



MODELO DE REGRESSÃO LINEAR E INTERPOLAÇÃO PARA PREVISÃO DE VENDA DIÁRIA

Autores Bronstrup FC*, Simoes WL.

Descritores: Ciência de Dados; Estatística; Regressão Linear; Interpolação; Previsão de Demanda

O objetivo da pesquisa é estimar a demanda de vendas prevista com base na temperatura registrada no período através do levantamento de dados históricos, análise de sazonalidade e aplicação de inferências estatísticas onde diversos são os fatores que interferem nas demandas de vendas nas organizações e podem auxiliar na previsão futura das demandas. A assertividade sobre as previsões de demanda impacta sobre toda a cadeia produtiva na redução das necessidades de estoque e produção para atendimento apenas da demanda específica do cliente, desta forma, possibilitam redução de custos em escala sobre toda a cadeia produtiva. Foram extraídos dados de um ano de venda diária em uma indústria de sorvetes devido a sensibilidade da demanda deste segmento pelas variações de temperatura. Utilizando a Linguagem de Programação Estatística R, foi realizado sobre estes dados o processo de limpeza dos dados identificando o período de sazonalidade onde foram removidos os seis meses de menor demanda entre abril e setembro e dias úteis, visto que, os faturamentos aos domingos ou feriados são esporádicos e não representativos ao modelo. Para identificação dos *outliers* nos dados foi utilizada a função *boxplot* que apresentou dados acima da amplitude, estes foram removidos para que os mesmos não interferissem sobre os algoritmos de previsão e garantissem a normalidade dos dados. Para análise dos dados foi realizado o teste de normalidade de Shapiro Wilk onde o resultado apresentado rejeitou a hipótese nula com um *P-Value* de 0.10, portanto, os dados apresentaram normalidade em sua distribuição e foi possível a sequência da aplicação dos algoritmos de previsão. Os dados foram separados entre treino e teste na proporção de 80% de treino e 20% para teste, o agrupamento das médias foi realizado sobre os dados de treino, consolidando as temperaturas a cada 1°C para utilização destes dados como referência nos algoritmos de previsão. Devido aos registros estarem concentrados entre os registros de 16° e 25°C, foi aplicado o método de Interpolação sobre os dados neste intervalo e para os registros acima e abaixo destas temperaturas, o método utilizado foi de Ajuste de Curvas. Os dados utilizados para esta comparação foram os registros separados para Teste onde os resultados estimados foram cruzados com os registros efetivos do modelo e aplicado o cálculo do Erro Absoluto e Erro Relativo, sendo este último com um resultado médio de 30%. Também foi possível observar que os resultados estimados estiveram em todos os registros entre o 1º e 3º quartil. Nos resultados foi possível observar que resultados estimados pelos métodos de interpolação e ajuste de curvas apresentaram para um mesmo registro de temperatura um erro 2% e de 40%, fato que pode ser explicado por outras variáveis que impactam sobre a demanda que precisam ser aplicados ao modelo para aumento da assertividade, estes temas de sequência do projeto através da aplicação de algoritmos de *Machine Learning*.