

## TRATAMENTO DE DADOS E ALGEBRA DE MAPAS DE PRODUTIVIDADE UTILIZANDO OS MÉTODOS DE ANÁLISE DE DESVIOS

GUANAIS, L. M. S.<sup>1</sup>, ARMACOLO, N. M.<sup>2</sup>, RALISCH, R.<sup>3</sup>, DELALIBERA, H. C.<sup>4</sup>

<sup>1</sup>Mestranda em Agronomia - Universidade Estadual de Londrina, Londrina, Paraná, Fone:(14)99620-4225  
laramarie\_guanais@hotmail.com

<sup>2</sup>Doutoranda em Agronomia, Universidade Estadual de Londrina - UEL, Londrina - PR

<sup>3</sup>Prof. Dr. do Programa de Pós Graduação em Agronomia, Universidade Estadual de Londrina - UEL, Londrina - PR.

<sup>4</sup>Doutor, pesquisador no Instituto Agrônomo do Paraná - IAPAR

Apresentado no  
XLVIII Congresso Brasileiro de Engenharia Agrícola - CONBEA 2019  
17 a 19 de setembro de 2019 - Campinas - SP, Brasil

**RESUMO:** Os mapas de produtividade são considerados os pontos de partida para visualização de possíveis zonas de manejo de uma área. Estes mapas são compostos por um grande conjunto de informações referenciadas no espaço, amostrados, mensurados e gravados por instrumentação instalada nas colhedoras e, como de praxe em instrumentação, permite uma grande densidade de dados. Nestes conjuntos de dados é comum existirem uma grande quantidade de valores que não são representativos ao fenômeno medido, devido a diversos fatores ligados a máquina, a área de trabalho e a própria instrumentação. Devido estes fatores o presente trabalho tem por objetivo estabelecer um método de tratamento de dados através de análises estatísticas consagradas, como a análise de resíduos Studentizados e Padronizados, com o objetivo de eliminar dados discrepantes ou não pertencentes a população, sem proporcionar alterações na distribuição e no comportamento do fenômeno medido (produtividade). Estes métodos também proporcionam escalonamento nas variáveis, permitindo que seja realizada álgebras entre mapas de colheita de culturas distintas e assim gerar informações baseado em históricos de dados, no qual efeitos de clima e cultura são diluídos, permitindo a assistência técnica tomar decisões mais assertivas ao zoneamento e manejo da fertilidade do solo das áreas agrícolas.

**PALAVRAS-CHAVE:** Limpeza de dados, Mapas de Colheita, Método Estatístico.

### DATA PROCESSING AND ALGEBRA OF PRODUCTIVITY MAPS USING DEVIATION ANALYSIS METHODS

**ABSTRACT:** The productivity maps are considered the starting points for visualization of possible management zones of an area. These maps consist of a large set of information referenced in space, sampled, measured and recorded by instrumentation installed in the harvesters and, as a standard in instrumentation, allows a high density of data. In these datasets it is common to have a large amount of values that are not representative to the measured phenomenon, due to various factors connected to the machine, the work area and the instrumentation itself. Due to these factors, the aim of this work is to establish a method of data processing through established statistical methods, such as the analysis of studentized and standardized deviation, with the objective of eliminating discrete data or not belonging to the population, without providing alterations in the distribution and behavior of the measured phenomenon (productivity). These methods also provide escalation in the variables, allowing to be performed algebras between crop maps of different crops and thus generate information based on data histories, in which effects of climate and culture are diluted,

Allowing technical assistance to make more assertive decisions to zoning and soil fertility management in agricultural areas.

**KEYWORDS:** Data cleansing, Harvest maps, Statistical method.

**INTRODUÇÃO:** Com a adoção da Agricultura de Precisão (AP) na colheita, foram agregadas tecnologias como sensores, sistema de geoposicionamento, coleta, armazenamento, tratamento e transmissão de dados (QUEIROZ et al., 2000). Para Molin (1997) os mapas de produtividade são o ponto de partida tido por usuários e pesquisadores para visualizar a variabilidade espacial das lavouras, por materializar a resposta da cultura. Estes mapas consistem em um conjunto de pontos, onde cada um representa a medida da quantidade e umidade dos grãos colhidos em uma área correspondente à largura da plataforma de colheita da colhedora pela distância no tempo percorrido entre duas leituras, referenciada pela coordenada geográfica (MOLIN, 2000). Esta informação é de grande valia, ajudando na tomada de decisões de gerenciamento e manejo da área (BRASE, 2006). Estes sistemas de monitoramento permitem configurar a frequência de aquisição dos dados de 250 a 1500 pontos por hectare, variando o tamanho em função da largura da plataforma. Os dados obtidos são interpolados, através de *softwares* do tipo Sistema de Informação Geográfica (SIG) para assim serem convertidos em informação útil (MOLIN, 2000), porem, nesta grande quantidade de dados adquiridos automaticamente, também são incorporados erros (MENEGATTI et al., 2004) que devem passar por tratamentos para remoção dos desvios e assim tornar estes representativos.

**MATERIAL E MÉTODOS:** Como entrada de dados devem ser utilizados valores de produtividade coletados por sistema de monitoramento de colheita/produtividade, exportados em formato para uma tabela do *software* Excel. Nesta tabela deve conter dados de produtividade, umidade do grão, coordenadas geográficas e velocidade de colheita. Pode ser feita remoção prévia de valores zerados ou nulos em todos as variáveis, eliminando-se a amostra (linha) inteira, para que estas não percam sua referência de dados. É também necessário agrupar os valores repetidos sequentes, onde o intervalo entre a obtenção de um ponto e outro é muito curto, pois, podem ser erros podendo afetar as medidas de tendência central. Após esta limpeza prévia, pode-se aplicar a análise de resíduos pelo método dos Desvios Padronizados (Equação 1), nos casos onde não há ou não se considera tratamentos distintos (como tratamentos pode-se considerar cultivares diferentes, culturas diferentes e até áreas diferentes, lembrando que é necessário análises separadas para cada tipo de cultura, isto é, não se pode colocar dados de milho dentro de uma análise com soja, devendo seus desvios ser calculados separadamente). A análise dos Desvios Padronizados pode ser utilizada como um filtro de base estatística para exclusão de amostras consideradas não pertencentes à população analisada, utilizando como parâmetros os valores da tabela “Z” com referência no valor crítico de  $\pm 3,09$  nos quais dados que apresentam índice de Desvio Padronizado fora deste intervalo são considerados como não pertencentes a população. A exclusão de dados dentro do intervalo citado, deve ser criteriosa e seguir os parâmetros da estatística “z”, recomendada para  $\sigma$  conhecido, pois podem alterar a distribuição dos dados e assim o comportamento da população estudada, ou seja, o conjunto de dados pode não mais representar a realidade do fenômeno estudado (SNEDECOR & COCHRAN, 1980). Nota-se que a aplicação do método dos Desvios Padronizados é comum para análises de dados de mapa de colheita, onde pode não haver cultivares distintas ou apenas um talhão no conjunto de dados. Quando se tem fatores que podem ser considerados distintos como tratamentos, recomenda-se utilizar o método dos Desvios Studentizados para a estimativa dos índices e a identificação de dados discrepantes. Este também conhecido como “Teorema do limite central” (Equação 2), que utiliza a distribuição “t” de Student para inferência sobre a amostra

com desvio padrão ( $\sigma$ ) desconhecido, onde índices fora do intervalo  $\pm 1,96$  (5% de significância para teste bilateral de valores críticos de t) podem ser considerados discrepantes e podem ser eliminados do conjunto de dados, lembrando que valores que apresentam índice fora do intervalo apresentado no método dos Desvios Padronizados são de exclusão obrigatória e, os valores que apresentam índice entre  $\pm 3,09$  e  $\pm 1,96$ , sua exclusão não é obrigatória, ficando esta a critério do analista, de acordo com o que se deseja com o ajuste da curva de distribuição dos dados e enquadramento à normal. É válido ressaltar que, para a análise de desvios Studentizados ser considerada robusta, é necessário o levantamento de no mínimo trinta (30) valores para cada tratamento. Caso o número amostras por tratamento seja menor, torna-se necessário verificar na tabela de valores críticos de “t”, o intervalo crítico adequado ao número de amostras do tratamento (SNEDECOR & COCHRAN, 1980).

$$Z = \frac{x_i - \bar{x}_i}{\sigma}$$

Eq. 1

$$t = \frac{x_i - \bar{x}_i}{\sigma / \sqrt{n}}$$

Eq. 2

Legenda

$x_i$  – Amostra pontual  
 $\bar{x}_i$  – Média amostral ou do tratamento  
 $\sigma / \sqrt{n}$  – Erro padrão (tratamento)  
 $\sigma$  – Desvio padrão

O tratamento pelo método dos Desvios Padronizados preserva o comportamento e a distribuição do fenômeno, propiciando a remoção de valores não pertencentes a população e, também, dados discrepantes, possibilitando a aplicação de métodos de análise paramétrica nos dados de saída. É importante que estes métodos de análise e remoção de dados seja realizada em planilha eletrônica se aplicando cálculos referenciados para que após cada remoção de valor, os índices sejam recalculados, pois, com a remoção de dados as medidas de tendência central se alteram, bem como seus índices finais. Também é de suma importância iniciar sempre a remoção pelos índices mais extremos, pois, a cada remoção de valores discrepantes, os índices podem ser movidos para o centro ou para a extremidade da distribuição, podendo valores que se apresentem no limiar de corte serem movidos para dentro do intervalo crítico de corte escolhido e, o caso inverso também é verdadeiro (DELALIBERA et al., 2017). A aplicação da análise de desvios deve ser feita apenas nos dados de produtividade e, durante a remoção de dados, a linha inteira deve ser excluída, para que este não perca a referência com os dados de posição espacial e seus demais vínculos. Ao final da aplicação da análise de desvios é possível observar que os índices gerados são um escalamento dos dados entre  $\pm 3,09$  ou  $\pm 1,96$ , independente da grandeza ou da unidade de entrada da base de dados. A partir dos mapas gerados a partir destes índices de desvio é possível gerar álgebras de mapas com safras, culturas, cultivares e climas distintos e, a partir da média resultante desta álgebra, é possível obter um mapa único final de resultados médios de colheita, no qual os fatores citados terão seus efeitos diluídos. A simples separação em quartis dos valores finais do mapa resultante da álgebra, permite identificação e segregação de zonas que apresentam produtividade baixa, média e alta, ficando assim ressaltado os efeitos de fertilidade físico-química do solo, que são fatores de intervenção antrópica mais fácil e, espera-se que, havendo um histórico de 5 a 6 colheitas abrangendo as fontes de variação já citadas, pode permitir demonstrar os efeitos de solo sobre as culturas, possibilitando intervenções mais assertivas para as questões de manejo do solo.

### CONSIDERAÇÕES FINAIS:

A aplicação dos métodos de análise de desvios sobre os dados é interessante, pois, pode-se aplicar remoção de dados discrepantes com base em critérios estatísticos consagrados. Ainda

a utilização dos índices gerados para o levantamento de efeitos de solo sobre a produtividade das culturas, na forma de zoneamento das áreas torna-se mais assertivo na medida que se acrescenta mais safras no sistema, permitindo que a assistência técnica levante a campo os fatores factíveis de intervenção, visando o aumento da produtividade ou a adequação do nível tecnológico e investimento para cada zona de manejo gerada.

**AGRADECIMENTOS:** À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro ao projeto e ao IAPAR pela concessão do método analítico de análise de dados e álgebra de mapas.

#### **REFERÊNCIAS:**

BRASE, T. A. **Precision agriculture** / Terry A. Brase., 1st ed., Clifton Park, NY: Thomson/Delmar Learning, 2006.

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

MENEGATTI, L. A.A.; MOLIN, J.P. Remoção de erros em mapas de produtividade via filtragem de dados brutos. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.8, n.1, p.126-134, 2004.

MOLIN, J.P. Geração e Interpretação de Mapas de Produtividade para Agricultura de Precisão. Borém, A.; Giúdice, M. P.; Queiroz, D. M.; Mantovani, E. C.; Ferreira, L. R.; Valle, F. X. R.; Gomide, R. L. **Agricultura de Precisão**. Viçosa: Universidade Federal de Viçosa, p. 237-258, 2000.

MOLIN J.P., CREMONINI L.C.M., MENEGATTI L., GIMENEZ L. **Acurácia de um monitor de produtividade com sensor de fluxo volumétrico**. In: CONGRESSO BRASILEIRO DE ENGENHARIA AGRÍCOLA, 2000, Fortaleza. Anais/CD-Rom Fortaleza: SBEA, 2000.

MOLIN, J.P. **Geração interpretação de mapas de produtividade para agricultura de precisão**. 1997. Disponível em: <http://www.ler.esalq.usp.br/download/CLP%202000.01.PDF>

QUEIROZ, D.M.; DIAS, G.P.; MANTOVANI, E.C. Agricultura de precisão na produção de grãos. Borém, A.B.; Giúdice, M.P.; Queiroz, D.M.; Mantovani, E.C.; Ferreira, Valle, F.X.R.; Gomide, R.T. **Agricultura de Precisão**. Viçosa: UFV, p.1-42. 2000.

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