

# Componentes principales comunes: una aproximación robusta al problema

Graciela Boente

CONICET y Departamento de Matemática e Instituto de Cálculo,  
Facultad de Ciencias Exactas y Naturales, UBA

Muchos autores, tales como Flury (1988), estudiaron modelos para estructuras de dispersión comunes. Dichos modelos se introdujeron para resolver el problema de un número excesivo de parámetros, cuando consideramos varias poblaciones en análisis multivariado. Por ejemplo, una situación en la que se involucran varias poblaciones es el caso de análisis discriminante donde usualmente, para realizar este análisis se supone igualdad de las matrices de covarianza. Sin embargo, a veces este supuesto no es válido y problemas relacionados con un número excesivo de parámetros ocurren si las matrices de escala,  $\Sigma_i$ , se estiman separadamente para cada población. A diferencia de lo que ocurre en el caso univariado y tal como menciona Flury (1988), las matrices de dispersión pueden diferir de muy distintas formas.

Una estructura común básica supone que las  $k$  matrices de covarianza tienen distintos autovalores pero idénticos autovectores, es decir, son conmutables, o sea,  $\Sigma_i = \beta \Lambda_i \beta^T$ ,  $1 \leq i \leq k$ , donde  $\Lambda_i$  son matrices diagonales,  $\beta$  es la matriz ortogonal de autovectores y  $\Sigma_i$  es la matriz de covarianza de la población  $i$ -ésima. Este modelo fue introducido en Flury (1984) y se conoce como el modelo de *componentes principales comunes* (CPC). En Flury (1988) se obtienen los estimadores de máxima verosimilitud y se estudian sus propiedades asintóticas así como test de hipótesis para estudiar la validez de cada una de las jerarquías de similaridad siguientes: (i) matrices arbitrarias, (ii) modelo de componentes principales comunes, (iii) matrices de escala proporcionales, o sea,  $\Sigma_i = \rho_i \Sigma_1$ ,  $1 \leq i \leq k$ ,  $\rho_1 = 1$  y (iv) matrices idénticas.

Es bien sabido que los estimadores de máxima verosimilitud bajo estos modelos pueden verse perturbados por la presencia de datos atípicos en la muestra. Por otra parte, también los tests de cociente de máxima verosimilitud ven afectados su nivel y potencia por datos anómalos. Es por lo tanto necesario, encontrar alternativas robustas a estos procedimientos.

En esta charla, se describirán diversos enfoques al problema de estimación robusta bajo el modelo CPC y proporcional y se presentarán tests robustos para las jerarquías de similaridad consideradas por Flury (1988).

## Referencias

- B. Flury (1984). Common principal components in  $k$  groups. *Journal of the American Statistical Association*, 79, 892–898.
- B. Flury (1988). *Common Principal Components and Related Multivariate Models*. New York: John Wiley.