

## MÉTODO PARA TRATAMENTO E ANÁLISE EM ZONAS DE POTENCIAL PRODUTIVO PARA DADOS DE MONITOR DE COLHEITA

Lara Marie Guanais Santos<sup>1</sup>, Ricardo Ralisch<sup>2</sup>, Otávio Jorge Grigoli Abi Saab<sup>2</sup>, Hevandro Colonhese Delalibera<sup>4</sup>

<sup>1</sup> Doutoranda em Agronomia, Universidade Estadual de Londrina - UEL, Londrina - PR, Fone: (14) 996204225, laramarie\_guanais@hotmail.com

<sup>2</sup> Docente Dr. do Programa de Pós Graduação em Agronomia, Universidade Estadual de Londrina - UEL, Londrina - PR.

<sup>3</sup> Pesquisador Dr. da Área de Engenharia Agrícola e Tecnologias Digitais do Instituto de Desenvolvimento Rural do Paraná - IDR-Paraná, Londrina - PR, Fone: (43) 3376 2423, hevandro@idr.pr.gov.br

Apresentado no  
XLIX Congresso Brasileiro de Engenharia Agrícola - CONBEA 2020  
23 a 25 de novembro de 2020 - Congresso On-line

**RESUMO:** Este trabalho tem por objetivo desenvolver uma metodologia para tratamento de dados de produtividade coletados através de monitor de colheita, com base em critérios estatísticos, visando eliminar dados discrepantes sem alterar o comportamento do fenômeno e propor a separação da área em zonas de potencial produtivo. Foram utilizados dados da cultura da soja que foram tratados pelo método *Jackknife* e teste de linearidade, o qual concentrou a distribuição entre os índices de -3,09 a 3,09 e resultou no aumento do grau de linearidade de 0,59 para 0,85. Após este, foram aplicadas classes para separação da área em zonas de potencial produtivo a partir dos intervalos de probabilidade de um desvio para mais e para menos, resultando na divisão da área em 7,5% com baixo índice de potencial produtivo, 87,4% com potencial médio normal local e 5,1% com alto índice de potencial produtivo. A aplicação do método de tratamento de dados e separação em zonas de potencial produtivo foi considerado simples e foi uma boa solução para adequar dados advindos de monitores de colheita através de métodos estatísticos consagrados.

**PALAVRAS-CHAVE:** Erro sistemático; dados discrepantes; método estatístico.

## METHOD FOR PROCESSING AND ANALYSIS OF PRODUCTIVE POTENTIAL AREAS FOR DATA MONITOR HARVEST

**ABSTRACT:** This work aims to develop a methodology for productivity processing data collected through a harvest monitor, based on statistical standard, which aims to eliminate discrepant data without changing the phenomenon behavior and proposing the separation of the area in potential productive zones. Soybean culture data were used, which were treated by the Jackknife method and linearity test, which concentrated the distribution between the indices from -3.09 to 3.09 and resulted in an increase in the degree of linearity from 0.59 to 0.85. After this, classes were applied to separate the area into potential productive zones from the probability intervals of one deviation upwards and downwards, resulting in the division of the area by 7.5% with a low index of productive potential, 87.4 % with average normal local potential and 5.1% with high index of productive potential. The application of the data processing and separation method in zones of productive potential was considered simple and was a good solution to adapt data from harvest monitors through established statistical methods.

**KEYWORDS:** Systematic error; outliers; statistical method.

**INTRODUÇÃO:** O método *Jackknife* (WU et al., 1986), é uma proposta que visa a remoção de dados consideradas não pertencentes a população e ou discrepantes, através da aplicação de análises de desvio sobre os dados e a sua comparação com distribuições teóricas, como o teorema de Laplace e o teorema do limite central (desvio padronizado e desvio Studentizado respectivamente), os quais são base de boa parte da estatística moderna. Estes são formas de escalamento dos dados de qualquer fenômeno contínuo ao intervalo destas distribuições teóricas (distribuição  $z$  e  $t$ ). Ressalva-se que existem outras distribuições teóricas que podem ser aplicadas como meio de comparação, porém, seu desenvolvimento é mais complexo e estas devem ser aplicadas conforme a natureza dos dados e, para este trabalho, não foram estudadas. Este trabalho tem por objetivo apresentar uma metodologia para tratamento de dados advindos de monitor de colheita, visando melhorar a qualidade dos dados e a separação da área em zonas de potencial produtivo, tendo como base métodos e critérios estatísticos consagrados.

**MATERIAL E MÉTODOS:** Foram utilizados dados referentes a cultura da soja (colhida em 07/02/2019) de área agrícola em sistema plantio direto, localizada no município de Cambé, PR (23° 8'29,99" S, 51°20'48,51" O e 521 m de altitude). A área é constituída por Latossolo Vermelho e está sob clima Cfa. Os dados foram obtidos por meio de uma colhedora John Deere modelo STS 9750 equipada com sensor de produtividade tipo placa de impacto, sensor de umidade por capacitância, sensor de velocidade magnético instalado nas rodas motrizes, sensor de inclinação tipo giroscópio e monitor modelo GS4 que apresenta e grava os dados.

Por meio do software RStudio, foi desenvolvido um script para remoção de dados zerados e nulos do conjunto de dados de produtividade da área, eliminando sempre a amostra inteira para que estes não perdessem sua referência de dados. Após este tratamento prévio, foi feita a correção da produtividade para 13% de umidade (AGUIAR, 1977). A planilha resultante foi submetida a aplicação do método *Jackknife* com escalamento para ajuste à distribuição “ $z$ ” pelo desvio padronizado. Com o uso deste método os dados de produtividade ( $\text{kg ha}^{-1}$ ) são convertidos para índices de desvio, sendo que, para estes, os valores abaixo de -3,09 ou acima de 3,09 são considerados não pertencentes a população conforme a distribuição “ $z$ ” e, portanto, foram removidos (WU *et al.*, 1986).

Deve-se iniciar a remoção sempre pelo índice mais extremo, um por vez, considerando o recálculo dos índices conforme recomendado por Delalibera *et al.* (2017). A exclusão de dados dentro do intervalo anteriormente citado deve ser criteriosa, pois, pode proporcionar a alteração do comportamento e da distribuição do fenômeno (SNEDECOR & COCHRAN, 1980), portanto, para tal, se aplicou um teste de linearidade ao conjunto de dados dos índices desvio, visando verificar o grau de linearidade do evento e, caso necessário, remover dados para melhorar a linearidade ( $r^2$ ) e, conseqüentemente, o ajuste a uma distribuição normal.

Após a aplicação deste tratamento no conjunto de dados, foi realizada a interpolação pelo método Redes de Triangulação Irregular (TIN) e a confecção do mapa de zonas de potencial produtivo, por meio do software SIG QGIS 3.4.6. A escolha do método TIN se deve a grande densidade de dados por unidade de área, proporcionada pelo monitor de colheita, pois, não se deseja suavizar e nem estimar informações, visto que, um método estatístico foi aplicado para excluir dados não representativos do conjunto e, portanto, a TIN se mostra mais adequada para tal, por ser uma representação fiel aos pontos amostrados (MIKHAIL *et al.*, 2001).

Após a interpolação, utilizou-se o modo de probabilidade de um desvio padrão da média da distribuição ajustada do evento para mais e para menos para realizar a classificação do mapa em três zonas de potencial produtivo. As áreas com índices menores que -1 desvio foram consideradas de baixo índice de potencial produtivo, entre -0,999 a 0,999, de potencial médio normal local e, acima de 1 desvio, alto índice de potencial produtivo. A partir dos mapas de zonas de potencial produtivo, se extraiu, via software QGIS, os histogramas dos dados de índices de produtividade. Para quantificar as áreas correspondentes aos potenciais produtivos,

foi efetuado a conversão de raster para vetor e, em seguida, selecionada cada classe de interesse para a obtenção dos valores em hectares para cada zona de potencial produtivo gerada.

**RESULTADOS E DISCUSSÃO:** Após a eliminação dos dados nulos e zerados e a aplicação do método *Jackknife* junto ao teste de linearidade, foi possível observar uma melhora no grau de linearidade do evento, demonstrado pelo  $r^2$ , o qual foi calculado a partir da frequência ordenada dos índices de desvio (se aplicou a remoção de dados extremos até a obtenção de  $r^2$  superior a 0,8). A Figura 1 apresenta que, com a aplicação do *Jackknife* junto a verificação do ajuste de linearidade, houve uma melhora no ajuste dos dados a distribuição teórica  $z$ . Ainda se ressalva que a melhora no ajuste da linearidade do diagrama de frequência ordenada favorece o ajuste da distribuição dos dados a normalidade, aumentando a representatividade das medidas de tendência central do fenômeno avaliado. Este procedimento de verificação da linearidade dos dados é similar aos procedimentos que alguns testes de normalidade utilizam na verificação do enquadramento a normalidade de conjuntos de dados, como por exemplo Lilliefors e Anderson-Darling, porém estes métodos não são recomendados para conjuntos muito grandes de dados (Figura 1B,  $\approx 6.000$ ), como é o caso de dados de colheita, devido as regras dos testes citados, portanto, se fez uma adaptação destes métodos, conforme proposto neste trabalho.

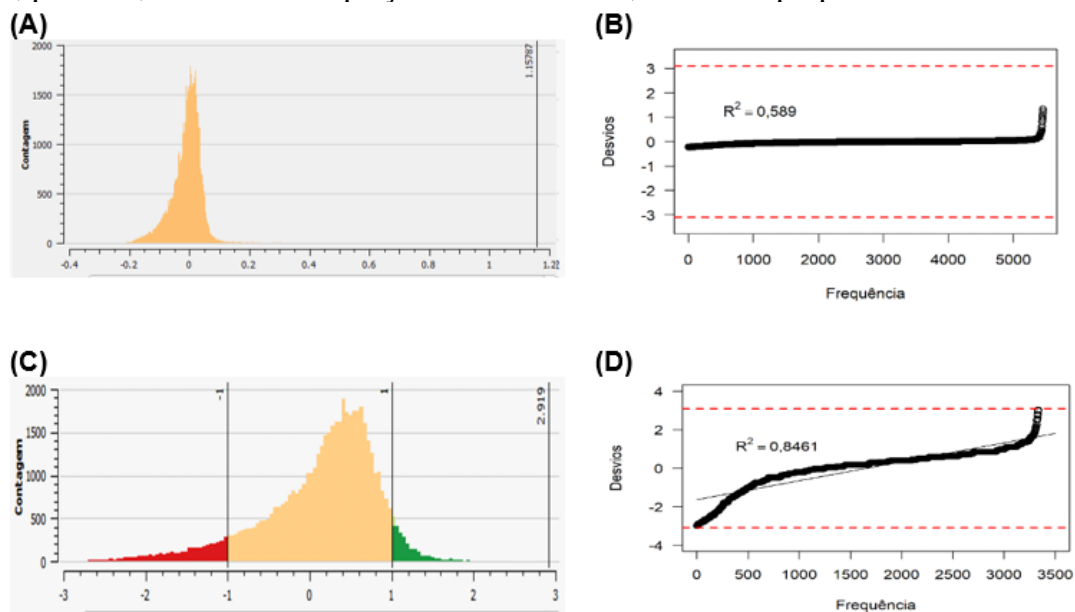


FIGURA 1. Histogramas dos índices de produtividade e, grau de linearidade dos diagramas de frequência ordenada dos índices de desvio padronizado, referente a produtividade de soja antes e após da aplicação do método *Jackknife*.

A - Histograma sem tratamento; B – Dados de soja sem tratamento; C – Histograma com tratamento; D – Dados de soja com tratamento Jackknife mais as exclusões para melhora do ajuste a linearidade.

Nas Figuras 1A e B, embora não apresente índices fora do intervalo crítico, este é proporcionado pelo elevado grau das medidas de dispersão estatística do fenômeno, o que leva a um baixo grau de ajuste a linearidade, portanto, foram excluídos os dados mais distantes do centro da distribuição, visando aumentar o grau de linearidade e melhor ajuste do evento a distribuição “ $z$ ”, aumentando também a representatividade das medidas de tendência central. É importante relatar que a exclusão de dados discrepantes não tem uma relação com sua distribuição espacial na área, e também, por se tratar de uma variável contínua, sua distribuição de frequência não deve se apresentar binomial ou multinomial, pois, caso isso ocorra, deve-se considerar duas populações estatísticas distintas e devem ser analisadas separadamente.

Os valores discrepantes excluídos são estimativas que estão sobre efeito de fatores não controlados randômicos e extrínsecos ao fenômeno avaliado. Dentre os fatores intrínsecos a

máquina pode-se citar colheita com meia plataforma e erros de sensores, dentre outros que são normais ao processo de colheita, ou seja, são sistemáticos e tendem a sempre se comportar da mesma forma e, portanto, não necessariamente devem ser excluídos, pois, fazem parte do evento mensurado. Por estes não apresentarem repetibilidade em alocação espacial entre uma safra e outra, ao longo do tempo, estes podem ser diluídos com uma álgebra de mapas.

A Figura 2 corresponde aos mapas de produtividade, convertidos para desvios padronizados como forma de visualização em uma mesma grandeza de apresentação, onde a Figura 2A corresponde a um mapa de produtividade sem a aplicação do *Jackknife*, porém, com a aplicação da classe de um desvio para mais e para menos e, a Figuras 2B, representa o mapa com a aplicação completa do tratamento proposto. Nestes é possível visualizar claramente como os dados discrepantes alteram o comportamento e a distribuição do evento, sendo no primeiro momento (Figura 2A), como as medidas de dispersão são muito grandes, praticamente, não há valores enquadrados fora do intervalo -1 a 1 desvio padronizado, mostrando que a produtividade da área é aparentemente homogênea, fato que é desmistificado no mapa B.

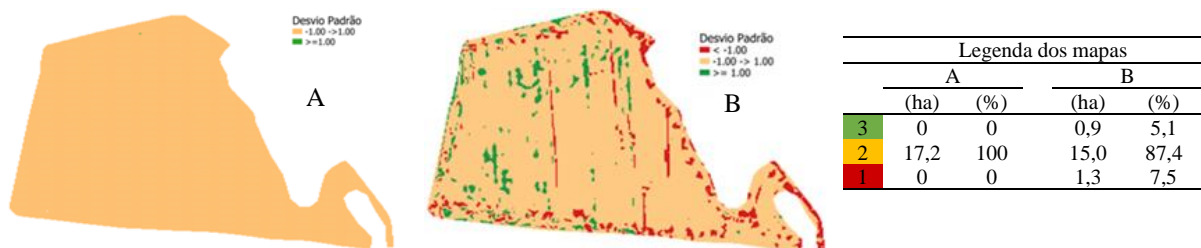


FIGURA 2. Mapas de desvios padronizados para o rendimento das culturas, sem e com a aplicação do método *jackknife* mais o ajuste a linearidade, classificados em três zonas de potencial produtivo por mais ou menos um desvio padrão da média da distribuição.

A - Dados sem tratamento; B - Dados com tratamento; 1- baixo índice de potencial produtivo; 2 - potencial médio normal local; 3 - alto índice de potencial produtivo

**CONCLUSÕES:** A aplicação do método proposto para análise de dados obtidos a partir de monitor de colheita é viável, pois, pode-se aplicar a remoção de dados discrepantes e/ou não pertencentes a população de forma fácil e, ainda, para o mapa resultante, é possível classificá-lo em zonas de potencial produtivo, utilizando meios e critérios estatísticos consagrados. Também, a partir do mapa de índices de potencial produtivo final, é possível realizar álgebras de mapas com culturas distintas, visto que, independente dos dados de entrada da análise, os dados de saída estão em uma mesma grandeza.

## REFERÊNCIAS:

AGUIAR, P.A.A. **Padronização da umidade dos grãos na experimentação agrícola.** In: Embrapa Semiárido-Resumo em anais de congresso (ALICE). In: Empresa Brasileira De Pesquisa Agropecuária. Centro de Pesquisa Agropecuária do Trópico Semi-Árido. Petrolina. Resumo de atividades de pesquisa. Petrolina, 1977.

DELALIBERA, H.C.; HIGASHIBARA, L.R.; RALISCH, R. Estimativa de perdas na colheita mecanizada de grãos: Adequação da metodologia para plataformas universais modernas. In: **Anais** do XLVI Congresso Brasileiro de Engenharia Agrícola – CONBEA, 2017.

MIKHAIL, E. M.; BETHEL, J. S.; MCGLONE, J. C. **Introduction to modern Photogrammetry.** John Wiley & Sons, p. 479, 2001.

SNEDECOR G.W., COCHRAN, W. G. **Statistical Methods.** Iowa State University Press. Biometry, ed.7, p.593, 1980.

WU, C.J. et al. Jackknife, bootstrap and other resampling methods in regression analysis. **The Annals of Statistics**, v.14, n.4, p.1261-1295, 1986.