# ¿QUÉ TIENEN DE COMÚN CIERTAS TÉCNICAS ESTADÍSTICAS MULTIVARIADAS CONOCIDAS?

Hernando E. Mutis G.*


## RESUMEN

*Este documento se propone presentar los elementos comunes que tienen las técnicas de regresión múltiple multivariada, análisis multivariado de varianza (manova), análisis de correlación canónica y análisis discriminante de variables canónicas, prestando particular atención al tipo de prueba estadística y, en alguna medida, a la matemática que está detrás de estas técnicas.*


**Palabras clave:** regresión múltiple multivariada, análisis multivariado de varianza (manova), análisis de correlación canónica y análisis discriminante de variable canónica, lambda de Wilks.


## ABSTRACT

*This paper presents the common elements of the multivariate multiple regression, multivariate analysis of variance (manova), canonical correlation analysis and discriminant analysis of canonical variates focusing on the statistical tests and, in some extent, on the mathematics behind the techniques*


## 1. INTRODUCCIÓN.

Dentro de los métodos estadísticos multivariados más conocidos se pueden mencionar al análisis de regresión múltiple multivariado, análisis multivariado de varianza, análisis de correlación canónica, análisis discriminante lineal, análisis de componentes principales, análisis de correspondencias y análisis de factores. Estos métodos, ordinariamente, se presentan de una manera tal que parecería que no hubiesen relaciones importantes entre ellos que los doten una estructura común más allá de mencionar que las técnicas multivariadas se refieren a ciertos procedimientos que tratan un determinado número de observaciones medidas sobre múltiples variables con objetivos muy particulares.

De esta manera, la clasificación de las técnicas obedece, por ejemplo, a los objetivos que se plantean (Johnson and Wichern, 2002; Sharma, 1996; Bilodeau and Brenner, 1999): dependencias, interdependencias, niveles de asociaciones, agrupamientos, clasificación, proyecciones, representaciones, generación de puntuaciones, comparaciones y reducciones de dimensionalidad. En ocasiones, también por ejemplo, a combinaciones de estos objetivos (Johnson, 1999) y, a veces, la clasificación obedece al tipo de variables que las componen (Hair, Anderson *et al*, 1999).

* Profesor Asociado, Depertamento de Ingeniería Industrial, Universidad de Los Andes, Bogotá, Colombia. e-mail: hemutis@uniandes.edu.co

Este documento se propone presentar los elementos comunes que tienen las técnicas de regresión múltiple multivariada, análisis multivariado de varianza (manova), análisis de correlación canónica y análisis discriminante de variable canónica, prestando particular atención al tipo de prueba estadística y, en alguna medida, a la matemática que está detrás de estas técnicas.

Este material está organizado de la siguiente manera: la primera sección presenta una síntesis de cada una de las técnicas mencionadas junto con las pruebas estadísticas más relevantes de cada método; la segunda sección discute las relaciones entre las distintas técnicas y la tercera sección concluye.

## 2. Los métodos en discusión

En esta sección se presenta brevemente el propósito, la solución y la prueba estadística respectiva de los métodos estadísticos regresión multiple multivariada, manova, análisis de correlación canónica y análisis discriminante de variable canónica.

### 2.1 Regresión múltiple multivariada y "univariada"

En la regresión lineal múltiple se trata de encontrar una estructura de dependencia entre un grupo de variables explicatorias (agrupadas en una matriz X) con respecto a una variable dependiente (y) a través de encontrar un conjunto de parámetros (β) que definan esta estructura. El modelo se expresa como $y = X\beta + \varepsilon$, donde y es un vector aleatorio (nx1), X una matriz –considerada en principio como no aleatoria- de p variables independientes sobre n observaciones de tamaño nxp y de rango completo y, finalmente, e es un vector aleatorio residual el cual podría suponerse que se distribuye normalmente con vector de medias igual a cero y varianza $\sigma^2 I$, donde I es la matriz identidad de orden n. El proceso de estimación, bien por cuadrados mínimos ordinarios (CMO), bien por máxima verosimilitud (MV) arroja que $b = (X^tX)^{-1} X^ty$ donde b es el estimador del vector β.

Ahora, al considerar la versión multivariada del asunto (Rencher, 1995 y 1998), el modelo se reformula como $Y = X\beta + \varepsilon$, donde ahora Y es una matriz ( n x p ) de variables aleatorias dependientes, X es una matriz ( n x q ), b otra matriz de estimadores (q x p) y ε una matriz (n x p).

Análogamente al caso univariado el estimador de CMO (o el MV) se obtiene
$B = (X^tX)^{-1} X^tY$ donde B es el estimador de β.

La matriz B minimiza $(Y - XB)^t (Y - XB)$

$B = (X^tX)^{-1} X^tY = (X^tX)^{-1} X^t(y_1, \ldots, y_p) = [(X^tX)^{-1} X^t y_1, \ldots, (X^tX)^{-1} X^t y_p]$

$B = [ b_1, b_2, \ldots, b_p ]$, de tal forma que un vector $b_j = (X^tX)^{-1} X^t y_j$, el cual corresponde al caso particular inicialmente planteado de $y = X\beta + \varepsilon$

Un estimador para la covarianza del error del modelo $Y = X\beta + \varepsilon$ es:

$$\frac{E}{n-q-1} = \frac{(Y-XB)^t(Y-XB)}{n-q-1} = \frac{Y^tY - B^tX^tY}{n-q-1}$$

Si el modelo considera a las variables X centradas, es decir restándoles el efecto de la media, entonces las formas definicionales no cambian. Asúmase, ahora, que las X están centradas.

Multiplicando y divídiendo por (n –1) el lado derecho de la expresión siguiente:

$$B = (n-1)\frac{(X^tX)^{-1}X^tY}{n-1} = \left[\frac{X^tX}{n-1}\right]^{-1}\frac{X^tY}{n-1}$$

Ahora, como las matrices de covarianzas de X ($S_x$) y de X con Y ($S_{xy}$) se pueden expresar como

$$S_x^{-1} = \left[\frac{X^tX}{n-1}\right]^{-1}$$

$$S_{xy}^{-1} = \frac{X^tY}{n-1}$$

Entonces, en conclusión, el estimador se puede obtener como $B = S_x^{-1} S_{xy}$

Para obtener la prueba de hipótesis sobre las pendientes Ho: B = 0 se parte de
$Y^tY = Y^tY$, al cual se le suma y se le resta $B^tX^tY$:

$$Y^tY = Y^tY - B^tX^tY + B^tX^tY$$

Agrupando términos

$$Y^tY = (Y^tY - B^tX^tY) + B^tX^tY$$

Nótese que la parte en paréntesis de la anterior expresión corresponde al numerador de

$$\frac{E}{n-q-1} = \frac{Y^tY - B^tX^tY}{n-q-1}$$

Para obtener el equivalente de la "Suma total de cuadrados con respecto a la media" (o de cuadrados corregidos), se debe recordar que, en el caso "univariado":

$$y^ty - n\,\bar{y}^2 = \sum (y_i - \bar{y})^2 = \Sigma y_i^2 - n\bar{y}^2$$

Así, por analogía, en el caso "multivariado" se encuentra que, si a ambos lados de la siguiente expresión se le resta $n\,\bar{y}\bar{y}^t$, encontramos:

$$Y^tY = (Y^tY - B^tX^tY) + B^tX^tY$$

$$Y^tY - n\,\bar{y}\bar{y}^t = (Y^tY - B^tX^tY) + (B^tX^tY - n\,\bar{y}\bar{y}^t)$$

$$= \quad E \quad + \quad H$$

Donde:

E: es la matriz de covarianzas del error o matriz de sumas de cuadrados del error.

H: matriz de Sumas de cuadrados y productos cruzados de la regresión.

$Y^tY - n\,\bar{y}\bar{y}^t$: matriz de suma total de cuadrados (corregidos)

La forma específica de la prueba Ho. B = 0 se presentará al final de la siguiente sección por la razón allí explicada.

## 2.2 Anova y Manova

En los modelos simples de Anova se consideraba que las observaciones respecto a los tratamientos se expresaban como $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, donde $Y_{ij}$ se refiere a la iésima observación sobre el j-ésimo tratamiento, (i = 1, 2, ..., t tratamientos y j = 1, 2, ..., n observaciones por tratamiento); $\mu$ es la media general; $\tau_i$ es el efecto del $\tau$-ésimo tratamiento; $\varepsilon_{ij}$ es el residual aleatorio el cual se asume normal con media cero y varianza constante.

Para probar los efectos de tratamiento se formulaba Ho: $\mu_1 = \mu_2 = \ldots = \mu_t$
Y la idea era comparar las sumas de cuadrados de las desviaciones **entre** tratamientos con las sumas de cuadrados de desviaciones **dentro** de tratamientos:

SCE = Sumas de cuadrados **dentro** de tratamientos $= \dfrac{\sum_i \sum_y (y_{ij} - \bar{y}_{i.})^2}{t(n-1)}$

o Sumas de cuadrados del Error

SCT = SCH: Sumas de cuadrados **entre** tratamientos $= \dfrac{n\sum_i (\bar{y}_{i.} - \bar{y}_{..})^2}{t-1}$

o Suma total de Cuadrados

Y, dada la independencia entre estas sumas, se probaba con la razón F:
$[SCH / (t-1)] / [SCE / t(n-1)] \sim F_{(t-1,\, N-t)}$, bajo Ho. El Rechazo de la hipótesis nula sustenta la validez general del conjunto de tratamientos inicialmente considerado.

En Manova (análisis de varianza multivariado), la hipótesis nula se refiere ahora a vectores de medias, en lugar de medias singulares por tratamiento. La hipótesis nula toma la misma forma anterior, pero las $\mu_i$ se refieren a los vectores de medias o centroides del iésimo tratamiento (o iésimo grupo ).

Análogamente al caso univariado, las matrices de sumas de cuadrados dentro de tratamientos ( E ) y entre tratamientos ( H ) se definen como:

$$E = \sum\sum (y_{ij} - \bar{y}_{i.})(y_{ij} - \bar{y}_{i.})'$$

$$H = \sum\sum (\bar{\mathbf{y}}_{i.} - \bar{\mathbf{y}}_{..})(\bar{\mathbf{y}}_{i.} - \bar{\mathbf{y}}_{..})'$$

Ahora el estadístico "análogo" para probar Ho se concentra en la razón de la variabilidad de la matriz de las sumas de cuadrados de los errores (E) con respecto a la variabilidad de la matriz de sumas de cuadrados totales, T, la cual se expresa como T = E + H.

Estas variabilidades se expresan en términos de las varianzas generalizadas. Recuérdese que la noción de varianza generalizada de una matriz de covarianzas corresponde a su determinante. De lo que se trata, entonces, es de comparar la varianza generalizada de E con respecto a la varianza generalizada total ( E+H ):

$$\frac{|E|}{|E+H|}$$

Y a este estadístico se le conoce como el lambda ( Λ ) de Wilks. Bajo Ho, si este es "pequeño" – en términos de su distribución asociada–, se rechaza Ho.

Una relación importante es

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{E}+\mathbf{H}|} = \Pi \frac{1}{1+\lambda_i}$$

Donde los $\lambda_i$ son los eigenvalores de la matriz $E^{-1}H$

De aquí se derivan algunos estadísticos asociados muy importantes: la traza de Pillai, la traza de Hotelling-Lawley y la mayor raíz característica de Roy, cuyas definiciones vienen a continuación.

$$Pillai = \sum \frac{\lambda_i}{1+\lambda_i} \qquad Hotelling = \sum \lambda_i$$

$$Roy = \frac{\lambda_1}{1+\lambda_1}$$

En la regresión múltiple multivariada, no se expresó el estadístico para probar Ho: B = 0 y para este propósito se utilizará el mismo estadístico de Manova. Esto se debe a que el modelo de Manova es equivalente al de regresión con el conjunto apropiado de variables indicadoras para los tratamientos, por esto la presentación de la prueba se aplazó hasta este punto. De esta manera:

$$\begin{aligned} Y^tY - n\bar{y}\bar{y}^t &= (Y^tY - B^tX^tY) + (B^tX^tY - n\bar{y}\bar{y}^t) \\ &= \quad E \quad + \quad H \end{aligned}$$

Así que $\Lambda = \dfrac{|E|}{|E+H|} = \dfrac{Y^tY - B^tX^tY}{Y^tY - n\bar{y}\bar{y}^t}$

Se decía que los $\lambda_i$ son los eigenvalores de la matriz $E^{-1}H$, es decir, se está afirmando que el eigensistema asociado tiene que ser de la forma

$[E^{-1}H - \lambda I] a = 0$

o, visto alternativamente, $E^{-1}H a = \lambda a$

Premultiplicando por E a $[E^{-1}H - l I] a = 0$

$[E E^{-1}H - \lambda E] a = [H - \lambda E] a = 0$

Esta última expresión será retomada más adelante.

## 2.3 Análisis de Correlación canónica

En el análisis de correlación canónica se busca encontrar y analizar las relaciones entre dos conjuntos de variables (X y Y) de una manera tal que se encuentren los pares de combinaciones lineales máximos posibles (Sharma, página 412). De esta manera, definiendo a W = b'X como una combinación lineal de las variables en X (vector aleatorio px1) siendo b el vector de ponderaciones y V= a'Y la otra combinación lineal, pero de las variables en Y (vector aleatorio qx1) con a el vector de ponderaciones, se trata de encontrar a y b tales que se maximice la correlación entre W y V, sujeta a un conjunto de restricciones (varianzas de W y de V sean iguales a la unidad).

La salida al problema se encuentra a través de la solución del eigensistema:

$[ S_y^{-1} S_{yx} S_x^{-1} S_{xy} - \rho^2 I ]\ a = 0$

$[ S_x^{-1} S_{xy} S_y^{-1} S_{yx} - \rho^2 I ]\ b = 0$

Donde el subíndice de las matrices S se refiere a las matrices de covarianzas de X ($S_x$), de Y ($S_x$), de X con Y ($S_{xy}$) y de Y con X ($S_{yx}$) y $\rho^2$ es el cuadrado de la correlación canónica entre W y V; $\rho^2$ el valor propio asociado a la matriz $S_y^{-1} S_{yx} S_x^{-1} S_{xy}$ o a la matriz $S_x^{-1} S_{xy} S_y^{-1} S_{yx}$ y a y b son los vectores propios de cada una de ellas, respectivamente.

### 2.4. Análisis de Variable Canónica

Como parte de los métodos del análisis discriminante, el análisis de variable canónica se propone enfatizar las diferencia entre g grupos en un conjunto de p variables sobre n elementos (Krzanowski, 1996).

Si B es la matriz de covarianzas entre grupos y W la matriz de covarianzas dentro de grupos, la idea es encontrar los vectores a que maximizan la razón a´Ba / a´Wa. La solución del problema es a través del eigensistema $W^{-1}Ba = fa$, con f el valor propio de la matriz $W^{-1}B$ y a su eigenvalor asociado.

El análisis de variable canónica o análisis discriminante canónico es equivalente a la correlación canónica entre un grupo de variables cuantitativas (X) y un grupo de variables cualitativas codificadas como dummy o indicadoras (Y). Expresando a la matriz muestral de varianzas covarianzas total como

$$S = \begin{bmatrix} S_x & S_{xy} \\ S_{yx} & S_y \end{bmatrix}$$

Si g es el número de grupos, $n_j$ el número de observaciones en el grupo j y $S_j$ a la matriz de covarianzas muestral de las variables x en el j-ésimo grupo, entonces la matriz ponderada de covarianzas dentro de clases, $S_{pooled} = S_p$ se puede escribir como:

$$S_p = \frac{1}{\sum n_j - g} \sum (n_j - 1) S$$

Entonces las correlaciones canónicas $r_i$ son las raíces cuadradas de los eigenvalores $l_i$ de la matriz $S_p^{-1/2} S_{xy} S_y^{-1} S_{yx} S_p^{-1/2}$

## 3. Las relaciones entre las técnicas.

En esta sección se presentan las asociaciones entre los métodos mencionados. Primero se discuten las relaciones de la correlación canónica con la regresión múltiple multivariada, luego con el título de otra forma del lambda de Wilks se muestran las relaciones entre las pruebas estadísticas del análisis de correlación canónica con las de regresión y manova. Posteriormente se muestra los vínculos entre las pruebas básicas del análisis de correlación canónica y del análisis de variable canónica

### 3.1 Correlación canónica y regresión múltiple multivariada, relaciones entre $\rho^2$ y $\lambda$

En el análisis de correlación canónica, se decía que el cuadrado de la correlación canónica, $\rho^2$, proviene de la solución del eigensistema:

$[ S_y^{-1} S_{yx} S_x^{-1} S_{xy} - \rho^2 I ]\ a = 0$

Es decir, el $\rho^2$ es el eigenvalor asociado a la matriz $S_y^{-1} S_{yx} S_x^{-1} S_{xy}$

Este eigensistema se obtuvo ( premultiplicando por $S_y$ ) de:

$$[S_y S_y^{-1} S_{yx} S_x^{-1} S_{xy} - \rho^2 S_y ]\, a = 0$$

$$[ S_{yx} S_x^{-1} S_{xy} - \rho^2 S_y ]\, a = 0$$

Es decir, el eigensistema correspondiente es $S_{yx} S_x^{-1} S_{xy} a = \rho^2 S_y\ a$

Ahora, en regresión múltiple multivariada se encontraba que

1) $B = S_x^{-1} S_{xy}$

2) $Y^tY - n\bar{y}\bar{y}^t = ( Y^tY - B^tX^tY ) + (B^tX^tY - n\bar{y}\bar{y}^t)$
$\qquad = \qquad E \qquad + \qquad H$

y la matriz E podía expresarse como:

$$\frac{E}{n-q-1} = \frac{Y^tY - B^tX^tY}{n-q-1}$$

$$\frac{E}{n-1} = \frac{Y^tY}{n-1} - \frac{B^tX^tY}{n-1}$$

Atención: si también las Ys están centradas, entonces:

$$\frac{E}{n-1} = S_y - B^tS_{xy}$$

$$\frac{E}{n-1} = S_y - [S_x^{-1} S_{xy}]^t S_{xy}$$

$$\frac{E}{n-1} = S_y - S_{yx}S_x^{-1}S_{xy}$$

Y la matriz H puede expresarse como:

$$\frac{H}{n-1} = \frac{B^tX^tX}{n-1} = S_{yx}S_x^{-1}S_{xy}$$

Como se afirmaba anteriormente (al final de las sección 2.2), en tanto Manova es equivalente a regresión cuando se usan las dummies apropiadas, entonces

$$[ H - \lambda E ]\, a = 0$$

$$[ S_{yx} S_x^{-1} S_{xy} - \lambda (S_y - S_{yx} S_x^{-1} S_{xy})]\, a = 0$$

Es decir: $S_{yx} S_x^{-1} S_{xy}\ a = \lambda\, [S_y - S_{yx} S_x^{-1} S_{xy} ]\, a$

Ahora, substrayendo a $\rho^2 S_{yx} S_x^{-1} S_{xy} a$ de $S_{yx} S_x^{-1} S_{xy} a = \rho^2 S_y\ a$

$$S_{yx} S_x^{-1} S_{xy} a - \rho^2 S_{yx} S_x^{-1} S_{xy} a = \rho^2 S_y\ a - \rho^2 S_{yx} S_x^{-1} S_{xy} a$$

$$[S_{yx} S_x^{-1} S_{xy} - \rho^2 S_{yx} S_x^{-1} S_{xy} ]\, a = [\rho^2 S_y - \rho^2 S_{yx} S_x^{-1} S_{xy}]\, a$$

$$(1 - \rho^2)\ S_{yx} S_x^{-1} S_{xy} a = \rho^2 [S_y - S_{yx} S_x^{-1} S_{xy}]\, a$$

$$S_{yx}S_x^{-1}S_{xy}a = \frac{\rho^2}{1-\rho^2}\left[S_y - S_{yx}S_x^{-1}S_{xy}\right]a$$

Y comparando con

$$S_{yx} S_x^{-1} S_{xy}\ a = \lambda\ [S_y - S_{yx} S_x^{-1} S_{xy} ]\, a$$

nos muestra que $\lambda_i = \dfrac{\rho_i^2}{1-\rho_i^2}$

Y/o que $\rho_i^2 = \dfrac{\lambda_i}{1+\lambda_i}$

Los estadísticos pueden presentarse como...

$$Pillai = \sum \frac{\lambda_i}{1+\lambda_i} = \sum \rho_i^2$$

$$Hotelling = \sum \lambda_i = \sum \frac{\rho_i^2}{1-\rho^2}$$

$$Roy = \frac{\lambda_1}{1+\lambda_1} = \rho_1^2$$

La matriz $E^{-1}H$ es, entonces..

$$E^{-1}H = [\, S_y - S_{yx} S_x^{-1} S_{xy} \,]^{-1} S_{yx} S_x^{-1} S_{xy}$$

Alternativamente, al premultiplicar por $S_y$ a E y por $S_y^{-1}$ a H:

$$E^{-1}H = [\, S_y^{-1} S_y - S_y^{-1} S_{yx} S_x^{-1} S_{xy} \,]^{-1} S_y^{-1} S_{yx} S_x^{-1} S_{xy}$$

$$E^{-1}H = [\, I - S_y^{-1} S_{yx} S_x^{-1} S_{xy} \,]^{-1} S_y^{-1} S_{yx} S_x^{-1} S_{xy}$$

## 3.2 **Otra forma del lambda de Wilks.**

Sean las matrices $\Sigma$ (covarianza poblacional), S (covarianza muestral) y $\hat{\Sigma}$:

$$\Sigma = \begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_y \end{bmatrix} \qquad S = \begin{bmatrix} S_x & S_{xy} \\ S_{yx} & S_y \end{bmatrix}$$

$$\hat{\Sigma} = \begin{bmatrix} S_x & 0 \\ 0 & S_y \end{bmatrix}$$

Una prueba acerca de la conveniencia de aplicar análisis de correlación canónica es que la matriz poblacional de covarianzas entre el grupo de variables X y el grupo de variables Y toma la forma (Mardia *et al*, 1982):

Ho: $S_{xy} = 0$

El determinante de

$$\hat{\Sigma}^{-1}S : |\,\hat{\Sigma}^{-1}S\,| = \quad \Lambda = \frac{|S|}{|\hat{\Sigma}|} = \frac{\begin{vmatrix} S_x & S_{xy} \\ S_{yx} & S_u \end{vmatrix}}{\begin{vmatrix} S_x & 0 \\ 0 & S_u \end{vmatrix}}$$

A $\Lambda$, también se le conoce como la razón de verosimilitud y esto se debe a que el numerador de la expresión anterior hace relación al estimador máximo-verosímil de la matriz de covarianzas $\Sigma$ sin la restricción de que $\Sigma_{xy} = 0$. Entre tanto, el denominador hace relación al estimador máximo-verosímil de la matriz de covarianzas S sujeto a la restricción de que $\Sigma_{xy} = 0$.

Es decir $\Lambda$ compara estos dos estimadores máximo-verosímiles; si $S_{xy}$ o $S_{yx}$ son de verdad "pequeñitos", entonces $\Lambda$ tiende a ser la unidad y no se rechaza Ho. Pero si no son "pequeñitos", entonces $\Lambda$ tiende a ser mucho menor que la unidad y, en ese caso, se rechaza Ho. (La medida de "pequeñez" se verá más adelante).

Entre tanto, véanse algunos resultados interesantes. $\hat{\Sigma}^{-1}S$ es:

$$\hat{\Sigma}^{-1}S = \begin{bmatrix} S_x & 0 \\ 0 & S_y \end{bmatrix}^{-1} \begin{bmatrix} S_x & S_{xy} \\ S_{yx} & S_y \end{bmatrix} = \begin{bmatrix} I & S_x^{-1}S_{xy} \\ S_y^{-1}S_{yx} & I \end{bmatrix}$$

El determinante de $\hat{\Sigma}^{-1}S$ se puede expresar como:

$$\Lambda = |\,\hat{\Sigma}^{-1}S\,| = |\, I - S_y^{-1}S_{yx}S_x^{-1}S_{xy} \,| = \Pi(1-\rho_i^2)$$

Donde los $\rho_i^2$ son los eigenvalores de $S_y^{-1}S_{yx}S_x^{-1}S_{xy}$

De otra parte el determinante de $\hat{\Sigma}^{-1}S$ también se puede escribir como:

$$\Lambda = |\,\hat{\Sigma}^{-1}S\,| = \frac{|S|}{|S_y||S_x|} = \frac{|S_Y - S_{YX}S_x^{-1}S_{XY}|}{|S_y|} = \frac{|S_x - S_{xy}S_y^{-1}S_{yx}|}{|S_x|} = |\, I - S_y^{-1}S_{yx}S_x^{-1}S_{xy} \,| = \Pi(1-\rho_i^2)$$

Lo cual se obtuvo haciendo uso de algunas propiedades de los determinantes (ver Apéndice)

Para demostrar por qué el $\Lambda = \Pi\,(1 - r_i^2)$, recordemos que

$$S_y^{-1} S_{yx} S_x^{-1} S_{xy}\, a = \rho^2 a$$

Además, $\lambda_i = \dfrac{\rho_i^2}{1-\rho_i^2}$ y/o $\rho_i^2 = \dfrac{\lambda_i}{1+\lambda_i}$

Expresemos el determinante de la siguiente manera:

$$| I - S_y^{-1} S_{yx} S_x^{-1} S_{xy} | = \Pi\, \phi_i$$

Donde los $\phi_i$ son los eigenvalores de la matriz $I - S_y^{-1} S_{yx} S_x^{-1} S_{xy}$

El eigensistema asociado se puede expresar como

$[I - S_y^{-1} S_{yx} S_x^{-1} S_{xy}]\, a = \phi\, a$

$a - S_y^{-1} S_{yx} S_x^{-1} S_{xy}\, a = \phi\, a$

y la sección en rojo $S_y^{-1} S_{yx} S_x^{-1} S_{xy}\, a = \rho^2 a$.

Por tanto: $a - \rho^2 a = \phi\, a$

Entonces $(1 - \rho^2)\, a = \phi\, a$, así que

$\phi = 1 - \rho^2$

Por eso $| I - S_y^{-1} S_{yx} S_x^{-1} S_{xy} | = \phi_i = \Pi(1 - \rho_i^2)$

Para determinar la "pequeñez" del estadístico se usa la aproximación de Bartlett en la cual $-[n-1/2(p+3)]\mathrm{Ln}\, \Lambda$ se distribuye como $c^2$ con grados de libertad dados por el producto del número de variables en X y el número de variables en Y.

### 3.3 Relaciones entre el análisis de variable canónica y correlación canónica

Se decía en la sección 2.4 que el análisis de variable canónica o análisis discriminante canónico es equivalente a la correlación canónica entre un grupo de variables cuantitativas (X) y un grupo de variables cualitativas codificadas como dummy o indicadoras (Y). Entonces las correlaciones canónicas $r_i$ son las raíces cuadradas de los eigenvalores $\lambda_i$ de la matriz $S_p^{-1/2} S_{xy} S_y^{-1} S_{yx} S_p^{-1/2}$.

Ahora, para la prueba estadística multivariada $E^{-1}H$ las matrices respectivas son

$$E = (n - 1)\, [\, S_y - S_{yx} S_x^{-1} S_{xy}\, ]$$

$$H = (n - 1)\, S_{yx} S_x^{-1} S_{xy}$$

Donde n es el número total de observaciones. Nótese la similitud de esta presentación con las formas anteriores del estadístico análogo.

## 4. Conclusión

El esfuerzo de sintesis de los métodos estadísticos multivariados que se esbozó en este documento permitió apreciar la fuerte similitud de las pruebas estadísticas presentadas para validar (o invalidar) los propósitos de cada procedimiento. En todos los casos, el problema central de cada método implica la descomposición espectral de las matrices en sus valores y vectores propios. Más aún, se establecieron claramente relaciones entre el lambda de Wilks, dos de sus presentaciones y las formas particulares que revisten las matrices E y H. También se mostraron las relaciones entre los eienvalores de diversos procedimientos ($\lambda$ y $\rho$) que, junto con lo anterior, dan lugar a pensar estas técnicas como variaciones sobre un mismo tema.

## Apéndice

Algunas propiedades de los determinantes:

$$\begin{vmatrix} S_x & S_{xy} \\ S_{yx} & S_y \end{vmatrix} = |S_x|\,|S_y - S_{yx}S_x^{-1}S_{xy}| = |S_y|\,|S_x - S_{xy}S_y^{-1}S_{yx}|$$

$$| AB | = | A |\, | B |$$

$$| A^{-1} | = 1 / | A |$$

Para A (pxn) y C(nxp) con A(pxp) no singular:

$$|A + BC| = |A|\,|I_p + A^{-1}BC| = |A|\,|I_n + CA^{-1}B|$$

## Referencias


Bilodeau, M. and Brenner, D. (1999). Theory of Multivariate Statistics, Springer-Verlag, New York

Hair, Anderson, Tatham y Black (1999) Análisis Multivariante, Quinta edición, Prentice Hall, España

Johnson, D. (2000). Métodos Multivariados aplicados al análisis de datos, Thomson, México.

Johnson, R. and Wichern, D. (2002). Applied Multivariate Statistical Analysis. Fifth Edition, Prentice Hall, New Jersey.

Krzanowski, W.J. (1996) Principles of Multivariate Analysis. A User´s Perspective, Clarendo Press, Oxford, Belfast

Mardia, K.V.J., Kent, J. T. and BibbY, J. M. (1982) Multivariate Analysis, Third Printing, Academic Press, Inc., Orlando.

Rencher, Alvin C. (1995). Methods of Multivariate Analysis, Wiley, New York.

Rencher, Alvin C. (1998). Multivariate Statistical Inference and Applications, Wiley, New York.

Sharma, S. (1996). Applied Multivariate Techniques, Wiley, New York.