

Ejemplo Análisis de Componentes principales

“decathlon”

El dataset “decathlon”, perteneciente al paquete “FactoMineR”, está formado por 41 filas y 13 columnas. Las columnas 1 a 12 son variables continuas: las primeras diez columnas corresponden al desempeño de los atletas para las 10 pruebas del decatón y las columnas 11 y 12 corresponden respectivamente al rango y los puntos obtenidos. La última columna es una variable categórica correspondiente a la competición atlética (2004 Olympic Game o 2004 Decastar). Las variables recogidas en la base de datos son las siguientes:

- 100m
- Long.jump
- Shot.put
- High.jump
- 400m
- 110m.hurdle
- Discus
- Pole.vault
- Javeline
- 1500m
- Rank
- Points
- Competition

Se desea realizar un Análisis de Componentes Principales con la intención de describir un conjunto de datos, deseamos conocer si existen relaciones lineales entre las variables y saber si hay similitudes entre los atletas. Es decir, se desea, resumir y reducir la dimensionalidad de la base de datos e interpretar los resultados. Ver si se puede medir el rendimiento de un atleta con un número reducido de variables, para ello, vamos a realizar los siguientes apartados:

1. Realiza un resumen descriptivo de los datos y análisis gráfico de la correlación de estos.
2. Realiza el Test de esfericidad de Bartlett e indica si es conveniente realizar un Análisis de Componentes Principales. Realiza ahora el test KMO. ¿Coinciden los resultados de los dos tests?
3. Obtén las Componentes Principales.
4. Decide cuál sería el número de Componentes Principales que resume mejor la información.
5. Realiza una representación de las variables respecto a las componentes principales y realiza una explicación de las mismas.
6. Indica qué variables son las que más han contribuido a la construcción de las componentes.
7. Realiza un estudio individual y representa las observaciones en las componentes. Ayúdate de una representación de los atletas y de las variables en un mismo gráfico. Interpreta los resultados.