

INFERÊNCIA ESTATÍSTICA NO CONTEXTO DE BIG DATA

XXVIII Encontro de Iniciação à Docência

Emerson Manoel Santos de Aguiar, Ronald Targino Nojosa

Com o avanço computacional das últimas décadas, facilitou-se a compreensão e o advento de ideias e práticas sobre a Estatística, bem como propiciou o aumento da quantidade de dados e também da necessidade de métodos mais apropriados para transformá-los em informação. Os chamados “big data” são grandes conjuntos de dados que podem conter, além de registros numéricos, sons e imagens, por exemplo. De modo geral, os “big data” para serem analisados necessitam de métodos estatísticos, computacionais e matemáticos. Importante também no estudo dos “big data” é o conjunto de conceitos conhecido como os quatro V's: volume, velocidade, variedade e veracidade. Objetiva-se expor neste trabalho as estratégias, com foco na Inferência Estatística, para proceder com estudos envolvendo grandes bases de dados, particularmente no que se refere à estimação de parâmetros populacionais e aos testes de hipóteses sobre estes parâmetros. Baseado nos quatro V's, tem-se métodos que vão dos descritivos (estatísticas, gráficos e tabelas) até o uso de algoritmos de mineração de dados. São apresentados resultados de aplicações que envolvem “big data” em diferentes áreas do conhecimento. Conclui-se que, dada a larga escala de conhecimentos envolvidos, o conhecimento multidisciplinar é de fundamental importância para os profissionais, incluindo o Estatístico, que almejam desenvolver estudos que envolvam “big data”.

Palavras-chave: OTIMIZAÇÃO. COMPLEXIDADE COMPUTACIONAL. APRENDIZAGEM ESTATÍSTICA. INTERDISCIPLINARIDADE.