

## AValiação DA REGRESSÃO PLS (*PARTIAL LEAST SQUARE*) CONSIDERANDO DIFERENTES ESTRUTURAS DE CORRELAÇÃO ENTRE AS COVARIÁVEIS

Fernanda Gomes da Silveira<sup>1</sup>, Lucas Monteiro Chaves<sup>2</sup>,  
Fabyano Fonseca e Silva<sup>3</sup>, Luzia Aparecida da Costa<sup>4</sup>

### RESUMO

O método de regressão por Quadrados Mínimos Parciais - PLS (Partial Least Square) é considerado especialmente útil para a construção de equações de predições em situações nas quais se tem um grande número de variáveis explicativas e um número relativamente pequeno de dados amostrais. A idéia geral do PLS é formar componentes que capturem a maior quantidade de informação, porém a determinação do número ótimo de tais componentes e a avaliação do método sob diferentes estruturas de dependência entre as covariáveis ainda se caracterizam como relevantes problemas para a aplicação do referido método. No presente trabalho objetivou-se realizar um estudo de simulação assumindo diferentes estruturas de correlação entre as covariáveis, de forma que matrizes de covariâncias bloco-diagonais foram construídas representando grupos de variáveis correlacionadas entre si, mas com independência entre estes grupos. Simularam-se dados de uma distribuição normal multivariada com média zero e matriz de covariância originada da mencionada estrutura de correlação considerando diferentes cenários com 100 repetições, 50 observações e 90 variáveis explicativas (30 em cada grupo). Estes cenários diferiram em relação ao grau de dependência entre as covariáveis dentro de cada um dos três blocos da matriz de correlação assumida na simulação. Portanto, o número ótimo de componentes a ser obtidos do PLS é conhecido de antemão como sendo três, permitindo verificar se diferenças nos cenários correspondem a uma maior ou menor precisão na determinação deste número ótimo. Concluiu-se que o PLS foi eficiente para reduzir a dimensão dos dados independentemente da estrutura de correlação entre as covariáveis, visto que a correlação entre os valores observados e estimados ficou próxima de 0,97 para a variável independente, próxima de 0,62 para os coeficientes de regressão e próxima de 4 para o número ótimo.

**Palavras-chave:** *Mínimos Quadrados Parciais, Correlação, Simulação.*

---

<sup>1</sup>Departamento de Ciências Exatas - Instituto Federal de Minas Gerais, Campus Bambuí, fernanda.gomes@ifmg.edu.br

<sup>2</sup>Departamento de Ciências Exatas - Universidade Federal de Lavras

<sup>3</sup>Departamento de Estatística - Universidade Federal de Viçosa

<sup>4</sup>Departamento de Ciências Exatas - Instituto Federal de Minas Gerais, Campus Formiga