

Uso do envelope simulado pelo *Rstudio* para detectar outliers em dados de contagem de *Euschistus heros*

Maria Cristina Neves de Oliveira¹; Samuel Roggia¹;
Elton Massahiro Saito Loures²

¹Embrapa Soja, Londrina, PR, Brasil. mcristina.oliveira@embrapa.br; ²Universidade Estadual de Londrina

Resumo

Uma fase muito importante na análise de dados é conhecer a origem dos dados antes de qualquer análise mais complexa. Em técnicas estatísticas, esse conhecimento diz respeito a identificar pontos discrepantes que indicam anormalidades do conjunto de dados ou erros na medição realizada. É primordial que sejam calculados corretamente, assegurando uma compreensão adequada dos dados e que levará a conclusões mais corretas do estudo. A presença de dados discrepantes ou *outliers* pode alterar e/ou comprometer a interpretação de um conjunto de dados. Vários são os testes e gráficos que podem ser utilizados para verificar se um conjunto de dados apresenta ou não dados discrepantes e essa metodologia é indicada por vários autores para que uma análise estatística seja eficiente e fidedigna. Entre eles destacam-se o teste de Grubbs, o histograma de frequências, estatísticas Resumo, o boxplot e a carta de controle. O objetivo desse trabalho foi analisar a ocorrência de *outliers* em dados de contagem de *Euschistus heros* na cultura da soja, na safra 2011/2012. O experimento foi realizado na área experimental da Embrapa Soja, na Fazenda Maravilha, com as coordenadas 23° 28' 44" S; 50° 59' 03" O e altitude média de 465 m, numa área de aproximadamente 3,38 ha. As amostragens foram sistematizadas, ou seja, georreferenciadas numa grade de 10 X 10 metros. Em cada ponto amostral, foram feitas quatro batidas de panos num metro linear e foi calculada a média dos insetos encontrados no estádio reprodutivo da soja R5.4. No estudo foi usado o gráfico de envelope simulado pela linguagem do *Rstudio*, para observar os *outliers*. Essa estrutura de programação é de um software livre que já vem sendo utilizado em resultados de diferentes áreas de pesquisa. Pode-se concluir que o envelope simulado permitiu detectar o distanciamento entre as amostras fora do conjunto de dados e com os critérios da assimetria e curtose, foi possível indicar a retirada dos outliers, mantendo-se os dados representativos da contagem de *Euschistus heros*.

Termos para indexação: *Euschistus heros*; *Rstudio*; dados anômalos