UNIVERSIDAD DE PANAMA
VICERRECTORIA DE INVESTIGACION Y POST-GRADO
PROGRAMA DE MAESTRIA EN MATEMATICA

COMPONENTES PRINCIPALES Y CORRELACIONES CANONICAS

POR:

GLADYS ESTHER SEGURA

Tesis presentada como uno de los requisitos para optar por el Grado de Maestro en Ciencias con Especialización en Matemática.

Panamá, República de Panamá

-1986-

UNIVERSIDAD DE PANAMA

VICERRECTORIA DE INVESTIGACION Y POSTGRADO

TM:

12 ENE. 1987

Obsequio del autor

Aprobado por:

Director de Tesis ______________________
DIMAS QUIEL, M.Sc.

Miembro del Jurado ______________________
JOSE ~~DEL R.~~ GARRIDO, Ph.D.

Miembro del Jurado ______________________
RICARDO PARKER, M.Sc.

Fecha 22/10/86

DEDICATORIA

A mis hijas MONICA Y VERONICA quienes con su bella existencia me ayudaron a no flaquear en momentos difíciles. A mi mamá y hermana quienes me brindaron su apoyo en todo momento. A mis alumnos y en particular aquellos que hoy son colegas míos pues siento haberles cumplido.

GLADYS

AGRADECIMIENTO

A mis profesores, J. POLTRONIERI, AUGUSTO ARRIAGADA, ADELA ABAD, RICARDO PARKER, VIRGINIA B. DE ALDANA, EGBERTO AGARD, TERESITA J. DE RUIZ y en particular al profesor y director de tesis, DIMAS QUIEL; gracias por sus enseñanzas.

A INESITA quién con mucha paciencia logró mecanografiar nitidamente la tesis, al Profesor M.Sc. JOSE FERNANDEZ, por la elaboración de los símbolos y a la Profesora EYDA JIMENEZ DE SANTIMATEO amiga y colega de quién siempre recibí ánimo para continuar.

CONTENIDO

PAGINA

INTRODUCCION

El tema de nuestra tesis está enmarcado dentro de Los Métodos del Análisis de Datos, que son procedimientos que se proponen resumir y sintetizar grandes conjuntos de datos, de los cuales tratamos, el Análisis en Componentes Principales (A.C.P.) y El Análisis de las Correlaciones Canónicas y Variables Canónicas (A.C.C.).

Estas ideas fueron desarrolladas desde hace mucho tiempo, por Karl-Pearson en 1900, luego por Harold Hottelling quién en 1933 estudió el A.C.P. y posteriormente en 1936 fundamenta el A.C.C.

El surgimiento de las computadoras, herramienta indispensable en los laboriosos cálculos del A.C.P. y el A.C.C. han influido en gran medida al estudio más profundo y en la actualización de gran parte de la teoría tratada por Pearson y Hottelling. Así el A.C.P. y el A.C.C. constituyen las herramientas fundamentales para la determinación del orden de importancia que tiene un conjunto de variables en la explicación de la variabilidad de un fenómeno en estudio.

La Tesis comprende tres capítulos, y se inicia con la demostración del teorema clásico de la diagonalización de una matriz simétrica, que desde un punto de vista puramente estadístico; consiste en la búsqueda de un conjunto de vectores α_i, de norma uno; tales que las combinaciones lineales $\alpha'_i X$ sean de varianza máxima y no-correlacionadas.

Uno de los resultados interesantes consiste en que los valores propios y los vectores propios de la matriz de covarianza poblacional Σ, son respectivamente, la varianza máxima y los vectores de los coeficientes de las combinaciones lineales $\alpha' X$. A continuación se aplica la misma metodología, considerando la matriz de las correlaciones, suponiendo en ambos casos conocida la matriz de covarianza.

A través de los estimadores de máxima verosimilitud, vemos que los estimadores de máxima verosimilitud de los valores y vectores propios de Σ , son respectivamente los valores y vectores propios de su estimador de máxima verosimilitud.

$$\hat{\Sigma} = \frac{1}{N} \quad (x_\alpha - \bar{x})(x_\alpha - \bar{x})' = A$$

En el capítulo 2, investigamos la existencia de dos vectores α y γ tales que $\alpha' X^{(1)}$ y $\gamma' X^{(2)}$ tengan correlación máxima; donde $X^{(1)}$ está integrado por las primeras s-componentes de X y $X^{(2)}$ las t-restantes componentes del vector aleatorio X, cuya matriz de covarianza se considera definida positiva, donde además $\alpha' X^{(1)}$ y $\gamma' X^{(2)}$ deben tener varianza igual a uno. Las soluciones son las correlaciones máximas y las variables canónicas $\alpha' X^{(1)}$ y $\gamma' X^{(2)}$ que son las raíces de:

$$\begin{bmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda\Sigma_{22} \end{bmatrix} = 0$$

y los vectores que satisfacen el siguiente sistema

$$\begin{bmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda\Sigma_{22} \end{bmatrix} \begin{bmatrix} \alpha \\ \gamma \end{bmatrix} = 0$$

respectivamente. Inmediatamente se demuestra que el número de correlaciones canónicas es a lo más s, que hay además s parejas de variables canónicas y (s-t) variables de correlación cero; donde s+t es la dimensión de X. De manera similar al de las componentes principales se obtienen los estimadores de las variables y correlaciones canónicas. Se termina este capítulo con la obtención de las variables y correlaciones canónicas a partir de un estimador de la matriz de correlación y en el modelo de regresión.

En el último capítulo tratamos los conceptos básicos del Análisis Lineal Multidimensional, para aplicarlo a las componentes principales. Este método general, se utiliza para otras técnicas de reducción de datos, como Análisis Factorial y de Correspondencia.

A través del isomorfismo entre un espacio vectorial E, p-dimensional sobre $\mathbb{R}$ y $\mathbb{R}^p$; se le hace corresponder a las columnas (filas) de una matriz de información de dimensión pxn, vectores de un espacio E(F), llamado Espacio de los individuos (Espacio de los caracteres). Se definen además los duales E^* y F^* como espacios de las representaciones de los caracteres e individuos respectivamente.

Utilizaremos el mismo símbolo M, (a veces la letra D) para representar:

a) Una forma bilineal simétrica definida positiva o la matriz M.

b) El isomorfismo inducido entre E y E^* por M.

c) La norma

$$M(x) = ||x||_M^2$$

d) La métrica

$$M(x, y) = ||x-y||_M$$

e) Una ortogonalidad

$$M(x,y) = 0 \Longleftrightarrow x \,{}_M\!\perp y$$

f) Un producto escalar:

$$M(x,y) = \langle Mx\ y \rangle = x'\ My$$

A través del siguiente diagrama se visualiza la relación entre estos espacios, la matriz X (o aplicación lineal X), y la métrica M.

$$W = X \circ M \circ X;\ V = X \circ D_p \circ X'$$

$$\begin{tikzcd}
E \arrow[d, shift right, "M"'] & F^* \arrow[l, "X"'] \arrow[d, shift right, "W"'] \\
E^* \arrow[u, shift right, "V"'] \arrow[r, "X"] & F \arrow[u, shift right, "D_p"']
\end{tikzcd}$$

donde W y V son formas bilineales simétricas semi-definidas positivas si y solo si (E,M) es una representación o imagen Euclideana de (I,d), es decir existe una aplicación f entre I y E tal que f(I) = $\boxed{M}$, siendo $\boxed{M}$ la imagen en E de los vectores columnas (elementos de $\mathbb{R}^p$) bajo el isomorfismo entre E_p y $\mathbb{R}^p$) y $d_I(i,i') = d_E(h(i), h(i'))$. Todo esto se realiza para poder cambiar a una métrica Euclideana; de manera que las distancias entre los individuos sea medida por la longitud de los vectores que los une.

Se utiliza el concepto de momento de Inercia con respecto a un punto para medir la dispersión. Encontrándose luego que el punto más próximo a la nube es su centro de gravedad (la media). Los ejes principales son los generados por los vectores propios normados asociados, a los valores propios de V o D; forma bilineal simétrica semi-definida positiva. Sucede que el primer eje principal, el plano principal y el subespacio principal son la recta, plano y subespacio más próximo a la nube de puntos, i.e. el momento de Inercia con respecto al eje principal es mínimo y máximo con respecto al espacio suplementario ortogonal.

La traza de Σ , denominada en el Capítulo I como la varianza total del sistema es en el Capítulo III; el momento de inercia con respecto al origen del Espacio, siendo este el centro de la nube.

En la última parte del capítulo definimos dos índice para medir la calidad de la representación de un índividuo x_i; el primero, como la razón de la norma de la proyección del vector x_i sobre el plano principal a la norma del vector x_i; El segundo como la razón entre la distancia del vector al plano principal y $\sqrt{\text{traza (VM)}-(\lambda_1 + \lambda_2)}$, "varianza promedio fuera del plano principal P".

En el desarrollo de este trabajo, una obra fundamental fue "An Introuction to Multivariate" Statistical Analysis del eminente científico T.W. Anderson. Citado en todo artículo o libro que trate algún tema de Análisis Multivariado.

En orden de importancia consultamos, otra obra inperecedera, la de S.S. Wilks, "Mathematical Statistical" de 1963; que apareció por primera vez en una impresión de la Princeton University Press en 1943 y a la cual Anderson hace referencia. Estos dos libros junto con el de D.F. Morrison; Multivariate Statistical Methods, me sirvieron de apoyo para el desarrollo de los dos primeros capítulos.

El último capítulo, en el que se desarrolla el Análisis Lineal Multidimensional (fundamentos), y el A.C.P. fueron de gran importancia las obras de Caillez - Pages y la de Lebart-Fenelon.

NOTACIONES Y ABREVIATURAS

c.l.	Combinaciones lineales.
A.C.P.	Análisis en Componentes Principales.
A.C.C.	Análisis de las Correlaciones Canónicas.
s.e.v.	Subespacio Vectorial.
v.c.	Variables Canónicas.
def. +	Definida positiva
$\cong$	Equivalente
l.i.	Linealmente independiente
$\Sigma \geqslant 0$	Matriz Simétrica semi-definida positiva.
$\Sigma > 0$	Matriz Simétrica definida positiva.

CAPITULO I

COMPONENTES PRINCIPALES

1.1 DETERMINACION DE LAS COMPONENTES PRINCIPALES A NIVEL POBLACIONAL.

1.1.1 Objetivos del Análisis en Componentes Principales y Definiciones generales.

El Análisis en Componentes Principales (A.C.P.) es uno de los métodos del Análisis de Datos. El Análisis de Datos, pretende suministrar representaciones sintetizadas de grandes conjuntos de datos.

Se utiliza cuando; nos interesa investigar las causas de variabilidad o cambio de algún fenómeno o proceso. Este método además de explicitar las causas de los cambios (C.P.) con su respectivos cargos o responsabilidad (valor de la varianza), finalmente reduce la dimensión al eliminarse las componentes de menor varianza; es decir aquellas que son menos importante en la explicación de las causas de cambio en el fenómeno de estudio. Además estas componentes resultan no correlacionadas; que en el caso de variables aleatorias normalmente distribuidas implica independencia y por lo tanto simplicidad en el análisis y estudio.

Tratar de lograr por un lado simplicidad para entender los datos, visualizarlos, interpretarlos y por otro lado retener suficiente información en detalle para una representación adecuada; es en su forma más simple el objetivo del Análisis en Componentes Principales.

Las ideas principales sobre el Análisis en Componentes Principales fueron originalmente desarrolladas por Pearson en 1900 y posteriormente por Harold Hotelling en 1933.

1.1.2 Teorema de Existencia.

La caracterización de las C.P. se logrará a través de la demostración de un teorema en el cual a partir de un vector aleatorio, de media cero y matriz de covarianza poblacional Σ (que se supone conocida); se obtendrá una transformación lineal de las componentes del vector X. El transformado tendrá componentes no correlacionadas; que son combinaciones lineales de las componentes de X. Los coeficientes de estas combinaciones son los vectores propios de Σ .

DEFINICIONES BASICAS.

Definición 1.1.1. Varianza Total del Sistema.

Se define como:

$$\operatorname{tr} \Sigma = s_1 + s_2 + \ldots + s_p = \operatorname{tr} \Sigma = \operatorname{tr} S$$

donde s_i es la varianza de la i-ésima componente principal y S es la matriz de covarianza del vector de componentes principales.

Definición 1.1.2. Varianza Generalizada.

Es la determinante de la matriz de covarianza.

Definición 1.1.3. Varianza Relativa de la C.P.

Es la parte de la varianza total explicada por la i-ésima C.P. y está dado por:

$$\frac{S_i}{\operatorname{tr} \Sigma} \times 100$$

Teorema de Existencia.

Dado un vector aleatorio X de esperanza cero y matriz de covarianza Σ , existe una transformación lineal $Y = P'X$ tal que $E(YY') = S$ con $S = P' \Sigma P$. Donde S y P son matrices diagonales y ortogonales respectivamente de los valores y vectores propios de Σ y además, la primera componente de Y es la componente principal Y_1; de varianza máxima s_1; Y_2 es la segunda C.P. de varianza $s_2 \leqslant s_1$ y así sucesivamente hasta Y_p que es la C.P. de menor varianza; s_p.

Demostración:

Deseamos determinar una combinación lineal normalizada con varianza máxima, i.e., un vector p tal que:

$$p'p = 1 \text{ y } E(p'x)^2 = p'\Sigma p \text{ sea máxima} \qquad 1.1.1$$

Resolviendo por Multiplicadores de Lagrange, p es una solución de:

$$(\Sigma - sI)p = 0 \qquad 1.1.2$$

donde s debe verificar:

$$|\Sigma - sI| = 0 \qquad 1.1.3$$

es decir un valor propio de Σ y $s = p'\Sigma p \geqslant 0$ 1.1.4

Sea s_1 el mayor de los valores propios y $Y_1 = p'_{(1)} X$ donde $p_{(1)}$ es un vector propio asociado a s_1.

Buscamos ahora una combinación lineal normalizada p'X de varianza máxima, dentro de los no-correlacionados con Y_1 (Las C.P. son no-correlacionadas; la matriz de covarianza de Y es diagonal); i.e., deseamos maximizar la siguiente función de Lagrange:

$$\phi = p'\Sigma p - s(p'p-1) - 2\gamma_1 p'\Sigma p \qquad 1.1.5$$

resultando p una solución normalizada de $(\Sigma - sI)p = 0$ donde s es el segundo valor propio de Σ según un orden decreciente.

Continuando este procedimiento, hasta la etapa r+1-ésima en que buscamos un vector p tal que p'X tenga varianza máxima de entre los no correlacionados con $Y_1, Y_2, \ldots, Y_r$; resolviendo por multiplicadores de Lagrange, debemos maximizar la siguiente función.

$$\phi = p'\Sigma p - s(p'p-1)-2\gamma_i \sum_{i=1}^{r} p'\Sigma p_{(i)} \qquad 1.1.6$$

Obteniéndose una solución semejante a 1.1.2 y 1.1.3. Si Σ fuese semi-definida positiva podría ocurrir que el r+1 valor propio fuera igual a cero y que coincidiese con otro valor propio (multiplicidad mayor que uno) i.e. si $s_{r+1} = 0$ y $s_j=0$ $(j \neq r+1)$, entonces $p \Sigma\ p_{(r+1)} = 0$ no implica $p_{(j')}\ p_{(r+1)} = 0$; pero P_{r+1} puede ser reemplazado por una combinación lineal de elementos de la base del espacio vectorial de los vectores propios asociados al valor propio cero, de manera que éste sea ortogonal a todos los p_j $(j = 1,2,\ldots,r)$, se continua hasta llegar a la etapa (m + 1) en que no se puede encontrar un vector tal que:

(a) $p'p = 1$, $(\Sigma - sI)p = 0$ y $E(p' X Y_j) = 0$ $j = 1,2,\ldots,m$

Sabemos que $m \leq p$. Se demuestra por el absurdo que $m = p$. Asumiendo que $m < p$, se demuestra que existe un vector que satisface 1.1.2 y 1.1.3.

Demostración:

Sea $P_* = (p_1,\ldots,p_m)$. Por las condiciones que verifican los p_i tenemos que:

$$P_*' P_* = I \qquad 1.1.7$$

P_* es p x m, de rango completo por columna ya que los p_i son linealmente independientes. Asi, por lo tanto tenemos que existe una matriz:

$$G = (g_{m+1}, \ldots, g_p) \qquad 1.1.8$$

donde G es px(p-m) tal que $(P_*\ G)$ es una matriz ortogonal; es decir

$$p_i' g_j = 0 \quad i = 1,2,\ldots,m;\ j = m+1,\ldots, p$$

$$g_k' g_l = \delta_{k,l} = m + 1,\ldots,p$$

A continuación veremos que el vector y el escalar que buscamos, es decir la nueva solución está dada por: Ge una combinación lineal de los vectores agregados y α es el valor propio de $G'\Sigma G$.

Sea α una raíz de

$$|G'\Sigma G - \alpha I| = 0$$

y e un vector propio asociado a α; i.e.

$$G'\Sigma Ge = \alpha e \qquad 1.1.9$$

tal que $e'e = 1$

Como ΣGe es ortogonal a $P_* = \{p_1, p_2, \ldots, p_m\}$ porque

$$p_i' \Sigma Ge = S_i p_i' \sum_{j=m+1}^{p} e_j g_j = S_i \sum_{j=m+1}^{p} e_j\, p_i' g_j = 0 \qquad 1.1.10$$

(por lo tanto Ge es no-correlacionado con los p_i, $i = 1,2,\ldots,m$) así ΣGe es generado por los $g_{m+1},\ldots,g_p$ y por lo tanto existe un vector h tal que:

$$\Sigma Ge = Gh \qquad 1.1.11$$

Asi tenemos que; por 1.1.9.

$$G'\Sigma Ge = G'Gh = h = \alpha e$$

Además; por 1.1.11. $\Sigma(Ge) = \alpha(Ge)$

de esta manera vemos que existe un vector Ge que satisface:

$$(\Sigma - \alpha I)Ge = 0.$$

Además Ge satisface:

$$(Ge)'(Ge) = e'G'Ge = e'e = 1.$$

y por lo tanto α verifica:

$$|\Sigma - \alpha I| = 0$$

siendo ésta una nueva solución.

Veamos que los $Y_i = p_i'X$ y $p_i'\Sigma p_i = s_i$ determinados por multiplicadores de Lagrange con las características exigidas son soluciones de 1.1.2 y 1.1.3.

Sea

$$P = (p_1,\ldots,p_p) \qquad y \qquad S = \begin{pmatrix} S_1 & & \\ & \ddots & \\ & & S_p \end{pmatrix}$$

la propiedad $\Sigma p_r = S_r p_r$ en forma matricial se expresa por:

$$\Sigma P = PS \qquad 1.1.12$$

Además $p_r'p_r = 1$ y $p_r'p_s = 0 \qquad r \neq s$.

en forma matricial es:

$$P'P = I \qquad 1.1.13$$

de 1.1.12 y 1.1.13 tenemos

$$P'\Sigma P = P'PS = S \qquad 1.1.14$$

Además los $S_i = p_i'\Sigma p_i$, verifican que

$$\begin{aligned} |\Sigma - sI| &= |P'|\ |\Sigma - sI|\ |P| \\ &= |P'\Sigma P - sI| \\ &= |S - sI| \\ &= \prod_{i=1}^{p} (s_i - s) = 0 \end{aligned}$$

Así las raices encontradas son las soluciones 1.1.2 y 1.1.3.

A continuación veremos que las C.P. son las únicas combinaciones normalizadas con la propiedad de varianza máxima.

Sean $Y_1, Y_2, \ldots, Y_p$ las C.P. de X y sea

$$Y^* = \sum_{i=1}^{p} C_i Y_I \qquad \text{donde} \sum_{i=1}^{p} C_i^2 = 1.$$

una combinación lineal normalizada de las componentes principales, que son a su vez una combinación lineal normalizada de las componentes de X,

$$C'Y = C'P'X; \text{ donde } C'P'PC = C'C = 1$$

Determinemos los valores de las c_i tales que Y^* tenga varianza máxima:

$$\text{Var}\ (Y^* = C'Y) = \text{Var}\left(\sum_{i=1}^{p} C_i Y_i\right) = \sum_{i=1}^{p} C_i^2\ \text{Var}(Y_i) = \sum_{i=1}^{p} C_i^2 S_i$$

Esta varianza puede ser expresada de la siguiente manera:

$$\begin{aligned}\text{Var}\ (Y^*) = S_1 + \sum_{i=2}^{p} C_i^2 (S_i - S_1) &= S_1 + \sum_{i=2}^{p} C_i^2\ S_i - S_1 \sum_{i=2}^{p} C_i^2 \\ &= S_1 + \sum_{i=2}^{p} C_i^2\ S_i - S_1(1 - C_1^2) \\ &= S_1 + \sum_{i=2}^{p} C_i^2\ S_i - S_1 + C_1^2\ S_1 \\ &= \sum_{i=1}^{p} C_i^2\ S_i\end{aligned}$$

Derivando parcialmente con respecto a c_j tenemos:

$$\frac{\partial \text{Var}\ (Y^*)}{\partial C_j} = 2C_j\ (S_j - S_1) = 0$$

$$C_j = 0; \text{ ya que } S_1 \neq S_j \quad j = 2, \ldots, p.$$

Obteniendo la segunda derivada parcial con respecto a c_k, resulta:

$$\begin{aligned}\frac{\partial^2\ \text{Var}\ (Y^*)}{\partial C_j\ \partial C_k} &= 0 \quad \text{si } j \neq k. \\ &= 2(S_k - S_1) < 0 \text{ si } S_1 \text{ es el máximo.}\end{aligned}$$

Así $c_j = 0$ para $j = 2, 3, \ldots, p$

y como $$\sum_{i=1}^{p} C_i = 1$$

entonces

$$C_1 = 1$$

por lo tanto

$$\mathrm{Var}\ (Y^{*} = Y_1) = S_1$$

Sea Y_1 no correlacionado con Y^{*}; i.e. $E(Y_1 Y^{*}) = 0$

$E(Y_1 Y^{*}) = E(Y_1(\sum C_i Y_i)) = \sum E(C_i Y_i Y_1') = \sum C_i E(Y_i Y_1) = 0$

$E(Y_i Y_1) = 0$ si $i \neq 1$, así $C_1 S_1 = 0 \Longrightarrow C_1 = 0.$

de donde $Y^{*} = \sum_{i=2}^{p} C_i Y_i$, repitiendo el proceso para las $Y_2, \ldots, Y_p$ vemos que $\mathrm{Var}\ (Y^{*})$ es máxima si $Y^{*} = Y_2$.

Por lo tanto queda demostrado lo anteriormente dicho.

1.1.3 Raíces de Multiplicidad m.

Proposición 1.1.1: Sea w una raíz de Σ, de multiplicidad m, i.e.

$$w_{r+1} = w_{r+2} = \cdots = w_{r+m} = w.$$

entonces $(\Sigma - wI)$ es de rango p-m. Ademas $P_0 = (p_{r+1}, \ldots, p_{r+m})$ está unívocamente determinado excepto por multiplicación a la derecha por una matriz ortogonal.

Demostración:

De la hipótesis tenemos que $p_{r+1}, \ldots, p_{r+m}$ son m soluciones linealmente independientes de $(\Sigma - wI)p = 0$. Consideremos $\sum_{i=1}^{p} x_i p_i$ donde los x_i son escalares, si esta combinación lineal es solución debe verificar que:

$$\Sigma\left(\sum_{i=1}^{p} x_i p_i\right) = w\left(\sum_{i=1}^{p} x_i p_i\right)$$

entonces:

$$\sum_{i=1}^{p} x_i \Sigma p_i = w(\sum_{i=1}^{p} x_i p_i)$$

$\sum_{i=1}^{p} x_i \lambda_i p_i = \sum_{i=1}^{p} w\, x_i p_i \Longrightarrow x_i \lambda_i = w x_i$ como los $\lambda_i \neq w$; $i = 1,\ldots,r$.

$\Longrightarrow x_i = 0$ para $i = 1,2,\ldots,r$.

p_* es una solución de $(\Sigma - wI)\, p = 0$ entonces cualquier otro conjunto solución es un conjunto de combinaciones lineales de p_* i.e. $p_* A$, donde A es una matriz no singular; la exigencia de que este nuevo conjunto sea ortogonal:

$$I = (p_* A)'\,(p_* A) = A'\, p_*'\, p_* A = A'A$$

va a dar como resultado la necesidad de que A sea ortogonal.

1.2 RELACION ENTRE LA VARIANZA GENERALIZADA, VARIANZA TOTAL DEL SISTEMA; DEL VECTOR X Y LAS DE SU VECTOR DE COMPONENTES PRINCIPALES.

Proposición 1.2.1. Las transformaciones ortogonales de un vector aleatorio X, dejan invariante la varianza generalizada y la varianza total del sistema o traza de la matriz de covarianza.

Demostración:

Varianza Generalizada:

Sea $Y = CX$, C matriz ortogonal:

$\mathrm{Var}\,(Y) = C \Sigma C'$

$|\mathrm{Var}\,(Y)| = |C|\ |\Sigma|\ |C'| = |\Sigma|\ |C|^2 = |\Sigma|$, ya que $|C| = |C'| = 1$ o -1.

Varianza total del Sistema:

$$\sum \mathrm{Var}(X_i) = t_r \Sigma = t_r(\Sigma C'C) = t_r(C \Sigma C') = \sum_{i=1}^{p} (C_i \Sigma C_i') = \sum_{i=1}^{p} \mathrm{Var}(Y_i)$$

donde $Y_i = \sum_{j=1}^{p} C_{ij} X_j = C_i X$; donde C_i es la i-ésima fila de C.

1.3 EL ANALISIS EN COMPONENTES PRINCIPALES SOBRE VECTORES CON TRANSFORMACIONES.

Las componentes principales se ven afectadas por las escalas de medición; i.e. al realizar cambios en un vector, variarán sus componentes principales; esto es necesario para que las componentes principales tengan más sentido, ya que si las escalas de medidas del vector X son muy variadas las componentes principales serán la suma de cualquier variedad de medidas. En la mayoría de los casos se estandarizan los valores; i.e. se les resta la media y se divide entre su error estandar, se trabaja entonces con la matriz de correlación. Asi las componentes de este primer transformado no tendrán unidades de medida. Sin duda alguna un conocimiento profundo de la situación que se analiza y mucha experiencia ayudan en la interpretación del Análisis en componentes principales en el problema.

Primero analizaremos una transformación cualquiera y luego el caso particular de la matriz de correlación.

Consideremos la transformación:

$$Z = DX$$

donde D es una matriz diagonal, i.e. cada componente de X es multiplicada por un valor con su respectiva unidad de medida; por ejemplo si una componente de X está en pies, la respectiva de Y estará en pulgadas, si otra está en libras, su respectiva en Y estará en onzas, etc. Varianza de Z:

$$\text{Var}(DX) = D'\Sigma D = \Delta$$

Así el Análisis en Componentes Principales sobre Z con matriz de covarianza Δ se reduce a considerar:

$\xi' z$, tal que $\xi'\Delta\xi$ sea máxima, sujeta a $\xi'\xi = 1$

cuyo planteamiento es

ξ debe satisfacer $(\Delta - \delta I)\xi = 0$

y δ verificar $|\Delta - \delta I| = 0$

Expresemos lo anterior en términos de X.

Primero observamos que; la varianza a maximizar es:

$$\xi'\Delta\xi = (\xi' D')\Sigma(D\xi) = (D\xi)'\Sigma(D\xi)$$

si hacemos $\alpha = D\xi$ entonces $\xi' z = \xi' DX = \alpha' X$.

y $\xi'\Delta\xi = \alpha'\Sigma\alpha$, la restricción $\xi'\xi = \alpha' D^{-1}D^{-1}\alpha = \alpha' D^{-2}\alpha = 1$.

Además si ξ satisface:
$$(\Delta - \delta I)\xi = 0$$
$$(D'\Sigma D - \delta I)\xi = 0$$
$$(\Sigma D - \delta D^{-1})\xi = 0$$
$$(\Sigma - \delta D^{-2})(D\xi) = 0$$

entonces $D\xi = \alpha$ satisface $(\Sigma - \delta D^{-2})\alpha = 0$, lo que resulta de maximizar $\alpha'\Sigma\alpha$ sujeta a $\alpha' D^{-2}\alpha = 1$

Por lo tanto podemos considerar sobre X la combinación lineal $\xi' DX$ y desarrollar lo expuesto anteriormente.

Ejemplo:

Sea

$$D = \begin{pmatrix} \frac{1}{\sqrt{\sigma_{11}}} & & & \\ & \frac{1}{\sqrt{\sigma_{22}}} & & \\ & & \ddots & \\ & & & \frac{1}{\sqrt{\sigma_{pp}}} \end{pmatrix}$$

así

$$Z = DX = \begin{pmatrix} \frac{x}{\sqrt{\sigma_{11}}} \\ \vdots \\ \frac{x_p}{\sqrt{\sigma_{pp}}} \end{pmatrix}$$

por lo tanto

$Var\ (Z) = D'\Sigma D$

esta matriz tiene por entrada en la (i,j) - ésima posición el valor:

$$\frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\ \sqrt{\sigma_{jj}}} = \rho_{ij}$$

que corresponde a las correlaciones simples entre X_i y X_j, concluimos asi que el análisis en componentes principales puede considerarse sobre Z con la matriz de correlación de X.

1.4 ESTIMACION DE LAS COMPONENTES PRINCIPALES.

A continuación veremos que las funciones inyectivas de los parámetros, tiene por estimadores de máxima verosimilitud la misma función pero de los estimadores de máxima verosimilitud de los parámetros.

Proposición sobre los Máximos.

Proposición 1.4.1:

Sea h una función sobre cierto conjunto K, que toma valores reales y sea w una función que toma un solo valor con inversa que también toma un solo valor, de K en K^0 y sea $\xi \in K$ y $\xi^0 \in K^0$ así:

$$f(\xi^0) = h\left[w^{-1}(\xi^0)\right] = f(w(\xi)) = h(\xi)$$

si $h(\xi)$ alcanza un máximo en $\xi = \xi_0$ entonces $f(\xi^0)$ alcanza un máximo en $\xi^0 = \xi^0_* = w(\xi_*)$. Si el máximo de h es único en ξ_*, igualmente es único el máximo de f en ξ^0_*.

Demostración:

Como h alcanza un máximo en ξ_*, se tiene que para todo $\xi \in k$.

$$h(\xi) \leqslant h(\xi_*).$$

por lo tanto, para toda $\xi^o \in k^o$

$f(\xi^o) = h(w^{-1}(\xi^o)) = h(\xi) \leqslant h(\xi_*) = f(w(\xi_*)) = f(\xi_*^o)$

así f alcanza un máximo en ξ_*^o.

Corolario 1.4.2:

Estimadores de Máxima Verosimilitud.

Sean $p_1, p_2, \ldots, p_n$ parámetros de una determinada distribución y sus estimadores de máxima verosimilitud $\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_n$ respectivamente. Si existe una función de $p_1, p_2, \ldots, p_n$ a $f_1(p_1, p_2, \ldots, p_n)$, $f_2(p_1, p_2, \ldots, p_n), \ldots, f_n(p_1, \ldots, p_n)$ que sea inyectiva, entonces $f_1(\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_n), f_2(\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_n), \ldots, f_n(\hat{p}_i, \ldots, \hat{p}_n)$ son estimadores de máxima verosimilitud de $f_1(p_1, \ldots, p_n), f_2(p_1, \ldots, p_n), \ldots, f_n(p_1, \ldots, p_n)$.

Demostración:

Como $\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_n$ son estimadores de máxima verosimilitud de $p_1, \ldots, p_n$ esto significa que la función de máxima verosimilitud:

$$L(p_1, p_2, \ldots, p_n).$$

alcanzó un máximo en $(\hat{p}_1, \ldots, \hat{p}_n)$

Sea $h(p_1, p_2, \ldots, p_n) = (f_1(p_1, \ldots, p_n), \ldots, f_n(p_1, \ldots, p_n))$ inyectiva, así cada f_i es inyectiva para todo $i = 1, 2, \ldots, n$

entonces tenemos:

$$\begin{aligned} L(p_1, p_2, \ldots, p_n) &= L(h^{-1}(f_1(p_1, \ldots, p_n), \ldots, f_n(p_1, \ldots, p_n)) \\ &= Q\ (f_1(p_1, \ldots, p_n), \ldots, f_n(p_1, \ldots, p_n)). \end{aligned}$$

donde por la proposición anterior la función Q alcanza un máximo en $f_1(\hat{p}_1, \ldots, \hat{p}_n), \ldots, f_n(\hat{p}_1, \ldots, \hat{p}_n)$ y por lo tanto son estimadores de máxima verosimilitud de $f_1(p_1, \ldots, p_n), \ldots, f_n(p_1, \ldots, p_n)$.

Teorema 1.4.3:

Estimadores de Máxima Verosimilitud de las componentes principales.

Sea $x_1, x_2, \ldots, x_N$ N- observaciones de $N(u, \Sigma)$ donde Σ es una matriz con p-valores propios distintos. Entonces un conjunto de estimadores de máxima verosimilitud de

$$S_1, S_2, \ldots, S_p \text{ y de } p_1, p_2, \ldots, p_p$$

son las raices $k_1 > \ldots > k_p$ de $|\hat{\Sigma} - kI| = 0$

y el conjunto de vectores $b_1, b_2, \ldots, b_p$ que satisfacen

$$(\hat{\Sigma} - k_i I)\, b^{(i)} = 0$$
$$b^{(i)\prime} b^{(i)} = 1.$$

donde $\hat{\Sigma}$ es el estimador de máxima verosimilitud de Σ.

Demostración:

Como cada una de las raices tiene orden de multiplicidad uno, su espacio asociado es generado por p_i, pero en este espacio solo hay dos con norma 1, p_i y $-p_i$ se puede tomar la decisión de escoger el que tenga la primera componente positiva; entonces u, S, P es una función de u, Σ que toma un solo valor, así por la proposición anterior el conjunto de estimadores de máxima verosimilitud de u, S, P es la misma función pero de $\hat{u}, \hat{\Sigma}$. Esta función está definida por las ecuaciones:

$$|\Sigma - kI| = 0$$
$$(\Sigma - KI)b = 0$$
$$b'b = 1 \quad \text{con } b \text{ de primera componente}$$

positiva.

1.4.1 Reconstrucción y Factorización de la matriz de Covarianza.

Dado que:

$$S = P'\Sigma P$$

tenemos que:

$$\Sigma = PSP'$$

Como:

$$PS = (S_1P_1, S_2P_2 \dots S_pP_p)$$

entonces $$\sum_{k=1}^{p} S_k P_k^{(i)} P_k^{(j)}$$

es la (i,j) - ésima entrada de la matriz PSP'

este desarrollo algebraico nos permite obtener a partir de las estimaciones de los valores y vectores propios, una estimación de $\hat{\Sigma}$

$$\hat{\Sigma} = \hat{P}\,\hat{S}\,\hat{P}'$$

Factorización de la matriz de Covarianza Muestral.

Sea $$\hat{S} = \hat{S}_{\sqrt{s_i}}\ \hat{S}_{\sqrt{s_i}}$$

donde $S_{\sqrt{s_i}}$ es la matriz de elementos $\sqrt{s_i}$ en la diagonal

y sea $$\hat{T} = \hat{P}\,S_{\sqrt{s_i}}$$

Así $$\hat{\Sigma} = \hat{T}\,\hat{T}'$$

Cuando los valores propios son todos distintos y se elige los $\hat{p}_i = b_i$ de primera componente positiva; $\hat{P}$ es única para cada estimación de $\hat{\Sigma}$ y por lo tanto su descomposición es única.

CAPITULO II

CORRELACIONES CANONICAS

2.1 LAS CORRELACIONES CANONICAS Y LAS VARIABLES CANONICAS.

El método de las correlaciones canónicas pretende obtener un conjunto de parejas (Z_i, W_i) de máxima correlación; $\rho(Z_i, W_i)$. De pareja a pareja la correlación debe ser cero. En el caso de distribución normal de las variables, esto equivale a independencia entre las parejas de máxima correlación. Estas parejas; denominadas variables canónicas son las combinaciones lineales de los componentes de los vectores aleatorios X, Y respectivamente. Estos vectores Aleatorios tienen una distribución conjunta y son de p_1 y p_2 componentes, respectivamente; con $p_1 \leqslant p_2$.

Al eliminar las parejas de menor correlación, se reduce la dimensión a un valor menor que p_1.

En el caso común de que se desconozca la matriz de covarianza y por lo tanto las variables y correlaciones canónicas, se pasa a la estimación; a través de una muestra de n-observaciones, realizándose un trabajo algebraico similar a la situación; con Σ conocida, a través de su estimador de máxima verosimilitud $\hat{\Sigma} = (1/N)A$ o su estimador insesgado $S = \frac{1}{N-1} A = \frac{N}{N-1} \hat{\Sigma}$

En algunas situaciones, la pareja de variables canónicas puede visualizarse como una variable y su mejor predictor.

2.2 DETERMINACION DE LAS CORRELACIONES Y VARIABLES CANONICAS POBLACIONALES.

A continuación daremos la definición de variables y correlaciones canónicas. Seguidamente un teorema que nos indica como obtener las correlaciones canónicas y las variables canónicas. Inmediatamente se observa que se obtienen p_1-parejas de variables canónicas, a través de otro teorema, y finalmente que estas correlaciones canónicas son positivas.

DEFINICIONES BASICAS.

Definición 2.2.1: k-ésimo par de variables canónicas:

es el par de combinaciones lineales:

$$T_k = a^{(k)'} X^{(1)} \text{ y } W_k = b^{(k)'} X^{(2)}$$

de varianza uno y de correlación máxima de entre las no-correlacionados con los primeros (k-1)-parejas de variables canónicas. Donde:

$$X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}$$

$X^{(1)}$ de p_1 componentes

$X^{(2)}$ de p_2 componentes, $p_1 + p_2 = p$ y $p_1 \leqslant p_2$

$E(X) = 0 \quad E(XX') = \Sigma$ matriz de covarianza poblacional. Se supone conocida y definida positiva.

Definición 2.2.2: k-ésima correlación canónica:

es la correlación entre el k-ésimo par de variables canónicas.

2.2.1 Teorema de Existencia de las Correlaciones Canónicas.

Teorema 2.2.1:

La k-ésima correlación canónica es la k-ésima mayor de las raíces de

$$\begin{vmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda\Sigma_{22} \end{vmatrix} = 0$$

y los coeficientes que caracterizan el par de variables canónicas satisfacen:

$$\begin{pmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda\Sigma_{22} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = 0$$

con

$$E(T^2) = a'\Sigma_{11}\, a = 1$$

$$E(W^2) = b'\Sigma_{22}\, b = 1$$

Demostración:

La partición de la matriz de covarianza Σ correspondiente a la que se hizo en X es:

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

como $\Sigma > 0$, entonces $\Sigma_{11}, \Sigma_{22}, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ son definidos positivos y por lo tanto invertibles.

Debemos determinar a y b tales que $T = a'X^{(1)}$ y $W = b'X^{(2)}$ sean de correlación máxima, i.e. como, $E(T) = E(W) = 0$; por maximizar:

$$\rho_{T,W} = E(TW) = a'\,\Sigma_{12}b \qquad (2.2.1)$$

Como además:

$$\rho_{T,W} = \frac{\operatorname{cov}(T,W)}{\sqrt{Var(T)\ Var(W)}} = \frac{\alpha\gamma \operatorname{cov}(T,W)}{\sqrt{Var(\alpha T)\ Var(\gamma W)}} = \rho_{\alpha T,\gamma W}$$

se considerará el caso en que:

$$E(T^2) = a'\Sigma_{11}\, a = 1 \qquad (2.2.2)$$

$$E(W^2) = b'\Sigma_{22}\, b = 1 \qquad (2.2.3)$$

La obtención de los puntos estacionarios, nos conduce al método de Multiplicadores de Lagrange:

$$f = a'\Sigma_{12}b - 1/2\,\lambda\,(a'\Sigma_{11}a-1)-1/2\,\mu(b'\Sigma_{22}b-1)$$

Derivando parcialmente con respecto a $\underline{a}$ y a $\underline{b}$ e igualando a cero; tenemos:

$$\frac{\partial f}{\partial a} = \Sigma_{12}b-\lambda\Sigma_{11}\,a = 0$$

$$\frac{\partial f}{\partial b} = \Sigma_{21}a-\mu\Sigma_{22}\,b = 0$$

obteniendo el sistema:

$$\Sigma_{12}b - \lambda\Sigma_{11}a = 0 \quad \text{i.e. } \Sigma_{12}b = \lambda\Sigma_{11}a \qquad (2.2.4)$$

$$\Sigma_{21}a - \mu\Sigma_{22}b = 0 \quad \text{i.e. } \Sigma_{21}a = \mu\Sigma_{22}b \qquad (2.2.5)$$

premultiplicando (2.2.4) y (2.2.5) por a' y b' respectivamente.

$$a'\Sigma_{12}b - \lambda\, a'\Sigma_{11}a = 0$$

por (2.2.2) tenemos que:

$$\lambda = a'\Sigma_{12}b. \quad \text{(Correlación a maximizar)}$$

$$b'\Sigma_{21}a - \mu\, b\,\Sigma_{22}b = 0$$

por (2.2.3) se da que:

$$\mu = b'\Sigma_{21}a = \lambda \qquad (2.2.6)$$

de donde los valores de λ que son soluciones de interés deberán cumplir con $|\lambda| \leqslant 1$.

Expresemos (2.2.4) y (2.2.5) en forma matricial:

$$\begin{pmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda\Sigma_{22} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = 0 \qquad (2.2.7)$$

Para obtener soluciones no-triviales.

$$\begin{vmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda\Sigma_{22} \end{vmatrix} = 0 \qquad (2.2.8)$$

Considerando $\mu = \lambda$ por (2.2.6), y premultiplicando (2.2.4) por Σ_{11}^{-1} y (2.2.5) por λ obtenemos:

$$\Sigma_{11}^{-1}\ \Sigma_{12}b = \lambda a \qquad (2.2.9)$$

$$\lambda\Sigma_{12}' a = \lambda^2 \Sigma_{22}b \qquad (2.2.10)$$

Reemplazando (2.2.9) en (2.2.10) tenemos:

$$\Sigma_{21}\ \Sigma_{11}\ \Sigma_{12}b = \lambda^2 \Sigma_{22}b.$$

por lo tanto

$$(\Sigma_{21}\ \Sigma_{11}^{-1}\ \Sigma_{12} - \lambda^2 \Sigma_{22})\ b = 0 \qquad (2.2.11)$$

Analizemos la matriz $\Sigma_{21}\ \Sigma_{11}^{-1}\ \Sigma_{12} = B.$ (2.2.12)

es $p_2 \times p_2$, de rango a lo más $p_1 \leqslant p_2$.

Si $r(\Sigma_{21}) = r(\Sigma_{12}) = k < p_1$, B es semi-definida positiva ($B \geqslant 0$) y sus raices $\lambda^2 \geqslant 0$, son $\lambda_1^2 \geqslant \lambda_2^2 \geqslant \ldots \geqslant \lambda_k^2 \geqslant 0$

si $r(\Sigma_{21}) = r(\Sigma_{21} = \Sigma_{12}') = p_1$; $B > 0$ y sus raíces $\lambda^2 > 0$.

Sea $\Sigma_{22} = A$ y $\lambda^2 = \ell$ (2.2.13)

Reemplazando (2.2.12) y (2.2.13) en (2.2.11)

$$(B - \ell A)b = 0$$

Para obtener soluciones no-triviales:

$$|B - \ell A| = 0 \qquad (2.2.14)$$

En forma similar obtenemos:

$$\Sigma_{12}\,\Sigma_{22}^{-1}\Sigma_{21}\,a = \lambda^2\,\Sigma_{11}\,a$$

$$(\Sigma_{12}\,\Sigma_{22}^{-1}\Sigma_{21} - \lambda^2\Sigma_{11})\,a = 0 \qquad (2.2.15)$$

Realizando el mismo análisis anterior y sustituciones semejantes tenemos:

$$\Sigma_{12}\,\Sigma_{22}^{-1}\,\Sigma_{21} = C$$
$$\Sigma_{11} = D;\quad \lambda^2 = \xi$$

Para obtener soluciones no-triviales:

$$|C - \xi D| = 0 \qquad (2.2.16)$$

si $C \geqslant 0$ las raíces de (2.2.16) son $\xi \geqslant 0$.

si $C > 0$ las raíces de (2.2.16) son $\xi > 0$.

como λ es real, las $p = p_1 + p_2$ raíces de(2.2.8) o (2.2.14) y (2.2.16) pueden ser ordenadas;

sea $\lambda_1 \geqslant \lambda_2 \geqslant \dots \geqslant \lambda_p$ dicha ordenación.

Así la correlación máxima será:

$$a'^{(1)}\,\Sigma_{12}b^{(1)} = \lambda_1$$ donde $a^{(1)}$ y $b^{(1)}$ satisfacen (2.2.7) para $\lambda = \lambda_1$.

Continuamos y buscamos ahora:

a y b tales que: $T = a'X^{(1)}$ $W = b'\,X^{(2)}$, tengan correlación máxima entre las combinaciones lineales no-correlacionados con $T_1 = a^{(1)'}X^{(1)}$ y $W_1 = b^{(1)'}X^{(2)}$ con las condiciones señaladas anteriormente.

Se debe maximizar $a'\Sigma_{12}b$

Sujeta a:

$$a'\,\Sigma_{11}a = 1 \qquad (2.2.17)$$

$$b'\Sigma_{22}b = 1 \qquad (2.2.18)$$

$$a'\Sigma_{11}a^{(1)} = 0 \qquad (2.2.19)$$

$$b'\Sigma_{22}b^{(1)} = 0 \qquad (2.2.20)$$

Las ecuaciones (2.2.17) y (2.2.18) expresan la no-correlación entre T_1 y T e W_1 y W; i.e.

$$E(T_1T) = a'\Sigma_{11}\, a^{(1)} = 0$$

$$E(W_1W) = b'\Sigma_{22}\, b^{(1)} = 0$$

A su vez (2.2.19) y (2.2.20) implican $E(TW_1) = 0$ y $E(WT_1) = 0$, pues por (2.2.4).

$a'\Sigma_{12}b^{(1)} = \lambda_1 a'\Sigma_{11}a^{(1)}$ y como $a'\Sigma_{11}\, a^{(1)} = 0$; tenemos que:

$E(T_1W) = 0$ en forma similar.

$$E(WW_1) = 0 \Longrightarrow E(WT_1) = 0.$$

Utilizando el método anterior:

$$g = a'\Sigma_{12}b - \frac{1}{2}\lambda(a'\Sigma_{11}a-1) -1/2\mu(b'\Sigma_{22}b-1)+ \xi_1 a'\Sigma_{11}\, a + \xi_2 b'\Sigma_{22}b.$$

Derivando parcialmente con respecto a a y después con respecto a b, e igualando a cero, resulta:

$$\frac{\partial g}{\partial a} = \Sigma_{12}b - \lambda\Sigma_{11}a + \xi_1\, \Sigma_{11}a^{(1)} = 0 \qquad (2.2.21)$$

$$\frac{\partial g}{\partial b} = \Sigma_{21}a - \mu\Sigma_{22}b + \xi_2\Sigma_{22}b^{(1)} = 0 \qquad (2.2.22)$$

premultiplicando por a' y b' (2.2.21) y (2.2.22) respectivamente

$$a'\Sigma_{12}b = \lambda$$

$$b'\Sigma_{21}a = \mu$$

Así

$$\lambda = u$$

en forma similar premultiplicando ahora a (2.2.21) y (2.2.22) por $a^{(1)'}$ y $b^{(1)'}$ respectivamente se tiene.

$$\underbrace{a^{(1)'}\Sigma_{12}b}_{=0} - \lambda\, \underbrace{a^{(1)'}\Sigma_{11}a}_{=0} + \xi_1 \underbrace{a^{(1)'}\Sigma_{11}a^{(1)}}_{=1} = 0$$

entonces

$$\xi_1 = 0$$

De manera similar, resulta $\xi_2 = 0$.

Por lo tanto hemos llegado a la ecuación (2.2.8) y como solución tomamos $\lambda = \lambda_2$ y $a^{(2)}$, $b^{(2)}$ vectores que satisfacen (2.2.7) para $\lambda = \lambda_2$.

En esta forma, en la k-ésima etapa tenemos k- parejas de combinaciones lineales de varianza 1, y de correlación máxima dos a dos y no correlacionadas con las parejas anteriores, son estas:

$$T_1 = a^{(1)'}X^{(1)} \qquad T_2 = a^{(2)'}X^{(1)} \quad \ldots \quad T_k = a^{(k)'}X^{(1)}$$

$$W_1 = b^{(1)'}X^{(2)} \qquad W_2 = b^{(2)'}X^{(2)} \quad \ldots \quad W_k = b^{(k)'}X^{(2)}$$

cuyas correlaciones son:

$$\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_k \geqslant 0.$$

2.2.1 EL CONJUNTO DE LAS VARIABLES CANONICAS.

Supongamos que hemos podido llegar hasta la m-ésima etapa, en el mismo proceso; veremos a continuación que $m = p_1$.

Teorema 2.2.2:

El número de parejas de variables canónicas que se pueden obtener es $p_1 \leqslant p_2$.

Demostración:

Sea $A = (a^{(1)}, a^{(2)}, \ldots, a^{(m)})$, $B_1 = (b^{(1)}, b^{(2)}, \ldots, b^{(m)})$ y

$$L = \begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ & & \ddots & \\ 0 & & & \lambda_m \end{pmatrix}$$

; donde A es $p_1$xm, B_1 es $p_2$xm y L es mxm.

Si expresamos para las m-parejas de variables canónicas y las máxima correlaciones canónicas las ecuaciones (2.2.17) a (2.2.20) nos da:

$$A'\textstyle\sum_{11}A = I \tag{2.2.23}$$

como $A'\sum_{11}A$ tiene rango m = rango I, entonces $m \leqslant p_1$ dado que $r(\sum_{11}) = p_1$, pues el rango de la matriz producto es menor o igual que el mínimo de los rangos de los factores, entonces menor que el rango de cualquier matriz factor.

Así dado que $m \leqslant p_1$, se considerará $m < p_1$, llegando a una contradicción; viéndose que existe otra solución.

Dado que $A'\sum_{11} = M$ es matriz mxp_1 de rango m, así su espacio nulo tiene dimensión (p_1-m), siendo una base de este espacio los vectores li de una matriz de (p_1-m) vectores p_1-dimensionales. (Es decir: el espacio generado por M tiene dimensión m, y así su suplementario ortogonal tiene dimensión p_1-m, siendo una base el conjunto de p_1-m vectores columnas li de una matriz $p_1x(p_1-m)$), sea esta la matriz E.

De donde

$$A'\sum_{11} E = 0 \qquad (2.2.24)$$

En forma similar, existe la matriz F de rango (p_2-m) tal que

$$B_1\sum_{22}F = 0 \qquad (2.2.25)$$

Al expresar en forma matricial (2.2.4)

$$B_1'\sum_{21} = LA'\sum_{11} \qquad (2.2.26)$$

Multiplicando (2.2.26) por E, tenemos:

$$B_1'\sum_{21} E = LA'\sum_{11}E = 0 \qquad (2.2.27)$$

(2.2.24) y (2.2.27) expresan que si los vectores columnas de E son los coeficientes de la combinaciones lineales $\mathcal{E} = e_i X^{(1)}$, ellas son no correlacionadas con los $T_i = a_i X^{(1)}$ y los $W_i = b_i X^{(2)}$, $i = 1,2,\ldots,m$

De igual forma expresando (2.2.5) en forma matricial y post-multiplicando esta por F tenemos:

$$A'\sum_{12} = LB_1'\sum_{22} \qquad (2.2.28)$$

$$A'\sum_{12}F = LB_1'\sum_{22}F = 0. \qquad (2.2.29)$$

Al igual que (2.2.24) y (2.2.27) las ecuaciones (2.2.25) y (2.2.29) expresan las no-correlaciones entre las c.l. fi $X^{(2)}$, con los $b_i X^{(2)}$ y los $a_i X^{(1)}$, $i = 1,2,\ldots,m$.

Dado que E es de rango por columnas, $E'\sum_{11}E$ es definida positiva y por lo tanto invertible; $|E'\sum_{11}E| \neq 0$. Así mismo $F'\sum_{22}F$ es definida positiva y por lo tanto invertible; $|F'\sum_{22}F| \neq 0$.

y por lo tanto existe al menos una raíz de

$$\begin{vmatrix} -rE'\sum_{11}E & E'\sum_{12}F \\ F'\sum_{21}E & -rF'\sum_{22}F \end{vmatrix} = 0 \qquad (2.2.30)$$

pues esta determinante es un polinomio en r de orden $(-r)^{p_1-m}$ $|E'\Sigma_{11}E|(-r)^{p_2-m}|F'\Sigma_{22}F| = (-r)^{p-2m}|E'\Sigma_{11}E||F'\Sigma_{22}F|$, p-2m, pues los otros términos tienen menor grado en r, porque una o más filas de las primeras p_1-columnas no contienen a r al desarrollar la determinante por menores de las primeras (p_1-m) columnas.

De acuerdo con (2.2.4),(2.2.5) y (2.2.8) existen α y β tales que:

$$E'\Sigma_{12}F\beta = r E'\Sigma_{11}E\alpha \qquad (2.2.31)$$

$$F'\Sigma_{21}E\alpha = r F'\Sigma_{22}F\beta \qquad (2.2.32)$$

Haciendo $E\alpha = g$, $F\beta = h$ se demostrará que r,g,h forman una nueva solución.

Sea $\Sigma_{11}^{-1}\Sigma_{12}h = \sigma$, entonces σ está en el espacio ortogonal (2.2.33) a $A'\Sigma_{11}$, pues; por (2.2.29).

$$A'\Sigma_{11}\sigma = A'\Sigma_{11}\Sigma_{11}^{-1}\Sigma_{12}F\beta = 0$$

y por lo tanto es una combinación lineal de los vectores columnas de E; i.e. $Ec = \sigma$. (2.2.34

así se tendrán las ecuaciones según (2.2.33)

$$\Sigma_{12}h = \Sigma_{11}\sigma \qquad (2.2.35)$$

De (2.2.34) tenemos

$$\Sigma_{12}F\beta = \Sigma_{11}Ec \qquad (2.2.36)$$

Premultiplicando (2.2.35) por la izquierda por E' y como $E'\Sigma_{11}E$ es invertible y por (2.2.31) tenemos que

$$E'\Sigma_{12}h = E'\Sigma_{11}\sigma = E'\Sigma_{11}Ec = r E'\Sigma_{11}E\alpha$$

Así $\quad c = r\alpha$

pues $Ec = \sigma = \gamma E \alpha = rg$; de (2.2.36) resulta:

$$\Sigma_{12} h = r\Sigma_{11} g \tag{2.2.37}$$

En igual forma

$$\Sigma_{21} g = r\Sigma_{22} h \tag{2.2.38}$$

y por lo tanto $r = \lambda^{(m+1)}$, $g = a^{(m+1)}$, $h = b^{m+1}$ es otra solución.

2.2.3 El Conjunto de las Correlaciones Canónicas.

Teorema 2.2.3:

Las raíces de (2.2.8) son p_1 raíces simétricas desde 1 a p_1 y una raíz $\lambda = 0$ de multiplicidad $p_2 - p_1$.

Demostración:

Las condiciones sobre varianza 1 y no-correlación sobre los p_1 parejas, se pueden resumir en forma matricial en la siguiente forma:

$$A'\Sigma_{11} A = I \tag{2.2.39}$$

$$A'\Sigma_{12} B_1 = L \tag{2.2.40}$$

$$B_1'\Sigma_{22} B_1 = I \tag{2.2.41}$$

Vamos a obtener de una manera más explicita las raíces de (2.2.7) a través de otra matriz no-singular.

Consideremos la matriz

$B_2 = (b^{(p_1+1)}, b^{(p_1+2)}, \dots, b^{p_2})$ de p_2 filas y $p_2 - p_1$ columnas satisfaciendo las siguientes condiciones.

$$B_2'\Sigma_{22} B_1 = 0 \tag{2.2.42}$$

$$B_2'\Sigma_{22} B_2 = I \tag{2.2.43}$$

Así los vectores de B_2 son ortogonales a los de $\sum_{22}B_1$, y también no correlacionados con los $b^{(p_1+i)}X^{(2)}$ y con $b^j\ X^{(2)}$ donde $i = 1,\ldots,p_2-p_1$, $j = 1,2,\ldots,p_1$ y los $b^{p_1+i}X^{(2)}$ tienen varianza uno. Sea $B = (B_1\ B_2)$ una matriz $p_2 \times p_2$, invertible; dado que $B'\sum_{22} B = I$ entonces $|B| = |B'| = \dfrac{1}{|\sum_{22}|} \neq 0$.

Verifiquemos que

$$B'\textstyle\sum_{22}B = I = \begin{pmatrix} B_1' \\ \\ B_2' \end{pmatrix} \sum_{22} (B_1,B_2) = \begin{pmatrix} B_1'\sum_{22} \\ \\ B_1'\sum_{22} \end{pmatrix} \quad (B_1\ B_2)$$

$$= \begin{pmatrix} B_1'\sum_{22} B_1 & B_1'\sum_{22} B_2 \\ B_2'\sum_{22} B_1 & B_2'\sum_{22} B_2 \end{pmatrix}$$

$$= \begin{pmatrix} I & 0 \\ \\ 0 & I \end{pmatrix} \quad = I_{p_2 \times p_2}$$

así el rango de B es p_2; pues es invertible y es $p_2 x p_2$.

Las raíces de (2.2.7) son las mismas que las de:

$$\begin{vmatrix} A' & 0 \\ 0 & B_1' \\ 0 & B_2' \end{vmatrix} \begin{vmatrix} -\lambda\sum_{11} & \sum_{12} \\ \sum_{21} & -\lambda\sum_{22} \end{vmatrix} \begin{vmatrix} A & 0 & 0 \\ 0 & B_1 & B_2 \end{vmatrix} \qquad (2.2.44)$$

ya que $\begin{vmatrix} A & 0 \\ 0 & B \end{vmatrix} = |A|\ |B| \neq 0$ pues $|A| \neq 0$, $|B| \neq 0$.

Realizando los productos matriciales indicados por (2.2.44) tenemos:

$$\begin{vmatrix} -\lambda A'\Sigma_{11} & A'\Sigma_{12} \\ B_1'\Sigma_{21} & -\lambda B_1'\Sigma_{22} \\ B_2'\Sigma_{21} & -\lambda B_2'\Sigma_{22} \end{vmatrix} \begin{vmatrix} A & 0 & 0 \\ 0 & B_1 & B_2 \end{vmatrix}$$

$$\begin{vmatrix} -\lambda A'\Sigma_{11} A & A'\Sigma_{12} B_1 & A'\Sigma_{12} B_2 \\ B_1'\Sigma_{21} A & -\lambda B_1'\Sigma_{22} B_1 & -\lambda B_1'\Sigma_{22} B_2 \\ B_2'\Sigma_{21} A & -\lambda B_2'\Sigma_{22} B_1 & -\lambda B_2'\Sigma_{22} B_2 \end{vmatrix}$$

$$\begin{vmatrix} -\lambda I & L & 0 \\ L & -\lambda I & 0 \\ 0 & 0 & -\lambda I \end{vmatrix} \qquad (2.2.45)$$

Veamos que $B_2'\Sigma_{21} A = 0$; significa por una parte, que los vectores columnas de B_2 son ortogonales a $\Sigma_{21} A$;

(2.2.45) es igual a: $\begin{vmatrix} Q & 0 \\ 0 & -\lambda I \end{vmatrix}$ (2.2.46)

donde $Q = \begin{pmatrix} -\lambda I & L \\ L & -\lambda I \end{pmatrix}$; Q es $2p_1 \times 2p_1$

Por lo tanto (2.2.46) es igual a:

$$|-\lambda I|\ |Q| \qquad (2.2.47)$$

Calculemos (2.2.47) factor a factor

$$|-\lambda I| = (-\lambda)^{p_2-p_1}$$

$$|Q| = |-\lambda I|\ |-\lambda I - L(-\lambda I)^{-1} L|$$

$$= |\lambda^2 I - L^2|$$

$$= \prod_{i=1}^{p_1} \lambda^2 - (\lambda^i)^2$$

Tenemos entonces que:

$$|-\lambda I|\ |Q| = (-\lambda)^{p_2-p_1} \prod(\lambda^2 - \lambda^{(i)2}) \qquad (2.2.48)$$

Así las raíces de (2.2.7) son las de (2.2.48) igualadas a cero.

$$(-\lambda)^{p_2-p_1} \prod(\lambda^2 - \lambda^{(i)2}) = 0 \Longleftrightarrow (\lambda^2 - \lambda^{(i)2}) = 0 \text{ ó } (-\lambda)^{p_2-p_1} = 0$$

Así

$$\lambda^2 = \lambda^{(i)^2} \Longrightarrow |\lambda| = |\lambda^{(i)}| \Longrightarrow \lambda = \pm\lambda^{(i)} \qquad i=1,2,\ldots,p_1$$

$(-\lambda)^{p_2-p_1} = 0 \Longrightarrow \lambda = 0$ de multiplicidad p_2-p_1.

Veamos que las raices $\lambda^{(i)}$ son iguales a los λ_i de (2.2.7) $i = 1,2,\ldots,p_1$ esto es así si los $\lambda^{(i)}$ son positivos, demostremos esto:

Sea $\lambda^{(i)}$, $a^{(i)}$, $b^{(i)}$ una solución i.e.

$$\Sigma_{12}\, b^{(i)} = -\lambda^{(i)}\Sigma_{11}\,(-a^{(i)})$$

$$\Sigma_{21}(-a^{(i)}) = -\lambda^{(i)}\Sigma_{22}\, b^{(i)}$$

Así también es una solución $-\lambda^{(i)}$, $-a^{(i)}$, $b^{(i)}$, pero se buscaba el máximo, por lo tanto se escoge como solución finalmente el positivo.

Así tenemos que:

$$(\lambda_1, \lambda_2, \ldots, \lambda_p) = (\lambda_1, \ldots, \lambda_{p_1}, 0, \ldots, 0, -\lambda_{p_2+1}, -\lambda_{p_2+2}, \ldots, -\lambda_p)$$

2.3 RESUMEN MATRICIAL DE LOS RESULTADOS OBTENIDOS.

Expresemos en forma matricial los resultados obtenidos con respecto a las c.c. y v.c.

$$T = A'X^{(1)} = \begin{pmatrix} T_1 \\ T_2 \\ \vdots \\ T_{p_1} \end{pmatrix} \qquad (2.3.1)$$

$$W_1 = B_1' \, X^{(2)} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_{p_1} \end{pmatrix} \qquad (2.3.2)$$

$$W_2 = B_2' \, X^{(2)} = \begin{pmatrix} w_{p_1+1} \\ w_{p_1+2} \\ \vdots \\ w_{p_2} \end{pmatrix} \qquad (2.3.3)$$

$$E\left[\begin{pmatrix} T \\ w_1 \\ w_2 \end{pmatrix} (T' \; W_1' \; W_2')\right] = \begin{pmatrix} E(TT') & E(TW_1') & E(TW_2') \\ E(w_1T') & E(W_1W_1') & E(W_1W_2') \\ E(w_2T') & E(W_2W_1') & E(W_2W_2') \end{pmatrix}$$

$$= \begin{pmatrix} I & L & 0 \\ L & I & 0 \\ 0 & 0 & I \end{pmatrix} \qquad (2.3.4)$$

observando (2.2. 44) y (2.2. 45) vemos que (2.3.4) se puede escribir como

$$\begin{pmatrix} A' & 0 \\ 0 & B_1' \\ 0 & B_2' \end{pmatrix} \begin{pmatrix} \sum_{11} & \sum_{21} \\ \sum_{12} & \sum_{22} \end{pmatrix} \begin{pmatrix} A & 0 & 0 \\ 0 & B_1 & B_2 \end{pmatrix} \qquad (2.3.5)$$

Observemos que (2.2.44) es la determinante de (2.3.4) y (2.3.5) siendo esta, la matriz de covarianza del vèctor transformado de X, llamémosle $Y = \phi X$; a esta transformación, donde:

$$\phi = \begin{pmatrix} A' & 0 \\ 0 & B_1' \\ 0 & B_2' \end{pmatrix} \text{, matriz pxp de rango } p_1$$

$$Y = \phi X = \begin{pmatrix} A' & 0 \\ 0 & B_1' \\ 0 & B_2' \end{pmatrix} \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} = \begin{pmatrix} T \\ W_1 \\ W_2 \end{pmatrix} = \begin{pmatrix} t_1 \\ \vdots \\ t_p \\ w_1^{(1)} \\ \vdots \\ w_{p_1}^{(1)} \\ w_1^{(2)} \\ \vdots \\ w_{p_2-p_1}^{(2)} \end{pmatrix}$$

donde las parejas de variables canónicas son:

$(t_1,w_1^{(1)}),\ (t_2,w_2^{(1)}),\ldots,(t_{p_1},w_{p_1}^{(1)})$.

$$E(YY') = \begin{pmatrix} I & L & 0 \\ L & I & 0 \\ 0 & 0 & I \end{pmatrix}$$

donde L es la matriz diagonal de las p_1 correlaciones canónicas.

2.4 Estimación de las Correlaciones y Variables Canónicas.

Para obtener estos estimadores, debe realizarse el mismo trabajo algebraico, hecho en el caso de Σ conocido, pero en este caso con su estimador de máxima verosimilitud.

$$\hat{\Sigma} = \frac{1}{N}\sum_{\alpha=1}^{N}(x_\alpha - \bar{x})(x_\alpha - \bar{x})'$$

Usaremos la proposición 1.4.1, para ellos debemos verificar que se cumple la hipótesis:

La función definida por las ecuaciones 2.2.2, 2.2.3, 2.2.7, 2.2.8, 2.2.42, 2.2.43; con las consideraciones de que:

(a) $B_1 = (b^{(1)} \ldots b^{p_1})$ sea la matriz de los $b^{(i)}$ tal que la primera entrada no-nula de cada columna sea positiva.

(b) $B_2 = (b^{p_1+1} \ldots b \quad)$ como son los vectores asociados a la raíz nula de multiplicidad $(p_1 - p_2)$, por proposición 1.1.1., B_2 A es solución; con A matriz ortogonal; esta falta de unicidad para B_2, se resuelve especificando una de ellas, por ejemplo que la submatriz $(p_1-p_2) \times (p_1-p_2)$ de las últimas filas sea triangular con los elementos de la diagonal positivos.

es inyectiva. Así la misma función de los estimadores de máxima verosimilitud es un estimador de máxima verosimilitud.

Por lo tanto tenemos que las raíces de:

$$\begin{vmatrix} -\lambda\hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & -\lambda\hat{\Sigma}_{22} \end{vmatrix} = 0$$

son estimadores de máxima verosimilitud de las correlaciones canónicas poblacionales.

Además $\hat{A}$ y $\hat{B}_1$ cumplen con

$$\begin{pmatrix} -\hat{\lambda}_j\hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & -\lambda_j\hat{\Sigma}_{22} \end{pmatrix} \begin{pmatrix} \hat{\alpha}^{(j)} \\ \hat{\gamma}^{(j)} \end{pmatrix} = 0$$

$$\hat{\alpha}^{(j)'} \hat{\Sigma}_{11} \alpha^{(j)} = 1$$

$$\hat{\gamma}^{(j)'} \hat{\Sigma}_{22} \hat{\gamma}^{(j)} = 1$$

y $\hat{\beta}_2$ satisface

$$\hat{\beta}_2' \hat{\Sigma}_{22} \hat{\beta}_1 = 0$$

$$\hat{\beta}_2' \hat{\Sigma}_{22} \hat{\beta}_2 = I$$

Así $\beta = (\beta_1 \ \beta_2)$, A y L

están unívocamente determinados

<u>Proposición 2.4.1.</u>

La determinante de

$$\begin{vmatrix} \beta\hat{\Sigma}_{11} & -\hat{\Sigma}_{12} \\ -\hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{vmatrix} = 0 \qquad (2.4.1)$$

es un polinomio en β de grado p_1 y sus raíces son reales y positivas.

Demostración:

Fórmula de Laplace:

$$\det A = (-1)^{i_1+\cdots+i_m} \sum_{j \in J} (-1)^{j_1+j_2+\cdots+j_m} D_I^j \, D_{I'}^{j'}$$

donde el conjunto $I = \{ i_1,\ldots,i_m \}$ es el conjunto fijo

$I = \{ 1,2,\ldots,p_1 \}$, $m = p_1$.

J es el conjunto de las partes de p_1 elementos de $[1,p]$ escritos en orden natural.

D_I^j es la determinante de la matriz donde se eliminan las filas I' y columnas j'. (matriz de las filas I y columnas J)

Así

$$\begin{vmatrix} \beta\hat{\Sigma}_{11} & -\hat{\Sigma}_{12} \\ -\hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{vmatrix} = (-1)^{\frac{p_1(p_1+1)}{2}} \left(\sum_{j \in J^*} (-1)^{j_1+\cdots+j_{p_1}} D_I^j \, D_{I'}^{j'} + (-1)^{\frac{p_1(p_1+1)}{2}} |\beta\hat{\Sigma}_{11}| \, |\hat{\Sigma}_{12}| \right)$$

donde J^* es el conjunto de las partes de tamaño p_1, menos la $\{ 1,2,\ldots,p_1 \}$.

Por lo tanto:

$$\begin{vmatrix} \beta\hat{\Sigma}_{11} & -\hat{\Sigma}_{12} \\ -\hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{vmatrix} = (-1)^{\frac{p_1(p_1+1)}{2}} \left(\sum_{j \in J^*} (-1)^{j_1+j_2+\cdots+j_{p_1}} D_I^j D_{I'}^{j'} + (-1)^{\frac{p_1(p_1+1)}{2}} \beta^{p_1} |\hat{\Sigma}_{11}| \, |\hat{\Sigma}_{22}| \right)$$

los otros sumandos $(-1)^{j_1+\cdots+j_{p_1}} D_I^j D_{I'}^{j'}$ serán determinantes de otras matrices en que algunas de las columnas de $\beta\hat{\Sigma}_{11}$ son reemplazadas por algunas columnas de $\hat{\Sigma}_{12}$, así son polinomios de grado menor que p_1 en β y además como $|\Sigma_{11}| \neq 0$ y $|\Sigma_{22}| \neq 0$ por ser submatrices de una matriz definida positiva. Así 2.4.1 es un polinomio a lo más de grado p_1

<u>Proposición 2.4.2.</u>

Las raíces de (2.4.1) son reales y positivas

<u>Demostración</u>:

$$\begin{vmatrix} \beta\hat{\Sigma}_{11} & -\hat{\Sigma}_{12} \\ -\hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{vmatrix} = |\hat{\Sigma}_{22}| \, |\beta\hat{\Sigma}_{11} - \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21}| = 0$$

como $|\hat{\Sigma}_{22}| \neq 0$, entonces:

$$|\beta\hat{\Sigma}_{11} - \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21}| = 0$$

Como $\hat{\Sigma}_{11}$ es definida positiva con probabilidad 1; existe E, no singular tal que

$$E' \hat{\Sigma}_{11} E = I$$

Si $\hat{\Sigma}_{12}$ es de rango completo por fila $r(\hat{\Sigma}_{12}) = p_1$

$\hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}$ es, definida positiva, simétrica

$$E'\hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}E, \text{ es definida positiva} \qquad (2.4.2)$$

y así

$$|\beta\hat{\Sigma}_{11} - \hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}| = |\beta I - E'\hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}E| = 0$$

y las raíces de (5) son los valores propios de(2.4.2)matriz simétrica def +, entonces sus valores propios son reales y positivos.

2.5 Obtención de las Variables Canónicas a partir de la Matriz de Correlación Muestral.

$$R = \frac{\sigma_{ij}}{\sqrt{\hat{\sigma}_{ii}}\sqrt{\hat{\sigma}_{jj}}} = \frac{S_{ij}}{\sqrt{S_{ii}S_{jj}}}$$

$$\hat{\sigma}_{ij} = \frac{1}{W}\sum(x_\alpha^{(1)} - \bar{x}^{(1)})(x_\alpha^{(2)} - \bar{x}^{(2)})' = \frac{N-1}{N}S_{ij}$$

Trataremos de expresar las relaciones (1), (2), (3), (4) en términos de R matrices de correlación muestral particionada en forma relacionada con $\begin{pmatrix} x_\alpha^{(1)} \\ x_\alpha^{(2)} \end{pmatrix} = x_\alpha$

$$R = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}$$

Además, consideremos la matriz:

$$S_1 = \begin{pmatrix} \sqrt{s_{11}} & & & 0 \\ & \sqrt{s_{22}} & & \\ 0 & & \ddots & \\ & & & \sqrt{s_{p_1p_1}} \end{pmatrix}$$

$$S_2 = \begin{pmatrix} \sqrt{s_{p_1+1,p_1+1}} & & & 0 \\ & \sqrt{s_{p_1+2,p_1+2}} & & \\ 0 & & \ddots & \\ & & & \sqrt{s_{pp}} \end{pmatrix}$$

$$s_{ij} = r_{ij}\sqrt{s_{ii}s_{jj}}\,, \quad r_{ii} = 1;\ i = 1,2,\ldots,p_1$$

$$S_{11} = \frac{1}{N-1} A_{11} \ddagger \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1p_1} \\ r_{21} & r_{22} & \cdots & r_{2p_1} \\ \vdots & & & \\ r_{p_1,} & r_{p_12} & \cdots & r_{p_1p_1} \end{pmatrix} \begin{pmatrix} \sqrt{s_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{s_{22}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sqrt{s_{p_1p_1}} \end{pmatrix}$$

$$= \begin{pmatrix} r_{11}\sqrt{s_{11}} & r_{12}\sqrt{s_{22}} & \cdots & r_{1p}\sqrt{s_{p_1p_1}} \\ r_{21}\sqrt{s_{11}} & r_{22}\sqrt{s_{22}} & \cdots & r_{2p}\sqrt{s_{p_1p_1}} \\ \vdots & & & \\ r_{p_1}\sqrt{s_{11}} & r_{p_12}\sqrt{s_{22}} & \cdots & r_{p_1p}\sqrt{s_{p_1p_1}} \end{pmatrix}$$

$$\begin{pmatrix} \sqrt{S_{11}} & \frac{S_{12}}{\sqrt{S_{11}}} & \cdots & \frac{S_{1p_1}}{\sqrt{S_{11}}} \\ \frac{S_{21}}{\sqrt{S_{22}}} & \sqrt{S_{22}} & \cdots & \frac{S_{2p_1}}{\sqrt{S_{22}}} \\ \vdots & \vdots & & \vdots \\ \frac{S_{p}}{\sqrt{S_{p_1p_1}}} & \frac{S_{p}}{\sqrt{S_{p_1p_1}}} & \cdots & \sqrt{S_{p_1p_1}} \end{pmatrix} = \begin{pmatrix} (S_{11} \quad S_{12} \cdots S_{1p_1}) \frac{1}{\sqrt{S_{11}}} \\ (S_{21} \quad S_{22} \cdots S_{rp_1}) \frac{1}{\sqrt{S_{22}}} \\ \vdots \\ (S_{p_1'} \quad S_{p_1'} \cdots S_{p_1p}) \frac{1}{\sqrt{S_{p_1}S_{p_1}}} \end{pmatrix}$$

$$S_1 a^{(j)} = \begin{pmatrix} \sqrt{S_{11}}\, a_1^{(j)} \\ \sqrt{S_{22}}\, a_2^{(j)} \\ \vdots \\ \sqrt{S_{p_1p_1}}\, a_{p_1}^{(j)} \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1p_1} \\ r_{21} & r_{22} & \cdots & r_{2p_1} \\ \vdots & & & \\ r_{p_1 2} & r_{p_1 2} & \cdots & r_{p_1p_1} \end{pmatrix} \begin{pmatrix} \sqrt{S_{11}}\, a_1^{(j)} \\ \sqrt{S_{22}}\, a_2^{(j)} \\ \vdots \\ \sqrt{S_{p_1p_1}}\, a_{p_1}^{(j)} \end{pmatrix}$$

$$R_{11}S_1 a^{(j)} = \begin{pmatrix} \sqrt{S_{11}}\, a_1^{(j)} + \frac{S_{12}}{\sqrt{S_{11}}} a_2^{(j)} + \cdots + \frac{S_{1p_1}}{\sqrt{S_{11}}} a_{p_1}^{(j)} \\ \frac{S_{21}}{\sqrt{S_{22}}} a_1^{(j)} + \sqrt{S_{22}}\, a_2^{(j)} \cdots\cdots + \frac{S_{2p_1}}{\sqrt{S_{22}}} a_{p_1}^{(j)} \\ \vdots \\ \frac{S_{p_1}}{\sqrt{S_{p_1p_1}}} a_1^{(j)} + \frac{S_{p_12}}{\sqrt{S_{p_1p_1}}} a_2^{(j)} + \cdots + \sqrt{S_{p_1p_1}}\, a_{p_1}^{(j)} \end{pmatrix}$$

$$\begin{pmatrix} \frac{1}{\sqrt{S_{11}}}(S_{11}a_1^{(j)} + \ldots + S_{1p_1}\, a_{p_1}^{(j)}) \\ \frac{1}{\sqrt{S_{22}}}(S_{21}a_1^{(j)} + S_{22}a_2^{(j)} + \ldots + S_{2p_1}\, a_{p_1}^{j}) \\ \vdots \\ \frac{1}{\sqrt{S_{p_1p_1}}}(S_{p_11}a_1^{(j)} + S_{p_12}a_2^{(j)} + \ldots + S_{p_1p_1}a_{p_1}^{j} \end{pmatrix}$$

$$R_{12}S_2C^{(j)} = \begin{pmatrix} r_{1p_1+1} & r_{1p_1+2} & \ldots & r_{1p} \\ r_{2p_1+1} & r_{2p_1+2} & & r_{2p} \\ \vdots & \vdots & & \vdots \\ r_{p_1p_1+1} & r_{p_1p_1+2} & & r_{p_1p} \end{pmatrix} \begin{pmatrix} \sqrt{S_{p_1+1,p_1+1}} & = 0 \ldots & 0 \\ 0 & \sqrt{S_{p_1+2p_1+2}} \ldots & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & \sqrt{S_{pp}} \end{pmatrix}$$

$$\begin{pmatrix} \frac{S_{1p_1+1}}{\sqrt{S_{11}}} & \frac{S_{1p_1+2}}{\sqrt{S_{11}}} & \ldots & \frac{S_1}{\sqrt{S_{11}}} \\ \vdots & \vdots \quad \vdots & \vdots & \vdots \end{pmatrix} C^{(j)} = \left(\sqrt{S_{ii}}\sum S_{ik}C_k^{(j)}\right)$$

Observamos que:

$l_j\, R_{11}\, S_1\, a^{(j)}$ es un vector cuyos componentes i-ésima es:

$$l_j\, \sqrt{S_{ii}}\left(\sum_{k=1}^{p} S_{ik}a_k^{(j)}\right) \quad i = 1,2,\ldots,p_1$$

y

$R_{12}\ S_2\ c^{(j)}$ es un vector cuya-ésima-componente es:

$$\sqrt{S_{ii}}\ \sum_{p_1+1}^{p} S_{ik} c_k^{(j)} \qquad\qquad i = 1,2,\ldots p_1$$

Así el vector de componentes

$l_j(\sqrt{S_{ii}} \sum_{k=1}^{p_1} S_{ik} a_k^{(j)})$ difiere del vector de componentes

$l_j \sum_{k=1}^{p_1} S_{ik} a_k^j$ en el factor $\sqrt{S_{ii}}$, y así para $i = 1,2,\ldots,p_1$

Igualmente: (Multiplicar una matriz diagonal por un vector $\equiv$

Multiplicar cada componentes del vector por el respectivo elemento de la diagonal).

Las i-ésimas componentes de los vectores difieren en:

$$\sum_{k=p_1+1}^{p} S_{ik}\ c_k^{(j)} \quad y \quad \sqrt{S_{ii}} \sum_{k=p_1+1}^{p} S_{ik}\ c_k^{(j)}$$

así $S_{12}\ c^j = l_j\ S_{11}\ a^{(j)}$ se multiplica ambos miembros por S_1.

$$S_1\ S_{12}\ C^{(j)} = l_j\ S_1\ S_{11}\ a^{(j)}$$

$$R_{12}\ S_2\ c^{(j)} = l_j\ R_{11}\ S_1\ a^{(j)}$$

entonces $S_1\ S_{12}\ c^{(j)} = R_{12}\ S_2\ c^{(j)}$ y $l_j\ S_1\ S_{11}\ a^{(j)} = l_j\ R_{11} S_1\ a^{(j)}$

En forma similar ocurre con el paso de:

$S_{21}\ a^{(j)} = l_j\ S_{22}\ c^{(j)}$ a $R_{21}(S_1 a^{(j)}) = l_j\ R_{22}(S_2 c^{(j)})$

$S_1\ S_{21}\ a^{(j)} = l_j\ S_1 S_{22} c^{(j)}$

donde:

$$S_1 S_{21} a^{(j)} = R_{21}\ S_1\ a^{(j)}$$

$$l_j\ S_1\ S_{22} c^{(j)} = l_j\ R_{22}\ (S_2\ c^{(j)})$$

Para pasar de $a^{(j)}\ S_{11}\ a^{(j)} = 1$, $a(S_1 a^{(j)})' R_{11}(S_1 a^{(j)}) = 1$

$$\text{Como } S_1\ a^{(j)} = \begin{pmatrix} \sqrt{S_{11}}\ a_1^{(j)} \\ \vdots \\ \sqrt{S_{p\,p}}\ a_p^{(j)} \end{pmatrix}$$

$$\text{Entonces } (S_1 a^{(j)})' R_{11}(S_1\ a^{(j)}) = \sum_{i,k} r_{ik}(\sqrt{S_{ii}}\ a_i^{(j)})(\sqrt{S_{kk}} a_k^{(j)})$$

$$a^{j'}\ S_{11} a^j = \sum_{i,k} S_{ik} a_i^{(j)}\ a_k^{(j)}$$

Entonces:

$$S_1'\ a^{(j)'} S_{11}\ a^{(j)} S_1 = (S_1 a^{(j)})'\ R_{11}(S_1 a^{(j)})$$

Similarmente para la obtención de $(S_2 c^{(j)})'\ R_{22}(S_2 c^{(j)}) = 1$

a partir de:

$$c^{(j)'} S_{22} c^{(j)} = 1$$

2.6 Correlaciones Canónicas en el Modelo de Regresión.

En el modelo de Regresión; donde $E(X^{(1)}) = u^{(1)} + \beta(x_{\Phi}^{(2)} - \bar{x}^{(2)})$ y $E(X_{\Phi}^{(1)} - E(X_{\Phi}^{(1)}))(X_{\Phi}^{(1)} - E(X_{\Phi}^{(1)}))' = \Psi$; estos parámetros son estimador por:

$$\beta = S_{12}\, S_{22}^{-1}$$

$$S_{11} - \beta S_{22} \beta' = S_{11} - S_{12}\, S_{22}^{-1}\, S_{21}$$

Como $\beta = \Sigma_{12} \Sigma_{22}^{-1}$

donde $\hat{\Sigma}_{ij} = \dfrac{N-1}{N}\, S_{ij}$

$$\hat{\beta} = \hat{\Sigma}_{12}\, \hat{\Sigma}_{22}^{-1} = \frac{N-1}{N}\, S_{12}\, \frac{N}{N-1}\, S_{22}^{-1} = S_{12}\, S_{22}^{-1}$$

En igual forma ocurre con

$$\hat{V}(x_{\alpha}^{(1)} / x_{\alpha}^{(2)}) = S_{11} - \beta S_{22} \beta' = S_{11} - S_{12} S_{22}^{-1}\, S_{21}$$

Sean las raíces de $|S_{12}S_{22}^{-1}S_{21} - r(S_{11} - S_{12}S_{22}^{-1}S_{21})|, \hat{r}_1, \hat{r}_2, \hat{r}_3, \ldots \hat{r}_{p_1}$ son los estimadores de $\gamma_1, \gamma_2, \ldots, \gamma_{p_1}$

Si $\tilde{a}_i$ es una solución de:

$$(S_{12}S_{22}^{-1}S_{21} - \tilde{\gamma}_i(S_{11} - S_{12}S_{22}^{-1}S_2))\tilde{a} = 0$$

donde $\tilde{a}'(S_{11} - S_{12}S_{22}^{-1}S_{21})\tilde{a} = 1$

como $|(1+\gamma_i)\, S_{12}S_{22}^{-1}S_{21} - \gamma_i S_{11}| = 0$

$$|\Sigma_{12}\, \Sigma_{22}^{-1} \Sigma_{21} - l_i^2\, \Sigma_{11}| = 0$$

Así $l_1^2 = \dfrac{\gamma_i}{1+\gamma_i} \Longrightarrow (\tilde{\gamma}_i + 1) l_i^2 = \tilde{\gamma}_i \Longrightarrow \tilde{\gamma}_i\, (l_i^2 - 1) = -l_i^2$

$$\tilde{\gamma}_i = \frac{l_i^2}{1 - l_i^2}$$

y $\tilde{a}^{(i)} = \left[1/(1-l_i^2) \right]^{1/2} a^i$, entonces $(1/\sqrt{\gamma_i})\, \beta' \tilde{a}_i = \tilde{c}_i$ y es por lo tanto el mismo c_i.

CAPITULO III

LAS COMPONENTES PRINCIPALES Y LAS CORRELACIONES CANONICAS EN EL LENGUAJE ALGEBRAICO

3.1 INDIVIDUOS - CARACTERES - CODIFICACION.

3.1.1 INTRODUCCION.

Partimos con unos datos, arreglados en forma de p-filas y n-columnas, a la cual llamaremos la matriz de información X. Para lo cual tiene sentido visualizarla como las respuestas de n-individuos a un cuestionario de p-preguntas. Bajo esta consideración está asociado a cada individuo un vector columna de la matriz de información. Podemos además a cada pregunta hacerle corresponder un caracter o característica al cuál se responderá con una sola respuesta dentro de las modalidades permisibles de cada caracter, de esta manera, cada fila de X representa una característica.

Ejemplos 3.1.1.

1. La característica "permanencia" está asociada a la pregunta: Cuántos años tiene de vivir en su actual residencia?
2. La característica "Estado Civil está asociada a la pregunta" Es usted casado o soltero?

   Las característica "Familia" está ligada a la pregunta. Si los padres están vivos, viven juntos en el hogar?

   Las modalidades o conjuntos de valores posibles asignados en el estudio, a las características de interés de los ejemplos anteriores son respectivamente:

$$\{0,1,2,\ldots,150\}$$
$$\{\text{casado, soltero}\}$$
$$\{\text{si, no}\}$$

3.1.2 Terminología y Simbolismo.

Denotaremos al conjunto de los individuos y al conjunto de los caracteres con los símbolos I y J respectivamente. Donde; cardinal (I) = n, cardinal (J) = p ambos finitos.

El conjunto de las modalidades de un determinado caracter j; se designará como Q_j.

Hacemos corresponder a cada caracter j, la aplicación h_j; que a cada individuo le asigna el valor de la característica j en él, i.e.

$$\begin{aligned} h_j : I &\longrightarrow Q_j \\ i &\longrightarrow h_j(i) \in Q_j \end{aligned}$$

y además a cada individuo le asociamos un punto de $Q=Q_1 x Q_2 x \ldots x Q_p$, que será su imagen a través de la aplicación:

$$\begin{aligned} h : I &\longrightarrow Q_1 x Q_2 x \ldots x Q_p = Q \\ i \in I &\longrightarrow h(i) = (h_1(i)\ h_2(i)\ \ldots\ h_p(i)) \in Q \end{aligned}$$

donde h(i) es el conjunto de las características del individuo i.

Si identificamos al caracter con su única aplicación asociada, tenemos la siguiente definición:

Definición 3.1.1.

Característica Cuantitativa: es un caracter cuyo conjunto de modalidades (Q) es un subconjunto de $\mathbb{R}$. En caso contrario se le llamará cualitativa.

Ejemplo 3.1.2.

Características cuantitativas: la edad, el peso, la altura.

Ejemplos 3.1.3.

Características Cualitativas: Familia, televisión, radio. Algunas características cualitativas como radio tienen sus modalidades ordenadas por ejemplo: 'nunca $<$ a veces $<$ con frecuencia.

3.1.3 Codificación.

Definición 3.1.2.

Codificación de una característica: es una aplicación "a"; que a cada modalidad del caracter j le hace corresponder un número real. Codificación de la característica j:

Sea Q_j el conjunto de las modalidades de la característica J:

$$a : Q_j \longrightarrow \mathbb{R}$$

Observación: Consideraremos como modalidad de un caracter solo aquell s para la cual existe al menos un individuo que la posee. Es decir las aplicaciones h_j son suryectivas para todo caracter.

Recordemos que toda aplicación induce una partición sobre el dominio de la aplicación; así

$$I_k^j = \{ i \mid i \in I;\ h_j(i) = q_k^j \in Q_j \}$$

por lo tanto $I_k^j \neq \phi$ para todo k y todo j y además

$I_k^j \cap I_w^j = \phi$, para todo j; y $k \neq w$; porque cada individuo solo responde con una de las posibles modalidades.

así

$I = \cup \{ I_k \mid k = 1,\dots,q \}$ donde $q \leq n$, y card $(Q_j) = q$.

Ejemplo 3.1.4.

Si la característica j es la edad, tenemos que la partición inducida por ella sobre el conjunto de los n-individuos es su clasificación o agrupación por edad. Así $I_k = \{$ los individuos que tienen 23 años $\}$

Si la característica j es televisión y las modalidades son las frecuencias con que se ve la televisión; los I_k son los individuos que ven con igual frecuencia la televisión ya sea, "a veces", "con regularidad" o " todo el tiempo".

Definición 3.1.3: Característica z menos fina que y.

se dice que z es menos fina que y sí

$$y(i) = y(i') \Longrightarrow z(i) = z(i')$$

es decir si conozco quienes integran una clase de y; se que ellos mismos forman parte de una misma clase por el caracter z; es decir a través de y, puedo conocer las clases que induce z sobre I. Los cardinales de las clases inducida por z son mayores o iguales que los de y; podría darse que en algunas clases de z, están agrupados varias clases de y.

Definición 3.1.4.

Características Equivalentes:

Sean y y z dos características diremos que ellas son equivalentes si inducen la misma partición sobre el conjunto I, es decir:

$$y \equiv z \Longrightarrow [y(i) = y(i') \Longleftrightarrow z(i) = z(i')]$$

Proposición 3.1.1. $z \leqslant y \Longleftrightarrow$ existe $\underline{a}$, aplicación de Q en Q', donde Q' es el conjunto de modalidades de z tal que:

$$\begin{tikzcd} I \arrow[r, "y"] \arrow[dr, "z"'] & Q \arrow[d, "a"] \\ & Q' \end{tikzcd}$$

<u>Demostración</u>:

$\Longrightarrow$)

Como $y(i) = y(i') \Longrightarrow z(i) = z(i')$, podemos definir la aplicación $\underline{a}$ de Q en Q' de la siguiente manera:

$$\underline{a}(q_j) = z(i); \text{ donde } q_j = y(i),$$

Esta aplicación está bién definida pues: para todo q_j
$I_j = \{ i : i \in I \mid y(i) = q_j \} \neq \emptyset$ (ya que y e z son suryectivas), así existe z(i) que es único pues z es una aplicación.

$\Longleftarrow$)

Veamos que z es una característica menos fina que y.

Sea $y(i) = y(i') \Longrightarrow z(i) = \underline{a}(q_j) = z(i')$

<u>Corolario 3.1.2</u>: z e y son equivalentes si Q y Q' tienen el mismo cardinal, i.e. $\underline{a}$ es biyectiva.

<u>Acuerdo</u>: en lo sucesivo consideramos asociada a cada variable caracter cualitativa y, una variable real ξ , menos fina que y. Así la aplicación $\underline{a}$ codifica las modalidades de la variable y y además si codificamos las modalidades de una variable y, estamos definiendo un carácter real menos fino que y, i.e el caracter ξ .

Definición 3.1.5. Potencial de Previsión de la característica y. es el conjunto de las características menos fina que y.

Recordemos que las características menos finas son los previsibles a partir de y; en este potencial están todas las que se pueden definir a partir de y, como menos fina.

Definición 3.1.6:

Variable Indicatriz de una modalidad del caracter j:

es la aplicación

$$y^j : I \longrightarrow \{0,1\} \subset \mathbb{R}$$

definida así:

$$y^j(i) = \ \text{si}\ y(i) = q_j$$
$$y^j(i) = 0 \ \text{si no; i.e.}\ y(i) \neq q_j$$

Proposición 3.1.3: los siguientes enunciados son equivalentes

(a) $\xi \leqslant y$

(b) existe $\underline{a}$ tal que $\xi = \underline{a} \circ y$.

(c) $\xi = \sum_{j=1}^{q} a_j\, y^j$; donde y^j es la función indicatriz de la modalidad j asociada al caracter y.

Demostración: (a)$\Longleftrightarrow$(b) se hizo en la proposición 3.1.1.

haremos (b)$\Longleftrightarrow$(c).

$\Longrightarrow$

por proposición 3.1.1. tenemos que:

$$\xi(i) = a(q_j);\ \text{hagamos}\ a(q_j) = a_j$$

como además

$$\text{si}\ y(i) = q_j,\ \Rightarrow y^j(i) = 1\ \text{e}\ y^k(i) = 0\ \text{para todo}\ k \neq j$$

tenemos entonces que para todo $i \in I$.

$$\xi(i) = (\sum_{j=1}^{q} a_j y^j)(i) = a_j$$

$\Longleftarrow$

por hipótesis $\xi(i) = (\sum a_j \ y^j)(i) = a_j$ si $y(i) = q_j$ sea $a_j = a(q_j)$, la aplicación a está bién definida pues para cada; $a_j = \xi$ (i) es única.

Observación: la proposición anterior nos dice que si una variable real es menos fina que otra, ella es una combinación lineal de sus funciones indicatrices y viceversa; toda combinación lineal de las funciones indicatrices de un caracter es un caracter menos fino que él.

Definición 3.1.7:

Aplicación individuo:

$$k_i : J \longrightarrow Q_1 \cup Q_2 \cup \ldots \cup Q_p$$
$$j \longrightarrow k_i(j) = h_j(i)$$

K_i: es la aplicación que a cada caracter le hace corresponder la modalidad escogida por el individuo i.

Caso Particular:

Si $Q_j = A$ para todo j;

entonces

$$k_i : J \longrightarrow A$$
$$j \longrightarrow k_i(j) = h_j(i).$$

Como ejemplo de este caso; tenemos cuando a todas las preguntas se puede responder con si o no; siendo esto así podemos definir la siguiente aplicación:

$$k : J \longrightarrow A^n$$

$$j \longrightarrow (k_1(j),\ k_2(j),\ldots,k_n(j))$$

donde k asocia a cada caracter las modalidades elegidas por los n-individuos.

3.2 ESQUEMA DE DUALIDAD.

Notación:

X es la matriz pxn; es la tabla de datos, la tabla de caracteres x individuos.

x_i^j: es el valor de la j-ésima característica (fila) en el individuo i-ésimo (columna).

Consideraremos asociada a cada columna de X; un vector, i.e. la i-ésima columna de X, la representaremos por el vector

$$x_i = \begin{pmatrix} x_i^1 \\ x_i^2 \\ \vdots \\ x_i^p \end{pmatrix} = \sum_{k=1}^{p} x_i^k\ e_k \in E = \mathbb{R}^p \qquad (3.2.1)$$

al vector X_i le llamaremos el vector individuo; i-ésimo;y $(e_1,\ldots,e_p)$ designará la base canónica del Espacio $\mathbb{R}^p$, espacio de los individuos. Así n-vectores p-dimensionales constituyen la matriz X; n-puntos en el espacio p-dimensional.

En forma simétrica:

A cada fila de X; sucesión de n-números reales; le asociamos el vector:

$$x^j = q(j) = \begin{pmatrix} x_1^j \\ x_2^j \\ \vdots \\ x_n^j \end{pmatrix} = \sum_{k=1}^{n} x_k^j \, f_k \in F = \mathbb{R}^n \qquad (3.2.2)$$

Al vector x^j le llamaremos el <u>vector caracter</u> j-ésimo; punto del espacio vectorial F de los caracteres y $(f_1, f_2, \ldots, f_n)$ será la base canónica del espacio $\mathbb{R}^n = F$ en consideración. Así X es la matriz de p-vectores n-dimensionales; la matriz de información constituye p-puntos en el espacio vectorial n-dimensional de los caracteres.

<u>Definición 3.2.1</u>: <u>Nube de Individuos</u>:

Es el conjunto de las n-columnas de X, vista como puntos del p-espacio E, i.e. $y(I) \subset E$. Se designará con la letra

$\boxed{M} = y(I) = \left\{ X_i / \ i = 1,2,\ldots.,\ n \right\} \subset E = \mathbb{R}^p$.

<u>Definición 3.2.2</u>: <u>Nube de Caracteres</u>.

Es el conjunto de las p-filas de X; vistos como p-puntos del espacio n-dimensional i.e. $k(J)$. Se designará con la letra

$N = k(J) = \left\{ x^j / j = 1,2,\ldots,p \right\} \subset F = \mathbb{R}^n$.

3.2.1 <u>Representación de los carácteres en F y E^* y de los individuos en E y F^*.</u>

Sean E^* y F^* los espacios duales de los espacios; E y F respectivamente. Consideremos en E^* y F^* las bases $(e_1^*, \ldots, e_p^*)$ y $(f_1^*, \ldots, f_n^*)$ respectivamente.

De acuerdo con las expresiones (3.2.1) y (3.2.2) tenemos:

$x_i^j = \langle e_j^*, x_i \rangle$ es la coordenada del vector X_i con respecto a e_j, cuando $\left\{ e_j | j = 1,2,\ldots,p \right\}$ es el sistema (3.2.3) de coordenados de E

$x_i^j = \langle f_i^*, x^j \rangle$ es la coordenada del vector X^j con respecto (3.2.4)

al eje f_i, cuando $\{ f_i / i = 1,2,\ldots,n \}$ es el sistema

de coordenadas de F.

Por las expresiones (3.2.3) y (3.2.4) tenemos que:

e_j^* es una representación del caracter j en E^*

f_i^* es una representación del individuo i en F^*

En el espacio de los individuos E, la base $\{ e_i / i=1,2,\ldots,p \}$ representa (cada e_j), el individuo que solo posee la característica j-ésima. Así $e_j^* : E \longrightarrow \mathbb{R}$ le hace corresponder a cada individuo el valor de la característica j en él. Así es lógico de acuerdo al sentido de la palabra llamarlo representante de la característica j. La interpretación es similar para f_i^*.

3.2.2 Aplicaciones Asociadas a las matrices X y X':

Sea:

$$X : F^* \longrightarrow E$$
$$f_i^* \longrightarrow x_i = x(f_i^*)$$

la aplicación que a la representación del individuo i en F^* le hace corresponder la representación del individuo i en E. Es de inmediato por la definición de la aplicación X; que la matriz asociada es la tabla de datos X; de acuerdo con las bases que estamos considerando.

La aplicación traspuesta:

$$X' : E^* \longrightarrow F$$
$$e_j^* \longrightarrow X^j = X'(e_j^*)$$

es la aplicación que a la representación del caracter j en E^* le hace corresponder la representación del caracter j en el espacio F. La matriz asociada a esta transformación en las bases anteriormente especificadas es la traspuesta de la matriz X.

3.3 METRICAS EUCLIDEANAS PARA MEDIR LA PROXIMIDAD.

3.3.1 "Métrica en el Espacio E de los individuos"

Sea D una métrica Euclineana en E; i.e.

$x \in E$, $y \in E$; $d(x,y) = \sqrt{(x-y)'D(x-y)} = ||x-y||_D$

Pretendemos definir una métrica W sobre F^* tal que:

$$\forall x_i, x_i' \in \boxed{M}$$

$$d(x_i, x_{i'}) = ||x_i - x_{i'}||_D = ||f_i^* - f_{i'}^*||_W = d(f_{i'}^*, f_i^*)$$

donde $X(f_i^*) = x_i$ y $X(f_{i'}^*) = x_{i'}$

Caracterizemos a W.

$$D(x_i, x_i') = \langle D(x_i), x_{i'} \rangle = \langle (DoX)(f_i^*), X(f_{i'}^*) \rangle$$

$$= \langle (X'oD\ o\ X)(f_i^*), f_{i'}^* \rangle = \langle W(f_i^*), f_{i'}^* \rangle$$

por la definición de aplicación traspuesta que damos a continuación.

$$E \xrightarrow{A} F$$

$$E^* \xrightarrow[A']{} F^*$$

sea $y \in F^*$, $x \in E$, tenemos entonces que:

$$\langle y, A(x) \rangle = y(A(x)) = [A'(y)](x) = \langle A'(y), x \rangle$$

Para que X' o D o X sea un isomorfismo X debe ser inyectiva y X es inyectiva $\Longleftrightarrow \ker(X) = \{0\} \Longleftrightarrow$ rang X = dim. F^* = rang X' es decir si $p < n$. (son p vectores n-dimensionales) entonces X' o D o X no es un isomorfismo.

Propiedades de $W = X'\circ D \circ X$.

Como $D(x_i,x_{i'}) = D(x_{i'},x_i)$; D es simétrica y bilineal y

$W(f_i^*,f_{i'}^*) = D(x_i,x_{i'}) = D(x_{i'},x_i) = W(f_{i'}^*,f_i^*)$

W es simétrica y bilineal, recordando además que $(F^*)^* \cong F$.

Dado que:

$W(f_i^*,f_i^*) = \left[(D(X(f_i^*),X(f_i^*))= 0 \Longleftrightarrow X(f_i^*) = 0\right]$ (D es definida)

Pero $X(f_i^*) = 0 \Longleftrightarrow f_i^* = 0 \Longleftrightarrow X$ es inyectiva. Por lo tanto W es definida si X es inyectiva, así si X no es inyectiva, W no es un isomorfismo, y tampoco está definida.

Como X es inyectiva si rang $(X) = n$ entonces:

Si $n > p$ y $\dim(\ker X) \geqslant n-p \Longrightarrow W$ no es definida.

Además como $W(f_i^*,f_i^*) = D(X(f_i^*), X(f_i^*) \geqslant 0$ entonces W es positiva.

Resumiendo W es bilineal simétrica semi-definición positiva.

Asociada a esta forma bilineal simétrica semi-definida positiva W, está la forma cuadrática semi-definida positiva W, que verifica:

$W(f_i^*) = ||x_i||_M^2$; donde $X(f_i^*) = x_i$. (3.3.1)

3.3.2 Métrica Euclidiana en el espacio de los caracteres:

A partir de una métrica k en el espacio F de los caracteres con la cual se medirá la proximidad y la colinealidad, deseamos una métrica en E^* (espacio de las representaciones de los caracteres), que verifique lo siguiente:

Para todo $a \in E^*$; $||a||_L = ||X'(a)||$.

De manera similar a lo anteriormente desarrollado esto equivale a asociar a el isomorfismo (métrica en F)

$$K : F \longrightarrow F^*$$

la aplicación lineal:

$$L : E^* \longrightarrow E$$

donde:

$$L = x \circ k \circ x'$$

L es una forma cuadrática semi-definida positiva con las siguientes propiedades; si $X'(e_j^*) = X^j$ (3.3.2)

$L(e_j^*-e_{j'}^*) = L(e_j^*-e_{j'}^*\,,\ e_j^*-e_{j'}^*) = ||x^j-x^{j'}||_k^2$ para todo $(j,j') \in J\times J$ (3.3.3)

si X' es inyectiva entonces L es una métrica Euclideana sobre E^*;

tal que:

$$||e_j^*-e_{j'}^*||_L = ||x^j-x^{j'}||_k$$

Resumiendo tenemos:

$$\begin{tikzcd}
E \arrow[r, no head, "X"] \arrow[d, shift right, "D"'] & F^* \arrow[d, shift right, "W"'] \\
E^* \arrow[r, "X'"] \arrow[u, shift right, "L"'] & F \arrow[u, shift right, "K"']
\end{tikzcd}$$

diagrama 3.3.1

$$W = x' \circ D \circ X \text{ y } L = x \circ k \circ x'$$

Además tenemos:

$$x(f_i^*) = x_i \qquad i = 1,2,\ldots,n$$

$$x'(e_j^*) = x^j \qquad j = 1,2,\ldots,p$$

A los isomorfismos D y K están asociadas: las métricas D y K en los espacios E;F de individuos y caracteres respectivamente; y además las formas cuadráticas y formas bilineales semi-definidos positivos W y L; que verifican la siguiente propiedad:

$$L(e_j^* - e_{j'}^*) = ||x^j - x^{j'}||_k^2 \qquad (3.3.4)$$

$$W(f_i^* - f_{i'}^*) = ||x_i - x_{i'}||_D^2 \qquad (3.3.5)$$

3.4 <u>FORMA CUADRATICA DE INERCIA Y CENTRO DE GRAVEDAD</u>.

Hasta el momento tenemos que cada individuo i está representado por un punto o vector x_i de $E = \mathbb{R}^p$; en particular a la tabla de datos, está asociada la nube $\boxed{M}$ de los n-individuos; $\boxed{M} \subset E = \mathbb{R}^p$.

Consideraremos además asociados a cada x_i de la nube los pesos p_i; donde $p_i > 0$; $\sum_{i=1} p_i = 1$

<u>Definición 3.4.1</u>

Proximidad de la nube $\boxed{M}$ a un punto $b \in E$: es el valor;

$$I_b = \sum p_i ||x_i - b||^2 = \sum p_i d^2(x_i, b) = \sum p_i D(x_i - b,\ x_i - b) \qquad (3.4.1)$$

El valor I_b también es conocido como <u>el valor de la dispersión de la muestra alrededor del punto pivotal b</u>.

<u>Definición 3.4.2</u>: <u>Centro de gravedad de la nube</u>: es el vector:

$$g = \sum_{i=1}^{n} p_i\, x_i; \quad x_i \in E;\ p_i > 0 \quad \sum_{i=1}^{n} p_i = 1$$

cuya j-ésima coordenada es la media ponderada del j-ésimo caracter $g^j = \sum_{k=1}^{n} p_k x_k^j$; en la base canónica $\{ e_i | i_1 = 1, 2, \ldots, p \}$.

<u>Proposición 3.4.1</u>: El punto del espacio de individuo E, que está más próximo a la nube de puntos es su centro de gravedad.

<u>Demostración</u>:

$$x_i - a = (x_i - g) + (g - a)$$

$$||x_i - a||^2 = ||(x_i - g) + (g - a)||^2 = D((x_i - g) + (g - a), (x_i - g) + (g - a))$$

$$= D((x_i - g), (x_i - g)) + D((x_i - g), (g - a)) + D((g - a), (x_i - g)) + D((g - a), (g - a))$$

$$= ||x_i - g||^2 + ||g - a||^2 + 2D((x_i - g), (g - a)) \qquad (3.4.2)$$

Al considerar: $\sum 2pi\ D((x_i-g),(g-a))=D(\sum 2pi(x_i-g),g-a)$

$$=D(0,g-a) = 0$$

Por lo tanto $I_a = \sum p_i ||x_i-g||^2 + ||g-a||^2$

$\frac{\partial I_a}{\partial a} = 2D(g-a) = 0$; donde D es la matriz asociada a la forma (3.4.3) cuadrática.

Sabemos que ella es definida positiva y por lo tanto.

3.3.4 implica que

$$g-a = 0$$

así

$$g = a$$

Valor mínimo del momento de inercia.

$$I_g = \sum pi\ ||x_i-g||^2$$

Si consideramos el espacio E, de los individuos con su origen colocado en el centro de gravedad de la nube; tenemos que el valor de I_g es:

$$I_g = \sum pi\ ||x_i||^2 \qquad (3.4.4)$$

3.4.1 Métrica de los Pesos:

Dados los isomorfismos D y K entre E y E^*; espacios de individuos y espacio de las representaciones de los caracteres y entre F y F^* espacio de los caracteres y espacio de las representaciones de los individuos. Todo lo tratado para el espacio de los individuos E, es **válido** para el espacio de las características.

Consideraremos a los caracteres centrados, es decir $g^j = 0$ para j = 1,2,...,p; así el centro de gravedad de la nube $\boxed{M}$ es el origen del espacio vectorial de los caracteres F.

Dado que estamos considerando características con conjunto de modalidades finitas y considerando los pesos p_i como las probabilidades de que la variable j-ésima caracter asuma el valor x_k^j tenemos:

$$\mathrm{Var}\,(X^j) = \sum_{k=1}^{n} p_k {X_k^j}^2 = E(X^{j2})-(E(X^j))^2 = 0) \qquad (3.4.5)$$

$$\mathrm{cov}\,(X^j,X^{j'}) = \sum_{k=1}^{n} p_k X_k^j\, X_k^{j'} = E(X^j X^{j'}) \qquad (3.4.6)$$

A continuación supondremos que el espacio de los caracteres F, está provisto de la métrica de los pesos definida así:

$$D_p(f_i,f_{i'}) = 0, \quad i \neq i' \qquad (3.4.7)$$

$$D_p(f_i) = ||f_i||_p^2 = p_i \qquad , \text{ donde } p_i > 0 \qquad (3.4.8)$$

la matriz de la forma bilineal simétrica definida positiva es según la expresión 3.3.3, 3.4.7 y 3.4.8

$$\begin{pmatrix} D_p(f_1,f_1) & D_p(f_1,f_2) & \cdots & D_p(f_1,f_n) \\ D_p(f_2,f_1) & D_p(f_2,f_2) & \cdots & D_p(f_2,f_n) \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ D_p(f_n,f_1) & D_p(f_n,f_2) & \cdots & D_p(f_n,f_n) \end{pmatrix} = \begin{pmatrix} p_1 & 0 & \cdots & 0 \\ 0 & p_2 & \cdots & 0 \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ 0 & 0 & \cdots & p_n \end{pmatrix}$$

Recordando el producto interno definido por D_p:

$$\begin{aligned} D_p(X^j,X^{j'}) &= \left\langle \sum_k X_k^j f_k, \sum_i X_i^{j'} f_i \right\rangle = \sum_{i,k} X_k^j X_i^{j'} \langle f_k, f_i \rangle \\ &= \sum_k X_k^j X_k^{j'} p_k \qquad (3.4.9) \\ &= \mathrm{cov}\,(X^j, X^{j'}) \end{aligned}$$

por lo tanto la distancia Euclidea, asociada a D_p es:

$$d(X^j,X^{j'}) = \sqrt{<X^j-X^{j'},X^j-X^{j'}>} = \sqrt{(X^j-X^{j'})' \; D_p(X^j-X^{j'})}$$
$$= \sqrt{\sum p_k(X_k^j-X_k^{j'})^2} \qquad (3.4.10)$$
$$= \sqrt{Var(X^j-X^{j'})} = ||X_j-X_{j'}||$$

desarrollando (3.4.1), aplicando (3.4.5) y (3.4.6)

$$d^2(X^j,X^{j'}) = Var(X^j-X^{j'}) = Var(X^j)+Var(X^{j'})-2\,cov(X^j,X^{j'}). \quad (3.4.11)$$

Por diagrama 3.3.1, y las expresiones 3.3.1, 3.3.2, 3.3.3

$$L(e_j^*,e_{j'}^*) = D_p(X^j,X^{j'}) = cov\,(X^j,X^{j'})$$
$$V(e_j^*) = V(e_j^*,e_j^*) = D_p(X^j) = Var(X^j) = ||X^j||^2$$

i.e considerando en el espacio de los caracteres la métrica de los pesos, la varianza del caracter j; es su norma y por lo tanto la distancia del caracter j al orígen, que según lo considerado es el centro de gravedad o centro de la nube que coincide con el origen del espacio.

3.4.2 Dualidad:

Considerando las matrices X,X' y D_p asociadas a las respectivas aplicaciones del mismo símbolo según esquema de dualidad 3.3.1 tenemos que la matriz asociada a V es:

$$V = X\,D_p\,X'$$

sea $X = (X_1\; X_2\; \dots\; X_n)$; donde X_i es el vector columna del i-ésimo individuo.

$$XD_p = (X_1\ X_2\ X_3 \cdots X_n) \begin{pmatrix} p_1 & 0 & \cdots & 0 \\ 0 & p_2 & \cdots & 0 \\ \vdots & \vdots & & \\ 0 & 0 & \cdots & p_n \end{pmatrix} = (p_1X_1\ p_2X_2 \cdots p_nX_n)$$

y

$$V = XD_pX' = (p_1X_1\ p_2X_2 \cdots p_nX_n) \begin{pmatrix} X_1' \\ X_2' \\ \vdots \\ X_n' \end{pmatrix} = \sum_{i=1}^{n} p_1X_iX_i' = \hat{V}(x) \quad (3.4.12)$$

Así V es una estimación de la matriz de covarianza del vector X (individuo) obtenida a través de n-valores (X_i) de la variable vectorial individuo x.

3.4.3 Relación de Finura entre caracteres cuantitativos: (cualitativos). Equivalencia entre caracteres cuantitativos. (cualitativos).

En la definición (3.1.3) se definió la relación de finura entre caracteres cualitativos con modalidades no-ordenadas. Definimos ahora la misma relación pero entre los caracteres cualitativos con modalidades ordenadas.

Definición 3.4.3:

Sea x e y dos caracteres cualitativos con modalidades ordenadas x <u>es más fino que</u> y si se verifica la siguiente condición:

$$x_i \leqslant x_j \Longrightarrow y_i \leqslant y_j \quad \text{para toda pareja } (i,j)$$

o

$$x_i \leqslant x_j \Longrightarrow y_j \leqslant y_i \quad \text{para toda pareja } (i,j).$$

i.e. si el orden inducido por el caracter x es igual u opuesto al inducido por y .

Definición 3.4.4: caracteres equivalentes

$x \equiv y \Longleftrightarrow$ x es menos fino que y y y es menos fino que x.

Definición 3.4.5:

Sean x y y dos caracteres cuantitativos.

x es menos fino que y si se verifica la siguiente condición: existe una función continua f tal que y = fox.

Definición 3.4.6: caracteres cuantitativos Equivalentes:

sean x e y caracteres cuantitativos, x e y son equivalentes si existe una función estrictamente creciente o decreciente es decir una biyección tal que y = fox.

3.4.4 Método Práctico para determinar la equivalencia entre caracteres cuantitativos.

x e y son equivalentes si verifican una de las dos condiciones.

Sean x_i e y_i realizaciones de los vectores x e y

(a) $y_i = x_i + b, \quad i = 1,2,\ldots,n$

(b) $y_i = ax_i + b \Longleftrightarrow r(x,y) = \pm 1; \; i = 1,2,\ldots,n.$

donde r(x,y) es el coeficiente de correlación entre el caracter x e y. Si consideramos la representación de los caracteres x e y como puntos o vectores del espacio $\mathbb{R}^n$, la correlación estimada es el coseno del ángulo comprendido por los vectores x e y. Así $\cos\theta = \pm 1$ sii los vectores x e y tienen ya sea la misma dirección, magnitud y sentido o igual magnitud y dirección pero sentido opuesto es decir; la condición (b) se expresa $y = \pm X$: donde y e x son los vectores x e y centrados, es decir $y = y - \bar{y}\varepsilon$ e $x = x - \bar{x}\varepsilon$; donde $\varepsilon = (1\ 1 \ldots 1)$.

Para la condición (a) tenemos que $y_i = x_i+b$; y_i es el trasladado de x_i, tienen como vectores la misma magnitud, dirección y sentido que x_i, así al proyectarse sobre $\mathcal{E}$; cada uno de ellos, esta coincide, (en magnitud y dirección y sentido); así son iguales, al trasladarse los vectores $y_i-\bar{y}_i j$, $x_i-\bar{x}_i j$ también son iguales, es decir $x = y$.

En la práctica las condiciones (a) y (b), se pueden referir a los casos siguientes:

Si x e y son vectores de alguna característica la misma de la cuál se está midiendo el tamaño con algún aparato.

(1) x e y; pueden diferir por la elección del cero del aparato (condición a).

(2) x e y pueden diferir por la elección de la escala (metros, centímetros, pulgadas). (condición b).

(3) x e y por cambio en los dos aspectos anteriores.

Equivalencia entre Individuos.

Solo se considerará el caso en que nos interesa la forma es decir si dos individuos tienen la misma forma, o tienen características semejantes en el sentido de homotecia, o como la traslación no cambia la forma, tenemos:

x_i e x_{i1}' son equivalentes sii $x_i = ax_i+c$.

3.5 ANALISIS EN COMPONENTES PRINCIPALES.

Notaciones:

$(\boxed{M}, E, M)$

$\boxed{M}$: nube de puntos de E.

E : espacio Euclideano.

M : métrica Euclideana sobre E.

(I, d_I):

I: conjunto de individuos.

d_I: distancia sobre I.

Recordaremos, de las notaciones anteriores, las que serán utilizadas, en esta sección.

Los n elementos i de I se considerarán provistos de masas p_i, donde $p_i > 0$, $\sum_{i=1}^{n} p_i = 1$.

El centro de gravedad de la nube $\boxed{M} \subset E$; coincide con el origen del espacio.

La métrica considerada en el espacio F de los caracteres es la de los pesos, D_p.

Esquema de Dualidad.

$$\begin{tikzcd}
E=\mathbb{R}^p \arrow[d, shift right, "M"'] & F^* \arrow[l, "X"'] \arrow[d, shift left, "W"] \\
E^* \arrow[u, shift right, "V"'] \arrow[r, "X'"] & F=\mathbb{R}^n \arrow[u, shift left, "D_p"]
\end{tikzcd}$$

donde $x_i = \sum_{j=1}^{p} x_i^j e_j \in E$

$$x^j = \sum_{i=1}^{n} x_i^j f_i \in F.$$

Definición 3.5.1. Imagen o Representación Euclideana.

($\boxed{M}$, E, M) es una representación euclideana de (I,d_I) si existe una aplicación.

$$I \xrightarrow{h} E$$

tal que

$$h(I) = \boxed{M}$$

y

$$d_I(i,i') = d_E(h(i),h(i')) = ||h(i)-h(i')||_M \forall (i,i') \in I x I$$

Dado que, si existe una imagen euclideana existe una infinidad, establecemos la siguiente relación de equivalencia.

Definición 3.5.2. Imágenes Euclideanas Equivalentes:

($\boxed{M}$, E, M) y ($\boxed{M_1}$, E, M_1) son equivalentes si existe una biyección.

$$\boxed{M} \xrightarrow{f} \boxed{M_1}$$

tal que:

$$||f(x_i)-f(x_{i'})||_{M_1} = ||x_1-x_{i'}||_M \text{ para todo } (x_i,x_{i'}) \in \boxed{M} \times \boxed{M} .$$

De este conjunto de imágenes Euclideanas equivalentes, la más simple, es la que tiene el espacio Euclideano más pequeño, ya que cada individuo será entonces, representado por pocas coordenadas.

Así veremos más adelante, que el A.C.P. tiene como objetivo simplificar una imagen Euclideana, asociada a priori.

De manera similar a la definición 3.5.1, $(\boxed{N}, F, D_p)$ es una imagen o representación Euclideana de (J, d_J). Si existe una biyección entre J y F, tal que $d_j(J,J') = ||x^j - x^{j'}||D_p$.

Recordemos además que: a la aplicación lineal $V = X \text{ o } D_p \text{ o } X'$ está asociada una forma cuadrática semi-definida positiva V sobre E^*, la forma cuadrática de Inercia sobre E; cuya matriz es una estimación de la <u>matriz de covarianza</u> del vector individuo x, donde enfatizamos que x_i es una realización del vector individuo x.

En forma semejante, a $W = X' \text{ o } M \text{ o } X$ aplicación lineal sobre F^*, está asociada la forma cuadrática semi-definida positiva W.

3.5.1 <u>Procedimiento en el Análisis en Componentes Principales</u>.

Este procedimiento fue ideado por Pearson en 1901 y se llama "planos de ajustes cerrados".

Para cambiar una métrica M sobre E, a una Euclideana, necesitamos considerar una aplicación lineal T; que es biyectiva, llamada isometría; que conserva las distancias; tal que; $M = T' \text{ o } I_p \text{ o } T$. Representada en el siguiente diagrama:

$$\begin{tikzcd}
E \arrow[r, "T"] \arrow[d, "M"'] & E \arrow[d, "I_p"] \\
E^* & E^* \arrow[l, "T'"']
\end{tikzcd}$$

$M = T' \text{ o } I_p \text{o} T$

de donde resulta que las dos representaciones

($\boxed{M}$, E, M) y (T $\boxed{M}$, E, I_p) son equivalente; siendo T la biyección f; (para ello basta que n sea mayor que p con la siguiente propiedad):

$||x_i-x_j||_M = ||z_i-z_j||_{I_p}$ donde $Z_i = Tx_i$

El cambiar la métrica a la Euclideana tiene la ventaja que las proximidades de los individuos va estar dada por la longitud de los vectores que los unen (los x_i). Pero además de esto buscamos una representación lo más simple posible (la de menor dimensión). Así determinaremos

(a) El punto de E más próximo a la nube.

(b) La recta afín más próxima a la nube.

(c) El plano afín más próximo a la nube.

y así sucesivamente. Generalizando lo que necesitamos es un índice, para medir la proximidad de la nube a un subespacio afín de E.

3.5.2 <u>El Subespacio Afín más próximo a la nube</u> $\boxed{M}$

El <u>punto</u> más próximo a la nube:

En la proposición 3.4.1 se demostró que este punto es el centro de gravedad de la nube. Como I_g mide la despersión alrededor de g; por (3.4.4) la nube $\boxed{M}$ está <u>concentrada</u> en el punto g sí

$$I_g = \sum p_i||x_i||^2 = 0$$

La expresión (3.4.4) la podemos también expresar de la siguiente forma:

dado que $||x_i||^2 = \langle x_i, x_i \rangle = M(x_i, x_i) = x_i' \; M \; x_i$; donde M es la forma bilineal simétrica definida positiva y a la vez la matriz asociada en la base canónica.

como $x_i' \; M \; x_i$ es un número:

$$\text{traza } (x_i' \; M \; x_i) = \text{traza } (x_i \; x_i' M) \tag{3.5.1}$$

así

$$I_g = \sum p_i ||x_i||^2 = \sum p_i \text{ traza } x_i \; x_i' M \tag{3.5.2}$$

$$= \text{traza } ((\sum p_i \; x_i \; x_i')M) \tag{3.5.3}$$

y por la expresión (3.4.7) resulta:

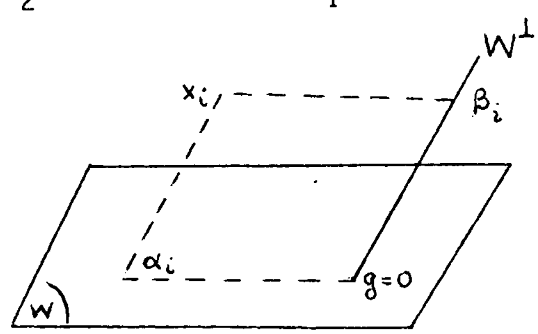
$$I_g = \text{traza (VM)}. \tag{3.5.4}$$

de donde podemos observar que $\underline{g}$ e I_g son las correspondientes definiciones de media y varianza pero para el caso multidimen sional; g es un vector, I_g es un número.

<u>Momento de Inercia con respecto a un subespacio afín.</u>

Sea $E = W \oplus W^{\perp}$; donde W es un subespacio de E. y $W^{\perp}$ es el suplementario M-ortogonal de W; es decir:

$M = (x_1, x_2) = 0$, para todo x_i W y todo x_2 W (3.5.5)



(3.5.1)

así $x_i = \alpha_i + \beta_i$ donde $\alpha_i \in W$ y $\beta_i \in W^{\perp}$ $i = 1,2,\ldots, n$

es decir

β_i es la proyección M-ortogonal de x_i sobre W según la dirección de $W^{\perp}$; así

$||x_i - \alpha_i|| = ||\beta_i||$ mide la distancia o proximidad del punto x_i al subespacio W.

Definición 3.5.3: Proximidad de la nube a un subespacio vectorial W es el valor numérico.

$$I_W = \sum_{i=1}^{n} p_i||\beta_i||^2; \text{ donde } \beta_i \in W^{\perp}$$

conocido como Momento de Inercia de $\boxed{M}$ con respecto a un subespacio Vectorial W de E.

De manera similar

$I_{W^{\perp}} = \sum_{i=1}^{n} p_i \; ||\alpha_i \;||^2$ mide el momento (la aproximidad) de inercia de la nube $\boxed{M}$ con respecto a W

Proposición 3.5.1: Sean A y B = (I-A) los proyectores asociados a la descomposición en suma directa de $E = W \oplus W^{\perp}$, entonces ellos son idempotentes y M-simétricas.

Demostración:

Sean $x, y \in E$; M-forma bilineal simétrica definida positiva, además $A(E) = W$, $B(E) = W^{\perp}$; así $A + B = I_n$, y $W^{\perp}$ es el suplementario M-ortogonal.

$$M(Ax,y) = M(Ax,\ In\ y) = M(Ax,\ (A + B)(y)).$$

$$= M(Ax,\ Ay) + M(Ax, B(y))$$

Como $Ax \in W$ y $B(y) \in W^{\perp}$ y ellos son M-ortogonales $M(Ax,By) = 0$.
así

$$M(Ax,y) = M(Ax,Ay)$$

Por la bilinealidad de M resulta también

$$M(x,Ay) = M(Ax,Ay) \Longrightarrow M(Ax,y) = M(x,Ay).$$

Veamos ahora que $M(Ax,y) = M(x,Ay)$ implica

$$M \circ A = (M \circ A)' = A' \circ M$$

Si consideramos a M como el isomorfismo asociado de E en E*.
tenemos que:

$$M(Ax,y) = \langle M(Ax),\ y \rangle = \langle x,\ M(Ay) \rangle$$

por el siguiente diagrama y dado que M es un isomorfismo.

$$\begin{tikzcd}
E \arrow[r, "A"] \arrow[d, "M"'] & E \arrow[d, "M"] \\
E^{*} & E^{*} \arrow[l, "A'"']
\end{tikzcd}$$

tenemos que $\langle (M \circ A)(x),y \rangle = \langle (A' \circ M)(x),y \rangle \quad \forall x,y$
así $M \circ A = A' \circ M = (M \circ A)'$ por lo tanto A es M-simétrica.

<u>Proposición 3.5.2:</u>

La nube $\boxed{M}$ es un subconjunto de W, está concentrada en W sii su momento de Inercia con respecto a W es cero, i.e. sii el valor de su proximidad a W es cero.

Demostración:

Si $I_w = \sum p_i ||\beta_i||^2 = 0. \Longrightarrow$ como $p_i > 0$, y $||\beta_i||^2 \geqslant 0$

entonces $||\beta_i||^2 = 0 \Longrightarrow \beta_i = x_i - \alpha_i = 0 \Longrightarrow x_i = \alpha_i \in W$

Si $\boxed{M} \subset W \Longrightarrow \beta_i = x_i - \alpha_i = 0$ para $i = 1,2,\dots,n$

$$||\beta_i|| = 0 \Longrightarrow I_w = 0$$

Proposición 3.5.3: El momento de inercia con respecto al origen del espacio E es suma de los momentos de inercia con respecto a W y $W^{\perp}$

Demostración:

Por teorema de Pitágoras.

$$I_g = \sum p_i ||x_i|| = \sum p_i ||\alpha_i + \beta_i||^2 = \sum p_i (||\alpha_i||^2 + ||\beta_i||^2)$$

$$= \sum p_i ||\alpha_i||^2 + \sum p_i ||\beta_i||^2$$

$$I_g = I_{W^{\perp}} + I_W$$

Definición 3.5.4: Aplicación M-simétrica.

Una aplicación A es M-simétrica si

$M(Ax,y) = M(x,Ay)$ para todo par (x,y)

i.e. $A' \circ M = (M \circ A)' = M \circ A$ i.e. A es M-simétrica si $M \circ A$ es simétrica.

Proposición 3.5.4. Sean A y $B = (I-A)$ los proyectores asociados a la descomposición de $E = W \oplus W^{\perp}$, entonces

$$I_W = \text{traza } (VM(I-A)) \text{ y } I_{W^{\perp}} = \text{traza } (VMA)$$

Demostración:

Dado que:

$$I_W = \sum p_i ||\beta_i||^2 = \sum p_i ||(I-A)x_i||^2$$

$$= \sum p_i ((I-A)x_i)' \; M(I-A)x_i$$

siguiendo el mismo procedimiento que en (3.5.1) a (3.5.3) tenemos que:

$$= \text{traza } ((\sum p_i (I-A)x_i x_i' \; (I-A)')M)$$

$$= \text{traza } ((I-A)V(I-A)'M)$$

$= \text{traza } (VM(I-A)(I-A))$ por proposición 3.5.3 y por propiedad de las trazas tenemos:

$$I_W = \text{traza } (VM(I-A))$$

$$I_{W^\perp} = \sum p_i \; ||\alpha_i||^2 = \sum p_i \; ||Ax_i||^2$$

$$= \sum p_i (Ax_i)'M(Ax_i)$$

$= \text{traza } AVA'M$ por propiedad de la traza

$$= \text{traza } VMAA$$

$$= \text{traza } VMA$$

Sea W_1 un espacio afín de dirección W; es decir

$W_1 = W + a$; donde $a \in W^\perp$

así todo $x \in E$, se puede expresar como suma de W_1 y $W^\perp$ donde

$W^\perp \cap W_1 = \{ a \}$

$x_i = \alpha_i + \beta_i = (\alpha_i + a) + (\beta_i - a)$

$(\alpha_i + a) \in W_1$ y $(\beta_i - a) \in W^\perp$

Definición 3.5.4:

Momento de Inercia con respecto a un subespacio afín:

$$I_{W_1} = \sum p_i \; ||\beta_i - a||^2; \qquad (\beta_i - a) \in W^{\perp}$$

Este número mide la proximidad de la nube al subespacio afín W_1.

Proposición 3.5.5:

$$I_{W_1} = I_W + ||a||^2$$

Demostración:

$$I_{W_1} = \sum_{i=1}^{n} p_i ||\beta_i||^2 = \sum_{i=1}^{n} p_i ||a||^2 + \sum_{i=1}^{n} p_i ||\beta_i||^2 - 2 \sum_{i=1}^{n} p_i M(\beta_i, a)$$

(por procedimiento similar al de la expresión 3.5.3).

como

$$\sum p_i \; ||a||^2 = ||a||^2 \text{ pues } \sum p_i = 1 \qquad (3.5.5)$$

Además, dado que

$$g = \sum p_i x_i = 0 \Longrightarrow \sum p_i \alpha_i + \sum p_i \beta_i = 0$$

$$\Longrightarrow \sum p_i \alpha_i = -\left(\sum p_i \beta_i \right)$$

dado que $\sum p_i \alpha_i \in W$ y $p_i \beta_i \in W^{\perp}$ y $W \cap W^{\perp} = \{0\}$

entonces

$$\sum p_i \alpha_i = \sum p_i \beta_i = 0$$

así

$$\sum p_i \; M(\beta_i, a) = M\left(\sum p_i \beta_i, a\right) = M(0,a) = 0 \qquad (3.5.6)$$

por (3.5.5 y (3.5.6) se da que:

$$I_{W_1} = I_W + ||a||^2 \qquad (3.5.7)$$

Este teorema nos dice que: el momento de Inercia de la nube $\boxed{M}$ con respecto a un subespacio afín W_1 es el momento de inercia con respecto a su dirección (W) más el cuadrado de la distancia recorrida en la traslación.

Corolario: El subespacio afín de dimensión k, más próximo a la nube $\boxed{M}$ es el que pasa por su centro de gravedad.

Demostración:

Por (3.5.5) tenemos que:

$$I_{W_1} \geqslant I_W$$

donde W subespacio de E; $g = 0 \in W$.

Por proposición 3.5.3 y dado que si $W = W_1 \oplus W_2$ donde $W_1 \underset{M}{\perp} W_2$; se da que $I_{W^\perp} = I_{W_1^\perp} + I_{W_2^\perp}$ (3.5.8)

por lo tanto $I_W = I_g - I_{W_1^\perp} - I_{W_2^\perp}$ (3.5.9)

Proposición 3.5.6:

Todo subespacio W_{k+1}, de dimensión k+1; para el cual $I_{W_{k+1}}$ es mínimo; contiene un subespacio W_k; de dimensión K para el cual I_{W_k} es mínimo.

Demostración:

Dado que:

$$\left.\begin{array}{c} \dim (W_k) = n-k \\ y \\ \dim (W_{k+1}) = k+1 \end{array}\right\} \Longrightarrow \dim (W_{k+1}) + \dim (W_k^\perp) = n+1$$

de donde dim $(W_{K+1} \cap W_k^{\perp}) \geqslant 1$

por lo tanto existe un $v \neq 0$ y que $v \in (W_{K+1} \cap W_k^{\perp})$ de esta manera podemos expresar

$W_{k+1} = \Delta_v \oplus \mathbb{R}$; donde Δ_v es la recta generada por v y $\mathbb{R}$ es un s.e.v.k-dimensional suplementario D-ortogonal a Δ_v en W_{k+1}.

Consideremos además otro espacio U de dimensión $K+1$ tal que:

$$U = \Delta_v + W_k$$

por la expresión (3.5.8) y (3.5.9) tenemos que:

$$\left.\begin{aligned} I_{w_{k+1}^{\perp}} &= I_{\Delta_v^{\perp}} + I_{\mathbb{R}^{\perp}} \\ &y \\ I_{U^{\perp}} &= I_{\Delta_v^{\perp}} + I_{w_k^{\perp}} \end{aligned}\right\} \Longleftrightarrow \begin{aligned} I_{w_{k+1}} &= I_g - I_{\Delta_v^{\perp}} - I_{\mathbb{R}^{\perp}} \\ \\ I_U &= I_g - I_{\Delta_v^{\perp}} - I_{w_k^{\perp}} \end{aligned}$$

Como $I_g = I_{w_k} + I_{w_k^{\perp}}$ entonces I_{w_k} es mínimo $\Longleftrightarrow$ $I_{w_k^{\perp}}$ es máximo, entonces $I_{w_k^{\perp}} \geqslant I_{\mathbb{R}^{\perp}}$ y así $I_{w_{k+1}} \geqslant I_U$.

y

$$I_{w_{k+1}} = I_U \Longleftrightarrow I_{w_k} = I_{\mathbb{R}}$$

De esta manera tenemos: $U = \Delta_v + W_k$ y $I_U = I_{w_{k+1}}$

ellos nos dicen que a partir de todo subespacio W_k de dimensión k, con momento de inercia mínimo, se puede construir adjuntándole una recta Δ_v; M-ortogonal, un subespacio U, de dimensión $(k+1)$ con momento de inercia mínimo.

3.5.3 Momento de Inercia con respecto a una recta Δ_u y con respecto al hiperplano M-ortogonal $\Delta_u^{\perp}$

Sea $u \in E$, $||u|| = \sqrt{M(u)} = 1$

Δ_u la recta engendrada por u

$E = \Delta_u + \Delta_u^{\perp}$; $x_i = \alpha_i + \beta_i$, $\alpha_i \in \Delta_u$ y $\beta_i \in \Delta_u^{\perp}$

Así

$$I_{\Delta_u} = \sum p_i \; || \beta_i ||^2 \quad y \qquad I_{\Delta_u^{\perp}} = \sum p_i \; ||\alpha_i||^2$$

son los momentos de inercia de la nube $\boxed{M}$ con respecto a la recta vectorial Δ_u y su hiperplano suplementario M-ortogonal $\Delta_u^{\perp}$ y por (3.5.3) tenemos que:

$$I_g = \text{traza } (VM) = I_{\Delta_u} + I_{\Delta_u^{\perp}}$$

además como $\alpha_i \in \Delta_u$, existe $c_i \in \mathbb{R}$ tal que $\alpha_i = c_i u$

$$D(x_i, u) = D(\alpha_i + \beta_i, u) = D(\alpha_i, u) + D(\beta_i, u)$$

$$D(x_i, u) = c_i \qquad (3.5.10)$$

(Dado que $D(\beta_i, u) = 0$ y $D(\alpha_i, u) = D(c_i u, u) = c_i D(u,u)$

$= c_i$

sea $L = Du \in E^*$ la forma lineal que a todo $x_i \in E$, asocia la coordenada de su proyección M-ortogonal sobre Δ_u según la dirección $\Delta_u^{\perp}$.

Sea

$$c = X'L = X' o Du \in F$$

que en el diagrama

$$\begin{tikzcd}[row sep=large, column sep=huge]
E \arrow[r, dash, "X"] \arrow[d, shift right, "D"'] & F^{*} \arrow[d, shift left, "W"] \\
E^{*} \arrow[r] \arrow[u, shift right, "L"'] & F \arrow[u, shift left, "D_p"]
\end{tikzcd}$$

Así tenemos caracterizada por la recta Δ_u a

L: la forma lineal $L = D_u$; $L(x_i) = L(\alpha_i + \beta_i) = L(c_i u + \beta_i)$

$$= c_i$$

y el vector $c = x'L = x'Du \in F$.

y así podemos expresar:

$$I_{\Delta_u^{\perp}} = \sum p_i ||\alpha_i||^2 = \sum p_i ||c_i u||^2 = \sum p_i \; c_i^2 = ||c||^2_{D_p}$$

Por (3.5.8) y (3.5.9) tenemos que:

$||c||^2_{D_p} = L(L) = D_p(c)$; donde L y D_p son las formas cuadráticas asociadas a las bilineales del mismo símbolo.

Observamos que:

$$L(L) = \langle L(L), L \rangle = \langle L \circ D_u, D_u \rangle = \langle D \circ L \circ D_u, u \rangle \; ; \; L \in E^*$$

así definimos la aplicación:

$E \xrightarrow{D \circ L \circ D} E^*$ a la cuál se le puede asociar la forma cuadrática definida sobre E^*.

$$L(L) = \langle D \circ L \circ Du, u \rangle = D \circ L \circ D(u) \qquad (3.5.11)$$

Proposición 3.5.6:

$$\boxed{M} \subset \Delta_u^{\perp} \Longleftrightarrow D(u,\ L \circ D_u) = 0 \Longleftrightarrow u \in \mathrm{Ker}(L \circ D)$$

Demostración:

por proposición 3.5.1 $\boxed{M} \subset \Delta_u^{\perp} \Longleftrightarrow I_{\Delta_u^{\perp}} = 0$

$$||c||^2_{D_p} = D_p(c) = L(L) = D(u, L \circ Du) = 0$$

$\Longleftrightarrow u \in \mathrm{Ker}\,(L \circ D)$

Proposición 3.5.7:

$$\boxed{M} \not\subset \Delta_u^{\perp} \Longrightarrow \begin{cases} I_{\Delta_{VDu}} \leqslant I_{\Delta_u} \\ I_{\Delta_{VDu}} = I_{\Delta_u} \Longleftrightarrow VDu = \lambda u \text{ con } \lambda \neq 0. \end{cases}$$

Demostración:

Si $||u|| \neq 1$

$$I_{\Delta_u^{\perp}} = \sum p_i\, ||\alpha_i||^2 = \sum p_i D(\alpha_i, \alpha_i) = \sum p_i\, D\left(\frac{c_i}{||u||} u, \frac{c_i}{||u||} u\right)$$

Por la expresión (3.5.7).

$$I_{\Delta_u^{\perp}} = \frac{\sum p_i c_i^2}{D^2(u,u)} \cdot D(u,u) = \frac{\sum p_i c_i^2}{D(u,u)} = \frac{D(u, L \circ Du)}{D(u,u)} = \frac{D \circ L \circ D(u,u)}{D(u,u)}$$

resumiendo, tenemos que:

si $D(u,u) \neq 1$ entonces $I_{\Delta_u} = \dfrac{D(u,\ L\ Du)}{D(u,u)}$ (3.5.12)

además, dado que: $L \circ Du \in E$

$$I_{\Delta^{\perp}_{L\circ Du}} = \frac{D(L \circ Du\,,\, LD\,LDu)}{D(LDu\,,\, LDu)} = \frac{DLD(LDu\,,\, LDu)}{D(LDu\,,\, LDu)}$$

Por la desigualdad de Schwarz.

$$(D(u,LDu))^2 \leqslant D(u,u).\,D(LDu,\, LDu)$$

$$(DLD(u,\, LDu))^2 \leqslant DLD(u,u).\,DLD(LDu,LDu)$$

σ: $D(u,LDu) = DLD(u,u) \neq o \neq D(LDu,\, LDu) = DLD(u,\, LDu)$

de donde:

$$\frac{D(u,LDu)}{D(u,u)} \leqslant \frac{DLD(u,\, LDu)}{DLD(u,u)} \leqslant \frac{DLD(LDu,\, LDu)}{D(LDu\,,\, LDu)}$$

y:

$$I_{\Delta^{\perp}_{LDu}} \geqslant I_{\Delta^{\perp}_{u}} \Longleftrightarrow I_{\Delta_{LDu}} \leqslant I_{\Delta_u}$$

Además: $I_{\Delta_{LDu}} = I_{\Delta_u} \Longrightarrow (D(u\,,\, LDu\,))^2 = D(u,u).D(LDu,LDu)$

Como la forma D es definida,entonces, $u \not\in$ Ker (VD)

se tiene así

$$I_{\Delta_{LDu}} = I_{\Delta_u} \Longrightarrow LDu = \lambda u \quad \text{con } \lambda \neq 0$$

por la proposición (3.5.6) y por (3.5.7) tenemos que

$$LDu = \lambda u \Longrightarrow I_{\Delta^{\perp}_{u}} = \lambda \qquad (3.5.13)$$

demostrandose que I_{Δ_u} es mínimo sí u pertenece al subespacio propio asociado al mayor de los valores propios de VD.

3.5.4 Estudio de los valores y vectores propios de la matriz VD.

Recordemos que:

Al isomorfismo $D: E \longrightarrow E^*$ están asociados:

(a) la forma bilineal simétrica definida positiva D.

(b) la forma cuadrática simétrica definida positiva D.

y que a la aplicación lineal $V = X \circ Dp \circ X' : E^* \longrightarrow E$.

están asociados:

(a) la forma bilineal simétrica semi-definida positiva V.

(b) la forma cuadrática simétrica semi-definida positiva V.

Así $V \circ D : E \xrightarrow{D} E^* \xrightarrow{V} E$;

Por lo tanto:

$V \circ D$ es una aplicación lineal de E en E con las siguientes propiedades.

1) $V \circ D$ es D-simétrica, i.e. $(V \circ D)' \circ D = D \circ V \circ D = (D \circ V \circ D)'$

2) Si V es regular o inversible entonces $V \circ D$ es V^{-1}-simétrica i.e. $(V^{-1} \circ (V \circ D))' = V^{-1} \circ V \circ D$.

Además, tenemos lo siguiente

(a) Los valores propios de $V \circ D$ son todos reales y se denotaran: $\lambda_1 \geqslant \lambda_2 \geqslant \dots \geqslant \lambda_p$.

(b) $E = E_1 \oplus E_2 \oplus \dots \oplus E_k = \oplus E_k = \oplus \{ E_i / i=1,\dots,k \}$

donde $\{ E_i / i = 1,2,\dots,k \}$ es el conjunto de los subespacios propios asociados a los k-valores propios diferentes.

Si $i \neq j \Longrightarrow E_i \underset{M}{\perp} E_j$

(c) $\sum_{i=1}^{k} \dim (E_i) = p$, $E_i \cap E_j = \{0\}$ si $i \neq j$

(d) Si V es invertible, $E_i \underset{V^{-1}}{\perp} E_j$ si $i \neq j$

(e) Los valores propios de V o D son todos positivos o nulos.

($VD\,u = \lambda u \Longrightarrow \lambda = I_{\Delta_u^{\perp}} \geqslant 0$)

(f) Se puede construir una base $\{u_i/i=1,2,\ldots,p\}$ D-ortonormado de E, con la ayuda de vectores propios de VoD.

3.5.5 Ejes y Planos Principales.

Definición 3.5.6.

Primer Eje Principal.

Es el eje Δ_u , generado por u , vector normado para el cual $I_{\Delta_u^{\perp}}$ es máximo.

Proposición 3.5.8.

El primer eje principal es el eje Δ_{u_1} generado por el vector u_1 , vector propio normado asociado al mayor de los valores propios λ_1, de V o D.

Demostración:

Fijada la base $\{u_i/i = 1,2,..,p\}$ D-ortonormada de los vectores propios de VoD, todo vector u se escribe de una única forma:

$$u = \sum_{i=1}^{p} \alpha_i u_i$$

donde:

$$||u||^2 = D(u,u) = D(\sum\alpha_i u_i, \sum\alpha_i u_i) \qquad (3.5.14)$$

y además por (3. 4.1)

$$I_{\Delta_V^{\perp}} = D(L \circ Du, u) = D(L \circ D(\sum \alpha_i u_i), \sum \alpha_i u_i) \qquad (3.5.15)$$

De (3.5.8) tenemos; por la bilinealidad de D y dado que la base es D-ortonormada lo siguiente

$$||u||^2 = \sum \alpha_i^2 = 1.$$

De la expresión (3.5.15) obtenemos lo siguiente:

$$D(L \circ D(\sum \alpha_i u_i), \sum \alpha_i u_i) =$$

$$D(\sum \alpha_i (L \circ D)(u_i), \sum \alpha_i u_i) =$$

$D(\sum \alpha_i \lambda_i u_i, \sum \alpha_i u_i)$ por ser los u_i vectores propios de L o D.

$\sum \alpha_i^2 \lambda_i = I_{\Delta_u^{\perp}}$ por la bilinealidad de D y porque la base es D-ortonormada.

donde $I_{\Delta_u^{\perp}} = \sum_{i=1}^{p} \alpha_i^2 \lambda_i$

con la restricción de que $\sum_{i=1}^{p} \alpha_i^2 = 1$

que de ser máximo; i.e. u_1 es el vector asociado a λ_1 el mayor de los valores propios y tal que $||u||^2 = 1$:

$$D(L \circ D(u), u) = D(\lambda_1 u_1, u) = \alpha_1 \lambda_1 \leqslant \lambda_1$$

debemos tener que $\alpha_1 = 1$, $\alpha_2 = \alpha_3 = \ldots = \alpha_p = 0$ por (3.5.13); así vemos que $u = \sum \alpha_i u_i = u_1$, i.e. u es uno de los vectores propios de λ_1.

Así el primer eje principal, es el generador por u_1; vector propio normado de V o D, asociado al mayor de los valores propios λ_1.

Esquemáticamente:

$$(\mathrm{Lo}\ D)(M_1) = \lambda_1 M_1 \text{ entonces } I_{\Delta_{u_1}^{\perp}} = \lambda_1$$

Parte de la Inercia explicada por Δ_{u_1}

$$I_{\Delta_{u_1}} = I_g - I_{\Delta_{u_1}^{\perp}} = t_r(VD) - \lambda_1$$

sabemos además que:

$$\boxed{M} \subset \Delta_{u_1} \Longleftrightarrow I_{\Delta_{u_1}} = 0 \Longleftrightarrow \lambda_1 = \text{traza } (VD) \Longleftrightarrow \frac{\lambda_1}{\text{traza } VD} = 1$$

$$M \not\subset \Delta_{u_1} \Longrightarrow I_{\Delta_{u_1}} > 0 \Leftrightarrow \text{traza } (VD) > \lambda_1 \Longleftrightarrow \frac{\lambda_1}{\text{traza } VD} < 1$$

<u>Definición 3.5.7.</u>

parte de la inercia explicada por el primer eje principal Δ_{u_1} es el valor

$$\frac{\lambda_1}{\text{traza } (VD)}$$

<u>Observaciones</u>.

Al eje principal Δ_{u_1} están asociados:

<u>en E^*</u> (espacio de las representaciones de los caracteres)

El primer factor principal (o forma lineal principal)

$$Du_1 = V_1 \in E^*$$

<u>en F</u>: (espacio de los caracteres)

La primera componente principal:

$$c_1 = X'V_1 = X' \circ Du_1$$

$c_1^i = D(u_1, x_i) = V_1(X_i)$ es el valor de la primera componente principal en el individuo i.

Proposición 3.5.9:

La primera componente principal centrada tiene varianza igual a λ_1.

Componente principal centrada:

$$\sum p_i \, c_1^i = D_p(j, c_1) = 0; \text{ donde } J = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in F$$

$$V(c_1 - E(c_1)) = \sum^{n} p_i (c_1^i)^2 = D_p(c_1) = V(v_1) = D \circ V \circ D \circ (u_1) = \lambda_1$$

3.5.6 Plano Principal.

Definición 3.5.8.

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

Por la proposición (3.5.5) tenemos que el plano principal P puede formarse de la siguiente manera:

(a) Con el eje principal Δ_{u_1}

(b) y el eje Δ_u, engendrado por el vector normado u, D-ortogonal a u_1, con momento de inercia mínimo.

Así para determinar P, lo que debemos buscar es el vector u, D-ortogonal a u_1 y tal que I_{Δ_u} sea mínimo.

Como I_{Δ_u} es mínimo $\Longleftrightarrow$ $I_{\Delta_u^\perp}$ es máximo, resulta que: $I_{\Delta_u^\perp} = D \circ V \circ D(u)$ máximo

con las restricciones $\begin{cases} \|u\|^2 = D(u) = 1 \\ D(u_1, u) = 0 \end{cases}$

Dado esto, es conveniente considerar en E, como base, la de los vectores propios D-ortonormales:

Sea

$$u = \sum_{i=1}^{p} \alpha_i u_i \Longrightarrow \begin{cases} I_{\Delta_u^{\perp}} = \sum^{p} \alpha_i^2 \lambda_i \\ D(u) = \sum_{i=1}^{p} \alpha_i^2 = 1 \\ D(u, u_1) = \alpha_1 = 0 \end{cases}$$

Por razones similares a las dadas en (3.5.14),(3.5.15), se da lo anteriormente expuesto.

Por lo tanto:

$$I_{\Delta_u^{\perp}} \leqslant \lambda_2 \qquad \text{y } I_{\Delta_u^{\perp}} \text{ es máximo e igual a } \lambda_2$$

si $\alpha_2 = 1$ y $\alpha_1 = \alpha_3 = \ldots = \alpha_p = 0$

así u es uno de los vectores D-ortonormales de V o D asociados a λ_2.

Así el segundo eje principal es el eje Δ_{u_2} engendrado por u_2, vector propio normado de V o D asociado al valor propio λ_2.

Resumiendo tenemos:

$$I_{\Delta_u^{\perp}} = \lambda_2, \quad (V \circ D)\, u_2 = \lambda_2 u_2$$

$$I_{p^{\perp}} = I_{\Delta_{u_1}^{\perp}} + I_{\Delta_{u_2}^{\perp}} = \lambda_1 + \lambda_2$$

$$I_p = \text{traza}(VD) - (\lambda_1 + \lambda_2)$$

Así

$$\boxed{M} \subset P \Longleftrightarrow I_p = 0 \Longleftrightarrow \frac{\lambda_1 + \lambda_2}{\text{traza}(VD)}$$ que es la parte de la Inercia explicada por el plano principal P.

Tenemos además que al segundo eje principal están asociados:

en $\underline{E^*}$ el segundo factor principal: $V_2 = Du_2$

en $\underline{F}$ la segunda componente principal: $c^2 = X'V_2$

3.5.7 Subespacio Principal de Dimensión k.

En forma semejante a todo lo anteriormente expuesto

El subespacio principal W, de dimensión k; el más próximo a la nube, en el sentido del momento de inercia, es el generado por los k-primeros ejes principales,

$\Delta u_1, \Delta u_2, \ldots, \Delta u_k$

donde $(V \circ D)(u_i) = \lambda_i u_i \quad i = 1,2,..,k$

y además:

$$I_{\Delta u_i^{\perp}} = \lambda_i \qquad i = 1,2,\ldots,k$$

$$I_{W^{\perp}} = \sum_{i=1}^{k} I_{\Delta u_i^{\perp}} = \sum_{i=1}^{k} \lambda_i$$

$$I_W = \text{traza } (VD) - \sum_{i=1}^{k} \lambda_i$$

y la parte de la inercia explicada por el subespacio W es

$$\frac{\sum_{i=1}^{k} \lambda_i}{\text{traza } (VD)}$$

Continuando los lineamientos de lo anterior.

$$\boxed{M} \subset W \Longleftrightarrow I_W = 0 \Longleftrightarrow \frac{\sum_{i=1}^{k} \lambda_i}{\text{traza } VD} = 1$$

A los ejes principales Δ_{u_i} estan asociados en E*.

Los factores principales $D u_i = V_i$

en F: Los componentes principales: $c^i = x'V_i = X' \circ D_{u_i}$

La componente centrada c^i tiene varianza λ_i

$$\sum_{j=1}^{n} p_i c_i^j = D_p(g, c_i) = 0 \qquad \sum_{j=1}^{n} p_j (c_i^j)^2 = D_p(c_i) = \lambda_i$$

3.5.8 Dualidad:

Para considerar la dualidad es necesario recordar lo siguiente:

Considerar en E, una base D-ortonormada de los vectores propios de (V o D); es considerar las siguientes propiedades:

$$(V \circ D)(u_i) = \lambda_i u_i ;\ D(u_i, u_i) = 1,\ D(u_i, u_j) = 0\ \forall i \neq j, i = 1,2,\dots p$$

Si V es regular o invertible, la base $\{u_i\ /i=1,2,\dots,p\}$ es además V^{-1} ortogonal.

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

Corresponden a los ejes Δ_{u_i};ejes principales contenidos en E lo siguiente.

en E^*: los factores principales que son los vectores propios de (D o V) de valor propio λ_i: D o V son D^{-1} y V simétricos $\{u_i/i=1,2,\dots,p\}$ es una base de E*, D^{-1} ortonormada y V ortogonal.

$$(2)\quad D \circ Vv_i = \lambda_i v_i ; \Longleftrightarrow (V \circ D)(u_i) = \lambda_i u_i$$

Como $D(u_i) = V_i \in E^*$

$$\Longleftarrow$$

$$M(VD(u_i)) = D(\lambda_i u_i) = \lambda_i D(u_i) = \lambda_i u_i$$

$$\Longrightarrow D \circ V \circ D(u_i) = \lambda_i D(u_i)$$

$$D^{-1}(v_i, v_i) = 1,\ D^{-1}(v_i, v_j) = 0 \text{ si } i \neq j$$

$$V(v_i, v_i) = \lambda_i ;\ V(v_i, v_j) = \text{ si } i \neq j$$

en F; los componentes principales: $c^i = X'V_i = X' \circ Du_i$

$$D \circ Vv_i = \lambda_i v_i \Longrightarrow D_o(X \circ D \circ X')v_i = D \circ X \circ D(c_i) = \lambda_i v_i$$

$$\Longrightarrow X' \circ D \circ X(c_i) = \lambda_i X'v_i$$

$$\Longrightarrow X' \circ D \circ X\, D_p \circ X'v_i = \lambda_i X'v_i$$

$$\Longrightarrow W \circ D_p\, c_i = \lambda_i c_i$$

Así los componentes principales son los vectores propios de la aplicación $W \circ D_p$ de los valores propios λ_i. Como además

$$D_p(c_i, c_j) = V(v_i, v_j) \qquad c_i \in F,\ v_i \in E^*$$

$W \circ D_p\, c_i = \lambda_i c_i$, $D_p(c_i, c_i) = \lambda_i$, $D_p(c_i, c_j) = 0$ si $i \neq j$

Enfasis sobre los rangos:

Si q es el número de valores propios no-nulos λ_i de $V \circ D$ (se puede dar que $\lambda_i = \lambda_j$ para todo $i \neq j$); se demostrará facilmente que:

$$q = \text{rang}(X) = \text{rang}(X') = \text{rang}(V) = \text{rang}(W) = \text{rang}(V \circ D) = \text{rang}(D \circ V)$$
$$= \text{rang}(W \circ D_p)$$

por lo tanto:

$$\dim(\text{Ker}(W \circ D_p)) = n-q$$

i.e. la dimensión del subespacio asociado al valor propio 0. es de dim $(n-q)$.

Así $\{ c_i \mid i = 1,2,\ldots,q \}$ es una base D_p-ortogonal de $X'(E^*) \subset F$ s.e.v. de F, engendrado por los caracteres x^j.

3.5.9 Descripción de la Nube de Individuos.

Definición: Soporte de la Nube:

Es el subespacio W generado por los q-ejes principales asociados a los velores propios no-nulos. Tiene por dimesión la del Ker (V o D); i.e. es un s.e.v. de V o D, de dimensión (p-q). Así la traza (V o D) = $\sum_{i=1}^{q} \lambda_i$

Una imagen Euclideana de ($\boxed{M}$, E, D,) es ($\boxed{M}$, W, D_W) donde D_W es la restricción de la métrica M al subespacio W; siendo esta la más simple.

Así como en la matriz X, un individuo está representado por los p-valores de una columna, en E, así en W; una columna de X, que representa un individuo está simbolizado por los q-valores en la componente principal.

Si la parte de la inercia explicada por el plano principal P, por ejemplo es próximo de 1, podemos lograr una descripción plana (una buena aproximación) de la nube a través de los α_i que son las proyecciones de los x_i sobre el plano P.

La Calidad Global de la representación de la nube $\boxed{M}$ en el plano principal P; se mide por: $\dfrac{\lambda_1 + \lambda_2}{\text{traza (VD)}}$

La calidad de la representación de un punto x_i:

Esta calidad vendrá dada por el índice

$L_1 = \cos\theta = \dfrac{||\alpha_i||}{||x_i||}$; coseno del ángulo entre x_i y α_i

donde α_i es la proyección de x_i sobre el plano principal P, y cada uno de estos vectores tiene por norma:

$$||x_i||^2 = \sum_{j=1}^{q} D^2(x_i,u_j) = \sum_{j=1}^{q} (c_j^i)^2$$

$$||\alpha_i||^2 = D^2(x_i,u_1) + D^2(x_i,u_2) = (c_1^i)^2 + (c_2^i)^2$$

Así, a mayor coseno mayor, menor ángulo, mayor proximidad ($0 \leq L_1 \leq 1$, $x_i \in P \Longrightarrow L_1 = 1$ que es el valor máximo de la proximidad.

Pero de acuerdo con este índice, si $x_i \in P^{\perp}$, entonces $\alpha_i = 0$, y él no nos informa, de la proximidad de x_i a P. Debido a esta pasamos a considerar el siguiente índice:

$$L_2 = \frac{||x_i - \alpha_i||}{\text{traza }(VD)-(\lambda_1+\lambda_2)}$$

que establece la razón entre la norma del vector $||x_i - \alpha_i||$ (distancia del vector x_i al plano P) y $\text{traza}\sqrt{(VD) - (\lambda_1 + \lambda_2)}$ que se puede considerar como una "varianza promedio" fuera del plano P.

Para distinguir en que semi-espacio está el vector x_i, L_1 ó L_2 llevarán el signo de la proyección de x_i sobre la tercera componente.

Estas proyecciones α_i de los puntos de la nube x_i considerando las calidades de las representaciones, nos permitiran agrupar en clase a los individuos semejantes, siendo éstos; los que tienen igual, ya sea L_1 ó L_2.

3.5.10 Descripción de los Caracteres:

Todo el trabajo realizado con los individuos, se pudo haber realizado con los caracteres; así tendríamos que maximizar.

$$\sum_{j=1}^{p} \sum_{j'=1}^{p} m_{jj'} Dp(x^j, c) D_p(x^{j'}, c)$$

Al igual que los componentes principales, estos caracteres $\{e^i, \ i = 1,2,\ldots,q\}$ forman una base D_p-ortonormal en subespacio de F; expresado como X'(E*). Este espacio es generado por los caracteres iniciales. Los caracteres principales son homotéticos a los componentes principales.

Dado que. $D_p(c^i, c^j) = 0 \quad i \neq j$

$$D_p(c^i) = D_p(c^i, c^i) = \lambda_i$$

todo caracter puede ser identificado por sus coordenadas en la base $\frac{1}{\sqrt{\lambda_j}} \quad c^j$

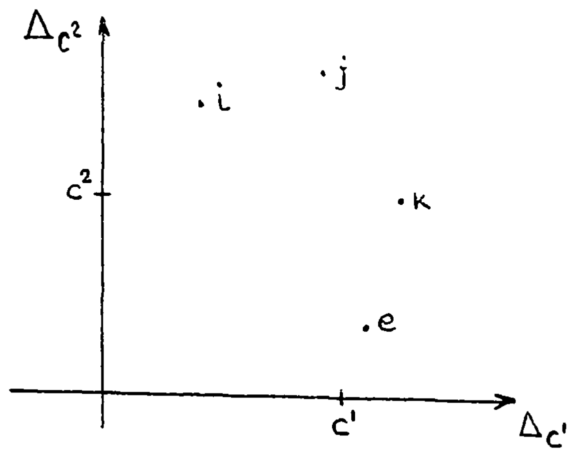
$$\frac{D_p(x^i, c^j)}{\sqrt{\lambda_j}} = \frac{\text{cov.}\ (x^i, c^j)}{\sqrt{\lambda_j}} \quad ; \quad j = 1,2,\ldots,q$$

De esta manera a la representación de los puntos individuos está asociada la representación de los puntos caracteres en el plano (c^1, c^2).

La intervención de un caracter en la descripción de la nube; se mide a través de la cercanía de su proyección, a los ejes c_1 y c_2) y se mide por.

$$\frac{D_p^2(x^i, c^1)}{\lambda_1} + \frac{D_p^2(x^i, c^2}{\lambda_2}$$

(i.e. al plano (c_1, c_2))



Así en este plano (c_1, c_2), sí las normas están bién construidas resulta:

La cercanía entre los caracteres x_i

La ortogonalidad entre los caracteres x_i

La proximidad y ortogonalidad entre los caracteres x_i y c_1 y c_2.

<u>Caracteres suplementarios</u>: son caracteres adicionales que pueden ser considerados en un momento dado; al proyectarse al plano (c^1, c^2) se puede evaluar la proximidad entre estos caracteres y los principales, para evaluar su importancia relativa.

CONCLUSIONES

Las conclusiones de este trabajo las hemos dividido en dos aspectos:

A.- Sobre la Utilidad del A.C.P. y el A.C.C.

1- El A.C.P. permite reducir la dimensionalidad, en un conjunto de datos, es decir constituye un conjunto mínimo de variables que ayuden a caracterizar una población.

2- A través del A.C.P. se logra un perfil de la población, un nuevo vector que tendrá una importancia mayor, con un nuevo significado y que representará una característica global.

3- El A.C.C. permite obtener niveles de asociación lineal entre grupos de variables.

4- En el A.C.C. se presenta como un conjunto de valores ordenados la correlación inicial del sistema. Tiene una gran utilidad en estudios exploratorios;así en la siguiente etapa solo se trabaja con las variables canónicas de mayor correlación.

Las dos componentes $X^{(1)}$ y $X^{(2)}$ de X; representan dos conjuntos de variables semejantes dentro de cada grupo pero deseamos averiguar las combinaciones lineales de máxima correlación.

B.- Sobre el desarrollo teórico analizado en el A.C.P. y A.C.C.

1- Dado un vector aleatorio X, de matriz de covarianza Σ , existe una transformación ortogonal P, tal que PX tiene matriz de covarianza diagonal cuyas entradas corresponden a los valores propios de Σ . Sí se desconoce Σ , también se garantiza la existencia de los estimadores de las componentes principales poblacionales que en este caso, se obtendrán a partir del estimador de máxima verosimilitud de Σ .

2- La varianza total del sistema y la varianza generalizada del vector X, son invariantes ante transformaciones ortogonales.

3- Las estimaciones de las componentes principales se pueden obtener a partir del estimador máxima verosimilitud de la matriz de correlación.

4- El eje y el plano principal son la recta y el plano más próximo a la nube de individuos o caracteres. Esta proximidad se expresa a través del momento de Inercia. Donde el recurso matemático básico utilizado fué: las formas bilineales simétricas definidas positivas y las cuatro aplicaciones que ella induce.

BIBLIOGRAFIA

[1] Anderson, T.W. — An Introduction to Multivariate Statistical Analysis. John Wiley and Sons, Inc. N.Y. 1958.

[2] Benzecri, J.P. — L'analyse Des Données. Tomo 1 y 2 Dunod, Paris 1973.

[3] Caillez, F. y Pages, J.P. — Introduction a l'Analyse des Données Smash, ASU, Buro, Paris 1976.

[4] Cooley, W.W — Multivariate Data Analysis J. Wiley and Son, N.Y. 1971.

[5] Gnadesekan, R — Methods for Statistical Data Analysis of Multivariate Observations. 1977.

[6] Kendall, M.G. y Stuart A. — The Advanced Theory of Statistic Vol. 1. Mc Millan, Hafner Press N.Y. 1969.

[7] Lebart L. y Fénelon J.P. — Statistique et Informatique Appliquées 2da Edition. Dunod, Paris.

[8] Morrison, D.E. — Multivariate Statistical Methods Ma Graw-Hill Book Co. 1976.

[9] Muirhead, R.J. — Aspects of Multivariate Statistical Theory. J. Wiley and Sons. 1982.

[10] Rao, C. R. — Linear Statistical Inference J. Wiley and Son N.Y. 1973.

[11] S.S. Wilks — Mathematical Statistics John Wiley and Sons, Inc. 1962.