

APLICAÇÃO DA ANÁLISE DE COMPONENTES PRINCIPAIS NA SELEÇÃO DE CARACTERÍSTICAS DE SOLO

Marina Rodrigues Maestre¹; Simone Daniela Sartorio^{1,2}; Cláudio Roberto Marciano³; Fernando Carvalho Oliveira⁴, Sônia Maria De Stefano Piedade⁵, César Gonçalves de Lima⁶.

¹ Doutoranda do PPG em Estatística e Experimentação Agronômica – ESALQ/USP, Piracicaba, SP.

E-mail: ninamaestre@yahoo.com.br

² Bolsista CAPES.

³ UENF/CCTA, LSOL, Campos dos Goytacazes, RJ;

⁴ Biossola Agricultura & Ambiente Ltda, Piracicaba, SP;

⁵ Docente do PPG em Estatística e Experimentação Agronômica – ESALQ/USP, Piracicaba, SP.

⁶ Docente do Departamento de Ciências Básicas, FZEA/USP, Pirassununga, SP.

INTRODUÇÃO

Melhorar e manter a fertilidade dos solos é uma das prioridades dos sistemas agroecológicos. Juntamente com a preservação da agrobiodiversidade e o uso eficiente da água e da energia, o equilíbrio adequado de nutrientes e da vida no solo é um fator necessário para que os sistemas agrícolas sejam sustentáveis.

Os atributos físicos do solo desempenham importante papel, senão o principal, dentre as propriedades do solo. Todas as práticas de manejo causam impacto sobre os atributos do solo que se traduzem em influência sobre o desenvolvimento das culturas.

A textura do solo depende da proporção de areia, silte e/ou argila na sua composição. Dependendo da proporção destes materiais, a textura pode influenciar na taxa de infiltração da água, no armazenamento da mesma, na aeração, na facilidade de mecanização e na distribuição de determinados nutrientes (fertilidade do solo). Na natureza, a distribuição dos percentuais de argila, silte e areia variam bastante ao longo de um terreno. A forma em que esses diferentes tipos de grãos se distribuem é de extrema importância na disseminação da água no solo¹. Já o carbono orgânico dissolvido é uma das principais frações da matéria orgânica, que se constitui em fonte de energia nos ambientes aquáticos, podendo também influenciar os mais diversos processos biogeoquímicos que neles se desenvolvem (KRÜGER, 2003).

Em muitas situações, os pesquisadores tendem a avaliar o maior número de características do solo, gerando acréscimo considerável de trabalho. Quando o número de características torna-se elevado, é possível que muitas contribuam pouco para a discriminação dos indivíduos avaliados, representando, conseqüentemente, aumento no trabalho de caracterização, sem melhoria na

¹ <http://videoseducacionais.cptec.inpe.br/swf/solo/3_1/01_erosao.swf>

precisão, e tornando mais complexa a análise e interpretação dos dados (LIBERATO et al., 1999; BARBOSA et al., 2006; SILVA et al., 2005).

Dessa forma, objetivou-se com este trabalho utilizar a técnica de análise de componentes principais no intuito de reduzir a dimensionalidade dos dados que caracterizam o solo, eliminando as informações redundantes.

MATERIAL E MÉTODOS

Os dados utilizados neste trabalho são provenientes da tese de Marciano (1999) que foi conduzido na área da Estação Experimental do Instituto Agrônomo de Campinas, em Piracicaba - SP.

Avaliaram-se as seguintes características: porcentagem de Areia (AREIA); porcentagem de Silte (SILTE); porcentagem de Argila (ARGILA); Densidade da partícula, medida em Mg/m^3 (DENS); Porosidade (PORO) e; concentração de Carbono orgânico (CORG). Essas seis variáveis foram observadas no início do experimento para caracterizar o solo, portanto, sem a aplicação de tratamentos.

Primeiramente, analisou-se a matriz de correlações e por meio da relação N_c (razão do maior autovalor pelo menor) visando identificar a presença de multicolinearidade. As variáveis mais correlacionadas entre si podem ser responsáveis por esse efeito, desta forma, elas devem ser uma a uma retiradas da análise. Recalcula-se a relação N_c e a combinação de variáveis que apresentar menor N_c deverá servir de-base para a aplicação da técnica das componentes principais.

Dados de 36 amostras de solo foram utilizados na análise. Cada variável foi padronizada (média igual a zero e variância igual à unidade), pois cada uma delas apresenta unidades de medidas diferentes. Neste caso, a estrutura de dependência das variáveis foi dada pela matriz de correlação.

O método de análise de componentes principais consiste em transformar um conjunto de variáveis originais X_1, X_2, \dots, X_p em um novo conjunto de variáveis Y_1 (CP₁), Y_2 (CP₂), \dots, Y_p (CP_p) (Jonhson and Wichern, 2007). Essas p variáveis são não correlacionadas entre si e estão arranjadas em ordem decrescente de variâncias. A idéia principal nesse procedimento é que poucas dentre as primeiras componentes principais incorporam a maior variabilidade dos dados originais, contudo, pode-se racionalmente descartar as demais componentes, reduzindo o número de variáveis.

Existem vários critérios para se determinar quantas componentes devem ser excluídas da análise, dentre eles o *scree-plot*, que é um gráfico dos autovalores em função da ordem das componentes principais, representando graficamente a porcentagem de variância explicada por componente. Quando esta porcentagem se reduz e a curva passa a ser quase paralela ao eixo das

abscissas, as componentes correspondentes podem ser excluídas. Vale salientar que as componentes obtidas nesta análise constituem as novas variáveis respostas que serão utilizadas nas análises subsequentes do estudo.

As variáveis foram submetidas à análise de componentes principais (ACP) através do procedimento PRINCOMP disponível no programa estatístico SAS (2003).

RESULTADOS E DISCUSSÃO

A multicolinearidade foi resolvida retirando-se a variável ARGILA, reduzindo o valor N_c de 5670762 para 4,66. A análise prosseguiu considerando apenas as variáveis: porcentagem de Areia (AREIA); porcentagem de Silte (SILTE); Densidade da partícula (DENS); Porosidade (PORO) e; concentração de Carbono orgânico (CORG).

A matriz de correlações e os respectivos valores-p referentes ao teste de que a correlação entre o par de variáveis é nula são apresentados na Tabela 1. Observou-se que as variáveis mais correlacionadas foram: porcentagem de Areia (AREIA) com concentração de Carbono orgânico (CORG) (-0,5071), Densidade da partícula (DENS) com Porosidade (PORO) (0,4970) e porcentagem de Silte (SILTE) com Porosidade (PORO) (-0,3328). A correlação negativa indica que as variáveis são inversamente correlacionadas.

Tabela 1 - Correlações e seus respectivos valores de significância (valor-p) das cinco características do solo: porcentagem de Areia (AREIA); porcentagem de Silte (SILTE); porcentagem de Argila (ARGILA); Densidade da partícula (DENS); Porosidade (PORO) e; concentração de Carbono orgânico (CORG).

Variáveis	Variáveis			
	AREIA	SILTE	DENS	PORO
SILTE	0,0062 (0,9712)			
DENS	0,0156 (0,9281)	-0,1207 (0,4830)		
PORO	-0,2401 (0,1584)	-0,3328 (0,0473)	0,4970 (0,0020)	
CORG	-0,5071 (0,0016)	-0,0792 (0,6462)	-0,0368 (0,8313)	0,1565 (0,3620)

Os resultados obtidos das componentes principais, seus respectivos autovalores e as porcentagens da variância explicada por essas componentes estão apresentados na Tabela 2. Considerando-se que a importância relativa das componentes principais decresce da primeira para a

última, tem-se que as últimas componentes são responsáveis pela explicação de uma fração mínima da variância total disponível.

Tabela 2 Estimativas dos autovalores (λ_i), das contribuições percentuais ($\% \lambda_i$) e os autovetores associados aos autovalores obtidos da matriz de correlação, envolvendo as cinco características do solo.

CP*	λ_i	$\% \lambda_i$	Autovetores				
			AREIA	SILTE	DENS	PORO	CORG
1	1,8027	0,3605	-0,4325	-0,5417	0,3991	0,6024	0,4073
2	1,4078	0,6421	0,5417	-0,2621	0,4953	0,2844	-0,5582
3	0,8939	0,8209	-0,1481	0,8498	0,4926	0,1020	-0,0537
4	0,5088	0,9226	0,5554	-0,0197	0,3402	-0,3071	0,6936
5	0,3868	1,0000	0,4348	0,2897	-0,4868	0,6719	0,1963

* Componentes principais.

Analisando o *scree-plot* (Figura 1) observa-se que apenas as três primeiras componentes são suficientes para explicar a maior parte (82,09%) da variação total dos dados (Tabela 2), sendo 36,05% da primeira componente, 28,16% da segunda e 17,88% da terceira. Estas três variáveis podem substituir as variáveis originais em análises subsequentes (se for de interesse do pesquisador).

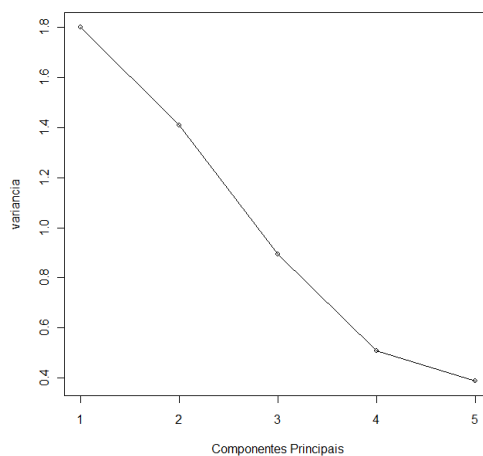


Figura 1. *Scree-plot* das 5 componentes principais

Analisando-se os coeficientes dos autovetores associados às componentes principais (Tabela 2) percebe-se que a primeira componente está mais relacionada às características SILTE e PORO; a segunda componente, com AREIA e CORG e a terceira, apenas com o SILTE.

Como a porosidade se refere à proporção de ar e água no solo, ela depende muito do tipo de argila e da estrutura do solo, que por sua vez influencia na taxa de infiltração da água. Logo, fica evidente que as componentes obtidas podem caracterizar de forma satisfatória o solo em estudo sem perda demasiada de informação.

CONCLUSÕES

Com base nos resultados, concluiu-se que 50% das variáveis analisadas foram consideradas redundantes e que a dimensão dos dados pode ser reduzida a três por meio do método das componentes principais.

REFERÊNCIAS

BARBOSA, L.; LOPES, P. L.; REGAZZI, A.J.; GUIMARÃES, S.E.F.; TORRES, R.A. Avaliação de características de qualidade da carne de suínos por meio de componentes principais. **Revista Brasileira de Zootecnia**, p. v.35, n.4, p.1639-1645, 2006.

JOHNSON, R.A.; WICHERN, D.W. Applied multivariate statistical analysis. 6th ed Upper Saddle River, N.J: Pearson Prentice Hall, 2007, 773 p.

KRÜGER, G.C.T.; CARVALHO, C.E.V.; FERREIRA, A.G.; GONÇALVES, G.M.; TRUCCOLO, E.C.; SCHETTINI, C.A.F. Dinâmica de carbono orgânico dissolvido no Estuário do Rio Paraíba do Sul, RJ, sob diferentes condições de maré e descarga fluvial. **Atlântica**, Rio Grande, 25(1): 27-33, 2003.

LIBERATO, J.R.; VALE, F.X.R.; CRUZ, C.D. Técnicas estatísticas de análise multivariada e a necessidade de o fitopatologista conhecê-las. **Fitopatologia Brasileira**, v.24, p.5-8, 1999.

MARCIANO, C.R. Incorporação de resíduos urbanos e as propriedades físico-hídricas de um latossolo vermelho amarelo. 1999. 93p. Tese (Doutorado) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 1999.

SILVA, I.C.; MESSAGE, D.; CRUZ, C.D.; SILVA, M.V. G.G. Aplicação de Análises Multivariadas para Determinação da Casta de Abelhas *Apis mellifera* L. (Africanizadas), Obtidas em Laboratório. **Revista Brasileira de Zootecnia**, p. v.34, n.2, p.635-640, 2005.