


## **Análise em Componentes Principais, Robustez e Detecção de Outliers: Uma Introdução em R**

**M. Rosário Oliveira<sup>1</sup>**

<sup>1</sup> CEMAT e Dep. Matemática, Instituto Superior Técnico, Universidade de Lisboa, [rosario.oliveira@tecnico.ulisboa.pt](mailto:rosario.oliveira@tecnico.ulisboa.pt)

A Análise de Componentes Principais (ACP) é uma das metodologias estatísticas mais usadas na resolução de problemas reais. No entanto, a existência de observações atípicas pode conduzir a conclusões enganadoras. A identificação destas observações é, em geral, uma tarefa difícil e a sua remoção pode traduzir-se em perdas inaceitáveis de informação. A solução natural reside no uso de procedimentos robustos. Estes têm a capacidade de detetar e acomodar as observações atípicas, propondo estimadores que, não sendo ideais no caso de se verificarem as hipóteses teóricas associadas ao modelo em estudo, dão melhores resultados quando ocorrem pequenos desvios destas mesmas hipóteses.

No estudo de problemas reais, a existência de outliers nos dados é muitas vezes encarada como um obstáculo, que se converte em efeitos perturbadores na análise. Mas por vezes o interesse principal do estudo reside na deteção dessas observações atípicas: a descoberta dos buracos na camada de ozono derivou da identificação de valores irregulares e inesperados nas medições efetuadas; um ataque numa rede de computadores pode ser evitado se o tráfego anómalo associado for identificado e bloqueado a tempo. A ACP robusta é um método de deteção de outliers que utiliza as potencialidades da ACP como método de redução da dimensionalidade em benefício da identificação de observações atípicas.

Neste curso introduzimos ACP e discutimos as limitações e vantagens dos estimadores clássicos e robustos. A deteção de outliers baseada na distância de Mahalanobis e em ACP é discutida. A formulação teórica e as propriedades dos métodos são exploradas, mas é dada igual importância à análise de dados reais usando o software . Traga o seu portátil para explorarmos, com um conjunto de dados, como estas metodologias funcionam.

**Palavras-chave:** Análise de Componentes Principais, Detecção de Outliers, Robustez, .