Anales de la Real Academia de Doctores de España
Volumen 16, n.º 2, 2012, pp. 157-180

ELEMENTOS BÁSICOS EN EL ANÁLISIS EN COMPONENTES PRINCIPALES (ACP)

[Cuarta parte: práctica (métricas no usuales y usuales en el ACP)]

DOCTOR D. FCO. JAVIER DÍAZ-LLANOS SAINZ-CALLEJA*
Académico de Número de la Sección de Ingeniería

M. YVES ESCOUFIER**

DRA. DÑA. M.ª DEL CARMEN CERMEÑO CARRASCO***

D. LUIS FELIPE GRAU SEGURA****

RESUMEN

El objetivo de este artículo se articula en dos partes. La primera consiste en presentar —de la forma más didáctica posible— cómo se calculan **los ejes principales, los factores principales** y **las componentes principales** bajo el concepto de las **métricas no usuales.** De los dos ejercicios de cálculo que proponemos se ha realizado el primero de forma completa mientras que, en el segundo, se proporcionarán los elementos suficientes para realizarlo hasta la consecución de los resultados finales con la ayuda de una simple calculadora.

En cuanto a la segunda parte, consiste en presentar —a partir de varios ejercicios extraídos de la bibliografía contenida en este artículo—, las diferencias existentes con respecto a los tres elementos básicos aludidos por el hecho de aplicar las **métricas usuales,** así como las **no usuales** en el **ACP.**

Palabras claves: valores propios, vectores propios, métricas, raíz cuadrada de una matriz simétrica definida positiva, simetrización de una matriz a diagonalizar, los ejes principales, los factores principales y las componentes principales.

---

* Medalla al Mérito Doctoral. Categoría de Plata de la Real Academia de Doctores de España.

** Professeur d'Analyse des Données. Ancien Président à l'Université de Montpellier II.

*** Miembro de Número de la Sociedad Española de Genética Humana. N.º 467. Antigua profesora de Genética y Citogenética en las Universidades Technische de München donde obtuvo «venia docendis», y Freie Angewandte, Berlin, Deutschland. Referee de artículos desde dichas Universidades.

**** Licenciado en Ciencias y Técnicas Estadísticas en la Universidad Carlos III de Getafe (Madrid).

ABSTRACT

The aim of this work have been to carry out —didactically— the calculation of the principal axies, principal factors and principal components (respectively), by using usual and non-usual metrics in the PCA, as well as: to establish the existing differences, found in the three basic elements above mentioned by applying such **metrics.** Some exercises are also presented and total or partially developed.

By comparing both, usual and non-usual metrics, there seems to be an improvement in the results obtained with the last one.

Key words: Characteristic roots and vectors, non-usual and usual metrics, square roots of symmetric matrices positive definite, procedure for making symmetric one matric, principal (axes, factors and components).

INTRODUCCIÓN

Se sabe —a ciencia cierta— que la aplicación de las **métricas usuales** está contemplada en muchos libros de «Analyse des Données». Sin embargo, este hecho se desconoce para las **métricas no usuales.**

Dado que el procedimiento que vamos a exponer —a nivel práctico— no está contenido en ningún libro de **análisis estadístico multidimensional lineal** —en sí mismo— y ni tan siquiera aplicado a las disciplinas de: **Investigación del Mercado, Psicología, Sociología, Química Experimental, Geología, Medicina, Farmacología, Biología,** etc., nos permitiremos mostrarlo amparándonos en la parte teórica contenida en (1, 2).

Entre otros libros, para la realización de tal desarrollo, hemos elegido un ejercicio contenido en el libro de J. Bon y P. Grégory (3(pp. 43-84)).

Aunque este ejercicio no sea el más idóneo para aplicar un **ACP** ya que, según apunta el profesor Jean-Paul Benzécri: a una **tabla de notas** debe aplicarse un **análisis de correspondencias (AFC)** (4(pp. 25-27)).

Además, según Thierry Foucart (5(p. 16)), en el caso hipotético que fuera una **tabla de datos** susceptible de la aplicación de un **ACP,** dicha **tabla** debe contener más de quince individuos y más de cuatro variables cuantitativas para obtener una interpretación adecuada.

Aún así, hemos elegido el ejercicio contenido en (3(pp. 43-84)) para poner de manifiesto la importancia de los **análisis de datos** en la **política de producto.** A su vez, aún no introduciendo de forma directa las **métricas usuales,** J. Bon y P. Grégory (3(pp. 43-84) desarrollan el **ACP** de una manera bastante aceptable, con respecto a otros libros de texto donde también se omiten las **métricas usuales** en R^p y en R^n.

La realización de dicho ejercicio es de carácter meramente ilustrativo con el objeto de familiarizar a los investigadores con **las métricas.**

No obstante, se presenta de la forma más didáctica posible con el fin de que estos puedan analizar correctamente sus propios datos empíricos.

Además, dado que la matriz de datos iniciales es de pequeñas dimensiones (3 filas y 4 columnas), este hecho nos permitirá exponer una presentación práctica —clara—, de la teoría contenida en (2).

Introduciremos —para cada uno de los ejercicios— en R^p dos **métricas** distintas de las **usuales**, permaneciendo la misma en R^n y distinta de la **usual.**

Con el fin de que los investigadores no pertenecientes al área de la Matemática puedan seguir —fácilmente— el procedimiento que presentamos, hemos partido de la misma tabla de datos ya aludida en (6) haciendo uso de las **métricas usuales.**

Aquellos responsables de proyectos de investigación y múltiples publicaciones en revistas (índice de impacto en numerosas ocasiones) y que aplican —sistemáticamente— el **ACP,** según las **métricas usuales,** programadas en los paquetes de programas comercializados de **Análisis Estadístico Multidimensional Lineal,** tendrán suficientes elementos para discernir que los resultados que obtengan mediante dichas **métricas** difieren considerablemente del obtenido por las **métricas no usuales.** Esto será fácilmente contrastable, tras la realización de los dos primeros ejercicios aquí propuestos. El contenido algebraico no supera al utilizado, en un primer curso, de una Escuela Técnica Superior de Ingeniería o Facultad de Ciencias.

Además, también presentamos —entre otras— una tabla de datos extraída del libro de Christian Bialès (7(pp. 224-238)) para que puedan verificar por sí mismos, los resultados obtenidos por este investigador, el cual omite las **métricas** en R^p y en R^n. Aunque en el libro de este autor no hace alusión a las **métricas**, hemos creído necesario hacer mención de él, ya que hace una serie de comentarios de interés sobre el **ACP** a pesar de contener ciertas imprecisiones a nivel de cálculo. La tabla de datos de los libros (3(pp. 43-84)) y (7(pp. 224-238)) son de pequeñas dimensiones, esto nos permitirá desarrollarlos de una forma sencilla.

MATERIAL

Material teórico (Primer ejercicio resuelto paso a paso hasta el final)

Elementos básicos y significado del contenido del **triplete estadístico.**

$$\left(\chi^{c}_{(n,p)},\ \ M_{(p,p)},\ \ N_{(n,n)}\right)$$

$$\chi^{c}_{(n,p)} = \left(I_{(n,n)} - 1_n\ 1_n^T\ \ N_{(n,n)}\right) X_{(n,p)}$$

$I_{(n,n)}$: es la matriz identidad de orden n.

1_n: es un vector columna constituido por n unos.

1_n^T: es el vector transpuesto de 1_n.

$N_{(n,n)}$: es una matriz cuadrada diagonal de orden n.

Los elementos de la diagonal principal son los pesos que los investigadores asignarán a cada uno de los individuos de sus datos empíricos. Se tiene que verificar la siguiente condición:

$$\sum_{i=1}^{i=n} p_i = 1 \qquad \text{Todos los } p_i \text{ son positivos.}$$

$M_{(p,p)}$: es una matriz cuadrada regular simétrica definida positiva.

Así como $N_{(n,n)}$ tiene que ser una matriz diagonal, $M_{(p,p)}$ no tiene porqué serlo.

Material práctico (Primer ejercicio resuelto paso a paso hasta el final)

Las matrices que figuran en el **triplete estadístico** son las que mostramos a continuación:

$$I_{(4,4)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$1_4 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \qquad 1_4^T = (1 \quad 1 \quad 1 \quad 1)$$

$$N_{(4,4)} = \frac{1}{10}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}$$

$$\chi_{(4,3)} = \begin{pmatrix} 5 & 1 & 3 \\ 1 & 6 & 5 \\ 7 & 8 & 3 \\ 7 & 5 & 9 \end{pmatrix}$$

$$\chi^c_{(4,3)} = \frac{1}{10}\begin{pmatrix} -6 & -47 & -28 \\ -49 & 3 & -8 \\ 14 & 23 & -8 \\ 14 & -7 & 32 \end{pmatrix}$$

$$\left(\chi^c_{(4,3)}\right)^T = \chi^F_{(3,4)}$$

$$\chi^F_{(3,4)} = \frac{1}{10}\begin{pmatrix} -6 & -49 & 14 & 14 \\ -47 & 3 & 23 & -7 \\ -28 & -8 & -8 & 32 \end{pmatrix}$$

$$M_{(3,3)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 16 \end{pmatrix}$$

Material práctico (Segundo ejercicio resuelto en el que proporcionamos al lector material suficiente para llegar con éxito al resultado final)

Para no resultar reiterativos tan sólo vamos a mostrar la matriz $M_{(3,3)}$, que es el único elemento del **triplete estadístico** que hemos cambiado con respecto al **primer ejercicio.**

$$M_{(3,3)} = \begin{pmatrix} 3 & 2 & 0 \\ 2 & 4 & -2 \\ 0 & -2 & 5 \end{pmatrix}$$

$M_{(3,3)}$ es una matriz simétrica real de orden 3 definida positiva.

Para aquellos profesionales no especialistas en el dominio de la Matemática que deseen entender los conceptos algebraicos básicos contenidos en este artículo, les recomendamos que consulten los libros de Jean-François Audroing (8(pp. 193-226)), y el de Bernard Guerrien (9(pp. 211-245)).

Ambos libros son muy didácticos y contienen —aunque no todos—, muchos de los elementos básicos de Álgebra Lineal que hemos introducido en este artículo.

Una vez definidos los elementos contenidos en el **triplete estadístico** procederemos a mostrar —paso a paso— el proceso metodológico para el primer ejercicio propuesto, en el cual se introducen en R^p y R^n **métricas no usuales** en el ACP, y para el segundo les proporcionaremos los elementos básicos para que los interesados en este tipo de metodología puedan llegar con éxito al resultado final.

MÉTODOS

MÉTODO (PRIMER EJERCICIO)

El método que hemos utilizado se basará únicamente en el cálculo de **los ejes principales, los factores principales y las componentes principales.**

1. Ejes principales:

Dado que los **ejes principales** son los vectores propios de:

$$\chi^F_{(p,n)} N_{(n,n)} \chi^c_{(n,p)} M_{(p,p)} \qquad M_{(p,p)} \text{ – \textbf{ortonormados;}}$$

es decir, los vectores propios $\vec{\boldsymbol{u}}^{\boldsymbol{j}}$ que verifican:

$$\chi^F_{(p,n)} N_{(n,n)} \chi^c_{(n,p)} M_{(p,p)}\ \vec{\boldsymbol{u}}^{\boldsymbol{j}} = \lambda_j\ \vec{\boldsymbol{u}}^{\boldsymbol{j}} \qquad M_{(p,p)} \text{ – \textbf{ortonormados.}}$$

Dicho de otra manera más usual, si los vectores $\vec{\boldsymbol{u}}^{\boldsymbol{j}}$ son $M_{(p,p)}$ **– ortonormados,** tienen que cumplirse estas dos condiciones:

$$\vec{\boldsymbol{u}}^{\boldsymbol{j}\ T} M_{(p,p)}\ \vec{\boldsymbol{u}}^{\boldsymbol{j'}} = 0 \qquad\qquad j \neq j'$$

$$\vec{\boldsymbol{u}}^{\boldsymbol{j}\ T} M_{(p,p)}\ \vec{\boldsymbol{u}}^{\boldsymbol{j}} = 1$$

Por lo tanto, lo primero que tenemos que calcular para conseguir nuestro objetivo es calcular:

$$\chi^{F}_{(p,n)}\; N_{(n,n)}\; \chi^{c}_{(n,p)}$$

Sustituyendo los valores numéricos de estas tres matrices ya definidas con anterioridad tenemos,

$$\chi^{F}_{(3,4)} N_{(4,4)}\, \chi^{c}_{(4,3)} = \begin{pmatrix} 6{,}210 & 0{,}562 & 2{,}408 \\ 0{,}562 & 4{,}010 & -0{,}180 \\ 2{,}408 & -0{,}180 & 5{,}200 \end{pmatrix}$$

De la mera observación de $\chi^{F}_{(3,4)} N_{(4,4)}\, \chi^{c}_{(4,3)}$ se concluye que dicho producto de matrices conlleva a una matriz simétrica real.

Ahora, nos falta multiplicar este resultado por la matriz $M_{(3,3)}$.

$$\chi^{F}_{(3,4)} N_{(4,4)}\; \chi^{c}_{(4,3)}\; M_{(3,3)} = \begin{pmatrix} 6{,}210 & 2{,}248 & 38{,}528 \\ 0{,}562 & 16{,}040 & -2{,}880 \\ 2{,}408 & -0{,}720 & 83{,}200 \end{pmatrix}$$

En esta ocasión observamos que la multiplicación de estas matrices conlleva que la matriz resultante ya no es una matriz simétrica real.

Por tal motivo, tendremos que proceder a la **simetrización de la matriz** (2,10)

$$\chi^{F}_{(3,4)} N_{(4,4)}\, \chi^{c}_{(4,3)} M_{(3,3)}$$

con el fin de que los **vectores propios** $\vec{u}^{j}$ sean $M_{(p,p)}$ **– ortonormados** y, por lo tanto, los $\vec{u}^{j}$ serán **los ejes principales.**

Pasos a seguir:

Teniendo en cuenta (2,10), calcularemos la siguiente expresión:

$$M^{\frac{1}{2}}_{(3,3)} \chi^{F}_{(3,4)} N_{(4,4)}\, \chi^{c}_{(4,3)} M^{\frac{1}{2}}_{(3,3)}$$

Sustituyendo estas matrices por sus valores reales y multiplicándolas, así como teniendo en cuenta que

$$S_{(3,3)} = M^{\frac{1}{2}}_{(3,3)}\, \chi^{F}_{(3,4)} N_{(4,4)}\, \chi^{c}_{(4,3)} M^{\frac{1}{2}}_{(3,3)}$$

estamos en condiciones de obtener la matriz $S_{(3,3)}$.

$$S_{(3,3)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 6,210 & 0,562 & 2,408 \\ 0,562 & 4,010 & -0,180 \\ 2,408 & -0,180 & 5,200 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

Así pues, $S_{(3,3)}$ es,

$$S_{(3,3)} = \begin{pmatrix} 6,210 & 1,124 & 9,632 \\ 1,124 & 16,040 & -1,440 \\ 9,632 & -1,440 & 83,200 \end{pmatrix}$$

De la mera observación de $S_{(3,3)}$ se desprende que dicha matriz es simétrica real definida positiva.

Observaciones de interés entre:

$$\chi^{F}_{(3,4)}\, N_{(4,4)}\, \chi^{c}_{(4,3)} M_{(3,3)} \qquad [1]$$

y

$$M^{\frac{1}{2}}_{(3,3)}\, \chi^{F}_{(3,4)}\, N_{(4,4)}\, \chi^{c}_{(4,3)} M^{\frac{1}{2}}_{(3,3)} \qquad [2]$$

1.ª) Los **valores propios** de [1] y de [2] son los mismos.

La diagonal principal de [1] es la misma que la de [2].

Valores propios	
$\chi^{F}_{(3,4)} N_{(4,4)}\ \chi^{c}_{(4,3)} M_{(3,3)}$	$M^{\frac{1}{2}}_{(3,3)}\, \chi^{F}_{(3,4)}\, N_{(4,4)}\, \chi^{c}_{(4,3)}\, M^{\frac{1}{2}}_{(3,3)}$
84,4111521610	84,4111521610
16,1652777790	16,1652777790
4,8735700593	4,8735700593

2.ª) Los **vectores propios** de [1] y de [2] son distintos.

Vectores propios		
$\chi^{F}_{(3,4)}\, N_{(4,4)}\, \chi^{c}_{(4,3)}\, M_{(3,3)}$		
–0,4409033660	–0,228530656	–0,9978128730
0,0341581759	–0,973534123	0,0582692973
–0,8969044420	–0,002247235	0,0312115246

Vectores propios		
$M_{(3,3)}^{\frac{1}{2}} \chi_{(3,4)}^{F} N_{(4,4)} \chi_{(4,3)}^{c} M_{(3,3)}^{\frac{1}{2}}$		
0,1219563911	0,1165702428	0,9856662808
–0,0188955870	0,9931718658	–0,1151199450
0,9923555791	0,0045851299	–0,1232263200

3.ª) Mientras que la matriz [1] no es simétrica, la matriz [2] sí lo es y, por lo tanto, los **vectores propios** de [2] son ortogonales.

4.ª) A partir de los **vectores propios** de [2] mediante una simple transformación:

$$\vec{\boldsymbol{u}}^{\boldsymbol{j}} = M_{(3,3)}^{-\frac{1}{2}} \vec{\boldsymbol{s}}^{\boldsymbol{j}} \qquad [3]$$

nos permitirá calcular los **ejes principales**.

1. Cálculo de los ejes principales

1.1. Primer eje principal

Aplicando [3] a nuestros datos concretos tenemos,

$$\vec{\boldsymbol{u}}^{\boldsymbol{1}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/4 \end{pmatrix} \begin{pmatrix} 0{,}1219563911 \\ -0{,}0188955870 \\ 0{,}9923555791 \end{pmatrix} = \begin{pmatrix} 0{,}1219563911 \\ -0{,}0094477930 \\ 0{,}2480888940 \end{pmatrix}$$

Por lo tanto, el **primer eje principal** es

$$\vec{\boldsymbol{u}}^{\boldsymbol{1}} = \begin{pmatrix} 0{,}1219563911 \\ -0{,}0094477930 \\ 0{,}2480888940 \end{pmatrix}$$

1.2. Segundo eje principal

Aplicando [3] a nuestros datos concretos tenemos,

$$\vec{\boldsymbol{u}}^{\boldsymbol{2}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/4 \end{pmatrix} \begin{pmatrix} 0{,}1165702428 \\ 0{,}9931718658 \\ 0{,}0045851299 \end{pmatrix} = \begin{pmatrix} 0{,}1165702428 \\ 0{,}4965859320 \\ 0{,}0011462820 \end{pmatrix}$$

Por lo tanto, el **segundo eje principal** es

$$\vec{\boldsymbol{u}}^{\boldsymbol{2}} = \begin{pmatrix} 0{,}1165702428 \\ 0{,}4965859320 \\ 0{,}0011462820 \end{pmatrix}$$

1.3. Tercer eje principal

Aplicando [3] a nuestros datos concretos tenemos,

$$\vec{\boldsymbol{u}}^{\mathbf{3}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/4 \end{pmatrix} \begin{pmatrix} 0{,}9856662805 \\ -0{,}1151199450 \\ -0{,}1232263200 \end{pmatrix} = \begin{pmatrix} 0{,}9856662805 \\ -0{,}0575599720 \\ 0{,}0308065800 \end{pmatrix}$$

Por lo tanto, el **tercer eje principal** es

$$\vec{\boldsymbol{u}}^{\mathbf{3}} = \begin{pmatrix} 0{,}9856662805 \\ -0{,}0575599720 \\ 0{,}0308065800 \end{pmatrix}$$

Los lectores, con interés de innovación, podrán contrastar —fácilmente— que los vectores propios $\vec{\boldsymbol{u}}^{\mathbf{1}}$, $\vec{\boldsymbol{u}}^{\mathbf{2}}$ *y* $\vec{\boldsymbol{u}}^{\mathbf{3}}$ son los **ejes principales** ya que verifican las siguientes condiciones:

$$\vec{\boldsymbol{u}}^{\boldsymbol{j}^T} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 16 \end{pmatrix} \vec{\boldsymbol{u}}^{\boldsymbol{j}'} = 0 \qquad\qquad j \neq j' \qquad j = 1, 2, 3$$

$$\vec{\boldsymbol{u}}^{\boldsymbol{j}^T} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 16 \end{pmatrix} \vec{\boldsymbol{u}}^{\boldsymbol{j}} = 1 \qquad\qquad\qquad j = 1, 2, 3$$

2. Cálculo de los factores principales

A partir de los **valores propios** de [2] una simple relación tal como la que mostramos a continuación:

$$M_{(3,3)}\, \vec{\boldsymbol{u}}^{\boldsymbol{j}} = \vec{\boldsymbol{v}}^{\boldsymbol{j}} \qquad\qquad \vec{\boldsymbol{u}}^{\boldsymbol{j}} \in R^{p} \qquad \vec{\boldsymbol{v}}^{\boldsymbol{j}} \in R^{p^*} \qquad [4]$$

nos permitirá calcular los **factores principales.**

2.1. Primer factor principal

Aplicando [4] a nuestros datos concretos tenemos,

$$\vec{\boldsymbol{v}}^{\mathbf{1}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 16 \end{pmatrix} \begin{pmatrix} 0{,}1219563911 \\ -0{,}0094477930 \\ 0{,}2480888940 \end{pmatrix} = \begin{pmatrix} 0{,}1219563911 \\ -0{,}0377911720 \\ 3{,}9694223040 \end{pmatrix}$$

Por lo tanto, el **primer factor principal** es

$$\vec{\boldsymbol{v}}^{\mathbf{1}} = \begin{pmatrix} 0{,}1219563911 \\ -0{,}0377911720 \\ 3{,}9694223040 \end{pmatrix}$$

2.2. Segundo factor principal

Aplicando [4] a nuestros datos concretos tenemos,

$$\vec{v}^2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 16 \end{pmatrix} \begin{pmatrix} 0{,}1165702428 \\ 0{,}4965859320 \\ 0{,}0011462820 \end{pmatrix} = \begin{pmatrix} 0{,}1165702428 \\ 1{,}9863437280 \\ 0{,}0183405120 \end{pmatrix}$$

Por lo tanto, el **segundo factor principal** es

$$\vec{v}^2 = \begin{pmatrix} 0{,}1165702428 \\ 1{,}9863437280 \\ 0{,}0183405120 \end{pmatrix}$$

2.3. Tercer factor principal

Aplicando [4] a nuestros datos concretos tenemos,

$$\vec{v}^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 16 \end{pmatrix} \begin{pmatrix} 0{,}9856662805 \\ -0{,}0575599720 \\ -0{,}0308065800 \end{pmatrix} = \begin{pmatrix} 0{,}9856662805 \\ -0{,}2302398880 \\ 0{,}4929058800 \end{pmatrix}$$

Por lo tanto, el **tercer factor principal** es

$$\vec{v}^3 = \begin{pmatrix} 0{,}9856662805 \\ -0{,}2302398880 \\ 0{,}4929058800 \end{pmatrix}$$

Los lectores podrán constatar —fácilmente— que los **vectores propios** $\vec{v}^1$, $\vec{v}^2$ y $\vec{v}^3$ son los **factores principales** ya que verifican las siguientes condiciones:

$$\vec{v}^{j^T} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/4 & 0 \\ 0 & 0 & 1/16 \end{pmatrix} \vec{v}^{j'} = 0 \qquad j \neq j' \quad j = 1, 2, 3$$

$$\vec{v}^{j^T} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/4 & 0 \\ 0 & 0 & 1/16 \end{pmatrix} \vec{v}^{j} = 1 \qquad j = 1, 2, 3$$

3. Cálculo de las componentes principales

Las **componentes principales** las podemos obtener haciendo uso de **dos procedimientos.** El **primer procedimiento** más simple que el **segundo.** Así como, el primero hace uso de los **factores principales**, el segundo de los **ejes principales.**

3.1. Primer procedimiento

Este procedimiento consiste en la aplicación de

$$\chi^{c}_{(4,3)}\,\vec{\boldsymbol{v}}^{\boldsymbol{j}}_{(\boldsymbol{3},\boldsymbol{1})} = \vec{\boldsymbol{c}}^{\boldsymbol{j}}_{(\boldsymbol{3},\boldsymbol{1})},\quad \vec{\boldsymbol{v}}^{\boldsymbol{j}} \in R^{p*},\quad \vec{\boldsymbol{c}}^{\boldsymbol{j}} \in R^{n}\quad j = 1,2,3 \qquad [5]$$

a nuestros datos concretos.

3.1.1. Primera componente principal

Sustituyendo los valores numéricos de $\chi^{c}_{(3,4)}$ y $\vec{\boldsymbol{v}}^{\boldsymbol{1}}_{(\boldsymbol{3},\boldsymbol{1})}$ en [5] tenemos,

$$\vec{\boldsymbol{c}}^{\boldsymbol{1}} = \frac{1}{10}\begin{pmatrix} -6 & -47 & -28 \\ -49 & 3 & -8 \\ 14 & 23 & -8 \\ 14 & -7 & 32 \end{pmatrix}\begin{pmatrix} 0{,}1219563911 \\ -0{,}0377911720 \\ 3{,}9694333040 \end{pmatrix} =$$

$$= \begin{pmatrix} -11{,}009968580 \\ -3{,}784470311 \\ -3{,}091727391 \\ 12{,}899379340 \end{pmatrix}$$

Por lo tanto, **la primera componente principal** es

$$\vec{\boldsymbol{c}}^{\boldsymbol{1}} = \begin{pmatrix} -11{,}009968580 \\ -3{,}784470311 \\ -3{,}091727391 \\ 12{,}899379340 \end{pmatrix}$$

3.1.2. Segunda componente principal

Sustituyendo los valores numéricos de $\chi^{c}_{(4,3)}$ y $\vec{\boldsymbol{v}}^{\boldsymbol{2}}_{(\boldsymbol{3},\boldsymbol{1})}$ en [5] tenemos,

$$\vec{\boldsymbol{c}}^{\boldsymbol{2}} = \frac{1}{10}\begin{pmatrix} -6 & -47 & -28 \\ 49 & 3 & -8 \\ 14 & 23 & -8 \\ 14 & -7 & 32 \end{pmatrix}\begin{pmatrix} 0{,}1165702428 \\ 1{,}9863437280 \\ 0{,}0183405120 \end{pmatrix} = \begin{pmatrix} -9{,}457111100 \\ 0{,}010036523 \\ 4{,}717116504 \\ -1{,}168552632 \end{pmatrix}$$

Por lo tanto, la **segunda componente principal** es

$$\vec{\boldsymbol{c}}^{\boldsymbol{2}} = \begin{pmatrix} -9{,}457111100 \\ 0{,}010036523 \\ 4{,}717116504 \\ -1{,}168552632 \end{pmatrix}$$

3.1.3. Tercera componente principal

Sustituyendo los valores numéricos de $\chi^{c}_{(4,3)}$ y $\vec{\boldsymbol{v}}^{\mathbf{3}}_{(\mathbf{3},\mathbf{1})}$ en [5] tenemos,

$$\vec{\boldsymbol{c}}^{\mathbf{3}} = \frac{1}{10}\begin{pmatrix} -6 & -47 & -28 \\ -49 & 3 & -8 \\ 14 & 23 & -8 \\ 14 & -7 & 32 \end{pmatrix}\begin{pmatrix} 0{,}9856662805 \\ -0{,}2322398880 \\ -0{,}4929058800 \end{pmatrix} = \begin{pmatrix} 1{,}8708641700 \\ -4{,}5045120340 \\ 1{,}2447057540 \\ -0{,}0361981024 \end{pmatrix}$$

Por lo tanto, la **tercera componente principal** es

$$\vec{\boldsymbol{c}}^{\mathbf{3}} = \begin{pmatrix} 1{,}8708641700 \\ -4{,}5045120340 \\ 1{,}2447057540 \\ -0{,}0361981024 \end{pmatrix}$$

Los lectores podrán constatar —fácilmente— que los **vectores propios** $\vec{\boldsymbol{c}}^{\mathbf{1}}$, $\vec{\boldsymbol{c}}^{\mathbf{2}}$ y $\vec{\boldsymbol{c}}^{\mathbf{3}}$ son las **componentes principales** ya que verifican las siguientes condiciones:

$$\vec{\boldsymbol{c}}^{\boldsymbol{j}^T}\frac{1}{10}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}\vec{\boldsymbol{c}}^{\boldsymbol{j}'} = 0 \qquad j \neq j' \qquad j = 1, 2, 3$$

$$\vec{\boldsymbol{c}}^{\boldsymbol{j}^T}\frac{1}{10}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}\vec{\boldsymbol{c}}^{\boldsymbol{j}} = \lambda_j \qquad j = 1, 2, 3$$

Dicho de otra manera, los cuadrados de la norma de los vectores $\vec{\boldsymbol{c}}^{\mathbf{1}}$, $\vec{\boldsymbol{c}}^{\mathbf{2}}$ y $\vec{\boldsymbol{c}}^{\mathbf{3}}$ en el sentido de la métrica $N_{(4,4)}$

$$N_{(4,4)} = \frac{1}{10}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}$$

son los que a continuación mostramos,

$$\left\|\vec{\boldsymbol{c}}^{\mathbf{1}}\right\|^2_{N_{(4,4)}} = 84{,}411612340$$

$$\left\|\vec{\boldsymbol{c}}^{\mathbf{2}}\right\|^2_{N_{(4,4)}} = 16{,}165277720$$

$$\left\|\vec{\boldsymbol{c}}^{\mathbf{3}}\right\|^2_{N_{(4,4)}} = 4{,}873450852$$

De lo que se desprende,

$$\sum_{j=1}^{j=3}\left\|\vec{\boldsymbol{c}}^{\boldsymbol{j}}\right\|^2_{N_{(4,4)}} = \sum_{j=1}^{j=3}\lambda_j = 105{,}4503409 \qquad j = 1{,}2{,}3$$

3.2. Segundo procedimiento

Este procedimiento consiste en el cálculo de $\vec{\boldsymbol{c}}^{\mathbf{1}}$, $\vec{\boldsymbol{c}}^{\mathbf{2}}$ y $\vec{\boldsymbol{c}}^{\mathbf{3}}$ a partir de,

$$\chi^{c}_{(4,3)} M_{(3,3)}\, \chi^{c}_{(4,3)} N_{(4,4)}\, \vec{\boldsymbol{c}}^{\,\boldsymbol{j}}_{(\mathbf{4},\mathbf{1})} = \lambda_j \ \vec{\boldsymbol{c}}^{\,\boldsymbol{j}}_{(\mathbf{4},\mathbf{1})} \quad [6]$$

bajo las restricciones:

$$\begin{array}{llll} \vec{\boldsymbol{c}}^{\boldsymbol{j}^T} N_{(4,4)} \ \vec{\boldsymbol{c}}^{\boldsymbol{j}'} = 0 & & j \neq j' & j = 1,2,3 \\ \vec{\boldsymbol{c}}^{\boldsymbol{j}^T} N_{(4,4)} \ \vec{\boldsymbol{c}}^{\boldsymbol{j}} = \lambda_j & & & j = 1,2,3 \end{array}$$

De la mera observación de las cuatro matrices contenidas en el primer miembro de [6] concluimos que mientras el resultado de las tres primeras es una matriz simétrica, el de la cuatro es una matriz no simétrica. Sólo cuando se verifique: $N_{(4,4)} = \frac{1}{4} I_{(4,4)}$ el resultado de la multiplicación de las cuatro matrices contenidas en [6] es una matriz simétrica definida positiva.

Sustituyendo los valores numéricos de las matrices $\chi^{c}_{(4,3)}$, $M_{(3,3)}, \chi^{F}_{(4,3)}$ y $N_{(4,4)}$ en [6] y multiplicando dichas matrices obtenemos,

$$\begin{pmatrix} 21{,}416 & 6{,}628 & -2{,}472 & -52{,}416 \\ 3{,}314 & 6{,}922 & 1{,}842 & -19{,}464 \\ -0{,}824 & 1{,}228 & 10{,}008 & -18{,}176 \\ -13{,}104 & -9{,}732 & -13{,}636 & 67{,}104 \end{pmatrix} \vec{\boldsymbol{c}}^{\boldsymbol{j}}_{(\mathbf{4},\mathbf{1})} = \lambda_j \,\vec{\boldsymbol{c}}^{\boldsymbol{j}}_{(\mathbf{4},\mathbf{1})} \quad [7]$$

De la mera observación de la matriz contenida en [7] se desprende el siguiente resultado:

$$Tr(\chi^{c}_{(4,3)} \ M_{(3,3)} \ \chi^{F}_{(3,4)} \ N_{(4,4)}) = 105{,}45$$

Si deseamos continuar con este procedimiento, ineludiblemente tenemos que someter a la matriz contenida en [7] al proceso de **simetrización de una matriz** (2,10) para asegurarnos de que los **vectores propios** $\vec{\boldsymbol{c}}^{\mathbf{1}}$, $\vec{\boldsymbol{c}}^{\mathbf{2}}$ y $\vec{\boldsymbol{c}}^{\mathbf{3}}$ cumplen la condición de ser $N_{(4,4)}$ **–ortogonales.**

Así pues, con el fin de que los profesionales se vayan familiarizando con nuestra innovadora propuesta, les proponemos que apliquen el proceso de **simetrización de la matriz** (2,10) contenida en [7] y, comprueben si los $\vec{\boldsymbol{c}}^{\mathbf{1}}$, $\vec{\boldsymbol{c}}^{\mathbf{2}}$ y $\vec{\boldsymbol{c}}^{\mathbf{3}}$ son —al menos— prácticamente iguales que aquellos obtenidos mediante el **primer procedimiento.**

No está de más recordar que mediante la aplicación del **primer procedimiento** se obtienen —fácilmente— no sólo cada uno de los tres **valores propios** cuya suma es 105,4503409, sino también, las **componentes principales:** $\vec{\boldsymbol{c}}^{\mathbf{1}}$, $\vec{\boldsymbol{c}}^{\mathbf{2}}$ y $\vec{\boldsymbol{c}}^{\mathbf{3}}$.

3.3. Recomendación en cuanto a la elección de uno de estos dos procedimientos

Si entre los **dos procedimientos** ya aludidos se optase por el segundo, la matriz que tendría que ser sometida al proceso de **simetrización** —en nuestro caso concreto— es de orden 4 y, por lo tanto, no se podría llegar al resultado final con una simple calculadora.

En el dominio de la práctica es aconsejable —según apunta Thierry Foucart (5(p. 16))—, que el orden de la matriz a **simetrizar** debería ser al menos de 16 para que se pudieran extraer conclusiones fidedignas de sus datos empíricos.

Dado que, en situaciones prácticas de interés, la matriz de datos estará constituida por muchas más de 4 variables cuantitativas y de 15 individuos para su análisis en **componentes principales con variables cuantitativas,** será necesario disponer de un programa actualizado que nos permita el cálculo de los **valores propios y vectores propios** con la mayor precisión posible. Por tal motivo, aconsejamos a los informáticos que implementen en el programa para el cálculo de **los ejes principales, los factores principales y las componentes principales,** los programas más actualizados y precisos para el cálculo de los **valores propios y los vectores propios.**

Desde nuestro punto de vista, les aconsejamos que consulten tres libros básicos: Householder (11), Ciarlet (12(pp. 90-94, 123-131)), Ralston (13), Puy Huarte (14(pp. 320-342, 399-402)).

Insistimos mucho sobre este punto ya que, en ciertas ocasiones, las representaciones gráficas de los puntos —individuos y los puntos— variables en los planos correspondientes van a variar sus posiciones según los programas que se utilicen para el cálculo de los **valores propios** y de los **vectores propios.**

En última instancia, los propios especialistas en **cálculo numérico** han podido constatar que, pequeñas perturbaciones en la matriz de los coeficientes de correlación lineal de Auguste Bravais-Karl Pearson conllevan grandes variaciones en los **valores propios.**

MÉTODO (SEGUNDO EJERCICIO)

Este ejercicio está reservado para los investigadores científicos que se dediquen a las ciencias experimentales con el fin de que se familiaricen con los tres artículos anteriores y con éste y, por lo tanto, puedan terminar con éxito el ejercicio que les propondremos de inmediato, consistente en el cambio de la métrica en R^3. Para conseguir dicho objetivo, les proporcionaremos los elementos básicos que van a necesitar para que puedan llegar con éxito al resultado final con una simple calculadora. La única particularidad, en cuanto al proceso metodológico de este **segundo ejercicio** con respecto al **primero,** es que el cálculo de la **raíz cuadrada de una matriz simétrica definida positiva** $M_{(3,3)}^{\frac{1}{2}}$ ya no es inmediato y, por lo tanto, tendremos que hacer uso tanto de su definición como de sus propiedades (2,15).

Aplicando la definición, a nuestro caso concreto, la matriz $M_{(3,3)}^{\frac{1}{2}}$ se expresa de la siguiente forma

$$M_{(3,3)}^{\frac{1}{2}} = U_{(3,3)}\ L_{(3,3)}^{\frac{1}{2}}\ U_{(3,3)}^{T}$$

Omitimos cómo se calculan las matrices contenidas en $M_{(3,3)}^{\frac{1}{2}}$ dado que no representa ninguna dificultad ni siquiera para cualquier estudiante de una Escuela Técnica Superior de Ingeniería o Facultad de Ciencias.

Para poder constatar unos resultados dados con respecto a los nuestros, se muestran cuales son las tres matrices necesarias para el cálculo de $M_{(3,3)}^{\frac{1}{2}}$

$$U_{(3,3)} = \frac{1}{3}\begin{pmatrix} -2 & 2 & -1 \\ 2 & 1 & -2 \\ 1 & 2 & 2 \end{pmatrix} \quad U_{(3,3)}^{T} = \frac{1}{3}\begin{pmatrix} -2 & 2 & -1 \\ 2 & 1 & 2 \\ -1 & -2 & 2 \end{pmatrix}$$

$$L_{(3,3)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 7 \end{pmatrix}$$

De esta manera se podrá comprobar no sólo que la descomposición ha sido correcta, sino también, calcular —mediante una simple multiplicación de matrices— la matriz $M_{(3,3)}$

$$M_{(3,3)}^{\frac{1}{2}} = \begin{pmatrix} 12+\sqrt{7} & 2\sqrt{7} & 6-2\sqrt{7} \\ 2\sqrt{7} & 6+2\sqrt{7} & 6-2\sqrt{7} \\ 6-\sqrt{7} & 6-4\sqrt{7} & 9+4\sqrt{7} \end{pmatrix}$$

Para no resultar reiterativos, dado que el cálculo de $M_{(3,3)}^{\frac{1}{2}}$ ya está resuelto, con el fin de que los lectores se familiaricen con este innovador proceso metodológico con respecto a los habituales en los libros de **análisis estadístico multidimensional lineal,** se proporciona la suficiente información para una resolución exitosa (con una simple calculadora) de los **ejes principales, los factores principales y las componentes principales.**

Información aportada para aquellos interesados en llevar a cabo el segundo ejercicio propuesto con una simple calculadora:

Valores propios	
$\chi_{(3,4)}^{F} N_{(4,4)}\, \chi_{(4,3)}^{c}\, M_{(3,3)}$	$M_{(3,3)}^{\frac{1}{2}}\, \chi_{(3,4)}^{F}\, N_{(4,4)}\, \chi_{(4,3)}^{c}\, M_{(3,3)}^{\frac{1}{2}}$
32,3677090550	32,3677090550
28,0677997880	28,0677997880
3,2024911574	3,2024911574

Vectores propios		
$\chi^{F}_{(3,4)} N_{(4,4)}\, \chi^{c}_{(4,3)}\ M_{(3,3)}$		
0,6281891820	–0,638861370	0,6436762071
–0,0556028070	–0,746260234	–0,6683905430
0,7760713108	0,186954040	–0,3727398960

Vectores propios		
$M^{\frac{1}{2}}_{(3,3)}\, \chi^{F}_{(3,4)}\ N_{(4,4)}\, \chi^{c}_{(4,3)}\ M^{\frac{1}{2}}_{(3,3)}$		
0,5102113403	0,5930145102	–0,6229110520
–0,0623058390	0,7478565559	0,6609300676
0,8577892344	–0,2984030200	0,4185131622

PROPUESTAS DE EJERCICIOS PARA APLICAR EL ACP A TRES TRIPLETES ESTADÍSTICOS

1. $\left(\left(I_{(n,n)} - \dfrac{1_n 1_n^T}{n}\right) X_{(n,p)},\ I_{(p,p)},\ \dfrac{1}{n}\, I_{(n,n)}\right)$

2. $\left(\left(I_{(n,n)} - \frac{1_n 1_n^T}{n}\right) X_{(n,p)} D_{\frac{1}{s^j}},\ I_{(p,p)},\ \frac{1}{n}\, I_{(n,n)}\right) \qquad j = 1, \dots, p$

3. $\left(\chi^{c}_{(n,p)}, M_{(p,p)},\ N_{(n,n)}\right)$

donde,

$I_{(n,n)}$: es la matriz identidad de orden n.

1_n: es un vector columna que contiene n unos.

1_n^T: es el vector transpuesto de 1_n.

$X_{(n,p)}$: es la matriz de datos constituida por n filas y p columnas.

$I_{(p,p)}$: es la matriz identidad de orden p.

$D_{\frac{1}{s^j}}$: es una matriz diagonal de orden p, donde las s^j son las desviaciones típicas de cada una de las p variables cuantitativas.

$$\chi^{c}_{(n,p)} = \left(I_{(n,n)} - 1_n 1_n^T\ N_{(n,n)}\right) X_{(n,p)}$$

$$\chi^{F}_{(p,n)} = X_{(p,n)} \left(I_{(n,n)} - 1_n 1_n^T\ N_{(n,n)}\right)$$

$M_{(p,p)}$: es la **métrica** introducida en R^p. Los responsables de introducir esta **métrica** son los propios investigadores en el dominio de las ciencias experimentales.

$N_{(n,n)}$: es la **métrica** introducida en R^n. Los responsables de introducir esta **métrica** son los mismos que los de introducir la **métrica** $M_{(p,p)}$ en R^p.

Una vez definidos los tres **tripletes estadísticos,** ya se está en condiciones para calcular los **ejes principales, los factores principales y las componentes principales** y, asimismo, para poder percatarse —al calcularlos— de las grandes discrepancias existentes según la elección del **triplete estadístico.**

Con el fin de facilitar la familiarización con los contenidos de los **cuatro artículos,** hemos creído conveniente presentar cinco ejercicios para que puedan constatar que los resultados que obtengan coinciden con los contenidos en dichos libros —al menos— para el primer **triplete estadístico.** De estos cuatro textos, dos de ellos escritos por los profesores Edwind Diday, Jacques Lemaire, Jean Pouget, Françoise Testu (16(pp. 194-196)) y por el profesor Christian Bialès (7(pp. 224-234)) presentan dos tablas de datos que, aunque no sean estrictamente iguales, presentan una característica común y otra de interés en cuanto a las **componentes principales** que mostramos a continuación:

Edwind Diday (1982) Chistian Bialès (1988)

$$X^{(1)}_{(6,3)} = \begin{array}{c} \begin{matrix} \vec{x}^1 & \vec{x}^2 & \vec{x}^3 \end{matrix} \\ \begin{pmatrix} 8 & 1 & 0 \\ 4 & 6 & 5 \\ 6 & 8 & 7 \\ 10 & 4 & 7 \\ 8 & 2 & 5 \\ 0 & 3 & 6 \end{pmatrix} \end{array} \qquad X^{(2)}_{(6,3)} = \begin{array}{c} \begin{matrix} \vec{x}^1 & \vec{x}^2 & \vec{x}^3 \end{matrix} \\ \begin{pmatrix} 4 & 5 & 6 \\ 6 & 7 & 8 \\ 8 & 0 & 1 \\ 0 & 6 & 3 \\ 10 & 7 & 4 \\ 8 & 5 & 2 \end{pmatrix} \end{array}$$

OBSERVACIONES RESPECTO AL ACP CONTENIDO EN LOS LIBROS DE EDWIND DIDAY ET COLLABORATEURS Y DE CHRISTIAN BIALÈS

1.ª) Edwind Diday *et collaborateurs* (3) (16(pp. 194-196)) consideran que los vectores columnas se comportan como variables cuantitativas y, por lo tanto, es totalmente correcto aplicar un **ACP del triplete estadístico**

$$\left(\left(I_{(6,6)} - \frac{1_6 1_6^T}{6}\right) X^{(1)}_{(6,3)},\ I_{(3,3)},\ \frac{1}{6}\, I_{(6,6)}\right)$$

aunque ellos lo hayan realizado de diferente manera que la nuestra.

2.ª) Christian Bialès en (7(pp. 224-234)) presenta la tabla de datos como una **tabla de notas** y, por lo tanto, según apunta el profesor Jean-Paul Benzécri *et collaborateurs* (34) (4(pp. 25-27)), en lugar de haber aplicado un **análisis en componentes principales,** obviando las **métricas,** se tendría que haber aplicado un **análisis de correspondencias (AFC).**

La aplicación del **ACP** del **triplete estadístico:**

$$\left(\left(I_{(6,6)} - \frac{1_6 1_6^T}{6}\right) X_{(6,3)}^{(2)},\ I_{(3,3)},\ \frac{1}{6}\, I_{(6,6)}\right)$$

lleva a los mismos resultados que muestra Christian Bialès en (7 (pp. 224-234)) obviando las **métricas.**

No obstante, a título ilustrativo hemos retenido el **ACP** contenido en el libro de Christian Bialès (7(pp. 224-234)) por considerarlo didáctico, aunque haga caso omiso de las **métricas** y además, la utilización de ambos textos puede aportarles o —en su caso— recordarles algunas informaciones estadísticas supuestamente no aprendidas o entendidas —total o parcialmente— o, en su caso, presuntamente olvidadas por diferentes motivos.

Es conveniente aconsejarles que así como en el **AFC** los puntos-frecuencia relativa (filas) y los puntos frecuencias relativas (columnas) juegan un papel simétrico, y por tal propiedad se pueden representar dichos puntos en un mismo plano factorial, el **ACP** no goza de esta propiedad y, por lo tanto, los puntos-filas tienen que representarse en otro plano factorial que los puntos-columnas.

3.ª) La característica común que presentan los dos **tripletes estadísticos** es que:

$$V_{(3,3)}^{(1)}\ I_{(3,3)} = V_{(3,3)}^{(2)}\ I_{(3,3)}$$

$$V_{(3,3)}^{(1)}\ I_{(3,3)} = \left(\left(I_{(6,6)} - \frac{1_6 1_6^T}{6}\right) X_{(6,3)}^{(1)}\right)^T \frac{1}{6}\, I_{(6,6)} \left(I_{(6,6)} - \frac{1_6 1_6^T}{6}\right) X_{(6,3)}^{(1)} I_{(3,3)}$$

$$V_{(3,3)}^{(2)}\ I_{(3,3)} = \left(\left(I_{(6,6)} - \frac{1_6 1_6^T}{6}\right) X_{(6,3)}^{(2)}\right)^T \frac{1}{6}\, I_{(6,6)} \left(I_{(6,6)} - \frac{1_6 1_6^T}{6}\right) X_{(6,3)}^{(2)} I_{(3,3)}$$

El hecho de que $V_{(3,3)}^{(1)}\ I_{(3,3)}$ sea igual $V_{(3,3)}^{(2)}\ I_{(3,3)}$ conlleva que los **ejes principales y los factores principales** obtenidos por la aplicación de cada uno de los **ACP** ya aludidos sean exactamente los mismos.

4.ª) En cuanto a los elementos contenidos en los vectores

$\vec{c}^{\,j}$ $j = 1, 2, 3$ del **primer triplete estadístico,**

podrán observar, que son los mismos que los elementos contenidos en los vectores

$\vec{c}^{\,j}$ $j = 1, 2, 3$ del **segundo triplete estadístico.**

La única particularidad es, que el orden de tales elementos no es el mismo.

Una vez asimilado lo ya expuesto sobre los **elementos básicos en el análisis en componentes principales** (1, 2, 6), les invitamos —haciendo uso de una simple calculadora— a que lleguen al resultado final; es decir, a que calculen los **ejes principales, los factores principales y las componentes principales.** Con el fin de permitir la propia constatación de la corrección de los resultados correctos, les presentamos los resultados finales a los que nosotros hemos llegado, los cuales, por supuesto, coinciden con los presentados por los profesores Edwind Diday *et collaborateurs* y con los expuestos por el profesor Christian Bialès.

RESULTADOS OBTENIDOS POR NOSOTROS:
ACP DEL TRIPLETE ESTADÍSTICO

$$\left(\left(I_{(6,6)} - \frac{1_6 1_6^T}{6}\right) X^{(1)}_{(6,3)}, I_{(3,3)}, \frac{1}{6} I_{(6,6)}\right)$$

Ejes principales:

$$\vec{u}^1 = \frac{1}{\sqrt{6}}\begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix} \qquad \vec{u}^2 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \qquad \vec{u}^3 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$

Factores principales:

$$\vec{v}^1 = \frac{1}{\sqrt{6}}\begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix} \qquad \vec{v}^2 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \qquad \vec{v}^3 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$

Componentes principales:

$$\vec{c}^1 = \sqrt{6}\begin{pmatrix} 2 \\ -1 \\ -1 \\ 1 \\ 1 \\ 2 \end{pmatrix} \qquad \vec{c}^2 = \sqrt{6}\begin{pmatrix} -2 \\ 0 \\ 2 \\ 2 \\ 0 \\ -2 \end{pmatrix} \qquad \vec{c}^3 = \sqrt{2}\begin{pmatrix} 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{pmatrix}$$

ACP del triplete estadístico:

$$\left(\left(I_{(6,6)} - \frac{1_6 1_6^T}{6}\right) X^{(2)}_{(6,3)}, I_{(3,3)}, \frac{1}{6} I_{(6,6)}\right)$$

Nuevamente, para no resultar reiterativos, omitimos **los ejes principales y los factores principales** por ser exactamente los mismos que en el primer caso. Así pues, simplemente haremos alusión a las **componentes principales:**

$$\vec{c}^{1} = \sqrt{6}\begin{pmatrix}-1\\-1\\2\\-2\\1\\1\end{pmatrix} \qquad \vec{c}^{2} = \sqrt{6}\begin{pmatrix}0\\2\\-2\\-2\\2\\0\end{pmatrix} \qquad \vec{c}^{3} = \sqrt{2}\begin{pmatrix}-1\\-1\\-1\\1\\1\\1\end{pmatrix}$$

OBSERVACIONES CRÍTICAS RESPECTO AL ACP CONTENIDO EN EL LIBRO DE JEAN-PIERRE VEDRINE

Aunque, por un lado, las dimensiones de la tabla de datos que presenta el profesor Jean-Pierre Vedrine en (17 (pp. 133-142)) no son suficientes para que sus resultados sean interpretables mediante un **ACP** y, además, por otra parte, su tabla de datos, en el supuesto de que hubiera contenido más de 4 variables cuantitativas y más de 15 individuos, no obstante, se adapta mejor el **AFC (análisis de correspondencias)** haciendo alusión a sus propias **métricas** (10).

Simplemente, a título ilustrativo, las dos formas que presenta se pueden expresar de la siguiente manera:

Mientras que la primera es exactamente lo mismo que aplicar un **ACP** al **triplete estadístico,**

$$\left(\left(I_{(15,15)} - \frac{1_{15}1_{15}^{T}}{15}\right) X_{(15,3)}, I_{(3,3)}, \ \frac{1}{15}\, I_{(15,15)}\right)$$

la segunda es, exactamente lo mismo que aplicar un **ACP** al **triplete estadístico,**

$$\left(\left(I_{(15,15)} - \frac{1_{15}1_{15}^{T}}{15}\right) X_{(15,3)} D_{\frac{1}{s^{j}}}, I_{(3,3)}, \ \frac{1}{15}\, I_{(15,15)}\right) \qquad j = 1, 2, 3$$

Si los investigadores en ciencias experimentales disponen de la documentación ya mencionada, así como de los programas para el cálculo de los **valores propios y vectores propios,** ellos mismos podrán darse cuenta de que los resultados que obtengan referentes a **los ejes principales, los factores principales y las componentes principales** tienen que ser casi iguales a los obtenidos por Jean-Pierre Vedrine (17 (pp. 133-142)), siempre y cuando hayan utilizado los mismos programas para el cálculo de los **valores propios y los vectores propios.**

OBSERVACIONES RESPECTO AL ACP CONTENIDO EN EL LIBRO DE CHRISTIAN GOUJET Y CLAIRE NICOLAS

Aunque la tabla de datos que se contempla en el libro de Christian Goujet y Claire Nicolas (18 (pp. 259-263)) es una **tabla de notas** en lugar de aplicar un **AFC (análisis factorial de correspondencias)** aplican un **ACP** sin hacer alusión a las **métricas**.

Tal como lo aplican es como si se aplicase un **ACP** del **triplete estadístico,**

$$\left(\left(I_{(5,5)} - \frac{1_5 1_5^T}{5}\right) X_{(5,3)}, I_{(3,3)}, \ \frac{1}{5} I_{(5,5)}\right)$$

Siguiendo fielmente el artículo (6) llegarán a los mismos resultados que apuntan los profesores Christian Goujet y Claire Nicolas (18 (pp. 259-263)).

Para poder familiarizarse aún más con la aplicación del **ACP (métricas no usuales)**, nosotros proponemos que se obtengan **los ejes principales, los factores principales y las componentes principales,** teniendo en cuenta el siguiente **triplete estadístico**:

$$\left(\chi^{c}_{(5,3)}, \ \ M_{(3,3)}, N_{(5,5)}\right)$$

donde:

$$X_{(5,3)} = \begin{pmatrix} 6 & 3 & 8 \\ 10 & 18 & 12 \\ 19 & 13 & 15 \\ 19 & 13 & 18 \\ 1 & 7 & 1 \end{pmatrix}$$

$$M_{(3,3)} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad N_{(5,5)} = \frac{1}{15} \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Nota de interés: la resolución de este caso concreto no lo hemos realizado, por lo tanto, no se podrán constatar si los resultados que obtengan son fidedignos a nivel de cálculo. Si se logra llegar con éxito a los resultados se relevarán —sin duda—, la existencia de grandes discrepancias entre los obtenidos mediante el **triplete estadístico (métricas usuales en el ACP)** y **el triplete estadístico (métricas no usuales en el ACP).**

UN BREVE COMENTARIO SOBRE LAS MÉTRICAS USUALES Y LAS NO USUALES CONTENIDAS EN LOS TRIPLETES ESTADÍSTICOS

1. Métricas usuales en el ACP: tripletes estadísticos

1.1.

$$\left(\left(I_{(n,n)} - \frac{1_n 1_n^T}{n}\right) X_{(n,p)}, \ I_{(p,p)}, \ \frac{1}{n} I_{(n,n)}\right)$$

1.2.

$$\left(\left(I_{(n,n)} - \frac{1_n 1_n^T}{n}\right) X_{(n,p)} D_{\frac{1}{s^j}},\ I_{(p,p)},\ \frac{1}{n} I_{(n,n)}\right) \quad j = 1, \ldots, p$$

El profesor Gibert Saporta (19 (p. 171)) dice: *«Cependant, les résultats obtenus (1.1) ne sont pas invariants si on change linéairement l'unité de mesure des variables. Les covariances sont multipliées par un facteurs k, la variance par un facteur* $\boldsymbol{k^2}$ *si on choisit une unité de mesure k fois plus petite pour une variable.*

Le choix de (1.2), est le plus communément fait, et a pour conséquence de rendre la distance entre individus invariants pour transformation linéaire séparée de chaque variable et de s'affranchir des unités de mesure ce qui est particulièrement intéressant lorque les variables sont hétérogenes».

En la mayoría de los libros escritos en castellano de Análisis Estadístico Multidimensional Lineal —en sí mismo— y aplicado a la Investigación del Mercado, se contempla el ACP sin hacer alusión a las métricas.

2. Métricas no usuales en el ACP: triplete estadístico

$$\left(\chi^{c}_{(n,p)},\ M_{(p,p)}, N_{(n,n)}\right)$$

Desde un punto de vista, tanto teórico como práctico, tal como lo hemos expuesto en nuestros cuatro artículos, que nosotros sepamos, ni se contempla en ningún libro comercializado escrito en castellano de **Análisis Estadístico Multidimensional Lineal** en sí mismo, ni aplicado a la Investigación del Mercado.

RECAPITULACIONES SOBRE EL ACP

En 1901, un año más tarde de que Karl Pearson introdujera el test de la Ji-cuadrado, haciendo uso de la ley Ji-cuadrado de Helmert (1875), ocurrieron dos acontecimientos de interés:

1.º El profesor Karl Pearson, nacido en Londres, introdujo el **ACP** (20).

2.º El profesor Karl Pearson, con la ayuda del profesor Francis Galton, nacido en Inglaterra, fundó y dirigió la revista **«Biométrica».**

Setenta y cinco años más tarde, los profesores Francis Cailliez y J-P. Pagès, con la contribución de 18 matemáticos de renombre internacional, escribieron el primer libro que contenía, no sólo los **elementos algebraicos** para entender y comprender los **análisis factoriales** utilizando los **esquemas de dualidad, las métricas,** sino también los **algoritmos de clasificación jerárquica y no jerárquica** (21).

En nuestro caso concreto, en este libro se contempla a parte de la componente teórica del **ACP** (21 (pp. 221-250)), la práctica (21 (pp. 271-302)).

Además de estas dos referencias que consideramos básicas en el **ACP,** es de interés mirar con atención el capítulo titulado: «*Vers l'analyse de tableaux à plus de deux dimensions*» (21 (pp. 467-492)). Dicho capítulo muestra los elementos básicos de ***la méthode Statis*** introducida por Yves Escoufier y contemplada en (22 (pp. 110-116)). No está demás recordar que la evolución de la programación de ***la méthode Statis*** ha ido mejorándose a lo largo de los años, y ahora es un programa de los muchos que contiene —al menos— el paquete de programas **SPAD.**

RECAPITULACIONES SOBRE EL AFDL

En 1936, el profesor Sir Ronald Aylmer Fisher, nacido en Inglaterra, introdujo el **análisis factorial discriminante lineal (AFDL)** (23).

Hacemos alusión a este análisis, no por simple capricho, sino porque dicho análisis es un caso particular del **ACP.**

Desde nuestro punto de vista, de entre toda la documentación escrita sobre Análisis de Datos como libros y apuntes de curso, pensamos que en los apuntes de curso aportados por el profesor Yves Escoufier (22 (pp. 88-98)) donde se encuentra reflejado la relación entre el **ACP** y el **análisis discriminante lineal** con el máximo rigor científico, haciendo uso de los **esquemas de dualidad** y de las **métricas.**

CONCLUSIONES

Cada profesional debe conocer perfectamente sus propios datos empíricos. Dichos datos pueden ser extraídos:

1.º Por la propia información que les proporciona la naturaleza.

2.º Haciendo uso de los anuarios estadísticos cuyo contenido haya sido no sólo actualizado sino contrastado.

3.º Utilización de técnicas instrumentales de precisión.

Hemos de advertir que una incorrecta introducción de las **métricas** en R^p y en R^n, en el **ACP** puede conducir a los investigadores científicos a transmitir en sus publicaciones —aunque éstas sean de índice de impacto, pues (esto no garantiza la calidad del artículo)— conclusiones erróneas y, por lo tanto, los resultados de sus investigaciones serán estériles conduciendo a una cadena de despropósitos.

Así como en el **análisis en componentes principales (ACP),** las métricas no están definidas *a priori,* en otros análisis de datos sí lo están **(AFDL, AFC).**

Finalmente, esperamos haber realizado una contribución útil a los profesionales interesados o que trabajen en este tema. Pensamos que sería necesario introducir un mayor número de elementos, así como la realización —por parte de los programadores—, de un programa con el fin de poner de manifiesto los puntos —individuos y los puntos— variables correspondientes para visualizar dichos puntos.

Recordaremos que cualquier **Análisis de Datos** carente de representaciones gráficas, dificulta grandemente la interpretación de los resultados constituyendo estos, por lo tanto, una herramienta de gran ayuda para la consecución de resultados certeros en cuanto a la interpretación de los mismos.

BIBLIOGRAFÍA

1. Díaz-Llanos, Fco. J.; Escoufier, Yves.; Cermeño, C., Grau, L. F. (2011): «Elementos básicos en el análisis en componentes principales (ACP) [Primera parte: teoría]». *Anales de la Real Academia de Doctores de España.* Vol. 15, n.º 1, pp. 51-75.
2. Díaz-Llanos, Fco. J.; Escoufier, Yves.; Cermeño, C., Grau, L. F. (2012): «Elementos básicos en el análisis en componentes principales (ACP) [Tercera parte: teoría (métricas distintas de las usuales)]». *Anales de la Real Academia de Doctores de España.* Vol. 16, n.º 1, pp. 51-64.
3. Bon, J., Grégory, P. (1986): *Techniques marketing.* Vuibert Gestion. Paris.
4. Benzécri, J.-P. *et collaborateurs* (1973): *L'Analyse des Données. 2. L'Analyse des Correspondances.* Dunod.
5. Foucart, Th. (1997): *L'Analyse des Données. Mode d'emploi.* Presses Universitaires de Rennes.
6. Díaz-Llanos, Fco. J.; Escoufier, Yves; Cermeño, C., Grau, L. F. (2011): «Elementos básicos en el análisis en componentes principales (ACP) [Segunda parte: práctica (métricas usuales en el ACP)]». *Anales de la Real Academia de Doctores de España.* Vol. 15, n.º 2, pp. 71-90.
7. Bialès, Chr. (1998): *L'analyse statistique des données. L'outil statistique appliqué au marketing et à la gestion.* Chotard et associés éditeurs.
8. Audroing, J.-F. (1977): «Mathématiques Linéaires». *Economica.*
9. Guerrien, B. (1980): «Álgèbre Linéaire pour economistes. Rappels de cours et exercices corrigés». *Economica.*
10. Lebart, L.; Morineau, A., Piron, M. (1995): *Statistique Exploratoire Multidimensionnelle.* Dunod.
11. Householder, A. S. (1953): *Principles of Numerical Analysis.* McGraw-Hill, New York.
12. Ciarlet, P. G. (1985): *Introduction à l'analyse numerique matricielle et à l'optimisation.* Ed. Masson.
13. Ralston, A. (1978): *Introducción al análisis numérico.* Ed. Limusa. México.
14. Puy Huarte, J. (1983): Cálculo numérico. Apuntes de Cátedra de Matemática III de la ETSI de Caminos, Canales y Puertos.
15. Denis, J.-B. y colaborador (1976): *Apuntes redactados con ocasión del cursillo impartido en la Sección de Proceso de Datos.* INIA (MAPA). Madrid.
16. Diday, E.; Lemaire, J.; Pouget, J., Testu, F. (1982): *Éléments d'analyse des données.* Dunod.
17. Vendrine, J.-P. (1991): *Le traiment des données en marketing en 10 questions, 15 applications, 27 exemples et exercices commentés et corrigés. Les editions d'organisation.*
18. Goujet, Ch., Nicolas, Cl. (1989): *Mathématiques apliqués à la gestion. Decisión en avenir incertain. Initiation à la recherche opérationnelle.* 4 édition. Masson.
19. Saporta, G. (1990): *Probabilités. Analyse des données et Statistique.* Éditions Technip.
20. Pearson, K. (1901): «On lines and planes of closest fit to systems of points in space». *Phil. Mag* 20; 6ème serie; pp. 557-572.
21. Cailliez, F., Pagès, J.-P. (1976): *Introduction à l'Analyse des Données.* SMASH.
22. Escoufier, Yves (1979): *Cours d'Analyse des Données.* RT. 7901. CRIG. Av. d'Occitanie 34075 Montpellier Cedex.
23. Fisher, R. A. (1936): «The use of multiple messurements in taxonomie problems». *Ann of Engenics,* 7, pp. 179-188.