

Misturas de regressões lineares: um novo teste de alteração da estrutura

Susana Faria

Officina Mathematica
Departamento de Matemática para a Ciência e Tecnologia
Universidade do Minho, 4800-058 Guimarães, Portugal

Gilda Soromenho

LEAD
Faculdade de Psicologia e Ciências da Educação
Universidade de Lisboa, Lisboa, Portugal

Resumo

No contexto do modelo de regressão linear, dispomos de uma relação estimada a partir de um certo número de observações, que permite explicar o comportamento de uma variável resposta em função de determinadas variáveis explicativas.

Por vezes, são obtidas observações adicionais dessas variáveis e procura-se testar se as novas observações são compatíveis com o modelo de regressão estimado ou se podem ser consideradas *outliers* a esse modelo. Os testes que permitem estudar este problema, designam-se por *testes de alteração da estrutura* (ver, por exemplo, Johnston (1991, Cap.5) e Murteira e al (2001, pp.576-578)).

Neste trabalho, descrevemos um novo teste deste tipo que desenvolvemos para o caso de modelos de mistura de regressões lineares. Este teste baseia-se na comparação entre o modelo de mistura estimado a partir do conjunto de observações iniciais e o modelo de mistura estimado a partir da totalidade das observações disponíveis (observações iniciais e observações novas).

Com o objectivo de ilustrar a aplicação deste teste em situações práticas onde as misturas de regressões lineares são adequadas, começamos por gerar amostras de pequena dimensão ($n=50$), média dimensão ($n=100$) e grande dimensão ($n=500$), variando a proporção de mistura e considerando três diferentes configurações para as verdadeiras rectas de regressão componentes da mistura: as rectas de regressão eram paralelas, perpendiculares ou concorrentes entre si. De seguida, estimámos um modelo de mistura de regressões lineares simples aplicando o algoritmo *Classification Expectation Maximization* (CEM) (Celeux and Govaert (1992)) e no final inserimos L novas observações que pretendíamos testar se eram compatíveis com o modelo de mistura estimado ou se constituíam *outliers* a esse modelo.

Três diferentes situações foram consideradas relativamente às L novas observações introduzidas:

i) as L novas observações constituem *outliers* ao modelo estimado;

ii) as L novas observações pertenciam ao modelo estimado;

iii) algumas das L novas observações pertenciam ao modelo estimado e outras constituem *outliers* a esse modelo.

Os resultados obtidos nos vários exemplos usados para ilustrar a aplicação do novo teste que desenvolvemos, permitem afirmar que é um teste adequado para identificar, em misturas de regressões lineares, se novas observações entretanto obtidas constituem *outliers* ao modelo de mistura estimado a partir do conjunto de observações iniciais. Podemos ainda afirmar que os resultados parecem não depender da proporção de mistura e da configuração das verdadeiras rectas de regressão.

BIBLIOGRAFIA:

Celeux, G., Govaert, G. (1992), *A classification EM Algorithm and two stochastic versions*, Computational Statistics & Data Analysis, 14, 315-332.

Johnston, J. (1991), *Econometric Methods*, McGraw-Hill International Editions

Murteira, B., Silva Ribeiro, C., Andrade e Silva, J., Pimenta, C., (2001) *Introdução à Estatística*, McGraw-Hill.