	35.20	21.20
N	3051.00	12.00	.80	14.70	36.80	2530.00	6.50	25.30	43.40
NL	12873.00	352.00	1.25	7.50	41.30	2190.00	7.00	25.50	41.90
P	9583.00	105.00	.90	31.50	35.50	600.00	17.70	18.40	24.00
RUN	55643.00	228.00	.65	2.90	46.80	1970.00	3.00	17.30	39.00
S	7969.00	18.00	.70	8.80	40.40	3230.00	5.90	23.60	48.10

(continua)

(continuació Taula II)

VARIABLES / PAISOS	10	11	12	13	14	15	16	17	18
D	10940.00	6.50	24926.00	29052.00	2990.00	8.60	3322.00	3.00	231.00
A	1563.00	5.00	2825.00	2412.00	2990.00	6.60	2647.00	4.40	134.00
B	2406.00	7.00	9984.00	10069.00	3150.00	5.00	2814.00	5.30	184.00
CON	3846.00	6.00	13137.00	13754.00	3160.00	8.20	8199.00	5.70	279.00
DK	384.00	9.00	3800.00	2958.00	3180.00	9.00	2413.00	6.00	244.00
E	1518.00	6.50	4233.00	1900.00	2750.00	6.40	1245.00	2.10	84.00
USA	12306.00	5.75	36052.00	37988.00	3210.00	7.70	7013.00	5.10	392.00
SF	379.00	6.00	2023.00	1985.00	2900.00	7.90	3836.00	6.30	193.00
F	4617.00	7.50	17392.00	15020.00	3160.00	8.20	2407.00	4.80	185.00
GR	290.00	6.50	1594.00	554.00	2910.00	10.10	823.00	2.40	9.00
SE	694.00	7.31	1413.00	891.00	3450.00	4.00	1577.00	4.20	111.00
I	4642.00	5.50	12450.00	11729.00	2940.00	5.10	1810.00	5.80	146.00
JAP	3072.00	6.00	15024.00	15990.00	2460.00	11.90	2734.00	4.50	190.00
N	607.00	4.50	2943.00	2203.00	2910.00	8.80	12976.00	5.80	175.00
NL	2621.00	6.00	10991.00	9965.00	3240.00	9.70	2565.00	6.70	197.00
P	1442.00	3.50	1231.00	823.00	2930.00	4.30	607.00	1.40	29.00
RUN	2469.00	7.00	19956.00	17515.00	3180.00	7.70	3680.00	4.20	263.00
S	506.00	7.00	5899.00	5698.00	2750.00	13.40	6803.00	7.40	288.00

TAULA III a)

MATRIU DE LES CORRELACIONS

	POP	DEN	TAT	AAS	AIN	PNB	PIB	FBC	REC	RES	TES	INP	EXP	CAL	LOG	ELE	EDU	TV
POP	1.00																	
DEN	.06	1.00																
TAT	.42	-.01	1.00															
AAS	-.32	-.32	-.33	1.00														
AIN	.05	.05	-.03	-.68	1.00													
PNB	.49	.49	.52	-.75	.19	1.00												
PIB	-.41	-.41	-.39	.92	-.67	-.76	1.00											
FBC	-.11	-.11	.04	.27	-.22	-.19	.00	1.00										
REC	-.28	-.28	-.08	-.60	.44	.49	-.54	-.19	1.00									
RES	.81	.81	.51	-.43	.31	.55	-.52	-.22	-.05	1.00								
TES	-.06	-.06	-.17	-.22	,11	.24	-.13	.05	.14	-.06	1.00							
INP	.87	.87	.52	-.60	,37	.64	-.68	-.19	.02	.91	.09	1.00						
EXP	.86	.86	.53	-.59	.35	.65	-.67	-.16	.03	.94	.06	.99	1.00					
CAL	-.05	-.05	-.03	-.30	.05	.26	-.10	-.59	.27	.16	.32	.17	.15	1.00				
LOG	.10	.10	.25	-.19	-.06	.38	-.33	.61	.32	-.03	.23	.12	.13	-.47	1.00			
ELE	.12	.12	.33	-.47	.00	.64	-.51	-.05	.56	.13	-.22	.18	.20	-.02	.33	1.00		
EDU	-.05	-.05	.11	-.59	.21	.60	-.47	.04	.71	-.08	.26	.07	.06	.15	.39	.50	1.00	
TV	.55	.55	.43	-.86	.35	.92	-.81	-.25	.50	.54	.29	.70	.70	.22	.37	.55	.62	1.00
	POP	DEN	TAT	AAS	AIN	PNB	PIB	FBC	REC	RES	TES	INP	EXP	CAL	LOG	ELE	EDU	TV

TAULA III b)

BREU DESCRIPCIO DE LES VARIABLES

	VARIABLE	MITJANA	DESVIACIO-TIPUS	MINIM	MAXIM
1	POP	36310,3311	48567.8198	2921,0000	***
2	DEN	125,2222	111,7266	2,0000	352,0000
3	TAT	0,8861	0,3483	0,2500	1,8500
4	AAS	17,3000	11,5364	2,9000	48,2000
5	AIN	37,8722	6,1872	22,5000	49,1000
6	PNB	2154,5555	1001,7437	600,0000	4660,0000
7	PIB	9,4167	5,5097	2,9000	20,3000
8	FBC	23,2056	4,2568	16,7000	35,2000
9	REC	33,9556	7,0692	21,2000	48,1000
10	RES	3016,7775	3345,4805	290,0000	12306,0000
11	TES	6,2533	1,1873	3,5000	9,0000
12	INP	10326,2771	9339,6686	1231,0000	36052,0000
13	EXP	10028,1106	10147,9950	554,0000	37988,0000
14	CAL	3014,4443	222,4174	2460,0000	3450,0000
15	LOG	7,9222	2,4204	4,0000	13,4000
16	ELE	3748,3887	3040,4936	607,0000	12976,0000
17	EDU	4,7278	1,5975	1,4000	7,4000
18	TV	185,2222	91,1102	9,0000	392,0000

TAULA IV a)

EDICIO DELS VALORS-PROPIS

- Suma dels valors-propis actius: 17.99999619

- Histograma dels primers valors propis:

	Valor-propi	Percentatge	Percentatge acumulat
1	7.11110353	39.51	39.51
2	3.05933398	17.00	56.50
3	2.31262597	12.85	69.35
4	1.98176439	11.01	80.36
5	1.28419745	7.13	87.49
6	.67594828	3.76	91.25
7	.43715972	2.43	93.68
8	.38712417	2.15	95.83
9	.28886601	1.60	97.43
10	.20426887	1.13	98.57

- Edició total dels valors-propis de 11 a 18:

.10502886 .06478972 .04496537 .02276048 .01581046 .00376381 .00048722 .00000009

TAULA IV b)

VARIABLES ACTIVES

	FACTOR 1	FACTOR 2	FACTOR 3
POP	-.631	.635	.210
DEN	-.167	.171	-.233
TAT	-.529	.273	.388
AAS	.886	.265	.223
AIN	-.493	-.102	-.447
PNB	-.891	-.195	.130
PIB	.903	.165	.023
FBC	.218	-.087	.722
REC	-.454	-.776	-.187
RES	-.726	.603	-.023
TES	-.168	-.256	-.203
INP	-.849	.494	.002
EXP	-.849	.496	.031
CAL	-.226	-.067	-.732
LOG	-.297	-.366	.741
ELE	-.512	-.441	.328
EDU	-.491	-.175	.065
TV	-.935	-.179	.035

TAULA V a)

OBSERVACIONS

	PES	FACTOR 1	FACTOR 2	FACTOR 3
ALE	1.000	-.734	.444	-.100
AUS	1.000	.318	-.168	-.160
BEL	1.000	-.145	-.125	-.492
CAN	1.000	-.542	-.139	.243
DIN	1.000	-.045	-.486	-.213
ESP	1.000	.717	.398	.060
USA	1.000	-1.451	.718	.096
FIN	1.000	.372	-.335	.122
FRA	1.000	-.272	.139	-.020
GRE	1.000	1.146	.218	.489
IRL	1.000	.796	-.017	-.552
ITA	1.000	.059	.214	-.243
JAP	1.000	.014	.398	.880
NOR	1.000	-.034	-.630	.327
HOL	1.000	-.295	-.271	-.097
POR	1.000	.953	.531	-.193
GB	1.000	-.498	-.020	-.498
SUE	1.000	-.359	-.868	.351

TAULA V b)

QUADRATS DELS COSINUS ACUMULATS DELS ANGLES QUE FORMEN LES OBSERVACIONS AMB ELS EIXOS PRINCIPALS

1	0.478	0.663	0.672
2	0.318	0.398	0.475
3	0.038	0.065	0.510
4	0.308	0.324	0.391
5	0.004	0.416	0.494
6	0.626	0.815	0.819
7	0.663	0.837	0.841
8	0.372	0.661	0.703
9	0.382	0.384	0.395
10	0.714	0.738	0.865
11	0.490	0.49	0.725
12	0.013	0.178	0.371
13	0.000	0.113	0.655
14	0.001	0.377	0.491
15	0.138	0.248	0.263
16	0.588	0.775	0.798
17	0.361	0.361	0.722
18	0.115	0.758	0.869

AGRAÏMENTS

Agraim a l'empresa Enher les facilitats informàtiques prestades.

Agraim al professor Albert Prat Bartés de la càtedra d'Estadística el seu assessorament al llarg del treball.

BIBLIOGRAFIA

-P. BERTIER - J.M. BOUROCHE. Analyse des données multidimensionnelles. Presses Universitaires de France. Paris, 1975.

-L. LEBART - A. MORINEAU - N. TABARD. Techniques de la description statistique. Dunod, Paris 1977.

-M. TENENHAUS. Analyse en composantes principales d'un ensemble de variables nominales ou numériques. Revue de statistique appliquée. 1977 - Vol. XXV. Nº 2