

# INTRODUCCIÓN

Uno de los aspectos más importantes de la Estadística Multivariante es el Análisis de Componentes Principales (ACP), cuyo objetivo principal es describir parsimoniosamente la variación total de la muestra en pocas dimensiones; es decir, se pretende reducir la dimensión del conjunto de datos original, minimizando la pérdida de información. Esas "pocas dimensiones" llamadas Componentes Principales son combinaciones lineales no correlacionadas de las variables originales, que en ocasiones son utilizadas como paso inicial en muchos análisis estadísticos, funcionando la metodología como un análisis exploratorio de datos multivariantes.

En muchos casos se asume la existencia de la(s) población(es), que está(n) caracterizada(s) por la distribución de probabilidad de las variables, y una observación (objeto) es un valor muestral o medida de tales variables en un individuo de una de las poblaciones.

La idea inicial del Análisis de Componentes Principales (ACP) fue sugerida por Pearson (1903) y más tarde desarrollada por Hotelling(1933), pero como muchos métodos estadísticos sólo comienza a ser utilizada con mayor frecuencia con el avance de las máquinas computadoras. Método multivariante, razonablemente sencillo de aplicar y que a la fecha se encuentra listo para su uso en muchos paquetes estadísticos tales como STATH GRAF, cuando se tiene la seguridad de que se está frente a un sólo grupo o población.

En general los métodos multivariantes están relacionados a datos donde para cada objeto en estudio se ha medido más de una variable. Una distinción muy importante en el análisis es saber si los  $n$  objetos representan muestras de una única población o de varias poblaciones.

Cabe hacer hincapié que el ACP es un método aplicado a una sola muestra y para lo cual se disponen de paquetes estadísticos como el que fue señalado líneas arriba, y, solamente, se conocen dos esfuerzos para la generalización de la metodología a más de una población. Krzanowsky (1979) compara los componentes principales de dos muestras diferentes calculando los ángulos entre los subespacios generados por las "q" primeras componentes para cada grupo, donde ( $q < p$ ) que es el número de variables del problema original y Flury (1983) usa los vectores característicos de la matriz  $\sum_1^{-1} \sum_2$  para obtener variables no correlacionadas en los dos grupos simultáneamente, donde  $\sum_g$  es la matriz de covarianzas poblacional del grupo  $g = 1, 2$ .

El objeto del presente trabajo, en el contexto del Análisis de Componentes Principales, es describir la generalización del Análisis de Componentes Principales (ACP) de un grupo a dos grupos —Componentes Principales Generalizadas (CPG), en base a la comparación de las matrices de covarianzas  $\sum_1$  y  $\sum_2$  de las dos poblaciones según la presentación de Flury (1983).

Flury (1983), basándose en el Principio de Unión e Intersección de Roy (1957), expone algunas propiedades de interés de los autovectores de la matriz  $\sum_1^{-1} \sum_2$  y las combinaciones lineales definidas con tales autovectores son presentados como una generalización del ACP a dos grupos; de mucha utilidad, toda vez que la hipótesis de igualdad de las matrices de covarianzas es rechazada. La metodología es contrastada con los mismos datos usados por Flury.

La conclusión del presente trabajo es que la metodología de Componentes Principales Generalizada a dos grupos es de mucha utilidad para pronunciarnos respecto a la heteroscedasticidad de las matrices de covarianzas de las dos poblaciones concurrentes,

además, usando el ACP generalizando a dos grupos se obtienen todas las combinaciones lineales no-correlacionadas simultáneamente en ambos grupos, y que algunas de las combinaciones lineales pueden ser muy útiles para futuros análisis estadísticos, tal como para hacer Análisis Discriminante.

Se sugiere que el método presentado sea introducido en el syllabus del curso de Análisis Multivariante como información y estímulo para los estudiantes interesados en las aplicaciones prácticas.

El trabajo se divide en tres capítulos. En el capítulo 1 se presenta una revisión bibliográfica bastante detallada de la técnica de ACP a un sólo grupo conforme se viene usando a la fecha. En el capítulo 2 se detalla la metodología de Análisis de Componentes Principales Generalizado a dos grupos (ACPG) y en el último capítulo presentamos la aplicación para dos conjuntos de datos relacionados a Billetes de Banco verdaderos y falsos utilizados en otras aplicaciones por Flury y Riedwiel (1983).