

## O USO DO DIAGRAMA DE CAIXA "BOX-PLOT" PARA DETECTAR OUTLIERS EM DADOS DE PESOS DE BOVINOS ZEBUÍNOS(OK)

Tiago Picon<sup>1</sup>, Patrícia Maria Buzzo<sup>1</sup>

Alfredo Ribeiro de Freitas<sup>2</sup>, Dorival Marcos Milani<sup>2</sup>

<sup>1 e 2</sup> Alunos e Professores do Centro Universitário Central Paulista - UNICEP

No estudo de desenvolvimento ponderal de bovinos de corte, aos dados de pesos dos animais obtidos ao longo do tempo (medidas repetidas), normalmente estão associadas características: a) correlação entre os erros, cuja magnitude depende do intervalo entre as medidas; b) vícios ou tendências na pesagens, acarretando dados bimodais ou multimodais, presença de *outliers*, etc; c) não-normalidade, com dados assimétricos e curtose do tipo platicúrtica ou mesocúrtica; d) heterogeneidade de variâncias. Os itens de a) a d) quando não devidamente analisados implicam na redução do progresso genético dos animais. No presente trabalho foram analisados dados do desenvolvimento ponderal de bovinos zebuínos, oriundos da Associação Brasileira de Criadores de Zebu. Além das variáveis que caracterizavam o animal (raça, sexo, informações do pedigree, estado de origem, produtor, entre outros), o arquivo abrangia nove pesagens de cada animal: ao nascimento (PN) e oito pesagens (P1 a P8), até os dois anos de idade. Para explorar e analisar este conjunto de dados de maneira rápida e ainda conhecer características importantes de sua estrutura como as medidas de tendência central e não central, como os percentis, decis e quartis, existência de heterogeneidade de variâncias, presença de *outliers*, entre outras, foi utilizado o diagrama de caixa (*Box-Plot*), por meio do módulo INSIGHT do SAS. No diagrama de caixa, a linha horizontal no meio da caixa representa a mediana; a parte inferior e superior da caixa central indica, respectivamente, o primeiro ( $Q_1$ ) e o terceiro ( $Q_3$ ) quartil, correspondendo aos elementos de posição 25ª e 75ª da amostra. As caixas estreitas (*whiskers*), acima e abaixo da caixa central, possuem distância não superior a 1,5 vezes à distância interquartilica ( $Q_3 - Q_1$ ). Os valores mais extremos isolados e que recebem marcações individuais são considerados dados discrepantes da amostra (*outliers*). A análise dos dados revelou assimetria positiva em todas as pesagens, presença de *outliers* e ainda variância crescente na resposta dos animais com a idade. Estes resultados sugeriram que a análise dos dados de pesagens, na forma de medidas repetidas, deve ser feita considerando-se nova escala - a transformação de potência por meio da família de *Box-Cox*. Esta transformação foi feita usando-se a macro BOXGLM do SAS, a qual foi efetiva na redução dos valores das assimetria e curtose e heterogeneidade de variâncias.