

ELEMENTOS BÁSICOS EN EL ANÁLISIS EN COMPONENTES PRINCIPALES (ACP)

[Segunda parte: práctica (métricas usuales en el ACP)]

DOCTOR D. FCO. JAVIER DÍAZ LLANOS SAINZ-CALLEJA

*Académico de Número de la Real Academia de
Doctores de España. Sección Ingeniería. N.º 118.
Medalla al Mérito Doctoral. Categoría de Plata.*

M. YVES ESCOUFIER

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

DOCTORA DÑA. M.ª DEL CARMEN CERMEÑO CARRASCO

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

DOCTOR D. LUIS FELIPE GRAU SEGURA

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

RESUMEN

El objetivo de este artículo, es el de mostrar -paso a paso- partiendo de una tabla de datos extraída del libro de los profesores J. Bon y P. Grégory (1 (pág. 70)), como se obtienen los elementos básicos de un **análisis en componentes principales**: los **ejes principales**, los **factores principales** y las **componentes principales**.

Veremos —con agrado— que, tal como desarrollan J. Bon y P. Grégory el ACP en su libro (1 (págs. 70-83)), éste es un caso particular del **ACP à la française**.

Aunque los autores de dicho libro hayan hecho caso omiso de las **métricas**, el ejercicio que exponen —a nivel didáctico—, lo consideramos muy necesario, —aunque no suficiente—, para los investigadores que se dediquen a otra disciplina que no sea, el **análisis estadístico multidimensional lineal**.

A título ilustrativo veremos que, los resultados que muestran J. Bon y P. Grégory (1 (págs. 70-83)), son los mismos a los que llegamos nosotros a partir de las **métricas** introducidas en

$$R^3 \text{ y } R^4$$

Estas métricas son:

$$M_{(3,3)} \text{ y } N_{(4,4)}$$

Que en el caso contenido en (1) son:

$$M_{(3,3)} = I_{(3,3)} \text{ y } N_{(4,4)} = \frac{1}{4}I_{(4,4)}$$

Palabras claves: valores propios, vectores propios, cuadrado de la norma de un vector en el sentido de una **métrica**, los **ejes principales**, los **factores principales** y las **componentes principales**.

Introducción:

El hecho de que, nos centremos —exclusivamente— en la exposición didáctica, haciendo uso de una tabla de datos de pequeñas dimensiones (4 x 3), no es por puro capricho sino porque, el cálculo de los **ejes principales**, los **factores principales** y los **componentes principales** [coordenadas de los puntos-fila (individuos) sobre los **factores principales** que —en nuestro caso concreto— serán los **ejes principales**, tal como veremos —más detenidamente— a lo largo de este artículo de carácter didáctico], está contemplado de forma -aunque restrictiva-, en cuanto a las **métricas** que se deben introducir en

$$R^p \text{ y } R^n$$

en multitud de libros de texto de **análisis estadístico multidimensional lineal**, sobre todo escritos en **castellano**.

Para poder comparar los resultados que obtendremos de la aplicación de la parte teórica con respecto a un ejercicio realizado, haciendo caso omiso de las **métricas**, hemos considerado oportuno, utilizar el contemplado por los profesores J. Bon y P. Grégory (1 (págs. 70-83)). La presentación del estudio del **análisis en componentes principales**, —por parte de dichos profesores— pensamos que, no obstante, aunque no sea de un rigor científico depurado (mediante una tabla de datos que contiene 4 individuos y tres variables cuantitativas), es extraordinariamente más didáctico que, la del resto de los libros de texto que hemos consultado que obvian las **métricas**.

La tabla de datos contenida en (1 (pág. 70)) al ser de dimensiones pequeñas, permitirá seguir el ejercicio —sin dificultad— con una simple calculadora.

El ejercicio que vamos a mostrar, tiene tan sólo carácter ilustrativo, pues, desde el punto de vista de la interpretación, carece de valor por lo ya expuesto por el profesor Thierry Foucart (2 (pág. 16)).

Dado que, la tabla de datos que vamos a tratar, haciendo uso del **esquema de dualidad**, es la misma que la contenida en (1 (pág. 70)), los investigadores podrán percatarse de que los resultados tanto de los **ejes principales, los factores principales y las componentes principales** son los mismos a los que llegamos nosotros mediante la utilización de las **métricas**:

$$M_{(3,3)} = I_{(3,3)} \quad \text{y} \quad N_{(4,4)} = \frac{1}{4}I_{(4,4)}$$

Sin lugar a dudas verán que, el procedimiento expuesto por J. Bon y P. Grégory (1 (págs. 70-83)), es un caso particular del contemplado por nosotros en la (Primera parte: teoría) y que desarrollaremos en esta [Segunda parte: práctica en el ACP (métricas usuales)], de la forma más didáctica posible.

Nos parece lamentable que, aunque nuestro procedimiento se conozca formalmente —al menos, desde el año 1972(3)— bajo el formato de un curso y el de un libro en 1976(4), aún no aparezca reflejado en la mayoría de los textos de **análisis estadístico multidimensional lineal** escritos en **castellano**.

No está demás recordar que, la idea que tienen numerosos profesores españoles sobre el **análisis en componentes principales**, es totalmente **errónea** pues, tan sólo se deciden a diagonalizar 1) la matriz de varianzas-covarianzas o bien 2) la de correlaciones de Auguste Bravais — Karl Pearson.

Nosotros nos inclinamos a pensar que, sistemáticamente se deciden por una de estas dos opciones ya que, son las que se contemplan en los paquetes de programas comercializados de **análisis estadístico multidimensional**. Así que, una vez leído nuestro artículo, les invitamos a que introduzcan la opción de **métricas** en su programa de ACP con el fin de subsanar los errores cometidos por la utilización sistemática de éstas y, en especial, de la primera en donde las **métricas** en

$$R^p \text{ y } R^n$$

$$\text{son: } M_{(p,p)} = I_{(p,p)} \quad \text{y} \quad N_{(n,n)} = \frac{1}{n}I_{(n,n)}, \text{ respectivamente}$$

La única manera de que puedan observar —los mismos investigadores— las diferencias existentes —en cuanto a los resultados obtenidos por los procedimientos que actualmente utilizan— y, utilizando las **métricas** es, que traten sus datos empíricos por ambos métodos y comparen cuales son los resultados que para ellos tienen mayor grado de fiabilidad.

En (5) el lector podrá informarse —más detenidamente— sobre la necesidad de las **métricas** y, de como pueden construir los propios profesionales las **métricas** que más se adapten a sus propios datos empíricos ya que, son los responsables de construirlas.

Material:

El **material** que necesitamos consiste en:

1. Disponer de una buena bibliografía tanto a nivel práctico como teórico.
 2. Una matriz de datos originales que contenga datos empíricos, obtenidos tanto de 1) anuarios que tengan un nivel de credibilidad aceptable como 2) análisis realizados mediante dispositivos de medida los más precisos posible.
 3. Un programa comercializado que disponga de un **ACP** en el cual se contemple la opción de las **métricas** o, en el caso de no estar comercializado, implementarlas en un **ACP** que no las contenga.
1. Disponer de una buena bibliografía tanto a nivel práctico como teórico:
 - 1.1. Aconsejamos que, de entre la multitud de libros de nivel práctico existentes en el mercado, consulten —preferentemente— los que mostraremos en orden cronológico: 1977(6), 1985(7), 1986(8), 1997(2), 1999(9) y 2000(10).
 - 1.2. Igualmente, aconsejamos que, de entre la multitud de libros de nivel teórico existentes en el mercado, consulten —preferentemente— los que mostraremos en orden cronológico: 1973(11), 1976(4), 1977(12), 1978(13), 1981(14, 15), 1982(16), 1984 (17), 1989(18,19), 1990(20,21), 1992(22), 1993(23), 1995(24), 1996(25), 1999(26) y 2007(27).

2. Matriz de datos:

Dado que, el objetivo de este artículo es el de exponer —de la forma más didáctica posible— tan sólo cómo se construyen los **ejes principales, los factores principales y las componentes principales**, partiremos de una matriz de datos de pequeñas dimensiones. Esto les permitirá a los profesionales que no se dediquen a la disciplina del **análisis estadístico multidimensional lineal**, seguir —fácilmente— el ejercicio ilustrativo que exponemos en este artículo. La matriz de datos ha sido extraída del libro de J. Bon y P. Grégory (1 (pág. 70)) y se mostrará en el apartado de **método**.

No está demás recordar que, cuando deseemos aplicar un **ACP** a una matriz de datos ésta, previamente debe de centrarse, en nuestro caso, por columnas.

3. Un programa comercializado que disponga de un **ACP en el cual se incluya la opción de métricas en:**

$$R^p \text{ y } R^n$$

Como lo apuntado anteriormente, nos parece inadmisibles que, aunque la aplicación de las **métricas** se conozca en forma de curso, desde el año 1972(3), aún en el año 2011 no exista ningún paquete de programas comercializado, de **análisis estadístico multidimensional lineal**, donde esté implementado la opción de **métricas** en un **ACP**.

Así pues, en el caso hipotético de que no se disponga de un **ACP** en el cual esté incluida la opción de **métricas**, nosotros les recomendamos que acudan a un informático que sin duda alguna, podrá implementarlas con ayuda tanto de la parte teórica como de la práctica.

Método:

En este apartado expondremos de la forma más didáctica posible las etapas a seguir para llegar a los **ejes principales**, los **factores principales** y las **componentes principales** en un **análisis en componentes principales à la française**.

Primera etapa: matriz de datos originales

Los tres vectores columna que van a formar parte de la **matriz de datos originales** son:

$$\vec{x}^1 = \begin{bmatrix} 5 \\ 1 \\ 7 \\ 7 \end{bmatrix} \quad \vec{x}^2 = \begin{bmatrix} 1 \\ 6 \\ 8 \\ 5 \end{bmatrix} \quad \vec{x}^3 = \begin{bmatrix} 3 \\ 5 \\ 3 \\ 9 \end{bmatrix}$$

Así pues, la **matriz de datos original** es:

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

Esta **matriz de datos** extraída de (1(pág. 70)) presenta las siguientes características:

1.^a Todas las variables cuantitativas tienen la misma media aritmética.

$$\vec{x}^j = 5 \quad j=1, 2, 3$$

2.^a Mientras que, la varianza de la primera variable es la misma que la de la tercera, la de la segunda es ligeramente mayor que la de la primera y de la tercera.

$$s^{2j} = 6 \quad j=1,3$$

$$s^{2j} = 6,5 \quad j=2$$

Segunda etapa: matriz de datos originales centrada por columnas

Así pues, la **matriz de datos original centrada por columnas** es:

$$\left(I_{(4,4)} - \frac{1_4 1_4^T}{4} \right) X_{(4,3)} = \begin{bmatrix} 0 & -4 & -2 \\ -4 & 1 & 0 \\ 2 & 3 & -2 \\ 2 & 0 & 4 \end{bmatrix}$$

Observación: La matriz de datos original centrada por columnas es la que hemos mostrado dado que, en nuestro caso particular se verifica que:

$$p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$$

El significado de los demás elementos de nuestra matriz de datos son los que indicamos a continuación:

$I_{(4,4)}$: es la matriz identidad de orden 4

1_4 : es un vector columna cuyos elementos son todos unos

$$1_4 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

1_4^T : es un vector fila cuyos elementos son todos unos

$$1_4^T = (1 \ 1 \ 1 \ 1)$$

$X_{(4,3)}$: es la matriz de datos original que contiene 4 individuos y 3 variables

Tercera etapa: matriz de datos originales centrada por columnas transpuesta.

$$\left(\left(I_{(4,4)} - \frac{1_4 1_4^T}{4} \right) X_{(4,3)} \right)^T = \begin{bmatrix} 0 & -4 & 2 & 2 \\ -4 & 1 & 3 & 0 \\ -2 & 0 & -2 & 4 \end{bmatrix}$$

Cuarta etapa:

*La métrica introducida en R^3
(espacio de los individuos)*

$$M_{(3 \times 3)} = I_{(3,3)}$$

Esta **métrica** nos va a permitir el cálculo de las distancias entre los puntos-columnas en

$$R^3$$

Quinta etapa:

*La métrica introducida en R^4
(espacio de los individuos)*

$$N_{(4 \times 4)} = \frac{1}{4} I_{(4,4)}$$

Esta métrica nos va a permitir el cálculo de las distancias entre los puntos-filas en

$$R^4$$

Sexta etapa: triplete estadístico asociado a nuestro caso concreto.

$$\left(\left(I_{(4,4)} - \frac{1_4 1_4^T}{4} \right) X_{(4,3)}, I_{(3,3)}, \frac{1}{4} I_{(4,4)} \right)$$

donde,

$$I_{(3,3)} \text{ y } I_{(4,4)}$$

son las matrices identidad de orden 3 y 4, respectivamente.

Séptima etapa: cálculo de:

$$V I_{(3,3)}$$

Dicho operador se calcula, en nuestro caso concreto, de la siguiente manera:

$$V I_{(3,3)} = \left(\left(I_{(4,4)} - \frac{1_4 1_4^T}{4} \right) X_{(4 \times 3)} \right)^T \frac{1}{4} I_{(4,4)} \left(I_{(4,4)} - \frac{1_4 1_4^T}{4} \right) X_{(4 \times 3)} I_{(3,3)}$$

De la mera observación del término de la derecha de esta igualdad se concluye que es la **matriz de varianzas-covarianzas** entre las tres variables cuantitativas.

Sustituyendo dichas matrices por sus términos numéricos tenemos,

$$V I_{(3,3)} = \frac{1}{4} \begin{bmatrix} 0 & -4 & 2 & 2 \\ -4 & 1 & 3 & 0 \\ -4 & 0 & -2 & 4 \end{bmatrix} \begin{bmatrix} 0 & -4 & -2 \\ -4 & 1 & 0 \\ 2 & 3 & -2 \\ 2 & 0 & 4 \end{bmatrix}$$

Finalmente, multiplicando estas dos **matrices** obtenemos la **matriz de varianzas-covarianzas** entre las tres variables cuantitativas.

$$V I_{(3,3)} = \frac{1}{4} \begin{bmatrix} 24 & 2 & 4 \\ 2 & 26 & 2 \\ 4 & 2 & 24 \end{bmatrix} = \begin{bmatrix} 6.0 & 0.5 & 1.0 \\ 0.5 & 6.5 & 0.5 \\ 1.0 & 0.5 & 6.0 \end{bmatrix}$$

Esta es la misma matriz que se contempla en (1(pág. 70)).

Octava etapa: cálculo de los valores propios de:

$$V I_{(3,3)}$$

Antes de proceder a mostrar las fases para el cálculo de los valores propios para aquellos futuros investigadores e investigadores que lo deseen, no está demás advertirles de que, la parte teórica para el cálculo de los mismos la pueden encontrar -de forma precisa- en el libro de los profesores P. Lascaux y R. Theodor (28(págs. 43-78)).

Fases para el cálculo de los valores propios:

$$\lambda_j, j=1, 2, 3$$

$$\det(V I_{(3,3)} - \lambda I_{(3,3)}) = 0$$

A continuación mostramos, **paso a paso**, las **tres fases** para el cálculo de los **valores propios**.

Primera fase: cálculo de la matriz

$$(V I_{(3,3)} - \lambda I_{(3,3)})$$

Esta matriz la podemos expresar más —detalladamente— de la siguiente manera:

$$\left(\left(\left(I_{(4,4)} - \frac{1_4 1_4^T}{4} \right) X_{(4 \times 3)} \right)^T I_{(4,4)} \left(I_{(4,4)} - \frac{1_4 1_4^T}{4} \right) X_{(4 \times 3)} \frac{1}{4} - \lambda I_{(3,3)} \right)$$

Sustituyendo las matrices contenidas en esta expresión por sus valores numéricos tenemos,

$$\begin{bmatrix} 6.000 & 0.500 & 1.000 \\ 0.500 & 6.000 & 0.500 \\ 1.000 & 0.500 & 6.000 \end{bmatrix} - \lambda \begin{bmatrix} 1.000 & 0.000 & 0.000 \\ 0.000 & 1.000 & 0.000 \\ 0.000 & 0.000 & 1.000 \end{bmatrix}$$

y, finalmente, restando ambas matrices llegamos al resultado final

$$\begin{bmatrix} 6.0 - \lambda & 0.500 & 1.000 \\ 0.500 & 6.5 - \lambda & 0.500 \\ 1.000 & 0.500 & 6.0 - \lambda \end{bmatrix}$$

A continuación veíamos que es diagonalizable dado que la multiplicidad geométrica de cada valor propio es igual a la multiplicidad algebraica. No está demás recordar que, la multiplicidad geométrica es la dimensión del sub-espacio propio asociado a .

Segunda fase: cálculo del polinomio característico de

$$VI_{(3,3)}$$

Se llama **polinomio característico** de una matriz

$$VI_{(3,3)}$$

al polinomio de grado 3

$$P_{VI_{(3,3)}}(\lambda) = \det(VI_{(3,3)} - \lambda I_{(3,3)})$$

Desarrollando el **determinante de esta matriz** llegamos –sin dificultad– al **polinomio característico** de grado 3

$$\lambda^3 - 18.5\lambda^2 + 112.5\lambda - 225.0$$

Tercera fase: cálculo de las raíces del polinomio característico.

Las raíces del **polinomio característico**, se obtienen igualando a cero dicho **polinomio característico**

$$\lambda^3 - 18.5\lambda^2 + 112.5\lambda - 225.0 = 0$$

Dichas raíces son los **valores propios** de la matriz

$$VI_{(3,3)}$$

Aplicando la regla de Ruffini obtenemos —sin dificultad— las tres raíces del polinomio característico; es decir, los tres **valores propios** de:

$$VI_{(3,3)}$$

(en nuestro caso particular, los tres valores propios de la **matriz de varianzas-covarianzas** entre las tres variables cuantitativas).

Puesto que la matriz

$VI_{(3,3)}$ es simétrica real y definida positiva

los 3 valores propios serán positivos y distintos.

Dichos valores propios los mostramos en orden decreciente:

$$\lambda_1 = 7,5$$

$$\lambda_2 = 6,0$$

$$\lambda_3 = 5,0$$

Indudablemente, los resultados obtenidos son correctos ya que se verifica:

$$Tr(VI_{(3,3)}) = \sum_{j=1}^{j=3} \lambda_j$$
$$Tr(VI_{(3,3)}) = 18,5 \quad \sum_{j=1}^{j=3} \lambda_j = 18,5$$

No está demás recordar que, dichos valores propios son los mismos que se encuentran en (1(pág. 70)).

Novena etapa: cálculo de los vectores propios de

$$VI_{(3,3)} \vec{u}^j = \lambda_j \vec{u}^j I_{(3,3)} - \text{ortonormados}$$

Estos **vectores propios** son los **ejes principales** en un ACP.

Décima etapa: cálculo de los ejes principales.

Para llegar a calcular los **ejes principales** desarrollaremos de forma didáctica las dos sub-etapas que mostramos a continuación:

$$VI_{(3,3)}$$

Primera sub-etapa: cálculo de los sub-espacios vectoriales propios.

Segunda sub-etapa: cálculo de los ejes principales en un ACP.

Primera sub-fase:

Para no resultar reiterativos dado que, el **sub-espacio vectorial propio** asociado a cada uno de los **valores propios** de

$$VI_{(3,3)}$$

se calcula de la misma manera para cada **valor propio**, tan sólo mostraremos —de forma didáctica- el del primero. Así pues, invitamos a los lectores que obtengan los otros dos **sub-espacios vectoriales propios** y con el fin de que sepan que han llegado al resultado **correcto** les indicamos los resultados.

Cálculo del sub-espacio vectorial propio asociado al valor propio:

$$\lambda_1 = 7,5$$

Dicho **sub-espacio vectorial** se obtiene mediante el **sistema de ecuaciones homogéneo**:

$$(VI_{(3,3)} - \lambda_1 I_{(3,3)}) \vec{u}^1 = \vec{0}$$

De lo que se desprende que para el valor propio

$$\lambda_1 = 7,5$$

el sistema de ecuaciones homogéneo que tenemos que resolver es,

$$\begin{bmatrix} -1.500 & 0.500 & 1.000 \\ 0.500 & -1.000 & 0.500 \\ 1.000 & 0.500 & -1.500 \end{bmatrix} \begin{bmatrix} u_1^1 \\ u_2^1 \\ u_3^1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

-Resolución del sistema de ecuaciones homogéneo:

La resolución del sistema de ecuaciones homogéneo que conlleva al sub-espacio vectorial de vectores propios asociado al primer valor propio la exponemos en dos fases:

Primera fase:

Pondremos el sistema de ecuaciones homogéneo bajo forma de tres ecuaciones con tres incógnitas:

$$-1.5u_1^1 + 0.5u_2^1 + 1.0u_3^1 = 0 \quad (1)$$

$$0.5u_1^1 - 1.0u_2^1 + 0.5u_3^1 = 0 \quad (2)$$

$$1.0u_1^1 + 0.5u_2^1 + 1.5u_3^1 = 0 \quad (3)$$

Segunda fase:

De las tres ecuaciones, retenemos la (1) y la (3).

$$-1.5u_1^1 + 0.5u_2^1 + 1.0u_3^1 = 0 \quad (1)$$

$$1.0u_1^1 + 0.5u_2^1 + 1.5u_3^1 = 0 \quad (3)$$

A la vista de estas dos ecuaciones nos interesa restar la (1) de la (3) ya que de esta manera nos desembarazamos de forma directa de la segunda componente.

Así pues, tenemos:

$$u_1^1 = u_3^1$$

A continuación, introduciendo este resultado en (3) se obtiene,

$$u_2^1 = u_3^1$$

Finalmente, haciendo $u_3^1 = k$

$$u_1^1 = k \text{ y } u_2^1 = k$$

Por lo tanto, el **sub-espacio vectorial propio asociado al valor propio**

$$\lambda_1 = 7,5$$

está constituido por **vectores propios** que tengan la siguiente forma:

$$\vec{u}^1 = \begin{bmatrix} k \\ k \\ k \end{bmatrix}$$

-Cálculo del sub-espacio vectorial propio asociado al valor propio

$$\lambda_2 = 6$$

Procediendo de la misma manera que hemos indicado con anterioridad el lector podrá comprobar el siguiente resultado:

$$\vec{u}^2 = \begin{bmatrix} k \\ -2k \\ k \end{bmatrix}$$

-Cálculo del sub-espacio vectorial propio asociado al valor propio

$$\lambda_3 = 5$$

Procediendo de la misma manera que hemos indicado con anterioridad el lector podrá comprobar el siguiente resultado:

$$\vec{u}^3 = \begin{bmatrix} -k \\ 0 \\ k \end{bmatrix}$$

Segunda sub-etapa: cálculo de los ejes principales en un ACP

Una vez que hemos obtenido los **sub-espacios vectoriales propios** asociados a los **valores propios**, el cálculo de los **ejes principales**, es casi inmediato.

Para no resultar reiterativos, como la manera de obtenerlos es la misma para todos los ejes principales, tan sólo detallaremos el cálculo para el primer eje principal. Para los dos restantes, nos limitaremos a poner los resultados finales con el fin de que, aquellos que deseen calcularlos puedan comprobar si han llegado con éxito al resultado final.

Cálculo del primer eje principal.

Para ello partimos del **sub-espacio vectorial propio** asociados al primer **vector propio** de:

$$VI_{(3,3)} : \vec{u}^1 = \begin{bmatrix} k \\ k \\ k \end{bmatrix}$$

Así pues, tenemos que encontrar el k que verifique la siguiente condición

$$\vec{u}^{1T} I_{3,3} \vec{u}^1 = 1 \quad (4)$$

De (4) se deduce que:

$$\begin{bmatrix} k & k & k \end{bmatrix} I_{(3,3)} \begin{bmatrix} k \\ k \\ k \end{bmatrix} = 1$$

Mediante una simple multiplicación de **matrices** llegamos al siguiente resultado:

$$k^2 = \frac{1}{3}$$

Por tanto,

$$k_1 = \frac{1}{\sqrt{3}} \quad \text{y} \quad k_2 = -\frac{1}{\sqrt{3}}$$

Reteniendo el valor positivo de k, concluimos que entre el **sub-espacio vectorial propio** asociado al **valor propio**

$$\lambda_1 = 7,5$$

el **vector propio** que verifica

$$\vec{u}^{1T} I_{3,3} \vec{u}^1 = 1 \text{ es: } \vec{u}^1 = \begin{bmatrix} 1 \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$$

Cálculo del segundo eje principal

Para no resultar reiterativos, dado que el cálculo del **sub-espacio vectorial propio** asociado al **valor propio**

$$\lambda_2 = 6.0$$

y la búsqueda del k para cada **sub-espacio vectorial** bajo la **métrica**.

$$I_{(3,3)}$$

se realiza mediante el mismo procedimiento ya indicado para

$$\vec{u}^1$$

nos limitaremos simplemente a presentar el resultado final:

$$\vec{u}^2 = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}$$

Cálculo de tercer eje principal

Actuando de la misma manera que para el **primer eje principal**, llegamos al resultado final:

$$\vec{u}^3 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

Observación de interés: Los tres vectores que hemos calculado no son más que los **ejes principales de un ACP** y presentan la **particularidad de ser**

$$I_{(3,3)} - \textit{ortonormados}$$

Los lectores podrán observar que, dichos **ejes principales** son los mismos que se contemplan en (1 (págs. 80)).

En el esquema de dualidad que hemos presentado en la parte teórica, estos tres vectores pertenecen a

$$R^3$$

Úndecima etapa: cálculo de los factores principales

Para pasar del ángulo superior izquierdo al inferior izquierdo en el esquema de dualidad que hemos presentado en la parte teórica.

$$R^3 \textit{ a } R^{3^*} \textit{ (espacio vectorial de las aplicaciones lineales de } R^3 \textit{ en } K)$$

es decir, para pasar de

$$R^3 \textit{ a } R^{3^*}$$

procedemos de la manera que mostramos a continuación:

$$I_{3,3} \vec{u}^1 = \vec{v}^1$$

$$I_{3,3} \vec{u}^2 = \vec{v}^2$$

$$I_{3,3} \vec{u}^3 = \vec{v}^3$$

Los tres vectores:

$$\vec{v}^1, \vec{v}^2 \textit{ y } \vec{v}^3$$

son los **factores principales** de un ACP y presentan la propiedad de ser

$$I_{(3,3)}^{-1} - \textit{ortonormados}$$

Observación de interés: En nuestro caso concreto los **factores principales:**

$$\vec{v}^j \in R^3 \quad j=1,2,3$$

coinciden con los **ejes principales**:

$$\vec{u}^j \in R^3 \quad j=1,2,3$$

Duodécima etapa: cálculo de las componentes principales

Para pasar del **ángulo inferior izquierdo al ángulo inferior derecho del esquema de dualidad**, contenido en la parte teórica, es decir, para pasar de

$$R^{3*} \text{ (espacio vectorial de las aplicaciones lineales de } R^3 \text{ en } K)$$

$$\quad \quad \quad a$$

$$R^4 \text{ (espacio de las variables)}$$

procedemos de la manera que mostramos a continuación:

$$\begin{bmatrix} I_{(4,4)} & -\frac{1_4 \cdot 1_4^T}{4} \end{bmatrix} X_{(4,3)} \vec{v}^1 = \vec{c}^1 \quad [1]$$

$$\begin{bmatrix} I_{(4,4)} & -\frac{1_4 \cdot 1_4^T}{4} \end{bmatrix} X_{(4,3)} \vec{v}^2 = \vec{c}^2 \quad [2]$$

$$\begin{bmatrix} I_{(4,4)} & -\frac{1_4 \cdot 1_4^T}{4} \end{bmatrix} X_{(4,3)} \vec{v}^3 = \vec{c}^3 \quad [3]$$

Los vectores pertenecientes a

$$R^4$$

son las **componentes principales** [coordenadas de los puntos-filas (individuos) sobre los ejes principales].

-Cálculo de la primera componente principal [coordenadas de los puntos-filas (individuos) sobre el primer eje principal]

$$\vec{c}^1 \in R^4$$

Sustituyendo en [1] nuestros datos concretos tenemos que,

$$\begin{bmatrix} 0 & -4 & -2 \\ -4 & 1 & 0 \\ 2 & 3 & -2 \\ 2 & 0 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ \sqrt{3} \\ \sqrt{3} \end{bmatrix} = \begin{bmatrix} -\frac{6}{\sqrt{3}} \\ \frac{3}{\sqrt{3}} \\ \frac{3}{\sqrt{3}} \\ \frac{6}{\sqrt{3}} \end{bmatrix}$$

-Cálculo de la segunda componente principal [coordenadas de los puntos-filas (individuos) sobre el segundo eje principal]

$$\vec{c}^2 \in R^4$$

Sustituyendo en [2] nuestros datos concretos tenemos que,

$$\begin{bmatrix} 0 & -4 & -2 \\ -4 & 1 & 0 \\ 2 & 3 & -2 \\ 2 & 0 & 4 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix} = \begin{bmatrix} \frac{6}{\sqrt{6}} \\ -\frac{6}{\sqrt{6}} \\ -\frac{6}{\sqrt{6}} \\ \frac{6}{\sqrt{6}} \end{bmatrix}$$

-Cálculo de la tercera componente principal [coordenadas de los puntos-filas (individuos) sobre el tercer eje principal]

$$\vec{c}^3 \in R^4$$

Sustituyendo en [3] nuestros datos concretos tenemos que,

$$\begin{bmatrix} 0 & -4 & -2 \\ -4 & 1 & 0 \\ 2 & 3 & -2 \\ 2 & 0 & 4 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} -\frac{2}{\sqrt{2}} \\ \frac{4}{\sqrt{2}} \\ -\frac{4}{\sqrt{2}} \\ \frac{2}{\sqrt{2}} \end{bmatrix}$$

Las **componentes principales** que hemos calculado son **correctas** dado que se verifica no sólo:

$$\vec{c}^{1'} \frac{1}{4} I_{4,4} \vec{c}^1 = 7,5$$

$$\vec{c}^{2'} \frac{1}{4} I_{4,4} \vec{c}^2 = 6,0$$

$$\vec{c}^{3'} \frac{1}{4} I_{4,4} \vec{c}^3 = 5,0$$

sino también,

$$c^{j'} \frac{1}{4} c^j = 0 \quad j \neq j' \quad j=1,2,3$$

Finalmente, observamos que J. Bon y P. Grégory (1 (pág. 83)) llegan a los mismo resultados que los nuestros, por lo tanto, el procedimiento general que hemos propuesto es totalmente **correcto**.

Observación de interés:

Según apunta el profesor Gilbert Saporta en (21 (pág. 171)), entre las dos opciones que están programadas en los paquetes de análisis estadístico multidimensional para aplicar un ACP, la desarrollada en este artículo, es la menos aconsejable desde un punto de vista práctico en el supuesto de que, se hubieran contemplado las condiciones ya mostradas por el profesor Thierry Foucart (2 (pág. 16)) ya que, no goza de las mismas propiedades de la consistente en diagonalizar la matriz de correlaciones de Auguste Bravais-Karl Pearson.

Agradecimientos:

Este artículo de carácter práctico no habría sido posible realizarlo sin la colaboración que tuve en el año 1976 de Jean-Baptiste Denis (29), investigador del INRA, y fue realizado durante su estancia en el INIA (Sección de Proceso de Datos) de Madrid.

BIBLIOGRAFÍA

- (1) Bon, J.; Grégory, D. (1986). *Techniques marketing*. Vuibert Gestion. Paris.
- (2) Foucart, Th. (1997). *L'Analyse des Données. Mode d'emploi*. Presses Universitaires de Rennes.
- (3) Cailliez, F. ; Pagès, J-P. ; Nakache, J-P. ; Mailles, J-P. (1972). *Cours d'Analyse Multidimensionnelles*. Polycopié du CEEE.
- (4) Cailliez, F. ; Pagès, J-P. (1976). *Introduction à l'Analyse des Données*. SMASH. Paris.
- (5) Díaz-Llanos, Fco. J.; Cayón, J.; Cermeño, C. (2009). ACP de un triplete estadístico (X, Q, D) antes de una regresión lineal múltiple en una situación límite. *Anales de la Real Academia de Doctores de España*. Volumen 13, Número 2, pp. 147-157.
- (6) Guigou, J-L. (1977). *Méthodes multidimensionnelles. Analyse des Données et choix à critères multiples*. Deuxième édition complétée. Dunod.
- (7) Roux, M. (1985). *Algorithmes de classifications*. Préface du Pr. Yves Escoufier. Masson.
- (8) Philippeau, G. (1986). *Comment interpréter les résultats d'une analyse en composantes principales ?* ITCF. Paris.
- (9) Aluja Banet, T. ; Morineau, A. (1999). *Aprender de los datos. El análisis de componentes principales. Una aproximación desde el Data Mining*. EUG, Barcelona.

- (10) Morineau, A.; Aluja Banet, T. (2000). Analyse en composantes principales (avec illustrations SPAD). CISIA-CE-RESTA, Montreuil.
- (11) Benzécri, J-P et collaborateurs (1973). L'Analyse des Données. 2. L'Analyse des correspondances. Dunod.
- (12) Bertier, P.; Bouroche, J-M. (1977). Analyse des données multidimensionnelles. Préface de Georges Morlat. Presses Universitaires de France. Paris.
- (13) Diday, E. et collaborateurs. (1978). Optimisation en classification automatique. Tome 1 et Tome 2. Dunod.
- (14) Volle, M. (1981). Analyse des données. Deuxième édition. Económica.
- (15) Nakache, J-P.; Chevalier, A.; Morice, V. (1981). Exercices commentés de mathématiques pour l'Analyse statistique des données. Dunod.
- (16) Diday, E. ; Lemaire, J.; Pouget, J. ; Testu, F. (1982). Éléments d'analyse des données. Dunod.
- (17) Cailliez, F. (1984). Analyse des données. Les Presses de l'Université de Montréal (Québec). Canada.
- (18) Robert, Cl. (1989). Analyse Descriptive Multivariée. Application à l'intelligence artificielle. Médecine-Sciences. Flammarion.
- (19) Celeux, G. ; Diday, E.; Govaert, G.; Lechevallier, Y.; Relambondrainy, H. (1989). Classification automatique des données. Environnement statistique et informatique. Dunod Informatique.
- (20) Pontier, J. ; Dufour, A-B.; Normand, M. (1990). Le modèle euclidien en Analyse des données. Éditions de l'Université de Bruxelles. Éditions Ellipses.
- (21) Saporta, G. (1990). Probabilités. Analyse des données et Statistique. Éditions Technip.
- (22) Dreesbeke, J-J. ; Fichet, B. ; Tassi, Ph (éditerus). (1992). Modèles pour l'Analyse de données multidimensionnelles. Economica.
- (23) Rouanet, H. ; Le Roux, B. (1993). Analyse des données multidimensionnelles. Dunod.
- (24) Lebart, L.; Morineau, A. ; Piron, M. (1995). Statistique exploratoire multidimensionnelle. Dunod.
- (25) Dazy, F.; Le Barzic, J-F. (1996). L'Analyse des données évolutives. Méthodes et applications. Édition Technip.
- (26) Jambu, M. (1999). Méthodes de base de l'Analyse des données. Eyrolles.
- (27) Tenenhaus, M. (2007). Statistique. Méthodes pour décrire, expliquer et prévoir. Dunod.
- (28) Lascoux, P.; Theodor, R. (1986). Analyse numérique matricielle appliquée à l'art de l'ingénieur. Tome 1. Masson.
- (29) Denis, J.B. y colaborador. (1976). Apuntes redactados con ocasión del cursillo impartido en la Sección de Proceso de Datos sobre el Tratamiento de Datos. INIA. MAPA.