Análise Multivariada II

Lista de Exercícios 3 – PCA - Heptatlo Olímpico de Seul (1988)

O pentatlo para mulheres foi realizado pela primeira vez na Alemanha, em 1928. Inicialmente a modalidade consistia do arremesso de peso, salto em distância, corrida de 100m, salto em altura e eventos de lançamento de dardo realizados durante dois dias. O pentatlo foi introduzido pela primeira vez em Jogos Olímpicos em 1964, consistindo dos 80 m com barreiras, tiros, salto em altura, salto em comprimento e corrida de 200 m. Em 1977, a corrida de 200 m foi substituída pela de 800 m e em 1981 a IAAF trouxe o heptatlo no lugar do pentatlo, com um dia que contém os eventos- corrida de 100 m com barreiras, tiro, salto em altura, corrida de 200 m e no segundo dia , o salto em comprimento, lançamento de dardo e corrida de 800 m. Um sistema de pontuação é utilizado para atribuir pontos aos resultados de cada evento, e o vencedor é a mulher que acumula mais pontos durante os dois dias. O evento fez sua primeira aparição olímpica em 1984.

Nos Jogos Olímpicos de 1988, em Seul, o heptatlo foi vencido por uma das estrelas do atletismo feminino, nos EUA, Jackie Joyner-Kersee. O pacote "HSAUR" contém os dados de 25 competidoras do heptatlo com 8 variáveis.

Variáveis:

hurdles: resultados de 100 m com barreiras

highjump: resultados de salto em altura

shot: resultados de arremesso de peso

run200m: resultados de 200 m rasos

longjump: resultados de salto em distância

javelin: resultados de lançamento de dardos

run800m: resultados de 800 m rasos

score: pontuação total

Realize a Análise de Componentes Principais visando a exploração da estrutura dos dados e avaliar como os escores das componentes principais se relacionam com os escores do sistema. Faça comentários sobre a matriz de correlação. Você observa alguma estrutura? Grupos se formando? Variáveis isoladas? Realize diagramas de dispersão para uma melhor visualização.

Obs.

* Algumas variáveis são medidas em segundos (quanto menor o número, melhor a performance) e outras são medidas em metros (quanto maior o número, melhor a performance). Então, para que os escores apontem na mesma direção é necessário transformar os tempos para que representem a diferença em relação ao tempo da corredora mais lenta (que é o maior valor).

Ex: hurdles2 = max(heptathlon$hurdles) - heptathlon$hurdles

Desse jeito os maiores números são melhores em todas as variáveis.

* Não esqueça de padronizar os dados.