



FACULTAD DE CIENCIAS

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

2 ej 41

ANALISIS DE COMPONENTES PRINCIPALES DESCRIPTIVO.

ENFOQUE FORMAL

TESIS PROFESIONAL

ACTUARIO

Presenta :

VICENTE ANGEL SORIANO RAMIREZ

TESIS CON
FALLA DE ORIGEN

Cd. Universitaria

1989



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

CAPITULO II

SUMARIO

	Pág.
Introducción General.	i, iv
CAPITULO I	
ANALISIS LINEAL	
Introducción .	1
1. Esquema de dualidad.	2
2. Centro de gravedad y momento de inercia.	8
2.1. Igualdad de Huyghens.	8
2.2. Momento de inercia en relación a un subespacio vectorial.	12
2.3. Momento de inercia en relación a un subespacio afín.	14
2.4. Momento de inercia en relación a una recta y en relación a su hiperplano M -ortogonal.	19
3. Construcción de la métrica de pesos.	24
4. Relación de correlación.	28
5. Análisis en Componentes Principales para una tabla de datos.	33
5.1. Ejes y planos principales.	33
5.2. Plano principal.	37
5.3. Subespacios principales de dimensión k .	38
5.4. Construcción y dualidad.	40
5.5. Descripción de la nube de individuos.	43
5.6. Individuos suplementarios.	45
5.7. Ejemplos.	48

ALGUNOS USOS DEL ACP.	
Introducción.	58
1. Escalamiento Multidimensional.	59
1.1. Caracterización de matrices euclidianas.	59
1.2. Algoritmo para no euclidianas con matriz simétrica.	62
1.3. Dualidad entre Análisis de Coordenadas Principales y Análisis en Componentes Principales.	64
2. Selección de variables con Componentes Principales.	66
3. Ejemplos	
3.1. Matrices de distancias.	74
3.1.1. No euclidianas.	74
3.1.2. Euclidianas.	77
3.2. Selección de variables en regresión.	82

CAPITULO III

ANALISIS CONJUNTO DE VARIAS MATRICES DE DATOS.	
Introducción.	87
1. Extensión del método a varias matrices de datos.	89
1.1. ACP para una tabla. Generalidades.	89
2. Los casos.	92
2.2.1. Definiciones de equivalencia.	95
2.2.2. Proyectores.	96
2.2.2.1. Más equivalencias en términos de proyectores.	97
2.2.3. El caso de las matrices de covarianza.	98
2.2.4. El caso 5 (caract. cualitativas).	98
3. Interestructura-Acuerdos-Intraestructura.	
3.1. Descripción del problema.	101
3.2. La Interestructura.	101
3.2.1. Análisis de la matriz \mathcal{E} .	102
3.2.2. Descripción de los estudios para "características".	105

3.2.3. Descripción de los estudios para "individuos".	108
3.2.4. Construcción de los acuerdos.	109
3.3. La Intraestructura.	116
3.3.1. Sobre un mismo conjunto de individuos.	116
3.3.2. Sobre un mismo conjunto de características.	117
3.3.3. Sobre los mismos conjuntos de individuos y características.	118
3.3.4. Caso donde se tienen las mismas características.	119
4. Tratamiento alternativo para características cualitativas.	121
5. Ejemplo.	127
5.1. Métodos que separan una variación ligada al tiempo de una estructura independiente del tiempo.	127
5.2. Método STATIS.	143

APÉNDICES :	
Apéndice 0.	A0.1-A0.6
Apéndice 1.	A1.1-A1.6
Apéndice 2.	A2.1-A2.2
Apéndice 3.	A3.1-A3.3
Apéndice 4.	A4.1-A4.2
Apéndice 5.	A5.1-A5.2
Apéndice 6.	A6.1-A6.3
Bibliografía.	152

INTRODUCCION

Existen diferentes enfoques para presentar el Análisis en Componentes Principales (ACP). Uno de ellos considera el ACP como un procedimiento para determinar los mejores ejes de un elipsoide derivado de una distribución normal multivariada (un problema de optimización en la varianza); los ejes son estimados a partir de la muestra. Este fue el contexto en el que Harold Hotelling (1933) originalmente lo describe y en muchos textos de análisis multivariado clásico usualmente se presenta bajo este enfoque (ver Anderson, 1958; Kendall y Stewart; Dempster 1969; Kshirsagar, 1972).

Más recientemente el Análisis de datos proporciona un punto de vista enteramente distinto acerca del ACP. En el Análisis de datos no se hacen suposiciones acerca de distribuciones o un modelo estadístico subyacente. Se propone, el uso de una técnica para describir los datos con algunos criterios geométricos y algebraicos de optimización. La rapidez del procesamiento de datos debido a un desarrollo tecnológico en el campo de la electrónica facilita los análisis desde este punto de vista.

El enfoque del Análisis de datos bajo sus principios fundamentales enunciados por Benzécri en *Histoire et Préhistoire de L'Analyse des Données* sitúa esta nueva disciplina en el contexto general de la estadística y muestra su potencia teórica para fundamentar un sinnúmero de métodos descriptivos multivariados. Dichos principios son:

La estadística no es probabilidad. En este principio se resalta la ambigüedad implícita en las aplicaciones de la probabilidad debido al hecho que dicha noción no es única. Para ello se propone el uso de un análisis estadístico independiente de interpretaciones particulares de probabilidad mediante el uso del álgebra y la geometría.

ello se propone el uso de un análisis estadístico independiente de interpretaciones particulares de probabilidad mediante el uso del álgebra y la geometría.

(i) El método debe seguir a los datos y no al revés. En la práctica existe una tendencia de ajustar modelos a los datos más que extraer estructuras de ellos.

(ii) Es conveniente tratar simultáneamente informaciones que conciernen al mayor número posible de dimensiones. El estudio de las relaciones entre las variables observadas permite proponer hipótesis ello implica considerar el mayor número posible de ellas.

(iii) En el análisis debe considerarse el uso de la computadora como una herramienta indispensable. El hecho de pretender analizar una gran cantidad de información crea la necesidad de implementar nuevos métodos de análisis y síntesis. La rapidez en el procesamiento de información permite elaborar dichos métodos.

(iv) El uso de la computadora implica abandonar las técnicas concebidas hasta antes de la aparición de la computadora. La dificultad en el cálculo frecuentemente derivaba en soluciones empíricas que supían a los argumentos matemáticos. El uso de la computadora impone la revisión de las técnicas usadas hasta antes de su aparición.

El Análisis de datos puede ser entendido como un conjunto de técnicas fundamentalmente descriptivas basadas en el álgebra y en la geometría que permiten el estudio de grandes volúmenes de datos multidimensionales de todo tipo (numéricos o no) por medio del uso intensivo de programas de cálculo electrónico. En su origen no tienen carácter inferencial en el sentido clásico de la estadística.

De acuerdo a Benzécri, la historia reciente del pensamiento estadístico en Análisis de datos se puede caracterizar por tres épocas claves:

La biometría que debe su origen a la fusión de dos corrientes

independientes; la antropometría cuyo principal exponente Quételet, habría incorporado el uso del modelo normal (y) por las tesis evolucionistas de Darwin. Galton movido por el deseo de "probar formalmente" las teorías evolutivas descubre de manera empírica la noción de correlación entre dos variables biológicas. Posteriormente Pearson, basado en los desarrollos matemáticos de Dickson sobre los descubrimientos de Galton, extiende los conceptos de correlación en un contexto del modelo normal multivariado.

De Fisher y sus discípulos (Kendall, Neyman, S. Pearson), Benzécri retiene para el Análisis de datos, el carácter geométrico que la fundamenta. Las teorías psicológicas vigentes hacia 1935 (particularmente las basadas en las tesis de Spearman) fundan la psicometría, la cual heredó al Análisis de datos el uso de métodos multidimensionales para el estudio de relaciones entre variables físicas de seres vivos, ésta por su parte permitió desarrollar técnicas que tratan de explicar dimensiones ocultas que en la década de los 40 representó un serio problema numérico, como consecuencia del intento de salvar dicho problema se generaron una serie de técnicas alternativas.

La última etapa es la relativa al análisis de correspondencias. Benzécri presenta en el Colegio de Francia la primera versión del análisis de correspondencias, mostrando sus propiedades geométricas y algebraicas fundamentales. Dicha versión permite tratar tablas de contingencias de dos criterios (variables cualitativas) de manera tal que es posible describir gráficamente las "proximidades" o asociaciones entre los niveles de criterios estudiados.

El Análisis de datos fue elaborado en el contexto general incluyendo dos grandes grupos de métodos; las técnicas de clasificación automática (Roux, 1985) y los análisis factoriales en donde se sitúa el análisis de correspondencias y el ACP.

En esta tesis el objetivo principal es presentar el ACP con la formalidad matemática necesaria desde el punto de vista del

análisis de datos, algunos ejemplos clásicos donde se aplica dicho análisis y finalmente realizar ACP para tablas cúbicas, por ejemplo considerando la evolución del fenómeno respecto al tiempo.

INTRODUCCION

En este capítulo se expone un conjunto de técnicas importantes basadas en el álgebra lineal conocida como Análisis Multidimensional lineal.

Se proporciona al iniciar el capítulo un tratamiento de las nociones de "individuos" y de "características" en el contexto más general posible.

En el marco del análisis multidimensional lineal para formalizar la presentación de las nociones antes mencionadas se requiere introducir métricas con las que sea posible medir la proximidad entre individuos y entre características.

En términos de un esquema de dualidad es posible describir las relaciones entre los diferentes espacios considerados, esto permite de hecho elaborar un resumen de dichas relaciones.

Por último se presenta un panorama de técnicas del análisis multidimensional lineal, además algunos requisitos para desarrollar el Análisis en Componentes Principales.

$\beta = (\hat{f}_t; t=1, \dots, n)$ respectivamente (no necesariamente ortogonales) con las cuales se escribe :

$$\underline{x}_i = \sum_{k=1}^p \alpha_k \hat{e}_k$$

$$\underline{x}^j = \sum_{t=1}^n \gamma_t \hat{f}_t$$

En ambos casos mediante un cambio de base se puede llevar a la base canónica respecto a la cual se tiene ;

-En E con la base canónica $\beta_1^* = (\hat{e}_k; k=1, \dots, p)$

$$\underline{x}_i = (x_i^1, \dots, x_i^1, \dots, x_i^p) = \sum_{k=1}^p x_i^k \hat{e}_k$$

En F con la base canónica $\beta_2^* = (\hat{f}_t; t=1, \dots, n)$

$$\underline{x}^j = (x_1^j, \dots, x_1^j, \dots, x_n^j) = \sum_{t=1}^n x_t^j \hat{f}_t$$

Esto es, la dirección k del espacio E corresponde a la característica k y la dirección t del espacio F corresponde al individuo t (en ambos casos salvo un cambio de base).

El estudio de los espacios E y F se puede efectuar a través de sus espacios duales.

La definición de espacio dual se discute en el apéndice 1, sin embargo es conveniente enunciarla aquí :

Definición : El espacio de los funcionales lineales de E al campo escalar K (R en este caso), se conoce como el espacio dual de E y se denota por E^* .

Esta definición implica que :

-Para cada elemento e_j de la base canónica β_1^* de E existe un único funcional lineal $e_j^* \in E^*$ tal que

$$e_j^*(e_k) = \delta_{jk}$$

a este funcional se le llama la representación funcional de la característica j.

-Para cada f_t de la base canónica β_2^* en F existe un único

I: ANALISIS LINEAL.

1/ Esquema de dualidad.

Suponiendo que se cuenta con n observaciones o individuos para realizar un estudio sobre p variables o características cuantitativas, los datos obtenidos de la medición de dichas características se resumen en la tabla ;

$$\begin{pmatrix} x_1^1 & \dots & x_1^j & \dots & x_1^p \\ \vdots & & \vdots & & \vdots \\ x_i^1 & \dots & x_i^j & \dots & x_i^p \\ \vdots & & \vdots & & \vdots \\ x_n^1 & \dots & x_n^j & \dots & x_n^p \end{pmatrix}$$

donde x_i^j es la medida de la característica j en el individuo i (de aquí y en adelante se hace la convención en llamar una variable cuantitativa como característica y a una observación como individuo).

La manifestación de la característica j en el conjunto de individuos se resume en la columna j de la tabla o en la representación vectorial $\underline{x}^j = (x_1^j, \dots, x_i^j, \dots, x_n^j)'$, de igual manera un individuo se puede caracterizar con respecto a las p características por sus medidas, esto es, la presencia del individuo i se resume en el renglón i de la tabla o en la representación vectorial $\underline{x}_i = (x_i^1, \dots, x_i^j, \dots, x_i^p)$.

Un individuo i se puede pensar como un punto en el espacio E euclidiano de dimensión p ($\underline{x}_i \in E$) ; y una característica j como un punto en el espacio F euclidiano de dimensión n ($\underline{x}^j \in F$).

En los espacios E y F existen bases $\beta_1 = (\hat{e}_k; k=1, \dots, p)$ y

funcional lineal $f_i^* \in F^*$ tal que

$$f_i^*(f_j) = \delta_{ij}$$

, es la representación funcional del individuo i .

Observación :

$$- \phi_j^*(x_i) = \sum_{k=1}^n x_i^k \phi_j^*(e_k) = x_i^j$$

$$- f_i^*(x_j) = \sum_{l=1}^n x_j^l f_i^*(f_l) = x_j^i$$

Se puede interpretar como la manifestación de la característica j en el individuo i .

Sin embargo surge una pregunta. ¿ Cómo saber cuando dos individuos o características son próximos ? La respuesta inmediata sería midiendo la "proximidad" entre sus representaciones.

Para ello es necesario dotar de una métrica a ambos espacios E y F , la construcción propuesta es :

-Considerar una forma bilineal simétrica definida positiva f ; se tiene entonces M la matriz $p \times p$ de entradas $f(e_i, e_j)$.

-El isomorfismo de E y E^* se hace explícito :

$$\begin{array}{ccc} E & \xrightarrow{g} & E^* \\ x & \longmapsto & (y \longmapsto f(x,y)) \end{array}$$

M es así la matriz de g en las bases canónicas.

Se hace la convención de notación en nombrar por M indiferentemente a :

- la forma f
- el isomorfismo g
- la matriz simétrica definida positiva M

-Con la matriz M se define un producto interno

$$\langle x, y \rangle_M = f(x, y) \quad \forall x, y \in E$$

-Con el producto interno es posible definir una norma

$$\|x\|_M^2 = f(x, x) \quad \forall x \in E$$

-Con la norma $\|\cdot\|_M$ es posible definir una métrica

$$d_M(x, y) = \|x - y\|_M \quad \forall x, y \in E$$

La matriz $M \in M_{p \times p}(\mathbb{R})$ permite medir las proximidades entre los individuos, considerados como elementos de E .

En F se puede considerar una construcción semejante que permita medir las proximidades entre las características, consideradas como elementos de F , es decir una matriz $N \in M_{n \times n}(\mathbb{R})$.

Generalmente se tiene que $N=D$, matriz diagonal definida positiva, donde los elementos diagonales son los pesos que afectan a los n individuos.

La matriz X es centrada por D si $DXD = 0$, donde

$$D = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

Retomando la observación donde se relacionan las representaciones vectorial y funcional.

$$- \phi_j^*(x_i) = x_i^j$$

$$- f_i^*(x_j) = x_j^i$$

es posible definir las aplicaciones lineales :

$$E^* \xrightarrow{X} F$$

$$\phi_j^* \longmapsto x(\phi_j^*) = (\phi_j^*(x_1), \dots, \phi_j^*(x_n)) = x_j^i$$

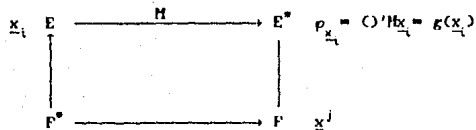
cuya matriz asociada es la generada por la tabla $X \in M_{n \times p}(\mathbb{R})$

$$F^* \xrightarrow{X'} E$$

$$f_i^* \longmapsto x'(f_i^*) = (f_i^*(x_1), \dots, f_i^*(x_p)) = x_i$$

cuya matriz asociada es $X' \in M_{p \times n}(\mathbb{R})$.

En resumen se ha construido el siguiente diagrama conmutativo :



$$\Lambda_{\underline{x}^j} = (C)' D \underline{x}^j = g'(\underline{x}^j)$$

Se puede observar la relación entre las representaciones ;
representación

	funcional	vectorial
individuo i	$\rho_{\underline{x}_i}$	\underline{x}_i
característica j	$\Lambda_{\underline{x}^j}$	\underline{x}^j

En los espacios E y F se cuenta con un producto interno propio \langle, \rangle_M y \langle, \rangle_D respectivamente, que definen una geometría en ellos, un ejemplo de este comentario es la noción de ortogonalidad pues en E se tiene ;

$$\frac{\underline{x}}{M} \perp \frac{\underline{y}}{M} \quad \text{si} \quad \langle \underline{x}, \underline{y} \rangle_M = 0.$$

Interesa saber si la medida de proximidad entre representaciones funcionales es la misma que la establecida entre representaciones vectoriales (en E^* y F^* aún no se ha definido una métrica).

Si considera la aplicación lineal $V = X' \cdot D \cdot X ; E^* \longrightarrow E$ con matriz asociada V (simétrica y definida positiva), se define una métrica en E^* para la cual se satisface ;

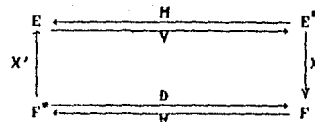
$$\| \underline{x}_i - \underline{x}_i' \|_M = \| \rho_{\underline{x}_i} - \rho_{\underline{x}_i'} \|_V.$$

Con lo cual se prueba que el isomorfismo entre E y E^* es isométrico.

Análogamente al proponer la aplicación lineal $W = X \cdot M \cdot X' ; F^* \longrightarrow F$ con matriz asociada W (simétrica y definida positiva) se define una métrica en F^* que satisface ;

$$\| \underline{x}^j - \underline{x}^j' \|_D = \| \Lambda_{\underline{x}^j} - \Lambda_{\underline{x}^j'} \|_W$$

esto es, el isomorfismo entre F y F^* es isométrico.
Se tiene el esquema siguiente :



Observaciones ;

- Si X está centrada, V es la matriz de covarianzas entre las p características de la tabla X.

- Se tiene $\langle \underline{x}^j, \underline{x}^j' \rangle_D = \| \underline{x}^j \|_D \| \underline{x}^j' \|_D \cos \theta$ donde θ es el ángulo entre \underline{x}^j y \underline{x}^j' . $\cos \theta$ es la correlación entre las características j y j'.

- El esquema de dualidad es determinado por

X la matriz de datos

M la métrica en E

D la métrica en F

por tanto se puede representar por la terna (X,M,D).

2. Centro de masa y momento de inercia.

2.1. Igualdad de Huyghens.

La matriz X de n renglones y p columnas corresponde en E a una nube \mathcal{M} de individuos:

$$\mathcal{M} = \{ x_i; i = 1, \dots, n \}$$

donde a cada individuo i se asocia un peso ρ_i ($\rho_i > 0$ se interpreta también como un porcentaje de la importancia total de los individuos, esto es $\sum \rho_i = 1$).

Definición: El momento de inercia respecto al punto $\underline{a} \in E$ de la nube \mathcal{M} denotado como $I_{\underline{a}}$ (también llamado la inercia de la nube \mathcal{M} respecto al centro de gravedad \underline{a}) se determina por

$$I_{\underline{a}} = \sum_{i=1}^n \rho_i \|x_i - \underline{a}\|_M^2.$$

Se observa que si \underline{a} es un punto variable, $I_{\underline{a}}$ significa la medida de la proximidad de \mathcal{M} al punto $\underline{a} \in E$.

Resulta fácil probar que $I_{\underline{a}} = 0$ implica que

$$\sum_{i=1}^n \rho_i x_i = \underline{a}.$$

En efecto:

Que $I_{\underline{a}} = 0$ es equivalente a que:

$$\begin{aligned} 0 &= \sum_{i=1}^n \rho_i \|x_i - \underline{a}\|_M^2 = \sum_{i=1}^n \rho_i \langle x_i - \underline{a}, x_i - \underline{a} \rangle_M \\ &= \sum_{i=1}^n \rho_i \langle x_i - \underline{a}, x_i - \underline{a} \rangle_M \\ &= \sum_{i=1}^n \langle \rho_i (x_i - \underline{a}), y_i \rangle_M \\ &= \left\langle \sum_{i=1}^n \rho_i (x_i - \underline{a}), y_i \right\rangle_M \end{aligned}$$

$$\begin{aligned} &= \left\langle \sum_{i=1}^n \rho_i x_i - \sum_{i=1}^n \rho_i \underline{a}, y_i \right\rangle_M \\ &= \left\langle \sum_{i=1}^n \rho_i x_i - \underline{a}, y_i \right\rangle_M \quad i=1, \dots, p \end{aligned}$$

con $y_i = x_i - \underline{a}$ y en general para cualquier combinación lineal \underline{t} de las y_i .

$$0 = \left\langle \sum_{i=1}^n \rho_i x_i - \underline{a}, \underline{t} \right\rangle_M$$

en particular $\underline{t} = (1, 1, \dots, 1)$ entonces

$$\sum_{i=1}^n \rho_i x_i = \underline{a}.$$

Esta prueba proporciona un método para encontrar el centro de gravedad de la nube \mathcal{M} denotado aquí y en adelante por:

$$\underline{g} = \sum_{i=1}^n \rho_i x_i$$

cuya j -ésima componente en la base canónica:

$$g_j = \sum_{i=1}^n \rho_i x_{ij} \quad j=1, \dots, p$$

es el promedio ponderado de la característica j .

Dada una nube $\mathcal{M} \subset \mathbb{R}^M$ (siempre es posible encontrar su centro de gravedad \underline{g} y para cualquier $\underline{a} \in E$ se tiene:

$$x_i - \underline{a} = x_i - \underline{g} + \underline{g} - \underline{a}$$

de aquí se infiere:

$$\begin{aligned} \|x_i - \underline{a}\|_M^2 &= \langle x_i - \underline{a}, x_i - \underline{a} \rangle_M \\ &= \langle x_i - \underline{g} + \underline{g} - \underline{a}, x_i - \underline{g} + \underline{g} - \underline{a} \rangle_M \\ &= \langle x_i - \underline{g}, x_i - \underline{g} \rangle_M + \langle \underline{g} - \underline{a}, \underline{g} - \underline{a} \rangle_M + 2 \langle x_i - \underline{g}, \underline{g} - \underline{a} \rangle_M \\ &= \|x_i - \underline{g}\|_M^2 + \|\underline{g} - \underline{a}\|_M^2 + 2 \langle x_i - \underline{g}, \underline{g} - \underline{a} \rangle_M \end{aligned}$$

en estos términos la fórmula de la inercia en \underline{a} adquiere la forma:

$$\begin{aligned}
 I_{\underline{a}} &= \sum_{i=1}^n \rho_i \|X_i - \underline{a}\|_M^2 \\
 &= \sum_{i=1}^n \rho_i (\|X_i - \underline{g}\|_M^2 + \|\underline{g} - \underline{a}\|_M^2 + 2 \langle X_i - \underline{g}, \underline{g} - \underline{a} \rangle_M) \\
 &= \sum_{i=1}^n \rho_i (\|X_i - \underline{g}\|_M^2 + \|\underline{g} - \underline{a}\|_M^2) + \sum_{i=1}^n \rho_i 2 \langle X_i - \underline{g}, \underline{g} - \underline{a} \rangle_M \\
 &= \sum_{i=1}^n \rho_i \|X_i - \underline{g}\|_M^2 + \|\underline{g} - \underline{a}\|_M^2 \sum_{i=1}^n \rho_i + 2 \left\langle \sum_{i=1}^n \rho_i X_i - \underline{g}, \underline{g} - \underline{a} \right\rangle_M \\
 &= \sum_{i=1}^n \rho_i \|X_i - \underline{g}\|_M^2 + \|\underline{g} - \underline{a}\|_M^2
 \end{aligned}$$

$$I_{\underline{g}} + \|\underline{g} - \underline{a}\|_M^2$$

esta conclusión es conocida como igualdad de Huyghens y $\|\underline{g} - \underline{a}\|_M^2$ es el cuadrado de la M-distancia de \underline{a} al centro de gravedad de los puntos en \mathcal{M} (la distancia de \underline{a} a \mathcal{M}). Un caso particular se tiene cuando la matriz M es la matriz identidad.

Observación : $I_{\underline{a}} : E \longrightarrow [0, \infty)$

$$\underline{a} \longrightarrow I_{\underline{a}} = \sum_{i=1}^n \rho_i \|X_i - \underline{a}\|_M^2$$

$I_{\underline{a}}$ mide la distancia de \underline{a} a la nube de puntos \mathcal{M} que se desea describir, esto obliga a situarse en el punto más cercano a \mathcal{M} , es decir, se requiere minimizar $I_{\underline{a}} : \underline{a} \in E$.

Para ello denota por \underline{g} el centro de gravedad de la nube \mathcal{M} e $I_{\underline{g}}$ denota el momento de inercia de \mathcal{M} de un punto $\underline{g} \in E$ ($I_{\underline{g}}$ es una función de $\underline{g} \in E$), entonces $I_{\underline{a}}$ es mínimo en $\underline{a} = \underline{g}$.

En efecto, minimizar $I_{\underline{a}} = I_{\underline{g}} + \|\underline{g} - \underline{a}\|_M^2$ es equivalente a minimizar $\|\underline{g} - \underline{a}\|_M^2$ lo cual ocurre en $\underline{a} = \underline{g}$.

Por tanto $\min_{\underline{a} \in E} I_{\underline{a}} = I_{\underline{g}}$ y \underline{g} es el punto más cercano de todos los puntos de E a la nube \mathcal{M} y la M-distancia más corta es

precisamente $I_{\underline{g}}$.

La nube \mathcal{M} está concentrada en \underline{g} si :

$$I_{\underline{g}} = \sum_{i=1}^n \rho_i \|X_i - \underline{g}\|_M^2 = 0$$

puesto que esta igualdad implica que $\|X_i - \underline{g}\|_M = 0$ y por tanto

$$X_i = \underline{g} = \underline{g} \quad \forall i=1, \dots, n.$$

Otra manera de medir la dispersión alrededor del punto \underline{g} de la misma forma que la varianza mide la dispersión alrededor del centro de gravedad requiere de las propiedades de la traza :

$$\text{tr}(AB) = \text{tr}(BA) \quad ; \quad \text{tr}(A+B) = \text{tr}(A) + \text{tr}(B)$$

$$I_{\underline{g}} = \sum_{i=1}^n \rho_i M(X_i, X_i) = \sum_{i=1}^n \rho_i X_i' M X_i$$

donde cada $X_i' M X_i = \text{tr}(X_i' M X_i) = \text{tr}(X_i X_i' M)$ de aquí se deduce que :

$$\begin{aligned}
 I_{\underline{g}} &= \sum_{i=1}^n \rho_i \text{tr}(X_i X_i' M) \\
 &= \sum_{i=1}^n \text{tr}(\rho_i X_i X_i' M) \\
 &= \text{tr} \left[\left(\sum_{i=1}^n \rho_i X_i X_i' \right) M \right] \\
 &= \text{tr} \left[\left(\sum_{i=1}^n \rho_i X_i X_i' \right) M \right] \\
 &= \text{tr}(VH)
 \end{aligned}$$

Por tanto $I_{\underline{g}}$ y $\text{tr}(VH)$ son las nociones generalizadas de media y varianza.

Observaciones :

-Si $M = I$ entonces $I_{\underline{g}} = \text{tr}(V) = \sum_{i=1}^n \text{var}(X_i)$.

-Si $M = D_{1 \times k}^{-2}$ entonces $I_{\mathcal{M}} = \text{tr} (V D_{1 \times k}^{-2})$
 $= \text{tr} (D_{1 \times k}^{-2} M D_{1 \times k}^{-2})$
 $= \text{tr} (C) = p,$

donde C es la matriz de correlación.

-Si M es métrica entonces $I_{\mathcal{M}} = \text{tr} (Y' M Y)$

$$= \sum_{i=1}^n \text{var}(Y_{X_i}).$$

para cualquier matriz Y asociada a una transformación $\psi: E \rightarrow E$.

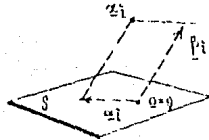
2.2 Momento de inercia en relación a un subespacio vectorial.

Sea S un subespacio vectorial de E y sea $S \oplus S^{\perp}$ la descomposición de E respecto a S , es decir para cada $x_i \in E$ se tiene

$$x_i = \alpha_i + \beta_i$$

para algún $\alpha_i \in S$ y $\beta_i \in S^{\perp}$, $i=1, \dots, n$.

Además α_i es la proyección M -ortogonal de x_i al subespacio S , y como se vio en la sección precedente la proximidad de la nube \mathcal{M} a S está dada por I_S con la interpretación geométrica:



$$I_S = \sum_{i=1}^n \rho_i \|\beta_i\|_M^2 \quad (1)$$

A la expresión (1) se le conoce como momento de inercia de \mathcal{M} en relación al subespacio vectorial S .

Análogamente:

$$I_{S^{\perp}} = \sum_{i=1}^n \rho_i \|\alpha_i\|_M^2$$

es el momento de inercia en relación al subespacio vectorial S^{\perp} , mide la proximidad de la nube \mathcal{M} al subespacio vectorial S^{\perp} .

$\mathcal{M} \subset S$ ssi $I_S = 0$.

En efecto $\mathcal{M} \subset S$ ssi $\forall i, x_i = \alpha_i$ y $\beta_i = 0$

$$\text{ssi } I_S = \sum_{i=1}^n \rho_i \|\beta_i\|_M^2 = 0.$$

-De acuerdo al teorema de Pitágoras,

$$I_{\mathcal{M}} = \sum_{i=1}^n \rho_i \|\alpha_i\|_M^2 + \sum_{i=1}^n \rho_i \|\beta_i\|_M^2$$

o bien: $I_{\mathcal{M}} = I_S + I_{S^{\perp}}$.

Se observa que si A y $(I-A)$ designan los proyectores asociados a la descomposición $E = S \oplus S^{\perp}$ entonces:

$$I_{S^{\perp}} = \text{tr} (VMA) \quad \text{y} \quad I_S = \text{tr} (VM(I-A)).$$

En efecto:

Si

$$x_i'(I-A)'M(I-A)x_i = \text{tr}(x_i'(I-A)'M(I-A)x_i) = \text{tr}((I-A)x_i x_i'(I-A)'M)$$

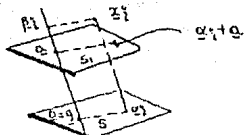
y se tiene:

$$\begin{aligned} I_S &= \sum_{i=1}^n \rho_i \|(I-A)x_i\|_M^2 = \sum_{i=1}^n \rho_i \langle (I-A)x_i, (I-A)x_i \rangle_M \\ &= \sum_{i=1}^n \rho_i x_i' (I-A)' M (I-A) x_i \\ &= \sum_{i=1}^n \rho_i \text{tr}((I-A)x_i x_i'(I-A)'M) \\ &= \text{tr} \left\{ \left(\sum_{i=1}^n \rho_i (I-A)x_i x_i'(I-A)' \right) M \right\} \end{aligned}$$

$$= L((VCI-A)H) = L((VHC)-A)$$

2.9 Momento de inercia en relación a un subespacio afín.

Sea S_1 un subespacio vectorial afín de E , \vec{a} un subespacio vectorial trasladado



Se observa que si $S_1 \cap S_1^\perp = \{\vec{a}\}$, donde $\vec{a} \in S_1^\perp$ entonces se obtiene una descripción conjuntista de S_1 :

$$S_1 = \{ \vec{y} \mid \vec{y} = \vec{x} + \vec{a}, \vec{x} \in S \}$$

, es decir S_1 es un elemento de la clase cuyo representante es S , los elementos en dicha clase son S salvo traslaciones hechas a través de elementos \vec{a} en S_1^\perp .

i.e. $S_1 = S + \vec{a}$.

Cada $\vec{x}_i \in E$ se escribe de manera única como:

$$\vec{x}_i = \vec{\alpha}_i + \vec{\beta}_i$$

con $\vec{\alpha}_i \in S$ y $\vec{\beta}_i \in S_1^\perp \quad \forall i=1, \dots, n$. Para $\vec{a} \in E$ fijo se escribe

$$\vec{x}_i = \vec{\alpha}_i + \vec{a} + \vec{\beta}_i - \vec{a}$$

donde $\vec{\alpha}_i + \vec{a} \in S + \vec{a} = S_1$ y $\vec{\beta}_i - \vec{a} \in S_1^\perp = S_1^\perp$, como lo muestra la figura:

Con estas observaciones se establece el momento de inercia en relación a S_1 :

$$I_{S_1} = \sum_{i=1}^n \rho_i \| \vec{\beta}_i - \vec{a} \|_M^2$$

que mide la proximidad de la nube M con el subespacio afín S_1 .

Proposición: Sea S_1 un subespacio afín de E (S_1 elemento de la clase cuyo representante es S) y $\vec{a} \in S_1^\perp$ fijo. Entonces

$$I_{S_1} = I_S + \| \vec{a} \|_M^2$$

Demostración:

En efecto:

$$\begin{aligned} I_{S_1} &= \sum_{i=1}^n \rho_i \| \vec{\beta}_i - \vec{a} \|_M^2 \\ &= \sum_{i=1}^n \rho_i \langle \vec{\beta}_i - \vec{a}, \vec{\beta}_i - \vec{a} \rangle_M \\ &= \sum_{i=1}^n \rho_i (\langle \vec{\beta}_i, \vec{\beta}_i \rangle_M + \langle \vec{a}, \vec{a} \rangle_M - 2 \langle \vec{\beta}_i, \vec{a} \rangle_M) \\ &= \sum_{i=1}^n \rho_i \| \vec{\beta}_i \|_M^2 + \| \vec{a} \|_M^2 - 2 \sum_{i=1}^n \rho_i \langle \vec{\beta}_i, \vec{a} \rangle_M \\ &= \| \vec{\beta} \|_M^2 + \| \vec{a} \|_M^2 - 2 \langle \sum_{i=1}^n \rho_i \vec{\beta}_i, \vec{a} \rangle_M \end{aligned}$$

Puesto que $\vec{a} = \vec{0}$ entonces:

$$\vec{a} = \sum_{i=1}^n \rho_i \vec{x}_i = \sum_{i=1}^n \rho_i \vec{\alpha}_i + \sum_{i=1}^n \rho_i \vec{\beta}_i = \vec{0}$$

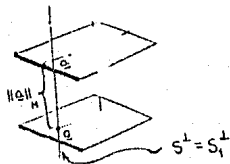
$$\left\{ \begin{array}{l} \sum_{i=1}^n p_i \underline{\alpha}_i = \underline{0} \\ \sum_{i=1}^n p_i \underline{\beta}_i = \underline{0} \end{array} \right.$$

$$\begin{aligned} 0 = H(\underline{0}, \underline{a}) &= H\left(\sum_{i=1}^n p_i \underline{\alpha}_i, \underline{a}\right) = \left\langle \sum_{i=1}^n p_i \underline{\alpha}_i, \underline{a} \right\rangle_M \\ &= \sum_{i=1}^n p_i \langle \underline{\alpha}_i, \underline{a} \rangle_M \end{aligned}$$

ello permite asegurar también que :

$$I_S^{\perp} = I_S + \|\underline{a}\|_M^2 \quad (2)$$

Observación: $\|\underline{a}\|_M$ es la distancia que se encuentran separados S y S_1^{\perp} .



Consecuencias de la expresión (2) :

-Si S_1 es un subespacio afín y S el subespacio vectorial asociado, bajo la suposición de que el centro de gravedad $\underline{g} = \underline{0}$ para la nube \mathcal{M} . Entonces es intuitivamente claro que S es más cercano a \mathcal{M} que el subespacio afín S_1 , esto es :

$$I_S \leq I_{S_1}$$

Esta desigualdad es trivial de la expresión (2).

-La interpretación geométrica a este hecho es que el subespacio afín de dimensión k más próximo a la nube \mathcal{M} pasa por el centro de gravedad \underline{g} , es un subespacio vectorial de E ($\underline{g} = \underline{0}$).

Consecuencias inmediatas de la construcción y del teorema de Pitágoras es el siguiente :

Teorema : Sea S un subespacio vectorial de E tal que S es

suma directa de subespacios ortogonales S_1^{\perp} y S_2^{\perp} , entonces

$$I_S^{\perp} = I_{S_1^{\perp}} + I_{S_2^{\perp}}$$

Demostración :

Si $S = S_1^{\perp} \oplus S_2^{\perp}$ se tiene que $S_1^{\perp} \perp S_2^{\perp}$ y por tanto :

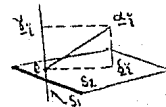
$$E = S \oplus S^{\perp} = S_1^{\perp} \oplus S_2^{\perp} \oplus S^{\perp}$$

por tanto \underline{x}_i se puede escribir como :

$$\underline{x}_i = \underline{\alpha}_i + \underline{\beta}_i \quad \text{con } \underline{\alpha}_i \in S \text{ y } \underline{\beta}_i \in S^{\perp}, \text{ además}$$

$\underline{\alpha}_i = \underline{\gamma}_i + \underline{\delta}_i$ con $\underline{\gamma}_i \in S_1^{\perp}, \underline{\delta}_i \in S_2^{\perp}$, por el teorema de Pitágoras :

$$I_S^{\perp} = \sum_{i=1}^n p_i \|\underline{\alpha}_i\|_M^2 = \sum_{i=1}^n p_i \|\underline{\gamma}_i\|_M^2 + \sum_{i=1}^n p_i \|\underline{\delta}_i\|_M^2 \quad (3)$$



donde $\underline{\gamma}_i$ y $\underline{\delta}_i$ son las proyecciones H -ortogonales de \underline{x}_i sobre S_1^{\perp} y S_2^{\perp} ; Por ello de (3) se obtiene que :

$$I_S = I_{S_1^{\perp}} + I_{S_2^{\perp}} \quad (4)$$

-Un hecho importante es que de encontrar un subespacio S tal que I_S es mínimo entonces hay un subespacio S_1 de S tal que también I_{S_1} es mínimo, así que es posible reducir la dimensión del subespacio más próximo hasta donde se desee, para proporcionar una representación de los individuos lo más sencilla posible aunque se corra el riesgo de no explicar totalmente la dispersión de la nube de puntos \mathcal{M} .

Teorema : Todo espacio vectorial $S_1 \subset E$ de dimensión l tal que I_{S_1} es mínima, entonces existe un subespacio vectorial S_{l-1} de dimensión $(l-1)$ tal que $I_{S_{l-1}}$ es mínima.

Prueba :

Sea S_1 de dimensión l , $l \geq 1$, entonces :

$$E = S_l \oplus S_l^\perp.$$

Si se considera :

$$E \xrightarrow{\rho_l} S_l$$

entonces ρ_l es lineal y $\ker \rho_l = S_l^\perp$ es un subespacio de E con base $\{e_1, \dots, e_{n-l}\}$ M -ortogonal, los posibles subespacios de dimensión $l+1$ que contienen a S_l son precisamente :

$$S_{l+1}^j = S_l \oplus \langle u_j \rangle \quad \text{donde } u_j = \rho_l^{-1}(e_j) \quad j=1, \dots, n-l$$

, sin embargo es necesario seleccionar el subespacio S_{l+1}^j de inercia mínima. Defina el conjunto

$$\mathcal{G} = \left\{ S_{l+1}^j \mid \min_j I_{S_{l+1}^j} \right\}$$

, se observa que \mathcal{G} es no vacío puesto que solo hay un número finito de S_{l+1}^j (exactamente $n-l$ subespacios de dimensión $l+1$), por lo tanto hay a lo más un número finito de $I_{S_{l+1}^j}$, esto es, $\exists S_{l+1}^j$ tal que $\min_j I_{S_{l+1}^j} = I_{S_{l+1}^j}$.

Se observa que en el caso en que $l=n-1$ el tal subespacio de dimensión $l+1$ es único y es E , en cualquier otro caso no se garantiza la unicidad de un tal subespacio, aunque en las aplicaciones es poco frecuente encontrarse en tal caso, el procedimiento para encontrar los vectores base de S_l^\perp es conocido como procedimiento de Gram-Schmidt, haciendo la salvedad de la métrica con la que se trabaja.

La búsqueda de un subespacio vectorial de dimensión l lo más próximo posible a \mathcal{M} requiere de una sucesión de "búsquedas".

- Iniciar con la búsqueda de la recta $\langle u_1 \rangle$ lo más próxima a \mathcal{M} .
- Buscar la recta $\langle u_2 \rangle$, M -ortogonal a $\langle u_1 \rangle$ lo más próxima a \mathcal{M} y así sucesivamente hasta hallar $\langle u_l \rangle$, M -ortogonal a $\langle u_1, \dots, u_{l-1} \rangle$ lo más próxima a \mathcal{M} .

2.4. Momento de inercia en relación a una recta y en relación a su hiperplano M -ortogonal.

Durante la búsqueda del eje $\langle u \rangle$ lo más próximo a \mathcal{M} se requiere tener conocimiento de la variación del momento de inercia $I_{\langle u \rangle}$ de la nube \mathcal{M} en relación a cualquier otra recta $\langle u \rangle$ generada por el vector unitario \underline{u} . ($\|\underline{u}\|_M^2 = M(\underline{u}, \underline{u}) = 1$).

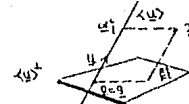
Esto es :

$$I_{\langle u \rangle} = \sum_{i=1}^n \rho_i \|\xi_i\|_M^2 \quad \text{y} \quad I_{\langle u \rangle^\perp} = \sum_{i=1}^n \rho_i \|\underline{u}\|_M^2$$

los momentos de inercia de la nube \mathcal{M} en relación con la recta $\langle u \rangle$ y el hiperplano suplementario M -ortogonal $\langle u \rangle^\perp$.

Por esta razón : $E = \langle u \rangle \oplus \langle u \rangle^\perp$ i.e. si para $\underline{x} \in E$ se escribe de manera única

$$\underline{x} = \underline{\alpha} + \underline{\beta} \quad \text{con } \underline{\alpha} \in \langle u \rangle \quad ; \quad \underline{\beta} \in \langle u \rangle^\perp$$



Sabemos que $I_{\underline{g}} = \text{tr}(YH) = I_{\langle u \rangle} + I_{\langle u \rangle^\perp}$, donde $\underline{\alpha}$ es colineal a \underline{u} , es decir, existe $\underline{c} \in \mathbb{R}$ tal que $\underline{\alpha} = \underline{c} \underline{u}$, donde $\underline{c} = M(\underline{x}, \underline{u}) = \underline{u}^*(\underline{x}) = \underline{x}_1 | \underline{H} \underline{u}$.

Observaciones :

En la última serie de igualdades se tienen para M los siguientes significados:

En el segundo término, M es forma bilineal.

En el cuarto término M es matriz.

La función \underline{u}^* es el único funcional lineal asociado a \underline{u} bajo el isomorfismo entre los espacios E y E^* , en este contexto \underline{u}^* es la función lineal que a todo $\underline{x} \in E$ le asocia la coordenada de su proyección M -ortogonal sobre la recta $\langle u \rangle$.

Ejemplo : Suponga $\underline{u} = (1, 0, 0)$ $\langle u \rangle = \{(x, 0, 0) | x \in \mathbb{R}\}$ y $M=I_3$ entonces

$$\underline{u}^*(y) = \underline{y}' I_2 (1,0,0) = (y_1, y_2, y_3) \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = y_1$$

la componente de la I-proyección sobre $\langle \underline{u} \rangle$.

$M(\underline{x}, \underline{u}) = \underline{x}' M \underline{u}$ define el isomorfismo entre E y E^* cuando fija una de las entradas de la forma bilineal $M(\underline{x}, \underline{u}) = M(\underline{x})$ pasa a ser un funcional lineal.

$$E \xrightarrow{M} E^*$$

$$\underline{u} \longmapsto M(\underline{u})$$

y la forma bilineal es exactamente el M-producto interior de los elementos φ_i y φ_k de la base.

$$M(\underline{x}, \underline{y}) = M \left(\sum_{i=1}^n x^i \varphi_i, \sum_{k=1}^n y^k \varphi_k \right) = \sum_{i=1}^n x^i y^k M(\varphi_i, \varphi_k) = \sum_{i=1}^n x^i y^k m_{ik}$$

$$m_{ik} = M(\varphi_i, \varphi_k).$$

$$\text{Notación: } \underline{c} = X(\underline{u}^*) = X(\underline{c})' M \underline{u} = \sum_{i=1}^n c_i \varphi_i.$$

- La forma lineal $\underline{u}^* \in E^*$, $\underline{u}^* = M(\underline{u})$ solo depende de \underline{u} .
- El vector de características $\underline{c} \in E^*$ y por la definición de la métrica D_p :

$$J_{(\underline{u}, \underline{c})} = D_p(\underline{c}) = \|\underline{c}\|_{D_p}^2.$$

-De la construcción del esquema de dualidad:

$$V(\underline{c}) = M(\underline{u}^*)$$

(Las formas cuadráticas V y M).

Además:

$V(\underline{u}^*) = \langle V \underline{u}^*, \underline{u}^* \rangle = \langle VM(\underline{u}), M(\underline{u}) \rangle = \langle MVH(\underline{u}), \underline{u} \rangle$
 esto es, al considerar la aplicación lineal $MVH: E \rightarrow E^*$ con lo cual se genera una forma bilineal, una cuadrática, un isomorfismo (los que se denotan por un símbolo común).

$$\text{i.e. } MVH(\underline{u}^*) = \langle MVH(\underline{u}), \underline{u} \rangle = V(\underline{u}),$$

-Donde el primer miembro de la igualdad es la forma

cuadrática asociada a la aplicación MVH.

-El segundo miembro de la igualdad es el I-producto interno de $V(\underline{u})$ y \underline{u} .

-El tercer miembro de la igualdad es la forma cuadrática asociada a V .

En resumen:

$$- J_{(\underline{u}, \underline{c})} = \text{tr}(VMH) = J_{(\underline{u}, \underline{c})},$$

y además:

$$J_{(\underline{u}, \underline{c})} = D_p(\underline{c}) = V(\underline{u}^*) = MVH(\underline{u}) = M(\underline{u}, VMH\underline{u})$$

$$\text{donde } \|\underline{u}\| = 1 \text{ y } \underline{c} = X'(\underline{u}^*) = X'(M(\underline{u})) = X'M(\underline{u}).$$

- Si $\|\underline{u}\| \neq 1$ se tiene:

$$J_{(\underline{u}, \underline{c})} = \frac{M(\underline{u}, VMH(\underline{u}))}{M(\underline{u}, \underline{u})} = \frac{MVH(\underline{u})}{M(\underline{u}, \underline{u})}.$$

- Si $M(\underline{u}, \underline{u}) \neq 0 \Leftrightarrow M(\underline{u}, VMH(\underline{u})) = 0 \Leftrightarrow \underline{u} \in \text{Ker}(VMH)$.

Es posible demostrar que en el espacio E con el producto interno generado por M una matriz simétrica definida positiva como se construye en la primera sección de este capítulo, vale la desigualdad de Schwarz.

Lema: Para toda x, y en un espacio con producto interno

(E, M) ,

$$\| \langle x, y \rangle \|_M \leq \|x\|_M \|y\|_M.$$

La igualdad es cierta si $x = \lambda y$ ó $y = 0$.

Prueba:

Si $y = 0$, la desigualdad es cierta trivialmente. Por tanto se supone que $y \neq 0$. Para todos los escalares s , se tiene

$$0 \leq M(\underline{x} - s\underline{y}, \underline{x} - s\underline{y}) = M(\underline{x}, \underline{x}) - sM(\underline{y}, \underline{x}) - sM(\underline{x}, \underline{y}) + |s|^2 M(\underline{y}, \underline{y}).$$

En particular, para $s = M(\underline{x}, \underline{y}) / M(\underline{y}, \underline{y})$, se tiene

$$0 \leq M(\underline{x}, \underline{x}) - \left[\frac{|M(\underline{x}, \underline{y})|^2}{M(\underline{y}, \underline{y})} \right] / M(\underline{y}, \underline{y})$$

$$\text{ó } \|M(\underline{x}, \underline{y})\| \leq \left\{ M(\underline{x}, \underline{x}) M(\underline{y}, \underline{y}) \right\}^{1/2} = \|x\|_M \|y\|_M.$$

Una consecuencia de este resultado se enuncia y se prueba en el siguiente:

Teorema : Si $M \prec \langle u \rangle^\perp \rightarrow \begin{cases} I_{(VM(u))} \geq I_{\langle u \rangle^\perp} \\ I_{(VM(u))} = I_{\langle u \rangle^\perp} \end{cases} \Leftrightarrow VMu = \lambda u \quad \lambda \geq 0.$

Prueba :

De la hipótesis se sigue que $M \prec \langle u \rangle^\perp$ y si se tienen en cuenta las igualdades, es claro que :

$$I_{(VM(u))} \geq I_{\langle u \rangle^\perp} \Leftrightarrow I_{(VM(u))} \leq I_{\langle u \rangle^\perp}$$

de manera que basta analizar la inercia de los complementos

ortogonales :

$$I_{\langle u \rangle^\perp} = \frac{M(u, VM(u))}{M(u, u)} \quad y$$

$$I_{(VM(u))} = \frac{M(VMu, VMVMu)}{M(VMu, VMu)} = \frac{MVM(u, VMVMu)}{MVM(u, VMu)}$$

De la aplicación del lema anterior se sigue :

$$(M(u, VM(u)))^2 \leq M(u, u) M(VM(u), VM(u)) \quad (5)$$

$$(MVM(u), VM(u))^2 \leq MVM(u, u) MVM(VM(u), VM(u)) \quad (6)$$

Si $u \neq 0$ entonces ambos los denominadores son no nulos puesto que si M es definida positiva entonces también lo es MVM , de la expresión (5) se obtiene :

$$\frac{M(u, VM(u))}{MVM(u, u)} \leq \frac{MVM(VM(u), VM(u))}{MVM(u, VM(u))} = \frac{MVM(u, VM(u))}{M(VM(u), VM(u))} \quad (7)$$

analogamente de la expresión (6) :

$$\frac{MVM(u, VM(u))}{M(u, u)} \leq \frac{MVM(u, VM(u))}{MVM(u, u)} = \frac{MVM(VM(u), VM(u))}{M(VM(u), VM(u))} \quad (8)$$

de (7) y (8), las expresiones en cocientes para la inercia de los espacios ortogonales y la observación inicial se infiere el primer resultado.

Si ocurre que $I_{(VM(u))} = I_{\langle u \rangle^\perp}$ entonces :

$(M(u, VM(u)))^2 = M(u, u) M(VM(u), VM(u))$, la forma M es definida positiva y $\ker(VM)$.

En efecto : $I_{(VM(u))} = I_{\langle u \rangle^\perp} \Leftrightarrow I_{(VM(u))} = I_{\langle u \rangle^\perp}$ y se observa que $VM(u) = \lambda u \quad \lambda \geq 0$ ($u \in \ker(VM)$). Y junto con el

teorema anterior $VMu = \lambda u \rightarrow I_{\langle u \rangle^\perp} = \frac{M(u, VM(u))}{M(u, u)}$ con

$$u \in \ker(VM)$$

$$= \frac{M(u, \lambda u)}{M(u, u)} = \lambda.$$

El resultado muestra que $I_{\langle u \rangle^\perp}$ es mínimo cuando u pertenece a un subespacio propio asociado al más grande eigenvalor de VM .

Observación :

$$\text{Si } u \perp u_1 \quad y \quad VMu_1 = \lambda_1 u_1 \rightarrow VMu \perp u_1.$$

En efecto : si $u \perp u_1$ entonces $M(u, u_1) = 0$.

$$\text{Por tanto : } 0 = \lambda_0 M(u, u_1) = M(u, \lambda_1 u_1)$$

$$= M(u, VMu_1) = MVM(u, u_1) = M(VMu, u_1) \rightarrow VMu \perp u_1.$$

2. / Construcción de la métrica de pesos.

La construcción formal de una métrica que tenga en cuenta la importancia o peso de los elementos se desarrollará bajo la suposición sin pérdida de generalidad que $\underline{g} = \mathbf{1}$ (si esto no ocurre, simplemente considere la traslación de \mathcal{A} de tal manera que esto ocurra, esto es $\mathcal{A} \setminus \underline{g}$).

Los valores que asumen las x_i son arreglos de números reales en este caso de características cuantitativas es por esto que se puede considerar como un vector aleatorio y su valor esperado es precisamente:

$$\underline{g} = \left[\sum_{i=1}^n \rho_i x_i^1, \dots, \sum_{i=1}^n \rho_i x_i^p \right]$$

donde la j -ésima coordenada de \underline{g} :

$$g_j = \sum_{i=1}^n \rho_i x_i^j$$

es el valor esperado de la característica j con las ponderaciones ρ_j .
La construcción matricial:

$$g_j = (\rho_1, \dots, \rho_n) \begin{bmatrix} x_1^j \\ \vdots \\ x_n^j \end{bmatrix} = \rho \cdot X^j$$

Por tanto:

$$\underline{g} = (\rho_1, \dots, \rho_n) \begin{bmatrix} x_1^1 & \dots & x_1^j & \dots & x_1^p \\ \vdots & & \vdots & & \vdots \\ x_n^1 & \dots & x_n^j & \dots & x_n^p \end{bmatrix} = \rho \cdot X$$

Mediante una construcción similar se obtiene a la covarianza

entre las características j y j' :

$$(x_1^j, \dots, x_n^j) \begin{bmatrix} \rho_1 & 0 & \dots & 0 \\ 0 & \rho_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \rho_n \end{bmatrix} \begin{bmatrix} x_1^{j'} \\ \vdots \\ x_n^{j'} \end{bmatrix} = \sum_{i=1}^n \rho_i x_i^j x_i^{j'}$$

Por tanto: $\text{Var}(j) = \sum_{i=1}^n \rho_i x_i^{j^2}$

y también la matriz de covarianzas se obtiene como:

$$X^j D X^{j'}$$

La matriz D como se puede observar define una métrica de pesos y tiene una forma bilineal asociada D_p tal que:

$$D_p(f_i, f_{i'}) = 0 \quad \text{si } i \neq i'$$

i.e. que la base $(f_i, i=1, \dots, n)$ permanezca ortogonal.

- La forma cuadrática asociada D_p

$$D_p(f_i) = D_p(f_i, f_i) = \# f_i^2 = \rho_i$$

i.e. que le asigne el peso que le corresponde a cada individuo, por ello es que la matriz asociada D_p es:

$$\begin{bmatrix} \rho_1 & 0 & \dots & 0 \\ 0 & \rho_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \rho_n \end{bmatrix}$$

- Puesto que $\rho_i > 0 \forall i=1, \dots, n$ $\text{Det } D_p = \prod_{i=1}^n \rho_i > 0$ y por tanto

D_p es invertible, con inversa $D_{1/p}$. Esto es, puede hacer el análisis con individuos de igual importancia y después transformar con D_p .

- Se recuerda al lector la convención en denotar indiferentemente por D_p a:

-el isomorfismo entre F^* y F .

- la forma bilineal.

- la forma cuadrática.

- La forma cuadrática proporciona la norma inducida con esta métrica :

$$D_p(C^j, C^j) = D_p(C^j) = \|\cdot\|_{D_p}^2 = \sum_{i=1}^n x_i^2 \rho_i = \text{var}(C^j).$$

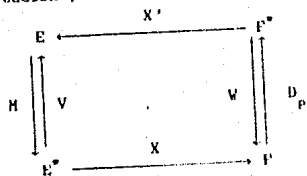
Con esta norma se define la métrica entre características ponderadas por ρ_k :

$$d_{D_p}^2(C^j, C^j) = \|C^j - C^j\|_{D_p}^2 = (\text{var}(C^j - C^j))^{1/2}$$

$$d_{D_p}^2(C^j, C^j) = \sum_{k=1}^n \rho_k (x_k^j - x_k^j)^2$$

$$d_{D_p}^2(C^j, C^j) = \text{var}(C^j) + \text{var}(C^j) - 2 \text{cov}(C^j, C^j).$$

En estas condiciones el esquema de dualidad adquiere la siguiente representación :



Observaciones :

- El isomorfismo de F a F^* se describe explícitamente por

$$D_p(C_j) = D_p(C_j) = (C_j)' D_p$$

- En la primera expresión D_p es isomorfismo, en la segunda D_p es forma bilineal y en la tercera D_p es la matriz de pesos.

- Asocia una aplicación lineal de E^* a E definida

$$Y = X' D_p^{-1} X$$

tal que :

$$V(C_j, C_j) = \text{Cov}(C_j, C_j)$$

$$V(C_j) = V(C_j, C_j) = (C_j)' V C_j = \text{var}(C_j)$$

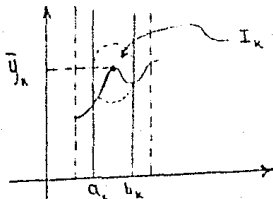
En la primera expresión V es forma cuadrática en la segunda V es forma bilineal y en la tercera V es la matriz de covarianzas del vector μ correspondiente a un individuo en n realizaciones afectadas por probabilidades. En otras palabras la matriz V se escribe como :

$$V = \sum_{i=1}^n \rho_i C_i C_i'$$

3. / Relación de correlación.

La relación de correlación permite juzgar en que medida se puede reconstruir una característica cuantitativa Z , a partir de una característica cuantitativa X , (incluso cualitativa) esto es, permite medir el grado de fineza de X en relación lineal con Z .

Para calcular el índice se recorta el intervalo de variación de las x_i en un número finito de intervalos disjuntos, digamos en q intervalos disjuntos



En el presente análisis se utiliza la notación siguiente :
Se define :

$$I_k = \{ i \mid a_k \leq x_i < b_k \}$$

$$\bar{y}_k = \sum (\rho_i y_i \mid i \in I_k)$$

$$\rho_k = \sum (\rho_i \mid i \in I_k) = P(I_k)$$

$$\bar{y}_k = \frac{1}{\rho_k} \sum (\rho_i y_i \mid i \in I_k)$$

donde I es el conjunto de los individuos, un subconjunto de índices finito correspondiente al número de individuos.

Se muestra sin dificultad que :

$$\sum_i (y_i - \bar{y})^2 \mid i \in I = \sum_{k=1}^q \sum_i (y_i - \bar{y}_k)^2 \mid i \in I_k + \sum_{k=1}^q (\bar{y}_k - \bar{y})^2$$

donde la suma $\sum (\rho_i (y_i - \bar{y})^2 \mid i \in I_k)$ se llama la suma total.

la suma $\sum_{k=1}^q \sum_i (\rho_i (y_i - \bar{y}_k)^2 \mid i \in I_k)$ suma intra y la suma

$\sum_{k=1}^q (\bar{y}_k - \bar{y})^2$ se le conoce como suma inter.

En efecto :

$$\begin{aligned} \sum \{ \rho_i (y_i - \bar{y})^2 \mid i \in I \} &= \sum \{ \rho_i (y_i - \bar{y}_k + \bar{y}_k - \bar{y})^2 \mid i \in I \} \\ &= \sum \{ \rho_i (y_i - \bar{y}_k)^2 + \rho_i (\bar{y}_k - \bar{y})^2 + 2 \rho_i (y_i - \bar{y}_k) (\bar{y}_k - \bar{y}) \mid i \in I \} \end{aligned}$$

donde $I = \bigcup_{k=1}^q I_k$ por lo cual se verifica la igualdad :

$$\begin{aligned} &= \sum_{k=1}^q \sum_i \{ \rho_i (y_i - \bar{y}_k)^2 \mid i \in I_k \} + \sum_{k=1}^q \sum_i \{ \rho_i (\bar{y}_k - \bar{y})^2 \mid i \in I_k \} \\ &\quad + 2 \sum_{k=1}^q \sum_i \{ \rho_i (y_i - \bar{y}_k) (\bar{y}_k - \bar{y}) \mid i \in I_k \} \end{aligned}$$

$$= \sum_{k=1}^q \sum_i \{ \rho_i (y_i - \bar{y}_k)^2 \mid i \in I_k \} + \sum_{k=1}^q (y_k - \bar{y})^2 \sum_i (\rho_i \mid i \in I_k)$$

Puesto que :

$$\begin{aligned} \sum_{k=1}^q \sum_i \rho_i (y_i - \bar{y}_k) (\bar{y}_k - \bar{y}) &= \sum_{k=1}^q (\bar{y}_k - \bar{y}) \left\{ \sum_{i \in I_k} \rho_i y_i - \bar{y} \sum_{i \in I_k} \rho_i \right\} \\ &= \sum_{k=1}^q (\bar{y}_k - \bar{y}) \left\{ \rho_k \cdot \frac{1}{\rho_k} \sum_{i \in I_k} \rho_i y_i - \bar{y}_k \rho_k \right\} \end{aligned}$$

donde $\bar{y}_k = \frac{1}{\rho_k} \sum_{i \in I_k} \rho_i y_i$, por lo tanto

$$= \sum_{k=1}^n (\bar{y}_k - \bar{y}) \{ \rho_1 \bar{y}_k - \rho_2 \bar{y}_k \} = 0$$

El conjunto de valores que asume una característica puede ser ordenado o no ordenado, ejemplos de características cualitativas con conjunto de respuestas :

Ordenado : La característica nivel escolar
primaria < secundaria < ... < posgrado.

No ordenado : La característica grupo racial
negroide, mongoloide, caucasoide , australoide.

Si se tienen dos características x y x' cualitativas con conjuntos de respuestas posibles Q y Q' no ordenables incluso, se dice que x es más fino que x' si existe una aplicación φ de Q en Q' tal que $x = \varphi \cdot x'$, i.e. que $x_i = x'_j \rightarrow x_i = x'_j \quad \forall i, j \in I$, i.e. x permite reconstruir exactamente a x' .

En efecto : Sea I el conjunto de individuos, Q y Q' los conjuntos de respuestas (también llamados de modalidades) de las características x y x' .

Si $x_i = x'_j + x'_k = x'_l \quad \forall i, j, k, l \in I$ entonces :

$$x^{-1}(q_i) = x^{-1}(x'_j) \subseteq x^{-1}(x'_l)$$

i.e. cada $k \in x^{-1}(x'_l)$ se cumple que $x(k) = x'_l$.

Se observa que $x^{-1}(q_i) \neq \emptyset$ y que para cada elemento k de la clase de equivalencia $I(i) = \{k \in I \mid x(k) = x'_l\} = q_i$ es posible definir :

$$x(k) = x(x^{-1}(q_i)) = \varphi(q_i)$$

obteniendo con ello una función φ en términos de x y x' tal que el diagrama conmuta :



En efecto : $x(k) = x(x^{-1}(q_i)) = \varphi(q_i) = \varphi(x(k)) = \varphi \cdot x(k)$
 $k \in I \quad \dots \quad x = \varphi \cdot x'$

inversamente si el diagrama (1) conmuta , todos los pares de elementos $j, l \in I$ tales que $x(j) = x(l)$ cumplen $\varphi \cdot x(j) = \varphi \cdot x(l)$ y por ello $x'(j) = x'(l)$.

Observación :

-Si Q y Q' son tales que $\#Q = \#Q' = r$, de la definición de φ se sabe que $\varphi(q_i) = x(x^{-1}(q_i))$, puesto que todos los valores de Q' se adoptan por los individuos I bajo x se tiene también que $\#(x(x^{-1}(q_i))) : \#I = r$, por ello es que $\#(\varphi(q_i)) : \#I = r$, por tanto el número de imágenes que asume la función φ es r y además $\varphi : Q \rightarrow Q'$ con $\#Q = r$, φ es biyección.

Si además los conjuntos de respuestas Q y Q' son ordenables entonces es posible definir la relación de fineza en términos de la relación de orden : x es más fina que x' si :

$$x'_i \leq x'_j \rightarrow x_i \leq x_j \quad \forall i, j \in I$$

$$\text{ó } x'_i \leq x'_j \rightarrow x_i \leq x_j$$

Hay también características cuyo conjunto de posibles respuestas es ordenable que además es cuantitativo, cuando estas propiedades se tienen en las características x y x' , se dice que x es más fino que x' si existe una función continua φ tal que $x = \varphi \cdot x'$, estas funciones pueden ser lo suficientemente regulares.

En este contexto se puede definir la relación de correlación entre las características x y x' como :

$$\rho_{xz} = \frac{\text{inter}}{\text{total}}$$

para $\text{inter} \neq \text{intra} \neq \text{total} \neq 0$.

De la relación $\text{inter} + \text{intra} = \text{total}$

por tanto :

$$\frac{\text{inter}}{\text{total}} + \frac{\text{intra}}{\text{total}} = 1$$

con ambos sumandos no negativos :

$$0 \leq \rho_{xz} \leq 1$$

permite medir la relación de las x_i a partir de las z_i a través de una función. Cuando el valor se acerca a 1 se tiene el indicio de la existencia de una tal función f que permite recuperar una característica a partir de la otra.

4. Análisis en Componentes Principales para una matriz de datos.

4.1. Ejes y planos principales.

Todo punto del se escribe de manera única como combinación lineal de los elementos de la base K -ortonormal $\{u_i; i=1, \dots, p\}$ de los eigenvectores de M , dicha base se puede construir porque la matriz asociada a M es simétrica y por los teoremas de descomposición espectral (apéndice 3) se sigue que :

$$M = \Gamma \Lambda \Gamma'$$

donde $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$ y Γ es una matriz ortogonal, i.e. Λ genera vectores ortogonales que se hacen unitarios si

$$\frac{1}{\|u_i\|_M} u_i = u_i'$$

El primer eje principal es el eje $\langle u_1 \rangle$ generado por vector unitario que proporciona la máxima $J_{\langle u_1 \rangle}$.

$$\text{i.e. } \|u_1\|_M^2 = M(u_1) = M\left[\sum_{i=1}^p \alpha_i u_i, \sum_{i=1}^p \alpha_i u_i\right] = 1$$

y también :

$$J_{\langle u_1 \rangle} = M(VH(u_1)) = M\left[VH\left[\sum_{i=1}^p \alpha_i u_i, \sum_{i=1}^p \alpha_i u_i\right]\right]$$

máxima, que por la bilinealidad de M y el hecho que la base de los eigenvectores es M -ortonormal, se deduce :

$$\|u_1\|_M^2 = \sum_{i=1}^p \alpha_i^2 = 1.$$

$$J_{\langle u_1 \rangle} = \sum_{i=1}^p \alpha_i^2 \lambda_i$$

En efecto :

$$1 = \mathbf{u}_i^T \mathbf{u}_i = \mathbf{H} \left(\sum_{i=1}^p \alpha_i^2 \mathbf{u}_i \right) \cdot \sum_{i=1}^p \alpha_i^2 \mathbf{u}_i$$

$$= \sum_{i=1}^p \alpha_i \mathbf{H} \left(\sum_{j=1}^p \alpha_j^2 \mathbf{u}_j \right) = \sum_{i=1}^p \alpha_i \alpha_j \delta_{ij} = \sum_{i=1}^p \alpha_i^2$$

$$I_{\langle \mathbf{u}_i \rangle} = \frac{H(\mathbf{u}_i, \mathbf{VH} \mathbf{u}_i)}{H(\mathbf{u}_i, \mathbf{u}_i)} = \mathbf{H} \left(\sum_{i=1}^p \alpha_i^2 \mathbf{u}_i, \mathbf{VH} \left(\sum_{i=1}^p \alpha_i^2 \mathbf{u}_i \right) \right)$$

$$= \mathbf{H} \left(\sum_{i=1}^p \alpha_i^2 \mathbf{u}_i, \sum_{i=1}^p \alpha_i \mathbf{VH} \mathbf{u}_i \right)$$

$$= \sum_{i=1}^p \alpha_i \alpha_j \delta_{ij} H(\mathbf{u}_i, \mathbf{VH} \mathbf{u}_i)$$

$$= \sum_{i=1}^p \alpha_i \alpha_j \delta_{ij} H(\mathbf{u}_i, \lambda_j \mathbf{u}_i)$$

$$= \sum_{i=1}^p \alpha_i \alpha_j \delta_{ij} \lambda_j H(\mathbf{u}_i, \mathbf{u}_i) = \sum_{i=1}^p \alpha_i^2 \lambda_i$$

De este resultado se obtiene una equivalencia para la restricción $\mathbf{u}_i^T \mathbf{u}_i = 1$ que es:

$$\sum_{i=1}^p \alpha_i^2 = 1$$

en el problema de optimización siguiente:

Proposición: Con la restricción anterior y si $\lambda_1 \geq \dots \geq \lambda_p \geq 0$ se

tiene que $\sum_{i=1}^p \alpha_i^2 \lambda_i \leq \lambda_1$.

Demostración:

Puesto que $\alpha_i^2 \lambda_i \geq \alpha_i^2 \lambda_j \quad \forall i=1, \dots, p$

$$\lambda_1 = \lambda_1 \sum_{i=1}^p \alpha_i^2 = \sum_{i=1}^p \alpha_i^2 \lambda_1 \geq \sum_{i=1}^p \alpha_i^2 \lambda_i$$

Donde se indica que $I_{\langle \mathbf{u}_i \rangle} = \sum_{i=1}^p \alpha_i^2 \lambda_i$ es máximo bajo la

restricción $\mathbf{u}_i^T \mathbf{u}_i = 1$ si $\alpha_1 = \alpha_2 = \dots = \alpha_p = 0$, esto es, si \mathbf{u}_i coincide con el primer eigenvector unitario de \mathbf{VH} con eigenvalor asociado λ_1 .

El problema de optimización tiene por objetivo explicar la mayor parte de la inercia, en otras palabras, la dispersión de la representación de los individuos, a través de este resultado se garantiza que si se elige la dirección generada por \mathbf{u}_1 el primer eigenvector de \mathbf{VH} se explica la inercia λ_1 , que no es la inercia total, siguiendo el mismo procedimiento se encuentra la siguiente dirección correspondiente a \mathbf{u}_2 que explica λ_2 , dirección \mathbf{H} -ortogonal a $\langle \mathbf{u}_1 \rangle$, se sigue el mismo procedimiento hasta explicar el total de la inercia o un porcentaje que se considere suficiente.

El primer eje principal es el eje $\langle \mathbf{u}_1 \rangle$ generado por \mathbf{u}_1 , el eigenvector unitario de \mathbf{VH} asociado al eigenvalor más grande λ_1 .

i.e. $\mathbf{VH} \mathbf{u}_1 = \lambda_1 \mathbf{u}_1$ entonces $I_{\langle \mathbf{u}_1 \rangle} = \sum_{i=1}^p \alpha_i^2 \lambda_i = \lambda_1$ es la máxima inercia, la parte de la inercia explicada por \mathbf{u}_1 .

La inercia que no es explicada por \mathbf{u}_1 es:

$$I_{\langle \mathbf{u}_1 \rangle^\perp} = \text{tr}(\mathbf{VH}) - \lambda_1$$

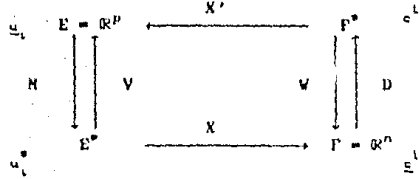
$$- \text{Si } \mathbf{H} \in \langle \mathbf{u}_1 \rangle \Leftrightarrow I_{\langle \mathbf{u}_1 \rangle} = 0 \Leftrightarrow \text{tr}(\mathbf{VH}) = \lambda_1 \Leftrightarrow \frac{\lambda_1}{\text{tr}(\mathbf{VH})} = 1$$

en cuyo caso la inercia acumulada es el 100% en la dirección $\langle \mathbf{u}_1 \rangle$.

$$- \text{Si } \mathbf{H} \notin \langle \mathbf{u}_1 \rangle \Leftrightarrow I_{\langle \mathbf{u}_1 \rangle^\perp} > 0 \Leftrightarrow \text{tr}(\mathbf{VH}) > \lambda_1 \Leftrightarrow \frac{\lambda_1}{\text{tr}(\mathbf{VH})} < 1$$

el índice estandarizado $\frac{\lambda_1}{\text{tr}(\mathbf{VH})}$ es la parte de la inercia explicada por el primer eje principal $\langle \mathbf{u}_1 \rangle$.

- En el esquema de dualidad :



A u_1 se le llama la primera componente principal. Para la

cual existe único $u_1^* = (C_1)'M u_1 \in E^*$, para lo cual $V((C_1)'M u_1) \in F$.

Al eje principal $\langle u_1 \rangle$, simbolizado en E como una nueva característica (una nueva característica que es combinación lineal de las características originales), está asociado ;

-En E^* un funcional lineal principal conocido como el "primer factor principal".

$$u_1^* = (C_1)'M u_1 = H(u_1)$$

-En F, la "primera componente principal" ;

$$c_1^* = X(u_1^*) = X((C_1)'M u_1)$$

donde $c_1^* = H(c_1, u_1) = (C_1)'M u_1 = X((C_1)'M u_1) = X((C_1)'M u_1)$ y

$c_1^* = H(c_1, u_1) = u_1^*(c_1)$ es por tanto el valor obtenido por la primera componente principal para el individuo i .

Observaciones ;

$-\sum_{i=1}^n p_i c_i^* = U_p(u_1^*) = 0$ donde $U = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in F$ indica que la primera componente principal está centrada en el origen.

$-\sum_{i=1}^n p_i (c_i^*)^2 = D_p(c_1^*) = V(c_1^*) = MVH(u_1) = \lambda_1$ esto es, la primera componente principal es λ_1 .

El plano principal P es el subespacio principal es el subespacio vectorial de dimensión 2 con momento de inercia mínimo.

El plano principal puede ser generado por ;

-El eje principal $\langle u_1 \rangle$.

-Y el eje $\langle v \rangle$, generado por el vector unitario v , H-ortogonal a u_1 , con momento de inercia mínimo.

La búsqueda del plano principal P coincide aquí con la búsqueda del eje $\langle v \rangle$ produciendo $l_{(v)\perp} = MVH(u_1)$ máxima, bajo las restricciones ;

$$\begin{cases} H(u_1)^2 = H(u_1) = 1 \\ H(c_1, u_1) = 0 \end{cases}$$

De este modo la base de los eigenvectores $\langle u_i \rangle; (i=1, \dots, p)$ es entonces elegida en E, y se obtiene la siguiente ;

Proposición : Si $u = \sum_{i=1}^p \alpha_i u_i \rightarrow$

$$\begin{cases} (i) l_{(u)\perp} = \sum_{i=1}^p \alpha_i^2 \lambda_i \\ (ii) H(c_1) = \sum_{i=1}^p \alpha_i^2 = 1 \\ (iii) H(c_1, u_1) = \alpha_1 = 0. \end{cases}$$

Demostración :

Basta probar la igualdad (iii), en efecto ;

$$\alpha_1 = \alpha_1 H(c_1, u_1) = \sum_{i=1}^p \alpha_i H(c_1, u_i) = H \left\{ \sum_{i=1}^p \alpha_i u_i, u_1 \right\} = H(c_1, u_1) = 0$$

De donde se infiere que $l_{(u)\perp} \leq \lambda_2$ y $l_{(u)\perp}$ es máximo si $\alpha_2=1$

, $\alpha_2 = \alpha_3 = \dots = \alpha_p = 0$, i.e. u coincide con alguno de los eigenvectores unitarios de VH con eigenvalor λ_2 .

-El segundo eje principal es el eje $\langle u_2 \rangle$ generado por el eigenvector unitario u_2 de VH con eigenvalor λ_2 ;

Así : $I_{(u_2)^\perp} = \lambda_2$

$$I_{P^\perp} = I_{(u_1)^\perp} + I_{(u_2)^\perp} = \lambda_1 + \lambda_2$$

$$I_P = \text{tr}(VH) - (\lambda_1 + \lambda_2).$$

- Si $\mathcal{M} \subset P \Leftrightarrow I_P = 0 \Leftrightarrow \frac{\lambda_1 + \lambda_2}{\text{tr}(VH)} = 1.$

El índice $\frac{\lambda_1 + \lambda_2}{\text{tr}(VH)}$ es la parte de la inercia explicada por el plano principal P.

Al eje principal $\langle u_2 \rangle$, simbolizado en E como una nueva característica (que que es combinación lineal de las características originales), está asociado :

-En E^* un funcional lineal principal (a esta función se le conoce como "segundo factor principal").

$$u_2^* = (.)' H u_2 = H(u_2).$$

-En P, la "segunda componente principal":

$$c_2^* = X(u_2^*) = X((.)' H u_2),$$

4.3. / Subespacios principales de dimensión l.

El subespacio principal de dimensión l es el subespacio vectorial de E que es más próximo a la nube \mathcal{M} en el sentido del momento de inercia, es el subespacio vectorial generado por los l primeros ejes principales :

$$\langle u_1 \rangle, \dots, \langle u_l \rangle ;$$

Donde el eje $\langle u_i \rangle$ es el generado por el vector unitario u_i de VH cuyo eigenvalor asociado es λ_i :

$$VH u_i = \lambda_i u_i \quad i=1, \dots, l.$$

Y es tal que $I_{(u_i)^\perp} = \lambda_i \quad i=1, \dots, l.$

Por ello se sigue que :

$$I_{P^\perp} = \sum_{i=1}^l I_{(u_i)^\perp} = \sum_{i=1}^l \lambda_i$$

Por tanto $I_S = \text{tr}(VH) - \sum_{i=1}^l \lambda_i$

El índice $\frac{\sum_{i=1}^l \lambda_i}{\text{tr}(VH)}$ es la parte de la inercia explicada por el subespacio S.

$$- \text{Si } \mathcal{M} \subset S \Leftrightarrow I_S = 0 \Leftrightarrow \frac{\sum_{i=1}^l \lambda_i}{\text{tr}(VH)} = 1.$$

A los ejes principales $\langle u_i \rangle$, simbolizados en E como una nueva característica (que es combinación lineal de las originales), que tiene asociados :

- En E^* un funcional lineal principal (a esta función se le conoce como el "i-ésimo factor principal").

$$u_i^* = (.)' H u_i = H(u_i).$$

-En P, la "i-ésima componente principal" :

$$c_i^* = X(u_i^*) = X((.)' H u_i),$$

donde la componente principal c_i^* está centrada y acumula inercia λ_i .

Esto es :

$$\sum_{i=1}^n \rho_i c_i^* = D_p (0, \dots, 0) = 0 \quad \text{donde } 0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\text{y} \quad \sum_{i=1}^n \rho_i (c_i^*)^2 = D_p (c_i^*) = \lambda_j$$

$$\text{si } \lambda_i \neq 0 \rightarrow \frac{1}{\lambda_i} = V^{-1}(u_i, u_i).$$

□

A los ejes principales $\langle u_i \rangle \in E$ corresponden ;

-En E^* los factores principales $u_i^* = (.)^* H u_i$, que son eigenvectores de VH con eigenvalor λ_i . HV es H y V simétrica y por tanto $\langle u_i; i=1, \dots, p \rangle$ es una base de E , H^{-1} -ortonormal y V -ortonormal.

Adh más (1) \leftrightarrow (2)

$$HV u_i^* = \lambda_i u_i^* \quad H^{-1}(u_i^*, u_i^*) = 1 \quad , \quad H^{-1}(u_i^*, u_j^*) = 0 \quad (i \neq j)$$

$$V(u_i^*, u_i^*) = 1 \quad ; \quad V(u_i^*, u_j^*) = 0 \quad (i \neq j)$$

$$VH u_i = \lambda_i u_i \Leftrightarrow HVH u_i = \lambda_i H u_i \Leftrightarrow HV u_i^* = \lambda_i u_i^*.$$

$$\delta_{ij} = H(u_i, u_j) = H(u_i, u_j) = HCH^{-1}H u_i H^{-1}H u_j$$

$$= H^{-1}HCH u_i H u_j$$

$$= H^{-1}(H u_i, H u_j)$$

$$= H^{-1}(u_i, u_j)$$

□

-En F las componentes principales :

$$\xi^j = X u_i^* = D^j H u_i.$$

$$MV u_i^* = \lambda_i u_i^* \rightarrow X \cdot H \cdot X^{-1} \cdot D_p \cdot X u_i^* = \lambda_i X u_i^*$$

lo cual implica :

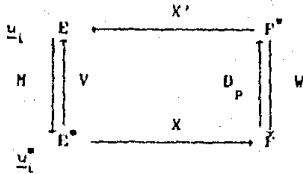
$$W \cdot D_p \xi^j = \lambda_j \xi^j.$$

-En F las componentes principales son eigenvectores de $V \cdot D_p$ con eigenvalor λ_j ; por otra parte :

$$D_p(\xi^j, \xi^j) = V(u_j, u_j)$$

4.4. Construcción y dualidad.

Sea $M \in S \subset E$, en el esquema :



$u_i^* = (.)^* H u_i$, se tiene que $X(u_i^*) = X((.)^* H u_i) = c^i$.

Se considera a E con la base ortonormal $\langle u_i; i=1, \dots, p \rangle$ formada a partir de los eigenvectores de VH .

$$1) \quad VH u_i = \lambda_i u_i.$$

$$2) \quad H(u_i, u_j) = \begin{cases} 1 & (i=j) \\ 0 & (i \neq j) \end{cases}.$$

En caso que V sea invertible, la base $\langle u_i; i=1, \dots, p \rangle$ es

además V^{-1} -ortonormal :

$$V^{-1}(u_i, u_j) = \frac{1}{\lambda_j} \quad V^{-1}(u_i, u_i) = 0 \quad (i \neq j)$$

$$\text{En efecto : } 1 = VV^{-1}(u_i, u_j) = V^{-1}(VH u_i, u_j) = V(\lambda_i u_i, u_j)$$

$$= \lambda_i V^{-1}(u_i, u_j)$$

$$V_i \| \begin{matrix} i \\ p \end{matrix} \|^2 = \lambda_j^{-1} \quad \| \begin{matrix} i \\ p \end{matrix} \|_i^2 = \begin{cases} \lambda_j^{-1} & j=j' \\ 0 & j \neq j' \end{cases}$$

Observaciones sobre los rangos :

Si q es el número de eigenvalores no nulos λ de VM (se puede

tener que $\lambda_j = \lambda_{j'} \quad j=j'$) se muestra que :

$$q = \text{rank}(X') = \text{rank}(X) = \text{rank}(V) = \text{rank}(W) = \text{rank}(VM) \\ = \text{rank}(MV) = \text{rank}(WD_p) \quad (3).$$

Demostración :

Puesto que $\text{rank}(L) = \text{rank}(L')$ solo hace falta probar la igualdad para alguna de estas matrices, además VM es M-simétrica entonces el rango de VM y MV es el mismo, por otra parte si la matriz VM tiene q eigenvalores no nulos implica que la dimensión de la imagen de VM (como transformación lineal) es precisamente q , finalmente se observa que :

$$\begin{aligned} \text{i)} \quad \text{rank}(V) &= \min(\text{rank}(X'), \text{rank}(D_p), \text{rank}(X)) \\ &= \min(\text{rank}(X), \text{rank}(D_p)) = \min(\text{rank}(X), n) \\ &= \text{rank}(X) \\ \text{ii)} \quad \text{rank}(W) &= \min(\text{rank}(X'), \text{rank}(M), \text{rank}(X)) \\ &= \min(\text{rank}(X), \text{rank}(M)) = \min(\text{rank}(X), p) \\ &= \text{rank}(X) \\ \text{iii)} \quad \text{rank}(WD_p) &= \min(\text{rank}(W), \text{rank}(D_p)) \\ &= \min(\text{rank}(X), n) \\ &= \text{rank}(X) \\ \text{iv)} \quad \text{rank}(VM) &= \min(\text{rank}(V), \text{rank}(M)) \\ &= \min(\text{rank}(X), p) \\ &= \text{rank}(X) \end{aligned}$$

de donde se concluye inmediatamente el resultado.

4.8. Descripción de la nube de individuos.

Si el núcleo de VM es de dimensión $p-q$, esto es, si el subespacio propio de VM de eigenvalor asociado $\lambda=0$ es de dimensión $(p-q)$, se tiene :

$$\text{tr}(VM) = \sum_{i=1}^q \lambda_i$$

La nube de puntos \mathcal{A} está entonces situada en el subespacio S de dimensión q generado por los q primeros ejes principales ;

A S se le llama soporte de la nube \mathcal{A} .

Por esta razón la imagen euclídeana (\mathcal{A}, S, d_v) del conjunto de los individuos I , donde d_v designa la restricción de la métrica \mathcal{A} al subespacio vectorial S ($M|_S$), es entonces la imagen euclídeana más simple equivalente a la imagen euclídeana original (\mathcal{A}, E, d_M) . Esto ocasiona una reducción de datos al pasar de (\mathcal{A}, E, d_M) a (\mathcal{A}, S, d_v) .

En E un renglón de datos X (un individuo) está simbolizado por p números (los valores asociados a las p características).

En S un renglón de X está simbolizada por una lista de q números (los valores asociados a las q componentes principales).

Si la parte de la inercia explicada por el plano principal P , es próxima a 1, se puede con ayuda de las proyecciones \underline{a}_i de los

puntos x_i sobre el plano P dar una buena descripción de la nube X , que proporciona una visualización plana aproximada de X .

En la representación cartesiana (los ejes principales son ortogonales, los vectores u_1 y u_2 tienen norma unitaria, los puntos-individuos se recuperan con la ayuda de las dos primeras componentes principales, un análisis de la representación sobre este plano será suficiente para hacer el balance del conjunto de proximidades entre individuos, esas proximidades están medidas por las longitudes de los segmentos que separan los puntos-individuos.

-El cociente $\frac{\lambda_1 + \lambda_2}{\text{tr}(\text{VMD})}$ es la medida de la "calidad global" de la representación de la nube X en el plano principal P.

La calidad de representación de un punto x_i por su proyección α_i sobre el plano principal puede ser medida por la comparación de las normas :

$$\|x_i\|_M^2 = \sum_{j=1}^q H^2(c_{ij}, u_j) = \sum_{j=1}^q (c_j^i)^2$$

$$\|\alpha_i\|_M^2 = H^2(c_{i1}, u_1) + H^2(c_{i2}, u_2) = (c_1^i)^2 + (c_2^i)^2$$

y con el índice $\cos \alpha = \frac{\|\alpha_i\|}{\|x_i\|}$ (el coseno del ángulo que forman los vectores α_i y x_i), se interpreta geométricamente :

(Handwritten note: $\cos(\alpha) = \frac{\|\alpha_i\|}{\|x_i\|}$)

Este coseno se afecta en signo por el valor c_1^i que asume la

tercera componente principal.

Observación :

El índice $\cos \alpha$ tiene la ventaja que pertenece al intervalo (0,1), pero tiene la desventaja que no expresa la proximidad del punto x_i al plano principal P cuando x_i está demasiado alejado de P :

$$\|x_i\|^2 \rightarrow \alpha_i = 0 \rightarrow \|\alpha_i\|_M = 0.$$

Sin embargo el índice $\frac{\|\alpha_i - \alpha_i\|_M}{\sqrt{\text{tr}(\text{VMD}) - (\lambda_1 + \lambda_2)}}$ permite comparar

la distancia de x_i al plano P con la dispersión promedio alrededor del plano P ($\text{tr}(\text{VMD}) - (\lambda_1 + \lambda_2)$)^{1/2} es análoga a un tipo de desviación ; este índice es similar al coseno, afectado por el signo que tome la tercera componente principal.

El examen en el plano principal de la nube de las proyecciones α_i de los puntos x_i , teniendo en cuenta la calidad de las representaciones permite reagrupar en clases de individuos semejantes.

(Aún mejor el plano principal puede ser "orientado" de tal manera que figuren las proyecciones M-ortogonales de los ejes $\langle u_1 \rangle$ asociados a las características iniciales).

Esta representación simultánea de los individuos y de los características permite precisar las generalidades de los grupos de individuos formados.

4.6. / Individuos suplementarios.

Todo $x_i \in E$ puede proyectarse sobre el plano principal $\langle u_1, u_2 \rangle$ en particular. Así que cualquier individuo que se agregue al conjunto I puede ser identificado por su proyección.

Descripción de las características.

El análisis de componentes principales se presentó como la búsqueda sucesiva de características c, D -unitarias que maximizan :

$$\sum_{j=1}^p \sum_{p=1}^p n_{jj} D_p(c_j^T, D_p(c_j^T, c_j^T))$$

Las características son homogéneas a las componentes principales puesto que son contracciones y expansiones.

El conjunto de las componentes principales $(\xi^j; j=1, \dots, q)$ forman una base D_p -ortogonal del subespacio vectorial $X(\mathbb{R}^p)$ en P generado por el conjunto de características iniciales.

$$D_p(\xi^1, \xi^j) = 0 \quad \text{si } (j \neq 1)$$

$$D_p(\xi^1) = D_p(\xi^1, \xi^1) = \lambda_1.$$

Por tanto toda característica ξ^j puede ser representada en esas coordenadas ;

$$\frac{D_p(\xi^j, \xi^j)}{\sqrt{\lambda_j}} = \frac{\text{Cov}(\xi^1, \xi^j)}{\sqrt{\lambda_j}} \quad j = 1, \dots, q$$

en la base de las $\frac{1}{\sqrt{\lambda_j}} \xi^j$.

A la representación de los puntos-individuos en el plano principal P está asociada la representación de los puntos-características en el plano $\langle \xi^1, \xi^2 \rangle$.

La característica ξ^j interviene tanto más en la descripción de la nube X por su proyección en el plano principal que ;

$$\| \xi^j \|^2_M = \sum_{j=1}^q D_p(\xi^j, \xi^j)$$

es cercano de ;

$$- \frac{D_p(\xi^j, \xi^1)}{\lambda_1} + \frac{D_p(\xi^j, \xi^2)}{\lambda_2}$$

Por ejemplo ξ^1 designa la proyección D_p -ortogonal de la característica ξ^1 sobre el plano $\langle \xi^1, \xi^2 \rangle$.

La interpretación de las componentes principales ξ^1 y ξ^2 se hace en vista de las proximidades de las proyecciones de las características ξ^j y las componentes ξ^1 y ξ^2 , en el plano $\langle \xi^1, \xi^2 \rangle$ donde los ejes D_p -ortogonales $\langle \xi^1, \xi^2 \rangle$ son representados ortogonalmente.

En el plano, se hace evidente si las normas están bien construidas ;

- las proximidades entre las características ξ^j entre las características ξ^1, ξ^1, ξ^2 .
- las ortogonalidades ; (si siguen siendo ortogonales) entre características ξ^j entre características ξ^1, ξ^1, ξ^2 .

Características suplementarias.

Todo elemento del espacio vectorial de las características P puede proyectarse en el plano $\langle \xi^1, \xi^2 \rangle$, en particular, es posible que aparezca en dicho plano a las características que no se tienen en cuenta en la búsqueda del plano principal, por ello basta evaluar visualmente las proximidades entre esas "características suplementarias" y las otras características.

/ EJEMPLO :

Los iris de Fisher.

Este ejemplo es clásico en estadística multivariada y pretende exhibir concretamente a través de los diferentes tratamientos que se manejan cuando se hace un ACP.

I. / Objetivos.

En el ejemplo considerado aquí, para el cuadro de estudio morfométrico de flores en siembra a los que se muestreo de un plantío considerando 50 elementos de cada tipo de iris (son tres tipos de iris los manejados aquí), se desea dar elementos para responder a las preguntas :

-En los tres tipos de iris hay una manera de caracterizar sus medidas morfométricas a partir del pétalo y del sépalo?

-La talla del iris depende del tipo de iris y de la relación entre las medidas morfométricas entre el pétalo y el sépalo?

El ACP no puede ser considerado más que como un análisis permitido para explorar los datos.

I.1. / Experiencia .

Las 150 flores fueron analizadas (50 de cada tipo) :

A1 : contiene las flores de 1 a 50.

A2 : contiene las flores de 51 a 100.

A3 : contiene las flores de 101 a 150.

Las características de talla que se desean analizar son :

- 1.-Longitud sépalo.
- 2.-Ancho sépalo.
- 3.-Longitud pétalo.
- 4.-Ancho pétalo.

Las 150 flores son descritas con la ayuda de 4 características (los datos fueron obtenidos para 3 tipos de iris), se dispone de :

-Una nube X de 150 puntos-individuos en $E = \mathbb{R}^4$.

-Una nube X' de 4 puntos-características supuestos centrados en $F = \mathbb{R}^{150}$.

Para dar la misma importancia a las características en el cálculo de las distancias entre flores, se tiene un precio para la métrica en E , la métrica D_{1/σ^2} de matriz :

$$D_{1/\sigma^2} = \begin{bmatrix} 1/\sigma_1^2 & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & 1/\sigma_4^2 \end{bmatrix}$$

donde σ_j^2 es la varianza de la característica j .

Así por ejemplo la distancia entre las flores 1 y 2

$$d_{12} = \left[\left(\frac{5.1-4.9}{\sigma_1} \right)^2 + \dots + \left(\frac{0.2-0.2}{\sigma_4} \right)^2 \right]^{1/2}$$

$$\text{ó} \quad d_{12}^2 = (x_1 - x_2)' D_{1/\sigma^2} (x_1 - x_2).$$

La fórmula :

$$\sum_{i=1}^n \sum_{j=1}^n \left(\frac{x_i^j - x_j^i}{\sigma_j} \right)^2 = (150)^2 = 22500$$

permite verificar, las sumas de los cuadrados de medidas desechadas sobre cada una de las características que son iguales, que las características tengan pesos idénticos en el cálculo de las distancias entre flores.

"Reducir" la característica x^j requiere de efectuar el cambio siguiente :

$$E \xrightarrow{D_{1/\sigma}} E$$

$$x_i \xrightarrow{\quad\quad\quad} z_i = D_{1/\sigma} x_i$$

-la flor "i" está entonces representada por el vector z_i cuyas coordenadas son los números x_i^j / σ_j .

la característica "j" está entonces representada por el vector $z_j = 1/\sigma_j x_j^i$.

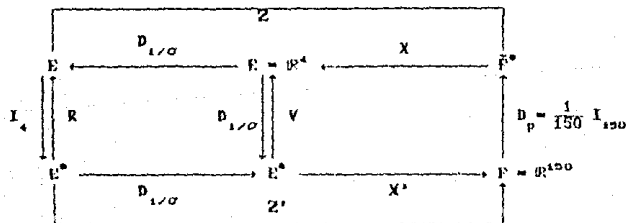
La igualdad :

$$\alpha_{ij}^2 = \sum_{j=1}^4 \left\{ \frac{x_i^j - \bar{x}_j}{\sigma_j} \right\}^2 = \sum_{j=1}^4 (z_{ij} - \bar{z}_j)^2$$

muestra que las dos imágenes euclidianas son equivalentes $(A, E, D_{1/\sigma})$ y $(D_{1/\sigma}, A, E, I_4)$.

El ACP se hace siempre sobre las características centradas ; X designa en lo sucesivo la tabla centrada de datos ;

Analizar la nube X en E dotado de la métrica $D_{1/\sigma}^2$ requiere analizar $D_{1/\sigma}(X)$, es decir la nube de los vectores z_i , en E provisto de la métrica euclídea clásica I_4 .



En E el coseno del ángulo entre z^j y $z^{j'}$, que es el coeficiente de correlación entre las características x^j y $x^{j'}$, no

es otra que $\rho_j(x^j, x^{j'})$.

La forma cuadrática de la inercia de la nube de puntos z que admite por matriz :

$$R = D_{1/\sigma} V D_{1/\sigma}$$

donde R es la matriz de correlación asociada al conjunto de las características x^j .

Los ejes principales generados por los eigenvectores de la matriz de correlación R.

Un vistazo a las correlaciones muestra que las características x^3 y x^4 están altamente correlacionadas (correlaciones del orden de 0,9628) de igual manera las características x^1 , x^2 y x^4 (correlaciones de 0,8707 y 0,8162 respectivamente).

Las características de longitud de sépalo y pétalo están altamente correlacionadas lo cual no resulta tan evidente entre las características de ancho de sépalo y pétalo.

1.3. / Estudio de la matriz de correlación, eigenvalores y eigenvectores.

La parte de la inercia explicada por el primer eje aparece en "contribución a la varianza", es próximo a 72,96 %.

El plano principal explica 98,72 % de la inercia, la imagen obtenida en proyección de los individuos en el plano principal podrá ser considerada globalmente buena.

1.4. / Cosenos cuadrados de los ángulos entre las características y sus proyecciones en los diferentes subespacios.

Esos cosenos permiten en principio medir la calidad de la representación de cada una de las flores en el plano principal ; desafortunadamente cuando una flor es próxima al centro de gravedad, y es el caso de las flores ;

1062,1069,1085,1098,1092,1067,1097 e 1065, este índice no tiene sentido.

1.1, / Resultados y conclusiones.

1.2.1, / Análisis de las características.

Característica	media	Tipos descartados	coef.de variación
1	5.0460	0.0374	
2	3.0573	0.4344	
3	3.7580	1.7594	
4	1.1993	0.7597	

La matriz de correlaciones entre las características es :

	x^1	x^2	x^3	x^4
x^1	1.0000			
x^2	-.1229	1.0000		
x^3	.0707	-.4204	1.0000	
x^4	.0163	-.3661	.9629	1.0000

La suma de los eigenvalores no nulos 1.3060400 ;

eje	eigen-valor	porcentaje inercia	porcentaje acum.
1	.953469	72.96	72.96
2	.397303	32.76	95.72
3	.421246E-01	3.76	99.47
4	.687185E-02	.53	100.00

Los eigenvectores ortonormales son :

eje	x^1	x^2	x^3	x^4
1	-.5209	.2707	-.5802	-.9644
2	.3752	.9239	.0267	.0693

Las correlaciones de las características con los dos primeros ejes :

eje	x^1	x^2	x^3	x^4
1	-.0099	.4624	-.0912	-.9644
2	.3900	.0015	.0255	.0661

1.3, / Representación de los individuos en el plano principal.

Las contribuciones absolutas de los individuos acordados con respecto a los dos primeros ejes aparecen en la Tabla no. 1. Aparecen también los cosenos cuadrados de los individuos acordados con respecto a los dos primeros ejes principales en la tabla no. 2.

La gráfica de los individuos acordados en el plano principal es generada por las siguientes coordenadas que aparecen en la tabla no 3. Por último aparece la gráfica de los individuos en el plano principal en la gráfica no 4.

Círculo de las correlaciones.

En la gráfica no. 5 aparece el círculo de las correlaciones de las características que intervienen en el estudio.

1.6, / Conclusión.

Un análisis visual de la representación gráfica en el plano principal permite asegurar que la nube X de los individuos está compuesta por tres subnubes correspondientes a los individuos de los tres tipos distintos de flores analizadas y estas nubes permiten diferenciar los individuos acordados entre los tres tipos de flores.

INTRODUCCION

En este capítulo se exponen :

- Una técnica conocida como Escalamiento multidimensional con la cual se obtiene una representación gráfica que ayude a interpretar relaciones entre los datos.
- Un procedimiento de selección de características concerniente al desarrollo de modelos de predicción usando las características correspondientes a un conjunto de datos.

La intuición geométrica arraigada principalmente en la geometría euclidiana permite interpretar relaciones entre los datos, en muchos casos dichas relaciones se pueden establecer a través de las distancias que separan entre sí a los individuos en estudio, sin embargo no siempre esto es posible y las relaciones obtenidas tienen propiedades más débiles que la función distancia. En este capítulo se expone un algoritmo con el cual se consigue una representación geométrica aún en estos casos. Por último se establece la dualidad entre ACP y el método de coordenadas principales para lo cual en todo este capítulo se trabaja con la matriz identidad como matriz de métrica ($B = D$) a menos que se especifique otra cosa.

Se presenta en este capítulo un procedimiento de selección de características con componentes principales para tratar el problema de multicolinealidad, se hace hincapié en el carácter subjetivo de la selección de las componentes principales y su significado dentro del problema.

1. Escalamiento multidimensional.

Esta técnica es útil para problemas en los que se requiere reconstruir una configuración de n puntos en un espacio euclidiano usando la información acerca de las distancias entre los n objetos (individuos). Previamente se han observado n puntos en un espacio de dimensión p los datos fueron extraídos directamente de los n puntos como valores de una función, en este caso la distancia.

Se puede reconstruir en base a distancias de ciudades en una guía de carreteras, un mapa geográfico de la República Mexicana, dado que la distancia de una ciudad a otra es la misma (salvo "pequeñas" perturbaciones) que la de x a y (ejemplo Distrito Federal-Queretaro).

En general las matrices de números que relacionan a los individuos entre sí no son distancias euclidianas y posiblemente no sean ni siquiera distancia según la definición del apéndice 1, pero aún en estos casos es posible representar las relaciones entre los individuos.

Se observa que la solución obtenida es única salvo transformaciones rígidas (traslaciones, rotaciones y reflexiones), esto es, para los puntos p_1, \dots, p_n , donde p_i tiene coordenadas (x_{i1}, \dots, x_{ip}) , $i=1, \dots, n$, se pueden identificar con (y_{i1}, \dots, y_{ip}) , $i=1, \dots, n$ por una isometría $\phi(_) = A_ + b$ donde A es una matriz ortogonal y $b \in \mathbb{R}^p$.

En general se busca obtener una representación gráfica que ayude a interpretar relaciones entre los datos, esto es, no se consideran técnicas probabilísticas sino solamente analíticas.

La pregunta obligada es: ¿Cómo se caracterizan las matrices que describen relaciones a través de distancias que son euclidianas?

1.1.1. Caracterización de matrices euclidianas.

Definición: $\mathbb{D} \in \mathbb{M}_{n \times n}(\mathbb{R})$ es una matriz de distancias si

$$(i) \mathbb{D} = \mathbb{D}^t$$

$$(ii) d_{ij} \geq 0 \quad \forall (i, j) \in \{1, \dots, n\}$$

Definición: \mathbb{D} es euclidiana si es la matriz de distancias de un conjunto de puntos en algún espacio euclidiano. (Esto es, si para algún $p \in \mathbb{R}^p$

$$x_1, \dots, x_n \in \mathbb{R}^p$$

$$y \quad d_{ij}^2 = D(x_i - x_j)$$

con D la forma cuadrática asociada a la matriz \mathbb{D} con la cual se define la distancia en \mathbb{R}^p .

En un problema particular es conveniente tener presente si se trata de una matriz euclidiana o hay que hacer consideraciones especiales para obtener una representación gráfica, en seguida se proporciona una caracterización de dichas matrices (ver 2).

Teorema: Si \mathbb{D} es una matriz de distancias y define $\mathbb{B} = \mathbb{B}(\mathbb{D})$ con $a_{ij} = -d_{ij}^2/2$ y $\mathbb{H} = \mathbb{I} - n^{-1}\mathbb{D}$ una matriz centrada.

Entonces \mathbb{D} es euclidiana si \mathbb{B} es semidefinida positiva.

Demostración:

La prueba no se incluye aquí pero se puede consultar en (2).

Observaciones:

- Cuando \mathbb{D} es matriz euclidiana de distancias entre puntos de una nube entonces $b_{ij} = (x_i - \bar{x})(x_j - \bar{x})'$ $(i, j) = 1, \dots, n$
 i.e. $\mathbb{B} = (\mathbb{D}\mathbb{H})(\mathbb{D}\mathbb{H})'$ y $\mathbb{B} \geq 0$. (Observación: \mathbb{D} es la matriz de productos internos centrada para \mathbb{B}).

En el caso que \mathbb{D} sea semidefinida positiva de rango p entonces la configuración correspondiente a \mathbb{B} se construye como sigue:

Sean $\lambda_1, \dots, \lambda_p$ los eigenvalores positivos ordenados en forma decreciente y \mathbb{V} la matriz cuyas columnas son los eigenvectores asociados a los eigenvalores, con norma la raíz cuadrada de su eigenvalor asociado.

$$v_j v_j' = \lambda_j \quad \forall j=1, \dots, p$$

Entonces los puntos p_j en \mathbb{R}^p cuyas coordenadas son $p_j = (y_{j1}, \dots, y_{jk})$ (el j -ésimo renglón de \mathbb{V}), tiene distancias

entre puntos dada por la ecuación tiene centro de gravedad $a=0$ y \mathbb{B} representa la matriz de producto interno.

En ese caso es posible construir una nube en \mathbb{R}^p a través de este producto interno siguiendo el procedimiento descrito en el esquema de dualidad del capítulo 1, con las consecuencias conocidas. En particular:

$$d_{ij}^2 = (x_i - \bar{x})(x_j - \bar{x})'$$

$$= \|x_i - \bar{x}\| \|x_j - \bar{x}\| \cos \theta$$

$$= \sigma_i \sigma_j \cos \theta$$

donde θ es el ángulo entre x_i y x_j .

En el caso en que algún eigenvalor sea negativo no se toma su eigenvector asociado pues la norma de este sería negativa según el procedimiento, por ello se excluyen dichos eigenvalores.

Además las distancias en la matriz \mathbb{D} puede ocurrir que no sea euclideana o que la dimensión del espacio en el que pueden ser representados los puntos sea mayor a dos o tres y no proporcionen una representación geométrica explícita de los datos y sus relaciones.

Una posibilidad es elegir l dimensiones es sugerida por el teorema 1, para ello hay que elegir una configuración en \mathbb{R}^l en la cual sus coordenadas están determinadas por los primeros l eigenvectores de \mathbb{B} . El argumento geométrico intuitivo de la prueba del teorema 1 es que si los l primeros eigenvalores de \mathbb{B} son grandes y positivos y el resto son cercanos a cero, ya sea positivos o negativos, entonces las distancias o similitudes entre los puntos de esta configuración pueden ser aproximados por \mathbb{D} , este procedimiento se puede describir aún de manera más simple si se piensa como una fotografía que mejor describa a la nube o como la fotografía de la nube desde el ángulo óptimo (este ejemplo para nada es formal pero sí bastante descriptivo). Esta configuración es la llamada solución clásica del escalamiento multidimensional en l dimensiones, es una solución métrica y tiene propiedades óptimas.

Una presentación del algoritmo es conveniente puesto que resume la discusión precedente:

1.2. Algoritmo.

i) A partir de la matriz D construir la matriz de similitudes transformadas :

$$A = (-d_{ij}^2 / 2)$$

ii) Obtener la matriz B con elementos

$$b_{ij} = a_{ij} - \bar{a}_{i.} - \bar{a}_{.j} + \bar{a}_{..}$$

donde :

$$\bar{a}_{i.} = \sum_n a_{in} ; \bar{a}_{.j} = \sum_m a_{mj} ; \bar{a}_{..} = \sum_{m,n} a_{nm}$$

iii) Considerar los l eigenvalores positivos $\lambda_1, \dots, \lambda_l$ de B , con sus eigenvectores correspondientes : $Y = (Y^1, \dots, Y^l)$ los cuales están normalizados por $Y^i Y^i = \lambda_i \quad \forall i=1, \dots, l$.

iv) Las coordenadas requeridas para los puntos son precisamente los renglones de Y .

o

Con este algoritmo se observa que basta con una rutina para cálculo de eigenvalores y eigenvectores para aplicar el método.

Algunas veces no es posible proporcionar una "medida" de comparación en términos de una métrica entre n individuos, en ese caso se propone una función de similitudes con matriz asociada S que cumple las siguientes propiedades :

- i) $u_{ij} = u_{ji} \quad \forall i, j$
- ii) $u_{ij} \leq u_{ii}$

Sin embargo aún en estos casos se pueden ser reducidos a trabajar con matrices de distancias D aplicando la transformación siguiente :

$$d_{ij} = (u_{ii} - 2u_{ij} + u_{jj})^{1/2}$$

Se observa que si S es matriz de similitudes, por ii) se tiene que $u_{ii} - 2u_{ij} + u_{jj}$ es no negativo y por lo tanto está bien definida la transformación propuesta, además $d_{ii} = 0$ con lo cual D es una matriz de distancias.

En caso de que la matriz de similitudes sea además semidefinida positiva

$$\sum_{i=1}^n x_i^2 u_{ii} + 2 \sum_{i=1}^n x_i x_j \geq 0$$

si se considera $x = \begin{pmatrix} 0 \\ \vdots \\ -1 \\ \vdots \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$, entonces se obtiene

$$d_{ij}^2 = u_{ii} + u_{jj} - 2u_{ij} \geq 0$$

sin embargo x se puede escribir como :

$$x = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix}_j - \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}_i = x_j - x_i$$

S es simétrica y semidefinida positiva por ello sus eigenvalores $\lambda_j, j=1, \dots, n$ son no negativos y reales, además existe una matriz ortogonal P tal que $P^T S P = \Lambda$ o $S = P \Lambda P^T$, donde $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, es decir que :

$$S = P \Lambda P^T = (P^T)^T \Lambda (P^T) = \|P^T\|_{\Lambda}$$

P está asociada a una transformación rígida i.e.

$$\|S\|_{\Lambda}$$

Si $x_j = x_i - \lambda_i$ entonces $d_{ij}^2 = \|x_j - x_i\|_{\Lambda}$ es euclidiana. Esta conclusión se resume en la siguiente proposición.

Proposición : Si S es matriz de similitudes semidefinida positiva, entonces la matriz de distancias D definida por la transformación (i) es euclidiana, con matriz de producto interno centrada

$$B = D^T S D$$

1.3. Dualidad entre ACP y A. de coordenadas principales.

Sin embargo en la práctica solo se cuenta con la matriz de datos $X_{n \times p}$ y hay que hacer una elección de la función distancia, muchas elecciones son posibles pero las más comunes son la euclídeana y sus variantes, en ese caso el trabajo desarrollado anteriormente se encuentra conectado con el ACP.

Sea $X_{n \times p}$ una matriz de datos y sea $\lambda_1 \geq \dots \geq \lambda_p > 0$ los eigenvalores de la matriz $X'BXn^{-1}$. La matriz de covarianzas muestral, para hacer más simple el análisis suponga que los eigenvalores son distintos entre sí. Los renglones de BX son los renglones centrados de X y por lo tanto se tiene que $B=BX'X'B$ representa la matriz de producto interno centrada y tiene los mismos eigenvalores que la matriz de covarianza.

$$b_{ij} = (\bar{x}_i - \bar{g})(\bar{x}_j - \bar{g})'$$

Ahora se establece la dualidad entre el ACP y el Análisis de Coordenadas Principales para ello es necesaria la siguiente definición:

Definición: Sea w_i el i -ésimo eigenvector de B con eigenvalor asociado λ_i , normalizado por $w_i'w_i = \lambda_i \quad \forall i=1, \dots, p$. Para l fijo ($1 \leq l \leq p$) los renglones de $B(l) = [w_1, \dots, w_l]$ son llamadas las coordenadas principales de X en l dimensiones.

Proposición: Las coordenadas principales de X en l dimensiones están dadas por los valores centrados de los n individuos sobre las l primeras componentes principales.

Demostración:

En efecto, si c^i es la i -ésima componente principal con

$$\langle c^i, c^i \rangle = 1$$

por el teorema de descomposición espectral (12) se tiene que

$$X'BX = VAV'$$

donde $V = [c^1, \dots, c^p]$ y $A = \text{diag}(\lambda_1, \dots, \lambda_p)$.

Por el teorema de descomposición en valores singulares (12) se puede elegir los signos de c^i y w_i , también BX puede ser escrita en términos de esos eigenvectores como:

$$BX = VU' = \sum_{k=1}^p \tau_k \quad \text{donde } U = (u_1, \dots, u_p)$$

Los valores de los n renglones de BX sobre la i -ésima componente principal están dados por los n elementos de BXc^i , los renglones de esta matriz son precisamente la componente de los renglones de BX en la dirección de c^i . Entonces, escribiendo

$$U(k) = [c^1, \dots, c^k]$$

una submatriz de k columnas de la matriz U , los registros de las k primeras componentes principales están dados por:

$$BXU(k) = VU'U(k) = V \begin{bmatrix} I_k \\ 0 \end{bmatrix} = V_k$$

Puesto que las componentes principales son ortogonales entre sí, $V_k = (w_1, \dots, w_k)$ son precisamente las coordenadas principales.

o

Observación:

-Puesto que las columnas de $U(k)$ son ortogonales entre sí, $U(k)'U(k) = I_k$ y se observa que $V_k'BXU(k)$ representa una proyección de X sobre un subespacio de k dimensiones generado por las columnas de $U(k)$ contenido en B^p .

- Este resultado es dual al de ACP en donde la inercia del subespacio generado por las k primeras componentes principales es la más grande de cualquier subespacio de k dimensiones.

2. / Selección de características con componentes principales.

2.1. / Comentarios.

En regresión múltiple, una de las mayores dificultades con las estimaciones usuales de mínimos cuadrados es el problema de multicolinealidad, el cual ocurre cuando se trata de funciones lineales casi constantes de dos o más de los predictores, o características regresoras. Recientemente una revisión al problema de multicolinealidad fue propuesta por Gunst (1983). Dichas multicolinealidades son a veces, aunque no siempre indicadas por las correlaciones grandes entre subconjuntos de características, y si existen las multicolinealidades entonces las varianzas de alguno de los coeficientes estimados de la regresión son muy grandes, pesados o inestables dentro de la ecuación de regresión. Otros métodos han sido propuestos. Una manera de salvar dicho problema es usar solo un subconjunto de características predictoras, elegido de tal manera que no contenga multicolinealidades. En la actualidad se cuenta con un asombroso catálogo de métodos para seleccionar un tal subconjunto (Draper y Smith 1981), entre los métodos propuestos hay una clase que merece particular atención, los basados en ACP. El uso se sigue de una segunda clase de aproximaciones al problema de multicolinealidad caracterizados por el uso de estimadores sesgados en regresión. Esta clase incluye a los estimadores de contracción y también son una aproximación basada en ACP. Pero la más conocida de todas las propuestas es la regresión con ACP, que simplemente establece en el uso de las CP de las características predictoras en lugar de las características predictoras. Puesto que las CP son no correlacionadas no hay multicolinealidades entre ellas, y los cálculos en la regresión se simplifican. Si todas las CP son incluidas en la regresión, entonces el modelo resultante es equivalente al obtenido por mínimos cuadrados, en cuyo caso la inercia causada por las

multicolinealidades es grande pero no son controladas. Además al se elimina alguna de las CP de la ecuación de regresión, los estimadores obtenidos para los coeficientes en la ecuación original de regresión son usualmente sesgado, pero reducen efectivamente las varianzas de los estimadores de los coeficientes, las cuales son causadas por las multicolinealidades. En esta sección se discute algunas estrategias de selección de características CP a eliminar en la ecuación de regresión.

Se sigue explotando la idea básica de selección a través de las CP, entre ellas se propone una que sugiere la eliminación parcial de una CP de las características predictoras junto con la característica dependiente. Esas CP pueden ser usados en la construcción de estimadores sesgados.

2.2. / Regresión con Componentes Principales.

Considere el modelo de regresión, como se define en la ecuación:

$$y = X\beta + \epsilon \quad (1)$$

donde y es un vector en E^n (de n observaciones) o es la medida dependiente de n individuos, considerada con respecto a su centro de gravedad y X es una matriz $n \times p$ cuya entrada x_i^k es el valor de la k -ésima característica predictora para el individuo i -ésimo ambas con respecto a su centro de gravedad, β es un vector de coeficientes de la regresión y ϵ es el vector de términos de error, los elementos de ϵ se suponen independientes y homocedásticos con varianzas σ^2 .

Conviene presentar (1) centrada para las características con respecto a su centro de gravedad, puesto que el caso general se obtiene mediante una traslación y la ganancia es que XX' es proporcional a la correlación entre las características predictoras, y esta convención se adopta en este trabajo.

Los valores de las CP para cada observación están dados por:

$$Z = XA \quad (2)$$

donde la entrada z_i^k de Z es el valor de la k -ésima CP para el i -ésimo individuo, y A es una matriz $p \times p$ cuya k -ésima columna es el k -ésimo eigenvector de XX' .

Puesto que A es ortogonal, $X\beta$ puede ser escrito como

$$XAA'\beta = Z\gamma$$

donde $\gamma = A'\beta$ y por ello la ecuación (1) adquiere la siguiente forma

$$y = Z\gamma + \varepsilon \quad (3)$$

la cual es simplemente un reemplazo de las características predictoras por sus CP en el modelo de regresión. Regresión con CP puede ser definido como el uso del modelo (3) o del modelo reducido

$$y = Z_m \gamma_m + \varepsilon_m \quad (4)$$

donde γ_m es un vector con m elementos que son un subconjunto de elementos de γ , Z_m es una matriz $n \times m$ cuyas columnas corresponden a un subconjunto de columnas de Z , y ε_m son los términos de error correspondientes. Por el uso de mínimos cuadrados para estimar γ en (3) es posible encontrar un estimador para $\hat{\beta}$ de la ecuación:

$$\hat{\beta} = A \hat{\gamma} \quad (5)$$

En efecto:

$$\begin{aligned} ZA'\hat{\beta} &= XAA'\hat{\beta} = Z\hat{\gamma} \rightarrow Z(A'\hat{\beta} - \gamma) = 0 \\ &\rightarrow Z'(Z(A'\hat{\beta} - \gamma)) = 0 \\ &\rightarrow A'\hat{\beta} - \gamma = 0 \\ &\rightarrow A\hat{\gamma} = \hat{\beta} \\ &\rightarrow \hat{\beta} = A \hat{\gamma} \end{aligned}$$

Observación:

-Entonces es equivalente a encontrar $\hat{\beta}$ por mínimos cuadrados en (1).

-Calcular $\hat{\gamma}$ en (3) es más sencillo que encontrar $\hat{\beta}$ en (1), ya que las columnas de Z son ortogonales,

también $\hat{\gamma}$ se obtiene:

$$\begin{aligned} \hat{\gamma} &= (Z'Z)^{-1}Z'y \\ &= I^{-2}Z'y \end{aligned} \quad (6)$$

con

$$I = \begin{bmatrix} \lambda_1^{-1/2} & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & \lambda_n^{-1/2} \end{bmatrix}$$

y $\lambda_1 > \lambda_2 > \dots > \lambda_k$ son eigenvalores de $X'X$. Además, si la ecuación de regresión es calculada para las CP, en lugar de hacerlo para las características predictoras, entonces las contribuciones de cada CP a la ecuación pueden ser fácilmente interpretadas, cosa que no ocurre con las características originales, por la falta de ortogonalidad. Así en el caso de tener presente el problema de multicolinealidad, el realizar una regresión con CP en lugar de las características originales trae consigo ventajas de cálculo e interpretación. Además es posible notar acerca de la interpretación de las contribuciones separadas de cada característica transformada es mejorada al tomar CP, aunque su interpretación en la ecuación de regresión puede no tener un significado muy claro.

-La principal ventaja en el uso de CP en regresión cuando las multicolinealidades están presentes. En este caso es borrando un subconjunto de las CP, especialmente aquellas varianzas pequeñas con lo cual se obtienen estimadores mucho más estables para $\hat{\beta}$. Para verificar esta observación hay que sustituir (6) en (5) para obtener

$$\begin{aligned} \hat{\beta} &= A(Z'Z)^{-1}Z'y \quad (7) \\ &= AL^{-2}Z'y \\ &= AL^{-2}A'X'y \\ &= \sum_{k=1}^p \lambda_k^{-2} u_k u_k' X'y \end{aligned} \quad (8)$$

con λ_k el k -ésimo elemento diagonal de L^{-2} y u_k es la k -ésima columna de A .

Esta última ecuación puede ser derivada directamente de

$$\hat{\beta} = (X'X)^{-1}X'y$$

por el uso de la descomposición espectral de la matriz $(X'X)^{-1}$ cuyos eigenvectores u_k y eigenvalores λ_k^{-2} , $k=1, \dots, p$.

Basando la suposición usual que los elementos de γ son no correlacionados, homocedásticos (i.e. $\text{cov}(\gamma) = \sigma^2 I_n$), de aquí que la matriz de covarianza de $\hat{\beta}$ es

$$\begin{aligned} \sigma^2 A(Z'Z)^{-1}Z'Z(Z'Z)^{-1}A' &= \sigma^2 A(Z'Z)^{-1}A' \\ &= \sigma^2 AL^{-2}A' \\ &= \sigma^2 \sum_{k=1}^p \lambda_k^{-1} u_k u_k' \end{aligned} \quad (9)$$

Esta última expresión proporciona otra forma de identificar multicolinealidades que produzca varianzas grandes para los elementos de $\hat{\beta}$. Si existe una multicolinealidad, entonces esta aparece como una CP con varianza pequeña, es decir, las CP que tienen varianzas pequeñas de λ_k (la varianza de la k -ésima CP es

$$\lambda_k / n)$$

, y por tanto para valores grandes de λ_k^{-1} . Entonces (9) muestra que cualquier característica predictora la cual tiene coeficientes grandes o medianos en cualquiera de las CP asociadas con los eigenvalores más pequeños que tienen varianza grande.

Un camino para reducir este efecto, es borrar los términos de (9) que correspondan a los más pequeños λ_k 's, si se queda solo con $m \leq k$ entonces se obtiene el estimador:

$$\tilde{\beta} = \sum_{k=1}^m \lambda_k^{-1} u_k u_k' X'Y \quad (10)$$

en este caso $\lambda_{m+1}, \dots, \lambda_p$ son eigenvalores más pequeños. Esto es equivalente a que los últimos $p-m$ elementos de γ sean iguales a cero.

-La matriz de covarianza $V(\tilde{\beta})$, para $\tilde{\beta}$ es

$$\sigma^2 \sum_{j=1}^m \lambda_j^{-1} u_j u_j' X'X \sum_{k=1}^m \lambda_k^{-1} u_k u_k'$$

Pero $X'X = \sum_{i=1}^n \lambda_i u_i u_i'$ por tanto de la descomposición espectral de $X'X$ se tiene

$$V(\tilde{\beta}) = \sigma^2 \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m \lambda_i^{-1} \lambda_j^{-1} u_i u_i' u_j u_j' u_k u_k'$$

Por otro lado los vectores $u_i, i = 1, \dots, p$ son ortonormales, los términos no nulos en la triple suma ocurren cuando $i=j=k$, y por ello

$$V(\tilde{\beta}) = \sum_{k=1}^m \lambda_k^{-1} u_k u_k' \quad (11)$$

En el caso en que ninguno de los primeros m λ_k 's sea muy pequeños entonces ninguna de las varianzas dadas por los elementos diagonales de (11) son grandes.

El decremento en varianza del estimador $\tilde{\beta}$ dado por (10) en relación al obtenido por mínimos cuadrados $\hat{\beta}$, se debe, como se esperaba a la introducción del estimador sesgado en el estimador $\hat{\beta}$. Es decir:

$$\tilde{\beta} = \hat{\beta} - \sum_{k=m+1}^p \lambda_k^{-1} u_k u_k' X'Y \quad \text{y} \quad E(\tilde{\beta}) = \beta$$

Y

$$E \left[\sum_{k=m+1}^p \lambda_k^{-1} u_k u_k' X'Y \right] = \sum_{k=m+1}^p E (\lambda_k^{-1} u_k u_k' X'Y)$$

$$= \sum_{k=m+1}^p \lambda_k^{-1} u_k u_k' X'X \beta$$

$$= \sum_{k=m+1}^p u_k u_k' \beta.$$

Este último término es, en general no nulo y por tanto $E(\hat{\beta}) \neq \beta$.

Cuando la multicolinealidad es un serio problema la reducción de la varianza puede ser sustancial, y el sesgo introducido es pequeño.

Esta afirmación es clara cuando los elementos correspondientes a las componentes eliminadas son nulos, entonces un tal sesgo no existe.

Además cuando se eliminan términos en (8) que correspondan a los eigenvalores más pequeños, es posible que estos estén asociados a un elemento de γ que no es significativamente diferente de cero, este problema es básicamente el problema de selección de características, con las CP (combinaciones lineales de las originales) y es un problema de estimación sesgada pero que permite optimizar en términos de explicación de la inercia.

La definición de regresión con CP proporcionada en términos de la ecuación (3) y (4), es equivalente al uso del modelo lineal (1) y estimando β por

$$\tilde{\beta} = \sum_k \lambda_k^{-1} u_k u_k' X' y \quad (12)$$

donde $S \subset J(p) = \{1, 2, \dots, p\}$. Así que (10) es un caso particular de la (12) en donde $S = J(m)$, en general S puede ser cualquier subconjunto de $J(p)$, por otra parte dado un subconjunto de coeficientes de γ , correspondientes al complemento de S , puede ocurrir que este sea igual a cero. La elección de el subconjunto S es un problema cuya respuesta ha sido un tema de discusión, un punto de vista considera la descomposición en valores singulares de la matriz X , y bajo el supuesto que $n \geq p$ y

$$X = ULA'$$

donde las columnas de U son los eigenvectores de XX' , los cuales corresponden a los eigenvalores no nulos, normalizados por $UU' = I_p$, y L se define como:

$$L = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{bmatrix}$$

Entonces $X\tilde{\beta}$ puede ser reescrito como:

$$ULA' \beta = U\delta,$$

con $\delta = LA'\beta$, tales que $\beta = AL^{-1}\delta$. En donde el estimador por mínimos cuadrados para δ es

$$\hat{\delta} = (U'U)^{-1}U'y = U'y,$$

de donde se obtiene que $\hat{\beta} = AL^{-1}\hat{\delta}$.

Por tanto a partir de las relaciones entre γ y δ se infiere la igualdad:

$$\gamma = \lambda'\beta = \lambda'(AL^{-1}\delta) = (\lambda'A)L^{-1}\delta = L^{-1}\delta,$$

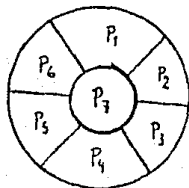
y en la cual es posible observar que considerar un subconjunto de elementos de δ iguales a cero es equivalente a elegir el mismo subconjunto de elementos de γ iguales a cero, proporcionando una formulación alternativa de la ecuación (12). Dicha formulación a través de la descomposición en valores singulares provee de una aproximación alternativa para estimar con ecuaciones de regresión basadas en CP.

1.4. Ejemplo .

En la presente sección se proporcionan ejemplos de aplicación del algoritmo descrito en la sección anterior.

1.4.1. Matrices no euclidianas.

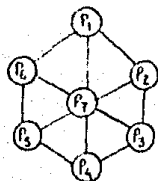
En una región del planeta hay 7 países cuya situación geográfica es la siguiente :



La red de transporte (ferrocarril) toma en cuenta únicamente el número de fronteras que atraviesa para ir de un país a otro, puesto que al cruzar una frontera entre dos países se cobra un arancel es deseable cruzar el mínimo número de fronteras posible (esto es, se considera que la distancia entre un país y otro, es el mínimo de fronteras cruzadas).

Es claro que la distancia así definida no es una métrica euclidiana.

El problema se puede trasladar a una gráfica \mathcal{G} del tipo :



En este contexto cada nodo p_i corresponde a un país y cada arista (p_i, p_j) corresponde a la frontera común de los países p_i y p_j .

La distancia entre los nodos p_i y p_j de la gráfica \mathcal{G} conexa, es la longitud de la cadena más corta entre p_i y p_j . Esta distancia se designa $\delta(p_i, p_j)$ y en el apéndice 1 se demuestra que δ , en efecto cumple la definición de distancia o métrica.

Se construye la matriz de distancias :

$$\delta(p_i, p_j) = \begin{matrix} & p_1 & p_2 & p_3 & p_4 & p_5 & p_6 & p_7 \\ \begin{matrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \\ p_7 \end{matrix} & \begin{pmatrix} 0 & 1 & 2 & 2 & 2 & 1 & 1 \\ & 0 & 1 & 2 & 2 & 2 & 1 \\ & & 0 & 1 & 2 & 2 & 1 \\ & & & 0 & 1 & 2 & 1 \\ & & & & 0 & 1 & 1 \\ & & & & & 0 & 1 \\ & & & & & & 0 \end{pmatrix} \end{matrix}$$

La matriz A resultante es :

$$A = \begin{pmatrix} 0 & -1/2 & -2 & -2 & -2 & -1/2 & -1/2 \\ & 0 & -1/2 & -2 & -2 & -2 & -1/2 \\ & & 0 & -1/2 & -2 & -2 & -1/2 \\ & & & 0 & -1/2 & -2 & -1/2 \\ & & & & 0 & -1/2 & -1/2 \\ & & & & & 0 & -1/2 \\ & & & & & & 0 \end{pmatrix}$$

La matriz B obtenida es :

$$B = \begin{pmatrix} 1.102 & 0.002 & -0.020 & -0.020 & -0.020 & -0.004 & 0.020 \\ & 1.102 & 0.002 & -0.020 & -0.020 & -0.020 & 0.020 \\ & & 1.102 & 0.002 & -0.020 & -0.020 & 0.020 \\ & & & 1.102 & 0.002 & -0.020 & 0.020 \\ & & & & 1.102 & 0.002 & 0.020 \\ & & & & & 1.102 & 0.020 \\ & & & & & & 1.102 \end{pmatrix}$$

Los eigenvalores asociados a los componentes correspondientes se listan a continuación :

Componentes	Eigenvalores
1	3.4999993
2	3.4999993
3	0.4999999
4	0.4999999
5	0.0000000
6	-0.142057
7	-0.9999998

La matriz de componentes ortogonales obtenida es :

$$V = \begin{pmatrix} 0.104 & -0.552 & 0.300 & -0.400 & 0.377 \\ 0.561 & -0.133 & 0.576 & -0.022 & 0.377 \\ 0.296 & 0.419 & -0.208 & 0.511 & 0.377 \\ -0.104 & 0.552 & -0.300 & -0.400 & 0.377 \\ -0.561 & 0.133 & 0.576 & -0.022 & 0.377 \\ -0.296 & -0.419 & -0.208 & 0.511 & 0.377 \\ -0.000 & 0.000 & 0.000 & 0.000 & 0.377 \end{pmatrix}$$

La representación obtenida en un subespacio de dos dimensiones es :

**

EJEMPLO :

Se propone reconstruir el mapa de la República Mexicana (de 9 ciudades de México) a partir de una matriz de distancias por carretera, se observa que dichas "similaridades" (más que distancias en el sentido de la definición en el apéndice 1, son similitudes pero no distancias euclidianas). Para ello se utiliza el algoritmo proporcionado en este capítulo.

Se utilizó el programa de cómputo *MATLAB* y en la máquina en que se trabajó solo aceptaba sin aparentes problemas una matriz de dimensiones relativamente pequeñas.

En las hojas siguientes se presentan los datos que intervinieron en el estudio.

-La matriz D tiene por elementos los cuadrados de las "distancias" entre ciudades.

$$D = (-1/2 \ d_{rs}^2)$$

-La matriz $E = -\frac{1}{2}I_D$, donde I_D es la identidad de orden 9.

-La matriz $A = DE$ es la matriz de distancias transformadas

$$A = (-\frac{1}{2} \ d_{rs}^2) \quad 1 \leq r, s \leq 9$$

-La matriz $H = I_D - \frac{1}{9} \ 00'$ la matriz usada para centrar.

-La matriz $B = H A H$ de distancias transformadas y centradas.

-La matriz X cuyas columnas son los eigenvectores de B .

-La matriz T cuya diagonal son los eigenvalores de B .

La calidad de la representación se refleja en los siguientes porcentajes :

Eje	Porcentaje explicado.
1	65.3 %
2	84.0 %
3	98.4 %

- Los eigenectores de B son unitarios.
- Las ciudades analizadas que aparecen en el orden de la matriz de similitudes :

- 1.-Campeche
- 2.-Chetumal
- 3.-Chihuahua
- 4.-Mérida
- 5.-Mexicali
- 6.-Monterrey
- 7.-Tepic
- 8.-Tuxtla Gutierrez
- 9.-La paz

Tienen por coordenadas en su representación :

Ciudad	Coordenadas
P_1	(0.2725,-0.0346)
P_2	(0.3259,0.1963)
P_3	(-0.0549,-0.4650)
P_4	(0.2956,0.0552)
P_5	(-0.3540,0.1368)
P_6	(0.0157,-0.4255)
P_7	(0.0043,-0.2901)
P_8	(0.2353,-0.1318)
P_9	(-0.7404,0.9667)

La representación obtenida aparece en la hoja 1 de resultados y sobreponiendo un mapa de la República Mexicana es posible trazar regiones guiándose por la posición geográfica de dichas ciudades en tales regiones, posteriormente se presenta la representación de México si hubiera carreteras rectas y transbordadores cuyo curso también es recto en la comunicación entre puertos.

Después se proporciona el mismo mapa reconstruido en donde aparecen cuadros regionales C_1, C_2, C_3 y C_4 que corresponden a las regiones originales de la República Mexicana.

Observación :

C_1 : Corresponde a la península de Baja California ubicada por La paz y Mexicali. (cambio de orientación).

C_2 : Corresponde al región que ocupan Sonora, Chihuahua y

Sinaloa, pero bajo una reflexión se puede voltear el mapa de México de la transparencia y sobreponerle en el dibujo.

C_3 : Corresponde a la península de Yucatán en el sentido y orientación usual.

C_4 : Corresponde al resto de la República Mexicana en el sentido usual (sin voltear la mica al revés) pero con orientación invertida. El norte es abajo y el sur arriba. (Esta región consta de casi todos los estados excepto los enunciados en C_2 y pequeñas porciones de otros, pero de cabeza).

Conclusión :

La representación obtenida es distinta de la original ello se debe principalmente a que la euclideanización de las "distancias" indica que en realidad las similitudes poco tienen que ver con la distancia euclideana usual.

Se observa que geográficamente dos ciudades pueden estar cercanas pero que entre ellas puede existir un muro formado por montañas, lagos o bordes curvos como es el caso del Istmo de Tehuantepec, o brazos de mar (o Golfos) como es el de Baja California.

Se observa que el análisis puede ser más exacto con ciudades dentro de un mismo estado o entidades cercanas donde no exista este tipo de deformaciones geográficas. Muy posiblemente se puede reconstruir una región mexicana donde no estén separadas las ciudades por montañas y la región sea convexa que el territorio total de la República Mexicana.

Problema :

Establecer un análisis en términos de convexidad de las regiones.

EJEMPLO :

Producción nacional azucarera.

Ahora se proporciona un ejemplo completo , concerniente a la producción nacional azucarera, el cual no ha sido discutido en ningún trabajo hasta el momento , en dicho ejemplo se aplica la discusión anteriormente expuesta. Los datos consisten de 10 características, las cuales son medidas observadas en 46 años (desde 1925 hasta 1970). El objetivo es construir una ecuación de predicción para una de las características (la producción de azúcar en toneladas , y) usando los valores de las otras 9 características, las cuales son medidas físicas.

Las características que intervienen en el estudio son ;

Característica	Clave
Superficie de caña (Ha.)	x1
Rendimiento de campo caña	x2
Rendimiento de azúcar (Ha.)	x3
Rendimiento en tab. (%)	x4
Toneladas de caña por toneladas de azúcar.	x5
Comercio exterior	x6
Importaciones (Ton.)	x7
Comercio exterior	x7
Exportaciones (Ton.)	x7
Consumo nacional (Ton.)	x8
Consumo Per-Cápita (Kg.)	x9
Producción de azúcar (Ton.)	y.

Los datos son listados en la página 1 de resultados y la matriz de covarianzas y las de correlaciones en las tablas 2 y 3 respectivamente.

Los coeficientes de cada una de las componentes principales

sobre las características es proporcionada en la tabla 4 , para el ACP sobre la matriz de covarianza y en la tabla 5 para el ACP sobre la matriz de correlación.

En el caso del ACP sobre la matriz de covarianzas es fácil interpretar dichos coeficientes, pues con la primer componente principal se tienen explicada 96.75 % de la inercia total, los coeficientes γ_k en una regresión de y sobre las componentes principales, y los valores de las T-estadísticas miden la importancia de cada componente principal en la regresión.

Juzgando unicamente el tamaño (porcentaje) de la inercia explicada debe aparecer que las últimas 7, o posiblemente 8 componentes principales deben ser eliminadas de la regresión. Además, observando los valores de las γ_k y las correspondientes T-estadísticas, se puede ver que basta retener cuando mucho la segunda componente principal en esta ecuación de regresión.

En el caso de la matriz de covarianzas el ejemplo es lo suficientemente claro para mostrar la selección de características con ACP.

INTRODUCCION

En este capítulo se presenta el método STATIS usado en el análisis de tablas cúbicas o 3-dimensionales o sobre un conjunto de tablas 2-dimensionales indexadas por el tiempo. En este método se considera que el orden de preferencia de cada arista del cubo de datos es irrelevante y por ello se pueden trabajar características (variables), individuos o tiempo de igual manera. La estrategia es separar un fenómeno de elección de una estructura invariante sobre el tiempo u otro parámetro, para ello se basa en la descomposición de la inercia total en inercia inter e inercia intra.

El método STATIS permite extraer información en forma de gráficas; resumiendo globalmente, las posiciones acordadas de los individuos en un sistema de ejes interpretables con la ayuda de las características, evolución de cada individuo autor de su posición acordada en el mismo sistema de ejes. Si se prefiere seguir la evolución de las características se debe usar el método dual.

El método STATIS de H.L'HERMIER DES PLANTES y Y.ESCOFFIER aplicable a varias situaciones descritas en detalle en el presente capítulo en la sección 2 y completado por el método "STAVA" de M. G. PLACE fue originalmente desarrollado y propuesto en Francia con técnicas propias del Análisis de datos y actualmente es considerado por su implementación computacional y programas como el software estadístico apropiado y dinámico en análisis multivariado para analizar un gran número de situaciones.

Existen otros métodos también desarrollados en Francia:

-El método de FOUCAULT.

-El método de JAFFRENOU.

Los cuales pueden ser consultados en (5) y (6) respectivamente.

STATIS puede ser clasificado en la familia de los métodos propuestos por R. Coppi (1966), basados en las medidas de asociaciones, cuyo interés principal es tratar tablas con dos

conjuntos de variables en función de un tercer conjunto invariante.

La presentación del método consiste de una exposición formal del fundamento teórico, para lo cual se requiere de algunos preliminares desarrollados en los capítulos I y II de esta tesis, un breve repaso del ACP para una tabla de datos cuyo objetivo es tener presente dicho desarrollo para comprender la generalización propuesta para el análisis conjunto de varias matrices de datos puede ser útil al lector.

Otro punto importante son las situaciones que se pueden presentar en un análisis de este tipo ya que de una tabla a otra no necesariamente se tienen las mismas características (individuos, matrices, etc.), desde un punto de vista formal esto se presenta a través de equivalencias entre situaciones.

Se presenta posteriormente en la sección 3 el esquema general del método en etapas:

Interestructura

Acuerdos

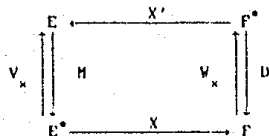
Intraestructura.

Con las cuales se obtiene una representación global de las tablas a través de una distancia para los elementos característicos, un resumen de las tablas que sea lo más representativo del conjunto total según la situación que se presente y una representación simultánea de todos los individuos o de todas las características que intervienen en el estudio.

Finalmente se proporciona un ejemplo de aplicación del método a un problema particular en el análisis de sectores del pequeño comercio en un grupo de ciudades de la República Mexicana, del cual se obtienen conclusiones importantes.

1.1 / Recapitulación de ACP.

En el capítulo I se describe el esquema de dualidad



$$V_x = XHX' \quad V_x = X'DX.$$

Con las restricciones del teorema del apéndice 1 se tiene:

-Las columnas de U son los eigenvectores normalizados de $V_x M$, que realizan, bajo las restricciones de M -ortonormalidad, el máximo de $\|XM\|_D^2$.

Sea $Y = \Delta V' = U'MX'$; los renglones de Y son las componentes principales, eigenvectores de $M'D$ de norma igual a la raíz cuadrada del eigenvector correspondiente.

Sea $\Lambda = \Delta^2$, matriz diagonal con eigenvectores de $M'M$ o de $V_x M$, y eigenvalores $\lambda_1, \dots, \lambda_r$.

Se tiene:

$$X' = U\Delta V' = UY = \sum_{i=1}^r u_i Y_i' \quad (Y_i \text{ es el } i\text{-ésimo renglón de } Y)$$

Según los resultados del apéndice 6 la aproximación óptima de M_x de rango $s \leq r$, en el sentido de la métrica $\| \cdot \|_{D,D}$ es obtenido por:

$$\sum_{i=1}^s Y_i Y_i'$$

que representa a los individuos de los primeros renglones de Y , tomando en cuenta en la interpretación los pesos que afectan a

cada uno de los individuos.

Los renglones de $Y = U'MX'$ son las coordenadas de las características en la base M -ortonormal de las columnas de U . Un individuo suplementario, vector $x \in E$, será representado por $U'Mx$.

Se representa análogamente las características de las primeras columnas de $Z = U\Lambda^{1/2} = X'DY'\Lambda^{1/2}$, coordenadas de las características en la base D -ortonormal de las columnas de $Y = Y'\Lambda^{1/2}$. Y en este caso las características suplementarias y vectores de F se pueden representar entonces por $y'DY'\Lambda^{1/2}$.

Se transforma el problema a uno equivalente que satisface las condiciones del último resultado del apéndice 6 ya que en este caso D es diagonal y las columnas de U asumen el máximo de:

$$\sum_{i=1}^n p_i \langle \cdot, \cdot \rangle_M^2$$

la inercia de la nube de los individuos en relación a los subespacios M -ortogonales a sus generados sucesivamente por los u_i ; como se observó detalladamente en el capítulo I las u_i generan los ejes principales de inercia de la nube.

A partir de esta propiedad, todos los resultados de ACP se valen aún cuando D no sea diagonal, pero que sea una métrica sobre F como lo demuestra el esquema de dualidad también parte del capítulo I.

Notación:

En el instante de tiempo k se consigue a través del esquema de dualidad la terna:

$$(X_k, M_k, D_k)$$

a partir de la cual es posible construir:

$$M_k = X_k M_k X_k' \quad \text{y} \quad V_k = X_k' D_k X_k$$

De aquí y en adelante denotaremos:

- Λ_k a la matriz diagonal de los eigenvalores no nulos de $M_k D_k$ o de $V_k D_k$ ordenados en forma decreciente
- Y_k es la matriz de dimensiones $r_k \times n_k$ cuyos renglones son los eigenvectores de $M_k D_k$, asociados a las λ_i con

D_k -norma igual a la raíz cuadrada de λ_k .

Z_k es la matriz de dimensiones $n \times n_k$ cuyas columnas son los eigenvectores de $Y_k M_k$, asociados a los λ_k , con M_k -norma igual a la raíz cuadrada de λ_k .

U_k se define como : $U_k = Z_k \lambda_k^{-1/2}$.

2/ Casos que se presentan en los estudios.

2.1/ Los casos.

Una tabla de datos X como se observó en el primer capítulo de esta tesis describe al fenómeno en estudio a través de las relaciones entre los individuos (características) en un instante del tiempo, en otras palabras es una fotografía del fenómeno, sin embargo, es deseable describir dicho fenómeno como un proceso evolutivo a distintos instantes del tiempo u otro parámetro, por ejemplo : medidas de parámetros físico-químicos a diversos grados de profundidad a lo largo de un transcurso de tiempo. En ese caso se dispone de K tablas de datos, cuantitativos o cualitativos, obtenidos en diferentes momentos o diferentes ocasiones y que se desean comparar.

Se pueden presentar los siguientes casos :

Caso 1 : Las K matrices de datos cuantitativos son construidas sobre los mismos individuos, con los mismos pesos. Es decir se cuenta con (X_k, M_k, D) , donde X_k es una matriz de $n \times p_k$,

centrada por D , y M_k una métrica para \mathbb{R}^{p_k} . Según el esquema de dualidad es posible construir $M_k = X_k M_k X_k'$.

Caso 2 : Las K matrices de similitudes, disimilitudes o de distancias entre los mismos individuos con los mismos pesos (matriz diagonal D). Se sustituye entonces en cada tabla una matriz M_k simétrica, semidefinida positiva, centrada por D , de manera que se puedan comparar las diferentes visualizaciones de los n individuos analizados por las K tablas, con el promedio de las M_k .

-Si se tiene una matriz de similitudes S_k , se construye

$\hat{S}_k = (I - 00' D) S_k (I - D 00')$; si \hat{S}_k es semidefinida positiva, se hace $M_k = \hat{S}_k$, si no, sea $\hat{S}_k^* = \hat{S}_k + |\lambda_m| I$, donde λ_m es el eigenvalor más pequeño de \hat{S}_k . Se hace entonces

$$M_k = (I - 00' D) \hat{S}_k^* (I - D 00'),$$

Se puede mostrar que \hat{S}_k^* es semidefinida positiva. W_k tiene las propiedades requeridas.

-Si se tienen matrices de disimilaridades, de distancias, sea \hat{D}_k la matriz cuyos elementos son los cuadrados de las disimilaridades; se construye:

$$\hat{S}_k = -\frac{1}{2} (I - DD' D) \hat{D}_k (I - DDD')$$

y se procede como se describe en el algoritmo descrito en el capítulo II de escalamiento multidimensional para obtener W_k .

Debido a los comentarios del caso 2, se considera en lo siguiente que se tiene únicamente (W_k, D) , donde W_k es una matriz $n \times n$ simétrica, semidefinida positiva centrada por D .

Caso 3: Las K matrices cuantitativas son resultado de la observación de las mismas variables. Se tienen K ternas (X_k, M_k, D_k) , donde X_k es una matriz $n \times p_k$ centrada por D_k , y M una métrica sobre \mathbb{R}^p .

Caso 4: Las K matrices son las covarianzas entre las mismas p características. Se tiene (V_k, M) , donde V_k es una matriz $p \times p$, simétrica, semidefinida positiva, y M es una métrica sobre \mathbb{R}^p .

Caso 5: Las K matrices de datos cuantitativos son resultado de la observación de las mismas características sobre los mismos individuos; es decir son registros de datos cronológicos.

Por lo que se vuelve a los casos 1 y 3. Se puede volver a interesarse en las posiciones relativas de los individuos, en la evolución de un individuo o de una característica.

En este caso se cuenta con K ternas (X_k, M_k, D) , donde X_k es una matriz $n \times p_k$.

Caso 6: Las K características cuantitativas medidas sobre los mismos individuos, con los mismos pesos. Se tienen entonces las ternas $(X_k, D_{i/p_k}, D)$ donde X_k es la matriz de indicadores de las modalidades de la k -ésima característica, D_{i/p_k} la matriz diagonal de los pesos de las diferentes modalidades ($D_{i/p_k} = X_k' D X_k$), y $D_{i/p_k} = D_{i/p_k}'$.

Se observa aquí que se pueden representar por una parte todas las características, por otra parte todas las modalidades de las diferentes características.

Caso 7: K tablas de contingencia, o de probabilidades, o de cambios, X_k , entre dos características cualitativas. (Esta situación es de suma importancia por la frecuencia con que se presenta en las aplicaciones).

Para la exposición se ha de considerar a las tablas de Burt:

$$H_k^* = \begin{bmatrix} D_p(k) & X_k' \\ X_k & D_q(k) \end{bmatrix}$$

donde $D_p(k)$ y $D_q(k)$ son las matrices diagonales de los pesos marginales deducidos de X_k .

En algunos casos se presentan dentro del mismo estudio características cuantitativas y cualitativas medidas sobre el mismo conjunto de individuos, una solución práctica es considerar como cualitativas a todas las características en esos casos con lo cual se plantea nuevamente en el caso anterior.

2.22 Tablas equivalentes.

Si se dispone de K tablas de datos del mismo tipo, se les puede comparar. Por ello se aborda el problema de definir equivalencia entre dichas tablas e identificar un elemento característico de cada clase de equivalencia. Dicha equivalencia depende de varias factores, el caso en que se encuentran ambas tablas, el objetivo del estudio, etc.

En los datos tabulados se puede comparar las relaciones entre las tablas con, entre tablas, es decir, comparar las matrices M_i . Así un elemento característico del estudio es M_i .

2.2.1.2 Definiciones y equivalencia.

Supóngase que (X_1, H_1, D) y (X_2, H_2, B) son equivalentes en $\mathbb{R}^p \times \mathbb{R}^q$.

Esto significa que si se ha hecho el ACP de dos series, y se obtiene la misma representación de los individuos:

$$Y_1 = Y_2 \quad A_1 = A_2 \quad (\text{con la notación de 1.1.1.})$$

Equivalencia (c) (X_1, H_1, D) y (X_2, H_2, B) son equivalentes si existe $U \in \mathbb{R}^q$ tal que $H_1 B = U H_2 D$.

Esto significa que si se ha hecho el ACP sobre dos series, y se obtienen representaciones de individuos coincidentes:

$$Y_1 = U^{-1} Y_2 \quad A_1 = U^{-1} A_2$$

Esta equivalencia se puede reescribir de la manera como lo demuestra la siguiente:

Proposición 2.2.1.3. (X_1, H_1, D) y (X_2, H_2, B) son equivalentes si y sólo si $H_1 B = U H_2 D$ con U una matriz de $\mathbb{R}^q \times \mathbb{R}^q$.

tal que $H_1 B = U H_2 D = U^{-1} H_1 X_1^{-1} H_1 X_2^{-1} H_2 B$.

Demostración:

$$\begin{aligned} \text{Si } (X_1, H_1, D) & \text{ y } (X_2, H_2, B) \text{ son equivalentes,} & \text{entonces } H_1 B &= U H_2 D \\ \text{entonces } H_1 B &= U H_2 D & \text{entonces } H_1 B &= U H_2 D \\ \text{entonces } H_1 B &= U H_2 D & \text{entonces } H_1 B &= U H_2 D \end{aligned}$$

Sea U_1 una matriz $p_1 \times p_2$, H_1 ortogonal, obtenida al condicionar el sistema H_2 ortogonal de los vectores columna de U_1 .

Sea U_2 una matriz $p_2 \times p_1$, H_2 ortogonal, obtenida de manera

similar, sea \tilde{Y} la matriz $p_2 \times n$, $Y' = \begin{pmatrix} Y \\ 0 \end{pmatrix}$.

se tiene $X_1' = U_1 Y = U_1^{-1} U_1 Y = U_1^{-1} U_1 \tilde{Y}$ $X_2' = U_2 H_2^{-1} U_2 \tilde{Y}$

$$X_1' = U_1^{-1} U_1 U_2 U_2^{-1} U_2 \tilde{Y} = U_1^{-1} U_1 U_2 H_2^{-1} X_2'$$

Se propone $H = U_1 U_2^{-1} H_2^{-1} U_2$, en efecto se cumple la igualdad requerida.

$$H' H = H_2 U_2^{-1} U_1 U_1^{-1} U_2 H_2^{-1} = H_2 U_2^{-1} U_2 H_2^{-1} = H_2^{-1}$$

□

Se observa que al restringir los métricos a la identidad I , la equivalencia significa que las dos nubes de puntos son equivalentes salvo rotaciones y simetrías, la equivalencia se hace por las similitudes descritas en el capítulo 11.

2.2.2.2 Proyectores.

En el caso de estudios sobre el mismo conjunto de individuos (caso 1.2.5 a 1.2.7) la relación de los individuos (unos con otros) a través de sus posiciones se lleva a cabo en un subespacio vectorial de \mathbb{R}^p , \mathcal{X}_k generado por las características individuales. Esto lleva a comparar los estudios promedio de los proyectores B-ortogonales sobre \mathcal{X}_1 y \mathcal{X}_2 .

Si X_k es de rango p_k y A_k se escribe

$$X_k' C X_k D X_k^{-1} X_k' B$$

con la notación de 1.1.1, como los componentes principales forman una base de \mathcal{X}_k se tiene que son aproximadamente al rango de X_k y $A_k = Y_k' A_k^{-1} Y_k B$.

2.2.2.1. Las equivalencias en términos de propiedades

Equivalencia (ii): (X_1, D) y (X_2, D) son equivalentes si $A_1 = A_2$.

Esta equivalencia es independiente de M_1 y M_2 . Ello puede

ser considerado en el caso 2 a partir de $M_k = Y_k^* Y_k$, se construye el

proyector: $A_k = Y_k^* A_k^{-1} Y_k D$.

Proposición: Si hay métricas M_1 y M_2 tales que (X_1, M_1, D) y (X_2, M_2, D) son equivalentes en el sentido (i) o también lo son en el sentido (ii). Inversamente, si (X_1, D) y (X_2, D) son

equivalentes en el sentido (ii), y si X_1 es de rango p_1 y X_2 es de

rango p_2 (aquí $p_1 = p_2$), entonces $(X_1, (X_1' D X_1)^{-1}, D)$ y

$(X_2, (X_2' D X_2)^{-1}, D)$ son equivalentes en sentido (i).

Demostración:

En efecto: si $M_1 D = \alpha M_2 D$, $Y_1 = \alpha^{1/2} Y_2$, $A_1 = \alpha A_2$,

$$A_1 = Y_1^* A_1^{-1} Y_1 D = \alpha Y_2^* A_2^{-1} Y_2 D = A_2$$

si ahora $A_1 = A_2$, $X_1 (X_1' D X_1)^{-1} X_1' D = X_2 (X_2' D X_2)^{-1} X_2' D$,

además cuando $A_1 = A_2$, $A_1^{-1/2} Y_1$ y $A_2^{-1/2} Y_2$ proporcionan (por renglones) dos bases D-ortonormales del mismo subespacio vectorial de F, entonces si que $M_1 = \alpha M_2$, esas bases son iguales.

La caso 6 (datos cualitativos).

Si X_k es la tabla de las características indicatrices de las p_k modalidades de la característica cualitativa, \mathcal{Y}_k es el conjunto de las características cuantitativas que se pueden reconstruir a partir de la característica cualitativa. Cada elemento de \mathcal{Y}_k es de la forma $X_k'(a)$, donde a es un registro de la característica.

Si X_k no está centrada, se tiene siempre $\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = 0$ en \mathcal{Y}_k .

Sea Δ_k la recta de F generada por 0,

$$\mathcal{Y}_k^* = \mathcal{Y}_k^* \oplus \Delta_k$$

donde \mathcal{Y}_k^* es el espacio vectorial generado por las características de $X_k^* = X_k'(I - B \mathbf{0}\mathbf{0}')$, es decir el conjunto de las características

cuantitativas centradas que se pueden reconstruir a partir de la característica cualitativa al considerar sus modalidades.

Si denota por A_k al proyector sobre \mathcal{Y}_k , A_k^* al proyector sobre \mathcal{Y}_k^* :

$$(A_1 = A_2) \Leftrightarrow (A_1^* = A_2^*).$$

En ese caso, el hecho de centrar o no las tablas no influye sobre la equivalencia.

Como aquí se consideró que las ternas $(X_k, D, \mathcal{Y}_k, D)$ donde $D_{1 \times p_k} = (X_k' D X_k)^{-1}$, las equivalencias (i) y (ii) coinciden.

2.2.3. / El caso de las matrices de Covarianzas.

Cuando los resultados son observaciones de las mismas características como ocurre en los casos 3, 4 y 5 se elige como elemento característico a la matriz V_k de covarianzas:

$$V_k = X_k' D X_k.$$

2.2.3.1. / Equivalencias para este caso.

Equivalencia (a): (X_1, M, D_1) y (X_2, M, D_2) son equivalentes si $V_1 = V_2$.

Si se hace el ACP de dos ternas, se logra la misma representación de las características: $Z_1 = Z_2$, $\Lambda_1 = \Lambda_2$. (con las notaciones de 1.1.)

Equivalencia (b): (X_1, M, D_1) y (X_2, M, D_2) son equivalentes si existe αD tal que $V_1 = \alpha V_2$.

Aquí se logra representaciones homotéticas de las características:

$$Z_1 = \alpha^{1/2} Z_2, \quad \Lambda_1 = \alpha \Lambda_2.$$

2.2.4. / Tablas de Burt.

En caso de encontrarse en el caso 7, se tiene una tabla de contingencia.

El análisis de correspondencias de una tabla de contingencias es equivalente a realizar el ACP de la tabla

$$\begin{pmatrix} X' \\ Y \end{pmatrix}$$

obtenida al yuxtaponer las tablas de características indicatrices de las modalidades de dos características cualitativas. La matriz de covarianzas de este estudio se conoce como tabla de Burt :

$$\begin{pmatrix} D_p(k) & X'_1 \\ X_k & D_q(k) \end{pmatrix}$$

donde $D_p(k) = X'DX$

$$D_q(k) = YDY'$$

Equivalencia (i) : $(X_1 = X_2)$ sii $(D_1 = D_2)$.

Para comparar las tablas de contingencia, se compara las tablas de Burt correspondientes. Esto fué hecho por FOUCARD (13).

2.2.8. / Si se encuentra en el caso 5. Observaciones sobre los mismos conjuntos de características e individuos.

Se tiene un conjunto de observaciones sobre el mismo conjunto de características y el mismo conjunto de individuos ; se puede comparar las M_k , A_k y V_k . Dependiendo del objetivo del estudio a realizarse, partiendo de este puede ser más preferida una y u otra de las equivalencias correspondientes, según la evolución de los individuos o características que se desee estudiar.

Si se quiere estudiar a la vez las relaciones entre individuos y las relaciones entre características, basta comparar directamente las tablas iniciales, esta idea se plasma formalmente en la siguiente :

Proposición : Sea (X_1, M, D) y (X_2, M, D) dos estudios sobre el mismo conjunto de individuos y el mismo conjunto de características. Si existe $\alpha > 0$ tal que $X_1 M X_1' = \alpha X_2 M X_2'$ y $\beta > 0$ tal que $X_1' D X_1 = \beta X_2' D X_2$, entonces :

$$\alpha = \beta \quad \vee \quad X_1' = \alpha^{-1} X_2'$$

Demostración :

En efecto :

$$\left. \begin{aligned} \text{si } M_1 = M_2, Y_1 = \alpha^{-1/2} Y_2 \quad \vee \quad \Lambda_1 = \alpha \Lambda_2 \\ \text{si } V_1 = \beta V_2, Z_1 = \beta^{1/2} Z_2 \quad \vee \quad \Lambda_1 = \beta \Lambda_2 \end{aligned} \right\} \rightarrow \alpha = \beta.$$

$$\begin{aligned} \text{Además :} \quad X_1' &= U_1 Y_1 = Z_1 \Lambda_1^{-1/2} Y_1 \\ &= \beta^{1/2} Z_2 \frac{1}{\beta^{1/2}} \Lambda_2^{-1/2} \alpha^{1/2} Y_2 \\ &= \alpha^{1/2} Z_2 \Lambda_2^{-1/2} Y_2 \\ X_1 &= \alpha^{1/2} X_2 \end{aligned}$$

o

2.2.9.1. / Equivalencias fuertes.

Equivalencia (ii) : $X_1 = X_2$ (se tiene simultáneamente las equivalencias i y io).

Equivalencia (iii) : $X_1 = \alpha^{1/2} X_2$ (se tiene simultáneamente las equivalencias ii y u).

3. / Inter-Intra estructuras y acuerdos.

3.1. / Descripción del problema.

Se plantea la necesidad de obtener representaciones globales de las tablas en conjunto, a partir de los elementos característicos de cada tabla, con los cuales se represente gráficamente a esas distancias, nuevamente aquí se está pensando en que se puede representar a una tabla en un cierto espacio euclídeo de dimensión finita (Ver (1)). Esta etapa es conocida como *Interestructura*.

En la práctica son limitados los recursos con que se cuenta al realizar un estudio por ello se desea también resumir las K tablas en una sola o un número reducido de tablas, que explique el comportamiento del conjunto original de tablas para hacer posible esto se requiere de cierto criterio de optimalidad, según los casos considerados, el objetivo del estudio y el criterio deducido se obtendrá muchas maneras de convenir. Esta etapa se conoce como etapa de *Acuerdos*.

Se busca representar simultáneamente a todos los individuos, o a todas las características utilizadas, de manera que sea posible observar su evolución gráficamente, esto es útil puesto que proporciona indicios de la tendencia en la evolución respecto a un parámetro (puede ser el tiempo) de dichos individuos (o características). Esta etapa es conocida como la *Intraestructura*.

3.2. / Interestructura.

Con objeto de comparar globalmente las k tablas que intervienen en el estudio se construye una matriz de productos escalares \mathcal{E} de $K \times K$, cuyas entradas son los productos escalares de los elementos característicos de las tablas. Los eigenvectores de \mathcal{E} proporcionan entonces una representación global, de la misma

manera que lo hacía el ACP en el caso de una sola tabla de datos. Nuevamente la dualidad está presente en los tratamientos alternativos con individuos o con características.

3.2.1. / Construcción de \mathcal{E} .

Los elementos característicos según los casos considerados son: matrices de productos escalares entre individuos, las matrices de proyectores, las matrices de covarianzas, las tablas de Burt o las tablas de datos iniciales. En cada caso son matrices asociadas a aplicaciones lineales definidas entre espacios de dimensión finita que sin embargo se pueden comparar (Ver apéndice 5). en cada caso se construye la entrada (k,l) de la matriz \mathcal{E} .

Cuando se trata de una matriz de productos escalares entre individuos se tiene que:

$$c_{kl} = \text{tr}(W_k D W_l D).$$

Cuando se trata de matrices de proyectores se tiene:

$$c_{kl} = \text{tr}(A_k A_l).$$

Cuando se trata de matrices de covarianzas V_k :

$$c_{kl} = \text{tr}(V_k H V_l H).$$

Cuando se trata de tablas de Burt H_k :

$$c_{kl} = \text{tr}(H_k H_l).$$

Cuando se trata de tablas iniciales de datos:

$$c_{kl} = \text{tr}(X_k' D X_l' H).$$

Más precisamente se tiene:

$$\text{tr}(W_k D W_l D) = \text{tr}(X_k' H_k X_k' D X_l' H_l X_l' D)$$

Proposición: $\text{tr}(W_k D W_l D) \geq 0$ y

$\text{tr}(W_k D W_l D) = 0 \Leftrightarrow X_k' D X_l = 0$, en este caso los característicos de las tabla X_k y X_l son no correlacionadas.

Demostración:

De acuerdo con la decomposición descrita en el apéndice 3 H_k y H_l se pueden escribir como $H_k = I_k I_k'$ y $H_l = I_l I_l'$ donde las matrices I_k y I_l son cuadradas de rango p_k y p_l respectivamente.

Además $\text{tr}(W_k D W_l D) = \text{tr}(I_k' X_k' D X_l I_l' X_l' D X_k I_k)$

$$= \sum_{k=1}^{P_1} \sum_{l=1}^{P_1} (L_k' X_l' D X_l L_k)' \geq 0$$

y se verifica que : $\text{tr}(M_k D M_k D) = 0$ ssi $L_k' X_k' D X_l L_k = 0$
 ssi $X_k' D X_l = 0$.

MIZERE desarrolla en (11) algunos casos particularmente importantes:

- Si $M_k = I_{P_k}$ y $M_l = I_{P_l}$ entonces se tiene

$$\text{tr}(M_k D M_l D) = \sum_{k=1}^{P_k} \sum_{l=1}^{P_l} \text{Cov}^2((X_k')^i, (X_l')^j).$$

- Si $M_k = D_{1, \sigma_k^2}$ y $M_l = D_{1, \sigma_l^2}$ matrices diagonales

inversas de las matrices de varianzas, entonces :

$$\text{tr}(M_k D M_l D) = \sum_{k=1}^{P_k} \sum_{l=1}^{P_l} \text{cor}^2((X_k')^i, (X_l')^j)$$

Si X_k es de rango p_k y X_l es de rango p_l , se tiene

$$\text{tr}(A_k A_l) = \text{tr}(X_k' (X_k' D X_k)^{-1} X_l' D X_l (X_l' D X_l)^{-1} X_l' D)$$

Si no, de cualquier manera se tiene siempre :

$$\text{tr}(A_k A_l) = \text{tr}(Y_k' A_k^{-1} Y_l D Y_l' A_l^{-1} Y_l D)$$

Proposición : $\text{tr}(A_k A_l) \geq 0$ y $\text{tr}(A_k A_l) = 0$ ssi $X_k' D X_l = 0$.

Demostración :

En efecto :

$$\text{tr}(A_k A_l) = \text{tr}(A_k^{-1} Y_k D Y_k' A_l^{-1} Y_l D Y_l') = \sum_{k=1}^{P_k} \sum_{l=1}^{P_l} \text{cor}^2((Y_k)^i, (Y_l)^j).$$

la suma de los cuadrados de los coeficientes de correlación entre las componentes principales de dos estudios. Por lo tanto

$$\text{tr}(A_k A_l) \geq 0 \text{ y } \text{tr}(A_k A_l) = 0 \Leftrightarrow Y_k D Y_l' = 0.$$

Pero puesto que $X_l' = U_l' Y_l$ para $l \in (k, l)$ entonces se tiene :

$$Y_k' D Y_l' = 0 \Leftrightarrow X_k' D X_l = U_k' Y_k D Y_l' U_l' = 0.$$

y como

$$Y_l = U_l' M_l X_l \quad \text{para } l \in (k, l)$$

$$X_k' D X_l = 0 \rightarrow Y_k D Y_l' = U_k' M_k X_k' D X_l M_l U_l' = 0$$

$$\Leftrightarrow Y_k D Y_l' = 0 \Leftrightarrow X_k' D X_l = 0$$

lo cual significa que los espacios vectoriales de F , Y_k y Y_l son D-ortogonales si y solo si las características de las dos tablas son no correlacionadas.

Se tiene también que :

$$\text{tr}(V_k M V_l M) = \text{tr}(X_k' D X_k M X_l' D X_l M)$$

Proposición : $\text{tr}(V_k M V_l M) \geq 0$ y $\text{tr}(V_k M V_l M) = 0$ ssi $X_k M X_l' = 0$, en el caso de los individuos generados por los subespacios vectoriales M-ortogonales de E.

Se tiene de la misma manera :

$$\text{tr}(B_k B_l) \geq 0 \text{ si } B_k \text{ y } B_l \text{ son tablas de Burt.}$$

Proposición : La matriz \mathcal{E} es definida positiva.

Prueba :

Sea 0_k el objeto característico del k-ésimo estudio, y ψ el producto escalar entre objetos :

$$c_{kl} = \langle 0_k, 0_l \rangle_\psi$$

Considere $w' = (w_1, \dots, w_k)$ por lo tanto :

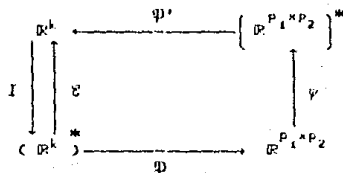
$$w' \mathcal{E} w = \sum_{k=1}^k \sum_{l=1}^k w_k w_l \langle 0_k, 0_l \rangle_\psi = \left\| \sum_{k=1}^k w_k 0_k \right\|_\psi^2 \geq 0.$$

3.2.2. Descripción de los estudios para características.

Cada O_k es una matriz $p_1 \times p_2$, a la cual es posible asociar un vector \tilde{O}_k en $R^{p_1 \times p_2}$ obtenido yuxtaponiendo las columnas de O_k . Se nota que ψ la métrica en $M_2 \otimes M_1$, donde M_i es la métrica sobre $E_i = R^{p_i}$, para $i=1,2$.

Denote por Ψ' la tabla $k \times p_1 \times p_2$; $\Psi' = \begin{pmatrix} O_1 \\ \vdots \\ O_k \end{pmatrix}$. De donde se

tiene el esquema:



Se tiene $\mathcal{E} = \Psi' \psi \Psi$. La matriz \mathcal{E} juega un papel similar al que tiene la matriz de covarianzas de las k "características" O_k en el ACP para una tabla de datos. Se tiene que $\mathcal{E} = Y Y'$, donde Y es una matriz $k \times r$ (donde $r = \text{rang}(\mathcal{E})$) cuyas columnas son precisamente los eigenvectores de \mathcal{E} , con norma la raíz cuadrada de su eigenvalor asociado. Los renglones de Y proporcionan una representación de los k estudios análogo a las características en el ACP; cada estudio está representado por un vector cuya norma es $\|O_k\|_{\psi}$, y nuevamente se cuenta con la herramienta necesaria para comparar estudios como en el caso de individuos o características en el esquema de dualidad para una tabla por la comparación a través del ángulo entre estos dos vectores según el producto interno ψ .

Observación:

-Si O_k es M_k o A_k , dos vectores O_k y O_l son ortogonales en el sentido de ψ si las características de X_k y las

de X_l son no correlacionadas.

-Dos estudios equivalentes en los sentidos ψ y ψ' serán representados por dos vectores colineales.

Si se tienen representados de manera idéntica los objetos en el sentido de las equivalencias ψ, ψ' y ψ'' se diagonaliza la matriz de los cosenos entre objetos X_k (no es otra que la normalización de las características):

$$R_{kl} = \frac{\langle O_k, O_l \rangle_{\psi}}{\|O_k\|_{\psi} \|O_l\|_{\psi}}$$

La interpretación de la representación que se obtiene con este procedimiento es análoga a la obtenida en un ACP sobre una matriz de correlaciones donde se elimina el problema de la escala de unidades de las diversas características que entran en el estudio, con la ventaja de que un punto es representativo si es próximo al círculo de radio unitario también conocido como círculo de correlación. Los cosenos de los ángulos entre los vectores asociados a los estudios representativos son próximos a los de R_{kl} que serán según los casos considerado s, de la forma siguiente:

Observación:

- En el caso en que O_k sea del tipo M_k .

$$R_{kl} = \frac{\text{tr}(M_k D M_l D)}{(\text{tr}(M_k D)^2 \text{tr}(M_l D)^2)^{1/2}} = R_M(M_k D, M_l D).$$

- En el caso en que O_k sea del tipo A_k .

$$R_{kl} = \frac{\text{tr}(A_k A_l)}{(\text{tr}(A_k)^2 \text{tr}(A_l)^2)^{1/2}} = \frac{\text{tr}(A_k A_l)}{\sqrt{r_k r_l}}$$

para r_k y r_l los rangos de X_k y X_l respectivamente.

- En el caso en que O_k sea del tipo V_k .

$$R_{kl} = \frac{\text{tr}(V_k M V_l M)}{(\text{tr}(V_k M)^2 \text{tr}(V_l M)^2)^{1/2}}$$

- En el caso en que O_k sea del tipo X_k .

$$K_{11} = \frac{\text{tr}(X_k' D X_k H)}{(\text{tr}(X_k' D X_k H))^2 \text{tr}(X_k' D X_k H)^{1/2}} = \frac{\text{tr}(X_k' D X_k H)}{(\ln(X_k') \ln(X_k'))^{1/2}}$$

donde $\ln(X_k')$ representa la inercia de la nube de individuos en relación a su centro de gravedad $g(c)$.

Observaciones :

- Si O_k es de la forma M_k , entonces $\psi = D \& D$.

- Si O_k es de la forma A_k , entonces $\psi = I$, ψ es por consiguiente una matriz diagonal de pesos, y las "características" O_k están centradas por ψ (cuando X_k está centrada);

$$\sum_{i=1}^r \sum_{j=1}^h \rho_{ij} \rho_{kij} = 0.$$

Por lo tanto \mathcal{E} cumple el papel de la matriz de covarianzas y K la matriz de correlación para las "características" O_k .

- Las anteriores observaciones son válidas si O_k es de la forma X_k' y si $M = I$, en cuyo caso $\psi = D \& I$.

En el mejor de los casos (el 5) donde se trabaja sobre los mismos conjuntos de individuos y de características, debido a que presenta las mejores condiciones pueden compararse los estudios de varias maneras :

- 1 - Construir \mathcal{E}_1 tal que su entrada (k,l) sea :
 $\text{tr}(M_k H V_k H)$.
- 2 - Construir \mathcal{E}_2 tal que su entrada (k,l) sea :
 $\text{tr}(V_k H V_k H)$.
- 3 - Construir \mathcal{E}_3 tal que su entrada (k,l) sea :
 $\text{tr}(X_k' D X_k H)$.

Cualquiera de las representaciones obtenidas por la diagonalización de las matrices \mathcal{E}_i muestra rasgos individuales que entre la naturaleza de los estudios, esto es, en general son distintas estas representaciones una de la otra.

- Dos estudios son colineales por 3 si lo son por 1 y por 2.

- Si dos estudios son ortogonales por 1 o por 2, ellos los son por 3.

\mathcal{E}_1 y \mathcal{E}_2 tienen la misma diagonal :

$$\text{tr}(M_k D)^2 = \text{tr}(X_k H X_k' D X_k H X_k' D) = \text{tr}(X_k' D X_k H X_k' D X_k H) = \text{tr}(V_k H)^2.$$

3.2.3. Descripción de los estudios.

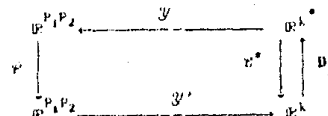
La matriz \mathcal{E} de los simétrica y semidefinida positiva también puede ser considerada como una matriz de productos escalares entre individuos. Para poder establecer un similitud con la matriz M del esquema de dualidad se le centra en relación a la matriz diagonal de pesos, D_k , que puede ser propuesta a priori en cada uno de los estudios, entonces se reemplaza a \mathcal{E} por la matriz :

$$\mathcal{E}^* = (I - \theta \theta' D_k) \mathcal{E} (I - D_k \theta \theta').$$

donde θ es un vector columna en \mathbb{R}^k , con todas las entradas iguales a uno.

Observación :

\mathcal{E}^* genera las mismas distancias entre estudios que la matriz \mathcal{E} , pero además posee una representación centrada. Entonces se tiene el esquema :



donde $\mathcal{E}^* = y' \mathcal{E} y$, donde $y = X(I - D_k \theta \theta') = (\tilde{O}_1, \dots, \tilde{O}_k)(I - D_k \theta \theta')$.

Esto es centrar \mathcal{E} requiere reemplazar cada objeto O_k por el objeto centrado

$$O_k - \sum_{i=1}^k \gamma_i O_i$$

donde los coeficientes γ_i son los pesos asociados al k -ésimo estudio.

Los primeros eigenvectores de \mathcal{E}_k^* de norma igual a la raíz

cuadrada del eigenvalor correspondiente, proveen de una representación de los K estudios. Tal que las distancias entre puntos representativos de los 0_i son las mejores aproximaciones posibles de las distancias

$$\alpha^2(0_k, 0_l) = \|0_k - 0_l\|_{\varphi}^2.$$

3.2.4. / Construcción de los acuerdos.

Se ven resumidos los K objetos $0_1, \dots, 0_k$ por uno solo, considerado como representativo del conjunto y ponderado por su importancia.

Si los pesos $q_i, i = 1, \dots, k$, son otorgados a los estudios a través de un conocimiento a priori es posible elegir como acuerdo

$$\tilde{0} = \sum_{i=1}^k q_i 0_i$$

es decir, el estudio en relación al cual la inercia de la nube de los estudios es la más pequeña. (aquí una tabla o estudio es ahora considerada como individuos o características y se le aplica el tratamiento analizado en el capítulo I de esta tesis, no es necesario hacer muy explícito lo siguiente puesto que es la aplicación de la teoría desarrollada en dicho capítulo).

Propiedad : $\tilde{0}$ realiza el mínimo de
$$\sum_{i=1}^k q_i \|0_i\|^2 - \sum_{i=1}^k q_i 0_i \|\tilde{0}\|^2.$$

Prueba : En efecto :

$$\sum_{i=1}^k q_i \|0_i\|^2 - \sum_{i=1}^k q_i 0_i \|\tilde{0}\|^2 = \sum_{i=1}^k q_i \left[\|0_i\|^2 - \|\tilde{0}\|^2 + \|\tilde{0}\|^2 - \sum_{i=1}^k q_i 0_i \|\tilde{0}\|^2 \right]$$

es mínimo para $q_i = q_i, \forall i$.

En particular cuando todos los estudios son igualmente importantes el peso asignado es $q_i = 1/k, \forall i$, por consiguiente

$$\tilde{0} = \frac{1}{k} \sum_{i=1}^k 0_i$$

es el objeto del acuerdo el promedio de los objetos característicos de los estudios.

Segundo acuerdo :

Otro acuerdo que se propone es $0 = \sum_{i=1}^k l_i 0_i$, donde los

coeficientes $l_i, i=1, \dots, k$, son los elementos del eigenvector de Σ asociado al eigenvalor más grande λ_1 , o se puede pensar también como la primera componente principal del análisis de la terna (X', I, φ) analizada en *****.

Propiedad : Para todo vector $u' = (u_1, \dots, u_k)$ tal que $u'u = 1$, se tiene :

$$i) \quad \left\| \sum_{i=1}^k u_i 0_i \right\|_{\varphi}^2 \leq \|0\|_{\varphi}^2 = \lambda_1.$$

$$ii) \quad \sum_{i=1}^k \left\langle \sum_{j=1}^k u_j 0_j, 0_i \right\rangle_{\varphi}^2 \leq \sum_{i=1}^k \langle 0, 0_i \rangle_{\varphi}^2 = \lambda_1^2.$$

Prueba : En efecto :

Con referencia al capítulo I o al apéndice 6, se sigue la demostración :

-Si $l = (l_1, \dots, l_k)$, entonces realiza el máximo de $\|X' I u\|_{\varphi}^2$ con valor máximo λ_1 .

$-\frac{0}{\sqrt{\lambda_1}}$ realiza el máximo de $\sum_{i=1}^k \left\langle \sum_{j=1}^k u_j 0_j, 0_i \right\rangle_{\varphi}^2$ con

$\sum_{i=1}^k \beta_i^2 = \frac{1}{\lambda_1}$, y el valor máximo es λ_1 , por consiguiente 0

realiza el máximo de $\sum_{j=1}^k \left\langle \sum_{i=1}^k u_i 0_i, 0_j \right\rangle_{\varphi}^2$ restringido a

$\sum_{i=1}^k u_i^2 = 1$, y ese máximo vale λ_1^2, λ_1^2 .

La construcción ofrece muchas otras posibles elecciones para acuerdos por ejemplo se puede tomar otras componentes principales de la terna (X', I, φ) , pero hay que tener precaución al momento de interpretar pues muchos de los acuerdos obtenidos carecen de

$$B = \frac{1}{K} \sum_{l=1}^K B_l$$

Observación: En este caso un planteamiento más general donde se contemplen pesos distintos entre los diversos estudios lleva a una inconsistencia puesto que la Tabla

$$\sum_{l=1}^K l B_l$$

no necesariamente es una Tabla de Burt.

Más han sido propuestos varios acuerdos, por ejemplo FOUCARD propuso:

$$B = \frac{1}{c} \sum_{l=1}^K l B_l \quad \text{con} \quad c = \sum_{l=1}^K l$$

se observa que esta propuesta básicamente busca preservar la convención sobre los pesos de los estudios i.e.

$$q_l = l \left(\sum_{l=1}^K l \right)^{-1}$$

tanto más pequeño sea el peso del estudio considerado menos se le tomará en cuenta.

Las posibles propuestas en este sentido del tipo $\sum_{l=1}^K \alpha_l B_l$ con

$\sum_{l=1}^K \alpha_l = 1, \alpha_l \geq 0$ parecen las más razonables para ser aplicadas como acuerdos.

Proposición: El máximo de $\| \sum_{l=1}^K \alpha_l O_l \|_p^2$, con $\alpha_l \geq 0$ y $\sum_{l=1}^K \alpha_l = 1$ se asume por O_l tal que:

$$\| O_l \|_p^2 = \max (\| O_l \|_p^2 : l=1, \dots, K)$$

Demostración:

En efecto:

$$\| \sum_{l=1}^K \alpha_l O_l \|_p^2 = \sum_{l=1}^K \sum_{l=1}^K \alpha_l \alpha_l \langle O_l, O_l \rangle$$

$$\leq \sum_{l=1}^K \sum_{l=1}^K \alpha_l \alpha_l \| O_l \|_p \| O_l \|_p$$

$$\leq \left\{ \max_{l=1, \dots, K} \| O_l \|_p^2 \right\} \sum_{l=1}^K \sum_{l=1}^K \alpha_l \alpha_l$$

$$= \left\{ \max_{l=1, \dots, K} \| O_l \|_p^2 \right\} \left\{ \sum_{l=1}^K \alpha_l \right\}^2$$

$$= \max_{l=1, \dots, K} \| O_l \|_p^2$$

De aquí se observa que es necesario conservar el objeto de norma máxima. Este es un procedimiento de selección alternativo al propuesto por FOUCARD.

-El caso de las mismas características-mismos individuos. Se considera:

$$X = \begin{pmatrix} X'_1 \\ \vdots \\ X'_k \end{pmatrix}$$

cuya diagonalización produce:

$$W_k D = \sum_{l=1}^k W_l D, \quad \text{donde } V_k = \begin{pmatrix} V_1' X'_1 D X_2 & \dots & X'_1 D X_k \\ \vdots & & \vdots \\ X'_k D X_1 & \dots & V_k \end{pmatrix}$$

Ello cuesta diagonalizar, para estudiar el conjunto de las características, el acuerdo "promedio" entre las W_k .

Sea $X = (X'_1, X'_2, \dots, X'_k)$, con diagonalización $V_k = \sum_{l=1}^k V_l$.

sentido. En este contexto es preferible tener como acuerdo de las O_k un objeto del mismo tipo que puede ser considerado como el objeto característico de un estudio.

En lo que sigue se analizan algunos casos particularmente importantes.

-El caso de las M_k : Puesto que cada elemento de τ es positivo, las componentes de τ son del mismo signo; sin pérdida de generalidad es posible suponerlas a todas positivas. De esta manera O es una combinación lineal cuyos coeficientes son positivos de matrices semidefinidas positivas, así que los dos acuerdos son semidefinidos positivos y centrados por D , y los acuerdos son del mismo tipo M .

Cada elección de los coeficientes $\alpha_1, \dots, \alpha_k$ positivos, se puede hacer corresponder, si se considera en los estudios (X_k, M_k, D) , el esquema de dualidad:

$$\begin{array}{ccc}
 \mathbb{R}^k = E & \xleftarrow{X'} & F^* \\
 \downarrow M & & \downarrow D \\
 E^* & \xrightarrow{X} & F = \mathbb{R}^n
 \end{array}$$

donde

$$X = \begin{pmatrix} X'_1 \\ \vdots \\ X'_k \end{pmatrix} \quad M = \begin{pmatrix} \alpha_1 M_1 & & 0 \\ & \ddots & \\ 0 & & \alpha_k M_k \end{pmatrix}$$

$$W = X' M X = \sum_{i=1}^k \alpha_i X_i M X_i' = \sum_{i=1}^k \alpha_i M_i$$

La segunda componente principal de (X', I, ρ) es también una combinación lineal de las M_k y entre sus coeficientes a veces aparecen algunos negativos (cosa que no pasaba con la primera, pues el teorema de Frobenius garantiza que en ese caso todas son del mismo signo), esta no es por tanto semidefinida positiva.

-El caso de A_k : En general una combinación lineal de proyectores no es un proyector, por tanto los acuerdos obtenidos

en este caso no son necesariamente del tipo A puesto que pueden ser M . En el caso de Análisis Canónico generalizado esto conduce a caracterizar las posiciones relativas de los K espacios vectoriales generados por las columnas de las tablas X_k , si se diagonaliza

$$\sum_{i=1}^k \lambda_i$$

en *****) y *****) generaliza el Análisis Canónico de dos tablas que pueden ser consideradas como el ACP de la terna:

$$\left[\begin{pmatrix} X'_1 \\ X'_2 \end{pmatrix}, \begin{pmatrix} (X'_1 D X_1)^{-1} & 0 \\ 0 & (X'_2 D X_2)^{-1} \end{pmatrix}, D \right]$$

Los eigenvectores normalizados de $\sum_{i=1}^k \lambda_i$ realizan

sucesivamente el máximo de $\sum_{i=1}^k \| \lambda_i X_{iD} \|^2$, restringido a $x'Dx=1$.

Esta técnica aplicada a el caso 6 produce el Análisis de Correspondencias Múltiples, y ayuda a describir las relaciones entre las K características cualitativas medidas sobre los mismos individuos.

Si se utilizara el segundo acuerdo propuesto entonces hay que realizar el Análisis de Correspondencias de la tabla

$$\begin{pmatrix} \sqrt{\alpha_1} X'_1 \\ \vdots \\ \sqrt{\alpha_k} X'_k \end{pmatrix}$$

(consultar *****)).

El caso de X_k : Este caso ha sido trabajado por P.A. JAFFRENOU y considera todas las componentes principales de la terna (X', I, ρ) ; de los elementos μ_{pkn} , es decir, obtiene tablas del mismo tipo que X_k centradas por D , que además son base del subespacio imagen de \mathbb{R}^k bajo el funcional lineal X .

El caso de las tablas de Burt: es posible definir aquí un acuerdo como el promedio de las tablas de contingencia:

$$\text{donde } W_X \left(\begin{matrix} D \\ \cdot \\ D \end{matrix} \right) = \begin{bmatrix} M_1 D & X_1 X_1' D & \dots & X_1 X_k' D \\ X_k' X_1 D & \dots & \dots & M_k D \end{bmatrix}$$

esto requiere estudiar el conjunto de individuos, de diagonalizar (considerando $\frac{1}{K}$ aproximadamente) el acuerdo promedio entre las matrices de covarianzas V_k .

-Sea $X = \sum_{i=1}^k X_i'$, el acuerdo "promedio" entre la tablas

iniciales mediante el cual se debe diagonalizar:

$$ND = \frac{1}{K} \sum_{i=1}^k \sum_{l=1}^k (X_k X_l' D) \quad \text{o} \quad V = \frac{1}{K^2} \sum_{i=1}^k \sum_{l=1}^k (X_i' D X_l)$$

Los acuerdos se proponen de acuerdo con los objetivos del análisis y en la práctica es posible contar con información a priori que le proporciona el sentido en algunas aplicaciones.

Observaciones:

-Es posible observar las posiciones de los estudios sobre un plano principal generado en la INTERESTRUCTURA.

-La proyección sobre el plano del baricentro y los puntos representativos de los estudios, afectados por los pesos q_i . En ese caso \bar{O} se localizará en el origen del plano principal, y la norma del vector representativo de un individuo O_k es una aproximación de la distancia real entre \bar{O} y O_k .

-La primera componente principal, está proyectada sobre el primer eje y permite ver cuales son las tablas lo más correlacionadas posibles con \bar{O} , en el sentido de la métrica p .

3.3. / La Infraestructura.

La diagonalización de los acuerdos produce en el caso de M_k y A_k una representación entre los individuos intermedia entre las que se puedan proporcionar en cada estudio, de manera semejante en el caso de las matrices de covarianzas V_k se proporciona una representación promedio de las características.

En el caso de los D_k , el Análisis de correspondencias de D proporciona una representación promedio de las modalidades de las características (Ver (12)). En resumen se presenta el problema de representar en un mismo plano todas las características, todos los individuos o todas las modalidades de las características cualitativas. Se mostrará en lo siguiente una posibilidad para resolver este problema.

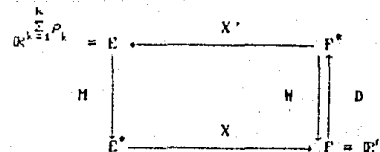
3.3.1. / Caso donde se tiene los mismos individuos.

Si se presentan los casos 1,2,5 ó 6. O_k es de la forma M_k o A_k . El tratamiento aquí expuesto se desarrollará solo para el caso 1, en donde se tiene $O_k = M_k$, entonces si

$$W = \sum_{j=1}^k \alpha_j W_j$$

es posible adaptar el mismo tratamiento a los demás casos y para $O_k = A_k$.

Se ha visto que el estudio de acuerdos lleva a considerar el esquema de dualidad:



donde:

$$X = \begin{pmatrix} X'_1 \\ \vdots \\ X'_k \end{pmatrix}, \quad M = \begin{pmatrix} \alpha_1 M_{11} & & \\ & \ddots & \\ & & \alpha_k M_{kk} \end{pmatrix}, \quad \text{las } \alpha_k \text{ son}$$

números positivos.

3.3.2. / Representación de las características.

Si se efectúa el ACP de la terna (X, M, D) , una representación de las características es dada naturalmente por:

$$\tilde{Z} = X' D Y' \Lambda^{-1/2}$$

donde Y es la matriz que contiene en sus renglones a las componentes principales, y $\Lambda = YDY'$.

$\tilde{Z}_k = X'_k D Y' \Lambda^{-1/2}$ representa las características del k -ésimo estudio, en la base D -ortonormal de los renglones de $\Lambda^{-1/2} Y$.

En base a esto, se puede representar las componentes principales de cada estudio por:

$$\tilde{Y} = Y_k D Y' \Lambda^{-1/2}$$

(Se puede en efecto considerar que $W = \sum_{j=1}^k \alpha_j W_j = \sum_{j=1}^k \alpha_j Y'_j Y_j$, se tiene a las mismas matrices Y y Λ que para el análisis de:

$$\left[\begin{pmatrix} Y_1 \\ \vdots \\ Y_k \end{pmatrix}, \begin{pmatrix} \alpha_1 I & & \\ & \ddots & \\ & & \alpha_k I \end{pmatrix}, D \right],$$

Si se guarda la matriz Y completa, se tiene se tiene así una representación exacta de todas las características y de todas las componentes principales, además las covarianzas entre las características se conservan:

$$\tilde{Z}_k D \tilde{Z}'_k = V_k \quad \text{y} \quad Y_k D Y'_k = \Lambda_k.$$

Si sólo se guardan los dos primeros renglones de Y , para tener una expresión plana, se tiene la mejor representación plana simultáneamente de todas las características (la mejor fotografía de la nube de características).

Observación:

-Esta representación plana es utilizada en el método *STATIS*, en dicho programa se obtiene una tal representación plana de la nube de características (individuos) cuya utilidad se puede observar en el algoritmo descrito en el capítulo II para recuperar representaciones gráficas planas a partir de matrices de similitudes.

En el caso ó (X_k es una tabla de características indicatrices), se puede así representar simultáneamente todas las modalidades de las características cualitativas consideradas.

3.3.3. / Representación de los individuos.

Esta sección bosqueja la representación simultánea de los n_k individuos, vistos para los K estudios, de manera que se pueda seguir la evolución de los individuos a través de los estudios.

En *STATIS* el desarrollo de H. L'HERMIER DES PLANTES, se presenta a los individuos por las columnas de:

$$\tilde{Y}_k = (\Lambda^{-1/2} Y D Y'_k \Lambda_k^{-1/2}) Y'_k.$$

Esto es justificado por la falta de los renglones de $\Lambda^{-1/2} Y$ para formar una base D -ortonormal de un subespacio vectorial de E^n generado por los renglones de todas las Y_k .

$\Lambda^{-1/2} Y D Y'_k \Lambda_k^{-1/2}$ representa el coseno entre las características de la tabla Y_k y estas tablas de Y . \tilde{Y}_k representa por consiguiente las características de Y_k en un espacio que engloba todas las componentes principales de todos los estudios.

Si se desea registrar conjuntamente a todos los individuos y nuevas características (componentes principales) de Y_k , se obtiene, para cada tabla, las mismas posiciones relativas de las características y de los individuos que aquellas de la representación conjunta resultante del ACP de la terna (X_k, M_k, D) .

a condición de guardar todas las componentes de Y.

Cuando las características son diferentes de un estudio a otro, parece difícil de comparar, para un individuo dado, los diferentes valores que adoptan las características en cualquier otro individuo, puesto que de un estudio a otro puede variar la presencia o ausencia de la característica.

3.3.4. / Caso donde se tienen las mismas características.

Si está en los casos 3.4 ó 5. Este análisis es el caso dual del precedente pues debido al esquema de dualidad considere las tablas transpuestas con

$$\begin{bmatrix} D_1 & & & 0 \\ & D_2 & & \\ & & \ddots & \\ 0 & & & D_k \end{bmatrix}$$

como matriz de pesos,

-los individuos son representados por :

$$\psi_k = \Delta^{-1/2} Z' M X_k', \text{ donde } Z \text{ es tal que } V = \sum_{i=1}^k V_i = Z Z' \text{ y } Z' M Z = \Delta.$$

-para la representación de las características, se tiene las mismas dificultades que para la representación de los individuos en el caso precedente.

3.3.5. / Caso donde se tienen los mismos individuos y las mismas características.

Si se encuentra en los casos 5 ó 7. Una buena representación de las características la proporcionan las Z_k y una representación de los individuos aceptable la proporcionan las ψ_k .

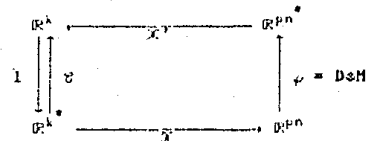
No parece haber relación entre las dos representaciones puesto que una describe el estudio de

$$\begin{bmatrix} X'_1 \\ \vdots \\ X'_k \end{bmatrix}$$

la otra al de : $\left\{ X'_1, \dots, X'_k \right\}$.

Observación :

Si $0_k = X'_k$, es decir si se considera como elemento característico de X'_k a ella misma, y si se lleva a cabo el esquema de la interestructura,



-la diagonalización de \mathcal{U} da la representación de las K características en la base de todos los acuerdos (que aquí son todos en un sentido ; mas se puede así representar los "individuos", es decir a la vez las características y los individuos de todas las tablas, con la ayuda de los autovectores de $XX'\phi$, generando con ello una especie de "representación" "promedio" de las características y de los individuos teniendo en cuenta muchos acuerdos y no solamente el primero.

-Se puede efectuar el ACP del primer acuerdo (primera componente principal de \mathcal{U}), proyectando las características y los individuos de todas las tablas en elementos suplementarios.

-En el caso de las tablas de Burt, una "representación" promedio de las modalidades de las características cualitativas es proporcionada por el Análisis de Correspondencias del acuerdo :

$$B = \sum_{i=1}^k \alpha_i B_i$$

A través de la proyección sobre elementos suplementarios, las leyes condicionales de cada tabla de contingencia, se tiene una representación de las modalidades tales que ellas son vistas para todos los estudios.

4. Tratamiento alternativo para características cualitativas.

Consideraremos tres variables aleatorias cualitativas X, Y, Z definidas sobre un mismo espacio de probabilidad (G.a.P.). Nosotros trabajaremos sobre el cubo de contingencia

$$N = (n_{ijk}) \begin{matrix} i \in I \\ j \in J \\ k \in K \end{matrix}$$

construida con la ayuda de esas tres variables cualitativas X, Y y Z cuyos conjuntos de modalidades respectivos I, J y K son tales que card I = n, card J = p y card K = m, donde n_{ijk} es la medida de la cardinalidad del conjunto

$$\{(\omega) / X(\omega)=i, Y(\omega)=j, Z(\omega)=k\}$$

En este artículo nos proponemos analizar la tabla N_k , mostraremos que la elección de la métrica apropiada permite medir no solamente la dependencia entre dos variables con la ayuda de una tabla marginal, mas también entre las tres. Para hacer eso descompondremos el cubo N fijando el conjunto K obteniendo así m tablas rectangulares de contingencia.

$$N_k = (n_{ijk}) \begin{matrix} i \in I \\ j \in J \end{matrix}$$

cada una de las N_k mide la dependencia de X e Y cuando Z es fijo e igual a k. Definimos :

$$p_{ijk} = \frac{n_{ijk}}{n_{...k}} \quad \text{donde} \quad n_{...k} = \sum_{i,j} n_{ijk}$$

$$p_{i..k} = \sum_j p_{ijk}, \quad p_{.jk} = \sum_i p_{ijk}, \quad p_{i.j.} = \sum_k p_{ijk}$$

$$p_{...k} = \sum_i p_{i..k}, \quad p_{.j.} = \sum_i p_{i.j.}, \quad p_{i..} = \sum_k p_{i..k}$$

El análisis de correspondencias de la tabla N_k es equivalente

(cf. BENZECRI (1)) a el análisis en componentes principales (ACP) de la tabla de los perfiles X_k donde los vectores columnas

$$X_k^{i,j} = (p_{ijk} / p_{.jk} : j \in J) \quad (i \in I)$$

miden la frecuencia condicional de ocurrencia de las modalidades j de Y, haciendo (X=i) y (Z = k). Cada uno de esos perfiles

tiene un peso $\frac{p_{.jk}}{p_{...k}}$ ($i \in I$) y puede ser representado en el espacio

vectorial euclideo de dimensión p por un punto. Las $(X_k^{i,j} / i \in I)$ lo que se acostumbra llamar la nube de "individuos"

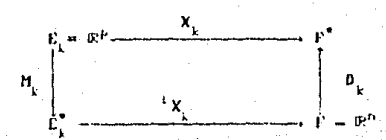
$$G_k = (X_k^{i,j} \cdot \frac{p_{.jk}}{p_{...k}} / i \in I)$$

los pesos $\frac{p_{.jk}}{p_{...k}}$ dependen a priori del indice k y miden la frecuencia condicional de la modalidad i de X para (Z = k).

Esa nube de centros de gravedad $G_k^{i,j} = (\frac{p_{.jk}}{p_{...k}} / j \in J)$ está contenida en el espacio euclideo

$$E_k = (E^p, H_k)$$

El análisis de la tabla X_k es resumido por el esquema de dualidad :



Analizar las proximidades de la nube de individuos G_k necesita la elección de :

una métrica sobre E_k ; nos proponemos utilizar la H_k métrica diagonal de KHI - 2 denotada así X_k^2 y definida por la matriz diagonal

$$X_k^2 = \text{diag} \left(\frac{p_{.jk}}{p_{.jk}} : j \in J \right)$$

por construcción la métrica X_k^2 depende a priori del indice k.

una métrica sobre F , D_k la métrica diagonal de pesos

$$D_k = \text{diag} \left(\frac{p_{1..k}}{p_{..k}} ; i = 1 \right)$$

se hace entonces el Análisis en Componentes Principales de la tabla X_k , D_k -centrada.

La inercia de la nube η_k está dada por :

$$(1) \quad I_k = \sum_{(i,j)} \frac{\left(\frac{p_{ijk}}{p_{..k}} - \left(\frac{p_{i..k}}{p_{..k}} \right) \left(\frac{p_{.jk}}{p_{..k}} \right) \right)^2}{\left(\frac{p_{i..k}}{p_{..k}} \right) \left(\frac{p_{.jk}}{p_{..k}} \right)}$$

en efecto :

$$I_k = \sum_{(i,j)} \frac{p_{ijk}}{p_{..k}} |k \Delta^{ij} - z^{ij}|^2 x_k^2$$

$$= \sum_{(i,j)} \left(\frac{p_{i..k}}{p_{..k}} \right)^2 \left(\frac{p_{.jk}}{p_{..k}} - \frac{p_{.jk}}{p_{..k}} \right)^2 x_k^2 \frac{1}{\left(\frac{p_{i..k}}{p_{..k}} \right) \left(\frac{p_{.jk}}{p_{..k}} \right)}$$

$$= \sum_{(i,j)} \frac{\left(\frac{p_{i..k}}{p_{..k}} \cdot \frac{p_{.jk}}{p_{..k}} - \left(\frac{p_{i..k}}{p_{..k}} \right) \left(\frac{p_{.jk}}{p_{..k}} \right) \right)^2}{\left(\frac{p_{i..k}}{p_{..k}} \right) \left(\frac{p_{.jk}}{p_{..k}} \right)}$$

I_k permite medir la dependencia de las variables X o Y condicionalmente con $(Z = K)$.

Entonces cuando m sea muy grande, es fastidioso hacer un análisis en Componentes Principales y la síntesis es bastante difícil, este método de análisis no presenta un interés práctico.

2.1 - Análisis de la nube global.

Denotaremos por $\eta = (X \Delta^{ij} / k \in K, i \in I) = \bigcup_k \eta_k$ a la nube global de "individuos".

Esa descripción tiene algunos problemas en la elección de la métrica de hecho las nubes η_k ($k \in K$) son descritas cada una con la ayuda de métricas diferentes $x_k^2(k \in K)$. Sin embargo construimos la nube

$$\eta_k = (X \Delta^{ij} / i \in I, j \in J)$$

de los tensores X_k^i (i.e.I) de matriz bloque

$$X_k^i = \begin{bmatrix} \dots & 0 & \dots \\ \dots & \vdots & \dots \\ \dots & 0 & \dots \\ \dots & \vdots & \dots \\ \dots & 0 & \dots \\ \dots & \vdots & \dots \\ \dots & 0 & \dots \\ \dots & \vdots & \dots \\ \dots & 0 & \dots \end{bmatrix} \begin{array}{l} \leftarrow \text{primer vector} \\ \\ \\ \leftarrow k\text{-ésimo vector} \\ \\ \\ \leftarrow m\text{-ésimo vector} \end{array}$$

X_k^i está incluido en el espacio vectorial $G \subseteq E$

($G \subseteq E^m$) que se descompone en suma directa $G \subseteq E = \bigoplus_k E_k$ entonces se define sobre $G \subseteq E$ la métrica M_k , que escrita en forma de bloque

$$M_k = [M_{kk}]_{k \in K}$$

Se verifica (cf. MIZEKE [3] pp 61) que se tiene

$$\langle X_k^i, X_k^j \rangle_{M_k} = \langle X_k^i, X_k^j \rangle_{M_{kk}}$$

donde M_{kk} es llamada la métrica bloque intra k . Se verifica además que las imágenes euclidianas (Pages [4] pp 222)

$$\left(\left(X_k^i, \frac{p_{i..k}}{p_{..k}} \right) / i \in I, \bigoplus_k E_k, M_k \right) \quad \forall$$

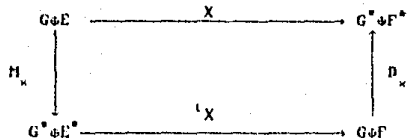
$$\left(\left(X_k^i, \frac{p_{i..k}}{p_{..k}} \right) / i \in I, E_k, M_{kk} \right) \quad \text{para } M_{kk} = p_{..k} X_k^i X_k^i$$

son equivalentes.

La nube global $\eta = \bigcup_k \eta_k$ es una representación euclidiana de las columnas de la tabla X diagonal por bloques definida por

$$X = \begin{bmatrix} X_{11} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & X_m \end{bmatrix}$$

donde los renglones son elementos del espacio vectorial euclideo $(G \oplus F, D_k)$. El análisis de la tabla X se resume en el esquema de dualidad



Ello muestra que la descripción de la nube η depende de la elección de las métricas M_k y D_k :

HIZERE propone las siguientes métricas en su artículo "L'ASPECT CHOIX DE METRIQUES DANS L'ANALYSE D'UN CUBE DE CONTINGENCE":

Elección de la métrica: M_k y D_k son bloques diagonales:

$$M_{kk} = 0, \quad D_{kk} = 0 \quad \text{para } k \neq k'$$

$$M_{kk} = p_{\cdot\cdot k} X_k^2, \quad D_{kk} = D_k$$

con la cual la inercia de la nube η se escribe como:

$$I = \sum_k p_{\cdot\cdot k} X_k^2$$

$$= \sum_{i,j;k} \frac{\left\{ p_{ijk} - \frac{p_{i\cdot k} p_{\cdot j k}}{p_{\cdot\cdot k}} \right\}}{\frac{p_{i\cdot k} p_{\cdot j k}}{p_{\cdot\cdot k}}}$$

La estadística I proporciona una medida de la dependencia de las características X e Y condicionalmente a Z. Ese resultado posiblemente mucho más interesante que el ACP de m tablas X_k ($k \in E$) resulta de un análisis global de las tablas X_k . Una fórmula de

restitución de los datos resultado directamente del método descrito en tal artículo es:

$$\forall i, j, k \quad p_{ijk} = \frac{p_{i\cdot k} p_{\cdot j k}}{p_{\cdot\cdot k}} \left(1 + \sum_{\alpha} F_{\alpha}(i, j, k) \alpha_{ijk} / p_{\cdot\cdot k} \right) \quad (4)$$

donde α_{ijk} es el término genérico del i -ésimo factor η^{α} del ACP de (X, H, D) . En efecto los α -ésimos ejes y factores principales de este ACP, denotados respectivamente por \underline{U}^{α} y η^{α} , se escriben como matrices de bloque:

$$\underline{U}^{\alpha} = \begin{bmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ k & \dots & 0 \\ \vdots & \ddots & \vdots \\ m & \dots & 0 \end{bmatrix} \quad \text{y} \quad \eta^{\alpha} = \begin{bmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ k & \dots & 0 \\ \vdots & \ddots & \vdots \\ m & \dots & 0 \end{bmatrix}$$

los tensores \underline{X}^{α} se escriben para toda i, j, k en la base de las \underline{U}^{α} ($\alpha = 1, \dots, mp$)

$$(5) \quad \underline{X}^{\alpha} = \sum_k F_{\alpha}(k, i) \underline{U}^{\alpha} \quad \eta^{\alpha} = H_k \underline{U}^{\alpha}$$

por lo cual es un hecho de la definición de H_k que:

$$k \underline{U}^{\alpha} = p_{\cdot\cdot k} X_k^2 \underline{U}^{\alpha}$$

de aquí se obtiene:

$$\alpha_{ijk} = p_{\cdot\cdot k} \times \frac{p_{\cdot\cdot k}}{p_{\cdot j k}} U_{ijk}$$

$$\text{y} \quad U_{ijk} = \frac{p_{\cdot j k}}{p_{\cdot\cdot k}} \times \frac{1}{p_{\cdot\cdot k}} \alpha_{ijk}$$

la relación (5) implica la igualdad:

$$\frac{p_{i\cdot k}}{p_{\cdot\cdot k}} \times \frac{p_{\cdot j k}}{p_{\cdot\cdot k}} = \sum_{\alpha} F_{\alpha}(k, i) \frac{p_{\cdot j k}}{p_{\cdot\cdot k}} \times \frac{1}{p_{\cdot\cdot k}} \alpha_{ijk}$$

y se asegura la fórmula (4).

EJEMPLO :

Métodos que separan una variación ligada al tiempo de una estructura independiente del tiempo.

Introducción :

Es posible distinguir :

I. / El método se puede interpretar a partir de la descomposición de la inercia total entre inercia inter e intra, que es la manera de proceder en el fondo en análisis discriminante;

$$V = B + W.$$

Entonces el conjunto J (el conjunto de las características) sirve de espacio de referencia y es necesario que las características obtenidas sobre el individuo sean las mismas a cada instante. La tabla de dos dimensiones considerada aquí tiene sus vectores renglón en \mathbb{R}^J (\mathbb{R}^1) asociados a las parejas (i,t) ((j,t) respectivamente) considerados bajo la relación individuo-instante (característica-instante).

El factor estructurante en los grupos puede ser entonces :

I.1. / El factor individuo (característica).

Entonces el conjunto T puede depender de I (J) (los registros temporales pueden ser hechos a instantes diferentes para cada elemento $t \in (j \in J)$). Un grupo G_i (G^j) está entonces constituido por las parejas (i,t) ((j,t)), donde t describe el tiempo en que se observó a los individuos (características). Su baricentro g_i tiene por coordenadas los promedios de las características obtenidas por el individuo i (por la característica j para g^j).

El análisis de la inercia inter (o ACP de una nube de baricentros g_i (g^j)) ((g^j)) , requiere entonces efectuar el análisis de las posiciones promedio de los individuos (características) que tienen una estructura independiente del tiempo, el análisis de la inercia intra requiere del estudio de

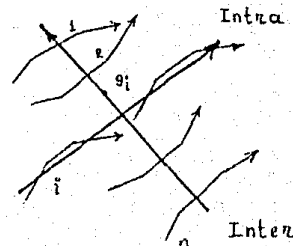
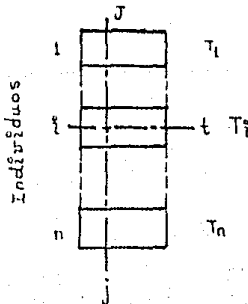
las trayectorias individuales. Se busca efectuar un análisis de las evoluciones (que se expone más adelante) y el análisis de las diferencias entre varias tablas de frecuencias.

I.1.1. / El factor tiempo.

Aquí puede ocurrir que el conjunto I dependa de T (el conjunto de los períodos en los cuales se realiza el estudio, más puede ocurrir que los individuos sean distintos). Los baricentros tienen por coordenadas los valores promedio de las características a cada instante. El estudio de la métrica de la inercia inter permite analizar la evolución promedio, el de la matriz intra el estudio de las relaciones entre individuos-características sin tener en cuenta la influencia del tiempo. Se busca el doble análisis en Componentes Principales propuesto por Bourroche en 1975. Esto es todavía el estudio de las interacciones de primer orden de T y J en el análisis inter, de I y J y de segundo orden entre I,J y T en el análisis intra.

I.1.2. / La elección de la métrica en \mathbb{R}^J .

En los incisos precedentes se precisa de una métrica en \mathbb{R}^J , para el objetivo del estudio. Entonces la elección de dicha métrica depende del conocimiento a priori que el investigador conjuntamente con el interesado tengan del fenómeno, una posibilidad es la de Mahalanobis (definida por la matriz V^{-1} en la base canónica de \mathbb{R}^J) (adaptada al objetivo del estudio) en ese caso los análisis de la inercia inter coinciden con los análisis discriminantes, y los grupos quedan definidos como se muestra a continuación :



Esto no toma en cuenta el tamaño de una ciudad al tiempo t . La j -ésima coordenada del renglón Y_t^j es entonces :

$$Y_t^j = x_t^j / x_t^i$$

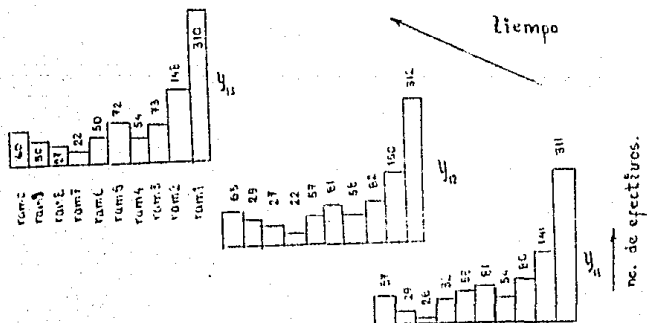
(donde un punto reemplaza a un índice, resumiendo a través de esta notación la suma sobre este índice de los elementos de la Tabla cúbica X). Entonces los pesos ρ de un renglón (t,t) pueden ser elegidos proporcionalmente a la entrada del renglón.

Si se busca estudiar la evolución, no solamente estructural, más así cuantitativa del comercio, es necesario hacer uso del registro en corte de perfil horizontal (sin hacer desaparecer el registro de porcentajes de la figura precedente)

$$Y_t^j = x_t^j / x_t^i$$

Aquí la entrada x_t^j es dividida por la suma $x_t^i = \sum_{j=1}^n x_t^j$ de las entradas de las ciudades observadas a cada registro comercial (la suma de la "rebanada" de la Tabla Y_t^j son conducidas como una sola).

En un tal registro, se guarda :



la información relativa a las variaciones de efectivos o número de establecimientos comerciales de la población (en este caso es la población de establecimientos comerciales) de una ciudad al transcurso del tiempo.

II.4. Registro de renglones-perfiles : los resultados.

La Tabla de las contribuciones a la inercia de las ciudades

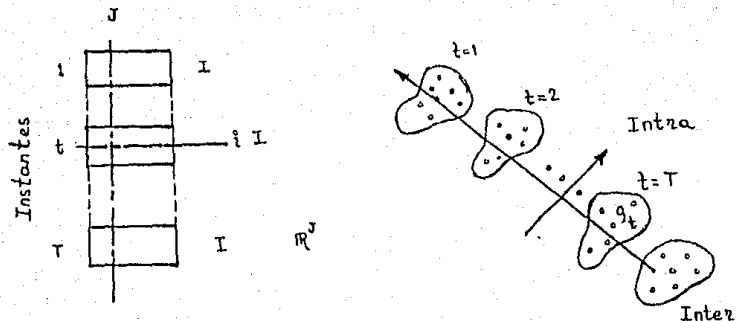
Ind.	ρ_t^j	new	d	eje1	eje2	ρ/d
11	92.44	7.400	.0033	.0468	11.5	
12	4.30	4.000	.0048	.0019	4.43	
13	1.26	37.80	.5446	.3477	34.9	
14	0.91	2.010	.0002	.0501	1.06	
15	0.54	0.200	.0293	.0013	0.10	
16	0.18	14.30	.2788	.4798	13.2	
17	0.13	2.010	.0095	.0073	1.06	
18	0.09	4.020	.0014	.0001	3.72	
19	0.05	0.000	.0309	.0024	0.00	
110	0.03	2.600	.0202	.0036	2.40	
111	0.02	2.800	.0001	.0266	2.59	
112	0.01	1.006	.0300	.0074	0.93	
113	0.01	3.400	.0098	.0124	3.14	
114	0.01	4.620	.0025	.0008	4.30	
115	0.00	7.040	.0076	.0070	6.51	
116	0.00	1.800	.0248	.0061	1.70	
117	0.00	4.100	.0013	.0002	3.00	

Las variaciones más altas se registran en las ciudades

13 ; Monterrey	34.9
16 ; Puebla	13.22
11 ; Cuernavaca	11.55

resultado que es posible observar gráficamente en el dibujo de la página EIII-1 de resultados en *INTRA*. En esa gráfica se aprecia la oposición de la nube de ciudades que excluye a los tres mencionadas en la lista precedente. Dicha gráfica (círculo de correlaciones *intra*) muestra las grandes variaciones de la

o por los instantes como se muestra :



Más los objetivos establecidos y los resultados obtenidos al invertir la definición de los grupos (factos individuo-factor instante) son diferentes, la estructuración individual pone en evidencia las posiciones promedio de los individuos (inter) y permite analizar las trayectorias individuales (intra), la estructuración de los instantes permite estudiar la evolución promedio de los individuos (inter), y ponen en evidencia sus posiciones respectivas "a tiempos constantes".

Se tiene como consecuencia de la relación ;

$$I = BV^{-1} + WV^{-1}$$

el hecho que los eigenvectores asociados a los eigenvalores más grandes de BV^{-1} , obtenidos en el análisis discriminantes (o análisis de inercia inter) son precisamente los eigenvectores de WV^{-1} asociados a sus más pequeños eigenvalores obtenidos en el análisis de la inercia intra.

Así los factores de los análisis inter e intra se obtienen por la diagonalización de una sola matriz y las componentes principales asociadas son no correlacionadas.

Por razones análogas, este resultado es todavía aceptable cuando J es el conjunto de modalidades de una característica cualitativa y la métrica elegida es la de T^2 . (Este punto es interesante comparar con MIZERE).***

$J=(1, \dots, 10)$ para los registros de comercio anuales en los años 1983, 1984 y 1985 (esto es, $T=(1,2,3)$ ó $(83,84,85)$).

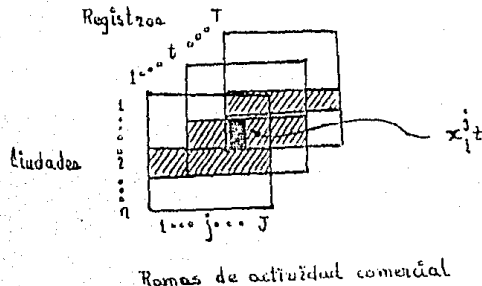
Las características son :

- Rama 1 ; Abarrotes y miscelaneas.
- Rama 2 ; Carnicerías y salchichonerías.
- Rama 3 ; Farmacias.
- Rama 4 ; Ferreterías y tiapalerías.
- Rama 5 ; Materiales para construcción.
- Rama 6 ; Muebles en general.
- Rama 7 ; Panaderías.
- Rama 8 ; Tiendas de ropa.
- Rama 9 ; Vinos y licores.
- Rama 10 ; Zapaterías.

II / Análisis evolutivo y clasificación de trayectorias.

Para ilustrar el método *STATIS* se presenta un ejemplo del Análisis de las evoluciones de una clasificación de trayectorias. Para simplificar, en el caso donde este análisis se reduce a un análisis de la inercia intra, los grupos están divididos por los individuos. (I.I.).

En un primer paso se presentan los datos a analizar, aquí las medidas son sincronizadas y se trabaja sobre los mismos individuos pues el programa impone esta restricción, las observaciones son hechas a los mismos instantes, la estructura de dichos datos es un caso de un cubo de datos como el la figura 1.



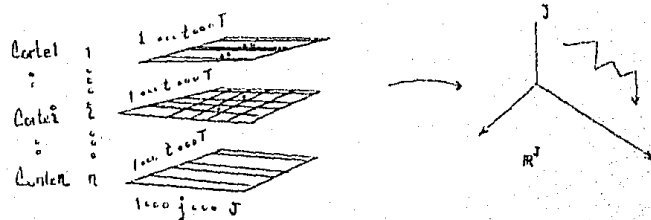
x_{it}^j es el registro comercial al año t de la ciudad i dedicada a la rama de actividad comercial j .

Este cubo de contingencia tridimensional está formado para 17 ciudades de la República Mexicana (conjunto de los individuos $I=(1, \dots, 17)$) cuyas ramas de actividad comercial (en artículos de necesidad primaria) se reparten en 10 características (o

La rama j se observa por ram_j dentro del listado de resultados de *STATIS*.

De modo que x_{it}^j designa la cantidad de establecimientos dedicados a la rama comercial j en la ciudad i durante el año t .

Un corte horizontal en la tabla cúbica de datos X genera una tabla de ramas comerciales y registros anuales para una ciudad :



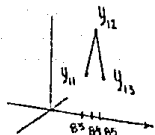
Cada tabla X_i representa el aumento o disminución del registro comercial por ramas de actividad, en la ciudad i .

El objetivo del análisis es estudiar la evolución de las ciudades por rama comercial de 1983 a 1985.

II.1. / Análisis de las evoluciones.

Observación :

$$d(c_1, c_2) = 13.527 \quad d(c_1, c_3) = 13.527 \quad d(c_2, c_3) = 16.1554$$



$$\mathbb{R}^J (= \mathbb{R}^{10})$$

La tabla analizada Y_i^j es la correspondiente a los años 83, 84 y 85 en las ramas comerciales para el individuo i (Cuernavaca , Morelos)

	HAM1	HAM2	HAM3	HAM4	HAM5	HAM6	HAM7	HAM8	HAM9	HAM10	
Y_i^1	1	311	141	80	74	81	52	22	26	29	57
Y_i^2	2	212	150	42	58	81	57	22	27	29	65
Y_i^3	3	310	144	73	54	72	50	22	27	30	60

II.1. / Notaciones.

El análisis de la tabla (Y_i^j) es obtenido por el registro de los cortes de un cubo (X_i^j) = X y se presenta en la figura 3 de la introducción. Los índices de las columnas están asociados a los j ($j \in J$), los de los renglones a las mediciones del individuo i en los tiempos t . $(i,t) \in I \times T$. Los renglones (unidades estadísticas) denotados por Y_i^j , son localizados en el espacio de referencia $\mathbb{R}^J (= \mathbb{R}^{10}$ en este caso). En lo siguiente nos interesaremos en la evolución de las diferentes ciudades, a través de analizar sus posiciones promedio.

II.2. / El análisis de la inercia intra (evolución por individuo).

Se denota por g_i el baricentro del grupo G_i donde los elementos son los individuos (X_i^j ; $t \in T$), en el ejemplo el

baricentro de Cuernavaca es :

$$(311, 146, 9, 78, 3, 55, 3, 78, 53, 3, 22, 26, 6, 29, 3, 60, 6) \in \mathbb{R}^{10} (= \mathbb{R}^J)$$

la j -ésima coordenada denotada por g_i^j es el promedio de la característica Y^j en el grupo i .

En general, la j -ésima coordenada denotada por g_i^j es el promedio de la característica Y^j en el grupo i . Se denota siguiendo la misma notación a g como el baricentro general y por g^j la j -ésima coordenada de este baricentro general g .

El analizar las evoluciones de los individuos (o análisis intra-individual) requiere reemplazar la notación usual de covarianza :

$$\text{Cov}(Y^j, Y^j) = \sum (\rho_{ii} (Y_i^j - g_i^j)(Y_i^j - g_i^j) \mid (i,t) \in I \times T)$$

por la notación de covarianza intra definida por :

$$\text{Cov}_i(Y^j, Y^j) = \sum \left\{ \sum (\rho_{ii} (Y_i^j - g_i^j)(Y_i^j - g_i^j) \mid (t,T) \in I \right\}$$

La correlación intra definido a partir de esta covarianza se interpreta como lo muestra la figura

$$r(x,y) = -0.7$$

correlación total

$$r_i(x,y) = 0.8$$

correlación intra

En lugar de interesarse en la variación general de las características, no se interesa en sus variaciones al interior de las submuestras individuales, siendo todavía "lo largo" de sus trayectorias (ver figura 3 de la introducción).

De este modo el análisis de la inercia intra es todavía un ACP por lo cual la matriz de covarianza total V tiene que ser reemplazada por la matriz de covarianza intra W , donde el elemento (j,j') es definido en (1). Los ejes principales de inercia son definidos por los eigenvectores de W . (Como se vio en el capítulo I).

II.3. / Elección de un sistema de ponderación y elección de la métrica.

Como en el Análisis de Correspondencias es posible apoyarse en propiedades naturales para la elección de métricas y ponderaciones; las propiedades de equivalencia distribucional clásica en Análisis de Correspondencias. En el ejemplo esto significa:

Las ciudades son muy diferentes en cuanto a la actividad comercial, las hay pequeñas y grandes urbes, el clima, la densidad de población, etc... por lo cual parece normal asignar pesos distintos. La propiedad de equivalencia distribucional implica que una mezcla de una ciudad en dos ciudades del mismo comportamiento (mismo porcentaje de las ramas comerciales en cada ciudad a cada instante) no cambia el análisis (mismas covarianzas intra, mismas inercias ...). Esta propiedad acarrea las restricciones en la elección de los pesos.

De igual modo para las ramas comerciales; ciertas ramas comerciales reagrupan más individuos que en otras (hay muchos más misceláneas y abarrotes que panaderías o vinaterías).

Esas mezclas son arbitrarias y no influyen mucho en el análisis. Para hacer este análisis sin embargo es necesario que la propiedad de equivalencia distribucional sea nuevamente satisfecha, esto es; se tiene mezclada o recortada una categoría en dos subcategorías que tienen el mismo comportamiento (mismo porcentaje en cada ciudad a cada instante) los resultados deben

ser inalterados.

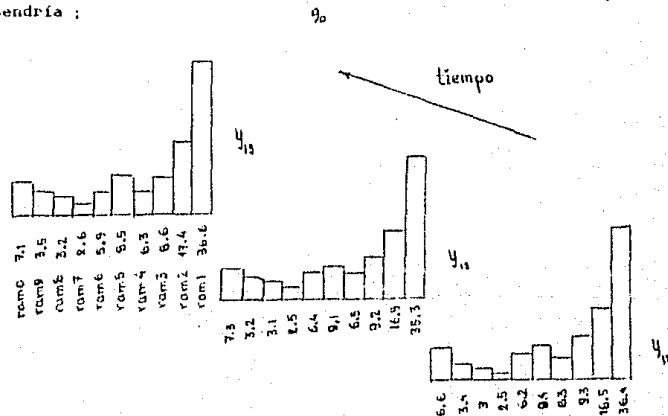
La elección de un sistema de ponderación y de la métrica en W deben ser hecha de manera que satisfaga esas propiedades. Se muestra que la métrica euclidiana es todavía una buena métrica, más hay otras, y en particular la métrica definida por la matriz:

$$M = \text{diag} (1/\sigma_{jj} | j \in J)$$

donde σ_{jj} es la desviación típica $INTRA$ de la característica J .

II.4. / La elección del registro.

Esta se hace en función de los objetivos del estudio; si se desea descubrir al estudiar la evolución de la estructura de ramas comerciales de las ciudades - estudio de la evolución de la repartición de las ramas de actividad comercial, se puede caracterizar una ciudad al tiempo t por el perfil de su repartición. En el ejemplo de la ciudad de Cuernavaca (I) se tendría;



estructura de los individuos para las ciudades con gran concentración de población o cercanas al DF.

En segundo nivel se puede considerar a :

I4 : Oaxaca

III: Culiacán

capitales de los estados de Oaxaca y Sinaloa respectivamente en donde la variación de dichas ciudades es apenas perceptible.

En tercer nivel se observa una tendencia del resto de las ciudades analizadas a estar en un pequeño globo donde las cercanas a la frontera son :

II4 : Los Mochis

II7 : Hermosillo

II8 : Queretaro

I7 : Tehuacán

ciudades con concentraciones de población en crecimiento o desarrollo industrial en desarrollo.

Sobre el plano principal (Figura 1 de TRAZA , entre inicio y pág. EIII-1), se presentan las trayectorias con efecto "spaghetti". Las ciudades a grandes rasgos siguen una evolución paralela exceptuando a Monterrey (C), Puebla (F) y Cuernavaca (A), en las cuales el número de establecimientos comerciales (o número de efectivos) sugiere una tendencia decreciente, mientras que para el resto de las ciudades se verifica una tendencia al crecimiento en número de establecimientos comerciales (las causas posibles se discuten ampliamente en el análisis conjunto de Tablas de datos , ejemplo 3).

La evolución de estas últimas fue paralela pero con menor intensidad para Queretaro, Mazatlán, Oaxaca, Los Mochis y Hermosillo, en las ciudades mencionadas aquí continúan sus actividades tradicionales , se observa poco desarrollo industrial y mayor atención a actividades turísticas (aunque tal vez exista un mercado fantasma o subempleo).

Las diferencias entre ciudades aquí no constituyen el primer objetivo del estudio pero pueden ser estudiadas con la ayuda de las correlaciones totales, entonces el primer eje se interpreta como un eje de oposición entre crecimiento de establecimientos comerciales (en número) en las diferentes ramas comerciales y el segundo como un eje de oposición respecto al volumen de población y las actividades predominantes (desarrollo industrial).

II.5, / Registro del corte de perfil ; los resultados.

El estudio de las correlaciones muestra que sigue vigente el análisis precedente más un factor económico puede estar ligado a las ciudades del último grupo mencionado en el inciso anterior (las ciudades contenidas en el globo más pequeño de la gráfica 1 en INTRA) caracterizadas por una tradición en las actividades comerciales fundamentalmente rurales y son relativamente recientes los establecimientos comerciales debido a esta tradición, o bien están localizadas en zonas del país importantes como zonas turísticas y poco industrializadas. En esas ciudades el crecimiento en el número de establecimientos es notorio aunque con distinta intensidad tal vez debido al tamaño del mercado al cual satisfacen su demanda. Contrariamente las grandes ciudades como Puebla y Monterrey observan un decremento brusco en su número de efectivos.

La visualización de 17 trayectorias proporciona una idea clara para un tomador de decisiones en el caso en que su objetivo sea montar un negocio en alguna de estas ciudades (y dentro de las ramas comerciales consideradas en el estudio).

II.6, / Clasificación de las trayectorias.

Se introduce un método de clasificación de las trayectorias, que va a permitir resolver el problema de la representación de las trayectorias , y facilita su interpretación.

II.6.1, / Definición de distancias entre trayectorias.

Lavit en () propone tres distancias a elegir :

$$d_1(i, i') = \sum (\mu_j (Y_{jt} - Y'_{jt})^2 \mid (j, t) \in J \times T)$$

conocida como la distancia instantánea promedio.

$$d_1(t, t') = \sum (\mu_1 (Y_t^j - Y_{t'}^j) - (Y_t^j - Y_{t'}^j)^2)$$

conocida como la distancia de las diferencias.

$$d_3(t, t') = \beta_1 d_1(t, t')^2 + \beta_2 d_2(t, t')^2$$

conocida como la distancia mixta.

En las expresiones de esas distancias, los coeficientes μ_1 son positivos y elegidos para satisfacer la propiedad de equivalencia distribucional sobre J. De igual manera, en la definición de la distancia d_3 , los coeficientes β_k ($k = 1, 2$) son elegidos para homogeneizar las contribuciones de las distancias d_1 y d_2 .

Note que d_2 no depende de la pendiente de las curvas :

$$t \longrightarrow Y_t^i \quad \text{y} \quad t \longrightarrow Y_t^j \quad \text{y que ,}$$

$$d_1(t, t') = 0 \text{ si } Y_t^i = Y_{t'}^i \quad \forall t;$$

$$d_2(t, t') = 0 \text{ si } (Y_t^i, t \in T) \text{ y } (Y_{t'}^i, t \in T)$$

son correspondientes salvo una traslación .

II.7. Los resultados obtenidos.

Se presenta aquí los resultados obtenidos por la elección de la tercer métrica. En efecto la primer elección tiende a clasificar el conjunto de las trayectorias correspondientes a los perfiles vecinos, más pueden evolucionar en el sentido contrario y en este sentido la segunda clasifica al contrario las ciudades que tienen estructuras diferentes más pueden evolucionar paralelamente.

Hay que hacer una clasificación para constituir las proyecciones sobre el plano principal a trayectorias seleccionadas.

Se cuenta con 7 clasificaciones básicas ;

- C1 ; Trayectoria de C.
- C2 ; Trayectoria de F.
- C3 ; Trayectoria de A.
- C4 ; Trayectoria de H y B.
- C5 ; Trayectoria de M, I, J y O.
- C6 ; Todas las demás.

Cuando se toman dos clases C5 y C6 se tiene un crecimiento en el número de efectivos (número de establecimientos comerciales y una relación de bajo desarrollo industrial, predominancia en labores agrícolas en la región o actividad turística (M, I, J y O) (Mazatlán, Chetumal, Cd. Valles y Tehuacán). Y crecimiento lento en el número de efectivos (Oaxaca, Guasave, Atlixco y Culiacán).

Cuando se consideran 3 clases se agregan ciudades con tendencias al crecimiento industrial y con aumento en su número de establecimientos comerciales (las ciudades que se agregan son Queretaro y Tepic).

Conclusión :

El método considerado aquí es una opción descriptiva a los métodos de series de tiempo. Su contribución principal consiste en la aproximación global de todos los individuos estando presentes las características (aquí la evolución de las ciudades desde el punto de vista del conjunto de las ramas comerciales RAM1, ..., RAMO). La interpretación es similar a la de un Análisis Factorial o de un Análisis Cluster. Y pone en evidencia las clases homogéneas en evolución, las etapas ulteriores del modelo pueden ser exploradas.

El de producción estadística entre las M_k . Se puede resumir el acuerdo como el resumen global de los K estudios.

Introducción :

Los datos se presentan en forma de un conjunto de tablas ; individuos x características cuantitativas, el método STATIS permite extraer información en forma gráfica ; un resumen global de las posiciones acordadas de los individuos en un sistema de ejes interpretables construidos con la ayuda de las características, evolución de cada individuo a través de sus acuerdos en el mismo sistema de ejes. Si se prefiere sobre la evolución de las características , se usa el método dual.

I. / Breve presentación del método.

Se parte de un conjunto de K tablas de individuos x características x^1, \dots, x^k concernientes a los mismos individuos ponderados por los elementos diagonales de una matriz de pesos D , sobre los cuales son medidas o ponderadas en importancia las características cuantitativas. Esas características pueden ser diferentes de una tabla a otra.

Para seguir la evolución de los individuos se trabaja sobre las tablas $M_k D = X^k M_k X_k D$. Se supondrá como acuerdos la tabla

$$\sum_{l=1}^k \alpha_l M_l D$$

que maximice :

$$\sum_{k=1}^K \langle M_k D \cdot \sum_{l=1}^K \alpha_l M_l D \rangle^2$$

donde el producto escalar entre matrices es el operador traza. Si las características son diferentes entre las tablas entonces hay que trabajar en las tablas normalizadas :

$$\frac{M_k D}{\sqrt{M_k D}}$$

Los coeficientes de la combinación lineal que dan el acuerdo son precisamente los elementos del primer eigenvalor de la matriz

Análisis global : Al efectuar la diagonalización de la matriz Y se proporciona una representación de las tablas de igual manera que se hace para las características en el ACP. Si las características están centradas (o centradas y reducidas) para cada una de las tablas, la suma de las covarianzas (o las correlaciones) esta dada por el producto escalar entre $M_k D$ y $M_l D$. De todas las características de la tabla X_k con todas las características de la tabla X_l . La norma y las proximidades de los puntos sobre la gráfica se interpretan en términos de sus covarianzas (o correlaciones) entre características de la matrices de datos.

Análisis individual : El ACP de la tabla

$$X = \left[\alpha_1^{1/2} X_1 \mid \dots \mid \alpha_k^{1/2} X_k \right]$$

proporciona la representación de los individuos y una representación de las características de las diferentes tablas.

El resto es, proporcionar una representación gráfica que permita comparar las posiciones respectivas de los individuos tales como son descritos por las diferentes tablas o de comparar sus evoluciones. Para ello se representa a los individuos en un sistema de ejes comunes (en los ejes del acuerdo $W D = X X' D$ que son los ejes principales del ACP de la matriz W) de la siguiente manera ; considere la matriz Y cuyas columnas son los eigenvalores del acuerdo, normalizados con el eigenvalor, y A la matriz diagonal con los eigenvalores ordenados decrecientemente, las coordenadas de los individuos descritos por la tabla k son precisamente los renglones de la matriz

$$M_k D Y^{-1} A.$$

Para un individuo fijo, se obtiene una posición acordada en el baricentro (sujeta a los pesos α_k) de los puntos de su trayectoria ya que $W D Y A^{-1} = Y$, el ACP de W que permite explicar la posición de los puntos bajo acuerdo con ayuda de sus características. Esta sitúa al individuo.

La trayectoria que se obtiene puede ser comparada con sus desviaciones del promedio; esto es, para un individuo fijo, cada punto k de la trayectoria sitúa este individuo en relación al "individuo" promedio de la tabla X_k , con las mismas reglas que para los individuos acordados, y el segmento que une entre dos puntos sucesivos de la trayectoria refleja en dirección y magnitud, la variación del individuo entre el estado k y el estado $k+1$ con respecto a la tendencia general, i.e. en relación a la desviación del individuo promedio de la tabla X_k y el individuo promedio de la tabla X_{k+1} . Como en todo plano principal, las variaciones se traducen mejor sobre las trayectorias de la gráfica que con respecto a las trayectorias "promedio" del centro.

Consideraremos las ciudades como los individuos, las ramas de actividad comercial como las características y los registros comerciales como las tablas.

El método *STATIS* debe ser aplicado sobre los perfiles de las ciudades, i.e. las fracciones en porcentaje de las diversas ramas comerciales por ciudad.

Se trabaja con características reducidas por tabla, (en *STATIS* es posible incluso normalizar las tablas $M_k D$).

II. Análisis global.

Como se puede observar en *INTER* en la página 2

$$M D = .3342 M_1 D + .3318 M_2 D + .3346 M_3 D$$

Los puntos se suceden en orden cronológico para los 2 primeros hacia arriba del gráfico y después para 84 y 85 ocurre un retraso notable. Considerando la distancia del conjunto con respecto al origen los tres puntos son cercanos. Aunque 85 está relativamente separado de 83 y 84, ello significa que los perfiles de las ciudades cambian notablemente entre 1984 y 1985 esto posiblemente se debe al recrudecimiento de la crisis económica predominante en los últimos años e incluso a la

aplicación de medidas por parte de la administración gubernamental entrante (después de un año se resiente la aplicación de sus medidas económicas) posiblemente se cerraron algunos establecimientos de las ramas de comercio o hubo cambio de giro en algunos establecimientos, como lo muestra la tabla de inauguración-quebra, en general de 84 a 85 se registraron 601 quiebras y 335 inauguraciones donde las ramas comerciales con mayor número de quiebra son:

$ram1, ram2, ram4, ram5, ram8, ram9$

y sólo se registró cierta tendencia a la apertura de establecimientos de las ramas $ram5$ y $ram0$ correspondientes a materiales para construcción y zapaterías.

Para las ramas $ram7$ y $ram3$ se puede considerar casi en igual proporción de quiebras y aperturas; estas corresponden a panaderías y farmacias respectivamente (ésto solo vale para las ciudades I_1, \dots, I_{17} no es un análisis de las ciudades de México).

Matriz de inauguración-quebra

individuo	ramas										Q	I
	1	2	3	4	5	6	7	8	9	0		
I1	-2	-2	-9	-4	-9	-7	0	0	1	-5	38	1
I2	3	2	3	3	14	-5	0	3	0	1	5	29
I3	65	-106	-10	-47	9	-71	-7	-20	-15	14	276	88
I4	-1	10	6	10	3	-2	-2	-3	-7	6	15	35
I5	0	0	0	0	1	0	0	1	0	0	0	1
I6	-58	-18	6	-16	-4	13	-3	-4	-2	9	102	31
I7	-6	-1	4	0	-2	1	1	-7	-1	1	17	7
I8	-12	2	-4	9	-5	-6	-4	-2	-5	7	38	18
I9	-1	1	0	0	0	0	0	1	0	-1	2	2
I10	1	3	3	4	2	-2	0	2	-1	1	3	16
I11	-2	4	5	2	6	-11	0	2	1	7	13	27
I12	2	0	0	-1	2	-1	0	-1	0	4	3	8
I13	-4	2	-5	-8	2	1	2	2	-5	-4	26	9
I14	1	12	0	3	0	-2	0	2	2	5	2	25
I15	-6	2	-11	-10	-2	-8	1	1	-1	-1	39	4
I16	4	2	1	-1	0	-6	0	-2	0	-7	16	7
I17	-6	3	1	8	0	3	3	1	0	7	6	26
Quiebras	98	127	30	87	22	121	13	39	37	18		
Inauguraciones	76	43	29	39	39	18	10	15	4	62		

Quiebras = 601

Inauguraciones = 335.

El año comercial que más asemeja al promedio corresponde al año 83.

Después de esta visión general de los registros comerciales es conveniente entrar en detalles.

Las ciudades en donde se registran mayor número de quiebras en relación a inauguraciones son :

- I3 : Monterrey
- I6 : Puebla
- I19: Cuernavaca
- I7 : Mazatlán
- I16: Guaymas

Las ciudades con mayor número de inauguraciones en relación a quiebras son :

- I2 : Tepic
- I14: Los Mochis
- I17: Hermosillo
- I11: Culiacán
- I10: Cd. Valles

Las causas posibles :

-Crisis económica aguda conforme transcurre el tiempo, posiblemente por alguna política económica implantada por la administración entrante.

-Disminución en la demanda de los artículos que se expenden, debido a una pérdida del poder adquisitivo de la población de las ciudades analizadas y en general de la población mexicana.

-Se registra menor cantidad de quiebras en ciudades de la región noroeste del país.

-Se registra mayor cantidad de quiebras en las ramas comerciales :

- ram5 : Muebles en general
- ram2 : Carnicerías y salchichonerías
- ram4 : Ferreterías y tlapalerías.

Si que la población de esas ciudades pueden dejar de construir y reparar o amueblar su vivienda y evitar la carne en la dieta alimenticia.

III / Análisis de fineza.

Sobre las gráficas proporcionadas por el ACP de

$$X = [\alpha_1^{1/2} X_1 | \dots | \alpha_n^{1/2} X_n]$$

se explica la repartición de las ciudades acordadas por sus categorías en ramas comerciales dominantes.

a) Repartición de las ciudades acordadas.

En INTRA (página 1) se observa la gráfica de ciudades (individuos) bajo acuerdos. En esta representación, los dos primeros ejes acumulan :

1er. eje	92.44 % inercia
2do. eje	4.30 % inercia

esto es, 96.74 % de la inercia de la nube X de las ciudades, por lo cual la gráfica es una excelente representación de dicha nube (la "mejor" fotografía de dicha nube fué bien tomada).

El primer eje opone las ciudades : Monterrey, Puebla y Cuernavaca principalmente, además de Tehuacán y Oaxaca con el resto de las ciudades analizadas, esto se debe quizás al gran número de establecimientos, puesto que son grandes núcleos de población comparados con el resto de las ciudades que son ciudades pequeñas, se nota que Queretaro está un poco abajo del eje horizontal.

Así las ciudades, Monterrey, Puebla y Cuernavaca se localizan en el lado positivo del primer eje, cabe hacer notar que por la proximidad de Cuernavaca a la ciudad capital (la ciudad más poblada de México) es posible suponer una influencia de la población del Distrito Federal en la existencia de la cantidad de establecimientos comerciales para las diversas ramas de actividad, esta ciudad (Cuernavaca) aparte de ser capital del Estado de Morelos es un centro recreativo y vacacional de ciertos sectores de la población capitalina. Puede pensarse que las ciudades separadas por el eje vertical son separadas según el desarrollo

industrial que sustenta una oferta de trabajo y consecuentemente mayor concentración de establecimientos comerciales. Cuernavaca por sí sola no se caracteriza por un alto grado de industrialización sin embargo por la cercanía a la zona metropolitana cuenta con un mercado vigoroso.

La tendencia de estas ciudades es claramente hacia el II cuadrante y la de las ciudades restantes hacia el IV cuadrante;

Las ciudades : Monterrey, Puebla y Cuernavaca llevan una trayectoria de descenso en el número de establecimientos en las diversas ramas comerciales. Mientras que el resto de las ciudades registran pequeños incrementos en su número de establecimientos.

IV / Trayectorias de las ciudades.

Un segundo nivel de análisis en detalle es atender a la representación de cada ciudad de 1983 a 1985 en el sistema de ejes que permita descubrir las ciudades bajo acuerdo. Una vez que se ha dado un nombre a cada uno de los ejes (en términos de categorías o ramas comerciales) la lectura de una trayectoria es fácil. Las características están reducidas para cada tabla, esto es, las desviaciones son usadas para proporcionar una "ciudad" promedio que se traza sobre una trayectoria.

La tabla siguiente proporciona la evolución de perfil de la "ciudad" promedio :

ramas	1983	1984	1985
ram1	186.1176	182.2353	180.7059
ram2	70.5294	74.7647	69.8235
ram3	73.8235	74.3529	73.7647
ram4	66.3529	64.2353	61.4118
ram5	61.2941	59.0000	60.0000
ram6	60.8235	54.0588	48.0000
ram7	18.1176	19.0000	18.8235
ram8	30.7647	29.8235	28.4118
ram9	21.2941	21.6471	19.7059
ram0	42.9412	46.0588	48.0588

Se registra sensible baja en las ramas :

- ram6 : Muebles en general
- ram1 : Miscelaneas y abarrotes
- ram4 : Ferreterías y tlapalerías

y en menor escala :

- ram2 : Carnicerías y salchichonerías
- ram3 : Farmacias
- ram5 : Materiales para construcción
- ram8 : Tiendas de ropa
- ram9 : Vinos y licores
- ram0 : Zapaterías

con pequeño aumento solo en ram7 ; Panaderías.

Nuevamente se puede pensar como causa al recrudescimiento de la crisis económica con consecuente baja en el poder adquisitivo y disminución en la demanda ocasionando el cierre de establecimientos (posiblemente de los establecimientos pequeños, incapaces de sostener una ley de oferta-demanda contra grandes almacenes dentro de una misma rama comercial). En panaderías ocurre que este es un producto alimenticio de primera necesidad.

Las trayectorias de las ciudades (gráficas 2 y 3 de la TRAZA) (gráfica 2 en la página 1 de TRAZA y gráfica 3 empieza en la página 8) son de tres tipos.

Las que se alejan del eje horizontal y corresponden por este criterio a :

- Monterrey (C)
- Puebla (F)

y en menor grado

- Cuernavaca (A).

Las que se aglutinan cerca de dicho eje ; pueden ser considerarse por subgrupos :

G1 : Oaxaca (D) y Cuicacán (K).

G2 : Queretaro (H), Hermosillo (Q) y Los Mochis (N).

G3 : Tepic (B) y Cd. Obregón (O).

G4 : Tehuacán (G) y Mazatlán (M).

Las que están bajo el eje horizontal y están relativamente separadas de la clasificación anterior :

GS1 : Guaymas (P) y Cd. Valles (J).

GS2 : Atlixco (E), Guasave (L) y Chetumal (I).

Estas relaciones describen un comportamiento distinto entre Monterrey y Puebla separados del resto posiblemente por el volumen de establecimientos activos en dichas ciudades, el caso similar es Cuernavaca aunque con menos importancia en el número de establecimientos comerciales.

En la segunda clasificación aparecen ciudades con núcleos de población importantes aunque sin llegar a ser comparables con los de la primera clasificación, dichas ciudades se caracterizan por ser centros turísticos o ciudades con un desarrollo industrial medio.

En la tercera clasificación predominan ciudades con poblaciones pequeñas por lo cual su evolución en cuanto a ramas comerciales es similar.

En la gráfica 3 se pueden apreciar las evoluciones pronunciadas que sufren las grandes ciudades como Monterrey y Puebla. Ciertas ciudades cambian completamente de perfil y sus trayectorias son bastante largas este es el caso de las grandes ciudades con concentraciones elevadas de población. Las ciudades menos industrializadas poseen trayectorias más semejantes y menos oscilantes (o más estables) cuyos cambios son más suaves). Esto se debe a que la cantidad de establecimientos en las ramas comerciales varía según el poder adquisitivo de la población, que depende de otras leyes económicas.

APENDICE O.

En este apéndice se discute el concepto de funcional lineal, espacio dual y base dual.

El concepto de funcional lineal es importante para el estudio de los espacios vectoriales (de dimensión finita en esta tesis), pues ayuda a presentar un estudio profundo de subespacios, ecuaciones lineales y coordenadas.

Definición : Sea V un espacio vectorial sobre un campo escalar K , una transformación lineal con dominio V y codominio K se denomina un funcional lineal sobre V .

Ejemplo 1 : Sobre el espacio \mathbb{R}^n un funcional lineal puede ser expresado en la forma

$$f(x) = \sum_{k=1}^n \eta_k \xi_k, \text{ para } x = (\xi_1, \xi_2, \dots, \xi_n)$$

donde $\eta_j, j=1, \dots, n$ son escalares fijos. Se puede ver que tales funcionales son lineales

$$\begin{aligned} f(x) &= f \left[\sum_{j=1}^n \xi_j e_j \right] \\ &= \sum_{j=1}^n \xi_j f(e_j) \\ &= \sum_{j=1}^n \xi_j \eta_j \quad \text{donde } \eta_j = f(e_j) \quad j=1, \dots, n \end{aligned}$$

f es el funcional representado por la matriz $[\eta_1, \eta_2, \dots, \eta_n]$ respecto a la base canónica de \mathbb{R}^n y a la base (1) de K . Además se puede verificar que todos los funcionales lineales son de esta forma.

Ejemplo 2 : Sea $M_{n \times n}(\mathbb{R})$ el conjunto de matrices de $n \times n$ con elementos en \mathbb{R} , si se define para $A \in M_{n \times n}(\mathbb{R})$:

$$\text{tr}(A) = \sum_{i=1}^n a_{ii}$$

la traza de A entonces $\text{tr} : \mathbb{R}^{n \times n} \longrightarrow \mathbb{R}$ y es lineal pues si $c \in \mathbb{R}$

$$\text{tr}(cA + B) = \sum_{i=1}^n (ca_{ii} + b_{ii}) = c \sum_{i=1}^n a_{ii} + \sum_{i=1}^n b_{ii} = c \text{tr}(A) + \text{tr}(B).$$

Ejemplo: Sea $C(0,1) = \{f \mid f \text{ es continua en } (0,1)\}$ y se define el funcional $f(x) = x(1/2)$ entonces f así definido es funcional lineal.

$$f : (0,1) \longrightarrow \mathbb{R}$$

y es lineal pues si $c \in \mathbb{R}$

$$f(cx+cy) = (cx+cy)(1/2) = cx(1/2) + cy(1/2) = cf(x) + f(y).$$

Definición: Sea V un espacio vectorial, el conjunto de todos los funcionales sobre V forman un nuevo espacio vectorial, es el espacio $L(V, \mathbb{R})$ y se le designa también por V^* , se le llama espacio dual de V .

Ejemplo 1: En el espacio \mathbb{R}^n cada vector es una n -ada de escalares reales $x = (\xi_1, \dots, \xi_n)$ con norma $\|x\| = \left(\sum_{i=1}^n \xi_i^2\right)^{1/2}$.

Cada funcional f de la forma $f(x) = \sum_{i=1}^n \eta_i \xi_i$ con cada $\eta_i \in \mathbb{R}$ es claramente lineal. También de la desigualdad de Cauchy-Schwarz para sumas finitas:

$$|f(x)| = \left| \sum_{i=1}^n \eta_i \xi_i \right| \leq \left[\sum_{i=1}^n \eta_i^2 \right]^{1/2} \left[\sum_{i=1}^n \xi_i^2 \right]^{1/2} = \left[\sum_{i=1}^n \eta_i^2 \right]^{1/2} \|x\| \quad (1)$$

$$\therefore \text{ si define } \|f\| = \sup_{\|x\| \leq 1} |f(x)|$$

se ve que f es acotada con $\|f\| \leq \left[\sum_{i=1}^n \eta_i^2 \right]^{1/2}$. Además para $x =$

(η_1, \dots, η_n) la igualdad se asume en la expresión (1) y de hecho se

$$\text{tiene } \|f\| = \left[\sum_{i=1}^n \eta_i^2 \right]^{1/2}.$$

Ahora sea f cualquier funcional lineal acotado sobre \mathbb{R}^n . Define los vectores base $e_i = (0, \dots, 0, 1, 0, \dots, 0)$ con la i -ésima componente 1 y el resto 0. Sea $\eta_i = x^*(e_i)$. Un vector $x = (\xi_1, \dots, \xi_n)$ puede ser expresado en estos términos en la base por

$$x = \sum_{i=1}^n \xi_i e_i$$

por lo tanto f es lineal y se tiene:

$$f(x) = \sum_{i=1}^n \xi_i f(e_i) = \sum_{i=1}^n \xi_i \eta_i.$$

Entonces el dual de \mathbb{R}^n es \mathbb{R}^n mismo en el sentido de que el espacio $(\mathbb{R}^n)^*$ consiste de todos los funcionales de la forma

$$f(x) = \sum_{i=1}^n \eta_i \xi_i \text{ y la norma sobre } (\mathbb{R}^n)^* \text{ es } \|f\| = \left[\sum_{i=1}^n \eta_i^2 \right]^{1/2}.$$

Ejemplo 2: El dual de \mathcal{C}_0 .

El espacio \mathcal{C}_0 es definido como el espacio de todas las sucesiones $x = (\xi_i)$ de números reales convergentes a cero. La norma sobre \mathcal{C}_0 es $\|x\| = \max |\xi_i|$ (\mathcal{C}_0 es un subespacio de l_∞), el dual de \mathcal{C}_0 es precisamente l_1 en el sentido usual de los funcionales lineales representados como:

$$f(x) = \sum_{i=1}^{\infty} \xi_i \eta_i \quad y = (\eta_1, \eta_2, \dots) \in l_1$$

y $\|f\| = \|y\|$.

Ejemplo 3: Sobre un espacio con un producto interior el funcional $f(x) = B(x, y)$ (donde B es una función bilineal simétrica definida positiva) para un y fijo es un funcional lineal en la variable x .

La desigualdad de Cauchy-Schwarz:

$$|B(x, y)| \leq \|x\| \|y\|$$

es posible mostrar que el funcional f es acotado con

$$\|f\| \leq \|y\|$$

(solo hay que tomar $x/|x|$ en lugar de x). La relación $f(y)=B(y,y)$ muestra que en efecto $|f| = |y|$. Obviamente vectores distintos producen funcionales distintos, esto es, en un espacio con producto interior los funcionales lineales son generados por elementos del espacio mismo. Para lograr cierto grado de formalidad hay que probar que todos los funcionales lineales son de esta forma.

En particular para E^n y una función bilineal con matriz asociada

$$B = \begin{pmatrix} 1 & 1/2 & 1/2 \\ 1/2 & 1 & 1/2 \\ 1/2 & 1/2 & 1 \end{pmatrix}$$

defina el producto interior de x e y como :

$$B(x,y) = xBy' = \langle x,y \rangle$$

así que para un vector $y=(y_1, y_2, y_3) \in \mathbb{R}^3$ fijo se tiene :

$$f(x) = B(x,y)$$

$$= (x_1, x_2, x_3) \begin{pmatrix} 1 & 1/2 & 1/2 \\ 1/2 & 1 & 1/2 \\ 1/2 & 1/2 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

$$= x \begin{pmatrix} y_1 + 1/2 y_2 + 1/2 y_3 \\ 1/2 y_1 + y_2 + 1/2 y_3 \\ 1/2 y_1 + 1/2 y_2 + y_3 \end{pmatrix}$$

$$= x_1(y_1 + 1/2 y_2 + 1/2 y_3) + x_2(1/2 y_1 + y_2 + 1/2 y_3) + x_3(1/2 y_1 + 1/2 y_2 + y_3)$$

con x como variable f es un funcional lineal sobre \mathbb{R}^3 . Por ejemplo si $x = (1,1,1)$ se tiene que $f(x)=2(x_1+x_2+x_3)$ es el funcional asociado a $y=(1,1,1)$, si ahora se tiene que $y=(1,0,0)$ se obtiene el funcional lineal $f(x)=x_1 + 1/2 x_2 + 1/2 x_3$.

Los elementos de la base canónica con sus respectivos funcionales asociados :

Vector	Funcional lineal
(1,0,0)	$f_1(x) = x_1 + 1/2 x_2 + 1/2 x_3$
(0,1,0)	$f_2(x) = 1/2 x_1 + x_2 + 1/2 x_3$
(0,0,1)	$f_3(x) = 1/2 x_1 + 1/2 x_2 + x_3$

Puesto que cualquier vector $y \in \mathbb{R}^3$ es combinación lineal de (e_1, e_2, e_3) entonces $y = a_1 e_1 + a_2 e_2 + a_3 e_3$ con $a_i \in \mathbb{R}$ $(i=1,2,3)$, el funcional lineal asociado a y es :

$$\begin{aligned} f(x) &= \\ &= x_1(a_1 + 1/2 a_2 + 1/2 a_3) + x_2(1/2 a_1 + a_2 + 1/2 a_3) + x_3(1/2 a_1 + 1/2 a_2 + a_3) \\ &= a_1(x_1 + 1/2 x_2 + 1/2 x_3) + a_2(1/2 x_1 + x_2 + 1/2 x_3) + a_3(1/2 x_1 + 1/2 x_2 + x_3) \\ &= a_1 f_1(x) + a_2 f_2(x) + a_3 f_3(x) \end{aligned}$$

Por lo tanto es suficiente conocer a los funcionales asociados a los elementos de la base para encontrar los funcionales asociados a cada elemento del espacio.

Si se requiere del procedimiento inverso, es decir, si se tiene un funcional lineal sobre \mathbb{R}^3 es posible hallar un vector asociado a este funcional :

Por ejemplo :

$$f(x) = 7x_1 + 2x_2 + \pi x_3$$

$$y \quad f(x) = a_1 f_1(x) + a_2 f_2(x) + a_3 f_3(x)$$

entonces basta igualar los coeficientes de ambas expresiones :

$$7 = a_1 + 1/2 a_2 + 1/2 a_3$$

$$+ \quad 2 = 1/2 a_1 + a_2 + 1/2 a_3$$

$$\pi = 1/2 a_1 + 1/2 a_2 + a_3$$

resolviendo el sistema de ecuaciones se obtiene el vector

$$y = (a_1, a_2, a_3)$$

del cual proviene bajo el isomorfismo entre \mathbb{R}^3 y $(\mathbb{R}^3)'$ definido por la matriz B .

La conclusión en el caso en que V es de dimensión finita es que se puede obtener una descripción explícita del espacio dual V' .

Pues el siguiente resultado, enunciado sin prueba garantiza que $\dim V' = \dim V$.

Teorema : Sea V espacio vectorial de dimensión finita n sobre el campo escalar K y W un espacio vectorial de dimensión finita m sobre K . Entonces el espacio $L(V,W)$ es de dimensión finita y

tiene dimensión nm. (Para la prueba consulte (3)).

Si $\beta = (\alpha_1, \dots, \alpha_n)$ es base de V entonces existe para cada α_i un funcional lineal único f_i en V^* tal que :

$$f_i(\alpha_j) = \delta_{ij}$$

De este modo obtiene n funcionales distintos (f_1, \dots, f_n) sobre V que son también linealmente independientes, ya que si

$$f = \sum_{i=1}^n c_i f_i \quad \text{entonces :}$$

$$f(\alpha_j) = \sum_{i=1}^n c_i f_i(\alpha_j) = \sum_{i=1}^n c_i \delta_{ij} = c_j$$

En particular si f es el funcional lineal cero, $f(\alpha_j) = 0$ para cada j y por tanto los escalares c_j son todos cero.

Entonces los f_1, \dots, f_n son n funcionales linealmente independientes, y como se sabe V^* tiene dimensión n , deben ser tales que $\beta^* = (f_1, \dots, f_n)$ es una base de V^* .

Definición : Si V es un espacio vectorial de dimensión finita con base $\beta = (\alpha_1, \dots, \alpha_n)$ y f_i es el funcional lineal asociado, entonces $\beta^* = (f_1, \dots, f_n)$ es una base de V^* y es llamada la base dual de β .

Ejemplo :

Sea $V = \mathbb{R}^2$ y $B = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. Para $\alpha \in \mathbb{R}^2$ fijo se tiene :

$$f_\alpha(\alpha) = (x_1, x_2) \begin{bmatrix} 2y_1 + y_2 \\ y_1 + 2y_2 \end{bmatrix} = x_1(2y_1 + y_2) + x_2(y_1 + 2y_2)$$

Vector	Funcional lineal
$u = (1, 1)$	$f_1(\alpha) = 3x_1 + 3x_2$
$w = (1, -1)$	$f_2(\alpha) = x_1 - x_2$

Los vectores u y w son linealmente independientes, por lo tanto son base de \mathbb{R}^2 , i.e. $\beta = (u, w)$ es base con base dual $\beta^* = (f_1, f_2)$.

APENDICE 1

En este apéndice se tratan los conceptos de : Métrica, producto escalar, forma bilineal, forma cuadrática, isomorfismo y norma.

Definición : Una métrica o función de distancia en un conjunto dado E es una función real d sobre $E \times E$ tal que para cualesquiera x, y, z pertenecientes a E se verifica

- (i) $d(x, x) = 0$
- (ii) $x \neq y \Rightarrow d(x, y) = d(y, x) > 0$.
- (iii) $d(x, y) + d(y, z) \geq d(x, z)$.

La condición (iii) se conoce como la desigualdad del triángulo (al espacio E con una función de distancia se le llama espacio métrico).

Ejemplos :

1.- E de dimensión n , $x = (x_1, \dots, x_n)$ $y = (y_1, \dots, y_n)$ elementos de E , se define la distancia

$$d_p = \left\{ \sum_{i=1}^n (x_i - y_i)^p \right\}^{1/p} \quad p \geq 1$$

es conocida como la p -distancia de E .

2.- E de dimensión n , x e y como en ejemplo anterior y B la matriz asociada a la forma bilineal B definida positiva, entonces

$$d_B(x - y) = \sqrt{(x - y)B(x - y)}$$

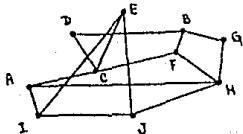
es conocida como la distancia elíptica de E .

3.- Se prueba que si en 1. considera

$$\lim_{p \rightarrow \infty} d_p(x, y) = d(x, y) = \max \{ |x_i - y_i| : i=1, \dots, n \}$$

resulta ser una nueva distancia. (Para detalles consulte (7)).

4.- Se define la distancia entre dos nodos p_i y p_j de una gráfica G conexa, como la longitud de la cadena más corta entre p_i y p_j . Esta distancia se designa por $\delta(p_i, p_j)$. Por ejemplo en la figura siguiente: $\delta(D, G) = 2$; $\delta(D, J) = 3$



Se prueba que en efecto es distancia:

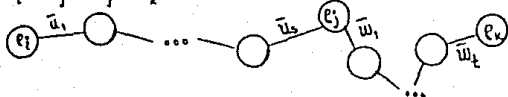
Demostración:

Sea γ una gráfica conexa y p_i, p_j y p_k nodos distintos.

Las condiciones (i) y (ii) son inmediatas. Para probar (iii):

$$\delta(p_i, p_j) + \delta(p_j, p_k) \geq \delta(p_i, p_k)$$

se considera (u_1, \dots, u_n) y (w_1, \dots, w_t) las cadenas más cortas entre los nodos p_i, p_j y p_j, p_k respectivamente.



De no ocurrir la desigualdad a probar se verifica que $s+t < \delta(p_i, p_k)$ y por tanto la cadena de mínima longitud entre p_i y p_k es mayor a $s+t$; sin embargo al considerar la cadena:

$$(u_1, \dots, u_s, w_1, \dots, w_t)$$

de p_i a p_k de longitud $s+t$ se verifica una contradicción.

□

Definición: Un producto escalar o función producto interno en un conjunto E es una función real $\langle \cdot, \cdot \rangle$ sobre $E \times E$ tal que para cualesquiera $x, y, z \in E$ y $\alpha \in \mathbb{R}$, se verifica:

$$(i) \quad \langle x, y \rangle = \langle y, x \rangle.$$

$$(ii) \quad \langle \alpha x, y \rangle = \alpha \langle x, y \rangle = \langle x, \alpha y \rangle.$$

$$(iii) \quad \langle x+y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$$

$$\langle x, y+z \rangle = \langle x, y \rangle + \langle x, z \rangle.$$

$$(iv) \quad \langle x, x \rangle \geq 0 \text{ y } \langle x, x \rangle = 0 \Leftrightarrow x = 0.$$

Ejemplos:

1.- E de dimensión n y B definida positiva y simétrica

$$\langle x, y \rangle_B = x^T B y.$$

-En particular si $B = I_n$ se tiene:

$$\langle x, y \rangle_1 = x^T I y = xy = \sum_{i=1}^n x_i y_i.$$

el producto interno usual.

2.- $P_1 = \{p \mid p \text{ es polinomio en } t\}$

$$\langle f, g \rangle = \int_a^b f(t)g(t)dt.$$

Se define el producto:

$$\langle x, y \rangle = \int_a^b x(t)y(t)dt$$

||

aquí es evidente que se cumplen (i), (ii) y (iii) y para (iv) suponga que

$\langle x, x \rangle = 0$ y suponga también que $x(t) \neq 0$ i.e. $\exists t_0 \in (a, b)$ tal que $x^2(t_0) > 0$ por la continuidad de $x \in V_c(t_0)$ abierto contenido en (a, b) tal que $x^2(t) > 0 \forall t \in V_c(t_0)$ por ello:

$$\int_a^b x^2(t)dt = \int_{(a, b) \setminus V_c(t_0)} x^2(t)dt + \int_{V_c(t_0)} x^2(t)dt \geq \int_{V_c(t_0)} x^2(t)dt > 0$$

$\therefore \langle x, x \rangle > 0$, lo cual evidentemente es una contradicción.

□

Definición: Una función bilineal en un conjunto dado E es una función real B sobre $E \times E$ tal que para cualesquiera $x, y, z \in E$ y $\alpha \in \mathbb{R}$ se tiene:

- (i) $B(x+y, z) = B(x, z) + B(y, z)$
 $B(x, y+z) = B(x, y) + B(x, z)$
(ii) $B(\alpha x, y) = \alpha B(x, y)$
 $B(x, \alpha y) = \alpha B(x, y)$

Definición: A la expresión de una función bilineal en coordenadas. En una base $\beta = (e_1, \dots, e_n)$ de E

$$B(x, y) = \sum_{i,j=1}^n b_{ij} x_i y_j$$

donde los coeficientes $b_{ik} = B(e_i, e_k)$ que dependen solo de la base y no de x e y , se llama forma bilineal.

(Observe que es una función bilineal con una base fija).

Ejemplos:

1.- Si $\langle \cdot, \cdot \rangle$ es un producto interior entonces es una función bilineal.

2.- Si $M_{2 \times 2}(\mathbb{R})$ entonces el determinante es una forma bilineal

$$B(x, y) = |x \ y|.$$

Definición: Una función cuadrática en un conjunto dado E es una función real no negativa sobre $E \times E$ tal que para $x \in E$,

$$C_B(x) = B(x, x).$$

donde B es una función bilineal simétrica.

Ejemplo:

$$1.- C_B(x) = \sum_{i=1}^n x_i^2$$

Definición: Un isomorfismo entre E y F es una correspondencia biyectiva entre los elementos de E y F que conserve las operaciones definidas en E, i.e. $x_1, y_1 \in E$ y $x_2, y_2 \in F$ y $x_1 \longleftrightarrow y_1$, $x_2 \longleftrightarrow y_2$ se tenga:

$$x_1 + y_1 \longleftrightarrow x_2 + y_2$$

Ejemplos:

1.- El conjunto de fuerzas coplanares y el de n-adas ordenadas de números reales para $n=2$ y que las operaciones de adición y multiplicación por un escalar se preservan; por lo

tanto ambos conjuntos son isomorfos.

2.- Otro ejemplo conocido es el isomorfismo entre $P^n(\mathbb{R})$ los polinomios de grado $\leq n$ de coeficientes reales y variable real t , y el conjunto \mathbb{R}^{n+1} dada por la correspondencia:

$$x = (x_1, \dots, x_{n+1}) \longleftrightarrow x_1 + x_2 t + \dots + x_{n+1} t^n.$$

3.- Si B es una forma bilineal simétrica definida positiva entonces la siguiente relación es isomorfismo

$$\begin{array}{ccc} E & \xrightarrow{\psi} & E^n \\ x & \longmapsto & \psi(x) = B_n^{-1} x = B(x, \cdot) = x B_n^{-1} \end{array}$$

donde la matriz asociada es la matriz simétrica B asociada a la forma bilineal B.

Definición: Una norma definida en un conjunto E es una función no negativa real $\| \cdot \|$ tal que para cualesquiera $x, y \in E$ y $\alpha \in \mathbb{R}$ se verifica:

- (i) $\|x\| \geq 0$.
- (ii) $\|x\| = 0$ si y sólo si $x = 0$.
- (iii) $\|\alpha x\| = |\alpha| \|x\|$.
- (iv) $\|x+y\| \leq \|x\| + \|y\|$.

(Un espacio vectorial E con una tal norma recibe el nombre de espacio normado).

Ejemplos:

1.- Normas elípticas. Si B es una función bilineal simétrica definida positiva se define la norma

$$\|x\|_B = \left\{ B(x, x) \right\}^{1/2}$$

para $x \in E$, y B definida sobre E.

2.- Normas p

Se define en E de dimensión n la norma:

$$\|x\|_p = \left\{ \sum_{i=1}^n |x_i|^p \right\}^{1/p}$$

3.- Sea $C[a,b]$ el espacio de las funciones continuas sobre $[a,b]$ se define una norma como :

$$|f| = \max_{t \in [a,b]} |f(t)|.$$

4.- Sea $L^p(0,1) = \left\{ f \mid \left\{ \int_0^1 |f(t)|^p dt \right\}^{1/p} < \infty \right\}$. Se define :

$$|x| = \left\{ \int_0^1 |f(t)|^p dt \right\}^{1/p}$$

como la norma en $L^p(0,1)$.

APENDICE 2.

En este apéndice se analiza el concepto de transformación ortogonal.

Definición : Una aplicación lineal L de un espacio euclideo E se llama ortogonal si

$$B(Lx, Ly) = B(x, y) \quad \forall x, y \in E.$$

En otras palabras toda aplicación ortogonal conserva el producto escalar y, por consiguiente, conserva las longitudes de los vectores y los ángulos entre estos.

Basta exigir sin embargo, que la transformación lineal conserve las longitudes de los vectores para que resulte ortogonal.

En efecto, si se supone que $|Lx|_V = |x|_V \quad \forall x \in E$.

En ese caso $|L(x+y)|_V = |x+y|_V$, ya que :

$$\begin{aligned} |L(x+y)|_V^2 &= B(L(x+y), L(x+y)) \\ &= B(Lx+Ly, Lx+Ly) \\ &= B(Lx, Lx) + 2 B(Lx, Ly) + B(Ly, Ly) \\ &= |Lx|_V^2 + |Ly|_V^2 + 2 B(Lx, Ly) \end{aligned}$$

donde también :

$$\begin{aligned} |x+y|_V^2 &= B(x+y, x+y) = B(x, x) + 2B(x, y) + B(y, y) \\ &= |x|_V^2 + |y|_V^2 + 2B(x, y) \end{aligned}$$

pero tomando en consideración las igualdades

$$|Lx|_V^2 = |x|_V^2 ; |Ly|_V^2 = |y|_V^2 ; |L(x+y)|_V^2 = |x+y|_V^2$$

se obtiene $B(Lx+Ly, Lx+Ly) = B(L(x+y), L(x+y))$.

La otra implicación es clara. □

Es también claro que toda aplicación ortogonal transforma bases ortonormales en bases ortonormales.

Aún más si se tiene una aplicación lineal L que transforma al menos una base ortonormal en una base ortonormal entonces L es ortogonal.

En efecto, sea $\{e_1, \dots, e_n\}$ una base ortonormal y $\{e'_1, \dots, e'_n\}$ base ortonormal tal que :

$$Le_i = e'_i \quad \forall i=1, \dots, n.$$

En cuyo caso $x, y \in E$ se escriben como :

$$x = \sum_{i=1}^n x_i e_i \quad y = \sum_{i=1}^n y_i e_i$$

y también

$$Lx = \sum_{i=1}^n Lx_i e_i = \sum_{i=1}^n x_i Le_i = \sum_{i=1}^n x_i e'_i$$

analogamente

$$Ly = \sum_{i=1}^n y_i e'_i \quad \text{que cumplen :}$$

$$B(Lx, Ly) = \sum_{i=1}^n x_i y_i b_{ii} = W(x, y).$$

Si además L^* es su aplicación conjugada se tiene :

$$W(x, y) = B(Lx, Ly) = W(x, L^*(Ly)) = W(x, (L^*L)y) \quad \forall x, y \in E$$

$\therefore L^*L = I$ y se sigue que $L^* = L^{-1}$.

Y se obtiene una condición necesaria y suficiente para que la aplicación L sea ortogonal.

Consecuencias :

- L es no degenerada.

- L^{-1} es también ortogonal $(L^{-1})^* = (L^*)^{-1} = (L^{-1})^{-1} = L$.

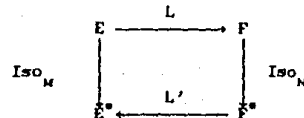
- Si L y T son aplicaciones lineales ortogonales entonces $L \cdot T$ es ortogonal, sean U y V las matrices asociadas, entonces $(LU)^* = U^*L^* = U^{-1}L^{-1} = (LU)^{-1}$.

Para los detalles consultar (3).

APENDICE 3.

En el presente apéndice se analiza el concepto de isometría.

Sean E y F dos espacios euclidianos de dimensión p y n respectivamente y M y N formas bilineales definidas positivas. Denote por $L(E, F)$ el conjunto de las funciones lineales de E a F . Si $L \in L(E, F)$ entonces la condición :



$$Iso_M(x)(y) = \{ (Iso_M \cdot L)(x)(L(y)) \} \quad ; (x, y) \in E \times E,$$

explica que se conservan los ángulos al pasar de E a F por la aplicación lineal L . Pero esta condición es equivalente a :

$$(Iso_M(x))(y) = \{ L' \cdot Iso_N \cdot L(x)(y) \}$$

se escribe entonces $Iso_M = L' \cdot Iso_N \cdot L$, puesto que Iso_M es un isomorfismo, la aplicación L es inyectiva; es aquí necesario exigir $n > p$.

Observación : -En el caso en que E y F sean espacios euclidianos de la misma dimensión finita, se tiene entonces automáticamente que L es un isomorfismo y se verifica la igualdad :

$$Iso_M \cdot L^{-1} = L' \cdot Iso_N = (Iso_N \cdot L)'$$

-Si además E y F así como Iso_M e Iso_N coinciden, esa igualdad se reescribe como :

$$Iso_M \cdot L^{-1} = L' \cdot Iso_M = (Iso_M \cdot L)'$$

el isomorfismo L es además M -ortogonal; esto es, conserva los M -ángulos.

-Pero si $E^* = F$ y $L = \text{Iso}_M$ se tiene el diagrama

$$E \begin{array}{c} \xrightarrow{\text{Iso}_M} \\ \xleftarrow{\text{Iso}_M^{-1}} \end{array} E^*$$

de donde se obtienen las igualdades :

$$\text{Iso}_M \cdot L^{-1} = L' \cdot \text{Iso}_N$$

se escribe : $I_P = \text{Iso}_M \cdot \text{Iso}_N \therefore \text{Iso}_M = \text{Iso}_N^{-1} \cdot I_P$

Si E es euclideo con métrica inducida por la bilineal M induce un isomorfismo Iso_M el cual a su vez produce en E^* la métrica inducida por Iso_M^{-1} o por M^{-1} .

-Se observa que si $\dim E = \dim F$ entonces la matriz asociada a la forma bilineal N es precisamente la matriz identidad I_P .

$$\begin{array}{ccc} E & \xrightarrow{L} & E \\ \text{Iso}_M \downarrow & & \downarrow \text{Iso}_{I_P} \\ E^* & \xrightarrow{L'} & E^* \end{array}$$

de aquí se obtiene que : $\text{Iso}_M = L' \cdot \text{Iso}_N \cdot L$, i.e. existe una composición de isomorfismos que permite pasar de M a I_P .

Descomposición propia :

Si M es la matriz asociada a Iso_M entonces M se puede escribir como :

$$M = U \cdot D_\lambda \cdot U^{-1} = (D_\lambda^{-1} U^{-1})' (D_\lambda U)$$

Donde las columnas de la matriz U son precisamente los eigenvectores de M y D_λ se define como la matriz diagonal con $\text{diag} D_\lambda = (\lambda_1, \lambda_2, \dots, \lambda_p)$ y λ_i son los eigenvalores en este caso

números positivos.

$$D_\lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \dots & \\ & & & \lambda_p \end{bmatrix}$$

(Ver (3) para detalles).

Descomposición de Doolittle-Choleski.

(Doolittle) .- Toda matriz simétrica puede ser escrita como un producto de dos matrices triangulares, pero si además M es positiva entonces la escritura de M en ese producto es única.

$$M = U' U \quad (\text{Choleski}).$$

y U es una matriz triangular.

En nivel matricial hay asociadas dos descomposiciones del isomorfismo Iso_M , la de Choleski es eficaz en la práctica y es la más usada.

apenduo?cht:

APENDICE 4.

En este apéndice se expone el teorema de descomposición, resultados preliminares en el análisis de matrices tridimensionales expuesto en el capítulo III de esta tesis.

Teorema de descomposición :

Sea E_1 un espacio vectorial euclideo real, de dimensión p_1 , de métrica M_1 , y sea E_2 un espacio vectorial real, de dimensión p_2 , de métrica M_2 . Sea X' la matriz de la aplicación lineal de E_2 en E_1 , de rango s .

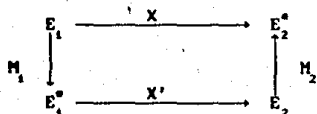
$X'M_1X'$ es M_2 -simétrica, positiva y a sus eigenvalores positivos se les denota por $\lambda_1 \geq \dots \geq \lambda_s > \lambda_{s+1} = \dots = \lambda_{p_2} = 0$. Sea (v_1, \dots, v_{p_2}) un sistema M_2 -ortonormal de eigenvectores asociados, se tiene :

$$V = (v_1, \dots, v_{p_2})$$

$$D = \begin{bmatrix} \lambda_1^{1/2} & & & 0 \\ & \lambda_2^{1/2} & & \\ & & \ddots & \\ 0 & & & \lambda_s^{1/2} \\ & & & & & & & 0 \end{bmatrix}$$

$$U = X'M_2VD^{-1}$$

se define :
dentro del esquema



Teorema : $X' = UDV'$, tal que $U'M_1U = I_{p_1}$, $V'M_2V = I_{p_2}$. D es una matriz diagonal positiva.

Demostración :

Defina $W = (V_1, \dots, V_{p_2})$, W' es entonces una matriz cuadrada tal que $W'M_2W = I_{p_2}$. M_2W es precisamente la inversa de W' donde $M_2W' = I_{p_2}$.

Por otra parte, para toda i , $X'M_1X'V_i = 0$, con

$$iX'M_2V_iM_2 = 0, \quad Y X'M_2V_i = 0.$$

$$\text{Con } X' = X'I_{p_2} = X'M_2WW' = X'M_2 \sum_{i=1}^{p_2} V_iV_i' = X'M_2 \sum_{i=1}^s V_iV_i'$$

$$= X'M_2VV' = UDV'$$

$$U'M_1U = D^{-1}V'M_2X'M_1X'M_2VD^{-1} = D^{-1}V'M_2VD^{-1} = I_{p_1}$$

Corolario : Las columnas de $U : U_1, \dots, U_{p_1}$, forman un sistema M_1 -ortonormal de eigenvectores de $X'M_2X'$, con eigenvalores asociados $\lambda_1, \dots, \lambda_s$ y $V = X'M_1UD^{-1}$.

Demostración :

$$\begin{aligned} \text{En efecto : } X'M_2X'V_i &= X'M_2X'M_1 \left(\frac{1}{\sqrt{\lambda_i}} X'M_2V_i \right) \\ &= \frac{1}{\sqrt{\lambda_i}} X'M_2(X\lambda_i V_i) \\ &= \lambda_i \left(\frac{1}{\sqrt{\lambda_i}} X'M_1UD^{-1} \right) \\ &= VDU'M_1U D^{-1} = V \end{aligned}$$

APENDICE 5.

Producto escalar de matrices.

En este apéndice se discute la construcción de un producto escalar entre matrices.

Sea (E_1, M_1) y (E_2, M_2) dos espacios euclidianos definidos como en el apéndice 4. Sea $\mathbb{R}(p_1, p_2)$ el espacio vectorial de las matrices $p_1 \times p_2$ isomorfo al espacio vectorial de las aplicaciones lineales de \mathbb{R}^{p_2} en \mathbb{R}^{p_1} .

Proposición : La aplicación

$$\begin{aligned} \gamma : \mathbb{R}(p_1, p_2) \times \mathbb{R}(p_1, p_2) &\longrightarrow \mathbb{R} \\ (A, B) &\longrightarrow \text{tr}(A' M_1 B M_2) \end{aligned}$$

es bilineal, simétrica, definida positiva.

Demostración :

La bilinealidad y la simetría resultan de las propiedades de la traza.

$$\text{Sea } A \in \mathbb{R}(p_1, p_2) : \gamma_{M_1 M_2}(A, A) = \text{tr}(A' M_1 A M_2).$$

Si $M_1 = L_1 L_1'$, donde L_1 es una matriz $p_1 \times p_1$ de rango p_1 , y $M_2 = L_2 L_2'$, donde L_2 es una matriz $p_2 \times p_2$ de rango p_2 , se tiene :

$$\text{tr}(A' M_1 A M_2) = \text{tr}(A' L_1 L_1' A L_2 L_2')$$

$$= \text{tr}(L_1' A L_2 L_2' A' L_1) = \sum_{i=1}^{p_1} \sum_{j=1}^{p_2} (L_1' A L_2)_{ij}^2 \geq 0$$

$$\gamma_{M_1 M_2}(A, A) = 0 \Leftrightarrow L_1' A L_2 = 0 \Leftrightarrow A = 0.$$

$\gamma_{M_1 M_2}$ es definida positiva.

Por consiguiente para toda elección de métricas M_1 sobre E_1 , y M_2 sobre E_2 , se puede definir un producto escalar $\gamma_{M_1 M_2}$, que permitirá definir la norma de una aplicación lineal de una matriz.

así como una distancia euclídeana. En el capítulo III como se utiliza ese producto escalar, según los espacios y las métricas consideradas.

Observación :

$\mathbb{R}(p_1, p_2)$ es isomorfo a $\mathbb{R}^{p_1 p_2}$, por el isomorfismo que a

toda matriz A , $\bar{A} = \begin{pmatrix} A \\ \vdots \\ A \\ P_2 \end{pmatrix}$ obtenida por la yuxtaposición de las columnas de la matriz A .

$\gamma_{M_1 M_2}$ puede generar una métrica sobre $\mathbb{R}^{p_1 p_2}$, con matriz :

$$\begin{pmatrix} \langle M_2 \rangle_{11} M_1 & \langle M_2 \rangle_{12} M_1 & \dots & \langle M_2 \rangle_{1 p_2} M_1 \\ \langle M_2 \rangle_{21} M_1 & \langle M_2 \rangle_{22} M_1 & \dots & \langle M_2 \rangle_{2 p_2} M_1 \\ \vdots & \vdots & \ddots & \vdots \\ \langle M_2 \rangle_{p_2 1} M_1 & \langle M_2 \rangle_{p_2 2} M_1 & \dots & \langle M_2 \rangle_{p_2 p_2} M_1 \end{pmatrix}$$

(mejor conocido como el producto de Kronecker de las matrices M_1 y M_2 , (para detalles consulte (2))).

APENDICE 6.

Se exponen los teoremas necesarios en la discusión de la construcción del método desarrollado en el capítulo III.

Teorema : Sea (E, M) un espacio vectorial euclideo real, de dimensión d y sea L una aplicación lineal de E^d en E , simétrica, semidefinida positiva. Sea $\{e_1, \dots, e_d\}$ una base ortonormal de eigenvectores de VM , con eigenvalores asociados $\lambda_1 \geq \dots \geq \lambda_d > 0$.

Se tiene $L = \sum_{i=1}^d \lambda_i e_i e_i'$. Entonces, para todo $k \leq d$, entre las aplicaciones lineales de E^d en E , simétricas, semidefinidas positivas, de rango a lo más k , el mínimo de $\|A - B\|_{VM}^2$ se asume

en $B = \sum_{i=1}^k \lambda_i e_i e_i'$, con valor mínimo $\sum_{i=1}^d \lambda_i^2$.

Teorema 2 : B definido como anteriormente, realiza el máximo de $\frac{\langle A, B \rangle_{VM}}{\|B\|_{VM}}$; con valor mínimo $(\lambda_1^2 + \dots + \lambda_k^2)^{1/2}$.

Teorema 3 : Sea A una aplicación lineal de E_2^d en E_1 , donde (E_1, M_1) y (E_2, M_2) son dos espacios vectoriales euclideos. Sea k el rango de A . Se tiene según el teorema del apéndice 5 que :

$$A = UDV' = \sum_{i=1}^k \sqrt{\lambda_i} V_i V_i'$$

Entonces la aplicación lineal B de E_2^d en E_1 , definida por

$T = \sum_{i=1}^k \sqrt{\lambda_i} V_i V_i'$ realiza el mínimo de $\|A - B\|_{M_1 M_2}^2$, por consiguiente las aplicaciones lineales de rango $\leq k$. Ese

mínimo vale $\sum_{i=1}^k \lambda_i$.

Teorema 4 : Sea $X' = UDV'$, entonces U_i es el vector normal de E_i realiza el máximo de $\|X M_i U_i\|_{M_2}^2$, y ese máximo vale λ_i .

Para $j \leq k$, U_j es el vector normal de E_j , M_1 -ortogonal a U_1, U_2, \dots y U_{j-1} , realiza el máximo de $\|X M_j U_j\|_{M_2}^2$ y ese máximo vale λ_j .

Demostración :

En efecto :

$\|X M_i U_i\|_{M_2}^2 = u'(M_i X' M_2 X M_i)u$ es máximo, con la restricción $u' M_i u = 1$, para $M_i X' M_2 X M_i u = \lambda M_i u$, donde $X' M_2 X M_i u = \lambda u$, con λ máximo; con $u = U_i$ eigenvector normalizado de $X' M_2 X M_i$ asociado al más grande eigenvalor. Por esto mismo $\|X M_i U_i\|_{M_2}^2$ es máximo con las restricciones :

$$u' M_i u = 1, u' M_i U_1 = u' M_i U_2 = \dots = u' M_i U_{j-1} = 0, \text{ para } u = U_i$$

$$\|X M_i U_i\|_{M_2}^2 = 1/\sqrt{\lambda_i} V_i \|M_2\|^2 = \lambda_i.$$

Observación :

-Cuando M_2 es diagonal, con elementos diagonales q_1, \dots, q_{p_2} ,

$$\|X M_i U_i\|_{M_2}^2 = \sum_{l=1}^{p_2} q_l \langle x_l, u \rangle_{M_1}^2$$

donde x_l es el renglón l de X .

Las U_i realizan en cada caso el máximo de

$$\sum_{l=1}^{p_2} q_l \langle x_l, u \rangle_{M_1}^2$$

-Análogamente se verifica que las V_i realizan el máximo de

$$\|X' M_2 V_i\|_{M_1}^2$$

bajo las restricciones de ortonormalidad de las soluciones V .

De igual manera si M_1 es diagonal, de elementos diagonales q'_1, \dots, q'_{p_1} , las V_i realizan el máximo de :

$$\sum_{i=1}^p q_i^2 (c_{ij}^1 v_j)^2$$

donde v^j es la j -ésima columna de X .

BIBLIOGRAFIA

- [1] TANKARD, E. - "Introduction to the theory of determinants and matrices". 1988, THE UNIVERSITY OF NORTH CAROLINA PRESS, U.S.A.
- [2] MARDIA, K.V., KENT, J.T. y BIBBY, J.M. "Multivariate Analysis", 1979, ACADEMIC PRESS, U.S.A.
- [3] HOFFMAN, K. - "Algebra lineal", 1977, Prentice/Hall, Madrid, España.
- [4] RAO, C.R. - "Linear Statistical Inference and Its Applications", 1973, WILEY, U.S.A.
- [5] FOUCART, T. - "Structure des tableaux de probabilités - Description et prevision", 1979, Université des Sciences et Techniques du Languedoc, Montpellier II
- [6] JAFFRENOU, P.A. - "Sur l'analyse des familles finies de variables vectorielles", 1978, Université Claude Bernard, Lyon I.
- [7] KOSAKU, Y. - "Functional Analysis" Springer-Verlag.
- [8] GLACON, F. - "Analyse conjointe de plusieurs matrices de données", Thèse de 3ème cycle. Unité de Biométrie, 9, Place P. Viala 34060 MONTPELLIER CEDEX.
- [9] LAVIT, C. - "METHODE STATIS", Unité de Biométrie, ENSA-INRA-USTL, Place Viala, 34060 MONTPELLIER CEDEX.
- [10] LAVIT, C. - "Analyse conjointe de tableaux quantitatifs", Masson, col. Méthode+Programme, 1988.
- [11] HIZERE, D. - "L'ASPECT CHOIX DE METRIQUES DANS L'ANALYSE D'UN CUBE DE CONTINGENCE". Faculté des Sciences, Département de Mathématiques. B.P., AFRIKA MATEMATIKA, 1984. BRAZZAVILLE (Congo).
- [12] LEBART, MORINEAU Y VARVICK. - "Multivariate Descriptive Statistical Analysis". Correspondence Analysis and Related Techniques for Large Matrices. WILEY, U.S.A.