

APLICAÇÃO DA ESTATÍSTICA MULTIVARIADA EM RESULTADOS DE ANÁLISE DE SOLOS NA FAZENDA GUAMIRIM, NA CIDADE DE ÁGUA DOCE, SC.

*Jorge Roberto Grobe*¹

Resumo: Este trabalho tem o objetivo de aplicar o uso da estatística multivariada no resultado das análises de solos. O estudo foi realizado na propriedade do Sr. José Antônio Bueno, localizado no município de Água Doce, SC, na região denominada “Campos de Palmas”. As variáveis que foram analisadas foram pH (potencial de hidrogênio), matéria orgânica (MO), alumínio (Al), acidez total (H+Al), cálcio (Ca), magnésio (Mg), potássio (K), fósforo (P), saturação por bases (V), amônia (NH₄) e nitrato (NO₃). Foram aplicadas técnicas de estatística multivariada análise fatorial e análise de agrupamento e para determinação dos resultados foram utilizadas funções no *software* STATISTIC 5.1.

Palavras-chave: Estatística multivariada, componentes principais, análise fatorial, fatorial, análise de agrupamento, análises de solos.

Abstract: This paper aims at applying the use of the multivariate statistics in the result of the soil analyses. The study was carried out in the property of Mr. José Antônio Bueno, located in the town of Água Doce, SC, in the region named “Campos de Palmas”. The analyzed variables were pH (potential of hydrogen), organic matter (MO), aluminum (Al), total acidity (H + Al), calcium (Ca), magnesium (Mg), potassium (K), phosphorus (P), saturation by bases (V), ammonia (NH₄) and nitrate (NO₃). Multivariate statistics techniques were applied as Factorial Analysis and Analysis of Group and functions in the software STATISTIC 5.1 were utilized for determination of the results.

Keywords: Multivariate statistics, main components, factorial analysis, factorial, analysis of group, analyses of soils.

¹ Graduado em Ciências, habilitação em Matemática. Especialista em Ciências, em Fundamentos de Matemática, em Ensino de Matemática Método da Modelagem, Especialista em Informática na Educação. Professor da Coordenação de Informática da UTFPR, Campus Pato Branco, Mestrando em Métodos Numéricos em Engenharia pela UFPR. jrgrobe@pb.cefetpr.br. grobe@ubbi.com.br.

Segundo Neto (2004), na vida cotidiana aparecem vários fatores que podem estabelecer várias decisões. Às vezes quando não se identifica o fator toma-se uma decisão a partir de uma intuição. Um grande número de variáveis são identificadas em todos os acontecimentos sejam eles culturais ou naturais, através de diversas ciências, como ciências humanas ou ciências naturais em que o homem pode conhecer a realidade e interpretar os acontecimentos. Para traduzir conhecimentos de um fenômeno analisado, há uma necessidade de controlar, manipular, medir as variáveis que são consideradas relevantes e traduzir essas informações. Algumas informações obtidas de conhecimentos podem existir uma dificuldade, pois a ciência não conhece a realidade, ela representa através de modelos e teorias de diversos estudos. Um ramo do conhecimento que aspira dificuldades da universalidade das explicações científicas e que implica na padronização dos dados é a avaliação estatística das informações. Outra maneira de fazer ciência reduzindo-se poucas variáveis chama-se estatística univariada. Para estatística univariada existem vantagens e desvantagens quando se tem um estudo frente a várias variáveis e a uma única variável. No caso restrito de variáveis independentes é possível com segurança, interpretar fenômenos usando medidas de tendência central como média, moda, etc. e dispersão dos dados. Na estatística multivariada o fenômeno depende de muitas variáveis, com isso não basta conhecer as variáveis isoladas, mas conhecê-las na sua totalidade, pois uma depende da outra e as informações são fornecidas pelo conjunto e não individualmente. Então para a estatística univariada ela analisa cada variável isoladamente e não o conjunto.

Para estatística a pesquisa tem significado se conhecermos o verdadeiro problema sobre o universo que elegemos, quanto às variáveis e às metodologias de análises.

Bueno et al (2004), em seu trabalho sobre pastagem nativa melhorada sob distintas intensidades de pastejo na região “CAMPOS DE PALMAS”, aplicou nas variáveis derivadas de animais a estatística univariada o que confirmou que o manejo de pastagem influenciou a produção individual de cada animal. Neste experimento agrônomo foram também realizadas análises de solos, variáveis medidas na pastagem. No resultado de análises de solos pretende-se aplicar técnicas de estatística multivariada para analisar o comportamento do conjunto dessas variáveis. Essas técnicas de estatística multivariada serão aplicadas neste experimento agrônomo como: análise fatorial e análise de agrupamento.

O objetivo geral deste trabalho é explicar o uso da aplicação da análise estatística multivariada em um experimento agrônomo. Para se alcançar estes objetivos é necessário:

- analisar quais as variáveis que explicam maior parte da variabilidade total dos dados;

- encontrar fatores interpretáveis;
- identificar e classificar uma amostra de indivíduos ou objetos em um pequeno número de grupos mutuamente exclusivos quanto à similaridade e dissimilaridade.

O período de execução deste experimento agrônômico foi efetuado entre julho de 2003 e abril de 2004. O experimento foi realizado na propriedade do Sr. José Antonio Bueno, localizada no município de Água Doce-SC, região denominada “Campos de Palmas” conforme figura 1. A área total destinada ao experimento foi aproximadamente 6,93 hectares dividida em 6 poteiros.

Para fins deste estudo tem-se uma matriz de resultados de análises de solos que caracteriza a área conforme tabela 1. As análises de rotinas de solos foram coletadas em três profundidades: 0-5 cm, 5-10cm e 10-20cm e os atributos químicos analisados foram: potencial de hidrogênio, matéria orgânica, alumínio, hidrogênio mais alumínio (acidez total), cálcio, fósforo, potássio e saturação por bases, nitrato e amônia.

Nas Análises de Agrupamento e Fatorial é possível também comparar os grupos homogêneos quanto à correlação das variáveis, distância média e os pesos atribuídos nos fatores. Os dados amostrais da população 1 foram padronizados para as análises fatorial e agrupamento.

FIGURA 1 – LOCAÇÃO² DO EXPERIMENTO AGRÔNÔMICO



Fonte:

<http://www.brasounds.hpg.ig.com.br/foto/sul.jpg>

² As cidades de Palmas e Água Doce foram adaptadas no mapa da figura original a fim de facilitar a observação.

TABELA 1 – MATRIZ DE DADOS DO RESULTADO DE ANÁLISES DE SOLOS DA POPULAÇÃO 1
 Projeto Guamirim / Água Doce - SC. Análise de Solo 30/08
 CARACTERIZAÇÃO DA ÁREA

Bloco	Identificação	Prof.	pH	MO	Al ³⁺	H+Al	Ca	Mg	K	P	V	NH4	NO3
		C m	CaCl ₂	g dm ⁻³	cmol _c ⁽⁺⁾ dm ⁻³					mg dm ⁻³		mg/kg	mg/kg
1	Trincheira	0-5	4,20	53,61	2,20	14,44	2,70	2,20	0,23	16,94	0,26	4,20	57,87
1	Trincheira	5-10	4,00	53,61	3,80	15,77	0,63	1,14	0,10	0,93	0,11	4,20	36,17
1	Trincheira	10-20	4,00	0,00	4,05	13,06	0,40	0,87	0,08	0,93	0,09	4,20	20,67
1	Composta	0-5	4,2	53,6	3,1	15,8	1,5	1,4	0,1	10,6	0,16	3,6	20,7
1	Composta	5-10	4,0	53,6	4,1	17,2	0,6	1,3	0,1	0,9	0,11	1,8	13,4
1	Composta	10-20	4,0	46,9	4,7	15,2	0,5	0,7	0,1	0,9	0,08	2,4	20,7
2	Trincheira	0-5	4,00	56,61	4,24	15,77	0,66	0,94	0,10	1,25	0,10	1,80	33,07
2	Trincheira	5-10	4,00	44,23	4,09	13,06	0,35	0,55	0,05	0,61	0,07	2,10	32,03
2	Trincheira	10-20	4,00	40,21	3,75	13,06	0,33	0,67	0,05	0,61	0,07	2,70	35,13
2	Composta	0-5	4,00	53,61	3,11	14,44	1,18	1,72	0,10	6,13	0,17	3,00	24,80
2	Composta	5-10	3,90	53,61	4,13	15,77	0,53	1,17	0,10	1,25	0,10	2,70	20,67
2	Composta	10-20	3,90	53,61	5,10	18,82	0,40	0,50	0,08	0,61	0,05	1,80	21,70
3	Trincheira	0-5	4,00	60,31	4,58	18,82	0,90	1,17	0,15	1,57	0,11	1,80	21,70
3	Trincheira	5-10	3,90	53,61	4,53	15,77	0,32	0,58	0,08	0,93	0,06	2,40	10,33
3	Trincheira	10-20	3,90	46,91	4,25	14,08	0,20	0,40	0,05	0,00	0,04	2,10	16,53
3	Composta	0-5	4,20	60,31	2,25	14,44	2,30	2,00	0,18	15,36	0,24	2,10	26,87
3	Composta	5-10	4,00	60,31	5,07	17,23	0,75	0,85	0,13	1,57	0,09	2,10	31,00
3	Composta	10-20	3,90	60,31	5,15	18,82	0,40	0,50	0,10	0,61	0,05	1,80	25,83

Fonte: Laboratório de Solos da Unidade do Cefet Pato Branco

No **tabela 2** apresentam-se os fatores da população 1 através análise fatorial utilizando o método da análise de componentes principais (ACP):

TABELA 2 – FATORES PELO MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 1

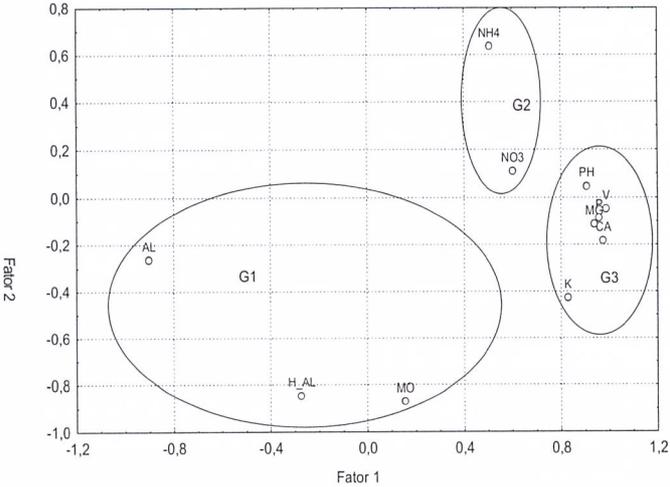
Variáveis	Método das componentes		Comunalidades
	Y_1	Y_2	
pH	0,9066	-0,0470	0,8242
MO	0,1552	0,8694	0,8850
Al	-0,9033	0,2627	0,7897
H+Al	-0,2755	0,8449	0,9817
Ca	0,9739	0,1821	0,8908
Mg	0,9372	0,1118	0,8697
K	0,8290	0,4272	0,9232
P	0,9570	0,0856	0,9756
V	0,9865	0,0480	0,6598
NH ₄	0,5064	-0,6351	0,3742
NO ₃	0,6017	-0,1101	
autovalores	6,7596	2,1940	
Proporção Acumulada da Variância	61,45%	81,40%	

Conforme os dados obtidos na tabela 2, o modelo ajustado levou em consideração dois fatores no Método das Componentes Principais Y_1 e Y_2 . No conjunto do modelo ajustado, explicaram 81,40% das variâncias das variáveis com autovalores maiores que 1. O percentual das comunalidades de cada variável que é explicado pelos fatores em conjunto mostra que as variáveis que melhor foram explicadas pelos fatores (apresentaram maior comunalidade) foi H+Al que teve 98,17% e P(fósforo) 97,56% de suas variações captadas pelos fatores, e a de menor valor percentual foi NO₃ (nitrato), apresentando 37,42% de sua variação explicada pelos fatores estudados. As demais variáveis apresentaram comunalidades entre 65% e 92%, o que mostra que suas variações foram, de modo geral, captadas de forma satisfatória pelos fatores ajustados.

Cada variável dentro do seu fator correspondente possui uma alta correlação entre elas e entre os fatores as variáveis possuem uma correlação mais fraca. Esta correlação varia de -1 até 1, sendo que perto de zero não existe correlação, portanto quanto mais perto de 1 ou -1 melhor é a correlação.

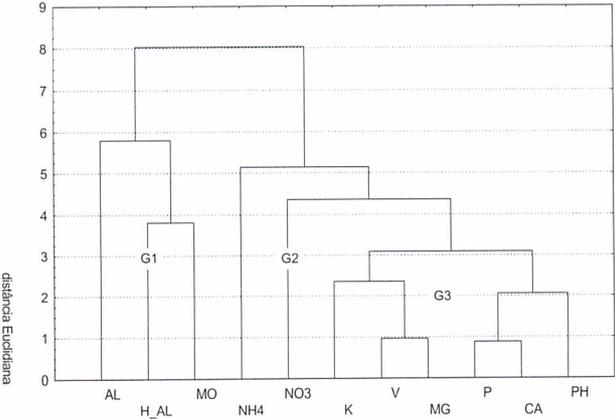
Os fatores calculados pelo método das componentes principais dispostas na tabela 2 da população 1 foram plotados no gráfico da figura 2, a seguir:

FIGURA 2 – ORDENAÇÃO DAS VARIÁVEIS PELO MÉTODO DAS COMPONENTES PRINCIPAIS DA POPULAÇÃO 1



Os fatores obtidos pelo método das componentes principais dos atributos químicos, são agrupados pela análise fatorial da população 1 por meio da figura 17.

FIGURA 3 – DENDROGRAMA DA POPULAÇÃO 1



As observações conjuntas da tabela 2 com as figuras 2 e 3 permitem que a análise de componentes principais confirme a análise de agrupamentos, sendo que as variáveis dos resultados das análises de solos foram sensíveis para discriminá-los.

Foram reunidos a partir de 3 grupos distintos constituídos pela semelhança (pode ser a correlação) e pela distância euclidiana dos atributos químicos. O Grupo G1 reuniu as variáveis Al, H+Al e MO com uma distância média 6 e o grupo G2 reuniu as variáveis NH₄ e NO₃ com uma distância média de 5 e o último grupo G3 tem uma distância média de 3 e estão alocadas neste grupo as variáveis K, V, Mg, P,Ca, pH

Pode-se concluir que a análise multivariada mostrou-se útil e adequado para identificação dos atributos químicos dos resultados de análises de solos. E que 61,45% da variabilidade dos dados estão grandes partes dos atributos químicos. Nas aplicações das duas técnicas análise fatorial e análise de agrupamento foi possível fazer uma comparação entre a correlação entre os atributos químicos e os grupos quanto à distância.

REFERÊNCIAS BIBLIOGRÁFICAS

BUENO E.A .C. et al. Pastagem Nativa Melhorada Sob Distintas Intensidades de Pastejo. In: IX Seminário de Iniciação Científica e Tecnológica, 2004. Curitiba PR *Anais...* Pato Branco.CEFET- PR 2004.

JOHNSON, R.A; WICHERN, D. W. *Applied multivariate statistical analysis*. 4 ed. New Jersey: Prentice-Hall, 1998.

NETO, J.M. MOITA. *Estatística Multivariada Uma Visão Didática Metodológica*. Disponível em <<http://www.criticanarede.com>> Acessado em: 21/09/2005.