

**MODELOS DE REGRESSÃO LINEAR PARA DADOS SIMBÓLICOS TIPO INTERVALO:
UMA APLICAÇÃO**Josemir Ramos de Almeida⁽¹⁾, Eufrásio de Andrade Lima Neto⁽³⁾.

Centro de Ciências Exatas e da Natureza/Departamento de Estatística/MONITORIA.

RESUMO

A Análise de Dados Simbólicos [4] foi introduzida como um novo domínio relacionado à análise multivariada, reconhecimento de padrões e inteligência artificial, estendendo as técnicas de análise exploratória de dados e métodos estatísticos para dados de natureza simbólica. Tais dados permitem múltiplos valores (às vezes ponderados) em suas células e novos tipos de variáveis foram introduzidos (intervalar, categórica multivalorada e modal). Tais variáveis permitem levar em consideração a variabilidade e/ou incerteza presente nos dados. Descrição: Neste trabalho apresentamos uma comparação entre o método proposto por [1] e os novos métodos de regressão para variáveis simbólicas propostos por [2] e [3], a uma base de dados real com as seguintes variáveis simbólicas de natureza intervalar: pressão arterial, % de gordura, IMC, entre outras, observados entre os habitantes de Campina Grande. Metodologia: Utilizaremos um procedimento Holdout para dividir a base de dados em dois conjuntos: treinamento/teste, avaliando a performance de predição dos métodos através de um teste de hipóteses t para amostras emparelhadas, ao nível $\alpha=5\%$, em relação as seguintes métricas: erro médio quadrático e quadrado do coeficiente de correlação. Resultados: Foram testadas 3 diferentes relações entre as variáveis analisadas. Em todos os casos, observamos que as novas abordagens apresentaram um erro de predição inferior ao modelo proposto por [1]. Conclusão: Os métodos propostos por [2] e [3] configuram-se como um avanço no ajuste de modelos de regressão para dados simbólicos.

Palavras-chave: Modelos de Regressão, Dados Simbólicos, Variáveis Intervalares.

¹⁾ Bolsista, ⁽²⁾ Voluntário/colaborador, ⁽³⁾ Orientador/Coordenador ⁽⁴⁾ Prof. colaborador, ⁽⁵⁾ Técnico colaborador.