

Aplicação da análise descritiva e espacial em dados de capacidade de troca de cátions

SIBALDELLI, R. N. R.¹; OLIVEIRA, M. C. N. de²,
CAPECHE, C. L.³, DA SILVA, E. F.³, HISSA, H. R.³,
MACEDO, J. R.³ | ¹Universidade Estadual de Londrina (UEL);
²Embrapa Soja; ³Embrapa Solos | rubson@cnpso.embrapa.br

Introdução

Os diagnósticos exploratórios são extremamente importantes em dados de pesquisa, nas diferentes áreas do conhecimento, sempre que se deseja fazer inferência estatística. Na inferência, pode-se realizar a análise de variância (ANOVA), utilização de hipóteses estatísticas e a comprovação das mesmas, teste de comparação múltipla de médias sempre atendendo todos os pressupostos da ANOVA e, em caso contrário, utilizando transformação de dados para atender esses pré-requisitos (OLIVEIRA, 1998).

Os dados deste trabalho referem-se à capacidade de troca de cátions (CTC $\text{cmol}_c \text{dm}^{-3}$), com a realização de análise descritiva e exploratória espacial, por meio de um diagnóstico exploratório com *box-plot*, histogramas, estatística descritiva, gráficos de controle de qualidade e de probabilidade normal, além de testes de correlação espacial. Tal diagnóstico detectará a presença de *outliers* de forma que todos os métodos estatísticos sejam aplicados com fidedignidade e precisão.

Objetivou-se analisar e interpretar os dados para avaliar a magnitude da variabilidade e uma posterior aplicação do método geoestatístico nos dados da CTC.

Material e Métodos

Os dados utilizados nesta pesquisa são procedentes de um levantamento realizado ao Norte do estado do Rio de Janeiro, à margem esquerda do Rio Paraíba do Sul, no município de Campos dos Goytacazes (Capeche et al., 1997), com as coordenadas geográficas 21°44'47''S e 41°18'24'' O. A amostragem do solo foi realizada com trado holandês, em uma malha regular, distando 900 metros para cada coordenada geográfica: norte e leste. Esta malha foi espaçada de 50 metros, totalizando 273 amostras. A variável agrônômica estudada foi a CTC ($\text{cmol}_c \text{dm}^{-3}$), na camada de 0-20 cm de profundidade. Foi realizada análise exploratória descritiva e espacial, por meio das estatísticas como: os coeficientes de assimetria e curtose, o teste de normalidade de Shapiro & Wilk, teste de normalidade de Kolmogorov-Smirnov, o gráfico da probabilidade normal, o de *box-plot*, o histograma de frequências, a média, a mediana, a moda, gráfico de correlação entre as coordenadas e a CTC, e os gráficos de controle de qualidade. Para a execução deste trabalho, foi utilizado o programa estatístico StatSoft, Inc. (STATSOFT, 1995).

Resultados e Discussão

A CTC ($\text{cmol}_c \text{dm}^{-3}$), apresenta valores próximos para a média e a mediana na presença e ausência de *outliers*, bem como valores de assimetria e curtose próximos a zero, sendo os valores do conjunto de dados com 240 amostras sem *outliers* (Tabela1).

A presença de *outliers* nos dados remete a resultados indesejáveis e, além disso, não atende as pressuposições da ANOVA (Sibaldelli, 2005). Embora a análise de *box-plot* (Figura 1) indique que no conjunto de dados com 250 amostras não existem *outliers*, isto não foi confirmado pelo teste de normalidade (Tabela 2 e Figura 2). Optou-se então pelo uso dos gráficos de controle (Figura 3), por este possuir maior rigor que o *box-plot*, onde verificou-se a necessidade de se retirar mais amostras resultando em um conjunto de dados com 240 observações.

Tabela 1. Resultados das estatísticas descritivas para a variável CTC ($\text{cmol}_c \text{dm}^{-3}$), 0-20cm do solo com e sem *outliers*.

Estatísticas	Com <i>Outliers</i>		Sem <i>Outliers</i>	
Nº Amostras	273	254	250	240
Média	12,86	12,98	13,02	13,10
Mediana	12,57	13,06	13,07	13,08
Moda	Múltipla	Múltipla	Múltipla	Múltipla
Mínimo	5,55	7,88	8,24	8,72
Máximo	21,30	17,82	17,64	17,32
Variância	6,64	4,19	3,85	3,14
CA	-0,24	-0,14	-0,09	0,06
CC	0,81	0,10	0,03	-0,13
Desvio Padrão	2,58	2,05	1,96	1,77
CV (%)	20,06	15,79	15,05	13,51

CA = Coeficiente de Assimetria; CC = Coeficiente de Curtose; CV = Coeficiente de Variação.

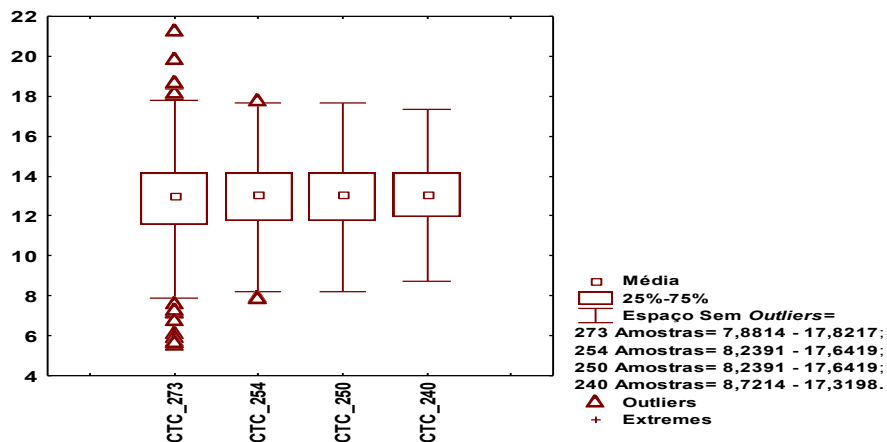


Figura 1. *Box-plot* para a variável CTC ($\text{cmol}_c \text{dm}^{-3}$), a 0-20 cm do solo, na presença de *outliers*, com 273 e 254 amostras, na ausência de *outliers*, com 250 e 240 amostras.

Tabela 2. Resultados dos testes de aderência para a variável CTC ($\text{cmol}_c \text{dm}^{-3}$), 0-20cm do solo com e sem *outliers*.

Estatísticas	Com <i>Outliers</i>						Sem <i>Outliers</i>	
	273 amostras		254 amostras		250 amostras		240 amostras	
	VT	Prob	VT	Prob	VT	Prob	VT	Prob
KS	d=0,0704	p<0,15000	d=0,0552	p>.20	d=0,0504	p>.20	d=0,0507	n.s.
SW	W=0,9782	p=0,00035	W=0,9868	p=0,0192	W=0,9883	p=0,040	W=0,9916	p=0,187

KS = Kolmogorov-Smirnov;

SW = Shapiro-Wilk;

VT = Valor do Teste;

Prob = Probabilidade

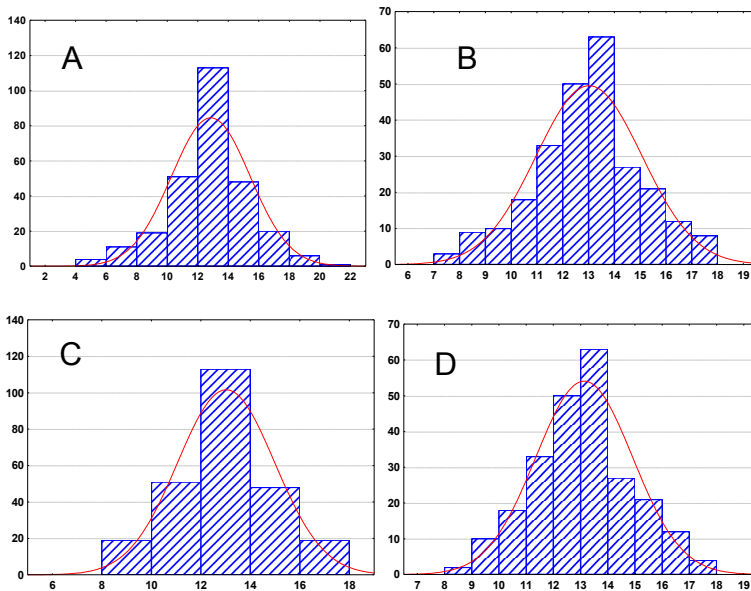


Figura 2. Distribuição de frequência sob a Curva Normal para a variável CTC ($\text{cmol}_c \text{dm}^{-3}$), a 0-20 cm. do solo, na presença de *outliers*, com 273 (a), 254 (b) e 250 amostras (c), sem a presença de *outliers* com 240 amostras (d).

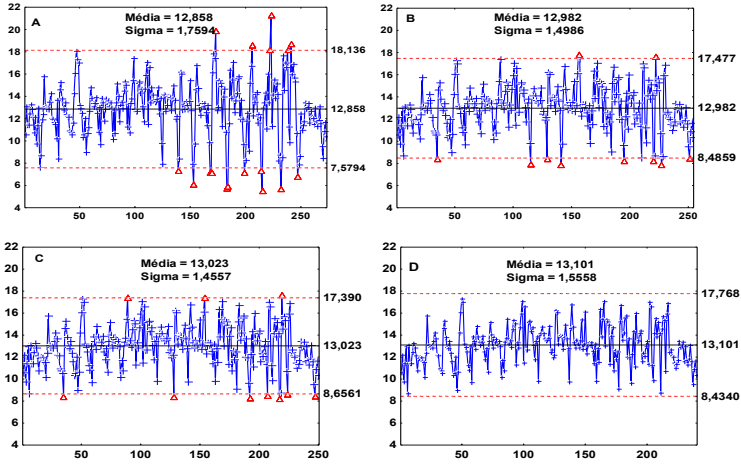


Figura 3. Gráfico de controle, com intervalo de segurança de 3σ , para a variável CTC ($\text{cmol}_c \text{dm}^{-3}$), a 0-20 cm do solo, com 273 (a), 254 (b) e 250 (c) amostras na presença de *outliers* e com 240 amostras (d), sem a presença de *outliers*.

Os valores da probabilidade normal (Figura 4) também mostram o ajuste a normal, inicialmente (Figura 4-A, B e C) observou-se, na presença de *outliers*, uma dispersão maior dos dados em relação à reta de probabilidade e no agrupamento com 240 amostras (Figura 4-D) um bom ajuste dos dados a reta.

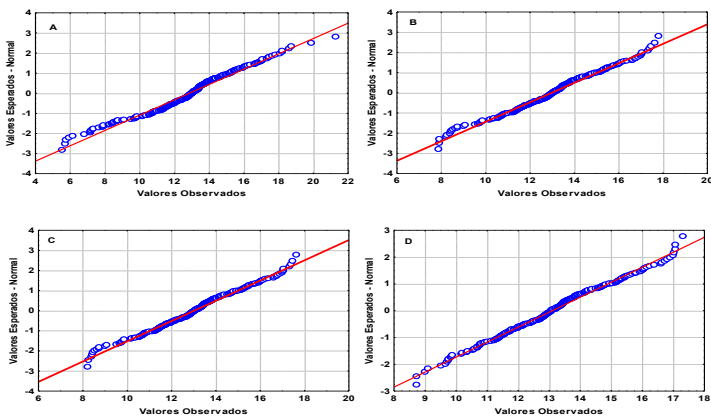


Figura 4: Gráfico da probabilidade normal para a variável CTC ($\text{cmol}_c \text{dm}^{-3}$), na camada de 0-20 cm do solo, com 273 (a), 254 (b) e 250 (c) amostras na presença de *outliers* e com 240 amostras (d), sem a presença de *outliers*.

Nas Figuras 5 e 6, foi obtida baixa tendência linear negativa em relação à direção norte e uma baixa tendência positiva em relação à direção leste, para todos os conjuntos de dados, tanto com *outliers* com 273, 254, 250 amostras, bem como sem *outliers*, no agrupamento com 240 amostras.

Os valores da correlação linear indicam que, após a retirada de *outliers* do conjunto de dados, a inclinação da reta aumenta, tanto para valores negativos em relação à direção norte, quanto para valores positivos em relação à direção leste. Com 273 amostras e na direção norte, o valor de $r = -0,1185$ enquanto que no conjunto com 240 amostras o valor de $r = -0,2972$. Já em relação a direção leste, no conjunto com 273 amostras o valor de $r = 0,0646$, enquanto para o conjunto com 240 amostras o valor de $r = 0,1087$ (Figuras 5 e 6).

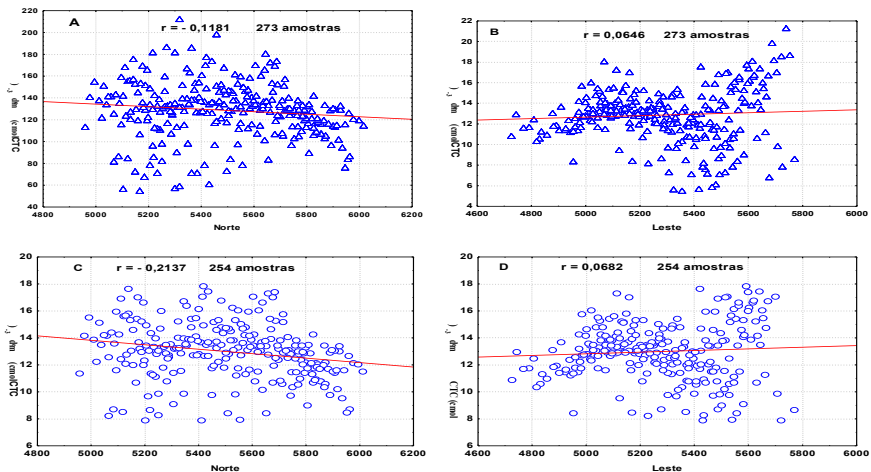


Figura 5: Gráfico de correlação para a posição geográfica Norte (a,c) e Leste (b,d), para a variável CTC (cmolc dm⁻³), na camada de 0-20 cm do solo, com 273 e 254 amostras.

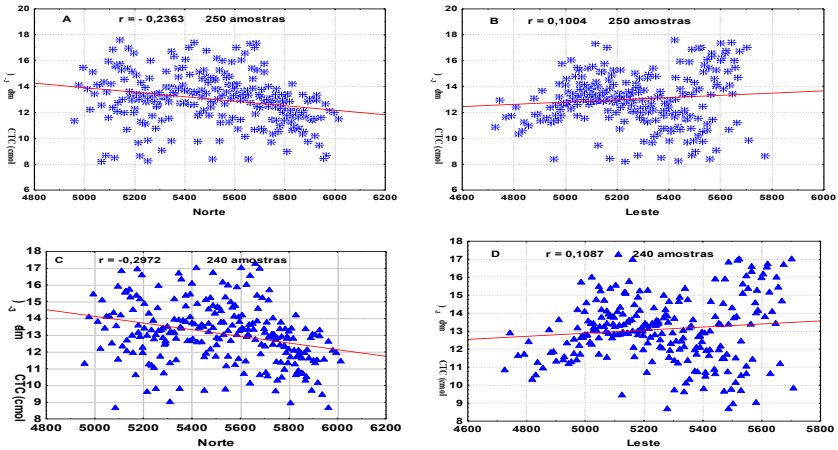


Figura 6: Gráfico de correlação para a posição geográfica Norte (a,c) e Leste (b,d), para a variável CTC (cmolc dm⁻³), na camada de 0-20 cm do solo, com 250 e 240 amostras.

Conclusões

Pôde-se concluir que: a) os dados são consistentes e robustos para o uso de ferramentas Geoestatísticas e b) o método da carta de controle de qualidade é eficiente para detectar *outliers* garantindo confiabilidade nos resultados.

Referências

CAPECHE, C. L.; MACEDO, J. R.; MANZATTO, H. R. H; SILVA, E. F. Caracterização pedológica da fazenda Angra – PESAGRO/RIO – Estação experimental de Campos (RJ). In: CONGRESSO BRASILEIRO DE CIÊNCIAS DO SOLO. 26., 1997, Rio de Janeiro. **Informação de solos na verbalização do conhecimento sobre o uso das terras: anais**. Rio de Janeiro: Sociedade Brasileira de Ciência do Solo, [1997]. 1 CD-ROM.

OLIVEIRA, M. C. N. de; CORRÊA-FERREIRA, B. S. Diagnóstico Exploratório na Análise de Dados Entomológicos: Efeito do Diflubenzurom no Consumo Alimentar de *Anticarsia gemmatilis* (Hubner). **Anais da Sociedade Entomológica Brasileira**. v. 27, n. 4. p. 627-637, dezembro 1998.

SIBALDELLI, R. N. R.; OLIVEIRA, M. C. N. de. A importância em detectar dados discrepantes em resultados de pesquisas com inoculantes da soja. In. REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 50.; SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 11., 2005, Londrina. [Anais...]. Londrina: UEL/Departamento de Estatística e Matemática Aplicada: IBS, 2005. 5p. 1 CD-ROM. Seção Trabalhos – Estatística Experimental.

STATSOFT, Inc. STATISTICA for windows (Computer program manual). Tulsa, Oklahoma, USA. 1995.