Anales de la Real Academia de Doctores de España
Volumen 15, n.º 1, 2011, pp. 51-75

ELEMENTOS BÁSICOS EN EL ANÁLISIS EN COMPONENTES PRINCIPALES (ACP) [Primera parte: Teoría]

DOCTOR D. FCO. JAVIER DÍAZ LLANOS SÁINZ-CALLEJA
Académico de Número de la sección de Ingeniería de la Real Academia de Doctores de España. N.º 118. Medalla al Mérito Doctoral. Categoría de Plata.

M. YVES ESCOUFIER
Professeur 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 en las Universidades Technische de München y Freie Angewandte, Berlín, Deutschland, donde obtuvo «venia docendi». Referee de artículos en dichas Universidades.

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 exponer —de la manera más didáctica posible— no solo un conjunto de elementos básicos en el **análisis en componentes principales**, tales como: **las métricas contenidas en el esquema de dualidad** sino también, cómo se obtienen los tres elementos básicos de un **análisis en componentes principales: los ejes principales, los factores principales y las componentes principales**. Que nosotros sepamos, el **ACP** bajo la opción de **(esquema de dualidad** y **métricas)** se conoce, en Francia, desde el año 1972 y, por tal motivo, tan solo demostraremos una propiedad de interés de dicho **análisis** en cuanto se refiere a **las componentes principales.**

1. MÉTRICAS CONTENIDAS EN EL ESQUEMA DE DUALIDAD

1.1. La **métrica** inducida en

$$R^{p^*}$$

por el hecho de introducir *a priori* en

$$R^{p^*} \textit{ la métrica } M_{(p,p)}$$
(matriz real definida positiva)

1.2. La **métrica** inducida en

$$R^{n^*}$$

por el hecho de introducir *a priori* en

$$R^{n} \textit{ la métrica } N_{(nxn)}$$
(matriz diagonal positiva)

$$N_{nxn} = diag(p_i) \qquad \sum_{i=1}^{i=n} p_i = 1 \quad i = 1,...,n$$

donde los

$$p_i$$

representan los pesos que pueden asociarse a los individuos.

1.3. Las **métricas** inducidas en

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

por el hecho de introducir *a priori* en

$$R^{p} \textit{ la métrica } M_{(p,p)}$$

1.4. Las **métricas** inducidas en

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

por el hecho de introducir *a priori* en

$$R^{n} \textit{ de la métrica } N_{(n,n)}$$

2. CÁLCULO DE LOS EJES PRINCIPALES, LOS FACTORES PRINCIPALES Y LAS COMPONENTES PRINCIPALES

3. DEMOSTRAR QUE SI EL CUADRADO DE LA NORMA DE LOS VECTORES

$$\vec{c}^{\,j}_{(n,1)} \in R^{n} \qquad j = 1,\ldots,p$$

en el sentido de la **métrica**

$$N_{(n,\ n)} \text{ es igual a } \lambda_j \quad j=1,\ldots,p$$

se tiene que verificar que el **cuadrado de la norma de los vectores**

$$\vec{v}^{\,j}_{(p,1)} \in R^{p^*}$$

en el sentido de la **métrica**

$$M^{-1}_{(p,p)} \text{ es igual a } 1.$$

Contemplaremos dos enfoques para la construcción del **esquema de dualidad:**

Enfoque 1:

Para el desarrollo de estos puntos haremos uso del **esquema de dualidad** propuesto en este artículo.

Dicho esquema es el que corresponde a la estructura de la matriz de datos contemplada en el año 1976 por Jean-Baptiste Denis y su colaborador Fco. J. Díaz-Llanos (1) y en los libros lanzados al mercado en los años 1975 y 1990 por: Patrice Bertier y Jean Bouroche (2), por un lado, y Gilbert Saporta (3), por otro, respectivamente.

Enfoque 2:

No está de más recordar que, el **esquema de dualidad** presentado en 1972 por los profesores Francis Cailliez, Jean-Paul Pagès, J-P. Nakache y J-P. Mailles (4), así como por el profesor Yves Escoufier (5) y, finalmente, en el libro de los profesores Francis Cailliez y Jean-Paul Pagès (6), hacen alusión al mismo **esquema de dualidad.**

El esquema que utilizamos nosotros [Enfoque 1], es distinto que el segundo enfoque, por el simple hecho de que en (1, 2, 3) se parte de una matriz de datos de dimensiones (n, p), donde n representa los individuos (filas), y p las variables (columnas) mientras que en (4, 5, 6) [Enfoque 2], se parte de una matriz de datos de dimensiones (p, n), donde p representa las variables (filas), y n los individuos (columnas).

Palabras clave: esquema de dualidad, métricas, cuadrado de la norma de un vector en el sentido de una métrica, vectores M-ortonormados, vectores N-ortogonales,

$$\textit{vectores } M^{-1}_{(p,p)} - \textit{ortonormados,}$$

los **ejes principales,** los **factores principales** y las **componentes principales.**

INTRODUCCIÓN

En este artículo, únicamente nos centraremos en la presentación de los elementos básicos en el **análisis en componentes principales,** basándonos en las **métricas** que deben introducir los investigadores.

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

Hemos tomado la decisión de mostrar tan solo los elementos básicos, por los motivos que mostraremos a continuación en el punto 1.

No está demás recordar que, en el **análisis de correspondencias** como **el análisis discriminante lineal,** cuyas variables explicativas son cuantitativas, son un caso particular del **análisis en componentes principales.**

Por otra parte, hemos de indicar que, **el análisis de datos à la française (métricas),** presenta la particularidad frente a los **métodos anglosajones** que rechazan las hipótesis distribucionales *a priori* y, para la contrastación de los resultados, huyen de las hipótesis distribucionales *a priori* centrándose —únicamente— en la **estabilidad observada de la representación sintética en los planos factoriales.** Así como en el **ACP,** los **puntos-fila** y los **puntos-columnas** no se pueden interpretar en el mismo **plano factorial**, en el AFC **(análisis factorial de correspondencias)** los **puntos-fila** (frecuencias relativas) y los **puntos-columnas** (frecuencias relativas) sí se puede realizar dicha interpretación. En el **AFC** las **métricas** introducidas en

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

se fijan *a priori* y, por tanto, no se necesita a este respecto de la opinión de los investigadores.

En última instancia, recordaremos que, tanto el análisis en **componentes principales, el análisis de correspondencias y el análisis discriminante lineal,** cuyas variables explicativas sean cuantitativas, son un caso particular del **análisis canónico.** Lamentablemente, el **análisis canónico** a nivel teórico es muy importante pero, a nivel de la ayuda a la interpretación, no proporciona resultados satisfactorios y por consiguiente, se debería desterrar de los libros de **análisis estadístico multidimensional** e incluir como sustitución el **ACPVI (análisis en componentes principales con respecto a variables instrumentales)** (5) y en última instancia, **la regresión PLS 2 (regresión por el método de mínimos cuadrados parciales).**

1. ORÍGENES DEL PCA (SIGLA ANGLOSAJONA) Y EL ACP (SIGLA FRANCESA)

El **análisis en componentes principales** merece una atención especial, puesto que fue propuesto en el año 1901 por Karl Pearson (7), y 32 años más tarde por H. Hotteling (8).

Recordamos de nuevo que en Francia, en el año 1972, los profesores F. Cailliez, J-P Pagès, J-P. Nakache y J-P Mailles (4) ya hacían referencia —en un curso que disponía de unos apuntes— sobre el **ACP (esquema de dualidad, métricas).**

Cuatro años más tarde, tal como ya se ha apuntado anteriormente en el resumen, apareció el libro (6) en el cual intervinieron un grupo de matemáticos (estadísticos) de gran reputación internacional tales como el profesor Yves Escoufier.

Fue entonces cuando, en Francia, en los años 1980 surgió **l'école française de l'analyse des données,** integrada no solo por los profesores que intervinieron en (4), sino también por el profesor Jean-Paul Benzécri y sus colaboradores (9).

Adicionalmente apuntaremos que en el año 2007, el profesor Michel Tenenhaus expuso de forma didáctica en (10) (pp. 152-180, 180-182), no solo la aproximación geométrica introducida en el año 1901 por Karl Pearson (7), sino también la aproximación introducida en el año 1933 por H. Hotelling (8), consistente en obtener más directamente las **componentes principales**, sin pasar por la etapa de la construcción de los **ejes principales.**

2. DISIMILITUDES ENTRE EL FACTOR ANALYSIS Y L'ANALYSE EN COMPOSANTES PRINCIPALES

Entre estos dos **métodos factoriales**, nos inclinamos por los de **l'école française de l'analyse des données,** que descarta las **hipótesis distribucionales** *a priori* y, por tanto, hace uso exclusivamente de la **geometría multidimensional (afín y euclidiana).**

La aplicación del **"factor análisis"** a unos datos empíricos por H. H. Harman en 1967 (11), cuyo **objetivo metodológico** es distinto al **ACP à la française,** nos conduce —sin duda alguna— a resultados menos satisfactorios, a nivel de ayuda a la interpretación de datos empíricos, que la aplicación de un **ACP à la française.**

Así como el elemento distorsionador en el **factor analysis** es la **rotación de los ejes,** que según se realice de una manera u otra nos dará un resultado u otro, en el **ACP à la française** es la **métrica.**

Nos parece importante apuntar que, mientras que la **métrica** tiene un **sentido práctico,** la **rotación de los ejes** es una mera especulación de los **datos empíricos** y, por lo tanto, no lo tiene.

Si los investigadores científicos conocen bien sus propios datos empíricos —condición indispensable para que puedan interpretarlos correctamente con ayuda del estadístico—, sin duda alguna es preferiblemente la aplicación del **ACP à la française** que el **«factor análisis»** (11).

3. OBSERVACIÓN DE INTERÉS

Los profesores Fréderic Dazy y Jean-François Labarzic en su libro (12 (p. 9)) bajo la dirección de los profesores Gilbert Saporta y Françoise Lavallar nos indican —de forma general— cómo debemos centrar por columnas la tabla de datos originales.

Por otro lado, la tabla de datos originales tiene la misma estructura que la de los libros (1, 2, 3) y, la mostraremos —con detalle— en el epígrafe de material no informático.

La tabla de datos para analizar que proponen Fréderic Dazy y Jean-François Labarzic (12 (p. 9)) presenta la siguiente estructura:

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

En el caso hipotético que todos los

$$p_i$$

fueran iguales, entonces dicha estructura sería:

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

4. OTRAS CONSIDERACIONES DE INTERÉS

Los lectores podrán comprobar —fácilmente— que el **ACP** que se contempla en la mayoría de los libros de **análisis estadístico multidimensional** y de **estadística** —escritos en **castellano**—, son dos casos particulares del **ACP *à la française*** que presentamos en este artículo.

Creemos que, puesto que ya estamos en el año 2011, la inmensa mayoría de los autores de los libros de texto escritos en **castellano** cuando presentan el **ACP,** hacen caso omiso de las **métricas** asociadas en

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

Puesto que sería de gran utilidad el uso de las **métricas en** R^p y R^n en el (ACP), consideramos fundamental la implementación de las métricas en el (ACP).

De no llevarlo a cabo de esta manera, esto implicaría una distorsionada interpretación de sus propios datos empíricos.

En los paquetes de programas actualmente comercializados —que nosotros sepamos—, se encuentra implícitamente plasmada para la construcción de los ejes principales la diagonalización de la matriz de varianzas-covarianzas y de correlaciones entre las variables cuantitativas originales.

La diagonalización primera aludida (varianzas-covarianzas) implica la introducción en

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

de las métricas: $M_{(p,\ p)} = I_{(p,\ p)}$ y $N_{(n,n)} = \frac{1}{n} I_{(n,n)}$ respectivamente, y la segunda, implica la introducción en R^p *y* R^n

de las métricas: $M_{(p,p)} = D_{\frac{1}{s^{2j}}}\quad j = 1,...,p\ \ y\ \ N_{(n,n)} = \frac{1}{n} I_{(n,n)}$

Hemos decidido que (después de escribir este primer artículo de contenido teórico), escribiremos otros dos más en **castellano: uno de contenido práctico y el otro de contenido teórico y práctico.** Los investigadores que deseen utilizar —debidamente— las **métricas,** deberían leer el artículo (13) donde indicamos la forma de construirlas, o bien, antes de aplicar un **ACP** a sus propios datos empíricos, consultar con un estadístico y, éste les indicará cómo realizarlo, teniendo en cuenta el conocimiento de los profesionales.

El **error** —producido por obviar las **métricas** en un **ACP**— sería parecido al de construir una **tipología concreta** basándonos en que se verifica la hipótesis de Laplace-Gauss Multidimensional.

En este sentido es necesario recordar el **primer principio** del **análisis de datos** contemplado en el año 1973 por el profesor Jean-Paul Benzécri y colaboradores (9) (pp. 3-6).

5. DOS CASOS PARTICULARES EN EL ANÁLISIS EN COMPONENTES PRINCIPALES (ACP)

Bajo nuestra nomenclatura, en los paquetes de programas comercializados **—al menos, hasta el año 2011, bajo las múltiples opciones de métricas que pueden incorporarse en un análisis en componentes principales—,** tan solo se contemplan dos opciones:

5.1. Primera opción

La **primera opción** es un caso particular de la presentación general a la que hacemos alusión en este artículo sobre el **conocimiento científico** en un **ACP**.

Partiendo del caso general cuyo **triplete estadístico** adopta la siguiente forma:

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

donde,

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

1_n: *es un vector columna. Todos sus elementos son unos.*

1^T_n: *es el vector transpuesto de* 1_n.

$N_{(n,n)}$: *es la métrica introducida en* R^n *(espacio de las variables).*

Esta métrica la definimos de la siguiente manera:

$N_{(n,n)}$ *= diag(p_i) bajo la condición:*

$$\sum_{i=1}^{i=n} p_i = 1.$$

Los p_i son los pesos que pueden asignarse a cada individuo según el investigador crea conveniente.

A título de ejemplo, los individuos pueden ser las provincias españolas. Es obvio que, a Madrid, no se le debe de **imputar el mismo peso que a Badajoz.**

$X_{(n,p)}$*: es la matriz que contiene los datos originales. El número de filas (individuos) lo representamos por n y el número de columnas (variables) lo representamos por p.*

$M_{(p,p)}$*: es la métrica introducida en R^p (espacio de los individuos).*

Por otra parte, las variables pueden ser las macromagnitudes más relevantes en todos los sectores de la economía nacional. Así, el número de teléfonos de una determinada provincia no aporta la misma relevancia que el valor añadido bruto al coste de los factores del sector servicio íntimamente relacionado con el valor añadido bruto del coste de los factores en el sector agrario. Por lo tanto, de alguna manera, hay que tener en cuenta en el análisis esta situación, pues sino, se están enmascarando los resultados del **análisis en componentes principales (ACP).**

Teniendo en cuenta lo expuesto, la primera opción consiste en:

$$N_{(n,n)} = \frac{1}{n} I_{(n,n)} \qquad M_{(p,p)} = I_{(p,p)}$$

donde,

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

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

Así pues, el **triplete estadístico** en la **primera opción** es:

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

Esto conlleva a que para el cálculo de los **valores propios** de

$$VI_{(p,p)},$$

elementos básicos para construir los **vectores propios** tendremos que **diagonalizar** la **matriz de varianzas-covarianzas** asociada a las p variables cuantitativas.

Según explica claramente el profesor Gilbert Saporta en (3 (p. 171)) dicha **opción** no es adecuada para realizar un **ACP,** ya que los resultados obtenidos no son **invariantes** si cambiamos linealmente la unidad de medida de las variables.

En esta **primera opción**

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

5.2. Segunda opción

La **segunda opción** es otro caso particular del caso general en el cual:

$$N_{(n,n)} = \frac{1}{n} I_{(n,n)} \qquad M_{(p,p)} = D_{\frac{1}{s^{2j}}}$$

donde,

$$D_{\frac{1}{s^{2j}}} : \textit{es la matriz diagonal de orden p}$$

Los elementos de la diagonal representan los $(s^{2^j})^{-1}$.

Las s^{2^j} *son las varianzas de las p variables cuantitativas* $j=1,\ldots,p$.

Así pues, el **triplete estadístico** en la **segunda opción** se puede expresar de la siguiente manera:

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

Esto conlleva a que para el cálculo de los **valores propios** de

$$VI_{(p,p)},$$

elementos básicos para construir los **vectores propios** tendremos que **diagonalizar** la matriz de correlaciones de Auguste Bravais-Karl Pearson asociada a las p variables cuantitativas.

Según explica claramente el profesor Gilbert Saporta en (3 (p. 171)), dicha opción es particularmente interesante cuando las p variables cuantitativas son heterogéneas.

En esta **segunda opción:**

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

MATERIAL Y MÉTODO

Material

1. *Material de base (no informático) para la realización de un ACP **à la française***

La necesidad de definir correctamente los tres elementos que contiene el **triplete estadístico** es básica para la ayuda a la interpretación de los datos empíricos aportados por los investigadores científicos.

El **triplete estadístico** lo definimos de la siguiente forma:

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

¿Cuál es el significado y cómo se definen estos tres elementos básicos?

1.1. *Significado de $X_{(n,p)}$*

La matriz de datos original de partida: $X_{(n,p)}$

Tendremos que definir:

1.1.1. Su **estructura.**

1.1.2. Sus **dimensiones.**

1.1.3. Algunas particularidades de la matriz de datos.

1.1.1. Su **estructura**

Consideramos que todas las variables contenidas en la tabla de datos de partida son **cuantitativas.**

Partimos de una **matriz de datos originales,** tal como mostramos a continuación.

$$X_{(n,p)} = \begin{pmatrix} x_1^1 & x_1^2 & \dots & \dots & x_1^p \\ x_2^1 & x_2^2 & \dots & \dots & x_2^p \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ x_n^1 & x_n^2 & \dots & \dots & x_n^p \end{pmatrix}$$

en donde x_i^j es el valor de la variable j sobre el individuo i

$j = 1,\dots,p \quad i = 1,\dots,n$

La **matriz de datos originales centrada por columna por los pesos que constituyen la diagonal de**

$$N_{(n,n)}$$

adopta la siguiente estructura:

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

1.1.2. Sus **dimensiones**

El profesor Thierry Foucart apunta en (14) (pág. 16) que la interpretación de un **ACP,** tan solo es válida si la matriz de datos originales contiene más de 15 **individuos** y más de cuatro **variables cuantitativas.**

No obstante, cuantas más variables e individuos contenga la tabla de datos, el nivel lógico de ayuda a la interpretación será de mejor calidad. Es fundamental que para que la interpretación de los datos empíricos sea fidedigna, haya una interacción entre el conocedor de los datos empíricos y el estadístico. Dicha interacción debe empezar desde el principio hasta el fin del trabajo que se esté realizando.

En principio, es mejor, como es lógico, que dispongamos de más variables cuantitativas que de individuos siempre que sea posible.

1.1.3. Algunas particularidades de la tabla de datos original

No obstante, según el tipo del problema que deseemos tratar en ciertas ocasiones dispondremos de más variables que de individuos.

Esto no impedirá, por supuesto, la incapacidad de la realización de un **análisis en componentes principales.**

Aunque, en ciertas ocasiones, es posible que la tabla de datos esté incompleta. Esta limitación tampoco impedirá que podamos realizar un **análisis en componentes principales.**

El que se verifiquen alguna —o las dos limitaciones a la vez—, tampoco impedirá que podamos realizar un **análisis en componentes principales.**

Como se puede apreciar, un caso claro se contempla en (15) (págs. 131-139) en donde hemos realizado un estudio sobre la **fluoxitina,** en el cual hay más variables que individuos.

¿Cuál es el significado de

$$M_{(p,p)} \text{ y } N_{(n,n)} \text{ ?}$$

1.2. Significado de

$$M_{(p,\,p)}$$

$M_{(p,\,p)}$ es la métrica introducida en R^p (espacio de los individuos)

Los responsables de la construcción de esta métrica —tal como ya hemos indicado en (13)—, son los propios investigadores que van a aportar sus datos a los estadísticos para que, conjuntamente de forma interactiva puedan extraer de sus datos la información menos distorsionada posible.

1.3. Significado de

$$N_{(n,\,n)}$$

$N_{(n,\,n)}$ es la métrica introducida en R^n (espacio de las variables).

Para no resultar reiterativos, los responsables son los mimos que para la construcción de la **métrica**

$$M_{(p,\,p)}$$

2. *Material de base (informático) para la realización de un ACP **à la française***

Es verdaderamente lamentable que, aunque el procedimiento que mostramos en este artículo —al menos, sus ideas básicas— se conocen formalmente —bajo estructura de libro editado— desde el año 1976 (6), que nosotros sepamos, aún en el año 2011 no hay ningún paquete de programa de **análisis estadístico multidimensional** comercializado que, contenga la inclusión de las **métricas** que más se ajusten a los datos empíricos de los investigadores.

No está demás recordar la idea que tienen numerosos investigadores de aplicar reiteradamente un **ACP** consistente: *a)* bien en la diagonalización de la matriz de varianzas-covarianzas entre las variables cuantitativas; *b)* en la diagonalización de la matriz de correlaciones de Auguste Bravais-Karl Pearson entre variables cuantitativas. Ya hemos apuntado que, entre estas dos opciones, la única que puede ser aceptada es la segunda (3 (p. 171)).

Así pues, los investigadores en sus análisis generalmente utilizan la opción de la diagonalización de la matriz de correlaciones, aunque la deseada para sus propios datos empíricos fuera otra, pero lo cierto es que sienten limitados para hacerlo por otro procedimiento indudablemente por dos motivos: 1) ausencia de conocimientos de Estadística; 2) no disponer del programa adecuado para hacerlo ya que —según nuestro conocimiento— no hay ningún programa de **ACP comercializado** en el mercado en el que se haya implementado la opción de **métricas.**

Por lo tanto, una vez asimilado nuestro artículo, les invitamos a que acudan a los informáticos de sus Instituciones para que, implementen la opción de **métricas** en

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

Indicaremos que sobre todo, la responsabilidad de construir **correctamente** dichas **métricas** en sus estudios es de los propios investigadores, que son los que deben conocer sus propios **datos empíricos,** y la forma de hacerlo ya está indicada en (12).

De no ser así, entre las múltiples opciones que presenta el **ACP *à la française,*** los investigadores tan solo podrán aplicar dos de ellas —que son las que se contemplan en los paquetes de programas de **análisis estadístico multidimensional**— y, por consiguiente, arrastrarán, sin remedio, las mismas **imprecisiones** en la interpretación de sus propios **datos empíricos.**

En cuanto a las dos **métricas** ya implementadas en

$$R^p$$

para la aplicación de un **ACP *à la française,*** el profesor Gilbert Saporta (3 (p. 171)) hace una reflexión sobre las dos únicas **métricas** implementadas en el **análisis en componentes principales.** Estas dos **métricas** nos conducirán a la **diagonalización** de la matriz de varianzas-covarianzas y la matriz de correlaciones de Auguste Bravais-Karl Pearson. Para cualquier información adicional indicaremos a los investigadores que contacten con el primer autor de este artículo.

Método

Como más arriba hemos apuntado, el elemento básico de nuestra metodología es el **esquema de dualidad** en el cual se encuentran contemplados los tres elementos básicos del **triplete estadístico**

$$\left(\left(I_{n,n} - 1_n 1_n^T \; N_{(n,n)}\right) X_{n,p}, M_{pxp}, N_{nxn}\right)$$

1. Estructura básica del esquema de dualidad y su utilidad

1.1. Posicionamiento de los elementos básicos del **triplete estadístico** en el **esquema de dualidad;** así como los cuatro espacios vectoriales. En dicho **esquema de dualidad** se contemplará también otros dos elementos de interés: V y W.

1.1.1. *Esquema de dualidad básico*

1.1.2. *Significado de los símbolos contenidos en el esquema de dualidad*

1.1.3. *¿Cómo se puede pasar de un espacio vectorial a otro haciendo uso del esquema de dualidad?*

1.1.1. *Esquema de dualidad básico*

$$\begin{tikzcd}[column sep=huge, row sep=huge]
R^{p} \arrow[d, shift right=2ex, "M_{(p,p)}"'] & R^{n*} \arrow[l, "\left(\left(I_{n,n} - 1_n 1_n^T N_{(n,n)}\right) X_{n,p}\right)^T"'] \arrow[d, "W"'] \\
M^{-1}_{(p,p)}\ R^{p*} \arrow[u, "V"'] \arrow[r, "\left(I_{n,n} - 1_n 1_n^T N_{(n,n)}\right) X_{n,p}"] & R^{n} \arrow[u, shift right=2ex, "N_{(n,n)}"']
\end{tikzcd}$$

1.1.2. *Significado de los símbolos contenidos en el esquema de dualidad*

$M_{(p,p)}$: *métrica en el espacio de los individuos.*

$\vec{u}^{\,j} \in R^p$ *(espacio de los individuos)*

Esta **métrica** nos permitirá el cálculo de la distancia entre los individuos en

R^p

$N_{(n,n)}$: *métrica en el espacio de las variables.*

$\vec{c}^{\,j} \in R^n$ *(espacio de las variables)*

Esta **métrica** nos permitirá el cálculo de las distancias entre las variables.

R^{p^*} *[espacio vectorial de las aplicaciones lineales de* R^p *en K]:* $R^{p^*} = L\ (R^p, K)$

$\vec{v}^{\,j} \in R^{p^*}$

R^{n^*} *[espacio vectorial de las aplicaciones lineales de* R^n *en K]:* $R^{n^*} = L\ (R^n, K)]$

$\vec{w}^{\,j} \in R^{n^*}$

En dimensiones finitas existe un isomorfismo canónico del bidual $R^{p^{**}}$ sobre el espacio R^p. El cuerpo K cumple la propiedad conmutativa.

Aquellos lectores que deseen saber más sobre el «Dual de un espacio vectorial» deberían consultar el libro de J. Rivaud (16 (pp. 51-58)).

$\left(I_{n,n} - 1_n 1_n^T\ N_{(n,n)}\right) X_{n,p}$ *: matriz de datos originales centrada por columnas por los pesos contenidos en la matriz diagonal de orden n:* $N_{(n,n)}$.

$$N_{(n,n)} = diag(p_i) \quad \sum_{i=1}^{i=n} p_i = 1 \quad p_{i:} \text{ son positivos}$$

$\left(\left(I_{n,n} - 1_n 1_n^T \; N_{(n,n)}\right) X_{n,p}\right)^T$: transpuesta de la matriz de datos originales centrada por columnas por los pesos que constituyen la diagonal de $N_{(n,n)}$.

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

$$W_{(n,\,n)} = \left(I_{n,n} - 1_n 1_n^T \; N_{(n,n)}\right) X_{n,p} M_{(p,p)} \left(\left(I_{n,n} - 1_n 1_n^T \; N_{(n,n)}\right) X_{n,p}\right)^T$$

1.1.3. *¿Cómo se puede pasar de un espacio vectorial a otro haciendo uso del esquema de dualidad?*

1.1.3.1. Para pasar los vectores

$$\vec{\boldsymbol{u}}^j \in R^p$$

$$\textit{a los vectores } \vec{\boldsymbol{v}}^j \in R^{p^*}$$

tendremos que hacer la siguiente operación:

$$\mathrm{M}\vec{\boldsymbol{u}}^j = \vec{\boldsymbol{v}}^j$$

1.1.3.2. Para pasar los vectores

$$\vec{\boldsymbol{v}}^j \in \mathrm{R}^{\mathrm{p}^*}$$

$$\text{a los vectores } \vec{\boldsymbol{c}}^j \in \mathrm{R}^{\mathrm{n}}$$

tendremos que hacer uso de la siguiente operación:

$$\left(I_{n,n} - 1_n 1_n^T \; N_{(n,n)}\right) X_{(n,p)} \; \vec{\boldsymbol{v}}^j_{j(p,1)} = \vec{c}^{\,j}_{(n,1)}$$

1.1.3.3. Para pasar los vectores

$$\vec{\boldsymbol{c}}^j \in R^n$$

$$\textit{a los vectores } \vec{\boldsymbol{w}}^j \in R^{n^*}$$

operaremos de la siguiente manera:

$$N\vec{\boldsymbol{c}}^j = \vec{\boldsymbol{w}}^j$$

1.1.3.4. Para pasar los vectores

$$\vec{w}^j \in R^{n^*}$$

$$a\ los\ vectores\ \vec{u}^j \in R^p$$

tendremos en cuenta la siguiente relación:

$$\left(\left(I_{n,n} - 1_n 1_n^T \ N_{(n,n)}\right) X_{n,p}\right)^T \vec{w}^j = \vec{u}^j$$

Métrica inducida en R^{p^*} por introducir
en R^p la métrica $M_{(p,p)}$.

Dicha **métrica** se obtiene igualando **el cuadrado de la norma de los vectores**

$$\vec{u}^j \ y \ \vec{v}^j$$

en el sentido de las métricas

$$M \ y \ M_*$$

Es decir, procedemos de la siguiente manera:

$$\left\|\vec{u}^j\right\|_M^2 = \left\|\vec{v}^j\right\|_{M_*}^2$$

Por tanto,

$$\vec{u}^{j^T} M \ \vec{u}^j = \vec{v}^{j^T} M_* \ \vec{v}^j \qquad [1]$$

Para la búsqueda de

$$M_*$$

partimos de

$$\vec{u}^{jT} M \, \vec{u}^j$$

y tendremos en cuenta

$$M \, \vec{u}^j = \vec{v}^j \qquad [2]$$

Pre-multiplicando por

$$\mathrm{M}^{-1}$$

el término de la izquierda y de la derecha de [2] se concluye que,

$$\vec{u}^{\,j} = M^{-1}\,\vec{v}^{\,j}$$

De lo que se desprende que,

$$\vec{u}^{\,j^T} M\ \vec{u}^{\,j} = \vec{v}^{\,j^T} M^{-1} M\ \ M^{-1}\ \vec{v}^{\,j} = \vec{v}^{\,j^T} M^{-1} \vec{v}^{\,j} \qquad [3]$$

De la mera observación de las ecuaciones [1] y [3] se desprende que,

$$M_* = M^{-1}$$

Métrica inducida en R^{n^*} por introducir en
R^n la métrica $N_{(n,n)}$

Dicha **métrica** se obtiene igualando **el cuadrado de la norma de los vectores**

$$\vec{c}^{\,j}\ y\ \ \vec{w}^{j}$$

en el sentido de las métricas

N y N_*, respectivamente.

Por lo tanto, realizaremos lo siguiente:

$$\left\|\vec{c}^{\,j}\right\|_N^2 = \left\|\vec{w}^{j}\right\|_{N_*}^2$$

Por lo tanto,

$$\vec{c}^{\,j^T}\ N\ \ \vec{c}^{\,j} = \vec{w}^{j^T}\ N_*\ \ \vec{w}^{j} \qquad [4]$$

Para la búsqueda de

$$N_*$$

partiremos de

$$\vec{c}^{\,j^T} N\ \ \vec{c}^{\,j}$$

y tendremos en cuenta

$$N\ \ \vec{c}^{\,j^T} = \vec{w}^{j} \qquad [5]$$

Pre-multiplicando por

$$\mathrm{N}^{-1}$$

el término de la izquierda y de la derecha de [5] se concluye que

$$\vec{c}^{\,j^T} = N^{-1}\vec{w}^{j}$$

De lo que se desprende que,

$$\vec{c}^{\,j^T} N\ \vec{c}^{\,j} = \vec{w}^{j^T} N^{-1} N N^{-1} \vec{w}^{j} = \vec{w}^{j^T} N^{-1} \vec{w}^{j} \qquad [6]$$

De la mera observación de las ecuaciones [4] y [6] se desprende que,

$$N_* = \mathrm{N}^{-1}$$

Métricas inducidas en R^{n*} y R^n
por introducir en R^p la métrica $M_{(p,p)}$

Dichas métricas se obtienen igualando **el cuadrado de la norma de los vectores**

$$\vec{u}^{\,j}, \vec{w}^{j} \text{ y } \vec{c}^{\,j}$$

en el sentido de las métricas

M, M_* y M_{**}, respectivamente.

Por lo tanto, procedemos de la siguiente manera:

$$\left\|\vec{u}^{\,j}\right\|_M^2 = \left\|\vec{w}^{j}\right\|_{M_*}^2 = \left\|\vec{c}^{\,j}\right\|_{M_{**}}^2$$

lo que nos arroja el siguiente resultado

$$\vec{u}^{\,j^T} M\ \vec{u}^{\,j} = \vec{w}^{j^T} M_*\ \vec{w}^{j} = \vec{c}^{\,j^T} M_{**}\ \vec{c}^{\,j} \qquad [7]$$

Búsqueda de

$$M_* \text{ y } M_{**}$$

La búsqueda de estas dos **métricas** la realizaremos en dos fases:

Primera fase:

Partiremos de

$$\vec{u}^{\,j^T} M\ \vec{u}^{\,j}$$

y, tendremos en cuenta la siguiente relación**,**

$$\left(\left(I_{n,n} - 1_n 1_n^T\ N_{(n,n)}\right) X_{(n,p)}^T\right)^T \vec{w}_{(n,1)}^{j} = \vec{u}_{(p,1)}^{\,j}$$

Así que, estamos en condiciones de llegar al siguiente resultado:

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

$$\vec{w}^{jT}\left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)}\right) X_{(n,p)} M_{(p,p)} \left(\left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)}\right) X_{(n,p)}\right)^T \vec{w}^{j} \quad [8]$$

Segunda fase:

Partiremos del término de la derecha de [8] y tendremos en cuenta que,

$$\vec{w}^{j} = N \; \vec{c}^{\,j}$$

Por lo tanto,

$$\vec{w}^{jT}\left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)}\right) X_{(n,p)} M_{(p,p)} \left(\left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)}\right) X_{(n,p)}\right)^T \vec{w}^{j} =$$

$$= \vec{c}^{\,jT} N_{(n,n)} \left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)}\right) X_{(n,p)} M_{(p,p)} \left(\left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)}\right) X_{(n,p)}\right)^T \vec{c}^{\,j}$$

Conclusión

De los resultados de estas dos fases se desprende lo siguiente,

$$\vec{u}^{j^T} M_{(p,p)} \; \vec{u}^{j} = \vec{w}^{jT}\left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)}\right) X_{(n,p)} M_{(p,p)} \left(\left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)}\right) X_{(n,p)}\right)^T \vec{w}^{j} =$$

$$= \vec{c}^{\,jT} N_{(n,n)} \left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)}\right) X_{(n,p)} M_{(p,p)} \left(\left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)}\right) X_{(n,p)}\right)^T N_{(n,n)} \; \vec{c}^{\,j} \quad [9]$$

De la mera observación [7] y [9] se concluirá que:

$$M_{*} = \left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)}\right) X_{(n,p)} M_{(p,p)} \left(\left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)}\right) X_{(n,p)}\right)^T$$

$$M_{**} = N_{(n,n)} \left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)}\right) X_{(n,p)} M_{(p,p)} \left(\left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)}\right) X_{(n,p)}\right)^T N_{(n,n)}$$

Métricas inducidas en R^{p^*} y R^p
por introducir en R^n la métrica $N_{(n,n)}$

Dichas **métricas** se obtienen igualando **el cuadrado de la norma de los vectores**

$$\vec{c}^{\,j}, \vec{v}^{j} \;\; y \;\; \vec{u}^{j}$$

en el sentido de las métricas

$N, \; N_{*} \; y \, N_{**}$, *respectivamente*

Es decir, procedemos de la siguiente manera:

$$\left\|\vec{c}^{\,j}\right\|_N^2 = \left\|\vec{v}^{\,j}\right\|_{N_*}^2 = \left\|\vec{u}^{\,j}\right\|_{N_{**}}^2$$

Por tanto,

$$\vec{c}^{\,j^T} N \; \vec{c}^{\,j} = \vec{v}^{\,j^T} N_* \vec{v}^{\,j} = \vec{u}^{\,j^T} N_{**} \; \vec{u}^{\,j} \qquad [10]$$

Búsqueda de

$$N_* \; y \; N_{**}$$

La búsqueda de estas dos **métricas** la haremos en dos fases:

Primera fase:

Partimos de

$$\vec{c}^{\,j^T} N \; \vec{c}^{\,j}$$

Teniendo en cuenta que:

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

se desprende que:

$$\vec{c}^{\,j^T} N \; \vec{c}^{\,j} =$$

$$= \vec{v}^{\,j^T} \left(\left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)}\right) X_{(n,p)}\right)^T N_{(n,n)} \left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)}\right) X_{(n,p)} \; \vec{v}^{\,j}$$

Segunda fase:

Partiendo del término de la derecha de [11] y teniendo en cuenta que,

$$M \; \vec{u}^{\,j} = \vec{v}^{\,j}$$

se desprende que,

$$\vec{v}^{\,j^T} \left(\left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)}\right) X_{(n,p)}\right)^T N_{(n,n)} \left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)}\right) X_{(n,p)} \vec{v}^{\,j} =$$
$$= \vec{u}^{\,j^T} M_{(p,p)} \left(\left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)}\right) X_{(n,p)}\right)^T N_{(n,n)} \left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)}\right) X_{(n,p)} M_{(p,p)} \vec{u}^{\,j}$$

Conclusión

De los resultados de estas dos fases se desprende que,

$$\vec{c}^{\,j^T} N_{(n,n)} \vec{c}^{\,j} = \vec{v}^{\,j^T} \left(\left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)} \right) X_{(n,p)} \right)^T N_{(n,n)} \left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)} \right) X_{(n,p)} \; \vec{v}^{\,j} =$$

$$= \vec{u}^{\,j^T} M_{(p,p)} \left(\left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)} \right) X_{(n,p)} \right)^T N_{(n,n)} \left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)} \right) X_{(n,p)} M_{(p,p)} \vec{u}^{\,j} \qquad [12]$$

De la mera observación de [10] y [12] se concluye que:

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

$$N_{**} = M_{(p,p)} \left(\left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)} \right) X_{(n,p)} \right)^T N_{(n,n)} \left(I_{(n,n)} - 1_n 1_n^T \; N_{(n,n)} \right) X_{(n,p)} M_{(p,p)}$$

Recordatorio

A título informativo hemos de recordar que:

$$M_* \textit{ se contempla en } R^{n^*}$$
$$M_{**} \textit{ se contempla en } R^{n}$$
$$N_* \textit{ se contempla en } R^{p^*}$$
$$N_{**} \textit{ se contempla en } R^{p}$$

De todo lo explicado se concluye que el **esquema de dualidad completo** en el **ACP à la française** presenta la estructura que mostraremos a continuación.

$$N_{**} = M_{(p,p)} \, \chi^F_{(p,n)} \, N_{(n,n)} \, \chi^C_{(n,p)} \, M_{(p,p)} \qquad M_* = \chi^C_{(n,p)} \, M_{(p,p)} \, \chi^F_{(p,n)}$$

$$\begin{tikzcd}
\vec{u}^{\,j} \in R^p \arrow[d, shift right, "M_{(p,p)}"'] & & \overset{N^{-1}}{\vec{w}^{\,j} \in R^{n^*}} \arrow[ll, "\chi^F_{(p,n)}"'] \arrow[d, shift right, "W_{(p,p)}"'] \\
\underset{M^{-1}_{(p,p)}}{\vec{v}^{\,j} \in R^{p^*}} \arrow[u, shift right, "V_{(p,p)}"'] \arrow[rr, "\chi^C_{(n,p)}"] & & \vec{c}_j \in R^n \arrow[u, shift right, "N_{(n,n)}"']
\end{tikzcd}$$

$$N_* = \chi^F_{(p,n)} \, N_{(n,n)} \, \chi^C_{(n,p)} \qquad M_{**} = N_{(n,n)} \, \chi^C_{(n,p)} \, M_{(p,p)} \, \chi^F_{(p,n)} \, N_{(n,n)}$$

Significado de la notación de este esquema de dualidad completo

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

$$\chi^{C}_{(n,p)} = \left(\chi^{F}_{(p,n)}\right)^T \qquad \chi^{F}_{(p,n)} = \left(\chi^{C}_{(n,p)}\right)^T$$

$$V_{(p,p)} = \chi^{F}_{(p,n)} \; N_{(n,n)} \; \chi^{C}_{(n,p)} \qquad W_{(n,n)} = \chi^{C}_{(n,p)} \; M_{(p,p)} \; \chi^{F}_{(p,n)}$$

El primer subíndice representa las filas y el segundo las columnas de las matrices.

Desarrollando la segunda parte que apuntábamos al principio del artículo, dentro del resumen (cálculo de los **ejes principales,** los **factores principales** y las **componentes principales)** tenemos los siguientes puntos:

2.1. Los ejes principales

Los **ejes principales** son los **vectores propios** de **VM M-ortonormados.**

$$VM \; \vec{u}^{\,j} = \lambda_j \vec{u}^{\,j} \qquad M_{(p\,,p)} - \textit{ortonormados } j=1,\ldots,p$$

Observación: El profesor Gilbert Saporta apunta que los ejes principales no tienen interés práctico (3) (p. 170).

2.2. Los factores principales

Los **factores principales** son los **vectores propios** de

$$MV \; \vec{v}^{\,j} = \lambda_j \vec{v}^{\,j} \quad M^{-1}_{(p,p)} \; - \textit{ortonormados} \quad j=1,\ldots,p$$

2.3. Las componentes principales

Las **componentes principales** son los **vectores propios** de

$$\chi^{C}_{(n,p)} M_{(p,p)} \chi^{F}_{(p,n)} N_{(n,n)} \vec{c}^{\,j}_{(n,1)} = \lambda_j \vec{c}^{\,j}_{(n,1)} \qquad N_{(n,n)} - \textit{ortogonales } j=1,\ldots,p$$

Así, en la práctica, para el cálculo de las **componentes principales** procederemos de la manera siguiente:

1.º Calcularemos los **vectores propios**

$$\vec{v}^{\,j}_{(p,1)} \qquad j=1,\ldots,p$$

mediante la **diagonalización** de la matriz **MV.**

2.º Calcularemos las **componentes principales** tal como mostramos a continuación:

$$\vec{c}^{\,j}_{(n,1)} = \chi^{C}_{(n,p)} \vec{v}^{\,j}_{(p,1)} \qquad j=1,\ldots,p$$

Con respecto al tercer punto del resumen: **demostrar que el cuadrado de la norma de los vectores**

$$\vec{c}^{\,j}_{(n,1)} \qquad j=1,\ldots,p$$

en el sentido de la métrica $N_{(n,\ n)}$ es igual a λ_j $j=1,\ldots,p$

bajo la condición de que el **cuadrado de la norma de los vectores**

$$\vec{v}^{\,j}_{(p,1)}$$

en el sentido de la

métrica $M^{-1}_{(p,p)}$ *es igual a 1.*

Demostración:

Partiremos de una expresión que es la que nos permitirá calcular los **ejes principales.**

$$\chi^{F}_{(p,n)} N_{(n,n)} \chi^{C}_{(n,p)} M_{(p,p)}\ \vec{u}^{\,j} = \lambda_j \vec{u}^{\,j} \qquad [1]$$

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

Pre-multiplicando ambos términos de [1] por $M_{(p,\ p)}$, tenemos que

$$M_{(p,p)} \chi^{F}_{(p,n)} N_{(n,n)} \chi^{C}_{(n,p)} M_{(p,p)} \vec{u}^{\,j} = \lambda_j M_{(p,p)} \vec{u}^{\,j}$$

y, además, teniendo en cuenta que se verifica

$$M\ \vec{u}^{\,j} = \vec{v}^{\,j}$$

llegamos a,

$$M_{(p,p)} \chi^{F}_{(p,n)} N_{(n,n)} \chi^{C}_{(n,p)}\ \vec{v}^{\,j}_{(p,1)} = \lambda_j \vec{v}^{\,j}_{(p,1)} \qquad [2]$$

Ahora, pre-multiplicando ambos términos de [2] por

$$M^{-1}_{(p,p)}$$

tendremos que,

$$\chi^F_{(p,n)} N_{(n,n)} \chi^C_{(n,p)} \ \vec{v}^j_{(p,1)} = \lambda_j M^{-1}_{(p,p)} \vec{v}^j \qquad [3]$$

Dado que se verifica:

$$\vec{c}^j_{(n,1)} = \chi^C_{(n,p)} \vec{v}^j_{(p,1)} \qquad [4]$$

haciendo uso de [3] y [4] estamos en condiciones de calcular

$$\vec{c}^{j^T}_{(1,n)} N_{(n,n)} \vec{c}^j_{(n,1)}$$

de la siguiente manera:

$$\vec{c}^{j^T}_{(1,n)} N_{(n,n)} \vec{c}^j_{(n,1)} = \vec{v}^{j^T} \chi^F_{(p,n)} N_{(n,n)} \chi^C_{(n,p)} \ \vec{v}^j = \vec{v}^{j^T} \lambda_j M^{-1}_{(p,p)} \vec{v}^j$$

Así pues, si se verifica

$$\vec{v}^{j^T} M^{-1}_{(p,p)} \vec{v}^j = 1$$

Concluimos que:

$$\vec{c}^{j^T}_{(1,n)} N_{(n,n)} \vec{c}^j_{(n,1)} = \lambda_j \qquad j=1,\ldots,p$$

CONSIDERACIONES FINALES

Esperamos que este artículo sirva para que los futuros investigadores dejen de aplicar —sistemáticamente— el **análisis en componentes principales** (ACP) a sus **datos empíricos,** teniendo en cuenta —tan solo—

*dos **métricas** en R^p y una en R^n*

Los investigadores deberán pedir a los informáticos que en su programa de **análisis en componentes principales** implementen la opción de **métricas.**

Esta será la única manera para que ellos mismos implementen la **métrica** que más se adecúe a sus datos empíricos con el fin de que obtengan unos resultados menos distorsionados con la realidad empírica.

AGRADECIMIENTOS

La realización de este artículo ha sido posible gracias a la colaboración de Jean-Baptiste Denis en los años setenta (1).

Agradecemos, a José María Navas Antón, Director del Departamento de Medio Ambiente, al cual pertenezco, la aceptación de la realización de este artículo que es, sin duda, básico para que tanto los investigadores de nuestro Departamento apliquen —fielmente— los conceptos que contiene, así como para cualquier otro Departamento en el que se aplique —correctamente—, el **análisis en componentes principales** en sus trabajos, tanto técnicos como, de investigación.

BIBLIOGRAFÍA

(1) Denis, J. B. y cols. (1976): *Apuntes redactados con ocasión del cursillo impartido en la Sección de Proceso de Datos sobre el Tratamiento de Datos.* INIA. MAPA.

(2) Bertier, P., y Bouroche, J.-M. (1975): *Analyse des données multidimensionnelles.* Préface de Georges Morlat. Presses Universitaires de France.

(3) Saporta, G. (1990): *Probabilités. Analyse des Données et Statistique.* Éditions Technip.

(4) Cailliez, F.; Pagès, J.-P.; Nakache, J.-P., y Mailles, J.-P. (1972): *Cours d'Analyse Multidimensionnelles.* Polycopié du CEEE.

(5) Escoufier, Y. (1979): *Cours d'Analyse des Données.* RT. 7901. CRIG av d'occitanie 34075 Montpellier Cedex.

(6) Cailliez, F., y Pagès, J.-P. et colls. (1976): *Introduction à l'Analyse des Données*. SMASH.

(7) Pearson, K. (1901): «On lines and planes of closest fit to systems of points in space». *Phil. Mag.,* n.º 11, pp. 1076-1076.

(8) Hotteling, H. (1933): «Analysis of a complex of Statistical variables into principal components». *J. Edud. Psy* 24, pp. 417-441, pp. 498-520.

(9) Benzécri, J.-P. et colls. (1973): *L'Analyse des Données. 2. L'Analyse des correspondances.* Dunod.

(10) Tenenhaus, M. (2007): *Statistique. Méthodes pour décrire, expliquer et prévoir.* Dunod.

(11) Harman, H. H. (1967): *Modern Factor Analysis* (2nd ed.). Chicago University Press, Chicago.

(12) Dazy, F., y Lebarzic, J.-F. (1996): *L'Analyse des données évolutives. Méthodes et applications.* Édition Technip.

(13) Díaz-Llanos, Fco. J.; Cayón, J., y Cermeño, C. (2009): «ACP de un triplete estadístico (X, Q, D) antes de una regresión lineal múltiple en una situación límite». *Anales de la RADE*, vol. 13, n.º 2, pp. 147-157.

(14) Foucart, Th. (1997): *L'Analyse des Données. Mode d'emploi.* Presses Universitaires de Rennes.

(15) Díaz-Llanos, Fco. J., y Valencia, J. L. (2004): *Métodos de predicción en situaciones límite.* Edit. LA MURALLA, S. A.

(16) Rivaud, J. (1978): *Algèbre linéaire. Classes Préparatoires et Université.* Tome 1. Exercises avec solutions. Vuilbert.