

EXPLORANDO GRANDES CONJUNTOS DE DADOS POR MEIO DE ANÁLISE GRÁFICA

Patrícia Maria Buzzo¹; Tiago Picon¹; Alfredo Ribeiro de Freitas² e Dorival Marcos Milani³

^{1,2,3} Alunos e Professores do Centro Universitário Central Paulista - UNICEP

² Pesquisador da Embrapa Pecuária Sudeste, São Carlos, SP

O uso adequado de gráficos e Figuras numa seqüência crescente de complexidade, resume e transmite de forma precisa e ágil uma quantidade grande de informações latentes existentes dentro de um grande conjunto de dados, como é o caso de dados de produção de bovinos de corte, em que grandes arquivos de dados, são mantidos pelas associações de criadores. Dentre algumas informações úteis que podem ser obtidas destes dados por meio do uso adequado de gráficos, tem-se: vícios de coleta de dados no campo decorrentes de treinamento inadequado de funcionários, presença de *outliers*, informações sobre a distribuição (assimetria, curtose, percentis, decis e quartis), heterogeneidade de variância, sugestões de transformação, entre outras. Estas informações em conjunto são importantes para mostrar a estrutura dos dados, detectar padrões globais e sugerir refinamentos metodológicos. O objetivo deste trabalho foi utilizar, pela ordem, histogramas, gráficos em linha x-y, diagrama de caixa (*box-plot*), *scatter plot* com duas e três dimensões, gráficos rotacionados do tipo *scatter plot*, com e sem ajuste de superfície em um arquivo de dados de bovinos Nelore, da Associação Brasileira de Criadores de Zebu, com mais de 3 milhões de observações). O software utilizado foi o módulo INSIGHT do SAS. Além de identificar vícios de pesagens dos animais no campo, resumo de estatísticas descritivas, visualização espacial do relacionamento entre três variáveis: peso, ganho de peso e idade dos animais, em função de sexo, ano de nascimento dos animais, local, etc. os gráficos sugeriram vários refinamentos metodológicos que possibilitaram otimizar as estimativas de parâmetros genéticos da raça Nelore.

