

# Métodos Exploratórios na Avaliação da Distribuição Espacial de Percevejos da Soja

Farias, C.H.<sup>1</sup>; Oliveira, M.C.N. de<sup>2</sup>; Roggia, S<sup>2</sup>.

<sup>1</sup>Universidade Norte do Paraná (Unopar), <sup>2</sup>Embrapa Soja | carlos@cnpso.embrapa.br

## Introdução

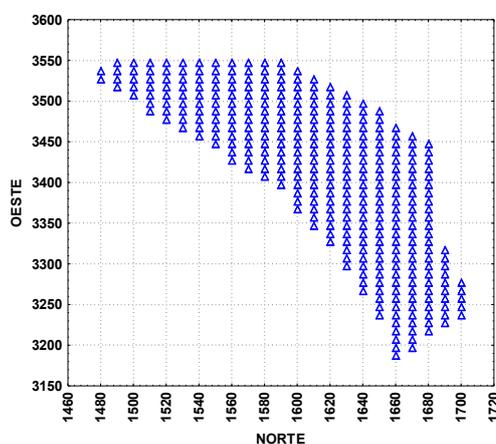
A grande demanda por alimentos para a população mundial estimulam a realização de estudos que intensificam o controle das pragas para aumento da produtividade sem aumentar a área plantada, como o caso da soja. As diferentes espécies de percevejos são pragas que vêm sendo estudadas e várias são as técnicas agronômicas e estatísticas para melhor avaliar e interpretar estas pesquisas de controle com acurácia. A eficiência destas pesquisas também está diretamente relacionada com os procedimentos anteriores a instalação do experimento como o uso do planejamento experimental adequado. O planejamento do experimento é de extrema importância para se obter conclusões de forma adequada evitando erros que podem ocorrer na obtenção dos dados, na digitação dos mesmos, obtenção de dados na leitura de equipamentos, entre outras situações. Muitos problemas são amenizados se esta prática for assessorada por profissionais da área de estatística, pois a falha na seleção de variáveis, na determinação do número de repetições e ou delineamento podem comprometer a verificação das hipóteses formuladas, para a obtenção das respostas aos questionamentos elaborados (BANZATTO & KRONKA, 1989). Além destes erros sistemáticos, os experimentos estão sempre sujeitos a erros ou variações, que são devidos a fatores não controlados como, pequenas variações no ambiente ou no material (ZIMMERMAN, 2004). A análise exploratória de dados nos fornece um extenso repertório de métodos para um estudo detalhado dos dados. Nessa abordagem, a finalidade é obter dos dados a maior quantidade possível de informação, que indique modelos plausíveis de serem utilizados numa fase posterior à análise conclusiva dos dados da pesquisa. O objetivo desse trabalho é informar a importância da análise exploratória na pesquisa agrícola e quais métodos selecionar para melhor interpretação dos dados.

## Materiais e Métodos

Os dados utilizados nesse estudo foram conduzidos no campo experimental da Embrapa Soja no Distrito de Maravilha (Londrina-PR). Localizam-se segundo as coordenadas 23° 28' 44" S; 50° 59' 03" O, e altitude média de 465 m, uma área de aproximadamente 3,38 ha para contagem de insetos da espécie *Piezodorus guildinii* (Figura 1).

Foram utilizadas as estatísticas como a média, a moda, a mediana, os coeficientes de assimetria e curtose, os gráficos como o boxplot, os histogramas de frequência, a superfície de resposta, a carta de controle e os testes de normalidade. Foi utilizado o boxplot e a carta de controle para detectar pontos discrepantes. Esses pontos foram encontrados considerando a mediana (q2),

o quartil inferior ( $q_1$ ), o quartil superior ( $q_3$ ) e o intervalo interquartil ( $IQR = q_3 - q_1$ ). A linha central da caixa marca a mediana do conjunto de dados. A parte inferior da caixa é delimitada pelo quartil inferior ( $q_1$ ) e a parte superior pelo quartil superior ( $q_3$ ). As hastes inferiores e superiores se estendem, respectivamente, do quartil inferior até o menor valor não inferior a  $q_1 - 1,5IQR$  e do quartil superior até o maior valor não superior a  $q_3 + 1,5IQR$ . Os valores inferiores a  $q_1 - 1,5IQR$  e superiores a  $q_3 + 1,5IQR$  são representados individualmente no gráfico sendo estes valores caracterizados como outliers. As quantidades  $q_1 - 1,5IQR$  e  $q_3 + 1,5IQR$  delimitam, respectivamente, os limites inferior e superior e se ocorrem além destes, são outliers. A carta de controle é uma ferramenta usada na indústria para avaliar se existem peças que estão fora do padrão recomendado e é possível detectar se existem pontos amostrais além das linhas inferior e superior aos três desvios padrões considerados neste trabalho. Esse critério é eficiente para o estudo da estabilidade estatística de um processo, a partir da observação de sequências aleatórias de amostras, coletadas a intervalos regulares (RAMOS, 1995). Foram utilizados três testes de normalidade para avaliar a distribuição da população de percevejos da espécie *Piezodorus guildinii*: o teste de Shapiro & Wilk (1965) deve ser aplicado para tamanho de amostras menor ou igual a 2000. Sua estatística  $W$  está no intervalo  $0 < W < 1$ , no caso em que  $W$  é igual a 1, afirma-se que os dados provêm de uma distribuição normal e, para pequenos valores de  $W$  indicam ausência de normalidade. O teste de Kolmogorov-Smirnov baseia-se na máxima diferença entre a distribuição acumulada da amostra e distribuição acumulada esperada. Se o valor calculado de  $D$  é estatisticamente significativo para ( $p \leq 0,05$ ) rejeita-se a hipótese que a distribuição estudada é normal. Não existe restrição para pequenas amostras e é mais rigoroso que o teste de  $\chi^2$ . É um teste que pode ser aplicado aos dados individualmente ou quando as observações estão estruturadas numa distribuição de frequência. O teste de Cramer-von Mises é uma alternativa do teste de Kolmogorov-Smirnov (Campos, 1979). Foi também avaliada a distribuição espacial dos insetos nas diferentes direções, pelo gráfico de superfície, a distribuição espacial dos percevejos.



**Figura 1.** Mapa de amostragem em grid regular para avaliação da distribuição espacial de percevejos na lavoura de soja, na Fazenda Maravilha, Embrapa Soja, safra 2011/2012.

## Resultados e Discussão

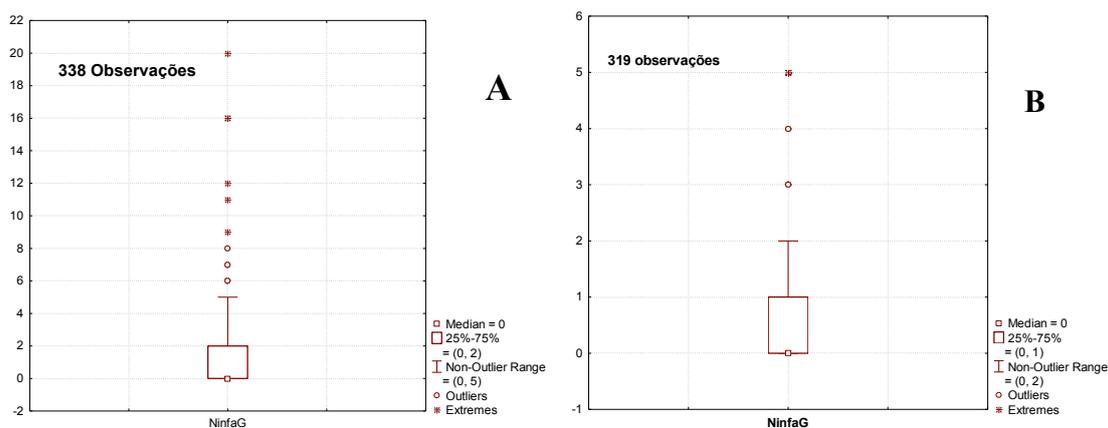
Pode-se verificar que inicialmente com 338 observações, na presença de vários valores discrepantes, os resultados para os coeficientes de assimetria, de curtose e de variação foram 3,65; 18,07 e 189,30 respectivamente, indicando ausência de normalidade, devido aos dois primeiros coeficientes afastar-se de zero. As medidas de assimetria permitem analisar uma distribuição conforme as relações entre suas medidas de moda, média e mediana, quando observadas graficamente. Considera-se neste caso dos percevejos é uma distribuição assimétrica, pois os valores para a moda, a média e a mediana são diferentes. Os teste de Shapiro & Wilk (1965), Kolmogorov-Smirnov (Campos, 1979) e Cramer von Mises (Campos,

1979) foram unânimes em rejeitar a hipótese de normalidade dos dados (Tabela 1). Além das estatísticas descritivas os métodos gráficos também foram usados para apresentar as respostas do levantamento populacional dos percevejos. O *boxplot* permitiu avaliar a simetria dos dados, sua dispersão e a existência de *outliers* (dados discrepantes). Na retirada dos *outliers* conforme indicado na metodologia houve uma expressiva redução de todas as estatísticas.

**Tabela 1.** Resultados das estatísticas descritivas para o número de ninfas grandes da espécie *Piezodorus guildinii*, amostradas com georreferenciamento no campo experimental da Fazenda Maravilha, da Embrapa Soja.

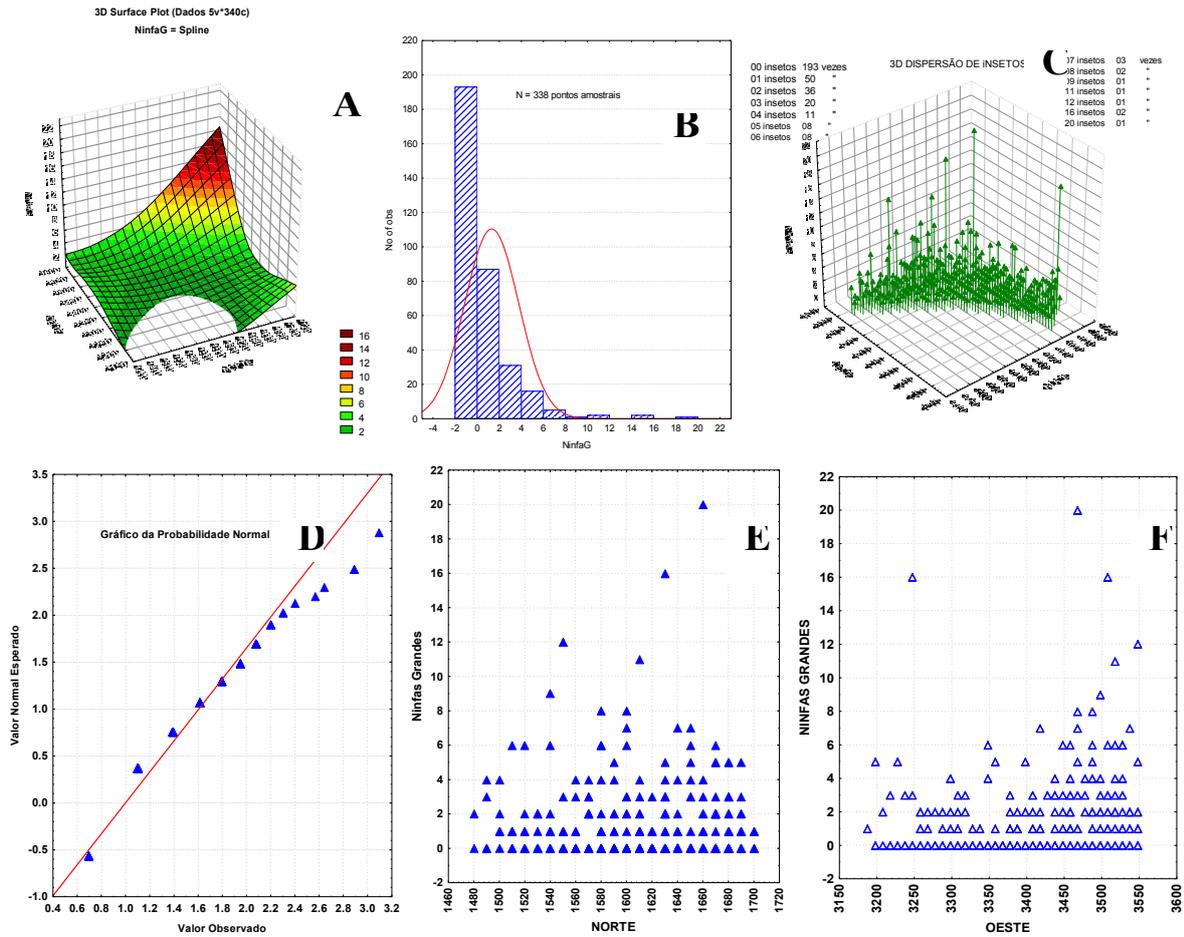
Estatísticas	Com outliers	Sem outliers
Nº Amostras	338	319
Média	1,28	0,83
Mediana	0	0
Moda	0	0
Mínimo	0	0
Máximo	20	5
Variância	5,96	1,65
Coeficiente de assimetria	3,65	1,58
Coeficiente de curtose	18,07	1,74
Coeficiente de variação (%)	189,30	153,88
Desvio padrão	2,44	1,28
Shapiro & Wilk	W= 0,574 Pr(W) =0,0001	W=0,694 Pr(W)=0,0001
Kolmogorov-Smirnov	D= 0,298 Pr > D <0,010	D= 0,347 Pr > D <0,010
Cramer-von Mises	W-Sq= 8,48 Pr> W-Sq < 0,005	W-Sq= 7,40 Pr> W-Sq < 0,005

Ainda nesta tabela, após a retirada de vários *outliers* os coeficientes de assimetria e curtose estão distantes de zero. Cabe aqui o bom senso do pesquisador da necessidade ou não da retirada de mais valores. Os três testes aplicados para avaliar a distribuição dos dados rejeitaram a hipótese de que a população de percevejos são provenientes de distribuição normal e este fato interfere nas inferências obtidas para o modelo georreferenciado a ser aplicado à posteriori (Tabela 1). O *boxplot* permitiu avaliar a simetria dos dados, sua dispersão e a existência de *outliers* (dados discrepantes). Na retirada dos *outliers* conforme indicado na metodologia houve uma expressiva redução de todas as estatísticas calculadas. O *boxplot* identificou a presença de *outliers*, cujos valores variaram de seis a 20 percevejos. Pode-se verificar que o conjunto de dados ainda apresenta variabilidade, bastante comum em estudos com população de insetos (Figura2 A,B).



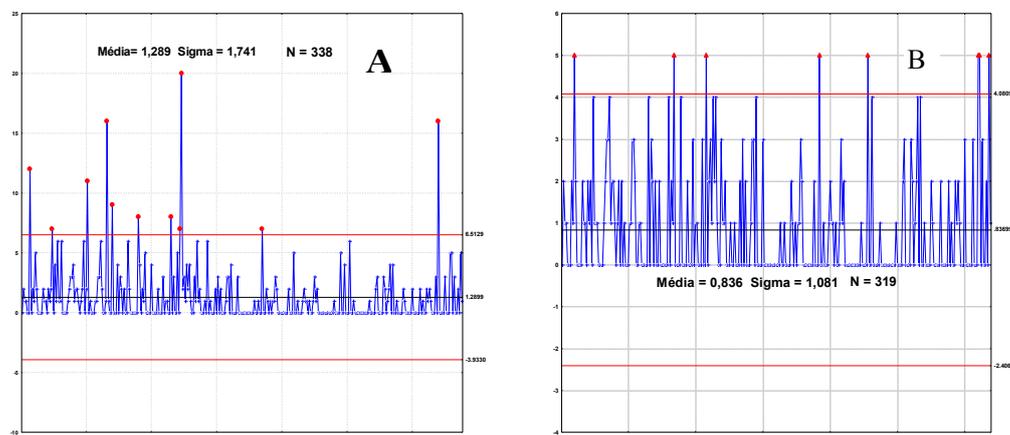
**Figura 2.** *Boxplot* (A) indicando os dados discrepantes com 338 observações e com um ponto discrepante com 319 amostras (B).

Outra opção de apresentar os dados é por meio dos gráficos de superfície, da probabilidade normal, do histograma de frequência, que indicaram a ocorrência dos percevejos em três dimensões com 338 pontos amostrais, o tipo da distribuição que seguem os mesmos dados e a dispersão de acordo com as coordenadas cardinais (Figs. 3A, 3B, 3C, 3D, 3E e 3F).



**Figura 3.** Dispersão de insetos tridimensional com 338 pontos (A, C) e distribuição de frequência dos percevejos (B), probabilidade normal (D), dispersão dos percevejos no lado Norte (E) e dispersão dos percevejos no lado Oeste da área experimental (F)

A carta de controle (Figura 4 A) detectou os mesmos pontos discrepantes já indicados na Figura 2 A. Esse gráfico em geral é mais rigoroso que os demais gráficos que detectam *outliers*. Na retirada dos pontos discrepantes foi observado que na carta de controle (Figura 4B), embora ainda apresentem alguns valores distantes da massa de dados, houve maior uniformidade para os demais dados entre os limites superiores e inferiores. Os métodos exploratórios numéricos e gráficos são essenciais antes de qualquer análise conclusiva. É por meio deste processo que avaliamos se existem valores discrepantes no conjunto de dados, e segundo Hoaglin et al (1922) contribui para aumentar a eficácia da análise estatística. Salienta-se que tanto para dados sistematizados (georreferenciados) quanto aos dados aleatorizados a análise exploratória é uma ferramenta importante na análise de dados.



**Figura 4.** Cartas de Controle para as mesmas situações dos boxplot, com 338 observações (A) e 319 observações (B) após a retiradas dos outliers.

## Conclusão

Pode-se concluir que:

- com o diagnóstico exploratório detectam-se pontos discrepantes nos dados com dependência espacial;
- os métodos gráficos são importantes na avaliação de resultados de pesquisas;
- a carta de controle de qualidade é mais rigorosa que o boxplot para detectar outliers;
- com a retirada dos valores discrepantes os resultados da análise estatística têm maior precisão e, a análise exploratória permite obter respostas com confiabilidade e eficácia.
- o diagnóstico exploratório é de grande importância para detectar pontos discrepantes que podem invalidar respostas de pesquisas.

## Referências

BANZATO, D. A.; KRONKA, S. N. Experimentação agrícola. 4. ed. Jaboticabal: Funep, 2006. 237 p.

ZIMMERMAN, F. J. P. Estatística aplicada à pesquisa agrícola. Santo Antônio de Goiás: Embrapa Arroz e Feijão, 2004. 402p.

HOAGLIN, D. C.; MOSTELLER, F.; TUKEY, J. W. Análise exploratória de dados. Técnicas Robustas. Edições Salamandra. Lisboa. 446 p.

VIEIRA, S.; WADA, R. Estatística: Introdução ilustrada. São Paulo: Atlas, 1986.

RAMOS, A. W. Controle estatístico de processo para pequenos lotes. São Paulo: Edgard Blucher, 1995. 151p.