

COMENTARIS DE LLIBRES

PRINCIPAL COMPONENT ANALYSIS

Autor: I.T. Jolliffe

Springer - Verlaq, N. York

26 ilustraciones, 271 p  g.

Comentario hecho por: CARLOS M. CUADRAS

El an  lisis de Componentes Principales cronol  gicamente es la primera t  cnica de An  lisis Multivariante y sigue siendo, 86 a  os despu  s de ser introducida por K.Pearson, - una importante herramienta del an  lisis estad  stico de datos. Adem  s, otras t  cnicas aparecidas posteriormente (an  lisis de coordenadas principales, an  lisis de correspondencias, an  lisis factorial, an  lisis can  nico de poblaciones, etc.), se fundamentan en el An  lisis de Componentes Principales.

La obra de I.T. Jolliffe es un tratado completo y bien presentado del tema. Se inicia con la definici  n y obtenci  n de las componentes principales (cap. 1), con una breve pero interesante referencia hist  rica. Los dos siguientes cap  tulos se destinan a presentar las propiedades matem  ticas y estad  sticas del An  lisis, con especial atenci  n a las propiedades optimales, algebraicas y geom  tricas de las componentes, distinguiendo el caso poblacional del muestral. Ciertos aspectos inferenciales, partiendo de una distribuci  n normal multivariante, son presentados (distribuci  n de la matriz de covarianzas, de sus valores propios, estimaci  n puntual, intervalos de confianza), con especial   nfasis en las propiedades asint  ticas. Los dos siguientes cap  tulos (4 y 5), se destinan a ilustrar con ejemplos las dos aplicaciones m  s importantes del An  lisis: interpretar con un peque  o n  mero de componentes un n  mero mucho mayor de variables, y la representaci  n gr  fica de datos a lo largo de los dos primeros ejes principales. A tal fin, el autor -- utiliza ejemplos bien conocidos de la literatura, abarcando un amplio campo de aplicaciones (antropolog  a, sociolog  a, meteorolog  a, qu  mica, marketing, etc.). El cap. 5 se destina tambi  n a describir otras t  cnicas

de representaci  n gr  fica muy relacionadas: biplot, an  lisis de coordenadas principales y an  lisis de correspondencias.

Un tema de gran inter  s para los investigadores es la decisi  n sobre el n  mero de componentes principales. Este tema queda debidamente tratado en el cap. 6, donde el autor, adem  s de exponer los criterios cl  sicos -- (porcentaje de varianza explicada, valores propios mayores que 1), presentando sus ventajas e inconvenientes, explica tambi  n las recientes t  cnicas cross-validatorias, basadas en la descomposici  n en valores singulares de la matriz de datos, utilizadas como medidas de predicci  n tras eliminar convenientemente algunos datos. En el mismo cap  tulo expone tambi  n el problema de elegir un subconjunto adecuado de m variables de -- otros m  s amplio de p variables. En todos los casos se dan ejemplos suficientemente explicativos.

Los cap  tulos 7, 8 y 9 se destinan a relacionar el An  lisis de Componentes Principales con otros m  todos estad  sticos, tales como el An  lisis Factorial, el modelo de regresi  n lineal, el An  lisis Discriminante, el An  lisis de Correlaci  n Can  nica, etc. quedando destacada la influencia de la t  cnica objeto de este libro. Tambi  n, pero ya en el cap. 10, se estudia como puede aplicarse a la detecci  n de outliers, por el sistema de representaci  n gr  fica sobre las primeras -- componentes. En este mismo cap  tulo se analiza la estimaci  n por m  todos robustos. Cabe destacar que los diferentes puntos de vista son comentados por el autor con acertado criterio.

En el cap. 11 se aborda el problema de apli-

car Análisis de Componentes Principales a datos en los que no se cumplan las condiciones iniciales del análisis (independencia de las observaciones y variables continuas). Así se explica como puede aplicarse en el caso de variables discretas, series temporales, datos composicionales (es decir, cada variable es una proporción y la suma vale 1), presencia de datos faltantes, etc. Incluso se presenta una aplicación al conocido test de Cramer-von Mises de bondad de ajuste.

El último capítulo es bastante heterogéneo. Se dedica a presentar diferentes generalizaciones (introducir pesos, análisis no lineal, datos centrados, variables instrumentales, etc.). La obra concluye con un apéndice sobre aspectos computacionales y numéricos del Análisis (básicamente ligado a la diagonalización y descomposición en valores singulares de una matriz), y una bien documentada relación de paquetes de programas que lo contienen, con comentarios sobre su eficiencia, incluyendo la incidencia de los mismos en los microcomputadores.

En definitiva, el libro de I.T. Jolliffe es una obra completa y bien documentada, con un nivel matemático muy asequible, en la que se dice casi todo sobre el tema y que es totalmente recomendable a cualquier estudioso del análisis multivariante.