

# ANÁLISE FATORIAL MULTIVARIADA APLICADA NA CARACTERIZAÇÃO DE CONTAMINANTES DE UM DEPÓSITO DE MINÉRIO DE FERRO

Daniel Geraldo da Cruz<sup>1</sup>  
Francielle Câmara Nogueira<sup>1</sup>  
Daniel Quinaud Rossi<sup>2</sup>  
Tiago Martins Pereira<sup>3</sup>  
Rodrigo Fina Ferreira<sup>4</sup>

## Resumo

A análise fatorial é uma técnica da estatística multivariada que tem o objetivo de diminuir o número de variáveis iniciais de uma análise, minimizando a perda de informação. A técnica cria um grupo reduzido de fatores comuns não observados a partir da interpretação das inter-relações existentes dentro do conjunto de variáveis observadas. Neste artigo, a análise fatorial foi aplicada utilizando o programa R com o objetivo de encontrar fatores que representem a relação entre os contaminantes de um minério de ferro. Para tanto, foram analisadas as variáveis Fe, SiO<sub>2</sub>, Al<sub>2</sub>O<sub>3</sub>, P, Mn, TiO<sub>2</sub>, CaO, MgO e PPC (perda por calcinação). Os fatores foram extraídos aplicando o método dos componentes principais, utilizando a matriz de correlação amostral. Observou-se que 83,5% do total da variância pôde ser explicada por 3 fatores. Utilizando a rotação ortogonal do tipo Varimax, observou-se que as variáveis PPC, CaO e MgO possuem cargas maiores no fator 1 (presença de carbonato), que as variáveis Al e TiO<sub>2</sub> têm cargas mais altas no fator 2 (intemperismo químico e/ou intrusões ígneas), e que o fator 3 está relacionado com a variável manganês.

**Palavras-chave:** Estatística multivariada; Análise fatorial; Minério de ferro.

## MULTIVARIATE FACTOR ANALYSIS APPLIED IN THE CHARACTERIZATION OF CONTAMINANTS OF AN IRON ORE DEPOSIT

### Abstract

Factorial analysis is the technique of multivariate statistics which aims at the reduction of the number of initial variables of analysis, in order to decrease information loss. The technique creates a small group of common factors not observed from the interpretation of the existing interrelations within the set of observed variables. In this article, the factor analysis was applied using the R program with the objective of finding factors that represent the relation between the contaminants of an iron ore. To this end, the variables Fe, SiO<sub>2</sub>, Al<sub>2</sub>O<sub>3</sub>, P, Mn, TiO<sub>2</sub>, CaO, MgO and PPC (loss by calcination) were analyzed. The factors were extracted using the Main Components method, using the sample correlation matrix. It was observed that 83.5% of the total variance could be explained by 3 factors. Using the orthogonal rotation of the Varimax type, it was observed that the PPC, CaO and MgO variables have higher loads in factor 1 (carbonate presence), Al and TiO<sub>2</sub> variables have higher loads in factor 2 (chemical weathering and/or igneous intrusions) and factor 3 is related to the Manganese variable.

**Keywords:** Multivariate statistics; Factor analysis; Iron ore.

## I INTRODUÇÃO

O tratamento da grande quantidade de dados obtidos a partir de prospecção mineral e levantamentos geoquímicos é enormemente facilitado pelo emprego

de técnicas e ferramentas da estatística multivariada. Estas técnicas propiciam o tratamento de diversas variáveis simultaneamente, permitindo a obtenção de informações

<sup>1</sup>Departamento de Engenharia de Minas, Escola de Minas, Universidade Federal de Ouro Preto – UFOP, Ouro Preto, MG, Brasil.  
E-mail: danielgc.ufop@gmail.com

<sup>2</sup>Gerência de Recursos Minerais Ferrosos, VALE S.A., Belo Horizonte, MG, Brasil.

<sup>3</sup>Departamento de Estatística, Instituto de Ciências Exatas e Biológicas, Universidade Federal de Ouro Preto – UFOP, Ouro Preto, MG, Brasil.

<sup>4</sup>Gerência de Desenvolvimento de Processo de Tratamento de Minério, Vale S/A, Mariana, MG, Brasil.



significativas, de maneira hierarquizada, apresentando comportamentos e tendências destacáveis e a relação entre as variáveis estudadas [1,2].

Ferreira [3] salienta que os métodos multivariados são promissores, uma vez que os estudos geológicos envolvem conjuntos complexos de fatores e variáveis, e uma grande quantidade de dados, dificultando a análise isolada dos fatores. Dos métodos multivariados usuais, merece especial atenção a análise fatorial, uma técnica que busca descrever a estrutura de covariâncias ou de correlações entre as variáveis originais, em termos de um menor número de variáveis (não observáveis), denominadas fatores [4,5].

Contessotto [6] destacou a necessidade de se conhecer o depósito e a sua natureza mineralógica. O autor apresentou uma linha de raciocínio na qual ressalta a grande importância da aplicação das técnicas de estatística multivariada a dados obtidos a partir de estudos de caracterização de minas, com o intuito de prever sua viabilidade de exploração e comportamento em etapas de concentração.

Celino e Rangel [1] aplicaram análise de agrupamento e análise de componentes principais a um conjunto de dados geoquímicos (análises químicas de amostras de sedimentos) obtendo evidências das relações entre os fatores com elevadas cargas. Os autores realizaram testes de correlação dos elementos traço, concluindo que os pares de elementos com maior correlação se justificavam por processos de adsorção em argilas (Cu/Pb, Pb/Zn) e pela presença de minerais ferromagnesianos (Ca/Mg, Mg/Ti).

De acordo com Souza [7], estudos sobre a geoquímica mineral de formações ferríferas bandadas do Quadrilátero Ferrífero, principalmente de elementos-traços, ainda são escassos. Neste contexto, desenvolveu-se o presente trabalho, que tem como objetivo encontrar fatores que representem o nível de dependência entre os contaminantes de um depósito de minério de ferro, utilizando a análise fatorial, aplicada com o uso da linguagem de programação R [8].

## 2 MATERIAIS E MÉTODOS

O banco de dados utilizados neste artigo é composto de 2285 amostras provenientes de uma mineração de ferro. O estudo considerou nove variáveis referentes aos percentuais de Fe, SiO<sub>2</sub>, Al<sub>2</sub>O<sub>3</sub>, P, Mn, TiO<sub>2</sub>, CaO, MgO e a perda por calcinação (PPC).

Primeiramente, foi realizada uma análise estatística descritiva para conhecer o perfil das variáveis analisadas. Em seguida, a análise multivariada foi aplicada para entender o inter-relacionamento destes dados. Para tanto, realizou-se uma análise fatorial utilizando a linguagem de programação R. Os *scripts* da análise fatorial foram elaborados utilizando os pacotes: *library(xlsx)*, *library(MVN)*, *library(goft)*, *library(psych)*, *library(Rcmdr)*, *library(MVar.pt)*, *library(FactoMineR)*.

A análise fatorial (AF) foi realizada de acordo com as seguintes etapas: construção da matriz de correlação, verificação da adequabilidade da base de dados através dos

testes KMO/MSA, determinação da matriz de cargas fatoriais utilizando o método dos componentes principais, extração do número de fatores a partir do gráfico *scree plot* (gráfico de autovalores *versus* número de fatores); rotação ortogonal dos fatores utilizando o método varimax, e interpretação dos fatores a partir da análise das cargas fatoriais das variáveis.

## 3 DISCUSSÃO DOS RESULTADOS

### 3.1 Análise Descritiva

Para realizar AF utilizando a matriz de covariâncias, é necessário que as variáveis possuam variâncias de magnitudes similares. Desta forma, a primeira análise do banco de dados referiu-se a uma análise estatística descritiva feita com objetivo de visualizar como os teores das amostras variavam. Analisando a Figura 1, pôde-se observar que as variáveis ferro (Fe) e sílica (SiO<sub>2</sub>) foram as que apresentaram teores com a maior variabilidade.

A análise fatorial é muito sensível a diferenças entre as variâncias. Portanto, foi necessário realizar a padronização das variáveis (média igual a zero e desvio padrão igual a 1). O resultado desta padronização está apresentado no gráfico *boxplot* da Figura 2.

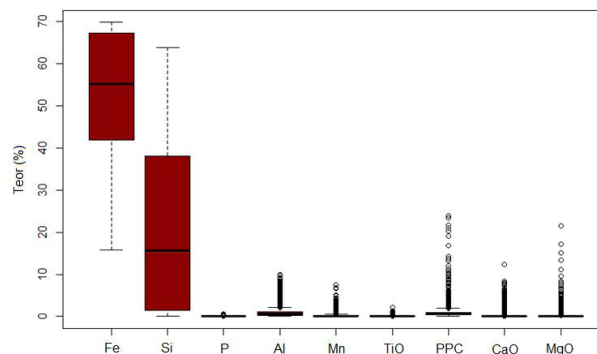


Figura 1. Gráfico boxplot das nove variáveis do banco de dados.

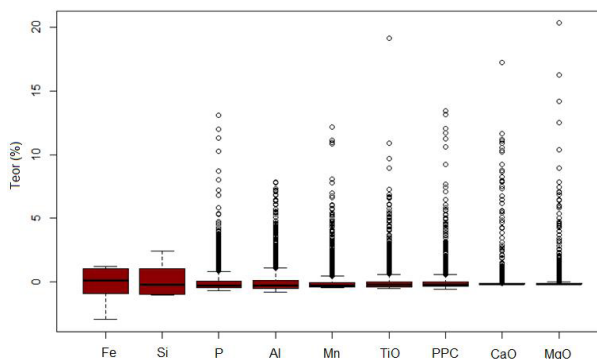


Figura 2. Gráfico boxplot padronizado das nove variáveis do banco de dados.

**Tabela 1.** Matriz de correlação dos dados

	Fe	Si	P	Al	Mn	TiO	PPC	CaO	MgO
Fe	1	-0,976554	-0,019387	-0,055690	0,035115	-0,051853	-0,088009	-0,108110	-0,126452
Si	-0,976554	1	-0,092753	-0,070394	-0,119066	-0,567898	-0,115463	-0,040272	-0,034761
P	-0,019387	-0,092753	1	0,424776	0,231858	0,431074	0,475830	0,238076	0,249576
Al	-0,055690	-0,070394	0,424776	1	0,305797	0,842073	0,418714	-0,003520	0,061330
Mn	0,035115	-0,119066	0,231858	0,305797	1	0,273868	0,232763	-0,010491	0,001715
TiO	-0,051853	-0,567898	0,431074	0,842073	0,273868	1	0,362725	-0,003618	0,034359
PPC	-0,088009	-0,115463	0,475830	0,418714	0,232763	0,362725	1	0,717460	0,751260
CaO	-0,108110	-0,040272	0,238076	-0,003520	-0,010491	-0,003618	0,717460	1	0,769982
MgO	-0,126452	-0,034761	0,249576	0,061330	0,001715	0,034359	0,751260	0,769982	1

### 3.2 Matriz de Correlação

Para realizar a AF, é necessário verificar o grau de relacionamento entre as variáveis. Para tanto, calculou-se a matriz de correlações (Tabela 1), o que possibilitou identificar subconjuntos de variáveis muito correlacionadas entre si, mas pouco relacionados a variáveis de outros subconjuntos. Coeficientes acima de 0,30 indicam correlação significativa entre as variáveis.

Estas correlações também foram apresentadas através do gráfico de elipses da Figura 3, onde a cor azul representa correlação positiva e os tons mais fortes representam as correlações mais altas. Observou-se que existe uma correlação forte e positiva entre os pares das variáveis Al-TiO<sub>2</sub>, PPC-CaO, CaO-MgO e MgO-PPC. Por outro lado, as variáveis Fe e Si apresentam forte correlação negativa, no entanto não se correlacionam com as demais variáveis.

### 3.3 KMO e MSA

Para realizar a AF é necessário analisar a adequabilidade da base de dados. Para esta verificação foi utilizado o teste do KMO (*Kaiser-Meyer-Olkin*), que verifica o ajuste dos dados utilizando todas as variáveis simultaneamente. Para mensurar a adequabilidade amostral de cada variável, foi utilizado o teste de MSA (*Measure of Sampling Adequacy*). A Tabela 2 apresenta uma síntese da adequação amostral. As variáveis podem ser utilizadas para realizar a AF caso os valores obtidos estejam entre 0,5 e 1,0 [9].

Para o banco de dados utilizado neste estudo foi encontrado o valor de KMO igual a 0,29. Com exceção das variáveis P (fósforo) e TiO<sub>2</sub> (óxido de titânio), todos os valores de MSA foram inferiores a 0,50 (Tabela 3), indicando uma correlação parcial muito baixa. Comparando os valores obtidos com os valores de referência (Tabela 2), observou-se que este conjunto de dados apresentou um grau de adequabilidade insuficiente para uma AF satisfatória.

Desta forma, as variáveis Fe e Si foram retiradas do banco de dados, pelo fato de não sofrerem variações em função das demais variáveis, e novamente aplicou-se o teste de KMO e MSA. Com a exclusão destas variáveis, o valor do KMO subiu de 0,29 para 0,72 e todos os valores de MSA foram superiores a 0,6 conforme apresentado na

**Tabela 2.** Adequação amostral segundo medidas de KMO e MSA

KMO/MSA	ADEQUABILIDADE
0,8-1,00	Excelente
0,70-0,80	Ótimo
0,60-0,70	Bom
0,50-0,60	Regular
0,00-0,50	Insuficiente

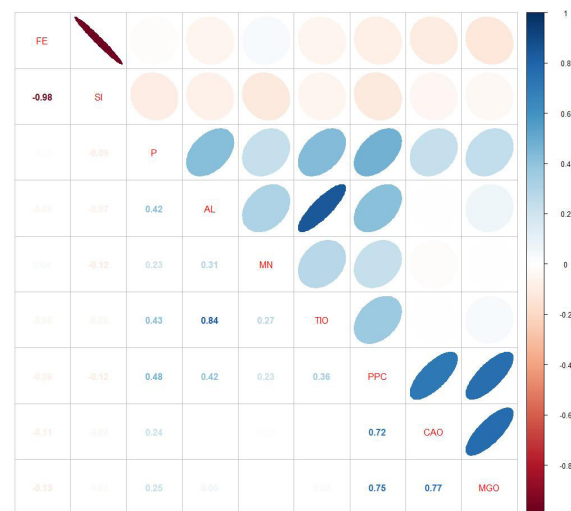
**Figura 3.** Gráfico em elipse das correlações amostrais entre as nove variáveis.

Tabela 4, indicando que essas variáveis apresentaram um grau de adequabilidade bom para a análise fatorial.

### 3.4 Método de Extração

Aplicando o método dos componentes principais, utilizando a matriz de correlação amostral, é possível determinar o menor número de fatores que expliquem a maior parte da variância do banco de dados original. A análise fatorial por componentes principais permite que as cargas fatoriais de um particular fator não se alteram quando acrescentamos novos fatores à análise. Desta forma, neste trabalho, os fatores foram extraídos aplicando o método dos Componentes Principais.

**Tabela 3.** Valores de MSA do banco de dados original

MSA	Fe	Si	P	Al	Mn	TiO	CaO	MgO	PPC
	0,19	0,19	0,79	0,27	0,08	0,85	0,37	0,31	0,37

**Tabela 4.** Valores de MSA do banco de dados após exclusão do Fe e Si

MSA	P	Al	Mn	TiO	CaO	MgO	PPC
	0,9	0,62	0,76	0,65	0,75	0,75	0,72

**Tabela 5.** Autovalores, Percentual de variância e da variância acumulada para os sete fatores

Fator	Autovalor	% Variância	% Acumulada da variância
1	3,0619642	43,742346	43,74235
2	1,9513607	27,876582	71,61893
3	0,8350473	11,929247	83,54817
4	0,6173208	8,818869	92,36704
5	0,2321499	3,316427	95,68347
6	0,1730245	2,471778	98,15525
7	0,1291326	1,844751	100,00000

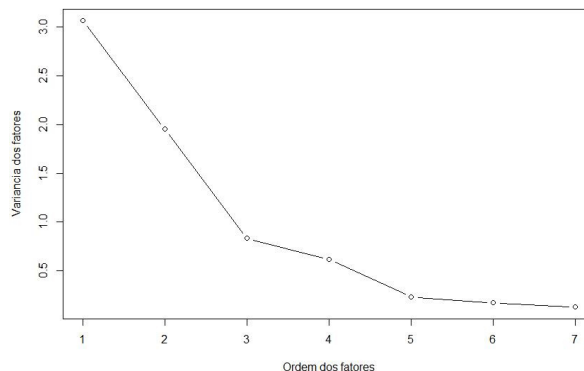
### 3.5 Número de Fatores

O número de fatores foi determinado utilizando o gráfico *scree plot*. Esta técnica considera como componentes principais os fatores anteriores ao ponto de inflexão da curva. Analisando o gráfico da Figura 4, foram identificados três fatores.

Na Tabela 5, foram apresentados os resultados dos autovalores, a porcentagem de variância explicada por cada variável e o percentual de variância acumulada pelas mesmas. Os três fatores definidos explicaram 83,5% da variação total dos dados originais.

### 3.6 Comunalidades

De acordo com Hair et al. [10], as variáveis originais podem ser definidas como uma combinação linear dos fatores comuns que explicam a parcela da variância de cada variável (comunalidade), somado a um desvio que é parcela da variância total não explicada por estes fatores (especificidade). Estes valores variam de 0 a 1, sendo que comunalidade próximas de 1 indicam que todas as variâncias são explicadas pelos fatores comuns. Analisando os valores da Tabela 6, observou-se que, com exceção da variável P (fósforo), todos os valores de comunalidade foram acima de 85%. Isto significa que mais de 85% da variabilidade das respectivas variáveis originais são explicadas pelos três fatores comuns. A permanência do P na base de dados, é explicada pela relação que esta variável tem com fenômenos intempéricos, fenômeno associado ao fator 2.

**Figura 4.** Gráfico *scree plot* dos fatores encontrados.

### 3.7 Matriz de Cargas Fatoriais e Rotação da Análise Fatorial

As cargas fatoriais representam a relação de cada uma das variáveis com seus respectivos fatores. Na análise fatorial, para facilitar a interpretação desses fatores, é comum realizar-se a rotação dos fatores. Neste trabalho, foi realizada a rotação ortogonal VARIMAX, que é um método que minimiza o número de variáveis, com altas cargas sobre o determinado fator. Analisando as cargas fatoriais das variáveis após a rotação (Tabela 7), observou-se que as variáveis PPC, CaO e MgO possuem cargas maiores, em valor absoluto, no fator 1 e que as variáveis Al, TiO têm cargas mais altas no fator 2. O manganês (Mn) está relacionado com o fator 3. Assim, o fator 1 foi denominado “presença de carbonato”, o fator 2 “intemperismo químico e/ou intrusões ígneas” e o fator 3 “manganês”.

**Tabela 6.** Comunalidades e especificidades das variáveis analisadas

	P	Al	Mn	TiO <sub>2</sub>	CaO	MgO	PPC
<b>Comunalidades</b>	0,49	0,87	0,99	0,88	0,86	0,86	0,89
<b>Especificidades</b>	0,51	0,122	0,01	0,12	0,14	0,14	0,11

**Tabela 7.** Cargas fatoriais estimadas para cada fator antes e depois da rotação Varimax

	SEM ROTAÇÃO			ROTAÇÃO VARIMAX		
	Fator 1	Fator 2	Fator 3	Fator 1	Fator 2	Fator 3
<b>P</b>	-0,6672135	<b>-0,2121502</b>	0,036561909	-0,33685055	<b>-0,5894319</b>	-0,1749858
<b>Al</b>	-0,6454498	<b>-0,6429837</b>	0,218117414	-0,01962165	<b>-0,9289927</b>	-0,1191477
<b>Mn</b>	-0,3524252	-0,3876516	<b>-0,84685084</b>	-0,01224156	-0,1886812	<b>-0,9776928</b>
<b>TiO</b>	-0,6192435	<b>-0,6546853</b>	0,260474589	0,008603346	<b>-0,9348467</b>	-0,0768758
<b>PPC</b>	<b>-0,9047324</b>	0,2673764	-0,02991671	<b>-0,84085663</b>	-0,3984082	-0,1586091
<b>CaO</b>	<b>-0,6417212</b>	0,6675760	-0,01187093	<b>-0,92498882</b>	0,03493348	0,02793035
<b>MgO</b>	<b>-0,6798324</b>	0,6299924	0,009797141	<b>-0,92616729</b>	-0,0227853	0,02921278

Observou-se que o fator 1, “presença de carbonato”, tornou as variáveis CaO, MgO e PPC fortemente relacionadas da seguinte forma: elevados teores de carbonato induzem a elevados valores de perda por calcinação (PPC) das amostras, resultando também em elevados teores de CaO e MgO, que são os óxidos resultantes da calcinação dos carbonatos de cálcio e de magnésio. Sugere-se a presença do mineral dolomita na paragénese mineral.

O fator 2, “intemperismo químico e/ou intrusões ígneas”, pode estar relacionado à dois processos geológicos: o primeiro seria a relação da presença de argilo-minerais ao intemperismo químico, ou alteração supergênica. O argilo-mineral mais comum em depósito de ferro é a caulinita, que é acompanhada por vários óxidos e hidróxidos de ferro, alumínio, titânio e fósforo; o segundo processo geológico, que explica a relação entre as variáveis Al e TiO<sub>2</sub>, seria intrusões de rochas meta-diabásicas contendo minerais ricos em alumina e titânio no depósito analisado.

O fator 3, “manganês”, está relacionado a concentrações residuais supergênicas de óxidos de manganês, que pode ocorrer associado às formações ferríferas bandadas e às rochas encaixantes do depósito analisado.

#### 4 CONCLUSÕES

A partir de uma análise estatística descritiva dos dados brutos do banco de dados estudado, foi possível concluir que os teores das variáveis ferro e sílica apresentaram elevada variabilidade, com valores variando de 15 a 69% e de 0,01 a 64%, respectivamente. Os demais compostos apresentaram teores muito baixos, em faixas correspondentes a de contaminantes, e a variabilidade desses dados foi baixa.

A partir da realização da análise fatorial, foram identificados 3 fatores responsáveis por explicar 83% da variabilidade das 9 variáveis estudadas. Os três fatores determinados foram identificados como fator 1 - presença de carbonato, fator 2- intemperismo químico e/ou intrusões ígneas e fator 3 - manganês.

O fator 1, cujas cargas fatoriais foram superiores a 0,8 (após a rotação ortogonal Varimax), permitiu relacionar as variáveis CaO, MgO e PPC à presença de carbonatos de cálcio e de magnésio, cuja presença em teores elevados, contribui para uma maior perda por calcinação (PPC) e também maiores teores dos óxidos de cálcio e de magnésio.

As variáveis Al e TiO<sub>2</sub> apresentaram cargas elevadas no fator 2. A relação destas variáveis pode ser explicada por dois processos geológicos: intemperismo químico e intrusões ígneas.

#### REFERÊNCIAS

- 1 Celino JJ, Rangel PA. Análise estatística multivariada aplicada a prospecção geoquímica de metais traços em sedimentos de corrente no semiárido do estado da Bahia. *Geochimica Brasiliensis*. 2007;21(2):193-211.
- 2 Macedo SG, Verdinelli MA, Stuker H. Modelo estatístico de análise na avaliação institucional: apresentando um recorte do desempenho docente. In: *Anais do 2º Colóquio Internacional sobre Gestão Universitária na América do Sul*; 2003 maio 7-9; Buenos Aires, Argentina. Buenos Aires: INPEAU. p. 1-17.
- 3 Ferreira MACM. Análise multivariada de dados geoquímicos aplicada à exploração mineral de ouro: estudo de caso no distrito aurífero de Almas, TO, Brasil [dissertação]. Brasília: Instituto de Geociências, Universidade de Brasília; 2015.

- 4 Silva NCN, Ferreira WL, Cirillo MA, Scallon JD. O uso da análise fatorial na descrição e identificação dos perfis característicos de municípios de Minas Gerais. *Revista Brasileira de Biometria*. 2014;32(2):201-215.
- 5 Lírio GSW. Métodos multivariados: uma metodologia para avaliar a satisfação dos clientes da RBS-TV na região noroeste do RS [dissertação]. Santa Maria: Programa de Pós-graduação em Engenharia de Produção, Universidade Federal de Santa Maria; 2004.
- 6 Contessotto R. Classificação de domínios mineralógicos e caracterização de minérios de níquel da mina de Santa Rita - BA, em apoio à geometalurgia [dissertação]. São Paulo: Escola Politécnica, Universidade de São Paulo; 2017.
- 7 Sousa DVM. Estudo geoquímico-mineral das formações ferríferas bandadas do Sinclinal Gandarela, Quadrilátero Ferrífero (MG) [dissertação]. Ouro Preto: Programa de Pós-graduação em Evolução Crustal e Recursos Naturais, Departamento de Geologia, Escola de Minas, Universidade Federal de Ouro Preto; 2016.
- 8 Projeto R [página da internet]. R: software livre. Versão 3.3.1 [acesso em 25 de mar. 2017]. Disponível em: <https://www.r-project.org>
- 9 Oliveira EC, Carvalho AA Fo, Firetti R. Identificação e elaboração de variáveis estatísticas através de técnicas multivariadas no estudo da produção de grãos e fibras, na região administrativa de Presidente Prudente-SP. *Revista Semina. Ciências Exatas e Tecnológicas*. 2014;35(2):159-180.
- 10 Hair JF, Anderson RE, Tatham RL, Black WC. *Multivariate data analysis*. 4th ed. New Jersey: Prentice Hall; 1998.

Recebido em: 24 Maio 2017

Aceito em: 22 Ago. 2017