

MÉDIAS WINSORIZADAS E TRIMMED APLICADAS A DADOS DE PRODUÇÃO ANIMAL

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

Quando *outliers* ou valores extremos estão presentes nos dados de uma amostra, as médias obtidas com a eliminação de uma percentagem de valores extremos, denominadas de médias *trimmed* e *Winsorized*, são estimativas robustas da média da população e são insensíveis à *outliers*; portanto, são importantes para reduzir os efeitos de valores extremos de um conjunto de dados. Após ordenar os dados de uma amostra, em ordem crescente, a média *trimmed* é calculada após deletar as k menores e as k maiores observações; a média *Winsorized* é calculada após substituir as k observações menores pela $(k+1)$ -ésima observação menor e as k maiores observações pela $(k+1)$ -ésima observação maior. Para uma distribuição simétrica, estas duas estatísticas são também simétricas e estimam a média da população. Entretanto, elas não possuem distribuição normal, mesmo quando a distribuição dos dados observados for normal. O objetivo do presente trabalho foi utilizar médias *trimmed* e *Winsorized* com o propósito de reduzir os efeitos de valores extremos em dados de nove pesagens de 541.000 bovinos zebuínos, da Associação Brasileira de Criadores de Zebu, nascidos de 1971 a 1978, coletados em condições de campo, do nascimento até dois anos de idade. A hipótese testada foi $H_0: m = m_0$ versus $H_a: m \neq m_0$, em que m é a média dos dados observados de cada variável; m_0 é a média teórica considerando-se que os dados são provenientes de uma distribuição normal. Em dados de pesos de bovinos, é comum a ocorrência de vícios de pesagens dos animais, além de problemas inerentes aos dados como presença de *outliers*, não-normalidade, com dados assimétricos e curtose do tipo platicúrtica ou mesocúrtica, heterogeneidade de variâncias, entre outros. Para cada variável foram feitas quatro etapas de cálculo, correspondendo a 0; 1,0; 2,5 e 5% dos dados deletados de cada extremo da amostra. Apenas quando foram deletados 10% dos dados de cada pesagem, ou seja, 5% de cada extremo da distribuição, o intervalo de confiança para m , ao nível de 5% de probabilidade, incluiu a média m_0 , sendo os resultados obtidos com as médias *trimmed*, ligeiramente superior.

CPPSE
AIN
SEPARATA