UNIVERSIDAD DE ALCALÁ

Análisis de Correlación Canónica

Un Ejemplo

- **Gastos y características socio-económicas de las Familias**

- **variables:**

– Gastos en Alimentación	- Número de miembros del hogar
– Gastos en Vestidos	- Nivel de ingresos del hogar
– Gastos en Vivienda	- Nivel de estudios del sustentador pral
– Gastos en Transporte	- Edad media de los miembros del hogar
– Gastos en Sanidad	- Número de miembros mayores de edad
– Gastos en Educación	- Numero de perceptores del hogar
–	- ....

- **¿Existe alguna relación de dependencia entre los hábitos de consumo de los hogares y las características socio-económicas de los mismos?**

Dpto. de Estadística Económica, Estructura Económica y O.E.I.
© Fco. Javier Callealta Barroso

UNIVERSIDAD DE ALCALÁ

Análisis de Correlación Canónica

Objetivo

- Mediante regresión lineal univariante se trata de explicar una variable Y, (*variable dependiente, endógena o explicada),* a partir devarias variables X_1, X_2, ..., X_p, (*variables independientes, exógenas o explicativas)*, mediante una relación lineal, en el caso general, de la forma:

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

  o, si las variables estuviesen centradas en el origen:

$$\hat{Y} = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

- El Análisis de Correlación Canónica trata de extender esta idea cuando, además del conjunto de variables exógenas $\{ X_1, X_2, \ldots, X_p \}$ hay, no sólo una variable Y, sino un conjunto de variables endógenas $\{Y_1, Y_2, \ldots, Y_q\}$, para lo que trata de encontrar unos coeficientes $\alpha_1^{(k)}, \alpha_2^{(k)}, \ldots, \alpha_q^{(k)}$ y $\beta_1^{(k)}, \beta_2^{(k)}, \ldots, \beta_P^{(k)}$, que proporcionen las mejores relaciones de aproximación del tipo:

$$\alpha_1^{(k)} Y_1 + \alpha_2^{(k)} Y_2 + \ldots + \alpha_q^{(k)} Y_q = Y^{(k)} \cong X^{(k)} = \beta_1^{(k)} X_1 + \beta_2^{(k)} X_2 + \ldots + \beta_P^{(k)} X_p$$

Análisis de Correlación Canónica

Variables canónicas

- Notando por $\alpha^{(k)}$ y $\beta^{(k)}$ los correspondientes vectores columna de coeficientes que intervienen abajo, el problema del Análisis de Correlación Canónica será encontrar una serie de nuevas variables tipificadas e incorrelacionadas en cada uno de los dos grupos, $Y^{(k)}$ y $X^{(k)}$, que llamaremos *variables canónicas,*

$$Y^{(k)} = \alpha^{(k)'} y = \alpha_1^{(k)} Y_1 + \alpha_2^{(k)} Y_2 + \ldots + \alpha_q^{(k)} Y_q$$
$$X^{(k)} = \beta^{(k)'} x = \beta_1^{(k)} X_1 + \beta_2^{(k)} X_2 + \ldots + \beta_P^{(k)} X_p$$

de forma que sea máxima su correlación en términos absolutos (el cuadrado del coeficiente de correlación lineal):

$$Max \quad \rho^2 (Y^{(k)}, X^{(k)})$$

Análisis de Correlación Canónica

Cálculo de la primera pareja de variables canónicas

- Notas previas:
  - Al ser centradas las variables $Y_1, Y_2, \ldots, Y_q$ y $X_1, X_2, \ldots, X_p$, lo serán en consecuencia las nuevas variables canónicas y , ya que son combinaciones lineales sin términos independientes de las anteriores
  - La matriz de varianzas y covarianzas de las variables toma la forma:

$$\Sigma = E[(x;y)(x;y)'] = E\begin{bmatrix} xx' & xy' \\ yx' & yy' \end{bmatrix} = \begin{pmatrix} E[xx'] & E[xy'] \\ E[yx'] & E[yy'] \end{pmatrix} = \begin{pmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_y \end{pmatrix}$$

  - Y al exigirles como condición que se encuentren tipificadas e incorrelacionadas, estaremos imponiendo las siguientes condiciones sobre los respectivos parámetros:

$$Var(Y^{(k)}) = 1 \Leftrightarrow E\left[(\alpha^{(k)'} y)(\alpha^{(k)'} y)'\right] = \alpha^{(k)'} E[yy'] \alpha^{(k)} = 1 \Leftrightarrow \alpha^{(k)'} \Sigma_y \alpha^{(k)} = 1$$
$$Cov(Y^{(l)}, Y^{(k)}) = 0 \Leftrightarrow E\left[(\alpha^{(l)'} y)(\alpha^{(k)'} y)'\right] = \alpha^{(l)'} E[yy'] \alpha^{(k)} = 0 \Leftrightarrow \alpha^{(l)'} \Sigma_y \alpha^{(k)} = 0, \quad l \neq k$$

$$Var(X^{(k)}) = 1 \Leftrightarrow E\left[(\beta^{(k)'} x)(\beta^{(k)'} x)'\right] = \beta^{(k)'} E[xx'] \beta^{(k)} = 1 \Leftrightarrow \beta^{(k)'} \Sigma_x \beta^{(k)} = 1$$
$$Cov(X^{(l)}, X^{(k)}) = 0 \Leftrightarrow E\left[(\beta^{(l)'} x)(\beta^{(k)'} x)'\right] = \beta^{(l)'} E[xx'] \beta^{(k)} = 0 \Leftrightarrow \beta^{(l)'} \Sigma_x \beta^{(k)} = 0, \quad l \neq k$$

Análisis de Correlación Canónica
Cálculo de la primera pareja de variables canónicas

- El coeficiente de correlación lineal de Fisher para las primeras variables canónicas será:

$$\rho^2(X,Y)=\frac{Cov^2(Y,X)}{Var(Y)\cdot Var(X)}=\frac{E^2\left[(\alpha' y)(\beta' x)'\right]}{(\alpha'\Sigma_y\alpha)(\beta'\Sigma_x\beta)}=\frac{(\alpha' E\left[yx'\right]\beta)^2}{(\alpha'\Sigma_y\alpha)(\beta'\Sigma_x\beta)}=\frac{(\alpha'\Sigma_{yx}\beta)^2}{(\alpha'\Sigma_y\alpha)(\beta'\Sigma_x\beta)}$$

- por lo que el problema de obtener las primeras variables canónicas puede expresarse como sigue:

$$\text{* Encontrar } \alpha \text{ y } \beta \text{ tales que : } \quad Var(Y)=Var(X)=1 \quad y \quad \rho^2(X,Y) \text{ sea máximo}$$

o equivalentemente:

$$\text{Maximizar } (\alpha'\Sigma_{yx}\beta)^2 \quad \text{sujeto a las restricciones : } \begin{cases}\alpha'\Sigma_y\alpha=1\\ \beta'\Sigma_x\beta=1\end{cases}$$

Análisis de Correlación Canónica
Cálculo de la primera pareja de variables canónicas

$$L=(\alpha'\Sigma_{yx}\beta)^2-\lambda\cdot(\alpha'\Sigma_y\alpha-1)-\mu\cdot(\beta'\Sigma_x\beta-1)$$

$$\left.\begin{aligned}\frac{\partial L}{\partial\alpha}&=2(\alpha'\Sigma_{yx}\beta)\Sigma_{yx}\beta-2\lambda\cdot\Sigma_y\alpha=0\Leftrightarrow(\alpha'\Sigma_{yx}\beta)^2=\lambda\alpha'\Sigma_y\alpha=\lambda\\ \frac{\partial L}{\partial\beta}&=2(\alpha'\Sigma_{yx}\beta)\alpha'\Sigma_{yx}-2\mu\beta'\Sigma_x=0\Leftrightarrow(\alpha'\Sigma_{yx}\beta)^2=\mu\beta'\Sigma_x\beta=\mu\end{aligned}\right\}\Rightarrow\lambda=\mu=(\alpha'\Sigma_{yx}\beta)^2=\rho^2$$

- de donde

$$\Sigma_{yx}\beta=(\alpha'\Sigma_{yx}\beta)\Sigma_y\alpha\Leftrightarrow\alpha=\frac{\Sigma_y^{-1}\Sigma_{yx}\beta}{(\alpha'\Sigma_{yx}\beta)}$$

$$\alpha'\Sigma_{yx}=(\alpha'\Sigma_{yx}\beta)\beta'\Sigma_x\Leftrightarrow\Sigma_{xy}\alpha=(\alpha'\Sigma_{yx}\beta)\Sigma_x\beta\Leftrightarrow\beta=\frac{\Sigma_x^{-1}\Sigma_{xy}\alpha}{(\alpha'\Sigma_{yx}\beta)}$$

- lo que conduce a

$$\beta'\Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx}\Sigma_x^{-1}=(\alpha'\Sigma_{yx}\beta)^2\beta'=\rho^2\beta'\Leftrightarrow(\Sigma_x^{-1}\Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx})\beta=\rho^2\beta$$

$$(\Sigma_y^{-1}\Sigma_{yx}\Sigma_x^{-1}\Sigma_{xy})\alpha=(\alpha'\Sigma_{yx}\beta)^2\alpha=\rho^2\alpha$$

Análisis de Correlación Canónica

Cálculo de la primera pareja de variables canónicas

- Por lo que:

$$\beta \text{ es un autovector de la matriz } \Sigma_x^{-1}\Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx}$$
$$\alpha \text{ es un autovector de la matriz } \Sigma_y^{-1}\Sigma_{yx}\Sigma_x^{-1}\Sigma_{xy}$$

- esas matrices, de dimensiones p·p y q·q respectivamente tienen el mismo número (Min(p,q)) de autovalores no nulos (ρ^2) y coinciden dos a dos

- se define el *primer coeficiente de correlación canónica* como el coeficiente de correlación lineal simple de Fisher, ρ, entre dichas dos primeras variables canónicas

- Además,

$$\alpha = \frac{\Sigma_y^{-1}\Sigma_{yx}\beta}{(\alpha'\Sigma_{yx}\beta)} = \frac{\Sigma_y^{-1}\Sigma_{yx}\beta}{\rho} \quad \text{y} \quad \beta = \frac{\Sigma_x^{-1}\Sigma_{xy}\alpha}{(\alpha'\Sigma_{yx}\beta)} = \frac{\Sigma_x^{-1}\Sigma_{xy}\alpha}{\rho}$$

Análisis de Correlación Canónica

Cálculo de las sucesivas variables canónicas

- Supongamos que ya disponemos de *r-1* parejas de variables canónicas, que ordenamos matricialmente de la siguiente forma:

$$A_{(r-1)} = \begin{pmatrix} \alpha_1^{(1)} & \alpha_1^{(2)} & \dots & \alpha_1^{(r-1)} \\ \alpha_2^{(1)} & \alpha_2^{(2)} & \dots & \alpha_2^{(r-1)} \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ \alpha_q^{(1)} & \alpha_q^{(2)} & \dots & \alpha_q^{(r-1)} \end{pmatrix} \quad , \quad B_{(r-1)} = \begin{pmatrix} \beta_1^{(1)} & \beta_1^{(2)} & \dots & \beta_1^{(r-1)} \\ \beta_2^{(1)} & \beta_2^{(2)} & \dots & \beta_2^{(r-1)} \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ \beta_p^{(1)} & \beta_p^{(2)} & \dots & \beta_p^{(r-1)} \end{pmatrix}$$

- Estas variables canónicas cumplirán, por las exigencias anteriores que sus correlaciones internas en cada grupo serán nulas; lo que podemos expresar como:

$$A'_{(r-1)}\Sigma_y A_{(r-1)} = I \quad \text{y que} \quad B'_{(r-1)}\Sigma_x B_{(r-1)} = I$$

- y que las correlaciones cruzadas con las anteriores sean nulas; lo que podemos expresar como

$$A'_{(r-1)}\Sigma_{yx}B_{(r-1)} = \begin{pmatrix} \rho_{(1)} & 0 & & 0 \\ 0 & \rho_{(2)} & \dots & 0 \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ 0 & 0 & \dots & \rho_{(r-1)} \end{pmatrix}_{(r-1)\cdot(r-1)} = \mathrm{P}_{(r-1)}$$

Análisis de Correlación Canónica

Cálculo de las sucesivas variables canónicas

- el problema de encontrar la r-ésima pareja de variables canónicas será:

$$* \text{Encontrar los vectores columnas } \alpha^{(r)} \text{ y } \beta^{(r)} \text{ tales que :}$$

$$A'_{(r)}\Sigma_y A_{(r)} = I\ ,\ \ B'_{(r)}\Sigma_x B_{(r)} = I\ ,\ A'_{(r)}\Sigma_{yx} B_{(r)} = \mathrm{P}_{(r)}$$

y que hagan máximo el cuadrado de la correlación lineal ρ^2 entre las variables $X^{(r)}, Y^{(r)}$ siendo :

$$Y^{(r)} = \alpha_1^{(r)} Y_1 + \alpha_2^{(r)} Y_2 + \ldots + \alpha_q^{(r)} Y_q \quad \text{y} \quad X^{(r)} = \beta_1^{(r)} X_1 + \beta_2^{(r)} X_2 + \ldots + \beta_P^{(r)} X_p$$

o, equivalentemente

$$\text{Maximizar } \rho^2_{(\mathrm{r})} = (\alpha^{(r)'} \Sigma_{yx} \beta^{(r)})^2 \quad \text{sujeto a las restricciones :} \begin{cases} A'_{(r)}\Sigma_y A_{(r)} = I \\ B'_{(r)}\Sigma_x B_{(r)} = I \\ A'_{(r)}\Sigma_{yx} B_{(r)} = \mathrm{P}_{(r)} \end{cases}$$

Análisis de Correlación Canónica

Cálculo de las sucesivas variables canónicas

- las nuevas parejas de variables canónicas vuelven a estar caracterizadas por los autovectores de las mismas matrices deducidas para la primera variable canónica, $\Sigma_x^{-1}\Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx}$ y $\Sigma_y^{-1}\Sigma_{yx}\Sigma_x^{-1}\Sigma_{xy}$, presentando siempre el mismo autovalor asociado para cada pareja.

- Así, si son $\rho^2_{(1)} \geq \rho^2_{(2)} \geq \ldots \geq \rho^2_{(\min(p,q))} \geq 0$ los mayores autovalores de ambas matrices, las sucesivas parejas de variables canónicas serán justamente aquéllas que toman por coeficientes los autovectores asociados a los autovalores, $\rho^2_{(\mathrm{r})}$, de sus correspondientes matrices anteriormente deducidas, ordenadamente de mayor a menor.

- $\rho^2_{(\mathrm{r})}$ será pues el coeficiente de determinación entre las r-ésimas variables canónicas; por lo que se define el *r-ésimo coeficiente de correlación canónica* como el coeficiente de correlación lineal simple de Fisher, $\rho_{(\mathrm{r})}$, entre dichas r-ésimas variables canónicas.

Análisis de Correlación Canónica

Interpretación geométrica

- En el espacio de los casos, el coseno del ángulo que forma cada pareja de variables canónicas puede calcularse como:

$$\cos(\phi) = \frac{\left(Y\alpha^{(k)}\right)'\left(X\beta^{(k)}\right)}{\left|Y\alpha^{(k)}\right| \cdot \left|X\beta^{(k)}\right|} \Leftrightarrow \cos^2(\phi) = \frac{\left(\left(Y\alpha^{(k)}\right)'\left(X\beta^{(k)}\right)\right)^2}{\left|Y\alpha^{(k)}\right|^2 \cdot \left|X\beta^{(k)}\right|^2} = \frac{\left(\left(Y\alpha^{(k)}\right)'\left(X\beta^{(k)}\right)\right)^2}{\left(Y\alpha^{(k)}\right)'\left(Y\alpha^{(k)}\right)\left(X\beta^{(k)}\right)'\left(X\beta^{(k)}\right)}$$

de donde

$$\cos^2(\phi) = \frac{\left(\alpha^{(k)'} \frac{1}{n}\left(Y'X\right)\beta^{(k)}\right)^2}{\left(\alpha^{(k)'} \frac{1}{n}\left(Y'Y\right)\alpha^{(k)}\right) \cdot \left(\beta^{(k)'} \frac{1}{n}\left(X'X\right)\beta^{(k)}\right)} = \frac{\left(\alpha^{(k)'} S_{yx} \beta^{(k)}\right)^2}{\left(\alpha^{(k)'} S_y^2 \alpha^{(k)}\right)\left(\beta^{(k)'} S_x^2 \beta^{(k)}\right)} = \hat{\rho}^2_{(k)}$$

Análisis de Correlación Canónica

Propiedades

- Las variables canónicas $(X^{(r)})$ están tipificadas y se encuentran incorrelacionadas entre sí para r≠s. Análogamente para las $(Y^{(r)})$.
- Las parejas de variables canónicas $(X^{(r)}, Y^{(s)})$, presentan una correlación máxima cuando r=s=min(r,s), y una correlación nula cuando r≠s.
- El cuadrado del coeficiente de correlación lineal, $\rho^2_{(r)}$, que presetan las parejas de variables canónicas $(X^{(r)}, Y^{(s)})$, es su autovalor (común) asociado.
- Ese autovalor representa la proporción de varianza de $X^{(r)}$ explicada por las $\{Y_1, Y_2, ..., Y_q\}$; así como la proporción de varianza de $Y^{(r)}$ explicada por las $\{X_1, X_2, ..., X_p\}$.

Análisis de Correlación Canónica

Propiedades

- los coeficientes de la r-ésima pareja de variables canónicas son los autovectores ligados al mismo r-ésimo autovalor (una vez ordenados de mayor a menor) de las matrices:

$$\Sigma_y^{-1}\Sigma_{yx}\Sigma_x^{-1}\Sigma_{xy} \quad (\text{sus autovector son los } \alpha^{(r)})$$
$$\Sigma_x^{-1}\Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx} \quad (\text{sus autovector son los } \beta^{(r)})$$

- La variable opuesta a una variable canónica, también lo es.

- Las correlaciones canónicas son invariantes ante cambios de origen y escala (transformaciones lineales) de las variables originales: Es decir, si

$$\tilde{y} = A'y + a \quad \text{y} \quad \tilde{x} = B'x + b$$

entonces, las correlaciones canónicas entre estas nuevas variables son las mismas que entre *y* y *x*; y sus vectores canónicos son:

$$\tilde{\alpha}^{(r)} = A^{-1}\alpha^{(r)} \quad \text{y} \quad \tilde{\beta}^{(r)} = B^{-1}\beta^{(r)}$$

- Las correlaciones canónicas no varían al sustituir las variables originales por un mismo número de combinaciones l.i. de las mismas.

Análisis de Correlación Canónica

Contrastación del modelo

Enfoque Muestral

- En la Práctica, podremos estimar las variables canónicas a partir de los autovalores y autovectores de las matrices muestrales:

$$S_y^{-1}S_{yx}S_x^{-1}S_{xy} \quad \text{y} \quad S_x^{-1}S_{xy}S_y^{-1}S_{yx}$$

siendo sus autovectores estimadores máximoverosímiles de los coeficientes $\alpha^{(r)}$ y $\beta^{(r)}$ recíprocamente, y sus autovalores, estimadores máximoverosímiles de los cuadrados de los correspondientes coeficientes de correlación canónica

UNIVERSIDAD DE ALCALÁ

Análisis de Correlación Canónica
Contrastación del modelo

- Adecuación del Modelo $\begin{cases} H_0 : \Sigma_{yx} = 0 & \text{(matriz nula)} \\ H_1 : \Sigma_{yx} \neq 0 \end{cases}$

- Bajo la hipótesis de Normalidad Multivariante N($\mathbf{0}$,Σ) para el conjunto de todas las variables observadas y, por tanto, de los vectores $y'=\{Y_1, Y_2, ..., Y_q\} \in N(\mathbf{0},\Sigma_y)$ y $x'=\{X_1, X_2, ..., X_p\} \in N(\mathbf{0},\Sigma_x)$, ocurrirá que:

$$\Lambda = \frac{|\Sigma|}{|\Sigma_y||\Sigma_x|} = \left| I - \Sigma_x^{-1}\Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx} \right| = \prod_{r=1}^{h}\left(1-\rho_{(r)}^2\right) \xrightarrow{Bajo\, H_0} \Lambda(p, n-1-q, q)$$

De donde se obtiene, a partir del test de razón de verosimilitudes

$$\lambda = -2\ln(\lambda^*) = -2\ln(\Lambda^{n/2}) = -n\ln(\Lambda) = -n\ln\left(\prod_{r=1}^{h}\left(1-\rho_{(r)}^2\right)\right) = -n\left(\sum_{r=1}^{h}\ln\left(1-\rho_{(r)}^2\right)\right) \xrightarrow[n\to\infty]{Bajo\, H_0} \chi_{pq}^2$$

- y por la aproximación de Bartlett $-\left(s - \frac{n-t+1}{2}\right)\ln\left(\Lambda(n,s,t)\right) \xrightarrow{s\to\infty} \chi_{nt}^2$

que $$-\left(n - \frac{3+p+q}{2}\right)\ln(\Lambda) = -\left(n - \frac{3+p+q}{2}\right)\left(\sum_{r=1}^{h}\ln\left(1-\rho_{(r)}^2\right)\right) \xrightarrow[n\to\infty]{Bajo\, H_0} \chi_{pq}^2$$

UNIVERSIDAD DE ALCALÁ

Análisis de Correlación Canónica
Contrastación del Modelo

- Análisis de la Dimensionalidad

$$\begin{cases} H_0 : \rho_{(k+1)}^2 = 0 & \left(\Rightarrow \rho_{(k+2)}^2 = ...\rho_{(h)}^2 = 0\right) \\ H_1 : \rho_{(k+1)}^2 > 0 \end{cases} \quad , k = 0,1,..., h-1$$

- Bajo la hipótesis de Normalidad Multivariante N($\mathbf{0}$,Σ) para el conjunto de todas las variables observadas y, por tanto, de los vectores $y'=\{Y_1, Y_2, ..., Y_q\} \in N(\mathbf{0},\Sigma_y)$ y $x'=\{X_1, X_2, ..., X_p\} \in N(\mathbf{0},\Sigma_x)$, el estadístico experimental del correspondiente contraste de razón de verosimilitudes será:

$$\lambda = -\left(n - \frac{3+p+q}{2}\right)\left(\sum_{r=k+1}^{h}\ln\left(1-\rho_{(r)}^2\right)\right) \xrightarrow[n\to\infty]{Bajo\, H_0} \chi_{(p-k)(q-k)}^2$$

- Aternativamente, el estadístico experimental de Bartlett-Lawley será:

$$L_k = -\left(n - k - \frac{3+p+q}{2} + \sum_{r=1}^{k}\rho_{(r)}^{-2}\right)\left(\sum_{r=k+1}^{h}\ln\left(1-\rho_{(r)}^2\right)\right) \xrightarrow[n\to\infty]{Bajo\, H_0} \chi_{(p-k)(q-k)}^2$$

Análisis de Correlación Canónica

Relación con otras técnicas de análisis multivariante

- Regresión simple
  - es el caso *p=q=1*
  - las submatrices Σ_y, Σ_{yx}, Σ_x son respectivamente los escalares varianza de *Y*, Covarianza de *Y* con *X* y varianza de *X*; por lo que el coeficiente de correlación canónica es:

$$\rho^2 = \Sigma_y^{-1}\Sigma_{yx}\Sigma_x^{-1}\Sigma_{xy} = \frac{\sigma_{xy}^2}{\sigma_x^2 \cdot \sigma_y^2}$$

- Regresión múltiple
  - es el caso *p=p, q=1*
  - las submatrices Σ_y, Σ_{yx}, Σ_x son respectivamente el escalar varianza de *Y*, el vector de dimensión *p* de las covarianzas de la variable *Y* con las *X*'s y la matriz de dimensión *p·p* de varianzas y covarianzas de las *X*'s; por lo que el coeficiente de correlación canónica es:

$$\rho^2 = \Sigma_y^{-1}\Sigma_{yx}\Sigma_x^{-1}\Sigma_{xy} = \frac{\Sigma_{yx}\Sigma_x^{-1}\Sigma_{xy}}{\sigma_y^2} = \frac{\sigma_{\hat{y}}^2}{\sigma_y^2}$$

Análisis de Correlación Canónica

Relación con otras técnicas de análisis multivariante

- Análisis Discriminante
  - Es el caso cuando definimos las *q* variables *Y*'s como indicadoras de pertenencia a cada uno de los *q+1* grupos de la variable categórica grupo del Análisis Discriminante, de la forma usual (*método indicador*)

$$Y_i = \begin{cases} 1 \text{ si el caso pertenece al grupo i-ésimo,} & i = 1,\dots,q \\ 0 \text{ si el caso pertenece a otro grupo distinto} \end{cases}$$

  - En este caso, el cuadrado del coeficiente de correlación canónica es:

$$\rho_r^2 = \frac{\lambda_r}{1+\lambda_r} = \frac{\text{varianza inter-grupos del eje discriminante r-ésimo}}{\text{varianza total del eje discriminante r-ésimo}} = \frac{u_r' B u_r}{u_r' S u_r} = \eta_r^2$$

  siendo λ_r el autovalor asociado a la función discriminante r-ésima

Análisis de Correlación Canónica

Relación con otras técnicas de análisis multivariante

- Tablas de Contingencia
  - Es el caso en que definimos las q variables Y's como indicadoras de la observación de cada una de las q modalidades de uno de los atributos de la tabla (por ejemplo, atributo-columna) y las p variables X's como indicadoras de ocurrencia de cada una de las p modalidades del otro atributo de la tabla (por ejemplo, atributo-fila).

$$Y_i = \begin{cases} 1 & \text{si el caso presenta el atributo columna i - ésimo} \\ 0 & \text{si el caso pertenece a otro grupo distinto} \end{cases}$$

$$X_i = \begin{cases} 1 & \text{si el caso presenta el atributo fila i - ésimo} \\ 0 & \text{si el caso pertenece a otro grupo distinto} \end{cases}$$

  - la asociación entre los atributos (variables cualitativas) puede estudiarse a partir de $h\text{-}1=\min(p,q)\text{-}1$ relaciones canónicas
  - Si la primera pareja de variables canónicas presente correlación canónica nula, ello indicaría la no asociación o independencia de los dos atributos enfrentados en la tabla.

Análisis de Correlación Canónica

Coeficientes de Redundancia

- Se define *Coeficiente de Redundancia* entre la variable $X^{(k)} = \beta^{(k)'}x = \beta_1^{(k)}X_1 + \beta_2^{(k)}X_2 + ... + \beta_p^{(k)}X_p$ y el conjunto de variables endógenas $Y_1, Y_2, ..., Y_q$, como el promedio de los cuadrados de las correlaciones entre la variable $X^{(k)}$ y cada una de las $Y_1, Y_2, ..., Y_q$.

- Supuestas las variables tipificadas, como

$$Corr(y, X^{(k)}) = Cov(y, X^{(k)}) = E\left[y\left(\beta^{(k)'}x\right)\right] = E[yx']\beta^{(k)} = \Sigma_{yx}\beta^{(k)} = \mathrm{P}_{yx}\beta^{(k)}$$

entonces:

$$CR(y \mid X^{(k)}) = \frac{1}{q}\left(\Sigma_{yx}\beta^{(k)}\right)'\left(\Sigma_{yx}\beta^{(k)}\right) = \frac{1}{q}\beta^{(k)'}\Sigma_{xy}\Sigma_{yx}\beta^{(k)} = \frac{1}{q}\beta^{(k)'}\mathrm{P}_{xy}\mathrm{P}_{yx}\beta^{(k)}$$

Análisis de Correlación Canónica

Coeficientes de Redundancia

- Se define *Coeficiente de Redundancia Total* como la suma de las redundancias anteriores para el conjunto de las *h* combinaciones lineales
- Supuestas las variables tipificadas,

$$CR(y\,|\,x)=\sum_{k=1}^{h}CR(Y\,|\,X^{(k)})=\frac{1}{q}\sum_{k=1}^{h}\beta^{(k)'}\Sigma_{xy}\Sigma_{yx}\beta^{(k)}=\frac{1}{q}\sum_{k=1}^{h}\beta^{(k)'}\mathrm{P}_{xy}\mathrm{P}_{yx}\beta^{(k)}$$

- Cuando se aplican sobre las variables canónicas obtenidas por el procedimiento anteriormente expuesto sobre las variables tipificadas, puede demostrarse que:

$$CR(y\,|\,x)=\frac{Traza(\mathrm{P}_{yx}\mathrm{P}_{xx}^{-1}\mathrm{P}_{xy})}{Traza(\mathrm{P}_{yy})}=\frac{1}{q}\sum_{k=1}^{q}R^2_{Y_k;X_1,X_2,\ldots,X_p}$$

Análisis de Correlación Canónica

Análisis Canónico Asimétrico

- Uno de los conjuntos de variables (por ejemplo las *X*'s) se consideran variables exógenas o explicativas de las variables del otro grupo (en este caso, las *Y*'s) que se consideran variables endógenas o explicadas.
- El Análisis Canónico Asimétrico, desarrollado por Stewart y Love (1968) y Gudmundsson (1977), trata de encontrar combinaciones lineales tipificadas de las variables exógenas

$$X^{(k)}=\beta^{(k)'}x=\beta_1^{(k)}X_1+\beta_2^{(k)}X_2+\ldots+\beta_P^{(k)}X_p$$

de forma que presenten las más altas correlaciones con las variables del conjunto de variables endógenas $Y_1, Y_2, \ldots, Y_q$ que se pretenden explicar, y utilizando como medida de correlación global el coeficiente de Redundancia Total; que en el caso de variables tipificadas es:

$$CR(y\,|\,x)=\sum_{k=1}^{h}CR(Y\,|\,X^{(k)})=\frac{1}{q}\sum_{k=1}^{h}\beta^{(k)'}\Sigma_{xy}\Sigma_{yx}\beta^{(k)}=\frac{1}{q}\sum_{k=1}^{h}\beta^{(k)'}\mathrm{P}_{xy}\mathrm{P}_{yx}\beta^{(k)}$$

Análisis de Correlación Canónica

Análisis Canónico Asimétrico

- Si todas las variables *X*'s e *Y*'s se encuentran tipificadas, el problema se plantea como:

$$*\ \text{Encontrar } \beta^{(1)} \text{ tal que: } \quad Var(X^{(1)}) = 1 \quad y \quad CR(y \mid X^{(1)}) \text{ sea máximo}$$

  o lo que es equivalente,

$$\text{Maximizar } \beta^{(1)'}\Sigma_{xy}\Sigma_{yx}\beta^{(1)} \quad \text{sujeto a la restricción: } \beta^{(1)'}\Sigma_{xx}\beta^{(1)} = 1$$

  lo que conduce a la ecuación:

$$\Sigma_{xy}\Sigma_{yx}\beta^{(1)} = \lambda\Sigma_{xx}\beta^{(1)} \quad \text{o equivalentemente} \quad \Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yx}\beta^{(1)} = \lambda\beta^{(1)}$$

- La solución sería el autovector de la matriz $H = \Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yx}$ asociado a su mayor autovalor λ.

- Imponiendo que las sucesivas variables canónicas (en este enfoque asimétrico) sean ortogonales a las anteriores, las obtendríamos como los autovectores de los sucesivos autovalores de la anterior matriz *H*.