QÜESTIIÓ, Vol 12, nº 1, pp. 59-76, 1988

APLICACIONES DE LOS TEOREMAS DE SEPARACIÓN PARA VALORES SINGULARES DE MATRICES AL ANÁLISIS DE LA REDUNDANCIA

FRANCISCO CARMONA
Universidad de Barcelona

El análisis de la redundancia constituye una alternativa al análisis de la correlación canónica en el estudio de la relación entre dos grupos de variables.

La utilización de normas invariantes por matrices unitarias permite generalizar la definición de índice de la redundancia. Con los teoremas de separación para valores singulares de matrices se obtienen caracterizaciones similares del análisis de la redundancia y del análisis de la correlación canónica.

En el problema de la regresión reduciendo el rango también puede considerarse a las variables del análisis de la redundancia como óptimas predictoras, puesto que los primeros factores minimizan la matriz de covarianza residual.

Separation theorems for singular values of matrices applied to redundancy analysis.

Keywords: Redundancy analysis, canonical correlation, unitarily invariant norm, multivariate linear regression.

1. INTRODUCCIÓN

El análisis de la correlación canónica es una generalización, debida a Hotelling (1935, 1936), del concepto de correlación y regresión entre dos variables a dos

—Francisco Carmona - Universitat de Barcelona - Dep. d'Estadística - Avda. Diagonal, 645 - 08028 Barcelona
—Article rebut al desembre del 1987.

grupos de varibles. Este análisis ha sido estudiado y desarrollado por otros autores, como Horst (1961), Meredith (1964), etc., hasta situarlo en casi todas las obras sobre análisis multivariante como una de las técnicas preferentes.

Las correlaciones canónicas dan una medida del grado de dependencia entre dos conjuntos de variables, concepto que se ha desarrollado con el nombre de medidas de asociación multivariante. El trabajo de Cramer y Nicewader (1979) proporciona una extensa relación de estas medidas.

En otro sentido, Steward y Love (1968) definen un índice como una medida de predicción o redundancia. El llamado índice de la redundancia es la media de los coeficientes al cuadrado de la correlación múltiple para predecir las variables de un conjunto por el otro. Este índice, a diferencia de las correlaciones canónicas, es una medida en general no simétrica.

Wollemberg (1977) desarrolla el análisis de la redundancia. Así como el análisis de la correlación canónica maximiza la correlación entre dos combinaciones lineales de ambos grupos de variables, el análisis de la redundancia obtiene combinaciones lineales incorrelacionadas de las variables predictoras, con varianza uno, tales que la media de las correlaciones al cuadrado con las variables del otro grupo sea máxima. Es decir, se trata de maximizar el índice de la redundancia entre una combinación lineal del grupo de variables predictor y el otro conjunto.

Hasta el presente varios autores han estudiado diversos aspectos relacionados con el índice y el análisis de la redundancia, entre ellos cabe destacar las aportaciones de Gleason (1976), Johanson (1981), DeSarbo (1981), Tyler Tatsuoka (1982) y Dawson-Saunders y Tatsuoka (1983).

Este trabajo resume brevemente, con una visión propia, lo ya conocido sobre el tema y desarrolla nuevas aportaciones que pretenden situar el análisis de la redundancia como una técnica más del análisis multivariante, al mismo nivel que el análisis de la correlación canónica.

Para ello, en el siguiente apartado se define el índice de la redundancia en sus diferentes versiones y sus principales propiedades. Entre éstas destacaremos, como ya hemos dicho, la no simetría y el hecho de que sólo es invariante por transformaciones afines del grupo predictor como demostraron Dawson-Saunders y Tatsuoka (1983).

Seguidamente se explica el análisis de Wollenberg (1977) y las versiones de Johanson (1981) y Tyler (1982). Al igual que el índice de la redundancia, el análisis también depende de la matriz que utilicemos, covarianzas o correlaciones, cosa que no ocurre con el análisis de la correlación canónica dada su invariancia por transformaciones lineales no singulares de ambos grupos de variables, en particular cambios de escala.

En el apartado 3, las normas invariantes por matrices unitarias de Rao (1979) permiten generalizar la definición de índice de la redundancia, aunque el análisis de la redundancia para estos índices coincide con el clásico.

En el mismo trabajo, Rao demuestra varios teoremas de separación para valores singulares de una matriz, similares al teorema de separación de Poincaré. Estos teoremas permiten obtener teoremas de caracterización del análisis de la redundancia y del análisis de la correlación canónica.

Varios autores, entre ellos Yokai y García Ben (1980), han estudiado el problema de predecir un grupo de variables en función de otro, pero utilizando un número menor de combinaciones lineales que el número de variables predictoras observables, el llamado "reduced rank regression problem". Las primeras variables canónicas del análisis de la correlación canónica minimizan algunas funciones de la matriz de covarianzas residual, por lo que han sido calificadas de óptimas predictoras. Sin embargo, si descomponemos la matriz de covarianzas en dos partes, donde la primera es la matriz de covarianzas fija del error de predicción y la segunda es el error adicional por la reducción a un número inferior de combinaciones lineales, se demuestra que los primeros factores del análisis de la redundancia minimizan la segunda parte. De modo que las variables del análisis de la redundancia también son buenas predictoras.

2. ÍNDICES Y ANÁLISIS DE LA REDUNDANCIA

Sean x e y dos vectores aleatorios de dimensiones p y q respectivamente. Sea $C = (C_{ij})$ $i, j = 1, 2$ la matriz de varianzas–covarianzas de x, y. $R = (R_{ij})$ $i, j = 1, 2$ es la matriz de correlaciones.

El llamado índice de la redundancia se define, en principio, como la varianza media de las variables de un conjunto que se explica por una variable canónica del otro conjunto. Se llama índice de la redundancia total a la suma de los índices parciales para todas las variables canónicas. Sin embargo, los trabajos de Gleason (1976), Dawson–Saunders y Tatsuoka (1983) y Carmona (1985) justifican la siguiente definición.

DEFINICIÓN 1

Según la matriz utilizada, el índice de la redundancia es:

$$R(y : x) = \frac{1}{q} tr \left(R_{21} R_{11}^{-1} R_{12}\right)$$

$$R_c(y : x) = \frac{tr \left(C_{21} C_{11}^{-1} C_{12}\right)}{tr \left(C_{22}\right)}$$

Esta definición formaliza el hecho de que el índice representa la proporción de la varianza total de y explicada por la predicción lineal de y por x:

$$R_c(y:x) = \frac{\sum\limits_{i=1}^{q} \text{Var}\,(y_i) R^2_{y_i \cdot x}}{\sum\limits_{i=1}^{q} \text{Var}\,(y_i)}$$

El índice $R(y:x)$ puede verse así como la media de las correlaciones múltiples al cuadrado de cada variable y_i con el vector x.

$$R(y:x) = \frac{1}{q} \sum_{i=1}^{q} R^2_{y_i \cdot x}$$

Con estas representaciones de los índices obtenemos fácilmente las siguientes propiedades:

a) $0 \leq R(y:x) \leq 1 \qquad 0 \leq R_c(y:x) \leq 1.$

b) $R(y:x) = 0 \quad \Leftrightarrow \quad R^2_{y_i \cdot x} = 0 \qquad \forall i = 1, ..., q$

Análogamente es válida para $R_c(y:x)$.

Respectivamente para el valor 1.

c) Cuando las variables y son de varianza 1, $R(y:x) = R_c(y:x)$.

d) En general, $R(y:x) \neq R(x:y)$ y $R_c(y:x) \neq R_c(x:y)$.

e) $$R(Uy + h : Tx + k) = R(y:x)$$

si T es una matriz $p \times p$ no singular, U es $q \times q$ diagonal no singular y h, k vectores constantes.

f) $$R_c(Qy : Px) = R_c(y:x)$$

si P es una matriz $p \times p$ no singular y Q una matriz $q \times q$ ortogonal.

Según Wollenberg (1977) el análisis de la redundancia es "una alternativa al análisis de la correlación canónica".

Dicho análisis obtiene combinaciones lineales incorrelacionadas de las variables independientes $w_i'x$, con varianza uno, tales que la suma de las correlaciones al cuadrado con las variables y sea máxima.

Se trata pues de maximizar el índice de redundancia entre $w'x$ e y

$$R(y : w'x) = \frac{1}{q} \sum_{i=1}^{q} R^2_{y_i \cdot w'x} = \frac{1}{q} w' R_{12} R_{21} w$$

con la restricción $w' R_{11} w = 1$, donde ahora $R^2_{y_i \cdot w'x}$ es la correlación al cuadrado entre las variables y_i y $w'x$.

TEOREMA 1 Wollenberg (1977)

Las combinaciones lineales $w_1'x, w_2'x, ..., w_r'x$ que corresponden a los vectores propios normalizados $(w_i'R_{11}w_i = 1)$ de

$$R_{11}^{-1}R_{12}R_{21}w_i = \lambda_i w_i$$

donde $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_r > 0$ son los respectivos valores propios de la matriz $R_{11}^{-1}R_{12}R_{21}$ y r = rango $(R_{11}^{-1}R_{12}R_{21})$ = rango (R_{12}), verifican que $R\,(y : w_i'x) = \lambda_i$ es máximo; además si $j < i \quad w_j'R_{11}w_i = 0$.

λ_i puede interpretarse como q veces la varianza media de las variables de y que se explica por la i-esima variable canónica de las x.

Simétricamente se puede resolver el problema para el vector y, a partir de la ecuación

$$\left(R_{22}^{-1}R_{21}R_{12} - \mu I_q\right)v = 0$$

El análisis de la redundancia puede también realizarse con la matriz de covarianzas C y el índice $R_c(y : x)$. Hay que resolver

$$C_{11}^{-1}C_{12}C_{21}w = \lambda w$$

La correlación múltiple es un caso particular del análisis de la redundancia. Si $q = 1$, el análisis de la redundancia tiene por ecuación

$$\left(R_{11}^{-1}R_{12}R_{21} - \lambda I_p\right)w = 0$$

donde R_{12} es un vector columna

Al sustituir $\lambda = R_{2\cdot 1}^2 = R_{21}R_{11}^{-1}R_{12}$ y $w = \beta = R_{11}^{-1}R_{12}$ se verifica la igualdad.

También el análisis de componentes principales puede verse como un caso especial del análisis de la redundancia.

Si $x = y$.

$$\left(R_{11}^{-1}R_{11}R_{11} - \lambda I_p\right)w = 0 \Rightarrow \left(R_{11} - \lambda I_p\right)w = 0$$

En este caso el índice de redundancia verifica:

$$R(x : w'x) = \frac{1}{p}w'R_{11}R_{11}w = \frac{1}{p}\lambda\left(w'R_{11}w\right) = \frac{\lambda}{p}$$

luego es proporcional a la varianza de la primera componente principal.

Así, maximizar la redundancia $R(x : w'x)$ equivale a maximizar la varianza de una combinación lineal $w'x$ con la restricción usual $w'w = 1$.

El análisis de la redundancia también está relacionado con el análisis factorial. El análisis factorial canónico se obtiene de forma mucho más natural considerando el A.R. de las variables y los factores. El análisis del factor principal puede verse como el A.R. de las variables que generan el espacio de los factores comunes consigo mismas, ya que en el fondo se trata de sus componentes principales.

Estos temas, así como la generalización del A.R. a variables categóricas pueden verse con mayor extensión en Carmona (1987).

El análisis de la redundancia de Wollenberg no es biortogonal como el análisis de la correlación canónica. Las ecuaciones características del análisis

$$(R_{12}R_{21} - \lambda R_{11})\, w = 0$$
$$(R_{21}R_{12} - \mu R_{22})\, v = 0$$

no están relacionadas y, en general, $\lambda \neq \mu$.

simplemente se verifica

$$\text{corr}\left(w_i'x, w_j'x\right) = \delta_{ij}$$
$$\text{corr}\left(v_i'y, v_j'y\right) = \delta_{ij}$$

Johanson (1981) sugiere dos transformaciones alternativas para el conjunto de las y asociadas de forma natural con la transformación del conjunto de variables x.

La primera consiste en hallar combinaciones lineales $v_1'y, ..., v_r'y$ que maximicen $v_i'R_{21}w_i$ con las restricciones

$$v_i'R_{22}v_j = \delta_{ij}$$

La segunda maximiza $v_i'R_{21}w_i$ con las restricciones

$$v_i'v_j = \delta_{ij}$$

Para este segundo caso obtenemos el siguiente

TEOREMA 2 Johanson (1981)

Los v_i tales que maximizan $v_i'R_{21}w_i$ con la restricción $v_i'v_j = \delta_{ij}$ vienen dados por la expresión

$$v_i = \eta_i^{-1} R_{21} w_i$$

donde $\eta_i^2 = w_i' R_{12} R_{21} w_i = \lambda_i$

Este teorema puede expresarse de la siguiente forma

COROLARIO 1.2

$$\begin{pmatrix} W' & 0 \\ 0 & V' \end{pmatrix} \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} \begin{pmatrix} W & 0 \\ 0 & V \end{pmatrix} = \begin{pmatrix} I_r & D \\ D & V' R_{22} V \end{pmatrix}$$

donde W y V son las matrices que tienen como columnas a los vectores w_i y v_i respectivamente y D es la matriz diagonal cuyos elementos son $\lambda_i^{\frac{1}{2}}, i = 1, ..., r$.

Las matrices V y W pueden ampliarse, obteniendo como resultado el teorema 1 del trabajo de Tyler (1982).

TEOREMA 3 Tyler (1982)

Existen dos matrices, una no singular $\overline{W}$ $p \times p$ y otra ortogonal $\overline{V} q \times q$, de forma que

$$\begin{pmatrix} \overline{W}' & 0 \\ 0 & \overline{V}' \end{pmatrix} \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} \begin{pmatrix} \overline{W} & 0 \\ 0 & \overline{V} \end{pmatrix} = \begin{pmatrix} I_p & \triangle \\ \triangle' & \overline{V}' R_{22} \overline{V} \end{pmatrix}$$

donde

$$\triangle = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \qquad \text{de orden } p \times q$$

y D es la matriz diagonal del corolario 1.

3. GENERALIZACIÓN DEL ÍNDICE DE LA REDUNDANCIA PARA NORMAS DE MATRICES INVARIANTES POR MATRICES UNITARIAS

Rao (1979) extendió la definción de norma invariante por matrices unitarias (n.i.m.u.), introducida por von Neumann (1937) para matrices cuadradas, a matrices rectangulares. Utilizando n.i.m.u.'s podemos generalizar la definición de índice de la redundancia, aunque para todos estos índices el A.R. es el mismo.

El índice de la redundancia puede definirse como

$$R_c(y:x) = \frac{\| C_{21}C_{11}^{-\frac{1}{2}} \|_E^2}{\| C_{22}^{-\frac{1}{2}} \|_E^2}$$

donde

$$\| X \|_E = (tr(XX'))^{\frac{1}{2}}$$

es la norma euclídea de matrices.

La definición de n.i.m.u. para matrices rectangulares dada por Rao (1979) es

DEFINICIÓN 2

Una función $\| \quad \|$ a valores reales del espacio de las matrices $p \times q$ se llama n.i.m.u. si satisface las siguientes condiciones

(i) $\| X \| > 0 \quad \text{si} \quad X \neq 0$

(ii) $\| cX \| = |c| \, \| X \|$

(iii) $\| X + Y \| \leq \| X \| + \| Y \|$

(iv) $\| VXU \| = \| X \|$ para matrices unitarias V y U de órdenes p y q respectivamente.

Si $X = PDQ'$ es la descomposición en valores singulares (d.v.s.) de X, donde P y Q son unitarias y D es la matriz diagonal $p \times q$ con los valores singulares de X, por la condición (iv), $\| X \| = \| D \|$. Es decir, una n.i.m.u. es una función simétrica de los valores singulares.

Un ejemplo es la norma espectral $\| X \|_2 = \sigma_1$, donde σ_1^2 es el mayor valor propio de $X'X$, que es compatible con la norma euclídea de vectores

$$\| Xv \|_E \leq \| X \|_2 \| v \|_E$$

Dada una n.i.m.u., podemos generalizar el índice de la redundancia.

DEFINICIÓN 3

$$R_g(y:x) = \frac{\| C_{21}C_{11}^{-\frac{1}{2}} \|^2}{\| C_{22}^{-\frac{1}{2}} \|^2}$$

TEOREMA 4

El A.R. es invariante para cualquier índice $R_g(y:x)$.

Demostración:

Se trata de hallar una combinación lineal $a'x$ que maximice $R_g(y:a'x)$ con alguna restricción.

$$R_g(y:a'x) = \frac{\| C_{21}a\,(a'C_{11}a)^{-\frac{1}{2}} \|^2}{\| C_{22}^{-\frac{1}{2}} \|^2}$$

Si prescindimos del denominador y utilizamos la propiedad (ii) queda

$$(a'C_{11}a)^{-1} \| C_{21}a \|^2$$

Pero el único valor singular de $C_{21}a$ es $(a'C_{12}C_{21}a)^{\frac{1}{2}}$ con lo que

$$R_g(y:a'x) = K\frac{a'C_{12}C_{21}a}{a'C_{11}a}$$

cuya maximización, con la restrición $a'C_{11}a = 1$, nos devuelve el A.R. de Wollenberg.

4. TEOREMA DE CARACTERIZACIÓN DEL ANÁLISIS DE LA REDUNDANCIA

En este apartado utilizaremos la siguiente notación:

Los valores singulares de una matriz A se designan $\sigma_1(A) \geq \sigma_2(A) \geq \ldots$ y los valores propios de A cuando es simétrica $\lambda_1(A) \geq \lambda_2(A) \geq \ldots$.

La descomposición de Jordan de una matriz simétrica A es $A = P\Lambda P'$. La matriz de las k primeras columnas de P (k primeros vectores propios) se describe $P_{(k)}$ mientras que la matriz de las k últimas columnas es $P_{[k]}$.

A continuación enunciamos el teorema de separación de Poincaré para valores propios de una matriz simétrica.

TEOREMA 5 (T.S.P.)

Sea A una matriz simsétrica de orden m y sea B una matriz $m \times k$ tal que $B'B = I_k$. Entonces

$$\lambda_{m-k+i}(A) \leq \lambda_i(B'AB) \leq \lambda_i(A) \qquad i = 1, \ldots, k$$

La cota superior se alcanza cuando $B = P_{(k)}$ y la inferior cuando $B = P_{[k]'}$ donde $A = P\Lambda P'$.

Rao (1979) demuestra una generalización de T.S.P. para valores singulares de una matriz.

TEOREMA 6 Rao (1979)

Sean las matrices A $m\times n$, B $m\times r$ y C $n\times k$ tales que $B'B = I_r$ y $C'C = I_k$. Entonces

$$\sigma_{t+i}(A) \leq \sigma_i(B'AC) \leq \sigma_i(A) \qquad i = 1, ..., \min(r, k)$$

donde $t = m + n - r - k$

La cota superior se alcanza cuando $B = P_{(r)}$ y $C = Q_{(k)}$ donde $A = PDQ'$ es la d.v.s. de A.

Vamos a utilizar estos teoremas para dar una caracterización del AR.

En un claro paralelismo con el teorema 3 de Tyler, consideramos dos matrices F $p \times p$ y G $q \times q$ tales que $F'C_{11}F = I_p$ y $G'G = I_q$ de forma que la matriz de varianzas–covarianzas de $F'x$ con $G'y$ es

$$\begin{pmatrix} I_p & F'C_{12}G \\ G'C_{21}F & G'C_{22}G \end{pmatrix}$$

Vamos a ver que maximizar cualquier función de los valores singulares de $F'C_{12}G$ (en particular una n.i.m.u.) o de los valores propios de $G'C_{21}FF'C_{12}G$ nos lleva a la solución del A.R.

TEOREMA 7

Sean F $p \times p$ y G $q \times q$ tales que $F'C_{11}F = I_p$ y $G'G = I_q$. Entonces $\| F'C_{12}G \|$ se maximiza cuando $F = \overline{W}$ y $G = \overline{V}$ para cualquier n.i.m.u.

Demostración:

Recordemos que en el teorema de Tyler hemos visto que

$$\overline{W}'C_{12}\overline{V} = \triangle$$

donde $\overline{W}'C_{11}\overline{W} = I_p$ y $\overline{V}$ es unitaria.

Multiplicando a la izquierda por $\overline{W}$ y a la derecha por $\overline{V}'$ tenemos

$$C_{11}^{-1}C_{12} = \overline{W}\triangle\overline{V}'$$

o también

$$C_{11}^{-\frac{1}{2}}C_{12} = C_{11}^{\frac{1}{2}}\overline{W}\triangle\overline{V}'$$

que es la d.v.s. de $C_{11}^{-\frac{1}{2}}C_{12}$ dado que $C_{11}^{\frac{1}{2}}\overline{W}$ y $\overline{V}$ son unitarias.

Los valores singulares de $F'C_{12}G$ verifican:

$$\sigma_i\left(F'C_{12}G\right) = \sigma_i\left(F'C_{11}^{\frac{1}{2}}C_{11}^{-\frac{1}{2}}C_{12}G\right) \leq \sigma_i\left(C_{11}^{-\frac{1}{2}}C_{12}\right) \quad = \quad \sigma_i(\triangle)$$

La cota superior se alcanza cuando $C_{11}^{\frac{1}{2}}F = C_{11}^{\frac{1}{2}}\overline{W}$, es decir, cuando $F = \overline{W}$ y $G = \overline{V}$.

Nota 1:

Por la demostración queda claro que, bajo las condiciones del teorema, el A.R. maximiza cualquier función de los valores singulares de $F'C_{12}G$ o de los valores propios de $G'C_{21}FF'C_{12}G$ por ejemplo

$$\begin{aligned} \det\ &\left(G'C_{21}FF'C_{12}G\right) \\ tr\ &\left(G'C_{21}FF'C_{12}G\right) \end{aligned}$$

Nota 2:

Es posible tomar matrices F $p \times k$ y G $q \times r$ tales que $F'C_{11}F = I_k$ y $G'G = I_r$ y la maximación se obtiene cuando $F = \overline{W}_{(k)}$ y $G = \overline{V}_{(r)}$. En particular cuando $k = r =$ rango $\left(C_{12}\right)$, $F = W$ y $G = V$.

Es posible dar un teorema de caracterización del A.C.C. en términos similares.

Si ... $C_{11}^{-\frac{1}{2}}C_{12}C_{22}^{-\frac{1}{2}} = PDQ'$ es la d.v.s. del A.C.C., donde $D =$ diag $\left(\rho_1, ..., \rho_k, 0, ..., 0\right)$, $k =$rango $\left(C_{12}\right)$, resulta que $C_{11}^{-\frac{1}{2}}P$ y $C_{22}^{-\frac{1}{2}}Q$ son los vectores de la correlación canónica.

Teorema 8

Sean F $p \times p$ y G $q \times q$ tales que $F'C_{11}F = I_p$ y $G'C_{22}G = I_q$. Entonces $\parallel F'C_{12}G \parallel$ se maximiza para cualquier n.i.m.u. cuando $F = C_{11}^{-\frac{1}{2}}P$ y $G = C_{22}^{-\frac{1}{2}}Q$.

Demostración:

$$\sigma_i\left(F'C_{12}G\right)=\sigma_i\left(F'C_{11}^{\frac{1}{2}}C_{11}^{-\frac{1}{2}}C_{12}C_{12}^{-\frac{1}{2}}C_{22}^{\frac{1}{2}}G\right)$$

$$\leq \sigma_i\left(C_{11}^{-\frac{1}{2}}C_{12}C_{22}^{-\frac{1}{2}}\right)=\sigma_i(D)=\begin{cases}\rho_i & i\leq k\\ 0 & i>k\end{cases}$$

La cota superior se alcanza cuando $C_{11}^{\frac{1}{2}}F=P$ y $C_{22}^{\frac{1}{2}}G=Q$ es decir, $F=C_{11}^{-\frac{1}{2}}P$ y $G=C_{22}^{-\frac{1}{2}}Q$.

En este caso $F'x, G'y$ son las variables canónicas y verifican

$$\begin{aligned}\text{cov}\,(F'x,F'x)&=I_p\\ \text{cov}\,(G'y,G'y)&=I_q\\ \text{cov}\,(F'x,G'y)&=F'C_{12}G=D\end{aligned}$$

Nota 1

Bajo las condiciones de este teorema se observa en la demostración que el *A.C.C.* maximiza cualquier función de los valores propios de $F'C_{12}GG'C_{21}F$.

Nota 2

También es posible tomar matrices F $p\times r$ y G $q\times s$ tales que $F'C_{11}F=I_r$ y $G'C_{22}G=I_s$ y la maximación se obtiene cuando $F=C_{11}^{-\frac{1}{2}}P_{(r)}$ y $G=C_{22}^{-\frac{1}{2}}Q_{(s)}$.

Con este teorema podemos obtener de otra forma un resultado del trabajo de Rao (1979).

TEOREMA 9 Rao (1979)

Sean F $p\times p$ y G $q\times r$ tales que $F'C_{11}F=I_p$ y $G'C_{22}G=I_r$ y C una matriz $r\times p$. Entonces

$$\lambda_{r-i+1}[E(G'y-CF'x)(G'y-CF'x)']\geq 1-\rho_i^2 \quad i=1,\ldots,r$$

y la cota inferior se alcanza cuando

$$F=C_{11}^{-\frac{1}{2}}P,\quad G=C_{22}^{-\frac{1}{2}}Q_{(r)}\quad \text{y}\quad C=G'C_{21}F$$

Demostración:

Sea $A = G'C_{21}F$ y $M = E(G'y - CF'x)(G'y - CF'X)'$

Dado que

$$\begin{aligned} M &= I_r - AC' - CA' + CC' \\ &= I_r + (C - A)(C - A)' - AA' \end{aligned}$$

resulta que

$$\lambda_i(M) \geq \lambda_i(I_r - AA')$$

y se verifica la igualdad cuando $C = A = G'C_{21}F$

Por el teorema 8

$$\lambda_i(AA') = \sigma_i^2(A') \leq \rho_i^2$$

y se verifica la igualdad tomando F y G del enunciado.

Luego

$$\lambda_{r-i+1}(M) \geq \lambda_{r-i+1}(I_r - AA') = 1 - \lambda_i(AA') \geq 1 - \rho_i^2$$

5. LAS VARIABLES DEL A.R. COMO ÓPTIMAS PREDICTORAS

En el trabajo de Yokai y García Ben (1980) se estudia la utilización del A.C.C. para predecir y en función de x, pero utilizando $k \leq p$ combinaciones lineales de las x

$$z = F'x$$

donde F es $p \times k$.

Sin pérdida de generalidad podemos suponer $F'C_{11}F = I_k$.

Sea $y(z)$ la mejor predicción de y basada en z por mínimos cuadrados:

$$y(z) = C_{21}F\left(F'C_{11}F\right)^{-1}z = C_{21}Fz.$$

Se sabe que $y(z)$ es la mejor predicción lineal de y por z minimizando

a) $E \parallel y - y(z) \parallel_E^2$

b) det $(E(y - y(z))(y - y(z))')$

sobre todas las predicciones de la forma $y(z) = Az$ donde A es $q \times k$.

El vector z que miniza (a) difiere, en general, de las variables canónicas del A.C.C. (Ver Rao (1973), cap. 8 problema 2). En cambio la elección de z minimizando (b) coincide con las k primeras variables del A.C.C..

En el mismo sentido Rao (1979) estudia el "Problema de la regresión reduciendo el rango".

Veamos que el siguiente teorema resume ambos trabajos.

Sea $E(y - y(z))(y - y(z))' = C_{22} - C_{12}FF'C_{21} = M(F)$

TEOREMA 10

Si F es $p \times k$ tal que $F'C_{11}F = I_k$, entonces

$$\lambda_{q-i+1}\left(C_{22}^{-\frac{1}{2}}M(F)C_{22}^{-\frac{1}{2}}\right) \geq 1 - \rho_i^2 \qquad i = 1, \dots, k$$

y se alcanza la igualdad cuando $F = C_{11}^{-\frac{1}{2}}P_{(k)}$, es decir, cuando $z = F'x$ son las k primeras variables canónicas del A.C.C.

Demostración:

Dado que $\quad C_{22}^{-\frac{1}{2}}M(F)C_{22}^{-\frac{1}{2}} = I_q - C_{22}^{-\frac{1}{2}}C_{21}FF'C_{12}C_{22}^{-\frac{1}{2}}$

$$\begin{aligned}
\lambda_{q-i+1}\left(C_{22}^{-\frac{1}{2}}M(F)C_{22}^{-\frac{1}{2}}\right) &= 1 - \lambda_i\left(C_{22}^{-\frac{1}{2}}C_{21}FF'C_{12}C_{22}^{-\frac{1}{2}}\right) \\
&= 1 - \lambda_i\left(F'C_{12}C_{22}^{-1}C_{21}F\right) \\
&= 1 - \lambda_i\left(F'C_{11}^{\frac{1}{2}}C_{11}^{-\frac{1}{2}}C_{12}C_{22}^{-1}C_{21}C_{11}^{-\frac{1}{2}}C_{11}^{\frac{1}{2}}F\right) \\
&\geq 1 - \lambda_i\left(C_{11}^{-\frac{1}{2}}C_{12}C_{22}^{-1}C_{21}C_{11}^{-\frac{1}{2}}\right) \\
&= 1 - \rho_i^2 \qquad i = 1, \dots, k
\end{aligned}$$

y se calcanza la cota inferior cuando $C_{11}^{\frac{1}{2}}F = P_{(k)}$.

COROLARIO 1.10 Rao (1979)

Las k primeras variables canónicas del A.C.C. minimizan

$$\| C_{22}^{-\frac{1}{2}}M(F)C_{22}^{-\frac{1}{2}} \|$$

para cualquier n.i.m.u.

COROLARIO 2.10 Yokai y García Ben (1980)

Las k primeras variables canónicas del A.C.C. minimizan

$$\det\left(M(F)\right) \geq \det\left(C_{22}\right)\prod_{i=1}^{k}\left(1-\rho_i^2\right)$$

Demostración:

Trivial, ya que

$$\det\left(M(F)\right) = \det\left(C_{22}\right)\det\left(C_{22}^{-\frac{1}{2}}M(F)C_{22}^{-\frac{1}{2}}\right) \geq$$
$$\geq \det\left(C_{22}\right)\prod_{i=1}^{k}\left(1-\rho_i^2\right)$$

Por otra parte podemos considerar la descomposición de $M(F)$

$$M(F) = C_{22} - C_{21}C_{11}^{-1}C_{12} + C_{21}\left(C_{11}^{-1} - FF'\right)C_{12}$$

donde la primera parte es la matriz de covarianzas del error al predecir y por todo x, mientras que la segunda parte es el error adicional por la reducción de x a un número inferior de combinaciones lineales.

Rao (1979) estudia la minimización de la segunda parte que nos permite un importante resultado para el AR.

TEOREMA 11

Sea F una matriz $p \times k$ tal que $F'C_{11}F = I_k$. Entonces

$$\| C_{21}\left(C_{11}^{-1} - FF'\right)C_{12} \|$$

se minimiza para cualquier n.i.m.u. cuando $F = \overline{W}_{(k)}$.

Demostración:

Se demuestra fácilmente que

$$\lambda_i\left[C_{21}\left(C_{11}^{-1} - FF'\right)C_{12}\right] = \lambda_i\left(U'C_{11}^{-\frac{1}{2}}C_{12}C_{21}C_{11}^{-\frac{1}{2}}U\right)$$
$$\geq \lambda_{k+i}\left(C_{11}^{-\frac{1}{2}}C_{12}C_{21}C_{11}^{-\frac{1}{2}}\right)$$

por el T.S.P., donde U es una matriz subortogonal de rango $p-k$ tal que

$$UU' = I_p - C_{11}^{\frac{1}{2}} F F' C_{11}^{\frac{1}{2}}$$

La igualdad se alcanza cuando $U = (C_{11}^{\frac{1}{2}} \overline{W})_{[p-k]}$, es decir, cuando $F = \overline{W}_{(k)}$.

6. CONCLUSIONES

Se ha presentado el análisis de la redundancia como una alternativa al análisis de la correlación canónica destacando las diferencias conceptuales entre las correlaciones canónicas como medidas de dependencia y los índices de redundancia o predicción.

Se han estudiado las principales propiedades de los índices y el análisis de la redundancia en sus diferentess versiones, indicando las relaciones naturales de este análisis con otros métodos multivariantes, en particular con la regresión múltiple, componentes principales y análisis factorial.

Utilizando normas invariantes por matrices unitarias se ha generalizado la definición de índice de la redundancia aunque el análisis resultante coincide con el de Wollenberg.

Con los teoremas de separación para valores singulares de matrices se han obtenido teoremas de caracterización del análisis de la redundancia y del análisis de la correlación canónica.

Finalmente, se ha demostrado que en el problema de la regresión reduciendo el rango no sólo las variables canónicas pueden calificarse de óptimas predictoras, puesto que las variables del análisis de la redundancia también minimizan, aunque de otra forma, la matriz de covarianzas residual.

7. BIBILIOGRAFÍA

[1] **Carmona, F.** (1985), "Sobre el análisis de la redundancia". Homenatge a F. d'A. Sales, Fac. de Matem., Univ. de Barcelona, 16-27.

[2] **Carmona, F.** (1987) "Análisis de la redundancia y su relación con otros métodos multivariantes". Tesis Doctoral. Publicaciones de la Univ. de Barcelona.

[3] **Cramer, E.M.** y **Nicewander, W.A.** (1979) "Some symetric, invariant measures of multivariate association. Psychometrika, 44, 43-54.

[4] **Cuadras, C.M.** (1981) "Métodos de Análisis Multivariante". Eunibar, Barcelona.

[5] **Dawson–Saunders, B.K.** y **Tatsouka, M.M.** (1983) "The effect of affine transformation on redundancy analysis". Psychometrika, 48, 299-302.

[6] **Gleason, T.C.** (1976) "On redundancy in canonical analysis". Psychol. Bull., 83, 1004-1006.

[7] **Horst, P.** (1961) "Relations among m sets of measures". Psychometrika, 26(2), 129-149.

[8] **Hotelling, H.** (1935) "The most predictable criterion". J. of Educ. Psych., 26, 139-142.

[9] **Hottelling, H.** (1936) "Relations between two sets of variables". Biometrika, 28, 321-377.

[10] **Johanson, J.K.** (1981) "An extensison of Wollenberg's redundancy analysis". Psychometrika, 46, 93-103.

[11] **Meredith, W.** (1964) "Canonical correlations with fallible data. Psychometrika, 29, 55-56.

[12] **Rao, C.R.** (1965-1973) "Linear statistical inference and its applications". Jhon Wiley, New York, 625 pp.

[13] **Rao, C.R.** (1979) "Separations theorems for singular values of matrices and their applications in multivariate analysis". J. Mult. Anal., 9, 362-377.

[14] **Stewart, D.** y **Love, W.** (1968) "A general canonical correlation index." Psychol. Bull., 70, 160-163.

[15] **Tyler, D.E.** (1982) "On the optimality of the simultaneous redundancy transformations". Psychometrika, 47, 77-86.

[16] **van den Wollenberg, A.L.** (1977) "Redundancy analysis: An alternative for canonical correlation analysis". Psychometrika, 42, 207-219.

[17] **von Neumann, J.** (1937) "Some matrix irregulaties and metrization of matrix spaces". Tomsk Univ. Rev., 1, 286-299.

[18] **Yokai, V.J.** y **García Ben, M.S.** (1980) "Canonical variables as optimal predictors". The Annals of Statistics, 8(4), 865-869.